Summation-By-Parts Operators for Time Discretisation: Initial Investigations

Jan Nordström [a,*] , Tomas Lundquist [a]

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

LiTH - MAT - R - - 2012 / 08 - - SE

Abstract

We develop a new high order accurate time-discretisation technique for initial value problems. We focus on problems that originate from a space discretisation 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-discretisation method is global and together with the approximation in space, it generates optimal fully discrete energy estimates, and efficient methods for both stiff and non-stiff problems. In particular, it is shown how stable fully discrete high order accurate approximations of the Maxwells' equations, the elastic wave equations and the linearised Euler and Navier-Stokes equations are 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.

Key words: time integration, initial value problems, weak initial conditions, high order accuracy, initial value boundary problems, weak boundary conditions, global methods, stability, convergence, summation-by-parts operators, energy estimates, 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 [5]. However, when the system of ordinary differential equations

* Corresponding author, *Email adress*: Jan.Nordstrom@liu.se

come from the spatial discretisation of an initial boundary value problem (IBVP), it gets more complicated.

For systems coming from 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 itself, as in chemical reaction problems or problems with boundary layers or shocks. It can also be generated by the spatial discretisation itself, and be due to nonuniform irregular meshes. A third major complication is non-linearity, often originating from the spatial discretisation. 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 [30], 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 [10],[9], implicit Runge-Kutta methods [19],[7], linear multi-step methods [16],[15] and various types of general linear methods [3],[4] 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 [17],[18] are used, so called IMEX methods.

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 final time T is considered. Global methods using collocation and spectral approximations have been considered previously (see [1],[11],[13],[34]) but are often 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 develop a new high order accurate time-discretisation technique for IVP. We aim for methods that are efficient for both stiff and non-stiff problems, but focus on the stiff case. In most cases we consider IBVP that are discretised 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 discretised by the SBP-SAT technique, we stress the the methodology is completely general and suitable for all semi-discrete

stable problems.

SBP operators [21,31,8,23] mimic integration by parts perfectly. Given the SBP discretisation, the boundary conditions are imposed weakly using penalty terms in the SAT method [6,8,32,33]. 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 [24,32,33,20,2,26,27] 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 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, efficiency, solvability as well as the particular organisation of the technique applied to IBVP will be explored. 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 [24,32,33,20,2,26,27], it leads to optimal fully discrete energy estimates. As was mentioned above, the methodology is completely general and suitable for all semi-discrete stable problems, not only the ones discretized by the SBP-SAT technique in space.

This initial paper is organized as follows. In Section 2 we deal with the scalar initial value problem. Optimal energy estimates are derived, the solvability question is discussed and numerical experiments are performed. Section 3 deals with the application of the technique to a representative scalar initial boundary value problems. 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 SBP-SAT formulation for time discretisation of initial value problems.

2.1 The continuous energy estimate

Consider the simplest possible first order initial value problem

$$u_t = \lambda u, \tag{1}$$

with initial condition $u(0) = f$ and $0 \leq t \leq T$. Let us consider the complex constant λ to represent a stable spatial discretisation of an IBVP. The stability implies that λ has a negative semi-definite real part. For hyperbolic problems, λ is proportional to the inverse of the space step, and for parabolic problems, the space step squared.

The energy method (multiplying with the complex conjugated solution and integrating over the domain) applied to (1) yields

$$|u(T)|^2 - 2Re(\lambda)||u||^2 = |f|^2, \tag{2}$$

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 $Re(\lambda) < 0$, also the norm of the solution is bounded in terms of the initial data.

2.2 The discrete energy estimate

The SBP-SAT approximation of (1) reads

$$P^{-1}Q\vec{U} = \lambda\vec{U} + P^{-1}(\sigma(U_0 - f))\vec{e_0}. \tag{3}$$

The vector $\vec{U}$ 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

$$P = P^T > 0, \quad Q + Q^T = E_N - E_0, \tag{4}$$

where $E_0 = diag(1, 0, \ldots, 0), E_N = diag(0, \ldots, 0, 1)$. The difference operators can be based on block norms P [31] for full accuracy but sometimes diagonal versions with lower accuracy must be used [25]. The extra (penalty) term on the right-hand-side of (3) enforces the initial condition weakly (it forces the discrete solution U_0 towards f) 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 σ 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 made to preserve the SBP properties of the difference operator which is necessary for the stability proof. For further details of this technique in space, see the references on the SBP-SAT work in the Introduction.*

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

$$|\vec{U}_N|^2 - 2Re(\lambda)||\vec{U}||_P^2 = (1 + 2\sigma)|\vec{U}_0|^2 - \sigma(\bar{U}_0 f + U_0 \bar{f}), \tag{5}$$

In (5), 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 (5) and making the choice $\sigma = -1$ we obtain

$$|\vec{U}_N|^2 - 2Re(\lambda)||\vec{U}||_P^2 = |f|^2 - |U_0 - f|^2. \tag{6}$$

By comparing the continuous estimate (2) with (6) 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).

Estimates like (6) are very hard to obtain 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, i.e. when the whole time interval is considered.

2.3 The question of solvability

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

$$P^{-1}(\tilde{Q} - \lambda I)\vec{U} = \vec{R}, \tag{7}$$

where $\tilde{Q} = Q - \sigma E_0$ and $\vec{R} = -\sigma P^{-1} f \vec{e_0}$. The matrix $\tilde{Q}$ must be non-singular with a low condition number for a well functioning procedure. We make the following assumption.

Assumption 1 *Let the SBP matrix Q be defined by (4). Then $\tilde{Q} = Q - \sigma E_0$ has eigenvalues with strictly positive real parts for $\sigma < -1/2$.*

Remark 2 *The estimate (6) and numerical tests indicate the validity Assumption 1.*

2.4 Numerical calculations for initial value problems

We compare the performance of the SBP-SAT technique described above with a selection of widely acknowledged explicit and implicit methods of various orders. We investigate both the non-stiff and the stiff case. 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 [30]. This pragmatic definition will be sufficient for our needs in this section.

Consider the initial value problem (1). Note that the matrix in the corresponding discretized system (7) only depends on the grid size and the length of the integrated time interval. This means that the problem can be solved by using successive intermediate time steps with forward and backward substitutions. The same LU factorization can be used each time.

The number of arithmetic operations required to solve the already factorized system using an SBP operator of order $2s$ can be estimated as $(3s+1)N$, where $N+1$ is the number of grid points in time. This estimate is conservative as it assumes a maximum number of pivotations during the LU factorization. According to the estimate above, the work associated with higher order operators grow relatively slowly when compared to low order operators.

The most important factor is the total work needed for a certain error level. We simply define work to be the measured cpu time required by the specific machine used. Long integration times are used to get reliable cpu time measurements, using the Fortran 90 routine *cpu_time*. Moreover, we use SBP operators with diagonal norms (i.e. where P is diagonal) as well as full block norms (P is diagonal in the interior, but not close to the boundaries). 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 [31]. We denote the corresponding SBP-SAT methods by SBP($2s$,s) and SBP($2s$,$2s-1$) respectively.

We have compared with the following time integration methods:

- The second order implicit backward differentiation formula, denoted BDF2.
- The classical explicit fourth order Runge-Kutta method, denoted RK4.
- A fourth order explicit singly diagonally implicit Runge-Kutta method, denoted ESDIRK4 [19].
- An eighth order embedded explicit Runge-Kutta method, denoted DOPRI8 [28].

We use constant step sizes, for all methods, also for the embedded Runge-Kutta schemes ESDIRK4 and DOPRI8.

2.4.1 Numerical calculations for non-stiff problems

As the non-stiff test problem we consider

$$\begin{aligned} u'(t) &= -u(t) + cos(t) - sin(t), \quad 0 < t < 10^4 \\ u(0) &= 1. \end{aligned} \tag{8}$$

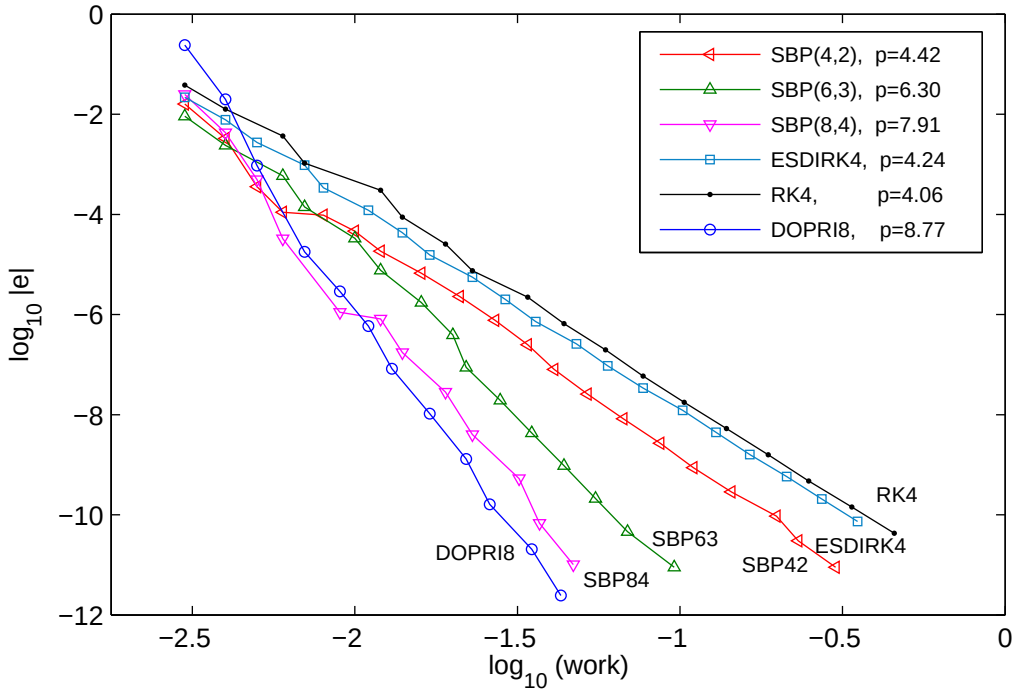

Fig. 1. Global error using diagonal norms at $t = 10^4$ versus work (measured in seconds of cpu time usage) for the non-stiff problem

The exact solution to this problem is a $u(t) = sin(t)$. Figures 1 and 2 show the global error (i.e. the difference between the numerical and the exact solution) at time $T = 10^4$ as a function of work, for diagonal norms and block norms respectively. The figures are in log-log scale and also show approximate convergence rates at the lower error levels. Figures 3 and 4 shows similarly the accumulated error up until $t = 10^4$ in the discrete L_2 norm.

2.4.2 Numerical calculations for stiff problems

For the stiff case we use again a sine function, but this time with a rapidly decaying exponential term added.

$$\begin{aligned} u'(t) &= -100u(t) + 100sin(t) + cos(t), \quad 0 < t < 10^4 \\ u(0) &= 1. \end{aligned} \tag{9}$$

The exact solution to this problem is $u(t) = e^{-100t} + sin(t)$. Figure 5 shows that the explicit schemes perform poorly on this problem, indeed confirming that this is a stiff case.

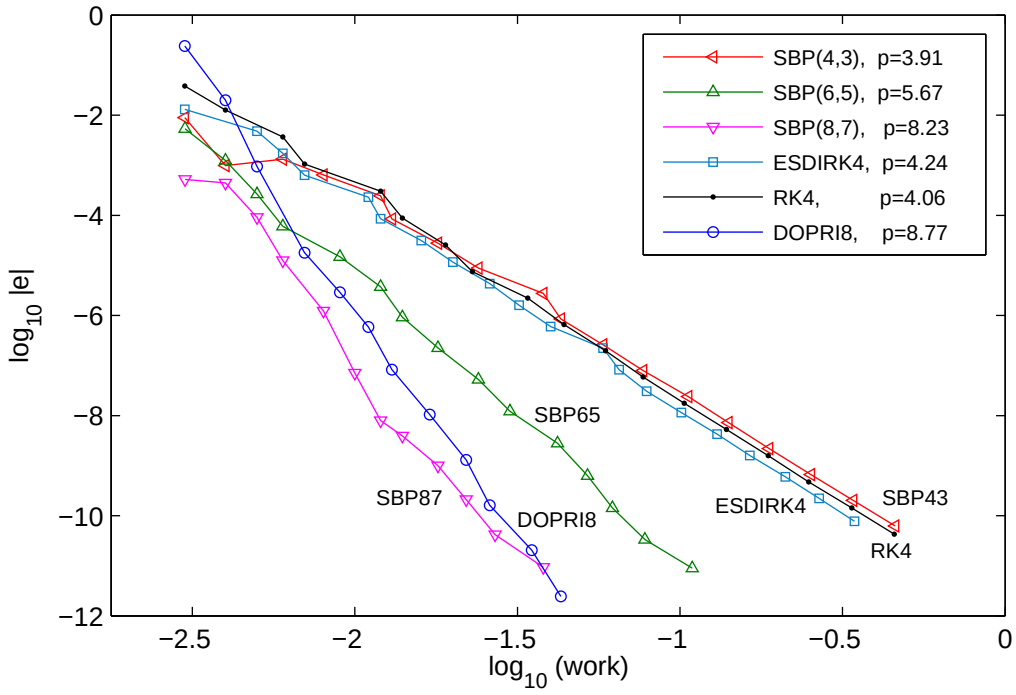

Fig. 2. Global error using block norms at $t = 10^4$ versus work (measured in seconds of cpu time usage) for the non-stiff problem

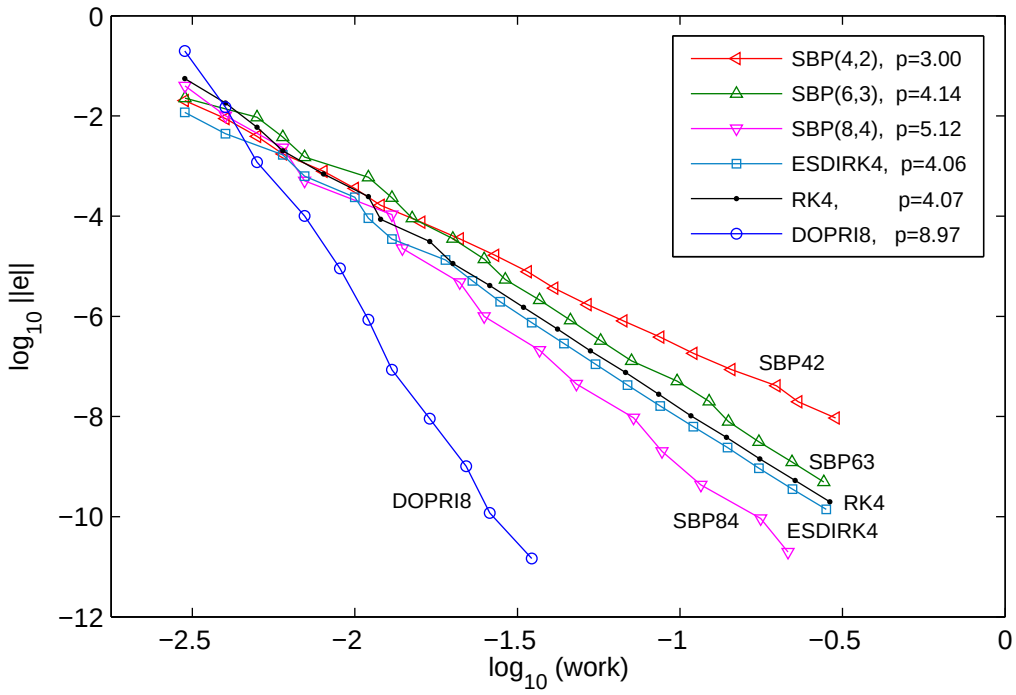

Fig. 3. L_2 error using diagonal norms versus work (measured in seconds of cpu time usage) for the non-stiff problem

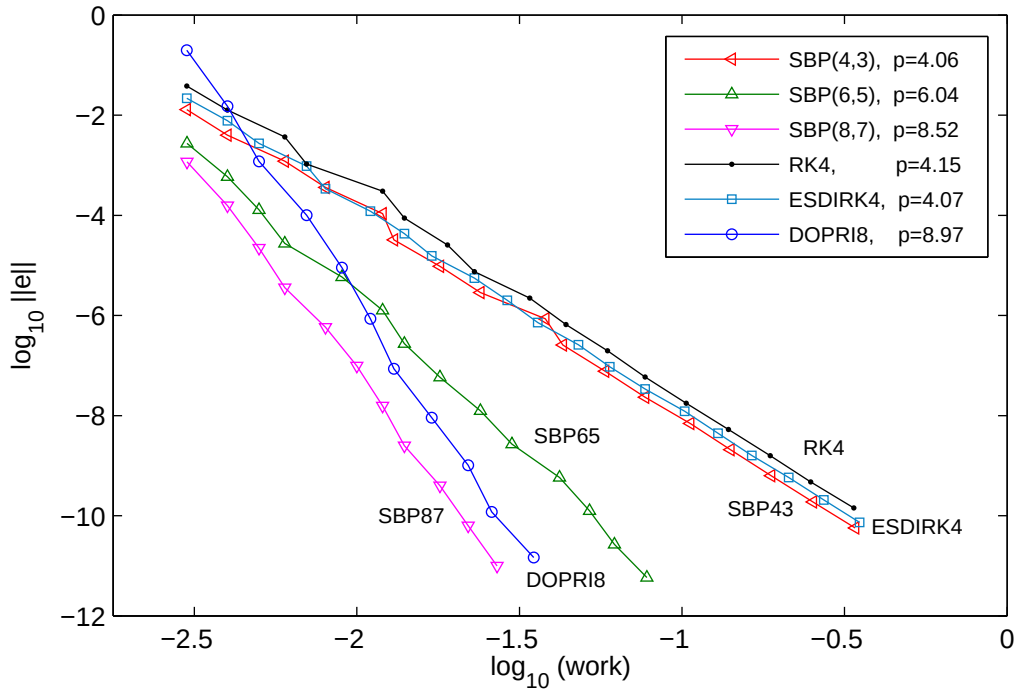

Fig. 4. L_2 error using block norms versus work (measured in seconds of cpu time usage) for the non-stiff problem

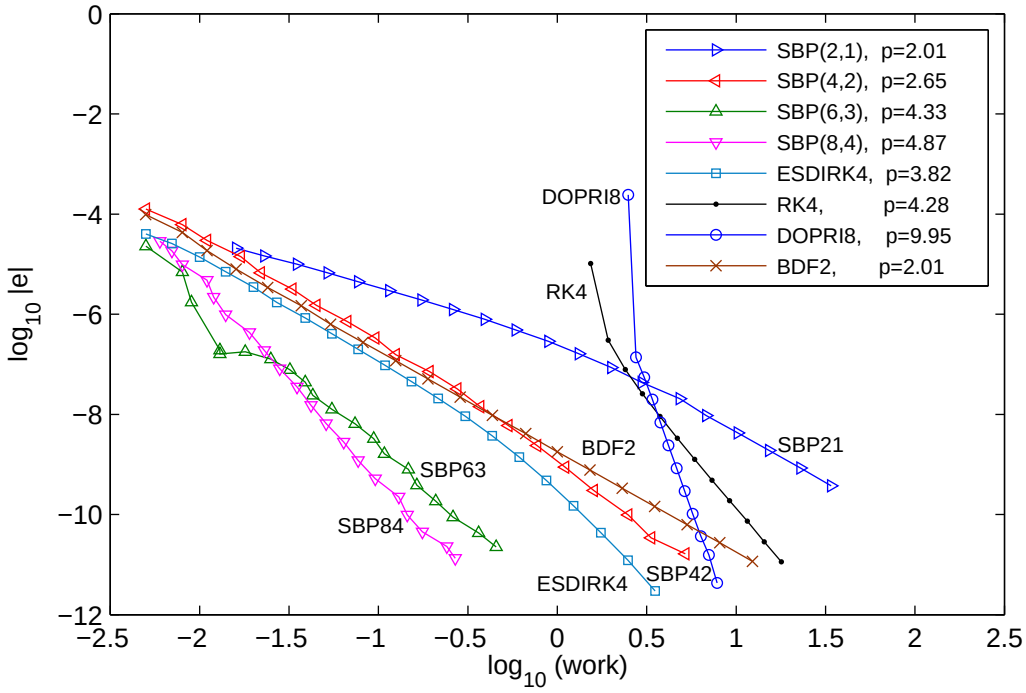

Fig. 5. Global error using diagonal norms at $t = 10^4$ versus work (measured in seconds of cpu time usage) for the stiff problem

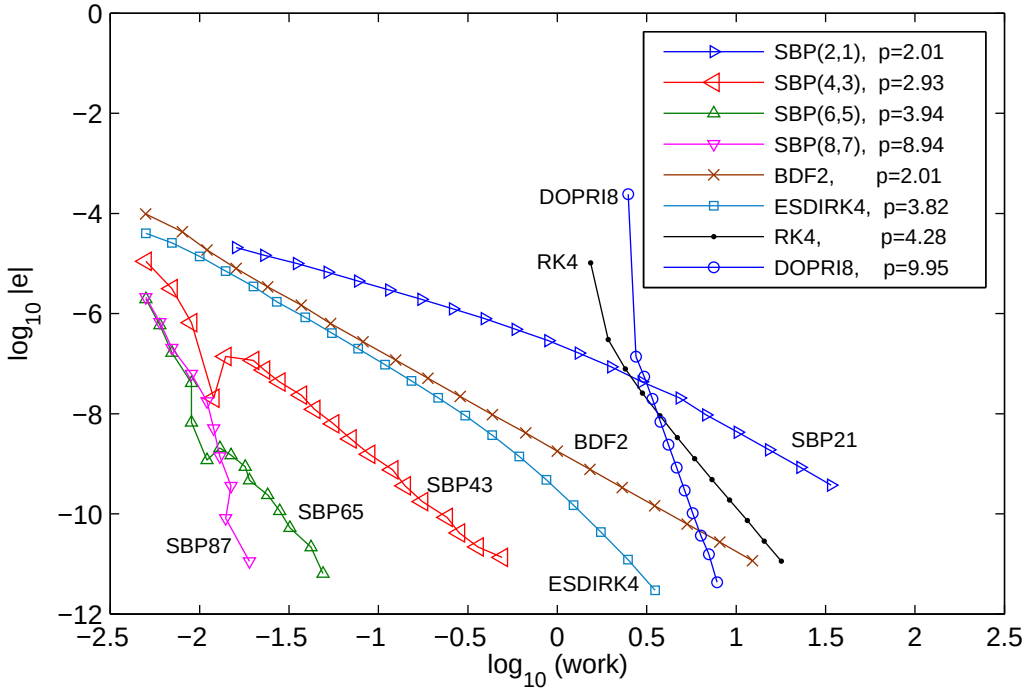

Fig. 6. Global error using block norms at $t = 10^4$ versus work (measured in seconds of cpu time usage) for the stiff problem

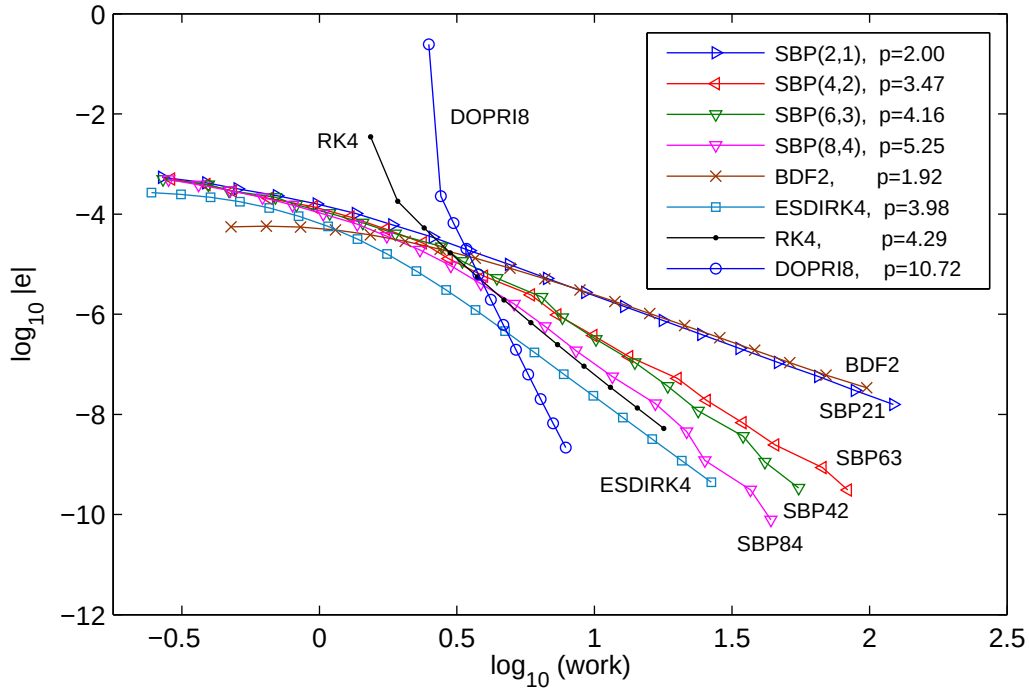

Fig. 7. L_2 error using diagonal norms versus work (measured in seconds of cpu time usage) for the stiff problem

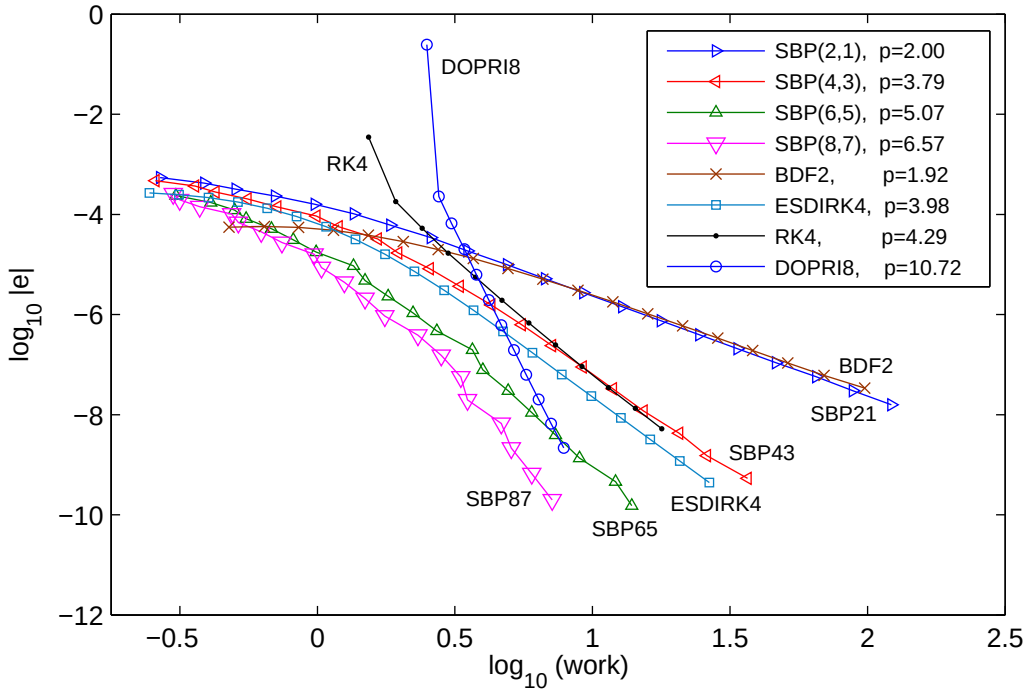

Fig. 8. L_2 error using diagonal norms versus work (measured in seconds of cpu time usage) for the stiff problem

2.4.3 Preliminary conclusions based on the numerical calculations

A bit surprisingly, the numerical calculations indicate that convergence rates for the global error are the same independent of whether diagonal or block norms where used in the non-stiff case. For the L_2 error, the higher convergence rate was only achieved by using block norms.

The stiff case is more difficult to analyze. It seems that the fourth and sixth order schemes with diagonal norms exhibit lower convergence rates than with block norms even for the global error, at least for the magnitudes of error studied here.

The combined result of the numerical calculations verify that the SBP-SAT technique applied to constant coefficient problems is highly competitive when compared with a particular selection of previously acknowledged methods. This conclusion seem to be especially true for medium to high order methods and for stiff problems.

3 The scalar initial boundary value problem

The time integration technique discussed above in (1),(3) can be extended to energy stable semi-discrete approximations of initial boundary value problems.

3.1 Preliminaries

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

$$\begin{aligned} U_t + \mathcal{P}^{-1}\mathcal{R}\mathcal{U} &= \mathcal{P}^{-1}G \\ U(0) &= F, \end{aligned} \tag{10}$$

where $\mathcal{P}^{-1}\mathcal{R}$ is an approximation of the spatial part of the PDE, $\mathcal{P}$ is the norm or mass matrix, $\mathcal{R}$ 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. $\mathcal{P}$ is symmetric and positiv definite.

Definition 1 *The approximation (10) is energy stable if*

$$\mathcal{R} + \mathcal{R}^T \geq 0 \tag{11}$$

We can now prove the following Proposition.

Proposition 1 *The system (10) under condition (11) has a bounded energy.*

Proof: The energy method (multiply from the left with $U^T\mathcal{P}$) applied to the semi-discrete approximation (10) with $G = 0$ leads by the use of (11) to the estimate

$$U^T\mathcal{P}U \leq F^T\mathcal{P}F. \tag{12}$$

□

Remark 3 *Note that many numerical methods methods using weak boundary conditions such as the SBP-SAT technique for finite differences, the finite/spectral element method, the discontinuous Galerkin method etc. have the general form (10) and satisfy (11).*

3.2 The continuous energy estimate

As example an initial boundary value problem we consider the one-dimensional advection-diffusion equation

$$\begin{aligned} u_t + au_x - \epsilon u_{xx} &= 0, & 0 \le x \le 1, \quad t \ge 0 \\ au - \epsilon u_x &= g_0(t), & x = 0, \quad t \ge 0 \\ \epsilon u_x &= g_1(t), & x = 1, \quad t \ge 0 \\ u &= f(x), & 0 \le x \le 1, \quad t = 0, \end{aligned} \tag{13}$$

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 (30) with u and integrating over the spatial domain we obtain,

$$||u||_t^2 + 2\epsilon||u_x||^2 = a^{-1}\left[(au - \epsilon u_x)^2 - (\epsilon u_x)^2\right]_{x=0} - a^{-1}\left[(au - \epsilon u_x)^2 - (\epsilon u_x)^2\right]_{x=1}$$

where $||u||^2 = \int_0^1 u^2 dx$. Next we insert the boundary conditions and arrive at the continuous energy rate

$$||u||_t^2 + 2\epsilon||u_x||^2 = a^{-1}\left[g_0^2 + g_1^2 - (au - g_0)^2 - (au - g_1)^2\right]. \tag{14}$$

Finally, by time integration, we have the final result for the continuous problem

$$\begin{aligned} ||u(\cdot,T)||^2 \, + \, 2\epsilon \int_0^T ||u_x(\cdot,t)||^2 dt = ||f||^2 + a^{-1} \int_0^T \left[g_0^2 + g_1^2\right] dt \\ - \, a^{-1} \int_0^T \left[(au - g_0)^2 + (au - g_1)^2\right] dt. \end{aligned} \tag{15}$$

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.

3.3 The semi-discrete energy estimate

The semi-discrete approximation of (30) using the SBP-SAT technique in space is

$$\begin{aligned} U_t + aDU - \epsilon D^2 U &= P^{-1}(\sigma_0(L_0 U - g_0)\vec{e_0} + \sigma_N(L_N U - g_1)\vec{e_N}) \\ U(0) &= F_0, \end{aligned} \tag{16}$$

where $D = P^{-1}Q$, $L_0 U = aU_0 - \epsilon(DU)_0$, $L_N U = \epsilon(DU)_N$ and $\vec{e_N} = (0, 0, ..., 0, 1)^T$. 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 (16) implements the boundary conditions weakly using the SAT technique.

By multiplying (16) 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

$$||U||_t^2 + 2\epsilon||DU||^2 = a^{-1}\left[g_0^2 + g_1^2 - (aU_0 - g_0)^2 - (aU_N - g_1)^2\right], \tag{17}$$

where $||U||_P^2 = U^T P U$ and $||DU||_P^2 = (DU)^T P (DU)$. Note that the semi-discrete energy rate (17) is almost identical to the corresponding continuous (14) one.

Finally, by time integration, we have the final result for the semi-discrete problem

$$\begin{aligned} ||U(\cdot,T)||^2 \, + \, 2\epsilon \int_0^T ||DU(\cdot,t)||^2 dt = ||F_0||^2 + a^{-1}\int_0^T \left[g_0^2 + g_1^2\right] dt \\ - \, a^{-1}\int_0^T \left[(aU_0 - g_0)^2 + (aU_N - g_1)^2\right] dt. \end{aligned} \tag{18}$$

The similarity of the semi-discrete estimate (18) with the continuous (15) one is striking.

For future use, we will rearrange the formulation (16) with $\sigma_0 = \sigma_N = -1$ to

$$\begin{aligned} U_t + P^{-1}R_x U &= P^{-1}G \\ U(0) &= F_0, \end{aligned} \tag{19}$$

where

$$R_x = a(Q + E_0) - \epsilon(Q + E_0 - E_1)D, \quad G = (g_0, 0, ..., 0, g_1)^T. \tag{20}$$

We have showed in (17) that (let $g_0 = g_1 = 0$)

$$R_x + R_x^T = a(E_0 + E_N) + 2\epsilon D^T P D. \tag{21}$$

Note that (21) mean that the symmetric part of R_x is positive semi-definit, i.e.

$$R_x + R_x^T \geq 0. \tag{22}$$

Clearly now the approximation (19) is energy stable according to Definition 1. It will be shown below that (21),(22) show that the eigenvalues of R_x cannot have negative real parts, i.e

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

$$A \otimes B = \begin{bmatrix} a_{1,1}B & \dots & a_{1,m}B \\ \vdots & \ddots & \vdots \\ a_{n,1}B & \dots & a_{m,n}B \end{bmatrix}. \tag{23}$$

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

$$(A \otimes B)(C \otimes D) = (AC \otimes BD) \tag{24}$$

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

$$(A \otimes B)^{-1,T} = A^{-1,T} \otimes B^{-1,T} \tag{25}$$

if the usual matrix inverse is defined.

The fully discrete version of (30) is obtained by discretising (19) in time using the SBP-SAT technique. The use of the Kronecker product rules (23-25) and (20) yield

$$(P_t^{-1}Q_t \otimes I_x)U + (I_t \otimes P_x^{-1}R_x)U = (I_t \otimes P_x^{-1})G + \sigma_t(P_t^{-1}E_0 \otimes I_x)(U - F), \tag{26}$$

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 σ_t which will be determined for stability. The organisation of the vectors are

$$\begin{aligned} U &= (U_0, U_1, ..., U_{M-1}, U_M)^T \quad , U_i = (U_{i0}, U_{i1}, ..., U_{iN-1}, U_{iN})^T \\ G &= (G_0, G_1, ..., G_{M-1}, G_M)^T \ , G_i = (g_0(i\Delta t), 0, ..., 0, g_1(i\Delta t))^T \\ F &= (F_0, U_1, ..., U_{M-1}, U_M)^T \quad , F_0 = (f_0, f_1, ..., f_{N-1}, f_N)^T \\ U^0 &= (U_{00}, U_{10}, ..., U_{M0})^T \qquad , G^0 = (g_0(0), g_0(\Delta t), ..., g_0(M\Delta t))^T \\ U^N &= (U_{0N}, U_{1N}, ..., U_{MN})^T \qquad , G^1 = (g_1(0), g_1(\Delta t), ..., g_1(M\Delta t))^T. \end{aligned} \tag{27}$$

By multiplying (26) with $U^T(P_t \otimes P_x)$ from the left, using the SBP properties (4), the relation (19), the Kronecker product rules (23-25), the relations (27)

and the choice $\sigma_t = -1$ we obtain

$$\begin{aligned} U_M^T P_x U_M \; &+ 2\epsilon (DU)^T (P_t \otimes P_x) DU = F_0^T P_x F_0 \\ &+ a^{-1} \left[(G^0)^T P_t G^0 + (G^1)^T P_t G^1 \right] \\ &- a^{-1} \left[(aU^0 - G^0)^T P_t (aU^0 - G^0) + (aU^N - G^1)^T P_t (aU^N - G^1) \right] \\ &- (U_0 - F_0)^T P_x (U_0 - F_0). \end{aligned} \tag{28}$$

Note the close similarity between the continuous estimate (15), the semi-discrete estimate (18) and the fully discrete one in (28). The fully discrete estimate has the additional damping term $-(U_0 - F_0)^T P_x (U_0 - F_0)$ also present in (6).

3.5 The question of solvability

By rearranging (26), we get the final equation to solve for U

$$BU = (B_t + B_x)U = \left[(P_t^{-1} \tilde{Q}_t \otimes I_x) + (I_t \otimes P_x^{-1} R_x) \right] U = H, \tag{29}$$

where $\tilde{Q}_t = Q_t - \sigma_t E_0$, R_x is defined in (20) and $H = (I_t \otimes P_x^{-1})G - \sigma_t (P_t^{-1} E_0 \otimes I_x)F$ is the data vector.

The following theorem is well-known.

Theorem 1 *Let A,B be diagonalizable. Then A and B commute if and only if they are simultaneously diagonalizable.*

Proof: See proof after theorem 1.3.12 in [14]. □

We need the following Assumption.

Assumption 2 *The eigenvectors of the matrix B_t in (29) are linearly independent. Also the matrix B_x in (29) have linearly independent eigenvectors.*

Remark 4 *We have presently no theoretical support for Assumption 2.*

We will need the following Lemma.

Lemma 1 *The matrices B_t, B_x and $B = B_t + B_x$ have the same eigenvectors.*

Proof: Let $B_t = (P_t^{-1} \tilde{Q}_t \otimes I_x) = (C_t \otimes I_x)$ and $B_x = (I_t \otimes P_x^{-1} R_x) = (I_t \otimes C_x)$. We have $B_t B_x = (C_t \otimes I_x)(I_t \otimes C_x) = (C_t \otimes C_x) = (I_t \otimes C_x)(C_t \otimes I_x) = B_x B_t$ i.e. the matrices commute and by Theorem 1 have the same eigenvectors. □

The following Lemma position the eigenvalues in the complex plane.

Lemma 2 *Let the matrix P be positive definite and the matrix A have a positive semi-definite symmetric part. Then, the matrix $P^{-1}A$ has eigenvalues with positive semi-definite real parts.*

Proof: Let λ and x be an eigenvalue and eigenvector to $P^{-1}A$, i.e. $P^{-1}Ax = \lambda x$. Elementary manipulations lead to $Re(\lambda) = x^*(A + A^T)x/(2x^*Px) \geq 0$.□

We are now ready to show

Proposition 2 *The matrix $B = \left[(P_t^{-1}\tilde{Q}_t \otimes I_x) + (I_t \otimes P_x^{-1}R_x)\right]$ in (29) have non-zero eigenvalues with positive real parts.*

Proof: Lemma 1 leads to $B = B_t + B_x = X(\Lambda_t + \Lambda_x)X^{-1}$ where X is the common eigenvector matrix. Assumption 2 together with (21),(22) and Lemma 2 show that the eigenvalues of B have positive real parts. □

The final result of the paper is summarized in the following Proposition.

Proposition 3 *The solution to (26) and (29) is unique and bounded.*

Proof: Proposition 2 and Assumption 2 leads to an invertible matrix B. □

3.6 Numerical calculations for the initial boundary value problems

We show preliminary results for the special case of periodic advection as well as the full advection-diffusion problem (30). The SBP-SAT technique is compared with the classical Runge-Kutta method. We stress that, in the end, the competitiveness will depend strongly on the exact technique used to solve the large linear equation system arising from the SBP-SAT approach. In this paper we exclude the full analysis of these problems and focus on convergence rates and on accuracy as a function of the spatial and temporal resolution.

3.6.1 Numerical calculations for the advection problem

First we consider the following periodic advection problem

$$\begin{aligned} u_t + u_x &= 0, && 0 \leq x \leq 1, && t \geq 0 \\ u(0,t) &= u(1,t), && && t \geq 0 \\ u &= f(x), && 0 \leq x \leq 1, && t = 0, \end{aligned} \tag{30}$$

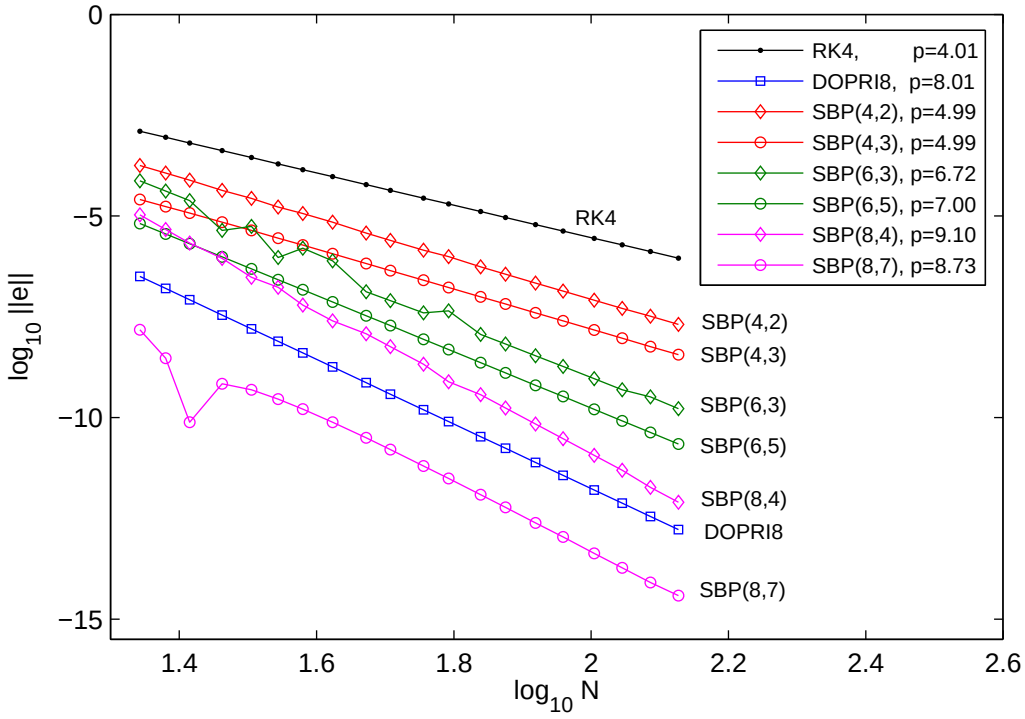

Fig. 9. L_2 error at t=1 for the periodic advection problem

where $f(x) = sin(2\pi x)$. In the discretisation we use a fully periodic operator in space, i.e. the SBP property changes to $Q_x + Q_x^T = 0$, and exclude boundary data. The system to solve is (29) with $H = (P_t^{-1} E_0 \otimes I_x)F$, and $\epsilon = 0$.

In the calculations we use the same order of accuracy in the spatial operator as the interior order of accuracy in the temporal operator. Moreover we use operators of the same size in both space and time, $N = N_x = N_t$. Figure 9 illustrate the convergence in terms of L_2 error at $t = 1$ as a function of N. The result are what could be expected, expect maybe that the SBP($2s$,s) schemes perform exceptionally well and converge at a higher than expected rate.

Even though the accuracy is high at times of order one, the errors might grow as times passes. Figures 10 and 11 shows the evolution of the L_2 error for long times. Here we can see that all the non SBP schemes ($RK4$, $DOPRI8$ and $ESDIRK4$) suffer from error growth in time while the SBP schemes does not. To illustrate the importance of that fact further, Figure 12 shows the numerical solution of $RK4$, $SBP(4,2)$ and $SBP(4,3)$ after a very long time ($t = 10000$) integration. The error growth in the $RK4$ method cause a phase shift and dispersion error, while no such problems exist for $SBP(4,2)$ and $SBP(4,3)$.

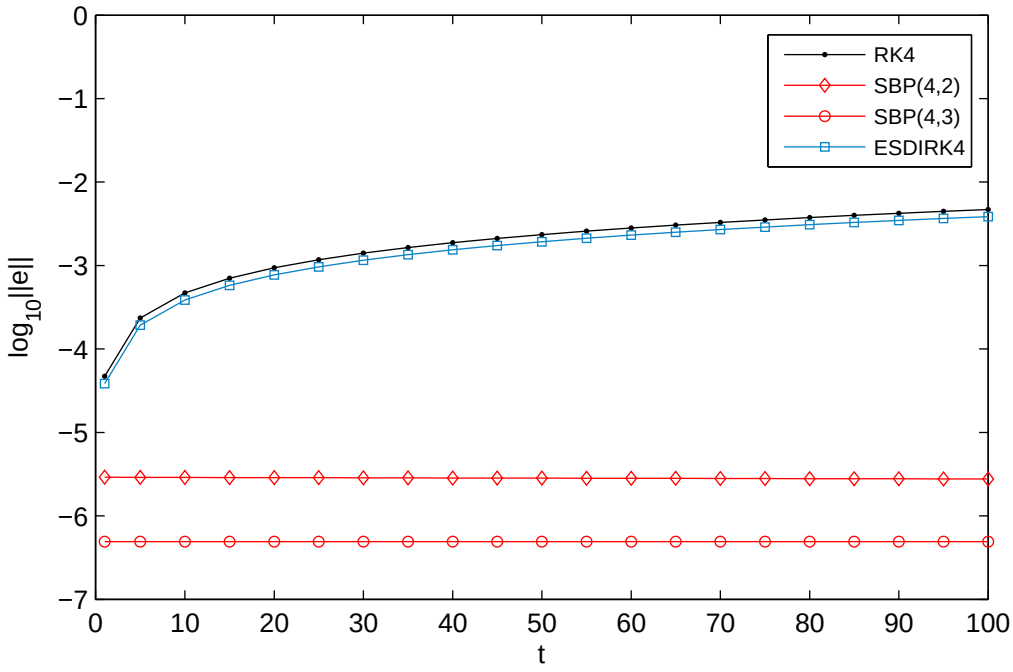

Fig. 10. L_2 error for long times of the periodic advection problem, fourth order methods, grid size N=100

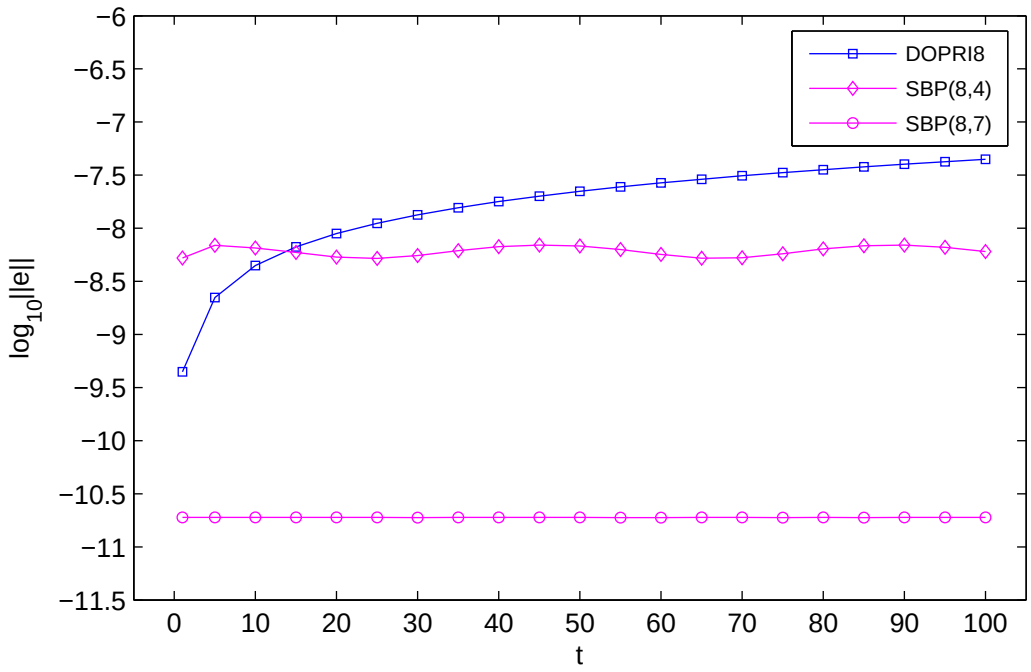

Fig. 11. L_2 error for long times of the periodic advection problem, eighth order methods, grid size N=100

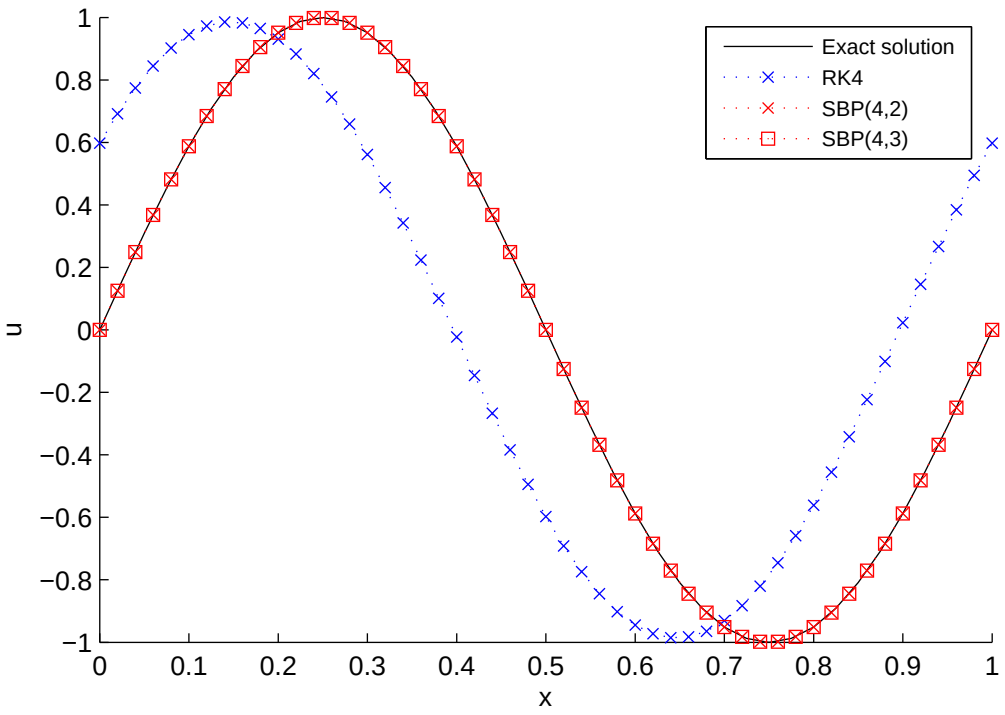

Fig. 12. Phase shift for three 4th order schemes with grid size N=50, at t=10000

3.6.2 Numerical calculations for the advection-diffusion problem

Next we consider the advection-diffusion equation (30) with the parameter choice $a = 1$ and $\epsilon = 1$. To check the accuracy we use the method of manufactured solutions, see [29],[22], and employ the following exact solution

$$u = sin(\frac{\sqrt{3}}{2}(x - 2t))e^{-\frac{1}{2}x}, \tag{31}$$

that was used in [12]. For this non-periodic spatial problem we must impose boundary conditions, which is done with the standard SBP-SAT technique. We use diagonal norms for the spatial SBP operators and choose spatial operators such that we match the accuracy in time of the time operators. As before, we let $N_t = N_x = N$.

Figure 13 shows the convergence at $t = 1$, while figures 14 and 15 show errors for a long time integration. The error growth that could be seen for the hyperbolic advection problem above cannot be seen in this case and the methods seem rather comparable.

Remark 5 *The error growth for the hyperbolic advection problem in Section 3.6.1 cannot be seen for the parabolic advection diffusion equation. It it well known that the error grows linearly in time for hyperbolic problems, unless*

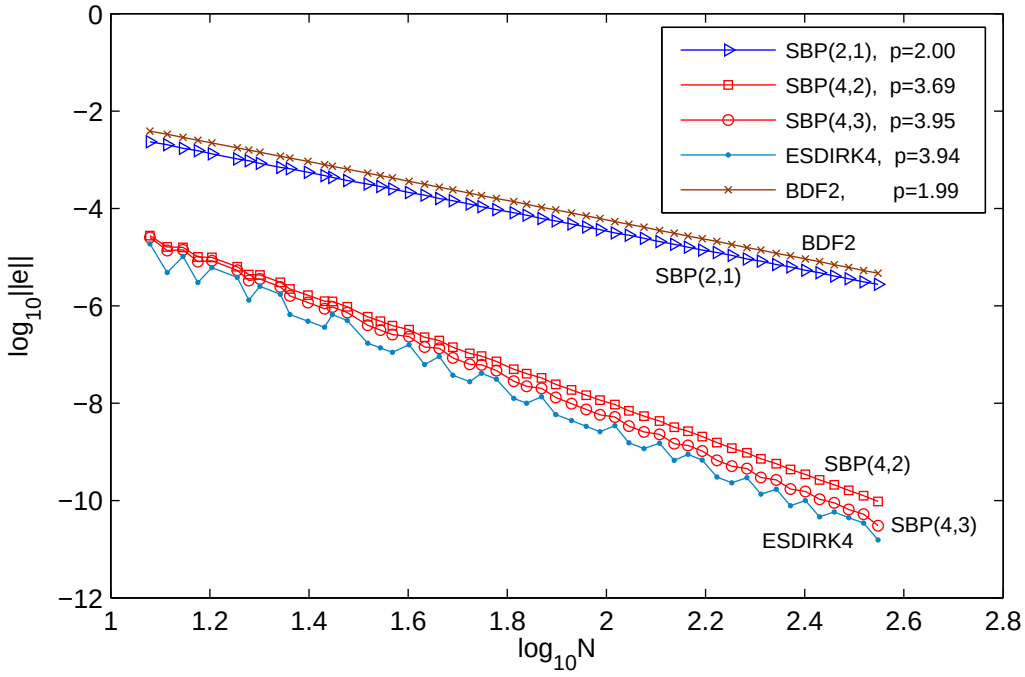

Fig. 13. L_2 error at t=1 for the advection-diffusion problem

specific boundary procedures are used, see [25], while parabolic problems have a natural error bound due to the second derivative.

3.6.3 Preliminary conclusions based on the numerical calculations

The most interesting results were obtained for hyperbolic periodic advection problem. The results indicate that the SBP-SAT technique gives as good accuracy as the established comparable methods for a given spatial and temporal resolution. An advantage with the SBP schemes are that they do not produce error growth in time. This advantage probably stems directly from the fact that there is a clear energy bound.

The results for the parabolic advection-diffusion equation were less interesting. Most methods of the same order or accuracy seem to perform equally well. As was mentioned above, error growth in time is normally not a big problem for parabolic equations. It remains to be seen how efficiently these methods can be implemented. This will be an obvious topic for future work.

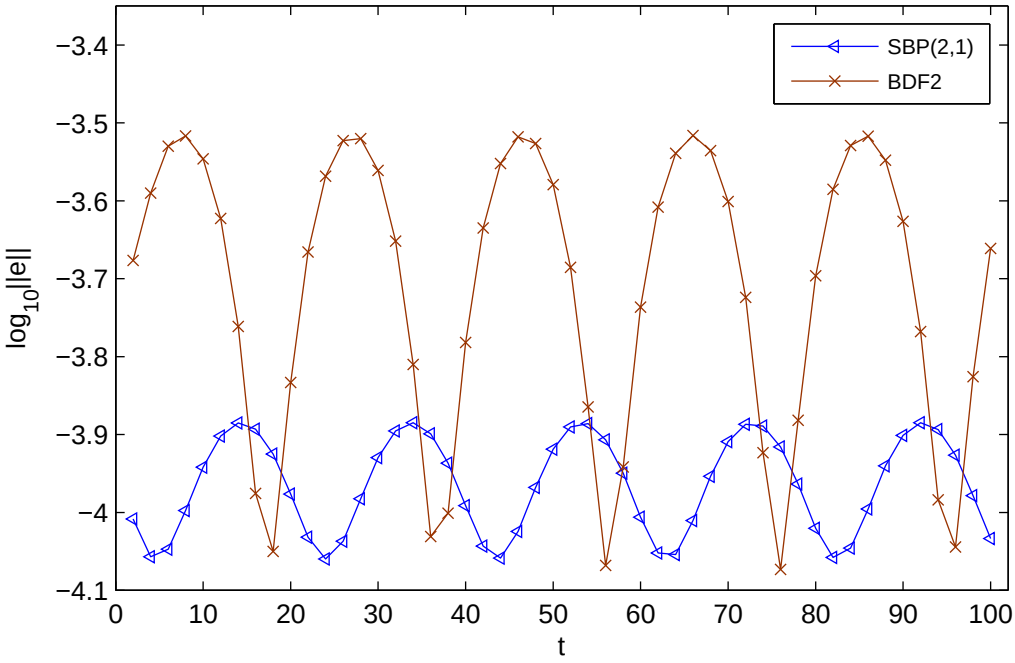

Fig. 14. L_2 error for long times of the advection-diffusion problem, 2nd order methods, grid size N=50

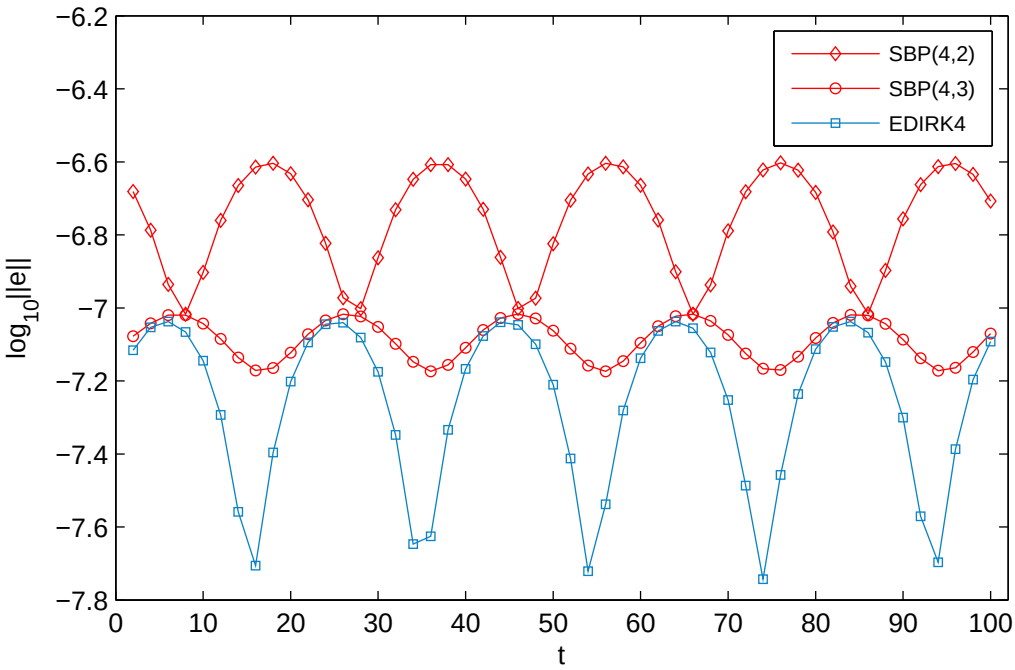

Fig. 15. L_2 error for long times of the advection-diffusion problem, 4th order methods, grid size N=50

4 Extension of the stability theory of the general case

The stability theory in section 3.4 and 3.5 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 [8,32,33,20,2,26] 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 (19) under condition (11) such that we have an energy stable approximation, see Proposition 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 (26) becomes

$$(P_t^{-1}Q_t \otimes I_s)U + (I_t \otimes \mathcal{P}^{-1}\mathcal{R})U = (I_t \otimes \mathcal{P}^{-1})G + \sigma_t(P_t^{-1}E_0 \otimes I_s)(U - F). \quad (32)$$

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 initialdata F_0). The matrix I_s is the identity matrix for the multi-index.

The energy method applied to (32) and the choice $\sigma_t = -1$ yield

$$U_M^T P_s U_M \leq F_0^T P_s F_0 - (U_0 - F_0)^T P_s (U_0 - F_0), \quad (33)$$

which correspond to (28) in the fully discrete one-dimensional case, or (12) in the general semi-discrete case. We have of course no details about the spatial part of the estimate.

Finally we generalize Proposition 3 for the one-dimensional scalar case to the multiple dimensional system case.

Proposition 4 *The solution to (32) is unique and bounded.*

Proof: The proof is analogous to the one-dimensional case in Section 3.5. □

Remark 6 *Proposition 4 above means that systems like the Maxwells' equations, the elastic wave equations and the linearised Euler and Navier-Stokes equations can be shown to be stable for fully discrete high order approximations. Stability can be 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-discretisation technique for initial value problems by extending the well known SBP-SAT technique for space discretisation in the time domain. We use summation-by-parts operators in time and a weak initial condition.

The new time-discretisation method is global and together with energy stable semi-discrete approximations, it generates optimal fully discrete energy estimates, and very efficient methods for both stiff and non-stiff problems. 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 and even superior in many cases. In particular, the SBP-SAT schemes har no error growth for long time integration of the hyperbolic advection problem.

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 linearised 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.

The investigations in this paper are of initial character, have shown great promise, but much research remains. Future work will include work on the theoretical foundation and on how to arrange and structure an efficient solution procedure. Also, we have focused on constant coefficient problems, time-dependent coefficients and nonlinear problems will be a future topic as well.

References

[1] O. Axelsson. Global integration of differential equations through lobatto quadrature. *BIT*, 4(2):69–86, 1964.

[2] J. Berg and J. Nordström. Stable Robin solid wall boundary conditions for the Navier-Stokes equations. *Journal of Computational Physics*, 230(19):7519–7532, 2011.

[3] J. C. Butcher. Initial value problems: numerical methods and mathematics. *Computers and Mathematics with Applications*, 28(10-12):1–16, 1994.

[4] J. C. Butcher. General linear methods for stiff differential equations. *BIT Numerical Mathematics*, 41(2):240–264, 2001.

[5] J.C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, Ltd, 2005.

[6] M. H. Carpenter, D. Gottlieb, and S. Abarbanel. Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: Methodology and application to high-order compact schemes. *Journal of Computational Physics*, 111(2):220–236, 1994.

[7] M. H. Carpenter, C. A. Kennedy, H. Bijl, S. A. Viken, and V. N. Vatsa. Fourth-order runge-kutta schemes for fluid mechanics applications. *Journal of Scientific Computing*, 25(1):157–194, 2005.

[8] M.H. Carpenter, J. Nordström, and D. Gottlieb. A stable and conservative interface treatment of arbitrary spatial accuracy. *Journal of Computational Physics*, 148:341–365, 1999.

[9] J. R. Cash. The integration of stiff initial value problems in odes using modified extended backward differentiation formulae. *Computers and Mathematics with Applications*, 9(5):645–657, 1983.

[10] J. R. Cash. Modified extended backward differentiation formulae for the numerical solution of stiff initial value problems in odes and daes. *Journal of Computational and Applied Mathematics*, 125(1-2):117–130, 2000.

[11] F. Costabile and A. Napoli. A method for global approximation of the initial value problem. *Numerical Algorithms*, 27(2):119–130, 2001.

[12] J. Gong and J. Nordström. Interface procedures for finite difference approximations of the advection-diffusion equation. *Journal of Computational and Applied mathematics*, 236:602–620, 2011.

[13] B. Guo and Z. Wang. Legendre-Gauss collocation methods for ordinary differential equations. *Advances in Computational Mathematics*, 30(3):249–280, 2009.

[14] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[15] W. Hundsdorfer, A. Mozartova, and M. N. Spijker. Stepsize restrictions for boundedness and monotonicity of multistep methods. *Journal of Scientific Computing*, 50(2):265–286, 2012.

[16] W. Hundsdorfer and S. J. Ruuth. On monotonicity and boundedness properties of linear multistep methods. *Mathematics of Computation*, 75(254):655–672, 2006.

[17] W. Hundsdorfer and S. J. Ruuth. Imex extensions of linear multistep methods with general monotonicity and boundedness properties. *Journal of Computational Physics*, 225(2):2016–2042, 2007.

[18] A. Kanevsky, M. H. Carpenter, D. Gottlieb, and J. S. Hesthaven. Application of implicit-explicit high order runge-kutta methods to discontinuous-galerkin schemes. *Journal of Computational Physics*, 225(2):1753–1781, 2007.

[19] C. A. Kennedy and M. H. Carpenter. Additive runge-kutta schemes for convection-diffusion-reaction equations. *Applied Numerical Mathematics*, 44(1-2):139–181, 2003.

[20] J. E. Kozdon, E. M. Dunham, and J. Nordström. Interaction of waves with frictional interfaces using summation-by-parts difference operators: Weak enforcement of nonlinear boundary conditions. *Journal of Scientific Computing*, 50(2):341–367, 2012.

[21] H.-O. Kreiss and 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 Equation.* Academic Press, New York, 1974.

[22] J. Lindström and J. Nordström. A stable and high-order accurate conjugate heat transfer problem. *Journal of Computational Physics*, 229(14):5440–5456, 2010.

[23] K. Mattsson and J. Nordström. Summation by parts operators for finite difference approximations of second derivatives. *Journal of Computational Physics*, 199(2):503–540, 2004.

[24] K. Mattsson, M. Svärd, M. Carpenter, and J. Nordström. High-order accurate computations for unsteady aerodynamics. *Computers and Fluids*, 36(3):636–649, 2007.

[25] J. Nordström. Error bounded schemes for time-dependent hyperbolic problems. *SIAM Journal on Scientific Computing*, 30:46–59, 2007.

[26] J. Nordström, J. Gong, E. van der Weide, and M. Svärd. A stable and conservative high order multi-block method for the compressible navier-stokes equations. *Journal of Computational Physics*, 228:9020–9035, 2009.

[27] J. Nordström and R. Gustafsson. High order finite difference approximations of electromagnetic wave propagation close to material discontinuities. *Journal of Scientific Computing*, 18(2):215–234, 2003.

[28] P.J. Prince and J.R. Dormand. High order runge-kutta formulae. *Journal of Computational and Applied Mathematics*, 7:67–75, 1981.

[29] L. Shunn, F. Ham, and P. Moin. Verification of variable-density flow solvers using manufactured solutions. *Journal of Computational Physics*, 231(9):3801 – 3827, 2012.

[30] M.N. Spijker. Stiffness in numerical initial-value problems. *Journal of Computational and Applied Mathematics*, 72(2):393–406, 1996.

[31] B. Strand. Summation by parts for finite difference approximation for d/dx. *Journal of Computational Physics*, 110(1):47–67, 1994.

[32] M. Svärd, M.H. Carpenter, and J. Nordström. A stable high-order finite difference scheme for the compressible Navier-Stokes equations: far-field boundary conditions. *Journal of Computational Physics*, 225(1):1020–1038, 2007.

[33] M. Svärd and J. Nordström. A stable high-order finite difference scheme for the compressible Navier-Stokes equations: No-slip wall boundary conditions. *Journal of Computational Physics*, 227(10):4805–4824, 2008.

[34] Z. Wang and B. Guo. Legendre-gauss-radau collocation method for solving initial value problems of first order ordinary differential equations. *Journal of Scientific Computing*, pages 1–30, 2011. Article in Press.