HIGH-ORDER STRONG-STABILITY-PRESERVING RUNGE-KUTTA METHODS WITH DOWNWIND-BIASED SPATIAL DISCRETIZATIONS

STEVEN J. RUUTH * AND RAYMOND J. SPITERI†

Abstract. Strong-stability-preserving Runge-Kutta (SSPRK) methods are a specific type of time discretization method that have been widely used for the time evolution of hyperbolic partial differential equations (PDEs). Under a suitable stepsize restriction, these methods share a desirable nonlinear stability property with the underlying PDE; e.g., stability with respect to total variation, maximum norm, or other convex functional. This is of particular interest when the solution exhibits shock-like or other nonsmooth behaviour. Many results are known for SSPRK methods with nonnegative coefficients. However, it has been recently shown that such methods cannot exist with order greater than four. In this paper, we give a systematic treatment of explicit SSPRK methods with general (i.e., possibly negative) coefficients up to order five. In particular, we show how to optimally treat negative coefficients (corresponding to a change in the upwind direction of the spatial discretization) in the context of *effective CFL coefficient* maximization and provide proofs of optimality of some explicit SSPRK methods of orders 1 to 4. We also give the first known explicit fifth-order SSPRK schemes and show their effectiveness in practice versus more well-known fifth-order explicit Runge-Kutta schemes.

Key words. downwinding, strong stability preserving, total variation diminishing, Runge-Kutta methods, high-order accuracy, time discretization

AMS subject classifications. 65L06, 65M20

1. Introduction.

Solutions to hyperbolic partial differential equations (PDEs) are commonly approximated by sequentially discretizing the spatial and temporal derivatives. For example, in the method of lines, a discretization of the spatial derivatives of the PDE is carried out to produce a large set of coupled time-dependent ordinary differential equations (ODEs). These ODEs can then be treated by suitable time-stepping techniques such as linear multi-step or Runge-Kutta methods.

In the numerical solution of hyperbolic PDEs, difficulties may arise due to the presence of shock waves or other discontinuous behaviour. In particular, the numerical solution to such problems often suffers from spurious oscillations or overshoots. This usually represents unphysical behaviour, and it is almost always desirable to use a numerical method that suppresses it. One of the first families of such schemes were called *total variation diminishing* (TVD); see [22, 21]. Following more recent work[4], we refer to them as *strong stability preserving* (SSP).

In particular, we are interested in the development, analysis, and optimization of SSP Runge-Kutta (SSPRK) time-stepping methods for the hyperbolic conservation law

$$u_t + f(u)_x = 0, \tag{1.1}$$

subject to appropriate initial conditions. When a SSPRK method with nonnegative coefficients is used it is convenient to consider a semi-discretization of (1.1) in space

*Department of Mathematics, Simon Fraser University, Burnaby, British Columbia, V5A 1S6 Canada (sruuth@sfu.ca). The work of this author was partially supported by a grant from NSERC Canada.

†Department of Computer Science, Dalhousie University, Halifax, Nova Scotia, B3H 1W5 Canada (spiteri@cs.dal.ca). The work of this author was partially supported by a grant from NSERC Canada.

to yield a large coupled set of ODEs

$$\dot{U} = F(U). \tag{1.2}$$

More generally, following [22, 21, 3, 4], upwind-biased ($F(U)$) and downwind-biased ($\tilde{F}(U)$) spatial discretizations may be applied in some combination to achieve favourable nonlinear stability properties for a given time-stepping scheme. For simplicity we refer to upwind-biased and downwind-biased spatial discretizations as upwind and downwind spatial discretizations respectively.

Optimal explicit SSPRK schemes with nonnegative coefficients and where the number of stages s is equal to the order p for $s = p = 1, 2$, and 3 have been known for some time. Gottlieb and Shu [3] showed that no such method exists with nonnegative coefficients when $s = p = 4$. In [25], Spiteri and Ruuth proposed a new class of explicit SSPRK methods with nonnegative coefficients with $s > p$. They gave optimal explicit SSPRK schemes with s stages and orders 1 and 2 (see also [21, 3]), as well as specific schemes for $p = 3$, $s = 4, 5$ and $p = 4$, $s = 5$. The advantage afforded by these high-stage schemes is that the increase in the CFL coefficient allows for a large enough increase in the stable time step to more than offset the increase in computational cost per step. However, in [20] they showed that it was impossible to have an explicit SSPRK method with order greater than 4 with nonnegative coefficients. In this paper, we give a unified treatment of all explicit SSPRK schemes with positive and/or negative coefficients of up to order 5 in terms of the effective CFL coefficient. We find that many of the optimal explicit SSPRK methods under the constraint of nonnegative coefficients are also optimal in terms of effective CFL coefficient when negative coefficients are allowed. We also present the first fifth-order explicit SSPRK methods.

We remark that explicit fifth-order SSP *multistep* schemes have been successfully constructed [21, 13, 4]. The most efficient scheme of this type that explicitly appears in the literature [21] is

$$\begin{aligned} U^{n+1} = {} & \frac{7}{20}U^n + \frac{3}{10}U^{n-1} + \frac{4}{15}U^{n-2} + \frac{7}{120}U^{n-4} + \frac{1}{40}U^{n-5} + \frac{291201}{108000}F(U^n) \\ & - \frac{198401}{86400}\tilde{F}(U^{n-1}) + \frac{88063}{43200}F(U^{n-2}) - \frac{17969}{43200}\tilde{F}(U^{n-4}) + \frac{73061}{432000}F(U^{n-5}) \end{aligned} \tag{1.3}$$

This six-step scheme involves evaluations of both upwind and downwind operators and has an effective CFL coefficient of 0.065. In this paper we construct explicit SSPRK methods with up to a 325% improvement in effective CFL coefficient over this scheme. Comparable gains are also shown to arise in practice.

We further note that in this paper we deal with explicit Runge-Kutta methods where the number of stages s can be substantially larger than the order p. These methods are optimized with respect to effective CFL coefficient, which is a theoretical measure of the stepsize restriction required for nonlinear stability. Although perhaps similar at first glance, this is not in general related to maximizing the area of the (linear) stability region of a Runge-Kutta method; see [3] for a counter-example. For work on the optimization of the linear stability regions of explicit Runge-Kutta methods, we refer to [15] and the references therein.

The remainder of the paper is organized as follows. In Section 2 we review some relevant results on SSP schemes as well as define important concepts such as effective CFL coefficient. In Sections 3 and 4 we use analytical as well as numerical techniques to find explicit SSPRK methods up to order five with optimal effective

CFL coefficients. In Section 5 we show the efficiency of the new optimized fifth-order explicit SSPRK methods versus the optimal fifth-order multistep method and a commonly used fifth-order explicit Runge-Kutta method. Finally in Section 6 we conclude by summarizing the main findings of the paper.

2. Background on SSP Schemes. In this section we give some theoretical background on SSPRK schemes. We begin by recalling the definition of strong stability:

DEFINITION 2.1. *A sequence* $\{U^n\}$ *is said to be* strongly stable *in a given semi-norm* $||\cdot||$ *if* $||U^{n+1}|| \leq ||U^n||$ *for all* $n \geq 0$.

Strong stability turns out to have an interesting relationship to the more classical concept of *contractivity* (see e.g., [23, 7, 8]). In this case for equations (1.2) satisfying a one-sided Lipschitz condition, we have that the distance between all exact solutions starting from different initial conditions is nonincreasing in time. It is reasonable to then require the same property of the numerical solution; i.e., $||\tilde{U}^{n+1} - U^{n+1}|| \leq ||\tilde{U}^n - U^n||$ for all $n \geq 1$. In classical stability analysis, $\tilde{U}^n$ is usually assumed to be a perturbation of U^n. It is interesting that many of the optimal SSP schemes found in [25] agree with optimal contractive schemes in [8]. In fact, recent work by Ferracina and Spijker [2] for schemes with positive coefficients shows that the step size coefficient C (see below) for strong stability is equivalent to the related quantity $R(A, b)$ [8] arising in contractivity studies.

To begin our analysis, assume that upwind spatial discretizations are appropriate and consider an s-stage, explicit Runge-Kutta method written in the form

$$U^{(0)} = U^n \tag{2.1a}$$

$$U^{(i)} = \sum_{k=0}^{i-1} (\alpha_{ik} U^{(k)} + \Delta t \beta_{ik} F(U^{(k)})), \qquad i = 1, 2, \ldots, s, \tag{2.1b}$$

$$U^{n+1} = U^{(s)}, \tag{2.1c}$$

where all the $\alpha_{ik} \geq 0$ and $\alpha_{ik} = 0$ only if $\beta_{ik} = 0$ [21].

For consistency, we must have that $\sum_{k=0}^{i-1} \alpha_{ik} = 1$, $i = 1, 2, \ldots, s$. Hence, if both sets of coefficients α_{ik}, β_{ik} are nonnegative, then (2.1) is a convex combination of forward Euler steps with various step sizes $\frac{\beta_{ik}}{\alpha_{ik}} \Delta t$. The strong stability property follows easily from this observation.

The Runge-Kutta scheme (2.1) is not written in standard Butcher array form; however, the representation (2.1) maps uniquely to a Butcher array. On the other hand, written in this form, it is particularly convenient to make use of the following result [22, 4]:

THEOREM 2.2. *If the forward Euler method is strongly stable under the CFL restriction* $\Delta t \leq \Delta t_{FE}$, *then the Runge-Kutta method (2.1) with* $\beta_{ik} \geq 0$ *is SSP provided*

$$\Delta t \leq C \Delta t_{FE},$$

where C *is the CFL coefficient*

$$C \equiv \min_{i,k} \frac{\alpha_{ik}}{\beta_{ik}}.$$

SSPRK schemes with negative coefficients β_{ik} are also possible with the appropriate interpretation. Following the procedure first suggested in [21], whenever $\beta_{ik} < 0$, the operator $\tilde{F}(\cdot)$ is used instead of $F(\cdot)$, where $\tilde{F}(\cdot)$ approximates the same derivatives as $F(\cdot)$ but is assumed to be strongly stable for Euler's method solved *backwards* in time under a suitable time-step restriction. In practice, this corresponds to a change in upwinding direction, or in other words, *downwinding*. This allows the following generalization of Theorem 2.2:

THEOREM 2.3. *Let Euler's method applied forward in time combined with the spatial discretization $F(\cdot)$ be strongly stable under the CFL restriction $\Delta t \leq \Delta t_{FE}$. Let Euler's method applied backward in time combined with the spatial discretization $\tilde{F}(\cdot)$ also be strongly stable under the same CFL restriction $\Delta t \leq \Delta t_{FE}$. Then the Runge-Kutta method (2.1) is SSP provided*

$$\Delta t \leq C \Delta t_{FE},$$

where C is the CFL coefficient

$$C \equiv \min_{i,k} \frac{\alpha_{ik}}{|\beta_{ik}|}, \tag{2.2}$$

where $\beta_{ik} F(\cdot)$ is replaced by $\beta_{ik}\tilde{F}(\cdot)$ whenever β_{ik} is negative.

We note that the assumptions on strong stability of Euler's method applied forward and backward in time restricts the theoretical advantages of this result to non-dissipative equations such as (1.1).

Irreducible explicit Runge-Kutta methods have one (new) function evaluation per stage. We note that if every coefficient β_{ik} is positive, then the number of stages is trivially equal to the number of function evaluations. However, if both $F(U^{(k)})$ and $\tilde{F}(U^{(k)})$ are required for some k, the Runge-Kutta method (2.1) has more function evaluations[1] than stages. So the first step in creating a fair comparison of the computational cost of a given Runge-Kutta method and in deriving optimal schemes is to consider general methods that allow only one (new) function evaluation per stage. A necessary and sufficient condition for this is that the non-zero coefficients β_{ik} for a given k are all of the same sign. To see this, let $\mathcal{K}_-$ be the set of levels k such that all $\beta_{ik} \leq 0$ and we consider

$$\begin{aligned}
U^{(0)} &= U^n \\
U^{(i)} &= \sum_{k=0}^{i-1} \alpha_{ik} U^{(k)} + \Delta t \begin{cases} \beta_{ik}\tilde{F}(U^{(k)})) & k \in \mathcal{K}_- \\ \beta_{ik} F(U^{(k)})) & \text{otherwise} \end{cases}, \qquad i = 1, 2, \ldots, s-1 \\
U^{n+1} &= U^{(s)}
\end{aligned} \tag{2.3}$$

For the remainder of the paper, we will tacitly assume that the schemes under consideration are of this form. Naturally schemes that are written combining positive and negative coefficients β_{ik} within a given level k can be augmented with additional stages to be of this form. Thus, without loss of generality, we have that the total number of evaluations of $F(\cdot)$ and $\tilde{F}(\cdot)$ is identically equal to the number of stages of the method.

We note that this formulation allows one to search for the optimal scheme for a given order and a *given number of stages* (function evaluations). This is a more

[1] The only difference between $\tilde{F}(\cdot)$ and $F(\cdot)$ is a change in the upwind direction; so $\tilde{F}(\cdot)$ can clearly be computed with the same cost as $F(\cdot)$ [4]. Indeed, recent studies make the assumption that if both $\tilde{F}(U^{(k)})$ and $F(U^{(k)})$ must be computed for some k, the cost as well as the storage requirements for that k doubles [3, 4, 25, 17]; i.e., each is given equal weight.

appropriate description of what should be optimized than has been considered in the literature thus far. For example, searching for the scheme with the largest CFL coefficient (or even effective CFL coefficient, see below) *for a given order* results in the number of stages tending to infinity.

Another advantage to this formulation is that schemes can be represented and implemented in Butcher array form using (3.6) since differences of the form $F(U^{(i)}) - \tilde{F}(U^{(i)})$ do not arise; i.e., the method can be implemented as

$$K_i = \begin{cases} F\left(U^n + \Delta t \sum_{j=1}^{i-1} a_{ij} K_j\right) & \text{if } b_i \geq 0, \\ \tilde{F}\left(U^n + \Delta t \sum_{j=1}^{i-1} a_{ij} K_j\right) & \text{otherwise,} \end{cases} \qquad i = 1, 2, \ldots, s,$$

$$U^{n+1} = U^n + \Delta t \sum_{i=1}^{s} b_i K_i.$$

This form is often desirable for implementing fifth-order schemes because the storage requirements can be reduced. We further remark that the differences $F(U^{(i)}) - \tilde{F}(U^{(i)})$ contribute to artificial dissipation and smearing. For example, this difference is proportional to the discrete Laplacian when first-order upwinding is applied to the linear advection equation. A natural consequence of our formulation is that during optimization these dissipative differences do not arise, leading to schemes with smaller errors and less smearing than would otherwise occur.

In Section 5, we compare the computational efficiencies of various Runge-Kutta methods. In order to make a fair comparison of the relative efficiencies of these methods and to derive optimal schemes we make the following definition.

DEFINITION 2.4. *The* effective CFL coefficient C_{eff} *of an SSPRK method is C/s where C is the CFL coefficient of the method and s is the number of stages (function evaluations) required for one step of the method.*

As conjectured in Shu and Osher [22] and subsequently proven in Gottlieb and Shu [3], the optimal two-stage, order-two explicit SSPRK scheme with nonnegative coefficients is the modified Euler scheme,

$$\begin{aligned} U^{(1)} &= U^n + \Delta t F(U^n), \\ U^{n+1} &= \frac{1}{2}U^n + \frac{1}{2}U^{(1)} + \frac{1}{2}\Delta t F(U^{(1)}). \end{aligned}$$

It has a CFL restriction $\Delta t \leq \Delta t_{FE}$, which implies a CFL coefficient of 1. Henceforth, we will refer to this scheme as SSP(2,2). In general, we adopt the convention of referring to the best (in terms of effective CFL coefficient) known s-stage, order-p explicit SSPRK scheme as SSP(s,p), where s is equal to the total number of function evaluations of $F(\cdot)$ and $\tilde{F}(\cdot)$. In [25] a class of s-stage, order-two explicit SSPRK schemes was given and proved to be optimal with a CFL coefficient of $s - 1$.

Shu and Osher [22] also conjectured that the optimal three-stage, order-three explicit SSPRK scheme with nonnegative coefficients is

$$\begin{aligned} U^{(1)} &= U^n + \Delta t F(U^n), \\ U^{(2)} &= \frac{3}{4}U^n + \frac{1}{4}U^{(1)} + \frac{1}{4}\Delta t F(U^{(1)}), \\ U^{n+1} &= \frac{1}{3}U^n + \frac{2}{3}U^{(2)} + \frac{2}{3}\Delta t F(U^{(2)}), \end{aligned}$$

which has a CFL coefficient of 1 as well. The optimality of this scheme was later proved by Gottlieb and Shu [3]. This scheme is commonly called the *Third-Order TVD Runge-Kutta scheme*, but we will refer to it as SSP(3,3).

In [20], Ruuth and Spiteri derived a linear bound that can be used to prove that the optimal four-stage, order-three explicit SSPRK scheme with nonnegative coefficients is

$$\begin{aligned}
U^{(1)} &= U^n + \frac{1}{2}\Delta t F(U^n),\\
U^{(2)} &= U^{(1)} + \frac{1}{2}\Delta t F(U^{(1)}),\\
U^{(3)} &= \frac{2}{3}U^n + \frac{1}{3}U^{(2)} + \frac{1}{6}\Delta t F(U^{(2)}),\\
U^{n+1} &= U^{(3)} + \frac{1}{2}\Delta t F(U^{(3)}),
\end{aligned}$$

which has a CFL coefficient of 2. This observation appears in [25]. Following [25] we will refer to this scheme as SSP(4,3).

Moving on to methods with five stages and order three gives a numerically optimized scheme, SSP(5,3), with a CFL coefficient of approximately 2.65. It can be proven that this is also the optimal explicit SSPRK scheme with five stages and order three via the following line of reasoning. The CFL coefficient C of SSP(5,3) is equal to the *radius of absolute monotonicity* $R(A, b)$ for linear constant-coefficient problems [7]. Because $C \leq R(A, b)$ [2] and C (and $R(A, b)$) for nonlinear problems cannot exceed that for linear problems, we conclude that SSP(5,3) is the optimal five-stage, third-order explicit SSPRK scheme. A similar line of reasoning can be applied to prove the optimality of SSP(3,3), SSP(4,3), as well as the first- and second-order SSP schemes. Indeed we have produced schemes of the form SSP(s,3), $s \leq 9$, with CFL coefficients equal to $R(A, b)$ for linear constant-coefficient problems; hence they are also optimal SSP schemes for nonlinear problems. It is worth mentioning, however, that this approach does not seem to be useful for proving the optimality of schemes of order greater than 3.

The main advantage offered by these high-stage schemes is that the additional computational cost incurred per step is more than offset by the increase in stable step size. For example, SSP(4,3) costs 33% more than SSP(3,3) but offers a 100% larger CFL coefficient. Thus for SSP(4,3), $C_{\text{eff}} = 2/4 = 1/2$, whereas for SSP(3,3), $C_{\text{eff}} = 1/3$. This translates into a $(1/2 - 1/3)/1/3 = 50\%$ increase in computational efficiency.

In [3], Gottlieb and Shu proved that it is impossible to have an explicit SSPRK method of order four in four stages having only nonnegative coefficients[2] . In Section 3.3 we prove the stronger result that it is in fact impossible to obtain a explicit SSPRK method of order four with any four (general) function evaluations. In [25] a five-stage, order-four explicit SSPRK scheme with nonnegative coefficients is given. It turns out that this scheme coincides with Kraaijevanger's optimal five-stage, order-four contractive scheme [8, 2]. A further study of explicit SSPRK methods of order four and $s = 6, 7, 8$ stages can be found in [24]. For the examples investigated in that paper, it was found that the increased stage number did lead to noteworthy improvements in practical performance. It is also worth mentioning that these high-stage schemes have modest storage requirements.

[2]We remark that a proof that it is also impossible to have a fourth-order with four stages and $R(A, b) > 0$ appeared earlier and independently of this [8].

In [20] it is shown that explicit SSPRK schemes with nonnegative coefficients do not exist with order greater than four. A similar restriction to orders four or less was proven for contractive schemes [8]. This means that the search for explicit schemes of order five and higher must involve evaluations of the downwinded operator $\tilde{F}(\cdot)$. In the remainder of the paper we present a unified treatment of explicit SSPRK schemes that use both upwinded and downwinded operators in terms of effective CFL coefficient and prove the optimality of several lower-order schemes.

In Section 4 we give the first fifth-order explicit SSPRK methods, optimized in terms of effective CFL coefficient. A fifth-order explicit SSPRK method in the form (2.1) was thought to have been found in [22], based on a fifth-order explicit Runge-Kutta scheme on page 143 of [9], which was in turn based on a particular choice from a family of fifth-order explicit Runge-Kutta schemes that appeared in [12]. The family of schemes in question is described by the Butcher tableau

$$\begin{array}{c|cccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\gamma & \gamma & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{4} & \frac{1}{4} - \frac{1}{32}\gamma^{-1} & \frac{1}{32}\gamma^{-1} & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} - 32\sigma - \frac{1}{8}\frac{1-64\sigma}{\gamma} & \frac{1}{8}\frac{1-64\sigma}{\gamma} & 32\sigma & 0 & 0 & 0 \\
\frac{3}{4} & -\frac{9}{16} + 24\sigma - \frac{6\sigma - \frac{3}{16}}{\gamma} & \frac{6\sigma-\frac{3}{16}}{\gamma} & \frac{3}{4} - 24\sigma & \frac{9}{16} & 0 & 0 \\
1 & \frac{11}{7} - \frac{384}{7}\sigma - \frac{1}{7}\frac{\frac{7}{2}-96\sigma}{\gamma} & \frac{1}{7}\frac{\frac{7}{2}-96\sigma}{\gamma} & \frac{384}{7}\sigma & -\frac{12}{7} & \frac{8}{7} & 0 \\
\hline
 & \frac{7}{90} & 0 & \frac{16}{45} & \frac{2}{15} & \frac{16}{45} & \frac{7}{90}
\end{array} \quad . \tag{2.4}$$

Unfortunately, although this family of explicit Runge-Kutta schemes (2.4) is indeed fifth order, there is an error in the particular member of this family upon which the reported fifth-order explicit SSPRK method was based. This proposed scheme has coefficient matrix A given by

$$A = \begin{bmatrix}
0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{8} & \frac{1}{8} & 0 & 0 & 0 & 0 \\
0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\
0 & -\frac{3}{16} & \frac{3}{8} & \frac{9}{16} & 0 & 0 \\
\frac{1}{7} & \frac{4}{7} & \frac{6}{7} & -\frac{12}{7} & \frac{8}{7} & 0
\end{bmatrix}.$$

This scheme was meant to correspond to the particular choice of $\sigma = \frac{1}{64}$ and (arbitrary) $\gamma = \frac{1}{2}$. However it is easily verified that this scheme does not belong to the family of explicit fifth-order schemes (2.4), differing in the coefficients a_{31} and a_{32}. In fact is only *second order*; e.g., it is easily seen that the third-order condition $b^T Ac = \frac{1}{6}$ is not satisfied.

3. Optimal SSPRK Methods. In this section we prove some optimality results in terms of effective CFL number for some low-order explicit SSPRK methods

($p = 1, 2, 3$). Optimal schemes for high-order methods ($p = 4, 5$) are determined numerically. We begin by describing the form of the optimization problem solved in all cases. We then prove some existence and optimality of some explicit SSPRK methods of up to order 4. Section 4 contains some numerical results for the first explicit SSPRK methods of order 5.

3.1. Formulation of the Optimization Problem. We seek to optimize an s-stage, order-p explicit SSPRK scheme by maximizing its effective CFL coefficient according to Theorem 2.3. That is, we seek the global maximum of the nonlinear programming problem

$$\max_{(\alpha_{ik},\beta_{ik})} \min \frac{\alpha_{ik}}{|\beta_{ik}|}, \tag{3.1}$$

where $\alpha_{ik},\ \beta_{ik}, k = 0, 1, \ldots, i-1,\ i = 1, 2, \ldots, s$ are real and $0 \le \alpha_{ik} \le 1$. As noted in Section 2, we insist that for each k and $i = k+1, k+2, \ldots, s$ that $\beta_{ik} \ge 0$ or $\beta_{ik} \le 0$ to ensure that the number of function evaluations corresponds to the number of stages. The case $\alpha_{ik} = \beta_{ik} = 0$ is defined as `NaN` in the sense that it is not included in the minimization process if it occurs. The objective function (3.1) is also subject to the constraints

$$\sum_{k=0}^{i-1} \alpha_{ik} = 1, \qquad i = 1, 2, \ldots, s, \tag{3.2}$$

$$\sum_{j=1}^{s} b_j \Phi_j(t) = \frac{1}{\gamma(t)}, \quad t \in T_q, \quad q = 1, 2, \ldots, p. \tag{3.3}$$

Here, the functions $\Phi_j(t)$ are nonlinear constraints that are polynomial in $\alpha_{ik},\ \beta_{ik}$ and that correspond to the order conditions for a Runge-Kutta method to be of order p (see e.g., [5]); i.e., T_q stands for the set of all rooted trees of order equal to q. The number of constraints represented by the Runge-Kutta order conditions is equal to

$$\sum_{q=1}^{p} \mathrm{card}(T_q),$$

where $\mathrm{card}(T_q)$ is the cardinality of T_q. Also, we use the notation b_j in the usual sense of the Butcher array representation of a Runge-Kutta method; again this would be a polynomial function of the coefficients α_{ik} and β_{ik}. It can be expected that the particular choice of coefficients $\alpha_{ik},\ \beta_{ik}$ that maximizes the quantity (2.2) for a given Runge-Kutta method will be naturally produced by the solution to this nonlinear programming problem; hence the result will be a sharp estimate of the CFL coefficient.

However, this formulation of the nonlinear programming problem does not lend itself easily to numerical solution; see [25] for further discussion. By introducing a dummy variable z, the nonlinear programming problem can be reformulated as

$$\max_{(\alpha_{ik},\beta_{ik})} z, \tag{3.4a}$$

subject to

$$\alpha_{ik} \geq 0, \tag{3.4b}$$

$$\beta_{k+1,k}, \beta_{k+2,k}, \ldots, \beta_{sk} \geq 0, \tag{3.4c}$$

$$\text{or} \quad \beta_{k+1,k}, \beta_{k+2,k}, \ldots, \beta_{sk} \leq 0, \quad k = 0, \ldots, s-1,$$

$$\sum_{k=0}^{i-1} \alpha_{ik} = 1, \qquad i = 1, 2, \ldots, s, \tag{3.4d}$$

$$\sum_{j=1}^{s} b_j \Phi_j(t) = \frac{1}{\gamma(t)}, \quad t \in T_q, \quad q = 1, 2, \ldots, p, \tag{3.4e}$$

$$\alpha_{ik} - z|\beta_{ik}| \geq 0, \quad k = 0, 1, \ldots, i-1, \ i = 1, 2, \ldots, s. \tag{3.4f}$$

Numerical optimization software may be applied to the reformulated problem (3.4) for various combinations of s and p. In our initial approach we considered using Matlab's Optimization Toolbox but found that it was nontrivial to determine an initial guess to start the nonlinear iteration. Subsequent efforts focussed on BARON [26], a deterministic, global optimization software package that uses algorithms of the branch-and-bound type. This approach was found to be superior to Matlab's Optimization Toolbox in the sense that it is faster, gives improved optima, and satisfies active constraints to 15 decimal digits.

In each of the cases $s = 7, 8, 9$ numerically optimal fifth-order SSPRK schemes were found in less than 90 minutes on a (shared) cluster of 96 dual 1.2 GHz Athlon processors with BARON. See [19] for further details on applying BARON to the optimization of SSPRK schemes.

3.2. Optimality of Some Low-Order Methods. We now give new results on optimal effective CFL coefficients for some low-order explicit SSPRK methods. Previous results primarily focus on optimizing raw CFL coefficients for methods with nonnegative coefficients. Here we give existence and optimality results in the context of effective CFL coefficients for methods with no sign restriction on their coefficients.

THEOREM 3.1. *For $s = 1, 2, 3, \ldots$, the optimal s-stage explicit SSPRK method of order 1 has effective CFL coefficient 1 and can be represented in the form of SSP(s,1); i.e.,*

$$\alpha_{ik} = \begin{cases} 1 & k = i-1, \\ 0 & \textit{otherwise}. \end{cases}, \qquad \beta_{ik} = \begin{cases} \frac{1}{s} & k = i-1, \\ 0 & \textit{otherwise}. \end{cases}, \qquad i = 1, 2, \ldots, s.$$

Before giving the proof of Theorem 3.1, we introduce the following notation and give two useful Lemmas. We find it convenient to write the general s-stage explicit Runge-Kutta method in the following form (cf. [3]):

$$U^{(0)} = U^n, \tag{3.5a}$$

$$U^{(i)} = U^{(0)} + \Delta t \sum_{k=0}^{i-1} \kappa_{ik} \begin{cases} F(U^{(k)}) & \text{if } \kappa_{ik} \geq 0 \\ \tilde{F}(U^{(k)}) & \text{otherwise} \end{cases}, \quad i = 1, 2, \ldots, s, \tag{3.5b}$$

$$U^{n+1} = U^{(s)}. \tag{3.5c}$$

Using the fact that the β_{ik} at a particular level are all of the same sign, the coefficients

κ_{ik} are related to the coefficients $\alpha_{ik},\ \beta_{ik}$ recursively by

$$\kappa_{ik} = \beta_{ik} + \sum_{j=k+1}^{i-1} \alpha_{ij}\kappa_{jk}. \tag{3.6}$$

We remark that the coefficients κ_{ik} can be related to the Butcher array quantities $a_{ik},\ b_k$ by

$$\begin{aligned} a_{ik} &= \kappa_{i-1,k-1}, \qquad & k &= 1,2,\ldots,i-1, \quad i = 1,2,\ldots,s-1,\\ b_k &= \kappa_{s,k-1}, & k &= 1,2,\ldots,s. \end{aligned}$$

It is also important to note that $\mathrm{sgn}(\kappa_{ik}) = \mathrm{sgn}(\beta_{ik})$, motivating the use of (3.5).

LEMMA 3.2. *If a method of the form (2.1) with $\alpha_{ik} \geq 0$ has a CFL coefficient $c \geq m > 0$, then $0 \leq |\kappa_{ik}| \leq \frac{1}{m}$ for all $k = 0,1,\ldots,i-1$, $i = 1,2,\ldots,s$.*

Proof. From Theorem 2.3, if $c \geq m > 0$, then $\alpha_{ik} \geq m|\beta_{ik}|$, or equivalently $|\beta_{ik}| \leq \frac{1}{m}\,\alpha_{ik}$, for all $i,\ k$ such that $\alpha_{ik} \neq 0$.

Now,

$$\alpha_{ik} \geq 0, \quad \sum_{k=0}^{i-1} \alpha_{ik} = 1, \quad i = 1,2,\ldots,s, \quad \Rightarrow \quad \alpha_{ik} \leq 1$$

for all $i,\ k$. Hence, $|\beta_{ik}| \leq \frac{1}{m}$ for all $i,\ k$. In particular, $|\kappa_{10}| = |\beta_{10}| \leq \frac{1}{m}$ for any valid explicit SSPRK method.

We now proceed by induction on stage ℓ of an s-stage method. Assume $|\kappa_{ij}| \leq \frac{1}{m}$ for $j = 0,1,\ldots,\ell-1$; $i = 1,2,\ldots,\ell$. (We have just shown that this result holds for $\ell = 1$.) Now consider stage $(\ell+1)$ of a valid explicit SSPRK method; i.e., consider coefficients $\kappa_{\ell+1,k}$ for $k = 0,1,\ldots,\ell$ with

$$\sum_{k=0}^{\ell} \alpha_{\ell+1,k} = 1.$$

Then using (3.6),

$$\begin{aligned} |\kappa_{\ell+1,0}| &= \left| \sum_{k=1}^{\ell} \alpha_{\ell+1,k}\kappa_{k0} + \beta_{\ell+1,0} \right| \\ &\leq \sum_{k=1}^{\ell} \alpha_{\ell+1,k}|\kappa_{k0}| + |\beta_{\ell+1,0}| \\ &\leq \frac{1}{m} \sum_{k=1}^{\ell} \alpha_{\ell+1,k} + \frac{1}{m}\,\alpha_{\ell+1,0} \\ &= \frac{1}{m}\,. \end{aligned}$$

Similar arguments can be used to show $|\kappa_{\ell+1,j}| \leq \frac{1}{m}$ for $j = 1,2,\ldots,\ell$. The Lemma now follows by induction. ∎

LEMMA 3.3. *Suppose a consistent s-stage explicit SSPRK method (2.1) has coefficients $\beta_{ik} \leq 0$ at ℓ distinct stages; i.e., $\beta_{ik} \leq 0$ for all i and $k = k_1, k_2, \ldots, k_\ell$ with $0 \leq k_1 < k_2 < \cdots < k_\ell \leq s-1$. Then the CFL coefficient C of the method satisfies $C \leq s - \ell$.*

Proof. Because the method is consistent, we have

$$\sum_{k=0}^{s-1} \kappa_{sk} = 1. \tag{3.7}$$

But by the definition of κ_{ik} it is clear that $\kappa_{sk_1}, \kappa_{sk_2}, \ldots, \kappa_{sk_\ell} \leq 0$. Thus

$$\sum_{\substack{k=0 \\ k\neq k_1,k_2,\ldots,k_\ell}}^{s-1} \kappa_{sk} \geq 1. \tag{3.8}$$

The desired result $C \leq s - \ell$ now follows immediately from applying Lemma 3.2 to (3.8). ■

Proof of Theorem 3.1. For nonnegative coefficients $\{\beta_{ik}\}$ the result for the raw CFL coefficient has been shown in [25]. By Lemma 3.3, a method containing any $\beta_{ik} < 0$ must have a CFL coefficient $C \leq s - 1 < s$, and thus we must have $C_{\text{eff}} < 1$. This completes the proof. ■

REMARK 1. *As noted [25], despite the increase in raw CFL coefficient, these first-order methods do not offer a theoretical computational advantage.*

THEOREM 3.4. *For $s = 2, 3, 4, \ldots$, the optimal s-stage explicit SSPRK method of order 2 has effective CFL coefficient $\frac{s-1}{s}$ and can be represented in the form of SSP(s,2); i.e.,*

$$\alpha_{ik} = \begin{cases} 1 & k = i-1, \\ 0 & \textit{otherwise.} \end{cases}, \quad \beta_{ik} = \begin{cases} \frac{1}{s-1} & k = i-1, \\ 0 & \textit{otherwise.} \end{cases}, \quad i = 1, 2, \ldots, s-1.$$

$$\alpha_{ik} = \begin{cases} \frac{1}{s} & k = 0, \\ \frac{s-1}{s} & k = s-1 \\ 0 & \textit{otherwise.} \end{cases}, \qquad \beta_{ik} = \begin{cases} \frac{1}{s} & k = s-1, \\ 0 & \textit{otherwise.} \end{cases}, \qquad i = s.$$

Proof. For nonnegative coefficients $\{\beta_{ik}\}$ the result for the raw CFL coefficient has been shown in [25]. By Lemma 3.3, any consistent, s-stage method with some $\beta_{ik} < 0$ must have a CFL coefficient $C \leq s - 1$, and thus we must have $C_{\text{eff}} \leq \frac{s-1}{s}$. This completes the proof. ■

REMARK 2. *As noted [25], in this case the theoretical increase in raw CFL coefficient more than offsets the increased work per step, leading to an overall computational advantage with increasing s. However, the effective CFL coefficient is bounded above by 1.*

We now give some specific optimality results for methods of order 3.

THEOREM 3.5. *The optimal 3-stage explicit SSPRK method of order 3 has effective CFL coefficient $C_{\text{eff}} = 1/3$, and an optimal representation is given by SSP(3,3).*

Proof. For nonnegative coefficients $\{\beta_{ik}\}$ the result for the raw CFL coefficient has been shown in [3].

Now suppose we allow $\beta_{ik} < 0$ in an attempt to improve the CFL coefficient. From the third-order condition $b^T Ac = 1/6$, we have $\beta_{10}\beta_{21}\beta_{32} = 1/6 > 0$; so the scheme must have $\beta_{ik} \leq 0$ at exactly two levels. But then we may apply Lemma 3.3 to show that $C \leq 1$, and hence its $C_{\text{eff}} \leq 1/3$. ■

THEOREM 3.6. *The optimal 4-stage explicit SSPRK method of order 3 has effective CFL coefficient $C_{\text{eff}} = 2/3$, and an optimal representation is given by SSP(4,3).*

Proof. For nonnegative coefficients $\{\beta_{ik}\}$ the result for the raw CFL coefficient has been shown in [25]. By Lemma 3.3 it is clear that $C \leq 2$ if the $\beta_{ik} \leq 0$ at two or more levels. So the only possibility for an improvement in the CFL coefficient over SSP(4,3) is if $\beta_{ik} \leq 0$ at precisely one level. But then by the third-order condition $b^T Ac = 1/6$, one of the following must hold:

$$\begin{aligned}
\kappa_{43}\kappa_{32}\kappa_{21} &\geq 1/6,\\
\kappa_{43}\kappa_{32}\kappa_{20} &\geq 1/6,\\
\kappa_{43}\kappa_{31}\kappa_{10} &\geq 1/6,\\
\kappa_{42}\kappa_{21}\kappa_{10} &\geq 1/6.
\end{aligned}$$

Supposing that the CFL coefficient is greater than 2 in any of these statements leads to a condition of the form $\kappa_{4i}\kappa_{ij}\kappa_{jk} \leq 1/8$, $i = 2,3$, $1 \leq j \leq i-1$, $0 \leq k \leq j-1$ by Lemma 3.2 and gives rise to a contradiction. Hence the optimal scheme must be SSP(4,3). ■

3.3. A Fourth-Order Result. In this section we demonstrate that, even allowing negative coefficients β_{ik}, there is no four-stage explicit SSPRK method of order 4. We begin with a lemma.

LEMMA 3.7. *If $s = p$, the β_{ik} at a particular level k, for some $0 \leq k \leq s-1$, $k+1 \leq i \leq s$, are all of the same sign, i.e., $\beta_{ik}\beta_{jk} \geq 0$ for $k+1 \leq i,j \leq s$, and the CFL coefficient is positive, then $\kappa_{ik} \neq 0$ for $k+1 \leq i \leq s$.*

Proof. From the order conditions, we have $\prod_{i=1}^{p} \beta_{i,i-1} = \frac{1}{p!}$, so each $\beta_{i,i-1} \neq 0, 1 \leq i \leq s$. Since the CFL coefficient is positive, this implies each $\alpha_{i,i-1} > 0, 1 \leq i \leq s$. Expanding κ_{ij} in terms of the α and β coefficients (see, e.g., [3]) it is easily seen that $|\kappa_{ij}| \geq |\beta_{j+1,j} \prod_{k=j+1}^{i-1} \alpha_{k+1,k}| > 0$ proving our result. ■

We note that Lemma 3.7 is only relevant for $s = p = 1,2,3,4$[3]. In this section, we will of course be interested specifically in the case with $s = p = 4$.

In proving the main result of this section, we will make extensive use of the following lemma, which follows immediately from Lemma 3.7 and the definition of the κ_{ij}.

LEMMA 3.8. *If $s = p$, the β_{ik} at a particular level k, for some $0 \leq k \leq s-1$, $k+1 \leq i \leq s$, are all of the same sign, and the CFL coefficient is positive, then $\kappa_{ik}, k+1 \leq i \leq s$ are also all of that same sign and are nonzero.*

We now give the main result of this section.

THEOREM 3.9. *There is no four-stage explicit SSPRK method of order 4 with a positive CFL coefficient.*

Proof. **General Case.** We proceed by contradiction. If two parameters u and v are such that $u \neq v$, $u \neq 0$, $u \neq 1/2$, $u \neq 1$, $v \neq 0$, $v \neq 1$, and $6uv - 4(u+v) + 3 \neq 0$,

[3]It is possible to have schemes with $s = p > 4$ for linear, constant-coefficient problems.

then the coefficients $\kappa_{ik} \neq 0$ may be written as functions of u and v [18]:

$$
\begin{aligned}
\kappa_{10} &= u, \\
\kappa_{20} &= v - \kappa_{21}, \\
\kappa_{21} &= \frac{v(v-u)}{2u(1-2u)}, \\
\kappa_{30} &= 1 - \kappa_{31} - \kappa_{32}, \\
\kappa_{31} &= \frac{(1-u)[u+v-1-(2v-1)^2]}{2u(v-u)[6uv-4(u+v)+3]}, \\
\kappa_{32} &= \frac{(1-2u)(1-u)(1-v)}{v(v-u)[6uv-4(u+v)+3]}, \\
\kappa_{40} &= \frac{1}{2} + \frac{1-2(u+v)}{12uv}, \\
\kappa_{41} &= \frac{2v-1}{12u(v-u)(1-u)}, \\
\kappa_{42} &= \frac{1-2u}{12v(v-u)(1-v)}, \\
\kappa_{43} &= \frac{1}{2} + \frac{2(u+v)-3}{12(1-u)(1-v)}.
\end{aligned}
$$

Similar to [3], there are five possibilities to consider:

1. $u < 0$. If $v < 0$ then $\kappa_{10}\kappa_{40} < 0$. Conversely if $v > 0$ then $\kappa_{10}\kappa_{20} < 0$. Both results contradict Lemma 3.8.
2. $0 < u < \frac{1}{2}$ and $v < u$.
   $\kappa_{21}\kappa_{41} > 0$ implies that $v < 0$. But this implies $\kappa_{10}\kappa_{20} < 0$, contradicting Lemma 3.8.
3. $0 < u < \frac{1}{2}$ and $v > u$.
   $\kappa_{21}\kappa_{41} > 0$ requires $v > \frac{1}{2}$. $\kappa_{20} > 0$ requires $v < 3u - 4u^2 \leq \frac{9}{16}$. $\kappa_{32}\kappa_{42} > 0$ and $\kappa_{31}\kappa_{41} > 0$ require that $u > 2 - 5v + 4v^2$. Since this is a decreasing function of v for $v \leq \frac{9}{16}$, we obtain $u > 2 - 5(3u-4u^2) + 4(3u-4u^2)^2$. Rearranging, we find that $0 > 2((2u-1)^2+4u^2)(2u-1)^2$, which is impossible.
4. $u > \frac{1}{2}$ and $v < u$. We can only have $\kappa_{32}\kappa_{42} > 0$ in one of two ways:
   (a) $1 - u > 0$ and $6uv - 4(u+v) + 3 > 0$.
   $\kappa_{21}\kappa_{41} > 0$ requires $0 < v < \frac{1}{2}$. Simple calculation yields

$$\kappa_{30} = \frac{(2-6u+4u^2) + (-5+15u-12u^2)v + (4-12u+12u^2)v^2}{2uv(6uv-4(u+v)+3)};$$

   hence $\kappa_{30} > 0$ requires

$$A+Bv+Cv^2 \equiv (2-6u+4u^2)+(-5+15u-12u^2)v+(4-12u+12u^2)v^2 > 0.$$

   It is easy to show that when $\frac{1}{2} < u < 1$ we have $A < 0, B < 0$, and $C > 0$. Thus for $0 < v < \frac{1}{2}$ we have

$$A+Bv+Cv^2 < \max\left(A, A + \frac{1}{2}B + \frac{1}{4}C\right) = \max(A, \frac{1}{2}(1-2u)(1-u)) < 0,$$

   resulting in a contradiction.

(b) $1-u<0$ and $6uv-4(u+v)+3<0$.
Suppose $v<0$. Then $\kappa_{10}\kappa_{20}>0$ implies $v<-u(4u-3)<-u$ and $\kappa_{21}\kappa_{31}>0$ implies $u+v-1-(2v-1)^2>0$. Together these yield a contradiction.
Now suppose $v>0$. $\kappa_{21}\kappa_{41}>0$ implies $v>\frac{1}{2}$. $\kappa_{31}\kappa_{41}>0$ requires $u+v-1-(2v-1)^2<0$ which implies

$$(1-4v)(1-v)=4v^2-5v+1>u-1>0.$$

Given the restrictions on v, this is true only if $v<\frac{1}{4}$, contradicting the requirement that $v>\frac{1}{2}$.

5. $u>\frac{1}{2}$ and $v>u$. $\kappa_{21}\kappa_{41}>0$ requires that $u>1$. This implies $\kappa_{42}>0$; so by Lemma 3.8 $\kappa_{32}>0$. It is now easily seen that $\kappa_{10}>0$, $\kappa_{21}<0$, and $\kappa_{43}>0$. Thus $\kappa_{10}\kappa_{21}\kappa_{32}\kappa_{43}<0$, contradicting the fourth-order condition $\kappa_{10}\kappa_{21}\kappa_{32}\kappa_{43}=\frac{1}{4!}$.

If $6uv-4(u+v)+3=0$, $u=0$, or $v=0$, then this method is not fourth order [18].
Special Cases. There remain three special cases [18, 5], namely

1. $u=\frac{1}{2},\ v=0$; $\kappa_{42}=w\neq 0,\ \kappa_{40}=\frac{1}{6}-w,\ \kappa_{41}=\frac{2}{3}$, and $\kappa_{43}=\frac{1}{6}$.
2. $u=v=\frac{1}{2}$; $\kappa_{40}=\frac{1}{6},\ \kappa_{42}=w\neq 0,\ \kappa_{41}=\frac{2}{3}-w$, and $\kappa_{43}=\frac{1}{6}$.
3. $u=1,\ v=\frac{1}{2}$; $\kappa_{43}=w\neq 0,\ \kappa_{41}=\frac{1}{6}-w,\ \kappa_{40}=\frac{1}{6}$, and $\kappa_{42}=\frac{2}{3}$.

In these cases, κ_{32} is obtained from

$$\kappa_{43}\kappa_{32}=\kappa_{42}(1-v).$$

The remaining coefficients κ_{21} and κ_{31} are then the solutions to the (nonsingular) linear system

$$\begin{aligned}\kappa_{42}\kappa_{21}uv+\kappa_{43}(\kappa_{31}u+\kappa_{32}v)&=\frac{1}{8},\\ \kappa_{42}\kappa_{21}+\kappa_{43}\kappa_{31}&=\kappa_{41}(1-u).\end{aligned}$$

It is easily verified that in each case the κ_{ik} fail to have the same sign at each level whenever negative β_{ik} are considered. ■

4. Fifth-Order Explicit SSPRK Methods. In this section, we give the results of the numerical optimization procedure outlined in Section 3. Examples of optimal explicit SSPRK methods of order 4 and up to 8 stages with positive coefficients appear in [24]. We have also constructed optimal explicit SSPRK methods of order 3 and up to 9 stages. Here we design the first optimized fifth-order explicit SSPRK methods. No formal proofs of optimality are given; however the methods described here are the results of extensive numerical testing. We now give the coefficients of the Butcher tableaus for SSP(7,5), SSP(8,5), and SSP(9,5) in Tables 4.1–4.3 respectively. The Butcher tableau format is provided because this is the more advantageous format for implementation.

5. Numerical Studies. In this section, we study the numerical behaviour of our fifth-order schemes and the optimal known fifth-order SSP multistep method (1.3) for a few test problems designed to capture solution features that pose particular difficulties to numerical methods. Our focus here is to illustrate the stability behaviour of various fifth-order schemes rather than to provide detailed accuracy study. If a study of the relative error constants was desired it would be more appropriate to consider systems where the spatial discretization errors are dominated by the time stepping

Table 4.1
Butcher tableau entries for SSP(7,5). CFL coefficient is 1.178508348471858.

entry	value
$a(2,1)$	0.392382208054010
$a(3,1)$	0.310348765296963
$a(3,2)$	0.523846724909595
$a(4,1)$	0.114817342432177
$a(4,2)$	0.248293597111781
$a(4,3)$	0
$a(5,1)$	0.136041285050893
$a(5,2)$	0.163250087363657
$a(5,3)$	0
$a(5,4)$	0.557898557725281
$a(6,1)$	0.135252145083336
$a(6,2)$	0.207274083097540
$a(6,3)$	-0.180995372278096
$a(6,4)$	0.326486467604174
$a(6,5)$	0.348595427190109
$a(7,1)$	0.082675687408986
$a(7,2)$	0.146472328858960
$a(7,3)$	-0.160507707995237
$a(7,4)$	0.161924299217425
$a(7,5)$	0.028864227879979
$a(7,6)$	0.070259587451358
$b(1)$	0.110184169931401
$b(2)$	0.122082833871843
$b(3)$	-0.117309105328437
$b(4)$	0.169714358772186
$b(5)$	0.143346980044187
$b(6)$	0.348926696469455
$b(7)$	0.223054066239366

Table 4.2
Butcher tableau entries for SSP(8,5). CFL coefficient is 1.875684961641323.

entry	value
$a(2,1)$	0.276409720937984
$a(3,1)$	0.149896412080489
$a(3,2)$	0.289119929124728
$a(4,1)$	0.057048148321026
$a(4,2)$	0.110034365535150
$a(4,3)$	0.202903911101136
$a(5,1)$	0.169059298369086
$a(5,2)$	0.326081269617717
$a(5,3)$	0.450795162456598
$a(5,4)$	0
$a(6,1)$	0.061792381825461
$a(6,2)$	0.119185034557281
$a(6,3)$	0.199236908877949
$a(6,4)$	0.521072746262762
$a(6,5)$	-0.001094028365068
$a(7,1)$	0.111048724765050
$a(7,2)$	0.214190579933444
$a(7,3)$	0.116299126401843
$a(7,4)$	0.223170535417453
$a(7,5)$	-0.037093067908355
$a(7,6)$	0.228338214162494
$a(8,1)$	0.071096701602448
$a(8,2)$	0.137131189752988
$a(8,3)$	0.154859800527808
$a(8,4)$	0.043090968302309
$a(8,5)$	-0.163751550364691
$a(8,6)$	0.044088771531945
$a(8,7)$	0.102941265156393
$b(1)$	0.107263534301213
$b(2)$	0.148908166410810
$b(3)$	0.105268730914375
$b(4)$	0.124847526215373
$b(5)$	-0.068303238298102
$b(6)$	0.127738462988848
$b(7)$	0.298251879839231
$b(8)$	0.156024937628252

TABLE 4.3
Butcher tableau entries for SSP(9,5). CFL coefficient is 2.695788289294857.

entry	value
$a(2, 1)$	0.234806766829933
$a(3, 1)$	0.110753442788106
$a(3, 2)$	0.174968893063956
$a(4, 1)$	0.050146926953296
$a(4, 2)$	0.079222388746543
$a(4, 3)$	0.167958236726863
$a(5, 1)$	0.143763164125647
$a(5, 2)$	0.227117830897242
$a(5, 3)$	0.240798769812556
$a(5, 4)$	0
$a(6, 1)$	0.045536733856107
$a(6, 2)$	0.071939180543530
$a(6, 3)$	0.143881583463234
$a(6, 4)$	0.298694357327376
$a(6, 5)$	-0.013308014505658
$a(7, 1)$	0.058996301344129
$a(7, 2)$	0.093202678681501
$a(7, 3)$	0.109350748582257
$a(7, 4)$	0.227009258480886
$a(7, 5)$	-0.010114159945349
$a(7, 6)$	0.281923169534861
$a(8, 1)$	0.114111232336224
$a(8, 2)$	0.180273547308430
$a(8, 3)$	0.132484700103381
$a(8, 4)$	0.107410821979346
$a(8, 5)$	-0.129172321959971
$a(8, 6)$	0.133393675559324
$a(8, 7)$	0.175516798122502
$a(9, 1)$	0.096188287148324
$a(9, 2)$	0.151958780732981
$a(9, 3)$	0.111675915818310
$a(9, 4)$	0.090540280530361
$a(9, 5)$	-0.108883798219725
$a(9, 6)$	0.112442122530629
$a(9, 7)$	0.147949153045843
$a(9, 8)$	0.312685695043563

entry	value
$b(1)$	0.088934582057735
$b(2)$	0.102812792947845
$b(3)$	0.111137942621198
$b(4)$	0.158704526123705
$b(5)$	-0.060510182639384
$b(6)$	0.197095410661808
$b(7)$	0.071489672566698
$b(8)$	0.151091084299943
$b(9)$	0.179244171360452

error. Experiments using the standard implementation of Fehlberg's fifth-order explicit Runge-Kutta method [5] are also included because this method is commonly used in method-of-lines discretizations of hyperbolic conservation laws. We remark that Fehlberg's scheme does not have a positive CFL coefficient in its standard implementation (using only $F(\cdot)$) because SSP methods of order greater than four require evaluations of $\tilde{F}(\cdot)$ [20].

We remark that tests using the popular Dormand-Prince scheme [5] gave results very similar to Fehlberg's scheme. For clarity, we do not include these simulations in our plotted results.

5.1. Test Problems. There are a variety of solution features in computational fluid dynamics that commonly cause numerical problems. For example, many numerical methods produce significant errors near sonic points (points where the wavespeed equals zero). Upwind methods in particular are forced to give sonic points special consideration since the upwind direction changes at sonic points. Shock waves, contact discontinuities, and expansion fans may also lead to a variety of serious problems including oscillations, overshoots, and smearing that can spread discontinuities over several cells. In particular, contact discontinuities do not have any physical compression and thus smearing increases progressively with the number of time steps. Even when approximating smooth solutions, most numerical methods exhibit obvious flaws. For example, many stable numerical methods continuously erode the solution, leading to amplitude and dissipation errors [11].

To investigate the behavior of our time-stepping schemes, we consider three of

Laney's five test problems [11]. These three problems involve all of the important flow features identified above: shocks, contacts, expansion fans, sonic points, and smooth solutions. Similar to Laney, we focus on the behaviour of the numerical scheme for interior regions rather than boundaries and impose periodic boundary conditions on the domain $[-1, 1]$. It is known that sometimes a conventional (and intuitive) treatment of the boundary data (especially in the case of inflow boundary conditions) within the stages of a Runge-Kutta method can lead to a deterioration in the overall accuracy of the integration. We refer to [1] and references therein for a discussion of this problem and a method for its resolution. The spatial discretization and the results of three test cases follow.

5.2. Spatial Discretization. Similar to [22, 25], we choose finite-difference Shu-Osher methods (ENO) to spatially discretize the equations. These methods are derived using flux reconstruction and have a variety of desirable properties. For example, they naturally extend to an arbitrary order of accuracy in space, and they are independent of the time discretization, thus allowing experimentation with different time discretization methods. Moreover, educational codes are also freely available [11, 10], an attribute which is desirable for standardizing numerical studies. Since we are focusing on fifth-order Runge-Kutta methods we carry out our simulations using a fifth-order spatial discretization. We further note that flux splitting is carried out according to

$$\begin{aligned} f^{+}(U) &= \frac{1}{2}(f(U) + \alpha^n_{i+1/2}U), \\ f^{-}(U) &= \frac{1}{2}(f(U) - \alpha^n_{i+1/2}U) \end{aligned}$$

where $\alpha^n_{i+1/2} = \max\{|f'(U^n_{i+1})|, |f'(U^n_i)|\}$. To evaluate $\tilde{F}(\cdot)$ we simply negate the discretization that arises when we apply the Shu-Osher finite difference method to the PDE evolved *backwards in time*[4] (see [21] for further details on the procedure). For further details on the underlying discretization as well as code for the spatial discretization, see [11, 10].

It is noteworthy that high-order, fully TVD spatial discretization schemes are also available; see Osher and Chakravarthy [16]. In these numerical studies, we choose Shu-Osher spatial discretization schemes rather than TVD schemes because TVD schemes only obtain between first- and second-order accuracy at extrema and they have "been largely superseded by Shu and Osher's class of high-order ENO methods" [11].

It is also noteworthy that recent variations on Shu-Osher methods such as methods based on WENO reconstructions (e.g., [14, 6]) also naturally combine with SSPRK schemes. See [11] for detailed discussions on these and other spatial discretizations appropriate for hyperbolic conservation laws.

[4]To illustrate the procedure, consider a first-order spatial discretization of $\tilde{F}$ for Burgers' equation. To proceed we need to construct an upwind spatial discretization for Burgers' equation evolved backwards in time, i.e.,

$$u_t = uu_x.$$

Carrying out a first-order upwind discretization with a uniform discretization step size h gives $-\tilde{F}$:

$$-\tilde{F} = \begin{cases} U_j(U_{j+1} - U_j)/h & \text{if } U_j > 0 \\ U_j(U_j - U_{j-1})/h & \text{otherwise} \end{cases}$$

from which $\tilde{F}$ is trivially obtained.

5.3. Test Case 1: Linear advection of a sinusoid. In this test case, the smooth initial conditions

$$u(x,0) = -\sin(\pi x)$$

are evolved to time $t = 30$ according to the linear advection equation

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0$$

using a constant grid spacing of $\Delta x = 1/80$. Because this evolution causes the initial conditions to travel around the periodic domain $[-1,1]$ exactly 15 times, it is clear that the exact solution is just $u(x,30) = -\sin(\pi x)$. Test Case 1 effectively illustrates the evolution of a smooth solution with no sonic points and is useful for verifying convergence rates for high-order schemes. Moreover, even on completely smooth solutions most numerical methods designed for hyperbolic conservation laws exhibit obvious flaws [11]. This test case is quite helpful for understanding phase and amplitude errors but should not be used to study dispersion because only one frequency is present in the exact solution. It is also informative to contrast these results with those derived for problems involving shocks and other discontinuities.

To quantify the accuracy of the computed solution, we use the logarithm of the l_1 errors, i.e.,

$$\log_{10}\left(\frac{1}{N}\sum_{i=1}^{N}|U_i - u(x_i,30)|\right),$$

where N is the number of grid points and x_i is the i^{th} grid node. A plot of the error is given in Figure 5.1. To ensure a fair comparison for methods with a different number of stages, the error is plotted as a function of the effective CFL number[5] rather than the CFL number itself. This implies that for a particular plot, the total number of function evaluations at a particular abscissa value will be the same for each scheme. We start calculating errors for an effective CFL number of 0.02 and continue until the numerical method is so unstable that a value of `NaN` is returned; i.e., the scheme has become completely unstable.

In this test example, the main conclusion is that Fehlberg's scheme and our new fifth-order explicit SSPRK schemes all outperform the multistep scheme (1.3) by more than 350%, with SSP(9,5) giving more than a 400% improvement. It is not surprising that Fehlberg's scheme performs well on this *smooth* problem because schemes based purely on a linear stability analysis are expected to perform well. SSP schemes are designed to outperform on problems involving discontinuities in the solution or its derivatives, so in this case there is no reason to expect that schemes derived using a nonlinear stability analysis will necessarily outperform classical schemes based on a linear stability analysis.

5.4. Test Case 2: Linear advection of a square wave. In this test case, the discontinuous initial conditions

$$u(x,0) = \begin{cases} 1 & \text{for } |x| < 1/3, \\ 0 & \text{for } 1/3 < |x| \le 1, \end{cases}$$

[5]Similar to the definition of effective CFL coefficient, the *effective CFL number* of an SSPRK method is $\left(\frac{1}{s}\right)\frac{\Delta t}{\Delta x}$, where s is the number of stages required for one step of the method.

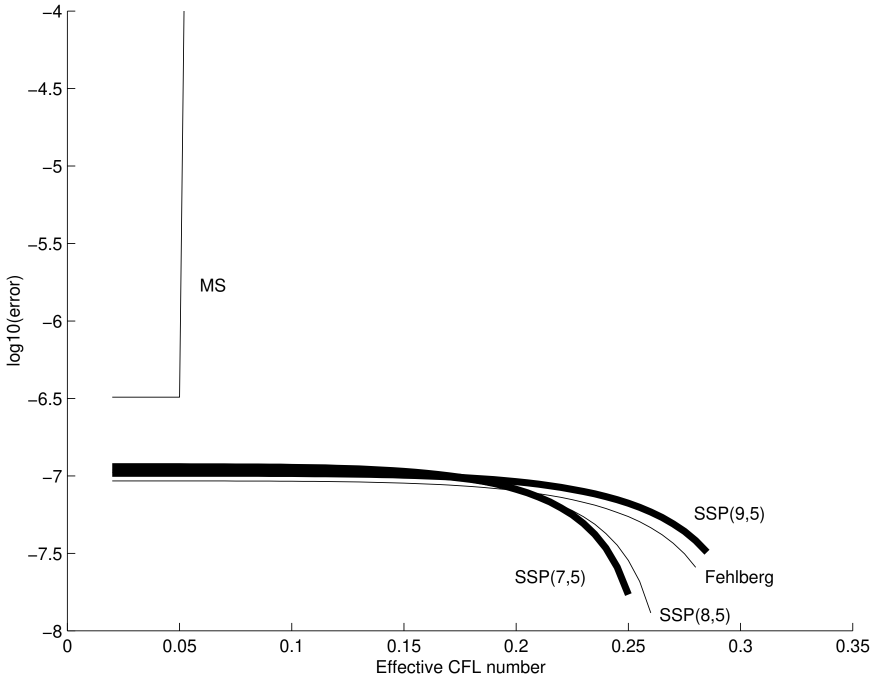

FIG. 5.1. *l_1 errors as a function of the effective CFL number for Test Case 1.*

are evolved to time $t = 4$ according to the linear advection equation

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0$$

using a constant grid spacing of $\Delta x = 1/320$. Because this evolution causes the initial conditions to travel around the periodic domain $[-1, 1]$ exactly 2 times, it is clear that the exact solution at the final time is just $u(x, 4) = u(x, 0)$. Test Case 2 exhibits two jump discontinuities in the solution that correspond to contact discontinuities. This test case nicely illustrates progressive contact smearing and dispersion.

The log of the l_1 errors as a function of the effective CFL number are plotted in Figure 5.2. Based on these plots, it is immediately clear that a material improvement in stability is obtained using our new fifth-order SSPRK schemes. Indeed, our schemes all outperform the multistep scheme (1.3) by 200% or more, with SSP(9,5) giving a 340% improvement. We also find that our schemes significantly outperform Fehlberg's scheme on this nonsmooth test. In particular, SSP(9,5) gives a 40% improvement over Fehlberg's scheme.

5.5. Test Case 3: Evolution of a square wave by Burgers' equation. In this test case, the discontinuous initial conditions

$$u(x, 0) = \begin{cases} 1 & \text{for } |x| < 1/3, \\ -1 & \text{for } 1/3 < |x| \le 1, \end{cases}$$

are evolved to time $t = 0.3$ according to Burgers' equation

$$\frac{\partial u}{\partial t} + \frac{\partial}{\partial x}\left(\frac{1}{2}u^2\right) = 0$$

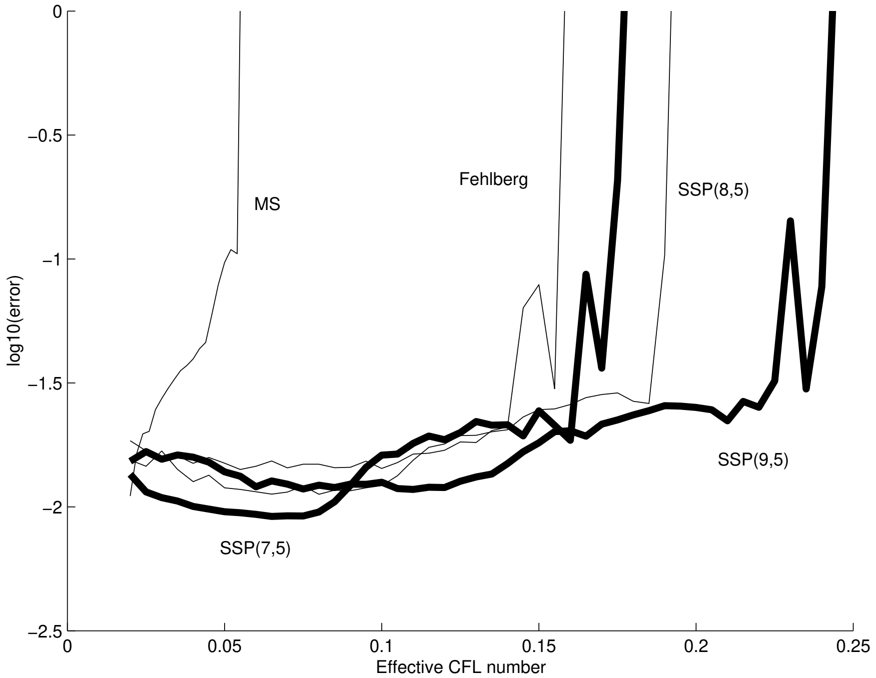

FIG. 5.2. *l_1 errors as a function of the effective CFL number for Test Case 2.*

using a constant grid spacing of $\Delta x = 1/320$. In this example, the jump at $x = -1/3$ creates a simple centered expansion fan and the jump at $x = 1/3$ creates a steady shock. Until the shock and expansion fan intersect (at time $t = 2/3$), the exact solution is

$$u(x,t) = \begin{cases} -1 & \text{for } -\infty < x < b_1, \\ -1 + 2\frac{x-b_1}{b_2-b_1} & \text{for } b_1 < x < b_2, \\ 1 & \text{for } b_2 < x < b_{shock}, \\ -1 & \text{for } b_{shock} < x < \infty, \end{cases}$$

where $b_1 = -1/3 - t$, $b_2 = -1/3 + t$ and $b_{shock} = 1/3$ [11]. Test Case 3 is particularly interesting because it illustrates the behaviors near sonic points ($u = 0$) that correspond to an expansion fan and a compressive shock.

The log of the l_1 errors as a function of the effective CFL number are plotted in Figure 5.3. In this nonlinear test case, we find a dramatic improvement for our new schemes over the multistep scheme (1.3). They all give more than a 350% improvement, with SSP(9,5) giving more than a 575% improvement. We also find that our schemes significantly outperform Fehlberg's scheme on this nonsmooth test. The SSP(9,5) scheme, in particular, gives more than a 150% improvement over Fehlberg's scheme.

6. Conclusions. We have studied high-order strong-stability-preserving explicit Runge-Kutta methods with downwind-biased spatial discretizations. We find that by requiring that the non-zero coefficients β_{ik} for a given k are all of the same sign we obtain a more appropriate description of what should be optimized. This leads to more efficient schemes with less smearing. When the order of the explicit Runge-Kutta

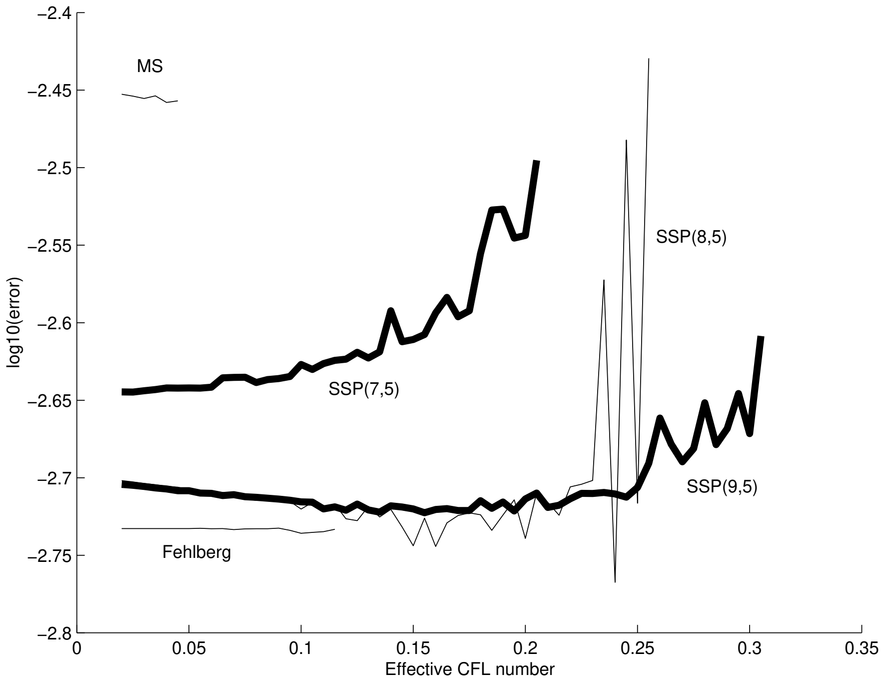

FIG. 5.3. l_1 *errors as a function of the effective CFL number for Test Case 3.*

method is less than or equal to four we prove in a variety of cases that there is no advantage in terms of effective CFL coefficient to using downwind-biased spatial discretizations. To achieve explicit SSPRK methods with fifth- or higher-order accuracy, however, downwind-biased discretizations are necessary. This paper provides the first examples of such schemes. We find that these new schemes are much more efficient than existing fifth-order explicit SSP multistep methods (both theoretically and in practice) and handily outperform classical explicit fifth-order schemes on nonsmooth problems. In particular, we found that in our marginally resolved test cases (involving shocks and contact discontinuities) larger time steps and improved efficiency were found as the effective CFL coefficient (and the number of stages) increased. In a well-resolved problem (Test Case 1), however, the practical performance of SSPRK schemes and classical Runge-Kutta schemes was very similar. This suggests that high-order SSPRK schemes with large effective CFL coefficients have the potential to provide high-order accuracy in smooth regions of the flow while still yielding large stable steps in marginally resolved regions. It is our hope that by providing numerically optimal schemes of this type we will stimulate further numerical studies and comparisons of SSPRK schemes against more classical approaches.

Acknowledgements. The authors thank J. Rusak for his help with the unconstrained global optimization. We also thank S. Gottlieb for interesting discussions on downwind-biased spatial discretizations.

REFERENCES

[1] S. ABARBANEL, D. GOTTLIEB, AND M. H. CARPENTER, *On the removal of boundary errors caused by Runge-Kutta integration of nonlinear partial differential equations*, SIAM J. Sci.

Comput., 17 (1996), pp. 777–782.

[2] L. Ferracina and M. Spijker, *Stepsize restrictions for the total-variation-diminishing property in general Runge-Kutta methods*, Technical Report MI 2002-21, University of Leiden, August 2002.

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

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

[5] E. Hairer, S. Norsett, and G. Wanner, *Solving Ordinary Differential Equations I*, Springer-Verlag, 1987.

[6] G.-S. Jiang and C. Shu, *Efficient implementation of weighted ENO schemes*, J. Comput. Phys., 126 (1996), pp. 202–228.

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

[8] ———, *Contractivity of Runge-Kutta methods*, BIT, 31 (1991), pp. 482–528.

[9] J. D. Lambert, *Computational methods in ordinary differential equations*, John Wiley & Sons, London-New York-Sydney, 1973. Introductory Mathematics for Scientists and Engineers.

[10] C. Laney, *CFD Recipes: Software for Computational Gasdynamics*. Web Address: http://capella.colorado.edu/∼laney/booksoft.htm.

[11] ———, *Computational Gasdynamics*, Cambridge University Press, 1998.

[12] J. Lawson, *An order five Runge-Kutta process with extended region of stability*, SIAM J. Numer. Anal., 3 (1966), pp. 593–597.

[13] H. W. J. Lenferink, *Contractivity preserving explicit linear multistep methods*, Numer. Math., 55 (1989), pp. 213–223.

[14] X.-D. Liu, S. Osher, and T. Chan, *Weighted essentially nonoscillatory schemes*, J. Comput. Phys., 115 (1994), pp. 200–212.

[15] A. A. Medovikov, *High order explicit methods for parabolic equations*, BIT, 38 (1998), pp. 372–390.

[16] S. Osher and S. Chakravarthy, *Very high order accurate TVD schemes*, in Oscillation Theory, Computation, and methods of Compensated Compactness. The IMA Volumes in Mathematics and Its Applications, C. Dafermos, J. Erikson, D. Kinderlehrer, and M. Slemrod, eds., vol. 2, Springer-Verlag, New York, 1986, pp. 229–271.

[17] S. Osher and R. Fedkiw, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, 2002.

[18] A. Ralston, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965.

[19] S. Ruuth, *Global optimization of strong-stability-preserving Runge-Kutta schemes*, unpublished manuscript (under review), 2003.

[20] S. Ruuth and R. Spiteri, *Two barriers on strong-stability-preserving time discretization methods*, J. Scientific Computation, 17 (2002), pp. 211–220.

[21] C.-W. Shu, *Total-variation-diminishing time discretizations*, SIAM J. Sci. Statist. Comput., 9 (1988), pp. 1073–1084.

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

[23] M. Spijker, *Contractivity in the numerical solution of initial value problems*, Numer. Math., 42 (1983), pp. 271–290.

[24] R. J. Spiteri and S. J. Ruuth, *Nonlinear evolution using optimal fourth-order strong-stability-preserving Runge-Kutta methods*, Mathematics and Computers in Simulation. Special issue on "Nonlinear Waves: Computation and Theory II" (to appear).

[25] ———, *A new class of optimal high-order strong-stability-preserving time-stepping schemes*, SIAM J. Numer. Anal., 40 (2002), pp. 469–491.

[26] M. Tawarmalani and N. V. Sahinidis, *Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications*, vol. 65 of Nonconvex Optimization and Its Applications, Kluwer Academic Publishers, Dordrecht, 2002.