Journal of Computational Physics 344 (2017) 451–464

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

On conservation and stability properties for summation-by-parts schemes

Jan Nordström *, Andrea A. Ruggiu

Department of Mathematics, Computational Mathematics, Linköping University, SE-581 83 Linköping, Sweden

ARTICLE INFO

Article history:
Received 26 December 2016
Received in revised form 1 May 2017
Accepted 2 May 2017
Available online 15 May 2017

Keywords:
Hyperbolic problems
Summation-by-parts
Boundary conditions
Interface conditions
Stability
Conservation

ABSTRACT

We discuss conservative and stable numerical approximations in summation-by-parts form for linear hyperbolic problems with variable coefficients. An extended setting, where the boundary or interface may or may not be included in the grid, is considered. We prove that conservative and stable formulations for variable coefficient problems require a boundary and interface conforming grid and exact numerical mimicking of integration-by-parts. Finally, we comment on how the conclusions from the linear analysis carry over to the nonlinear setting.

1. Introduction

High order methods for partial differential equations provide accurate numerical solutions with a limited computational effort [1]. The main requirement for this efficiency is that stable and accurate implementations of boundary and interface conditions exist. Summation-By-Parts (SBP) operators [2,3], together with a weak imposition of the boundary and interface conditions through Simultaneous-Approximation-Term (SAT) techniques [4–6], meet this challenge for finite difference methods. The SBP-SAT technique also enables generalizations to curvilinear domains [7–9] and multi-block schemes [10,11]. In addition to finite difference methods, other discretization techniques such as discontinuous Galerkin [12,13], finite volume [14,15] and pseudo-spectral methods [16,17] can be enclosed in the SBP-SAT framework.

An extension of the SBP-SAT technique was presented in [18], where it was shown that approximations of the first derivative on SBP form can be obtained starting from a quadrature rule. The authors proceed to show that this fact generalizes the construction of SBP operators to grids which do not include the domain boundaries. These operators are referred to as Generalized Summation-By-Parts (GSBP) operators.

In this paper, our aim is to compare conservation and stability properties of SBP and GSBP based approximations. We will analyze a linear scalar conservation law with a spatially varying coefficient on single and multiple domains. This model problem can be seen as a building block for more demanding nonlinear cases. The conditions for discrete conservation and stability of SBP-SAT formulations will be specified in detail. The results limit the use of GSBP formulations as general building blocks in schemes, and stress the need for exact numerical mimicking of integration-by-parts.

The article proceeds as follows. Section 2 deals with energy boundedness and conservation for a model problem with a variable coefficient. In section 3, the SBP and GSBP operators are introduced. In section 4 and 5 we present SBP-SAT and

* Corresponding author.
E-mail addresses: jan.nordstrom@liu.se (J. Nordström), andrea.ruggiu@liu.se (A.A. Ruggiu).

http://dx.doi.org/10.1016/j.jcp.2017.05.002
0021-9991/© 2017 Elsevier Inc. All rights reserved.

GSBP-SAT single-domain approximations and determine the requirements for discrete conservation and stability. In section 6 and 7, extensions to multi-domain approximations are considered. Section 8 describes a slight modification of the current GSBP operators. Finally, in section 9 we discuss the implications of the linear analysis on nonlinear problems. Conclusions are drawn in section 10.

2. The continuous problem, conservation and energy boundedness

Consider the following initial–boundary value problem on conservation form

$$\begin{aligned} u_t + f_x &= 0, & \alpha < x < \beta, \quad & t > 0, \\ u &= h(x), & \alpha < x < \beta, \quad & t = 0, \\ B_\alpha u &= g_\alpha(t), & x = \alpha, \quad & t > 0, \\ B_\beta u &= g_\beta(t), & x = \beta, \quad & t > 0, \end{aligned} \tag{1}$$

where u, $f = f(u, x)$, B_α, B_β is the solution, flux function, left boundary operator and right boundary operator respectively. The initial data h and boundary data g_α, g_β are collectively referred to as the data of the problem.

For two real-valued functions v and w, we define a scalar product and norm in $L^2(\alpha, \beta)$

$$(v, w)_2 = \int_\alpha^\beta v w \, \mathrm{d}x, \qquad \|v\|_2 = \sqrt{(v, v)_2}.$$

This formalism allows us to introduce the concepts of energy boundedness and conservation [19].

Definition 2.1. The problem (1) is said to be energy-bounded if the estimate

$$\|u(\cdot, t)\|_2^2 \le K(t) \left[\|h\|_2^2 + \int_0^t g_\alpha(\tau)^2 + g_\beta(\tau)^2 \mathrm{d}\tau \right] \tag{2}$$

holds with $K(t)$ independent of g_α, g_β and h, and bounded for finite times.

Having defined energy boundedness, we next define conservation.

Definition 2.2. The problem (1) is in conservative form since

$$\frac{\mathrm{d}}{\mathrm{d}t}(1, u)_2 = f(u(\alpha, t), \alpha) - f(u(\beta, t), \beta)$$

holds. The integral of u changes only by the flux through the boundaries.

2.1. The model problem

To illustrate most of our points in this paper, it is sufficient to consider the following linear variable coefficient advection problem

$$\begin{aligned} u_t + (au)_x &= 0, & \alpha < x < \beta, \quad & t > 0, \\ u(x, 0) &= h(x), & \alpha < x < \beta, & \\ u(\alpha, t) &= g_\alpha(t), & t > 0, & \end{aligned} \tag{3}$$

where $a = a(x)$ and $a(\alpha) > 0$, $a(\beta) > 0$ is assumed. The problem is in conservative form since integration yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(1, u)_2 = a(\alpha) g_\alpha(t) - a(\beta) u(\beta, t), \tag{4}$$

where the Dirichlet boundary condition has been imposed at $x = \alpha$.

Energy-boundedness of the solution in terms of the data follows by applying the energy-method to (3) (multiplying by u and integrating over $[\alpha, \beta]$). The Integration-By-Parts (IBP) rule

$$(v, w_x)_2 = v(\beta) w(\beta) - v(\alpha) w(\alpha) - (v_x, w)_2 \tag{5}$$

leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\|u\|_2^2 = a(\alpha) g_\alpha^2(t) - a(\beta) u^2(\beta, t) - (u, a_x u)_2. \tag{6}$$

The energy-rate (6), in combination with the assumption $a_x \in L^\infty(\alpha,\beta)$, leads to an energy estimate for (3), including a limited exponential growth given by

$$|(u, a_x u)_2| \leq \|a_x\|_\infty \|u\|_2^2, \qquad \|a_x\|_\infty = \max_{x\in[\alpha,\beta]} |a_x(x)|. \tag{7}$$

In particular, we find

$$\|u(\cdot,t)\|_2^2 \leq e^{\|a_x\|_\infty t}\left[\|h\|_2^2 + \int_0^t e^{-\|a_x\|_\infty \tau}\left(a(\alpha)g_\alpha^2(\tau) - a(\beta)u^2(\beta,\tau)\right)\mathrm{d}\tau\right],$$

which is of the form (2) in Definition 2.1.

3. Standard and generalized Summation-By-Parts discretizations

Here, we define the discrete operators which mimic the IBP rule (5).

3.1. Summation-By-Parts operators

Consider the discrete grid $\mathbf{x} = [x_0, \ldots, x_N]^T$, with the ordering of nodes $\alpha = x_0 < \cdots < x_N = \beta$. Furthermore, let the spatial derivative of a function φ be approximated through the matrix D, i.e. $\varphi_x \approx D\boldsymbol{\varphi}$, with $\boldsymbol{\varphi} = [\varphi(x_0), \ldots, \varphi(x_N)]^T$.

Definition 3.1. An operator D is a qth order accurate approximation of the first derivative on SBP form if

i) $D\mathbf{x}^j = P^{-1}Q\mathbf{x}^j = j\mathbf{x}^{j-1}$, $j \in [0,q]$,
ii) P is a symmetric positive definite matrix,
iii) $Q + Q^T = \mathbf{e}_\beta \mathbf{e}_\beta^T - \mathbf{e}_\alpha \mathbf{e}_\alpha^T$, where $\mathbf{e}_\alpha = [1, 0, \ldots, 0]^T$ and $\mathbf{e}_\beta = [0, \ldots, 0, 1]^T$.

Condition i) in Definition 3.1 implies that the operator D exactly mimics the first derivative for the grid monomials $\mathbf{x}^j = [x_0^j, \ldots, x_N^j]^T$ up to the qth order. The matrix P in condition ii) defines a discrete scalar product and norm

$$(\mathbf{v},\mathbf{w})_P = \mathbf{v}^T P\mathbf{w}, \qquad \|\mathbf{v}\|_P = \sqrt{(\mathbf{v},\mathbf{v})_P}.$$

To avoid well known stability issues for variable coefficients and nonlinear problems [20–22], we consider P to be diagonal in the remainder of this paper. Finally, condition iii) ensures that D mimics the IBP rule (5)

$$(\mathbf{v}, D\mathbf{w})_P = v_N w_N - v_0 w_0 - (D\mathbf{v}, \mathbf{w})_P. \tag{8}$$

Remark 3.2. SBP operators were originally developed for finite difference methods on equidistant grids [2]. It is also possible to build SBP operators starting from any given quadrature rule, e.g. Gauss–Lobatto [12,18].

3.2. Generalized Summation-By-Parts operators

A first generalization of SBP operators was given in [23,24], where condition iii) in Definition 3.1 was modified by introducing an almost skew-symmetric matrix Q with the exception of $(k+1)\times(k+1)$ large boundary blocks. The definition of SBP operators can also be extended to non-uniform discrete grids $\mathbf{x} = [x_0, \ldots, x_N]^T$ that do not include one or both boundary nodes [18].

Definition 3.3. An operator D is a qth order accurate approximation of the first derivative with the Generalized SBP (GSBP) property if

i) $D\mathbf{x}^j = P^{-1}Q\mathbf{x}^j = j\mathbf{x}^{j-1}$, $j \in [0,q]$,
ii) P is a symmetric positive definite matrix,
iii) $Q + Q^T = E$, where $(\mathbf{x}^i)^T E\mathbf{x}^j = \beta^{i+j} - \alpha^{i+j}$, $i, j = 0, \ldots, r$, $r \geq q$.

Remark 3.4. For standard SBP operators in Definition 3.1, the matrix $E = Q + Q^T$ is such that $(\mathbf{x}^i)^T E\mathbf{x}^j = \beta^{i+j} - \alpha^{i+j}$, $\forall i, j \in \mathbb{N}$. Consequently, SBP operators can be seen as particular GSBP operators with α, β being nodes on the grid. To avoid ambiguity, henceforth D will be called a GSBP operator if one or both boundary nodes are excluded from the discrete grid. Operators with both boundary nodes included are called SBP operators.

The GSBP operators can be constructed from a quadrature rule and may have a non-repeating interior stencil. As an example, consider $(\alpha,\beta)=(-1,1)$ and the Legendre–Gauss quadrature on the three-point grid $\mathbf{x}=[-\sqrt{15}/5, 0, \sqrt{15}/5]^T$. We find

$$P=\frac{1}{9}\begin{bmatrix}5 & 0 & 0\\ 0 & 8 & 0\\ 0 & 0 & 5\end{bmatrix}, \quad Q=\frac{\sqrt{15}}{54}\begin{bmatrix}-15 & 20 & -5\\ -8 & 0 & 8\\ 5 & -20 & 15\end{bmatrix}.$$

It is easy to verify that the GSBP operator $D=P^{-1}Q$ exactly differentiates second degree polynomials.

The matrix E in Definition 3.3 can be written in terms of boundary interpolants of degree r. In [18], $\mathbf{t}_\alpha$ and $\mathbf{t}_\beta$ were introduced such that

$$\mathbf{t}_\phi^T\mathbf{u}\approx u(\phi): \quad \mathbf{t}_\phi^T\mathbf{x}^j=\phi^j, \quad j\in[0,r], \quad \phi\in\{\alpha,\beta\}. \tag{9}$$

This gives rise to $E=\mathbf{t}_\beta\mathbf{t}_\beta^T-\mathbf{t}_\alpha\mathbf{t}_\alpha^T$ and to the GSBP property

$$(\mathbf{v}, D\mathbf{w})_P=(\mathbf{t}_\beta^T\mathbf{v})(\mathbf{t}_\beta^T\mathbf{w})-(\mathbf{t}_\alpha^T\mathbf{v})(\mathbf{t}_\alpha^T\mathbf{w})-(D\mathbf{v},\mathbf{w})_P. \tag{10}$$

In the example above, the interpolants are $\mathbf{t}_\alpha=[(5+\sqrt{15})/6, -2/3, (5-\sqrt{15})/6]^T$ and $\mathbf{t}_\beta=[(5-\sqrt{15})/6, -2/3, (5+\sqrt{15})/6]^T$, with $r=2$.

Remark 3.5. In the SBP case, we have $\mathbf{t}_\alpha=\mathbf{e}_\alpha$, $\mathbf{t}_\beta=\mathbf{e}_\beta$ and $r=\infty$.

4. Conservation of single-domain discretizations

Our general goal is to find an approximation of the model problem (3) which is both conservative and stable. Consider the standard SBP operators in Definition 3.1. A straightforward semi-discrete approximation of (3) with the SBP-SAT technique [5] gives

$$\mathbf{u}_t+P^{-1}QA\mathbf{u}=\sigma_\alpha AP^{-1}\mathbf{e}_\alpha(\mathbf{e}_\alpha^T\mathbf{u}-g_\alpha(t)), \tag{11}$$

where $A=\operatorname{diag}(a(x_0),\ldots,a(x_N))$. Let $\sigma_\alpha=-1$, $\mathbf{1}=(1,1,\ldots,1,1)^T$, multiply (11) from the left by $\mathbf{1}^TP$, and use $Q=\mathbf{e}_\beta\mathbf{e}_\beta^T-\mathbf{e}_\alpha\mathbf{e}_\alpha^T-Q^T$. We find

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},u)_P=a(\alpha)g_\alpha(t)-a(\beta)u_N, \tag{12}$$

which mimics the continuous conservation relation (4) perfectly.

Definition 4.1. A numerical scheme discretizing the model problem (3) is said to be conservative if (12) holds.

4.1. The conventional skew-symmetric SBP approximation

In [21] it was shown that the formulation (11) leads to stability problems. To overcome this issue, we split the continuous spatial operator into a symmetric and anti-symmetric part

$$(au)_x=\frac{1}{2}(au)_x+\frac{1}{2}au_x+\frac{1}{2}a_xu. \tag{13}$$

The first two terms constitute the anti-symmetric part, while the third one forms the symmetric portion [21].

By using (13), an SBP-SAT discretization of (3) in skew-symmetric form can be written as

$$\mathbf{u}_t+\frac{1}{2}P^{-1}(QA+AQ)\mathbf{u}+\frac{1}{2}A_x\mathbf{u}=\sigma_\alpha AP^{-1}\mathbf{e}_\alpha(\mathbf{e}_\alpha^T\mathbf{u}-g_\alpha(t)), \tag{14}$$

where A_x is a matrix which consistently represents a_x. Let for now $A_x=\operatorname{diag}(a_x(x_0),\ldots,a_x(x_N))$. The formulation (14) is energy-stable [21], but not conservative, since for $\sigma_\alpha=-1$ the conservation relation becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},\mathbf{u})_P=a(\alpha)g_\alpha(t)-a(\beta)u_N+\frac{1}{2}(\mathbf{1},(P^{-1}AQ^T-A_x)\mathbf{u})_P.$$

By comparing with (12), we realize that a conservative formulation requires that $A_x=P^{-1}AQ^T$. However, this does not lead to a consistent representation of a_x since

$$\begin{aligned}A_x\mathbf{1} &= P^{-1}AQ^T\mathbf{1} = P^{-1}A(\mathbf{e}_\beta\mathbf{e}_\beta^T - \mathbf{e}_\alpha\mathbf{e}_\alpha^T - Q)\mathbf{1} = P^{-1}A(\mathbf{e}_\beta - \mathbf{e}_\alpha)\\ &= \left[-P_{00}^{-1}a(x_0), 0, \dots, 0, P_{NN}^{-1}a(x_N)\right]^T \neq [a_x(x_0), \dots, a_x(x_N)]^T .\end{aligned}$$

We have proved

Proposition 4.2. *The conventional skew-symmetric approximation (14) of (3), can not be both consistent and conservative.*

Consequently, an alternative formulation should be considered.

4.2. Conservation for the skew-symmetric SBP approximation

We start by introducing an alternative way to write a vector [25].

Definition 4.3. Let $\boldsymbol{\phi} = [\phi_0, \dots, \phi_N]^T$ be a general vector. We denote with the capital letter Φ the matrix $\Phi = \text{diag}\,(\phi_0, \dots, \phi_N)$ such that $\boldsymbol{\phi} = \Phi\mathbf{1}$.

Next, let $U = \text{diag}(u_0, \dots, u_N)$ and $\mathbf{a}$ be the grid function representing $a(x)$ on $\mathbf{x}$. The symmetric part of (13) can now be consistently represented by $UD\mathbf{a} \approx a_x u$ and the semi-discrete SBP-SAT approximation becomes

$$\mathbf{u}_t + \frac{1}{2}P^{-1}(QA + AQ)\mathbf{u} + \frac{1}{2}UD\mathbf{a} = \sigma_\alpha AP^{-1}\mathbf{e}_\alpha(\mathbf{e}_\alpha^T\mathbf{u} - g_\alpha(t)). \tag{15}$$

This formulation leads to the following result.

Proposition 4.4. *The discretization (15) of (3) with a general $a(x)$ using the SBP operators in Definition 3.1 is conservative for $\sigma_\alpha = -1$.*

Proof. Consider

$$\mathbf{u}_t + \frac{1}{2}P^{-1}(QAU + AQU + UQA)\mathbf{1} = \sigma_\alpha AP^{-1}\mathbf{e}_\alpha\left(\mathbf{e}_\alpha^T\mathbf{u} - g_\alpha(t)\right), \tag{16}$$

which is equivalent to (15) by Definition 4.3. By multiplying (16) from the left with $\mathbf{1}^TP$ we find

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1}, \mathbf{u})_P + \frac{1}{2}[\mathbf{1}^T(QAU + AQU + UQA)\mathbf{1}] = \sigma_\alpha(\mathbf{e}_\alpha^T\mathbf{a})(\mathbf{e}_\alpha^T\mathbf{u} - g_\alpha(t)). \tag{17}$$

The terms on the left-hand side in (17) can be rewritten by using $Q\mathbf{1} = \mathbf{0}$ and property iii) in Definition 3.1 as

$$\begin{aligned}\mathbf{1}^TQAU\mathbf{1} &= \mathbf{1}^T(\mathbf{e}_\beta\mathbf{e}_\beta^T - \mathbf{e}_\alpha\mathbf{e}_\alpha^T - Q^T)AU\mathbf{1} = a(\beta)u_N - a(\alpha)u_0,\\ \mathbf{1}^T(AQU + UQA)\mathbf{1} &= \mathbf{1}^TA(Q + Q^T)U\mathbf{1} = a(\beta)u_N - a(\alpha)u_0.\end{aligned}$$

Thus, the relation (17) becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1}, \mathbf{u})_P = a(\alpha)u_0 - a(\beta)u_N + \sigma_\alpha a(\alpha)(u_0 - g_\alpha(t)),$$

which exactly mimics the conservation relation (12) if $\sigma_\alpha = -1$. □

4.3. Conservation for the skew-symmetric GSBP approximation

The discretization of (3) using GSBP operators and the split in (13) is

$$\begin{aligned}\mathbf{u}_t + \frac{1}{2}P^{-1}(QA + AQ)\mathbf{u} + \frac{1}{2}UD\mathbf{a} = \sigma_\alpha^I AP^{-1}\mathbf{t}_\alpha\left(\mathbf{t}_\alpha^T\mathbf{u} - g_\alpha(t)\right)\\ + \sigma_\alpha^{II}P^{-1}\mathbf{t}_\alpha\left(\mathbf{t}_\alpha^TA\mathbf{u} - a(\alpha)g_\alpha(t)\right),\end{aligned} \tag{18}$$

where we use two penalty terms to weakly impose the boundary condition.

Following the proof of Proposition 4.4, the conservation relation associated to (18) becomes

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1}, \mathbf{u})_P = \left(\frac{1}{2} + \sigma_\alpha^I\right)(\mathbf{t}_\alpha^T\mathbf{a})(\mathbf{t}_\alpha^T\mathbf{u}) + \left(\frac{1}{2} + \sigma_\alpha^{II}\right)\mathbf{t}_\alpha^TA\mathbf{u} - \frac{1}{2}(\mathbf{t}_\beta^T\mathbf{a})(\mathbf{t}_\beta^T\mathbf{u})\\ - \frac{1}{2}\mathbf{t}_\beta^TA\mathbf{u} - \sigma_\alpha^I(\mathbf{t}_\alpha^T\mathbf{a})g_\alpha(t) - \sigma_\alpha^{II}a(\alpha)g_\alpha(t).\end{aligned}$$

By choosing $\sigma_\alpha^I = \sigma_\alpha^{II} = -1/2$ we find

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1}, \mathbf{u})_P = \frac{1}{2}\left[\mathbf{t}_\alpha^T \mathbf{a} + a(\alpha)\right] g_\alpha(t) - \frac{1}{2}\left[(\mathbf{t}_\beta^T \mathbf{a})(\mathbf{t}_\beta^T \mathbf{u}) + \mathbf{t}_\beta^T A\mathbf{u}\right]. \tag{19}$$

We say that the single domain formulation (18) is *weakly* conservative since (19) approximately mimics the conservation relation (12).

Remark 4.5. Letting either σ_α^I or σ_α^{II} be equal to zero does not lead to any form of conservation.

The choices $\sigma_\alpha^I = \sigma_\alpha^{II} = -1/2$ are optimal, and we have proved

Proposition 4.6. *The discretization (18) of (3) with a general $a(x)$ using the GSBP operators in Definition 3.3 is not conservative.*

5. Stability of single-domain discretizations

In concert with Definition 2.1, we define energy stability [19].

Definition 5.1. A semi-discrete approximation of (1) is energy-stable if

$$\|\mathbf{u}(t)\|_d^2 \le K_d(t)\left[\|\mathbf{h}\|_d^2 + \max_{\tau\in[0,t]}(g_\alpha(\tau)^2) + \max_{\tau\in[0,t]}(g_\beta(\tau)^2)\right] \tag{20}$$

holds. In (20), $\|\cdot\|_d$ denotes a suitable discrete norm and $\mathbf{u}$, $\mathbf{h}$ are grid functions representing u, h in (1), respectively. The function $K_d(t)$ is independent of g_α, g_β, $\mathbf{h}$ and bounded for finite times and all spatial mesh sizes.

Energy-stable discretizations for the model problem (3) should mimic the continuous energy-rate (6) and lead to an energy-estimate of the form (20).

5.1. Stability of SBP discretizations

We begin by considering the conservative SBP-SAT discretization (15) with $\sigma_\alpha = -1$. By multiplying (15) from the left with $\mathbf{u}^T P$ and using the SBP property (8) we find

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}\|_P^2 = a(\alpha)g_\alpha^2(t) - a(\beta)u_N^2 - (\mathbf{u}, UD\mathbf{a})_P - a(\alpha)(u_0 - g_\alpha(t))^2. \tag{21}$$

This energy-rate is similar to (6), except for an extra damping term. The relation (21), together with a discrete analogue of (7), i.e.

$$|(\mathbf{u}, UD\mathbf{a})_P| \le \|D\mathbf{a}\|_\infty \|\mathbf{u}\|_P^2, \qquad \|D\mathbf{a}\|_\infty = \max_{i=0,\dots,N}|(D\mathbf{a})_i|,$$

leads to an energy-estimate of the form (20) and proves

Proposition 5.2. *The discretization (15) of (3) with a general $a(x)$ using the SBP operators in Definition 3.1 is energy-stable.*

5.2. Stability of GSBP discretizations

Next, we move on to the GSBP-SAT approximation (18), which satisfies the weak conservation property (19) with $\sigma_\alpha^I = \sigma_\alpha^{II} = -1/2$. For simplicity, let $g_\alpha(t) = 0$. The discrete energy-method using the GSBP property (10), gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}\|_P^2 = -\mathbf{u}^T A\mathbf{t}_\alpha \mathbf{t}_\alpha^T \mathbf{u} - \mathbf{u}^T A\mathbf{t}_\beta \mathbf{t}_\beta^T \mathbf{u} - (\mathbf{u}, UD\mathbf{a})_P. \tag{22}$$

The relation (22) does not lead to a bound on $\|\mathbf{u}\|_P^2$ since the boundary terms are indefinite even if A is positive definite. We state the result as

Proposition 5.3. *The discretization (18) of (3) with a general $a(x)$ using the GSBP operators in Definition 3.3 is not energy-stable.*

Proof. The symmetric parts of the matrices $A\mathbf{t}_\alpha\mathbf{t}_\alpha^T$ and $A\mathbf{t}_\beta\mathbf{t}_\beta^T$ are indefinite for a general A. □

Remark 5.4. Proposition 5.3 holds independently of the penalty terms in (18), since $\mathbf{u}A\mathbf{t}_\beta\mathbf{t}_\beta^T\mathbf{u}$ in (22) may be negative. In the very special case with GSBP operators which include the right-boundary node, stability can be obtained by the choice $\sigma_\alpha^{II} = -1/2 - \sigma_\alpha^I$ in (18). However, with that choice, the weak conservation result (19) does not hold anymore.

As a first example, consider the Legendre–Gauss GSBP operators of order 2 introduced previously and the matrix $A = \operatorname{diag}(1, 1, 2)$. In this case both symmetric parts of the matrices $A\mathbf{t}_\alpha\mathbf{t}_\alpha^T$ and $A\mathbf{t}_\beta\mathbf{t}_\beta^T$ (indicated with the superscript S below) are indefinite, as can be seen from their eigenvalues

$$\lambda\left((A\mathbf{t}_\alpha\mathbf{t}_\alpha^T)^S\right) \in \{-8.5631\cdot 10^{-3}, 0, 2.7105\},$$
$$\lambda\left((A\mathbf{t}_\beta\mathbf{t}_\beta^T)^S\right) \in \{-5.3450\cdot 10^{-2}, 0, 4.9071\}.$$

As a consequence, the GSBP approximation (18) is not stable.

As a second example, consider the domain $(\alpha, \beta) = (-1, 1)$ and the GSBP operators based on the Legendre–Gauss–Radau quadrature on $\mathbf{x} = [-1, (1-\sqrt{6})/5, (1+\sqrt{6})/5]^T$ with the Lagrange interpolants $\mathbf{t}_\alpha = [1, 0, 0]^T$ and $\mathbf{t}_\beta = [1/3, (2-3\sqrt{6})/6, (2+3\sqrt{6})/6]^T$, as in [18]. In this case only the left boundary node $x = \alpha$ is included on $\mathbf{x}$. If $A = \operatorname{diag}(1, 2, 4)$, the right boundary term is indefinite and we find

$$\lambda\left((A\mathbf{t}_\beta\mathbf{t}_\beta^T)^S\right) \in \{-2.1993\cdot 10^{-1}, 0, 11.631\},$$

which again leads to an unstable scheme.

6. Conservation for multi-domain discretizations

Multi-block schemes make the standard SBP-SAT technique more useful by allowing for more complex geometries and parallel computations [11,26,27]. For GSBP-SAT approximations, the extension to multi-elements becomes necessary even for model equations, since typically only a few nodes are used to construct the discrete operators. We treat the two-domain case and subsequently generalize to multi-domains.

6.1. Conservation for SBP multi-domain approximations

For clarity, the terms associated with the boundaries $\{\alpha, \beta\}$ will be neglected below, since the related boundary procedures are identical to the ones in section 4. Let x_I be a point in the interior of the domain $[\alpha, \beta]$ and define the discrete grids $\mathbf{x}_L \in \mathbb{R}^{N_L+1}$ and $\mathbf{x}_R \in \mathbb{R}^{N_R+1}$ on the subintervals $[\alpha, x_I]$ and $[x_I, \beta]$, respectively. Furthermore, consider $\mathbf{e}_{\alpha,L} = [1, 0, \ldots, 0]^T$, $\mathbf{e}_{x_I,L} = [0, \ldots, 0, 1]^T$ which exactly project grid functions on $\mathbf{x}_L$ to $x = \alpha$ and $x = x_I$, respectively. For the domain $\mathbf{x}_R$ the interpolants $\mathbf{e}_{x_I,R}$ and $\mathbf{e}_{\beta,R}$ are defined likewise. By indicating the solution vector and the discrete operators on $\mathbf{x}_L$ and $\mathbf{x}_R$ with the subscript L and R, we write

$$\begin{aligned}\mathbf{u}_{L,t} + \frac{1}{2}P_L^{-1}(Q_LA_L + A_LQ_L)\mathbf{u}_L + \frac{1}{2}U_LD_L\mathbf{a}_L &= \sigma_LP_L^{-1}\mathbf{e}_{x_I,L}(\mathbf{e}_{x_I,L}^TA_L\mathbf{u}_L - a(x_I)\mathbf{e}_{x_I,R}^T\mathbf{u}_R),\\ \mathbf{u}_{R,t} + \frac{1}{2}P_R^{-1}(Q_RA_R + A_RQ_R)\mathbf{u}_L + \frac{1}{2}U_RD_R\mathbf{a}_R &= \sigma_RP_R^{-1}\mathbf{e}_{x_I,R}(\mathbf{e}_{x_I,R}^TA_R\mathbf{u}_R - a(x_I)\mathbf{e}_{x_I,L}^T\mathbf{u}_L),\end{aligned} \tag{23}$$

where $D_L = P_L^{-1}Q_L$, $D_R = P_R^{-1}Q_R$ and $\sigma_L, \sigma_R \in \mathbb{R}$ are penalty parameters.

Next, we rewrite (23) in terms of a compact operator and follow the notation in [29] by introducing the discrete solution $\mathbf{u} = [\mathbf{u}_L^T, \mathbf{u}_R^T]^T$ and the boundary interpolants $\mathbf{e}_\alpha = [\mathbf{e}_{\alpha,L}^T, \mathbf{0}_R^T]^T$, $\mathbf{e}_\beta = [\mathbf{0}_L^T, \mathbf{e}_{\beta,R}^T]^T$. The interface penalty terms in (23) can now be represented in matrix form as

$$\begin{bmatrix}\sigma_LP_L^{-1}\mathbf{e}_{x_I,L}(\mathbf{e}_{x_I,L}^TA_L\mathbf{u}_L - a(x_I)\mathbf{e}_{x_I,R}^T\mathbf{u}_R)\\ \sigma_RP_R^{-1}\mathbf{e}_{x_I,R}(\mathbf{e}_{x_I,R}^TA_R\mathbf{u}_R - a(x_I)\mathbf{e}_{x_I,L}^T\mathbf{u}_L)\end{bmatrix} = a(x_I)\mathcal{P}^{-1}E_{x_I}\Sigma E_{x_I}^T\mathbf{u}, \tag{24}$$

since $\mathbf{e}_{x_I,L}^TA_L\mathbf{u}_L = a(x_I)\mathbf{e}_{x_I,L}^T\mathbf{u}_L$ and $\mathbf{e}_{x_I,R}^TA_R\mathbf{u}_R = a(x_I)\mathbf{e}_{x_I,R}^T\mathbf{u}_R$.

In (24), we have introduced the block diagonal norm $\mathcal{P} = \operatorname{diag}(P_L, P_R)$. Moreover, E_{x_I} is a $(N_L + N_R + 2)\times 2$ matrix that projects grid functions on $\mathbf{x}$ to the interface points, while Σ is a 2×2 penalty matrix:

$$E_{x_I} = \begin{bmatrix}\mathbf{e}_{x_I,L} & \mathbf{0}\\ \mathbf{0} & \mathbf{e}_{x_I,R}\end{bmatrix}, \quad \Sigma = \begin{bmatrix}\sigma_L & -\sigma_L\\ -\sigma_R & \sigma_R\end{bmatrix}.$$

By also letting $\mathcal{U} = \operatorname{diag}(U_L, U_R)$ and $\mathbf{a} = [\mathbf{a}_L^T, \mathbf{a}_R^T]^T$, the approximation (23) can be expressed in compact form as

$$\mathbf{u}_t + \frac{1}{2}\mathcal{P}^{-1}\mathcal{Q}\mathbf{u} + \frac{1}{2}\mathcal{U}\mathcal{D}\mathbf{a} = \mathbf{0}, \tag{25}$$

where $\mathcal{D} = \text{diag}(D_L, D_R)$ and

$$\mathcal{Q} = \begin{bmatrix} Q_L A_L + A_L Q_L & 0 \\ 0 & Q_R A_R + A_R Q_R \end{bmatrix} - 2a(x_I) E_{x_I} \Sigma E_{x_I}^T. \tag{26}$$

We will now prove

Proposition 6.1. *The multi-domain discretization* (25) *of* (3) *with a general* $a(x)$ *using the standard SBP operators is conservative for* $\sigma_L = \sigma_R + 1$.

Proof. Let $\mathbf{1} = [\mathbf{1}_L^T, \mathbf{1}_R^T]^T$, $\mathcal{A} = \text{diag}(\mathbf{a})$ and consider the following problem

$$\mathbf{u}_t + \frac{1}{2}\mathcal{P}^{-1}\mathcal{Q}\mathcal{U}\mathbf{1} + \frac{1}{2}\mathcal{U}\mathcal{D}\mathcal{A}\mathbf{1} = \mathbf{0}, \tag{27}$$

which by Definition 4.3 is equivalent to (25). By multiplying (27) from the left with $\mathbf{1}^T\mathcal{P}$ we find

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1}, \mathbf{u})_{\mathcal{P}} + \frac{1}{2}\mathbf{1}^T\mathcal{Q}\mathcal{U}\mathbf{1} + \frac{1}{2}\mathbf{1}^T\mathcal{U}\mathcal{P}\mathcal{D}\mathcal{A}\mathbf{1} = 0. \tag{28}$$

The second term in (28) can be rewritten using property iii) in Definition 3.1) as

$$\begin{aligned}
\mathbf{1}^T\mathcal{Q}\mathcal{U}\mathbf{1} &= \mathbf{1}_L^T\left[\left(\mathbf{e}_{x_I,L}\mathbf{e}_{x_I,L}^T - Q_L^T\right)A_L U_L + A_L Q_L U_L\right]\mathbf{1}_L \\
&+ \mathbf{1}_R^T\left[-\left(\mathbf{e}_{x_I,R}\mathbf{e}_{x_I,R}^T + Q_R^T\right)A_R U_R + A_R Q_R U_R\right]\mathbf{1}_R \\
&- 2a(x_I)\left(E_{x_I}^T\mathbf{1}\right)^T \Sigma \left(E_{x_I}^T\mathbf{u}\right) \\
&= (\mathbf{e}_{x_I,L}^T\mathbf{1}_L)\mathbf{e}_{x_I,L}^T A_L U_L \mathbf{1}_L - (Q_L\mathbf{1}_L)^T A_L U_L \mathbf{1}_L + \mathbf{1}_L^T A_L Q_L U_L \mathbf{1}_L \\
&- (\mathbf{e}_{x_I,R}^T\mathbf{1}_R)\mathbf{e}_{x_I,R}^T A_R U_R \mathbf{1}_R - (Q_R\mathbf{1}_R)^T A_R U_R \mathbf{1}_R + \mathbf{1}_R^T A_R Q_R U_R \mathbf{1}_R \\
&- 2a(x_I)\begin{bmatrix} \mathbf{e}_{x_I,L}^T\mathbf{1}_L \\ \mathbf{e}_{x_I,R}^T\mathbf{1}_R \end{bmatrix}^T \begin{bmatrix} \sigma_L & -\sigma_L \\ -\sigma_R & \sigma_R \end{bmatrix} \begin{bmatrix} \mathbf{e}_{x_I,L}^T\mathbf{u}_L \\ \mathbf{e}_{x_I,R}^T\mathbf{u}_R \end{bmatrix} \\
&= a(x_I)(1 + 2\sigma_R - 2\sigma_L)(u_{x_I,L} - u_{x_I,R}) + \mathbf{1}^T\mathcal{A}\mathcal{P}\mathcal{D}\mathcal{U}\mathbf{1}.
\end{aligned} \tag{29}$$

The substitution of (29) into (28) and the use of $\mathbf{1}^T\mathcal{U}\mathcal{P}\mathcal{D}\mathcal{A}\mathbf{1} + \mathbf{1}^T\mathcal{A}\mathcal{P}\mathcal{D}\mathcal{U}\mathbf{1} = \mathbf{1}^T\mathcal{U}\left(\mathcal{P}\mathcal{D} + (\mathcal{P}\mathcal{D})^T\right)\mathcal{A}\mathbf{1}$ leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1}, \mathbf{u})_{\mathcal{P}} = a(x_I)(\sigma_L - \sigma_R - 1)(u_{x_I,L} - u_{x_I,R}), \tag{30}$$

and conservation follows. □

6.2. Conservation for GSBP multi-domain approximations

Consider the two-domain GSBP discretization of (3)

$$\begin{aligned}
\mathbf{u}_{L,t} + \frac{1}{2}P_L^{-1}(Q_L A_L + A_L Q_L)\mathbf{u}_L + \frac{1}{2}U_L D_L \mathbf{a}_L &= \sigma_L P_L^{-1}\mathbf{t}_{x_I,L}(\mathbf{t}_{x_I,L}^T A_L \mathbf{u}_L - a(x_I)\mathbf{t}_{x_I,R}^T\mathbf{u}_R), \\
\mathbf{u}_{R,t} + \frac{1}{2}P_R^{-1}(Q_R A_R + A_R Q_R)\mathbf{u}_L + \frac{1}{2}U_R D_R \mathbf{a}_R &= \sigma_R P_R^{-1}\mathbf{t}_{x_I,R}(\mathbf{t}_{x_I,R}^T A_R \mathbf{u}_R - a(x_I)\mathbf{t}_{x_I,L}^T\mathbf{u}_L),
\end{aligned}$$

where we have used the same interface penalties as in (23) and continue to omit the boundary terms. This problem can be rewritten as

$$\mathbf{u}_t + \frac{1}{2}\mathcal{P}^{-1}\mathcal{Q}\mathbf{u} + \frac{1}{2}\mathcal{U}\mathcal{D}\mathbf{a} = \mathbf{0}, \tag{31}$$

where now

$$\mathcal{Q} = \begin{bmatrix} Q_L A_L + A_L Q_L & 0 \\ 0 & Q_R A_R + A_R Q_R \end{bmatrix} - 2\begin{bmatrix} \sigma_L \mathbf{t}_{x_I,L}\mathbf{t}_{x_I,L}^T A_L & -a(x_I)\sigma_L \mathbf{t}_{x_I,L}\mathbf{t}_{x_I,R}^T \\ -a(x_I)\sigma_R \mathbf{t}_{x_I,R}\mathbf{t}_{x_I,L}^T & \sigma_R \mathbf{t}_{x_I,R}\mathbf{t}_{x_I,R}^T A_R \end{bmatrix}.$$

Following the steps in the proof of Proposition 6.1, we find that the conservation relation associated to (31) is

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},\mathbf{u})_{\mathcal{P}} &= \tfrac{1}{2}\{(2\sigma_L-1)\mathbf{t}^T_{x_I,L}A_L\mathbf{u}_L+(2\sigma_R+1)\mathbf{t}^T_{x_I,R}A_R\mathbf{u}_R \\ &\quad - \left[\mathbf{t}^T_{x_I,L}\mathbf{a}_L+2a(x_I)\sigma_R\right]\mathbf{t}^T_{x_I,L}\mathbf{u}_L \\ &\quad +\left[\mathbf{t}^T_{x_I,R}\mathbf{a}_R-2a(x_I)\sigma_L\right]\mathbf{t}^T_{x_I,R}\mathbf{u}_R\}.\end{aligned}\tag{32}$$

The most natural choice, proposed also in [28], is to set $\sigma_L = -\sigma_R = 1/2$ in order to make the first two terms in (32) vanish. This gives

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},\mathbf{u})_{\mathcal{P}} = -\frac{1}{2}\left\{\left[\mathbf{t}^T_{x_I,L}\mathbf{a}_L-a(x_I)\right]\mathbf{t}^T_{x_I,L}\mathbf{u}_L-\left[\mathbf{t}^T_{x_I,R}\mathbf{a}_R-a(x_I)\right]\mathbf{t}^T_{x_I,R}\mathbf{u}_R\right\}.$$

In concert with the single domain case we say that the approximation is *weakly* conservative since the interface terms only vanish if $\mathbf{t}^T_{x_I,L}\mathbf{a}_L = \mathbf{t}^T_{x_I,R}\mathbf{a}_R = a(x_I)$. This requirement holds for polynomial advection coefficients of order at most N, see (9).

For other choices of σ_L and σ_R, the right-hand side of (32) can not be made identically zero. Hence, we have proved

Proposition 6.2. *The multi-domain discretization* (31) *of* (3) *with a general* $a(x)$ *using the GSBP operators in* Definition 3.3 *is not conservative.*

Remark 6.3. In order to prove conservation, one may consider an augmented set of penalty terms as in the single-domain case, i.e.

$$\begin{aligned}&P_L^{-1}\left[\sigma_L^I A_L\mathbf{t}_{x_I,L}\left(\mathbf{t}^T_{x_I,L}\mathbf{u}_L-\mathbf{t}^T_{x_I,R}\mathbf{u}_R\right)+\sigma_L^{II}\mathbf{t}_{x_I,L}\left(\mathbf{t}^T_{x_I,L}A_L\mathbf{u}_L-a(x_I)\mathbf{t}^T_{x_I,R}\mathbf{u}_R\right)\right],\\ &P_R^{-1}\left[\sigma_R^I A_R\mathbf{t}_{x_I,R}\left(\mathbf{t}^T_{x_I,R}\mathbf{u}_R-\mathbf{t}^T_{x_I,L}\mathbf{u}_L\right)+\sigma_R^{II}\mathbf{t}_{x_I,R}\left(\mathbf{t}^T_{x_I,R}A_R\mathbf{u}_R-a(x_I)\mathbf{t}^T_{x_I,L}\mathbf{u}_L\right)\right].\end{aligned}$$

The resulting conservation relation becomes

$$\begin{aligned}\tfrac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},\mathbf{u})_{\mathcal{P}} &= \tfrac{1}{2}\{(2\sigma_L^{II}-1)\mathbf{t}^T_{x_I,L}A_L\mathbf{u}_L+(2\sigma_R^{II}+1)\mathbf{t}^T_{x_I,R}A_R\mathbf{u}_R \\ &\quad + \left[\sigma_L^I\mathbf{t}^T_{x_I,L}\mathbf{a}_L-\sigma_R^I\mathbf{t}^T_{x_I,R}\mathbf{a}_R-\mathbf{t}^T_{x_I,L}\mathbf{a}_L-2a(x_I)\sigma_R^{II}\right]\mathbf{t}^T_{x_I,L}\mathbf{u}_L \\ &\quad - \left[\sigma_L^I\mathbf{t}^T_{x_I,L}\mathbf{a}_L-\sigma_R^I\mathbf{t}^T_{x_I,R}\mathbf{a}_R-\mathbf{t}^T_{x_I,R}\mathbf{a}_R+2a(x_I)\sigma_L^{II}\right]\mathbf{t}^T_{x_I,R}\mathbf{u}_R\}.\end{aligned}$$

As before, the first two terms are identically zero only if $\sigma_L^{II} = -\sigma_R^{II} = 1/2$, while the remaining part vanishes if, and only if,

$$\begin{bmatrix}1 & -1\\ -1 & 1\end{bmatrix}\begin{bmatrix}\sigma_L^I\mathbf{t}^T_{x_I,L}\mathbf{a}_L\\ \sigma_R^I\mathbf{t}^T_{x_I,R}\mathbf{a}_R\end{bmatrix}=\begin{bmatrix}\mathbf{t}^T_{x_I,L}\mathbf{a}_L-a(x_I)\\ \mathbf{t}^T_{x_I,R}\mathbf{a}_R-a(x_I)\end{bmatrix}.$$

This is a rank-one 2 × 2 system solvable only when $\mathbf{t}^T_{x_I,L}\mathbf{a}_L+\mathbf{t}^T_{x_I,R}\mathbf{a}_R = 2a(x_I)$. As in the previous case, this condition does not hold for a general $a(x)$ and the augmented formulation is not conservative. For simplicity we will only continue to study the two-parameters GSBP formulation shown above.

7. Stability of multi-domain discretizations

In the previous sections we proved that the SBP-SAT single-domain discretization (15) is stable and that the corresponding two-domain formulation (25) is conservative. The same conclusions does not hold for the GSBP-SAT approach. In this section, we will study the stability properties of the inter-element coupling procedure for both types of discretization. Again, the two-domain case is studied and the conclusions are generalized to multi-domains.

7.1. Stability for multi-domain SBP approximations

We start with the standard SBP approach and consider the two-domain SBP-SAT discretization (25) with the conservative choice $\sigma_L = \sigma_R + 1$. By using the parametrization proposed in [30], i.e. $\sigma_L = \sigma + 1/2$ and $\sigma_R = \sigma - 1/2$, the matrix $\mathcal{Q}$ in (26) satisfies the following Summation-By-Parts property

$$\frac{\mathcal{Q}+\mathcal{Q}^T}{2} = a(\beta)\mathbf{e}_\beta\mathbf{e}_\beta^T - a(\alpha)\mathbf{e}_\alpha\mathbf{e}_\alpha^T - 2\sigma a(x_I)E_{x_I}ME_{x_I}^T,\tag{33}$$

where the matrix

$$M = \begin{bmatrix} +1 & -1 \\ -1 & +1 \end{bmatrix}$$

has the eigenvalues $\{0, 2\}$.

We apply the discrete energy-method to (25) with an added penalty term for the boundary condition. The relation (33) yields

$$\tfrac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}\|^2_{\mathcal{P}} \ = a(\alpha)g_\alpha^2(t) - a(\beta)u_N^2 - (\mathbf{u}, \mathcal{U}\mathcal{D}\mathbf{a})_{\mathcal{P}} - a(\alpha)(u_0 - g_\alpha(t))^2 + 2\sigma a(x_I)(E_{x_I}^T\mathbf{u})^T M(E_{x_I}^T\mathbf{u}). \tag{34}$$

The estimate (34) differs from the single-block energy-rate (21) only by the interface contribution $2\sigma a(x_I)(E_{x_I}^T\mathbf{u})^T M(E_{x_I}^T\mathbf{u})$. Since the matrix M is symmetric and positive semi-definite, we conclude that the two-domain formulation is stable if σ and $a(x_I)$ have opposite signs.

We have proved

Proposition 7.1. *The multi-domain SBP discretization (25) of (3) with a general $a(x)$ is energy-stable for $\sigma = 0$ or $sgn(\sigma) = -sgn(a(x_I))$.*

Note that the interface term in (34) adds dissipation for $\sigma \neq 0$.

7.2. Stability for multi-domain GSBP approximations

Next, we study the stability of the two-domain GSBP-SAT discretization (31) with an added penalty term for the left boundary condition. As in the previous case, we write the Summation-By-Parts property associated to the matrix $\mathcal{Q}$ as

$$\begin{aligned}
\frac{\mathcal{Q} + \mathcal{Q}^T}{2} =& \begin{bmatrix} -(\mathbf{t}_{\alpha,L}\mathbf{t}_{\alpha,L}^T A_L)^S & 0 \\ 0 & (\mathbf{t}_{\beta,R}\mathbf{t}_{\beta,R}^T A_R)^S \end{bmatrix} \\
&+ \begin{bmatrix} (\mathbf{t}_{x_I,L}\mathbf{t}_{x_I,L}^T A_L)^S & 0 \\ 0 & -(\mathbf{t}_{x_I,R}\mathbf{t}_{x_I,R}^T A_R)^S \end{bmatrix} \\
&- 2\begin{bmatrix} \sigma_L\left(\mathbf{t}_{x_I,L}\mathbf{t}_{x_I,L}^T A_L\right)^S & -a(x_I)(\sigma_L + \sigma_R)\mathbf{t}_{x_I,L}\mathbf{t}_{x_I,R}^T \\ -a(x_I)(\sigma_L + \sigma_R)\mathbf{t}_{x_I,R}\mathbf{t}_{x_I,L}^T & \sigma_R\left(\mathbf{t}_{x_I,R}\mathbf{t}_{x_I,R}^T A_R\right)^S \end{bmatrix},
\end{aligned} \tag{35}$$

where the superscript S denotes the symmetric part. The first matrix in (35) is related to the boundary nodes α and β, while the next two relate to the interface at x_I. The energy-method for $g_\alpha = 0$ leads to

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}\|^2_{\mathcal{P}} = -\mathbf{u}^T A\mathbf{t}_\alpha\mathbf{t}_\alpha^T\mathbf{u} - \mathbf{u}^T A\mathbf{t}_\beta\mathbf{t}_\beta^T\mathbf{u} - (\mathbf{u}, \mathcal{U}\mathcal{D}\mathbf{a})_{\mathcal{P}} - \mathbf{u}^T \mathrm{IT}\mathbf{u},$$

which is similar to the estimate (22), except for the interface term related to

$$\mathrm{IT} = \begin{bmatrix} (1 - 2\sigma_L)A_L\mathbf{t}_{x_I,L}\mathbf{t}_{x_I,L}^T & 2a(x_I)\sigma_L\mathbf{t}_{x_I,L}\mathbf{t}_{x_I,R}^T \\ 2a(x_I)\sigma_R\mathbf{t}_{x_I,R}\mathbf{t}_{x_I,L}^T & -(1 + 2\sigma_R)A_R\mathbf{t}_{x_I,R}\mathbf{t}_{x_I,R}^T \end{bmatrix}.$$

This matrix is skew-symmetric if, and only if, $\sigma_L = -\sigma_R = 1/2$. Otherwise, Sylvester's criterion [31] implies that the symmetric part of IT is indefinite, since both $(A_L\mathbf{t}_{x_I,L}\mathbf{t}_{x_I,L}^T)^S$ and $(A_R\mathbf{t}_{x_I,R}\mathbf{t}_{x_I,R}^T)^S$ are in general indefinite.

We have proved

Proposition 7.2. *The coupling procedure of the multi-domain GSBP discretization (31) of (3) with a general $a(x)$ is stable for $\sigma_L = -\sigma_R = 1/2$.*

Remark 7.3. Additional dissipation can be introduced by using upwind SATs, first proposed in [32] (see also [28,33]) of the form

$$\begin{aligned}
&\frac{1}{2}P_L^{-1}\mathbf{t}_{x_I,L}\left[\mathbf{t}_{x_I,L}^T A_L\mathbf{u}_L - a(x_I)\mathbf{t}_{x_I,R}^T\mathbf{u}_R - |a(x_I)|\left(\mathbf{t}_{x_I,L}^T\mathbf{u}_L - \mathbf{t}_{x_I,R}^T\mathbf{u}_R\right)\right], \\
-&\frac{1}{2}P_R^{-1}\mathbf{t}_{x_I,R}\left[\mathbf{t}_{x_I,R}^T A_R\mathbf{u}_R - a(x_I)\mathbf{t}_{x_I,L}^T\mathbf{u}_L - |a(x_I)|\left(\mathbf{t}_{x_I,L}^T\mathbf{u}_L - \mathbf{t}_{x_I,R}^T\mathbf{u}_R\right)\right].
\end{aligned}$$

8. A possible remedy for GSBP operators

We have shown that the GSBP operators in Definition 3.3 can not be used to construct conservative and stable discretizations for variable coefficient advection problems. The main issue with this methodology is that the Summation-By-Part property does not involve the advection coefficient. In this section we propose a modified version of GSBP operators inspired by the splitting in (13) and with dependency on $a(x)$.

Rather than approximating the first derivative, we build a discretization for the anti-symmetric portion of $(au)_x$, i.e. we aim for

$$P^{-1}\Theta\mathbf{u} \approx \frac{1}{2}\left[(au)_x + au_x\right] = \frac{1}{2}a_x u + au_x. \tag{36}$$

We refer to these modified operators as a-Generalized Summation-By-Parts (a-GSBP) operators.

Definition 8.1. An operator $P^{-1}\Theta$ is a qth order accurate approximation of the anti-symmetric portion of $(au)_x$ in (36) with the a-GSBP property if

i) $P^{-1}\Theta\mathbf{x}^j = (1/2)A_x\mathbf{x}^j + jA\mathbf{x}^{j-1}$, $j \in [0, q]$,
ii) P is a symmetric positive definite matrix,
iii) $\Theta + \Theta^T = a(\beta)\mathbf{t}_\beta\mathbf{t}_\beta^T - a(\alpha)\mathbf{t}_\alpha\mathbf{t}_\alpha^T$.

The symmetric part of the operator $(au)_x$, i.e. $(1/2)a_x u$, can be consistently represented by $UP^{-1}\Theta\mathbf{1}$, due to (36) and condition i). Condition iii) implies that a-GSBP discretizations are based on the continuous property

$$\frac{1}{2}(v, (aw)_x + aw_x)_2 = a(\beta)v(\beta)w(\beta) - a(\alpha)v(\alpha)w(\alpha) - \frac{1}{2}(av_x + (av)_x, w)_2.$$

For a constant coefficient advection problem, a-GSBP operators represent the first derivative and mimic the IBP rule (5).

Remark 8.2. The standard SBP operators satisfy iii) in Definition 8.1.

8.1. Conservation and stability for a-GSBP operators

It is easy to verify that the conservation and stability analysis made for the SBP-SAT discretizations apply with minor modifications to

$$\mathbf{u}_t + P^{-1}\Theta\mathbf{u} + UP^{-1}\Theta\mathbf{1} = \sigma_\alpha a(\alpha)P^{-1}\mathbf{t}_\alpha(\mathbf{t}_\alpha^T\mathbf{u} - g_\alpha(t)). \tag{37}$$

We can prove

Proposition 8.3. *The discretization (37) of (3) with a general $a(x)$ using the a-GSBP operators in Definition 8.1 is conservative for $\sigma_\alpha = -1$.*

Proof. By multiplying (37) from left with $\mathbf{1}^T P$ we find

$$\frac{\mathrm{d}}{\mathrm{dt}}(\mathbf{1}, \mathbf{u})_P + \mathbf{1}^T\left(\Theta U + U\Theta\right)\mathbf{1} = \sigma_\alpha a(\alpha)(\mathbf{t}_\alpha^T\mathbf{u} - g_\alpha(t)). \tag{38}$$

The second term on the left-hand side of (38) can be rewritten as $\mathbf{1}^T\left(\Theta U + U\Theta\right)\mathbf{1} = \mathbf{1}^T\left(\Theta + \Theta^T\right)U\mathbf{1} = a(\beta)\mathbf{t}_\beta^T\mathbf{u} - a(\alpha)\mathbf{t}_\alpha^T\mathbf{u}$. This leads to

$$\frac{\mathrm{d}}{\mathrm{dt}}(\mathbf{1}, \mathbf{u})_P = a(\alpha)\mathbf{t}_\alpha^T\mathbf{u} - a(\beta)\mathbf{t}_\beta^T\mathbf{u} + \sigma_\alpha a(\alpha)(\mathbf{t}_\alpha^T\mathbf{u} - g_\alpha(t)),$$

which exactly mimics the conservation relation (12) if $\sigma_\alpha = -1$. □

Next, we prove

Proposition 8.4. *The discretization (37) of (3) with a general $a(x)$ using the a-GSBP operators in Definition 8.1 is energy-stable.*

Proof. Let $\sigma_\alpha = -1$. Multiplying (37) from left by $\mathbf{u}^T P$ and applying the a-GSBP property leads to

$$\frac{\mathrm{d}}{\mathrm{dt}}\|\mathbf{u}\|_P^2 = a(\alpha)g_\alpha^2(t) - a(\beta)\left(\mathbf{t}_\beta^T\mathbf{u}\right)^2 - 2(\mathbf{u}, U\Theta\mathbf{1})_P - a(\alpha)(\mathbf{t}_\alpha^T\mathbf{u} - g_\alpha(t))^2.$$

This energy-rate is analogous to (21) and energy-stability follows. □

The a-GSBP approach also enables conservative and stable multi-element formulations. The proofs are straightforward and left to the reader.

8.2. Examples of a-GSBP operators

As a first example, consider $a(x) = \sqrt{1-x^2}$ and the domain $(\alpha, \beta) = (-1, 1)$. A first order accurate a-GSBP operator on the nodes $x_j = \cos((4-j)\pi/5)$, $j = 0, \dots, 3$ is given for the norm $P = \text{diag}[p_0, p_1, p_1, p_0]$ with $p_0 = 1$, $p_1 = 0.6180339887$. The associated skew-symmetric matrix Θ is

$$\Theta = \begin{bmatrix} 0 & \theta_{12} & \theta_{13} & \theta_{14} \\ -\theta_{12} & 0 & \theta_{23} & \theta_{24} \\ -\theta_{13} & -\theta_{23} & 0 & \theta_{34} \\ -\theta_{14} & -\theta_{24} & -\theta_{34} & 0 \end{bmatrix}$$

where

$$\begin{aligned} \theta_{12} = \theta_{34} &= -0.5257311121, & \theta_{14} &= -1.0131106564, \\ \theta_{13} = \theta_{24} &= 2.2270327288, & \theta_{23} &= -2.6523581330. \end{aligned}$$

As a second example, a 2nd order accurate a-GSBP operator for $a(x) = e^x \cos(x)$ on $(\alpha, \beta) = (-1, 1)$ is given on the nodes $x_j = \cos((6-j)\pi/7)$, $j = 0, \dots, 5$. The a-GSBP norm is $P = \text{diag}[p_0, p_1, p_2, p_3, p_4, p_5]$ with $p_0 = 0.1807450623$, $p_1 = 0.4493684668$, $p_2 = 0.3060513005$, $p_3 = 0.5586874298$, $p_4 = 0.2615743465$, $p_5 = 1/4$. By considering the Lagrange interpolants $\mathbf{t}_\alpha$ and $\mathbf{t}_\beta$ for $r = 2$ in (9), the matrix $a(\beta)\mathbf{t}_\beta \mathbf{t}_\beta^T - a(\alpha)\mathbf{t}_\alpha \mathbf{t}_\alpha^T$ is block diagonal with 3×3 blocks and we find

$$\Theta = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} & \theta_{14} & \theta_{15} & \theta_{16} \\ \theta_{21} & \theta_{22} & \theta_{23} & \theta_{24} & \theta_{25} & \theta_{26} \\ \theta_{31} & \theta_{32} & \theta_{33} & \theta_{34} & \theta_{35} & \theta_{36} \\ -\theta_{14} & -\theta_{24} & -\theta_{34} & \theta_{44} & \theta_{45} & \theta_{46} \\ -\theta_{15} & -\theta_{25} & -\theta_{35} & \theta_{54} & \theta_{55} & \theta_{56} \\ -\theta_{16} & -\theta_{26} & -\theta_{36} & \theta_{64} & \theta_{65} & \theta_{66} \end{bmatrix}$$

where

$$\begin{aligned} &\theta_{11} = -0.2402977716, && \theta_{12} = 0.3953075927, && \theta_{13} = -0.0993532508, \\ &\theta_{14} = -0.0253873146, && \theta_{15} = 0.0236280213, && \theta_{16} = -0.0023318272, \\ &\theta_{21} = -0.1814224545, && \theta_{22} = -0.0475939207, && \theta_{23} = 0.3606468273, \\ &\theta_{24} = 0.0868083292, && \theta_{25} = -0.0140669134, && \theta_{26} = -0.0362624660, \\ &\theta_{31} = 0.0569906760, && \theta_{32} = -0.3417937080, && \theta_{33} = -0.0018670458, \\ &\theta_{34} = 0.6013867421, && \theta_{35} = -0.2230998806, && \theta_{36} = 0.05489342718, \\ &\theta_{44} = 0.01379570584, && \theta_{45} = 0.9039654995, && \theta_{46} = 0.0083927376, \\ &\theta_{54} = -1.0432722564, && \theta_{55} = 0.3516741501, && \theta_{56} = 0.5336790144, \\ &\theta_{64} = 0.3046267038, && \theta_{65} = -2.1140882996, && \theta_{66} = 1.7755737146. \end{aligned}$$

Remark 8.5. The examples above show that it is possible to construct non-boundary conforming a-GSBP operators. However, the procedure is not practical for time-dependent coefficients $a(x,t)$, multi-element formulations and nonlinear problems which require frequent reconstruction of operators. The same result can be obtained with SBP operators without reconstruction.

9. Implication of the linear analysis on nonlinear problems

The conclusions drawn from the previous linear analysis carry over in a straightforward way to smooth nonlinear problems, with minor modifications. As an example, one must split the Burgers equation $u_t + \left(u^2/2\right)_x = 0$ as $u_t + (u^2)_x/3 + uu_x/3 = 0$ in order to obtain an energy-estimate. The conclusions regarding stability and conservation remain the same, i.e. stability and conservation can be proved for SBP operators but not for GSBP operators.

We end the paper by briefly commenting on how the standard SBP-SAT formulation can be applied to nonlinear conservation laws of the form (1) with the boundary condition at $x = \alpha$. The conservation law (1) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(1, u)_2 = f(g_\alpha, \alpha) - f(u(\beta, t), \beta), \tag{39}$$

where $u(\alpha, t) = g_\alpha$.

Consider a flux of the form $f(u) = v(u)w(u)$ and split it in an arbitrary way as $f_x = \gamma f_x + (1-\gamma)(vw_x + v_x w)$. The analysis of conservation in [22] of the split form was based on a detailed investigation of the difference operators and the use of so called flux points, in addition to the standard grid points. Here we show that Definition 4.3 simplifies the proof of conservation considerably. In particular, for the SBP-SAT discretization of (1)

$$\mathbf{u}_t + \gamma D\mathbf{f} + (1-\gamma)(VD\mathbf{w} + WD\mathbf{v}) = \sigma_\alpha P^{-1}\mathbf{e}_\alpha(\mathbf{e}_\alpha^T\mathbf{f} - f(g_\alpha)) \tag{40}$$

we can prove

Proposition 9.1. *The approximation (40) is conservative for any γ and $\sigma_\alpha = -1$.*

Proof. By multiplying (40) from the left by $\mathbf{1}^T P$ and using property iii) in Definition 4.3 we find

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},\mathbf{u})_P + \gamma\mathbf{1}^T Q\mathbf{f} + (1-\gamma)\mathbf{1}^T(VQW + WQV)\mathbf{1} = \sigma_\alpha(\mathbf{e}_\alpha^T\mathbf{f} - f(g_\alpha)). \tag{41}$$

The second term in (41) can be rewritten as $\mathbf{1}^T Q\mathbf{f} = \mathbf{1}^T(\mathbf{e}_\beta\mathbf{e}_\beta^T - \mathbf{e}_\alpha\mathbf{e}_\alpha^T - Q^T)\mathbf{f} = \mathbf{e}_\beta^T\mathbf{f} - \mathbf{e}_\alpha^T\mathbf{f} = f_N - f_0$ while the third one becomes

$$\mathbf{1}^T(VQW + WQV)\mathbf{1} = \mathbf{1}^T[V(Q+Q^T)W]\mathbf{1} = v_N w_N - v_0 w_0 = f_N - f_0.$$

Consequently, the resulting conservation relation becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbf{1},\mathbf{u})_P + \gamma(f_N - f_0) + (1-\gamma)(f_N - f_0) = \sigma_\alpha(f_0 - f(g_\alpha))$$

and by letting $\sigma_\alpha = -1$ we find that it perfectly mimics (39). □

10. Conclusions

We have discussed numerical approximations on Summation-By-Parts form for a linear advection problem with a variable coefficient. It was shown that the standard SBP-SAT formulation is conservative and stable for single and multi-element formulations.

The same problem was also studied with GSBP operators which do not include (the whole or part of) the boundary or interface and do not mimic the continuous integration-by-parts rule exactly. It was shown that the single-block GSBP-SAT formulation applied to variable coefficient problems is unstable and not conservative. Furthermore, the coupling between two or more of such blocks is stable but does not lead to a conservative scheme.

We have generalized the definition of GSBP operators. The generalization allows for conservative and stable schemes by approximating the anti-symmetric part of the continuous spatial operator, rather than the first derivative. However, the new GSBP operators are impractical for time-dependent coefficients, multi-element formulations and nonlinear problems.

These results limit the use of generalized Summation-By-Parts formulations as general building blocks in schemes, and stress the need for exact numerical mimicking of integration-by-parts.

Acknowledgement

The second author was financed by VINNOVA contract 2013-01209.

References

[1] H.O. Kreiss, J. Oliger, Comparison of accurate methods for the integration of hyperbolic equations, Tellus 24 (1972) 199–215.

[2] H.O. Kreiss, G. Scherer, Finite element and finite difference methods for hyperbolic partial differential equations, in: C. de Boor (Ed.), Mathematical Aspects of Finite Elements in Partial Differential Equations, Academic Press, 1974.

[3] B. Strand, Summation by parts for finite difference approximations for d/dx, J. Comput. Phys. 110 (1994) 47–67.

[4] M.H. Carpenter, D. Gottlieb, S. Abarbanel, Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: methodology and application to high-order compact schemes, J. Comput. Phys. 111 (2) (1994) 220–236.

[5] M. Svärd, J. Nordström, Review of summation-by-parts schemes for initial–boundary-value problems, J. Comput. Phys. 268 (2014) 17–38.

[6] D.C. Del Rey, J.E. Hicken, D.W. Zingg, Review of summation-by-parts operators with simultaneous approximation terms for the numerical solution of partial differential equations, Comput. Fluids 95 (2014) 171–196.

[7] J. Nordström, M.H. Carpenter, High-order finite difference methods, multidimensional linear problems and curvilinear coordinates, J. Comput. Phys. 173 (2001) 149–174.

[8] S. Nikkar, J. Nordström, Fully discrete energy stable high order finite difference methods for hyperbolic problems in deforming domains, J. Comput. Phys. 291 (2015) 82–98.

[9] J. Nordström, S. Nikkar, Hyperbolic systems of equations posed on erroneous curved domains, J. Comput. Phys. 308 (2016) 438–442.

[10] M.H. Carpenter, J. Nordström, D. Gottlieb, Revisiting and extending interface penalties for multi-domain summation-by-parts operators, J. Sci. Comput. 45 (2010) 118–150.

[11] J. Nordström, J. Gong, E. van der Weide, M. Svärd, A stable and conservative high order multi-block method for the compressible Navier–Stokes equations, J. Comput. Phys. 228 (2009) 9020–9035.

[12] G.J. Gassner, A skew-symmetric discontinuous Galerkin spectral element discretization and its relation to SBP-SAT finite difference methods, J. Sci. Comput. 35 (2013) 1233–1253.

[13] G.J. Gassner, A kinetic energy preserving nodal discontinuous Galerkin spectral element method, Int. J. Numer. Methods Fluids (2014) 28–50.

[14] J. Nordström, M. Björck, Finite volume approximations and strict stability for hyperbolic problems, Appl. Numer. Math. 38 (2001) 237–255.

[15] J. Nordström, K. Forsberg, C. Adamsson, P. Eliasson, Finite volume methods, unstructured meshes and strict stability for hyperbolic problems, Appl. Numer. Math. 45 (2003) 453–473.

[16] M.H. Carpenter, D. Gottlieb, Spectral methods on arbitrary grids, J. Comput. Phys. 129 (1996) 74–86.

[17] D.A. Kopriva, A.R. Winters, M. Bohm, G.J. Gassner, A provably stable discontinuous Galerkin spectral element approximation for moving hexahedral meshes, Comput. Fluids 139 (2016) 148–160.

[18] D.C. Del Rey, P.D. Boom, D.W. Zingg, A generalized framework for nodal first derivative summation-by-parts operators, J. Comput. Phys. 266 (2014) 214–239.

[19] B. Gustafsson, H.O. Kreiss, J. Oliger, Time Dependent Problems and Difference Methods, John Wiley & Sons, Inc., 1995.

[20] M. Svärd, On coordinate transformation for summation-by-parts operators, J. Sci. Comput. 20 (1) (2004).

[21] J. Nordström, Conservative finite difference formulations, variable coefficients, energy estimates and artificial dissipation, J. Sci. Comput. 29 (2006) 375–404.

[22] T.C. Fisher, M.H. Carpenter, J. Nordström, N.K. Yamaleev, C. Swanson, Discretely conservative finite-difference formulations for nonlinear conservation laws in split form: theory and boundary conditions, J. Comput. Phys. 234 (2013) 353–375.

[23] S. Abarbanel, A. Chertock, Strict stability of high-order compact implicit finite-difference schemes: the role of boundary conditions for hyperbolic PDEs, I, J. Comput. Phys. 160 (2000) 42–66.

[24] S. Abarbanel, A. Chertock, A. Yefet, Strict stability of high-order compact implicit finite-difference schemes: the role of boundary conditions for hyperbolic PDEs, II, J. Comput. Phys. 160 (2000) 67–87.

[25] C. Sorgentone, C. La Cognata, J. Nordström, A new high order energy and enstrophy conserving Arakawa-like Jacobian differential operator, J. Comput. Phys. 301 (2015) 167–177.

[26] M. Osusky, D. Zingg, Parallel Newton–Krylov–Schur flow solver for the Navier–Stokes equations, AIAA J. 51 (12) (2012) 2833–2851.

[27] E. Schnetter, P. Diener, E.N. Dorband, M. Tiglio, A multi-block infrastructure for three-dimensional time-dependent numerical relativity, Class. Quantum Gravity 23 (16) (2006) S553–S578.

[28] D.C. Del Rey, J.E. Hicken, D.W. Zingg, Simultaneous approximation terms for multi-dimensional summation-by-parts operators, J. Sci. Comput. (2017), submitted for publication; preprint, arXiv:1605.03214v2, 2016.

[29] J. Nordström, M.H. Carpenter, High-order finite difference methods, multidimensional linear problems, and curvilinear coordinates, J. Comput. Phys. 173 (2001) 149–174.

[30] S. Eriksson, Q. Abbas, J. Nordström, A stable and conservative method for locally adapting the design order of finite difference schemes, J. Comput. Phys. 230 (2011) 4216–4231.

[31] G.T. Gilbert, Positive definite matrices and Sylvester's criterion, Am. Math. Mon. 98 (1991) 44–46.

[32] K. Mattson, M.H. Carpenter, Stable and accurate interpolation operators for high-order multi-block finite-difference methods, SIAM J. Sci. Comput. 32 (4) (2010).

[33] J.E. Kozdon, L.C. Wilcox, Stable coupling of nonconforming, high-order finite difference methods, SIAM J. Sci. Comput. 38 (2016) A923–A952.