JOURNAL OF OPTIMIZATION THEORY AND APPLICATIONS: Vol. 22, No. 1, MAY 1977

Existence of Optimal Controls for the Diffusion Equation

D. A. WILSON[1] AND J. E. RUBIO[2]

Communicated by G. Leitmann

Abstract. The existence is considered of a boundary control which drives a system governed by the one-dimensional diffusion equation from the zero state to a given final state, and at the same time minimizes a given functional. The problem is first modified to one in which the minimum is sought of a functional defined on a set of Radon measures. The existence of a minimizing measure is demonstrated, and it is shown that this measure may be approximated by a piecewise constant control. Finally, conditions are given under which a minimizing measurable control exists for the unmodified problem.

Key Words. Optimal control, Radon measures, existence theory, diffusion equation.

1. Introduction

In this note, we consider the existence of a class of optimal controls for the one-dimensional diffusion equation

$$y_{xx}(x, t) = y_t(x, t), \qquad (x, t) \in (0, 1) \times (0, T), \tag{1}$$

with boundary conditions

$$y_x(0, t) = 0, \qquad t \in [0, T],$$
$$y_x(1, t) = u(t), \qquad t \in [0, T],$$
$$y(x, 0) = 0, \qquad x \in [0, 1],$$

[1] Lecturer, Department of Electrical and Electronic Engineering, University of Leeds, Leeds, England.
[2] Lecturer, School of Mathematics, University of Leeds, Leeds, England.

where $u(t)$, $t \in [0, T]$, is the control. The control u will be termed admissible if it is a measurable function on $[0, T]$ and

(a) $u(t) \in [-1, 1]$ a.e. for $t \in [0, T]$,
(b) $y(x, T) = g(x)$, a.e. for $x \in [0, 1]$.

$g \in L_2(0, 1)$ is then the desired final state. The set of all admissible controls, which is assumed nonempty, will be denoted by U.

The control problem consists of finding a $u \in U$ which minimizes the functional

$$J(u) = \int_0^T f^0(t, u(t))\, dt,$$

where $f^0 \in C(\Omega)$, the space of continuous functions on

$$\Omega = [0, T] \times [-1, 1]$$

with the uniform topology.

In Section 2, the above control problem is restated in terms of a moment problem, which is then modified by admitting Radon measures on Ω as solutions. This approach automatically guarantees the existence of a minimizing solution. In Section 3, it is shown that the optimal measure may be approximated by a piecewise constant control. Section 4 deals with the existence of solutions to the unmodified original problem.

2. Modified Control Problem

We consider the solution of Eq. (1), in the sense defined by Fattorini and Russell (Ref. 1), in which case

$$y(x, T) = \int_0^T u(t)\, dt + \sum_1^\infty 2(-1)^n \int_0^T \exp[-n^2\pi^2(T-t)]u(t)\, dt \cos(n\pi x)$$

$$= \sum_0^\infty \int_0^T \psi_n(t, u(t))\, dt \cos(n\pi x),$$

where

$$\psi_0(t, u(t)) = u(t),$$

$$\psi_n(t, u(t)) = 2(-1)^n \exp[-n^2\pi^2(T-t)]u(t),$$

$$t \in [0, T], \qquad n = 1, 2, \ldots.$$

Since $g \in L_2(0, 1)$, it possesses a half-range Fourier series

$$\sum_0^\infty a_n \cos(n\pi x).$$

Hence, the control problem reduces to finding a measurable control

$$u(t) \in [-1, 1], \qquad t \in [0, T],$$

which satisfies

$$\int_0^T \psi_n(t, u(t))\, dt = a_n, \qquad n = 0, 1, 2, \ldots, \tag{2}$$

and minimizes

$$\int_0^T f^0(t, u(t))\, dt. \tag{3}$$

In general, a minimizing solution to the problem defined by (2)–(3) may not exist; thus, following the work of Young (Ref. 2), we replace this problem by one in which the minimum of a linear functional is sought over a set of Radon measures on Ω.

We notice that, for a fixed u, the mapping

$$f(\,.\,,\,.\,) \to \int_0^T f(t, u(t))\, dt$$

defines a positive linear functional on $C(\Omega)$. Thus, by the Riesz representation theorem (Ref. 3), there exists a unique positive Radon measure μ on Ω such that

$$\int_0^T f(t, u(t))\, dt = \int_\Omega f\, d\mu = \mu(f) \tag{4}$$

for all $f \in C(\Omega)$, and in particular for $f = f^0$.

We now replace the original minimization problem by one in which we seek the minimum of $\mu(f^0)$ over a set Q of positive Radon measures Ω. Measures in Q are required to have certain properties which are abstracted from those satisfied by admissible controls.

First, from (4),

$$|\mu(f)| \leqslant T \sup_\Omega |f(t, u)|;$$

hence,

$$\int_\Omega d\mu \leqslant T.$$

Next, measures in Q must satisfy an abstracted version of Eq. (2):

$$\mu(\psi_n) = a_n, \qquad n = 0, 1, \ldots.$$

Note that this is possible, since

$$\psi_n \in C(\Omega), \qquad n = 0, 1, 2, \ldots.$$

Finally, suppose that $h \in C(\Omega)$ does not depend on u, that is,

$$h(t, u_1) = h(t, u_2)$$

for all $t \in [0, T]$, $u_1, u_2 \in [-1, 1]$. Then, measures in Q must satisfy

$$\int_\Omega h\, d\mu = \int_0^T h(t, u)\, dt = a_h,$$

where u is an arbitrary number in $[-1, 1]$, and a_h is the Lebesgue integral of $h(\,.\,, u)$ independent of u. This property of Q is needed in the next section in order to use a theorem due to Ghouila-Houri (Ref. 4).

To summarize, Q may be written as

$$Q = S \cap P \cap M \cap N,$$

where

$$S = \left\{\mu : \int_\Omega d\mu \leq T\right\},$$

$$P = \{\mu : \mu(f) \geq 0, f \in C(\Omega), f \geq 0\},$$

$$M = \{\mu : \mu(\psi_n) = a_n, n = 0, 1, 2, \ldots\},$$

$$N = \{\mu : \mu(h) = a_h, h \in C(\Omega), h \text{ independent of } u\}.$$

We topologize the space of all Radon measures on Ω by the weak star topology. S is then compact in this topology. M can be written as

$$M = \bigcap_{n=0}^{\infty} \{\mu : \psi_n(\mu) = a_n\} = \bigcap_{n=0}^{\infty} M_n.$$

Each M_n is the inverse image of a closed set on the real line (the single point a_n) under a continuous map. Hence, M is closed. By similar arguments, it is easy to see that both N and P are closed. Therefore, Q is compact. It is also obvious that S, P, M, and N are convex; thus, Q is convex. Therefore, by the Krein–Milman theorem (Ref. 5), Q, a compact convex set, has extreme points.

Consider now the functional I: $Q \to R$ defined by

$$I(\mu) = \int_\Omega f^0\, d\mu, \qquad \mu \in Q.$$

I is a continuous linear functional defined on a convex compact set Q, and will therefore attain its minimum at one or more of the extreme points of Q. We have shown the following proposition.

Proposition 2.1. The modified optimal control problem, which consists of finding the minimum of I over Q, possesses a minimizing solution μ_0 (say) which belongs to Q.

In the next section, we show that μ_0 may be approximated by a piecewise constant control.

3. Approximation to the Optimal Measure

With each piecewise constant control

$$u(t) \in [-1, 1], \qquad t \in [0, T],$$

we may associate a measure μ_u in $S \cap P \cap N$ which satisfies

$$\int_0^T f(t, u(t))\, dt = \int_\Omega f\, d\mu_u = \mu_u(f)$$

for all $f \in C(\Omega)$.

Let W be the set of all such measures μ_u. Then, Theorem 1 of Ref. 4 shows that, when the space of Radon measures on Ω has the weak star topology, W is dense in $S \cap P \cap N$. A basis of closed neighborhoods in this topology is given by sets

$$\{\mu : |\mu(f_n)| \leq \varepsilon,\ n = 1, 2, \ldots, k+2,\ \varepsilon > 0\},$$

where k is an integer,

$$f_n \in C(\Omega), \qquad n = 1, 2, \ldots, k+2,$$

and $\varepsilon > 0$. In any weak star neighborhood of μ_0 (the minimizing measure of Section 2), it is then possible to find a μ_u corresponding to a piecewise constant control. In particular, if

$$f_1 = f^0, \qquad f_2 = \psi_0, \ldots, f_{k+2} = \psi_k,$$

a piecewise constant control u may be found such that

$$\left| \int_0^T f^0(t, u(t))\, dt - I(\mu_0) \right| \leq \varepsilon,$$

$$\left| \int_0^T \psi_n(t, u(t))\, dt - a_n \right| \leq \varepsilon, \qquad n = 0, 1, \ldots, k.$$

Thus, using this control u, we get within ε of the minimum value $I(\mu_0)$, and we attain a final state $y(x, T)$, $x \in [0, 1]$, the Fourier coefficients of which are

$$b_n = \int_0^T \psi_n(t, u(t))\, dt, \qquad n = 0, 1, \ldots,$$

where

$$|b_n - a_n| \leqslant \varepsilon, \qquad n = 0, 1, \ldots, k.$$

Since the piecewise constant control u has its range in $[-1, 1]$ for all $t \in [0, T]$, the Fourier coefficients b_n of $y(\,.\,, T)$ satisfy

$$|b_n| \leqslant 2 \int_0^T \exp[-n^2\pi^2(T-t)]|u(t)|\, dt \leqslant 2/(n\pi)^2, \qquad n = 1, 2, \ldots.$$

Similarly since it is assumed that g is reachable with an admissible control, $|a_n|$ satisfies the same inequality as $|b_n|$. We now demonstrate the following proposition.

Proposition 3.1. Given any $\delta > 0$, we may choose k and $\varepsilon > 0$ such that

(i) $\varepsilon \leqslant \delta$,

(ii) $\int_0^1 (y(x, T) - g(x))^2\, dx \leqslant \delta$.

Proof. Using the above inequalitites on b_n and a_n,

$$\int_0^1 (y(x, T) - g(x))^2\, dx = \sum_0^L (b_n - a_n)^2 + \sum_{L+1}^{\infty} (b_n - a_n)^2$$

$$\leqslant \sum_0^L (b_n - a_n)^2 + 16 \sum_{L+1}^{\infty} 1/(n\pi)^4. \tag{5}$$

Since the last summation in (5) is the tail of a convergent series, we may choose L such that

$$16 \sum_{L+1}^{\infty} 1/(n\pi)^4 \leqslant \delta/2.$$

The integer k can now be chosen as that satisfying

$$k \geqslant \max(L, (1/2\delta) - 1). \tag{6}$$

Then,

$$16 \sum_{k+1}^{\infty} 1/(n\pi)^4 \leqslant \delta/2; \tag{7}$$

we choose

$$\varepsilon = [\delta/(2(k+1))]^{\frac{1}{2}}.$$

From (6), it follows that

$$1+k \geq 1/(2\delta),$$

from which we derive

$$[\delta/(2(k+1))]^{\frac{1}{2}} = \varepsilon \leq \delta;$$

thus, (i) is satisfied by this choice of ε. In the neighborhood defined by choosing ε and k as above, a μ_u exists which corresponds to a piecewise constant control u for which we must have

$$|b_n - a_n| \leq \varepsilon, \qquad n = 0, 1, \ldots, k;$$

hence,

$$\sum_0^k (b_n - a_n)^2 \leq (k+1)\varepsilon^2 = \delta/2.$$

Combining this last relation with (7) completes the proof of Proposition 3.1.

The piecewise constant control which approximates to μ_0 in the above series will depend on δ. From the proof of Theorem 1 of Ref. 4, it can be deduced that, in general, for small δ, the approximating control will remain constant, with values in $[-1, 1]$, for short time intervals in $[0, T]$, switching then in general rapidly from one level to another.

4. Unmodified Control Problem

In the previous section, we demonstrated that the optimal control measure μ_0 of the modified problem may be approximated by a piecewise constant control. In this section, we discuss briefly some conditions under which the original problem defined by (2) and (3) has a solution.

We make the following additional assumptions on f^0:

(i) f^0 is differentiable in $u \in (-1, 1)$ for all $t \in [0, T]$;

(ii) $f^0_u \in C(\Omega)$, that is, f^0_u exists and is uniformly continuous in the interior of Ω;

(iii) f^0 is convex in $u \in [-1, 1]$ for all $t \in [0, T]$.

Let

$$\rho = \inf_{u \in U} J(u).$$

Since $f^0 \in C(\Omega)$, J is bounded below; therefore, there exists a sequence $\{u_n\}$, $u_n \in U$, with

$$\lim_{n \to \infty} J(u_n) = \rho. \tag{8}$$

We show the following proposition.

Proposition 4.1. When $f^0 \in C(\Omega)$ satisfies the properties (i), (ii), and (iii) above, there exists a control $\bar{u} \in U$ for which

$$J(\bar{u}) = \rho.$$

Proof. Since each control in the sequence $\{u_n\}$ in (8) belongs to U,

$$\int_0^T u_n^2(t)\,dt \leq T.$$

Hence,

$$\|u_n\| \leq T^{\frac{1}{2}}, \qquad n = 1, 2, \ldots,$$

where $\|\cdot\|$ is the norm in $L_2[0, T]$. We endow $L_2[0, T]$ with the weak topology, which means that the set

$$V = \{u : \|u\| \leq T^{\frac{1}{2}}\}$$

is compact, and $\{u_n\}$ has a weakly convergent subsequence, which we again denote by $\{u_n\}$. Let

$$\lim_{n\to\infty} u_n = \bar{u} \qquad \text{(weak limit)}.$$

We claim that

(a) $\bar{u} \in L_2[0, T]$,

(b) $\int_0^T \psi_k(t, \bar{u}(t))\,dt = a_k, \qquad k = 0, 1, \ldots.$

(c) $|\bar{u}(\cdot)| \leq 1$ a.e.,

and hence that $\bar{u} \in U$.

Condition (a) follows directly from the weak compactness of V. If (b) were false for some k, then an $\varepsilon > 0$ would exist with

$$\int_0^T \exp[-k^2\pi^2(T-t)](\bar{u}(t) - u_n(t))\,dt > \varepsilon$$

for all n. But, since

$$\exp(\cdot) \in L_2[0, T],$$

this contradicts the fact that $\{u_n\}$ converges weakly to $\bar{u}$.

To prove (c), suppose that

$$|\bar{u}(\cdot)| > 1$$

on some subset of $[0, T]$ having finite measure. Let p be the function on $[0, T]$ defined by

$$p(t)=1, \qquad t \in \{s: \bar{u}(s)>1\},$$
$$p(t)=-1, \qquad t \in \{s: \bar{u}(s)<-1\},$$
$$p(t)=0, \qquad t \in \{s: |\bar{u}(s)| \leq 1\}.$$

Since $\bar{u}$ is measurable, $p \in L_2[0, T]$ and

$$\int_0^T p(t)\bar{u}(t)\, dt > \int_0^T p(t)u_n(t)\, dt$$

for all n. This contradicts the fact that $\{u_n\}$ converges weakly to $\bar{u}$. Hence, $\bar{u} \in U$.

We now show that

$$J(\bar{u}) \leq \rho = \lim_{n\to\infty} J(u_n). \tag{9}$$

This implies that, when $L_2[0, T]$ has the weak topology, J is lower semicontinuous and that

$$J(\bar{u}) = \rho.$$

The convexity and differentiability assumption on f^0 imply that (Ref. 6)

$$f^0(t, v_1) \geq f^0(t, v_2) + (v_1 - v_2) f_u^0(t, v_2)$$

for every

$$v_1, v_2 \in [-1, 1], \qquad t \in [0, T].$$

Hence,

$$\int_0^T f^0(t, u_n(t))\, dt \geq \int_0^T f^0(t, \bar{u}(t))\, dt + \int_0^T (u_n(t) - \bar{u}(t)) f_u^0(t, \bar{u}(t))\, dt.$$

Therefore,

$$\rho = \lim_{n\to\infty} J(u_n) = \lim_{n\to\infty} \int_0^T f^0(t, u_n(t))\, dt$$
$$\geq J(\bar{u}) + \lim_{n\to\infty} \int_0^T (u_n(t) - \bar{u}(t)) f_u^0(t, \bar{u}(t))\, dt. \tag{10}$$

By assumption, $f_u^0 \in C(\Omega)$, and hence f_u^0 is bounded on Ω. Since t, $\bar{u}(t)$, $t \in [0, T]$, are both measurable,

$$f_u^0(\cdot, \bar{u}(\cdot)) \in L_2[0, T]$$

(see, for example, Ref. 7). Since $\{u_n\}$ converges weakly to $\bar{u}$, the last limit in (10) reduces to zero and the required result of (9) has been demonstrated. The proof of Proposition 4.1 is complete.

5. Discussion

We have considered the existence of a class of optimal controls for the one-dimensional diffusion equation. In the first problem (Section 2), the only requirement on the function f^0 in (3) was that it was continuous on Ω. Then, the existence of a minimizing measure was demonstrated. In the second problem (Section 4), additional convexity and differentiability conditions were put on f^0, and the existence of a minimizing control $\bar{u} \in U$ was demonstrated. Of course, when f^0 satisfies these additional conditions, we may still define a modified control problem, as in Section 2, and obtain an optimal measure μ_0. The question then arises as to the relation between $I(\mu_0)$ and $J(\bar{u})$. Since each $u \in U$ can be identified with a measure, we have

$$J(\bar{u}) \geqslant I(\mu_0).$$

Again, a piecewise constant control may be found which approximates to μ_0; however, this control does not exactly satisfy the moments given by (2), and thus does not belong to U. It may be that, in general, $I(\mu_0)$ is strictly less than $J(\bar{u})$; and, only if we relax our requirement of reaching the final state g exactly, may we get arbitrarily close to $I(\mu_0)$ using a measurable control.

References

1. FATTORINI, H. O., and RUSSELL, D. L., *Exact Controllability Theorems for Linear Parabolic Equations in One Space Dimension*, Archive for Rational Mechanics and Analysis, Vol. 4, pp. 272–292, 1971.
2. YOUNG, L. C., *Calculus of Variations and Optimal Control Theory*, W. B. Saunders Company, Philadelphia, Pennyslvania, 1969.
3. RUDIN, W., *Real and Complex Analysis*, McGraw-Hill Book Company, New York, New York, 1966.
4. GHOUILA-HOURI, A., *Sur la Généralization de la Notion de Commande d'un Systeme Guidable*, Revue Française d'Automatique, Informatique et Recherche Operationelle, Vol. 1, pp. 7–32, 1967.
5. ROBERTSON, A. P., and ROBERTSON, W. J., *Topological Vector Spaces*, Cambridge University Press, Cambridge, England, 1973.
6. WALSH, G. R., *Methods of Optimization*, John Wiley and Sons, London, England, 1975.

7. EWING, G. M., *Calculus of Variations with Applications*, Norton, New York, New York, 1969.