NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Singly diagonally implicit Runge-Kutta methods with an explicit first stage

by

Anne Kværnø

PREPRINT
NUMERICS NO. 1/2004

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL
`http://www.math.ntnu.no/preprint/numerics/2004/N1-2004.ps`
Address: Department of Mathematical Sciences, Norwegian University of Science and
Technology, N-7491 Trondheim, Norway.

Singly diagonally implicit Runge-Kutta methods with an explicit first stage

Anne Kværnø

Department of Mathematical Sciences
The Norwegian University of Science and Technology
N-7491 Trondheim
Norway
E-mail: anne@@math.ntnu.no

January 19, 2004

Dedicated to Syvert P. Nørsett on the occasion of his 60th birthday

Abstract

The purpose of this paper is to construct methods for solving stiff ODEs, in particular singular perturbation problems. We consider embedded pairs of singly diagonally implicit Runge-Kutta methods with an explicit first stage (ESDIRKs). Stiffly accurate pairs of order 3/2, 4/3 and 5/4 are constructed.

AMS Subject Classification: 65L05
Key Words: Stiff ODEs, Singular perturbation problems, Runge-Kutta methods.

1 Introduction

Singly diagonally implicit Runge-Kutta methods (SDIRKs) have been quite popular for solving stiff ordinary differential equations (ODEs) since their introduction in the beginning of the seventies, [3, 7, 16, 18], see also [2, 5, 6]. They can be efficient, and they can preserve the excellent stability properties of implicit Runge-Kutta methods. SIMPLE by Nørsett and Thomsen [19, 20, 21] as well as SDIRK4 by Hairer and Wanner [13] are nice examples of ODE-solvers based on SDIRK methods. In addition, SDIRK-methods are reasonably easy to implement, which makes them attractive as time-integrators for partial differential equations. More recently, there is a renewed interest in SDIRK methods as the implicit part of implicit-explicit (IMEX) RK-methods for solving convection-diffusion-reaction problems [4, 15, 26]. The main restriction of SDIRK methods is their relatively low order. In addition they might suffer from order reduction when applied to stiff ODEs. In particular, the order of B-convergence can in general not exceed 1 [8]. But as we will see, this problem can be avoided for large classes of stiff ODEs.

The aim of this paper is to construct reliable and efficient Runge-Kutta methods with error estimates. In Section 2 a set of design criteria for the method are developed. Singular perturbation problems (SPP) serves as model equations for stiff ODEs here. SDIRK methods with an explicit first stage (ESDIRK) are constructed in Section 3. These methods are of order 3/2, 4/3 and 5/4. The methods have already been available in an unpublished note by the author [17], and used in different contexts. Some relevant references are given in Section 4.

2 Runge-Kutta methods for singular perturbation problems.

As a model equation for stiff ODEs we consider the singular perturbation problem (SPP)

$$y' = f(y,z), \qquad \varepsilon z' = g(y,z), \qquad 0 < \varepsilon << 1, \tag{1}$$

where f and g are smooth functions. We assume $\mu(\partial g/\partial z) < -1$ in some ε-independent neighbourhood of the solution, where μ denotes the logarithmic norm with respect to some inner product. The initial values $y(0), z(0)$ are assumed to admit a smooth solution of (1). The theory of RK-method applied to (1) is developed by Hairer et al. [11], see also [13, Section VI.3].

The exact solution of (1) can be expressed as a power series of ε:

$$\begin{aligned} y(x) &= y_0(x) + \varepsilon y_1(x) + \varepsilon^2 y_2(x) + \cdots \\ z(x) &= z_0(x) + \varepsilon z_1(x) + \varepsilon^2 z_2(x) + \cdots \end{aligned} \tag{2}$$

where $y_l(x), z_l(x)$ are ε-independent functions. The dominating terms $y_0(x), z_0(x)$ is the solution of the index 1 differential algebraic equation (DAE),

$$y_0' = f(y_0, z_0), \qquad 0 = g(y_0, z_0), \tag{3}$$

also called the reduced problem. Further, $y_1(x), z_1(x)$ is the solution of an index 2 DAE and so on. For stepsizes $h \geq Const \cdot \varepsilon$ the numerical solution can be expressed in similar series. The consequence is illustrated in Figure 1, showing plots of the local error and error estimate as functions of h, see Example 1 for details. Similar plots can be found in [12, 13]. The stepsizes can roughly be divided into three intervals:

- $h < Const \cdot \varepsilon$, the index 0 or ODE interval. For such small stepsizes the problem is nonstiff, and no extra caution is needed.

- $h >> \varepsilon$, the index 1 interval. The solutions are dominated by the index 1 solutions. This is the preferred situation, and the methods will be designed to work well here.

- Between these two intervals, the errors are dominated by contributions from the index 2 components $y_1(x), z_1(x)$. Low order is not the worst problem here, but propagation of errors from the last solution point.

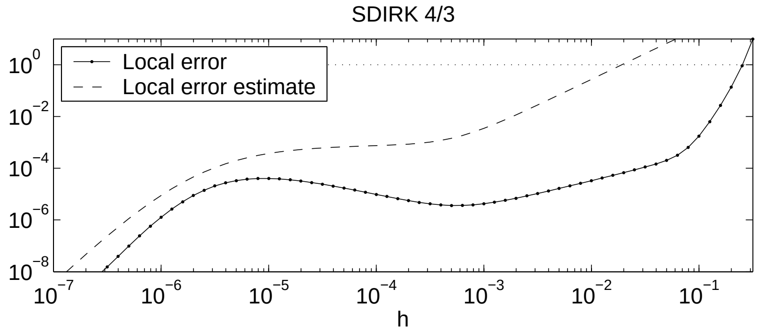

Figure 1: The exact local error (—) and the local error estimate (– –) as functions of h.

> These contributions do not decay with reduced stepsizes. They are almost constant, and might even increase. This situation should if possible be avoided, by making the errors well below the error tolerance.

In the following, we review some theory for Runge-Kutta methods applied to DAEs. This will give the basis for design criteria of our methods.

Runge-Kutta methods. An s-stage RK method is characterised by its Butcher tableau

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & \cdots & a_{2s} \\ \vdots & \vdots & \vdots & & \vdots \\ c_s & a_{s1} & a_{s2} & \cdots & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_s \end{array} \qquad \text{or} \qquad \begin{array}{c|c} c & A \\ \hline & b^T \end{array}.$$

The method is *stiffly accurate* if $b_j = a_{sj},\ j = 1, 2, \cdots, s$. In this paper, the following classes of implicit RK methods are considered:

$\mathbb{M}1$**:** Methods for which the coefficient matrix A is nonsingular.

$\mathbb{M}2$**:** Methods for which

- $a_{1j} = 0,\ j = 1, \cdots, s$,
- the submatrix $\tilde{A} = (a_{ij})_{i,j=2}^{s}$ is nonsingular,
- $b_j = a_{sj},\ j = 1, \cdots, s$.

Order and convergence results for these methods applied to DAEs can be found in [12, 27] ($\mathbb{M}1$) and [14] ($\mathbb{M}2$).

An RK method applied to (1) is given by

$$\begin{aligned}
\dot{Y}_i &= f(Y_i, Z_i), & \varepsilon \dot{Z}_i &= g(Y_i, Z_i), & & \\
Y_i &= y_n + h\sum_{j=1}^{s} a_{ij}\dot{Y}_j & Z_i &= z_n + h\sum_{j=1}^{s} a_{ij}\dot{Z}_j, & & i = 1, \cdots, s \\
y_{n+1} &= y_n + h\sum_{i=1}^{s} b_i\dot{Y}_i & z_{n+1} &= z_n + h\sum_{i=1}^{s} b_i\dot{Z}_i & &
\end{aligned}$$

where Y_i, Z_i are the stage values and $\dot{Y}_i, \dot{Z}_i$ the stage derivatives. For stiffly accurate methods, the new solution point is simply

$$y_{n+1} = Y_s, \quad z_{n+1} = Z_s.$$

The method is of stage order q if the simplifying assumption

$$C(q): \qquad \sum_{j=1}^{s} a_{ij}c_j^{k-1} = \frac{1}{k}c_i^k, \qquad i = 1, 2, \cdots, s, \qquad k = 1, 2, \cdots, q$$

is satisfied. The stability function is given by

$$R(z) = 1 + zb^T(I - zA)^{-1}\mathbb{1}$$

where $\mathbb{1} = (1, 1, \cdots, 1) \in \mathbb{R}^s$, and the stability constant $R(\infty) = \lim_{z\to\infty} R(z)$ is

$$R(\infty) = \begin{cases} 1 - b^T A^{-1}\mathbb{1} & \text{for methods in } \mathbb{M}1 \\ 0 & \text{for stiffly accurate methods in } \mathbb{M}1 \\ -e_{s-1}^T \tilde{A}^{-1} A_1 & \text{for methods in } \mathbb{M}2, \end{cases}$$

where $e_{s-1} = (0, 0, \cdots, 0, 1)^T \in \mathbb{R}^{s-1}$ and $A_1 = (a_{21}, a_{31}, \cdots, a_{s1})^T$. When RK-methods are applied to DAEs, it is well known that $|R(\infty)| \leq 1$ is required for convergence. An A-stable method with $R(\infty) = 0$ is called L-stable.

An estimate for the local error is given by

$$err_y = h\sum_{i=1}^{s}(b_i - \hat{b}_i)\dot{Y}_i, \quad err_z = h\sum_{i=1}^{s}(b_i - \hat{b}_i)\dot{Z}_i, \tag{4}$$

where $\hat{b}^T$ together with A and c forms a method of order $\hat{p} \neq p$, so that

$$err = \mathcal{O}(h^{\min(p,\hat{p})+1}).$$

Error estimators of order $\hat{p} = p - 1$ are common, most SDIRK-methods are equipped with such. But high order estimators are not available for fully implicit methods like Gauss-Kuntzmann-Butcher (order $p = 2s$) and Radau methods (order $p = 2s - 1$) without adding stages. This, together with their relatively high implementation costs, are severe restrictions on fully implicit methods.

The endeavour to get high order error estimators may cause other important properties to be lost. For instance the error estimators used by Hairer and Wanner in their codes SDIRK4 and RADAU5 are not A-stable. In the latter case, $|R(\infty)| = \infty$. To deal with this, the error estimate given by (4) is multiplied by $(I - h\gamma_0 J)^{-1}$ where γ_0 is some constant and J is the jacobian of the equation, see [13, Sec. IV.8].

In the following, all expressions referring to the error estimating method is marked with a "hat", thus $R(z)$ is the stability function of the advancing method, $\hat{R}(z)$ that of the error estimating method, and so on.

Index 1 DAEs. Theory for RK-methods applied to (3) is given by e.g. Deuflhard et al. [9], Griepentrog et al. [10] and Roche [27]. Several methods suffer from a severe order reduction in the z_0-component, while the error of the y_0-component behaves as expected. The order reduction do not occur for stiffly accurate methods, since the use of such methods is equivalent to solve the (nonstiff) problem $y_0' = f(y_0, G(y_0))$, where $z_0 = G(y_0)$ is the solution of $g(y_0, z_0) = 0$. So stiffly accurate methods are particularly suited for solving index 1 DAEs, and thus also stiff SPPs.

Most RK-methods for stiff problems are stiffly accurate. But their error estimators are not, causing the order reduction illustrated in Figure 1.

Index 2 DAEs. The main objective in the index 2 area is to keep the contribution from the index 2 components to the total error well below the error tolerance. To see how, we will quote some results from Hairer et al. [12, Section 8], treating index 2 DAEs solved by methods in the class $\mathbb{M}1$.

Consider the index 2 DAE

$$y' = f(y, z), \qquad 0 = g(y), \qquad g_y f_z \text{ nonsingular.} \tag{5}$$

Exact solutions of this equation have to satisfy the hidden constraint, $(g_y f)(y, z) = 0$. However, numerical solutions will usually not, and this might cause trouble in the next step.

Let y_0, z_0 be the starting values, and suppose they satisfy the assumptions $||((g_y f_z)^{-1} g)(y_0, z_0)|| \leq h\delta$ and $||((g_y f_z)^{-1} g_y f)(y_0, z_0)|| \leq \theta$ with h, δ and θ sufficiently small. Then the local errors are given by

$$\begin{aligned} y_1 - y(x_0 + h) &= R(\infty)(f_z(g_y f_z)^{-1} g)(y_0, z_0) + \mathcal{O}(h^2\delta + h\delta^2) + \mathcal{O}(h^{q_y+1}), \\ z_1 - z(x_0 + h) &= R(\infty)((g_y f_z)^{-1} g_y f)(y_0, z_0) + \mathcal{O}(h\delta + \theta^2) \\ &- \sigma\frac{1}{h}((g_y f_z)^{-1} g)(y_0, z_0) + \mathcal{O}(h\delta + \delta^2) + \mathcal{O}(h^{q_z}). \end{aligned}$$

Here $\sigma = b^T A^{-2} \mathbb{1}$ and q_y, q_z are usually equal to the stage order q. The drift from the constraint $0 = g(y)$ can be restricted by using stiffly accurate methods and by solving algebraic equations with sufficiently accuracy. The constraint $(g_y f)(y, z) = 0$ is in general not satisfied by the numerical solution, but its contributions to the local error can be suppressed by choosing $R(\infty) = 0$. If $R(\infty) \neq 0$ this term might dominate the error, at least for small stepsizes. Results for methods of class $\mathbb{M}2$, see [14], leads to similar conclusions.

The index 2 system arising from the SPP is a special case of (5). Its contributions are multiplied by a factor ε, thus we might expect a significant influence from these terms only for mildly stiff problems.

The following classical example clearly illustrates the theoretical results:

Example 1 Consider Van der Pol equation

$$\begin{aligned} y' &= z & y(0) &= 2, \\ \varepsilon z' &= (1-y^2)z - y & z(0) &= -\frac{2}{3} + \frac{10}{81}\varepsilon + \frac{292}{2187}\varepsilon^2 \end{aligned} \tag{6}$$

with $\varepsilon = 10^{-6}$. The problem is solved with an L-stable, stiffly accurate SDIRK method of order 4 with an error estimator of order 3, given by Hairer and Wanner [13, Table 6.5]. The system is integrated to the first step exceeding $x = 0.3$. At this point we measure the exact local error and the local error estimate of the next step as functions of h. The results are given in Figure 1. The plot shows the 2-norm of the errors, scaled with the tolerance. Thus the next step will be accepted if the local error estimate is less than 1.

The method is a stiffly accurate SDIRK method, thus the stability constant $R(\infty) = 0$ and the stage order $q = 1$. The index 2 interval of h goes from approximately $5 \cdot 10^{-6}$ to 0.1. Here the exact local error is dominated by contribution from the index 2 z_1-component, first by the term $\mathcal{O}(1/h)$, then by the local truncation error term $\mathcal{O}(h)$. This somewhat peculiar behaviour do not cause any problems since the error is well below the error tolerance. For $h > 0.1$, the index 1 interval, the error is dominated by contributions from the index 1 terms, and no order reduction of the exact local error occur.

The error estimating method is not stiffly accurate, and $\hat{R}(\infty) = 10/3$. The error estimate is dominated by contributions from the z-component. In the index 2 interval, the numerical values from the last step do not satisfy the hidden constraint, this error is multiplied by $\hat{R}(\infty)$, causing the $\mathcal{O}(1)$ behaviour. Again, the index 2 error causes no problem, since it is well below the error tolerance. There are however situations where this contribution might exceed one, causing a dramatically reduction of the operating stepsize. The behaviour of the error estimate for larger stepsizes is more problematic. The local order of the z-component of the error estimating method applied to index 1 problems is only 2. The low order causes the method to operate on smaller stepsizes than really necessary. Also, stepsize selection algorithms designed on the assumption of error estimates of classical order will fail to work properly unless some remedial action is taken.

The methods of the next section are constructed according to the following list of requirements:

1. Stiffly accuracy in both the advancing and the error estimating methods. In this case there is no order reduction in the index 1 interval.

2. $R(\infty) = 0$, and $|\hat{R}(\infty)|$ as small as possible, at least less than one. This to reduce the influence of inconsistent numerical values from the last step.

3. A-stability, or at least $A(\alpha)$-stability for both methods.

4. As high stage order as possible.

Remark 1 Methods satisfying requirements 1 and 2 can be directly adapted to index 1 DAEs of the form

$$My' = f(y)$$

where M is a constant, singular matrix.

Remark 2 A-stability and high stage order are strictly speaking not required for solving very stiff SPPs. By adding these requirements, the methods are capable to solve more general classes of stiff problems.

Remark 3 The list could have included other properties like B-stability, high order of B-convergence, etc. And if the method is adapted to problems with discontinuities, we would require $c_i \leq 1,\ i = 1, \cdots, s$ and high order interpolants. Special problems requires special methods. We do however believe that methods satisfying requirements 1-4 are capable to solve a large class of stiff ODEs satisfactorily.

3 Singly diagonally implicit Runge-Kutta methods with an explicit first stage

This section is devoted to the construction of methods of SDIRK type, according to the specifications given in Section 2. For simplicity, we consider ODEs of the generic form

$$y' = f(y)$$

here. Requirement 1, stiffly accuracy for both the advancing and the error estimating method is fulfilled by using the last two stage values. Thus, if the complete method is stiffly accurate and of order p, we want the first $s-1$ stages to form a stiffly accurate method of order $p-1$. The error estimate (4) is then simply given by

$$err = Y_s - Y_{s-1}.$$

Both Y_s and Y_{s-1} can be used to advance the solution. But, if requirement 2 is to be fulfilled, then these two options leads to slightly different methods. Both possibilities will be considered.

The highest attainable stage order of SDIRK methods is 1. Although this may not be a serious objections to the methods, the stage order can be increased to 2 by using an explicit first stage.

Such methods will in the following be denoted as ESDIRK $p/p-1$, followed by a for methods using local extrapolation ($y_{n+1} = Y_s$), b otherwise.

Order conditions. The Butcher tableau of ESDIRK $p/p-1$ is given by

$$\begin{array}{c|cccccccc}
0 & 0 & & & & & & & \\
c_2 & a_{21} & \gamma & & & & & & \\
c_3 & a_{31} & a_{32} & \gamma & & & & & \\
\vdots & \vdots & & & \ddots & & & & \\
\vdots & \vdots & & & & \ddots & & & \\
c_{s-2} & a_{s-2,1} & a_{s-2,2} & a_{s-2,3} & \cdots & \cdots & \gamma & & \\
1 & a_{s-1,1} & a_{s-1,2} & a_{s-1,3} & \cdots & \cdots & a_{s-1,s-2} & \gamma & \\
1 & a_{s1} & a_{s2} & a_{s3} & \cdots & \cdots & a_{s,s-2} & a_{s,s-1} & \gamma
\end{array} . \tag{7}$$

Stage order 2 is given by the condition

$$C(2): \qquad \sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k \qquad i = 1, \cdots, s, \qquad k = 1, 2.$$

This is used to find the first two columns of the coefficient matrix A, thus

$$\begin{aligned}
a_{i1} &= c_i - \sum_{j=2}^{i-1} a_{ij} - \gamma, && i = 2, \cdots, s, \\
a_{i2} &= \frac{1}{c_2}\left(\frac{1}{2}c_i^2 - \sum_{j=3}^{i-1} a_{ij} c_j - \gamma c_i\right), && i = 2, \cdots, s.
\end{aligned}$$

In particular, for $i = 2$ we get

$$c_2 = 2\gamma, \qquad a_{21} = \gamma.$$

With $C(2)$ satisfied, the remaining conditions for ESDIRK methods of order $p \leq 5$ are

$$B(p): \quad \sum_{i=1}^{s} b_i c_i^{k-1} = \tfrac{1}{k}, \quad k = 3, \cdots, p,$$

in addition to

$$\begin{aligned}
&\text{Order 4:} \quad \sum_{i,j=1}^{s} b_i a_{ij} c_j^2 = \frac{1}{12}. \\
&\text{Order 5:} \quad \sum_{i,j=1}^{s} b_i c_i a_{ij} c_j^2 = \frac{1}{15}, \quad \sum_{i,j=1}^{s} b_i a_{ij} c_j^3 = \frac{1}{20}, \quad \sum_{i,j,k=1}^{s} b_i a_{ij} a_{jk} c_k^2 = \frac{1}{60},
\end{aligned}$$

where b_i is either a_{si} or $a_{s-1,i}$.

Based on these order conditions, methods of order 3/2, 4/3 and 5/4 are constructed. When the order conditions are satisfied, the remaining free parameter is tuned to satisfy the stability requirements.

s	A-stability	L-stability
3	$1/4 \leq \gamma < \infty$	$\gamma = \frac{2\pm\sqrt{2}}{2}$
4	$1/3 \leq \gamma \leq 1.06860$	$\gamma = 0.4358665215$
5	$0.39434 \leq \gamma \leq 1.28060$	$\gamma = 0.5728160625$

Table 1: Stability of stiffly accurate ESDIRK methods of order $p \geq s-1$.

Stability. The stability function of a RK-method can be written as

$$R(z) = \frac{\det(I - zA + z\mathbb{1}b^T)}{\det(I - zA)} = \frac{P(z)}{Q(z)}.$$

For a stiffly accurate ESDIRK method the polynomial $P(z)$ is at most of degree $s-1$, and $Q(z) = (1-\gamma z)^{s-1}$. As a consequence, ESDIRK methods of order $p \geq s-1$ have the same stability functions as $s-1$ stage SDIRK-methods of order $p \geq s-1$. Such functions was first studied by Nørsett [18], see [13, Sec. IV.6] for a more available reference. The numerator is

$$P(z) = (-1)^{s-1} \sum_{j=0}^{s-1} L_{s-1}^{(s-j)}\left(\frac{1}{\gamma}\right)(\gamma z)^j,$$

where

$$L_{s-1}(x) = \sum_{j=0}^{s-1} (-1)^j \binom{s-1}{j} \frac{x^j}{j!}$$

are the Laguerre-polynomials, and $L_s^{(k)}(x)$ denotes their k-th derivative. The requirement $R(\infty) = 0$ is then satisfied when

$$\gamma^{s-1} L_{s-1}\left(\frac{1}{\gamma}\right) = 0$$

The regions of γ for A- and L-stability are given in Table 1.

ESDIRK 3/2 in 4 stages. Such methods are given by the Butcher-tableau

$$\begin{array}{c|cccc}
0 & 0 & 0 & 0 & 0 \\
2\gamma & \gamma & \gamma & 0 & 0 \\
1 & \frac{-4\,\gamma^2+6\,\gamma-1}{4\,\gamma} & \frac{-2\,\gamma+1}{4\,\gamma} & \gamma & 0 \\
1 & \frac{6\,\gamma-1}{12\,\gamma} & \frac{-1}{(24\,\gamma-12)\gamma} & \frac{-6\,\gamma^2+6\,\gamma-1}{6\,\gamma-3} & \gamma
\end{array} .$$

The free parameter γ is chosen according to the stability properties given in Table 1. Thus:

a) If $y_{n+1} = Y_4$ then $\gamma = 0.4358665215$ with $|\hat{R}(\infty)| = 0.9569$.
b) If $y_{n+1} = Y_3$ then $\gamma = 0.2928932188$ with $|\hat{R}(\infty)| = 1.609$.

The first of these choices satisfies all the requirements. But the error estimator of the second method fails to satisfy $|R(\infty)| < 1$. If this method is used, the error estimate could be handled by some trick similar to the one used by Hairer and Wanner.

These methods were first presented by Alexander and Coyle [1] for solving index 2 DAEs. Their choice of γ differs from ours, since their concern is to satisfy order conditions rather than stability requirements.

ESDIRK 4/3 in 5 stages. The coefficients of these methods are given by

$$\begin{aligned}
a_{21} &= \gamma, \\
a_{31} &= \frac{144\,\gamma^5-180\,\gamma^4+81\,\gamma^3-15\,\gamma^2+\gamma}{(12\,\gamma^2-6\,\gamma+1)^2}, \\
a_{32} &= \frac{-36\,\gamma^4+39\,\gamma^3-15\,\gamma^2+2\,\gamma}{(12\,\gamma^2-6\,\gamma+1)^2}, \\
a_{41} &= \frac{-144\,\gamma^5+396\,\gamma^4-330\,\gamma^3+117\,\gamma^2-18\,\gamma+1}{12\,\gamma^2(12\,\gamma^2-9\,\gamma+2)}, \\
a_{42} &= \frac{72\,\gamma^4-126\,\gamma^3+69\,\gamma^2-15\,\gamma+1}{12\,\gamma^2(3\,\gamma-1)}, \\
a_{43} &= \frac{\left(-6\,\gamma^2+6\,\gamma-1\right)\left(12\,\gamma^2-6\,\gamma+1\right)^2}{12\,\gamma^2(12\,\gamma^2-9\,\gamma+2)(3\,\gamma-1)}, \\
a_{51} &= \frac{288\,\gamma^4-312\,\gamma^3+120\,\gamma^2-18\,\gamma+1}{48\,\gamma^2(12\,\gamma^2-9\,\gamma+2)}, \\
a_{52} &= \frac{24\,\gamma^2-12\,\gamma+1}{48\,\gamma^2(3\,\gamma-1)}, \\
a_{53} &= \frac{-\left(12\,\gamma^2-6\,\gamma+1\right)^3}{48\,\gamma^2\,(3\,\gamma-1)(12\,\gamma^2-9\,\gamma+2)(6\,\gamma^2-6\,\gamma+1)} \\
a_{54} &= \frac{-24\,\gamma^3+36\,\gamma^2-12\,\gamma+1}{24\,\gamma^2-24\,\gamma+4}.
\end{aligned}$$

The parameter γ are chosen as:

$a)$ If $y_{n+1} = Y_5$ then $\gamma = 0.5728160625$ with $|\hat{R}(\infty)| = 0.5525$.
$b)$ If $y_{n+1} = Y_4$ then $\gamma = 0.4358665215$ with $|\hat{R}(\infty)| = 0.7175$.

The first of these methods has the drawback that $\gamma > 1/2$, giving $c_2 > 1$, which makes it less suitable for some problems.

ESDIRK 5/4 in 6 stages do not exist. This can be proved by using $C(2)$ to find the first two columns of the coefficient matrix. The coefficients a_{53} and a_{54} are found by $B(4)$ with $b_i = a_{5i}$, and a_{63}, a_{64} and a_{65} by $B(5)$ with $b_i = a_{6i}$. The coefficient a_{43} is solved by the order 4 condition, using $b_i = a_{5i}$. The same order condition, but now with $b_i = a_{6i}$ is used to solve for c_4. The first of the order 5 conditions is used to solve for c_3. The last two equations now become

$$\begin{aligned}
-\frac{\left(6\,\gamma^3-12\,\gamma^2+6\,\gamma-1\right)\gamma}{3} &= \frac{1}{60} \\
-\frac{\left(1440\,\gamma^6-4128\,\gamma^5+4224\,\gamma^4-2080\,\gamma^3+526\,\gamma^2-65\,\gamma+3\right)\gamma}{360\,\gamma^3-228\,\gamma^2+48\,\gamma-3} &= \frac{1}{20}
\end{aligned}$$

which have no common solutions.

ESDIRK 5/4 in 7 stages. By adding one more stage, there are several possibilities of choosing methods. Two methods satisfying the four requirements are presented in Table 2.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.26000000000000000 & 0.26000000000000000 & 0 & 0 & 0 & 0 & 0 \\ 0.13000000000000000 & 0.84033320996790809 & 0.26000000000000000 & 0 & 0 & & \\ 0.22371961478320505 & 0.47675532319799699 & -0.06470895363112615 & 0.26000000000000000 & 0 & 0 & 0 \\ 0.16648564323248321 & 0.10450018841591720 & 0.03631482272098715 & -0.13090704451073998 & 0.26000000000000000 & 0 & 0 \\ 0.13855640231268224 & 0 & -0.04245337201752043 & 0.02446657898003141 & 0.61943039072480676 & 0.26000000000000000 & 0 \\ 0.13659751177640291 & 0 & -0.05496908796538376 & -0.04118626728321046 & 0.62993304899016403 & 0.06962479448202728 & 0.26 \end{bmatrix}$$

ESDIRK 5/4a using $y_{n+1} = Y_7$, with $|\hat{R}(\infty)| = 0.7483$.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.27000000000000000 & 0.27000000000000000 & 0 & 0 & 0 & 0 & 0 \\ 0.13500000000000000 & 0.87265371804359686 & 0.27000000000000000 & 0 & 0 & 0 & 0 \\ 0.24814211234447322 & 0.13282088522859322 & -0.03886686658917771 & 0.27000000000000000 & 0 & 0 & 0 \\ 0.25494479822150471 & 0.13106196422347200 & -0.04522093930235708 & 0.03389121682051642 & 0.27000000000000000 & 0 & 0 \\ 0.17549975523182941 & 0 & -0.01641725931492383 & 3.59357175290010625 & -3.02265424881701182 & 0.27000000000000000 & 0 \\ 0.15847612643670410 & 0 & -0.07384703732094983 & 5.26056776397634893 & -4.83946947758407500 & 0.22427262449197180 & 0.27 \end{bmatrix}$$

ESDIRK 5/4b using $y_{n+1} = Y_6$, with $|\hat{R}(\infty)| = 0.8732$.

Table 2: ESDIRK methods of order 5/4 .

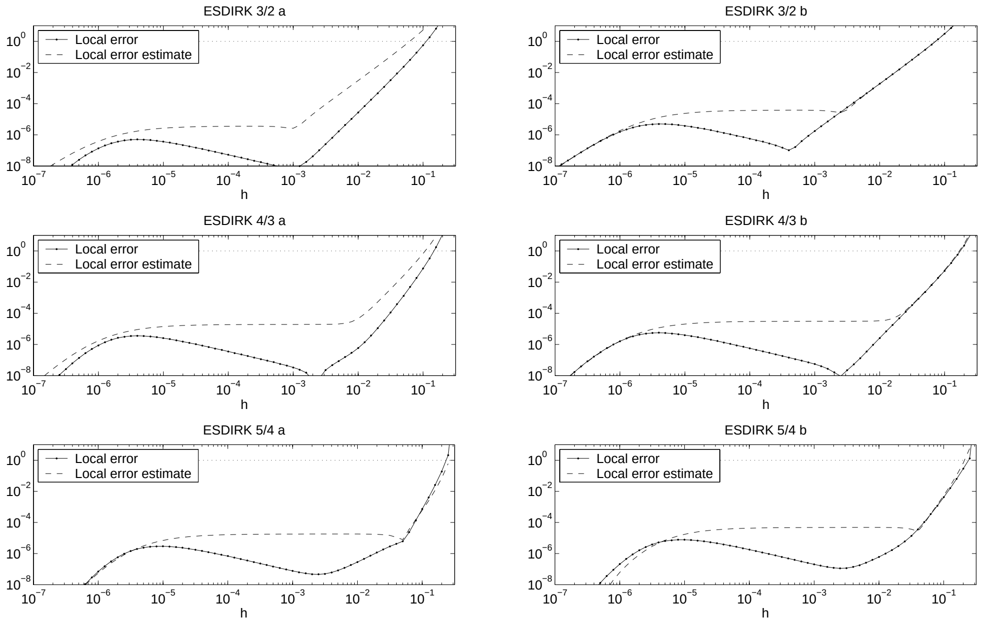

Figure 2: Local error and error estimates for the ESDIRK-methods.

At the conclusion of this section, the experiments described in Example 1 is carried out again, this time by using the ESDIRK methods. The results are given in Figure 2. We observe that the operating stepsize is in the index 1 interval for all the methods, and neither the local error nor the error estimate suffer from order reduction in this area. As expected, the methods using local extrapolation ($y_{n+1} = Y_s$) overestimate the local error. For the remaining methods, the correspondence between the error estimate and the error is quite good.

The error estimate in the index 2 area is a factor of about 10^{-5} of the tolerance, and will usually not cause any problems.

4 Further reading

In this paper we focus on the theoretical background for the choice of methods through the study of SPPs. The ESDIRK methods were originally developed as a part of a now terminated ODE software project, and have been available in an unpublished note by the author [17]. For experience with the methods, as well as implementation issues, see [22, 23, 24, 25]. The ESDIRK methods have been used as the implicit part of IMEX-methods [15]. The methods can work well also in real-life simulations. This has been demonstrated in [28], where the low order pair ESDIRK 3/2a has been successfully used for solving ODEs modelling electrical activity in cardiac cells.

Acknowledgements

The author is in great debt to Syvert P. Nørsett for introducing her to the subject of numerical ODEs/DAEs, and for being a great teacher, supervisor and colleague over the years. She will also give her thanks to the rest of the GODESS team, in particular Per Grove Thomsen, Gustaf Söderlind and Hans Olsson.

References

[1] R. K. Alexander and J. J. Coyle. Runge-Kutta methods and differential-algebraic systems. *SIAM J. Numer. Anal.*, 27(3):736–752, 1990.

[2] R.K. Alexander. Diagonally implicit Runge-Kutta methods for stiff O.D.Es. *SIAM J. Numer. Anal.*, 14:1006–1024, 1977.

[3] R. Alt. *Méthodes A-stables pour l'intégration de systémes différentielles mal conditionneés.* PhD thesis, Université Paris, 1971.

[4] U. M. Ascher, S. J. Ruuth, and R. J. Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *App. Numer. Math.*, 25:151–167, 1997.

[5] J.C. Butcher. Implicit Runge-Kutta processes. *Math. Comp.*, 18:50–64, 1964.

[6] J.R. Cash. Diagonally implicit Runge-Kutta formulae with error estimates. *J. Inst. Maths. Applics.*, 24:293–301, 1979.

[7] M. Crouzeix. *Sur l'approximation des équations différentielles opérentielles lineéaires par des méthodes de Runge-Kutta.* PhD thesis, Université Paris, 1975.

[8] K. Dekker and J.G. Verwer. *Stability of Runge-Kutta methods for stiff nonlinear differential equations.* CWI Monographs. North-Holland, 1984.

[9] P. Deuflhard, E. Hairer, and J. Zugck. One-step and extrapolation methods for differential-algebraic systems. *Numer. Math.*, 51:501–516, 1987.

[10] E. Griepentrog and R. März. *Differential-algebraic equations and their numerical treatment.* Band 88. Teubner-Texte zur Mathematik, 1986.

[11] E. Hairer, Ch. Lubich, and M. Roche. Error of Runge-Kutta methods for stiff problems studied via differential algebraic equations. *BIT*, 28:678–700, 1988.

[12] E. Hairer, Ch. Lubich, and M. Roche. *The numerical solution of differential-algebraic systems by Runge-Kutta methods.* Number 1409 in Lecture Notes in Mathematics. Springer Verlag, 1989.

[13] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II, Stiff and Differential-Algebraic Problems.* Springer, Berlin, Second revised edition, 2002.

[14] L. Jay. Convergence of a class of Runge-Kutta methods for differential-algebraic systems of index 2. *BIT*, 33:137–150, 1993.

[15] C. A. Kennedy and M. H. Carpenter. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *App. Numer. Math.*, 44:139–181, 2003.

[16] M.A. Kurdi. *Stable high order methods for time discretization of stiff differential equations.* PhD thesis, University of California, 1974.

[17] A. Kværnø. More, and to be hoped, better DIRK-methods for the solution of stiff ODEs. Unpublished note.

[18] S.P. Nørsett. *Numerical solution of ordinary differential equations.* PhD thesis, University of Dundee, 1975.

[19] S.P. Nørsett and P.G. Thomsen. Embedded SDIRK-methods of basic order three. *BIT*, 24:634–646, 1984.

[20] S.P. Nørsett and P.G. Thomsen. Local error control in SDIRK-methods. *BIT*, 26:100–113, 1986.

[21] S.P. Nørsett and P.G. Thomsen. Switching between modified and fix-point iteration for implicit ODE-solvers. *BIT*, 26:339–348, 1986.

[22] H. Olsson. Practical implementation of Runge-Kutta methods for initial value problems. Lic. Thesis, Lund University, Sweden, 1995.

[23] H. Olsson. *Runge-Kutta solution of initial value problems.* PhD thesis, Lund University, 1998.

[24] H. Olsson and G. Söderlind. Stage value predictors and efficient Newton iterations in implicit Runge-Kutta methods. *SIAM J. Sci. Comput.*, 20(1):185–202, 1998.

[25] H. Olsson and G. Söderlind. The approximate Runge-Kutta computational process. *BIT*, 40(2):351–373, 2000.

[26] L. Pareshci and G. Russo. Implicit-explcit Runge-Kutta schemes for stiff systems of differential equations. *Recent Trends in Numerical Analysis*, 3:269–289, 2000.

[27] M. Roche. Implicit Runge-Kutta methods for differential algebraic equations. *SIAM J. Numer. Anal.*, 26:963–975, 1989.

[28] J. Sundnes, G.T. Lines, and A. Tveito. Efficient solution of ordinary differential equations modeling electrical activity in cardiac cells. *Math. Biosci.*, 172(2):55–72, 2001.