CWI Quarterly Volume 11 (1) 1998, pp. 33 – 53

CWI Contributions to the Development of Parallel Runge-Kutta Methods

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

This paper will concentrate on contributions of CWI to the development of parallel Runge-Kutta (RK) methods. We shall describe two approaches to construct such methods. In both approaches, a conventional implicit RK method is used as a corrector equation whose solution is approximated by an iterative method. In the first approach, the iteration method uses a fixed number of iterations without solving the corrector. Assuming that a one-step predictor is used, this approach again results in an RK method, however, an RK method possessing a lot of intrinsic parallelism. In the second approach, the corrector is solved by modified Newton iteration and the linear systems arising in each Newton iteration are solved by a parallel iteration process which is tuned to the special form of these linear systems. Furthermore, we apply the parallel iteration process in a step-parallel fashion which further enhances the amount of parallelism. Finally, the application of parallel RK methods within the framework of waveform relaxation is briefly discussed.

1. Introduction

We will be concerned with the solution of the initial-value problem (IVP)

$$\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}), \quad \mathbf{y}(t_0) = \mathbf{y}_0, \quad \mathbf{y}, \mathbf{f} \in \mathbb{R}^d \tag{1.1}$$

by Runge-Kutta (RK) methods on parallel computers. Our starting point is the RK method

$$\begin{aligned} &\mathbf{y}_n = \mathbf{y}_{n-1} + h(\mathbf{b}^T \otimes I)\mathbf{F}(\mathbf{Y}_n), \\ &\mathbf{R}(\mathbf{Y}_n) := \mathbf{Y}_n - h(A \otimes I)\mathbf{F}(\mathbf{Y}_n) - \mathbf{e} \otimes \mathbf{y}_{n-1} = \mathbf{0}. \end{aligned} \tag{1.2}$$

Here, A is the s-by-s Butcher matrix, $\mathbf{b}$ is an s-dimensional vector containing the step point weights, $\mathbf{e}$ is the s-dimensional vector with unit entries, I is the d-by-d identity matrix, h is the stepsize $t_n - t_{n-1}$, and $\otimes$ denotes the Kronecker product. The s components $\mathbf{Y}_{ni}$ of the sd-dimensional solution vector $\mathbf{Y}_n$ (the *stage vector*) represent s numerical approximations to the s exact solution vectors $\mathbf{y}(t_{n-1} + c_i h)$ where $\mathbf{c} = (c_i) := A\mathbf{e}$ denotes the abscissa vector. Furthermore, for any vector $\mathbf{V} = (\mathbf{V}_i), \mathbf{F}(\mathbf{V})$ contains the derivative values $(\mathbf{f}(\mathbf{V}_i))$. It is assumed that the components of $\mathbf{c}$ are distinct and arranged in increasing order. In the following, we shall use the notation I for any identity matrix. However, its order will always be clear from the context.

This paper will concentrate on contributions of CWI to the development of parallel RK methods. We shall describe two approaches to construct such methods. In both approaches, (1.2) is used as a corrector equation whose solution is approximated by an iterative method. In the first approach, the iteration method uses a *fixed number* of iterations and (1.2) is not necessarily solved. Assuming that a one-step predictor is used, this approach again results in an RK method, however, an RK method possessing a lot of intrinsic parallelism. In the second approach, (1.2) is solved by modified Newton iteration and the linear systems arising in each Newton iteration are solved by a parallel iteration process which is tuned to the special form of these linear systems. Sections 2 and 3 describe the construction and analysis of the parallel RK methods and the parallel iterated RK methods. In Section 4, the parallel iteration process is applied in a step-parallel fashion which further enhances the amount of parallelism. Finally, the application of parallel RK methods within the framework of waveform relaxation is briefly discussed in Section 5.

2. Parallel RK methods

Consider the method

$$\mathbf{Y}_n^{(0)} = \mathbf{e} \otimes \mathbf{y}_{n-1} + h(B \otimes I)\mathbf{F}(\mathbf{Y}_n^{(0)}) + h(C \otimes I)\mathbf{F}(\mathbf{e} \otimes \mathbf{y}_{n-1}), \tag{2.1}$$

$$\mathbf{Y}_n^{(j)} = \mathbf{e} \otimes \mathbf{y}_{n-1} + h(B \otimes I)\mathbf{F}(\mathbf{Y}_n^{(j)}) + h((A-B) \otimes I)\mathbf{F}(\mathbf{Y}_n^{(j-1)}), \quad j = 1, \ldots, m, \tag{2.2}$$

$$\mathbf{y}_n = \mathbf{y}_{n-1} + h(\mathbf{b}^T \otimes I)\mathbf{F}(\mathbf{Y}_n^{(m)}), \tag{2.3a}$$

where B and C are appropriately chosen matrices and m is a fixed integer. This method can be interpreted as an iterative method with a fixed number of iterations. Evidently, if $m \to \infty$ and if $\mathbf{Y}_n^{(j)}$ converges, then $\mathbf{Y}_n^{(j)}$ converges to the solution $\mathbf{Y}_n$ of (1.2). However, for m fixed, we may also interpret {(2.1),(2.2),(2.3a)} as an RK method with Butcher tableau as given in Figure 2.1a.

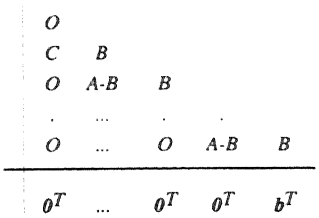

FIGURE 2.1a.

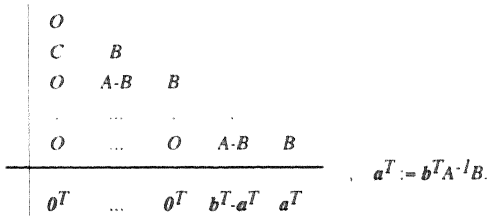

FIGURE 2.1b.

In the case of stiff problems it is recommendable to replace (2.3a) by the formula (see SHAMPINE [49])

$$\mathbf{y}_n = \mathbf{y}_{n-1} + (\mathbf{b}^T A^{-1} \otimes I)(\mathbf{Y}_n^{(m)} - \mathbf{e} \otimes \mathbf{y}_{n-1}), \tag{2.3b}$$

provided that A is nonsingular (we remark that for stiffly accurate RK methods, where $\mathbf{b}^T = \mathbf{e}_s^T A$, formula (2.3b) reduces to $\mathbf{y}_n = (\mathbf{e}_s^T \otimes I)\mathbf{Y}_n^{(m)}$). The step-point formulas (2.3a) and (2.3b) will be referred to as the *conventional* and the *Shampine* step-point formula. The Butcher tableau for {(2.1),(2.2), (2.3b)} is given in Figure 2.1b. Methods of the type (2.2) can also be based on more general correctors than RK formulas (for a survey we refer to BURRAGE [10, 11, 12] and to [51]).

The order of accuracy, the linear stability and the amount of intrinsic parallelism of the methods {(2.1),(2.2),(2.3)} are determined by the matrices A, B and C. We have the following result for the (nonstiff) order of accuracy (see e.g. JACKSON and NØRSETT [35], JACKSON, KVÆRNØ and NØRSETT [34], BURRAGE [8,9], VAN DORSSELAER [17], and the CWI papers [22, 27]).

THEOREM 2.1. *The orders of accuracy of the RK methods {(2.1),(2.2),(2.3a)} and {(2.1),(2.2),(2.3b)} are respectively given by* $p := \min\{p^*, m+q+1\}$ *and* $p := \min\{p^*, m+q\}$, *where* p^* *and* q *denote the orders of the corrector (1.2) and of the predictor formula for* $\mathbf{Y}_n^{(0)}$. *If* $(B+C)\mathbf{e} = \mathbf{c}$, *then* $q \geq 1$ *and if also* $B\mathbf{c} = A\mathbf{c}$, *then* $q \geq 2$. □

Results for the *stiff* order of accuracy are given in [25]. From now on, the order of a method is always meant to be the *nonstiff* order of accuracy.

The linear stability properties are obtained by applying {(2.1),(2.2),(2.3)} to the basic stability test equation $y' = \lambda y$. For the step point formulas (2.3a) and (2.3b), this leads to the respective stability functions

$$\begin{aligned} R_m(z) &= R(z) + z\mathbf{b}^T Z^m(z) Q(z)\mathbf{e}, \\ R_m(z) &= R(z) + \mathbf{b}^T A^{-1} Z^m(z) Q(z)\mathbf{e}, \end{aligned} \tag{2.4}$$

where $z := h\lambda$ and where the matrices Z, Q and the function R are given by

$$\begin{aligned} Z(z) &:= z(I - zB)^{-1}(A - B), \\ Q(z) &:= (I - zB)^{-1}(I + zC) - (I - zA)^{-1}, \\ R(z) &:= 1 + z\mathbf{b}^T(I - zA)^{-1}\mathbf{e}. \end{aligned} \tag{2.5}$$

Here, $R(z)$ is the stability function of the corrector (1.2). In the following sections, we discuss the cases where B vanishes and where B is diagonal.

2.1. Explicit RK methods

For nonstiff problems, we may set $B = C = O$ to obtain an explicit $s(m+1)$-stage RK method requiring $sm + 1$ right-hand side evaluations (in [22] we called such methods Parallel Iterated RK methods, or briefly PIRK methods, in order to indicate that they are based on iterating an RK method). Since in nonstiff situations it is natural to use the conventional step-point formula (2.3a), Theorem 2.1 implies that the order of accuracy is given by

$$p := \min\{p^*, m + q + 1\} = \min\{p^*, m + 1\}.$$

Each block of s stages of this PIRK method can be computed in parallel, so that for $m \leq p^* - 1$, we effectively have an $(m+1)$-stage method of order $m + 1$ (provided that s processors are available). Hence, for $m \leq p^* - 1$, {(2.1),(2.2),(2.3a)} generates an explicit RK method (ERK method) the order of which equals its number of effective (or sequential) stages. ISERLES and NØRSETT [33] showed that this is an optimal result, because the order p of explicit RK methods cannot exceed the number s_{seq} of sequential stages (see also NØRSETT and SIMONSEN [47]). If we choose for the underlying corrector, the s-stage Gauss-Legendre method, then $p^* = 2s$, so that the number of processors is half the order. The stability polynomials of optimal ERK methods are given by truncated Taylor expansions of $\exp(z)$, the stability regions of which can be found in the literature (cf., e.g., [19]). Experiments on four-processor Alliant computers were performed at the University of Trondheim [36, 40] and at CWI [22]. These experiments showed that parallel RK methods of the above type are quite efficient.

REMARK 2.1. Optimal ERK methods can also be generated by Richardson extrapolation (see, e.g., [23] and [50]). In particular, extrapolation of the explicit midpoint rule generates an optimal ERK of order p which only needs $[1 + p/4]$ processors (here, [.] denotes the integer part function). However, the experiments in [23] indicate that they are more expensive than Gauss-Legendre-based methods. □

2.2. Diagonally implicit RK methods

Parallel diagonally implicit RK methods arise if B is a diagonal matrix D. Because of the 'diagonal' implicitness, each block of s stages can be computed in parallel, so that effectively, we only have $m + 1$ implicit stages. The stability

Order	p_{st}	s_{seq}	K	Stability	Remarks
$p = 3$	1	p-1	1	A	DIRK, NØRSETT [44]
$p = 3$	2	p-1	1	$> A$	DIRK, CROUZEIX [16]
$p = 4$	1	p-1	1	A	DIRK, CROUZEIX [16], ALEXANDER [1]
$p = 4$	1	p-2	2	L	PARK, ISERLES & NØRSETT [33]
$p = 3, 4, 5$	s	p-1	s	$> A$	Parallel DIRK, $C = O$, $D = diag(c)$, [27]
$p = 6, 7$	s	p-1	s	$> A(\alpha)$	Parallel DIRK, $C = O$, $D = diag(c)$, [27]
$p = 3, 5, 7$	s	p	s	$> A$	Parallel DIRK, $C = A$-D, $\rho(I$-$D^{-1}A) = 0$, [24]
$p \le 6, p = 8$	s	p	s	L	Parallel DIRK, $C = O$, $D = \delta I$, [27], [45]
$p \le 8, p = 10$	s	p+1	s	L	Parallel DIRK, $C = O$, $D = \delta I$, [27], [45]

TABLE 2.1. Characteristics of DIRK, PARK and Parallel DIRK methods.

regions can be computed from the stability functions (2.4). In [27] and [24], this has been done for several choices of $B = D$ and C. Table 2.1 specifies the main characteristics of a number of these parallel DIRK methods. For reasons of comparison, we also list characteristics of conventional DIRK methods and a parallel RK (PARK) method of ISERLES and NØRSETT [33]. In this table, p_{st} denotes the block-stage order, s_{seq} the number of implicit sequential stages, and K the number of processors needed. Furthermore, A-stability, $A(\alpha)$-stability, L-stability, and strong A-stability and $A(\alpha)$-stability are respectively indicated by $A, A(\alpha), L, > A$, and $> A(\alpha)$. All these methods need only one LU-decomposition per processor. The methods referred to in the fifth and sixth row of this table use either Gauss-Legendre or Radau IIA as the underlying corrector, both with step-point formula (2.3a). In the methods of the last three rows, the corrector is Radau IIA with step-point formula (2.3b), and in the methods of the last two rows, D is determined by the restricted Padé approximants of NØRSETT [45] (see also WOLFBRANDT [54]). With respect to its order, the PARK method of ISERLES and NØRSETT needs a surprisingly low number of sequential stages and yet it is L-stable. The parallel DIRK methods have the advantage of a relatively high stage order and step-point order.

3. PARALLEL ITERATED RK METHODS

The conventional approach of solving the corrector equation (1.2) is the modified (or simplified) Newton iteration scheme

$$(I - A \otimes hJ_n)(\mathbf{Y}_n^{(j)} - \mathbf{Y}_n^{(j-1)}) = -\mathbf{R}(\mathbf{Y}_n^{(j-1)}), \quad j = 1, \ldots, m, \tag{3.1}$$

where J_n is the Jacobian of the right-hand side function $\mathbf{f}$ at t_n and $\mathbf{Y}_n^{(0)}$ is the initial iterate to be provided by some predictor formula. The most powerful RK methods with respect to order of accuracy and stability (such as those based on Gaussian quadrature) possess a full Butcher matrix A, so that each iteration with (3.1) requires the solution of an sd-dimensional linear system for

the Newton correction $\mathbf{Y}_n^{(j)} - \mathbf{Y}_n^{(j-1)}$. If direct solution methods are used, then the costs for solving the linear systems usually are extremely high, particularly for large values of sd, because of the expensive LU-decompositions. As pointed out by Butcher in 1976, LU-costs can be reduced by using a transformation $\mathbf{Y}_n^{(j)} = (Q \otimes I)\mathbf{U}_n^{(j)}$ to obtain transformed linear systems with a matrix of coefficients of the form $I - Q^{-1}AQ \otimes hJ_n$ (assuming that Q is nonsingular). Hence, by choosing Q such that $Q^{-1}AQ$ has a (block) diagonal or (block) triangular structure, the transformed systems can be split into subsystems of dimension less than sd (see [13, 14]). Unfortunately, RK methods of Gauss-Legendre and Radau type possess a Butcher matrix with at most one real eigenvalue, so that the best we can achieve is either complex-valued subsystems of dimension d or real-valued subsystems of dimension $2d$ (cf. HAIRER and WANNER [19, p.130]). To circumvent this overhead in the linear algebra part, NØRSETT [46] introduced RK methods with an A-matrix possessing a real, one-point spectrum. Using the BUTCHER transformation [13], these methods can be implemented in such a way that only real-valued systems of dimension d have to be solved. This work was then extended by BURRAGE [7], who also derived reference formulas for error control. These so-called SIRK methods are particularly suitable for implementation on sequential computer systems, since they require only one LU-decomposition of dimension d per Jacobian or stepsize update. On parallel computer systems, we may drop the 'one-point spectrum' requirement, because the LU-decompositions needed in the transformed subsystems can be computed in parallel. Hence, the Butcher transformation is a means to introduce parallelism into RK schemes. For example, if A has a *real spectrum* such as the multi-implicit RK methods of NØRSETT [46] and OREL [48]. Effectively, these methods require only one LU-decomposition of dimension d per Jacobian or stepsize update.

At CWI we did not change the RK method, but we changed the iteration process for solving the corrector equation (1.2). We designed parallel iteration processes with the property that only real-valued, linear systems of dimension d are to be solved.

3.1. PDIRK and PTIRK methods

In the preceding section, we used (2.2) with a fixed number of iterations and B was chosen to achieve high accuracy and good stability. Let us now use (2.2) as a (nonlinear) iterative solver for approximating the solution of the corrector (1.2) and let B be chosen to achieve fast convergence to the corrector solution. We remark that in an actual application, the solution $\mathbf{Y}_n^{(j)}$ of (2.2) is often approximated by just one modified Newton iteration. In that case, the iteration method (2.2) reduces to a process of the form (3.1) with A replaced by B.

3.1.1. PDIRK methods. As in Section 2.2, we may choose B diagonally. This results into the Parallel Diagonal-implicit Iterated RK (PDIRK) methods anal-

ysed in [24]. They possess the same parallel features as the parallel DIRK methods. It turned out that in the case of stiff IVPs, it is crucial for a fast convergence that $\rho(Z(\infty)) = \rho(I - D^{-1}A)$ is small (here, $Z(z)$ is defined as in (2.5)). In [24], the matrix $B = D$ was determined by minimizing the value of $\rho(I - D^{-1}A)$ by a computer search. The resulting matrices D indicated that it is highly likely that there exist matrices D such that $\rho(Z(\infty))$ actually vanishes. This led us to pose the problem:

PROBLEM 3.1. For what class of Butcher matrices A do there exist diagonal matrices D with positive diagonal entries such that $D^{-1}A$ has a one-point spectrum at 1? □

If such a matrix D exists, then the diagonal entries of D are determined by the (nonlinear) system that is obtained by requiring that the equation

$$\det(D^{-1}A - \mu I) = 0$$

has only zeros equal to 1. In this way, LIOEN [41] showed the following result:

THEOREM 3.1. *For s-stage Radau IIA correctors with $s = 2, \ldots, 8$, there do exist diagonal matrices D with positive diagonal entries such that $D^{-1}A$ has a one-point spectrum at 1.* □

The matrices D derived by Lioen all generate A-convergent PDIRK methods (here, a method is called A-convergent if its region of convergence contains the whole left halfplane, see Section 3.2.1.). However, the convergence in the initial phase of the PDIRK iteration process may be rather slow.

3.1.2. PTIRK methods. One of the research issues of the ODE group at CWI has been the improvement of the rate of convergence of PDIRK methods, particularly with respect to the initial phase of the iteration process. One option is to choose the matrix $B = T$ where T is lower triangular with positive diagonal entries. Such methods were called PTIRK methods [30]. The LU-decomposition of $I - T \otimes hJ_n$ again splits into s parallel LU-decompositions of dimension d. If T is nondefective, then we may perform a Butcher transformation $\mathbf{Y}_n^{(j)} = (Q \otimes I)\tilde{\mathbf{Y}}_n^{(j)}$, with nonsingular Q, such that $Q^{-1}TQ$ is diagonal. In this way, we can obtain 'diagonal' implicitness as in the PDIRK methods. As for the PDIRK methods, it is again crucial that $\rho(Z(\infty)) = \rho(I - T^{-1}A)$ is as small as possible. The following result was proved in [20, 30]:

THEOREM 3.2. *Let A be defined by any collocation method with positive abscissae and let $A = TU$ be the Crout decomposition of A with T lower triangular and U unit upper triangular. Then, $\rho(I - T^{-1}A)$ vanishes and T has positive diagonal entries.* □

For a large number of RK methods, we computed the convergence regions of the generated PTIRK method which were all found A-convergent. Moreover,

the rate of convergence in the initial phase is considerably improved (see the experiments reported in [30]).

3.2. PILSRK methods

The iteration processes described in the preceding section are *nonlinear* solvers for RK systems. Quite recently, we started an alternative approach. Our point of departure is the modified Newton method (3.1). In order to avoid linear systems of dimension 2d, we solved the linear Newton systems in (3.1) iteratively by an *inner* iteration process which only requires the solution of d-dimensional systems. This leads to the inner-outer iteration method

$$
\begin{aligned}
&\mathbf{Y}_n^{(0,r)} = \text{initial approximation to } \mathbf{Y}_n \\
&\textbf{For } j = 1 \textbf{ to } m \\
&\quad \mathbf{Y}_n^{(j,0)} = \mathbf{Y}_n^{(j-1,r)} \\
&\quad \textbf{For } \nu = 1 \textbf{ to } r \\
&\qquad (I - B \otimes hJ_n)(\mathbf{Y}_n^{(j,\nu)} - \mathbf{Y}_n^{(j,\nu-1)}) = \\
&\qquad -(I - A \otimes hJ_n)(\mathbf{Y}_n^{(j,\nu-1)} - \mathbf{Y}_n^{(j-1,r)}) - \mathbf{R}(\mathbf{Y}_n^{(j-1,r)}), \\
&\qquad \textbf{either } \mathbf{y}_n^{(j,\nu)} = \mathbf{y}_{n-1}^{(m,r)} + h(\mathbf{b}^T \otimes I)\mathbf{F}(\mathbf{Y}_n^{(j,\nu)}) \\
&\qquad \textbf{or} \qquad \mathbf{y}_n^{(j,\nu)} = \mathbf{y}_{n-1}^{(m,r)} + (\mathbf{b}^T A^{-1} \otimes I)(\mathbf{Y}_n^{(j,\nu)} - \mathbf{e} \otimes \mathbf{y}_{n-1}^{(m,r)}),
\end{aligned}
\tag{3.2}
$$

where B is a free matrix with real entries and positive eigenvalues. The inner loop of (3.2) represents the inner iteration process with inner iterates $\mathbf{Y}_n^{(j,\nu)}$ and $\mathbf{y}_n^{(j,\nu)}, \nu = 1, \ldots, r$. We shall refer to this process as a Parallel Iterative Linear System method for RK systems (PILSRK method). The process defining the *outer* iterates $\mathbf{Y}_n^{(j,r)}$ and $\mathbf{y}_n^{(j,r)}, j = 1, \ldots, m$, will be called the *outer iteration process*. Obviously, if the inner iterates converge as $r \to \infty$, then they converge to the solution $\mathbf{Y}_n^{(j)}$ of (3.1). Before discussing the choice of suitable matrices B, we consider convergence and stability aspects of the iteration process (3.2).

3.2.1. The region of convergence. In order to derive convergence conditions, let $\mathbf{U}_n^{(j)}$ be the solution of the equation

$$
(I - A \otimes hJ_n)(\mathbf{U}_n^{(j)} - \mathbf{Y}_n^{(j-1,r)}) = -\mathbf{R}(\mathbf{Y}_n^{(j-1,r)}), \tag{3.1'}
$$

and define the inner iteration error $\delta_{j,\nu}$, the modified Newton error θ_j, and the total iteration error $\varepsilon_{j,\nu}$, i.e.,

$$
\delta_{j,\nu} := \mathbf{Y}_n^{(j,\nu)} - \mathbf{U}_n^{(j)}, \quad \theta_j := \mathbf{Y}_n^{(j)} - \mathbf{Y}_n, \quad \varepsilon_{j,\nu} := \mathbf{Y}_n^{(j,\nu)} - \mathbf{Y}_n \tag{3.3}
$$

with $\varepsilon_{j,0} := \varepsilon_{j-1,r}$. Furthermore, we need

$$\begin{aligned} \mathbf{G}(\Delta) &:= \mathbf{F}(\mathbf{Y}_n + \Delta) - \mathbf{F}(\mathbf{Y}_n) - (I \otimes J_n)\Delta, \\ M &:= (I - B \otimes hJ_n)^{-1}((A - B) \otimes hJ_n), \\ N_1 &:= (I - B \otimes hJ_n)^{-1}(A \otimes I),\ N_2 := (I - A \otimes hJ_n)^{-1}(A \otimes I). \end{aligned} \tag{3.4}$$

From (3.1) and (3.2) we derive the error recursions

$$\delta_{j,\nu} = M\delta_{j,\nu-1},\ \theta_j = hN_2\mathbf{G}(\theta_{j-1}),\ \varepsilon_{j,\nu} = M\varepsilon_{j,\nu-1} + hN_1\mathbf{G}(\varepsilon_{j-1,r}), \tag{3.5}$$

where $j = 1, \ldots, m$ and $\nu = 1, \ldots, r$. From the relation for $\delta_{j,\nu}$ we see that the inner iteration process converges if the spectral radius $\rho(M)$ of M is less than 1. Since the spectrum $\sigma(M)$ of M is given by that of the matrix $Z(z)$ defined in (2.5) with $z \in \sigma(hJ_n)$, we are led to define the *region of convergence* of the inner iteration process by $\Gamma := \{z : \rho(Z(z)) < 1\}$. We shall call $Z(z)$ the *amplification* matrix at the point z and $\rho(Z(z))$ the *(asymptotic) amplification factor* at z. Its maximum in the nonpositive halfplane $\mathrm{Re}(z) \leq 0$ will be denoted by ρ. If $\rho < 1$, i.e. Γ contains the whole nonpositive halfplane, then the inner iteration process will be called *A-convergent.*

THEOREM 3.3. *The PILSRK method converges as $r \to \infty$ if $\sigma(hJ_n) \in \Gamma$.* □

A simple manipulation reveals that

$$\varepsilon_{j,r} = M^r\varepsilon_{j-1,r} + h(I - M^r)N_2\mathbf{G}(\varepsilon_{j-1,r}),\ j = 1, \ldots, m. \tag{3.6}$$

Hence, if $\sigma(hJ_n) \in \Gamma$ and if $\varepsilon_{1,0} = \theta_0$, then it follows from (3.5) and (3.6) that $\varepsilon_{j,\infty}$ and θ_j satisfy the same error recursion. Thus, if the modified Newton method (3.1) converges and if $\sigma(hJ_n) \in \Gamma$, then the iteration process (3.2) converges as $m, r \to \infty$.

3.2.2. The order of accuracy of the iterates. To obtain further insight into the convergence behaviour, we consider the order of accuracy of the method (3.2) after a finite number of inner and outer iterations. Let $\varepsilon_{j,r} = O(h^{p(j)})$. Then it follows from (3.4) and (3.6) that $p(j)$ satisfies the recursion

$$p(0) = q + 1,\ \ p(j) = p(j-1) + \min\{r, 2\},\ j = 1, \ldots, m, \tag{3.7}$$

where q is the order of the predictor. Since $\mathbf{Y}_n^{(m,r)} = \mathbf{Y}_n + \varepsilon_{m,r}$, we derive from (3.7) the result:

THEOREM 3.4. *Let $p_0 = \min\{r, 2\}$ and let p^* and q denote the orders of the corrector (1.2) and of the predictor formula for $\mathbf{Y}_n^{(0,r)}$. Irrespective the structure of B, the order of accuracy of the method (3.2) is given by $p := \min\{p^*, q + 1 + mp_0\}$ when using the conventional step-point formula and by $p := \min\{p^*, q + mp_0\}$ when using the Shampine step-point formula.* □

This theorem shows that with respect to order of accuracy, it is recommendable to perform at least two inner iterations, so that for the step-point formulas (2.3a) and (2.3b) the order of the corrector is reached within $[(p^* - q)/2]$ and

$[(p^* - q + 1)/2]$ outer iterations (we recall that the order of accuracy is understood to be the *nonstiff* order; for *stiff* order considerations of Newton-like processes we refer to the work of VAN DORSSELAER and SPIJKER [17, 52]). For example, for Radau IIA correctors with extrapolation predictor of order $q = s - 1$ (see (3.11) below) and step-point formula (2.3b), we find that for at least two inner iterations (i.e., $r \geq 2$), the order of accuracy is given by $p := \min\{2s - 1, s - 1 + 2m\}$. Thus, the order of the corrector is attained for $[(s + 1)/2]$ or more outer iterations.

3.2.3. The PILSRK amplification factors. Next we address the speed of convergence of the PILSRK method. Since M is not expected to be a normal matrix, the asymptotic amplification factor ρ defined above only gives information on the speed of convergence after many inner iterations and does not give insight into the convergence behaviour in the initial phase of the iteration process. However, by using a generalization of a theorem of Von Neumann due to NEVANLINNA [43] (see also [19, p.356]), we can prove the theorem:

THEOREM 3.5. *Let* $||.||_2$ *denote the Euclidean matrix norm, and let* $\mu_2[.]$ *be the corresponding logarithmic norm. If* $\mu_2[J_n] \leq 0$, *then* $||M^r||_2 \leq \max_{\mathrm{Re}(z)\leq 0} ||Z^r(z)||_2$. □

This theorem suggests characterizing the convergence behaviour of PILSRK methods by the (averaged) amplification factors

$$\rho^{(r)} = \max_{\mathrm{Re}(z)\leq 0} \rho^{(r)}(z), \quad \rho^{(r)}(z) := \sqrt[r]{||Z^r(z)||_2}. \tag{3.8}$$

3.2.4. Stability. Finally, we discuss the stability of the method after a finite number of inner and outer iterations. Stability also plays an important role, because stability for small values of r and m implies that we can produce stable results at low computational costs. This is particularly important in step-parallel applications of the scheme (3.2) (cf. Section 4). Therefore, it is of interest to know the minimal number of iterations in order to ensure that (3.2) is sufficiently stable. For the test equation $y' = \lambda y$, we have $\mathbf{Y}_n = (I - zA)^{-1}\mathbf{e}y_{n-1}^{(m,r)}$, so that we deduce from (3.3) and (3.6)

$$\begin{aligned} \mathbf{Y}_n^{(m,r)} &= \mathbf{Y}_n + Z^{mr}(\mathbf{Y}_n^{(0,r)} - \mathbf{Y}_n) \\ &= (I - Z^{mr})(I - zA)^{-1}\mathbf{e}y_{n-1}^{(m,r)} + Z^{mr}\mathbf{Y}_n^{(0,r)}. \end{aligned} \tag{3.9}$$

The stability behaviour is highly dependent on the predictor formula for $\mathbf{Y}_n^{(0,r)}$ used. We shall consider last step-point value (LSV) predictors that are only based on $\mathbf{y}_{n-1}^{(m,r)}$ and extrapolation (EPL) predictors based on $\mathbf{y}_{n-1}^{(m,r)}$ and $\mathbf{Y}_{n-1}^{(m,r)}$. They can both be cast into the form

$$\mathbf{Y}_n^{(0,r)} = \mathbf{p} \otimes \mathbf{y}_{n-1}^{(m,r)} + (P \otimes I)\mathbf{Y}_{n-1}^{(m,r)}, \tag{3.10}$$

B	$s=4$	r^*	$s=8$	r^*
PDIRK	[3.6, 2.5, 1.6, ... , 0.52]	5	[20, 12, 7.7, ... , 0.90]	>40
PTIRK	[0.6, 0.5, 0.5, ... , 0.50]	1	[1.0, 0.9, 0.9, ... , 0.86]	2
(3.14), $\tilde{A}_{ij}=O, \gamma=1$	[2.2, 1.0, 0.8, ... , 0.44]	3	[14, 2.6, 1.6, ... , 0.64]	7

TABLE 3.1. PILSRK amplification factors $[\rho^{(1)}, \rho^{(2)}, \rho^{(3)}, \ldots, \rho^{(\infty)}]$ for Radau IIA correctors.

where the s-by-s matrix P and the s-dimensional vector $\mathbf{p}$ are determined by order conditions. For LSV predictors we have $\mathbf{p} = \mathbf{e}, P = O$ and order $q = 0$. If $c_s = 1$, then we have for EPL predictors

$$\begin{aligned} &\mathbf{p} := \mathbf{0}, \quad P := WV^{-1}, \\ &V := ((\mathbf{c}-\mathbf{e})^{i-1}), \quad W := (\mathbf{c}^{i-1}), \quad i = 1, \ldots, s; \; q = s-1, \end{aligned} \tag{3.11}$$

where powers of vectors are defined componentwise. On substitution of (3.10) into (3.9), we find for the test equation the relation

$$\begin{aligned} \mathbf{Y}_n^{(m,r)} &= (I - zA)^{-1}\mathbf{e}y_{n-1}^{(m,r)} - Z^{mr}(I - zA)^{-1}\mathbf{e}y_{n-1}^{(m,r)} \\ &\quad + Z^{mr}\mathbf{p}y_{n-1}^{(m,r)} + Z^{mr}P\mathbf{Y}_{n-1}^{(m,r)}. \end{aligned}$$

Together with the step-point formulas (2.3), we obtain a linear recursion for the pair $(y_n^{(m,r)}, \mathbf{Y}_n^{(m,r)})$. The stability is determined by the magnitude of the characteristic roots of this recursion. In the particular case of stiffly accurate RK methods (as in Radau IIA correctors) where $c_s = 1$ and $\mathbf{b}^T = \mathbf{e}_s^T A$, we have $\mathbf{y}_n^{(m,r)} = (\mathbf{e}_s^T \otimes I)\mathbf{Y}_n^{(m,r)}$, so that

$$\begin{aligned} \mathbf{Y}_n^{(m,r)} &= S_{mr}(z)\mathbf{Y}_{n-1}^{(m,r)}, \\ S_{mr}(z) &:= (I - Z^{mr})(I - zA)^{-1}\mathbf{e}\mathbf{e}_s^T + Z^{mr}(\mathbf{p}\mathbf{e}_s^T + P), \end{aligned} \tag{3.12}$$

the characteristic roots of which are given by the spectrum $\sigma(S_{mr}(z))$ of the stability matrix $S_{mr}(z)$. In applications, it is advantageous to have an L-stable method. Since A-stability automatically implies L-stability if $\rho(S_{mr}(\infty))$ vanishes, we are led to consider $S_{mr}(\infty) = Z^{mr}(\infty)(\mathbf{p}\mathbf{e}_s^T + P)$. Since B is nonsingular (because B is assumed to have positive eigenvalues), $S_{mr}(\infty) = (I - B^{-1}A)^{mr}(\mathbf{p}\mathbf{e}_s^T + P)$. By observing that $(I - B^{-1}A)^{mr}$ vanishes for $mr \geq s$ if $\rho(I - B^{-1}A)$ vanishes, we have the result:

THEOREM 3.6. *Let $c_s = 1$, $mr \geq s$, $\rho(I - B^{-1}A) = 0$, let the predictor formula be defined by {(3.10),(3.11)} and let the Shampine step-point formula be used. Then, the method (3.2) is L-stable whenever it is A-stable.* □

Using (3.12), we can compute the maximal spectral radius in the left halfplane $\mathrm{Re}(z) \leq 0$ of the stability matrix $S_{mr}(z)$. This maximum value will be denoted

by $\rho(S_{mr})$. We have A-stability or L-stability if $\rho(S_{mr}) = 1$. The following subsections are devoted to the region of convergence, the amplification factors $\rho^{(\nu)}$, and to stability for a few special choices of the matrices B. The starting point for choosing B is that the linear systems in (3.2) are more efficiently solved than the linear system (3.1) when implemented on a parallel computer system.

3.2.5. PDIRK matrices. Suppose that we choose $B = D$, where D is a diagonal matrix with nonnegative diagonal entries. The linear system in (3.2) is only 'diagonally implicit' and splits into s subsystems, each of dimension d, which can be solved in parallel. In particular, if a direct linear solver is used, then the s LU-decompositions can be computed in parallel, so that effectively only one decomposition is required. Similarly, in each iteration, the s components of the right-hand side and the s forward-backward substitutions can also be computed in parallel.

Evidently, we may use the PDIRK matrices D employed in the PDIRK methods discussed in Section 3.1. However, since the PDIRK methods exhibit a poor initial convergence, we may expect that the inner amplification factors $\rho^{(r)}$ associated with the generated linear solver are relatively large for small r, particularly for larger values of s. In the first row of Table 3.1, these factors are listed for the four-stage and eight-stage Radau IIA correctors (note that $\rho^{(\infty)}$ equals ρ). In addition, we listed the value r^* of r for which the PILSRK amplification factor becomes less than 1. The relatively large values of r^* indicate that the number of iterations needed to achieve sufficient stability is expected to be high when using PDIRK matrices. The value of mr for which $\rho(S_{mr})$ becomes and remains less than or equal to a given number γ will be denoted by $(mr)_\gamma$. For a few values of γ, Table 3.2 lists $(mr)_\gamma$ for the LSV and EPL predictor and for a number of Radau IIA correctors (in order to demonstrate how fast the $(mr)_\gamma$-values increase with s, we have included all correctors with $s \leq 8$). These values show that for $s = 4$ the $(mr)_\gamma$-values are acceptable, but for $s = 8$, PDIRK becomes stable only after a dramatically large number of iterations.

3.2.6. PTIRK matrices. Next we use the PTIRK matrices T used in the PTIRK methods of Section 3.1.2. For the four-stage and eight-stage Radau IIA correctors, the range of inner amplification factors is given in Table 3.1. These figures clearly show the superior convergence behaviour obtained by the PTIRK matrices for small r. Moreover, for finite mr-values the stability is also much better as can be concluded from the $(mr)_\gamma$-values listed in Table 3.2.

3.2.7. Matrices with positive eigenvalues. Our most recent attempt to improve the convergence chooses for B a matrix with the only requirement that its eigenvalues are positive. By performing a Butcher transformation, it is possible to transform the PILSRK method into

B	Predictor	γ	$s=2$	$s=3$	$s=4$	$s=5$	$s=6$	$s=7$	$s=8$
PDIRK	LSV	1.1	1	4	6	10	17	28	58
		1.01	1	4	6	10	18	29	58
		1	1	5	8	10	18	29	59
	EPL	1	1	5	7	14	27	>40	>61
PTIRK	LSV	1.1	1	1	1	1	1	1	1
		1.01	1	1	1	7	8	9	11
		1	1	5	8	11	14	17	20
	EPL	1	1	3	4	10	14	26	>43
(3.14), $\tilde{A}_{ij}=O$	LSV	1.1			2				4
		1.01			2				6
		1			8				19
	EPL	1			5				14

TABLE 3.2. Values of $(mr)_\gamma$.

$$
\begin{aligned}
(I-\tilde{T}\otimes hJ_n)(\tilde{\mathbf{Y}}_n^{(j,\nu)}-\tilde{\mathbf{Y}}_n^{(j,\nu-1)}) = \\
-(I-\tilde{A}\otimes hJ_n)(\tilde{\mathbf{Y}}_n^{(j,\nu-1)}-\tilde{\mathbf{Y}}_n^{(j-1,r)}) \qquad (3.13)\\
-(Q^{-1}\otimes I)\mathbf{R}((Q\otimes I)\tilde{\mathbf{Y}}_n^{(j-1,r)}),
\end{aligned}
$$

where $\tilde{A}=Q^{-1}AQ$ and where $\tilde{T}=Q^{-1}BQ$ is triangular or even diagonal if B is nondefective. A first result is [31]:

THEOREM 3.7. *Let A have eigenvalues $\xi_k \pm i\eta_k$ with $\xi_k > 0$ and define $\alpha_k := \sqrt{\xi_k^2+\eta_k^2}$, $K=\{k : \eta_k \neq 0\}$ and $B=Q\tilde{T}Q^{-1}$, where Q is such that the diagonal blocks of $(\tilde{A})=Q^{-1}AQ=(\tilde{A}_{jk})$ are given by*

$$
\tilde{A}_{kk} := \begin{pmatrix} \gamma\alpha_k & \dfrac{\gamma(\gamma^2\alpha_k-2\gamma\xi_k+\alpha_k)}{1+\gamma^2} \\ -\dfrac{1+\gamma^2}{\gamma}\alpha_k & 2\xi_k-\gamma\alpha_k \end{pmatrix} \text{ if } k\in K, \qquad (3.14a)
$$

$$
\tilde{A}_{kk} := \xi_k \text{ if } k\notin K,
$$

and where

$$
\tilde{T} := \begin{pmatrix} \tilde{T}_{11} & O & O & O & \cdots \\ \tilde{A}_{21} & \tilde{T}_{22} & O & O & \cdots \\ \tilde{A}_{31} & \tilde{A}_{32} & \tilde{T}_{33} & O & \cdots \\ . & . & . & . & . \end{pmatrix},
$$

$$\tilde{T}_{kk} := \begin{pmatrix} \gamma\alpha_k & 0 \\ -\frac{1+\gamma^2}{\gamma}\alpha_k & \frac{\alpha_k}{\gamma} \end{pmatrix} \text{ if } k \in K, \quad \tilde{T}_{kk} := \xi_k \text{ if } k \notin K. \tag{3.14b}$$

Then:

(a) The PILSRK method is A-convergent for all $\gamma > 0$.
(b) In the left halfplane $\mathrm{Re}(z) \leq 0$, *the asymptotic amplification factor vanishes at infinity and is bounded by*

$$\rho = \max\{|1 - 2\gamma(\gamma^2+1)^{-1}\xi_k\alpha_k^{-1}| : k \in K\}. \qquad \square$$

The value of the asymptotic amplification factor ρ is minimized for $\gamma = 1$. However, if $\gamma = 1$, then $\tilde{T}_{kk}$ is defective for $k \in K$, so that B cannot be diagonalized. At the cost of a slight increase of ρ, the defectness of $\tilde{T}_{kk}$ can be removed by choosing γ close to but different from 1. The resulting values of ρ are smaller than for the PILSRK method generated by the PDIRK and PTIRK matrices (see Table 3.1 and recall that $\rho = \rho^{(\infty)}$). In [31] we analysed the case where the lower triangular blocks $\tilde{A}_{ij}$ in $\tilde{T}$ vanish. Using a numerical search, transformation matrices Q with minimal condition number (with respect to the Euclidean norm) were determined for $\gamma \approx 1$. The averaged inner amplification factors and the $(mr)_\gamma$-values corresponding with these matrices Q are listed in the Tables 3.1 and 3.2. Table 3.1 shows that the initial and asymptotic amplification factors are respectively larger and smaller than those associated with PTIRK matrices, while Table 3.2 implies that for $s = 8$ block-diagonal matrices of the form (3.14) are much more stable than the PTIRK matrices.

4. Step-parallel iteration

In methods employing *step-parallel* iteration, the iteration procedure is concurrently applied at a number of step-points, that is, the iteration process at the point t_n is already started without waiting until the iterates $\mathbf{Y}^{(j)}$ at t_{n-1} have converged. Step-parallel methods and its various versions (also called frontal methods) have been discussed and analysed in a number of papers, among which Miranker and Liniger [42], Bellen [3], Bellen et al. [4, 6], Burrage [12], Gear and Xu Xuhai [18], Chartier [15], and Augustyn and Überhuber [2]. Further references can be found in Burrage [12].

In the following, we survey step-parallel methods developed at CWI. These methods can be seen as step-parallel versions of the scheme (3.2). The 'step-parallelization' of (3.2) consists of a modification of the predictor formula and of the residual function $\mathbf{R}$. In order to specify this modified scheme, we write $\mathbf{R}(\mathbf{Y}_n, \mathbf{y}_{n-1})$ instead of $\mathbf{R}(\mathbf{Y}_n)$. Then the step-parallel version of (3.2) is defined by

$$
\begin{aligned}
&\textbf{For } n = 1 \textbf{ to } N \\
&\quad \mathbf{Y}_n^{(0,r)} = \mathbf{p} \otimes \mathbf{y}_{n-1}^{(1,r)} + (P \otimes I)\mathbf{Y}_{n-1}^{(1,r)} \\
&\quad \textbf{For } j = 1 \textbf{ to } m \\
&\qquad \mathbf{Y}_n^{(j,0)} = \mathbf{Y}_n^{(j-1,r)} \\
&\qquad \textbf{For } \nu = 1 \textbf{ to } r \\
&\qquad\quad (I - B \otimes hJ_n)(\mathbf{Y}_n^{(j,\nu)} - \mathbf{Y}_n^{(j,\nu-1)}) = \\
&\qquad\qquad -(I - A \otimes hJ_n)(\mathbf{Y}_n^{(j,\nu-1)} - \mathbf{Y}_n^{(j-1,r)}) \\
&\qquad\qquad -\mathbf{R}(\mathbf{Y}_n^{(j-1,r)}, \mathbf{y}_{n-1}^{(j,r)}), \\
&\qquad\quad \textbf{either} \quad \mathbf{y}_n^{(j,\nu)} = \mathbf{y}_{n-1}^{(j,r)} + h(\mathbf{b}^T \otimes I)\mathbf{F}(\mathbf{Y}_n^{(j,\nu)}) \\
&\qquad\quad \textbf{or} \qquad \mathbf{y}_n^{(j,\nu)} = \mathbf{y}_{n-1}^{(j,r)} + (\mathbf{b}^T A^{-1} \otimes I)(\mathbf{Y}_n^{(j,\nu)} - \mathbf{e} \otimes \mathbf{y}_{n-1}^{(j,r)}),
\end{aligned}
\tag{4.1}
$$

where N denotes the number of integration steps. In the case of one inner iteration ($r = 1$), this scheme has been analysed in [28, 29, 53]. The coupling between the iterates $\mathbf{Y}_n^{(j,r)}$ and $\mathbf{y}_n^{(j,r)}$ in (4.1) allows us to start the iteration process at the point t_n already after just one outer iteration at t_{n-1}, that is, as soon as $\mathbf{Y}_{n-1}^{(1,r)}$ is computed, we can compute $\mathbf{Y}_{n-1}^{(2,r)}$ and $\mathbf{Y}_n^{(1,r)}$ concurrently. In this way, the outer iterates $\mathbf{Y}_1^{(j,r)}, \mathbf{Y}_2^{(j-1,r)}, \mathbf{Y}_3^{(j-2,r)}, \ldots$ can all be computed concurrently. In fact, we may write:

$$
\begin{aligned}
&\textbf{For } j = 1 \textbf{ to } m \\
&\quad \textbf{For } i = 1 \textbf{ to } \min\{j, N\} \\
&\qquad \textbf{If } i = j \textbf{ then } \mathbf{Y}_i^{(0,r)} = \mathbf{p} \otimes \mathbf{y}_{i-1}^{(1,r)} + (P \otimes I)\mathbf{Y}_{i-1}^{(1,r)} \\
&\qquad \mathbf{Y}_i^{(j-i+1,0)} = \mathbf{Y}_i^{(j-i,r)} \\
&\qquad \textbf{For } \nu = 1 \textbf{ to } r \\
&\qquad (I - B \otimes hJ_n)(\mathbf{Y}_i^{(j-i+1,\nu)} - \mathbf{Y}_i^{(j-i+1,\nu-1)}) = \\
&\qquad\quad -(I - A \otimes hJ_i)(\mathbf{Y}_i^{(j-i+1,\nu-1)} - \mathbf{Y}_i^{(j-i,r)}) \\
&\qquad\quad -\mathbf{R}(\mathbf{Y}_i^{(j-i,r)}, \mathbf{y}_{i-1}^{(j-i+1,r)}),
\end{aligned}
\tag{4.2}
$$

where we assumed $m \geq N$ and where we omitted the step-point formula. Hence, effectively, only $N + m - 1$ outer iterates have to be computed, instead of Nm outer iterates as required by (3.2). The sequential (or effective) number of outer iterations per step becomes $m_{\text{seq}} = (m + N - 1)N^{-1} \approx mN^{-1} + 1$. However, the step-parallel approach requires that the predictor formula needed to start the iteration at t_n is based on a sufficiently "safe" iterate $\mathbf{Y}_{n-1}^{(1,r)}$. This requirement implies that we should perform sufficiently many inner iterations in the first outer iteration. The condition $m \geq N$ imposed on (4.2) implies that

we need N processors for a parallel implementation. In practice, the number of steps may be much larger than the number of processors available. This can be accounted for by dividing the integration interval in subintervals (windows) and by applying the integration process successively on these subintervals.

For $r = 1$, a convergence analysis of (4.2) and related versions can be found in [28, 29, 53]. Here, we shall consider the case $r \geq 1$. For simplicity, we only consider step-point formulas of the form $\mathbf{y}_n^{(j,r)} = (\mathbf{e}_s^T \otimes I)\mathbf{Y}_n^{(j,r)}$. An elementary derivation reveals that for the usual test equation the iteration error $\varepsilon_\nu^{(j,n)} := \mathbf{Y}_n^{(j,\nu)} - \mathbf{Y}_n$ satisfies the relation

$$\varepsilon_n^{(j,\nu)} = M\varepsilon_n^{(j,\nu-1)} + (I - B \otimes hJ_n)^{-1}(\mathbf{e}\mathbf{e}_s^T \otimes I)\varepsilon_{n-1}^{(j,r)}, \tag{4.3}$$

where M is defined in (3.4). This leads to

$$\begin{aligned} &\varepsilon_n^{(j,r)} - L_r\varepsilon_{n-1}^{(j,r)} = M^r\varepsilon_n^{(j-1,r)}, \\ &L_r := (I - M^r)(I - A \otimes hJ_n)^{-1}(\mathbf{e}\mathbf{e}_s^T \otimes I), \end{aligned} \tag{4.4}$$

or equivalently,

$$\begin{aligned} &\varepsilon^{(j+1)} = M_{N,r}\varepsilon^{(j)},\ \varepsilon^{(j)} := \begin{pmatrix} \varepsilon_1^{(j,r)} \\ \varepsilon_2^{(j,r)} \\ \cdots \\ \varepsilon_N^{(j,r)} \end{pmatrix}, \\ &M_{N,r} := \begin{pmatrix} M^r & O & O & O & \cdots \\ L_rM^r & M^r & O & O & \cdots \\ L_r{}^2M^r & L_rM^r & M^r & O & \cdots \\ L_r{}^3M^r & L_r{}^2M^r & L_rM^r & M^r & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}. \end{aligned} \tag{4.5}$$

Evidently, we have convergence whenever the spectrum $\sigma(M)$ of M is within the unit circle. In Section 3 we already saw that this is precisely the convergence condition for the inner iteration process. However, since $M_{N,r}$ is defective for $N > 1$, the inner amplification factors may be large for small r. Proceeding as in Section 3, we define the *inner* amplification factors (cf. (3.8))

$$\rho_N^{(r)} = \max_{\mathrm{Re}(z)\leq 0} \rho_N^{(r)}(z), \quad \rho_N^{(r)}(z) := \sqrt[r]{||Z_{N,r}(z)||_2}, \tag{4.6}$$

where

$$Z_{N,r}(z) := \begin{pmatrix} Z^r & O & O & O & \cdots \\ K_rZ^r & Z^r & O & O & \cdots \\ K_r{}^2Z^r & K_rZ^r & Z^r & O & \cdots \\ K_r{}^3Z^r & K_r{}^2Z^r & K_rZ^r & Z^r & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{pmatrix}, \tag{4.7}$$

and

B	N	$s=4$	r^*	$s=8$	r^*
PDIRK	2	[3.7, 2.9, 1.8, ..., 0.52]	6	[21, 22, 18, ..., 0.90]	>40
	3	[4.1, 3.4, 2.0, ..., 0.52]	6	[23, 47, 48, ..., 0.90]	>40
	4	[4.7, 4.0, 2.3, ..., 0.52]	6	[25, 107, 128, ..., 0.90]	>40
PTIRK	2	[0.8, 0.7, 0.6, ..., 0.50]	1	[1.2, 1.0, 1.0, ..., 0.86]	3
	3	[0.9, 0.8, 0.7, ..., 0.50]	1	[1.3, 1.1, 1.0, ..., 0.86]	4
	4	[1.0, 0.8, 0.7, ..., 0.50]	2	[1.5, 1.2, 1.1, ..., 0.86]	5
(3.14), $\hat{A}_{ij}=O\gamma=1$	2	[3.6, 1.3, 0.9, ..., 0.44]	3	[136, 4.4, 2.2, ..., 0.64]	9
	3	[5.1, 1.4, 1.0, ..., 0.44]	3	[1352, 6.7, 2.8, ..., 0.64]	9
	4	[6.6, 1.5, 1.0, ..., 0.44]	4	[13432, 9.7, 3.5, ..., 0.64]	10

TABLE 4.1. Inner amplification factors $[\rho_N^{(1)}, \rho_N^{(2)}, \rho_N^{(3)}, \ldots, \rho_N^{(\infty)}]$ for Radau IIA correctors.

$$K_r := (I - Z^r)\,(I - zA)^{-1}\mathbf{e}\mathbf{e}_s^T.$$

The analogue of Table 3.1 where the inner amplification factors for $N = 1$ are listed, is given by Table 4.1 where $N = 2, 3, 4$ (note that $\rho_1^{(r)} = \rho^{(r)}$). This table shows the same trends as Table 3.1, but much more pronounced.

5. WAVEFORM RELAXATION

The derivation of waveform relaxation (WR) methods starts with representing the IVP (1.1) in the form

$$\frac{d\mathbf{y}}{dt} = \boldsymbol{\phi}(\mathbf{y}, \mathbf{y}), \quad \mathbf{y}(t_0) = \mathbf{y}_0, \quad \mathbf{y}, \boldsymbol{\phi} \in \mathbb{R}^d, \tag{5.1}$$

where $\boldsymbol{\phi}(\mathbf{u}, \mathbf{v})$ is a *splitting function* satisfying $\boldsymbol{\phi}(\mathbf{y}, \mathbf{y}) = \mathbf{f}(\mathbf{y})$. This splitting function is chosen such that the Jacobian matrix $J^* = \partial\boldsymbol{\phi}/\partial\mathbf{u}$ has a simple structure, so that, given an approximation $\mathbf{y}^{(k-1)}$ to the solution $\mathbf{y}$ of (5.1), a next approximation $\mathbf{y}^{(k)}$ is more easily solved from the system

$$\frac{d\mathbf{y}^{(k)}}{dt} = \boldsymbol{\phi}(\mathbf{y}^{(k)}, \mathbf{y}^{(k-1)}), \quad \mathbf{y}^{(k)}, \mathbf{y}^{(k-1)}, \boldsymbol{\phi} \in \mathbb{R}^d \tag{5.2}$$

than $\mathbf{y}$ is solved from (5.1). Here, $k = 1, 2, \ldots, q$, and $\mathbf{y}^{(0)}$ denotes an initial approximation to the solution of (5.1). The iteration process (5.2) is called *continuous WR iteration* with WR iterates $\mathbf{y}^{(k)}$. This approach was introduced by LELARASMEE [38] and LELARASMEE, RUEHLI and SANGIOVANNI-VINCENTELLI [39] in 1982 and since then has intensively been analysed and applied to IVPs (see e.g. [37]). WR iteration has a lot of potential parallelism. For example, a popular choice for the splitting function $\boldsymbol{\phi}$ is such that the matrix J^* is σ-by-σ block-diagonal (block-Jacobi WR method). Then, each iteration of the WR method (5.2) requires the integration of s uncoupled IVP systems which can be done in parallel on s processors. For a detailed survey of the potential

for parallelism of WR methods, we refer to the recent book of BURRAGE [12]. Here, we present a brief description of a WR approach based on RK methods and its relation with the step-parallel methods of the preceding section.

Let us integrate the IVP for (5.2) numerically by the RK method (1.2). Then, we obtain the scheme

$$\begin{aligned} &\textbf{For } k = 1 \textbf{ to } q \\ &\textbf{For } n = 1 \textbf{ to } N \\ &\quad \mathbf{Y}_n^{(k)} = \mathbf{e} \otimes \mathbf{y}_{n-1}^{(k)} + h(A \otimes I)\Phi(\mathbf{Y}_n^{(k)}, \mathbf{Y}_n^{(k-1)}). \end{aligned} \tag{5.3}$$

Here, $\mathbf{y}_n^{(k)}$, $\mathbf{Y}_n^{(k)}$, and $\mathbf{\Phi}$ are the analogues of $\mathbf{y}_{n-1}$, $\mathbf{Y}_n$ and $\mathbf{F}$ occurring in (1.2). As soon as $\mathbf{Y}_n^{(k)}$ is computed, the step-point value $\mathbf{y}_n^{(k)}$ can be obtained by one of the following two formulas (cf. (2.3))

$$\begin{aligned} \mathbf{y}_n^{(k)} &= \mathbf{y}_{n-1}^{(k)} + h(\mathbf{b}^T \otimes I)\mathbf{F}(\mathbf{Y}_n^{(k)}), \\ \mathbf{y}_n^{(k)} &= \mathbf{y}_{n-1}^{(k)} + (\mathbf{b}^T A^{-1} \otimes I)(\mathbf{Y}_n^{(k)} - \mathbf{e} \otimes \mathbf{y}_{n-1}^{(k)}). \end{aligned} \tag{5.4}$$

The scheme {(5.3),(5.4)} is called the *discrete WR iteration process* with (discrete) WR iterates $\mathbf{Y}_n^{(k)}$ and $\mathbf{y}_n^{(k)}$. Its stability and convergence properties have recently been investigated by BELLEN, JACKIEWICZ and ZENNARO [4, 5] and by IN 'T HOUT [21].

Observe that (5.3) has a substantial amount of parallelism, irrespective the structure of the splitting function $\mathbf{\Phi}$. It has a similar type of step-parallelism as (4.2), because for given k, all iterates $\mathbf{Y}_1^{(k)}, \mathbf{Y}_2^{(k-1)}, \mathbf{Y}_3^{(k-2)}, \ldots$ can be computed in parallel (see also [5]). Hence, effectively, (5.3) does not require the computation of qN iterates, but only of $N + q - 1$ iterates.

Finally, we remark that the nonlinear system for $\mathbf{Y}_n^{(k)}$ in (5.3) is of the same type as the system (1.2), so that it can be solved by modified Newton using the iterative linear system solver as described in Section 3. First results are published in [26]. Extensions to general implicit differential equations (including DAEs) are subject of current research [32].

REFERENCES

1. ALEXANDER, R. (1977). Diagonally implicit Runge-Kutta methods for stiff ODEs, *SIAM J. Numer. Anal.* **14**, 1006–1021.
2. AUGUSTYN, R. & ÜBERHUBER, C.W. (1992). *Parallel defect correction algorithms for ordinary differential equations*, ACPC/TR 92–22, Austrian Centre for Parallel Computation, Vienna, Austria.
3. BELLEN, A. (1987). Parallelism across the steps for difference and differential equations, *Numerical methods for ordinary differential equations*. A. BELLEN, C.W. GEAR & E. RUSSO (eds.), *Proceedings L'Aquila 1987*, Lecture Notes in Mathematics **1386**, Springer-Verlag, Berlin, 22–35.

4. BELLEN, A., JACKIEWICZ, Z. & ZENNARO, M. (1993). Time-point relaxation Runge-Kutta methods for ordinary differential equations, *J. Comput. Appl. Math.* **45**, 121–137.
5. BELLEN, A., JACKIEWICZ, Z. & ZENNARO, M. (1994). Contractivity of waveform relaxation Runge-Kutta iterations and related limit methods for dissipative systems in the maximum norm, *SIAM J. Numer. Anal.* **41**,499–523.
6. BELLEN, A., VERMIGLIO, R. & M. ZENNARO (1990). Parallel ODE-solvers with step size control, *J. Comput. Appl. Math.* **31**, 277–293.
7. BURRAGE, K. (1978). A special family of Runge-Kutta methods for solving stiff differential equations, *BIT* **18**, 22–41.
8. BURRAGE, K. (1991). The error behaviour of a general class of predictor-corrector methods, *Appl. Numer. Math.* **8**, 201–216.
9. BURRAGE, K. (1993). Efficient block predictor-corrector methods with a small number of corrections, *J. Comput. Appl. Math.* **45**, 139–150.
10. BURRAGE, K. (1993) Parallel methods for initial value problems, *Appl. Numer. Math.* **11**, 5–25.
11. BURRAGE, K. (1993). The search for the Holy Grail, or: Predictor-corrector methods for solving ODEIVPs, *Appl. Numer. Math.* **11**, 125–141.
12. BURRAGE, K. (1995). *Parallel and sequential methods for ordinary differential equations*, Clarendon Press, Oxford.
13. BUTCHER, J.C. (1976). On the implementation of implicit Runge-Kutta methods, *BIT* **16**, 237–240.
14. BUTCHER, J.C. (1979). A transformed implicit Runge-Kutta method, *J. Assoc. Comput. Mach.* **26**, 731–738.
15. CHARTIER, P. (1993). *Parallelism in the numerical solution of initial value problems for ODEs and DAEs*, Ph.D. Thesis, Université de Rennes I, France.
16. CROUZEIX, M. (1975). *Sur l'approximation des équations différentielles opérationnelles linéaires par des méthodes de Runge-Kutta*, Ph.D. Thesis, Université de Paris.
17. DORSSELAER, J.L.M. VAN (1994). *Theoretical aspects of numerical methods for initial value problems*, Ph.D. Thesis, University of Leiden, The Netherlands.
18. GEAR, C.W. & XU XUHAI (1993). Parallelism across time in ODEs, *Appl. Numer. Math.* **11**, 45–68.
19. HAIRER, E. & WANNER, G. (1991). *Solving ordinary differential equations, II. Stiff and differential-algebraic problems*, Springer-Verlag, Berlin.
20. HOFFMANN, W. & SWART, J.J.B. DE (1997). *Approximating Runge-Kutta matrices by triangular matrices*, *BIT Numer. Math.* **37**, 346–354.
21. HOUT, K.J. IN 'T (1995). On the convergence of waveform relaxation methods for stiff nonlinear ordinary differential equations, *Appl. Numer. Math.* **18**, 175–190.
22. HOUWEN, P.J. VAN DER & SOMMEIJER, B.P. (1990). Parallel iteration of high-order Runge-Kutta methods with step size control, *J. Comput. Appl.*

Math. **29**, 111–127.

23. HOUWEN, P.J. VAN DER & SOMMEIJER, B.P. (1990). Parallel ODE solvers, *Proceedings of the International Conference on Supercomputing*, ACM Press, 71–81.
24. HOUWEN, P.J. VAN DER & SOMMEIJER, B.P. (1991). Iterated Runge-Kutta methods on parallel computers, *SIAM J. Sci. Statist. Comput.* **12**, 1000–1028.
25. HOUWEN, P.J. VAN DER & SOMMEIJER, B.P (1993). Analysis of parallel diagonally implicit iteration of Runge-Kutta methods, *Appl. Numer. Math.* **11**, 169–188.
26. HOUWEN, P.J. VAN DER & SOMMEIJER, B.P. (1995). Iteration of Runge-Kutta methods with block triangular Jacobians, *Z. Angew. Math. Mech.* **76**, 367–375.
27. HOUWEN, P.J. VAN DER, SOMMEIJER, B.P. & COUZY, W. (1992). Embedded diagonally implicit Runge-Kutta algorithms on parallel computers, *Math. Comp.* **58**, 135–159.
28. HOUWEN, P.J. VAN DER, SOMMEIJER, B.P. & VEEN, W.A. VAN DER (1995). Parallel iteration across the steps of high-order Runge-Kutta methods for nonstiff initial value problems, *J. Comput. Appl. Math.* **60**, 309–329.
29. HOUWEN, P.J. VAN DER, SOMMEIJER, B.P. & VEEN, W.A. VAN DER (1994). Parallelism across the steps in iterated Runge-Kutta methods for stiff initial value problems, *Numerical Algorithms* **8**, 293–312.
30. HOUWEN, P.J. VAN DER & SWART, J.J.B. DE (1996). Triangularly implicit iteration methods for ODE-IVP solvers, *SIAM J. Sci. Comput.*, **18**, 41–55.
31. HOUWEN, P.J. VAN DER & SWART, J.J.B. DE (1997). Parallel linear system solvers for Runge-Kutta methods, *Advances in Computational Mathematics* **7**, 157–181.
32. HOUWEN, P.J. VAN DER & VEEN, W.A. VAN DER (1997). Waveform relaxation methods for implicit differential equations, *Advances in Computational Mathematics* **7**, 183–197.
33. ISERLES, A. & NØRSETT, S.P. (1990). On the theory of parallel Runge-Kutta methods, *IMA J. Numer. Anal.* **10**, 463–488.
34. JACKSON, K.R., KVÆRNØ, A. & NØRSETT, S.P. (1994). The use of Butcher series in the analysis of Newton-like iterations in Runge-Kutta formulas, *Appl. Numer. Math.* **15**, 341–356.
35. JACKSON, K.R. & NØRSETT, S.P. (1995) The potential for parallelism in Runge-Kutta methods, Part 1: RK formulas in standard form, *SIAM J. Numer. Anal.* **32**, 49–82.
36. KALVENES, J. (1989). *Experiments with parallel ODE solvers*, Master's thesis, University of Trondheim.
37. LEIMKUHLER, B. & RUEHLI, A.E. (1993). Rapid convergence of waveform relaxation, *Appl. Numer. Math.* **11**, 211–224.
38. LELARASMEE, E. (1982). *The waveform relaxation method for the time domain analysis of large scale nonlinear dynamical systems*, Ph.D. Thesis, Univ. of California, Berkeley.

39. LELARASMEE, E., RUEHLI, A.E. & SANGIOVANNI-VINCENTELLI, A.L. (1982), The waveform relaxation method for time domain analysis of large scale integrated circuits, *IEEE Trans. CAD IC Syst.* **1**, 131–145.
40. LIE, I. (1987). *Some aspects of parallel Runge-Kutta methods*, Report No. 3/87, Division Numerical Mathematics, University of Trondheim.
41. LIOEN, W.M. (1996). On the diagonal approximation of full matrices, *J. Comput. Appl. Math.* **75**, 35–42.
42. MIRANKER, W.L. & LINIGER, W. (1967). Parallel methods for the numerical integration of ordinary differential equations, *Math. Comp.* **21**, 303–320.
43. NEVANLINNA, O. (1985). Matrix valued versions of a result of von Neumann with an application to time discretization, *J. Comput. Appl. Math.* **12** & **13**, 475–489.
44. NØRSETT, S.P. (1974). *Semi-explicit Runge-Kutta methods*, Report Mathematics and Computation No.6/74, Depart. of Mathematics, University of Trondheim.
45. NØRSETT, S.P. (1975) *Numerical solution of ordinary differential equations*, Ph.D. Thesis, University of Dundee.
46. NØRSETT, S.P. (1976). Runge-Kutta methods with a multiple real eigenvalue only, *BIT* **16**, 388–393.
47. NØRSETT, S.P. & SIMONSEN, H.H. (1989). Aspects of parallel Runge-Kutta methods, A. BELLEN, C.W. GEAR & E. RUSSO (eds.), *Numerical methods for ordinary differential equations, Proceedings L'Aquila 1987*, Lecture Notes in Mathematics **1386**, Springer-Verlag, Berlin, 103–117.
48. OREL, B. (1993). Parallel Runge-Kutta methods with real eigenvalues, *Appl. Numer. Math.* **11**, 241–250.
49. SHAMPINE, L.F. (1980). Implementation of implicit formulas for the solution of ODEs, *SIAM J. Sci. Statist. Comput.* **1**, 103–118.
50. SIMONSEN, H.H. (1990). *Extrapolation methods for ODE's: continuous approximations, a parallel approach*, Ph.D. Thesis, University of Trondheim.
51. SOMMEIJER, B.P. (1993). *Parallelism in the numerical integration of initial value problems*, CWI Tract 99, CWI, Amsterdam.
52. SPIJKER, M.N. (1995). The effect of the stopping of the Newton iteration in implicit linear multistep methods, Appl. Numer. Math. **18**, 367–386.
53. VEEN, W.A. VAN DER, HOUWEN, P.J. VAN DER & SWART, J.J.B. DE (1995). Convergence aspects of step-parallel iteration of Runge-Kutta methods, *Appl. Numer. Math.* **18**, 397–411.
54. WOLFBRANDT, A. (1977). *A study of Rosenbrock processes with respect to order conditions and stiff stability*, Ph.D. Thesis, Chalmers University of Technology, Göteborg.