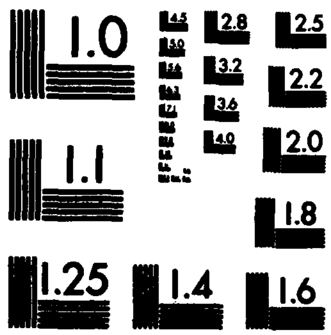
AD-A142 905 GLOBALLY CONVERGENT PROCEDURES FOR SOLVING NONLINEAR 1/1
MINIMIZATION PROBLEMS(U) WISCONSIN UNIV-MADISON
MATHEMATICS RESEARCH CENTER A EYDELAND MAY 84
UNCLASSIFIED MRC-TSR-2699 DAAG29-80-C-0041 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

MRC Technical Summary Report #2699

GLOBALLY CONVERGENT PROCEDURES FOR SOLVING NONLINEAR MINIMIZATION PROBLEMS

Alexander Eydeland

Mathematics Research Center
University of Wisconsin—Madison
610 Walnut Street
Madison, Wisconsin 53705

May 1984

(Received April 11, 1984)

Sponsored by

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

UNIVERSITY OF WISCONSIN-MADISON
MATHEMATICS RESEARCH CENTER

GLOBALLY CONVERGENT PROCEDURES FOR SOLVING NONLINEAR MINIMIZATION PROBLEMS

Alexander Eydeland

Technical Summary Report #2699
May 1984

ABSTRACT

In this work we describe globally convergent iterative procedures for generating a minimizing sequence for the problem of finding the infimum for the function $\Phi(x) = \sum f_i(x)$ on a certain set D. The minimizing sequence consists of points of the infimum for functions of the type $\sum [\beta_i(g_i(x,f_i(x)) + \langle\gamma_i,x\rangle]$, where the transforming functions $g_i(x,\cdot)$ are chosen in such a way that the compositions $g_i(x,f_i(x))$ are simpler than the given functions $f_i(x)$, and where the coefficients $\beta_i > 0$, $\gamma_i \in R^n$, $i = 1,\ldots,m$, are determined by the choice of g_i. Various classes of the functions g_i are considered and global convergence results are proved. It is shown that many well-known algorithms, for example, Weiszfeld's algorithm and Newton's method are particular cases of the general method.

AMS (MOS) Subject Classifications: 49D37, 65K05

Key Words: Nonlinear minimization problem, globally convergent minimization procedure, transformation of the objective function

Work Unit Number 5 (Optimization and Large Scale Systems)

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant Nos. MCS-8210950 and MCS-8301628.

SIGNIFICANCE AND EXPLANATION

The method developed in this paper applies to a number of nonlinear minimization problems. The bibliography includes the numerical implementation for several applied problems. The main advantages of our approach are: generality of the technique, global convergence of the method, its ease of formulation and implementation, and its numerical efficiency. The method can be extended to variational problems.

A1

The responsibility for the wording and views expressed in this descriptive summary lies with MRC, and not with the author of this report.

GLOBALLY CONVERGENT PROCEDURES FOR SOLVING NONLINEAR MINIMIZATION PROBLEMS

Alexander Fydeland

1. In this paper we discuss the method of nonlinear transformation of the objective function. This approach has proved to be a convenient tool for generating and analyzing optimization procedures. The method is based on the following simple observation. The minimization problem for a given function $f(x)$, $x \in R^n$, can be replaced by the problem of minimizing the composition $g(f(x))$, where $g(\alpha)$ is a monotone function, provided that the minimization problem for $g(f(x))$ can be solved with less effort than the original problem of minimizing $f(x)$. Of course the minimum points for both functions coincide.

Now we consider a more complicated case, when we have to find a sequence $\{x^k\}$ such that

$$\lim_{k\to\infty} \Phi(x^k) = \inf_{x\in D} \Phi(x) ;$$

(1) $$\Phi(x) = \sum_{i=1}^{m} f_i(x) ;$$

$$f_i(x) \geq 0, \quad i = 1,\dots,m ,$$

where D is a convex set in R^n. The assumption that the given functions $f_i(x)$ are non-negative is made for convenience and does not lead to the loss of generality. Assume that we can find a set of functions $g_i(x,\alpha)$, $i = 1,\dots,m$, such that the compositions $g_i(x,f_i(x))$ are simpler than the functions $f_i(x)$ (for example, the $g_i(x,f_i(x))$ are quadratic or linear functions). Can we use the existence of the simplifying functions g_i in order to construct an iterative procedure which determines a minimizing sequence for the problem (1)? In this paper we make an attempt to answer this question. We construct a minimizing sequence for the problem (1) which consists of points of infimum of functions, whose general form is $\sum_{i=1}^{m} (\beta_i g_i(x,f_i(x)) + \langle\gamma_i,x\rangle)$, where $\beta_i > 0$ and $\gamma_i \in R^n$. By

---

Sponsored by the United States Army under Contract No. DAAG29-80-C-0041. This material is based upon work supported by the National Science Foundation under Grant Nos. MCS-8210950 and MCS-8301628.

virtue of the assumption above, the problem of finding points of infimum for these linear combinations is simpler than the original problem (1).

In the sections 2, 3, 4 and 6 we consider various classes of functions $\Phi(x)$ and $g_i(x,\alpha)$, construct the corresponding iterative procedures and prove convergence results. Some of those results are from the papers [3], [5] and we only state them without proofs. Other results are new and we consider them in detail. We do not discuss numerical applications of the method in this paper, although they are numerous. For some numerical examples we refer the reader to the papers [1], [2]. Other important applied problems will be considered in forthcoming papers. In Section 5 we show that some of the well-known optimization procedures, such as Weiszfeld's algorithm and Newton's method, are particular cases of the method of nonlienar transformation of the objective function. Throughout this paper the superscript k denotes the iteration number.

2. We start with the simplest case when the functions g_i depend on α only and for every $i = 1,\dots,m$

$$
\begin{aligned}
&g_i(\alpha) > 0 \quad \text{for} \quad \alpha > 0 \,, \\
(2) \qquad &g_i(\alpha) \quad \text{is convex for} \quad \alpha > 0 \,, \\
&g_i(\alpha) \in C^1[0,\infty), \quad g_i'(f_i(x)) > \delta > 0 \quad \forall\, x \in D \,,
\end{aligned}
$$

where δ is some positive constant. We assume that for every set of positive constants c_i, $i = 1,\dots,m$, the set $\operatorname{Arg}\inf_{x\in \mathrm{cl} D} \sum_{i=1}^{m} c_i g_i(f_i(x)) \neq \emptyset$, where for any function $\psi(x)$ and any set B the set $\operatorname{Arg}\inf_{x\in B} \psi(x)$ $(\operatorname{Arg}\min_{x\in B} \psi(x))$ is the set of all points of infimum (minimum) for the function $\psi(x)$ in the set B. Now we consider the following iterative procedure:

$$
\begin{aligned}
&x^0 \in D \quad \text{is an arbitrary initial point;} \\
(3) \qquad &\text{for} \quad k = 0,1,2,\dots \\
&x^{k+1} \in \operatorname{Arg}\inf_{x\in \mathrm{cl} D} \sum_{i=1}^{m} g_i(f_i(x))/g_i'(f_i(x^k)) \,.
\end{aligned}
$$

For this procedure we can prove the following results.

Theorem 1. Let the function $\Phi(x)$, defined in (1), be convex, let the functions $f_i(x)$, $i = 1,\ldots,m$, be continuous on $c\ell D$, non-negative and finite in D and let the functions $g_i(\alpha)$ satisfy the conditions (2). If the sequence $\{x^k\}$, determined by the procedure (2), has a bounded subsequence then

$$\Phi(x^k) \downarrow \inf_{x \in D} \Phi(x) \quad \text{as} \quad k \to \infty .$$

Proof. We introduce the function

$$\phi(x,y) = \Phi(y) + \sum_{i=1}^{m} [g_i(f_i(x)) - g_i(f_i(y))]/g_i'(f_i(y)), \quad x,y \in D . \tag{4}$$

From (3) and the definition (4) of the function $\phi(x,y)$ if follows that

$$\phi(x^{k+1},x^k) < \phi(x^k,x^k) = \Phi(x^k) . \tag{5}$$

On the other hand

$$\phi(x^{k+1},x^k) = \Phi(x^{k+1}) + (\Phi(x^k) - \Phi(x^{k+1}))$$

$$+ \sum_{i=1}^{m} [g_i(f_i(x^{k+1})) - g_i(f_i(x^k))]/g_i'(f_i(x^k))$$

$$= \Phi(x^{k+1}) + \sum_{i=1}^{m} [g_i(f_i(x^{k+1})) - g_i(f_i(x^k))$$

$$- g_i'(f_i(x^k))(f_i(x^{k+1}) - f_i(x^k))]/g_i'(f_i(x^k)) .$$

Each term in the square brackets in the last sum in (6) is non-negative by virtue of the convexity of the functions $g_i(\alpha)$. Since, by (2), $g_i'(f_i(x^k)) > \delta > 0$, we finally have that

$$\phi(x^{k+1},x^k) > \Phi(x^{k+1}) = \phi(x^{k+1},x^{k+1}) . \tag{7}$$

Combining the inequalities (6) and (7) together we obtain the sequence of inequalities

$$\ldots \phi(x^{k+1},x^{k+1}) < \phi(x^{k+1},x^k) < \phi(x^k,x^k) < \ldots , \tag{8}$$

or

$$\ldots \le \Phi(x^{k+1}) \le \phi(x^{k+1},x^k) \le \Phi(x^k) \le \ldots .$$

The function $\Phi(x)$ is bounded from below. Therefore there exists

$$\lim_{k\to\infty} \Phi(x^k) = \bar{\Phi} . \tag{9}$$

We shall prove now that $\bar{\Phi} = \inf_{x\in D} \Phi(x)$. By the assumption of the theorem there exists a subsequence $\{x^{k_\nu}\}$ of the sequence $\{x^k\}$ such that $x^{k_\nu} \to \bar{x} \in c\ell D$ as $\nu \to \infty$. It is clear that

$$\Phi(\bar{x}) = \bar{\Phi} = \lim_{k\to\infty} \Phi(x^k) = \lim_{k\to\infty} \phi(x^{k+1},x^k) . \tag{10}$$

Let us assume that

$$\bar{\Phi} = \Phi(\bar{x}) > \inf_{x\in D} \Phi(x) . \tag{11}$$

Then there eixsts a point $x^* \in D$ such that the right directional derivative

$$\frac{d^+}{d\tau} \Phi(\bar{x} + \tau(x^* - \bar{x}))\Big|_{\tau=0} = \lim_{\tau\downarrow 0} \frac{\Phi(\bar{x} + \tau(x^* - x)) - \Phi(\bar{x})}{\tau} < 0 . \tag{12}$$

Taking into account that for every pair of points $x_1,x_2 \in c\ell D$

$$\frac{d^+}{d\tau} \phi(x_1 + \tau(x_2 - x_1),x_1)\Big|_{\tau=0} = \frac{d^+}{d\tau} \Phi(x_1 + \tau(x_2 - x_1))\Big|_{\tau=0}$$

we obtain from (12) that

$$\frac{d^+}{d\tau} \phi(\bar{x} + \tau(x^* - \bar{x}),\bar{x})\Big|_{\tau=0} < 0 . \tag{13}$$

This inequality, convexity of the function $\phi(x,y)$ with respect to x for every fixed $y \in c\ell D$ and the results of §24 of [8] imply that there exists a point $z = \bar{x} + \bar{\tau}(x^* - \bar{x})$, $\bar{\tau} \in [0,1]$, such that

$$\phi(z,\bar{x}) < \phi(\bar{x},\bar{x}) .$$

Since the function $\phi(x,y)$ is continuous and $x^{k_\nu} \to \bar{x}$ as $\nu \to \infty$, from the inequality above it follows that there exists ν_0 such that

$$\phi(z,x^{k_\nu}) < \phi(\bar{x},\bar{x})$$

for $\nu > \nu_0$. By (3) from this inequality it follows that

$$\phi(x^{k_\nu+1}, x^{k_\nu}) < \phi(\bar{x},\bar{x}),\ \nu > \nu_0 ,$$

which contradicts the fact that by (8) and (10), $\phi(x^{k+1},x^k) \geq \phi(\bar{x},\bar{x}) = \bar{\Phi}$ for all $k = 0,1,\ldots$. Therefore the assumption (11) was false and

$$\Phi(\bar{x}) = \lim_{k\to\infty} \Phi(x^k) = \inf_{x\in D} \Phi(x) .$$

The theorem is proved.

In the following theorem we remove the condition of the existence of a bounded subsequence of the sequence $\{x^k\}$.

<u>Theorem 2</u>. Let the functions $\Phi(x)$ and $f_i(x)$, $i = 1,\ldots,m$, satisfy the conditions of Theorem 1. Moreover, let the functions $f_i(x)$ be convex outside a circle of some radius R. Then

$$\Phi(x^k) \downarrow \inf_{x\in D} \Phi(x) \quad \text{as} \quad k \to \infty ,$$

where the sequence $\{x^k\}$ is determined by (3).

<u>Proof</u>. If the sequence $\{x^k\}$ has a bounded subsequence then the assertion of the theorem follows from Theorem 1. Let us assume now that $|x^k| \to \infty$ as $k \to \infty$. We consider seperately two cases.

A) In this case we assume that there exists a point $x^* \in \mathrm{c}\ell D$ such that $\Phi(x^*) = \inf_{x\in D} \Phi(x)$. Since the sequence $\{(x^k - x^*)/|x^k - x^*|\}$ is bounded for all k there exists a recession direction

$$\xi = \lim_{\nu\to\infty} (x^{k_\nu+1} - x^*)/|x^{k_\nu+1} - x^*| \tag{14}$$

for some subsequence $\{x^{k_\nu+1}\}$, $\nu = 0,1,\ldots$. Consider now the set

$$\omega = K_\xi^\varepsilon \cap H_R \cap \mathrm{c}\ell D ,$$

where ε is a small positive constant, the cone $K_\xi^\varepsilon = \{x^* + \tau\eta | \tau > 0;$

$\eta \in R^n, |\eta| = 1, |\eta - \xi| \leq \varepsilon\}$, a closed half-space H_R is chosen in such a way that for all $\eta \in K_\xi^\varepsilon \cap H_R$ $|\eta| > R$. It is clear that the set ω is convex if ε is sufficiently small, and that ω is relatively closed. By the assumptions of the theorem the functions $f_i(x)$ are convex in ω. Now we introduce the set

$$\Omega = c\ell\{\beta \in R^m \mid \exists x \in \omega \text{ such that } \beta_i > f_i(x),\ i = 1,\dots,m\} .$$

The set Ω is convex. Indeed let β' and β'' belong to Ω. Then there exist points $x',x'' \in \omega$ such that $\beta_i' > f_i(x')$, $\beta_i'' > f_i(x'')$, $i = 1,\dots,m$. Hence the point $\lambda\beta' + (1 - \lambda)\beta''$, $\lambda \in [0,1]$, also belongs to Ω since, by convexity of the functions $f_i(x)$ in ω,

$$\lambda\beta_i' + (1 - \lambda)\beta_i'' > f_i(\lambda x' + (1 - \lambda)x''),\quad i = 1,\dots,m .$$

On the set $\Omega \times \Omega$ we introduce the function

$$\phi_1(\alpha,\beta) = \sum_{i=1}^{m} \beta_i + \sum_{i=1}^{m} [g_i(\alpha) - g_i(\beta)]/g_i'(\beta) . \tag{15}$$

It is obvious that if $\alpha_i = f_i(x)$, $\beta_i = f_i(y)$, $x,y \in D$, $i = 1,\dots,m$, then $\phi_1(\alpha,\beta) = \phi(x,y)$, see (4).

By the definition of the set ω and by (14) we may assume, without the loss of generality, that $x^{k_\nu+1} \in \omega$, $\nu = 0,1,\dots$. As in Theorem 1 we can prove that $\Phi(x^k) \to \bar{\Phi}$ as $k \to \infty$, where $\bar{\Phi}$ is some positive constant. Therefore

$$\Phi(x^{k_\nu}) \to \bar{\Phi},\quad \phi(x^{k_\nu+1},x^{k_\nu}) \to \bar{\phi} \quad \text{as} \quad \nu \to \infty . \tag{16}$$

Let us assume that

$$\bar{\Phi} > \inf_{x\in D} \Phi(x) = \Phi(x^*) . \tag{17}$$

Consider now a new sequence $\{z^{k_\nu+1}\}$, $\nu = 0,1,\dots,$ defined as follows:

$$z^{k_\nu+1} = [x^{k_\nu+1},x^*] \cap \partial w,\quad \nu = 0,1,\dots ,$$

where by $[x^{k_\nu+1},x^*]$ we denote a segment of a straight line between $x^{k_\nu+1}$ and x^*. Since x^* is the vertex of the cone K_ξ^ε, the point $z^{k_\nu+1}$ always exists and, by closedness of ω, $z^{k_\nu+1} \in \omega$, $\nu = 0,1,\dots$. Moreover from the definition of the set ω

it follows that the sequence $\{|z^{k_\nu+1} - x^*|\}$ is bounded by some constant $T(R,\varepsilon)$. By convexity of the function $\Phi(x)$

$$0 < \frac{\Phi(z^{k_\nu+1}) - \Phi(x^*)}{\Phi(x^{k_\nu+1}) - \Phi(x^*)} \le \frac{|z^{k_\nu+1} - x^*|}{|x^{k_\nu+1} - x^*|}, \quad \nu = 0,1,\dots .$$

As $\nu \to \infty$ the denominator $\Phi(x^{k_\nu+1}) - \Phi(x^*) \to \bar{\Phi} - \Phi(x^*)$, which is not equal to zero by the assumption (17), and $\frac{|z^{k_\nu+1} - x^*|}{|x^{k_\nu+1} - x^*|} \le \frac{T(R,\varepsilon)}{|x^{k_\nu+1} - x^*|} \to 0$, since $|x^{k_\nu+1} - x^*| \to \infty$. Therefore

$$\Phi(z^{k_\nu+1}) \downarrow \Phi(x^*) \quad \text{as} \quad \nu \to \infty . \tag{18}$$

Thus, there exist in ω two sequences $\{x^{k_\nu+1}\}$ and $\{z^{k_\nu+1}\}$ such that $\Phi(x^{k_\nu}) \downarrow \bar{\Phi}$ and $\Phi(z^{k_\nu}) \downarrow \inf\limits_{x \in D} \Phi(x)$ as $\nu \to \infty$. We consider the *corresponding* sequences $\{\beta^{k_\nu+1}\}$ and $\{\gamma^{k_\nu+1}\}$ in Ω:

$$\beta_i^{k_\nu+1} = f_i(x^{k_\nu+1}), \quad \gamma_i^{k_\nu+1} = f_i(z^{k_\nu+1}) , \quad i = 1,\dots,m; \ \nu = 0,1,\dots . \tag{19}$$

From (16) we have that

$$\sum_{i=1}^{m} \beta_i^{k_\nu+1} \downarrow \bar{\Phi} \quad \text{as} \quad \nu \to \infty , \tag{20}$$

and from (18)

$$\sum_{i=1}^{m} \gamma_i^{k_\nu+1} \downarrow \Phi(x^*) \quad \text{as} \quad \nu \to \infty . \tag{21}$$

From (20) and (21) and from the fact that $\beta_i^{k_\nu+1} > 0$, $\gamma_i^{k_\nu+1} > 0$, $i = 1,\dots,m$, $\nu = 0,1,\dots$, we obtain, taking smaller subsequences if necessary, that

$$\beta_i^{k_\nu+1} \to \bar{\beta}_i, \quad \gamma_i^{k_\nu+1} \to \bar{\gamma}_i \quad \text{as} \quad \nu \to \infty, \quad i = 1,\dots,m . \tag{22}$$

Since Ω is a closed set, $\bar{\beta} \in \Omega$ and $\bar{\gamma} \in \Omega$. From (20), (21) and (22) it follows that

$$(23) \qquad \sum_{i=1}^{m} \bar{\beta}_i = \bar{\Phi}, \quad \sum_{i=1}^{m} \bar{\gamma}_i = \Phi(x^*) .$$

Finally from the assumption (17) it follows that

$$(24) \qquad \sum_{i=1}^{m} \bar{\beta}_i > \sum_{i=1}^{m} \bar{\gamma}_i .$$

Let us prove now that

$$(25) \qquad \beta^{k_\nu+1} \in \operatorname{Arg}\min_{\beta\in\Omega} \phi_1(\beta,\beta^{k_\nu}), \quad \nu = 0,1,\dots ,$$

where $\beta_i^{k_\nu} = f_i(x^{k_\nu})$, $i = 1,\dots,m$ and the function $\phi_1(\alpha,\beta)$ is defined in (15). Indeed, if there exists a vector $\alpha \in \Omega$ such that $\phi_1(\alpha,\beta^{k_\nu}) < \phi_1(\beta^{k_\nu+1},\beta^{k_\nu})$ then, by monotonicity of the $g_i(\alpha)$ and by the definition of Ω, there exists a point $z \in \omega$ such that $\phi_1(F(z),\beta^{k_\nu}) < \phi_1(\beta^{k_\nu+1},\beta^{k_\nu})$, where for every $z \in \omega$ the vector function $F(z)$ is defined as follows: $F_i(z) = f_i(z)$. Therefore, recalling the definitions of the functions ϕ_1 and ϕ, we obtain that $\phi(z,x^{k_\nu}) < \phi(x^{k_\nu+1},x^{k_\nu})$. The inequality obtained contradicts the definition of $x^{k_\nu+1}$, see (3). Thus (25) is proved.

Let us prove now that

$$(26) \qquad \frac{d^+}{d\tau} \phi_1(\bar{\beta} + \tau(\bar{\gamma} - \bar{\beta}),\bar{\beta})\Big|_{\tau=0} \geq 0 ,$$

where the definition of the right derivative $d^+/d\tau$ is given in (12). If we assume that this directional derivative is negative then, by convexity of the function $\phi_1(\alpha,\beta)$ with respect to α for any fixed β, there exists a point $\bar{\alpha} \in \Omega$ such that $\phi_1(\bar{\alpha},\bar{\beta}) < \phi_1(\bar{\beta},\bar{\beta})$. Therefore, since $\phi_1(\alpha,\beta)$ is continuous with respect to β, it follows from (22) that for some large ν, $\phi_1(\bar{\alpha},\beta^{k_\nu}) < \phi_1(\bar{\beta},\bar{\beta})$. By (25) this means that $\phi_1(\beta^{k_\nu+1},\beta^{k_\nu}) < \phi_1(\bar{\beta},\bar{\beta}) = \sum_{i=1}^{m} \bar{\beta}_i$. Recalling (23) and the fact that $\phi_1(\beta^{k_\nu+1},\beta^{k_\nu}) = \phi(x^{k_\nu+1},x^{k_\nu})$ we obtain that

$$\phi_1(x^{k_\nu+1},x^{k_\nu}) < \bar{\Phi} ,$$

which contradicts (16). Thus the inequality (26) holds. Calculating the derivative in

(26) we obtain that

$$\sum_{i=1}^{m} \bar{\gamma}_i - \sum_{i=1}^{m} \bar{\beta}_i \geq 0 , \tag{27}$$

which is in contradiction with the assumption (24). Therefore the assumption (17) is false and

$$\Phi(x^k) \downarrow \bar{\Phi} = \Phi(x^*) = \inf_{x \in D} \Phi(x) \quad \text{as} \quad k \to \infty .$$

Thus the part A) of the theorem is proved. Consider now the second case.

B) The minimizing sequence $\{y^\ell\}$, $\ell = 0,1,\ldots,$ such that $\Phi(y^\ell) \downarrow \inf_{x \in D} \Phi(x)$ as $\ell \to \infty$, is unbounded. Taking y^ℓ instead of x^* and repeating the arguments of the part A) we can obtain in the same manner as we have obtained the inequality (27) that $\Phi(y^\ell) \geq \bar{\Phi}$, $\ell = 0,1,\ldots$. From this it follows that $\lim_{\ell \to \infty} \Phi(y^\ell) = \inf_{x \in D} \Phi(x) \geq \bar{\Phi}$. On the other hand $\inf_{x \in D} \Phi(x) \leq \bar{\Phi}$, since $\bar{\Phi} = \lim_{k \to \infty} \Phi(x^k)$. Therefore $\inf_{x \in D} \Phi(x) = \bar{\Phi}$ and the theorem is proved.

To complete the investigation of the procedure (3) we prove several results concerning the rate of convergence of that procedure.

It is clear that the rate of convergence for procedure (3) depends on the choice of the functions $g_i(\alpha)$. Below we shall prove several estimates for the convergence rate under various conditions on the $g_i(\alpha)$. We shall always assume that there exists a point $x^* \in \operatorname{Arg} \inf_{x \in D} \Phi(x) \subset c\ell D$.

First of all we observe that, by (7), (3) and (4),

$$\Phi(x^{k+1}) \leq \inf_{\tau \in [0,1]} \phi(x^k + \tau(x^* - x^k), x^k) =$$

$$= \inf_{\tau \in [0,1]} \left\{ \Phi(x^k) + \sum_{i=1}^{m} [g_i(f_i(x^k + \tau(x^* - x^k))) - g_i(f_i(x^k))]/g_i'(f_i(x^k)) \right\} .$$

It is obvious that by continuity we can replace inf by min above. Hence

$$(28)\qquad \Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} \Big\{\Phi(x^k + \tau(x^* - x^k)) - \Phi(x^k)$$

$$+ \sum_{i=1}^{m} \int_{f_i(x^k)}^{f_i(x^k+\tau(x^*-x^k))} ds[g_i'(s) - g_i'(f_i(x^k))]/g_i'(f_i(x^k))\Big\} .$$

Now we use the inequality (28) to prove the following lemma.

Lemma 3. Let $\Phi(x)$ be a convex function and there exists $x^* \in \operatorname{Arg\,inf} \Phi(x)$. Then the following convergence estimates hold.

A) If the functions $f_i(x)$ and $g_i'(f_i(x))$, $i = 1,\dots,m$, are convex then

$$\Phi(x^n) - \Phi(x^*) \le (C_1 + C_2 n)^{-1}, \quad n = 0,1,\dots ,$$

where C_1 and C_2 are positive constants.

B) If $g_i(\alpha) = G_i\alpha^{p_i}$, $i = 1,\dots,m$, $G_i > 0$, $1 < p_i \le 2$, and if the functions $f_i(x)$ are convex then

$$\Phi(x^n) - \Phi(x^*) \le (C_3 + C_4 n)^{-1}, \quad n = 0,1,\dots ,$$

where C_3 and C_4 are positive constants.

C) If the functions $f_i(x)$, $i = 1,\dots,m$, are convex, if the sequence $\{x^k\}$ from (3) is bounded, if the functions $g_i \in C^2(\mathbb{R}_+)$, $i = 1,\dots,m$, and $g_i''(f_i(x)) \le M$, $i = 1,\dots,m$, $x \in D$, $M > 0$, and if there exists a positive q, $1 \le q \le 2$, such that

$$(29)\qquad \Phi(x) - \Phi(x^*) \ge \eta|x - x^*|^q, \quad x \in D ,$$

where η is a positive constant, then

i) if $q = 2$

$$(30)\qquad \Phi(x^n) - \Phi(x^*) = (\Phi(x^0) - \Phi(x^*))Q_1^n, \quad n = 0,1,\dots ,$$

where $0 < Q_1 < 1$;

ii) if $1 \le q < 2$ there exists a number n_0 such that

$$(31)\qquad \Phi(x^n) - \Phi(x^*) \le Q_2^{(2/q)^{n-n_0}-a(n-n_0)}, \quad n = n_0, n_0 + 1,\dots ,$$

where $0 < Q_2 < 1$ and a is some constant.

Proof of A). The function $g_i'(\alpha)$, $i = 1,\ldots,m$, are increasing functions. Therefore the inequality (28) can be replaced by the following inequality:

$$\Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} \Big\{\Phi(x^k + \tau(x^* - x^k)) - \Phi(x^k)$$

$$+ \sum_{i=1}^{m} [(g_i'(f_i(x^k + \tau(x^* - x^k))) - g_i'(f_i(x^k))) \cdot$$

$$(f_i(x^k + \tau(x^* - x^k)) - f_i(x^k))]/g_i'(f_i(x^k))\Big\} .$$

By convexity of the functions $\Phi(x)$, $g_i'(f_i(x))$ and $f_i(x)$, $i = 1,\ldots,m$, we obtain that

$$\Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} [\tau(\Phi(x^*) - \Phi(x^k))$$

(32)

$$+ \sum_{i=1}^{m} \frac{\tau^2}{\delta} [g_i'(f_i(x^*)) - g_i'(f_i(x^k))][f_i(x^*) - f_i(x^k)] ,$$

where δ is determined in (2). From Theorem 2 we know that the sequence $\{\Phi(x^k)\}$ converges. Therefore, since the functions $f_i(x)$ are non-negative, the sequences $\{f_i(x^k)\}$, $i = 1,\ldots,m$, are bounded. Hence from (32) if follows that

(33) $$\Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} [\tau(\Phi(x^*) - \Phi(x^k)) + M\tau^2] .$$

where M is the bound for $\frac{1}{\delta}\sum_{i=1}^{m} [g_i'(f_i(x^*)) - g_i'(f_i(x^k))][f_i(x^*) - f_i(x^k)]$ $k = 0,1,\ldots$.

Denoting $\Phi(x^k) - \Phi(x^*)$ by μ^k we can rewrite (33) as

$$\mu^{k+1} \le \min_{\tau\in[0,1]} [(1 - \tau)\mu^k + \tau^2 M] .$$

Optimal τ is $\mu^k/2M$. Therefore

(34) $$\mu^{k+1} \le \mu^k - (\mu^k)^2/(2M) .$$

By the standard technique, for example see [1], we obtain from (34) that

$$\mu^n \le (C_1 + C_2 n)^{-1}, \quad n = 0,1,\dots ,$$

where $C_1 = \mu^0 = \Phi(x^0) - \Phi(x^*)$, $C_2 = 1/2M$.

Proof of B). We consider this case separately because it appears frequently in applications. First we observe that from (2) and from the condition that $g_i(\alpha) = G_i \alpha^{p_i}$ it follows that for every $x \in D$, $f_i(x) \ge \delta_i$, $i = 1,\dots,m$, where $\delta_i = (\delta/G_i)^{1/p_i}$. From (28), from convexity of $\Phi(x)$ and concavity of the functions $g_i'(\alpha)$ it follows that

$$\Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} [\tau(\Phi(x^*) - \Phi(x^k)) + \frac{(p_i - 1)}{2\delta}\tau^2 \sum_{i=1}^m (f_i(x^*) - f_i(x^k))^2 .$$

After this the proof of B) can be completed in the same way as was done in part A).

Proof of C). From (28), convexity of $\Phi(x)$ and the condition that $g_i''(f_i(x)) \le M$ it follows that

$$\Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} [\tau(\Phi(x^*) - \Phi(x^k)) + \frac{M}{2\delta}\sum_{i=1}^m (f_i(x^k + \tau(x^* - x^k)) - f_i(x^k))^2 , \tag{35}$$

where δ is defined in (2). Since by the condition of the lemma there exists a bounded convex set $S \subset cl D$ such that $x^k \in S$ for every $k \ge 0$, we can use the results of §24 from [8] to prove that $\max_{s\in[0,1]} \frac{d^+}{ds} f_i(x^k + s(x^* - x^k)) \le F$ for $k = 0,1,\dots$ and $i = 1,\dots,m$, where F is some positive constant. Therefore it follows from (35) that

$$\Phi(x^{k+1}) \le \Phi(x^k) + \min_{\tau\in[0,1]} [\tau(\Phi(x^*) - \Phi(x^k)) + \frac{MF}{2\delta}\tau^2(x^* - x^k)^2] .$$

From the condition that $(x^* - x^k)^2 \le [\frac{1}{\eta}(\Phi(x^k) - \Phi(x^*))]^{2/q}$ we have, denoting as usual $\Phi(x^k) - \Phi(x^*)$ by μ^k, that

$$\mu^{k+1} \le \mu^k + \min_{\tau\in[0,1]} [-\tau\mu^k + L(\mu^k)^{2/q}\tau^2] , \tag{36}$$

where $L = MF/(2\delta\eta^{2/q})$. Without loss of generality we assume that $L \ge 1/2$. If $q = 2$, which is the case of a twice differentiable $\Phi(x)$, then we take in (36) the optimal $\tau = 1/2L$ and obtain

$$\mu^{k+1} \leq \mu^k(1 - 1/4L) .$$

Hence

$$\mu^n \leq \mu^0 \rho_1^n, \quad n = 0,1,\ldots ,$$

where $\rho = 1 - 1/4L < 1$, $\rho > 0$ since $L \geq 1/2$. Thus we have proved the inequality (30). If $1 \leq q < 2$ then there exists, by Theorem 2, a number n_0 such that $\mu^k < 1$ for all $k \geq n^0$. We take $\tau^k = 1$ for $k \geq n_0$ in (36) and obtain

$$\mu^{k+1} \leq L(\mu^k)^{2/q}, \quad k = n_0, n_0 + 1,\ldots .$$

From this the inequality (31) follows with $a = \log_b L$, $b = \mu^{n_0}$. The lemma is proved.

In Lemma 3 we have investigated only the convergence of the sequence $\{\Phi(x^k)\}$ to $\Phi(x^*)$. If one is interested in convergence of the sequence $\{x^k\}$ to x^*, provided that this convergence takes place, one can easily obtain the necessary estimates using the standard technique, which combines the results of Lemma 3 and inequalities similar to the inequality (29).

Remark. In Lemma 3 we have considered only few possible cases, although frequently encountered in applications. Our intention was not to investigate all cases, but rather to suggest an approach to the problem of finding the rate of convergence for the procedure (3).

3. We now consider a wider than in Section 1 class of problems to which we can apply the method of nonlinear transformation of the objective function. In this section we remove the last condition in (2) and assume that each function $g_i(\alpha)$ satisfies the following conditions for every $i = 1,\ldots,m$:

(37a) $g_i(\alpha) \geq 0$, $g_i(\alpha)$ is convex for $\alpha \geq 0$;

(37b) $g_i'(\alpha) > 0$ for $\alpha > 0$; there exist an interval $[0,r]$ and continuous functions $F_i(u)$ and $G_i(v)$ such that

(37c) the function $F_i(u) > 0$ for $u > 0$;

(37d) the function $G_i(v) > 0$ for $v \in (0, g_i'(r)]$ and is a monotone non-decreasing function on this interval;

for every $\alpha', \alpha'' \in (0,r]$

$$(37e)\qquad \frac{g_i(\alpha') - g_i(\alpha'')}{g_i'(\alpha'')} \geq (\alpha' - \alpha'') + G_i(g_i'(\alpha''))F_i\left(\left|1 - \frac{g_i'(\alpha')}{g_i'(\alpha'')}\right|\right) .$$

Remark. For a given function $g_i(\alpha)$ the existence of the functions F_i, G_i and of the number r, which satisfy the conditions (37c-e) does not seem to be obvious. However the results of [5] indicate that the cases in which F_i, G_i and r do not exist are exceptional. Indeed in [5] we have explicitly constructed the functions G_i and F_i for every function $g_i(\alpha)$ which satisfies conditions (37a,b) and for which one can find $r > 0$ and $p_i > 0$ such that $\int_0^r |g_i''(\alpha)|^{p_i} d\alpha < \infty$. Of course the standard functions, such as α^p, $\exp(\alpha)$, etc., which are used the most in applications, satisfy the condition above.

Now we consider a procedure, which is similar to the procedure (3), but where we take into account that the denominator $g_i'(f_i(x^k))$ can be equal to zero at those points x^k where $f_i(x^k) = 0$. The modified procedure is as follows.

$x^0 \in D$ is an arbitrary initial point;

for $k = 0,1,2,\ldots$

$$(38)\qquad \alpha_i^k = \max\{f_i(x^k), \delta_i^k\}, \quad i = 1,\ldots,m ;$$

$$x^{k+1} = \operatorname{Arg} \inf_{x \in c\ell D} \sum_{i=1}^{m} g_i(f_i(x))/g_i'(\alpha_i^k) ,$$

where the sequences $\{\delta_i^k\}$, $i = 1,\ldots,m$, are chosen to satisfy the conditions

$$(39)\qquad \delta_i^k > 0 \text{ for } k = 0,1,\ldots \text{ and } i = 1,\ldots,m; \quad \sum_{k=0}^{\infty} G_i(g_i'(\delta_i^k)) = \infty, \quad i = 1,\ldots,m .$$

Note that the sequences $\{\delta_i^k\}$ which satisfy (39) always exist.

As in Section 2 we assume that the set $\operatorname{Arg} \inf_{x \in c\ell D} \sum_{i=1}^{m} c_i g_i(f_i(x)) \neq \emptyset$ for every collection of positive coefficients $c_1, \ldots, c_m$ and that the problem of finding a point in this set is easier numerically than solving the original problem (1). For the procedure (38) we can establish the following results (see proofs in [5]).

Theorem 4. Assume that the functions $f_i(x)$, $i = 1, \ldots, m$, are convex, finite and non-negative in $c\ell D$. If the functions $g_i(\alpha)$, $i = 1, \ldots, m$, satisfy the conditions (37a-e) then

$$\Phi(x^k) \downarrow \inf_{x \in D} \Phi(x) \quad \text{as} \quad k \to \infty ,$$

where the sequence $\{x^k\}$ is determined in (38).

Theorem 5. Let the functions $f_i(x)$, $i = 1, \ldots, m$, be non-negative and let their right directional derivatives be upper-semicontinuous in $c\ell D \times R^n$. Let the function $\Phi(x)$ be pseudo-convex in D and let its right directional derivative be upper semicontinuous in $c\ell D \times R^n$. Assume moreover that for every $\gamma > 0$ the sets $\mu_\gamma = \{x \in D | \Phi(x) < \gamma\}$ are bounded. If the functions $g_i(\alpha)$, $i = 1, \ldots, m$, satisfy the conditions (33a-e) then

$$\Phi(x^k) \downarrow \inf_{x \in D} \Phi(x) \quad \text{as} \quad k \to \infty$$

where the sequence $\{x^k\}$ is determined in (38).

4. In this section we consider a new class of transforming functions. The functions g_i, $i = 1, \ldots, m$, depend now both on x and α. Moreover we shall allow the functions g_i to change with k. We assume that for $i = 1, \ldots, m$ and for $k = 0, 1, \ldots$

(39)

$$g_i^k(x,\alpha) > 0 \quad \text{for } x \in c\ell D, \ \alpha > 0 \ ;$$

$$g_i^k(x,\alpha) \quad \text{are convex in} \quad c\ell D \times R_+ \ ;$$

$$g_i^k(x,\alpha) \in C^2(c\ell D \times R_+), \qquad \frac{\partial g_i^k}{\partial \alpha}(x, f_i(x)) \geq \delta > 0 \quad \forall x \in D \ ;$$

$$\left\| \frac{\partial^2}{\partial x_i \partial x_i} g_i^k(x, f_i(x)) \right\| \leq C_0, \quad x \in D ,$$

where C_0 is some positive constant. Note that C_0, which is the upper bound for the norm of the Hessian of $g_i^k(x,f_i(x))$, does not depend on k and i. We consider now the following iterative procedure:

$x^0 \in D$ is an arbitrary initial point,

for $k = 0,1,\ldots$

$$x^{k+1} \in \operatorname{Arg}\inf_{x\in D} \sum_{i=1}^{m} [g_i^k(x,f_i(x)) \tag{40}$$

$$- \langle x, \frac{\partial g_i^k}{\partial x}(x^k,f_i(x^k))\rangle / \frac{\partial g_i^k}{\partial \alpha}(x^k,f_i(x^k)) \;,$$

where $k = 0,1,\ldots$. For the procedure (40) we can prove the following theorem.

<u>Theorem 6</u>. Let the $C^2(D)$ functions $\Phi(x)$ and $f_i(x)$, $i = 1,\ldots,m$, satisfy the conditions of Theorem 1 and let the functions $g_i^k(x,\alpha)$ satisfy the conditions (39). If the sequence $\{x^k\}$, determined in (40), has a bounded subsequence $\{x^{k_\nu}\}$ then

$$\Phi(x^k) \to \inf_{x\in D} \Phi(x) \quad \text{as} \quad k \to \infty \;.$$

<u>*Proof*</u>. *As in Theorem 1* we introduce the function

$$\phi^k(x,y) = \Phi(y) + \sum_{i=1}^{m} [g_i^k(x,f_i(x)) - g_i^k(y,f_i(y))$$

$$- \langle x - y, \frac{\partial g_i^k}{\partial x}(y,f_i(y))\rangle] / \frac{\partial g_i^k}{\partial \alpha}(y,f_i(y)), \quad k = 0,1,\ldots \;.$$

From (40) it follows that

$$\phi^k(x^{k+1},x^k) \le \phi^k(x^k,x^k) = \Phi(x^k) \;.$$

On the other hand

$$(41)\qquad \phi^k(x^{k+1},x^k) = \Phi(x^{k+1}) + \sum_{i=1}^{m} [g_i^k(x^{k+1},f_i(x^{k+1}))$$

$$- g_i^k(x^k,f_i(x^k)) - \langle x^{k+1} - x^k, \frac{\partial g_i^k}{\partial x}(x^k,f_i(x^k))\rangle$$

$$- (f_i(x^{k+1}) - f_i(x^k)) \frac{\partial g_i^k}{\partial \alpha}(x^k,f_i(x^k))] \Big/ \frac{\partial g_i^k}{\partial \alpha}(x^k,f_i(x^k)) .$$

Since the functions $g_i^k(x,\alpha)$, $i = 1,\ldots,m$, are convex, from (41) it follows that $\phi^k(x^{k+1},x^k) \geq \Phi(x^{k+1}) = \phi^{k+1}(x^{k+1},x^{k+1})$. Thus, as in Theorem 1, we have the sequence of inequalities

$$\ldots \leq \Phi(x^{k+1}) = \phi^{k+1}(x^{k+1},x^{k+1}) \leq \phi^k(x^{k+1},x^k) \leq \phi^k(x^k,x^k) = \Phi(x^k) \leq \ldots$$

Therefore there exists

$$\lim_{k\to\infty} \Phi(x^k) = \lim_{k\to\infty} \phi^k(x^k,x^k) = \lim_{k\to\infty} \phi^k(x^{k+1},x^k) = \bar{\Phi} .$$

We assume now that

$$\bar{\Phi} > \inf_{x\in D} \Phi(x) .$$

Then as in Theorem 1 there exist $x^* \in c\ell D$ and $\varepsilon > 0$ such that

$$\frac{d}{d\tau} \Phi(x^{k_\nu} + \tau(x^* - x^{k_\nu}))\Big|_{\tau=0} \leq -\varepsilon < 0$$

for every $\nu = 0,1,\ldots,$ where by the assumption of the lemma, the subsequence $\{x^{k_\nu}\}$ is bounded. Therefore for every ν

$$\frac{d}{d\tau} \phi^{k_\nu}(x^{k_\nu} + \tau(x^* - x^{k_\nu}),x^{k_\nu})\Big|_{\tau=0} \leq -\varepsilon < 0$$

By the last condition in (39) and by the boundedness of the subsequence $\{x^{k_\nu}\}$ there exists a constant $M > 0$ such that

$$\max_{\tau\in[0,1]} \frac{d^2}{d\tau^2} \phi^{k_\nu}(x^{k_\nu} + \tau(x^* - x^{k_\nu}),x^{k_\nu}) \leq M, \quad \nu = 0,1,\ldots .$$

Combining the two inequalities above we obtain that

$$\phi^{k_\nu}(x^{k_\nu} + \tilde{\tau}(x^* - x^{k_\nu}), x^{k_\nu}) \leq \phi(x^{k_\nu}, x^{k_\nu}) - \varepsilon^2/2M ,$$

where $\tilde{\tau} = \varepsilon/M$. Therefore, recalling that $x^{k_\nu+1} \in \operatorname{Arg}\inf_{x\in D} \phi^{k_\nu}(x, x^{k_\nu})$, we have that

$$\phi^{k_\nu}(x^{k_\nu+1}, x^{k_\nu}) \leq \phi^{k_\nu}(x^{k_\nu}, x^{k_\nu}) - \varepsilon^2/2M$$

for every ν. It is clear that this contradicts the fact that $\lim_{\nu\to\infty} \phi^{k_\nu}(x^{k_\nu+1}, x^{k_\nu}) = \lim_{\nu\to\infty} \phi^{k_\nu}(x^{k_\nu}, x^{k_\nu}) = \bar{\Phi}$. Thus our assumption that $\bar{\Phi} > \Phi(x^*)$ was false. The theorem is proved.

Remark 1. As follows from the proof of Theorem 6, the second condition in (39) is excessive. We only need to require that for every k the functions $g_i^k(x,\alpha)$ are convex in the set $M^k = \{x \in D | \Phi(x) < \Phi(x^k)\}$ for $i = 1,\ldots,m$.

Remark 2. If the functions $g_i(x,\alpha)$, $i = 1,\ldots,m$, do not depend on k, then we can prove for the procedure (40) the exact analog of Theorem 1 with only first three conditions in (39) and requiring that $g_i(x,\alpha) \in C^1(c\ell D \times R_+)$. The proof will be just a repetition of the proof of Theorem 1 with the function $\phi(x,y)$ defined as in Theorem 6. The estimates of the convergence rates of the procedure (40) are also analogous to those of Section 2 and we shall not discuss them here. Instead we consider several important particular cases of the procedures introduced in Sections 2, 3 and 4.

5. Example 1. We start with the well-known Weber's problem of finding

$$\min_{x\in R^n} \sum_{i=1}^{m} w_i |x - a_i| , \tag{42}$$

where w_i are positive constants, the a_i, $i = 1,\ldots,m$, are fixed points in R^n, $|x - a_i| = [\sum_{i=1}^{n} (x_j - a_{ij})^2]^{1/2}$. Weiszfeld in [10] suggested the following procedure to determine the minimizing sequence for the problem (42):

x^0 is an arbitrary point;

for $k = 0,1,\ldots$

(43)
$$x^{k+1} = \sum_{i=1}^{m} \frac{w_i a_i}{|x^k - a_i|} \Big/ \sum_{i=1}^{m} \frac{w_i}{|x^k - a_i|} .$$

Weiszfeld suggested a proof of the convergence of the procedure (43) which had several drawbacks, see [6], related to the behavior of the procedure if one of the points x^k coincides with a_i. Since that time a number of papers have appeared where the procedure (43) and its modifications were considered. We mention only [4], where the proof of the linear rate of convergence for (43) was established if the solution of (42) $x^* \neq a_i$, $i = 1,\ldots,m$, and if the initial guess is close to x^*, [6], where the procedure (43) is modified so that the proof of convergence can be carried out even if $x^k = a_i$; in [6] the convergence was proved for almost all initial points. We refer the reader to [9] for generalizations and bibliography.

The method of nonlinear transformation of the objective function allows us to suggest a modification of Weiszfeld's algorithm which will converge to x^* from arbitrary initial point. First we observe that the original procedure (43) is simply the procedure (3) with $f_i(x) = w_i|x - a_i|$, $g_i(\alpha) = \alpha^2$, $i = 1,\ldots,m$. However we cannot apply the convergence results of Section 2 because the last condition in (2) does not hold if $x = a_i$. Thus, we have to consider the modified procedure (38) with $g_i(\alpha) = \alpha^2$. We obtain the following procedure:

x^0 is an arbitrary point;

for $k = 0,1,\ldots$

(44)
$$\alpha_i^k = \max\{w_i|x^k - a_i|, \delta_i^k\} ,$$

$$x^{k+1} \in \operatorname{Arg}\min_{x \in \mathbb{R}^n} \sum_{i=1}^{m} \frac{w_i^2 (x - a_i)^2}{2\alpha_i^k} = \sum_{i=1}^{m} \frac{w_i^2 a_i}{\alpha_i^k} \Big/ \sum_{i=1}^{m} \frac{w_i}{\alpha_i^k} .$$

As was shown in [5] we can take

$$\delta_i^k = \delta^k = \frac{1}{k}, \quad k = 1,2,\ldots; \quad \delta^0 = 1 .$$

It is obvious that if the point x^k lies outside a small neighborhood of the points a_i, $i = 1,\dots,m$, then the iteration (44) coincides with the iteration of Weiszfeld's algorithm (43). From the results of Section 3 the global convergence of the procedure (44) to the solution x^* of (42) follows. As for the rate of this convergence, if $x^* \neq a_i$, $i = 1,\dots,m$, then by virtue of the above-said and by Lemma 3, part B), the sequence $\{x^k\}$ converges to x^* with the linear rate of convergence. Thus we have reproduced the result of [4]. Moreover, if $x^* = a_\ell$ then also by Lemma 3, part C) starting from some iteration n_0 we have the quadratic convergence until we approach very closely to x^*, so that $\alpha_i^k = \delta^k$ in (44) for some i. Then the convergence slows down and, as can be shown, $\mu^k = O\left(\frac{1}{k}\right)$. Note that procedure (44) allows us to avoid difficulties with x^k being equal to a_i. Of course, the procedure (43) (or (44)) is neither the only nor the best procedure to solve the problem (42), for example, see [7]. However we have considered it because it allows us to demonstrate the generality of our method. The next example is considered for the same purpose.

Example 2. In this example we consider the method of steepest descent and Newton's method. We show that both of these methods are particular cases of the procedure (40). Therefore the results of Section 4 enable us to obtain some sufficient conditions for the global convergence of these methods. We first consider the problem of finding

$$\min_{x \in \mathbb{R}^n} \Phi(x) \ , \tag{45}$$

where $\Phi(x) \in C^1(\mathbb{R}^n)$ and is convex and non-negative. Consider $g_1^k(x,\alpha) = \alpha - \Phi(x) + h^k x^2$ where h^k is a positive constant which will be determined later. In this case the procedure (40) has the form

$$\begin{aligned} &x^0 \text{ is an arbitrary point} \\ &x^{k+1} \in \operatorname{Arg}\min_{x \in \mathbb{R}^n} \{h^k x^2 - (2h^k x^k - \nabla\Phi(x^k))x\} \end{aligned} \tag{46}$$

or

$$x^{k+1} = x^k - \frac{1}{h^k} \nabla\Phi(x^k) .$$

This is an iteration of the steepest descent method. By the results of Section 4, in order for the procedure (46) to converge it is sufficient that the coefficients h^k are such that the functions $g_1^k(x,\alpha)$ are convex and $|h^k| < C_0$, $k = 0,1,\ldots,$ for $\alpha \in R_+$ and for $x \in M^k$, where $M^k = \{x \in R^n | \Phi(x) < \Phi(x^k)\}$.

Now again we consider the problem (45) and assume that $\Phi(x) \in C^2(R^n)$. We take $g_1^k(x,\alpha) = \alpha - \Phi(x) + \langle H^k x,x\rangle$, where H^k is some symmetric positive definite matrix. Then the procedure (40) will have the form

(46)
$$\begin{aligned} &x^0 \text{ is an arbitrary point ,} \\ &x^{k+1} = x^k - (H^k)^{-1}\nabla\Phi(x^k) . \end{aligned}$$

From the results of Section 4 it follows that the procedure (46) will converge to the solution of (45) if $\|H^k\| < C_0$ and $g_1^k(x,\alpha)$ are convex for $k = 0,1,\ldots,$ $\alpha > 0$, $x \in M^k = \{x|\Phi(x) < \Phi(x^k)\}$. Of course if H^k is $H(x^k)$, where $H(x)$ is the Hessian matrix of the function $\Phi(x)$, (46) is an iteration of Newton's method. Note that the fixed point method is also a particular case of the procedure (40).

6. Finally we consider a version of the method of nonlinear transformation of the objective function without which our presentation of that method undoubtedly would not be complete. In this section we consider the case when the transforming functions $g_i(\alpha)$ are concave. We assume that

(47)
- D is a closed bounded convex set in R^n;
- the functions $g_i(\alpha) \in C^2[0,\infty)$ and are convex;
- $g_i'(f_i(x)) > \delta > 0$, $|g_i''(f_i(x))| < \gamma$, $x \in D$, $i = 1,\ldots,m$;
- the functions $g_i(f_i(x))$ are convex for $x \in D$;

where δ and γ are some positive constants. Note that these conditions will allow us to consider in our method such functions $g_i(\alpha)$ as $\ln \alpha$, α^p, $0 < p < 1$, etc. As usual we assume that the functions $g_i(f_i(x))$, $i = 1,\dots,m$, are simple (for example, quadratic) and that the linear combination $\sum_{i=1}^{m} c_i g_i(f_i(x))$ can be minimized on D with less effort than $\Phi(x)$ for every set of positive coefficients $c_1,\dots,c_m$. Consider now the following procedure:

$x^0 \in D$ is an arbitrary point; for $k = 0,1,\dots$

$$\xi^k = \min \sum_{i=1}^{m} g_i(f_i(x))/g_i'(f_i(x^k)) ,$$

$$M_i^k = \min_{\beta \in [f_i(x^k), f_i(\xi^k)]} |g_i''(\beta)|/g_i'(f_i(x^k)), \quad i = 1,\dots,m ,$$

$$J^k = \left\{ j \in \{1,\dots,m\} \,\middle|\, d_j^k = \frac{d^+}{d\tau} g_j(f_j(x^k + \tau(\xi^k - x^k)))\Big|_{\tau=0} < 0 \right\} \tag{48}$$

$$\sigma^k = \sum_{i \in J^k} M_i^k d_i^k (\xi^k - x^k)^2 + \sum_{i \notin J^k} M_i^k (f_i(\xi^k) - f_i(x^k))^2$$

$$\alpha^k = \min\Big\{1, \frac{1}{\sigma^k} \sum_{i=1}^{m} [g_i(f_i(x^k)) - g_i(f_i(\xi^k))]/g_i'(f_i(x^k)),$$

$$x^{k+1} = x^k + \alpha^k(\xi^k - x^k) .$$

<u>Theorem 7</u>. Assume that the conditions (47) hold and that the functions $f_i(x)$, $i = 1,\dots,m$, are non-negative and convex in D. Then

$$\Phi(x^k) \downarrow \Phi(x^*) \quad \text{as} \quad k \to \infty ,$$

where $x^* \in \operatorname{Arg}\min_{x \in D} \Phi(x)$.

<u>Proof</u>. We introduce the function

$$G_k(x) = \sum_{i=1}^{m} g_i(f_i(x))/g_i'(f_i(x^k)) .$$

Since for every $k = 0,1,\dots$ the point ξ^k is a point of minimum of the function $G_k(x)$,

$$G_k(\xi^k) - G_k(x^k) \le G_k(x^*) - G_k(x^k)$$

$$= \sum_{i=1}^{m} [g_i(f_i(x^*)) - g_i(f_i(x^k))]/g_i'(f_i(x^k)) .$$

By concavity of $g_i(\alpha)$, $i = 1,\ldots,m$,

$$\sum_{i=1}^{m} g_i(f_i(x^*)) - g_i(f_i(x^k))/g_i'(f_i(x^k))$$

$$\le \sum_{i=1}^{m} [f_i(x^*) - f_i(x^k)] = \Phi(x^*) - \Phi(x^k).$$

Thus

$$G_k(\xi^k) - G_k(x^k) \le \Phi(x^*) - \Phi(x^k) . \tag{49}$$

On the other hand, by convexity of the functions $f_i(x)$, $i = 1,\ldots,m$,

$$\begin{aligned} G_k(x^{k+1}) - G_k(x^k) &= \Phi(x^{k+1}) - \Phi(x^k) \\ &+ \sum_{i=1}^{m} [g_i(f_i(x^k + \alpha^k(\xi^k - x^k))) - g_i(f_i(x^k)) \\ &- g_i'(f_i(x^k))(f_i(x^{k+1}) - f_i(x^k))]/g_i'(f_i(x^k)) \\ &\ge \Phi(x^{k+1}) - \Phi(x^k) - \frac{1}{2}\sum_{i=1}^{m} M_i^k[f_i(x^k + \alpha^k(\xi^k - x^k)) \qquad (50) \\ &- f_i(x^k)]^2 \ge \Phi(x^{k+1}) - \Phi(x^k) \\ &- \frac{1}{2}(\alpha^k)^2\{\sum_{i\in J^k} M_i^k d_i^k(\xi^k - x^k)^2 + \sum_{i\notin J^k} M_i^k(f_i(\xi^k) - f_i(x^k))^2\} , \end{aligned}$$

where the set J^k and constants M_i^k, d_i^k and α^k are defined in (48). Recalling the definition of σ^k in (48) we obtain from (50) the following inequality:

$$\Phi(x^{k+1}) - \Phi(x^k) \le G_k(x^{k+1}) - G_k(x^k) + \frac{1}{2}\sigma^k(\alpha^k)^2, \quad k = 0,1,\ldots .$$

The function $G_k(x)$ is convex for every k. Therefore, since $x^{k+1} = x^k + \alpha^k(\xi^k - x^k)$,

$$(51) \qquad \Phi(x^{k+1}) - \Phi(x^k) < \alpha^k(G_k(\xi^k) - G_k(x^k)) + \frac{1}{2}\sigma^k(\alpha^k)^2, \quad k = 0,1,\ldots .$$

By definition $\alpha^k = \min\{1,(G_k(x^k) - G_k(\xi^k))/\sigma^k\}$, and is non-negative since $G(\xi^k) < G_k(x^k)$. If $(G_k(x^k) - G_k(\xi^k))/\sigma^k < 1$ we obtain from (51) that

$$(52) \qquad \Phi(x^{k+1}) - \Phi(x^k) < -[G_k(x^k) - G_k(\xi^k)]^2/(2\sigma^k)$$

If $(G_k(x^k) - G_k(\xi^k))/\sigma^k > 1$ then it follows from (51) that

$$(53) \qquad \Phi(x^{k+1}) - \Phi(x^k) < -\frac{1}{2}(G_k(x^k) - G_k(\xi^k)) .$$

Combining (52), (53) with (49) we obtain that

$$(54) \qquad \begin{aligned} &\mu^{k+1} < \mu^k - \frac{1}{2\sigma^k}(\mu^k)^2 \quad \text{if} \quad (G_k(x^k) - G_k(\xi^k))/\sigma^k < 1 , \\ &\mu^{k+1} < \frac{1}{2}\mu^k \quad \text{if} \quad (G_k(x^k) - G_k(\xi^k))/\sigma^k > 1 , \end{aligned}$$

$k = 0,1,\ldots,$ $\mu^k = \Phi(x^k) - \Phi(x^*)$. Since D is bounded, $\sigma^k < T$, $k = 0,1,\ldots,$ where T is a positive constant. Moreover, from (54) it follows that $\mu^k < \mu^0$ for $k = 0,1,\ldots$. Therefore from the second inequality in (54) it follows that

$$(55) \qquad \mu^{k+1} < \mu^k - \frac{1}{2\mu^0}(\mu^k)^2 \quad \text{if} \quad (G_k(x^k) - G_k(\xi^k))/\sigma^k > 1 .$$

From (54) and (55) we obtain that

$$(56) \qquad \mu^{k+1} < \mu^k - \frac{1}{2T_1}(\mu^k)^2, \quad k = 0,1,\ldots ,$$

where $T_1 = \max\{\mu^0,T\}$. The inequality (56) is the same as in Section 2. It follows from (56) that $\mu^k \to 0$ as $k \to \infty$. The theorem is proved.

An interesting application of the procedure (48), the geometric programming problem, will be considered in the forthcoming paper.

REFERENCES

[1] Eydeland, A., "A Method of Solving Nonlinear Variational Problems by Nonlinear Transformation of the Objective Functional, Part I", Numer. Math. 43 (1984), 59-82.

[2] Eydeland, A., "A Method of Solving Nonlinear Variational Problems by Nonlinear Transformation of the Objective Functional, Part II", Numer. Math., to appear.

[3] Eydeland, A., "Method of Transformation of the Objective Function", In: Problems of math. simulation, Radio and Electronics Inst. of Acad. of Sci. of the USSR (1979), 79-86 (in Russian).

[4] Katz, N., "On the Convergence of a Numerical Scheme for Solving Some Locational Equilibrium Problems", SIAM J. Appl. Math. 17, No. 6 (1969), 1224-1231.

[5] Korobochkin, B. and A. Eydeland, "Application of the objective function transformation method to one class of optimization problems", In: Math. methods of solving econom. problems, Nauka, Moscow (1980), 72-84 (in Russian).

[6] Kuhn, H. W., "A Note on Fermat Problem", Math. Progr. 4 (1973), 98-107.

[7] Overton, M., "A Quadratically Convergent Method for Minimizing a Sum of Euclidean Norms", Techn. Report No. 30, Courant Inst. of Math. Sci., New York (1981).

[8] Rockafellar, R. T., "Convex Analysis", Princeton Univ. Press, Princeton, N.J. (1972).

[9] Voss, H. and U. Eckhardt, "Linear Convergence of Generalized Weiszfeld's Method", Computing 25 (1980), 243-257.

[10] Weiszfeld, E., "Sur le point par lequel la somme des distances de n points donnés est minimum", Tohuku Math. Journal 43 (1937), 355-386.

AB/ed

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: 2699

2. GOVT ACCESSION NO.: AD-A142 905

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): GLOBALLY CONVERGENT PROCEDURES FOR SOLVING NONLINEAR MINIMIZATION PROBLEMS

5. TYPE OF REPORT & PERIOD COVERED: Summary Report - no specific reporting period

6. PERFORMING ORG. REPORT NUMBER:

7. AUTHOR(s): Alexander Eydeland

8. CONTRACT OR GRANT NUMBER(s): MCS-8301628, DAAG29-80-C-0041, MCS-8210950

9. PERFORMING ORGANIZATION NAME AND ADDRESS: Mathematics Research Center, University of Wisconsin, 610 Walnut Street, Madison, Wisconsin 53706

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: Work Unit Number 5 - Optimization and Large Scale Systems

11. CONTROLLING OFFICE NAME AND ADDRESS: (See Item 18 below)

12. REPORT DATE: May 1984

13. NUMBER OF PAGES: 25

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

U. S. Army Research Office
P. O. Box 12211
Research Triangle Park
North Carolina 27709

National Science Foundation
Washington, DC 20550

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Nonlinear minimization problem, globally convergent minimization procedure, transformation of the objective function

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In this work we describe globally convergent iterative procedures for generating a minimizing sequence for the problem of finding the infimum for the function $\Phi(x) = \sum f_i(x)$ on a certain set D. The minimizing sequence consists of points of the infimum for functions of the type

(cont.)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT (cont.)

$\sum [\beta_i(g_i(x,f_i(x)) + \langle\gamma_i,x\rangle]$, where the transforming functions $g_i(x,\cdot)$ are chosen in such a way that the compositions $g_i(x,f_i(x))$ are simpler than the given functions $f_i(x)$, and where the coefficients $\beta_i > 0$, $\gamma_i \in R^n$, $i = 1,\ldots,m$, are determined by the choice of g_i. Various classes of the functions g_i are considered and global convergence results are proved. It is shown that many well-known algorithms, for example, Weiszfeld's algorithm and Newton's method are particular cases of the general method.

END

FILMED

8 - 84

DTIC