

Summation by Parts Formula for Noncentered Finite Differences

R. Bodenmann

Research Report No. 95-07
September 1995

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Summation by Parts Formula for Noncentered Finite Differences

R. Bodenmann

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Research Report No. 95-07

September 1995

Abstract

Boundary conditions for general high order finite difference approximations to $\frac{d}{dx}$ are constructed. It is proved that it is always possible to find boundary conditions and a weighted norm in such a way that a summation by parts formula holds.

Keywords: high order finite differences, summation by parts formula, energy method, linear stability, initial-boundary value problem

1. INTRODUCTION

For analysing numerical methods applied to partial differential equations, a priori estimates of the growth rate of the solution are necessary. Such stability estimates ensure that small discretisation errors cannot grow arbitrarily.

Analysing convergence behaviour of finite difference approximations to nonlinear problems, it turns out that stability inequalities of the linearisation of the methods have to be proved [1], [8]. Since schemes which are obtained by linearisation have variable coefficients, the energy method turns out to be one of the most effective devices to derive bounds of the growth rate [2], [5], [7].

One of the most important components of the energy method in the continuous case is the integration by parts formula. To get an energy estimate for the discretised problem, the integration by parts formula has to be replaced by its discrete analog, the summation by parts formula. This is accomplished by constructing the difference operator including numerical boundary conditions in an appropriate manner.

Kreiss and Scherer proved that in general new scalar products and norms have to be introduced to make possible the existence of difference approximations which fulfil a summation by parts formula. They constructed weighted diagonal or even more general types of norms [3] and proved the existence of difference approximations to $\frac{d}{dx}$ of almost antisymmetric type such that the summation by parts formula holds. These results, which are reviewed in the next section, can also be found in [9] together with some examples.

In Section 3, the theory is developed for more general difference operators, which do not need to be antisymmetric. It is pointed out in Section 4 that the results from Section 3 are optimal, even though they are weaker than the ones which can be proved in the antisymmetric case.

An algorithm to compute proper boundary conditions for the diagonal norm case is presented in Section 5 and some of its results are displayed in Section 6.

2. REVIEW OF THE ANTISYMMETRIC CASE

The goal is to construct difference approximations Q to $\frac{d}{dx}$ on the interval $[0, 1]$, which are of arbitrarily high order in the interior as well as on the boundaries of the interval and which fulfil a summation by

parts formula:

$$(u, Qv)_h = -(Qu, v)_h + u_J v_J - u_0 v_0$$

for any two grid functions u and v and for a certain discrete scalar product $(\cdot, \cdot)_h$.

Obviously it is enough to investigate the halfline problem. So let us consider the halfline $[0, \infty)$, which we divide into intervals of length $h > 0$. Denote the gridpoints by $x_j = jh$, $j = 0, 1, \dots$ and let $u_j = u(x_j)$ be a scalar grid function in l_2 , i.e. $h \sum_{j=0}^{\infty} |u_j|^2 < \infty$. A discrete scalar product and a norm are defined by

$$\begin{aligned} (u, v)_h &= \langle u^I, H v^I \rangle_h + h \sum_{j=r}^{\infty} u_j v_j \\ &= h \sum_{i,j=0}^{r-1} h_{i,j} u_i v_j + h \sum_{j=r}^{\infty} u_j v_j, \\ \|u\|_h^2 &= (u, u)_h. \end{aligned} \tag{1}$$

Here $u^I = (u_0, u_1, \dots, u_{r-1})^T$ and $H = H^T > 0$ is a symmetric, positive definite $(r \times r)$ -matrix. We want to construct matrices H and a difference approximation Q to $\frac{d}{dx}$, such that the summation by parts formula

$$(u, Qv)_h = -(Qu, v)_h - u_0 v_0 \tag{2}$$

holds.

The following theorems are proved in [3], [6], and [9].

Theorem 2.1 (Diagonal Norms). *There exist scalar products (1), with H diagonal, and difference operators Q of accuracy p at the boundary and $2p$ in the interior, $1 \leq p \leq 4$, such that the summation by parts formula (2) holds.*

Theorem 2.2 (Full Norms). *There exist scalar products (1) and difference operators Q of accuracy $2p - 1$ at the boundary and $2p$ in the interior, $p > 0$, such that the summation by parts formula (2) holds.*

The proofs are constructive and examples are computed in [9]. The simplest example for the diagonal norm case is given by

$$hQ = \begin{pmatrix} -1 & 1 & & & \\ -0.5 & 0 & 0.5 & & \\ & -0.5 & 0 & 0.5 & \\ & & & \ddots & \ddots & \ddots \end{pmatrix}$$

and $H = \text{diag}(0.5, 1, 1, \dots)$.

3. NONCENTERED DIFFERENCES

We now want to construct pairs of finite difference operators (Q_-, Q_+) and scalar products (1) such that the summation by parts formula

$$(u, Q_-v)_h = -(Q_+u, v)_h - u_0v_0 \quad (3)$$

holds for any scalar grid functions u and v .

Since u and v are considered as infinite vectors, Q_{\mp} are represented as infinite matrices. Assume that Q_{\mp} have the form

$$hQ_{\mp} = \begin{pmatrix} (Q_{\mp})_{11} & (Q_{\mp})_{12} \\ C_{\mp}^T & D_{\mp} \end{pmatrix},$$

with

$$(Q_{\mp})_{11} = \begin{pmatrix} q_{0,0} & q_{0,1} & \cdots & q_{0,r-1} \\ \vdots & \vdots & & \vdots \\ q_{r-1,0} & q_{r-1,1} & \cdots & q_{r-1,r-1} \end{pmatrix}_{\mp},$$

$$(Q_{\mp})_{12} = \begin{pmatrix} q_{0,r} & \cdots & q_{0,m} & 0 & \cdots \\ \vdots & & \vdots & \vdots & \\ q_{r-1,r} & \cdots & q_{r-1,m} & 0 & \cdots \end{pmatrix}_{\mp},$$

$$D_- = \begin{pmatrix} \alpha_0 & \alpha_1 & \cdots & \alpha_R & 0 & 0 & \cdots & 0 & 0 & \cdots \\ \alpha_{-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_R & 0 & \cdots & 0 & 0 & \cdots \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \vdots & \vdots & \\ \alpha_{-L} & \cdots & \alpha_{-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_R & 0 & 0 & \cdots \\ 0 & \alpha_{-L} & \cdots & \alpha_{-1} & \alpha_0 & \alpha_1 & \cdots & \alpha_R & 0 & \cdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \ddots & & \ddots & \ddots \end{pmatrix},$$

$$D_+ = -D_-^T,$$

$$C_{\mp} = \begin{pmatrix} C_{\mp}^0 & 0 & \cdots \\ \tilde{C}_{\mp} & 0 & \cdots \end{pmatrix},$$

$$\tilde{C}_- = \begin{pmatrix} \alpha_{-L} & 0 & 0 & \cdots & 0 \\ \alpha_{-L+1} & \alpha_{-L} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha_{-2} & \alpha_{-3} & \cdots & \alpha_{-L} & 0 \\ \alpha_{-1} & \alpha_{-2} & \cdots & \alpha_{-L+1} & \alpha_{-L} \end{pmatrix},$$

$$\tilde{C}_+ = - \begin{pmatrix} \alpha_R & 0 & 0 & \cdots & 0 \\ \alpha_{R-1} & \alpha_R & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \alpha_2 & \alpha_3 & \cdots & \alpha_R & 0 \\ \alpha_1 & \alpha_2 & \cdots & \alpha_{R-1} & \alpha_R \end{pmatrix},$$

C_-^0 and C_+^0 are $(r-L) \times L$ and $(r-R) \times R$ matrices respectively with only zeroes.

Theorem 3.1. *The operators Q_- and Q_+ satisfy the relation (3) if and only if they can be written as*

$$hQ_- = \begin{pmatrix} H^{-1}A & -H^{-1}C_+ \\ C_-^T & D_- \end{pmatrix}, \quad hQ_+ = \begin{pmatrix} H^{-1}B & -H^{-1}C_- \\ C_+^T & D_+ \end{pmatrix}, \quad (4)$$

where A and B are $(r \times r)$ -matrices of the form $A = A_1 + A_2$, $B = B_1 + B_2$ with

$$A_1 = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad B_1 = \begin{pmatrix} b_{11} & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \quad a_{11} + b_{11} = -1,$$

and $A_2 + B_2^T = 0$.

Proof. We can formally write (3) as

$$\begin{aligned} -u_0 v_0 &= \langle u^I, H(Q_-)_{11} v^I \rangle + \langle u^I, H(Q_-)_{12} v^{II} \rangle \\ &\quad + \langle u^{II}, C_-^T v^I \rangle + \langle u^{II}, D_- v^{II} \rangle + \langle (Q_+)_{11} u^I, H v^I \rangle \\ &\quad + \langle (Q_+)_{12} u^{II}, H v^I \rangle + \langle C_+^T u^I, v^{II} \rangle + \langle D_+ u^{II}, v^{II} \rangle \end{aligned}$$

and by making use of the symmetry of H

$$\begin{aligned} -u_0 v_0 &= \langle u^I, H(Q_+)_{11} v^I \rangle + \langle u^I, H(Q_+)_{12} v^{II} \rangle \\ &\quad + \langle u^{II}, C_+^T v^I \rangle + \langle u^{II}, D_+ v^{II} \rangle + \langle (Q_-)_{11} u^I, H v^I \rangle \\ &\quad + \langle (Q_-)_{12} u^{II}, H v^I \rangle + \langle C_-^T u^I, v^{II} \rangle + \langle D_- u^{II}, v^{II} \rangle, \end{aligned}$$

where $u^I = (u_0, u_1, \dots, u_{r-1})^T$ and $u^{II} = (u_r, u_{r+1}, \dots)^T$. For D_{\mp} we have the relation $\langle u^{II}, D_- v^{II} \rangle + \langle D_+ u^{II}, v^{II} \rangle = 0$. Letting $v^I = 0$, we get

$$\langle u^I, H(Q_-)_{12} v^{II} \rangle + \langle C_+^T u^I, v^{II} \rangle = \langle u^I, (H(Q_-)_{12} + C_+) v^{II} \rangle = 0$$

and

$$\langle u^I, H(Q_+)_{12} v^{II} \rangle + \langle C_-^T u^I, v^{II} \rangle = \langle u^I, (H(Q_+)_{12} + C_-) v^{II} \rangle = 0$$

for all vectors u^I and v^{II} . Therefore we have

$$(Q_-)_{12} = -H^{-1}C_+ \quad \text{and} \quad (Q_+)_{12} = -H^{-1}C_-.$$

Only the relation

$$\langle u^I, (H(Q_-)_{11} + (Q_+^T)_{11}H)v^I \rangle = -u_0v_0$$

remains. Therefore $H(Q_-)_{11} = A$ and $H(Q_+)_{11} = B$ must be of the required form. \square

Let $h = 1$ and denote by

$$w_l = \begin{pmatrix} e_l \\ f_l \end{pmatrix}, \quad e_l = (-1)^l \begin{pmatrix} r^l \\ (r-1)^l \\ \vdots \\ 1^l \end{pmatrix}, \quad f_l = \begin{pmatrix} 0^l \\ 1^l \\ \vdots \end{pmatrix}, \quad l = 0, 1, \dots$$

the discretisation of $(x-r)^l$, with the conventions $0^0 = 1$, and $e_{-1} = 0$.

The following lemma characterises the accuracy of Q_{\mp} in the interior.

Lemma 3.2. *The operators $h^{-1}(C_{\mp}^T, D_{\mp})$ approximate $\frac{d}{dx}$ with order q at the points x_j , $j = r, r+1, \dots$ if and only if*

$$\sum_{\nu=-L}^R \alpha_{\nu} \nu^p = \begin{cases} 1, & p = 1, \\ 0, & p = 0, 2, 3, \dots, q. \end{cases} \quad (5)$$

\square

Accuracy conditions for Q_{\mp} at boundary points are given by the next lemma.

Lemma 3.3. *The operators $h^{-1}((Q_{\mp})_{11}, (Q_{\mp})_{12})$ approximate $\frac{d}{dx}$ with order τ at the points x_j , $j = 0, 1, \dots, r-1$ if and only if*

$$le_{l-1} = (Q_{\mp})_{11}e_l + (Q_{\mp})_{12}f_l, \quad l = 0, 1, \dots, \tau. \quad (6)$$

\square

We now want to calculate the coefficients of Q_{\mp} at the boundary points. To do so, we first rewrite (6). By (4) we get

$$\begin{aligned} A_2 e_l &= g_l^-, & g_l^- &= lH e_{l-1} - A_1 e_l + C_+ f_l, & l &= 0, 1, \dots, \tau, \\ B_2 e_l &= g_l^+, & g_l^+ &= lH e_{l-1} - B_1 e_l + C_- f_l, & l &= 0, 1, \dots, \tau. \end{aligned} \quad (7)$$

Since $A_2 + B_2^T = 0$, the following compatibility conditions must hold

$$\langle e_l, g_m^- \rangle + \langle e_m, g_l^+ \rangle = 0, \quad 0 \leq m, l \leq \tau. \quad (8)$$

If these conditions are fulfilled, the system (7) can be resolved as it is expressed in Lemma 3.4.

Lemma 3.4. *Assume that $r \geq \tau + 1$ and that the relations (8) hold. Then there are matrices A_2 and B_2 with $A_2 + B_2^T = 0$ such that (7) is valid. \square*

We have to show that there exists a symmetric, positive definite matrix H such that the compatibility conditions (8) hold. By (7) we know that such a matrix H has to fulfil

$$m \langle e_l, H e_{m-1} \rangle + l \langle e_m, H e_{l-1} \rangle = M_{l,m}, \quad 0 \leq l, m \leq \tau, \quad (9)$$

where

$$\begin{aligned} M_{l,m} &= \langle e_l, A_1 e_m \rangle - \langle e_l, C_+ f_m \rangle + \langle e_m, B_1 e_l \rangle - \langle e_m, C_- f_l \rangle \\ &= \langle e_l, (A_1 + B_1) e_m \rangle - \langle e_l, C_+ f_m \rangle - \langle e_m, C_- f_l \rangle. \end{aligned}$$

Lemma 3.5. $M_{l,m}$ can be written as

$$\begin{aligned} M_{l,m} &= -(-r)^{l+m} + J_{l,l+m} \\ J_{l,\sigma} &= - \sum_{\nu=1}^L \alpha_{-\nu} \sum_{\mu=0}^{\nu-1} (\mu - \nu)^{\sigma-l} \mu^l \\ &\quad + \sum_{\nu=1}^R \alpha_{\nu} \sum_{\mu=0}^{\nu-1} (\mu - \nu)^l \mu^{\sigma-l}, \quad \sigma = l + m. \end{aligned} \quad (10)$$

Proof. The scalar products can be calculated as

$$\begin{aligned} \langle e_l, (A_1 + B_1) e_m \rangle &= (-1)^{l+m+1} r^{l+m}, \\ \langle e_m, C_- f_l \rangle &= (-1)^m \sum_{\nu=1}^L \alpha_{-\nu} \sum_{\mu=0}^{\nu-1} (\nu - \mu)^m \mu^l \\ \langle e_l, C_+ f_m \rangle &= (-1)^l \sum_{\nu=1}^R \alpha_{\nu} \sum_{\mu=0}^{\nu-1} (\nu - \mu)^l \mu^m \end{aligned}$$

The proof is concluded by using these expressions in (9) and by introducing $\sigma = l + m$. \square

If we introduce the notation $\rho_{l,m} = \langle e_l, H e_m \rangle$, then we can write (9) as

$$m \rho_{l,m-1} + l \rho_{m,l-1} = M_{l,m}, \quad 0 \leq l, m \leq \tau. \quad (11)$$

Here $\rho_{l,-1} = 0$ by the convention for e_{-1} and $\rho_{l,m} = \rho_{m,l}$ by the symmetry of H . From the solution of this system in the $\rho_{l,m}$ we obtain the elements of H . The matrix H has to be positive definite in order to be used as a norm matrix. The equivalence between this condition and the positive definiteness of the matrix defined by the $\rho_{l,m}$ is stated by Lemma 3.9 below.

In the next step we derive conditions for $M_{l,m}$ such that the system (11) has a solution, and therefore the compatibility conditions hold. These conditions are resumed in Lemma 3.7.

For $l = m = 0$ we obtain from (11) and (10)

$$0 = 0\rho_{0,-1} + 0\rho_{0,-1} = M_{0,0} = -1 + \sum_{\nu=-L}^R \nu\alpha_\nu, \quad \text{i.e.} \quad \sum_{\nu=-L}^R \nu\alpha_\nu = 1. \quad (12)$$

By Lemma 3.2 this is fulfilled if the approximation in the interior is at least of order 1.

Since the left hand side of (11) is symmetric, the same has to be true for the right hand side. The following lemma gives conditions for $M_{l,m}$ to be symmetric.

Lemma 3.6. *If the approximation in the interior is at least of order 2τ , then*

$$M_{l,m} = M_{m,l}, \quad 0 \leq l, m \leq \tau.$$

Proof. We have to show that $J_{l,\sigma} = J_{m,\sigma}$ for $0 \leq l, m \leq \tau, l+m = \sigma$. We introduce the following abbreviations:

$$N_{l,\sigma}(\nu) = \sum_{\mu=0}^{\nu-1} \mu^{\sigma-l} (\mu - \nu)^l + \sum_{\mu=1}^{\nu} (\mu - \nu)^l \mu^{\sigma-l}, \quad 0 \leq l \leq \sigma$$

and

$$\widetilde{N}_{l,\sigma}(\nu) = \sum_{\mu=0}^{\nu-1} \mu^{\sigma-l} (\mu - \nu)^l - \sum_{\mu=1}^{\nu} (\mu - \nu)^l \mu^{\sigma-l}, \quad 0 \leq l \leq \sigma.$$

With

$$\sum_{\mu=0}^{\nu-1} \mu^l (\mu - \nu)^{\sigma-l} = (-1)^\sigma \sum_{\mu=1}^{\nu} (\mu - \nu)^l \mu^{\sigma-l}$$

and letting $s = \max(L, R)$, $\alpha_{-\nu} = 0$, $\nu > L$, and $\alpha_\nu = 0$, $\nu > R$, we get

$$J_{l,\sigma} - J_{m,\sigma} = \begin{cases} \sum_{\nu=1}^s (\alpha_\nu + \alpha_{-\nu}) \widetilde{N}_{l,\sigma}(\nu), & \text{if } \sigma \text{ even,} \\ \sum_{\nu=1}^s (\alpha_\nu + \alpha_{-\nu}) N_{l,\sigma}(\nu), & \text{if } \sigma \text{ odd.} \end{cases} \quad (13)$$

In [3] and [9] it is proved that $N_{l,\sigma}(\nu)$ is of the following form:

$$N_{l,\sigma}(\nu) = \gamma_1^{(l,\sigma)} \nu^{\sigma+1} + \gamma_3^{(l,\sigma)} \nu^{\sigma-1} + \dots, \quad \gamma_1^{(l,\sigma)} \neq 0.$$

For $\widetilde{N}_{l,\sigma}(\nu)$ we get

$$\widetilde{N}_{l,\sigma}(\nu) = 0^{\sigma-l}(-\nu)^l - 0^l\nu^{\sigma-l} = \begin{cases} 0, & l \neq 0 \text{ and } l \neq \sigma, \\ \nu^\sigma, & l = 0, \\ (-\nu)^\sigma, & l = \sigma. \end{cases}$$

The proof is completed by using these expressions for $N_{l,\sigma}(\nu)$ and $\widetilde{N}_{l,\sigma}(\nu)$ in (13) and by Lemma 3.2. \square

We can proceed now as in [3] and [9] for the antisymmetric case. Suppose that the conditions of Lemma 3.6 are fulfilled, then $M_{l,m} = M_{m,l}$, $l, m \leq \tau$ and therefore (11) has to be considered only for $l \leq m \leq \tau$. For $l = 0$ we get the following condition from (11)

$$\rho_{0,m-1} = \frac{1}{m}M_{0,m}, \quad m = 1, 2, \dots, \tau.$$

If $l > 0$, then $\rho_{l,m-1}$ can be calculated explicitly

$$\rho_{l,m-1} = \frac{1}{m}M_{l,m} - \frac{l}{m}\rho_{l-1,m}.$$

Now (11) can be written recursively as

$$\begin{aligned} \rho_{l,m} &= \rho_{m,l}, \quad \rho_{l,-1} = \rho_{-1,l} = 0, \\ \rho_{l,m-1} &= \frac{1}{m}M_{l,m} + \sum_{k=1}^{\alpha} (-1)^k \frac{l!(m-1)!}{(l-k)!(m+k)!} M_{l-k,m+k} \\ &\quad + (-1)^{\alpha+1} \frac{l!(m-1)!}{(l-\alpha-1)!(m+\alpha)!} \rho_{l-\alpha-1,m+\alpha}, \quad (14) \\ &\quad 0 < l \leq m \leq \tau, \end{aligned}$$

where $\alpha = \min(l-1, \tau-m)$.

If $l-1 < \tau-m$, i.e. $l+m \leq \tau$, then $\alpha = l-1$ and the following condition is obtained

$$\rho_{l-\alpha-1,m+\alpha} = \rho_{0,l+m-1} = \frac{1}{l+m}M_{0,l+m}. \quad (15)$$

Thus (12) and (14) imply that $\rho_{l,m-1}$ is completely determined by the $M_{l,m}$ provided $l+m \leq \tau$.

If $l+m > \tau$ then $\alpha = \tau-m$ and

$$\rho_{l-\alpha-1,m+\alpha} = \rho_{l+m-1-\tau,\tau} = \rho_{\nu,\tau}, \quad \nu = l+m-1-\tau. \quad (16)$$

There are no further relations that $\rho_{\nu,\tau}$ needs to satisfy.

If we use equation (14) with $l = n$, $m = n$ and $l = n-1$, $m = n+1$, for $n \leq \tau$, then we obtain representations for $\rho_{n,n-1}$ and $\rho_{n-1,n}$ but by (11) $\rho_{n,n-1} = \rho_{n-1,n}$. If $l+m = 2n > \tau$, then by (16) these two

relations determine $\rho_{\nu,\tau}$, $\nu = 2n - 1 - \tau$ and no conditions for the $M_{l,m}$ result. If $l + m = 2n \leq \tau$ we get from (14) and (15)

$$\begin{aligned} \sum_{k=0}^n (-1)^k \frac{n!(n-1)!}{(n+k)!(n-k)!} M_{n-k,n+k} &= \rho_{n,n-1} = \rho_{n-1,n} \\ &= \sum_{k=1}^n (-1)^{k-1} \frac{(n-1)!n!}{(n+k)!(n-k)!} M_{n-k,n+k} \end{aligned}$$

This relation can be written as

$$\frac{1}{2n} M_{n,n} + \sum_{k=1}^n (-1)^k \frac{(n-1)!n!}{(n+k)!(n-k)!} M_{n-k,n+k} = 0 \quad (17)$$

for $n = 1, 2, \dots$, with $2n \leq \tau$.

We obtain

Lemma 3.7. *If the difference approximation in the interior is accurate at least to the order of 2τ and the $M_{l,m}$ satisfy the relations (17), then the system (11) has a solution. \square*

If the system (11) has a solution, then the $\rho_{l,m}$ are determined for $0 \leq l, m \leq \tau$ if one specifies those $\rho_{\nu,\tau}$, $\nu = 0, 1, \dots, \tau$, which are not determined by the system. They can be used to define a symmetric matrix $R_\tau = (\rho_{l,m})_{0 \leq l, m \leq \tau}$. The following lemma states that it is always possible to choose the parameters $\rho_{\nu,\tau}$ so that the matrix R_τ becomes positive definite.

Lemma 3.8. *If $R_{\tau-1} = (\rho_{l,m})_{0 \leq l, m \leq \tau-1} = R_{\tau-1}^T$ is positive definite then one can choose $\rho_{\tau,\tau}$ such that R_τ is symmetric and positive definite as well, independently of the values of $\rho_{\nu,\tau}$, $\nu = 0, 1, \dots, \tau - 1$. \square*

We now have to determine a positive definite $(r \times r)$ -matrix H such that

$$\langle e_l, H e_m \rangle = \rho_{l,m}, \quad 0 \leq l, m \leq \tau. \quad (18)$$

In fact the following lemma holds.

Lemma 3.9. *If $r \geq \tau + 1$ and the matrix R_τ is positive definite, then there are $H = H^T > 0$ such that (18) holds. In particular if $r = \tau + 1$, then H is uniquely defined by*

$$E^T H E = R_\tau, \quad E = (e_0, e_1, \dots, e_{r-1}).$$

\square

Now we can state the main result.

Theorem 3.10. *For every $q = 2, 3, \dots$ and for a sufficiently large r there is a symmetric, positive definite matrix H , a scalar product of the form (1), and approximations Q_{\mp} of $\frac{d}{dx}$ that are accurate with order $\tau = \lfloor \frac{q}{2} \rfloor$ for $x = x_j$, $j = 0, 1, \dots, r-1$, and accurate with order q for $x = x_j$, $j \geq r$ such that (3) holds. \square*

To prove stability of difference approximations to systems of partial differential equations, it is more convenient to have so-called restricted full norms [4], defined by a symmetric and positive definite matrix of the form

$$H = \begin{pmatrix} \lambda_0 & 0 & \cdots & 0 \\ 0 & h_{1,1} & \cdots & h_{1,r-1} \\ \vdots & \vdots & & \vdots \\ 0 & h_{r-1,1} & \cdots & h_{r-1,r-1} \end{pmatrix}. \quad (19)$$

The following lemma holds.

Lemma 3.11. *If R_{τ} is positive definite, then one can choose H in the form (19) such that (18) holds. In general one has to choose $r \geq \tau + 2$. \square*

Finally we can rewrite Theorem 3.10 for restricted full norms.

Theorem 3.12. *For every $q = 2, 3, \dots$ there is a symmetric, positive definite matrix H of the form (19), a scalar product of the form (1), and approximations Q_{\mp} of $\frac{d}{dx}$, which are accurate with order $\tau = \lfloor \frac{q}{2} \rfloor$ for $x = x_j$, $j = 0, 1, \dots, r-1$, and accurate with order q for $x = x_j$, $j \geq r$ such that (3) holds. \square*

The proofs of the above lemmata and theorems exactly follow those in [3] and in [9].

4. CONCLUDING REMARKS

Comparing the results of the antisymmetric case in Theorem 2.2 with those for noncentered schemes – Theorem 3.10 and Theorem 3.12 – we see that for the antisymmetric case we only need a scheme of order $\tau + 1$ in the interior to get boundary conditions of order τ , whereas for noncentered schemes, we need order 2τ (respectively $2\tau + 1$) in the interior to get order τ at the boundary. The reason is that for an antisymmetric scheme we always have $\alpha_{\nu} + \alpha_{-\nu} = 0$ by definition. Therefore (13) reads $J_{l,\sigma} - J_{m,\sigma} = 0$ and the statement of Lemma 3.6 is always true without any accuracy assumptions on the interior scheme and the only interior accuracy assumptions have to be made in order to fulfil the conditions (17), cf. [9].

5. ALGORITHMS

We now want to present algorithms to compute the norm matrix H and the boundary conditions A respectively B . For the sake of simplicity we assume H to be diagonal. From Theorem 2.1 we know that in the antisymmetric case such norms exist for low order schemes. It is still possible to construct diagonal norms for higher order schemes but then one has to take a scheme with order $q \gg 2\tau$ in the interior to get boundary conditions of order τ . As we have seen in Section 3, the theory for antisymmetric schemes can be generalised to noncentered schemes and similar results can be proved. That is why one can expect results for noncentered schemes that are similar to Theorem 2.1, but we should reckon with the same problems with schemes of higher order.

The following notation is used in the algorithms, which are written in Maple:

$$a_{ij} = (Q_-)_{ij}, \quad b_{ij} = (Q_+)_{ij}, \quad 1 \leq i, j \leq r + s, \quad s = \max(L, R).$$

The conditions $A_{11} + B_{11} = -1$ and $A_2 + B_2^T = 0$ are designated as ‘structure assumptions’ and the accuracy conditions are the ones from formula (5). The matrices C_{\mp} are called ‘interfaces’. This is where the interior scheme enters the computations.

5.1. Third order scheme. The first algorithm computes boundary conditions of first order for the scheme with $\alpha_{-2} = \frac{1}{6}$, $\alpha_{-1} = -1$, $\alpha_0 = \frac{1}{2}$, $\alpha_1 = \frac{1}{3}$, which is accurate of third order.

```

r:=2:
s:=2:
# norm
normout := { seq (lambda[i]=1, i=r+1..r+s) }:
eqns := normout:
# interfaces
inta := { seq (seq (a[i,j]=0,j=1..r-s), i=r+1..r+s)}:
intb := { seq (seq (b[i,j]=0,j=1..r-s), i=r+1..r+s)}:
eqns := eqns union inta union intb:
# essential interfaces
intea := { a[r+1,r-1]=1/6, a[r+1,r]=-1, a[r+2,r-1]=0, a[r+2,r]=1/6}:
inteb := { b[r+1,r-1]=0, b[r+1,r]=-1/3, b[r+2,r-1]=0, b[r+2,r]=0}:
eqns := eqns union intea union inteb:
# structure assumptions
bdterm := { lambda[1]*(a[1,1]+b[1,1])=-1}:
adj1 := { seq( seq( lambda[i]*a[i,j]=-lambda[j]*b[j,i],i=1..r), j=2..r+s)}:
adj2 := { seq( seq( lambda[i]*a[i,j]=-lambda[j]*b[j,i],i=2..r+s), j=1..r)}:
eqns := eqns union bdterm union adj1 union adj2:
# accuracy conditions
ord0a := { seq (value(Sum(a[i,l], l=1..r+s)=0),i=1..r)}:
ord0b := { seq (value(Sum(b[i,l], l=1..r+s)=0),i=1..r)}:
ord1a := { seq (value(Sum((1-i)*a[i,l], l=1..r+s)=1),i=1..r)}:
ord1b := { seq (value(Sum((1-i)*b[i,l], l=1..r+s)=1),i=1..r)}:
eqns := eqns union ord0a union ord0b union ord1a union ord1b;
solve(eqns);

```

5.2. Fourth order scheme. Now we are going to take a fourth order scheme in the interior in order to get boundary conditions of second order. The interior scheme is defined by $\alpha_{-3} = -\frac{1}{12}$, $\alpha_{-2} = \frac{1}{2}$, $\alpha_{-1} = -\frac{3}{2}$, $\alpha_0 = \frac{5}{6}$, $\alpha_1 = \frac{1}{4}$. Therefore we have to change the first two lines in the algorithm to

```

r:=4:
s:=3:

```

The interfaces now read

```

# interfaces
inta := { seq (seq (a[i,j]=0,j=1..r-s), i=r+1..r+s)}:
intb := { seq (seq (b[i,j]=0,j=1..r-s), i=r+1..r+s)}:

```

```
# essential interfaces
intea := { a[r+1,r-2]=-1/12, a[r+1,r-1]=1/2, a[r+1,r]=-3/2, a[r+2,r-2]=0,
  a[r+2,r-1]=-1/12, a[r+2,r]=1/2, a[r+3,r-2]=0, a[r+3,r-1]=0,a[r+3,r]=-1/12}:
inteb := { b[r+1,r-2]=0, b[r+1,r-1]=0, b[r+1,r]=-1/4,
  seq (seq (b[r+i,r-j]=0,j=0..s-1), i=2..s)}:
```

Finally the accuracy conditions are changed to

```
# accuracy conditions
ord0a := { seq (value(Sum(a[i,l], l=1..r+s)=0),i=1..r)}:
ord0b := { seq (value(Sum(b[i,l], l=1..r+s)=0),i=1..r)}:
ord1a := { seq (value(Sum((l-i)*a[i,l], l=1..r+s)=1),i=1..r)}:
ord1b := { seq (value(Sum((l-i)*b[i,l], l=1..r+s)=1),i=1..r)}:
ord2a := { seq (value(Sum((l-i)^2*a[i,l], l=1..r+s)=0),i=1..r)}:
ord2b := { seq (value(Sum((l-i)^2*b[i,l], l=1..r+s)=0),i=1..r)}:
```

6. NUMERICAL RESULTS

For the third order scheme from Section 5 we get the following normmatrix H and boundary conditions A and B , which are of first order

$$H = \text{diag}\left(\frac{5}{12}, \frac{13}{12}\right), \quad H^{-1}A = \begin{pmatrix} -1 & 1 \\ -\frac{9}{13} & \frac{5}{13} \end{pmatrix}, \quad H^{-1}B = \begin{pmatrix} -\frac{7}{5} & \frac{9}{5} \\ -\frac{5}{13} & -\frac{5}{13} \end{pmatrix}.$$

The whole difference approximations Q_- and Q_+ are

$$hQ_- = \begin{pmatrix} -1 & 1 & & & & \\ -\frac{9}{13} & \frac{5}{13} & \frac{4}{13} & & & \\ \frac{1}{6} & -1 & \frac{1}{2} & \frac{1}{3} & & \\ & \frac{1}{6} & -1 & \frac{1}{2} & \frac{1}{3} & \\ & & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}$$

and

$$hQ_+ = \begin{pmatrix} -\frac{7}{5} & \frac{9}{5} & -\frac{2}{5} & & & \\ -\frac{5}{13} & -\frac{5}{13} & \frac{12}{13} & -\frac{2}{13} & & \\ & -\frac{1}{3} & -\frac{1}{2} & 1 & -\frac{1}{6} & \\ & & -\frac{1}{3} & -\frac{1}{2} & 1 & -\frac{1}{6} \\ & & & \ddots & \ddots & \ddots & \ddots \end{pmatrix}.$$

The algorithm for the fourth order difference approximations from Section 5 gives a one parameter family of pairs of difference operators.

- [6] G. SCHERER, *Numerical computations with energy estimates schemes*. Technical report, Dept. of Scientific Computing, Uppsala University, 1977.
- [7] ———, *On energy estimates for difference approximations to hyperbolic partial differential equations*. Ph.D. thesis, Dept. of Scientific Computing, Uppsala University, 1977.
- [8] H. J. SCHROLL, *Convergence of implicit finite difference methods applied to nonlinear mixed systems*. Preprint, Dep. of Informatics, Univ. of Oslo, Norway 1993; to appear in SIAM J. Numer. Anal.
- [9] B. STRAND, *Summation by parts for finite difference approximations for $\frac{d}{dx}$* , J. Comp. Phys., 110 (1994).

Research Reports

No.	Authors	Title
95-07	R. Bodenmann	Summation by parts formula for noncentered finite differences
95-06	M.D. Buhmann	Neue und alte These über Wavelets
95-05	M. Fey, A.-T. Morel	Multidimensional method of transport for the shallow water equations
95-04	R. Bodenmann, H.J. Schroll	Compact difference methods applied to initial-boundary value problems for mixed systems
95-03	K. Nipp, D. Stoffer	Invariant manifolds and global error estimates of numerical integration schemes applied to stiff systems of singular perturbation type - Part II: Linear multistep methods
95-02	M.D. Buhmann, F. Derrien, A. Le Méhauté	Spectral Properties and Knot Removal for Interpolation by Pure Radial Sums
95-01	R. Jeltsch, R. Renaut, J.H. Smit	An Accuracy Barrier for Stable Three-Time-Level Difference Schemes for Hyperbolic Equations
94-13	J. Waldvogel	Circuits in Power Electronics
94-12	A. Williams, K. Burrage	A parallel implementation of a deflation algorithm for systems of linear equations
94-11	N. Botta, R. Jeltsch	A numerical method for unsteady flows
94-10	M. Rezny	Parallel Implementation of ADVISE on the Intel Paragon
94-09	M.D. Buhmann, A. Le Méhauté	Knot removal with radial function interpolation
94-08	M.D. Buhmann	Pre-wavelets on scattered knots and from radial function spaces: A review
94-07	P. Klingenstein	Hyperbolic conservation laws with source terms: Errors of the shock location
94-06	M.D. Buhmann	Multiquadric Pre-Wavelets on Non-Equally Spaced Centres
94-05	K. Burrage, A. Williams, J. Erhel, B. Pohl	The implementation of a Generalized Cross Validation algorithm using deflation techniques for linear systems
94-04	J. Erhel, K. Burrage, B. Pohl	Restarted GMRES preconditioned by deflation
94-03	L. Lau, M. Rezny, J. Belward, K. Burrage, B. Pohl	ADVISE - Agricultural Developmental Visualisation Interactive Software Environment
94-02	K. Burrage, J. Erhel, B. Pohl	A deflation technique for linear systems of equations
94-01	R. Sperb	An alternative to Ewald sums