MATEMATIČKI VESNIK
МАТЕМАТИЧКИ ВЕСНИК
73, 2 (2021), 119–130
June 2021

research paper
оригинални научни рад

AN IMPLICIT-EXPLICIT METHOD OF THIRD ORDER FOR STIFF ODEs

Anton Tuzov

Abstract. This paper develops an Implicit-Explicit method (IMEX) of third order for solving stiff system of ordinary differential equations (ODEs). The method is L-stable with respect to the implicit part and allows the use of an arbitrary approximation of the Jacobian matrix. Order and stability conditions are derived and then solved analytically. Automatic stepsize selection based on local error estimation and stability control is made. The estimations for local error and stability control are obtained without significant additional computational cost. The results of numerical experiments confirm the reliability and efficiency of the implemented integration algorithm.

1. Introduction

For many systems of differential equations modeling problems in science and engineering, the right hand side is naturally split into two parts: one part is non-stiff, and the other one is stiff. Such systems can be efficiently integrated by a class of implicit-explicit (IMEX) methods, where the stiff part is treated by an implicit method, and the non-stiff part is treated by an explicit method. A class of IMEX methods is an important special case of additive methods, where the number of elemental methods $N = 2$.

Various general classes of methods can be used as a basis for constructing IMEX methods. For example, singly diagonally implicit Runge-Kutta methods with an explicit first stage (ESDIRKs) and explicit Runge-Kutta methods were used in [1,12,13] for the treatment of the stiff and non-stiff parts correspondingly. IMEX methods based on linear multistep methods (LMMs) [10], diagonally implicit multistage integration methods (DIMSIMs) [11,19] and two-step Runge-Kutta methods (TSRKs) [20] were constructed.

2020 Mathematics Subject Classification: 65L04, 65L05, 65L20

Keywords and phrases: Stiff systems of ODEs; implicit-explicit methods; L-stability, local error estimation and stability control; embedded methods.

In the current paper an IMEX method is constructed on the basis of Runge-Kutta methods and the non-stiff part is treated by an explicit RK method, as in [1, 12, 13]. In contrast to IMEX RK methods [1, 12, 13], the stiff part is treated here by a so-called (m,k)-method [15–17]. Class (m,k)-methods belongs to linearly implicit Runge-Kutta methods which avoid nonlinear systems and replace them by a sequence of linear systems. These (m,k)-methods are as simple in realization as Rosenbrock-type methods but have better accuracy and stability properties [17].

The focus of this paper is on methods with low computational cost per step due to appropriate choice of the class of methods for the treatment of the stiff part, using the stages of the main scheme in auxiliary formulas for local error estimation and stability control, suitable Jacobian approximation and re-using the same Jacobian over several steps.

Spatial discretization of continuum mechanics problems in partial differential equations by finite difference or finite element methods results in the Cauchy problem for the system of ordinary differential equations with an additively split right-hand side function of the form

$$y' = \varphi(t, y) + g(t, y), \quad y(t_0) = y_0, \quad t_0 \leq t \leq t_k,$$

where $\varphi(t, y)$ is a non-symmetric term obtained from the discretization of the first-order differential operator, $g(t, y)$ is a symmetric term obtained from the discretization of the second-order differential operator, t is an independent variable. It is assumed that in the problem the vector-function g is a stiff term and φ is a non-stiff term.

Explicit Runge-Kutta methods have a bounded stability region and are suitable for non-stiff and mildly stiff problems only. L-stable methods are usually used for solving stiff problems. In the case of large-scale problems the overall computational cost of L-stable methods is almost completely dominated by evaluations and inversions of the Jacobian matrix of the right-hand side vector function. The overall computational cost can be significantly reduced by re-using the same Jacobian matrix over several integration steps (freezing the Jacobian).

Freezing the Jacobian in iterative methods affects the convergence speed of an iterative process only and doesn't lead to loss of accuracy. So, this approach is extensively used in the implementation of these methods. For Rosenbrock type methods and their modifications [2, 5, 9] an approximation of the Jacobian matrix can lead to the reduction of the consistency order.

The system $y' = f(t, y)$ can be written in the form $y' = [f(t, y) - By] + By$, where B is some approximation of the Jacobian matrix. Assuming that stiffness is fully concentrated in the term $g(t, y) = By$, the expression $\varphi(t, y) = f(t, y) - By$ can be interpreted as the non-stiff term [4]. If the Cauchy problem is considered in the form $y' = [f(t, y) - By] + By$ in the construction of additive methods, then an arbitrary approximation of the Jacobian matrix can be used without the reduction of the order of these methods. Additive methods constructed in this way allow both analytical and numerical computations of the Jacobian matrix. Note that the approximation of the Jacobian by a diagonal matrix is suitable for some mildly stiff problems.

In this paper we construct a six-stage third order IMEX method that allows the use of different kinds of approximation of the Jacobian matrix. The estimations of the local error and the maximum absolute eigenvalue of the Jacobian matrix have been obtained without significant additional computational cost. Indeed, the error estimation has been obtained on the basis of an embedded IMEX method and the maximum absolute eigenvalue estimation has been obtained by a power method using only two additional computations of $\varphi(y)$. These estimations are used for local error and stability control correspondingly. In contrast to [15,16], we impose an additional condition of consistency of the explicit and implicit methods.

2. An IMEX scheme

Consider the Cauchy problem for an autonomous system of ordinary differential equations

$$y' = \varphi(y) + g(y), \quad y(t_0) = y_0, \quad t_0 \le t \le t_k, \tag{1}$$

where y, φ and g are N-dimensional smooth vector-functions, t is an independent variable. In the following, we assume that g is a stiff term and φ is a non-stiff term. Consider a six-stage numerical scheme for solving (1):

$$\begin{aligned} y_{n+1} &= y_n + \sum_{i=1}^{6} p_i k_i, \\ k_1 &= h\varphi(y_n), \qquad D_n k_2 = h\varphi(y_n) + hg(y_n), \qquad D_n k_3 = k_2, \\ D_n k_4 &= h\varphi(y_n + \sum_{j=1}^{3} \beta_{4j} k_j) + hg(y_n + \sum_{j=1}^{3} \alpha_{4j} k_j), \\ D_n k_5 &= k_4 + \gamma k_3, \qquad k_6 = h\varphi(y_n + \sum_{j=1}^{5} \beta_{6j} k_j), \end{aligned} \tag{2}$$

where $D_n = I - ahg'_n$, $g'_n = \partial g(y_n)/\partial y$ is the Jacobian matrix of the function $g(y)$, I is the identity matrix, k_i, $1 \le i \le 6$, are stages, $a, p_i, \alpha_{4j}, \beta_{4j}, \beta_{6j}, \gamma$ are coefficients that affect accuracy and stability properties of the scheme (2).

In the IMEX scheme (2) the stiff term g is treated by a $(4,2)$-method [17] (the implicit part), while the non-stiff term φ is treated by the three-stage explicit Runge-Kutta method (the explicit part).

3. Derivation of the third order conditions

We expand the approximate solution in a Taylor series up to terms in h^3

$$\begin{aligned} y_{n+1} = y_n &+ \big(p_1+p_2+p_3+p_4+(\gamma+1)p_5+p_6\big)h\varphi + \big(p_2+p_3+p_4+(\gamma+1)p_5\big)hg \\ &+ \Big((\beta_{41}+\beta_{42}+\beta_{43})(p_4+p_5) + \big(\beta_{61}+\beta_{62}+\beta_{63}+\beta_{64}+(\gamma+1)\beta_{65}\big)p_6\Big)h^2\varphi'\varphi \end{aligned}$$

$$
\begin{aligned}
&+\Big((\beta_{42}+\beta_{43})(p_4+p_5)+\big(\beta_{62}+\beta_{63}+\beta_{64}+(\gamma+1)\beta_{65}\big)p_6\Big)h^2\varphi' g\\
&+\big[a\big(p_2+2p_3+p_4+(3\gamma+2)p_5\big)+(\alpha_{41}+\alpha_{42}+\alpha_{43})(p_4+p_5)\big]h^2 g'\varphi\\
&+\big[a\big(p_2+2p_3+p_4+(3\gamma+2)p_5\big)+(\alpha_{42}+\alpha_{43})(p_4+p_5)\big]h^2 g' g\\
&+1/2\big[(\beta_{41}+\beta_{42}+\beta_{43})^2(p_4+p_5)+\big(\beta_{61}+\beta_{62}+\beta_{63}+\beta_{64}+(\gamma+1)\beta_{65}\big)^2 p_6\big]h^3\varphi''\varphi^2\\
&+1/2\big[(\beta_{42}+\beta_{43})^2(p_4+p_5)+\big(\beta_{62}+\beta_{63}+\beta_{64}+(\gamma+1)\beta_{65}\big)^2 p_6\big]h^3\varphi'' g^2\\
&+\big[(\beta_{42}+\beta_{43})(\beta_{41}+\beta_{42}+\beta_{43})(p_4+p_5)\\
&+\big(\beta_{61}+\beta_{62}+\beta_{63}+\beta_{64}+(\gamma+1)\beta_{65}\big)\big(\beta_{62}+\beta_{63}+\beta_{64}+(\gamma+1)\beta_{65}\big)p_6\big]h^3\varphi''\varphi g\\
&+(\beta_{41}+\beta_{42}+\beta_{43})(\beta_{64}+\beta_{65})p_6 h^3{\varphi'}^2\varphi+(\beta_{42}+\beta_{43})(\beta_{64}+\beta_{65})p_6 h^3{\varphi'}^2 g\\
&+\Big[a\Big((\beta_{42}+2\beta_{43})(p_4+p_5)+\big(\beta_{62}+2\beta_{63}+\beta_{64}+(3\gamma+2)\beta_{65}\big)p_6\Big)\\
&+(\alpha_{41}+\alpha_{42}+\alpha_{43})(\beta_{64}+\beta_{65})p_6\Big]h^3\varphi' g'\varphi+\Big[a\Big((\beta_{42}+2\beta_{43})(p_4+p_5)\\
&+\big(\beta_{62}+2\beta_{63}+\beta_{64}+(3\gamma+2)\beta_{65}\big)p_6\Big)+(\alpha_{42}+\alpha_{43})(\beta_{64}+\beta_{65})p_6\Big]h^3\varphi' g' g\\
&+1/2(\alpha_{41}+\alpha_{42}+\alpha_{43})^2(p_4+p_5)h^3 g''\varphi^2+1/2(\alpha_{42}+\alpha_{43})^2(p_4+p_5)h^3 g'' g^2\\
&+(\alpha_{42}+\alpha_{43})(\alpha_{41}+\alpha_{42}+\alpha_{43})(p_4+p_5)h^3 g''\varphi g+a(\beta_{41}+\beta_{42}+\beta_{43})(p_4+2p_5)h^3 g'\varphi'\varphi\\
&+a(\beta_{42}+\beta_{43})(p_4+2p_5)h^3 g'\varphi' g+a\big[a\big(p_2+3p_3+p_4+(6\gamma+3)p_5\big)\\
&+(\alpha_{41}+2\alpha_{42}+3\alpha_{43})p_4+(2\alpha_{41}+2\alpha_{42}+3\alpha_{43})p_5\big]h^3{g'}^2\varphi\\
&+a\big[a\big(p_2+3p_3+p_4+(6\gamma+3)p_5\big)+(2\alpha_{42}+3\alpha_{43})p_4+(2\alpha_{42}+3\alpha_{43})p_5\big]h^3{g'}^2 g+O(h^4),
\end{aligned}
$$

where the corresponding elementary differentials are evaluated at y_n.

Expanding the exact solution in a Taylor series up to terms in h^3, we obtain

$$
\begin{aligned}
y(t_{n+1}) = y(t_n)&+h(\varphi+g)+\frac{h^2}{2}(\varphi'\varphi+\varphi' g+g'\varphi+g' g)+\frac{h^3}{6}(\varphi''\varphi^2+\varphi'' g^2+2\varphi''\varphi g+{\varphi'}^2\varphi\\
&+{\varphi'}^2 g+\varphi' g'\varphi+\varphi' g' g+g''\varphi^2+g'' g^2+2g''\varphi g+g'\varphi'\varphi+g'\varphi' g+{g'}^2\varphi+{g'}^2 g)+O(h^4), \qquad (3)
\end{aligned}
$$

where the corresponding elementary differentials are evaluated at $y(t_n)$.

Comparing the successive terms in the Taylor series expansion of the approximate and the exact solutions up to third order terms under the assumption $y_n = y(t_n)$, we have the third order conditions (Tables 1 and 2) of the scheme (2):

Table 1: Order conditions for explicit Runge-Kutta method and (4,2)-method

explicit Runge-Kutta method	(4,2)-method
$p_1+p_2+p_3+\widetilde{p}_4+\gamma p_5+p_6=1$	$p_2+p_3+\widetilde{p}_4+\gamma p_5=1$
$\beta_4\widetilde{p}_4+(\beta_6+\gamma\beta_{65})p_6=1/2$	$a(p_2+2p_3+\widetilde{p}_4+(3\gamma+1)p_5)+\widetilde{\alpha}_4\widetilde{p}_4=1/2$
$\beta_4^2\widetilde{p}_4+(\beta_6+\gamma\beta_{65})^2 p_6=1/3$	$\widetilde{\alpha}_4^2\widetilde{p}_4=1/3$
$\beta_4\widetilde{\beta}_{64}p_6=1/6$	$a\big[a\big(p_2+3p_3+\widetilde{p}_4+2(3\gamma+1)p_5\big)+(2\widetilde{\alpha}_4+\alpha_{43})\widetilde{p}_4\big]=1/6$

Table 2: Coupling conditions for IMEX scheme

$$
\begin{aligned}
&\widetilde{\beta}_4\widetilde{p}_4 + (\beta_6 - \beta_{61} + \gamma\beta_{65})p_6 = 1/2,\\
&a(p_2 + 2p_3 + \widetilde{p}_4 + (3\gamma+1)p_5) + \alpha_4\widetilde{p}_4 = 1/2,\\
&\widetilde{\beta}_4^2\widetilde{p}_4 + (\beta_6 - \beta_{61} + \gamma\beta_{65})^2 p_6 = 1/3,\\
&\beta_4\widetilde{\beta}_4\widetilde{p}_4 + (\beta_6 + \gamma\beta_{65})(\beta_6 - \beta_{61} + \gamma\beta_{65})p_6 = 1/3,\\
&\widetilde{\beta}_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&a\Big((\widetilde{\beta}_4 + \beta_{43})\widetilde{p}_4 + \big(\widetilde{\beta}_6 + \beta_{63} + (3\gamma+1)\beta_{65}\big)p_6\Big) + \alpha_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&a\Big((\widetilde{\beta}_4 + \beta_{43})\widetilde{p}_4 + \big(\widetilde{\beta}_6 + \beta_{63} + (3\gamma+1)\beta_{65}\big)p_6\Big) + \widetilde{\alpha}_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&\alpha_4^2\widetilde{p}_4 = 1/3,\\
&\alpha_4\widetilde{\alpha}_4\widetilde{p}_4 = 1/3,\\
&a\beta_4(\widetilde{p}_4 + p_5) = 1/6,\\
&a\widetilde{\beta}_4(\widetilde{p}_4 + p_5) = 1/6,\\
&a\big[a\big(p_2 + 3p_3 + \widetilde{p}_4 + 2(3\gamma+1)p_5\big) + (2\alpha_4 - \alpha_{41} + \alpha_{43})\widetilde{p}_4 + \alpha_{41}p_5\big] = 1/6.
\end{aligned}
$$

Here $\alpha_4 = \sum_{j=1}^{3}\alpha_{4j}$, $\beta_4 = \sum_{j=1}^{3}\beta_{4j}$, $\beta_6 = \sum_{j=1}^{5}\beta_{6j}$, $\widetilde{p}_4 = p_4 + p_5$, $\widetilde{\beta}_{64} = \beta_{64} + \beta_{65}$, $\widetilde{\alpha}_4 = \alpha_4 - \alpha_{41}$, $\widetilde{\beta}_4 = \beta_4 - \beta_{41}$, $\widetilde{\beta}_6 = \beta_6 - \beta_{61}$. After simplification the third order conditions take the form:

$$
\begin{aligned}
&\alpha_{41} = \beta_{41} = \beta_{61} = 0, \quad p_1 = -p_6,\\
&p_2 + p_3 + \widetilde{p}_4 + \gamma p_5 = 1, \quad \beta_4\widetilde{p}_4 + (\beta_6 + \gamma\beta_{65})p_6 = 1/2,\\
&a(p_2 + 2p_3 + \widetilde{p}_4 + (3\gamma+1)p_5) + \alpha_4\widetilde{p}_4 = 1/2,\\
&\beta_4^2\widetilde{p}_4 + (\beta_6 + \gamma\beta_{65})^2 p_6 = 1/3, \quad \beta_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&a\big[(\beta_4 + \beta_{43})\widetilde{p}_4 + \big(\beta_6 + \beta_{63} + (3\gamma+1)\beta_{65}\big)p_6\big] + \alpha_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&\alpha_4^2\widetilde{p}_4 = 1/3, \quad a\beta_4(\widetilde{p}_4 + p_5) = 1/6,\\
&a\big[a\big(p_2 + 3p_3 + \widetilde{p}_4 + 2(3\gamma+1)p_5\big) + (2\alpha_4 + \alpha_{43})\widetilde{p}_4\big] = 1/6.
\end{aligned}
\tag{4}
$$

4. Linear stability analysis

Let us investigate the stability properties of the IMEX scheme (2) with respect to the scalar linear test problem

$$y' = \lambda_1 y + \lambda_2 y, \quad y(0) = y_0, \quad t \geq 0,\ \Re(\lambda_1) \leq 0,\ \Re(\lambda_2) \leq 0,\ |\Re(\lambda_1)| \ll |\Re(\lambda_2)|, \tag{5}$$

where the free parameters $\lambda_1,\ \lambda_2$ can be interpreted as eigenvalues of the Jacobian matrices of the functions φ (the non-stiff term) and g (the stiff term) correspondingly.

Applying the scheme (2) to the problem (5), we obtain $y_{n+1} = R(x,z)y_n$, where $x = \lambda_1 h,\ z = \lambda_2 h$ and $R(x,z)$ is a stability function (its analytical expression is omitted here for brevity).

The necessary condition of L-stability of the IMEX scheme (2) with respect to the stiff term has the form $\lim\limits_{z\to-\infty} R(x,z) = 0$. It is satisfied if the following two conditions hold:

$$a^2(p_1+p_6) + \big((\alpha_{42}-a)\beta_{64} - a\beta_{62}\big)p_6 = 0, \quad a(a-p_2) + (\alpha_{42}-a)p_4 = 0. \tag{6}$$

5. An analytical solution of the order and stability conditions

Solving the system (4), (6), we assume that $\sum_{j=1}^{3}\alpha_{4j} = \sum_{j=1}^{3}\beta_{4j}$, $\sum_{j=1}^{5}\beta_{6j} = 1$, $\alpha_{42} = a$, $\beta_{42} = a$. The first relation means the consistency of explicit and implicit methods of IMEX scheme (2) [3,4], i.e., in the fourth stage φ and g are evaluated at the same point. The second relation ensures that in the sixth stage $\varphi(y_n + \sum_{j=1}^{5}\beta_{6j}k_j)$ approximates $\varphi(y(t_{n+1}))$, the other ones improve the stability properties of the intermediate numerical formulas. Obvious simplifications of the system (4), (6) yields

$$\begin{aligned}
&\alpha_{41} = \beta_{41} = \beta_{61} = \beta_{62} = 0,\ \beta_6 = 1,\ \alpha_{42} = \beta_{42} = p_2 = a,\\
&\alpha_{43} = \beta_{43} = \alpha_4 - a,\ p_1 = -p_6,\ p_3 + \widetilde{p}_4 + \gamma p_5 = 1-a,\\
&\alpha_4\widetilde{p}_4 + (\gamma\beta_{65}+1)p_6 = 1/2,\ a^2 + a(2p_3 + \widetilde{p}_4 + (3\gamma+1)p_5) + \alpha_4\widetilde{p}_4 = 1/2,\\
&\alpha_4^2\widetilde{p}_4 + (\gamma\beta_{65}+1)^2p_6 = 1/3,\ \alpha_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&a\big[(2\alpha_4 - a)\widetilde{p}_4 + \big(\beta_{63} + (3\gamma+1)\beta_{65} + 1\big)p_6\big] + \alpha_4\widetilde{\beta}_{64}p_6 = 1/6,\\
&\alpha_4^2\widetilde{p}_4 = 1/3,\ a\alpha_4(\widetilde{p}_4 + p_5) = 1/6,\\
&a\big[a^2 + a\big(3p_3 + \widetilde{p}_4 + 2(3\gamma+1)p_5\big) + (3\alpha_4 - a)\widetilde{p}_4\big] = 1/6.
\end{aligned} \tag{7}$$

Now, the coefficients of the L -stable third order scheme (2) are given by

$$\begin{aligned}
&\alpha_{41} = \beta_{41} = \beta_{61} = \beta_{62} = 0, \quad \alpha_4 = 2/3,\\
&\alpha_{42} = \beta_{42} = p_2 = a, \quad \alpha_{43} = \beta_{43} = 2/3 - a,\\
&\gamma = (4a^2 - 2a - 1)/(1-3a), \quad u = (\gamma+1)/(3(1-a)\gamma),\\
&p_4 = (6a-1)/(4a), \quad p_5 = 3/4 - p_4\\
&p_3 = 1/4 - a - \gamma p_5, \quad p_6 = 1/(4u), \quad p_1 = -p_6,\\
&\beta_{65} = -1/\gamma, \quad \beta_{63} = 1-u, \quad \beta_{64} = u - \beta_{65},
\end{aligned} \tag{8}$$

where the coefficient a is determined from the equation $4a^2 - 9a + 3 = 0$. This equation has two real roots $a_1 = (9-\sqrt{33})/8$ and $a_2 = (9+\sqrt{33})/8$. Numerical experiments show that the method with a_1 gives more accurate results. The coefficients, corresponding to $a = a_1$, are

$$\begin{aligned}
&\alpha_{41} = \beta_{41} = \beta_{61} = \beta_{62} = 0, && p_4 = 0.885643223060915,\\
&a = 0.406929669182746, && p_5 = -0.135643223060915,\\
&\gamma = 5.21535165408627, && \alpha_{43} = \beta_{43} = 0.259736997483920,\\
&p_1 = -p_6 = -0.373237570007449, && \beta_{63} = 0.330185329427018,\\
&p_2 = \alpha_{42} = \beta_{42} = 0.406929669182746, && \beta_{64} = 0.861556295361886,\\
&p_3 = 0.550497438573592, && \beta_{65} = -0.191741624788904,
\end{aligned} \tag{9}$$

6. Local error estimation

For the local error estimation we construct an embedded method of second order of the form

$$
\begin{aligned}
y_{n+1,\,2} &= y_n + \sum_{i=1}^{4} r_i k_i + r_5 \widetilde{k_5}\ ,\\
k_1 &= h\varphi(y_n), \quad D_n k_2 = h\varphi(y_n) + hg(y_n), \quad D_n k_3 = k_2, \qquad (10)\\
D_n k_4 &= h\varphi(y_n + \sum_{j=1}^{3} \beta_{4j} k_j) + hg(y_n + \sum_{j=1}^{3} \alpha_{4j} k_j),\\
D_n \widetilde{k_5} &= k_4,
\end{aligned}
$$

where the coefficients r_i, $1 \le i \le 5$, should be determined, and parameters α_{4j}, β_{4j} are given by (9). This form is chosen to avoid additional computational cost, associated with evaluating of right-hand side, evaluating and inverting of the Jacobian matrix. Note that there is no sixth stage in (10) and γk_3 in the fifth stage, in contrast to (2).

We expand the approximate solution computed by the scheme (10) in a Taylor series up to terms in h^2

$$
\begin{aligned}
y_{n+1,\,2} &= y_n + (r_1 + r_2 + r_3 + r_4 + r_5)h\varphi + (r_2 + r_3 + r_4 + r_5)hg\\
&+ \big(a(r_2 + 2r_3 + r_4 + 2r_5) + r_4 + r_5\big)h^2 g'\varphi + \big(a(r_2 + 2r_3 + r_4 + 2r_5) + r_4 + r_5\big)h^2 g' g\\
&+ \beta_4(r_4 + r_5)h^2\varphi'\varphi + \beta_4(r_4 + r_5)h^2\varphi' g + O(h^3),
\end{aligned}
$$

where the elementary differentials are evaluated at y_n. Comparing successive terms in the Taylor series expansion of the approximate and the exact solutions up to second order terms under the assumption $y_n = y(t_n)$, we obtain the second order conditions of the scheme (10):

$$
\begin{aligned}
&r_1 + r_2 + r_3 + r_4 + r_5 = 1, \quad r_2 + r_3 + r_4 + r_5 = 1, \qquad (11)\\
&\beta_4(r_4 + r_5) = 1/2, \quad a(r_2 + 2r_3 + r_4 + 2r_5) + r_4 + r_5 = 1/2,
\end{aligned}
$$

Now we analyze the stability of the scheme (10). Its application to numerically solving the equation (5) yields $y_{n+1,\,2} = R_2(x,z)\ y_{n,\,2}$, where $x = \lambda_1 h,\ z = \lambda_2 h$ and the stability function $R_2(x,z)$ has the form

$$
\begin{aligned}
R_2(x,z) = {}& [a^3(a-r_2)z^4 - a^3(r_2-r_4)xz^3 - a\big(4a^2 - a(3r_2+r_3+2r_4)+r_4\big)z^3\\
&+ a^3 r_4 x^2 z^2 + a\big(a(3r_2+r_3+r_4-r_5) - r_4(\beta_4+1)\big)xz^2\\
&+ \big(6a^2 - a(3r_2+2r_3+3r_4+2r_5)+r_4+r_5\big)z^2 - a\big(a(r_4+r_5)+r_4\beta_4\big)x^2 z\\
&+ \big(-a(3r_2+2r_3+3r_4+2r_5)+(r_4+r_5)(\beta_4+1)\big)xz\\
&+ (-4a+r_2+r_3+r_4+r_5)z + \beta_4(r_4+r_5)x^2 + (r_2+r_3+r_4+r_5)x+1]/(1-az)^4.
\end{aligned}
$$

From the necessary condition of L-stability of the auxiliary scheme (10) with respect to the stiff term $\lim_{z\to-\infty} R_2(x,z) = 0$ we have $r_2 = a$. Then, solving (11), we obtain the coefficients of the L stable embedded method (10) of second order

$$
r_1 = 0,\ r_2 = a,\ r_3 = 1 - a - v,\ r_4 = 2 - a + (v - 1/2)/a,\ r_5 = v - r_4,
$$

where $v = 1/(2\beta_4)$. The coefficients, corresponding to (9), are

$$r_1 = 0, \ r_2 = 0.406929669182746, \ r_3 = -0.156929669182746,$$
$$r_4 = 2.20742710775634, \ r_5 = -1.45742710775634.$$

The embedded method (10) requires, at each integration step, only one additional backward substitution steps of Gauss elimination method and doesn't require additional evaluations of right-hand side, evaluations and inversions of the Jacobian matrix. In the case of large-scale problems the overall computational cost of the method (10) is almost completely dominated by evaluations and inversions of the Jacobian matrix. So, we have obtained the local error estimation based on the embedded method (10) without significant additional computational cost.

Let us denote the local error estimation by

$$err_n = \max_{1\le i\le N} \frac{|y_n^i - y_{n,\,2}^i|}{Atol_i + Rtol_i|y_n^i|},$$

where $Atol_i$ and $Rtol_i$ are the desired tolerances prescribed by the user. If $err_n \le 1$, then the computed step is accepted, else the step is rejected and computations are repeated. When $Rtol_i = 0$, the absolute error is controlled on the i-th component of the solution with the desired tolerance $Atol_i$. If $Atol_i = 0$ then the relative error is controlled on the i-th component with the tolerance $Rtol_i$.

7. Stability control and stepsize selection

In the IMEX scheme (2) the stiff term g is treated by the L-stable $(4,2)$-method [17] (the implicit part), while the non-stiff term φ is treated by the three-stage explicit Runge-Kutta method (the explicit part). In the general case there is no guarantee that the function $\varphi(y) = f(y) - By$ is the non-stiff term in reducing $y' = f(y)$ to $y' = [f(y) - By] + By$. If some stiffness is in $\varphi(y) = f(y) - By$ (i.e., stiffness leakage phenomenon occurs) then the additional stability control of the explicit part of the scheme (2) can increase efficiency of computations for many problems. In some cases it does not have a significant effect on the efficiency of the integration algorithm because of the good stability properties of the scheme (2). Therefore the choice of using or not using the additional stability control of the explicit part is given to the end-user.

We perform the stability control of the explicit part of the scheme (2) by analogy with [15, 16]. Let us consider the additional stages d_1, d_2 of the form

$$d_1 = h\varphi(y_n + \alpha_{21}k_1), \quad d_2 = h\varphi(y_n + \alpha_{31}k_1 + \alpha_{32}d_1).$$

Denote $\varphi(y) = Ay + b$, where A and b are matrix and vector with constant coefficients correspondingly, then we have

$$k_1 = h(Ay_n + b), \quad d_1 = k_1 + \alpha_{21}hAk_1, \quad d_2 = k_1 + (\alpha_{31} + \alpha_{32})hAk_1 + \alpha_{21}\alpha_{32}h^2A^2k_1.$$

Assuming $\alpha_{21} = \alpha_{31} + \alpha_{32}$, we obtain

$$d_2 - d_1 = \alpha_{21}\alpha_{32}h^2A^2k_1, \quad d_1 - k_1 = \alpha_{21}hAk_1.$$

The maximum absolute eigenvalue $v_n = h|\lambda_{n\ max}|$ of the matrix hA can be approximated using the power method by the following formula

$$v_n = |\alpha_{32}^{-1}| \max_{1 \le i \le N} \frac{|d_2^i - d_1^i|}{|d_1^i - k_1^i|},$$

then the stability control can be made by $v_n \le 2$, where number 2 is an approximate length of the stability interval of the three-stage explicit Runge-Kutta method.

In the general case this estimation is quite crude because of the small number of iterations of the power method and the nonlinearity of the function $\varphi(y)$. Therefore the stability control is used for limiting the stepsize growing only.

Let the approximate solution y_n be computed with the stepsize h_n. For the stepsize selection we use $err_n = O(h_n^3)$. The stepsize h_{acc} predicted by accuracy we compute by the formula $h_{acc} = q_1 h_n$, where q_1 is a root of the equation $q_1^3 err_n = 1$. In view of $v_n = O(h_n)$, the stepsize h_{st} predicted by stability is computed by $h_{st} = q_2 h_n$, where q_2 is a root of the equation $q_2 v_n = 2$. Then the stepsize h_{n+1} predicted by accuracy and stability is selected by the formula $h_{n+1} = \max[h_n, \min(h_{acc}, h_{st})]$.

The stability control of the explicit part of the scheme (2) requires, at each integration step, two additional computations of $\varphi(y)$. This computational cost is negligible for large-scale problems, but if one is sure that all stiffness is in $g(y)$ then one can turn off stability control to save computational cost.

8. Numerical experiments

In what follows, the numerical code based on the third order IMEX method (2) constructed under the consistency condition and with local error estimation and stability control as well as with diagonal Jacobian approximation is called IMEX3.

The test problems given below have been reduced to the form $y' = (f(y) - By) + By$. All numerical computations have been performed in double precision arithmetic with the desired tolerances of the local error $Atol = Rtol = Tol = 10^{-m},\ m = 2, 4$. The scheme (2) is of third order, therefore it is unreasonable to perform numerical computations with higher tolerance.

The following four test problems are considered:

Test problem 1 [6]

$$\begin{aligned} y_1' &= -0.013y_1 - 1000y_1y_3, \\ y_2' &= -2500y_2y_3, \\ y_3' &= -0.013y_1 - 1000y_1y_3 - 2500y_2y_3, \\ & t \in [0, 50], \quad y_1(0) = 1, \quad y_2(0) = 1, \quad y_3(0) = 0, \quad h_0 = 2.9 \cdot 10^{-4}. \end{aligned}$$

Test problem 2 [8]

$$\begin{aligned} y_1' &= 77.27(y_2 - y_1y_2 + y_1 - 8.375 \cdot 10^{-6}y_1^2), \\ y_2' &= (-y_2 - y_1y_2 + y_3)/77.27, \\ y_3' &= 0.161(y_1 - y_3), \end{aligned}$$

$$t \in [0,300], \quad y_1(0) = 4, \quad y_2(0) = 1.1, \quad y_3(0) = 4, \quad h_0 = 2 \cdot 10^{-3}.$$

Test problem 3

$$\begin{aligned}
y_1' &= -0.04y_1 + 0.01y_2y_3,\\
y_2' &= 400y_1 - 100y_2y_3 - 3000y_2^2,\\
y_3' &= 30y_2^2,\\
&t \in [0,40], \quad y_1(0) = 1, \quad y_2(0) = y_3(0) = 0, \quad h_0 = 10^{-5}.
\end{aligned}$$

Test problem 4

$$\begin{aligned}
y_1' &= y_3 - 100y_1y_2,\\
y_2' &= y_3 + 2y_4 - 100y_1y_2 - 2 \cdot 10^4 y_2^2,\\
y_3' &= -y_3 + 100y_1y_2,\\
y_4' &= -y_4 + 10^4 y_2^2,\\
&t \in [0,20], \quad y_1(0) = y_2(0) = 1, \quad y_3(0) = y_4(0) = 0, \quad h_0 = 2.5 \cdot 10^{-5}.
\end{aligned}$$

The approximation of the Jacobian by a diagonal matrix is used when solving the test problems by IMEX3, ASODE3-1 [15] and ASODE3-2 [16]. In this case computational cost of additive methods is dominated by the number of right-hand side function evaluations, the same is true for explicit Runge-Kutta methods. Therefore, IMEX3 is compared with the following numerical codes based on well-known explicit Runge-Kutta methods:

RKM4	– 5-stage	Merson method of order 4 [14],
RKF5	– 6-stage	Felberg method of order 5 [7],
RKF7	– 13-stage	Felberg method of order 7 [7],
DP8	– 13-stage	Dormand and Prince method of order 8 [18],

and additive methods:

ASODE3-1	– 6-stage	method of order 3 [15],
ASODE3-2	– 6-stage	method of order 3 [16].

The overall computational cost (measured by the number of right-hand side function evaluations over the integration interval) is given in the Table 3 and Table 4.

As can be seen from Tables 3 and Table 4, the developed integration algorithm IMEX3 is more efficient than the other additive and explicit Runge-Kutta methods considered in this paper.

9. Conclusions

So, in this paper we have constructed the third order IMEX method, which combines a (4,2)-method with an explicit Runge–Kutta method and includes an embedded method for error control. This method is L-stable with respect to the implicit part and allows the use of an arbitrary approximation of the Jacobian matrix without loss

Table 3: Computational cost of the explicit Runge-Kutta methods and IMEX3

№	Tol	RKM4	RKF5	RKF7	DP8	IMEX3
1	10^{-2}	401 716	401 005	982 536	717 526	90
	10^{-4}	400 627	400 656	982 150	717 287	2 232
2	10^{-2}	13 391 594	15 694 434	38 429 196	27 998 053	3 951
	10^{-4}	13 384 132	15 691 105	38 429 976	27 993 793	76 092
3	10^{-2}	204 889	237 942	587 509	431 591	417
	10^{-4}	206 647	240 676	565 396	430 823	3 297
4	10^{-2}	10 832	11 874	29 991	23 052	123
	10^{-4}	10 236	11 366	28 819	23 354	5 766

Table 4: Computational cost of the additive methods ASODE3-1, ASODE3-2, IMEX3

№	Tol	ASODE3-1	ASODE3-2	IMEX3
1	10^{-2}	3 129	243	90
	10^{-4}	16 361	5 253	2 232
2	10^{-2}	63 430	4 245	3 951
	10^{-4}	367 411	89 993	76 092
3	10^{-2}	9 351	1 278	417
	10^{-4}	37 338	7 908	3 297
4	10^{-2}	1 589	174	123
	10^{-4}	7 711	7 938	5 766

of accuracy. Order and stability conditions were derived and then solved analytically. Automatic stepsize selection based on local error estimation and stability control is performed and the auxiliary formulas for performing this were obtained without significant additional computational cost.

The aim of the numerical computations was to test the reliability and efficiency of the implemented integration algorithm with local error estimation and stability control as well as with diagonal Jacobian approximation. Solving specific applied problems is beyond the scope of the current computations.

Numerical experiments show reliability and efficiency of the presented method for solving mildly stiff problems and that the test problems considered turned out to be rather stiff for the explicit Runge-Kutta methods.

Acknowledgement. The author thanks Russian Fond of Fundamental Research for financial support (project № 05-01-00579-a)

References

[1] S. Boscarino, *On an accurate third order implicit-explicit Runge-Kutta method for stiff problems*, Appl. Numer. Math., **59(7)** (2009), 1515–1528.

[2] J. C. Butcher, *Numerical methods for ordinary differential equations*, John Wiley&Sons, Chichester, 2016.

[3] G. J. Cooper, *Additive methods for the numerical solution of ordinary differential equations*, Math. Comput., **35(152)** (1980), 1159–1172.

[4] G. J. Cooper, *Additive Runge – Kutta methods for Stiff Ordinary Differential Equations*, Math. Comput., **40(161)** (1983), 207–218.

[5] K. Dekker, J. G. Verwer, *Stability of Runge-Kutta methods for stiff nonlinear differential equations*, North-Holland Publishing Co., Amsterdam, 1984.

[6] W. H. Enright, T. E. Hull, *Comparing numerical methods for the solutions of stiff systems of ODE's*, BIT, **15** (1975), 10–48.

[7] E. Fehlberg, *Classical fifth-, sixth-, seventh- and eighth order Runge – Kutta formulas with step size control*, Computing, **4** (1969), 93–106.

[8] C. W. Gear, *The automatic integration of stiff ordinary differential equations*, Proc. IFIP Congress, **1** (1968), 187—193.

[9] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Springer-Verlag, Berlin, 2010.

[10] W. Hundsdorfer, S. J. Ruuth, *IMEX extensions of linear multistep methods with general monotonicity and boundedness properties*, J. Comput. Phys., **225** (2007), 2016—2042.

[11] Z. Jackiewicz, H. Mittelmann, *Construction of IMEX DIMSIMs of high order and stage order*, Appl. Numer. Math., **121** (2017), 234–248.

[12] C.A. Kennedy, M.H. Carpenter, *Higher-order additive Runge–Kutta schemes for ordinary differential equations*, Appl. Numer. Math. **136** (2019), 183–205.

[13] C.A. Kennedy, M.H. Carpenter, *Diagonally implicit Runge-Kutta methods for stiff ODEs*, Appl. Numer. Math., **146** (2019), 221–244.

[14] R. H. Merson, *An operational methods for integration processes*, Proc. Symp. on Data Proc., **1**, 1957.

[15] E. Novikov, A. Tuzov, *A third-order nonhomogeneous method for additive stiff systems*, Math. Models Comput. Simul., **19(6)** (2007), 61-–70.

[16] E. Novikov, A. Tuzov, *A six-stage third order additive method for stiff ordinary differential equations*, Sib. Zh. Vychisl. Mat., **10(3)** (2007), 307–316.

[17] E. A. Novikov, Yu. A. Shitov, Yu. I. Shokin, *On a class of* (m, k) *- methods for solving stiff systems*, Comput. Math. Math. Phys., **29(2)** (1989), 194–201.

[18] P. J. Prince, J. R. Dormand, *High order embedded Runge – Kutta formulae*, J. Comp. Appl. Math, **7** (1981), 67–75.

[19] H. Zhang, A. Sandu, S. Blaise, *Partitioned and implicit–explicit general linear methods for ordinary differential equations*, J. Sci. Comput., **61(1)** (2014), 119-–144.

[20] E. Zharovsky, A. Sandu, H. Zhang, *A class of implicit-explicit two-step Runge–Kutta methods*, SIAM J. Numer. Anal., **53** (2015), 321-–341.

(received 22.10.2019; in revised form 02.05.2020; available online 19.12.2020)

Department of Applied Mathematics, Siberian State Aerospace University, Krasnoyarsk, 660037, Russian Federation

E-mail: tuzov@sibsau.ru