An Embedded Fourth Order Method for Solving Structurally Partitioned Systems of Ordinary Differential Equations

Igor V. Olemskoy and Alexey S. Eremin

Faculty of Applied Mathematics and Control Processes
Saint-Petersburg State University
Universitetskii prospekt 35, Petergof, Saint-Petersburg, Russia, 198504

Copyright © 2015 Igor V. Olemskoy and Alexey S. Eremin. This article is distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

An embedded one-step numerical method for structurally partitioned systems of ordinary differential equations (ODEs) is considered. Two-parametric families of methods of order four with automatic step-control are constructed for systems of ODEs of first and second order. The methods have fewer stages than classic Runge–Kutta methods.

Mathematics Subject Classification: 65L05, 65L06

Keywords: initial value problem, embedded methods, partitioned systems, control term

1 Introduction

In the paper [12] there was presented the algorithm of partitioning a system of ordinary differential equations (ODEs)

\[ z'_s = \varphi_s(x, z_1, \ldots, z_M), \quad s = 1, \ldots, M, \] (1)

and the way of its reducing to the canonical form of structurally partitioned systems

\[ y'_0 = f_0(x, y_0, \ldots, y_n), \] (2)
\[ y'_i = f_i(x, y_0, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n), \quad i = 1, \ldots, l, \] (3)
\[ y'_j = f_j(x, y_0, \ldots, y_{j-1}), \quad j = l + 1, \ldots, n, \] (4)
The groups of equations (3) and (4) are structurally identical. Equations in them are ordered in such way that in every equation the right-hand side depends only on the previous functions of the same group. The group (2) contains all the equations that cannot be placed into (3) or (4), so it can be called general group. Sometimes it can be missing, as well as the group (3).

In works [6, 9, 10] a generalisation of explicit Runge–Kutta methods (RKs) for structurally partitioned systems (SPS) of ODEs (2)–(4) was suggested. The effectiveness of the methods presented there is provided by the fact, that the number of stages \( m \) for the general group (3) is equal to that for classic RKs \( (m = Q, q \leq 4) \), but for the groups (3) and (4), taking in account their special structure, it can be reduced while keeping the method’s order. Moreover, when the general group is absent the number of necessary stages can be reduced even more significantly. Thus, a fifth order method for a system with only groups (3) and (4) was constructed with four stages [9] (while classic RKs have at least six).

The considered generalisation of RKs is based on usage of the system’s special structure in the algorithm. It is natural to apply the idea to embedded RKs. Dormand–Prince type methods [2] are widely used nowadays. Combining their approach with the First Same As Last (FSAL) idea — the last stage of a step is the first of the next step [3, 4] — fast explicit one-step methods with stable step-size control were constructed. For systems with groups (3) and (4) an embedded Dormand–Prince type method of order five with automatic step-size control was constructed with five stages [11].

2 The integration method

In the paper an explicit embedded fourth order method with automatic step-size control for systems

\[
\begin{align*}
x \in [X_0, X_k] \subset \mathbb{R}, \\
y_s : [X_0, X_k] \rightarrow \mathbb{R}^{r_s}, \quad s = 0, \ldots, n, \\
f_0 : [X_0, X_k] \times \mathbb{R}^M \rightarrow \mathbb{R}^{r_0}, \quad \sum_{s=0}^{n} r_s = M, \\
f_i : [X_0, X_k] \times \mathbb{R}^{M-r_i} \rightarrow \mathbb{R}^{r_i}, \quad \hat{r}^i = \sum_{s=i}^{l} r_s, \quad i = 1, \ldots, l, \\
f_j : [X_0, X_k] \times \mathbb{R}^{M-r_j} \rightarrow \mathbb{R}^{r_j}, \quad \hat{r}^j = \sum_{s=j}^{n} r_s, \quad j = l + 1, \ldots, n.
\end{align*}
\]

\[
\begin{align*}
y'_i &= f_i(x, y_1, \ldots, y_{i-1}, y_{i+1}, \ldots, y_n), \quad i = 1, \ldots, l, \\
y'_j &= f_j(x, y_1, \ldots, y_{j-1}), \quad j = l + 1, \ldots, n, \quad n \geq 3
\end{align*}
\]
An embedded fourth order method for SPS of ODEs

4845

is constructed.

Such systems appear, for example, in problems of optimal control and stabilisation [7, 8], medical modelling [13], celestial mechanics or high-energy physics. For instance, large-scale oscillations of systems with rotational symmetry [6] are described (after some reducing) with equations

\[
\begin{align*}
x'' &= K - \frac{x}{(2x + z)^{3/2}}, \\
z'' &= L - \frac{z}{(2x + z)^{3/2}}, \\
K' &= -\frac{1}{(2x + z)^{3/2}} x', \\
L' &= -\frac{1}{(2x + z)^{3/2}} z'.
\end{align*}
\]  

(6)

Here \( x \) is a dimensionless moment of inertia about the rotational axis; \( z \) is a dimensionless moment of inertia about the equatorial plane; \( K \) is a kinetic energy of motions parallel to the equatorial plane; and \( L \) is a kinetic energy of vertical motions. Change of variables \( \xi_1 = x', \xi_2 = z', \xi_3 = x, \xi_4 = z, \xi_5 = K, \xi_6 = L \) reduces the system to the form (5), which is an SPS without the general group.

The approximated solution of (5) is found as

\[
y_s(x + h) \approx z_s = y_s(x) + \sum_{v=1}^{m_s} b_{sv} k_{sv}(h), \quad s = 1, \ldots, n,
\]  

(7)

where \( k_{sv} \equiv k_{sv}(h) \) are calculated in the strict order

\[
k_{11}, k_{21}, \ldots, k_{n1}, k_{12}, k_{22}, \ldots,
\]

as

\[
k_{iv} = hf_i \left( x + c_{iv} h, y_1(x) + \sum_{w=1}^{v} a_{i,v,w} k_{1w} + \ldots + y_{i-1}(x) + \sum_{w=1}^{v} a_{i,v,i-1,w} k_{i-1,w}, y_{i+1}(x) + \sum_{w=1}^{v-1} a_{i,v,l+1,w} k_{l+1,w} + \ldots + y_{n}(x) + \sum_{w=1}^{v-1} a_{ivnw} k_{nw} \right),
\]  

(8)

\[
k_{jv} = hf_j \left( x + c_{jv} h, y_1(x) + \sum_{w=1}^{v} a_{jv1w} k_{1w} + \ldots + y_{j-1}(x) + \sum_{w=1}^{v} a_{j,v,j-1,w} k_{j-1,w} \right),
\]

\[
c_{i1} = 0, \quad i = 1, \ldots, l, \quad c_{j1} \neq 0, \quad j = l + 1, \ldots, n.
\]

The method will be referred as structural.

For fast and effective practical implementation of numerical methods for ODEs one needs a step-size control algorithm. Here the embedded Dormand–Prince type method of order four, corresponding to the form (7)–(8), is constructed on base of three-stage fourth order methods from [6]. In Dormand–Prince methods two approximations of different orders are found simultaneously; the higher order result is used in further computations, and the lower
order result is used only for the step-size control (technically, it is not the error being estimated, but the last term equal in the Taylor series of the solution and the approximation).

In [6] a number of important relations between the method’s parameters was obtained. They combine the parameters \( a_{spv}, b_{sp} \) and \( c_{sp} \) into quite few groups and can be referred as group equalities:

\[
\begin{align*}
    c_{1v} &= \ldots = c_{lv} : c_{iv}, \\
    c_{i+1,v} &= \ldots = c_{nv} : c_{jv}, \\
    b_{1v} &= \ldots = b_{lv} : b_{iv}, \\
    b_{i+1,v} &= \ldots = b_{nv} : b_{jv}, \\
    a_{iviw} &= \ldots = a_{ivi-1,w} : a_{iviw}, \\
    a_{i,v,i+1,w} &= \ldots = a_{ivnw} : a_{ivjw}, \\
    a_{jv1w} &= \ldots = a_{jvlw} : a_{jviw}, \\
    a_{j,v,j+1,w} &= \ldots = a_{j,v,j-1,w} : a_{jv}.
\end{align*}
\] (9)

Here and further the indices \( i, j, i', j' \) of the parameters \( c_{iv}, c_{jv}, b_{iv}, b_{jv}, a_{iviw}, a_{jv}, a_{jviw}, a_{jv} \) correspond to the group attribute, not to a certain number. Namely, the index at first place connects parameters to the system (5) structural groups — \( i \) to the first and \( j \) to the second; and the third index shows which argument of \( f_i \) or \( f_j \) in (8) the parameter corresponds — \( i \) and \( i' \) up to \( l \)-th function, \( j \) and \( j' \) from \( l + 1 \) to \( n \).

The use of restrictions (9) significantly simplifies the design of order conditions and methods construction and realisation.

We use the slightly modified Butcher tableau to present the embedded structural method (see table 1).

The control term estimation is done with the same \( k_{iv} \) and \( k_{jv} \) values as the result but with \( d_{iv} \) and \( d_{jv} \) parameters

\[
\tilde{z}_s = y_s(x) + \sum_{v=1}^{m_s} d_{sv} k_{sv}. \tag{10}
\]

The parameters of an \( m = (m_i, m_j) \)-stage embedded method (7)–(8) of order \( p \) and estimator order \( q \) must provide that \( y_s(x + h) - z_s = O(h^{p+1}) \) and \( y_s(x + h) - \tilde{z}_s = O(h^{q+1}) \) for all \( s = 1, \ldots, n \).

We consider methods where \( p > q \). The difference \( z_s - \tilde{z}_s \) is used for the step-size control.

The method’s (7)–(8), (10) abbreviation RKSp\( (q)(m_i, m_j) \)F or RKSp\( (q)m \)F tells about its type — RK (Runge–Kutta); its generalisation to systems with special structure (5) — S; its main order \( p \) and the estimator order \( (q) \); number of stages \( m \) if \( m_i = m_j = m \) or \( m_i, m_j \); and about the FSAL idea used — F.

We construct fourth order methods with third order estimators. FSAL idea is also used.
New methods will be compared to the classic four-stages Runge–Kutta method ("The" Runge–Kutta method) with five-stages third-order estimator, named RK4(3)\(T\) in [2]; and to the same Runge–Kutta method but with an embedded second-order estimator (named in Russian mathematical tradition "Egorov control term" after Prof. Vsevolod A. Egorov [1]) — RK4(2).

### 3 Order conditions

The order conditions for the fourth order method (7)–(8), (10) with use of (9) and standard simplifying conditions

\[
\sum_{e=1}^{w} a_{swve} = c_{sw}
\]  

form the system of 52 nonlinear algebraic equations:

\[
\sum_{w=1}^{3} b_{svw} c_{svw}^{v} = \frac{1}{v + 1}, \quad v = 0, \ldots, 3
\]
\[
\sum_{w=1}^{3} b_{iw} c_i^v \sum_{r=2}^{w} a_{iwir} c_{ir} = \frac{1}{2 \cdot (3 + v)}, \quad v = 0, 1, \quad (13)
\]
\[
\sum_{w=2}^{3} b_{iw} c_{iw} \sum_{r=1}^{w-1} a_{iwjr} c_{jr} = \frac{1}{2 \cdot (3 + v)}, \quad v = 0, 1, \quad (14)
\]
\[
\sum_{w=2}^{3} b_{iw} \sum_{r=2}^{w} a_{iwir} c_{ir}^2 = \frac{1}{12}, \quad \sum_{w=2}^{3} b_{iw} \sum_{r=1}^{w-1} a_{iwjr} c_{jr}^2 = \frac{1}{12}, \quad (15)
\]
\[
\sum_{w=2}^{3} b_{iw} \sum_{r=2}^{w} a_{iwir} a_{irie} c_{ie} = \frac{1}{24}, \quad \sum_{w=2}^{3} b_{iw} \sum_{r=2}^{w} a_{iwir} \sum_{e=1}^{r-1} a_{irje} c_{je} = \frac{1}{24}, \quad (16)
\]
\[
\sum_{w=3}^{3} b_{iw} \sum_{r=2}^{w} a_{iwjr} a_{jrie} c_{ie} = \frac{1}{24}, \quad \sum_{w=3}^{3} b_{iw} \sum_{r=1}^{w-1} a_{iwjr} \sum_{e=1}^{r} a_{jrje} c_{je} = \frac{1}{24}, \quad (17)
\]
\[
\sum_{w=2}^{3} b_{jw} c_{jw}^v \sum_{r=2}^{w} a_{jwir} c_{ir} = \frac{1}{2 \cdot (3 + v)}, \quad v = 0, 1, \quad (18)
\]
\[
\sum_{w=1}^{3} b_{jw} c_{jw}^v \sum_{r=2}^{w} a_{jwjr} c_{jr} = \frac{1}{2 \cdot (3 + v)}, \quad v = 0, 1, \quad (19)
\]
\[
\sum_{w=2}^{3} b_{jw} \sum_{r=2}^{w} a_{jwir} c_{ir}^2 = \frac{1}{12}, \quad \sum_{w=1}^{3} b_{jw} \sum_{r=1}^{w} a_{jwjr} c_{jr}^2 = \frac{1}{12}, \quad (20)
\]
\[
\sum_{w=2}^{3} b_{jw} \sum_{r=2}^{w} a_{jwir} a_{irie} c_{ie} = \frac{1}{24}, \quad \sum_{w=2}^{3} b_{jw} \sum_{r=2}^{w} a_{jwir} \sum_{e=1}^{r-1} a_{irje} c_{je} = \frac{1}{24}, \quad (21)
\]
\[
\sum_{w=2}^{3} b_{jw} \sum_{r=2}^{w} a_{jwjr} a_{jrie} c_{ie} = \frac{1}{24}, \quad \sum_{w=1}^{3} b_{jw} \sum_{r=1}^{w} a_{jwjr} \sum_{e=1}^{r} a_{jrje} c_{je} = \frac{1}{24}, \quad (22)
\]
\[
\sum_{w=1}^{4} d_{sw} c_{sw}^v = \frac{1}{v + 1}, \quad v = 0, \ldots, 2, \quad (23)
\]
\[
\sum_{w=2}^{4} d_{iw} \sum_{r=2}^{w} a_{iwir} c_{ir} = \frac{1}{6}, \quad \sum_{w=2}^{4} d_{iw} \sum_{r=1}^{w-1} a_{iwjr} c_{jr} = \frac{1}{6}, \quad (24)
\]
\[
\sum_{w=2}^{4} d_{jw} \sum_{r=2}^{w} a_{jwjr} c_{jr} = \frac{1}{6}, \quad \sum_{w=2}^{4} d_{jw} \sum_{r=1}^{w} a_{jwjr} c_{ir} = \frac{1}{6}. \quad (25)
\]
Table 2: RKS4(3)4F

<table>
<thead>
<tr>
<th>$c_{is}$</th>
<th>$a_{isiv}$</th>
<th>$a_{isju}$</th>
<th>$b_{is}$</th>
<th>$d_{is}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10} - \frac{2}{5} \xi$</td>
</tr>
<tr>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{1}{2} + \xi$</td>
</tr>
<tr>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5} - \frac{8}{5} \xi$</td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{2} - \frac{2}{5} \xi$</td>
<td>$0$</td>
<td>$\xi$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c_{js}$</th>
<th>$a_{jsiv}$</th>
<th>$a_{jsju}$</th>
<th>$b_{js}$</th>
<th>$d_{js}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{1}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5} + \frac{1}{15} \eta$</td>
</tr>
<tr>
<td>$\frac{2}{3}$</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{2}{3}$</td>
<td>$\frac{2}{3} - \frac{2}{5} \eta$</td>
<td></td>
</tr>
<tr>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6} - \frac{2}{5} \eta$</td>
</tr>
<tr>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
</tr>
</tbody>
</table>

The system consists of three mutually connected blocks of equations. First are the conditions of the fourth order — 28 equations (12)–(22). Second gives the third order of the estimator — 10 equations (23)–(25). The last block (26) consists of 18 restrictions to provide the FSAL idea for the method (7)–(8), (10):

\[
a_{i4iw} = a_{j4iw} = b_{iw}, \quad a_{i4jw} = a_{j4jw} = b_{jw}, \quad w = 1, \ldots, 3, \\
c_{i4} = 1, \quad c_{j4} = 1 + \sigma, \quad a_{i4i4} = a_{i4j4} = 0, \quad a_{j4i4} = a_{j4j4} = \sigma \neq 0.
\]  

The number of unknown parameters is 56.

**Theorem 3.1** There exists an embedded RKS4(3)4F method (7)–(8), (10) that satisfies conditions (9), (26).

The theorem is proved by solving the system (11)–(26). The fourth order method is determined uniquely; the third order estimator exist only when $\sigma = -\frac{1}{6}$ and has two free parameters $d_{i4} \neq 0$ and $d_{j4} \neq 0$. The table 2 contains the parameters of RKS4(3)4F.

The constructed method has better number of stages to order ratio than RK4(3)T and RK4(2).

Often the implementation of Runge–Kutta methods is done so, that the step-size is changed every step. If so, $k_{j1}$ values of RKS4(3)4F cannot be used as $k_{j1}$ of the next step. It reduces the effectiveness of the method, but still it excels both RK4(3)T and RK4(2).
There is an option to control the step-size on base of error estimation of only first $l$ functions (which can work fine when two groups correspond to two different physical values or are scaled differently [1]). The fourth stage for the second group of (5) can be omitted then. Such method RKS4(4)(4,2)F (Table 3) in fact takes only three stages per step, while RK4(3)T and RK4(2) both take four.

Table 3: RKS4(3)(4,3)F

<table>
<thead>
<tr>
<th>$c_{is}$</th>
<th>$a_{isiv}$</th>
<th>$a_{isjv}$</th>
<th>$b_{is}$</th>
<th>$d_{is}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{10} - \frac{7}{6} \xi$</td>
</tr>
<tr>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{3}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>$\frac{5}{6}$</td>
<td>$\frac{1}{24}$</td>
<td>$\frac{5}{8}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{5}{12}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{1}{10}$</td>
<td>$\frac{1}{2}$</td>
<td>$\frac{2}{5}$</td>
<td>$\frac{2}{5}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$c_{js}$</th>
<th>$a_{jsiv}$</th>
<th>$a_{jsjv}$</th>
<th>$b_{js}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{1}{6}$</td>
<td>$\frac{2}{5}$</td>
</tr>
<tr>
<td>$\frac{2}{3}$</td>
<td>$-\frac{1}{12}$</td>
<td>$\frac{3}{4}$</td>
<td>$\frac{1}{2}$</td>
</tr>
<tr>
<td>1</td>
<td>$\frac{3}{4}$</td>
<td>$-\frac{5}{12}$</td>
<td>$\frac{2}{3}$</td>
</tr>
</tbody>
</table>

4 Testing

The constructed methods RKS4(3)4F and RKS4(3)(4,3)F with parameters $\xi = \eta = 3$ were compared to RK4(3)T and RK4(2) solving the problem

$$\begin{aligned}
\begin{cases}
x'' = -2x + \frac{1}{2}y', \\
y'' = -\frac{1}{2}x' - 2y,
\end{cases}
\quad x(0) = x'(0) = 1,
\quad y(0) = 2, 
\quad y'(0) = 3,
\end{aligned}$$

for $t \in [0, 2\pi]$. The general solution of the problem is

$$\begin{aligned}
x(t) &= C_1 \cos(\alpha t) + C_2 \sin(\alpha t) + C_3 \cos(\beta t) + C_4 \sin(\beta t), \\
y(t) &= -C_1 \sin(\alpha t) + C_2 \cos(\alpha t) - C_3 \sin(\beta t) + C_4 \cos(\beta t),
\end{aligned}$$

$$\alpha = \frac{1 - \sqrt{33}}{4}, \quad \beta = \frac{1 + \sqrt{33}}{4}.$$
Since the practical effectiveness of Runge–Kutta methods depends on the algorithm of step-size choosing, and all the methods we compare have the same order properties, all were implemented within a common algorithmic “driver”. The new step-size was determined as $h_{\text{new}} = c \cdot h(\text{tol}/\text{err})^{1/(q+1)}$ with the guarantee factor $c = 0.9$. A series of computations with global error $\delta$ varying from $10^{-6}$ to $10^{-11.5}$ was held.

All four methods passed the convergence test, but it should be noted, that the step-size choosing algorithm was more sensitive to the solution behaviour in case of the new methods RKS4(3)4F, RKS4(3)(4,3)F.

In the table 4 number of necessary steps $N_h$ and the number of right-hand side evaluations $N_g + N_f$ ($N_g$ in the first equation and $N_f$ in the second) over the whole integration interval are shown for the desired norm of the global error.

The results confirm the theoretical expectations. For the same global error RKS4(3)4F and RKS4(3)(4,3)F takes less steps and less right-hand side evaluations than RK4(3)T and RK4(2) do. And vice versa for the same amount of computations RK4(3)T and RK4(2) have larger global error.


delicacies

References


Received: June 5, 2015; Published: July 15, 2015