Convergence Properties of the Runge-Kutta-Chebyshev Method

J.G. Verwer, W.H. Hundsdorfer and B.P. Sommeijer
Centre for Mathematics and Computer Science (CWI)
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

1. Introduction

This paper is devoted to an examination of convergence properties of a class of Runge-Kutta-Chebyshev (RKC) schemes. These schemes have been designed by van der Houwen and Sommeijer (1980) for the *explicit* time integration of stiff systems of ODEs,

$$\dot{U}(t) = F(t, U(t)),\ 0 < t \leqslant T,\ U(0) \text{ given}, \tag{1.1}$$

which originate from spatial discretization of *parabolic* partial differential equations (Method of Lines). For the time being, it is not necessary to define a particular class of parabolic problems or to specify the space discretization technique. The only restrictions for application of the RKC schemes are (i) The eigenvalue spectrum of the Jacobian matrix $\partial F(t, U)/\partial U$ should lie in a narrow strip along the negative axis of the complex plane, and (ii) The Jacobian matrix should 'not deviate too much from a normal matrix'. These two conditions trivially hold if $\partial F(t, U)/\partial U$ is symmetric and negative definite, properties frequently encountered when discretizing elliptic operators.

The RKC method is a typical example of an *explicit, stabilized RK method*. The method has been designed such that it possesses an *extended real stability interval*. Its real stability boundary β is in fact proportional to s^2, s being the number of stages, while its main characteristic is that s can be taken arbitrarily large. This is made possible by an intelligent use of Chebyshev polynomials, thus explaining the name of the method. The possibility of using arbitrarily large values for s is of practical relevance due to the fact that the effective real stability boundary β/s linearly increases with s. Hence, in applications it is possible and advantageous to choose the stepsize on the basis of accuracy and to adjust s to meet the demand of (linear) stability

Van der Houwen and Sommeijer (1980) have developed a 1-st and 2-nd order RKC scheme. In this paper we examine both these schemes (the coefficients of our 2-nd order schemes slightly differ; they have been taken from Sommeijer and Verwer (1980)). While these schemes have been developed along the lines of the classical ODE theory, the purpose of the present examination is to analyse their *full convergence properties*. Full convergence means convergence of the fully discrete solution with respect to the solution of the PDE upon simultaneous refinement of the space-time mesh. For linear PDE problems, whose semi-discretizations take the form

$$\dot{U}(t) = MU(t) + g(t),\ 0 \leqslant t \leqslant T,\ U(0) \text{ given}, \tag{1.2}$$

with M a *symmetric*, constant coefficient matrix possessing *non-positive eigenvalues*, we prove convergence under the sole condition that the necessary time-step restriction $\tau\sigma(M) \leqslant \beta$ is satisfied, where $\sigma(M)$ is the spectral radius of M. Of interest is that the derived error bounds are independent of $\sigma(M)$ and valid for arbitrarily large s, the number of stages, thus showing that in applications the best strategy is to have the size of τ determined by the desired accuracy level and s by the stability demand. It is stipulated that this result is quite uncommon for an explicit method. We owe this to the favourable *internal stability* property of the RKC method. Internal stability has to do with the propagation of errors over the stages within one single integration step.

The convergence analysis presented in this paper is akin to the analysis of Sanz-Serna, Verwer and Hundsdorfer (1987) and Hundsdorfer and Verwer (1989) which, in turn, was inspired by the B-convergence analysis from the stiff ODE field (see Dekker and Verwer (1984), Ch.7). The present paper is a condensed version of [13].

2. DESCRIPTION OF THE METHOD

For the ODE system (1.1), the RKC formula considered here is of the form [6, 12]

$$\begin{aligned} Y_0 &= U_n, \\ Y_1 &= Y_0 + \tilde{\mu}_1 \tau F_0, \\ Y_j &= \mu_j Y_{j-1} + \nu_j Y_{j-2} + (1-\mu_j-\nu_j) Y_0 + \tilde{\mu}_j \tau F_{j-1} + \tilde{\gamma}_j \tau F_0 \quad (2 \leqslant j \leqslant s), \\ U_{n+1} &= Y_s, \quad n = 0, 1, \ldots, \end{aligned} \tag{2.1}$$

where $F_j = F(t_n + c_j\tau, Y_j)$; U_n represents the approximation to the exact solution U of (1.1) at time $t = t_n$ and $\tau = t_{n+1} - t_n$ is the stepsize. Throughout it is assumed that the increment parameters c_j are defined by the integration coefficients $\mu_j, \nu_j, \tilde{\mu}_j$ and $\tilde{\gamma}_j$ in the following way,

$$\begin{aligned} c_0 &= 0, \; c_1 = \tilde{\mu}_1, \\ c_j &= \mu_j c_{j-1} + \nu_j c_{j-2} + \tilde{\mu}_j + \tilde{\gamma}_j \quad (2 \leqslant j \leqslant s). \end{aligned} \tag{2.2}$$

Then, if we bring Y_j in the standard RK form

$$Y_j = U_n + \tau \sum_{l=0}^{j-1} a_{jl} F(t_n + c_l\tau, Y_l) \quad (0 \leqslant j \leqslant s),$$

where the coefficients a_{jl} are expressions in $\mu_j, \nu_j, \tilde{\mu}_j, \tilde{\gamma}_j$, it is readily seen that the usual condition

$$c_j = \sum_{l=0}^{j-1} a_{jl}$$

is satisfied. Hence, (2.1) is an *explicit, s-stage RK method* and Y_j is an intermediate approximation at the intermediate point $t = t_n + c_j\tau$. Due to the specific recursive nature of the method, as shown in the formula defining Y_j, formula (2.1) is more convenient to work with than the common RK formula. The rationale behind the specific form (2.1) is that this form is easily identified with stable three-term Chebyshev recursions. This will become clear later on. Note that irrespective the number of stages, the number of required storage arrays is maximal 6.

Let us determine the *consistency conditions* (in the classical ODE sense) for order 1 and 2. Suppose $U_n = U(t_n)$, where $U(t), t \geqslant t_n$ is a sufficiently smooth solution of (1.1). By definition of c_j it then holds that all Y_j satisfy an expansion

$$Y_j = U(t_n) + c_j\tau \dot{U}(t_n) + X_j \tau^2 U^{(2)}(t_n) + O(\tau^3), \tag{2.3}$$

where, similar as c_j, X_j is determined by the integration coefficients. Substitution of this expression into (2.1) gives

$$X_0 = X_1 = 0, \; X_j = \mu_j X_{j-1} + \nu_j X_{j-2} + \tilde{\mu}_j c_{j-1} \quad (2 \leqslant j \leqslant s). \tag{2.4}$$

We conclude that the RKC method is *consistent of order 1* if

$$c_s = 1, \tag{2.5}$$

and note that the *j-th* stage formula is consistent of order 1 at $t = t_n + c_j\tau$.

It follows from (2.3) that each stage formula is consistent of order 2 at $t=t_n+c_j\tau$, for $2\leqslant j\leqslant s$, if $X_j=\frac{1}{2}c_j^2$. In terms of c_j this gives

$$c_2^2=2\tilde{\mu}_2c_2,\ c_3^2=\mu_3c_2^2+2\tilde{\mu}_3c_2, \tag{2.6}$$
$$c_j^2=\mu_jc_{j-1}^2+\nu_jc_{j-2}^2+2\tilde{\mu}_jc_{j-1}\quad(4\leqslant j\leqslant s).$$

As pointed out in [12], it is possible to satisfy this condition in a satisfactory way for all $2\leqslant j\leqslant s$. We here adopt this condition and hence the 2-nd order scheme derived below has all its stages consistent of order 2 at the intermediate step points $t=t_n+c_j\tau$, except the first one. The original 2-nd order scheme from [6] is only consistent of order 2 at the main step points.

For future reference, it is stipulated that the derivation of the current consistency conditions follows the lines of the classical numerical ODE theory [3,5], as it is based on expanding F-terms. This means that it is tacitly assumed here that F satisfies a Lipschitz condition so that $\tau\|F\|=O(\tau)$. For stiff problems this is unduly restrictive and particularly so for semi-discrete parabolic equations for which $\|F\|\to\infty$ upon grid refinement. In Section 4 we will re-examine the consistency properties of the RKC scheme. The derivation presented there is inspired by the B-convergence theory for stiff ODEs, the central theme of which is the derivation of error bounds which do not depend on the stiffness of the problem (see [2], Ch. 7 and [8,9,10]).

Finally, a natural condition is that all (intermediate) step points lie within the step interval $[t_n,t_{n+1}]$ and increase monotonically with j:

$$0=c_0<c_1<c_2<\cdots<c_{s-1}<c_s=1. \tag{2.7}$$

It will turn out this condition is satisfied for the two selected schemes.

We proceed with the *stability function*. Because the RKC method is an s-stage, explicit RK method, application to the scalar test equation $\dot{U}(t)=\lambda U(t)$ leads to the linear, one-step recursion

$$U_{n+1}=P_s(z)U_n,\quad z=\tau\lambda, \tag{2.8}$$

where the stability function $P_s:\mathbf{C}\to\mathbf{C}$ is a polynomial of degree s. P_s itself is also defined recursively as follows:

$$P_0(z)=1,\ P_1(z)=1+\tilde{\mu}_1z, \tag{2.9}$$
$$P_j(z)=(1-\mu_j-\nu_j)+\tilde{\gamma}_jz+(\mu_j+\tilde{\mu}_jz)P_{j-1}(z)+\nu_jP_{j-2}(z)\quad(2\leqslant j\leqslant s).$$

In fact, all polynomials P_j are of degree j and satisfy

$$Y_j=P_j(z)U_n\quad(0\leqslant j\leqslant s). \tag{2.10}$$

Therefore we will also call the intermediate polynomials P_j stability functions, but note that they play no role in the step-by-step stability like P_s.

According to (2.3), each stability function $P_j(z)$ approximates the exponential e^{c_jz} for $z\to0$ as

$$P_j(z)=1+c_jz+X_jz^2+O(z^3). \tag{2.11}$$

Hence, each P_j is consistent of order 1 (with the exponential e^{c_jt}) and consistent of order 2 if, in addition, $X_j=c_j^2/2$. Substitution of this expansion into (2.9) and equating powers of z then reveals relations (2.2) and condition (2.6). Hence, if we select the coefficients $\mu_j,\nu_j,\tilde{\mu}_j,\tilde{\gamma}_j$ in the recursion (2.9) such that P_j is of order 1 or 2 in the sense of (2.11), then the 1-st or 2-nd order conditions associated to expansion (2.3) are automatically satisfied. This is very convenient since it enables us to concentrate entirely on the choice of the stability functions.

The choice of the stability functions P_j is the central issue in the development of the RKC method. This choice underlies two *design rules:*
(I) The coefficients $\mu_j, \nu_j, \tilde{\mu}_j, \tilde{\nu}_j$ in the recursion (2.9) are chosen such that the *real stability boundary*

$$\beta(s) = \max\{-z : z \leqslant 0,\ |P_s(z)| \leqslant 1\} \tag{2.12}$$

of the genuine stability function P_s is as large as possible, so as to obtain good stability properties for parabolic equations. This requirement leads to the *Chebyshev polynomial of the first kind*

$$T_s(x) = \cos(s \arccos x),\quad -1 \leqslant x \leqslant 1,$$

to which we owe the quadratic increase of $\beta(s)$ with s (van der Houwen (1977), p. 89). For example, within the class of 1-st order consistent polynomials, the shifted Chebyshev polynomial

$$P_s(z) = T_s(1 + \frac{z}{s^2}),\quad -\beta(s) \leqslant z \leqslant 0, \tag{2.13}$$

yields the largest possible value for $\beta(s)$, viz., $\beta(s) = 2s^2$. □

(II) The second design rule has to do with the desirability of applying the method with an arbitrary number of stages which means that, given s, all coefficients $\mu_j, \nu_j, \tilde{\mu}_j, \tilde{\gamma}_j$ must be known in analytic form. Further, and this is most important, it should be possible to let s arbitrarily large without severe accumulation of errors within one single step *(internal stability).* The notion of internal stability will be discussed in Section 3. Here we mention that both these requirements are fulfilled by adjusting the three-term recursion (2.9) for P_j to the known three-term recursion of appropriately chosen shifted Chebyshev polynomials. For example, the polynomials $P_j(z) = T_j(1 + z/s^2)$ satisfy the recursion

$$P_0(z) = 1,\ P_1(z) = 1 + \frac{z}{s^2},\ P_j(z) = 2(1 + \frac{z}{s^2})P_{j-1} - P_{j-2},\ j \geqslant 2, \tag{2.14}$$

and adjusting (2.9) gives

$$\tilde{\mu}_1 = 1/s^2,\ \mu_j = 2,\ \tilde{\mu}_j = 2/s^2,\ \nu_j = -1,\ \tilde{\gamma}_j = 0 \quad (2 \leqslant j \leqslant s).$$

Note that $|P_j(z)| \leqslant 1$ for all $j \leqslant s$ as long as z lies within the real stability interval $[-2s^2, 0]$ of the genuine stability function P_s. □

Having outlined these two design rules, we are now ready to specify the stability functions P_j with the associated coefficient sets for the 1-st and 2-nd order RKC schemes examined in this paper. They all fit in the general form

$$P_j(z) = a_j + b_j\, T_j(w_0 + w_1 z),\quad 0 \leqslant j \leqslant s, \tag{2.15}$$

where the parameters a_j, b_j, w_0 and w_1 have been chosen in accordance with the design rules (I) and (II). Before specifying them, there is one point left that should be mentioned (to save space we must refer to [5,6] for more details). This point concerns the parameter w_0. Consider the polynomial (2.13) where $w_0 = 1$. This polynomial alternates between $+1$ and -1, i.e., $|P_s(z)| = 1$ at $s+1$ points $z \in [-\beta, 0]$. It is desirable to introduce some damping in P_s, i.e., to let P_s alternate between values $\simeq 1-\epsilon$ and $\simeq -1+\epsilon$ for all $z \in [-\beta, 0]$ (with the exception of a small neighbourhood of $z = 0$), where ϵ is a small positive number. The damping is obtained by choosing $w_0 = w_0(\epsilon)$, called the *damping parameter,* slightly larger than 1. By introducing this damping in the stability function, we achieve that the *stability region* becomes a long, narrow strip around

the negative axis of the complex plane. On the other hand, the real stability boundary slightly decreases [5]. There is practical evidence that with damping the RKC method becomes more robust for nonlinear problems.

The 1-st order case : RKC1 [6]

$$a_j=0,\ b_j=T_j^{-1}(w_0),\ w_0=1+\frac{\epsilon}{s^2},\ w_1=\frac{T_s(w_0)}{T_s'(w_0)} \qquad (0\leqslant j\leqslant s). \tag{2.16}$$

It can be shown that with this choice of parameters

$$\beta(s)\simeq\frac{(w_0+1)T_s'(w_0)}{T_s(w_0)}\simeq(2-\tfrac{4}{3}\epsilon)s^2 \quad \text{for} \quad \epsilon\to 0.$$

A suitable values for ϵ is 0.05. Since $T_s^{-1}(w_0)\simeq 1-\epsilon$, this yields about 5% damping with only a very little decrease in $\beta(s)$, $\beta(s)\simeq 1.90s^2$. Note that with $\epsilon=0$ we recover the polynomials (2.14). Adjusting recursion (2.9) to the current choice for P_j completely defines the general 1-st order scheme (2.1):

$$\begin{aligned}
&\tilde{\mu}_1=\frac{w_1}{w_0},\\
&\mu_j=2w_0\frac{b_j}{b_{j-1}},\quad \nu_j=-\frac{b_j}{b_{j-2}},\\
&\tilde{\mu}_j=2w_1\frac{b_j}{b_{j-1}},\quad \tilde{\gamma}_j=0 \qquad (2\leqslant j\leqslant s).
\end{aligned} \tag{2.17}$$

Note that each value of s defines a different coefficient set. Also note that $\mu_j+\nu_j=1$ and that the increment parameters

$$c_j=\frac{T_s(w_0)}{T_s'(w_0)}\,\frac{T_j'(w_0)}{T_j(w_0)}\simeq j^2/s^2 \tag{2.18}$$

satisfy condition (2.7). For more details we refer to [5,6].

The 2-nd order case : RKC2 [6,12]

$$\begin{aligned}
&a_j=1-b_jT_j(w_0),\ b_j=\frac{T_j''(w_0)}{(T_j'(w_0))^2},\ w_0=1+\frac{\epsilon}{s^2},\ w_1=\frac{T_s'(w_0)}{T_s''(w_0)} \qquad (2\leqslant j\leqslant s),\\
&a_0=1-b_0,\ a_1=1-b_1w_0,\ b_0=b_1=b_2.
\end{aligned} \tag{2.19}$$

For this choice of parameters one can prove that

$$\beta(s)\simeq\frac{(w_0+1)T_s''(w_0)}{T_s'(w_0)}\simeq\tfrac{2}{3}(s^2-1)(1-\tfrac{2}{15}\epsilon) \quad \text{for} \quad \epsilon\to 0.$$

A suitable value for ϵ is 2/13. This gives about 5% damping ($a_s+b_s\simeq 1-\frac{2}{3}\epsilon$) with a reduction in $\beta(s)$ of about 2%. The current choice of stability polynomials covers roughly 80% of the optimal real stability interval for 2-nd order consistent polynomials (van der Houwen (1982)). Adjusting again recursion (2.9), completely defines the general 2-nd order scheme (2.1):

$$\begin{aligned}
&\tilde{\mu}_1=b_1w_1,\\
&\mu_j=2w_0\frac{b_j}{b_{j-1}},\quad \nu_j=-\frac{b_j}{b_{j-2}},
\end{aligned} \tag{2.20}$$

$$\tilde{\mu}_j = 2w_1 \frac{b_j}{b_{j-1}}, \quad \tilde{\gamma}_j = -(1-b_{j-1}T_{j-1}(w_0))\,\tilde{\mu}_j \qquad (2 \leqslant j \leqslant s).$$

The increment parameters are

$$c_1 = \frac{c_2}{T_2'(w_0)} \simeq \frac{c_2}{4}, \; c_j = \frac{T_s'(w_0)}{T_s''(w_0)} \frac{T_j''(w_0)}{T_j'(w_0)} \simeq \frac{j^2-1}{s^2-1} \qquad (2 \leqslant j \leqslant s) \tag{2.21}$$

and thus satisfy conditions (2.7). For more details, see [5,6].

3. Convergence analysis: Internal stability.

The remainder of the paper is devoted to the full convergence analysis of the schemes defined by the coefficient sets (2.17), (2.20) when applied to the linear problem class (1.2). Hence, it is supposed that the $m \times m$ constant coefficient matrix M is *symmetric* and *possesses nonpositive eigenvalues* $\lambda(M)$. This covers many linear parabolic problems with time-independent coefficients in the elliptic operator. We stipulate that the RKC method is very well applicable to *nonlinear* parabolic problems, provided the spectrum of the Jacobian $F'(t, U)$ is located in a long, narrow strip around the negative axis of the complex plane and $F'(t, U)$ does not 'deviate too much from a normal matrix'. A nonlinear analysis of the RKC method is likely to become very complicated, if feasible at all. Our convergence analysis for the linear problem gives also insight in handling nonlinear problems.

Throughout, $\|\cdot\|$ denotes the common (appropriately weighted) Euclidean norm in $\mathbf{R}^m$, or the associated spectral matrix norm. Recall that, since M is normal, $\|M\| = \sigma(M)$, σ being the spectral radius. Further, for any polynomial $P(z)$, the spectrum of the matrix polynomial $P(\tau M)$ is the set of values $P(\tau\lambda)$, where $\tau\lambda$ runs through the spectrum of τM; $P(\tau M)$ is also normal and

$$\|P(\tau M)\| = \sigma(P(\tau M)) = \max_{\tau\lambda} |P(\tau\lambda)|.$$

By assumption on M,

$$-\tau\sigma(M) \leqslant \tau\lambda(M) \leqslant \max(\tau\lambda(M)) \leqslant 0.$$

Hence, if we select the stepsize τ and the number of stages s such that *the stability condition* $\tau\sigma(M) \leqslant \beta$ is satisfied, then $\|P_s(\tau M)\| \leqslant 1$ for the genuine stability function P_s of the scheme under consideration.

In the analysis of the RKC scheme (2.1), the notion of internal stability plays an important role. Internal stability is investigated with the *perturbed scheme*

$$\begin{aligned}
&\tilde{Y}_0 = \tilde{U}_n, \\
&\tilde{Y}_1 = \tilde{Y}_0 + \tilde{\mu}_1 \tau \tilde{F}_0 + \tilde{r}_1, \\
&\tilde{Y}_j = \mu_j \tilde{Y}_{j-1} + \nu_j \tilde{Y}_{j-2} + (1-\mu_j-\nu_j)\tilde{Y}_0 + \tilde{\mu}_j \tau \tilde{F}_{j-1} + \tilde{\gamma}_j \tau \tilde{F}_0 + \tilde{r}_j \qquad (2 \leqslant j \leqslant s), \\
&\tilde{U}_{n+1} = \tilde{Y}_s, \quad n = 0, 1, \ldots,
\end{aligned} \tag{3.1}$$

where now

$$\tilde{F}_j \equiv F_j(t_n + c_j\tau, \tilde{Y}_j) = M\tilde{Y}_j + g(t_n + c_j\tau) \tag{3.2}$$

and $\tilde{r}_j$ represents a perturbation introduced at stage j (e.g. round off). Likewise, $\tilde{U}_n$ represents a perturbation of U_n.

Let

$$e_n = \tilde{U}_n - U_n, \quad d_j = \tilde{Y}_j - Y_j \quad (0 \leqslant j \leqslant s) \tag{3.3}$$

represent the errors introduced by these perturbations. Note that, by definition, $d_0 = e_n$ and $e_{n+1} = d_s$. If we subtract the non-perturbed scheme (2.1) from (3.1), we get the *error scheme*

$$\begin{aligned} d_0 &= e_n, \\ d_1 &= d_0 + \tilde{\mu}_1 \tau M d_0 + \tilde{r}_1, \\ d_j &= \mu_j d_{j-1} + \nu_j d_{j-2} + (1-\mu_j-\nu_j) d_0 + \tilde{\mu}_j \tau M d_{j-1} + \tilde{\gamma}_j \tau M d_0 + \tilde{r}_j \qquad (2 \leqslant j \leqslant s), \\ e_{n+1} &= d_s, \quad n = 0, 1, \ldots . \end{aligned} \tag{3.4}$$

Due to the linearity, d_j can be written as

$$d_j = P_j(\tau M) e_n + \sum_{k=1}^{j} Q_{jk}(\tau M) \tilde{r}_k \qquad (1 \leqslant j \leqslant s), \tag{3.5}$$

where P_j are the previously introduced stability functions (cf. (2.10)) and Q_{jk} are new polynomials of degree $j-k$. Of importance is that these new polynomials determine the propagation of all internal perturbations over the stages within one single integration step. We therefore call them *internal stability functions*. In particular, together with the stability function P_s, the internal stability functions $Q_{s1}, \ldots, Q_{sk}$ occurring in the final stage error formula

$$e_{n+1} = P_s(\tau M) e_n + \sum_{k=1}^{s} Q_{sk}(\tau M) \tilde{r}_k, \tag{3.6}$$

determine the error e_{n+1} of U_{n+1}. In order to avoid large contributions $Q_{sk}(\tau M)\tilde{r}_k$, the polynomials $Q_{sk}(z)$ should mimic, in some sense, the behaviour of the stability function $P_s(z)$ for all $z = \tau\lambda(M) \in [-\tau\sigma(M), 0]$. This is particularly important in applications where both the number of stages s and the spectral radius $\tau\sigma(M)$ are large.

One can show that [13]

$$Q_{jk}(z) = \frac{b_j}{b_k} S_{j-k}(w_0 + w_1 z) \qquad (1 \leqslant k \leqslant s,\ k \leqslant j \leqslant s), \tag{3.7}$$

where $S_i(x)$ is the i-th degree Chebyshev polynomial of the second kind (in literature usually denoted by $U_i(x)$ [1]). The error scheme (3.6), then reads

$$e_{n+1} = P_s(\tau M) e_n + \sum_{k=1}^{s} \frac{b_s}{b_k} S_{s-k}(w_0 I + w_1 \tau M) \tilde{r}_k. \tag{3.8}$$

This error scheme gives a complete description of the stability of the RKC schemes under examination. To proceed with it, we briefly recall a few properties of the second kind Chebyshev polynomial [1]. As opposed to $T_i(x)$, $S_i(x)$ is not bounded by ± 1 for $-1 \leqslant x \leqslant 1$. There holds $S_i(\pm 1) = (\pm 1)^i (i+1)$ and $i+1$ is also the maximal value for $-1 \leqslant x \leqslant 1$. On the greater part of this interval, $S_i(x)$ alternates between (approximately) $+1$ and -1. The slope of $S_i(x)$ near $x = 1$ is also larger than that of $T_i(x)$. There holds $S_i'(1) = i(i+1)(i+2)/3$.

The following theorem is proved in [13]:

THEOREM 3.1. *Suppose that τ and s are such that the stability time-step restriction $\tau\sigma(M) \leqslant \beta$ is satisfied. Then the following error bound is valid,*

$$\|e_{n+1}\| \leqslant \|e_n\| + C \sum_{k=1}^{s} (s-k+1) \|\tilde{r}_k\| \leqslant \|e_n\| + \frac{1}{2} s(s+1) C \max_k \|\tilde{r}_k\|, \tag{3.9}$$

where C is a constant of moderate size independent of M, τ and s. □

This result shows that within one full RKC step the accumulation of internal perturbations, such as round-off errors, is *independent of the spectrum of M* as long as $\tau\sigma(M)\leqslant\beta$. As far as rounding errors are concerned, the *quadratic increase with the number of stages* renders no problem. For example, if $s=1000$, which for a serious application is of course a hypothetical value, the local perturbation is at most $\sim 10^6 \max\|\tilde{r}_j\|$. If the machine precision of the computer is about 14 digits, a common value, this local perturbation still leaves 8 digits for accuracy which for *PDEs* is more than enough.

Van der Houwen and Sommeijer (1980) also discuss two different stabilized, explicit RK methods. These two methods possess the same stability function P_s as the RKC method, but show a very strong, spectrum dependent accumulation of internal perturbations (see their numerical experiment). They also conclude that for the RKC scheme the accumulation of internal perturbations is negligible and almost independent of s. Their conclusion is not quite correct since it is based on the assumption that this accumulation is governed by the stability functions P_j, rather than by the internal stability functions Q_{jk}. See [13] for a numerical illustration.

4. Convergence analysis: The local defects

We continue the convergence analysis with the computation of the local defects which arise if an exact PDE solution is inserted in the Runge-Kutta scheme. Consider the semi-discrete PDE problem [9, 8, 10]

$$\dot{u}_h(t)=F(t,u_h(t))+\alpha_h(t),\quad 0<t\leqslant T,\quad u_h(0)\ \text{given},\tag{4.1}$$

that is associated to the ODE system (1.1). Hence, $u_h(t)$ represents an *exact PDE solution* restricted to some space grid parametrized by h, and $\alpha_h(t)$ is the *local space truncation error* that originates from replacing the original PDE problem by its exact, semi-discrete counterpart (4.1). The derivation of the local defects applies to any initial-boundary value problem whose semi-discretization can be put in the generic form (4.1). In particular, in this section F is allowed to be *nonlinear* and merely smoothness assumptions on $u_h(t)$ will be made (cf. the B-convergence theory).

In the previous section we have introduced the perturbed scheme (3.1) for examining the internal propagation of local, arbitrary perturbations $\tilde{r}_j$. If we set in the perturbed scheme $\tilde{Y}_j$ equal to $u_h(t_n+c_j\tau)$, then the $\tilde{r}_j$ represent residual (local) discretization errors, which will be called the *local defects*. These defects will be denoted by r_j in order to distinguish them from the general $\tilde{r}_j$. The local defects are thus defined by

$$\begin{aligned}
&u_h(t_n+c_1\tau)=u_h(t_n)+\tilde{\mu}_1\tau F(t_n,u_h(t_n))+r_1,\\
&u_h(t_n+c_j\tau)=\mu_j u_h(t_n+c_{j-1}\tau)+\nu_j u_h(t_n+c_{j-2}\tau)+(1-\mu_j-\nu_j)u_h(t_n)+\\
&\qquad+\tilde{\mu}_j\tau F(t_n+c_{j-1}\tau,u_h(t_n+c_{j-1}\tau))+\tilde{\gamma}_j\tau F(t_n,u_h(t_n))+r_j\quad(2\leqslant j\leqslant s).
\end{aligned}\tag{4.2}$$

Let $p\in\mathbb{N}$ and assume $u_h\in C^{p+1}[0,T]$. From (4.1) and the Taylor series expansion of $u_h,\dot{u}_h$ at the *intermediate step point* $t_n+c_{j-1}\tau$, it follows that

$$\begin{aligned}
r_j=&\tau\theta_{1j}\dot{u}_h(t_n+c_{j-1}\tau)+\cdots+\tau^p\theta_{pj}u_h^{(p)}(t_n+c_{j-1}\tau)+\tau^{p+1}\rho_j+\\
&+\tau\tilde{\mu}_j\alpha_h(t_n+c_{j-1}\tau)+\tau\tilde{\gamma}_j\alpha_h(t_n)\quad(2\leqslant j\leqslant s),
\end{aligned}\tag{4.3}$$

where the coefficients θ_{qj} and *remainder term* ρ_j are given by

$$\begin{aligned}
\theta_{1j}&=(c_j-c_{j-1})-\nu_j(c_{j-2}-c_{j-1})+(1-\mu_j-\nu_j)c_{j-1}-\tilde{\mu}_j-\tilde{\gamma}_j,\\
\theta_{qj}&=\frac{1}{q!}(c_j-c_{j-1})^q-\frac{1}{q!}\nu_j(c_{j-2}-c_{j-1})^q-\frac{1}{q!}(1-\mu_j-\nu_j)(-c_{j-1})^q-
\end{aligned}\tag{4.4}$$

$$\rho_j = \begin{cases} \dfrac{1}{(q-1)!}\tilde{\gamma}_j(-c_{j-1})^{q-1} & (2 \leqslant q \leqslant p), \\ \dfrac{1}{(p+1)!}(c_j - c_{j-1})^{p+1} u_h^{(p+1)}(*) - \dfrac{1}{(p+1)!}\nu_j(c_{j-2} - c_{j-1})^{p+1} u_h^{(p+1)}(*) - \\ \dfrac{1}{(p+1)!}(1-\mu_j-\nu_j)(-c_{j-1})^{p+1} u_h^{(p+1)}(*) - \dfrac{1}{p!}\tilde{\gamma}_j(-c_{j-1})^p u_h^{(p+1)}(*). \end{cases}$$

In the remainder term, $u_h^{(p+1)}$ is evaluated in various points in (t_n, t_{n+1}). The formulas (4.3), (4.4) also hold for $j=1$ if we set $\mu_1 = 1, c_0 = c_{-1} = 0$ and $\nu_1 = \tilde{\gamma}_1 = 0$.

The coefficients θ_{1j}, θ_{2j} can be written in a more convenient form. We have

$$\theta_{1j} = c_j - \mu_j c_{j-1} - \nu_j c_{j-2} - \tilde{\mu}_j - \tilde{\gamma}_j. \tag{4.5}$$

Relations (2.2) then imply that $\theta_{1j} = 0$ $(1 \leqslant j \leqslant s)$ and thus the contribution of the *temporal errors* to all defects r_j is always $0(\tau^2)$. Furthermore, by inserting the expression for $\tilde{\gamma}_j$ that follows from (2.2) into (4.4), we get

$$\theta_{2j} = \tfrac{1}{2}(c_j^2 - \mu_j c_{j-1}^2 - \nu_j c_{j-2}^2) - \tilde{\mu}_j c_{j-1} \quad (1 \leqslant j \leqslant s). \tag{4.6}$$

For the second order scheme, the conditions (2.6) then imply

$$\theta_{2j} = 0 \ (4 \leqslant j \leqslant s), \quad \theta_{21} = \tfrac{1}{2}c_1^2, \quad \theta_{22} = -\tfrac{1}{2}\mu_2 c_1^2, \quad \theta_{23} = -\tfrac{1}{2}\nu_3 c_1^2. \tag{4.7}$$

In the next section these results will be used to prove convergence. Formula (4.3) will be applied with $p = 1,2$ for the 1-st and 2-nd order schemes, respectively. For the convergence analysis *an upper bound for the remainder terms* ρ_j is needed. For the sake of simplicity, such a bound has been derived in [13] for the undamped schemes ($\epsilon = 0$). For $p = 1$ (RKC1) there holds

$$\|\rho_j\| \leqslant 4s^{-2} \max_{t_n \leqslant t \leqslant t_{n+1}} \|u_h^{(2)}(t)\| \quad (1 \leqslant j \leqslant s). \tag{4.8}$$

while for $p = 2$ (RKC2)

$$\|\rho_j\| \leqslant Cs^{-2} \max_{t_n \leqslant t \leqslant t_{n+1}} \|u_h^{(3)}(t)\| \quad (1 \leqslant j \leqslant s), \tag{4.9}$$

where $C > 0$ is a constant independent of s.

Of interest is to observe that the two bounds (4.8) and (4.9) are proportional to s^{-2}. This means that at each stage the *remainder term* contained in the defect (4.3) diminishes with s^{-2} for increasing s. This is also true for the *spatial error* part in (4.3), i.e.,

$$\|\tilde{\mu}_j \alpha_h(t_n + c_{j-1}\tau) + \tilde{\gamma}_j \alpha_h(t_n)\| \leqslant Cs^{-2} \max_{t_n \leqslant t \leqslant t_{n+1}} \|\alpha_h(t)\|, \tag{4.10}$$

since the coefficients $\tilde{\mu}_j, \tilde{\gamma}_j$ are bounded by Cs^{-2} with $C > 0$ a constant independent of s and j. We have strong numerical evidence that these results are also valid in case of damping ($\epsilon > 0$). However, the derivation of the bounds (4.8), (4.9) then becomes rather technical and lengthy, while no more insight is obtained.

5. Convergence analysis: A bound for the full global error

The results of the two previous sections are now combined so as to derive a bound for the full global error. Hence we again consider the *linear problem class* (1.2) (cf. Section 3) and, for simplicity, restrict ourselves to the *undamped schemes* (cf. Section 4). In our analysis, the time step τ

and the grid distances in space, parametrized by h, are allowed to tend to zero simultaneously and independently of each other. Usually, convergence for explicit methods applied to parabolic equations requires a stepsize restriction $\tau\sigma(M) \leqslant$ const., $\sigma(M) \sim h^{-2}$, due to stability. With the RKC schemes $\tau\sigma(M)$ is allowed to become arbitrarily large, stability being achieved by taking s sufficiently large. This advantage over standard explicit methods is fully justified by our *unconditional convergence* analysis where the assumption $\tau\sigma(M) \leqslant \beta$ is to be interpreted as a condition on s, rather than as a restriction on τ.

Let $e_n = u_h(t_n) - U_n$ be the *full global error*. For these errors we have (cf. (3.6) or (3.8)) the recursion

$$e_{n+1} = P_s(Z)e_n + \sum_{k=1}^{s} Q_{sk}(Z)\, r_k, \tag{5.1}$$

where $Z = \tau M$ and the vectors r_k are the local defects due to discretization. Upper bounds for $\|e_n\|$ will be derived by elaborating this recursion with the help of our estimates of the local defects and our results on internal stability.

In the following, C will denote a positive constant independent of τ, M and s, not necessarily always with the same value.

The RKC1 method ($\epsilon = 0$). Consider the method defined by (2.16)-(2.18) with $\epsilon = 0$. This method was constructed such that the temporal ODE order is one. With temporal ODE order we mean the order obtained from an analysis where the dimension of the problem, and thus the space grid, is fixed. We will show that we have also temporal order one for any value of $\sigma(M)$ and s, hence for any spatial grid refinement, provided $\tau\sigma(M) \leqslant \beta$.

Suppose $u_h \in C^2[0,T]$. From the results of Section 4 (see (4.8), (4.10)) it directly follows that there is a $C > 0$ such that

$$\|r_k\| \leqslant C\tau s^{-2}\Big(\tau \max_{t_n \leqslant t \leqslant t_{n+1}} \|u_h^{(2)}(t)\| + \max_{t_n \leqslant t \leqslant t_{n+1}} \|\alpha_h(t)\|\Big) \quad (1 \leqslant k \leqslant s). \tag{5.2}$$

Using Theorem 3.1, we then immediately obtain the following bound for the global errors:

THEOREM 5.1. *Assume $u_h \in C^2[0,T]$ and $\tau\sigma(M) \leqslant \beta$. Let $U_0 = u_h(0)$. Then the global errors of the undamped RKC1 scheme satisfy*

$$\|e_n\| \leqslant C\Big(\tau \max_{0 \leqslant t \leqslant T} \|u_h^{(2)}(t)\| + \max_{0 \leqslant t \leqslant T} \|\alpha_h(t)\|\Big) \quad (n = 1,2,\ldots;\ n\tau \leqslant T)$$

with a constant $C > 0$ independent of τ, M and s. □

The RKC2 method ($\epsilon = 0$) Consider the method defined by (2.19)-(2.21) with $\epsilon = 0$. This method was constructed such that the temporal ODE order is two. The following theorem presents an error bound which proves that RKC2 has 'almost' order 2 in time for any value of $\sigma(M)$ and s, provided $\tau\sigma(M) \leqslant \beta$ (the proof is somewhat lengthy and therefore omitted; see [13]).

THEOREM 5.2. *Assume $u_h \in C^3[0,T]$ and $\tau\sigma(M) \leqslant \beta$. Let $U_0 = u_h(0)$. Then the global errors of the undamped RKC2 scheme satisfy*

$$\|e_n\| \leqslant C\Big(s^{-3}\tau \max_{0 \leqslant t \leqslant T} \|u_h^{(2)}(t)\| + \tau^2 \max_{0 \leqslant t \leqslant T} \|u_h^{(3)}(t)\| + \max_{0 \leqslant t \leqslant T} \|\alpha_h(t)\|\Big)$$

(for $n = 1,2,\ldots;\ n\tau \leqslant T$) with a constant $C > 0$ independent of τ, M and s. □

Theorems 5.1 and 5.2 prove convergence of RKC1 and RKC2, respectively, irrespective the size of s or $\tau\sigma(M)$. The analysis also shows that the use of many stages within one single step does not adversely affect the accuracy. The temporal error is merely determined by τ and the smoothness of u_t as a function of t. The spatial error is merely determined by the size of the local space truncation error, the common situation. Theorem 5.2 shows that RKC2 is of 'almost' temporal order 2 and that with s large order 2 will be observed. Theorem 5.2 does not reveal the classical order 2 for fixed M (fixed space grid) and s. However, this property can be proved with the above analysis (see [13]).

6. Numerical example

We consider Fisher's equation

$$u_t = u_{xx} + u^2(1-u), \quad 0 \leqslant x,t \leqslant 1, \tag{6.1}$$

with the exact solution $u(x,t) = (1+\exp(v(x-vt)))^{-1}$, $v = \frac{1}{2}\sqrt{2}$. We use this equation to illustrate the convergence behaviour in a non-model situation (A 2D-example is presented in [13]). The second derivative is approximated with 2-nd order central differences on a uniform grid with gridsize h. The schemes are applied with damping. For RKC1 the damping parameter $\epsilon = 0.05$ (see (2.16)) and for RKC2 $\epsilon = 2/13$ (see (2.19)). In the experiment we let $\tau = h$ decrease and s is chosen to satisfy the stability condition $\tau\sigma \leqslant \beta(s)$, while s is taken as small as possible. We put $\sigma = 4h^{-2}+4$ and

$$\begin{aligned} s &= 1 + \text{entier}\,[(1+\tau\sigma/1.90)^{1/2}] \quad \text{for RKC1,} \\ s &= 1 + \text{entier}\,[(1+\tau\sigma/0.65)^{1/2}] \quad \text{for RKC2.} \end{aligned} \tag{6.2}$$

The number 4 in the expression for the spectral radius σ serves as a (conservative) upperbound for the derivative of the inhomogeneous term in (6.1). Note that we select s slightly larger than necessary to satisfy the condition $\tau\sigma \leqslant \beta(s)$.

Table 6.1 lists maximum errors at $t = 1$ for a sequence of $\tau = h$ values. As expected, RKC1 converges with order 1 and RKC2 with order 2. We owe the high level of accuracy to the high degree of smoothness of u.

	RKC1		RKC2	
$(\tau=h)^{-1}$	s	error	s	error
5	4	$.63\ 10^{-4}$	6	$.15\ 10^{-4}$
10	5	$.26\ 10^{-4}$	8	$.25\ 10^{-5}$
20	7	$.13\ 10^{-4}$	12	$.54\ 10^{-6}$
40	10	$.44\ 10^{-5}$	16	$.15\ 10^{-6}$
80	14	$.21\ 10^{-5}$	23	$.33\ 10^{-7}$
160	19	$.99\ 10^{-6}$	32	$.77\ 10^{-8}$
320	26	$.48\ 10^{-6}$	45	$.19\ 10^{-8}$

Table 6.1 Convergence test on Fisher's equation

Acknowledgement We gratefully acknowledge Joke Blom for her assistance in programming the numerical examples.

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of mathematical functions*, National Bureau of Standards, Applied Mathematics Series 55, Washington, 1964.

[2] K. DEKKER and J.G. VERWER, *Stability of Runge-Kutta methods for stiff nonlinear differential equations*, North-Holland, Amsterdam-New York (1984).
[3] E. HAIRER, S.P. NØRSETT and G. WANNER, *Solving ordinary differential equations I: Nonstiff problems*, Springer Series in Computational Mathematics 8, Springer-Verlag, 1987.
[4] P.J. VAN DER HOUWEN, *Explicit Runge-Kutta formulas with increased stability boundaries*, Numer. Math. 20, 149-164 (1972).
[5] P.J. VAN DER HOUWEN, *Construction of integration formulas for initial value problems*, North-Holland, Amsterdam-New York (1977).
[6] P.J. VAN DER HOUWEN and B.P. SOMMEIJER, *On the internal stability of explicit, m-stage Runge-Kutta methods for large m-values*, ZAMM 60, 479-485 (1980).
[7] P.J. VAN DER HOUWEN, *On the time integration of parabolic differential equations*, in: *Numerical Analysis*, Procs. 9th Biennial Conference, Dundee, Scotland, 1981. Lecture Notes in Mathematics 912, G.A. WATSON (ed.), Springer, Berlin, 157-168 (1982).
[8] W.H. HUNDSDORFER and J.G. VERWER, *Stability and convergence of the Peaceman-Rachford ADI method for initial-boundary value problems*, Math. Comp., to appear (1989).
[9] J.M. SANZ-SERNA, J.G. VERWER and W.H. HUNDSDORFER, *Convergence and order reduction of Runge-Kutta schemes applied to evolutionary problems in partial differential equations*, Numer. Math. 50, 405-418 (1987).
[10] J.M. SANZ-SERNA and J.G. VERWER, *Stability and convergence at the PDE/stiff ODE interface*, Appl. Numer. Math. 5, 117-132 (1989).
[11] P.B. SOMMEIJER and P.J. VAN DER HOUWEN, *On the economization of stabilized Runge-Kutta methods with applications to parabolic initial value problems*, ZAMM 61, 105-114 (1981).
[12] B.P. SOMMEIJER and J.G. VERWER, *A performance evaluation of a class of Runge-Kutta-Chebyshev methods for solving semi-discrete parabolic differential equations*, Report NW 91/80, Centre for Mathematics and Computer Science (CWI), Amsterdam (1980).
[13] J.G. VERWER, W.H. HUNDSDORFER and B.P. SOMMEIJER, *Convergence properties of the Runge-Kutta-Chebyshev method*, Report NM-R8907, Centre for Mathematics and Computer Science (CWI), Amsterdam (1989).