A GLOBALLY CONVERGENT QUADRATIC APPROXIMATION FOR INEQUALITY CONSTRAINED MINIMAX PROBLEMS

K. Kiwiel

March 1981
CP-81-9

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

A GLOBALLY CONVERGENT QUADRATIC APPROXIMATION FOR INEQUALITY CONSTRAINED MINIMAX PROBLEMS

K. Kiwiel

1. INTRODUCTION

In this paper we present an implementable algorithm for solving optimization problems of the following type: minimize $f_0(x)$, subject to $f(x) \leqslant 0$, where $x \in R^N$ and f_0 and f are real-valued functions that are the pointwise maxima of two families of continuously differentiable functions.

Our algorithm combines, extends and modifies ideas contained in Wierzbicki [1978], Pschenichnyi and Danilin [1975] and Huard [1968]. Its derivation is based on the application of quadratic approximation methods to the improvement function used in the modified method of centers. In fact, when an initial approximation to the solution is feasible, the algorithm works as a feasible direction method [Polak 1971] and the objective function f_0 need not be evaluated at infeasible points. When the initial approximation is infeasible, the algorithm decreases constraint violation at each iteration and its accumulation points are stationary [Demianov, Malozemov 1972] if some regularity assumption on the gradients of constraints functions outside the feasible set holds. However, we do not require that the optimization problem be normal [Clarke 1976], which is necessary for convergence of quadratic approximation methods using exact penalty functions [Han 1977; Conn, Pietnykowski 1977]. The algorithm may use variable metric techniques to speed up convergence as in [Wierzbicki 1978]; this point is a subject of on-going research.

A further modification of the ideas presented in this paper has lead to a new implementable algorithm [Kiwiel forthcoming] for solving problems of a more general nature, when f_0 and f are semismooth [Mifflin 1979].

Our algorithm has search direction finding subproblems that are quadratic programming problems involving convex combinations of problem function gradients with a linear form in the subproblem objective related to complimentary slackness [Wierzbicki 1978]. These subproblems are discussed in para. 3. The algorithm is defined in para. 4 and in para. 5 we show stationarity of the algorithms accumulation points. In para. 6 we present conditions under which the algorithm converges linearly.

The scalar product of $x = (x_1,...,x_N)$ and $y = (y_1,...,y_N)$ in R^N, defined by $\sum_{i=1}^{N} x_i y_i$ is denoted $\langle x,y\rangle$ and the Euclidean norm of x, defined by $|\langle x,x\rangle^{\frac{1}{2}}|$ is denoted $|x|$. If A is an $N \times N$ symmetric positive definite matrix, $\langle Ax,y\rangle$ is denoted $\langle x,y\rangle_A$ and $|x|_A^2$ denotes $\langle Ax,x\rangle$. If $h = R^N \rightarrow R^1$ is twicely continuously differentiable, $h'(x)$ denotes its gradient at $x \in R^N$, and $h''(x)$ its hessian.

2. PROBLEM STATEMENT

Consider the following optimization problem:

(2.1) $$\min f_0(x) \quad \text{s.t. } f(x) \leqslant 0 ,$$

where

(2.2) $$f_0(x) = \max_{i=1..,n} f_{0,i}(x) , \quad f(x) = \max_{i=1,m} f_i(x) , \quad i = 1,m \text{ or } i = 1,\ldots,m$$

and $f_{0,i} : R^N \to R \quad i = 1,\ldots,n$, $f_i : R^N \to R \quad i = 1,\ldots,m$ are continously differentiable; $n,m < +\infty$.

The necessary conditions of optimality for some $\hat{x}$ to be a solution of (2.1) are as follows [Clarke 1976]: there exists a collection of numbers $\{\lambda_i\}_{i=1}^{n+m}$ satisfying:

(2.3) $$\lambda_i \leqslant 0 \quad i = 1; n + m \quad \sum_{i=1}^{m+n} \hat{\lambda}_i = 1 ,$$

(2.4) $$\sum_{i=1}^{n} \hat{\lambda}_i f_{0,i}(\hat{x}) + \sum_{i=1}^{m} \hat{\lambda}_{i+n} f_i'(\hat{x}) = 0 ,$$

(2.5) $$\hat{\lambda}_i [f_0, i(\hat{x}) - f_0(\hat{x})] = 0 \quad i = 1;n ,$$

(2.6) $$\hat{\lambda}_{i+n} [f_i(\hat{x}) - f^+(\hat{x})] = 0 \quad i = 1;m .$$

Note that $f^+(\hat{x}) = 0$ since $\hat{x}$ is feasible. Consider also an auxilliary problem:

(2.7) $$\min f(x) .$$

If $\bar{x}$ is its solution, then there exist numbers $\{\bar{\lambda}_i\}_{i=1}^{m}$ satisfying:

(2.8) $\bar{\lambda}_i \geq 0 \quad i = 1;m, \quad \sum_{i=1}^{m} \bar{\lambda}_i = 1$,

(2.9) $\sum_{i=1}^{m} \bar{\lambda}_i f_i'(\bar{x}) = 0$, $\sum \bar{\lambda}_i f_i'(\bar{x}) = 0$

(2.10) $\bar{\lambda}_i [f_i(\bar{x}) - f(x)] = 0$.

3. DIRECTION FINDING PROBLEM

The algorithm presented in the next section uses search directions generated as follows. Let $x \in R^N$ and $\delta > 0$ be given. Introduce two activity sets:

(3.1) $I_0(x,\delta) = \{i : f_{0,i}(x) - f_0(x) \geq f^+(x) - \delta, \ 1 \leq i \leq n\}$,

$I_c(x,\delta) = \{i : f_i(x) \geq f^+(x) - \delta, \ 1 \leq i \leq m\}$.

Let A be an $N \times N$ symmetric positive definite matrix. Then the following problem with respect to variables $\beta \in R^1$ and $p \in R^N$:

(3.2) $\min \{\beta + \frac{1}{2}|p|_A^2\}$

$f_{0,i}(x) - f_0(x) - f^+(x) + \langle f_{0,1}(x), p \rangle \leq \beta \quad i \in I_0(x,\delta)$

$f_i(x) - f^+(x) + \langle f_i(x), p \rangle \leq \beta \quad i \in I_c(x,\delta)$

satisfies Slater's condition [Pschenicnyi and Danilin 1975:259] its solutions $\beta(x)$ and $p(x)$ exist and are uniquely determined by the following set of conditions:

(3.3) $p(x) = -A^{-1}(\sum_{i \in I_0(x,\delta)} \lambda_i f_{0,i}(x) + \sum_{i \in I_c(x,\delta)} \lambda_{i+n} f_i(x))$

$$(3.4)\qquad -\beta(x) = |p(x)|_A^2 + \sum_{i\in I_0(x,\delta)} \lambda_i[f_0(x) - f_{0,i}(x) + f^+(x)] + \sum_{i\in I_c(x,\delta)} \lambda_{i+n}[f^+(x) - f_i(x)]$$

where $\{\lambda_i\}$ satisfy:

$$(3.5)\qquad \lambda_i \geqslant 0 \quad i\in I_0(x,\delta) \quad \lambda_{i+n} \geqslant 0 \quad i\in I_c(x,\delta) \quad ,$$

$$(3.6)\qquad \sum_{i\in I_0(x,\delta)} \lambda_i + \sum_{i\in I_c(x,\delta)} \lambda_{i+n} = 1 \quad ,$$

$$(3.7)\qquad \lambda_i[f_{0,i}(x) - f_0(x) - f^+(x) + \langle f_{0,i}(x), p(x)\rangle - \beta(x)] = 0 \qquad i\in I_0(x,\delta) \quad ,$$

$$(3.8)\qquad \lambda_{i+n}[f_i(x) - f^+(x) + \langle f_i'(x), p(x)\rangle - \beta(x)] = 0 \qquad i\in I_c(x,\delta) \quad .$$

Note that when $I_0(x,\delta)$ is empty, the direction $p(x)$ is computed as in Pshenichnyi's method of linearization for solving the problem (2.7) if $A = I$ is used [Pschenicnyi and Danilin 1975]; when A approximates the Hessian of the Lagrange function for (2.7), $p(x)$ is equal to the direction obtained in the quadratic approximation method for (2.7) [Wierzbicki 1978]. In general, (3.2) may be viewed as a quadratic approximation problem for the function

$$(3.9)\qquad \zeta(x') = \max\,\{f_0(x') - f_0(x), f(x)\} \quad .$$

4. ALGORITHM

Step 0. Choose a starting point $x^0 \in R^N$, an $N \times N$ symmetric positive definite matrix A_0 (e.g., $A_0 = I$), a final accuracy parameter ε_f, an activity bound $\underline{\delta} > 0$, a desired rate of convergence parameter $\gamma \in [0,1)$, line search parameters $\varepsilon_i \in (0,1)$ and $0 < m_1 < m_2 < 1$. Choose initial values of a convergence variable $\eta^0 \geqslant \underline{\delta}$.

Set $k = 0$

Step 1. Compute $p^k = p(x^k)$ and $\beta^k = \beta(x^k)$ solving (3.2) with $A = A_k$ and $\delta = \delta^k$.

Step 2. If $\beta^k \geqslant -\varepsilon_f$, stop.

Step 3. Let an improvement function be given by

$$\phi_k(x) = \max \{f_0(x) - f_0(x^k), f(x)\} \quad . \tag{3.1}$$

If $\phi_k(x^k + p^k) < \phi_k(x^k)$ and $\beta^k \geqslant \gamma\eta^k$, set $\alpha^k = 1$ and go to Step 5. (Direct prediction).

Step 4. Compute a step-size coefficient $\alpha^k > 0$ satisfying one of the following conditions:

Step 4i. $\alpha^k = 2^{-i_k}$ where i_k is the first number $i = 0,1...$ for which:

$$\phi_k(x^k + 2^{-1_k}p) \leqslant \phi_k(x^k) + \varepsilon_i 2^{-i_k}\beta \quad .$$

(Armigo's rule).

Step 4ii. $\phi_k(x^k) + m_2\alpha^k\beta^k \leqslant \phi_k(x^k + \alpha^k p^k) \leqslant \phi_k(x^k) + m_1\alpha^k\beta^k$, (Goldstein's rule). The line search of [Wierzbicki 1978] is recommentded for the exception of this step.

Step 4iii. $\phi_k(x^k + \alpha^k p^k) \leqslant \phi_k(x^h + \tilde{\alpha}^k p^k)$ for some $\tilde{\alpha}^k > 0$ satisfying either of the above requirements (approximate or exact minimization).

Step 5. Set $x^{k+1} = x^k + \alpha^k p^k$, choose new symmetric positive definite A_{k+1} and $\delta^{k+1} \geqslant \underline{\delta}$, set $\eta^{k+1} = \max \{\eta^k, \beta^k\}$. Replace k by k + 1 and go to Step 1.

A few comments on the implementation of the algorithm are presented below. In order to compute $p(x)$ and $\beta(x)$ it is more efficient to solve the dual of (3.2), viz.

$$(4.2)\quad \begin{cases} \min \{\tfrac{1}{2}\left|\sum_{i\in I_0(x,\delta)} \lambda_i f'_{0,i}(x) + \sum_{i\in I_c(x,\delta)} \lambda_{i+n} f'_i(x)\right|^2_{A^{-1}} + \\ \qquad \sum_{i\in I_0(x,\delta)} \lambda_i [f_0(x) - f_{0,i}(x)] + \sum_{i\in I_c(x,\delta)} \lambda_{i+n}[f^+(x) - f_i(x)] \\ \lambda_i \geqslant 0 \; i\in I_0(x,\delta) \quad \lambda_{i+n} \geqslant 0 \; i\in I_c(x,\delta) \; , \\ \sum_{i\in I_0(x,\delta)} \lambda_i + \sum_{i\in I_c(x,\delta)} \lambda_{i+n} = 1 \end{cases}$$

If $\{\lambda_i\}$ solves (4.2), let

$$(4.3)\quad d(x) = -\Big(\sum_{i\in I_0(x,\delta)} \lambda_i f'_{0,i}(x) + \sum_{i\in I_c(x,\delta)} \lambda_{i+n} f'_i(x)\Big) \quad .$$

Then $p(x) = \bar{A}^1 d(x)$ by (3.3) and $\beta(x)$ is determined by (3.4) with $|p(x)|^2_A = |d(x)|^2_{A-1}$. Thus we see that it may be easier to work with $H = A^{-1}$ rather than with A.

In this paper we do not consider the important questions of the choice of $\{A_k\}$ (or $\{H_k\}$). Our global convergence analysis requires this sequence to be uniformly positive definite and bounded. However, in order to obtain fast local convergence results we conjecture, by analogy to [Wierzbicki 1978], that A^k should approximate the Hessian of the Lagrange function for (2.1). Therefore, some quasi-Newton updating formula [Han 1977] could be used, based on data

(3.5) $$s^{k+1} = x^{k+1} - x^k \quad ,$$

(3.6) $$r^{k+1} = \sum_{i\in I_0} \lambda_i^{k+1} [f_{0,i}'(x^{k+1}) - f_{0,i}'(x^k)] + \sum_{i\in I_c} \lambda_i^{k+1} [f_i'(x^{k+1}) - f_i'(x^k)] \quad ,$$

where $\{\lambda_i^{k+1}\}$ denotes the solution of (4.2) with $x = x^k$. We leave that question open for future research.

The value of δ^k controls the size of the direction finding problem (3.2) and $\underline{\delta}$ establishes a threshold for determining the functions probably active at the solution. Note that if x^0 is infeasible, i.e., $f(x^0) = f^+(x^0) > 0$, the algorithm reduces to the quadratic approximation method for minimizing the constraint violation $f(x)$ until $f(x^k) \leqslant \delta^k$, since $I_o(x^i,\delta^i) = \emptyset$ for $i = 0,1,\ldots,k-1$. This suggests the following strategy for changing δ:

(3.7) $$\delta^{k+1} = \max \{\underline{\delta}, \zeta\sqrt{-\eta^{k+1}}\} \quad ,$$

with ζ being a scaling parameter.

The existence of a finite i_k in Step 4i follows from the results of the next section. Under an additional assumption that the function $\psi_k(\alpha) = \phi_k(x^k + \alpha p^k)$ is bounded from below for $\alpha > 0$, finite termination of the line search of [Wierzbicki 1978] (which is based on geometric expansion, contraction and bisection) may be easily proved, thus providing a method for Step 4ii.

A nice feature of the algorithm is that it decreases constraint violation at each iteration. To see this, note that due to the line search rules

$$(4.8) \qquad F(x^{k+1}) \leqslant \zeta_k(x^{k+1}) < \zeta_k(x^k) = f^+(x^k) \quad ,$$

Since $\beta^k < 0$ at Step 4 owing to (3.4). Observe that if some x^k is feasible, $f^+(x^k) = 0$ and (4.8) imply that all consecutive points are feasible.

5. CONVERGENCE

In this section we analyze convergence of the proposed algorithm. Since we do not assume that the initial x^0 is feasible, it is not suprising that we have to impose additional assumptions on the gradients of the constraint functions outside the feasible set. Namely, consider the following assumption:

(A1) If $x \in R^N$ is such that $f(x^0) \geqslant f(x) > 0$, then x is not stationary for (2.7), i.e., there are no $\{\bar{\lambda}_i\}_1^m$ satisfying (2.8) - (2.10) with $\bar{x} = x$.

We think that (A1) is a natural requirement for the problem computing a feasible point to be well-posed. Note, however, that we do not assume that the original problem (2.1) is normal.

Naturally, convergence results assume $\varepsilon_f = 0$. We first consider the case when the algorithm terminates.

Proposition 5.1

If the algorithm stops at iteration k, then x^k is either feasible and stationary for (2.1) or infeasible and stationary for (2.7). If (A1) holds, then x^k is feasible.

Proof. Since $0 = \varepsilon_f \leqslant \beta^k \leqslant 0$, (3.4) and (3.3) imply that $p^k = 0$, hence $d^k = A_k p^k = 0$. Since $f_{0,i}(x^k) \leq f_0(x^k)$, $f_i(x^k) \leq f^+(x^k)$, (3.4) implies

$$(5.1) \qquad \lambda_i^{k+1}[f_0(x^k) - f_{0,i}(x^k) + f^+(x^k)] = 0 \quad i \in I_0(x^k,\delta^k) \quad ,$$

$$(5.2) \qquad \lambda_{i+n}^{k+1}[f^+(x^k) - f_i(x^k)] = 0 \qquad i \in I_c(x^k,\delta^k) \quad .$$

Now, if x^k is feasible, $f^+(x^k) = 0$. If $f^+(x^k) > 0$, then (5.1) implies that $\lambda_i^{k+1} = 0$ $i \in I_0(x^k,\delta^k)$. Noting that

$$0 = d^k = \sum_{i \in I_0(x^k,\delta^k)} \lambda_i^{k+1} f_{0,i}'(x^k) + \sum_{i \in I_c(x^u,\delta^k)} \lambda_{i+n}^{k+1} f_i'(x^k) \quad ,$$

we see that $\{\lambda_i^{k+1}\}$ satisfy either (2.3) - (2.6) or (2.8) - (2.10), which ends the proof.

From now on we assume that the algorithm does not stop. We shall also assume that $\{A_k\}$ are uniformly positive definite and bounded, i.e., that there exist two constants γ_1 and γ_2, $0 < \gamma_1 \leq \gamma_2$:

$$(5.3) \qquad \gamma_1 |x|^2 \leq |x|^2_{A_k} \leq \gamma_2 |x|^2 \text{ for all } k \text{ and all } x \in R^N \quad .$$

THEOREM 5.2

Every accumulation point of $\{x^k\}$ is either feasible and stationary for (2.1), or infeasible and stationary for (2.7). If (A1) holds, then any accumulation point of $\{x^k\}$ is feasible. In particular, if x^0 is feasible, then every accumulation point is feasible.

Proof: Let $\bar{x}$ be some accumulation point of $\{x^k\}$, i.e., $x^k \to \bar{x}$ $k \in k_1$. Since $\beta^k < 0$ by the rules of the algorithm, we shall consider two cases, depending on whether lim sup $\{\beta^k; k \in k_1\}$ equals zero or not.

A) Suppose that $\beta^k \to 0$ for $k \in K_2 \subset K_1$. Then $|p^k|^2_{A_k} \to 0$ for $k \in K_2$ by (3.4) hence $p^k \to 0$ $k \in K_2$, since $|p^k|^2 \leq \frac{1}{\gamma_1} |p^k|^2_{A_k}$ by (5.3). As $d^k = A^k p^k$, $|p^k|^2_{A_k 1} \geq \frac{1}{\gamma_2} |d^k|^2$ by (5.3), hence $d^k \to 0$ $k \in K_2$. As $\{\lambda_i^k\}$ satisfy (3.5) and (3.6), we may introduce additional $\lambda_i^k = 0$ to get

$$(5.4) \qquad -d^k = \sum_{i=1}^{n} \lambda_i^{k+1} f'_{0,i}(x^k) + \sum_{i=1}^{m} \lambda_{i+n}^{k+1} f'_i(x^k) \quad ,$$

$$(5.5) \qquad \lambda_i^{k+1} \geq 0 \quad i = 1; m+n \quad \sum_{i=1}^{m+n} \lambda_i^{k+1} = 1 \quad ,$$

$$(5.6) \qquad \lambda_i^{k+1} [f_{0,i}(x^k) - f_0(x^k) - f^+(x^k) + \langle f'_{0,i}(x^k), p^k \rangle - \beta^k] = 0$$
$$i = 1; n \quad .$$

$$(5.7) \qquad \lambda_{i+n}^{k+1} [f_i(x^k) - f^+(x^k) + \langle f'_i(x^k), p^k \rangle - \beta^k] = 0$$
$$i = 1; m \quad .$$

Using (5.5) and passing to further subsequences, if necessary, we may write that $\lambda_i^{k+1} \to \bar{\lambda}_i$ $k \in K_2$ with

$$(5.8) \qquad \bar{\lambda}_i \geq 0 \quad i = 1; m+n \text{ and } \sum_{i=1}^{m+n} \bar{\lambda}_i = 1 \quad .$$

Since $f_{0,i}$, f_i, are continuously differentiable, we may pass to to the limit in (5.4) through (5.7) with $k \in K_2$ and get

$$(5.9) \qquad \sum_{i=1}^{n} \bar{\lambda}_i f_{0,i}(\bar{x}) + \sum_{i=1}^{m} \bar{\lambda}_{i+n} f_i'(\bar{x}) = 0 \quad ,$$

$$(5.10) \qquad \bar{\lambda}_i [f_{0,i}(\bar{x}) - f_0(\bar{x}) - f^+(\bar{x})] = 0 \qquad i = 1;n \quad ,$$

$$(5.11) \qquad \bar{\lambda}_{i+n} [f_i(\bar{x}) - f^+(\bar{x})] = 0 \qquad i = 1;m \quad .$$

Next the argument proceeds exactly as in the proof of Proposition 1.

B) Now suppose that there exists $\underline{\beta}(\bar{x}) < 0$ such that $\beta^k \leqslant \underline{\beta}(\bar{x})$, $k \in k_1$. By (5.4),(5.5) and the coninuity of problem function gradients, d^k are uniformly bounded for $k \in K_1$, and thus $p^k, k \in K_1$, are uniformly bounded by (5.3). Combining this with the continuous differentiability of the problem functions [5, Appendix III, para. 3, Note 2], we infer that for almost all $k \in K_1$ the following expansions are valid:

$$(5.12) \qquad f_{0,i}(x^k + \alpha p^k) = f_{0,i}(x^k) + \alpha \langle f_{o,i}(x^k), p^k \rangle + o(\alpha)$$

$$i = 1;n \quad ,$$

$$f_i(x^k + \alpha p^k) = f_i(x^k) + \alpha \langle f_i(x^k), p^k \rangle + o(\alpha)$$

$$i = 1;m \quad ,$$

where $o(\alpha)/\alpha \to 0$ when $\alpha \downarrow 0$.

By (5.12) and (3.2), if $i \in I_0(x^k, \alpha^k)$ then for almost all

$k \in K_1$

$$(5.13)\quad f_{0,i}(x^k+\alpha p^k) \leq f_{0,i}(x^k) + \alpha[\beta^k - f_{0,i}(x^k) + f_0(x^k) + f^+(x^k)] + 0(\alpha)$$
$$\leq f_0(x^k) + f^+(x^k) + \alpha\beta^k + 0(\alpha) \quad .$$

Similar arguments show that for $i \in I_c(x^k, \delta^k)$ and almost all $k \in K_1$

$$(5.14)\quad f_i(x^k+\alpha p^k) \leq f^+(x^k) + \alpha\beta^k + 0(\alpha) \quad .$$

Using (5.12) and (3.1), we see that for $i \notin I_0(x^k, \delta^k)$

$$(5.15)\quad f_{0,i}(x^k+\alpha p^k) \leq f_0(x^k) + f^+(x^k) - \delta(\alpha) \quad ,$$

and that for $i \notin I_c(x^k, \delta^k)$

$$(5.16)\quad f_i(x^k+\alpha p^k) \leq f^+(x^k) - \delta + o(\alpha)$$

for almost all $k \in K_1$, where $o(\alpha) \to 0$ as $\alpha \downarrow 0$.

Let $\kappa = \max\{\varepsilon_1, m_2\}$. (5.13) through (5.16) imply that there exists a number $\underline{\alpha}(\bar{x}) > 0$ such that for almost all $k \in K_1$

$$(5.17)\quad f_{0,i}(x^k+\alpha p^k) \leq f_0(x^k) + f^+(x^k) + \alpha\kappa\underline{\beta}(\bar{x}) \qquad i = 1;n \quad ,$$

$$(5.18)\quad f_i(x^k+\alpha p^k) \leq f^+(x^k) + \alpha\kappa\underline{\beta}(\bar{x}) \qquad i = 1;m \quad ,$$

for all $\alpha \in [0, \underline{\alpha}(\bar{x})]$. Since $\phi_k(x^k+\alpha p^k) = \max\{f_0(x^k+\alpha p^k) - f_0(x^k), f(x^k)\}$ and $\phi_k(x^k) = f^+(x^k)$, the line search rules of Step 4 together with (5.17), (5.18) imply that

$$(5.19)\quad f_0(x^k+\alpha^k p^k) \leq f_0(x^k) + f^+(x^k) + \kappa\underline{\alpha}(\bar{x})\underline{\beta}(\bar{x})/2 \quad ,$$

(5.20) $\quad f(x^k + \alpha^k p^k) \leqslant f^+(x^k) + \kappa\underline{\alpha}(\bar{x})\underline{\beta}(\bar{x})/2 \quad ,$

for almost all $k \in K_1$. Here we have used the fact that $\alpha^k = 1$ at Step 3 may be accepted for $k \in K_1$ only finitely often; otherwise $\eta^{k+1} = \beta^k \geqslant \gamma\eta^k$ with $\gamma \in (0,1)$, taking place infinitely often would contradict $\beta^k \leqslant \underline{\beta}(\bar{x}) > 0$ for $k \in K_1$ (note that $\eta^{k+1} \geqslant \eta^k$ by construction).

We shall now consider two cases:

B1) Suppose that $f(x^k) \leqslant 0$ for some k. Then due to line search rules $f^+(x^k) = 0$ for almost all k, and for these $k \in K_1$ (5.19) gives

(5.21) $\quad f_0(x^{k+1}) \leqslant f_0(x^k) + \kappa\underline{\alpha}(\bar{x})\underline{\beta}(\bar{x}) \quad .$

Noting that $f^+(x^k) = 0$ and the line search rules give

$$f_0(x^{k+1}) - f_0(x^k) \leqslant \phi_k(x^{k+1}) \leqslant \phi_k(x^k) = 0 \quad ,$$

and using the continuity of f_0, we should have $f_0(x^k) \to f_0(\bar{x}), k \in K_1$, which contradicts (5.21), since $\kappa\underline{\alpha}(\bar{x})\underline{\beta}(\bar{x}) < 0$.

B2) If $f(x^k) > 0$ for all k, then the above argument of B1 with f_0 substituted by f and (5.19) by (5.20) also leads to contradiction, thus ending the proof.

Remark 5.3: Since $\beta^k < 0$ at Step 4, the above argument leading to (5.17) and (5.18) may be repeated to show the existence of some $\underline{\alpha}^k > 0$ such that

$$\phi_k(x^k + \alpha p^k) \leqslant \phi_k(x^k) + \alpha\varepsilon_l\beta^k \quad \text{for } \alpha \in [0, \underline{\alpha}^k] \quad ,$$

which proves finite termination of Step 4i. Similar approach may be used for proving finite termination of the line search of [Wierzbicki 1978] used at Step 4ii.

6. RATE OF CONVERGENCE

In this section we shall show that under favorable conditions our algorithm converges at least F-linearly (see [Pironneau and Polak 1972]). Since our analysis generalizes the results of Pironneau and Polak, we constantly refer to [Pironneau and Polak 1972] providing here essential modifications only.

Throughout this section, the functions $\{f_{0,i}\}$ and $\{f_i\}$ are assumed to be convex and twice continuously differentiable. We shall consider the problem (2.1) and our algorithm under the following hypotheses.

(6.1) Assumptions. If $f^+(x^0) = 0$, let $B = \{x \in \mathbb{R}^N: f_0(x) \leq f_0(x^o), f^+(x) = 0\}$. If $f(x) > 0$,

let $B = \{x \in \mathbb{R}^N: f(x) \leq f(x^o)\}$. We shall assume that

i) B is compact

ii) f_0 is strictly convex in B (e.g., $f_{0,i}$ $i = 1;n$ are strictly convex in B).

iii) $C' = \{x \in \mathbb{R}^N: f(x) < 0\}$ is not empty.

It follows that there exists a unique $\hat{x} \in \mathbb{R}^N$ solving (2.1). Since (6.1iii) implies (A1) of para. 5 and $\{x^k\} \subset B$ by construction, the results of para. 5 show that $x^k \to \hat{x}$. Let $\Lambda(\hat{x})$ denote the set of Lagrange multipliers of (2.1) at $\hat{x}$, i.e., $\hat{\lambda} \in \Lambda(\hat{x})$ if it satisfies (2.3 - 6). It is easy to prove (of [13, Lemma B.11]) that

(6.2) $$\underline{\lambda}^{\circ} = \min \{\sum_{i=1}^{n} \lambda_i : \lambda \in \Lambda(\hat{x})\} > 0 \quad .$$

We shall also assume that there exist constraints $\varepsilon > 0$ and $m^{\circ} \in (0,1)$ such that

(6.3) $$m^{\circ} |y - x|^2 \leq \langle y - x, L''(x,\lambda)(y - x)\rangle \text{ for all } x,y \in B(\hat{x},\varepsilon)$$
$$\text{and } \lambda \in N(\Lambda(\hat{x}),\varepsilon) \quad ,$$

where the Lagrangian L for (2.1) is defined by

(6.4) $$L(x,\lambda) = \sum_{i=1}^{n} \lambda_i f_{0,i}(x) + \sum_{i=1}^{m} \lambda_{i+n} f_i(x) \quad .$$

Since the multipliers $\{\lambda^k\}^{\infty}$ remain in a compact set and $x^k \to \hat{x}$, a closer inspection of the proof of Theorem 5.2 shows that

(6.5) $$\lambda^k \in N(\Lambda(\hat{x}),\varepsilon) \quad \text{for almost all } k \quad ,$$

(6.6) $$\beta^k \to 0 \quad .$$

We assume that the algorithm constructs an infinite sequence $\{x^k\}$ with $\gamma = 0$, i.e., that no direct prediction steps are taken. We shall start by estimating

(6.7) $$\delta^k = \phi_k(x^{k+1}) - \phi_k(x^k) \quad .$$

LEMMA 6.1

There exists a constant $\underline{\alpha} > 0$ such that

(6.8) $$\delta^k \leq \underline{\alpha}\beta^k \quad \text{for all } k.$$

Proof. Let $M = \max\{\|f''_{0,i}(x)\|, \|f''_j(x)\| : x \in B; i=1,\ldots n; j=1,\ldots,m\}$. Since

$$f_{0,i}(x^k+\alpha p^k) = f_0(x^k) + \alpha\langle p^k, f'_{0,i}(x^k)\rangle + \alpha\langle p^k, f'_{0,i}(\theta_{0,i}) - f_{0,i}(x^k)\rangle \quad ,$$

where $\theta_{0,i} = x^k + \alpha_{0,i}p^k$ and $\alpha_{0,i} \in [0,\alpha]$, by (3.2) and (5.3) we get for $\alpha \in [0,1]$ and $i \in I_0(x^k,\delta^k)$

$$f_{0,i}(x^k+\alpha p^k) \leqslant f_{0,i}(x^k) + \alpha[\beta^k - f_{0,i}(x^k) + f_0(x^k) + f^+(x^k)] + \alpha^2 L|p^k|^2$$

$$\leqslant f_0(x^k) + f^+(x^k) + \alpha\beta^k + \alpha^2 L\,|p^k|^2_{A_k}/\gamma_1 \quad .$$

In the same manner we have for $\alpha \in [0,1]$ and $i \in I_c(x^k,\delta^k)$

$$f_i(x^k+\alpha p^k) \leqslant f^+(x^k) + \alpha\beta^k + \alpha^2 L|p^k|^2_{A_k}|\gamma_1 \quad .$$

Now let $k = \max\{|f'_{0,i}(x)|, |f'_j(x)| : x \in B;\ i = 1,\ldots,n;\ j = 1,\ldots,m\}$. By (3.1) and (5.3) together with $\delta^k \geqslant \underline{\delta}$ and (3.4), if $i \notin I_0(x^k,\delta^k)$ then

$$f_{0,i}(x^k+\alpha p^k) = f_{0,i}(x^k) + \alpha\langle p^k, f'_{0,i}(\theta_{0,i})\rangle \leqslant f_0(x^k) + f^+(x^k) - \underline{\delta}$$
$$+ \alpha K\sqrt{-\beta^k}/\sqrt{\gamma_1} \ .$$

In the same way we prove for $i \notin I_c(x^k,\delta^k)$ that

$$f_i(x^k+\alpha p^k) \leqslant f^+(x^k) - \underline{\delta} + \alpha K\sqrt{-\beta^k}/\sqrt{\gamma_1} \ .$$

By definition of ϕ_k and the above estimates, $\phi_k(x^k+\alpha p^k) \leqslant \phi_k(x^k) + \kappa\alpha\beta^k$ for $\alpha \leqslant \underline{\alpha}$, where

(6.9) $\quad \underline{\alpha}^k = \min\{1, \underline{\delta}/[\sqrt{-\beta^\kappa}(K/\sqrt{\gamma_1} + \kappa\sqrt{-\beta^k})], (1-\kappa)\gamma_1/M\} \quad .$

By the rules of Step 4 of the algorithm, if α^k is accepted at Step 4i or Step 4ii, then $\alpha^k \geq \underline{\alpha}^k/2$; by the same token $\tilde{\alpha}^k \geq \underline{\alpha}^k/2$ at Step 4iii. Therefore $\phi_k(x^k + \alpha^k p^k) = \phi_k(x^k) + \min\{\varepsilon_\nu m_1\}\, \underline{\alpha}^k \beta^k/2$. Since $\beta^k \to 0$, (6.9) implies that there exists a constant $\underline{\alpha} > 0$ such that $\alpha < \min\{\varepsilon_i, m_1\}\, \underline{\alpha}^k/2$ and $\phi_k(x^{k+1}) \geq \phi_k(x^k) + \underline{\alpha}\beta^k$, which ends the proof.

Proceeding as in [Pironneau and Polak 1972], let

$$\text{(6.10)}\quad \sigma(x^k) = \min\{\sigma : f_{0,i}(x) - f_0(x^k) - \sigma \leq 0,\ i = 1,\dots,n;\ f_i(x) - \sigma \leq 0,\ i = 1,\dots,m;\ x \in B\}\ .$$

The following proposition is a trvial extension of [Pironneau and Polak 1972, Lemma 2.7].

LEMMA 6.2

Let $\bar{u}^k \in \mathbb{R}^{n+m}$ by any solution of the dual of (6.10), i.e., of

$$\text{(6.11)}\quad \max_{u \geq 0}\ \{\min_{\substack{\sigma, x \\ x \in B}} \{(1 - \sum_{i=1}^{m+n} u_i)\sigma + \sum_{i=1}^{n} u_i [f_{0,i}(x) - f_0(x^k)] + \sum_{i=1}^{m} u_{i+n} f_i(x)\}\}\ .$$

Then any accumulation point of $\{\bar{u}^k\}_0^\infty$ belongs to the set $\Lambda(\hat{x})$, and $\sum_{i=1}^{m+n} \bar{u}_i^k = 1$ for all k.

Following [Pironneau and Polak 1972, Theorem 2.11], we obtain

$$\text{(6.12)}\quad \sigma(x^k) = \min_{x \in B}\ \{\sum_{i=1}^{n} \bar{u}_i^k [f_{0,i}(x) - f_0(x^k)] + \sum_{i=1}^{m} \bar{u}_{i+n}^k f_i(x)\}\ .$$

Upon replacing x by $\hat{x}$ in (6.12) and noting that $f(\hat{x}) \leq 0$, we obtain

(6.13) $$\sigma(x^k) \leq \sum_{i=1}^{n} \bar{u}_i^k [f_0(\hat{x}) - f_0(x^k)] \quad .$$

Next, from Lemma 6.2 and (6.2) we deduce that

$$\liminf_{k \to \infty} \sum_{i=1}^{n} \bar{u}_i^k \geq \underline{\lambda}^0 > 0 \quad ,$$

which implies that, given any $\tau \in (0,1)$, there exists a $k_0(\tau)$ such that

(6.14) $$\sum_{i=1}^{n} \bar{u}_i^k \geq \underline{\lambda}^0 (1 - \tau) \text{ for all } k \geq k_0(\tau) \quad .$$

Combining (6.14) with (6.13) we now obtain

(6.15) $$\sigma(x^k) \geq \underline{\lambda}^0 (1 - \tau) [f_0(\hat{x}) - f_0(x^k)] \quad .$$

Generalizing [Pironneau and Polak 1972, Theorem 3.16], we get

LEMMA 6.3

Assume (with no loss of generality) that $m^0 \leq \gamma_{20}$. Then

(6.16) $$\sigma(x^k) \geq \frac{\gamma_2}{m^0} [\beta^k + \frac{1}{2}|p^k|^2_{A_k}] \text{ for almost all } k.$$

Proof: From [Pironneau and Polak 1972, (3.23)] we obtain that that for almost k.

(6.17) $$\sigma(x^k) = \max_{\nu \geq 0} \{ \inf_{y \in B(\hat{x},\varepsilon)} \{ \sum_{i=1}^{n} \nu_i [f_{0,i}(y) - f_0(x^k)] + \sum_{i=1}^{m} \nu_{i+n} f_i(y) \} : \sum_{i=1}^{m+n} \nu_i = 1 \}$$

Therefore for almost k

$$(6.18)\qquad \sigma(x^k) \geq \inf_{y\in B(\hat{x},\epsilon)} \{\sum_{i=1}^{n} \lambda_i^{k+1}[f_{0,i}(x^k) - f^+(x^k)] + \sum_{i=1}^{m} \lambda_{i+n}^{k+1}[f_i(x^k) - f^+(x^k)]$$

$$+ \sum_{i=1}^{n} \lambda_i^{k+1}[f_{0,i}(y) - f_{0,i}(x^k)] + \sum_{i=1}^{m} \lambda_{i+n}^{k+1}[f_i(y) - f_i(x^k)]\} \quad .$$

Expanding $f_{0,i}(y) - f_{0,i}(x^k)$ and $f_i(y) - f_i(x^k)$ to second order terms and making use of (6.3) and (6.5), we obtain that for almost all k

$$(6.19)\qquad \sigma(x^k) \geq \inf_{y\in B(\hat{x},\varepsilon)} \{\sum_{i=1}^{n} \lambda_i^{k+1}[f_{0,i}(x^k) - f_0(x^k) - f^+(x^k)] + \sum_{i=1}^{m} \lambda_{i+n}^{k+1}[f_i(x^k - f^+(x^k)]$$

$$+ \langle \sum_{i=1}^{n} \lambda_i^{k+1} f'_{0,i}(x^k) + \sum_{i=1}^{m} \lambda_{i+n}^{k+1} f'_i(x^k), y - x^k\rangle + \frac{1}{2} m^0 |y - x^k|^2\}$$

By deleting the constraint $y \in B(\hat{x},\varepsilon)$ in (6.19) and using (4.3) and (5.3), we obtain

$$(6.20)\qquad \sigma(x^k) \geq \sum_{i=1}^{n} \lambda_i^{k+1}[f_{0,i}(x^k) - f_0(x^k) - f^+(x^k)] + \sum_{i=1}^{m} \lambda_{i+n}^{k+1}[f_i(x^k) - f^+(x^k)]$$

$$- \frac{\gamma_2}{2m^0} |p^k|^2_{A_k} \quad .$$

Since $\gamma_2/m^0 \geq 1$ by assumption and the first two terms in (6.20) are nonpositive, (3.4) and (6.20) imply (6.17), which proves the lemma.

We are now ready to state the main convergence result.

THEOREM 6.4

Suppose that $f(x^k) \leqslant 0$ for some k. Then given any $\tau \in (0,1)$,

$$(6.21)\qquad f_0(x^{k+1}) - f_0(\hat{x}) \leqslant \left[1 - \frac{\underline{\alpha}\, m_0 \lambda^0 (1-2)}{\gamma_2}\right] [f_0(x^k) - f_0(\hat{x})]$$

Proof: By the rules of the algorithm, we have $f^+(x^k) = 0$ for all k, which in turn implies $\phi_k(x^k) = f^+(x^k) = 0$ for those k. From (6.7) we now obtain $\sigma^k = \phi_k(x^{k+1})$, Lemma 6.1 implies

$$(6.22)\qquad f_0(x^{k+1}) - f_0(x^k) \leqslant \phi_k(x^{k+1}) = \sigma^k \leqslant \underline{\alpha}\beta^k \quad .$$

From (6.15) and (6.16) we obtain

$$(6.23)\qquad \beta^k \leqslant \frac{m_0}{\gamma_2}\, \sigma(x^k) \leqslant \frac{m_0 \underline{\lambda}^0 (1-\tau)}{\gamma_2}\, [f_0(\hat{x}) - f_0(x^k)] \quad ,$$

for almost all k. Finally, from (6.22) and (6.23)

$$(6.24)\qquad f_0(x^{k+1}) - f_0(x^k) \leqslant \frac{m_0 \underline{\lambda}^0 \underline{\alpha}(1-\tau)}{\gamma_2} [f_0(\hat{x}) - f_0(x^k)] \quad ,$$

for almost all k. Rearranging (6.24), we obtain (6.21), which completes our proof.

COROLLARY 6.5

Suppose that $f(x^k) \leqslant 0$ for some k, then $\{x^k\}_{k=0}^{\infty}$ converges to $\hat{x}$ linearly.

Proof: Let $\bar{\lambda} \in \Lambda(\hat{x})$. According to the Taylor expansion formula, for any x^k there exist a $\theta(x^k) \in (0,1)$ such that

$$(6.25)\quad \sum_{i=1}^{n}\bar{\lambda}_i[f_{0,i}(x^k)-f_{0,i}(\hat{x})]+\sum_{i=1}^{m}\bar{\lambda}_{i+n}[f_i(x^k)-f_i(\hat{x})]=$$

$$\langle x^k-\hat{x}, \sum_{i=1}^{n}\bar{\lambda}_i f_{0,i}'(\hat{x})+\sum_{i=1}^{m}\bar{\lambda}_{i+n}f_i'(\hat{x})\rangle+\frac{1}{2}\langle x^k-\hat{x},(\sum_{i=1}^{n}\bar{\lambda}_i f_{0,i}''(\xi))(x^k-\hat{x})\rangle \quad ,$$

with $\zeta = \theta(x^k)x^k + [1-\theta(x^k)]\hat{x}$. Since $\bar{\lambda}$ satisfies (2.3) through (2.6) and $x^k \to \hat{x}$, (6.3) and (6.25) give

$$(6.26)\quad \sum_{i=1}^{n}\bar{\lambda}_i[f_0(x^k)-f_0(\hat{x})]+\sum_{i=1}^{m}\bar{\lambda}_{i+n}[f^+(x^k)-f^+(\hat{x})]\geq\frac{m_0}{2}|x^k-\hat{x}|^2 \quad .$$

Therefore, for almost all k

$$(6.27)\quad |x^k-\hat{x}|^2\leq\frac{2\sum_{i=1}^{n}\bar{\lambda}_i}{m_0}[f_0(x^k)-f_0(\hat{x})] \quad ,$$

and our thesis follows from (6.21) and (6.27), thus ending the proof.

Remark 6.6: If the initial point x^0 is feasible, we may modify the algorithm to obtain a feasible direction method generalizing [Pironneau and Polak 1973]. It suffices to re-define the improvement function ϕ_k by putting $\phi_k(x) = f_0(x) - f_0(x^k)$ and then to include an additional step-size requirement that $f(x^{k+1}) \leq 0$ in the algorithm's description. One may easily check that all results of this paper remain valid for this modification; in particular - linear convergence is retained.

REFERENCES

Clarke, F.H. 1976. A new approach to Lagrange multipliers. Math. Oper. Res. 1:165-174.

Conn, A.R., and T. Pietnykowski. 1977. A penalty function method converging directly to a constrained optimum. SIAM Journal on Num. Analysis 14:348-378.

Demianov, V.F., and V.N. Malozemov. 1972. Vvedenie v Minimaks. Moskua: Nauka (in Russian).

Han, S.P. 1977. A globally convergent method for mathematical programming. JOTA 22:297-309.

Huard, P. 1968. Programmation matematique convexe. Rev. Fr. Inform. Rech. Operation. 7:43-59.

Kiwiel, K.C. 1979. Algorithms for Non-Smooth Optimization and their Applications. Ph.D. thesis, Technical University of Warsaw, Warsaw, Poland. Forthcoming.

Mifflin, R. 1979. An algorithm for constrained optimization with semismooth functions. Mathematics of Operations Research 2:191-207.

Pironneau, O., and E. Polak. 1972. On the rate of convergence of certain methods of centers. Mathematical Programming 2:230-257.

Pironneau O., and E. Polak. 1973. Rate of convergence of a class of methods of feasible directions. SIAM Journal on Numerical Analysis. 10(1):161-174.

Polak, E. 1971. Computational Methods in Optimization. A Unified Approach. New York and London: Academic Press.

Pschenicnyi, B.N., and Yu. M. Danilin. 1975. Cislennyie Metodi v Ekstremalnih Zadacah. Moskua: Nauka (in Russian).

Wierzbicki, A.P. 1978. Lagrangian Functions and Nondifferentiable Optimization. WP-78-63 International Institute for Applied Systems Analysis, Laxenburg, Austria.