ELSEVIER

Applied Numerical Mathematics 36 (2001) 113–128

APPLIED NUMERICAL MATHEMATICS

www.elsevier.nl/locate/apnum

Factorization in block-triangularly implicit methods for shallow water applications ☆

P.J. van der Houwen, B.P. Sommeijer *

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

Abstract

The systems of first-order ordinary differential equations obtained by spatial discretization of the initial-boundary value problems modeling phenomena in shallow water in three spatial dimensions have right-hand sides of the form $f(t, y) := f_1(t, y) + f_2(t, y) + f_3(t, y) + f_4(t, y)$, where f_1, f_2 and f_3 contain the spatial derivative terms with respect to the x_1, x_2 and x_3 directions, respectively, and f_4 represents the forcing terms and/or reaction terms. It is typical for shallow water applications that the function f_4 is nonstiff and that the function f_3 corresponding with the vertical spatial direction is much more stiff than the functions f_1 and f_2 corresponding with the horizontal spatial directions. In order to solve the initial value problem for the system of ordinary differential equations numerically, we need a stiff solver. In a few earlier papers, we considered fully implicit Runge–Kutta methods and block-diagonally implicit methods. In the present paper, we analyze Rosenbrock type methods and the related DIRK methods (diagonally implicit Runge–Kutta methods) leading to block-triangularly implicit relations. In particular, we shall present a convergence analysis of various iterative methods based on approximate factorization for solving the triangularly implicit relations. Finally, the theoretical results are illustrated by a numerical experiment using a 3-dimensional shallow water transport model. © 2001 IMACS. Published by Elsevier Science B.V. All rights reserved.

Keywords: Numerical analysis; Shallow water applications; Iteration methods; Approximate factorization; Parallelism

1. Introduction

We consider initial-boundary value problems modeling phenomena in shallow water in three spatial dimensions. The systems of ordinary differential equations (ODEs) obtained by spatial discretization (method of lines) of the governing partial differential equations can be written in the form

$$\frac{dy(t)}{dt} = f(t, y(t)), \quad f(t, y) := f_1(t, y) + f_2(t, y) + f_3(t, y) + f_4(t, y), \quad y, f_k \in \mathbb{R}^N, \quad (1.1)$$

☆ The investigations reported in this paper were partly supported by the Dutch HPCN Program.

* Corresponding author.
E-mail address: bsom@cwi.nl (B.P. Sommeijer).

PII: S0168-9274(99)00050-1

where f_1, f_2 and f_3 contain the spatial derivative terms with respect to the x_1, x_2 and x_3 directions, respectively, f_4 represents the forcing terms and/or reaction terms, and N is a large integer proportional to the number of spatial grid points used for the spatial discretization. It is typical for shallow water applications that the function f_4 is nonstiff and that the function f_3 corresponding with the vertical spatial direction is much more stiff than the functions f_1 and f_2 corresponding with the horizontal spatial directions. As a consequence, the spectral radius of the Jacobian matrix $\partial f_3/\partial y$ is much larger than the spectral radius of $\partial f_1/\partial y$ and $\partial f_2/\partial y$. The reason is that in shallow seas the gridsize in the vertical direction is several orders of magnitude smaller than in the horizontal directions.

In order to solve the initial value problem (IVP) for the system (1.1) numerically, we need a stiff IVP solver, because the Lipschitz constants with respect to y associated with the functions f_1, f_2 and f_3 become increasingly large as the spatial resolution is refined. Stiff IVP solvers are necessarily implicit, requiring the solution of large systems of implicit relations. In a few earlier papers [3,8,10, 12], we considered the approximate factorization iteration of implicit Runge–Kutta methods leading to *fully coupled*, implicit systems whose dimension is a multiple of N, and of *block-diagonally* implicit methods in which the implicit relations can be decoupled into subsystems of dimension N (like backward differentiation formulas and block-diagonally implicit general linear methods). In these papers, it was shown that the spectral radius of the Runge–Kutta matrix, or of its equivalent in the case of *block-diagonally* implicit methods, determines the maximal convergent timestep and that convergence is faster as the spectral radius is smaller.

In the present paper, we analyze the approximate factorization iteration of Rosenbrock type methods and of the related DIRK methods (diagonally implicit Runge–Kutta methods) leading to *block-triangularly* implicit relations (this is also the case for the DIRK methods, in spite of the terminology "diagonally implicit"). Rosenbrock type methods are quite popular in air pollution simulations (see, e.g., [7,13,14]). This motivated us to look whether Rosenbrock and the related DIRK methods can also be useful in shallow water modeling.

First we show that in shallow water applications, where the eigenvalues of $\partial f_1/\partial y$, $\partial f_2/\partial y$ and $\partial f_3/\partial y$ are essentially purely imaginary, the so-called *factorized Rosenbrock methods*, which arise after performing just one approximate factorization iteration of the Rosenbrock method, are less suitable due to an extremely small imaginary stability boundary (see Section 2.3). However, continuing the approximate factorization iteration improves the stability considerably, because the convergence condition allows relatively large timesteps. Hence, if the underlying Rosenbrock method is unconditionally stable, the overall stability is largely determined by the convergence condition (see the numerical results in Section 5). A similar argument applies to the iterated DIRK methods. In fact, by an appropriate choice of the underlying Rosenbrock or DIRK method, the convergence region can be made much larger than in the case of the backward differentiation formulas (see Section 3.3.3).

In the numerical experiments, we use a three-dimensional shallow water transport model with chemical interactions and we compare the convergence of an iterated Rosenbrock method with that of the iterated two-step backward differentiation method applied in [8]. Both methods are second-order accurate and L-stable, so that in the case of convergence, their accuracy and stability properties are comparable. It turns out that the maximal convergent stepsize of the iterated Rosenbrock method is larger than that of the iterated backward differentiation method by a factor of about 2.

2. Rosenbrock methods and their factorization

We start with an example of a family of two-stage Rosenbrock methods:

$$\begin{aligned} &y_{n+1} = y_n + bk_1 + (1-b)k_2,\\ &(I-\kappa_1\Delta t J)k_1 = \Delta t f(y_n),\\ &(I-\kappa_2\Delta t J)k_2 = \Delta t f(y_n+\mu k_1) + \nu\Delta t J k_1, \qquad \kappa_i > 0, \quad \mu := \frac{1/2 - b\kappa_1 + (b-1)\kappa_2}{1-b} - \nu. \end{aligned} \tag{2.1}$$

Here, b, κ_1, κ_2 and ν are free parameters and J is an approximation to the Jacobian matrix $\partial f/\partial y$ at t_n. For simplicity of notation, we assumed the ODE of autonomous form. The nonautonomous version can be obtained by applying (2.1) to the augmented system $\{y' = f(y_0, y),\ y_0' = 1\}$. The method (2.1) is triangularly implicit, that is, k_1 and k_2 can be computed by successively solving two linear systems of dimension N.

If $J = \partial f/\partial y(t_n) + \mathrm{O}(\Delta t)$, then the formulas (2.1) are all second-order accurate Rosenbrock methods. The stability function for (2.1) is given by

$$R(z) = \frac{1 + (1-\kappa_1-\kappa_2)z + (1/2)(1-2\kappa_1-2\kappa_2+2\kappa_1\kappa_2)z^2}{(1-\kappa_1 z)(1-\kappa_2 z)}. \tag{2.2}$$

From this expression it follows that the methods (2.1) are A-stable if $\frac{1}{2} \leqslant \kappa_1+\kappa_2 \leqslant 2\kappa_1\kappa_2 + \frac{1}{2}$ and L-stable if $\kappa_1 + \kappa_2 = \kappa_1\kappa_2 + \frac{1}{2}$.

The first examples of Rosenbrock methods were given by Rosenbrock [6] in 1962 and are obtained by choosing in (2.1)

$$b = 0, \qquad \kappa_1 = \kappa_2 = \kappa := 1 \pm \tfrac{1}{2}\sqrt{2}, \qquad \nu = 0. \tag{2.3}$$

Of particular interest are the methods which remain second-order accurate if we choose an arbitrary matrix for J. Such methods are called Rosenbrock-W methods and were proposed by Steihaug and Wolfbrandt [9]. If we choose in (2.1) $\kappa_1 = \kappa_2 = \kappa$ and $\nu = -\kappa(1-b)^{-1}$, then (2.1) becomes a W-method (see Dekker and Verwer [2, p. 233]). The special case

$$b = \tfrac{1}{2}, \qquad \kappa_1 = \kappa_2 = \kappa := 1 \pm \tfrac{1}{2}\sqrt{2}, \qquad \nu = -2\kappa \tag{2.4}$$

was used by Verwer et al. [14] for solving atmospheric transport problems. Note, however, that for stability reasons, J should be a reasonably close approximation to the true Jacobian $\partial f/\partial y$ at t_n.

2.1. General Rosenbrock methods

More generally, we consider Rosenbrock methods of the form (cf. [4, p. 111])

$$y_{n+1} = y_n + (b^{\mathrm{T}} \otimes I)K, \qquad (I - T \otimes \Delta t J)K = \Delta t F(e \otimes y_n + (L \otimes I)K), \tag{2.5}$$

where b is an s-dimensional vector, $K := (k_1^{\mathrm{T}}, \ldots, k_s^{\mathrm{T}})^{\mathrm{T}}$, and T and L are lower and strictly lower triangular s-by-s matrices, respectively. This property of T and L implies that (2.5) is triangularly implicit, so that the components k_i of K can be computed by successively solving s linear systems of dimension N with system matrices $I - \kappa_i\Delta t J$, where the κ_i denote the diagonal entries of T. If the order of the method (2.5) is independent of the choice of the Jacobian approximation J, then (2.5) is again called a Rosenbrock-W method.

If T is not diagonal (as in (2.4)), then for an actual implementation one often transforms the linear system for $\boldsymbol{K}$ by a Butcher similarity transformation $\boldsymbol{U} = (T \otimes I)\boldsymbol{K}$, where T is assumed invertible (cf. [4, p. 120]). Writing $T^{-1} = S + D^{-1}$ with S strictly lower triangular and $D = \mathrm{diag}(T)$, (2.5) becomes

$$\begin{aligned} &\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + (\boldsymbol{b}^{\mathrm{T}}T^{-1} \otimes I)\boldsymbol{U}, \\ &(I - D \otimes \Delta t J)\boldsymbol{U} = \Delta t(D \otimes I)\boldsymbol{F}(\boldsymbol{e} \otimes \boldsymbol{y}_n + (LT^{-1} \otimes I)\boldsymbol{U}) - (DS \otimes I)\boldsymbol{U}. \end{aligned} \tag{2.6}$$

As in (2.5) the components $\boldsymbol{u}_i$ of $\boldsymbol{U}$ can be computed by again successively solving s linear systems of dimension N. As an example of a transformed Rosenbrock method, we give the transformation of the method (2.4):

$$\begin{aligned} &\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + \tfrac{1}{2}\kappa^{-1}(3\boldsymbol{u}_1 + \boldsymbol{u}_2), \\ &(I - \kappa\Delta t J)\boldsymbol{u}_1 = \kappa\Delta t \boldsymbol{f}(\boldsymbol{y}_n), \quad \kappa = 1 \pm \tfrac{1}{2}\sqrt{2}, \\ &(I - \kappa\Delta t J)\boldsymbol{u}_2 = \kappa\Delta t \boldsymbol{f}(\boldsymbol{y}_n + \kappa^{-1}\boldsymbol{u}_1) - 2\boldsymbol{u}_1. \end{aligned} \tag{2.4′}$$

Note that unlike (2.5), no Jacobian multiplications are involved in transformed Rosenbrock methods. In general, this is considered as an advantage because such Jacobian multiplications can be quite expensive. However, it should be remarked that in the case of shallow water applications the matrix J is extremely sparse, so that Jacobian multiplications are not so costly.

2.2. *Factorized Rosenbrock methods*

In order to further reduce the linear algebra costs in the method (2.4), Sandu [7] and Verwer et al. [13] applied to the system matrix $I - \kappa\Delta t J$ the technique of *approximate factorization* based on some splitting $\sum J_k$ of the Jacobian J. This leads to the *factorized Rosenbrock method.*

This technique goes back to Peaceman and Rachford [5] who used it for approximately solving the linear systems originating from a finite difference discretization of two-dimensional parabolic problems. In such problems, the system matrix is of the form $I - \frac{1}{2}\Delta t J$, where J is the discretization of the Laplace operator $\partial^2/\partial x_1^2 + \partial^2/\partial x_2^2$. By writing $J = J_1 + J_2$, where J_1 and J_2 correspond with $\partial^2/\partial x_1^2$ and $\partial^2/\partial x_2^2$, respectively, Peaceman and Rachford replaced $I - \frac{1}{2}\Delta t J$ by the approximate factorization $(I - \frac{1}{2}\Delta t J_1)(I - \frac{1}{2}\Delta t J_2)$.

The same approximate factorization technique can be applied to the matrix $I - T \otimes \Delta t J$ in (2.5) or to the matrix $I - D \otimes \Delta t J$ in (2.6). We shall illustrate this for the case (2.6). Since we are concerned with shallow water applications, we use the splitting $J = J_1 + J_2 + J_3$, where the matrices J_k denote the Jacobian matrices of the terms $\boldsymbol{f}_k$ at t_n occurring in the right-hand side function $\boldsymbol{f}$ in (1.1) and where the nonstiff interaction terms are ignored. This leads to the factorized method

$$\begin{aligned} &\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + (\boldsymbol{b}^{\mathrm{T}}T^{-1} \otimes I)\boldsymbol{V}, \\ &\Pi\boldsymbol{V} = \Delta t(D \otimes I)\boldsymbol{F}(\boldsymbol{e} \otimes \boldsymbol{y}_n + (LT^{-1} \otimes I)\boldsymbol{V}) - (DS \otimes I)\boldsymbol{V}, \end{aligned} \tag{2.7}$$

where Π is defined by

$$\Pi := (I - D \otimes \Delta t J_1)(I - D \otimes \Delta t J_2)(I - D \otimes \Delta t J_3), \quad D = \mathrm{diag}(T). \tag{2.8}$$

Each step of the factorized Rosenbrock method (2.7) requires the solution of $3s$ one-dimensional, linear systems. The three LU-decompositions needed in (2.8) can be computed in parallel, but the $3s$ forward–backward substitutions have to be done sequentially.

We may interpret the factorized method as the original Rosenbrock method with a perturbed matrix J. In the space spanned by the nonstiff eigenvectors of J, we even have $\Pi = I - D \otimes \Delta t J + \mathrm{O}((\Delta t)^2)$, showing only an order Δt perturbation of the matrix J. Hence, factorization will not affect the order of Rosenbrock-W methods. Furthermore, any factorized Rosenbrock method has at least order two if the original Rosenbrock method also has at least order two.

2.3. *Stability*

Next, we define the stability region $\mathbb{S}$ for the factorized versions of the methods (2.5) and (2.6). We first define the stability function by applying them to the test equation $y' = (J_1 + J_2 + J_3)y$. Assuming that the matrices J_k commute and ignoring the interaction terms in $\boldsymbol{F}$, the factorized versions of the methods (2.5) and (2.6) will reduce to recursions of the form

$$y_{n+1} = R(\Delta t J_1, \Delta t J_2, \Delta t J_3) y_n,$$

where $R(z_1, z_2, z_3)$ is a rational function of its arguments. Using the identities

$$1 + \boldsymbol{p}^{\mathrm{T}} M^{-1} \boldsymbol{q} = \frac{\det(M + \boldsymbol{q}\boldsymbol{p}^{\mathrm{T}})}{\det(M)}, \qquad \det(M + \widetilde{S}) = \det(M),$$

where M is a square, nonsingular matrix, $\widetilde{S}$ a square strictly lower triangular matrix, and $\boldsymbol{p}$ and $\boldsymbol{q}$ are vectors of the same dimension as the matrices M and $\widetilde{S}$ (the proof of the first identity can be found in, e.g., [1, p. 475]), we find that the stability functions corresponding to the factorized versions of (2.5) and (2.6) can be respectively expressed as

$$R(z_1, z_2, z_3) = \frac{\det(P + z(\boldsymbol{e}\boldsymbol{b}^{\mathrm{T}} - L))}{\det(P)}, \qquad P := (I - z_1 T)(I - z_2 T)(I - z_3 T), \quad (2.9)$$

$$R(z_1, z_2, z_3) = \frac{\det(P + DS + zD(\boldsymbol{e}\boldsymbol{b}^{\mathrm{T}} - L)T^{-1})}{\det(P)}, \qquad P := (I - z_1 D)(I - z_2 D)(I - z_3 D), \quad (2.10)$$

where $z := z_1 + z_2 + z_3$.

For the test equation defined above, the *stability region* is defined by the region $\mathbb{S}$ in the (z_1, z_2, z_3)-space where $|R(z_1, z_2, z_3)| \leqslant 1$. The method (2.7) is called *stable* if all eigenvalue triplets $(\Delta t\lambda(J_1), \Delta t\lambda(J_2), \Delta t\lambda(J_3))$ are in $\mathbb{S}$. Since in shallow water applications, many of the eigenvalues of J_k, $k = 1, 2, 3$, are close to the imaginary axis, we are particularly interested in the most critical case where the eigenvalues of J_k are purely imaginary, i.e., $z_k = \mathrm{i}y_k$ with y_k real-valued. Let us introduce for a given value of y_3 the *stability boundary* $\beta(y_3)$. This boundary is defined such that the method is stable at the points (y_1, y_2, y_3) with $y_3 \in \mathbb{R}$ and (y_1, y_2) in the region

$$\mathbb{S}(y_3) := \{(y_1, y_2): |y_k| \leqslant \beta(y_3), \ k = 1, 2\}. \quad (2.11)$$

Since the spectral radius of $\Delta t J_1$ and $\Delta t J_2$ is much smaller than that of $\Delta t J_3$, we like to have stability independent of the value of y_3. This is the case if $|y_k| \leqslant \min_{y_3} \beta(y_3)$ for $k = 1, 2$.

The corresponding time step condition is given by

$$\Delta t \leqslant \frac{\beta}{\max\{\rho(J_1), \rho(J_2)\}}, \qquad \beta := \min_{y_3} \beta(y_3). \quad (2.12)$$

Let us consider the stability of the factorized versions of (2.3) and (2.4′). It is easily verified that their stability functions respectively take the form

$$R_1(z_1,z_2,z_3) := 1 + \frac{z}{(1-\kappa z_1)(1-\kappa z_2)(1-\kappa z_3)} + \frac{(1/2)(1-2\kappa)z^2}{(1-\kappa z_1)^2(1-\kappa z_2)^2(1-\kappa z_3)^2}, \tag{2.13}$$

$$R_2(z_1,z_2,z_3) := 1 + \frac{2z}{(1-\kappa z_1)(1-\kappa z_2)(1-\kappa z_3)} + \frac{(1/2)z^2 - z}{(1-\kappa z_1)^2(1-\kappa z_2)^2(1-\kappa z_3)^2} \tag{2.14}$$

and that $|R_1(0,0,\mathrm{i}y_3)| < 1$, $|R_2(0,0,\mathrm{i}y_3)| < 1$ for $y_3 \neq 0$. Hence, we have a nonzero stability boundary β. However, a numerical calculation reveals that β is quite small (less than $\frac{1}{10}$). Hence, the factorized versions of (2.3), (2.4) and (2.4') are less suitable in shallow water applications where the Jacobians J_1, J_2 and J_3 have purely imaginary eigenvalues.

Although in practice spatially discretized shallow water problems in general do not lead to ODE systems of the model form $y' = (J_1 + J_2 + J_3)y$ with commuting J_k, the stability results based on the model problem are at least indicative for actual shallow water simulations (see Section 5).

3. Approximate factorization iteration

In this section, we will improve the quite poor stability properties of the factorized Rosenbrock methods along the imaginary axes. Here, the aim is to really solve the Rosenbrock method, for which we need an iteration method. One possibility would be to use a Krylov-type iterative solver (e.g., GMRES). This approach, however, requires a problem-dependent preconditioner to be efficient.

As an alternative, we analyze in this paper an iteration method based on the approximate factorization technique as described in the preceding section. Evidently, if this iteration process converges, then we retain the properties (like accuracy and stability) of the underlying method (henceforth called the corrector). Thus, if we restrict our considerations to A-stable (or even L-stable) correctors, then the stability region of the iteration method is the same as its convergence region.

Apart from the Rosenbrock corrector, we shall also study the iterative solution of the related implicit methods

$$y_{n+1} = y_n + \Delta t(b^{\mathrm{T}} \otimes I)F(X), \qquad X - \Delta t(A \otimes I)F(X) = e \otimes y_n, \tag{3.1}$$

where A is a lower triangular matrix. In [4, p. 97] these methods are called DIRK methods (diagonally implicit Runge–Kutta methods). Like Rosenbrock methods, DIRK methods are triangularly implicit (in spite of the terminology "diagonally implicit" now commonly accepted in the literature).

In the following sections we will discuss the iterative solution of Rosenbrock and DIRK correctors by means of approximate factorization and we will see that this approach leads to acceptably large convergence regions.

3.1. Iterative solution of the Rosenbrock equations

We consider two iterative approximate factorization approaches for actually solving the implicit Rosenbrock relations. The first approach solves the components u_i from (2.6) one after another by *repeated* application of a linear system solver, the second approach solves all components k_i from (2.5) *simultaneously* by a nonlinear system solver. We shall refer to these iteration methods as repeated and simultaneous approximate factorization iteration of the Rosenbrock method, briefly, the RAF-Rosenbrock and SAF-Rosenbrock processes, respectively.

3.1.1. The RAF-Rosenbrock process

The s linear systems in (2.6) have the form

$$\begin{aligned} &(I-\kappa_i\Delta t J)\boldsymbol{u}_i = \boldsymbol{g}_i, \quad i=1,\ldots,s,\\ &\boldsymbol{g}_i := (\boldsymbol{e}_i^{\mathrm{T}}\otimes I)\big(\Delta t(D\otimes I)\boldsymbol{F}(\boldsymbol{e}\otimes y_n + (LT^{-1}\otimes I)\boldsymbol{U}) - (DS\otimes I)\boldsymbol{U}\big), \end{aligned} \tag{3.2}$$

where κ_i is the ith diagonal entry of T. Since L and S are strictly lower triangular, these s systems can be solved successively. We solve the ith linear system by the linear solver

$$(\boldsymbol{e}_i^{\mathrm{T}}\otimes I)\Pi(\boldsymbol{u}_i^{(j)} - \boldsymbol{u}_i^{(j-1)}) = \boldsymbol{g}_i - (I-\kappa_i\Delta t J)\boldsymbol{u}_i^{(j-1)}, \quad j=1,2,\ldots,m,\ i=1,\ldots,s, \tag{3.3}$$

where Π is defined in (2.8). In this RAF-Rosenbrock process the initial iterates $\boldsymbol{u}_i^{(0)}$ should be provided by some predictor formula and the number of iterations m is assumed to be determined by some iteration strategy such that $\boldsymbol{u}_i^{(m)}$ is sufficiently close to the solution $\boldsymbol{u}_i$ of (3.2). In our numerical experiments in Section 4, we used the predictor $\boldsymbol{u}_i^{(0)} = \boldsymbol{0}$, $i=1,\ldots,s$. The effect of this choice is that the first iterate $\boldsymbol{u}_i^{(1)}$ is identical with the result of the factorized Rosenbrock method (2.7). Hence, this method can be considered as a predictor for the iterative approach.

If the iterates $\boldsymbol{u}_i^{(j)}$ converge, then they can only converge to the solution $\boldsymbol{u}_i$ of (3.2). Each iteration in (3.3) requires the solution of 3 linear systems with system matrices $I-\kappa_i\Delta t J_k$, $k=1,2,3$, each of order N. Note that the three LU-decompositions of these system matrices can again be done in parallel. These LU-decompositions and the corresponding forward-backward substitutions are relatively cheap, because the matrices J_k each correspond with a one-dimensional differential operator.

The convergence is determined by the error recursion satisfied by the iteration error vector $\boldsymbol{\varepsilon}^{(j)}$:

$$\boldsymbol{\varepsilon}^{(j)} := (\boldsymbol{u}_1^{(j)} - \boldsymbol{u}_1, \ldots, \boldsymbol{u}_s^{(j)} - \boldsymbol{u}_s), \tag{3.4}$$

$$\boldsymbol{\varepsilon}^{(j)} = Z_1\boldsymbol{\varepsilon}^{(j-1)}, \quad Z_1 := I - \Pi^{-1}(I - D\otimes\Delta t J), \qquad j=1,2,\ldots,m. \tag{3.5}$$

Before analyzing the matrix Z_1, we first derive the error recursion for the other iterative approaches.

3.1.2. The SAF-Rosenbrock process

Instead of solving the linear systems in the Rosenbrock method (2.6) successively for the components $\boldsymbol{u}_i$ of $\boldsymbol{U}$, we may iterate them simultaneously. Since in such an approach it is more convenient to go back to the untransformed method (2.5), we shall solve the components $\boldsymbol{k}_i$ of $\boldsymbol{K}$ simultaneously from (2.5). Consider the SAF-Rosenbrock process

$$\begin{aligned} &\Pi(\boldsymbol{K}^{(j)} - \boldsymbol{K}^{(j-1)}) = -\big((I - T\otimes\Delta t J)\boldsymbol{K}^{(j-1)} - \Delta t\boldsymbol{F}(\boldsymbol{e}\otimes y_n + (L\otimes I)\boldsymbol{K}^{(j-1)})\big),\\ &\quad j=1,2,\ldots m, \end{aligned} \tag{3.6}$$

where Π is again defined by (2.8). Note that this method is a *nonlinear* system solver.

Evidently, if the iterates $\boldsymbol{K}^{(j)}$ converge and if (2.5) has a unique solution $\boldsymbol{K}$, then they can only converge to this solution $\boldsymbol{K}$. Each SAF-Rosenbrock iteration requires the solution of three linear systems with system matrices $I - D\otimes\Delta t J_k$, $k=1,2,3$, each of order sN. The $3s$ LU-decompositions and the s forward–backward substitutions corresponding with each matrix $I - D\otimes\Delta t J_k$ can be done in parallel. Again, the LU-decompositions and the forward–backward substitutions are relatively cheap, because J_k corresponds with a one-dimensional differential operator. A drawback is the matrix–vector multiplication

in the right-hand side of (3.6). Note that applying the SAF-Rosenbrock iteration process to (2.6) instead of (2.5) does not avoid such a matrix–vector multiplication.

Let us consider the iteration error $\boldsymbol{\varepsilon}^{(j)} := \boldsymbol{K}^{(j)} - \boldsymbol{K}$. From (2.5) and (3.6) it follows that

$$\begin{aligned} &\boldsymbol{\varepsilon}^{(j)} = Z_2\boldsymbol{\varepsilon}^{(j-1)} + \Delta t\Pi^{-1}\boldsymbol{G}(\boldsymbol{\varepsilon}^{(j-1)}),\\ &Z_2 := I - \Pi^{-1}(I - (T+L)\otimes \Delta t J), \quad j = 1,2,\ldots,m,\\ &\boldsymbol{G}(\boldsymbol{\varepsilon}) := \boldsymbol{F}(\boldsymbol{e}\otimes \boldsymbol{y}_n + (L\otimes I)(\boldsymbol{K}+\boldsymbol{\varepsilon})) - \boldsymbol{F}(\boldsymbol{e}\otimes \boldsymbol{y}_n + (L\otimes I)\boldsymbol{K}) - (L\otimes J)\boldsymbol{\varepsilon}. \end{aligned} \tag{3.7}$$

Since $\boldsymbol{G}(\boldsymbol{\varepsilon})$ has a small Lipschitz constant in the neighbourhood of the origin, the error recursion (3.7) essentially behaves as the linearized recursion

$$\boldsymbol{\varepsilon}^{(j)} \approx Z_2\boldsymbol{\varepsilon}^{(j-1)}, \quad Z_2 := I - \Pi^{-1}(I - (T+L)\otimes \Delta t J), \qquad j = 1,2,\ldots,m. \tag{3.8}$$

3.2. Iterative solution of DIRK equations

As for Rosenbrock methods, we may consider repeated and simultaneous approximate factorization iteration of the DIRK method (3.1). These processes are respectively given by

$$\begin{aligned} &(\boldsymbol{e}_i^{\mathrm{T}}\otimes I)\Pi(\boldsymbol{x}_i^{(j)} - \boldsymbol{x}_i^{(j-1)}) = \boldsymbol{g}_i, \quad j = 1,2,\ldots,m,\ i = 1,\ldots,s,\\ &\boldsymbol{g}_i := (\boldsymbol{e}_i^{\mathrm{T}}\otimes I)\big((\boldsymbol{e}\otimes I)\boldsymbol{y}_n - \boldsymbol{X}^{(j-1)} + \Delta t(A\otimes I)\boldsymbol{F}(\boldsymbol{X}^{(j-1)})\big), \end{aligned} \tag{3.9}$$

and

$$\Pi(\boldsymbol{X}^{(j)} - \boldsymbol{X}^{(j-1)}) = -\big(\boldsymbol{X}^{(j-1)} - \Delta t(A\otimes I)\boldsymbol{F}(\boldsymbol{X}^{(j-1)}) - (\boldsymbol{e}\otimes I)\boldsymbol{y}_n\big), \quad j = 1,2,\ldots,m, \tag{3.10}$$

where Π is again defined by (2.8) with $D := \mathrm{diag}(A)$. They will be referred to as the RAF-DIRK and SAF-DIRK processes. A comparison with (3.3) and (3.6) shows that we have the same iteration costs except for the Jacobian multiplication.

Defining the iteration error $\boldsymbol{\varepsilon}^{(j)} := \boldsymbol{X}^{(j)} - \boldsymbol{X}$, we can write down the linearized error recursions. We find that the linearized error recursions associated with the RAF-DIRK method (3.9) and the SAF-DIRK method (3.10) are respectively given by

$$\boldsymbol{\varepsilon}^{(j)} = Z_3\boldsymbol{\varepsilon}^{(j-1)}, \quad Z_3 := Z_1, \qquad j = 1,2,\ldots,m \tag{3.11}$$

$$\boldsymbol{\varepsilon}^{(j)} \approx Z_4\boldsymbol{\varepsilon}^{(j-1)}, \quad Z_4 := I - \Pi^{-1}(I - A\otimes \Delta t J), \qquad j = 1,2,\ldots,m. \tag{3.12}$$

3.3. Convergence

The convergence of the iterated Rosenbrock methods (3.2) and (3.6), and of the iterated DIRK methods (3.9) and (3.10) is determined by the amplification matrices

$$\begin{aligned} &Z_1 := I - \Pi^{-1}(I - D\otimes \Delta t J), \qquad Z_2 := I - \Pi^{-1}(I - (T+L)\otimes \Delta t J),\\ &Z_3 = Z_1, \qquad\qquad\qquad\qquad Z_4 := I - \Pi^{-1}(I - A\otimes \Delta J), \end{aligned}$$

occurring in the error recursions (3.5), (3.8), (3.11) and (3.12), respectively. They only differ by the matrix in front of $\Delta t J$ (we recall that $D = \mathrm{diag}(T) = \mathrm{diag}(A)$). In the following subsections we respectively discuss the region of convergence where $\rho(Z_r) < 1$, the rate of convergence of the nonstiff iteration error components, and the stability of the iterated methods.

3.3.1. The region of convergence

The matrices Z_r are lower triangular block matrices with the same diagonal blocks

$$I-(I-\kappa_j\Delta t J_1)^{-1}(I-\kappa_j\Delta t J_2)^{-1}(I-\kappa_j\Delta t J_3)^{-1}(I-\kappa_j\Delta t J), \quad j=1,\ldots,s,$$

for all r. Here, the κ_i denote the diagonal entries of D. Hence, the eigenvalues of the matrices Z_r are identical. They act as amplification factors for the eigenvalue components of the iteration error. For the test problem also used in the stability analysis (see Section 2.3), they are given by

$$\alpha_j = C(\kappa_j z_1, \kappa_j z_2, \kappa_j z_3), \quad C(x_1,x_2,x_3) := 1-\frac{1-x_1-x_2-x_3}{(1-x_1)(1-x_2)(1-x_3)}, \tag{3.13}$$

where $j=1,\ldots,s$ and where z_k runs through the eigenvalues of $\Delta t J_k$. Evidently, we have convergence if $|\alpha_j|<1$, $j=1,\ldots,s$. We consider the most critical case where the eigenvalues of J_k are purely imaginary, that is we consider the values of $|\alpha_j| = |C(i\kappa_j y_1, i\kappa_j y_2, i\kappa_j y_3)|$. Recalling that the spectral radius of $\Delta t J_1$ and $\Delta t J_2$ is much smaller than that of $\Delta t J_3$, we are interested in convergence regions of the form (cf. (2.11))

$$\mathbb{C}(y_3) := \left\{(y_1,y_2):\ |y_k| \leqslant \frac{\gamma(y_3)}{\rho(D)},\ k=1,2\right\}, \quad |y_3| \leqslant \infty. \tag{3.14}$$

Theorem 3.1. *Let the function $g(x)$ be defined by the relation*

$$4xg^3+2(x^2-1)g^2-x^2-1=0. \tag{3.15}$$

Then, the convergence boundary $\gamma(y_3)$ in (3.14) is given by

$$\gamma(y_3) = \rho(D)\min_j \frac{g(\kappa_j|y_3|)}{\kappa_j} \tag{3.16}$$

and the minimal value of $\gamma(y_3)$ is given by the positive root of the equation $4\gamma^4(\gamma^2+1)=1$.

Proof. Using Maple, we verified that for given values of y_3, $|C(i\kappa_j y_1, i\kappa_j y_2, i\kappa_j y_3)|$ increases most rapidly along the line $y_1 = y_2$ (the length of the formulas prevents us from writing out the various derivative expressions), so that we may restrict our considerations to the values of

$$|\alpha_j| = |C(i\kappa_j y_1, i\kappa_j y_1, i\kappa_j y_3)| = \kappa_j^2|y_1|\left(\frac{y_1^2(\kappa_j^2y_3^2+1)+4y_1y_3+4y_3^2}{(1+\kappa_j^2y_1^2)^2(1+\kappa_j^2y_3^2)}\right)^{1/2}.$$

If we set $|\alpha_j|=1$, $x=\kappa_j y_3$ and $y=\kappa_j y_1$, then we find the relation

$$4xy^3+2(x^2-1)y^2-x^2-1=0.$$

This relation determines a real-valued function $y=g(x)$. Hence, for given values of κ_j and y_3, i.e., of x, we have $|\alpha_j|\leqslant 1$ provided that both $\kappa_j|y_1|$ and $\kappa_j|y_2|$ are bounded by $g(\kappa_j|y_3|)$. This proves (3.16).

In order to find the minimal value of $\gamma(y_3)$, we look at the plot of the function $g(x)$ (see Fig. 1). Let $x_1(y)$ and $x_2(y)$ denote the two solutions of the equation $4xy^3+2(x^2-1)y^2-x^2-1=0$. Then, the minimal value of $g(x)$ is determined by the relation $x_1(y)=x_2(y)$. This leads to the equation $4y^4(y^2+1)=1$ whose only positive root determines the minimal value of $\gamma(y_3)$. □

Since the positive root of the equation $4\gamma^4(\gamma^2+1)=1$ is given by $\gamma=0.647\ldots$ we derive from this theorem the following convergence condition:

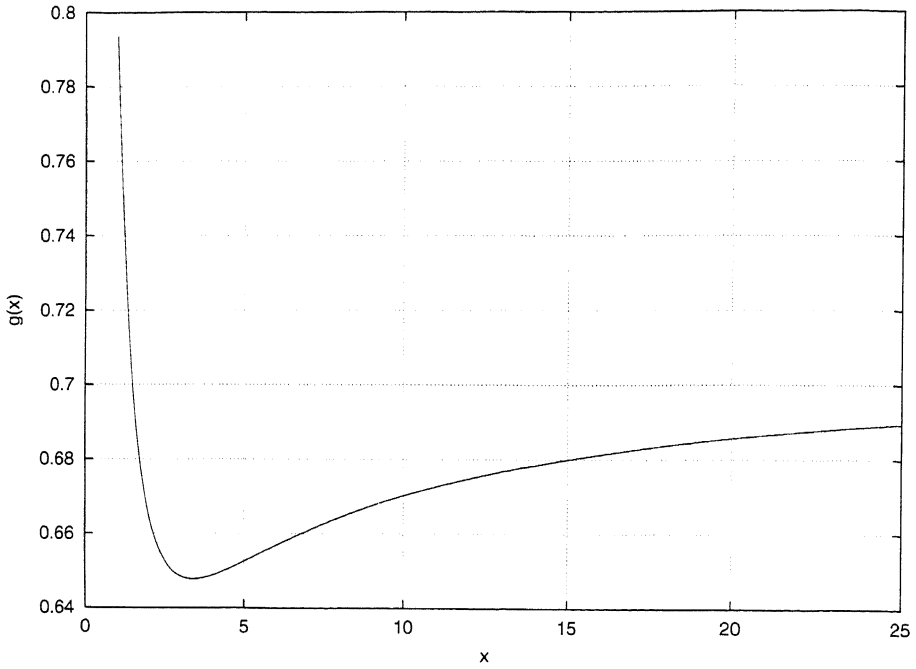

Fig. 1. The function $g(x)$ defined by (3.15).

Theorem 3.2. *Let $\lambda(J_k)$, $k = 1, 2, 3$, be purely imaginary. Then, a sufficient condition for convergence of the iterated Rosenbrock methods* (3.3) *and* (3.6), *and of the iterated DIRK methods* (3.9) *and* (3.10) *is given by*

$$\Delta t \leqslant \frac{\gamma}{\rho(D)\max\{\rho(J_1), \rho(J_2)\}}, \quad \gamma = 0.647\ldots.$$

3.3.2. The rate of convergence of the nonstiff error components

The rate of convergence of the *nonstiff* error components can be studied by the behaviour of the nonstiff amplification factors, that is, the eigenvalues of Z_r corresponding with small values of $\Delta t\lambda(J_k)$. From (3.13) it can be deduced that

$$\alpha_j = \kappa_j^2(z_1z_2 + z_1z_3 + z_2z_3) + O\big((\Delta t)^3\big), \quad z_k = \Delta t\lambda(J_k), \qquad j = 1, \ldots, s.$$

Hence, after m iterations the amplification factors behave as $O((\Delta t)^{2m})$ for all m and irrespective the value of r. However, this is not true for the amplification matrices Z_r^m.

Theorem 3.3. *The amplification matrices Z_r satisfy the relations*

$$r = 1, 3: \quad Z_r^m = O\big((\Delta t)^{2m}\big) \quad \textit{for all } m,$$

$$r = 2, 4: \quad \begin{cases} Z_r^m = O\big((\Delta t)^m\big) & \textit{for } m \leqslant s-1, \\ Z_r^m = O\big((\Delta t)^{2m+1-s}\big) & \textit{for } m \geqslant s. \end{cases}$$

Proof. The relation $Z_1 = Z_3 = O((\Delta t)^2)$ immediately follows from the definition of Z_1 and Z_3 in (3.5) and (3.11). For Z_2 and Z_4 it follows from (3.8) and (3.12) that

$$Z_2 = I - (I + D \otimes \Delta t J)(I - (T + L) \otimes \Delta t J) + O((\Delta t)^2) = (T + L - D) \otimes \Delta t J + O((\Delta t)^2).$$
$$Z_4 = I - (I + D \otimes \Delta t J)(I - A \otimes \Delta t J) + O((\Delta t)^2) = (A - D) \otimes \Delta t J + O((\Delta t)^2).$$

Hence, we certainly have $Z_r^m = O((\Delta t)^m)$ for $r = 2, 4$. However, writing $Z_r = A_r + B_r$ with $A_2 := (T + L - D) \otimes \Delta t J$ and $A_4 := (A - D) \otimes \Delta t J$, and observing that A_2 and A_4 are strictly lower block triangular, so that A_2^j and A_4^j vanish for $j \geqslant s$, we obtain for $m \geqslant s$

$$Z_r^m = \binom{m}{m-s+1} A_r^{s-1} B_r^{m-s+1} + \cdots + \binom{m}{m} B_r^m, \quad r = 2, 4.$$

Since $A_r = O(\Delta t)$ and $B_r = O((\Delta t)^2)$, we find that

$$Z_r^m = O((\Delta t)^{2m-s+1}), \quad r = 2, 4. \qquad \square$$

From this theorem it follows that in all four approaches the *nonstiff* error components are rapidly removed from the iteration error. However, we may expect that the RAF processes (3.5) and (3.9) damp these nonstiff components stronger than the SAF processes (3.8) and (3.12).

3.3.3. The region of stability

Evidently, if the iteration process converges, then the stability of the iterated method is determined by the stability of the underlying integration method. Hence, with respect to the stability test equation, the stability region of the iterated method converges to the intersection of the convergence region and the stability region of the integration method, that is, to

$$\mathbb{S} := \mathbb{S}_0 \cap \mathbb{C}, \qquad \mathbb{C} := \bigcap_{y_3} \mathbb{C}(y_3), \quad |y_3| \leqslant \infty,$$

where $\mathbb{S}_0$ is the stability region of the integration method and $\mathbb{C}(y_3)$ is defined by (3.14). For A-stable integration methods, the stability region $\mathbb{S}$ equals the convergence region $\mathbb{C}$, so that the stability condition is given by the stepsize condition in Theorem 3.2. Thus, for iterated, A-stable integration methods we may define the stability boundary $\beta := \gamma \rho^{-1}(D)$.

For example, if the Rosenbrock methods (2.3) and (2.4) (with $\kappa = 1 - \frac{1}{2}\sqrt{2}$) are iterated using the iteration matrix Π, then we find in both cases the stability boundary $\beta \approx 2.20$. If we choose $\kappa_1 = \kappa_2 = \frac{1}{4}$ in (2.1), then (2.1) is still A-stable with a slightly larger stability boundary $\beta \approx 2.59$. As a comparison, we mention that the second-order backward differentiation formula used in [8] has $\beta \approx 0.97$.

4. Explicit treatment of the horizontal terms

The modest values of the stability boundary β raises the question whether it is necessary to treat the horizontal terms fully implicitly. Afterall, when applying the standard, explicit, fourth-order Runge–Kutta method, we have an imaginary stability boundary of comparable size, viz. $\beta = 2\sqrt{2}$.

Let us define, in addition to the iteration matrix Π, the matrices

$$\begin{aligned} \Pi_3 &:= I - D \otimes \Delta t J_3, \\ \Pi_{13} &:= (I - D \otimes \Delta t J_1)(I - D \otimes \Delta t J_3), \\ \Pi_{23} &:= (I - D \otimes \Delta t J_2)(I - D \otimes \Delta t J_3). \end{aligned} \tag{4.1}$$

Table 1
Main characteristics of the iteration strategies

Process	Π	Π_3	$\Pi_3\Pi$	$\Pi_3\Pi^2$	$\Pi_{13}\Pi_{23}$
$\gamma \approx$	0.65	0.5	0.72	0.75	2.19
α_j	$O((\Delta t)^2)$	$O(\Delta t)$	$O((\Delta t)^{3/2})$	$O((\Delta t)^{5/3})$	$O(\Delta t)$
FBS per step	3 ms	1 ms	2 ms	7 ms/3	2 ms

Once again, consider the iteration methods (3.3), (3.6), (3.9) and (3.10), and let us replace the iteration matrix Π with the matrix Π_3 (the Π_3 process) or alternatingly with the iteration matrices $\Pi_3, \Pi, \Pi_3, \Pi, \ldots$ (the $\Pi_3\Pi$ process), or with $\Pi_3, \Pi, \Pi, \Pi_3, \Pi, \Pi, \ldots$ (the $\Pi_3\Pi^2$ process), or with Π_{13} and Π_{23} (the $\Pi_{13}\Pi_{23}$ process), etc. (the iteration strategy described in the preceding sections will be called the Π process). In each iteration of these processes, the vertical direction is treated implicitly, but not all horizontal directions are treated implicitly. After each update of the Jacobian, the Π_3 process requires only 1 LU-decomposition, whereas all other variants need 3 LU-decompositions. The number of forward/backward substitutions (FBSs) for the various approaches is given in Table 1.

In the institute report version of the present paper [11] the main properties of iteration processes of this type were analyzed. Here, we only summarize a few results. Table 1 lists for a number of iteration strategies (i) the convergence boundary γ in the timestep condition in Theorem 3.2, (ii) the order behaviour of the amplification factors α_j as $\Delta t \to 0$ (i.e., the eigenvalues of the analogues of the matrices Z_r as $\Delta t \to 0$, see Section 3.3.2), and (iii) the number forward/backward substitutions (FBS) after m iterations. This table reveals that the $\Pi_{13}\Pi_{23}$ process allows much larger convergent timesteps than the other strategies. However, if we look at the behaviour of the amplification factors α_j as a function of the eigenvalues of $\Delta t J_k$, then it turns out that for the $\Pi_{13}\Pi_{23}$ process the *averaged* amplification factor (averaged over the eigenvalue region) is considerably larger than for, e.g., the equally expensive $\Pi_3\Pi$ process. In [11] amplification factor profiles are given which indicate that on the basis of these profiles the $\Pi_3\Pi$ process is quite promising. A numerical comparison of the various iteration strategies will be subject of future research. In this paper, we only give numerical results for the RAF-Rosenbrock process (3.3) using the iteration matrix Π.

5. Numerical results

For our numerical experiments we chose a transport model for two interacting species of the form as used in [8]. This problem consists of two PDEs (for the concentrations c_1 and c_2) in three spatial dimensions,

$$\begin{aligned} \frac{\partial c_1}{\partial t} + \boldsymbol{V}\cdot\nabla c_1 &= \varepsilon \Delta c_1 - k_1 c_1 c_2, \\ \frac{\partial c_2}{\partial t} + \boldsymbol{V}\cdot\nabla c_2 &= \varepsilon \Delta c_2 - k_1 c_1 + k_2(1 - c_2), \end{aligned} \tag{5.1}$$

defined on $\mathbb{D} := \{(x_1, x_2, x_3)\colon 0 \leqslant x_1, x_2 \leqslant L_h, -L_v \leqslant x_3 \leqslant 0\}$, $0 \leqslant t \leqslant T$, with L_h, L_v, and T specified below. Here, ∇ is the three-dimensional gradient operator, $\boldsymbol{V} = (u, v, w)^{\mathrm{T}}$ denotes a divergence free

velocity field, ε is a diffusion constant, and k_1, k_2 are reaction constants. For V we took the same velocity field as in [8]:

$$\begin{aligned} u(t,x_1,x_2,x_3) &= \{\tilde{x}_2 + 3(\tilde{x}_3+\tfrac{1}{2})[(\tilde{x}_1-\tfrac{1}{2})^2 + (\tilde{x}_2-\tfrac{1}{2})^2 - p^2]\},\\ v(t,x_1,x_2,x_3) &= \{-\tilde{x}_1 + 3(\tilde{x}_3+\tfrac{1}{2})[(\tilde{x}_1-\tfrac{1}{2})^2 + (\tilde{x}_2-\tfrac{1}{2})^2 - p^2]\},\\ w(t,x_1,x_2,x_3) &= -3L_v\tilde{x}_3(\tilde{x}_3+1)\{(\tilde{x}_1-\tfrac{1}{2})/L_h + (\tilde{x}_2-\tfrac{1}{2})/L_h\}, \end{aligned} \tag{5.2}$$

where p is a given constant, $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ are the scaled co-ordinates $\tilde{x}_1 := x_1/L_h$, $\tilde{x}_2 := x_2/L_h$, $\tilde{x}_3 := x_3/L_v$. The boundary conditions are given by $c_1 = c_2 = 0$ on the vertical boundaries and $\partial c_1/\partial x_3 = \partial c_2/\partial x_3 = 0$ at the surface and at the bottom. The initial condition is of the form

$$c_i(t=0,x_1,x_2,x_3) = \exp\left\{\frac{\tilde{x}_3}{i} - \gamma_i\left[\left(\tilde{x}_1-\frac{1}{2}\right)^2 + \left(\tilde{x}_2-\frac{1}{2}\right)^2\right]\right\}, \quad i=1,2. \tag{5.3}$$

In our experiments, we take the following values for the various parameters (mks units):

$$\begin{aligned} &\varepsilon = 0.5, \quad k_1 = k_2 = 10^{-4}, \quad L_h = 20000, \quad L_v = 100, \quad T = 36000,\\ &p = \tfrac{1}{10}, \quad \gamma_1 = 80, \quad \gamma_2 = 20. \end{aligned} \tag{5.4}$$

The above test problem was discretized on a spatial grid with $N_1 = 51$, $N_2 = 51$ and $N_3 = 11$ grid points in the x_1-, x_2- and x_3-direction, respectively, using symmetric finite differences for the diffusion terms and upwind discretizations for the convection terms (for details we refer to [8]). The resulting ODE system is of the form (1.1) with $N \approx 57000$.

5.1. Convergence test

In order to show that block-triangularly implicit methods like Rosenbrock methods can be made convergent for larger timesteps than block-diagonally implicit methods like the backward differentiation formulas, we compared the convergence behaviour of the RAF-Rosenbrock method {(2.4′), (3.3), $\kappa = 1 - \frac{1}{2}\sqrt{2}$}, denoted by RAF-ROS, with that of approximate factorization iteration applied to the two-step backward differentiation formula (AF-BDF2) used in [8]. For RAF-ROS we used the predictor formula $\boldsymbol{u}^{(0)} = \boldsymbol{0}$. Recall that by this predictor, the first iteration result of RAF-ROS is identical with the factorized Rosenbrock method {(2.4′), (2.7)}. For AF-BDF2 we used the last-step-value predictor.

Table 2
Values of $r(m)$ for AF-BDF1 in the first step

Δt	$m=1$	$m=10$	$m=50$	$m=100$	$m=200$
900	0.56	0.0040	0.00082	0.0013	0.00094
950	0.59	0.0047	0.0016	0.0034	0.0039
1000	0.63	0.0057	0.0026	0.0078	0.015
1500	0.94	0.045	0.055	0.81	25.1
2400	1.50	0.32	2.1	10.7	840

Table 3
Values of $r(m)$ for RAF-ROS in the first step

Δt	$m=1$	$m=10$	$m=50$	$m=100$	$m=200$
2400	0.44	0.0069	0.00018	0.00005	0.00001
3000	0.55	0.019	0.004	0.0053	0.0028
3400	0.65	0.037	0.038	0.048	0.078
3500	0.69	0.074	0.079	0.11	0.15
4000	1.52	1.39	1.60	2.42	7.06

Table 4
Values of $r(m)$for AF-BDF2 in the second step

Δt	$m=1$	$m=10$	$m=50$	$m=100$	$m=200$
1470	12.3	8.6	2.4	0.76	0.45
1500	15.9	11.4	3.4	1.11	0.78
1600	34.1	26.3	10.4	3.73	3.15
1650	47.3	37.7	16.5	7.1	6.3

Table 5
Values of $r(m)$ for RAF-ROS in the second step

Δt	$m=1$	$m=10$	$m=50$	$m=100$	$m=200$
2400	0.44	0.0046	0.00012	0.000033	0.0000059
3000	0.57	0.012	0.0088	0.0091	0.0058
3400	0.66	0.39	0.32	0.43	0.32
3500	1.04	0.81	0.70	0.95	0.63
4000	69.3	70.8	59.8	27.4	73.0

The BDF2 itself can only be applied in the second and subsequent integration steps. In the first step, we applied the one-step BDF or implicit Euler rule (AF-BDF1). Tables 2 and 3 list values of the maximal absolute difference between two successive iterates after m iterations, denoted by $r(m)$, for AF-BDF1 and RAF-ROS in the first integration step. These results show that AF-BDF1 converges for timesteps smaller than 950, whereas RAF-ROS converges for $\Delta t < 3000$. Tables 4 and 5 present convergence results for the second integration step. Now, the critical stepsizes are roughly $\Delta t = 1470$ for AF-BDF2 and again $\Delta t = 3000$ for RAF-ROS. In this connection, we recall that the convergence boundaries for these methods are $\beta \approx 0.97$ and $\beta \approx 2.20$, respectively (see Section 3.3.3). Hence, the theory predicts a factor 2.2 larger convergent steps, whereas the practical gain factor is about a factor 2.0.

Table 6
Values of cd and $\tilde{m}$ (in brackets) for RAF-ROS

Δt	$m_1 = m_2 = 1$	TOL $= 10^{-1}$	TOL $= 10^{-2}$	TOL $= 10^{-3}$	TOL $= 10^{-4}$
1200	*	1.7 (2.6)	2.5 (4.2)	2.0 (7.2)	2.1 (10.9)
900	*	1.8 (1.9)	2.7 (2.7)	2.7 (3.3)	2.7 (5.0)
720	*	1.8 (1.3)	2.7 (1.9)	3.0 (2.5)	3.0 (3.6)
600	*	1.8 (1.1)	2.9 (1.6)	3.3 (2.2)	3.3 (3.0)
500	2.6	2.6 (1.0)	3.6 (1.6)	3.6 (2.0)	3.6 (2.7)

5.2. *Dynamic iteration strategy*

We conclude this paper with experiments showing that a dynamic iteration strategy based on approximate factorization improves the robustness of the integration process. We illustrate this by applying the two-stage RAF-ROS method $\{(2.4'),\ (3.3),\ \kappa = 1 - \frac{1}{2}\sqrt{2}\}$ with a stopping strategy based on the criterion $r_i(m_i) \geqslant$ TOL, where TOL is a given tolerance and $r_i(m_i)$ is the maximal absolute difference between two successive iterates after m_i iterations in the ith stage of the RAF-ROS method. Table 6 lists the number of correct digits in the end point $t = T$, i.e., the value of

$$\text{cd} := \text{minimum}\big(-{}^{10}\log(\text{absolute end point error})\big),$$

taken over all grid points and over both species, and in brackets the averaged number of iterations $\tilde{m}$ (over all steps and both stages) needed in the integration process. In order to illustrate the stabilizing effect of approximate factorization iteration, we also list the results obtained by the factorized Rosenbrock method, that is, the results obtained for $m_1 = m_2 = 1$. Negative cd-values are indicated by *. Table 6 clearly shows that the instabilities produced by factorized Rosenbrock can be removed if the dynamic iteration strategy is applied. We remark that the stable behaviour of factorized Rosenbrock for still reasonably large stepsizes is due to the diffusion terms and the upwind discretizations which introduce negative real parts in the eigenvalues of the Jacobians J_k. In general, stable results are already obtained for the values of $\tilde{m}$ between 1 and 3. Taking into account that in transport problems with a time-dependent velocity field the LU-decompositions need regular updating, we may conclude that the total iteration costs per step increase sublinearly with $\tilde{m}$. Hence, the introduction of a dynamic iteration strategy is quite effective, provided that the tolerance parameter TOL is appropriately chosen. In fact, this parameter should be related to a tolerance parameter which controls the local truncation error as is done in implementations of ODE solvers. Such more sophisticated implementations of the methods proposed in this paper will be subject of future research.

References

[1] H. Brunner, P.J. van der Houwen, The Numerical Solution of Volterra Equations, North-Holland, Amsterdam, 1986.

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

[3] C. Eichler-Liebenow, P.J. van der Houwen, B.P. Sommeijer, Analysis of approximate factorization in iteration methods, Appl. Numer. Math. 28 (1998) 245–258.
[4] E. Hairer, G. Wanner, Solving Ordinary Differential Equations, II. Stiff and Differential–Algebraic Problems, Springer, Berlin, 1991.
[5] D.W. Peaceman, H.H. Rachford Jr., The numerical solution of parabolic and elliptic differential equations, J. Soc. Indust. Appl. Math. 3 (1955) 28–41.
[6] H.H. Rosenbrock, Some general implicit processes for the numerical solution of differential equations, Computer J. 5 (1962–1963) 329–330.
[7] A. Sandu, Numerical aspects of air quality modeling, Ph.D. Thesis, University of Iowa, 1997.
[8] B.P. Sommeijer, The iterative solution of fully implicit discretizations of three-dimensional transport models, in: C.A. Lin, A. Ecer, J. Periaux, P. Fox, N. Satofuka (Eds.), Parallel Computational Fluid Dynamics—Development and Applications of Parallel Technology, Proceedings of the 10th Int. Conf. on Parallel CFD, Hsinchu, Taiwan, May 1998, Elsevier, Amsterdam, 1999, pp. 67–74.
[9] T. Steihaug, A. Wolfbrandt, An attempt to avoid exact Jacobian and nonlinear equations in the numerical solution of stiff differential equations, Math. Comp. 33 (1979) 521–534.
[10] P.J. van der Houwen, B.P. Sommeijer, Approximate factorization in shallow water applications, Report MAS R9835, CWI, Amsterdam, 1998, submitted for publication.
[11] P.J. van der Houwen, B.P. Sommeijer, Factorization in block-triangularly implicit methods for shallow water applications, Report MAS R9906, CWI, Amsterdam, 1999.
[12] P.J. van der Houwen, B.P. Sommeijer, J. Kok, The iterative solution of fully implicit discretizations of three-dimensional transport models, Appl. Numer. Math. 25 (1997) 243–256.
[13] J.G. Verwer, W. Hundsdorfer, J.G. Blom, Numerical time integration for air pollution models, Report MAS R9825, CWI, Amsterdam, 1998, submitted for publication.
[14] J.G. Verwer, E.J. Spee, J.G. Blom, W.H. Hundsdorfer, A second-order Rosenbrock method applied to photochemical dispersion problems, SIAM J. Sci. Comput. 20 (1999) 1456–1480.