Strong stability preserving explicit Runge–Kutta methods of maximal effective order

Yiannis Hadjimichael* Colin B. Macdonald† David I. Ketcheson*
James H. Verner‡

March 3, 2013

Abstract

We apply the concept of effective order to strong stability preserving (SSP) explicit Runge–Kutta methods. Relative to classical Runge–Kutta methods, methods with an effective order of accuracy are designed to satisfy a relaxed set of order conditions, but yield higher order accuracy when composed with special starting and stopping methods. We show that this allows the construction of four-stage SSP methods with effective order four (such methods cannot have classical order four). However, we also prove that effective order five methods—like classical order five methods—require the use of non-positive weights and so cannot be SSP. By numerical optimization, we construct explicit SSP Runge–Kutta methods up to effective order four and establish the optimality of many of them. Numerical experiments demonstrate the validity of these methods in practice.

1 Introduction

Strong stability preserving time discretization methods were originally developed for the solution of nonlinear hyperbolic partial differential equations (PDEs). Solutions of such PDEs may contain discontinuities even when the initial conditions are smooth. Many numerical methods for their solution are based on a method-of-lines approach in which the problem is first discretized in space to yield a system of ODEs. The spatial discretization is often chosen to ensure the solution is total variation diminishing (TVD), in order to avoid the appearance of spurious oscillations near discontinuities, *when coupled with first-order forward Euler time integration*. Strong stability preserving (SSP) time discretizations (also known as TVD discretizations [25]) are high-order time discretizations that guarantee the TVD property (or other convex functional bounds), with a possibly different step-size restriction [10]. Section 2 reviews Runge–Kutta methods and the concept of strong stability preserving methods.

Explicit SSP Runge–Kutta methods cannot have order greater than four [23]. However, a Runge–Kutta method may achieve an effective order of accuracy higher than its classical order by the use of special starting and stopping procedures. The conditions for a method to have effective order q are in general less restrictive than the conditions for a method to have classical order q. Section 3 presents a brief overview of the algebraic representation of Runge–Kutta methods, following Butcher [5]. This includes the concept of effective order and a list of effective order conditions.

*Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science & Technology (KAUST), P.O. Box 4700, Thuwal 23955, Saudi Arabia (yiannis.hadjimichael@kaust.edu.sa, david.ketcheson@kaust.edu.sa). The work of these authors is supported by Award No. FIC/2010/05, made by King Abdullah University of Science and Technology (KAUST).

†Mathematical Institute, University of Oxford, OX1 3LB, UK (macdonald@maths.ox.ac.uk). The work of this author was supported by NSERC Canada and by Award No KUK-C1-013-04 made by King Abdullah University of Science and Technology (KAUST).

‡Department of Mathematics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6, Canada (jverner@pims.math.ca). The work of this author was supported by Simon Fraser University.

We examine the SSP properties of explicit Runge–Kutta methods whose effective order is greater than their classical order. Previous studies of SSP Runge–Kutta methods have considered only the classical order of the methods. Three natural questions are:

- Can an SSP Runge–Kutta method have effective order of accuracy greater than four?
- If we only require methods to have *effective* order q, is it possible to achieve larger SSP coefficients than those obtained in methods with *classical* order q?
- SSP Runge–Kutta methods of order four require at least five stages. Can SSP methods of effective order four have fewer stages?

We show in Section 4 that the answer to the first question is negative. We answer the second question by numerically solving the problem of optimizing the SSP coefficient over the class of methods with effective order q; see Section 5. Most of the methods we find are shown to be optimal, as they achieve a certain theoretical upper bound on the SSP coefficient that is obtained by considering only linear problems [19]. We answer the last question affirmatively by construction, also in Section 5. The paper concludes with numerical experiments in Section 6 and conclusions in Section 7.

2 Strong stability preserving Runge–Kutta methods

Strong stability preserving (SSP) time-stepping methods were originally introduced for time integration of systems of hyperbolic conservation laws [27]

$$\boldsymbol{U}_t + \nabla \cdot \boldsymbol{f}(\boldsymbol{U}) = 0, \tag{1}$$

with appropriate initial and boundary conditions. A spatial discretization gives the system of ODEs

$$\boldsymbol{u}'(t) = \boldsymbol{F}(\boldsymbol{u}(t)), \tag{2}$$

where $\boldsymbol{u}$ is a vector of continuous-in-time grid values approximating the solution $\boldsymbol{U}$ at discrete grid points. Of course, (2) can arise in many ways and $\boldsymbol{F}$ need not necessarily represent a spatial discretization. Particularly, $\boldsymbol{F}$ may be time-dependent, but we can always make a transformation to an autonomous form. In any case, a time discretization then produces a sequence of solutions $\boldsymbol{u}^n \approx \boldsymbol{u}(t_n)$. This work studies explicit Runge–Kutta time discretizations. An explicit s-stage Runge–Kutta method takes the form

$$\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \Delta t \sum_i^s b_i \boldsymbol{F}(\boldsymbol{Y}_i),$$

where

$$\boldsymbol{Y}_i = \boldsymbol{u}^n + \Delta t \sum_j^{i-1} a_{ij} \boldsymbol{F}(\boldsymbol{Y}_j).$$

Such methods are characterized by the coefficient matrix $A = (a_{ij}) \in \mathbb{R}^{s\times s}$, the weight vector $\boldsymbol{b} = (b_i) \in \mathbb{R}^s$ and the abscissa $\boldsymbol{c} = (c_i) \in \mathbb{R}^s$, where $c_i = \sum_{j=1}^{i-1} a_{ij}$. The accuracy and stability of the method depend on the coefficients of the Butcher tableau $(A, \boldsymbol{b}, \boldsymbol{c})$ [5].

In some cases, the solutions of hyperbolic conservation laws satisfy a monotonicity property. For example, if (1) is scalar then solutions are monotonic in the total variation semi-norm [15]. For this reason, many popular spatial discretizations are designed such that, for a suitable class of problems, the solution $\boldsymbol{u}$ in (2) computed with the forward Euler scheme is non-increasing (in time) in some norm, semi-norm, or convex functional; i.e.,

$$\|\boldsymbol{u} + \Delta t \boldsymbol{F}(\boldsymbol{u})\| \le \|\boldsymbol{u}\|, \quad \text{for all } \boldsymbol{u} \text{ and for } 0 \le \Delta t \le \Delta t_{\mathrm{FE}}. \tag{3}$$

If this is the case, then an SSP method also generates a solution whose norm is non-increasing in time, under a modified time-step restriction.

Definition 2.1 (Strong Stability Preserving)**.** *A Runge–Kutta method is said to be* strong stability preserving *with* SSP coefficient $\mathcal{C} > 0$ *if, whenever the forward Euler condition* (3) *holds and*

$$0 \leq \Delta t \leq \mathcal{C}\Delta t_{\mathrm{FE}},$$

the Runge–Kutta method generates a monotonic sequence of solution values $\boldsymbol{u}^n$ *satisfying*

$$\|\boldsymbol{u}^{n+1}\| \leq \|\boldsymbol{u}^n\|.$$

Note that Δt_{FE} is a property of the spatial discretization $\boldsymbol{F}$ and is independent of $\boldsymbol{u}$. The SSP coefficient $\mathcal{C}$ is a property of the particular time-stepping method and quantifies the allowable time step size relative to that of the forward Euler method. Generally we want the SSP coefficient to be as large as possible for efficiency. To allow a fair comparison of explicit methods with different number of stages, we consider the *effective SSP coefficient*

$$\mathcal{C}_{\mathrm{eff}} = \frac{\mathcal{C}}{s}.$$

Note that the use of the word *effective* here is unrelated to the concept of *effective order* introduced in Section 3.

2.1 Optimal SSP schemes

We say that an SSP Runge–Kutta method is optimal if it has the largest possible SSP coefficient for a given order and a given number of stages. The search for these optimal methods was originally based on expressing the Runge–Kutta method as combinations of forward Euler steps (the Shu–Osher form) and solving a non-linear optimization problem [25, 26, 28, 29, 24, 22]. However, the SSP coefficient is related to the *radius of absolute monotonicity* [20] and, for irreducible Runge–Kutta methods, the two are equivalent [8, 14]. This gives a simplified algebraic characterization of the SSP coefficient [9]; it is the maximum value of r such that the following conditions hold:

$$K(I + rA)^{-1} \geq 0 \tag{4a}$$

$$\boldsymbol{e}_{s+1} - rK(I + rA)^{-1}\boldsymbol{e}_s \geq 0, \tag{4b}$$

provided that $I + rA$ is invertible. Here

$$K = \begin{pmatrix} A \\ \boldsymbol{b}^{\mathsf{T}} \end{pmatrix},$$

while $\boldsymbol{e}_s$ denotes the vector of ones of length s and I is the $s \times s$ identity matrix. The inequalities are understood component-wise.

The optimization problem of finding optimal SSP Runge–Kutta methods can thus be written as follows:

$$\max_{A,\boldsymbol{b},r} \; r \quad \text{subject to (4) and } \Phi(K) = 0. \tag{5}$$

Here $\Phi(K)$ represents the order conditions.

Following [15, 18], we will numerically solve the optimization problem (5) to find optimal explicit SSP Runge–Kutta methods for various effective orders of accuracy. However, we first need to define the order conditions $\Phi(K)$ for these methods. This is discussed in the next section.

3 The effective order of Runge–Kutta methods

The definition, construction, and application of methods with an effective order of accuracy relies on the use of starting and stopping methods. Specifically, we consider a *starting method* S, a *main method* M, and a *stopping method* S^{-1}. The successive use of these three methods results in a method $P = S^{-1}MS$,

Table 1: Elementary weights $\alpha(t_i)$ of trees t_i up to order five for a Runge–Kutta method with Butcher tableau $(A, \boldsymbol{b}, \boldsymbol{c})$. Here C is a diagonal matrix with components $c_i = \sum_{j=1}^{i-1} a_{ij}$ and exponents of vectors represent component exponentiation. By convention $\alpha_0 = \alpha(t_0) = 1$, where t_0 denotes the empty tree.

i	tree t_i	$\alpha(t_i)$	$\gamma(t_i)$	i	tree t_i	$\alpha(t_i)$	$\gamma(t_i)$
0	$\emptyset$	1	0	9		$\boldsymbol{b}^T\boldsymbol{c}^4$	5
1		$\boldsymbol{b}^T\boldsymbol{e}$	1	10		$\boldsymbol{b}^T C^2 A\boldsymbol{c}$	10
2		$\boldsymbol{b}^T\boldsymbol{c}$	2	11		$\boldsymbol{b}^T CA\boldsymbol{c}^2$	15
3		$\boldsymbol{b}^T\boldsymbol{c}^2$	3	12		$\boldsymbol{b}^T CA^2\boldsymbol{c}$	30
4		$\boldsymbol{b}^T A\boldsymbol{c}$	6	13		$\boldsymbol{b}^T (A\boldsymbol{c})^2$	20
5		$\boldsymbol{b}^T\boldsymbol{c}^3$	4	14		$\boldsymbol{b}^T A\boldsymbol{c}^3$	20
6		$\boldsymbol{b}^T CA\boldsymbol{c}$	8	15		$\boldsymbol{b}^T ACA\boldsymbol{c}$	40
7		$\boldsymbol{b}^T A\boldsymbol{c}^2$	12	16		$\boldsymbol{b}^T A^2\boldsymbol{c}^2$	60
8		$\boldsymbol{b}^T A^2\boldsymbol{c}$	24	17		$\boldsymbol{b}^T A^3\boldsymbol{c}$	120

which denotes the application of method S, followed by method M, followed by method S^{-1}. The method S^{-1} is an "inverse" of method S. We want P to have order q, whereas M might have lower classical order $p < q$. We then say M has *effective order* q.

When the method P is used for n steps,

$$P^n = (S^{-1}MS)^n = (S^{-1}MS)\cdots(S^{-1}MS)(S^{-1}MS),$$

it turns out that only M need be used repeatedly, as in $S^{-1}M^nS$, because SS^{-1} leaves the solution unchanged up to order q. The starting method introduces a perturbation to the solution, followed by n time steps of the main method M, and finally the stopping method is used to correct the solution. In Section 5.2, we propose alternative starting and stopping procedures which allow the overall procedure to be SSP.

The effective order of a Runge–Kutta method is defined in an abstract algebraic context introduced by Butcher [2] and developed further in [3, 13, 6, 4] and others. We follow the book [5] in our derivation of the effective order conditions.

3.1 The algebraic representation of Runge–Kutta methods

According to Butcher's algebraic theory, irreducible Runge–Kutta methods are placed in one-to-one correspondence with elements of a group G, consisting of real-valued functions on the set of rooted trees [5, Theorem 384A]. A Runge–Kutta method corresponds to the map that takes each rooted tree t to the corresponding elementary weight $\Phi(t)$ of that Runge–Kutta method. Table 1 lists the elementary weights for trees of up to degree five; a general recursive formula can be found in [5, Definition 312A]. The ordering of trees given in Table 1 is used throughout the remainder of this work; thus t_9 refers to the tree with elementary weight $\boldsymbol{b}^T\boldsymbol{c}^4$. For a function $\alpha \in G$ we write the values of the elementary weights as $\alpha_i = \alpha(t_i)$ for tree t_i. A special element of the group $E \in G$ corresponds to the (hypothetical) method that evolves the solution exactly. The values of $E(t)$ are denoted $1/\gamma(t)$ [5] and the values of $\gamma(t)$ are included in Table 1. Classical order conditions are obtained by comparing the elementary weights of a method with these values.

Let $\alpha, \beta \in G$ correspond to Runge–Kutta methods M_1 and M_2 respectively. The application of method M_1 followed by method M_2 corresponds to the multiplicative group operation $\alpha\beta$.[1] This product is defined by partitioning the input tree and computing over the resulting forest [5, § 383].

[1] We write M_2M_1 to mean the application of M_1 followed by the application of M_2 (following matrix and operator ordering convention) but when referring to products of elements of G we use the reverse ordering $(\alpha\beta)$ to match the convention in [5].

Table 2: Effective order five conditions on α (main method M) in terms of order conditions on β (starting method S). See also [5, § 389]. Recall that α_i and β_i are the elementary weights associated with the index i in Table 1. We assume that $\beta_1 = 0$ (see Section 3.2.1).

q	Effective order conditions	
1	$\alpha_1 = 1.$	
2	$\alpha_2 = \frac{1}{2}.$	
3	$\alpha_3 = \frac{1}{3} + 2\beta_2, \quad \alpha_4 = \frac{1}{6}.$	
4	$\alpha_5 = \frac{1}{4} + 3\beta_2 + 3\beta_3, \quad \alpha_6 = \frac{1}{8} + \beta_2 + \beta_3 + \beta_4, \quad \alpha_7 = \frac{1}{12} + \beta_2 - \beta_3 + 2\beta_4, \quad \alpha_8 = \frac{1}{24}.$	
5	$\alpha_9 = \frac{1}{5} + 4\beta_2 + 6\beta_3 + 4\beta_5,$	$\alpha_{10} = \frac{1}{10} + \frac{5}{3}\beta_2 - 2\beta_2^2 + \frac{5}{2}\beta_3 + \beta_4 + \beta_5 + 2\beta_6,$
	$\alpha_{11} = \frac{1}{15} + \frac{4}{3}\beta_2 + \frac{1}{2}\beta_3 + 2\beta_4 + 2\beta_6 + \beta_7,$	$\alpha_{12} = \frac{1}{30} + \frac{1}{3}\beta_2 - 2\beta_2^2 + \frac{1}{2}\beta_3 + \frac{1}{2}\beta_4 + \beta_6 + \beta_8,$
	$\alpha_{13} = \frac{1}{20} + \frac{2}{3}\beta_2 - \beta_2^2 + \beta_3 + \beta_4 + 2\beta_6,$	$\alpha_{14} = \frac{1}{20} + \beta_2 + 3\beta_4 - \beta_5 + 3\beta_7,$
	$\alpha_{15} = \frac{1}{40} + \frac{1}{3}\beta_2 + \frac{3}{2}\beta_4 - \beta_6 + \beta_7 + \beta_8,$	$\alpha_{16} = \frac{1}{60} + \frac{1}{3}\beta_2 - \frac{1}{2}\beta_3 + \beta_4 - \beta_7 + 2\beta_8, \ \alpha_{17} = \frac{1}{120}.$

Two Runge–Kutta methods M_1 and M_2, are equivalent up to order p if their corresponding elements in G, α and β, satisfy $\alpha(t) = \beta(t)$, for every tree t with $r(t) \leq p$, where $r(t)$ denotes the order of the tree (number of vertices). We denote this equivalence relation by

$$M_1 \underset{p}{\simeq} M_2.$$

In this sense, methods have inverses: the product of α^{-1} and α must match the identity method up to order p. Note that inverse methods up to order p are not unique and inverse methods of explicit methods need not be implicit. We can then define the effective order of accuracy of a method M with starting method S and stopping method S^{-1}.

Definition 3.1. *[5, § 389] Suppose M is a Runge–Kutta method with corresponding $\alpha \in G$. Then the method M is of effective order q if there exist methods S, S^{-1} (with corresponding $\beta, \beta^{-1} \in G$) such that*

$$(\beta\alpha\beta^{-1})(t) = E(t), \text{ for every tree with } r(t) \leq q, \tag{6}$$

where β^{-1} is an inverse of β up to order q; i.e.

$$(\beta^{-1}\beta)(t) = 1(t), \text{ for every tree with } r(t) \leq q.$$

Here $1 \in G$ is the identity element and $E \in G$ is the exact evolution operator.

3.2 Effective order conditions

For the main method M to have effective order q, its coefficients and those of the starting and stopping methods must satisfy a set of algebraic conditions. These *effective order conditions* can be found by rewriting (6) as $(\beta\alpha)(t) = (E\beta)(t)$ and applying the group product operation. For trees up to order five these are tabulated in Table 2 (and also in [5, § 389]). In general, the effective order conditions allow more degrees of freedom for method design than do the classical order conditions. Note that the effective order conditions match the classical order conditions up to second order.

Remark 3.2. *The effective order conditions of the main method for the "tall" trees $t_1, t_2, t_4, t_8, t_{17}, \ldots$ match the classical order conditions and these are precisely the order conditions for linear problems. This follows from inductive application of the group product on the tall trees. Therefore, methods of effective order q have classical order at least q for linear problems.*

3.2.1 Order conditions of the main and starting methods

As recommended in [5], we consider the elementary weights β_i of the starting method as free parameters when determining the elementary weights α_i of the main method. The relationship in Table 2 between the α_i and β_i is mostly linear (although there are a few β_2^2 terms). It is thus straightforward to (mostly) isolate the equations for α_i and determine the β_i as linear combination of the α_i. This separation provides maximal degrees of freedom and minimizes the number of constraints when constructing the method M. The resulting effective order conditions for the main method M are given in Table 3 (up to effective order five). For a specified classical and effective order, these are the equality constraints $\Phi(K)$ in the optimization problem (5) for method M.

Constructing the main method M then determines the α values and we obtain a set of order conditions on β (for that particular choice of M). These are given in the right-half of Table 3. We can also find the order conditions on S^{-1} in terms of the β_i (see [5, Table 386(III)]). We note that increasing the classical order of the main method requires $\alpha_i = 1/\gamma(t_i)$ and thus by Table 2 requires more of the β_i to be zero.

Tables 2 and 3 both assume that $\beta_1 = 0$ (i.e., the starting and stopping methods perturb the solution but do not advance the solution in time). This assumption is without loss of generality following [5, Lemma 389A], the proof of which shows that we can always find starting procedures with $\beta_1 = 0$ for which the main method has effective order q, whenever this holds for a starting method with $\beta_1 \neq 0$.

4 Explicit SSP Runge–Kutta methods have effective order at most four

The classical order of any explicit SSP Runge–Kutta method cannot be greater than four [23]. It turns out that the effective order of any explicit SSP Runge–Kutta method also cannot be greater than four, although the proof of this result is more involved. We begin by recalling a well-known result.

Lemma 4.1. *(see [20, Theorem 4.2],[23, Lemma 4.2]) Any irreducible Runge–Kutta method with positive SSP coefficient $\mathcal{C} > 0$ must have positive weights $\boldsymbol{b} > \boldsymbol{0}$.*

Irreducibility [7] is technically important in this result and those that follow because a reducible SSP method might not have positive weights (but it would be reducible to one that does, as per the lemma). The main result of this section is:

Theorem 4.2. *Any explicit Runge–Kutta method with positive weights $\boldsymbol{b} > \boldsymbol{0}$ has effective order at most four.*

Proof. Any method of effective order five must have classical order at least two (see [5] or Table 3). Thus it is sufficient to show that any method with all positive weights cannot satisfy the conditions of effective order five and classical order two.

Let $(A, \boldsymbol{b}, \boldsymbol{c})$ denote the coefficients of an explicit Runge–Kutta method with effective order at least five, classical order at least two, and positive weights $\boldsymbol{b} > \boldsymbol{0}$. The effective order five and classical order two conditions (see Table 3 with $q = 5$ and $p = 2$) include the following:

$$\boldsymbol{b}^T \boldsymbol{e} = 1, \tag{7a}$$

$$\boldsymbol{b}^T A\boldsymbol{c} = \frac{1}{6}, \tag{7b}$$

$$\frac{1}{2}\boldsymbol{b}^T \boldsymbol{c}^2 - \frac{1}{6} = \beta_2, \tag{7c}$$

$$\frac{1}{4}\boldsymbol{b}^T \boldsymbol{c}^4 - \boldsymbol{b}^T C^2 A\boldsymbol{c} + \boldsymbol{b}^T (A\boldsymbol{c})^2 = \beta_2^2, \tag{7d}$$

where the powers on vectors are understood component-wise. Let $\boldsymbol{v} = \frac{1}{2}\boldsymbol{c}^2 - A\boldsymbol{c}$. Then substituting (7b)

Table 3: Effective order q, classical order p conditions on α and β for the main and starting methods, M and S respectively.

q	p	Order conditions for main method M	Order conditions for starting method S
3	2	$\alpha_1 = 1, \quad \alpha_2 = \frac{1}{2}, \quad \alpha_4 = \frac{1}{6}.$	$\beta_1 = 0, \quad \beta_2 = -\frac{1}{6} + \frac{1}{2}\alpha_3.$
4	2	$\alpha_1 = 1, \quad \alpha_2 = \frac{1}{2}, \quad \alpha_4 = \frac{1}{6},$ $\frac{1}{4} - \alpha_3 + \alpha_5 - 2\alpha_6 + \alpha_7 = 0,$ $\alpha_8 = \frac{1}{24}.$	$\beta_1 = 0, \quad \beta_2 = -\frac{1}{6} + \frac{1}{2}\alpha_3,$ $\beta_3 = \frac{1}{12} - \frac{1}{2}\alpha_3 + \frac{1}{3}\alpha_5,$ $\beta_4 = -\frac{1}{24} - \frac{1}{3}\alpha_5 + \alpha_6.$
4	3	$\alpha_1 = 1, \quad \alpha_2 = \frac{1}{2}, \quad \alpha_3 = \frac{1}{3}, \quad \alpha_4 = \frac{1}{6},$ $\frac{1}{12} - \alpha_5 + 2\alpha_6 - \alpha_7 = 0, \quad \alpha_8 = \frac{1}{24}.$	$\beta_1 = 0, \quad \beta_2 = 0, \quad \beta_3 = -\frac{1}{12} + \frac{1}{3}\alpha_5,$ $\beta_4 = -\frac{1}{24} - \frac{1}{3}\alpha_5 + \alpha_6.$
5	2	$\alpha_1 = 1, \quad \alpha_2 = \frac{1}{2}, \quad \alpha_4 = \frac{1}{6},$ $\alpha_8 = \frac{1}{24}, \quad \alpha_{17} = \frac{1}{120},$ $\frac{1}{4} - \alpha_3 + \alpha_5 - 2\alpha_6 + \alpha_7 = 0,$ $\beta_2^2 = \frac{1}{4}\alpha_9 - \alpha_{10} + \alpha_{13}, \quad \beta_2 = -\frac{1}{6} + \frac{1}{2}\alpha_3,$ $6\beta_2^2 = \frac{3}{10} - \frac{3}{2}\alpha_3 + \alpha_5 + \frac{1}{2}\alpha_9 - 3\alpha_{10} + 3\alpha_{11}$ $-\alpha_{14},$ $2\beta_2^2 = \frac{1}{15} - \frac{1}{2}\alpha_3 + \alpha_6 + \frac{1}{2}\alpha_9 - 2\alpha_{10} + \alpha_{11}$ $+\alpha_{12} - \alpha_{15},$ $4\beta_2^2 = \frac{19}{60} - \alpha_3 + \alpha_5 - 2\alpha_6 + \alpha_{11} - 2\alpha_{12} + \alpha_{16}.$	$\beta_1 = 0, \quad \beta_2 = -\frac{1}{6} + \frac{1}{2}\alpha_3,$ $\beta_3 = \frac{1}{12} - \frac{1}{2}\alpha_3 + \frac{1}{3}\alpha_5,$ $\beta_4 = -\frac{1}{24} - \frac{1}{3}\alpha_5 + \alpha_6,$ $\beta_5 = -\frac{1}{120} + \frac{1}{4}\alpha_3 - \frac{1}{2}\alpha_5 + \frac{1}{4}\alpha_9,$ $\beta_6 = \frac{7}{720} + \beta_2^2 + \frac{1}{12}\alpha_3 - \frac{1}{2}\alpha_6 - \frac{1}{8}\alpha_9$ $+\frac{1}{2}\alpha_{10},$ $\beta_7 = \frac{8}{45} - 2\beta_2^2 - \frac{7}{12}\alpha_3 + \frac{1}{2}\alpha_5 - \alpha_6$ $+\frac{1}{4}\alpha_9 - \alpha_{10} + \alpha_{11},$ $\beta_8 = -\frac{1}{120} + \beta_2^2 + \frac{1}{8}\alpha_9 - \frac{1}{2}\alpha_{10} + \alpha_{12}.$
5	3	$\alpha_1 = 1, \alpha_2 = \frac{1}{2}, \alpha_3 = \frac{1}{3}, \alpha_4 = \frac{1}{6}, \alpha_8 = \frac{1}{24},$ $\alpha_{17} = \frac{1}{120}, \quad \frac{1}{12} - \alpha_5 + 2\alpha_6 - \alpha_7 = 0,$ $\frac{1}{4}\alpha_9 - \alpha_{10} + \alpha_{13} = 0,$ $\frac{1}{5} - \alpha_5 - \frac{1}{2}\alpha_9 + 3\alpha_{10} - 3\alpha_{11} + \alpha_{14} = 0,$ $\frac{1}{10} - \alpha_6 - \frac{1}{2}\alpha_9 + 2\alpha_{10} - \alpha_{11} - \alpha_{12} + \alpha_{15} = 0,$ $\frac{1}{60} - \alpha_5 + 2\alpha_6 - \alpha_{11} + 2\alpha_{12} - \alpha_{16} = 0.$	$\beta_1 = 0, \quad \beta_2 = 0, \quad \beta_3 = -\frac{1}{12} + \frac{1}{3}\alpha_5$ $\beta_4 = -\frac{1}{24} - \frac{1}{3}\alpha_5 + \alpha_6,$ $\beta_5 = \frac{3}{40} - \frac{1}{2}\alpha_5 + \frac{1}{4}\alpha_9,$ $\beta_6 = \frac{3}{80} - \frac{1}{2}\alpha_6 - \frac{1}{8}\alpha_9 + \frac{1}{2}\alpha_{10},$ $\beta_7 = -\frac{1}{60} + \frac{1}{2}\alpha_5 - \alpha_6 + \frac{1}{4}\alpha_9 - \alpha_{10} + \alpha_{11},$ $\beta_8 = -\frac{1}{120} + \frac{1}{8}\alpha_9 - \frac{1}{2}\alpha_{10} + \alpha_{12}.$
5	4	$\alpha_1 = 1, \alpha_2 = \frac{1}{2}, \alpha_3 = \frac{1}{3}, \alpha_4 = \frac{1}{6}, \alpha_5 = \frac{1}{4},$ $\alpha_6 = \frac{1}{8}, \quad \alpha_7 = \frac{1}{12}, \quad \alpha_8 = \frac{1}{24}, \quad \alpha_{17} = \frac{1}{120},$ $\frac{1}{4}\alpha_9 - \alpha_{10} + \alpha_{13} = 0,$ $\frac{1}{20} + \frac{1}{2}\alpha_9 - 3\alpha_{10} + 3\alpha_{11} - \alpha_{14} = 0,$ $\frac{1}{40} + \frac{1}{2}\alpha_9 - 2\alpha_{10} + \alpha_{11} + \alpha_{12} - \alpha_{15} = 0,$ $\frac{1}{60} - \alpha_{11} + 2\alpha_{12} - \alpha_{16} = 0.$	$\beta_1 = 0, \quad \beta_2 = 0,$ $\beta_3 = 0, \quad \beta_4 = 0,$ $\beta_5 = -\frac{1}{20} + \frac{1}{4}\alpha_9,$ $\beta_6 = -\frac{1}{40} - \frac{1}{8}\alpha_9 + \frac{1}{2}\alpha_{10},$ $\beta_7 = -\frac{1}{60} + \frac{1}{4}\alpha_9 - \alpha_{10} + \alpha_{11},$ $\beta_8 = -\frac{1}{120} + \frac{1}{8}\alpha_9 - \frac{1}{2}\alpha_{10} + \alpha_{12}.$

into (7c) and expressing (7d) in terms of $\boldsymbol{v}$ gives

$$\boldsymbol{b}^T\boldsymbol{v} = \beta_2,$$
$$\boldsymbol{b}^T\boldsymbol{v}^2 = \beta_2^2.$$

Each of these is a strictly convex combination. Jensen's inequality (with a strictly convex function, as is the case with the square function here) then states $\boldsymbol{b}^T\boldsymbol{v}^2 \leq (\boldsymbol{b}^T\boldsymbol{v})^2$ with equality if and only if all components of $\boldsymbol{v}$ are equal [1, Theorem 12, pg 31]. Now $v_1 = 0$ for every explicit method so we deduce that $\boldsymbol{v} = \boldsymbol{0}$. That implies the method has stage order two, which is not possible for explicit methods [23]. This contradiction completes the proof. □

Corollary 4.3. *Let M denote an irreducible explicit Runge–Kutta method with $\mathcal{C} > 0$. Then M has effective order at most four.*

Proof. This follows immediately from Lemma 4.1 and Theorem 4.2. □

Remark 4.4. *It is worth noting here an additional result that follows directly from what we have proved. Using Theorem 4.2 and [7, Theorem 4.1], it follows that any irreducible explicit Runge–Kutta method with positive radius of circle contractivity has effective order at most four.*

5 Optimal explicit SSP Runge–Kutta schemes with maximal effective order

In this section, we use the SSP theory and Butcher's theory of effective order (Sections 2 and 3) to find optimal explicit SSP Runge–Kutta schemes with prescribed effective order and classical order. According to Corollary 4.3, there are no explicit SSPRK methods of effective order five, and therefore we need only consider methods with effective order up to four.

Recall from Section 3 that the methods with an effective order of accuracy involve a main method M as well as starting and stopping methods S and S^{-1}. In Section 5.2 we introduce a novel approach to construction of starting and stopping methods in order to allow them to be SSP.

We denote by ESSPRK(s, q, p) an s-stage explicit SSP Runge–Kutta method of effective order q and classical order p. Also we write SSPRK(s, q) for an s-stage explicit SSP Runge–Kutta method of order q.

5.1 The main method

Our search is carried out in two steps, first searching for optimal main methods M and then for possible corresponding methods S and S^{-1}. For a given number of stages, effective order, and classical order, our aim is thus to find an optimal main method, meaning one with the largest possible SSP coefficient $\mathcal{C}$.

To find a method ESSPRK(s, q, p) with Butcher tableau $(A, \boldsymbol{b}, \boldsymbol{c})$, we consider the optimization problem (5) with $\Phi(K)$ representing the conditions for effective order q and classical order p (as per Table 3). The methods are found through numerical search, using MATLAB's optimization toolbox. Specifically, we use `fmincon` with a sequential quadratic programming approach [15, 18]. This process does not guarantee a global minimizer, so many searches from random initial guesses are performed to help find methods with the largest possible SSP coefficients.

5.1.1 Optimal SSP coefficients

Useful bounds on the optimal SSP coefficient can be obtained by considering an important relaxation. In the relaxed problem, the method is required to be accurate and strong stability preserving only for linear, constant-coefficient initial value problems. This leads to a reduced set of order conditions and a relaxed absolute monotonicity condition [19, 15, 16]. We denote the maximal SSP coefficient for linear problems (maximized over all methods with order q and s stages) by $\mathcal{C}^{\text{lin}}_{s,q}$.

Table 4: Effective SSP coefficients $\mathcal{C}_{\text{eff}} = \mathcal{C}/s$ of the best known ESSPRK(s, q, p) methods. Entries in bold achieve the bound $\mathcal{C}^{\text{lin}}_{s,q}$ given by the linear SSP coefficient and are therefore optimal. If no positive $\mathcal{C}$ can be found, we use "–" to indicate non-existence. The optimal fourth-order linear SSP coefficients are $\mathcal{C}^{\text{lin}}_{4,4} = 0.25$, $\mathcal{C}^{\text{lin}}_{5,4} = 0.40$ and $\mathcal{C}^{\text{lin}}_{6,4} = 0.44$.

q	p	stages s										
		1	2	3	4	5	6	7	8	9	10	11
3	2	–	–	**0.33**	**0.50**	**0.53**	**0.59**	**0.61**	**0.64**	**0.67**	**0.68**	**0.69**
4	2	–	–	–	0.22	0.39	**0.44**	**0.50**	**0.54**	**0.57**	**0.60**	**0.62**
4	3	–	–	–	0.19	0.37	0.43	**0.50**	**0.54**	**0.57**	**0.60**	**0.62**

Let $\mathcal{C}_{s,q}$ denote the maximal SSP coefficient (relevant to non-linear problems) over all methods of s stages with order q. Let $\mathcal{C}_{s,q,p}$ denote the object of our study, i.e. the maximal SSP coefficient (relevant to non-linear problems) over all methods of s stages with effective order q and classical order p. From Remark 3.2 and the fact that the ESSPRK(s, q, p) methods form a super class of the SSPRK(s, q) methods, we have

$$\mathcal{C}_{s,q} \leq \mathcal{C}_{s,q,p} \leq \mathcal{C}^{\text{lin}}_{s,q}. \tag{8}$$

The effective SSP coefficients for methods with up to eleven stages are shown in Table 4. Recall from Section 4 that $q = 5$ implies a zero SSP coefficient and from Section 3 that for $q = 1, 2$, the class of explicit Runge–Kutta methods with effective order q is the simply the class of explicit Runge–Kutta methods with order q. Therefore we consider only methods of effective order $q = 3$ and $q = 4$. Exact optimal values of $\mathcal{C}^{\text{lin}}_{s,q}$ are known for many classes of methods; for example see [19, 15, 16]. Those results and (8) allow us to determine the optimal value of $\mathcal{C}_{s,q,p}$ *a priori* for the cases $q = 3$ (for any s) and for $q = 4, s = 10$, since in those cases we have $\mathcal{C}_{s,q} = \mathcal{C}^{\text{lin}}_{s,q}$.

5.1.2 Effective order three methods

Since $\mathcal{C}_{s,q} = \mathcal{C}^{\text{lin}}_{s,q}$ for $q = 3$, the optimal effective order three methods have SSP coefficients equal to the corresponding optimal classical order three methods. In the cases of three and four stages, we are able to determine exact coefficients for families of optimal methods of effective order three.

Theorem 5.1. *A family of optimal three-stage, effective order three SSP Runge–Kutta methods of classical order two, with SSP coefficient $\mathcal{C}_{3,3,2} = 1$, is given by*

$$\begin{aligned}
\boldsymbol{Y}_1 &= \boldsymbol{u}^n, \\
\boldsymbol{Y}_2 &= \boldsymbol{u}^n + \Delta t \boldsymbol{F}(\boldsymbol{Y}_1), \\
\boldsymbol{Y}_3 &= \boldsymbol{u}^n + \gamma \Delta t \boldsymbol{F}(\boldsymbol{Y}_1) + \gamma \Delta t \boldsymbol{F}(\boldsymbol{Y}_2), \\
\boldsymbol{u}^{n+1} &= \boldsymbol{u}^n + \frac{5\gamma - 1}{6\gamma} \Delta t \boldsymbol{F}(\boldsymbol{Y}_1) + \frac{1}{6} \Delta t \boldsymbol{F}(\boldsymbol{Y}_2) + \frac{1}{6\gamma} \Delta t \boldsymbol{F}(\boldsymbol{Y}_3),
\end{aligned}$$

where $1/4 \leq \gamma \leq 1$ is a free parameter.

Theorem 5.2. *A family of optimal four-stage, effective order three SSP Runge–Kutta methods of classical*

order two, with SSP coefficient $\mathcal{C}_{4,3,2} = 2$ is given by

$$\begin{aligned}
\boldsymbol{Y}_1 &= \boldsymbol{u}^n, \\
\boldsymbol{Y}_2 &= \boldsymbol{u}^n + \frac{1}{2}\Delta t \boldsymbol{F}(\boldsymbol{Y}_1), \\
\boldsymbol{Y}_3 &= \boldsymbol{u}^n + \frac{1}{2}\Delta t \boldsymbol{F}(\boldsymbol{Y}_1) + \frac{1}{2}\Delta t \boldsymbol{F}(\boldsymbol{Y}_2), \\
\boldsymbol{Y}_4 &= \boldsymbol{u}^n + \gamma \Delta t \boldsymbol{F}(\boldsymbol{Y}_1) + \gamma \Delta t \boldsymbol{F}(\boldsymbol{Y}_2) + +\gamma \Delta t \boldsymbol{F}(\boldsymbol{Y}_3), \\
\boldsymbol{u}^{n+1} &= \boldsymbol{u}^n + \frac{8\gamma - 1}{12\gamma}\Delta t \boldsymbol{F}(\boldsymbol{Y}_1) + \frac{1}{6}\Delta t \boldsymbol{F}(\boldsymbol{Y}_2) + \frac{1}{6}\Delta t \boldsymbol{F}(\boldsymbol{Y}_3) + \frac{1}{12\gamma}\Delta t \boldsymbol{F}(\boldsymbol{Y}_4),
\end{aligned}$$

where $1/6 \leq \gamma \leq 1/2$ is a free parameter.

Proof. In either theorem, feasibility can be verified by direct calculation of the conditions in problem (5). Optimality follows because $\mathcal{C}_{s,3,2} = \mathcal{C}^{\text{lin}}_{s,3}$. □

Theorem 5.1 gives a *family* of three-stage methods. The particular value of $\gamma = 1/4$ corresponds to the classical Shu–Osher SSPRK$(3,3)$ method [25]. Similarly, in Theorem 5.2 the particular value of $\gamma = 1/6$ corresponds to the usual SSPRK$(4,3)$ method. It seems possible that for each number of stages, the ESSPRK$(s,3,2)$ methods may form a family in which an optimal SSPRK$(s,3)$ method is a particular member.

5.1.3 Effective order four methods

The ESSPRK$(s,4,p)$ methods can have classical order $p = 2$ or 3. In either case, for stages $7 \leq s \leq 11$ the methods found are optimal because the SSP coefficient attains the upper bound of $\mathcal{C}^{\text{lin}}_{s,q}$. For fewer stages, the new methods still have SSP coefficients up to 30% larger than that of explicit SSPRK(s,q) methods. In the particular case of four-stage methods we have the following:

Remark 5.3. *In contrast with the non-existence of an SSPRK(4, 4) method [25, 23], we are able to find ESSPRK(4, 4, 2) and ESSPRK(4, 4, 3) methods. The coefficients of these methods are found in Tables 6 and 7.*

Additionally, we find two families of methods with effective order four, for which $\mathcal{C}_{\text{eff}}$ asymptotically approaches unity. The families consist of second order methods with $s = n^2+1$ stages and SSP coefficient $\mathcal{C}_{s,4,2} = n^2 - n$. They are optimal since $\mathcal{C}_{s,4,2} = \mathcal{C}^{\text{lin}}_{s,4}$ [19, Theorem 5.2(c)]. It is convenient to express the coefficients in the modified Shu–Osher form [10]

$$\begin{aligned}
\boldsymbol{Y}_i &= v_i \boldsymbol{u}^n + \sum_{j=1}^{i-1} \big(\alpha_{ij}\boldsymbol{Y}_j + \Delta t \beta_{ij} \boldsymbol{F}(\boldsymbol{Y}_j)\big), \ 1 \leq i \leq s+1 \\
\boldsymbol{u}^{n+1} &= \boldsymbol{Y}_{s+1},
\end{aligned}$$

because of the sparsity of the matrices $\alpha, \beta \in \mathbb{R}^{(s+1)\times s}$ and vector $\boldsymbol{v} \in \mathbb{R}^s$. For $n \geq 3$ the non-zero elements are given by

$$\begin{aligned}
v_1 &= 1, \qquad v_{n^2+2} = \frac{2}{(n^2+1)\big((n-1)^2+1\big)}, \\
\alpha_{n^2-2n+4,(n-2)^2} &= \frac{n^2 - 1 \pm \sqrt{n^3 - 3n^2 + n + 1}}{4n^2 - 6n + 2}, \\
\alpha_{n^2+2,n^2+1} &= \frac{n(n-1)^2}{(2n-1)(n^2+1)(1 - \alpha_{n^2-2n+4,(n-2)^2})}, \\
\alpha_{n^2+2,n^2-2n+2} &= 1 - v_{n^2+2} - \alpha_{n^2+2,n^2+1}, \\
\alpha_{i+1,i} &= \begin{cases} 1 - \alpha_{i+1,(n-2)^2}, & i = n^2 - 2n + 3 \\ 1, & \text{otherwise}, \end{cases}
\end{aligned}$$

Table 5: Order conditions on ρ and τ up to effective order four for starting and stopping methods R and T, respectively. The upper block represents the effective order three conditions. As in Table 2 and Table 3 we assume $\beta_1 = 0$.

$\rho(t) = (\beta\alpha)(t)$	$\tau(t) = (\alpha\beta^{-1})(t)$
$\rho_1 = \alpha_1$	$\tau_1 = \alpha_1$
$\rho_2 = \alpha_2 + \beta_2$	$\tau_2 = \alpha_2 - \beta_2$
$\rho_3 = \alpha_3 + \beta_3$	$\tau_3 = \alpha_3 - 2\alpha_1\beta_2 - \beta_3$
$\rho_4 = \alpha_4 + \alpha_1\beta_2 + \beta_4$	$\tau_4 = \alpha_4 - \alpha_1\beta_2 - \beta_4$
$\rho_5 = \alpha_5 + \beta_5$	$\tau_5 = \alpha_5 - 3\alpha_1^2\beta_2 - 3\alpha_1\beta_3 - \beta_5$
$\rho_6 = \alpha_6 + \alpha_2\beta_2 + \beta_6$	$\tau_6 = \alpha_6 - (\alpha_1^2 + \alpha_2 - \beta_2)\beta_2 - \alpha_1\beta_3 - \alpha_1\beta_4 - \beta_6$
$\rho_7 = \alpha_7 + \alpha_1\beta_3 + \beta_7$	$\tau_7 = \alpha_7 - 2\alpha_1\beta_4 - \alpha_1^2\beta_2 - \beta_7$
$\rho_8 = \alpha_8 + \alpha_1\beta_4 + \alpha_2\beta_2 + \beta_8$	$\tau_8 = \alpha_8 - \alpha_1\beta_4 - \alpha_2\beta_2 + \beta_2^2 - \beta_8$

where $1 \leq i \leq n^2$ and

$$\beta_{i,j} = \frac{\alpha_{i,j}}{n^2 - n}, \quad 1 \leq i \leq n^2 + 2, \;\; 1 \leq j \leq n^2 + 1.$$

In [10, § 6.2.2], a similar pattern was found for SSPRK$(s, 3)$ methods.

5.2 Starting and stopping methods

Provided an ESSPRK(s, q, p) scheme that can be used as the main method M, we want to find perturbation methods S and S^{-1} such that the Runge–Kutta scheme $S^{-1}MS$ attains classical order q, equal to the effective order of method M. We also want the resulting overall process to be SSP. However at least one of the S and S^{-1} methods is not SSP: if $\beta_1 = 0$ then $\sum_i b_i = 0$ implies the presence of at least one negative weight and thus neither scheme can be SSP. Even if we consider methods with $\beta_1 \neq 0$, one of S or S^{-1} must step backwards and thus that method cannot be SSP (unless we consider the downwind operator [24, 11, 17]).

In order to overcome this problem and achieve "bona fide" SSPRK methods with an effective order of accuracy, we need to choose different starting and stopping methods. We consider methods R and T which each take a positive step such that $R \underset{q}{\simeq} MS$ and $T \underset{q}{\simeq} S^{-1}M$. That is, the order conditions of R and T must match those of MS and $S^{-1}M$, respectively, up to order q. This gives a new $TM^{n-2}R$ scheme which is equivalent up to order q to the $S^{-1}M^nS$ scheme and attains classical order q. Each starting and stopping procedure now takes a positive step forward in time.

To derive order conditions for the R and T methods, consider their corresponding functions in group G to be ρ and τ respectively. Then the equivalence is expressed as

$$\rho(t) = (\beta\alpha)(t) \text{ and } \tau(t) = (\alpha\beta^{-1})(t), \quad \text{for all trees } t \text{ with } r(t) \leq q. \tag{9}$$

Rewriting the second condition in (9) as $(\tau\beta)(t) = \alpha(t)$, the order conditions for the starting and stopping methods can be determined by the usual product formula and are given in Table 5. These conditions could be constructed more generally but here we have assumed $\beta_1 = 0$ (see Section 3.2.1); this will be sufficient for constructing SSP starting and stopping conditions.

5.2.1 Optimizing the starting and stopping methods

It turns out that the order conditions from (9) do not contradict the SSP requirements. We can thus find methods R and T using the optimization procedure described in Section 2.1 with the order conditions given by Table 5 for $\Phi(K)$ in (5).

Table 6: ESSPRK(4,4,2): an effective order four SSPRK method with four stages and classical order two with its associated starting and stopping methods.

0				
0.730429885783319	0.730429885783319			
0.644964638145795	0.251830917810810	0.393133720334985		
1.000000000000000	0.141062771617064	0.220213358584678	0.638723869798257	
	0.384422161080494	0.261154113377550	0.127250689937518	0.227173035604438

(a) Main method M, ESSPRK$(4,4,2)$

0					
0.545722177514735	0.545722177514735				
0.842931687441527	0.366499989048164	0.476431698393363			
0.574760809487828	0.135697968350722	0.176400587890242	0.262662253246864		
0.980872743236632	0.103648417776838	0.134737771331049	0.200625899485633	0.541860654643112	
	0.233699169638954	0.294263351266422	0.065226988215286	0.176168374199685	0.230642116679654

(b) Starting method R

0				
0.509877496215340	0.509877496215340			
0.435774135529007	0.182230305923759	0.253543829605247		
0.933203341300203	0.148498121305090	0.206610981494095	0.578094238501017	
	0.307865440399752	0.171863794704750	0.233603236964822	0.286667527930676

(c) Stopping method T

The values of α_i are determined by the main method M. Also note that for effective order q, the algebraic expressions on β up to order $q-1$ are already found by the optimization procedure of the main method (see Table 3). However, the values of the order q elementary weights on β are not known; these are β_3 and β_4 for effective order three and β_5, β_6, β_7 and β_8 for effective order four. From Table 5, we see that both the R and T methods depend on these parameters. Our approach is to optimize for both methods at once: we solve a modified version of the optimization problem (5) where we simultaneously maximize both SSP coefficients subject to the constraints given in (9) and conditions on β given by Table 3. The unknown elementary weights on β are used as free parameters. In practice, we maximize the objective function $\min(r_1, r_2)$, where r_1 and r_2 are the radii of absolute monotonicity of the methods R and T.

We were able to construct starting and stopping schemes for each main method, with an SSP coefficient at least as large as that of the main method. This allows the usage of a uniform time-step $\Delta t \leq \mathcal{C}\Delta t_{\text{FE}}$, where $\mathcal{C}$ is the SSP coefficient of the main method. The additional computational cost of the starting and stopping methods is minimal: for methods R and T associated with an s-stage main method, at most $s+1$ and s stages, respectively, appear to be required. Tables 6 and 7 show the coefficients of the schemes where the main method is ESSPRK$(4,4,2)$ and ESSPRK$(4,4,3)$, respectively.

It is important to note that in practice, if accurate values are needed at any time other than the final time, the computation must invoke the stopping method to obtain them. Furthermore, changing step-size would require first applying the stopping method with the old step-size and then applying the starting method with the new step-size.

6 Numerical experiments

Having constructed strong stability preserving $TM^{n-2}R$ schemes in the previous section, we now numerically verify their properties. Specifically, we use a convergence study to show that the procedure attains order of accuracy q, the effective order of M. We also demonstrate on Burgers' equation that the SSP coefficient accurately measures the maximal time-step for which the methods are strong stability preserving.

Table 7: ESSPRK(4,4,3): an effective order four SSPRK method with four stages and classical order three with its associated starting and stopping methods.

0				
0.601245068769724	0.601245068769724			
0.436888719886063	0.139346829159954	0.297541890726109		
0.747760163757110	0.060555450075478	0.129301708677891	0.557903005003740	
	0.220532078662434	0.180572397883936	0.181420582644840	0.417474940808790

(a) Main method M, ESSPRK$(4,4,3)$

0					
0.438463764036947	0.438463764036947				
0.639336395725557	0.213665532574654	0.425670863150903			
0.434353425654020	0.061345094040860	0.122213530726218	0.250794800886942		
0.843416464962307	0.039559973266996	0.078812561688700	0.161731525131914	0.563312404874697	
	0.154373542967849	0.307547588471376	0.054439037790856	0.189611674483496	0.294028156286422

(b) Starting method R

0				
0.556337718891090	0.556337718891090			
0.428870688216872	0.166867537553458	0.262003150663414		
0.815008947642716	0.104422177204659	0.163956032598547	0.546630737839510	
	0.203508169408374	0.096469758967330	0.321630956102914	0.378391115521382

(c) Stopping method T

6.1 Convergence study

We consider the van der Pol system [12]

$$\begin{aligned} u_1'(t) &= u_2(t), \\ u_2'(t) &= \mu\big(1 - u_1^2(t)\big)u_2(t) - u_1(t), \end{aligned} \tag{10}$$

over the time interval $t \in [0, 50]$ with $\mu = 2$ and initial values $u_1(0) = 2$ and $u_2(0) = 1$. The reference solution for the convergence study is calculated by MATLAB's `ode45` solver with relative and absolute tolerances set to 10^{-13}.

We solve the initial value problem (10) using SSP $TM^{n-2}R$ schemes. The solution is computed using $n = 100 \cdot 2^k$ time steps for $k = 2, \ldots, 7$. The error at $t = 50$ with respect to time-step is shown in Figure 1 on a logarithmic scale. The convergence study is performed for $TM^{n-2}R$ schemes with various number of stages s and the results show that the schemes attain an order of accuracy equal to the effective order of their main method M. It is important in doing this sort of convergence study that the effective order of accuracy can only be obtained after the stopping method is applied. Intermediate steps will typically only be order p accurate (the classical order of the main method). Finally, we note that the methods with more stages generally exhibit smaller errors (for a given step size).

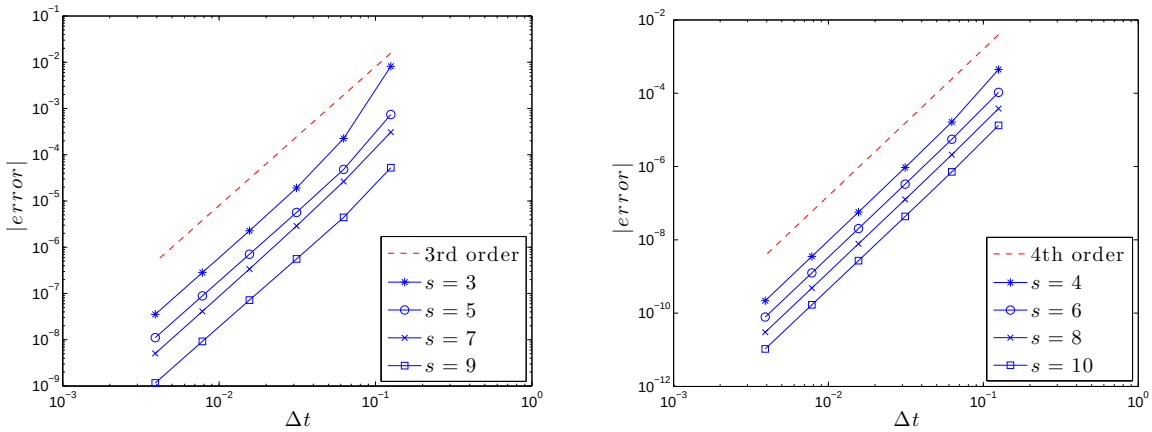
6.2 Burgers' equation

The inviscid Burgers' equation consists of the scalar hyperbolic conservation law

$$U_t + f(U)_x = 0, \tag{11}$$

with flux function $f(U) = \frac{1}{2}U^2$. We consider initial data $U(0, x) = \frac{1}{2} - \frac{1}{4}\sin \pi x$, on a periodic domain $x \in [0, 2)$. The solution advances to the right where it eventually exhibits a shock. We perform a semi-discretization using an upwind approximation to obtain the system of ODEs

$$\frac{\mathrm{d}}{\mathrm{d}t} u_i = -\frac{f(u_i) - f(u_{i-1})}{\Delta x}.$$



(a) ESSPRK$(s, 3, 2)$ (b) ESSPRK$(s, 4, 2)$

Figure 1: Convergence study for the van der Pol system using $TM^{n-2}R$ Runge–Kutta schemes when (a) M is an ESSPRK$(s, 3, 2)$ method and (b) M is an ESSPRK$(s, 4, 2)$ method.

This spatial discretization is total-variation-diminishing (TVD) when coupled with the forward Euler method under the restriction [21]

$$\Delta t \leq \Delta t_{\text{FE}} = \Delta x / \|U(0, x)\|_\infty.$$

Recall that a time discretization with SSP coefficient $\mathcal{C}$ will give a TVD solution for $\Delta t \leq \mathcal{C}\Delta t_{\text{FE}}$.

Burgers' equation was solved using an SSP $TM^{n-2}R$ scheme with time-step restriction $\Delta t = \sigma \Delta t_{\text{FE}}$, where σ indicates the size of the time step. We integrate to roughly time $t_{\text{f}} = 1.62$ with 200 points in space. Figure 2 shows that if σ is chosen less than the SSP coefficient of the main method, then no oscillations are observed. If this stability limit is violated, then oscillations may appear, as shown in Figure 1b. We measure these oscillations by computing the total variation of the numerical solution.

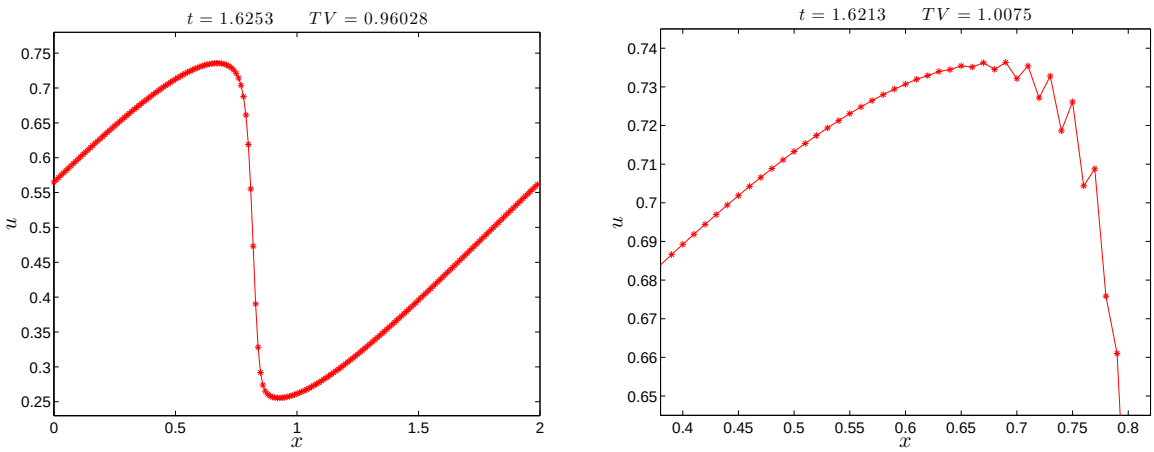
We also consider Burgers' equation with a discontinuous square wave initial condition

$$U(0, x) = \begin{cases} 1, & 0.5 \leq x \leq 1.5 \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

The solution consists of a rarefaction (i.e., an expansion fan) and a moving shock. Again we use 200 points in space and we compute the solution until roughly time $t_{\text{f}} = 0.6$, using a time-step $\Delta t = \sigma \Delta t_{\text{FE}}$. Figure 3 shows the result of solving the discontinuous problem using an SSP $TM^{n-2}R$ scheme, where M is an ESSPRK$(5, 4, 2)$ method with SSP coefficient $\mathcal{C} = 1.95$. In this case, $\sigma = 1.98$ appears to be the largest value for which the total variation is monotonically decreasing during the calculation. This is only 2% larger than the value of the SSP coefficient. Figure 2b shows part of the solution exhibiting oscillations when σ is larger than the SSP coefficient.

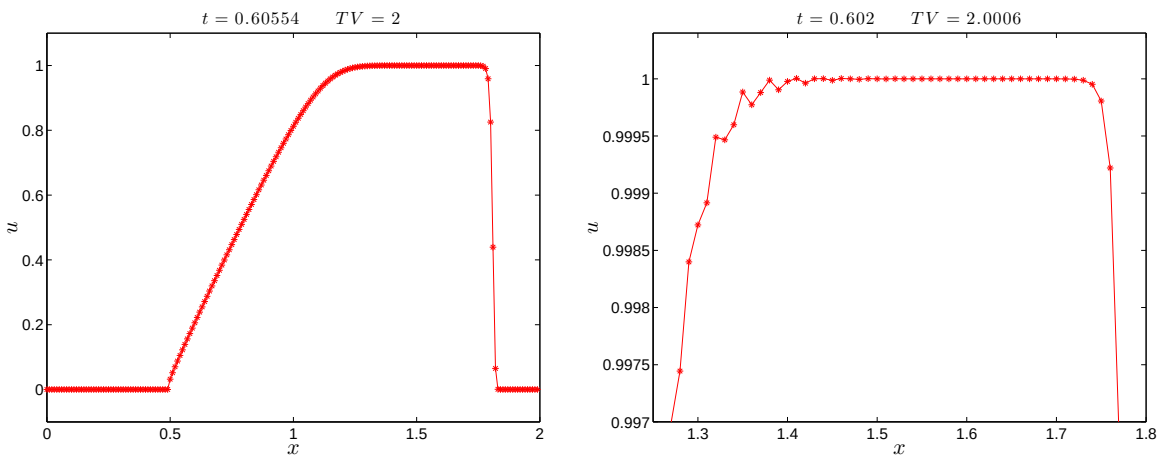
For various schemes, Table 8 shows the maximum observed values of σ for which the numerical solution is total variation decreasing for the entire computation. With the exception of the four-stage effective order four methods, we note good agreement between these experimental values and the SSP coefficients predicted by the theory.

We also note the necessity of our modified starting and stopping methods in the $RM^{n-2}T$ approach: in this example if we use the original approach of S and S^{-1}, the solution exhibits oscillations immediately following the application of the starting perturbation method S.



(a) $\sigma = 0.88$ (b) $\sigma = 1.60$

Figure 2: Solution of Burgers' equation at the final time with continuous initial data, using a $TM^{n-2}R$ scheme, where M is the optimal ESSPRK$(4, 4, 2)$. The time-step used is $\Delta t = \sigma \Delta t_{\text{FE}}$: in (a) we take σ as the SSP coefficient $\sigma = \mathcal{C} = s \times 0.22 = 0.88$ (see Table 4) and no oscillations are observed. However, in (b) we take $\sigma > \mathcal{C}$ and we observe oscillations (note (b) is magnified to show these). Here TV denotes the discrete total variation of the solution at the final time: a value greater than 1 (the total variation of the initial condition) indicates a violation of the TVD condition.



(a) $\sigma = 1.95$ (b) $\sigma = 2.15$

Figure 3: Solution of Burgers' equation at the final time with discontinuous initial data, using a $TM^{n-2}R$ scheme, where M is ESSPRK$(5, 4, 2)$ method. The time-step used is $\Delta t = \sigma \Delta t_{\text{FE}}$: in (a) we take σ as the SSP coefficient $\sigma = \mathcal{C} = s \times 0.39 = 1.95$ and no oscillations are observed. However, in (b) we take $\sigma > \mathcal{C}$ and, when magnified, we observe oscillations. Here TV denotes the total variation semi-norm of the solution at the final time: a value greater than 2 indicates an increase in total variation.

Table 8: Maximum observed coefficients exhibiting the TVD property on the Burgers' equation example with discontinuous data (12). The numbers in parenthesis indicate the increase relative to the corresponding SSP coefficients.

q	p	stages s								
		3	4	5	6	7	8	9	10	11
3	2	1.04(4%)	2.00(0%)	2.65(0%)	3.52(0%)	4.29(0%)	5.11(0%)	6.00(0%)	6.79(0%)	7.63(0%)
4	2	–	1.07(22%)	1.98(2%)	2.69(2%)	3.56(1%)	4.33(1%)	5.16(1%)	6.05(1%)	6.84(1%)
4	3	–	1.05(35%)	1.89(3%)	2.63(2%)	3.53(1%)	4.31(1%)	5.16(1%)	6.04(1%)	6.85(1%)

7 Conclusions

We use the theory of strong stability preserving time discretizations with Butcher's algebraic interpretation of order to construct explicit SSP Runge–Kutta schemes with an effective order of accuracy. These methods, when accompanied by starting and stopping methods, attain an order of accuracy higher than their (classical) order. We propose a new choice of starting and stopping methods to allow the overall procedure to be SSP. We prove that explicit Runge–Kutta methods with strictly positive weights have at most effective order four. This extends the barrier already known in the case of classical order explicit SSPRK methods.

SSP Runge–Kutta methods of effective order three and four are constructed by numerical optimization. Most of the methods found are optimal because they achieve the upper bound on the SSP coefficient known from linear problems. Also, despite the non-existence of four-stage, order four explicit SSPRK methods, we find effective order four methods with four stages (of classical order two and three). We perform numerical tests which confirm the accuracy and SSP properties of the new methods.

The ideas here are applied to explicit Runge–Kutta methods, but they could also be applied to other classes of methods including implicit Runge–Kutta methods, general linear methods, and Rosenbrock methods.

Acknowledgments

The authors would like to thank the anonymous referees for their helpful and valuable suggestions on the paper.

References

[1] P. S. Bullen, *Handbook of Means and their Inequalities*, Kluwer, 2003.

[2] J. C. Butcher, *The effective order of Runge–Kutta methods*, in Conf. on Numerical Solution of Differential Equations (Dundee, 1969), Springer, 1969, pp. 133–139.

[3] J. C. Butcher, *An algebraic theory of integration methods*, Math. Comp., 26 (1972), pp. 79–106.

[4] J. C. Butcher, *Order and effective order*, Appl. Numer. Math., 28 (1998), pp. 179–191. Eighth Conference on the Numerical Treatment of Differential Equations (Alexisbad, 1997).

[5] J. C. Butcher, *Numerical methods for ordinary differential equations*, Wiley, second ed., 2008.

[6] J. C. Butcher and J. M. Sanz-Serna, *The number of conditions for a Runge–Kutta method to have effective order p*, Appl. Numer. Math., 22 (1996), pp. 103–111.

[7] G. Dahlquist and R. Jeltsch, *Reducibility and contractivity of Runge–Kutta methods revisited*, BIT, 46 (2006), pp. 567–587.

[8] L. Ferracina and M. N. Spijker, *Stepsize restrictions for the total-variation-diminishing property in general Runge–Kutta methods*, SIAM J. Numer. Anal., 42 (2004), pp. 1073–1093.

[9] L. Ferracina and M. N. Spijker, *An extension and analysis of the Shu–Osher representation of Runge–Kutta methods*, Math. Comp., 74 (2005), pp. 201–219.

[10] S. Gottlieb, D. I. Ketcheson, and C. W. Shu, *Strong Stability Preserving Runge–Kutta and Multistep Time Discretizations*, World Scientific, Jan. 2011.

[11] S. Gottlieb and S. J. Ruuth, *Optimal strong-stability-preserving time-stepping schemes with fast downwind spatial discretizations*, J. Sci. Comput., 27 (2006), pp. 289–303.

[12] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations I: Nonstiff problems*, vol. 8 of Springer Series in Computational Mathematics, Springer-Verlag, 1987.

[13] E. Hairer and G. Wanner, *On the Butcher group and general multi-value methods*, Computing, 13 (1974), pp. 1–15.

[14] I. Higueras, *On strong stability preserving time discretization methods*, J. Sci. Comput., 21 (2004), pp. 193–223.

[15] D. I. Ketcheson, *Highly efficient strong stability-preserving Runge–Kutta methods with low-storage implementations*, SIAM J. Sci. Comput., 30 (2008), pp. 2113–2136.

[16] D. I. Ketcheson, *Computation of optimal monotonicity preserving general linear methods*, Math. Comp., 78 (2009), pp. 1497–1513.

[17] D. I. Ketcheson, *Step sizes for strong stability preservation with downwind-biased operators*, SIAM J. Numer. Anal., 49 (2011), pp. 1649–1660.

[18] D. I. Ketcheson, C. B. Macdonald, and S. Gottlieb, *Optimal implicit strong stability preserving Runge–Kutta methods*, Appl. Numer. Math., 59 (2009), pp. 373–392.

[19] J. F. B. M. Kraaijevanger, *Absolute monotonicity of polynomials occurring in the numerical solution of initial value problems*, Numer. Math., 48 (1986), pp. 303–322.

[20] J. F. B. M. Kraaijevanger, *Contractivity of Runge–Kutta methods*, BIT, 31 (1991), pp. 482–528.

[21] C. B. Laney, *Computational gasdynamics*, Cambridge University Press, 1998.

[22] S. J. Ruuth, *Global optimization of explicit strong-stability-preserving Runge–Kutta methods*, Math. Comp., 75 (2006), pp. 183–207.

[23] S. J. Ruuth and R. J. Spiteri, *Two barriers on strong-stability-preserving time discretization methods*, J. Sci. Comput., 17 (2002), pp. 211–220. Proceedings of the Fifth International Conference on Spectral and High Order Methods (ICOSAHOM-01).

[24] S. J. Ruuth and R. J. Spiteri, *High-order strong-stability-preserving Runge–Kutta methods with downwind-biased spatial discretizations*, SIAM J. Numer. Anal., 42 (2004), pp. 974–996.

[25] Gottlieb S and C. W. Shu, *Total variation diminishing Runge–Kutta schemes*, Math. Comp., 67 (1998), pp. 73–85.

[26] Gottlieb S, C. W. Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Rev., 43 (2001), pp. 89–112.

[27] C. W. Shu and S. Osher, *Efficient implementation of essentially nonoscillatory shock-capturing schemes*, J. Comput. Phys., 77 (1988), pp. 439–471.

[28] R. J. Spiteri and S. J. Ruuth, *A new class of optimal high-order strong-stability-preserving time discretization methods*, SIAM J. Numer. Anal., 40 (2002), pp. 469–491.

[29] R. J. Spiteri and S. J. Ruuth, *Non-linear evolution using optimal fourth-order strong-stability-preserving Runge–Kutta methods*, Math. Comput. Simulation, 62 (2003), pp. 125–135.