On the Construction of Splitting Methods by Stabilizing Corrections with Runge-Kutta Pairs

Willem Hundsdorfer*

Abstract

In this technical note a general procedure is described to construct internally consistent splitting methods for the numerical solution of differential equations, starting from matching pairs of explicit and diagonally implicit Runge-Kutta methods. The procedure will be applied to suitable second-order pairs, and we will consider methods with or without a mass conserving finishing stage. For these splitting methods, the linear stability properties are studied and numerical test results are presented.

2000 Mathematics Subject Classification: 65L06, 65M06, 65M20.
Keywords and Phrases: splitting methods, stability

1 Introduction

In this note we will discuss a class of splitting methods for solving initial value problems for ordinary differential equations (ODEs)

$$u'(t) = F(t, u(t)) \,, \qquad u(0) = u_0 \,, \tag{1.1}$$

with given $u_0 \in \mathbb{R}^M$, $F : \mathbb{R} \times \mathbb{R}^M \to \mathbb{R}^M$ and dimension $M \geq 1$. For many practical problems there is a natural decomposition

$$F(t, u) \;=\; F_0(t, u) + F_1(t, u) + \cdots + F_s(t, u) \tag{1.2}$$

in which the separate component functions F_j are more simple than the whole F, and where F_0 is a non-stiff or mildly stiff term that can be treated explicitly in a time stepping method. For such problems we will study a class of stabilizing correction splitting methods, where explicit predictions are followed by corrections that are implicit in one of the F_j terms, $j = 1, 2, \ldots, s$. The methods will be constructed such that all intermediate stages yield consistent approximations to the exact solution.

1.1 Stabilizing corrections: general procedure

Consider a pair of Runge-Kutta methods, consisting of a diagonally implicit method with coefficients a_{ik} $(k \leq i)$, and an explicit method with coefficients $\hat{a}_{ik}$ $(k < i)$, and assume these two methods have the same abscissae $c_i = \sum_{k \leq i} a_{ik} = \sum_{k<i} \hat{a}_{ik}$. If the methods are applied to (1.1), with known $u_n \approx u(t_n)$, $t_n = n\Delta t$, the i-th stage of the implicit method reads

$$y_i \;=\; u_n \;+\; \Delta t \sum_{k=1}^{i} a_{ik} F(t_n + c_k \Delta t, y_k) \,, \tag{1.3a}$$

*CWI, Science Park 123, Amsterdam, The Netherlands. E-mail: willem.hundsdorfer@cwi.nl

and for the explicit method it reads

$$y_i \;=\; u_n \;+\; \Delta t \sum_{k=1}^{i-1} \hat{a}_{ik} F(t_n + c_k \Delta t, y_k)\,. \tag{1.3b}$$

We will combine these methods for problems with decomposition (1.2) by using the explicit formula as a predictor, followed by correction steps for the implicit terms. The general procedure is:

$$\left\{ \begin{array}{rcl} x_{i,0} &=& u_n \;+\; \Delta t \displaystyle\sum_{k=1}^{i-1} \hat{a}_{ik} F(t_n + c_k \Delta t, y_k)\,, \\ x_{i,j} &=& x_{i,j-1} \;+\; \Delta t \displaystyle\sum_{k=1}^{i-1} \big(a_{ik} - \hat{a}_{ik}\big) F_j(t_n + c_k \Delta t, y_k) \\ && \;+\; \Delta t\, a_{ii} F_j(t_n + c_i \Delta t, x_{i,j}) \qquad\qquad (j = 1, 2, \ldots, s)\,, \\ y_i &=& x_{i,s}\,. \end{array} \right. \tag{1.4}$$

The implicit stages, where the $x_{i,j}$ are computed, mainly serve to stabilize the process, allowing the F_j terms to be stiff. Since the two Runge-Kutta methods have the same abscissae c_k, the vectors $x_{i,0}, x_{i,1}, \ldots, x_{i,s}$ will all be consistent approximations to $u(t_n + c_i \Delta t)$. Following the terminology of [16], we will call (1.4) a stabilizing correction procedure.

The best known method of this type is obtained by combining the explicit Euler method with the implicit trapezoidal rule or the implicit Euler method. This method is known as the Douglas method because of the close relation to ADI methods developed by J. Douglas Jr. and co-workers [4, 5] for multi-dimensional parabolic problems with dimension splitting, cf. also [13, p. 373]. A class of methods with two stabilizing correction stages of the form (1.4) has been derived in [12]. In the present paper we will consider a more general approach, starting with Runge-Kutta pairs of order two with two or three stages.

1.2 Outline of the paper

In Section 2 we will derive stabilizing correction schemes based on suitable pairs of Runge-Kutta methods of order two. After the stabilizing correction stages, a finishing stage can be appended to guarantee the preservation of linear invariants, for example mass conservation.

The stability properties of the methods are examined in Section 3 for scalar linear test equations. It will be seen that the methods with the appended finishing stage become unstable in general for stiff problems with $s \geq 2$. In Section 4 some numerical test results are presented for a 2D reaction-diffusion problem, where we will consider $s = 1$ (splitting of reaction and diffusion) as well as $s = 2$ (with dimension splitting). The final Section 5 contains remarks on generalizations and conclusions.

2 Stabilizing correction methods of order two

2.1 Implicit and explicit Runge-Kutta pairs of order two

As a starting point for a stabilizing correction method, one needs a suitable pair of implicit and explicit methods. Here we consider pairs of second-order Runge-Kutta methods. The implicit method is taken to be diagonally implicit and stiffly accurate,

with three stages and abscissae $0, \kappa, 1$. For the explicit method we take a two-stage method with the same κ as abscissa. This pair can be represented in tableau form as

$$\begin{array}{c|ccc} 0 & 0 & & \\ \kappa & a_{21} & \theta & \\ 1 & b_1 & b_2 & \theta \\ \hline & b_1 & b_2 & \theta \end{array} \qquad \begin{array}{c|ccc} 0 & & & \\ \kappa & \hat{a}_{21} & & \\ 1 & \hat{b}_1 & \hat{b}_2 & \\ \hline & \hat{b}_1 & \hat{b}_2 & 0 \end{array} \,, \tag{2.1}$$

where the explicit method is written with a reducible extra stage to make it more similar to the implicit method. This stage will not be used in computations. Further it will be assumed that

$$\kappa \;=\; \hat{a}_{21} \;=\; a_{21} + \theta \,, \tag{2.2a}$$

so that the coefficients match the abscissae. Then the conditions for order two are

$$b_1 + b_2 + \theta = 1 \,, \quad b_2\kappa + \theta = \tfrac{1}{2} \,, \quad \hat{b}_1 + \hat{b}_2 = 1 \,, \quad \hat{b}_2\kappa = \tfrac{1}{2} \,. \tag{2.2b}$$

This leaves us with two free parameters.

As an alternative we will also consider an augmented explicit method where the finishing stage of the implicit method is copied, giving

$$\begin{array}{c|ccc} 0 & 0 & & \\ \kappa & a_{21} & \theta & \\ 1 & b_1 & b_2 & \theta \\ \hline & b_1 & b_2 & \theta \end{array} \qquad \begin{array}{c|ccc} 0 & & & \\ \kappa & \hat{a}_{21} & & \\ 1 & \hat{a}_{31} & \hat{a}_{32} & \\ \hline & b_1 & b_2 & \theta \end{array} \,. \tag{2.3}$$

Together with the matching conditions

$$\kappa \;=\; \hat{a}_{21} \;=\; a_{21} + \theta \,, \qquad \hat{a}_{31} + \hat{a}_{31} = 1 \,, \tag{2.4a}$$

we will impose order two, leading to the conditions

$$b_1 + b_2 + \theta = 1 \,, \quad b_2\kappa + \theta = \tfrac{1}{2} \,, \quad \hat{a}_{31} + \hat{a}_{32} = 1 \,. \tag{2.4b}$$

This gives three degrees of freedom in the parameters.

2.2 The stabilizing correction methods

In the stabilizing correction stages (1.4) the difference between the coefficients appear. For the second-order pair (2.1) we have $a_{21} - \hat{a}_{21} = -\theta$, $b_1 - \hat{b}_1 = \theta/\kappa - \theta$ and $b_2 - \hat{b}_2 = -\theta/\kappa$. This leads to the following stabilizing correction method:

$$\left\{ \begin{array}{rcll} v_0 &=& u_n + \kappa\,\Delta t\, F(t_n, u_n) \,, & \\ v_j &=& v_{j-1} + \theta\Delta t\big(F_j(t_{n+\kappa}, v_j) - F_j(t_n, u_n)\big) & (j = 1, 2, \ldots, s) \,, \\ w_0 &=& u_n + \hat{b}_1\Delta t\, F(t_n, u_n) + \hat{b}_2\Delta t\, F(t_{n+\kappa}, v_s) \,, & \\ w_j &=& w_{j-1} + \theta\Delta t\big(F_j(t_{n+1}, w_j) - \big(1 - \frac{1}{\kappa}\big)F_j(t_n, u_n) - \frac{1}{\kappa}F_j(t_{n+\kappa}, v_s)\big) & \\ & & & (j = 1, 2, \ldots, s) \,, \\ u_{n+1} &=& w_s \,. & \end{array} \right. \tag{2.5}$$

Here the $v_i \approx u(t_{n+\kappa})$ and $w_i \approx u(t_{n+1})$, $i = 0, 1, \ldots, s$, are internal vectors and the intermediate time level is $t_{n+\kappa} = t_n + \kappa\Delta t$. Using θ and κ as free parameters we have $\hat{b}_1 = 1 - 1/(2\kappa)$ and $\hat{b}_2 = 1/(2\kappa)$. To distinguish this method from a variant with an

extra finishing stage, to be introduced next, we will often refer to (2.5) as a stabilizing correction method of *type-A*.

For the implicit and explicit pair (2.3), the final stage of the methods can be appended to the stabilizing correction stages. This leads to the following method:

$$\left\{\begin{array}{rcll} v_0 &=& u_n + \kappa\,\Delta t\,F(t_n,u_n)\,, & \\ v_j &=& v_{j-1} + \theta\Delta t\big(F_j(t_{n+\kappa},v_j) - F_j(t_n,u_n)\big) & (j=1,2,\ldots,s)\,, \\ w_0 &=& u_n + \hat{a}_{31}\Delta t\,F(t_n,u_n) + \hat{a}_{32}\Delta t\,F(t_{n+\kappa},v_s)\,, & \\ w_j &=& w_{j-1} + \theta\Delta t\big(F_j(t_{n+1},w_j) - \mu_1 F_j(t_n,u_n) - \mu_2 F_j(t_{n+\kappa},v_s)\big) & \\ & & & (j=1,2,\ldots,s)\,, \\ u_{n+1} &=& u_n + b_1F(t_n,u_n) + b_2\Delta tF(t_{n+\kappa},v_s) + \theta F(t_n,w_s)\,, & \end{array}\right. \tag{2.6}$$

where

$$\mu_1 \;=\; \frac{1}{\theta}(\hat{a}_{31}-b_1)\,, \qquad \mu_2 \;=\; \frac{1}{\theta}(\hat{a}_{32}-b_2)\,. \tag{2.7}$$

Note that if $\hat{a}_{31} = \hat{b}_1$, $\hat{a}_{32} = \hat{b}_2$, then the formulas for the internal vectors are the same as in (2.5), but for this variant (2.6) there is still a finishing stage with the whole function F. We will refer to (2.6) as a *type-B* stabilizing correction method.

It will be examined how this finishing stage influences the local accuracy and stability of the schemes. Here we can already mention two important properties of the above methods: internal consistency for type-A and type-B methods, and mass conservation for type-B methods.

Internal consistency: The internal vectors v_j, w_j are consistent approximations to the exact solution at time levels $t_{n+\kappa}$ and t_{n+1}, respectively. This important property guarantees that steady state solutions of autonomous problems are returned without any error. Methods with this property are called *well-balanced* in shallow water applications. Many other splitting methods, based on Lie splitting or Strang splitting, do not share this property.

Mass conservation: Consider a type-B method, and suppose the ODE system (1.1) is such that $h^T u(t)$ is constant for any solution, with a weight vector $h \in \mathbb{R}^m$, so the system has a linear invariant. This is equivalent with $h^T F(t,v) = 0$ for all $v \in \mathbb{R}^m$. For a splitting of F it may happen that $h^T F_j(t,v) \neq 0$ for some j. In that case, due to finishing stage with the whole function F, the type-B method (2.6) will still preserve the linear invariant, $h^T u_n = h^T u_{n-1}$, but this property may be lost with the type-A method (2.5). In particular for mass conservation it can be important to maintain linear invariants in a numerical method.

If $s = 1$ the above stabilizing correction methods reduce to implicit-explicit (IMEX) methods. This special case often occurs in practice and it will be closely examined in this paper. Further it is noted that the same approach with stabilizing corrections could be used for Runge-Kutta methods with more stages. This will lead, however, to classes of methods with many free coefficients, from which it will not be easy to chose (embarrassment of riches). Finally we mention that the two classes of splitting methods (2.5) and (2.6) do not include the modified Craig-Sneyd methods constructed by in 't Hout and Welfert [10]. These methods contain explicit stages where the coefficients for F_0 are different from those for the implicit component functions F_j. Comments on this are given in the last section of this note.

2.3 Examples

In these notes, we will focus on classes of methods that are obtained by specific choices of either the explicit or the implicit methods. In the first two examples we will have a connection between the explicit methods in (2.1) and (2.3) by taking

$$\hat{a}_{31} = \hat{b}_1\,, \qquad \hat{a}_{32} = \hat{b}_2\,. \tag{2.8}$$

Example 2.1 One of the best known explicit two-stage Runge-Kutta method is the explicit trapezoidal rule, also known as the modified Euler method, with coefficients

$$\kappa \;=\; 1\,, \qquad \hat{b}_1 \;=\; \hat{b}_2 \;=\; \tfrac{1}{2}\,. \tag{2.9a}$$

From the conditions for order two, it follows that the coefficients of the corresponding implicit method are

$$a_{21} \;=\; 1-\theta\,, \quad b_1 \;=\; \tfrac{1}{2}\,, \quad b_2 \;=\; \tfrac{1}{2}-\theta\,, \tag{2.9b}$$

where we will use the diagonal coefficient θ as free parameter. The resulting stabilizing correction method (2.5) was introduced in [12]. In this note it will be examined whether more favourable methods can be found within the classes (2.5) or (2.6).

To identify interesting methods, stability regions will play an important role. The explicit trapezoidal rule has the familiar stability function

$$r_{\mathrm{expl},A}(z) \;=\; 1 + z + \tfrac{1}{2}z^2\,. \tag{2.10}$$

For the explicit method with the extra stage in (2.3) we consider (2.8), that is, $\hat{a}_{31} = \hat{a}_{32} = \frac{1}{2}$, which yields the stability function

$$r_{\mathrm{expl},B}(z) \;=\; 1 + z + \tfrac{1}{2}z^2 + \tfrac{1}{2}\theta z^3\,. \tag{2.11}$$

The stability function of the implicit method is given by

$$r_{\mathrm{impl}}(z) \;=\; \frac{1 + (1-2\theta)z + (\frac{1}{2} - 2\theta + \theta^2)z^2}{(1-\theta z)^2}\,. \tag{2.12}$$

This implicit method is A-stable if $\theta \geq \frac{1}{4}$ and it is L-stable for the parameter values $\theta = 1 \pm \frac{1}{2}\sqrt{2}$. The stability regions $\mathcal{E}$ of the explicit methods (2.10) and (2.11) are presented in Figure 1 for three parameter values. The dotted lines are contour lines for $|r_{\mathrm{expl},A}(z)|$ and $|r_{\mathrm{expl},B}(z)|$ at the levels $0.1, \ldots, 0.9$. Since the implicit methods are

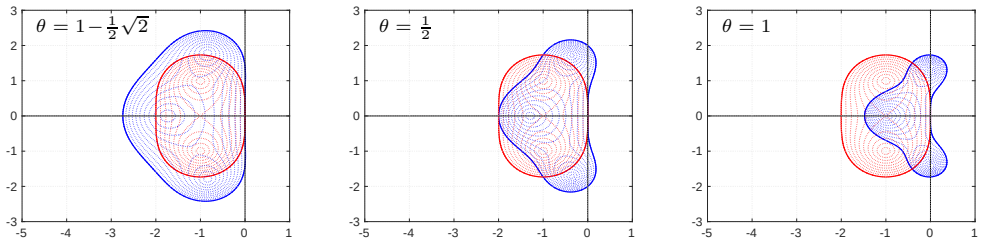

Figure 1: Stability regions of the explicit trapezoidal rule (2.10) indicated with red lines, and the augmented explicit method (2.11) with blue lines. Parameter value $\theta = 1 - \frac{1}{2}\sqrt{2}$ [left panel], $\theta = \frac{1}{2}$ [middle panel] and $\theta = 1$ [right panel].

A-stable, the corresponding plots for these methods are less interesting, and therefore these are not shown.

Further we note that linearization of the diagonally implicit method with coefficients (2.9b) leads to a well-known Rosenbrock-type method, or W-method [8], which is such that its order remains two with arbitrary approximations for the Jacobian matrix $A \approx \frac{\partial}{\partial u} F(t_n, u_n)$. For such methods one can apply an approximate matrix factorization where the matrix $I - \theta \Delta t A$ in the Rosenbrock method is replaced by a product $\prod_{j=1}^{s} (I - \theta \Delta t A_j)$ with $A_j \approx \frac{\partial}{\partial u} F_j(t_n, u_n)$. The resulting method can then be viewed as a linearized version of the splitting method (2.5); see for instance [13, p.400]. ◇

Example 2.2 The choice $a_{21} = \theta$ gives a popular diagonally implicit method where the first nontrivial stage consists of a scaled step with the implicit trapezoidal rule; see e.g. [13, p. 144]. Using θ as free parameter, this implicit method has order two if

$$a_{21} \,=\, \theta\,, \quad \kappa = 2\theta\,, \quad b_1 \,=\, \tfrac{3}{2} - \theta - \tfrac{1}{4\theta}\,, \quad b_2 \,=\, -\tfrac{1}{2} + \tfrac{1}{4\theta}\,. \tag{2.13a}$$

Then, requiring order two for the two-stage explicit method gives the coefficients

$$\hat{b}_1 \,=\, 1 - \tfrac{1}{4\theta}\,, \qquad \hat{b}_2 \,=\, \tfrac{1}{4\theta}\,. \tag{2.13b}$$

The implicit method has again stability function (2.12), because the coefficients of $r_{\rm impl}$ are determined by the order two conditions. Likewise, the stability function of the two-stage explicit method (2.1) is given by (2.10). For the augmented three-stage explicit method we consider (2.8), that is, $\hat{a}_{31} = 1 - \frac{1}{4\theta}$, $\hat{a}_{32} = \frac{1}{4\theta}$. It is seen by some calculations that this gives again the stability function (2.11).

So, even though the methods are different if $\theta \neq \frac{1}{2}$, the stability functions are the same as in the previous example and the stability regions are as in Figure 1 for $\theta = \frac{1}{2}$ and $\theta = 1 - \frac{1}{2}\sqrt{2}$. We will see in the next section that this is a common property of the methods based on second-order pairs (2.1) as well as for the related methods (2.3) with coefficients specified by (2.8). ◇

Example 2.3 The implicit method of the previous example was used to construct implicit-explicit (IMEX) methods in [6, 7]. In these references, the implicit method (2.13a) was combined with the three-stage explicit methods (2.3) with parameters

$$\theta = 1 - \tfrac{1}{2}\sqrt{2}\,, \qquad \hat{a}_{31} = \tfrac{1}{2} - \omega\,, \qquad \hat{a}_{32} = \tfrac{1}{2} + \omega\,. \tag{2.14a}$$

The values $\omega = 0$ and $\omega = \frac{1}{3}\sqrt{2}$ correspond to the choices made in [6] and [7], respectively. The other parameters are

$$a_{21} = 1 - \tfrac{1}{2}\sqrt{2}\,, \qquad \kappa = 2 - \sqrt{2}\,, \qquad b_1 = \tfrac{1}{4}\sqrt{2}\,, \qquad b_2 = -\tfrac{1}{4}\sqrt{2}\,, \tag{2.14b}$$

in agreement with (2.13a). Plots of the stability regions of these explicit method can be found in Figure 5.

The IMEX methods of [6, 7] can be obtained from (2.6) with $s = 1$. We will also study these methods for the case $s > 1$, but it will be seen that stability then becomes problematic. Finally we note that the type-A methods with $\hat{b}_j = \hat{a}_{3j}$, $j = 1, 2$, are less interesting in this example since the two-stage explicit methods are then only of order one, except for the choice $\omega = \frac{1}{4}\sqrt{2}$. ◇

For the implicit method (2.13a), used in the above examples, the order will be three if $\theta = \frac{1}{2} \pm \frac{1}{6}\sqrt{3}$. This can be interesting for IMEX applications where the dominant error is caused by the implicit term.

3 Linear stability properties

3.1 Stability functions

Stability and convergence will be analyzed for linear systems of differential equations where $F_j(t,u) = A_j u + g_j(t)$. As a first step we consider the test equation

$$u'(t) \;=\; (\lambda_0 + \lambda_1 + \ldots + \lambda_s)u(t)\,. \tag{3.1}$$

Let $z_j = \Delta t \lambda_j$. The stabilizing correction methods will then give a relation $u_{n+1} = r(z_0, z_1, \ldots, z_s)u_n$. Similar as for Runge-Kutta methods, such a function r will be called the stability function. It will be seen that the variables $z_0, z_1, \ldots, z_s$ appear in the stability functions only in the combinations

$$z = z_0 + z_1 + \ldots + z_s\,, \qquad \varpi \;=\; \prod_{j=1}^{s}(1-\theta z_j)\,. \tag{3.2}$$

To have a clear distinction between the stabilizing correction methods of type-A and type-B, we will use sub-indices A or B for these stability functions. First we will derive a relation $w_s = q(z_0, z_1, \ldots, z_s)u_n$. For the type-A method (2.5) this will already provide the stability function.

Consider the stabilizing correction methods (2.5) and (2.6). To get the same notation we set $\mu_1 = 1 - 1/\kappa$, $\mu_2 = 1/\kappa$ for (2.5), and $\hat{b}_1 = \hat{a}_{31}$, $\hat{b}_2 = \hat{a}_{32}$ for (2.6). Then, application to the test equation gives

$$\begin{aligned}
v_0 \;&=\; u_n + \kappa\, z\, u_n\,,\\
v_j \;&=\; v_{j-1} + \theta z_j (v_j - u_n) \qquad (j = 1, \ldots, s)\,,\\
w_0 \;&=\; u_n + \hat{b}_1 z\, u_n + \hat{b}_2 z\, v_s\,,\\
w_j \;&=\; w_{j-1} + \theta z_j (w_j - \mu_1 u_n - \mu_2 v_s) \qquad (j = 1, \ldots, s)\,.
\end{aligned}$$

To derive a suitable expression for the stability function, it is convenient to introduce $\bar{v}_j = v_j - u_n$ and $\bar{w}_j = w_j - \mu_1 u_n - \mu_2 v_s$. Then

$$\begin{aligned}
\bar{v}_0 \;&=\; \kappa\, z\, u_n\,, & \bar{v}_j \;&=\; \tfrac{1}{1-\theta z_j}\,\bar{v}_{j-1}\,,\\
\bar{w}_0 \;&=\; z\, u_n - (\mu_2 - \hat{b}_2 z)\bar{v}_s\,, & \bar{w}_j \;&=\; \tfrac{1}{1-\theta z_j}\,\bar{w}_{j-1}\,.
\end{aligned}$$

Hence $\bar{v}_s = \bar{v}_0/\varpi$ and $\bar{w}_s = \bar{w}_0/\varpi$. Combining these relations and using $b_1 + b_2 + \theta = 1$ gives $\bar{w}_s = \frac{1}{\varpi} z u_n - \frac{1}{\varpi^2}(\mu_2 - \hat{b}_2 z)\kappa z u_n$, which finally leads to $w_s = q(z_0, \ldots, z_s)u_n$ with

$$q(z_0, z_1, \ldots, z_s) \;=\; 1 + (1 + \mu_2\kappa)\frac{z}{\varpi} - \mu_2\kappa\frac{z}{\varpi^2} + \hat{b}_2\kappa\frac{z^2}{\varpi^2}\,. \tag{3.3}$$

For method (2.5) the stability function is $r_A = q$. By use of the conditions (2.2b) for order two, the following result is obtained:

Proposition 3.1 *For all pairs (2.1) with (2.2a), (2.2b), the type-A method (2.5) has stability function*

$$r_A(z_0, z_1, \ldots, z_s) \;=\; 1 \;+\; 2\frac{z}{\varpi} \;-\; \frac{z}{\varpi^2} \;+\; \frac{1}{2}\,\frac{z^2}{\varpi^2}\,. \tag{3.4}$$

This expression follows immediately from (3.3): if (2.2b) then $\kappa\mu_2 = 1$ and $\hat{b}_2\kappa = \frac{1}{2}$. Further it should be noted that the stability function does not depend on the parameter κ, and the other free parameter θ only enters through $\varpi = \prod_{j=1}^{s}(1-\theta z_j)$.

For the type-B method (2.6), with the additional finishing stage, application to the test equation gives $u_{n+1} = (1 + b_1 z)u_n + b_2 z v_s + \theta z w_s$, where we can use the above expressions for v_s and for w_s with $\hat{a}_{32} = \hat{b}_2$. After a little calculation this leads to

$$r_B(z_0, z_1, \ldots, z_s) \;=\; 1 + z + \big(\tfrac{1}{2} + \theta\mu_2\kappa\big)\frac{z^2}{\varpi} - \theta\mu_2\kappa\frac{z^2}{\varpi^2} + \theta\hat{a}_{32}\kappa\frac{z^3}{\varpi^2}\,. \tag{3.5}$$

Use of the conditions (2.4b) for order two gives the following result:

Proposition 3.2 *For all pairs (2.3) with (2.4a), (2.4b), the type-B method (2.6) has stability function*

$$r_B(z_0, z_1, \ldots, z_s) \;=\; 1 \;+\; z \;+\; \big(\tfrac{1}{2} + \nu\big)\frac{z^2}{\varpi} \;-\; \nu\frac{z^2}{\varpi^2} \;+\; \big(\tfrac{1}{2} - \theta + \nu\big)\theta\frac{z^3}{\varpi^2}\,, \tag{3.6}$$

where $\nu = \theta\kappa\mu_2$. If (2.2b) and (2.8) also hold, then

$$r_B(z_0, z_1, \ldots, z_s) \;=\; 1 \;+\; z \;+\; \big(\tfrac{1}{2} + \theta\big)\frac{z^2}{\varpi} \;-\; \theta\frac{z^2}{\varpi^2} \;+\; \tfrac{1}{2}\theta\frac{z^3}{\varpi^2}\,. \tag{3.7}$$

Formula (3.6) directly follows from $\kappa\hat{a}_{32} = \frac{1}{2} - \theta + \theta\kappa\mu_2$, which is a consequence of (2.4b). Moreover, if (2.2b) with $\hat{b}_2 = \hat{a}_{32}$ is also valid, then $\nu = \theta$, giving (3.7).

Note that in these stability functions there are now terms z^{k+1}/ϖ^k, with power in the numerator higher than in the denominator, so it not very surprising that stability is often harder to achieve for these type-B methods. This will be discussed in detail in the next sections.

3.2 Stability domains

In the following it will be assumed that the implicit arguments z_j, $j \geq 1$, are in a wedge $\mathcal{W}_\alpha$ in the left-half plane,

$$\mathcal{W}_\alpha \;=\; \{\zeta \in \mathbb{C} : \arg(-\zeta) \leq \alpha \;\text{ or }\; \zeta = 0\}$$

with angle $\alpha \in [0, \frac{1}{2}\pi]$. So $\mathcal{W}_0 = \mathbb{R}^-$ is the non-positive real axis and $\mathcal{W}_{\pi/2} = \mathbb{C}^-$ is the left-half plane. For a given stability function r and angle α, we will then study the following domains for the explicit argument z_0:

$$\mathcal{D}_\alpha \;=\; \{z_0 \in \mathbb{C} :\; |r(z_0, z_1, \ldots, z_s)| \leq 1 \;\text{ for all } z_j \in \mathcal{W}_\alpha,\, j = 1, 2, \ldots, s\}\,. \tag{3.8}$$

3.2.1 Necessary conditions for stability with $\alpha = 0$

Necessary stability conditions can be obtained by studying the limit case $z_s \to -\infty$, with the other implicit arguments $z_1, \ldots, z_{s-1}$ real and non-positive. For a method with stability function r it will be required that

$$\lim_{z_s \to -\infty} |r(z_0, z_1, \ldots, z_s)| \;\leq\; 1 \quad \text{for all } z_1, \ldots, z_{s-1} \in \mathbb{R}^-\,. \tag{3.9}$$

Further we will use the notation

$$\sigma_s \;=\; \sum_{j=0}^{s-1} z_j\,, \qquad \pi_s \;=\; \prod_{j=1}^{s-1}(1 - \theta z_j)^{-1}\,. \tag{3.10}$$

With real $z_1, \ldots, z_{s-1} \leq 0$ we have $\pi_s \in [0, 1]$, and if $s = 1$ then $\pi_s = 1$.

Proposition 3.3 *For the type-A methods (2.5), with stability function $r = r_A$ given by (3.4), the stability condition (3.9) holds iff $\theta \geq \frac{1}{4}$.*

Proof. For large z_s we have $z/\varpi = -\pi_s/\theta + \mathcal{O}(z_s^{-1})$, $z/\varpi^2 = \mathcal{O}(z_s^{-1})$. Consequently

$$\lim_{z_s \to -\infty} r_A(z_0, \ldots, z_s) \;=\; \phi_A(\pi_s) \;=\; 1 \;-\; \tfrac{2}{\theta}\pi_s \;+\; \tfrac{1}{2\theta^2}\pi_s^2 \,. \tag{3.11}$$

For this limit function ϕ_A it is easily seen that

$$\begin{aligned} \phi_A(0) = 1\,, \qquad \phi_A(1) &= \tfrac{1}{\theta^2}\big(\tfrac{1}{2} - 2\theta + \theta^2\big)\,, \\ \phi_A'(\pi_s) = 0 \;\text{ for }\; \pi_s = 2\theta\,, \;\text{ and }\quad \phi_A(2\theta) &= -1\,. \end{aligned}$$

Furthermore $|\frac{1}{2} - 2\theta + \theta^2| \leq \theta^2$ iff $\theta \geq \frac{1}{4}$, which provides the proof. [To be done more carefully and more clearly, separating the cases $s = 1$ and $s \geq 2$.] □

Observe that $\theta \geq \frac{1}{4}$ gives exactly the parameter range for which the implicit method is A-stable. So this is the only requirement to be fulfilled for stability in the limit case $z_s \to -\infty$.

As can be expected from the form of the stability functions for type-B methods, stability is more delicate for such methods. To formulate the result, we define

$$\phi_B(z_0) \;=\; \tfrac{1}{\theta^2}\big(\tfrac{1}{2} - 2\theta + \theta^2\big) + \tfrac{1}{\theta}\big(\tfrac{1}{2} - 2\theta + \nu\big) z_0 \,. \tag{3.12}$$

Proposition 3.4 *Consider the type-B methods (2.6) with stability function $r = r_B$ given by (3.6).*
(a) *If $s = 1$, then the stability condition (3.9) holds iff $|\phi_B(z_0)| \leq 1$.*
(b) *If $s \geq 2$, then the stability condition (3.9) cannot hold.*

Proof. For large $|z_s|$ we have

$$\begin{aligned} z \;&=\; z_s + \sigma_s \;=\; z_s\big(1 + \mathcal{O}(z_s^{-1})\big)\,, \\ \frac{z}{\varpi} \;&=\; -\tfrac{1}{\theta}\,\pi_s \;-\; \tfrac{1}{\theta^2}(1 + \theta\sigma_s)\pi_s z_s^{-1} \;+\; \mathcal{O}(z_s^{-2})\,, \end{aligned}$$

and therefore

$$\begin{aligned} \frac{z^2}{\varpi} \;&=\; -\tfrac{1}{\theta}\,\pi_s z_s - \tfrac{1}{\theta^2}(1 + 2\theta\sigma_s)\pi_s + \mathcal{O}(z_s^{-1})\,, \\ \frac{z^3}{\varpi^2} \;&=\; \tfrac{1}{\theta^2}\pi_s^2 z_s + \Big(\tfrac{1}{\theta^2}\sigma_s\pi_s^2 + \tfrac{2}{\theta^3}(1 + \theta\sigma_s)\Big) + \mathcal{O}(z_s^{-1})\,. \end{aligned}$$

Inserting these expansions in (3.6) we obtain

$$\begin{aligned} r_B(z_0, \ldots, z_s) \;=\; &\Big(1 - \tfrac{1}{\theta}\big(\tfrac{1}{2} + \nu\big)\pi_s + \tfrac{1}{\theta}\big(\tfrac{1}{2} - \theta + \nu\big)\pi_s^2\Big) z_s \\ &+ \Big(1 + \sigma_s - \tfrac{1}{\theta^2}\big(\tfrac{1}{2} + \nu\big)(1 + 2\theta\sigma_s)\pi_s - \tfrac{\nu}{\theta^2}\pi_s^2 \\ &\quad + \big(\tfrac{1}{2} - \theta + \nu\big)\big(\tfrac{1}{\theta}\sigma_s\pi_s^2 + \tfrac{2}{\theta^2}(1 + \theta\sigma_s)\big)\Big) + \mathcal{O}(z_s^{-1})\,. \end{aligned}$$

If $z_s \to -\infty$ then $|r_B|$ will tend to a finite limit value iff the first term on the right vanishes, that is,

$$1 - \tfrac{1}{\theta}\big(\tfrac{1}{2} + \nu\big)\pi_s + \tfrac{1}{\theta}\big(\tfrac{1}{2} - \theta + \nu\big)\pi_s^2 \;=\; 0\,. \tag{3.13}$$

If $s \geq 2$, then π_s may take on any value between 0 and 1, in which case this equality cannot be satisfied. On the other hand, if $s = 1$ we simply have $\pi_s = 1$, in which case (3.13) holds trivially, and it then also follows that $\lim_{z_1 \to -\infty} r_B(z_0, z_1) = \phi_B(z_0)$. □

To establish the connection between part (a) of this proposition and Proposition 3.3, note that $|\phi_B(0)| \leq 1$ iff $\theta \geq \frac{1}{4}$, which is the parameter range for which the implicit method is A-stable. Furthermore, it is clear from the negative result in part (b) that the type-B are not suited for problems with $s \geq 2$.

3.2.2 Stability domains $\mathcal{D}_\alpha$

We now consider the stability domains $\mathcal{D}_\alpha$ for the explicit argument z_0. Useful analytic results can be very hard to derive. The main objective of this section is the presentation and discussion of plots of these domains, obtained by taking for each z_0 a large number of points $z_1, \ldots, z_s$ on the boundary of the wedge $\mathcal{W}_\alpha$ with a given angle α. The plots will mostly be presented only for the angles $\alpha = 0$ and $\alpha = \frac{1}{2}\pi$, with comments on the stability domains for the intermediate angle $\alpha = \frac{1}{4}\pi$ given in the text.

For a given stability function r, it will be convenient in the discussion to refer to the function

$$\psi_\alpha(z_0) = \sup_{z_1, \ldots, z_s \in \mathcal{W}_\alpha} |r(z_0, z_1, \ldots, z_s)| . \tag{3.14}$$

The set $\mathcal{D}_\alpha$ then consists of those $z_0 \in \mathbb{C}$ for which $\psi_\alpha(z_0) \leq 1$. In the plots of the stability domains, also contour lines $\psi_\alpha(z_0) = c$ will be drawn, with dotted lines, for the contour levels $c = 0.1, 0.2, \ldots, 0.9$.

Methods (2.9) with $s = 1$: First we consider the methods from Example 2.1 and Example 2.2, with three values of the parameter θ. The stability functions are $r = r_A$ and $r = r_B$, as given by the equations (3.4), (3.7) for the type-A and type-B methods, respectively. The domains $\mathcal{D}_\alpha$ with angles $\alpha = 0$ and $\alpha = \frac{1}{2}\pi$ are shown in Figure 2 for the case $s = 1$.

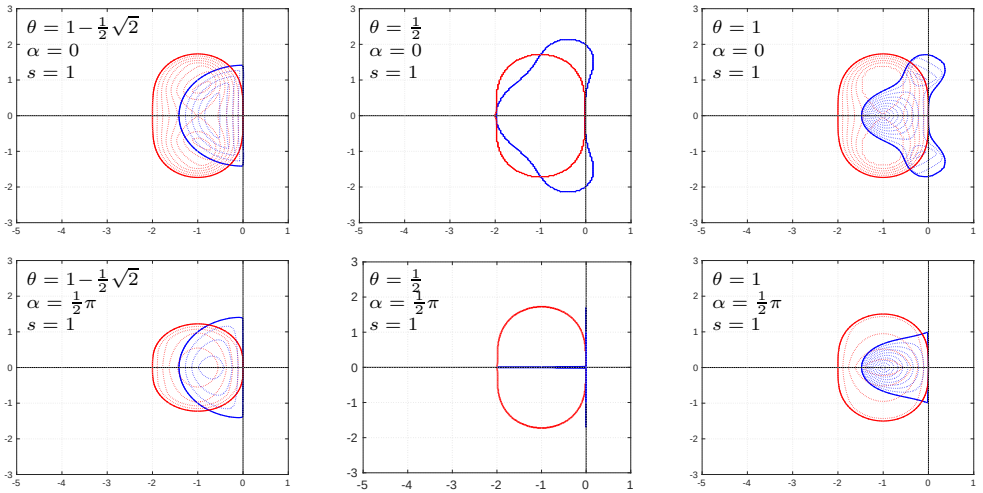

Figure 2: Stability domains for $s = 1$. Methods from Example 2.1, 2.2. Red lines for domains with $r = r_A$ [eq. (3.4)], blue for $r = r_B$ [eq. (3.7)]. From left to right: $\theta = 1 - \frac{1}{2}\sqrt{2}$ [left], $\theta = \frac{1}{2}$ [middle] and $\theta = 1$ [right]. Top row for angle $\alpha = 0$, bottom row for $\alpha = \frac{1}{2}\pi$.

These domains $\mathcal{D}_0$ and $\mathcal{D}_{\pi/2}$ can now be compared with the stability regions $\mathcal{E}$ of the explicit methods, as given in Figure 1. It is seen that the domains $\mathcal{D}_0$ are equal to $\mathcal{E}$ for the type-A methods. For the type-B methods this also holds if $\theta = \frac{1}{2}$ and $\theta = 1$, but for the smallest parameter value, $\theta = 1 - \frac{1}{2}\sqrt{2}$, the domain $\mathcal{D}_0$ is considerably smaller than $\mathcal{E}$.

The domains $\mathcal{D}_{\pi/2}$ are in general smaller than $\mathcal{E}$. In particular, for the type-B method with parameter $\theta = \frac{1}{2}$ the domain is reduced to a small set containing the segment $[-2, 0]$ of the negative real axis and a part of the imaginary axis, roughly $[-2i, 2i]$ (not well visible on the scale used in these plots). On the other hand, for corresponding type-A method with $\theta = \frac{1}{2}$, we again obtain the full stability region $\mathcal{E}$. This somewhat surprising result is quite easy to derive analytically, see [13, p. 402].

Similar plots have been made for $\alpha = \frac{1}{4}\pi$, showing that the stability domains for this angle are only slightly smaller than for $\alpha = 0$.

Finally it should be mentioned that the (dotted) contour lines $\psi_\alpha(z_0) = c$ with levels $c = 0.1, \ldots, 0.9$ are absent in the figure for the parameter $\theta = \frac{1}{2}$. With this parameter value we always have $\psi_\alpha(z_0) \geq 1$, due to the fact that $|r(z_0, z_1)| \to 1$ as $z_1 \to -\infty$. Furthermore, as a consequence, $\psi_\alpha(z_0)$ will not be differentiable at the boundary of the domain $\mathcal{D}_\alpha$, and this non-smooth behaviour causes the (Matlab) plotting routine to draw staircase-shaped lines instead of smooth curves for the boundaries of the stability domains.

Methods (2.9) with $s = 2$: Next we consider the methods from Example 2.1 and 2.2 for the case $s = 2$. For this case the stability domains are empty for the type-B methods, in agreement with Proposition 3.4. The domains $\mathcal{D}_\alpha$ for the type-A methods are presented in Figure 3 for angles $\alpha = 0, \frac{1}{2}\pi$.

With angle $\alpha = 0$ the domains still coincide with the stability region of the underlying explicit method. However, for $\alpha = \frac{1}{2}\pi$ only a very small stability domain remains for $\theta = 1$, and for the parameters $\theta = 1 - \frac{1}{2}\sqrt{2}$ and $\theta = \frac{1}{2}$ the domains are

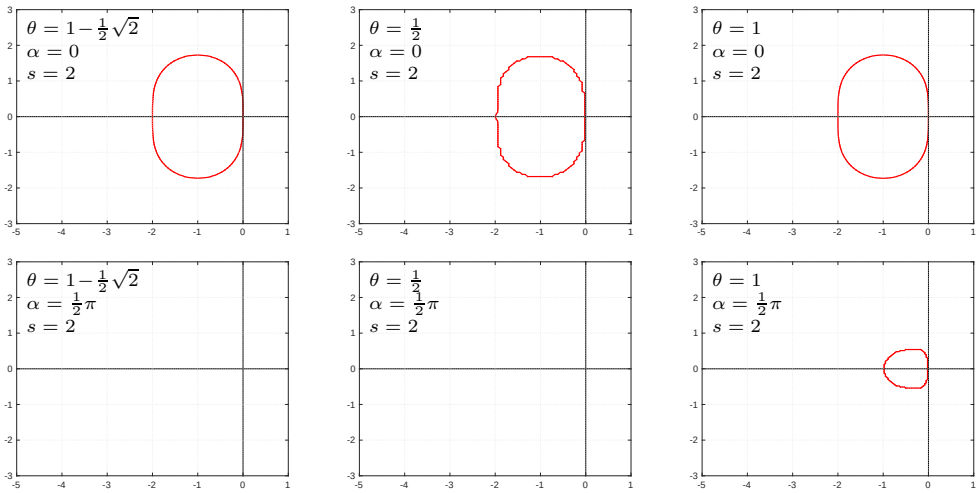

Figure 3: Stability domains for $s = 2$. Methods from Example 2.1, 2.2. Red lines for domains with $r = r_A$ [eq. (3.4)]; the stability domains for $r = r_B$ [eq. (3.7)] are empty. From left to right: $\theta = 1 - \frac{1}{2}\sqrt{2}$ [left], $\theta = \frac{1}{2}$ [middle] and $\theta = 1$ [right]. Top row for angle $\alpha = 0$, bottom row for $\alpha = \frac{1}{2}\pi$.

empty; cf. [15] where it was shown that $\mathcal{D}_{\pi/2}$ is non-empty for $s = 2$ iff $\theta \geq \frac{1}{2} + \frac{1}{6}\sqrt{3}$.

In the figure the (dotted) contour lines of $\psi_\alpha(z_0) = c$ with $c = 0.1, \ldots, 0.9$ are absent. This is due to the fact that $|r(z_0, z_1, z_2)|$ tends to 1 as $z_1, z_2 \to -\infty$, which implies that $\psi_\alpha(z_0) \geq 1$.

The domains $\mathcal{D}_\alpha$ were also examined for $\alpha = \frac{1}{4}\pi$. For that angle, the domains with parameter values $\theta = \frac{1}{2}, 1$ are much larger than for $\alpha = \frac{1}{2}\pi$, namely equal to the stability region $\mathcal{E}$ of the corresponding explicit method, but for the smallest parameter $\theta = 1 - \frac{1}{2}\sqrt{2}$ the domain is still empty for this intermediate angle. In fact, by considering $z_1 = \frac{2\theta-1}{2\theta^2} + i\varepsilon$ and $z_2 \to -\infty$ it can be shown that $0 \in \mathcal{D}_\alpha$ for some $\alpha > 0$ requires $\theta \geq \frac{1}{2}$.

Methods (2.9) with $s = 3$: For general stabilizing correction methods and $s \geq 3$ it can be shown that stability domains $\mathcal{D}_\alpha$ can only be non-empty if $\alpha \leq \frac{1}{s-1}\frac{\pi}{2}$, see [12, p. 224]. Therefore, for the case $s = 3$, the plots of the stability domains are presented in Figure 4 with angles $\alpha = 0$ and $\frac{1}{4}\pi$. Moreover, since we know that the type-B methods are not stable for $s > 1$, we only consider the type-A method with $r = r_A$ given by equation (3.4).

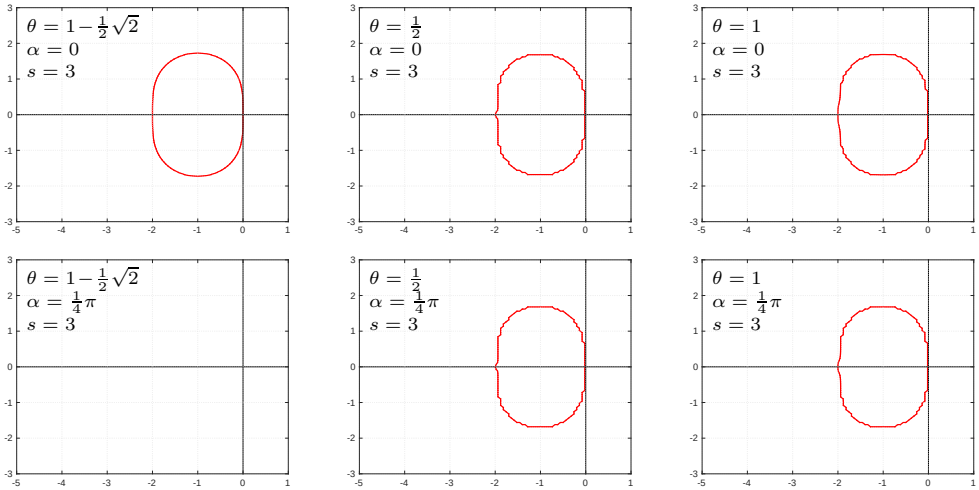

Figure 4: Stability domains for $s = 3$. Methods from Example 2.1, 2.2, with the function $r = r_A$ [eq. (3.4)]. From left to right: $\theta = 1 - \frac{1}{2}\sqrt{2}$ [left], $\theta = \frac{1}{2}$ [middle] and $\theta = 1$ [right]. Top row for angle $\alpha = 0$, bottom row for $\alpha = \frac{1}{4}\pi$.

If $s = 3$ then $\mathcal{D}_0$ equals the explicit stability region $\mathcal{E} = \{z_0 \in \mathbb{C} : |1 + z_0 + \frac{1}{2}z_0| \leq 1\}$ for all three parameters $\theta = 1 - \frac{1}{2}\sqrt{2}, \frac{1}{2}, 1$. Further it is seen that $\mathcal{D}_{\pi/4}$ is again equal to $\mathcal{E}$ for $\theta = \frac{1}{2}$, it is slightly smaller (not well visible on this scale) for $\theta = 1$, but for the parameter value $\theta = 1 - \frac{1}{2}\sqrt{2}$ the domain is now empty. These experimental findings are in agreement with those of [12, Fig. 1,2].

Methods (2.9) with $s > 3$: For the type-A methods with $s > 3$ the above mentioned angle bound $\alpha \leq \frac{1}{s-1}\frac{\pi}{2}$ restricts the eigenvalues of the (linearized) implicit terms. We do have the following result for arbitrary s with $\alpha = 0$:

Proposition 3.5 *Consider a type-A method with stability function $r = r_A$ given by (3.4), and suppose $\theta \geq \frac{1}{4}$, $s \geq 1$ and $z_0 = 0$. Then*

$$|r(0, z_1, z_2, \ldots, z_s)| \leq 1 \quad \textit{for all } z_1, \ldots, z_s \in \mathbb{R}^- . \tag{3.15}$$

Proof. Since $z_0 = 0$ and the other $z_j \leq 0$, we have $z \leq 0$ and $\varpi \geq 1 - \theta z \geq 1$. For the stability function (3.4), with dependence on $z_0, \ldots, z_s$ through z and ϖ, we can write

$$r \;=\; 1 \;+\; \frac{z}{\varpi^2}\big(2\varpi - 1\big) \;+\; \frac{1}{2}\,\frac{z^2}{\varpi^2} .$$

Therefore $r \leq 1$ iff

$$a \leq 0 \qquad \text{with} \qquad a = z(2\varpi - 1) + \tfrac{1}{2} z^2 .$$

Since $a \leq z\big(2(1-\theta z) - 1\big) + \frac{1}{2}z^2 = z + (\frac{1}{2} - 2\theta)z^2$ it follows that $a \leq 0$ whenever $\theta \geq \frac{1}{4}$. Further we have $r \geq -1$ iff

$$b \geq 0 \qquad \text{with} \qquad b = 2\varpi^2 + z\big(2\varpi - 1\big) + \tfrac{1}{2}\, z^2 .$$

Since $b = \frac{1}{2}(2\varpi + z)^2 - z$, it is seen that $b \geq 0$ irrespective of the value of θ. □

The above result is equivalent to the statement that $0 \in \mathcal{D}_0$ for any $s \geq 1$. Such a result for $z_0 = 0$ will provide stability on finite intervals for non-stiff explicit terms, for which we have $|z_0| \leq \Delta t\, L$ with a fixed Lipschitz constant L.

Methods (2.14) with $s = 1$: As a final example we consider the methods from Example 2.3 with $s = 1$. With these methods we have ω as free parameter, and $\theta = 1 - \frac{1}{2}\sqrt{2}$

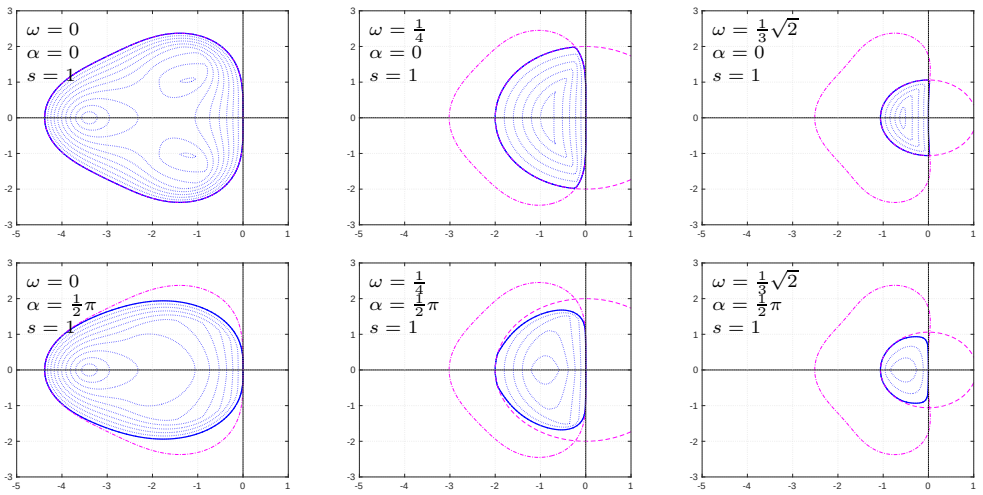

Figure 5: Stability domains for $s = 1$. Methods from Example 2.3, with the function $r = r_B$ [eq. (3.6)] for $\theta = 1 - \frac{1}{2}\sqrt{2}$ and $\nu = 2\theta(\hat{a}_{32} - b_2)$, $b_2 = -\frac{1}{2} + \frac{1}{4\theta}$, $\hat{a}_{32} = \frac{1}{2} + \omega$. From left to right: $\omega = 0$ [left], $\omega = \frac{1}{4}$ [middle] and $\omega = \frac{1}{3}\sqrt{2}$[right]. Top row for angle $\alpha = 0$, bottom row for $\alpha = \frac{1}{2}\pi$. For reference: the dash-dotted lines indicate the stability boundaries of the explicit methods from eq. (2.11), and the dashed lines give the necessary condition $|\phi_B(z_0)| \leq 1$, with ϕ_B given by eq. (3.12). [Note: $\phi_B = 0$ for $\omega = 0$.]

is fixed. Since it concerns here type-B methods, the stability domains are empty for $s = 2$ or larger. In contrast to this, we do get rather large domains for $s = 1$, as shown in Figure 5.

In this figure, along with the domains $\mathcal{D}_\alpha$, also the curves are drawn that are obtained from the necessary condition in Proposition 3.3(a) for $s = 1$. Stability of the explicit method can be viewed as another necessary condition, for the special case $z_1 = 0$. Together these two necessary conditions give a quite good inclusion for $\mathcal{D}_\alpha$ for the parameters $\omega = \frac{1}{4}$ and $\omega = \frac{1}{3}\sqrt{2}$. For $\omega = 0$ the function ϕ_B from (3.12) is equal to zero, so then the necessary condition from Proposition 3.3(a) vanishes.

Remark 3.6 For parabolic problems with mixed derivatives on Cartesian grids, one can apply dimension splitting with explicit treatment of the mixed derivatives. Stability results for such problems can be found in [9, 10], for example, together with applications for option pricing in financial mathematics.

4 Numerical illustration

In this section we present some numerical test results for a 2D reaction-diffusion problem. This problem will be examined with $s = 1$ and $s = 2$ to illustrate the differences between the type-A and type-B methods.

The test results will be presented for the following methods:

$$\begin{aligned}
&\text{SCM-A1 : type-A method, (2.9) with } \theta = 1 - \tfrac{1}{2}\sqrt{2},\\
&\text{SCM-A2 : type-A method, (2.9) with } \theta = \tfrac{1}{2} + \tfrac{1}{6}\sqrt{3},\\
&\text{SCM-B1 : type-B method, (2.14) with } \theta = 1 - \tfrac{1}{2}\sqrt{2},\ \omega = 0,\\
&\text{SCM-B2 : type-B method, (2.14) with } \theta = 1 - \tfrac{1}{2}\sqrt{2},\ \omega = \tfrac{1}{3}\sqrt{2}.
\end{aligned}$$

The tests were also performed with the type-A methods (2.13), as well as with the variant with $\kappa = \frac{1}{2}$, but the results with these methods differed only very slightly from the ones with the methods (2.9). In the error plots this would have led to lines and markers visually coinciding with those for (2.9).

4.1 A reaction-diffusion problem with pattern formation

As test problem we consider the so-called Schnackenberg model for the interaction of two chemical species, consisting of the following system of reaction-diffusion equations

$$\begin{aligned}
u_t &= D_1(u_{xx} + u_{yy}) + \kappa(a - u + u^2 v)\,,\\
v_t &= D_2(v_{xx} + v_{yy}) + \kappa(b - u^2 v)\,.
\end{aligned} \tag{4.1}$$

This system is considered on the spatial domain $\Omega = [0,1]^2$ and time interval $[0,T]$. The initial condition is

$$u(x,y,0) = a + b + 10^{-3} e^{-100\left(\left(x-\frac{1}{4}\right)^2 + \left(y-\frac{1}{6}\right)^2\right)}, \qquad v(x,y,0) = b/(a+b)^2,$$

and at the boundaries homogeneous Neumann conditions are imposed. The parameter values are $D_1 = 0.05$, $D_2 = 1$, $\kappa = 100$, $a = 0.1305$ and $b = 0.7695$. The initial condition consists of a small Gaussian perturbation added to the chemical steady state $u \equiv a + b = 0.90$, $v \equiv b/(a+b)^2 = 0.95$. This small perturbation is then amplified and spread, leading to the formation of patterns with spots.

The space derivatives are approximated by standard second-order finite differences on a uniform Cartesian mesh. This problem was used as numerical test in [13] with

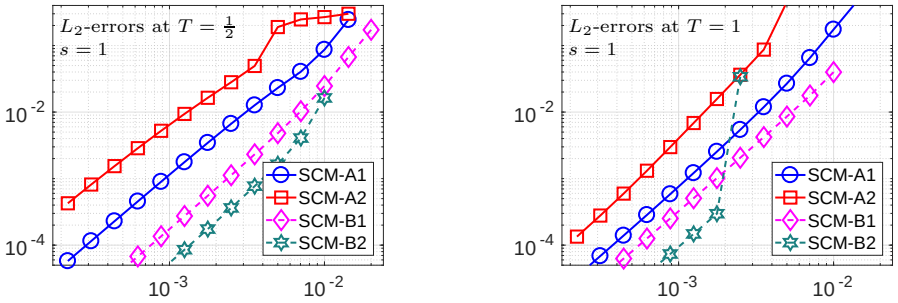

Figure 6: Problem (4.1) with $s = 1$: L_2-errors versus Δt at output time $T = \frac{1}{2}$ [left] and $T = 1$ [right] with 100×100 spatial grid and $\Delta t = 1/N$, $N = 50 \cdot 2^{(j-1)/2}$ rounded to even numbers ($j = 1, 2, \ldots, 14$). The methods are unstable with $\Delta t = \frac{1}{50}$, except for SCM-B1 with $T = \frac{1}{2}$; if $T = 1$, method SCM-B2 already appears unstable with step-size $\Delta t = \frac{1}{400}$.

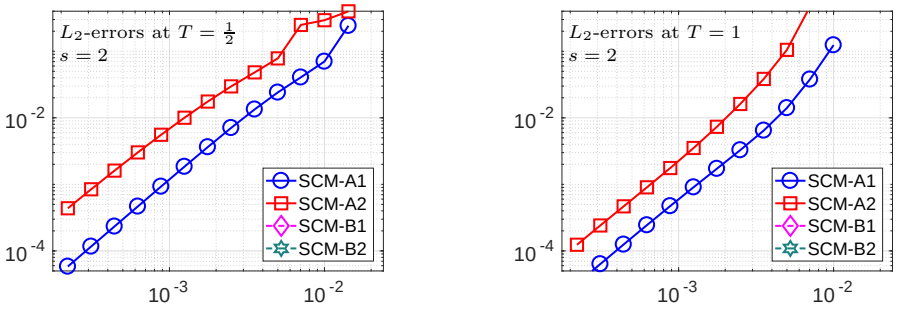

Figure 7: Problem (4.1) with $s = 2$: L_2-errors versus Δt at output time $T = \frac{1}{2}$ [left] and $T = 1$ [right] with 100×100 spatial grid and $\Delta t = 1/N$, $N = 50 \cdot 2^{(j-1)/2}$ rounded to even numbers ($j = 1, 2, \ldots, 14$). The methods SCM-A1, SCM-A2 are unstable with $\Delta t = \frac{1}{50}$; the methods SCM-B1, SCM-B2 are unstable for all the step-sizes.

$s = 1$ and F_1 the discretized 2D diffusion operator. Here we will also consider $s = 2$ with dimension splitting, such that F_1 contains the discretized x-derivatives and F_2 the y-derivatives. Computationally the problem becomes much easier with this dimension splitting, because all linear systems to be solved are then essentially three-diagonal. Tests in [1] for this problem on a hexagonal spatial region with domain decomposition splitting revealed that the errors at output time $T = \frac{1}{2}$ can be quite different from those at $T = 1$. Therefore the errors will be presented here for both these output times.

The results with a uniform spatial 100×100 grid are shown in Figure 6 for $s = 1$ and Figure 7 for $s = 2$. In these figures the temporal L_2-errors for the u-component are plotted as function of the step-size. To compute these temporal errors, a reference solution was found on this fixed spatial grid by using a very small step-size. Unstable results are not shown in these error plots. Consequently, the type-B methods are listed in the legends in Figure 7 even though these methods turned out to be unstable for $s = 2$, in agreement with Proposition 3.4.

As seen in the figures, the type-A methods can be used for the case of dimension splitting with $s = 2$. In fact, for this problem the dimension splitting hardly leads to an increase of the errors. The type-B methods, on the other hand, can only be used with $s = 1$, but for that case the errors can be significantly smaller than those of the

type-A methods.

With respect to stability for $s = 1$, the type-B method with $\omega = 0$ does allow larger steps than the method with $\omega = \frac{1}{3}\sqrt{2}$, in agreement with the stability domains shown in Figure 5. However, once the step-size is small enough to have stability, the errors for the method with $\omega = \frac{1}{3}\sqrt{2}$ are smaller than for $\omega = 0$.

5 Concluding remarks

5.1 Generalizations

The general stabilizing correction procedure can be generalized by treating F_0 different from the other F_j in the prediction steps. Based the formulas (1.3a) and (1.3b) for the explicit and the implicit method, with coefficients $\hat{a}_{ik}$ and a_{ik} respectively, we can proceed in the following way:

$$\left\{\begin{array}{rcl} x_{i,0} &=& u_n + \Delta t \sum_{k=1}^{i-1} \Big(\hat{a}_{ik} F_0(t_n + c_k\Delta t, y_k) + \sum_{j=1}^{s} \check{a}_{ik} F_j(t_n + c_k\Delta t, y_k) \Big) , \\ x_{i,j} &=& x_{i,j-1} + \Delta t \sum_{k=1}^{i-1} \big(a_{ik} - \check{a}_{ik}\big) F_j(t_n + c_k\Delta t, y_k) \\ && \quad + \Delta t\, a_{ii} F_j(t_n + c_i\Delta t, x_{i,j}) \qquad\qquad (j = 1, 2, \ldots, s) , \\ y_i &=& x_{i,s} . \end{array}\right. \tag{5.1}$$

Here the new coefficients $\check{a}_{ik}$ should be such that

$$\sum_{k=1}^{i-1} \check{a}_{ik} = \sum_{k=1}^{i-1} \check{a}_{ik} \tag{5.2}$$

in order to preserve the internal consistency of the scheme, by which all vectors $x_{i,j}$, $j = 0, 1, \ldots, s$, are to be consistent approximations to $u(t_n + c_i\Delta t)$.

By allowing such generalizations one can include for example the modified Craig-Sneyd methods of [10] and the modified Douglas method of [1]. However, the general formulas that can be obtained this way contain new free parameters $\check{a}_{ik}$ for which it is not easy to make suitable a priori choices.

This last point also applies for Runge-Kutta pairs with three or more stages. It seems, at present, no attempts have been made in that direction. Suitable pairs of methods might be found within the classes of IMEX Runge-Kutta methods derived in [2, 3, 14], for example.

Other generalization are possible by considering multistep methods. It is then easy to obtain methods of high order, but stability becomes more problematic. Application of such methods to special classes of parabolic problems from mathematical finance will be discussed in a separate report.

5.2 Conclusions

In this technical note splitting methods have been derived, using the idea of stabilizing corrections, starting from suitable pairs of explicit and diagonally-implicit Runge-Kutta methods with two stages. In the resulting splitting methods all internal stages provide fully consistent approximations to the exact solutions. Consequently, steady states of the ODE system are preserved as steady states of the splitting methods.

Linear invariance properties can be ensured in the splitting methods by performing a finishing stage involving the whole function F (the type-B methods). However, it

was found that for multiple implicit terms, $s > 1$, unconditional stability properties are then lost.

For the derived splitting methods, linear stability properties have been studied and some numerical tests were performed, but a full convergence analysis is still lacking. For practical relevance, such an analysis should be valid for (semi-discrete) PDEs and stiff ODEs, but this will make the analysis rather complicated because the local errors must be studied together with the error propagation, see for example the results obtained in [12] for $s = 1$ and [11] for $s = 2$ with some type-A methods.

In the numerical example it was seen that type-B methods may perform well for $s = 1$, but the additional finishing stage makes these methods unsuitable for problems with multiple implicit terms. Among the type-A methods with θ fixed, the influence of the second parameter κ was marginal in the tests, producing error plots with almost identical lines. Stronger nonlinearities might be needed to see significant differences.

References

[1] A. Arrarás, K.J. in 't Hout, W. Hundsdorfer, L. Portero, *Modified Douglas Splitting Methods for Reaction-Diffusion Equations.* BIT Numer. Math. 57 (2017), 261–285.

[2] U.M. Ascher, S.J. Ruuth, R.J. Spiteri, *Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations.* Appl. Numer. Math. 25 (1997), 151–167.

[3] S. Boscarino, L. Pareschi, *On the asymptotic properties of IMEX Runge-Kutta schemes for hyperbolic balance laws.* J. Comp. Appl. Math. 316 (2017), 60–73.

[4] J. Douglas, *Alternating direction methods for three space variables.* Numer. Math. 4 (1962), 41–63

[5] J. Douglas, J.E. Gunn, *A general formulation of alternating direction methods.* Numer. Math. 6 (1964), 428–453

[6] F.X. Giraldo, J.F. Kelly, E.M. Constantinescu, *Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere.* SIAM J. Sci. Comput. 35 (2013), B1162–B1194.

[7] D. Ghosh, E.M. Constantinescu, *Semi-implicit time integration of atmospheric flows with characteristic-based flux partitioning.* SIAM J. Sci. Comput. 38 (2016), A1848–A1875.

[8] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II – Stiff and Differential-Algebraic Problems.* Second edition, Springer, 1996.

[9] K.J. in 't Hout, C. Mishra, *Stability of ADI schemes for multidimensional diffusion equations with mixed derivative terms.* Appl. Numer. Math. 74 (2013), 83–94.

[10] K.J. in 't Hout, B.D. Welfert, *Unconditional stability of second-order ADI schemes applied to multi-dimensional diffusion equations with mixed derivative terms.* Appl. Numer. Math. 59 (2009), 677–692.

[11] K.J. in 't Hout, M. Wyns, *Convergence of the Hundsdorfer-Verwer scheme for two-dimensional convection-diffusion equations with mixed derivative term.* Proceedings ICNAAM-2014, T.E. Simos, C. Tsitouras (eds.), AIP Conf. Proc. 1648, 850054 (2015).

[12] W. Hundsdorfer, *Accuracy and stability of splitting with stabilizing corrections.* Appl. Numer. Math. 42 (2002), 213–233.

[13] W. Hundsdorfer, J.G. Verwer, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations.* Springer, 2003.

[14] C.A. Kennedy, M.H. Carpenter, *Additive Runge-Kutta schemes for convection-diffusion-reaction equations*. Appl. Numer. Math. 44 (2003), 139–181.

[15] D. Lanser, J.G. Blom, J.G. Verwer, *Time integration of the shallow water equations in spherical geometry*. J. Comput. Phys. 171 (2001), 373–393.

[16] G.I. Marchuk, *Splitting and alternating direction methods*. In: *Handbook of Numerical Analysis I*. Eds. P.G. Ciarlet, J.L. Lions, North-Holland, 1990, 197–462.