Parallel Two-Step W-Methods with Peer Variables

Bernhard A. Schmitt, Rüdiger Weiner

Abstract

A new class of methods for the solution of stiff initial value problems is introduced that is parallel by design. It has a two-step character and propagates s different 'peer' solution variables with essentially identical characteristics from step to step. The main work lies in the solution of s independent linear stage equations which may be solved in parallel. Convergence of order $s-1$ and stability for general stepsize sequences are proved. Conditions for order s and stronger stability criteria are addressed, as well. Promising methods up to order 7 are identified by numerical tests with some widely used stiff test problems. Some of these are competitive with existing software even in sequential computations.

Key words Stiff initial value problems, parallel two-step methods, parallel peer W-methods.

AMS subject classifications 65L06, 65Y05

1 Introduction

Parallel solution methods in large and stiff initial value problems

$$y' = f(t,y), \quad y(t_0) = y_0 \in \mathbb{R}^n \tag{1}$$

$t_0 \le t \le t_e$, may be based on many different strategies. In this paper we will consider a new class of time integration schemes with inherent 'method parallelism', [8]. This feature is independent of 'parallelism across the system' since both kinds of parallelism may be applied simultaneously in many situations. Approaches for method parallelism may start with well-known classical methods and try to parallelize expensive parts of its numerical components, like the solution of the large nonlinear systems in Implicit Runge-Kutta methods (e.g., [1]). Similarly, parallel iteration schemes may be applied to classical one-step or multistep methods as in some work of van der Houwen, et.al., [11, 12]. and Burrage, et.al., [3]. For a recent textbook on parallel methods see [2]. Other approaches are based on new kinds of methods that are parallel by design. Within the wide class of General Linear Methods (GLMs) the subclass of DIMSIM methods suitable for parallel implementation has been proposed by Butcher [5]. Here, the equations for the internal stages of the GLM may be solved in parallel since the corresponding part of the coefficient matrix is diagonal. The paper [7] is concerned with implementation issues

of such methods. Unfortunately, the practical experience with these methods is not overly positive compared to RADAU [10] or PSIDE [12].

In order to circumvent some of the theoretical bottlenecks in the structure of classical one-step and multistep methods the class of linearly implicit *Parallel Two-Step W-methods* (PTSW-methods) has been discussed recently by H. Podhaisky and the authors [14]. These PTSW-methods are similar to ordinary ROW or W-methods but use s stage increments from the previous time step only. So, all s current stages can be processed parallel. The PTSW methods are already quite competitive [15, 18] even in sequential computations. They are particularly attractive in very stiff or singularly perturbed problems since they do not suffer an order reduction due to high stage orders. One of their weaknesses, however, is some critical dependence on stepsize ratios in the stiff case.

Literally speaking, many parallel and classical time integration methods have a set of distinguished or 'master' variables and compute additional 'slave' variables to improve accuracy or stability properties of the masters. In fact, these methods usually employ only one n-dimensional master approximation for the solution in each time interval with distinguished accuracy and stability properties. In contrast to these schemes we now consider methods having only peer variables sharing the same accuracy and stability properties (with minor modifications). An immediate advantage of this approach is the existence of a continuous extension for these methods. For the time being we also restrict the discussion to two-step methods that compute several solution approximations $Y_{mi},\ i = 1,\dots,s$, associated with a time interval $[t_m, t_{m+1}]$ from the information contained in the variables $Y_{m-1,i}$ from the previous interval. Generalization to methods using even earlier information is obvious but will not be considered here. Moreover we will concentrate on linearly implicit methods avoiding the solution of nonlinear systems of equations. In the following class of Parallel 'Peer' two-Step W-Methods (PPSW methods), the solutions $Y_{mi},\ i = 1,\dots,s$, are related to points $t_{mi} := t_m + h_m c_i$ associated with the time interval $[t_m, t_{m+1}]$ but not necessarily contained in it. The methods are given by

$$(I - \gamma_i h_m T_m) Y_{mi} = \sum_{j=1}^{s} (b_{ij} + h_m \gamma_{ij} T_m) Y_{m-1,j} + h_m \sum_{j=1}^{s} a_{ij} f(t_{m-1,j}, Y_{m-1,j}), \tag{2}$$

$i = 1,\dots,s$. The terms $\gamma_i > 0$, $b_{ij}, \gamma_{ij}, a_{ij}$ are the parameters of the method. The matrix T_m should be an approximation of the Jacobian $f_y(t_m, y(t_m))$ for stability reasons only. In fact, the accuracy of these methods is derived for arbitrary T_m in the sense of W-methods. In this context it is no essential restriction to consider autonomous problems and we will do this for simplicity. A subclass of these methods with $\gamma_i = 0$, $\gamma_{ij} = 0$, may also be attractive for nonstiff problems but will not be discussed here.

Introducing the stage vectors $Y_m = (Y_{mi})_{i=1}^{s} \in \mathbb{R}^{sn}$ and coefficient matrices $G := \mathrm{diag}(\gamma_i)$, $B = (b_{ij})$, $\Gamma = (\gamma_{ij})$ a compacter version of the PPSW method is

$$(I - h_m G \otimes T_m) Y_m = (B \otimes I + h_m \Gamma \otimes T_m) Y_{m-1} + h_m (A \otimes I) f(Y_{m-1}). \tag{3}$$

Here it is easily seen that for methods using $G = \gamma I$ the matrix $I - h_m G \otimes T_m$ commutes with all other matrices in the scheme. Hence, the stability analysis simplifies considerably and we will concentrate on this kind of methods in the present paper. Similar to the notion in Implicit Runge-Kutta methods we will also call methods using $\gamma_i \equiv \gamma$ singly-implicit. Such methods may be attractive already in a sequential computer environment since the matrix decomposition of $I - \gamma h_m T_m$ may be used in all stages. A situation of interest on parallel machines is the use of expensive parallel preconditioners for this matrix. Still, multi-implicit methods with a general diagonal matrix G offer additional design options and will be discussed in [17].

The emphasis of our discussion lies on higher-order methods with an order near the number of stages s. Here, a critical source of problems is the fact that the coefficients of the methods depend on the current stepsize ratio

$$\sigma_m := h_m / h_{m-1}. \tag{4}$$

When discussing nonlocal effects we will occasionally add an additional step-index m to the coefficient matrices A, B, Γ to indicate this dependence. Stepsize increases will be restricted by an upper bound $\bar{\sigma} \geq \sigma_m$ for all stepsize ratios, but no lower bound is assumed since this might be a severe restriction in practice. Convergence results are formulated in terms of the maximal stepsize $H := \max_{j \geq 0} h_j$. Further abreviations are $\beta := A + \Gamma$, $\mathbb{1} = (1, \ldots, 1)^{\mathsf{T}}$, e_i for the ith unit vector. The spectral radius of a matrix is denoted by ρ.

After this introduction the paper continues with basic aspects on stability and accuracy of the schemes. Conditions for stability with general stepsizes and the structure of the stability matrix are derived as well as basic order conditions. The conditions for order $s - 1$ lead to explicit representations of the coefficients of the scheme. We concentrate on these high order schemes in Section 3 and identify one class where the general stability conditions from Section 2 may be established easily. We also discuss conditions for improving the order to s. For the singly-implicit methods discussed in this paper it is unlikely that this may be achieved by improving all local errors by one order. But we show that there is some superconvergence effect for the global error for certain parameter choices. Unfortunately, this effect may be conveniently exploited for constant stepsizes only. In Section 4 we finally discuss implementation issues and present numerical results with different test problems and methods with up to 8 stages.

2 Basic properties

2.1 Stability issues

Due to the two-step structure of the scheme and the σ-dependence of its coefficients the stability analysis encounters many of the difficulties of multistep methods. In addition to the step recursion (3) we consider a second one with additional perturbations $h_m g_m = h_m (g_{mi})_{i=1}^s \in \mathbb{R}^{sn}$

and solutions $Y_m + X_m$. So, for the error X_m we have the recursion

$$\begin{aligned}(I - h_m G_m \otimes T_m)X_m &= (B_m \otimes I + h_m \Gamma_m \otimes T_m)X_{m-1} \\ &\quad + h_m (A_m \otimes I)\Big(f(Y_{m-1} + X_{m-1}) - f(Y_{m-1})\Big) + h_m g_m,\end{aligned} \tag{5}$$

where the coefficient matrices are supplemented by the step index m. The stability of the recursion (5) is covered by the theory for multistep methods, see [9]. For ease of reference we formulate it in the following lemma. The crucial assumption (6) will be verified for our methods in section 3.1.

Lemma 1 *Assume, that for some fixed $\bar{\sigma} > 1$ and stepsize sequences with $h_m \le \bar{\sigma} h_{m-1}$ a uniform bound $\bar{b}$ exists for all products*

$$\|B_{m+k} \cdots B_{m+1} B_m\| \le \bar{b}, \quad 0 \le m \le m + k, \tag{6}$$

where $\sum_{j=m}^{m+k} h_j \le t_e - t_0$. Let the maps $X \mapsto (I - h_m G_m \otimes T_m)^{-1}((G_m \beta_m + \Gamma_m) \otimes T_m)X + (A_m \otimes I)(f(Y_{m-1} + X) - f(Y_{m-1})))$ be uniformly Lipschitz continuous in some neighbourhood of zero. Then there exits a constant C such that

$$\|X_m\| \le C(\|X_0\| + \max_{j=1}^{m} \|g_j\|) \quad \textit{for} \quad t_0 \le t_m \le t_e.$$

The standard application of this stability lemma concerns the convergence of the scheme, where the 'perturbed' solution $Y_m + X_m = y(t_m)$ is the solution of the initial value problem and the perturbations g_m are the corresponding residuals defined in (8) below. Here, Lemma 1 shows that the scheme (2) converges with order p, i.e., $X_m = O(H^p)$, if the local error is of order p, too, i.e., $g_m = O(H^p)$. So we do not need to distinguish between the local and global order of the method except in the case of superconvergence, cf. Section 3.2.

In the stiff context it is appropriate to consider the linear autonomous problem $y' = Jy$ as a test equation, where the eigenvalues of J lie in the left complex halfplane. With the choice $T_m = J$ the scheme (2) reduces to the recursion

$$Y_m = M_m(h_m J)Y_{m-1}, \quad M(z) := (I - zG_m)^{-1}(B_m + z\beta_m), \tag{7}$$

where $M_m(\cdot)$ is the stability matrix of the scalar test equation $y' = \lambda y$ and $z = h\lambda$. Unfortunately, M_m depends on the current stepsize ratio σ_m and only for constant stepsizes the spectral radius of M is an appropriate measure for the longtime behaviour in the recursion (7). In the case $\sigma \equiv 1$ we may apply some standard stability notions to the scalar function $\rho(M(z))$.

Definition 2 *Let $M(.)$ be the stability matrix of a PPSW method (2) as defined in (7) with $\sigma = 1$. Then, the method is called* zero-stable, *if $\rho(M(0)) = 1$ and eigenvalues on the unit circle are simple. It is A-stable if $\rho(M(z)) < 1$ in the open left complex halfplane, $z \in \mathbb{C}_- := \{z \in \mathbb{C} : Re z < 0\}$ and L-stable if, additionally, $\rho(M(\infty)) = 0$.*

Remarks a) Any consistent method should reproduce constant solutions for $\lambda = 0$. Since this requirement leads to the identity $B\mathbb{1} = \mathbb{1}$ the spectral radius of $M(0) = B$ cannot be smaller than one.

b) For singly-implicit methods we have $M(\infty) = -\frac{1}{\gamma}\beta$ and L-stability requires $\rho(\beta) = 0$.

c) It is convenient to introduce the variable $w = z/(1-\gamma z)$ satisfying $(1+\gamma w)(1-\gamma z) = 1$. For $z \in \mathbb{C}_-$ the variable w is contained in the circle centered at $-1/(2\gamma)$ and going through $w = 0$. Now, the stability matrix of singly-implicit methods is a linear function of w given by

$$M(z) = (1+\gamma w)B + w\beta = B + w(\gamma B + \beta).$$

This form may be conveniently employed to check the A-stability of the scheme since M has an eigenvalue $\lambda = e^{i\theta}$ on the unit circle if w is an eigenvalue of the generalized eigenvalue problem, [16],

$$(e^{i\theta} I - B)x = w(\gamma B + \beta)x, \quad x \in \mathbb{C}^s.$$

A thorough and general analysis of stability properties of our methods is quite difficult since the stability matrix M depends on the actual stepsize ratio σ_m for accuracy reasons. But this is a common problem for multistep methods where even proving zero-stability is nontrivial for general stepsize sequences. For PPSW methods this case will be dealt with in §3.1 by verifying (6). A similar result covering the stiff limit $z = h\lambda \to \infty$ for arbitrary stepsize ratios is presented there, too. But stronger, z-uniform stability results for these methods with nonconstant stepsize sequences are not yet available.

2.2 Accuracy conditions

The structure of the coefficient matrices is determined to a large extent by accuracy requirements. The accuracy of the PPSW methods may be analyzed in a standard way by considering the residuals Δ_{mi} obtained when the exact solution is put into the method. For the methods (2) we can use the information $f(t_{m-1,j}, y(t_{m-1,j})) = y'(t_{m-1,j})$ and consider

$$h_m\Delta_{mi} := (I - \gamma h_m T_m)y(t_{m,i}) - \sum_{j=1}^{s}(b_{ij} + h_m\gamma_{ij}T_m)y(t_{m-1,j}) - h_m\sum_{j=1}^{s} a_{ij}y'(t_{m-1,j}), \qquad (8)$$

$i = 1, \ldots, s$. We will use Taylor expansion at t_m but write the error in terms of h_{m-1} since σ is bounded from above only, so $h_m = O(h_{m-1})$ but not vice versa. Separating the terms depending on T_m from the others we see that

$$h_m T_m(\gamma y(t_{mi}) + \sum_{j=1}^{s}\gamma_{ij}y(t_{m-1,j})) = O(h_{m-1}^{q+1})$$

holds if condition

$$\Gamma(q): \quad \gamma c_i^k + \sum_{j=1}^{s}\gamma_{ij}(c_j - 1)^k\sigma^{-k} = 0, \quad k = 0, \ldots, q-1 \qquad (9)$$

with $\sigma = \sigma_m$ is satisfied for a sufficiently smooth solution. Secondly, it holds that

$$y(t_{mi}) - \sum_{j=1}^{s} b_{ij} y(t_{m-1,j}) - h_m \sum_{j=1}^{s} a_{ij} y'(t_{m-1,j}) = O(h_{m-1}^q)$$

if condition

$$AB(q): \quad c_i^k - \sum_{j=1}^{s} b_{ij}(c_j - 1)^k \sigma^{-k} - k \sum_{j=1}^{s} a_{ij}(c_j - 1)^{k-1} \sigma^{1-k} = 0, \quad k = 0, \ldots, q-1, \qquad (10)$$

is true with $\sigma = \sigma_m$ and the tacit convention that the second sum is not evaluated for $k = 0$ and canceled by the factor k. In both conditions q denotes the number of equations.

Lemma 3 *If the conditions $\Gamma(q)$ and $AB(q+1)$ are satisfied with $q \geq 1$ and the solution y of (1) is sufficiently smooth then the residuals (8) for the scheme (2) are of order q, i.e., $\|\Delta_m\| = O(h_{m-1}^q)$.*

With respect to the order, the most critical condition seems to be $\Gamma(q)$ in (9) since it depends on the $s^2 + 1$ coefficients γ, γ_{ij} only. In fact, requiring $\Gamma(s)$ leads to the explicit relation

$$\Gamma = -\gamma\Theta, \quad \Theta := V S_m P V^{-1}, \qquad (11)$$

with the Vandermonde matrix $V = (c_i^{j-1})_{i,j=1}^s$, the Pascal matrix $P = \left(\binom{j-1}{i-1}\right)_{i,j=1}^s$ and the diagonal matrix $S_m = \mathrm{diag}(1, \sigma_m, \ldots, \sigma_m^{s-1})$. The matrix Θ will be encountered quite often since it describes the polynomial extrapolation from the subgrid $\{t_{m-1} + h_{m-1} c_i,\ i = 1, \ldots, s\}$ to $\{t_m + h_m c_i,\ i = 1, \ldots, s\}$.

In a similar way, explicit solutions for the order conditions (10) may be obtained by requiring $AB(s)$. In this case, (10) has the matrix form $0 = V - BVP^{-1}S_m^{-1} - AVP^{-1}S_m^{-1}F_0^{\mathsf{T}}\,\mathrm{diag}(i-1)$. The matrix $F_0 = (\delta_{i-1,j})$ describes the shift to the right. This equation may be solved for the matrix B yielding

$$B = \Theta - \sigma_m A V\,\mathrm{diag}(i) F_0^{\mathsf{T}} V^{-1}. \qquad (12)$$

Here, the identity $PF_0^{\mathsf{T}}\,\mathrm{diag}(i-1) = \mathrm{diag}(i)F_0^{\mathsf{T}}P$ was used. Since stability properties depend on the matrix $\beta = A + \Gamma$ by (7) it is convenient to replace the coefficient A in the last equation by using β and the first order condition (9), as well.

Lemma 4 *If the PPSW method (2) satisfies $AB(s)$ and $\Gamma(s-1)$, its coefficient matrices B and β are related through*

$$B = (I + \gamma E)\Theta + \sigma_m \beta E, \quad E = -V\,diag(i) F_0^{\mathsf{T}} V^{-1}. \qquad (13)$$

The matrices E and Θ commute for $\sigma_m = 1$, since $E\Theta = \sigma_m \Theta E$.

Proof. In matrix form the condition $\Gamma(s-1)$ amounts to $\gamma V S_m P + \Gamma V = u e_s^{\mathsf{T}}$ with an arbitrary last column u. Replacing $A = \beta - \Gamma$ in (12) gives

$$B = \Theta - \sigma_m(\beta V + \gamma V S_m P)\,\mathrm{diag}(i) F_0^{\mathsf{T}} V^{-1},$$

since the last column of ΓV drops out after multiplication with F_0^{T}. Using $P F_0^{\mathsf{T}}\,\mathrm{diag}(i-1) = \mathrm{diag}(i) F_0^{\mathsf{T}} P$ again, the factorizations $\sigma_m \Theta E = \sigma_m V S_m P\,\mathrm{diag}(i) F_0^{\mathsf{T}} V^{-1} = V\,\mathrm{diag}(i) F_0^{\mathsf{T}} S_m P V^{-1} = E\Theta$ are possible. □

3 High order methods

3.1 Convergence for variable stepsize

In the case of one single parameter value γ considered here all coefficient matrices have a simpler structure in the monomial basis (c^{j-1}) contained in the Vandermonde matrix V. The transformed version of the matrix B of (13) is given by

$$\tilde{B} := V^{-1} B V = S_m P - \gamma\,\mathrm{diag}(i) F_0^{\mathsf{T}} S_m P - \sigma_m \tilde{\beta}\,\mathrm{diag}(i) F_0^{\mathsf{T}},\ \ \tilde{\beta} := V^{-1}\beta V. \tag{14}$$

Note, that its first column is the first unit vector. According to (7) the stability matrix at infinity is $M(\infty) = -\frac{1}{\gamma}\beta$ and optimal stiff damping could be obtained by the choice $\beta = 0$. But (14) shows, that this choice violates the condition of zero stability for $\sigma \geq 1$, since the leading matrix $S_m P - \gamma\,\mathrm{diag}(i) F_0^{\mathsf{T}} S_m P$ is upper triangular and has the eigenvalues σ^{i-1}. Hence, $\tilde{\beta}$ is needed for the stabilization of B. The following theorem provides a convenient compromise between stability for stiff problems and zero-stability.

Theorem 5 *Let the method (2) satisfy the conditions $AB(s)$ and $\Gamma(s-1)$ and use one single $\gamma = \gamma_i$, $i = 1, \ldots, s$. Then, with the choice*

$$V^{-1}\beta V = F_0 S_m\, diag(\frac{1}{i}) \tag{15}$$

the method is both zero-stable and $L(\alpha)$-stable, if it is $A(\alpha)$-stable. In fact, both matrices $B - \mathbb{1} e_1^{\mathsf{T}}$ and β are nilpotent.

Proof. The only non-zero elements of $\tilde{\beta}$ are $\tilde{\beta}_{i+1,i} = \frac{1}{i}\sigma_m^{i-1}$, $i = 1, \ldots, s-1$. So, clearly, $\tilde{\beta}$ is nilpotent. This choice leads to $\sigma_m \tilde{\beta}\,\mathrm{diag}(i) F_0^{\mathsf{T}} = S_m F_0 F_0^{\mathsf{T}} = S_m\,\mathrm{diag}(0, 1, \ldots, 1)$ in (14) and cancels the main diagonal of $\tilde{B}$ beyond the first entry. In fact, we have

$$\tilde{B} := V^{-1} B V = e_1 e_1^{\mathsf{T}} + S_m(P - I) - \gamma\,\mathrm{diag}(i) F_0^{\mathsf{T}} S_m P, \tag{16}$$

and $\tilde{B} - e_1 u^{\mathsf{T}}$ is nilpotent for any vector u with $u^{\mathsf{T}} e_1 = 1$. So, $V(\tilde{B} - e_1 u^{\mathsf{T}})V^{-1} = B - \mathbb{1} e_1^{\mathsf{T}}$ is nilpotent with $u = e_1^{\mathsf{T}} V = (1, c_1, \ldots)$. □

By Theorem 5 the transformed coefficient matrix $\tilde{B}$ is fully specified. Due to the upper triangular structure, for any s, this matrix really is part of the large matrix

$$\begin{pmatrix} 1 & 1-\gamma\sigma & 1-2\gamma\sigma & 1-3\gamma\sigma & \cdots \\ & 0 & 2\sigma(1-\gamma\sigma) & 3\sigma(1-2\gamma\sigma) & \cdots \\ & & 0 & 3\sigma^2(1-\gamma\sigma) & \cdots \\ & & & \ddots & \end{pmatrix}.$$

Although σ (and even γ) may change between intervals this representation shows that the nilpotency discussed in Theorem 5 is structurally stable. It really holds for general stepsize sequences since the Vandermonde transformation used in (16) is the same everywhere. Since we again have to track the situation over different time steps we indicate the dependence on σ_m by the additional matrix index m.

Lemma 6 *Let the matrices B_m, $m \geq 1$, be given by (14) and the matrices $\beta = \beta_m$ as defined in Theorem 5. Then each product of at least $s-1$ matrices β_j resp. B_j has rank one, at most. In fact, it holds that*

$$\begin{aligned} \beta_m\beta_{m-1}\cdots\beta_{m-k} &= 0, \quad m,k \geq s-1, \\ B_mB_{m-1}\cdots B_{m-k} &= \mathbb{1}v^{\mathsf{T}}, \quad m,k \geq s-2, \end{aligned} \tag{17}$$

with $v^{\mathsf{T}}\mathbb{1} = 1$.

Proof. Without restriction we may consider the matrices $\tilde{\beta}_j, \tilde{B}_j$ since they are all transformed by the same matrix V. The result for the matrices $\tilde{\beta}_j$ is trivial due to their common triangular structure. Equation (16) shows that each $\tilde{B}_j$ has the structure

$$\tilde{B}_j = \begin{pmatrix} 1 & v_j^{\mathsf{T}} \\ 0 & \tilde{B}_{22}^{(j)} \end{pmatrix},$$

where $\tilde{B}_{22}^{(j)}$ is *strictly* upper triangular. Hence, $\tilde{B}_{22}^{(m)}\tilde{B}_{22}^{(m-1)}\cdots\tilde{B}_{22}^{(m-k)} = 0$ for $k \geq s-2$. □

Remark: We note that the row vector v^{T} in (17) depends on the stepsize ratios of the $s-1$ right-most matrices in the product.

But, since stepsize changes are bounded above by $\bar{\sigma}$ the vector v in (17) is bounded by some fixed power of σ (and γ). This verifies the main assumption of the stability Lemma 1 and we may summarize most of the previous results in the following theorem.

Theorem 7 *Let a grid (t_m) be given with bounded stepsize changes $\sigma_m \leq \bar{\sigma}$ and let $0 \leq \gamma \leq \bar{\gamma}$. Let the method (2) satisfy $AB(s)$ and $\Gamma(s-1)$ and let β be chosen by (15). If the initial values are accurate, i.e., $\|Y_{0i} - y(t_{0i})\| = O(h_0^{s-1})$, $i = 1,\ldots,s$, then the PPSW method converges with order $s-1$, i.e.,*

$$\|Y_{mi} - y(t_{mi})\| = O(H^{s-1}), \quad t_0 \leq t_m \leq t_e, \ i = 1,\ldots,s.$$

Proof. For $X_{mi} = y(t_{mi}) - Y_{mi}$ the recursion (5) holds with $g_m = \Delta_m$ defined in (8) which is of order $O(h_{m-1}^{s-1})$ by assumption and Lemma 3. Since the main assumption (6) of the stability Lemma 1 has been verified in Lemma 6 with the present assumptions the assertion follows. □

The representations (13) and (16) show, that the extrapolation matrix Θ is an important contribution in both coefficients B and Γ. Applying a standard reformulation of W-methods to the scheme (2) avoiding unnecessary multiplications with T_m reveals a different interpretation. For high-order methods satisfying $\Gamma(s)$ this version reads

$$(I - \gamma h_m I \otimes T_m)(Y_m - (\Theta \otimes I)Y_{m-1}) = ((B - \Theta) \otimes I)Y_{m-1} + h_m(A \otimes I)f(Y_{m-1}), \qquad (18)$$

$A = \gamma\Theta + \beta$. It may be interpreted as a corrector equation for the predictor $(\Theta \otimes I)Y_{m-1}$ obtained by polynomial extrapolation. The implementation of PPSW methods will be based on this formulation in Section 4.

3.2 Superconvergence of PPSW methods

So far, we have considered methods of order $s - 1$ only. It would be of interest, of course, to achieve even higher orders in PPSW methods. In the light of Lemma 3 order s requires that the conditions $\Gamma(s)$ and $AB(s + 1)$ hold. While the first condition is already satisfied by $\Gamma = -\gamma\Theta$ defined in (11) the second one, $AB(s + 1)$, leads to s additional nonlinear restrictions. Before formulating this result we point out that the representation (16) for the coefficient matrix B may be written as the rule

$$\sum_{j=1}^{s} b_{ij}\psi(c_j) = \psi(1 + \sigma_m c_i) - \psi(\sigma_m c_i) + \psi(0) - \gamma\sigma_m\psi'(1 + \sigma_m c_i), \ i = 1, \ldots, s, \qquad (19)$$

for any polynomial ψ of degree $s - 1$.

Lemma 8 *Consider the PPSW method with $\Gamma(s)$ and the coefficients defined in Theorem 5, let $\phi(t) := \prod_{i=1}^{s}(t - c_i)$ be the knot polynomial of the scheme and let the solution y be sufficiently smooth. Then, for each i, $1 \le i \le s$, the condition*

$$\gamma\sigma_m\phi'(1 + \sigma_m c_i) = \phi(1 + \sigma_m c_i) - \phi(\sigma_m c_i) + (\sigma_m c_i)^s + \phi(0) \qquad (20)$$

implies $\Delta_{mi} = O(h_{m-1}^s)$.

Proof. With $\Gamma(s)$ and (16) the residual (8) has the form

$$h_m\Delta_{mi} = \frac{h_{m-1}^s}{s!}\Big((1 + \sigma c_i)^s - \sum_{j=1}^{s} b_{ij}c_j^s - \sigma s\sum_{j=1}^{s} a_{ij}c_j^{s-1}\Big)\, y^{(s)}(t_{m-1}) + O(h_{m-1}^{s+1}), \qquad (21)$$

where Taylor expansion at t_{m-1} was used for convenience and $\sigma = \sigma_m$. For the terms $a_{ij}c_j^{s-1} = (\beta_{ij} - \gamma_{ij})c_j^{s-1}$ the condition $\Gamma(s)$ gives $\sum_j \gamma_{ij}c_j^{s-1} = -\gamma(1 + \sigma c_i)^{s-1}$. But since the last column

of $\tilde{\beta}$ in Theorem 5 is zero the contribution $\sum_{j=1}^{s} \beta_{ij} c_j^{s-1} = 0$ is missing. So, the leading bracket in (21) has the form

$$(1+\sigma c_i)^s - \sigma s\gamma (1+\sigma c_i)^{s-1} - \sum_{j=1}^{s} b_{ij} c_j^s =: u_i.$$

We may get rid of the highest power $c_j^s = c_j^s - \phi(c_j) = -\psi(c_j)$ in the sum by considering

$$u_i = (1+\sigma c_i)^s - \sigma s\gamma (1+\sigma c_i)^{s-1} + \sum_{j=1}^{s} b_{ij} \psi(c_j), \tag{22}$$

with the polynomial $\psi(t) := \phi(t) - t^s$ of degree $s-1$. Here, condition (19) applies and we obtain

$$\begin{aligned} u_i &= (1+\sigma c_i)^s - \sigma s\gamma (1+\sigma c_i)^{s-1} + \psi(1+\sigma c_i) - \psi(\sigma c_i) + \psi(0) - \sigma\gamma\psi'(1+\sigma c_i) \\ &= \phi(1+\sigma c_i) + (\sigma c_i)^s - \phi(\sigma c_i) + \phi(0) - \sigma\gamma\phi'(1+\sigma c_i). \end{aligned}$$

So, $u_i = 0$ implies $\Delta_{mi} = O(h_{m-1}^s)$ in (21). □

Since the recursion (5) combines all previous errors the gain of one order from $s-1$ to s in the global error requires that condition (20) holds for all stages $i = 1, \ldots, s$. But these s conditions are very severe restrictions for the remaining $s+1$ parameters $\gamma, c_1, \ldots, c_s$ and are unlikely to be satisfied for some interesting range of σ-values. So, we will not pursue this line of research further. However, later on, we will discuss an application of using one single superconsistency condition (20) in error control. And for multi-implicit methods the corresponding conditions (20) may easily satisfied by using different γ_i, see [17].

However, for PPSW-1 methods there is a different global effect leading to order-s convergence in the error. Unfortunately, this super-convergence may be conveniently exploited for constant stepsizes only.

By a careful choice of parameters the structural information (17) may be used to eliminate the leading term in the global error. In fact, adding the error residuals Δ_m from (8) the exact solution $y_m = (y_{mi})$ obeys the modified time step recursion

$$y_m = I \otimes (I - \gamma h_m T_m)^{-1} \Big((B_m \otimes I + h_m \Gamma_m \otimes T_m) y_{m-1} + h_m (A \otimes I) f(y_{m-1}) + h_m \Delta_m \Big)$$

and the error $X_m = (Y_{mi} - y(t_{m,i}))_i$ satisfies equation (5) that we now write as

$$X_m = (B_m \otimes I + h_m \mathbf{D_m}) X_{m-1} - h_m (I \otimes (I - \gamma h_m T_m)^{-1}) \Delta_m. \tag{23}$$

The matrix $\mathbf{D_m} = I \otimes (I - \gamma h_m T_m)^{-1} ((\Gamma_m + \gamma B_m) \otimes T_m + (A_m \otimes I)\mathbf{J_m})$ is obtained by using the representation

$$f(Y_{mi}) - f(y_{mi}) = \mathbf{J_{mi}} (Y_{mi} - y_{mi}), \quad \mathbf{J_{mi}} = \int_0^1 f'(y_{mi} + t(Y_{mi} - y_{mi}))\, dt$$

with the block diagonal matrix $\mathbf{J_m} = \mathrm{diag}(\mathbf{J_{mi}})$.

The structural result in Lemma 6 shows that the following computations may still be performed for nonconstant stepsizes. However, the dependence on the different stepsize ratios becomes overly complicated for higher-order methods and may be of no practical use. So we will restrict this discussion to the case $\sigma_m \equiv \sigma = 1$. Then, all matrices B_m are identical and we drop the index again. And the vector v in (17) is a fixed vector, namely the left eigenvector of B to the eigenvalue one, and will be given explicitly below. As a first step we simplify the recursion (23) by assuming $O(\|\mathbf{D_m} X_{m-1}\|) = O(\|X_{m-1}\|)$. Since our main argument is independent of the dimension n we write down only the scalar case $n = 1$, for simplicity. We obtain

$$\begin{aligned} X_m &= BX_{m-1} - h\Delta_m + O(h\|X_{m-1}\| + h^2\|\Delta_m\|) \\ &= -h\sum_{j=0}^{m-1} B^j \Delta_{m-j} + B^m X_0 + \sum_{j=1}^{m-1} O(h\|X_{m-j}\| + h^2\|\Delta_{m-j+1}\|) \\ &= -h\mathbb{1}\sum_{j=s-1}^{m-1} v^{\mathsf{T}}\Delta_{m-j} - h\sum_{j=0}^{s-2} B^j\Delta_{m-j} + B^m X_0 + \sum_{j=1}^{m} O(h\|X_{m-j}\| + h^2\|\Delta_{m-j+1}\|). \end{aligned}$$

Now, if the method has order $s-1$ and the initial error X_0 is of appropriate size, all terms in the last line are of order h^s except the first one. But, by an appropriate choice of the vector v in $B^\infty = \mathbb{1}v^{\mathsf{T}}$ the leading error term in $v^{\mathsf{T}}\Delta_{m-j}$ may be eliminated, as well.

Theorem 9 *Let the method satisfy $AB(s)$, $\Gamma(s)$ and use the matrix β from Theorem 5. Then, for constant stepsizes the vector v^{T} in (17) depends only on γ and the knots c_i. In fact, $v^{\mathsf{T}}V$ depends on γ only and is given by the first s components of the vector*

$$\tilde{v}^{\mathsf{T}} = (1, 1-\gamma, 3-6\gamma+2\gamma^2, 13-39\gamma+30\gamma^2-6\gamma^3, \ldots).$$

If the solution y of (1) is smooth enough, then

$$v^{\mathsf{T}}\Delta_m = \tilde{v}_{s+1}\frac{h^{s-1}}{s!}\, y^{(s)}(t_{m-1}) + O(h^s).$$

Proof. Since $\tilde{B}$ is upper triangular and explicitly given by (16) the component $\tilde{v}_i$ depends on the first i columns of $\tilde{B}$ only and may be computed by computer algebra. Introducing the s-vector $\varphi = (\phi_0, \ldots, \phi_{s-1})^{\mathsf{T}}$ of coefficients of the knot polynomial $\phi(t) = (t-c_1)\cdots(t-c_s) = \sum_{k=0}^{s}\phi_k t^k$ defined above we may write $0 = (\phi(c_i))_{i=1}^s = V\varphi + (c_i^s)_i$. So, a different representation of the local residual (21) is obtained by writing (22) in vector form

$$\begin{aligned} u &= \Big((1+c_i)^s - \phi(c_i) - s\gamma(1+c_i)^{s-1} + \sum_{j=1}^{s} b_{ij}(\phi(c_j) - c_j^s)\Big)_{i=1}^{s} \\ &= \Big(\sum_{k=1}^{s}\binom{s}{k-1}c_i^{k-1}\Big)_i - V\varphi - s\gamma VPe_s + BV\varphi \\ &= V\Big((\tilde{B}-I)\varphi + (\mathbf{P}e_{s+1})_1^s - \gamma PDe_s\Big) \\ &= V\Big((\tilde{\mathbf{B}}-\mathbf{I})\begin{pmatrix}\varphi\\1\end{pmatrix}\Big)_1^s. \end{aligned} \tag{24}$$

Here, the bold-faced symbols denote the $(s+1)\times(s+1)$ versions of matrices given by explicit rules but only their first s rows are used as indicated, of course. Now, the first thing we see from (24) is that $v^{\mathsf{T}}u$ is independent of φ. The vector $v^{\mathsf{T}}V$ contains the first s components of the left eigenvector of $\tilde{\mathbf{B}}$, and so, φ is canceled. But, since $\tilde{\mathbf{B}}-\mathbf{I}$ is upper triangular with diagonal elements -1 below the first row, we get $v^{\mathsf{T}}u = -\tilde{v}_{s+1}e_{s+1}^{\mathsf{T}}(\tilde{\mathbf{B}}-\mathbf{I})e_{s+1} = \tilde{v}_{s+1}$. □

Remark The theorem shows, that the leading error constant of the scheme is a multiple of the component $\tilde{v}_{s+1}$ which is a polynomial in the parameter γ. So there are a few exceptional values for γ where the method has global order s for any set of off-step knots c_i. In Table 1 we present explicit formulas for $\tilde{v}_{s+1}(\gamma)$ and some of the parameters $\hat{\gamma}$ with superconvergence that were used in specific PPSW methods. The fourth column contains a numerical estimate for the angle of $A(\alpha)$ stability in degrees and the last column the name of the corresponding method used in the numerical tests in Section 4.

s	$\tilde{v}_{s+1}$	$\hat{\gamma}:\ \tilde{v}_{s+1}(\hat{\gamma})=0$	α	Method
2	$3-6\gamma+2\gamma^2$	$\frac{1}{2}(3\pm\sqrt{3})$	90°	
3	$13-39\gamma+30\gamma^2-6\gamma^3$	1.32088	90°	
4	$75-300\gamma+372\gamma^2-168\gamma^3+24\gamma^4$	0.468147	52.4°	4a
		0.912763	89.9°	4b
5	$541-2705\gamma+4660\gamma^2-3420\gamma^3+1080\gamma^4-120\gamma^5$	0.722499	79.6°	5b
6	$4683-28098\gamma+62130\gamma^2-64200\gamma^3+32760\gamma^4$	0.619411	57.5°	6b
	$-7920\gamma^5+720\gamma^6$	1.087080	87.9°	6c
7	$47293-331051\gamma+894810\gamma^2-201410\gamma^3$	0.557132	23.9°	7b
	$+864360\gamma^4-335160\gamma^5+65520\gamma^6-5040\gamma^7$	0.885444	81.3°	7c
8	$545835-4366680\gamma+13959176\gamma^2-23146032\gamma^3$			
	$+21724080\gamma^4-11847360\gamma^5+3689280\gamma^6$			
	$-604800\gamma^7+40320\gamma^8$	0.758671	67.2°	8c

Table 1: Superconvergence conditions and methods

The superconvergence property can easily be observed in numerical computations with constant stepsizes. For variable stepsizes a similar effect might still be obtained by clever strategies for a step-dependent choice of the parameter γ. However, for higher order methods such strategies are probably too complicated to be of practical use. Still, we expect that superconvergence for $\sigma=1$ has some beneficial effect on the general performance especially for sharp tolerances where stepsize ratios are clustered around one.

4 Implementation and numerical tests

For practical implementations the version (18) is preferable. Since the coefficients of the method depend on the actual stepsize ratio they have to be recomputed before the equations (18) may be solved. But since the expense for these computations is $O(s^2)$ only it is negligible compared to the solution of the stage equations for large dimensions n. We also note that in case of a step rejection the function evaluations $f(Y_{m-1})$ may be re-used.

Estimates for the local error are required for stepsize selection procedures. As mentioned before, (18) may be interpreted as a corrector equation and its solution $K_m := Y_m - (\Theta \otimes I)Y_{m-1}$ is an obvious candidate for an error estimate. Yet, it is of the same order $s-1$ as the error in Y_m that it should estimate. However, this situation can be improved by requiring order s of consistency (20) for one single stage only, $i = s$, for instance. Then, K_{ms} becomes an asymptotically correct order-s estimate for the local error of $Y_{m,s}$. Unfortunately, only for low-order methods ($s \leq 4$) it is possible to find methods satisfying this additional condition and minimal stability requirements, i.e., A(0)-stability for $\sigma = 1$. So, only Method 4a uses this error estimate with the superconvergence value $\gamma = 0.468147$ and knots c_i yielding (20) for $i = s$ and $\sigma = 1$. In order to obtain a sound error estimate in higher order methods, too, we compare Y_m with a predictor of order $s-2$ by ignoring the oldest approximate solution and using only $Y_{m-1,i}$, $i = 2, ..., s$.

Presently the stage equations (18) are solved by LU decomposition. Since this decomposition is the most expensive part and a severe sequential bottleneck for large systems a remarkable parallel speed-up with PPSW methods is not likely to be achieved yet. From our experience with PTSW methods we expect superior parallel performance for large problems only in combination with Krylov solvers for the s independent linear stage systems (see [15, 16]). So we decided to present tests with sequential computations only but we compare with the state-of-the-art linearly implicit one-step code RODAS [10]. We think that this already gives a good insight into the properties and the potential of these methods and assists in choosing suitable ones. We used Delphi 5 for comfortable programming. Since the reference results for the original RODAS code are in Fortran we compare the number of steps instead of the computing time. Note that the computational amount of work per step in our methods is comparable to that of RODAS in sequential runs. An advantage of our W-type methods over RODAS is that the Jacobian could be kept constant for several steps but we did not exploit this possibility here.

The following stiff test problems from [10] were used in our numerical tests: OREGO with $t_e = 360$, ROBER with $t_e = 10^8$, VDPOL with $t_e = 11$ and PLATE. A problem with strongly varying Jacobian is KREISS from [13] defined by

$$y' = R(t)\Lambda(\varepsilon)R^{-1}(t), \quad y(0) = \begin{pmatrix} 1 \\ 2.6 \end{pmatrix}, \quad 0 \leq t \leq 1,$$

$$R(t) = \begin{pmatrix} \cos(-\theta t) & \sin(-\theta t) \\ -\sin(-\theta t) & \cos(-\theta t) \end{pmatrix}, \quad \Lambda(\varepsilon) = \begin{pmatrix} -\frac{1}{\varepsilon} & 0 \\ 0 & -1 \end{pmatrix}.$$

We used $\varepsilon = 10^{-6}$ and $\theta = 1$ here. A reference solution for these problems was computed with RADAU5 ([10]) and high accuracy.

With the exception of Method 4a the knots are nearly equidistributed in $[-1, 1]$ and $c_s = 1$ was used always. The starting values Y_1 for the PPSW methods were obtained with RODAS. Based on the error estimates described above the new step size was computed in a standard manner, cf. [15]. The maximally allowed increase of the stepsize was set to $\sigma_{max} = 1.5$ for all PPSW methods. Computations were performed for $atol = rtol = 10^{-2}...10^{-9}$ except for ROBER where $atol = 10^{-6} rtol$ was set (cf. [10]). In the following figures we present results for some special methods from Table 1. The logarithm of the final error at the endpoint for $Y_{m,s}$ is shown versus the number of steps. A general observation for most PPSW methods is that both the error and the angle α of $L(\alpha)$-stability increase with the parameter γ. So for all examples except PLATE we present the results for the methods with smallest γ from Table 1 since these produced slightly better results. The problem PLATE, however, requires a stability angle $\alpha \geq 71^\circ$ (cf.[10]). Some of the previously mentioned methods had difficulties here, namely, Method 4a for all tolerances and 6b for some of them. By chosing methods with larger γ from Table 1 having larger stability angles suitable PPSW methods for this problem could be identified, too. Surprisingly, methods 7b and 8c solved this problem without diffculties for all tolerances.

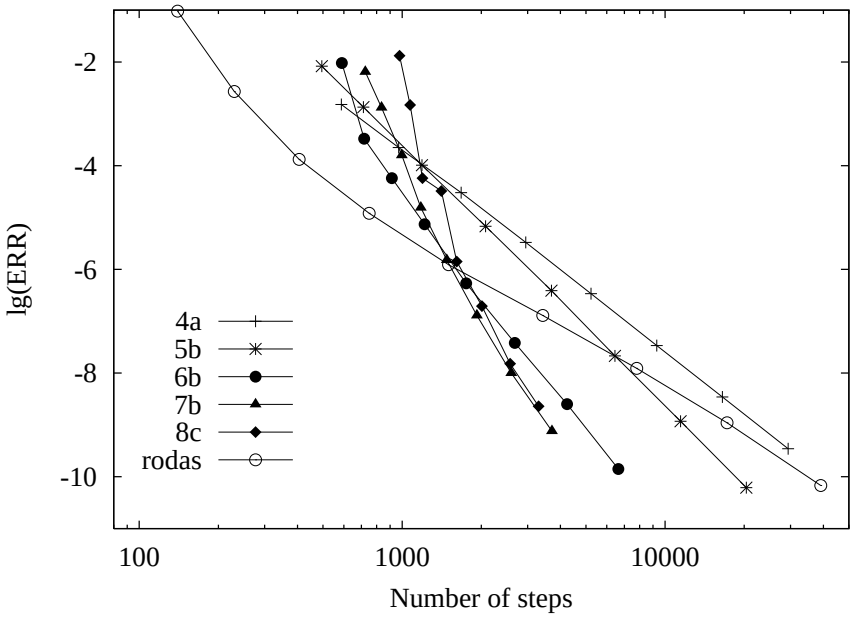

Figure 1: Results for OREGO

From these figures it is seen that the presented methods are interesting and have a large potential for stiff equations. The smooth curves show that the stepsize control works reliable, but it may still be improved. The irregular behaviour of the 7- and 8-stage methods in the problem ROBER is an exception that is probably caused by rounding errors due to large values in the coefficients.

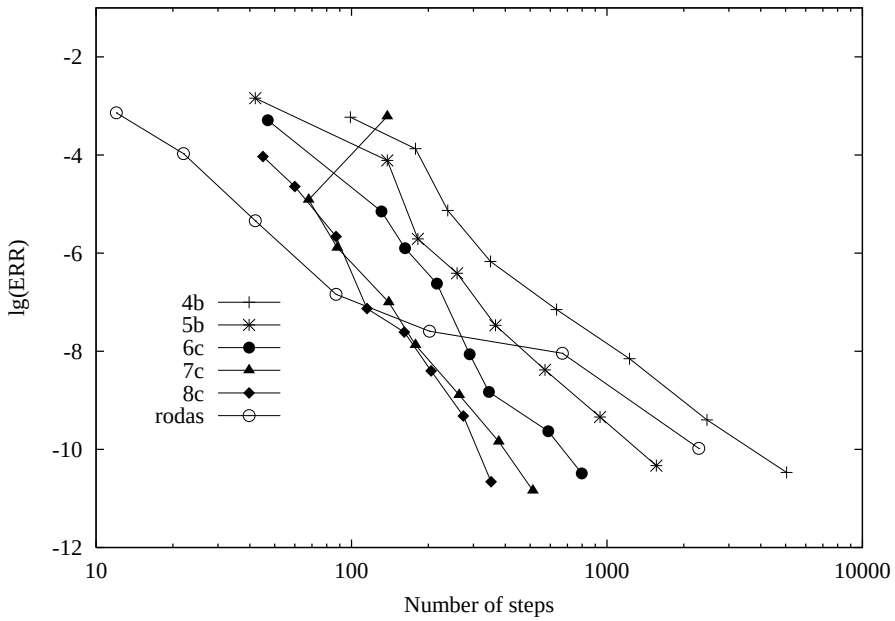

Figure 2: Results for PLATE

This problem has been observed with other methods [4], too, and we tried to ameliorate it by distributing the knots c_i over the interval $[-1, 1]$ instead of $[0, 1]$. In comparison to PPSW methods RODAS is clearly superior for weak tolerances. This is mainly due to the ability of RODAS to increase the stepsize faster than PPSW-methods. Another reason may be that the initial stepsize for the PPSW methods were too small. But for medium tolerances and more stringent ones especially the higher order PPSW methods become comparable and even superior to RODAS. The choice of the knots c_i is still heuristic and their influence on accuracy and robustness of the methods needs further research.

5 Conclusions

A new class of linearly implicit methods for stiff initial value problems has been presented with a nearly optimal potential of method parallelism. Methods of order $s - 1$ with s stages have been derived possessing good stability properties. The global error improves to order s for special parameter values and constant stepsizes since a certain superconvergence property has been identified. In a sequential implementation using automatic stepsize control and LU-decompositions several methods with up to $s = 8$ stages have been tested and were found to be competitive with the code RODAS. The full potential of parallelization for PPSW methods in the solution of large stiff ODEs is only expected in combination with Krylov techniques. Here the use of multi-implicit methods using different parameters γ_i offers additional options and will be investigated in [17].

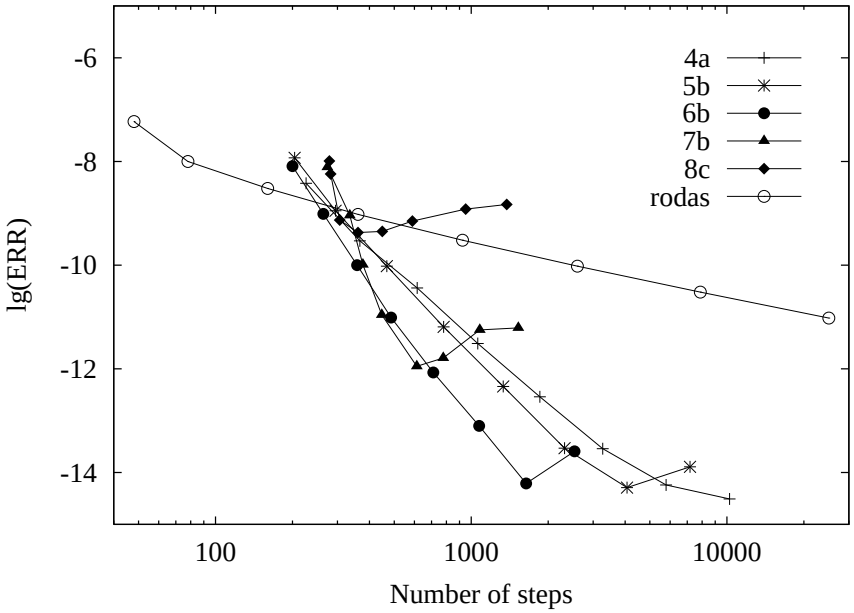

Figure 3: Results for ROBER

References

[1] C. Bendtsen, *A parallel stiff ODE solver based on MIRKs*, Adv. Comput. Math. 7 (1997), 27-36.

[2] K. Burrage, *Parallel and sequential methods for ordinary differential equations*, Clarendon Press, Oxford, 1995.

[3] K. Burrage, and H. Suhartanto, *Parallel iterated methods based on variable step-size multistep Runge-Kutta methods of Radau type for stiff problems*, Adv. Comput. Math. 13 (2000), 257-270.

[4] K. Burrage, and T. Tian, *Parallel half-block methods for initial value problems*, Appl. Numer. Math. 32 (2000), 255-271.

[5] J.C. Butcher, *Diagonally-implicit multi-stage integration methods*, Appl. Numer. Math. 11 (1993), 347-363.

[6] J.C. Butcher and Z. Jackiewicz, *Implementation of diagonally implicit multistage integration methods for ordinary differential equations*, SIAM J. Numer. Anal. **34** (1997), 2119–2141.

[7] J.C. Butcher and A.D. Singh, *The choice of parameters in parallel general linear methods for stiff problems*, Appl. Numer. Math. 34 (2000), 59-84.

[8] C.W. Gear and X. Xuhai: *Parallelism across time in ODEs*, Appl. Numer. Math. 11 (1993), 45-68.

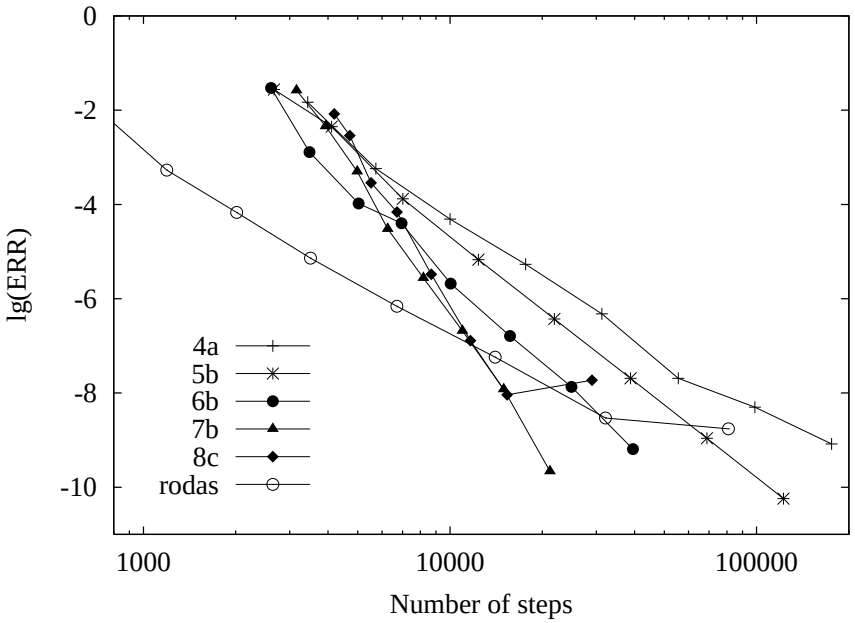

Figure 4: Results for VDPOL

[9] E. Hairer, S.P. Nørsett and G. Wanner, *Solving ordinary differential equations I*, 2.ed., Springer 1993.

[10] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Diffferential–Algebraic Problems*. Springer Verlag, Berlin, Heidelberg, New York 1996.

[11] P.J. van der Houwen and B.P. Sommeijer, *CWI contibutions to the development of parallel Runge-Kutta methods*, Appl. Numer. Math. 22 (1996), 327-344.

[12] P.J. van der Houwen and J.J.B. de Swart, *Parallel Linear System Solvers for Runge-Kutta Methods*, Adv. Comput. Math. 7 (1997), 157–181.

[13] H.O. Kreiss, *Difference methods for stiff ordinary differential equations*, SIAM J. Numer. Anal. 15 (1978), 21–58.

[14] H. Podhaisky, B.A. Schmitt, and R. Weiner, *Two-step W-methods with parallel stages*, Tech. Report 22, Universität Halle, 1999.

[15] H. Podhaisky, B.A. Schmitt, and R. Weiner, *Design, analysis and testing of some parallel two–step W–methods for stiff systems*, Appl. Numer. Math. 42 (2002), 381-395.

[16] H. Podhaisky, *Parallele Zweischritt–W–Methoden*, Ph.D. Thesis, Martin–Luther–Universität, Halle–Wittenberg 2002.

[17] B.A. Schmitt, R. Weiner, and H. Podhaisky, *Multi-implicit Peer Two-Step W-Methods for parallel time integration*, in preparation

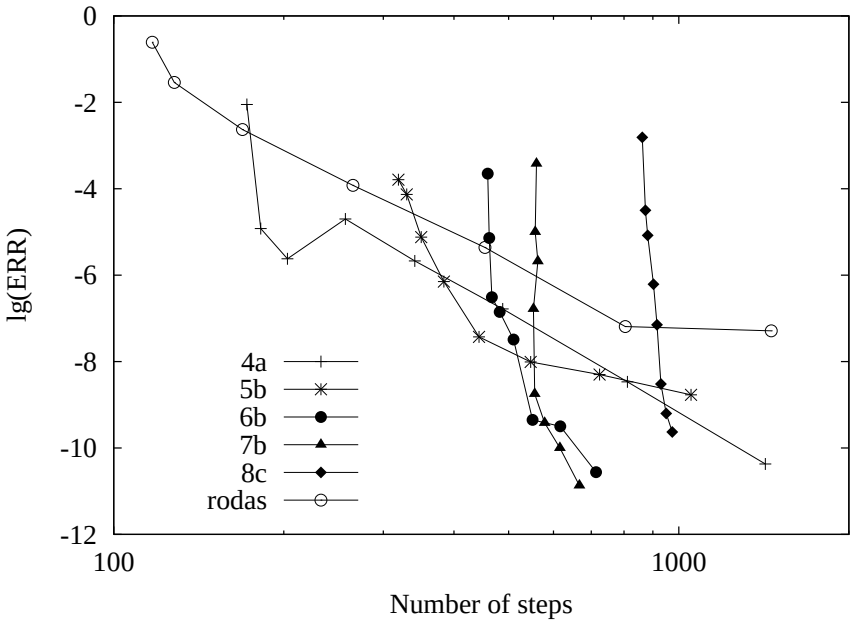

Figure 5: Results for the Kreiss problem

[18] R. Weiner, B.A. Schmitt, and H. Podhaisky, *Two–step W–methods on singular perturbation problems*, to appear in Lecture Notes in Computer Science.