Mathematical Programming 2 (1972) 258–260. North-Holland Publishing Company

NOTE

EXPLICIT SOLUTION OF AN OPTIMIZATION PROBLEM *

Philip WOLFE

IBM Thomas J. Watson Research Center, Yorktown Heights, N.Y., U.S.A.

Received 17 August 1971
Revised manuscript received 23 December 1971

A simply stated n-variable constrained optimization problem, useful as a test problem, is explicitly solved. It has a large number of easily described local optima.

The following optimization problem was posed by David Feder ** in connection with a maximum-likelihood estimation problem: Let X denote the n-component real vector $(X_1, ..., X_n)$, and $S_k = \Sigma_{j=1}^n X_j^k$ (for $k = 2, 3, 4$). It is required to find X satisfying the constraints $0 \leq X_j \leq 1$ (all j) which maximize the function $f(X) = S_2 S_4 - S_3^2$.

Our solution: Let X be any vector m of whose components equal 1 and p of whose components equal $\frac{1}{2}$, with $m + p = n > 1$. The vector X yields a strict local maximum of f if and only if X is of the above form with $n > m > \frac{1}{9} n$, and yields the global maximum of f if and only if it is of the above form and either m or p is $[\frac{1}{2} n]$.

These facts are useful: letting v be the n-vector given by $v_j = X_j^2$, we have $f(X) = |X|^2 |v|^2 - (X, v)^2$, so by the Schwartz inequality f is nonnegative and vanishes only when X is proportional to v, that is, when all the nonzero components of X are equal; so $f(X) > 0$ for some X. Since f is a nontrivial polynomial it cannot vanish identically on a set of positive volume, so any local maximum value is positive and thus any local maximizing point must have at least two positive components. Since f is

* This work was supported (in part) by the Office of Naval Research under contract number N00014-71-C-0112.

** Department of Statistics, Princeton University, speaking at the Probability and Statictics Seminar at IBM Research Center, Yorktown Heights, N.Y. in October 1970.

homogeneous (of degree 6), its maximum is assumed on the "outer" boundary of the constraint set: $X_j = 1$ for some j.

Let x denote any of the X_j, and denote by $\overline{S}_k$ the sum of the k-th powers of the remaining variables, so that $S_k = \overline{S}_k + x^k$. We find that

$$f(X) = \overline{S}_2\overline{S}_4 - \overline{S}_3^2 + \overline{S}_4 x^2 - 2\overline{S}_3 x^3 + \overline{S}_2 x^4 ,$$

$$\partial f(X)/\partial x = 2\overline{S}_4 x - 6\overline{S}_3 x^2 + 4\overline{S}_2 x^3 = 2S_4 x - 6S_3 x^2 + 4S_2 x^3 ,$$

$$\partial^2 f(x)/\partial x^2 = 2\overline{S}_4 - 12\overline{S}_3 x + 12\overline{S}_2 x^2 .$$

Now suppose that the point X yields a local maximum of f: $S_k > 0$ for all k. The first derivative vanishes at $x = 0$, but the second derivative is positive there, so every component of X is positive. The first expression for the first derivative shows it to be cubic in x; the second expression shows that its (possibly) three real roots, $0, r, s$, say, are independent of the index j. If $X_j < 1$, then X_j is one of the roots r, s. If there is only one root between 0 and 1, then we have it; otherwise – say $0 < r < s < 1$ – since the first derivative is increasing at 0, it is decreasing at r and increasing at s, so that s is not a candidate: $X_j = r$. Thus X_j is either 1 or r, for all j.

Let X have m ones and p r's; then $S_k = m + pr^k$ so $f(X) = mpr^2(1-r)^2$. We must have $r = \frac{1}{2}$ in order for X to give a local maximum, and mp must be maximal for the global maximum, which happens when either m or p equals $[\frac{1}{2}n]$. Thus the maximum value is $[\frac{1}{2}n]\ [\frac{1}{2}(n+1)]/16$, assumed for any X whose components are all either $\frac{1}{2}$ or 1 in as nearly equal numbers as possible.

Now let X be any point having m components 1 and p components $\frac{1}{2}$, with $m + p = n$. Letting as before $x = x_j$ for any j, we find

$$\partial f/\partial x = x(1 - 2x)(2m + \tfrac{1}{8}p - [2m + \tfrac{1}{2}p]\,x) ,$$

which vanishes for $x = \frac{1}{2}$ and has the value $3p/8$ for $x = 1$. Thus X is a (constrained) stationary point, even if m or p is zero; and if $p > 0$, the behavior of f around X is determined by the second-order dependence of f on those components – say $x_1, \ldots, x_p$ – having the value $\frac{1}{2}$ in X

(since their first-order variation vanishes, while the other components may not be varied). Letting T_k be the sum of the k-th powers of those components we have $S_k = m + T_k$ and $f(X) = m(T_2 - 2T_3 + T_4) + T_2 T_4 - T_3^2$. The Hessian of f with respect to the first p variables is readily calculated to be

$$-\tfrac{1}{8}\,[(8m - p)I_p + e_p e_p^{\mathrm{T}}]\ ,$$

where I_p is the identity matrix of order p and e_p the p-component column vector consisting entirely of ones. If $8m - p > 0$, that is, if $m > \frac{1}{9}n$, then the Hessian is negative definite and f is strictly concave in the neighborhood of X, so f has a strict local maximum there, while if $m < \frac{1}{9}n$ the Hessian is indefinite and X is not a local maximum. To handle the case $p = 8m$ set $x_1 = x_2 = t$, $x_3 = \frac{3}{2} - 2t$, $x_4, \ldots, x_p = \frac{1}{2}$, and the remaining variables to one. Direct calculation shows that the first and second derivatives of f with respect to t at $t = \frac{1}{2}$ vanish, but that the third derivative has the value $-72m$, so that again X is not a local maximum.

When $n = 8$ there are, by the above results, $C_4^8 = 70$ global maxima and $2^8 - 1 - 70 = 185$ (or 72%) other local maxima. Fifty runs from random starting points using the penalty-function SUMT routine (see eg. [1]) reached a nonglobal local maximum in 67% of the cases. [The computations were performed by the members of our Fall, 1970 course on nonlinear programming at Columbia University.]

Reference

[1] A.V. Fiacco and C. McCormick, *Nonlinear programming: sequential unconstrained minimization techniques* (Wiley, New York, 1968).