Fundamental Solutions and Optimal Control of Neutral Systems

Kai Liu

Division of Statistics and Probability, Department of Mathematical Sciences,
The University of Liverpool, Peach Street, Liverpool, L69 7ZL, U.K.
k.liu@liv.ac.uk

Abstract. In this work, we shall consider standard optimal control problems for a class of neutral functional differential equations in Banach spaces. As the basis of a systematic theory of neutral models, the fundamental solution is constructed and a variation of constants formula of mild solutions is established. Necessary conditions in terms of the solutions of neutral adjoint systems are established to deal with the fixed time integral convex cost problem of optimality. Based on optimality conditions, the maximum principle for time varying control domain is presented.

1 Introduction

Let X be a separable Banach space with norm $\|\cdot\|_X$. For any fixed constant $r > 0$, we denote by $L_r^2 = L^2([-r,0];X)$ the space of all X-valued equivalence classes of measurable functions which are square integrable on $[-r,0]$. Let $\mathscr{X}$ denote the product Banach space $X \times L_r^2$ with the norm

$$\|\phi\|_{\mathscr{X}} = \sqrt{\|\phi_0\|_X^2 + \|\phi_1\|_{L_r^2}^2} \quad \text{for all} \quad \phi = (\phi_0, \phi_1) \in \mathscr{X}. \tag{1}$$

Consider the following neutral functional differential equation on X,

$$\frac{d}{dt}\left[y(t) - \int_{-r}^0 D(\theta)y(t+\theta)d\theta\right] = Ay(t) + \int_{-r}^0 d\eta(\theta)y(t+\theta) + f(t)$$
$$\text{for any } t > 0, \tag{2}$$
$$y(0) = \phi_0, \ y_0(\cdot) = \phi_1(\cdot), \ \phi = (\phi_0, \phi_1) \in \mathscr{X},$$

where $A : X \to X$ with domain $\mathscr{D}(A) \subset X$ is the infinitesimal generator of a C_0-semigroup $\{S(t); t \geq 0\}$ on X and $y_t(\theta) := y(t+\theta)$ for any $\theta \in [-r,0]$ and $t \geq 0$. Here $f(\cdot)$ is some properly given function in X and η is the Stieltjes measure given by

$$\eta(\tau) = -\sum_{i=1}^m \chi_{(-\infty, -r_i]}(\tau)A_i - \int_\tau^0 B(\theta)d\theta, \qquad \tau \in [-r,0]. \tag{3}$$

It is assumed that $0 < r_1 < r_2 < \cdots < r_m \leq r$, $A_i \in \mathscr{L}(X)$, $i = 1, \cdots, m$, the family of all bounded, linear operators on X and $B(\cdot), D(\cdot) \in L^2([-r,0]; \mathscr{L}(X))$, the Banach space of all $\mathscr{L}(X)$-valued equivalence classes of square integrable functions on $[-r,0]$.

The abstract formulation (2) in an infinite dimensional space has been well motivated both theoretically and practically by such systems as neutral partial functional differential equations. There exists an extensive literature which deals with various problems of the so-called distributed parameter systems with time delays in a Banach space (see the monograph [19], for instance, and references cited therein for a systematic statement). The system (2) in finite or infinite dimensions was considered in a systematic way by [8] and [19] in spaces of continuous functions, e.g., in $C([-r,0];X)$, the space of all continuous functions from $[-r,0]$ into X. The same phase spaces were also used by many works such as [7,9] and references cited therein. Although this choice certainly has its advantages, it is often useful to work in L^p space instead. There are at least two good reasons for this setting. First, it allows ones to work in a Hilbert space. This is particularly important in control, stability and optimality theory in which an adjoint theory for the system (2) is essentially needed, e.g., [1,5,6,11,17]. Unfortunately, as indicated in Hale and Lunel [9], it is generally difficult to have an adjoint theory in a continuous functions space setting. Second, the choice of L^p-phase space allows us to consider equations having discontinuous solutions. This is the case when ones consider discontinuous initial functions for (2) or the system is perturbed by some random resources with jumps (cf. [14]).

Although there is some work, e.g., [13], on optimal control of functional differential equations in infinite dimensions available in the existing literature, there exist however few results on the same topic for functional differential equations of neutral type. The current work will devote itself to exploring some basic material. We shall consider the fixed time integral convex cost problem for mild solutions of a class of functional differential equations of neutral type in Banach spaces. To this end, we first formulate and study a class of linear neutral systems to derive fundamental results. Precisely, as the basis of whole theory we shall construct fundamental solutions or Green operators and establish a variation of constants formula. Adjoint neutral systems are considered and the associated representation formulae of adjoint states in terms of fundamental solutions are established. All of these will allow us to present results on the existence of optimal controls, necessary conditions of optimality and maximum principle.

The following are some notations and terminologies to be used in the work. The symbol $\mathbb{R}_+$ denotes the set of all nonnegative numbers and $\mathbb{R}^n$ denotes the n-dimensional real vector space with the usual Euclidean norm $\|\cdot\|_{\mathbb{R}^n}$. For any $\lambda \in \mathbb{C}^1$, the symbols $\Re(\lambda)$ and $\Im(\lambda)$ denote the real and imaginary parts of complex number λ, respectively. For any separable Banach space U, we use U^* to denote its adjoint space and $\langle\cdot,\cdot\rangle_{U,U^*}$ the dual pairing, respectively. We use $\mathscr{L}(U,X)$ to denote the space consisting of all bounded linear operators T from U into X with domain U. When $X = U$, $\mathscr{L}(X,X)$ is denoted by $\mathscr{L}(X)$. Every operator norm is simply denoted by $\|\cdot\|$ when there is no danger of confusion. The symbols $\mathscr{D}(T)$ and $\mathscr{R}(T)$ will be used to denote the domain and range of operator T, respectively. For a closed linear operator A on a dense domain $\mathscr{D}(A) \subset X$ into X, its adjoint operator is denoted by A^*. Given an interval $E \subset \mathbb{R}^1$, the function χ_E denotes the characteristic function on the interval E. For a measurable function $f : \mathbb{R}^1 \to X$, its Laplace transform $\hat{f}$ is defined by

$$\hat{f}(\lambda) = \int_0^\infty e^{-\lambda t} f(t) dt$$

whenever the Bochner integral exists.

2   Fundamental Solutions

In this section, we shall consider a class of linear autonomous neutral functional differential equations on X which are defined formally by

$$\frac{d}{dt}\left[y(t) - \int_{-r}^0 D(\theta)y(t+\theta)d\theta\right] = Ay(t) + \int_{-r}^0 d\eta(\theta)y(t+\theta) \text{ for any } t > 0,$$
$$y(0) = \phi_0, \ y_0(\cdot) = \phi_1(\cdot), \quad \phi = (\phi_0, \phi_1) \in \mathcal{X}, \tag{4}$$

where $A : X \to X$ and η are defined as in (2) and (3). Generally, it is quite restrictive to find a solution in the usual sense for the equation (4). Instead, it is hoped to consider an "integrated" form of the system (4). To this end, we further assume that for each i, $i = 1, \cdots, m$, and $\theta \in [-r,0]$, $\mathcal{R}(D(\theta)) \subset \mathcal{D}(A)$ such that $AD(\cdot) \in L^2([-r,0]; \mathcal{L}(X))$.

Consider the following integral equation on X

$$y(t,\phi) = \int_{-r}^0 D(\theta)y(t+\theta,\phi)d\theta + S(t)\left[\phi_0 - \int_{-r}^0 D(\theta)\phi_1(\theta)d\theta\right]$$
$$+ \int_0^t S(t-s)\left[\int_{-r}^0 d\eta(\theta)y(s+\theta,\phi) + \int_{-r}^0 AD(\theta)y(s+\theta,\phi)d\theta\right]ds, \ \forall t > 0,$$
$$y(0,\phi) = \phi_0, \ y_0(\cdot,\phi) = \phi_1(\cdot), \quad \phi = (\phi_0,\phi_1) \in \mathcal{X}. \tag{5}$$

For simplicity, we sometimes denote $y(t,\phi)$ and $y_t(\cdot,\phi)$ by $y(t)$ and $y_t(\cdot)$, respectively, in the remainder of this work. The following existence and uniqueness of solutions of Equation (5) can be established in the spirit of Datko [5] and Wu [18].

Theorem 1. *For arbitrary $T \geq 0$, $\phi = (\phi_0,\phi_1) \in \mathcal{X}$, (i) there exists a unique solution $y(t,\phi) \in L^2([-r,T];X)$ of (5); (ii) for arbitrary $t \in [0,T]$, $\|y(t,\phi)\|_X \leq Ce^{\gamma t}\|\phi\|_{\mathcal{X}}$ almost everywhere for some constants $\gamma \in \mathbb{R}^1$ and $C > 0$.*

The solution $y(t,\phi)$ of the equation (5) is called a *mild solution* of (4). For any $x \in X$, let $\phi_0 = x$, $\phi_1(\theta) = 0$ for $\theta \in [-r,0)$ and $\phi = (x,0)$, we define the *fundamental solution* $G(t) \in \mathcal{L}(X)$, $t \in \mathbb{R}^1$, of (5) with such an initial datum by

$$G(t)x = \begin{cases} y(t,\phi), \ t \geq 0, \\ 0, \quad\quad t < 0. \end{cases} \tag{6}$$

The term (6) implies that $G(t)$ is a unique solution of the equation

$$G(t) = S(t) + \int_{-r}^0 D(\theta)G(t+\theta)d\theta + \int_0^t S(t-s)\left[\int_{-r}^0 d\eta(\theta)G(s+\theta)\right.$$
$$\left. + \int_{-r}^0 AD(\theta)G(s+\theta)d\theta\right]ds, \quad \text{if } t \geq 0,$$
$$G(t) = O, \quad \text{if } t < 0, \tag{7}$$

where O is the null operator on X. It is immediate to see that $G(t)$ is strongly continuous on $\mathbb{R}^1$ and satisfies

$$\|G(t)\| \leq C \cdot e^{\gamma t}, \qquad t \geq 0, \tag{8}$$

for some $C > 0$ and $\gamma \in \mathbb{R}^1$.

2.1 Variation of Constants Formula

Consider a class of non-autonomous neutral functional differential equations on X

$$\frac{d}{dt}\left[y(t) - \int_{-r}^{0} D(\theta)y(t+\theta)d\theta\right] = Ay(t) + \int_{-r}^{0} d\eta(\theta)y(t+\theta) + f(t)$$
$$\text{for any } t > 0, \tag{9}$$
$$y(0) = \phi_0, \, y_0(\cdot) = \phi_1(\cdot), \quad \phi = (\phi_0, \phi_1) \in \mathscr{X},$$

where $A : X \to X$, η are defined as in the last subsection and $f(\cdot) \in L^2_{loc}(\mathbb{R}_+; X)$, the Fréchet space of functions which belong to $L^2([0,T]; X)$ for any $T \geq 0$. Once again, we intend to consider the following integral equation of (9) on X,

$$y(t, f, \phi) = \int_{-r}^{0} D(\theta)y(t+\theta, f, \phi)d\theta + S(t)\left[\phi_0 - \int_{-r}^{0} D(\theta)\phi_1(\theta)d\theta\right]$$
$$+ \int_{0}^{t} S(t-s)\left[\int_{-r}^{0} d\eta(\theta)y(s+\theta, f, \phi)\right.$$
$$\left. + \int_{-r}^{0} AD(\theta)y(s+\theta, f, \phi)d\theta + f(s)\right]ds, \ \forall t > 0, \tag{10}$$
$$y(0, \phi) = \phi_0, \, y_0(\cdot, \phi) = \phi_1(\cdot), \quad \phi = (\phi_0, \phi_1) \in \mathscr{X}.$$

It is extremely useful to find an explicit representation for the solution $y(t, f, \phi)$ of (10) in applications, e.g., in the optimal control theory. This is possible if we restrict the initial data of (10) to some proper subset of $\mathscr{X}$. Indeed, let $W^{1,2}([-r, 0]; X)$ denote the Sobolev space of X-valued functions $x(\cdot)$ on $[-r, 0]$ such that $x(\cdot)$ and its distributional derivative belong to $L^2([-r, 0]; X)$, and define $\mathscr{W}^{1,2} = X \times W^{1,2}([-r, 0]; X)$.

The following *variation of constants formula* (11) provides a representation for solutions of (10) in terms of the fundamental solution $G(t) \in \mathscr{L}(X)$.

Theorem 2. *For arbitrary $\phi = (\phi_0, \phi_1) \in \mathscr{W}^{1,2}$, the solution $y(t, f, \phi)$ of (10) can be represented almost everywhere by*

$$y(t, f, \phi) = G(t)\phi_0 - V(t, 0)\phi_1(0) + \int_{-r}^{0} U(t, \theta)\phi_1(\theta)d\theta$$
$$+ \int_{-r}^{0} V(t, \theta)\phi_1'(\theta)d\theta + \int_{0}^{t} G(t-s)f(s)ds, \quad t \geq 0, \tag{11}$$

where for any $t \geq 0$, the kernels

$$U(t, \theta) = \int_{-r}^{\theta} G(t-\theta+\tau)d\eta(\tau) \in L^2([-r, 0]; \mathscr{L}(X)), \qquad \theta \in [-r, 0], \tag{12}$$

and, similarly,

$$V(t,\theta) = \int_{-r}^{\theta} G(t-\theta+\tau)D(\tau)d\tau \in L^2([-r,0];\mathscr{L}(X)), \qquad \theta \in [-r,0]. \qquad (13)$$

Proof. We first prove (11) by assuming $f \in L^2(\mathbb{R}_+;X) \cap L^1(\mathbb{R}_+;X)$. To this end, define

$$x(t) = G(t)\phi_0 - V(t,0)\phi_1(0) + \int_{-r}^{0} U(t,\theta)\phi_1(\theta)d\theta + \int_{-r}^{0} V(t,\theta)\phi_1'(\theta)d\theta$$
$$+ \int_{0}^{t} G(t-s)f(s)ds, \quad t \geq 0, \qquad (14)$$

and $x(t) = \phi_1(t)$ for $t \in [-r,0)$. It is easy to see that $x(t) \in L^2([0,T];X)$ and $x(t)$ is almost everywhere continuous on $[0,T]$. For $\lambda \in \mathbb{C}^1$ with $\Re(\lambda)$ large enough, define

$$M(\lambda) = I - \int_{-r}^{0} e^{\lambda\theta}D(\theta)d\theta,$$
$$N(\lambda) = R(\lambda,A)\left[\int_{-r}^{0} e^{\lambda\theta}d\eta(\theta) + \int_{-r}^{0} e^{\lambda\theta}D(\theta)d\theta\right], \qquad (15)$$

where I denotes the identity operator on X and $R(\lambda,A) = (\lambda I - A)^{-1}$ is the resolvent operator of A. From the structure of both $M(\lambda)$ and $N(\lambda)$, we see that $M(\lambda) \to I$ and $\|N(\lambda)\| \to 0$ as $\Re(\lambda) \to +\infty$. This implies that there exists a real number λ_0 such that for $\Re(\lambda) \geq \lambda_0$, both $M^{-1}(\lambda)$ and $[I - N(\lambda)M^{-1}(\lambda)]^{-1}$ exist. Therefore, we can apply the convolution theorem on Laplace transforms to (7) to obtain that

$$\hat{G}(\lambda) = R(\lambda,A) + \int_{-r}^{0} e^{\lambda\theta}D(\theta)d\theta \cdot \hat{G}(\lambda) + R(\lambda,A)\int_{-r}^{0} e^{\lambda\theta}d\eta(\theta) \cdot \hat{G}(\lambda)$$
$$+ R(\lambda,A)\int_{-r}^{0} e^{\lambda\theta}AD(\theta)d\theta \cdot \hat{G}(\lambda), \qquad (16)$$

where $\hat{G}(\lambda)$ denotes the Laplace transform of $G(\cdot)$. This yields that for $\Re(\lambda) \geq \lambda_0$,

$$\hat{G}(\lambda) = M^{-1}(\lambda)\left[I - N(\lambda)M^{-1}(\lambda)\right]^{-1}R(\lambda,A). \qquad (17)$$

Note that the Laplace transform $\hat{x}(\lambda)$ of $x(\cdot)$ makes sense for sufficiently large $\Re(\lambda)$. Therefore, we apply the convolution theorem on Laplace transforms to (14) and use Fubini's theorem to obtain that

$$\hat{x}(\lambda) = \hat{G}(\lambda)\left[\phi_0 - \int_{-r}^{0} D(\theta)\phi_1(\theta)d\theta + \left(\sum_{i=1}^{m} A_i e^{-\lambda r_i}\int_{-r_i}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau\right)\right.$$
$$+ \left(\int_{-r}^{0} e^{\lambda\theta}B(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau d\theta\right)$$
$$+ \int_{-r}^{0} \lambda e^{\lambda\theta}D(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau d\theta + \hat{f}(\lambda)\right] \qquad (18)$$
$$= \hat{G}(\lambda)\left[\phi_0 - \int_{-r}^{0} e^{\lambda\theta}D(\theta)\phi_1(0)d\theta + \int_{-r}^{0} e^{\lambda\theta}D(\theta)\int_{\theta}^{0} e^{-\lambda\tau}d\phi_1(\tau)d\theta\right.$$
$$+ \int_{-r}^{0} e^{\lambda\theta}d\eta(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau + \hat{f}(\lambda)\right].$$

On the other hand, since $y(\cdot)$ satisfies the equation (10) and note that $y(t) = \phi_1(t)$ for $t \in [-r,0)$, we can use Fubini's theorem again to calculate the Laplace transform $\hat{y}(\lambda)$ of $y(\cdot)$ which is given by

$$
\begin{aligned}
\hat{y}(\lambda) = {}& \int_{-r}^0 e^{\lambda\theta} D(\theta) d\theta \hat{y}(\lambda) + \int_{-r}^0 e^{\lambda\theta} D(\theta) \int_\theta^0 e^{-\lambda\tau} \phi_1(\tau) d\tau d\theta \\
& + R(\lambda,A) \Big\{ \Big[\phi_0 - \int_{-r}^0 D(\theta)\phi_1(\theta) d\theta \Big] + \Big[\int_{-r}^0 e^{\lambda\theta} d\eta(\theta) + \int_{-r}^0 e^{\lambda\theta} AD(\theta) d\theta \Big] \hat{y}(\lambda) \\
& + \Big[\int_{-r}^0 e^{\lambda\theta} d\eta(\theta) \int_\theta^0 e^{-\lambda\tau}\phi_1(\tau) d\tau + \\
& \int_{-r}^0 e^{\lambda\theta} AD(\theta) \int_\theta^0 e^{-\lambda\tau}\phi_1(\tau) d\tau d\theta \Big] + \hat{f}(\lambda) \Big\},
\end{aligned}
$$

$$\tag{19}$$

for sufficiently large $\Re(\lambda)$. In terms of (15), we can rewrite (19) as

$$
\begin{aligned}
(M(\lambda) &- N(\lambda))\hat{y}(\lambda) \\
= {}& \Big\{ \Big(I - \int_{-r}^0 e^{\lambda\theta} D(\theta) d\theta \Big) - R(\lambda,A) \Big[\int_{-r}^0 e^{\lambda\theta} d\eta(\theta) + \int_{-r}^0 e^{\lambda\theta} AD(\theta) d\theta \Big] \Big\} \hat{y}(\lambda) \\
= {}& \int_{-r}^0 e^{\lambda\theta} D(\theta) \int_\theta^0 e^{-\lambda\tau}\phi_1(\tau) d\tau d\theta + R(\lambda,A) \Big[\phi_0 - \int_{-r}^0 D(\theta)\phi_1(\theta) d\theta \Big] \\
& + R(\lambda,A) \Big[\int_{-r}^0 d\eta(\theta) \int_\theta^0 e^{\lambda(\theta-\tau)}\phi_1(\tau) d\tau + \\
& \int_{-r}^0 AD(\theta) \int_\theta^0 e^{\lambda(\theta-\tau)}\phi_1(\tau) d\tau d\theta + \hat{f}(\lambda) \Big],
\end{aligned}
$$

which immediately yields that

$$
\begin{aligned}
\hat{y}(\lambda) = {}& M^{-1}(\lambda) \big[I - N(\lambda) M^{-1}(\lambda) \big]^{-1} \Big\{ \int_{-r}^0 e^{\lambda\theta} D(\theta) \int_\theta^0 e^{-\lambda\tau}\phi_1(\tau) d\tau d\theta \\
& + R(\lambda,A) \Big[\phi_0 - \int_{-r}^0 D(\theta)\phi_1(\theta) d\theta \Big] \\
& + R(\lambda,A) \Big[\int_{-r}^0 e^{\lambda\theta} d\eta(\theta) \int_\theta^0 e^{-\lambda\tau}\phi_1(\tau) d\tau + \\
& \int_{-r}^0 e^{\lambda\theta} AD(\theta) \int_\theta^0 e^{-\lambda\tau}\phi_1(\tau) d\tau d\theta \Big] \\
& + R(\lambda,A)\hat{f}(\lambda) \Big\}.
\end{aligned}
$$

$$\tag{20}$$

Then, by virtue of (17), it is immediate to see that for large $\Re(\lambda)$,

$$
\begin{aligned}
\hat{y}(\lambda) &= \hat{G}(\lambda)\Big[(\lambda I - A)\Big(\int_{-r}^{0} e^{\lambda\theta} D(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau d\theta\Big) + \\
&\quad \phi_0 - \int_{-r}^{0} D(\theta)\phi_1(\theta)d\theta \\
&\quad + \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau + \\
&\quad \int_{-r}^{0} e^{\lambda\theta} AD(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau d\theta + \hat{f}(\lambda)\Big] \\
&= \hat{G}(\lambda)\Big[\phi_0 - \int_{-r}^{0} e^{\lambda\theta} D(\theta)\phi_1(0)d\theta + \int_{-r}^{0} e^{\lambda\theta} D(\theta)\int_{\theta}^{0} e^{-\lambda\tau}d\phi_1(\tau)d\theta \\
&\quad + \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta)\int_{\theta}^{0} e^{-\lambda\tau}\phi_1(\tau)d\tau + \hat{f}(\lambda)\Big].
\end{aligned}
\tag{21}
$$

Therefore, from (18) it follows that

$$
\hat{y}(\lambda) = \hat{x}(\lambda)
$$

for sufficiently large $\Re(\lambda)$. By the uniqueness of Laplace transforms and the almost everywhere strong continuity of $y(t)$ and $x(t)$ on $\mathbb{R}_+$, we obtain that

$$
y(t) = x(t) \quad \text{for almost all } t \in \mathbb{R}_+,
$$

which proves the desired result. Lastly, we shall prove (11) for $f \in L^2_{loc}(\mathbb{R}^+;X)$. To this end, it suffices to prove (11) for $t \in [0,T]$ with any fixed $T \geq 0$. For a given $f \in L^2_{loc}(\mathbb{R}^+;X)$ and $T \geq 0$, we define the truncated function $f_T(t) = \chi_{[0,T]}(t)f(t)$. Then $f_T(\cdot) \in L^2(\mathbb{R}^+;X) \cap L^1(\mathbb{R}^1;X)$ and the corresponding solution $y_T(t)$ of (5) satisfies (11) for all $t \geq 0$. Since $y_T(t) = y(t)$ for $t \in [0,T]$, then (11) is true for all $f(\cdot) \in L^2_{loc}(\mathbb{R}^+;X)$. The proof is now complete. $\qquad\square$

3   Neutral Resolvent Operators

For each $\lambda \in \mathbb{C}^1$, we define the densely defined, closed linear operator $\Delta(\lambda,A,\eta,D)$ by

$$
\Delta(\lambda,A,\eta,D) = \lambda I - A - \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) - \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta)d\theta.
$$

The neutral resolvent set $\rho(A,\eta,D)$ is defined as the set of all values λ in $\mathbb{C}^1$ for which the operator $\Delta(\lambda,A,\eta,D)$ has a bounded inverse in X.

Proposition 1. *(i) Let $x \in X$, then $\int_0^t G(s)xds \in \mathscr{D}(A)$ for almost all $t \in \mathbb{R}_+$, and the relation*

$$
\begin{aligned}
A \int_0^t G(s)xds &= G(t)x - x - \int_0^t \int_{-r}^{0} d\eta(\theta)G(s+\theta)xds \\
&\quad - \int_{-r}^{0} D(\theta)G(t+\theta)xd\theta, \quad t \in \mathbb{R}_+, \quad x \in X,
\end{aligned}
\tag{22}
$$

holds almost everywhere.

(ii) Let $x \in \mathscr{D}(A)$, then $\int_0^t G(s)Ax\,ds \in X$ for almost all $t \in \mathbb{R}_+$, and the commutative relation

$$\int_0^t G(s)Ax\,ds = G(t)x - x - \int_0^t \int_{-r}^0 G(s+\theta)d\eta(\theta)x\,ds$$

$$- \int_{-r}^0 G(t+\theta)D(\theta)x\,d\theta, \quad t \in \mathbb{R}_+, \quad x \in \mathscr{D}(A), \tag{23}$$

holds almost everywhere.

Proof. We first prove the claim (i). For any $x \in X$ and $\varepsilon > 0$, let $A(\varepsilon) = \varepsilon^{-1}(S(\varepsilon) - I)$. We calculate $A(\varepsilon)\int_0^t G(s)x\,ds$ as follows:

$$A(\varepsilon)\int_0^t G(s)x\,ds = \varepsilon^{-1}\int_0^t \Big\{ S(\varepsilon)\Big[S(s) + \int_{-r}^0 D(\theta)G(s+\theta)d\theta$$

$$+ \int_0^s S(s-u)\Big(\int_{-r}^0 d\eta(\theta)G(u+\theta) +$$

$$\int_{-r}^0 AD(\theta)G(u+\theta)d\theta\Big)du\Big] - G(s)\Big\}x\,ds$$

$$= \varepsilon^{-1}\int_0^t \Big\{ G(s+\varepsilon) - G(s) + (S(\varepsilon) - I)\int_{-r}^0 D(\theta)G(s+\theta)d\theta - \tag{24}$$

$$\int_{-r}^0 D(\theta)G(s+\theta+\varepsilon)d\theta$$

$$+ \int_{-r}^0 D(\theta)G(s+\theta)d\theta - \int_0^\varepsilon S(\varepsilon-v)\Big[\int_{-r}^0 d\eta(\theta)G(s+v+\theta)$$

$$+ \int_{-r}^0 AD(\theta)G(s+v+\theta)d\theta\Big]dv\Big\}x\,ds, \quad t \geq 0.$$

Note the strong continuity of $G(t)$ and it is not difficult for us to deduce that

$$\lim_{\varepsilon \to 0+} A(\varepsilon)\int_0^t G(s)x\,ds = \lim_{\varepsilon \to 0+} \varepsilon^{-1}\int_0^\varepsilon \Big(G(t+s) - G(s)\Big)x\,ds$$

$$+ \int_0^t \int_{-r}^0 AD(\theta)G(s+\theta)x\,d\theta\,ds - \lim_{\varepsilon \to 0+} \varepsilon^{-1}\Big(\int_0^t \int_{-r}^0 D(\theta)G(s+\theta+\varepsilon)d\theta\,ds$$

$$- \int_0^t \int_{-r}^0 D(\theta)G(s+\theta)d\theta\Big)x\,ds - \int_0^t \int_{-r}^0 d\eta(\theta)G(s+\theta)x\,ds -$$

$$\int_0^t \int_{-r}^0 AD(\theta)G(s+\theta)x\,d\theta\,ds \tag{25}$$

$$= G(t)x - x - \lim_{\varepsilon \to 0+} \varepsilon^{-1}\Big(\int_0^t \int_{-r}^0 D(\theta)G(s+\theta+\varepsilon)d\theta\,ds -$$

$$\int_0^t \int_{-r}^0 D(\theta)G(s+\theta)d\theta\,ds\Big)x$$

$$- \int_0^t \int_{-r}^0 d\eta(\theta)G(s+\theta)x\,ds.$$

However, it is easy to see that

$$
\lim_{\varepsilon \to 0+} \varepsilon^{-1} \left(\int_0^t \int_{-r}^0 D(\theta) G(s + \theta + \varepsilon) d\theta ds - \int_0^t \int_{-r}^0 D(\theta) G(s + \theta) d\theta ds \right) x
$$
$$
= \int_{-r}^0 D(\theta) \lim_{\varepsilon \to 0+} \left\{ \int_0^\varepsilon \varepsilon^{-1} G(t + s + \theta) ds - \int_0^\varepsilon \varepsilon^{-1} G(s + \theta) ds \right\} x d\theta = \quad (26)
$$
$$
\int_{-r}^0 D(\theta) G(t + \theta) x d\theta.
$$

Therefore, $\int_0^t G(s) x ds \in \mathscr{D}(A)$ for almost all $t \in \mathbb{R}_+$, and

$$
A \int_0^t G(s) x ds = G(t) x - x - \int_0^t \int_{-r}^0 d\eta(\theta) G(s + \theta) x ds -
$$
$$
\int_{-r}^0 D(\theta) G(t + \theta) x d\theta, \quad t \in \mathbb{R}_+, \quad (27)
$$

almost everywhere. The proof of (i) is complete.

Next, we intend to prove the relation (23). Firstly, by definition note that for sufficiently large $\mathfrak{R}(\lambda)$,

$$
\Delta(\lambda, A, \eta, D) = (\lambda I - A) \left[I - R(\lambda, A) \int_{-r}^0 e^{\lambda\theta} d\eta(\theta) - R(\lambda, A) \int_{-r}^0 \lambda e^{\lambda\theta} D(\theta) d\theta \right],
$$

and by using (15), we have that

$$
\left\| R(\lambda, A) \int_{-r}^0 e^{\lambda\theta} d\eta(\theta) + R(\lambda, A) \int_{-r}^0 \lambda e^{\lambda\theta} D(\theta) d\theta \right\|
$$
$$
= \| N(\lambda) - M(\lambda) + I \| \leq \| N(\lambda) \| + \| I - M(\lambda) \| \to 0 \text{ as } \mathfrak{R}(\lambda) \to +\infty. \quad (28)
$$

Hence, the bounded inverse

$$
\left[I - R(\lambda, A) \int_{-r}^0 e^{\lambda\theta} d\eta(\theta) - R(\lambda, A) \int_{-r}^0 \lambda e^{\lambda\theta} D(\theta) d\theta \right]^{-1}
$$

exists for sufficiently large $\mathfrak{R}(\lambda)$. For such a $\mathfrak{R}(\lambda)$ it is easy to deduce that

$$
\hat{G}(\lambda) x = \left[I - R(\lambda, A) \int_{-r}^0 e^{\lambda\theta} d\eta(\theta) - R(\lambda, A) \int_{-r}^0 \lambda e^{\lambda\theta} D(\theta) d\theta \right]^{-1} R(\lambda, A) x, \quad \forall x \in X.
$$
$$
(29)
$$

For any $x \in \mathscr{D}(A)$, it then follows from (29) that

$$
\lambda \hat{G}(\lambda)x - x =
$$

$$
\lambda \left[I - R(\lambda,A) \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) - R(\lambda,A) \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right]^{-1} R(\lambda,A)x
$$

$$
- \left[I - R(\lambda,A) \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) - R(\lambda,A) \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right]^{-1} R(\lambda,A)(\lambda I - A)
$$

$$
\cdot \left[I - R(\lambda,A) \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) - R(\lambda,A) \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right] x \tag{30}
$$

$$
= \left[I - R(\lambda,A) \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) - R(\lambda,A) \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right]^{-1} R(\lambda,A)
$$

$$
\cdot \left\{ \lambda I - \left(\lambda I - A - \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) - \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right) \right\} x
$$

$$
= \hat{G}(\lambda) \left(A + \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) + \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right) x.
$$

On the other hand, note that the following Laplace transform holds

$$
\int_{0}^{\infty} e^{-\lambda t} \left(\int_{0}^{t} \left\{ G(s)Ax + \int_{-r}^{0} G(s+\theta) d\eta(\theta)x \right\} ds + \int_{-r}^{0} G(t+\theta) D(\theta)xd\theta \right) dt
$$

$$
= \lambda^{-1} \hat{G}(\lambda) \left(A + \int_{-r}^{0} e^{\lambda\theta} d\eta(\theta) + \int_{-r}^{0} \lambda e^{\lambda\theta} D(\theta) d\theta \right) x, \tag{31}
$$

for sufficiently large $\Re(\lambda)$. Therefore, by (30), (31) and the uniqueness of Laplace transforms, we have that for all $x \in \mathscr{D}(A)$,

$$
G(t)x - x = \int_{0}^{t} \left\{ G(s)Ax + \int_{-r}^{0} G(s+\theta) d\eta(\theta)x \right\} ds +
$$

$$
\int_{-r}^{0} G(t+\theta) D(\theta)xd\theta, \quad t \in \mathbb{R}_+, \tag{32}
$$

almost everywhere. This completes the proof of (ii). □

4 Adjoint Theory

In the sequel we shall assume that X is reflexive. We intend to establish an adjoint theory of neutral functional differential equations. Let $\psi^* = (\psi_0^*, \psi_1^*) \in \mathscr{X}^*$. The "formal" transposed neutral system of (4) on X^* is defined by

$$
\frac{d}{dt} \left[y^*(t) - \int_{-r}^{0} D^*(\theta) y^*(t+\theta) d\theta \right] = A^* y^*(t) + \int_{-r}^{0} d\eta^*(\theta) y^*(t+\theta) + f^*(t),
$$

$$
t > 0, \tag{33}
$$

$$
y^*(0) = \psi_0^*, \quad y_0^*(\cdot) = \psi_1^*(\cdot), \psi^* = (\psi_0^*, \psi_1^*) \in \mathscr{X}^*,
$$

where $\eta^*(\theta)$, $D^*(\theta)$ and A^* denote the adjoint operators of $\eta(\theta)$, $D(\theta)$ and A, respectively, and $f^* \in L^1([0,T];X^*)$. It is well known that A^* generates a C_0-semigroup

$S^*(t)$ on X^* which is the adjoint of $S(t)$, $t \geq 0$. Hence, we can construct a fundamental solution $G_*(t)$ which is characterized as the unique solution of

$$G_*(t) = S^*(t) + \int_{-r}^0 D^*(\theta)G_*(t+\theta)d\theta + \int_0^t S^*(t-s)\left[\int_{-r}^0 d\eta^*(\theta)G_*(\theta+s)\right.$$
$$\left. + \int_{-r}^0 A^*D^*(\theta)G_*(s+\theta)d\theta\right]ds, \quad t \geq 0, \tag{34}$$

$$G_*(t) = \mathrm{O}, \quad t < 0.$$

We denote by $G^*(t)$ the adjoint of $G(t)$, $t \in \mathbb{R}^1$. The following theorem shows that $G^*(t) = G_*(t)$ for all $t \in \mathbb{R}^1$.

Theorem 3. *Let $G_*(t)$ be the solution of (34). Then*

$$G^*(t) = G_*(t) \quad \text{for almost all } t \in \mathbb{R}^1.$$

Proof. Since $G(t)$ satisfies (7), then

$$G^*(t) = S^*(t) + \int_{-r}^0 G^*(t+\theta)D^*(\theta)d\theta + \int_0^t \left(\int_{-r}^0 G^*(s+\theta)d\eta^*(\theta)\right.$$
$$\left. + \int_{-r}^0 G^*(s+\theta)D^*(\theta)A^*d\theta\right)S^*(t-s)ds, \quad t \geq 0. \tag{35}$$

Note that $S^*(t)$ is strongly continuous on $\mathbb{R}_+$. Then by using (34), (35) and the Lebesgue dominated convergence theorem, $G^*(t)x^*$ and $G_*(t)x^*$ are of exponential order for each $x^* \in X^*$. Hence, both $G^*(t)$ and $G_*(t)$ are Laplace transformable. Taking Laplace transform on both sides of (34), we have for sufficiently large $\mathfrak{R}(\lambda)$ that

$$\hat{G}_*(\lambda) = R(\lambda, A^*) + \int_{-r}^0 e^{\lambda\theta}D^*(\theta)d\theta \cdot \hat{G}_*(\lambda) + R(\lambda, A^*)\int_{-r}^0 e^{\lambda\theta}d\eta^*(\theta) \cdot \hat{G}_*(\lambda)$$
$$+ R(\lambda, A^*)\int_{-r}^0 e^{\lambda\theta}A^*D^*(\theta)d\theta \cdot \hat{G}_*(\lambda), \tag{36}$$

where $R(\lambda, A^*)$ denotes the resolvent of A^*. Similarly to (15), ones can define

$$M_*(\lambda) = I - \int_{-r}^0 e^{\lambda\theta}D^*(\theta)d\theta,$$
$$N_*(\lambda) = R(\lambda, A^*)\left[\int_{-r}^0 e^{\lambda\theta}d\eta^*(\theta) + \int_{-r}^0 e^{\lambda\theta}A^*D^*(\theta)d\theta\right], \tag{37}$$

for arbitrary $\lambda \in \mathbb{C}^1$ with $\mathfrak{R}(\lambda)$ large enough. It is easy to see that $M_*(\lambda) \to I$ and $\|N_*(\lambda)\| \to 0$ as $\mathfrak{R}(\lambda) \to +\infty$. This implies that for sufficiently large $\mathfrak{R}(\lambda)$, both $M_*(\lambda)^{-1}$ and $[I - M_*(\lambda)^{-1}N_*(\lambda)]^{-1}$ exist. Therefore, by virtue of (37) and (36), we can rewrite $\hat{G}_*(\lambda)$ as

$$\hat{G}_*(\lambda) = \left[I - M_*(\lambda)^{-1}N_*(\lambda)\right]^{-1}M_*(\lambda)^{-1}R(\lambda, A^*). \tag{38}$$

On the other hand, the following equality holds

$$\hat{G}(\lambda)\left(\lambda I - A - \int_{-r}^{0} e^{\lambda\theta}d\eta(\theta) - \int_{-r}^{0}\lambda e^{\lambda\theta}D(\theta)d\theta\right) = I \tag{39}$$

for sufficiently large $\mathfrak{R}(\lambda)$. Substituting $\lambda = \bar{\lambda}$ (complex conjugate) into (39) and taking its adjoint, we obtain

$$\begin{aligned}I &= \left(\bar{\lambda}I - A - \int_{-r}^{0} e^{\bar{\lambda}\theta}d\eta(\theta) - \int_{-r}^{0}\bar{\lambda}e^{\bar{\lambda}\theta}D(\theta)d\theta\right)^{*}(\hat{G}(\bar{\lambda}))^{*}\\ &= \left(\lambda I - A^{*} - \int_{-r}^{0} e^{\lambda\theta}d\eta^{*}(\theta) - \int_{-r}^{0}\lambda e^{\lambda\theta}D^{*}(\theta)d\theta\right)\hat{G}^{*}(\lambda),\end{aligned} \tag{40}$$

so that

$$\hat{G}^{*}(\lambda) = \left(I - A^{*} - \int_{-r}^{0} e^{\lambda\theta}d\eta^{*}(\theta) - \int_{-r}^{0}\lambda e^{\lambda\theta}D^{*}(\theta)d\theta\right)^{-1}. \tag{41}$$

Note that we have

$$\begin{aligned}R(\lambda,A^{*})^{-1}&M_{*}(\lambda)\left[I - M_{*}(\lambda)^{-1}N_{*}(\lambda)\right] = R(\lambda,A^{*})^{-1}[M_{*}(\lambda) - N_{*}(\lambda)]\\ &= R(\lambda,A^{*})^{-1}\left\{I - \int_{-r}^{0} e^{\lambda\theta}D^{*}(\theta)d\theta - \right.\\ &\quad \left. R(\lambda,A^{*})\left[\int_{-r}^{0} e^{\lambda\theta}d\eta^{*}(\theta) + \int_{-r}^{0} e^{\lambda\theta}A^{*}D^{*}(\theta)d\theta\right]\right\}\\ &= \lambda I - A^{*} - (\lambda I - A^{*})\int_{-r}^{0} e^{\lambda\theta}D^{*}(\theta)d\theta - \int_{-r}^{0} e^{\lambda\theta}d\eta^{*}(\theta) - \int_{-r}^{0} e^{\lambda\theta}A^{*}D^{*}(\theta)d\theta\\ &= \lambda I - A^{*} - \int_{-r}^{0}\lambda e^{\lambda\theta}D^{*}(\theta)d\theta - \int_{-r}^{0} e^{\lambda\theta}d\eta^{*}(\theta).\end{aligned} \tag{42}$$

Thus, ones have by virtue of (41) and (42) that

$$\hat{G}^{*}(\lambda) = \left[I - M_{*}(\lambda)^{-1}N_{*}(\lambda)\right]^{-1}M_{*}(\lambda)^{-1}R(\lambda,A^{*}),$$

which, together with (38), immediately implies that for sufficiently large $\mathfrak{R}(\lambda)$,

$$\hat{G}^{*}(\lambda) = \hat{G}_{*}(\lambda)$$

and then by the uniqueness of Laplace transforms,

$$G^{*}(t) = G_{*}(t), \quad t \in \mathbb{R}^{+},$$

almost everywhere. Since $G^{*}(t) = G_{*}(t) = O$ if $t < 0$, the desired result follows now.

Note that the adjoint $B^{*}(\theta)$, $D^{*}(\theta)$ of $B(\theta)$, $D(\theta)$ defined in (33) satisfy

$$B^{*}(\cdot),\ D^{*}(\cdot) \in L^{2}([-r,0];\mathscr{L}(X^{*})),$$

respectively. For arbitrary $\psi^* = (\psi_0^*, \psi_1^*) \in X^* \times W^{1,2}([-r,0];X^*)$, the mild solution $y^*(t,f^*,\psi^*)$ of the adjoint equation (33) exists and may be represented by

$$
\begin{aligned}
y^*(t,f^*,\psi^*) = {} & G^*(t)\psi_0^* + V_*(t,0)\psi_1^*(0) + \int_{-r}^0 U_*(t,\theta)\psi_1^*(\theta)d\theta \\
& + \int_{-r}^0 V_*(t,\theta)\psi_1^*(\theta)'d\theta + \int_0^t G^*(t-s)f^*(s)ds,
\end{aligned}
\tag{43}
$$

where

$$
U_*(t,\theta) = \int_{-r}^\theta G^*(t-\theta+\tau)d\eta^*(\tau) \in L^2\big([-r,0];\mathscr{L}(X^*)\big), \quad \theta \in [-r,0], \tag{44}
$$

and

$$
V_*(t,\theta) = \int_{-r}^\theta G^*(t-\theta+\tau)D^*(\tau)d\theta \in L^2\big([-r,0];\mathscr{L}(X^*)\big), \quad \theta \in [-r,0]. \tag{45}
$$

Note that the operators $U_*(t,\theta)$ and $V_*(t,\theta)$ in (44) and (45) are not necessarily identical with the adjoints $U^*(t,\theta)$ and $V^*(t,\theta)$ of $U(t,\theta)$ and $V(t,\theta)$, respectively. In particular, by an argument similar as for Proposition 1 ones can easily obtain the following theorem.

Theorem 4. *(i) Let $x^* \in X^*$, then $\int_0^t G^*(s)x^*ds \in \mathscr{D}(A^*)$ for almost all $t \in \mathbb{R}_+$, and the relation*

$$
\begin{aligned}
A^* \int_0^t G^*(s)x^*ds = {} & G^*(t)x^* - x^* - \int_0^t \int_{-r}^0 d\eta^*(\theta)G^*(s+\theta)x^*ds \\
& - \int_{-r}^0 D^*(\theta)G^*(t+\theta)x^*d\theta, \quad t \in \mathbb{R}_+, \quad x^* \in X^*,
\end{aligned}
\tag{46}
$$

holds almost everywhere.
(ii) Let $x^* \in \mathscr{D}(A^*)$, then $\int_0^t G^*(s)A^*x^*ds \in X^*$ for almost all $t \in \mathbb{R}_+$, and the commutative relation

$$
\begin{aligned}
\int_0^t G^*(s)A^*x^*ds = {} & G^*(t)x^* - x^* - \int_0^t \int_{-r}^0 G^*(s+\theta)d\eta^*(\theta)x^*ds \\
& - \int_{-r}^0 G^*(t+\theta)D^*(\theta)x^*d\theta, \quad t \in \mathbb{R}_+, \quad x^* \in \mathscr{D}(A^*),
\end{aligned}
\tag{47}
$$

holds almost everywhere.

5   Optimal Control

Let $T > 0$ and U be a separable Banach space. Consider the following neutral hereditary controlled system on X:

$$
\frac{d}{dt}\Big[y(t) - \int_{-r}^0 D(\theta)y(t+\theta)d\theta\Big] = Ay(t) + \int_{-r}^0 d\eta(\theta)y(\theta+t) + f(t) + Q(t)u(t),
\tag{48}
$$
$$
t \in [0,T],
$$
$$
y(0) = \phi_0, \ y_0 = \phi_1, \ \phi = (\phi_0,\phi_1) \in \mathscr{W}^{1,2}, \ u \in U_{ad},
$$

where A, η, D are given as in (4), $f \in L^2([0,T];X)$, $U_{ad} \subset L^2([0,T];U)$ and $Q \in L^{\infty}([0,T];\mathscr{L}(U,X))$.

The quantities $y(\cdot)$, $u(\cdot)$, Q and U_{ad} in (48) denote a system state, a control, a controller and a class of admissible controls, respectively. It is known by virtue of Theorem 2 that the following form

$$
\begin{aligned}
y(t) &= G(t)\left[\phi_0 - \int_{-r}^{0} D(\theta)\phi_1(\theta)d\theta\right] + \int_{-r}^{0} U(t,\theta)\phi_1(\theta)d\theta + \int_{-r}^{0} V(t,\theta)\phi_1'(\theta)d\theta \\
&\quad + \int_{0}^{t} G(t-s)f(s)ds + \int_{0}^{t} G(t-s)Q(s)u(s)ds \\
&= y(t,f,\phi) + \int_{0}^{t} G(t-s)Q(s)u(s)ds, \quad t \geq 0,
\end{aligned}
$$

(49)

is the mild solution of (48) where $y(t,f,\phi)$ is given by (11).

5.1 Existence of Optimal Control

In what follows, the admissible set U_{ad} is assumed to be closed and convex in $L^2([0,T];U)$. Let $J = J(u)$ be the integral convex cost given by

$$
J = R(y(T)) + \int_{0}^{T} \Big(P(y(t),t) + L(u(t),t)\Big)dt,
$$

(50)

where $R : X \to \mathbb{R}^1$, $P : X \times [0,T] \to \mathbb{R}^1$ and $L : U \times [0,T] \to \mathbb{R}^1$. We are interested in the following control problem on the finite interval $I = [0,T]$: find a control $u \in U_{ad}$ which minimizes the cost J subject to (48).

Assumption $\mathbf{A}_1$:

(1) $R : X \to \mathbb{R}^1$ is continuous and convex, and there exists a constant $c_0 > 0$ such that $R(x) \geq -c_0$ on X;
(2) $P : X \times [0,T] \to \mathbb{R}^1$ is measurable in $t \in [0,T]$ for each $x \in X$ and continuous, convex in $x \in X$ for $t \in [0,T]$, and there exists a constant $c_1 > 0$ such that $P(x,t) \geq -c_1$ on $X \times [0,T]$;
(3) $L : U \times [0,T] \to \mathbb{R}^1$ satisfies that for any $u \in U_{ad}$, $L(u(t),t)$ is integrable on $[0,T]$ and the functional $\Gamma : U_{ad} \to \mathbb{R}^1$ given by

$$
\Gamma(u) = \int_{0}^{T} L(u(t),t)dt
$$

is continuous and convex. Moreover, there exists a monotone increasing function $\theta_0 \in C(\mathbb{R}^+;\mathbb{R}^1)$ such that $\lim_{r \to \infty} \theta_0(r) = \infty$ and

$$
\Gamma(u) = \int_{0}^{T} L(u(t),t)dt \geq \theta_0(\|u\|_{L^2([0,T];U)}) \quad \text{for } u \in U_{ad}.
$$

Theorem 5. *Assume that the assumption* $\mathbf{A}_1$ *is satisfied. Then there exists a control* $u_0 \in U_{ad}$ *that minimizes the cost J in (50).*

Proof. Let $\{u_n\}$ be a minimizing sequence of J such that

$$\inf_{u \in U_{ad}} J(u) = \lim_{n \to \infty} J(u_n) = m_0.$$

By virtue of $\mathbf{A}_1$, it follows that

$$J(u) \geq \theta_0(\|u\|_{L^2([0,T];U)}) - c_0 - c_1 T \quad \text{for } u \in U_{ad}.$$

Hence, a standard argument with $\lim_{r \to \infty} \theta_0(r) = \infty$ yields that the minimizing sequence $\{u_n\}$ is bounded in $L^2([0,T];U)$, which, together with the closedness of U_{ad}, implies that there exists a subsequence (which we denote it again by $\{u_n\}$) of $\{u_n\}$ and a $u_0 \in U_{ad}$ such that

$$u_n \to u_0 \quad \text{weakly in} \quad L^2([0,T];U). \tag{51}$$

We denote $y^{u_n}(t)$ and $y^{u_0}(t)$ the mild solutions of (48) corresponding to u_n and u_0, respectively. For any fixed $x^* \in X^*$ and $t \in [0,T]$, since $G(t) = O$ if $t < 0$, then we have that for any $t \geq 0$,

$$\langle y^{u_n}(t), x^* \rangle_{X,X^*} = \langle y(t,f,\phi), x^* \rangle_{X,X^*} + \int_0^t \langle u_n(s), Q^*(s)G^*(t-s)x^* \rangle_{U,U^*} ds, \tag{52}$$

where $y(t,f,\phi)$ is the mild solution of (48) corresponding to $Q(\cdot) = O$. Since $Q(t) \in \mathcal{L}(U,X)$ and $G(t)$ is piecewise strongly continuous on $[0,T]$, it is easy to see that the function $Q^*(\cdot)G^*(t-\cdot)x^*$ belongs to $L^2([0,T];U^*)$. Hence, by virtue of (51) and (52), it follows that

$$\langle y^{u_n}(t), x^* \rangle_{X,X^*} \to \langle y(t,f,\phi), x^* \rangle_{X,X^*} + \int_0^t \langle u_0(s), Q^*(s)G^*(t-s)x^* \rangle_{U,U^*} ds$$

$$= \langle y(t,f,\phi), x^* \rangle_{X,X^*} + \left\langle \int_0^t G(t-s)Q(s)u_0(s)ds, x^* \right\rangle_{X,X^*} \tag{53}$$

$$= \langle y^{u_0}(t), x^* \rangle_{X,X^*} \quad \text{as } n \to \infty,$$

i.e.,

$$y^{u_n}(t) \to y^{u_0}(t) \quad \text{weakly in } X \text{ as } n \to \infty. \tag{54}$$

It is well known that continuity plus convexity imply weak lower semi-continuity. Then the condition (1) in Assumption $\mathbf{A}_1$ and (54) with $t = T$ imply

$$\lim_{n \to \infty} R(y^{u_n}(T)) \geq R(y^{u_0}(T)). \tag{55}$$

In a similar way, we have

$$\lim_{n \to \infty} P(y^{u_n}(t),t) \geq P(y^{u_0}(t),t), \qquad t \in [0,T]. \tag{56}$$

It follows via Fatou's lemma that

$$\lim_{n \to \infty} \int_0^T P(y^{u_n}(t),t)dt \geq \int_0^T \lim_{n \to \infty} P(y^{u_n}(t),t)dt \geq \int_0^T P(y^{u_0}(t),t), \qquad t \in [0,T]. \tag{57}$$

As for the term $\int_0^T L(u_n(t),t)dt$, it is clear from (3) in Assumption $\mathbf{A}_1$ that

$$\varliminf_{n\to\infty} \Gamma(u_n) \geq \Gamma(u_0) = \int_0^T L(u_0(t),t)dt. \tag{58}$$

Therefore, by (55), (57) and (58) we have

$$\begin{aligned}
m_0 = \inf_{u\in U_{ad}} J(u) &\geq \varliminf_{n\to\infty} R(y^{u_n}(T)) + \varliminf_{n\to\infty} \int_0^T P(y^{u_n}(t),t)dt + \varliminf_{n\to\infty} \Gamma(u_n) \\
&\geq R(y^{u_0}(T)) + \int_0^T \Big[P(y^{u_0}(t),t) + L(u_0(t),t) \Big] dt \\
&= J(u_0) > -\infty,
\end{aligned} \tag{59}$$

so that $m_0 = J(u_0)$. This proves that u_0 is the optimal solution for J.

5.2 Optimality Condition

In this subsection, we shall seek necessary optimality conditions of the optimal solution u for J in (50). The existence of optimal solutions is assumed but not the closedness of U_{ad}.

Assumption $\mathbf{A}_2$:

(1) $R : X \to \mathbb{R}^1$ is continuous and Gâteau differentiable, and the Gâteau derivative $R'(x) \in X^*$ for each $x \in X$;
(2) $P : X \times [0,T] \to \mathbb{R}^1$ is measurable in $t \in [0,T]$ for each $x \in X$ and continuous, convex on X for $t \in [0,T]$, and furthermore there exist functions $\partial_x P : X \times [0,T] \to X^*$, $\theta_1 \in L^1([0,T];\mathbb{R}^1)$, $\theta_2 \in C(\mathbb{R}^+;\mathbb{R}^1)$ such that:
  (a) $\partial_x P(x,t)$ is measurable in $t \in [0,T]$ for each $x \in X$ and continuous in $x \in X$ for $t \in [0,T]$ and the value $\partial_x P(x,t)$ is the Gâteau derivative of $P(x,t)$ in the first argument for $(x,t) \in X \times [0,T]$, and
  (b) $\|\partial_x P(x,t)\|_{X^*} \leq \theta_1(t) + \theta_2(\|x\|_X)$ for $(x,t) \in X \times [0,T]$;
(3) $L : U \times [0,T] \to \mathbb{R}^1$ is measurable in $t \in [0,T]$ for each $z \in U$ and continuous, convex on U for each $t \in [0,T]$. Moreover, there exist function $\theta_3(\cdot) \in L^1([0,T];\mathbb{R}^1)$ and constant $M > 0$ such that

$$|L(z,t)| \leq \theta_3(t) + M\|z\|_U^2 \qquad \text{for } (z,t) \in U \times [0,T].$$

Lemma 1. [12] *Consider the function*

$$J(v) = J_1(v) + J_2(v)$$

for any $v \in L^2([0,T];U)$ where we assume that the functions $J_i(v)$, $i = 1, 2$, are continuous and convex. Further assume that the function $v \to J_1(v)$ is differentiable. Then the unique element u in U_{ad} satisfying $J(u) = \inf_{v\in U_{ad}} J(v)$ is characterized by

$$J_1'(u)(v-u) + J_2(v) - J_2(u) \geq 0 \qquad \textit{for all} \ \ v \in U_{ad}. \tag{60}$$

Now we are in a position to state one of the main theorems in this section.

Theorem 6. *Suppose that the assumption* $\mathbf{A}_2$ *holds and* $u \in U_{ad}$ *is an optimal solution for J in (50). Then the integral inequality*

$$\int_0^T \langle v(t) - u(t), -Q^*(t)p(t) \rangle_{U,U^*} dt + \int_0^T \Big(L(v(t),t) - L(u(t),t) \Big) dt \geq 0$$

$$\text{for all } v \in U_{ad}$$

$$(61)$$

holds, where

$$p(t) = -G^*(T-t)R'(y^u(T)) - \int_t^T G^*(s-t)\partial_x P(y^u(s),s)ds \qquad (62)$$

satisfies that for $t \in [0,T]$,

$$\frac{d}{dt}\Big[p(t) - \int_{-r}^0 D^*(\theta)p(t+\theta)d\theta\Big] + A^*p(t) +$$

$$\int_{-r}^0 d\eta^*(\theta)p(t-\theta) - \partial_x P(y^u(t),t) = 0, \qquad (63)$$

$$p(T) = -R'(y^u(T)), \qquad p(t) = 0, \ t \in (T,T+r],$$

in the weak sense.

Proof. By virtue of the assumption $\mathbf{A}_2$, we have by Lebesgue's dominated convergence theorem that

$$(J - \Gamma)'(u)(v-u) = \Big\langle \int_0^T G(T-s)Q(s)(v(s)-u(s))ds, R'(y^u(T)) \Big\rangle_{X,X^*}$$

$$+ \int_0^T \Big\langle \int_0^s G(s-\tau)Q(\tau)(v(\tau)-u(\tau))d\tau, \partial_x P(y^u(s),s) \Big\rangle_{X,X^*} ds.$$

$$(64)$$

Note that all integrals in (64) are well defined by making use of the assumption $\mathbf{A}_2$. The first term of (64) can be written as

$$\int_0^T \langle v(s) - u(s), Q^*(s)G^*(T-s)R'(y^u(T)) \rangle_{U,U^*} ds. \qquad (65)$$

Using the standard Fubini lemma, the second term of (64) is transformed as

$$\int_0^T \int_0^s \langle G(s-\tau)Q(\tau)(v(\tau)-u(\tau)), \partial_x P(y^u(s),s) \rangle_{X,X^*} d\tau ds$$

$$= \int_0^T \langle v(\tau) - u(\tau), Q^*(\tau)\int_\tau^T G^*(s-\tau)\partial_x P(y^u(s),s)ds \rangle_{U,U^*} d\tau.$$

$$(66)$$

If we define $p(t)$ by (62) and apply Lemma 1 to the mapping $J = (J-\Gamma) + \Gamma$, then the relations (64), (65) and (66) yield the inequality (61). For the last statement, note that by virtue of (43), the function

$$-G^*(t)R'(y^u(0)) - \int_0^t G^*(t-s)\partial_x P(y^u(T-s),T-s)ds, \ t \in [0,T],$$

satisfies

$$\frac{d}{dt}\left[y^*(t) - \int_{-r}^{0} D^*(\theta)y^*(t+\theta)d\theta\right] = A^*y^*(t) + \int_{-r}^{0} d\eta^*(\theta)y^*(t+\theta)$$
$$- \partial_x P(y^u(T-t), T-t), \quad t \in [0,T], \tag{67}$$
$$y^*(0) = -R'(y^u(0)), \quad y^*(\theta) = 0, \quad \theta \in [-r,0).$$

A change of variable $t \to T-t$ yields the desired result. The proof is now complete. □

5.3   Maximum Principle

In view of Theorem 6, we can obtain from (61) the following "integral" maximum principle:

$$\max_{v \in U_{ad}} \int_0^T \left(\langle v(t), Q^*(t)p(t)\rangle_{U,U^*} - L(v(t),t)\right)dt =$$
$$\int_0^T \left(\langle u(t), Q^*(t)p(t)\rangle_{U,U^*} - L(u(t),t)\right)dt. \tag{68}$$

It is possible to improve this result to establish the so-called "pointwise" maximum principle for the convex cost (50). To this end, the assumption $\mathbf{A_2}$ is assumed at the moment. Let the admissible set U_{ad} be

$$U_{ad} = \left\{u \in L^2([0,T];U) : u(t) \in U(t), t \in [0,T]\right\}, \tag{69}$$

where the (time varying) control domain $U(t) \subset U, t \in [0,T]$, satisfies

Assumption $\mathbf{A_3}$:

(1) $U(t)$ is closed and convex in U for each $t \in [0,T]$;
(2) For any $t \in [0,T]$, $z \in \mathrm{Int}\, U(t)$, the interior of $U(t)$, there exists an $\varepsilon_0 > 0$ such that

$$z \in \left(\bigcap_{s \in (t,t+\varepsilon_0)} U(s)\right) \cup \left(\bigcap_{s \in (t-\varepsilon_0,t)} U(s)\right). \tag{70}$$

Theorem 7. *Let $u \in U_{ad}$ be an optimal solution for J in (50). Then*

$$\max_{z \in U(t)} \left\{\langle Q(t)z, p(t)\rangle_{X,X^*} - L(z,t)\right\} = \langle Q(t)u(t), p(t)\rangle_{X,X^*} - L(u(t),t), \quad t \in [0,T],$$

where $p(t)$ is given by (62).

Proof. Let $t \in [0,T]$ and $z \in \mathrm{Int}\, U(t)$. Since z satisfies (70), we suppose, for instance, $z \in \cap_{s \in (t,t+\varepsilon_0)} U(s)$. Then it is easy to see that for any $\varepsilon > 0$ the function

$$v_\varepsilon(s) = \begin{cases} u(s), & s \in [0,t) \text{ or } (t+\varepsilon, T], \\ z, & s \in [t,t+\varepsilon], \end{cases} \tag{71}$$

belongs to U_{ad} for the u in (61). Substituting v_ε for v in (61) and dividing the resulting inequality by ε, we obtain

$$\frac{1}{\varepsilon} \int_t^{t+\varepsilon} \left\{ \langle z - u(s), -Q^*(s)p(s) \rangle_{U,U^*} + \Big(L(z,s) - L(u(s),s) \Big) \right\} ds \geq 0. \quad (72)$$

Since all the integrands in (72) are integrable on $[0,T]$ by virtue of the assumption $\mathbf{A}_2$, the Lebesgue density theorem can apply. Then by letting $\varepsilon \to 0$ in (72), we have

$$\langle z, Q^*(t)p(t) \rangle_{U,U^*} - L(z,t) \leq \langle u(t), Q^*(t)p(t) \rangle_{U,U^*} - L(u(t),t), \quad t \in [0,T]. \quad (73)$$

Since the duality pairing $\langle z, Q^*(t)p(t) \rangle_{U,U^*}$ is continuous in z, we see from (73) that the maximum principle is true for such $t \in [0,T]$. $\qquad \square$

References

1. Ahmed, N.U., Teo, K.L.: Optimal Control of Distributed Parameter Systems. North-Holland, New York (1981)
2. Bernier, C., Manitius, A.: On semigroups in $R^n \times L^p$ corresponding to differential equations with delays. Canadian J. Math. 30, 897–914 (1978)
3. Bhat, K., Koivo, H.: Modal characterization of controllability and observability for time delay systems. IEEE Trans. Auto. Control AC-21, 292–293 (1976)
4. Datko, R.: Linear autonomous neutral differential equations in a Banach space. J. Differential Eqns. 25, 258–274 (1977)
5. Datko, R.: Representation of solutions and stability of linear differential-difference equations in a Banach space. J. Differential Eqns. 29, 105–166 (1978)
6. Delfour, M., Mitter, S.: Controllability, observability and optimal feedback control of affine hereditary differential systems. SIAM J. Control & Optim. 10, 298–327 (1972)
7. Diekmann, O., Gils, S., Lunel, S., Walther, H.: Delay Equations: Functional-, Complex- and Nonlinear Analysis. In: Applied Mathematical Sciences, vol. 110. Springer, New York (1995)
8. Hale, J.: Theory of Functional Differential Equations. Springer, Heidelberg (1977)
9. Hale, J., Lunel, S.: Introduction to Functional Differential Equations. Applied Mathematical Sciences, vol. 99. Springer, Heidelberg (1993)
10. Ito, K., Tarn, T.: A linear quadratic optimal control for neutral systems. Nonlinear Analysis 9, 699–727 (1985)
11. Jeong, J.: Stabilizability of retarded functional differential equation in Hilbert spaces. Osaka J. Math. 28, 347–365 (1991)
12. Lions, J.L.: Optimal Control of Systems Governed by Partial Differential Equations. Springer, Heidelberg (1971)
13. Nakagiri, S.: Optimal control of linear retarded systems in Banach spaces. J. Math. Anal. Appl. 120, 169–210 (1986)
14. Peszat, S., Zabczyk, J.: Stochastic Partial Differential Equations with Lévy Noise: An Evolution Equation Approach. In: Encyclopedia of Mathematics and its Applications. Cambridge Univ. Press, Cambridge (2007)
15. Salamon, D.: Control and Observation of Neutral Systems. Research Notes in Mathematics, vol. 91. Pitman Advanced Publishing Program, Boston (1984)

16. Tanabe, H.: Equations of Evolution. Pitman, New York (1979)
17. Travis, C., Webb, G.: Existence and stability for partial functional differential equations. Trans. Amer. Math. Soc. 200, 395–418 (1974)
18. Wu, J.: Semigroup and integral form of a class of partial differential equations with infinite delay. Differential and Integral Equations 6, 1325–1351 (1991)
19. Wu, J.: Theory and Applications of Partial Functional Differential Equations. In: Applied Mathematical Sciences, vol. 119. Springer, Berlin (1996)