

## Single Step Galerkin Approximations for Parabolic Problems

By Garth A. Baker, James H. Bramble and Vidar Thomée

**Abstract.** In this paper we construct and analyze classes of single step methods of arbitrary order for homogeneous linear initial boundary value problems for parabolic equations with time-independent coefficients. The spatial discretization is done by means of general Galerkin-type methods.

**1. Introduction.** In this paper we shall consider the approximate solution of the initial boundary value problem

$$(1.1) \quad \begin{aligned} D_t u &= -Lu \equiv \sum_{j,k=1}^N \frac{\partial}{\partial x_j} \left( a_{jk} \frac{\partial u}{\partial x_k} \right) - a_0 u \quad \text{in } \Omega \times (0, t^*], \\ u(x, t) &= 0 \quad \text{on } \partial\Omega \times (0, t^*], \\ u(x, 0) &= v(x) \quad \text{for } x \in \Omega. \end{aligned}$$

Here  $\Omega$  is a bounded domain in  $R^N$  with sufficiently smooth boundary  $\partial\Omega$ ,  $a_{jk}$  and  $a_0$  are sufficiently smooth functions which are independent of  $t$ , the matrix  $(a_{jk})$  is symmetric and uniformly positive definite and  $a_0$  is nonnegative on  $\bar{\Omega}$ . All functions considered are real valued.

In Bramble, Schatz, Thomée and Wahlbin [3] the discretization of this problem in the space variables was considered by means of Galerkin's method, using a family  $\{S_h\}$  of finite element spaces with certain approximation properties, thus replacing (1.1) by an approximating system of ordinary differential equations in time. Special emphasis was placed on deriving optimal order error estimates for time bounded away from zero, under weak regularity assumptions on the initial data. This pursued a point of view taken previously in Blair [1], Helfrich [8] and Thomée [12] (cf. [3] for further references on semidiscrete Galerkin methods).

In the present paper the purpose is to study discretization also of the time variable. In particular, we shall formulate single step (in time) completely discrete schemes based on the above discretization in space and prove estimates analogous to those of [3] for such schemes.

In order to describe our investigations, let  $\{\lambda_j\}_1^\infty$  and  $\{\varphi_j\}_1^\infty$  be the eigenvalues (in nondecreasing order) and  $(L_2$  orthonormal) eigenfunctions of the operator  $L$  (with boundary values zero). For  $s$  real we let  $\dot{H}^s = \dot{H}^s(\Omega)$  denote the space of functions  $w \in L_2$  for which (with  $(\cdot, \cdot)$  the inner product in  $L_2(\Omega)$ )

---

Received December 23, 1976.

AMS (MOS) subject classifications (1970). Primary 65N30.

Copyright © 1977, American Mathematical Society

$$\|w\|_s = \left( \sum_j \lambda_j^s(w, \varphi_j)^2 \right)^{1/2} < \infty.$$

Recall that for  $s \geq 0$ , in order that  $w \in \dot{H}^s$ , it is required not only that  $w \in H^s(\Omega) = W_2^s(\Omega)$  but also that certain boundary conditions be satisfied (cf. [4]). For  $s = 0$  we shall normally omit the subscript and denote the norm in  $L_2 = L_2(\Omega)$  by  $\|\cdot\|$ .

Now let  $T$  be the solution operator of the elliptic problem

$$Lw = f \quad \text{in } \Omega, \quad w = 0 \quad \text{on } \partial\Omega,$$

so that

$$Tf = \sum_j \lambda_j^{-1}(f, \varphi_j)\varphi_j.$$

This operator is a bounded operator from  $\dot{H}^s$  into  $\dot{H}^{s+2}$ . In terms of  $T$  we can write the parabolic problem (1.1) as

$$D_t Tu + u = 0 \quad \text{for } 0 < t \leq t^*, \quad u(0) = v,$$

and its exact solution as

$$(1.2) \quad u(t) = \sum_j e^{-t\lambda_j}(v, \varphi_j)\varphi_j \equiv \exp(-tL)v.$$

For the purpose of discretizing this problem in space, let  $\{S_h\}$  ( $h$  small and positive) be a family of finite dimensional subspaces of  $L_2$ . As in [3] we shall assume that we are given a corresponding family of operators  $T_h: L_2 \rightarrow S_h$  with the properties:

- (i)  $T_h$  is selfadjoint, positive semidefinite on  $L_2$  and positive definite on  $S_h$ ;
- (iia) There is a positive integer  $r \geq 2$  and a constant  $C$  such that

$$\|(T_h - T)v\| \leq Ch^{q+2}\|v\|_q \quad \text{for } 0 \leq q \leq r-2,$$

or, in some cases, alternatively,

- (iib) There is a positive integer  $r > 2$  and a constant  $C$  such that

$$\|(T_h - T)v\|_{-p} \leq Ch^{p+q+2}\|v\|_q \quad \text{for } 0 \leq p, q \leq r-2.$$

We then consider the semidiscrete problem: Find  $u_h(t) \in S_h$  such that

$$D_t T_h u_h(t) + u_h(t) = 0 \quad \text{for } 0 < t \leq t^*,$$

$$u_h(0) = P_0 v \in S_h,$$

where  $P_0$  denotes the projection onto  $S_h$  in  $L_2(\Omega)$ . Introducing the inverse  $L_h = T_h^{-1}$  of  $T_h$  on  $S_h$ , this can also be written

$$D_t u_h(t) + L_h u_h(t) = 0 \quad \text{for } 0 < t \leq t^*, \quad u_h(0) = P_0 v.$$

In analogy to (1.2), its solution admits the representation

$$(1.3) \quad u_h(t) = \sum_j e^{-t\Lambda_j}(v, \Phi_j)\Phi_j = \exp(-tL_h)P_0 v,$$

where  $\{\Lambda_j\}$  and  $\{\Phi_j\}$  are the eigenvalues and eigenfunctions of  $L_h$ . Notice that the

$\Lambda_j$  are the inverses of the positive eigenvalues of  $T_h$ .

An important feature of this formulation is that it encompasses several different procedures for dealing with the homogeneous boundary condition (cf. [3]).

In [3], assuming (i) and (ia) above, it was first proved, by the energy method, that

$$(1.4) \quad \|u_h(t) - u(t)\| \leq Ch^s \|v\|_s, \quad 0 \leq s \leq r,$$

uniformly on  $(0, t^*]$ . It was then shown by means of spectral representations that

$$(1.5) \quad \|u_h(t) - u(t)\| \leq Ch^r t^{-r/2} \|v\|,$$

so that the error is in fact of optimal order in  $h$  for  $t$  bounded away from zero, even with  $v$  only in  $L_2$ . After that, estimates also in maximum norm were derived provided such estimates were known for the stationary problem. These results thus extended to the present level of generality results known before in special cases (cf. the references in [3]). Further, particular estimates were obtained for the case in which  $u_h$  satisfies the interior equation

$$(D_t u_h, \chi) + A(u_h, \chi) = 0, \quad \forall \chi \in S_h^0(\Omega_0),$$

where  $S_h^0(\Omega_0)$  denotes the elements of  $S_h$  with supports in  $\Omega_0$  and

$$(1.6) \quad A(v, w) = \int_{\Omega} \left( \sum_{j,k} a_{jk} \frac{\partial v}{\partial x_j} \frac{\partial w}{\partial x_k} + a_0 v w \right) dx,$$

and where the elements of  $S_h$  are assumed to satisfy a condition of translation invariance on  $\Omega_0$ . Specifically, it was shown that arbitrary spatial derivatives of  $u$  can be approximated to optimal order in the interior of  $\Omega_0$  by difference quotients of  $u_h$  and finally that by applying the averaging process of Bramble and Schatz [2], superconvergent  $O(h^{2r-2})$  approximations can be defined provided the assumption (iib) holds.

We now turn to the purpose of the present paper, the construction and analysis of single step arbitrary order completely discrete schemes for the parabolic problem (1.1).

The discrete approximations  $U^n$ ,  $n = 0, 1, \dots$ , to the exact solution  $u(t)$  of the parabolic problem at  $t = nk$ , with  $k$  the step size in time, will be defined in terms of a rational approximation  $r(\tau)$  to the exponential  $e^{-\tau}$  by the recursion formula

$$(1.7) \quad \begin{aligned} U^{n+1} &= r(kL_h)U^n, \quad n = 0, 1, \dots, nk \leq t^*, \\ U^0 &= P_0 v. \end{aligned}$$

The rational functions employed will be assumed to satisfy, for some  $\nu \geq 1$ ,

$$(1.8) \quad r(\tau) = e^{-\tau} + O(\tau^{\nu+1}) \quad \text{as } \tau \rightarrow 0,$$

which will signify that the accuracy in the time discretization is of order  $\nu$ . In addition they will be classified according to their behavior on the positive real line. In particular,  $r(k\lambda)$  will have no pole for  $\lambda$  an eigenvalue of  $L_h$ , so that (1.7) defines  $U^n$  uniquely as an element of  $S_h$ .

Since  $T_h$  rather than  $L_h$  is considered given, it is sometimes convenient to express  $r(kL_h)$  as a rational function of  $T_h$ , or equivalently, to consider the rational function  $r(k/\mu)$  of  $\mu$ . Assuming that this function has the form

$$r(k/\mu) = \alpha_0 \prod_j (\mu - \beta_j) / \prod_j (\mu - \gamma_j),$$

the recursion formula (1.7) for  $U^n$  may be written

$$\prod_j (T_h - \gamma_j) U^{n+1} = \alpha_0 \prod_j (T_h - \beta_j) U^n.$$

Hence, in order to determine  $U^{n+1}$  from  $U^n$ , with  $T_h$  given, it is sufficient to solve a sequence of equations of the form

$$(1.9) \quad (\alpha T_h + \beta)W = (\gamma T_h + \delta)V,$$

for  $W$ , with  $V$  given. Here, although  $U^n$  and  $U^{n+1}$  are both real,  $\alpha, \beta, \gamma, \delta$  and hence also  $V$  and  $W$  may be complex valued, with  $T_h f$  defined for complex  $f$  by linearity. In applications,  $T_h$  is often defined by means of a positive definite bilinear form (cf. [3, Section 8]),

$$B_h(T_h f, \chi) = (f, \chi), \quad \forall \chi \in S_h.$$

In this case, (1.9) takes the form

$$\alpha(W, \chi) + \beta B_h(W, \chi) = \gamma(V, \chi) + \delta B_h(V, \chi), \quad \forall \chi \in S_h.$$

We will return below to convenient choices of such rational functions.

Under our different assumptions on the time discretization schemes, to which we shall return presently, we shall derive error estimates corresponding to the ones quoted above for the semidiscrete situation. The analysis relies on the fact that similarly to  $u_h(t)$ ,  $U^n$  can be represented in terms of the discrete eigenfunctions and eigenvalues as

$$(1.10) \quad U^n = r(kL_h)^n P_0 v = \sum_j r(k\Lambda_j)^n (v, \Phi_j) \Phi_j.$$

Notice in particular that the scheme (1.7) will be stable in  $L_2$  if  $\max_j |r(k\Lambda_j)| \leq 1$ ; we then immediately find by Parseval's relation that

$$(1.11) \quad \|U^n\| \leq \|P_0 v\| \leq \|v\|.$$

This condition will be satisfied for all schemes employed below.

We shall now describe our estimates. In Section 2, we derive such estimates in  $L_2$ . First we prove the analogue of (1.5) which reads

$$(1.12) \quad \|U^n - u(nk)\| \leq C(h^r t^{-r/2} + k^\nu t^{-\nu}) \|v\|, \quad 0 < t = nk \leq t^*.$$

The proof of this estimate is based on the eigenfunction expansions (1.2) and (1.3). After that we derive an analogue of (1.4) for smooth data,

$$(1.13) \quad \|U^n - u(nk)\| \leq C(h^r \|v\|_r + k^\nu \|v\|_{2\nu}).$$

In contrast to the semidiscrete case this latter estimate is now a consequence of the estimate for nonsmooth data and the following representation of smooth initial data, namely

$$v = \sum_{j=0}^{\nu-1} T_h^j (T - T_h) L^{j+1} v + T_h^\nu L^\nu v.$$

Section 2 also contains estimates for less smooth data which are uniform for small  $t$  but not of the highest order  $O(h^r + k^\nu)$ , as well as estimates for approximations to time derivatives.

In Section 3 we derive the analogue in the completely discrete case to the maximum norm estimates of [3]. Section 4 is devoted to interior estimates on regular elements for difference quotients. It is shown here how arbitrary derivatives of  $u$  can be approximated to optimal order  $O(h^r + k^\nu)$  by difference quotients of  $U^n$ , and also that (for  $T_h$  satisfying (iib)) the effect of the Bramble-Schatz averaging operator is not spoiled by time discretization so that  $O(h^{2r-2} + k^\nu)$  approximations can also be found.

After giving examples, in Section 5, of specific rational functions satisfying our assumptions, we conclude the paper by deriving in Section 6,  $L_2$  error estimates for schemes using successively a finite number of different choices of rational functions, time steps and approximating function spaces, and also for schemes using initial data other than  $P_0 v$ .

We now turn to the classification of the discretization in time. First, the rational approximations to  $e^{-\tau}$  are said to be of types I, II, III or IV, respectively, if

- I:  $|\mathcal{r}(\tau)| < 1$  for  $0 < \tau < \alpha$ , for some  $\alpha > 0$ ;
- II:  $|\mathcal{r}(\tau)| < 1$  for  $\tau > 0$ ;
- III:  $\sup_{\tau > \delta} |\mathcal{r}(\tau)| < 1$  for all  $\delta > 0$ ;
- IV:  $|\mathcal{r}(\tau)| < 1$  for  $\tau > 0$  and  $\lim_{\tau \rightarrow \infty} \mathcal{r}(\tau) = 0$ .

Notice that these conditions are successively more restrictive.

The discretization scheme (1.7) is now classified by these conditions on  $\mathcal{r}(\tau)$ , assuming always the conditions (i), (iia) placed previously on  $T_h$  (when (iib) is needed this will be explicitly stated). In certain cases, we shall need to impose, in addition, restrictions on the relation between the parameters  $h$  and  $k$  occurring in (1.7). More precisely, we shall say that the scheme is of type I', II' or III' if, with  $\Lambda_{\max}$  the maximal eigenvalue of  $L_h$ ,

- I':  $\mathcal{r}(\tau)$  is of type I and  $k\Lambda_{\max} \leq \alpha_0$  for some  $\alpha_0$ ,  $0 < \alpha_0 < \alpha$ ;
- II':  $\mathcal{r}(\tau)$  is of type II and  $k\Lambda_{\max} \leq \alpha_1$  for some  $\alpha_1$ ,  $0 < \alpha_1 < \infty$ ;
- III':  $\mathcal{r}(\tau)$  is of type III and  $k^s \Lambda_{\max} \leq \alpha_s$  for some  $\alpha_s$ ,  $0 < \alpha_s < \infty$ , and  $s \geq 1$ .

A scheme of type III or IV will simply be one for which  $\mathcal{r}(\tau)$  is of type III or IV, respectively, with no restriction on the relation between  $k$  and  $\Lambda_{\max}$ .

Notice that for schemes of types I' and II', setting  $\tau_0 = \alpha_0$  and  $\alpha_1$ , respectively, we have  $0 < k\lambda \leq \tau_0$  for  $\lambda$  in  $\sigma_h$ , the spectrum of  $L_h$ , and hence  $|\mathcal{r}(k\lambda)| < 1$  for  $\lambda \in \sigma_h$  so that, in fact,

$$|r(k\lambda)| \leq e^{-ck\lambda} \quad \text{for } \lambda \in \sigma_h, \text{ with } c > 0.$$

This fact will be used repeatedly in the proofs.

In order that the condition  $k\Lambda_{\max} \leq \tau_0$  be satisfied it is sufficient that

$$(1.14) \quad \Lambda_{\max} \leq \kappa_0 h^{-2},$$

and that the mesh ratio condition  $k/h^2 \leq \tau_0/\kappa_0$  hold. For schemes of type I' this is analogous to the mesh ratio restrictions necessary for explicit difference schemes whereas for schemes of type II' this mesh ratio is only required to be bounded above. For the standard Galerkin method, for example, (1.14) would follow from an inverse assumption

$$\|\chi\|_1 \leq Ch^{-1} \|\chi\|_0, \quad \forall \chi \in S_h.$$

The condition for schemes of type III' would be satisfied if  $\Lambda_{\max} \leq \kappa_0 h^{-\sigma_0}$  holds and  $k/h^{\sigma_1}$  is bounded, with  $\sigma_0$  and  $\sigma_1$  arbitrary positive ( $s = \sigma_0/\sigma_1$ ).

Of the results quoted above for nonsmooth data, the  $L_2$  estimate (1.12) is proved for schemes of types I', II' and III, the maximum-norm estimates for schemes of types I', II', III' and IV, and the interior estimates for difference quotients for types I', II' and IV. The estimate (1.13) for smooth data holds for schemes of types I' and II.

Particular fully discrete Galerkin methods for parabolic problems have been treated in the literature by various authors. For instance, using energy methods and assuming a smooth solution, Douglas and Dupont [6], Dupont [7] and Wheeler [14] considered the Crank-Nicolson scheme, Zlámal [15] the so-called Calahan scheme (see below) and Crouzeix [5] schemes based on implicit Runge-Kutta methods for ordinary differential equations. Further, a theory for schemes employing general rational functions, allowing also subspaces with elements not necessarily satisfying the boundary conditions, was developed in Bramble and Thomée [4]. Except for the case of rational functions with numerators and denominators at most linear, these schemes are different from the ones studied here. The case of nonsmooth data was considered in Blair [1] for the simplest backward difference method in time and in Thomée [12], where classifications of the above types and techniques similar to the ones of the present paper were employed in one space dimension. The third author wishes to point out that the proof of the main result concerning completely discrete schemes with nonsmooth data in [12] (Theorem 4.2) is incorrect, in that the analysis relies on an eigenfunction expansion of the form (1.10) and hence pertains to the appropriate special case of the presently studied scheme (1.7), rather than to the scheme proposed in [4].

Let us emphasize again that our formulation and analysis here only requires the availability of an approximate solution operator  $T_h$  for the elliptic problem, with the above properties, and that no further smoothness and other requirements are necessary for the subspaces  $S_h$ . This is in contrast to the investigations in Bramble and Thomée [4] where higher accuracy in the time discretization necessitated higher smoothness of the approximating functions.

Specific examples of rational functions of types I, II and IV are provided by the

above diagonal, diagonal and below diagonal entries of the Padé table for  $e^{-\tau}$ . Other examples of completely discrete schemes and their computational implementation are described in Section 5 below. In particular we consider some classes of schemes of types III and IV where the rational functions are defined by taking appropriate truncations of a series expansion of the form

$$e^{-\tau} = 1 - \sum_{j=0}^{\infty} P_j(b) \left( \frac{\tau}{1+b\tau} \right)^{j+1} \quad \text{with } b > 0,$$

where the coefficients  $P_j$  are real polynomials associated with the Laguerre polynomials of the first order. For given  $\nu$ , the rational function thus derived will have denominator  $(1+b\tau)^{\nu-1}$ . This, as follows from the discussion following the introduction of the scheme (1.7), means that computationally,  $\nu-1$  systems, all with the same matrix will have to be solved at each time step. Such a property should be desirable from a practical point of view. These rational functions were studied and used for stiff systems of ordinary differential equations by Nørsett [11]; the particular case  $\nu=3$  was applied by Zlámal [15] to parabolic problems and referred to as the Calahan scheme (cf. also Makinson [9]).

Throughout this paper,  $C$  and  $c$  will denote positive constants, not necessarily the same at different occurrences.

**2. Error Estimates in  $L_2$ .** Our first result in this section is an  $L_2$  estimate for the difference between the semidiscrete and completely discrete solutions, with initial data only assumed to belong to  $L_2(\Omega)$ . Together with the estimates from [3] for the error in the semidiscrete solution, this results in a complete error estimate for  $U^n$ . This estimate has optimal order of accuracy for  $t$  bounded away from zero, but contains a negative power of  $t$  on the right.

We now state this first result.

**THEOREM 2.1.** *Let the discretization scheme be of type I', II' or III. Then for  $0 < t = nk \leq t^*$ ,*

$$\|U^n - u_h(nk)\| \leq Ck^\nu t^{-\nu} \|v\|.$$

Before giving the proof of this result we shall introduce some notation. For an arbitrary function  $g(\tau)$  defined for  $\tau \in \sigma_h$ , the spectrum of  $L_h = T_h^{-1}$ , we set for  $v \in L_2(\Omega)$ ,

$$(2.1) \quad g(L_h)v = \sum_{\Lambda_j \in \sigma_h} g(\Lambda_j)(v, \Phi_j)\Phi_j = g(L_h)P_0v.$$

In particular, with

$$(2.2) \quad F_n(\tau) = r(\tau)^n - e^{-n\tau},$$

we can write

$$(2.3) \quad U^n - u_h(nk) = F_n(kL_h)v.$$

By (2.1),  $L_h$  is defined on all of  $L_2$  and not just as the inverse of  $T_h$  on  $S_h$ . By Parseval's relation we have for the operator norm of  $g(L_h)$  on  $L_2$ ,

$$(2.4) \quad \|g(L_h)\| = \sup_{\tau \in \sigma_h} |g(\tau)|.$$

In view of (2.3), Theorem 2.1 is a consequence of the following lemma:

LEMMA 2.1. *Let the discretization scheme be of type I', II' or III. Then for  $0 < t = nk \leq t^*$ ,*

$$\|F_n(kL_h)\| \leq Ck^\nu t^{-\nu}.$$

*Proof.* We have by (2.4),

$$\|F_n(kL_h)\| = \sup_{\lambda \in \sigma_h} |F_n(k\lambda)| = \sup_{\tau/k \in \sigma_h} |F_n(\tau)|,$$

and hence it suffices to show that

$$|F_n(\tau)| \leq C(k/t)^\nu = Cn^{-\nu} \quad \text{for } \tau/k \in \sigma_h.$$

Let  $\tau_0$  be a positive number such that

$$(2.5) \quad |r(\tau)| < 1 \quad \text{for } 0 < \tau \leq \tau_0.$$

Then for  $0 < \tau \leq \tau_0$  by (1.8),

$$|r(\tau) - e^{-\tau}| \leq C\tau^{\nu+1},$$

and

$$(2.6) \quad |r(\tau)| \leq e^{-c\tau} \quad \text{with } 0 < c < 1.$$

Hence

$$(2.7) \quad |F_n(\tau)| = \left| (r(\tau) - e^{-\tau}) \sum_{l=0}^{n-1} r(\tau)^{n-1-l} e^{-l\tau} \right| \\ \leq Cn\tau^{\nu+1} e^{-c(n-1)\tau} \leq C\tau^\nu e^{-cn\tau} \leq Cn^{-\nu}.$$

Since in cases I' and II',  $k\lambda \leq \tau_0$  for all  $\lambda \in \sigma_h$  if  $\tau_0$  is appropriately chosen, this completes the proof for such schemes. For type III schemes we also have to consider  $\tau$  large,  $\tau \geq 1$ , say. We then have

$$e^{-n\tau} \leq e^{-n} \leq Cn^{-\nu}.$$

Further,

$$(2.8) \quad \sup_{\tau \geq 1} |r(\tau)| = e^{-c} \quad \text{with } c > 0,$$

so that similarly

$$|r(\tau)^n| \leq e^{-cn} \leq Cn^{-\nu} \quad \text{for } \tau \geq 1.$$

This completes the proof.

Combining Theorem 2.1 with the case  $j = 0$  of Theorem 3.2 in [3], which states that

$$(2.9) \quad \|u_h(t) - u(t)\| \leq Ch^\nu t^{-\nu/2} \|v\| \quad \text{for } 0 < t \leq t^*,$$



we conclude immediately:

**THEOREM 2.2.** *Let the discretization scheme be of type I', II' or III. Then for  $0 < t = nk \leq t^*$ ,*

$$\|U^n - u(nk)\| \leq C\{h^r t^{-r/2} + k^\nu t^{-\nu}\} \|v\|.$$

We shall now turn to estimates which hold uniformly down to  $t = 0$ . In this case, in order to obtain optimal order results, smoothness has to be required of the initial data. In Theorem 2.1 of [3], before proving the estimate (2.9), the energy method was used to derive such an error estimate for the semidiscrete solution; in particular

$$(2.10) \quad \|u_h(t) - u(t)\| \leq Ch^s \|u\|_s \quad \text{for } 0 \leq s \leq r, 0 \leq t \leq t^*.$$

In order to obtain now a similar result for the completely discrete situation we shall combine Theorem 2.1 with the following representation of the initial data, namely

$$(2.11) \quad v = \sum_{j=0}^{\nu-1} T_h^j (T - T_h) L^{j+1} v + T_h^\nu L^\nu v.$$

The following lemma will be needed.

**LEMMA 2.2.** *Let the discretization scheme be of type I' or II. Then for  $0 \leq j \leq \nu$*

$$\|T_h^j F_n(kL_h)\| \leq Ck^j.$$

*Proof.* By (2.4),

$$\|T_h^j F_n(kL_h)\| = \sup_{\lambda \in \sigma_h} |\lambda^{-j} F_n(k\lambda)| = \sup_{\tau/k \in \sigma_h} |k^j \tau^{-j} F_n(\tau)|,$$

and hence it suffices to show that

$$|\tau^{-j} F_n(\tau)| \leq C \quad \text{for } \tau/k \in \sigma_h.$$

As in the proof of Lemma 2.1, let  $\tau_0$  be a positive number such that  $|\tau(\tau)| < 1$  for  $0 < \tau \leq \tau_0$ . Then for such  $\tau$ , by (1.8)

$$|\tau(\tau) - e^{-\tau}| \leq C\tau^{j+1}, \quad 0 \leq j \leq \nu,$$

and (2.6) holds.

Hence

$$|\tau^{-j} F_n(\tau)| = \left| \tau^{-j} (\tau(\tau) - e^{-\tau}) \sum_{l=0}^{n-1} \tau(\tau)^{n-1-l} e^{-l\tau} \right| \leq Cn\tau e^{-c(n-1)\tau} \leq C.$$

For  $\tau > \tau_0$  the desired inequality follows trivially.

The result in the smooth data case is then the following:

**THEOREM 2.3.** *Let the discretization scheme be of type I' or II. Then for  $0 \leq t = nk \leq t^*$ ,*

$$\|U^n - u(nk)\| \leq C(h^r \|v\|_r + k^\nu \|v\|_{2\nu}).$$

*Proof.* We first note that if we set

$$v_k = \sum_{k\lambda_j \leq 1} (v, \varphi_j) \varphi_j$$

(recalling that  $\varphi_j$  and  $\lambda_j$  are the eigenfunctions and eigenvalues of the differential operator  $L$ ) then by the definition of the norms in  $\dot{H}^s$ ,

$$(2.12) \quad \|v - v_k\| \leq k^\nu \|v\|_{2\nu},$$

$$(2.13) \quad \|v_k\|_{2\nu} \leq \|v\|_{2\nu},$$

$$(2.14) \quad \|v_k\|_{r+2j} \leq k^{-j} \|v\|_r, \quad j = 0, 1, \dots$$

Applying the identity (2.11) to  $v_k$  and setting for brevity  $F_n = F_n(kL_h)$  as defined by (2.1) and (2.2), we may write

$$F_n v_k = \sum_{j=0}^{\nu-1} F_n T_h^j (T - T_h) L^{j+1} v_k + F_n T_h^\nu L^\nu v_k.$$

Here, by Lemma 2.2 and (2.13),

$$\begin{aligned} \|F_n T_h^\nu L^\nu v_k\| &= \|T_h^\nu F_n L^\nu v_k\| \\ &\leq C k^\nu \|L^\nu v_k\| = C k^\nu \|v_k\|_{2\nu} \leq C k^\nu \|v\|_{2\nu}. \end{aligned}$$

Further, using also (iia) and (2.14), we obtain for  $0 \leq j \leq \nu - 1$ ,

$$\begin{aligned} \|F_n T_h^j (T - T_h) L^{j+1} v_k\| &\leq C k^j \|(T - T_h) L^{j+1} v_k\| \\ &\leq C k^j h^r \|L^{j+1} v_k\|_{r-2} = C k^j h^r \|v_k\|_{r+2j} \\ &\leq C h^r \|v\|_r \leq C h^r \|v\|_r. \end{aligned}$$

Together, these estimates imply

$$\|F_n v_k\| \leq C(h^r \|v\|_r + k^\nu \|v\|_{2\nu}).$$

Since obviously, by stability (cf. (1.11)) and (2.12),

$$\|F_n(v - v_k)\| \leq 2\|v - v_k\| \leq C k^\nu \|v\|_{2\nu},$$

we conclude that

$$\|U^n - u_h(nk)\| = \|F_n v\| \leq C(h^r \|v\|_r + k^\nu \|v\|_{2\nu}).$$

By (2.10) (with  $s = r$ ) this completes the proof.

It is also possible to prove estimates which are uniform in  $t$  for less regular initial data. We have for instance:

**THEOREM 2.4.** *Let the discretization scheme be of type I' or II and let  $s \geq 0$ . Then for  $0 \leq t = nk \leq t^*$ ,*

$$\|U^n - u(nk)\| \leq C\{h^{\min(r,s)} + k^{\min(\nu, s/2)}\} \|v\|_s.$$

*Proof.* We first notice that in Theorem 2.3 and its proof we may replace  $r$  by any  $s_1$  with  $2 \leq s_1 \leq r$ . Using also the fact that the function  $v_k$  satisfies, for  $0 \leq s_2 \leq 2\nu$ ,

$$\|v - v_k\| \leq k^{s_2/2} \|v\|_{s_2}, \quad \|v_k\|_{2\nu} \leq k^{-(\nu-s_2/2)} \|v\|_{s_2},$$

we obtain

$$\|U^n - u(nk)\| \leq C\{h^{s_1}\|v\|_{s_1} + k^{s_2/2}\|v\|_{s_2}\}.$$

In particular, this proves the desired result for  $s \geq 2$ . Interpolation between the result for  $s = 2$  and the stability estimate

$$\|U^n - u(nk)\| \leq 2\|v\|,$$

now completes the proof.

We shall conclude this section by considering a finite difference approximation for a time derivative  $D_t^l u$  of the solution of the parabolic problem, of the form

$$q_k U^n = k^{-l} q(E) U^n,$$

where

$$q(E) = \sum_{j=j_0}^{j_1} q_j E^j, \quad E^j U^n = U^{n+j}.$$

Notice that this approximation will be defined only for  $n + j_0 \geq 0$ . Recall that the operator  $q_k$  is an approximation of order  $\nu$  to  $D_t^l$  if

$$(2.15) \quad q(e^{-\tau}) = \sum_{j=j_0}^{j_1} q_j e^{-j\tau} = (-\tau)^l + O(\tau^{l+\nu}) \quad \text{as } \tau \rightarrow 0.$$

Notice that then, by (1.7),

$$(2.16) \quad q(r(\tau)) = (-\tau)^l + O(\tau^{l+\nu}) \quad \text{as } \tau \rightarrow 0.$$

We then have the following optimal order error estimate for  $t$  bounded away from zero and initial data in  $L_2$ .

**THEOREM 2.5.** *Assume that  $q_k$  is an approximation of order  $\nu$  to  $D_t^l$  and that the scheme is of type I', II' or III. Then for  $0 < t = nk \leq t^*$ , with  $n + j_0 \geq 0$ ,*

$$\|q_k U^n - D_t^l u(nk)\| \leq C(h^r t^{-r/2-l} + k^\nu t^{-\nu-l})\|v\|.$$

*Proof.* By Theorem 3.2 of [3] we have

$$\|D_t^l(u_h(t) - u(t))\| \leq Ch^r t^{-r/2-l}\|v\|,$$

and so it remains to prove

$$\|q_k U^n - D_t^l u_h(nk)\| \leq Ck^\nu t^{-\nu-l}\|v\|.$$

By the spectral representation we have

$$q_k U^n - D_t^l u_h(nk) = \sum_j (k^{-l} q(r(k\Lambda_j)) r(k\Lambda_j)^n - (-\Lambda_j)^l e^{-t\Lambda_j})(v, \Phi_j) \Phi_j,$$

so that setting

$$s_n(\tau) = q(r(\tau)) r(\tau)^n - (-\tau)^l e^{-n\tau},$$

our result would follow from

$$\sup_{\lambda \in \sigma_h} |k^{-l} s_n(k\lambda)| \leq C k^\nu t^{-\nu-l}$$

or

$$(2.17) \quad |s_n(\tau)| \leq C n^{-\nu-l} \quad \text{for } \tau/k \in \sigma_h.$$

Now let  $\tau_0$  be as in the proof of Lemma 2.1 (cf. (2.5)) and write

$$s_n(\tau) = (q(r(\tau)) - (-\tau)^l)r(\tau)^n + (-\tau)^l(r(\tau)^n - e^{-n\tau}).$$

Using (2.16) and (2.6) we obtain for  $\tau \leq \tau_0$ ,

$$(2.18) \quad |(q(r(\tau)) - (-\tau)^l)r(\tau)^n| \leq C \tau^{l+\nu} e^{-cn\tau} \leq C n^{-l-\nu},$$

and by (2.7),

$$(2.19) \quad |\tau^l(r(\tau)^n - e^{-n\tau})| \leq C \tau^{\nu+l} e^{-cn\tau} \leq C n^{-l-\nu},$$

so that

$$|s_n(\tau)| \leq C n^{-l-\nu} \quad \text{for } \tau \leq \tau_0.$$

This completes the proof of (2.17) for schemes of types I' and II'. In the case of type III we also consider  $\tau \geq 1$ , say. For such  $\tau$ ,

$$\tau^l e^{-n\tau} \leq \tau^{l+\nu} e^{-n\tau} \leq C n^{-l-\nu},$$

and

$$|q(r(\tau))r(\tau)^n| \leq C e^{-cn} \leq C n^{-l-\nu}.$$

This completes the proof of (2.17) for  $\tau \geq 1$  and hence of the theorem.

**3. Error Estimates in the Maximum Norm.** In this section we shall derive maximum norm estimates, analogous to those of Section 4 of [3] for the semidiscrete problem, for our completely discrete approximations. In [3] it was shown that if certain maximum norm error estimates were available for the associated elliptic problem, then corresponding maximum norm error estimates could be derived for  $t$  bounded away from zero for the parabolic problem.

The assumption about the elliptic problem, which will also be made here, was that there is a function  $\gamma(h)$  such that

$$(i)' \quad |T_h w| \leq C |Tw|_1, \quad \|T_h w\| \leq C \|Tw\|_1,$$

$$(ii)' \quad |(T_h - T)w| \leq \gamma(h) |Tw|_r,$$

where  $|\cdot|$  and  $|\cdot|_r$  denote the norms in  $C(\bar{\Omega})$  and  $C^r(\bar{\Omega})$ , respectively. The conclusion for our semidiscrete problem was [3, Theorem 4.2] that

$$(3.1) \quad |u_h(t) - u(t)| \leq C \{\gamma(h) + h^r\} \|v\| \quad \text{for } 0 < t_* \leq t \leq t^*.$$

In special cases discussed in Section 8 of [3],  $\gamma(h) = Ch^r$  for  $r > 2$  and  $\gamma(h) = Ch^2 \log h^{-1}$  for  $r = 2$ .

To obtain our error estimate for the completely discrete problem we shall again use the spectral representation

$$(3.2) \quad U^n - u_h(nk) = F_n(kL_h)v = \sum_j F_n(k\Lambda_j)(v, \Phi_j)\Phi_j.$$

The technique will now be to derive first a maximum norm estimate for  $\varphi_j$ , and then an estimate for  $F_n(k\Lambda_j)$ .

We first have:

LEMMA 3.1. *Under the above assumptions (i)', (ii)' on  $T_h$  there exists a  $J$  depending only on  $N$  such that*

$$|\Phi_j| \leq C\Lambda_j^J.$$

*Proof.* By Lemma 4.1 of [3] we have for  $2 \leq p \leq \infty$ ,  $0 < 1/q - 1/p < 1/N$ ,

$$\|T_h w\|_{L_p} \leq C\|w\|_{L_q}.$$

In particular, then

$$\|\Phi_j\|_{L_p} = \Lambda_j \|T_h \Phi_j\|_{L_p} \leq C\Lambda_j \|\Phi_j\|_{L_q}.$$

Hence, by repeated application, using a suitable decreasing sequence  $\{p_j\}_0^J$  with  $p_0 = \infty$ ,  $p_J = 2$ ,  $1/p_{j+1} - 1/p_j < 1/N$  we obtain

$$|\Phi_j| \leq C\Lambda_j^J \|\Phi_j\| = C\Lambda_j^J,$$

which proves the lemma.

The next lemma, which we shall also use in the next section, contains the needed estimate for  $F_n(k\Lambda_j)$ .

LEMMA 3.2. *Let the scheme be of type I', II', III' or IV. Then for  $0 < t = nk \leq t^*$  and  $j \geq 0$ ,*

$$\|L_h^j F_n(kL_h)\| = \sup_{\lambda \in \sigma_h} |\lambda^j F_n(k\lambda)| \leq Ck^\nu t^{-\nu-\rho j}$$

where  $\rho = 1$  for types I', II' and IV and  $\rho = s$  for type III' (with parameter  $s$ ).

*Proof.* Let  $\tau_0$  be as in (2.5) and consider first  $\tau = k\lambda \leq \tau_0$ . We want to show that then

$$(3.3) \quad |\lambda^j F_n(k\lambda)| = |k^{-j} \tau^j F_n(\tau)| \leq Ck^\nu t^{-j-\nu}.$$

But, applying (2.7) with  $j = 0$ , we obtain at once

$$|\tau^j F_n(\tau)| \leq C\tau^{j+\nu} e^{-cn\tau} \leq Cn^{-j-\nu} = Ck^{j+\nu} t^{-j-\nu},$$

which proves (3.3). This completes the proof for schemes of types I' and II'.

For schemes of types III' and IV we also have, as usual, to consider  $\tau = k\lambda \geq 1$ , for  $\lambda \in \sigma_h$ . In both cases we have, by obvious estimates,

$$\lambda^j e^{-n\tau} = k^{-j} \tau^j e^{-n\tau} \leq Ck^{-j} n^{-j} e^{-cn\tau} \leq Ct^{-j} n^{-\nu} = Ck^\nu t^{-j-\nu}.$$

Now, in case III' we have by definition and (2.8), for  $\lambda \in \sigma_h$ ,

$$|\lambda^j r(\tau)^n| \leq Ck^{-sj} e^{-cn} \leq Ck^{-sj} n^{-\nu-sj} = Ck^\nu t^{-\nu-sj}.$$

Since  $s \geq 1$ , the above estimates complete the proof for case III'.

Finally, for type IV, the degree of the numerator of  $r(\tau)$  is less than that of the denominator and it follows that there is a  $c > 0$  such that

$$|r(\tau)| \leq \frac{1}{1 + c\tau} \quad \text{for } \tau \geq 1.$$

Hence for these  $\tau$ ,

$$(3.4) \quad \begin{aligned} |k^{-j} \tau^j r(\tau)^n| &\leq k^{-j} (\tau/(1 + c\tau))^j (1 + c\tau)^{-(n-j)} \\ &\leq Ck^{-j} (1 + c)^{-n} \leq Ck^{-j} n^{-\nu-j} = Ck^\nu t^{-\nu-j}. \end{aligned}$$

This completes the proof in case IV and hence the lemma is proved.

In order to state and prove our main result of this section we shall make the additional assumption that the eigenvalues  $\Lambda_j$  of  $L_h$  tend to infinity fast enough that for some  $J_0$ , uniformly for small  $h$ ,

$$(iii) \quad \sum_j \Lambda_j^{-J_0} \leq C < \infty.$$

For instance, if we consider the standard Galerkin method, with the elements of  $S_h$  vanishing on  $\partial\Omega$  (cf. Section 8.1 of [3]), it is well known that  $\Lambda_j \geq \lambda_j$ , and (iii) then follows from known estimates for the spectrum of the elliptic operator  $L$ .

**THEOREM 3.1.** *Assume that (i)', (ii)' and (iii) hold and let the discretization scheme be of type I', II', III' or IV. Then for  $0 < t_* \leq t = nk \leq t^*$ ,*

$$|U^n - u(nk)| \leq C(\gamma(h) + h^r + k^\nu) \|v\|.$$

*Proof.* By (3.1) it suffices to show that for  $t = nk$  bounded away from zero and infinity,

$$(3.5) \quad |U^n - u_h(nk)| \leq Ck^\nu \|v\|.$$

But by (3.2) and Lemma 3.1,

$$(3.6) \quad |U^n - u_h(nk)| \leq \left( \sum_j |F_n(k\Lambda_j)| |\Phi_j| \right) \|v\| \leq C \left( \sum_j \Lambda_j^J |F_n(k\Lambda_j)| \right) \|v\|.$$

In view of (iii) it therefore suffices to notice that by Lemma 3.2,

$$|F_n(k\Lambda_j)| \leq Ck^\nu t^{-\nu-\rho(J+J_0)} \Lambda_j^{-(J+J_0)}.$$

This completes the proof of the lemma.

Notice that we actually have the somewhat stronger statement than (3.5), namely

$$|U^n - u_h(nk)| \leq Ck^\nu \|v\|_{L_1(\Omega)}.$$

For, instead of (3.6) we can similarly derive

$$\begin{aligned} |U^n - u_h(nk)| &\leq \left( \sum_j |F_n(k\Lambda_j)| |\Phi_j|^2 \right) \|v\|_{L_1(\Omega)} \\ &\leq C \left( \sum_j \Lambda_j^{2J} |F_n(k\Lambda_j)| \right) \|v\|_{L_1(\Omega)} \leq Ck^\nu t^{-\nu-\rho(J_0+2J)} \|v\|_{L_1(\Omega)}. \end{aligned}$$

**4. Interior Estimates for Difference Quotients.** In this section we shall derive  $L_2$  error estimates for difference quotients of  $U^n - u_h(nk)$  on some interior subdomain  $\Omega_0$  of  $\Omega$ , under certain additional assumptions on  $S_h$  and  $T_h$ . We shall then apply these estimates to extend to the completely discrete case the study of Sections 6 and 7 of [3] concerning approximation in the maximum norm of derivatives of the solution of the continuous problem, including approximations of superconvergent order  $O(h^{2r-2})$ .

Here, in addition to the norms in  $\dot{H}^s(\Omega)$  we shall use for  $s \geq 0$  the norm  $\|\cdot\|_{s,\Omega_0}$  on  $H^s(\Omega_0) = W_2^s(\Omega_0)$  (denoted  $\|\cdot\|_{\Omega_0}$  when  $s = 0$ ) and the norm of negative order  $-s$ ,

$$\|w\|_{-s,\Omega_0} = \sup_{\varphi \in C_0^\infty(\Omega_0)} \frac{(w, \varphi)}{\|\varphi\|_{s,\Omega_0}}.$$

Further,  $|\cdot|_{\Omega_0}$  and  $|\cdot|_{s,\Omega_0}$  will be used to denote the norms in  $C(\bar{\Omega}_0)$  and  $C^s(\bar{\Omega}_0)$ , respectively.

The first of our additional assumptions is that  $T_h$  satisfies the “interior” equation

$$(4.1) \quad A(T_h f, \chi) = (f, \chi), \quad \forall \chi \in S_h^0(\Omega_0),$$

where  $S_h^0(\Omega_0)$  is the set of functions in  $S_h$  with supports in  $\Omega_0$  and  $A(\cdot, \cdot)$  is the bilinear form (1.6). This implies that the solution  $u_h$  of the semidiscrete problem satisfies the interior parabolic Galerkin equation

$$(D_t u_h, \chi) + A(u_h, \chi) = 0, \quad \forall \chi \in S_h^0(\Omega_0), \quad t > 0.$$

We shall further require that on  $\Omega_0$  the functions in  $S_h$  be piecewise polynomials on a uniform partition. The precise conditions needed, which will be referred to as  $r$ -regularity of  $\{S_h\}$  on  $\Omega_0$ , are stated in Section 6 of [3], and rather than restating them here, we shall quote only the consequences needed for our present purposes.

In particular, setting

$$\partial_h^\alpha = \partial_{h,1}^{\alpha_1} \cdots \partial_{h,N}^{\alpha_N} \quad \text{with } \partial_{h,j} w(x) = h^{-1}(w(x + he_j) - w(x)),$$

where  $e_j$  is the unit vector in the direction of  $x_j$ , it follows under these assumptions from Lemma 6.4 and Theorem 3.2 of [3] that for  $\Omega_1 \subset\subset \Omega_0$  and for  $0 < t_* \leq t \leq t^*$ ,

$$(4.2) \quad \|\partial_h^\alpha D_t^j(u_h(t) - u(t))\|_{\Omega_1} \leq Ch^r \|v\|.$$

Using also a discrete Sobolev inequality this could be seen to imply (cf. [3, Lemma 6.5]) that for  $\Omega_1 \subset\subset \Omega_0$  and  $0 < t_* \leq t \leq t^*$ ,

$$(4.3) \quad |\partial_h^\alpha D_t^j(u_h(t) - u(t))|_{\Omega_1} \leq Ch^r \|v\|.$$

Similarly, using Lemma 7.3 and Theorems 3.1 and 3.2 of [3] we find for  $\Omega_1 \subset\subset \Omega_0$ ,  $0 < t_* \leq t \leq t^*$ ,

$$(4.4) \quad \|\partial_h^\alpha D_t^j(u_h(t) - u(t))\|_{-(r-2),\Omega_1} \leq Ch^{2r-2} \|v\|_{r-2}.$$

Such estimates were the basis for deriving the maximum norm and superconvergent order error estimates in [3] referred to above. These estimates will be stated more precisely below.

In the technical work we shall need the following special case of Theorem 5.2, (5.6) in Nitsche and Schatz [10].

LEMMA 4.1. *Let  $\{S_h\}$  be  $r$ -regular on  $\Omega_0 \subset \Omega$  and let  $\Omega_2 \subset\subset \Omega_1 \subset\subset \Omega_0$ . Assume that  $v_h \in S_h$  satisfies*

$$A(v_h, \chi) = f_h(\chi), \quad \forall \chi \in S_h^0(\Omega_0),$$

where  $f_h$  is a linear functional on  $H_0^1(\Omega_1)$ . Then

$$\|v_h\|_{1, \Omega_2} \leq C(\|f_h\|_{-1, \Omega_1} + \|v_h\|_{\Omega_1}),$$

where

$$\|f_h\|_{-1, \Omega_1} = \sup_{\varphi \in C_0^\infty(\Omega_1)} \frac{|f_h(\varphi)|}{\|\varphi\|_1}.$$

The subsequent proofs of the present analogues of (4.2) and (4.3) depend on the following lemma.

LEMMA 4.2. *Assume that  $\{S_h\}$  is  $r$ -regular in  $\Omega_0 \subset \Omega$  and that  $T_h$  satisfies the interior equation (4.1). Then for  $|\alpha| \leq 2m$ ,  $\Omega_1 \subset\subset \Omega_0$ ,*

$$(4.5) \quad \|\partial_h^\alpha \chi\|_{\Omega_1} \leq C\|L_h^m \chi\|, \quad \forall \chi \in S_h.$$

*Proof.* As a first step we shall show that if  $\Omega_2 \subset\subset \Omega_1 \subset\subset \Omega_0$  and  $n \geq 0$ , then for  $|\alpha| \leq n + 1$ ,

$$(4.6) \quad \|\partial_h^\alpha T_h f\|_{1, \Omega_2} \leq C \left\{ \sum_{|\beta| \leq n} \|\partial_h^\beta f\|_{\Omega_1} + \|T_h f\|_{\Omega_1} \right\}.$$

Set  $v_h = T_h f$  and let  $\Omega'_0 \subset\subset \Omega_0$ . We have for  $\chi \in S_h^0(\Omega'_0)$  and small  $h$ ,

$$\begin{aligned} A(\partial_h^\alpha v_h, \chi) &= \int_{\Omega} \left( \sum_{j,k} a_{jk} \frac{\partial}{\partial x_j} (\partial_h^\alpha v_h) \frac{\partial}{\partial x_k} \chi + c \partial_h^\alpha v_h \chi \right) dx \\ (4.7) \quad &= \int_{\Omega} \left( \sum_{j,k} \partial_h^\alpha \left( a_{jk} \frac{\partial}{\partial x_j} v_h \right) \frac{\partial}{\partial x_k} \chi + \partial_h^\alpha (c v_h) \cdot \chi \right) dx + \tilde{f}_h(\chi) \\ &= (-1)^{|\alpha|} A(v_h, \partial_h^\alpha \chi) + \tilde{f}_h(\chi), \end{aligned}$$

where for  $\varphi \in C_0^\infty(\Omega'_1)$ , with  $\Omega'_1 \subset\subset \Omega_1$  (cf. [10, (6.7)]),

$$(4.8) \quad |\tilde{f}_h(\varphi)| \leq C \sum_{|\beta| < |\alpha|} \|\partial_h^\beta v_h\|_{1, \Omega_1} \cdot \|\varphi\|_{1, \Omega'_1}.$$

(For the case of constant coefficients,  $\tilde{f}_h = 0$ .) Since  $\partial_h^\alpha \chi \in S_h^0(\Omega_0)$  for  $\chi \in S_h^0(\Omega'_0)$  and small  $h$ , we have

$$(4.9) \quad A(v_h, \partial_h^\alpha \chi) = (f, \partial_h^\alpha \chi),$$

and since  $|\alpha| \leq n + 1$ , for  $\varphi \in C_0^\infty(\Omega'_1)$ ,



$$(4.10) \quad |(f, \partial_h^\alpha \varphi)| \leq C \sum_{|\beta| \leq n} \|\partial_h^\beta f\|_{\Omega_1} \|\varphi\|_{1, \Omega_1'}.$$

Together, (4.7), (4.8), (4.9) and (4.10) show by Lemma 4.1,

$$\|\partial_h^\alpha v_h\|_{1, \Omega_2} \leq C \left\{ \sum_{|\beta| \leq n} \|\partial_h^\beta f\|_{\Omega_1} + \sum_{|\beta| < |\alpha|} \|\partial_h^\beta v_h\|_{1, \Omega_1} \right\}.$$

Hence by repeated application, for any  $\Omega_2 \subset \subset \Omega_1 \subset \subset \Omega_0$ ,

$$\|\partial_h^\alpha v_h\|_{1, \Omega_2} \leq C \left\{ \sum_{|\beta| \leq n} \|\partial_h^\beta f\|_{\Omega_1} + \|v_h\|_{1, \Omega_1} \right\}.$$

Using Lemma 4.1 once more to estimate the last term we conclude for  $\Omega_2 \subset \subset \Omega_1$ ,

$$\|\partial_h^\alpha v_h\|_{1, \Omega_2} \leq C \left\{ \sum_{|\beta| \leq n} \|\partial_h^\beta f\|_{\Omega_1} + \|v_h\|_{\Omega_1} \right\},$$

which is (4.6).

As a consequence of (4.6), we have also, now for  $|\alpha| \leq n + 2$ , again with  $\Omega_2 \subset \subset \Omega_1 \subset \subset \Omega_0$ ,

$$(4.11) \quad \|\partial_h^\alpha T_h f\|_{\Omega_2} \leq C \left\{ \sum_{|\beta| \leq n} \|\partial_h^\beta f\|_{\Omega_1} + \|T_h f\|_{\Omega_1} \right\}.$$

We are now in a position to prove (4.5) by induction over  $m$ . The statement being obvious for  $m = 0$  we consider the step from  $m$  to  $m + 1$ . For  $|\alpha| \leq 2m + 2$  we then obtain by (4.11) for  $\chi \in S_h$ ,

$$\|\partial_h^\alpha \chi\|_{\Omega_1} = \|\partial_h^\alpha T_h L_h \chi\|_{\Omega_1} \leq C \left\{ \sum_{|\beta| \leq 2m} \|\partial_h^\beta L_h \chi\|_{\Omega_2} + \|\chi\|_{\Omega_2} \right\},$$

and hence by the induction assumption and the boundedness of  $T_h$ ,

$$\|\partial_h^\alpha \chi\|_{\Omega_1} \leq C \{ \|L_h^m L_h \chi\| + \|\chi\| \} \leq C \|L_h^{m+1} \chi\|,$$

which completes the proof.

We can now state and prove the present analogues of (4.2) and (4.3).

**THEOREM 4.1.** *Assume that  $\{S_h\}$  is  $r$ -regular in  $\Omega_0 \subset \Omega$  and that (4.1) holds. Let  $q_k = k^{-1}q(E)$  approximate  $D_t^l$  with accuracy  $\nu$  in the sense of (2.15). Then for for completely discrete schemes of types I', II' and IV, we have for any  $\alpha$ ,  $\Omega_1 \subset \subset \Omega_0$  and  $0 < t = nk \leq t^*$ ,*

$$(4.12) \quad \|\partial_h^\alpha (q_k U^n - D_t^l u_h(nk))\|_{\Omega_1} \leq C k^\nu t^{-\nu-l-|\alpha|/2} \|v\|,$$

and with  $N_0 = [N/2] + 1$ ,

$$(4.13) \quad |\partial_h^\alpha (q_k U^n - D_t^l u(nk))|_{\Omega_1} \leq C k^\nu t^{-\nu-l-(|\alpha|+N_0)/2} \|v\|.$$

*Proof.* We first show (4.12) for  $\alpha$  of even order,  $|\alpha| = 2m$ . We obtain by Lemma 4.2 and (2.4),

$$\begin{aligned}
\|\partial_h^\alpha(q_k U^n - D_t^l u_h(nk))\|_{\Omega_1} &\leq C \|L_h^m(q_k U^n - D_t^l u_h(nk))\| \\
&= \sup_{\lambda \in \sigma_h} |\lambda^m (k^{-l} q(r(k\lambda)) r(k\lambda)^n - (-\lambda)^l e^{-nk\lambda})| \|v\| \\
&= C k^{-l-m} \sup_{\tau/k \in \sigma_h} |\tau^m s_n(\tau)| \|v\|,
\end{aligned}$$

where as in the proof of Theorem 2.5,

$$s_n(\tau) = q(r(\tau)) r(\tau)^n - (-\tau)^l e^{-n\tau}.$$

It therefore suffices to show that

$$(4.14) \quad |\tau^m s_n(\tau)| \leq C n^{-\nu-l-m} \quad \text{for } \tau/k \in \sigma_h.$$

Now let  $\tau_0$  be as in (2.5). Then by (2.18) and (2.19),

$$|\tau^m s_n(\tau)| \leq C \tau^{m+l+\nu} e^{-cn\tau} \leq C n^{-l-\nu-m} \quad \text{for } \tau \leq \tau_0.$$

This completes the proof of (4.14) for schemes of types I' and II'. For schemes of type IV we also need to consider  $\tau \geq 1$ , say. For such  $\tau$ ,

$$\tau^{l+m} e^{-n\tau} \leq \tau^{l+\nu+m} e^{-n\tau} \leq C n^{-l-\nu-m},$$

and, appealing now to (3.4),

$$|\tau^{l+m} q(r(\tau)) r(\tau)^n| \leq C |\tau^{l+m} r(\tau)^{n-j_0}| \leq C n^{-\nu-l-m},$$

which implies (4.14). This concludes the proof of (4.12) for  $|\alpha|$  even. Notice that it is at the last step that we have to use type IV rather than type III, as was the case in Theorem 2.5.

In order to show (4.12) for  $|\alpha|$  odd, we choose  $j$  such that  $\alpha_j > 0$  and apply summation by parts to obtain the inequality

$$\|\partial_h^\alpha \chi\|_{\Omega_1} \leq \|\partial_h^{\alpha-e} j \chi\|_{\Omega_2}^{1/2} \|\partial_h^{\alpha+e} j \chi\|_{\Omega_2}^{1/2}, \quad \Omega_1 \subset \subset \Omega_2 \subset \subset \Omega_0.$$

With  $\chi = q_k U^n - D_t^l u_h(nk)$  the result for  $|\alpha|$  even hence implies the result for  $|\alpha|$  odd. This concludes the proof of (4.12).

The maximum norm estimate (4.13) now follows at once by the Sobolev inequality

$$|\chi|_{\Omega_2} \leq C \sum_{|\beta| \leq N_0} \|\partial_h^\beta \chi\|_{\Omega_1} \quad \text{for } \chi \in S_h, \Omega_2 \subset \subset \Omega_1,$$

since  $q_k U^n - D_t^l u_h(nk) \in S_h$ . This completes the proof of the theorem.

We now turn to the maximum norm approximation of derivatives of the solution  $u(t)$ . We recall that the finite difference operator

$$Q_h w(x) = \sum_{\beta, \eta} q_{\beta\eta} \partial_h^\beta w(x - \eta h)$$

approximates the derivative  $D^\alpha w$  of order  $|\alpha| = m$  with accuracy  $r$  if for  $\Omega_2 \subset \subset \Omega_1$ ,

$$|Q_h w - D^\alpha w|_{\Omega_2} \leq Ch^r |w|_{r+m, \Omega_1}.$$

Combining this for  $w = D_t^l u(t)$  with (4.3) we concluded in [3] (cf. Theorem 6.2) that under the above assumptions on  $S_h$  and  $T_h$ , for  $\Omega_1 \subset\subset \Omega_0$  and  $0 < t_* \leq t \leq t^*$ ,

$$(4.15) \quad |Q_h D_t^l u_h(t) - D_x^\alpha D_t^l u(t)|_{\Omega_1} \leq Ch^r \|v\|.$$

The following is the corresponding result for our completely discrete approximation.

**THEOREM 4.2.** *Under the assumptions of Theorem 4.1, with  $Q_h$  approximating  $D^\alpha$  with accuracy  $r$  we have for any  $\Omega_1 \subset\subset \Omega_0$ ,  $0 < t_* \leq t = nk \leq t^*$ ,*

$$|Q_h q_k U^n - D_x^\alpha D_t^l u(nk)|_{\Omega_1} \leq C(h^r + k^\nu) \|v\|.$$

*Proof.* We have by Theorem 4.1, with  $\Omega_1 \subset\subset \Omega_2 \subset\subset \Omega_0$ ,

$$|Q_h q_k U^n - Q_h D_t^l u_h(nk)|_{\Omega_1} \leq C \sum_{|\beta| \leq |\alpha|} |\partial_h^\beta (q_k U^n - D_t^l u_h(nk))|_{\Omega_2} \leq Ck^\nu \|v\|.$$

In view of (4.15) this proves the theorem.

We now turn to the superconvergence estimates. For the semidiscrete case this was done in Section 7 of [3] for the approximation of the continuous solution itself. Here we shall generalize this to cover also derivatives. For this purpose, we will use the following lemma which is a special case of Theorem 3 in Thomée [13]. Here we denote by  $\psi$  the  $N$ -dimensional analogue of the  $B$ -spline of order  $r - 2$ , that is, the convolution  $\chi * \cdots * \chi$  with  $r - 2$  factors, where  $\chi$  is the characteristic function of the cube  $[-\frac{1}{2}, \frac{1}{2}]^N$ .

**LEMMA 4.3.** *Let  $\alpha$  be a given multi-index. Then there exists a function  $K_h = K_h^{(\alpha)}$  of the form*

$$K_h(x) = h^{-N} \sum_{\beta} k_{\beta} \psi(h^{-1}x - \beta),$$

with  $k_{\beta} = 0$  when  $|\beta_j| \geq r - 1$  such that if  $\Omega_2 \subset\subset \Omega_1 \subset\subset \Omega_0$  and  $w \in C^{2r-2+|\alpha|}(\Omega_0)$ ,  $w_h \in C(\Omega_0)$ ,

$$\begin{aligned} & |K_h * \partial_h^\alpha w_h - D^\alpha w|_{\Omega_2} \\ & \leq C \left\{ h^{2r-2} |w|_{2r-2+|\alpha|, \Omega_1} + \sum_{|\beta| \leq r-2+N_0} \|\partial_h^{\alpha+\beta}(w_h - w)\|_{-(r-2), \Omega_1} \right. \\ & \quad \left. + h^{r-2} \sum_{|\beta| \leq r-2} |\partial_h^{\alpha+\beta}(w_h - w)|_{\Omega_1} \right\}. \end{aligned}$$

Applying this to  $w_h = D_t^l u_h$  and  $w = D_t^l u$  we obtain by the negative norm estimate (4.4) and the maximum norm estimate (4.3) that for  $\Omega_1 \subset\subset \Omega_0$ ,  $0 < t_* \leq t \leq t^*$ ,

$$(4.16) \quad |K_h * \partial_h^\alpha D_t^l u_h - D_x^\alpha D_t^l u|_{\Omega_1} \leq Ch^{2r-2} \|v\|_{r-2}.$$

For  $\alpha = l = 0$  this was Theorem 7.2 of [3].

We can now show that this high order estimate in space is not spoiled by discretization in time.

**THEOREM 4.3.** *Under the assumptions of Theorem 4.1, with  $K_h = K_h^{(\alpha)}$  as in*

*Lemma 4.3.* we have for any  $\Omega_1 \subset\subset \Omega_0$ ,  $0 < t_* \leq t = nk \leq t^*$ ,

$$|K_h * \partial_h^\alpha q_k U^n - D_x^\alpha D_t^l u(nk)|_{\Omega_1} \leq C\{h^{2r-2}\|v\|_{r-2} + k^\nu\|v\|\}.$$

*Proof.* We have by Theorem 4.1, with  $\Omega_1 \subset\subset \Omega'_1 \subset\subset \Omega_0$ ,

$$|K_h * \partial_h^\alpha(q_k U^n - D_t^l u_h(nk))|_{\Omega_1} \leq C|\partial_h^\alpha(q_k U^n - D_t^l u_h(nk))|_{\Omega'_1} \leq Ck^\nu\|v\|.$$

In view of (4.16) this proves the theorem.

**5. Examples.** In this section we present examples of schemes defined by the four classes of rational functions I, II, III and IV and discuss their computational implementation, in particular in the case when  $T_h$  is defined by means of a bilinear form (cf. [3, Section 8]),

$$(5.1) \quad B_h(T_h f, \chi) = (f, \chi), \quad \forall \chi \in S_h.$$

As was pointed out in the introduction, since  $T_h$  rather than  $L_h = T_h^{-1}$  is given in applications, it is now natural to study  $r(kL_h)$  as a rational function of  $T_h$ , or equivalently, to consider the rational function  $r(k/\mu)$  of  $\mu$ .

*Padé Schemes.* Examples of rational functions of types I, II and IV are provided by the above diagonal, diagonal and below diagonal entries of the Padé table for  $e^{-\tau}$ , respectively. In fact, the general entry in this Padé table is given by

$$r_{p,q}(\tau) = n_{p,q}(\tau)/d_{p,q}(\tau),$$

where

$$(5.2) \quad n_{p,q}(\tau) = \sum_{j=0}^q \frac{(p+q-j)!q!}{(p+q)!j!(q-j)!} (-1)^j \tau^j,$$

and

$$(5.3) \quad d_{p,q}(\tau) = \sum_{j=0}^p \frac{(p+q-j)!p!}{(p+q)!j!(p-j)!} \tau^j.$$

By the definition of the Padé approximant,

$$(5.4) \quad r_{p,q}(\tau) = e^{-\tau} + O(\tau^{p+q+1}) \quad \text{as } \tau \rightarrow 0,$$

so that (1.7) holds with  $\nu = p + q$ . It is well known and obvious from (5.2) and (5.3) that  $r_{p,q}$  is of type II and IV for  $p = q$  and  $p > q$ , respectively, and clearly, by (5.4),  $r_{p,q}$  is of type I for  $p < q$ .

In particular, we have  $r_{0,1}(\tau) = 1 - \tau$ , which is a rational function of type I with  $\alpha = 1$ . This defines

$$U^{n+1} = U^n - kL_h U^n,$$

and with (5.1),

$$(U^{n+1}, \chi) = (U^n, \chi) - kB_h(U^n, \chi), \quad \forall \chi \in S_h.$$

If  $B_h$  is bounded in  $H^1(\Omega)$  and the inverse assumption

$$\|\chi\|_{H^1(\Omega)} \leq Ch^{-1} \|\chi\|, \quad \forall \chi \in S_h,$$

holds, then we have

$$(L_h \chi, \chi) = B_h(\chi, \chi) \leq C_1 \|\chi\|_{H^1(\Omega)}^2 \leq C_2 h^{-2} \|\chi\|^2,$$

so that for the maximum eigenvalue of  $L_h$ ,

$$\Lambda_{\max} \leq C_2 h^{-2}.$$

Hence, a type I' scheme is defined here for

$$k/h^2 \leq \alpha_0/C_2 \quad \text{with } \alpha_0 < 1.$$

The subdiagonal and diagonal Padé approximants with linear denominators are

$$r_{1,0}(\tau) = \frac{1}{1+\tau} \quad \text{and} \quad r_{1,1}(\tau) = \frac{1 - \frac{1}{2}\tau}{1 + \frac{1}{2}\tau},$$

which are of types IV and II, respectively. They correspond respectively to the purely implicit scheme

$$(U^{n+1}, \chi) + kB_h(U^{n+1}, \chi) = (U^n, \chi), \quad \forall \chi \in S_h,$$

and the trapezoidal (Crank-Nicolson) scheme

$$(U^{n+1}, \chi) + \frac{1}{2}kB_h(U^{n+1}, \chi) = (U^n, \chi) - \frac{1}{2}kB_h(U^n, \chi), \quad \forall \chi \in S_h.$$

We consider also the subdiagonal and diagonal Padé approximants with quadratic denominators,

$$r_{2,0}(\tau) = \frac{1}{1 + \tau + \frac{1}{2}\tau^2}, \quad r_{2,1}(\tau) = \frac{1 - \frac{1}{3}\tau}{1 + \frac{2}{3}\tau + \frac{1}{6}\tau^2},$$

and

$$r_{2,2}(\tau) = \frac{1 - \frac{1}{2}\tau + \frac{1}{12}\tau^2}{1 + \frac{1}{2}\tau + \frac{1}{12}\tau^2}.$$

In all of these cases the poles are complex. Hence in the use of these functions to obtain the time stepping procedure it is convenient to use complex matrices even though the approximations  $U^n$  will be real. We shall see now that it is sufficient to solve one complex system at each time.

Consider first  $r_{2,1}$ . Setting

$$\beta_1 = \frac{1}{3} \left( 1 + i\frac{\sqrt{2}}{2} \right), \quad \gamma_1 = 1 - i\frac{\sqrt{2}}{2}$$

we have

$$r_{2,1}(k/\mu) = \frac{\mu^2 - \frac{1}{3}k\mu}{\mu^2 + \frac{2}{3}k\mu + \frac{1}{6}k^2} = 1 - k \operatorname{Re} \left( \frac{\gamma_1}{\mu + \beta_1 k} \right),$$

and hence, with  $A_h = T_h + k\beta_1 I$ , (1.7) reduces to

$$U^{n+1} = U^n - k \operatorname{Re}(\gamma_1 A_h^{-1} U^n),$$

where  $A_h^{-1}$  exists since  $\operatorname{Re} \beta_1 > 0$ . Note that in order to compute  $W = k\gamma_1 A_h^{-1} U^n$  we solve the complex elliptic problem

$$(W, \chi) + k\beta_1 B_h(W, \chi) = k\gamma_1 B_h(U^n, \chi), \quad \forall \chi \in S_h,$$

which amounts to solving a nondegenerate complex linear system of equations.

The analogous statements for  $r_{2,0}$  and  $r_{2,2}$  follow from

$$r_{2,0}(k/\mu) = \frac{\mu^2}{\mu^2 + k\mu + \frac{1}{2}k^2} = 1 - k \operatorname{Re} \left( \frac{1}{\mu + \beta_0 k} \right)$$

and

$$r_{2,2}(k/\mu) = \frac{\mu^2 - \frac{1}{2}\mu k + \frac{1}{12}k^2}{\mu^2 + \frac{1}{2}\mu k + \frac{1}{12}k^2} = 1 - k \operatorname{Re} \left( \frac{\gamma_2}{\mu + \beta_2 k} \right),$$

where

$$\beta_0 = \frac{1}{2}(1 + i), \quad \beta_2 = \frac{1}{4}(1 + i\sqrt{3}/3), \quad \gamma_2 = 1 - i\sqrt{3}.$$

We shall now turn to the schemes of type III and IV mentioned in the introduction, for which the rational functions employed have denominators a power of  $(1 + b\tau)$  with  $b$  positive. For their analysis we shall use some simple properties of Laguerre polynomials.

Recall that the Laguerre polynomials  $\{L_n^\alpha\}$  of order  $\alpha$ ,  $\alpha > -1$ ,

$$L_n^\alpha(x) = (n + \alpha)! \sum_{j=0}^n \frac{(-1)^j x^j}{j!(n-j)!(j + \alpha)!},$$

are the polynomials which are orthogonal with respect to the weight function  $x^\alpha e^{-x}$  on  $(0, \infty)$ . For their generating function we have

$$(5.5) \quad (1 - t)^{-1-\alpha} e^{-xt/(1-t)} = \sum_{n=0}^{\infty} L_n^\alpha(x) t^n \quad \text{for } |t| < 1, x \geq 0.$$

We recall also the three term recursion formula

$$(5.6) \quad nL_n^\alpha(x) = (-x + 2n + \alpha - 1)L_{n-1}^\alpha(x) - (n + \alpha - 1)L_{n-2}^\alpha(x),$$

and finally the relation

$$(5.7) \quad \frac{d}{dx} L_n^\alpha(x) = x^{-1} \{nL_n^\alpha(x) - (n + \alpha)L_{n-1}^\alpha(x)\}, \quad n \geq 1.$$

*Some Type III Schemes.* Setting  $\tau = xt/(1 - t)$  in (5.5) with  $\alpha = 1$  we obtain

$$e^{-\tau} = (1-t)^2 \sum_{n=0}^{\infty} L_n^1(x) t^n,$$

and hence, using also the recursion relation (5.6),

$$e^{-\tau} = 1 - \sum_{n=0}^{\infty} \frac{x L_n^1(x)}{n+1} t^{n+1} = 1 - \sum_{n=0}^{\infty} \frac{x L_n^1(x)}{n+1} \left( \frac{\tau}{x+\tau} \right)^{n+1}.$$

Consequently, with  $b = 1/x$  any positive number, we have

$$(5.8) \quad e^{-\tau} = 1 - \sum_{n=0}^{\infty} P_n(b) \left( \frac{\tau}{1+b\tau} \right)^{n+1} \quad \text{for } \tau > 0,$$

where  $P_n$  is the polynomial of degree  $n$  defined by

$$P_n(b) = (n+1)^{-1} b^n L_n^1(1/b).$$

Note that since  $L_n^1$  has its zeros positive and distinct the same holds for  $P_n$ .

Now let  $b_n = 1/\beta_n$  denote the largest zero of  $P_n$  (so that  $\beta_n$  is the smallest zero of  $L_n^1$ ). We define for  $\nu \geq 2$  the rational function

$$r_\nu(\tau) = 1 - \sum_{j=0}^{\nu-2} P_j(b_{\nu-1}) \left( \frac{\tau}{1+b_{\nu-1}\tau} \right)^{j+1}.$$

In view of (5.8) and the choice of  $b_{\nu-1}$  it is obvious that

$$r_\nu(\tau) = e^{-\tau} + O(\tau^{\nu+1}) \quad \text{as } \tau \rightarrow 0.$$

In order to see that these rational functions are of type III for  $\nu \geq 3$  we shall establish that for those  $\nu$ ,

$$(5.9) \quad \sup_{\tau > \delta} |r_\nu(\tau)| < 1 \quad \text{for any } \delta > 0.$$

Since  $r_\nu(\tau)$  is decreasing it suffices to show

$$\lim_{\tau \rightarrow \infty} r_\nu(\tau) > -1.$$

But, with  $\beta_j = 1/b_j$ , we have

$$\begin{aligned} \lim_{\tau \rightarrow \infty} r_\nu(\tau) &= 1 - \sum_{j=0}^{\nu-2} P_j(b_{\nu-1}) b_{\nu-1}^{-(j+1)} \\ &= 1 - \beta_{\nu-1} \sum_{j=0}^{\nu-2} \frac{L_j^1(\beta_{\nu-1})}{j+1} = f_\nu, \end{aligned}$$

and we shall prove that for  $\nu \geq 2$ ,  $f_{\nu+1} > f_\nu$ . Since  $f_2 = 1 - \beta_1 = 1 - 2 = -1$ , this will complete the proof. Set

$$f(\beta) = 1 - \beta \sum_{j=0}^{\nu-1} \frac{L_j^1(\beta)}{j+1},$$

so that

$$f_\nu = f(\beta_{\nu-1}), \quad f_{\nu+1} = f(\beta_\nu).$$

Since  $\beta_{j-1} > \beta_j$  it suffices to show that  $f(\beta)$  is decreasing in  $(0, \beta_{\nu-1})$ . But, using (5.7) with  $\alpha = 1$  we obtain

$$f'(\beta) = -1 - \sum_{j=1}^{\nu-1} (L_j^1(\beta) - L_{j-1}^1(\beta)) = -L_{\nu-1}^1(\beta),$$

which is negative in  $(0, \beta_{\nu-1})$ . This completes the proof of (5.9).

We shall now consider the computational implementation of the scheme corresponding to these rational functions. We have

$$r_\nu(k/\mu) = 1 - \sum_{j=0}^{\nu-2} k^{j+1} P_j(b_{\nu-1}) \frac{1}{(\mu + kb_{\nu-1})^{j+1}},$$

so that with  $A_h = T_h + kb_{\nu-1}I$ , (1.7) reduces to

$$U^{n+1} = U^n - \sum_{j=0}^{\nu-2} k^{j+1} P_j(b_{\nu-1}) A_h^{-(j+1)} U^n.$$

Setting

$$U_0^n = U^n, \quad U_{j+1}^n = kA_h^{-1} U_j^n,$$

we find

$$U^{n+1} = U^n - \sum_{j=0}^{\nu-2} P_j(b_{\nu-1}) U_{j+1}^n.$$

Hence  $U^{n+1}$  can be obtained from  $U^n$  by solving  $\nu - 1$  linear systems, all with the same matrix  $A_h$ . When  $T_h$  is defined by (5.1) this system takes the form

$$(U_{j+1}^n, \chi) + kb_{\nu-1} B_h(U_{j+1}^n, \chi) = kB_h(U_j^n, \chi), \quad \forall \chi \in S_h.$$

For  $\nu = 2$  we obtain

$$r_2(\tau) = 1 - \frac{\tau}{1 + \frac{1}{2}\tau} = \frac{1 - \frac{1}{2}\tau}{1 + \frac{1}{2}\tau},$$

which defines the trapezoidal type II scheme discussed above. For  $\nu = 3$ , we have

$$r_3(\tau) = 1 - \frac{\tau}{1 + b_2\tau} - \frac{\sqrt{3}}{6} \left( \frac{\tau}{1 + b_2\tau} \right)^2, \quad b_2 = \frac{1}{2} \left( 1 + \frac{\sqrt{3}}{3} \right),$$

which corresponds to the Calahan scheme analyzed by Zlámal [15]. In the appendix we have listed the coefficients for the schemes with  $\nu \leq 10$ .

*Some Type IV Schemes.* In a way analogous to the derivation of the type III schemes above we can generate a class of type IV schemes. By taking  $\alpha = 0$  in (5.5) and making the same subsequent change of variables as above, we have that for  $b > 0$ ,

$$e^{-\tau} = \frac{1}{1 + b\tau} \sum_{j=0}^{\infty} Q_j(b) \left( \frac{\tau}{1 + b\tau} \right)^j \quad \text{for } \tau \geq 0,$$

where  $Q_j(b) = b^j L_j^0(1/b)$ . Defining now for  $\nu \geq 1$ ,

$$r_\nu(\tau) = \frac{1}{1 + b_\nu\tau} \sum_{j=0}^{\nu-1} Q_j(b_\nu) \left( \frac{\tau}{1 + b_\nu\tau} \right)^j$$



where  $b_\nu$  is the largest zero of  $Q_\nu$ , we clearly have

$$r_\nu(\tau) = e^{-\tau} + O(\tau^{\nu+1}) \quad \text{as } \tau \rightarrow 0.$$

In order to show that  $r_\nu(\tau)$  induces a type IV scheme it remains only to establish that

$$0 < r_\nu(\tau) < 1 \quad \text{for } \tau > 0,$$

since clearly  $\lim_{\tau \rightarrow \infty} r_\nu(\tau) = 0$ . Setting, as above,  $\beta_j = 1/b_j$  we have

$$r_\nu(\tau) = \frac{1}{1 + b_\nu \tau} \sum_{j=0}^{\nu-1} L_j^0(\beta_\nu) \left( \frac{\tau b_\nu}{1 + \tau b_\nu} \right)^j,$$

and hence, with  $\tau b_\nu / (1 + \tau b_\nu) = \lambda$  it is enough to prove that for each  $\lambda$  in  $(0, 1)$ ,

$$0 < f(\lambda) = (1 - \lambda) \sum_{j=0}^{\nu-1} L_j^0(\beta_\nu) \lambda^j < 1.$$

Since  $f(0) = 1$  and  $f(1) = 0$  it suffices to show that  $f$  is decreasing in  $(0, 1)$ . But, using (5.7) in the last step we obtain

$$\begin{aligned} f'(\lambda) &= \sum_{j=1}^{\nu-1} j L_j^0(\beta_\nu) \lambda^{j-1} - \sum_{j=0}^{\nu-1} (j+1) L_j^0(\beta_\nu) \lambda^j \\ &= \sum_{j=1}^{\nu-1} j (L_j^0(\beta_\nu) - L_{j-1}^0(\beta_\nu)) \lambda^{j-1} - \nu L_{\nu-1}^0(\beta_\nu) \lambda^{\nu-1} \\ &= \sum_{j=1}^{\nu-1} \beta_\nu \frac{d}{dx} L_j^0(\beta_\nu) \lambda^{j-1} - \nu L_{\nu-1}^0(\beta_\nu) \lambda^{\nu-1}. \end{aligned}$$

Since  $L_j^0$  is nonincreasing in  $(0, \beta_j)$  and  $L_{\nu-1}^0(\beta_\nu)$  is positive, the result follows.

We have here

$$r_\nu(k/\mu) = \frac{\mu}{\mu + b_\nu k} \sum_{j=0}^{\nu-1} k^j Q_j(b_\nu) \frac{1}{(\mu + k b_\nu)^j},$$

so that now with  $A_h = T_h + k b_\nu I$ , (1.6) reduces to

$$U^{n+1} = \sum_{j=0}^{\nu-1} Q_j(b_\nu) U_j^n,$$

where the  $U_j^n$ 's are defined by

$$U_0^n = A_h^{-1} T_h U^n, \quad U_{j+1}^n = k A_h^{-1} U_j^n, \quad j = 0, \dots, \nu-2.$$

When  $T_h$  is defined by (5.1), this is

$$\begin{aligned} (U_0^n, \chi) + k b_\nu B_h(U_0^n, \chi) &= (U^n, \chi), \quad \forall \chi \in S_h, \\ (U_{j+1}^n, \chi) + k b_\nu B_h(U_{j+1}^n, \chi) &= k B_h(U_j^n, \chi), \quad j = 0, \dots, \nu-2. \end{aligned}$$

The values of  $b_\nu$  and  $Q_j(b_\nu)$  for  $0 \leq j \leq \nu-1$  and  $\nu \leq 10$  are given in the appendix.

**6. Variations.** In this section we shall briefly focus our attention first on schemes in which both the time and space discretizations are allowed to change a finite number of times and then where initial data other than  $P_0 v$  are chosen.

We consider thus a finite set of rational functions  $\{r_j(\tau)\}_1^M$ , a corresponding set of time steps  $\{k_j\}_1^M$  and a finite set of finite dimensional subspaces  $\{S_{j,h_j}\}_1^M$  of  $L_2(\Omega)$  with corresponding operators  $T_{j,h_j}$  and  $L_{j,h_j} = T_{j,h_j}^{-1}$ . Denoting the  $L_2$  projection onto  $S_{j,h_j}$  by  $P_{j,0}$  we then define a discretization scheme by successively applying the  $j$ th scheme  $n_j$  times with time step  $k_j$ , for  $j = 1, \dots, M$ . More precisely, setting  $U_j^n \in S_{j,h_j}$ ,  $n = 0, \dots, n_j$ , the approximation at  $t = \sum_{l=1}^{j-1} n_l k_l + nk_j = t_{j-1} + nk_j$  (with  $t_0 = 0$ ), we define

$$U_j^n = r_j(k_j L_{j,h_j})^n P_{j,0} U_{j-1}^{n-1} \quad \text{for } n = 0, \dots, n_j,$$

$$U_1^0 = P_{1,0} v.$$

We then have the following  $L_2$  error estimates:

**THEOREM 6.1.** *Assume that each of the schemes is of type I' or II. Then for  $0 \leq t = \sum_{l=1}^{j-1} n_l k_l + nk_j \leq t^*$ ,  $n = 1, \dots, n_j$ ,  $j = 1, \dots, M$ ,*

$$(6.1) \quad \|U_j^n - u(t)\| \leq C \sum_{j=1}^M \{h_j^{r_j} \|v\|_{r_j} + k_j^{v_j} \|v\|_{2v_j}\}.$$

Further, if  $t_1 = n_1 k_1 \geq t_*$  and  $0 < t_* \leq t = \sum_{l=1}^{j-1} n_l k_l + nk_j \leq t^*$  then

$$(6.2) \quad \|U_j^n - u(t)\| \leq C(t_*) \sum_{j=1}^n (h_j^{r_j} + k_j^{v_j}) \|v\|.$$

*Proof.* We denote by  $\tilde{U}_j^n$  the fully discrete Galerkin approximation at time  $t = t_{j-1} + nk_j$  with initial values  $u(t_{j-1})$  for  $n = 0$ , that is

$$\tilde{U}_j^n = r_j(k_j L_{j,h_j})^n P_{j,0} u(t_{j-1}), \quad n = 0, \dots, n_j.$$

We then have, by Theorem 2.2, for  $n = 1, \dots, n_j$ ,

$$\|\tilde{U}_j^n - u(t_{j-1} + nk_j)\| \leq C \{h_j^{r_j} \|u(t_{j-1})\|_{r_j} + k_j^{v_j} \|u(t_{j-1})\|_{2v_j}\}.$$

In particular, setting  $j = 1$  this shows (6.1) for  $t \leq t_1$ . For  $t > t_1$ , on the other hand, using the stability of the scheme (cf. (1.11)),

$$\|U_j^n - \tilde{U}_j^n\| \leq \|U_{j-1}^{n-1} - u(t_{j-1})\|,$$

so that

$$(6.3) \quad \|U_j^n - u(t_{j-1} + nk_j)\|$$

$$\leq \|U_{j-1}^{n-1} - u(t_{j-1})\| + C \{h_j^{r_j} \|u(t_{j-1})\|_{r_j} + k_j^{v_j} \|u(t_{j-1})\|_{2v_j}\}.$$

Now, since the solution operator of the continuous problem is bounded in  $\dot{H}^s$ , we obtain

$$\|U_j^n - u(t_{j-1} + nk_j)\|$$

$$\leq \|U_{j-1}^{n-1} - u(t_{j-1})\| + C \{h_j^{r_j} \|v\|_{r_j} + k_j^{v_j} \|v\|_{2v_j}\}.$$

The first result (6.1) now follows at once by recursion.

In order to prove (6.2) we use the fact that for any  $s \geq 0$ , and  $t \geq t_*$ ,

$$\|u(t)\|_s \leq Ct^{-s/2} \|v\| \leq C(t_*) \|v\|,$$

so that (6.3) implies for  $j = 2, \dots, M$ ,

$$\begin{aligned} \|U_j^n - u(t_{j-1} + nk_j)\| \\ \leq \|U_{j-1}^{n_{j-1}} - u(t_{j-1})\| + C(t_*)(h_j^r + k_j^\nu) \|v\|. \end{aligned}$$

Since now by Theorem 2.1,

$$\|U_1^{n_1} - u(t_1)\| \leq C(t_*)(h_1^r + k_1^\nu) \|v\|,$$

the result (6.2) also follows by recursion. This completes the proof.

Let us notice that if only the time discretization varies and not the spaces  $\{S_{j,h_j}\}$  then with  $T_h$  and  $L_h = T_h^{-1}$  corresponding to  $S_h$ , we have

$$U_j^n = \prod_{l=1}^{j-1} r(k_l L_h)^{n_l} r(k_j L_h)^n P_0 v,$$

where now the operators in the product commute. Consequently, the order in which the time discretizations have been applied is insignificant. It follows, for instance, that the finitely many different time discretizations can be applied cyclically, to give the analogues of the results of (6.1) and (6.2), the latter for any time bounded away from zero and infinity.

We now return to the completely discrete problem (1.7), but with initial data  $v_h$  other than  $P_0 v$ , so that

$$U^n = r(k L_h)^n v_h.$$

Following the lines of the proof of Theorem 3.3 in [3] we then have the following  $L_2$  error estimate:

**THEOREM 6.2.** *Assume that the scheme is of type I', II' or III. Then for each  $j \geq 0$  and for  $0 < t = nk \leq t^*$ ,*

$$\|U^n - u(nk)\| \leq C(h^r t^{-r/2} + k^\nu t^{-\nu}) \|v_h\| + Ct^{-j/2} \|v - v_h\|_{-j}.$$

*Proof.* Let  $\tilde{u}$  be the solution of the continuous problem with initial data  $v_h$ . Then by Theorem 2.2,

$$\|U^n - \tilde{u}(nk)\| \leq C\{h^r t^{-r/2} + k^\nu t^{-\nu}\} \|v_h\|.$$

On the other hand, since  $u - \tilde{u}$  is the solution of the continuous problem with initial data  $v - v_h$  we have

$$\|u(t) - \tilde{u}(t)\| \leq Ct^{-j/2} \|v - v_h\|_{-j}.$$

Together these estimates prove the theorem.

In particular, if  $v_h$  is chosen boundedly in  $L_2$  and approximating  $v$  to order  $O(h^r)$  in some negative norm, the error is  $O(h^r + k^\nu)$  for  $t$  positive. Recall from [3] that this is satisfied, for example, for  $v_h$  the elliptic projection  $P_1 v = T_h L v$  of  $v \in \dot{H}^2$  if (iib) (or (iia) if  $r = 2$ ) holds.

## Appendix

(a) Tables of constants associated with the rational functions  $r_\nu(\tau) = 1 - \sum_{j=0}^{\nu-2} P_j(b_{\nu-1})(\tau/(1 + b_{\nu-1}\tau))^{j+1}$ ,  $\nu \leq 10$ , rounded to eight figures.

$$P_0 \equiv 1.$$

$$b_1 = .5$$

$$b_2 = .78867513$$

$j$	$P_j(b_2)$
1	.28867513

$$b_3 = 1.0685790$$

$j$	$P_j(b_3)$
1	.56857902
2	.23994877

$$b_4 = 1.3453664$$

$j$	$P_j(b_4)$
1	.84536642
2	.63131105
3	.35112809

$$b_5 = 1.6206645$$

$j$	$P_j(b_5)$
1	1.1206645
2	1.1725556
3	1.0855973
4	.75003511

$$b_6 = 1.8951306$$

$j$	$P_j(b_6)$
1	1.3951306
2	1.8630560
3	2.3250180
4	2.5702151
5	2.1228872

$$b_7 = 2.1690834$$

$j$	$P_j(b_7)$
1	1.6690834
2	2.7025062
3	4.1908611
4	6.0773015
5	7.8122875
6	7.5096538

$$b_8 = 2.4426969$$

$j$	$P_j(b_8)$
1	1.9426969
2	3.6907381
3	6.8045360
4	12.020296
5	19.865792
6	29.119416
7	31.936476

$$b_9 = 2.7160742$$

$j$	$P_j(b_9)$
1	2.2160742
2	4.8276516
3	10.287422
4	21.280435
5	42.191381
6	78.287509
7	128.88852
8	158.83591

(b) Tables of constants associated with the rational functions  $r_\nu(\tau) = [1/(1 + b_\nu\tau)] \sum_{j=0}^{\nu-1} Q_j(b_\nu)(\tau/(1 + b_\nu\tau))^j$ ,  $\nu \leq 10$ , rounded to eight figures.

$$Q_0 \equiv 1.$$

$$b_1 = 1$$

$$b_2 = 1.7071068$$

$j$	$Q_j(b_2)$
1	.70710678

$$b_3 = 2.4051496$$

$j$	$Q_j(b_3)$
1	1.4051496
2	1.4744453

$$b_4 = 3.1003167$$

$j$	$Q_j(b_4)$
1	2.1003167
2	3.9113304
3	5.4480493

$$b_5 = 3.7941979$$

$j$	$Q_j(b_5)$
1	2.7941979
2	7.3075442
3	16.957853
4	29.458888

$$b_6 = 4.4873917$$

$j$	$Q_j(b_6)$
1	3.4873917
2	11.661901
3	36.515558
4	101.50142
5	211.31583

$$b_7 = 5.1801749$$

$j$	$Q_j(b_7)$
1	4.1801749
2	16.973862
3	66.106869
4	241.14213
5	781.11612
6	1895.6814

$$b_8 = 5.8726930$$

$j$	$Q_j(b_8)$
1	4.8726930
2	23.243137
3	107.71731
4	478.88829
5	1994.4686
6	7377.7839
7	20451.648

$$b_9 = 6.5650300$$

$j$	$Q_j(b_9)$
1	5.5650300
2	30.469559
3	163.33232
4	850.73985
5	4251.4423
6	19905.841
7	82793.242
8	258096.14

$$b_{10} = 7.2572379$$

$j$	$Q_j(b_{10})$
1	6.2572379
2	38.653026
3	234.93728
4	1398.1894
5	8086.1025
6	44871.439
7	233319.27
8	1077827.0
9	3732247.5

The Aiken Computation Laboratory  
Harvard University  
Cambridge, Massachusetts 02138

Department of Mathematics  
Cornell University  
Ithaca, New York 14853

Department of Mathematics  
Chalmers University of Technology  
S-402 20 Göteborg, Sweden

1. J. BLAIR, *Approximate Solution of Elliptic and Parabolic Boundary Value Problems*, Thesis, University of California, Berkeley, 1970.
2. J. H. BRAMBLE & A. H. SCHATZ, "Higher order local accuracy by averaging in the finite element method," *Math. Comp.*, v. 31, 1977, pp. 94–111.
3. J. H. BRAMBLE, A. H. SCHATZ, V. THOMÉE & L. B. WAHLBIN, "Some convergence estimates for Galerkin type approximations for parabolic equations," *SIAM J. Numer. Anal.*, v. 14, 1977, pp. 218–241.

4. J. H. BRAMBLE & V. THOMÉE, "Discrete time Galerkin methods for a parabolic boundary value problem," *Ann. Mat. Pura Appl.* (4), v. 101, 1974, pp. 115–152. MR 52 #9639.
5. M. CROUZEIX, *Sur l'Approximation des Équations Différentielles Opérationnelles Linéaires par des Méthodes de Runge-Kutta*, Thesis, University of Paris VI, 1975.
6. J. DOUGLAS, JR. & T. DUPONT, "Galerkin methods for parabolic equations," *SIAM J. Numer. Anal.*, v. 7, 1970, pp. 575–626. MR 43 #2863.
7. T. DUPONT, "Some  $L^2$  error estimates for parabolic Galerkin methods," *The Mathematical Foundations of the Finite Element Method with Applications to Partial Differential Equations* (A. K. Aziz, Editor), Academic Press, New York, 1972, pp. 491–504. MR 53 #7067.
8. H.-P. HELFRICH, "Fehlerabschätzungen für das Galerkinverfahren zur Lösung von Evolutionsgleichungen," *Manuscripta Math.*, v. 13, 1974, pp. 219–235. MR 50 #8983.
9. G. J. MAKINSON, "Stable high order implicit methods for the numerical solution of systems of differential equations," *Comput. J.*, v. 11, 1968/69, pp. 305–310. MR 38 #4040.
10. J. A. NITSCHKE & A. H. SCHATZ, "Interior estimates for Ritz-Galerkin methods," *Math. Comp.*, v. 28, 1974, pp. 937–958. MR 51 #9525.
11. S. P. NØRSETT, "One-step methods of Hermite type for numerical integration of stiff systems," *BIT*, v. 14, 1974, pp. 63–77. MR 49 #1787.
12. V. THOMÉE, "Some convergence results for Galerkin methods for parabolic boundary value problems," *Mathematical Aspects of Finite Elements in Partial Differential Equations* (Carl de Boor, Editor), Academic Press, New York, 1974, pp. 55–88.
13. V. THOMÉE, "High order local approximations to derivatives in the finite element method," *Math. Comp.*, v. 31, 1977, pp. 652–660.
14. M. F. WHEELER, "A priori  $L_2$  error estimates for Galerkin approximations to parabolic partial differential equations," *SIAM J. Numer. Anal.*, v. 10, 1973, pp. 723–759. MR 50 #3613.
15. M. ZLÁMAL, "Finite element methods for parabolic equations," *Math. Comp.*, v. 28, 1974, pp. 393–404. MR 52 #9647.