# A penalty function proof of a Lagrange multiplier theorem with application to linear delay systems

**Fritz Colonius**

**Angaben zur Veröffentlichung / Publication details:**

Colonius, Fritz. 1981. "A penalty function proof of a Lagrange multiplier theorem with application to linear delay systems." *Applied Mathematics & Optimization* 7 (1): 309–34. https://doi.org/10.1007/bf01442124.

# A Penalty Function Proof of a Lagrange Multiplier Theorem with Application to Linear Delay Systems

Fritz Colonius

Forschungsschwerpunkt Dynamische Systeme, FB 4, Universität Bremen, Bibliothekstrasse, Postfach 330 440, D-2800 Bremen 33, Federal Republic of Germany

**Abstract.** Using a penalty function method, a Lagrange multiplier theorem in dual Banach spaces is proved. This theorem is applied to the optimal control of linear, autonomous time-delay systems with function space equality end condition and pointwise control restrictions. Under an additional regularity condition, the resulting Lagrange multiplier can be identified with an element of $L_\infty$.

## 1. Introduction

In this paper, penalty functions are used for proving necessary optimality conditions. Section 3 exhibits a proof of an abstract Lagrange multiplier theorem, while in section 4 this result is applied to an optimal control problem for linear autonomous time-lag systems with function space end condition and pointwise control restrictions.

Penalty function methods for a proof of necessary optimality conditions are intuitively appealing, since they are closely related to the interpretation of Lagrange multipliers as measuring the sensitivity against small deviations from the constraints. The idea was introduced in lecture notes by Courant [17] and was further developed by Beltrami [8]. In his ε-technique, Balakrishnan [1, 2, 3] used penalization in order to prove a maximum principle for optimal control problems. McShane [27] showed how to obtain the finite dimensional Kuhn–Tucker theorem for all optimal solutions, not only for limits of optimal points of the approximating problems. This was used by Berkovitz [7] for a proof of the maximum principle for optimal relaxed controls. His proof has been extended to hereditary differential systems with finite dimensional equality end condition by Bates [6].

This study was motivated by optimal control problems for hereditary differential systems with function space end condition. The crucial point in penalty function proofs is the convergence of the approximating Lagrange multipliers. McShane used the compactness of the unit sphere $\mathfrak{S} := \{x : |x| = 1\}$ in finite dimensions. Because of the consideration of finite dimensional end conditions, Berkovitz and Bates could use essentially the same argument. In the case of infinite dimensional end conditions (not necessarily of equality type), to which attention is focused here, this device is not applicable, since $\mathfrak{S}$ is not compact in infinite dimensional space. Instead, we impose an additional regularity condition that may be considered as a generalized Mangasarian–Fromowitz constraint qualification; then we use Alaoglu's theorem in order to establish existence of a weak* cluster point $l$ of the approximating Lagrange multipliers. The cluster point $l$ turns out to be a Lagrange multiplier for the original problem. The finite dimensional part of the implicit constraint is split off and treated in the manner introduced by McShane.

In the resulting Lagrange multiplier theorem it is not assumed that either the explicit constraint set or the cone used for the formulation of the implicit constraint contains interior points. This is opposed to standard Lagrange multiplier theorems (see, e.g. [25]). However, it is similar to a recent result by Zowe and Kurcyusz [33]. Zowe and Kurcyusz proved a generalized open mapping theorem and used a result by Robinson [29] on the relation between the sequential cone and the linearizing cone for their proof in general Banach spaces. The proof given in this paper also uses—at least in the general case—the open mapping theorem of Zowe and Kurcyusz. However, it does not rely on Robinson's result. Our assumptions are somewhat stronger than those of Zowe and Kurcyusz, since we need certain properties related to weak* topology. In turn, however, we gain the following:

(i) no regularity assumption concerning the finite dimensional part of the implicit constraint is required, and

(ii) the Lagrange multiplier is obtained "constructively" as a weak* cluster point of the approximate Lagrange multipliers.

The second part of this paper is concerned with the maximum principle for linear time-lag systems with function space end condition and pointwise control constraints. Here, in a first step, the general Lagrange multiplier theorem yields only a multiplier $l$ in $(W^{1,\infty}([-h,0],\mathbb{R}^n))^*$. Under an additional assumption $l$ can be identified with an element of $W^{1,\infty}([-h,0],\mathbb{R}^n)$. Further discussion of this and previous results are given in section 4.

## 2. Notation

Let $Z$ be a Banach space and let $Z^*$ denote its topological dual space. We define the symbol $\langle z^*, z\rangle_Z$ by $\langle z^*, z\rangle_Z := z^*(z)$, where the right-hand side is the value of the linear form $z^* \in Z^*$ at the point $z \in Z$. The weak topology on $Z$ is the weakest topology such that all maps $\langle z^*, \cdot\rangle_Z : Z \to \mathbb{R}$ are continuous; the weak* topology on $Z^*$ is the weakest topology such that all maps $\langle \cdot, z\rangle_Z : Z^* \to \mathbb{R}$ are continuous. For a map $F: Z_1 \to Z_2$, $Z_1, Z_2$ Banach spaces, $DF(z)$ denotes the Fréchet derivative of $F$ at $z$.

For $1 \leqslant p \leqslant \infty$, $W^{1,p}([t_0, t_1], \mathbb{R}^n)$ is the Banach space of absolutely continuous functions $x:[t_0, t_1] \rightarrow \mathbb{R}^n$, $t_0, t_1 \in \mathbb{R}$, $t_0 < t_1$, with derivative $\dot{x} \in L^p([t_0, t_1], \mathbb{R}^n)$ and $\|x\| := |(x(t_0), \|\dot{x}\|_{L^p})|$; $|\cdot|$ denotes the Euclidean norm in finite dimensional space.

For $t_0 < t_1 - h$, we abbreviate $T := [t_0, t_1]$, $T_0 := [t_0, t_1 - h]$, $T_1 := [t_1 - h, t_1]$.

## 3. Lagrange Multiplier Theorem in Dual Banach Spaces

First, we describe an abstract optimization problem. Then, simpler approximating problems where the implicit constraint is substituted by a penalty term are analyzed. Finally, a limiting procedure yields the desired Lagrange multiplier theorem. Consider the following:

*Problem 1:* Minimize

$$g_0(q) \tag{3.1}$$

s.t.

$$g(q) \in -K \subset \mathbb{R}^n \times Z \tag{3.2}$$

$$q \in Q \subset Y \tag{3.3}$$

where $Y$ and $Z$ are dual Banach spaces, $Z$ can densely and weakly* continuously be embedded into a Hilbert space $H$, $K = K_1 \times K_2$, $K_1 \subset \mathbb{R}^n$, $K_2 \subset Z$, $K$ is a weakly* closed convex cone, $Q$ is weakly* closed and convex, $\mathcal{Y} \supset Q$ is an open subset of $Y$, and $g_0: \mathcal{Y} \rightarrow \mathbb{R}$, $g = (g_1, g_2): \mathcal{Y} \rightarrow \mathbb{R}^n \times Z$ are continuously Fréchet differentiable, $g_0$ is weakly* lower semicontinuous, and $g$ is weakly* continuous.

The assumptions concerning properties in the weak* topology are usually satisfied in optimal control problems with relaxed controls or in problems where $L_\infty$-controls—taking values in a compact set—appear linearly. On the range space $Z$ of the implicit constraint (3.2) we need a Hilbert space structure in order to define a differentiable penalization. Therefore, $Z$ is assumed to be embedded into a Hilbert space $H$. The following identifications will be used throughout:

$$\mathbb{R}^n \times Z \subset \mathbb{R}^n \times H = (\mathbb{R}^n)^* \times H^* \subset \mathbb{R}^n \times Z^*.$$

In fact, only for the third part of this identification, which simplifies notation, we have assumed that $Z$ is *densely* embedded into $H$. Furthermore, the finite dimensional part $\mathbb{R}^n$ is split off, because it is much easier to handle. The regularity condition (not the nondegeneracy condition, however!), which we have to impose on the optimal solution, refers only to the infinite dimensional part.

Let $q^0$ be an optimal solution of Problem 1. We shall prove a necessary optimality condition for $q^0$.

Define the cone $D \subset \mathbb{R}^n \times H$ as the closure of $K$ with respect to the norm topology of $\mathbb{R}^n \times H$. Then $D = K_1 \times D_2$, where $D_2 \subset H$ is a convex and closed, hence weakly closed, cone in $H$. $D$ is a weakly closed convex cone in the Hilbert space $\mathbb{R}^n \times H$. Observe that the dual cone

$$D^* := \{y \in \mathbb{R}^n \times H : \langle y, d \rangle_{\mathbb{R}^n \times H} \geqslant 0 \text{ for all } d \in D\}$$

is contained in the dual cone $K^*$ of $K$.

The projections $h^D, h^{D^*}$ on $D$ respectively $D^*$ are defined for all $h \in \mathbb{R}^n \times H$ (see Wierzbicki/Kurcyusz [31, §2]). We assume throughout that the following consistency condition for $K_2$ and $D_2$ holds:

$$D_2 \cap Z = K_2.$$

This holds trivially, if $K = \{0\}$. It is also true, e.g., for the natural positivity cones $K_2$ in $L_\infty(T, \mathbb{R}^n)$ and $D_2$ in $L_2(T, \mathbb{R}^n)$. Define a penalty functional

$$P: \mathcal{Y} \times \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}$$

by

$$P(q, \varepsilon, K) := g_0(q) + K/2 \left\| (g(q))^{D^*} \right\|^2_{\mathbb{R}^n \times H} + \varepsilon \| q - q^0 \|_Y$$

and let

$$\mathcal{D}^\varepsilon := \{ q \in Q : \| q - q^0 \|_Y \leq \varepsilon \}.$$

**Lemma 3.1.** *For each $\varepsilon \in \mathbb{R}_+$, $\mathcal{D}^\varepsilon$ is convex and weakly* compact.*

*Proof.* Convexity of $\mathcal{D}^\varepsilon$ follows from convexity of $Q$. Weak* compactness is a consequence of Alaoglu's theorem [21, p. 37] taking into account that $Q$ is weakly* closed. □

Since $q^0$ is an optimal solution, $q^0 \in \mathcal{Y}$ and, because $\mathcal{Y}$ is open, there is $\varepsilon_1 > 0$ such that $\mathcal{D}^\varepsilon \subset \mathcal{Y}$ for all $0 < \varepsilon \leq \varepsilon_1$.

**Lemma 3.2.** *For $0 < \varepsilon \leq \varepsilon_1$ and $K \in \mathbb{R}_+$, each summand of $P(\cdot, \varepsilon, K): \mathcal{D}^\varepsilon \to \mathbb{R}$ is weakly* lower semicontinuous.*

*Proof.* For $g_0$ this is true by assumption. Furthermore, $g$ is weakly* continuous and the embedding $\mathbb{R}^n \times Z \to \mathbb{R}^n \times H$ is weakly* continuous; $\mathbb{R}^n \times H$ is a Hilbert space and the map $h \mapsto \| h^{D^*} \|^2_{\mathbb{R}^n \times H}$ is weakly lower semicontinuous by [31, Lemma 3.3i]. The assertion for the last summand follows by weak* lower semicontinuity of the norm in dual Banach space [32, Theorem V.1.9].

**Lemma 3.3.** *For $0 < \varepsilon \leq \varepsilon_1$, there is $K^\varepsilon > 0$ such that*

$$P(q, \varepsilon, K^\varepsilon) > g_0(q^0)$$

*for all $q \in \mathcal{D}^\varepsilon$ with $\| q - q^0 \|_Y = \varepsilon$.*

*Proof.* If the assertion is false, there exist $0 < \varepsilon \leq \varepsilon_1$ and sequences $K^i \to \infty$, $q^i \in \mathcal{D}^\varepsilon$, $i \in \mathbb{N}$, such that

$$\| q^i - q^0 \|_Y = \varepsilon \quad \text{and} \quad P(q^i, \varepsilon, K^i) \leq g_0(q^0),$$

i.e.,

$$g_0(q^i)-g_0(q^0)\leqslant -K^i/2\left\|\left(g(q^i)\right)^{D^*}\right\|^2-\varepsilon\|q^i-q^0\|$$
$$\leqslant 0. \tag{3.4}$$

Then by compactness of $\mathfrak{D}^\varepsilon$ and lower continuity of $g_0$ there is a cluster point $q^*\in\mathfrak{D}^\varepsilon$ of $(q^i)$ with

$$g_0(q^*)\leqslant \liminf g_0(q^i)\leqslant g_0(q^0).$$

Division of (3.4) by $-K^i/2$ yields in the limit

$$0\leqslant \limsup\left\|\left(g(q^i)\right)^{D^*}\right\|^2\leqslant 0,$$

since $g_0(q^i)$ is bounded.

Then by weak* lower continuity of the norm and weak* closedness of $Q$ we find that $q^*\in Q$ and $\|(g(q^*)^{D^*}\|=0$, i.e.,

$$g(q^*)\in -D\cap\{g(q): q\in Q\}\subset -D\cap(\mathbb{R}^n\times Z)$$

(see [31, Theorem 2.4]).

Using the consistency condition for $K_2$ and $D_2$, we find that $g(q^*)\in -K$. Thus $q^*$ satisfies the problem constraints, and since $q^0$ is optimal

$$g_0(q^*)\geqslant g_0(q^0);$$

however, (3.4) implies

$$g_0(q^*)-g_0(q^0)\leqslant -\varepsilon^2.$$

This is a contradiction. □

Now for $0<\varepsilon\leqslant\varepsilon_1$, we can define the

*ε-Problem:* Minimize $P(q,\varepsilon,K^\varepsilon)$

s.t. $q\in\mathfrak{D}^\varepsilon$.

**Theorem 3.1.** *For* $0<\varepsilon\leqslant\varepsilon_1$, *the ε-problem has a solution* $q^\varepsilon$.

*Proof.* Follows from Lemma 3.1 and Lemma 3.2. □

**Remark 3.1.** We have chosen $K^\varepsilon$ such that

$$\|q^\varepsilon-q^0\|_Y<\varepsilon.$$

Then necessary optimality conditions for $q^\varepsilon$ can be proved by arbitrary variations in $Q$. This is performed in the following Theorem 3.2.

**Theorem 3.2** (ε-Lagrange Multiplier Theorem. *There is* $l^\varepsilon \in D^* \subset \mathbb{R}^n \times H$ *such that*

(i) $[Dg_0(q^\varepsilon)+Dg(q^\varepsilon)^* l^\varepsilon][q-q^\varepsilon] \geqslant -\varepsilon \| q-q^0 \|_Y$ *for all* $q \in Q$;

(ii) $\langle l^\varepsilon, g(q^\varepsilon)\rangle_{\mathbb{R}^n \times H} \geqslant 0$.

*Proof.* For $q \in Q$ the Gateaux derivative of $P(\cdot, \varepsilon, K^\varepsilon)$ in $q^\varepsilon$ in direction $q-q^\varepsilon$ exists, is nonnegative and has the form

$$\left[Dg_0(q^\varepsilon)+K^\varepsilon Dg(q^\varepsilon)^*\left(g(q^\varepsilon)\right)^{D^*}\right][q-q^\varepsilon]+\varepsilon \frac{d^+}{d\theta}\gamma(\theta)\Big|_{\theta=0}, \tag{3.5}$$

where

$$\gamma(\theta) := \| q^\varepsilon - q^0 + \theta(q-q^\varepsilon)\|_Y, \qquad 0 \leqslant \theta \leqslant 1,$$

and

$$\frac{d^+}{d\theta}\gamma(\theta)\Big|_{\theta=0}$$

means the right-hand derivative of the convex function $\gamma$ in $\theta=0$; $g_0$ and $g$ are Fréchet differentiable by assumption; existence and form of the derivative of the projection on $D^*$ result from [31, Lemma 3.3iii].

Furthermore, by convexity of $\gamma$,

$$\begin{aligned}\frac{d^+}{d\theta}\gamma(\theta)\Big|_{\theta=0} &= \lim_{\theta\to 0^+}\{1/\theta[\gamma(\theta)-\gamma(0)]\}\\ &\leqslant \gamma(1)-\gamma(0)\\ &\leqslant \| q-q^0\|_Y.\end{aligned}$$

Define

$$l^\varepsilon := K^\varepsilon\left(g(q^\varepsilon)\right)^{D^*}$$

Then (i) holds and by [31, Lemma 2.3 and Theorem 2.4]

$$\begin{aligned}\langle l^\varepsilon, g(q^\varepsilon)\rangle &= \langle l^\varepsilon, g(q^\varepsilon)^{-D}\rangle + \langle l^\varepsilon, g(q^\varepsilon)^{D^*}\rangle\\ &= K^\varepsilon |g(q^\varepsilon)^{D^*}|^2\\ &\geqslant 0.\end{aligned}$$

□

**Corollary 3.1.** *There are* $(l_1^\varepsilon, l_2^\varepsilon) \in K_1^* \times D_2^* = D^* \subset \mathbb{R}^n \times H$ *such that*

(i) $[Dg_0(q^\varepsilon)+Dg_1(q^\varepsilon)^* l_1^\varepsilon + Dg_2(q^\varepsilon)^* l_2^\varepsilon][q-q^\varepsilon] \geqslant -\varepsilon \| q-q^0\|_Y$ *for all* $q \in Q$;

(ii) $\langle l_1^\varepsilon, g_1(q^\varepsilon)\rangle_{\mathbb{R}^n} \geqslant 0$ *and* $\langle l_2^\varepsilon, g_2(q^\varepsilon)\rangle_H \geqslant 0$.

*Proof.* Remember $g=(g_1, g_2)$ and define

$$l_1^\varepsilon := K^\varepsilon g_1(q^\varepsilon)^{K_1^*} \quad \text{and} \quad l_2^\varepsilon := K^\varepsilon g_2(q^\varepsilon)^{D_2^*}.$$

Then

$$\begin{aligned} l^\varepsilon &= K^\varepsilon g(q^\varepsilon)^{D^*} \\ &= K^\varepsilon\big[g_1(q^\varepsilon)^{K_1^*} + g_2(q^\varepsilon)^{D_2^*}\big] \\ &= l_1^\varepsilon + l_2^\varepsilon, \end{aligned}$$

and (i) follows by Theorem 3.2. (ii) follows as in the proof of Theorem 3.2. □

Next we formulate a regularity condition for the optimal solution. This condition requires local attainability in $Z$. It will allow the establishing of "convergence" of the approximate optimality conditions.

There exists a neighborhood $W$ of $0 \in Z$ such that

$$W \subset \big\{Dg_2(q^0)(q-q^0) : q \in Q\big\} + K_2\big(g_2(q^0)\big), \tag{3.6}$$

where for $z \in Z$

$$K_2(z) := \{\alpha(k_2 + z) : \alpha \geq 0,\ k_2 \in K_2\}.$$

**Lemma 3.4.** *If* (3.6) *holds, there exists a neighborhood $W_1$ of $0 \in Z$ such that for all $\varepsilon > 0$ small enough*

$$\begin{aligned} W_1 \subset &\big\{Dg_2(q^\varepsilon)(q-q^\varepsilon) : q \in Q,\ \|q-q^0\|_Y \leq 1\big\} \\ &+ \big(K_2 + g_2(q^\varepsilon)\big) \cap \{z \in Z : \|z\|_Z \leq 1\}. \end{aligned}$$

*Proof.* Obviously, (3.6) implies

$$Z = \big\{\alpha Dg_2(q^0)(q-q^0) : q \in Q,\ \alpha \geq 0\big\} + K_2\big(g_2(q^0)\big).$$

By a generalized open mapping theorem obtained from Zowe and Kurcyusz [33, Theorem 2.1], it follows that there is a neighborhood $W_2$ of $0 \in Z$ such that

$$\begin{aligned} W_2 \subset &\big\{Dg_2(q^0)(q-q^0) : q \in Q,\ \|q-q^0\|_Y \leq 1\big\} \\ &+ \big(K_2 + g_2(q^0)\big) \cap \{z \in Z : \|z\|_Z \leq 1\}. \end{aligned}$$

This is the assertion for $\varepsilon = 0$.

For the general assertion, it suffices to prove that for $\varepsilon \to 0$

$$\begin{aligned} d\Big(&\big\{Dg_2(q^\varepsilon)(q-q^\varepsilon) : q \in Q,\ \|q-q^0\|_Y \leq 1\big\}, \\ &\big\{Dg_2(q^0)(q-q^0) : q \in Q,\ \|q-q^0\|_Y \leq 1\big\}\Big) \to 0 \end{aligned}$$

and

$$d\Big(\big(K_2+g_2(q^\varepsilon)\big)\cap\{z\in Z:\|z\|_Z\leqslant 1\},$$
$$\big(K_2+g_2(q^0)\big)\cap\{z\in Z:\|z\|_Z\leqslant 1\}\Big)\to 0,$$

where $d$ is the Hausdorff metric for bounded closed sets in a normed space [19, p. 205].

But this is a consequence of $q^\varepsilon\to q^0$ in norm, the continuous Fréchet differentiability of $g$, and the convexity of $Q$ and $K_2$. □

**Remark 3.2.** Observe that in the proof above, application of the generalized open mapping theorem is not necessary, if $Q$ is bounded and $K=\{0\}$, or if $Dg(q^0)$ is weakly* continuous: In the latter case,

$$\{Dg_2(q^0)(q-q^0):q\in Q,\ \|q-q^0\|\leqslant 1\}$$

is closed and the assertion for $\varepsilon=0$ is a direct consequence of Baire's category theorem.

The following theorem is the main result of this section.

**Theorem 3.3** (Lagrange Multiplier Theorem). *If* (3.6) *holds for an optimal solution* $q^0$ *of Problem* 1, *there are* $l_0\leqslant 0$, $l=(l_1,l_2)\in\mathbb{R}^n\times Z$, *such that* $(l_0,l_1)\neq(0,0)$ *and*

(i) $l\in K^*=K_1^*\times K_2^*$;

(ii) $[-l_0Dg_0(q^0)+Dg(q^0)^*l][q-q^0]\geqslant 0$ *for all* $q\in Q$;

(iii) $\langle l_1,g_1(q^0)\rangle_{\mathbb{R}^n}=0$ *and* $\langle l_2,g_2(q^0)\rangle_Z=0$.

*If* (3.6) *is replaced by the stronger condition*: *There is a neighborhood* $V$ *of* $0\in\mathbb{R}^n\times Z$ *such that*

$$V\subset\{Dg(q^0)(q-q^0):q\in Q\}+K\big(g(q^0)\big),$$

*it follows that* $l_0\neq 0$.

*Proof.* Divide (i) and (ii) in Corollary 3.1 by $1+|l_1^\varepsilon|$ and define

$$\tilde l_0^\varepsilon:=-1/(1+|l_1^\varepsilon|) \tag{3.7}$$

$$\tilde l_1^\varepsilon:=l_1^\varepsilon/(1+|l_1^\varepsilon|) \tag{3.8}$$

$$\tilde l_2^\varepsilon:=l_2^\varepsilon/(1+|l_1^\varepsilon|) \tag{3.9}$$

$$\tilde l^\varepsilon:=l^\varepsilon/(1+|l_1^\varepsilon|)$$
$$=(\tilde l_1^\varepsilon,\tilde l_2^\varepsilon)\in D^*. \tag{3.10}$$

Then one obtains

$$\left[-\tilde{l}_0^\varepsilon Dg_0(q^\varepsilon)+Dg_1(q^\varepsilon)^*\tilde{l}_1^\varepsilon+Dg_2(q^\varepsilon)^*\tilde{l}_2^\varepsilon\right][q-q^\varepsilon] \geqslant -\varepsilon\|q-q^0\|/(1+|l_1^\varepsilon|), \tag{3.11}$$

$$\langle \tilde{l}_1^\varepsilon, g_1(q^\varepsilon)\rangle_{\mathbb{R}^n} \geqslant 0 \quad \text{and } \langle \tilde{l}_2^\varepsilon, g_2(q^\varepsilon)\rangle_H \geqslant 0. \tag{3.12}$$

Obviously, $|\tilde{l}_0^\varepsilon|\leqslant 1$ and $|\tilde{l}_1^\varepsilon|\leqslant 1$. Thus there are a sequence $\varepsilon_k \to 0$ and $l_0 \leqslant 0$, $l_1 \in \mathbb{R}^n$ such that

$$\lim \tilde{l}_0^{\varepsilon_k} = l_0 \tag{3.13}$$

$$\lim \tilde{l}_1^{\varepsilon_k} = l_1. \tag{3.14}$$

Then $(l_0, l_1) \neq (0,0)$: If $\tilde{l}_0^{\varepsilon_k} \to 0$, then $|\tilde{l}_1^{\varepsilon_k}| \to 1$. Since $l_1^{\varepsilon_k}$ lies in a finite dimensional space, we conclude $(l_0, l_1) \neq (0,0)$.

It follows from assumption (3.6), Lemma 3.4, and (3.11) that for all $z \in W$ and all $\varepsilon>0$ small enough there are $q \in Q$ with $\|q-q^0\|_Y \leqslant 1$ and $k_2 \in K_2$ with $\|k_2+g_2(q^\varepsilon)\|_Z \leqslant 1$, such that

$$\begin{aligned}\langle \tilde{l}_2^\varepsilon, z\rangle_Z &= \langle \tilde{l}_2^\varepsilon, Dg_2(q^\varepsilon)(q-q^\varepsilon)\rangle_Z + \langle \tilde{l}_2^\varepsilon, k_2+g_2(q^\varepsilon)\rangle_Z \\ &\geqslant -\varepsilon\|q-q^0\|_Y + \left(\tilde{l}_0^\varepsilon Dg_0(q^\varepsilon) - Dg_1(q^\varepsilon)^*\tilde{l}_1^\varepsilon\right)(q-q^\varepsilon) \\ &\quad + \langle \tilde{l}_2^\varepsilon, k_2\rangle_Z + \langle \tilde{l}_2^\varepsilon, g_2(q^\varepsilon)\rangle_Z \\ &\geqslant -\varepsilon + \left(\tilde{l}_0^\varepsilon Dg_0(q^\varepsilon) - Dg_1(q^\varepsilon)^*\tilde{l}_1^\varepsilon\right)(q-q^\varepsilon).\end{aligned} \tag{3.15}$$

Since the right-hand side is bounded, $\langle \tilde{l}_2^\varepsilon, z\rangle$ is bounded below. The same argument for $-z$ shows, that $\langle \tilde{l}_2^\varepsilon, z\rangle_Z$ is bounded. By Alaoglu's theorem [21, p. 37], $(\tilde{l}_2^{\varepsilon_k})$ has a weak* cluster point $l_2 \in Z^*$. Then $l := (l_1, l_2) \in K^*$, since $l$ is a weak* cluster point of $(\tilde{l}_1^{\varepsilon_k}, \tilde{l}_2^{\varepsilon_k}) \in D^* \subset K^*$ and $K^*$ is weakly* closed. This proves assertion (i).

Application of the following lemma to condition (i) of Theorem 3.2 proves assertion (ii) (define

$$x_k^* := \tilde{l}^{\varepsilon_k} \in R^n \times Z^*,$$

$$c_k := \left(-\varepsilon_k\|q-q^0\|_Y + l_0^{\varepsilon_k} Dg_0(q^{\varepsilon_k})\right)/\left(1+|l_1^{\varepsilon_k}|\right)\big).$$

**Lemma 3.5.** *Suppose a sequence $(x_k)$ in a Banach space $X$ converges to $x_0 \in X$ and a bounded sequence $(x_k^*)$ in the dual space $X^*$ has the weak* cluster point $x_0^*$. Let $(c_k) \subset \mathbb{R}$ converge to $c_0 \in \mathbb{R}$. Then*

$$\langle x_k^*, x_k\rangle_X \geqslant c_k, \qquad k \in \mathbb{N},$$

*implies*

$$\langle x_0^*, x_0\rangle_X \geqslant c_0.$$

The proof of this lemma is straightforward (see [13]) and therefore omitted.

Assertion (iii) follows by application of Lemma 3.5 to condition (ii) of Corollary 3.1.

The last assertion of the theorem follows by contradiction: If $l_0=0$, $l$ is nonnegative on a neighborhood of $0\in\mathbb{R}^n\times Z$. Thus $l=0$ contradicting the nontriviality condition $(l_0,l_1)\neq(0,0)$. □

**Remark 3.3.** We have shown that the Lagrange multiplier $(l_0,l)\in\mathbb{R}\times K^*$ is obtained as a weak* cluster point of the sequence $(\tilde{l}_0^{\varepsilon_k},\tilde{l}^{\varepsilon_k})\subset\mathbb{R}\times D^*\subset\mathbb{R}\times K^*$. If $Z$ is separable, this implies that there exists a subsequence converging to $(l_0,l)$.

**Remark 3.4.** For the nondegeneracy part of Theorem 3.3 (i.e., the condition for $l_0\neq 0$), it is not necessary to assume that the final space splits into a product $\mathbb{R}^n\times Z$.

**Remark 3.5.** The nondegeneracy part of Theorem 3.3 follows also—under weaker assumptions—from [33, Theorem 3.1]. This paper also contains a useful discussion on the relation between the nondegeneracy condition and boundedness of the set of Lagrange multipliers (see also [11]). This topic is important for stability results. Furthermore, Zowe and Kurcyusz show that the nondegeneracy condition is invariant under "small perturbations of the problem." The same arguments apply to the regularity condition (3.6).

**Remark 3.6.** Suppose that $Y$ and $Z$ are Hilbert spaces. Then one can choose $H=Z$ and the properties, with respect to weak* topology required in the formulation of Problem 1, coincide with the analogous properties with respect to weak topology.

**Remark 3.7.** A specific example of the situation mentioned in the last remark is the optimal control problem for linear functional differential systems with function space end condition and *energy constrained* controls treated in [16, Problem 3.2 and Theorem 3.1]. Here, $Y=L_2([t_0,t_1],\mathbb{R}^r)$, $Z=H=\mathcal{A}\subset W^{1,2}([t_0,t_1],\mathbb{R}^n)$, where $\mathcal{A}$ is the attainable subspace of the unconstrained system; $\mathcal{A}$ is assumed to be closed, i.e., $\mathcal{A}$ is a Hilbert space. The required weak continuity properties are satisfied since the state $x$ and the control $u$ appear linearly in the system equation and the performance index is convex (compare, however [16, Remark 3.1]). An application of Theorem 3.3 presupposes that the explicit constraint set $Q$ is weakly closed, which had not to be assumed in [16, Theorem 3.1]. At the other hand, the nondegeneracy condition in [16, Theorem 3.1] (that is assumption (e)) can be weakened by an application of Theorem 3.3: In the notation of this paper, assumption (e) means that there exists $\tilde{q}\in\operatorname{int}Q$ with

$$Dg(q^0)(\tilde{q}-q^0)=g(q^0)\tilde{q}=0$$

($K$ and the finite dimensional part of the equality constraint are trivial, and $g$ is affine). Since $Dg(q^0)$ is surjective onto $\mathcal{A}=Z$, the open mapping theorem implies that there is a neighborhood $W$ of $0\in\mathcal{A}$ with

$$W\subset\{Dg(q^0)(q-q^0):q\in Q\},$$

i.e., the condition for $l_0 \neq 0$ in Theorem 3.3 is satisfied. Obviously, this latter condition is weaker than assumption (e), since it does not presuppose $\operatorname{int} Q \neq \varnothing$.

**Remark 3.8.** At first sight, the weak* continuity properties required in Theorem 3.3 may appear very restrictive. However, consider the nonlinear version of the problem discussed in the last remark. One can prove that nonlinearity of the system equation in $x$ does not destroy the required weak continuity property (similarly as [7, Lemma 2]). Furthermore, the theory given in [16] had also to assume that the control $u$ appears linearly in the systems equation (see [16, Remark 3.2]).

**Remark 3.9.** The maximum principle for nonlinear, ordinary differential systems with relaxed controls follows from Theorem 3.3, since the results derived by Berkovitz in [7] for his proof of the maximum principle imply that the required weak* continuity properties are satisfied (in fact, his paper was a starting point for the proof given above).

A similar theory for relaxed *hereditary* differential systems with *function space end condition* can be developed on the basis of Theorem 3.3 [15].

**Remark 3.10.** It seems possible to apply Theorem 3.3 also to control problems with inequality constraints as considered by Buehler [12] and Mersky [28].

## 4. Optimal Control of Linear Delay Systems

In this section, the results obtained above are applied to linear autonomous delay systems with function space end condition and pointwise control restrictions.

First we state the problem formally. Then, as a corollary to Theorem 3.3, we obtain an abstract maximum principle with Lagrange multiplier $l$ in $(W^{1,\infty})^*$. This result is not satisfactory, since $l$ may not be identifiable with a real function. Only an additional regularity condition allows proof of the existence of a more regular Lagrange multiplier. This yields a pointwise concrete maximum principle with adjoint variable in $L_\infty$. An example shows that this regularity condition cannot be dispensed with.

The problem we consider has the following form:

*Problem 2:* Minimize

$$\int_{t_0}^{t_1} f(x(t), u(t))\, dt \tag{4.1}$$

subject to

$$\dot{x}(t) = A_1 x(t-h) + A_0 x(t) + B_0 u(t), \qquad \text{for a.e. } t \in T := [t_0, t_1] \tag{4.2}$$

$$x_{t_0} = \varphi_0 \tag{4.3}$$

$$x_{t_1} = \varphi_1 \tag{4.4}$$

$$u(t) \in \Omega(t) \qquad \text{a.e. } t \in T, \tag{4.5}$$

where $x_t(s):=x(t+s)$, $s\in[-h,0]$, and $h>0$, $f:\mathbb{R}^n\times\mathbb{R}^m\to\mathbb{R}$ continuously differentiable, $\phi\neq\Omega(t)\subset\mathbb{R}^m$ closed, convex and uniformly bounded, $t\mapsto\Omega(t)$ measurable in the Hausdorff metric, $A_0$, $A_1$, $B_0$ matrices of appropriate dimensions, $\varphi_0:[-h,0]\to\mathbb{R}^n$ an integrable function with $\varphi_0(0)=0$ and $\varphi_1\in W^{1,1}([-h,0],\mathbb{R}^n)$.

$h$ denotes the length of the system memory and $x_t$ is the state of the system at time $t$. For a fixed control function $u$, the future development of the system is not uniquely determined by $x(t)$, but by the function segment $x_t$. Thus the function space end condition (4.4) appears appropriate, if the behavior of the system after $t_1$ is of interest.

Because of their infinite dimensional character, problems with function space end condition present great difficulties (see [4, 5]). Concerning the important case of pointwise control restrictions, [9, 10, 14], and [15] contain maximum principles with nontriviality of the adjoint variable guaranteed. However, certain regularity conditions had to be assumed which imply for the considered Problem 2 that Rank $B_0=n$. In the sequel, this condition will be considerably weakened.

Suppose that $(x^0,u^0)$ is an optimal solution of Problem 2. We shall prove a (necessary) optimality condition for $(x^0,u^0)$. For an application of the theory developed in section 3, the following notations are convenient.

Let $2\leq p\leq\infty$, and $\underline{W}^{1,p}(T,\mathbb{R}^n):=\{x\in W^{1,p}(T,\mathbb{R}^n)\colon x(t_0)=0\}$. Define

$F\colon \underline{W}^{1,p}(T,\mathbb{R}^n)\times L_p(T,\mathbb{R}^m)\to\mathbb{R}$

$$F(x,u):=\int_T f(x(t),u(t))\,dt; \tag{4.6}$$

$A\colon \underline{W}^{1,p}(T,\mathbb{R}^n)\to\underline{W}^{1,p}(T,\mathbb{R}^n)$

$$(Ax)(t):=\int_{t_0}^{t}[A_1x(s-h)+A_0x(s)]\,ds,\qquad t\in T, \tag{4.7}$$

where we define for $x\in\underline{W}^{1,p}(T,\mathbb{R}^n)$

$$x(t_0+s):=\varphi_0(s),\qquad s\in[-h,0];$$

$B\colon L_p(T,\mathbb{R}^m)\to\underline{W}^{1,p}(T,\mathbb{R}^n)$,

$$(Bu)(t):=\int_{t_0}^{t}B_0u(s)\,ds,\qquad t\in T; \tag{4.8}$$

$C\colon \underline{W}^{1,p}(T,\mathbb{R}^n)\to W^{1,p}([-h,0],\mathbb{R}^n)$

$$Cx:=x_{t_1}-\varphi_1; \tag{4.9}$$

$$\mathfrak{U}_{ad}:=\{u\in L_p(T,\mathbb{R}^m)\colon u(t)\in\Omega(t)\text{ a.e. }t\in T\}. \tag{4.10}$$

$B$ is continuous and linear, $A$ and $C$ are affine. Thus the Fréchet derivatives $DA$ and $DC$ of $A$ and $C$, respectively, exist. $DA$ is a compact linear operator with

Kernel $(Id-DA)=\{0\}$. Then

$$x(u) = (Id-A)^{-1}Bu \tag{4.11}$$

is well defined and associates with each control function $u$ the corresponding trajectory $x(u)$ of (4.2) with initial condition (4.3). $(Id-DA)^{-1}Bu$ is the trajectory of (4.2) corresponding to the control $u$ and initial condition $x_{t_0}=0$.

Define the attainable subspace of the unconstrained system (4.2) with initial condition $x_{t_0}=0$ by

$$\mathfrak{A}_p := \Big\{\varphi\in W^{1,p}([-h,0],\mathbb{R}^n)\colon \text{there is } u\in L_p(T,\mathbb{R}^m) \text{ such that } \varphi=DC(Id-DA)^{-1}Bu\Big\}. \tag{4.12}$$

Let

$$\pi_1\mathfrak{A}_p := \{\alpha\in\mathbb{R}^n\colon \text{There is } \varphi\in\mathfrak{A}_p \text{ with } \alpha=\varphi(-h)\}$$

$$\pi_2\mathfrak{A}_p := \{z\in L_p([-h,0],\mathbb{R}^n)\colon \text{There is } \varphi\in\mathfrak{A}_p \text{ with } z=\dot\varphi\}.$$

Then, using the identification

$$W^{1,p}([-h,0],\mathbb{R}^n) \cong \mathbb{R}^n\times L_p([-h,0],\mathbb{R}^n)$$

$$\varphi\mapsto(\varphi(-h),\dot\varphi),$$

we have

$$\mathfrak{A}_p \subset \pi_1\mathfrak{A}_p\times\pi_2\mathfrak{A}_p.$$

Now Problem 2 can be rephrased as

*Problem 2′:* Minimize $F(x(u),u)$

$$\text{s.t.}\quad C[Id-A]^{-1}Bu = 0$$

$$u\in\mathfrak{U}_{\text{ad}}.$$

Here we consider $C[Id-A]^{-1}B$ as a mapping defined on $L_p(T,\mathbb{R}^m)$ with values in $\pi_1\mathfrak{A}_p\times\pi_2\mathfrak{A}_p$. This is possible, since this mapping is affine and $\varphi_1=x^0_{t_1}$.

For an application of the results obtained in section 3, let

$$Y := L_\infty(T,\mathbb{R}^m),\qquad Z := \pi_2\mathfrak{A}_p,\qquad H := \pi_2\mathfrak{A}_2,\qquad K := \{0\}, \tag{4.13}$$

$$g_0(u) := F(x(u),u) \tag{4.14}$$

$$g=(g_1,g_2) := C[Id-A]^{-1}B\colon L_p(T,\mathbb{R}^m)\to\pi_1\mathfrak{A}_p\times\pi_2\mathfrak{A}_p, \tag{4.15}$$

$$Q := \mathfrak{U}_{\text{ad}}. \tag{4.16}$$

It remains to specify $2 \leqslant p \leqslant \infty$. For this purpose, consider condition (3.6). For Problem 2′ (3.6) takes the following form:

There exists a neighborhood $W$ of $0 \in \pi_2 \mathcal{C}_p$ with

$$\begin{aligned} \dot{\varphi}_1 + W \subset & \{ (A_1 x(u)(t_1+s-h) + A_0 x(u)(t_1+s) \\ & + B_0 u(t_1+s), s \in [-h,0] ) : u \in \mathcal{U}_{ad} \} \\ = & \{ \dot{x}(u)_{t_1} : u \in \mathcal{U}_{ad} \}. \end{aligned} \tag{4.17}$$

Since the functions at the right-hand side are essentially bounded, this condition can only be satisfied for $p = \infty$. Thus the theory will be developed for $p = \infty$.

We need that $\pi_2 \mathcal{C}_\infty$ is a Banach space and $\pi_2 \mathcal{C}_2$ is a Hilbert space. The following theorem summarizes some known facts on the attainability subspaces of linear time-delay systems. In particular, this theorem gives a useful algebraic characterization for the closedness of $\mathcal{C}_p$ in $W^{1,p}([-h,0],\mathbb{R}^n)$.

**Lemma 4.1.** *Via* (4.13)–(4.16), *problem* 2′ *is a special case of Problem* 1, *provided that* $\mathcal{C}_\infty$ *is closed in* $W^{1,\infty}([-h,0],\mathbb{R}^n)$ *and the following condition is satisfied:*

(i) $\mathcal{C}_p$ *is closed in* $W^{1,p}([-h,0],\mathbb{R}^n)$ *for any* $1 \leqslant p \leqslant \infty$;
(ii) $A_1 A_0^i \mathcal{B} \subset \mathcal{B}$ *for* $i = 0, 1, \ldots, n-1$;
(iii) *For any* $1 \leqslant p \leqslant \infty$ $\varphi \in \mathcal{C}_p$ *iff there exists* $u \in L_p([-h,0],\mathbb{R}^m)$ *such that*

$$\dot{\varphi}(t) = A_0 \varphi(t) + B_0 u(t) \text{ for a.e. } t \in [-h,0]$$

$$\varphi(-h) \in \operatorname{span}\{ A_0^i \mathcal{B}, i = 0, 1, \ldots, n-1 \}.$$

*If* (i) *holds,* $\pi_2 \mathcal{C}_p$ *is closed in* $L_p([-h,0],\mathbb{R}^n)$ *for any* $1 \leqslant p \leqslant \infty$.

*Proof.* The proof follows from [24, Corollaries 1 and 2]. □

Thus in the following we assume that $\mathcal{C}_\infty$ is closed in $W^{1,\infty}([-h,0],\mathbb{R}^n)$. Then $\pi_2 \mathcal{C}_\infty$ is a Banach space and $\pi_2 \mathcal{C}_2$ is a Hilbert space.

$(x^0, u^0)$ is an optimal solution of Problem 2′. The sets $\mathcal{D}^\varepsilon$ and the functionals $P$ are defined as in section 3.

**Lemma 4.1.** *Via* (4.13)–(4.16), *problem* 2′ *is a special case of Problem* 1, *provided that* $\mathcal{C}_\infty$ *is closed in* $W^{1,\infty}([-h,0],\mathbb{R}^n)$ *and the following condition is satisfied*:

$$\text{The functional } F(x(\cdot),\cdot) \text{ is weakly* lower semicontinuous on } \mathcal{D}^{\varepsilon_1} \subset L_\infty(T,\mathbb{R}^m). \tag{4.18}$$

*Proof.* It is easily seen that $g$ is weakly* continuous. It only remains to prove that $\mathcal{U}_{ad} = \{ u \in L_\infty(T,\mathbb{R}^m) : u(t) \in \Omega(t) \text{ a.e. } t \in T \}$ is weakly* closed. Since $\Omega(\cdot)$ is measurable, the map

$$t \mapsto \Omega(t) \cap \mathrm{cl}(C\Omega(t)) =: \delta\Omega(t)$$

is measurable, where $C\Omega(t)$ denotes the complement of $\Omega(t)$ in a fixed compact set $\Omega_0$ in $\mathbb{R}^m$ containing $\Omega(t)$ for a.e. $t \in T$.

By Castaing's theorem [30, Theorem I.7.8], there exists a denumerable collection $(\theta_i)$ of measurable selections of $\delta\Omega(\cdot)$ such that $\{\theta_i(t)\}$ is dense in $\delta\Omega(t)$ for a.e. $t\in T$.

Define for $i\in\mathbb{N}$ and $t\in T$

$$\Gamma_i(t) := \{y\in\mathbb{R}^m : |y|=1, y\omega \leq y\theta_i(t) \text{ f.a. } \omega\in\Omega(t)\}.$$

By the Filippov–Castaing theorem [30, Theorem I.7.10] the $\Gamma_i$ are measurable functions. Then, again by Castaing's theorem, there are denumerable collections $(\xi_{ij})_{j\in\mathbb{N}}$ of measurable selections of $\Gamma_i$ such that $\{\xi_{ij}(t) : j\in\mathbb{N}\}$ is dense in $\Gamma_i(t)$ for a.e. $t\in T$ and all $i\in\mathbb{N}$.

Since $\Omega(t)$ is closed and convex,

$$\Omega(t) = \bigcap_{i,j\in\mathbb{N}} \{\omega\in\mathbb{R}^m : \omega[\xi_{ij}(t)-\theta_i(t)]\leq 0\}, \qquad t\in T.$$

Thus

$$\begin{aligned}\mathcal{U}_{\text{ad}} &= \bigcap_{i,j\in\mathbb{N}} \{u\in L_\infty(T,\mathbb{R}^m) : u(t)[\xi_{ij}(t)-\theta_i(t)]\leq 0 \text{ a.e. } t\in T\}\\ &= \bigcap_{i,j\in\mathbb{N}} \Big\{u\in L_\infty(T,\mathbb{R}^m) : \int_T f(t)u(t)[\xi_{ij}(t)-\theta_i(t)]\,dt\leq 0\\ &\qquad \text{f.a. } f\in L_1(T,\mathbb{R}) \text{ with } f\geq 0\Big\}\end{aligned}$$

Thus $\mathcal{U}_{\text{ad}}$ as an intersection of weakly* closed sets is weakly* closed. □

**Remark 4.1.** By Alaoglu's theorem, $\mathcal{U}_{\text{ad}}$ is even weakly* compact.

We have the following sufficient condition for the validity of (4.18).

**Theorem 4.2.** *Suppose* $F(x,\cdot) : \mathcal{U}_{\text{ad}}\to\mathbb{R}$ *is weakly* lower semicontinuous. Then condition* (4.18) *is satisfied. In particular, this is true if* $f(y,\cdot)$ *is convex for* $y\in\mathbb{R}^n$.

The proof of this theorem uses the same arguments as [20, Satz 2]. See also [13, Satz 5.1].

The $\varepsilon$-problems are defined in the same way as in section 3. They have optimal solutions satisfying necessary optimality conditions as in Theorem 3.2.

If we want to establish convergence of these approximate optimality conditions, we have to assume that the attainability condition (3.6) in Theorem 3.3 is satisfied. For Problem 2, this condition requires local attainability with respect to velocity:

There exists a neighborhood $W$ of $\dot{\varphi}_1\in\pi_2\mathcal{C}_\infty$, such that for all $\xi\in W$ there is a trajectory $x$ of (4.2) with initial condition (4.3) and control $u\in\mathcal{U}_{\text{ad}}$ such that

$$\xi(t) = \dot{x}(t+t_1), \qquad \text{a.e. } t\in[-h,0]. \tag{4.19}$$

Under this assumption, Theorem 3.3 establishes the existence of nontrivial Lagrange multipliers $(l_0, l_1, l_2)$. However, $l_2$ may be of a very complicated form, since it is only known to lie in $(\pi_2\mathcal{Q}_\infty)^*$, which is isomorphic to a factor space of $(L_\infty)^*$. Therefore, further analysis is required, and we have to impose the following stronger assumption.

$$B_0 u^0(t) \in \operatorname{int}^\delta_{\mathfrak{B}} B_0\Omega(t)$$
$$\text{for a.e. } t \in T_1 \text{ and some fixed constant } \delta > 0; \tag{4.20}$$

here $\operatorname{int}^\delta_{\mathfrak{B}} B_0\Omega(t)$ denotes the set of all interior points of $\{B_0\omega: \omega \in \Omega(t)\}$ relative to $\mathfrak{B} = \operatorname{Im} B_0$ having at least distance $\delta$ to the relative boundary of $B_0\Omega(t)$.

**Remark 4.2.** If $\operatorname{rank} B_0 = n$, then one can show that condition (4.20) implies condition (4.19). However, in general this implication is not valid. This can be seen, e.g., by analyzing [22, Example 1].

Conversely, also (4.19) does not imply (4.20) (see the counterexample at the end of this section).

The next theorem is a pointwise maximum principle for Problem 2 and represents the main result of this section.

**Theorem 4.3.** *Assume that* $(x^0, u^0)$ *is an optimal solution of Problem* 2, *that* (4.18), (4.19), *and* (4.20) *are satisfied and* $\mathcal{Q}_\infty$ *is closed in* $W^{1,\infty}([-h,0],\mathbb{R}^n)$. *Then there are* $l_0 \leqslant 0$, $\psi \in L_\infty(T,\mathbb{R}^n)$, *such that* $\psi|T_0$ *is absolutely continuous*, $(0,0,0) \neq (l_0, \psi(t_1-h), \psi|T_1) \in \mathbb{R}\times\mathbb{R}^n\times L_\infty(T_1,\mathbb{R}^n)$, *and*

$$\text{(i)}\quad \dot{\psi}(t) = l_0 D_1 f\big(x^0(t), u^0(t)\big) - A_1^*\psi(t+h) - A_0^*\psi(t)$$

*for a.e.* $t \in T_0$;

$$\text{(ii)}\quad \big[l_0 D_2 f\big(x^0(t), u^0(t)\big) - \psi(t) B_0\big]\big[u^0(t) - \omega\big] \geqslant 0$$

*for all* $\omega \in \Omega(t)$ *and a.e.* $t \in T$. *If, additionally, a neighborhood of* $\varphi_1$ *in* $\mathcal{Q}_\infty$ *is completely attainable,* $l_0$ *can be chosen as* $-1$.

*Proof.* By Lemma 4.1 the assumptions of Theorem 3.3 are satisfied. Hence there are $l_0 \leqslant 0$, $l = (l_1, l_2) \in (\pi_1\mathcal{Q}_\infty)^* \times (\pi_2\mathcal{Q}_\infty)^*$ such that $(l_0, l_1) \neq (0,0)$ and

$$\begin{aligned}
&\big[-l_0 D_1 F(x^0, u^0)(Id - DA)^{-1}B - l_0 D_2 F(x^0, u^0) + l\cdot DC(Id - DA)^{-1}B\big]\\
&[u - u^0] = -l_0 D_1 F(x^0, u^0) z(u - u^0) - l_0 D_2 F(x^0, u^0)[u - u^0]\\
&\qquad\qquad + l_1 z(u - u^0)(t_1 - h) + l_2\big(\dot{z}(u - u^0)_{t_1}\big)\\
&\qquad \geqslant 0,
\end{aligned} \tag{4.21}$$

where $z(u - u^0)$ is defined by

$$\begin{aligned}
&\dot{z}(t) = A_0 z(t) + A_1 z(t-h) + B_0\big(u(t) - u^0(t)\big), \qquad \text{a.e. } t \in T,\\
&z_{t_0} = 0.
\end{aligned} \tag{4.22}$$

Let

$$\pi_2\mathcal{Q}_p^0 := \{z \in L_p([-h,0],\mathbb{R}^n):\ \text{There is } \varphi \in \mathcal{Q}_p \text{ with } z = \dot{\varphi} \text{ and } \varphi(-h) = 0\}, \qquad p = 1, \infty.$$

We show first that there exists a dense subspace $E_\infty$ of $\pi_2\mathcal{Q}_\infty^0$ such that $l_2 | E_\infty$ is continuous with respect to $L_1$-norm on $E_\infty$. Then it can be extended to a continuous linear functional on $\pi_2\mathcal{Q}_\infty^0$ in $L_1$-norm. Since $E_\infty$ is dense in $L_\infty$-norm, this extension coincides with $l_2$ on $\pi_2\mathcal{Q}_\infty^0$. By Theorem 4.1(iii), $\pi_2\mathcal{Q}_\infty = \pi_2\mathcal{Q}_\infty^0 + V$, where $V$ is a finite dimensional vector space. Since all norms on finite dimensional vector spaces are equivalent, $l_2$ is also continuous on $\pi_2\mathcal{Q}_\infty$ in $L_1$-norm and hence can be identified with a function $\rho \in L_\infty([-h,0],\mathbb{R}^n)$:

$$\langle l_2, \xi \rangle = \int_{-h}^{0} \rho(t)\xi(t)\,dt \qquad \text{for all } \xi \in \pi_2\mathcal{Q}_\infty. \tag{4.23}$$

We first construct $E_\infty$. Consider the dense subspace $S$ of simple functions in $L_\infty(T_1, \mathfrak{B})$. For $p = 1, \infty$ define

$$\xi_p: L_p(T_1, \mathfrak{B}) \to \pi_2\mathcal{Q}_p^0 \tag{4.24}$$

as the continuous map associating with each $y \in L_p$ the function segment $\dot{x}_{t_1}$ where $x$ is the unique solution of

$$\dot{x}(t) = A_0 x(t) + y(t) \qquad \text{a.e. } t \in T_1,$$
$$x_{t_1 - h} = 0.$$

Then one sees easily that $\xi_p$ is an isomorphism and that

$$E_p := \xi_p(S)$$

is dense in $\pi_2\mathcal{Q}_p^0$ for $p = 1, \infty$. For $e \in E_\infty$ there is a unique $s \in S$ with

$$e = \xi_\infty(s) = \xi_1(s).$$

We can write $s$ in the form

$$s(t) = \sum_{i=1}^{k} \sum_{j=1}^{N} s_{ij} \chi_{A_i}(t) y_j(t), \qquad t \in T_1,$$

where $N$ is the dimension of $\mathfrak{B}$, $\{A_j\}$ is a measurable decomposition of $T_1$, $s_{ij} \in \mathbb{R}$, and $y_j: T_1 \to \mathfrak{B} = \mathbb{R}^N$ are constant functions having value 0 in all components $y_{jl}$ for $j \neq l$ and $y_{jj} > 0$.

We can choose $y_j$ such that $|y_j(t)|<\delta$, with $\delta$ as in (4.20). Thus there are by [30, Theorem I.7.10] $u_j^{\pm}\in\mathcal{U}_{\text{ad}}$ such that for a.e. $t\in T_1$

$$y_j(t) = B_0\left[u_j^+(t)-u^0(t)\right]$$

$$-y_j(t) = B_0\left[u_j^-(t)-u^0(t)\right].$$

Let $s_{ij}^{\pm}:=\max(0,\pm s_{ij})$. Then

$$s(t) = \sum_{i=1}^{k}\sum_{j=1}^{N}\chi_{A_i}(t)\left[s_{ij}^+ y_j(t)-s_{ij}^- y_j(t)\right],$$

and since $\xi_1$ is an isomorphism,

$\|e\|_{L_1}\to 0$ implies for $j=1,\dots,N$

$$\sum_{i=1}^{k}\lambda(A_i)\left(s_{ij}^+ + s_{ij}^-\right)\to 0. \tag{4.25}$$

Define for $i=1,\dots,k$, $j=1,\dots,N$ $w_{ij}^{\pm}\in\mathcal{U}_{\text{ad}}$ by

$$w_{ij}^{\pm}(t):=\begin{cases} u_j^{\pm}(t) & \text{for } t\in A_i\\ u^0(t) & \text{for } t\in T\setminus A_i.\end{cases}$$

Then

$$\begin{aligned}\langle l_2,e\rangle &= \langle l_2,\xi_1(s)\rangle\\ &= \sum_{j=1}^{N}\sum_{i=1}^{k}s_{ij}^+(l_2\cdot\xi_1)\left(B_0\left(w_{ij}^+-u^0\right)\right)\\ &\quad+\sum_{j=1}^{N}\sum_{i=1}^{k}s_{ij}^-(l_2\cdot\xi_1)\left(B_0\left(w_{ij}^--u^0\right)\right).\end{aligned}$$

By definition (4.24) of $\xi_1$

$$\xi_1\left(B_0\left(w_{ij}^{\pm}-u^0\right)\right) = \left(\dot z\left(w_{ij}^{\pm}-u^0\right)\right)_{t_1},$$

where $z$ is defined as in (4.22).

The variation of constants formula implies

$$\|z\left(w_{ij}^{\pm}-u^0\right)\|_\infty\leqslant c_0\lambda(A_i) \tag{4.26}$$

for a constant $c_0>0$ which is independent of $e$.

Apply (4.21) $2Nk$ times in order to obtain

$$\begin{aligned}
&\langle l_2, e\rangle \\
&= \sum_{j=1}^{N}\sum_{i=1}^{k}\left[s_{ij}^{+}\langle l_2, \dot z\left(w_{ij}^{+}-u^0\right)_{t_1}\rangle + s_{ij}^{-}\langle l_2, z\left(w_{ij}^{-}-u^0\right)_{t_1}\rangle\right] \\
&\geqslant -\sum_{j=1}^{N}\sum_{i=1}^{k} s_{ij}^{+}\left\{l_0 D_1F(x^0,u^0)z\left(w_{ij}^{+}-u^0\right)+l_0D_2F(x^0,u^0)\left(w_{ij}^{+}-u^0\right)\right. \\
&\qquad \left.+l_1 z\left(w_{ij}^{+}-u^0\right)(t_1-h)\right\} \\
&\quad -\sum_{j=1}^{N}\sum_{i=1}^{k} s_{ij}^{-}\left\{l_0 D_1F(x^0,u^0)z\left(w_{ij}^{-}-u^0\right)+l_0D_2F(x^0,u^0)\left(w_{ij}^{-}-u^0\right)\right. \\
&\qquad\qquad \left.+l_1 z\left(w_{ij}^{-}-u^0\right)(t_1-h)\right\} \\
&\geqslant -c_1\sum_{j=1}^{N}\sum_{i=1}^{k}\left(s_{ij}^{+}+s_{ij}^{-}\right)\lambda(A_i).
\end{aligned}$$

for a constant $c_1>0$. This follows from (4.26) and the properties of $D_2F$ (cp. (4.31), below).

By (4.19) this last expression converges to 0 for $\|e\|_{L_1}\to 0$.

The same arguments for $-e$ prove that $l_2(e)\to 0$ for $\|e\|_{L_1}\to 0$.

Hence, $l_2$ can be identified with $\rho\in L_\infty([-h,0],\mathbb{R}^n)$ satisfying (4.23). $l_1$ is identified with an element of $\mathbb{R}^n$.

Define $y\in(W^{1,\infty}(T,\mathbb{R}^n))^*$ by

$$y := (Id-DA^*)^{-1}\left(DC^*l+l_0D_1F(x^0,u^0)\right).$$

Then by (4.21)

$$\langle l_0D_2F(x^0,u^0)+B^*y, u-u^0\rangle_{L_2}\geqslant 0$$

for all $u\in\mathcal{U}_{\text{ad}}$.

Computation of the adjoint operators yields the following (cp. [16, p. 871]): For $x\in \underline{W}^{1,\infty}(T,\mathbb{R}^n)\subset(\underline{W}^{1,\infty}(T,\mathbb{R}^n))^*$

$$\frac{d}{dt}[DA^*(x)](t)=\begin{cases}\int_{t+h}^{t_1}A_1^*x(s)\,ds+\int_t^{t_1}A_0^*x(s)\,ds, & t\in T_0\\ \int_t^{t_1}A_0^*x(s)\,ds, & t\in T_1.\end{cases} \tag{4.27}$$

For $l=(l_1,\rho)\in(\pi_1\mathcal{C}_\infty)^*\times(\pi_2\mathcal{C}_\infty)^*$

$$DC^*(l)(t)=\begin{cases}(t-t_0)l_1, & t\in T_0\\ (t_1-h-t_0)l_1+\int_{-h}^{t-t_1}\rho(s)\,ds, & t\in(t_1-h,t_1].\end{cases} \tag{4.28}$$

$D_1F(x(u^0), u^0)$ can be identified with the following element of $\underline{W}^{1,\infty}(T, \mathbb{R}^n)$:

$$\int_{t_0}^{t}\int_{\tau}^{t_1} D_1 f\big(x(u^0)(s), u^0(s)\big)\, ds\, d\tau, \qquad t \in T. \tag{4.29}$$

For $x \in \underline{W}^{1,\infty}(T, \mathbb{R}^n)$

$$(B^*x)(t) = B_0^*\dot{x}(t), \qquad t \in T. \tag{4.30}$$

$D_2F(x(u^0), u^0)$ can be identified with the following element of $L_\infty(T, \mathbb{R}^n)$:

$$D_2 f\big(x(u^0)(t), u^0(t)\big), \qquad t \in T. \tag{4.31}$$

Hence $y$ can be identified with an element of $\underline{W}^{1,\infty}(T, \mathbb{R}^n)$ and the derivative $\psi$ satisfies the adjoint equation (i) and the maximum condition (ii) in integral form. By standard arguments the pointwise form follows. Now suppose that $(l_0, \psi(t_1 - h), \psi|T_1)$ is trivial. Then the adjoint equation (see (4.28)) yields

$$\begin{aligned} 0 = \psi(t_1 - h) &= l_0 \int_{T_1} D_1 f\big(x^0(s), u^0(s)\big)\, ds \\ &\quad + \int_{T_1} A_0^* \psi(s)\, ds + l_1 \\ &= l_1. \end{aligned}$$

This contradicts the nontriviality condition $(l_0, l_1) \neq (0,0)$.

If a neighborhood of $\varphi_1 \in \mathcal{C}_\infty$ is completely attainable, we can apply the nondegeneracy part of Theorem 3.3 with $Z = \mathcal{C}_\infty$, $H = \mathcal{C}_2$, and trivial finite dimensional part. This allows to choose $l_0 = -1$. □

**Remark 4.3.** The idea for the construction in the first part of the proof is taken from [26].

**Remark 4.4.** By essentially the same proof, Theorem 4.3 holds also if $\Omega(t) \equiv \Omega$ is an *unbounded* closed convex set.

**Remark 4.5.** Another sufficient nondegeneracy condition is the following:

There exists a neighborhood $V$ of $\varphi_1(-h) \in \pi_1 \mathcal{C}_\infty$ such that for each $\alpha \in V$ there is a trajectory $x$ of (4.2) with initial condition (4.3) and control $u \in \mathcal{U}_{ad}$ such that $x(t_1 - h) = \alpha$.

A slight modification of Lemma 4.2 shows that this condition together with (4.20) implies the local attainability condition which by Theorem 4.3 is sufficient for nondegeneracy.

**Remark 4.6.** If the relaxed version of Problem 2 (in the sense of Warga) is considered, the semicontinuity property (4.18) is always satisfied, and a global maximum principle is obtained. Relaxed problems which appear more realistic are treated in [15].

**Remark 4.7.** It is obvious from the proof that the nontriviality condition can be strengthened to:

$$(0,0,0) \neq \left( l_0, \psi(t_1-h), \int_{T_1} A_0^* \psi(s)\, ds \right) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n.$$

**Remark 4.8.** In essence, Theorem 4.3 states the existence of a Lagrange multiplier in $W^{1,2}([-h,0],\mathbb{R}^n)$. Formally, this might have been obtained by working in the Hilbert spaces

$$Y = L_2(T,\mathbb{R}) \quad \text{and} \quad Z = H = W^{1,2}([-h,0],\mathbb{R}^n).$$

However, the regularity condition (4.19) required for an application of Theorem 3.3 (i.e., local attainability) is never achievable in $L_2$-norm with controls taking values in a bounded set. Thus it is not possible to prove Theorem 4.3 by a direct application of Theorem 3.3.

**Remark 4.9.** Consider an implicit control constraint of the form

$$q(u(t),t) \leq 0 \qquad \text{f.a.e. } t \in T,$$

where $q: \mathbb{R}^m \times T \to \mathbb{R}^r$ is a function which has a derivative with respect to the first argument such that $D_1 q$: $R^m \times T \to \mathbb{R}^{mr}$ is continuous.

This is a special case of (4.5) if $q(\cdot,t)$ is convex for a.e. $t \in T$: define

$$\Omega(t) := \{\omega \in \mathbb{R}^m : q(\omega,t) \leq 0\}, \qquad t \in T.$$

Bien and Chyung [10] consider implicit control constraints of the form $q(u(t)) \leq 0$ for $t \in T$. For the system (4.2) their regularity condition specializes to

$$L_\infty(T_1,\mathbb{R}^n) = \Big\{ (B_0 \alpha u(t), t \in T_1) : u \in L_\infty(T_1,\mathbb{R}^m), \alpha \geq 0, \\ \operatorname*{ess\,sup}_{t \in T_1} \big[ q(u^0(t)) + Dq(u^0(t))u(t) \big] < 0 \Big\}.$$

Obviously, this condition implies rank $B_0 = n$ and $\mathcal{Q}_\infty = W^{1,\infty}([-h,0],\mathbb{R}^n)$. It is satisfied if

$$B_0 u^0(t) \in \operatorname{int}^{\delta}_{\mathbb{R}^n} B_0 \Omega \qquad \text{for a.e. } t \in T_1,$$

for a constant $\delta > 0$. Consequently, this condition implies the regularity condition (4.20) where only projections onto $\mathfrak{B}$ are considered.

Similarly, also the regularity condition in [15] implies (4.20) and rank $B_0 = n$.

In the absence of control constraints, the closedness of $\mathcal{Q}_\infty$ in $W^{1,\infty}$ is sufficient and in the following sense also *necessary* for the validity of the maximum principle:

If it is not satisfied, there exists a performance index such that the maximum principle holds only in trivial form [22, 23].

The following example shows that also condition (4.20) cannot be omitted. In this example the assumptions of Theorem 3.3 are satisfied, in particular, condition (4.19) holds, while (4.20) is violated. It turns out, that the optimality conditions of Theorem 4.3 hold only trivially with $l_0=0$, $l_1=0$, $\psi=0$.

Thus there exists a Lagrange multiplier $l=W^{1,\infty}([-h,0],\mathbb{R}^n)^*$ which, however, is not in $W^{1,2}([-h,0],\mathbb{R}^n)$. This leads us to the general (unproven) conjecture:

If a neighborhood of $\varphi_1\in W^{1,\infty}([-h,0],\mathbb{R}^n)$ is completely attainable and condition (4.20) is violated, there exists a performance index such that the conditions of the maximum principle hold only trivially.

*Example:* Minimize $-\int_{-1}^{0} u(t)\,dt$

$$\text{s.t.}\quad \dot{x}(t)=x(t-2)+u(t),\quad t\in[-1,4]$$

$$x(t)=-1 \qquad t\in[-3,-1]$$

$$x(t)=\varphi_1(t-4)=\begin{cases} 1/2t^2-2t+4, & t\in[2,3] \\ 1/2t^2-2, & t\in[3,4] \end{cases}$$

$$u(t)\in\Omega:=[0,9/2],\qquad t\in[-1,4].$$

We claim:

(i) There exists an optimal solution $(x^0,u^0)$ defined by

$$u^0(t):=\begin{cases} 2, & t\in[-1,1] \\ -t+3, & t\in(1,2] \\ 0, & t\in(2,3] \\ 2, & t\in(3,4] \end{cases}$$

$$x^0(t):=\begin{cases} -1, & t\in[-3,-1] \\ t, & t\in(-1,2] \\ \varphi_1(t-4), & t\in(2,4] \end{cases}$$

(ii) A neighborhood of $\varphi_1\in W^{1,\infty}([-2,0],\mathbb{R})$ is completely attainable, in particular, condition (4.19) is satisfied.

(iii) Condition (4.20) is violated.

(iv) The maximum principle holds only trivially.

ad (i): Suppose $x^1=x(u^1)$ is any trajectory attaining $\varphi_1$ with $u^1(t)\in\Omega$, $t\in[-1,4]$. We have to show that

$$\int_{-1}^{0} u^1(s)\,ds\leqslant\int_{-1}^{0} u^0(s)\,ds.$$

For a.e. $t\in[2,3]$

$$u^0(t)=0;$$

thus

$$\begin{aligned} x_1(t-2) &= \dot{x}^1(t)-u^1(t)\\ &\leqslant \dot{x}^0(t)-u^0(t)\\ &= x^0(t-2), \qquad t\in[2,3], \end{aligned}$$

in particular $x^1(0)\leqslant x^0(0)$, i.e.,

$$-1+\int_{-1}^{0}\left[x^1(t-2)+u^1(t)\right]dt = x^1(0)$$
$$\leqslant x^0(0) = -1+\int_{-1}^{0}\left[x^0(t-2)+u^0(t)\right]dt.$$

Since

$$x^1(t-2)=x^0(t-2) \qquad \text{for } t\in[-1,0],$$
$$\int_{-1}^{0}u^1(t)\,dt\leqslant\int_{-1}^{0}u^0(t)\,dt.$$

ad (ii): Let $\varphi\in W^{1,\infty}([-2,0],\mathbb{R})$ satisfy

$$\|\varphi-\varphi_1\|_{W^{1,\infty}}<1/2.$$

Then $\varphi$ is attainable by

$$u(t)=\begin{cases}1, & t\in[-3,0]\\ 2, & t\in(0,1]\\ \varphi(-2)+1, & t\in(1,2]\\ \dot{\varphi}(t-4)-t+3, & t\in(2,3]\\ \dot{\varphi}(t-4)-\varphi(-2)(t-3), & t\in(3,4]\end{cases}$$

$$x(t)=\begin{cases}-1, & t\in[-3,0]\\ t-1, & t\in(0,1]\\ \varphi(-2)(t-1), & t\in(1,2]\\ \varphi(t-4), & t\in(2,4].\end{cases}$$

$u$ takes values only in $\Omega=[0,9/2]$ and $x$ satisfies the initial condition and the systems equation with control $u$.

ad (iii): Observe that for $t\in(2,3]$

$$u^0(t)=0\in\delta\Omega=\{0,9/2\}.$$

Thus on $(2,3]$ (4.20) is not satisfied.

ad (iv): Assume that the maximum principle holds in a nontrivial form. Then without loss of generality, $l_0 = -1$. This follows from the nondegeneracy part of Theorem 4.3, since by (ii) a neighborhood of $\varphi_1 \in W^{1,\infty}([-h,0],\mathbb{R})$ is completely attainable.

The maximum condition (see Theorem 4.3) has the following form:

$$\left[-D_2 f\left(x^0(t), u^0(t)\right) - \psi(t)\right]\left[u^0(t) - \omega\right] \geq 0$$

for all $\omega \in \Omega$ and a.e. $t \in [-1,4]$, i.e. for all $\omega \in [0,9/2]$

$$\psi(t)(\omega - 2) \geq 2 - \omega, \qquad \text{a.e. } t \in [-1,0]$$

and

$$\psi(t)(\omega - 2) \geq 0, \qquad \text{a.e. } t \in [0,1].$$

Then $\psi(t) = 0$ for $t \in [0,1]$. The maximum condition on $[-1,0]$ implies for $\omega = 3$

$$\psi(t) \geq -1$$

and for $\omega = 1$

$$-\psi(t) \geq 1$$

Thus $\psi(t) = -1$ for $t \in [-1,0]$.

Since $\psi$ is continuous on $[-1,2]$, this contradicts $\psi(t) = 0$ for $t \in [0,1]$. Then (iv) holds.

It follows from the example that the nondegeneracy condition in Theorem 3.3 (i.e., local attainability) does not guarantee that $P$ is an exact penalty function: If for an $\varepsilon > 0$, $u^0$ is an optimal solution of the $\varepsilon$-Problem, $l^\varepsilon$ would be a Lagrange multiplier in $H = W^{1,2}([-h,0],\mathbb{R}^n)$ for the original problem, which cannot happen in the example.

This is of interest in connection with the work by Dolecki and Rolewicz [18], who prove that in case of local attainability a large class of (nondifferentiable) penalty functionals are exact (for example, penalization in terms of the norm: [18, Theorem 22]). The example above shows that in general $P$ is not exact even if local attainability in $W^{1,\infty}$-norm is assumed.

## Acknowledgments

I thank Professor Diederich Hinrichsen for his constructive criticism during the preparation of this material.

## References

1. A. V. Balakrishnan, The epsilon technique—a constructive approach to optimal control, in: *Control Theory and Calculus of Variations*, ed. by L. W. Neustadt, Academic Press, New York, 1967.

2. A. V. Balakrishnan, On a new computing technique in optimal control, *SIAM J. Control*, 6, 149–173 (1968).
3. A. V. Balakrishnan, A computational approach to the maximum principle, *J. Computer Syst. Sci.*, 5, 163–191 (1971).
4. H. T. Banks, and G. A. Kent, Control of functional differential equations of retarded and neutral type with target sets in function space, *SIAM J. Control*, 10, 567–593 (1972).
5. H. T. Banks, and A. Manitius, Application of abstract variational theory to hereditary systems—a survey, *IEEE Transactions on Automatic Control*, *AC-19*, 524–533 (1974).
6. G. R. Bates, *Hereditary Optimal Control Problems*, Ph.D. Thesis, Purdue University, West Lafayette, Ind., 1977.
7. L. D. Berkovitz, A penalty function proof of the maximum principle, *Appl. Math. Optim.*, 2, 291–303 (1975/6).
8. E. J. Beltrami, A constructive proof of the Kuhn-Tucker multiplier rule, *J. Math. Anal. Appl.*, 26, 297–306 (1969).
9. Z. Bien, *Optimal Control of Delay Systems*, Ph.D. Thesis, University of Iowa, Iowa, 1975.
10. Z. Bien and D. H. Chyung, Optimal control of delay systems with a final function condition, *Int. J. Control*, 32, 539–560 (1980).
11. M. Brokate, A regularity condition for optimization in Banach spaces: Counterexamples, *Appl. Math. Optim.*, 6, 189–192 (1980).
12. H. H. Buehler, Application of Neustadt's theory of extremals to an optimal control problem with a functional differential equation and a functional inequality constraint, *Appl. Math. Optim.*, 2, 34–74 (1975/6).
13. F. Colonius, *Hereditäre differenzierbare Systeme mit Funktionenraum–Endbedingung und punktweisen Steuerbeschränkungen*: *Notwendige Optimalitätsbedingungen und Erreichbarkeit*, Dissertation, Universität Bremen, Bremen, 1979.
14. F. Colonius, Necessary optimality conditions for non-linear hereditary differential systems with function space end constraints, in: *Functional Differential Systems and Related Topics*, ed. by M. Kisielewicz, The Higher College of Engineering in Zielona Gora, Zielona Gora, 62–71, 1980.
15. F. Colonius, Regularization of Lagrange multipliers for time delay systems with fixed final state, in: *Optimization and Optimal Control*, ed. by W. Oettli et al., Springer-Verlag, 163–177, 1981.
16. F. Colonius, and D. Hinrichsen, Optimal control of functional differential systems, *SIAM J. Control Optim.*, 16, 861–879 (1978).
17. R. Courant, *Calculus of Variations and Supplementary Notes and Exercises* (mimeographed notes), supplementary notes by Martin Kruskal and Hanan Rubin, revised and amended by J. Moser, New York University, 1956–1957.
18. S. Dolecki, and S. Rolewicz, Exact penalties for local minima, *SIAM J. Control Optim.*, 17, 596–606 (1979).
19. J. Dugundjii, *Topology*, Allyn and Bacon, Boston-London-Sydney-Toronto, 1966.
20. K. Glashoff, Schwache Stetigkeit bei nichtlinearen Kontrollproblemen, Numerische Methoden bei Optimierungsaufgaben, Tagung Oberwolfach 1971 = *Internationale Schriftenreihe zur Numerischen Mathematik*, Vol. 17, Birkhäuser Verlag, 1973.
21. E. Hille, and R. S. Philips, *Functional Analysis and Semi-groups*, American Mathematical Society, Providence, R. I., 1974.
22. S. Kurcyusz, A local maximum principle for operator constraints and its application to systems with time lag, *Control Cyber.*, 2, 99–125 (1973).
23. S. Kurcyusz, On the existence and non-existence of Lagrange multipliers in Banach spaces, *J. Optim. Th. Appl.*, 20, 81–110 (1976).
24. S. Kurcyusz, and A. W. Olbrot, On the closure in $W^{1,q}$ of the attainable subspace of linear time lag systems, *J. Diff. Eq.*, 24, 29–50 (1977).
25. D. G. Luenberger, *Optimization by Vector Space Methods*, J. Wiley, New York, 1968.
26. K. Makowski, and L. W. Neustadt, Optimal control problems with mixed control phase variable equality and inequality constraints, *SIAM J. Control*, 12, 184–228 (1974).
27. E. J. McShane, The Lagrange multiplier rule, *Am. Math. Monthly*, 80, 922–925 (1973).
28. J. W. Mersky, *An Application of the Epsilon Technique to Control Problems with Inequality Constraints*, Ph.D. Thesis, University of California, Los Angeles, 1973.

29. S. M. Robinson, Stability theory for systems of inequalities in nonlinear programming; part II: Differentiable nonlinear systems, *SIAM J. Num. Anal.*, 13, 497–513 (1976).
30. J. Warga, *Optimal Control of Differential and Functional Equations*, Academic Press, New York, 1972.
31. A. P. Wierzbicki, and S. Kurcyusz, Projection on a cone, penalty functionals, and duality theory for problems with inequality constraints in a Hilbert space, *SIAM J. Control Optim.*, 15, 25–56 (1977).
32. K. Yosida, *Functional Analysis*, 4th ed., Springer-Verlag, New York, 1974.
33. J. Zowe, and S. Kurcyusz, Regularity and stability for the mathematical programming problem in Banach spaces, *Appl. Math. Optim.*, 5, 49–62 (1979).