JOURNAL OF DIFFERENTIAL EQUATIONS **68**, 22–35 (1987)

Optimal Control of Nonlinear Hereditary Systems*

DOUGLAS C. REBER

E-Systems, Inc., Garland Division, Dallas, Texas, 75266

Received May 1, 1986

A function space Λ is introduced for the study of nonlinear hereditary differential equations. The properties of Λ include: it is a Banach space under the supremum norm, the continuous functions constitute a closed proper subspace, and the unit ball is sequentially compact in the weak-* topology. Existence, uniqueness, and continuous dependence results are obtained for solutions of a broad class of initial value problems. An optimization problem is formulated for systems which are affine in the control, and solutions are approximated by means of a sequence of problems which are finite-dimensional in the control.

1. INTRODUCTION

Various state spaces have been employed successfully in the study of hereditary differential equations (HDE) with finite delay. The elements of these spaces are (in part, equivalence classes of) functions defined on a compact interval K. The reason for this choice is that a function φ defined on K generally determines a unique solution $x(\varphi)$ of a given HDE. The most common state spaces are $C(K)$, the continuous functions defined on K under the supremum norm, and $L_p(K) \times R$, where frequently $p = 2$. (In this report it is assumed that solutions are real-valued; the above comments and the results presented below are readily generalized to higher dimensions.)

The space $C(K)$ is very convenient for the study of HDE if the initial state φ is itself an element of $C(K)$ and the control function (if any) is continuous. The advantage of using $C(K)$ rests partially on the fact that point evaluation is continuous (i.e., the map $\varphi \to \varphi(\theta)$ from $C(K)$ to R is continuous for all θ in K); this permits the treatment of a broad class of HDE. However, it is sometimes necessary to allow for discontinuous initial states, and it is almost always necessary to consider discontinuous control

* This research was conducted while the author was a visiting faculty member at the Department of Mathematical Sciences, University of Delaware, Newark, Delaware 19711 and at the Department of Mathematics, Calvin College, Grand Rapids, Michigan 49506.

0022-0396/87 $3.00

functions when optimizing the response of a system. Moreover, the HDE is equivalent to an ordinary differential equation involving $C(K)$-valued functions only under severe restrictions on the initial data. (Further comments on this topic will be made in Sect. 5.)

In contrast, the spaces $L_p(K) \times R$ allow for a broader range of initial data and require only that a control function be an element of L_p. This permits a satisfactory treatment of optimization problems. Unfortunately, the class of HDE which may be considered is greatly restricted because point evaluation is not continuous (or even well defined). This drawback may be overcome in many linear and in some nonlinear problems, but most nonlinear HDE are not amenable to investigation in such state spaces.

To avoid these disadvantages of $C(K)$ and $L_p(K) \times R$, a state space $\Lambda(K)$ is introduced in Section 2. $\Lambda(K)$ is isometrically isomorphic to $L_\infty(K) \times R$, and is therefore isometrically isomorphic to the dual of the space $AC(K)$ of absolutely continuous functions on K under the norm $|\varphi| = |\varphi(d)| + \text{var}(\varphi, K)$. (The expression $\text{var}(\varphi, K)$ denotes the total variation of φ on K; d denotes the right endpoint of K.) Furthermore, $C(K)$ is a closed proper subspace of $\Lambda(K)$, and point evaluation of elements of $\Lambda(K)$ is continuous.

A class of HDE which may be successfully treated in the state space $\Lambda(K)$ is described in Section 3. Existence and uniqueness of solutions of initial value problems, as well as continuous dependence with respect to initial data and control functions, are then established for such systems.

An optimization problem is stated in Section 4 for systems governed by HDE which are affine in the control function. Solutions are shown to exist and sufficient conditions for uniqueness are given. Finally, solutions of a sequence of approximate, finite-dimensional optimization problems are shown to converge to a solution of the original problem.

The results mentioned above are compared in Section 5 with analogous results in the state spaces $C(K)$ and $L_p(K) \times R$.

Certain notational conventions are employed throughout:

- N^+ — positive integers;
- R^n — for any $n \in N^+$, n-dimensinal euclidean space ($R = R^1$);
- $B(E)$ — for any interval E, the Banach space of Lebesgue measurable, bounded functions under the supremum norm;
- $\Sigma(E)$ — for any interval E, the σ-algebra of Lebesgue measurable subsets of E;
- χ_E — characteristic function of $E \subset R$;
- μ — Lebesgue measure.

Let $r > 0$ be given. Define $I_r = [-r, 0]$, $S = [a, b]$ and $S_r = [a-r, b]$. For a function $x \in B(S_r)$ and for $t \in S$, define $x_t \in B(I_r)$ by $x_t(\theta) = x(t+\theta)$. If $x \in B(E)$, let $[x]$ denote the element of $L_p(E)$ which contains x (where the value of p is to be taken from the context).

2. Description and Properties of the State Space

A Banach space Λ of real-valued functions defined on the interval I_r will serve as state space for the HDE to be considered. Since the properties of Λ are also appropriate for control functions, the material in this section will be developed for functions defined on a compact interval $K=[c, d]$; the corresponding Banach space will be denoted $\Lambda(K)$. The state and control spaces will be $\Lambda(I_r)$ and $\Lambda(S_r)$, respectively.

Let $K_0=[c, d)$. The definition of $\Lambda(K)$ rests upon the existence of a map $\beta\colon L_\infty(K)\to B(K_0)$ which is an isometric isomorphism onto its range. Once the existence of such a map has been established, $\Lambda(K)$ may be considered as the class of all functions from K to R whose restriction to K_0 is an element of $\beta(L_\infty(K))$, under the supremum norm.

The presentation given below in terms of the function β is based upon an idea of Dieudonné [6]. Throughout this section, Φ will denote the Fréchet filter on N^+ and Ψ will denote a fixed nonprincipal ultrafilter on N^+.

To begin, define:

$$J\colon K_0\times N^+\to 2^K \qquad \text{by } J(t, n)=[t, t+1/n]\cap K,$$

$$\alpha\colon L_\infty(K)\times K_0\times N^+\to R \qquad \text{by } \alpha(\xi, t, n)=\left[\int_{J(t,n)}\xi\,d\mu\right]\Big/\mu(J(t, n)).$$

For all $t\in K_0$ and $n\in N^+$ the measure of $J(t, n)$ is positive. For all $\xi\in L_\infty(K)$ and $n\in N^+$ the maps $t\to\int_{J(t,n)}\xi\,d\mu$ and $t\to\mu(J(t, n))$ are continuous on K_0. Therefore $\alpha(\xi, \cdot, n)$ is continuous.

Observe that $|\alpha(\xi, t, n)|\leqslant|\xi|$ for all ξ, t, n. Consequently the sets $\alpha(\xi, t, A)$, $A\in\Psi$, form the base of an ultrafilter of subsets of the compact interval $[-|\xi|, |\xi|]$. It follows that for all ξ and t, $\lim_\Psi\alpha(\xi, t, \cdot)$ exists in $[-|\xi|, |\xi|]$.

Suppose $\xi\in L_\infty(K)$ and $x\in\xi$. Let Δ denote the set of Lebesgue points of x. For all $t\in\Delta\cap K_0$, $\lim_\Phi\alpha(\xi, t, \cdot)$ exists and equals $x(t)$, which implies that $\lim_\Psi\alpha(\xi, t, \cdot)=x(t)$. Since $\mu(\Delta)=\mu(K)$ the function $t\to\lim_\Psi\alpha(\xi, t, \cdot)$ is measurable. Define $\beta\colon L_\infty(K)\to B(K_0)$ by $\beta(\xi)(t)=\lim_\Psi\alpha(\xi, t, \cdot)$. Then $|\xi|\leqslant|\beta(\xi)|$ because $\beta(\xi)\in\xi$. Previous comments imply that $|\beta(\xi)|\leqslant|\xi|$; consequently $|\beta(\xi)|=|\xi|$. The linearity of β follows directly from the linearity for all n of $\xi\to\alpha(\xi, \cdot, n)$ and the appropriate properties of limits with respect to a filter. Let $X=\beta(L_\infty(K))$. Then since $\beta([x])=x$ for all $x\in X$ and $[\beta(\xi)]=\xi$ for all $\xi\in L_\infty(K)$, β is an isometric isomorphism between $L_\infty(K)$ and X.

Define $\gamma\colon L_\infty(K)\times R\times K\times N^+\to R$ by

$$\gamma(\xi, \xi_0, t, n)=\begin{cases}\alpha(\xi, t, n), & t\in[c, d_n]\\ \alpha(\xi, d_n, n)+[\xi_0-\alpha(\xi, d_n, n)](t-d_n)/(d-d_n), & t\in(d_n, d],\end{cases}$$

where $d_n = d - (d-c)/(n+1)$. Then $\gamma(\xi, \xi_0, \cdot, n) \in C(K)$ for all ξ, ξ_0, n. Now define ρ: $L_\infty(K) \times R \to B(K)$ by $\rho(\xi, \xi_0)(t) = \lim_{\Psi} \gamma(\xi, \xi_0, t, \cdot)$. Clearly $\rho(\xi, \xi_0) = \beta(\xi)$ on K_0 for all ξ, and $\rho(\xi, \xi_0)(d) = \xi_0$. Thus if the norm on $L_\infty(K) \times R$ is given by $|(\xi, \xi_0)| = \max\{|\xi|, |\xi_0|\}$, then ρ: $L_\infty(K) \times R \to B(K)$ is an isometric isomorphism onto its range. Let $\Lambda(K)$ be defined as $\rho(L_\infty(K), R)$, under the supremum norm. These results are summarized in the following theorem.

THEOREM 1. *$\Lambda(K)$ is a Banach space; it is isometrically isomorphic to $L_\infty(K) \times R$ and is consequently isometrically isomorphic to* $\mathrm{AC}(K)^*$.

Observe that $\Lambda(K)$ is defined by means of functions which are continuous on K, not just on K_0. This permits the statement of the following result, which will be useful in establishing the measurability of certain functions in Section 3.

THEOREM 2. *For every $x \in \Lambda(K)$ there is a sequence $\{x_n\}$ in $C(K)$ which converges to x pointwise a.e. on K and is such that $|x_n| \leqslant |x|$ for all n.*

Proof. Observe that $\gamma([x], x(d), \cdot, n) \in C(K)$ for all n and that $\lim_{\Phi} \gamma([x], x(d), t, \cdot) = x(t)$ at all Lebesgue points t of x. Furthermore, $|\gamma([x], x(d), t, n)| \leqslant \max\{|[x]|, |x(d)|\} \leqslant |x|$ for all t, n.

In this paper, Dieudonné [6] employed integral averages over the intervals $[t - 1/n, t + 1/n]$ instead of $[t, t + 1/n] \cap K$ (as above). The intersection with K accounts for the shortening of $[t, t + 1/n]$ which is necessary to stay within the domain of ξ when $t + 1/n > d$; this is merely a technical modification.

More important, however, is the fact that the integral averages do not extend over $[t - 1/n, t)$. The significance of this difference is most easily seen by an example. Let x: $[0, 1] \to R$ be the characteristic function of $\{1\}$ and let y: $[0, 2] \to R$ be the characteristic function of $[1, 2]$. Note that $x \in \Lambda([0, 1])$. If integral averages had been taken over $[t - 1/n, t + 1/n]$ in the definition of $\Lambda([c, d])$, then y would not be an element of $\Lambda([0, 2])$ because $y(1) = 1 \neq \frac{1}{2}$. It would thus be possible to continuously extend an element in $\Lambda([0, 1])$ to $[0, 2]$ without obtaining an element of $\Lambda([0, 2])$.

Solutions of initial value problems involve the continuous extension of an element of $\Lambda([a - r, a])$ to the interval $[a - r, b]$. Thus if solutions are to be elements of $\Lambda([a - r, b])$, the integral averages cannot be taken over $[t - 1/n, t + 1/n]$. The following lemma indicates that the desired property does hold for Λ as defined above; its proof is immediate.

LEMMA 1. *Let $x \in \Lambda([a - r, a])$ and let y: $[a - r, b] \to R$ be an extension of x which is continuous on $[a, b]$. Then $y \in \Lambda([a - r, b])$, and $y_t \in \Lambda([-r, 0])$ for all $t \in [a, b]$.*

The next property of $\Lambda(K)$ to be considered provides the basis for relating optimization over finite-dimensional subspaces of $\Lambda(K)$ to optimization over $\Lambda(K)$ itself. The result is given below in Theorem 4.

The proof of Theorem 3, which is an intermediate result, includes an argument (different from the one given above) that $\lim_\Phi \alpha([x], t, \cdot)$ exists and equals $x(t)$ a.e. in K. The following discussion is modeled on Rudin's [8] treatment of differentiation of a Borel measure.

For $t \in K_0$ and $n \in N^+$ define

$$\mathscr{J}(t, n) = \{[p, q]: c \leqslant p \leqslant t < q \leqslant d \quad \text{and} \quad q - p \leqslant 1/n\},$$

and let $\omega^+, \omega^-: L_\infty(K) \times K_0 \times N^+ \to R$ be given by

$$\omega^+(\xi, t, n) = \sup\left\{\left[\int_E \xi d\mu\right] \Big/ \mu(E): E \in \mathscr{J}(t, n)\right\},$$

$$\omega^-(\xi, t, n) = \inf\left\{\left[\int_E \xi d\mu\right] \Big/ \mu(E): E \in \mathscr{J}(t, n)\right\}.$$

Note that $J(t, n) \in \mathscr{J}(t, n)$ so that $\omega^-(\xi, t, n) \leqslant \alpha(\xi, t, n) \leqslant \omega^+(\xi, t, n)$ for all ξ, t, n. Note also that $\omega^-(\xi, t, n) \leqslant \omega^-(\xi, t, n+1)$ and $\omega^+(\xi, t, n) \geqslant \omega^+(\xi, t, n+1)$ for all ξ, t, n. Let

$$\sigma^+(\xi, t) = \lim_\Phi \omega^+(\xi, t, \cdot),$$

$$\sigma^-(\xi, t) = \lim_\Phi \omega^-(\xi, t, \cdot).$$

Observe that for all ξ, t,

$$\sigma^-(\xi, t) \leqslant \liminf_\Phi \alpha(\xi, t, \cdot) \leqslant \beta(\xi)(t) \leqslant \limsup_\Phi \alpha(\xi, t, \cdot) \leqslant \sigma^+(\xi, t),$$

since Ψ finer than Φ implies

$$\begin{aligned}\liminf_\Phi \alpha(\xi, t, \cdot) &\leqslant \liminf_\Psi \alpha(\xi, t, \cdot) = \beta(\xi)(t)\\ &= \limsup_\Psi \alpha(\xi, t, \cdot) \leqslant \limsup_\Phi \alpha(\xi, t, \cdot).\end{aligned}$$

The proof of Theorem 3 appears immediately after Lemmas 2 and 3.

THEOREM 3. *For all* $\xi \in L_\infty(K)$, $\mu\{t \in K_0: \sigma^-(\xi, t) < \sigma^+(\xi, t)\} = 0$.

LEMMA 2. *For all* ξ, n *both* $\omega^+(\xi, \cdot, n)$ *and* $\omega^-(\xi, \cdot, n)$ *are measurable.*

Proof. Given $k \in R$, let $E_k^+ = \{t \in K_0: \omega^+(\xi, t, n) > k\}$. If $t \in E_k^+$ then there is an interval $[p, q] \in \mathscr{J}(t, n)$ with $\int_p^q \xi d\mu > k(q - p)$. Since $[p, q] \in$

$\mathscr{J}(s, n)$ for $t \leqslant s < q$, it follows that $[t, q) \subset E_k^+$. Thus $E_k^+ \in \Sigma(K_0)$. A similar argument shows that $\omega^-(\xi, \cdot, n)$ is measurable.

LEMMA 3. *Suppose $x \in B(K)$ and $\mu(\{t \in K: x(t) < 0\}) = 0$. Let $\xi = [x]$. If $A \in \Sigma(K)$ and $\int_A \xi d\mu = 0$ then $\mu(\{t \in A: \sigma^+(\xi, t) > 0\}) = 0$.*

Proof. Let $P = \{t \in K_0: \sigma^+(\xi, t) > 0\}$. If $\mu(A \cap P) > 0$ then there is a positive number k and a compact set $F \subset A \cap P$ such that $\mu(F) > 0$ and $\sigma^+(\xi, t) > k$ for all $t \in F$. For each $t \in F$ and $n \in N^+$ there is a set $E \in \mathscr{J}(t, n)$ such that $\int_E \xi d\mu > k\mu(E)$; let $\mathscr{E}_m$ denote the class of all such sets for $t \in F$ and $n \geqslant m$. Each class $\mathscr{E}_m$ is a Vitali covering of F. Hence there is a finite subclass $\mathscr{F}_m$ of disjoint sets for which $\mu(F \backslash \bigcup_{E \in \mathscr{F}_m} E) < \mu(F)/2$. Define $F_m = \{t \in K: \text{dist}(t, F) \leqslant 1/m\}$. Then since $\xi \geqslant 0$ a.e. and $E \subset F_m$ for all $E \in \mathscr{F}_m$,

$$\int_{F_m} \xi d\mu \geqslant \sum_{E \in \mathscr{F}_m} \int_E \xi d\mu > \sum_{E \in \mathscr{F}_m} k\mu(E) \geqslant k\mu(F)/2.$$

Since $F = \bigcap_{m \geqslant 1} F_m$, $\int_F \xi d\mu \geqslant k\mu(F)/2 > 0$. This contradicts the assumption $\int_A \xi d\mu = 0$. Consequently $\mu(A \cap P) = 0$.

Proof of Theorem 3. Choose $x \in B(K)$ with $[x] = \xi$. First, it will be shown that $\mu(\{t \in K_0$: $x(t) < \sigma^+(\xi, t)\}) = 0$. Let q be a rational number. Define $E = \{t \in K: x(t) < q\}$ and $F = K \backslash E$. Let $y \in B(K)$ be given by $y(t) = [x(t) - q]\, \chi_F(t)$ and let $\eta = [y]$. Note that y is nonnegative on K, and that $\int_G \xi d\mu - q\mu(G) \leqslant \int_G \eta d\mu$ for all $G \in \Sigma(K)$. This implies that $\sigma^+(\xi, t) - q \leqslant \sigma^+(\eta, t)$ for all t. Since $\int_E \eta d\mu = 0$, Lemma 3 implies that $\mu(\{t \in E$: $\sigma^+(\eta, t) > 0\}) = 0$. Let $A_q = \{t \in K_0$: $x(t) < q < \sigma^+(\xi, t)\}$. It follows from the above comments that $\mu(\{t \in E$: $\sigma^+(\xi, t) > q\}) = 0$, i.e., $\mu(A_q) = 0$. If $A = \{t \in K_0$: $x(t) < \sigma^+(\xi, t)\}$ then $A = \bigcup_q A_q$, so $\mu(A) = 0$.

Observe that for all $t \in K_0$, $\sigma^-(\xi, t) = -\sigma^+(-\xi, t)$. The above argument implies that

$$\begin{aligned}\mu(\{t \in K_0: \sigma^-(\xi, t) < x(t)\}) &= \mu(\{t \in K_0: -\sigma^+(-\xi, t) < x(t)\}) \\ &= \mu(\{t \in K_0: -x(t) < \sigma^+(-\xi, t)\}) \\ &= 0.\end{aligned}$$

The conclusion follows immediately.

A sequence of projection operators, each having finite-dimensional range, will now be defined on $A(K)$. Theorem 3 will be used to establish a convergence property of this sequence.

For $n \in N^+$ and $i = 1, 2, \ldots, n$, define $K(n, i) \subset K$ by

$$K(n, i) = [c + (i-1)(d-c)/n,\ c + i(d-c)/n).$$

Let $\pi_n: \Lambda(K) \to B(K)$ be given by

$$(\pi_n x)(t) = \begin{cases} n/(d-c) \int_{K(n,i)} x\,d\mu & \text{for} \quad t \in K(n, i) \\ x(d) & \text{for} \quad t = d. \end{cases}$$

Observe that $\pi_n x$ is right-continuous and bounded on K_0, which implies that $\pi_n x \in \Lambda(K)$.

THEOREM 4. *For every $x \in \Lambda(K)$ the sequence $\pi_n x$ converges to x pointwise a.e. on K and satisfies: $|\pi_n x| \leqslant |x|$ for all n.*

Proof. Let $\xi = [x]$ and define $E = \{t \in K_0: \sigma^-(\xi, t) = \sigma^+(\xi, t)\}$. Then $\mu(E) = \mu(K)$ by Theorem 3. Define $\tau: K \times N^+ \to N^+$ by $\tau(t, n) = \max\{i: c + (i-1)(d-c)/n \leqslant t\}$; observe that $t \in K(n, \tau(t, n))$ for all t, n. Let $t \in E$ be fixed and let $\varepsilon > 0$ be given. Choose $n_\varepsilon \in N^+$ such that $\omega^+(\xi, t, n_\varepsilon) - x(t) = \omega^+(\xi, t, n_\varepsilon) - \sigma^+(\xi, t) < \varepsilon$, and $x(t) - \omega^-(\xi, t, n_\varepsilon) = \sigma^-(\xi, t) - \omega^-(\xi, t, n_\varepsilon) < \varepsilon$. Choose $m_\varepsilon \geqslant n_\varepsilon$ so that $(d-c)/m_\varepsilon < 1/n_\varepsilon$. Then for $m \geqslant m_\varepsilon$, the closure of $K(m, \tau(t, m))$ is in $\mathscr{I}(t, n_\varepsilon)$. Consequently $\omega^-(\xi, t, n_\varepsilon) \leqslant (\pi_m x)(t) \leqslant \omega^+(\xi, t, n_\varepsilon)$ for all $m \geqslant m_\varepsilon$. Therefore

$$|(\pi_m x)(t) - x(t)| < \varepsilon \qquad \text{for} \quad m \geqslant m_\varepsilon.$$

The inequalities $|\pi_n x| \leqslant |x|$ are obvious.

It should perhaps be noted that $\Lambda(K)$ depends upon the ultrafilter Ψ in the following sense: if $\xi \in L_\infty(K)$, Γ is a nonprincipal ultrafilter on N^+ different from Ψ, and $\liminf_\Phi \alpha(\xi, t, \cdot) < \limsup_\Phi \alpha(\xi, t, \cdot)$ then it may happen that $\lim_\Psi \alpha(\xi, t, \cdot) \neq \lim_\Gamma \alpha(\xi, t, \cdot)$.

The relationship between $\Lambda(K)$ and $\mathrm{AC}(K)^*$ may be sketched as follows. Let $\lambda \in \mathrm{AC}(K)^*$ be given. For $E \in \Sigma(K)$ define $h_E: K \to R$ by $h_E(t) = -\int_t^d \chi_E\,d\mu$; $h_E \in \mathrm{AC}(K)$ for all E. Define the set function ν_λ on $\Sigma(K)$ by $\nu_\lambda(E) = \lambda(h_E)$; ν_λ is countably additive and $\nu_\lambda \ll \mu$. Let $\xi_\lambda \in L_1(K)$ be the Radon–Nikodym derivative of ν_λ; ξ_λ is in fact an element of $L_\infty(K)$ since $|\nu_\lambda|(E) \leqslant |\lambda|\, \mu(E)$ for all E. Let $u \in \mathrm{AC}(K)$ be given by $u(t) = 1$ for all t. Finally, let $x_\lambda = \rho(\xi_\lambda, \lambda(u))$. Clearly, $|x_\lambda| \leqslant |\lambda|$. Moreover, $\lambda(h) = x_\lambda(d)\, h(d) + \int_K x_\lambda\, dh$ for all $h \in \mathrm{AC}(K)$. Since this equation implies that $|\lambda| \leqslant |x_\lambda|$, it follows that $|x_\lambda| = |\lambda|$. Conversely, for every $x \in \Lambda(K)$ the above equation defines an element λ_x of $\mathrm{AC}(K)^*$, and x is the element of $\Lambda(K)$ associated as above with λ_x.

The separability of $\mathrm{AC}(K)$ implies that the unit ball is sequentially compact in the weak-* topology on $\Lambda(K)$. This fact will be instrumental in Section 4 in the proof of the existence of optimal controls.

3. Description and Characteristics of HDE

Consider the initial value problem (IVP): for $\varphi \in \Lambda(I_r)$ and $u \in \Lambda(S_r)$ find $x \in \Lambda(S_r)$ such that

$$x_a = \varphi,\ x(t) = \varphi(0) + \int_a^t f(s, x_s, u_s)\, ds \qquad \text{for} \quad t \in S.$$

It will be shown below that IVP has a unique solution whenever $f\colon S \times \Lambda(I_r) \times \Lambda(I_r) \to R$ satisfies

(a) $s \to f(s, x_s, u_s)$ is measurable for all x, $u \in \Lambda(S_r)$,

(b) $\operatorname{ess\,sup}\{|f(s, 0, 0)| : s \in S\} < \infty$,

(c) $u^n \to u$ (weak-*) implies that

$$\int_a^t f(s, x_s, u_s^n)\, ds \to \int_a^t f(s, x_s, u_s)\, ds$$

for all t, x,

(d) for all $c > 0$ there is a number $M(c) > 0$ such that

$$|\varphi|, |\psi|, |\zeta|, |\eta| \leqslant c \qquad \text{implies that for a.a. } s,$$
$$|f(s, \varphi, \zeta) - f(s, \psi, \eta)| \leqslant M(c)\{|\varphi - \psi| + |\zeta - \eta|\}.$$

This property of IVP, together with some continuous dependence results, is established in

Theorem 1. *The above* IVP *has a unique solution* $x(\varphi, u)$ *for every* $\varphi \in \Lambda(I_r)$ *and every* $u \in \Lambda(S_r)$. *Furthermore, if* $\varphi^n \to \varphi$ *(norm) and* $u^n \to u$ *(weak-*) then* $x(\varphi^n, u^n) \to x(\varphi, u)$ *(norm).*

Proof. Let $k = \sup\{|\varphi^n| + |u^n| : n \in N^+\}$ and $K = \operatorname{ess\,sup}\{|f(s, 0, 0)| : s \in S\}$; both k and K are finite. Let $c = 2k + 1$. The proof will proceed under the assumption that $[cM(c) + K](b - a) \leqslant \frac{1}{2}$. Afterward, the general case $b > a$ will be considered.

Define $e\colon \Lambda(I_r) \to \Lambda(S_r)$ by $(e\psi)(s) = \psi(s - a)$ for $s \in [a - r, a)$ and $(e\psi)(s) = \psi(0)$ for $s \in S$. Define $T\colon \Lambda(S_r) \times \Lambda(I_r) \times \Lambda(S_r) \to \Lambda(S_r)$ by

$$T(y, \psi, v)(t) = \begin{cases} 0 & t \in [a - r, a) \\ \int_a^t f(s, y_s + (e\psi)_s, v_s)\, ds, & t \in [a, b]. \end{cases}$$

Suppose $|y| \leqslant 1$ and $|\psi|$, $|v| \leqslant k$. Then for all s, $|y_s + (e\psi)_s|$, $|v_s| \leqslant c$. Hence for all t,

$$|T(y, \psi, v)(t)| \leqslant \int_a^t |f(s, y_s + (e\psi)_s, v_s) - f(s, 0, 0)|\, ds + \int_a^t |f(s, 0, 0)|\, ds$$
$$\leqslant [cM(c) + K](t-a)$$
$$\leqslant \tfrac{1}{2}.$$

That is, $T(\cdot, \psi, v)$ maps $A = \{y \in A(S_r): |y| \leqslant 1\}$ into itself for all ψ, v satisfying $|\psi|$, $|v| \leqslant k$. Similarly, for $|y|$, $|z| \leqslant 1$ and $|\psi|$, $|v| \leqslant k$, $|T(y, \psi, v)(t) - T(z, \psi, v)(t)| \leqslant \frac{1}{2}|y - z|$ for all t. Therefore $T(\cdot, \psi, v)$ is a contraction on A uniformly with respect to $|\psi|$, $|v| \leqslant k$; let $y(\psi, v) \in A$ denote its fixed point.

For $|y| \leqslant 1$, $T(y, \varphi^n, v) \to T(y, \varphi, v)$ uniformly with respect to $|v| \leqslant k$. By property (c) of f, $T(y, \psi, u^n)(t) \to T(y, \psi, u)(t)$ for all ψ, t. Since $\{T(y, \psi, u^n): n \in N^+\}$ is bounded and uniformly equicontinuous for $|\psi| \leqslant k$, it follows that $T(y, \psi, u^n) \to T(y, \psi, u)$. Consequently $T(y, \varphi^n, u^n) \to T(y, \varphi, u)$.

Let z_n: $S_r \to R$ be given by $z_n(t) = \sup\{|y(\varphi^n, u^n)(s) - y(\varphi, u)(s)|: a - r \leqslant s \leqslant t\}$. By writing $T(y(\varphi^n, u^n), \varphi^n, u^n) - T(y(\varphi, u), \varphi, u)$ as $[T(y(\varphi^n, u^n), \varphi^n, u^n) - T(y(\varphi, u), \varphi^n, u^n)] + [T(y(\varphi, u), \varphi^n, u^n) - T(y(\varphi, u), \varphi, u)]$ and applying Gronwall's inequality to z_n, it follows that $z_n \to 0$ in $A(S_r)$. Therefore $y(\varphi^n, u^n) \to y(\varphi, u)$. The desired solution of the IVP is given by $x(\varphi, u) = y(\varphi, u) + e\varphi$.

In general, if $b > a$ choose $m \in N^+$ such that $[cM(c) + K](b - a) \leqslant m/2$. Then divide S into m subintervals of equal length. The above procedure may be applied to each of these subintervals and the solution $x(\varphi, u)$ pieced together.

The utility of the above theorem depends on the types of functions which satisfy the conditions (a)–(d). Lemmas 1 and 2 below specify a readily identifiable class of such functions.

Let A be a finite subset of I_r. Define Γ as the Banach space of all bounded measures v on $\Sigma(I_r)$ such that $v = v_1 + v_2$, where $v_1 \ll \mu$ and $\mathrm{var}(v_2, I_r \setminus A) = 0$, under the total variation norm.

LEMMA 1. *Suppose γ: $S \to \Gamma$ is strongly measurable and essentially bounded, and let $u^n \to u$ (weak-*) in $A(S_r)$. Then for all measurable and essentially bounded functions y: $S \to R$,*

$$\int_a^t y(s) \left(\int_{-r}^0 u_s^n \, d\gamma(s) \right) ds \to \int_a^t y(s) \left(\int_{-r}^0 u_s \, d\gamma(s) \right) ds$$

for all t.

Proof. Let $\gamma(s) = \gamma_1(s) + \gamma_2(s)$, where $\gamma_1(s) \ll \mu$ and $\mathrm{var}(\gamma_2(s), I_r \setminus A) = 0$ for all s. The weak-* convergence of u^n to u implies that $\int_{-r}^0 u_s^n \, d\gamma_1(s) \to$

$\int_{-r}^{0} u_s \, d\gamma_1(s)$ for all s. Since $\sup\{|u^n|: n \in N^+\}$ and $\operatorname{ess\,sup}\{|\gamma_1(s)|: s \in S\}$ are both finite, $\int_a^t y(s)(\int_{-r}^{0} u_s^n \, d\gamma_1(s)) \, ds \to \int_a^t y(s)(\int_{-r}^{0} u_s d\gamma_1(s)) \, ds$ by the dominated convergence theorem.

Let $\theta \in A$. The function $\alpha_\theta(s) = \int_{-r}^{0} \chi_{\{\theta\}} \, d\gamma_2(s)$ is measurable and essentially bounded, and $\sum_{\theta \in A} \alpha_\theta(s) \, \varphi(\theta) = \int_{-r}^{0} \varphi d\gamma_2(s)$ for all φ, s. Thus $\int_a^t y(s)$ $(\int_{-r}^{0} u_s^n \, d\gamma_2(s)) \, ds = \sum_{\theta \in A} \int_a^t y(s) \, \alpha_\theta(s) \, u^n(s+\theta) \, ds$. The weak-* convergence of u^n to u and the finiteness of A imply that $\int_a^t y(s)(\int_{-r}^{0} u_s^n \, d\gamma_2(s)) \, ds \to$ $\int_a^t y(s)(\int_{-r}^{0} u_s \, d\gamma_2(s)) \, ds$ for all t.

LEMMA 2. *Suppose* $f\colon S \times A(I_r) \times A(I_r) \to R$ *is of the form* $f(s, \varphi, \zeta) = f_1(s, \varphi) + f_2(s, \varphi) \, g(s, \zeta)$ *where*

(i) *for all* $c > 0$ *there are numbers* $M_j(c) > 0$ *such that* $|\varphi|$, $|\psi| \leqslant c$ *implies that for a.a.* s *and for* $j = 1, 2$,

$$|f_j(s, \varphi) - f_j(s, \psi)| \leqslant M_j(c) \left[\max\{|\varphi(\theta) - \psi(\theta)|: \theta \in A\} \right.$$
$$\left. + \int_{-r}^{0} |\varphi - \psi| \, d\mu \right],$$

(ii) $f_j(\cdot, \varphi)$ *is measurable for every* φ, *and* $\operatorname{ess\,sup}\{|f_j(s, 0)|: s \in S\} < \infty$ $(j = 1, 2)$,

(iii) *there is a strongly measurable and essentially bounded function* $\gamma\colon S \to \Gamma$ *such that* $g(s, \zeta) = \int_{-r}^{0} \zeta \, d\gamma(s)$.

Then f *satisfies properties* (a)–(d).

Proof. (a) Suppose x, $u \in A(S_r)$. Let x^n be a sequence in $C(S_r)$ such that $|x^n| \leqslant |x|$ for all n and $x^n \to x$ pointwise a.e. (The existence of such a sequence is established by Theorem 2 of Sect. 2.) Then $s \to f_j(s, x_s^n)$ is measurable for $j = 1, 2$. Condition (i) implies that $f_j(s, x_s^n) \to f_j(s, x_s)$ pointwise a.e. on S. Consequently $s \to f_j(s, x_s)$ is measurable for $j = 1, 2$. It follows from the proof of Lemma 1 that $s \to g(s, u_s)$ is measurable. Therefore $s \to f(s, x_s, u_s)$ is measurable.

(b) $f(s, 0, 0) = f_1(s, 0)$, so $\operatorname{ess\,sup}\{|f(s, 0, 0)|: s \in S\} < \infty$ by condition (ii).

(c) Let $x \in A(S_r)$ be given. Then $s \to f_2(s, x_s)$ is measurable and essentially bounded. If $u^n \to u$ (weak-*) then Lemma 1 implies that $\int_a^t f_2(s, x_s) \, g(s, u_s^n) \, ds \to \int_a^t f_2(s, x_s) \, g(s, u_s) \, ds$ for all t.

(d) Let $K = \operatorname{ess\,sup}\{\max\{|f_1(s, 0)|, |f_2(s, 0)|, |\gamma(s)|\}: s \in S\}$. For $c > 0$ let $M(c) = (1 + r) \, M_1(c) + (1 + r) \, cKM_2(c) + K^2$. If φ, ψ, ζ, $\eta \in A(I_r)$ with $|\varphi|$, $|\psi|$, $|\zeta|$, $|\eta| \leqslant c$ then for a.a. $s \in S$,

$$|f_1(s, \varphi) - f_1(s, \psi)| \leqslant (1 + r) \, M_1(c) |\varphi - \psi|,$$

$$\begin{aligned}
&|f_2(s,\varphi)\,g(s,\zeta)-f_2(s,\psi)\,g(s,\eta)| \\
&\quad \leqslant |f_2(s,\varphi)-f_2(s,\psi)|\,|g(s,\zeta)|+|f_2(s,\psi)|\,|g(s,\zeta)-g(s,\eta)| \\
&\quad \leqslant (1+r)\,M_2(c)|\varphi-\psi|\;K|\zeta|+[|f_2(s,\psi)-f_2(s,0)|+|f_2(s,0)|]\;K|\zeta-\eta| \\
&\quad \leqslant (1+r)\,cKM_2(c)|\varphi-\psi|+[(1+r)\,M_2(c)|\psi|+K]\;K|\zeta-\eta| \\
&\quad \leqslant (1+r)\,cKM_2(c)|\varphi-\psi|+[(1+r)\,cKM_2(c)+K^2]|\zeta-\eta|.
\end{aligned}$$

Hence $|f(s,\varphi,\zeta)-f(s,\psi,\eta)|\leqslant M(c)[|\varphi-\psi|+|\zeta-\eta|]$.

It should be noted that condition (i) was necessary to establish property (a) as well as property (d).

4. Optimization

Having obtained continuous dependence results for solutions of HDE, it is now possible to consider an associated optimization problem. In particular, let $g\colon A(S_r)\to R$ and $h\colon L_2(S_r)\to R$ be given. Throughout this section, $[\cdot]$ will be considered as a map from $A(S_r)$ to $L_2(S_r)$. Define $Q\colon A(I_r)\times A(S_r)\to R$ by $Q(\varphi,u)=g(x(\varphi,u))+h([u])$. The basic optimization problem is

$$\mathscr{P}\text{: given } \varphi\in A(I_r) \text{ and } U\subset A(S_r), \text{ minimize } Q(\varphi,\cdot) \text{ over } U.$$

If the right-hand side f of HDE satisfies conditions (i)–(iii) of Lemma 2 in Section 3, then the optimization problem $\mathscr{P}$ is tractable under the hypotheses:

(a) g, h are continuous,

(b) h is quasiconvex,

(c) $U\subset(S_r)$ is convex and sequentially closed (weak-*),

(d) (i) U is bounded, or

(ii) g, h are bounded below and h is radially unbounded (i.e., $|v|\to\infty$ implies $h(v)\to\infty$).

The existence of an optimal control may be established as follows. Fix φ. Let $q=\inf\{Q(\varphi,u)\colon u\in U\}$ and choose $u_n\in U$ such that $Q(\varphi,u_n)\to q$. Hypothesis (d) implies that the sequence $\{u_n\}$ is bounded; let u^* be a weak-* limit point of (a subsequence of) $\{u_n\}$. Observe that $u^*\in U$ by hypothesis (c).

The continuity of g and Theorem 1 of Section 3 imply that $g(x(\varphi,u_n))\to g(x(\varphi,u))$. It follows from Mazur's theorem and the fact that $[u_n]\to[u]$ (weak) if $u_n\to u$ (weak-*) that $h([\cdot])$ is weak-* lower semicontinuous on U. Consequently $h([u^*]) \leqslant \liminf h([u_n])$. Therefore $q \leqslant Q(\varphi, u^*) \leqslant \liminf Q(\varphi, u_n) = q$.

The question of uniqueness is more difficult and will be addressed at present only in the event that: $f(s, \cdot, \zeta)$ is affine for all s and ζ; g is quasiconvex; and either g or h is strictly quasiconvex. Under these assumptions $Q(\varphi, \cdot)$ is strictly quasiconvex, so $Q(\varphi, u^*) = q = Q(\varphi, v^*)$ implies that $Q(\varphi, (u^* + v^*)/2) < q$ if $u^* \neq v^*$.

The projection operators π_n of Section 2 above may be employed (with $K = S_r$) to define a sequence of approximate optimization problems. Each such problem involves optimization in finite-dimensional subspaces of $A(S_r)$ and is consequently amenable to solution by numerical procedures.

Let $U_n = \{\pi_n u : u \in U \text{ and } |u| \leqslant n\}$ and let $\{\varphi_n\}$ be a sequence in $A(I_r)$ which converges to φ. For each n, define the approximate optimization problem $\mathscr{P}_n$: minimize $Q(\varphi_n, \cdot)$ over U_n. Lemma 1 below verifies that U_n satisfies hypothesis (c), and thereby guarantees the existence of a solution of each $\mathscr{P}_n$.

LEMMA 1. *If $U \subset A(S_r)$ is convex and sequentially closed (weak-*) then $\pi_n U$ has the same properties for all n.*

Proof. The convexity of each U_n is obvious. Suppose $\{u_i\} \subset U_n$ and $u_i \to u_0$ (weak-*). For each i there is a $v_i \in U$ such that $u_i = \pi_n v_i$ and $|v_i| \leqslant n$. Let v_0 be a weak-* limit point of (a subsequence of) the sequence $\{v_i\}$. Then $|\pi_n v_i| \leqslant n$ for $i = 0, 1, 2, \ldots$, and $(\pi_n v_i)(t) \to (\pi_n v_0)(t)$ for all $t \in S_r$; in particular, $(\pi_n v_i)(b) \to (\pi_n v_0)(b)$. The dominated convergence theorem implies that $\int_{S_r} \pi_n v_i \, dh \to \int_{S_r} \pi_n v_0 dh$ for all $h \in AC(S_r)$. Thus $u_i = \pi_n v_i \to \pi_n v_0$ (weak-*), so $u_0 = \pi_n v_0 \in U_n$.

Let u_n^* denote a solution of $\mathscr{P}_n$. Theorem 1 establishes the connection between problem $\mathscr{P}$ and problems $\mathscr{P}_n$.

THEOREM 1. *Assume hypotheses* (a)–(d); *assume further that $\pi_n U \subset U$ for all n. Then (a subsequence of) the sequence $\{u_n^*\}$ converges (weak-*) to a solution of problem $\mathscr{P}$. The corresponding trajectories and payoffs also converge.*

Proof. Since $\{u_n^*\} \subset U$, (a subsequence of) the sequence has a weak-* limit point in U if $\{u_n^*\}$ is norm-bounded. This is obvious under hypothesis (d)(i). So, assume hypothesis (d)(ii) and choose any $u \in U$. Then $\pi_n u \to u$ (weak-*) by Theorem 4 of Section 2 and the dominated convergence theorem. Therefore $Q(\varphi_n, \pi_n u) \to Q(\varphi, u)$ by Theorem 1 of Section 3, hypothesis (a), and the fact that $[\pi_n u] \to [u]$ in $L_2(S_r)$. Since $Q(\varphi_n, u_n^*) \leqslant Q(\varphi_n, \pi_n u)$ for all $n > |u|$ the sequence $\{Q(\varphi_n, u_n^*)\}$ is bounded above. The norm-boundedness of $\{u_n^*\}$ follows from the assumption that g is bounded below and h is radially unbounded.

Suppose u^* is a weak-* limit point of $\{u_n^*\}$. Then for any $u \in U$, $Q(\varphi, u^*) \leqslant \liminf Q(\varphi_n, u_n^*) \leqslant \limsup Q(\varphi_n, u_n^*) \leqslant \limsup Q(\varphi_n, \pi_n u) = Q(\varphi, u)$. Therefore u^* solves $\mathscr{P}$, and $\lim Q(\varphi_n, u_n^*) = Q(\varphi, u^*)$. The convergence of the trajectories $x(\varphi_n, u_n^*)$ to $x(\varphi, u^*)$ is a direct result of Theorem 1 of Section 3.

Given φ in problem $\mathscr{P}$, any convenient sequence $\{\varphi_n\}$ of initial data which converges to φ in $A(I_r)$ may be employed for $\mathscr{P}_n$. Each problem $\mathscr{P}_n$ may be solved numerically using standard search techniques in conjunction with an integration algorithm (see Cryer [4]) for HDE.

5. Comparison with Other Approaches

As noted in Section 1, the most popular state spaces for the study of HDE are $C(I_r)$ and $L_p(I_r) \times R$. $C(I_r)$ and $A(I_r)$ share the advantage that point evaluation of both trajectory and control is well defined and continuous. Delayed point evaluation of the trajectory may be dealt with in $L_p(I_r) \times R$ when the HDE is linear; see Reber [7]. $L_p(I_r) \times R$ and $A(I_r)$ share the advantage that the unit ball has compactness properties which permit the solution of optimal control problems.

A pleasant feature of $L_p(I_r) \times R$ is that for certain initial data the initial value problem is equivalent to: $z(t) = \varphi + \int_a^t \{A(s, z(s)) + (0, u(s))\}\, ds$, where $z(t) = (x_t, x(t))$ and $A(s, (\psi, \eta)) = (\dot{\psi}, f(s, \psi))$. A development along these lines may be found, e.g., in Banks and Burns [2] for the linear case and in Banks [1], Webb [9] for some nonlinear equations. Lamm [5] has extended the results of Banks [1] to nonlinear HDE with discrete delays in the state.

The corresponding ordinary differential equation in $C(I_r)$, with $z(t) = x_t$, is equivalent to an HDE only if $\dot{\varphi}(0^-) = f(a, \varphi) + u(a)$. This greatly restricts the utility of such an equation because the allowable data depend on the system itself as well as the control function.

There are still several details which must be investigated before a similar equivalence for $A(I_r)$-valued functions could be stated. In both of the above cases the indicated integral may be taken in the sense of Bochner. This is feasible because strong measurability of the integrand may be established without great difficulty. However, not only is $A(I_r)$ nonseparable, but $t \to (\dot{x})_t$ fails to have an essentially separable range even for relatively well-behaved functions $x \in A(S_r)$. Consequently, it appears that weak measurability is the best that one can hope for.

The next step would be to establish the Pettis integrability of $t \to (\dot{x})_t$. Then if equivalence could indeed be proved, it might be possible to obtain approximation results similar to those of Banks and Kappel [3].

One feature of the approach taken in this report is the presence of discrete and distributed delays in the control. The author is not aware of other investigations which treat this aspect of optimization of nonlinear hereditary systems.

REFERENCES

1. H. T. BANKS, Approximation of nonlinear functional differential equation control systems, *J. Optim. Theory Appl.* **29** (1979), 383–408.
2. H. T. BANKS AND J. A. BURNS, An abstract framework for approximate solutions to optimal control problems governed by hereditary systems, *in* "International Conference on Differential Equations" (H. Antosiewicz, Ed.), pp. 10–25, Academic Press, New York, 1975.
3. H. T. BANKS AND F. KAPPEL, Spline approximations for functional differential equations, *J. Differential Equations* **34** (1979), 496–522.
4. C. W. CRYER, Numerical methods for functional differential equations, *in* "Delay and Functional Differential Equations and Their Applications, Proc. Conf., Park City, Utah, 1972," pp. 17–101, Academic Press, New York, 1972.
5. P. K. LAMM, Spline approximations for nonlinear hereditary control systems, *J. Optim. Theory Appl.* **44** (1984), 585–624.
6. J. DIEUDONNÉ, Sur le théorème de Lebesgue–Nikodým, IV, *J. Indian Math. Soc.* (*N.S.*) **15** (1951), 77–86.
7. D. C. REBER, A finite difference technique for solving optimization problems governed by linear functional differential equations, *J. Differential Equations* **32** (1979), 193–232.
8. W. RUDIN, "Real and Complex Analysis," McGraw-Hill, New York, 1966.
9. G. F. WEBB, Functional differential equations and nonlinear semigroups in L^p-spaces, *J. Differential Equations* **20** (1976), 71–89.