The SBP-SAT Technique for Time-Discretization

Tomas Lundquist and Jan Nordström

Linköping University Post Print

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

Original Publication:

Tomas Lundquist and Jan Nordström, The SBP-SAT Technique for Time-Discretization, 2013, AIAA Aerospace Sciences – Fluid Sciences Event.

<http://dx.doi.org/10.2514/6.2013-2834>

From the 21st AIAA Computational Fluid Dynamics Conference 24 - 27 June 2013 San Diego, California

Postprint available at: Linköping University Electronic Press

<http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-96448>

The SBP-SAT Technique for Time-Discretization

Tomas Lundquist*

Linköping University, Linköping, Sweden

Jan Nordström†

Linköping University, Linköping, Sweden

We show how the *SBP–SAT* technique can be used to derive a new family of methods for the time integration of initial value problems producing optimally sharp energy estimates. Some stability properties relevant for these methods are studied in detail, and accuracy results for both non-stiff and stiff problems are presented. We show that the technique is particularly suitable for the time integration of energy stable semi-discrete problems.

I. Introduction

A. Background

In this paper we study the properties of a new family of numerical methods, previously described in,¹² for initial value problems of the following form:

$$\begin{aligned} u_t + F(t, u) &= 0, \quad 0 < t \leq T \\ u(0) &= f. \end{aligned} \tag{1}$$

The goal of a numerical method is to find an approximate solution $\vec{U}$ to such a problem at discrete points in time between 0 and T :

$$\vec{U} = (U_0 \ U_1 \ \dots \ U_N)^T \approx (u(0) \ u(t_1) \ \dots \ u(T))^T. \tag{2}$$

If the initial value problem (1) is non-stiff, then the time steps can be selected with accuracy as the only consideration. In this case explicit time-stepping methods are most often sufficient.⁴

If the problem is stiff on the other hand, an implicit method is generally required to ensure stability of the solution at the desired level of accuracy. Time-stepping methods such as implicit one-step Runge-Kutta or linear multi-step methods are often used in this case. A serious drawback of the latter group is the lack of A –stability for methods with order higher than two, while methods in the former group are generally more demanding with respect to cost of implementation.³

To complicate things further, less computationally expensive implicit Runge-Kutta methods can suffer from serious reduction of order in the presence of large stiffness. For instance, the fourth order explicit, singly diagonally implicit Runge-Kutta method (ESDIRK4)^{2,8} has gained popularity in recent years for combining L –stability (and thus A –stability) and relatively low cost of implementation. However, the convergence order for very stiff problems can drop to two.

Even though stiffness is non-trivial to define for a general non-linear problem (1),¹⁴ it can in many cases be characterized through the following scalar constant coefficient test equation:¹³

$$\begin{aligned} u_t + \lambda u &= \psi_t + \lambda \psi, \quad 0 < t \leq 1 \\ u(0) &= \psi(0), \end{aligned} \tag{3}$$

where ψ is the (manufactured) exact solution. The constant λ can be used as a stiffness parameter.

The methods studied in this paper were first described in,¹² where optimal energy estimates were derived both for scalar initial value problems and for problems resulting from energy stable semi-discrete systems. High orders of stiff and non-stiff convergence were confirmed in numerical calculations.

*Ph.D student, Department of Mathematics, Linköping University, SE-581 83, Linköping, Sweden

†Professor, Department of Mathematics, Linköping University, SE-581 83, Linköping, Sweden

B. A-stability

We first consider the situation where the linear system (3) sufficiently characterizes the stiffness in the general set of equations (1). For the purpose of well-posedness, it is sufficient to consider the homogeneous form of (3), obtained by choosing the exact solution $\psi = fe^{-\lambda t}$:

$$\begin{aligned} u_t + \lambda u &= 0, \quad 0 < t \leq 1 \\ u(0) &= f, \end{aligned} \quad (4)$$

The energy method (multiplying with the complex conjugated solution and integrating over the domain) applied to (4) yields

$$|u(T)|^2 + 2\operatorname{Re}(\lambda)\|u\|^2 = |f|^2, \quad (5)$$

where $\|u\|^2 = \int_0^T |u|^2 dt$. Thus the solution at the final time is bounded in terms of the initial data if $\operatorname{Re}(\lambda) \geq 0$. It is desirable that this property holds also for the discrete solution. We therefore introduce the following definition.

Definition 1 *A numerical method that produces the approximate solution (2) when applied to (4) is said to be A-stable if $\operatorname{Re}(\lambda) \geq 0$ implies that*

$$|U_N|^2 \leq |f|^2 \quad (6)$$

If we replace u_N by u_1 in (6), then Definition 1 becomes equivalent to Definition 3.3 in⁷ for A-stability of Runge-Kutta methods. In the following two sections we will proceed toward more general stability definitions that apply for nonlinear equations as well.

C. B-stability

We now consider a general nonlinear problem (1), but where the function F satisfies a one-sided Lipschitz condition:

$$(u - v, F(t, u) - F(t, v)) \geq L\|u - v\|^2. \quad (7)$$

Then, suppose that u and v are two solutions to (1) with initial data f and g respectively. This gives:

$$\begin{aligned} u_t - v_t + F(t, u) - F(t, v) &= 0, \quad 0 < t \leq T \\ (u - v)(0) &= f - g. \end{aligned}$$

The energy method (multiplying from the left with $(u - v)^*$ and adding the complex conjugate) then yields:

$$\|u - v\|_t^2 + 2\operatorname{Re}(u - v, F(t, u) - F(t, v)) = 0.$$

If moreover $L \geq 0$ then

$$\operatorname{Re}(u - v, F(t, u) - F(t, v)) \geq 0. \quad (8)$$

The difference is then bounded by

$$\|u - v\| \leq \|f - g\|.$$

This motivates the following definition:

Definition 2 *A numerical method that produces the approximate solution (2) when applied to (1) is said to be B-stable if the contractivity condition (8) implies that*

$$\|U_N - V_N\| \leq \|f - g\|, \quad (9)$$

where U_N and V_N are numerical approximations of the solution at $t = 1$, using initial values f and g respectively.

Similarly to the definition of A-stability in the previous section, if we replace $U_N - V_N$ by $U_1 - V_1$ in (9), then Definition (2) is identical to Definition 12.2 in⁷ for B-stability of Runge-Kutta methods.

D. Energy stability

The one-sided Lipschitz condition (7), although sufficient for well-posedness of any nonlinear initial value problem (1), is not always satisfied in problems that are of interest for solving numerically. An alternative is to consider problems where the time growth of the solution is limited. For instance, solutions to initial boundary value problems as well as the semi-discrete spatial approximations can often be shown to be bounded in time by the problem data.⁶

We consider especially the case where the function F in the initial value problem (1) satisfies:

$$\operatorname{Re}(u, F(t, u)) \geq 0. \quad (10)$$

in an appropriate scalar product (Note that this condition does not in general guarantee well-posedness of (1)). If (10) holds, then the solution is bounded by initial data as

$$\|u(T)\| \leq \|f\|$$

The condition (10) is a more general form of the condition given in¹² defining energy stability of semi-discrete approximations to initial boundary value problems. It is desirable that this property of boundedness of the solution is preserved by the time integration procedure as well. For this reason we introduce the concept of energy stability also for the time discretization technique.

Definition 3 *A numerical method that produces the approximate solution (2) when applied to (1) is said to be energy stable if condition (10) implies that*

$$\|U_N\| \leq \|f\|.$$

II. The SBP-SAT technique

A. SBP operators

A discrete difference operator on summation-by-parts form approximating the first derivative of $(u(0) \ u(\Delta t) \ \dots \ u(T))^T$ can be written as the product $P^{-1}Q$ between two matrices with the following properties:^{5,9,15}

$$P = P^T > 0, \quad Q + Q^T = E_N - E_0,$$

where $E_0 = \operatorname{diag}(1, 0, \dots, 0)$, $E_N = \operatorname{diag}(0, \dots, 0, 1)$. The so called norm matrix P is of the form $P = \Delta t \tilde{P}$, where $\tilde{P}$ has entries of order one.

As an example, consider the operator consisting of second order central differences in the interior of the computational domain and first order one-sided differences at the boundaries. Then the matrices P and Q can be chosen as follows

$$P = \Delta t \begin{bmatrix} \frac{1}{2} & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & \frac{1}{2} \end{bmatrix}, Q = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & & & \\ -\frac{1}{2} & 0 & \frac{1}{2} & & \\ & \ddots & \ddots & \ddots & \\ & & -\frac{1}{2} & 0 & \frac{1}{2} \\ & & & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Difference operators on *SBP* form have been constructed with order $2s$ for interior stencil points, $s = 1, 2, 3, 4$, but the order is generally lower in a limited number of points close to the boundaries of the time domain. If P is restricted to be diagonal, the operators have order s at the boundaries^{10,11} whereas if blocks of limited size at the upper left and lower right corners of P are allowed, the order can be raised to $2s - 1$.¹⁵ We will refer to these two cases as operators with diagonal norms and block norms respectively, and the corresponding *SBP - SAT* methods as *SBP*($2s, s$) and *SBP*($2s, 2s - 1$) respectively.

B. The scalar constant coefficient problem

In order to show *A*-stability for the *SBP - SAT* methods with the energy method, we follow the path set in¹² where these methods were first introduced. The *SBP - SAT* approximation of (4) is:

$$P^{-1}Q\vec{U} + \lambda\vec{U} = P^{-1}(\sigma(U_0 - f))\vec{e}_0. \quad (11)$$

where $\vec{e}_0 = (1 \ 0 \ \dots \ 0)^T$. The term in the right-hand-side of (11) forces the solution at $t = 0$ toward the initial condition weakly using the SAT technique. By choosing $\sigma = -1$, the discrete approximation (11) becomes dual consistent,¹ which will prove to be a useful property.

The discrete energy method applied to (11) with $\sigma = -1$ (multiplying from the left with $\vec{U}^*P$ and adding the conjugate transpose) leads to (see¹² for details)

$$|\vec{U}_N|^2 + 2\text{Re}(\lambda)\|\vec{U}\|_P^2 = |f|^2 - |U_0 - f|^2. \quad (12)$$

Compare (12) to the continuous estimate (5). The discrete bound perfectly mimics the continuous one, but has in addition the small (to the order of accuracy of the method) damping term $-|U_0 - f|^2$. This kind of optimally strict estimates is to our knowledge never obtained using local time-stepping methods.

The next two propositions give the order of accuracy of different components of the solution for both non-stiff and stiff problems.

Proposition 1 *The pointwise order of accuracy for $SBP(2s, 2s-1)$ and $SBP(2s, s)$ when applied to problem (3) in the non-stiff case ($|\lambda|h \ll 1$) is $2s - 1$ and s respectively. If moreover the scheme is dual consistent ($\sigma = -1$), then the order of accuracy is $2s$ for linear functionals of the form $\int_0^T g(t)u(t)dt$ of the solution as well as for $u(T)$.*

Thus, if only linear functionals of the solution or the solution in the last time step is sought, then methods based on diagonal norms or block norms have the same order of accuracy. The situation changes in the stiff case however.

Proposition 2 *The pointwise order of accuracy for $SBP(2s, 2s-1)$ and $SBP(2s, s)$ when applied to problem (3) in the stiff case ($|\lambda|h \gg 1$) is $2s - 1$ and s respectively for boundary points, and $2s$ for interior points.*

In other words, Proposition (2) states that the order of accuracy of the $SBP - SAT$ methods coincides with the order of the truncation error of the corresponding operators. There is no gain for linear functionals or the solution in the last time step. The higher order of accuracy for interior points are also only obtained with this simple scalar test equation, and should not be expected for more general problems in higher dimension.

The full proofs of propositions 1 and 2 will be presented in a future article.

The accuracy results in this section are illustrated in Figures 1 through 3 (first shown in¹²), for the manufactured solution $\psi = e^{-t}$ to (3). In the stiff case, the $SBP - SAT$ methods are compared with two popular L -stable time-stepping methods: The $BDF2$ (second order backward differentiation formula) and the previously mentioned $ESDIRK4$, where the former does not suffer from any order reduction. The non-stiff case is represented by $\lambda = 1$, and the stiff case by $\lambda = 1000$.

C. Nonlinear problems

We return to the general non-linear problem (1). An $SBP - SAT$ discretization can be written

$$(P^{-1}Q \otimes I)\vec{U} + \begin{pmatrix} F(t_0, U_0) \\ \vdots \\ F(t_N, U_N) \end{pmatrix} = P^{-1}\sigma\vec{e}_0 \otimes (U_0 - f),$$

where I is a unit matrix with the same dimension as u , and $\vec{e}_0 = (1 \ 0 \ \dots \ 0)^T$ has length $N + 1$. The stability properties of the $SBP - SAT$ methods, if diagonal norms are used, can be summarized in the following proposition:

Proposition 3 *The $SBP(2s, s)$ family of methods are always B -stable and energy stable.*

This result makes the $SBP(2s, s)$ methods particularly useful for non-linear problems in general, and for spatial discretizations of energy stable initial boundary value problems in particular. Unfortunately, the same result does not hold for block norms.

The full proof of Proposition 3 will be presented in a future article.

Figure 1. Convergence of solution at $t = 1$ for the $SBP - SAT$ technique applied to a non-stiff problem ($\lambda = 1$) with solution $u = \exp(-t)$. As can be seen, both $SBP(2s,s)$ and $SBP(2s,s-1)$ converge with order $2s$.

Figure 2. Convergence for $SBP(2s,s)$ at $t = 1$ compared to other implicit methods applied to a stiff problem ($\lambda = 1000$) with solution $u = \exp(-t)$.

D. Multi-block implementation

Potentially, e.g if the problem (1) results from the spatial discretization of an initial boundary value problem, the fully discrete system obtained with an $SBP - SAT$ discretization can be prohibitively big to solve on the whole time interval of interest all at once. It can also be of advantage to be able to use varying time step sizes on different time intervals. To facilitate this we can split the time domain into smaller parts and solve for each part individually.

Assume that the time domain is split into two blocks with an interface at $t = a$, where $0 < a < T$ and introduce corresponding numerical approximations.

$$\vec{U} = (U_o \ U_1 \ \dots \ U_N)^T \approx (u(0) \ u(\Delta t_1) \ \dots \ u(a))^T, \quad \vec{V} = (V_o \ V_1 \ \dots \ V_M)^T \approx (u(a) \ u(a + \Delta t_2) \ \dots \ u(T))^T$$

With the $SBP - SAT$ technique, we can construct a discrete approximation using different operators in the

Figure 3. Convergence for $SBP(2s, 2s-1)$ at $t = 1$ compared to other implicit methods applied to a stiff problem ($\lambda = 1000$) with solution $u = \exp(-t)$.

two blocks as follows:

$$\begin{aligned} P_l^{-1} Q_l \vec{U} + \lambda \vec{U} &= P^{-1}(\sigma(U_0 - f)) \vec{e}_0 + P^{-1}(\tau_l(U_N - V_0)) \vec{e}_N \\ P_r^{-1} Q_r \vec{V} + \lambda \vec{V} &= P^{-1}(\tau_r(V_0 - U_N)) \vec{e}_{N+1}, \end{aligned} \quad (13)$$

where τ_l and τ_r are *SAT* penalty parameters forcing the two solutions U_N and V_0 at $t = a$ toward each other. $\vec{e}_0$, $\vec{e}_N$ and $\vec{e}_{N+1}$ are unit vectors with zeros everywhere except at position 0, N and $N + 1$ respectively. We also define the fully discrete solution $\vec{W} = (U_0 \dots U_N \ V_0 \dots V_M)^T$ as well as the discrete L_2 norm $\|\vec{W}\|_{\bar{P}}^2 = \vec{W}^* \bar{P} \vec{W}$, where

$$\bar{P} = \begin{pmatrix} P_l & \mathbf{0} \\ \mathbf{0} & P_r \end{pmatrix}.$$

The energy method (multiplying from the left with $\vec{W}^* \bar{P}$ and adding the conjugate transpose) then yields

$$|V_M|^2 + 2\operatorname{Re}(\lambda) \|\vec{W}\|_{\bar{P}}^2 + \begin{pmatrix} U_N \\ V_0 \end{pmatrix}^T \begin{pmatrix} 1 - 2\tau_l & \tau_l + \tau_r \\ \tau_l + \tau_r & -1 - 2\sigma_r \end{pmatrix} \begin{pmatrix} U_N \\ V_0 \end{pmatrix} = |U_0|^2 - |U_0 - f|^2. \quad (14)$$

It was originally shown in⁵ that the matrix in expression (14) is non-negative definite if and only if

$$\tau_l = \tau_r + 1, \quad \tau_r \leq -\frac{1}{2}$$

Consider the discrete problem (13) again. The choice $\tau_r = -1$ clearly makes the solution to the first equation in (13) independent of the solution to the second. After the first equation is solved, the solution component U_0 can then be used as initial condition to the second equation.

This technique can easily be expanded to include an arbitrary number of subdomains, on which the solution can then be computed one after the other. This also opens for the possibility of using an alternative method of grid refinement, where the number of blocks is increased while the size of each block remains constant. No order of accuracy is lost even if a minimal number of points is used in each block. Figure 4 compares the performance of this grid refinement technique to the standard one applied to the non-stiff test equation. The minimal number of grid points was always used in each block. This number is determined by the size of the boundary closures in the *SBP* operators.

Coupled with an appropriate error estimation strategy, the blocking technique also allows for the introduction of schemes with adaptive step sizes.

Figure 4. Convergence for $SBP(2s,s)$ at $t = 1$ with the alternative grid refinement strategy involving multi-block implementation, applied on a non-stiff problem ($\lambda = 1$) with solution $u = \exp(-t)$.

III. Conclusions

The $SBP - SAT$ family of methods represents a new technique for high order time-discretization of initial value problems, based on the well known $SBP - SAT$ technique for spatial discretizations of initial value problems. The technique utilizes first derivative operators on summation-by-parts form, and imposes initial conditions weakly.

The concepts of A -stability and B -stability are introduced for this type of methods, that mimic existing definitions for conventional time-stepping methods. The additional concept of energy stability is also introduced for methods that preserve energy estimates for general non-linear problems.

The $SBP - SAT$ methods are then shown to be A -stable and, if operators with diagonal norms are used, also B -stable and energy stable. They can be constructed to give high orders of stiff and non-stiff convergence, but the order of stiff convergence is higher for operators using block norms. The property of energy stability makes methods using diagonal norm operators especially interesting for energy stable spatial discretizations of initial boundary value problems.

The $SBP - SAT$ methods are global and generate optimal energy estimates. Using a multi-domain approach, the solution process can for practical reasons be split into several smaller time domains, the size of each only limited downwards by the minimal size of the operator.

References

- ¹J. Berg and J. Nordström. Superconvergent functional output for time-dependent problems using finite differences on summation-by-parts form. *Journal of Computational Physics*, 231(20):6846–6860, 2012.
- ²H. Bijl, M. Carpenter, Vatsa N.V., and Kennedy C.A. Implicit time integration schemes for the unsteady compressible navier-stokes equations: Laminar flow. *Journal of Computational Physics*, 179(1):313–329, 2002.
- ³J. C. Butcher. General linear methods for stiff differential equations. *BIT Numerical Mathematics*, 41(2):240–264, 2001.
- ⁴J.C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, Ltd, 2005.
- ⁵M.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.
- ⁶B. Gustafsson. *High Order Difference Methods for Time Dependent PDE*. Springer-Verlag, 2008.
- ⁷E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Springer-Verlag, 1980.
- ⁸C. A. Kennedy and M. H. Carpenter. Additive runge-kutta schemes for convection-diffusion-reaction equations. *Applied Numerical Mathematics*, 44(1-2):139–181, 2003.
- ⁹K. Mattsson and J. Nordström. Summation by parts operators for finite difference approximations of second derivatives. *Journal of Computational Physics*, 199(2):503–540, 2004.

- ¹⁰J. Nordström. Conservative finite difference formulations, variable coefficients, energy estimates and artificial dissipation. *Journal of Scientific Computing*, 29(3):375–404, 2006.
- ¹¹J. Nordström. Error bounded schemes for time-dependent hyperbolic problems. *SIAM Journal on Scientific Computing*, 30:46–59, 2007.
- ¹²J. Nordström and T. Lundquist. Summation-by-parts in time. *Accepted in Journal of Computational Physics*, 2013.
- ¹³A. Prothero and A. Robinson. On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations. *Siam Journal on Scientific Computing*, 28(125):145–162, 1974.
- ¹⁴M.N. Spijker. Stiffness in numerical initial-value problems. *Journal of Computational and Applied Mathematics*, 72(2):393–406, 1996.
- ¹⁵B. Strand. Summation by parts for finite difference approximation for d/dx . *Journal of Computational Physics*, 110(1):47–67, 1994.