

A GRADIENT TECHNIQUE FOR AN OPTIMAL CONTROL PROBLEM GOVERNED BY A SYSTEM OF NONLINEAR FIRST ORDER PARTIAL DIFFERENTIAL EQUATIONS

MOHAMMAD A. KAZEMI¹

(Received 24 September 1992; revised 3 March 1993)

Abstract

In this paper a class of optimal control problems with distributed parameters is considered. The governing equations are nonlinear first order partial differential equations that arise in the study of heterogeneous reactors and control of chemical processes. The main focus of the present paper is the mathematical theory underlying the algorithm. A conditional gradient method is used to devise an algorithm for solving such optimal control problems. A formula for the Fréchet derivative of the objective function is obtained, and its properties are studied. A necessary condition for optimality in terms of the Fréchet derivative is presented, and then it is shown that any accumulation point of the sequence of admissible controls generated by the algorithm satisfies this necessary condition for optimality.

1. Introduction

In this paper, a conditional gradient technique is used to develop and study a computational method for a class of optimal control problems with distributed parameters. The partial differential equations involved form a system of nonlinear first order partial differential equations that arise in the study of heterogeneous reactors and control of chemical processes (cf. [5], [8]). In [4] we have studied the existence and uniqueness of L_p -solution, norm and pointwise estimates, as well as the stability of this class of partial differential equations. The problem of developing computational algorithms for solving optimal control problems with distributed parameter has been the subject of a number of recent studies (cf. [3], [7], [11], [13]).

Pontryagin-type necessary conditions for optimality of the control process considered in this paper have been obtained in [12] and by the author in [2].

In the present study we impose control constraints but no state or terminal state constraints. The approach of this paper is motivated by those of [6], [9], [10], [11]

¹Department of Mathematics, University of North Carolina at Charlotte, Charlotte, NC 28223, U.S.A.
© Australian Mathematical Society, 1994, Serial-fee code 0334-2700/94

and [13]. As will be seen from the presentation, the technique employed here is by no means particular to the governing equations (see also [11]). What is required is a rather thorough understanding of the governing equations.

In Section 2, we describe the optimal control problem, notation, and the hypotheses required. We will also recall certain results from [4] that are required in subsequent developments. In Section 3, the adjoint system of equations (the linear conjugate problem) is introduced. In Section 4 we obtain an integral representation of the increment of the functional as well as a bound on the remainder term. In Section 5 the properties of the objective function are studied, in particular, it is shown that the cost functional is Fréchet differentiable, and its Fréchet derivative is found. Finally, in Section 6, we describe the algorithm and study its convergence properties.

2. Problem statement

In the rectangular region $G := [a, a + h] \times [b, b + k]$, we consider a controlled process described by a system of first order partial differential equations of the form

$$\begin{aligned} \frac{\partial x}{\partial s} &= f(s, t, x(s, t), y(s, t), u(s, t)), & \text{a.e. } (s, t) \in G, \\ \frac{\partial y}{\partial t} &= g(s, t, x(s, t), y(s, t), u(s, t)), & \text{a.e. } (s, t) \in G, \end{aligned} \tag{2.1}$$

$$\begin{aligned} x(a, t) &= \varphi(t) & b \leq t \leq b + k, \\ y(s, b) &= \psi(s) & a \leq s \leq a + h, \end{aligned} \tag{2.2}$$

where $z(s, t) := (x(s, t), y(s, t)) \in \mathbb{R}^{n_1+n_2}$, $u(s, t) = (u_1, u_2, \dots, u_m) \in \mathbb{R}^m$ are the state and control of the system, respectively; $\varphi(t)$ and $\psi(s)$ are given vector functions in the respective intervals.

Let U be a fixed compact and convex subset of \mathbb{R}^m , and let Ω be defined by

$$\Omega := \{u \in L_\infty(G, \mathbb{R}^m) \mid u(s, t) \in U \text{ for all } (s, t) \in G\}. \tag{2.3}$$

Any function in Ω will be called an admissible control.

The optimal control problem consists of finding a control $u \in \Omega$ that minimizes the cost functional

$$J(u) := \int_a^{a+h} P(s, y(s, b + k)) ds + \int_b^{b+k} R(t, x(a + h, t)) dt, \tag{2.4}$$

where $z(s, t) := (x(s, t), y(s, t))$ is the solution of (2.1)–(2.2) corresponding to u .

NOTATION. For the vector $X = (X_1, \dots, X_n) \in \mathbb{R}^n$, we write $|X| := \sum_{i=1}^n |X_i|$; for a matrix $M = (m_{ij})$, we write $|M| := \sum_{i,j} |m_{ij}|$. For $z \in L_p(G)$, $1 \leq p \leq \infty$,

$\|z\|_p$ denotes the usual L_p norm, in particular, $\|z\|_\infty := \text{ess sup } |z|$. As usual, $L_p(G, \mathbb{R}^n)$ will denote the Banach space of n -vector functions $F(s, t) = (F_1, \dots, F_n)$ such that $F_i \in L_p(G)$. For such a function, we write $\|F\|_p := \sum_{i=1}^n \|F_i\|_p$, and $F_\beta := (F_{1\beta}, \dots, F_{n\beta})$ for the generalized derivatives of F with respect to β .

We require the following assumptions:

- A₁: The vector functions $\varphi(t) = (\varphi_1, \dots, \varphi_{n_1}), \psi(s) = (\psi_1, \dots, \psi_{n_2})$ belong to L_p for some $p, 1 \leq p \leq \infty$.
- A₂: For fixed z and u , the functions $f(s, t, z, u)$ and $g(s, t, z, u)$ are measurable in G . For each $u \in \Omega$, the functions $f_1(s, t) := f(s, t, 0, u(s, t)), g_1(s, t) := g(s, t, 0, u(s, t))$ belong to $L_p(G)$ with p as in (A₁).
- A₃: The partial derivatives of f and g with respect to z, u exist and furthermore there are constants K_1, K_2 such that on $G \times \mathbb{R}^n \times \mathbb{R}^m$,

$$\begin{aligned}
 |f_d| &\leq K_1, |g_d| \leq K_1, \\
 |f_d(t, s, z, u) - f_d(t, s, \zeta, v)| &\leq K_2\{|z - \zeta| + |u - v|\}, \\
 |g_d(t, s, z, u) - g_d(t, s, \zeta, v)| &\leq K_2\{|z - \zeta| + |u - v|\},
 \end{aligned}$$

where f_d denotes any of the derivative matrices $f_z := (f_{iz_j}), f_u := (f_{iu_j})$. Similarly for g_d .

- A₄: The function $P(s, y)$ is measurable in s for fixed y and twice continuously differentiable in y for fixed s ; $R(t, x)$ is measurable in t for fixed x and twice continuously differentiable in x for fixed t .

We require the following results, proved in [4]. Let X denote the subspace of $L_p(G, \mathbb{R}^{n_1+n_2})$ consisting of all vector functions $z(s, t) = (x(s, t), y(s, t))$ for which the generalized derivatives x_s and y_t exist and belong, respectively, to $L_p(G, \mathbb{R}^{n_1})$ and $L_p(G, \mathbb{R}^{n_2})$.

THEOREM 2.1. *Under assumptions (A₁)–(A₃), for each $u \in L_\infty(G, \mathbb{R}^m)$, the system of partial differential equations (2.1)–(2.2) has a unique solution $z := z(u)$ in X .*

Let $u, v \in L_\infty(G, \mathbb{R}^m)$ and let $z(u), z(v)$ be the corresponding solutions of (2.1)–(2.2). For $z(u)(s, t) := (x(u)(s, t), y(u)(s, t))$, we set

$$\Delta_{u,v}f(s, t) := f(s, t, z(u)(s, t), v(s, t)) - f(s, t, z(u)(s, t), u(s, t))$$

with similar definition for $\Delta_{v,u}g(s, t)$.

THEOREM 2.2. *Under the assumptions (A₁)–(A₃), there exists a nonnegative constant C such that*

$$\begin{aligned}
 |x(v)(s, t) - x(u)(s, t)| &\leq C \left[\int_a^s |\Delta_{v,u} f(\alpha, t)| d\alpha + \right. \\
 &\quad \left. \int \int_G (|\Delta_{v,u} f(\alpha, \beta)| + |\Delta_{v,u} g(\alpha, \beta)|) d\alpha d\beta \right], \\
 |y(v)(s, t) - y(u)(s, t)| &\leq C \left[\int_b^t |\Delta_{v,u} g(s, \beta)| d\beta + \right. \\
 &\quad \left. \int \int_G (|\Delta_{v,u} f(\alpha, \beta)| + |\Delta_{v,u} g(\alpha, \beta)|) d\alpha d\beta \right].
 \end{aligned}$$

REMARK. Under the assumption (A₄), it follows from Theorem 2.2 that the cost functional J defined in (2.4) is well-defined and bounded.

3. The adjoint equations

To obtain the adjoint system of equations (the linear conjugate problem) for the optimal control problem, we first linearize the state equations (2.1) around a fixed pair $(u, z(u))$, find the adjoint linear operator corresponding to the linearized differential system, and then introduce suitable boundary conditions using (2.2) and (2.3). We omit the details. The resulting adjoint equations are:

$$\begin{aligned}
 \frac{\partial \lambda}{\partial s} &= -\frac{\partial H}{\partial x}(s, t, z(u)(s, t), u(s, t), \lambda(s, t), \mu(s, t)), \\
 \frac{\partial \mu}{\partial t} &= -\frac{\partial H}{\partial y}(s, t, z(u)(s, t), u(s, t), \lambda(s, t), \mu(s, t)),
 \end{aligned} \tag{3.1}$$

$$\lambda(a+h, t) = \frac{\partial R(t, x(u)(a+h, t))}{\partial x}, \quad \mu(s, b+k) = \frac{\partial P(s, y(u)(s, b+k))}{\partial y}, \tag{3.2}$$

where

$$H(s, t, z, u, \lambda, \mu) := \lambda \cdot f(s, t, z, u) + \mu \cdot g(s, t, z, u).$$

It follows from Theorem 2.1 that for a given pair $u, z(u)$, the linear system of equations (3.1)–(3.2) has a unique solution $(\lambda, \mu) \in L_p(G, \mathbb{R}^{n_1+n_2})$.

4. Integral representation of the increment

In order to find the gradient of the cost functional and study its properties for later developments, we need an integral representation of the increment.

Thus let $(u, z(u))$ and $(v, z(v))$ be two admissible pairs, and let $\lambda = \lambda(u)$, $\mu = \mu(u)$, be the solution of (3.1)–(3.2) corresponding to $(u, z(u))$. Then

$$\begin{aligned}
 J(v) - J(u) &= \int_a^{a+h} [P(s, y(v)(s, b+k)) - P(s, y(u)(s, b+k))] ds \\
 &\quad + \int_b^{b+k} [R(t, x(v)(a+h, t)) - R(t, x(u)(a+h, t))] dt \\
 &= \int_a^{a+h} [P_y(s, y(u)(s, b+k) \cdot (y(v) - y(u))(s, b+k)] ds \\
 &\quad + \int_b^{b+k} [R_x(t, x(u)(a+h, t)) \cdot (x(v) - x(u))(a+h, t)] dt \\
 &\quad + o(|\Delta x|) + o(|\Delta y|), \tag{3.3}
 \end{aligned}$$

where $\Delta x := x(v) - x(u)$, $\Delta y := y(v) - y(u)$.

In what follows we will also use the following notations. We use $\lambda(s, t) := \lambda(u)(s, t)$, $\mu(s, t) := \mu(u)(s, t)$; we write $\Delta_{v,u}H(s, t)$ for the difference

$$H(s, t, z(u)(s, t), v(s, t), \lambda(s, t), \mu(s, t)) - H(s, t, z(u)(s, t), u(s, t), \lambda(s, t), \mu(s, t)),$$

and denote by η the remainder term

$$\begin{aligned}
 &\iint_G [H_z(s, t, \tilde{z}(s, t), v(s, t), \lambda(s, t), \mu(s, t)) \\
 &\quad - H_z(s, t, z(u)(s, t), u(s, t), \lambda(s, t), \mu(s, t))] ds dt,
 \end{aligned}$$

where $\tilde{z} := z(u) + \theta(s, t)(z(v) - z(u))$, for some $\theta(s, t) \in (0, 1)$. Using (3.1)–(3.2), integration by parts and boundary conditions (2.2), we continue (3.3):

$$\begin{aligned}
 J(v) - J(u) &= \iint_G [\mu_t \cdot \Delta y + \mu \cdot \Delta y_t] ds dt \\
 &\quad + \iint_G [\lambda_s \cdot \Delta x + \lambda \cdot \Delta x_s] ds dt + o(|\Delta x|) + o(|\Delta y|) \\
 &= \iint_G \Delta_{v,u}H(s, t) ds dt \\
 &\quad + \iint_G H_z(s, t, \tilde{z}(s, t), v(s, t), \lambda(s, t), \mu(s, t)) \cdot \Delta z ds dt \\
 &\quad - \iint_G H_z(s, t, z(u)(s, t), u(s, t), \lambda(s, t), \mu(s, t)) \cdot \Delta z ds dt \\
 &\quad + o(|\Delta x|) + o(|\Delta y|) \\
 &= \iint_G \Delta_{v,u}H(s, t) ds dt + \eta + o(|\Delta x|) + o(|\Delta y|).
 \end{aligned}$$

The remainder term η can be estimated as

$$\begin{aligned} |\eta| &= \iint_G |H_z(s, t, \tilde{z}, v, \mu) - H_z(s, t, z(u), u, \lambda, \mu)| |\Delta z| \, ds \, dt \\ &\leq \iint_G |H_z(s, t, \tilde{z}, v, \lambda, \mu) - H_z(s, t, z(u), v, \lambda, \mu)| |\Delta z| \, ds \, dt \\ &\quad + \iint_G |H_z(s, t, z(u), v, \lambda, \mu) - H_z(s, t, z(u), u, \lambda, \mu)| |\Delta z| \, ds \, dt. \end{aligned}$$

Using (A₃), and noting that $|\tilde{z} - z(u)| \leq |z(v) - z(u)|$, we continue the above estimate:

$$\begin{aligned} |\eta| &\leq C \left[\iint_G |z(v) - z(u)|^2 \, ds \, dt + \iint_G |v - u| |z(v) - z(u)| \, ds \, dt \right] \\ &\leq C \left[\iint_G |z(v) - z(u)|^2 \, ds \, dt + \|v - u\|_\infty \iint_G |z(v) - z(u)| \, ds \, dt \right]. \end{aligned}$$

Now the estimates of Theorem 2.2, the mean value theorem, and the boundedness of $|f_u|, |g_u|$ in (A₃) yield

$$|\eta| \leq K \|v - u\|_\infty^2.$$

Note also, from Theorem 2.2, that $o(|\Delta x|)$ and $o(|\Delta y|)$ can be replaced by $o(|v - u|)$.

In conclusion, we have the following integral representation of the increment of the cost functional

$$J(v) - J(u) = \iint_G \Delta_{v,u} H(z(u)(s, t)) \, ds \, dt + \eta \tag{3.4}$$

where, for some nonnegative constant C,

$$|\eta| \leq C \|v - u\|_\infty^2. \tag{3.5}$$

5. Properties of the cost functional

In this section we show that the cost functional J is Fréchet differentiable, find its Fréchet derivative and study its properties.

We first recall the definition of Fréchet differentiability.

DEFINITION 5.1. A functional $J : L_\infty(G, \mathbb{R}^m) \rightarrow \mathbb{R}$ is said to be Fréchet differentiable at $u^0 \in L_\infty(G, \mathbb{R}^m)$, if there exists a continuous linear functional $J'(u) : L_\infty(G, \mathbb{R}^m) \rightarrow \mathbb{R}$ such that

$$\lim_{\|u - u^0\|_\infty \rightarrow 0} \frac{|J(u) - J(u^0) - J'(u^0)(u - u^0)|}{\|u - u^0\|_\infty} = 0. \tag{5.1}$$

The linear functional $J'(u^0)$ is called the Fréchet derivative of J at $u^0 \in L_\infty(G, \mathbb{R}^m)$.

THEOREM 5.1. *The cost functional $J : L_\infty(G, \mathbb{R}^m) \rightarrow \mathbb{R}$ defined in (2.4) is Fréchet differentiable everywhere on $L_\infty(G, \mathbb{R}^m)$. Furthermore, the derivative of J at $u^0 \in L_\infty(G, \mathbb{R}^m)$ is given by*

$$J'(u^0) = \iint_G [\lambda(u^0)(s, t) \cdot f_u(s, t, z(u^0)(s, t), u^0(s, t)) + \mu(u^0)(s, t) \cdot g_u(s, t, z(u^0)(s, t), u^0(s, t))] u(s, t) \, ds \, dt. \tag{5.2}$$

PROOF. Let $u, u^0 \in L_\infty(G, \mathbb{R}^m)$. Using (3.4)–(3.5), we have

$$\begin{aligned} J(u) - J(u^0) - J'(u^0)(u - u^0) &= \iint_G [\Delta_{u,u^0} H(z(u^0)(s, t)) - H(z(u^0)(s, t)) \cdot (u - u^0)] \, ds \, dt + \eta. \end{aligned}$$

Applying the mean value theorem to the term $\Delta_{u,u^0} H(z(u^0)(s, t))$ in the above integral we obtain

$$\begin{aligned} |J(u) - J(u^0) - J'(u^0)(u - u^0)| &\leq \iint_G |H_u(s, t, z(u^0)(s, t), \lambda(s, t), \mu(s, t), \tilde{u}(s, t)) \\ &\quad - H_u(s, t, z(u^0)(s, t), \lambda(s, t), \mu(s, t), u^0(s, t))| |u(s, t) - u^0(s, t)| \, ds \, dt \\ &\quad + C \|u - u^0\|_\infty^2, \end{aligned}$$

where $u(s, t) := u^0(s, t) + \theta(s, t)(u(s, t) - u^0(s, t))$, for some $\theta(s, t)$, $0 \leq \theta(s, t) \leq 1$. Using assumption (A₃), and noting that $|\tilde{u}(s, t) - u^0(s, t)| \leq |u(s, t) - u^0(s, t)|$, we have

$$|J(u) - J(u^0) - J'(u^0)(u - u^0)| \leq C' \|u - u^0\|_\infty^2,$$

from which the theorem follows.

We next investigate the continuity of the derivative $J'(u)$. We will require this property in establishing the convergence of the algorithm, to be presented in the next section.

LEMMA 5.1. *Let $f_d : L_\infty(G, \mathbb{R}^m) \rightarrow L_\infty(G, \mathbb{R}^{n \times r})$ be defined by*

$$f_d(u) := f_d(\cdot, \cdot, z(u)(\cdot, \cdot), u(\cdot, \cdot)). \tag{5.3}$$

Then

$$\|f_d(u)\|_\infty \leq K, \text{ for all } u \in L_\infty(G, \mathbb{R}^m), \tag{5.4}$$

where K is a nonnegative constant independent of u and $z(u)$. Furthermore, $f_d(u)$ is uniformly continuous on $L_\infty(G, \mathbb{R}^{n \times r})$.

REMARK. The same statement as in Lemma 5.1 holds for g_d .

PROOF OF LEMMA 5.1. The estimate (5.4) is an immediate consequence of assumption (A₃). To prove the second part of the lemma, note from assumption (A₃) that for $u, v \in L_\infty(G, \mathbb{R}^m)$,

$$\begin{aligned} & |f_d(s, t, z(v)(s, t), v(s, t)) - f_d(s, t, z(u)(s, t), u(s, t))| \\ & \leq K_2(|z(v)(s, t) - z(u)(s, t)| + |v(s, t) - u(s, t)|). \end{aligned}$$

Using Theorem 2.2,

$$\begin{aligned} |z(v)(s, t) - z(u)(s, t)| & \leq K_3 \left[\int_a^s |\Delta_{v,u} f(\alpha, t)| d\alpha + \int_b^t |\Delta_{v,u} g(s, \beta)| d\beta \right. \\ & \quad \left. + \iint_G (|\Delta_{v,u} f(\alpha, \beta)| + |\Delta_{v,u} g(\alpha, \beta)|) d\alpha d\beta \right]. \end{aligned}$$

In view of assumption (A₃), we have

$$|\Delta_{v,u} f(s, t)|, |\Delta_{v,u} g(s, t)| \leq C|v(s, t) - u(s, t)| \leq C\|v - u\|_\infty,$$

for almost all $(s, t) \in G$. Therefore, for almost all $(s, t) \in G$,

$$|f_d(s, t, z(v)(s, t), v(s, t)) - f_d(s, t, z(u)(s, t), u(s, t))| \leq C'\|v - u\|_\infty,$$

where C' denotes a nonnegative constant depending only on the dimensions of the rectangular region G . Finally, it follows from the last inequality that

$$\|f_d(v) - f_d(u)\| \leq C'\|v - u\|_\infty,$$

completing the proof of the lemma.

Recall from (3.1)–(3.2) that the adjoint variables λ and μ depend on u . In the following lemma we shall prove the uniform boundedness and uniform continuity of λ and μ with respect to u .

LEMMA 5.2.

- (a) $\|\lambda(u)\|_\infty \leq K, \|\mu(u)\|_\infty \leq K$ for some constant K , and for all $u \in L_\infty(G, \mathbb{R}^m)$.
- (b) $\lambda(u)$ and $\mu(u)$ are uniformly continuous on $L_\infty(G, \mathbb{R}^m)$.

PROOF. Note from (3.1)–(3.2) that (λ, μ) is the unique solution of a linear system of partial differential equations of the same form as (2.1)–(2.2). Thus Theorem 2.1 applies to the system (3.1)–(3.2). Noting in this case that $f_1 \equiv 0, g_1 \equiv 0$ (see (A₂))

for definitions of f_1 and g_1), Theorem 2.1 shows that the pointwise norms of λ and μ depend on the norms of P_y and R_x . Thus (a) follows from assumption (A₄).

To see the uniform continuity of λ and μ with respect to u , apply Theorem 2.2 (along with assumption (A₃)) to the adjoint system (3.1)–(3.2), which is a linear version of (2.1)–(2.2). This completes the proof of the lemma.

Let us finally check the uniform continuity of the derivative $J'(u)$ on $L_\infty(G, \mathbb{R}^m)$. Let $u, v \in L_\infty(G, \mathbb{R}^m)$. Then it follows from Theorem 5.1 that, for any $w \in L_\infty(G, \mathbb{R}^m)$

$$|J'(v)(w) - J'(u)(w)| \leq \int \int_G \left[|\lambda(v)(s, t) \cdot f_u(s, t, z(v)(s, t), v(s, t)) - \lambda(u)(s, t) \cdot f_u(s, t, z(u)(s, t), u(s, t))| + |\mu(v)(s, t) \cdot g_u(s, t, z(v)(s, t), v(s, t)) - \mu(u)(s, t) \cdot g_u(s, t, z(u)(s, t), u(s, t))| \right] \|w\|_\infty ds dt.$$

Using assumption (A₃) and Lemma 5.2, we have

$$|J'(v)(w) - J'(u)(w)| \leq \|w\|_\infty [\|\lambda(v) - \lambda(u)\|_\infty + \|\mu(v) - \mu(u)\|_\infty] + C_2 [\|\Delta_{v,u} f_u\|_\infty + \|\Delta_{v,u} g_u\|_\infty],$$

where C_1 and C_2 are constants independent of u, v , and w . It then follows that

$$\|J'(v)(w) - J'(u)(w)\| \leq C_1 [\|\lambda(v) - \lambda(u)\|_\infty + \|\mu(v) - \mu(u)\|_\infty] + C_2 [\|\Delta_{v,u} f_u\|_\infty + \|\Delta_{v,u} g_u\|_\infty], \tag{5.5}$$

where

$$\|J'(v) - J'(u)\| := \sup_{w \in L_\infty(G, \mathbb{R}^m)} \frac{|J'(v)(w) - J'(u)(w)|}{\|w\|_\infty}.$$

Applying Lemmas 5.2 and 5.3 to the above estimate (5.5) we can state the following

THEOREM 5.2. *The derivative $J'(u)$ of the cost functional $J(u)$, which maps $L_\infty(G, \mathbb{R}^m)$ into its dual space $L_\infty^*(G, \mathbb{R}^m)$ is uniformly continuous on $L_\infty(G, \mathbb{R}^m)$.*

6. The algorithm

Based on the analysis of the previous section, following the lines of the algorithm devised in [11, pages 148–149] we can now devise the following conditional gradient algorithm for the optimal control problem (2.1)–(2.4).

STEP 1: Choose $\alpha, \beta \in (0, 1)$, and let $u^0 \in \Omega$ be an initial control. Set $k = 0$.

STEP 2: Solve (2.1)–(2.2) for $z(u^k)$, and (3.1) for $\lambda(u^k), \mu(u^k)$. Using (5.2), compute $J'(u^k)$.

STEP 3: Find a control \bar{u}^k such that

$$\begin{aligned} & [\lambda(u^k)(s, t) \cdot f_u(s, t, z(u^k)(s, t), u^k(s, t)) \\ & + \mu(u^k)(s, t) \cdot g_u(s, t, z(u^k)(s, t), u^k(s, t))] \bar{u}^k(s, t) \\ & \leq [\lambda(u^k)(s, t) \cdot f_u(s, t, z(u^k)(s, t), u^k(s, t)) \\ & + \mu(u^k)(s, t) \cdot g_u(s, t, z(u^k)(s, t), u^k(s, t))] u^k(s, t) \end{aligned} \tag{6.1}$$

for all $u \in \Omega$ and for a.a. $(s, t) \in G$.

STEP 4: If $J'(u^k)(\bar{u}^k - u^k) = 0$ then set $u^{k+l} := u^k$ for $l = 1, 2, \dots$, and stop; otherwise go to the next step.

STEP 5: Choose α^m to be the first element in the sequence $1, \alpha, \alpha^2, \dots$, such that

$$J(u^k + \alpha^k(\bar{u}^k - u^k)) - J(u^k) \leq \alpha^k \beta J'(u^k)(\bar{u}^k - u^k). \tag{6.2}$$

STEP 6: Set $u^{k+1} := u^k + \alpha^k(\bar{u}^k - u^k)$. Go to Step 2 with $k = k + 1$.

Our next aim is to show that the above algorithm is well-defined. Steps 1 and 2 are well-defined. In Lemma 6.1 below we will prove the existence of a control function \bar{u}^k defined in Step 3. Recall that Ω is convex and so for any $0 \leq \theta \leq 1$, $u^k + \theta(\bar{u}^k - u^k) \in \Omega$. We have

$$\lim_{\theta \rightarrow 0} [J(u^k + \theta(\bar{u}^k - u^k)) - J(u^k)]/\theta = J'(u^k)(\bar{u}^k - u^k).$$

Thus there does, in fact, exist α^k satisfying (6.2). (Note from (6.1) that terms on both sides of (6.2) are negative.) Finally, Step 6 is well-defined by the convexity of Ω , that is, $u^{k+l} \in \Omega$.

LEMMA 6.1. *For each $u \in \Omega$, there exists a control $\bar{u} \in \Omega$ that minimizes the linear functional $J'(u)(\cdot)$.*

PROOF. Proof is based on the Fillipov implicit function theorem, cf. [11, Lemma 4.6.1].

The following theorem gives a necessary condition for optimality of the optimal control problem (2.1)–(2.4), in terms of the derivative of the cost functional.

THEOREM 6.1. *Let $\bar{u} \in \Omega$ be an optimal control for problem (2.1)–(2.4). Then*

$$J'(\bar{u})\bar{u} \leq J'(\bar{u})u, \quad \text{for all } u \in \Omega. \tag{6.3}$$

PROOF. Let $u \in \Omega$ and $\alpha \in (0, 1]$. We have $(1 - \alpha)\bar{u} + \alpha u \in \Omega$, and so $J(1 - \alpha)\bar{u} + \alpha u \geq J(\bar{u})$, from which we have $[J((1 - \alpha)\bar{u} + \alpha u) - J(\bar{u})]/\alpha \geq 0$. Letting $\alpha \rightarrow 0$, we obtain $J'(\bar{u})(u - \bar{u}) \geq 0$. This completes the proof.

In the following theorem, we establish a convergence property of the algorithm.

THEOREM 6.2. *Let $\{u^k\}$ be a sequence of admissible controls generated by the above algorithm. If u^* is an accumulation point of $\{u^k\}$, then u^* satisfies the optimality condition (6.3).*

PROOF. The proof is similar to the one given for [11, Theorem IV.6.2]. If, for some k , u^k satisfies the optimality condition (6.3), then by Step 4 of the algorithm, we have $u^{k+i} = u^k$, $i = 1, 2, \dots$. In this case, the theorem is obvious. We thus assume that $J'(u^k)(\bar{u} - u^k) \neq 0$ for all $k = 1, 2, \dots$. Since u^* is an accumulation point of the sequence $\{u^k\}$, there is a subsequence of it, again denoted by $\{u^k\}$, converging to u^* in the norm of $L_\infty(G, \mathbb{R}^m)$.

According to (6.2),

$$\alpha^k \beta |J'(u^k)(\bar{u} - \bar{u}^k)| \leq |J(u^{k+1}) - J(u^k)|, \tag{6.4}$$

and by the definition of α^k ,

$$J\left(u^k + \left(\frac{\alpha^k}{\alpha}\right)(\bar{u}^k - u^k)\right) - J(u^k) > \left(\frac{\alpha^k}{\alpha}\right) \beta J'(u^k)(\bar{u}^k - u^k), \tag{6.5}$$

whenever $\alpha^k < 1$. Applying the mean value theorem to the left hand side of (6.5), we have

$$J'(\hat{u}^k)(\bar{u}^k - u^k) - J'(u^k)(\bar{u}^k - u^k) > (\beta - 1)J'(u^k)(\bar{u}^k - u^k), \tag{6.6}$$

or

$$\begin{aligned} 0 < J'(u^k)(\bar{u}^k - u^k) &< \frac{1}{1 - \beta} |J'(\hat{u}^k) - J'(u^k)| \|\bar{u}^k - u^k\| \\ &\leq \frac{1}{1 - \beta} |J'(\hat{u}^k) - J'(u^k)| (\|\bar{u}^k\|_\infty + \|u^k\|_\infty) \\ &\leq \frac{2C}{1 - \beta} |J'(\hat{u}^k) - J'(u^k)|, \end{aligned} \tag{6.7}$$

where C is an upper bound for the set $\{\|u\|_\infty : u \in \Omega\}$.

Noting that $\{J(u^k)\}$ is a bounded decreasing sequence, we have

$$J(u^k) - J(u^{k+1}) \rightarrow 0, \text{ as } k \rightarrow \infty. \tag{6.8}$$

It thus follows from (6.8) and (6.4) that

$$\alpha^k J'(u^k)(\bar{u}^k - u^k) \rightarrow 0, \text{ as } k \rightarrow \infty. \quad (6.9)$$

To complete the proof, we can show as in Theorem IV.6.2 that

$$J'(u^k)(\bar{u}^k - u^k) \rightarrow 0, \text{ as } k \rightarrow \infty. \quad (6.10)$$

Thus by the uniform continuity of $J'(\cdot)$ and the fact that $\|u^k - u^*\|_\infty \rightarrow 0$, we have $J'(u^k)(u^k) \rightarrow J'(u^*)(u^*)$. Thus by (6.9),

$$\lim_{k \rightarrow \infty} J'(u^k)(\bar{u}^k) = J'(u^*)(u^*). \quad (6.11)$$

By the definition of \bar{u}^k given in Step 3, we have $J'(u^k)(\bar{u}^k) \leq J'(u^k)(u)$ for all $u \in \Omega$. Thus it follows from (6.11) that $J'(u^*)(u^*) \leq J'(u^*)(u)$ for all $u \in \Omega$. In other words, u^* satisfies the necessary condition for optimality, and the proof is complete.

REMARK. Theorem 6.1 assumes that the generated sequence of admissible controls $\{u^k\}$ has an accumulation point in the strong topology of $L_\infty(G, U)$. However, there is no guarantee of the existence of such an accumulation point. In order to improve upon Theorem 6.2, one can introduce (as in [11, Chapter V]) the notion of relaxed controls. The topology associated with the space of relaxed controls is a more natural one in the sense that it is weak enough to ensure the existence of accumulation points for any infinite sequence. Yet it is strong enough to ensure that the accumulation points satisfy a meaningful necessary condition for optimality of the associated relaxed optimal control problem. For details, we refer the reader to [11].

7. Acknowledgements

The author is thankful to the referee for helpful comments.

This work was supported in part by a grant from the UNCC Foundation.

References

- [1] L. Cesari, *Optimization — Theory and applications* (Springer-Verlag, New York, 1983).
- [2] M. A. Kazemi, "Optimal control of systems governed by partial differential equations with integral inequality constraints", *Nonlinear Analysis, Theory, Methods, Applications* **8** (1984) 1409–1425.
- [3] M. A. Kazemi, "A method of successive approximations for optimal control of distributed parameter systems", *J. Math. Anal. Appl.* **133** (1988) 484–497.

- [4] M. A. Kazemi, " L_p -solution for a class of partial differential equations", *J. Math. Anal. Appl.* **171** (1992) 14–26.
- [5] H. H. Lee, *Heterogeneous reactor design* (Butterworth Publishers, 1985).
- [6] S. J. Leese, "Convergence of gradient methods for optimal control problems", *J. Optim. Theory Appl.* **21** (1977) 329–337.
- [7] A. T. Lukyanov and S. Ya. Serovalski, "The method of successive approximations in a problem of optimal control of a non-linear parabolic system", *U.S.S.R. Comput. Math. and Math. Phys.* **24** (1984) 23–30.
- [8] A. F. Oguny and W. H. Ray, "Optimal control policies for tubular reactors experiencing catalyst decay", *AIChE. J.* **17** (1971) 43–51.
- [9] B. N. Pshenichny and Yu. M. Danilin, *Numerical methods in external problems* (Mir, Moscow, 1978).
- [10] D. W. Reid, "On the computational methods for the optimal control of distributed parameter systems", Ph. D. Thesis, University of New South Wales, Australia, 1980.
- [11] K. L. Teo and Z. S. Wu, *Computational methods for optimizing distributed systems* (Academic Press, 1984).
- [12] O. V. Vasil'ev and V. Srochoko, "Optimization of a class of controlled processes with distributed parameters", *Siberian Math. J.* **14** (1988) 466–470.
- [13] Z. S. Wu and K. L. Teo, "A conditional gradient method for an optimal control problem involving a class of nonlinear second-order hyperbolic partial differential equations", *J. Math. Anal. Appl.* **91** (1983) 376–393.