Summation-By-Parts in Time

Jan Nordström and Tomas Lundquist

Linköping University Post Print

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

Original Publication:

Copyright: Elsevier http://www.elsevier.com/

Postprint available at: Linköping University Electronic Press http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-94641
Summation-By-Parts in Time

Jan Nordström\textsuperscript{a}, Tomas Lundquist\textsuperscript{b}

\textsuperscript{a}Department of Mathematics, Computational Mathematics, Linköping University, SE-581 83 Linköping, Sweden (jan.nordstrom@liu.se).
\textsuperscript{b}Department of Mathematics, Computational Mathematics, Linköping University, SE-581 83 Linköping, Sweden (tomas.lundquist@liu.se).

Abstract

We develop a new high order accurate time-integration technique for initial value problems. We focus on problems that originate from a space approximation using high order finite difference methods on summation-by-parts form with weak boundary conditions, and extend that technique to the time-domain. The new time-integration method is global, high order accurate, unconditionally stable and together with the approximation in space, it generates optimally sharp fully discrete energy estimates. In particular, it is shown how stable fully discrete high order accurate approximations of the Maxwells’ equations, the elastic wave equations and the linearized Euler and Navier-Stokes equations can obtained. Even though we focus on finite difference approximations, we stress that the methodology is completely general and suitable for all semi-discrete energy-stable approximations. Numerical experiments show that the new technique is very accurate and has limited order reduction for stiff problems.

Keywords: time integration, initial value problems, high order accuracy, initial value boundary problems, boundary conditions, global methods, stability, convergence, summation-by-parts operators, stiff problems

1. Introduction

For time integration of non-stiff initial value problems (IVP), the time-step limitation is moderate and dictated by accuracy requirements only. Explicit methods such as various forms of Runge-Kutta or linear multi-step methods often suffice [1]. However, when the system of ordinary differential equations come from the spatial approximation of an initial boundary value problem (IBVP), it gets more complicated.
For systems coming from an IBVP there are two major complications and sometimes three. Firstly, the number of equations increase with increasing resolution of the spatial domain. Secondly, the ratio of the largest eigenvalue to the smallest eigenvalue often increases without bound. When this happens, the problem is called stiff. Stiffness can be generated by the physics considered, as in chemical reaction problems or problems with boundary layers or shocks. It can also be generated by the spatial approximation itself, and be due to non-uniform irregular meshes. A third major complication is non-linearity, often originating from the spatial approximation. Typical examples include the Navier-Stokes equations in fluid dynamics, the Black-Scholes equation in finance and the nonlinear Schrödinger equation in optics.

Stiffness, although hard to define [2], forces the use of implicit methods in order to reduce the stability requirements on the time-step. Methods such as the BDF (backward differentiation) methods [3],[4], implicit Runge-Kutta methods [5],[6], linear multi-step methods [7],[8] and various types of general linear methods [9],[10] are used. Both single and multi-step as well as multi-stage methods exist. Roughly speaking, the linear multi-step methods are cheap and efficient but lacks certain stability properties. On the other hand, implicit Runge-Kutta methods have good stability and accuracy properties but can be very expensive. Often, the efficiency can be increased if combinations of implicit and explicit methods, so called IMEX methods [11],[12] are used.

All the previously mentioned methods are local, i.e. the solution at the next time level is computed by using one or a few previously computed time levels. In global methods, the whole time interval from zero to the final time $T$ is considered. Global methods using collocation and spectral approximations have been considered previously (see [13],[14],[15],[16]) but have often been considered unpractical. However, the unconditional stability in combination with the very high accuracy cannot be matched by the local methods. Also, energy estimates which precisely match the continuous estimates can be obtained. This is seldom (if ever) possible for local methods.

The goal of this paper is to present a new high order accurate time-integration technique for IVP. The new time-approximation method is global, unconditionally stable and together with the approximation in space, it generates optimally sharp fully discrete energy estimates. The methodology is highly accurate for both stiff and non-stiff problems. In most cases we consider an IBVP that is discretized in space with high order finite difference methods on summation-by-parts (SBP) form complemented with weak
boundary conditions using the simultaneous approximation term method (SAT). Even though we focus on problems discretized by the SBP-SAT technique, we stress the methodology is completely general and suitable for all semi-discrete energy stable problems.

SBP operators [17, 18, 19, 20] mimic integration by parts perfectly. Given the SBP discretisation, the boundary conditions are imposed weakly using penalty terms in the SAT method [21, 19, 22, 23]. The combination of this technique together with well posed boundary conditions for the IBVP guarantees semi-discrete stability via the energy-method. For application of the SBP-SAT technique to high order finite difference methods see [24, 22, 23, 25, 26, 27, 28, 29, 30, 31] where many different problems including fluid flow, wave propagation and conjugate heat transfer have been considered. In the sequel of this paper we will assume that the reader is reasonably familiar with the SBP-SAT technique presented in the references above.

In this paper we will explore the use of this technique in time. The stability and the spectral content as well as the particular organisation of the technique applied to IBVPs will be considered. We will limit ourselves to constant coefficient problems in this initial study. In particular we will show that together with the energy-stable semi-discrete approximations in [24, 22, 23, 25, 26, 27, 28, 29, 30, 31], it leads to sharp fully discrete energy estimates. As was mentioned above, the methodology is completely general and suitable for all semi-discrete energy stable problems, not only the ones discretized by the SBP-SAT technique in space.

The paper is organized as follows. In Section 2 we deal with the scalar initial value problem. Optimal energy estimates are derived and the spectrum of the operator is discussed. Section 3 deals with the application of the technique to a representative scalar initial boundary value problem. In Section 4 we generalize the one-dimensional theory for a scalar partial differential equation developed in Section 3 to multiple dimensions and systems of partial differential equations. Finally in Section 5 we draw conclusions.

2. The initial value problem

We start by discussing the discretisation of initial value problems.

2.1. The continuous energy estimate

Consider the simplest possible first order initial value problem

\[ u_t = \lambda u, \]  \hspace{1cm} (1)
with initial condition \(u(0) = f\) and \(0 \leq t \leq T\). Let us consider the complex constant \(\lambda\) to represent an energy stable spatial discretisation of an IBVP. The stability implies that \(\lambda\) has a negative semi-definite real part. For hyperbolic problems, \(\lambda\) is proportional to the inverse of the mesh size, and for parabolic problems, the mesh size squared.

The energy method (multiplying with the complex conjugated solution and integrating over the domain) applied to (1) yields
\[
|u(T)|^2 - 2\text{Re}(\lambda)|u|^2 = |f|^2,
\]
where \(|u|^2 = \int_0^T |u|^2 dt\). Note that the solution at the final time is bounded in terms of the initial data. If \(\text{Re}(\lambda) < 0\), also the norm of the solution is bounded.

2.2. The discrete energy estimate

The SBP-SAT approximation of (1) reads
\[
P^{-1}Q\tilde{U} = \lambda\tilde{U} + P^{-1}(\sigma(U_0 - f))e_0.
\]
The vector \(\tilde{U}\) contains the numerical approximation of \(u\) at all grid points in time. The matrices \(P, Q\) form the differentiation matrix \(D = P^{-1}Q\). The SBP properties are
\[
P = P^T > 0, \quad Q + Q^T = E_N - E_0,
\]
where \(E_0 = \text{diag}(1, 0, \ldots, 0), E_N = \text{diag}(0, \ldots, 0, 1)\). The difference operators can be based on block norms \(P\) [18] for full accuracy but sometimes diagonal versions with lower accuracy must be used [32],[33],[34]. The norm matrix is of the form \(P = \Delta t\tilde{P}\) where \(\tilde{P}\) has entries of order one. As an example, consider the second order operator with the matrices \(P\) and \(Q\) as
\[
P = \Delta t\begin{bmatrix}
\frac{1}{2} & 1 & \cdots & 1 \\
1 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & 1 \\
\frac{1}{2} & \cdots & 1 & \frac{1}{2}
\end{bmatrix}, \quad Q = \begin{bmatrix}
-\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \cdots & \cdots \\
\frac{1}{2} & 0 & \frac{1}{2} & \cdots & \cdots \\
\frac{1}{2} & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{2} & \cdots & \cdots & \cdots & \cdots \\
\frac{1}{2} & \cdots & \cdots & \cdots & \cdots
\end{bmatrix}.
\]
The extra (penalty) term on the right-hand-side of (3) enforces the initial condition weakly using the SAT technique and position it at grid point zero.
by the unit vector \( \vec{e}_0 = (1,0,...,0,0)^T \). The penalty parameter \( \sigma \) will be decided by stability requirements.

**Remark 1:** The penalty term in (3) forces the discrete solution towards the initial data, i.e. \( U_0 \neq f \) in general, but it is close. This technique is used in order to preserve the SBP properties of the difference operator \( D = P^{-1}Q \) which is necessary for the stability proof.

The discrete energy method applied to (3) (multiplying from the left with \( \vec{U}^* P \) and using the SBP properties (4)) leads to

\[
|\vec{U}_N|^2 - 2Re(\lambda)||\vec{U}||_P^2 = (1 + 2\sigma)|\vec{U}_0|^2 - \sigma(U_0f + U_0\bar{f}),
\]

(6)

In (6), the overbar denotes a complex conjugated quantity, \( \vec{U}^* \) is the complex conjugate of \( \vec{U}^T \), and \( ||\vec{U}||_P^2 = \vec{U}^* P \vec{U} \). The method is obviously stable for \( \sigma \leq -1/2 \). By adding and subtracting \( |f|^2 \) to the right hand side of (6) and making the choice \( \sigma = -1 \) we obtain

\[
|\vec{U}_N|^2 - 2Re(\lambda)||\vec{U}||_P^2 = |f|^2 - |U_0 - f|^2.
\]

(7)

The choice \( \sigma = -1 \) also makes the discretization dual consistent \([35],[36],[37]\). By comparing the continuous estimate (2) with (7) we see that the discrete bound is slightly more strict than the continuous counterpart due to the term \(-|U_0 - f|^2 \) (which goes to zero with increasing accuracy).

**Remark 2:** Note that the estimate (7) is independent of the size of the time-step, i.e. the method is unconditionally stable.

**Remark 3:** Sharp estimates like (7) can hardly be obtained using conventional local methods where only a few time levels are involved. One can argue, although no proof exist, that it can be done only with global methods.

### 2.3. The spectrum of the IVP

By rearranging (3), the final equation to solve for \( \vec{U} \) becomes

\[
(P^{-1}\tilde{Q} - \lambda I)\vec{U} = \vec{R},
\]

(8)

where \( \tilde{Q} = Q - \sigma E_0 \) and \( \vec{R} = -\sigma P^{-1}f\vec{e}_0 \). For example, the second order operator given by (5), leads to the matrix

\[
(P^{-1}\tilde{Q} - \lambda I) = \begin{bmatrix}
-\frac{1-2\sigma}{\Delta x} - \lambda & \frac{1}{\Delta x} \\
-\frac{1}{2\Delta x} - \lambda & \frac{1}{2\Delta x} \\
\vdots & \vdots \\
-\frac{1}{\Delta x} - \lambda & \frac{1}{\Delta x} \\
-\frac{1}{\Delta x} & \frac{1}{\Delta x} - \lambda
\end{bmatrix}.
\]
The matrix $P^{-1}\tilde{Q} - \lambda I$ must be non-singular for a well functioning procedure, which is guaranteed if $P^{-1}\tilde{Q}$ has eigenvalues with strictly positive real parts. The following Lemma is useful.

**Lemma 1.** Let the matrix $B$ be symmetric positive definite and the matrix $A$ have a positive semi-definite symmetric part. Then, the matrix $B^{-1}A$ has eigenvalues with positive semi-definite real parts.

**Proof.** Let $\lambda$ and $x$ be an eigenpair to $B^{-1}A$, i.e. $B^{-1}Ax = \lambda x$. Elementary manipulations lead to $\text{Re}(\lambda) = x^*(A + A^T)x/(2x^*Bx) \geq 0$. □

We can prove the following Proposition.

**Proposition 1.** Let the SBP matrix $Q$ be defined by (4). Then $P^{-1}\tilde{Q}$ with $\tilde{Q} = Q - \sigma E_0$ in the second order case (SBP(2,1)) has eigenvalues with strictly positive real parts for $\sigma < -1/2$.

**Proof.** Consider the eigenvalue problem $\tilde{Q}\vec{x} = \lambda P\vec{x}$. The proof of Lemma 1 leads directly to $\text{Re}(\lambda) = (|x_N|^2 - (1 + 2\sigma)|x_0|^2)/(2||\vec{x}||_P^2)$. Consequently, for stable approximations ($(1+2\sigma) < 0$) we have no eigenvalues on the imaginary axis if $x_0, x_N \neq 0$. Next we assume that the eigenvector is of the form $\vec{x} = (0, \tilde{x}, 0)^T$ with a corresponding imaginary eigenvalue $i\xi$. The reduced eigenvalue problem now becomes $\tilde{Q}\tilde{x} = \lambda P\tilde{x}$, where $\tilde{Q}, P$ have $N+1$ rows and $N-1$ columns. This results in a system with $N+1$ equations and $N$ unknowns ($\tilde{x}$ and $\xi$). For the second order SBP operator, the reduced eigenvalue problem can be solved analytically and the eigenvector $\tilde{x}$ is identically zero for all $\xi$. Hence, no eigenvalues on the imaginary axis. □

For higher order SBP operators we have not been able to solve the reduced eigenvalue problem analytically (due to the algebraic complexity) and show that the eigenvalues have strictly positive real parts for $\sigma < -1/2$. However, in Figure 1 we see the eigenvalues for the spectrum of $P^{-1}\tilde{Q}$ with $\sigma = -1/2$ (which is right on the stability limit) for second, fourth, sixth and eighth order operators with diagonal norms. The eigenvalues for all the different SBP operators are distinctly separated from zero, making the matrix in (8) non-singular. Furthermore, we have through extensive computational investigations shown that all SBP operators have eigenvalues in the right half plane for $\sigma > -1/2$. Based on the computational results and Proposition 1 for the second order case, we make the following assumption for all the higher order SBP operators.
Assumption 1: Let $P$ and $Q$ be defined by (4). Then $P^{-1}\tilde{Q}$ with $\tilde{Q} = Q - \sigma E_0$ has eigenvalues with strictly positive real parts for $\sigma < -1/2$.

2.4. Numerical calculations for initial value problems

In this section we investigate the performance of the $SBP-SAT$ method described above for scalar initial value problems. We use both a non-stiff and a stiff test equation. The stiff case is especially interesting since some popular A-stable methods experience so called order reduction when applied to stiff problems. Order reduction might occur for operators with eigenvalues $\lambda$ such that $|\lambda\Delta t| >> 1$ while $\Delta t << 1$, see [38]. Various definitions of stiffness exist, the most common one simply states that stiffness occurs if the largest time step guaranteeing stability for an explicit method is larger than the step size needed for the local discretization error to be small enough [2]. This pragmatic definition will be sufficient for our needs in this section.

For the stiff case we compare the SBP based methods with the second order implicit backward differentiation formula (BDF2, see [3],[4]), as well as a fourth order explicit singly diagonally implicit Runge-Kutta method, (ES-DIRK4, see [5]). We use SBP operators with diagonal norms as well as full block norms. Operators with discretization error of order $2s$ in the interior have order $s$ at the boundaries in the case of diagonal norms and $2s - 1$ in the case of full block norms [18]. The corresponding SBP-SAT methods are
denoted SBP(2s,s) and SBP(2s,2s − 1) respectively. In this initial investigation we focus on accuracy and stability and in all the numerical calculations we use the standard direct linear equation solver ("backslash") in MATLAB, which implements a sparse LU-factorization algorithm.

2.4.1. Numerical calculations for non-stiff problems

As a non-stiff test equation we consider the following

\[ u'(t) = -u(t), \quad 0 < t \leq 1 \]
\[ u(0) = 1. \]  
(9)

The exact solution to this problem is \( u(t) = e^{-t} \). Figure 2 shows the global convergence at \( t = 1 \). The rate of convergence for all \( SBP - SAT \) methods is approximately \( 2s \), which suggests that it is independent of the lower order at the boundaries.
2.4.2. Numerical calculations for stiff problems

To investigate the stiff case we use the following equation with the same exact solution \( u(t) = e^{-t} \) as before:

\[
\begin{align*}
  u'(t) &= \lambda(u(t) - e^{-t}) - e^{-t}, \quad 0 < t \leq 1 \\
  u(0) &= 1.
\end{align*}
\]  

(10)

Using a test equation on this form with \( \lambda \Delta t \ll -1 \) to characterize stiffness was first proposed in [38]. In the numerical calculations we used \( \lambda = -1000 \). As can be seen in figures 3 and 4, the convergence rate of the \( SBP-SAT \) methods is reduced to the order of accuracy at the boundaries, rather than the order in the interior as in the non-stiff case. The order reduction is therefore less severe for block norms, for whom only one order is lost.

Note that \( ESDIRK4 \) starts with a convergence rate close to two, which coincides with the so called stage order of the method. The order of convergence in stiff cases for implicit Runge-Kutta methods is often limited by the orders of the intermediate stages rather than by the design order of accuracy [39]. To our knowledge, no Runge-Kutta method exist which combine the diagonally implicit structure with a stage order higher than two. Linear multistep methods like \( BDF2 \) do not have a problem with order reduction, but they are on the other hand not A-stable for higher orders than two, making them less attractive in practice.

2.4.3. Preliminary conclusions based on the numerical calculations

Our calculations show that the convergence rate of the \( SBP(2s, s) \) and \( SBP(2s, 2s-1) \) operators is \( 2s \) in the non-stiff case, i.e. the same as the order of accuracy in the interior of the domain. In the stiff case, the convergence rate is reduced to \( 2s \) and \( 2s-1 \) respectively, i.e. to the order of accuracy at the boundaries of the domain. The optimal stability property of the \( SBP-SAT \) technique combined with the high orders of convergence for the stiff case, especially for \( SBP(2s, 2s-1) \), is a very promising feature of these methods.

3. The initial boundary value problem for a scalar equation

Here we apply the new technique to initial boundary value problems.
3.1. Preliminaries

We consider numerical approximations of well-posed partial differential equations (PDE’s) on the general form

\[ U_t + P^{-1} RU = P^{-1} G \]

\[ U(0) = F, \]

(11)

where \( P^{-1} R \) is an approximation of the spatial part of the PDE, \( P \) is the norm or mass matrix, \( R \) is a general operator and \( G \) denotes the generalized boundary data. \( G \) includes a possible forcing function in the original PDE and \( F \) is the initial data. \( P \) is symmetric and positive definite. We now make the following Definition.

Definition 1. The approximation (11) is energy stable if \( R + R^T \geq 0 \).

The definition is easy to understand since the energy method (multiply from the left with \( U^T P \)) applied to (11) with \( G = 0 \) leads by the use of \( R + R^T \geq 0 \) directly to the estimate \( U^T P U \leq F^T P F \).
Remark 4: Note that many numerical methods using weak boundary conditions such as the SBP-SAT technique for finite differences, the spectral element method, the finite volume method, the finite element method and the discontinuous Galerkin method have the general form (11).

Proposition 2. Energy stable schemes of the form (11) have the spectrum in the right half of the complex plane, possibly including the imaginary axis.

Proof. Definition 1 together with Lemma 1 leads directly to the result. \(\square\)

3.2. The continuous energy estimate

As example, we consider the one-dimensional advection-diffusion equation

\[
\begin{align*}
    u_t + au_x - \epsilon u_{xx} &= 0, & 0 \leq x \leq 1, & t \geq 0 \\
    au - \epsilon u_x &= g_0(t), & x = 0, & t \geq 0 \\
    \epsilon u_x &= g_1(t), & x = 1, & t \geq 0 \\
    u &= f(x), & 0 \leq x \leq 1, & t = 0,
\end{align*}
\]

where the boundary data \(g_0, g_1\) and the initial function \(f\) are the data of the problem and \(a, \epsilon > 0\).
By multiplying (12) with \( u \) and integrating in space we obtain,
\[
||u||_t^2 + 2\epsilon||u_x||_t^2 = a^{-1} \left[ (au - \epsilon u_x)^2 - (\epsilon u_x)^2 \right]_x = a^{-1} \left[ (au - \epsilon u_x)^2 - (\epsilon u_x)^2 \right]_x = 0
\]
where \( ||u||_t^2 = \int_0^1 u^2 dx \) and \( ||u_x||_t^2 = \int_0^1 u_x^2 dx \). Next we insert the boundary conditions and arrive at the continuous energy rate
\[
||u||_t^2 + 2\epsilon||u_x||_t^2 = a^{-1} \left[ g_0^2 + g_1^2 - (au - g_0)^2 - (au - g_1)^2 \right]. \quad (13)
\]
Finally, by time integration, we have the result for the continuous problem
\[
||u(\cdot, T)||_t^2 + 2\epsilon \int_0^T ||u_x(\cdot, t)||_t^2 dt = ||f||_t^2 + a^{-1} \int_0^T [g_0^2 + g_1^2] dt - a^{-1} \int_0^T [(au - g_0)^2 + (au - g_1)^2] dt. \quad (14)
\]
Note that the norm of the solution at the final time and the time integral of the first derivative is bounded by initial data and boundary data. The estimate (14) is the target for the fully discrete energy estimate that we will derive below.

3.3. The semi-discrete energy estimate

The semi-discrete approximation of (12) using the SBP-SAT technique is
\[
U_t + aDU - \epsilon D^2U = P^{-1}(\sigma_0(L_0U - g_0)\hat{e}_0 + \sigma_N(L_NU - g_1)\hat{e}_N)
\]
\[
U(0) = F_0,
\]
where \( D = P^{-1}Q, \) \( L_0U = aU_0 - \epsilon(DU)_0, \) \( L_NU = \epsilon(DU)_N \) and \( \hat{e}_N = (0, 0, ..., 0, 1)^T. \) The vector \( U(t) = (U_0(t), U_1(t), ..., U_{N-1}, U_N(t))^T \) contains the numerical approximation of \( u \) at all grid points in space and \( F_0 = (f_0, f_1, ..., f_{N-1}, f_N)^T. \) The right hand side of (15) implements the boundary conditions weakly using the SAT technique.

By multiplying (15) with \( U^T P \) from the left, using the SBP properties (4) and \( \sigma_0 = \sigma_N = -1 \) we obtain the semi-discrete energy rate
\[
||U||_t^2 + 2\epsilon||DU||_t^2 = a^{-1} \left[ g_0^2 + g_1^2 - (aU_0 - g_0)^2 - (aU_N - g_1)^2 \right], \quad (16)
\]
where \( ||U||_t^2 = U^T P U \) and \( ||DU||_t^2 = (DU)^T P(DU) \). Note that the semi-discrete energy rate (16) is almost identical to the corresponding continuous (13) one. Finally, by time integration, we have the result for the semi-discrete problem
\[
||U(\cdot, T)||_t^2 + 2\epsilon \int_0^T ||DU(\cdot, t)||_t^2 dt = ||F_0||_t^2 + a^{-1} \int_0^T [g_0^2 + g_1^2] dt - a^{-1} \int_0^T [(aU_0 - g_0)^2 + (aU_N - g_1)^2] dt. \quad (17)
\]
The similarity of the semi-discrete estimate (17) with the continuous estimate (14) is striking and demonstrates the strength of the SBP-SAT method in space.

For future use, we rearrange the formulation (15) with $\sigma_0 = \sigma_N = -1$ to

\[
U_t + P^{-1}R_x U = P^{-1}G \\
U(0) = F_0,
\]

(18)

where

\[
R_x = a(Q + E_0) - \epsilon(Q + E_0 - E_1)D, \quad G = (g_0, 0, ..., 0, g_1)^T.
\]

(19)

We have showed in (16) that (let $g_0 = g_1 = 0$)

\[
R_x + R_x^T = a(E_0 + E_N) + 2\epsilon D^T P D \geq 0.
\]

(20)

Clearly now the approximation (18) is energy stable since Definition 1 holds.

3.4. The fully discrete energy estimate

Here it is convenient to introduce the Kronecker product $A \otimes B$ for arbitrary matrices $A \in R^{m \times n}$ and $B \in R^{p \times q}$ as

\[
A \otimes B = \begin{bmatrix}
a_{1,1}B & \hdots & a_{1,m}B \\
\vdots & \ddots & \vdots \\
a_{n,1}B & \hdots & a_{m,n}B
\end{bmatrix}.
\]

(21)

The Kronecker product is bilinear, associative and obeys the mixed product property

\[
(A \otimes B) (C \otimes D) = (AC \otimes BD)
\]

(22)

if the usual matrix products are defined. For inversion and transposing we have

\[
(A \otimes B)^{-1,T} = A^{-1,T} \otimes B^{-1,T}
\]

(23)

if the usual matrix inverse is defined.

The fully discrete version of (12) is obtained by discretizing (15) or equivalently (18) in time using the SBP-SAT technique. The use of the Kronecker product rules (21-23) and (19) yield

\[
(P_t^{-1}Q_t \otimes I_x)U + (I_t \otimes P_x^{-1}R_x)U = (I_t \otimes P_x^{-1})G \\
+ \sigma_t(P_t^{-1}E_0 \otimes I_x)(U - F),
\]

(24)
where the first index correspond to time and the second to space. We have indicated the operators and vectors that belong to \( t, x \) with subscripts where appropriate. The second penalty term on the right hand side include the unknown coefficient \( \sigma_t \) which will be determined for stability. The organisation of the vectors in (24) are

\[
\begin{align*}
U &= (U_0, U_1, \ldots, U_M)^T, \quad U_i = (U_{i0}, U_{i1}, \ldots, U_{iN})^T \\
G &= (G_0, G_1, \ldots, G_M)^T, \quad G_i = (g_0(i\Delta t), 0, \ldots, 0, g_1(i\Delta t))^T \\
F &= (F_0, U_1, \ldots, U_M)^T, \quad F_0 = (f_0, f_1, \ldots, f_N)^T \\
U^0 &= (U_{00}, U_{10}, \ldots, U_{M0})^T, \quad G^0 = (g_0(0), g_0(\Delta t), \ldots, g_0(M\Delta t))^T \\
U^N &= (U_{0N}, U_{1N}, \ldots, U_{MN})^T, \quad G^1 = (g_1(0), g_1(\Delta t), \ldots, g_1(M\Delta t))^T.
\end{align*}
\]

By multiplying (24) with \( U^T(P_t \otimes P_x) \) from the left, using the SBP properties (4), the relation (18), the Kronecker product rules (21-23) and the choice \( \sigma_t = -1 \) we obtain

\[
\begin{align*}
U^T P_x U_M + 2\epsilon(DU)^T(P_t \otimes P_x)DU &= F_0^T P_x F_0 \\
+ a^{-1} \left[ (G^0)^T P_x G^0 + (G^1)^T P_x G^1 \right] \\
+ a^{-1}(aU^0 - G^0)^T P_x (aU^0 - G^0) \\
+ a^{-1}(aU^N - G^1)^T P_x (aU^N - G^1) \\
+ (U_0 - F_0)^T P_x (U_0 - F_0).
\end{align*}
\]

Note the close similarity between the continuous estimate (14), the semidiscrete estimate (17) and the fully discrete one in (25). The fully discrete estimate has the additional damping term \(-(U_0 - F_0)^T P_x (U_0 - F_0)\) from the time-discretisation, also present in (7).

By rearranging (24), we get the final equation to solve for \( U \)

\[
BU = (B_t + B_x)U = \left[ (P_t^{-1} \tilde{Q}_t \otimes I_x) + (I_t \otimes P_x^{-1} R_x) \right] U = H, \tag{26}
\]

where \( \tilde{Q}_t = Q_t - \sigma_t E_0, R_x \) is defined in (19) and \( H = (I_t \otimes P_x^{-1}) G - \sigma_t (P_t^{-1} E_0 \otimes I_x) F \) is the data vector.

For example, by using the second order operator in (5), the matrix \( B \) in (26) becomes:

\[
B = \begin{bmatrix}
-\frac{1-\epsilon}{\Delta x} I_x + P_x^{-1} R_x & 1 \frac{1}{\Delta x} I_x \\
-\frac{1}{2\Delta x} I_x & P_x^{-1} R_x & 1 \frac{1}{2\Delta x} I_x \\
& & \ddots & \ddots & \ddots \\
& & -\frac{1}{2\Delta x} I_x & P_x^{-1} R_x & 1 \frac{1}{2\Delta x} I_x \\
& & & & -\frac{1}{\Delta x} I_x & 1 \frac{1}{\Delta x} I_x + P_x^{-1} R_x
\end{bmatrix}.
\]
For higher order of accuracy in time, the banded structure of $B$ becomes wider and with higher order in space, the bandwidth of $R_x$ becomes larger. The size of the blocks $I_x$ and $P_x^{-1}R_x$ increase with higher resolution in space.

The following theorem will be needed.

**Theorem 1.** Let $A$ and $B$ be two matrices. If $A$ and $B$ commute, they are simultaneously triangularizable, i.e. there exists a unitary matrix $X$ such that $A = XT_A X^*$ and $B = XT_B X^*$, where $T_A$ and $T_B$ are both upper triangular.

**Proof.** See for example the proof of theorem 40.4 in [40].

We will also need the following Lemma.

**Lemma 2.** The matrices $B_t$ and $B_x$ commute, i.e. $B_t B_x = B_x B_t$.

**Proof.** Let $C_t = P_t^{-1} \tilde{Q}_t$ and $C_x = P_x^{-1} R_x$. We have $B_t B_x = (C_t \otimes I_x)(I_t \otimes C_x) = (C_t \otimes C_x)(I_t \otimes I_x) = B_x B_t$.

We are now ready to show

**Proposition 3.** Given that Assumption 1 holds, then the matrix $B = B_t + B_x$ in (26) have eigenvalues with strictly positive real parts.

**Proof.** Theorem 1 and Lemma 2 leads to $B = B_t + B_x = XT_t X^{-1} + XT_x X^{-1} = X(T_t + T_x)X^{-1}$ where $X$ is unitary, and $T_t$ and $T_x$ are (upper) triangular matrices. Any eigenvalue $\lambda$ of $B$ can thus be written as $\lambda = \lambda_t + \lambda_x$, where $\lambda_t$ and $\lambda_x$ are eigenvalues of $B_t$ and $B_x$ respectively. The relation (20) and Lemma 1 show that the eigenvalues of $B_x$ have positive semi-definite real parts. If Assumption 1 holds, then the eigenvalues of $B_t$ have strictly positive real parts and the result follows.

**Remark 5:** Proposition 3 show that (26) has a unique bounded solution.

### 3.5. Numerical calculations for the initial boundary value problems

We use the method of manufactured solutions, see [41],[42],[43] to employ the exact solution $u = \sin(2\pi(x - t))$ for the advection-diffusion problem (12). The boundary conditions are imposed with the standard $SBP-SAT$ technique and data from the exact solution. As was mentioned above, in the present calculations, we used the linear equations solver ("backslash") in MATLAB.
As a non-stiff case we use the parameters $a = 1$, $\epsilon = 0.01$. We use a second order coarse ($\Delta x = 1/16$) spatial discretization. Figure 5 shows the eigenvalue distribution of the semi-discrete problem, while Figure 7 shows convergence to the exact solution of the semi-discrete problem. The exact solution of the semi-discrete problem was obtained by using a very fine time integration scheme. The convergence rates are approximately $2s$, just as we saw with the non-stiff initial value problem in section 2.4.1.

Next we consider a stiff IBVP calculation by choosing $a = 1$ and $\epsilon = 10$. This time we use an eighth order, high resolution ($\Delta x = 0.01$) spatial discretization to make sure that the error in the fully discrete solution is dominated by the temporal discretization. Figure 6 shows the semi-discrete spectrum of this problem and it can be seen by comparing with the spectrum in Figure 5 that the problem is indeed stiff. Figures 8 and 9 show the convergence to the exact solution. We see that ESDIRK4 shows order reduction also in this case, especially in the lower range of temporal resolution. The $SBP(2s, s)$ methods show convergence rates between $s$ and $s + 1$ for lower resolution, while the $SBP(2s, 2s - 1)$ methods converge with orders between...
Figure 6: The spectrum of the stiff advection-diffusion problem using an eighth order spatial discretization with $\Delta x = 0.01$. The real part of the eigenvalues are strictly positive.

2s – 1 and 2s. In the latter case, the order reduction is minor for the higher order methods, if at all present. Note also the similarities between these results and those in section 2.4.2.

We conclude that block norms should be favored over diagonal norms whenever possible, due to the relatively minor order reduction experienced in the stiff case. We also stress that the practical efficiency of the $SBP-SAT$ technique will greatly depend on the technique used to solve the large linear equation system arising from the spatial and temporal discretization.

4. The initial boundary value problem for systems of equations

The stability theory in section 3.4 above can be extended to all energy stable semi-discrete systems of equations in multiple dimensions. Such energy stable semi-discrete approximations can for example be found in [24, 22, 23, 25, 26, 27, 28, 29, 30, 31] where the SBP-SAT technique in space was used. However, the methodology is completely general and suitable for all semi-discrete energy stable problems. We consider formulations on the form (11) such that we have an energy stable approximation, see Definition 1. The
ambition in this section is to give an overview of the general stability theory, and hence some details will not be scrutinized.

The multi-dimensional fully discrete approximation analogous to (24) is

\[(P^{-1}_t Q_t \otimes I_s) U + (I_t \otimes P^{-1} R) U = (I_t \otimes P^{-1}) G + \sigma_t (P^{-1}_t E_0 \otimes I_s) (U - F). \tag{27}\]

The first index correspond to time and the second one is a multi-index corresponding to the number of dimensions in space and the number of equations in the system. The vectors \(G\) and \(F\) are the boundary and initial data organized in an appropriate way (\(F\) now contains the initial data \(F_0\)). The matrix \(I_s\) is the identity matrix for the multi-index. The energy method applied to (27) and the choice \(\sigma_t = -1\) yield

\[U_M^T P_s U_M \leq F_0^T P_s F_0 - (U_0 - F_0)^T P_s (U_0 - F_0), \tag{28}\]

which correspond to (25) in the fully discrete one-dimensional case.

**Remark 6:** The estimate (28) above means that systems like the Maxwells’ equations, the elastic wave equations and the linearized Euler and Navier-Stokes equations can be shown to be stable for fully discrete high order ap-
proximations. **Stability is obtained in an almost automatic way if the systems are energy stable in a semi-discrete sense.**

5. Conclusions

We develop a new high order accurate time-integration technique for initial value problems by extending the well known SBP-SAT technique for space discretisation into the time domain. We use summation-by-parts operators in time and a weak initial condition.

The new time-integration method is global, high order accurate, unconditionally stable and together with energy stable semi-discrete approximations, it generates optimal fully discrete energy estimates. Even though we focus on finite difference approximations, we stress that the methodology is completely general and suitable for all semi-discrete energy-stable approximations.

We have derived optimal energy estimates for the scalar initial value problem, the scalar advection-diffusion problem and done numerical experiments. The experiments verify that the SBP-SAT schemes in time are comparable to and sometimes better than other popular methods.
Figure 9: Temporal convergence in the discrete $L_2$ norm at $t = 1$ for the stiff advection-diffusion equation. Results for BDF2, ESDIRK4 and SBP(2s,2s-1) operators are shown.

The theoretical work on the initial value problem and the scalar advection-diffusion problem was generalized to energy stable multi-dimensional system problems such as the Maxwells’ equations, the elastic wave equations and the linearized Euler and Navier-Stokes equations. It was shown how fully discrete energy estimates for high order approximations can be obtained in an almost automatic way.

In this paper we focused on the excellent stability properties of the method. Future work will include work on how to construct an efficient solution procedure. Also, we have considered constant coefficient problems. Time-dependent coefficients and nonlinear problems will be a future topic as well.

References


