Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

Approximate factorization in shallow water applications

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

Modelling, Analysis and Simulation (MAS)

MAS-R9835 December 1998

Report MAS-R9835
ISSN 1386-3703

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

Approximate Factorization in Shallow Water Applications

P.J. van der Houwen & B.P. Sommeijer
CWI
P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT
We consider the numerical integration of problems modelling phenomena in shallow water in 3 spatial dimensions. If the governing partial differential equations for such problems are spatially discretized, then the righthand side of the resulting system of ordinary differential equations can be split into terms $\mathbf{f}_1$, $\mathbf{f}_2$, $\mathbf{f}_3$ and $\mathbf{f}_4$, respectively representing the spatial derivative terms with respect to the x, y and z directions, and the interaction terms. It is typical for shallow water applications that the interaction term $\mathbf{f}_4$ is nonstiff and that the function $\mathbf{f}_3$ corresponding with the vertical spatial direction is much more stiff than the functions $\mathbf{f}_1$ and $\mathbf{f}_2$ corresponding with the horizontal spatial directions. 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 these systems numerically, we need a stiff IVP solver, which is necessarily implicit, requiring the iterative solution of large systems of implicit relations. The aim of this paper is the design of an efficient iteration process based on approximate factorization. Stability properties of the resulting integration method are compared with those of a number of integration methods from the literature. Finally, a performance test on a shallow water transport problem is reported.

1991 Mathematics Subject Classification: 65L06
Keywords and Phrases: numerical analysis, partial differential equations, iteration methods, approximate factorization, parallelism.
Note. The investigations reported in this paper were partly supported by the Dutch HPCN Program.

1. Introduction

We consider initial-boundary value problems modelling phenomena in shallow water in 3 spatial dimensions, such as the transport of pollutants in shallow seas including the mutual chemical interactions of these species. In [16] a full description of the partial differential equations (PDEs) describing various shallow water applications can be found (see also Section 4 of this paper for a model transport problem with chemical interactions). The systems of ordinary differential equations (ODEs) obtained by spatial discretization of these PDEs (method of lines) can be written in the form

$$(1.1)\qquad \frac{d\mathbf{y}(t)}{dt} = \mathbf{f}(t,\mathbf{y}(t)), \quad \mathbf{f}(t,\mathbf{y}) := \mathbf{f}_1(t,\mathbf{y}) + \mathbf{f}_2(t,\mathbf{y}) + \mathbf{f}_3(t,\mathbf{y}) + \mathbf{f}_4(t,\mathbf{y}), \qquad \mathbf{y}, \mathbf{f}_k \in \mathbb{R}^N,$$

where $\mathbf{f}_1$, $\mathbf{f}_2$ and $\mathbf{f}_3$ contain the spatial derivative terms with respect to the x, y and z directions, respectively, $\mathbf{f}_4$ 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 $\mathbf{f}_4$ is nonstiff and that the function $\mathbf{f}_3$ corresponding with the vertical spatial direction is much more stiff than the functions $\mathbf{f}_1$ and $\mathbf{f}_2$ corresponding with the horizontal spatial directions. As a consequence, the spectral radius of the Jacobian matrix $\partial\mathbf{f}_3/\partial\mathbf{y}$ is much larger than the spectral radius of $\partial\mathbf{f}_1/\partial\mathbf{y}$ and $\partial\mathbf{f}_2/\partial\mathbf{y}$. The reason is that in shallow seas the vertical gridsize is several orders of magnitude smaller than the horizontal gridsize.

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 $\mathbf{f}_1$, $\mathbf{f}_2$ and $\mathbf{f}_3$ become increasingly large as the spatial resolution is refined. Stiff IVP solvers are necessarily implicit, requiring the solution of large systems of implicit relations. We distinguish two basic types of implicit IVP solvers, viz. partially implicit methods and fully implicit methods (see Section 2). The aim of this paper is the design of an iteration process for use in fully implicit methods such that shallow water problems can be solved more efficiently than by partially implicit methods. Section 3 focuses on iteration processes based on factorization of the system matrix in the Newton method according to the splitting in (1.1). A detailed discussion of the convergence of the iteration process and the stability of the resulting integration method is presented. A performance evaluation of this integration method for a shallow water transport problem is presented in Section 4.
Finally, we remark that all plots and many of the formulas appearing in this paper were obtained using the software package Maple [12].

2. Implicit IVP solvers

We briefly discuss the suitability of partially implicit methods and fully implicit methods in shallow water applications. In particular, we discuss the stability properties of these methods. Ignoring the nonstiff interaction term $\mathbf{f}_4$, we consider stability with respect to the linear test equation $\mathbf{y}' = J\mathbf{y}$, where $J = J_1 + J_2 + J_3$ with J_k denoting an approximation to the Jacobian matrix of $\mathbf{f}_k$ at t_n, and where the matrices J_k are assumed to share the same eigenspace, the usual assumption if a normal mode analysis is applied. This equation will be referred to as the *stability test equation*. Writing $z_k = \Delta t\lambda(J_k)$, the region $\mathbb{S}$ in the (z_1, z_2, z_3)-space is called the *stability region* if the integration method is stable with respect to all test equations $\mathbf{y}' = J\mathbf{y}$ with the eigenvalue triples $(\Delta t\lambda(J_1), \Delta t\lambda(J_2), \Delta t\lambda(J_3))$ in $\mathbb{S}$. Since in shallow water applications many of the eigenvalues of J_k are close to the imaginary axis, we are interested in the most critical case where the eigenvalues of J_k are purely imaginary, i.e. $\Delta t\lambda(J_k) = iy_k$ with y_k real-valued. Furthermore, the spectral radius of ΔtJ_1 and ΔtJ_2 is much smaller than that of ΔtJ_3, so that we want stability in regions of the form

$$\text{(2.1a)} \qquad \mathbb{S} := \{(y_1,y_2,y_3): |y_k| \le \beta,\ k = 1, 2; |y_3| \le \infty\},$$

where the *stability boundary* β is not too small. The corresponding timestep condition is then given by

$$\text{(2.2)} \qquad \Delta t \le \frac{\beta}{\max\{\rho(J_1),\rho(J_2)\}}.$$

It may happen that the value of β is determined by a small set of critical y_3-values, that is, ignoring these critical values on the y_3-axis would lead to substantially greater values of β. To get insight into this situation, we introduce for a given value of y_3 the stability boundary $\beta(y_3)$ which is such that the method is stable in a region of the form

(2.1b) $\mathbb{S}(y_3) := \{(y_1,y_2): |y_k| \le \beta(y_3),\ k = 1, 2\}$.

Evidently, $\beta = \min_{y_3} \beta(y_3)$ and $\mathbb{S} = \cap_{y_3} \mathbb{S}(y_3)$ for $|y_3| \le \infty$.

2.1. Partially implicit methods

We confine our considerations to the so-called *implicit-explicit methods* which recently got renewed attention (see e.g. [1], [2] and [7]), the *stabilizing corrections method* of Douglas [4], and the *approximating corrections method* of Yanenko [17].

2.1.1. Implicit-explicit methods. Implicit-explicit methods arise if in a fully implicit method the 'implicit' righthand side evaluations are split into an explicit and an implicit part. A typical example is the implicit-explicit version of the two-step backward differentiation formula (BDF), advocated in [7] and [15]. When applied to (1.1) and taking into account that we want to treat the $\mathbf{f}_3$ term implicitly, this method becomes

$$\begin{aligned} &\mathbf{y}_{n+1}^{(0)} = 2\mathbf{y}_n - \mathbf{y}_{n-1}, \\ &\mathbf{y}_{n+1} = \tfrac{4}{3}\mathbf{y}_n - \tfrac{1}{3}\mathbf{y}_{n-1} + \tfrac{2}{3}\Delta t \left(\mathbf{f}(t_{n+1},\mathbf{y}_{n+1}^{(0)}) - \mathbf{f}_3(t_{n+1},\mathbf{y}_{n+1}^{(0)}) + \mathbf{f}_3(t_{n+1},\mathbf{y}_{n+1})\right). \end{aligned} \tag{2.3}$$

Here, Δt is the stepsize $t_{n+1} - t_n$ and $\mathbf{y}_n$ represents the numerical approximation to $\mathbf{y}(t_n)$. The method (2.3) is second-order accurate and requires the solution of one one-dimensionally implicit system per step. This system can be solved by Newton iteration using a band solver to handle the linear Newton systems. Hence, from a computational point of view, the costs per step are quite modest.
Applying (2.3) to the stability test equation defined above, we obtain a linear two-step recursion whose characteristic equation is given by

$$(1 - \tfrac{2}{3}z_3)w^2 - \tfrac{4}{3}(1 + z_1 + z_2)w + \tfrac{1}{3}(1 + 2z_1 + 2z_2) = 0.$$

The method (2.3) will be called stable if this equation has its roots on the unit disk. According to the boundary locus method, the boundary of the stability region in the (y_1,y_2,y_3)-space is defined by

$$\begin{aligned} &4(y_1 + y_2)\sin(\phi) + 2y_3\sin(2\phi) = 4\cos(\phi) - 3\cos(2\phi) - 1, \\ &\qquad\qquad 0 \le \phi \le 2\pi. \\ &2(y_1 + y_2)(1 - 2\cos(\phi)) - 2y_3\cos(2\phi) = 4\sin(\phi) - 3\sin(2\phi), \end{aligned}$$

Solving this system for $y_1 + y_2$ and y_3 yields the solution

$$y_1 + y_2 = \frac{1 - \cos(\phi)}{2\sin(\phi)}, \quad y_3 = 3\frac{1 - \cos(\phi)}{2\sin(\phi)}.$$

Hence, the boundary of the stability region is given by the plane $3y_1 + 3y_2 - y_3 = 0$, so that the stability boundary $\beta(y_3)$ introduced in (2.1b) is given by $\beta(y_3) = \frac{1}{6}|y_3|$. Hence, we have stability in

regions of the form (2.1a) with $\beta = 0$, showing that the implicit-explicit BDF is less suitable for shallow water problems.

2.1.2. Douglas and Yanenko methods. In [9], where a number of splitting methods for three-dimensional transport in shallow water were compared, the second-order methods from the family of stabilizing corrections methods of Douglas and of approximating corrections methods of Yanenko turned out to be the most efficient ones. When applied to (1.1), these families are given by

$$\begin{aligned} &\mathbf{y}_{n+1}^{(0)} = \mathbf{y}_n + \Delta t\, \mathbf{f}(t_n,\mathbf{y}_n), \\ &\mathbf{y}_{n+1}^{(k)} = \mathbf{y}_{n+1}^{(k-1)} + \theta\Delta t\big(\mathbf{f}_k(t_{n+1},\mathbf{y}_{n+1}^{(k)}) - \mathbf{f}_k(t_n,\mathbf{y}_n)\big), \quad k = 1, \dots, 4, \\ &\mathbf{y}_{n+1} = \mathbf{y}_{n+1}^{(4)} \end{aligned} \tag{2.4}$$

and

$$\begin{aligned} &\mathbf{y}_{n+1}^{(0)} = \mathbf{y}_n, \\ &\mathbf{y}_{n+1}^{(k)} = \mathbf{y}_{n+1}^{(k-1)} + \theta\Delta t\, \mathbf{f}_k(t_n+\theta\Delta t,\mathbf{y}_{n+1}^{(k)}), \quad k = 1, \dots, 4, \\ &\mathbf{y}_{n+1} = \mathbf{y}_n + \Delta t\, \mathbf{f}(t_n+\theta\Delta t,\mathbf{y}_{n+1}^{(4)}), \end{aligned} \tag{2.5}$$

where θ is a positive parameter. The implicit relations for $\mathbf{y}_{n+1}^{(k)}$, $k = 1, 2, 3$, are of the same type as in the implicit-explicit BDF (2.3). Note, however, that the LU-decompositions needed in the Newton process can be computed in parallel, but all forward-backward substitutions have to be done sequentially. Since the reaction term $\mathbf{f}_4$ is nonstiff, $\mathbf{y}_{n+1}^{(4)}$ can be solved by fixed point iteration.
For $\theta = \frac{1}{2}$ the methods are second-order accurate, otherwise first-order accurate. Furthermore, for all θ, the Douglas method (2.4) has stage order 1 (i.e. the order of the stage values $\mathbf{y}_{n+1}^{(k)}$ equals one), whereas the Yanenko method (2.4) has zero stage order. This stage order property is important, because it improves the actual accuracy of the method (this was confirmed by the experiments in [9]).
In order to specify the stability properties of (2.4) and (2.5), we apply them to the stability test equation, to obtain the recursion

$$\mathbf{y}_{n+1} = R(\Delta tJ_1, \Delta tJ_2, \Delta tJ_3)\, \mathbf{y}_n, \quad R(z_1,z_2,z_3) := 1 + \frac{z_1 + z_2 + z_3}{(1 - \theta z_1)(1 - \theta z_2)(1 - \theta z_3)}\ . \tag{2.6}$$

The methods (2.4) and (2.5) are stable if $|R(z_1,z_2,z_3)| \le 1$ with z_k running through the eigenvalues of $\Delta t\, \lambda(J_k)$. Hundsdorfer [11] showed that $|R(z_1,z_2,z_3)| \le 1$ on the infinite wedge $|\arg(-z_k)| \le \frac{\pi}{4}$, $k = 1, 2, 3$. However, since we assumed that the eigenvalues of J_k are purely imaginary, we are more interested in stability in regions of the form (2.1a). Unfortunately, requiring $|R(iy_1,iy_2,iy_3)| \le 1$ in $\mathbb{S}$ yields $\beta = 0$ for $\theta \le \frac{1}{2}$ and quite small β - values for $\theta > \frac{1}{2}$. We also looked at plots for the stability boundary $\beta(y_3)$ introduced in (2.1b). For $\theta = \frac{1}{2}$ we found that $\beta(0) = \infty$ and $\beta(y_3) = 0$ for $|y_3| > 0$. This makes the $(\theta = \frac{1}{2})$ - method unsuitable for shallow water applications. More interesting are the plots for $\theta > \frac{1}{2}$. Figure 2.1a presents a plot for $\theta = \frac{3}{5}$ (this value turned out to be almost optimal as far as stability is concerned). Details in the neighbourhood of the origin are given in Figure 2.1b. Evidently, the $(\theta = \frac{3}{5})$ - method possesses considerably better stability properties than the $(\theta = \frac{1}{2})$ -

method. This is also reflected by the values of β_ε defined by requiring $|R(iy_1,iy_2,iy_3)| \le 1 + \varepsilon$ in the region $\mathbb{S}$ given in (2.1a). Here, ε is a small positive parameter, so that $|R| \le 1 + \varepsilon$ implies a very mild form of instability. It can be shown that for $\theta = \frac{1}{2}$

(2.7a) $$\beta_\varepsilon = \sqrt{2\varepsilon}\left(1 + \tfrac{1}{2}\varepsilon + O(\varepsilon^2)\right).$$

For example, for $\varepsilon = 10^{-5}$ we obtain $\beta_\varepsilon \approx 0.0045$ which is of course unacceptably small. However, the $\left(\theta = \frac{3}{5}\right)$ - method yields

(2.7b) $$\beta_\varepsilon = \tfrac{5}{3}\sqrt[6]{\frac{\varepsilon}{10}}\left(1 - \sqrt[3]{\frac{\varepsilon^2}{10}} + O(\varepsilon)\right),$$

so that we obtain for $\varepsilon = 10^{-5}$ the quite reasonable value $\beta_\varepsilon \approx 0.167$. Nevertheless, we do not recommend the $\left(\theta = \frac{3}{5}\right)$ - method for shallow water applications.

2.2. Fully implicit methods

We consider fully implicit IVP solvers that fit into the wide class of General Linear Methods introduced by Butcher in 1966 (see [3, p. 335] for a detailed discussion). These methods are given by

(2.8) $$\mathbf{Y}_{n+1} - \Delta t(A\otimes I)\mathbf{F}(\mathbf{e}t_n + \mathbf{c}\Delta t,\mathbf{Y}_{n+1}) = (B\otimes I)\mathbf{Y}_n + \Delta t(C\otimes I)\mathbf{F}(\mathbf{e}t_{n-1} + \mathbf{c}\Delta t,\mathbf{Y}_n),\ \ n \ge 0.$$

Here A, B and C denote s-by-s matrices, I is the identity matrix whose order equals that of the system (1.1), **e** is an s-dimensional vector with unit entries, $\mathbf{c} = (c_i)$ is an s-dimensional abscissae vector, and $\otimes$ denotes the Kronecker product, i.e. if $A = (a_{ij})$, then $A\otimes I$ denotes the matrix of matrices $(a_{ij}I)$. Furthermore, for any vector $\mathbf{Y}_n = (\mathbf{y}_{ni})$, $\mathbf{F}(\mathbf{e}t_{n-1} + \mathbf{c}\Delta t,\mathbf{Y}_n)$ contains the derivative values $(\mathbf{f}(t_{n-1} + c_i\Delta t,\mathbf{y}_{ni}))$. The s vector components $\mathbf{y}_{n+1,i}$ of $\mathbf{Y}_{n+1}$ represent numerical approximations to the s exact solution vectors $\mathbf{y}(t_n + c_i\Delta t)$. The quantities $\mathbf{Y}_n$ are usually called the *stage vectors* and their components $\mathbf{y}_{ni}$ the *stage values*. We assume that the step point value $\mathbf{y}_n$ is defined by the last component of $\mathbf{Y}_n$, i.e. $\mathbf{y}_n := (\mathbf{e}_s^T\otimes I)\mathbf{Y}_n$, where $\mathbf{e}_s$ is the sth unit vector.

Each step by the method (2.8) requires the solution a nonlinear system. Let us define the residue function

(2.9) $$\mathbf{R}(\mathbf{Y}) := \mathbf{Y} - \Delta t(A\otimes I)\mathbf{F}(\mathbf{e}t_n + \mathbf{c}\Delta t,\mathbf{Y}) - (B\otimes I)\mathbf{Y}_n - \Delta t(C\otimes I)\mathbf{F}(\mathbf{e}t_{n-1} + \mathbf{c}\Delta t,\mathbf{Y}_n)$$

and consider the Newton-type iteration process for solving $\mathbf{Y}$ from $\mathbf{R}(\mathbf{Y}) = \mathbf{0}$:

(2.10) $$M\left(\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)}\right) = -\mathbf{R}(\mathbf{Y}^{(j-1)}),\quad j = 1, 2, \dots ,$$

where M is an approximation to the Jacobian matrix of the stiff part of $\mathbf{R}(\mathbf{Y})$, i.e.

(2.11) $$M = I - \Delta t\, A\otimes(J_1 + J_2 + J_3),$$

where J_1, J_2 and J_3 are defined above. This expression shows that solving the linear Newton systems (2.10) by a direct method is quite costly. It is the aim of this paper to reduce these costs by replacing the matrix M by a suitable approximation based on the splitting of the righthand side function given in (1.1). A detailed analysis of the convergence of the resulting iteration process will be the subject of Section 3. This analysis reveals that the magnitude of the spectral radius $\rho(A)$ of the matrix A plays a role. Convergence turns out to be faster as $\rho(A)$ is smaller. For future reference, we specify the method parameters and the $\rho(A)$-value for a few methods that are suitable in shallow water computations, viz. the second-order trapezoidal rule, the second-order BDF, and the third-order Radau IIA method, respectively given by:

(2.12) $\quad s = 1, \quad \mathbf{c} = 1, \quad A = \frac{1}{2}, \quad B = 1, \quad C = \frac{1}{2}, \quad \rho(A) = \frac{1}{2},$

(2.13) $\quad s = 2, \quad \mathbf{c} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad A = \frac{1}{3}\begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}, \quad B = \frac{1}{3}\begin{pmatrix} 0 & 3 \\ -1 & 4 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \rho(A) = \frac{2}{3},$

(2.14) $\quad s = 2, \quad \mathbf{c} = \begin{pmatrix} \frac{1}{3} \\ 1 \end{pmatrix}, \quad A = \frac{1}{12}\begin{pmatrix} 5 & -1 \\ 9 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad \rho(A) = \frac{1}{6}\sqrt{6}.$

These methods are all A-stable, that is, the eigenvalues of $\partial\mathbf{f}/\partial\mathbf{y}$ are allowed to be anywhere in the left halfplane. Hence, there are no stepsize restrictions due to stability requirements. This property of unconditional stability is particularly important in shallow water applications, because (as already remarked) many of the eigenvalues of J_k, k = 1, 2, 3, are close to the imaginary axis. Furthermore, the BDF (2.13) has the greatest $\rho(A)$-value, the Radau IIA method (2.14) the smallest. This raises the question whether we can construct (A-stable) integration methods whose $\rho(A)$-values are as small as possible. Although this topic is outside the scope of this paper, we want to illustrate by a simple example that the construction of integration methods with reduced $\rho(A)$-values is certainly feasible. Consider the one-parameter family of three-step, second-order BDF type methods

(2.15) $$\mathbf{y}_{n+1} = (3 - \tfrac{5}{2}b_0)\mathbf{y}_n + (4b_0 - 3)\mathbf{y}_{n-1} + (1 - \tfrac{3}{2}b_0)\mathbf{y}_{n-2} + b_0\Delta t\mathbf{f}(t_{n+1},\mathbf{y}_{n+1}).$$

When written in the form (2.8), the matrix A has zero entries except for its last diagonal entry which equals b_0. Hence, we are looking for the smallest value of b_0 such that (2.15) is still A-stable (and hence L-stable). However, firstly, we should impose the condition of zero-stability. This leads to the condition $0 < b_0 < 1$. Next we consider the stability region of (2.15). Proceeding as in Section 2.1.1, we apply (2.15) to the stability test equation and we use the boundary locus method to find that the boundary of the stability region in the complex z-plane is defined by

$$z = z(\phi,b_0) := \frac{1}{b_0}\left(1 - (3 - \tfrac{5}{2}b_0)e^{-i\phi} - (4b_0 - 3)e^{-2i\phi} - (1 - \tfrac{3}{2}b_0)e^{-3i\phi}\right), \quad 0 \le \phi \le 2\pi.$$

We have A-stability if $\mathrm{Re}(z(\phi,b_0)) \ge 0$ for $0 \le \phi \le 2\pi$. An elementary calculation reveals that this is true if $\frac{3}{5} \le b_0 < 1$. Thus, the minimal A-stable and zero-stable value of $b_0 = \rho(A)$ equals $\frac{3}{5}$. Although

this is only 10% smaller than the BDF (2.13), it indicates that further reductions might be obtained by considering more general linear multistep methods.

3. Iteration methods based on factorization

We shall construct and analyse iterative methods for solving the linear system (2.10) exploiting the fact that the matrix M can be factorized using the splitting of the righthand side function in (1.1).

3.1. Approximate factorization

In [10], [13] and [14] we used the iteration process

$$\Pi\left(\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)}\right) = -\mathbf{R}(\mathbf{Y}^{(j-1)}), \quad j = 1, 2, \ldots, m, \tag{3.1}$$
$$\Pi := \left(I - A\otimes\Delta t J_1\right)\left(I - A\otimes\Delta t J_2\right)\left(I - A\otimes\Delta t J_3\right),$$

for solving the implicit relations resulting from the application of the backward differentiation and Radau methods (2.13) and (2.14) to the three-dimensional chemistry-transport problem in shallow water. In (3.1) the initial iterate $\mathbf{Y}^{(0)}$ should be provided by some predictor formula and the number of iterations m is assumed to be determined by some iteration strategy such that $\mathbf{Y}^{(m)}$ may be considered as the solution of $\mathbf{Y}_{n+1}$ of (2.8). The matrix Π can be seen as an approximate factorization of the matrix M used in the full Newton process (2.10). Therefore, we shall call (3.1) *approximate factorization iteration*, briefly, the AF process.

If the iterates $\mathbf{Y}^{(j)}$ converge and if (2.8) has a unique solution $\mathbf{Y}_{n+1}$, then they can only converge to $\mathbf{Y}_{n+1}$. Each iteration in (3.1) requires the solution of 3 linear systems with system matrices $I - A\otimes\Delta t J_k$, k = 1, 2, 3, each of order sN. Note that the three LU-decompositions of these system matrices can be done in parallel. Of particular interest is the case where the matrix A is diagonal (as in (2.12) and (2.13)). Then, the LU-decompositions of $I - A\otimes\Delta t J_k$ and the corresponding forward-backward substitutions are relatively cheap, because the matrices J_k each correspond with a one-dimensional differential operator. If A is a full matrix as in (2.14), then it is recommendable to replace the matrix A in the iteration matrix Π by a diagonalizable matrix A^*. This approach was considered in [10] and [6]. The case of lower triangular A will be considered in a forthcoming paper [8]. In this paper, it will from now on be assumed that A is diagonal with nonnegative diagonal entries.

In the following section we present convergence and stability results for the AF process.

3.1.1. Convergence results. Let us consider the AF iteration error $\varepsilon^{(j)} := \mathbf{Y}^{(j)} - \mathbf{Y}_{n+1}$. From (2.8), (2.9), (2.11) and (3.1) it follows that

$$\varepsilon^{(j)} = Z\,\varepsilon^{(j-1)} + \Delta t\,\Pi^{-1}(A\otimes I)\left(\Phi_1(\varepsilon^{(j-1)}) + \Phi_2(\varepsilon^{(j-1)}) + \Phi_3(\varepsilon^{(j-1)}) + \Phi_4(\varepsilon^{(j-1)})\right),$$
$$Z := I - \Pi^{-1}M,$$
$$\Phi_k(\varepsilon) := \mathbf{F}_k(\mathbf{e}t_n + \mathbf{c}\Delta t, \mathbf{Y}_{n+1} + \varepsilon) - \mathbf{F}_k(\mathbf{e}t_n + \mathbf{c}\Delta t, \mathbf{Y}_{n+1}) - (I\otimes J_k)\varepsilon, \quad k = 1, 2, 3,$$
$$\Phi_4(\varepsilon) := \mathbf{F}_4(\mathbf{e}t_n + \mathbf{c}\Delta t, \mathbf{Y}_{n+1} + \varepsilon) - \mathbf{F}_4(\mathbf{e}t_n + \mathbf{c}\Delta t, \mathbf{Y}_{n+1}).$$

Here, the functions $\mathbf{F}_k$ are defined in a similar way as the function **F**. Evidently,

(3.2) $\Phi_k(\varepsilon) = K_k\varepsilon + O(\varepsilon^2)$,

where the matrices K_k are s-by-s block-diagonal matrices with blocks of the same dimension as the matrix J. Hence, we obtain

(3.3) $\varepsilon^{(j)} = \left(Z + \Delta t\, \Pi^{-1}(A\otimes I)(K_1 + K_2 + K_3 + K_4\right) \varepsilon^{(j-1)} + O\left((\varepsilon^{(j-1)})^2\right)$.

The matrices K_k may be assumed to be of small magnitude. Hence, the error recursion (3.3) is essentially given by

(3.3') $\varepsilon^{(j)} \approx Z\, \varepsilon^{(j-1)}$, $j = 1, 2, \ldots, m$,

so that our first concern is the matrix Z. Part of the properties of Z given below were derived in [6] and are reproduced here (without proof) for reference reasons.

Theorem 3.1. The amplification matrix Z in (3.3') satisfies $Z = O((\Delta t)^2)$. ♦

This theorem shows that we always have convergence if Δt is sufficiently small, that is, the convergence region is never empty. Furthermore, Theorem 3.1 indicates that the nonstiff error components (corresponding with eigenvalues of J_k of modest magnitude) are rapidly removed from the iteration error.
The following theorem provides information on the size of the convergence region, that is, the region in the space $\Delta t(\lambda(J_1),\lambda(J_2),\lambda(J_3))$, where the amplification factors $\lambda(Z)$ are within the unit circle. As before, we consider the case where the Jacobian matrices J_k share the same eigensystem.

Theorem 3.2. Necessary and sufficient conditions for $\rho(Z) < 1$ are given by $|\arg(-\lambda(J_k))| \le \frac{\pi}{4}$, $k = 1, 2, 3$. ♦

Recalling that the eigenvalues of J_1, J_2 and J_3 are often close to the imaginary axis, we need more information on the amplification factors $\lambda(Z)$ than provided by Theorem 3.2. We again consider the most critical case where the eigenvalues of J_k are purely imaginary. Let us write $\Delta t\lambda(A)\lambda(J_k) = i\zeta_k$ and $\lambda(Z) = C(\zeta_1,\zeta_2,\zeta_3)$, where ζ_k is real-valued. Then $|C|$ is given by

(3.4) $$\left| C(\zeta_1,\zeta_2,\zeta_3) \right| = \left(\frac{(\zeta_1 + \zeta_2)^2\zeta_3^2 + \zeta_1^2\zeta_2^2(\zeta_3^2 + 1) + 2\zeta_1\zeta_2\zeta_3(\zeta_1 + \zeta_2)}{(1 + \zeta_1^2)(1 + \zeta_2^2)(1 + \zeta_3^2)} \right)^{1/2} .$$

This expression enables us to derive an upperbound for the amplification factors $\lambda(Z)$ which is quite accurate for the nonstiff $\lambda(Z)$.

Theorem 3.3. The amplification factors $\lambda(Z)$ corresponding with eigenvalues $\lambda(J_k)$ lying in the interval i[-δ,δ] satisfy the estimate

$$|\lambda(Z)| \le 3\zeta^2\sqrt{1 + \tfrac{1}{9}\zeta^2}, \quad \zeta := \rho(A)\delta\Delta t. \blacklozenge$$

This theorem shows that the nonstiff amplification factors $\lambda(Z)$, that is, amplification factors corresponding with small values of δ, are quite small.

Next, we consider the overall convergence. Taking into account that we do not want to restrict the magnitude of the eigenvalues $\lambda(J_3)$ on the imaginary axis, we are interested in the region in the (ζ_1,ζ_2)-plane where the amplification factor

$$\alpha(\zeta_1,\zeta_2) := \max_{-\infty\le\zeta_3\le\infty} \left| C(\zeta_1,\zeta_2,\zeta_3) \right|$$

is less than 1. If $\alpha(\zeta_1,\zeta_2) < 1$ is satisfied for $-\gamma < \zeta_k < \gamma$, k = 1, 2, then in the imaginary $\Delta t\big(\lambda(J_1), \lambda(J_2), \lambda(J_3)\big)$ - space the corresponding region of convergence of (3.1) is given by

(3.5a) $$\mathbb{C}_{AF} := \{(y_1,y_2,y_3): \ |y_k| \le \frac{\gamma}{\rho(A)}, \ k = 1, 2; |y_3| \le \infty\},$$

where γ will be called the *convergence boundary*. We found that $\alpha(\zeta_1,\zeta_2)$ is given by

$$\alpha(\zeta_1,\zeta_2) = \max\left\{ \left| C(\zeta_1,\zeta_2,0) \right|, \left| C(\zeta_1,\zeta_2,\zeta_\pm(\zeta_1,\zeta_2)) \right|, \left| C(\zeta_1,\zeta_2,\infty) \right| \right\},$$

$$\zeta_\pm(\zeta_1,\zeta_2) := \frac{\zeta_1 + \zeta_2 \pm \sqrt{(\zeta_1 + \zeta_2)^2 + 4\zeta_1{}^2\zeta_2{}^2}}{2\zeta_1\zeta_2} \ .$$

It is easily verified that $|C(\zeta_1,\zeta_2,0)| < 1$ for all ζ_1 and ζ_2, and that $|C(\zeta_1,\zeta_2,\infty)| < 1$ on the domain $\zeta_1\zeta_2 < \frac{1}{2}$, which contains the square $-\frac{1}{2}\sqrt{2} < \zeta_k < \frac{1}{2}\sqrt{2}$, k = 1, 2. Furthermore, $|C(-\zeta_1,-\zeta_2,\zeta_+(\zeta_1,\zeta_2))| = |C(\zeta_1,\zeta_2,\zeta_-(\zeta_1,\zeta_2))|$, so that we may confine our considerations to $|C(\zeta_1,\zeta_2,\zeta_-(\zeta_1,\zeta_2))|$ in the square $-\frac{1}{2}\sqrt{2} < \zeta_k < \frac{1}{2}\sqrt{2}$, k= 1, 2. Figure 3.1 shows the function $|C(\zeta_1,\zeta_2,\zeta_-(\zeta_1,\zeta_2))|$. This picture indicates that this function increases most rapidly along the line $\zeta_1 = \zeta_2$. Hence, the convergence boundary $\gamma = \min\{\frac{1}{2}\sqrt{2}, \gamma_0\}$, where γ_0 is the smallest positive root of the equation $\alpha(\zeta,\zeta) = 1$. This equation is given by $4\zeta^8 + 8\zeta^6 + 4\zeta^4 - \zeta^2 = 1$, so that $\gamma_0 = 0.647...$.

Theorem 3.4. Let the eigenvalues of J_k, k = 1, 2, 3, be purely imaginary. Then, a sufficient condition for $\rho(Z) < 1$ is:

$$\Delta t \le \frac{\gamma}{\rho(A)\max\{\rho(J_1),\rho(J_2)\}}, \quad \gamma = 0.647... \ . \blacklozenge$$

In order to get more detailed information on the actual region of convergence, we define for a given value of y_3 the region (compare the region $\mathbb{S}(y_3)$ introduced in (2.1b))

(3.5b) $$\mathbb{C}_{AF}(y_3) := \{(y_1,y_2):\ |y_k| \le \frac{\gamma(y_3)}{\rho(A)},\ k = 1, 2\},$$

where the convergence boundary $\gamma(y_3)$ is defined by the requirement that $\rho(Z) < 1$ in $\mathbb{C}_{AF}(y_3)$. We verified that $\gamma(y_3)$ is determined on the line $\zeta_1 = \zeta_2$, that is, by the equation $\left| C(\zeta_1,\zeta_1,\zeta_3) \right| = 1$. This equation can be reduced to $4\zeta_1^3\zeta_3 + 2(\zeta_3^2 - 1)\zeta_1^2 - \zeta_3^2 - 1 = 0$. Hence, defining the function g(x) by the relation $4xg^3(x) + 2(x^2 - 1)g^2(x) - x^2 - 1 = 0$, it follows that (recall that A is diagonal with nonnegative diagonal entries, so that $\lambda(A) \ge 0$)

(3.5c) $$\gamma(y_3) = \rho(A) \min_{\lambda(A)} \frac{g(\lambda(A)\left| y_3 \right|)}{\lambda(A)}.$$

Figures 3.2 presents a plot of g(x) on the interval [0,50].

3.1.2. Stability results. Evidently, if AF iteration converges, then the stability is determined by the stability of the underlying method (2.8). Hence, with respect to the stability test equation, the stability region of the iterated method converges to the cross section of the convergence region and the stability region of (2.8). Thus, the stability region is given by

(3.6) $$\mathbb{S}_{AF} := \mathbb{S}_0 \cap \mathbb{C}_{AF},$$

where $\mathbb{S}_0$ is the stability region of (2.8) and $\mathbb{C}_{AF}$ is defined by (3.5a) with $\gamma\rho^{-1}(A) \approx 0.647\rho^{-1}(A)$ (see Theorem 3.4). For A-stable integration methods, the stability region $\mathbb{S}_{AF}$ equals the convergence region $\mathbb{C}_{AF}$. Hence, the stability region $\mathbb{S}_{AF}$ and the corresponding stability condition are therefore respectively of the form (2.1a) and (2.2) with $\beta = \gamma\rho^{-1}(A)$. For example, for the trapezoidal, BDF and Radau methods (2.12), (2.13) and (2.14), we find $\beta \approx 1.29$, 0.97 and 1.58, respectively. Thus, the stability of these AF iterated methods compares favourably with the stability of the implicit-explicit BDF and the Douglas-Yanenko methods.

In (3.6) it is assumed that the AF process is iterated until convergence. However, by virtue of Theorem 3.1, the order of the underlying integration formula is already reached after a few iterations. Therefore, it is tempting to stop in each step the iteration process after two or three iterations, rather than trying to solve (2.8). Unfortunately, after a fixed number of iterations, the stability of the iterated method is still rather poor. In [6] it was shown that the AF iterated BDF (2.13) is stable in a region of the form (2.1a) with $\beta(m) \approx 0.3$ for $m \le 4$. Hence, in order to rely on the stability boundary valid for the 'converged' method, one should not iterate with a *fixed* number of iterations, but employ some form of iteration strategy which guarantees that sufficiently many iterations are performed to remain close to the solution of (2.8). In practice, the averaged number of iterations per step will still be modest (in the range of 2 until 4 iterations), but the iteration strategy serves to perform additional iterations when necessary.

Although AF iteration leads to stability regions of the form (2.1a) with reasonable values for β, it would be desirable to have still larger stability boundaries. Therefore, we look for iteration methods yielding larger stability regions than AF iteration.

3.2. Safety net iteration process

We shall consider an iteration strategy where we perform only a few AF iterations with (3.1) and where we continue with a second iteration process subject to less severe convergence conditions than AF iteration. Since by virtue of Theorem 3.1 we have $Z = O((\Delta t)^2)$, the first few iterations serve to achieve an acceptable order of accuracy with respect to Δt with an order constant of reasonable size. The second iteration process should serve to achieve convergence in a larger eigenvalue domain, that is, a less severe timestep restriction than the one given in Theorem 3.4. Thus, in a dynamic iteration strategy, where the number of iterations is determined by for example the size of the residue term, this second iteration process acts as a safety net that should ensure a more or less monotonic convergence. Therefore, we shall call it a *safety net* iteration process, or briefly, SN iteration.

Let us define $\Pi_{k3} := (I - A\otimes\Delta t J_k)(I - A\otimes\Delta t J_3)$. Then, SN iteration is defined by

$$\begin{aligned}\Pi_{23}\left(\mathbf{Y}^{(j-1/2)} - \mathbf{Y}^{(j-1)}\right) &= -\mathbf{R}\left(\mathbf{Y}^{(j-1)}\right) - \omega\Delta t(A\otimes I)\left[\mathbf{F}_1\left(\mathbf{e}t_n+\mathbf{c}\Delta t,\mathbf{Y}^{(j-1)}\right) - \mathbf{F}_1\left(\mathbf{e}t_n+\mathbf{c}\Delta t,\mathbf{Y}^{(m)}\right)\right],\\ \Pi_{13}\left(\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1/2)}\right) &= -\mathbf{R}\left(\mathbf{Y}^{(j-1/2)}\right) - \omega\Delta t(A\otimes I)\left[\mathbf{F}_2\left(\mathbf{e}t_n+\mathbf{c}\Delta t,\mathbf{Y}^{(j-1/2)}\right) - \mathbf{F}_2\left(\mathbf{e}t_n+\mathbf{c}\Delta t,\mathbf{Y}^{(m)}\right)\right]\end{aligned} \tag{3.7}$$

for $j = m+1, m+2, \ldots, m^*$. Here, $\mathbf{Y}^{(m)}$ is the last iterate obtained by the AF method (3.1) and ω may be considered as a relaxation parameter which is assumed in [0,1]. Note that the matrices Π_{k3} are less accurate factorizations of the matrix M than the matrix Π in the AF method.

If the iterates $\mathbf{Y}^{(j)}$ converge to a vector $\mathbf{U}_{n+1} = \mathbf{Y}^{(\infty)}$, then $\mathbf{U}_{n+1}$ is a zero of the new residue function $\mathbf{R}^*(\mathbf{Y})$. This function is obtained on substitution of $\mathbf{Y}^{(j-1)} = \mathbf{Y}^{(j)} = \mathbf{Y}$ into (3.7) and on elimination of $\mathbf{Y}^{(j-1/2)}$. Evidently, $\mathbf{U}_{n+1}$ depends on ω and $\mathbf{Y}^{(m)}$. For $\omega = 0$ we have $\mathbf{U}_{n+1} = \mathbf{Y}_{n+1}$. For $\omega \neq 0$ we have an iteration defect $\mathbf{U}_{n+1} - \mathbf{Y}_{n+1}$. In the next section we shall derive an estimate for this defect.

Compared with the iterations in the AF method, the iterations in the SN method are more expensive because in each iteration we have to solve four instead of three linear systems and to evaluate twice instead of once a residue function. Moreover, it requires more storage, because $\mathbf{F}_1(\mathbf{e}t_n + \mathbf{c}\Delta t,\mathbf{Y}^{(m)})$ and $\mathbf{F}_2(\mathbf{e}t_n + \mathbf{c}\Delta t,\mathbf{Y}^{(m)})$ have to be saved. To be more precise, in the case of the two-species application described in Section 4, the amount of storage increases by about 10% (the percentage of additional storage increases to at most 25% as the number of species increases). Thus, when compared with the AF process, the main drawback of the SN process is the relatively expensive iteration cost. However, we shall see that SN iteration converges much faster than AF iteration.

3.2.1. Convergence. In order to get insight into the convergence of SN iteration, we consider the corresponding error recursion. Omitting second-order error terms, we find

$$\varepsilon^{(j-1/2)} = \left(\bar{Z}_1(\omega) + \Delta t \Pi_{23}^{-1}(A\otimes I)\left((1-\omega)K_1 + K_2 + K_3 + K_4\right)\right)\varepsilon^{(j-1)} + \omega\left(Q_1 + \Delta t\, \Pi_{23}^{-1}(A\otimes I)K_1\right)\varepsilon^{(m)},$$

(3.8)
$$\varepsilon^{(j)} = \left(\bar{Z}_2(\omega) + \Delta t \Pi_{13}^{-1}(A\otimes I)\left(K_1 + (1-\omega)K_2 + K_3 + K_4\right)\right)\varepsilon^{(j-1/2)} + \omega\left(Q_2 + \Delta t\, \Pi_{13}^{-1}(A\otimes I)K_2\right)\varepsilon^{(m)},\ j > m,$$

where

$$\bar{Z}_1(\omega) := I - \Pi_{23}^{-1}M - \omega Q_1, \quad \bar{Z}_2(\omega) := I - \Pi_{13}^{-1}M - \omega Q_2,$$

$$Q_1 := \Delta t\, \Pi_{23}^{-1}(A\otimes J_1), \qquad Q_2 := \Delta t\, \Pi_{13}^{-1}(A\otimes J_2).$$

Since the matrices K_k are of small magnitude, the error recursion (3.8) essentially behaves as

(3.8')
$$\varepsilon^{(j)} \approx \bar{Z}(\omega)\,\varepsilon^{(j-1)} + \omega Q(\omega)\,\varepsilon^{(m)}, \quad \bar{Z}(\omega) := \bar{Z}_2(\omega)\bar{Z}_1(\omega),\ Q(\omega) := Q_2 + \bar{Z}_2(\omega)Q_1,\ j > m.$$

It is easily verified that $\bar{Z}_k(\omega) = \Delta t\,(A\otimes(1-\omega)J_k) + O((\Delta t)^2)$, leading to the result (cf. Theorem 3.1):

Theorem 3.5. The amplification matrix $\bar{Z}(\omega)$ in (3.8') satisfies

$$\bar{Z}(\omega) = (\Delta t)^2\left((1-\omega)^2\,(A^2\otimes J_2J_1) + (1-\omega)O(\Delta t) + O((\Delta t)^2)\right).\ \blacklozenge$$

Hence, for the nonstiff error components we always have $O((\Delta t)^2)$ convergence and even $O((\Delta t)^4)$ convergence as $\omega \to 1$.

Next, we look at the region of convergence. A necessary and sufficient condition for convergence requires the eigenvalues of $\bar{Z}(\omega)$ within the unit circle. Again, we write $\Delta t\lambda(A)\lambda(J_k) = i\zeta_k$ and $\lambda(\bar{Z}(\omega)) = C(\zeta_1,\zeta_2,\zeta_3)$. Then,

(3.9)
$$\left|\, C(\zeta_1,\zeta_2,\zeta_3)\,\right| = \left(\frac{((1-\omega)^2\zeta_1^2 + \zeta_3^2\zeta_2^2)((1-\omega)^2\zeta_2^2 + \zeta_3^2\zeta_1^2)}{(1+\zeta_1^2)(1+\zeta_2^2)(1+\zeta_3^2)^2}\right)^{1/2}.$$

The analogue of Theorem 3.3 becomes

Theorem 3.6. The amplification factors $\lambda(\bar{Z}(\omega))$ corresponding with eigenvalues $\lambda(J_k)$ lying in the interval $i[-\delta,\delta]$ satisfy the estimate

$$|\lambda(\bar{Z}(\omega))| \le \zeta^2\,((1-\omega)^2 + \zeta^2), \quad \zeta := \rho(A)\delta\Delta t.\ \blacklozenge$$

A comparison with Theorem 3.3 shows that the rate of convergence of SN iteration is considerably larger than that of AF iteration, particularly as $\omega \to 1$. Since in the Theorems 3.3 and 3.6 δ refers to an arbitrary large eigenvalue interval, this statement also applies to the stiff error components.

Next we define the amplification factor

$$\alpha(\zeta_1,\zeta_2,\omega) := \max_{-\infty\le\zeta_3\le\infty} \left|\, C(\zeta_1,\zeta_2,\zeta_3)\,\right|.$$

If $\alpha(\zeta_1,\zeta_2,\omega) < 1$ is satisfied for $-\gamma(\omega) < \zeta_k < \gamma(\omega)$, k = 1, 2, then in the $\Delta t\big(\lambda(J_1), \lambda(J_2), \lambda(J_3)\big)$ - space the region of convergence of (3.7) is given by

$$(3.10) \qquad \mathbb{C}_{SN}(\omega) := \{(y_1,y_2,y_3):\ |y_k| < \frac{\gamma(\omega)}{\rho(A)},\ k = 1, 2;\ |y_3| \le \infty\}.$$

It turns out that

$$\alpha(\zeta_1,\zeta_2,\omega) = \max\ \big\{|\,C(\zeta_1,\zeta_2,0)|,\ |\,C(\zeta_1,\zeta_2,\zeta_0(\zeta_1,\zeta_2,\omega))|,\ |\,C(\zeta_1,\zeta_2,\infty)|\big\}, \qquad 0 \le \omega \le 1.$$

$$\zeta_0^2(\zeta_1,\zeta_2,\omega) := \frac{\zeta_1^4 - 2(1-\omega)^2\zeta_1^2\zeta_2^2 + \zeta_2^4}{\zeta_1^4 + \zeta_2^4},$$

It can be verified that $|\,C(\zeta_1,\zeta_2,0)| < 1$ and $|\,C(\zeta_1,\zeta_2,\infty)| < 1$ for all ζ_1 and ζ_2, provided $0 \le \omega \le 1$, so that the convergence boundary $\gamma(\omega)$ is determined by the inequality $|\,C(\zeta_1,\zeta_2,\zeta_0(\zeta_1,\zeta_2,\omega))| < 1$. We also verified that the function $|\,C(\zeta_1,\zeta_2,\zeta_0(\zeta_1,\zeta_2,\omega))|$ increases most rapidly along the ζ_1 and ζ_2 axes for all ω in [0,1]. In Figure 3.3 this is illustrated for $\omega = \frac{9}{10}$. Hence, $\gamma(\omega)$ is determined by the smallest positive root of the equation $|\,C(\zeta_1,0,\zeta_0(\zeta_1,0,\omega))| = 1$. This leads to the following analogue of Theorem 3.4:

Theorem 3.7. Let the eigenvalues of J_k, k = 1, 2, 3, be purely imaginary. Then, a sufficient condition for $\rho(\bar{Z}(\omega)) < 1$ is:

$$\Delta t \le \frac{\gamma(\omega)}{\rho(A)\max\{\rho(J_1),\rho(J_2)\}},\quad \gamma(\omega) = \frac{\sqrt{2 + 2\sqrt{1 + (1-\omega)^2}}}{1 - \omega}.\ \blacklozenge$$

For a few values of ω the convergence boundary $\gamma(\omega)$ is listed in Table 3.1. A comparison with the convergence boundary $\gamma \approx 0.64\rho^{-1}(A)$ for AF iteration given by Theorem 3.4 reveals that SN iteration has considerably larger convergence boundaries.

Table 3.1. Values of $\gamma(\omega)$, $\rho_{max}(\omega S(\omega))$ and $\rho_{aver}(\omega S(\omega))$.

σ	ω	=	0	.1	.25	.50	.75	.90	1.0
	$\gamma(\omega)$	$\approx$	2.19	2.40	2.82	4.11	8.06	20.0	∞
[0,∞]	$\rho_{max}(\omega S(\omega))$	$\le$	0	0.66	1.85	5.25	14.4	40.0	∞
10	$\rho_{aver}(\omega S(\omega))$	$\approx$	0	0.04	0.09	0.20	0.35	0.50	
20	$\rho_{aver}(\omega S(\omega))$	$\approx$	0	0.02	0.06	0.12	0.21	0.31	
40	$\rho_{aver}(\omega S(\omega))$	$\approx$	0	0.01	0.03	0.07	0.13	0.19	

However, a drawback is that we have a nonzero iteration defect $\mathbf{U}_{n+1} - \mathbf{Y}_{n+1}$, unless $\omega = 0$. From (3.8') we derive

$$(3.11) \qquad \mathbf{U}_{n+1} - \mathbf{Y}_{n+1} = \mathbf{Y}^{(\infty)} - \mathbf{Y}_{n+1} = \varepsilon^{(\infty)} \approx \omega S(\omega)\varepsilon^{(m)},\ S(\omega) := \big(I - \bar{Z}(\omega)\big)^{-1}Q(\omega).$$

We consider the effect of the SN method (3.7) on the error $\varepsilon^{(m)}$. Let us assume that the eigenvalues of J_k are more or less equally distributed and that $\rho(J_1) = \rho(J_2) = \sigma^{-1}\rho(J_3)$, where σ is the factor by which the vertically discretized terms are more stiff than the horizontally discretized terms. Because the meshsize along the vertical will be much smaller than the horizontal meshsizes, we have $\sigma >> 1$. We define for given values of ω and σ the maximum and averaged values $\rho_{max}(\omega S(\omega))$ and $\rho_{aver}(\omega S(\omega))$ of the spectral radius of the matrix $\omega S(\omega)$ in the domain

$$\mathbb{E}(\omega,\sigma) := \{(\zeta_1,\zeta_2,\zeta_3):\ |\zeta_1| \le \gamma(\omega),\ |\zeta_2| \le \gamma(\omega),\ |\zeta_3| \le \sigma\gamma(\omega)\}.$$

Table 3.1 presents upperbounds for $\rho_{max}(\sigma,\gamma)$, irrespective the value of σ, computed for a grid in $\mathbb{E}(\omega,\sigma)$ with meshsizes $\Delta\zeta_1 = \Delta\zeta_2 = 0.1$ and $\Delta\zeta_3 = 0.1\sigma$. These values are quite alarming. However, it seems more realistic to look at the values of $\rho_{aver}(\sigma,\gamma)$. These values are also listed in Table 3.1 and indicate that for, say $0 < \omega \le \frac{1}{2}$, we may expect a substantial reduction of the error $\varepsilon^{(m)}$ by applying the SN iteration process (3.7), particularly for larger values of σ, so that the iteration defect $\mathbf{U}_{n+1} - \mathbf{Y}_{n+1}$ is expected to be quite small. For $\omega > \frac{1}{2}$, the iteration defect is expected to become increasingly larger with ω. Similarly, if we look at the *nonstiff* components of the iteration defect, that is, the components which correspond with eigenvalues $\lambda(J_k)$ lying in the interval $i[-\delta,\delta]$ with δ of modest magnitude, then the nonstiff iteration defect increases about linearly with ω. This can be concluded from the following estimate for the nonstiff eigenvalues of $\omega S(\omega)$:

$$\left|\lambda(\omega S(\omega))\right| \le \omega\zeta\sqrt{1 + (10 - 6\omega + \omega^2)\zeta^2 + O(\zeta^4)}, \qquad \zeta := \rho(A)\delta\Delta t.$$

Thus, the preceding considerations lead to the conclusion that the convergence of SN iteration improves as $\omega \to 1$, but the iteration defect becomes worse.
Finally, we remark that by virtue of the above estimate for $\left|\lambda(\omega S(\omega))\right|$ and by Theorem 3.1, the order in Δt of the nonstiff components of the iteration defect $\mathbf{U}_{n+1} - \mathbf{Y}_{n+1} \approx \omega S(\omega) Z^m \varepsilon^{(0)}$ is given by $\omega O((\Delta t)^{2m+q+1})$, where q is the order of $\varepsilon^{(0)}$ with respect to Δt. Thus, even if we perform only a few AF iterations, then we already achieve a high order with respect to Δt. For example, if m = 3 and $\mathbf{Y}^{(0)} = (\mathbf{e}\mathbf{e}_s^T \otimes I)\mathbf{U}_n$, i.e. q = 1, then the nonstiff components of $\mathbf{U}_{n+1} - \mathbf{Y}_{n+1}$ are $\omega O((\Delta t)^8)$. This implies that for all ω the *smooth* part of the final solution $\mathbf{U}_{n+1}$ is very close to the smooth part of the solution $\mathbf{Y}_{n+1}$ of the underlying integration method (2.8).

3.2.2. Stability. We consider the linear stability properties of the sequence $\{\mathbf{U}_n\}$ with respect to the test equation $\mathbf{y}' = J\mathbf{y}$. Let the predictor be given by $\mathbf{Y}^{(0)} = (P \otimes I)\mathbf{U}_n$, where P is the predictor matrix. Furthermore, since now $\mathbf{Y}_{n+1}$ is the solution of (2.9) with $\mathbf{Y}_n$ replaced by $\mathbf{U}_n$, we obtain

$$\mathbf{Y}_{n+1} = M^{-1}N\mathbf{U}_n, \quad N := B \otimes I + \Delta t(C \otimes J).$$

Using (3.11), we find that

$$\begin{aligned}\mathbf{U}_{n+1} &= \mathbf{Y}_{n+1} + \varepsilon^{(\infty)} = \mathbf{Y}_{n+1} + \omega S(\omega) Z^m \varepsilon^{(0)} \\ &= \left(I - \omega S(\omega) Z^m\right)\mathbf{Y}_{n+1} + \omega S(\omega) Z^m (P \otimes I)\mathbf{U}_n = R_m(\omega)\mathbf{U}_n,\end{aligned}$$

$$R_m(\omega) := \left(I - \omega S(\omega) Z^m\right)M^{-1}N + \omega S(\omega) Z^m (P \otimes I).$$

We have stability if the stability matrix $R_m(\omega)$ has its eigenvalues on the unit disk. These eigenvalues are given by the eigenvalues of the matrix

$$\tilde{R}_m(\omega) := (I - \omega\tilde{S}\tilde{Z}^m)\tilde{M}^{-1}\tilde{N} + \omega\tilde{S}\tilde{Z}^m P, \tag{3.12}$$

where

$$\tilde{M} = I - z\,A, \quad z = z_1 + z_2 + z_3, \quad \tilde{Z} = I - (I - z_1 A)^{-1}(I - z_2 A)^{-1}(I - z_3 A)^{-1}\tilde{M},$$
$$\tilde{N} = B + z\,C, \quad \tilde{S} = \left(I - \tilde{Z}_2\tilde{Z}_1\right)^{-1}\left(\tilde{Q}_2 + \tilde{Z}_2\tilde{Q}_1\right),$$
$$\tilde{Z}_1 = I - (I - z_2 A)^{-1}(I - z_3 A)^{-1}(\tilde{M} + \omega z_1 A), \quad \tilde{Z}_2 = I - (I - z_1 A)^{-1}(I - z_3 A)^{-1}(\tilde{M} + \omega z_2 A),$$
$$\tilde{Q}_1 = z_1 A\,(I - z_2 A)^{-1}(I - z_3 A)^{-1}, \quad \tilde{Q}_2 = z_2 A\,(I - z_1 A)^{-1}(I - z_3 A)^{-1}.$$

Again confining our considerations to the most critical case where the eigenvalues of J_k are purely imaginary without restrictions on the magnitude of the eigenvalues of J_3 on the imaginary axis, the corresponding stability region will contain a domain of the form

$$\mathbb{S}_{SN}(m,\omega) := \mathbb{S}_0 \cap \mathbb{C}_{SN} \cap \mathbb{S}(m,\omega), \quad \mathbb{S}(m,\omega) := \{(y_1,y_2,y_3): |y_k| \le \beta^*,\ k = 1, 2;\ |y_3| \le \infty\}, \tag{3.13}$$

where $\mathbb{S}_0$ is the stability region of (2.8), $\mathbb{C}_{SN}$ is the convergence region of the SN method (3.7) given by (3.10) and Table 3.1, and where $\mathbb{S}(m,\omega)$ follows from the requirement $\rho(\tilde{R}_m(\omega)) \le 1$. The boundary β^* in $\mathbb{S}(m,\omega)$ depends not only on m and ω, but also on the predictor matrix P and the underlying integration method (2.8).

Table 3.2. Values of β^*, $\gamma\rho^{-1}(A)$ and β for (2.13) with m = 3 and $P := \mathbf{e}\mathbf{e}_s^T$.

ω	0	0.001	0.01	0.1	0.25	0.50	0.75	0.90	1.0
β^*	∞	3.3	0.8	0.9	0.4	0.9	1.2	3.1	78.9
$\gamma\rho^{-1}(A)$	3.2	3.2	3.2	3.6	4.2	6.1	12.0	30.0	∞
β	3.2	3.2	0.8	0.9	0.4	0.9	1.2	3.1	78.9

Let us consider the regions $\mathbb{C}_{SN}$ and $\mathbb{S}(m,\omega)$ associated with the second-order backward differentiation formula (2.13) and the predictor matrix $P := \mathbf{e}\mathbf{e}_s^T$. For m = 3 and a number of ω - values, Table 3.2 lists approximations to the values of β^*. Furthermore, we listed the values of $\gamma\rho^{-1}(A) = 3\gamma/2$ determining the convergence region $\mathbb{C}_{SN}$ (note that the convergence region $\mathbb{C}_{SN}$ is much larger than the region $\mathbb{S}(3,\omega)$, even for $\omega \approx 0$).

We are now in a position to compare the stability regions $\mathbb{S}_{AF}$ and $\mathbb{S}_{SN}(m,\omega)$ associated with the AF and AF-SN iteration processes. For A-stable integration methods, $\mathbb{S}_{SN}(m,\omega)$ and the corresponding timestep condition are respectively of the form (2.1a) and (2.2) with $\beta = \min\{\beta^*, \gamma\rho^{-1}(A)\}$. For example, for the second-order BDF (2.13), we find the values listed in Table 3.2. Since for AF iteration we have $\beta = \gamma\rho^{-1}(A) = 0.97$, we conclude that for $\omega \approx 0$ and $\omega \ge 0.9$, the AF-SN stability boundary is substantially greater than the AF stability boundary. Note that the stability boundaries for $\omega \approx 0$ and $\omega = 0.9$ are comparable. However, for $\omega \approx 0$ it is essentially determined by the convergence boundary, whereas for $\omega = 0.9$ it is determined by the value of β^*. Since convergence plays

a role in each integration step and the value of β^* is connected with the propagation of errors through the steps, it is expected that the ($\omega = 0.9$) - method will show a more robust performance than the ($\omega = 0$) - method. This is confirmed by the numerical experiments reported in the next section.

4. Numerical experiments

For our numerical experiments we use a simple transport model for two interacting species proposed in [13]. This problem consists of two PDEs in three spatial dimensions,

$$\begin{aligned} \frac{\partial c_1}{\partial t} + \mathbf{V}\cdot\nabla c_1 &= \varepsilon \Delta c_1 + g_1(t,x,y,z) - k_1 c_1 c_2, \\ \frac{\partial c_2}{\partial t} + \mathbf{V}\cdot\nabla c_2 &= \varepsilon \Delta c_2 + g_2(t,x,y,z) - k_1 c_1 + k_2(1 - c_2), \end{aligned} \tag{4.1}$$

defined on $\mathbb{D} := \{(x,y,z): 0 \le x, y \le L_h, -L_v \le z \le 0\}$, $0 \le t \le T$ with L_h, L_v, and T specified below. Here $\mathbf{V} = (u, v, w)^T$ denotes the velocity field, taken from the literature (see [5]), ε is a diffusion constant, and k_1, k_2 are reaction constants. $\mathbf{V}$ is divergence free and given in analytical form by

$$\begin{aligned} u(t,x,y,z) &= \{ \quad \tilde{y} + 3(\tilde{z} + 1/2)\,[\,(\tilde{x} - 1/6)^2 + (\tilde{y} - 1/6)^2 - p^2]\,\}\, d(t), \\ v(t,x,y,z) &= \{ -\tilde{x} + 3(\tilde{z} + 1/2)\,[\,(\tilde{x} - 1/6)^2 + (\tilde{y} - 1/6)^2 - p^2]\,\}\, d(t), \\ w(t,x,y,z) &= -3 L_v \tilde{z}(\tilde{z} + 1)\,\{(\tilde{x} - 1/6)/L_h + (\tilde{y} - 1/6)/L_h\}\, d(t), \end{aligned} \tag{4.2}$$

where we used the scaled co-ordinates $\tilde{x} := x/L_h$, $\tilde{y} := y/L_h$, $\tilde{z} := z/L_v$, and $d(t) = \cos(2\pi t/T_p)$ with p and T_p given constants. The Dirichlet boundary conditions, the initial condition and the functions g_1 and g_2 are chosen in accordance with a prescribed analytical solution, which is of the form

$$c_i(t,x,y,z) = \exp\{\, \tilde{z}/i - f_i(t) - \gamma_i\,[(\tilde{x} - r(t))^2 + (\tilde{y} - s(t))^2\,]\}, \quad i = 1, 2, \tag{4.3}$$

with $f_2(t) = t/(T_b + t)$, $f_1(t) = 4f_2(t)$, $r(t) = 1/6 + \cos(2\pi t/T_p)/40$, and $s(t) = 1/6 + \sin(2\pi t/T_p)/40$.
In our experiments, we take the following values for the various parameters (mks units):

$$\begin{aligned} &\varepsilon = 0.5, \quad k_1 = k_2 = 10^{-4}, \quad L_h = 20\,000, \quad L_v = 100, \quad T = 36000, \\ &p = 1/10, \quad T_p = 43200, \quad T_b = 32400, \quad \gamma_1 = 80, \quad \gamma_2 = 20. \end{aligned} \tag{4.4}$$

The domain $\mathbb{D}$ was subdivided into four domains separated by the planes $\tilde{x} := 1/3$ and $\tilde{y} := 1/3$. The above test problem was discretized on these four domains, each containing a spatial grid with $N_x = 61$, $N_y = 61$ and $N_z = 31$ grid points in the x-, y- and z-direction, respectively. The resulting ODE system consists of over 900 000 equations.
The semidiscrete system was integrated by the BDF (2.13). The implicit relations were approximately solved by using two iteration strategies, viz. (i) only iterating with the AF iteration process (3.1), and (ii) first performing 3 AF iterations and then continuing with the SN iteration process (3.7). The total number of iterations is denoted by m^*. To start the iteration, we use a 'trivial' prediction, i.e. $\mathbf{Y}^{(0)} := \mathbf{Y}_n$. This prediction proved to be more robust than using an extrapolation formula. The accuracy of the numerical solution is measured by the number of correct digits in the end point $t = T$, that is, by

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

taken over all grid points and over both species. Notice that the numerical solution is compared with the exact solution (4.3) of the PDE and hence comprises both spatial errors and temporal errors.
For various values of Δt, Table 4.1 and 4.2 list the cd-values obtained by iterating the BDF (2.13) with the AF process (3.1) and the AF-SN iteration process {(3.1),(3.7), m = 3}, respectively.
Our first concern is the convergence of the combined iteration process AF-SN for stepsizes that are not determined by convergence and stability requirements, but only by accuracy requirements. Of course, this property depends on the range of accuracies we are interested in. In shallow water applications, a precision of about 1% is quite realistic, hence we are aiming at cd-values of about 2.

Table 4.1. Values of cd by AF iterated BDF for problem (4.1) - (4.4).

Δt	Strategy	m = 1	m = 2	m = 3	m = 5	m = 7	m = 9	m = 11	…	m = 21
60 min.	AF	1.6	1.8	2.0	2.1	2.2	2.2	1.8	…	divergence
30 min.	AF	2.0	2.2	2.4	2.7	2.9	3.0	2.3	…	overflow
15 min.	AF	2.4	2.8	3.1	3.7	4.0	4.1	4.1	…	overflow
7.5 min.	AF	2.7	3.5	4.1	4.7	4.7	4.7	4.7	…	4.7

Table 4.2. Values of cd by AF-SN iterated BDF with m = 3 for problem (4.1) - (4.4).

Δt	Strategy	ω	$m^* = 4$	$m^* = 5$	$m^* = 6$	$m^* = 7$	…	$m^* = 12$
60 min.	AF-SN	0	2.3	- 1.2	overflow			
		0.5	2.2	- 0.5	overflow			
		0.9	1.1	1.1	1.4	1.8	…	1.5
		1.0	0.8	0.7	0.7	0.7	…	0.7
30 min.	AF-SN	0	3.1	3.5	overflow			
		0.5	2.9	2.9	2.9	0.8	…	overflow
		0.9	2.6	2.6	2.6	2.6	…	2.6
		1.0	2.6	2.6	2.5	2.5	…	2.3
15 min.	AF-SN	0	4.1	4.1	4.1	4.1	…	4.1
		0.5	3.5	3.5	3.5	3.5	…	3.5
		0.9	3.3	3.3	3.3	3.3	…	3.3
		1.0	3.3	3.3	3.3	3.3	…	3.3
7.5 min.	AF-SN	0	4.7	4.7	4.7	4.7	…	4.7
		0.5	4.2	4.2	4.2	4.2	…	4.2
		0.9	4.0	4.0	4.0	4.0	…	4.0
		1.0	4.0	4.0	4.0	4.0	…	4.0

The figures in Table 4.1 indicate that the AF process has the property of accuracy-dictated-stepsizes only in the range of about 4 or more digits accuracy, whereas Table 4.2 shows that AF-SN iteration with $\omega = 0.9$ already has this property for accuracies of 1.5 digits. Note that the ($\omega = 0.9$) - method

behaves much more stably than the ($\omega = 0$) - method, in spite of the fact that both methods possess a comparable stability boundary (see Table 3.2). As already observed, this can be explained by the better convergence properties of the ($\omega = 0.9$) - method.

Secondly, we are interested in the monotony of the convergence of AF-SN iteration. Since SN iteration should act as a savety net procedure in a dynamic iteration strategy, it would be desirable that changing from AF to SN preserves the monotony of the convergence. In Figure 4.1 we have plotted the maximum norm of $\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)}$ as a function of j for the AF and AF-SN iteration processes with $\omega = 0.9$ within a single step of 60 min. In this plot the AF process is described by the graph that initially decreases and that starts to increase with the 10th iteration. The graph of the AF-SN process shows a minor dismonotony at the third iteration. Apparently, the fourth AF-SN iteration is less accurate than the fourth AF iteration. However, the fifth AF-SN iteration is already more accurate than the fifth AF iteration. Continuing the iteration processes demonstrates increasing divergence for AF, whereas SN ($\omega = 0.9$) nicely converges to 0.

A third issue is the effect of ω on the accuracy of the solution. We recall that for $\omega = 0$ the iterated solution converges to the BDF solution (provided, of course, that the iteration process does not diverge). For $\omega > 0$, it will converge to a different solution. On the basis of the ρ_{max} and ρ_{aver} values listed in Table 3.1, we should be prepared that for larger stepsizes the difference with the BDF accuracy may be considerable for $\omega > 0.5$. This is confirmed in Table 4.2.

Summarizing, we conclude that for low accuracy computations (about 1 or 2 digits accuracy) and a dynamic iteration strategy, we should use the AF-SN process with $\omega = 0.9$. Also note that AF iteration requires about 8 times smaller stepsizes in order to remain stable.

5. Conclusions

We conclude this paper by summarizing the main properties of the second-order integration methods analysed in the preceding sections. For these methods, we have listed in Table 5.1 the order of accuracy p, the stability boundary β occurring in the stepsize condition (2.2), the number of LU-decompositions (LUDs) per update of the various Jacobian matrices (we recall that these LUDs can be done in parallel on a parallel computer system), the number of forward-backward substitutions (FBSs) per step, and the number of righthand sides (RHSs) per step in the case where the terms $\mathbf{f}_k$ in (1.1) are equally expensive. The parameter ε occurring in the stability boundary of the Douglas-Yanenko methods denotes the stability defect introduced in Section 2.1.2. Furthermore, m and m^* denote the number of iterations used in the AF and AF-SN processes. The specification dynamic m or dynamic m^* means that the iteration process is stopped if some residual tolerance is reached.

Taking into account the size of the stability boundary, we may draw the conclusion that

(i) implicit-explicit BDF and Douglas-Yanenko are not suitable for shallow water applications.
(ii) AF iterated BDF and AF-SN iterated BDF are both suitable for shallow water applications.
(iii) AF-SN iterated BDF is by far superior to AF iterated BDF.

Table 5.1. Properties of various numerical methods for integrating shallow water problems.

Method	p	β	LUDs	FBSs	RHSs
Implicit-explicit BDF (2.3)	2	0	1	m	$\frac{1}{4}$(m + 4)
Douglas-Yanenko (2.4) - (2.5) with $\theta = \frac{1}{2}$	2	$\sqrt{(2\varepsilon)}$	3	3m	1 + m
AF-iterated BDF (2.13) - (3.1) with fixed m = 3	2	0.30	3	9	3
AF-iterated BDF (2.13) - (3.1) with dynamic m	2	0.97	3	3m	m
AF-SN-iterated BDF {(2.13),(3.1),(3.7), $\omega = 0.9$} with m = 3 and dynamic m^*	2	3.10	3	$4m^* - 3$	$2m^* - 3$

References

[1] Ascher, U.M., Ruuth, S.J. & Wetton, B. [1997]: Implicit-explicit methods for time-dependent PDEs, SIAM J. Numer. Anal. 32, 797-823.

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

[3] Butcher, J.C. [1987]: The Numerical Analysis of Ordinary Differential Equations, Runge-Kutta and General Linear Methods, Wiley.

[4] Douglas, J. Jnr. [1962]: Alternating direction methods for three space variables, Num. Math. 4, 41-63.

[5] Dunsbergen, D. [1994]: Particle models for transport in three-dimensional shallow water flow, Ph. D. Thesis, Delft Technical University.

[6] Eichler-Liebenow, C., Houwen, P.J. van der & Sommeijer, B.P. [1998]: Analysis of approximate factorization in iteration methods, Appl. Numer. Math. 28, 245-258.

[7] Frank, J., Hundsdorfer, W. & Verwer, J.G. [1997]: On the stability of implicit-explicit linear multistep methods, Appl. Numer. Math. 25, 193-205.

[8] Houwen, P.J. van der & Sommeijer [1999]: Factorization in block-triangular implicit methods for shallow water applications, in preparation.

[9] Houwen, P.J. van der & Sommeijer [1997]: Splitting methods for three-dimensional transport models with interaction terms, J. Scientific Comput. 12, 215-231.

[10] Houwen, P.J. van der, Sommeijer, B.P. & Kok, J. [1997]: The iterative solution of fully implicit discretizations of three-dimensional transport models, Appl. Numer. Math. 25, 243-256.

[11] Hundsdorfer, W. [1998]: A note on the stability of the Douglas splitting method, Math. Comp. 67, 183-190.

[12] Maple for Macintosh V 5.4 [1996], Waterloo Maple Inc., Canada.

[13] Sommeijer, B.P. [1998]: The iterative solution of fully implicit discretizations of three-dimensional transport models, Proceedings of the 10th Int. Conf. on Parallel CFD, May 1998, Hsinchu, Taiwan.

[14] Sommeijer, B.P. & Kok, J. [1997]: Domain decomposition for an implicit shallow-water transport solver. In: B.Hertzberger & P. Sloot (eds.), Proceedings of the HPCN Europe 1997 Conference, April 1997, Vienna, Lect. Notes in Comp. Science 1225, Springer, 379-388.

[15] Verwer, J.G., Hundsdorfer, W. & Blom, J.G. [1998]: Numerical time integration for air pollution models, in preparation.

[16] Vreugdenhil, C.B. [1994]: Numerical methods for shallow-water flow, Kluwer.

[17] Yanenko, N.N. [1971]: The method of fractional steps, Springer-Verlag, Berlin.

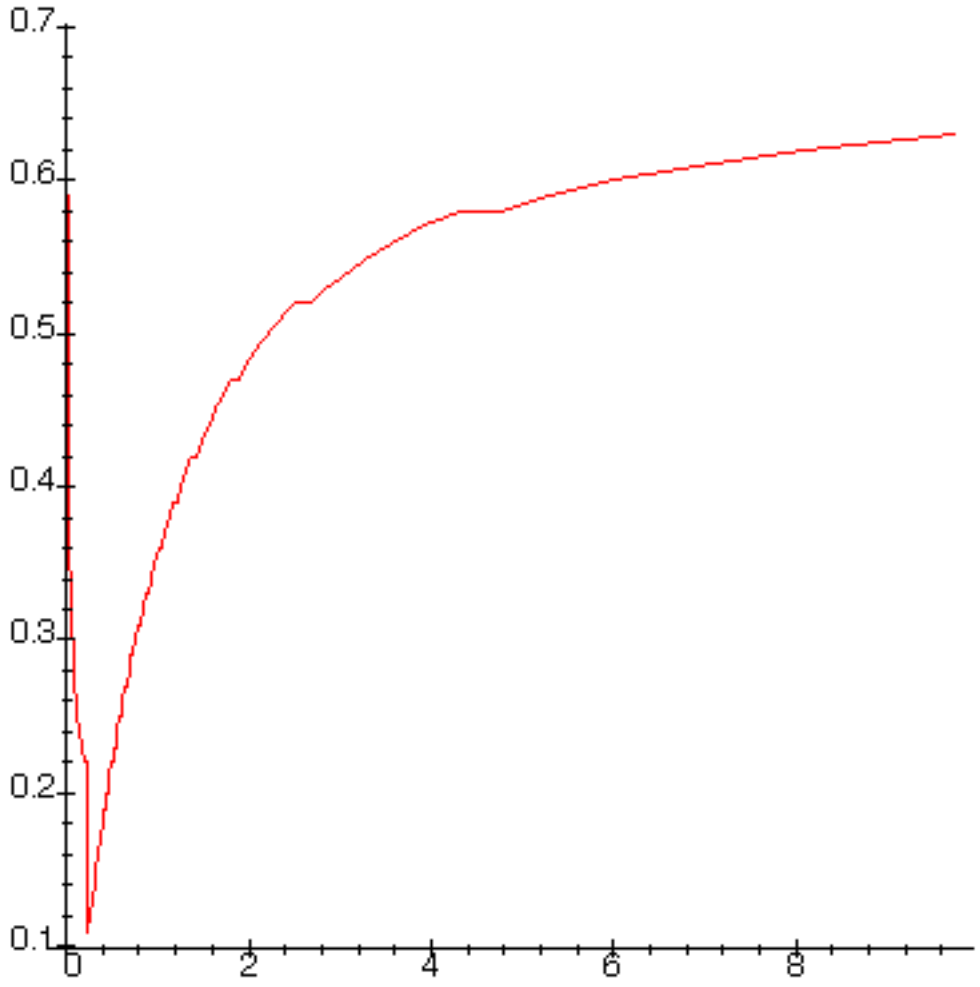

Figure 2.1a. Overall view of $\beta(y_3)$ for the Douglas-Yanenko methods (2.4) - (2.5) with $\theta = \frac{3}{5}$.

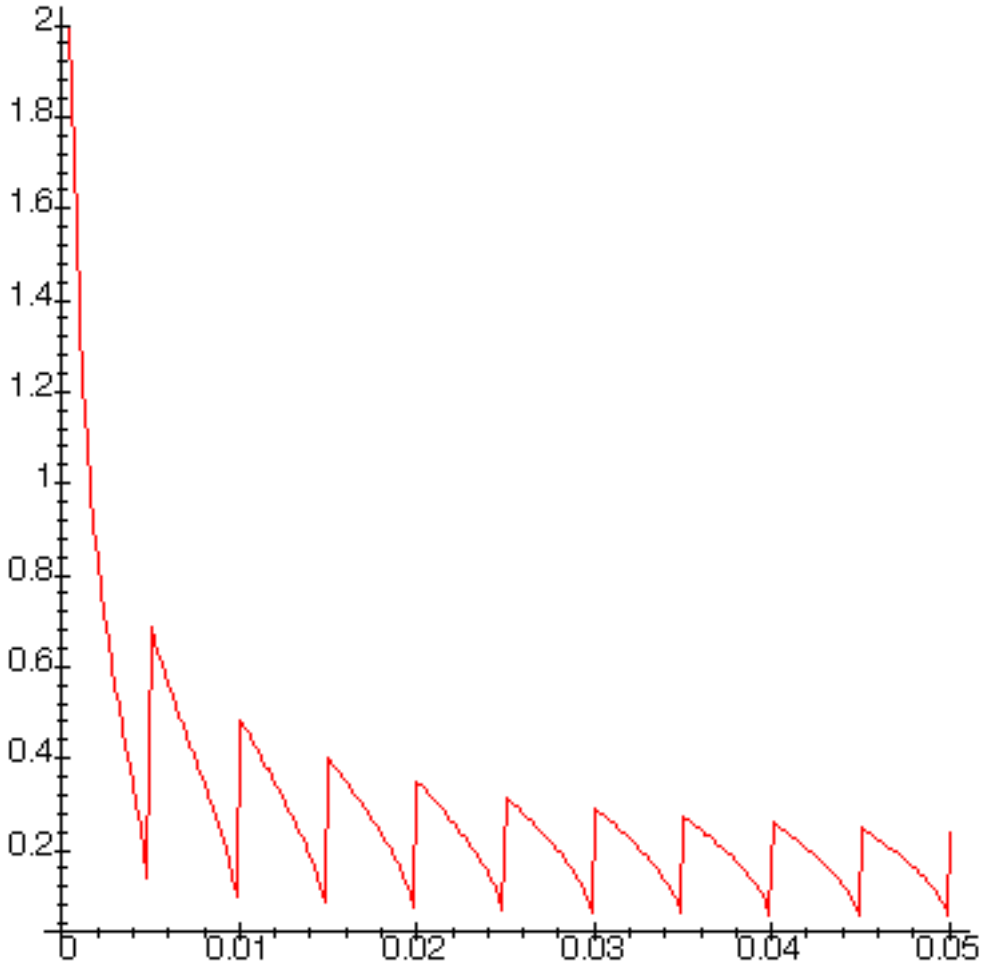

Figure 2.1b. $\beta(y_3)$ at the origin for the Douglas-Yanenko methods (2.4) - (2.5) with $\theta = \frac{3}{5}$.

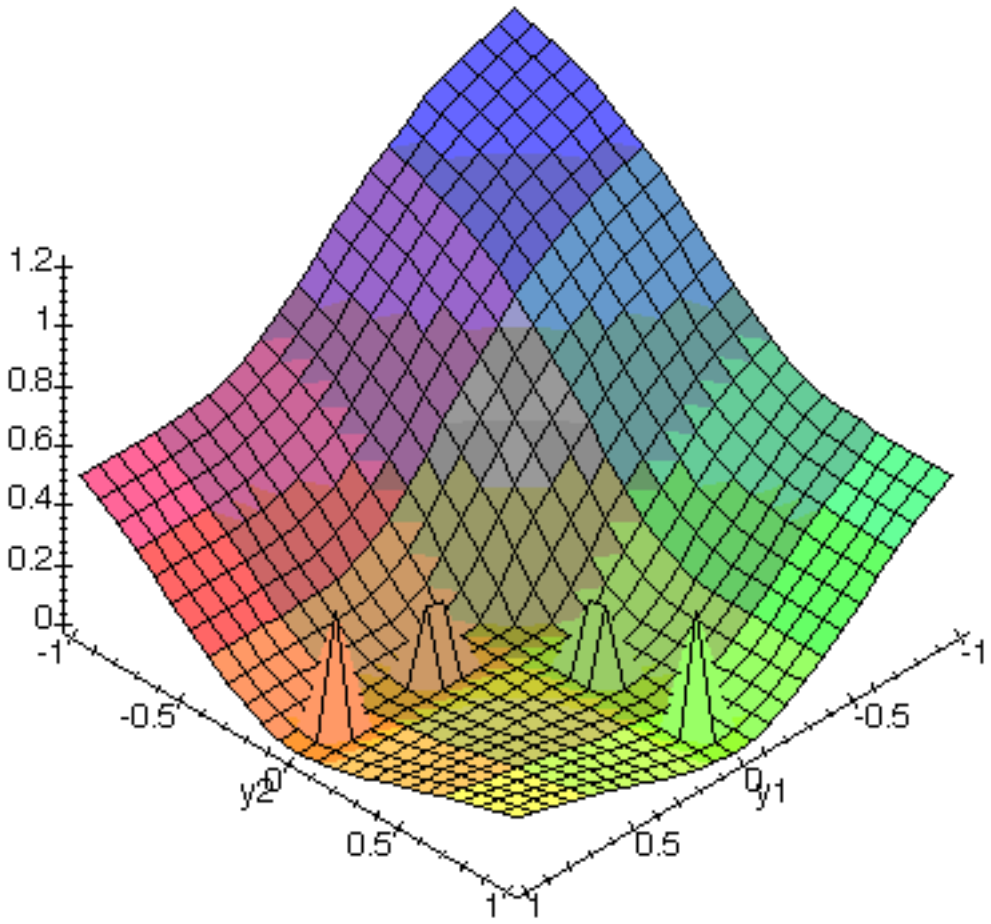

Figure 3.1. The function $|C(y_1,y_2,\zeta_-(y_1,y_2))|$.

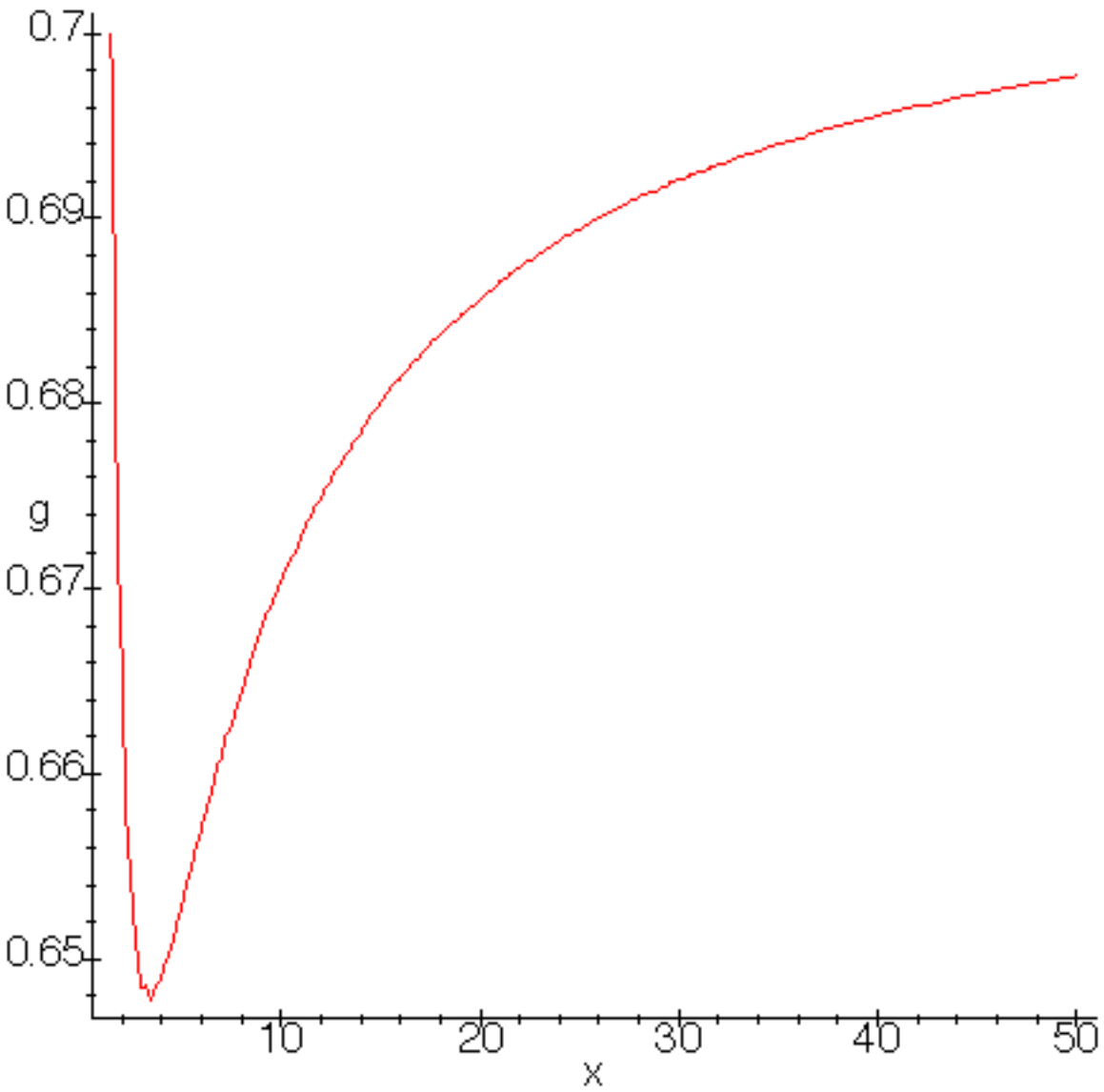

Figure 3.2. Overall view of the function g(x).

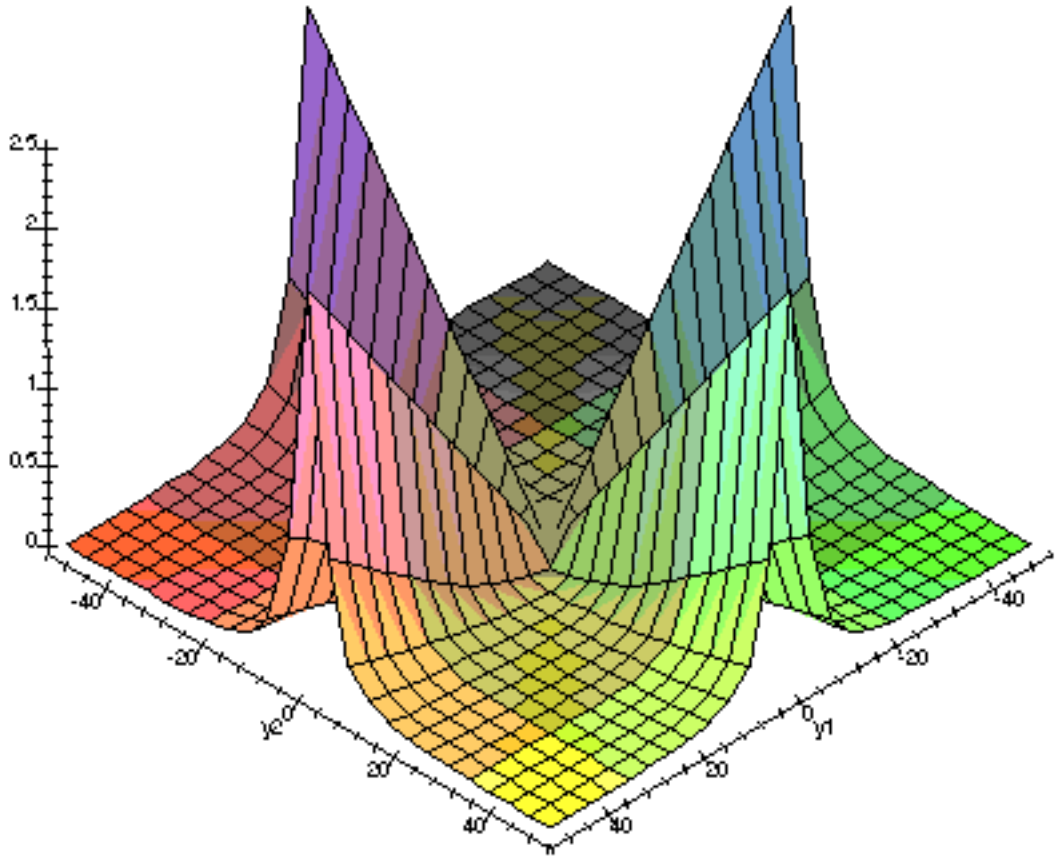

Figure 3.3. The function $|C(y_1,y_2,\zeta_0(y_1,y_2,\omega))|$.

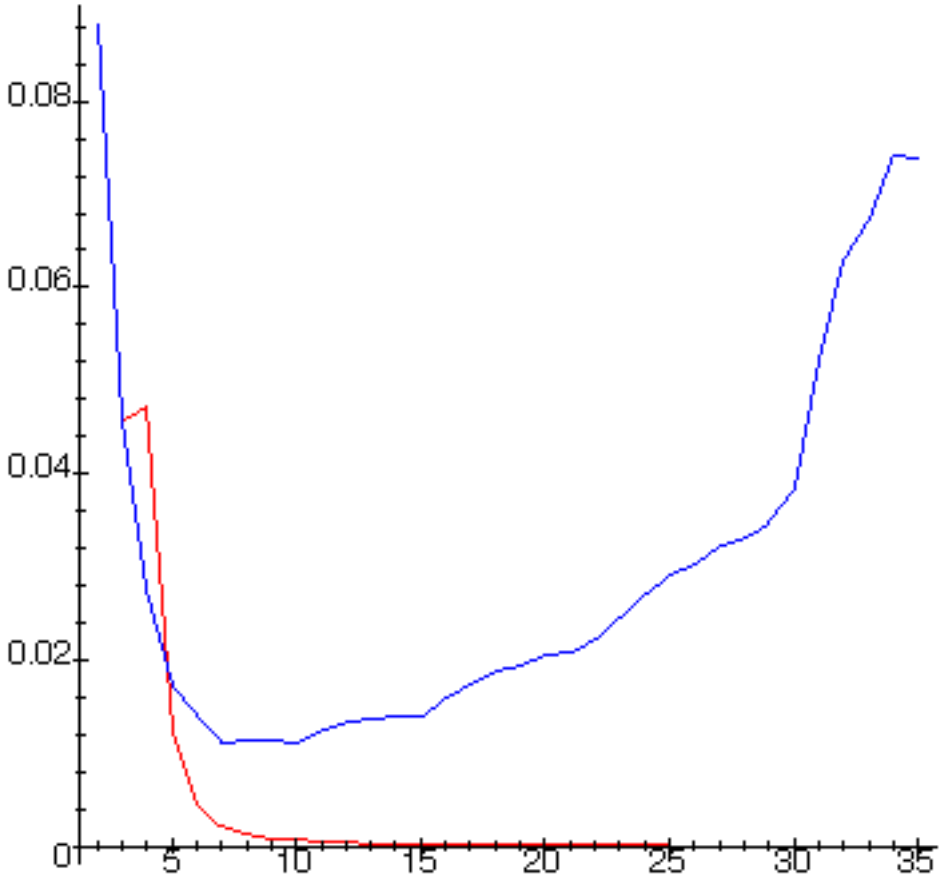

Figure 4.1. $\| \mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)} \|_{\infty}$ as a function of j for AF and AF-SN iteration.