VARIABLE-STEP SIZE EXPLICIT TWO-STEP RUNGE-KUTTA METHODS

Z. JACKIEWICZ AND M. ZENNARO

Abstract. Variable-step explicit two-step Runge-Kutta methods for the numerical solution of ordinary differential equations are studied. Order conditions are derived and the results about the minimal number of stages required to attain a given order are established up to order five. The existence of embedded pairs of continuous Runge-Kutta methods and two-step Runge-Kutta methods of order \( p - 1 \) and \( p \) is proved. This makes it possible to estimate local discretization error of continuous Runge-Kutta methods without any extra evaluations of the right-hand side of the differential equation. An algorithm to construct such embedded pairs is described, and examples of (3, 4) and (4, 5) pairs are presented. Numerical experiments illustrate that local error estimation of continuous Runge-Kutta methods based on two-step Runge-Kutta methods appears to be almost as reliable as error estimation by Richardson extrapolation, at the same time being much more efficient.

1. Introduction

In this paper we study explicit two-step Runge-Kutta (TSRK) methods for the numerical solution of systems of ordinary differential equations (ODE’s)

\[
\begin{align*}
y'(x) &= f(y(x)), \quad x \in [a, b], \\
y(a) &= y_0,
\end{align*}
\]

where the function \( f: \mathbb{R}^q \to \mathbb{R}^q \) is assumed to be sufficiently smooth. Let a nonuniform grid \( a = x_0 < x_1 < x_2 < \cdots < x_N, \quad x_N \geq b, \) be given, and let \( h_i = x_{i+1} - x_i, \quad i = 0, 1, \ldots, N, \quad \xi = h_i/h_{i-1} \). We consider an \( m \)-stage explicit TSRK method defined on the nonuniform mesh \( \{x_i\}_{i=0}^N \) by the formulas

\[
\begin{align*}
y_{i+1} &= (1 - \xi \theta)y_i + \xi \theta y_{i-1} + h_{i-1} \sum_{j=1}^{m} a_{j}f(Y_{i-1}^{j}) + \xi w_j f(Y_{i}^{j}), \\
y_{i} &= y_{i-1} + h_{i-1} \sum_{s=1}^{j-1} a_{j}f(Y_{i-1}^{s}), \\
Y_{i}^{j} &= y_{i} + \xi h_{i-1} \sum_{s=1}^{j-1} a_{j}f(Y_{i}^{s}),
\end{align*}
\]

\( i = 1, 2, \ldots, N - 1 \). Here, \( y_i \) is an approximation to \( y(x_i) \), where \( y \) is the unique solution to (1.1). The method (1.2) requires initial value \( y_0 \) and starting value \( y_1 \). This value should be computed by some starting procedure.
based on a self-starting method, for example, a Runge-Kutta (RK) method of appropriate order.

It is convenient to represent (1.2) by the following table of coefficients of the TSRK method:

\[
\begin{array}{c|ccccc}
\theta & c_1 & c_2 & c_3 & \ldots & c_m \\
\hline
\theta & v_1 & v_2 & \ldots & v_{m-1} & v_m \\
A & a_{21} & a_{31} & a_{32} & \ldots & a_{mm-1} \\
\theta & w_1 & w_2 & \ldots & w_{m-1} & w_m \\
\end{array}
\]

where \( c_i = \sum_{j=1}^{i-1} a_{ij} \). Here, \( \theta \), \( c_i \), and \( a_{ij} \) are constants, and the weights \( v_i \) and \( w_i \) depend on the ratio of stepsizes \( \xi = h_i/h_{i-1} \). The method (1.2) is constructed in such a way that each step requires only \( m \) new evaluations of the function \( f \) associated with the computation of \( Y_j^i \), \( j = 1, 2, \ldots, m \). Observe that it is not necessary to compute \( Y_{j-1}^i \), \( j = 1, 2, \ldots, m \), at the current step, since these values have already been computed in the previous step.

Explicit TSRK methods have found applications in the numerical solution of systems of ODE's arising from semidiscretizations of parabolic and hyperbolic partial differential equations and have been studied by Byrne and Lambert [3], Renaut [16, 17], and Verwer [18–20]. Related results for explicit \( k \)-step \( m \)-stage RK methods are given in van der Houwen and Sommeijer [7, 8] and van der Houwen [6]. Jackiewicz, Renaut, and Feldstein [10] investigated implicit TSRK methods. Order conditions were formulated and \( A \)-stable two-step two-stage methods of order four were found by extensive computer search.

All the papers mentioned above deal with the methods defined on uniform meshes. However, to implement these methods with variable stepsizes, or to use these methods to estimate local discretization error of continuous RK methods (see §5), it is necessary to consider nonuniform meshes. There exist TSRK methods whose order is lower on nonuniform meshes than on uniform meshes (see [11]). Therefore, we consider in this paper the order conditions for nonuniform meshes. In the next section, the theorem on order conditions for variable-stepsize TSRK methods (1.2) is formulated. This theorem is then used to generate order conditions up to the order five. For the rooted tree \( t \), the order condition involves elementary weights \( \overline{v}^n(t) \) and \( \overline{w}^n(t) \) corresponding to \( v \) and \( w \). For \( \theta = 0 \) and \( v = 0 \), this condition reduces to the order condition (in a slightly different representation than that in [2]) for the RK method:

\[
\begin{array}{c|cc}
\theta & A \\
\hline
\theta & w^T \\
\end{array}
\]

The understanding of this relationship makes it possible to construct TSRK methods (1.2) of any order, based on some simplifying assumptions. This is described in §3. In §4, examples of TSRK methods up to order five are presented. These examples are constructed in such a way that they have embedded within them continuous RK methods of order one less. As a consequence, they provide
error estimates for continuous RK methods, which do not require any extra evaluations of the function \( f \). This is remarkable, since the construction of embedded pairs of continuous RK methods always leads to extra stages (compare \([13-15]\)). In §5 order barriers, i.e., results about the minimum number of stages required to attain a given order for variable-stepsize TSRK methods up to order five, are presented. These results relate the minimal number of stages for TSRK method (1.2) of order \( p \) to the minimal number of stages of continuous RK methods of order \( p - 1 \) for \( p = 2, 3, 4, \) and 5. We also prove a theorem about the existence of a variable-step TSRK method of order \( p \), if a continuous RK method of order \( p - 1 \) is known. This theorem is of great practical value, since it demonstrates the possibility of estimating the local discretization error of continuous RK methods by TSRK methods without any extra evaluations of the function \( f \). Finally, the results of some numerical experiments presented in the technical report \([12]\) are discussed. These experiments indicate, among other things, that local error estimation of continuous RK methods by TSRK methods appears to be almost as reliable as error estimation by Richardson extrapolation, at the same time being much more efficient.

2. Structure of order conditions

Denote by \( \tau \) the unique tree of order one and by \( t = [t_1, t_2, \ldots, t_s] \) the tree formed from the trees \( t_1, t_2, \ldots, t_s \) by adding a new root and joining it with the roots of \( t_1, t_2, \ldots, t_s \). Let \( \rho(t) \) and \( \gamma(t) \) be the order and the density of the tree \( t \) (compare \([2]\)).

To obtain order conditions corresponding to the tree \( t \), we observe that the method (1.2) is a special case of general linear methods introduced by Butcher \([2]\) (compare also \([1]\)). Using the theory of order conditions for general linear methods presented in \([5, \text{pp. 396–399}]\) and the composition theorem for B-series (compare \([5, \text{§11.11}]\)), we obtain the following theorem.

**Theorem 1.** The order condition corresponding to \( \tau \) is
\[
(v^T + \xi w^T)u = \xi + \xi \theta,
\]
and if \( t = [t_1, t_2, \ldots, t_s] \), then the order condition corresponding to it is
\[
v^T \prod_{j=1}^{s} \left( A \overline{w}^v(t_j) + \frac{(-1)^{\rho(t_j)} u}{\gamma(t_j)} \right) + \xi^\rho(t) w^T \prod_{j=1}^{s} (A \overline{w}^w(t_j)) = \xi^\rho(t) - \frac{(-1)^{\rho(t)} \xi \theta}{\gamma(t)},
\]
where \( \overline{w}^v(t_j) \) and \( \overline{w}^w(t_j) \) are elementary weights corresponding to \( v \) and \( w \) and the trees \( t_j, j = 1, 2, \ldots, s \).

In \([12]\) a detailed proof of this theorem is given, based on the results of \([4]\).

This theorem can be used recursively to generate conditions corresponding to trees of any order. For example, if \( t = [\tau^{p-1}] \) (so-called primary tree of order \( \rho(t) = p \)), then \( \gamma(t) = p \), and it follows from Theorem 1 and the relation \( Au = c \) that the order condition corresponding to this tree (so-called primary condition) is
\[
v^T (c - u)^{\rho-1} + \xi^\rho w^T c^{\rho-1} = \frac{\xi^\rho - (-1)^{\rho} \xi \theta}{\rho},
\]
or

\[(v^T + \xi^p w^T)c^{p-1} + v^T \sum_{r=1}^{p-1} (-1)^r \binom{p-1}{r} c^{p-1-r} = \frac{\xi^p - (-1)^p \xi \theta}{p}\]

The conditions up to the order five are listed in Table 1.

**Table 1. Order conditions for variable step TSRK methods**

<table>
<thead>
<tr>
<th>(t)</th>
<th>(p(t))</th>
<th>Order condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>(t_{1,1} = \tau)</td>
<td>1</td>
<td>((v^T + \xi w^T)u = \xi + \xi \theta)</td>
</tr>
<tr>
<td>(t_{2,1} = [\tau])</td>
<td>2</td>
<td>((v^T + \xi^2 w^T)c - v^T u = \frac{\xi^2 - \xi \theta}{2})</td>
</tr>
<tr>
<td>(t_{3,1} = [\tau^2])</td>
<td>3</td>
<td>((v^T + \xi^3 w^T)c^2 - v^T (2c - u) = \frac{\xi^3 + \xi \theta}{3})</td>
</tr>
<tr>
<td>(t_{3,2} = [2\tau]_2)</td>
<td>3</td>
<td>((v^T + \xi^3 w^T)Ac - v^T \left(c - \frac{1}{2}u\right) = \frac{\xi^3 + \xi \theta}{6})</td>
</tr>
<tr>
<td>(t_{4,1} = [\tau^3])</td>
<td>4</td>
<td>((v^T + \xi^4 w^T)c^3 - v^T (3c^2 - 3c + u) = \frac{\xi^4 - \xi \theta}{4})</td>
</tr>
<tr>
<td>(t_{4,2} = [\tau])</td>
<td>4</td>
<td>((v^T + \xi^4 w^T)(c \cdot Ac) - v^T \left(c^2 + Ac - \frac{3}{2}c + \frac{1}{2}u\right) = \frac{\xi^4 - \xi \theta}{8})</td>
</tr>
<tr>
<td>(t_{4,3} = [2\tau^2]_2)</td>
<td>4</td>
<td>((v^T + \xi^4 w^T)Ac^2 - v^T \left(2Ac - c\frac{1}{3}u\right) = \frac{\xi^4 - \xi \theta}{12})</td>
</tr>
<tr>
<td>(t_{4,4} = [3\tau]_3)</td>
<td>4</td>
<td>((v^T + \xi^4 w^T)A^2c - v^T \left(Ac - \frac{1}{2}c + \frac{1}{6}u\right) = \frac{\xi^4 - \xi \theta}{24})</td>
</tr>
<tr>
<td>(t_{5,1} = [\tau^4])</td>
<td>5</td>
<td>((v^T + \xi^5 w^T)c^4 - v^T (4c^3 - 6c^2 + 4c - u) = \frac{\xi^5 + \xi \theta}{5})</td>
</tr>
<tr>
<td>(t_{5,2} = [\tau^2][\tau])</td>
<td>5</td>
<td>((v^T + \xi^5 w^T)(c^2 \cdot Ac) - v^T \left(c^3 + 2c \cdot Ac - \frac{5}{2}c^2 - Ac + 2c - \frac{1}{2}u\right) = \frac{\xi^5 + \xi \theta}{10})</td>
</tr>
</tbody>
</table>
Order conditions for TSRK methods can also be generated using the approach by Burrage [1].

When \( \theta = 0 \) and \( v = 0 \), these order conditions reduce to the order conditions for the \( m \)-stage RK method

\[
\begin{align*}
\frac{c}{w^T} &\quad \frac{A}{w^T} \\
\end{align*}
\]

(in a slightly different representation than that in [2]).

Denote by \( T_r \) the set of rooted trees of order \( r \). Then \( T = \bigcup_{r=1}^{\infty} T_r \). The set \( T_r \) has \( n_r \) elements; they will be denoted by \( t_{r,1} , t_{r,2} , \ldots , t_{r,n_r} \). The following result follows from the theorem on the composition of B-series (compare [5, Theorem II.11.6, p. 245]). The details of the proof are given in [12].
Theorem 2. The following relation holds:
\[
\psi^u(t) = \psi^{\bar{w}}(t) + \sum_{r=1}^{\rho(t)-1} \sum_{l=1}^{n_r} \alpha_{r,l}^{(t)} \psi^{u}(t_r,l),
\]
where
\[
\sum_{l=1}^{n_r} \frac{\alpha_{r,l}^{(t)}}{\gamma(t_r,l)} = (-1)^{\rho(t)-r} \frac{\rho(t)}{r\gamma(t)} \left( \frac{\rho(t) - 1}{r - 1} \right),
\]
r = 1, 2, ..., \rho(t) - 1. □

Theorem 2 plays a fundamental role in justifying the construction of TSRK methods described in the next section.

3. Construction of TSRK methods

Denote by \( T_P \) the set of all primary trees defined by
\[
T_P = \{ t \in T : t = [t_r], r = 0, 1, \ldots \},
\]
where \([t_0] := t\). The set \( T_S = T - T_P \) is called the set of secondary trees. We will call the order conditions corresponding to the trees \( t \in T_P \) the primary conditions, and the order conditions corresponding to \( t \in T_S \) the secondary conditions. Let \( e^{(1)} = [1, 0, \ldots, 0] \in \mathbb{R}^m \). We have the following theorem.

Theorem 3. Assume that \( v + \xi^p w = f(\xi)e^{(1)} \), and that \( v \) satisfies all the secondary conditions up to order \( p - 1 \) of the one-step RK method with modified right-hand side as follows:
\[
v_{T} c^{p-1} + \xi c^{p-1} w = f_{p}(\xi)e^{(1)},
\]
t \( \in T_S \), \( \rho(t) \leq p - 1 \). Then all the secondary conditions up to order \( p \) reduce to primary conditions of the same order.

Proof. The assertion follows from Theorems 1 and 2, using induction on the order of the trees. The details of the proof are given in [12]. □

Theorem 3 leads to the following algorithm for constructing TSRK methods of order \( p \geq 2 \). Let
\[
v + \xi^p w = f_p(\xi)e^{(1)},
\]
where \( f_p(\xi) \) is a function which will be determined later. Consider the systems
\[
\begin{align*}
v^T c^{p-1} + \xi^p w^T c^{p-1} & = f_p(\xi)\delta_{p-r}, \quad r = 1, p - 1, \\
v^T c^{p-1} + \xi \rho w^T c^{p-1} & = \frac{\xi p - (-1)^{p-1} p \xi \theta}{\rho} - v^T \sum_{r=1}^{p-1} (-1)^r \left( \frac{p - 1}{r} \right) c^{p-1-r},
\end{align*}
\]
\( p = 1, 2, \ldots, p - 1 \), where the first equation of (3.2) is obtained by multiplying (3.1) by \( c^{p-1} \), and the second equation of (3.2) is the primary condition of order \( p \). This gives
\[
v^Tc^{p-1} = \left[\xi^{p-\rho} \left(\frac{\xi^{p} - (-1)^{p-1}\xi\theta}{\rho} \right) \sum_{r=1}^{p-1} (-1)^r \left(\frac{p-1}{r}\right) v^Tc^{p-1-r} \right] - f_p(\xi)\delta_{p-p+1} \right] / (\xi^{p-\rho} - 1),
\]

which express \( v^Tc^{p-1} \) in terms of \( f_p(\xi) \) and \( v^Tc^r \) for \( r = 0, 1, \ldots, p-2 \). We can compute \( f_p(\xi) \) substituting (3.3) into the primary condition of order \( p \),

\[
\sum_{r=1}^{p-1} (-1)^r \left(\frac{p-1}{r}\right) v^Tc^{p-1-r} = \frac{\xi^p - (-1)^p\xi\theta}{p}.
\]

Substituting \( f_p(\xi) \) computed from the above equation back into (3.3), we can express \( v^Tc^{p-1} \) in the form

\[
v^Tc^{p-1} = f^p_{p-1}(\xi),
\]

\( p = 1, 2, \ldots, p-1 \), where \( f^p_{p-1}(\xi) \) are functions of \( \xi \) only. The function \( f_p(\xi) \) and the functions \( f^p_{p-1}(\xi) \) computed by the above algorithm for \( \theta = 0 \) are listed below for \( p = 1, 2, \ldots, p-1 \) and \( p = 2, 3, 4, 5 \).

\( p = 2 \):

\[
\begin{align*}
f_2(\xi) &= \frac{1}{2} \xi^2 (\xi + 1), \\
g_2(\xi) &= -\xi^2 / 2;
\end{align*}
\]

\( p = 3 \):

\[
\begin{align*}
f_3(\xi) &= \frac{1}{2} \xi^3 (\xi^2 + \xi + 1), \\
f_2(\xi) &= g_3(\xi)(\xi + 2), \\
g_2(\xi) &= g_3(\xi)(\xi + \frac{3}{2}), \\
g_3(\xi) &= -\xi^3 / (3(\xi + 1));
\end{align*}
\]

\( p = 4 \):

\[
\begin{align*}
f_4(\xi) &= \frac{1}{4} \xi^4 (\xi^3 + \xi^2 + \xi + 1), \\
f_3(\xi) &= g_4(\xi)(\xi^2 + 2\xi + 3), \\
g_3(\xi) &= g_4(\xi)(\xi^2 + 2\xi + 2), \\
f_2(\xi) &= g_4(\xi)(\xi^2 + \frac{3}{2}), \\
g_2(\xi) &= g_4(\xi)(\xi^2 + \frac{3}{2}), \\
g_3(\xi) &= g_4(\xi)(\xi^2 + \frac{3}{2}), \\
g_4(\xi) &= -\xi^4 / (4(\xi^2 + \xi + 1));
\end{align*}
\]

\( p = 5 \):

\[
\begin{align*}
f_5(\xi) &= \frac{1}{2} \xi^5 (\xi^4 + \xi^3 + \xi^2 + \xi + 1), \\
g_5(\xi) &= g_5(\xi)(\xi^5 + 3\xi^4 + 6\xi^3 + 9\xi^2 + 7\xi + 4), \\
g_5(\xi) &= g_5(\xi)(\xi^5 + 3\xi^4 + 6\xi^3 + \frac{15}{2}\xi^2 + 5\xi + \frac{5}{2}), \\
g_5(\xi) &= g_5(\xi)(\xi^5 + 3\xi^4 + \frac{15}{4}\xi^3 + \frac{17}{3}\xi^2 + \frac{10}{3}\xi + \frac{5}{3}), \\
g_5(\xi) &= g_5(\xi)(\xi^5 + \frac{11}{4}\xi^4 + \frac{17}{4}\xi^3 + \frac{4}{3}\xi^2 + \frac{9}{4}\xi + \frac{5}{4}), \\
g_5(\xi) &= -\xi^5 / (5(\xi^2 + \xi + 1)(\xi^3 + \xi^2 + \xi + 1)).
\end{align*}
\]
4. Examples of TSRK methods and embedded pairs of continuous RK methods and TSRK methods

In this section we construct TSRK methods of order \( p = 2, 3, 4, \) and \( 5. \) It will always be assumed that \( \theta = 0. \) The coefficient matrix \( A \) and the vector \( c = Au \) are chosen to correspond to a continuous RK method of order \( p - 1. \) As a result, we obtain embedded pairs of a continuous RK method of order \( p - 1 \) and a TSRK method of order \( p. \) This pair will be denoted by the following table of coefficients:

<table>
<thead>
<tr>
<th>( c )</th>
<th>( A )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( c_1 = 0 )</td>
<td>( a_{21} )</td>
</tr>
<tr>
<td>( c_2 )</td>
<td>( a_{31} )</td>
</tr>
<tr>
<td>( \vdots )</td>
<td>( a_{m1} )</td>
</tr>
<tr>
<td>( c_m )</td>
<td>( a_{m2} )</td>
</tr>
</tbody>
</table>

Here, \( b_i(\eta), \ i = 1, 2, \ldots, m, \) are continuous weights of RK methods, and \( v_i(\xi) \) and \( w_i(\xi), \ i = 1, 2, \ldots, m, \) are weights of a TSRK method.

Consider the local solution \( u(t) \) defined by

\[
\begin{align*}
\cases{ u'(t) = f(u(t)), \\ u(t_i) = y_i, } \\
\end{align*}
\]

\( t \in [t_{i-1}, t_{i+1}] \)

Then the local error \( l_{i+1} \) of the RK method at \( t_{i+1} \) is given by \( l_{i+1} = u(t_{i+1}) - y_{i+1}, \) where

\[
(4.1) \begin{align*}
y_{i+1} &= y_i + h_i \sum_{j=1}^{m} b_j(1)f(Y_i^j), \\
Y_i^j &= y_i + h_i \sum_{s=1}^{j-1} a_jsf(Y_s^j), \\
j = 1, 2, \ldots, m. 
\end{align*}
\]

Denote by \( \hat{y}_{i+1} \) the numerical approximation computed by the TSRK method

\[
(4.2) \begin{align*}
\hat{y}_{i+1} &= y_i + h_{i-1} \sum_{j=1}^{m} (v_j f(Y_{i-1}^j) + \xi w_j f(Y_i^j)), \\
Y_{i-1}^j &= y_{i-1} + h_{i-1} \sum_{s=1}^{j-1} a_jsf(Y_{i-1}^s), \\
Y_{i}^j &= y_{i} + \xi h_{i-1} \sum_{s=1}^{j-1} a_jsf(Y_{i-1}^s), \\
j = 1, 2, \ldots, m. 
\end{align*}
\]

We have the following theorem.

**Theorem 4.** Assume that the RK method \((4.1)\) has order \( p - 1 \) and the TSRK method \((4.2)\) has order \( p. \) Then

\[
(4.3) \begin{align*}
l_{i+1} &= \hat{y}_{i+1} - y_{i+1} + O(h^{p+1}) \\
\end{align*}
\]

as \( h \to 0, \) where \( h = \max h_i. \)
Proof. Define \( y_{i+1}^* \) by

\[
\begin{align*}
y_{i+1}^* &= y_i + h_{i-1} \sum_{j=1}^{m}(v_j f(U^j_{i-1}) + \xi w_j f(Y^j_i)), \\
U^j_{i-1} &= u(t_{i-1}) + h_{i-1} \sum_{s=1}^{j-1} a_{js} f(U^j_{i-1}), \\
Y^j_i &= y_i + \xi h_{i-1} \sum_{s=1}^{j-1} a_{js} f(Y^j_s),
\end{align*}
\]

for \( j = 1, 2, \ldots, m \), where \( u(t_{i-1}) \) is the local solution at \( t_{i-1} \). In general, \( u(t_{i-1}) \neq y_{i-1} \). However, we can view \( y_{i-1} \) as computed from \( y_i \) by the reflection of the method (4.1), and since the reflection of an RK method of order \( p - 1 \) is also of order \( p - 1 \) (compare Theorem 344D in [2]), we have \( u(t_{i-1}) = y_{i-1} + O(h^p), \ h \to 0 \). Since,

\[
Y^j_i - U^j_i = h_{i-1} \sum_{s=1}^{j-1} a_{js} (f(Y^j_s) - f(U^j_s)) + O(h^p),
\]

it is easy to see that \( Y^j_i - U^j_i = O(h^p) \) and \( f(Y^j_i) - f(U^j_i) = O(h^p) \) as \( h \to 0 \). Hence,

\[
y_{i+1}^* - y_{i+1} = h_{i-1} \sum_{j=1}^{m} v_j (f(Y^j_i) - f(U^j_i)) = O(h^{p+1})
\]

as \( h \to 0 \). The TSRK method has order \( p \), and we have also

\[
y_{i+1}^* - u(t_{i+1}) = O(h^{p+1}) \quad \text{as} \ h \to 0.
\]

Therefore,

\[
y_{i+1}^* - y_{i+1} = \dot{y}_{i+1} - y_{i+1}^* + y_{i+1}^* - u(t_{i+1}) + u(t_{i+1}) - y_{i+1} = l_{i+1} + O(h^{p+1}) \quad \text{as} \ h \to 0,
\]

and the proof is complete. \( \square \)

Remark. Observe that if \( \theta \neq 0 \), the above perturbation arguments are not valid. In this case,

\[
\begin{align*}
\dot{y}_{i+1} &= (1 - \theta)y_{i-1} + \theta y_i + h_{i-1} \sum_{j=1}^{m}(v_j f(Y^j_i) + \xi w_j f(Y^j_i)), \\
y_{i+1}^* &= (1 - \theta)u(t_{i-1}) + \theta y_i + h_{i-1} \sum_{j=1}^{m}(v_j f(U^j_{i-1}) + \xi w_j f(Y^j_i)),
\end{align*}
\]

and we can only conclude that \( y_{i+1}^* - \dot{y}_{i+1} = O(h^p) \) as \( h \to 0 \).

In what follows we will list specific examples of embedded pairs of continuous RK methods of order \( p - 1 \) and TSRK methods of order \( p \), \( p = 2, 3, 4, \) and \( 5 \), constructed using the algorithm described in §3.
Example 1. Pair of order \((1, 2)\), \(m = 1\):
\[
\begin{array}{c|ccc}
0 & 0 \\
\hline
\xi & \frac{\eta}{2} & -
\end{array}
\]

Example 2. Pair of order \((2, 3)\), \(m = 2\):
\[
\begin{array}{c|ccc}
0 & c & c \\
\hline
b_1(\eta) & b_2(\eta) & v_1(\xi) & v_2(\xi) \\
0 & w_1(\xi) & w_2(\xi)
\end{array}
\]

Here, \(c \neq 0\), the \(b_i(\eta)\) satisfy order conditions for the continuous RK method (see [20]), and \(v_i(\xi)\) and \(w_i(\xi)\) satisfy the system (3.1) and (3.4). This gives
\[
\begin{align*}
b_1(\eta) &= \eta - \frac{\eta^2}{2c}, \\
b_2(\eta) &= \frac{\eta^2}{2c}, \\
v_1(\xi) &= \frac{\xi^3(2(1 - c)\xi + 3 - 4c)}{6c(\xi + 1)}, \\
v_2(\xi) &= -\frac{\xi^3(2\xi + 3)}{6c(\xi + 1)}, \\
w_1(\xi) &= \frac{2c\xi^3 + 4c\xi^2 + 2(3c - 1)\xi + 6c + 3}{6c(\xi + 1)}, \\
w_2(\xi) &= \frac{2\xi + 3}{6c(\xi + 1)}.
\end{align*}
\]

Example 3. Pair of order \((3, 4)\), \(m = 4\): Consider the RK method of order three with three stages:
\[
\begin{array}{c|ccc}
0 & c_2 & a_{21} \\
c_3 & a_{31} & a_{32} \\
\hline
b_1 & b_2 & b_3
\end{array}
\]

\(a_{21} = c_2, \ a_{31} + a_{32} = c_3\). These methods are characterized in Butcher [2]. Assuming that \(0, c_2, c_3\) are all different, and that \(c_2 \neq \frac{2}{3}\), which corresponds to Case 1 in [2], we obtain for the coefficients of this method
\[
\begin{align*}
b_2 &= \frac{c_3 - \frac{2}{3}}{2c_2(c_3 - c_2)}, \\
b_3 &= \frac{\frac{2}{3} - c_2}{2c_3(c_3 - c_2)}, \\
b_1 &= 1 - b_2 - b_3, \\
a_{32} &= \frac{c_3(c_3 - c_2)}{c_2(2 - 3c_2)}.
\end{align*}
\]
This method can be extended to a 4-stage RK method with stage reuse (compare [5]):

\[
\begin{array}{cccc}
0 & a_{21} & & \\
c_2 & a_{31} & a_{32} & \\
c_3 & b_1 & b_2 & b_3 \\
1 & b_1 & b_2 & b_3 & 0
\end{array}
\]

which has a continuous extension with weights given by (see [13])

\[
\begin{align*}
b_1(\eta) &= (1 - 2b_1)\eta^3 + (3b_1 - 2)\eta^2 + \eta, \\
b_2(\eta) &= -b_2(2\eta^3 - 3\eta^2), \\
b_3(\eta) &= -b_3(2\eta^3 - 3\eta^2), \\
b_4(\eta) &= \eta^3 - \eta^2.
\end{align*}
\]

Observe that this continuous extension is a Hermite interpolant at the endpoints \(t_i\) and \(t_{i+1}\).

Choosing \(c_2 = \frac{1}{2}\) and \(c_3 = 1\) leads to the following (3, 4) pair:

\[
\begin{array}{cccc}
0 & \frac{1}{2} & 2 & 1 \\
\frac{1}{2} & 1 & \frac{2}{3} & \frac{1}{6} \\
1 & \frac{1}{6} & b_1(\eta) & b_2(\eta) \\
& & b_3(\eta) & b_4(\eta) \\
& & v_1(\xi) & v_2(\xi) \\
& & v_3(\xi) & v_4(\xi) \\
0 & w_1(\xi) & w_2(\xi) & w_3(\xi) \\
& & w_4(\xi)
\end{array}
\]

Here,

\[
\begin{align*}
b_1(\eta) &= \frac{1}{2}\eta^3 - \frac{3}{2}\eta^2 + \eta, \\
b_2(\eta) &= -\frac{4}{3}\eta^3 + 2\eta^2, \\
b_3(\eta) &= -\frac{1}{2}\eta^3 + \frac{1}{2}\eta^2, \\
b_4(\eta) &= \eta^3 - \eta^2,
\end{align*}
\]

and

\[
\begin{align*}
v_1(\xi) &= \frac{\xi^4(2\xi + 1)}{12(\xi^2 + \xi + 1)}, \\
v_2(\xi) &= -\frac{\xi^4(\xi + 2)}{3(\xi^2 + \xi + 1)}, \\
v_3(\xi) &= -\frac{\xi^4(\xi + 2)}{12(\xi^2 + \xi + 1)}, \\
v_4(\xi) &= -\frac{\xi^4(\xi^2 + \xi)}{4(\xi^2 + \xi + 1)},
\end{align*}
\]
\begin{align*}
  w_1(\xi) &= \frac{1}{4}(\xi^3 + \xi^2 + \xi + 1) - \frac{2\xi + 1}{12(\xi^2 + \xi + 1)}, \\
  w_2(\xi) &= -\frac{\xi + 2}{3(\xi^2 + \xi + 1)}, \\
  w_3(\xi) &= -\frac{\xi + 2}{12(\xi^2 + \xi + 1)}, \\
  w_4(\xi) &= -\frac{\xi^2 + \xi}{4(\xi^2 + \xi + 1)}.
\end{align*}

**Example 4.** Pair of order \((4, 5), m = 6\): Assuming that \(A\) and \(c\) correspond to the 6-stage continuous RK method of order four with stage reuse given in [13, Example 2.21] leads to the embedded pair

\[
\begin{array}{cccccc}
0 & 1 & 2 & 3 & 4 & 5 & 6 \\
\frac{1}{2} & \frac{1}{2} & 0 & -1 & 2 & 0 & 0 \\
\frac{1}{4} & \frac{1}{4} & 0 & \frac{9}{16} & 0 & 0 & 0 \\
1 & 0 & 0 & \frac{2}{3} & \frac{1}{6} & 0 & 0 \\
\end{array}
\]

Here,

\[
\begin{align*}
  b_1(\eta) &= \eta(-\frac{3}{2}\eta^3 + 2\eta^2 - \frac{13}{6}\eta + 1), \\
  b_2(\eta) &= 0, \\
  b_3(\eta) &= \eta^2(4\eta^2 - \frac{28}{3}\eta + 6), \\
  b_4(\eta) &= \eta^2(\eta^2 - \frac{3}{2}\eta + \frac{1}{2}), \\
  b_5(\eta) &= \eta^2(-\frac{16}{3}\eta^2 + \frac{32}{3}\eta - \frac{16}{3}), \\
  b_6(\eta) &= \eta^3(\eta - 1),
\end{align*}
\]

and

\[
\begin{align*}
  v_1(\xi) &= g_5(\xi)\left(\frac{2}{3}\xi^4 + \frac{2}{3}\xi^3 - \frac{1}{6}\xi^2 - \frac{2}{3}\xi - \frac{1}{6}\right), \\
  v_2(\xi) &= 0, \\
  v_3(\xi) &= g_5(\xi)\left(-4\xi^4 - \frac{28}{3}\xi^3 - \frac{14}{3}\xi^2 + \frac{8}{3}\xi + \frac{10}{3}\right), \\
  v_4(\xi) &= \frac{1}{4}v_3(\xi), \\
  v_5(\xi) &= g_5(\xi)\left(\frac{16}{3}\xi^4 + 16\xi^3 + 16\xi^2 + \frac{16}{3}\xi\right), \\
  v_6(\xi) &= g_5(\xi)\left(\xi^5 + 2\xi^4 + \xi^3 - \xi^2 - \xi\right).
\end{align*}
\]
where
\[ g_5(\xi) = -\frac{\xi^5}{5(\xi^2 + \xi + 1)(\xi^3 + \xi^2 + \xi + 1)} \, . \]

We have also
\[ w_1(\xi) = (f_5(\xi) - v_1(\xi))/\xi^5, \quad w_i(\xi) = -v_i(\xi)/\xi^5, \quad i = 2, 3, 4, 5, 6. \]

5. Order barriers for TSRK methods

In this section we investigate the minimum number of stages of variable-step explicit TSRK methods required to attain a given order \( p \). This number will be denoted by \( TVSEN(p) \). Obviously, \( TVSEN(p) \leq EN(p) \), where \( EN(p) \) is the minimum number of stages for a RK method of order \( p \). It turns out that for low orders \( p \leq 5 \) this number is closely related to the minimum number of stages of continuous explicit RK methods required to attain order \( p - 1 \). This number will be denoted by \( CEN(p - 1) \). It was proved by Owren and Zennaro [15] (see also [14]) that
\[ CEN(1) = 1, \quad CEN(2) = 2, \quad CEN(3) = 4, \quad CEN(4) = 6, \quad CEN(5) = 8. \]

We have the following theorem.

**Theorem 5.** The minimum number of stages for a variable-step explicit TSRK method required to attain a given order \( p \) satisfies
\[ TVSEN(p) = CEN(p - 1) \]
for \( p = 2, 3, 4, \) and
\[ TVSEN(5) \leq CEN(4) = 6. \]

**Proof.** The examples of TSRK methods presented in the previous section show that
\[ TVSEN(p) \leq CEN(p - 1), \]
\( p = 2, 3, 4, \) and 5, and obviously,
\[ TVSEN(2) = CEN(1) = 1. \]

It is easy to check that the system of order conditions up to order three for a TSRK method cannot be satisfied for \( m = 1 \). Hence,
\[ TVSEN(3) = CEN(2) = 2. \]

Consider now the case \( p = 4 \). We have to show that \( TVSEN(4) = 4 \). Assume to the contrary that \( TVSEN(4) < 4 \). It can be shown that the simplifying assumptions \( u + \varepsilon^4 w = f(\xi)e^{(1)} \) and \( v^T Ac = \frac{1}{2} v^T c^2 \) (compare Theorem 3) are, in fact, necessary for a TSRK method of order four with three stages. This requires some tedious algebra, and for the sake of brevity the details of the derivation are omitted. The interested reader is referred to [12].
The vector $v$ satisfies the system of equations

$$
\begin{align*}
v^Tu &= g_4(\xi)(\xi^2 + 2\xi + 3), \\
v^Tc &= g_4(\xi)(\xi^2 + 2\xi + 2), \\
v^Tc^2 &= g_4(\xi)(\xi^2 + \frac{2}{3}\xi + \frac{4}{5}), \\
v^TAc &= \frac{1}{2}g_4(\xi)(\xi^2 + \frac{2}{3}\xi + \frac{4}{5}), \\
g_4(\xi) &= -\xi^4/(4(\xi^2 + \xi + 1)),
\end{align*}
$$

compare §3. Observe that the first three equations correspond to primary conditions, and the fourth equation to the secondary condition. Let

$$
v_i(\xi) = g_4(\xi)(v_{i,0} + v_{i,1}\xi + v_{i,2}\xi^2),
$$

$i = 1, 2, 3$. Then the above system of order conditions can be written in the form

$$
\begin{bmatrix}
\Phi_p \\
\Phi_s
\end{bmatrix} V = \begin{bmatrix}
F_p \\
F_s
\end{bmatrix},
$$

where $V = [v_{i,j}]_{i=1, j=0}^{3, 2}$ and

$$
\begin{align*}
\Phi_p &= \begin{bmatrix}
u^T \\
c^T \\
(c^2)^T
\end{bmatrix}, \\
F_p &= \begin{bmatrix}3 & 2 & 1 \\2 & 2 & 1 \\\frac{4}{5} & \frac{4}{5} & 1\end{bmatrix}, \\
\Phi_s &= (Ac)^T, \\
F_s &= \begin{bmatrix}\frac{3}{2} & \frac{5}{6} & \frac{1}{2}\end{bmatrix}.
\end{align*}
$$

Multiplying (5.1) from the right by the matrix $Q$ given by

$$
Q = \begin{bmatrix}1 & -\frac{1}{2} & 0 \\
-2 & \frac{5}{2} & -1 \\
2 & -\frac{7}{2} & 2
\end{bmatrix},
$$

and putting $B = VQ$, leads to the system

$$
\begin{bmatrix}
\Phi_p \\
\Phi_s
\end{bmatrix} B = \begin{bmatrix}
F_p Q \\
F_s Q
\end{bmatrix}.
$$

Since

$$
F_p Q = \begin{bmatrix}1 & 0 & 0 \\
0 & \frac{1}{2} & 0 \\
0 & 0 & \frac{1}{3}
\end{bmatrix}
$$

and $F_s Q = [0, 0, 1/6]$, the system (5.2) has exactly the form encountered in the theory of explicit continuous RK methods of order 3 (compare [13, 15]). Since it is known that the minimal number of stages for an explicit continuous RK method of order three is four (see [13–15]), it follows that the system (5.2), and as a consequence, the system (5.1), cannot have a solution. This completes the proof. □

Our next result shows that, using the algorithm based on Theorem 3, we cannot obtain a variable-step explicit TSRK method of order five with fewer than six stages.

**Theorem 6.** Assume that $v + \xi^5w = f(\xi)e^{(1)}$ and $v^T\overline{w}^u(t) = \rho(t)v^Tc^{\rho(t)-1}/\gamma(t)$, $t \in T_1$, $\rho(t) \leq 4$. Then the minimal number of stages of the TSRK method (1.2) of order five is $m = 6$.

License or copyright restrictions may apply to redistribution; see http://www.ams.org/journal-terms-of-use
Proof. The proof of this theorem is similar to the proof of Theorem 5 and is therefore omitted. The interested reader is referred to [12]. ⊓⊔

Remark. Theorem 6 does not rule out the existence of a TSRK method of order five with fewer than six stages. It only says that such methods cannot satisfy the simplifying conditions given in Theorem 3.

In the report [12] we also investigated the necessity of the simplifying conditions given in Theorem 3 for TSRK methods of order five. It was proved that

\[ v + \xi^5 w = f(\xi)e^{(1)} \]

holds if and only if

\[ \text{rank}(M) = m - 1, \]

where the matrix \( M \) is given by

\[
M = \begin{bmatrix}
(A^2c^2 - 2A^3c)^T \\
(Ac^3 + 3(Ac)^2 - 3c \cdot Ac^2)^T \\
((Ac)^2 - 2c \cdot Ac^2 + A^2c^2)^T \\
(c^4 - 4Ac \cdot c^2 + 4(Ac)^2)^T \\
(2Ac^3 - 6A(c \cdot Ac) + 3A^2c^2)^T
\end{bmatrix}.
\]

This matrix corresponds to a subset of conditions of order five. We also proved that if (5.3) is satisfied, then the conditions

\[ v^T \psi_w(t) = \frac{\rho(t)}{\gamma(t)} c^{\rho(t)-1}, \]

\( t \in T_s, \rho(t) \leq 4 \), are automatically satisfied.

Summing up the above discussion, we conclude that the conditions (5.3) and (5.5) are necessarily satisfied for a TSRK method of order five with \( m \leq 6 \) if and only if (5.4) holds. Using this fact, and Theorem 6, we can assert that if a TSRK of order five with \( m < 6 \) exists, then

\[ \text{rank}(M) < m - 1. \]

Unfortunately, this condition is difficult to analyze.

We conclude this section with the following theorem.

Theorem 7. For each continuous \( m \)-stage RK method of order \( p - 1 \geq 2 \), there exists an \( m \)-stage variable-step TSRK method of order \( p \) with the same coefficient matrix \( A \).

Proof. Put \( B = [b_{i,j}]_{i=1, j=1}^{m, p-1} \), where the \( b_{i,j} \) are the coefficients of the continuous weight \( b_i(\eta) \), i.e.,

\[ b_i(\eta) = \sum_{j=1}^{p-1} b_{i,j} \eta^j. \]

It was demonstrated by Owren and Zennaro [15] (see also [13]) that the order conditions for an \( m \)-stage continuous RK method of order \( p - 1 \) can be written in the form

\[ \begin{bmatrix}
\Phi_p \\
\Phi_S
\end{bmatrix} B = \begin{bmatrix}
G_p \\
G_S
\end{bmatrix}, \]

where
where

\[ \Phi_p = \begin{bmatrix} u^T & c^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (c^{p-2})^T \end{bmatrix}, \quad G_p = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \frac{1}{2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{p-1} \end{bmatrix}, \]

\[ \Phi_S = \left[ (w^u(t))^T \right]_{t \in T_S, \rho(t) \leq p-1}, \quad G_S = \left[ e^{(\rho(t))/\gamma(t)} \right]_{t \in T_S, \rho(t) \leq p-1}, \]

and \( e^{(\rho(t))} = [0, \ldots, 0, 1, 0, \ldots, 0] \) with 1 in position \( \rho(t) \).

Consider now the system of equations for the vector \( v \) corresponding to a TSRK method of order \( p \). The first part of the system corresponds to primary conditions

\[ v^T c^{i-1} = f^p_i(\xi), \quad i = 1, 2, \ldots, p-1, \]

where the functions \( f^p_i(\xi), \) \( i = 1, 2, \ldots, p-1, \) can be computed by the algorithm based on Theorem 3 and described in §3. These functions always have the form

\[ f^p_i(\xi) = g_p(\xi)r^p_i(\xi), \]

where \( g_p(\xi) \) is a rational function and the \( r^p_i(\xi) \) are polynomials of degree \( k \geq p - 2 \). The second part of the system corresponding to the secondary conditions reads

\[ v^T w^u(t) = \frac{\rho(t)}{\gamma(t)} f^p_\rho(t)(\xi), \quad t \in T_S, \rho(t) \leq p-1. \]

Let

\[ v_i(\xi) = g_p(\xi) \sum_{j=0}^k v_{i,j} \xi^j, \quad i = 1, 2, \ldots, m, \]

\[ f^p_i(\xi) = g_p(\xi) \sum_{j=0}^k f^p_{i,j} \xi^j, \quad i = 1, 2, \ldots, p-1, \]

and denote by \( f^p_i \) the row vector \([f^p_{i,j}]_{j=0}^k \). Then the system of order conditions can be written in the form

\[ (5.7) \quad \begin{bmatrix} \Phi_p \\ \Phi_S \end{bmatrix} V = \begin{bmatrix} F_p \\ F_S \end{bmatrix}, \]

where

\[ V = [v_{i,j}]_{i=1,j=0}^m, \quad F_p = [f^p_{i,j}]_{i=1,j=0}^{p-1}, \quad F_S = \left[ \frac{\rho(t)}{\gamma(t)} f^p_\rho(t) \right]_{t \in T_S, \rho(t) \leq p-1}. \]

Let \( \tilde{G}_p \) and \( \tilde{G}_S \) be \((p - 1) \times (k + 1)\) matrices defined by

\[ \tilde{G}_p = [G_p|0], \quad \tilde{G}_S = [G_S|0], \]

and let us partition the matrix \( F_p \) into

\[ F_p = [F_p^1|F_p^2], \]

where \( F_p^1 \) stands for the first \( p - 1 \) columns of \( F_p \), and \( F_p^2 \) for the last \( k - p + 2 \) columns of \( F_p \). Define the matrix \( T \) by
where $T_1 = G_p^{-1} F_p$ and $T_2 = G_p^{-1} F_p^2$, and where 0 stands for the $(k-p+2) \times (p-1)$ zero matrix and $I$ for the $(k-p+2) \times (k-p+2)$ identity matrix. Let $\bar{B} = [B|0]$, where 0 stands for the $m \times (k+1)$ zero matrix. Then the system (5.6) can be written as

\begin{equation}
\begin{pmatrix}
\Phi_p \\
\Phi_S
\end{pmatrix}
\bar{B} =
\begin{pmatrix}
\tilde{G}_p \\
\tilde{G}_S
\end{pmatrix}.
\end{equation}

We have $\tilde{G}_p T = F_p$, and since

\[
\frac{1}{\gamma(t)} e^{\rho(t)} T = \begin{bmatrix}
\frac{1}{\gamma(t)} e^{\rho(t)} & 0
\end{bmatrix} = \frac{\rho(t)}{\gamma(t)} \rho^{p},
\]

it follows that $\tilde{G}_S T = F_S$. Hence, postmultiplying (5.8) by $T$, this system reduces to (5.7) with $V = \bar{B} T$. Therefore, if (5.8) has a solution $\bar{B}$, then the system (5.7) also has a solution $V = \bar{B} T$, and the theorem follows. \square

The matrices $F_p, T, \bar{B}$, and $V = \bar{B} T$ are listed in [12] for $p = 2, 3, 4$, and 5.

We will now discuss the necessity of the simplifying conditions given in Theorem 3 for TSRK methods of order $p$. It follows from Theorem 5 that these conditions are necessary for $p = 2, 3$, and 4. We also suspect that these conditions are necessary for $p = 5$ (compare the discussion following Theorem 6), but so far we are unable to prove this. Since in view of Theorem 7, the minimal number of stages for TSRK methods of order six satisfying the simplifying assumptions of Theorem 3 is CEN(5) = 8, and since there exist RK methods of order six with only seven stages, it follows that these simplifying assumptions are not necessary for $p \geq 6$.

In the technical report [12] the results of some numerical experiments were presented. In these experiments we have used some of the test problems from the paper by Hull et al. [9]. We implemented the continuous RK methods of order 3 and 4 with step control strategy based on the estimation of local discretization error by TSRK methods of order 4 and 5, respectively, and by Richardson extrapolation. We have listed in [12] numerical data such as the number of steps, the number of rejected steps, the number of function evaluations, the percentage of steps for which the local error exceeded the tolerance, the maximum local truncation error in units of tolerance, and the global error at the endpoint of integration. These results are given for three tolerances $TOL = 10^{-3}, 10^{-6}, 10^{-9}$, and for the two strategies of estimating the local discretization error. To obtain further insight into the reliability of error estimation based on TSRK methods, as compared with error estimation based on Richardson extrapolation, we also presented in [12] the percentage of steps for which the local error was within 1%, 5%, 10%, 25%, and 50% of the local error estimate computed by the above two methods. All these results indicate that the error estimation of continuous RK methods by TSRK methods appears to be almost as reliable as local error estimation based on Richardson extrapolation, at the same time being much more efficient.
Bibliography


