DISCUSSION PAPER NO. 261

EFFICIENT ACCELERATION TECHNIQUES FOR FIXED POINT ALGORITHMS

by

R. Saigal
and
M. J. Todd

December, 1976

Efficient Acceleration Techniques for Fixed Point Algorithms

R. Saigal[2]
M. J. Todd[1]

ABSTRACT

Recently, Saigal has presented an acceleration technique whereby the fixed point algorithms based on complementary pivoting can be made to converge quadratically. In this paper, we study the efficiency of this acceleration when, along with the fixed point steps, a series of Newton type steps are performed. We show that this efficiency is comparable to that of Shamanskii's method and that the "global" convergence properties of the original fixed point algorithm are not destroyed.

---

[1]Department of Operations Research, Cornell University, Ithaca, New York.
This author's work was partially supported by NSF Grant GK-42092.

[2]Supported by NSF Grant MCS-7703472.

Efficient Acceleration Techniques for Fixed Point Algorithms

R. Saigal
M. J. Todd

1. Introduction

Let R^n be the n-dimensional Euclidean space, and for a vector $x = (x_1, \ldots, x_n)$ in R^n, let $\|x\| = \max_i |x_i|$. Given a continuous function ℓ from R^n into R^n, in this paper we consider the recent fixed point algorithms pioneered by Scarf [17] and extended by Merrill [9] and Eaves and Saigal [6]. These algorithms have the capability to compute fixed points of point-to-set mappings as well. Under a variety of conditions on f, which include the conditions of the Brouwer's fixed point theorem [2] and the Leray-Schauder theorem [10, 6.3.3], these algorithms are successful in computing a fixed point. Appropriate generalizations to point-to-set mappings of the above fixed point theorems exist, and these algorithms are effective in computing fixed points of such mappings, see for example [6, Theorem 4.2].

In this paper, we consider the application of the algorithm of [6], which is called the continuous deformation method (see [14]), for computing fixed points or, equivalently, computing zeros of a differentiable mapping ℓ whose derivative, ℓ', is Lipschitz continuous. For such mappings, it is shown in [13] that the continuous deformation method can be accelerated to converge quadratically. In addition, as observed by Todd [20] and explicitly used in [13, 21], a finite difference approximation to the derivative ℓ' is readily available. The aim of this paper is to use this matrix in Newton type steps and, hopefully, make the acceleration scheme more "efficient" (according to the measures introduced by Ostrowski [11, 6.11] and Brent [1]). Thus, the aim of this paper is to study a hybrid algorithm in which, between consecutive acceleration steps of the fixed point algorithm, a series of m Newton-type steps are performed, using the readily available finite difference approximation to the derivative. The efficiency of our hybrid algorithm compares favorably with that for Shamanskii's method found by Brent [1, Table 1]. In addition, asymptotically, the acceleration step of the fixed point algorithm can be viewed as a version of the finite difference form of Newton's method.

In section 2, we present the Eaves-Saigal algorithm; in section 3, we present the acceleration scheme of Saigal [13] and its extensions; in section 4, we study its asymptotic convergence and the efficiency; and in section 5, we study its global convergence.

2. The Fixed Point Algorithm

In this section, we describe the Eaves-Saigal algorithm [6], [14]. In this method, an initial one-to-one affine mapping r and its zero x^1 are chosen. This mapping is then deformed in a piecewise linear manner to the mapping, ℓ. Also, starting with x^1, a piecewise linear path, x^t, is traced, such that, under certain general conditions on r and ℓ, this path converges to a zero of ℓ. This algorithm shares many similarities with the recent "globalization schemes" for Newton's method proposed by Kellogg, Li and Yorke [8] and Smale [18], where a similar path is traced via a solution to a differential equation.

The usual implementation of this algorithm is on the triangulation J_3 of $R^n \times (0,1]$. One creates a piecewise-linear homotopy, L, on $R^n \times [0,1]$, such that L restricted to $R^n \times \{1\}$ is r, and restricted to $R^n \times \{0\}$ is ℓ. The algorithm then traces a component of $L^{-1}(0)$, which contains $(x^1,1)$. We now give a brief overview of this method.

The triangulation, J_3, is a collection of (n + 1) - simplexes, which together with all their faces partition $R^n \times (0,1]$. The vertices of these simplexes lie in $R^n \times \{2^{-k}\}$ for some k = 0,1, . . . It follows that $\theta J_3 = \{\theta\sigma : \sigma \text{ in } J_3\}$ triangulates $R^n \times (0,\theta]$. Further, any simplex of θJ_3 in $R^n_x(0, 2^{-k}\theta]$ has diameter at most $2^{-k}\theta$. The aforementioned deformation and path tracing on J_3 can now be described as follows.

With r and x^1 as above, let $\alpha : R_+ \rightarrow R_+$ be a continuous increasing function with $\alpha(0) = 0$. For a vertex $(v, 2^{-k})$ in J_3, define

$$L(v, 2^{-k}) = \begin{cases} \ell(v) & \text{if } \alpha(\|x - x'\|) < k \\ r(v) & \text{if not} \end{cases}$$

Now, extend L linearly on each simplex σ of J_3, that is, if $(x,t) = \sum \lambda_i (v^i, t)$, $\sum \lambda_i = 1$ and $\lambda_i \geqq 0$ for each i, where (v^i, t) are vertices of σ, $L(x,t) = \sum \lambda_i \, L(v^i, t)$. If we further set $L(x,0) = \ell(x)$, then $L : R^n \times [0,1] \to R^n$ is a piecewise linear homotopy from ℓ to r.

The method then traces the unique component of $L^{-1}(0)$ which contains $(x^1, 1)$ in the following manner. Starting with the (usually) unique simplex σ_0 of J_3 containing $(x^1, 1)$, it traces the line $L^{-1}_{\sigma_0}(0)$ where L_σ is the restriction of the mapping L to a simplex σ of J_3. This line crosses some other facet of σ_0 (assumed unique by a non-degeneracy assumption) at some point (x', t'). This procedure is then continued by finding the other simplex σ' containing (x', t') and tracing the line $L^{-1}_{\sigma'}(0)$, which passes through (x', t'). Under certain general conditions on the mappings r and ℓ (see, for example, [6, theorem 4.2]), this piecewise linear map thus traced converges to a zero of ℓ. For more details of this method, see, for example, [15], [22].

3. The Accelerated Fixed Point Algorithm

In this section, we describe the acceleration technique of Saigal [13]. We assume that the mapping, ℓ, is continuously differentiable with Jacobean $\ell'(x) = (\partial \ell_i(x)/\partial x_j)$ satisfying a Lipschitz condition with constant K;

$$\|\ell'(x) - \ell'(y)\| \leqq K \|x - y\| . \tag{3.1}$$

For a n x n matrix A, let $\|A\| = \max_i \sum_j |a_{ij}|$. We could replace (3.1) with the assumption of Hölder continuity on R^n with straightforward changes in our results (as is shown in section 4).

Lemma 3.1: For all x, y ϵ R^n, define $e(y,x) = \ell(y) - \ell(x) - \ell'(x)(y - x)$. Then $\|e(y,x)\| \leqq \frac{1}{2} K \|y - x\|^2$.

Proof: See [10, 3.2.12].

We now describe this acceleration technique applied to the triangulation J_3 of $R^n \times (0,1]$. Let $x^1 \epsilon R^n$, $\epsilon_1 > 0$, and A_0 be a nonsingular matrix. Ideally, x^1 is close to a zero of ℓ, A_0 to the Jacobean $\ell'(x^1)$; ϵ^1 should have the same order as the distance from x^1 to a zero of ℓ.
Define $r^1(x) = A_0(x - x^1)$ for all $x \in R^n$. Also choose a continuous nondecreasing function $\alpha : R_+ \to R_+$ with $\alpha(0) = 0$ and $\alpha(1) < 1$. We frequently pick α identically zero. Finally, translate the triangulation $2\epsilon_1 J_3$ by $x^1 - \epsilon_1\left(\frac{2n-1}{2n}, \frac{2n-3}{2n}, \ldots, \frac{1}{2n}\right)^T$ to get $2\epsilon_1 J_3'$, say. The effect is to put (x^1, ϵ_1) in the center of a facet of $2\epsilon_1 J_3'$ and will be important for the acceleration techniques.

Now define $L : R^n \times (0, 2\epsilon_1] \to R^n$ as follows. If (v, t) is a vertex of $2\epsilon_1 J_3'$ with $t = 2\epsilon_1 \cdot 2^{-k}$, then set

$$L(v, t) = \begin{cases} r^1(v) & \text{if } \alpha\left(\frac{\|v - x^1\|}{2\epsilon_1}\right) \geq k \\ \ell(v) & \text{if not .} \end{cases}$$

If $\alpha \equiv 0$, this rule reduces to

$$L(v, t) = \begin{cases} r^1(v) & \text{if } t = 2\epsilon_1 \\ \ell(v) & \text{if } t < 2\epsilon_1 . \end{cases}$$

Create the piecewise linear homolopy by extending L linearly on each simplex of $2\,\epsilon_1\, J_3'$, and starting with the unique zero $(x^1, 2\,\epsilon_1)$ of L in $R^n \times \{2\,\epsilon_1\}$, trace the piecewise linear path in $L^{-1}(0)$. The algorithm thus generates a sequence of distinct facets of $2\epsilon_1 J_3'$ with each containing a zero of L and with the convex hull of each consecutive pair of facets being a simplex of $2\epsilon_1 J_3'$. Moving from one facet to the next requires one function evaluation and one linear programming pivot. Under suitable conditions on ℓ and A_0 (see section 5), the facets remain in a bounded region. In this case, the set of cluster points of the facets is the projection on $R^n \times \{0\}$ of a closed connected set of zeroes of ℓ.

When a facet τ of $2\epsilon_1 J_3'$ in $R^n \times \{2^{-k+1}\epsilon_1\}$ is generated for the first time, we say that a new level is penetrated. If the vertices of τ are (v^i, ϵ), $i = 0, 1, \ldots, n$, where $\epsilon = 2^{-k+1}\epsilon_1$, and if $L(v^i, \epsilon) = \ell(v^i)$ for all i (for example, if $\alpha \equiv 0$), then let $(\overline{x}, \epsilon)$ be the zero of L in τ. Then we have $\overline{x} = \sum \lambda_i v^i$ and $\sum \lambda_i \ell(v^i) = 0$ for $\sum \lambda_i = 1$ and all $\lambda_i \geq 0$. However, the fixed point algorithm has available also the matrix $B = \begin{bmatrix} 1 & \ldots & 1 \\ \ell(v^0) & \ldots & \ell(v^n) \end{bmatrix}$ and its inverse, and from these further information can be deduced. Since each $v^i - v^{i-1}$ is $\pm\,\epsilon$ times some unit vector, it is easy to obtain from B and B^{-1} the matrices A and A^{-1} where $\ell(v^i) = A(v^i - \overline{x})$ for all i. A is a finite difference approximation to the Jacobian of ℓ at $\overline{x}$ (see [13], [20].

<u>Lemma 3.2</u> Under assumption (3.1), $\|\ell(\overline{x})\| \leq \frac{1}{2} K \epsilon^2$ and $\|A - \ell'(\overline{x})\| \leq Kn\epsilon$.

Proof See [12], [13].

We summarize the information obtained from τ by writing $(\bar{x}, A) = FP(x^1, A_0, \epsilon_1, k)$.

Saigal [13] showed that the basic fixed-point algorithm could be accelerated by restarting with a much finer grid size. Let $\epsilon_2 = \theta\|A^{-1}\ell(\bar{x})\|$, where θ is a sufficiently large constant, say 3n. If $\epsilon_2 < \epsilon_1/2$, then the algorithm is restarted with $x^2 = \bar{x}$ replacing x^1, $A_1 = A$ replacing A_0 and ϵ_2 replacing ϵ_1. The accelerated algorithm proceeds in this way to generate $x^3, x^4, \ldots$ At each iteration, k is as small as possible so that $\epsilon_{i+1} < \epsilon/2$ as above. Suppose $x^i \to x^*$ with $\ell(x^*) = 0$ and $\ell'(x^*)$ nonsingular. Then Saigal showed that for large enough i, $(x^{i+1}, A_i) = FP(x^i, A_{i-1}, \epsilon_i, 1)$; that is, the algorithm accelerates with each level penetrated. Note that A_i is then a finite difference approximation to $\ell'(x^{i+1})$ based on a step size of ϵ_i, and x^{i+1} is an approximate zero of ℓ based on a simplex of diameter ϵ_i. Further, for large enough i, only n + 1 function evaluations and pivots are required to obtain x^{i+1} from x^i. (It is to guarantee this property that the triangulation $2\epsilon_i J_3$ is translated.) Lastly, $x^i \to x^*$ with Q-order at least 2, i.e., for some constant β, $\|x^{i+1} - x^*\| \le \beta\|x^i - x^*\|^2$, i = 1, 2,...

The accelerated version considered here builds on Saigal's method by interposing Newton-like steps between successive applications of the fixed-point algorithm. It can be described as follows:

<u>Accelerated Fixed Point Algorithm with Parameter $m \ge 1$: ($AFPA_m$)</u>

<u>Step 0</u> Choose $y^1 \in R^n$, $\epsilon_1 > 0$ and a nonsingular matrix A_0. Set i = 1.

<u>Step 1</u> Let $(z^i, A_i) = FP(y^i, A_{i-1}, \epsilon_i, k)$, where k = k(i) is as small as possible so that $\|A_i^{-1}\ell(z^i)\| < 2^{-k}\epsilon_i$.

Step 2 Set $z^{i,1} = z^i$, $s^{i,1} = A_i^{-1}\ell(z^{i,1})$ and $z^{i,2} = z^{i,1} - s^{i,1}$. If $m = 1$, set $j = 2$ and go to step 3. Otherwise, for $j = 2, 3, \ldots, m$:

Set $s^{i,j} = A_i^{-1}\ell(z^{i,j})$. If $\|s^{i,j}\| > \|s^{i,j-1}\|$, go to step 3. Otherwise, set $z^{i,j+1} = z^{i,j} - s^{i,j}$, and increase j by one.

Step 3 Set $y^{i+1} = z^{i,j}$ and $x^{i+1} = z^{i,j-1}$. Set $\varepsilon_{i+1} = \theta_i\|y^{i+1} - x^{i+1}\| = \theta_i\|A_i^{-1}\ell(x^{i+1})\|$, where $0 < \theta_i \leq 1$. (The choice of θ_i will be discussed below.) Increase i by one and return to step 1.

Note that step 1 contains the "fixed-point algorithm" step while step 2 contains "parallel chord" [10,7.1] or Newton-like steps. We have incorporated protection against these latter steps increasing in size so that $\varepsilon_{i+1} < \varepsilon_i/2$ for all i.

We first remark that $AFPA_1$ coincides with Saigal's method with one exception. When restarting the algorithm, we use the linear function $r^{i+1}(x) = A_i(x - y^{i+1})$ rather than $r^{i+1}(x) = A_i(x - x^{i+1})$, where $y^{i+1} = x^{i+1} - A_i^{-1}\ell(x^{i+1})$. (Note that in $AFPA_1$, $x^{i+1} = z^i$ is the same as in Saigal's algorithm.) Since $A_i^{-1}\ell(x^{i+1})$ must be computed in Saigal's method to determine ε_{i+1}, negligible extra work is involved in obtaining the (usually better) point y^{i+1} from x^{i+1}.

We defer the question of global convergence of the complete algorithm. For now we suppose that each of the fixed-point steps 1 converges. Henceforth, assume that there is a constant μ such that $\ell(w) = 0$ implies $\ell'(w)$ nonsingular with $\|\ell'(w)^{-1}\| < \mu$.

Lemma 3.3 Let $\left(z^{(k)}, A_{(k)}\right) = FP(y, A, \varepsilon, k)$ and assume $\{z^{(k)}\}$ remains bounded. Then for some k, $\|A_{(k)}^{-1}\ell(z^{(k)})\| < 2^{-k}\varepsilon$.

Proof Let w be a limit point of the $z^{(k)}$'s. Then Lemma 2.2 implies that $\ell(w) = 0$ and $\|\ell'(w)^{-1}\| < \mu$. For k sufficiently large, $z^{(k)}$ is close to some such limit point, and Lemmas 3.2 and 3.1 then imply that $A_{(k)}$ is nonsingular with $\|A_{(k)}^{-1}\| < 2\mu$ for sufficiently large k. Then $\|A_{(k)}^{-1}\ell(z^{(k)})\| \le \mu K(2^{-k+1}\varepsilon)^2$ and for sufficiently large k the condition holds.

Lemma 3.3 shows that the sequences z^i, x^i and y^i are well-defined if the fixed-point algorithm steps 1 converge.

<u>Theorem 3.4</u> Suppose that A is nonsingular and $\|A^{-1}\ell(y)\|/\varepsilon + \|A^{-1}\|\left(4\|A - \ell'(y)\| + 8K\varepsilon\right) < \frac{1}{2n}$. Then $(z, A') = FP(y, A, \varepsilon, 1)$ is obtained in exactly $n+1$ pivots and $\|z - y\| < \frac{\varepsilon}{2n}$.

<u>Proof</u> Let $\tau = \langle (v^0, t_0), \ldots, (v^n, t_n) \rangle$ be a facet encountered in the fixed-point algorithm, and let $f(v^i) = L(v^i, t_i)$. For some $\lambda_i \ge 0$, $\sum \lambda_i = 1$, we have $\sum \lambda_i f(v^i) = 0$; let $x = \sum \lambda_i v^i$. Then $x - y = A^{-1}\left(A(x - y)\right) = A^{-1} \sum \lambda_i r(v^i) = A^{-1} \sum \lambda_i \left(r(v^i) - f(v^i)\right)$, where $r(v^i) = A(v^i - y)$.

Now $r(v^i) - f(v^i)$ is either zero or $r(v^i) - \ell(v^i) = -\ell(y) + \left(A - \ell'(y)\right) \cdot (v^i - y) - e(v^i, y)$. Hence, $\|x - y\| \le \|A^{-1}\ell(y)\| + \|A^{-1}\|\|A - \ell'(y)\|\max_i\|v^i - y\| + \|A^{-1}\| \frac{1}{2}K(\max_i\|v^i - y\|)^2$.

Suppose all $\|v^i - y\| \le 4\varepsilon$. Then by the condition of the theorem, $\|x - y\| < \frac{\varepsilon}{2n}$. It follows from the construction of J_3 [19] and the fact that $2\varepsilon J_3$ has been suitably translated that only $n+1$ simplices contain a point (x, t) with $\|x - y\| < \frac{\varepsilon}{2n}$, and all of them have all vertices within 2ε of y. Since no simplex with vertices more than 4ε from y can be generated without passing through a simplex with all vertices within 4ε of y, only these $n+1$ simplices are encountered. Taking τ to be the final facet, we have $x = z$ and $\|z - y\| < \frac{\varepsilon}{2n}$.

<u>Lemma 3.5</u> Suppose A is nonsingular and $y = x - A^{-1}\ell(x) \ne x$. Then

$$\|A^{-1}\ell(y)\|/\|A^{-1}\ell(x)\| \le \|A^{-1}\|\left(\|A - \ell'(x)\| + \tfrac{1}{2}K\|A^{-1}\ell(x)\|\right).$$

Proof With $e(y, x)$ as in Lemma 3.1, we have

$$\ell(y) = \ell(x) + \ell'(x)(y - x) + e(y, x) = \left(A - \ell'(x)\right)A^{-1}\ell(x) + e(y, x) .$$

Hence, using Lemma 3.1,

$$\|A^{-1}\ell(y)\| \le \|A^{-1}\|\|A - \ell'(x)\|\|A^{-1}\ell(x)\| + \tfrac{1}{2}K\|A^{-1}\|\|A^{-1}\ell(x)\|^2$$

from which the result follows.

Theorem 3.6 If all the fixed-point algorithms in step 1 converge, then the sequences $\{z^i\}$, $\{x^i\}$, $\{y^i\}$ are well-defined. Now suppose $\{z^i\}$ is bounded and for some λ, $\theta_i \ge -\lambda\epsilon_i \log \epsilon_i$ for all i. Then for all sufficiently large i, $k(i) = 1$, z^i is obtained in $n+1$ pivots from y^i, and all m parallel chord steps in step 2 are performed. Further, $x^i \to x^*$, $y^i \to x^*$ and $z^i \to x^*$ for some zero x^* of ℓ.

Proof Lemma 3.3 shows that the sequences are well-defined. Now $\epsilon_{i+1} < \frac{\epsilon_i}{2}$ for all i, so $\epsilon_i \to 0$. Lemma 3.2 then implies that any limit point of the sequence $\{z^i\}$ is a zero of ℓ. For sufficiently large i, z^i will be close enough to a zero of ℓ and A_i close enough to $\ell'(z^i)$ that $\|A_i^{-1}\| < 2\mu$; suppose this is true for $i \ge i^*$.

Since $\epsilon_i \to 0$, Lemma 3.2 shows that for sufficiently large $i \ge i^*$, $k(i) = 1$. Now $\|A_i - \ell'(z^i)\| \le Kn\epsilon_i$ and for $j = 1, 2, \ldots, m+1$, $\|z^{i,j} - z^i\| \le m\epsilon_i$; hence using (3.1), $\|A_i - \ell'(z^{i,j})\| \le K(m+n)\epsilon_i$. Now apply Lemma 3.5 inductively when $i \ge i^*$ and $2\mu\left(K(m+n)\epsilon_i + \frac{1}{2}K\epsilon_i\right) < 1$ to show that $\|A_i^{-1}\ell(z^{i,j})\|$ is decreasing in j for sufficiently large $i \ge i^*$. Thus all m parallel chord steps are eventually performed.

Lemma 3.5 also gives $\|A_i^{-1}\ell(y^{i+1})\| / \|A_i^{-1}\ell(x^{i+1})\| \le 2\mu\left(K(m+n)\epsilon_i + \frac{1}{2}K\epsilon_i\right)$. Hence if $\epsilon_{i+1} = \theta_i\|A_i^{-1}\ell(x^{i+1})\| > -\lambda\epsilon_i \log \epsilon_i\|A_i^{-1}\ell(x^{i+1})\|$, we obtain

$\|A_i^{-1}\ell(y^{i+1})\|/\epsilon_{i+1} \le -2\mu K\lambda^{-1}(m+n+\frac{1}{2})/\log \epsilon_i \to 0$. Theorem 3.4 then implies that for sufficiently large $i \ge i^*$, $\|z^i - y^i\| < \frac{\epsilon_i}{2n}$ and z^i is obtained in $n+1$ pivots from y^i.

Now since $\|z^i - y^i\| < \frac{\epsilon_i}{2n}$ and $\|y^{i+1} - z^i\| < m\epsilon_i$ we get $\|y^{i+1} - y^i\| < (m+1)\epsilon_i$ for sufficiently large $i \ge i^*$. With $\epsilon_{i+1} < \frac{\epsilon_i}{2}$, we find $\|y^j - y^i\| < 2(m+1)\epsilon_i$ for all $j > i$. Hence $\{y^i\}$ is a Cauchy sequence with limit x^*. Since $\|z^i - y^i\| < \frac{\epsilon_i}{2n}$ and $\|x^i - y^i\| < \epsilon_{i-1}$ for sufficiently large $i \ge i^*$, $x^i \to x^*$ and $z^i \to x^*$ also. Hence x^* is a zero of ℓ and the theorem is proved.

4. Asymptotic Convergence of $AFPA_m$

The statement of our accelerated fixed-point algorithm in section 2 is not conducive to an analysis of its convergence rate. However, Theorem 3.6 shows that for large enough i all parallel chord steps are performed and z^i is obtained from y^i in exactly $n+1$ pivots. It is thus possible to describe how z^i is obtained from y^i without reference to the fixed-point algorithm; z^i is the zero of an approximation to ℓ based on $n+1$ specified points. Below we state an asymptotic form of the algorithm and show that it coincides with $AFPA_m$ for sufficiently large i. We then analyze the convergence rate of $AFPA_m$ in this asymptotic algorithm, and finally discuss the efficiency of $AFPA_m$ in the spirit of Brent [1]. From this discussion we deduce good values of m.

Asymptotic Algorithm with Parameter $m \geq 1$ (AA_m)

Step 0 Choose $y^{i'} \in R^n$ and $\varepsilon_{i'} > 0$, and let $w = \left(\frac{2n-1}{2n}, \frac{2n-3}{2n}, \ldots, \frac{1}{2n}\right)^T \in R^n$. Set $i = i'$.

Step 1 Set $v^{i,0} = y^i - \varepsilon_i w$ and $v^{i,q} = v^{i,q-1} + \varepsilon_i e^q$ for $q = 1, 2, \ldots, n$, where e^q is the qth unit vector in R^n. Define A_i by

$$A_i e^q = [\ell(v^{i,q}) - \ell(v^{i,q-1})]/\varepsilon_i \quad \text{for } q = 1, 2, \ldots, n .$$

Step 2 Let $z^{i,0} = v^{i,0}$. For $j = 0, 1, \ldots, m$, set $z^{i,j+1} = z^{i,j} - A_i^{-1}\ell(z^{i,j})$.

Step 3 Set $x^{i+1} = z^{i,m}$, $y^{i+1} = z^{i,m+1}$, and $\varepsilon_{i+1} = \theta_i \|y^{i+1} - x^{i+1}\| = \theta_i \|A_i^{-1}\ell(x^{i+1})\|$, where $0 < \theta_i \leq \theta$. Stop if ε_{i+1} is within desired bounds; otherwise increase i by one and return to step 1. (The choice of θ_i will be discussed below.)

Note that we start AA_m with $y^{i'}$ and $\epsilon_{i'}$, with i' possibly greater than one. We now claim that for sufficiently large i', if $AFPA_m$ generates $\{x^i\}$ and $\{y^i\}$ and AA_m is started with $y^{i'}$, $\epsilon_{i'}$, then AA_m also generates $\{x^i\}$ and $\{y^i\}$ for $i > i'$. Indeed, choose i' sufficiently large so that the conclusions of Theorem 3.6 hold. Then z^i is the zero of the approximation to ℓ based on the facet of $2\epsilon_i J_3'$ containing (y^i, ϵ_i). But this facet is just $\langle (v^{i,0}, \epsilon_i), \ldots, (v^{i,n}, \epsilon_i) \rangle$. Hence the A_i of $AFPA_m$ is exactly the A_i of AA_m. Also, $z^i = v^{i,0} - A_i^{-1}\ell(v^{i,0})$, so that the $z^{i,1}$ of $AFPA_m$ is the $z^{i,1}$ of AA_m. Hence the algorithms produce identical sequences.

We first state a version of Lemma 3.2.

<u>Lemma 4.1</u> Under assumption (3.1), $\|A_i - \ell'(y^i)\| \leq Kn\epsilon_i$.

<u>Proof</u> From the definition of A_i, we have

$$[A_i - \ell'(y^i)]e^q = \Big(\Big[\ell(v^{i,q}) - \ell(y^i) - \ell'(y^i)(v^{i,q} - y^i)\Big] - \Big[\ell(v^{i,q-1}) - \ell(y^i) - \ell'(y^i)(v^{i,q-1} - y^i)\Big]\Big)\Big/\epsilon_i .$$

Using the triangle inequality and Lemma 3.1, $\|[A_i - \ell'(y^i)]e^q\| \leq K\epsilon_i$ and the result follows.

Now we suppose $x^i \to x^*$ and $y^i \to x^*$ with $\ell(x^*) = 0$ and $\ell'(x^*)$ nonsingular. Define $\beta = K(\frac{5}{2} + 2n)\|\ell'(x^*)^{-1}\|$, and let $\epsilon < \frac{1}{\beta(2\theta + 1)} < \beta^{-1}$. Our approach follows that of Brent [1].

<u>Lemma 4.2</u> For all $i > i'$, let $\gamma_i = \|x^i - x^*\|$, $\delta_i = \|y^i - x^*\|$ and $\eta_i = \delta_i + \epsilon_i$. Suppose $\eta_i < \epsilon$. Then for $j = 0, 1, \ldots, m$, $z^{i,j}$ is defined and

$$\|z^{i,j} - x^*\| \le \eta_i \tag{4.1}$$

and

$$\|z^{i,j+1} - x^*\| \le \beta\eta_i \|z^{i,j} - x^*\| . \tag{4.2}$$

Further, $\gamma_{i+1} \le (\beta\eta_i)^m \eta_i$, $\delta_{i+1} \le (\beta\eta_i)^{m+1}\eta_i$, $\epsilon_{i+1} \le 2\theta_i(\beta\eta_i)^m\eta_i$ and $\eta_{i+1} \le (2\theta_i + \beta\eta_i)(\beta\eta_i)^m\eta_i$.

Proof From (3.1), $\|\ell'(y^i) - \ell'(x^*)\| \le K\delta_i \le K\eta_i$ and from Lemma 4.1, $\|A_i - \ell'(y^i)\| \le Kn\epsilon_i \le Kn\eta_i$, so

$$\|A_i - \ell'(x^*)\| \le (n+1)K\eta_i .$$

Using $\beta\eta_i \le \beta\epsilon < 1$ and the Banach Perturbation Lemma [10, 2.3.2], A_i is nonsingular and

$$\|A_i^{-1}\| \le \frac{\frac{5}{2} + 2n}{\frac{3}{2} + n} \|\ell'(x^*)^{-1}\| .$$

Hence $z^{i,j}$ exists for all j. Now suppose (4.1) holds for some j. Then

$$\begin{aligned}
\|A_i(z^{i,j+1} - x^*)\| &= \|A_i(z^{i,j} - x^*) - \ell(z^{i,j})\| \\
&\le \|\ell'(x^*)(z^{i,j} - x^*) - \ell(z^{i,j})\| + \|A_i - \ell'(x^*)\|\|z^{i,j} - x^*\| \\
&\le \tfrac{1}{2}K\|z^{i,j} - x^*\|^2 + (n+1)K\eta_i\|z^{i,j} - x^*\| \\
&\le (\tfrac{3}{2} + n)K\eta_i\|z^{i,j} - x^*\| .
\end{aligned}$$

Hence $\|z^{i,\ +1} - x^*\| \le \|A_i^{-1}\|(\frac{3}{2} + n)K\eta_i\|z^{i,j} - x^*\| \le \beta\eta_i\|z^{i,j} - x^*\|$; that is, (4.2) holds for j. Since $\beta\eta_i < \beta\epsilon < 1$, (4.1) then holds for $j+1$. Now (4.1) holds for $j = 0$, establishing (4.1) and (4.2) for all j by induction.

The bounds on γ_{i+1} and δ_{i+1} follow immediately. Also, $\epsilon_{i+1} = \theta_i \|y^{i+1} - x^{i+1}\| \le \theta_i(\gamma_{i+1} + \delta_{i+1})$ and $\eta_{i+1} = \delta_{i+1} + \epsilon_{i+1}$, so that the proof is complete.

Theorem 4.3 Suppose $\|y^{i'} - x^*\| + \epsilon_i < \epsilon$. The algorithm AA_m is well-defined and generates sequences $\{x^i\}$, $\{y^i\}$ converging to x^*. Further, $x^i \to x^*$ with Q-order at least $m+1$, that is, for some constant ρ, $\|x^{i+1} - x^*\| \le \rho\|x^i - x^*\|^{m+1}$, $i = i', i'+1, \ldots$ If we choose $\theta_i = -\lambda\epsilon_i \log\epsilon_i$, $y^i \to x^*$ with R-order at least $m+2$; that is, if $p < m+2$, $\overline{\lim}\ \|y^{i'+i} - x^*\|^{1/p^i} = 0$.

Proof The first part follows from Lemma 4.2, since $\eta_{i+1} \le (2\theta_i + \beta\eta_i) \cdot (\beta\eta_i)^m\eta_i \le (2\theta + 1)\beta\epsilon\eta_i$ and so $\eta_i \to 0$. Now $\eta_i = \delta_i + \epsilon_i \le \gamma_i + 2\theta_i\gamma_i \le (2\theta + 1)\gamma_i$. Hence Lemma 4.2 gives $\gamma_{i+1} \le \beta^m(2\theta + 1)^{m+1}\gamma_i^{m+1}$, and $x^i \to x^*$ with Q-order at least $m+1$. Now let $p = m+2-\zeta$, $\zeta > 0$ and note that $\eta_{i+1} \le (2\theta_i + \beta\eta_i)(\beta\eta_i)^m\eta_i \le (\beta^{m+1}\eta_i^\zeta - 2\lambda\beta^m\eta_i^\zeta \log\eta_i)\eta_i^p$, since $\theta_i \le -\lambda\eta_i \log\eta_i$. Now the term in parentheses tends to 0 for any $\zeta > 0$ as $\eta_i \to 0$. Hence, $\eta_i \to 0$ with R-order at least $m+2$. Since $\delta_i \le \eta_i$, $\delta_i \to 0$ with R-order at least $m+2$.

Suppose now that the Jacobian of ℓ is Hölder continuous, i.e., that (3.1) is replaced by $\|\ell'(x) - \ell'(y)\| \le K\|x - y\|^p$ for some $p \in (0, 1]$. Then in Lemma 3.1 we have $\|e(y, x)\| \le (p+1)^{-1}K\|x - y\|^{p+1}$, and in Lemma 3.2, $\|\ell(\bar{x})\| \le (p+1)^{-1}K\epsilon^{p+1}$ and $\|A - \ell'(\bar{x})\| \le 2(p+1)^{-1}Kn\epsilon^p$. Lemmas 3.3 - 3.5 are unchanged, while in Theorem 3.6 we must require $\theta_i \ge \lambda(-\epsilon_i \log\epsilon_i)^p$ for all i. In Lemma 4.2, we obtain $\|A_i - \ell'(y^i)\| \le 2(p+1)^{-1}Kn\epsilon^p$. For Lemma 4.2 we let $\beta = 3K(n+1)\|\ell'(x^*)^{-1}\|$ and $\epsilon < [\beta(2\theta + 1)]^{-1/p} < \beta^{-1/p}$. Then Lemma 4.2 is valid with $(\beta\eta_i)^p$ replacing $\beta\eta_i$ everywhere. In Theorem 4.3, $x^i \to x^*$ with Q-order at least $mp+1$. If $\theta_i = \lambda(-\epsilon_i \log\epsilon_i)^p$ for all i, then $y^i \to x^*$ with R-order at least $mp+p+1$. The proofs are almost identical.

Instead of keeping A_i constant in step 2 of the algorithms, we could update it using a quasi-Newton update. For the Broyden update [3], [4], the bounded deterioration of A_i as shown in (3.5) and (4.8) of [3] is sufficient to keep our results valid. Quasi-Newton updates allow a superlinear rate of convergence to be attained without resorting to step 1 [3]; one may therefore wish to continue quasi-Newton steps until they deteriorate, rather than setting an a priori bound of m.

We now discuss the choice of the parameter m. We choose m to maximize the efficiency of the sequence $\{y^i\}$; as defined by Brent, the efficiency of the sequence w^i converging to w^* with R-order ρ is $k^{-1} \log \rho$ where k function evaluations are needed to compute w^{i+1} from w^i. We will use the lower bounds on the R-order of convergence of $\{y^i\}$ and $\{x^i\}$ rather than their true unknown values. Then the efficiency of $\{x^i\}$ is $(n+m+1)^{-1} \log(m+1)$ and that of $\{y^i\}$ is $(n+m+1)^{-1} \log(m+2)$. Brent's Table I [1] gives optimal values of $m+1$ for $\{y^i\}$ for various n; for example m = 6 for n = 10, 21 for n = 50 and 36 for n = 100.

5. Global Convergence

The algorithm $AFPA_m$ as stated in section 2 is probably the way one would implement such an accelerated fixed-point method. However, such an implementation possibly sacrifices one of the main advantages of fixed-point techniques; that is, under suitable conditions on ℓ, one obtains global convergence. The problem is that, although A_0 can be chosen so that z^1 is found in a finite number of steps, the change in artifical map to $r^2(x) = A_1(x - y^1)$ may permit the algorithm to diverge, so that z^2 is not found.

In this section we show how the algorithm can be modified so that global convergence is maintained. The basic idea is to change the map to $A_i(x - y^i)$ only locally. We consider the algorithm as being a special case of the Eaves-Saigal algorithm [6] with a modified triangulation T and labelling rule L defined on the set T^0 of vertices of T. While T and L depend on the behavior of the algorithm, this does not invalidate the argument.

For simplicity we assume $2\varepsilon_1 = 1$. The triangulation T is then the image of J_3 under a "piecewise shearing" homeomorphism h of $R^n \times (0, 1]$ onto itself. The function h is defined by a sequence $\{a^k\}_0^\infty$ in R^n. If $2^{-k} \geq t \geq 2^{-k-1}$ for some integer $k \geq 0$, then $h(x, t) = (x + a, t)$, where $a = (2^{k+1}t - 1)a^k + (2 - 2^{k+1}t)a^{k+1}$. In other words, $R^n \times \{2^{-k}\}$ is translated by $(a^k, 0)$ for each integer $k \geq 0$, and the layers between are mapped by a linear interpolation.

The sequence $\{a^k\}$ will be chosen so that $\|a^k - a^{k+1}\| \leq (m+1)2^{-k}$ for all k. It follows that if σ is a simplex of T lying in $R^n \times (0, t]$, then diam $p(\sigma) \leq (m+2)t$, where p projects $R^n \times (0, 1]$ onto R^n.

We now choose a nondecreasing function $\alpha : R_+ \to R_+$ with $\alpha(0) = 0$, $\alpha(1) < 1$, and $\alpha(s) \to \infty$ as $s \to \infty$. The labelling $L : T^0 \to R^n$ depends on ℓ, r^1, α and the past behavior of the algorithm. Let $L_k(v) = L(v, 2^{-k})$ for all k. We must now describe how the sequence $\{a^k\}$ and the labels L_k are recursively defined.

To start, define $L_0(v) = r^1(v) = A_0(r - y^1)$ and $a^0 = y^1 - \frac{1}{2}\left(\frac{2n-1}{2n}, \ldots, \frac{1}{2n}\right)^T$. Now define $a^1 = a^0$ and

$$L_1(v) = \begin{cases} r^1(v) & \text{for } \alpha(\|v - y^1\|) \geq 1 \\ \ell(v) & \text{if not.} \end{cases}$$

Suppose $a^0, \ldots, a^k$ and $L_0, \ldots, L_k$ have been defined in such a way that $\|a^j - a^{j+1}\| \leq (m+2)2^{-j}$ for $0 \leq j < k$ and that $L_j(v) = r^1(v)$ for $0 \leq j \leq k$ and $\|v - y^1\|$ sufficiently large. These properties certainly hold for $k = 1$. We now define $G : R^n \times [2^{-k}, 1] \to R^n$ to agree with L on T^0 and to be linear on each simplex of T. Then the Eaves-Saigal algorithm constructs a piecewise linear path of zeroes of G. Since all labels are r^1 outside a compact region and since the projected diameter of any simplex of T is at most $m+2$, this path cannot diverge. Hence it hits $R^n \times \{2^{-k}\}$ for the first time at a point z^k, say, lying in a simplex τ_k. (Note that this z^k is not the same as the z^k of $AFPA_m$.)

If some vertex of τ_k is not labelled according to ℓ, then we set $a^{k+1} = a^k$ and define

$$L_{k+1}(v) = \begin{cases} r^1(v) & \text{if } \alpha(\|v - y^1\|) \geq k+1 \\ \ell(v) & \text{if not.} \end{cases}$$

Otherwise we compute A_k so that $A_k(v - z^k) = \ell(v)$ for vertices v of τ_k and then $\|A_k^{-1}\ell(z^k)\|$. If this quantity is at least 2^{-k-1}, then define a^{k+1} and L_{k+1} as above.

If $\|A_k^{-1}\ell(z^k)\| < 2^{-k-1}$, we compute x^{k+1} and y^{k+1} as in steps 2 and 3 of $AFPA_m$. Note that $\|y^{k+1} - z^k\| < m2^{-k-1}$. Now pick $0 < \theta_i \leq \frac{1}{2}$ so that $\theta_i\|y^{k+1} - x^{k+1}\| = 2^{-j-1}$, $j > k$.

Choose $\hat{z}^k$ so that $(\hat{z}^k - a^k, 2^{-k})$ and $(z^k - a^k, 2^{-k})$ lie in the same facet of J_3 and so that $(\hat{z}^k - a^k, 2^{-j-1})$ lies in the center of a facet of J_3. For example, if the vertices of the facet containing $(z^k - a^k, 2^{-k})$ are $(y^i, 2^{-k})$, $0 \leq i \leq n$, in natural order, choose $\hat{z}^k - a^k = y^0\left(1 + \frac{4n-1}{2n}\beta\right) + \sum_{i=1}^{n-1} \frac{\beta}{n} y^i + \frac{2n+1}{2n}\beta\, y^n$, with $\beta = 2^{k-j-1}$. Then define $a^{k+1} = \ldots = a^{j+1} = a^k + y^{k+1} - \hat{z}^k$. Note that $\|a^{k+1} - a^k\| = \|y^{k+1} - z^k + z^k - \hat{z}^k\| \leq (m+1)2^{-k}$.

Now we define $L_{k+1}, \ldots, L_{j+1}$. For $\|v - z^k\| \leq (m+2)2^{-k}$, $L_{k+1}(v) = \ldots = L_j(v) = A_k(v - y^{k+1})$. Otherwise, for each $q = k+1, \ldots, j+1$,

$$L_q(v) = \begin{cases} r^1(v) & \text{if } \alpha(\|v - y^1\|) \geqq q \\ \ell(v) & \text{if not.} \end{cases}$$

The sequences $a^0, \ldots, a^{j+1}$, $L_0, \ldots, L_{j+1}$ satisfy the above properties. Hence the inductive definition is complete. See also Figure 1.

Note that part of the path $G^{-1}(0)$ can be predicted. In the notation above, every simplex of T meeting the line segment joining $(z^k, 2^{-k})$ and $(y^{k+1}, 2^{-k-1})$ has all vertices labelled according to $A_k(v - z^k)$ or $A_k(v - y^{k+1})$. Hence this line segment lies in $G^{-1}(0)$, as does the line segment from $(y^{k+1}, 2^{-k-1})$ to $(y^{k+1}, 2^{-j})$. The algorithm can therefore be reinitiated at $(y^{k+1}, 2^{-j})$. If the layer $R^n \times [2^{-j}, 2^{-k}]$ is never encountered again, the algorithm coincides with $AFPA_m$. However, it is possible (when the algorithm accelerates when it shouldn't) that this layer is encountered again.

We now give conditions under which the algorithm converges. Note that, because of our use of α and r^1, each z^k exists. We must show that $\{z^k\}$ remains in a bounded set under appropriate conditions. One such condition is the following:

There exist $\delta > 0$, $w \in R^n$, $C \in L(R^n)$ and compact $S \subseteq R^n$ such that (i) $x \notin S$, $\|y - x\| \le \delta$ imply $\ell(x)^T C(y - w) > 0$; (ii) $A_0^T C$ is positive definite.

Merrill [9] has shown that many applications satisfy this condition with $C = -I$.

<u>Theorem 5.1</u> If the condition above holds, all z^k lie in a bounded set.

<u>Proof</u> We first find j so that $2^{-j} \le \delta$. Then z^k for $k \le j$ lie in a bounded set. Suppose $k \ge j$. Then $(z^k, 2^{-k})$ lies in a simplex τ_k of projected diameter at most δ. Let the vertices of τ_k be (v^i, t_i), $i = 0, \dots, n$. Let $\|v^0 - w\| = \rho$; we obtain a bound on ρ independent of k. Let $\eta = \min\{x^T A_0^T C x \mid \|x\| = 1\} > 0$.

Now 0 is in the convex hull of $L(v^i, t_i)$, $i = 0, \dots, n$. Suppose first that each $L(v^i, t_i)$ is $\ell(v^i)$ or $r^1(v^i)$. We show that, if ρ is too large, $L(v^i, t_i)^T C(v^0 - w) > 0$ for all i. Indeed, if ρ is so large that v^i does not lie in S, then the inequality holds by the assumed condition if $L(v^i, t_i) = \ell(v^i)$. Suppose $L(v^i, t_i) = r^1(v^i) = A_0(v^i - y^1)$. Then

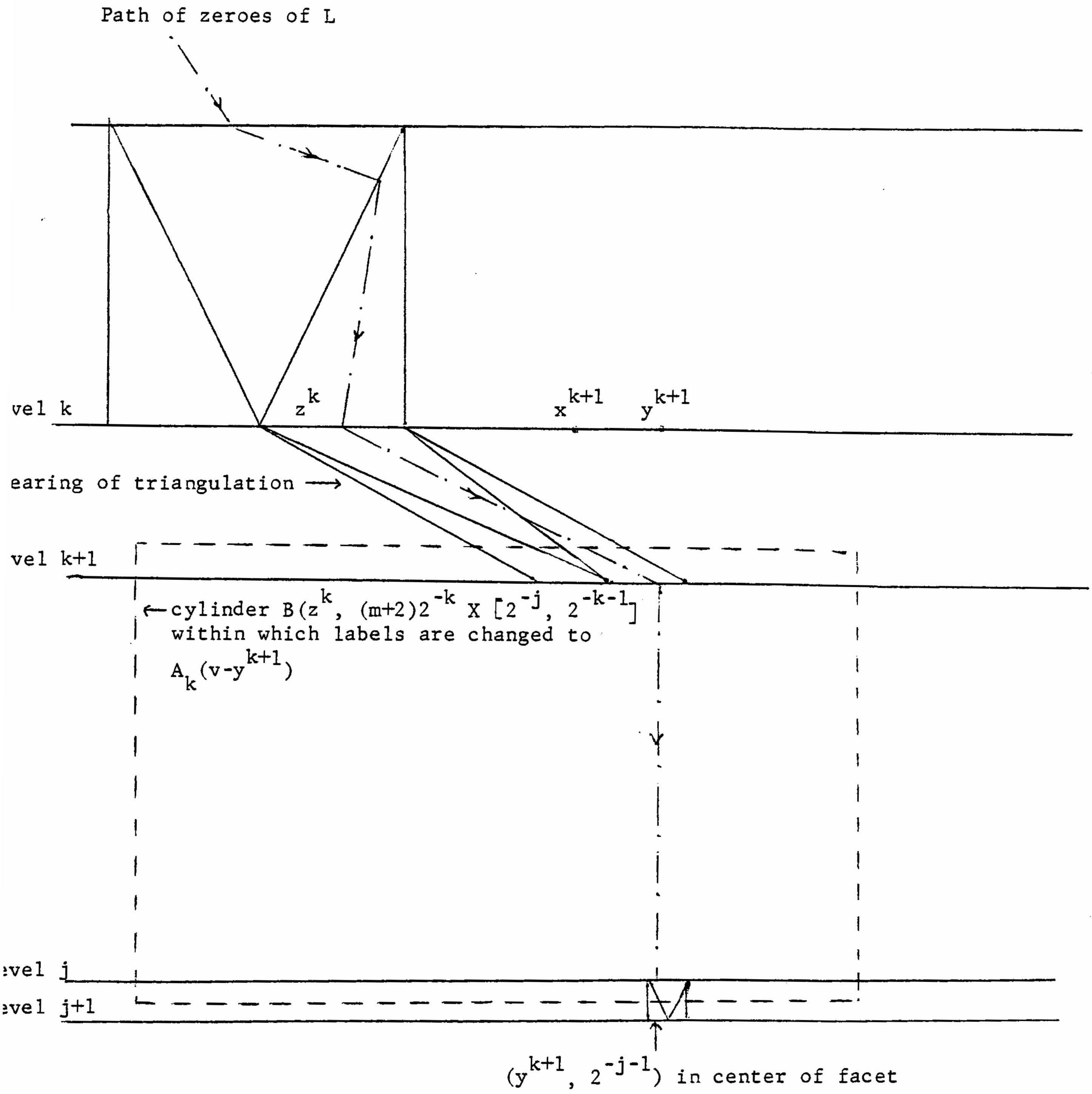
$$L(v^i, t_i)^T C(v^0 - w) = (v^i - y^1)^T A_0^T C(v^0 - w)$$

$$= (v^0 - w) A_0^T C(v^0 - w) + (v^i - v^0) A_0^T C(v^0 - w) + (w - y^1) A_0^T C(v^0 - w)$$

$$\ge \eta\rho^2 - \|A_0^T C\| \rho (m + 2 + \|w - y^1\|).$$

Hence if ρ is too large, $L(v^i, t_i)^T C(v^0 - w) > 0$, and τ_k does not contain a zero of G.

Suppose now some $L(v^i, t_i)$ is neither $\ell(v^i)$ nor $r^1(v^i)$. Then by the construction of L, v^i is within a distance of $m+2$ of some z^q lying in a simplex whose every vertex is labelled by ℓ. Since we have shown above that all such z^q lie in a bounded set, the same is true for z^k. The theorem is proved.



Acceleration Step in Globally Convergent Algorithm

Figure 1

References

1. Brent, R. P., "Some Efficient Algorithms for Solving Systems of Nonlinear Equations," SIAM J. Numer. Anal., 10 (1973), 327-344.

2. Brouwer, L. E. J., "Über eineindentige, stetige Transformation von Flächen in Sich," Mathematische Annalen, 67 (1910), 176-180.

3. Broyden, G. C., J. E. Dennis and J. J. Moré, "On the Local and Superlinear Convergence of Quasi-Newton Methods," J. Inst. Math. Applics., 12 (1973), 223-245.

4. Dennis, J. E. and J. J. Moré, "Quasi-Newton Methods, Motivation and Theory," SIAM Review, 19 (1977), 46-89.

5. Eaves, B. C., "Homotopies for Computing Fixed Points," Mathematical Programming, 3 (1972), 1-22.

6. Eaves, B. C. and R. Saigal, "Homotopies for Computing Fixed Points on Unbounded Regions," Mathematical Programming, 3 (1972), 225-237.

7. Eaves, B. C. and H. Scarf, "The Solution of Systems of Piecewise Linear Equations," Math. of O.R., 1 (1976), 1-27.

8. Kellogg, R. B., T. Y. Li and J. Yorke, "A Constructive Proof of the Brouwer Fixed Point Theorem," SIAM J. Numerical Analysis, 13 (1976), 473-483.

9. Merrill, O. H., "Applications and Extensions of an Algorithm that Computes Fixed Points of Certain Upper Semi-Continuous Point to Set Mappings," Ph.D. Dissertation, University of Michigan, Ann Arbor, Michigan, 1972.

10. Ortega, J. M. and W. C. Rheinboldt, Iterative Solutions of Nonlinear Equations in Several Variables, New York: Academic Press, 1970.

11. Ostrowski, A. M., Solution of Equations and Systems of Equations, New York: Academic Press, 1960.

12. Saigal, R., "Investigations into the Efficiency of Fixed Point Algorithms," Fixed Points - Algorithms and Applications, S. Karamardian, Editor, New York: Academic Press, 1977.

13. Saigal, R., "On the Convergence Rate of Algorithms for Solving Equations that are based on Methods of Complementary Pivoting," to appear in Mathematics of Operations Research, 2 (1977).

14. Saigal, R., "On Paths Generated by Fixed Point Algorithms," Mathematics of Operations Research, 1 (1976), 359-380.

15. Saigal, R., "Fixed Point Computing Methods," Encyclopedia of Computer Science and Technology, New York: Marcel-Dekker Inc., to appear.

16. Scarf, H., "The Approximation of Fixed Points of a Continuous Mapping," SIAM J. Appl. Math., 15 (1967), 1328-1343.

17. Scarf, H. E., Computation of Economic Equilibria (in collaboration with T. Hansen), Yale University Press, New Haven, Conn. (1973), 249 pp.

18. Smale, S., "A Convergent Process of Price Adjustment and Global Newton Methods," Journal of Math. Econ., 3 (1976), 107-120.

19. Todd, M. J., "On Triangulations for Computing Fixed Points," Math. Prog., 10 (1976), 322-346.

20. Todd, M. J., "Orientation in Complementary Pivot Algorithms," Mathematics of Operations Research, 1 (1976), 54-66.

21. Todd, M. J., "Improving the Convergence of Fixed Point Algorithms," Tech. Report No. 276, Department of Operations Research, Cornell University, Ithaca, New York, 1975.

22. Todd, M. J., The Computation of Fixed Points and Applications, New York: Springer-Verlag, 1976.

23. Varian, H., "A Remark on the Boundary Restrictions in the Global Newton Method," (unpublished communication).