



A New Class of A Stable Summation by Parts Time Integration Schemes with Strong Initial Conditions

Hendrik Ranocha^{1,2} · Jan Nordström^{3,4}

Received: 3 June 2020 / Revised: 23 December 2020 / Accepted: 26 February 2021
© The Author(s) 2021

Abstract

Since integration by parts is an important tool when deriving energy or entropy estimates for differential equations, one may conjecture that some form of summation by parts (SBP) property is involved in provably stable numerical methods. This article contributes to this topic by proposing a novel class of A stable SBP time integration methods which can also be reformulated as implicit Runge–Kutta methods. In contrast to existing SBP time integration methods using simultaneous approximation terms to impose the initial condition weakly, the new schemes use a projection method to impose the initial condition strongly without destroying the SBP property. The new class of methods includes the classical Lobatto IIIA collocation method, not previously formulated as an SBP scheme. Additionally, a related SBP scheme including the classical Lobatto IIIB collocation method is developed.

Keywords Summation by parts · Projection method · Runge–Kutta methods · Time integration schemes · Energy stability · A stability

Mathematics Subject Classification 65L06 · 65L20 · 65N06 · 65M06 · 65M12 · 65N35 · 65M70

✉ Jan Nordström
jan.nordstrom@liu.se

Hendrik Ranocha
mail@ranocha.de

¹ Computer Electrical and Mathematical Science and Engineering Division (CEMSE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia

² Present Address: Mathematics Münster, University of Münster, Münster, Germany

³ Department of Mathematics, Applied Mathematics, Linköping University, SE-58183 Linköping, Sweden

⁴ Department of Mathematics and Applied Mathematics, University of Johannesburg, P.O. Box 524, Johannesburg, Auckland Park 2006, South Africa

1 Introduction

Based on the fact that integration by parts plays a major role in the development of energy and entropy estimates for initial boundary value problems, one may conjecture that the summation by parts (SBP) property [8,47] is a key factor in provably stable schemes. Although it is complicated to formulate such a conjecture mathematically, there are several attempts to unify stable methods in the framework of summation by parts schemes, starting from the origin of SBP operators in finite difference methods [16,43] and ranging from finite volume [23,24] and discontinuous Galerkin methods [9] to flux reconstruction schemes [37].

Turning to SBP methods in time [2,19,25], a class of linearly and nonlinearly stable SBP schemes has been constructed and studied in this context, see also [18,41,42]. If the underlying quadrature is chosen as Radau or Lobatto quadrature, these Runge-Kutta schemes are exactly the classical Radau IA, Radau IIA, and Lobatto IIIA methods [29]. Having the conjecture “stability results require an SBP structure” in mind, this article provides additional insights to this topic by constructing new classes of SBP schemes, which reduce to the classical Lobatto IIIA and Lobatto IIIB methods if that quadrature rule is used. Consequently, all A stable classical Runge-Kutta methods based on Radau and Lobatto quadrature rules can be formulated in the framework of SBP operators. Notably, instead of using simultaneous approximation terms (SATs) [5,6] to impose initial conditions weakly, these new schemes use a strong imposition of initial conditions in combination with a projection method [22,26,27].

By mimicking integration by parts at a discrete level, the stability of SBP methods can be obtained in a straightforward way by mimicking the continuous analysis. All known SBP time integration methods are implicit and their stability does not depend on the size of the time step. In contrast, the stability analysis of explicit time integration methods can use techniques similar to summation by parts, but the analysis is in general more complicated and restricted to sufficiently small time steps [36,44,46]. Since there are strict stability limitations for explicit methods, especially for nonlinear problems [30,31], an alternative to stable fully implicit methods is to modify less expensive (explicit or not fully implicit) time integration schemes to get the desired stability results [10,15,32,33,39,45].

This article is structured as follows. At first, the existing class of SBP time integration methods is introduced in Sect. 2, including a description of the related stability properties. Thereafter, the novel SBP time integration methods are proposed in Section 3. Their stability properties are studied and the relation to Runge-Kutta methods is described. In particular, the Lobatto IIIA and Lobatto IIIB methods are shown to be recovered using this framework. Afterwards, results of numerical experiments demonstrating the established stability properties are reported in Sect. 4. Finally, the findings of this article are summed up and discussed in Sect. 5.

2 Known Results for SBP Schemes

Consider an ordinary differential equation (ODE)

$$\forall t \in (0, T): \quad u'(t) = f(t, u(t)), \quad u(0) = u_0, \quad (1)$$

with solution u in a Hilbert space. Summation by parts schemes approximate the solution on a finite grid $0 \leq \tau_1 < \dots < \tau_s \leq T$ pointwise as $\mathbf{u}_i = u(\tau_i)$ and $\mathbf{f}_i = f(\tau_i, \mathbf{u}_i)$. Although the grid does not need to be ordered for general SBP schemes, we impose this restriction to simplify the presentation. (An unordered grid can always be transformed into an ordered

one by a permutation of the grid indices.) The SBP operators can be defined as follows, cf. [7,8,47].

Definition 2.1 A first derivative SBP operator of order p on $[0, T]$ consists of

- a discrete operator D approximating the derivative $D\mathbf{u} \approx u'$ with order of accuracy p ,
- a symmetric and positive definite discrete quadrature matrix M approximating the L^2 scalar product $\mathbf{u}^T M \mathbf{v} \approx \int_0^T u(\tau)v(\tau)d\tau$,
- and interpolation vectors $\mathbf{t}_L, \mathbf{t}_R$ approximating the boundary values as $\mathbf{t}_L^T \mathbf{u} \approx u(0)$, $\mathbf{t}_R^T \mathbf{u} \approx u(T)$ with order of accuracy at least p , such that the SBP property

$$MD + (MD)^T = \mathbf{t}_R \mathbf{t}_R^T - \mathbf{t}_L \mathbf{t}_L^T \tag{2}$$

holds.

Remark 2.2 There are analogous definitions of SBP operators for second or higher order derivatives [20,21,34]. In this article, only first derivative SBP operators are considered.

Remark 2.3 The quadrature matrix M is sometimes called norm matrix (since it induces a norm via a scalar product) or mass matrix (in a finite element context).

Because of the SBP property (2), SBP operators mimic integration by parts discretely via

$$\underbrace{\mathbf{u}^T M(D\mathbf{v}) + (D\mathbf{u})^T M \mathbf{v}}_{\approx} = \underbrace{(\mathbf{t}_R^T \mathbf{u})^T (\mathbf{t}_R^T \mathbf{v}) - (\mathbf{t}_L^T \mathbf{u})^T (\mathbf{t}_L^T \mathbf{v})}_{\approx} \tag{3}$$

$$\int_0^T u(\tau)v'(\tau)d\tau + \int_0^T u'(\tau)v(\tau)d\tau = \underbrace{u(T)v(T) - u(0)v(0)}.$$

However, this mimetic property does not suffice for the derivations to follow. Nullspace consistency will be used as an additional required mimetic property. This novel property was introduced in [48] and has been a key factor in [18,38].

Definition 2.4 A first derivative SBP operator D is nullspace consistent, if the nullspace (kernel) of D satisfies $\ker D = \text{span}\{\mathbf{1}\}$.

Here, $\mathbf{1}$ denotes the discrete grid function with value unity at every node.

Remark 2.5 Every first derivative operator D (which is at least first order accurate) maps constants to zero, i.e. $D\mathbf{1} = \mathbf{0}$. Hence, the kernel of D always satisfies $\text{span}\{\mathbf{1}\} \leq \ker D$. Here and in the following, \leq denotes the subspace relation of vector spaces. If D is not nullspace consistent, there are more discrete grid functions besides constants which are mapped to zero (which makes it inconsistent with ∂_t). Then, $\ker D \neq \text{span}\{\mathbf{1}\}$ and undesired behavior can occur, cf. [18,29,48,49].

An SBP time discretization of (1) using SATs to impose the initial condition weakly is [2,19,25]

$$D\mathbf{u} = \mathbf{f} + M^{-1} \mathbf{t}_L (u_0 - \mathbf{t}_L^T \mathbf{u}). \tag{4}$$

The numerical solution u_+ at $t = T$ is given by $u_+ = \mathbf{t}_R^T \mathbf{u}$, where \mathbf{u} solves (4).

Remark 2.6 The interval $[0, T]$ can be partitioned into multiple subintervals/blocks such that multiple steps of this procedure can be used sequentially [19].

In order to guarantee that (4) can be solved for a linear scalar problem, $D + \sigma M^{-1} \mathbf{t}_L \mathbf{t}_L^T$ must be invertible, where σ is a real parameter usually chosen as $\sigma = 1$. The following result has been obtained in [18, Lemma 2].

Theorem 2.7 *If D is a first derivative SBP operator, $D + M^{-1} \mathbf{t}_L \mathbf{t}_L^T$ is invertible if and only if D is nullspace consistent.*

Remark 2.8 In [41,42], it was explicitly shown how to prove that $D + M^{-1} \mathbf{t}_L \mathbf{t}_L^T$ is invertible in the pseudospectral/polynomial and finite difference case.

As many other one-step time integration schemes, SBP-SAT schemes (4) can be characterized as Runge-Kutta methods, given by their Butcher coefficients [4,11]

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}, \tag{5}$$

where $A \in \mathbb{R}^{s \times s}$ and $b, c \in \mathbb{R}^s$. For (1), a step from u_0 to $u_+ \approx u(\Delta t)$ is given by

$$u_i = u_0 + \Delta t \sum_{j=1}^s a_{ij} f(c_j \Delta t, u_j), \quad u_+ = u_0 + \Delta t \sum_{i=1}^s b_i f(c_i \Delta t, u_i). \tag{6}$$

Here, u_i are the stage values of the Runge-Kutta method. The following characterization of (4) as Runge-Kutta method was given in [2].

Theorem 2.9 *Consider a first derivative SBP operator D . If $D + M^{-1} \mathbf{t}_L \mathbf{t}_L^T$ is invertible, (4) is equivalent to an implicit Runge-Kutta method with the Butcher coefficients*

$$\begin{aligned} A &= \frac{1}{T} (D + M^{-1} \mathbf{t}_L \mathbf{t}_L^T)^{-1} = \frac{1}{T} (MD + \mathbf{t}_L \mathbf{t}_L^T)^{-1} M, \\ b &= \frac{1}{T} M \mathbf{1}, \quad c = \frac{1}{T} (\tau_1, \dots, \tau_s)^T. \end{aligned} \tag{7}$$

The factor $\frac{1}{T}$ is needed since the Butcher coefficients of a Runge-Kutta method are normalized to the interval $[0, 1]$.

Next, we recall some classical stability properties of Runge-Kutta methods for linear problems, cf. [12, Section IV.3]. The absolute value of solutions of the scalar linear ODE

$$u'(t) = \lambda u(t), \quad u(0) = u_0 \in \mathbb{C}, \quad \lambda \in \mathbb{C}, \tag{8}$$

cannot increase if $\text{Re } \lambda \leq 0$. The numerical solution after one time step of a Runge-Kutta method with Butcher coefficients A, b, c is $u_+ = R(\lambda \Delta t) u_0$, where

$$R(z) = 1 + z b^T (I - z A)^{-1} \mathbf{1} = \frac{\det(I - z A + z \mathbf{1} b^T)}{\det(I - z A)} \tag{9}$$

is the *stability function* of the Runge-Kutta method. The stability property of the ODE is mimicked discretely as $|u_+| \leq |u_0|$ if $|R(\lambda \Delta t)| \leq 1$.

Definition 2.10 A Runge-Kutta method with stability function $|R(z)| \leq 1$ for all $z \in \mathbb{C}$ with $\text{Re}(z) \leq 0$ is *A stable*. The method is *L stable*, if it is *A stable* and $\lim_{z \rightarrow \infty} R(z) = 0$.

Hence, *A stable* methods are stable for every time step $\Delta t > 0$ and *L stable* methods damp out stiff components as $|\lambda| \rightarrow \infty$.

The following stability properties have been obtained in [2,19].

Theorem 2.11 Consider a first derivative SBP operator D . If $D + M^{-1}\mathbf{t}_L\mathbf{t}_L^T$ is invertible, then the SBP-SAT scheme (4) is both A and L stable.

Corollary 2.12 The SBP-SAT scheme (4) is both A and L stable if D is a nullspace consistent SBP operator.

Proof This result follows immediately from Theorem 2.7 and Theorem 2.11. □

3 The New Schemes

The idea behind the novel SBP time integration scheme introduced in the following is to mimic the reformulation of the ODE (1) as an integral equation

$$u(t) = u_0 + \int_0^t f(\tau, u(\tau))d\tau. \tag{10}$$

Taking the time derivative on both sides yields $u'(t) = f(t, u(t))$. The initial condition $u(0) = u_0$ is satisfied because $\int_0^0 f(\tau, u(\tau))d\tau = 0$. Hence, the solution u of (1) can be written implicitly as the solution of the integral equation (10). Note that the integral operator $\int_0^t \cdot d\tau$ is the inverse of the derivative operator $\frac{d}{dt}$ with a vanishing initial condition at $t = 0$. Hence, a discrete inverse (an integral operator) of the discrete derivative operator D with a vanishing initial condition will be our target.

Definition 3.1 In the space of discrete grid functions, the scalar product induced by M is used throughout this article. The adjoint operators with respect to this scalar product will be denoted by \cdot^* , i.e. $D^* = M^{-1}D^T M$. The adjoint of a discrete grid function \mathbf{u} is denoted by $\mathbf{u}^* = \mathbf{u}^T M$.

By definition, the adjoint operator D^* of D satisfies

$$\langle \mathbf{u}, D^*\mathbf{v} \rangle_M = \mathbf{u}^T M D^* \mathbf{v} = \mathbf{u}^T D^T M \mathbf{v} = (D\mathbf{u})^T M \mathbf{v} = \langle D\mathbf{u}, \mathbf{v} \rangle \tag{11}$$

for all grid functions \mathbf{u}, \mathbf{v} . The adjoint \mathbf{u}^* is a discrete representation of the inverse Riesz map applied to a grid function \mathbf{u} [40, Theorem 9.18] and satisfies

$$\mathbf{u}^* \mathbf{v} = \mathbf{u}^T M \mathbf{v} = \langle \mathbf{u}, \mathbf{v} \rangle_M. \tag{12}$$

The following lemma and definition were introduced in [38].

Lemma 3.2 For a nullspace consistent first derivative SBP operator D , $\dim \ker D^* = 1$.

Definition 3.3 A fixed but arbitrarily chosen basis vector of $\ker D^*$ for a nullspace consistent SBP operator D is denoted as \mathbf{o} .

The name \mathbf{o} is intended to remind the reader of (grid) oscillations, since the kernel of D^* is orthogonal to the image of D [40, Theorem 10.3] which contains all sufficiently resolved functions. Several examples are given in [38]. To prove $\ker D^* \perp \text{im } D$, choose any $D\mathbf{u} \in \text{im } D$ and $\mathbf{v} \in \ker D^*$ and compute

$$\langle D\mathbf{u}, \mathbf{v} \rangle_M = \langle \mathbf{u}, D^*\mathbf{v} \rangle_M = \langle \mathbf{u}, \mathbf{0} \rangle_M = 0. \tag{13}$$

Example 3.4 Consider the SBP operator of order $p = 1$ defined by the $p + 1 = 2$ Lobatto-Legendre nodes $\tau_1 = 0$ and $\tau_2 = T$ in $[0, T]$. Then,

$$D = \frac{1}{T} \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}, \quad M = \frac{T}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \mathbf{t}_L = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad \mathbf{t}_R = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{14}$$

Therefore,

$$D^* = M^{-1}D^T M = \frac{1}{T} \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix} \tag{15}$$

and $\ker D^* = \text{span}\{\mathbf{o}\}$, where $\mathbf{o} = (-1, 1)^T$. Here, \mathbf{o} represents the highest resolvable grid oscillation on $[\tau_1, \tau_2]$ and \mathbf{o} is orthogonal to $\text{im } D$, since $\mathbf{o}^T M D = \mathbf{0}^T$.

The following technique has been used in [38] to analyze properties of SBP operators in space. Here, it will be used to create new SBP schemes in time. Consider a nullspace consistent first derivative SBP operator D on the interval $[0, T]$ using s grid points and the corresponding subspaces

$$V_0 = \left\{ \mathbf{u} \in \mathbb{R}^s \mid \mathbf{u}(t = 0) = \mathbf{t}_L^T \mathbf{u} = 0 \right\}, \quad V_1 = \left\{ \mathbf{u} \in \mathbb{R}^s \mid \exists \mathbf{v} \in \mathbb{R}^s : \mathbf{u} = D\mathbf{v} \right\}. \tag{16}$$

Here and in the following, $\mathbf{u}(t = 0)$ denotes the value of the discrete function \mathbf{u} at the initial time $t = 0$. For example, $\mathbf{u}(t = 0) = \mathbf{t}_L^T \mathbf{u} = \mathbf{u}^{(1)}$ is the first coefficient of \mathbf{u} if $\tau_1 = 0$ and $\mathbf{t}_L = (1, 0, \dots, 0)^T$.

V_0 is the vector space of all grid functions which vanish at the left boundary point, i.e. $V_0 = \ker \mathbf{t}_L^T$. V_1 is the vector space of all grid functions which can be represented as derivatives of other grid functions, i.e. $V_1 = \text{im } D$ is the image of D .

Remark 3.5 From this point in the paper, D denotes a nullspace consistent first derivative SBP operator.

Lemma 3.6 *The mapping $D: V_0 \rightarrow V_1$ is bijective, i.e. one-to-one and onto, and hence invertible.*

Proof Given $\mathbf{u} \in V_1$, there is a $\mathbf{v} \in \mathbb{R}^s$ such that $\mathbf{u} = D\mathbf{v}$. Hence,

$$D(\mathbf{v} - (\mathbf{t}_L^T \mathbf{v})\mathbf{1}) = D\mathbf{v} - (\mathbf{t}_L^T \mathbf{v})D\mathbf{1} = D\mathbf{v} = \mathbf{u} \tag{17}$$

and

$$\mathbf{t}_L^T (\mathbf{v} - (\mathbf{t}_L^T \mathbf{v})\mathbf{1}) = \mathbf{t}_L^T \mathbf{v} - \mathbf{t}_L^T \mathbf{v} = 0, \tag{18}$$

since $\mathbf{t}_L^T \mathbf{v}$ is a scalar. Hence, (18) implies that $\mathbf{v} - (\mathbf{t}_L^T \mathbf{v})\mathbf{1} \in V_0$. Moreover, (17) shows that an arbitrary $\mathbf{u} \in V_1$ can be written as the image of a vector in V_0 under D . Therefore, $D: V_0 \rightarrow V_1$ is surjective (i.e. onto).

To prove that D is injective (i.e. one-to-one), consider an arbitrary $\mathbf{u} \in V_1$ and assume there are $\mathbf{v}, \mathbf{w} \in V_0$ such that $D\mathbf{v} = \mathbf{u} = D\mathbf{w}$. Then, $D(\mathbf{v} - \mathbf{w}) = \mathbf{0}$. Because of nullspace consistency, $\mathbf{v} - \mathbf{w} = \alpha\mathbf{1}$ for a scalar α . Since $\mathbf{v}, \mathbf{w} \in V_0 = \ker \mathbf{t}_L^T$,

$$0 = \mathbf{t}_L^T (\mathbf{v} - \mathbf{w}) = \alpha \mathbf{t}_L^T \mathbf{1} = \alpha. \tag{19}$$

Thus, $\mathbf{v} = \mathbf{w}$. □

Remark 3.7 V_0 is isomorphic to the quotient space $\mathbb{R}^s / \ker D$, since D is nullspace consistent. Hence, Lemma 3.6 basically states that D is a bijective mapping from $V_0 \cong \mathbb{R}^s / \ker D$ to $V_1 = \text{im } D$.

Definition 3.8 The inverse operator of $D: V_0 \rightarrow V_1$ is denoted as $J: V_1 \rightarrow V_0$.

The inverse operator J is a discrete integral operator such that $J\mathbf{v} \approx \int_0^t v(\tau)d\tau$. In general, there is a one-parameter family of integral operators given by $\int_{t_0}^t v(\tau)d\tau$. Here, we chose the one with $t_0 = 0$ to be consistent with (10).

Example 3.9 Continuing Example 3.4, the vector spaces V_0 and V_1 are

$$V_0 = \left\{ \mathbf{u} = \alpha \begin{pmatrix} 0 \\ 1 \end{pmatrix} \mid \alpha \in \mathbb{R} \right\}, \quad V_1 = \left\{ \mathbf{u} = \beta \begin{pmatrix} 1 \\ 1 \end{pmatrix} \mid \beta \in \mathbb{R} \right\}. \tag{20}$$

This can be seen as follows. For $V_0 = \ker \mathbf{t}_L^T$, using $\mathbf{t}_L^T = (1, 0)$ implies that the first component of $\mathbf{u} \in V_0$ is zero and that the second one can be chosen arbitrarily. For $V_1 = \text{im } D$, note that both rows of D are identical. Hence, every $\mathbf{u} = D\mathbf{v} \in V_1$ must have the same first and second component.

At the level of \mathbb{R}^2 , the inverse J of D can be represented as

$$J = T \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{21}$$

Indeed, if $\mathbf{u} = \begin{pmatrix} 0 \\ \mathbf{u}_2 \end{pmatrix} \in V_0$, then

$$J D \mathbf{u} = T \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{T} \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_2 \\ \mathbf{u}_2 \end{pmatrix} = \begin{pmatrix} 0 \\ \mathbf{u}_2 \end{pmatrix} = \mathbf{u}. \tag{22}$$

Similarly, if $\mathbf{u} = \mathbf{u}_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \in V_1$, then

$$D J \mathbf{u} = \frac{1}{T} \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} T \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ \mathbf{u}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_1 \end{pmatrix} = \mathbf{u}. \tag{23}$$

Hence, $J D = \text{id}_{V_0}$ and $D J = \text{id}_{V_1}$, where id_{V_i} is the identity on V_i .

Note that the matrix representation of J at the level of \mathbb{R}^s is not unique since J is only defined on $V_1 = \ker \boldsymbol{\sigma}^* = \text{im } D$. In general, a linear mapping from \mathbb{R}^s to \mathbb{R}^s is determined uniquely by s^2 real parameters (the entries of the corresponding matrix representation). Since J is defined as a mapping between the $(s - 1)$ -dimensional spaces V_1 and V_0 , it is given by $(s - 1)^2$ parameters. Requiring that J maps to V_0 yields s additional constraints $\mathbf{t}_L^T J = \mathbf{0}^T$. Hence, $s - 1$ degrees of freedom remain for any matrix representation of J at the level of \mathbb{R}^s . Indeed, adding $\mathbf{v}\boldsymbol{\sigma}^*$ to any matrix representation of J in \mathbb{R}^s results in another valid representation if $\mathbf{v} \in V_0 = \ker \mathbf{t}_L^T$. In this example, another valid representation of J at the level of \mathbb{R}^2 is

$$J = T \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \tag{24}$$

which still satisfies $JD = \text{id}_{V_0}$, $DJ = \text{id}_{V_1}$, and yields the same results as the previous matrix representation when applied to any $\mathbf{v} \in V_1$.

Now, we have introduced the inverse J of $D: V_0 \rightarrow V_1$, which is a discrete integral operator $J: V_1 \rightarrow V_0$. However, the integral operator J is only defined for elements of the space $V_1 = \text{im } D$. Hence, one has to make sure that a generic right hand side vector \mathbf{f} is in the range of the derivative operator D in order to apply the inverse J . To guarantee this, components in the direction of grid oscillations \mathbf{o} must be removed. For this, the discrete projection/filter operator

$$F = I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} \tag{25}$$

will be used.

Lemma 3.10 *The projection/filter operator F defined in (25) is an orthogonal projection onto the range of D , i.e. onto $V_1 = \text{im } D = (\ker D^*)^\perp$. It is symmetric and positive semidefinite with respect to the scalar product induced by M .*

Proof Clearly, $\frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2}$ is the usual orthogonal projection onto $\text{span}\{\mathbf{o}\}$ [40, Theorems 9.14 and 9.15]. Hence, F is the orthogonal projection onto the orthogonal complement $\text{span}\{\mathbf{o}\}^\perp = (\ker D^*)^\perp = \text{im } D = V_1$. In particular, for a (real or complex valued) discrete grid function \mathbf{u} ,

$$\langle \mathbf{u}, F\mathbf{u} \rangle_M = \langle \mathbf{u}, \mathbf{u} \rangle_M - \frac{|\langle \mathbf{u}, \mathbf{o} \rangle_M|^2}{\|\mathbf{o}\|_M^2} \geq 0 \tag{26}$$

because of the Cauchy-Schwarz inequality [40, Theorem 9.3]. □

Now, all ingredients to mimic the integral equation (10) have been provided. Applying at first the discrete projection operator F and second the discrete integral operator J to a generic right hand side \mathbf{f} results in $JF\mathbf{f}$, which is a discrete analog of the integral $\int_0^t f d\tau$. Additionally, the initial condition has to be imposed, which is done by adding the constant initial value as $u_0 \otimes \mathbf{1}$. Putting it all together, a new class of SBP schemes mimicking the integral equation (10) discretely is proposed as

$$\mathbf{u} = u_0 \otimes \mathbf{1} + JF\mathbf{f}, \quad u_+ = \mathbf{t}_R^T \mathbf{u}. \tag{27}$$

For a scalar ODE (1), the first term on the right-hand side of the proposed scheme (27) is $u_0 \otimes \mathbf{1} = (u_0, \dots, u_0)^T \in \mathbb{R}^s$. Note that (27) is an implicit scheme since $f = f(t, u)$.

Example 3.11 Continuing Examples 3.4 and 3.9, the adjoint of \mathbf{o} is

$$\mathbf{o}^* = \mathbf{o}^T M = (-1, 1) \frac{T}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \frac{T}{2} (-1, 1). \tag{28}$$

Hence, $\|\mathbf{o}\|_M^2 = \mathbf{o}^* \mathbf{o} = T$, and the projection/filter operator (25) is

$$F = I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{1}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{29}$$

Thus, F is a smoothing filter operator that removes the highest grid oscillations and maps a grid function into the image of the derivative operator D . Hence, the inverse J , the discrete integral operator, can be applied after F , resulting in

$$JF = J \left(I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} \right) = T \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \frac{T}{2} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}. \tag{30}$$

Finally, for an arbitrary $\mathbf{u} \in \mathbb{R}^2$,

$$JF\mathbf{u} = \frac{T}{2} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{pmatrix} = \frac{T}{2} \begin{pmatrix} 0 \\ \mathbf{u}_1 + \mathbf{u}_2 \end{pmatrix} \in V_0. \tag{31}$$

A more involved example of the development presented here is given in Appendix B.

3.1 Summarizing the Development

As stated earlier, the SBP time integration scheme (27) mimics the integral reformulation (10) of the ODE (1). Instead of using J as discrete analog of the integral operator $\int_0^t \cdot d\tau$ directly, the projection/filter operator F defined in (25) must be applied first in order to guarantee that the generic vector \mathbf{f} is in the image of D . Finally, the initial condition is imposed strongly.

Note that

$$\mathbf{t}_L^T \mathbf{u} = \mathbf{t}_L^T (u_0 \otimes \mathbf{1}) + \mathbf{t}_L^T J \left(I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} \right) \mathbf{f} = u_0. \tag{32}$$

The second summand vanishes because J returns a vanishing value at $t = 0$, i.e. $\mathbf{t}_L^T J = \mathbf{0}^T$, since $J: V_1 \rightarrow V_0$ maps onto $V_0 = \ker \mathbf{t}_L^T$.

Note that the projection/filter operator F is required in (27), since the discrete integral operator J only operates on objects in V_1 . Note also that the matrix representations of J in \mathbb{R}^s given in the examples above are constructed such that they should be applied only to vectors $\mathbf{v} \in V_1$. As explained in Example 3.9, the matrix representation of J in \mathbb{R}^s is not unique. Thus, choosing any of these representations without applying the filter/projection operator F would result in undefined/unpredictable behavior. The projection/filter operator F is necessary to make (27) well-defined. Indeed, the product JF is well-defined, i.e. it is the same for any matrix representation of J , since F maps to $V_1 = \text{im } D$ and the action of J is defined uniquely on this space. In particular, JF itself is a valid matrix representation of J in \mathbb{R}^s . As an example, JF in (30) is a linear combination of the possible representations (21) and (24) of J in \mathbb{R}^s and thus also a representation of J in \mathbb{R}^s .

Another argument for the necessity of the filter/projection operator F can be derived using the following result.

Lemma 3.12 *Let D be a nullspace consistent first derivative SBP operator. Then, $DJF = F$ and the solution \mathbf{u} of (27) satisfies*

$$D\mathbf{u} = F\mathbf{f}. \tag{33}$$

Before proving Lemma 3.12, we discuss its meaning here. In general, it is not possible to find a solution \mathbf{u} of $D\mathbf{u} = \mathbf{f}$ for an arbitrary right-hand side \mathbf{f} , since D is not invertible on \mathbb{R}^s . Multiplying \mathbf{f} by the orthogonal projection operator F ensures that the new right-hand side $F\mathbf{f}$ of (33) is in the image of D and hence that (33) can be solved for any given \mathbf{f} .

This projection F onto $V_1 = \text{im } D$ is necessary in the discrete case because of the finite dimensions.

Proof of Lemma 3.12 Taking the discrete derivative on both sides of (27) results in

$$D\mathbf{u} = DJF\mathbf{f}, \tag{34}$$

since $D(u_0 \otimes \mathbf{1}) = u_0 \otimes (D\mathbf{1}) = \mathbf{0}$. Hence, (33) holds if $DJF = F$. To show $DJF = F$, it suffices to show $DJF\mathbf{f} = F\mathbf{f}$ for arbitrary \mathbf{f} . Write \mathbf{f} as $\mathbf{f} = D\mathbf{v} + \alpha\mathbf{o}$, where $\mathbf{t}_L^T\mathbf{v} = 0$ and $\alpha \in \mathbb{R}$. This is always possible since D is nullspace consistent. Then,

$$DJF\mathbf{f} = DJFD\mathbf{v} + \alpha DJF\mathbf{o} = DJD\mathbf{v} = D\mathbf{v}, \tag{35}$$

where we used $FD = D$, $F\mathbf{o} = \mathbf{0}$, $JD = \text{id}_{V_0}$, and $\mathbf{v} \in V_0 = \ker \mathbf{t}_L^T$. Using again $FD = D$ and $F\mathbf{o} = \mathbf{0}$, we get

$$D\mathbf{v} = FD\mathbf{v} + \mathbf{0} = F(D\mathbf{v} + \alpha\mathbf{o}) = F\mathbf{f}. \tag{36}$$

Hence, $DJF\mathbf{f} = F\mathbf{f}$. □

Remark 3.13 In the context of SBP operators, two essentially different interpretations of integrals arise. Firstly, the integral $\int_0^T \cdot d\tau$ gives the L^2 scalar product, approximated by the mass matrix M which maps discrete functions to scalar values. Secondly, the integral $\int_0^t \cdot d\tau$ is the inverse of the derivative with vanishing values at $t = 0$. This operator is discretized as J on its domain of definition $V_1 = \text{im } D$ and maps a discrete grid function in $V_1 = \text{im } D$ to a discrete grid function in $V_0 = \ker \mathbf{t}_L^T$.

3.2 Linear Stability

In this section, linear stability properties of the new scheme (27) are established.

Theorem 3.14 *For nullspace consistent SBP operators, the scheme (27) is A stable.*

Proof For the scalar linear ODE (8) with $\text{Re } \lambda \leq 0$, the energy method will be applied to the scheme (27). We write $\bar{\cdot}$ to denote the complex conjugate. Using $\mathbf{t}_L^T\mathbf{u} = u_0$ from (32) and $u_+ = \mathbf{t}_R^T\mathbf{u}$ from the definition of the scheme (27), the difference of the energy at the final and initial time is

$$\begin{aligned} |u_+|^2 - |u_0|^2 &= \left| \mathbf{t}_R^T\mathbf{u} \right|^2 - \left| \mathbf{t}_L^T\mathbf{u} \right|^2 = \bar{\mathbf{u}}^T \mathbf{t}_R \mathbf{t}_R^T \mathbf{u} - \bar{\mathbf{u}}^T \mathbf{t}_L \mathbf{t}_L^T \mathbf{u} \\ &= \bar{\mathbf{u}}^T (MD + (MD)^T)\mathbf{u}, \end{aligned} \tag{37}$$

where the SBP property (2) has been used in the last equality. As shown in Lemma 3.12, the scheme (27) yields $D\mathbf{u} = F\mathbf{f}$. For the scalar linear ODE (8), $\mathbf{f} = \lambda\mathbf{u}$. Hence, we can replace $D\mathbf{u}$ by $F\mathbf{f} = \lambda F\mathbf{u}$ in (37), resulting in

$$\begin{aligned} |u_+|^2 - |u_0|^2 &= 2 \text{Re}(\bar{\mathbf{u}}^T MD\mathbf{u}) \\ &= 2 \text{Re}(\lambda \bar{\mathbf{u}}^T M F \mathbf{u}) = 2 \underbrace{\text{Re}(\lambda)}_{\leq 0} \underbrace{\bar{\mathbf{u}}^T M F \mathbf{u}}_{\geq 0} \leq 0. \end{aligned} \tag{38}$$

The second factor is non-negative because F is positive semidefinite with respect to the scalar product induced by M , cf. Lemma 3.10. Therefore, $|u_+|^2 \leq |u_0|^2$, implying that the scheme is A stable. □

In general, the novel SBP scheme (27) is not L stable, cf. Remark 3.19.

3.3 Characterization as a Runge-Kutta Method

Unsurprisingly, the new method (27) can be characterized as a Runge-Kutta method.

Theorem 3.15 *For nullspace consistent SBP operators that are at least first order accurate, the method (27) is a Runge-Kutta method with Butcher coefficients*

$$A = \frac{1}{T} J F = \frac{1}{T} J \left(I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} \right), \quad b = \frac{1}{T} M \mathbf{1}, \quad c = \frac{1}{T} (\tau_1, \dots, \tau_s)^T. \tag{39}$$

Proof First, note that one step (6) from zero to T of a Runge-Kutta method with coefficients A, b, c can be written as

$$\mathbf{u} = u_0 \otimes \mathbf{1} + T A \mathbf{f}, \quad u_+ = u_0 + T b^T \mathbf{f}, \tag{40}$$

where the right hand side vector \mathbf{f} is given by $\mathbf{f}_i = f(c_i T, \mathbf{u}_i)$. Comparing this expression with

$$\mathbf{u} = u_0 \otimes \mathbf{1} + J F \mathbf{f}, \quad u_+ = \mathbf{t}_R^T \mathbf{u}, \tag{27}$$

where the right hand side is given by $\mathbf{f}_i = f(\tau_i, \mathbf{u}_i)$, the form of A and c is immediately clear. The new value u_+ of the new SBP method (27) is

$$u_+ = \mathbf{t}_R^T \mathbf{u} = \mathbf{1}^T \mathbf{t}_R \mathbf{t}_R^T \mathbf{u}. \tag{41}$$

Using the SBP property (2) and $D \mathbf{1} = \mathbf{0}$,

$$u_+ = \mathbf{1}^T \mathbf{t}_L \mathbf{t}_L^T \mathbf{u} + \mathbf{1}^T M D \mathbf{u} + \mathbf{1}^T D^T M \mathbf{u} = \mathbf{t}_L^T \mathbf{u} + \mathbf{1}^T M D \mathbf{u}. \tag{42}$$

Inserting $\mathbf{t}_L^T \mathbf{u} = u_0$ and $D \mathbf{u}$ from (33) results in

$$u_+ = u_0 + \mathbf{1}^T M \left(I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} \right) \mathbf{f} = u_0 + \mathbf{1}^T M \mathbf{f} - \frac{\langle \mathbf{1}, \mathbf{o} \rangle_M \langle \mathbf{o}, \mathbf{f} \rangle_M}{\|\mathbf{o}\|_M^2}. \tag{43}$$

Because of $\mathbf{1} \in \text{im } D \perp \ker D^* \ni \mathbf{o}$, we have $\langle \mathbf{1}, \mathbf{o} \rangle_M = 0$. Hence,

$$u_+ = u_0 + \mathbf{1}^T M \mathbf{f}. \tag{44}$$

Comparing this expression with (40) yields the final assertion for b . □

Lemma 3.16 *For nullspace consistent SBP operators, the first row of the Butcher coefficient matrix A in (39) of the method (27) is zero if $\mathbf{t}_L = (1, 0, \dots, 0)^T$.*

Proof By definition, J yields a vector with vanishing initial condition at the left endpoint, i.e. $\mathbf{t}_L^T J = \mathbf{0}^T$. Because of $\mathbf{t}_L = (1, 0, \dots, 0)^T$, we have $\mathbf{0}^T = \mathbf{t}_L^T J = J[1, :]$, which is the first row of J , where a notation as in Julia [1] has been used. □

3.4 Operator Construction

To implement the SBP scheme (27), the product $J F$ has to be computed, which is (except for a scaling by T^{-1}) the matrix A of the corresponding Runge-Kutta method, cf. Theorem 3.15. Since the projection operator F maps vectors into $V_1 = \text{im } D$, the columns of F are in the image of the nullspace consistent SBP derivative operator D . Hence, the matrix equation

$DX = F$ can be solved for X , which is a matrix of the same size as D , e.g. via a QR factorization, yielding the least norm solution. Then, we have to ensure that the columns of JF are in $V_0 = \ker \mathbf{t}_L^T$, since J maps V_1 into V_0 . This can be achieved by subtracting $\mathbf{t}_L^T X[:, j]$ from each column $X[:, j]$ of X , $j \in \{1, \dots, s\}$, where a notation as in Julia [1] has been used. After this correction, we have $X = JF$. Finally, we need to solve (27) for \mathbf{u} for each step, using the operator JF constructed as described above.

3.5 Lobatto IIIA Schemes

General characterizations of the SBP-SAT scheme (4) on Radau and Lobatto nodes as classical collocation Runge-Kutta methods (Radau IA and IIA or Lobatto IIIC, respectively) have been obtained in [29]. A similar characterization will be obtained in this section.

Theorem 3.17 *If the SBP operator D is given by the nodal polynomial collocation scheme on Lobatto-Legendre nodes, the SBP method (27) is the classical Lobatto IIIA method.*

Proof The Lobatto IIIA methods are given by the nodes c and weights b of the Lobatto-Legendre quadrature, just as the SBP method (27). Hence, it remains to prove that the classical condition $C(s)$ is satisfied [4, Section 344], where

$$C(\eta): \quad \sum_{j=1}^s a_{i,j} c_j^{q-1} = \frac{1}{q} c_i^q, \quad i \in \{1, \dots, s\}, \quad q \in \{1, \dots, \eta\}. \quad (45)$$

In other words, all polynomials of degree $\leq p = s - 1$ must be integrated exactly by A with vanishing initial value at $t = 0$. By construction of A , see (39), this is satisfied for all polynomials of degree $\leq p - 1$, since the grid oscillations are given by $\mathbf{o} = \varphi_p$, where φ_p is the Legendre polynomial of degree p , cf. [38, Example 3.6].

Finally, it suffices to check whether φ_p is integrated exactly by A . The left hand side of (45) yields

$$A\varphi_p = \frac{1}{T} J \left(\mathbf{I} - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} \right) \varphi_p = 0, \quad (46)$$

since $\mathbf{o} = \varphi_p$. On the right hand side, transforming the time domain to the standard interval $[-1, 1]$, the analytical integrand of φ_p is

$$\int_{-1}^x \varphi_p(s) ds = \frac{1}{p(p+1)} \int_{-1}^x \partial_s \left[(s^2 - 1)\varphi_p'(s) \right] ds = \frac{1}{p(p+1)} (x^2 - 1)\varphi_p'(x), \quad (47)$$

since the Legendre polynomials satisfy Legendre’s differential equation

$$\partial_x((1 - x^2)\partial_x \varphi_p(x)) + p(p+1)\varphi_p(x) = 0. \quad (48)$$

Hence, $\int_{-1}^x \varphi_p(s) ds$ vanishes exactly at the $s = p + 1$ Legendre nodes for polynomials of degree p , which are ± 1 and the roots of φ_p' . Thus, the analytical integral of φ_p vanishes at all grid nodes. \square

Example 3.18 Continuing Examples 3.4 and 3.11, the nodes $c = \frac{1}{T}(\tau_1, \tau_2)^T = (0, 1)^T$ are the nodes of the Lobatto-Legendre quadrature with two nodes in $[0, 1]$. Moreover, the corresponding weights are given by

$$b = \frac{1}{T} M\mathbf{1} = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \quad (49)$$

Finally, the remaining Butcher coefficients are given by

$$A = \frac{1}{T} JF = \frac{1}{T} J \left(I - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}^*}{\|\boldsymbol{\sigma}\|_M^2} \right) = \frac{1}{2} \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad (50)$$

which are exactly the coefficients of the Lobatto IIIA method with $s = 2$ stages.

Remark 3.19 Since the Lobatto IIIA methods are neither L nor B stable, the new SBP method (27) is in general not L or B stable, too.

Remark 3.20 Because of Lemma 3.16, the classical Gauss, Radau IA, Lobatto IIIB, and Lobatto IIIC methods cannot be expressed in the form (27). The classical Radau I, Radau II, and Radau IIA methods are also not included in the class (27). For example, for two nodes, these methods have the A matrices

$$\begin{pmatrix} 0 & 0 \\ 1/3 & 1/3 \end{pmatrix}, \quad \begin{pmatrix} 1/3 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 5/12 & -1/12 \\ 3/4 & 1/4 \end{pmatrix} \quad (51)$$

while the methods (27) on the left and right Radau nodes yield the matrices

$$\begin{pmatrix} 0 & 0 \\ 1/6 & 1/2 \end{pmatrix}, \quad \begin{pmatrix} 1/4 & 1/12 \\ 3/4 & 1/4 \end{pmatrix}. \quad (52)$$

We develop a related SBP time integration scheme that includes the classical Lobatto IIIB collocation method in the appendix. Additionally, we mention why it seems to be difficult to describe Gauss collocation methods in a general SBP setting, cf. Appendix C.

3.6 Order of Accuracy

Next, we establish results on the order of accuracy of the new class of SBP time integration methods.

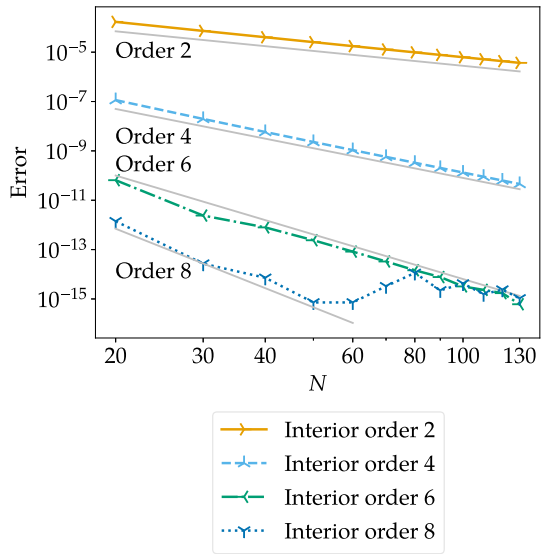
Theorem 3.21 *For nullspace consistent SBP operators that are p th order accurate with $p \geq 1$, the Runge-Kutta method (39) associated to the SBP time integration scheme (27) has an order of accuracy of*

- a) at least p for general mass matrices M .
- b) at least $2p$ for diagonal mass matrices M .

The technical proof of Theorem 3.21 is given in Appendix A.

Remark 3.22 The result on the order of accuracy given in Theorem 3.21 may appear counter-intuitive at first when looking from the perspective of classical finite difference SBP operators, since diagonal norm matrices are usually less accurate in this context. Indeed, finite difference SBP operators for the first derivative with a diagonal norm matrix have an order of accuracy of $2q$ in the interior and $r \leq q$ at the boundaries [17], where usually $r = q$. In contrast, the corresponding dense norm operators have an order of accuracy $2q$ in the interior and $2q - 1$ at the boundaries. Hence, the total order of accuracy $p = q$ for diagonal mass matrices is smaller than the order of accuracy $p = 2q - 1$ for dense norms. However, dense norms are not guaranteed to result in the same high order of accuracy when used as a quadrature rule. Thus, the total order of accuracy can be smaller even if the pointwise accuracy as a derivative operator is higher (which basically corresponds to the *stage order* in the context of Runge-Kutta methods).

Fig. 1 Convergence behavior of the SBP scheme (27) using the diagonal norm operators of [21] for the non-stiff test problem (53)



4 Numerical Experiments

Numerical experiments corresponding to the ones in [19] will be conducted. The novel SBP methods (27) have been implemented in Julia v1.5 [1] and Matplotlib [14] has been used to generate the plots. The source code for all numerical examples is available online [35]. After computing the operators $A = \frac{1}{T} JF$ as described in Section 3.4 and inserting the right-hand sides f of the ODEs considered in the following into the scheme (27), the resulting linear systems are solved using the backslash operator in Julia.

Numerical experiments are shown only for new SBP methods (27) based on finite difference SBP operators and not for methods based on Lobatto quadrature, since the classical Lobatto IIIA and IIIB schemes are already well-known in the literature. The diagonal norm finite difference SBP operators of [21] use central finite difference stencils in the interior of the domain and adapted boundary closures to satisfy the SBP property (2). The Butcher coefficients of some of these methods are given in Appendix D.

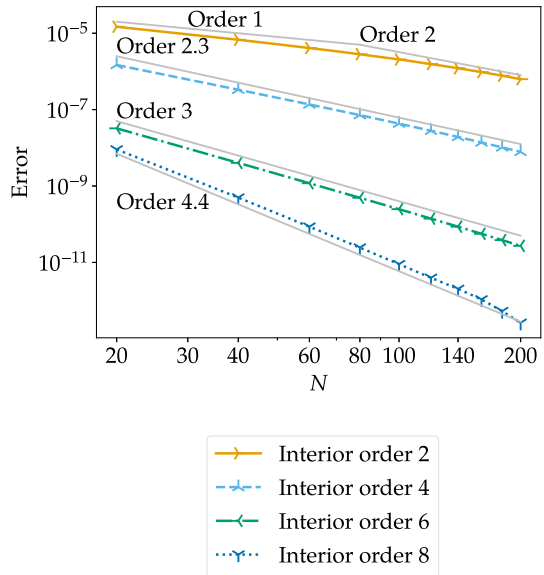
4.1 Non-stiff Problem

The non-stiff test problem

$$u'(t) = -u(t), \quad u(0) = 1, \tag{53}$$

with analytical solution $u(t) = \exp(-t)$ is solved in the time interval $[0, 1]$ using the SBP method (27) with the diagonal norm operators of [21]. The errors of the numerical solutions at the final time are shown in Fig. 1. As can be seen, they converge with an order of accuracy equal to the interior approximation order of the diagonal norm operators. For the operator with interior order eight, the error reaches machine precision for $N = 50$ nodes and does not decrease further. These results are comparable to the ones obtained by SBP-SAT schemes in [19] and match the order of accuracy of the corresponding Runge-Kutta methods guaranteed by Theorem 3.21.

Fig. 2 Convergence behavior of the SBP scheme (27) using the diagonal norm operators of [21] for the stiff test problem (54)



4.2 Stiff Problem

The stiff test problem

$$u'(t) = \lambda(u(t) - \exp(-t)) - \exp(-t), \quad u(0) = 1, \tag{54}$$

with analytical solution $u(t) = \exp(-t)$ and parameter $\lambda = 1000$ is solved in the time interval $[0, 1]$. The importance of such test problems for stiff equations has been established in [28]. Using the diagonal norm operators of [21] for the method (27) yields the convergence behavior shown in Fig. 2. Again, the results are comparable to the ones obtained by SBP-SAT schemes in [19]. In particular, the order of convergence is reduced to the approximation order at the boundaries, exactly as for the SBP-SAT schemes of [25]. Such an order reduction for stiff problems is well-known in the literature on time integration methods, see [28] and [12, Chapter IV.15].

5 Summary and Discussion

A novel class of A stable summation by parts time integration methods has been proposed. Instead of using simultaneous approximation terms to impose the initial condition weakly, the initial condition is imposed strongly. Similarly to previous SBP time integration methods, the new schemes can be reformulated as implicit Runge-Kutta methods.

Compared to the SAT approach, some linear and nonlinear stability properties such as L and B stability are lost in general. On the other hand, well-known A stable methods such as the Lobatto IIIA schemes are included in this new SBP framework. Additionally, a related SBP time integration method has been proposed which includes the classical Lobatto IIIB schemes.

This article provides new insights into the relations of numerical methods and contributes to the discussion of whether SBP properties are necessarily involved in numerical schemes for differential equations which are provably stable.

5.1 Final Reflections on Obtained Results

We have concentrated on classical collocation Runge-Kutta methods when looking for known schemes in the new class of SBP time integration methods, since these have direct connections to quadrature rules, which are closely connected to the SBP property [13]. We are not aware of other classical Runge-Kutta methods that are contained in the new class of SBP methods proposed and analyzed in this article besides Lobatto IIIA and IIIB schemes. The implicit equations that need to be solved per time step for the new methods can be easier to solve than the ones occurring in SBP-SAT methods, e.g. since the first stage does not require an implicit solution at all for some methods (Lemma 3.16). On the other hand, the new classes of methods do not necessarily have the same kind of nonlinear stability properties as previous SBP-SAT methods. Hence, a thorough parameter search and comparison of the methods would be necessary for a detailed comparison, which is beyond the scope of this initial article.

In general, methods constructed using SBP operators often imply certain stability properties automatically, which are usually more difficult to guarantee when numerical methods are constructed without these restrictions. On the other hand, not imposing SBP restrictions can possibly result in more degrees of freedom which can be used to construct more flexible and possibly a larger number of numerical methods. From a practical point of view, the availability of numerical algorithms in standard software packages and the efficiency of the implementations are also very important. In this respect, established time integration methods have definitely many advantages, since they are widespread and considerable efforts went into the available implementations. Additionally, a practitioner can choose to make a trade-off between guaranteed stability properties and the efficiency of schemes that “just work” in practice, although only weaker stability results might be available. For example, linearly implicit time integration schemes such as Rosenbrock methods can be very efficient for certain problems.

Having said all that, it is important to note that the process of discretizing differential equations is filled with pitfalls. Potentially unstable schemes may lead to results that seem correct but are in fact erroneous. A provably stable scheme can be seen as a quality stamp.

Acknowledgements Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST). Jan Nordström was supported by Vetenskapsrådet, Sweden grant 2018-05084 VR and by the Swedish e-Science Research Center (SeRC) through project ABL in SESSI.

Funding Open access funding provided by Linköping University.

Data Availability The source code and datasets generated and analyzed for this study are available in [35].

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the

article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Proof of Theorem 3.21

Here, we present the technical proof of Theorem 3.21.

Proof of Theorem 3.21 Consider the classical simplifying assumptions

$$B(\xi): \quad \mathbf{b}^T \mathbf{c}^{q-1} = \frac{1}{q}, \quad q \in \{1, \dots, \xi\}, \quad (55)$$

$$C(\eta): \quad A \mathbf{c}^{q-1} = \frac{1}{q} \mathbf{c}^q, \quad q \in \{1, \dots, \eta\}, \quad (56)$$

$$D(\zeta): \quad A^T B \mathbf{c}^{q-1} = \frac{1}{q} B(\mathbf{1} - \mathbf{c}^q), \quad q \in \{1, \dots, \zeta\}, \quad (57)$$

where $B = \text{diag}(\mathbf{b})$ is a diagonal matrix. To prove Theorem 3.21, we will use the following result of Butcher [3] for Runge-Kutta methods.

If $B(\xi), C(\eta), D(\zeta), \xi \leq 1 + \eta + \zeta$, and $\xi \leq 2 + 2\eta$, then the Runge-Kutta method has an order of accuracy at least ξ .

(a) General mass matrices M

– Proving $B(p)$

The quadrature rule given by the weights $\mathbf{b} = \frac{1}{T} M \mathbf{1}$ of the p th order accurate SBP operator is exact for polynomials of degree $p - 1$ for a general norm matrix M [7, Theorem 1]. Hence, $B(p)$ is satisfied.

– Proving $C(p)$

As in the proof of Theorem 3.17, $C(p)$ is satisfied by construction of $A = \frac{1}{T} J F$, since all polynomials of degree $\leq p - 1$ are integrated exactly by A with vanishing initial value at $t = 0$.

– Concluding

Since only $B(p)$ is satisfied in general, the order of the Runge-Kutta method is limited to p and we do not need further conditions on the simplifying assumption $D(\zeta)$. Hence, it suffices to consider the empty condition $D(0)$. Since $B(p), C(p - 1)$, and $D(0)$ are satisfied, the result of Butcher cited above implies that the Runge-Kutta method has at least an order of accuracy at least p .

(b) Diagonal mass matrices M

– Proving $B(2p)$

For a diagonal norm matrix M , the quadrature rule given by the weights $\mathbf{b} = \frac{1}{T} M \mathbf{1}$ of the p th order accurate SBP operator is exact for polynomials of degree $2p - 1$ [7, Theorem 2]. Hence, $B(2p)$ is satisfied.

– Proving $C(p)$

The proof of $C(p)$ is exactly the same as for general mass matrices M .

– Proving $D(p - 1)$

It suffices to consider a scaled SBP operator such that $T = 1, \mathbf{t}_L^T \boldsymbol{\tau} = 0, \mathbf{t}_R^T \boldsymbol{\tau} = 1$.

Then, inserting the Runge-Kutta coefficients (39), $D(p - 1)$ is satisfied if for all $q \in \{1, \dots, p - 1\}$

$$F^T J^T M \boldsymbol{\tau}^{q-1} = \frac{1}{q} M (\mathbf{1} - \boldsymbol{\tau}^q). \tag{58}$$

This equation is satisfied if and only if for all $\mathbf{u} \in \mathbb{R}^s$

$$\mathbf{u}^T F^T J^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{u}^T M (\mathbf{1} - \boldsymbol{\tau}^q) = 0. \tag{59}$$

Every \mathbf{u} can be written as $\mathbf{u} = D\mathbf{v} + \alpha\mathbf{o}$, where $\mathbf{t}_L^T \mathbf{v} = 0$ and $\alpha \in \mathbb{R}$, since D is nullspace consistent. Hence, it suffices to consider

$$\begin{aligned} & \mathbf{v}^T D^T F^T J^T M \boldsymbol{\tau}^{q-1} + \alpha \mathbf{o}^T F^T J^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{v}^T D^T M \mathbf{1} - \frac{1}{q} \alpha \mathbf{o}^T M \mathbf{1} \\ & + \frac{1}{q} \mathbf{v}^T D^T M \boldsymbol{\tau}^q + \frac{1}{q} \alpha \mathbf{o}^T M \boldsymbol{\tau}^q \\ & = \mathbf{v}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{v}^T D^T M \mathbf{1} + \frac{1}{q} \mathbf{v}^T D^T M \boldsymbol{\tau}^q. \end{aligned} \tag{60}$$

Here, we used that $JFD\mathbf{v} = \mathbf{v}$ by definition of $\mathbf{v} \in V_0 = \ker \mathbf{t}_L^T$. The filter F removes grid oscillations, i.e. $F\mathbf{o} = \mathbf{0}$. Additionally, grid oscillations are orthogonal to the constant $\mathbf{1}$ for SBP operators that are at least first-order accurate and orthogonal to $\boldsymbol{\tau}^q$ for $q \leq p - 1$ in general (since $D\boldsymbol{\tau}^p = p\boldsymbol{\tau}^{p-1}$).

Using the SBP property (2), the expression above can be rewritten as

$$\begin{aligned} & \mathbf{v}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{v}^T D^T M \mathbf{1} + \frac{1}{q} \mathbf{v}^T D^T M \boldsymbol{\tau}^q \\ & = \mathbf{v}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{t}_R^T \mathbf{v} + \frac{1}{q} \mathbf{v}^T \mathbf{t}_R - \frac{1}{q} \mathbf{v}^T M D \boldsymbol{\tau}^q = 0, \end{aligned} \tag{61}$$

where we used $\mathbf{t}_L^T \mathbf{v} = 0$, $\mathbf{t}_R^T \boldsymbol{\tau}^q = 1$, and the accuracy of D . This proves $D(p - 1)$.

– Concluding

Since $B(2p)$, $C(p)$, and $D(p - 1)$ are satisfied, the result of Butcher cited above guarantees an order of accuracy of at least $2p$.

□

Remark A.1 The simplifying assumptions used in the proof of Theorem 3.21 can be satisfied to higher order of accuracy. For example, the Lobatto IIIA methods satisfy $C(p + 1)$ instead of only $C(p)$, cf. the proof of Theorem 3.17. In that case, the order of the associated quadrature still limits the order of the corresponding Runge-Kutta method to $2p$.

Similarly, the method based on left Radau quadrature mentioned in Example 3.20 has Butcher coefficients

$$A = \begin{pmatrix} 0 & 0 \\ 1/6 & 1/2 \end{pmatrix}, \quad b = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 2/3 \end{pmatrix}. \tag{62}$$

Thus, the quadrature condition is satisfied to higher order of accuracy, i.e. $B(3)$ holds instead of only $B(2)$ for $p = 1$. Nevertheless, the Runge-Kutta method is only second-order accurate, i.e. it satisfies the order conditions

$$b^T \mathbf{1} = 1, \quad b^T A \mathbf{1} = \frac{1}{2}, \tag{63}$$

but violates one of the additional conditions for a third-order accurate method, i.e.

$$b^T(A\mathbf{1})^2 = \frac{1}{3}, \quad b^T A^2\mathbf{1} = \frac{1}{4} \neq \frac{1}{6}, \tag{64}$$

since neither $C(p + 1)$ nor $D(p)$ is satisfied.

B An Example Using Gauss Quadrature

Here, we follow the derivation of the scheme presented in Sect. 3 using a more complicated SBP operator that does not include any boundary node. Consider the SBP operator of order $p = 2$ induced by classical Gauss-Legendre quadrature on $[0, T]$, using the nodes

$$\tau_1 = \frac{T}{10}(5 - \sqrt{15}), \quad \tau_2 = \frac{T}{2}, \quad \tau_3 = \frac{T}{10}(5 + \sqrt{15}). \tag{65}$$

The associated SBP operator exactly differentiating polynomials of degree $p = 2$ is given by

$$D = \frac{\sqrt{15}}{3T} \begin{pmatrix} -3 & 4 & -1 \\ -1 & 0 & 1 \\ 1 & -4 & 3 \end{pmatrix}, \quad M = \frac{T}{18} \begin{pmatrix} 5 & & \\ & 8 & \\ & & 5 \end{pmatrix}, \quad \mathbf{t}_L = \frac{1}{6} \begin{pmatrix} 5 + \sqrt{15} \\ -4 \\ 5 - \sqrt{15} \end{pmatrix}, \quad \mathbf{t}_R = \frac{1}{6} \begin{pmatrix} 5 - \sqrt{15} \\ -4 \\ 5 + \sqrt{15} \end{pmatrix}. \tag{66}$$

Therefore,

$$D^* = M^{-1}D^T M = \frac{\sqrt{15}}{30T} \begin{pmatrix} -30 & -16 & 10 \\ 25 & 0 & -25 \\ -10 & 16 & 30 \end{pmatrix} \tag{67}$$

and $\ker D^* = \text{span}\{\mathbf{o}\}$, where $\mathbf{o} = (4, -5, 4)^T$ represents the highest resolvable grid oscillation; \mathbf{o} is orthogonal to $\text{im } D$, since $\mathbf{o}^T M D = \mathbf{0}^T$. The adjoint of \mathbf{o} is

$$\mathbf{o}^* = \mathbf{o}^T M = \frac{10T}{9}(1, -2, 1). \tag{68}$$

Thus, the filter/projection operator (25) is

$$F = I - \frac{\mathbf{o}\mathbf{o}^*}{\|\mathbf{o}\|_M^2} = \frac{1}{18} \begin{pmatrix} 14 & 8 & -4 \\ 5 & 8 & 5 \\ -4 & 8 & 14 \end{pmatrix}. \tag{69}$$

The vector spaces V_0, V_1 associated to the given SBP operator are

$$V_0 = \left\{ \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ -(4 + \sqrt{15})\alpha_1 + 2(5 + \sqrt{15})\alpha_2/5 \end{pmatrix} \mid \alpha_1, \alpha_2 \in \mathbb{R} \right\},$$

$$V_1 = \left\{ \begin{pmatrix} \beta_1 \\ \beta_2 \\ -\beta_1 + 2\beta_2 \end{pmatrix} \mid \beta_1, \beta_2 \in \mathbb{R} \right\}. \tag{70}$$

To verify this, observe that both vector spaces are two-dimensional, V_0 is the nullspace of \mathbf{t}_L^T , and V_1 is the nullspace of \mathbf{o}^* .

At the level of \mathbb{R}^3 , the inverse J of D can be represented as

$$J = \frac{3T}{24\sqrt{15}} \begin{pmatrix} 2 & 4(\sqrt{15}-3) & -2 \\ 5 & 4\sqrt{15} & -5 \\ 2 & 4(\sqrt{15}+3) & -2 \end{pmatrix}. \tag{71}$$

Indeed, for $\mathbf{u} = (\alpha_1, \alpha_2, -(4 + \sqrt{15})\alpha_1 + 2(5 + \sqrt{15})\alpha_2/5)^T$, $\alpha_1, \alpha_2 \in \mathbb{R}$,

$$J D \mathbf{u} = \frac{1}{6} \begin{pmatrix} 1 - \sqrt{15} & 4 & -5 + \sqrt{15} \\ -5 - \sqrt{15} & 10 & -5 + \sqrt{15} \\ -5 - \sqrt{15} & 4 & 1 + \sqrt{15} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ -(4 + \sqrt{15})\alpha_1 + 2(5 + \sqrt{15})\alpha_2/5 \end{pmatrix} = \mathbf{u} \tag{72}$$

and similarly for $\mathbf{u} = (\beta_1, \beta_2, -\beta_1 + 2\beta_2)^T$, $\beta_1, \beta_2 \in \mathbb{R}$,

$$D J \mathbf{u} = \frac{1}{2} \begin{pmatrix} 1 & 2 & -1 \\ - & 2 & 0 \\ -1 & 2 & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ -\beta_1 + 2\beta_2 \end{pmatrix} = \mathbf{u}. \tag{73}$$

Hence, $J D = \text{id}_{V_0}$ and $D J = \text{id}_{V_1}$, where id_{V_i} is the identity on V_i .

Since the filter F removes the highest grid oscillations and maps a grid function into the image of the derivative operator D , the inverse J can be applied after F , resulting in

$$J F = \frac{T}{72\sqrt{15}} \begin{pmatrix} -12 + 10\sqrt{15} & -48 + 16\sqrt{15} & -48 + 10\sqrt{15} \\ 45 + 10\sqrt{15} & 16\sqrt{15} & -45 + 10\sqrt{15} \\ 48 + 10\sqrt{15} & 48 + 16\sqrt{15} & 12 + 10\sqrt{15} \end{pmatrix}. \tag{74}$$

Thus, the Butcher coefficients associated to the new SBP projection method given by Theorem 3.15 are

$$A = \frac{1}{T} J F = \frac{1}{72\sqrt{15}} \begin{pmatrix} -12 + 10\sqrt{15} & -48 + 16\sqrt{15} & -48 + 10\sqrt{15} \\ 45 + 10\sqrt{15} & 16\sqrt{15} & -45 + 10\sqrt{15} \\ 48 + 10\sqrt{15} & 48 + 16\sqrt{15} & 12 + 10\sqrt{15} \end{pmatrix}, \tag{75}$$

$$b = \frac{1}{T} M \mathbf{1} = \frac{1}{18} \begin{pmatrix} 5 \\ 8 \\ 5 \end{pmatrix}, \quad c = \frac{1}{T} (\tau_1, \tau_2, \tau_3)^T = \frac{1}{10} \begin{pmatrix} 5 - \sqrt{15} \\ 5 \\ 5 + \sqrt{15} \end{pmatrix}.$$

Note that the coefficients in A are different from those of the classical Runge-Kutta Gauss-Legendre collocation method (while b and c are the same by construction), see also Remark C.4 below. In particular, this method is fourth-order accurate while the classical Gauss-Legendre collocation method with the same number of stages is of order six.

C Lobatto IIIB Schemes

Another scheme similar to (27) can be constructed by considering $-D$ as bijective operator acting on functions that vanish at the right endpoint. The equivalent of J , the inverse of D mapping $\text{im } D$ to the space of grid functions vanishing at the left endpoint, in this context

is written as \tilde{J} , which is the inverse of $-D$ mapping $\text{im}(-D)$ to the space of grid functions vanishing at the right endpoint. The corresponding matrix A of the Runge-Kutta method becomes

$$A = A^{**} = M^{-1}(A^*)^T M, \quad A^* = \frac{1}{T} \tilde{J} \left(I - \frac{\mathbf{o}\mathbf{o}^*}{|\mathbf{o}|_M^2} \right). \tag{76}$$

Using A of (76) and b, c as in (39), the scheme is defined as the Runge-Kutta method (6) with these Butcher coefficients A, b, c .

Theorem C.1 *If the SBP derivative operator D is given by the nodal polynomial collocation scheme on Lobatto-Legendre nodes, the Runge-Kutta method (6) with A as in (76) and b, c as in (39) is the classical Lobatto IIIB method.*

Proof The Lobatto IIIB methods are given by the nodes c and weights b of the Lobatto-Legendre quadrature, just as the SBP method. Hence, it remains to prove that the classical condition $D(s)$ is satisfied [4, Section 344], where

$$D(\zeta): \quad \sum_{i=1}^s b_i c_i^{q-1} a_{i,j} = \frac{1}{q} b_j (1 - c_j^q), \quad j \in \{1, \dots, s\}, \quad q \in \{1, \dots, \zeta\}. \tag{77}$$

In matrix vector notation, this can be written as

$$A^T M \mathbf{c}^{q-1} = \frac{1}{q} M (\mathbf{1} - \mathbf{c}^q) \iff \underbrace{M^{-1} A^T M}_{=A^*} \mathbf{c}^{q-1} = \frac{1}{q} (\mathbf{1} - \mathbf{c}^q), \tag{78}$$

where the exponentiation \mathbf{c}^q is performed pointwise.

In other words, all polynomials of degree $\leq p$ must be integrated exactly by A^* with vanishing final value at $t = 1$. By construction of A , this is satisfied for all polynomials of degree $\leq p - 1$, and the proof can be continued as the one of Theorem 3.17. \square

Remark C.2 Remarks 3.19 and 3.20 hold analogously: Schemes based on (76) are also in general not L or B stable and other classical schemes on Gauss, Radau, or Lobatto nodes are not included in this class.

Similarly to Theorem 3.21, we present some results on the order of accuracy of the new SBP methods. Since this class of methods is made to satisfy the simplifying condition $D(\zeta)$ instead of $C(\eta)$ and stronger results on $C(\eta)$ are necessary to apply the results of Butcher [3], we concentrate on diagonal norms.

Theorem C.3 *For nullspace consistent SBP operators that are p th order accurate with $p \geq 1$ and a diagonal norm matrix M , the Runge-Kutta method associated to the SBP time integration scheme (76) has an order of accuracy of at least $2p$.*

Proof This proof is very similar to the one of Theorem 3.21. By construction, the simplifying conditions $B(2p)$ and $D(p)$ are satisfied. Hence, the order of accuracy is at least $2p$ if $C(p - 1)$ is satisfied.

Again, suffices to consider a scaled SBP operator such that $T = 1, \mathbf{t}_L^T \boldsymbol{\tau} = 0, \mathbf{t}_R^T \boldsymbol{\tau} = 1$. Then, inserting the Runge-Kutta coefficients (76), $C(p - 1)$ is satisfied if for all $q \in \{1, \dots, p - 1\}$

$$M^{-1} F^T \tilde{J} M \boldsymbol{\tau}^{q-1} = \frac{1}{q} \boldsymbol{\tau}^q. \tag{79}$$

This equation is satisfied if and only if for all $\mathbf{u} \in \mathbb{R}^s$

$$\mathbf{u}^T F^T \tilde{J}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{u}^T M \boldsymbol{\tau}^q = 0. \tag{80}$$

Every \mathbf{u} can be written as $\mathbf{u} = D\mathbf{v} + \alpha\mathbf{o}$, where $\mathbf{t}_R^T \mathbf{v} = 0$ and $\alpha \in \mathbb{R}$, since D is nullspace consistent. Hence, it suffices to consider

$$\begin{aligned} & \mathbf{v}^T D^T F^T \tilde{J}^T M \boldsymbol{\tau}^{q-1} \alpha \mathbf{o}^T F^T \tilde{J}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{v}^T D^T M \boldsymbol{\tau}^q - \frac{1}{q} \alpha \mathbf{o}^T M \boldsymbol{\tau}^q \\ &= -\mathbf{v}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{v}^T D^T M \boldsymbol{\tau}^q \\ &= -\mathbf{v}^T M \boldsymbol{\tau}^{q-1} - \frac{1}{q} \mathbf{v}^T (\mathbf{t}_R \mathbf{t}_R^T - \mathbf{t}_L \mathbf{t}_L^T - MD) \boldsymbol{\tau}^q \\ &= -\mathbf{v}^T M \boldsymbol{\tau}^{q-1} \frac{1}{q} \mathbf{v}^T - MD \boldsymbol{\tau}^q = 0. \end{aligned} \tag{81}$$

This proves $C(p - 1)$. □

Remark C.4 Up to now, it is unclear whether the classical collocation Runge-Kutta schemes on Gauss nodes can be constructed as special members of a family of schemes that can be formulated for (more) general SBP operators. The problem seems to be that SBP schemes rely on differentiation, while the conditions $C(s)$ and $D(s)$ describing the Runge-Kutta methods rely on integration. Thus, special compatibility conditions as in the proof of Theorem 3.17 are necessary. To the authors' knowledge, no insights in this direction have been achieved for Gauss methods.

D Butcher Coefficients of some Finite Difference SBP Methods

Here, we provide the Butcher coefficients given by Theorem 3.15 for some of the finite difference SBP methods used in the numerical experiments in Sect. 4.

– Interior order 2, $N = 3$ nodes

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 3/8 & 1/4 & -1/8 \\ 1/4 & 1/2 & 1/4 \end{pmatrix}, \quad b = \begin{pmatrix} 1/4 \\ 1/2 \\ 1/4 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 1/2 \\ 1 \end{pmatrix} \tag{82}$$

– Interior order 2, $N = 9$ nodes

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 15/128 & 1/64 & -1/64 & 1/64 & -1/64 & 1/64 & -1/64 & 1/64 & -1/128 \\ 1/64 & 7/32 & 1/32 & -1/32 & 1/32 & -1/32 & 1/32 & -1/32 & 1/64 \\ 13/128 & 3/64 & 13/64 & 3/64 & -3/64 & 3/64 & -3/64 & 3/64 & -3/128 \\ 1/32 & 3/16 & 1/16 & 3/16 & 1/16 & -1/16 & 1/16 & -1/16 & 1/32 \\ 11/128 & 5/64 & 11/64 & 5/64 & 11/64 & 5/64 & -5/64 & 5/64 & -5/128 \\ 3/64 & 5/32 & 3/32 & 5/32 & 3/32 & 5/32 & 3/32 & -3/32 & 3/64 \\ 9/128 & 7/64 & 9/64 & 7/64 & 9/64 & 7/64 & 9/64 & 7/64 & -7/128 \\ 1/16 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/8 & 1/16 \end{pmatrix}, \quad b = \begin{pmatrix} 1/16 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/8 \\ 1/16 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 1/8 \\ 1/4 \\ 3/8 \\ 1/2 \\ 5/8 \\ 3/4 \\ 7/8 \\ 1 \end{pmatrix} \tag{83}$$

– Interior order 4, $N = 9$ nodes

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 13/180 & 18/385 & 1/2044 & 2/211 & -3/371 & 1/124 & -3/317 & 2/215 & -1/267 \\ 5/434 & 60/271 & 3/103 & -7/283 & 3/118 & -7/283 & 3/103 & -20/699 & 5/434 \\ 17/265 & 37/361 & 37/228 & 13/230 & -4/157 & 7/244 & -7/211 & 3/92 & -4/305 \\ 11/408 & 11/56 & 47/689 & 99/614 & 1/16 & -11/327 & 13/297 & -8/187 & 5/289 \\ 7/122 & 42/347 & 109/751 & 37/374 & 31/206 & 29/408 & -8/159 & 20/391 & -7/352 \\ 15/458 & 39/214 & 29/350 & 39/256 & 24/241 & 39/256 & 29/350 & -21/310 & 15/458 \\ 23/479 & 55/381 & 17/140 & 41/343 & 35/263 & 43/364 & 32/287 & 31/290 & -9/322 \\ 12/271 & 57/371 & 43/384 & 43/337 & 1/8 & 43/337 & 43/384 & 57/371 & 12/271 \end{pmatrix},$$

$$b = \begin{pmatrix} 17/384 \\ 59/384 \\ 43/384 \\ 49/384 \\ 1/8 \\ 49/384 \\ 43/384 \\ 59/384 \\ 17/384 \end{pmatrix}, \quad c = \begin{pmatrix} 0 \\ 1/8 \\ 1/4 \\ 3/8 \\ 1/2 \\ 5/8 \\ 3/4 \\ 7/8 \\ 1 \end{pmatrix} \quad (84)$$

References

1. Bezanson, J., Edelman, A., Karpinski, S., Shah, V.B.: Julia: a fresh approach to numerical computing. *SIAM Rev.* **59**(1), 65–98 (2017). <https://doi.org/10.1137/141000671>. [arxiv:1411.1607](https://arxiv.org/abs/1411.1607) [cs.MS]
2. Boom, P.D., Zingg, D.W.: High-order implicit time-marching methods based on generalized summation-by-parts operators. *SIAM J. Sci. Comput.* **37**(6), A2682–A2709 (2015). <https://doi.org/10.1137/15M1014917>
3. Butcher, J.: Implicit Runge-Kutta processes. *Math. Comput.* **18**(85), 50–64 (1964). <https://doi.org/10.1090/S0025-5718-1964-0159424-9>
4. Butcher, J.C.: Numerical methods for ordinary differential equations. Wiley, Chichester (2016)
5. Carpenter, M.H., Gottlieb, D., Abarbanel, S.: Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes. *J. Comput. Phys.* **111**(2), 220–236 (1994). <https://doi.org/10.1006/jcph.1994.1057>
6. Carpenter, M.H., Nordström, J., Gottlieb, D.: A stable and conservative interface treatment of arbitrary spatial accuracy. *J. Comput. Phys.* **148**(2), 341–365 (1999). <https://doi.org/10.1006/jcph.1998.6114>
7. Fernández, D.C.D.R., Boom, P.D., Zingg, D.W.: A generalized framework for nodal first derivative summation-by-parts operators. *J. Comput. Phys.* **266**, 214–239 (2014). <https://doi.org/10.1016/j.jcp.2014.01.038>
8. Fernández, D.C.D.R., Hicken, J.E., Zingg, D.W.: Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations. *Comput. Fluid* **95**, 171–196 (2014). <https://doi.org/10.1016/j.compfluid.2014.02.016>
9. Gassner, G.J.: A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods. *SIAM J. Sci. Comput.* **35**(3), A1233–A1253 (2013). <https://doi.org/10.1137/120890144>
10. Glaubitz, J., Öffner, P., Ranocha, H., Sonar, T.: Artificial viscosity for correction procedure via reconstruction using summation-by-parts operators. In: C. Klingenberg, M. Westdickenberg (eds.) *Theory, numerics and applications of hyperbolic problems II*, *Springer Proceedings in Mathematics & Statistics*, vol. 237, pp. 363–375. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-91548-7_28

11. Hairer, E., Nørsett, S.P., Wanner, G.: Solving ordinary differential equations I: nonstiff problems, Springer series in computational mathematics. Springer-Verlag, Berlin Heidelberg (2008)
12. Hairer, E., Wanner, G.: Solving ordinary differential equations II: stiff and differential-algebraic problems, Springer series in computational mathematics. Springer-Verlag, Berlin Heidelberg (2010)
13. Hicken, J.E., Zingg, D.W.: Summation-by-parts operators and high-order quadrature. *J. Comput. Appl. Math.* **237**(1), 111–125 (2013). <https://doi.org/10.1016/j.cam.2012.07.015>
14. Hunter, J.D.: Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**(3), 90–95 (2007). <https://doi.org/10.1109/MCSE.2007.55>
15. Ketcheson, D.I.: Relaxation Runge-Kutta methods: conservation and stability for inner-product norms. *SIAM J. Num. Anal.* **57**(6), 2850–2870 (2019). <https://doi.org/10.1137/19M1263662>. [arxiv:1905.09847](https://arxiv.org/abs/1905.09847) [math.NA]
16. Kreiss, H.O., Scherer, G.: Finite element and finite difference methods for hyperbolic partial differential equations. In: de Boor, C. (ed.) *Mathematical aspects of finite elements in partial differential equations*, pp. 195–212. Academic Press, New York (1974)
17. Linders, V., Lundquist, T., Nordström, J.: On the order of accuracy of finite difference operators on diagonal norm based summation-by-parts form. *SIAM J. Num. Anal.* **56**(2), 1048–1063 (2018). <https://doi.org/10.1137/17M1139333>
18. Linders, V., Nordström, J., Frankel, S.H.: Properties of Runge-Kutta-summation-by-parts methods. *J. Comput. Phys.* **419**, 109,684 (2020). <https://doi.org/10.1016/j.jcp.2020.109684>
19. Lundquist, T., Nordström, J.: The SBP-SAT technique for initial value problems. *J. Comput. Phys.* **270**, 86–104 (2014). <https://doi.org/10.1016/j.jcp.2014.03.048>
20. Mattsson, K.: Diagonal-norm summation by parts operators for finite difference approximations of third and fourth derivatives. *J. Comput. Phys.* **274**, 432–454 (2014). <https://doi.org/10.1016/j.jcp.2014.06.027>
21. Mattsson, K., Nordström, J.: Summation by parts operators for finite difference approximations of second derivatives. *J. Comput. Phys.* **199**(2), 503–540 (2004). <https://doi.org/10.1016/j.jcp.2004.03.001>
22. Mattsson, K., Olsson, P.: An improved projection method. *J. Comput. Phys.* **372**, 349–372 (2018). <https://doi.org/10.1016/j.jcp.2018.06.030>
23. Nordström, J., Björck, M.: Finite volume approximations and strict stability for hyperbolic problems. *Appl. Num. Math.* **38**(3), 237–255 (2001). [https://doi.org/10.1016/S0168-9274\(01\)00027-7](https://doi.org/10.1016/S0168-9274(01)00027-7)
24. Nordström, J., Forsberg, K., Adamsson, C., Eliasson, P.: Finite volume methods, unstructured meshes and strict stability for hyperbolic problems. *Appl. Num. Math.* **45**(4), 453–473 (2003). [https://doi.org/10.1016/S0168-9274\(02\)00239-8](https://doi.org/10.1016/S0168-9274(02)00239-8)
25. Nordström, J., Lundquist, T.: Summation-by-parts in time. *J. Comput. Phys.* **251**, 487–499 (2013). <https://doi.org/10.1016/j.jcp.2013.05.042>
26. Olsson, P.: Summation by parts, projections, and stability. I. *Math. Comput.* **64**(211), 1035–1065 (1995). <https://doi.org/10.1090/S0025-5718-1995-1297474-X>
27. Olsson, P.: Summation by parts, projections, and stability. II. *Math. Comput.* **64**(212), 1473–1493 (1995). <https://doi.org/10.1090/S0025-5718-1995-1308459-9>
28. Prothero, A., Robinson, A.: On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations. *Math. Comput.* **28**(125), 145–162 (1974). <https://doi.org/10.1090/S0025-5718-1974-0331793-2>
29. Ranocha, H.: Some notes on summation by parts time integration methods. *Result Appl. Math.* **1**, 100,004 (2019). <https://doi.org/10.1016/j.rinam.2019.100004>. [arxiv:1901.08377](https://arxiv.org/abs/1901.08377) [math.NA]
30. Ranocha, H.: On strong stability of explicit Runge-Kutta methods for nonlinear semibounded operators. *IMA J. Num. Anal.* (2020). <https://doi.org/10.1093/imanum/drz070>
31. Ranocha, H., Ketcheson, D.I.: Energy stability of explicit Runge-Kutta methods for nonautonomous or nonlinear problems. *SIAM J. Num. Anal.* **58**(6), 3382–3405 (2020). <https://doi.org/10.1137/19M1290346>. [arxiv:1909.13215](https://arxiv.org/abs/1909.13215) [math.NA]
32. Ranocha, H., Ketcheson, D.I.: Relaxation Runge-Kutta methods for Hamiltonian problems. *J. Sci. Comput.* (2020). <https://doi.org/10.1007/s10915-020-01277-y>
33. Ranocha, H., Lóczy, L., Ketcheson, D.I.: General relaxation methods for initial-value problems with application to multistep schemes. *Numer. Math.* **146**, 875–906 (2020). <https://doi.org/10.1007/s00211-020-01158-4>. [arxiv:2003.03012](https://arxiv.org/abs/2003.03012) [math.NA]
34. Ranocha, H., Mitsotakis, D., Ketcheson, D.I.: A broad class of conservative numerical methods for dispersive wave equations (2020). Accepted in communications in computational physics. [arxiv:2006.14802](https://arxiv.org/abs/2006.14802) [math.NA]
35. Ranocha, H., Nordström, J.: SBP-projection-in-time-notebooks. A new class of A stable summation by parts time integration schemes with strong initial conditions. <https://github.com/ranocha/SBP-projection-in-time-notebooks> (2020). <https://doi.org/10.5281/zenodo.3699173>

36. Ranocha, H., Öffner, P.: L_2 stability of explicit Runge-Kutta schemes. *J. Sci. Comput.* **75**(2), 1040–1056 (2018). <https://doi.org/10.1007/s10915-017-0595-4>
37. Ranocha, H., Öffner, P., Sonar, T.: Summation-by-parts operators for correction procedure via reconstruction. *J. Comput. Phys.* **311**, 299–328 (2016). <https://doi.org/10.1016/j.jcp.2016.02.009>. arxiv:1511.02052 [math.NA]
38. Ranocha, H., Ostaszewski, K., Heinisch, P.: Discrete vector calculus and Helmholtz Hodge decomposition for classical finite difference summation by parts operators. *Commun. Appl. Math. Comput.* (2020). <https://doi.org/10.1007/s42967-019-00057-2>
39. Ranocha, H., Sayyari, M., Dalcin, L., Parsani, M., Ketcheson, D.I.: Relaxation Runge-Kutta methods: fully-discrete explicit entropy-stable schemes for the compressible Euler and Navier-Stokes equations. *SIAM J. Sci. Comput.* **42**(2), A612–A638 (2020). <https://doi.org/10.1137/19M1263480>. arxiv:1905.09129 [math.NA]
40. Roman, S.: *Advanced linear algebra*. Springer, New York (2008). <https://doi.org/10.1007/978-0-387-72831-5>
41. Ruggiu, A.A., Nordström, J.: On pseudo-spectral time discretizations in summation-by-parts form. *J. Comput. Phys.* **360**, 192–201 (2018). <https://doi.org/10.1016/j.jcp.2018.01.043>
42. Ruggiu, A.A., Nordström, J.: Eigenvalue analysis for summation-by-parts finite difference time discretizations. *SIAM J. Num. Anal.* **58**(2), 907–928 (2020). <https://doi.org/10.1137/19M1256294>
43. Strand, B.: Summation by parts for finite difference approximations for d/dx . *J. Comput. Phys.* **110**(1), 47–67 (1994). <https://doi.org/10.1006/jcph.1994.1005>
44. Sun, Z., Shu, C.W.: Stability of the fourth order Runge-Kutta method for time-dependent partial differential equations. *Annal Math. Sci. Appl.* **2**(2), 255–284 (2017). <https://doi.org/10.4310/AMSA.2017.v2.n2.a3>
45. Sun, Z., Shu, C.W.: Enforcing strong stability of explicit Runge-Kutta methods with superviscosity (2019). arxiv:1912.11596 [math.NA]
46. Sun, Z., Shu, C.W.: Strong stability of explicit Runge-Kutta time discretizations. *SIAM J. Num. Anal.* **57**(3), 1158–1182 (2019). <https://doi.org/10.1137/18M122892X>. arxiv:1811.10680 [math.NA]
47. Svärd, M., Nordström, J.: Review of summation-by-parts schemes for initial-boundary-value problems. *J. Comput. Phys.* **268**, 17–38 (2014). <https://doi.org/10.1016/j.jcp.2014.02.031>
48. Svärd, M., Nordström, J.: On the convergence rates of energy-stable finite-difference schemes. *J. Comput. Phys.* **397**, 108,819 (2019). <https://doi.org/10.1016/j.jcp.2019.07.018>
49. Svärd, M., Nordström, J.: Convergence of energy stable finite-difference schemes with interfaces. *J. Comput. Phys.* (2020). <https://doi.org/10.1016/j.jcp.2020.110020>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.