

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/258240259>

# Accuracy of Runge–Kutta Methods Applied to Stiff Problems

Article in Computational Mathematics and Mathematical Physics · September 2003

CITATIONS

20

READS

1,541

1 author:



[Leonid Skvortsov](#)

Bauman Moscow State Technical University

37 PUBLICATIONS 265 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:



SimInTech [View project](#)

# Accuracy of Runge–Kutta Methods Applied to Stiff Problems

L. M. Skvortsov

Bauman State Technical University, Vtoraya Baumanskaya ul. 5, Moscow, 107005 Russia

e-mail: v401@sm.bmstu.ru

Received April 29, 2002

**Abstract**—The behavior of the error in the numerical solution of stiff problems by Runge–Kutta methods is investigated for simple model equations. Error functions are proposed that demonstrate the dependence of the error on the problem stiffness and the coefficients of the method. It is shown that the minimization of the error functions allows one to improve the accuracy of solving nonlinear stiff problems by explicit and implicit Runge–Kutta methods.

## 1. INTRODUCTION

We consider the Cauchy problem for the system of ordinary differential equations

$$y' = f(x, y), \quad y(x_0) = y_0.$$

One step of the solution of this problem by the  $s$ -stage Runge–Kutta method is described by the formulas

$$y_1 = y_0 + h \sum_{i=1}^s b_i k_i, \quad (1.1)$$

$$k_i = f\left(x_0 + c_i h, y_0 + h \sum_{j=1}^s a_{ij} k_j\right), \quad i = 1, 2, \dots, s.$$

Every particular method can be represented by the table of coefficients

$$\begin{array}{c|c} c & A \\ \hline b^T & \begin{array}{c} c_1 \mid a_{11} \dots a_{1s} \\ \dots \mid \dots \\ c_s \mid a_{s1} \dots a_{ss} \\ \hline b_1 \dots b_s \end{array} \end{array}$$

According to the classical views, the accuracy of the Runge–Kutta method is determined by the approximation order and the values of the error coefficients [1]. As a consequence, the method parameters are usually designed so as to ensure the desired order of accuracy and minimize the error coefficients. However, for stiff problems, the classical notions are not applicable. For example, the determination of the approximation order is based on the asymptotic behavior of the error of the numerical solution at  $h \rightarrow 0$ . However, in this case, we have  $h\lambda_i \rightarrow 0$  for all eigenvalues  $\lambda_i$  of the Jacobi matrix; i.e., the problem ceases to be stiff. In the stiff case, the behavior of the error should be analyzed at large values of  $h\lambda_i$  for the stiff spectrum.

When solving stiff problems by implicit Runge–Kutta methods, the actual order can be less than the classical one. The simplest model that explains the reduced order phenomenon is the Prothero–Robinson equation (see [2])

$$y' = \lambda(y - \varphi(x)) + \varphi'(x), \quad y(t_0) = \varphi_0 = \varphi(x_0), \quad \operatorname{Re} \lambda \leq 0, \quad (1.2)$$

with the solution  $y(x) = \varphi(x)$ . In [2–4], the error of the numerical solution of this equation was investigated for  $h \rightarrow 0$  and  $z = h\lambda \rightarrow \infty$ . Under such assumptions, the asymptotic behavior of the local error is given by the formula

$$\delta = O(z^{-k} h^{q+1}), \quad (1.3)$$

where  $q$  is the stage order of the method and  $k$  is the damping order of the error as  $z \rightarrow \infty$ . For stiffly accurate methods,  $k > 0$ , and then  $\delta \rightarrow 0$  as  $z \rightarrow \infty$ .

Recall that the stage order of the Runge-Kutta method is defined as the maximal integer  $q$  for which the following equalities hold:

$$c^i - iAc^{i-1} = [0, \dots, 0]^T, \quad 1 - ib^T c^{i-1} = 0, \quad i = 1, 2, \dots, q. \quad (1.4)$$

Here and in what follows, raising a vector to a power is interpreted componentwise. A method is called stiffly accurate if the last row of the matrix  $A$  coincides with  $b^T$ .

Stiff problems are conventionally solved by implicit  $A(\alpha)$ -stable methods for which the stability function satisfies the inequality  $|R(\infty)| < 1$ . In this case, the asymptotic behavior of the global error coincides with that of the local error and is also expressed by formula (1.3). Studies of the behavior of the error of the numerical solution to the Prothero-Robinson equation at  $h \rightarrow 0$  and  $h\lambda \rightarrow \infty$  revealed the importance of the stage order and stiff accuracy concepts for implicit Runge-Kutta methods.

The asymptotic estimate (1.3) can be used if  $|h\lambda_i|$  are sufficiently large for the entire stiff spectrum. However, this estimate is inapplicable in the opposite case, i.e., for moderately stiff problems and problems with a uniformly filled spectrum. Note that solving such problems with an improved accuracy is very difficult.

The error of the numerical solution to Eq. (1.2) as a function of  $z = h\lambda$  was investigated in [5] for explicit Runge-Kutta methods and in [6] for implicit ones. The local error was obtained in the form

$$\varphi(x_0 + h) - y_1 = \sum_{i=1}^{\infty} e_i(z) \frac{d^i \varphi(x_0) h^i}{dx_i i!}, \quad (1.5)$$

where  $e_i(z)$ , which are called error functions, are determined by the formulas

$$e_i(z) = zb^T(I - zA)^{-1}(c^i - iAc^{i-1}) + (1 - ib^T c^{i-1}), \quad i = 1, 2, \dots \quad (1.6)$$

If the stage order of a method is  $q$ , then the first  $q$  terms in expansion (1.5) vanish; the leading term of the error is proportional to  $e_{q+1}(z)$  and is equal to the corresponding term in the Taylor expansion of  $\varphi(x)$ . Equation (1.2) provides a model of the behavior of stiff components when the Cauchy problem is solved. Thus, we can conclude that the accuracy of these components is determined by the stage order and the error functions of the method.

In this paper, we develop and generalize the results presented in [5, 6]. It is shown that the minimization of the error functions yields an effect similar to increasing the stage order of the Runge-Kutta method.

## 2. MODEL EQUATIONS

The error functions (1.6) were obtained as a result of the investigation of the numerical solution of the Prothero-Robinson equation. The question arises as to whether it is justifiable to use the results obtained for the particular equation (1.2) in the more general case of nonlinear stiff differential equations. To answer this question, we consider simple equations that model the behavior of various components of the error. We begin with nonstiff problems.

The determination of the approximation order and coefficients of the error of Runge-Kutta methods is reduced to the comparison of the Taylor series for the exact and numerical solutions (the derivation of the order conditions can be found in [1, 7]). For the pictorial representation of the elementary differentials obtained in the expansion in a series, rooted trees are used. There exists a one-to-one correspondence between the set of elementary differentials and the set of trees.

The construction of model equations is based on the following fact (see [1]): for any tree, a system of equations can be constructed such that the Taylor expansion of one of the variables involves only a single elementary differential corresponding to this tree. The system of equations is constructed as follows. The tree consisting of a single vertex is assigned the variable described by the equation

$$y_1' = 1, \quad y_1(0) = 0.$$

The tree  $t_{ij}$  (where  $i$  is the order of the tree, i.e., the number of its vertices, and  $j$  is the serial number of this tree among all trees of order  $i$ ) is assigned the variable  $y_{ij}$ , and the equations are recurrently defined by the formula (see [1])

$$y_{\text{predecessor}}' = \prod y_{\text{descendants}},$$

where the descendents are the trees obtained by eliminating the root vertex of the predecessor tree along with all branches that are incident to this vertex. The initial values for all variables are set to zero. The equations thus obtained and their solutions at the first step (the exact solution  $y(h)$  and the numerical solution  $\tilde{y}(h)$ ) for the trees up to the fourth order inclusive are presented in Table 1, where  $\mathbf{1} = [1, \dots, 1]^T$  and the dot in one of the formulas denotes the componentwise multiplication of vectors. The coefficients of the error are calculated as the relative error of the numerical solution at the first step:

$$e(t_{ij}) = \frac{y_{ij}(h) - \tilde{y}_{ij}(h)}{y_{ij}(h)}. \quad (2.1)$$

Similarly, one can set up model equations for differential algebraic problems of index 1 (see [4]):

$$y' = f(y, u), \quad 0 = g(y, u), \quad (2.2)$$

where the matrix  $\partial g / \partial u$  is invertible. In this case, trees with vertices of two types are used (dot vertices correspond to differential equations and circle vertices correspond to algebraic ones). In [4], trees of this type were used for Rosenbrock methods. For Runge–Kutta methods, it is sufficient to consider trees consisting of a single circle-type vertex (see [4]). Table 2 presents model algebraic equations and their solutions for problems of index 1.

The differential algebraic problem (2.2) can be used to construct a stiff singular perturbed system. Similarly, on the basis of model algebraic equations, one can construct stiff model equations. To make the exact solution independent of stiffness, we add the exact value of the derivative to the right-hand side. Some model equations for stiff problems are presented in Table 3. The error functions are determined by analogy with (2.1):

$$e_{ij}(z) = \frac{u_{ij}(h) - \tilde{u}_{ij}(h)}{u_{ij}(h)},$$

where  $u_{ij}(h)$  and  $\tilde{u}_{ij}(h)$  are the exact and numerical solutions, respectively, and  $z = h\lambda$ . Note that  $e_{i1}(z) = e_i(z)$ , where  $e_i(z)$  are the error functions (1.6) obtained for the Prothero–Robinson equation.

If the stage order of a method is greater than unity, then  $e_{i1}(z) \equiv 0$  for  $i \leq q$ , and all functions  $e_{q+1,j}(z)$  are equal to  $e_{q+1}(z)$ . In this case, the number of different higher order error functions also decreases. For example, for  $q = 2$ , we have

$$e_{31}(z) = e_{32}(z), \quad e_{41}(z) = e_{42}(z) = e_{44}(z), \quad e_{43}(z) = e_{45}(z).$$

When a model equation in the variable  $u_{q+1,j}$  is solved by a method with the stage order  $q$  and a constant

**Table 1**









Graph	Equation	$y(h)$	$\tilde{y}(h)$
	$y'_1 = 1$	$h$	$b^t \mathbf{1} h$
	$y'_{21} = y_1$	$h^2/2$	$b^t c h^2$
	$y'_{31} = y_1^2$	$h^3/3$	$b^t c^2 h^3$
	$y'_{32} = y_{21}$	$h^3/6$	$b^t A c h^3$
	$y'_{41} = y_1^3$	$h^4/4$	$b^t c^3 h^4$
	$y'_{42} = y_1 y_{21}$	$h^4/8$	$b^t (c \cdot A c) h^4$
	$y'_{43} = y_{31}$	$h^4/12$	$b^t A c^2 h^4$
	$y'_{44} = y_{32}$	$h^4/24$	$b^t A^2 c h^4$

Table 2









Graph	Equation	$u(h)$	$\tilde{u}(h)$
	$0 = u_{21} - y_1^2$	$h^2$	$b^t A^{-1} c^2 h^2$
	$0 = u_{31} - y_1^3$	$h^3$	$b^t A^{-1} c^3 h^2$
	$0 = u_{32} - y_1 y_{21}$	$h^3/2$	$b^t A^{-1} (c \cdot Ac) h^3$
	$0 = u_{41} - y_1^4$	$h^4$	$b^t A^{-1} c^4 h^4$
	$0 = u_{42} - y_1^2 y_{21}$	$h^4/2$	$b^t A^{-1} (c^2 \cdot Ac) h^4$
	$0 = u_{43} - y_1 y_{31}$	$h^4/3$	$b^t A^{-1} (c \cdot Ac^2) h^4$
	$0 = u_{44} - y_{21}^2$	$h^4/4$	$b^t A^{-1} (Ac^2) h^4$
	$0 = u_{45} - y_1 y_{32}$	$h^4/6$	$b^t A^{-1} (c \cdot A^2 c) h^4$

Table 3

Equation	Error function
$u'_{21} = \lambda(u_{21} - y_1^2) + 2x$	$zb^t(I - zA)^{-1}(c^2 - 2Ac) + (1 - 2b^t c)$
$u'_{31} = \lambda(u_{31} - y_1^3) + 3x^2$	$zb^t(I - zA)^{-1}(c^3 - 3Ac^2) + (1 - 3b^t c^2)$
$u'_{32} = \lambda(u_{32} - y_1 y_{21}) + (3/2)x^2$	$zb^t(I - zA)^{-1}(2c \cdot Ac - 3Ac^2) + (1 - 3b^t c^2)$
$u'_{41} = \lambda(u_{41} - y_1^4) + 4x^3$	$zb^t(I - zA)^{-1}(c^4 - 4Ac^3) + (1 - 4b^t c^3)$
$u'_{43} = \lambda(u_{43} - y_1 y_{31}) + (4/3)x^3$	$zb^t(I - zA)^{-1}(3c \cdot Ac^2 - 4Ac^3) + (1 - 4b^t c^3)$

step, the global error is given by the formula

$$\Delta_{n+1} = R(z)\Delta_n + \delta,$$

where  $R(z)$  is the stability function and  $\delta$  is the local error, which is the same for all steps. When  $n \rightarrow \infty$  and  $|R(z)| < 1$ , we have  $\Delta_\infty = \delta/[1 - R(z)]$ . Therefore, along with (local) error functions  $e_{ij}(z)$ , we introduce the global error functions

$$E_{ij}(z) = e_{ij}(z)/[1 - R(z)], \quad (2.3)$$

which account for error accumulation. We will use the notation  $E_i(z) = E_{i1}(z)$ .

Thus, we have derived equations that model the behavior of nonstiff (Table 1) and stiff (Table 3) components of the error. According to these models and the classical theory, nonstiff components of the global error are proportional to  $h^p$  and to the error coefficients  $e(t_{p+1,j})$ , where  $p$  is the order of the method. Stiff components are proportional to  $h^{q+1}$  and to the values of the error function  $E_{i+1}(z)$  at the points  $z_i = h\lambda_i$  of the stiff spectrum. If  $q < p - 1$ , then the contribution of the stiff components to the total error is significant, and the free parameters of the method should be determined by minimizing functions (2.3) for  $i = q + 1, \dots, p - 1$ . Experiments with test problems showed that methods based on this idea have an improved accuracy when solving stiff nonlinear problems.

In this paper, we present results of the numerical solution of the Kaps problem

$$\begin{aligned} y_1' &= -(\mu + 2)y_1 + \mu y_2^2, & y_1(0) &= 1, \\ y_2' &= y_1 - y_2 - y_2^2, & y_2(0) &= 1, & 0 \leq x \leq 1. \end{aligned} \quad (2.4)$$

This problem has the smooth solution  $y_1(x) = \exp(-2x)$ ,  $y_2(x) = \exp(-x)$ , which is independent of the stiffness parameter  $\mu$ . For large  $\mu$ , the maximum (in absolute value) eigenvalue varies insignificantly on the integration interval and is approximately equal to  $-\mu$ , which makes it possible to investigate the numerical integration error as a function of  $h\mu \approx -h\lambda_{\max}$ . In [3], the Kaps problem was used to investigate the decreased order phenomenon.

### 3. IMPLICIT METHODS

Among implicit Runge–Kutta methods, diagonally implicit Runge–Kutta (DIRK) methods are the simplest to implement. In these methods, matrix  $A$  has the lower triangular form. Often, it is also required that all the diagonal elements of  $A$  be equal, which makes it possible to perform a single LU factorization per integration step. Such methods are called singly diagonally implicit (SDIRK). SDIRK methods can have only the first stage order, which limits their accuracy as applied to stiff problems. For this reason, among the DIRK methods of order three or higher, the so-called FSAL-DIRK methods are most often used (see [6–8]). In these methods, the first stage is explicit and coincides with the last stage of the previous step; hence the acronym FSAL, which stands for *first same as last*. Due to this property, such methods can have the second stage order. FSAL-DIRK methods are as easily implemented as SDIRK methods (the implementation is presented in detail in [6]).

Consider fourth-order DIRK methods SDIRK4 (see [4]) and two FSAL-DIRK methods defined by the following tables of coefficients:

0	0				
1/2	1/4	1/4			
4/5	31/100	6/25	1/4		
1	21/64	7/24	25/192	1/4	
2/15	-109/675	77/225	-55/108	143/675	1/4
1	1/96	4/11	25/96	-7/39	675/2288
	1/96	4/11	25/96	-7/39	675/2288

(3.1)

0	0				
1/2	1/4	1/4			
1/4	1/16	-1/16	1/4		
3/4	1/16	-1/16	1/2	1/4	
1	-9/62	-77/124	143/124	45/124	1/4
1	7/90	2/15	16/45	16/45	-31/180
	7/90	2/15	16/45	16/45	-31/180

(3.2)

All three methods are  $L$ -stable and stiffly accurate, and they have the same stability function. The free parameters of SDIRK4 were determined in [4] from the minimization condition of the error coefficients. Methods (3.1) and (3.2) have the second stage order, and the free parameters of method (3.1) were also determined from the minimization condition of the error coefficients. Method (3.2) was proposed in [6]. Its parameters were chosen so as to ensure fast damping of the error function  $e_3(z)$  as  $z \rightarrow \infty$  (as  $O(z^{-2})$ ) and its small absolute values in the left half-plane. Accuracy characteristics of these methods are presented in Table 4, where  $\|e(t_5)\|_2$  is the Euclidean norm of the fifth-order error coefficients and  $\|E_{q+1}(z)\|_\infty$  is the maximum of the error function absolute value in the left half-plane.

The absolute values of the error functions ( $E_2(z)$  for SDIRK (solid line),  $E_3(z)$  for FDIRK4a (dotted line), and FDIRK4b (dashed line)) for negative values of the argument are shown in Fig. 1. The experimental results obtained when solving problem (2.1) with a step of  $h = 1/20$  and various values of  $\mu$  are shown in

Table 4

Method	Formula	$\ e(t_5)\ _2$	$\ E_{q+1}(z)\ _\infty$
SDIRK4	(6.16) @ [4]	0.134	0.155
FDIRK4a	(3.1)	0.144	0.0405
FDIRK4b	(3.2)	0.233	0.00328

Fig. 2, where  $\varepsilon_{20}$  is the maximal relative error on the entire interval when it is covered in 20 steps and  $p_r$  is the actual order estimated by the formula

$$p_r = \frac{\varepsilon_{N-1} - \varepsilon_{N+1}}{2\varepsilon_N} N$$

for  $N = 20$ . This formula is obtained as a result of finite-difference approximation of the derivative in the expression  $p_r = \varepsilon'(h)h/\varepsilon(h)$ , where  $\varepsilon(h)$  is the global error considered as a function of the step size.

The comparison of the results yielded by the SDIRK and FSAL-Dirk methods provides a strong evidence of the advantages of a higher stage order. The behavior of the error and actual order is completely explained by the behavior of the error function and the influence of the nonstiff error component, which dominates at small and large values of  $h\mu$ . Experiments showed that the minimization of the error function yields an effect similar to that of an increased stage order. When constructing fifth-order FSAL-Dirk methods, it is reasonable to minimize the error functions  $E_{41}(z)$  and  $E_{43}(z)$  as well.

#### 4. CLASSICAL EXPLICIT METHODS

The stage order of an explicit Runge–Kutta method cannot exceed unity; however, one can construct explicit schemes that ensure approximate equalities (1.4) for  $q > 1$ . It was shown in [5] that methods of this type have an improved accuracy when solving moderately stiff problems; such methods can be constructed on the basis of well-known methods. Consider the Merson fourth-order method and its variants.

The first variant is obtained by considering the abscissa  $c_2$  as a free parameter (in the Merson method,  $c_2 = 1/3$ ). As a result, we obtain the following table of coefficients:

$$\begin{array}{c|ccc}
 0 & & & \\
 c_2 & c_2 & & \\
 \frac{1}{3} & \frac{1}{2} - \frac{1}{18c_2} & \frac{1}{18c_2} & \\
 \frac{1}{2} & \frac{1}{8} & 0 & \frac{3}{8} \\
 1 & \frac{1}{2} & 0 & -\frac{3}{2} \quad 2 \\
 \hline
 & \frac{1}{6} & 0 & 0 \quad \frac{2}{3} \quad \frac{1}{6}
 \end{array} \quad (4.1)$$

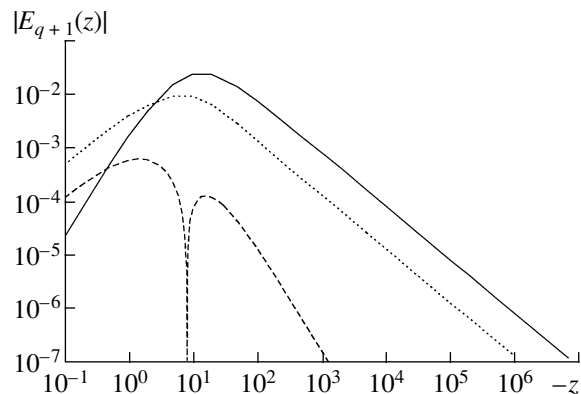


Fig. 1.

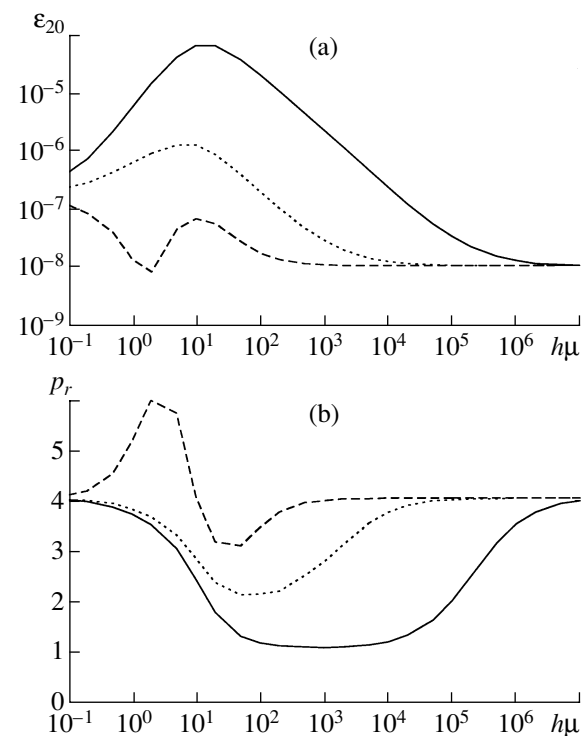


Fig. 2.

For this method,  $c_2 - 2Ac = [0, c_2^2, 0, 0, 0]^t$  and  $e_2(z) = \frac{1}{144} c_2 z^4$ . Formally, method (4.1) has the first stage order; however, for small  $c_2$ , it has properties characteristic of second stage order methods.

The approach suggested here makes it possible to construct explicit schemes having properties of even higher stage orders. We modify method (4.1) by considering  $c_3$  as an additional free parameter. The corresponding coefficients are as follows:

$$\begin{aligned} c_4 &= 1/2, \quad c_5 = 1, \quad b_1 = b_5 = 1/6, \quad b_2 = b_3 = 0, \quad b_4 = 2/3, \\ a_{32} &= \frac{c_3(c_3 - c_2)}{2c_2(1 - 3c_2)}, \quad a_{31} = c_3 - a_{32}, \\ a_{42} &= \frac{3c_3 - 1}{24c_2(c_3 - c_2)}, \quad a_{43} = \frac{1 - 3c_2}{24c_3(c_3 - c_2)}, \quad a_{41} = \frac{1}{2} - a_{42} - a_{43}, \\ a_{52} &= -4a_{42}, \quad a_{53} = -4a_{43}, \quad a_{54} = 2, \quad a_{51} = 1 - a_{52} - a_{53} - a_{54}. \end{aligned} \quad (4.2)$$

For this method, the error functions have the form

$$e_2(z) = \frac{1}{144} c_2 z^4, \quad e_{31}(z) = \frac{1}{144} z^3 (2c_3 - c_2 - 6c_2 c_3 + c_2^2 z), \quad e_{32}(z) = \frac{1}{144} z^3 (2c_3 - 3c_2).$$

For small  $c_2$ , this method has the “almost second” stage order, and, for small  $c_2$  and  $c_3$ , its stage order is “almost third.” Note that the stability function and the fifth-order error coefficients of methods (4.1) and (4.2) are independent of  $c_2$  and  $c_3$ .

Let us demonstrate how these parameters affect the accuracy of solving problem (2.4). We use the step  $h = 1/20$  and two values of  $\mu$  that are characteristic of nonstiff problems ( $h\mu = 0.1$ ) and moderately stiff problems ( $h\mu = 2$ ). The results are presented in Table 5. At  $c_2 = c_3 = 1/3$ , the classical Merson method was used. Table 6 presents the results obtained by higher order explicit methods with the coefficients presented in [1, 4]. These results show that when moderately stiff problems are solved by explicit methods, the minimization of error functions (which is actually equivalent to increasing the stage order) is more effective than increasing the classical order.

## 5. EXPLICIT METHODS WITH AN EXTENDED STABILITY REGION

Explicit methods with the stability region extended along the real axis can be very effective for solving certain stiff problems [4, 9, 10]. As a rule, these are first- or second-order methods; however, recent results show that higher order methods can have certain advantages (see [4]). We have already seen the importance of increasing the stage order for constructing methods with an improved accuracy. Now, we show that this is also true for explicit methods with an extended stability region.

Consider the construction of such methods on the bases of Merson-type methods. This name is used for explicit schemes satisfying the following conditions: (1)  $c_s = 1$ ; (2)  $R(z) - \hat{R}(z) = d_s z^s$ ,  $d_s \neq 0$ , where  $\hat{R}(z)$  is the stability function of the last internal stage (embedded method); (3) the order of the embedded method is not less than  $p - 1$ , where  $p$  is the order of the main method. The simplest method of this type is the Euler method with correction; more complicated examples are the Merson method and its modifications (4.1) and

**Table 5**

$c_2$	$c_3$	$\varepsilon_{20}(h\mu = 0.1)$	$\varepsilon_{20}(h\mu = 2)$
1/3	1/3	$1.51 \times 10^{-7}$	$1.51 \times 10^{-4}$
1/30	1/3	$2.10 \times 10^{-7}$	$2.10 \times 10^{-5}$
1/300	1/3	$2.16 \times 10^{-7}$	$8.06 \times 10^{-6}$
1/3000	1/3	$2.17 \times 10^{-7}$	$6.76 \times 10^{-6}$
1/3000	1/30	$2.41 \times 10^{-7}$	$8.50 \times 10^{-7}$
1/3000	1/300	$2.43 \times 10^{-7}$	$2.59 \times 10^{-7}$
1/3000	1/2000	$2.43 \times 10^{-7}$	$2.03 \times 10^{-7}$



Table 6

Method	Order	$\varepsilon_{20}(h\mu = 0.1)$	$\varepsilon_{20}(h\mu = 2)$
Fehlberg's	5	$9.62 \times 10^{-8}$	$7.72 \times 10^{-5}$
Dormand-Prince's	5	$4.05 \times 10^{-8}$	$9.84 \times 10^{-5}$
Higham-Hall's	5	$6.42 \times 10^{-8}$	$3.96 \times 10^{-5}$
Butcher's	6	$2.18 \times 10^{-8}$	$3.52 \times 10^{-4}$
Fehlberg's	8	$1.27 \times 10^{-12}$	$1.11 \times 10^{-6}$

(4.2). The error estimate for Merson-type methods is  $e = y_1 - \hat{y}_1$ , where  $\hat{y}_1$  is the result of the last internal stage. For convenience, we define  $Y_0 = y_1$  and  $\hat{Y}_0 = \hat{y}_1$ . In the linear approximation, we have the relations

$$Y'_0 - \hat{Y}'_0 \approx J(Y_0 - \hat{Y}_0) \approx d_s(hJ)^s y'_0, \quad (5.1)$$

where  $J$  is the Jacobi matrix, which allows one to use the power method to obtain an estimate of the largest (in absolute value) eigenvalue.

Let

$$R_0(z) = R(z) = 1 + d_1 z + \dots + d_{s-1} z^{s-1} + d_s z^s, \quad \hat{R}_0(z) = \hat{R}(z) = 1 + d_1 z + \dots + d_{s-1} z^{s-1}, \quad (5.2)$$

where  $d_i = 1/i!$  for  $i \leq p$ . Set up the polynomial

$$R_r(z) = 1 + d_1 z + d_2 z^2 + \dots + d_s z^s + \dots + d_{s+r} z^{s+r}, \quad (5.3)$$

with the coefficients  $d_1, \dots, d_s$  coinciding with the coefficients of the polynomial  $R_0(z)$ ; the other coefficients are chosen so as to ensure the desired size of the stability region. This polynomial can be uniquely determined by  $r$  negative roots  $z_1, \dots, z_r$ ; then, it can be constructed by the recurrent formulas

$$\begin{aligned} R_i(z) &= R_{i-1}(z) + \alpha_i z [R_{i-1}(z) - \hat{R}_{i-1}(z)], \\ \hat{R}_i(z) &= \hat{R}_{i-1}(z) + \beta_i [R_{i-1}(z) - \hat{R}_{i-1}(z)], \quad i = 1, 2, \dots, r, \end{aligned} \quad (5.4)$$

where the coefficients  $\alpha_i$  and  $\beta_i$  are chosen from the conditions

$$R_i(z_i) = 0, \quad \hat{R}_i(z_i) = 0, \quad i = 1, 2, \dots, r.$$

Polynomials (5.4) can be also represented in the form

$$\begin{aligned} R_i(z) &= (1 + d_1 z + \dots + d_{s-1} z^{s-1} + d_s z^s) \prod_{j=1}^i (1 + \gamma_j z), \\ \hat{R}_i(z) &= (1 + d_1 z + \dots + d_{s-1} z^{s-1}) \prod_{j=1}^i (1 + \gamma_j z), \quad \gamma_j = -z_j^{-1}, \end{aligned}$$

whence we derive the following recurrent relations for determining  $\alpha_i$  and  $\beta_i$ :

$$\begin{aligned} \beta_i &= d_{s-1,i} \gamma_i / d_{s,i-1}, \quad \alpha_i = (1 - \beta_i) \gamma_i, \\ d_{1i} &= d_{1,i-1} - \gamma_i, \quad d_{ji} = d_{j,i-1} - \gamma_i d_{j-1,i}, \quad i = 1, 2, \dots, r, \quad j = 2, 3, \dots, s. \end{aligned} \quad (5.5)$$

If the estimate of the largest (in absolute value) eigenvalue obtained on the basis of (5.1) satisfies the stability condition, then  $Y_0$  can be taken as the solution at the current step; otherwise, additional stages must be used to extend the stability region. Additional stages are performed in accordance with (5.4), whence

$$\begin{aligned} Y_i &= Y_{i-1} + \alpha_i h (Y'_{i-1} - \hat{Y}'_{i-1}), \quad Y'_i = f(x_i, Y_i), \\ \hat{Y}_i &= \hat{Y}_{i-1} + \beta_i (Y_{i-1} - \hat{Y}_{i-1}), \quad \hat{Y}'_i = \hat{Y}'_{i-1} + \beta_i (Y'_{i-1} - \hat{Y}'_{i-1}). \end{aligned} \quad (5.6)$$

As a result of performing  $r$  stages by formulas (5.5) and (5.6), we obtain a solution  $y_1 = Y_r$  at one step, which has the desired stability function (5.3) and the error estimate  $e = Y_r - \hat{Y}_r$ . A drawback of the method of

extending the stability region is the instability of internal stages; thus, it can be recommended only for moderately stiff problems.

For our experiments, we used methods (4.1) and (4.2) with  $r = 5$ . Requiring that the polynomial values be within  $-0.9$  and  $0.9$  on the real axis in the stability interval, we obtain the following values for its first five negative roots:

$$z_1 = -29.1870, \quad z_2 = -27.5066, \quad z_3 = -24.2963, \quad z_4 = -19.8304, \quad z_5 = -14.4221$$

(they were found by the search optimization technique). The stability polynomial and the error function  $E_2(z)$  for  $c_2 = c_3 = 1/3$  are shown in Fig. 3. As for methods (4.1) and (4.2), the decrease in  $c_2$  and  $c_3$  results in decreasing error functions. The results of experiments with problem (2.4) (the dependence of the error of solution on the stiffness for various  $c_2$  and  $c_3$ ) are presented in Fig. 4 (the solid line for  $c_2 = c_3 = 1/3$ , the dotted line for  $c_2 = 1/3000$  and  $c_3 = 1/3$ , and the dashed line for  $c_2 = 1/3000$  and  $c_3 = 1/2000$ ).

## 6. EXPLICIT ADAPTIVE METHODS

Another technique used to construct explicit methods for stiff problems is based on deriving estimates of the largest (in absolute values) eigenvalues and the subsequent stabilization of the stability function at the resulting points of a stiff spectrum. Among such techniques are the exponential and fractional rational methods with a componentwise implementation, which were suggested in [11, 12] and other studies. Sometimes, explicit nonlinear methods solved stiff problems using a large step without losing stability; however, the accuracy of integrating slow components was unacceptable in these cases (see [13]). The reason is that the actual order decreases down to zero for stiff problems, which was demonstrated in [14, 15] using the Kaps problem as an example. The elimination of this drawback made it possible to develop efficient explicit adaptive one-step and multistep methods [14, 15]. These methods are competitive with implicit ones when applied to certain stiff problem; among those problems are five of the twelve benchmarks used in [4] to test various methods (VDPOL, ROBER, OREGO, HIRE, and CUSP).

The technique proposed in [14, 15] for constructing one-step methods does not allow one to obtain a classical order greater than three; moreover, for stiff problems, the actual order does not exceed two. In this respect, multistep adaptive methods are more interesting; they were used to develop an efficient computer program that implements a variable step and order algorithm. One-step adaptive methods with improved accuracy can be constructed on the basis of Merson-type schemes. In this case, relations (5.1) are used to obtain a vector of componentwise estimates for the largest eigenvalue:

$$z_1 = h\lambda_1 = \frac{h(Y'_0 - \hat{Y}'_0)}{Y_0 - \hat{Y}_0} \quad (6.1)$$

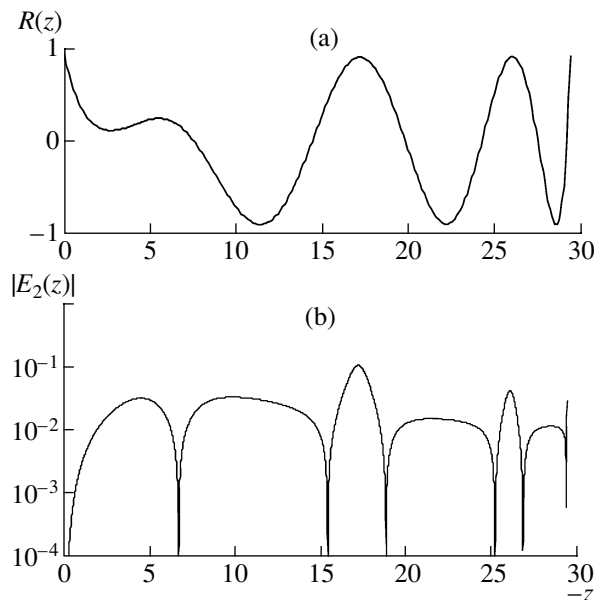


Fig. 3.

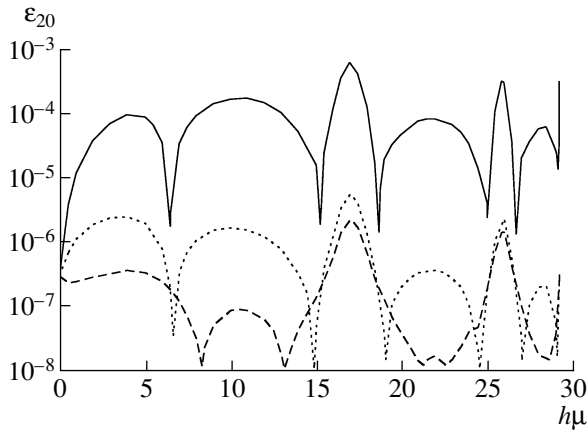


Fig. 4.

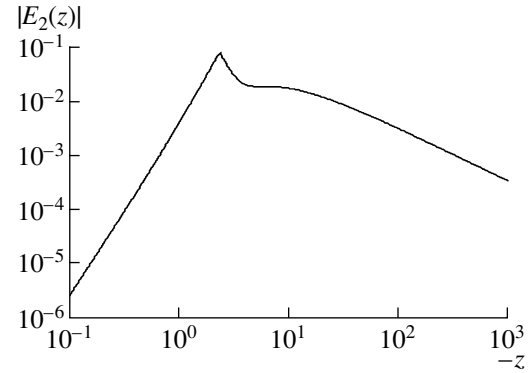


Fig. 5.

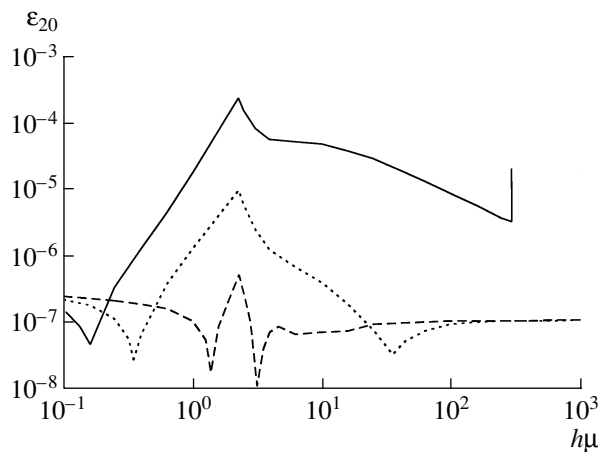


Fig. 6.

(here and in what follows, all operations on vectors are performed componentwise). The final formula for the integration step has the form

$$y_1 = \hat{Y}_0 + \alpha(Y_0 - \hat{Y}_0), \quad (6.2)$$

where the vector of adjustable parameters  $\alpha$  is chosen from the minimum error condition for nonstiff components and the condition

$$\hat{R}_0(z_1) + \alpha[R_0(z_1) - \hat{R}_0(z_1)] = 0$$

for stiff components.

For the method constructed on the basis of schemes (4.1) and (4.2), the componentwise tuning of the vector  $\alpha$  in (6.2) is performed by the formula

$$\alpha = \begin{cases} 1, & z_1 \geq -2.416, \\ -z_1^{-1}(6 + 24z_1^{-1} + 72z_1^{-2} + 144z_1^{-3} + 144z_1^{-4}), & z_1 < -2.416, \end{cases}$$

which ensures that  $\alpha$  continuously depends on  $z_1$ . For large negative values of  $z_1$ , this method is very sensitive to errors in the eigenvalue estimates (6.1); because of this, we were unable to obtain a solution to problem (2.4) with the step  $h = 1/20$  for  $h\mu > 50$ . To eliminate this drawback, we performed additional stages (no more than two), which extend the stability region in the neighborhood of  $z_1$ . This enabled us to obtain a solution for  $h\mu$  up to 1000. The error function of this method (for  $c_2 = c_3 = 1/3$ ) is shown in Fig. 5. The

results of experiments for various  $c_2$  and  $c_3$  are presented in Fig. 6 (the line styles are the same as in Fig. 4). For  $c_2 = c_3 = 1/3$ , a solution was obtained only for  $h\mu < 300$ .

In this paper, we presented the results of solving only one test problem (2.4). Experiments with other nonlinear stiff problems completely support the qualitative pattern obtained for the Kaps problem. Some other results can be found in [5, 6].

#### ACKNOWLEDGMENTS

I am grateful to S.S. Filippov for his interest in my work.

#### REFERENCES

1. Hairer, E., Norsett, S.P., and Wanner, G., *Solving Ordinary Differential Equations. I: Nonstiff Problems*, Berlin: Springer, 1987. Translated under the title *Reshenie obyknovennykh differentsial'nykh uravnenii. Nezhestkie zadachi*, Moscow: Mir, 1990.
2. Prothero, A. and Robinson, A., On the Stability and Accuracy of One-Step Methods for Solving Stiff Systems of Ordinary Differential Equations, *Math. Comput.*, 1974, vol. 28, no. 1, pp. 145–162.
3. Dekker, K. and Verwer, J.G., *Stability of Runge–Kutta Methods for Stiff Nonlinear Differential Equations*, Amsterdam: North-Holland, 1984. Translated under the title *Ustoichivost' metodov Runge–Kutty dlya zhestkikh nelineinykh differentsial'nykh uravnenii*, Moscow: Mir, 1988.
4. Hairer, E. and Wanner, G., *Solving Ordinary Differential Equations. II: Stiff and Differential–Algebraic Problems*, Berlin: Springer, 1991. Translated under the title *Reshenie obyknovennykh differentsial'nykh uravnenii. Zhestkie i differentsial'no-algebraicheskie zadachi*, Moscow: Mir, 1999.
5. Skvortsov, L.M., Improvement of the Accuracy of Explicit Runge–Kutta Methods in Solving Moderately Stiff Problems, *Dokl. Akad. Nauk*, 2001, vol. 378, no. 5, pp. 602–604.
6. Skvortsov, L.M., Diagonal-Implicit FSAL Runge–Kutta Methods for Stiff and Differential–Algebraic Systems, *Mat. Model.*, 2002, vol. 14, no. 2, pp. 3–17.
7. *Modern Numerical Methods for Ordinary Differential Equations*, Hall, G. and Watt, J., Eds., Oxford, U.K.: Oxford Univ. Press, 1976. Translated under the title *Sovremennye chislennye metody resheniya obyknovennykh differentsial'nykh uravnenii*, Moscow: Mir, 1979.
8. Hosea, M.E. and Shampine, L.F., Analysis and Implementation of TR-BDF2, *Appl. Numer. Math.*, 1996, vol. 20, nos. 1–3, pp. 21–37.
9. Lebedev, V.I., How to Solve Stiff Systems of Differential Equations by Explicit Methods, in *Vychislitel'nye protsessy i sistemy* (Computational Processes and Systems), Marchuk, G.I., Ed., Moscow: Nauka, 1991, issue 8, pp. 237–291.
10. Novikov, E.A., *Yavnye metody dlya zhestkikh sistem* (Explicit Methods for Stiff Systems), Novosibirsk: Nauka, 1997.
11. Fowler, M.E. and Warten, R.M., A Numerical Integration Technique for Ordinary Differential Equations with Widely Separated Eigenvalues, *IBM J. Res. Develop.*, 1967, vol. 11, no. 5, pp. 537–543.
12. Bobkov, V.V., New Explicit A-Stable Methods for Numerically Solving Differential Equations, *Differ. Uravn.*, 1978, vol. 14, no. 12, pp. 2249–2251.
13. Zavorin, A.N., Application of Nonlinear Methods to Compute Transitional Processes in Electrical Circuits, *Izv. Vyssh. Uchebn. Zaved. Radioelektron.*, 1983, vol. 26, no. 3, pp. 35–41.
14. Skvortsov, L.M., Adaptive Numerical Integration Methods in Problems of Modeling Dynamical Systems, *Izv. Ross. Akad. Nauk, Teor. Sistemy Upravl.*, 1999, no. 4, pp. 72–78.
15. Skvortsov, L.M., Explicit Adaptive Methods for Numerical Solution of Stiff Systems, *Mat. Model.*, 2000, vol. 12, no. 12, pp. 97–107.