FINDING EVERETT'S LAGRANGE MULTIPLIERS BY LINEAR PROGRAMMING†

R. Brooks

The Rand Corporation, Santa Monica, California

and A. Geoffrion

University of California, Los Angeles, California

(Received March 31, 1966)

IN AN article[4] in this JOURNAL in 1963, EVERETT observed that if x^0 is optimal in

$$\text{maximize}_{x\varepsilon X}\, f(x) - \textstyle\sum_{i=1}^{i=m} u_i g_i(x), \qquad (1)$$

where the m constants u_i are nonnegative 'multipliers' and f and the g_i are arbitrary real-valued functions defined over an arbitrary set X, then x^0 also maximizes $f(x)$ over all $x\varepsilon X$ satisfying $g_i(x) \leqq g_i(x^0)$ $(i=1, \cdots, m)$.† Thus to solve

$$\text{maximize}_{x\varepsilon X}\, f(x) \text{ subject to } g_i(x) \leqq b_i \qquad (i=1, \cdots, m), \qquad (2)$$

where the b_i are given constants (it is convenient to think of the b_i as the amounts of available resources) it is sufficient to find nonnegative multipliers u_i^0 such that—and this we call *Everett's Condition*—a corresponding optimal solution x^0 of (1) can be found that satisfies $g_i(x^0) = b_i (i=1, \cdots, m)$. If such multipliers exist and a convenient mechanism for finding them is available, then solving (2) by solving (1) may be computationally convenient. For many problems of practical interest, however, such multipliers do not exist; but there may still be multipliers for which the $g_i(x^0)$ approximate the b_i closely enough for x^0 to be a useful approximate solution to (2). This approach amounts to reducing (2) to a problem without the g_i constraints.

The kth step $(k \geqq 2)$ of the iterative procedure implicitly suggested by Everett for finding an (approximate) solution to (2) is [here $u \equiv (u_i, \cdots, u_m)$]:

(k.1) Based on knowledge of $u^1, x^1, \cdots, u^{k-1}, x^{k-1}$, choose multipliers $u_i^k \geqq 0$ $(i=1, \cdots, m)$ in an attempt to satisfy Everett's Condition.
(k.2) Solve (1) with $u = u^k$ for an optimal solution x^k.
(k.3) If $g_i(x^k)$ is 'sufficiently near' b_i, $i=1, \cdots, m$, then stop; x^k is sufficiently near to being optimal in (2). Otherwise, go to step $k+1$.

Step 1 is the same as the general step, except that it begins with an arbitrary u^1 (guessed on the basis of past experience with a similar problem, say). It is assumed

† This observation is Everett's 'Main Theorem' [reference 4, p. 401].

that some method is available for performing substep (k.2).† How to perform substep (k.1) when $m \geqq 2$ was left largely unresolved by Everett, and stimulated the present note.‡

The main purpose of this note is to indicate how one might approximate the desired multipliers by means of linear programming. First, however, we weaken Everett's Condition slightly so that his approach can be applicable to problems with ineffectual constraints. A relation then becomes apparent to the saddle-point condition of KUHN AND TUCKER[8] for nonlinear programming. Since Everett's approach seems most competitive with other known methods for certain discrete allocation problems, we consider this case in some detail. It will be seen that Everett's method, when the multipliers are found by linear programming, becomes essentially the Simplex method with a 'column-generating' feature applied to an approximation of (2). Finally, we point out a relation to the so-called decomposition method of concave programming[3,10] for continuous allocation problems.

WEAKENING EVERETT'S CONDITION

IN CERTAIN problems with ineffectual constraints, Everett's Condition is unnecessarily restrictive in that, when a multiplier is zero, it is not necessary to require that the corresponding constraint be satisfied with strict equality. All that is needed is to find x^0 and u^0 such that

(i) x^0 is optimal in (1) with $u=u^0$, and

(ii) $u^0 \geqq 0$ and $u_i^0>0$ (resp. $=0$) implies $g_i(x^0)=b_i$ (resp. $\leqq b_i$), $i=1, \cdots, m$.

It is easily shown that if these conditions are satisfied, then x^0 is optimal in (2). We shall henceforth deal with this slightly modified version of Everett's condition.

It is of interest to note that (i) and (ii) are equivalent to the requirement that (x^0, u^0) be a *saddle-point* of the Lagrangian

$$L(x, u) \equiv f(x) - \textstyle\sum_{i=1}^{i=m} u_i[g_i(x) - b_i], \text{ i.e.,}$$

$$L(x, u^0) \leqq L(x^0, u^0) \leqq L(x^0, u) \qquad \text{for all} \qquad x \varepsilon X \text{ and } u \geqq 0.$$

Thus Everett's approach is seen to be essentially the attempt to construct a saddle-point for $L(x, u)$. Kuhn and Tucker[8] and others have given conditions on (2) that guarantee the existence of such a saddle-point. The basic condition for Euclidean spaces is that X be a convex set, f a concave function, and the g_i convex functions which satisfy any one of a number of mild qualifications.[1] Similar conditions for more general spaces are known (*see*, e.g., reference 7). Unfortunately, such conditions do not cover the case in which X is discrete, the situation of greatest interest to Everett and perhaps the one in which his (modified) approach is most promising.

† Throughout this paper we assume, as Everett did implicitly in his, that (1) achieves its maximum for any set of nonnegative multipliers. A sufficient condition for this when $X \subset R^n$ is that X be closed and bounded and f and the g_i be continuous. Similar sufficient conditions exist for more general spaces (e.g., reference 2, p. 69). *See also* the footnote, on p. 1151.

‡ We would like to thank DAVID MCGARVEY for encouraging our interest in this question.

FINDING THE MULTIPLIERS BY LINEAR PROGRAMMING

WHEN (2) IS a *linear* programming problem, i.e., when X is the nonnegative orthant of E^n and f and the g_i are linear functions, then it is not difficult to show that (x^0, u^0) satisfies conditions (i) and (ii) above if and only if x^0 solves (2) and u^0 solves the dual of (2). The u_i^0 are often interpreted as the 'dual prices' associated with (2), and are produced as an automatic by-product of the computational solution of (2). Dropping the assumption of linearity now, and observing that the burden of substep (k.1) is to approximate such prices on the basis of the data $u^1, x^1, \cdots, u^{k-1}, x^{k-1}$, it seems natural to use linear programming to compute the prices corresponding to a linearized version of (2) over the convex hull of the grid $\langle x^1, \cdots, x^{k-1}\rangle$. The resulting linear program, the dual prices of which are required at substep (k.1), is:

$$\text{maximize}_{\lambda_t \geqq 0} \sum_{t=1}^{k-1} \lambda_t f(x^t) \quad \text{subject to} \quad \sum_{t=1}^{k-1} \lambda_t = 1, \qquad \sum_{t=1}^{k-1} \lambda_t g_i(x^t) \leqq b_i \qquad (i=1, \cdots, m.). \tag{3}$$

Substep (*k.1 LP*): Solve (3) for the dual prices $u_0^k, u_1^k, \cdots, u_m^k$ corresponding to the $m+1$ constraints.

By linear programming theory, $u_i^k \geqq 0$ $(i=1, \cdots, m)$. The significance of u_0^k will become apparent below.

Discrete Case

If $X = \{\xi_1, \cdots, \xi_N\}$, where N is a finite positive integer, then Everett's procedure using (k.1 *LP*) is very close to the Simplex method for the linear programming problem

$$\text{maximize}_{\lambda_j \geqq 0} \sum_{j=1}^{j=N} \lambda_j f(\xi_j) \quad \text{subject to} \quad \sum_{j=1}^{j=N} \lambda_j = 1, \qquad \sum_{j=1}^{j=N} \lambda_j g_i(\xi_j) \leqq b_i \qquad (i=1, \cdots, m). \tag{4}$$

The subproblem (1), which now takes the form

$$\text{maximize}_{\xi \epsilon \{\xi_1, \cdots, \xi_N\}} f(\xi) - \sum_{i=1}^{i=m} u_i^k g_i(\xi), \tag{5}$$

does nothing more than determine (by the usual Simplex criterion) which new variable to bring into the basis at the kth iteration.† This permits the economy of carrying explicitly at one time no more than $m+1$ of the N columns corresponding to the ξ_j. The usual Simplex termination signal occurs at the first step k_0 such that

$$\text{maximum}_{\xi \epsilon \{\xi_1, \cdots, \xi_N\}} \left[f(\xi) - \sum_{i=1}^{i=m} u_i^k g_i(\xi)\right] \leqq u_0^k \tag{6}$$

(actually the maximum will $= u_0^k$). Thus in the finite discrete case Everett's procedure becomes precisely the Simplex method applied to (4) with a 'column-generation' feature if substep (k.3) is replaced by

Substep (*k.3 LP*): If (6) holds, stop. Otherwise, go to step $k+1$.

† In practice one probably would not solve (5) completely at every step, particularly in the early steps or when N is very large; from the theory of the Simplex method it is known that it is enough to find a ξ_j that gives a value greater than u_0^k to the maximand.

Since (4) is a finite linear program, Everett's procedure with substeps (*k*.1 *LP*) and (*k*.3 *LP*) is finitely convergent to the optimal solution $\lambda_j^*, j=1, \cdots, N$, of (4).

This method has been used to advantage by GILMORE AND GOMORY.[6] In their problem a ξ_j was a cutting pattern, and the subproblem a knapsack problem.

The question arises regarding the relation of the optimal solution of (4) to the original problem (2). Harking back to Everett's discussion of his method in terms of 'payoff-constraint space,' we see that if the points $[f(\xi_j), g_1(\xi_j), \cdots, g_m(\xi_j)] \varepsilon R^{m+1} (j=1, \cdots, N)$ are sufficiently dense near the boundary of their convex hull, then some of the policies ξ_j corresponding to $\lambda_j^*>0$ (and there will be no more than $m+1$ of these) will be good approximate solutions to (2).

Note that it is not necessary to store all of the ξ_j corresponding to the basic λ_j as the calculations proceed, but only the corresponding $f(\xi_j)$ and $g_i(\xi_j)$, $i=1, \cdots, m$. After termination, the 'basic' ξ_j can be recovered if desired by utilizing the fact that they 'price out' to 0. That is, they achieve the maximum $u_0^* \equiv u_0^{k_0}$ in (6). In fact all of the ξ_j that achieve u_0^* in (6) are 'used' by some optimal solution of (4). If it is desired to examine the ξ_j used in near-optimal solutions of (4), then one should recover the ξ_j that satisfy

$$f(\xi_j) - \textstyle\sum_{i=1}^{i=m} u_i^* g_i(\xi_j) \geqq u_0^* - \epsilon \qquad (7)$$

for some suitably small $\epsilon > 0$.

A possibly useful interpretation of (4) is the following: it is the extension of (2) from pure to mixed (randomized) strategies with f and the g_i replaced by their expectations. In this interpretation, λ_j^* is the probability of utilizing allocation ξ_j. When mixed strategies have a legitimate and acceptable interpretation, then (2) should have been written as (4) in the first place.†

Continuous Case

If X is not a finite discrete set, then the analysis of the previous case is complicated by the fact that there are an infinite number of variables in (4). Nevertheless, Everett's procedure using substep (*k*.1 *LP*) is almost exactly the so-called decomposition procedure for nonlinear programming.[3,10] When X is a bounded convex set and f is concave and the g_i are convex functions, then the sequence $\langle \sum_{t=1}^{k-1} \lambda_t x^t \rangle$ converges[3,9] to an optimal solution of (2) as $k \to \infty$.

REFERENCES

1. K. J. ARROW, L. HURWICZ, AND H. UZAWA, "Constraint Qualifications in Maximization Problems," *Naval Res. Log. Quart.* **8,** 175–190 (1961).
2. C. BERGE, *Topological Spaces*, Macmillan, New York, 1963.
3. G. DANTZIG, *Linear Programming and Extensions*, Princeton University Press, Princeton, N. J., 1963.
4. H. EVERETT, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Opns. Res.* **11,** 399–471 (1963).
5. S. FROMOVITZ, "Nonlinear Programming with Randomization," *Management Sci.* **11,** 831–846 (1965).

† Cf. FROMOVITZ,[5] and Gilmore and Gomory[6].

6. P. Gilmore and R. Gomory, "A Linear Programming Approach to the Cutting-Stock Problem," *Opns. Res.* **9**, 849–859 (1961).
7. L. Hurwicz, "Programming in Linear Spaces," *Studies in Linear and Nonlinear Programming* (K. J. Arrow, L. Hurwicz, and H. Uzawa, eds.), Stanford University Press, 1958.
8. H. K. Kuhn and A. W. Tucker, "Nonlinear Programming," *Proc. Second Berkeley Symposium on Mathematical Stat. and Prob.*, pp. 481–492, University of California Press, 1951.
9. M. Roubault, "De la dualité des programmes mathématiques convexes," Abstract, 1964 International Symposium on Mathematical Programming, London, July 6–10, 1964.
10. P. Wolfe, "Methods of Nonlinear Programming," in *Recent Advances in Mathematical Programming* (R. Graves and P. Wolfe, eds.), McGraw-Hill, New York, 1963.

SOME INVALID PROPERTIES OF MARKOV CHAINS

Paul J. Schweitzer

Institute for Defense Analyses, Arlington, Virginia

(Received June 6, 1966)

Let, P_A, P_B, and P_AP_B denote the transition probability matrices of finite Markov chains A, B, and the 'product chain' AB. It is demonstrated by counterexamples that the sets of recurrent, periodic, and transient states of A and B cannot be simply related to the corresponding sets of AB.

LET R_A, R_B, and R_{AB} denote the recurrent states of finite Markov chains with transition probability matrices P_A, P_B, and P_AP_B, respectively. Then the following statements are, in general, false:

$$R_A \cap R_B \subseteq R_{AB}, \tag{1}$$

$$R_A \cup R_B \subseteq R_{AB}, \tag{2}$$

$$R_{AB} \subseteq R_A \cap R_B, \tag{3}$$

$$R_{AB} \subseteq R_A \cup R_B, \tag{4}$$

$$R_A \cap R_{AB} \neq 0, \tag{5}$$

$$R_B \cap R_{AB} \neq 0, \tag{6}$$

$$R_{AB} \cap (R_A \cup R_B) \neq 0, \tag{7}$$

$$R_A = R_B \quad \text{implies} \quad R_A = R_{AB}. \tag{8}$$

The case

$$P_A=\begin{pmatrix}0&1&0\\1&0&0\\0&1&0\end{pmatrix},\quad P_B=\begin{pmatrix}0&0&1\\0&0&1\\1&0&0\end{pmatrix},\quad P_AP_B=\begin{pmatrix}0&0&1\\0&0&1\\0&0&1\end{pmatrix},$$

with $R_A=(1, 2)$, $R_B=(1, 3)$, and $R_{AB}=(3)$, provides a counterexample to (1–3).