Centrum voor Wiskunde en Informatica
**REPORT***RAPPORT*

# Solving implicit differential equations on parallel computers

P.J. van der Houwen and W.A. van der Veen

Department of Numerical Mathematics

**NM-R9526 1995**

Report NM-R9526
ISSN 0169-0388

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

# Solving Implicit Differential Equations on Parallel Computers

P.J. van der Houwen & W.A. van der Veen
*CWI*
*P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*

**Abstract**

We construct and analyse three methods for solving initial value problems for implicit differential equations (IDEs) on parallel computer systems. The first IDE method can be applied to general IDEs of higher index, the other two methods can be applied to partitioned (or semi-explicit) IDEs. The partitioned IDE methods both exploit the special form of the problem and often converge faster than the general IDE method. The first partitioned IDE method is suitable for higher-index problems, the second partitioned IDE method only applies to index 1 problems, but possesses more parallelism across the method. The convergence of these methods is illustrated by solving implicit IDEs of index 0 until 3 that are taken from the literature.

*CR Subject Classification (1991):* G.1.7
*Keywords and Phrases:* numerical analysis, implicit ODEs, DAEs, Runge-Kutta methods, parallelism.

*Note:* The research reported in this paper was partly supported by the Technology Foundation (STW) in the Netherlands.

## 1. Introduction

We consider initial value problems (IVPs) for systems of implicit differential equations (IDEs)

(1.1) $\phi(\dot{\mathbf{y}}(t),\mathbf{y}(t)) = \mathbf{0}, \qquad \mathbf{y}, \phi \in \mathbf{R}^d.$

It will always be assumed that the initial conditions for $\mathbf{y}(t_0)$ and $\dot{\mathbf{y}}(t_0)$ are consistent and that the IVP has a unique solution. Let us define the Jacobian matrices $K := \phi_{\mathbf{u}}(\mathbf{u},\mathbf{v})$ and $J := -\phi_{\mathbf{v}}(\mathbf{u},\mathbf{v})$ (in the definition of J, the minus sign is inserted so that *explicit* differential equations with $\phi = \mathbf{y}' - \mathbf{f}(t,\mathbf{y})$ yield the familiar formula $J = \partial\mathbf{f}/\partial\mathbf{y}$). Then, the IVP is said to be *stable* if in the neighbourhood of the solution, the eigenspectrum $\sigma(K,J)$ of the pencil $J - \lambda K$ is in the nonpositive halfplane, that is, $\det(\lambda K - J)$ has only zeros in the nonpositive halfplane. More generally, any pair of matrices $\{K,J\}$ is said to be a *stable pair* if they satisfy this requirement. In the analysis of iteration methods for solving the numerical discretization of (1.1), the stability of matrix pairs will play a central role. Note that the

stability of the matrix pair {K,J} associated with (1.1) implies that its solution is bounded as $t \to \infty$. Furthermore, we remark that higher-index problems are stable whenever their index 1 representation are stable.

In particular, this paper will consider the *partitioned* case (or *semi-explicit* case, cf. [2]) where

$$(1.2) \qquad K = \begin{pmatrix} K_{11} & O \\ O & O \end{pmatrix}, \quad J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}.$$

Here, $K_{11}$, $J_{11}$ and $J_{22}$ are square matrices, with respective dimensions $d_1$, $d_1$ and $d_2 = d - d_1$. Stability conditions for the pair (1.2) in terms of its submatrices can be obtained by writing the eigenvalue equation $J\mathbf{w} = \lambda K\mathbf{w}$ in the componentwise form

$$(1.3) \qquad J_{11}\mathbf{u}_1 + J_{12}\mathbf{u}_2 = \lambda K_{11}\mathbf{u}_1, \quad J_{21}\mathbf{u}_1 + J_{22}\mathbf{u}_2 = \mathbf{0}.$$

We briefly discuss two special cases that will arise later in this paper. Let us first consider the case where $J_{22}$ is nonsingular. Then, $\mathbf{u}_2$ can be eliminated from the eigenvalue equations (1.3) to obtain the relation $S\mathbf{u}_1 = \lambda K_{11}\mathbf{u}_1$, where S is the Schur complement of J. Thus, the pair (1.2) is stable if

$$(1.4) \qquad J_{22} \text{ is nonsingular, } \{K_{11},S\} \text{ is stable, } \quad S := J_{11} - J_{12}J_{22}^{-1}J_{21}.$$

If $J_{22}$ is not necessarily nonsingular, but if it commutes with $J_{12}$ (see Example 2.3), then $\mathbf{u}_2$ can be eliminated by premultiplying the first equation of (1.3) with $J_{22}$ (here, we assume $d_1 = d_2$). This leads to the relation $\Delta\mathbf{u}_1 = \lambda J_{22}K_{11}\mathbf{u}_1$, where $\Delta$ is the 'determinant' of J. Thus, (1.2) is stable if

$$(1.5) \qquad d_1 = d_2, \; J_{22} \text{ commutes with } J_{12}, \; \{J_{22}K_{11}, \Delta\} \text{ is stable, } \quad \Delta := J_{22}J_{11} - J_{12}J_{21}.$$

It may happen that the IDE is not given in the semi-explicit form {(1.1),(1.2)}, but in the form

$$(1.6) \qquad Q\dot{\mathbf{y}}(t) = \mathbf{f}(\mathbf{y}(t)), \quad \mathbf{y}, \mathbf{f} \in R^d.$$

Suppose that Q is a constant, singular matrix of rank $d_1$. Then, there exist nonsingular matrices $S_1$ and $S_2$ such that (1.6) can be represented as the equation (see [4, p.406])

$$(1.7) \qquad S_1QS_2\dot{\mathbf{z}} - S_1\mathbf{f}(S_2\mathbf{z}) = \mathbf{0}, \quad \mathbf{z} = S_2^{-1}\mathbf{y}, \quad S_1QS_2 = \begin{pmatrix} I & O \\ O & O \end{pmatrix}, \qquad \mathbf{y}, \mathbf{z}, \mathbf{f} \in R^d$$

where the dimension of I equals the rank of Q. This problem is of the partitioned form {(1.1),(1.2)} with $K_{11} = I$. Introducing the partitioning $\mathbf{z} = (\mathbf{u}^T,\mathbf{v}^T)^T$ with $\mathbf{u}$ and $\mathbf{v}$ respectively of dimension $d_1$ and $d_2$, it can be written as an IVP for the semi-explicit differential-algebraic equation (DAE)

$$(1.8) \qquad \dot{\mathbf{u}}(t) = \mathbf{f}(\mathbf{u},\mathbf{v}), \qquad \mathbf{g}(\mathbf{u},\mathbf{v}) = \mathbf{0}, \quad \mathbf{u}, \mathbf{f} \in R^{d1}, \quad \mathbf{v}, \mathbf{g} \in R^{d2}.$$

In this paper, we shall analyse integration methods for solving (1.1) that can be efficiently used on parallel computer systems. We construct an IDE method for general IDEs of higher index, and two

methods that can be applied to partitioned problems of the type (1.2). These partitioned IDE methods both exploit the special form of the problem and converge faster than the general IDE method. The first partitioned IDE method is suitable for higher-index problems, the second partitioned IDE method only applies to index 1 problems, but is often considerably less expensive. The performance of the methods is illustrated by test problems from the literature.

## 2. The numerical scheme

Let us start with the case where (1.1) is an (explicit) ordinary differential equation (ODE)

$$\dot{\mathbf{y}}(t)) = \mathbf{f}(\mathbf{y}(t)); \quad \mathbf{y}, \mathbf{f} \in R^d.$$

A large class of implicit step-by-step methods for solving ODEIVPs can be represented in the form

$$(2.1) \qquad \mathbf{y}_{n+1} = (\mathbf{e}_s^T \otimes I)\mathbf{Y}, \quad \mathbf{Y} - h(A \otimes I)\mathbf{F}(\mathbf{Y}) = \mathbf{W}.$$

Here, A denotes a nonsingular s-by-s matrix, **W** is an sd-dimensional vector containing information computed in preceding integration steps, I is the d-by-d identity matrix, h is the stepsize $t_{n+1} - t_n$, and $\otimes$ denotes the Kronecker product. The s vector components $\mathbf{Y}_i$ of the sd-dimensional solution vector **Y** represent numerical approximations to the exact solution vectors $\mathbf{y}(\mathbf{e}t_n + \mathbf{c}h)$, **c** being the abscissa vector with $c_s = 1$ and **e** representing the s-dimensional vector with unit entries. Furthermore, $\mathbf{e}_s$ is the sth unit vector and $\mathbf{y}_n$ is the numerical approximation to $\mathbf{y}(t_n)$. In (2.1), $\mathbf{F}(\mathbf{V})$ contains the derivative values $\big(\mathbf{f}(\mathbf{V}_i)\big)$ for any vector $\mathbf{V} = \big(\mathbf{V}_i\big)$. In the following, I will denote the identity matrix and its dimension will always be clear from the context.

In order to derive the analogue of (2.1) for IDEs, we observe that if $\phi$ is invertible with repect to $\dot{\mathbf{y}}$, then this analogue should be equivalent to (2.1). This leads us to use (2.1) for expressing the derivative stage vector $\mathbf{F}(\mathbf{Y})$ in terms of **Y**, and to substitute this expression into the equation $\Phi\big(\mathbf{F}(\mathbf{Y}),\mathbf{Y}\big) = \mathbf{0}$, with $\Phi$ defined in the same way as **F**. This yields the method

$$(2.2) \qquad \mathbf{y}_{n+1} = (\mathbf{e}_s^T \otimes I)\mathbf{Y}, \quad \mathbf{R}(\mathbf{Y}) = \mathbf{0}, \quad \mathbf{R}(\mathbf{Y}) := \Phi\Big(\big(h^{-1}A^{-1} \otimes I\big)(\mathbf{Y} - \mathbf{W}),\mathbf{Y}\Big).$$

Thus, the method (2.2) is completely specified by the pair {A,**W**}.

An important class of methods leading to implicit relations of the form (2.1) are the stiffly accurate Runge-Kutta (RK) methods which arise for $\mathbf{W} := \mathbf{e} \otimes \mathbf{y}_n$. As an example, we consider the equation (1.6). If we apply (2.2), then

$$(2.3) \qquad \mathbf{y}_{n+1} = (\mathbf{e}_s^T \otimes I)\mathbf{Y}, \quad \big(h^{-1}A^{-1} \otimes Q\big)\big(\mathbf{Y} - \mathbf{W}\big) - \mathbf{F}\big(\mathbf{Y}\big) = \mathbf{0},$$

which is equivalent to the RK method discussed in [4, p.406].

**Remark 2.1.** As explained in [4, p.407], the RK solution **Y** defined by (2.3) is algebraically identical to $(I\otimes S_2)\mathbf{Z}$, where **Z** is the RK solution obtained by applying (2.2) with $\mathbf{W} := \mathbf{e}\otimes\mathbf{z}_n$ to (1.7), or equivalently, to (1.8). This equivalence holds for any method $\{A,\mathbf{W}\}$. ♦

The implicit equation in (2.2) will be solved iteratively by generating sequences of iterates $\{\mathbf{Y}^{(j)}\}$. Our starting point is the iteration method

$$(2.4)\qquad N\big(\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)}\big) = -\big(hA\otimes I\big)\mathbf{R}\big(\mathbf{Y}^{(j-1)}\big),\ \ j = 1, \dots , m,$$

where N is a nonsingular matrix. The iteration error associated with (2.4) satisfies the recursion

$$N\big(\mathbf{Y}^{(j)} - \mathbf{Y}\big) = N\big(\mathbf{Y}^{(j-1)} - \mathbf{Y}\big) - \big(hA\otimes I\big)\big(\mathbf{R}\big(\mathbf{Y}^{(j-1)}\big) - \mathbf{R}\big(\mathbf{Y}\big)\big),$$

so that ignoring second-order terms leads to

$$(2.5)\qquad \mathbf{Y}^{(j)} - \mathbf{Y} = M\big(\mathbf{Y}^{(j-1)} - \mathbf{Y}\big),\ \ M := N^{-1}\big(N - N_0\big),\ \ N_0 := I\otimes K - A\otimes hJ,$$

where the Jacobian matrices K and J are both evaluated at the step point $t_n$. The conventional choice for N is the modified Newton iteration matrix $N_0$ resulting in a zero amplification matrix M. The advantage of the choice $N = N_0$ is that, even in strongly nonlinear problems, a few iterations usually suffice to solve the implicit system in (2.2). However, a disadvantage is that solving the linear Newton systems can be quite expensive. For example, when direct methods are used, the LU-decomposition of the sd-by-sd matrix $N_0$ requires as many as $O(s^3d^3)$ arithmetic operations.
In this paper, we shall consider several choices of more 'convenient' iteration matrices N. A necessary and sufficient condition for linear convergence of the iteration method (2.4) requires the spectrum $\sigma(M)$ of M (to be referred to as the set of *amplification factors*) within the unit circle. Therefore, we shall try to combine a spectral radius $\rho(M) < 1$ with a reduction of the complexity of the linear Newton systems. In particular, we shall look for matrices N that reduce the computational complexity on parallel computer systems (for example, matrices N with a block-triangular structure).

**Remark 2.2.** In an actual implementation of (2.4), it may be recommendable to remove the $h^{-1}$ factor in the residual in (2.2) by defining the 'derivative' iterate $\dot{\mathbf{Y}}^{(j)} := \big(h^{-1}A^{-1}\otimes I\big)\big(\mathbf{Y}^{(j)} - \mathbf{W}\big)$. Then, the iteration scheme becomes

$$(2.4')\qquad N\big(A\otimes I\big)\big(\dot{\mathbf{Y}}^{(j)} - \dot{\mathbf{Y}}^{(j-1)}\big) = -\big(A\otimes I\big)\Phi\big(\dot{\mathbf{Y}}^{(j-1)},\mathbf{W} + \big(hA\otimes I\big)\dot{\mathbf{Y}}^{(j-1)}\big),\ \ \mathbf{Y}^{(j)} = \mathbf{W} + h\big(A\otimes I\big)\dot{\mathbf{Y}}^{(j)}.$$

The sequences $\{\mathbf{Y}^{(j)}\}$ generated by the schemes (2.4) and (2.4') are algebraically identical, but (2.4') can be used as $h \to 0$. ♦

### 2.1. General IDE method

In the case of IVPs for the IDE (1.1) with *general* Jacobians K and J, the Newton process defined by (2.4) with $N = N_0 = I\otimes K - A\otimes hJ$ can be 'simplified' by replacing A by a lower triangular approximation B, so that

$$(2.6)\qquad N = I\otimes K - B\otimes hJ = \begin{pmatrix} K - hB_{11}J & O & O & \dots \\ -hB_{21}J & K - hB_{22}J & O & \dots \\ -hB_{31}J & -hB_{32}J & K - hB_{33}J & \dots \\ . & . & . & . \end{pmatrix}.$$

In addition, one may also replace J by an approximation $J^*$ that is tuned to the problem to be solved. In the case of RK methods for ODEIVPs, a first analysis of such an approach can be found in [7], and for IDEIVPs, results will be given in a forthcoming paper.

Substitution of (2.6) into (2.4) yields

$$(2.7)\qquad \big(I\otimes K - B\otimes hJ\big)\big(\mathbf{Y}^{(j)} - \mathbf{Y}^{(j-1)}\big) = -\big(hA\otimes I\big)\mathbf{R}(\mathbf{Y}^{(j-1)}).$$

Formally, the iteration method (2.7) can be applied to problems of any index if the matrix $N = I\otimes K - B\otimes hJ$ is nonsingular. Thus, the s blocks $K - hB_{ii}J$ should be nonsingular. The method (2.7) will be referred to as the *general IDE method*. Furthermore, we shall say that the method is in *diagonal mode* if B is a diagonal matrix D and in *triangular mode* if B is a lower triangular matrix T.

Each iteration with (2.7) requires the solution of a linear system with the block-triangular matrix N. Hence, the system splits into s subsystems of dimension d, reducing the computational costs considerably. In this paper, we assume that the subsystems are solved by a direct solution method. Then, apart from solving these s linear subsystems, each update of the matrix $I\otimes K - B\otimes hJ$ implies the LU-decomposition of the s blocks $K - hB_{ii}J$ associated with the s subsystems. However, this only requires $O(sd^3)$ flops which is a factor $s^2$ less than the number of flops needed when the modified Newton matrix $N_0$ is used. Moreover, these LU-decomposutions can be done in parallel, so that the effective costs on a parallel system are a factor $s^3$ smaller. Likewise, the solution of the s subsystems can also be done in parallel, both in the case of the diagonal-implicit and the triangular-implicit approach. The triangular approach requires additional costs for performing similarity transformations, but as will be demonstrated in our numerical experiments in Section 3, it usually converges much faster. We remark that the linear system in (2.7) has similarities with the linear systems occurring in the parallel diagonal-implicitly iterated RK methods and the parallel triangular-implicitly iterated RK methods analysed in [6] and [8] for solving IVPs for ODEs. In particular, reference [8] presents a detailed discussion of the advantages and disadvantages of the diagonal-implicit and the triangular-implicit approach which to a large extent also applies to the IDE case.

Next, we consider the convergence of the iteration method (2.7). If $I\otimes K - B\otimes hJ$ is nonsingular, then it follows from (2.5) that the error amplification matrix corresponding to (2.7) is given by

(2.8) $M = \big(I\otimes K - B\otimes hJ\big)^{-1}\big((A - B)\otimes hJ\big)$.

Let us denote the eigenvectors and eigenvalues of M by $\mathbf{a}\otimes\mathbf{w}$ and $\mu$. Then, we derive the relation

$$h(A - B + \mu B)\mathbf{a}\otimes J\mathbf{w} = \mu\mathbf{a}\otimes K\mathbf{w}.$$

This shows that $J\mathbf{w}$ and $K\mathbf{w}$ are related by the generalized eigenvalue equation $J\mathbf{w} = \lambda K\mathbf{w}$, where $\lambda$ is a generalized eigenvalue. On substitution of $J\mathbf{w} = \lambda K\mathbf{w}$ and by defining $z := \lambda h$, we obtain

$$z(A - B)\mathbf{a}\otimes K\mathbf{w} = \mu\big(I - zB\big)(\mathbf{a}\otimes K\mathbf{w}).$$

Thus, if $K\mathbf{w} \neq \mathbf{0}$, then $\mu = \mu(z)$ is an eigenvalue of the matrix

(2.9) $Z(z) := z\big(I - zB\big)^{-1}\big(A - B\big)$,

provided that $I - zB$ is nonsingular. If $K\mathbf{w} = \mathbf{0}$ with $\mathbf{w} \neq \mathbf{0}$, then $\lambda = \infty$, so that $\mu$ is an eigenvalue of the matrix $I - B^{-1}A$. Matrices of the type (2.9) have extensively been studied in [6] and [8]. For a large number of RK matrices A, diagonal and lower triangular matrices B have been found such that the eigenvalues $\mu(z)$ of $Z(z)$ are within the unit circle whenever $Re(z) \le 0$. Lower triangular matrices B that possess this property will be said to lie in the set $\mathbf{B}(A)$ associated with the matrix A (in the following, it will be assumed that $B \in \mathbf{B}(A)$, unless stated otherwise). The condition $Re(z) \le 0$ is satisfied if, and only if, the eigenspectrum $\sigma(K,J)$ of the pencil $J - \lambda K$ is in the nonpositive halfplane, that is, if $\{K,J\}$ is a stable pair. Hence, stability of $\{K,J\}$ implies $|\mu(z)| < 1$. Furthermore, stability implies the nonsingularity of the matrices $K - hB_{ii}J$, because if the matrix $B \in \mathbf{B}(A)$, then its diagonal entries $B_{ii}$ are necessarily positive, otherwise $Z(\infty)$ would be singular. Thus, we have proved the convergence theorem:

**Theorem 2.1.** Let $B \in \mathbf{B}(A)$. Then, the general IDE method (2.7) converges for all $h > 0$ if, and only if, $\{K,J\}$ is a stable pair. ♦

**Example 2.1.** Consider equation (1.6) with $Q = Q(\mathbf{y})$. Then, $K = Q(\mathbf{y})$ and $J = \big(\mathbf{f}(\mathbf{y}) - Q(\mathbf{y})\dot{\mathbf{y}}\big)_{\mathbf{y}}$. Thus, Theorem 2.1 implies convergence for $h > 0$ if $\{K,J\} = \big\{Q(\mathbf{y}),\big(\mathbf{f}(\mathbf{y}) - Q(\mathbf{y})\dot{\mathbf{y}}\big)_{\mathbf{y}}\big\}$ is a stable pair. ♦

**Example 2.2.** Consider equation (1.8) with $\mathbf{g} = \mathbf{g}(\mathbf{u})$, that is, the equation

(1.8') $\dot{\mathbf{u}}(t) = \mathbf{f}(\mathbf{u},\mathbf{v}), \quad \mathbf{g}(\mathbf{u}) = \mathbf{0}, \quad \mathbf{v}, \mathbf{g} \in R^{d2}$.

If $\mathbf{f_v g_u}$ is assumed nonsingular, then this equation is of index 2. Evidently, it is of the form (1.2) with

$$K = \begin{pmatrix} I & O \\ O & O \end{pmatrix}, \quad J = \begin{pmatrix} \mathbf{f_u} & \mathbf{f_v} \\ \mathbf{g_u} & O \end{pmatrix}.$$

Recalling that higher-index problems are stable whenever their index 1 representation is stable, we write (1.8') in the index 1 form $\dot{\mathbf{u}}(t) = \mathbf{f}(\mathbf{u},\mathbf{v})$, $\mathbf{g_u}(\mathbf{u})\mathbf{f}(\mathbf{u},\mathbf{v}) = \mathbf{0}$. This equation is again of the form (1.2) with

$$K = K^* = \begin{pmatrix} I & O \\ O & O \end{pmatrix}, \quad J = J^* = \begin{pmatrix} \mathbf{f_u} & \mathbf{f_v} \\ \mathbf{g_u f_u} & \mathbf{g_u f_v} \end{pmatrix}.$$

Hence, we have convergence for all $h > 0$ if $\{K^*,J^*\}$ is a stable pair. ♦

### 2.2. IDE methods for partitioned problems

In this section, we consider problems of the partitioned form {(1.1),(1.2)}. For such problems, it is convenient to write $\mathbf{y} = (\mathbf{u}^T,\mathbf{v}^T)^T$, where $\mathbf{u}$ and $\mathbf{v}$ are respectively of dimension $d_1$ and $d_2$, and to replace the stage vector $\mathbf{Y}$ by the permuted stage vector $\tilde{\mathbf{Y}} = P\mathbf{Y} := (\mathbf{U}^T,\mathbf{V}^T)^T$, where $\mathbf{U}$ and $\mathbf{V}$ are stage vectors associated with $\mathbf{u}$ and $\mathbf{v}$ in the same way as $\mathbf{Y}$ is associated with $\mathbf{y}$. Let us introduce the permuted iterates $\tilde{\mathbf{Y}}^{(j)} := P\mathbf{Y}^{(j)}$, then the permuted versions of (2.4) and (2.5) respectively become

$$(2.10) \qquad \tilde{N}\left(\tilde{\mathbf{Y}}^{(j)} - \tilde{\mathbf{Y}}^{(j-1)}\right) = -P(hA\otimes I)\mathbf{R}(P^{-1}\tilde{\mathbf{Y}}^{(j-1)}), \quad \tilde{N} := PNP^{-1},$$

$$(2.11) \qquad \tilde{\mathbf{Y}}^{(j)} - \tilde{\mathbf{Y}} = \tilde{M}\left(\tilde{\mathbf{Y}}^{(j-1)} - \tilde{\mathbf{Y}}\right), \quad \tilde{M} := PMP^{-1} = \tilde{N}^{-1}\left(\tilde{N} - \tilde{N}_0\right), \quad \tilde{N}_0 := P(I\otimes K - A\otimes hJ)P^{-1}.$$

Before selecting suitable matrices $\tilde{N}$ for the problem {(1.1),(1.2)} we consider the general IDE method when applied to {(1.1),(1.2)}. By replacing in (2.7) $\mathbf{Y}^{(j)}$ by $P^{-1}\tilde{\mathbf{Y}}^{(j)}$ and by observing that for any matrix C and any matrix J with a partitioning as in (1.2), the permutation matrix P satisfies

$$P(C\otimes J)P^{-1} = \begin{pmatrix} C\otimes J_{11} & C\otimes J_{12} \\ C\otimes J_{21} & C\otimes J_{22} \end{pmatrix},$$

it can be verified that the general IDE method (2.7) takes the form

$$(2.7') \qquad \begin{pmatrix} I\otimes K_{11} - B\otimes hJ_{11} & -B\otimes hJ_{12} \\ -B\otimes hJ_{21} & -B\otimes hJ_{22} \end{pmatrix}\left(\tilde{\mathbf{Y}}^{(j)} - \tilde{\mathbf{Y}}^{(j-1)}\right) = \begin{pmatrix} -hA\otimes I & O \\ O & -hA\otimes I \end{pmatrix} P\mathbf{R}(P^{-1}\tilde{\mathbf{Y}}^{(j-1)}).$$

Since

$$\tilde{N}_0 = \begin{pmatrix} I\otimes K_{11} - A\otimes hJ_{11} & -A\otimes hJ_{12} \\ -A\otimes hJ_{21} & -A\otimes hJ_{22} \end{pmatrix},$$

the matrix $\tilde{N}$ associated with (2.7') can be written as

$$(2.6') \qquad \tilde{N} = \begin{pmatrix} I\otimes K_{11} - B\otimes hJ_{11} & -B\otimes hJ_{12} \\ -B\otimes hJ_{21} & -B\otimes hJ_{22} \end{pmatrix} = \tilde{N}_0 + h\begin{pmatrix} (A-B)\otimes J_{11} & (A-B)\otimes J_{12} \\ (A-B)\otimes J_{21} & (A-B)\otimes J_{22} \end{pmatrix}.$$

We now want to 'improve' this matrix by exploiting the special form of the equation {(1.1),(1.2)}. From (2.11) it follows that convergence is expected to be faster as the magnitude of $\tilde{N} - \tilde{N}_0$ is smaller, that is, if the generated method is 'closer' to the true (modified) Newton process. Of the various possibilities, we shall consider the cases

$$\text{(2.12a)}\quad \tilde{N} = \begin{pmatrix} I\otimes K_{11} - B\otimes hJ_{11} & -B\otimes hJ_{12} \\ -A\otimes hJ_{21} & -A\otimes hJ_{22} \end{pmatrix} = \tilde{N}_0 + h \begin{pmatrix} (A-B)\otimes J_{11} & (A-B)\otimes J_{12} \\ O & O \end{pmatrix}.$$

$$\text{(2.12b)}\quad \tilde{N} = \begin{pmatrix} I\otimes K_{11} - A\otimes hJ_{11} & -A\otimes hJ_{12} \\ O & -A\otimes hJ_{22} \end{pmatrix} = \tilde{N}_0 + h \begin{pmatrix} O & O \\ A\otimes J_{21} & O \end{pmatrix}.$$

Notice that (2.12b) requires the matrix $J_{22}$ to be nonsingular, that is, the equation {(1.1),(1.2)} should be of index 1. This condition is satisfied by many IVPs (e.g. all IVPs for the equation (1.8) with $\mathbf{g_v}$ nonsingular), so that it is relevant to analyse the case (2.12b). The iteration matrices (2.12a) and (2.12b) respectively generate the methods

$$\text{(2.13a)}\quad \begin{pmatrix} I\otimes K_{11} - B\otimes hJ_{11} & -B\otimes hJ_{12} \\ I\otimes J_{21} & I\otimes J_{22} \end{pmatrix} \left(\tilde{\mathbf{Y}}^{(j)} - \tilde{\mathbf{Y}}^{(j-1)}\right) = \begin{pmatrix} -hA\otimes I & O \\ O & I \end{pmatrix} P\mathbf{R}\left(P^{-1}\tilde{\mathbf{Y}}^{(j-1)}\right),$$

$$\text{(2.13b)}\quad \begin{pmatrix} I\otimes K_{11} - A\otimes hJ_{11} & -A\otimes hJ_{12} \\ O & I\otimes J_{22} \end{pmatrix} \left(\tilde{\mathbf{Y}}^{(j)} - \tilde{\mathbf{Y}}^{(j-1)}\right) = \begin{pmatrix} -hA\otimes I & O \\ O & I \end{pmatrix} P\mathbf{R}\left(P^{-1}\tilde{\mathbf{Y}}^{(j-1)}\right),$$

and will be referred to as the *partitioned IDE method I* and the *partitioned IDE method II*. Note that for $d_2 = 0$ (that is, if there is no partitioning), the methods (2.13a) and (2.13b) respectively reduce to the general IDE method (2.7') and to the modified Newton method with $\tilde{N} = \tilde{N}_0$. From (2.12) it follows that we may expect (2.13a) to converge faster than (2.7') if $(A-B)\otimes J_{11}$ and $(A-B)\otimes J_{12}$ are of small magnitude, whereas (2.13b) will converge faster if $J_{21}$ is of small magnitude.
In the Subsections 2.2.1 and 2.2.2, the computational efficiency and convergence conditions for the partitioned IDE methods (2.13a) and (2.13b)will be discussed.

### 2.2.1. Partitioned IDE method I

Each iteration with (2.13a) requires the solution of s systems of dimension $d := d_1 + d_2$ whose matrices of coefficients are of the form

$$\text{(2.14)}\quad N_i := \begin{pmatrix} K_{11} - hB_{ii}J_{11} & hB_{ii}J_{12} \\ J_{21} & J_{22} \end{pmatrix},\quad i = 1, \ldots, s.$$

Thus, in order to apply the partitioned IDE method I, the matrices $N_i$ should be nonsingular. Let $D_i$ be the diagonal matrix with $d_1$ diagonal entries 1 and $d_2$ diagonal entries $-hB_{ii}$. Then we may write $N_i = D_i^{-1}(K - hB_{ii}J)$. For stable IVPs, the matrices $K - hB_{ii}J$ are nonsingular, and therefore the matrices $N_i$. Furthermore, it follows from (2.7') and (2.13a) that for partitioned problems, the computational complexities of the general IDE and the partitioned IDE method I are comparable.

In the convergence analysis, we proceed as in the proof of Theorem 2.1. The matrix $\tilde{M}$ is given by

$$\tilde{M} = h \begin{pmatrix} I\otimes K_{11} - B\otimes hJ_{11} & -B\otimes hJ_{12} \\ -A\otimes hJ_{21} & -A\otimes hJ_{22} \end{pmatrix}^{-1} \begin{pmatrix} (A-B)\otimes J_{11} & (A-B)\otimes J_{12} \\ O & O \end{pmatrix}.$$

Let the eigenvalues and eigenvectors of $\tilde{M}$ be denoted by $\mu$ and $\begin{pmatrix} \mathbf{b}\otimes\mathbf{u} \\ \mathbf{b}\otimes\mathbf{v} \end{pmatrix}$. Then,

$$h(A-B)\mathbf{b}\otimes(J_{11}\mathbf{u} + J_{12}\mathbf{v}) = \mu\big(\mathbf{b}\otimes K_{11}\mathbf{u} - B\mathbf{b}\otimes(hJ_{11}\mathbf{u} + hJ_{12}\mathbf{v})\big),$$
$$\mu\big(A\mathbf{b}\otimes(J_{21}\mathbf{u} + J_{22}\mathbf{v})\big) = \mathbf{0}.$$

Again, we use the generalized eigenvalue equation $J\mathbf{w} = \lambda K\mathbf{w}$. Writing this equation in the componentwise form (1.3), we obtain upon substitution

$$z(A-B)\mathbf{b}\otimes K_{11}\mathbf{u} = \mu\big(I - zB\big)(\mathbf{b}\otimes K_{11}\mathbf{u}), \quad z := \lambda h.$$

If $K_{11}\mathbf{u} \neq \mathbf{0}$, then $\mu = \mu(z)$ is again an eigenvalue of the matrix $Z(z)$ defined in (2.9). If $K_{11}\mathbf{u} = \mathbf{0}$ with $\mathbf{u} \neq \mathbf{0}$, then $\mu$ is an eigenvalue of the matrix $I - B^{-1}A$. The analogue of Theorem 2.1 becomes:

**Theorem 2.2.** Let $B \in \mathbf{B}(A)$. Then, the partitioned IDE method I defined by (2.13a) converges for all $h > 0$ if, and only if, (1.2) is a stable pair. ♦

Thus, a comparison with Theorem 2.1 reveals that for partitioned problems where K is of the form (1.2), the two theorems impose the same convergence conditions.

**Example 2.3.** The IDE (1.6) considered in Example 2.1 can be written as the DAE (cf. [4, p.486])

(1.7) $\quad \dot{\mathbf{u}}(t) = \mathbf{v}, \quad Q(\mathbf{u})\mathbf{v} - \mathbf{f}(\mathbf{u}) = \mathbf{0}, \quad \mathbf{u}, \mathbf{v}, \mathbf{f} \in R^{d_1}.$

K and J are of the form (1.2) with $K_{11} = I$ and $J_{11} = O$, $J_{12} = I$, $J_{21} = \big(Q(\mathbf{u})\mathbf{v} - \mathbf{f}(\mathbf{u})\big)_{\mathbf{u}}$, $J_{22} = Q(\mathbf{u})$. Since $J_{22}$ commutes with $J_{12}$, it follows from Theorem 2.2 and (1.5) that the partitioned method I converges for $h > 0$ whenever $\{Q(\mathbf{u}), \big(\mathbf{f}(\mathbf{u}) - Q(\mathbf{u})\mathbf{v}\big)_{\mathbf{u}}\}$ is a stable pair (compare Example 2.1). ♦

### 2.2.2. Partitioned IDE method II

From (2.13b) it follows that in each iteration we can first solve in parallel the s (uncoupled) $d_2$-dimensional systems for the last $sd_2$ components $\mathbf{V}^{(j)}$ of $\tilde{\mathbf{Y}}^{(j)}$ (requiring the nonsingularity of $J_{22}$), and next the s systems of dimension $d_1$ for the first $sd_1$ components $\mathbf{U}^{(j)}$ of $\tilde{\mathbf{Y}}^{(j)}$, that is, the system

(2.15) $\quad \big(I\otimes K_{11} - A\otimes hJ_{11}\big)\big(\mathbf{U}^{(j)} - \mathbf{U}^{(j-1)}\big) = \mathbf{Q}^{(j)}, \quad \mathbf{Q}^{(j)} := h\big(-A\otimes I \quad A\otimes J_{12}J_{22}^{-1}\big)P\mathbf{R}\big(P^{-1}\tilde{\mathbf{Y}}^{(j-1)}\big).$

The iterate $\mathbf{U}^{(j)}$ as defined by (2.15) will be computed iteratively by an *inner* iteration method. The recursion (2.13b) itself will be called *outer* iteration.

For the inner iteration, we shall use a method which is very much like the general IDE method used for solving $\mathbf{R}(\mathbf{Y}) = \mathbf{0}$ in (2.2). Denoting the inner iterates by $\mathbf{Z}^{(k)}$, we have

$$(2.16)\quad \left(I\otimes K_{11} - B\otimes hJ_{11}\right)\left(\mathbf{Z}^{(k)} - \mathbf{Z}^{(k-1)}\right) = \mathbf{Q}^{(j)} - \left(I\otimes K_{11} - A\otimes hJ_{11}\right)\left(\mathbf{Z}^{(k-1)} - \mathbf{U}^{(j-1)}\right),\ k = 1,..., r,$$

requiring the nonsingularity of the matrices $K_{11} - hB_{ii}J_{11}$. In this inner iteration process we may use as initial approximation $\mathbf{Z}^{(0)} = \mathbf{U}^{(j-1)}$ and, after r iterations, we set $\mathbf{U}^{(j)} = \mathbf{Z}^{(r)}$.
If B is diagonal, then the s linear systems to be solved in each iteration of (2.16) can be treated in parallel. For triangular B, the degree of parallelism depends on the linear solver used. Let us consider the case of direct solvers. Then, the LU-decompositions of $J_{22}$ and the matrices $K_{11} - hB_{ii}J_{11}$, $i = 1, \ldots, s$, can be done in parallel. It is here where the sequential (or effective) costs of the partitioned method II may be substantially less than those for the general IDE method and the partitioned method I. The LU-costs of the latter two methods are $\approx 2(d_1 + d_2)^3 / 3$ flops, whereas the partitioned method II requires only $\approx 2(\max\{d_1,d_2\})^3 / 3$ flops, yielding a speed-up factor for the (often dominating) LU costs of $\approx (d_1 + d_2)^3(\max\{d_1,d_2\})^{-3} = (1 + \min\{d_1d_2^{-1},d_2d_1^{-1}\})^3$. Hence, if $d_1$ equals $d_2$, then a maximum speed-up factor of 8 is obtained. If $d_2 > d_1$ and if the matrix $K_{11}$ is allowed to be singular, then we may repartition the partitioned equation {(1.1),(1.2)} by adding algebraic equations to the 'differential equation' part so that $d_1$ is increased. This would reduce the computational complexity of the method when implemented on a parallel system.
In the next two subsections, the convergence of the outer and inner iteration will be analysed.

**2.2.2.1. Convergence of the outer iteration.** A convergence result for the outer iteration (2.13b) is obtained by elimination of $\mathbf{V}^{(j)}$ from (2.15) and by writing down the error recursion for $\mathbf{U}^{(j)} - \mathbf{U}$. The corresponding amplification matrix becomes

$$\tilde{M}_{11} := -\left(I\otimes K_{11} - A\otimes hJ_{11}\right)^{-1}\left(A\otimes hJ_{12}J_{22}^{-1}J_{21}\right).$$

Denoting the eigenvectors and eigenvalues of $\tilde{M}_{11}$ by $\mathbf{a}\otimes\mathbf{w}$ and $\tilde{\mu}$, we obtain

$$\left(A\mathbf{a}\otimes h(S - J_{11} + \tilde{\mu}J_{11})\right)\mathbf{w} = \left(\mathbf{a}\otimes\tilde{\mu}K_{11}\right)\mathbf{w}$$

Let $\mathbf{a}$ be an eigenvector of A with eigenvalue $\alpha$. Then,

$$\left(\tilde{\mu}(K_{11} - \alpha hJ_{11}) + \alpha hJ_{12}J_{22}^{-1}J_{21}\right)\mathbf{w} = \mathbf{0}.$$

Suppose that $\{K_{11},J_{11}\}$ is stable, that is, the 'differential equation' part of the IVP is required to be stable, which seems to be a quite natural requirement. Then, it follows that the matrix $K_{11} - \alpha hJ_{11}$, and therefore $I\otimes K_{11} - A\otimes hJ_{11}$, is nonsingular for all $\alpha h$ in the positive halfplane. Hence, for $\mathrm{Re}(\alpha h) > 0$, the eigenvalues $\tilde{\mu}$ are also eigenvalues of the matrix

$$(2.17)\quad \tilde{Z}(\alpha h) := -\alpha h\left(K_{11} - \alpha hJ_{11}\right)^{-1}\left(J_{12}J_{22}^{-1}J_{21}\right).$$

This leads us to the convergence theorem:

**Theorem 2.3.** Assume that $J_{22}$ is nonsingular and that A has its spectrum $\sigma(A)$ in the positive halfplane. Then, the outer iteration (2.13b) converges if, and only if, (i) $\{K_{11},J_{11}\}$ is stable and (ii) $\tilde{Z}(\alpha h)$ has its eigenvalues within the unit circle for all $\alpha \in \sigma(A)$. ♦

Using properties of the logarithmic matrix norm $\mu[.]$, the following corollary from Theorem 2.3 can be proved.

**Corollary 2.1.** Let the assumptions of Theorem 2.3 be satisfied and let $\{K_{11},J_{11}\}$ be stable. Then, each of the following conditions are sufficient for convergence of the outer iteration (2.13b):

(2.18a) $h > 0,\ \mu[K_{11}^{-1}J_{11}] < -\|K_{11}^{-1}J_{12}J_{22}^{-1}J_{21}\|,\quad K_{11}$ nonsingular

(2.18b) $$0 < h < \frac{\mathrm{Re}(\alpha)}{|\alpha|^2}\,\frac{1}{\|K_{11}^{-1}J_{12}J_{22}^{-1}J_{21}\| + \mu[K_{11}^{-1}J_{11}]}\ ,\quad \alpha \in \sigma(A),\quad K_{11}\text{ nonsingular.}$$

(2.18c) $$0 < h < -\frac{\mu[-K_{11} + \alpha h J_{11}]}{|\alpha|\,\|J_{12}J_{22}^{-1}J_{21}\|}\ ,\quad \alpha \in \sigma(A).$$

**Proof.** To prove assertions (2.18a) and (2.18b) we use the inequality

(2.19) $$\rho(\tilde{Z}(\alpha h)) \le h\,|\alpha|\,\|H^{-1}\|\,\|K_{11}^{-1}J_{12}J_{22}^{-1}J_{21}\|,$$

where $H := I - \alpha h K_{11}^{-1}J_{11}$. Obviously, $\rho(\tilde{Z}(\alpha h)) < 1$ if $h\,\|K_{11}^{-1}J_{12}J_{22}^{-1}J_{21}\| < \|(\alpha^{-1}H)^{-1}\|^{-1}$. By virtue of a property of the logarithmic norm, we have that for nonsingular, complex matrices C both $-\mu[C]$ and $-\mu[-C]$ are less than $\|C^{-1}\|^{-1}$ (the proof given in [3] for real C, is easily generalized for complex C). Hence, setting $C = \alpha^{-1}H$, we conclude that satisfying $h\,\|K_{11}^{-1}J_{12}J_{22}^{-1}J_{21}\| < -\mu[-\alpha^{-1}H] = -\mu[-\alpha^{-1} + hK_{11}^{-1}J_{11}]$ certainly implies that $\rho(\tilde{Z}(\alpha h))$ is less than 1. Since (cf. [10])

$$\mu[-\alpha^{-1} + hK_{11}^{-1}J_{11}] = -\frac{\mathrm{Re}(\alpha)}{|\alpha|^2} + h\mu[K_{11}^{-1}J_{11}],$$

we are led to the conditions (2.18a) or (2.18b). For proving (2.18c), we use the inequality

(2.19') $$\rho(\tilde{Z}(\alpha h)) \le h\,|\alpha|\,\|L^{-1}\|\,\|J_{12}J_{22}^{-1}J_{21}\|,$$

where $L := K_{11} - \alpha h J_{11}$. We have $\rho(\tilde{Z}(\alpha h)) < 1$ if $h\,|\alpha|\,\|J_{12}J_{22}^{-1}J_{21}\| < \|L^{-1}\|^{-1}$. Since $\|L^{-1}\|^{-1}$ is bounded below by $\max\{-\mu[-L],-\mu[L]\}$, we conclude that $\rho(\tilde{Z}(\alpha h))$ is certainly less than 1 if the inequality $h\,|\alpha|\,\|J_{12}J_{22}^{-1}J_{21}\| < -\mu[-L]$ is satisfied. This leads to the condition (2.18c). ♦

Condition (2.18a) implies unconditional convergence and is satisfied if $\mu[K_{11}^{-1}J_{11}]$ is sufficiently negative (i.e. {(1.1),(1.2)} is sufficiently dissipative). If the problem is not sufficiently dissipative, then we have to impose a step restiction and we may use the conditions (2.18b) or (2.18c). A practical advantage of (2.18b) over (2.18c) is the separation of quantities defined by the method (the eigenvalues $\alpha$ of A) and quantities defined by the problem (the matrices $K_{11}$, $J_{11}$, etc.). On the other hand, (2.18b) cannot be applied if $K_{11}$ is singular and it may be unnecessary restrictive in cases where $\mathrm{Re}(\alpha)$ is small (note that (2.18b) and (2.18c) are equivalent if $\alpha > 0$ and $K_{11} = I$).

**Example 2.4.** Let $K_{11} = I$ and let $J_{11}$ be symmetric. Then, using the Euclidean norm $\|.\|_2$, we have for any complex matrix C (see e.g. [10]), $\mu_2[C] = \frac{1}{2}\lambda_{max}(C+C^*)$. Hence, the conditions (2.18) become

(2.18a') $\quad \lambda_{max}(J_{11}) < - \|J_{12}J_{22}^{-1}J_{21}\|,$

(2.18b') $\quad 0 < h < \dfrac{\mathrm{Re}(\alpha)}{|\alpha|^2} \dfrac{1}{\|J_{12}J_{22}^{-1}J_{21}\| + \lambda_{max}(J_{11})}, \qquad \alpha \in \sigma(A),$

(2.18c') $\quad 0 < h < \dfrac{1}{|\alpha|\,\|J_{12}J_{22}^{-1}J_{21}\| + \mathrm{Re}(\alpha)\,\lambda_{max}(J_{11})}, \quad \alpha \in \sigma(A).$ ♦

**2.2.2.2. Convergence of the inner iteration.** Finally, we have to prove that the inner iteration process (2.16) converges to the outer iterate $\mathbf{U}^{(j)}$ defined by (2.15). From (2.15) and (2.16) it follows that

$$\left(I\otimes K_{11} - B\otimes hJ_{11}\right)\left(\mathbf{Z}^{(k)} - \mathbf{Z}^{(k-1)}\right) = - \left(I\otimes K_{11} - A\otimes hJ_{11}\right)\left(\mathbf{Z}^{(k-1)} - \mathbf{U}^{(j)}\right).$$

It is easily verified that this relation yields the error equation

$$\mathbf{Z}^{(k)} - \mathbf{U}^{(j)} = \left(I\otimes K_{11} - B\otimes hJ_{11}\right)^{-1}\left((A - B)\otimes hJ_{11}\right)\left(\mathbf{Z}^{(k-1)} - \mathbf{U}^{(j)}\right).$$

The corresponding amplification matrix is of the same structure as the matrix M defined in (2.8). Hence, proceeding as in the proof of Theorem 2.1, we are led to the convergence result:

**Theorem 2.4.** Let $B \in \mathbf{B}(A)$. Then, the inner iteration method (2.16) converges for all $h > 0$ to the solution $\mathbf{U}^{(j)}$ of (2.15) if, and only if, $\{K_{11},J_{11}\}$ is stable. ♦

Recall that $B \in \mathbf{B}(A)$ implies that the $B_{ii}$ are positive, so that the convergence condition implies that the matrices $K_{11} - hB_{ii}J_{11}$ are nonsingular as required for applying the inner iteration (2.16). Furthermore, we recall that convergence of the outer iteration method (2.15) requires the eigenvalues of $\tilde{Z}(\alpha h)$ within the unit circle for all $\alpha$ (see Theorem 2.3). In cases where this condition imposes a stepsize restriction ({(1.1),(1.2)} not sufficiently dissipative), one may wonder whether we should

require the inner iteration to be unconditionally convergent with respect to h, that is, why should we choose $B \in \mathbf{B}(A)$. For example, we may simply take B = O, to obtain the result:

**Theorem 2.5.** Let B = O. The inner iteration method (2.16) converges to the solution $\mathbf{U}^{(j)}$ of (2.15) if

(2.20) $$h < \frac{1}{|\alpha|\,\rho(K_{11}^{-1}J_{11})}\,, \quad \alpha \in \sigma(A), \quad K_{11} \text{ nonsingular.}$$ ♦

**Example 2.5.** Consider the case of Example 2.4 where $K_{11} = I$, $J_{11}$ is symmetric. A comparison of (2.18b') and (2.18c') with (2.20) shows that (2.20) is less restrictive if, respectively, $\|J_{12}J_{22}^{-1}J_{21}\| > |\alpha|^{-1}\mathrm{Re}(\alpha)\,\rho(J_{11}) - \lambda_{\max}(J_{11})$ and $\|J_{12}J_{22}^{-1}J_{21}\| > \rho(J_{11}) - |\alpha|^{-1}\mathrm{Re}(\alpha)\,\lambda_{\max}(J_{11})$ for all $\alpha \in \sigma(A)$. This situation can easily occur if in J the entries of large magnitude are concentrated in the blocks $J_{12}$ and $J_{21}$. ♦

Finally, we remark that in actual computation, it is often sufficient to perform only a few inner iterations. In the extreme case where just one inner iteration is used (i.e., r = 1 in (2.16)), the partitioned IDE method {(2.13b),(2.16)} reduces to

(2.21) $$\begin{pmatrix} I\otimes K_{11} - B\otimes hJ_{11} & -A\otimes hJ_{12} \\ O & I\otimes J_{22} \end{pmatrix} \left(\tilde{\mathbf{Y}}^{(j)} - \tilde{\mathbf{Y}}^{(j-1)}\right) = \begin{pmatrix} -hA\otimes I & O \\ O & I \end{pmatrix} \mathbf{PR}\left(P^{-1}\tilde{\mathbf{Y}}^{(j-1)}\right).$$

In fact, in our experiments, we did apply the partitioned IDE method II with a single inner iteration.

### 3. Numerical experiments

The aim of this section is to compare (i) algorithmic properties of the general IDE method and the partitioned IDE methods (if applicable), and (ii) the convergence behaviour of the diagonal and the triangular mode of the IDE methods. This comparison is carried out for IDEIVPs, taken from the literature, with index varying from 0 until 3.

In all cases, the corrector is defined by the four-stage Radau IIA corrector. The predictor formula is given by $\mathbf{Y}^{(0)} := E\mathbf{Y}_n^{(m)}$, where $\mathbf{Y}_n^{(m)}$ denotes the final approximation to the stage vector obtained in the preceding step, and E is the extrapolation matrix of maximal order. As already remarked, the partitioned iteration method II is applied in the one-inner-iteration mode (2.21). The matrix B occurring in the three iteration schemes is either the diagonal matrix D derived in [6] or the lower triangular matrix T derived in [8]. Thus,

$$B = D := \begin{pmatrix} 0.3205 & 0 & 0 & 0 \\ 0 & 0.0892 & 0 & 0 \\ 0 & 0 & 0.1817 & 0 \\ 0 & 0 & 0 & 0.2334 \end{pmatrix}, \quad B = T := \begin{pmatrix} 0.1130 & 0 & 0 & 0 \\ 0.2344 & 0.2905 & 0 & 0 \\ 0.2167 & 0.4834 & 0.3083 & 0 \\ 0.2205 & 0.4668 & 0.4414 & 0.1176 \end{pmatrix}.$$

Both matrices have the property that $\rho(Z(z)) < 1$ whenever $\mathrm{Re}(z) \le 0$. However, for B = D the maximal value of $\| Z(z)^j \|$ in the lefthand halfplane is greater than 1 for $j \le 4$, whereas for B = T, it is less than 1 for all j. As a consequence, B = T should lead to a much more robust iteration scheme. Furthermore, the matrices A - B that play a role in how close the iteration methods are to the true Newton iteration process are given by

$$A - D = \begin{pmatrix} -0.2075 & -0.0403 & 0.0258 & -0.0099 \\ 0.2344 & 0.1177 & -0.0479 & 0.0160 \\ 0.2167 & 0.4061 & 0.0073 & -0.0242 \\ 0.2205 & 0.3882 & 0.3288 & -0.1709 \end{pmatrix}, \quad A - T = -\begin{pmatrix} 0 & 0.0403 & -0.0258 & 0.0099 \\ 0 & 0.0836 & 0.0479 & -0.0160 \\ 0 & 0.0773 & 0.1192 & 0.0242 \\ 0 & 0.0786 & 0.1126 & 0.0551 \end{pmatrix}.$$

Evidently, the magnitude of A - T is considerably smaller than that of A - D. Finally, we observe that for this Radau IIA corrector the quantity $|\alpha|^{-2}\,\mathrm{Re}(\alpha)$ in the convergence condition (2.18b) is bounded below by 3.2.

Since this paper aims at a comparison of algorithmic properties of the three IDE methods and the effect of the diagonal and triangular modes, we avoided effects of stepsize and iteration strategies by performing the experiments with fixed stepsizes h and fixed numbers of iterations m. Furthermore, the Jacobian and the LU-decompositions were computed in each integration step.

The tables of results in Section 3.5 list the values $\mathrm{csd}_D$ / $\mathrm{csd}_T$, where D and T refer to the diagonal and triangular mode, respectively, and where csd is the minimal number of correct significant digits:

$$\text{(3.1)} \qquad \mathrm{csd} := -\log_{10} \left\| \frac{\mathbf{y}_N - \mathbf{y}(t_N)}{\mathbf{y}(t_N)} \right\|_\infty .$$

Here $\mathbf{y}_N$ denotes the numerical solution at the end point $t_N$, and where the division of vectors should be understood as componentwise division. Divergence will be denoted by csd = -.

### 3.1. The Colpitts oscillator (index 0)

Our first test problem is the IVP for the Colpitts oscillator specified in [9] on the interval [0,1.8]. This IVP of index 0 is described by an implicit ODE system of the form (1.6) with four linear differential equations and with constant, nonsingular capacity matrix Q. For such problems, the general IDE method and the two partitioned IDE methods are identical, so that only results for the general IDE method are listed. The results in Table 3.1 show, surprisingly, that the diagonal mode performs much better than the triangular mode. In all other examples, it is, as expected, the other way around.

### 3.2. The transistor amplifier (index 1)

The second test problem is an IVP for the transistor amplifier given in [4] on the interval [0,0.2] (see also [11]). This nonlinear, eight-dimensional problem of index 1 can be represented in the implicit form (1.6) with a constant, nonpartitioned (but singular) capacity matrix Q, as well as in the semi-explicit form (1.8) with $d_1 = 5$ and $d_2 = 3$. In the implicit form (1.6), only the general IDE method can be applied, whereas the semi-explicit form (1.8) allows application of all three IDE methods. In

order to facilitate a mutual comparison, the csd-values in the Tables 3.2 until 3.5 all refer to the accuracies of the numerical solution of the untransformed problem (1.6).
Table 3.2 lists results for the general IDE method when applied to the implicit form (1.6). It clearly shows the greater robustness of the triangular mode.
Next, we compare the three IDE methods when applied to the semi-explicit form (1.8) of the transistor amplifier. The general IDE method produces the same results as listed in Table 3.2 (algebraically, the two approaches are equivalent). The Tables 3.3 and 3.4 present the results obtained by the partitioned IDE methods. As expected, the triangular mode is again superior to the diagonal mode. Furthermore, we see that for larger stepsizes and low numbers of iterations, the partitioned IDE methods are more robust than the general IDE method. Nevertheless, we may conclude that the general IDE method in triangular mode performs best for the transistor problem.
In order to appreciate these results, we give results obtained in [11] by means of the RADAU5 code of Hairer and Wanner [4]. In this table, TOL denotes the tolerance parameter (the absolute and relative error tolerance both equal TOL), $h_{av}$ is the averaged (accepted) integration step, and $m_{av}$, $JE_{av}$, $LU_{av}$ are the averaged values per (accepted) step of the number of iterations, Jacobian evaluations and LU decompositions, respectively. Table 3.5 shows that RADAU5 computed the Jacobian and the LU decomposition in (at least) each integration step. Thus, in this respect, both methods are comparable. However, the LU decompositions needed in the general IDE method has more intrinsic parallelism than those needed in RADAU5. Hence, on parallel computer systems, the sequential (or effective) LU costs of the IDE method will be less. On the other hand, a comparison with the Tables 3.2, 3.3 and 3.4 reveals that the parallel IDE methods need about twice as many iterations than RADAU5 to get the corrector equation more or less solved.

### 3.3. The Arnold-Strehmel-Weiner problem (index 2)

In the paper [1] of Arnold, Strehmel and Weiner, we find the index 2 test problem

$$
\begin{aligned}
&\dot{u} = u^2 - \tfrac{1}{2} v - \tfrac{1}{4} uw - \tfrac{3}{4} w^2, \\
&\dot{v} = \tfrac{1}{2} u^2 w + \tfrac{3}{4} uw^2 + \tfrac{3}{4} w^3 + \tfrac{1}{2} v^2 w, \qquad 0.5 \le t \le 0.6, \\
&0 = 4u^2 + v^2 - 4.
\end{aligned} \tag{3.2}
$$

with exact solution u = w = cos(t), v = 2sin(t). Only the general IDE method and the partitioned IDE method I can be applied. The Tables 3.6 and 3.7 give the results. Evidently, it is here where the partitioned method is by far superior to the general IDE method, particularly, for small m.

### 3.4. The pendulum problem (index 3)

A familiar higher-index test problem is the mathematical pendulum. In index 3 form it reads [5]

$$
\begin{aligned}
&\dot{p} = u, && p(0) = 1, \\
&\dot{q} = v, && q(0) = 0,
\end{aligned}
$$

(3.3) $\dot{u} = -p\lambda, \quad u(0) = 0, \quad 0 \le t \le 10.$

$\dot{v} = -q\lambda - 1, \quad v(0) = 0,$

$0 = p^2 + q^2 - 1, \quad \lambda(0) = 0.$

The Tables 3.8 and 3.9 again show that the partitioned method is faster than the general IDE method.

**Table 3.1.** General IDE method (2.7)
Colpitts oscillator (form (1.6))

| $h.10^6$ | m = 3 | m = 4 | m = 5 | m = 6 |
|---|---|---|---|---|
| 7.2 | 4.0/3.0 | 5.7/4.2 | 6.1/5.6 | 7.4/6.4 |
| 3.6 | 4.8/4.2 | 7.8/5.8 | 7.9/7.3 | 9.4/8.7 |

**Table 3.2.** General IDE method (2.7)
Transistor amplifier (form (1.6))

| $h.10^6$ | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| 0.04 | - | - /6.8 | - /6.5 | 5.7/6.5 |
| 0.02 | 6.0/8.0 | 7.9/8.6 | 8.0/8.8 | 8.6/9.3 |

**Table 3.3.** Partitioned IDE method I (2.13a)
Transistor amplifier (form (1.8))

| h | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| $4_{10}{}^{-4}$ | 3.3/3.8 | 3.6/4.0 | 3.9/4.3 | 4.2/4.7 |
| $2_{10}{}^{-4}$ | 5.1/5.4 | 5.7/6.2 | 6.3/6.9 | 7.0/7.6 |

**Table 3.4.** Partitioned IDE method II (2.21)
Transistor amplifier (form (1.8))

| h | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| $4_{10}{}^{-4}$ | 4.1/4.1 | 5.1/4.5 | 4.8/5.0 | 5.6/5.6 |
| $2_{10}{}^{-4}$ | 5.5/5.9 | 6.3/6.4 | 8.1/7.0 | 7.3/7.7 |

**Table 3.5.** RADAU5 applied to the transistor amplifier (form (1.6))

| TOL | csd | $h_{av}$ | $m_{av}$ | $JE_{av}$ | $LU_{av}$ |
|---|---|---|---|---|---|
| $10^{-4}$ | 4.6 | $3.6\ 10^{-4}$ | 3.0 | 0.99 | 1.31 |
| $10^{-7}$ | 8.3 | $0.7\ 10^{-4}$ | 2.6 | 0.99 | 1.01 |

**Table 3.6.** General IDE method (2.7)
Arnold-Strehmel-Weiner problem

| h | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| 0.02 | - /2.8 | - /5.3 | 2.5/6.7 | 5.6/7.7 |
| 0.01 | - /5.5 | 4.4/6.0 | 5.9/7.4 | 7.2/8.8 |

**Table 3.7.** Partitioned IDE method I (2.13a)
Arnold-Strehmel-Weiner problem

| h | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| 0.02 | - /6.7 | 5.1/7.4 | 6.8/8.2 | 7.5/9.1 |
| 0.01 | 3.8/9.0 | 6.0/9.2 | 8.0/9.9 | 8.9/10.9 |

**Table 3.8.** General IDE method (2.7)
Pendulum problem

| h | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| 0.1 | - | - | - /1.9 | - /2.1 |
| 0.05 | - /2.7 | - /1.3 | 2.4/3.6 | - /3.6 |
| 0.025 | - /3.8 | 3.2/3.8 | 3.1/4.6 | 4.0/5.4 |

**Table 3.9.** Partitioned IDE method I (2.13a)
Pendulum problem

| h | m = 4 | m = 5 | m = 6 | m = 7 |
|---|---|---|---|---|
| 0.1 | - /4.3 | - /3.6 | 3.7/3.2 | 4.1/3.7 |
| 0.05 | - /4.9 | - /4.5 | 4.9/4.3 | 4.6/4.5 |
| 0.025 | - /5.5 | 4.7/5.4 | 6.1/5.4 | 5.4/5.4 |

**Acknowledgement.** The authors are grateful to Dr. B.P. Sommeijer for his careful reading of and the many valuable discussions on this paper.

## References

[1] Arnold, M., Strehmel, K. & Weiner, R. (1995): Errors in the numerical solution of nonlinear differential-algebraic systems of index 2, Report No. 11, Faculty of Mathematics and Computer Science, Martin-Luther-Universität, Halle-Wittenberg.

[2] Brenan, K.E., Campbell, S.L. & Petzold, L.R. (1989): Numerical solution of initial-value problems in differential-algebraic equations, North-Holland, New York.

[3] Desoer, C.A. & Haneda, H. (1972): The measure of a matrix as a tool to analyze computer algorithms for circuit analysis, IEEE Transactions on Circuit Theory 19, 480-486.

[4] Hairer, E. & Wanner, G. (1991): Solving ordinary differential equations, II. Stiff and differential-algebraic problems, Springer-Verlag, Berlin.

[5] Hairer, E., Lubich, Chr, & Roche, M. (1989): The numerical solution of differential-algebraic systems by Runge-Kutta methods, LNM 1409, Springer, Berlin.

[6] Houwen, P.J. van der, & Sommeijer, B.P. (1991): Iterated Runge-Kutta methods on parallel computers, SIAM J. Sci. Stat. Comput. 12, 1000-1028.

[7] Houwen, P.J. van der, & Sommeijer, B.P. (1996): Iteration of Runge-Kutta methods with block triangular Jacobians, to appear in ZAMM.

[8] Houwen, P.J. van der, & Swart, J.J.B. de (1995): Triangularly implicit iteration methods for ODE-IVP solvers, Preprint NM-R9510, CWI, Amsterdam.

[9] Kampowsky, W., Rentrop, P. & Schmidt, W. (1992): Classification and numerical simulation of electric circuits, Surveys on Mathematics for Industry 2, 23-65.

[10] Ström, T. (1975): On logarithmic norms, SIAM J. Numer. Anal. 12, 741-753.

[11] Swart, J.J.B. de, Veen, W.A. van der, & Sommeijer, B.P. (1995): Test set for IVPODE and IVPDAE solvers, available via anonymous ftp at the site *ftp.cwi.nl* in the directory *pub/IVPtestset* in the (compressed) file *testset.ps.Z.*