JOURNAL OF MATHEMATICAL ANALYSIS AND APPLICATIONS 78, 328–356 (1980)

Galerkin Approximations for Initial Value Problems with Known End Time Conditions

JAMES C. CAVENDISH

Mathematics Department, General Motors Research, Warren, Michigan 48090

AND

CHARLES A. HALL AND THOMAS A. PORSCHING

Institute for Computational Mathematics and Applications, Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

Submitted by W. F. Ames

Galerkin's method is used to approximate the transient solutions of intial value problems in which a steady state or advanced time state is known. A convergence theorem is established and choices of basis functions are discussed. The method is then applied to systems arising from nuclear reactor kinetics theory and the semi-discretization of parabolic two-point boundary value problems.

1. INTRODUCTION

Numerical methods for solving systems of ordinary differential equations (ode's) subject to an initial condition typically require the solution of systems of algebraic equations a each time step [1, 15, 20, 21, 29]. However, a recent paper [3] suggests a new approach for handling such problems when an end time condition is also known. The new approach produces an approximation for all time and requires only the solution of a *single* system of equations. See [6] for another effort to avoid time stepping.

The "blended infinite element method" as presented in [3] is an algorithm for the numerical solution of two-point parabolic initial-boundary value problems. However, we observe here that in fact the blended infinite element method as described in [3, 4] is mathematically equivalent to two applications of Galerkin's method: first, a standard semi-discrete Galerkin approximation $\hat{u}(x, t) = \sum_{i=0}^{N+1} \hat{u}_i(t)\, \varphi_i(x)$ is formed, where the functions $\hat{u}_i(t)$, $1 \leqslant i \leqslant N$, are characterized as solutions of a system of linear or nonlinear ordinary differential equations in time t [11, 12, 26]; second, Galerkin's method is again applied to this system of ode's to obtain approximations to

0022-247X/80/110328–29$02.00/0

the functions $\hat{u}_i(t)$, $1 \leqslant i \leqslant N$. It is worth emphasizing that these two steps are combined into one step in [3, 4].

The convergence of the semi-discrete Galerkin approximation is well documented (see, for example, [5, 11, 12, 26]). It is the convergence of the second application of Galerkin's method above which we now investigate in the context of its application to more general systems of ode's. The systems of ode's inherent in the infinite element method presented in [3] are a special case of those considered in Section 2. Error bounds are given in Theorem 1 below. Their from requires the establishment of new approximation theoretic results concerning simultaneous approximation on the semi-infinite interval $[0, \infty)$. For the Duffin–Whidden exponomials [14] and the Laguerre functions such results are given in Section 3 providing a proof of convergence of the respective Galerkin approximations under suitable conditions (Theorems 4 and 7).

For our purposes in this paper we consider the initial value problem

$$\frac{dy}{dt} = -F(t, y), \qquad 0 < t < b \leqslant \infty, \tag{1}$$

$$y(0) = y_0,$$

where y is an N-vector of continuously differentiable functions and F is a continuous function

$$F\colon (0, b) \times R^N \to R^N.$$

We assume that (1) possesses a unique solution $y(t)$ on $[0, b)$ such that $-\infty < \lim_{t \uparrow b} y(t) < \infty$. Then, we define $y(b)$ to be this limit. Systems such as (1) are usually solved numerically by methods which approximate the evolutionary behavior of $y(t)$ in a step-by-step manner [20, 21]. In this paper we are concerned with the situation when $y(b)$ is known. Usually, this occurs when $b = \infty$ and the known value is a steady state of y. However, there are certain cases (for example, the exact boundary control problem of [16, 18]) for which $b < \infty$.

The method suggested by [3, 4], hereafter referred to as the Time–Galerkin Method, approximates $y(t)$ on the entire interval $[0, b]$ by an expansion of the form

$$Y(t) = y(b) + \sum_{n=1}^{M} c_n \theta_n(t), \qquad c_n \in R^N. \tag{2}$$

Here the expansion functions $\theta_n(t)$ are members of $W^{1,2}(0, b)$ which satisfy $\theta_1(0) = 1$, $\theta_n(0) = 0$, $n = 2, 3, \ldots, M$; $\lim_{t \uparrow b} \theta_n(t) \equiv \theta_n(b) = 0$, $1 \leqslant n \leqslant M$.

Furthermore, $c_1 = y_0 - y(b)$ and the remaining vector coefficients c_n, $n = 2, 3, \ldots, M$, are determined by the Galerkin equations

$$\int_0^b \left[\frac{dY}{dt} + F(t, Y) \right]_i \theta_j(t)\, dt = 0, \qquad j = 2, 3, \ldots, M; \quad i = 1, 2, \ldots, N. \tag{3}$$

Equations (3) constitute a nonlinear system of $(M-1)N$ scalar equations for the $(M-1)N$ unknown components of $c_2, c_3, \ldots, c_M$. Note that by construction $Y(0) = y_0$, $Y(b) = y(b)$.

In Section 2 we prove a result establishing a bound on the continuous L_2-norm of the error $Y - y$ in the Time–Galerkin method. Specific classes of expansion functions which are applicable to the infinite interval ($b = \infty$) are studied in Section 3. Section 4 contains numerical results of applying the Time–Galerkin Method to a system which arises in the study of reactor kinetics and in the final section, we apply this scheme to linear and nonlinear systems of ordinary differential equations arising from the semi-discretization of parabolic partial differential equations.

2. Error Bounds for the Time-Galerkin Method

In this section we establish a result which bounds the L_2-norm of the discretization error $Y - y$ by a function of the approximation error. We assume that system (3) possesses a unique solution (for linear constant coefficient problems, necessary and sufficient conditions for this conclusion are established in the Appendix).

For $x(t) \in R^N$, we let $|x|_2 = (x^T x)^{1/2}$ and $\|x\|_2 = \left(\int_0^b |x|_2^2\, dt\right)^{1/2}$, provided that the integral exists in the extended real number system. Furthermore, we let

$$S_M = \left\{ w \;\middle|\; w = y(b) + \sum_{n=1}^{M} a_n \theta_n(t),\ a_1 = y_0 - y(b) \right\},$$

and we make the following definitions:

A continuously differentiable map F: $(0, b) \times R^N \to R^N$ is said to be *uniformly monotone* if for F_y, the Jacobian matrix of $F(t, \cdot)$, there is a positive constant γ such that $x^T F_y(t, y)\, x \geqslant \gamma x^T x$ for all $t \in (0, b)$ and y, x in R^N. F is said to be *uniformly Lipschitz* in t if, for each $x \in R^N$ the elements of F_y are bounded for all $t \in (0, b)$.

We note that for any such F, if $t \in (0, b)$ and $x, z \in R^N$, then

$$\begin{aligned}(x - z)^T [F(t, x) - F(t, z)] &= \int_0^1 (x - z)^T F_y(t, z + s(x - z))(x - z)\, ds \\ &\geqslant \gamma (x - z)^T (x - z). \end{aligned} \tag{4}$$

Our main result is contained in the following theorem:

THEOREM 1. *Let $F: (0, b) \times R^N \to R^N$ be a continuously differentiable uniformly monotone map on R^N which is also uniformly Lipschitz in t. Let $D \subset R^N$ be any compact set which contains $y(t)$, $t \in [0, b]$, where $y(t)$ is the solution of* (1). *If W is any element of S_M which is also in D for $t \in [0, b]$, and if Y is the Time–Galerkin approximation determined by* (3), *then*

$$\|Y - y\|_2 \leqslant \alpha_1 \|W - y\|_2 + \alpha_2 \left\| \frac{d}{dt}(W - y) \right\|_2, \tag{5}$$

where the constants α_1 and α_2 do not depend on W or Y.

Remark. If (1) is autonomous, then under the above hypotheses on F, we can prove the existence and uniqueness of a solution of (1) on $[0, b)$ such that $\lim_{t \uparrow b} y(t)$ is finite. To see this recall [7, p. 14] that if $K > 0$ is a constant, then a unique solution of (1) exists in some neighborhood of $t = 0$, and this solution may be continued to $t = t^*$, where $t^* = \inf\{\xi \mid |y(\xi) - y_0|_2 = K\}$. Now the hypotheses on F are sufficient to guarantee that the nonlinear system $F(y) = 0$ has a unique solution, say, y_∞ (see [27, p. 143]). Let $\rho(t) = |y(t) - y_\infty|_2^2$. Then,

$$\frac{d\rho}{dt} = 2\left(\frac{dy}{dt}\right)^T (y - y_\infty) = -2(F(y) - F(y_\infty))^T (y - y_\infty).$$

But, by (4),

$$(F(y) - F(y_\infty))^T (y - y_\infty) \geqslant \gamma |y - y_\infty|^2.$$

Hence, $d\rho/dt \leqslant -2\gamma\rho$ and so $\rho(t) \leqslant \rho(0) e^{-2\gamma t}$. It follows that with $K = 3|y_0 - y_\infty|_2$, there is a unique solution of (1) which may be continued to any finite value of t. Moreover, $\lim_{t \to \infty} y(t) = y_\infty$.

Proof of Theorem 1. Let $t^* \in (0, b)$. Since $t^* < \infty$,

$$k_j \equiv \int_0^{t^*} \left[\frac{dW}{dt} + F(t, W) \right] \theta_j(t)\, dt, \qquad j = 2, 3, \ldots, M,$$

is well defined. Moreover, since

$$\int_0^{t^*} \left[\frac{dy}{dt} + F(t, y) \right] \theta_j(t)\, dt = 0, \qquad j = 2, 3, \ldots, M,$$

we have

$$k_j \equiv \int_0^{t^*} \left[\frac{d}{dt}(W - y) + F(t, W) - F(t, y) \right] \theta_j(t)\, dt, \qquad j = 2, 3, \ldots, M. \tag{7}$$

Also, from (3) and (6) it follows that

$$k_j = \int_0^{t'} \left[\frac{d}{dt}(W - Y) + F(t, W) - F(t, Y) \right] \theta_j(t)\, dt$$

$$- \int_{t'}^{b} \left[\frac{dY}{dt} + F(t, Y) \right] \theta_j(t)\, dt, \qquad j = 2, 3, \ldots, M, \tag{8}$$

the last integral on the right side of (8) being finite by assumption. Noting that $W - Y$ may be expressed as $\sum_{j=2}^{M} \delta_j \theta_j(t)$ for some vectors $\{\delta_j\}$ we multiply (7) and (8) by δ_j^T and sum from $j = 2$ to $j = M$ to get

$$\sum_{j=2}^{M} \delta_j^T k_j = \int_0^{t'} (W - Y)^T \left[\frac{d}{dt}(W - y) + F(t, W) - F(t, y) \right] dt, \tag{9}$$

and

$$\sum_{j=2}^{M} \delta_j^T k_j = \int_0^{t'} (W - Y)^T \left[\frac{d}{dt}(W - Y) + F(t, W) - F(t, Y) \right] dt$$

$$- \int_{t'}^{b} (W - Y)^T \left[\frac{dY}{dt} + F(t, Y) \right] dt. \tag{10}$$

But, in (10)

$$F(t, W) - F(t, Y) = \int_0^1 F_y[t, Y + s(W - Y)]\, ds (W - Y) = C(W - Y), \tag{11}$$

where the matrix $C \equiv \int_0^1 F_y[t, Y + s(W - Y)]\, ds$ is positive definite and so, by hypothesis, $(W - Y)^T C(W - Y) \geqslant \gamma |W - Y|_2^2$. Let β be a positive real number such that $\beta < (\gamma)^{1/2}$. Also, let

$$V \equiv \beta^{-1} \left[\frac{d}{dt}(W - y) + F(t, W) - F(t, y) \right]$$

and

$$U \equiv \beta(W - Y).$$

Then equating (9) and (10) and using (11) and the elementary inequality

$$\int_0^{t'} U^T V\, dt \leqslant \frac{1}{2} \int_0^{t'} (|U|_2^2 + |V|_2^2)\, dt,$$

we obtain

$$\int_0^{t^*} (W-Y)^T C(W-Y)\,dt$$
$$\leqslant \frac{1}{2}\int_0^{t^*} (|U|_2^2 + |V|_2^2)\,dt - \frac{1}{2}|W(t^*) - Y(t^*)|_2^2$$
$$+ \int_{t^*}^{b} (W-Y)^T \left[\frac{dY}{dt} + F(t, Y)\right] dt.$$

Therefore,

$$\gamma \int_0^{t^*} |W-Y|_2^2\,dt$$
$$\leqslant \int_0^{t^*} (W-Y)^T C(W-Y)\,dt$$
$$\leqslant \frac{1}{2}\beta^2 \int_0^{t^*} |W-Y|_2^2\,dt + \frac{1}{2}\beta^{-2}\int_0^{t^*} \left|\frac{d}{dt}(W-y) + F(t, W) - F(t, y)\right|_2^2 dt$$
$$- \frac{1}{2}|W(t^*) - Y(t^*)|_2^2 + \int_{t^*}^{b} (W-Y)^T \left[\frac{dY}{dt} + F(t, Y)\right] dt. \tag{12}$$

It then follows that

$$\int_0^{t^*} |W-Y|_2^2\,dt \leqslant \frac{1}{\gamma\beta^2}\int_0^{t^*} \left|\frac{d}{dt}(W-y) + F(t, W) - F(t, y)\right|_2^2 dt + \Delta, \tag{13}$$

where

$$\Delta = \frac{2}{\gamma}\int_{t^*}^{b} (W-Y)^T \left[\frac{dY}{dt} + F(t, Y)\right] dt - \frac{1}{\gamma}|W(t^*) - Y(t^*)|_2^2.$$

If $W \equiv y$, then (13) shows that $\int_0^{t^*} |y-Y|_2^2\,dt = \Delta$ and clearly $\Delta \to 0$ as $t^* \uparrow b$. In this case (5) is a trivial equality. Otherwise, suppose $W \not\equiv y$. Then with $\alpha = \gamma^{1/2}\beta$

$$\left[\int_0^{t^*} |W-Y|_2^2\,dt\right]^{1/2}$$
$$\leqslant \frac{1}{\alpha}\left[\int_0^{t^*} \left|\frac{d}{dt}(W-y) + F(t, W) - F(t, y)\right|_2^2 dt\right]^{1/2} \left(1 + \frac{\Delta}{K}\right)^{1/2}, \tag{14}$$

where $K > 0$ is any lower bound for the first term on the right side of (13). As before,

$$F(t, W) - F(t, y) = \int_0^1 F_y[t, y + s(W - y)](W - y)\, ds.$$

Now let

$$\Gamma \equiv \sup_{\substack{x \in D \\ t \in (0,b)}} |F_y(t, x)|_2.$$

Since F is uniformly Lipschitz in t and D is compact, $\Gamma < \infty$ and

$$|F(t, W) - F(t, y)|_2 \leqslant \Gamma |W - y|_2.$$

Combining this with (14), we obtain by the triangle inequality

$$\int_0^{t^*} ||Y - y|_2^2\, dt|^{1/2} \leqslant \frac{1}{\alpha}\left(1 + \frac{\Delta}{K}\right)^{1/2}\left[\int_0^{t^*}\left|\frac{d}{dt}(W - y)\right|_2^2 dt\right]^{1/2} + \left[1 + \frac{\Gamma}{\alpha}\left(1 + \frac{\Delta}{K}\right)^{1/2}\right]\left[\int_0^{t^*} |W - y|_2^2\, dt\right]^{1/2}.$$

Inequality (5) now follows by letting $t^* \uparrow b$ and noting that $\Delta \to 0$. Q.E.D.

In the finite interval case, $b < \infty$, there are many ways to choose the expansion set $\{\theta_n(t)\}_{n-1}^M$ such that the right side of (5) is arbitrarily small for M sufficiently large. For example, let $\theta_1(t) = 1 - t/b$ and partition $[0, b]$ into M equal subintervals, say, $[0, b] = \bigcup_{n-1}^M [t_n, t_{n-1}]$. Then, for $n = 2,..., M$ let $\theta_n(t)$ be the function which is linear on each subinterval, taking the value of unity at t_n and zero at t_i, $i \neq n$. If the hypotheses of Theorem 1 hold, then $y \in C^2(0, b)$. Also, $z \in C^2(0, b)$, where $z(t) \equiv y(t) - y(b) - a_1\theta_1(t)$. But it is known [2] that there are vectors $a_2,..., a_M$ such that if $w(t) = \sum_{n-2}^M a_n\theta_n(t)$, then $\sup_{t\in(0,b)} |z(t) - w(t)|_\infty = O(1/M^2)$, and $\sup_{t\in(0,b)} |(d/dt)(z(t) - w(t))|_\infty = O(1/M)$, where $|\cdot|_\infty$ is the infinity norm in R^N. Hence, if $W \equiv y(b) + a_1\theta_1(t) + w$, it is clear that the right side of (5) is $O(1/M)$. As the regularity of y improves, the choice of more elaborate piecewise polynomial functions for the θ_n yields higher order bounds for the right side of (5) (see, for example, [2]).

When $b < \infty$ and $\theta_1(t) = 1 - t/b$, we may also choose $\theta_n(t)$, $n = 2,..., M$ as a polynomial of degree $n - 2$. It is well known [9] that there exist polynomials, e.g., Bernstein polynomials, which simultaneously approximate z and dz/dt as closely as desired for M sufficiently large.

Contrasted to the above, the case $b = \infty$ is less routine, and is examined in the next section.

3. Simultaneous Approximation on $[0, \infty)$: Exponomials and Laguerre Functions

We assume the notation of the previous sections with $b = \infty$. Let $L_2(0, \infty) = \{x \mid \|x\|_2 < \infty\}$, and $\|x\|_\infty = \sup_{t \in (0,\infty)} |x(t)|_\infty$. In this section we investigate the problem of determining S_M such that for $\varepsilon > 0$ there is an M for which $\|W - y\|_2$ and $\|(d/dt)(W - y)\|_2$ are both less than ε for a suitable W in S_M. In determining S_M it clearly suffices to consider only the scalar case $N = 1$.

Lemma 1. *Let x be contained in $C^1(0, \infty)$ and suppose that x and dx/dt vanish at ∞ and belong to $L_2(0, \infty)$. If $\varepsilon > 0$, then there is a g in $C^1(0, \infty)$ and a t_0 such that $g(0) = x(0)$, $g \equiv 0$ for $t \geqslant t_0$, and*

$$\|x - g\|_2 \leqslant \varepsilon \qquad \text{and} \qquad \left\| \frac{d}{dt}(x - g) \right\|_2 \leqslant \varepsilon$$

Proof. Define g to equal x on $[0, t_0 - 1]$ and take $g \equiv 0$ on $[t_0, \infty)$. On $[t_0 - 1, t_0]$ define g to equal the appropriate cubic Hermite polynomial [9] so as to guarantee the continuity of g and dg/dt on $(0, \infty)$. For t_0 sufficiently large the conclusion of the lemma follows from the existence of the improper integrals and the vanishing of x and dx/dt at ∞. Q.E.D.

3.1. *Exponentials*

Now let $\alpha > 0$ and consider the set of polynomials in $e^{-\alpha t}$ which vanish at ∞. These "exponomials" were studied by Duffin and Whidden in [14]. Obviously, $h(t)$ is an exponomial if and only if $h(t) = \sum_{n=1}^{M} c_n e^{-n\alpha t}$ for some constants $c_1, c_2, \ldots, c_M$.

Lemma 2. *Let g be in $C^1(0, \infty)$ and vanish for all t sufficiently large. If $\varepsilon > 0$, then there is an exponential $h(t)$ such that $h(0) = g(0)$, and*

$$\|e^{\alpha t} g - h\|_\infty \leqslant \frac{3\varepsilon}{\alpha} \qquad \text{and} \qquad \left\| e^{\alpha t} \frac{d}{dt}(e^{\alpha t} g - h) \right\|_\infty \leqslant 2\varepsilon.$$

Proof. Let $f(t) = e^{\alpha t}(d/dt)(e^{\alpha t} g(t))$ and $\xi = e^{-\alpha t}$. Then the function $\Phi(\xi)$, defined by setting

$$\Phi(0) = 0, \qquad \Phi(\xi) = f(-(1/\alpha) \ln \xi) = f(t), \qquad \xi \in (0, 1],$$

is continuous on $[0, 1]$. Therefore, by the Weierstrass approximation theorem, Φ can be uniformly approximated by a polynomial of the form $\sum_{n=1}^{M} c_n \xi^n$. Hence, the exponomial $h^*(t) = \sum_{n=1}^{M} c_n e^{n\alpha t}$ satisfies $\|f(t) - h^*(t)\|_\infty \leqslant \varepsilon$.

Set

$$\hat{h}(t) = \sum_{n=1}^{M} -\frac{c_n}{\alpha(n+1)} e^{(n+1)\alpha t}.$$

Then, $h^* = e^{\alpha t}\, d\hat{h}/dt$. Hence,

$$\left\| e^{\alpha t} \frac{d}{dt}(e^{\alpha t} g - \hat{h}) \right\|_\infty = \| f - h^* \|_\infty \leqslant \varepsilon.$$

Moreover,

$$\begin{aligned} |(e^{\alpha t} g(t) - \hat{h}(t)) - (g(0) - \hat{h}(0))| &= \left| \int_0^t \frac{d}{ds}(e^{\alpha s} g - \hat{h})\, ds \right| \\ &\leqslant \int_0^t e^{-\alpha s} \left| e^{\alpha s} \frac{d}{ds}(e^{\alpha s} g - \hat{h}) \right| ds \\ &\leqslant \varepsilon/\alpha. \end{aligned}$$

Since the inequality holds for all t, we must have $|g(0) - \hat{h}(0)| \leqslant \varepsilon/\alpha$ and then $\|e^{\alpha t} g - \hat{h}\|_\infty \leqslant 2\varepsilon/\alpha$. Letting $h(t) = \hat{h}(t) + [g(0) - \hat{h}(0)]\, e^{-\alpha t}$, the conclusion of the lemma readily follows from the above estimates. Q.E.D.

Using Lemma 2, we can obtain estimates in the L_2-norm:

LEMMA 3. *Let g be in $C^1(0, \infty)$ and vanish for all t sufficiently large. If $\varepsilon > 0$, then there is an exponential $h(t)$ such that $h(0) = g(0)$, and*

$$\| g - h \|_2 \leqslant \frac{\varepsilon}{(2\alpha)^{1/2}} \quad \text{and} \quad \left\| \frac{d}{dt}(g - h) \right\|_2 \leqslant \frac{1+\alpha}{(2\alpha)^{1/2}} \varepsilon.$$

Proof. According to Lemma 2, there is an exponomial $h^*(t)$ such that $h^*(0) = g(0)$ and $\|e^{\alpha t} g - h^*\|_\infty \leqslant \varepsilon$ and $\|e^{\alpha t}(d/dt)(e^{\alpha t} g - h^*)\|_\infty \leqslant \varepsilon$. Set $h(t) = e^{-\alpha t} h^*$. Clearly, $h(0) = g(0)$. Also

$$\begin{aligned} \| g - h \|_2 &= \| e^{\alpha t}(e^{\alpha t} g - h^*) \|_2 \\ &= \left[\int_0^\infty e^{2\alpha s} |e^{\alpha s} g - h^*|^2\, ds \right]^{1/2} \leqslant \frac{\varepsilon}{(2\alpha)^{1/2}}. \end{aligned}$$

Furthermore,

$$\left\| \frac{d}{dt}(g-h) \right\|_2 - \|\alpha(g-h)\|_2 \leqslant \left\| \frac{d}{dt}(g-h) + \alpha(g-h) \right\|_2$$

$$\leqslant \left\| e^{\alpha t} \left[\frac{d}{dt}(g-h) + \alpha(g-h) \right] \right\|_2$$

$$= \left\| \frac{d}{dt}(e^{\alpha t} g - h^*) \right\|_2 \leqslant \frac{\varepsilon}{(2\alpha)^{1/2}}$$

Thus,

$$\left\| \frac{d}{dt}(g-h) \right\|_2 \leqslant \frac{\varepsilon}{(2\alpha)^{1/2}} + \frac{\alpha\varepsilon}{(2\alpha)^{1/2}}. \qquad \text{Q.E.D.}$$

By combining the results of Lemma 1 and Lemma 3 we obtain the following theorem on approximation by exponomials:

THEOREM 2. *Let x be contained in $C^1(0, \infty)$ and suppose that x and dx/dt vanish at ∞ and belong to $L_2(0, \infty)$. If $\varepsilon > 0$, then there is an exponomial $h(t)$ such that $h(0) = x(0)$, and*

$$\|x-h\|_2 \leqslant \varepsilon \qquad \text{and} \qquad \left\| \frac{d}{dt}(x-h) \right\|_2 \leqslant \varepsilon.$$

Now let us define

$$\theta_1(t) = e^{-\alpha t}, \qquad \theta_n(t) = e^{-n\alpha t} - e^{-\alpha t}, \quad n = 2, 3, \ldots, M. \tag{15}$$

As shown by the next theorem, the associated set S_M is appropriate for the construction of Time–Galerkin approximations on $[0, \infty)$.

THEOREM 3. *Let $y(t)$ be the unique solution of (1) on $[0, \infty)$ and let $y(\infty)$ be its steady state value. If $y - y(\infty)$ and dy/dt belong to $L_2(0, \infty)$, and if S_M is generated by the functions (15), then for $\varepsilon > 0$ there is an M and a W in S_M such that*

$$\|y-W\|_2 \leqslant \varepsilon \qquad \text{and} \qquad \left\| \frac{d}{dt}(y-W) \right\|_2 \leqslant \varepsilon.$$

Proof. Consider $x(t) = y(t) - y(\infty)$. By Theorem 2, there is an exponential $h(t) = \sum_{n=1}^{M} c_n e^{-n\alpha t}$ such that $y(0) - y(\infty) = \sum_{n=1}^{M} c_n = h(0)$, and $\|x-h\|_2$, $\|(d/dt)(x-h)\|_2 \leqslant \varepsilon$. Hence, if

$$W = y(\infty) + (y_0 - y(\infty))\,\theta_1 + \sum_{n=2}^{M} c_n \theta_n,$$

then

$$\|y - W\|_2 = \left\| (y - y(\infty)) - \sum_{n=1}^{M} c_n e^{-\alpha t} - \sum_{n=2}^{M} c_n (e^{-n\alpha t} - e^{-\alpha t}) \right\|_2$$
$$= \|x - h\|_2 \leqslant \varepsilon,$$

and similarly for $\|(d/dt)(y - W)\|_2$. Q.E.D.

Obviously, we can combine Theorem 1 and Theorem 3 to obtain a convergence result for the exponomial Galerkin approximation:

THEOREM 4. *Under the hypotheses of Theorem* 1 *and Theorem* 3 *if* y *is the true solution to* (1) *and* Y *is the exponomial Time–Galerkin approximation determined by* (3) *and* (15), *then*

$$\int_0^{\infty} |Y - y|^2 \, dt \to 0 \qquad \text{as} \quad M \to \infty.$$

We remark that if (1) is autonomous and the hypotheses of Theorem 1 hold, then $y - y(\infty)$ and dy/dt automatically belong to $L_2(0, \infty)$. That $y - y(\infty)$ belongs to $L_2(0, \infty)$ follows from the remarks after the statement of Theorem 1. Furthermore,

$$\left| \frac{dy}{dt} \right|_2 = |F(y) - F(y(\infty))|_2 \leqslant \Gamma |y - y(\infty)|_2,$$

where $\Gamma = \sup_{x \in D} |F_y(x)|_2$. Therefore, since $y - y(\infty) \in L_2(0, \infty)$,

$$\left\| \frac{dy}{dt} \right\|_2 \leqslant \Gamma \|y - y(\infty)\|_2 < \infty.$$

3.2. *Laguerre Functions*

A second set of expansion functions appropriate to the infinite interval results from considering the Laguerre functions. These are functions of the form $p(t)\, e^{-\alpha t}$, where $p(t)$ is a polynomial and $\alpha > 0$. According to Stone [28, Theorem 18], any continuous function which vanishes at ∞ can be uniformly approximated by a Laguerre function. This leads to

LEMMA 4. *Let* g *be contained in* $C^1(0, \infty)$ *and vanish for all* t *sufficiently large. If* $\varepsilon > 0$, *then there is a Laguerre function* $p(t)\, e^{-2\alpha t/3}$ *such that* $p(0) = g(0)$, *and*

$$\|e^{\alpha t/3} g - p e^{-2\alpha t/3}\|_{\infty} \leqslant \frac{9\varepsilon}{\alpha}$$

and

$$\left\| e^{\alpha t/3} \frac{d}{dt} (e^{\alpha t/3} g - p e^{-2\alpha t/3}) \right\|_\infty \leqslant 3\varepsilon.$$

Proof. Choose [28, Theorem 18] the polynomial $q(t)$ such that

$$\left\| e^{\alpha t/3} \frac{d}{dt} (e^{\alpha t/3} g) - q e^{-\alpha t/3} \right\|_\infty \leqslant \varepsilon,$$

and note that there is a polynomial $p_\varepsilon(t)$ such that $(d/dt)(p_\varepsilon e^{-2\alpha t/3}) = q e^{-2\alpha t/3}$. Hence

$$\left\| e^{\alpha t/3} \frac{d}{dt} (e^{\alpha t/3} g - p_\varepsilon e^{-2\alpha t/3}) \right\|_\infty$$
$$= \left\| e^{\alpha t/3} \frac{d}{dt} (e^{\alpha t/3} g) - q e^{-\alpha t/3} \right\|_\infty \leqslant \varepsilon.$$

Moreover,

$$|(e^{\alpha t/3} g(t) - p_\varepsilon(t)\, e^{-2\alpha t/3}) - (g(0) - p_\varepsilon(0))|$$
$$= \left| \int_0^t \frac{d}{ds} (e^{\alpha s/3} g - p_\varepsilon e^{-2\alpha s/3})\, ds \right| \leqslant \varepsilon \int_0^t e^{-\alpha s/3}\, ds \leqslant \frac{3\varepsilon}{\alpha}.$$

As in the proof of Lemma 2, this implies that $|g(0) - p_\varepsilon(0)| \leqslant 3\varepsilon/\alpha$ and that $\| e^{\alpha t/3} g - p_\varepsilon(t)\, e^{-2\alpha t/3} \|_\infty \leqslant 6\varepsilon/\alpha$. Now let $p(t) = p_\varepsilon(t) + g(0) - p_\varepsilon(0)$. Then $p(0) = g(0)$ and

$$\left\| e^{\alpha t/3} \frac{d}{dt} (e^{\alpha t/3} g - p e^{-2\alpha t/3}) \right\|_\infty$$
$$= \left\| e^{\alpha t/3} \frac{d}{dt} (e^{\alpha t/3} g) - q e^{-\alpha t/3} + \frac{2\alpha}{3} (g(0) - p_\varepsilon(0))\, e^{-\alpha t/3} \right\|_\infty \leqslant 3\varepsilon.$$

Finally, it follows from the above estimates that $\| e^{\alpha t/3} g - p e^{-2\alpha t/3} \|_\infty \leqslant 9\varepsilon/\alpha$. Q.E.D.

The L_2 estimates now follow as before.

LEMMA 5. *Let g be contained in $C^1(0, \infty)$ and vanish for all t sufficiently large. If $\varepsilon > 0$, then there is a Laguerre function $p(t)\, e^{-\alpha t}$ such that $p(0) = g(0)$, and*

$$\| g - p e^{-\alpha t} \|_2 \leqslant \varepsilon \left(\frac{3}{2\alpha} \right)^{1/2}$$

and

$$\left\| \frac{d}{dt}(g - pe^{-\alpha t}) \right\|_2 \leqslant \varepsilon \left(1 + \frac{\alpha}{3}\right)\left(\frac{3}{2\alpha}\right)^{1/2}.$$

Proof. By Lemma 4 there is a Laguerre function $p(t)e^{-2\alpha t/3}$ such that $g(0) = p(0)$, $\|e^{\alpha t/3}g - pe^{-2\alpha t/3}\|_\infty \leqslant \varepsilon$, and also $\|e^{\alpha t/3}(d/dt)(e^{\alpha t/3}g - pe^{-2\alpha t/3})\|_\infty \leqslant \varepsilon$. Then, $\|g - pe^{-\alpha t}\|_2 = \|e^{-\alpha t/3}(e^{\alpha t/3}g - pe^{-2\alpha t/3})\|_2 \leqslant \varepsilon(3/2\alpha)^{1/2}$. Also,

$$\begin{aligned}
&\left\| \frac{d}{dt}(g - pe^{-\alpha t}) \right\|_2 - \left\| \frac{\alpha}{3}(g - pe^{-\alpha t}) \right\|_2 \\
&\quad \leqslant \left\| \frac{d}{dt}(g - pe^{-\alpha t}) + \frac{\alpha}{3}(g - pe^{-\alpha t}) \right\|_2 \\
&\quad \leqslant \left\| e^{\alpha t/3}\left[\frac{d}{dt}(g - pe^{-\alpha t}) + \frac{\alpha}{3}(g - pe^{-\alpha t}) \right] \right\|_2 \\
&\quad = \left\| \frac{d}{dt}(e^{\alpha t/3}g - pe^{-2\alpha t/3}) \right\|_2 \leqslant \varepsilon \left(\frac{3}{2\alpha}\right)^{1/2}.
\end{aligned}$$

Thus,

$$\left\| \frac{d}{dt}(g - pe^{-\alpha t}) \right\|_2 \leqslant \varepsilon \left(1 + \frac{\alpha}{3}\right)\left(\frac{3}{2\alpha}\right)^{1/2}. \qquad \text{Q.E.D.}$$

The following analog of Theorem 2 is an immediate consequence of Lemma 1 and Lemma 5.

THEOREM 5. *Let x be in $C^1(0, \infty)$ and suppose that x and dx/dt vanish at ∞ and belong to $L_2(0, \infty)$. If $\varepsilon > 0$, then there is a Laguerre function $p(t)e^{-\alpha t}$ such that $p(0) = x(0)$, and*

$$\|x - pe^{-\alpha t}\|_2 \leqslant \varepsilon, \qquad \left\| \frac{d}{dt}(x - pe^{-\alpha t}) \right\|_2 \leqslant \varepsilon.$$

If we now define the expansion functions

$$\theta_n(t) = t^{n-1}e^{-\alpha t}, \qquad n = 1, 2, 3, \ldots, M, \tag{16}$$

then the corresponding set S_M generated by the functions in (16) is again proper for Time–Galerkin approximations on $[0, \infty)$.

THEOREM 6. *Let $y(t)$ be the unique solution of* (1) *on $[0, \infty)$ and let $y(\infty)$ be its steady state value. If $y - y(\infty)$ and dy/dt both belong to*

$L_2(0, \infty)$ and if S_M is generated by the functions in (16), *then for $\varepsilon > 0$, there is an M and a W in S_M such that $\|y - W\|_2 \leqslant \varepsilon$ and $\|(d/dt)(y - W)\|_2 \leqslant \varepsilon$.*

Proof. In Theorem 5 set $x(t) \equiv y(t) - y(\infty)$, and let the associated Laguerre function be $\sum_{n=1}^{M} c_n t^{n-1} e^{-\alpha t} \equiv \sum_{n=1}^{M} c_n \theta_n(t)$. Since $c_1 = y(0) - y(\infty)$, the conclusion is immediate with

$$W(t) = y(\infty) + (y(0) - y(\infty))\,\theta_1(t) + \sum_{n=2}^{M} c_n \theta_n(t). \qquad \text{Q.E.D.}$$

Combining Theorem 1 and Theorem 6 we obtain the following convergence result for the Laguerre Galerkin approximation:

THEOREM 7. *Under the hypotheses of Theorem* 1 *and Theorem* 6 *if y is the true solution of* (1) *and Y is the Laguerre Time–Galerkin approximation determined by* (3) *and* (16), *then*

$$\int_0^{\infty} |Y - y|^2 \, dt \to 0 \qquad \text{as} \quad M \to \infty.$$

4. An Application to a Linear Reactor Kinetics Problem

A simple system of ode's which arises in nuclear reactor kinetics is [8, 25]

$$dy/dt = -Ay, \qquad t > 0, \tag{17}$$

where

$$y = (y_0, y_1, \ldots, y_6)^T$$

and

$$-A = \begin{bmatrix} (\rho - \beta)/\Lambda & \lambda_1 & \lambda_2 & \cdots & \lambda_6 \\ \beta_1/\Lambda & -\lambda_1 & & \ddots & \\ \vdots & & \ddots & & \\ \beta_6/\Lambda & 0 & & & -\lambda_6 \end{bmatrix}.$$

Here Λ, λ_i, β_i, $i = 1, 2, \ldots, 6$, are positive numbers, and $\beta = \sum_{i=1}^{6} \beta_i$. The initial condition

$$y(0) = \left(1, \frac{\beta_1}{\Lambda\lambda_1}, \frac{\beta_2}{\Lambda\lambda_2}, \ldots, \frac{\beta_6}{\Lambda\lambda_6}\right)^T \tag{18}$$

is such that when the reactivity $\rho = 0$, $y(0)$ is the steady state solution. We consider the following data from a thermal reactor [8]:

i	λ_i	β_i
1	3.87	1.950 (−4)
2	1.40	9.600 (−4)
3	0.311	3.052 (−3)
4	0.115	1.410 (−3)
5	0.0317	1.597 (−3)
6	0.0127	2.850 (−4)

$\Lambda = 5.0\ (-4)$, $\beta = 7.5\ (-3)$.

Moreover, we assume $\rho = -0.07$. This defines a subcritical state of the reactor.

It is known [25] that when $\rho < 0$, the matrix $-A$ has all real negative eigenvalues and hence $y(\infty) = (0, 0, \ldots, 0)^T$. Furthermore, according to the results of the Appendix, the Time–Galerkin approximation is uniquely defined by Eq. (3).

For the data given one finds that, rounded to five places, the eigenvalues of $-A$ are: $\mu_0 = -155.04$, $\mu_1 = -3.8601$, $\mu_2 = -1.3827$, $\mu_3 = -0.29867$, $\mu_4 = -0.11275$, $\mu_5 = -0.03099$, $\mu_6 = -0.01265$. Thus, the system is moderately stiff.

To apply Theorem 1 to the linear system (17) we need to verify that $x^T A x \geqslant \gamma x^T x$ for some $\gamma > 0$ and all $x \in R^7$. Since A is a matrix of constants, this is equivalent to the condition that $\tilde{A} \equiv (A + A^T)/2$ has only positive eigenvalues. For the given data, this is the case and is a consequence of the following general result:

THEOREM 8. *Let A be the coefficient matrix of* (17). *Then $\tilde{A}$ has positive eigenvalues if and only if*

$$\rho < -\frac{\Lambda}{4} \sum_{i=1}^{6} (\beta_i \Lambda^{-1} - \lambda_i)^2 / \lambda_i \tag{19}$$

Proof. From the eigenvector equation

$$\tilde{A}x = \omega x, \tag{20}$$

$x = (x_0, x_1, \ldots, x_6)^T$, we deduce that

$$x_i = -\frac{\beta_i \Lambda^{-1} + \lambda_i}{2(\omega - \lambda_i)} x_0, \qquad i = 1, 2, \ldots, 6. \tag{21}$$

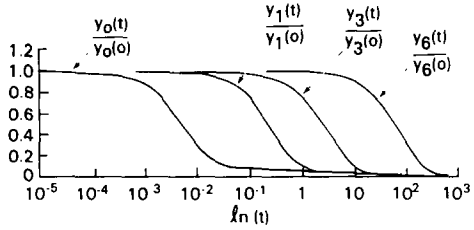

FIG. 1. Profiles for kinetics problem.

Note that $\omega = \lambda_i$ implies that $x = 0$, so the above division is proper. Substituting the expressions (21) in the equation of (20) which occurs when $i = 0$, we find that $f(\omega)\, x_0 = 0$, where

$$f(\omega) = \frac{\beta - \rho}{A} - \omega + \frac{1}{4} \sum_{i=1}^{6} \left(\frac{(\beta_i A^{-1} + \lambda_i)^2}{\omega - \lambda_i} \right). \tag{22}$$

Since we cannot have $x_0 = 0$, we conclude that the eigenvalues of $\tilde{A}$ are the zeros of $f(\omega)$. But it is easy to see that six of the zeros of $f(\omega)$ are always positive, and furthermore that the remaining zero will be positive if and only if $f(0) > 0$. After simplification this condition reduces to (19). Q.E.D.

Remark. According to (19), $\tilde{A}$ has only positive eigenvalues if and only if $\rho < -0.06437....$

The exact solution to (17) (see Fig. 1) can be obtained by use of the matrix exponential

$$y(t) = e^{At} y(0) = P^{-1} e^{Dt} P y(0) = B(e^{\mu_0 t}, e^{\mu_1 t}, ..., e^{\mu_6 t})^T, \tag{23}$$

where D is the diagonal matrix of eigenvalues of A and the columns of the matrix P are the corresponding eigenvectors. The matrix B in (23) is strongly diagonally dominant and this results in the component $y_i(t)$ being driven by the term $e^{\mu_i t}$ in (23).

Time–Galerkin approximations to (17) were calculated from the exponomial space S_M described in Section 3.1. For this example we chose a decay constant $\alpha = 0.01$ so that each component $y_i(t)$ in (17) is represented by an approximation of the form

$$Y_i(t) = \sum_{n=1}^{M} c_n e^{-0.01nt}, \qquad 0 \leqslant i \leqslant 6.$$

We define the component error e_i by

$$e_i \equiv \sup_{[0,\infty]} \left| \frac{Y_i(t) - y_i(t)}{y_i(0)} \right|, \qquad 0 \leqslant i \leqslant 6.$$

Table I shows the convergence behavior of e_i with respect to M.

TABLE I

Convergence of Exponomials for Kinetics Problem

M	e_0	e_2	e_4	e_6
2	9.03 (1)	8.40 (−1)	4.28 (−1)	2.80 (−2)
4	8.99 (1)	6.96 (−1)	1.16 (−1)	7.16 (−3)
6	8.94 (−1)	5.40 (− 1)	1.28 (−2)	3.14 (3)
8	8.88 (− 1)	4.11 (−1)	1.70 (−3)	1.60 (−3)
10	8.82 (−1)	3.11 (−1)	1.08 (−3)	8.21 (−4)

As illustrated in Table I, convergence to the slowly decaying components (y_3 to y_6) is reasonably fast while convergence to the rapidly decaying components is painfully slow.

In a simple linear problem such as (17) it is expedient (particularly for stiff problems) to use Time–Galerkin approximations of the form

$$Y_i(t) = \sum_{n=1}^{M} c_n e^{-\sigma_n t}, \tag{24}$$

where $\{\sigma_n\}_{n=1}^{M}$ is close in some sense to $\{\mu_n\}_{n=0}^{6}$. Suppose, for example, we estimate μ_n by σ_n^+ and σ_n^-, where

$$\sigma_n^+ = \mu_{n-1}(1+P), \qquad \sigma_n^- = \mu_{n-1}(1-P), \qquad \sigma \leqslant P < 1, \quad 1 \leqslant n \leqslant 7. \tag{25}$$

Hence, if $P = 0.5$, then σ_n^+ is a 50% overestimate of μ_{n-1} while σ_n^- is a 50% underestimate of μ_{n-1}. Sometimes such estimates are known a priori or can be made by inspection. (For example, the Gerschgorin circle theorem [29] tells us that the matrix A in (17) has exactly one eigenvalue in the interval $[-160.7404, -149.2596]$). In Table II we give the results of approximating $y(t)$ in (17) by the Time–Galerkin approximation (24), where the decay rates σ_n are determined from $\{\sigma_n^+, \sigma_n^-\}_{n=1}^{7}$ in (25). Table II indicates that the

TABLE II

Time–Galerkin Approximation to Kinetics Problem Using (24), (25)

P (%)	e_0	e_2	e_4	e_6
0	0	0	0	0
5	4.55 (−4)	4.55 (−5)	2.47 (−5)	8.33 (−5)
10	1.82 (−3)	1.83 (−4)	9.64 (−5)	3.26 (−4)
30	1.65 (−2)	1.41 (−3)	6.97 (−4)	2.26 (−3)
50	4.65 (−2)	2.17 (−3)	9.06 (−4)	2.40 (−3)
80	1.24 (−1)	1.36 (−2)	3.41 (−3)	9.27 (−3)

slower decaying components of $y(t)$ are more forgiving of over- and underestimation of the eigenvalues than are the rapidly decaying components. Note also that estimates of the eigenvalues which are in error by as much as 50% still produce numerical approximations to $y(t)$ with less than 5% relative error for all $t \geqslant 0$.

5. Applications to Parabolic Initial-Boundary Value Problems

Let us construct in the manner of [11, 12, 26] a semi-discrete approximation to the initial-boundary value problem

$$\frac{\partial u}{\partial t} = L[u] + f(u), \qquad 0 < x < 1, \quad t > 0, \tag{26}$$

$$\alpha_0 \frac{\partial u}{\partial x}(0, t) + \beta_0 u(0, t) = \delta_0, \qquad \alpha_1 \frac{\partial u}{\partial x}(1, t) + \beta_1 u(1, t) = \delta_1, \tag{27}$$

$$u(x, 0) = g(x), \qquad 0 < x < 1, \tag{28}$$

where $L[u] = \nabla(a(x)\nabla u)$, and there exist constants η and c_0 such that $0 < \eta \leqslant a(x) \leqslant c_0$ for all $0 < x < 1$. The α_i, β_i, and δ_i are constants, and the function f is assumed to be continuously differentiable with respect to u.

We consider basis functions $\{\varphi_i(x)\}_{i=0}^{N+1}$ which are piecewice linear and continuous on the uniform mesh of gauge $h = 1/(N+1)$ such that

$$\varphi_i(x_j) = \delta_j^i, \qquad 0 \leqslant i, \quad j \leqslant N+1. \tag{29}$$

The *semi-discrete Galerkin* approximation is defined by

$$\hat{u}(x, t) = \sum_{i=0}^{N+1} \hat{u}_i(t)\, \varphi_i(x),$$

where the $\hat{u}_i(t)$ are determined from applying integration by parts to

$$\int_0^1 \left(\frac{\partial u}{\partial t} - L[u] - f(u) \right) \varphi_i(x)\, dx = 0, \qquad 0 \leqslant i \leqslant N+1, \tag{30}$$

and replacing the u by $\hat{u}$. This approach is an example of the method of lines [24] and the equations for $i = 0$ and $i = N+1$ may need to be modified to reflect the boundary conditions (27). This leads to a system of ordinary differential equations of the form (1) [12, 24, 26]. Specifically, homogeneous Dirichlet boundary conditions in (27) yield (1) with

$$F(t, y) \equiv -T_1^{-1}(T_2 y + S(y)), \tag{31}$$

where $y = (\hat{u}_1(t), \hat{u}_2(t), \ldots, \hat{u}_N(t))^T$, T_1 and T_2 are $N \times N$ matrices with entries $[T_1]_{ij} = \int_0^1 \varphi_i(x)\, \varphi_j(x)\, dx$ and $[T_2]_{ij} = \int_0^1 a(x)\, \varphi_i'(x)\, \varphi_j'(x)\, dx$. For $a(x) \equiv 1$ we

have $T_1 = h/6$ Tridiag$\{1, 4, 1\}$ and $T_2 = 1/h$ Tridiag$\{-1, 2, -1\}$. The N-vector $S(y)$ has as its ith component $[S(y)]_i = \int_0^1 f(\hat{u})\,\varphi_i(x)\,dx$. (We could also use finite difference [27] or collocation methods [13] to obtain similar spatial discretizations.) We recall that it is well known [11, 12, 26] that under suitable conditions $\|\hat{u} - u\|_T \to 0$ as $N \to \infty$, where $\|\cdot\|_T^2 = \int_0^T \int_0^1 |\hat{u} - u|^2\,dx\,dt$.

For ease of exposition we assume Dirichlet boundary conditions in (27). As described in Section 1, the infinite element approximation [3] results when Galerkin's method is applied again, this time with respect to the t-variation to obtain approximations to the functions $\hat{u}_i(t)$, $1 \leqslant i \leqslant N$. In the notation of Section 1, $y(t) \equiv (\hat{u}_1(t), \hat{u}_2(t), \ldots, \hat{u}_N(t))^T$ is approximated by $Y(t) \equiv y(\infty) + \sum_{n=1}^{M} c_n \theta_n(t)$, where the vectors c_n are determined by (3). The resulting *Time–Galerkin approximation* to $\hat{u}(x, t)$, or *infinite element approximation* to u, is

$$u(M; x, t) \equiv \sum_{i=0}^{N+1} Y_i(t)\,\varphi_i(x), \tag{32}$$

where $Y_0(t) = u(0, t)$ and $Y_{N+1}(t) = u(1, t)$.

The convergence of the infinite element method is then extablished by combining Theorems 4 and 7 of the present paper with Theorem 3.1 of [12] to obtain:

THEOREM 9. *Let $u(x, t)$ be the solution of* (26)–(28) *and let $\hat{u}(x, t)$ be the semi-discrete Galerkin approximation to u defined by* (30). *Further, let $u(M; \cdot, \cdot)$ be the infinite element approximation to u as given in* (32) *with $\theta_n(t)$ chosen according to* (15) *or* (16). *If $F(t, y)$ in* (31) *is uniformly monotone, then for arbitrary $\varepsilon > 0$ and $T > 0$, there exists an N and M_0 such that for all $M > M_0$,*

$$\|u - u(M; \cdot, \cdot)\|_T < \varepsilon. \tag{33}$$

Proof. By the triangle inequality,

$$\|u - u(M; \cdot, \cdot)\|_T \leqslant \|\hat{u} - u\|_T + \|\hat{u} - u(M; \cdot, \cdot)\|_T.$$

Given $\varepsilon > 0$, from [12, Theorem 3.1], there exists an N such that $\|\hat{u} - u\|_T \leqslant \varepsilon/2$.

Now for a fixed, but arbitrary, $t = t^*$ and for all x, $0 \leqslant x \leqslant 1$,

$$|\hat{u} - u(M; x, t^*)|^2 = \left| \sum_{i=0}^{N+1} (y_i(t^*) - Y_i(t^*))\,\varphi_i(x) \right|^2$$

$$\leqslant \max_j |y_j(t^*) - Y_j(t^*)|^2 \leqslant |y(t^*) - Y(t^*)|_2^2$$

since $1 \geqslant \varphi_i(x) \geqslant 0$ and $\sum_{i=0}^{N+1} \varphi_i(x) = 1$. Hence

$$\|\hat{u} - u(M; \cdot, \cdot)\|_T^2 = \int_0^T \int_0^1 |\hat{u} - u(M; x, t^*)|^2 \, dx \, dt^*$$

$$\leqslant \int_0^T |y(t^*) - Y(t^*)|_2^2 \, dt^*$$

$$\leqslant \int_0^\infty |y(t^*) - Y(t^*)|_2^2 \, dt^* \equiv \|Y - y\|_2^2.$$

Finally, by Theorem 4 or Theorem 7, there exists an M_0 such that for all $M > M_0$,

$$\|\hat{u} - u(M; \cdot, \cdot)\|_T \leqslant \|Y - y\|_2 \leqslant \varepsilon/2. \qquad \text{Q.E.D.}$$

Remark. In the notation of Section 1 we now have $F_y(t, y) = T_1^{-1}(T_2 - S_y(y))$. But, with $u^* \equiv u(M; \cdot, t)$,

$$[S_y(y)]_{ij} = \int_0^1 \frac{df}{du}(u^*(x)) \, \varphi_i(x) \, \varphi_j(x) \, dx$$

$$= \frac{df}{du}(u^*(\xi_{ij})) \int_0^1 \varphi_i(x) \, \varphi_j(x) \, dx$$

for some intermediate value ξ_{ij}. If df/du is uniformly bounded, then $[S_y(y)]_{ij}$ is $O(h)$ as $h \to 0$. Combining this with $[T_2]_{ij} = O(1/h)$, we have for h sufficiently small $F_y(t, y) \sim T_1^{-1} T_2$.

Now, for example, if $a(x) \equiv 1$, $T_1^{-1} T_2$ is a positive definite matrix whose eigenvalues are bounded below by 4 [5, (2.30)]. That is, F is uniformly monotone.

If df/du is not uniformly bounded, then we can define a new source term $\tilde{f}$ as the indefinite integral of $\tilde{f}'$, where for L sufficiently large

$$\tilde{f}' = \frac{df}{du} \qquad \text{if} \quad |u| \leqslant K$$

$$= \frac{df}{du}(K) \qquad \text{if} \quad u > K$$

$$= \frac{df}{du}(-K) \qquad \text{if} \quad u < -K.$$

As above, the conditions of Theorem 1 are met, and the Time–Galerkin approximation $u(M; x, t)$ converges to the solution of this modified problem. Now, if the modified problem has solution $|w| \leqslant K$, then in fact $w = u$, otherwise we choose a larger K. If u is bounded, then such a K must exist.

We illustrate the kind of computational problem which arises for the simple linear problem

$$\frac{\partial u}{\partial t} = L|u| + f(x), \qquad L|u| \equiv \frac{\partial^2 u}{\partial x^2},$$
$$u(0, t) = 0, \qquad u(1, t) = 0, \tag{34}$$
$$u(x, 0) = g(x)$$

if, as was done in [3], we combine both applications of Galerkin's method. With $Y_i(t) = u(x_i, \infty) + (g(x_i) - u(x_i, \infty))\theta_1(t) + \sum_{j=2}^{M} a_{ij}\theta_j(t)$, $0 \leqslant i \leqslant N+1$, where $a_{0j} = a_{N+1,j} = 0$ from (34), we find that the other a_{ij} are determined by the system of equations

$$\sum_{i=1}^{N}\sum_{j=2}^{M} a_{ij}\int_0^1\int_0^\infty \{\varphi_i\dot{\theta}_j\varphi_k\theta_l + \varphi_i'\theta_j\varphi_k'\theta_l\}\,dx\,dt$$
$$= \int_0^1\int_0^\infty \left\{-\frac{\partial v}{\partial t}\varphi_k\theta_l - \frac{\partial v}{\partial x}\varphi_k'\theta_l + f(x)\varphi_k\theta_l\right\} dx\,dt.$$
$$1 \leqslant k \leqslant N, \quad 2 \leqslant l \leqslant M, \tag{35}$$

where φ_i and θ_l are functions of x and t, respectively, the "dot" and "prime" mean differentiation, and

$$v(x, t) = \sum_{i=0}^{N+1} \{u(x_i, \infty) + (g(x_i) - u(x_i, \infty))\theta_1(t)\}\varphi_i.$$

If we order the unknowns as $\bar{a} \equiv (a_{12}, a_{13}, \ldots, a_{1M}, a_{22}, a_{23}, \ldots, a_{2M}, \ldots, a_{N2}, a_{N3}, \ldots, a_{NM})^T$, then (35) becomes

$$\mathcal{A}\bar{a} = \bar{b}, \tag{36}$$

where $\mathcal{A}$ is an $N(M-1)$ by $N(M-1)$ banded unsymmetric matrix. Moreover, if $\bar{a}$ is partitioned as $\bar{a} = (\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_N)^T$, where $\bar{a}_i = (a_{i2}, a_{i3}, \ldots, a_{iM})^T$, then $\mathcal{A}$ has an associated block tridiagonal partitioning of the form

$$\mathcal{A} = \begin{bmatrix} B & C & & & \\ C & B & C & & \\ & \ddots & \ddots & \ddots & \\ & & C & B & C \\ & & & C & B \end{bmatrix}, \tag{37}$$

where B and C are $(M-1)$ by $(M-1)$ full matrices. Block Gauss elimination [22] proved to be an efficient method for solving (36).

For the linear problem (34) the system of ordinary differential equations (30) determining the semi-discrete Galerkin approximation is actually of the form

$$T_1 \frac{dy}{dt} = -T_2 y + k,$$

where as before T_1 and T_2 are symmemtric, tridiagonal, positive definite matrices. To conform to (1) we should multiply through by T_1^{-1} to obtain a system of the form (1). Then the matrix $P = T_1^{-1}T_2$ cannot have any pure imaginary eigenvalues. For if $Pz = \lambda z$, where $\lambda = -\bar{\lambda}$, then we easily obtain $z^*T_2 z = \lambda z^*T_1 z$ and $z^*T_2 z = -\bar{\lambda} z^*T_1 z$. (Here z^* denotes conjugate transpose of z.) Thus $z^*T_2 z = 0$, and this is a contradiction. It follows from a result of the Appendix that the associated Time–Galerkin system (3) has a unique solution. Since system (36) is equivalent to that of (3), C is nonsingular and the use of (36) avoids the need to invert T_1.

The use of Galerkin's method is a straightforward matter when (26)–(28) represents a linear problem. When (26) is nonlinear, several practical questions of implementation arise which we now discuss.

Estimating the steady state, $u(x, \infty)$. In developing the finite parameter representation in (2) it was assumed that the steady state solution $u(x, \infty)$ of (26)–(27) exists and could be found or accurately approximated, say, by solving the steady state two-point boundary value problem. If $f(u)$ is nonlinear in u, then it may not be a simple computational task to estimate $u(x, \infty)$. Perhaps what is more important is that for nonlinear problems, (26)–(27) may have multiple steady state solutions. The steady state reached is governed by the initial distribution $g(x)$ in (28). If there are several steady states, then the task of approximating that particular steady state which corresponds to the given initial data may be difficult. This will provide some limit to the class of nonlinear problems for which the proposed approach is useful.

Choice of expansion functions. We have shown in Section 3 that choosing the expansion functions to be exponomials or Laguerre functions leads to convergent schemes. However, there are other choices which reflect the fact that $F(t, y)$ in (1) may involve the discretization of a differential operator about which prior information may be known (see Section 5) These latter choices enhance the accuracy of the approximation.

For linear problems, $f(u) = au$ in (26), one choice of expansion functions $\theta_j(t)$ is particularly appropriate. In this case the true solution to (26)–(28) can be represented by

$$u(x, t) = u(x, \infty) + \sum_{n=1}^{\infty} a_n \omega_n(x)\, e^{(\lambda_n^2 - a)t},$$

where λ_n^2 and $\omega_n(x)$ are eigenvalues and eigenfunctions of the problem

$$L[\omega] + \lambda^2\omega = 0,$$
$$\alpha_0 \frac{\partial\omega}{\partial x}(0) + \beta_0\omega(0) = 0, \qquad \alpha_1 \frac{\partial\omega}{\partial x}(1) + \beta_1\omega(1) = 0. \tag{38}$$

In the linear case, then, we select the first M expansion functions $\gamma_j(t) = \{e^{-(\lambda_j^2-\alpha)t}\}_{j=1}^M$, where $\lambda_1 < \lambda_2 < \lambda_3 \cdots < \lambda_M$. The functions $\theta_j(t)$ are then constructed as

$$\theta_1(t) = \frac{\gamma_1(t)}{\gamma_1(0)}, \qquad \theta_j(t) = \gamma_j(t) - \gamma_j(0)\,\theta_1(t), \qquad j = 2, 3, \ldots, M. \tag{39}$$

If the λ's are not known exactly, they can be approximated by approximating the eigenvalue problem (38). See [10] for other spectral matching schemes.

For nonlinear problems the choice of expansion functions is not so simple. Experience has shown that selection of the eigenvalues of the linear problem (38) offers little. An approach which has proven useful in practice is predicated upon the assumption that (26) holds at $t = 0$. If this is the case, $(\partial u/\partial t)(x, 0)$ can be evaluated from (26) and the initial distribution (28). We use decaying exponentials $\{e^{-\mu_j t}\}_{j=1}^M$ to represent the t-variation of $u(x, t)$. If $M = 1$, then

$$\hat{u}(x_i, t) = u(x_i, 0)\, e^{-\mu_1 t} + u(x_i, \infty)(1 - e^{-\mu_1 t}) \tag{40}$$

represents the time variation of the Galerkin approximation at the mesh point $x = x_i$. We determine μ_1 by solving

$$\frac{d}{dt}[\hat{u}(x_i, t)]_{t=0} = \mu_1[u(x_i, \infty) - u(x_i, 0)] = \frac{\partial u}{\partial t}(x_i, 0) \tag{41}$$

for some mesh point x_i (we have found it best to choose that value of x_i which gives the smallest value of μ_1 in (41)). Equation (41) will yield a positive value of μ_1 if $u(x, t)$ is monotone in t for any x in $[0, 1]$. Such is the case for the nonlinear example we consider here. Once μ_1 has been defined, we can define $\mu_2, \mu_3, \ldots, \mu_M$ to be positive numbers in $[\mu_1 - \varepsilon, \mu_1 + \varepsilon]$, where ε is some fraction of μ_1. This method of selecting decay constants for the exponential expansion functions is somewhat arbitrary; however, it is based upon known information about the time variation of $u(x, t)$ at $t = 0$.

5.1. *Numerical Examples*

Let $u(M; x, t)$ denote the infinite element approximation based on M expansion functions in (39). To study the accuracy of the approximation

$u(M; x, t)$ to $u(x, t)$ we consider either the uniform time error at a mesh point x_i,

$$\|u(M; x_i, \cdot) - u(x_i, \cdot)\|_{L_\infty[0,\infty)} \equiv \sup_{0 \leqslant t < \infty} |u(M; x_i, t) - u(x_i, t)|, \tag{42}$$

or the maximum relative mesh point error $\varepsilon(t)$,

$$\varepsilon(t) \equiv \max_{x_i} |(\hat{u}(M; x_i, t) - u(x_i, t))/u(x_i, t)|, \qquad t \geqslant 0. \tag{43}$$

EXAMPLE 1. Consider the simple linear problem

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \quad t > 0,$$

$$u(0, t) = u(1, t) = 0, \qquad t > 0,$$

$$u(x, 0) = \sin^2 \pi x, \qquad 0 < x < 1.$$

Solution:

$$u(x, t) = \sum_{j=1}^{\infty} c_j e^{-(2j-1)^2 \pi^2 t}.$$

We discretize the spatial variable by a uniform mesh of $N + 1 = 50$ mesh points. Several sets of expansion functions $\gamma_j(t)$ were used to generate infinite element approximations and the detailed results are reported in [3, 4]. We include here Table III, which summarizes the convergence results for $x = 0.5$. This example satisfies the conditions of Theorem 9, so convergence is assured for the case of exponomials and Laguerre functions.

Our next example, similar to those considered in [19], is a nonlinear problem in spherical geometry for which it is not obvious that the conditions in Theorem 9 are satisfied. However, it appears as though convergence is still achieved.

EXAMPLE 2.

$$\frac{\partial u}{\partial t} = \frac{1}{x^2} \frac{\partial}{\partial x} \left(x^2 \frac{\partial u}{\partial x} \right) + u^5, \qquad 0 < x < 1, \quad t > 0, \tag{44}$$

$$\frac{\partial u}{\partial x}(0, t) = 0, \qquad u(1, t) = 3^{1/2}/4, \qquad t > 0, \tag{45}$$

$$u(x, 0) = 3^{1/2}/4. \tag{46}$$

One steady state of (44), (45) is $u(x, \infty) = 1/(1 + x^2/3)^{1/2}$, which is also the steady state solution of (44)–(46). For this example we used exponomial

TABLE III

Infinite Element Approximation Results for Example 1

$\gamma_j(t)$, $1 \leqslant j \leqslant M$	M	$\|u(M; 0.5, \cdot) - u(0.5, \cdot)\|$
$e^{(2j-1)\pi t}$	2	5.38 (−3)
	3	5.35 (−4)
	4	1.20 (−4)
$e^{j\pi t}$	2	3.13 (−2)
	3	3.71 (−3)
	4	5.53 (−4)
e^{-t}	2	4.50 (−1)
	3	2.65 (−1)
	4	1.59 (−1)
$t^{j-1}e^{-t}$	2	5.54 (−1)
	3	4.50 (−1)
	9	1.82 (−1)
	15	1.11 (−1)
	25	6.74 (−2)

expansion functions of the form $\{e^{-jt}\}_{j=1}^{M}$, that is, $\alpha = 1$ in (15). This choice of α comes from solving (41) to obtain $\mu_1 \sim 3$. When $M \geqslant 3$, the exponomial e^{-3t} is contained in the expansion set. Newton's method was used to solve the nonlinear algebraic equations coming from the Time–Galerkin formulation. The matrix equations that must be solved during iteration in Newton's method are block tridiagonal with each block a $(M-1) \times (M-1)$ matrix (cf. (37)). These well-structured matrix equations were solved by block Gauss elimination [29]. For assessment of accuracy, the exact solution to the transient problem (44)–(46) was approximated as a piecewise linear semi-discrete (continuous time-discrete space) Galerkin approximation defined on a fine spatial mesh of 200 points. The associated system of ode's was integrated using the GEARIB code [20].

The numerical results using exponomial expansion functions, $\{e^{-jt}\}_{j=1}^{4}$, for the Time–Galerkin approximations are shown in Figs. 2 and 3. In Fig. 2 we have plotted $\varepsilon(t)$, the maximum relative error incurred at the mesh points for a given t (that is, the maximum component error for the Time–Galerkin approximation). From Fig. 2, $\varepsilon(t) < 0.5\%$ for all $t \geqslant 0$. In Fig. 3 we compare the spatial profiles $u(x, t)$ and $u(4; x, t)$ for $t = 0, 0.1, 0.2, 0.3$ and $t = \infty$; again the agreement between approximation and solution is quite good.

Computation Times

It is difficult to make meaningful comparisons between computation times required by standard numerical methods for solving (26)–(28) and the

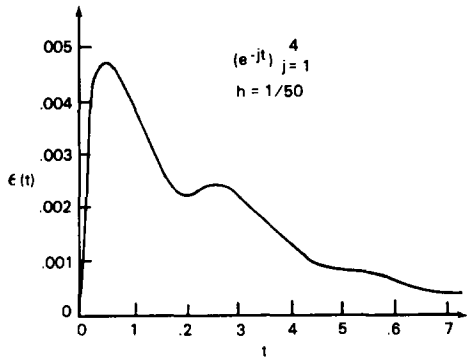

FIG. 2. Example 2. Relative error.

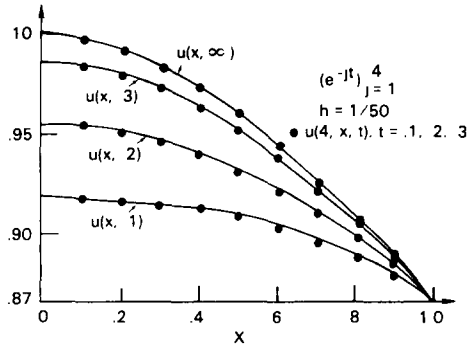

FIG. 3. Example 2. Exact aand approximate spatial profiles.

corresponding times required using the two stage Galerkin, infinite element approach; so much depends upon the numerical methods being compared and the particular problem at hand. For the simple problems presented in this paper we have made comparisons between results obtained using the infinite element method and computational results obtained using the computer code PDEPACK [23], a carefully engineered FORTRAN program designed specifically for the numerical solution of transient, nonlinear two-point boundary value problems of the form (26)–(28). This program first discretizes the spatial operator in (26) by second order centered finite differences, and then calls upon the efficient GEAR codes [21] to integrate the resulting system of nonlinear initial value problems.

For the linear problems we have tested there is no question about the superiority of the infinite element approach. For example, while it required 0.064 sec of IBM 370/168 computer time to generate the infinite element approximation to Example 1 based on $\gamma_j(t) = e^{-j^2\pi^2 t}$ and $M = 3$ in Table III, it required 0.715 sec of CPU time for PDEPACK to generate a numerical approximation of similar accuracy—about 0.4% relative error. The improvement comes, of course, from the ability to easily identify optimal

expansion functions $\gamma_j(t)$ in (39) for such simple linear problems. Substantial improvement in computational efficiency using infinite elements was also realized for the nonlinear problem treated in Example 2. Using PDEPACK with the same uniform spatial mesh of 50 points (the spatial truncation for both PDEPACK and the infinite element is the same; $O(h^2)$), 1.237 sec of computer time was needed to provide an approximation with a relative error of about 0.5%. This is to be contrasted with 0.218 sec of computer time required by the infinite element approach. For details on other examples and implementation, the reader is referred to [4].

APPENDIX: THE NONSINGULARITY OF THE TIME–GALERKIN EQUATIONS

In this Appendix we establish conditions under which the Time–Galerkin equations (3) associated with the *linear* system

$$\frac{dy}{dt} = -Py + s(t), \qquad t > 0, \tag{47}$$
$$y(0) = y_0,$$

possess a unique solution. In (47), P is assumed to be an $N \times N$ constant matrix and $s(t)$ is a known N-vector. Since in this case $F(t, y) = Py - s(t)$, the Time–Galerkin equations (3) may be written as the matrix equation

$$CH_2 + PCH_1 = S. \tag{48}$$

Here the coefficient matrix C is $N \times (M-1)$ with the unknown vectors $c_2, c_3, \ldots, c_M$ as its columns (see (2)), and H_1 and H_2 are square matrices of order $M-1$, their (i, j)th entries being respectively

$$\int_0^b \theta_{i+1}\theta_{j+1}\,dt \quad \text{and} \quad \int_0^b \frac{d\theta_{i+1}}{dt}\theta_{j+1}\,dt, \qquad i, j = 1, 2, \ldots, M-1.$$

Finally, S is an $N \times (M-1)$ matrix whose jth column is given by

$$\int_0^b \left[s - c_1 \frac{d\theta_1}{dt} - P(y(b) + c_1\theta_1) \right] \theta_{j+1}\,dt, \qquad 1 \leqslant j \leqslant M-1.$$

Since the expansion functions $\theta_2, \theta_3, \ldots, \theta_M$ are assumed to be linearly independent, we see that H_1 is a symmetric, positive definite matrix. Furthermore, it is easy to verify that H_2 is skew symmetric. In view of this we can write (48) as

$$PC + CH_2H_1^{-1} = SH_1^{-1}. \tag{49}$$

But, it is well known [17, p. 225] that (49) has a unique solution C if and only if the spectrums of P and $H_2H_1^{-1}$ are disjoint.

This necessary and sufficient condition does not appear to be easy to verify. However, as we now show, $H_2H_1^{-1}$ has only pure imaginary eigenvalues. In the first place, the eigenvalues of $H_2H_1^{-1}$ coincide with those of $H_1^{-1}H_2$. Thus, from the eigenvector equation $H_1^{-1}H_2z = \lambda z$ we obtain $z^*H_2z = \lambda z^*H_1z$, where z^* denotes the conjugate transpose of z. However, since H_1 is Hermitian and H_2 skew Hermitian, we also have that $-z^*H_2z = \bar{\lambda}z^*H_1z$. Therefore, $(\lambda + \bar{\lambda})(z^*H_1z) = 0$; that is, λ is pure imaginary. Combining these results, we obtain the simple sufficient condition of the following Corollary:

COROLLARY 1. *If P has no eigenvalues on the imaginary axis, then the Time–Galerkin equations* (47) *have a unique solution matrix C.*

REFERENCES

1. G. A. BAKER, J. H. BRAMBLE, AND V. THOMEE, Single step Galerkin approximations for parabolic problems, *Math. Comput.* **31** (1977), 818–847.
2. G. BIRKHOFF, M. SCHULTZ, AND R. S. VARGA, Piecewise Hermite interpolation in one and two variables with applications to partial differential equations, *Numer. Math.* **11** (1968), 232–256.
3. J. C. CAVENDISH, C. A. HALL, AND O. C. ZIENKIEWIZ, Blended infinite elements for parabolic boundary value problems, *Internat. J. Numer. Meth. Engrg.* **12** (1978), 1841–1851.
4. J. C. CAVENDISH AND C. A. HALL, "Blended Infinite Elements for Parabolic Boundary Value Problems," General Morots Research Publication GMR-121, 1978.
5. J. C. CAVENDISH AND C. A. HALL, L_∞-Convergence of collocation and galerkin approximations to linear two-point parabolic problems, *Aequationes Math.* **11** (1974), 230–249.
6. W. J. CODY, G. MEINARDUS, AND R. S. VARGA, Chebyshev rational approximation to e^{-x} on $[0, \infty)$ and applications to heat-conduction problems, *J. Approx. Theory* **2** (1969), 50–65.
7. W. A. COPPEL, "Stability and Asymptotic Behavior of Differential Equations," Heath, Boston, 1965.
8. J. A. W. DA NOBREGA, A new solution of the point kinetics equations, *Nucl. Sci. Engrg.* **46** (1971), 366–375.
9. P. J. DAVIS, "Interpolation and Approximation," Blaisdell, London, 1963.
10. J. DEVOOGHT AND E. MUND, *A*-Stable algorithms for neutron kinetics, *in* "Proceedings, NEARCRP/CSNI Meeting on New Developments in Three Dimensional Neutron Kinetics and Review of Kinetics Benchmark Calculations, Garehing–Munich, 1975."
11. J. DOUGLAS, JR., AND T. DUPONT, Galerkin methods for parabolic equations with nonlinear boundary conditions, *Numer. Math.* **20** (1973), 213–237.
12. J. DOUGLAS, JR., AND T. DUPONT, Galerkin methods for parabolic equations, *SIAM J. Numer. Anal.* **7** (1970), 575–626.
13. J. DOUGLAS, JR., AND T. DUPONT, Finite element collocation methods, *Math. Comput.* **27** (1973), 17–25.

14. R. J. DUFFIN AND P. WHIDDEN, An exponomial extrapolator, *J. Math. Anal. Appl.* 3 (1961), 526–536.
15. G. FORSYTHE AND W. WASOW, "Finite-Difference Methods for Partial Differential Equations," Wiley, New York, 1960.
16. H. O. FATTORNINI AND D. L. RUSSELL, Exact controllability theorems for linear parabolic equations in one space dimension, *Arch. Rational Mech. Anal.* **43** (1971), 272–292.
17. F. R. GANTMAKHER, "The Theory of Matrices," Vol. I, Moscow, 1954.
18. P. J. HARLEY AND A. R. MITCHELL, A finite element collocation method for the exact control of a parabolic problem, *Internat. J. Numer. Meth. Engrg.* **11** (1977), 345–353.
19. L. L. HEGEDUS AND J. C. CAVENDISH, Intrapellet diffusivities from integral reactor models and experiments, *I&EC Fund.* **16** (1977), 356–361.
20. A. C. HINDMARSH, "GEARIB: Solution of Implicit Systems of Ordinary Differential Equations with Banded Jacobian," Report UCID-30130, Lawrence Livermore Lab., Livermore, Calif., February 1976.
21. A. C. HINDMARSH, "GEAR: Ordinary Differential Equation Solver;" Report UCID-3001, Lawrence Livermore Lab., Livermore, Calif. December 1974.
22. E. ISAACSON AND H. B. KELLER, "Analysis of Numerical Methods," Wiley, New York, 1966.
23. N. K. MADSEN AND R. F. SINCOVEC, "PDEPACK: Partial Differential Equations Package," Scientific Computing Consulting Services, 531 Zircon Way, Livermore, California.
24. N. K. MADSEN AND R. F. SINCOVEC, The numerical method of lines for the solution of nonlinear partial differential equations, *in* "Computational Methods in Nonlinear Mechanics" (J. T. Oden *et al.*, Eds.), Texas Institute for Computational Mechanics, Austin, 1974.
25. T. A. PORSCHING, On the spectrum of a matrix Arising from a problem in reactor kinetics, *SIAM J. Appl. Math.* **16** (1968), 301–317.
26. H. S. PRICE AND R. S. VARGA, Error bounds for semidiscrete Galerkin approximations of parabolic problems with applications to pretroleum reservoir mechanics, *in* "Numerical Solution of Field Problems in Continuum Physics" (G. Birkhoff and R. S. Varga, Eds.), pp. 79–94, SIAM–AMS Proceedings II, 1970.
27. J. M. ORTEGA AND W. C. RHEINBOLDT, "Iterative Solution of Nonlinear Equations in Several Variables," Academic Press, New York, 1970.
28. M. H. STONE, A generalized Weierstrass approximation theorem, *in* "*Studies in Modern Analysis*" (R. C. Buck, Ed.), Vol. 1, MAA/Prentice–Hall, Englewood Cliffs, N. J., 1962.
29. R. S. VARGA, "Matrix Iterative Analysis," Prentice–Hall, Englewood Cliffs, N. J., 1962.

Printed in Belgium