MATHEMATICS OF COMPUTATION
VOLUME 64, NUMBER 211
JULY 1995, PAGES 1035–1065

SUMMATION BY PARTS, PROJECTIONS, AND STABILITY. I

PELLE OLSSON

ABSTRACT. We have derived stability results for high-order finite difference approximations of mixed hyperbolic-parabolic initial-boundary value problems (IBVP). The results are obtained using summation by parts and a new way of representing general linear boundary conditions as an orthogonal projection. By rearranging the analytic equations slightly, we can prove strict stability for hyperbolic-parabolic IBVP. Furthermore, we generalize our technique so as to yield stability on nonsmooth domains in two space dimensions. Using the same procedure, one can prove stability in higher dimensions as well.

1. INTRODUCTION

When solving a partial differential equation numerically it is necessary to have some bound of the growth rate of the solution, since otherwise roundoff errors could grow arbitrarily fast. This upper bound can be established by ensuring some kind of stability. We have elected to use the energy method, because it can be applied to the continuous as well as the discrete model. Furthermore, it can be applied to general domains, which is important when studying multidimensional problems.

Stability of the continuous problem is established by means of an integration-by-parts procedure introducing boundary terms, some of which must be eliminated to ensure stability. For the finite difference model integration by parts is replaced by summation by parts. This amounts to designing the discrete difference operator ensuring that, in addition to the accuracy requirements, certain conditions of antisymmetry are met. As a consequence, the common problem of finding proper "numerical" boundary conditions will be eliminated; they will be built in the discrete difference operator.

The analytic boundary conditions are yet to be incorporated. We propose a certain projection operator, which interacts with the difference operator so as to generate boundary terms that are completely analogous to those of the continuous problem. This can be done for any type of linear boundary conditions. Thus, an energy estimate is obtained for the discrete problem, provided there is one for the analytic model. This conclusion remains true for domains in several space dimensions, even if the boundary is nonsmooth. Furthermore, using this projection operator allows us to derive stability results for a larger class of finite difference operators than those considered in [5]. Stability will be proved

Received by the editor January 19, 1994 and, in revised form, August 5, 1994.
1991 *Mathematics Subject Classification.* Primary 65M06, 65M12.
This work has been sponsored by NASA under contract No. NAS 2-13721.

©1995 American Mathematical Society

for high-order finite difference approximations of mixed hyperbolic-parabolic variable-coefficient systems subject to general boundary conditions.

1.1. **An introductory example.** To illustrate the underlying principles of the energy method, we consider the convection-diffusion equation

$$\begin{aligned} &u_t = u_{xx} + u_x\,, \quad x \in (0\,,\,1)\,, \quad t > 0\,,\\ &u(x\,,\,0) = f(x)\,,\\ &u(0\,,\,t) = 0\,,\\ &u_x(1\,,\,t) = g(t)\,. \end{aligned}$$

In the sequel we shall use the standard L^2-scalar product

$$(u\,,\,v) = \int_0^1 uv\,dx$$

with the corresponding norm defined as $||u||^2 = (u\,,\,u)$.

We can obtain an a priori estimate for this example using the following tools.

(i) Integration by parts:

$$\frac{d}{dt}||u||^2 = 2(u\,,\,u_{xx}) + 2(u\,,\,u_x) = -2||u_x||^2 + 2(u\,,\,u_x) + 2uu_x|_0^1\,.$$

(ii) Boundary conditions:

$$\frac{d}{dt}||u||^2 = -2||u_x||^2 + 2(u\,,\,u_x) + 2u(1\,,\,t)g(1\,,\,t)\,.$$

(iii) Cauchy-Schwarz inequality:

$$\frac{d}{dt}||u||^2 \le -2||u_x||^2 + 2||u||||u_x|| + 2u(1\,,\,t)g(1\,,\,t)\,.$$

(iv) Algebraic inequality:

$$2|xy| \le \epsilon x^2 + \epsilon^{-1}y^2$$

implies ($\epsilon = 1$)

$$\frac{d}{dt}||u||^2 \le -||u_x||^2 + ||u||^2 + u(1\,,\,t)^2 + g(1\,,\,t)^2\,.$$

(v) Sobolev inequality:

$$|u|_\infty^2 \le \epsilon||u_x||^2 + (\epsilon^{-1} + 1)||u||^2$$

is used to eliminate $u(1\,,\,t)$ ($\epsilon = 1$)

$$\frac{d}{dt}||u||^2 \le 3||u||^2 + g(1\,,\,t)^2\,,$$

which can be solved analytically to yield

$$||u(\cdot\,,\,t)||^2 \le e^{3t}\left(||f||^2 + \int_0^t g(\tau)^2 d\tau\right).$$

If we are to obtain such an estimate for a system of equations we will also need

(vi) The adjoint of A:

$$(u\,,\,Av) = (A^T u\,,\,v)\,.$$

Summing up, the energy method boils down to the six basic "tools" above. In the subsequent sections we shall see how these principles can be modified so as to give an energy estimate for the semidiscrete system.

2. General principles for the semidiscrete case

In this section the basic principles of the energy method will be transferred to the semidiscrete case. Furthermore, a number of lemmas, which will be needed later, will be stated. Throughout this section grid vectors will be denoted by $v^T = (v_0^T \dots v_\nu^T)$, $v_j \in \mathbf{R}^d$. Difference operators approximating $\partial/\partial x$ will be designated by

$$D = \frac{1}{h}\begin{pmatrix} d_{00}I & \dots & d_{0\nu}I \\ \vdots & & \vdots \\ d_{\nu 0}I & \dots & d_{\nu\nu}I \end{pmatrix}, \quad I \in \mathbf{R}^{d\times d},$$

where D is written as a square matrix for convenience; in reality D will be a banded matrix, where the bandwidth is independent of the mesh size $h = 1/\nu$.

2.1. **Summation by parts.** In the semidiscrete case we employ summation by parts instead of integration by parts. The basic idea is to use difference operators satisfying

$$(u, Dv)_h = u_\nu^T v_\nu - u_0^T v_0 - (Du, v)_h \tag{2.1}$$

with respect to a *weighted* scalar product

$$(u, v)_h = h\sum_{i,j=0}^{\nu} \sigma_{ij} u_i^T v_j .$$

It should be remarked that the usual Euclidean scalar product cannot be used. To prove the existence of summation by parts, it suffices to consider scalar products of the form

$$\Sigma = \begin{pmatrix} \Sigma^{(1)} & & \\ & I & \\ & & \Sigma^{(2)} \end{pmatrix}, \quad \Sigma^{(l)} \in \mathbf{R}^{(r_l+1)d\times(r_l+1)d}, \quad l = 1, 2, \tag{2.2}$$

where the blocks of Σ are given by $\Sigma_{ij} = \sigma_{ij} I$, $I \in \mathbf{R}^{d\times d}$; r_l and the elements of $\Sigma^{(l)}$, $l = 1, 2$, are independent of h. The following existence proof can be found in [5].

Proposition 2.1. *There exist scalar products* (2.2) *and difference operators* D *of accuracy* $2p-1$ *at the boundaries and* $2p$ *in the interior,* $p > 0$, *such that the summation-by-parts property* (2.1) *holds.*

Confining ourselves to the case where $\Sigma^{(1)}$ and $\Sigma^{(2)}$ are diagonal, we have the following existence theorem [4].

Proposition 2.2. *There exist diagonal scalar products* (2.2) *and difference operators* D *of accuracy* p *at the boundaries and* $2p$ *in the interior,* $1 \le p \le 4$, *such that the summation-by-parts property* (2.1) *holds.*

Remark 2.0. If one omits the requirement that the boundary stencils be at least accurate of order p for a given interior accuracy $2p$, it is possible to prove summation by parts for diagonal scalar products and difference operators D of arbitrary order of accuracy [8]. For a given boundary accuracy p, however, it may be necessary to resort to interior stencils of accuracy $q \gg 2p$, which may render these operators useless in practice.

The actual computation of the operators above is ill-conditioned, since it involves the solution of a rank-deficient problem. Using a symbolic language, one can solve for D exactly, the elements of which in general will depend on one or more parameters. Explicit examples can be found in [6]. For details on the algorithms we refer to [9]. The simplest example is furnished by

(2.3)
$$D = \frac{1}{h}\begin{pmatrix} -1 & 1 & & & \\ -0.5 & 0 & 0.5 & & \\ & \ddots & \ddots & \ddots & \\ & & -0.5 & 0 & 0.5 \\ & & & -1 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0.5 & & & & \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ & & & & 0.5 \end{pmatrix}.$$

Summation by parts can be generalized to several space dimensions if we restrict ourselves to diagonal norms. To simplify the notation, we consider only the two-dimensional case. A general proof is given in [6]. The grid function u_{ij} is partitioned as $u^T = (u_0^T \dots u_{\nu_2}^T)$, $u_j^T = (u_{0j}^T \dots u_{\nu_1 j}^T)$, $j = 0, \dots, \nu_2$. Define the weighted scalar product as

$$(u, v)_h = h \sum_{i=0}^{\nu_1} \sum_{j=0}^{\nu_2} \sigma_i \sigma_j u_{ij}^T v_{ij}, \tag{2.4}$$

where $h = h_1 h_2$ is the cell area. Let D_1 and D_2 denote the difference operators approximating $\partial/\partial x_1$ and $\partial/\partial x_2$. Define

$$(D_1 u)_{ij} = \frac{1}{h_1} \sum_{k=0}^{\nu_1} d_{ik} u_{kj}, \qquad (D_2 u)_{ij} = \frac{1}{h_2} \sum_{k=0}^{\nu_2} d_{jk} u_{ik}, \tag{2.5}$$

where it is assumed that the σ's and d's satisfy (2.1). Hence

$$(u, D_1 v)_h = h_2 \sum_{j=0}^{\nu_2} \sigma_j \left(\sum_{i=0}^{\nu_1} \sigma_i u_{ij}^T \sum_{k=0}^{\nu_1} d_{ik} v_{kj} \right),$$

and a similar expression holds for $(u, D_2 v)_h$. The parenthetical expression satisfies (2.1) for each j. We thus arrive at

Proposition 2.3. *Let the discrete difference operators D_1 and D_2 be defined by* (2.5). *Summation by parts then holds in both dimensions,*

$$(u, D_1 v)_h = h_2 \sum_{j=0}^{\nu_2} \sigma_j u_{\nu_1 j}^T v_{\nu_1 j} - h_2 \sum_{j=0}^{\nu_2} \sigma_j u_{0j}^T v_{0j} - (D_1 u, v)_h,$$
$$(u, D_2 v)_h = h_1 \sum_{i=0}^{\nu_1} \sigma_i u_{i\nu_2}^T v_{i\nu_2} - h_1 \sum_{i=0}^{\nu_1} \sigma_i u_{i0}^T v_{i0} - (D_2 u, v)_h,$$

where $(\cdot, \cdot)_h$ is defined by (2.4).

Remark 2.1. This is the discrete counterpart of the two-dimensional divergence theorem. With a general domain Ω we assume that there is a smooth map $\xi = \xi(x)$ taking Ω onto the unit cube where Proposition 2.3 can be applied. The assumption of such a map ξ is necessary in order for finite difference methods to apply to curvilinear domains. Consequently, integration by parts can always be replaced with summation by parts in the discrete case. It is presently unknown if it is possible to obtain the summation-by-parts property in more than one dimension using nondiagonal norms.

2.2. **Projections.** Suppose that the model equation of §1.1 were discretized as

$$\begin{aligned} v_t &= D^2 v + Dv \,, \\ v(0) &= f \,, \end{aligned} \tag{2.6}$$

where we have assumed homogeneous Neumann data for convenience; in Part II [7] it will be shown how to treat inhomogeneous boundary conditions. For every fixed h the problem above is a constant-coefficient ODE system with a unique analytic solution. Consequently, there is little hope that the discretized boundary conditions $v_0(t) = (Dv)_\nu(t) = 0$ are fulfilled, since they have not been accounted for so far.

Denote by $V \subset \mathrm{R}^{\nu+1}$ the vector space where $v_0(t) = (Dv)_\nu(t) = 0$, and let P be a projection of v onto V. Multiplying (2.6) by P yields

$$(Pv)_t = P\left(D^2 v + Dv\right) .$$

Any solution of (2.6) satisfying the boundary conditions must obey $v = Pv$, whence

$$v_t = P\left(D^2 v + Dv\right) . \tag{2.7}$$

Conversely, we have

Proposition 2.4. *Let $P \in R^{s\times s}$ be a given projection independent of t, and suppose that $v(t) \in R^s$ is a solution of the nonlinear ODE system*

$$\begin{aligned} v_t &= PR(t, v) + (I - P)g_t \,, \\ v(0) &= f \,, \end{aligned} \tag{2.8}$$

where f satisfies $f = Pf + (I - P)g(0)$. Then

$$v(t) = Pv(t) + (I - P)g(t) \,, \quad t > 0 .$$

Proof. Since P is independent of t, premultiplication of (2.8) gives ($P^2 = P$)

$$(Pv)_t = PR(t, v) .$$

Using this equality in (2.8) implies

$$v_t = (Pv + (I - P)g)_t .$$

Hence, by integration,

$$(I - P)(v(t) - g(t)) = (I - P)(f - g(0)) \,,$$

which proves the proposition. □

Remark 2.2. The function $g(t)$ represents the boundary data, and $(I - P) \cdot (v - g) = 0$ is the extension of $(I - P)v = 0$ to inhomogeneous boundary data. Proposition 2.4 thus tells us that any solution to (2.8) will satisfy the boundary conditions if the initial data do.

In general, P is not uniquely defined. Consider the vector space $V = \{v \in \mathrm{R}^{\nu+1} | v_0 = 0, v_\nu = v_{\nu-1}\}$. Then

$$P = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \\ & & & 1 \quad 0 \end{pmatrix}, \quad P = \begin{pmatrix} 0 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 0 \quad 1 \\ & & & \quad 1 \end{pmatrix}$$

both imply $Pv \in V$. To shed some light on how to choose P, we apply the energy method to (2.7):

$$\frac{d}{dt}||v||_h^2 = 2(v, P(D^2v + Dv))_h .$$

If P were selfadjoint with respect to $(\cdot, \cdot)_h$, then

$$\frac{d}{dt}||v||_h^2 = 2(Pv, D^2v + Dv)_h = 2(v, D^2v + Dv)_h ,$$

where the last equality follows from Proposition 2.4. The crucial condition to obtain this equality is expressed by

$$(u, Pv)_h = (Pu, v)_h , \tag{2.9}$$

which states that P is an orthogonal projection (using the *weighted* scalar product $(\cdot, \cdot)_h$).

Suppose that $u(x, t) \in \mathrm{R}^d$, $x \in \mathrm{R}^n$, is a solution to

$$\begin{aligned} u_t &= F(x, t, \partial)u , \quad && x \in \Omega , \\ L(x, \partial)u &= 0 , && x \in \Gamma , \end{aligned}$$

where ∂ denotes the n-dimensional gradient; Γ is the boundary of Ω. This system is discretized in space, possibly requiring a coordinate mapping onto the unit cube,

$$v_t = PG(t, D)v .$$

The projection P should be such that v fulfills

$$L^T v = 0 ,$$

where L now represents a discretization of the analytic boundary conditions. Let $V = \{v \in \mathrm{R}^m | L^T v = 0\}$. According to the preceding discussion, P is taken to be the orthogonal projection onto V (with respect to $(\cdot, \cdot)_h$). The boundary conditions can be written as

$$Q^T \Sigma v = 0 ,$$

where $Q = \Sigma^{-1} L$. Hence, the boundary conditions are fulfilled for all vectors v that are orthogonal to the column space of Q, the orthogonal projection onto which reads $Q(Q^T \Sigma Q)^{-1} Q^T \Sigma$. In case $\Sigma = I$, this is the standard projection.

The desired boundary projection is thus given by

$$P = I - Q(Q^T\Sigma Q)^{-1}Q^T\Sigma ,$$

or

$$P = I - \Sigma^{-1}L(L^T\Sigma^{-1}L)^{-1}L^T . \tag{2.10}$$

Remark 2.3. In order for the projection to be well defined, the inverse of $L^T\Sigma^{-1}L$ must exist, which follows if and only if L has full column rank. The latter will follow from assumptions on the analytic boundary conditions (consistency arguments).

Proposition 2.5. *Suppose that L has full column rank, and let P be defined by* (2.10). *Then*

(i) $P^2 = P$,
(ii) $\Sigma P = P^T\Sigma$,
(iii) $v = Pv \iff L^Tv = 0$.

Proof. All statements are immediate consequences of (2.10). □

Remark 2.4. The second statement of Proposition 2.5 is equivalent to (2.9).

2.3. **Some technical lemmas.** In this subsection we have gathered some technical results that will be needed in the subsequent presentation. Their proofs have been deferred to the Supplement. As seen in §1.1, it is necessary to have a Sobolev inequality. The following proposition shows that there is a discrete Sobolev inequality for the norms that we are interested in. We present it in a form suitable for proving *strict* stability.

Proposition 2.6. *Let $||\cdot||_h$ and D be defined by* (2.2) *and* (2.1), *respectively. Then*

$$|v|_\infty^2 \le \epsilon||Dv||_h^2 + \left(\epsilon^{-1} + 1 + \mathcal{O}(h)\right)||v||_h^2 ,$$

where $\epsilon > 0$.

Let

$$A = \begin{pmatrix} A_0 & & \\ & \ddots & \\ & & A_\nu \end{pmatrix} , \quad A_j = A(jh) , \quad j = 0, \dots , \nu , \quad h\nu = 1 , \tag{2.11}$$

denote the grid matrix representation of $A(x_j) \in \mathbf{R}^{d\times d}$, $x_j = hj \in [0, 1]$. Smoothness will be assumed as needed.

Lemma 2.1. *Let Σ and A be defined by* (2.2) *and* (2.11), *respectively. Then*

$$|(u, Av)_h - (A^Tu, v)_h| \le \mathcal{O}(h)||u||_h||v||_h .$$

Remark 2.5. According to Lemma 2.1, the transpose of A is an approximate adjoint with respect to $(\cdot, \cdot)_h$; the perturbation consists of lower-order terms.

The following assumption will be crucial when proving strict stability for hyperbolic systems.

Assumption 2.1. Let A and Σ be given by (2.11) and (2.2). Then one of the conditions below is assumed to hold:

(i) Σ is diagonal $\operatorname{diag}(\sigma_0 I \dots \sigma_\nu I)$;

(ii) The blocks of A satisfy $A_0 = \cdots = A_{r_1}$ and $A_{\nu - r_2} = \cdots = A_\nu$, where r_1 and r_2, $r_1 \geq 0$, $0 \leq r_2 < \nu - r_1$, are defined by equation (2.2).

Corollary 2.1. *If Assumption* 2.1 *holds, then* $(u, Av)_h = (A^T u, v)_h$.

Proof. In both cases, Σ and A commute. The corollary follows immediately. □

Remark 2.6. The latter criterion is satisfied if there is a $\delta > 0$ such that $A(x) = \text{const}$ for $0 \leq x < \delta$ and $1 - \delta < x \leq 1$, and if h is chosen such that $hr < \delta$, where $r = \max(r_1, r_2)$.

Lemma 2.2. *Let* Σ *and* A *be defined by* (2.2) *and* (2.11). *Then*

$$|(u, Av)_h| \leq |A|_\infty (1 + \mathcal{O}(h)) \|u\|_h \|v\|_h ,$$

where $|A|_\infty = \sup |A(x)|$.

Corollary 2.2. *If, in addition to the hypotheses of Lemma* 2.2, *Assumption* 2.1 *is fulfilled, then*

$$|(u, Av)_h| \leq |A|_\infty \|u\|_h \|v\|_h .$$

Remark 2.7. Lemma 2.2 states that the growth rate induced by low-order terms is the same (modulo $\mathcal{O}(h)$-terms) in the continuous and the semidiscrete case.

Denote by $[D, A]$ the commutator of D and A. It is well known that $(u, [D, A]v)_h \leq \|[D, A]\|_h \|u\|_h \|v\|_h$, where $\|[D, A]\|_h$ can be bounded independently of h. This result can be sharpened under certain circumstances.

Lemma 2.3. *Let* D *be a difference approximation satisfying the summation-by-parts rule* (2.1) *with respect to a weighted norm* (2.2), *and define* A *by* (2.11). *Suppose that Assumption* 2.1 *holds. If* A *is symmetric, then*

$$(u, [D, A]v)_h \leq \rho([D, A]) \|u\|_h \|v\|_h ,$$

where $\rho([D, A])$ *is the spectral radius of* $[D, A]$, *i.e.,* $\rho([D, A]) = \sup |\lambda_k|$, λ_k *an eigenvalue of* $[D, A]$.

3. One-dimensional problems

We shall successively consider hyperbolic, parabolic and mixed hyperbolic-parabolic systems. Variable-coefficient matrices will be allowed. To simplify the presentation, we shall only deal with the lower boundary $x = 0$, which is justified if we take the solution to have compact support. In general, the upper boundary $x = 1$ is treated in a fashion similar to the procedure at the lower boundary.

3.1. **Hyperbolic systems.** Consider the hyperbolic system

(3.1)

$$\begin{aligned} &u_t = \Lambda u_x + Bu + F , \quad && x \in (0, 1) , \\ &u(x, 0) = f(x) , \\ &u_-(0, t) = L u_+(0, t) , \quad && L \in \mathbf{R}^{d_1 \times d_2} , \end{aligned} \qquad \Lambda(x, t) = \begin{pmatrix} \Lambda_-(x, t) & \\ & \Lambda_+(x, t) \end{pmatrix} ,$$

where $u \in \mathbf{R}^d$, $d_1 + d_2 = d$; Λ_-, Λ_+ is the partitioning of Λ into negative and positive eigenvalues. It is assumed that the elements of the diagonal matrix Λ never change sign at the boundaries $x = 0$ and $x = 1$, and that there is a constant $\gamma > 0$ such that $\Lambda_-(j, t) \leq -\gamma$ and $\Lambda_+(j, t) \geq \gamma$, $j = 0, 1$. This implies that the rank of L is constant. Furthermore, L is assumed to be "small".

The discrete boundary conditions are written as $L^T v = 0$, where

$$L^T = (L_0^T \quad 0 \quad \dots \quad 0) \in \mathbf{R}^{d_1 \times (\nu+1)d}. \tag{3.2}$$

Here, $L_0^T = (I \quad -L) \in \mathbf{R}^{d_1 \times d}$, the latter L being the analytic boundary operator. It follows immediately that $\operatorname{rank}(L) = \operatorname{rank}(I) = d_1$. The hypothesis of Proposition 2.5 is thus satisfied, and we have the semidiscrete system

$$\begin{aligned} &v_t = P(\Lambda D v + Bv + F), \\ &v(0) = f, \end{aligned} \qquad \Lambda = \begin{pmatrix} \Lambda(0, t) & & \\ & \ddots & \\ & & \Lambda(1, t) \end{pmatrix}. \tag{3.3}$$

Proposition 3.1. *Let* $(\cdot, \cdot)_h$ *be given by* (2.2) *and suppose that* D *satisfies the conclusion of Proposition* 2.1. *If* P *is defined by* (2.10) *and* (3.2), *then the solution of* (3.3) *satisfies an energy estimate*

$$\|v(t)\|_h^2 + \int_0^t \left(|v_0(\tau)|^2 + |v_\nu(\tau)|^2\right) d\tau \leq K e^{(\alpha' + \mathscr{O}(h))t} \left(\|f\|_h^2 + \int_0^t \|F(\tau)\|_h^2 d\tau\right).$$

Proof. The energy method yields (using Propositions 2.5, 2.4)

$$\begin{aligned} \frac{d}{dt}\|v\|_h^2 &= 2(v, v_t)_h = 2(v, P(\Lambda D v + Bv + F))_h \\ &= 2(v, \Lambda D v)_h + 2(v, Bv)_h + 2(v, F)_h. \end{aligned}$$

Summation by parts implies ($v_\nu = 0$)

$$(v, \Lambda D v)_h = -v_0^T \Lambda_0 v_0 - (Dv, \Lambda v)_h - (v, [D, \Lambda] v)_h.$$

Hence, by Lemma 2.1,

$$(v, \Lambda D v)_h \leq -\frac{1}{2} v_0^T \Lambda_0 v_0 + \frac{1}{2}\left(K_0 \|v\|_h \|hDv\|_h + \|[D, \Lambda]\|_h \|v\|_h^2\right),$$

where hD is a *bounded* operator, i.e.,

$$(v, \Lambda D v)_h \leq -\frac{1}{2} v_0^T \Lambda_0 v_0 + \frac{1}{2}(K_1 + \|[D, \Lambda]\|_h)\|v\|_h^2.$$

Now, according to Propositions 2.4, 2.5 we have $L^T v = 0$, which is equivalent to $v_- = L v_+$ (the latter L denoting the analytic boundary operator). Thus,

$$v_0^T \Lambda_0 v_0 = v_-^T \Lambda_- v_- + v_+^T \Lambda_+ v_+ = v_+^T (\Lambda_+ + L^T \Lambda_- L) v_+ \geq \frac{\gamma}{2}|v_0|^2,$$

where the last inequality follows from the boundary conditions and the assumptions on L and Λ. Note that the analytic problem would result in exactly the

same inequality. Hence,

$$(v, \Lambda D v)_h \leq -\frac{\gamma}{4}|v_0|^2 + \frac{1}{2}(K_1 + |||[D, \Lambda]|||_h)\,||v||_h^2 .$$

Lemma 2.2 shows that

$$(v, Bv)_h \leq (|B|_\infty + \mathcal{O}(h))\,||v||_h^2 .$$

Consequently,

$$\frac{d}{dt}||v||_h^2 + |v_0|^2 \leq \frac{1}{\min(1, \gamma/2)}\left((|||[D, \Lambda]|||_h + 2|B|_\infty + 1 + K_1 + \mathcal{O}(h))||v||_h^2 + ||F||_h^2\right).$$

Integration with respect to t proves the proposition with $K = \max(1, 2/\gamma)$. □

Definition 3.1. A semidiscrete approximation to the initial-boundary value problem $u_t = F(x, t, \partial)u$ is said to be strictly stable, if the semidiscrete solution satisfies an energy estimate that is exponentially bounded by $\exp(\alpha' t)$, $\alpha' = \alpha + \mathcal{O}(h)$, where α is the exponential growth factor of the analytic estimate.

Remark 3.1. If Λ (or $(\cdot, \cdot)_h$) satisfies Assumption 2.1, it follows that $K_1 = 0$. Also, by Lemma 2.3, $|||[D, \Lambda]|||_h = \rho([D, \Lambda])$. Equation (3.3) would thus be strictly stable if $\rho([D, \Lambda]) \leq |\Lambda'|_\infty$. In particular, (3.3) is strictly stable if $\Lambda(x) = \text{const}$, since this implies $[D, \Lambda] = 0$. We also point out that the proportionality constant K is completely independent of the discretization. In case the estimate of the boundary integral is not needed, one may take $K = 1$. For variable-coefficient problems we have the following result.

Corollary 3.1. *Let D and $(\cdot, \cdot)_h$ be given by* (2.3). *Then* (3.3) *is strictly stable.*

Proof. According to the preceding remark, the corollary follows if we can show that $\rho([D, \Lambda]) \leq |\Lambda'|_\infty$. But

$$[D, \Lambda] = \frac{1}{h}\begin{pmatrix} 0 & \Lambda_1 - \Lambda_0 & & & \\ 0.5(\Lambda_1 - \Lambda_0) & 0 & 0.5(\Lambda_2 - \Lambda_1) & & \\ & \ddots & \ddots & \ddots & \\ & & 0.5(\Lambda_{\nu-1} - \Lambda_{\nu-2}) & 0 & 0.5(\Lambda_\nu - \Lambda_{\nu-1}) \\ & & & \Lambda_\nu - \Lambda_{\nu-1} & 0 \end{pmatrix}.$$

If $\Lambda(x)$ is assumed C^1, the mean value theorem gives $\Lambda_i - \Lambda_j = \Lambda'(\xi_{ij})(i - j)h$ for some $\xi_{ij} \in (ih, jh)$. The corollary thus follows from the Gershgorin disk theorem. □

3.2. **Parabolic systems.** We consider the parabolic system

(3.4)
$$\begin{aligned} &u_t = Au_{xx} + Bu_x + Cu + F, \quad x \in (0, 1), \\ &u(x, 0) = f(x), \\ &L_0 u(0, t) + L_1 u_x(0, t) = 0, \end{aligned} \qquad L_0 = \begin{pmatrix} L_0^I \\ L_0^{II} \end{pmatrix}, \; L_1 = \begin{pmatrix} L_1^I \\ 0, \end{pmatrix},$$

where $L_0^I, L_1^I \in \mathrm{R}^{d_1 \times d}$, $L_0^{II} \in \mathrm{R}^{d_2 \times d}$, $d_1 + d_2 = d$; $\operatorname{rank}(L_1^I) = d_1$, $\operatorname{rank}(L_0^{II}) = d_2$; A, B, C, and F depend smoothly on x and t. It is assumed that the system is strongly parabolic, i.e., $A(x, t) + A(x, t)^T \geq 2\delta I$.

The following lemma, a proof of which can be found in [3, Lemma 7. 2. 1, p. 215], will be crucial when proving an energy estimate for the solution of (3.4) and its semidiscrete counterpart.

Lemma 3.1. *Let* $A \in R^{d\times d}$ *be arbitrary and let* $L_0, L_1 \in R^{d\times d}$ *be of the form* (3.4). *The following conditions are equivalent*:

(i) *There exists a constant* $c > 0$ *such that*

$$|u^T A u_x| \le c|u|^2$$

for all u, $u_x \in R^d$ *that satisfy*

$$L_0 u + L_1 u_x = 0.$$

(ii) *If* a, $b \in R^d$ *are vectors such that*

$$L_1^I b = 0, \quad L_0^{II} a = 0,$$

then

$$a^T A b = 0.$$

Assumption 3.1. Given the boundary matrices L_0, L_1, the matrix $A(x, t)$ is supposed to be such that the second condition of Lemma 3.1 holds for $x = 0, 1$.

Remark 3.2. Except for Dirichlet and Neumann conditions, Assumption 3.1 imposes severe restrictions on A. Lemma 3.1 states that the assumption above is necessary in order to obtain an energy estimate. The computations that follow will show how the second condition, which holds by assumption, implies the first.

Before deriving the energy estimates, one more lemma is needed [3, Lemma 7. 2. 3, p. 217].

Lemma 3.2. *Suppose that Assumption* 3.1 *holds and that* $A(x, t) + A(x, t)^T \ge 2\delta I$. *Then the* $d \times d$ *matrix*

$$\begin{pmatrix} L_1^I \\ L_0^{II} \end{pmatrix}$$

is nonsingular.

As usual, the boundary conditions are written as $L^T v = 0$, where

$$(3.5)\quad L^T = \left(L_0 + \frac{d_{00}}{h} L_1 \quad \frac{d_{01}}{h} L_1 \quad \cdots \quad \frac{d_{0r}}{h} L_1 \quad 0 \quad \cdots 0 \right) \in \mathrm{R}^{d\times(\nu+1)d},$$

and where d_{0j}/h are the nonzero elements of the first row of D, which is a difference operator satisfying the conclusion of Proposition 2.1 or 2.2. We have

$$L_0 + \frac{d_{00}}{h} L_1 = \begin{pmatrix} (d_{00}/h) I^I & \\ & I^{II} \end{pmatrix} \left[\begin{pmatrix} L_1^I \\ L_0^{II} \end{pmatrix} + \frac{h}{d_{00}} \begin{pmatrix} L_0^I \\ 0 \end{pmatrix} \right].$$

Thus, Lemma 3.2 implies that $L_0 + (d_{00}/h)L_1$ is nonsingular for $h > 0$ sufficiently small. From (3.5) it follows immediately that $\operatorname{rank}(L) = d$. According to Proposition 2.5, the corresponding projection operator is well defined, and

we obtain

$$(3.6)\quad \begin{matrix} v_t = P(AD^2v + BDv + Cv + F)\,, \\ v(0) = f\,, \end{matrix} \qquad A = \begin{pmatrix} A(0,t) & & \\ & \ddots & \\ & & A(1,t) \end{pmatrix},$$

with similar expressions for B, C, F.

Proposition 3.2. *Let* $(\cdot,\cdot)_h$ *be given by* (2.2) *and suppose that* D *satisfies the conclusion of Proposition* 2.1. *If* P *is defined by* (2.10) *and* (3.5), *then the solution of* (3.6) *satisfies an energy estimate*

$$\|v(t)\|_h^2 + \int_0^t \left(|v_0(\tau)|^2 + |v_\nu(\tau)|^2\right) d\tau \le e^{(\alpha' + \mathcal{O}(h))t} \left(\|f\|_h^2 + \int_0^t \|F(\tau)\|_h^2 d\tau\right).$$

Proof. By Propositions 2.5, 2.4, 2.1 we have ($A_0 = A(0,t)$)

$$(v\,, PAD^2v)_h = (v\,, AD^2v)_h = -v_0^T A_0 (Dv)_0 - (Dv\,, ADv)_h - (v\,, [D\,, A]Dv)_h\,,$$

where we have assumed homogeneous Dirichlet conditions at the upper boundary for convenience. From Proposition 2.5 it follows that

$$(3.7)\qquad L_0 v_0 + L_1 \frac{1}{h} \sum_{j=0}^{r} d_{0j} v_j = L_0 v_0 + L_1 (Dv)_0 = 0\,.$$

Partition $v_j = v_j' + v_j''$, $v_j' \in \ker L_1^I$, $v_j'' \in (\ker L_1^I)^\perp$. Equation (3.7) implies $L_0^{II} v_0 = 0$ and by construction $L_1^I (Dv')_0 = 0$. Hence, according to Assumption 3.1,

$$-v_0^T A_0 (Dv)_0 = -v_0^T A_0 (Dv'')_0\,.$$

Equation (3.7) can be rewritten as

$$\begin{pmatrix} L_1^I \\ 0 \end{pmatrix} (Dv'')_0 = -L_0 v_0\,.$$

Since $(Dv'')_0 \in (\ker L_1^I)^\perp$, we get

$$\tilde{L}_1 (Dv'')_0 = -L_0 v_0\,, \qquad \tilde{L}_1 = \begin{pmatrix} L_1^I \\ s_1^T \\ \vdots \\ s_{d_2}^T \end{pmatrix},$$

where $\{s_j\}$ is a basis in $\ker L_1^I$. Thus, $\tilde{L}_1$ is nonsingular, and one obtains

$$-v_0^T A_0 (Dv)_0 = v_0^T A_0 \tilde{L}_1^{-1} L_0 v_0 \le \gamma |v_0|^2\,, \qquad \gamma = |A_0 \tilde{L}_1^{-1} L_0|_\infty\,.$$

This is exactly the same expression as one would get in the analytic case ($A_0 = A(0,t)$). Thus,

$$(3.8)\qquad (v\,, PAD^2v)_h \le \gamma |v_0|^2 - \delta \|Dv\|_h^2 + \|[D\,, A]\|_h \|v\|_h \|Dv\|_h\,.$$

Furthermore,

$$(3.9)\qquad \begin{aligned} (v\,, PBDv)_h &\le (|B|_\infty + \mathcal{O}(h)) \|v\|_h \|Dv\|_h\,, \\ (v\,, PCv)_h &\le (|C|_\infty + \mathcal{O}(h)) \|v\|_h^2\,. \end{aligned}$$

Finally, Proposition 2.6 and the algebraic inequality yield

$$\frac{d}{dt}||v||_h^2 + |v_0|^2 \leq (\alpha' + \mathscr{O}(h))\,||v||_h^2 + ||F||_h^2 .$$

Integration with respect to time proves the proposition. □

Remark 3.3. All coefficients, except $||[D, A]||_h$, appearing in (3.8) and (3.9) are identical (modulo $\mathscr{O}(h)$-terms) to those of the analytic estimate. Since the discrete Sobolev inequality 2.6 introduces the same growth rate as the analytic Sobolev inequality, it follows that (3.6) is strictly stable if we have the estimate $||[D, A]||_h \leq |A'|_\infty$, which is true if $A(x) = \text{const}$.

For variable coefficients one can prove

Corollary 3.2. *Let* D *and* $(\cdot, \cdot)_h$ *be given by* (2.3). *Then* (3.6) *is strictly stable if* A *is symmetric.*

Proof. Same as for Corollary 3.1. □

3.3. **Hyperbolic-parabolic systems.** Consider the mixed hyperbolic-parabolic system

$$\begin{aligned} &u_t = Au_{xx} + B_{11}u_x + B_{12}v_x + C_{11}u + C_{12}v + F, \quad x \in (0, 1), \\ &v_t = \Lambda v_x + B_{21}u_x + C_{21}u + C_{22}v + G, \\ &L_1 u_x(0, t) + L_0 u(0, t) + M_0 v(0, t) = 0, \\ &v_-(0, t) = S_0 v_+(0, t) + R_0 u(0, t), \\ &u(x, 0) = f(x), \\ &v(x, 0) = \phi(x), \end{aligned} \tag{3.10}$$

where $u \in \mathbf{R}^{d_1}$, $v \in \mathbf{R}^{d_2}$, $v_- \in \mathbf{R}^{d_2'}$, $v_+ \in \mathbf{R}^{d_2''}$, and

$$M_0 = \begin{pmatrix} M_0^I \\ 0 \end{pmatrix}.$$

As usual, we assume $u = v \equiv 0$ in a neighborhood of $x = 1$ for convenience; L_0, L_1 are as in §3.2, and S_0 satisfies the hypotheses of the boundary operator in §3.1. The coefficient matrices and the forcing functions of the differential equations may depend on x and t.

The discretized boundary conditions are written as $L^T w = 0$, where $L^T \in \mathbf{R}^{d' \times (\nu+1)d}$, $d = d_1 + d_2$, $d' = d_1 + d_2'$, is given by

$$\begin{aligned} &\left(\begin{pmatrix} L_0 + \dfrac{d_{00}}{h}L_1 & M_0 \\ -R_0 & (I \quad -S_0) \end{pmatrix} \begin{pmatrix} \dfrac{d_{01}}{h}L_1 & 0 \\ 0 & 0 \end{pmatrix} \cdots \right. \\ &\qquad \left. \begin{pmatrix} \dfrac{d_{0r}}{h}L_1 & 0 \\ 0 & 0 \end{pmatrix} 0 \ \ldots \ 0 \right). \end{aligned} \tag{3.11}$$

We want to show that L^T has full rank. The first block of L^T can be rewritten as

$$\begin{pmatrix} D(h) & 0 \\ 0 & I \end{pmatrix} \left[\begin{pmatrix} \tilde{L} & 0 & 0 \\ -R_0 & I & 0 \end{pmatrix} + \begin{pmatrix} (h/d_{00})\hat{L} & (h/d_{00})M_{00} & (h/d_{00})M_{01} \\ 0 & 0 & -S_0 \end{pmatrix} \right],$$

where

$$D(h) = \begin{pmatrix} (d_{00}/h)I^I & \\ & I^{II} \end{pmatrix}, \quad \tilde{L} = \begin{pmatrix} L_1^I \\ L_0^{II} \end{pmatrix}, \quad \hat{L} = \begin{pmatrix} L_0^I \\ 0 \end{pmatrix}.$$

Since $\tilde{L}$ is invertible, it follows that

$$\begin{pmatrix} \tilde{L} & 0 \\ -R_0 & I \end{pmatrix} + \begin{pmatrix} (h/d_{00})\hat{L} & (h/d_{00})M_{00} \\ 0 & 0 \end{pmatrix}$$

is invertible, i.e., has full rank for $h > 0$ sufficiently small. The expression enclosed by the square brackets thus has linearly independent rows, which in turn implies that the first block of L^T has full rank. Hence, L has full rank, and the corresponding projection is well defined.

The semidiscrete system is formulated as

(3.12)

$$\begin{aligned} &w_t = P\left(\tilde{A}\tilde{D}^2 w + \tilde{\Lambda}\tilde{D}w + \tilde{C}w + \tilde{F}\right), \quad w_j = \begin{pmatrix} u_j \\ v_j \end{pmatrix}, \quad j = 0, \dots, \nu, \\ &w(0) = \psi, \end{aligned}$$

where

$$\tilde{A} = \operatorname{diag}\left[\begin{pmatrix} A(jh, t) & 0 \\ 0 & 0 \end{pmatrix}\right],$$

$$\tilde{\Lambda} = \operatorname{diag}\left[\begin{pmatrix} B_{11}(jh, t) & B_{12}(jh, t) \\ B_{21}(jh, t) & \Lambda(jh, t) \end{pmatrix}\right], \quad j = 0, \dots, \nu.$$

$$\tilde{C} = \operatorname{diag}\left[\begin{pmatrix} C_{11}(jh, t) & C_{12}(jh, t) \\ C_{21}(jh, t) & C_{22}(jh, t) \end{pmatrix}\right].$$

The forcing function $\tilde{F}$ and the initial data ψ are defined analogously.

Proposition 3.3. *Let* $(\cdot, \cdot)_h$ *be given by* (2.2) *and suppose that* $\tilde{D}$ *satisfies the conclusion of Proposition* 2.1. *If* P *is defined by* (2.10) *and* (3.11), *then the solution of* (3.12) *satisfies an energy estimate*

$$\begin{aligned} &\|u(t)\|_h^2 + \|v(t)\|_h^2 + \sum_{j=0 \text{ and } \nu} \int_0^t \left(|u_j(\tau)|^2 + |v_j(\tau)|^2\right) d\tau \\ &\quad \le K e^{(\alpha' + \mathcal{O}(h))t} \left(\|f\|_h^2 + \|\phi\|_h^2 + \int_0^t \left(\|F(\tau)\|_h^2 + \|G(\tau)\|_h^2\right) d\tau\right). \end{aligned}$$

Proof. The energy method applied to (3.12) yields

$$\begin{aligned} \frac{d}{dt}\|w\|_h^2 &= 2(w, P(\tilde{A}\tilde{D}^2 w + \tilde{\Lambda}\tilde{D}w + \tilde{C}w + \tilde{F}))_h \\ &= 2(w, (\tilde{A}\tilde{D}^2 w + \tilde{\Lambda}\tilde{D}w + \tilde{C}w + \tilde{F}))_h. \end{aligned}$$

Now

$$\begin{aligned} (w, \tilde{A}\tilde{D}^2 w)_h &= h \sum_{i,j=0}^{\nu} \sigma_{ij}(u_i^T v_i^T) \begin{pmatrix} A_j & 0 \\ 0 & 0 \end{pmatrix} \frac{1}{h^2} \sum_{k,l=0}^{\nu} d_{jk} d_{kl} \begin{pmatrix} u_l \\ v_l \end{pmatrix} \\ &= h \sum_{i,j=0}^{\nu} \sigma_{ij} u_i^T A_j \frac{1}{h^2} \sum_{k,l=0}^{\nu} d_{jk} d_{kl} u_l = (u, AD^2 u)_h, \end{aligned}$$

where D is the difference operator of (3.6). The remaining terms are handled in a similar manner. One has

(i) $(w, \tilde{A}\tilde{D}^2 w)_h = (u, AD^2 u)_h$,
(ii) $(w, \tilde{\Lambda}\tilde{D}w)_h = (u, B_{11}Du)_h + (u, B_{12}Dv)_h + (v, B_{21}Du)_h + (v, \Lambda Dv)_h$,
(iii) $(w, \tilde{C}w)_h = (u, C_{11}u)_h + (u, C_{12}v)_h + (v, C_{21}u)_h + (v, C_{22}v)_h$,
(iv) $(w, \tilde{F})_h = (u, F)_h + (v, G)_h$.

For convenience we use the same symbol D to denote the difference operators acting on u and v. As far as the energy estimate is concerned, the hyperbolic-parabolic system has now been reduced to the previously treated hyperbolic and parabolic systems.

Items (iii) and (iv) consist only of lower-order terms, and can be estimated using Lemma 2.2. Thus, the coefficients of the estimates are identical to the corresponding analytic estimate (modulo $\mathcal{O}(h)$-terms). In item (ii) the potentially "dangerous" terms are those containing Dv. Using exactly the same technique as in the proof of Proposition 3.1, we get

$$(v, \Lambda Dv)_h \le -\frac{\gamma}{4}|v_0|^2 + |S_0^T \Lambda_- R_0|_\infty |v_0||u_0| + \frac{1}{2}|R_0^T \Lambda_- R_0|_\infty |u_0|^2 + \frac{1}{2}(K_1 + |||[D, \Lambda]|||_h)\,||v||_h^2,$$

i.e., by means of the algebraic inequality

$$(v, \Lambda Dv)_h \le -\frac{\gamma}{6}|v_0|^2 + \gamma_1|u_0|^2 + \frac{1}{2}(K_1 + |||[D, \Lambda]|||_h)\,||v||_h^2.$$

Furthermore,

$$(u, B_{12}Dv)_h \le \frac{1}{2}|B_{12}|_\infty \left(\epsilon_1|v_0|^2 + \epsilon_1^{-1}|u_0|^2\right) + |||[D, B_{12}]|||_h ||u||_h ||v||_h - (Du, B_{12}v)_h.$$

Finally, in item (i) the term $(u, AD^2u)_h$ is treated as in the proof of Proposition 3.2, the only difference being that

$$-u_0^T A_0 (Du)_0 = u_0^T A_0 \tilde{L}_1^{-1} L_0 u_0 + u_0^T A_0 \tilde{L}_1^{-1} M_0 v_0 \le \gamma_2 \left[\epsilon_2|v_0|^2 + \left(\epsilon_2^{-1} + 1\right)|u_0|^2\right].$$

We point out that the coefficients of the boundary terms in the inequalities above are identical to those of the analytic estimate. Choosing ϵ_1 and ϵ_2 sufficiently small, we thus arrive at

$$\frac{d}{dt}||w||_h^2 + \frac{\gamma}{4}\left(|u_0|^2 + |v_0|^2\right) \le (\alpha' + \mathcal{O}(h))\,||w||_h^2 + ||F||_h^2 + ||G||_h^2,$$

where we have used $||w||_h^2 = ||u||_h^2 + ||v||_h^2$; in the right member we have used Proposition 2.6 and the algebraic inequality to eliminate $|u_0|^2$ and $||Du||_h$. Integration proves the proposition with $K = \max(1, 4/\gamma)$. □

Remark 3.4. In case no estimate of $|v_0|^2$ is needed, one may take $K = 1$. Also, only the coefficients $|||[D, A]|||_h$, $|||[D, \Lambda]|||_h$, $|||[D, B_{12}]|||_h$ and K_1 will be larger than their analytic counterparts. If either of the conditions of Assumption 2.1 is

met, then $K_1 = 0$ and the operator norms can be replaced by the corresponding spectral radii (cf. Lemma 2.3). In particular, if A, Λ, B_{12} are constant, then (3.12) is strictly stable.

As before, for variable coefficients we have

Corollary 3.3. *Let $\tilde{D}$ and $(\cdot, \cdot)_h$ be given by* (2.3). *Then* (3.12) *is strictly stable if A and B_{12} are symmetric.*

Proof. Same as for Corollary 3.1. □

3.4. **Strict stability.** So far we have obtained strict stability under special circumstances, such as constant-coefficient problems or second-order methods. The crux of the matter lies in estimating the commutator $[D, A]$. Only in the previous cases were we able to prove that $\|[D, A]\|_h \leq |A'|_\infty$. In fact, numerical experiments show that $\|[D, A]\|_h \geq \rho([D, A]) = K|A'|_\infty$, $K > 1$, for high-order methods. Typical values for D's corresponding to diagonal norms are $K = 1.67$, $K = 2.55$, and $K = 35.8$, where the operator accuracy increases from three to five. One would still obtain $K > 1$ even if one considered only the interior operator. This indicates that the commutator should be avoided, which can be achieved if the analytic problem is reformulated.

The hyperbolic system (3.1) can be rewritten in skew-symmetric form as

$$\begin{aligned} &u_t = \frac{1}{2}(\Lambda u)_x + \frac{1}{2}\Lambda u_x + \left(B - \frac{1}{2}\Lambda'\right)u + F\,, \quad x \in (0, 1)\,, \\ &u(x, 0) = f(x)\,, \\ &u_-(0, t) = Lu_+(0, t)\,, \qquad\qquad L \in \mathbf{R}^{d_1 \times d_2}\,. \end{aligned}$$

The corresponding semidiscrete system becomes

$$\begin{aligned} &v_t = P\left(\frac{1}{2}D\Lambda v + \frac{1}{2}\Lambda D v + \left(B - \frac{1}{2}\Lambda'\right)v + F\right)\,, \\ &v(0) = f\,. \end{aligned} \tag{3.13}$$

Proposition 3.4. *Let $(\cdot, \cdot)_h$ be given by* (2.2) *and suppose that D satisfies the conclusion of Proposition* 2.1. *Define P by* (2.10) *and* (3.2). *If either Λ or Σ fulfills Assumption* 2.1, *then* (3.13) *is strictly stable.*

Proof. The energy method implies

$$\begin{aligned} \frac{d}{dt}\|v\|_h^2 \;=\; &-v_0^T\Lambda_0 v_0 - (Dv\,,\,\Lambda v)_h + (v\,,\,\Lambda Dv)_h - (v\,,\,\Lambda' v)_h \\ &+2(v\,,\,Bv)_h + 2(v\,,\,F)_h\,. \end{aligned}$$

The boundary terms are treated exactly as in the proof of Proposition 3.1. Because of Corollary 2.1 we have $(Dv\,,\,\Lambda v)_h = (v\,,\,\Lambda Dv)_h$. Thus, by Lemma 2.2,

$$\frac{d}{dt}\|v\|_h^2 + \frac{\gamma}{2}|v_0|^2 \leq (|\Lambda'|_\infty + 2|B|_\infty + 1 + \mathcal{O}(h))\|v\|_h^2 + \|F\|_h^2\,,$$

which is identical (neglecting $\mathcal{O}(h)$-terms) to the analytic estimate. □

Remark 3.5. If Σ is diagonal, then the $\mathcal{O}(h)$-terms vanish identically (Corollary 2.2).

The parabolic system (3.4) is altered in a slightly different manner. The modified system reads

$$\begin{aligned} &u_t = (Au_x)_x + (B - A')u_x + Cu + F\,, \quad x \in (0,1)\,,\\ &u(x,0) = f(x)\,,\\ &L_0u(0,t) + L_1u_x(0,t) = 0\,, \end{aligned}$$

which is discretized as

$$\begin{aligned} &v_t = P\,(DADv + (B - A')\,Dv + Cv + F)\,,\\ &v(0) = f\,. \end{aligned} \tag{3.14}$$

Proposition 3.5. *Let* $(\cdot,\cdot)_h$ *be given by* (2.2) *and suppose that* D *satisfies the conclusion of Proposition* 2.1. *If* P *is defined by* (2.10) *and* (3.5), *then* (3.14) *is strictly stable.*

Proof. Left to the reader. □

Finally, the mixed hyperbolic-parabolic system is reformulated as

$$\begin{aligned} u_t &= (Au_x)_x + (B_{11} - A')u_x + (B_{12}v)_x + C_{11}u + (C_{12} - B_{12}')v + F\,,\\ v_t &= \frac{1}{2}(\Lambda v)_x + \frac{1}{2}\Lambda v_x + B_{21}u_x + C_{21}u + (C_{22} - \frac{1}{2}\Lambda')v + G\,, \qquad x \in (0,1)\,, \end{aligned}$$

where the initial data and the boundary conditions are identical to those of (3.10). In semidiscrete form we have

$$\begin{aligned} &w_t = P\left(\tilde{D}\tilde{A}\tilde{D}w + \tilde{D}\tilde{\Lambda}w + \tilde{B}\tilde{D}w + (\tilde{C} - \tilde{\Lambda}')w + \tilde{F}\right)\,,\\ &w(0) = \psi\,, \end{aligned} \tag{3.15}$$

where $\tilde{\Lambda}$ and $\tilde{B}$ are block-diagonal matrices of the form

$$\tilde{\Lambda} = \operatorname{diag}\left[\begin{pmatrix} 0 & B_{12}(jh,t) \\ 0 & \Lambda(jh,t)/2 \end{pmatrix}\right]\,,$$

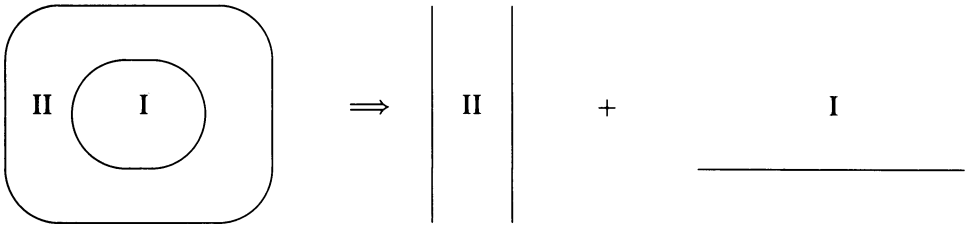
$$\tilde{B} = \operatorname{diag}\left[\begin{pmatrix} B_{11}(jh,t) - A'(jh,t) & 0 \\ B_{21}(jh,t) & \Lambda(jh,t)/2 \end{pmatrix}\right]\,, \qquad j = 0, \dots, \nu\,.$$

Proposition 3.6. *Let* $(\cdot,\cdot)_h$ *be given by* (2.2) *and suppose that* $\tilde{D}$ *satisfies the conclusion of Proposition* 2.1. *Define* P *by* (2.10) *and* (3.11). *If either* Λ *or* Σ *fulfills Assumption* 2.1, *then* (3.15) *is strictly stable.*

Proof. Left to the reader. □

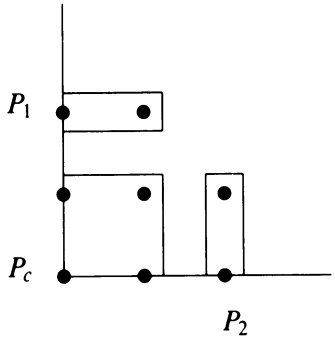
4. Two-dimensional problems

The results of §3 will now be generalized to two space dimensions. If the boundary is smooth, the original problem can be decomposed into two problems via a partition of unity, one of which is a Cauchy problem. The second problem is an initial-boundary value problem that is periodic in one space dimension, see figure below.



Consequently, summation by parts is needed only in one dimension, and the generalization of Propositions 3.1, 3.2, 3.3 to two dimensions follows immediately. For details on the decomposition we refer to [3, sec. 8. 1. 4 and sec. 8. 2. 6]. The situation is different if the boundary is nonsmooth, which is the case in the presence of corners. As mentioned at the end of §2.1, it is not known how to extend norms of type (2.2) so as to obtain summation by parts in several space dimensions. We thus limit ourselves to diagonal norms, in which case we have Proposition 2.3.

All boundary conditions considered so far are *local.* In case of characteristic and Dirichlet conditions no new difficulties are presented in two dimensions, because each boundary point can be treated individually. Boundary conditions involving derivatives increase the complexity significantly. Therefore, we shall only allow normal derivatives in the boundary operator. This is no serious restriction from the application point of view. Thus, away from the corners these boundary conditions are locally one-dimensional. For each such boundary point we obtain a projection operator of the previous section. In particular, these operators commute since they affect disjoint sets of grid points. At corners the situation is more complicated, because there are two different normal derivatives, which implies that the corresponding projection no longer is locally one-dimensional.



P_1, P_2, P_c commute

Throughout this section we shall focus our interest on the origin, and assume that the solutions are supported only in a neighborhood of $(0, 0)$. The remaining boundary conditions will be accounted for by applying the projection

operators corresponding to the boundary point in question. Since these operators commute, the resulting product is the uniquely defined boundary projection. The domain of definition is taken to be $\Omega = (0, 1) \times (0, 1)$ with boundary Γ. It will be shown in Part II how to extend the results to curvilinear domains. In order to simplify the presentation, all lower-order terms will be omitted.

4.1. **Symmetric hyperbolic systems.** Consider

$$\begin{aligned} &u_t = \sum_{i=1}^{2} A_i u_{x_i} + F , && x \in \Omega = (0, 1) \times (0, 1) , \quad u \in \mathbf{R}^d , \\ &u(x, 0) = f(x) , && x = (x_1, x_2) , \\ &\varphi_I(x, t) = S(x)\varphi_{II}(x, t) , && x \in \Gamma , \end{aligned} \tag{4.1}$$

where φ_I, φ_{II} denote the locally ingoing and outgoing characteristic variables; $A_i = A_i(x, t)$, $i = 1, 2$, are symmetric and $S(x)$ is assumed to be "small". It should be noted that $\varphi_I \in \mathbf{R}^{d_1(x)}$, $\varphi_{II} \in \mathbf{R}^{d_2(x)}$, where $d_1(x) + d_2(x) = d$, $x \in \Gamma$. The matrix

$$A(x, t) \equiv \sum_{i=1}^{2} n_i(x) A_i(x, t) \tag{4.2}$$

can be diagonalized for every $x \in \Gamma$; $n(x) = (n_1(x), n_2(x))$ is the outward unit normal of Γ. Hence,

$$\Lambda(x, t) = Q^T(x) A(x, t) Q(x) , \quad x \in \Gamma . \tag{4.3}$$

Note that we allow the eigenvalues to be time-dependent, whereas the eigenvectors are assumed to be time-independent to make the resulting projection operator independent of time. It will be shown in a future paper how this technicality can be overcome. The characteristic variables are only needed at the boundary, and they are defined as $\varphi(x, t) = Q^T(x) u(x, t)$. It will be assumed that $\Lambda(x, t)$ is uniformly nonsingular for $x \in \Gamma$, i.e., the eigenvalues are bounded away from zero. However, the number of positive and negative eigenvalues may differ from one boundary point to another. The analytic boundary conditions can thus be expressed as

$$L(x) u(x, t) = 0 , \quad L(x) = Q_I^T(x) - S(x) Q_{II}^T(x). \tag{4.4}$$

Clearly, $L(x)$ has full rank for every $x \in \Gamma$. Strictly speaking, $L(0, 0)$ is not defined so far, because the normal $n(0, 0)$ is not well defined. It will soon be shown how to define $L(0, 0)$, and we can formally consider $L(x)$ as being defined for every $x \in \Gamma$.

Let v_{ij}, $i = 0, \dots, \nu_1$, $j = 0, \dots, \nu_2$ be a grid function. Define $v^T = (v_0^T \dots v_{\nu_2}^T)$, $v_j^T = (v_{0j}^T \dots v_{\nu_1 j}^T)$. The discretized boundary conditions are written as

$$L_{ij}^T v_j = 0 \tag{4.5}$$

for $i = 0, \nu_1$, $j = 1, \dots, \nu_2 - 1$ and $j = 0, \nu_2$, $i = 0, \dots, \nu_1$, where

$$L_{ij}^T = (0 \quad \dots \quad 0 \quad L(ih_1, jh_2) \quad 0 \quad \dots \quad 0) \in \mathbf{R}^{d_1(i,j) \times (\nu_1 + 1)d} ,$$

with the nonzero element being the ith entry. At the origin we define

$$L(0,0) = Q_I^T(0,0) - S(0,0)Q_{II}^T(0,0)\,, \tag{4.6}$$

where $Q(0,0)$ fulfills

$$Q^T A_{00} Q = \Lambda_{00}\,, \quad A_{00} = \sum_{i=1}^{2} n_i A_i(0,0,t)\,, \quad n_1 = -\frac{h_2}{h}\,,\ n_2 = -\frac{h_1}{h}\,,$$

where $h = \sqrt{h_1^2 + h_2^2}$. The motive for defining $L(0,0)$ this way will be evident later. Furthermore, A_{00} is supposed to be nonsingular. Let

$$\begin{aligned}
L_0 &= (L_{00} \ \dots \ L_{\nu_1 0}) && \in \mathbf{R}^{(\nu_1+1)d\times s_0}\,, \ s_0 = \sum_{i=0}^{\nu_1} d_1(i,0)\,,\\
L_j &= (L_{0j} \ \ L_{\nu_1 j}) && \in \mathbf{R}^{(\nu_1+1)d\times s_j}\,, \ s_j = \sum_{i=0 \text{ and } \nu_1} d_1(i,j)\,,\\
L_{\nu_2} &= (L_{0\nu_2} \ \dots \ L_{\nu_1\nu_2}) && \in \mathbf{R}^{(\nu_1+1)d\times s_{\nu_2}}\,, \ s_{\nu_2} = \sum_{i=0}^{\nu_1} d_1(i,\nu_2),
\end{aligned}$$

where $j = 1, \dots, \nu_2 - 1$. The boundary conditions may thus be expressed as

$$L^T v = 0\,, \quad L = \begin{pmatrix} L_0 & & \\ & \ddots & \\ & & L_{\nu_2} \end{pmatrix} \in \mathbf{R}^{(\nu_1+1)(\nu_2+1)d\times s}\,, \ s = \sum_{j=0}^{\nu_2} s_j\,. \tag{4.7}$$

Obviously, $\operatorname{rank}(L) = s$, i.e., L has full rank. Hence, the corresponding boundary projection is well defined, and is given by

$$P = I - \Sigma^{-1} L (L^T \Sigma^{-1} L)^{-1} L^T\,,$$

where

$$\Sigma = \begin{pmatrix} \sigma_0\Sigma_1 & & \\ & \ddots & \\ & & \sigma_{\nu_2}\Sigma_1 \end{pmatrix}\,, \quad \Sigma_1 = \begin{pmatrix} \sigma_0 I & & \\ & \ddots & \\ & & \sigma_{\nu_1} I \end{pmatrix}\,, \quad I \in \mathbf{R}^{d\times d}\,.$$

It is possible to simplify the expression for P in this case. We have

$$\Sigma^{-1} L = \begin{pmatrix} \Sigma_1^{-1} L_0/\sigma_0 & & \\ & \ddots & \\ & & \Sigma_1^{-1} L_{\nu_2}/\sigma_{\nu_2} \end{pmatrix}\,.$$

But $\Sigma_1^{-1} L_j = L_j H_j$, where

$$\begin{aligned}
H_j &= \begin{pmatrix} I/\sigma_0 & \\ & I/\sigma_{\nu_1} \end{pmatrix} && \in \mathbf{R}^{s_j\times s_j}\,, \ j = 1, \dots, \nu_2 - 1\,,\\
H_j &= \begin{pmatrix} I/\sigma_0 & & \\ & \ddots & \\ & & I/\sigma_{\nu_1} \end{pmatrix} && \in \mathbf{R}^{s_j\times s_j}\,, \ j = 0, \nu_2\,.
\end{aligned}$$

Hence

$$\Sigma^{-1}L = LH\,, \quad H = \begin{pmatrix} H_0/\sigma_0 & & \\ & \ddots & \\ & & H_{\nu_2}/\sigma_{\nu_2} \end{pmatrix} \in \mathbf{R}^{s\times s}\,.$$

Clearly, H is invertible. We therefore arrive at

$$P = I - LH(L^TLH)^{-1}L^T = I - L(L^TL)^{-1}L^T\,,$$

i.e., P is independent of Σ.

The semidiscrete system can now be defined as

$$\begin{aligned} v_t &= P\left(\sum_{i=1}^{2} A_iD_iv + F\right)\,, \\ v(0) &= f\,. \end{aligned} \tag{4.8}$$

It will next be shown that the solution to the system above satisfies an energy estimate.

Proposition 4.1. *Let* $(\cdot\,,\cdot)_h$ *be given by* (2.4) *and suppose that* D_1 *and* D_2 *satisfy the conclusion of Proposition* 2.3. *If* P *is defined by* (2.10) *and* (4.7), *then the solution of* (4.8) *satisfies an energy estimate*

$$\|v(t)\|_h^2 + \int_0^t \|v(\tau)\|_\Gamma^2 d\tau \le Ke^{\alpha' t}\left(\|f\|_h^2 + \int_0^t \|F(\tau)\|_h^2 d\tau\right)\,,$$

where the boundary energy $\|\cdot\|_\Gamma$ *is given by* ($\nu_1 = \nu_2 = \nu$ *for convenience*)

$$\|v(\tau)\|_\Gamma^2 = h_2\sum_{j=0}^{\nu}\sigma_j\left(|v_{0j}|^2 + |v_{\nu j}|^2\right) + h_1\sum_{i=0}^{\nu}\sigma_i\left(|v_{i0}|^2 + |v_{i\nu}|^2\right)\,.$$

Proof. From Propositions 2.5 and 2.4 we obtain

$$\frac{d}{dt}\|v\|_h^2 = 2\sum_{i=1}^{2}(v\,,\,A_iD_iv)_h + 2(v\,,\,F)_h\,.$$

From Proposition 2.3 and Corollary 2.1 it follows that (v is only supported in a neighborhood of $(0\,,\,0)$)

$$\begin{aligned} (v\,,\,A_1D_1v)_h &= -\frac{1}{2}\left(h_2\sum_{j=0}^{\nu}\sigma_j v_{0j}^T(A_1v)_{0j} + (v\,,\,[D_1\,,\,A_1]v)_h\right)\,, \\ (v\,,\,A_2D_2v)_h &= -\frac{1}{2}\left(h_1\sum_{i=0}^{\nu}\sigma_i v_{i0}^T(A_2v)_{i0} + (v\,,\,[D_2\,,\,A_2]v)_h\right)\,. \end{aligned}$$

Thus, by Lemma 2.3 we have

$$\begin{aligned} \frac{d}{dt}\|v\|_h^2 \le{}& -h_1\sum_{i=0}^{\nu}\sigma_i v_{i0}^T(A_2v)_{i0} - h_2\sum_{j=0}^{\nu}\sigma_j v_{0j}^T(A_1v)_{0j} \\ &+ \left(\sum_{i=1}^{2}\rho([D_i\,,\,A_i]) + 1\right)\|v\|_h^2 + \|F\|_h^2\,. \end{aligned}$$

In the first sum the outward unit normal is $n = (0, -1)$, and in the second $n = (-1, 0)$. Except for the origin, the boundary terms are of exactly the same form as in the one-dimensional case. Eqs. (4.2), (4.3) thus imply that

$$-v_{i0}^T(A_2 v)_{i0} = \varphi_{i0}^T \Lambda_{i0} \varphi_{i0} \le -\frac{\gamma_{i0}}{2}|\varphi_{0i}|^2 = -\frac{\gamma_{i0}}{2}|v_{0i}|^2 , \quad i \ge 1 .$$

A similar inequality holds for the other terms. At the origin we get

$$-h_2\sigma_0 v_{00}^T(A_1 v)_{00} - h_1\sigma_0 v_{00}^T(A_2 v)_{00} = h\sigma_0 \varphi_{00}^T \Lambda_{00}\varphi_{00} \le -h\sigma_0 \frac{\gamma_{00}}{2}|\varphi_{00}|^2 .$$

But $h \ge (h_1 + h_2)/\sqrt{2}$. Hence,

$$-h_2\sigma_0 v_{00}^T(A_1 v)_{00} - h_1\sigma_0 v_{00}^T(A_2 v)_{00} \le -h_1\sigma_0 \frac{\gamma_{00}}{2\sqrt{2}}|v_{00}|^2 - h_2\sigma_0 \frac{\gamma_{00}}{2\sqrt{2}}|v_{00}|^2 .$$

Since $\Lambda(x)$ is uniformly nonsingular it follows that $\gamma \equiv \inf(\gamma_{00}/\sqrt{2}, \gamma_{i0}, \gamma_{0j}) > 0$. Because of γ_{00}, the constant γ will in general be smaller than the corresponding constant of the analytic energy estimate. We thus arrive at

$$\frac{d}{dt}||v||_h^2 + \frac{\gamma}{2}||v||_\Gamma^2 \le \left(\sum_{i=1}^{2} \rho([D_i, A_i]) + 1\right) ||v||_h^2 + ||F||_h^2 ,$$

which proves the proposition ($K = \max(1, 2/\gamma)$). □

4.2. **The heat equation.** The analysis of homogeneous Dirichlet conditions is straightforward, even if the domain of definition Ω is nontrivial. The problem lies in discretizing the Neumann conditions properly. This was clear in one space dimension. In two dimensions the occurrence of corners certainly complicates the analysis. To gain insight, we shall begin by looking at a simple model problem.

The two-dimensional heat equation reads

$$\begin{aligned} &u_t = u_{x_1x_1} + u_{x_2x_2} , && x \in \Omega = (0, 1) \times (0, 1) , \\ &u_n(x, t) = 0 , && x \in \Gamma , \\ &u(x, 0) = f(x) , \end{aligned}$$

where u_n is the normal derivative of u. Again, we focus our attention to a neighborhood of $(0, 0)$. The boundary conditions are discretized as

(4.9)
$$\frac{1}{h_1}\sum_{k=0}^{r} d_{0k} v_{kj} = 0 , \quad j = 0, \dots , r , \qquad \frac{1}{h_2}\sum_{k=0}^{r} d_{0k} v_{ik} = 0 , \quad i = 0, \dots , r ,$$

or, equivalently,

(4.10) $$(D_1 v)_{0j} = 0 , \quad j = 0, \dots , r , \qquad (D_2 v)_{i0} = 0 , \quad i = 0, \dots , r ,$$

where D_1 and D_2 are defined by Proposition 2.3. The conditions above imply that two boundary conditions are prescribed at the origin for the discrete problem. This approach is natural from the intuitive point of view, in that gradients at the origin may be interpreted as one-sided limits from the interior. For the time being we ignore this technicality. It will later be shown how it can be overcome. When deriving the projection operator it is convenient to cast the

boundary conditions into yet another form. Define the boundary operators L_{1j} and L_{2i} through

$$L_{1j}^T v \equiv (D_1 v)_{0j} = 0\,, \qquad L_{2i}^T v \equiv (D_2 v)_{i0} = 0 \tag{4.11}$$

for $i, j = 0, \dots, r$, where

$$L_{1j}^T = \left(0 \quad \dots \quad 0 \quad \frac{1}{h_1}\sum_{k=0}^{r} d_{0k} e_k^T \quad 0 \quad \dots \quad 0\right) \in \mathbf{R}^{1\times(\nu_1+1)(\nu_2+1)}\,,$$

$$L_{2i}^T = \left(\frac{d_{00}}{h_2} e_i^T \quad \dots \quad \frac{d_{0r}}{h_2} e_i^T \quad 0 \quad \dots \quad 0\right) \in \mathbf{R}^{1\times(\nu_1+1)(\nu_2+1)}\,,$$

where $i, j = 0, \dots, r$; $\{e_i\}$ is the canonical basis in $\mathbf{R}^{\nu_1+1}$. The boundary conditions can thus be written in standard form $L^T v = 0$, where

$$L = (L_{10} \quad \dots \quad L_{1r} \quad L_{20} \quad \dots \quad L_{2r}) \in \mathbf{R}^{(\nu_1+1)(\nu_2+1)\times 2(r+1)}\,. \tag{4.12}$$

We know that the corresponding projection operator is well defined if and only if $\operatorname{rank}(L) = 2(r+1)$.

Lemma 4.1. *The columns of* L (4.12) *are linearly dependent. Thus,* $\operatorname{rank}(L) \le 2r+1$.

Proof. To investigate linear dependence, we study

$$\sum_{j=0}^{r} \alpha_j L_{1j} + \sum_{j=0}^{r} \beta_j L_{2j} = 0\,,$$

which is equivalent to

$$\sum_{k=0}^{r} (\alpha_j h_2 d_{0k} + \beta_k h_1 d_{0j})\, e_k = 0\,, \qquad j = 0, \dots, r\,.$$

Since $\{e_k\}$ is an orthonormal system, it follows that

$$\alpha_j h_2 d_{0k} + \beta_k h_1 d_{0j} = 0\,, \qquad j, k = 0, \dots, r\,,$$

which obviously has the nontrivial solution

$$\alpha_j = d_{0j}\,, \qquad \beta_j = -\frac{h_2}{h_1} d_{0j}\,, \qquad j = 0, \dots, r\,.$$

The lemma is proved. □

As a consequence of Lemma 4.1, the projection formulation breaks down. If, however, we change the boundary condition at the origin to

$$L_{0\chi}^T v \equiv ((1-\chi) L_{10}^T + \chi L_{20}^T)\, v = 0\,, \qquad 0 \le \chi \le 1\,, \tag{4.13}$$

and leave the boundary conditions at the remaining points unchanged, we get a well-defined projection operator, since

$$L = (L_{11} \quad \dots \quad L_{1r} \quad L_{0\chi} \quad L_{21} \quad \dots \quad L_{2r}) \in \mathbf{R}^{(\nu_1+1)(\nu_2+1)\times 2r+1} \tag{4.14}$$

has full rank.

Lemma 4.2. *The columns of* L (4.14) *are linearly independent. In particular,* $\operatorname{rank}(L) = 2r + 1$.

Proof. Again we study

$$\sum_{j=1}^{r} \alpha_j L_{1j} + \gamma L_{0\chi} + \sum_{j=1}^{r} \beta_j L_{2j} = 0 ,$$

which is the same as

$$\frac{d_{00}}{h_2} \sum_{k=1}^{r} \beta_k e_k + \gamma \left((1-\chi) \frac{1}{h_1} \sum_{k=0}^{r} d_{0k} e_k + \chi \frac{d_{00}}{h_2} e_0 \right) = 0 ,$$
$$\frac{d_{0j}}{h_2} \sum_{k=1}^{r} \beta_k e_k + \alpha_j \frac{1}{h_1} \sum_{k=0}^{r} d_{0k} e_k + \gamma\chi \frac{d_{0j}}{h_2} e_0 = 0 , \quad j = 1, \dots , r .$$

The first component of the first equation yields $\gamma(h_2(1-\chi) + h_1\chi)d_{00} = 0$. Since $d_{00} \neq 0$ for any operator satisfying Proposition 2.3, and since $h_i > 0$, $0 \leq \chi \leq 1$, necessarily $\gamma = 0$. From the remaining components of the first equation we then obtain $\beta_j = 0$, $j = 1, \dots , r$, which in turn implies $\alpha_j = 0$, $j = 1, \dots , r$. The columns of L are thus linearly independent, i.e., L has full rank. □

Before proceeding with the energy estimate, one more lemma is needed. Let $L_{0\chi_1}$ and $L_{0\chi_2}$ be defined by (4.13), and define $L \in \mathbf{R}^{(\nu_1+1)(\nu_2+1)\times 2(r+1)}$ by

$$L = (L_{11} \quad \dots \quad L_{1r} \quad L_{0\chi_1} \quad L_{0\chi_2} \quad L_{21} \quad \dots \quad L_{2r}) . \tag{4.15}$$

Lemma 4.3. *The columns of* L (4.15) *are linearly dependent. Thus,* $\operatorname{rank}(L) \leq 2r + 1$.

Proof. Consider

$$\sum_{j=1}^{r} \alpha_j L_{1j} + \gamma_1 L_{0\chi_1} + \gamma_2 L_{0\chi_2} + \sum_{j=1}^{r} \beta_j L_{2j} = 0 .$$

Obviously, the lemma is true for $\chi_1 = \chi_2$. In the following we thus assume $\chi_1 \neq \chi_2$. The equation above can be rewritten as

$$\sum_{j=0}^{r} \alpha_j L_{1j} + \sum_{j=0}^{r} \beta_j L_{2j} = 0 , \tag{4.16}$$

where

$$\begin{pmatrix} 1-\chi_1 & 1-\chi_2 \\ \chi_1 & \chi_2 \end{pmatrix} \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \beta_0 \end{pmatrix} .$$

According to Lemma 4.1, equation (4.16) has the nontrivial solution

$$\alpha_j = d_{0j} , \qquad \beta_j = -\frac{h_2}{h_1} d_{0j} , \qquad j = 0, \dots , r ,$$

whence

$$\gamma_1 = d_{00}\left(\chi_2 + \frac{h_2}{h_1}(1-\chi_2)\right)/(\chi_2-\chi_1), \qquad \alpha_j = d_{0j}, \qquad \beta_j = -\frac{h_2}{h_1}d_{0j},$$
$$\gamma_2 = -d_{00}\left(\chi_1 + \frac{h_2}{h_1}(1-\chi_1)\right)/(\chi_2-\chi_1),$$

solves the original equation. The lemma is proved. □

Proposition 4.2. *Let P be given by Proposition* 2.5, *where L is defined by* (4.14). *Then $L_{10}^T P = L_{20}^T P = 0$.*

Proof. Clearly, $L^T P = 0$. Furthermore, L_{10}, $L_{20} = L_{0\chi}$ for $\chi = 0, 1$, respectively. But then, by Lemma 4.3,

$$L_{10} = L\alpha_1, \qquad L_{20} = L\alpha_2,$$

for some vectors $\alpha_1, \alpha_2 \in \mathbf{R}^{2r+1}$. This proves the proposition. □

Remark 4.1. Suppose that v is a vector such that $v = Pv$, where P is as in the previous proposition. Then $L_{10}v = L_{20}v = 0$, i.e., $(D_1 v)_{00} = (D_2 v)_{00} = 0$. In other words, by requiring that the boundary condition at the origin hold for a specific convex combination, we actually get the stronger result $(D_1 v)_{00} = (D_2 v)_{00} = 0$. Thus, we need not overspecify at the corners, cf. equation (4.9). In the Supplement we give a direct proof that $L_{10}^T P = 0$ for $L_{0\chi}$ with $\chi = 0.5$.

The semidiscrete heat equation is given by

$$\begin{aligned} v_t &= P\left(D_1^2 + D_2^2\right) v, \\ v(0) &= f. \end{aligned} \tag{4.17}$$

Proposition 4.3. *Let $(\cdot, \cdot)_h$ be given by* (2.4) *and suppose that D_1 and D_2 satisfy the conclusion of Proposition* 2.3. *If P is defined by* (2.10) *and* (4.14), *then the solution of* (4.17) *satisfies an energy estimate*

$$\|v(t)\|_h \le \|f\|_h.$$

Proof. The energy method gives

$$\frac{d}{dt}\|v\|_h^2 = 2(v, D_1^2 v)_h + 2(v, D_2^2 v)_h.$$

By Proposition 2.3 (v is supported only in a neighborhood of the origin),

$$(v, D_1^2 v)_h = -h_2 \sum_{j=0}^{r} \sigma_j v_{0j} (D_1 v)_{0j} - \|D_1 v\|_h^2.$$

According to Propositions 2.4, 2.5 and 4.2 we have

$$(D_1 v)_{0j} = L_{1j}^T v = 0, \qquad j = 0, \dots, r.$$

The remaining term $(v, D_2^2 v)_h$ is treated similarly, and the proposition follows. □

4.3. **Parabolic systems.** Consider

(4.18)

$$\begin{aligned} &u_t = \sum_{i,j=1}^{2} A_{ij} u_{x_i x_j} + F, && x \in \Omega = (0,1) \times (0,1),\ u \in \mathbf{R}^d, \\ &u(x,0) = f(x), && x = (x_1, x_2), \\ &L_0(x)u(x,t) + L_1(x)u_n(x,t) = 0, && x \in \Gamma. \end{aligned}$$

The assumptions on L_0, L_1 in (3.4) are supposed to hold pointwise for each $x \in \Gamma$. Furthermore, we require that Assumption 3.1 with $A = A_{ii}$ be valid on $x_i = 0$, $i = 1, 2$. In particular, the conclusion of Lemma 3.2 holds for each boundary point. It will be assumed that (4.18) is strongly parabolic, i.e., for all vectors $u_i(x,t) \in \mathbf{R}^d$, $i = 1, 2$, one has

$$\sum_{i.j=1}^{2} u_i(x,t)^T A_{ij}(x,t) u_j(x,t) \geq 2\delta \sum_{i=1}^{2} |u_i(x,t)|^2$$

for all $x \in \Omega$, $t \geq 0$. If the matrices $A_{ij} \neq 0$, $i \neq j$, then the assumptions must be strengthened. The energy method applied to one of the cross terms yields (u is supported only at the origin, $A_{12} = \text{const}$ for simplicity, Ω is the unit square)

$$(u, A_{12}u_{x_1x_2}) = -\int_{x_1=0} u^T A_{12} u_{x_2} dx_2 - (u_{x_1}, A_{12}u_{x_2}).$$

In general we cannot get an estimate of $u_{x_2}(0, x_2, t)$ in the boundary integral. It is therefore natural to require

Assumption 4.1. $A_{ij}^T = A_{ij}$, $i \neq j$.

Remark 4.2. Neglecting scaling factors, we have

$$A_{12} = A_{21} = \frac{1}{\rho} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

for the Navier-Stokes equations (ρ denotes the density). Clearly, Assumption 4.1 is fulfilled.

If Assumption 4.1 holds, one can integrate by parts once more to obtain

$$(u, A_{12}u_{x_1x_2}) = \frac{1}{2} u^T A_{12}(0,0,t) u - (u_{x_1}, A_{12}u_{x_2}).$$

In two dimensions we cannot eliminate the boundary terms by means of Sobolev inequalities, since they would involve L^2-norms of $u_{x_1x_1}$ and so forth. This motivates

Assumption 4.2. Let $u(x,t)$ satisfy

$$L_0(0,0)u + L_1(0,0)u_n = 0$$

at the origin. Then

$$u^T A_{ij}(0,0,t)u = 0, \quad i \neq j.$$

Remark 4.3. This assumption ensures an energy estimate for the continuous problem in case of a nonsmooth boundary, and couples the cross terms of the differential operator to the boundary conditions at the origin. In case of the Navier-Stokes equations one has zero velocity at the origin. Hence, the state vector becomes $u^T = (\rho \quad 0 \quad 0 \quad p)$, which implies Assumption 4.2.

The discrete boundary conditions are formulated as (D_1 and D_2 are defined by Proposition 2.3)

$$
\begin{aligned}
&L_{1j}^T v \equiv L_0(0, jh_2)v_{0j} + L_1(0, jh_2)(D_1 v)_{0j} = 0, \quad j = 0, \dots, r, \\
&L_{2i}^T v \equiv L_0(ih_1, 0)v_{i0} + L_1(ih_1, 0)(D_2 v)_{i0} = 0, \quad i = 0, \dots, r,
\end{aligned} \tag{4.19}
$$

where

$$
L_{1j}^T = \left(0 \quad \dots \quad 0 \quad L_0(0, jh_2)e_0^T + L_1(0, jh_2)\frac{1}{h_1}\sum_{k=0}^{r} d_{0k}e_k^T \quad 0 \quad \dots \quad 0\right),
$$

$$
L_{2i}^T = \left(\left(L_0(ih_1, 0) + L_1(ih_1, 0)\frac{d_{00}}{h_2}\right)e_i^T \quad \dots \quad L_1(ih_1, 0)\frac{d_{0r}}{h_2}e_i^T \quad 0 \quad \dots \quad 0\right),
$$

and $e_i^T = (0 \quad \dots \quad 0 \quad I \quad 0 \quad \dots \quad 0) \in \mathbf{R}^{d \times (\nu_1+1)d}$. The boundary conditions can be expressed in the usual form $L^T v = 0$, where $L \in \mathbf{R}^{(\nu_1+1)(\nu_2+1)d \times (2r+1)d}$ is given by

$$
L = (L_{11} \quad \dots \quad L_{1r} \quad L_{0\chi} \quad L_{21} \quad \dots \quad L_{2r}), \tag{4.20}
$$

and

$$
L_{0\chi} \equiv (1 - \chi)L_{10} + \chi L_{20}, \quad 0 \le \chi \le 1.
$$

Lemma 4.4. *The columns of* L (4.20) *are linearly independent for sufficiently small step lengths* h_1 *and* h_2. *In particular,* $\operatorname{rank}(L) = (2r + 1)d$.

Proof. Imitating the proof of Lemma 4.2 gives

$$
\gamma\left[L_0(0, 0) + d_{00}\left(\frac{1 - \chi}{h_1} + \frac{\chi}{h_2}\right)L_1(0, 0)\right] = 0.
$$

By Lemma 3.2 the expression inside the brackets is nonsingular for h_1, h_2 sufficiently small. Hence, $\gamma = 0$, which in turn implies $\alpha_j = \beta_j = 0$, $j = 1, \dots, r$. Since the columns of each block L_{1j}, L_{2j} and $L_{0\chi}$ are linearly independent, the lemma follows. □

The semidiscrete parabolic system reads

$$
\begin{aligned}
v_t &= P\left(\sum_{i,j=1}^{2} A_{ij}D_iD_jv + F\right), \\
v(0) &= f,
\end{aligned} \tag{4.21}
$$

where P is defined by Proposition 2.5 and by (4.20). Unfortunately, Assumption 4.2 is not sufficient for the semidiscrete problem. We need

Assumption 4.3. Let v satisfy

$$L_0(0,0)v_{00} + L_1(0,0)((1-\chi)(D_1v)_{00} + \chi(D_2v)_{00}) = 0\,, \qquad 0 \le \chi \le 1\,,$$

at the origin. Then

(i)

$$v_{00}^T A_{ij}(0,0,t)v_{00} = 0\,, \quad i \neq j\,,$$

(ii)

$$v_{00}^T A_{11}(0,0,t) = v_{00}^T A_{22}(0,0,t)\,.$$

Remark 4.4. The first requirement is identical to that of Assumption 4.2. The second, however, appears only in the discrete case. We note that Assumption 4.3 holds for the Navier-Stokes equations, since A_{11} and A_{22} are given by

$$A_{11} = \frac{1}{\rho}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & c_1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -c_2 p/\rho & 0 & 0 & c_2 \end{pmatrix}, \qquad A_{22} = \frac{1}{\rho}\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & c_1 & 0 \\ -c_2 p/\rho & 0 & 0 & c_2 \end{pmatrix}.$$

Hence, $v_{00}^T A_{11} = v_{00}^T A_{22} = c_2(-p^2/\rho^2 \quad 0 \quad 0 \quad p/\rho)$.

Proposition 4.4. *Let* $(\cdot,\cdot)_h$ *be given by* (2.4) *and suppose that* D_1 *and* D_2 *satisfy the conclusion of Proposition* 2.3. *If* P *is defined by* (2.10) *and* (4.20), *and if Assumptions* 4.1 *and* 4.3 *hold, then the solution of* (4.21) *satisfies an energy estimate*

$$\|v(t)\|_h^2 + \int_0^t \|v(\tau)\|_\Gamma^2 d\tau \le e^{(\alpha' + \mathcal{O}(h))t}\left(\|f\|_h^2 + \int_0^t \|F(\tau)\|_h^2 d\tau\right).$$

Proof. The energy method yields

$$\begin{aligned}\frac{d}{dt}\|v\|_h^2 \;\le\; & -2\sum_{j=1}^{2}\left(h_2\sum_{k=0}^{r}\sigma_k v_{0k}^T(A_{1j}D_jv)_{0k} + h_1\sum_{k=0}^{r}\sigma_k v_{k0}^T(A_{2j}D_jv)_{k0}\right) \\ & -2\delta\sum_{i=1}^{2}\|D_iv\|_h^2 + (K_0 + \mathcal{O}(h))\|v\|_h^2 + \|F\|_h^2\,,\end{aligned}$$

where K_0 depends on $\|[D_i, A_{ii}]\|_h$, $i = 1, 2$ and $\rho([D_i, A_{ij}])$, $i \neq j$. The first cross term can be written as (v has compact support)

$$\begin{aligned}-h_2\sum_{k=0}^{r}\sigma_k v_{0k}^T(A_{12}D_2v)_{0k} &= -h_2\sum_{k=0}^{\nu}\sigma_k v_{0k}^T A_{12}(0, kh_2, t)\left(\frac{1}{h_2}\sum_{l=0}^{\nu}d_{kl}v_{0l}\right) \\ &\equiv -(v_0, \tilde{A}_{12}\tilde{D}_2v_0)_{h_2}\,,\end{aligned}$$

where $v_0^T = (v_{00}^T \ldots v_{0\nu}^T)$, and where $\tilde{D}_2$ satisfies (2.1) with respect to the one-dimensional scalar product $(\cdot,\cdot)_{h_2}$. Hence,

$$-(v_0, \tilde{A}_{12}\tilde{D}_2v_0)_{h_2} = \frac{1}{2}v_{00}^T A_{12}(0,0,t)v_{00} + \frac{1}{2}(v_0, [\tilde{D}_2, \tilde{A}_{12}]v_0)_{h_2}\,.$$

By Assumption 4.3 the boundary terms vanish. The remaining cross term is treated in a similar manner.

Next, we take care of the boundary terms corresponding to the pure second differences. Only the origin needs to be analyzed, since the other boundary points are treated exactly as in the proof of Proposition 3.2. At the origin we get ($A_{ii} = A_{ii}(0, 0, t)$)

$$
\begin{aligned}
&- h_2\sigma_0 v_{00}^T A_{11}(D_1 v)_{00} - h_1\sigma_0 v_{00}^T A_{22}(D_2 v)_{00} \\
&\quad = -(h_1 + h_2)\sigma_0 v_{00}^T \left((1-\chi)A_{11}(D_1 v)_{00} + \chi A_{22}(D_2 v)_{00}\right) , \qquad \chi = \frac{h_1}{h_1 + h_2} ,
\end{aligned}
$$

and, by Assumption 4.3,

$$
\begin{aligned}
&- h_2\sigma_0 v_{00}^T A_{11}(D_1 v)_{00} - h_1\sigma_0 v_{00}^T A_{22}(D_2 v)_{00} \\
&\quad = -(h_1 + h_2)\sigma_0 v_{00}^T A_{11} \left((1-\chi)(D_1 v)_{00} + \chi(D_2 v)_{00}\right) .
\end{aligned}
$$

But $v = Pv$ implies $L_{0\chi} v = 0$, i.e., by (4.19),

$$
L_0(0, 0)v_{00} + L_1(0, 0)\left((1-\chi)(D_1 v)_{00} + \chi(D_2 v)_{00}\right) = 0 .
$$

In particular, $L_0^{II} v_{00} = 0$. Partition $v_{ij} = v'_{ij} + v''_{ij}$, where $v'_{ij} \in \ker(L_1^I)$, $v''_{ij} \in \ker(L_1^I)^\perp$. Assumption 3.1 then gives

$$
\begin{aligned}
&- h_2\sigma_0 v_{00}^T A_{11}(D_1 v)_{00} - h_1\sigma_0 v_{00}^T A_{22}(D_2 v)_{00} \\
&\quad = -(h_1 + h_2)\sigma_0 v_{00}^T A_{11} \left((1-\chi)(D_1 v'')_{00} + \chi(D_2 v'')_{00}\right) .
\end{aligned}
$$

By construction,

$$
L_0(0, 0)v_{00} + L_1(0, 0)\left((1-\chi)(D_1 v'')_{00} + \chi(D_2 v'')_{00}\right) = 0 ,
$$

which can be solved in exactly the same way as the corresponding equation in the proof of Proposition 3.2. Hence,

$$
\begin{aligned}
&- h_2\sigma_0 v_{00}^T A_{11}(D_1 v)_{00} - h_1\sigma_0 v_{00}^T A_{22}(D_2 v)_{00} \\
&\quad = h_2\sigma_0 v_{00}^T A_{11}\tilde{L}_1^{-1} L_0 v_{00} + h_1\sigma_0 v_{00}^T A_{22}\tilde{L}_1^{-1} L_0 v_{00} ,
\end{aligned}
$$

where we again have invoked Assumption 4.3. We thus arrive at

$$
\begin{aligned}
\frac{d}{dt}||v||_h^2 + ||v||_\Gamma^2 \le{}& \left(2|A_{11}\tilde{L}_1^{-1} L_0|_{2,\infty} + \rho([\tilde{D}_2, \tilde{A}_{12}]) + 1\right) h_2 \sum_{k=0}^{r} \sigma_k |v_{0k}|^2 \\
&+ \left(2|A_{22}\tilde{L}_1^{-1} L_0|_{1,\infty} + \rho([\tilde{D}_1, \tilde{A}_{21}]) + 1\right) h_1 \sum_{k=0}^{r} \sigma_k |v_{k0}|^2 \\
&- 2\delta \sum_{i=1}^{2} ||D_i v||_h^2 + (K_0 + \mathscr{O}(h))\,||v||_h^2 + ||F||_h^2 ,
\end{aligned}
$$

where $|v|_{1,\infty} = \sup(|v_{k0}|)$ and $|v|_{2,\infty} = \sup(|v_{0k}|)$. Replacing $\rho([\tilde{D}_i, \tilde{A}_{ji}])$ by $|A'_{ji}|_{i,\infty}$, $i \neq j$, one obtains the coefficients of the boundary terms of the analytic energy estimate. They are thus identical if the coefficient matrices are constant or if we use the standard second-order method. Finally, the boundary terms of the right hand are eliminated by applying the one-dimensional Sobolev inequality 2.6 in the x_1- and x_2-directions, respectively. This proves the proposition. □

Remark 4.5. It is clear from the proof that (4.21) is strictly stable if the coefficient matrices are constant, or if $A_{ii}^T = A_{ii}$, $i = 1, 2$, and the second-order method (2.3) is used.

5. Summary and conclusions

We have demonstrated that for a given finite-dimensional scalar product $(\cdot, \cdot)_h$ any linear discretized boundary condition can be written as an orthogonal projection operator P that satisfies $(u, Pv)_h = (Pu, v)_h$. It should be noted that the projection is well defined if the corresponding analytic problem is well posed. For general boundary conditions one may also have to require that the discretization parameter h be small enough (consistency). The projections P, the summation-by-parts property, and Proposition 2.4 constitute the main tools needed to obtain an energy estimate for the semidiscrete case. For a large class of problems it has been established that existence of an energy estimate for the continuous problem implies the same for the semidiscrete system.

In one space dimension we are no longer required to consider restricted full norms

$$\Sigma = \begin{pmatrix} 1 & & & \\ & \Sigma^{(1)} & & \\ & & 1 & \\ & & & \ddots \end{pmatrix},$$

which were used in [4] to prove stability for symmetric hyperbolic systems subject to homogeneous boundary conditions. The main result is the stability proof for mixed hyperbolic-parabolic systems subject to general linear boundary conditions. Reformulating the analytic problem makes it possible to obtain strict stability, i.e., we have a time stable semidiscrete approximation that is bounded by the same exponential growth rate (modulo $\mathcal{O}(h)$) as the analytic problem. For the parabolic part the excess growth rate is induced by the discrete Sobolev inequality. Furthermore, for the hyperbolic part we have used Assumption 2.1. In particular, strict stability is obtained for diagonal norms and variable-coefficient problems, and for general norms and constant-coefficient problems. The stability results hold for finite difference approximations of arbitrary order.

In two space dimensions we are forced to consider diagonal norms in order to have summation by parts in both dimensions. Stability of high-order schemes is obtained for general hyperbolic and parabolic initial-boundary value problems. All results obtained for two dimensions generalize to higher dimensions. Furthermore, the stability results are valid even if there are corners present. Although there are no general existence proofs for hyperbolic-parabolic problems on nonsmooth domains, it may still be useful to have stability results that allow for corners since they appear in most multi-dimensional finite difference implementations.

The methods presented in this paper are similar to finite element methods in that stability for the semidiscrete system follows more or less directly from the corresponding continuous one. There is, however, one major difference: The FEM technique often results in implicit space discretization, whereas the discretized space operators reported in this paper always are explicit.

There are other ways of imposing boundary conditions so as to ensure time stability (strict stability) when using difference operators satisfying a summation-by-parts property. An elegant technique is proposed in [2]. A so-called Simultaneous Approximation Term, SAT for short, is added to the semidiscrete scheme. The SAT will act as a penalty function to enforce an approximation of the discrete boundary conditions. In [2] this approach is used to prove time stability for high-order finite difference approximations of one-dimensional constant-coefficient hyperbolic systems. Also, it is not necessary to consider identical difference stencils in the interior. A new and interesting class of such difference operators can be found in [1].

ACKNOWLEDGMENT

The author wishes to thank Professor Joseph Oliger for many stimulating discussions on the topics of this paper.

BIBLIOGRAPHY

1. M. Carpenter and J. Otto, *High-order cyclo-difference techniques: A new methodology for finite differences*, Tech. report, ICASE, NASA Langley Research Center, Hampton, VA 23681-0001, 1993.
2. D. Gottlieb, M. Carpenter, and S. Abarbanel, *Time-stable boundary conditions for finite-difference schemes solving hyperbolic systems: Methodology and application to high-order compact schemes*, Tech. Report 93-21, ICASE, NASA Langley Research Center, Hampton, VA 23681-0001, 1993.
3. H.-O. Kreiss and J. Lorenz, *Initial-boundary value problems and the Navier-Stokes equations*, Pure and Appl. Math., vol. 136, Academic Press, San Diego, CA, 1989.
4. H.-O. Kreiss and G. Scherer, *Finite element and finite difference methods for hyperbolic partial differential equations*, Mathematical Aspects of Finite Elements in Partial Differential Equations (C. de Boor, ed.), Academic Press, New York, 1974, pp. 195–212.
5. ———, *On the existence of energy estimates for difference approximations for hyperbolic systems*, Tech. Report, Dept. of Scientific Computing, Uppsala University, 1977.
6. P. Olsson, *Stable approximation of symmetric hyperbolic and parabolic equations in several space dimensions*, Tech. Report 138, Dept. of Scientific Computing, Uppsala Univ., Uppsala, Sweden, December 1991.
7. ———, *Summation by parts, projections, and stability.* II, Math. Comp., to appear.
8. G. Scherer, *Numerical computations with energy estimates schemes*, Tech. report, Dept. of Scientific Computing, Uppsala Univ., Uppsala, Sweden, April 1977. In PhD thesis by G. Scherer, 1977.
9. B. Strand, *Summation by parts for finite difference approximations for d/dx*, J. Comput. Phys. **110** (1994), 47–67.

RIACS, MAIL STOP T20G-5, NASA AMES RESEARCH CENTER, MOFFETT FIELD, CALIFORNIA 94035-1000

E-mail address: pelle@riacs.edu

MATHEMATICS OF COMPUTATION
VOLUME 64, NUMBER 211
JULY 1995, PAGES S23–S26

Supplement to
SUMMATION BY PARTS, PROJECTIONS, AND STABILITY. I

PELLE OLSSON

Let $P = I - \Sigma^{-1}L(L^T\Sigma^{-1}L)^{-1}L^T$ where $L = \begin{pmatrix} L_{11} & \dots & L_{1\tau} & L_{0\chi} & L_{21} & \dots & L_{2\tau} \end{pmatrix}$ represents homogeneous Neumann conditions (localized to the origin), cf. section on the heat equation. For convenience we have $\chi = 0.5$ and $h_1 = h_2 = 1$. We shall show that $L_{10}^T P = 0$, which will follow if we can prove that $L_{10}^T\Sigma^{-1}L(L^T\Sigma^{-1}L)^{-1}L^T = L_{10}^T$. Straightforward computations show that

$$L^T\Sigma^{-1}L = \begin{pmatrix} D_{11} & D_{12} & D_{13} \\ D_{12}^T & D_{22} & D_{23} \\ D_{13}^T & D_{23}^T & D_{33} \end{pmatrix},$$

where

$$D_{11} = D_{33} = \kappa \begin{pmatrix} \frac{1}{\sigma_1} & & \\ & \ddots & \\ & & \frac{1}{\sigma_\tau} \end{pmatrix}, \quad D_{12} = D_{23}^T = \tau^{-1} \begin{pmatrix} \frac{d_{01}}{\sigma_1} \\ \vdots \\ \frac{d_{0\tau}}{\sigma_\tau} \end{pmatrix},$$

and

$$D_{13} = \tau^2 D_{12}D_{23}, \quad D_{22} = \frac{1}{2}\left(\frac{\kappa}{\sigma_0} + \frac{d_{00}^2}{\sigma_0^2}\right), \quad \kappa = \sum_{k=0}^{\tau} \frac{d_{0k}^2}{\sigma_k}, \quad \tau = \frac{2\sigma_0}{d_{00}}.$$

The inverse is given by

$$(L^T\Sigma^{-1}L)^{-1} = \begin{pmatrix} T_{11} & T_{12} & T_{13} \\ T_{12}^T & T_{22} & T_{23} \\ T_{13}^T & T_{23}^T & T_{33} \end{pmatrix}$$

with

$$T_{11} = T_{33} = D_{11}^{-1} + \frac{\mu}{1-\mu\sigma} D_{11}^{-1}D_{12}D_{12}^T D_{11}^{-1},$$

$$T_{12} = T_{23}^T = \frac{-(\mu - \sigma\tau^4)}{(1-\mu\sigma)(1-\sigma\tau^2)} D_{11}^{-1}D_{12},$$

$$T_{13} = T_{13}^T = \frac{\mu - \tau^2}{(1-\mu\sigma)(1-\sigma\tau^2)} D_{11}^{-1}D_{12}D_{23}D_{33}^{-1},$$

$$T_{22} = \frac{(\mu - \sigma\tau^4)(1+\sigma\tau^2)}{(1-\mu\sigma)(1-\sigma\tau^2)},$$

and

$$\sigma = D_{12}^T D_{11}^{-1} D_{12},$$
$$\mu = \sigma\tau^4 + \frac{1}{\nu}\left(1 - \sigma\tau^2\right)^2,$$
$$\nu = D_{22} - \sigma.$$

© 1995 American Mathematical Society

Let $\dot{L} = \begin{pmatrix} L_{10} & \dots & L_{1r} & L_{20} & \dots & L_{2r} \end{pmatrix}$. Then

$$L_{10}^T \Sigma^{-1} L (L^T \Sigma^{-1} L)^{-1} L^T = L_{10}^T \Sigma^{-1} \dot{L} \begin{pmatrix} R & S \\ S & R \end{pmatrix} \dot{L}^T,$$

where (using $T_{12}^T = T_{23}$ and $T_{11} = T_{33}$)

$$R = \begin{pmatrix} T_{22}/4 & T_{12}^T/2 \\ T_{12}/2 & T_{11} \end{pmatrix}, \quad S = \begin{pmatrix} T_{22}/4 & T_{12}^T/2 \\ T_{12}/2 & T_{13} \end{pmatrix}.$$

Furthermore,

$$L_{10}^T \Sigma^{-1} \dot{L} = \begin{pmatrix} \kappa/\sigma_0 & 0 & d_{00}^2/\sigma_0^2 & 2D_{12}^T \end{pmatrix}.$$

Using $D_{22} = (\kappa/\sigma_0 + d_{00}^2/\sigma_0^2)/2$, we have

$$L_{10}^T \Sigma^{-1} \dot{L} \begin{pmatrix} R & S \\ S & R \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{2} D_{22} T_{22} + D_{12}^T T_{12} & D_{22} T_{12}^T + 2 D_{12}^T T_{13} & \frac{1}{2} D_{22} T_{22} + D_{12}^T T_{12} & D_{22} T_{12}^T + 2 D_{12}^T T_{11} \end{pmatrix}.$$

But

$$\frac{1}{2} D_{22} T_{22} + D_{12}^T T_{12} = \frac{1}{2} (D_{12}^T T_{12} + D_{22} T_{22} + D_{23} T_{23}^T) = \frac{1}{2},$$

and

$$D_{22} T_{12}^T + 2 D_{12}^T T_{13} = 2 \left(D_{12}^T T_{13} + D_{22} T_{12}^T + D_{12}^T T_{11} \right) - \left(D_{22} T_{12}^T + 2 D_{12}^T T_{11} \right).$$

Observing that $T_{12}^T = T_{23}$, $D_{12}^T T_{11} = D_{23} T_{33}$, we conclude that the first parenthetical expression vanishes. Thus,

$$L_{10}^T \Sigma^{-1} \dot{L} \begin{pmatrix} R & S \\ S & R \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & D_{22} T_{12}^T + 2 D_{12}^T T_{13} & \frac{1}{2} & -\left(D_{22} T_{12}^T + 2 D_{12}^T T_{13} \right) \end{pmatrix}.$$

Also,

$$D_{22} T_{12}^T + 2 D_{12}^T T_{13} = \rho D_{12}^T D_{11}^{-1}, \quad \rho = \frac{2\sigma(\mu - \tau^2) - (\sigma + \nu)(\mu - \sigma\tau^4)}{(1 - \mu\sigma)(1 - \sigma\tau^2)}.$$

Substituting $1 - \sigma\tau^2 = d_{00}^2/(\kappa\sigma_0)$ and the expression for μ yields

$$\rho = -\frac{\kappa\sigma_0}{d_{00}^2},$$

and so

$$\rho D_{12}^T D_{11}^{-1} = -\frac{1}{2d_{00}} \begin{pmatrix} d_{01} & \dots & d_{0r} \end{pmatrix}.$$

Consequently,

$$L_{10}^T \Sigma^{-1} L (L^T \Sigma^{-1} L)^{-1} L^T = \frac{1}{2d_{00}} \begin{pmatrix} d_{00} & -d_{01} & \dots & -d_{0r} & d_{00} & d_{01} & \dots & d_{0r} \end{pmatrix} \dot{L}^T.$$

The jth block column of $\dot{L}^T$ reads

$$c_j^T = \begin{pmatrix} 0 & \dots & 0 & \sum_{k=0}^{r} d_{0k} e_k^T & 0 & \dots & 0 & d_{0j} e_0^T & \dots & d_{0j} e_r^T \end{pmatrix},$$

where the sum is the jth block element of c_j, $j = 0, \dots, r$. Hence,

$$\frac{1}{2d_{00}} \begin{pmatrix} d_{00} & -d_{01} & \dots & -d_{0r} & d_{00} & d_{01} & \dots & d_{0r} \end{pmatrix} c_j = \delta_{j0} \sum_{k=0}^{r} d_{0k} e_k^T,$$

i. e.,

$$L_{10}^T \Sigma^{-1} L (L^T \Sigma^{-1} L)^{-1} L^T - L_{10}^T.$$

In a similar fashion one shows that L_{20} also satisfies the above equation.

The simplest example is obtained by discretizing the Neumann conditions using the standard divided difference D_+ in both coordinate directions, i. e.,

$$\begin{aligned} &v_{11} - v_{01} = 0, \\ &v_{11} - v_{10} = 0, \\ &(v_{10} - v_{00})/2 + (v_{01} - v_{00})/2 = 0, \end{aligned} \tag{1}$$

which leads to

$$\Sigma^{-1} L (L^T \Sigma^{-1} L)^{-1} L^T = \frac{1}{18} \begin{pmatrix} 16 & -4 & 0 & \dots & 0 & -4 & -8 & 0 & \dots & 0 \\ -2 & 14 & 0 & \dots & 0 & -4 & -8 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ -2 & -4 & 0 & \dots & 0 & 14 & -8 & 0 & \dots & 0 \\ -2 & -4 & 0 & \dots & 0 & -4 & 10 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}. \tag{2}$$

Evidently, any vector $v = Pu$, $P = I - \Sigma^{-1} L (L^T \Sigma^{-1} L)^{-1} L^T$, satisfies (1). Furthermore, by (2), $v_{10} - v_{00} = v_{01} - v_{00} = 0$, which also follows directly from (1).

We conclude this Supplement by proving a number of technical lemmas that were used in the paper. Let D be a difference operator satifying a summation-by-parts rule with respect to the scalar product $(u, v)_h = h u^T \Sigma v$, where

$$\Sigma = \begin{pmatrix} \Sigma^{(1)} & & \\ & I & \\ & & \Sigma^{(2)} \end{pmatrix}, \quad \Sigma^{(l)} \in \mathrm{R}^{(r_l+1)d \times (r_l+1)d}, \quad l = 1.2. \tag{3}$$

Proposition 1.1. *Let D be as above, and define the norm $\|\cdot\|_h = \sqrt{(\cdot,\cdot)_h}$. Then*

$$|v|_\infty^2 \le \epsilon \|Dv\|_h^2 + \left(\epsilon^{-1} + 1 + \mathcal{O}(h)\right) \|v\|_h^2 ,$$

where $\epsilon > 0$.

Proof. Choose k, l such that

$$|v_k|^2 = \min_j \left(|v_j|^2\right) ,$$
$$|v_l|^2 = \max_j \left(|v_j|^2\right) \equiv |v|_\infty^2 .$$

Eq. (3) implies that

$$\|v\|_h^2 \ge h \left(\lambda_1 |v^{(1)}|^2 + \lambda_2 |v^{(2)}|^2\right) + h \sum_{j=r_1+1}^{\nu-r_2} |v_j|^2 ,$$

where $\lambda_{1,2} > 0$ are the smallest eigenvalues of $\Sigma^{(1,2)}$. Note that $\lambda_{1,2}$ are independent of h. Hence,

$$\|v\|_h^2 \ge \left(1 - h\left(r_1(1-\lambda_1) + r_2(1-\lambda_2)\right)\right) |v_k|^2 ,$$

where we have used $h\nu = L = 1$. If $c \equiv r_1(1-\lambda_1) + r_2(1-\lambda_2) \le 0$, one immediately gets $|v_k|^2 \le \|v\|_h^2$. Otherwise, we choose h such that $hc < 1$. Hence,

$$|v_k|^2 \le \frac{1}{1-hc} \|v\|_h^2 \le (1 + Kh) \|v\|_h^2 , \qquad K = \frac{c}{1 - h_0 c} , \tag{4}$$

for $h \le h_0$, where h_0 is a fixed number such that $h_0 c < 1$.

Next, we define a family of norms, which is obtained by shrinking the interior of (3); $\Sigma^{(1,2)}$ remain constant. Allowing a slight abuse of notation, we write these norms as

$$(u, v)_{h,r,s} = h \sum_{j=r}^{s} \sigma_{ij} u_i^T v_j ,$$

where $r \ge 0$ and $s \le \nu$. Shrinking the interior of D accordingly, one has

$$(v, Dv)_{h,k,l} = |v_l|^2 - |v_k|^2 - (Dv, v)_{h,k,l} ,$$

i. e.,

$$|v|_\infty^2 \le |v_k|^2 + 2 \|Dv\|_{h,k,l} \|v\|_{h,k,l} .$$

Obviously, $\|v\|_{h,k,l} \le \|v\|_{h,0,\nu} \equiv \|v\|_h$, whence

$$|v|_\infty^2 \le \epsilon \|Dv\|_h^2 + \left(\epsilon^{-1} + 1 + \mathcal{O}(h)\right) \|v\|_h^2 ,$$

where (4) and the standard algebraic inequality have been used. □

Lemma 1.1. *Let $A = \mathrm{diag}(A_0 \ \dots \ A_\nu)$, $A_j = A(jh) \in \mathbb{R}^d$. Then*

$$|(u, Av)_h - (A^T u, v)_h| \le \mathcal{O}(h) \|u\|_h \|v\|_h .$$

Proof. Denote the commutator of Σ and A by $[\Sigma, A]$. Then

$$(u, Av)_h = (A^T u, v)_h + h u^T [\Sigma, A] v ,$$

where

$$[\Sigma, A] = \begin{pmatrix} [\Sigma^{(1)}, A^{(1)}] & & \\ & 0 & \\ & & [\Sigma^{(2)}, A^{(2)}] \end{pmatrix}$$

with

$$[\Sigma^{(1)}, A^{(1)}] = \begin{pmatrix} 0 & \sigma_{01}(A_1 - A_0) & \dots & \sigma_{0r}(A_r - A_0) \\ -\sigma_{01}(A_1 - A_0) & 0 & \dots & \sigma_{1r}(A_r - A_1) \\ \vdots & & & \vdots \\ -\sigma_{0r}(A_r - A_0) & -\sigma_{1r}(A_r - A_1) & \dots & 0 \end{pmatrix} .$$

The other nonzero block has a similar structure. Assuming that $A(x)$ is differentiable, we can apply the mean value theorem:

$$[\Sigma^{(1)}, A^{(1)}] = h \begin{pmatrix} 0 & \sigma_{01} A'_{10} & \dots & \sigma_{0r} A'_{r0} \\ -\sigma_{01} A'_{10} & 0 & \dots & \sigma_{1r}(r-1) A'_{r1} \\ \vdots & & & \vdots \\ -\sigma_{0r} A'_{r0} & -\sigma_{1r}(r-1) A'_{r1} & \dots & 0 \end{pmatrix} .$$

Hence,

$$|(u, Av)_h - (A^T u, v)_h| \le c |A'|_\infty h^2 \left(|u^{(1)}||v^{(1)}| + |u^{(2)}||v^{(2)}|\right) \le \mathcal{O}(h) \|u\|_h \|v\|_h ,$$

which proves the lemma. □

Lemma 1.2. *Let A be as in the previous lemma. Then*

$$|(u, Av)_h| \le |A|_\infty (1 + \mathcal{O}(h)) \|u\|_h \|v\|_h .$$

where $|A|_\infty = \sup |A(x)|$.

Proof. The definition of $(\cdot,\cdot)_h$ implies that $(u, Av)_h = h \tilde{u}^T \tilde{A} \tilde{v}$, where $\tilde{u} = \Sigma^{1/2} u$, $\tilde{v} = \Sigma^{1/2} v$, and $\tilde{A} = \Sigma^{1/2} A \Sigma^{-1/2}$. Taylor expansion yields $\tilde{A} = \Lambda + R$,

$$R = \begin{pmatrix} R^{(1)} & & \\ & 0 & \\ & & R^{(2)} \end{pmatrix}$$

with $R^{(l)} = \mathcal{O}(h)$, $l = 1, 2$. Thus,

$$|(u, Av)_h| \le |A|_\infty \|\tilde{u}\| \|\tilde{v}\| + \mathcal{O}(h) \left(\|\tilde{u}^{(1)}\| \|\tilde{v}^{(1)}\| + \|\tilde{u}^{(2)}\| \|\tilde{v}^{(2)}\|\right) ,$$

i. e.,

$$|(u, Av)_h| \leq (|A|_\infty + \mathcal{O}(h)) \|\tilde{u}\| \|\tilde{v}\| ,$$

where $\|\cdot\|$ denotes the standard Euclidean norm. Since $\|\tilde{u}\| = |u|_h$, $\|\tilde{v}\| = \|v\|_h$, the lemma follows. □

Corollary 1.1. *If, in addition to the hypotheses of Lemma 1.2, one of the following conditions holds:*

(i) Σ *is diagonal* $\operatorname{diag}(\sigma_0 I \ldots \sigma_\nu I)$,

(ii) *The blocks of* A *satisfy* $A_0 = \ldots = A_{r_1}$ *and* $A_{\nu - r_2} = \ldots = A_\nu$.

then

$$|(u, Av)_h| \leq |A|_\infty \|u\|_h \|v\|_h .$$

Proof. The hypotheses imply that $\dot{A} = A$, and the corollary follows. □

Lemma 1.3. *Let* A *be as in Lemma 1.1 and assume that one of the following conditions holds:*

(i) Σ *is diagonal* $\operatorname{diag}(\sigma_0 I \ldots \sigma_\nu I)$,

(ii) *The blocks of* A *satisfy* $A_0 = \ldots = A_{r_1}$ *and* $A_{\nu - r_2} = \ldots = A_\nu$.

If A *is symmetric, then*

$$(u, [D, A]v)_h \leq \rho([D, A]) \|u\|_h \|v\|_h .$$

Proof. According to the definition of the operator norm we have

$$\|[D, A]\|_h^2 = \max_{\|v\|_h = 1} \|[D, A]v\|_h^2 = \max_{\|w\| = 1} h w^T C^T C w ,$$

where $w = \Sigma^{1/2} v$, $C = \Sigma^{1/2}[D, A]\Sigma^{-1/2}$. Because of the assumptions on A (or Σ) we have $C = \dot{D}A - A\dot{D}$, where $\dot{D} = \Sigma^{1/2} D \Sigma^{-1/2}$. Summation by parts implies that

$$\Sigma D = D_s + D_a , \quad D_s = \frac{1}{2} \begin{pmatrix} -I & & \\ & 0 & \\ & & I \end{pmatrix} , \quad I \in \mathbb{R}^{d \times d} ,$$

and D_a is an anti-symmetric matrix. Consequently,

$$C = [\Sigma^{-1/2} D_a \Sigma^{-1/2}, A] ,$$

where we have used $[\Sigma^{-1/2} D_s \Sigma^{-1/2}, A] = 0$. Since D_a is anti-symmetric and A symmetric, we have $C^T = C$, i. e.,

$$\|[D, A]\|_h^2 = \max_{\|w\| = 1} h w^T C^2 w - \rho(C)^2 .$$

Finally, $C = \Sigma^{1/2}[D, A]\Sigma^{-1/2}$ implies that

$$\|[D, A]\|_h = \rho([D, A]) ,$$

which proves the lemma. □