ICASE

DIFFERENCE EQUATION STATE APPROXIMATIONS FOR NONLINEAR HEREDITARY CONTROL PROBLEMS

(NASA-CR-172785) DIFFERENCE EQUATION STATE APPROXIMATIONS FOR NONLINEAR HEREDITARY CONTROL PROBLEMS (NASA) 52 p HC A04/MF A01 CSCL 12A
N83-33589
Unclas
G3/64 10975

I. G. Rosen

Report No. 82-16
June 30, 1982

INSTITUTE FOR COMPUTER APPLICATIONS IN SCIENCE AND ENGINEERING
NASA Langley Research Center, Hampton, Virginia 23665

Operated by the

UNIVERSITIES SPACE RESEARCH ASSOCIATION

DIFFERENCE EQUATION STATE APPROXIMATIONS FOR NONLINEAR HEREDITARY CONTROL PROBLEMS

I. G. Rosen
Department of Mathematics, Bowdoin College
Brunswick, MAINE 04011

ABSTRACT

Discrete approximation schemes for the solution of nonlinear hereditary control problems are constructed. The methods involve approximation by a sequence of optimal control problems in which the original infinite dimensional state equation has been approximated by a finite dimensional discrete difference equation. Convergence of the state approximations is argued using linear semigroup theory and is then used to demonstrate that solutions to the approximating optimal control problems in some sense approximate solutions to the original control problem. Two schemes, one based upon piecewise constant approximation, and the other involving spline functions are discussed. Numerical results are presented, analyzed and used to compare the schemes to other available approximation methods for the solution of hereditary control problems.

Work supported in part by the Air Force Office of Scientific Research under Contract AFOSR 76-3092D, in part by the National Science Foundation under Grant NSF-MCS-7905774-02; in part by the U. S Army Research Office under Contract ARO-DAAG29-79-C-0161; and in part by the Bowdoin College Faculty Research Fund. Additional support was provided by the National Aeronautics and Space Administration under NASA Contract Nos. NAS1-15810 and NAS1-16394 while the author was in residence at the Institute for Computer Applications in Science and Engineering, NASA Langely Research Center, Hampton, VA 23665.

1. Introduction

The purpose of this paper is two-fold. It first serves to describe how the abstract approximation framework developed for the integration of linear functional differential equation (FDE) initial value problems in [24] can be extended so as to be applicable to certain nonlinear problems as well. Secondly, the application of the resulting approximation schemes to the generation of approximate solutions to optimal control problems in which the dynamics of the underlying system are governed by nonlinear FDE is discussed.

The approach we take is not new. We consider the nonlinear FDE in an equivalent form, i.e., as an implicit abstract evolution equation in an infinite dimensional Hilbert space Z. We then construct a sequence of finite dimensional approximating discrete difference equations by approximating the solution semigroup of operators (and its infinitesimal generator) defined by the linear part of the equation using piecewise constant or spline based subspaces of Z. Linear Semigroup Theory and discrete analogs of the Trotter-Kato Theorem [18] and the well-known Gronwall inequality are then used to argue convergence. Approximate solutions to the optimal control problem are generated by considering a sequence of approximating optimal control problems in each of which the infinite dimensional FDE state equation has been approximated in the spirit of the discussions above. Using the fact that the state approximations converge, we are then able to demonstrate thac solutions to the approximating optimal control problems (which can be solved by conventional methods) in some sense approximate solutions to the original control problem.

Banks and Burns [3][4] were among the first to propose the idea of approximating hereditary control problems by a sequence of finite dimensional approximating control problems. The semi-discrete methods for problems with linear state equations which they developed were later extended by Banks [2] so as to be applicable to nonlinear problems as well. Using similar approaches, Reber [21] and Rockey [23] developed fully discrete schemes for the approximation of FDE which they then applied to the solution of control problems. Reber developed first order convergent schemes for linear non-autonomous equations. In the constant coefficient case, his work becomes a special case of the more general theory to be presented below. In [23], the linear or nonlinear FDE is first recast as an equivalent Volterra integral equation in L_2, and is then discretized using piecewise constant or spline subspaces. An algebraic system for the Fourier coefficients of the solution results which is then solved using standard methods. Recently, within the context of the framework developed in [3], Gibson [13] and Kunisch [20] have formulated semi-discrete approximation schemes which yield approximating closed loop solutions to the linear quadratic control problem with hereditary system dynamics.

The discussion of our results below closely parallels the presentation in [2]. The treatment in [2] relies heavily upon the linear theory developed in [3] and [4] by considering the nonlinearities (which are assumed to satisfy local Lipschitz and affine growth conditions) to be a perturbation of the linear part of the equation. Since the basis for our approximation schemes involves the approximation of the solution semigroup e^{At} using rational function approximations to the exponential and finite dimensional approximation of A, we too depend heavily upon the linear theory, and hence consider precisely the same class of equations which are studied in [2]. An unfortunate consequence, however, is that this precludes the inclusion of nonlinearities in the discrete delay terms (i.e. terms of the form x(t-r)). This is in contrast to the work of Kappel and Schappacher [17] and Kappel [16] and the recent paper by Daniel [10] in which the nonlinearities in the equation are handled more directly and which do permit discrete delay terms to enter into the equation in a nonlinear fashion. The convergence arguments for the approximation schemes developed in [16] and [17] are based largely on ideas from nonlinear semigroup theory and approximation results analogous to those used in the linear case. Daniel, on the other hand, avoids the semigroup approach entirely and relies instead, directly upon the dissipative properties of the nonlinear operators arising in the abstract formulation of the FDE in order to argue the convergence of spline based semidiscrete approximation schemes. These results are obtained, however, at the expense of requiring somewhat stronger assumptions (global Lipschitz and additional smoothness) on the nonlinearities in the equation and the placement of additional restrictions on the class of admissible controls.

We conclude this section with an outline of the rest of the paper and a brief description of our notation. In Section 2 we define the nonlinear FDE with which we shall be concerned and state the hypotheses it must satisfy in order for us to carry out our analysis. We also state fundamental existence and uniqueness results and describe the equivalent formulation of the FDE as an abstract evolution equation in the Hilbert space Z. In Section 3 we first recall the abstract approximation results for linear equations discussed in [24]. We then extend them so that they are applicable to the nonlinear equation as well and state and prove the fundamental convergence result. In Section 4 we briefly describe the details involved in the construction of actual schemes to which our general convergence results apply. We also outline two specific schemes, one using piecewise constant functions and the other using splines. Section 5 contains the results pertaining to the application of the approximation schemes to the solution of optimal control problems while in Section 6 we demonstrate the feasibility of our methods by presenting and analyzing several numerical examples.

The notation we use, is, for the most part, standard. The superscripts on the Lebesgue spaces $L_p^n(a,b)$, the space of functions with p continuous derivatives $C_p^n(a,b)$ and the Sobolev spaces $H_p^n(a,b)$ denote that they consist of functions (or equivalence classes of functions) defined on (a,b) with range in R^n. The symbol L_∞^{loc} is used to denote the class of functions which are locally essentially bounded. The space of continuous functions from an interval (a,b) with range in the abstract space Z is denoted by $C([a,b],Z)$. We assume that this space is endowed with the usual supremum norm. For a linear operator A and a complex number λ contained in the resolvent set of A we denote the resolvent of A at λ by $R(\lambda;A)$.

2. Nonlinear Hereditary Control Systems and Their Abstract Formulation

In this paper we consider nonlinear hereditary control systems which are governed by functional differential state equations of retarded type of the form

$$\dot{x}(t) = Lx_t + f(t,x(t),x_t,u(t)) \qquad t \in [0,T] \tag{2.1}$$

with initial conditions given by

$$x(0) = \eta \qquad x_0 = \varphi \tag{2.2}$$

where $\eta \in R^n$, $\phi \in L_2^n(-r,0)$ and x_t denotes the function on $[-r,0]$ defined by $x_t(\theta) = x(t+\theta)$, $-r \le \theta \le 0$. The linear part of the equation, given by the linear operator $L:L_2^n(-r,0) \to R^n$ will be assumed to be of the form

$$L\phi = \sum_{j=0}^{\nu} A_j \phi(-\tau_j) + \int_{-r}^{0} A(\theta)\phi(\theta)d\theta$$

where the A_j are $n\times n$ matrices, $A(\cdot)$ is a square integrable $n\times n$ matrix valued function defined on the interval $(-r,0)$ and $0 = \tau_0 < \tau_1 < \tau_2 \ldots < \tau_\nu = r$. Strictly speaking, $L\phi$ is not well defined for all $\phi \in L_2^n(-r,0)$ in that point evaluations of ϕ are required. We remedy this situation by insuring that in any instance in which the operator L appears below, either it is being applied to an element in $L_2^n(-r,0)$ for which the value of $L\phi$ is well defined, or $L\phi$ appears either explicitly or implicitly beneath an integral sign in a reference to x, the solution to the initial value problem (2.1)(2.2) above.

In addition, we assume that the nonlinear perturbation term $f:R^1\times R^n\times L_2^n(-r,0)\times R^m \to R^n$ satisfies the following hypotheses

(H1) The mapping $(t,\eta,\phi,v) \to f(t,\eta,\phi,v)$ is continuous on $R^1\times R^n\times L_2^n(-r,0)\times R^m$

(H2) For any bounded subset $\mathcal{D}$ of $R^n\times L_2^n(-r,0)$ there exist $m_i = m_i(\mathcal{D})$, $m_i \in L_\infty^{loc}$ $i = 1,2$ such that for $v \in R^m$, $t \in R^1$ and $(\eta,\phi),(\xi,\psi) \in \mathcal{D}$ one has

$$|f(t,\eta,\phi,v) - f(t,\xi,\psi,v)| \le \{m_1(t)+m_2(t)|v|\}\{|\eta-\xi|+|\phi-\psi|\}$$

(H3) There exists a continuous $n\times m$ matrix valued mapping $t \to B(t)$ such that $f(t,0,0,v) = B(t)v$ for all $t \in R^1$ and $v \in R^m$. In addition there exist functions $\hat{m}_i \in L_\infty^{loc}$ $i = 1,2$ such that

$$|f(t,\eta,\phi,v)| \le \{\hat{m}_1(t) + \hat{m}_2(t)|v|\}\{|\eta| + |\phi|\} + |B(t)||v|$$

for all $t \in R^1$, $v \in R^m$ and all $(\eta,\phi) \in R^n \times L_2^n(-r,0)$ with $|(\eta,\phi)|^2 = |\eta|^2 + |\phi|^2$ sufficiently large.

(H4) There exists a continuous function $g: R^1 \times R^n \times L_2^n(-r,0) \to R^1$ such that

$$|f(t,\eta,\phi,v) - f(t,\eta,\phi,w)| \le g(t,\eta,\phi)|v-w|$$

for all $t \in R^1$, $\eta \in R^n$, $\phi \in L_2^n(-r,0)$ and $v,w \in R^m$.

Hypotheses (H2) and (H3) together yield the following growth condition satisfied by f:

(G) There exist functions $\tilde{m}_1, \tilde{m}_2 \in L_\infty^{loc}$ such that

$$|f(t,\eta,\phi,v)| \in \{\tilde{m}_1(t) + \tilde{m}_2(t)|v|\}\ \{|\eta| + |\phi|\} + |B(t)||v|$$

for all $t \in R^1$, $\eta \in R^n$, $\phi \in L_2^n(-r,0)$ and $v \in R^m$.

A solution $x(t) = x(t;\eta,\phi,u)$ to (2.1)(2.2) is defined to be a function $x \in L_2^n(-r,T)$ such that the mapping $t \to x(t)$ is absolutely continuous on $(0,T)$, (2.1) is satisfied a. e. on $(0,T)$ and for which $x(0) = \eta$ $x_0 = \phi$. Using standard arguments, the following existence, uniqueness and continuous dependence result for solutions to the initial value problem (2.1) (2.2) can be established.

Theorem 2.1 Under hypotheses (H1) - (H4), given $u \in L_2^m(0,T)$ and $(\eta,\phi) \in R^n \times H_1^n(-r,0)$ with $\eta = \phi(0)$, there exists a unique solution to the initial value problem (2.1)(2.2) on $[0,T]$. Moreover, the mapping $(\phi,u) \to (x(t;\phi(0),\phi,u), x_t(\phi(0),\phi,u))$ from $H_1^n(-r,0) \times L_2^m(0,T)$ into $R^n \times L_2^n(-r,0)$ where x is the unique solution to (2.1)(2.2) corresponding to $u \in L_2^m(0,T)$

and $x(0) = \phi(0)$, $x_0 = \phi$, is continuous with respect to the topology on $H_1^n(-r,0) \times L_2^m(0,T)$ induced by the supremum norm on $H_1^n(-r,0)$ and the standard L_2 norm on $L_2^m(0,T)$.

Fundamental to the development of our approximation schemes below will be the equivalence which exists between the FDE initial value problem (2.1)(2.2) above and an abstract evolution equation set in the Hilbert space $Z = R^n \times L_2^n(-r,0)$ with inner product $\langle \cdot,\cdot \rangle_Z = \langle \cdot,\cdot \rangle_{R^n} + \langle \cdot,\cdot \rangle_{L_2^n(-r,0)}$.

For each $t \geq 0$ let $S(t):Z \to Z$ denote the solution operator for the associated linear homogeneous initial value problem corresponding to (2.1)(2.2). That is, for $(\eta,\phi) \in Z$ we have

$$S(t)(\eta,\phi) = (x(t),x_t)$$

where x is the unique solution to (2.1)(2.2) with $f \equiv 0$. Based upon existence, uniqueness and continuous dependence results for the linear homogeneous problem (see [3][4][24]) one may conclude that $\{S(t):t \geq 0\}$ represents a parameterized family of well defined bounded linear transformations forming a C_0-semigroup of operators on Z. The infinitesimal generator of $\{S(t):t \geq 0\}$, A, and its domain of definition $D(A)$ may be calculated and are given by

$$D(A) = \{(\eta,\phi) \in Z:\eta=\phi(0),\phi \in H_1^n(-r,0)\}$$

$$A(\phi(0),\phi) = (L\phi,\dot{\phi}).$$

If we define the inner product $\langle \cdot,\cdot \rangle_g$ on Z by

$$\langle (\eta,\phi),(\xi,\psi) \rangle_g = \eta^T\xi + \int_{-r}^{0} \phi(\theta)^T\psi(\theta)g(\theta)d\theta$$

where

$$g(\theta) = \begin{cases} 1 & -r \le \theta < \tau_{\nu-1} \\ 2 & -\tau_{\nu-1} \le \theta < \tau_{\nu-2} \\ \vdots & \\ \nu & \tau_1 \le \theta \le 0 \end{cases}$$

then it clearly follows that for $(\eta,\phi) \in Z$

$$|(\eta,\phi)|_Z \le |(\eta,\phi)|_g \le \sqrt{\nu}\,|(\eta,\phi)|_Z \, .$$

Furthermore, it can be shown that the operator A satisfies the following dissipative inequality with respect to the g inner product:

$$\text{(2.3)} \qquad \langle Az_0, z_0\rangle_g \le \omega \langle z_0, z_0\rangle_g$$

with $\omega = \frac{\nu+1}{2} + |A_0| + 1/2 \sum_{i=1}^{\nu} |A_i|^2 + 1/2 \int_{-r}^{0} |A(\theta)|^2 d\theta$

and hence $A \in G(\sqrt{\nu},\omega)$ -- that is, the semigroup of operators $\{S(t): t \ge 0\}$ satisfies the exponential bound given by

$$|S(t)| \le \sqrt{\nu}\, e^{\omega t}.$$

Let $\pi_1 : Z \to R^n$ and $\pi_2 : Z \to L_2^n(-r,0)$ denote the two coordinate projections of Z onto R^n and $L_2^n(-r,0)$ respectively. That is for $(\eta,\phi) \in Z$, we have

$$\pi_1(\eta,\phi) = \eta \quad \pi_2(\eta,\phi) = \phi.$$

Let the mapping $F: R^1 \times Z \times R^m \to Z$ be defined by

$$F(t,z,v) = (f(t,\pi_1 z,\pi_2 z,v),0).$$

Hypotheses (H1)-(H4) imposed upon f naturally imply that the mapping F defined above will have the following properties:

(P1) For any $z \in C([0,T],Z)$ and $u \in L_2^m(0,T)$, the mapping $t \to |F(t,z(t),u(t))|$ is in $L_2^1(0,T)$.

(P2) For any bounded subset $\mathcal{D}$ of Z, there exist M_1, M_2 (depending on $\mathcal{D}$) in L_∞^{loc} such that $|F(t,z,v) - F(t,w,v)| \le \{M_1(t) + M_2(t)|v|\}\ |z-w|$ for all $z,w \in \mathcal{D}$, $t \in R^1$ and $v \in R^m$.

For $z_0 \in Z$ and $u \in L_2^m(0,T)$, let the mapping $z:[0,T] \to Z$ be defined implicitly by the following expression

$$(2.4) \qquad z(t) = z(t;z_0,u) = S(t)z_0 + \int_0^t S(t-\sigma)F(\sigma,z(\sigma),u(\sigma))d\sigma \ .$$

Using hypotheses (H1)-(H4) and properties (P1) and (P2) above together with standard arguments involving Picard iterates and the Gronwall inequality, Banks [2] is able to establish the following lemma.

<u>Lemma</u> 2.1 Under hypotheses (H1)-(H4), equation (2.4) above, defines for each $z_0 \in Z$ and $u \in L_2^m(0,T)$ a unique function $t \to z(t;z_0,u) \in C([0,T],Z)$. Moreover, the mapping $(\phi(0),\phi,u) \to z(t,(\phi(0),\phi),u)$ is continuous on $D(A)\times L_2^m(0,T)$ with respect to the $Z\times L_2$ and $R^n\times C^n\times L_2$ topologies.

Finally, using the above results, the equivalence which we desire between the FDE initial value problem (2.1)(2.2) and an abstract evolution equation set in Z, in particular the system given by (2.4), can be established.

Theorem 2.2 For f satisfying hypotheses (H1)-(H4), $z_0 = (\phi(0),\phi) \in D(A)$ and $u \in L_2^m(0,T)$ we have

$$z(t;z_0,u) = (x(t;\phi(0),\phi,u),x_t(\phi(0),\phi,u)) \qquad t \in [0,T] \tag{2.5}$$

where $z(t;z_0,u)$ is the unique solution to (2.4) guaranteed to exist by Lemma 2.1 and $x(t;\phi(0),\phi,u)$ is the unique solution to the FDE initial value problem (2.1)(2.2) guaranteed to exist by Theorem 2.1.

We shall only briefly outline the essential aspects of the arguments necessary to verify Theorem 2.2. The details of the proof can be found in [2]. The equivalence described by (2.5) is first established on a restricted class of initial data z_0 and input functions u. For $(\phi(0),\phi) \in Z$ with $\phi \in C_1^n(-r,0)$ and $u \in C^m(0,T)$ it can be shown that

$$w(t) = (x(t;\phi(0),\phi,u),x_t(\phi(0),\phi,u))$$

is the unique strong solution to the abstract evolution equation in Z given in differentialed form by

$$\dot{y}(t) = Ay(t) + F(t,y(t),u(t)) \tag{2.6}$$

$$y(0) = (\phi(0),\phi) \tag{2.7}$$

However, it can also be shown that any strong solution of (2.6) (2.7) corresponding to $\phi \in C_1^n(-r,0)$ and $u \in C^m(0,T)$ must satisfy

$$y(t) = S(t)y(0) + \int_0^t S(t-\sigma)F(\sigma,y(\sigma),u(\sigma))d\sigma$$

and hence by Lemma 2.1 must be the unique solution in $C([0,T],Z)$ of (2.4).

Therefore, we have that

$$z(t,(\phi(0),\phi),u) = w(t) = (x(t,\phi(0),\phi,u),x_t(\phi(0),\phi,u))$$

With the desired equivalence now established for $\phi \in C_1^n(-r,0)$ and $u \in C^m(0,T)$, it is easily extended to the more general class of initial data and input functions described in the statement of the theorem through the use of the continuous dependence results given in Theorem 2.1 and Lemma 2.1 together with standard density arguments.

3. An Abstract Approximation Framework

In this section we develop an abstract approximation framework under which approximation schemes applicable to the abstract evolution equation given by (2.4) can be constructed. In addition, we establish conditions which are sufficient to conclude convergence of schemes constructed within the framework. The approach we take is based upon, and an extension of, the discrete approximation framework for the integration of linear FDE initial value problems described in [24]. Indeed our schemes will be based upon the approximation of the semigroup of operators $\{S(t):t \geq 0\}$ defined on Z by a sequence of discrete semigroups (see [18]) which are defined on finite dimensional approximating subspaces of Z and which are constructed using rational function approximations to the exponential and finite dimensional approximations to the infinitesimal generator A of $\{S(t):t \geq 0\}$. The fundamental convergence results for these constructions are given in Theorem 3.1 to follow and are used extensively throughout our discussions below.

For each $N = 1,2\ldots$ let Z_N be a finite dimensional subspace of Z of dimension k_N and let $P_N: Z \to Z_N$ be the associated orthogonal (not necessarily with respect to the standard inner product on Z) projection of Z onto Z_N. Define $A_N: Z_N \to Z_N$ to be a bounded linear operator on Z_N and let $S_N(t) = e^{A_N t}$ for all $t \geq 0$.

<u>Theorem</u> 3.1 Suppose

(1) $P_N z \to z$ as $N \to \infty$ for each $z \in Z$

(2) There exist constants M, β, independent of N for which $A, A_N \in G(M,\beta)$, $N = 1,2\ldots$ (i.e. $|S(t)| \leq Me^{\beta t}, |S_N(t)| \leq Me^{\beta t} \quad N = 1,2\ldots$)

(3) There exists $D_1 \subset D(A)$, a dense subset of Z for which $|A_N P_N z - Az| \to 0$ as $N \to \infty$ for each $z \in D_1$.

(4) There exist $\lambda \in C$ with $\text{Re}\lambda > \beta$ and D_2 a dense subset of Z for which $R(\lambda;A)D_2 \subseteq D_1$

(5) $C(z)$ is a rational function of the complex variable z for which

(a) $|C(z) - e^z| = O(|z|^{q+1})$ as $|z| \to 0$ with $q > 0$

(b) if $C(z) = n(z)/d(z)$ then

degree $C(z) \equiv$ degree $n(z)$ - degree $d(z) \leq q + 1$

(c) $C(z)$ has no poles in $\{z \in C: \text{Re}\, z \leq 0\}$

Then the operators $C(\frac{r}{N}A_N) = n(\frac{r}{N}A_N)d(\frac{r}{N}A_N)^{-1}$ exist for all N sufficiently large. If, in addition, for ρ_N, that positive integer for which $\rho_N \frac{r}{N} \leq T < (\rho_N + 1)\frac{r}{N}$, we have that the infinite collection of operators on Z_N, $\left\{C(\frac{r}{N}A_N)^k\right\}_{k=0}^{\rho_N}$ are uniformly bounded with respect to N then

(3.1) $\quad |C(\frac{r}{N} A_N)^k P_N z - S(t_k^N)z| \to 0$

as $N \to \infty$ for each $z \in Z$ uniformly in k, $k = 0,1,2\ldots\rho_N$ where $t_k^N = k\frac{r}{N} \quad k = 0,1,2\ldots\rho_N$.

Theorem 3.1 is based primarily upon a result due to Hersh and Kato [14] and is in fact a fully discrete analog of the well known Trotter-Kato results which are commonly used in establishing the convergence of semi-discrete approximations to semigroups of operators (see [18]). That it is possible to actually construct schemes (i.e. Z_N, P_N, A_N, $C(z)$) which satisfy the hypotheses of Theorem 3.1 is exhibited in the next section.

Remark 3.1 As a corollary to Theorem 3.1, it is possible to estimate the rate of convergence in (3.1). Indeed if for $z \in S$, a particular subset of Z which is defined in Theorem 4.17 of [24] we have that the convergence in Hypothesis (3) is $O(\frac{r}{N})^p$ for some $p > 0$, then the rate of convergence in (3.1) will be $O(\frac{r}{N})^p + O(\frac{r}{N})^q$ for $z \in S$.

Before we can proceed to apply the results of Theorem 3.1 in the development of approximation schemes for the nonlinear system (2.4), we must first consider the linear nonhomogeneous problem. We shall require the following result from [24]. For $f \in L_2^n(0,T)$ and $z_0 \in Z$, let $z \in C([0,T],Z)$ be given by

$$z(t) = S(t)z_0 + \int_0^t S(t-\sigma)(f(\sigma),0)d\sigma$$

and let $\left\{ z_k^N \right\}_{k=0}^{\rho^N} \subset Z_N$ be given by

$$z_k^N = C(\frac{r}{N} A_N)^k P_N z_0 + \frac{r}{N} \sum_{j=1}^{k} C(\frac{r}{N} A_N)^{k-j} D(\lambda \frac{r}{N} A_N) P_N (f_j^n , 0)$$

where $f_j^N = \frac{N}{r} \int_{(j-1)\frac{r}{N}}^{j\frac{r}{N}} f(\sigma)d\sigma$, $D(z)$ is a rational function of the complex variable z and $0 \le \lambda \le 1$.

Theorem 3.2 Suppose that Z_N, P_N, A_N, $C(z)$ satisfy the hypotheses of Theorem 3.1. Suppose further that

(1) The infinite collection of operators on Z_N, $\left\{ C(\frac{r}{N} A_N)^k \right\}_{k=0}^{\rho_N}$ are uniformly bounded with respect to N

and

(2) The operators $D(\lambda \frac{r}{N} A_N)$ exist for all N sufficently large and satisfy $|D(\lambda \frac{r}{N} A_N)P_N z - z| \to 0$ as $N \to \infty$ for each $z \in Z$.

Then

$$|z_k^N - z(t_k^N)| \to 0$$

as $N \to \infty$ for each $z_0 \in Z$ uniformly in k, $k = 0,1,2\ldots\rho_N$ and uniformly in f for f in bounded subsets of $L_2^n(0,T)$.

Several of our arguments below rely upon an application of the following lemma. The result given in Lemma 3.1 is a discrete analog of the well known generalized Gronwall differential inequality. Since we have been unable to locate a suitable reference in the literature, a proof of the result has been included.

Lemma 3.1 Suppose that $\{\alpha_j\}_{j=0}^{\infty}$ and $\{\beta_j\}_{j=0}^{\infty}$ are sequences of non-negative real numbers and that $\{\phi_j\}_{j=0}^{\infty}$ is a sequence of real numbers which satisfy

$$\phi_n \le \alpha_n + \sum_{k=0}^{n-1} \beta_k \phi_k \qquad n = 1,2\ldots$$

Then we have that

$$\phi_n \le \alpha_n + \sum_{j=0}^{n-1} \beta_j \alpha_j (\exp(\sum_{k=j+1}^{n-1} \beta_k)) \qquad n = 1,2\ldots$$

If in addition $\alpha_j = \alpha \ge 0 \quad j = 0,1,2 \ldots$ then

$$\phi_n \le \alpha \exp(\sum_{k=0}^{n-1} \beta_k) \qquad n = 1,2, \ldots$$

Proof: Let $s_n = \sum_{k=0}^{n-1} \beta_k \phi_k \qquad n = 1,2 \ldots$ Then

$$s_n - s_{n-1} = \beta_{n-1}\phi_{n-1} \le \beta_{n-1}(\alpha_{n-1} + \sum_{k=0}^{n-2} \beta_k \phi_k) = \beta_{n-1}(\alpha_{n-1} + s_{n-1})$$

and therefore

$$s_n \le (1 + \beta_{n-1})s_{n-1} + \beta_{n-1}\alpha_{n-1} .$$

This implies that

$$(3.2) \qquad s_n \le \prod_{k=1}^{n-1} (1 + \beta_k) s_1 + \sum_{j=1}^{n-1} \prod_{k=j+1}^{n-1} (1 + \beta_k)\beta_j \alpha_j$$

$$= \prod_{k=1}^{n-1} (1 + \beta_k)\beta_0 \alpha_0 + \sum_{j=1}^{n-1} \prod_{k=j+1}^{n-1} (1+\beta_k)\beta_j \alpha_j$$

$$= \sum_{j=0}^{n-1} \prod_{k=j+1}^{n-1} (1+\beta_k)\beta_j \alpha_j .$$

Now $\beta_k \ge 0$ implies that $(1 + \beta_k) \le \exp(\beta_k) \qquad k = 1,2\ldots n-1$ and hence

$$s_n \le \sum_{j=0}^{n-1} \beta_j \alpha_j (\prod_{k=j+1}^{n-1} \exp(\beta_k)) = \sum_{j=0}^{n-1} \beta_j \alpha_j (\exp(\sum_{k=j+1}^{n-1} \beta_k)) .$$

from which we conclude that

$$\phi_n \le \alpha_n + s_n \le \alpha_n + \sum_{j=0}^{n-1} \beta_j \alpha_j (\exp(\sum_{k=j+1}^{n-1} \beta_k)) .$$

In the case that $\alpha_j = \alpha, \quad j = 0,1,2\ldots,$ (3.2) implies that

$$s_n \le \alpha \sum_{j=0}^{n-1} \beta_j \prod_{k=j+1}^{n-1} (1+\beta_k) = \alpha \sum_{j=0}^{n-1} \left(\prod_{k=j}^{n-1} (1+\beta_k) - \prod_{k=j+1}^{n-1} (1+\beta_k)\right)$$

$$= \alpha\left(\left(\prod_{k=0}^{n-1} (1+\beta_k)\right) - (1+\beta_{n-1})\right)$$

$$\le \alpha\left(\left(\prod_{k=0}^{n-1} \exp(\beta_k)\right) - (1+\beta_{n-1})\right)$$

$$= \alpha\left(\exp\left(\sum_{k=0}^{n-1} \beta_k\right) - (1+\beta_{n-1})\right).$$

Therefore

$$\phi_n \le \alpha + s_n \le \alpha\left(1 + \exp\left(\sum_{k=0}^{n-1} \beta_n\right) - 1 - \beta_{n-1}\right)$$

$$= \alpha\left(\exp\left(\sum_{k=0}^{n-1} \beta_n\right) - \beta_{n-1}\right)$$

$$\le \alpha \exp\left(\sum_{k=0}^{n-1} \beta_k\right).$$

For Z_N, P_N, A_N, $C(z)$, $D(z)$ and ρ_N as described above, $z_0 \in Z$, $u \in L_2^m(0,T)$ and f satisfying hypotheses (H1) – (H4) we define the collection $\left\{z_k^N\right\}_{k=0}^{\rho_N} \subset Z_N$ by

$$(3.3) \qquad z_k^N = z_k^N(z_0,u) = C\left(\frac{r}{N} A_N\right)^k P_N z_0 + \frac{r}{N} \sum_{j=1}^{k} C\left(\frac{r}{N} A_N\right)^{k-j} D\left(\lambda \frac{r}{N} A_N\right) P_N F_j^N$$

$$k = 0,1,2\ldots\rho_n$$

where $F_i^N \equiv \frac{N}{r} \int_{(i-1)\frac{r}{N}}^{i\frac{r}{N}} F(\sigma, z_{i-1}^N, u(\sigma))d\sigma = \left(\frac{N}{r} \int_{(i-1)\frac{r}{N}}^{i\frac{r}{N}} f(\sigma, \pi_1 z_{i-1}^N, \pi_2 z_{i-1}^N, u(\sigma))d\sigma, 0\right)$

$$i = 1,2\ldots\rho_N.$$

Lemma 3.2 For Z_N, P_N, A_N, $C(z)$, $D(z)$ and ρ_N satisfying the hypotheses of Theorem 3.2 and $u \in E$, a bounded subset of $L_2^m(0,T)$, the collection $\left\{z_k^N\right\}_{k=0}^{\rho_N}$ defined by (3.3) above are bounded in $(\overset{\rho_N}{\underset{0}{\times}} Z, \|\cdot\|_\infty^N)$ uniformly in N for all N sufficiently large and uniformly in $u \in E$ where $\left\|\left\{z_k^N\right\}_{k=0}^{\rho_N}\right\|_\infty^N \equiv \max_{0 \le k \le \rho_n} |z_k^N|_Z$.

Proof Let K_0 denote the uniform bound on the operators $\left\{C(\frac{r}{N} A_N)^k\right\}_{k=0}^{\rho_N}$ for all N sufficiently large which is assumed to exist in hypothesis (1) of Theorem 3.2, and let K_1 denote the uniform bound on the operators $D(\lambda \frac{r}{N} A_N)$ for all N sufficiently large whose existence can be argued using hypothesis (2) of Thecrem 3.2 and the uniform boundedness principle. Then, for $k = 0,1,2 \ldots \rho_N$

$$|z_k^N| \le K_0|z_0| + \frac{r}{N} K_0 K_1 \sum_{j=1}^{k} |P_N F_j^N|$$

$$\le K_0|z_0| + K_0 K_1 \sum_{j=1}^{k} \int_{j-1\frac{r}{N}}^{j\frac{r}{N}} |f(\sigma, \pi_1 z_{j-1}^N, \pi_2 z_{j-1}^N, u(\sigma))| d\sigma.$$

Applying the growth condition (G) satisfied by f, we find

$$|z_k^N| \le K_0|z_0| + K_0 K_1 \sum_{j=1}^{k} \int_{(j-1)\frac{r}{N}}^{j\frac{r}{N}} (\tilde{m}_1(\sigma) + \tilde{m}_2(\sigma)|u(\sigma)|) \{ |\pi_1 z_{j-1}^N| + |\pi_2 z_{j-1}^N| \} + |B(\sigma)||u(\sigma)| d\sigma$$

$$\le K_0|z_0| + K_0 K_1 |B|_{L_2^{n\times m}(0,T)} |u|_{L_2^m(0,T)}$$

$$+ K_0 K_1 \sum_{j=1}^{\rho_N} \{|\pi_1 z_{j-1}^N| + |\pi_2 z_{j-1}^N|\} \int_{j-1\frac{r}{N}}^{j\frac{r}{N}} (\tilde{m}_1(\sigma) + \tilde{m}_2(\sigma)|u(\sigma)|) d\sigma$$

$$\leq K_0|z_0| + K_0K_1|B|_{L_2^{n\times m}(0,T)} |u|_{L_2^m(0,T)} +$$

$$\sqrt{2}\, K_0K_1 \sum_{j=1}^{\rho_N} |z_{j-1}^N| \int_{j-1\frac{r}{N}}^{j\frac{r}{N}} (\tilde{m}_1(\sigma) + \tilde{m}_2(\sigma)|u(\sigma)|)\, d\sigma,$$

and hence by Lemma 3.1

$$|z_k^N| \leq (K_0|z_0| + K_0K_1|B|_{L_2^{n\times m}(0,T)} |u|_{L_2^m(0,T)}) \cdot$$

$$\exp(\sqrt{2}\, K_0K_1 \int_0^T \tilde{m}_1(\sigma) + \tilde{m}_2(\sigma)|u(\sigma)|d\sigma)$$

$$\leq (K_0|z_0| + K_0K_1|B|_{L_2^{n\times m}(0,T)} |u|_{L_2^m(0,T)}).$$

$$\exp(\sqrt{2}\, K_0K_1(|\tilde{m}_1|_{L_\infty} T + |\tilde{m}_2|_{L_\infty} T^{1/2}|u|_{L_2^m(0,T)}).$$

<u>Theorem</u> 3.3 For Z_N, P_N, A_N, $C(z)$, $D(z)$ and ρ_N satisfying the hypotheses of Theorem 3.2, $u \in E$, a bounded subset of $L_2^m(0,T)$, $\left\{z_k^N\right\}_{k=0}^{\rho_N}$ given by (3.3), and z given by (2.4) we have

$$|z_k^N(z_0,u) - z(t_k^N;z_0,u)| \to 0$$

as $N \to \infty$ for each $z_0 \in Z$ uniformly in k, $k = 0,1,2\ldots\rho_N$ and uniformly in u for $u \in E$.

<u>Proof</u> For $z \in C([0,T],Z)$, the unique solution to (2.4) guaranteed to exist by Lemma 2.1, and $t \in [0,T]$ define the function h by

$$h(t) = h(t,u) = f(t,\pi_1 z(t),\pi_2 z(t),u(t)).$$

Then

$$z(t) = S(t)z_0 + \int_0^t S(t-\sigma)(h(\sigma),0)d\sigma$$

and using hypotheses (H1)-(H4) it is easily verified that for $u \in E$, h lies in a bounded subset of $L_2^n(0,T)$. If we define $\left\{\tilde{z}_k^N\right\}_{k=0}^{\rho_N} \subset Z_N$ by

$$\tilde{z}_k^N = C(\tfrac{r}{N} A_N)^k P_N z_0 + \frac{r}{N} \sum_{j=1}^{k} C(\tfrac{r}{N} A_N)^{k-j} D(\tfrac{r}{N} A_N) P_N (h_j^N, 0)$$

where $h_i^N = \frac{N}{r} \int_{t_{i-1}^N}^{t_i^N} h(\sigma)d\sigma$ then it follows that

$$|z_k^N - z(t_k^N)| \le |z_k^N - \tilde{z}_k^N| + |\tilde{z}_k^N - z(t_k^N)|$$

$$\le |\tilde{z}_k^N - z(t_k^N)| + \sum_{j=1}^{k} K_0 K_1 \int_{t_{j-1}^N}^{t_j^N} |f(\sigma,\pi_1 z_{j-1}^N,\pi_2 z_{j-1}^N,u(\sigma)) - f(\sigma,\pi_1 z(\sigma),\pi_2 z(\sigma),u(\sigma))|d\sigma$$

$$= |\tilde{z}_k^N - z(t_k^N)| + \sum_{j=0}^{k-1} K_0 K_1 \int_{t_j^N}^{t_{j+1}^N} |F(\sigma,z_j^N,u(\sigma)) - F(\sigma,z(\sigma),u(\sigma))|d\sigma$$

where K_0 and K_1 are as they were defined in the proof of Lemma 3.2. Since $\{z(t;z_0,u):t\in[0,T],u \in E\}$ lies in a bounded subset of Z (see [2]) as does $\{z_k^N(z_0,u),\ k = 0,1,2\ldots\rho_n\ u \in E\}$ uniformly in N for all N sufficiently large, property (P2) implies

$$(3.4) \qquad |z_k^N - z(t_k^N)| \le |\tilde{z}_k^N - z(t_k^N)| + \sum_{j=0}^{k-1} K_0 K_1 \int_{t_j^N}^{t_{j+1}^N} (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)|z_j^N - z(\sigma)|d\sigma$$

$$\leq \quad |\tilde{z}_k^N - z(t_k^N)| + K_0K_1 \sum_{j=0}^{k-1} \int_{t_j^N}^{t_{j+1}^N} (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)\cdot$$

$$|z(t_j^N) - z(\sigma)|d\sigma$$

$$+ \quad \sum_{j=0}^{k-1} K_0K_1 \int_{t_j^N}^{t_{j+1}^N}(M_1(\sigma) + M_2(\sigma)|u(\sigma)|)d\sigma|z_j^N - z(t_j^N)|.$$

Let $\varepsilon > 0$ be given. Theorem 3.2 implies that $|\tilde{z}_k^N - z(t_k^N)| < \varepsilon$ for all N sufficiently large uniformly in k, $k = 0,1,2\ldots\rho_N$ and uniformly in u for $u \in E$. Furthermore by Theorem 3.2 of [3], the operator $F{:}L_2^n(0,T) \to C([0,T],Z)$ defined by

$$F(f)(t) = S(t)z_0 + \int_0^t S(t-\sigma)(f(\sigma),0)d\sigma$$

is a compact affine operator. Since $u \in E$, a bounded subset of $L_2^m(0,T)$ implies $h(\cdot,u)$ lies in a bounded subset of $L_2^n(0,T)$, it follows that $\{z(\cdot;z_0u):u \in E\}$ is a relatively compact subset of $C([0,T],Z)$. Therefore, the mappings $t \to z(t;z_0,u)$, $u \in E$ are uniformly equicontinuous on $[0,T]$ and $|z(t_k^N) - z(\sigma)|<\varepsilon, \sigma \in \left[t_k^N, t_{k+1}^N\right]$ for all N sufficiently large uniformly in k, $k = 0,1,2\ldots\rho_N$ and uniformly in u, $u \in E$. The above arguments together with the inequalities given by (3.4) imply

$$(3.5) \qquad |z_k^N - z(t_k^N)| \leq \varepsilon(1 + K_0K_1 \int_0^T (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)d\sigma)$$

$$+ \quad \sum_{j=0}^{k-1} K_0K_1 \int_{t_j^N}^{t_{j+1}^N} (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)d\sigma|z_j^N - z(t_j^N)|$$

$$\leq \quad \varepsilon(1 + K_0K_1T|M_1|_{L_\infty} + K_0K_1|M_2|_{L_\infty} T^{1/2}|u|_{L_2})$$

$$+ \sum_{j=0}^{k-1} K_0K_1 \int_{t_j^N}^{t_{j+1}^N} (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)d\sigma |z_j^N - z(t_j^N)|$$

for all N sufficiently large. If we now apply Lemma 3.1 to (3.5) it then follows that

$$|z_k^N - z(t_k^N)| \le \varepsilon(1+\gamma)\exp(K_0K_1 \sum_{j=0}^{k-1} \int_{t_j^N}^{t_{j+1}^N} (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)d\sigma)$$

$$\le \varepsilon(1+\gamma)\exp(K_0K_1 \int_0^T (M_1(\sigma) + M_2(\sigma)|u(\sigma)|)d\sigma)$$

$$\le \varepsilon(1+\gamma)\exp(\gamma)$$

for all N sufficiently large, where

$$\gamma = K_0K_1T|M_1|_{L_\infty} + K_0K_1|M_2|_{L_\infty} T^{1/2}|u|_{L_2}$$ and the theorem is proven.

<u>Corollary</u> 3.1 For $\left\{z_k^N\right\}_{k=0}^{\rho_N}$ generated by an approximation scheme satisfying the hypotheses of Theorem 3.3 it follows that

$$|\pi_1 z_k^N((\eta,\phi),u) - x(t_k^N;\eta,\phi,u)| \to 0$$

as $N \to \infty$ uniformly in k, $k = 0,1,2...\rho_N$ and uniformly in u, $u \in E$ where x denotes the unique solution to the initial value problem (2.1)(2.2).

4. <u>Construction of Convergent Approximation Schemes</u>

In this section we construct approximation schemes which are based upon the framework described in the previous section and which satisfy the hypotheses of Theorem 3.3. Since the schemes described below and the verification of the fact that they satisfy the hypotheses of Theorem 3.3 have appeared elsewhere ([2],[4],[8],[24]), the relevant results are outlined and the details

Each of our approximation schemes is composed of two interrelated components, the state discretization, as is characterized by the choice of Z_N, P_n, and A_n, and the temporal discretization which is determined by the rational functions $C(z)$ and $D(z)$. The interrelation which exists between the two components is a consequence of the conditions under which our fundamental convergence result, Theorem 3.3, applies. We begin with a description of two state approximations and then discuss and characterize families of rational functions which, when coupled with these state approximations lead to convergent approximation schemes.

The averaging state approximation (AVE), the more primitive of the two state approximations to be discussed is based upon finite difference approximations and is defined as follows. For each $N = 1,2...$ let χ_j^N $j = 1,2...N$ denote the characteristic function on the interval $[-j\frac{r}{N}, -(j-1)\frac{r}{N})$ and let

$$Z_N = \{(\eta,\phi) \in Z: \phi = \sum_{j=1}^{N} v_j\chi_j^N, v_j \in R^n\}.$$

We note that $\dim Z_N = n(N+1)$ and that the orthogonal (with respect to the standard Z inner product) projections $P_N:Z \to Z_N$ are given by

$$P_N(\eta,\phi) = (\eta, \sum_{j=1}^{N} \phi_j^N\chi_j^N)$$

where $\phi_j^N = \frac{N}{r}\int_{-j\frac{r}{N}}^{-(j-1)\frac{r}{N}} \phi(\theta)d\theta$. It is not difficult to show that $P_N z \to z$ as $N \to \infty$ for each $z \in Z$. Let the operators $L_N:Z_N \to R^n$ and $D_N:Z_N \to L_2^n(-r,0)$ be given by

$$L_N(\eta, \sum_{j=1}^{N} v_j \chi_j^N) = A_0\eta + \sum_{i=1}^{\nu} \sum_{j=1}^{N} A_i v_j \chi_j^N(-\tau_i)$$

$$+ \frac{r}{N} \sum_{j=1}^{N} A_j^N v_j$$

where $A_j^N = \frac{N}{r} \int_{-j\frac{r}{N}}^{-(j-1)\frac{r}{N}} A(\theta)d\theta \quad j = 1,2\ldots N$ and

$$D_N(\eta, \sum_{j=1}^{N} v_j \chi_j^N) = \sum_{j=1}^{N} \frac{N}{r} (v_{j-1} - v_j)\chi_j^N$$

where $v_0 = \eta$ respectively. Define $\mathcal{A}_N : Z_N \to Z_N$ by

$$\mathcal{A}_N(\eta,\phi) = (L_N(\eta,\phi), D_N(\eta,\phi))$$

and for $t \geq 0$ let $S_N(t) = e^{\mathcal{A}_N t}$. A sequence of inner products on Z, $\langle\cdot,\cdot\rangle_N$, can be constructed for which there exists an $M > 0$, independent of N, such that

(4.1) $$|(\eta,\phi)|_N \leq |(\eta,\phi)| \leq M|(\eta,\phi)|_N$$

for all $(\eta,\phi) \in Z_N$. Furthermore, there exists a $\beta > 0$ independent of N, for all N sufficiently large for which the operators $\mathcal{A}_N - \beta I$ are maximal dissipative with respect to the $\langle\cdot,\cdot\rangle_N$ inner product on Z_N. It follows therefore that $\mathcal{A}_N \in G(M,\beta)$ and $|S_N(t)| \leq Me^{\beta t}$ for all N sufficiently large. It is in fact the case that the $\mathcal{A}_N$ as defined above satisfy a somewhat stronger condition. It can be shown that there exist an $\alpha > 0$ for which

(4.2) $$|I + \frac{r}{N}\mathcal{A}_N|_N \leq 1 + \alpha\frac{r}{N}.$$

apparent when we discuss the choice of the rational function component of the approximation scheme below.

If we let $D_1 = D(A^2)$ and $D_2 = D(A)$ then $D_1 \subset D(A)$ is a dense subset of Z and for all $\lambda \in C$ with $\mathrm{Re}\lambda > \beta, R(\lambda;A)D_2 = D_1$. Moreover, it can be shown that for each $z \in D_1$

$$|A_N P_N z - Az| = O(N^{-1/2})$$

as $N \to \infty$.

We shall next describe a spline based state approximation. The discussions which follow will be restricted to constructions involving linear, or first order spline functions. However, the results given below are easily generalized so as to be applicable to state approximations employing higher order spline functions. For each $N = 1,2\ldots$ and $\theta \in [-r,0]$ let

$$\phi_0^N(\theta) = \begin{cases} \frac{N}{r}(\theta - t_1^N) & t_1^N \le \theta \le 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\phi_j^N(\theta) = \begin{cases} \frac{N}{r}(t_{j-1}^N - \theta) & t_j^N \le \theta \le t_{j-1}^N \\ \frac{N}{r}(\theta - t_{j+1}^N) & t_{j+1}^N \le \theta \le t_j^N \\ 0 & \text{otherwise} \end{cases} \qquad j = 1,2\ldots N-1$$

$$\phi_N^N(\theta) = \begin{cases} \frac{N}{r}(t_{N-1}^N - \theta) & -r \le \theta \le t_{N-1}^N \\ 0 & \text{otherwise} \end{cases}$$

where $t_j^N = -j\frac{r}{N}$ $j = 0,1,2\ldots N$ and define Z_N by

$$Z_N = \{(\phi(0),\phi) \in Z: \phi = \sum_{j=0}^{N} v_j \phi_j^N, \ v_j \in R^n\}.$$

It is immediately clear that $\dim Z_N = r(N + 1)$, $Z_N \subset D(A)$ and Z_N consists of all those elements $(\eta,\phi) \in Z$ for which $\eta = \phi(0)$ and ϕ is a first order spline function with knots at $\{t_j^N\}_{j=0}^N$. Let $P_N: Z \to Z_N$ denote the orthogonal projection from Z onto Z_N computed with respect to the weighted inner product on Z, $\langle\cdot,\cdot\rangle_g$, defined in Section 2. Finally we define the operators $A_N: Z_N \to Z_N$ by

$$A_N = P_N A .$$

Using the fact that the P_N are orthogonal projections it follows from (2.3) that for $z_N \in Z_N$

$$(4.3) \qquad \langle A_N z_N, z_N\rangle_g = \langle P_N A z_N, z_N\rangle_g = \langle A z_N, P_N z_N\rangle_g$$
$$= \langle A z_N, z_N\rangle_g \le \omega \langle z_N, z_N\rangle_g$$

and hence that $A_N \in G(\sqrt{\nu},\omega)$. Furthermore, using the properties of interpolatory splines it is not difficult to show that $P_N z \to z$ as $N \to \infty$ for each $z \in Z$ and that

$$(4.4) \qquad |A_N P_N z - Az| = O(N^{-1})$$

for each $z \in D_1 \equiv D(A^3)$. If we choose $D_2 = D(A^2)$ then all of the hypotheses and conditions of Theorem 3.3 concerning the state approximation only hold for the linear spline scheme defined above. We note that for state approximations employing higher order spline functions the order of convergence in (4.4) and therefore in the integration method itself (see Remark 3.1) can be increased.

For either the AVE or spline based state approximations, a rational function $C(z)$ satisfying conditions of Theorem 3.1 must be chosen for which the operators $\left\{C(\frac{r}{n} A_N)^{\ell}\right\}_{\ell=0}^{\rho_N}$ are uniformly bounded in N for all N sufficiently large. It is clear from condition (5a) that we are seeking rational function approximations to the exponential. While there are many families of approximating rational functions from which to choose, we have restricted our attention to the well known Padé approximants [26] which are given by $P_{jk}(z) = N_{jk}(z)/D_{jk}(z)$ where

$$(4.5) \qquad N_{jk}(z) = \sum_{i=0}^{k} \frac{(j+k-i)!k!}{(j+k)!i!(k-i)!} z^i$$

and

$$(4.6) \qquad D_{jk}(z) = \sum_{i=0}^{j} \frac{(j+k-i)!j!}{(j+k)!i!(j-i)!} (-z)^i .$$

It can be shown that

$$|P_{jk}(z) - e^z| = O(|z|^{j+k+1}) \text{ as } |z| \to 0$$

and hence the Padé approximants satisfy condition (5b) since $\deg P_{jk}(z) = k - j \le k + j + 1$. It is immediately clear from (4.5), (4.6) that $\left\{P_{0k}(z)\right\}_{k=0}^{\infty}$ are the Maclaurin polynomials for e^z and therefore satisfy condition (5c). Furthermore, Ehle [11], in his study of the use of the Padé approximants in the construction of A-stable integration schemes for stiff systems of ordinary differential equations has shown that for $z \in \{z \in C: \text{Re}\, z \le 0\}$

$$(4.7) \qquad |P_{jk}(z)| \le 1 \qquad j = k, k+1, k+2 \quad k = 0,1,2\ldots$$

Consequently, from the standpoint of the constraint that condition 5 of Theorem 3.1 be satisfied, $C(z)$ can be chosen from among the entries in the top row, the principal diagonal and the first two subdiagonals of the Padé table. However, the convergence of approximation schemes constructed using these rational functions and the AVE or spline based state approximations defined above is guaranteed by Theorem 3.3 only if the uniform boundedness of the operators $\left\{P_{jk}(\frac{r}{N}A_N)^{\ell}\right\}_{\ell=0}^{\rho_N}$ can be demonstrated.

Using the von Neumann theory of spectral sets [22] and a result due to Hersh and Kato [14] the following result can be obtained.

Theorem 4.1 Let T be a bounded linear operator on a Hilbert space H for which there exist a $\beta > 0$ such that $\langle Tx,x\rangle \le \beta\langle x,x\rangle$ for all $x \in H$, and let $r(z)$ be a rational function satisfying condition 5 of Theorem 3.1. Then if $|r(z)| \le 1$ for all $z \in \{z \in C: \mathrm{Re}\, z \le 0\}$ we have

$$|r(hT)| \le 1 + \beta K h$$

where K is a positive constant independent of h and T.

It follows immediately from the dissipative properties of the operators A_N defined as a part of the AVE state approximation, (4.1),(4.7), and Theorem 4.1 that for $j = k, k+1, k+2, k = 1,2\ldots$ and $\ell = 0,1,2\ldots\rho_N$

$$|P_{jk}(\tfrac{r}{N}A_N)^{\ell}| \le M(1 + \beta K_{jk}\tfrac{r}{N})^{\ell}$$
$$\le Me^{\beta K_{jk}\rho_N \frac{r}{N}} \quad Me^{\beta K_{jk}T} .$$

Similarly, for the spline based state approximations it follows from (4.3) that

$$|P_{jk}(\tfrac{r}{N}A_N)^{\ell}| \le \sqrt{\nu}\, e^{\omega K_{jk}T}$$

In addition, for the AVE state approximation which satisfies (4.2), it can be shown independently of Theorem 4.1 that for $k = 1,2\ldots$ and

$\ell = 0,1,2\ldots\rho_N$

$$|P_{0k}(\tfrac{r}{N} A_N)^{\ell}| \le Me^{\alpha T} .$$

Although, as far as the convergence of the approximation scheme is concerned, it would suffice to choose $D(z) \equiv 1$, and hence $D(\lambda\frac{r}{N} A_N) = I$ (see Theorem 3.2), empirical evidence can be given, and an intuitive argument can be made for choosing $D(z)$ as a rational function approximation to the exponential. It is easily verified that any rational function approximation to the exponential which is a suitable choice for $C(z)$ is a suitable choice for $D(z)$ as well. In addition, for the spline based state approximations and $k = 1,2\ldots$ it can be shown that $|P_{0k}(\lambda\frac{r}{N} A_N)P_N z - z| \to 0$ as $N \to \infty$ for each $z \in Z$. A more detailed description of the role played by the rational function $D(z)$ and its effect upon the overall performance of the approximation scheme can be found in [24].

The results in this section are summarized in the following theorem.

Theorem 4.2 For $\{Z_N, P_N, A_N\ C(z), D(z)\}$ an approximation scheme for the initial value problem (2.1)(2.2), the hypotheses and conditions of Theorem 3.3 are satisfied if

(1) Z_N, P_N, A_N is an AVE state approximation and $C(z), D(z) \in \mathscr{D}_p \cup \mathscr{M}_p$

or

(2) Z_N, P_N, A_N is a spline based state approximation, $C(z) \in \mathscr{D}_p$ and $D(z) \in \mathscr{D}_p \cup \mathscr{M}_p$

where $\mathscr{D}_p = \{P_{jk}(z)\}$ $j = k,k+1,k+2,$ $k = 1,2\ldots$ and $\mathscr{M}_p = \{P_{0k}(z)\}$ $k = 1,2\ldots$.

5. Application to Optimal Control Problems

In this section we consider the application of the approximation results discussed above to the solution of optimal control problems in which the state is governed by a nonlinear hereditary system of the form (2.1). In particular let $\phi_1 : R^n \to R^1, \phi_2 : L_2^n(0,T) \to R^1$ be continuous and let $\phi_3 : L_2^m(0,T) \to R^1$ be continuous and convex. Let U be a closed convex subset of $L_2^m(0,T)$ and define problem (P) as follows

(P)

$$\text{Minimize } \Phi(u) = \phi_1(x(T;\eta,\phi,u)) + \phi_2(x(\cdot;\eta,\phi,u)) + \phi_3(u)$$

over all $u \in U$ where $x(\cdot;\eta,\phi,u)$ denotes the unique solution to (2.1),(2.2) corresponding to $u \in U$.

The approach we take is to consider a sequence of approximating optimal control problems $\{(P_N)\}$, in each of which the governing state equation is a finite dimensional discrete difference equation constructed in accordance with the approximation framework developed in Section 3. Let $\{Z_N, P_N, A_N, C(z), D(z)\}$ be an approximation scheme for (2.1),(2.2) which satisfies the hypotheses of Theorem 3.3 and for $z_0 = (\eta,\phi)$, $u \in L_2^m(0,T)$ and $k = 0,1,2\ldots\rho_N$ let

$$z_k^N(z_0,u) = (x_k^N(z_0,u), y_k^N(z_0,u))$$

where $\left\{z_k^N(z_0,u)\right\}_{k=0}^{\rho_N}$ are given by (3.3) with $x_k^N(z_0,u) \in \mathbb{R}^n$ and $y_k^N(z_0,u) \in L_2^n(0,T)$. Define $x^N \in L_2^n(0,T)$ by

$$x^N(\theta) = x^N(\theta;\eta,\phi,u) = \sum_{j=0}^{\rho_N} x_j^N(z_0,u)\chi_{[j\frac{r}{N},(j+1)\frac{r}{N})}(\theta)$$

and for each $N = 1,2\ldots$ let problem (P_N) be given by

(P_N)

$$\text{Minimize} \quad \Phi_N(u) = \phi_1(x^N_{\rho_N}(z_0,u)) + \phi_2(x^N(\cdot;\eta,\phi,u)) + \phi_3(u)$$

over all $u \in U$.

Remark 5.1 While it is true that for each $N = 1,2\ldots$ problem (P_N) is not fully discrete in that the minimization of Φ is being considered over a function space, it is in fact possible to define the problem in a form which is directly suitable for solution on the computer. Indeed, if we consider the minimization over the set $U_N \equiv Q_N U \subset \overset{\rho_N-1}{\underset{0}{\times}} R^m$ where $Q_N : L_2^m(0,T) \to \overset{\rho_N-1}{\underset{0}{\times}} R^m$ is defined by

$$(Q_N u)_j = \frac{N}{r} \int_{j\frac{r}{N}}^{j+1\frac{r}{N}} u(\tau)d\tau \qquad j = 0,1,2\ldots\rho^N-1,$$

then by placing relatively minor restrictions on the choice of the set U, all of the convergence results for the solutions to the sequence of problems $\{(P_N)\}$ to be discussed below can be shown to hold for the fully discrete problems as well. In order to simplify the presentation, however, we shall restrict our attention to the approximating problems as given.

It is our ultimate goal to demonstrate that in some sense, solutions to problem (P_N) approximate solutions to problem (P). However before this can be accomplished, the existence of solutions to problems (P) and (P_N) must be considered. In order to insure the convexity of Φ and Φ_N with respect to u it is necessary that we restrict f, the nonlinear part of the state equation to be affine in the controls. Following Banks [2], henceforth we shall assume that $f:R^1 \times R^n \times L_2^n(-r,0) \times R^m \to R^n$ is of the form

$$(5.1) \qquad f(t,\eta,\phi,v) = f_1(t,\eta,\phi) + (f_2(t,\eta,\phi) + B(t))v$$

where B is continuous and $f_1:R^1 \times R^n \times L_2^n(-r,0) \to R^n$ and $f_2:R^1 \times R^n \times L_2^n(-r,0) \to R^{n\times m}$ satisfy the following hypotheses.

(1) The mappings $(t,\eta,\phi) \to f_i(t,\eta,\phi)$ $\quad i = 1,2$ are continuous on $R^1 \times R^n \times L_2^n(-r,0)$.

(2) For any bounded subset $\mathcal{D}$ of $R^n \times L_2^n(-r,0)$ there exist $m_i = m_i(\mathcal{D})$, $m_i \in L_\infty^{loc}$ $\quad i = 1,2$ such that for $t \in R^1$ and (η,ϕ), $(\xi,\psi) \in \mathcal{D}$ one has
$|f_i(t,\eta,\phi) - f_i(t,\xi,\psi)| \le m_i(t)\{|\eta-\xi| + |\phi-\psi|\}$

(3) For $i = 1,2$, $f_i(t,0,0) = 0$ and there exist functions $\hat{m}_i \in L_\infty^{loc}$ such that for $t \in R^1$
$|f_i(t,\eta,\phi)| \le \hat{m}_i(t)\{|\eta|+|\phi|\}$

for $(\eta,\phi) \in R^n \times L_2^n(-r,0)$ with $|\eta| + |\phi|$ sufficiently large. It is immediately clear that any function f of the form (5.1) satisfying (1) - (3) above will also satisfy hypotheses (H1)-(H4).

In addition, it is necessary that we make either one or the other of the following two assumptions

(A1) The set U is bounded

(A2) The mappings ϕ_i $\quad i = 1,2,3$ satisfy

(i) $\phi_i \ge 0$ $\quad i = 1,2$

(ii) $\phi_3(u) \to \infty$ if $|u| \to \infty$.

We note that problem (P) is most commonly stated with $U = L_2^m(0,T)$ and Φ a quadratic of the form

(5.2) $$\Phi(u) = x(t;\eta,\phi,u)^T G x(T;\eta,\phi,u) + \int_0^T x(s;\eta,\phi,u)^T Q x(s;\eta,\phi,u)ds + \int_0^T u(s)^T R u(s)ds$$

where G and Q are positive semi-definite $n\times n$ matrices and R is a positive definite $m\times m$ matrix. In this case, assumption (A2) holds.

<u>Lemma</u> 5.1 For f of the form (5.1) satisfying hypotheses (1)-(3), $z_0 \in Z$ and $u_\ell \to u$ weakly in $L_2^m(0,T)$ we have

(5.3) $$|z(t;z_0,u_\ell) - z(t;z_0,u)| \to 0$$

as $\ell \to \infty$ uniformly in t for $t \in [0,T]$ and for $N = 1,2\ldots$ fixed we have

(5.4) $$|z_k^N(z_0,u_\ell) - z_k^N(z_0,u)| \to 0$$

as $\ell \to \infty$ uniformly in k for $k = 0,1,2\ldots\rho_N$ where $z(t;z_0,u)$ and $z_k^N(z_0,u)$ are given by (2.4) and (3.3) respectively.

The proof of (5.3) follows from Theorem 3.2 of [2] while similar arguments and Lemma 3.1 can be used to verify (5.4).

<u>Corollary</u> 5.1 Under the hypotheses of Lemma 5.1, if $\{u_N\}$ is a sequence in $L_2^m(0,T)$ for which $u_N \to u$ weakly then

$$|z_k^N(z_0,u_N) - z(t_k^N;z_0,u)| \to 0$$

as $N \to \infty$ uniformly in k, $k = 0,1,2\ldots\rho_N$.

<u>Proof</u> Since

$$(5.5) \qquad |z_k^N(z_0,u_N) - z(t_k^N;z_0,u)|$$
$$|z_k^N(z_0,u_N) - z(t_k^N;z_0,u_N)| + |z(t_k^N;z_0,u_N) - z(t_k^N;z_0,u)|$$

and $u_N \to u$ weakly implies that $\{u_N\}$ lies in a bounded subset of $L_2^m(0,T)$, the first term on the right hand side of (5.5) tends toward zero as $N \to \infty$ uniformly in k, $k = 0,1,2\ldots\rho_N$ as a consequence of Theorem 3.3 while Lemma 5.1 insures that the second term tends toward zero in the stated manner as well.

<u>Theorem</u> 5.1 If either assumption (A1) or (A2) hold and f is of the form (5.1) satisfying hypotheses (1)-(3) then problems (P) and (P_N) have solutions.

<u>Proof</u> Lemma 5.1, ϕ_i continuous, $i = 1,2,3$ and ϕ_3 convex imply that Φ and Φ_N are weakly semi-continuous from below. Therefore, if U is bounded, Φ and Φ_N will assume their infimum on U (see [19] Existence Theorem, page 90) and the theorem is proven.

On the other hand, suppose Assumption (A2) holds, and let $\{u_i\} \in U$ be such that

$$\Phi(u_i) \to \alpha = \inf\{\Phi(u) : u \in U\}.$$

Note that $\phi_i \geq 0$ $i = 1,2,3$ implies that $0 \leq \alpha < \infty$. Since U is closed and convex (and therefore weakly sequentially closed) and $\{u_i\}$ is bounded (Assumption (A2)), $\{u_i\}$ must contain a weakly convergent subsequence $\{u_{i_j}\}$, $u_{i_j} \to \bar{u} \in U$, weakly. However, Φ weakly semi-continuous from below implies that

$$\alpha \le \Phi(\bar{u}) \le \liminf \Phi(u_{i_j}) = \alpha .$$

and hence $\Phi(\bar{u}) = \alpha$, and $\bar{u}$ is a solution to problem (P). A similar argument may be used to demonstrate the existence of a solution $\bar{u}_N \in U$ to problem (P_N).

<u>Theorem</u> 5.2 Suppose that the hypotheses of Theorem 5.1 hold and for each $N = 1,2\ldots,\bar{u}_N$ denotes a solution to problem (P_N). Then $\{\bar{u}_N\}$ contains a subsequence $\{\bar{u}_{N_k}\}$ for which $\bar{u}_{N_k} \to \bar{u} \in U$ weakly. Moreover, $\bar{u}$ is a solution to problem (P) and $\Phi_{N_k}(\bar{u}_{N_k}) \to \Phi(\bar{u})$ as $k \to \infty$.

<u>Proof</u> Under either assumption (A1) or (A2) the sequence $\{\bar{u}_N\}$ is bounded. It therefore must contain a weakly convergent subsequence $\{\bar{u}_{N_k}\}$. If $\bar{u} \in U$ is such that $\bar{u}_{N_k} \to \bar{u}$ weakly as $k \to \infty$ then Corollary 3.1, Corollary 5.1 and the weak semi-continuity from below of ϕ_3 (it being continuous and convex) imply that

$$\Phi(\bar{u}) = \phi_1(x(T;\eta,\phi,\bar{u})) + \phi_2(x(\cdot;\eta,\phi,\bar{u})) + \phi_3(\bar{u})$$

$$\le \lim_{k\to\infty} \phi_1(x^N_{\rho_N}((\eta,\phi),\bar{u}_{N_k})) + \lim_{k\to\infty} \phi_2(x^N(\cdot;\eta,\phi,\bar{u}_{N_k}))$$

$$+ \liminf_{k\to\infty} \phi_3(\bar{u}_{N_k})$$

$$= \liminf_{k\to\infty} \Phi_{N_k}(\bar{u}_{N_k}) \le \limsup_{k\to\infty} \Phi_{N_k}(\bar{u}_{N_k})$$

$$\le \limsup_{k\to\infty} \Phi_{N_k}(u) = \lim_{k\to\infty} \Phi_{N_k}(u) = \Phi(u)$$

for arbitrary $u \in U$, and hence that $\bar{u}$ is a solution to problem (P). The fact that $\Phi_{N_k}(\bar{u}_{N_k}) \to \Phi(\bar{u})$ as $k \to \infty$ follows from

$$\Phi(\bar{u}) \le \liminf_{k\to\infty} \Phi_{N_k}(\bar{u}_{N_k}) \le \limsup_{k\to\infty} \Phi_{N_k}(\bar{u}_{N_k})$$

$$\le \limsup \Phi_{N_k}(\bar{u}) = \lim \Phi_{N_k}(\bar{u}) = \Phi(\bar{u}).$$

Remark Since it is difficult to determine the convexity properties of the functional Φ it is not possible to say anything about the uniqueness of solutions to problem (P). However, if in fact problem (P) has a unique solution, then the sequence itself, $\{\bar{u}_N\}$ will converge to $\bar{u}$ weakly as $N \to \infty$.

Remark If Φ is of the form (5.2) then it is possible to show that $|\bar{u}_{N_k}| \to |\bar{u}|$ as well, and hence that $\bar{u}_{N_k} \to \bar{u}$ strongly as $k \to \infty$. Once again if problem (P) admits a unique solution $\bar{u}$, then $\bar{u}_N \to \bar{u}$ strongly as $N \to \infty$.

6. Analysis of Numerical Results

In this section we present numerical results obtained through the implementation of the approximation schemes described above. The schemes employed have been constructed using the AVE and spline based (SPL) state approximations together with the Padé rational function approximations to the exponential. In all of the examples below, however, we have chosen $C(z) = D(z) = P_{22}(z)$ and $\lambda = \frac{1}{2}$. The effect of varying the choice of the rational function components of the approximation scheme (from among those in the Padé table for which the hypotheses of Theorem 3.3 are satisfied) was studied extensively in [24].

We have included one example involving the integration of an initial value problem of the form (2.1)(2.2) only and three other examples which involve the solution of an optimal control problem of the form given by problem (P) in Section 5. We have deliberately chosen to include examples which have been used by other authors to test other approximation schemes for the integration of FDE and the solution of FDE control problems so that our methods can be compared to theirs. The other places where each example has appeared has been so noted.

All programming was done in FORTRAN and implemented on the Digital Equipment Corporation DEC system 10 computer at Bowdoin College. The optimization in each of the approximating problems (P_N) was carried out using the IMSL [15] routine ZXMIN, an iterative quasi-Newton algorithm for finding the minimum of a scalar valued function of several variables. The discretization of the admissible control space U in the approximating optimal control problems (P_N) was done in two different ways. One involved the use of the space $U_N = \overset{\rho_N}{\underset{0}{\times}} R^m$ as an approximation to $L_2^m(0,T)$ (see Remark 5.1). In this case the number of parameters over which the minimization takes place increases with the degree of approximation N. The second approach was to minimize over the space $\tilde{U} = \overset{L}{\underset{0}{\times}} R^m$ where L is a fixed constant independent of N. A cubic spline interpolation scheme was then used to obtain the values of the control which are required to evaluate (3.3). The approximate solutions resulting from the two methods were virtually indistinguishable. However, the number of iterations required to obtain the minimizing control increased like $O(N)$ for the first method; while the iteration count remained essentially constant for all values of N for the second method.

Since, with the exception of Example 6.2 which has a linear state equation, it is impossible to obtain exact solutions to the optimal control problems below, we have included approximate solutions which were obtained using methods independent from our own. These alternate approximate solutions, which can be used for comparision, were computed by Daniel [10] using a fourth order integration scheme for FDE developed by Tavernini [25] to solve the mixed retarded/advanced two point boundary value problem which

Example 6.1 (Banks [2], Example 4.1)

We consider the integration of the equation

$$\dot{x}(t) = -1.5x(t) - 1.25x(t-1) + x(t)\sin x(t)$$

on the interval $0 \le t \le 5$ with initial data

$$x(0) = 1 \qquad x_0(s) = 10s + 1 \qquad -1 \le s \le 0$$

The approximate solutions generated by the AVE and SPL state approximations are given in Tables 6.1 and 6.2 respectively. The values in the last column of each of the tables were computed using the method of steps [12] together with a fourth order Runge-Kutta routine for ordinary differential equations and may be used for comparison purposes.

t	$x_4^{AVE}(t)$	$x_8^{AVE}(t)$	$x_{16}^{AVE}(t)$	$x_{32}^{AVE}(t)$	$x(t)$
0.0	1.0	1.0	1.0	1.0	1.0
.5	3.0954	3.1924	3.2531	3.2840	3.3142
1.0	2.1375	2.2051	2.2522	2.2841	2.3317
1.5	.9759	.7151	.5163	.3877	.2294
2.0	- .2258	- .6233	- .8116	- .9020	- .9909
2.5	- .5984	- .6920	- .7221	- .7331	- .7399
3.0	- .3491	- .2599	- .1715	- .1073	- .0245
3.5	- .0573	.1091	.2409	.3251	.4259
4.0	.1024	.2389	.3244	.3711	.4195
4.5	1.2229	.1598	.1532	.1370	.1081
5.0	.0634	.0150	- .0469	- .0919	- .1480

Table 6.1

t	$x_4^{SPL}(t)$	$x_8^{SPL}(t)$	$x_{16}^{SPL}(t)$	$x_{32}^{SPL}(t)$	$x(t)$
0.0	1.0038	1.0010	1.0003	1.0001	1.0
.5	3.5036	3.3623	3.3344	3.3236	3.3142
1.0	2.1694	2.2636	2.2992	2.3157	2.3317
1.5	.3642	.2834	.2538	.2405	.2294
2.0	-1.0308	- .9972	- .9929	- .9919	- .9909
2.5	- .7248	- .7332	- .7345	- .7367	- .7399
3.0	- .0612	- .0218	- .0188	- .0205	- .0245
3.5	.4055	.4166	.4230	.4251	.4259
4.0	.4180	.4145	.4157	.4173	.4195
4.5	.1572	.1187	.1099	.1081	.1081
5.0	- .1124	- .1391	- .1454	- .1473	- .1480

Table 6.2

Example 6.2 (Banks, Burns, Cliff [5], Example C7, Rockey [23], Test Problem 5.6)

In this example we consider an optimal control problem whose state equation is a linear harmonic oscillator with delayed damping.

$$\text{Minimize} \quad \Phi(u) = 5y(2)^2 + \frac{1}{2}\int_0^2 u(s)^2 ds$$

over $u \in U = L_2^1(0,2)$ subject to

(6.1) $$\ddot{y}(t) + \dot{y}(t-1) + y(t) = u(t)$$

with initial conditions

(6.2) $$y(0) = 10 \qquad y_0(s) = 10 \qquad -1 \le s \le 0$$

(6.3) $$\dot{y}(0) = 0 \qquad \dot{y}_0(s) = 0 \qquad -1 \le s \le 0$$

For this problem, the true optimal control $\bar{u}$ may be computed, and is given

$$\bar{u}(t) = \begin{cases} \delta \sin(2-t) + \frac{\delta}{2}(1-t)\sin(t-1) & 0 \le t \le 1 \\ \delta \sin(2-t) & 1 \le t \le 2 \end{cases}$$

where $\delta \approx 2.5599$, with $\Phi(\bar{u}) = 3.3991$. This example may be put in the form of problem (P) by transforming (6.1)(6.2)(6.3) into an equivalent first order system, which is given by

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} x(t-1) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

$$x(0) = \begin{bmatrix} 10 \\ 0 \end{bmatrix} \qquad x_0(s) = \begin{bmatrix} 10 \\ 0 \end{bmatrix} \qquad -1 \le s \le 0$$

where $x(t) = \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix}$. The payoff functional Φ would now take the form

$$\Phi(u) = x^T(2) \begin{bmatrix} 5 & 0 \\ 0 & 0 \end{bmatrix} x(2) + \int_0^2 u(s)^2 ds .$$

Tables 6.3 and 6.4 contain the resulting approximating optimal controls.

t	$\bar{u}_4^{AVE}(t)$	$\bar{u}_8^{AVE}(t)$	$\bar{u}_{16}^{AVE}(t)$	$\bar{u}_{32}^{AVE}(t)$	$\bar{u}(t)$
0.0	1.2757	1.2797	1.2746	1.2336	1.2506
.25	1.4515	1.6358	1.7463	1.8024	1.8645
.50	1.7195	1.9506	2.0888	2.1706	2.2467
.75	1.8076	2.0427	2.1827	2.2642	2.3501
1.00	1.7070	1.9094	2.0263	2.0911	2.1541
1.25	1.4333	1.5844	1.6641	1.7075	1.7449
1.50	1.0255	1.1216	1.1718	1.2018	1.2273
1.75	.5324	.5794	.6043	.6164	.6333
2.0	.2708	.1473	.0776	[illegible]	[illegible]

t	$\bar{u}_4^{SPL}(t)$	$\bar{u}_8^{SPL}(t)$	$\bar{u}_{16}^{SPL}(t)$	$\bar{u}_{32}^{SPL}(t)$	$\bar{u}(t)$
0.0	1.6887	1.4468	1.3456	1.2776	1.2506
.25	1.9415	1.8856	1.8748	1.8686	1.8645
.50	2.3024	2.2635	2.2553	2.2501	2.2467
.75	2.3675	2.3570	2.3583	2.3521	2.3501
1.00	2.1634	2.1573	2.1482	2.1539	2.1541
1.25	1.7592	1.7489	1.7458	1.7443	1.7449
1.50	1.2238	1.2269	1.2261	1.2276	1.2273
1.75	.61496	.6301	.6285	.6339	.6333
2.0	.2999	.1548	.0798	.0392	0.000
$\Phi_N(\bar{u}_N)$	3.5664	3.4438	3.411	3.4021	3.3991

Table 6.4

Example 6.3 (Banks [2], Example 4.4, Banks, Burns, Cliff [5], Example C11, Daniel [10], Example 4.5, Rockey [23], Test Problem 5.10)

In this example we consider an optimal control problem with a one dimensional nonlinear state equation

$$\text{Minimize} \quad \Phi(u) = \frac{1}{2} x(2)^2 + \frac{1}{2} \int_0^2 x(s)^2 + u(s)^2 ds$$

over $u \in L_2^1(0,2)$ subject to

$$\dot{x}(t) = x(t-1) + x(t) \sin x(t) + u(t)$$

with initial conditions given by

$$x(0) = 0 \qquad x_0(s) = \begin{cases} -10s & -\frac{1}{2} \le s \le 0 \\ 10(s+1) & -1 \le s \le \frac{1}{2} \end{cases}$$

The approximating minimizing controls for the AVE and SPL state approximations are given in Tables 6.5 and 6.6 respectively while Tables 6.7 and 6.8 contain the corresponding optimal trajectories.

t	$\bar{u}_4^{AVE}(t)$	$\bar{u}_8^{AVE}(t)$	$\bar{u}_{16}^{AVE}(t)$	$\bar{u}_{32}^{AVE}(t)$	$\bar{u}(t)$
0.0	-2.1967	-2.2417	-2.2681	-2.2817	-2.3028
.25	-2.0860	-2.1699	-2.2295	-2.2662	-2.3164
.50	-1.8082	-1.9655	-2.0971	-2.1893	-2.3189
.75	-1.4605	-1.5635	-1.6386	-1.6853	-1.7470
1.00	-1.1242	-1.1470	-1.1443	-1.1333	-1.1031
1.25	- .8467	- .8273	- .7972	- .7747	- .7483
1.50	- .6332	- .6072	- .5838	- .5708	- .5619
1.75	- .4665	- .4484	- .4376	- .4349	- .4440
2.0	- .3921	- .3477	- .3282	- .3223	- .3230
$\Phi_N(\bar{u}_N)$	1.9914	2.1673	2.3020	2.3953	

Table 6.5

t	$\bar{u}_4^{SPL}(t)$	$\bar{u}_8^{SPL}(t)$	$\bar{u}_{16}^{SPL}(t)$	$\bar{u}_{32}^{SPL}(t)$	$\bar{u}(t)$
0.0	-2.2389	-2.2372	-2.2596	-2.2741	-2.3028
.25	-2.3139	-2.3019	-2.3023	-2.3041	-2.3164
.50	-2.1999	-2.2596	-2.2908	-2.3022	-2.3189
.75	-1.6929	-1.7129	-1.7295	-1.7364	-1.7470
1.00	-1.1384	-1.1198	-1.1117	-1.1070	-1.1031
1.25	- .7830	- .7702	- .7610	- .7569	- .7483
1.50	- .6151	- .5874	- .5749	- .5682	- .5619
1.75	- .5032	- .4543	- .4531	- .4469	- .4440
2.0	- .4417	- .3676	- .3461	- .3351	- .3230
$\Phi_N(\bar{u}_N)$	2.5119	2.4996	2.5103	2.5133	

Table 6.6

t	$\bar{x}_4^{AVE}(t)$	$\bar{x}_8^{AVE}(t)$	$\bar{x}_{16}^{AVE}(t)$	$\bar{x}_{32}^{AVE}(t)$	$\bar{x}(t)$
0.0	0.0	0.0	0.0	0.0	0.0
.25	– .0087	– .1034	– .1672	– .2051	– .2473
.50	.1537	.1434	.1357	.1271	.1078
.75	.2757	.3540	.4329	.4931	.5663
1.00	.3282	.4199	.5009	.5562	.6186
1.25	.3368	.3901	.4182	.4259	.4127
1.50	.3314	.3393	.3233	.3006	.2474
1.75	.3337	.3139	.2886	.2687	.2272
2.00	.3486	.3264	.3165	.3159	.3053

Table 6.7

t	$\bar{x}_4^{SPL}(t)$	$\bar{x}_8^{SPL}(t)$	$\bar{x}_{16}^{SPL}(t)$	$\bar{x}_{32}^{SPL}(t)$	$\bar{x}(t)$
0.0	– .0034	– .0010	– .0003	– .0001	0.0
.25	– .2538	– .2415	– .2425	– .2440	– .2473
.50	.1721	.1259	.1155	.1136	.1078
.75	.6048	.5994	.5749	.5737	.5663
1.00	.6607	.6234	.6257	.6278	.6186
1.25	.4115	.4222	.4222	.4231	.4127
1.50	.2005	.2599	.2665	.2629	.2474
1.75	.2398	.2617	.2494	.2456	.2272
2.0	.4021	.3485	.3359	.3282	.3053

Table 6.8

Example 6.4 (Daniel [10], Example 4.2)

In this example we consider an inertial control problem (see [9])

$$\text{Minimize} \quad \Phi(u) = \frac{1}{2} y(2)^2 + \frac{1}{2} \int_0^2 \dot{u}(s)^2 ds$$

over $u \in U = \{u \in H_1^1(0,2) : u(0) = 0\}$ subject to

$$\dot{y}(t) = y(t-1) + \frac{1}{2} t^2 \sin y(t) + u(t)$$

with initial conditions

$$y(0) = 1 \qquad y_0(s) = 1 \qquad -1 \le s \le 0.$$

Although this example is not in the form of problem (P) it can be transformed into an equivalent optimal control problem to which the theory developed above applies. If we let

$$x(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix} \qquad v(t) = \dot{u}(t)$$

then the problem becomes

$$\text{Minimize } \Phi(v) = x(2)^T \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 0 \end{bmatrix} x(2) + \frac{1}{2} \int_0^2 v(s)^2 ds$$

over $v \in L_2^1\ (0,2)$ subject to

$$\dot{x}(t) = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x(t) + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x(t-1) + \begin{bmatrix} \frac{1}{2} t^2 \sin x_1(t) \\ v(t) \end{bmatrix}$$

with initial conditions

$$x(0) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad x_0(s) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \qquad -1 \le s \le 0$$

The approximating optimal controls $\bar{u}_N = (\bar{x}_N)_2$ are given in Tables 6.9 and 6.10, and the corresponding optimal trajectories $\bar{y}_N = (\bar{x}_N)_1$ are given in Tables 6.11 and 6.12.

t	$\bar{u}_4^{AVE}(t)$	$\bar{u}_8^{AVE}(t)$	$\bar{u}_{16}^{AVE}(t)$	$\bar{u}_{32}^{AVE}(t)$	$\bar{u}(t)$
0.0	0.0	0.0	0.0	0.0	0.0
.25	- .6687	- .6836	- .6870	- .6880	- .6858
.50	-1.2308	-1.2329	-1.2332	-1.2336	-1.2291
.75	-1.6634	-1.6569	-1.6555	-1.6558	-1.6494
1.00	-1.9840	-1.9723	-1.9707	-1.9714	-1.9645
1.25	-2.2085	-2.1937	-2.1932	-2.1950	-2.1891
1.50	-2.3520	-2.3349	-2.3352	-2.3382	-2.3332
1.75	-2.4293	-2.4095	-2.4098	-2.4132	-2.4087
2.00	-2.4644	-2.4342	-2.4320	-2.4349	-2.4303
$\Phi_N(\bar{u}_N)$	2.5484	2.4804	2.4617	2.4570	

Table 6.9

t	$\bar{u}_4^{SPL}(t)$	$\bar{u}_8^{SPL}(t)$	$\bar{u}_{16}^{SPL}(t)$	$\bar{u}_{32}^{SPL}(t)$	$\bar{u}(t)$
0.0	0.0	0.0	0.0	0.0	0.0
.25	- .6491	- .6799	- .6860	- .6877	- .6858
.50	-1.2161	-1.2282	-1.2319	-1.2331	-1.2291
.75	-1.6506	-1.6524	-1.6539	-1.6553	-1.6494
1.00	-1.9727	-1.9692	-1.9699	-1.9716	-1.9645
1.25	-2.2070	-2.1957	-2.1950	-2.1966	-2.1891
1.50	-2.3639	-2.3436	-2.3404	-2.3411	-2.3332
1.75	-2.4483	-2.4207	-2.4157	-2.4167	-2.4087
2.00	-2.4816	-2.4447	-2.4382	-2.4388	-2.4303
$\Phi_N(\bar{u}_N)$	2.5826	2.4986	2.4719	2.4624	

Table 6.10

t	$\bar{x}_4^{AVE}(t)$	$\bar{x}_8^{AVE}(t)$	$\bar{x}_{16}^{AVE}(t)$	$\bar{x}_{32}^{AVE}(t)$	$\bar{x}(t)$
0.0	1.0000	1.0000	1.0000	1.0000	1.0000
.25	1.1698	1.1654	1.1638	1.1633	1.1636
.50	1.2019	1.1911	1.1881	1.1873	1.1889
.75	1.1454	1.1273	1.1213	1.1193	1.1228
1.00	1.0462	1.0208	1.0095	1.0034	1.0041
1.25	.9371	.9066	.8919	.8835	.8862
1.50	.8376	.8031	.7878	.7805	.7927
1.75	.7590	.7178	.7008	.6932	.7167
2.00	.7103	.6561	.6315	.6195	.6564

Table 6.11

t	$\bar{x}_4^{SPL}(t)$	$\bar{x}_8^{SPL}(t)$	$\bar{x}_{16}^{SPL}(t)$	$\bar{x}_{32}^{SPL}(t)$	$\bar{x}(t)$
0.0	1.0000	1.0000	1.0000	1.0000	1.0000
.25	1.1703	1.1662	1.1639	1.1633	1.1636
.50	1.2102	1.1932	1.1887	1.1875	1.1889
.75	1.1500	1.1279	1.1214	1.1195	1.1228
1.00	1.0418	1.0111	1.0017	.9984	1.0041
1.25	.9340	.8944	.8824	.8775	.8862
1.50	.8507	.8055	.7875	.7799	.7927
1.75	.7949	.7332	.7084	.6968	.7162
2.00	.7674	.6820	.6440	.6252	.6564

Table 6.12

Based upon the examples presented here, and several others which we have looked at, the following observations can be made.

(1) The schemes which we have proposed represent feasible and relatively efficient approximation methods for solving certain classes of non-linear hereditary control problems.

(2) Since the resulting approximating problems are governed by discrete difference equations the programming required is relatively simple. Moreover, since no additional discretization is necessary when the schemes are implemented on the computer, no further stability analysis is required in order to guarantee convergence of the approximating solutions.

(3) The spline based schemes, although somewhat more difficult to program and costlier to run, out-perform the averaging schemes. However, tne difference appears to be more pronounced in the case of simple integration of initial value problems as opposed to the solution of optimal control problems.

(4) The accuracy of the approximating optimal controls and trajectories is quite good even for relatively small values of N. This is especially true for the schemes employing the spline based state approximation.

(5) Our results are comparable to those obtained by Rockey [25] and to those obtained via the semi-discrete schemes developed by Banks [2] [5] and Daniel [10].

We have also applied our schemes to the design of an open loop controller for the mach number - guide vane angle control loop of the National Transonic Wind Tunnel Facility (NTF) at the NASA Langley Research Center in Hampton, Virginia (see [1],[10]). Although the operation of the NTF is best described by a complex system of nonlinear partial differential equations, the dynamics of the system near steady state operating conditions can be modeled by a linear hereditary system in which either the guide vane angle actuator, or the guide vane angle actuator rate act as a control. If we assume that a disturbance has occurred at time $t = 0$, the problem is to choose the control so as to drive the system back to equilibrium as quickly as possible without exceeding the physical limitations of the components of the system. This leads to a linear quadratic optimal control problem in which the dynamics are governed by a linear FDE of the form (2.1) with $f(t,\eta,\phi,u) = Bu$. While an approximation to the closed loop solution to this problem (in the form of approximating feedback gains matrices) would be more desirable (and is accessable through the techniques discussed in [7] [13] and [20]) we have generated approximating open loop solutions using the schemes developed above. This permitted us to test our methods on systems of higher dimension ($n = 3$ and 4) with the optimization being carried out over an extended time interval ($T = 30$). Both the averaging and spline based state approximations were employed with values of N as large as 24. We compared our results to the open loop solutions to this problem which appear in [10] and to the open loop form of the closed loop solutions computed in [1] and [7]. Our schemes performed comparably, both qualatatively and quantatatively, and provided acceptable approximating solutions for all values of $N \geq 4$.

References

[1] E. S. Armstrong and J. S. Tripp, An application of multivariable design techniques to the control of the National Transonic Facility, NASA Technical Paper 1887, Langley Research Center, Hampton, VA, August 1981.

[2] H. T. Banks, Approximation of nonlinear functional differential equation control systems, J. Optimization Theory Appl., 29 (1979) 383 - 408.

[3] H. T. Banks and J. A. Burns, An abstract framework for approximate solutions to optimal control problems governed by hereditary systems, Proc. International Conf. on Differential Equations, (Univ. So. Calif., Sept. 1974), H. A. Antosiewicz, ed., Academic Press, New York, 1975, 10 - 25.

[4] H. T. Banks and J. A. Burns, Hereditary control problems: numerical methods based on averaging approximations, SIAM J. Control and Optimization, 16 (1978), 169 - 208.

[5] H. T. Banks, J. A. Burns and E. M. Cliff, A comparison of numerical methods for identification and optimization problems involving control systems with delays, Brown Univ. LCDS Tech Rep. 79-7, Providence, RI, 1979.

[6] H. T. Banks, J. A. Burns, E. M. Cliff and P. R. Thrift, Numerical solutions of hereditary control problems via an approximation technique, Brown Univ. LCDS Technical Report 75-6, 1975, Providence, RI.

[7] H. T. Banks, K. Ito, and I. G. Rosen, A spline based technique for computing Riccati operators and feedback controls in regulator problems for delay equations, to appear.

[8] H. T. Banks and F. Kappel, Spline approximations for functional differential equations, J. Differential Equations, 34 (1979), 496 - 522.

[9] P. L. Daniel, Spline-based approximation methods for the identification and control of nonlinear functional differential equations, Ph.D. Thesis, Brown University, Providence, RI, June, 1981.

[10] P. L. Daniel, Spline approximations for nonlinear hereditary control systems, ICASE Report number 82-[illegible], Institute for Computer Applications in Science and Engineering, 198[illegible], Hampton, VA.

[11] R. L. Ehle, A-stable methods and Padé approximations to the exponential, SIAM J. Math. Anal., 4 (1973), [illegible] - 680.

[12] L. E. El'sgol'ts, Introduction to the Theory of Differential Equations with Deviating Arguments, Holden-Day, San Francisco, 1966.

[13] J. S. Gibson, Linear quadratic optimal control of hereditary differential systems: infinite dimensional Riccati equations and numerical approximations, preprint, Sept., 1980, submitted to SIAM J. Control Optimization.

[14] R. Hersh and T. Kato, High-accuracy stable difference schemes for well-posed initial-value problems, SIAM J. Numer. Anal., 16 (1979), 670 - 682.

[15] International Mathematical and Statistical Libraries, Library 1, Edition 6, (1977), ZXMIN-1-3.

[16] F. Kappel, An approximation scheme for delay equations, Proc. International Conf. on Nonlinear Phenomena in the Math Sciences, (Univ. Texas, Arlington, June 16-20, 1980), Academic Press, to appear.

[17] F. Kappel and W. Schappacher, Autonomous nonlinear functional differential equations and averaging approximations, Nonlinear Analysis: Theory, Methods and Appl., 2 (1978), 391 - 422.

[18] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, New York, 1966.

[19] W. Krabs, Optimization and Approximation, John Wiley and Sons, New York, 1979.

[20] K. Kunisch, Approximation schemes for the linear quadratic optimal control problem associated with delay equations, SIAM J. Control Optimization, to appear.

[21] D. C. Reber, A finite difference technique for solving optimization problems governed by linear functional differential equations, J. Differential Equations, 32 (1979), 192 - 232.

[22] F. Reisz and B. Sz.-Nagy, Functional Analysis, Ungar, New York, 1975.

[23] S. A. Rockey, Discrete methods of state approximation, parameter identification and optimal control for hereditary systems, Ph.D. thesis, Brown University, Providence, RI, June 1982.

[24] I. G. Rosen, A discrete approximation framework for hereditary systems, J. Differential Equations, 40 (1981), 377 - 499.

[25] L. Tavernini, One-step methods for the numerical solution of Volterra functional differential equations, SIAM J. Numer. Anal., 8 (1971), 786 - 795.

[26] R. S. Varga, Matrix Iterative Analysis, Prentice-Hall, Englewood Cliffs, NJ, 1962, Chapter 8.3.