Journal of the Korean Statistical Society (2001), **30**: *3, pp 367–385*

Characterization of the Smoothest Density with Given Moments

Changkon Hong[1]

ABSTRACT

In this paper, we characterize the smoothest density with prescribed moments. Hong and Kim(1995) proved the existence and uniqueness of such a density. We introduce the general optimal control problem and prove some theorems on the characterization of the minimizer using the optimal control problem techniques.

Keywords: Moment, Optimal control problem, Roughness penalty, Sobolev space

1. Introduction

In the physical sciences, the moments for finite discrete distributions have been used for a long time. The modern theory of moments begins in the late nineteenth century with the work of Markov(1898), Stieltjes(1884) and Tchebysheff (1874). Modern versions of this theory are in Kreĭn and Nudel′man(1977) and Akhiezer(1961). The method of moments in statistical theory began about the same time with the work of Pearson(1894) and Edgeworth(1886, 1887).

Let c_i be the i-th moment of a probability measure μ and let M_n denote the convex set of all possible first n moments from probability measures on $[a, b]$, that is,

$$\mathsf{M}_n = \{(c_1, \cdots, c_n) \mid \text{for all probability measure } \mu\}$$

It is well known that for every $\mathbf{c} = (c_1, \cdots, c_n) \in \mathsf{M}_n$ there exist finite discrete measures with these moments.

In this paper, we consider the problem of finding the smoothest density with given moments $c_1, \cdots, c_n$. Without loss of generality we will assume $[a, b] = [0, 1]$. We formulate the problem as follows;

[P]

[1]Department of statistics, Pusan National University, Pusan 609-735, Korea

$$\begin{array}{lll} \textit{Minimize} & J(g) \quad \textit{on} \quad H & \\ \textit{subject to}: & L_i g = \int_0^1 t^i f(t)\,dt = c_i\ , & i = 0, \cdots, n \\ & \text{and } g(t) \geq 0 \quad \forall t \in [0,1], & \end{array}$$

where $c_1, \cdots, c_n$ are given moments, J is a roughness penalty functional of g and H is a normed linear space of functions on [0,1]. We only consider the penalty functional $J(g) = \int_0^1 (g^{(m)}(t))^2\,dt$ for $m \geq 1$. The natural space H for this penalty functional will be Sobolev space W_m^2 defined as:

$$\begin{aligned} W_m^2 &:= W_m^2[0,1] \\ &= \{g \text{ on } [0,1] \mid g^{(i)} \text{ is absolutely continuous}, i = 0, \cdots, m-1, \\ &\qquad \text{and} \quad g^{(m)} \in L^2[0,1] \}, \end{aligned}$$

with inner product $<,>$,

$$< f, g > = \sum_{i=0}^{m-1} f^{(i)}(1) g^{(i)}(1) + \int_0^1 f^{(m)}(t) g^{(m)}(t)\,dt.$$

(cf. Adams (1975)). In the followings we use the notation [P] to refer our optimizaton problem. Hong and Kim(1995) show that there exist infinitely many densities in W_m^2 with the given moments $\mathbf{c} = (c_1, \cdots, c_n)$ as long as $\mathbf{c}$ is an interior point of M_n. They also show that there exists unique minimizer for [P].

Section 2 contains some necessary conditions for the minimizer. In section 3, we introduce the general optimal control problem and formulate our optimization problem as an optimal control problem. We prove some lemmas and theorems about the minimizers for optimization problems in the form of optimal control problem. Finally, in section 4 we state and prove the main theorem on the characterization of our minimizer.

2. Some necessary conditions for the minimizer

In this section, we will derive some necessary conditions for the minimizer. Consider the optimization problem : Optimize the objective functional J defined on a normed linear space $(X, ||\cdot||)$ with the constraint set $M = \{x \in X \mid g_i(x) = 0,\ i = 1, \cdots, k\}$, where g_i's are functionals defined on X. Below we assume that all the functionals, $J, g_1, \cdots, g_k$, are Fréchet differentiable with Fréchet derivative $J', g_1', \cdots, g_k'$.

Definition *We say that $x_0 \in M$ is a regular point of the constraints $g_i = 0$, $i = 1, \cdots, k$, if $g_1'(x_0), \cdots, g_k'(x_0)$ are linearly independent.*

When X is a Hilbert space, every bounded linear functional on X can be represented as an inner product with an element in X. In other words, let X^* be the space of all bounded linear functionals on X, then for every given $x^* \in X^*$ there exist unique $y \in X$ such that $x^*(x) =< y, x >$, $\forall x \in X$. This$y$ is the Riesz representer of the functional x^*. The Riesz representer of the bounded linear functional $g'(x)$ is called the gradient of g at x, denoted by $\nabla g(x)$. Note $g'(x)(h) =< \nabla g(x), h >$, $\forall h \in X$. We can easily check that the linear independence of the $g_i'(x_0)$'s is equivalent to that of the $\nabla g_i(x_0)$'s. The following proposition can be found in many optimization books. (cf. Luenberger(1968), pp.187-189 and Hestenes(1966), p.36).

Proposition 2.1. *Suppose that x_0 is a minimizer and is a regular point of the constraints. Then there exist unique Lagrange multipliers $\lambda_1, \cdots, \lambda_k$ such that*

$$J'(x_0)\,(h) + \sum_{i=0}^{k} \lambda_i g_i'(x_0)\,(h) = 0, \qquad \forall h \in X.$$

Proposition 2.1 plays a very important role in finding necessary conditions for the minimizers in optimization problems with constraint sets like M. Using this proposition one can find what the minimizer for [P] would look like.

Theorem 2.1. *Suppose that $n \geq m - 1$. If the minimizer for [P] is positive on an interval, then in this interval it is a polynomial of degree $\leq 2m + n$.*

Proof : Let g_* be the minimizer for [P]. Suppose g_* is positive on $[\alpha, \beta] \subset [0, 1]$. Consider the restricted Sobolev space $W^2_{m,0}[\alpha,\beta] = \{g \in W^2_m[\alpha,\beta] \mid g^{(i)}(\alpha) = g^{(i)}(\beta) = 0, \ \forall\, i = 0, \cdots, m-1\}$. We can extend any h in$W^2_{m,0}[\alpha,\beta]$ to W^2_m. Define $\bar{h}$ as follows:

$$\bar{h}(t) = \begin{cases} h(t) & \text{if } t \in [\alpha, \beta] \\ 0 & \text{otherwise.} \end{cases}$$

Then $\bar{h}(t)$ belongs to W^2_m. Define a functional G on $W^2_{m,0}[\alpha,\beta]$ as:

$$G(h) = J(f_* + \bar{h}).$$

Let us consider a new optimization problem $[P_{\alpha,\beta}]$:

$$\begin{array}{llll} \textit{Minimize} & G(h) & on & W^2_{m,0}[\alpha,\beta] \\ \textit{Subject to}: & G_i(h) = 0, & i = 0, \cdots, n, & \end{array}$$

where $G_i = L_i(\bar{h}) = \int_\alpha^\beta t^i h(t)\, dt$.

We claim that this problem has a unique minimizer and it is $h_* = 0$. The existence and uniqueness can readily be shown as in Hong and Kim(1995). Suppose

$h_* \neq 0$ is the solution to this problem. Then $g_* + \delta\bar{h}_*$ obviously belongs to W_m^2 and satisfies the moment constraints $L_i g = c_i$, $i = 0, \cdots, n$. Since g_* is positive on $[\alpha, \beta]$, $g_* + \delta\bar{h}_*$ is nonnegative on [0,1] for sufficiently small δ. This shows that $g_* + \delta\bar{h}_*$ belongs to the constraint set S of the problem [P]. Choose a sufficiently small $\delta \in (0,1)$ so that $f_* + \delta\bar{h}_* \in S$. Then by strict convexity of J in S,

$$\begin{aligned} J(g_* + \delta\bar{h}_*) &= J((1-\delta)g_* + \delta(g_* + \bar{h}_*)) \\ &< (1-\delta)J(g_*) + \delta J(g_* + \bar{h}_*) \\ &= (1-\delta)G(0) + \delta G(h_*) \\ &\leq G(0) \\ &= J(g_*). \end{aligned}$$

The last inequality holds because h_* is the minimizer for problem $[P_{\alpha,\beta}]$ by assumption. This contradicts the optimality of g_*, and our claim was proved. Since $G_i'(0)h = \int_0^1 t^i h(t)\, dt$, the linear independence of $G_i'(0)$'s is equivalent to that of the functions t^i's. And the latter is obvious. Thus we have shown that 0 is regular point of the constraints $G_i = 0$, $i = 0, \cdots, n$. By Proposition 2.1 there exists unique multipliers $\lambda_0, \cdots, \lambda_n$ such that

$$G'(0) + \sum_{i=0}^{n} \lambda_i G_i'(0) = 0,$$

or equivalently,

$$\nabla G(0) + \sum_{i=0}^{n} \lambda_i \nabla G_i(0) = 0. \tag{2.1}$$

For $h \in W_{m,0}^2[\alpha, \beta]$,

$$G'(0)h = 2\int_\alpha^\beta g_*^{(m)}(t) h^{(m)}(t)\, dt.$$

There exists unique $\phi \in \Pi_{2m-1}[0,1]$ such that $g_*(t) + \phi(t)$ restricted on $[\alpha, \beta]$ belongs to $W_{m,0}^2[\alpha, \beta]$. In the space $W_{m,0}^2[\alpha, \beta]$, the inner product $<, >_{\alpha,\beta}$ is given by $< g, h >_{\alpha,\beta} = \int_\alpha^\beta g^{(m)}(t) h^{(m)}(t)\, dt$. Since $< \phi, h >_{\alpha,\beta} = 0$ for all $h \in W_{m,0}^2[\alpha, \beta]$, $G'(0)h = < g_* + \phi, h >_{\alpha,\beta}$ for all $h \in W_{m,0}^2[\alpha, \beta]$, that is, $g_* + \phi$ is the gradient $\nabla G(0)$ of the functional G at 0. On the other hand,

$$\begin{aligned} G_i'(0)h &= \int_\alpha^\beta t^i h(t)\, dt \\ &= \int_\alpha^\beta g_i^{(m)}(t) h^{(m)}(t)\, dt, \end{aligned}$$

where $g_i(t) = (-1)^m(\prod_{j=1}^{2m}(i+j))^{-1}t^{2m+i}$. There exits $\phi_i \in \Pi_{2m-1}[0,1]$ such that $\nabla G_i(0) = g_i + \phi_i$. Equation (2.1) yields

$$g_* = \phi + \sum_{i=0}^{n} \lambda_i(g_i + \phi_i) \quad \text{on} \quad [\alpha, \beta].$$

Therefore, g_* is a polynomial of degree $\leq 2m + n$ on $[\alpha, \beta]$. □

By the same argument as in the above proof, one can easily show that the following corollary holds.

Corollary 2.1. *Without the nonnegativity constraint, the problem [P] has a unique minimizer which is a polynomial of degree $\leq 2m+n$ on the whole interval [0,1].*

Let *Supp*(g) be the support of g, then one can prove the following theorem about the minimizer g_* for [P].

Theorem 2.2. *Let g_* be the minimizer for [P]. Then there exists a unique polynomial ϕ_* of degree $\leq n$ such that*

$$g_*^{(2m)}(t) = \phi_*(t), \qquad \textit{for any} \quad t \in \textit{Supp}(g_*)$$

Proof : Assume that g_* is positive on $E_{(k)} = \cup_{i=1}^{k}[\alpha_i, \beta_i]$, where (α_i, β_i)'s are disjoint subintervals of [0,1]. We will show that $g_*^{(2m)}(t) = \phi_*(t)$ on $E_{(k)}$, where $\phi_*(t)$ is a polynomial of degree $\leq n$. Consider the product space $H = W_{m,0}^2[\alpha_1, \beta_1] \times \cdots \times W_{m,0}^2[\alpha_k, \beta_k]$ with norm $\| h \|_H = \max(\| h_1 \|_1, \cdots, \| h_k \|_k)$, where $h = (h_1, \cdots, h_k) \in H$ and $\| h_i \|_i = \| h_i^{(m)} \|_{L^2[\alpha_i, \beta_i]}$. Then H is also a Banach space.(See Devito(1978), p.30). Define $\bar{h}_i$ and G as follows:

$$\bar{h}_i(t) = \begin{cases} h_i(t) & \text{if } t \in [\alpha_i, \beta_i] \\ 0 & \text{otherwise,} \end{cases}$$

and

$$G(h) = J(g_* + \bar{h}_1 + \cdots + \bar{h}_k),$$

where $h = (h_1, \cdots, h_k) \in H$. Consider another optimization problem $[P_k]$:

$$\textit{Minimize} \quad G(h) \quad \textit{on} \quad H$$
$$\textit{subject to}: \quad G_i(h) = L_i(\textstyle\sum_{j=1}^{k} \bar{h}_j) = 0, \quad i = 0, \cdots, n$$

Similarly as in the proof of Theorem 2.1, one can easily show that the problem $[P_k]$ has the unique minimizer 0. Also can one show that 0 is a regular point of the constraints $G_i(h) = 0, i = 0, \cdots, n$. By Proposition 2.1 there exist unique multipliers $\lambda_0, \cdots, \lambda_n$, such that

$$G'(0) + \sum_{j=0}^{n} \lambda_j G_j'(0) = 0. \tag{2.2}$$

For $h \in H$,

$$G'(0)h = \sum_{j=1}^{k} \int_{\alpha_j}^{\beta_j} g_*^{(m)}(t) h_j^{(m)}(t)\, dt.$$

Choose ϕ in $\Pi = \Pi_{2m-1}[\alpha_1, \beta_1] \times \cdots \times \Pi_{2m-1}[\alpha_k, \beta_k]$, such that

$$(g_* \delta_{(\alpha_1,\beta_1)}, \cdots, g_* \delta_{(\alpha_k,\beta_k)}) + \phi \quad \in H,$$

where

$$\delta_A(t) = \begin{cases} 1 & \text{if} \quad t \in A \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\nabla G(0) = (g_* \delta_{(\alpha_1,\beta_1)}, \cdots, g_* \delta_{(\alpha_k,\beta_k)}) + \phi. \tag{2.3}$$

Similarly,

$$\nabla G_i(0) = (g_i \delta_{(\alpha_1,\beta_1)}, \cdots, g_i \delta_{(\alpha_k,\beta_k)}) + \phi_i, \tag{2.4}$$

for suitably chosen $\phi_i \in \Pi$ and

$$g_i(t) = (-1)^m \prod_{j=1}^{2m} (i+j)^{-1} t^{2m+i}. \tag{2.5}$$

By using the Equations (2.2) $\sim$ (2.5) and differentiating both sides of Equation (2.2) $2m$ times, we get

$$g_*^{(2m)}(t) = \sum_{i=0}^{n} \lambda_i t^i, \qquad \forall t \in E_{(k)}.$$

This proves the theorem. $\square$

Let

$W_m^\infty = \{g$ on $[0,1] \mid g^{(i)}$ is absolutely continuous, $i = 0, \cdots, m-1,$ and $g^{(m)} \in L^\infty[0,1] \}$.

From Theorem 2.2, we can say that $g_*^{(m)}$ is a bounded function in [0,1]. This makes it sufficient to investigate functions in W_m^∞ only, in order to find the minimizer for [P]. Let us consider the problem:

$$\begin{array}{llll} \textit{Minimize} & J(g) & \text{on} & W_m^\infty \\ \textit{subject to}: & L_i g = c_i, & i = 0, \cdots, n, & \\ & \text{and } g(t) \geq 0, & \forall t \in [0,1]. & \end{array}$$

Since the minimizer g_* for [P] belongs to W_m^∞, this new problem is equivalent to [P]. From now on, we will consider this new problem instead of the original problem [P] and refer this as [P]. Accordingly the constraint set S will be:

$$S = \{g \in W_m^\infty \mid \text{ i) } L_i g = c_i, \quad i = 0, \cdots, n, \quad \text{ii) } g(t) \geq 0, \quad \forall t \in [0,1] \ \}.$$

The reason we consider the space W_m^∞ instead of W_m^2 is that the subset B_1 of W_m^∞, $B_1 = \{g \in W_m^\infty \mid g(t) \geq 0, \}$ has interior points, while the subset B_2 of W_m^2, $B_2 = \{g \in W_m^2 \mid g(t) \geq 0, \}$ has no interior point. Later we will see that it is very important that this subset B_1 has interior points, since this fact makes it possible to use fundamental theorems in mathematics.

3. Formulation of the problem as an optimal control problem

Problem [P] can be thought as an 'optimal control problem'. We introduce the problem of optimal control. Let $x : [0,1] \to R^{n_1}$ be a vector valued function. For notational convenience, we use $\dot{x}(t)$ to denote $\frac{d}{dt}x(t)$ and G_u to denote $\frac{\partial}{\partial u}G$, the partial derivative of G with respect to u.

3.1. Introduction to optimal control problems

The dynamic behavior of a system to be controlled is assumed to be described by a set of ordinary differential equations of the form

$$\dot{x}(t) = f(x(t), u(t), t) \qquad a.e. \ t \in [0,1], \tag{3.1}$$

where x is an n_1-vector function on [0,1] called the state variable and u is an n_2-vector function on [0,1] called the control variable. Equation (3.1) is called the state equation. Since the state equation (3.1) must only hold almost everywhere on [0,1], the differential equation is interpreted as the integral equation

$$x(t) = x(0) + \int_0^t f(x(s), u(s), s) \, ds.$$

For this integral equation to make sense, $x(t)$ must be absolutely continuous and

$$\phi(t) = f(x(t), u(t), t) \ \in L^1[0,1]. \tag{3.2}$$

Suppose that f is continuously Fréchet differentiable. A sufficient condition for Equation (3.2) to hold is that $u(t) \in L^{\infty}[0,1]$. Suppose that we are interested in the problems where the initial states $x(0)$ and the terminal state $x(1)$ are fixed and constraints on the control u and the state x are imposed. The following three types of constraints can be imposed;

Control constraints:	$u(t) \in U$	*a.e.*	$0 \le t \le 1$,
Mixed state control constraints:	$G_1(x(t), u(t), t) \le 0$	*a.e.*	$0 \le t \le 1$,
State constraints:	$G_2(x(t), t) \le 0$		$0 \le t \le 1$.

It may seem strange to distinguish between the 2-nd and the 3-rd types, because the 3-rd one is a special case of the 2-nd. But in order for some principles to be developed, the full rank condition of G_{1u} should be assumed as long as the constraint $G_1(x(t), u(t), t) \le 0$ contains x, u explicitly. G_2 does not include the control variable u explicitly, and because of this some important theorems can still be derived even though full rank condition of G_{2u} is obviously violated. Let $L^{\infty}[0,1]^m$ be the space of functions $\nu : [0,1] \to R^m$, which is measurable and essentially bounded. For every ν in $L^{\infty}[0,1]^m$, the ∞-norm is defined by :

$$\| \nu \|_{\infty} := ess\ sup\ \| \nu(t) \|,$$

where $\| \ \|$ is the Euclidian vector norm on R^m. $L^{\infty}[0,1]^m$ with the ∞-norm is a Banach space. Analogously, the space $W_1^{\infty}[0,1]^n$ is a Banach space when equipped with the norm

$$\| x \|_{W_1^{\infty}} = \max\{\| x \|_{\infty}, \| \dot{x} \|_{\infty}\} \ \ for \ \ all \ \ x \in W_1^{\infty}[0,1]^n.$$

(cf. Adams(1975) and Kirsch et al.(1978)).

A general form of optimal control problems can be formulated as follows: Determine a control function $\hat{u} \in L^{\infty}[0,1]^{n_2}$, a state trajectory $\hat{x} \in W_1^{\infty}[0,1]^{n_1}$ which minimize the functional

$$h_0(x(0)) + \int_0^1 f_0(x(t), u(t), t)\, dt + g_0(x(1)),$$

subject to the constraints:

$\dot{x}(t) = f(x(t), u(t), t)$ a.e. $t \in [0,1]$,
$D(x(0)) = 0$,
$E(x(1)) = 0$,
$u(t) \in U$,
$G_1(x(t), u(t), t) \le 0$ a.e. $t \in [0,1]$,

$G_2(x(t),t) \leq 0,$

where

$h_0 : R^{n_1} \to R$; $f_0 : R^{n_1} \times R^{n_2} \times R \to R$;

$g_0 : R^{n_1} \times R \to R$; $D : R^{n_1} \to R^d$;

$f : R^{n_1} \times R^{n_2} \times R \to R^{n_1}$; $E : R^{n_1} \to R^q$;

$G_1 : R^{n_1} \times R^{n_2} \times R \to R^{k_1}$; $G_2 : R^{n_1} \times R \to R^{k_2}$;

$U \subset R^{n_2}$, is a convex set with nonempty interior.

For all $x \in R^{n_1}$, $u \in R^{n_2}$, $\quad \operatorname{rank} G_{1u}(x,u,t) = k_1 \quad$ a.e. $t \in [0,1]$.

The operators, $h_0, f_0, g_0, f, D, E, G_1, G_2$ are continuously Fréchet differentiable.

Some necessary conditions for the solution to this problem can be found in several books. (cf. Machielsen(1988), Tikhomirov(1986), Girsanov(1972) and Hestenes(1966)). But this form is so general that the related theorems will be very complicated. Our problem can be formulated as a rather simple optimal control problem. So we only consider rather simple optimal control problems.

[OCP]:

Determine a control function $\hat{u} \in L^\infty[0,1]^{n_2}$, a state trajectory $\hat{x} \in W_1^\infty[0,1]^{n_1}$ which minimize the functional

$$\int_0^1 f_0(x(t),u(t),t)\, dt,$$

subject to the constraints:

$$\dot{x}(t) = f(x(t),u(t),t) \qquad \text{a.e. } t \in [0,1], \tag{3.3}$$

$D(x(0)) = 0,$

$E(x(1)) = 0,$

$G(x(t),t) \leq 0 \qquad \forall t \in [0,1],$

We say that (x,u) in $W_1^\infty[0,1]^{n_1} \times L^\infty[0,1]^{n_2}$ is admissible if it satisfies all the constraints in [OCP].

We consider the problem [P] with additional end point constraints, $\mathbf{g}(0) = \mathbf{a}$ and $\mathbf{g}(1) = \mathbf{b}$, where $\mathbf{g} = (g, g^{(1)}, \cdots, g^{(m-1)})$. Later this will be referred as problem [P{**a**,**b**}]. From now on we use $x_1(t)$ to denote the elements in W_m^∞. We introduce more variables $x_2, \cdots, x_{m+n+1}$ and u to express problem [P{**a**,**b**}] in the form of the problem [OCP]. Define differential equations

$$\begin{aligned} \dot{x}_i &= x_{i+1}, \quad i = 1, \cdots, m-1, \\ \dot{x}_m &= u, \\ \dot{x}_{m+j+1} &= t^j x_1, \quad j = 0, \cdots, n. \end{aligned}$$

Then the nonnegativity constraint will be expressed as $-x_1 \leq 0$.
Let $x = (x_1, \cdots, x_{m+n+1})^T$, then the end point conditions and the moment constraints become $x(0)^T = (\mathbf{a}^T, \mathbf{0}^T)$, $x(1)^T = (\mathbf{b}^T, \mathbf{c}^T)$. Now one can see that problem [P{**a**,**b**}] is an optimal control problem of the form [OCP] with

$$\begin{aligned} f_0(x,u,t) &= u^2, \\ f(x,u,t) &= (x_2, \cdots, x_m, u, x_1, tx_1, \cdots, t^n x_1)^T, \\ D(x)^T &= x^T - (\mathbf{a}^T, \mathbf{0}^T), \qquad \mathbf{0} \text{ is of dimension } (n+1) \times 1, \\ E(x)^T &= x^T - (\mathbf{b}^T, \mathbf{c}^T), \\ G(x,t) &= -x_1, \\ n_1 &= d = q = m+n+1, \\ n_2 &= k_2 = 1. \end{aligned}$$

Let [TP] be the problem [P{**a**,**b**}] stated in the form of problem [OCP]. If $(\hat{x}, \hat{u})$ is the solution to [TP], then $\hat{x}_1$ is the solution to [P{**a**,**b**}].

3.2. Some related theorems

Now we introduce some important theorems for [OCP]. Here an important role is played by the Hamiltonian, which is defined as

$$H(x,u,\rho,\lambda,t) = \rho f_0(x,u,t) + \lambda^T f(x,u,t).$$

In the theorems below the notation $[t]$ is used to replace $\hat{x}(t)$, $(\hat{x}(t), \hat{u}(t), t)$, $(\hat{x}(t), \hat{u}(t), \hat{\rho}, \hat{\lambda}(t), t)$, or $(\hat{x}(t), \hat{u}(t), \hat{\rho}, \hat{\lambda}(t), \hat{\xi}(t), t)$ appropriately, and $NBV[0,1]^k$ is the normalized space of k-vector valued functions on [0,1] of bounded variation (cf. Luenberger(1968) and Nachbin(1981)).

Theorem 3.1. *If $(\hat{x}, \hat{u})$ is a solution to [OCP], then there exit a real number $\hat{\rho} \geq 0$, and vector-valued functions $\hat{\lambda} \in NBV[0,1]^{n_1}, \hat{\xi} \in NBV[0,1]^{k_2}$ and vectors $\hat{\sigma} \in R^d, \hat{\mu} \in R^q$, not all zero, such that*

$$\begin{aligned} \hat{\lambda}(t_1)^T - \hat{\lambda}(t_0)^T &= -\int_{t_0}^{t_1} H_x[t]\,dt - \int_{t_0}^{t_1} d\hat{\xi}(t)^T G_x[t], \quad \forall t_0, t_1 \in [0,1], &(3.4) \\ \hat{\lambda}(0)^T &= -\hat{\sigma}^T D_x[0], &(3.5) \\ \hat{\lambda}(1)^T &= \hat{\mu}^T E_x[1], &(3.6) \\ H_u[t](u - \hat{u}(t)) &\geq 0 \quad \textit{for all } u \in R^{n_2} \quad \textit{a.e. } 0 \leq t \leq 1, &(3.7) \end{aligned}$$

$\hat{\xi}_i(t)$ *is nondecreasing on [0,1] and is constant on intervals where* $G_i[t] < 0$, (3.8)

where $H_x[t] = (H_{x_1}[t], \cdots, H_{x_{n_1}}[t]),$
$G_x[t] = (G_{x_1}[t], \cdots, G_{x_{n_1}}[t]).$

Proof: See Machielsen(1988), Theorem 3.11 in Chapter 3. (Also see Grisanov(1972) and Kurcyusz(1976) for the details of the proof of the theorem). □
The above theorem can be derived by considering the Lagrangian,

$$L_1(x, u, \rho, \lambda, \eta, t) = H(x, u, \rho, \lambda, t) + \eta\, G(x, t).$$

(The relationship between $\hat{\xi}$ in Theorem 3.1 and $\hat{\eta}$ will be $\dot{\hat{\xi}}(t) = \hat{\eta}$.) This method is called the direct adjoining approach. For the case when $m = 1$, stronger results can be derived by considering the following type of Lagrangian :

$$L_2(x, u, \rho, \lambda, \xi, t) = H(x, u, \rho, \lambda, t) + \xi\, \phi(x, u, t),$$

where $\phi(x(t), u(t), t) = \dot{G}(x(t), t) = G_x[t]\, f_u[t]$. This method is called the indirect adjoining approach. In the following theorem we assume that ϕ_u has full row rank.

Theorem 3.2. *Let* $(\hat{x}, \hat{u})$ *be the solution to [OCP]. Then there exit a real number* $\hat{\rho} \geq 0$, *and vector-valued functions* $\hat{\lambda}, \hat{\xi} \in NBV[0,1]^{k_2}$ *such that*

(i) $\hat{\lambda}(t)$ *is continuous on* $0 \leq t \leq 1$, *and* $\hat{\xi}(t)$ *is continuous on each interval of continuity of* $\hat{u}(t)$. *Furthermore on each interval where* $\hat{u}(t)$ *is continuous, the following equations holds*

$$\begin{aligned} \dot{x}(t)^T &= L_{2\lambda}[t] = f[t], \\ \dot{\hat{\lambda}}(t)^T &= -L_{2x}[t], \\ L_{2u} &= \mathbf{0}_{1\times n_2}. \end{aligned}$$

(ii) $\hat{\xi}_i(t)$ *is nondecreasing on* $0 \leq t \leq 1$, *and is constant on every interval in which* $G_i[t] < 0$. *It is continuous whenever* $\hat{u}(t)$ *is continuous and is continuous at every point at which* $\phi[t] = \dot{G}[t]$ *is discontinuous.*

(iii) At no point t on [0,1] are the multipliers $\hat{\rho} = 0$, $\hat{\lambda}(t)^T = c^T G_x[t]$.

Proof: See Hestenes(1966), Theorem 2.1. in Chapter 8. □

3.3. Characterization of the minimizer with additional end point constraints

We now use Theorem 3.1 to find a necessary condition for the minimizer for [TP]. The Hamiltonian H for [TP] is,

$$\begin{aligned} H(s,u,\rho,\lambda,t) &= \rho f_0(x,u,t)+\lambda^T f(x,u,t) \\ &= \rho u^2+\sum_{i=1}^{m-1}\lambda_i x_{i+1}+\lambda_m u+\sum_{j=0}^{n}\lambda_{m+j+1}t^j\,x_1. \end{aligned}$$

Let $(\hat{x},\hat{u})$ be a solution to [TP], of which existence and uniqueness can be proved as in Hong and Kim(1995). Equation (3.4) yields

$$\lambda(t)^T-\lambda(0)^T=-\int_0^t H_x[s]\,ds-\int_0^t d\hat{\xi}(s)\,G_x[s]\quad \forall t\in[0,1]. \tag{3.9}$$

Since $H_x[t]=(\sum_{j=0}^n\lambda_{m+j+1}t^j,\lambda_1,\cdots,\lambda_{m-1},\mathbf{0}^T)$ and $G_x[t]=(1,\mathbf{0}^T)$,

$$\hat{\lambda}_1(t) = \hat{\lambda}_1(0)-\int_0^t\sum_{j=0}^{n}\hat{\lambda}_{m+j+1}(s)s^j\,ds+\int_0^t d\hat{\xi}(s), \tag{3.10}$$

$$\hat{\lambda}_i(t) = \hat{\lambda}_i(0)-\int_0^t\hat{\lambda}_{i-1}(s)\,ds,\qquad i=2,\cdots,m, \tag{3.11}$$

$$\hat{\lambda}_i(t) = \hat{\lambda}_i(0),\qquad i=m+1,\cdots,m+n+1. \tag{3.12}$$

Equations (3.5), (3.6) and (3.7) yield

$$\hat{\lambda}(0) = -\hat{\sigma}, \tag{3.13}$$

$$\hat{\lambda}(1) = \hat{\mu}, \tag{3.14}$$

$$0 = H_u[t] \tag{3.15}$$

$$= 2\,\hat{\rho}\,\hat{u}(t)+\hat{\lambda}_m. \tag{3.16}$$

claim : $\hat{\rho}>0$.
Proof of the claim : Suppose $\hat{\rho}=0$, then by Equation (3.16) $\hat{\lambda}_m=0$. This and Equation (3.11) yield

$$\hat{\lambda}_i=0,\qquad i=1,\cdots,m. \tag{3.17}$$

Since $\hat{\xi}\in NBV[0,1]$, it is almost everywhere differentiable. Let $\hat{\eta}=\dot{\hat{\xi}}(t)$. Then the condition (3.8) leads to

$$\hat{\eta}(t) = 0,\qquad \text{if } \hat{x}_1(t)>0, \tag{3.18}$$

$$\hat{\eta}(t) \ge 0,\qquad \forall t. \tag{3.19}$$

By differentiating both sides of Equation (3.10), and using Equation (3.12) and (3.13), one can get

$$\dot{\hat{\lambda}}_1 = \hat{\eta}(t) + \sum_{j=0}^{n} \hat{\sigma}_{m+j+1} t^j. \tag{3.20}$$

From Equations (3.17) and (3.20), it follows that $\hat{\eta}$ is a polynomial of degree n and is given by

$$\hat{\eta}(t) = -\sum_{j=0}^{n} \hat{\sigma}_{m+j+1} t^j. \tag{3.21}$$

If $\hat{\eta}(t)$ is the zero function, then

$$\hat{\sigma}_{m+j+1} = 0, \qquad j = 0, \cdots, n. \tag{3.22}$$

Equations (3.12), (3.13), (3.17) and (3.19) together result in

$$\hat{\lambda} = 0, \quad \hat{\sigma} = 0, \quad \hat{\mu} = 0, \tag{3.23}$$

and this contradicts that not all of $\hat{\rho}, \hat{\lambda}, \hat{\xi}, \hat{\sigma}$ and $\hat{\mu}$ are zero. If $\hat{\eta}(t)$ is not the zero function, then it has at most n zeros. Equation (3.18) says that $\hat{x}_1(t) > 0$ at most n points, which is impossible. This shows that $\hat{\rho} > 0$. □

Without loss of generality one can assume that $\hat{\rho} = 1/2$. Equation (3.16) yields

$$\hat{u}(t) = -\hat{\lambda}_m.$$

By Equations (3.10), (3.11), (3.12), (3.13), and (3.16), we get

$$\begin{aligned}
\hat{u}(t) &= \hat{\sigma}_m - \int_0^t \hat{\lambda}_{m-1}(s)\, ds \\
&= \sum_{i=0}^{m-1} (-1)^i \hat{\sigma}_{m-i} t^i \\
&\quad +(-1)^m \int_0^t \int_0^{t_1} \cdots \int_0^{t_{m-1}} \sum_{j=0}^{n} \hat{\sigma}_{m+j+1} t_m^j \, dt_m \, dt_{m-1} \cdots dt_1 \\
&\quad +(-1)^m \int_0^t \int_0^{t_1} \cdots \int_0^{t_{m-2}} \hat{\xi} \, dt_{m-1} dt_{m-2} \cdots dt_1 \\
&= \phi(t) + (-1)^m I_{m-1}(\hat{\xi}),
\end{aligned}$$

where

$$\phi(t) = \sum_{i=0}^{m-1} (-1)^i \hat{\sigma}_{m-1} t^i + (-1)^m \sum_{j=0}^{n} \frac{\hat{\sigma}_{m+j+1} t^{m+i}}{\prod_{j=1}^{m} (i+j)},$$

$$I_m(g)(t) = \int_0^t \int_0^{t_1} \cdots \int_0^{t_{m-1}} g(t_m)\, dt_m \cdots dt_1$$

with $I_0(g)(t) = g(t)$. The state variable $\hat{x}(t)$ can be determined by the differential equations

$$\dot{\hat{x}}(t) = f(\hat{x}(t), \hat{u}(t), t)$$

with boundary conditions

$$\hat{x}(0)^T - (\mathbf{a}^T, \mathbf{0}^T) = \mathbf{0}^T$$

and

$$\hat{x}(1)^T - (\mathbf{b}^T, \mathbf{c}^T) = \mathbf{0}^T.$$

Thus we have proved the following theorem.

Theorem 3.3. *If $(\hat{x}, \hat{u})$ is the solution to [TP], then $\hat{u}(t)$ is of the from*

$$\hat{u}(t) = \phi(t) + (-1)^m I_{m-1}(\hat{\xi})(t), \tag{3.24}$$

where $\phi(t)$ is a polynomial of degree $\leq n+m$ and $\hat{\xi}(t)$ is a nondecreasing function on [0,1] and is constant on each interval where $\hat{x}(t) > 0$. Furthermore, we can conclude that $\hat{u}(t)$ is $m-2$ times continuously differentiable for $m \geq 2$, i.e., $\hat{u}(t) \in C^{m-2}[0,1]$.

Remark: When $m = 1$, using Theorem 3.2 we can show that $\hat{\xi}(t)$ is continuous on the whole interval [0,1]. Thus $\hat{u}(t)$ is continuous on $[0,1]$.
The condition that $\hat{u}(t)$ is of the form (3.24) is not only necessary condition , but also sufficient condition for the solution to [TP]. We have the following theorem.

Theorem 3.4. *Any admissible (x^*, u^*), such that*

$$u^*(t) = \phi(t) + (-1)^m I_{m-1}(\xi^*)(t)$$

is a solution to [TP], where ϕ is a polynomial of degree $\leq n+m$ and ξ^* is a nondecreasing function on [0,1] and constant on each interval where $x_1^*(t) > 0$.

Before we prove the theorem we prove the following useful lemma. Let $J_0(u) = \int_0^1 (u(t))^2\, dt$.

Lemma 3.1. *An admissible $(\hat{x}, \hat{u})$ is the solution to [TP] if and only if $J_0'(\hat{u})(u - \hat{u}) \geq 0$, $\forall u$ such that (x, u) is admissible for some x.*

Proof : If $(\hat{x}, \hat{u})$ is solution to problem [TP], then $J_0(\alpha u + (1-\alpha)\hat{u}) \geq J_0(\hat{u})\ \forall \alpha \in (0,1)$, for all admissible (x, u). This implies $J_0'(\hat{u})(u - \hat{u}) \geq 0$. For sufficiency, suppose that $J_0'(\hat{u})(u - \hat{u}) \geq 0$, then

$$\begin{aligned} J_0(u) - J_0(\hat{u}) &= J_0(u - \hat{u}) + 2\int_0^1 \hat{u}(t)\,(u(t) - \hat{u}(t))\,dt \\ &\geq 0, \end{aligned}$$

thus $(\hat{x}, \hat{u})$ is the solution to [TP]. □

Remark: Without the nonnegativity constraint, the condition $J_0'(\hat{u})(u - \hat{u}) = 0$ is sufficient and necessary for $\hat{u}$ to be the minimizer.

Proof of Theorem 3.4 : Let (x, u) be any admissible pair. If u^* is given as in the theorem, then

$$\begin{aligned} J_0'(u^*)(u - u^*) &= \int_0^1 (u(t) - u^*(t))\,u^*(t)\,dt \\ &= \sum_{i=0}^{m-1} (-1)^i (x_{m-i}(t) - x_{m-i}^*(t))\,u^{*(i)}(t)\,|_0^1 \\ &\quad +(-1)^m \int_0^1 (x_1(t) - x_1^*(t)\,du^{*(m-1)} \\ &= (-1)^m \int_0^1 (x_1(t) - x_1^*(t))\,du^{*(m-1)} \end{aligned}$$

The last equality holds because of the endpoint conditions. Now

$$\int_0^1 (x_1(t) - x_1^*(t))\,\psi(t)\,dt = 0, \quad \forall \psi \in \Pi_n[0,1].$$

Since $x_1(t) \geq 0$ and $\hat{\xi}$ is nondecreasing,

$$\int_0^1 x_1(t)\,d\hat{\xi}(t) \geq 0$$

and since $\xi^*(t)$ is constant on intervals where $x_1^*(t) > 0$,

$$\int_0^1 x_1^*(t)\,d\xi^*(t) = 0.$$

So we have

$$\begin{aligned} J_0'(u^*)(u - u^*) &= (-1)^m \int_0^1 (x_1(t) - x_1^*(t)\,\phi^{(m)}(t)\,dt \\ &\quad +(-1)^{2m} \int_0^1 (x_1(t) - x_1^*(t))\,d\xi^*(t) \\ &= \int_0^1 x_1(t)\,d\xi^*(t) \\ &\geq 0. \end{aligned}$$

By Lemma 3.1, (x_*, u_*) is the solution to [TP]. □

4. Characterization of the minimizer

Now we think about the difference between the problem [P] and the problem [TP]. If we know the values $\mathbf{x}_*(0)$ and $\mathbf{x}_*(1)$, where x_* is the minimizer for [P] and $\mathbf{x}_* = (x_*, \ldots, x_*^{(m-1)})^T$, then finding the minimizer for [TP] with $\mathbf{a} = \mathbf{x}_*(0)$ and $\mathbf{b} = \mathbf{x}_*(1)$ is equivalent to finding the minimizer for [P]. But we don't know those values. Therefore, the sufficient condition in Theorem 3.4 is not directly helpful to find the minimizer for [P]. The following necessary and sufficient condition suggests a useful way to find the minimizer for [P].

Theorem 4.1. *Let f_* satisfy all the constraints in the problem [P]. Then f_* is the minimizer for [P] if and only if $f_*^{(m)}$ is of the form (3.24) and satisfies the following end point conditions;*

$$f_*^{(i)}(0) = f_*^{(i)}(1) = 0 \quad i = m, \cdots, 2m-1. \tag{4.1}$$

Furthermore, $f_* \in C^{2m-2}[0,1]$.

Before we prove the theorem we restate Lemma 3.1 in terms of problem [P].

Lemma 4.1. *Let S be the constraint set of the problem [P]. Then $\hat{f}(t) \in S$ is the solution to [P] if and only if $J'(\hat{f})(f-\hat{f}) \geq 0, \forall f \in S$. Without the nonnegativity constraint, the necessary and sufficient condition would be $J'(\hat{f})(f-\hat{f}) = 0, \forall f \in W_m^2$ satisfying the moment constraints.*

Proof of Theorem 4.1 :
(Sufficiency): Let $J(f) = \int_0^1 (f^{(m)}(t))^2\, dt$. Then

$$\begin{aligned} J'(f_*)(f-f_*) &= \int_0^1 (f^{(m)}(t) - f_*^{(m)}(t))\, f_*^{(m)}(t)\, dt \\ &= \sum_{i=0}^{m-1} (-1)^i (f^{(m-i-1)}(t) - f_*^{(m-i-1)}(t))\, f_*^{(m+i)}(t) \,\Big|_0^1 \\ &\quad +(-1)^m \int_0^1 (f(t) - f_*(t))\, df_*^{(2m-1)}(t) \\ &= (-1)^m \int_0^1 (f(t) - f_*(t))\, df_*^{(2m-1)}(t) \end{aligned}$$

The last equation holds because of condition (4.1). By the same reasoning as in the proof of Theorem 3.4,

$$J'(f_*)(f - f_*) \geq 0.$$

This proves the sufficiency.

(Necessity): It suffices to show that the end point conditions (4.1) hold. First, we will show that (4.1) holds for the minimizer g_* for [P] without the nonnegativity constraint. We already know that g_* is a polynomial of degree $\leq 2m+n$ and by Lemma 4.1 $J'(g_*)(g - g_*) = 0$, $\forall g \in W_m^2[0,1]$ such that $L_i g = c_i$. Therefore we have

$$\sum_{i=0}^{m-1} (-1)^{m-i-1} (g^{(i)}(t) - g_*^{(i)}(t))\, g_*^{(2m-i-1)}(t) \mid_0^1 = 0,$$

$\forall g \in W_m^2[0,1]$ such that $L_i g = c_i$. Let $g = \sum_{j=1}^{l} b_j t^j$, where $l \geq 2m+n$. Let $g_{0\,0}, \cdots, g_{0\,m-1}, g_{1\,0}, \cdots, g_{1\,m-1}$ be any fixed values. Let $\mathbf{g}_0 = (g_{0\,0}, \cdots, g_{0\,m-1})^T$ and $\mathbf{g}_1 = (g_{1\,0}, \cdots, g_{1\,m-1})^T$. Consider the linear system

$$\begin{aligned} g^{(j)}(0) &= g_{0\,j}, \quad j = 0, \cdots, m-1, \\ g^{(j)}(1) &= g_{1\,j}, \quad j = 0, \cdots, m-1, \\ L_i g &= c_i \quad\ i = 0, \cdots, n \end{aligned}$$

or equivalently

$$\begin{pmatrix} \mathbf{I}_{m\times m}\ \mathbf{0} \\ \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{b} = \begin{pmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \mathbf{c} \end{pmatrix}, \tag{4.2}$$

where $(j+1)$-th row $\mathbf{A}_{j+1}$ of $\mathbf{A}$ is given by

$$\mathbf{A}_{j+1} = (\mathbf{0}_{1\times j}, j!, \frac{(j+1)!}{1!}, \cdots, \frac{l!}{(l-j)!}), \quad j = 0, \cdots, m-1$$

and $(\mathbf{B})_{ij} = 1/(i+j-1)$, $i = 1, \cdots, n+1$, $j = 1, \cdots, m$. It can be shown that the matrix in the left hand side of (4.2) has full row rank provided $l \geq 2m+n$ and thus Equation (4.2) has at least one solution. Now choose $\mathbf{g}_0$ and $\mathbf{g}_1$ in such a way that

$$\begin{aligned} (g_{0\,j_0} - g_*^{(j_0)}(0))\, sgn[(-1)^{m-j_0-1} g_*^{(2m-j_0-1)}(0)] &> 0, \\ (g_{1\,j_1} - g_*^{(j_1)}(0))\, sgn[(-1)^{m-j_1-1} g_*^{(2m-j_1-1)}(0)] &> 0, \end{aligned}$$

where $j_0, j_1 \in \{0, \cdots, m-1\}$ are such that $g_*^{(2m-j_0-1)}(0) \neq 0,\ \ g_*^{(2m-j_1-1)}(1) \neq 0$. If there exist such j_0 or j_1, then $J'(g_*)(g - g_*)$ is strictly positive, which is impossible. This proves the theorem without the nonnegativity constraint. Now we apply this result to the problem [P] with the nonnegativity constraint. Consider the functions f which satisfies the moment constraints and $f(t) = 0,\ \ \forall t \in (Supp(f_*))^c$. Note that f need not be nonnegative. Since $f_*^{(2m)}(t) = \phi^{(m)}(t),\ \forall t \in Supp(f_*)$ and $\phi^{(m)}$ is a polynomial of degree $\leq n$, we can use the same argument as in the case without nonnegativity constraint. This proves the theorem. □

References

Adams, R. A. (1975). *Sobolev Spaces,* Academic Press, New York.

Akhiezer, N. I. (1961). *The Classical Moment Problem and Some Related Topics in Analysis*, Fizmatgiz, Moscow; English Transl., Hafner, New York. MR 27 #4028; 32 #1518.

Devito, C. L. (1978). *Functional Analysis*, Academic Press, New York.

Edgeworth, F. Y. (1886). "The law of error and the elimination of chance," *Philosophical Magazine, 5th series* 21, 308-324.

Edgeworth, F. Y. (1887). "The empirical proof of the law of error," *Philosophical Magazine, 5th series* 24, 330-342.

Girsanov, I. V. (1972). *Lectures on Mathematical Theory of Extremum Problems*, Lecture Notes in Economics and Mathematical Systems, vol. 67. Springer-Berlag, New York.

Hestenes, M. R. (1966). *Calculus of Variations and Optimal Control Theory*, Wiley, New York.

Hong, C. and Kim, C. (1995). "The existence and uniqueness of the smoothest density with prescribed moments," *Journal of the Korean Statistical Society*, 24, 233-242.

Kirsch, A., Warth, W., and Werner J. (1978). *Notwendige Optimalitätsbedingungen und ihre Anwendung*, Lecture Notes in Economics and Mathematical System, vol. 152, Springer-Verlag, New York.

Kreĭn, M. G., Nudel'man, A. A. (1977). *The Markov Problem and Extremal Problems*, Translations of Mathematical Monographs, vol. 50, American Mathematical Society, Providence, Rhode Island.

Kurcyusz, S. (1976). "On the existence and nonexistence of Lagrange multipliers in Banach spaces," *Journal of Optimization Theory and Applications*, 20, 81-110.

Luenberger, D. G. (1968). *Optimization by Vector Space Methods*, Wiley, New York.

Machielsen, K. C. P. (1988). *Numerical Solution of Optimal Control Problems with State Constraints by Sequential Quadratic Programming in Function Space*, CWI Tract, vol. 53, Centrum voor Wiskunde en Informatica (Centre for Mathematics and Computer Science), Amsterdam.

Markov, A. A. (1898). "Sur les valeurs extrèmes des intégrales et l'interpolation," *Mémoires de l'Académie des Sciences de St.-Petersburg, classe physico-mathématique*, (8) 6 (5), 69.

Pearson, K. (1894). "Contributions to the mathematical theory of evolution," *Philosophical Transactions of the Royal Society of London (A)* 185, 71-110, *Page references are to the reprint in Pearson*, 1956; 1-40.

Pearson, K. (1956). *Karl Pearson's Early Statistical Papers*, Cambridge: Cambridge University Press (first issued, 1948).

Stieltjes, T. J. (1884). "Quelques recherches sur les quadratures dites mécaniques," *Annales Scientifiques de l'Ecold Normale Supérieure* 1 (3), 409-426.

Tchebycheff, P. (1874). "Sur les valeurs limits des intégrales," *Journal des Mathématiques pures et eppliquées* 19 (2), 157-160. Collected Papers 2, No 10, 181-185.

Tikhomirov, V. M. (1986). *Fundamental Principles of the Theory of Extremal Problems*, Wiley, Chichester-New York.