J. Austral. Math. Soc. **21** (Series B) (1979), 37–44

CHARACTERIZATIONS OF OPTIMALITY FOR CONTINUOUS CONVEX MATHEMATICAL PROGRAMS. PART I. LINEAR CONSTRAINTS

C. H. SCOTT and T. R. JEFFERSON

(Received 24 April 1978)

(Revised 22 November 1978)

Abstract

Recently we have developed a completely symmetric duality theory for mathematical programming problems involving convex functionals. Here we set our theory within the framework of a Lagrangian formalism which is significantly different to the conventional Lagrangian. This allows various new characterizations of optimality.

1. Introduction

A completely symmetric duality theory for mathematical programs involving functionals has recently been given by Scott and Jefferson [4, 5]. This theory is essentially an extension of generalized geometric programming to function spaces and, as such, relies heavily on modern ideas of convexity. An alternative approach to continuous programs, based on conjugate function theory, has been given by Rockafellar [3] who embeds mathematical programs in a certain parameterized family of closely related programs. In this paper we embed our theory in a Lagrangian formalism taking the Lagrangian for the dual program as opposed to that for the primal program as is conventional. It is shown that a saddle point condition holds for this Lagrangian and this paves the way to determine five equivalent characterizations of optimality for continuous convex mathematical programs. An example involving a quadratic functional is given to illustrate the theory.

Specifically we concentrate on linearly constrained functional programs. A sequel to this paper will consider the case of non-linear convex constraints [6].

https://doi.org/10.1017/S0334270000001892 Published online by Cambridge University Press

As a preliminary, we give the necessary ideas of convex analysis needed in this paper.

2. Convexity

Let X and Y be real vector spaces in duality with respect to a certain real bilinear function $\langle\cdot,\cdot\rangle$. We assume that X and Y have been assigned locally convex Hausdorff topologies compatible with this duality, so that elements of each space can be identified as continuous linear functionals on the other. Then X and Y are topologically paired spaces.

DEFINITION. *The function g: $X \to [-\infty, +\infty]$ is convex if its epigraph*

$$\operatorname{epi}(g) = \{(x, \mu) \mid x \in X,\ \mu \in R,\ \mu \geqslant g(x)\}$$

is a convex set in $X \times R$.

DEFINITION. *The set,*

$$\operatorname{dom}(g) = \{x \in X \mid g(x) < +\infty\}$$

is the effective domain of g.

DEFINITION. *A convex function g on X is said to be proper if $g(x) > -\infty$ for all $x \in X$ and $g(x) < +\infty$ for at least one $x \in X$.*

If g is a proper convex function, then $\operatorname{dom}(g)$ is a non-empty convex set and g is finite there.

DEFINITION. *A convex function g on X is lower semi-continuous (l.s.c.) if, for each $\mu \in R$, the convex level set*

$$\{x \in X \mid g(x) \leqslant \mu\}$$

is a closed set in X.

DEFINITION. *Let g be a proper convex function on X. Its conjugate function h on Y is defined by*

$$h(y) = \sup_{x \in X} (\langle x, y\rangle - g(x)) \quad \text{for all } y \in Y.$$

The function h is a l.s.c. convex function but not necessarily proper. However if g is a l.s.c. proper convex function, then h is also l.s.c. proper convex and

$$g(x) = \sup_{y \in Y} (\langle x, y\rangle - h(y)) \quad \text{for all } x \in X.$$

The conjugate functions by definition satisfy for $x \in X$ and $y \in Y$ the conjugate inequality:

$$g(x)+h(y)\geqslant\langle x,y\rangle$$

DEFINITION. *An element $y\in Y$ is said to be a subgradient of the convex function g at the point x if*

$$g(x_0)\geqslant g(x)+\langle x_0-x,y\rangle \quad \text{for all } x_0\in X.$$

The set of all subgradients at x, denoted by $\partial g(x)$ is a weak star-closed convex set in X which may be empty. If $\partial g(x)$ is non-empty, the convex function $g(x)$ is said to be subdifferentiable at x. If g is differentiable in the sense of Frechet, $\partial g(x)$ consists of a single point, namely, the gradient $\nabla g(x)$ of g at x.

If g is a l.s.c. proper convex function on X, then

$$y\in\partial g(x)\Leftrightarrow g(x)-h(y)=\langle x,y\rangle.$$

In this paper we consider integrals of convex functions of the form

$$G(x)=\int_T g(t,x(t))\,\nu(dt), \quad x\in\mathscr{L}$$

where $g(t,x)$ is a l.s.c. proper convex function of x for each t. Here T is a measure space with complete σ-finite measure ν and $\mathscr{L}$ is a real vector space of measurable functions x from T to a separable Hilbert space $\mathscr{H}$ and $g\colon T\times\mathscr{H}\to[-\infty,+\infty]$. g is termed a normal convex integrand if it satisfies the conditions:

(i) $g(t,x)$ is a l.s.c. proper convex function on $\mathscr{H}$ for each fixed t.

(ii) There is a countable collection $\mathscr{S}$ of measurable functions x from T to $\mathscr{H}$ such that

(a) for each $x\in\mathscr{S}$, $g(t,x(t))$ is measurable in t,

(b) for each t, $\mathscr{S}_t\cap\operatorname{dom}(g)$ is dense in $\operatorname{dom}(g)$, where

$$\mathscr{S}_t=\{x(t)\,|\,x\in\mathscr{S}\}$$

Normal convex integrands may be recognized from the following known results [1]:

(i) Suppose $g(t,x)=\hat{g}(x)$ for all t, where $\hat{g}$ is a l.s.c. proper convex function on $\mathscr{H}$. Then $\hat{g}$ is a normal convex integrand.

(ii) Let the function $g(t,x)$ on $T\times\mathscr{H}$ have values in $[-\infty,+\infty]$ such that $g(t,x)$ is measurable in t for each fixed x and for each t, $g(t,x)$ is a l.s.c. proper convex function in x with interior points in its effective domain. Then g is a normal convex integrand.

DEFINITION. *$\mathscr{L}$ is said to be decomposable if it satisfies the following conditions:*

(i) $\mathscr{L}$ contains every bounded measurable function from T to $\mathscr{H}$ which vanishes outside a set of finite measure.

(ii) If $x \in \mathscr{L}$ and E is a set of finite measure in T, then $\mathscr{L}$ contains $\chi_E x$ where χ_E is the characteristic function of E.

An important class of functions which are decomposable in this sense are the $L^p(0, T;\ R^n)$ spaces.

THEOREM 1. *Let $\mathscr{L}$ and $\mathscr{K}$ be topologically paired by means of the summable inner product*

$$\langle x, y\rangle = \int_T x(t)\, y(t)\, \mu(dt) \quad \text{for all } x \in \mathscr{L},\ y \in \mathscr{K}$$

and suppose $\mathscr{L}$ and $\mathscr{K}$ are decomposable. Let g be a normal convex integrand such that $g(t, x(t))$ is summable in t for at least one $x \in \mathscr{L}$ and $h(t, y(t))$ is summable in t for at least one $y \in \mathscr{K}$. Then the functionals G on $\mathscr{L}$ and H on $\mathscr{K}$, where

$$G(x) = \int_T g(t, x(t))\, \mu(dt), \quad H(y) = \int_T h(t, y(t))\, \mu(dt)$$

are proper convex functions conjugate to each other. Here g and h are conjugate functions as defined previously.

PROOF. See [2].

3. Duality

We consider the following functional program, termed the primal program.

$$\textit{Minimize} \quad G(x)$$

subject to implicit constraints $x \in C$, where $C \subset \mathscr{L}$ is closed and convex, and cone condition $x \in \chi$, where χ is a non-empty closed convex cone in $\mathscr{L}$.

In [4], we show that the above program has an associated dual program.

$$\textit{Minimize} \quad H(y)$$

subject to implicit constraints $y \in D \subset \mathscr{K}$ and the polar cone condition $y \in \chi^* \subset \mathscr{K}$, where

$$\chi^* = \{y \mid \langle x, y\rangle \geq 0 \textit{ for all } x \in \chi\}$$

and

$$D = \{y \in Y \mid \sup_{x \in C} (\langle x, y\rangle - G(x)) < +\infty\}.$$

At optimality, the following relations have been shown to hold:

$$G(x^*) + H(y^*) = 0, \quad x^* \in \partial H(y^*), \quad y^* \in \partial G(x^*).$$

In fact [4] deals with subspaces rather than cones but the generalization is straightforward.

For purposes of exposition, we define

$$\Phi = G(x^*).$$

4. Characterizations of optimality

We introduce a Lagrangian

$$L(x;y) = \langle x, y\rangle - H(y) \tag{1}$$

which, although significantly different to the conventional Lagrangian for the problem, has a saddle-point property. In particular, this is the Lagrangian for the dual program treated as a maximization problem. We show this in the following theorem.

THEOREM 2. *If $x^* \in \chi$ and $y^* \in D$, then*

(i) *x^* is optimal for the primal program, and*

(ii) $\Phi = \inf_{x\in\chi} L(x;y^*)$

if and only if $(x^*;y^*)$ is a saddle point for the Lagrangian, that is,

$$L(x^*;y) \leqslant L(x^*;y^*) \leqslant L(x;y^*) \quad \textit{for all } y \in D \quad \textit{and} \quad x \in \chi.$$

In addition, the value of the saddle point

$$L(x^*;y^*) = G(x^*) = \Phi.$$

PROOF. Initially we show that (i) and (ii) imply that $(x^*;y^*)$ is a saddle point for the Lagrangian. By definition, we have that

$$\begin{aligned} L(x^*;y) &= \langle x^*, y\rangle - H(y) \\ &\leqslant \sup_{z\in D}(\langle x^*, z\rangle - H(z)) \\ &= G(x^*) && \text{by Theorem 1} \\ &= \Phi && \text{by definition} \\ &= \inf_{x\in\chi} L(x;y^*) && \text{by (ii)} \\ &\leqslant L(x;y^*) && \text{for all } x \in \chi. \end{aligned}$$

https://doi.org/10.1017/S0334270000001892 Published online by Cambridge University Press

From the above, we see that

$$L(x^*; y^*) \leqslant G(x^*) \quad \text{and} \quad G(x^*) \leqslant L(x^*; y^*).$$

Hence we have that

$$L(x^*: y^*) = G(x^*) = \Phi$$

and $(x^*; y^*)$ is a saddle point for L.

Since the argument is reversible (the conjugate transform of a convex functional is a symmetric operation), the theorem is proved.

COROLLARY 1.

$$\sup_{y \in D} L(x^*; y) = L(x^*; y^*)$$

if and only if the following relations hold:

(i) $x^* \in C$,

(ii) $y^* \in \partial G(x^*)$ (*subgradient condition*).

PROOF. Straightforward from the proof of Theorem 2.

COROLLARY 2.

$$L(x^*; y^*) = \inf_{x \in \chi} L(x; y^*)$$

if and only if the following relations hold:

(i) $y^* \in \chi^*$ (*polar cone condition*),

(ii) $\langle x^*, y^* \rangle = 0$ (*orthogonality condition*).

Then, we have that

$$L(x^*; y^*) = -H(y^*).$$

PROOF.

$$\begin{aligned} \inf_{x \in \chi} L(x; y^*) &= \inf_{x \in \chi} [\langle x, y^* \rangle - H(y^*)] \\ &= -H(y^*) \quad (2) \text{ for } y^* \in \chi^* \end{aligned}$$

By hypothesis, equation (2) is equal to $L(x^*; y^*)$.
Once again the argument can be reversed.

We are now in the position of being able to give five equivalent characterizations of optimality. For $x^* \in \chi$ and $y^* \in D$, these are:

(1) x^* is optimal for the primal program and

$$\Phi = \inf_{x \in \chi} L(x; y^*).$$

(2) $(x^*; y^*)$ is a saddle point for the Lagrangian

$$\text{that is } L(x^*; y) \leqslant L(x^*, y^*) \leqslant L(x, y^*) \quad \text{for all } y \in D \quad \text{and} \quad x \in \chi.$$

This is a consequence of the Theorem 2.

(3) $$x^* \in C, \quad y^* \in \partial G(x^*), \quad L(x^*; y^*) = \inf_{x \in \chi} L(x; y^*).$$

This is a consequence of Corollary 1.

(4) $$y^* \in \chi^*, \quad \langle x^*, y^* \rangle = 0, \quad \sup_{y \in D} L(x^*; y) = L(x^*, y^*).$$

This is a consequence of Corollary 2.

(5) $$x^* \in C, \quad y^* \in \partial G(x^*), \quad y^* \in \chi^*, \quad \langle x^*, y^* \rangle = 0.$$

This is a consequence of Corollaries 1 and 2.

5. Example

Consider the following quadratic functional program.

$$\text{Minimize} \quad G(x) = \frac{1}{2} \int_T x^2(t)\, \nu(dt)$$

subject to $A(t)\, x(t) \leqslant b(t)$, ν almost everywhere on T, where $A(t)$ is $n \times m$, $b(t)$ are given functions and $x \in L^2[0, T; R^n]$. This may be written as a primal program in Section 3 as

$$\text{minimize} \quad G(x)$$

subject to implicit constraints

$$x(t) \in R^n, \quad \alpha(t) \in \{b(t)\} \in R^m,$$

and cone condition

$$\chi = \left\{ (x, \alpha) \,\middle|\, (A(t) - I) \begin{pmatrix} x(t) \\ \alpha(t) \end{pmatrix} \leqslant 0, \quad \nu \text{ almost everywhere on } T \right\}.$$

Following the prescription in Section 2, the dual program is given by

$$\text{Minimize} \quad H(y) = \frac{1}{2} \int_T y^2(t)\, \nu(dt) + \int_T b(t)\, \beta(t)\, \nu(dt)$$

subject to implicit constraints

$$y(t) \in R^n, \quad \beta(t) \in R^m$$

and polar cone condition

$$\chi^* = \{(y,\beta) \mid y(t) = A^{\mathrm{T}} z(t), \beta(t) = -z(t), z(t) \leqslant 0, \ \nu \text{ almost everywhere on } T\}.$$

The subgradient condition relates the primal and dual variables at optimality by

$$x^*(t) = y^*(t).$$

The Lagrangian, defined by equation (1), is in this case

$$L(x;\, y) = \int_T x(t)\, y(t)\, \nu(dt) - \frac{1}{2}\int_T y^2(t)\, \nu(dt) - \int_T b(t)\, \beta(t)\, \nu(dt)$$

since

$$\langle x, y\rangle = \int_T x(t)\, y(t)\, \nu(dt) \quad \text{for } x, y \in L^2[0, T].$$

Hence, it is straightforward to obtain the five equivalent characterizations of optimality as given in Section 3.

References

[1] R. T. Rockafellar, "Integrals which are convex functionals", *Pacific J. Math.* 24 (1968), 525–539.

[2] R. T. Rockafellar, "Convex integral functionals and duality", in *Contributions to nonlinear functional analysis* (New York: Academic Press, 1971), pp. 215–236.

[3] R. T. Rockafellar, "Conjugate duality and optimization", *SIAM Regional Conference Series in Applied Mathematics*, 16 (1974).

[4] C. H. Scott and T. R. Jefferson, "A generalization of geometric programming with an application to information theory", *Information Sciences* 12 (1977), 263–269.

[5] C. H. Scott and T. R. Jefferson, "Duality in infinite-dimensional mathematical programming: Convex integral functionals", *J. Math. Anal. Appl.* 61 (1977), 251–261.

[6] C. H. Scott and T. R. Jefferson, "Characterizations of optimality for continuous convex mathematical programs. Part 2. Nonlinear constraints" (in preparation).

School of Mechanical and Industrial Engineering
University of N.S.W.
Kensington, N.S.W. 2033