On pseudo-spectral time discretizations in summation-by-parts form

Andrea A Ruggiu and Jan Nordström

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):
http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-145083

N.B.: When citing this work, cite the original publication.

Original publication available at:
https://doi.org/10.1016/j.jcp.2018.01.043

Copyright: Elsevier
http://www.elsevier.com/
On pseudo-spectral time discretizations in summation-by-parts form

Andrea A. Ruggiu*, Jan Nordström

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

Abstract

Fully-implicit discrete formulations in summation-by-parts form for initial-boundary value problems must be invertible in order to provide well functioning procedures. We prove that, under mild assumptions, pseudo-spectral collocation methods for the time derivative lead to invertible discrete systems when energy-stable spatial discretizations are used.

1. Introduction

Well-posedness of partial differential equations is a key concept in mathematical modeling. A well-posed problem has a unique solution which is bounded by the data of the problem [1, 2]. Once initial and boundary conditions that lead to well-posedness are determined, a discrete approximation of the continuous problem can be formulated and solved, if it is stable. An additional requirement for fully-implicit discrete problems is the invertibility of the resulting system of equations. If this condition is not met, an iterative solver applied to such problems may converge to an erroneous solution [3].

Stable and high-order accurate discretizations for well-posed linear problems can be achieved in a straightforward way by combining Summation-By-Parts (SBP) operators [4, 5] and weak boundary and initial procedures using Simultaneous-Approximation-Terms (SATs) [6, 7, 8]. For initial value problems, the dual-consistent [9] SBP-SAT formulations are A- and L-stable
implicit Runge-Kutta schemes [10, 11]. However, the invertibility of fully-discrete stable approximations have so far only been conjectured in this setting [7, 11] (it does not follow from any known stability property). For further reading on these topics, see for example [12, 13].

In this work, we focus on pseudo-spectral collocation methods on finite domains [14] for initial value problems. We prove that these discretizations, which can be recast as SBP-SAT schemes with diagonal norms [15, 16], have eigenvalues with strictly positive real parts for specific choices of the penalty parameter. As a natural consequence, pseudo-spectral time discretizations combined with energy stable spatial approximations of initial-boundary value problems lead directly to invertible fully discrete approximations.

This paper is organized as follows. In section 2, the pseudo-spectral methods in SBP form are introduced for initial value problems. We also reformulate the invertibility issue in terms of the eigenvalues of the time discretization. In section 3, we show that pseudo-spectral SBP-SAT approximations yield discrete operators which can be easily studied through a coordinate transformation to Legendre polynomials. The eigenvalues of these matrices are analyzed in section 4, where the main result of the paper is stated and proved. Next, the properties of fully discrete systems with pseudo-spectral time discretizations are exemplified in section 5. Section 6 contains our conclusions.

2. Pseudo-spectral SBP-SAT time discretizations

The aim of this section is two-fold: firstly, we introduce the pseudo-spectral SBP-SAT discretizations for initial value problems [7]; secondly, we outline the main theoretical issue in this paper.

2.1. SBP operators

Consider a set of \( n \) nodes \( \mathbf{x} = [x_1, \ldots, x_n]^T \subset [\alpha, \beta] \) which may or may not include the endpoints of the interval. A Summation-By-Parts operator discretizing the first derivative on \( \mathbf{x} \) can be defined as follows [16]:

**Definition 2.1.** A discrete operator \( D \) is a \( q \)th order accurate approximation of the first derivative with the Summation-By-Parts (SBP) property if

\[ i) \quad D\mathbf{x}^j = P^{-1}Q\mathbf{x}^j = j\mathbf{x}^{j-1}, \quad j \in [0, q], \]

\[ ii) \quad P \text{ is a symmetric positive definite matrix}, \]

2
\[ Q + Q^T = E, \text{ where } E \text{ is such that } (x^i)^T E x^j = \beta^{i+j} - \alpha^{i+j}, \ i, j = 0, \ldots, r, \ r \geq q. \]

Condition i) in Definition 2.1 implies that the operator \( D \) exactly mimics the first derivative for the grid monomials \( x^j = [x_1^j, \ldots, x_n^j]^T \) up to the \( q \)th order. Condition ii) defines a discrete scalar product and a norm
\[
(v, w)_P = v^T P w, \quad \|v\|_P = \sqrt{(v, v)_P}
\]
which mimic the continuous \( L^2 \) counterparts. Furthermore, condition iii) implies that the integration-by-parts rule
\[
\int_\alpha^\beta w_x dx = u(\beta)v(\beta) - u(\alpha)v(\alpha) - \int_\alpha^\beta u_x v dx
\]
is exactly mimicked for \( u \) and \( v \) being polynomials of at most order \( r \). In particular,
\[
(x^i, Dx^j)_P = \beta^{i+j} - \alpha^{i+j} - (Dx^i, x^j)_P, \quad i, j = 0, \ldots, r. \tag{1}
\]
The discrete operators that we consider in this paper are based on pseudo-spectral collocation methods with accuracy \( q = n - 1 \) that satisfy (1) with \( r = n - 1 \).

For pseudo-spectral methods on finite domains, the grid is usually a subset of the reference interval \( [\alpha, \beta] = [-1, 1] \).

**Example 2.2.** Consider a 2nd order accurate SBP operator \( D = P^{-1}Q \) based on the Legendre-Gauss-Radau quadrature. On the three-point grid \( x = [-1, (1 - \sqrt{6})/5, (1 + \sqrt{6})/5]^T \), the SBP operator is given by
\[
P = \frac{1}{18} \begin{bmatrix} 4 & 0 & 0 \\ 0 & 16 + \sqrt{6} & 0 \\ 0 & 0 & 16 - \sqrt{6} \end{bmatrix}, \tag{2}
\]
\[
Q = \frac{1}{108} \begin{bmatrix} -48 & 24 + 14\sqrt{6} & 24 - 14\sqrt{6} \\ -12 - 32\sqrt{6} & 87 - 18\sqrt{6} & -75 + 50\sqrt{6} \\ -12 + 32\sqrt{6} & -75 - 50\sqrt{6} & 87 + 18\sqrt{6} \end{bmatrix}. \tag{3}
\]

The matrix \( E \) in Definition 2.1 can be written in terms of boundary interpolants of degree \( r \):
\[
e^T \gamma x^j = \gamma^j, \quad j = 0, \ldots, r, \ \gamma \in \{\alpha, \beta\}. \tag{4}
\]
where \( \mathbf{e}_T^T \mathbf{u} \approx u(\gamma) \). This gives rise to the relation \( E = \mathbf{e}_1 \mathbf{e}_1^T - \mathbf{e}_{-1} \mathbf{e}_{-1}^T \). In the example above, the interpolants are \( \mathbf{e}_{-1} = [1, 0, 0]^T \) and \( \mathbf{e}_1 = [1/3, (2 - 3\sqrt{6})/6, (2 + 3\sqrt{6})/6]^T \). The interpolation at \( x = -1 \) is exact, since this point is included in \( \mathbf{x} \). At \( x = 1 \) the interpolation is exact only for 2nd order polynomials.

Remark 2.3. Pseudo-spectral collocation methods typically involve only a few nodes and, hence, are suitable for multi-stage discretizations in time.

2.2. Restrictions for pseudo-spectral SBP operators based on diagonal norms

Next, we recall the necessary conditions on the quadrature rule \( P \) in order to satisfy the SBP property in Definition 2.1 [16]. In particular, we focus on diagonal norms \( P \). In this case, \( (\mathbf{x}^i, \mathbf{x}^j)_P = (\mathbf{x}^0, \mathbf{x}^{i+j})_P \) and by setting \( k = i + j - 1 \), the integration-by-parts rule in (1) leads to

\[
(x^0, x^k)_P = \beta^{k+1} - \alpha^{k+1} \frac{k + 1}{k + 1}, \quad k = 0, \ldots, 2r - 1.
\]

In order to satisfy this condition, the quadrature rule involved must be exact to order \( p \) where \( p \geq 2r - 1 = 2n - 3 \).

On the other hand, it is well-known that the maximum order which is attainable for a quadrature rule is \( 2n - 1 \) [17]. Hence, a diagonal norm must be based on a quadrature rule with \( p \in \{2n - 3, 2n - 2, 2n - 1\} \). Relevant examples of discretizations which satisfy this constraint are based on the Legendre-Gauss-Lobatto (here denoted by LGL, \( p = 2n - 3 \), both boundary nodes are included), Legendre-Gauss-Radau (LGR, \( p = 2n - 2 \), one boundary node is included) and Legendre-Gauss (LG, \( p = 2n - 1 \), no boundary nodes are included) quadratures.

For simplicity, we will restrict our analysis to these three diagonal norm based SBP operators. Our findings will later be extended to include all quadrature approximations (diagonal and non-diagonal) with \( p \in \{2n - 3, 2n - 2, 2n - 1\} \).

2.3. The initial value problem and its discretization

Consider the initial value problem

\[
\begin{align*}
    u_t + \lambda u &= 0, & 0 < t < T, \\
    u(0) &= f,
\end{align*}
\] (5)
where the complex constant \( \lambda \) represents a suitable spatial discretization of an initial-boundary value problem. In particular, energy-stability of the corresponding semi-discrete problem implies that \( \text{Re}(\lambda) \geq 0 \) \cite{7}. To discretize the initial value problem (5) in \([0, T]\) with the pseudo-spectral SBP operators \( P \) and \( Q \), we use the linear mapping

\[
t = t(\xi) := \frac{T}{2}(1 + \xi), \quad \xi \in [-1, 1].
\]  

The coordinate transformation (6) applied to (5) yields

\[
u_\xi + \lambda_\xi u = 0, \quad -1 < \xi < 1,
\]

\[
u(-1) = f,
\]

where \( \lambda_\xi = (dt/d\xi)\lambda = (T/2)\lambda \). To obtain an estimate of the solution at \( T = t(1) \), we apply the energy-method (multiplying by the complex conjugate of \( u \), integrating in time and using integration by parts) and get

\[|u(1)|^2 + 2\text{Re}(\lambda)\|u\|^2 = |f|^2.\]  

In (8), \( \|u\|^2 = \int_0^T |u|^2 dt = \int_{-1}^1 (T/2)|u|^2 d\xi \) and the solution at the final time is bounded by the initial data.

Next, consider the grid \( t = [t_1, \ldots, t_n]^T = [t(\xi_1), \ldots, t(\xi_n)]^T \), where \( \xi = [\xi_1, \ldots, \xi_n]^T \) are the nodes of a suitable quadrature rule. The SBP discretization of (7) with a weakly imposed initial condition using SAT reads

\[
P^{-1}Q u + \lambda_\xi u = \sigma P^{-1} e_{-1}(e_{T-1}^Tu - f).
\]

In (9), the vector \( u = [u_1, \ldots, u_n]^T \) contains the numerical approximations of \( u \) at each \( t_i, \ i = 1, \ldots, n \), i.e. \( u_i \approx u(t_i) \). Moreover, \( \sigma \) is a penalty parameter and \( e_{-1} \) is the interpolant at \( \xi = -1 \).

By applying the discrete-energy method to (9) (multiplying by \( u^*P \), where \( * \) denotes the conjugate transpose, and using the SBP property), we find

\[|e_{i}^Tu|^2 + 2\text{Re}(\lambda)\|u\|^2_{(T/2)P} = (1 + 2\sigma)|e_{T-1}^Tu|^2 - \sigma((e_{T-1}^Tu)f + (e_{T-1}^Tu)f).\]  

In (10), the bar indicates the complex conjugate of a scalar. For \( \sigma \leq -1/2 \) the discretization is stable. The (dual consistent \cite{9, 18}) choice \( \sigma = -1 \) leads to

\[|e_{i}^Tu|^2 + 2\text{Re}(\lambda)\|u\|^2_{(T/2)P} = |f|^2 - |e_{T-1}^Tu - f|^2,
\]
which mimics the continuous estimate (8). The additional term on the right-hand side adds numerical dissipation, which vanishes as accuracy increases.

Finally, we discuss the invertibility of the operator in (9). The discrete problem (9) can be recast as

$$(D + \lambda \xi I)u = F,$$

where $D = P^{-1}(Q - \sigma e_{-1}e_{1}^T)$ and $F = -\sigma f P^{-1}e_{-1}$. Recalling that the only constraint on $\lambda \xi$ is that $\text{Re}(\lambda \xi) \geq 0$, we conclude that the following proposition holds.

**Proposition 2.4.** The discrete problem (9) leads to an invertible system if the operator $D$ has eigenvalues with strictly positive real parts.

To avoid ambiguity, the eigenvalues of the discrete operator $D$ will henceforth be denoted by $\mu$.

### 3. SBP operators on modal basis

In this section, we show that, for pseudo-spectral methods, $D$ is similar to a matrix with a particular structure suitable for analysis. This result, which can be obtained through a coordinate transformation to Legendre polynomials, is based on a previous study in [19] and will enable the eigenvalue analysis in the next section.

#### 3.1. First derivative on the Legendre polynomial basis

To start with, we recall the Legendre basis representation of the derivative $D$. Consider an arbitrary set of $n$ nodes $x$ in the interval $[-1,1]$, and the Legendre basis functions $\phi_n(x)$, where the first four are

$$\phi_0(x) = 1, \quad \phi_1(x) = x, \quad \phi_2(x) = \frac{1}{2}(3x^2 - 1), \quad \phi_3(x) = \frac{1}{2}(5x^3 - 3x).$$

Each basis function $\phi_n(x)$ is a $n^{th}$ degree polynomial of the form

$$\phi_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n} [(x^2 - 1)^n].$$

We will use the following properties of the Legendre polynomials [20]:

i) $\phi_n(-1) = (-1)^n$, $\phi_n(1) = 1$ for $n = 0, 1, 2, \ldots$
ii) The derivative of the \((n + 1)\)th degree Legendre polynomial can be expressed in terms of lower degree polynomials using the relation
\[
\frac{d}{dx} \phi_{n+1}(x) = (2n + 1)\phi_n(x) + \frac{d}{dx} \phi_{n-1}(x), \quad n = 1, 2, \ldots
\]

iii) The Legendre polynomials are orthogonal in \(L^2(-1, 1)\) and satisfy
\[
\int_{-1}^{1} \phi_i(x) \phi_j(x) dx = \frac{2}{2j + 1} \delta_{ij}, \quad i, j = 0, 1, 2, \ldots
\]
where \(\delta_{ij}\) is equal to 1 if \(i = j\) and 0 otherwise.

Let \(\phi_i = \phi_i(x)\) be the grid function corresponding to the \(i\)th degree Legendre polynomial and define the Vandermonde matrix \(V = [\phi_0, \phi_1, \phi_2, \phi_3, \ldots]\). A \(q\)th order accurate first derivative operator \(D\) can be represented using the Legendre basis by imposing the \(q + 1\) accuracy conditions in Definition 2.1 in a slightly different way. More precisely, we can write
\[
DV = \begin{bmatrix} 0, 1, 3x, \frac{3}{2}(5x^2 - 1), \ldots \end{bmatrix} = [0, \phi_0, 3\phi_1, \phi_0 + 5\phi_2, \cdots]\]
\[
= [\phi_0, \phi_1, \phi_2, \phi_3, \cdots]
\begin{bmatrix}
0 & 1 & 0 & 1 & 0 & \cdots \\
0 & 0 & 3 & 0 & 3 & \cdots \\
0 & 0 & 0 & 5 & 0 & \cdots \\
0 & 0 & 0 & 0 & 7 & \cdots \\
0 & 0 & 0 & 0 & 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}
= V\hat{D},
\]
where \(V \in \mathbb{R}^{n \times (q+1)}\), \(\hat{D} \in \mathbb{R}^{(q+1) \times (q+1)}\). We stress that \(\hat{D}\) is a square matrix with a banded structure that does not depend on the grid nodes. Moreover, the last row of \(\hat{D}\) is filled with zero elements. As an example, the matrix \(\hat{D}\) for \(q = 3\) consists of the \(4 \times 4\) upper left submatrix in (13).

Pseudo-spectral collocation discretizations for the first derivative have accuracy \(q = n - 1\). Under this constraint, the Vandermonde matrix \(V \in \mathbb{R}^{n \times n}\) is invertible, since it contains \(n\) linearly independent column vectors \(\{\phi_i\}_{i=0,\ldots,n-1}\) with \(n\) components. Hence, the operator \(D\) is similar to \(\hat{D}\) since, from (13), the invertibility of \(V\) leads to \(D = V\hat{D}V^{-1}\).

**Remark 3.1.** Two \(n \times n\) matrices \(A\) and \(B\) are similar if \(B = XAX^{-1}\) for an invertible \(n \times n\) matrix \(X\). The eigenvalues of two similar matrices are identical [21].
3.2. The discrete norm for the Legendre polynomial basis

Following [19], we present the Legendre basis representation of the norm $P$ associated with the SBP operator $D$ in Definition 2.1. For the Legendre polynomial basis, (12) holds, and hence, for a $p$th order exact quadrature rule $P$, 

$$
\phi_i^T P \phi_j = \frac{2}{2j+1} \delta_{ij}, \quad i + j \leq p.
$$

(14)

By defining $\hat{P}_{ij} = \phi_i^T P \phi_j$, one gets

$$
\hat{P} = V^T P V = \begin{bmatrix}
\phi_0^T P \phi_0 & \phi_0^T P \phi_1 & \cdots & \phi_0^T P \phi_{n-1} \\
\phi_1^T P \phi_0 & \phi_1^T P \phi_1 & \cdots & \phi_1^T P \phi_{n-1} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{n-1}^T P \phi_0 & \phi_{n-1}^T P \phi_1 & \cdots & \phi_{n-1}^T P \phi_{n-1}
\end{bmatrix} \in \mathbb{R}^{n \times n},
$$

which represents the discrete norm in the Legendre basis. Note that also $\hat{P}$ defines a norm since it is symmetric and positive definite.

**Remark 3.2.** The matrix $\hat{P}$ is diagonal since we consider the Legendre polynomial basis and assume that $p \geq 2n - 3$.

**Remark 3.3.** The relation between the norms $P$ and $\hat{P}$ can be written as $P = V^{-T} \hat{P} V^{-1}$ and used to construct an accurate norm from an arbitrary set of nodes. As an example, the matrix $P$ in (2) can be derived from $\hat{P} = \text{diag}(2, 2/3, 2/5)$ by the constraints (14) with $p = 2n - 2 = 4$. The resulting matrix $PD$ with $D = V \hat{D} V^{-1}$ produces $Q$ in (3).

The norm $\hat{P}$ depends on the order $p$ of the quadrature. In general, one can prove that for all the quadratures with $p \in \{2n - 3, 2n - 2, 2n - 1\}$, there exists $\kappa > 0$ such that $\hat{P}_\kappa = \text{diag}(2, 2/3, \ldots, 2/(2n - 3), \kappa)$ [22].

**Example 3.4.** For the LGR and LG quadratures ($p = 2n - 2$ and $p = 2n - 1$, respectively) the numerical integration rules $\phi_i^T P \phi_j$ are exact for all the polynomials $\{\phi_k\}_{k=0,\ldots,n-1}$ and, from (14), it follows that $\hat{P}_{LGR} = \hat{P}_{LG} = \text{diag}(2, 2/3, \ldots, 2/(2n - 3), 2/(2n - 1))$. For the LGL quadrature rule ($p = 2n - 3$), the accuracy conditions (14) provide all the entries of $\hat{P}$ except for the one in the bottom corner, since $\phi_{n-1}^T P \phi_{n-1}$ is not exact. It can be shown that $\hat{P}_{LGL} = \text{diag}(2, 2/3, \ldots, 2/(2n - 3), 2/(n - 1))$. 

8
3.3. The SBP-SAT operator based on the Legendre polynomial basis

The modal representations of $D$ and $P$ can be used to formulate the SBP-SAT operator $D = P^{-1}(Q - \sigma e_{-1}e_{-1}^T)$ as

$$D = D - \sigma P^{-1}e_{-1}e_{-1}^T = V\hat{D}V^{-1} - \sigma V\hat{P}_{\kappa}^{-1}V^Te_{-1}e_{-1}^TV^{-1} =$$

$$= V[\hat{D} - \sigma \hat{P}_{\kappa}^{-1}(e_{-1}^TV)^T(e_{-1}^TV)]V^{-1} = V\hat{D}_\kappa V^{-1},$$

where $\hat{D}_\kappa = \hat{D} - \sigma \hat{P}_{\kappa}^{-1}(e_{-1}^TV)^T(e_{-1}^TV)$ is the representation of $D$ in the modal basis. The relation (15) proves

**Lemma 3.5.** $\hat{D}_\kappa$ and $D$ are similar and have the same spectrum.

Our aim is to get an explicit form of the matrix $\hat{D}_\kappa$ and thereby analyze the eigenvalues of $D$. Recalling that

$$e_{-1}^TV = [e_{-1}^T\phi_0, e_{-1}^T\phi_1, \ldots, e_{-1}^T\phi_{n-1}] = [\phi_0(-1), \phi_1(-1), \ldots, \phi_{n-1}(-1)] = [1, -1, \ldots, (-1)^{n-1}],$$

since the interpolation is exact for polynomials of degree at most $(n - 1)$, the matrix $\hat{D}_\kappa$ reads

$$\hat{D}_\kappa = \hat{D} - \sigma \hat{P}_{\kappa}^{-1}(e_{-1}^TV)^T(e_{-1}^TV) =$$

$$= \begin{bmatrix} 0 & 1 & 0 & 1 & \ldots \\ 0 & 0 & 3 & 0 & \ldots \\ 0 & 0 & 0 & 5 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} - \sigma \begin{bmatrix} \frac{1}{2} & 0 & 0 & \ldots \\ 0 & \frac{3}{2} & 0 & \ldots \\ 0 & 0 & \frac{5}{2} & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} 1 & -1 & 1 & -1 & \ldots \\ -1 & 1 & -1 & 1 & \ldots \\ 1 & -1 & 1 & -1 & \ldots \\ \vdots & \vdots & \vdots & \ddots & \ddots \end{bmatrix}$$

$$= -\frac{1}{2}\begin{bmatrix} \sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) & \sigma & \ldots \\ -3\sigma & 3\sigma & -3(2 + \sigma) & 3\sigma & -3(2 + \sigma) & \ldots \\ 5\sigma & -5\sigma & 5\sigma & -5(2 + \sigma) & 5\sigma & \ldots \\ -7\sigma & 7\sigma & -7\sigma & 7\sigma & -7(2 + \sigma) & \ldots \\ 9\sigma & -9\sigma & 9\sigma & -9\sigma & 9\sigma & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. $$

(16)

The parameter $\kappa$ affects only the last row of $\hat{D}_\kappa$. 9
Example 3.6. For \( n = 4 \) we have

\[
\hat{D}_\kappa = -\frac{1}{2} \begin{bmatrix}
\sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) \\
-3\sigma & 3\sigma & -3(2 + \sigma) & 3\sigma \\
5\sigma & -5\sigma & 5\sigma & -5(2 + \sigma) \\
-2\sigma/\kappa & 2\sigma/\kappa & -2\sigma/\kappa & 2\sigma/\kappa
\end{bmatrix}.
\]

The difference between \( \hat{D}_{LGR} (\kappa = 2/(2n - 1)) \) and \( \hat{D}_{LGL} (\kappa = 2/(n - 1)) \) is shown below:

\[
\hat{D}_{LGR} = -\frac{1}{2} \begin{bmatrix}
\sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) \\
-3\sigma & 3\sigma & -3(2 + \sigma) & 3\sigma \\
5\sigma & -5\sigma & 5\sigma & -5(2 + \sigma) \\
-7\sigma & 7\sigma & -7\sigma & 7\sigma
\end{bmatrix},
\]

\[
\hat{D}_{LGL} = -\frac{1}{2} \begin{bmatrix}
\sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) \\
-3\sigma & 3\sigma & -3(2 + \sigma) & 3\sigma \\
5\sigma & -5\sigma & 5\sigma & -5(2 + \sigma) \\
-3\sigma & 3\sigma & -3\sigma & 3\sigma
\end{bmatrix}.
\]

These two matrices are identical, except for the fourth row.

Remark 3.7. The SBP-SAT discretizations based on Legendre-Gauss-Radau and Legendre-Gauss quadratures have the same spectra, since the modal representation of \( \mathcal{D} \) is the same.

Figure 1 illustrates the spectrum of \( \mathcal{D} \) for the second order LGR discretization in (2), (3). The eigenvalues are identical to the ones of \( \hat{D}_{LGR} \) as is illustrated for several values of \( \sigma \).

4. Invertibility of pseudo-spectral first derivative discretizations

Before the invertibility for discretizations of (5) is discussed, we recall the following theorem [23, 24]

**Theorem 4.1.** Given a matrix \( A \in \mathbb{R}^{n \times n} \), suppose that \( H = \frac{1}{2}(GA + A^TG) \) and \( S = \frac{1}{2}(GA - A^TG) \) for some positive definite matrix \( G \) and some positive semidefinite matrix \( H \). Then \( A \) has eigenvalues with strictly positive real parts if, and only if, no eigenvector of \( G^{-1}S \) lies in the null space of \( H \).

Moreover, we will make use of the following
Numerical validation of similarity for LGR with $n = 3$

Figure 1: The spectrum of the second order Legendre-Gauss-Radau (LGR) operator is compared with the one of its modal representation $\tilde{D}_{\text{LGR}}$ (Modal) for $\sigma = -2, -1, -1/2, 1/2, 1$. Note that the eigenvalues coincide.

Lemma 4.2. The determinant of $\tilde{D}_\kappa$ is equal to $-(2n - 3)!! \sigma/\kappa$.

Proof. The matrix $\tilde{D}_\kappa$ can be rewritten as

$$\tilde{D}_\kappa = -\tilde{P}_\kappa^{-1} \begin{bmatrix} \sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) & \ldots \\ -\sigma & \sigma & -(2 + \sigma) & \sigma & \ldots \\ \sigma & -\sigma & \sigma & -(2 + \sigma) & \ldots \\ -\sigma & \sigma & -\sigma & \sigma & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} = -\tilde{P}_\kappa^{-1} \Sigma$$

and its determinant computed with the Cauchy-Binet formula for matrix products, i.e.

$$\det(\tilde{D}_\kappa) = (-1)^n (\det(\tilde{P}_\kappa))^{-1} \det(\Sigma). \quad (17)$$

It is straightforward to verify that the determinant of the diagonal matrix $\tilde{P}_\kappa$ is $2^{n-1} \kappa/(2n - 3)!!$. Gaussian elimination (without row switches) on $\Sigma,
gives
\[
\begin{bmatrix}
\sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) & \ldots \\
0 & -2 & -2 & -2 & \ddots \\
0 & 0 & -2 & -2 & \ddots \\
0 & 0 & 0 & -2 & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
\end{bmatrix}
\]

implying that \( \det(\Sigma) = (-2)^{n-1}\sigma \). As a consequence, (17) leads to \( \det(\hat{D}_\kappa) = -(2n - 3)!! \sigma/\kappa \).

Next, we prove

**Theorem 4.3.** Let \( \text{Re}(\lambda) \geq 0 \) and \( \sigma < -1/2 \). Pseudo-spectral SBP-SAT discretizations (9) of the initial value problem (5) are invertible and have eigenvalues with strictly positive real parts if the SBP norm \( P \) is based on a quadrature rule with order \( p \in \{2n - 3, 2n - 2, 2n - 1\} \).

**Proof.** Firstly, we find the inverse of
\[
\hat{D}_\kappa = -\frac{1}{2}
\begin{bmatrix}
\sigma & -(2 + \sigma) & \sigma & \ldots \\
-3\sigma & 3\sigma & -3(2 + \sigma) & \ddots \\
5\sigma & -5\sigma & 5\sigma & \ddots \\
\vdots & \ddots & \ddots & \ddots \\
(-1)^n\sigma(2n - 3) & (-1)^{n+1}\sigma(2n - 3) & (-1)^n\sigma(2n - 3) & \ddots \\
(-1)^{n+1}(2\sigma)/\kappa & (-1)^n(2\sigma)/\kappa & (-1)^n(2\sigma)/\kappa & \ddots \\
\end{bmatrix},
\]
i.e. the matrix \( R \) such that \( R\hat{D}_\kappa \) and \( \hat{D}_\kappa R \) give the identity matrix. Note that for \( \sigma \neq 0 \) the inverse of \( \hat{D}_\kappa \) exists, due to Lemma 4.2. For \( n = 4 \), the products between the two matrices
\[
\begin{bmatrix}
\sigma & -(2 + \sigma) & \sigma & -(2 + \sigma) \\
-3\sigma & 3\sigma & -3(2 + \sigma) & 3\sigma \\
5\sigma & -5\sigma & 5\sigma & -5(2 + \sigma) \\
-(2\sigma)/\kappa & (2\sigma)/\kappa & -(2\sigma)/\kappa & (2\sigma)/\kappa \\
\end{bmatrix},
\]
and
\[
R = \begin{bmatrix}
1 & -1/3 & \kappa(1 + \sigma)/\sigma \\
1 & 0 & -1/5 \\
1/3 & 0 & -\kappa/2 \\
1/5 & \kappa/2 \\
\end{bmatrix}
\]

12
give the identity matrix. A straightforward generalization yields that for a general \( n \), the matrix

\[
R = \begin{bmatrix}
1 & -\frac{1}{3} & \\
\frac{1}{3} & 0 & -\frac{1}{5} \\
\frac{1}{2n-7} & 0 & -\frac{1}{2n-3} \\
0 & 0 & -\frac{1}{2n-5} \\
\frac{1}{2n-3} & \frac{1}{2n-5} & -\frac{\kappa}{2} \\
\frac{1}{2n-3} & \frac{1}{2n-5} & -\frac{\kappa}{2}
\end{bmatrix}
\]

is the inverse of \( \hat{D}_\kappa \), i.e., \( \hat{D}_\kappa^{-1} = R \).

Secondly, due to Lemma 3.5, the matrix \( \hat{D}_\kappa^{-1} \) has the same eigenvalues as the inverse of \( D \). If \( \mu \in \mathbb{C} \) is an eigenvalue of \( D \), the corresponding eigenvalue of the inverse is \( \mu^{-1} = \overline{\mu}/|\mu|^2 \), where \( \overline{\mu} \) is the complex conjugate of \( \mu \). In particular, \( \text{Re}(1/\mu) = \text{Re}(\mu)/|\mu|^2 \) and this implies that the sign of the real parts of \( \mu \) is determined by the eigenvalues of \( \hat{D}_\kappa^{-1} \).

Thirdly, to conclude the proof, we show that the matrix \( \hat{D}_\kappa^{-1} \) have eigenvalues with strictly positive real parts for \( \sigma < -1/2 \). Recalling that matrix transposition does not modify the spectrum, we apply Theorem 4.1 to \( A = \hat{D}_\kappa^{-T} \) with \( G = 2\hat{P}_\kappa^{-1} = \text{diag}(1, 3, 5, \ldots, 2n - 3, 2/\kappa) \). The auxiliary matrix

\[
H = (\hat{P}_\kappa \hat{D}_\kappa)^{-T} + (\hat{P}_\kappa \hat{D}_\kappa)^{-1} = \begin{bmatrix}
1 & 0 & \cdots & 0 & (-1)^n \frac{1+\sigma}{\sigma} \\
0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 \\
(-1)^n \frac{1+\sigma}{\sigma} & 0 & \cdots & 0 & 1
\end{bmatrix}
\]

is positive semidefinite for \( \sigma < -1/2 \). Furthermore, it is clear that any vector \( \mathbf{v} = [v_1, v_2, \ldots, v_{n-1}, v_n]^T \) belonging to the null space of \( H \) is such that \( v_1 = v_n = 0 \). Hence, \( \mathbf{v} \) is not an eigenvector of

\[
G^{-1} S = \begin{bmatrix}
0 & 1 & (-1)^n \frac{1+\sigma}{\sigma} \\
-\frac{1}{3} & 0 & \frac{1}{3} \\
& \ddots & \ddots & \ddots \\
& \frac{1}{2n-3} & 0 & \frac{1}{2n-3} \\
(-1)^n \frac{1+\sigma}{\sigma} & \frac{1}{2n-3} & -\frac{\kappa}{2} & 0
\end{bmatrix}
\]
since \((G^{-1}S - tI)v = 0\) with \(v_1 = v_n = 0\) is only solved by \(v = 0\) for any \(t \in \mathbb{C}\). This proves that the matrix \(\tilde{D}_\kappa^{-1}\) has eigenvalues with strictly positive real parts for \(\sigma < -1/2\). The invertibility of the discrete system (9) follows directly from Proposition 2.4.

\(\square\)

Remark 4.4. Since we have not considered the distribution of the grid nodes in the proof, this statement holds in general for all discretizations which are based on quadrature rules with order \(p \geq 2n - 3\). Note that for nodes different from the ones prescribed by LGR, LG or LGL quadratures, the matrix \(P\) is not guaranteed to be diagonal.

5. An illustrative example

Consider the linearized one dimensional symmetrized form of the compressible Euler equations [25]

\[
U_t + AU_x = F, \quad 0 < x < 1, \quad 0 < t < T,
\]

\[
U(x, 0) = f, \quad 0 < x < 1,
\]

(18)

with subsonic characteristic boundary conditions [26]. In (18) we have

\[
U = \left[ \frac{\pi}{\sqrt{\gamma}} \bar{\rho}, \frac{1}{\sqrt{\gamma(\gamma-1)}} \bar{\theta} \right]^T,
\]

\[
A = \begin{bmatrix}
\frac{\pi}{\sqrt{\gamma}} & \frac{\pi}{\sqrt{\gamma}} & 0 \\
\frac{\pi}{\sqrt{\gamma}} & \frac{\pi}{\sqrt{\gamma}} & \sqrt{\gamma^{-1} \bar{c}} \\
0 & \sqrt{\gamma^{-1} \bar{c}} & \frac{\pi}{\sqrt{\gamma}}
\end{bmatrix}
\]

with \(\bar{\pi} = 1, \bar{c} = 2, \bar{\rho} = 1\) and \(\gamma = 1.4\). The variables \(\rho, u\) and \(\theta\) are the density, velocity and temperature perturbations of the fluid, respectively.

A fully-discrete approximation of (18) with a pseudo-spectral collocation approximation in time can be written as

\[
\left( \mathcal{D} \otimes I_x \right) + \frac{T}{2} \left( I_t \otimes \mathcal{L} \right) U = F,
\]

(19)

where \(\otimes\) denotes the Kronecker product defined by

\[
A \otimes B = \begin{bmatrix}
a_{11}B & \cdots & a_{1n}B \\
\vdots & \ddots & \vdots \\
a_{m1}B & \cdots & a_{mn}B
\end{bmatrix} \in \mathbb{R}^{mr \times ns}, \quad A = \{a_{ij}\} \in \mathbb{R}^{m \times n}, \quad B \in \mathbb{R}^{r \times s}.
\]
The discrete operator $\mathcal{L}$ approximates the spatial operator augmented with weak boundary conditions. The temporal discrete operator $\mathcal{D}$ includes the initial condition weakly.

If the spatial discretization is energy-stable, then $\mathcal{L}$ has eigenvalues $\lambda$ with $\text{Re}(\lambda) \geq 0$ [7]. In the previous section we have proved that the eigenvalues of $\mathcal{D}$, $\mu$, have strictly positive real parts. Moreover, it can be shown [27] that the eigenvalues of the matrix defining the fully-discrete approximation (19) are given by $\mu + (T/2)\lambda$. Hence, the eigenvalues of the fully-discrete operator have strictly positive real parts. This implies that the discrete problem is invertible, as predicted by Theorem 4.3, and it also indicates that a dual time-stepping procedure [28, 29, 30] applied to (19) converges. Eventually, the convergence will be delayed by transient growth, due to the nonnormality of the time operator $\mathcal{D}$ [31].

Figure 2 shows the spectrum of the spatial discrete operator $\mathcal{L}$, together with the spectrum of the fully-discrete problem (19). For the spatial discretization, we have considered the $(6, 3)$ finite difference SBP-SAT operators [32] with 51 equidistant nodes. The 4th order LGR pseudo-spectral method has been used in time with $T = 0.2$ and $\sigma = -1$. As predicted by our analysis, the fully-discrete operator is invertible and has eigenvalues with strictly positive real parts, well separated from the imaginary axis.
6. Conclusions

The invertibility of fully discrete approximations for initial boundary value problems has been discussed. We have proved that the SBP-SAT pseudo-spectral methods in time lead to invertible formulations for energy-stable spatial discretizations.

We have also shown that the eigenvalues of these fully discrete operators have strictly positive real parts, allowing for the use of iterative techniques such as dual time-stepping. These results hold in general, irrespectively of the number of grid nodes, if the quadrature used for the time approximation has order $p \geq 2n - 3$.

An example using the linearized compressible Euler equations, discretized with the 4th order LGR pseudo-spectral method in time and the (6, 3) finite-difference method in space, corroborates our findings.

7. Acknowledgments

This work was supported by VINNOVA, the Swedish Governmental Agency for Innovation Systems, under contract number 2013-01209.

References


