NONLINEAR PROGRAMMING:

THE GLOBAL USE OF THE LAGRANGIAN

by

Philip B. Zwart

This research was supported in part by the Atomic Energy Commission under Research Contract No. A(11-1)-1493 and by the Department of Defense under Themis Grant No. F44620-69-C-0116.

Report No. COO-1493-29

LEGAL NOTICE

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

ABSTRACT

In some constrained nonlinear programming problems possessing multi local optima, a local optimum can be recognized as the global optimum by looking closely at the Lagrangian, the augmented function. Similarly, classes of constrained optimization problems, such as geometric programming problems, can be recognized as possessing at most a single local optimum.

I. Introduction:

The continuous nonlinear constrained optimization problem is to find $x = (x_1, \dots, x_n)$ in order to

$$\begin{array}{ll} \text{Minimize} & F(x) \\ \text{Subject to:} & G_j(x) \leq b_j, j=1, \dots, m. \end{array} \quad (1)$$

Here the F and G_j 's are assumed to be twice differentiable and the b_j 's are real numbers. Fig. 1 illustrates a two dimensional problem. Notice that x^* has the property that it is better than any nearby feasible point. x^* is a local optimum. x^{**} is another local optimum. Moreover, $F(x^{**}) > F(x^*)$. x^{**} is the global optimum.

Nonlinear programming algorithms, such as appear in [3], [4], [7], and [8], for solving (1) have the capability of numerically locating a local optimum. However, the recognition of whether or not a local optimum is actually the global optimum cannot be handled numerically in the general case. In the absence of knowledge of special properties of the F and G_j 's, one can only resort to computing the F and G_j 's at points of a fine grid which covers the feasible region. Global optimization can be effectively accomplished by other methods in some special

problems which possess many local optima. See, for example [2], [6], and [10].

A well-known result (see for instance [5], p. 93) tells us that a large class of problems of (1) possess only a single local optimum. Problems of the form (1) in which the F and G_j 's are convex functions, have the property that any local minimum is a global minimum. As a consequence, current algorithms are truly capable of solving such problems.

Figure 2 illustrates a problem in which the G_j 's are not all convex. The problem possesses a single local optimum, x^* . This local optimum would be found by current methods, but would not be recognized as the global optimum.

A simple observation regarding Lagrange multipliers and the augmented function, the Lagrangian, yields a method for sometimes recognizing a global minimum. This is described in section II. This same observation yields results concerning families of functions which generate problems of the type (1) for which any local minimum must be a global minimum. These results are described in section IV, after a few unimodality definitions are reviewed in section III. Section V shows how these results relate to geometric programming.

Figure 1

Minimize $F(x) = -(x_1 - \frac{1}{2})^2 - (x_2 - \frac{1}{2})^2$

Subject to: $G_1(x) = x_1^2 + 99x_2^2 - 100 \leq 0$

$G_2(x) = 99x_1^2 + x_2^2 - 100 \leq 0$

Figure 2

Maximize $F(x) = (x_1 - 2)^2 + (x_2 - 10)^2$

Subject to: $G_1(x) = x_1^2 + x_2^2 - 73 \leq 0$

$$G_2(x) = -x_1^2 - x_2^2 - 2x_1x_2 + x_1 + 11x_2 - 36 \leq 0$$

$$G_3(x) = -(x_1 - 10)^2 - (x_2 - 100)^2 + 9901 \leq 0$$

II. Recognition of a Global Minimum:

We first make the following observation: If x^* is a global minimum for $\bar{F} = F + \sum_{j=1}^m \lambda_j G_j$, where the λ_j are non-negative real numbers, then x^* is the global minimum of F subject to: $G_j(x) \leq G_j(x^*)$, $j=1, \dots, m$.

Proof: By assumption, for any x ,

$$\bar{F}(x) \geq \bar{F}(x^*). \quad \text{That is,}$$

$$F(x) + \sum_{j=1}^m \lambda_j G_j(x) \geq F(x^*) + \sum_{j=1}^m \lambda_j G_j(x^*).$$

$$\text{Or, } F(x) \geq F(x^*) + \sum_{j=1}^m \lambda_j (G_j(x^*) - G_j(x)).$$

If x satisfies $G_j(x) \leq G_j(x^*)$, $j=1, \dots, m$,

$$\text{then } \sum_{j=1}^m \lambda_j (G_j(x^*) - G_j(x)) \geq 0. \quad \text{Thus,}$$

$$F(x) \geq F(x^*).$$

Q.E.D.

If x^* is a local optimum for (1) then Kuhn-Tucker theory tells us that there exist non-negative λ_j , $j=1, \dots, m$, such that (i) $\lambda_j (G_j(x^*) - b_j) = 0$, $j=1, \dots, m$ and (ii) $\nabla \bar{F}(x^*) = 0$, where $\bar{F} = F + \sum_{j=1}^m \lambda_j G_j$. Let $J(x^*) = \{j | G_j(x^*) = b_j\}$. (i) says that $\lambda_j = 0$, $j \notin J(x^*)$. So $\bar{F}$ can be expressed as $\bar{F} = F + \sum_{j \in J(x^*)} \lambda_j G_j$. If x^* is a global minimum for $\bar{F}$, then according to the above observation x^* is a global minimum of F , subject to: $G_j(x) \leq b_j$, $j \in J(x^*)$. And x^* must, in such a case, be a global minimum for (1).

Of course, x^* generally would not be the global minimum for $\bar{F}$. But in some cases this may be true and in these cases we can recognize that x^* is the global minimum for the constrained problem. Let us consider the examples of section I.

Problem 1 is:

$$\begin{aligned} \text{minimize} \quad & F(x) = -(x_1 - \frac{1}{2})^2 - (x_2 - \frac{1}{2})^2 \\ \text{subject to:} \quad & G_1(x) = x_1^2 + 99x_2^2 - 100 \leq 0 \\ & G_2(x) = 99x_1^2 + x_2^2 - 100 \leq 0. \end{aligned}$$

A steepest descent search begun at $(0,0)$ would lead to the local minimum $(-1,-1)$. At this point we have

$$\nabla F = (3,3), \quad \nabla G_1 = (-2,-198), \quad \nabla G_2 = (-198,-2).$$

The Lagrange multipliers are found to be $\lambda_1 = .015$, $\lambda_2 = .015$. And,

$$\bar{F}(x) = -(x_1 - \frac{1}{2})^2 - (x_2 - \frac{1}{2})^2 + 1.5x_1^2 + 1.5x_2^2 - 3.$$

Now, $\bar{F}$ is a strictly convex function and $(-1,-1)$, being a point where $\nabla \bar{F} = (0,0)$ must be the global minimum of $\bar{F}$.

Thus, $(-1,-1)$ must be the global minimum of F for the constrained problem. Of course, if such analysis were carried out at any other local minimum, the resulting $\bar{F}$ would not be convex.

Problem 2 is:

$$\begin{aligned} \text{minimize} \quad & F(x) = (x_1 - 2)^2 + (x_2 - 10)^2 \\ \text{subject to:} \quad & G_1(x) = x_1^2 + x_2^2 - 73 \leq 0 \\ & G_2(x) = -x_1^2 - x_2^2 - 2x_1x_2 + x_1 + 11x_2 - 36 \leq 0 \\ & G_3(x) = -(x_1 - 10)^2 - (x_2 - 100)^2 + 9901 \leq 0. \end{aligned}$$

A steepest descent search begun at any feasible point would lead to the local minimum (3,8). At this point we have

$$\nabla F = (2, -4), \quad \nabla G_1 = (6, 16), \quad \nabla G_2 = (-15, -5).$$

The Lagrange multipliers are found to be $\lambda_1 = \frac{1}{3}$, $\lambda_2 = \frac{4}{15}$.

And,

$$\bar{F}(x) = \frac{1}{15}(16x_1^2 + 16x_2^2 - 8x_1x_2 - 56x_1 - 256x_2 + 991).$$

The matrix of second derivatives of $\bar{F}$ is positive definite.

So $\bar{F}$ is a convex function, and (3,8), being a point where

$\nabla \bar{F} = (0, 0)$, must be the global minimum of $\bar{F}$. We can conclude

that (3,8) is the global minimum of F for the constrained

problem. Notice that our analysis does not tell us that

(3,8) is the only local minimum for the constrained problem.

II. Unimodal Functions:

D. J. Wilde and C.S. Beichtler discuss unimodal functions in [9], chapter 6. Let us review a few of their definitions.

We describe the various classes of unimodality in terms of minimization, whereas Wilde and Beichtler usually refer to

maximization. A function F is said to be:

- (i) unimodal (n variables) if F possesses a unique local minimum.
- (ii) strictly unimodal (one variable) if F is strictly decreasing at all points on the left of its minimum and strictly increasing at all points on the right of its minimum.
- (iii) differentiably unimodal (one variable) if F is differentiable and $\frac{dF}{dx}$ vanishes only at the minimum of F .
- (iv) strongly unimodal (n variables) if F is differentiable and F is differentiably unimodal on all lines through its minimum point.

Wilde and Beightler define "regular strict unimodality (one variable)" and "rectangular unimodality (n variables)" also. These unimodality classes, which are strongly related to special numerical schemes for locating the optimum, need not be discussed here.

In this paper we define differentiable unimodality (n variables) (in terms of minimization) as follows:

F is said to be differentiably unimodal (n variables) on the open set $R \subseteq E^n$ if F is differentiable on R and $\nabla F(x^*)=0$, $x^* \in R$, implies that x^* is a global minimum for F

restricted to R . (It is not required that such an x^* exist.) In cases where $R = E^n$, we will merely say F is differentially unimodal.

Notice that this definition is a natural extension of (iii) except that this definition includes the possibility that F has no minimum point. Such a possibility is unreasonable in unconstrained optimization, which is the primary concern of Wilde and Beightler in [9], chapter 5. This is probably why it was excluded by way of (i).

Also notice that any function which is "strongly unimodal" must be "differentially unimodal". That is, the class of differentially unimodal functions includes the class of strongly unimodal functions. As mentioned in [9] rectangular unimodality implies strong unimodality. So this last statement holds with "rectangular" in place of "strong".

Wilde and Beightler's discussion is concerned with methods of finding the optimum of unconstrained problems. These results, though of much interest, do not concern us. We will study classes of constrained optimization problems.

IV. Classes of Problems For Which Local Minimality Implies Global Minimality.

As observed in I, if F_i is contained in the family of convex functions, $i=0,1,\dots,m$; then the problem:

$$\begin{array}{ll} \text{minimize} & F_0(x) \\ \text{subject to:} & F_j(x) \leq b_j \quad j=1, \dots, m. \end{array}$$

has the property that any local minimum must be a global minimum. This property is valuable because it enables us to conclude that we have the global minimum once we have found a local minimum. Thus it is of interest to try to find other families of functions which generate problems possessing this property. The concept of differentiable unimodality helps to describe some such families.

We have the following Proposition: Suppose F is a family of functions for which

- (i) $F \in F \implies F$ is differentiable unimodal on R
- (ii) $F \in F \implies \alpha F \in F$ for any positive real number α , and
- (iii) $F_1 \in F, F_2 \in F \implies F_1 + F_2 \in F$,

then any problem of the form:

$$\begin{array}{ll} \text{minimize} & F_0(x) \\ \text{subject to:} & F_j(x) \leq b_j, \quad j=1, \dots, m \\ & x \in R \end{array} \quad (2)$$

where $F_i \in F, i=0,1, \dots, m$, must have the property that any local minimum is a global minimum.

Proof: Suppose x^* is a local minimum to (2). Since R is an open set, x^* is also a local minimum to (2) with the restriction $x \in R$ removed. By Kuhn-Tucker theory there exist $\lambda_j \geq 0$ such that $\nabla \bar{F}(x^*) = 0$, where $\bar{F} = F_0 + \sum_{j=1}^m \lambda_j F_j$. But

$\bar{F} \in F$ because of (ii) and (iii). Consequently, by (i), x^* is the minimum of $\bar{F}$ over all $x \in R$. Finally, by the observation in II, x^* must be a global minimum for (2). (Note: The observation in II clearly generalizes to the case where it is also required that x be contained in some set R .)

Q.E.D.

V. Posynomials and Geometric Programming:

In [1], Duffin, Peterson, and Zener describe the beautiful theory of Geometric Programming, a numerical method for attacking the problem:

$$\begin{aligned} &\text{Minimize} && P_0(x) \\ &\text{Subject to:} && P_j(x) \leq 1, \quad j=1, \dots, m, \\ &&& x_i > 0, \quad i=1, \dots, n, \end{aligned}$$

where each $P_i(x)$, $i=0,1,\dots,m$ is a "posynomial". A posynomial is any function of the form

$$\sum_i C_i x_1^{a_{i1}} x_2^{a_{i2}} \dots x_n^{a_{in}}, \text{ where each}$$

C_i is a positive real number, and each a_{ij} , $j=1,\dots,n$, is a real number.

Posynomials are not contained in the family of convex functions. This provides a nice example of the use of the proposition in section IV on a class of problems which are not included in the standard results of convex programming.

We will show that the family of posynomials satisfies (i), (ii), and (iii) of section IV with $R=\{x|x_i>0, i=1, \dots, n\}$. Clearly, (ii) and (iii) are satisfied. To see (i) we will use an argument related to the material in [1], p. 7.

Suppose

$$P(x) = \sum_{i=1}^r C_i x_1^{a_{i1}} \dots x_n^{a_{in}}, \text{ and } \nabla P(x^*)=0.$$

Then

$$0 = x_j \frac{\partial P}{\partial x_j} = \sum_{i=1}^r C_i a_{ij} x_1^{a_{i1}} \dots x_n^{a_{in}}. \quad (2)$$

$$\text{Set } \delta_i = \frac{C_i x_1^{a_{i1}} \dots x_n^{a_{in}}}{P(x^*)}, \quad i = 1, \dots, r.$$

It follows from (2) that

$$\sum_{i=1}^k a_{ij} \delta_i = 0. \quad (3)$$

And, by the definition of $P(x)$,

$$\sum_{i=1}^k \delta_i = 1. \quad (4)$$

The geometric inequality (using (4) and the fact that $\delta_i > 0$, $i=1, \dots, r$) gives

$$\delta_1 y_1 + \delta_2 y_2 + \dots + \delta_r y_r \geq y_1^{\delta_1} y_2^{\delta_2} \dots y_r^{\delta_r}, \quad (5)$$

for any positive real numbers y_i , $i=1, \dots, r$.

In particular taking $y_i = \frac{C_i x_1^{a_{i1}} \dots x_n^{a_{in}}}{\delta_i}$ in (5) gives

$$P(x) \geq \prod_{i=1}^r \left(\frac{C_i x_1^{a_{i1}} \dots x_n^{a_{in}}}{\delta_i} \right)^{\delta_i}$$

And, using (3),

$$P(x) \geq \prod_{i=1}^r \left(\frac{C_i}{\delta_i} \right)^{\delta_i}. \quad (6)$$

But for $x = x^*$,

$$\frac{C_i x_1^{*a_{i1}} \dots x_n^{*a_{in}}}{\delta_i} = P(x^*) .$$

$$\text{So, } \prod_{i=1}^r \left(\frac{C_i}{\delta_i} \right)^{\delta_i} = \prod_{i=1}^r (P(x^*))^{\delta_i} = P(x^*),$$

because of (4). Substitution in (6) yields $P(x) \geq P(x^*)$,

which shows that x^* is a global minimum of P over R . That is, (iii) holds.

Our proposition says that the problems of geometric programming possess at most a single local minimum. Thus any local optimizer could be used to solve geometric programming problems numerically. Of course, the geometric programming method is probably most suitable in view of its other special qualities.

These comments do not apply to the more general form of geometric programming in which the c_i 's are permitted to be negative.

BIBLIOGRAPHY

1. Duffin, R.J., Peterson, E.L., and Zener, C.L. -
"Geometric Programming", Wiley, New York, N.Y., 1962.
2. Hoang Tui, "Concave Programming under Linear Constraints",
Dokl. Akad. Nauk SSR 159 (1964) 32-35; English
translation in Soviet Math. Dokl. 5 (1964) 1437-1440.
3. Glass, H. and Cooper, L. = "Sequential Search: A
Method for Solving Constrained Optimization Problems",
Journal of the ACM 12 (1965) 71-82.
4. Griffith, R.E. and Stewart, R.A. = "A Non-linear
Programming Technique for the Optimization of Continuous
Processing Systems", Management Science 7 (1961) 379-392.
5. Hadley, G. ' "Nonlinear and Dynamic Programming",
Addison-Wesley, Reading, Mass., 1964.
6. Ritter, K., "A Method for Solving Maximum-Problems
with a Nonconcave Quadratic Objective Function",
Z. Wahrscheinlichkeitstheorie verw. Geb. 4, (1966) 340-351.
7. Rosen, J.B. - "The Gradient Projection Method for
Non-linear Programming, Part I", SIAM Journal 8, (1960)
181-217.
8. Rosen, J.B. - "The Gradient Projection Method for
Non-linear Programming, Part II", SIAM Journal 9,
(1961) 414-432.

9. Wilde, D.J. and Beichtler, C.S. - Foundations of Optimization, Prentice-Hall, Englewood Cliffs, N.J., 1967.
10. Zwart, P.B. - "Global Maximization of a Convex Function with Linear Inequality Constraints", AEC Report No. COO-1493-24, (1969) (Recently submitted to ORSA Journal).