Extending the Farkas Lemma Approach to Necessity Conditions to Infinite Programming

R. A. Tapia
M. W. Trosset

CRPC-TR91279
October 1991

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

Extending the Farkas Lemma Approach to Necessity Conditions to Infinite Programming

R.A. Tapia *   M.W. Trosset †

October 22, 1991

*Department of Mathematical Sciences and Center for Research in Parallel Computation, Rice University, Houston, Texas 77251-1892. This author was supported in part by NSF Coop. Agr. No. CCR-8809615, AFOSR 89-0363, DOE DEFG05-86ER25017 and ARO 9DAAL03-90-G-0093.

†Consultant, P.O. Box 40993, Tucson, Arizona 85717-0993. This author was supported in part by NSF Coop. Agr. No. CCR-8809615 as a visiting member of the Center for Research in Parallel Computation, Rice University, Houston, Texas 77251-1892, July 1991.

Abstract

Under mild assumptions, the classical Farkas lemma approach to Lagrange multiplier theory is extended to an infinite programming formulation. The main result generalizes the usual first-order necessity conditions to address problems in which the domain of the objective function is Hilbert space and the number of constraints is arbitrary. The result is used to obtain necessity conditions for a well-known problem from the statistical literature on probability density estimation.

Key words: Lagrange multiplier theory, Farkas lemma, infinite programming, mathematical programming.

1 Introduction

Kuhn and Tucker [1] developed a Lagrange multiplier theory for mathematical programming problems in which the domain of the objective function is Euclidean space and the constraint functionals are indexed by a finite set. The standard development of the Kuhn-Tucker theory, explicated and popularized by Fiacco and McCormick ([4], Chapter 2), invokes the classical Farkas lemma to generate a vector (the Lagrange multipliers) that can be viewed as a "weighting" of the finite set of constraints. For years we have taught this material, and each time we cannot avoid pondering the extent to which this development of Lagrange multiplier theory depends on the finite dimensionality of Euclidean space and the finiteness of the constraint set. Somewhat recently, our interest was enhanced when we learned of an important infinite programming problem in the statistics literature for which we could state a formal generalization of the Kuhn-Tucker conditions with no known theoretical justification for doing so. In the present study of Lagrange multiplier theory, we have not only succeeded in generalizing the Farkas lemma approach, but have also acquired new insight into the essential features of that approach.

The objective of this paper is to extend the classical Farkas lemma approach to mathematical programming problems in which the domain of the objective function is Hilbert space and the constraint functionals are indexed by an arbitrary set. Our approach carefully mimics the finite programming development. It is based on a generalized Farkas lemma, and replaces the Lagrange multiplier vector with a measure on the (possibly infinite) index set. If this measure is absolutely continuous, then it can be represented as a (density) function on the index set. Because our point of view may seem unnatural to some readers otherwise familiar with Lagrange multiplier theory, we briefly digress to motivate it.

Consider vectors $x_1, \ldots, x_k \in \mathrm{R}^n$, scalar weights $u_1, \ldots, u_k \in \mathrm{R}^n$, and the weighted sum

$$\sum_{i \in I} u_i x_i ,$$

where the index set $I = \{1, \ldots, k\}$. By defining a measure μ on the Borel sets of R^n that concentrates on $\{x_1, \ldots, x_k\}$ and satisfies $\mu(\{x_i\}) = u_i$, we can write

$$\sum_{i \in I} u_i x_i = \int_{\mathrm{R}^n} x \mu(dx) .$$

Thus, a set of weights can be viewed as a measure and a weighted sum can be viewed as a (Lebesgue) integral with respect to that measure. When the weights are nonnegative and sum to unity, μ is a probability measure and probabilists call the integral an expectation.

Now consider the index map $i \mapsto x_i$, which embeds I in R^n. The measure μ induces a measure u on the subsets of I by $u(\{i\}) = \mu(\{x_i\})$. This allows us to further write

$$\sum_{i \in I} u_i x_i = \int_{\mathrm{R}^n} x \mu(dx) = \int_I x_i u(di) \ ;$$

hence, our point of view that a set of weights is a measure on an index set. It is this perspective that will lead to a manageable statement of generalized first-order conditions.

The flavor of our generalization of Lagrange multiplier theory is not entirely new. Semi-infinite programming is also concerned with problems in which the constraint functionals are indexed by an infinite set. However, the domain of the objective function is still assumed to be Euclidean space. A multiplier theorem of precisely the sort that we seek has been published by Goberna *et al.* [2]. Unfortunately, their result depends critically on the finite dimensionality of Euclidean space.

It should be noted that a number of authors have published multiplier theorems in very abstract settings. The standard formulation is that of Guignard [3], who derived both necessity and sufficiency conditions for the problem

$$
\begin{aligned}
\text{maximize} \quad & \psi(x) \\
\text{subject to} \quad & x \in C \subset X \\
& a(x) \in B \subset Y ,
\end{aligned}
$$

where X and Y are real Banach spaces and $\psi : X \to (-\infty, +\infty)$ and $a : X \to Y$ are Fréchet differentiable. Guignard's multiplier is an element of the topological dual space of Y, and her entire approach is markedly different from ours.

The primary purpose of the present paper is perhaps pedagogical. That is, we wish to demonstrate that by (i) interpreting the vector of Lagrange multipliers as a measure on the index set of constraints and by (ii) utilizing tools from functional analysis and probability theory, the standard finite-dimensional approach to Kuhn-Tucker theory (Fiacco and McCormick [4]) can be successfully generalized to infinite programming in Hilbert space. This exercise, however, is not entirely pedagogical, for we also believe that there are important infinite programming problems to which the new theory can be profitably applied. Therefore, after in Section 2 deriving first-order necessity conditions for general infinite programming problems, in Section 3 we will consider results that facilitate the use of these conditions. In Section 4, by way of an example, we will also apply this theory to obtain necessity conditions for a constrained optimization problem from the statistical literature on probability density estimation. However, we will defer to another paper an investigation of the statistical consequences of these conditions.

2 Main Theorem

We begin with a real Hilbert space X with inner product $\langle \cdot\, , \cdot \rangle$. By the general nonlinear programming problem — problem (NLP) for short — we mean the constrained optimization problem

$$\begin{array}{lll} \text{maximize} & f(x) & \\ \text{subject to} & g_\alpha(x) \geq 0 & \forall \alpha \in I \\ & h_\beta(x) = 0 & \forall \beta \in J\ , \end{array}$$

where $f, g_\alpha, h_\beta :\ X \rightarrow (-\infty, +\infty)$. We assume that the index sets I and J have corresponding sigma fields $\mathbf{I}$ and $\mathbf{J}$ such that the pairs $(I, \mathbf{I})$ and $(J, \mathbf{J})$

are measure spaces. Measures on these spaces will be denoted by t, u, λ, etc. At times, we will also endow I and J with topologies. Typically, I and J will be subsets of Euclidean space. For each $x \in X$, we define the index subset $I_0(x) := \{\alpha \in I : g_\alpha(x) = 0\}$.

We assume that $f, g_\alpha, h_\beta \in C^1(X)$. For each $x \in X$, the sets $\nabla A_0(x) := \{\nabla g_\alpha(x) : \alpha \in I_0(x)\}$ and $\nabla B(x) := \{\nabla h_\beta(x) : \beta \in J\}$ are assumed to be Borel measurable. We also assume that the index maps $\alpha \mapsto \nabla g_\alpha(x)$ and $\beta \mapsto \nabla h_\beta(x)$ are Borel bimeasurable functions. This will enable measures on I and J to induce measures on X, and also conversely. Measures on X will be denoted $\nabla F, \nabla G$, etc.

For technical reasons, we will sometimes further assume that the functions g_α and h_β are elements of a real Hilbert space Γ. In that event, we will assume that the sets $A_0(x) := \{g_\alpha : \alpha \in I_0(x)\}$ and $B := \{h_\beta : \beta \in J\}$ are Borel measurable. We will also assume that the index maps $\alpha \mapsto g_\alpha$ and $\beta \mapsto h_\beta$ are Borel measurable functions. This will enable measures on I and J to induce measures on Γ. Such measures will be denoted by F, G, etc.

In this section we will derive necessary conditions for a point $x^* \in X$ to be a local solution of problem (NLP). To do so, we generalize Fiacco's and McCormick's [4] presentation of the Kuhn-Tucker theory for the finite-dimensional case. The key to this generalization is the concept of the expectation of a measure on a Hilbert space. Toward this end, in what follows H will denote a real Hilbert space with inner product $(\cdot\,,\cdot)$. Following Parthasarathy [5] we have the following definition.

Definition 2.1 *Let μ be a measure on H. If the linear functional $L(y) := \int (y,x)\mu(dx)$ is continuous, then the expectation of μ, which we denote by $\int x\mu(dx)$, is defined to be the Riesz representer of L.*

At this point it will be of value to introduce some basic notation. Let $M(K)$ denote the family of totally finite positive measures that concentrate on the set $K \subset H$, and let $M_1(K)$ denote the family of probability measures that concentrate on the set $K \subset H$. We are interested in the sets of expectations

$$C(K) = \left\{ \int x\mu(dx) : \mu \in M(K) \right\}$$

and

$$C_1(K) = \left\{ \int x\mu(dx) : \mu \in M_1(K) \right\} .$$

The set $C_1(K)$ is essentially the convex hull of K, and the set $C(K)$ is essentially the half-cone generated by $C_1(K)$. It should be clear that $C(K)$ and $C_1(K)$ are convex. In the next section we will demonstrate that $C_1(K)$ is also compact. The closedness of $C(K)$ will be of fundamental importance in our theory. In the next section we will construct a condition which guarantees that $C(K)$ is closed. However, for the moment we will assume that it is closed.

We now generalize a famous result. It is possible that our lemma can be obtained as a special case of the very abstract Farkas lemma recently given by Swartz [7]. However, after some thought we are convinced that a demonstration of this fact would require considerably more effort and be less illuminating than the elementary proof we give below, specific to Hilbert space.

Lemma 2.1 *(Generalized Farkas Lemma): Let H denote a real Hilbert space with inner product $(\cdot\,,\cdot)$. Let $x_0 \in H$. Assume:*

A1: $K \subset H$ *is compact;*

A2: $C(K)$ *is closed.*

Then the following are equivalent:

(i) $\forall\, y \in H\,,\ (y, x) \geq 0 \qquad \forall\, x \in K$ *entails* $(y, x_0) \geq 0\,;$

(ii) $\exists\, \mu \in \mathbf{M}(K)$ *such that* $x_0 = \int x\mu(dx)\,.$

Proof: Suppose that $x_0 = \int x\mu(dx)$. If $(y, x) \geq 0\ \forall\, x \in K$, then it follows from the definition of expectation and the positivity of μ that $(y, x_0) =$

$(y, \int x\mu(dx)) = \int(y, x)\mu(dx) \geq 0$. This proves that (ii) implies (i).

To demonstrate the converse, let $\hat{x}_0$ denote the projection of x_0 into $C(K)$, which we know to be closed and convex. Then $\hat{x}_0$ solves the optimization problem $\min_{x \in C(K)} \frac{1}{2}\|x - x_0\|^2$; hence, $(\hat{x}_0 - x_0, x - \hat{x}_0) \geq 0 \; \forall x \in C(K)$.

Since $C(K)$ is a half-cone, if $\hat{x}_0 \neq 0$, then $(1+r)\hat{x}_0 \in C(K)$ for r in a neighborhood of zero. Then $\phi(r) := \|\hat{x}_0 + r\hat{x}_0 - x_0\|^2$ is minimized by $r = 0$; hence, $\phi'(0) = (\hat{x}_0 - x_0, \hat{x}_0) = 0$. Of course, if $\hat{x}_0 = 0$, then this conclusion is immediate. It follows that $(\hat{x}_0 - x_0, x) \geq 0 \; \forall \, x \in C(K)$, and in particular $\forall \, x \in K$. Taking $y = \hat{x}_0 - x_0$, we infer from (i) that $(\hat{x}_0 - x_0, x) \geq 0$. Then $0 \leq (\hat{x}_0 - x_0, x_0) = -(\hat{x}_0 - x_0, \hat{x}_0 - x_0) = -\|\hat{x}_0 - x_0\|^2 \leq 0$, so $x_0 = \hat{x}_0$ and we have derived (ii). □

Associated with problem (NLP) is the generalized Lagrangian gradient

$$\ell'(x, u, \lambda) := \nabla_x f(x) - \int_I \nabla_x g_\alpha(x) u(d\alpha) + \int_J \nabla_x h_\beta(x) \lambda(d\beta)\,,$$

which is guaranteed to exist if the sets $\nabla A(x)$ and $\nabla B(x)$ are compact and the measures u and λ are totally finite. Our goal is to derive necessary conditions for solving problem (NLP) that involve this expression. We are now in a position to characterize some of these conditions.

Suppose that x is a feasible point of problem (NLP). Let

$$\begin{aligned} Z_1(x) &:= \{z \in X : \langle z, \nabla g_\alpha(x)\rangle \geq 0 \ \forall \ \alpha \in I_0(x), \langle z, \nabla h_\beta(x)\rangle \\ &\qquad = 0 \ \forall \ \beta \in J, \langle z, \nabla f(x)\rangle \geq 0\}, \end{aligned}$$

$$\begin{aligned} Z_2(x) &:= \{z \in X : \langle z, \nabla g_\alpha(x)\rangle \geq 0 \ \forall \ \alpha \in I_0(x), \langle z, \nabla h_\beta(x)\rangle \\ &\qquad = 0 \ \forall \ \beta \in J, \langle z, \nabla f(x)\rangle < 0\} . \end{aligned}$$

Proposition 2.1 *Let $K = \nabla A_0(x^*) \cup \nabla B(x^*)$. Assume*

A1 : $\quad$ *K is compact;*

A2 : $\quad$ *$C(K)$ is closed.*

If x^* is a feasible point of problem (NLP), then the following are equivalent:

(i) $Z_2(x^*) = \phi$.

(ii) There exist totally finite measures u^* on $(I, \mathbf{I})$ and λ^* on $(J, \mathbf{J})$ such that

(a) $\ell'(x^*, u^*, \lambda^*) = 0$,

(b) $g_\alpha(x^*) \geq 0 \ \forall \ \alpha \in I$,

(c) $h_\beta(x^*) = 0 \ \forall \ \beta \in J$,

(d) $u^*(I') = 0 \;\; \forall \;\; I'$ *measurable* $\subset I \sim I_0(x^*)$,

(e) $u^*(I') \geq 0 \;\; \forall \;\; I'$*measureable* $\subset I$.

We will refer to the conditions (a)–(e) in (ii) as the generalized first-order conditions.

Proof: Assume (ii) and suppose that $z \in Z_2(x^*)$. Then

$$\begin{aligned} 0 &> \langle z, \nabla f(x^*)\rangle \\ &= \langle z, \int_I \nabla g_\alpha(x^*) u^*(d\alpha) - \int_J \nabla h_\beta(x^*)\rangle \lambda^*(d\beta)\rangle \\ &= \int_I \langle z, \nabla g_\alpha(x^*)\rangle u^*(d\alpha) - \int_J \langle z, \nabla h_\beta(x^*)\rangle \lambda^*(d\beta) \\ &\geq 0\,, \end{aligned}$$

which is a contradiction. This proves that (ii) implies (i).

Conversely, suppose that $Z_2(x^*) = \phi$. Then, if z satisfies

$$\begin{aligned} &\langle z, \nabla g_\alpha(x^*)\rangle \geq 0 \qquad \forall\, \alpha \in I_0(x^*) \\ &\langle z, \nabla h_\beta(x^*)\rangle \geq 0 \qquad \forall\, \beta \in J \\ &\langle z, -\nabla h_\beta(x^*)\rangle \geq 0 \qquad \forall\, \beta \in J\,, \end{aligned}$$

z must also satisfy $\langle z, \nabla f(x^*)\rangle \geq 0$. But this implication is (i) in Lemma 2.1, so we may conclude that

$$\nabla f(x^*) = \int_{\nabla A_0(x^*)} y \nabla F_0(dy) + \int_{\nabla B(x^*)} y \nabla F'(dy) - \int_{\nabla B(x^*)} y \nabla F''(dy)$$

$$= \int_{I_0(x^*)} \nabla g_\alpha(x^*) t_0(d\alpha) + \int_J \nabla h_\beta(x^*) t'(d\beta) - \int_J \nabla h_\beta(x^*) t''(d\beta) .$$

We now obtain conditions (a)–(e) by setting $u^* = t_0$ on $I_0(x^*), u^* = 0$ on $I \sim I_0(x^*)$, and $\lambda^* = -(t' - t'')$ on J. □

Our statement of first-order condition (a) is somewhat nontraditional. Suppose that the g_α and h_β are elements of a real Hilbert space Γ. Assuming that the indicated expectations exist (which, of course, they may not), define the generalized Lagrangian function to be

$$\ell(x, u, \lambda) := f(x) - \int_I g_\alpha(x) u(d\alpha) + \int_J h_\beta(x) \lambda(d\beta) .$$

To conform to common practice, we would write condition (a) as

$$\nabla_x \ell(x^*, u^*, \lambda^*) = \ell'(x^*, u^*, \lambda^*) = 0 .$$

The following result establishes circumstances in which this representation is legitimate.

Proposition 2.2 *Fix $x \in X$. Let u and λ denote totally finite measures on $(I, \mathbf{I})$ and $(J, \mathbf{J})$. Assume that the expectations $\bar{g} := \int_I g_\alpha u(d\alpha)$ and $\bar{h} := \int_J h_\beta \lambda(d\beta)$ both exist. If the sets of functions $A := \{g_\alpha : \alpha \in I\}$ and B are each uniformly Lipschitz continuous, then $\nabla_x \ell(x, u, \lambda) = \ell'(x, u, \lambda)$.*

Proof: We must establish that

$$\nabla \int_I g_\alpha(x)u(d\alpha) = \nabla \bar{g}(x) = \int_{\nabla A(x)} y \nabla F(dy) = \int_I \nabla g_\alpha(x) u(d\alpha)\,, \qquad (1)$$

where ∇F is the measure on X induced by u. Clearly it suffices to prove that

$$\langle \eta, \nabla \bar{g}(x)\rangle = \langle \eta, \int_{\nabla A(x)} y \nabla F(dy)\rangle \quad \forall\ \eta \in X\,. \qquad (2)$$

We note that

$$\begin{aligned}
\langle \eta, \nabla \bar{g}(x)\rangle &= \bar{g}'(x)(\eta) \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon}\{\bar{g}(x+\epsilon\eta) - \bar{g}(x)\} \\
&= \lim_{\epsilon \to 0} \frac{1}{\epsilon}\{\int_I [g_\alpha(x+\epsilon\eta) - g_\alpha(x)]u(d\alpha)\} \\
&= \lim_{\epsilon \to 0} \int_I \phi_\epsilon(x) u(d\alpha)
\end{aligned}$$

and that

$$\begin{aligned}
\langle \eta, \int_{\nabla A(x)} y \nabla F(dy)\rangle &= \int_{\nabla A(x)} \langle \eta, y\rangle \nabla F(dy) \\
&= \int_I \langle \eta, \nabla g_\alpha(x)\rangle u(d\alpha) \\
&= \int_I g'_\alpha(x)(\eta) u(d\alpha) \\
&= \int_I \lim_{\epsilon \to 0} \frac{1}{\epsilon}[g_\alpha(x+\epsilon\eta) - g_\alpha(x)]u(d\alpha) \\
&= \int_I \lim_{\epsilon \to 0} \phi_\epsilon(\alpha) u(d\alpha)\,.
\end{aligned}$$

Since the g_α's are uniformly Lipschitz continuous, i.e. $\exists\ M < \infty$ such that $|g_\alpha(y) - g_\alpha(z)| \leq M\|y - z\|\ \ \forall\ \ x, y \in X$ and $\forall\ \alpha \in I$, we have

$$|\phi_\epsilon(\alpha)| \leq \frac{M}{\epsilon}\|\epsilon\eta\| = M\|\eta\| < \infty\ .$$

Then, since u is totally finite, we can apply the Dominated Convergence Theorem to interchange $\lim_{\epsilon\to 0}$ and $\int_I$. This establishes (2); hence, (1). The identical argument establishes that

$$\nabla \int_J h_\beta(x)\lambda(d\beta) = \int_J \nabla h_\beta(x)\lambda(d\beta)\ ,$$

and the result follows. □

We now return to problem (NLP). As in the finite-dimensional case, in order to derive a necessity condition from Proposition 2.1, we must supplement the first-order conditions with a constraint qualification.

Definition 2.2 *Suppose that x^* is a feasible point of problem (NLP). We say that x^* satisfies the constraint qualification for problem (NLP) if:*

for each nonzero $z \in X$ *satisfying* $\langle z, \nabla g_\alpha(x^*)\rangle \geq 0 \ \forall\, \alpha \in I_0(x^*)$ *and* $\langle z, \nabla h_\beta(x^*)\rangle = 0 \ \forall\, \beta \in J$, $\exists \tau > 0$ *and a continuous arc* $C : [0,\tau) \to X$ *satisfying*

$$
\begin{aligned}
&C(0) = x^* \, , \\
&C'(0) = z \, , \\
&g_\alpha(C(t)) \geq 0 \quad \forall \ t \in [0,\tau) \ \textit{and} \ \forall\, \alpha \in I \, , \\
&\beta(C(t)) = 0 \quad \forall \ t \in [0,\tau) \ \textit{and} \ \forall\, \beta \in J \, .
\end{aligned}
$$

Our main result now follows precisely as in the finite-dimensional case.

Theorem 2.1 *Let* $K = \nabla A_0(x^*) \cup \nabla B(x^*)$. *Assume*

A1: K *is compact;*

A2: $C(K)$ *is closed.*

If x^* *satisfies the constraint qualification for problem (NLP), then a necessary condition for* x^* *to be a local solution of problem (NLP) is that the first-order conditions hold.*

Proof: We invoke Proposition 2.1. Suppose that x^* is a local solution and that $z \in Z_2(x^*)$. Clearly it must be that $z \neq 0$, so there exists a feasible continuous arc $C : [0,\tau) \to X$. Since x^* is a local solution, for $t > 0$

sufficiently small it must be that

$$f(C(t)) - f(C(0)) \geq 0$$

and therefore

$$\frac{1}{t}[f(C(t)) - f(C(0))] \geq 0 \, .$$

But this implies that

$$[f \circ C]'(0) = \langle \nabla f(C(0)), C'(0) \rangle = \langle \nabla f(x^*), z \rangle \geq 0 \, ,$$

which is a contradiction. □

3 Discussion of Hypotheses

Let us now examine assumptions A1 and A2 of Theorem 2.1. To begin, recall that in the finite programming case they automatically hold. Specifically, a finite set is obviously compact, and it is well-known that a finitely generated cone is closed. Hence, assumptions A1 and A2 are exactly the price one must pay to extend the Farkas lemma approach to necessity conditions from finite programming to infinite programming. Of course this extension will be of no value if we cannot find reasonable conditions that imply assumptions A1, A2, the constraint qualification, and a meaningful example that satisfies our conditions. These concerns are the subject of the remainder of this section and the next section where our example is presented.

We first consider assumption A1, the compactness assumption for the sets $\nabla A_0(x^*)$ and $\nabla B(x^*)$. Since problem (NLP) is stated without reference to these sets, it is obviously cumbersome to check hypotheses involving them. Fortunately, many problems will not require this.

Lemma 3.1 *Suppose that the g_α and h_β are elements of a real Hilbert space Γ. Fix $x \in X$ and let V_x denote evaluation at x. Suppose that the g_α and h_β are uniformly continuous and that V_x is a continuous functional on A and B. Assume that the index sets I and J have been topologized. If I and J are compact and the index maps $\alpha \mapsto g_\alpha$ and $\beta \mapsto h_\beta$ are continuous, then the sets $\nabla A_0(x)$ and $\nabla B(x)$ are compact.*

Proof: We argue in terms of the g_α. Given a sequence $\{\alpha_n\} \subset I_0(x)$, we claim that there exists $\alpha_0 \in I_0(x)$ and a subsequence $\{\alpha_{n'}\}$ such that $\nabla g_{\alpha_{n'}}(x) \to \nabla g_{\alpha_0}(x)$.

The indexing assumptions imply that A is compact. Since V_x is continuous on A, it follows that the level set $A_0(x) = \{g_\alpha : g_\alpha(x) = 0\} = \{g_\alpha : V_x(g_\alpha) = 0\}$ is closed, hence compact itself. Therefore there exists

$\alpha_0 \in I_0(x)$ and a subsequence $\{\alpha_{n'}\}$ such that $g_{\alpha_{n'}} \to g_{\alpha_0}$.

The convergence indicated is in norm. However, since the g_α are uniformly continuous, the convergence must also be uniform. But this allows us to write

$$\lim_{n' \to \infty} \nabla g_{\alpha_{n'}}(x) = \nabla \lim_{n' \to \infty} g_{\alpha_{n'}}(x) = \nabla g_{\alpha_0}(x).$$

$\square$

We now derive conditions which imply that assumption A2 holds. We first derive a technical lemma about expectations that will be used to show that K compact and $0 \notin C_1(K)$ implies A2. This lemma derives from probability theory. An excellent reference for the requisite material is Billingsley [6] .

Lemma 3.2 *Let H denote a real Hilbert space with inner product $(\cdot\,,\cdot)$. Let $\mathbf{M}_1(K)$ denote the family of probability measures that concentrate on the set $K \subset H$. If K is compact, then the set of expectations $C_1(K) := \{\int x\mu(dx) : \mu \in \mathbf{M}_1(K)\}$ is convex and compact.*

Remark: As mentioned before, the set $C_1(K)$ is essentially the convex hull of K.

Proof: Since K is compact, $\int (y, x)\mu(dx) \leq \|y\| \int \|x\|\mu(dx) \leq \|y\| \sup_{x \in K} \|x\| < \infty$; we are therefore assured that the expectations exists. The convexity of $C_1(K)$ follows immediately from the linearity of expectation.

To demonstrate compactness, consider the sequence $\{x_n = \int x\mu_n(dx) : \mu_n \in \mathbf{M}_1(K)\}$. Since K is compact, $\mathbf{M}_1(K)$ is tight. It follows from Prohorov's Theorem that there exists a weakly convergent subsequence of $\{\mu_n\}$, i.e. that there exists $\mu_0 \in \mathbf{M}_1(K)$ and a subsequence $\{\mu_{n'}\}$ such that

$$\int \phi(x)\mu_{n'}(dx) \to \int \phi(x)\mu_0(dx)$$

for all bounded continuous functions $\phi : H \to (-\infty, +\infty)$. Since K is compact, $(y, \cdot)$ is such a function; hence

$$\int (y, x)\mu_{n'}(dx) \to \int (y, x)\mu_0(dx) \quad \forall y \in H .$$

Then it must be that the Riesz representers

$$x_{n'} = \int x\mu_{n'}(dx) \to x_0 := \int x\mu_0(dx) ,$$

so the arbitrary sequence $\{x_n\}$ has a convergent subsequence. □

We now remove the restriction that the positive measures used to form expectations have a total mass of unity.

Lemma 3.3 *Let H denote a real Hilbert space with inner product $(\cdot, \cdot)$ and origin 0. Let $\mathbf{M}(K)$ denote the family of totally finite, positive measures that concentrate on the set $K \subset H$. If K is compact and $0 \notin C_1(K)$, then $C(K) := \{\int x\mu(dx) : \mu \in \mathbf{M}(K)\}$ is convex and closed.*

Remark: As mentioned before, the set $C(K)$ is essentially the half-cone generated by the convex hull of K.

Remark: The conditions that K is compact and $0 \notin C_1(K)$ are sufficient but not necessary for the conclusion. To illustrate, let S be a closed subspace of H and let $K \subset S$ be any set such that 0 is an interior point of $C_1(K) \subset S$ relative to S, e.g. $\{x \in S : \|x\| < 1\}$. Then $C(K) = S$ is automatically convex and closed. However, the simple conditions stated in the lemma have a natural analog in the finite-dimensional theory and are entirely adequate for the example of Section 4.

Proof: Writing $C(K) = \{rx : x \in C_1(K), r = [0, +\infty)\}$, it follows from the convexity of $C_1(K)$ that $C(K)$ is a convex half-cone. We claim that $C(K)$ is also closed.

Toward that end, suppose that $\{y_n\} \subset C(K)$ with $\|y_n - \bar{y}\| \to 0$. Write $y_n = r_n x_n$ with $x_n \in C_1(K)$. By the compactness of $C_1(K)$, $\{x_n\}$ contains a subsequence $\{x_{n'}\}$ with $\|x_{n'} - \bar{x}\| \to 0$ for some $\bar{x} \in C_1(K)$. Moreover, since $0 \notin C_1(K)$, $\|\bar{x}\| > 0$.

Now let $\epsilon > 0$ be arbitrary. By construction, there exists $N(\epsilon)$ such that $n' \geq N(\epsilon)$ entails $\|x_{n'} - \bar{x}\| \leq \epsilon$ and $\|r_{n'} x_{n'} - \bar{y}\| \leq \epsilon$. It follows that, if

$\epsilon < \|\bar{x}\|$ and $n' \geq N(\epsilon)$, then

$$\frac{\|\bar{y}\| - \epsilon}{\|\bar{x}\| + \epsilon} \leq r_{n'} \leq \frac{\|\bar{y}\| + \epsilon}{\|\bar{x}\| - \epsilon} \quad ,$$

so that $r_{n'} \to \bar{r} := \|\bar{y}\|/\|\bar{x}\|$. Hence,

$$\begin{aligned} \|y_{n'} - \bar{r}\bar{x}\| &= \|r_{n'}x_{n'} - r_{n'}\bar{x} + r_{n'}\bar{x} - \bar{r}\bar{x}\| \\ &\leq r_{n'}\|x_{n'} - \bar{x}\| + |r_{n'} - \bar{r}|\ \|\bar{x}\| \\ &\to 0\,. \end{aligned}$$

By the uniqueness of limits, $\bar{y} = \bar{r}\bar{x} \in C(K)$. □

Notice that the hypothesis that $0 \notin C_1(K)$ is closely related to the oft-imposed (in finite programming) condition of regularity. A feasible point x^* is said to be *regular* if the set K is linearly independent, i.e. if no finite nonzero linear combination of the constraint gradients at x^* can vanish. Our condition is somewhat stronger in one respect, but much weaker in another. On the one hand, we consider arbitrary measures (weights) on K, not just finitely supported ones. This is analogous to infinite linear combinations, hence stronger; on the other hand, we only consider probability measures (nonnegative weights totalling unity); this is analogous to convex combinations instead of linear combinations, hence weaker.

In finite programming, if x^* is a regular point, then x^* must satisfy the constraint qualification. This pleasant property does not hold in infinite programming; in fact, since the number of linearly independent gradients cannot exceed the dimension of the space X, the notion of regularity is wholly inappropriate for the case of semi-infinite programming and somewhat inappropriate in the case of infinite programming. Accordingly, we will search for other conditions that will imply the constraint qualification.

The simplest situation is the one in which all of the constraints are linear. If x^* and z are as in Definition 2.2, then the arc $C(t) = x^* + tz$ satisfies

$$\begin{aligned} C(0) &= x^* ; \\ C'(0) &= z ; \\ g_\alpha(C(t)) &= 0 \quad \forall\, t \geq 0, \quad \forall\, \alpha \in I_0(x^*) ; \\ h_\beta(C(t)) &= 0 \quad \forall\, t \geq 0, \quad \forall\, \beta \in J . \end{aligned}$$

Moreover, for each $\alpha \in I \sim I_0(x^*)$ (the nonbinding constraints), there exists $\tau(\alpha) > 0$ such that

$$g_\alpha(C(t)) \geq 0 \; \forall\, t \in [0, \tau(\alpha)).$$

If the number of nonbinding constraints is finite, then we can take $\tau = \inf_\alpha\{\tau(\alpha)\} > 0$ and the constraint qualification is automatically satisfied.

Otherwise, it may be that $\inf_\alpha\{\tau(\alpha)\} = 0$ and the constraint qualification may not hold. We are therefore content to establish that the constraint qualification holds for one important family of examples.

Both control theory and statistics abound with constraints of the sort that a function be bounded by certain values. The following result addresses the prototypical case; we hope that the method of proof will suffice for a variety of applications.

Lemma 3.4 *Let X denote a real Hilbert space of functions $x : I \to (-\infty, +\infty)$. Let g_α denote evaluation at $\alpha \in I$. If X is a proper functional Hilbert space, i.e. if the g_α are continuous, then the collection of inequality constraints*

$$g_\alpha(x) = x(\alpha) \geq 0 \quad \forall\, \alpha \in I$$

satisfies the constraint qualification.

Proof: Since the g_α are continuous and linear, $\nabla g_\alpha(x)$ exists $\forall\, x \in X$. Suppose that $x^* \in X$ is a feasible point and that a nonzero $z \in X$ satisfies

$$\langle z, \nabla g_\alpha(x^*)\rangle = g'_\alpha(x^*)(z) = z(\alpha) \geq 0 \;\; \forall\, \alpha \in I_0(x^*),$$

$$\text{i.e.} \;\; \forall\, \alpha \;\text{ such that }\; x(\alpha) = 0\,.$$

For $t > 0$, let $z^*(t) := P_t z$, where P_t denotes projection into the closed convex set

$$K(t) := \{y \in X : x^*(\alpha) + ty(\alpha) \geq 0 \quad \forall \quad \alpha \in I\} .$$

We note that the sets $K(t)$ are nested, for suppose that $t_0 < t_1$. If $0 \leq x^*(\alpha) + t_1 y(\alpha) \quad \forall \ \alpha \in I$, i.e. $y \in K(t_1)$, then

$$0 \leq \frac{t_0}{t_1} x^*(\alpha) + t_0 y(\alpha) \leq x^*(\alpha) + t_0 y(\alpha) \ \forall \ \alpha \in I ,$$

i.e. $y \in K(t_0)$. Thus, $K(t_1) \subset K(t_0)$. We also note that, if $t_n \to t_0$ and $y_0 \in K(t_0)$, then there exists $y_n \in K(t_n)$ such that $\|y_n - y_0\| \to 0$ as $t_n \to t_0$. This follows upon setting $y_n = (t_0/t_n)y_0$, and means that the point-to-set map $t \mapsto K(t)$ is open. (See Hogan [8] for an introduction to this subject in the context of mathematical programming.)

Next suppose that $t_n \downarrow t_0$. Since the $K(t)$ are nested,

$$\|z^*(t_n) - z\| \downarrow r := \lim_{n \to \infty} \|z^*(t_n) - z\|$$

and $\|z^*(t_0) - z\| \leq r$. Let $B(z, r)$ denote the closed ball of radius r centered at z. The point $z^*(t_0)$ must lie on the boundary of $B(z, r)$, for suppose that it lies in the interior. Then there exist $y_n \in K(t_n)$ such that $\|y_n - z^*(t_0)\| \to 0$ as $t_n \to t_0$ and it follows from the triangle inequality that $\lim_{n \to \infty} \|y_n - z\| < r$.

But

$$\|y_n - z\| \geq \|z^*(t_n) - z\| \quad \forall\, n\, ;$$

hence $\inf\{\|y_n - z\|\} \geq r$, which is a contradiction.

Thus, $B(z,r) \cap K(t_0) = \{z^*(t_0)\}$. Since $z^*(t_n) \in K(t_n) \subset K(t_0)$, it follows that $\|z^*(t_n) - z^*(t_0)\| \to 0$ as $t_n \downarrow t_0$. A similar argument establishes the same fact for $t_n \uparrow t_0$, and we conclude that $z^*(t)$ is a continuous arc for $t > 0$. Moreover, since $x^*(\alpha) = 0$ entails $z(\alpha) \geq 0$, z is contained in the closure of $\cup_{t>0} K(t)$. We can therefore close the arc by setting $z^*(0) = z$.

Now let $C(t) := x^* + tz^*(t)$. By construction, C is a feasible continuous arc with $C(0) = x^*$. Moreover,

$$\lim_{t\to 0} \frac{1}{t} \|C(t) - C(0) - tz\| = \lim_{t\to 0} \frac{1}{t} \|tz^*(t) - tz\| = \lim_{t\to 0} \|z^*(t) - z\| = 0\,,$$

so $C'(0) = z$. This verifies the conditions specified by Definition 2.2. □

4 An Example

We now apply our results to obtain necessity conditions for a well-known problem from the statistical literature on probability density estimation. Watson and Leadbetter [9] sought to minimize the mean integrated squared error of a kernel probability density estimator. Specifically, given independent and identically distributed random variables $X_1, \ldots, X_n$ with probability density function δ, they analyzed the optimization problem

$$\underset{K_n \in L^2(-\infty,+\infty)}{\text{minimize}} \quad E \int_{-\infty}^{\infty} \left[\frac{1}{n} \sum_{i=1}^{n} K_n(x - X_i) - \delta(x) \right]^2 dx \ .$$

It turns out that solutions are typically not everywhere nonnegative, which results in estimates that are not themselves probability densities. This is a matter of taste, but if we prefer to estimate densities with densities, then we must confront the constrained optimization problem

$$\begin{array}{ll} \underset{K_n \in L^2(-\infty,+\infty)}{\text{minimize}} & E \int_{-\infty}^{\infty} \left[\frac{1}{n} \sum_{i=1}^{n} K_n(x - X_i) - \delta(x) \right]^2 dx \\ \text{subject to} & K_n(x) \geq 0 \ \ x \in (-\infty, +\infty) \\ & \int_{-\infty}^{\infty} K_n(x) dx = 1 \ . \end{array}$$

This problem does not yield to variational methods, making it a natural candidate for the application of our multiplier theory. We proceed to formulate it in that context.

Consider the Sobolev space $H^1[\alpha_1, \alpha_2]$, which is defined by endowing the vector space

$$\{x : x^{(j)} \in L^2[\alpha_1, \alpha_2] \text{ for } j = 0, 1\}$$

with the inner product

$$\langle x, y\rangle = \sum_{j=0}^{1} \langle x^{(j)}, y^{(j)}\rangle_{L^2[\alpha_1,\alpha_2]}.$$

It should be noted that the derivatives in the definition of $H^1[\alpha_1, \alpha_2]$ are taken in the sense of distributions. It is well known that $H^1[\alpha_1, \alpha_2]$ is a proper functional Hilbert space and that each element of $H^1[\alpha_1, \alpha_2]$ is absolutely continuous. See Appendix I of Tapia and Thompson [10] for a discussion of the analogous Sobolev space, $H^1(-\infty, +\infty)$. Notice that, if $\delta \in H^1(-\infty, +\infty)$, then the restriction of δ to $[\alpha_1, \alpha_2]$ is an element of $H^1[\alpha_1, \alpha_2]$.

We now return to the problem of Watson and Leadbetter, which we reformulate as problem (WL):

$$\text{(WL)} \qquad \begin{aligned} &\underset{x_n \in X}{\text{minimize}} && f(x_n) = E\int_{-\infty}^{\infty}\left[\frac{1}{n}\sum_{i=1}^{n} \bar{x}_n(\alpha - \chi_i) - \delta(\alpha)\right]^2 d\alpha \\ &\text{subject to} && g_\alpha(x_n) = x_n(\alpha) \geq 0 \ \ \forall\, \alpha \in I \\ & && h(x_n) = \int_I x_n(\alpha)d\alpha - 1 = 0\ , \end{aligned}$$

where $I = [\alpha_1, \alpha_2]$, $X = H^1[\alpha_1, \alpha_2]$, and $\bar{x}_n$ denotes the extension of x_n to $(-\infty, +\infty)$ defined by $\bar{x}_n(\alpha) = 0$ if $\alpha \notin I$; and where the expectation is taken with respect to the independent and identically distributed random variables $\chi_i, i = 1, \ldots, n$, having probability density function $\delta \in H^1(-\infty, +\infty)$. We have modified the original problem in two ways. First, we have demanded some additional smoothness. Second, we have restricted attention to kernels supported on $[\alpha_1, \alpha_2]$. We proceed to verify that Theorem 2.1 can be applied to problem (WL).

The point evaluation functionals $g_\alpha \in \Gamma = X^*$ are both linear and continuous, hence continuously differentiable and also uniformly continuous. It is also easily checked that $f, h \in C^1(X)$. Furthermore, the set $\nabla B(x) = \{\nabla h(x)\}$ is obviously compact. We also have

Lemma 4.1 *For problem (WL), the set* $\nabla A(x) = \{\nabla g_\alpha(x) : \alpha \in I\}$ *is compact.*

Proof: We apply Lemma 3.1. The point evaluation functionals $V_x \in \Gamma^*$ are continuous, since $V_x(g_\alpha) = g_\alpha(x) = x(\alpha)$. Since I is compact, it remains only to demonstrate that the index map $\alpha \mapsto g_\alpha$ is continuous.

Consider the optimization problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \int_{a_1}^{a_2} [x'(\alpha)]^2 d\alpha \\ \text{subject to} & x(a_1) = b_1, x(a_2) = b_2 \,. \end{array}$$

It is a trivial exercise in the calculus of variations to establish that the minimizer is a straight line with slope

$$x'(\alpha) = (b_2 - b_1)/(a_2 - a_1).$$

This yields a minimum objective function value of $|b_2 - b_1|^2/|a_2 - a_1|$. It follows that any $x \in X$ with $x(a_1) = b_1$ and $x(a_2) = b_2$ must satisfy

$$\|x\|^2 \geq |b_2 - b_1|^2/|a_2 - a_1|. \tag{3}$$

Now suppose that $\alpha_n \to \alpha_0$ as $n \to \infty$. Then (3) allows us to write

$$\begin{aligned}
\|g_{\alpha_n} - g_{\alpha_0}\| &= \sup_{\|x\| \leq 1} |g_{\alpha_n}(x) - g_{\alpha_0}(x)| = \sup_{\|x\| \leq 1} |x(\alpha_n) - x(\alpha_0)| \\
&\leq \sup_{\|x\| \leq 1} |\alpha_n - \alpha_0|^{\frac{1}{2}} \|x\| = |\alpha_n - \alpha_0|^{\frac{1}{2}} \to 0 \text{ as } n \to \infty . \quad \square
\end{aligned}$$

Next, we show that our conditions on K hold.

Lemma 4.2 *For problem (WL), let $K = \nabla A(x) \cup \nabla h(x)$. Then $C_1(K)$ does not contain the origin of $X = H^1[\alpha_1, \alpha_2]$.*

Proof: We exploit the fact that the gradient is the Riesz representer of the directional derivative. Let $\eta \in X$; then

$$\begin{aligned}
g'_\alpha(x)(\eta) &= \lim_{t \to 0} \frac{1}{t} [g_\alpha(x + t\eta) - g_\alpha(x)] \\
&= \lim_{t \to 0} \frac{1}{t} [x(\alpha) + t\eta(\alpha) - x(\alpha)] \\
&= \eta(\alpha)
\end{aligned}$$

and

$$h'(x)(\eta) = \lim_{t \to 0} \frac{1}{t} [h(x + t\eta) - h(x)]$$

$$
\begin{aligned}
&= \lim_{t\to 0} \frac{1}{t}\left[\int_I [x+t\eta](\alpha)d\alpha - 1 - \int_I x(\alpha)d\alpha + 1\right] \\
&= \int_I \eta(\alpha)d\alpha \ .
\end{aligned}
$$

Hence, $\nabla g_\alpha(x)$ must satisfy

$$\langle \nabla g_\alpha(x)\,,\eta\rangle = \eta(\alpha) \quad \forall\, \eta \in X$$

and $\nabla h(x)$ must satisfy

$$\langle \nabla h(x)\,,\eta\rangle = \int_I \eta(\alpha)d\alpha \quad \forall\, \eta \in X\ .$$

Now suppose that there exists $\mu \in \mathbf{M}_1(K)$ such that $\int_K y\mu(dy) = 0$. Let $\lambda = \mu(\nabla h(x))$ and let $(1-\lambda)u$ denote the measure on $(I, \mathbf{I})$ induced by μ. Then it must be that, $\forall\, \eta \in X$,

$$
\begin{aligned}
0 &= \langle 0, \eta\rangle \\
&= \langle \int_K y\mu(dy), \eta\rangle \\
&= \int_K \langle y, \eta\rangle \mu(dy) \\
&= \int_{\nabla A(x)} \langle y, \eta\rangle \mu(dy) + \lambda\langle \nabla h(x), \eta\rangle \\
&= \int_{\nabla A(x)} \eta(\alpha)\mu(dy) + \lambda \int_I \eta(\alpha)d\alpha \\
&= (1-\lambda)\int_I \eta(\alpha)u(d\alpha) + \lambda \int_I \eta(\alpha)d\alpha\ . \qquad (4)
\end{aligned}
$$

But the last expression in (4) is strictly positive if $\eta \in X$ is strictly positive on I; hence, $C_1(K)$ cannot contain the origin of X. $\square$

Remark: If u is a finitely supported signed measure, say $u = \sum_{i=1}^{m} u_i \mathbf{1}(\alpha_i)$, where $\mathbf{1}$ denotes point-mass, then (4) reduces to

$$0 = (1-\lambda)\sum_{i=1}^{m} u_i \eta(\alpha_i) + \lambda \int_I \eta(\alpha)d\alpha \,.$$

If $\lambda = 1$, this equality fails for (say) $\eta(\alpha) \equiv 1$; if $\lambda \neq 1$, this equality fails for any η satisfying $\eta(\alpha_i) = -u_i$ and $\int_I \eta(\alpha)d\alpha = 0$. Thus, the condition of regularity also holds for problem (WL). Notice, however, that the restriction to finite linear combinations in the definition of linear independence is crucial to this conclusion. If arbitrary signed measures are allowed, then take u to be the negative uniform measure on I and put $\lambda = 1/(\alpha_2 - \alpha_1 + 1)$. Then the last expression in (4) is

$$\begin{aligned}
&(1-\lambda)\int_I \eta(\alpha)u(d\alpha) + \lambda \int_I \eta(\alpha)d\alpha \\
= \;& \left[1 - \frac{1}{\alpha_2 - \alpha_1 + 1}\right]\frac{1}{\alpha_2 - \alpha_1}\int_I \eta(\alpha)d\alpha + \frac{1}{\alpha_2 - \alpha_1 + 1}\int_I \eta(\alpha)d\alpha \,,
\end{aligned}$$

which does indeed vanish $\forall\, \eta \in X$. This distinction should not be surprising. Roughly stated, finitely many values do not determine a function's Lebesgue integral, but all values together do.

Finally, the equality constraint in problem (WL) is easily incorporated into the proof of Lemma 3.4. This provides a means of verifying that any feasible point for problem (WL) satisfies the constraint qualification. Theorem 2.1 therefore applies: a necessary condition for x_n^* to be a local solution of problem (WL) is that the first-order conditions hold.

Let us make some further observations concerning problem (WL). The objective function is strictly convex and the constraint set is convex. It follows that any local solution will be the unique global solution. It is well known that the variational inequality which serves as a necessity condition when the constraint set is convex serves as a sufficiency condition when the objective function is also convex. A rather straightforward argument can be used to show that, in the case of a convex constraint set, condition (i) of Proposition 2.1, namely $Z_2(x^*) = \phi$, implies the variational inequality necessity condition. These comments say that, in the case of a convex program where the constraint qualification holds (as is the case for problem (WL)), the existence of Lagrange multipliers (Proposition 2.1) is both necessary and sufficient for x^* to be a global minimizer.

Our theory, the above comments and some straightforward computations lead us to the following result concerning problem (WL): x_n^* is the unique global minimizer if and only if there exists a totally finite measure concentrating on $[\alpha_1, \alpha_2]$, with density function u_n^*, and a real number λ_n^*, such that

(a) $u_n^*(\alpha) = 2\frac{n-1}{n}[x_n^* * \delta * \bar{\delta}](\alpha) - 2[\delta * \bar{\delta}](\alpha) + \frac{2}{n}x_n^*(\alpha) + \lambda_n^* \quad \forall\, \alpha \in [\alpha_1, \alpha_2]$,

(b) $x_n^*(\alpha) \geq 0 \quad \forall\, \alpha \in [\alpha_1, \alpha_2]$,

(c) $\int_{\alpha_1}^{\alpha_2} x_n^*(\alpha) d\alpha = 1$,

(d) $x_n^*(\alpha) u_n^*(\alpha) = 0 \quad \forall\, \alpha \in [\alpha_1, \alpha_2]$,

(e) $u_n^*(\alpha) \geq 0 \quad \forall\, \alpha \in [\alpha_1, \alpha_2]$.

In condition (a), $\bar{\delta}(\alpha) := \delta(-\alpha)$, and $*$ denotes convolution.

Since problem (WL) is highly nontrivial, it is not surprising that the corresponding necessity conditions are somewhat complicated. It is not the purpose of this paper to attempt a detailed analysis of these conditions, although we intend to do so in later work. It is evident, however, that the theory developed here may be productively applied to a body of problems admitting an infinite programming formulation.