Z. Wahrscheinlichkeitstheorie verw. Geb. 11, 327—337 (1969)

A Relaxed Interior Approach to Nonlinear Programming

JAMES E. FALK

Received February 28, 1968

Summary. This paper contains the mathematical validation of a new approach to mathematical programming problems based on a penalty function method. The given problem is replaced by a second „auxiliary" problem which, in many cases may be solved by standard methods since it involves the maximization of a concave function of a single variable over an interval. The auxiliary problem is defined implicitly in therms of the constituents of the original problem. Examples are presented in order to illustrate the theoretical results.

1. Introduction

The various "penalty function" algorithms which have been introduced in recent years may be placed into two broad categories which might be termed parametric and nonparametric. The class of parametric methods includes the Sequential Unconstrained Minimization Technique (SUMT) of FIACCO and MCCORMICK [3] and the algorithms described by ZANGWILL [8]. Parametric methods are distinguished by the presence of a variable which weights the penalty term associated with the constraints of the problem and which is varied by the problem solver in order to attain the desired solution. Nonparametric methods, such as HUARD's "Methods of Centers" [6] and FIACCO and MCCORMICK's "SUMT Without Parameters" [4], treat the objective function as an additional constraint. Convergence is obtained by a successive tightening of this artificial constraint and the amount of tightening is determined from problem data alone; the problem solver does not independently vary any external parameter to attain the desired solution.

The parametric techniques may be further subdivided into "exterior" and "interior" methods. If the sequence of points generated is feasible, the method is termed interior. By contrast, the exterior methods generate a nonfeasible sequence and the method yields the solution if the sequence ever becomes feasible. An example of an interior method is SUMT whereas the techniques of PIETRZYKOWSKI [7], ZANGWILL [8], and FIACCO and MCCORMICK [5] are exterior algorithms. Only recently nonparametric exterior methods have been introduced [2].

Lagrangian techniques [e. g., see 1] are related to parametric penalty function procedures in that both types deal with the constraints by incorporating them, together with the objective function, into an "auxiliary" function whose unconstrained minima must be computed. The parameter(s) in the auxiliary function are then varied in such a way that the minimizing points of the auxiliary function approach the solution of the original problem.

An excellent survey which includes a description of the methods mentioned above in addition to several others may be found in [2].

The method to be described in this paper is a parametric interior algorithm which has some of the features of the methods of centers and some of the features of a lagrangian approach. Comparisions to existing techniques will be offered from time to time in the sections that follow. Section 2 describes the general method of approach while Section 3 contains the details of the method together with proofs. The structure of the "auxiliary problem" is investigated and it is shown that the method is valid from a mathematical standpoint with the proper assumptions on the constituents of the original problem.

Some examples are given in Section 7.

2. Definitions and Assumptions

The problem which we are interested in solving has the form

$$\begin{aligned} &\text{minimize} && \Phi(x) \\ &\text{subject to} && g(x) \geqq 0 \\ &&& x \in C \end{aligned} \tag{2.1}$$

where $g(x) = (g_1(x), \ldots, g_m(x))$, $x \in E^n$ and C is some nonempty subset of E^n (possibly E^n itself). Let $G = \{x: g(x) \geqq 0\}$ and $S = G \cap C$. For any given problem the decomposition of S into G and C is determined by the nature of the technique to be described. In general, C will be E^n, $(E^n)^+$ or a linear polyhedron in E^n, that is to say, we will generally be interested in letting the nonlinear constraints describe G. Computational considerations will dictate that C should be chosen in such a way that nonlinear functions may be easily minimized over it. The mathematical theory of the method does not depend on the choice of C.

It is necessary to make a certain number of assumptions on the constituents of (2.1) in order to validate the theory. We therefore assume that

(2.2)

a) Φ is convex,

b) g_i is concave $(i = 1, \ldots, m)$,

c) C is closed and convex,

d) $\operatorname{int} G \cap C \neq \Phi$,

e) problem (2.1) has a solution, and

f) the function $\theta(\cdot, r)$ defined in Section 3 possesses a unique minimizing point over C if it is minimized at all.

Let x^* denote the solution of (2.1). Assumption (2.2-f) will imply that x^* is unique.

The following notation will be used throughout the paper.

$(E^n)^+$	$\{x: x = (x_1, \ldots, x_n)^T$ and $x \geqq 0\}$
∂G	boundary of G
$\operatorname{int} G$	interior of G
$x(\cdot)$	alternate notation for the function $x: \tau \to x(\tau)$
$D[\varrho]$	domain of the function ϱ
$N(x^*; \varepsilon)$	open sphere about x^* of radius ε
$\bar{N}$	closure of N.

3. The General Approach

Associated with problem (2.1), we require a function $\pi\colon (E^m)^+ \to (E^1)^+$ such that the composite function $\pi(g(\cdot))$ has the properties

$$\begin{aligned}
&\text{a) } \pi(g(x)) = 0 \quad \text{if} \quad x \in \partial G \cap C,\\
&\text{b) } \pi(g(x)) > 0 \quad \text{if} \quad x \in \operatorname{int} G \cap C,\\
&\text{c) } \pi(g(\,)) \text{ is concave and continuous over } G \cap C.
\end{aligned} \tag{3.1}$$

We shall refer to such functions as penalty terms. Examples of such functions are

$$\begin{aligned}
&\text{a) } \pi(g(x)) = \min\{g_1(x), \ldots, g_m(x)\}\\
&\text{b) } \pi(g(x)) = \prod_{i=1}^{m} (g_i(x))^{\alpha_i} \quad \text{where} \quad \alpha_i > 0, \quad \sum_{i=1}^{m} \alpha_i = 1\\
&\text{c) } \pi(g(x)) = \begin{cases} 0 & \text{if} \quad g_i(x) \quad \text{for some } i \\ \dfrac{1}{\sum\limits_{i=1}^{m} 1/(g_i(x))} & \text{if} \quad g(x) > 0 \end{cases}
\end{aligned} \tag{3.2}$$

We note in passing that similar definitions are made in constructing other parametric penalty function approaches. The SUMT method of Fiacco and McCormick essentially required a function $\tau\colon \operatorname{int}(E^m)^+ \to E^1$ such that the composite function $\tau(g(\cdot))$ has the properties

$$\begin{aligned}
&\text{a) } \tau(g(x^k)) \to -\infty \qquad \text{if} \quad x^k \to \partial G \cap C\\
&\qquad \tau(g(x)) \text{ finite} \qquad \text{if} \quad x \in \operatorname{int} G \cap C\\
&\text{b) } \tau(g(\,)) \text{ concave over } \operatorname{int} G \cap C
\end{aligned} \tag{3.3}$$

in place of (3.1). We also note that Huard's methods of centers (which are non-parametric algorithms) make use of functions of the same type as described in (3.1).

We now define the *penalty function* $\theta\colon E^m \times (E^1)^+ \to E^1$ by means of the equation

$$\theta(x, r) = \Phi(x) - r\,\pi(g(x)) \tag{3.4}$$

where $r \geqq 0$ is a scalar parameter. We note that $\theta(\cdot, r)$ is convex as a function of x and

$$\theta(x, r) \begin{cases} = \Phi(x) & \text{if} \quad x \in \partial G \cap C, \\ < \Phi(x) & \text{if} \quad x \in \operatorname{int} G \cap C. \end{cases} \tag{3.5}$$

Thus it would appear that $\theta(\cdot, r)$ is more likely than Φ to have a minimum in the set $\operatorname{int} G \cap C$. This is desirable from a computational standpoint if the constraints defining G are troublesome, for knowing that the minimizing point avoids all constraints of the form $g_i(x) \geqq 0$ essentially means that an unconstrained minimization technique may be employed in determining this minimizing point. Of course, it could happen that $\theta(\cdot, r)$ is not minimized over S at all for certain values of r.

When dealing with SUMT, the function $\theta(\cdot, r)$ has the properties

$$\theta(x, r) = \begin{cases} +\infty & x \in \partial G \cap C, \\ \text{finite} & x \in \operatorname{int} G \cap C. \end{cases}$$

Thus the function $\theta(\cdot, r)$ generally attains its minimum over S for fixed r and this

minimum *must* occur in the set $\operatorname{int} G \cap C$. As above, the constraints $g_i(x) \geqq 0$ essentially may be ignored in the determination of this minimum.

We now define the *auxiliary function* $\varrho\colon D[\varrho] \to E^1$ by means of the relations

$$D[\varrho] = \{r \geqq 0\colon \theta(\cdot, r) \text{ attains a minimum over } S\},$$
$$\varrho(r) = \min_{x \in S} \theta(x, r). \tag{3.6}$$

The point $X(r)$ which minimizes $\theta(\cdot, r)$ over S is unique. Hence $X\colon r \to X(r)$ is a single valued mapping of $D[\varrho]$ into S and we shall refer to this as the minimizing function.

In the following section, we describe the properties of the auxiliary function and the minimizing function. It will be shown that there is a close relationship between problem (2.1) and the auxiliary function and in order to solve (2.1) it suffices to maximize ϱ over $D[\varrho]$. The function ϱ turns out to be concave and differentiable over $D[\varrho]$ and hence should be easy to maximize. A difficulty arises in the fact that the determination of ϱ at any particular value of r requires the solution of a nonlinear program. It is shown, however, that this nonlinear program is relatively easy to solve if r is sufficiently large, for in this case the minimization of $\theta(\cdot, r)$ is effectively taken over C rather than $G \cap C$. On the other hand, if r is too large, the term $-r\pi(g(x))$ dominates in θ and it could happen that θ does not attain a minimum over S.

The similarity between this method and the lagrangian method [see e. g., 1] is apparent here, for both types of algorithms attempt to maximize an auxiliary function over its domain. In general, the lagrangian method makes use of several "multipliers" and does not destroy the structure of the constraint functions, whereas the present method uses a single parameter and lumps all of the constraints together. A more important difference lies in the fact that the determination of the value of the auxiliary function is *always* unconstrained (with respect to C) using the lagrangian method whereas here this calculation is unconstrained only for certain values of r.

4. The Structure of the Auxiliary Function

In this section we describe the properties of the auxiliary function ϱ defined by (3.5) in detail and present the mathematical validation of these properties. Often it will be convenient to derive certain relationships that exist between ϱ and Problem (2.1) and we shall do this when the occasion arises, rather than wait until the properties of ϱ have been established completely.

We first characterize the set $D[\varrho]$ by showing that it is not empty and is, in fact, an interval. Let

$$r^* = \inf_{x \in \operatorname{int} G \cap C} \frac{\Phi(x) - \Phi(x^*)}{\pi(g(x))},$$

where x^* is the solution of Problem (2.1). Note that $r^* \geqq 0$.

Theorem 1. *Under the hypotheses of Section 2, the point* $r^* \in D[\varrho]$ *so that* $D[\varrho] \neq \Phi$.

Proof. From the definition of r^* we have

$$r^* \pi(g(x)) \leqq \Phi(x) - \Phi(x^*)$$

i.e.,

$$\Phi(x^*) \leqq \Phi(x) - r^* \pi(g(x))$$

for all $x \in S$. In particular, since $x^* \in S$, we have

$$r^* \pi(g(x^*)) \leqq 0.$$

But r^* and $\pi(g(x^*))$ are both nonnegative so that

$$r^* \pi(g(x^*)) = 0.$$

Hence

$$\begin{aligned}\theta(x^*, r^*) &= \Phi(x^*) - r^* \pi(g(x^*)) \\ &\leqq \Phi(x) - r^* \pi(g(x))\end{aligned}$$

for all $x \in S$ and the theorem is proved.

The next corollary follows directly from the above proof and the definition of $X(r)$.

Corollary. *Under the hypotheses of Section* 2,

a) $X(r^*) = x^*$,

b) $r^* \pi(g(x^*)) = 0$.

We now show that $D[\varrho]$ contains more than the single point r^*.

Theorem 2. *Under the hypotheses of Section* 2, *the interval* $[r^*, r^* + \delta]$ *is contained in* $D[\varrho]$ *for some* $\delta > 0$.

Proof. Let $N = N(x^*; \varepsilon)$ be an open neighborhood of x^* of radius $\varepsilon > 0$ where

$$\partial N \cap \operatorname{int} G \cap C \neq \Phi.$$

Such a neighborhood exists since we have assumed that $\operatorname{int} G \cap C \neq \Phi$ and $x^* \in S$. Let r^* be given by (4.1) and define

$$\begin{aligned}\mu_1 &= \theta(x^*, r^*) \\ \mu_2 &= \min_{\partial N \cap S} \theta(x, r^*) \\ \nu_1 &= \pi(g(x^*)) \\ \nu_2 &= \max_{\partial N \cap S} \pi(g(x)).\end{aligned}$$

Note that $\mu_2 > \mu_1$ since the unique minimum of θ occurs at x^*. Also note that $\nu_2 > 0$ since $\partial N \cap \operatorname{int} G \cap C \neq \Phi$ and $\pi(g(x)) > 0$ when $x \in \operatorname{int} G \cap C$.

Choose $\delta > 0$ so that

$$\delta < \begin{cases} \dfrac{\mu_2 - \mu_1}{3\nu_2} & \text{if} \quad \nu_1 = 0 \\ \min\left(\dfrac{\mu_2 - \mu_1}{3\nu_1}, \dfrac{\mu_2 - \mu_1}{3\nu_2}\right) & \text{if} \quad \nu_1 \neq 0. \end{cases}$$

Now when $r = r^* + \delta$ we have

$$\begin{aligned}|\theta(x^*, r) - \theta(x^*, r^*)| &= \delta \pi(g(x^*)) \\ &= \delta \nu_1 \\ &< \frac{\mu_2 - \mu_1}{3}\end{aligned}$$

so that

$$\theta(x^*, r) < \mu_1 + \frac{\mu_2 - \mu_1}{3}.$$

In words, the value of $\theta(\cdot, r)$ at x^* differs from μ_1 by at most $\frac{\mu_2 - \mu_1}{3}$.

On the other hand, on the set $\partial N \cap S$ we have

$$\begin{aligned} |\theta(x, r^*) - \theta(x, r)| &= \delta\, \pi(g(x)) \\ &\leqq \delta\, \nu_2 \\ &< \frac{\mu_2 - \mu_1}{3}, \end{aligned}$$

so that

$$\theta(x, r^*) < \frac{\mu_2 - \mu_1}{3} + \theta(x, r).$$

This implies that

$$\mu_2 - \frac{\mu_2 - \mu_1}{3} < \theta(x, r),$$

which is to say that $\theta(\cdot, r)$ differs from μ_2 by at most $\frac{\mu_2 - \mu_1}{3}$ at any point in $\partial N \cap S$.

Now $\theta(\cdot, r)$ must be minimized somewhere in $\bar{N} \cap S$ since this set is compact. It cannot attain this minimum along $\partial N \cap S$ since there is a point (i.e., x^*) in $N \cap S$ which gives a lower value to $\theta(\cdot, r)$ than any point in $\partial N \cap S$. Hence $\theta(\cdot, r)$ attains its minimum over $\operatorname{int} N \cap S$. By assumption (2.2f) this minimum is attained at a unique point and the proof is complete.

Note that the special character of r^* was not used in the proof of this theorem. Indeed, we could have replaced r^* and x^* by any point $r' \in D[\varrho]$ and $x(r')$ to obtain the property that $D[\varrho]$ is either an interval or a collection of intervals each open on the right. It will be shown that $D[\varrho]$ is, in fact, an interval.

We know that $D[\varrho]$ always contains some points to the right of r^*. Without any additional assumptions, it is not generally true that the entire interval $[r^*, +\infty]$ lies in $D[\varrho]$. Nor is it true that there need be points to the left of r^* for, in particular, r^* could be zero but $D[\varrho] \subset (E^1)^+$.

The corollary which follows is of prime importance since it implies that the constraints $g_1(x) \geqq 0$ can be effectively ignored in computing $\varrho(r)$ for $r > r^*$.

Corollary. *If* $r^+ \in D[\varrho]$, $r^+ > r^*$, *then*

$$X(r^+) \in \operatorname{int} G \cap C.$$

Proof. If $x^+ = X(r^+) \in \partial G \cap C$, then

$$\begin{aligned} \Phi(x^+) &= \min_{x \in S} \{\Phi(x) - r^+ \pi(g(x))\} \\ &\leqq \Phi(x^*) - r^+ \pi(g(x^*)) \\ &\leqq \Phi(x^*). \end{aligned}$$

But since x^* is the unique solution of (2.1), either $x^+ = x^*$ or we have a contradiction. Suppose, then, that $x^+ = x^*$. It follows that

$$\Phi(x^+) = \Phi(x^*) \leqq \Phi(x) - r^+ \pi(g(x))$$

for all $x \in S$. Thus

$$r^+ \leqq \frac{\Phi(x) - \Phi(x^*)}{\pi(g(x))}$$

for all $x \in \operatorname{int} G \cap C$. But this is the defining property of r^* so that we arrive at a contradiction and the proof is complete.

It can be shown that $0 \leqq r < r^*$ implies that $X(r) = X(r^*)$, so that it is as difficult to obtain $\varrho(r)$ for such r's as it is to obtain the solution x^* since an unconstrained minimization technique cannot be employed in this instance.

The next theorem characterizes $D[\varrho]$ as an interval. The proof of this theorem is essentially identical to the *proof* of Theorem 8 in reference [1] and hence is omitted here.

Theorem 3. *Under the hypotheses of Section 2, $D[\varrho]$ is convex.*

Having established the nature of $D[\varrho]$ we now turn to the properties of ϱ. The *proof* of the following theorem is straightforward.

Theorem 4. *Under the hypotheses of Section 2, ϱ is concave and monotone decreasing.*

Eventually we whish to establish the differentiability of ϱ. To do this, it is convenient to first establish the continuity of $X(\cdot)$.

Theorem 5. *$X(\cdot)$ is continuous over $D[\varrho]$.*

Proof. Fix $r^+ \in D[\varrho]$ and $\varepsilon > 0$. We must show that there is a $\delta > 0$ such that

$$|r - r^+| < \delta \quad \text{and} \quad r \in D[\varrho]$$

implies that

$$\| X(r) - x^+ \| < \varepsilon,$$

where $x^+ = X(r^+)$. Assume that $\partial N(x^+; \varepsilon) \cap S$ is not empty for otherwise the theorem is trivial. Set $N = N(x^+; \varepsilon)$ and

$$M > \max_{x \in \partial N \cap S} |\pi(g(x)) - \pi(g(x^+))|$$

and let $\delta > 0$ be any number such that

$$\frac{1}{M} \{\Phi(x) - \Phi(x^*) - r^+(\pi(g(x)) - \pi(g(x^+)))\} > \delta$$

for all $x \in \partial N \cap S$. Pick $r \in D[\varrho] \cap N(r^+; \delta)$.

Then

$$\begin{aligned} \{\Phi(x) - r^+ \pi(g(x))\} - \{\Phi(x^+) - r^+ \pi(g(x))\} &> M\delta \\ &> |r - r^+| \, |\pi(g(x)) - \pi(g(x^+))| \\ &\geqq (r - r^+) \{\pi(g(x)) - \pi(g(x^+))\} \end{aligned}$$

so that

$$\Phi(x) - r\pi(g(x)) > \Phi(x^+) - r\pi(g(x^+))$$

for all $x \in \partial N \cap S$. Since $r \in D[\varrho]$, θ attains its minimum over S. The last inequality shows that it cannot be minimized over $\partial N \cap S$ so that the minimum

must occur in $\operatorname{int} N \cap S$, i.e.,

$$\| X(r) - x^+ \| < \varepsilon$$

and the proof is complete.

We are now in a position to prove that ϱ is differentiable over $D[\varrho]$ and exhibit its derivative.

Theorem 6. *Under the hypotheses of Section 2, ϱ is continuously differentiable over $D[\varrho]$ with*

$$\frac{d\varrho}{dr} = -\pi(g(x(\cdot))).$$

Proof. Fix $r^0 \in D[\varrho]$ and $\eta > 0$. The definition of ϱ allows us to write

$$\begin{aligned} \frac{\varrho(r^0+\eta) - \varrho(r^0)}{\eta} &\leqq \frac{\theta(x^0, r^0+\eta) - \theta(x^0, r^0)}{\eta} \\ &= -\frac{(r^0+\eta)\,\pi(g(x^0)) - r^0\,\pi(g(x^0))}{\eta} \\ &= -\pi(g(x^0)), \end{aligned}$$

where $x^0 = X(r^0)$.

On the other hand,

$$\begin{aligned} \frac{\varrho(r^0+\eta) - \varrho(r^0)}{\eta} &\geqq \frac{\theta(X(r^0+\eta), r^0+\eta) - \theta(X(r^0+\eta), r^0)}{\eta} \\ &= -\pi(g(X(r^0+\eta))). \end{aligned}$$

The continuity of π, g and $X(\cdot)$ then implies that the right hand derivative of ϱ is

$$\frac{\mathrm{d}\varrho}{d\eta^+}\bigg|_{r^0} = -\pi(g(x^0)).$$

Similar reasoning establishes that the left hand derivative also equals the above expression and the proof is complete.

The same method of proof will establish the next

Corollary. *The right derivative of ϱ exists at the left endpoint of $D[\varrho]$ and is equal to $-\pi(g(\cdot))$ evaluated there.*

Thus, up to this point we have shown that ϱ is a differentiable concave function defined over an interval in $(E^1)^+$. The next two theorems relate this auxiliary function to problem (2.1) and show that the problem of maximizing ϱ over $D[\varrho]$ will yield the solution of (2.1). We shall term the problem of maximizing ϱ over $D[\varrho]$ the *auxiliary problem.*

Theorem 7. *Under the hypotheses of Section 2, the following relationship holds:*

$$\varrho(r) \leqq \Phi(x)$$

for all $r \in D[\varrho]$ and $x \in S$.

Proof. Follows directly from the definitions of ϱ and π.

Theorem 8. *Under the hypotheses of Section 2, the point r^* given by (4.1) maximizes ϱ over $D[\varrho]$. Furthermore, $X(r^*) = r^*$ and $\varrho(r^*) = \Phi(x^*)$.*

Proof. From the proof of Theorem 1 we know that $\theta(\cdot, r^*)$ attains its minimum over S at x^*. The corollary to this theorem states that $X(r^*) = r^*$ and

$r^* \pi(g(x^*)) = 0$. Hence $\varrho(r^*) = \Phi(x^*)$. Since

$$\varrho(r) \leqq \Phi(x^*)$$

for any $r \in D[\varrho]$, it follows that ϱ is maximized at r^* and the proof is complete.

Having established the relationship between the auxiliary problem and problem (2.1), we now describe the behavior of the functions $\pi(g(X(\cdot)))$ and $\Phi(X(\cdot))$ as functions of r in order to be able to describe the type of convergence obtained as $r \to r^*$.

Theorem 9. *Under the hypotheses of Section 2 the functions* $\pi(g(X(\cdot))$ and $\Phi(X(\cdot))$ *are monotone increasing.*

Proof. From the definition of ϱ, we have

$$\Phi(X(r')) - r' \pi(g(X(r'))) \leqq \Phi(X(r^0)) - r' \pi(g(X(r^0)))$$

and

$$\Phi(X(r^0)) - r^0 \pi(g(X(r^0))) \leqq \Phi(X(r')) - r^0 \pi(g(X(r'))).$$

Adding these two inequalities, we obtain

$$(r^0 - r') \pi(g(X(r'))) \leqq (r^0 - r') \pi(g(X(r^0))).$$

Since $r^0 > r'$, we have

$$\pi(g(X(r'))) \leqq \pi(g(X(r^0)))$$

so that $\pi(g(X(\cdot)))$ is monotone increasing.

Using this fact, and the inequality

$$\Phi(X(r')) - r' \pi(g(X(r'))) \leqq \Phi(X(r^0)) - r' \pi(g(X(r^0)))$$

we obtain

$$\Phi(X(r')) \leqq \Phi(X(r^0))$$

and the proof is complete.

5. Some Computational Considerations

The problem of maximizing ϱ over $D[\varrho]$ is a one dimensional maximization problem to which a number of algorithms could be applied. Any such algorithm will, however, require the evaluation of ϱ at a number of points in $D[\varrho]$. Each such evaluation requires the solution of a non-linear program. With C chosen properly, and with $r' > r^*$, we know from the corollary to Theorem 2 that the problem of minimizing $\theta(\cdot, r')$ over S is essentially the problem of minimizing $\theta(\cdot, r')$ over C. However, unlike other interior methods (such as SUMT), the set $D[\varrho] \neq \{r > 0\}$ in general. For if too large a value of r is chosen, the penalty term dominates in the expression for $\theta(\cdot, r)$ and thus $\theta(\cdot, r)$ may not possess a minimum over C. On the other hand, if $r < r^*$ the determination of ϱ is as difficult as finding the solution of the original problem. A possible method for avoiding these difficulties would be to choose C compact and carry out all auxiliary computations over this set.

To initiate the algorithm, it is possible to use a "SUMT-like" algorithm to calculate a value of $r \in D[\varrho]$. Suppose the functions Φ and $\pi(g(\cdot))$ are diffrentiable and let t be any positive number. The function

$$\tau(g(\cdot)) = -\frac{1}{\pi(g(\cdot))}$$

satifies the properties (3.3) if $\pi(g(\cdot))$ satisfies (3.2). If assumptions (2.2) hold, the function

$$\Phi(x) - t\,\tau(g(x))$$

will attain a minimum at some point $x(t) \in \operatorname{int} G \cap C$. Since Φ and $\tau(g(\cdot))$ are differentiable, we have

$$\nabla \Phi(x) - t\,\nabla \tau(g(x)) = 0$$

at $x(t)$. In terms of π, we have

$$\nabla \Phi(x) - \frac{t}{(\pi(g(x)))^2}\,\nabla \pi(g(x)) = 0$$

at $x(t)$ so that the number $r^0 = t/\pi(g(x(t)))$ is in $D[\varrho]$ and exceeds r^*. Moreover, with this choice of r^0, it is clear that $X(r^0) = x(t)$.

If problem (2.1) fails to have a solution because Φ is unbounded from below, Theorem 7 implies that $D[\varrho] = \Phi$. Thus it would be impossible to initiate the algorithm as in the above paragraph since $\Phi(x) - t\,\tau(g(x))$ would also be unbounded from below over C. The same situation holds if $G \cap C = \Phi$ for in this case also, SUMT cannot produce an initial minimum of its penalty function.

Methods for choosing r^0 and a "maximizing sequence" of r's are under investigation. As compared to SUMT, these questions are more critical since $D[\varrho]$ is limited and since r^* is unknown at the outset. In SUMT, $D[\varrho] = \{r > 0\}$ and it is required to evaluate

$$\lim_{r \to 0} X(r) \qquad \lim_{r \to 0} \varrho(r).$$

Any sequence of r's tending to zero will accomplishing this from a mathematical viewpoint.

On the other hand, in SUMT the evaluation of $X(r)$ and $\varrho(r)$ are relatively difficult for small values of r since $\theta(\cdot, r)$ is very large near $\partial G \cap C$. In the present method, $r\pi(g(X(r)))$ approaches 0 as r approaches r^* and $\theta(\cdot, r) = \Phi(x)$ for $x \in \partial G \cap C$ so that $\theta(\cdot, r)$ should be relatively easy to minimize for r near r^* (assuming Φ is well behaved).

6. Example

Minimize

$$\Phi(x) = (x_1 - 1)^2 + (x_2 - 2)^2$$

subject to

$$x_1 - x_2^2 \geqq 0$$
$$2 - x_1 - x_2 \geqq 0.$$

Using the penalty term π given by (3.2b) we have

$$\varrho(r) = \min_{x \in S} \{(x_1 - 1)^2 + (x_2 - 2)^2 - r(x_1 - x_2^2)^{1/2}(2 - x_1 - x_2)^{1/2}\}.$$

This function, together with the function $\Phi(X(\cdot))$ are depicted in Fig. 1. The determination of $X(\cdot)$ for various values of r was obtained by using a variation of the Fletcher-Reeves conjugate gradient method. Note the irregular curvature

of the graph of $\Phi(X(\cdot))$. The solution of this problem is

$$X_1^* = X_2^* = 1$$

and

$$\varrho(r^*) = \Phi(x^*) = 1 .$$

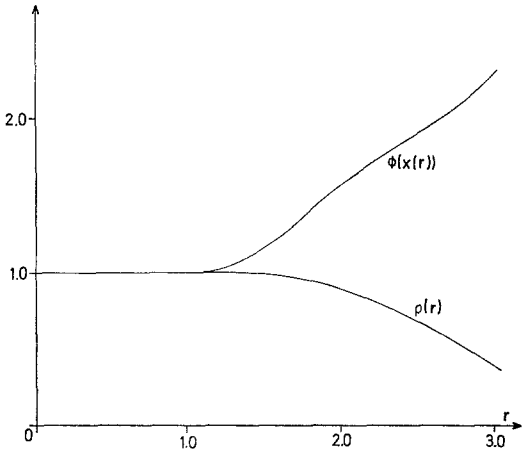

Fig. 1. The auxiliary function of example 2

References

1. Falk, J. E.: A Lagrangian approach to nonlinear programming. Journal of Mathematical Analysis and Applications **19**, 141—159 (1967).
2. Fiacco, A. V.: Sequential unconstrained minimization methods for nonlinear programming. Doctoral thesis, Northwestern University, 1967.
3. —, and G. P. McCormick: The sequential unconstrained minimization technique for nonlinear programming, a primal-dual method. Management Science **10**, 360—366 (1964).
4. — — SUMT without parameters. System Research Memo. No. 121, Tech. Inst. Evanston, Illinois: Northwestern University 1965.
5. — — The slacked unconstrained minimization technique for convex programming. J. Soc. industr. appl. Math. (To appear.)
6. Huard, P.: Resolution des P. M. a contraintes non linear par la methode des centres. Note E. D. F. no. HR 5-690 du 6 Mai 1964.
7. Pietrzykovski, F.: Application of the steepest descent method to concave programming. Proceedings of the IFIPS Congress, Munich. Amsterdam, Holland: North Holland Company 1962.
8. Zangwill, W. I.: Nonlinear programming via penalty functions. Management Science **13**, 344—358 (1967).

Dr. James E. Falk
Department of Industrial Engineering
and Management Sciences
The Technological Institute
Northwestern University
Evanston, Ill. 60201, USA