RATE OF CONVERGENCE OF A CLASS OF METHODS OF FEASIBLE DIRECTIONS

by

O. Pironneau and E. Polak

Memorandum No. ERL-M301

26 July 1971

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Abstract

This paper deals with the rate of convergence of four methods of feasible directions: the Zoutendijk procedures 1 and 2 and two modifications of these procedures due to the authors. It is shown that of these methods, the two due to the authors converge linearly under convexity assumptions, that the Zoutendijk procedure 2 converges sublinearly under these assumptions, and that the Zoutendijk procedure 1 converges linearly provided the solution of the problem is a vertex of the constraint set.

Research sponsored by the Joint Services Electronics Program, Contract F44620-71-C-0087 and the National Aeronautics and Space Administration, Grant NGL-05-003-016.

1. INTRODUCTION

Most of the currently used methods of feasible directions, such as the Frank and Wolfe method for problems with affine constraints [3] the Zoutendijk methods [11], the Zukhovitskii-Polyak-Primak method [12], and the Polak method [8] degenerate to first-order gradient methods when the number of constraints is zero (i.e. in the unconstrained case). Thus it is clear that these algorithms cannot converge better than linearly in the general case. However, linear convergence for the general case is not ensured by the behavior of a constrained optimization algorithm on unconstrained problems. Thus, it was shown by Canon and Cullum [2] that the Frank and Wolfe method converges sublinearly on constrained problems and a similar result was established by the authors in [7] for the Topkis and Veinott version of the Zoutendijk procedure 2 algorithm [9]. In both cases, the cost and constraint functions were assumed to be convex.

Whereas the above mentioned results were negative in nature, this paper presents a few positive results on the rate of convergence of some methods of feasible directions. Thus, it will be shown that in some cases a method due to Zoutendijk [11], converges R-linearly* when the solution to the problem is a vertex of the constraint set. Finally it will be shown that two methods due to the authors converge

*A discussion of Root-order of convergence can be found in [6] Section 9.2. For our purposes it suffices to know that if a sequence $\{x_i\}_{i=0}^{\infty}$ satisfies $\|x_i-\hat{x}\| \leq K\gamma^i$, $i = 0, 1, 2, \ldots$, with $\gamma \in (0,1)$ and $K < \infty$, then $x_i \to \hat{x}$ as $i \to \infty$ at least R-linearly (i.e. with a root order at least equal to 1).

R-linearly under reasonably general assumptions, without any restrictions on the location of the optimal point. The first of these methods is closely related to the Zoutendijk Procedure 2 (p. 74 [11]), while the second one is a cross between the Zoutendijk Procedure 2 and the Zoutendijk Procedure 1, (p. 73 [11]). These methods have not been described before. As we shall see, when the number of constraints active at the solution is small relative to the number of variables then both of these new methods are superior to their progenitors, because in that case their rate of convergence and computational complexity do not depend upon the dimension of the space in which the problem is defined.

2. PRELIMINARIES

Throughout this paper we shall consider algorithms for solving problems of the form

2.1 $\min\{f^0(z) \mid f^j(z) \leq 0,\ j = 1, 2, \ldots, m\}$,

where $f^j : \mathbb{R}^n \to \mathbb{R}^1$, $j = 0, 1, 2, \ldots, m$, are continuously differentiable functions. To establish a rate of convergence for the algorithms to be considered, it will be necessary to postulate the following hypotheses (these are considerably stronger than the ones needed to establish convergence to a stationary point).

2.2 Assumptions

(i) The problem (2.1) has a unique solution $\hat{z}$.

(ii) The functions $f^j(\cdot)$, $j = 0, 1, 2, \ldots, m$, are convex and twice continuously differentiable.

(iii) The set $C' = \{z \mid f^j(z) < 0,\ j = 1, 2, \ldots, m\}$ is not empty.

(iv) There exist constants $\ell > 0$ and $\rho > 0$ such that

2.3 $$\ell\|y\|^2 \leq \langle y, \frac{\partial^2 f^0(z)}{\partial z^2} y \rangle$$

for all $y \in \mathbb{R}^n$, for all $z \in B(\hat{z},\rho) \triangleq \{z \mid \|z-\hat{z}\| \leq \rho\}$. ¤

The algorithms we are about to discuss differ from one another mainly in the subprocedure for finding a feasible direction. Because of this, we can combine their statements into a single algorithm with a parameter whose value determines the particular direction finding subprocedure to be used.

2.4 Algorithm (Methods of Feasible Directions)

Step 0: Select scale factors $\varepsilon_0 > 0$, $\beta \in (0,1)$ and an integer $q \geq 1$; select a direction finding subprocedure indicator $p \in \{Z1, Z2, PP1, PP2\}$; select a normalization set $S \subset \mathbb{R}^n$, which is a compact, convex neighborhood of the origin.

Comment: The indicators Z1, Z2, PP1 and PP2 designate the Zoutendijk procedure 1 (p. 73 [11]), the Zoutendijk procedure 2 (p. 74 [11]) the two procedures introduced in this paper by the authors.

Step 1: Compute a $z_0 \in \{z \mid f^i(z) \leq 0,\ j = 1, 2, \ldots, m\}$, and set $i = 0$, $\varepsilon = \varepsilon_0$.

Step 2: Set $z = z_i$, set

2.5 $$I(z,\varepsilon) = \{j \in \{1, 2, \ldots, m\} \mid f^j(z) \geq -\varepsilon\},$$

2.5' $$I(z,0) = \{j \in \{1, 2, \ldots, m\} \mid f^j(z) = 0\}$$

and compute $(h^0(z,\varepsilon),\ h(z,\varepsilon)) \in \mathbb{R}^{n+1}$ and $(h^0(z,0),\ h(z,0)) \in \mathbb{R}^{n+1}$ as solutions of (2.5p) below (where p was selected in Step 0) for $\bar{\varepsilon} = \varepsilon,\ 0$ (Note that (2.6Z2) and (2.6 PP1) do not depend on ε).[+]

2.6 Z1 $$h^0(z,\bar{\varepsilon}) = \min_{h \in S} \max \{\langle \nabla f^j(z),\ h \rangle \,|\, j \in I(z,\bar{\varepsilon}) \cup \{0\}\}$$

2.6 Z2 $$h^0(z,\bar{\varepsilon}) = \min_{h \in S} \max \{\langle \nabla f^0 z), h \rangle ;\ f^j(z) + \langle \nabla f^j(z),\ h \rangle ,\ j = 1,\ 2,\ \ldots,\ m\}$$

2.6 PP1 $$h^0(z,\bar{\varepsilon}) = \min \{\tfrac{1}{2} \|h\|^2 + \max \{\langle \nabla f^0(z),\ h \rangle ;\ f^j(z) + \langle \nabla f^j(z),\ h \rangle ,\ j = 1,\ 2,\ \ldots,\ m\}\}.$$

2.6 PP2 $$h^0(z,\bar{\varepsilon}) = \min \{\tfrac{1}{2} \|h\|^2 + \max \{\langle \nabla f^0(z),\ h \rangle ;\ f^j(z) + \langle \nabla f^j(z),\ h \rangle ,\ j \in I(z,\varepsilon)\}\}.$$

<u>Step 3</u>: If $h^0(z,0) = 0$, set $\hat{z} = z$ and stop; else go to Step 4.

<u>Step 4</u>: If $p \in \{Z1,\ PP2\}$, go to Step 5; else, go to Step 6.

<u>Comment</u>: For p = Z1, PP2, the direction finding subprocedure must also find a correct value for ε. This is done by means of the test in Step 5.

<u>Step 5</u>: If $h^0(z,\varepsilon) \leq -\varepsilon^q$

$\varepsilon = \beta\varepsilon$ and go to Step 2.[*]

<u>Step 6</u>: Compute μ_i to be the solution of

[+]In practice the calculation for $\bar{\varepsilon} = 0$ need only be carried out when $\varepsilon \leq \varepsilon'$, a precision parameter.

[*]ε_i is defined here only for use in the proofs to follow.

2.7 $\min \{f^0(z+\mu h(z,\varepsilon)) \mid f^j(z+\mu h(z,\varepsilon)) \leq 0, \quad j = 1, 2, \ldots, m\}$,

and set $z_{i+1} = z + \mu_i h(z,\varepsilon)$.

<u>Step 7</u>: Set $i = i + 1$, and go to Step 2. ¤

<u>Note</u>: To indicate the particular version of algorithm (2.4) under discussion, we shall use the self explanatory notation (2.4 Z1) (2.4 Z2), etc.

Before proceding any further, it may be interesting to observe that it is much more efficient to solve the duals of (2.6 PP1) and (2.6 PP2) than the primals. Thus, a solution $(h^0(z,\varepsilon), h(z,\varepsilon))$ of (2.6 PP1) is given by

$$h(z,\varepsilon) = -\sum_{j=0}^{m} u^j \nabla f^j(z)$$

with $u = (u^0, u^1, \ldots, u^m)$ any solution of the quadratic programming problem

$$h^0(z,\varepsilon) = \max \{ \sum_{j=1}^{m} u^j f^j(z) - \frac{1}{2} \| \sum_{j=0}^{m} u^i \nabla f^i(z) \|^2 \mid u^j \geq 0,$$

$$j = 0, 1, \ldots, m, \sum_{j=0}^{m} u^j = 1\}$$

the expressions for the dual of (2.6 PP2) are entirely analogous. Note that the above quadratic program is quite simple when the number of constraints is small. When the dimension n of z is more than twice the number of constraints m ($n \geq 2m$), this quadratic program will be easier to solve (by means of the Lemke algorithm [4]) than the linear program (2.6 Z2). A similar statement holds in a comparison of (2.6Z1) with (2.6 PP2). In this case, however, the above mentioned relation $n \geq 2m$ can be replaced by $n \geq 2\bar{m}$ where $\bar{m}$ is the cardinality of the set $I(z,\varepsilon)$.

2.8 Theorem: Let $\{z_i\}$ be a sequence constructed by algorithm (2.4) in solving (2.1), and suppose that the assumptions (2.2) are satisfied. Then, either $\{z_i\}$ is finite and its last element, $\hat{z}$, is the unique solution of (2.1), or else $z_i \to \hat{z}$ as $i \to \infty$, where $\hat{z}$ is the unique solution of (2.1). Furthermore, if $p \in \{Z1, PP2\}$ and $\{\varepsilon_i\}$ is infinite, then $\varepsilon_i \to 0$ as $i \to \infty$. ¤

For a proof of this theorem, the reader is referred to [11] for p = Z1, Z2, or, alternatively, to Polak [8] where all these cases are considered. The proof for p = PP1 follows trivially from the case of p = Z2, while the proof for p = PP2 can be obtained by suitably adapting the proof for the case p = Z1.

The following lemma (see [7]) shows that under the assumptions (2.2), R-linear convergence in cost implies R-linear convergence in norm.

2.9 Lemma: Consider problem (2.1) and suppose that the assumptions (2.2) are satisfied. Suppose that in the process of solving (2.1), an algorithm constructs a sequence $\{z_i\}_{i=0}^{\infty}$ such that for some integer $i_0 \geq 0$, there exist a $\gamma \in (0,1)$ and a $K > 0$ satisfying

2.10 $$f^0(\hat{z}_i) - f^0(z) \leq K\gamma^i \quad \text{for all } i \geq i_0,$$

where $\hat{z}$ is the unique solution of (2.1). Then

2.11 $$\|z_i - \hat{z}\| \leq \sqrt{\frac{2K}{\ell}}\,(\sqrt{\gamma}\,)^i \quad \text{for all } i \geq i_0,$$

where ℓ is as in (2.3). ¤

3. THE ZOUTENDIJK PROCEDURE 1

We shall now consider algorithm (2.4) when the direction finding subprocedure Z1, is used. The results in this section will be seen to be qualitatively independent of the normalization set S used. In addition to the assumptions (2.2), we shall need the following one.

3.1 Assumption: Let $\hat{z}$ be the unique solution of (2.1). Then there exists a $\hat{\rho} > 0$ such that if $I \subset \{1, 2, \ldots, m\}$* satisfies $\theta \in \text{co}\,\{\nabla f^j(\hat{z})\}_{j \in I \cup \{0\}}$, then $\theta \in \text{co}\,\{\nabla f^j(z)\}_{j \in I \cup \{0\}}$ for all $z \in \{z \mid \|z-\hat{z}\| \leq \hat{\rho};\ f^j(z) \leq 0,\ j = 1, 2, \ldots, m\}$, where θ denotes the zero vector in $\mathbb{R}^n$ and co denotes the convex hull of the set. ¤

Thus, (3.1) states that there exists a $\hat{\rho} > 0$ such that if $\sum_{j \in I \cup \{0\}} \hat{\mu}^j \nabla f^j(\hat{z}) = 0$, with $\hat{\mu}^j \geq 0$ and $\sum_{j \in I \cup \{0\}} \hat{\mu}^j = 1$, then for all feasible z satisfying $\|z-\hat{z}\| \leq \hat{\rho}$, there exist $\mu^j(z) \geq 0$, $j \in I \cup \{0\}$, with $\sum_{j \in I \cup \{0\}} \mu^j(z) = 1$, such that $\sum_{j \in I \cup \{0\}} \mu^j(z)\nabla f^j(z) = 0$. Since the solution $\hat{z}$ of (2.1) is unique, assumption (3.1) implies that $\nabla f^0(\hat{z}) \neq 0$. Assumption (3.1) will be satisfied when $\hat{z}$ is a "vertex" of the constraint set $C = \{z \mid f^j(z) \leq 0,\ j = 1, 2, \ldots, m\}$, the gradients

* Note that the empty set is a possible value for I.

$\nabla f^j(\hat{z})$, $j \in I(\hat{z},0)$, are linearly independent, and $\nabla f^0(\hat{z}) \neq 0$.

3.2 Lemma: Suppose that $\{z_i\}_{i=0}^{\infty}$ and $\{\varepsilon_i\}_{i=0}^{\infty}$ are two corresponding sequences constructed by algorithm (2.4 Z1), with $q = 1$ and S an arbitrary normalization set, and suppose that assumptions (2.2) and (3.1) are satisfied. Then, given any $\alpha \in (0,1)$, there exists an integer $i_0(\alpha)$ such that

3.3 $$h^0(z_i,\varepsilon_i) \leq \eta(\hat{z})\alpha < 0 \text{ for all } i \geq i_0(\alpha),$$

where $h^0(z_i,\varepsilon_i)$ is defined by (2.6 Z1) and

3.4 $$\eta(\hat{z}) = \max \{\eta_I(\hat{z}) \mid I \subset \{1, 2, \ldots, m\},$$
$$\theta \notin \text{co } \{\nabla f^j(\hat{z})\}_{j \in I \cup \{0\}}\},$$

with

3.5 $$\eta_I(\hat{z}) = \min_{h \in S} \max_{j \in I \cup \{0\}} \langle \nabla f^j(\hat{z}), h \rangle$$

(S is the "normalization" set appearing in (2.6 Z1)).

Proof: We begin by showing that $\eta(\hat{z}) < 0$. For every $z \in C$, let

3.6 $$\mathcal{G}(z) = \{I \subset 1, 2, \ldots, m\} \mid \theta \notin \text{Co } \{\nabla f^j(z)\}_{j \in I \cup \{0\}}\}$$

Let $I \in \mathcal{G}(\hat{z})$ be arbitrary and suppose that $\eta_I(\hat{z}) = 0$. We shall show that this leads to a contradiction. Thus, first rewriting (3.5) in convex hull form and then applying Von Neumann's minmax theorem [5], we obtain, 3.7 $$\eta_I(\hat{z}) = \min_{h \in S} \max_{u \in U_I} \sum_{j=0}^{m} u^j \langle \nabla f^j(\hat{z}), h \rangle$$

$$= \max_{u \in U_I} \min_{h \in S'} \sum_{j=0}^{m} u^j \langle \nabla f^j(z), h \rangle ,$$

where

3.8 $$U_I = \{(u^0, u^1, \dots, u^m) | u^j \geq 0, j = 0, 1, \dots, m, \sum_{j \in I \cup \{0\}} u^j = 1, \text{ and } u^j = 0 \text{ for all } j \notin I \cup \{0\}\}.$$

Hence, for some $\bar{u} \in U_I$, we must have

3.9 $$0 = \eta_I(\hat{z}) = \min_{h \in S} \sum_{j=1}^{m} \bar{u}^j \langle \nabla f^j(\hat{z}), h \rangle .$$

But S is a neighborhood of the origin and hence (3.9) implies that $\sum_{j \in I \cup \{0\}} \bar{u}^j \nabla f^j(\hat{z}) = 0$, i.e. that $\theta \in \text{Co } \{\nabla f^j(\hat{z})\}_{j \in I \cup \{0\}}$, which contradicts our assumption that $I \in \mathcal{J}(\hat{z})$. Consequently we must have $\eta_I(\hat{z}) < 0$ for all $I \in \mathcal{J}(\hat{z})$, i.e., $\eta(\hat{z}) < 0$.

Now, since the functions $\eta_I : C \to \mathbb{R}^n$, defined by (3.5) are continuous, and since by assumption (3.1) there exists a $\hat{\rho} > 0$ such that $\mathcal{J}(\hat{z}) \supset \mathcal{J}(z)$ for all $z \in C$, $\|z-\hat{z}\| \leq \hat{\rho}$, we conclude that $\eta : C \to \mathbb{R}^1$, defined by (3.4) is upper semi-continuous at $\hat{z}$. Now, by construction (see Step 4 of (2.4)), for $i = 0, 1, 2, \dots$, we have $h^0(z_i, \varepsilon_i) \leq -\varepsilon_i < 0$ and hence, by comparing (2.6 Z1) with (3.7), we must have $I(z_i, \varepsilon_i) \in \mathcal{J}(z_i)$.

Consequently, $h^0(z_i,\varepsilon_i) \leq \eta(z_i)$ for all i, and hence (because η is u.s.c.), for any $\alpha \in (0,1)$, there exists an integer $i_0(\alpha)$ such that for all $i \geq i_0(\alpha)$,

3.10 $$h^0(z_i,\varepsilon_i) \leq \eta(z_i) \leq \eta(\hat{z})\alpha < 0,$$

which completes our proof. ¤

3.11 <u>Lemma</u>: Consider the corresponding sequences $\{z_i\}_{i=0}^{\infty}$, $\{\varepsilon_i\}_{i=0}^{\infty}$ constructed by algorithm (2.4 Z1), with q = 1 and S an arbitrary normalization set, and suppose that assumptions (2.2) and (3.1) are satisfied. Let K be the infinite subset of the integers such that

3.12 $$\varepsilon_i < \varepsilon_{i-1} \text{ for all } i \in K.$$

Then, given any $\alpha \in (0,1)$, there exists an integer $i_1(\alpha)$ such that

3.13 $$\varepsilon_i \geq \beta\alpha\, \bar{\lambda}^0[f^0(z_i) - f^0(\hat{z})] \text{ for all } i \in K,\ i \geq i_1(\alpha),$$

where $\hat{z}$ is the unique solution of (2.1) and

3.14 $$\bar{\lambda}^0 = \min \{\lambda^0 \mid \sum_{j=0}^{m} \lambda^j\, \nabla f^j(\hat{z}) = 0, \sum_{j=1}^{m} \lambda^j f^j(z) = 0,$$
$$\sum_{j=0}^{m} \lambda^j = 1,\ \lambda^j \geq 0,\ j = 0, 1, \ldots, m\} > 0.$$

<u>Proof</u>: First we note that because of assumption (2.2), $\bar{\lambda}^0$ in (3.14) is indeed strictly positive (see (Lemma 1.12) in [7]). Next (see 2.5), let

3.15 $\quad I_i = I(z_i, \frac{\varepsilon_i}{\beta}), \; i = 0, 1, 2, \ldots$

Then, by definition of I_i

3.16 $$\min_{j \in I_i} f^j(z_i) \geq \frac{-\varepsilon_i}{\beta} .$$

Now, since the f^j, $j = 0, 1, \ldots, m$, are convex and $f^j(\hat{z}) \leq 0$ for $j = 1, 2, \ldots, m$, we have for $i = 0, 1, 2, \ldots,$ $f^0(z_i) - f^0(\hat{z}) \leq \langle \nabla f^0(z_i), z_i - \hat{z} \rangle ; f^j(z_i) \leq \langle \nabla f^j(z_i), z_i - \hat{z} \rangle$, $j = 1, 2, \ldots, m$. Consequently,

3.17 $$\min_{j \in I_i} f^j(z_i) = \min_{u \in U_{I_i}} \sum_{j \in I_i} u^j f^j(z_i)$$

$$= \min_{u \in U_{I_i}} \{ \sum_{j \in I_i} u^j f^j(z_i) + u^0[f^0(z_i) - f^0(\hat{z})]$$

$$- u^0[f^0(z_i) - f^0(\hat{z})]\}$$

$$\leq \min_{u \in U_{I_i}} \{ \sum_{j \in I_i \cup \{0\}} u^j \langle \nabla f^j(z_i), z_i - \hat{z} \rangle$$

$$- u^0[f^0(z_i) - f^0(\hat{z})]\},$$

where U_{I_i} is defined by (3.8). We shall now show that for every $i \in K$ there is a $\bar{u}_i \in U_{I_i}$ such that $\sum_{j \in I_i \cup \{0\}} \bar{u}_i^j \nabla f^j(z_i) = 0$. First we recall that for $i \in K$, $\varepsilon_i < \varepsilon_{i-1}$ and hence, by construction of ε_i, we must have

3.18 $$0 \geq h^0(z_i, \frac{\varepsilon_i}{\beta}) > -\frac{\varepsilon_i}{\beta}$$

Since $\varepsilon_i \to 0$ as $i \to \infty$, for any $\alpha \in (0,1)$, there exists an integer $i_1'(\alpha) \geq i_0(\alpha)$, such that $-\varepsilon_i > \eta(\hat{z})\alpha$ for all $i \geq i_1'(\alpha)$, where $\eta(\hat{z})$ is defined by (3.4) and $i_0(\alpha)$ is as in (3.3). Now, from the proof of Lemma (3.2) we conclude that given any $\alpha \in (0,1)$, and any $\varepsilon > 0$ we must have either $h^0(z_i,\varepsilon) = 0$ in which case $I(z_i,\varepsilon) \notin \mathcal{J}(z_i)$, or $h^0(z_i,\varepsilon) \leq \eta(\hat{z})\alpha$, for all $i \geq i_0(\alpha)$. Since $-\varepsilon_i > \eta(\hat{z})\alpha$ for $i \geq i_1(\alpha) \geq i_0(\alpha)$, we conclude that $h^0(z_i, \frac{i}{}) = 0$ for all $i \in K$, $i \geq i_1(\alpha)$. However, for this to be true, since $h^0(z_i,\varepsilon_{i-1}) = \min_{h \in S} \max_{u \in U_{I_i}} \sum_{j=0}^{m} u^j \langle \nabla f^j(z_i), h \rangle$,

there must exist a $\bar{u}_i \in U_{I_i}$ such that $\sum_{j \in I_i \cup \{0\}} \bar{u}_i^j \nabla f^j(z_i) = 0$.

Hence, returning to (3.17), we find that given any $\alpha \in (0,1)$, there exists an $i_1'(\alpha)$ such that

3.19 $$\min_{j \in I_i} f^j(z_i) \leq -\bar{u}_i^0 [f^0(z_i) - f^0(\hat{z})] \text{ for all } i \in K,\ i \geq i_1'(\alpha),$$

where $\bar{u}_i^0$ is the first component of a vector satisfying

3.20 $$\bar{u}_i \in U_{I_i}, \quad \sum_{j \in I_i \cup \{0\}}^{m} \bar{u}_i^j \nabla f^j(z_i) = 0.$$

Now let $\{\bar{u}_i\}_{i \in K}$, be such that (3.19), (3.20) holds and let K' be an infinite subset of K such that $\lim_{\substack{i \in K' \\ i \to \infty}} \bar{u}_i^0 = \liminf \bar{u}_i^0$. Let K" be an infinite subset of K' and $\hat{I}$ a subset of $\hat{I}(z,0)$ such that $\hat{I}_i = I$ for all $i \in K''$. It is not difficult to see that such a K'' exists. Since the set valued map $\Gamma_{\hat{I}} : C \to 2^{U_{\hat{I}}}$ defined by

$$3.21 \qquad \Gamma_{\hat{I}}(z_i) = \{u \in U_{\hat{I}} | \sum_{j=0}^{m} u^j \nabla f^j(z_i) = 0\}$$

is closed (see [1], p. 111) and not empty for all $i \in K''$. Hence, since $\hat{I} \subset I(\hat{z},0)$, and $z_i \to \hat{z}$ as $i \to \infty$, $i \in K''$, the set $\Gamma_{\hat{I}}(\hat{z})$ is a non empty subset of the optimal multipliers at $\hat{z}$, i.e.,

$$3.22 \qquad \Gamma_{\hat{I}}(\hat{z}) \subset \Lambda(\hat{z}) \triangleq \{\lambda \in U_{I(\hat{z},0)} | \sum_{j=0}^{m} \lambda^j \nabla f^j(\hat{z}) = 0\}.$$

Consequently,

$$\liminf \bar{u}_i^0 = \lim_{\substack{i \in K'' \\ i \to \infty}} \bar{u}_i^0 \geq \min \{\lambda^0 | \lambda \in \Lambda(\hat{z})\} = \bar{\lambda}^0 > 0.$$

We therefore conclude that, given any $\alpha \in (0,1)$ there exists an $i_1(\alpha) \geq i_1'(\alpha)$ such that

$$3.23 \qquad \bar{u}_i^0 \geq \bar{\lambda}^0(1-\alpha) \qquad \text{for all } i \geq i_1(\alpha),$$

with $i_1(\alpha)$ independent of the particular sequence $\{\bar{u}_i\}$ chosen. Combining (3.16), (3.19) and (3.23), we obtain (3.13), which completes our proof. ¤

3.24 __Theorem:__ Let $\{z_i\}_{i=0}^{\infty}$ be a sequence generated by algorithm (2.4 Z1) (i.e. by the Zoutendijk procedure 1), in solving (2.1) with $q = 1$ and S an arbitrary normalization set, and suppose that assumptions (2.2) (i)-(iii) and (3.1) are satisfied. Then given any $\alpha \in (0,1)$, there exists an integer $i(\alpha)$ such that for all $j \_ 0$,

$$3.25 \quad f^0(z_{i(\alpha)+j}) - f^0(\hat{z}) \le (1 - \frac{\beta\bar{\lambda}^0|\eta(z)|\alpha^4}{LW} \times [f^0(z_{i(\alpha)} - f^0(\hat{z})]$$

$$\times [f^0(z_{i(\alpha)} - f^0(\hat{z})] \text{ for all } j \ge 0,$$

where

$$3.26 \quad W = \max \{\|h\| \,|\, h \in S\}$$

$$3.27 \quad L = \max \{\|\nabla f^j(\hat{z})\|, \; j = 0, 1, \ldots, m; \; 1\},$$

and $\bar{\lambda}^0$, $\eta(\hat{z})$ are defined by (3.4), (3.14) respectively, and $\hat{z}$ is the solution of (2.1).

<u>Proof</u>: Let $\{\varepsilon_i\}_{i=0}^{\infty}$ be the sequence associated with $\{z_i\}_{i=0}^{\infty}$, with ε_i defined as in Step 4 of (2.4), and let $\{h(z_i,\varepsilon_i)\}_{i=0}^{\infty}$ be defined as in Step 2 of (2.4 Z1) (for $z = z_i$, $\varepsilon = \varepsilon_i$). Then, because of convexity, for any $\mu \ge 0$ and $i = 0, 1, 2, \ldots$

$$3.28 \quad f^j(z_i + \mu h(z_i,\varepsilon_i)) \le f^j(z_i) + \mu\langle \nabla f^j(z_i + \mu h(z_i,\varepsilon_i)), h(z_i,\varepsilon_i) \rangle,$$

$$j = 0, 1, 2, \ldots, m.$$

Next, since $z_i \to \hat{z}$, the unique solution of (2.1), and $\varepsilon_i \to 0$, given any $\alpha \in (0,1)$,* there exists, by continuity, an integer $i_2(\alpha)$ such that

*Throughout this proof we assume that $\alpha \in (0,1)$ is arbitrary, but fixed.

$$3.29 \qquad \|\nabla f^j(z_i + \mu h(z_i,\varepsilon_i))\| \leq L/\alpha, \; j = 0, 1, 2, \ldots, m,$$

for all $i \geq i_2(\alpha)$ and for all $\mu \in [0, \frac{\varepsilon_i}{LW}]$. Now, suppose that for some $i \geq i_2(\alpha)$, $j \in \{1, 2, \ldots, m\}$, but $j \notin I(z_i,\varepsilon_i)$ (see (2.5)). Then $f^j(z_i) \leq -\varepsilon_i$ and hence (3.28) and (3.29) together with the Schwartz inequality give, for this i and j,

$$3.30 \qquad f^j(z_i + \mu h(z_i,\varepsilon_i)) \leq -\varepsilon_i + \frac{\mu L}{\alpha}\|h(z_i,\varepsilon_i)\| \text{ for all } \mu \in [0, \frac{\varepsilon_i}{LW}].$$

It therefore follows from (2.7) and (3.30) that for all $i \geq i_2(\alpha)$

$$3.31 \qquad f^0(z_{i+1}) - f^0(z_i) \leq \min \{f^0(z_i + \mu h(z_i,\varepsilon_i)) - f^0(z_i) \mid f^j(z_i + \mu h(z_i,\varepsilon_i)) \leq 0,\; j \in I(z_i,\varepsilon_i);\; \mu \in [0, \alpha\varepsilon_i/L\|h(z_i,\varepsilon_i)\|]\}$$

$$\leq \min_{0 \leq \mu \leq \frac{\alpha\varepsilon_i}{L\|h(z_i,\varepsilon_i)\|}} \max \{f^0(z_i + \mu h(z_i,\varepsilon_i)) - f^0(z_i);\; f^j(z_i + \mu h(z_i,\varepsilon_i)),\; j \in I(z_i,\varepsilon_i)\}$$

$$= \min_{0 \leq \mu \leq \frac{\alpha\varepsilon_i}{L\|h(z_i,\varepsilon_i)\|}} \max_{u \in U_{I(z_i,\varepsilon_i)}} \{\sum_{j=0}^{m} u^j f^j(z_i + \mu h(z_i,\varepsilon_i)) - u^0 f^0(z_i)\},$$

where $U_{I(z_i,\varepsilon_i)}$ is defined by (3.8) for $I = I(z_i,\varepsilon_i)$. Let

$$3.32 \quad M = \max \{ \| \frac{\partial^2 f^j(z_i + \bar{\mu}_i h(z_i,\varepsilon_i))}{\partial z^2} \| \mid j = 0, 1, \ldots, m,$$

$$i \geq i_2(\alpha),\ \bar{\mu}_i \in [0, \alpha\varepsilon_i / L\|h(z_i,\varepsilon_i)\|]\}$$

Then, expanding the last expression of (3.31) to second order terms and making use of (3.32), (2.6 Z1) and (3.8), we obtain for all $i \geq i_2(\alpha)$,

$$3.33 \quad f^0(z_{i+1}) - f^0(z_i) \leq \min_{0 \leq \mu \leq \frac{\alpha\varepsilon_i}{L\|h(z_i,\varepsilon_i)\|}} \max_{u \in U_{I(z_i,\varepsilon_i)}} \{ \sum_{j=1}^{m} u^i f^j(z_i)$$

$$+ \mu \sum_{j=0}^{m} u^j \langle \nabla f^j(z_i), h(z_i,\varepsilon_i) \rangle + \frac{\mu^2}{2} M\|h(z_i,\varepsilon_i)\|^2\}$$

$$\leq \min_{0 \leq \mu \leq \frac{\alpha\varepsilon_i}{L\|h(z_i,\varepsilon_i)\|}} \{\mu h^0(z_i,\varepsilon_i) + \frac{\mu^2}{2} M\|h(z_i,\varepsilon_i)\|^2\},$$

since $\sum_{j=1}^{m} u^j f^j(z_i) \leq 0$ for all $i \geq 0$ and for all $u \in U_{I(z_i,\varepsilon_i)}$.

Next, let $i_0(\alpha) \geq i_2(\alpha)$ be such that (3.3) is satisfied. Then (3.33), (3.26), (3.32) and (3.3) combine to give for all $i \geq i_0(\alpha)$

$$3.34 \quad f^0(z_{i+1}) - f^0(z_i) \leq \min_{0 \leq \mu \leq \frac{\alpha\varepsilon_i}{LW}} \{\mu\eta(\hat{z})\alpha + \frac{\mu^2}{2} MW^2\}.$$

Since $\varepsilon_i \to 0$ as $i \to \infty$, there exists an integer $i_3(\alpha) \geq i_0(\alpha)$ such that the min in (3.34) is achieved at $\mu = \alpha\varepsilon_i/LW$ for all $i \geq i_3(\alpha)$, and hence for all $i \geq i_3(\alpha)$, we have

3.35 $$f^0(z_{i+1}) - f^0(z_i) \leq \frac{\varepsilon_i \eta(\hat{z})\alpha^2}{LW} + \frac{\varepsilon_i^2 \alpha^2 M}{L^2} .$$

Again since $\varepsilon_i \to 0$ as $i \to \infty$, we conclude from (3.35) that there exists an $i_4(\alpha) \geq i_3(\alpha)$ such that for all $i \geq i_4(\alpha)$

3.36 $$f^0(z_{i+1}) - f^0(z_i) \leq \frac{\varepsilon_i \eta(\hat{z})\alpha^3}{LW} .$$

We are now ready to make use of lemma (3.11), where we set $i_1(\alpha) \geq i_4(\alpha)$. For $i = 0, 1, 2, \ldots$, let $k(i)$ be the integer satisfying

3.37 $$\varepsilon_i = \varepsilon_{i-1} = \ldots = \varepsilon_{i-k(i)} < \varepsilon_{i-k(i)-1}$$

Then (3.36) and (3.13) combine to give (after adding and subtracting terms)

3.38 $$f^0(z_{i+1}) - f^0(z_{i-k(i)}) \leq \frac{\eta(\hat{z})\alpha^3}{LW} (\varepsilon_i + \varepsilon_{i-1} + \ldots + \varepsilon_{i-k(i)})$$

$$= - \frac{\eta(\hat{z})\alpha^3}{LW} (1 + k(i))\varepsilon_{i-k(i)}$$

$$\leq - \frac{\beta\eta(\hat{z})\alpha^{4-0}\lambda}{LW} (1 + k(i))[f^0(z_{i-k(i)}) - f^0(\hat{z})]$$

for all $i \geq i_1(\alpha)$. Rearranging (3.38) we obtain, for all $i \geq i_1(\alpha)$,

3.39 $$f^0(z_{i+1}) - f^0(\hat{z}) \leq [1 + \frac{\beta\eta(\hat{z})\alpha^{4-0}\lambda}{LW} (1 + k(i))][f^0(z_{i-k(i)}) - f^0(\hat{z})].$$

Note that since $f^0(z_{i+1}) - f^0(\hat{z}) > 0$, (3.39) implies that $k(i)$ is bounded. Also, since $0 < |\beta\eta(\hat{z})\alpha^4\bar{\lambda}^0/LW| < 1$, we must have

$$3.40 \quad 1 + \frac{\beta\eta(\hat{z})\alpha^4\bar{\lambda}^0}{LW}(1 + k(i)) \leq \left(1 + \frac{\beta\eta(\hat{z})\alpha^4\bar{\lambda}^0}{LW}\right)^{1+k(i)}$$

But (3.39) and (3.40) imply (3.25), and hence we are done. ¤

To conclude our discussion of the Zoutendijk procedure 1, we shall comment on the effect of the normalization set S on the rate of convergence. Thus, suppose that m = 0 in (2.1) and that $S = \{h \mid |h|^i \leq 1, i = 1, 2, \ldots, m\}$. This is the most popular choice for S. Then algorithm (2.4 Z1) becomes a linearly convergent method of steepest descent, with the sup norm on $\mathbb{R}^n$, for which a bound on the rate of convergence is given by (see [8]),

$$3.41 \quad f^0(z_{i+1}) - f^0(\hat{z}) \leq \left(1 - \frac{\ell_\alpha^2}{2M^{0^2} n}\right)^i [f^0(z_i) - f^0(\hat{z})] \text{ for all } i \geq 0$$

where ℓ is as in (2.3) and

$$3.42 \quad M^0 = \left\| \frac{\partial^2 f^0(\hat{z})}{\partial z^2} \right\|$$

Note that the exponentiation constant $(1 - \ell/2M^0 n) \to 1$ as $n \to \infty$ and hence that the algorithm may deteriorate as n, the number of variables in the problem, increases. In fact, this prediction of deterioration is supported by experimental evidence. Thus, a choice of $S = \{h \mid |h|^i \leq 1,$

$i = 1, 2, \ldots, n\}$ makes (2.4 Z1) highly sensitive to the number of variables i the problem. The reason why the avove choice for S is popular is that it makes (2.4 Z1) a linear program.

An alternative choice for a normalization set is $S = \{h \mid \|h\|^2 \leq 1\}$, which at least in the unconstrained case removes the dependence on n of the exponentiation constant governing the linear rate. This choice of S results in direction finding problems of the form

$$3.43 \quad h^0(z,\varepsilon) = \min_{\|h\|^2 \leq 1} \max_{j \in I(z,\varepsilon) \cup \{0\}} \langle \nabla f^j(z), h \rangle$$

$$= \min_{\|h\|^2 \leq 1} \max_{u \in U_{I(z,\varepsilon)}} \sum_{j=0}^{m} u^j \langle \nabla f^j(z), h \rangle$$

To compute a solution of (3.43), we can make use of the Von Neumann minimax theorem and of the Kuhn-Tucker conditions to conclude that the minimizer $h(z,\varepsilon)$ is of the form

$$3.44 \quad h(z,\varepsilon) = -(1/\| \sum_{j=0}^{m} \bar{u}^j \nabla f^j(z) \|) \sum_{j=0}^{m} \bar{u}^j \nabla f^j(z),$$

for some $\bar{u} \in U_{I(z,\varepsilon)}$ which is a solution of

$$3.45 \quad \max_{u \in U_{I(z,\varepsilon)}} \| \sum_{j=0}^{m} u^j \nabla f^j(z) \|.$$

Now it so happens that (3.45) is equivalent to

$$3.46 \qquad \max_{u \in U_{I(z,\varepsilon)}} \left\| \sum_{j=0}^{m} u^j \nabla f^j(z) \right\|^2,$$

which is a very simple quadratic program. In fact, (3.46) is so simple that it will often be easier to solve than the linear program resulting from the use of the normalization set $S = \{h \mid |h^i| \leq 1,\ i = 1, 2, \ldots, n\}$. Thus, it appears that the most efficient version of (2.4 Z1) is the one which uses the normalization set $S = \{h \mid \|h\|^2 \leq 1\}$ and computes a direction $h(z,\varepsilon)$ by means of (3.46) and (3.45).

4. THE ZOUTENDIJK PROCEDURE 2 AND THE PIRONNEAU-POLAK METHODS.

In appendix A of [7], there is a counter example which shows that the Zoutendijk procedure 2, with $S = \{h \mid |h^j| \leq 1\}$ i.e., algorithm (2.4 Z2), does not converge linearly even under convexity assumptions such as (2.2). Hence this algorithm does not appear to be of particular interest. We have not been able to obtain any results for algorithm (2.4 Z2) with $S = \{h \mid \|h\|^2 \leq 1\}$.

The Pironneau-Polak algorithms (2.4 P1) and (2.4 P2) were derived from the algorithms (2.4 Z1) and (2.4 Z2) by replacing the normalization set $S = \{h \mid \|h\| \leq 1\}$ by the added term $\frac{1}{2}\|h\|^2$ in the cost of the direction finding subproblem. As we shall see, this modification results in linearly convergent algorithms. As we have already pointed out in Section 3, the algorithms (2.4 PP1) and (2.4 PP2) are insensitive to the number of variables in the problem and their direction finding subproblems are usually easy to solve.

The rate of convergence of algorithm (2.4 PP1) follows directly from the rate of convergence of a modified method of centers using the same direction finding subprocedure (2.6 PP1), but which computes the step length by minimizing a distance function along the given direction. The relevent result is given by theorem (3.20) in [7] and the following theorem is a straightforward corollary to it.

4.1 **Theorem:** Let $\{z_i\}_{i=0}^{\infty}$ be a sequence generated by algorithm (2.4 PP1) in solving (2.1), and suppose that assumption (2.2) is satisfied. Then given any $\alpha \in (0, 1)$, there exists an integer $i(\alpha)$ such that

4.2 $$f^0(z_{i+1}) - f^0(\hat{z}) \leq [1 - \frac{\ell\bar{\lambda}^{0^2}\alpha}{\tilde{M}}][f^0(z_i) - f^0(\hat{z})] \quad \text{for all } i \geq i(\alpha),$$

where ℓ, $\bar{\lambda}^0$ are as in (2.3) and (3.14), respectively, z is the solution of (2.1) and $\tilde{M} = \max \{\| \frac{\partial^2 f^j(z^2)}{\partial z^2} \| \,|\, j = 0, 1, \ldots, m\}$. ¤

Finally, we turn to algorithm (2.4 PP2).

4.3 **Theorem:** Let $\{z_i\}_{i=0}^{\infty}$ be a sequence generated by algorithm (2.4 PP2), with $q = 2$, and suppose that assumption (2.2) is satisfied. Then, given any $\alpha \in (0,1)$ there exists an integer $i(\alpha)$ such that for all $i \geq i(\alpha)$,

4.4 $$f^0(z_{i(\alpha)+j}) - f^0(\hat{z}) \leq [1 - \frac{\bar{\lambda}^{0^2}\ell\alpha}{\beta L'}]^j\,[f^0(z_{i(\alpha)}) - f^0(\hat{z})]$$

where ℓ, $\bar{\lambda}^0$ and β are as in (2.3), (3.14) and step 0 of (2.4 PP1), respectively, and $L' = \max \{L,M\}$, with L and M as in (3.27) and (3.32).

Proof: We begin by recalling that (2.6 PP2) can also be written as

4.5 $$h^0(z_i,\varepsilon_i) = \min\{\tfrac{1}{2}\|h\|^2 + k^0 | k^0 \geq \langle \nabla f^0(z_i), h \rangle ;$$

$$k^0 \geq f^j(z_i) + \langle \nabla f^j(z_i), h \rangle, \quad j \in I(z_i,\varepsilon_i)\}$$

Hence, if we define k^0 by

4.6 $$k^0(z_i,\varepsilon_i) = h^0(z_i,\varepsilon_i) - \tfrac{1}{2}\|h(z_i,\varepsilon_i)\|^2,$$

we must have

4.7 $$\langle \nabla f^0(z_i), h(z_i,\varepsilon_i) \rangle \leq k^0(z_i,\varepsilon_i),$$

4.8 $$f^j(z_i) + \langle \nabla f^j(z_i), h(z_i,\varepsilon_i) \rangle \leq k^0(z_i,\varepsilon_i), \; j \in I(z_i,\varepsilon_i).$$

We can now repeat the steps followed in deriving (3.33) from (3.31) to obtain that for every $\alpha \in (0,1)$, there exists an integer $i_2(\alpha)$ such that

4.9 $$f^0(z_{i+1}) - f^0(z_i) \leq \min_{0 \leq \mu \leq \frac{\alpha\varepsilon_i}{L'\|h(z_i,\varepsilon_i)\|}} \{\mu k^0(z_i,\varepsilon_i) + \frac{\mu^2}{2} M\|h(z_i,\varepsilon_i)\|^2\}$$

$$\leq \frac{\alpha\varepsilon_i k^0(z_i,\varepsilon_i)}{L'\|h(z_i,\varepsilon_i)\|} + \frac{\alpha^2\varepsilon_i^2 M}{2L'^2}$$

$$\leq \frac{\alpha\varepsilon_i}{L'}\left(\frac{k^0(z_i,\varepsilon_i)}{\|h(z_i,\varepsilon_i)\|} + \frac{\alpha\varepsilon_i}{2}\right).$$

Now applying the Wolfe strong duality theorem [10] to (4.5) we conclude that

$$4.10 \quad h^0(z_i,\varepsilon_i) = \sum_{j \in I(z_i,\varepsilon_i)} u_i^j f^j(z_i) - \frac{1}{2} \| \sum_{j \in I(z_i,\varepsilon_i) \cup \{0\}} u_i^j \nabla f^j(z_i) \|^2,$$

$$4.11 \quad h(z_i,\varepsilon_i) = - \sum_{j \in I(z_i,\varepsilon_i) \cup \{0\}} u_i^j \nabla f^j(z_i),$$

where $u_i = (u_i^0, u_i^1, \ldots, u_i^m)$ is a solution of

$$4.12 \quad \max_{u \in U_{I(z_i,\varepsilon_i)}} \{ \sum_{j=1}^{m} u^j f^j(z_i) - \frac{1}{2} \| \sum_{j=0}^{m} u^j \nabla f^j(z_i) \|^2 \}.$$

Now (4.6), (4.10), (4.11) and the fact that $\sum_{j \in I(z_i,\varepsilon_i)} u_i^j f^j(z_i) \le 0$ imply that

$$4.13 \quad k^0(z_i,\varepsilon_i) \le - \|h(z_i,\varepsilon_i)\|^2.$$

By construction of ε_i, $h^0(z_i,\varepsilon_i) \le - \varepsilon_i^2$ which implies that

$$4.14 \quad k^0(z_i,\varepsilon_i) \leq -\varepsilon_i^2 .$$

Making use of (4.13) and (4.14) we now obtain

$$4.15 \quad \frac{k^0(z_i,\varepsilon_i)^2}{\|h(z_i,\varepsilon_i)\|^2} \geq -k^0(z_i,\varepsilon_i) \geq \varepsilon_i^2 ,$$

which yields

$$4.16 \quad k^0(z_i,\varepsilon_i)/\|h(z_i,\varepsilon_i)\| \leq -\varepsilon_i .$$

Since $\alpha\varepsilon_i < \varepsilon_i$, (4.9) and (4.16) combine to give

$$4.17 \quad f^0(z_{i+1}) - f^0(z_i) \leq -\frac{\alpha\varepsilon_i^2}{2L'} \quad \text{for all } i \geq i_2(\alpha).$$

Now, it was shown in lemma 3-53 of [7] (which dealt with a modified method of centers using (2.6 PP1)) that, given any $\alpha \in (0,1)$ and a sequence $z_i \to \hat{z}$, there exists an integer $i_3(\alpha) \geq i_2(\alpha)$ such that

$$4.18 \quad h^0(z_i,\infty) \leq \ell\bar{\lambda}^{0^2}\alpha^2[f^0(\hat{z}) - f^0(z_i)],$$

where $h^0(z_i,\infty)$ is defined as in (2.4 PP2) with $\varepsilon = \infty$, i.e., with $I(z_i,\infty) = \{1, 2, \ldots, m\}$. Since for any $\varepsilon < \infty$, we must have $h^0(z_i,\varepsilon) \leq h^0(z_i,\infty)$, (4.18) implies that for any $\varepsilon \in [0,\infty)$,

$$4.19 \quad h^0(z_i,\varepsilon) \leq \ell\bar{\lambda}^{0^2}\alpha^2[f^0(\hat{z}) - f^0(z_i)] \quad \text{for all } i \geq i_3(\alpha).$$

Now, for $i = 0, 1, 2, \ldots$, let $k(i)$ be the integer satisfying (c.f. (3.37))

$$4.20 \quad \varepsilon_i = \varepsilon_{i-1} = \dots = \varepsilon_{i-k(i)} < \varepsilon_{i-k(i)-1}.$$

Then, by the rules of the algorithm, since $q = 2$, and $q_{i-k(i)} < \varepsilon_{i-k(i)-1}$, we must have

$$4.21 \quad h^0(z_i, \tfrac{1}{\beta}\varepsilon_{i-k(i)}) > -\tfrac{1}{\beta}\varepsilon_{i-k(i)}.$$

Next, from (4.17), by adding and subtracting terms, we obtain that for all $i \geq i_2(\alpha)$,

$$4.22 \quad f^0(z_{i+1}) - f^0(z_{i-k(i)}) \leq -\frac{\alpha}{2L'}(\varepsilon_i^2 + \varepsilon_{i-1}^2 + \dots + \varepsilon_{i-k(i)}^2)$$

$$= -\frac{\alpha}{2L'}(1 + k(i))\varepsilon_{i-k(i)}^2$$

Combining (4.22) with (4.21) and (4.18), we finally obtain that for all $i \geq i_3(\alpha)$

$$4.23 \quad f^0(z_{i+1}) - f^0(z_{i-k(i)}) \leq -\frac{\beta\ell\bar{\lambda}^{0\,2}\alpha^3}{2L'}(1 + k(i))[f^0(z_{i-k(i)}) - f^0(\hat{z})].$$

Rearranging terms in (4.23), we get

$$4.24 \quad f^0(z_{i+1}) - f^0(\hat{z}) \leq \left(1 - \frac{\beta\ell\bar{\lambda}^{0\,2}\alpha^3(1 + k(i))}{2L'}\right)$$

$$\times [f^0(z_{i-k(i)}) - f^0(\hat{z})] \text{ for all } i \geq i_3(\alpha).$$

Setting $i(\alpha) = i_3(\alpha)$ (4.4) can now be obtained from (4.24) in the same manner as we have obtained (3.25) from (3.39), which completes our proof. ¤

5. CONCLUSION

The results in this paper lead us to the following tentative rendering of the four algorithms considered. The Zoutendijk Procedure 2 appears to be the least attractive one because of its poor rate of convergence. The algorithm (2.4 PP1) which is a modification of the Zoutendijk procedure 2, has a good rate of convergence, but uses a direction finding subproblem of unnecessarily large dimension. Thus, the real choice is between the Zoutendijk Procedure 1 and our procedure (2.4 PP2). When the number of "active" constraints $\bar{m}$ is much smaller than the number of variables n ($\bar{m} \leq n/2$) and n is large, (2.4 PP2) is clearly superior because then is has a much better rate of convergence as well as a simpler direction finding subprocedure. When the number of "active" constraints increases, the better rate of convergence of (2.4 PP2) becomes offset by the increased number of operations needed to solve the direction finding subproblem. Although the exact break point is difficult to estimate, it can be shown that when $\bar{m} \geq n$, the Zoutendijk Procedure 1 is clearly superior because of its simpler direction finding subproblem, even though its rate may be quite bad compared to that of (2.4 PP2). To extend the usefulness of the method (2.4 PP2), it will be necessary to develop a special quadratic programming algorithm which exploits the structure of the problem (2.6 PP2). It is to be hoped that such an algorithm will be found.

REFERENCES

[1] BERGE, C., Topological Spaces, MacMillan, New York. 1963.

[2] CANON, M. and C. CULLUM, "A Tight Upper Bound on the Rate of Convergence of the Frank and Wolfe algorithm," SIAM J. on Control, Vol. 6, No. 4, 1968.

[3] FRANK, M., and P. WOLFE, "An Algorithm for Quadratic Programming," Naval Log, Quart., Vol. 3, pp. 95-110. 1956.

[4] LEMKE, C., Bimatrix Equilibrium points and Mathematical Programming, Management Science, 11, pp. 681-689. 1965.

[5] VON NEUMANN, J., Zur Theorie der Gesellshaftspiele, Mathematische Annalen, 100, pp. 295-320. 1928.

[6] ORTEGA, J. and W. RHEINBOLDT, Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York. 1970.

[7] PIRONNEAU, O. and E. POLAK, "On the Rate of Convergence of Certain Method of Centers," Memorandum No. ERL-M296, University of California, Berkeley 1971.

[8] POLAK, E., Computational Method in Optimization, Academic Press, New York, 1971.

[9] TOPKIS, D., and A. VEINNOTT, "On the Convergence of some Feasible Direction Algorithms for Nonlinear Programming," SIAM J. on Control, 5 (2), pp. 268-279. 1967.

[10] WOLFE, P., "A Duality Theorem for Nonlinear Programming," Quarterly of Applied Mathematics, 1961, pp. 239-244.

[11] ZOUTENDIJK, G., Method of Feasible Direction, Elsevier Publishing Co., Amsterdam, 1960.

[12] ZUKHOVISKII, S., R. POLYAK and M. PRIMAK, An Algorithm for the Solution of Convex Programming Problems, DAN USSR 153 (5), pp. 991-1000. 1963.