STOCHASTIC NON-LINEAR PROGRAMMING

M. A. HANSON

(received 27 October 1963)

1. Introduction

Although many varied techniques have been proposed for handling deterministic non-linear programming problems there appears to have been little success in solving the more realistic problem of stochastic non-linear programming, despite the many results that have been obtained for stochastic linear programming. In this paper the stochastic non-linear problem is treated by means of an adaptation of a method used by Berkovitz [1] in obtaining an existence theorem for a type of inequality constrained variational problem involving one independent variable. The stochastic programming problem of course involves many independent variables. Necessary conditions are obtained for the existence of a solution of a fairly general type of non-linear problem, and these conditions are shown to be also sufficient for the convex problem. A duality theorem is given for the latter problem.

An interesting result is that the necessary and sufficient conditions thus obtained for the stochastic problem are in fact of the form of deterministic non-linear programming problems, so that much of the existing theory for deterministic problems can be applied to the stochastic problems, where these conditions apply.

The term *stochastic* is used here to mean generally either that the data in the programming problem is subject to random errors of measurement or subject to unpredictable fluctuations in time.

2. Statement of the problem

Let the vector $y \in Y \subset E_m$ be the data of a programming problem, and let y have the scalar distribution $\psi(y)$ which is continuously twice differentiable on Y. Let $x \in E_n$ be an activity vector chosen to minimise the expected value of the scalar objective function $\phi(x, y)$ subject to the two types of constraints:

(i) the expected value of the function $g(x, y)$ is non-negative, where $g \in E_r$, and

(ii) the function $h(x, y)$ is non-negative where $h \in E_s$.

https://doi.org/10.1017/S1446788700024113 Published online by Cambridge University Press

Both types of constraints are important in practice; the first for example may refer to long term contractual requirements whose average short term requirements may be violated, and the second may refer to the more common physical limitations of the system which cannot be violated at all.

In the following the symbol $\int dy$ is used to indicate the m-dimensional definite integral over Y; superscripts denote vector components; and subscripts denote vector differentiation, that is, g_x is the matrix $[\partial g^j/\partial x^i]$. There is no notational distinction between row and column vectors.

So the problem may be written in the form:

(1) minimise $\int \phi(x, y)\psi(y)dy$

(2) subject to $\int g(x, y)\psi(y)dy \geqq 0$

(3) and $h(x, y) \geqq 0.$

Let the set of values of x defined by (2) and (3) be denoted by X. It will be supposed that ϕ, g, and h are continuously twice differentiable on $S = X \times Y$.

We further require the *constraint* condition that if $r+s > n$ then at most n components of g and h together can vanish at each point of S, and for those components g^j say $j = 1, 2, \cdots J$, and h^k, say $k = 1, 2, \cdots K$, that do vanish at some point of S the matrix

$$\left[\frac{\partial g^1}{\partial x^i} \cdots \frac{\partial g^J}{\partial x^i}\, \frac{\partial h^1}{\partial x^i} \cdots \frac{\partial h^K}{\partial x^i}\right]$$

has maximum rank. Further it is desired to impose a certain smoothness on the constraint boundary, namely that if $g^j(x) = 0$ then $||x^* - x||^2 = o(||g^j(x^*) - g^j(x)||)$ for feasible x^*, and similarly for h. These conditions are not very restrictive from the practical viewpoint, but they are required in the following proof to ensure that specific changes in the value of x can be made which induce specific changes in all the zero components g^j, h^k, $j = 1, 2, \cdots J$, $k = 1, 2, \cdots K$.

3. Necessary and sufficient conditions

THEOREM 1. If x is to be a solution of the problem (1), (2), (3) above, it is necessary that there exist an r-dimensional vector $\lambda \leqq 0$ and an s-dimensional vector function $\mu(y) \leqq 0$ such that

(4) $$\phi_x(x, y)\psi(y) + \lambda g_x(x, y)\psi(y) + \mu(y)h_x(x, y) = 0$$

(5) $$\mathscr{E}\lambda g(x, y) = 0$$

and

https://doi.org/10.1017/S1446788700024113 Published online by Cambridge University Press

$$\mu(y)h(x, y) = 0 \tag{6}$$

PROOF. We shall follow the usual method of converting inequality constraints into equations by introducing slack variables. Standard Euler-Lagrange theory (see for example [3]) will then be used to show the necessity of (4), (5) and (6), after which the slack variables will be eliminated and a separate argument will show the necessity of the semi-negativity of λ and $\mu(y)$.

Let

$$\eta^2 = [(\eta^1)^2(\eta^2)^2 \cdots (\eta^r)^2]$$

and

$$\xi^2 = [(\xi^1)^2(\xi^2)^2 \cdots (\xi^s)^2]$$

be r-dimensional and s-dimensional vectors respectively such that

$$\int g(x, y)\psi(y)\mathrm{d}y - \eta^2 = 0 \tag{7}$$

and

$$h(x, y) - \xi^2 = 0. \tag{8}$$

Introducing the Lagrange multipliers λ and $\mu(y)$ we immediately derive from the Lagrangian function

$$L = \phi\psi + \lambda(g\psi - \eta^2) + \mu(h - \xi^2)$$

the Euler equations:

$$L_{x^i} = 0, \qquad i = 1, 2, \cdots n \tag{9}$$

$$L_{\eta^j} = 0, \qquad j = 1, 2, \cdots r \tag{10}$$

$$L_{\xi^k} = 0, \qquad k = 1, 2, \cdots s \tag{11}$$

that is,

$$\phi_x\psi + \lambda g_x + \mu h_x = 0 \tag{12}$$

$$\lambda^j\eta^j = 0, \qquad j = 1, 2, \cdots r \tag{13}$$

and

$$\mu^k\xi^k = 0, \qquad k = 1, 2, \cdots s \tag{14}$$

Equations (7) and (13) imply

$$\mathscr{E}\lambda^j g^j = 0 \tag{15}$$

and equations (8) and (14) imply

$$\mu^k h^k = 0. \tag{16}$$

https://doi.org/10.1017/S1446788700024113 Published online by Cambridge University Press

The results (12), (15) and (16) thus give the equations (4), (5) and (6). It remains to prove the semi-negativity of λ and μ.

From (15) it follows that if some $g^j > 0$ then the corresponding $\lambda^j = 0$, and similarly from (16) for $\mu(y)$. If some $g^j = 0$ consider a special variation from the minimising x to a neighbouring feasible x^* to be defined in the following proof. Taylor's theorem to first order with remainder term $\varepsilon(x^*, x, y) = o(||x^*-x^*||^2)$ gives

$$\begin{aligned}\int [(\phi(x^*, y)\psi(y)+\lambda g(x^*, y)\psi(y)+\mu(y)h(x^*, y)\\ -(\phi(x, y)\psi(y)+\lambda g(x, y)\psi(y)+\mu(y)h(x, y))]\mathrm{d}y\\ = \int [(x^*-x)(\phi_x(x, y)\psi(y)+\lambda g_x(x, y)\psi(y)+\mu(y)h_x(x, y))+\varepsilon]\mathrm{d}y\\ = \int \varepsilon\, \mathrm{d}y\end{aligned}$$

by (4), as would be expected from the usual Euler-Lagrange theory.

Therefore

$$\begin{aligned}&\int \phi(x^*, y)\psi(y)\mathrm{d}y-\phi(x, y)\psi(y)\mathrm{d}y\\ (17)\qquad &= -\int [\lambda(g(x^*, y)-g(x, y))\psi(y)+\mu(y)(h(x^*, y)-h(x, y))-\varepsilon]\mathrm{d}y\\ &= -\int [\lambda g(x^*, y)\psi(y)+\mu(y)h(x^*, y)-\varepsilon]\mathrm{d}y\end{aligned}$$

by (5) and (6).

Now let x^* be such that ε is sufficiently small not to affect the sign of the term in square brackets and such that all zero components of g, except the jth, and all zero components of h are unchanged in value. Such a choice of x^* is possible under the constraint conditions.

Then the sign of

$$\left[\int \phi(x^*, y)\psi(y)\mathrm{d}y - \int \phi(x, y)\psi(y)\mathrm{d}y\right]$$

is the sign of

$$\left[-\int \lambda^j g^j(x^*, y)\psi(y)\mathrm{d}y\right].$$

Hence since $g^j(x^*, y) > 0$ by hypothesis, $\lambda^j \leqq 0$.

Similarly $\mu(y) \leqq 0$; for if $\mu^k(y) > 0$ for some values of y then the variation from x to x^* can be chosen so that $-\int \mu^k(y)h^k(x^*, y)\psi(y)\mathrm{d}y$ is strictly negative, and all other zero components of g and h are unchanged, giving from (17) the contradiction

$$\int \phi(x^*, y)\psi(y)\mathrm{d}y - \int \phi(x, y)\psi(y)\mathrm{d}y < 0.$$

Thus equations (4), (5) and (6) provide $n+r+s$ necessary conditions for the solution of the problem involving the $n+r+s$ unknowns x, λ, and μ, in terms of the parameter y. The additional necessary inequalities $\lambda \leqq 0$, $\mu \leqq 0$ render the problem a non-linear programming problem, although deterministic in form.

The foregoing theorem may be considered to be a stochastic generalization of the Kuhn-Tucker theorem [2] for deterministic non-linear programming. In general the conditions obtained will not be sufficient to determine a minimum, of course, but as in the Kuhn-Tucker theorem, they are sufficient for convex programming, as will be shown in Theorem 2.

THEOREM 2. If in addition to the properties given above $\phi(x, y)$ is convex with respect to x, and $g(x, y)$ and $h(x, y)$ are concave with respect to x, then the necessary conditions obtained in Theorem 1 are also sufficient for a minimum of the stochastic programming problem.

PROOF. Since $\phi(x, y)$ is convex with respect to x,

$$\int \phi(x^*, y)\psi(y)\,\mathrm{d}y \geqq \int (\phi(x, y)+(x^*-x)\phi_x(x, y))\psi(y)\mathrm{d}y$$
$$= \int (\phi(x, y)\psi(y)-(x^*-x)[\lambda g_x(x, y)\psi(y)+\mu(y)h_x(x, y)])\mathrm{d}y$$

by (4)

$$\geqq \int \phi(x, ty)\psi(y)\mathrm{d}y - \int \lambda(g(x^*, y)-g(x, y))\psi(y)\mathrm{d}y - \int \mu(y)(h(x^*, y)-h(x, y))\mathrm{d}y,$$

since $\lambda g(x, y)$ and $\mu(y)h(x, y)$ are convex with respect to x,

$$= \int \phi(x, y)\psi(y)\mathrm{d}y - \int \lambda g(x^*, y)\psi(y)\mathrm{d}y - \int \mu(y)h(x^*, y)\mathrm{d}y$$

by (5) and (6)

$$\geqq \int \phi(x, y)\psi(y)\mathrm{d}y$$

by (2) and (3) and the fact that $\lambda \leqq 0$, $\mu(y) \leqq 0$.

Hence $\phi(x, y)$ is minimal.

4. Duality

Duality theorems are useful in mathematical programming for establishing upper and lower bounds for the objective function when an approximately optimal activity is known, and hence providing a measure of the accuracy of the approximation. If the primal problem involves

https://doi.org/10.1017/S1446788700024113 Published online by Cambridge University Press

minimisation, say, then the dual involves maximisation, and bounds are obtained by solving a linear problem.

THEOREM 3. If ϕ, ψ, g, and h are functions as defined in Theorem 2, and if x_0 is optimal in the primal problem (1), (2), (3) then x_0 is also optimal in the dual problem:

$$\text{(18)} \qquad \text{maximise} \qquad \int (\phi(x, y)\psi(y) + \lambda g(x, y)\psi(y) + \mu(y)h(x, y))\,dy$$

$$\text{(19)} \qquad \text{subject to} \qquad \int (\phi_x(x, y)\psi(y) + \lambda g_x(x, y)\psi(y) + \mu(y)h_x(x, y))\,dy = 0$$

$$\text{(20)} \qquad \lambda \leqq 0$$

and

$$\text{(21)} \qquad \mu(y) \leqq 0.$$

The optimal λ, $\mu(y)$ satisfy (4), (5) and (6), and the optimal objective functions in both problems are equal in value.

PROOF. Since x_0 is optimal in problem (1), (2), (3) then there exist vectors $\lambda_0 \leqq 0$, $\mu_0(y) \leqq 0$ satisfying (4), (5) and (6) and hence satisfying the constraints of the dual problem. Let x^*, λ^*, $\mu^*(y)$ be any feasible solution of the dual problem.

Then

$$\begin{aligned}
&\int (\phi(x_0, y)\psi(y) + \lambda_0 g(x_0, y)\psi(y) + \mu_0(y)h(x_0, y)dy) \\
&\qquad - \int (\phi(x^*, y)\psi(y) + \lambda^* g(x^*, y)\psi(y) + \mu^*(y)h(x^*, y))\,dy \\
&\qquad = \int (\phi(x_0, y)\psi(y) - \phi(x^*, y)\psi(y) - \lambda^* g(x^*, y)\psi(y) - \mu^*(y)h(x^*, y))\,dy \\
&\qquad\quad \text{by (5) and (6),} \\
&\qquad \geqq \int ([\phi(x_0, y) - \phi(x^*, y)]\psi(y) + \lambda^*[g(x_0, y) - g(x^*, y)]\psi(y) \\
&\qquad\quad + \mu^*(y)[h(x_0, y) - h(x^*, y)])\,dy \\
&\qquad\quad \text{by (2) and (3)} \\
&\qquad \geqq \int (x_0 - x^*)[\phi_x(x^*, y)\psi(y) + \lambda^* g_x(x^*, y)\psi(y) + \mu^*(y)h_x(x^*, y)]\,dy \\
&\qquad\quad \text{since } \phi,\ \lambda^* g, \text{ and } \mu^* h \text{ are convex,} \\
&\qquad = 0, \text{ by (19).}
\end{aligned}$$

Hence x_0, λ_0 and $\mu_0(y)$ are optimal in the dual problem.

Further,

https://doi.org/10.1017/S1446788700024113 Published online by Cambridge University Press

$$\int (\phi(x_0, y)\psi(y) + \lambda_0 g(x_0, y)\psi(y) + \mu_0(y)h(x_0, y))\mathrm{d}y$$
$$= \int \phi(x_0, y)\psi(y)\mathrm{d}y$$
$$\text{by (5) and (6),}$$

so the optimal objective functions in both problems are equal.

References

[1] Berkovitz, L. D., Variational Methods in Problems of Control and Programming, J. Math. Anal. App. 3 (1961) 145—169.

[2] Kuhn, H. W. and Tucker, A. W., Non-Linear Programming, Proc. Second Berkeley Symp. on Math. Stats. and Prob., Univ. Calif. Press 1951.

[3] Bliss, G. A., Lectures on the Calculus of Variations, University of Chicago Press, Phoenix Edition 1961.

Department of Statistics,
University of New South Wales.