Some embedded pairs for optimal implicit strong stability preserving Runge–Kutta methods

Imre Fekete and Ákos Horváth

Abstract We construct specific embedded pairs for second and third order optimal strong stability preserving implicit RungeKutta methods with large absolute stability regions. These pairs offer adaptive implementation possibility for strong stability preserving (SSP) methods and maintain their inherent nonlinear stability properties, too.

1 Introduction and SSP Runge–Kutta methods

Let us consider an initial value problem (IVP)

$$y'(t) = f(t, y(t)), \qquad y(t_0) = y_0. \tag{1}$$

The numerical solution of (1) at each time step with an implicit s-stage Runge–Kutta (RK) method RK(A, b^T) is given by

$$y_{n+1} = y_n + \Delta t \sum_{j=1}^{s} b_j f(t_n + c_j \Delta t, Y_j) \tag{2}$$

and the internal stages are computed as

Imre Fekete
Institute of Mathematics, Eötvös Loránd University, MTA-ELTE Numerical Analysis and Large Networks Research Group, Pázmány P. Sétány 1/C, Budapest H-1117, Hungary, e-mail: feipaat@cs.elte.hu

Ákos Horváth
Institute of Mathematics, Eötvös Loránd University, Pázmány P. Sétány 1/C, Budapest H-1117, Hungary, e-mail: horvathakos723@gmail.com

$$Y_i = y_n + \Delta t \sum_{j=1}^{s} a_{ij} f(t_n + c_j \Delta t, Y_j), \qquad i = 1, \ldots, s \tag{3}$$

where y_n is an approximation to the solution of (1) at time $t_n = t_0 + n\Delta t$, $A = (a_{ij})$ and $b^T = (b_j)$ are the coefficient of the method. By using the method-of-line approach, spatial discretization of hyperbolic partial differential equations (PDEs) lead to a large system of ordinary differential equations (ODEs)

$$u_t = F(u), \tag{4}$$

where u is a vector of approximations to the exact solution of the PDE. SSP time discretization methods were designed to ensure nonlinear stability properties in (4). We assume that the semi-discretization (4) and a convex functional $||\cdot||$ (or norm, semi-norm) are given, and that there exists a Δt_{FE} such that the forward Euler condition

$$||u + \Delta t F(u)|| \leq ||u|| \text{ for } 0 \leq \Delta t \leq \Delta t_{\mathrm{FE}} \tag{5}$$

holds for all u. An implicit Runge–Kutta (IRK) method is called SSP if the estimate

$$||u_{n+1}|| \leq ||u_n||$$

holds for the numerical solution of (4), whenever (5) holds and $\Delta t \leq \mathcal{C} \Delta t_{\mathrm{FE}}$. The constant $\mathcal{C}$ is called the SSP coefficient. For a complete introduction into the SSP theory we recommend monograph [2]. Below we give the main results which will be used in this paper.

Theorem 1 ([2], Theorem 3.2.). *Let us consider the matrix*

$$K = \begin{pmatrix} A & 0 \\ b^T & 0 \end{pmatrix}$$

and the SSP conditions

$$K(I + rK)^{-1} \geq 0 \tag{6a}$$

$$rK(I + rK)^{-1} e \leq e. \tag{6b}$$

Then, the SSP coefficient of the IRK method is

$$\mathcal{C}(A, b^T) = \sup\left\{ r : (I + rK)^{-1} \textit{exists and conditions } (6a)\text{-}(6b) \textit{ hold} \right\}.$$

Theorem 2 ([2], Observation 5.2.). *Consider an IRK method. If the method has positive SSP coefficient $\mathcal{C}(A, b^T)$, then $A \geq 0$ and $b^T \geq 0$.*

It has been showed that IRK methods with positive $\mathcal{C}$ cannot exist for $p > 6$ [1]. Therefore, we are interested in taking into account order conditions up to order of six.

By using embedded pairs we could allow adaptive step-size control based on local truncation error estimation [3]. The general s-stage IRK pair $\mathrm{RK}(A, b^T, \tilde{b}^T)$ of order $p(p-1)$ has the following extended Butcher tableau.

$$\begin{array}{c|c} c & A \\ \hline & b^T \\ & \tilde{b}^T \end{array}$$

As usual, $c = (c_1, c_2, \ldots, c_s)^T$ is given by $c = A\mathbf{e}$ with $\mathbf{e} = (1, \ldots, 1)^T \in \mathbb{R}^s$. The vectors b^T, $\tilde{b}^T$ define the coefficients of the p-th and $(p-1)$-th order approximations, respectively. Motivation for providing embedded pairs for SSP methods is that several optimal implicit SSP methods have useful stability regions, small error coefficients, big absolute monotonicity radius and are frequently used even when SSP theory cannot be applied. In the next section, we give the analytical framework that enables us to construct the new family of embedded pairs and construct the embedded pairs analytically and numerically for second and third order optimal implicit SSP RK methods.

2 Embedded pairs for second and third order implicit SSP RK methods

We introduce the notation SSPIRK(s, p) for optimal implicit SSP RK methods, where s and p refer to the number of stages and order, respectively. We give below the desired properties for embedded pairs.

(i), The embedded method is order of $p-1$.

(ii), The embedded method is non-defective, i.e. it violates all of the p-th order conditions.

(iii), The embedded method has rational coefficients and simple structure.

(iv), The embedded method has maximum SSP coefficient $\tilde{\mathcal{C}}$, where $\tilde{\mathcal{C}}$ is the SSP coefficient of the the optimal SSPIRK method; if this is not the case, then we are looking for embedded SSPIRK methods with smaller SSP coefficient or simply embedded IRK methods.

Taking into account the desired properties (i)-(iv), we seek an embedded pair $\tilde{b}^T$, with the stage coefficient A from a SSPIRK method such that these satisfy the following optimization problem

$$\text{the appropriate order conditions and property (ii) are fulfilled,} \tag{7}$$

$$\begin{pmatrix} A & 0 \\ \tilde{b}^T & 0 \end{pmatrix} \left(I + \tilde{\mathcal{C}} \begin{pmatrix} A & 0 \\ \tilde{b}^T & 0 \end{pmatrix} \right)^{-1} \geq 0, \tag{8}$$

$$\left|\left|\tilde{\mathcal{C}}\begin{pmatrix} A & 0 \\ \tilde{b}^T & 0 \end{pmatrix}\left(I+\tilde{\mathcal{C}}\begin{pmatrix} A & 0 \\ \tilde{b}^T & 0 \end{pmatrix}\right)^{-1}\right|\right|_{\infty} \leq 1, \tag{9}$$

where (8)-(9) are equivalent with (6a)-(6b) and $||\cdot||_\infty$ denotes the induced matrix norm. Since we fix $\tilde{\mathcal{C}}$ therefore we have a simplified optimization problem (7)-(9). Due to Theorem 2 and the first order condition $\tilde{b}^T\mathbf{e} = 1$ we have the componentwise condition $0 \leq \tilde{b}^T \leq \mathbf{e}$. The newly constructed pairs should satisfy desired properties (i)-(iv) and should have large absolute stability regions.

2.1 Embedded pairs for SSPIRK(s,2) methods

The s-stage second order characterization was given by Gottlieb, Ketcheson and Macdonald [4]. The methods have $\mathcal{C} = 2s$. The Butcher form of SSPIRK$(s,2)$ methods is given in Table 1. Taking into account desired properties (i)-(iv) it turns out that for general s we cannot find embedded pairs with maximal $\tilde{\mathcal{C}}$.

Table 1 Butcher form of SSPIRK$(s,2)$ methods.

$\frac{1}{2s}$	$\frac{1}{2s}$				
$\frac{3}{s}$	$\frac{1}{s}$	$\frac{1}{2s}$			
$\frac{5}{s}$	$\frac{1}{s}$	$\frac{1}{s}$	$\frac{1}{2s}$		
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\ddots$	
$\frac{2s-1}{2s}$	$\frac{1}{s}$	$\frac{1}{s}$	$\cdots$	$\frac{1}{s}$	$\frac{1}{2s}$
	$\frac{1}{s}$	$\frac{1}{s}$	$\cdots$	$\frac{1}{s}$	$\frac{1}{s}$

Theorem 3. *There is no first order embedded pair for SSPIRK*$(2,2)$ *with properties (i)-(iv).*

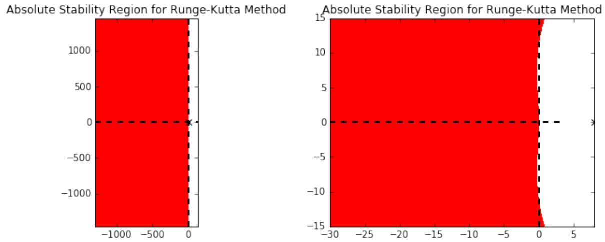
Based on Theorem 3 and its generalization one can conclude that there isn't first order embedded pair with $\tilde{\mathcal{C}} = 2s$ for SSPIRK$(s,2)$. Therefore we are interested in giving embedd pairs with smaller $\tilde{\mathcal{C}}$. Namely we are looking for $\tilde{\mathcal{C}} = s$ and our numerical search suggested the following pairs satisfying the desired properties (i)-(iv).

$$\tilde{b}_1^T = \left(\frac{2}{s+1},\ldots,\frac{2}{s+1},\frac{3}{s+1}\right)^T,\ \tilde{b}_2^T = \left(\frac{1}{s},\ldots,\frac{1}{s},\frac{5}{4s},\frac{3}{4s}\right)^T$$

$$\tilde{b}_3^T = \left(\frac{1}{s}, \dots, \frac{1}{s}, \frac{13}{12s}, \frac{10}{12s}, \frac{10}{12s}, \frac{15}{12s}\right)^T$$

Based on absolute stability region measurements it is obvious that embedded pair $\tilde{b}_2^T$ is reccommended. Below we present a result for $s = 4$ on Fig. 1 but as we are increasing the number of stages we can see similar results

Fig. 1 The left and right plots correspond to the absolute stability region of SSPIRK$(4, 2)$ and its $\tilde{b}_2^T$ embedded pair.



2.2 Embedded pairs for SSPIRK(s,3) methods

The s-stage third order characterization was also given by Gottlieb, Ketcheson and Macdonald [4]. The methods have $\mathcal{C} = s - 1 + \sqrt{s^2 - 1}$. The Butcher form of SSPIRK$(s, 3)$ methods is given in Table 2.

Table 2 Butcher form of SSPIRK$(s, 3)$ methods.

$$\begin{array}{c|ccccc}
\beta_1 & \beta_1 & & & & \\
2\beta_1 + \beta_2 & \beta_1 + \beta_2 & \beta_1 & & & \\
3\beta_1 + 2\beta_2 & \beta_1 + \beta_2 & \beta_1 + \beta_2 & \beta_1 & & \\
\vdots & \vdots & \vdots & \ddots & \ddots & \\
s\beta_1 + (s-1)\beta_2 & \beta_1 + \beta_2 & \beta_1 + \beta_2 & \dots & \beta_1 + \beta_2 & \beta_1 \\
\hline
 & \frac{1}{s} & \frac{1}{s} & \dots & \frac{1}{s} & \frac{1}{s}
\end{array}$$

where

$$\beta_1 = \frac{1}{2}\left(1 - \sqrt{\frac{s-1}{s+1}}\right) \text{ and } \beta_2 = \frac{1}{2}\left(\sqrt{\frac{s+1}{s-1}} - 1\right).$$

Similarly to the SSPIRK$(s, 2)$ case after tedious calculations one can see for lower stages that the desired properties (i)-(iv) cannot be satisfied with the maximal $\tilde{\mathcal{C}}$ coefficient. However, if we consider $\tilde{\mathcal{C}} = \mathcal{C}/2$ then we could give general form for SSPIRK$(s,3)$ methods with desired properties (i)-(iv). These pairs are

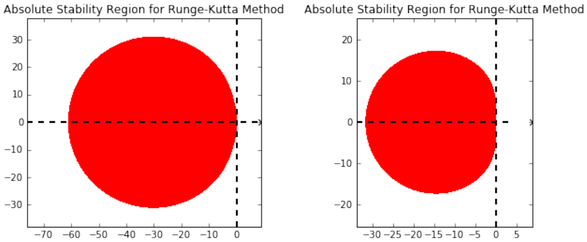
$$\tilde{b}_1^T = \left(\frac{1}{\sqrt{s^2-1}}, \dots, \frac{1}{\sqrt{s^2-1}}, \frac{s-1-\frac{s-2}{s-1}\sqrt{s^2-1}}{2}, \frac{3-s+\frac{s-2}{s+1}\sqrt{s^2-1}}{2} \right)$$

and

$$\tilde{b}_2^T = \left(\frac{1}{s}, \dots, \frac{1}{s}, \frac{21s+39-3\sqrt{s^2-1}}{16s^2+34s}, \frac{3s+12+3\sqrt{s^2-1}}{8s^2+17s}, \frac{21s+39-3\sqrt{s^2-1}}{16s^2+34s} \right).$$

Based on absolute stability region measurements we reccommend embedded pair $\tilde{b}_2^T$. Here we present a result for $s = 4$ on Fig. 2. As we are increasing the number of stages we can see similar results.

Fig. 2 The left and right plots correspond to the absolute stability region of SSPIRK$(4,3)$ and its $\tilde{b}_2^T$ embedded pair.



Acknowledgements The project has been supported by the European Union, co-financed by the European Social Fund (EFOP-3.6.3-VEKOP-16-2017-00002).

References

1. S. Gottlieb (2015) *Strong Stability Preserving Time Discretizations: A Review*, ICOSAHOM 2014, Springer International Publishing, Cham, 17–30.
2. S. Gottlieb, D. Ketcheson, C.-W. Shu. (2011). *Strong stability preserving Runge-Kutta and multistep time discretizations*, World Scientific Publishing
3. E. Hairer, S.P. Nørsett, G. Wanner, *Solving ordinary differential equations. I*, Springer (1993)
4. D. Ketcheson, C. Macdonald, S. Gottlieb. (2009). *Optimal implicit strong stability preserving Runge-Kutta methods*, Appl. Num. Math., 59(2), 373–392