ON GLOBALLY STABILIZED QUASI-NEWTON METHODS FOR INEQUALITY CONSTRAINED OPTIMIZATION PROBLEMS

by

E. Polak and A.L. Tits

Memorandum No. UCB/ERL M81/91

30 November 1981

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Introduction

Over the last several years there have been a number of successful attempts to construct superlinearly converging algorithms for the solution of constrained optimization problems. A common starting point in the construction of these new methods is the use of Newton's method, in some form, for solving the Kuhn-Tucker first order optimality condition equations and inequalities. These methods can be grouped into two categories: those traceable to R W. Wilson's successive quadratic programming method (SQP) [14], and those which emanate from the ordinary Newton method for the solution of equations.

Wilson's method is a form of Newton's method which solves a quadratic program with equality and inequality constraints at each iteration. For optimization problems of the form $\min\{f(x) \mid h(x) = 0\}$, it yields exactly the same iterates (x_i, λ_i) as the ordinary Newton method does when applied to the optimality equations $h(x) = 0$, $\nabla f(x) + (\partial h(x)/\partial x)^T\lambda = 0$; for optimization problems of the form $\min\{f(x) \mid g(x) \leqq 0\}$, it yields iterates which differ only by a second order term from those constructed by the extended Newton method, developed by Robinson [12], when applied to the Kuhn-Tucker optimality equations and inequalities, viz. $\mu^j g^j(x) = 0$, $\nabla f(x) + (\partial g(x)/\partial x)^T\lambda = 0$, $g(x) \leqq 0$, $\lambda \geqq 0$. It was shown by Robinson [11] that when intialized sufficiently closely to a "strong" Kuhn-Tucker pair (x,λ), the SQP method was quadratically convergent. SQP was extended to a quasi-Newton version by Han [4,5,6]. Han also globalized the local method, i.e., extended its domain of convergence,as well as eliminated the possibility of convergence to a local maximum instead of to a local minimum, by using an exact penalty function for step size determination: a technique subsequently refined and improved upon by Powell [10] and

Mayne and Polak [8]. The main drawback of successive quadratic programming is that it is difficult to find reliable quadratic programming codes, capable of solving non-positive-semidefinite problems, that find a solution of smallest norm, as required by Robinson's theory [11] for superlinear convergence.

The extended Newton method was never tried for solving the Kuhn-Tucker relations of general optimization problems because of a persisting erroneous belief that it would fail because the relations did not satisfy the Robinson LI conditions [12] and because it was not clear how it could be globalized. However, it was considered for problems of the form $\min\{f(x) : h(x) = 0\}$ by Tapia [13] and by Bertsekas [1]. Furthermore, Bertsekas was able to globalize Newton's method by using an exact differentiable penalty function, proposed by DiPillo and Grippo [2], as a descent function in step size determination. He showed that Newton's method yields a direction which, asymptotically, approaches the Newton direction for the DiPillo and Grippo penalty function. For problems with both equality and inequality constraints, Bertsekas has proposed an "active set" strategy, as a means of removing the need to solve inequalities as well as equations. The obvious advantage of the ordinary Newton method over successive quadratic programming is that it only needs to solve a linear equation at each iteration.

In the present paper, we show that when a sufficiently good initial approximation to a "strong" Kuhn-Tucker triplet is available, optimization problems with both equality and inequality constraints can be solved <u>without</u> using an active set strategy, by applying Newton's method, or a quasi-Newton method, only to the equations part of the Kuhn-Tucker conditions. The resulting local method is superlinearly convergent.

For problems with inequality constraints only, we show that globally convergent methods with excellent overall properties can be obtained by combining quasi-Newton methods with a phase I - phase II method of feasible directions.

2. Local Methods

Consider the problem

$$\min\{f(x)\ g(x) \leq 0,\ h(x) = 0\} \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$ and $h : \mathbb{R}^n \to \mathbb{R}^\ell$ are all twice continuously differentiable. Let x^* be a local minimizer for (1) such that the triplet $z^* = (x^*,\mu^*,\lambda^*)$ satisfies the Kuhn-Tucker first order conditions:

$$\nabla_x L(x,\mu,\lambda) = 0 ; \tag{2a}$$

$$h(x) = 0 ; \tag{2b}$$

$$\mu^j g^j(x) = 0, \quad j \in \underline{m} ; \tag{2c}$$

$$g(x) \leqq 0 ; \tag{2d}$$

$$\mu \geqq 0 ; \tag{2e}$$

where $L(x,\mu,\lambda) = f(x) + \langle \mu,g(x)\rangle + \langle \lambda,h(x)\rangle$ and $\underline{m} = \{1,2,\ldots,m\}$.

Assumption 1: With $J^* \triangleq \{j \in \underline{m} | g^j(x^*) = 0\}$, we assume that

$$\langle y, \frac{\partial^2 L(x^*,\mu^*,\lambda^*)}{\partial x^2} y\rangle > 0 \quad \forall y \in \{y' | \frac{\partial h(x^*)}{\partial x}^T y' = 0;\ \langle \nabla g^j(x^*),y'\rangle = 0$$

$$\forall j \in J^*,\ \|y'\| = 1\}. \tag{3}$$

(ii) that $\mu^{*j} > 0$ for all $j \in J^*$, and

(iii) that the vectors $\nabla h^k(x^*)$, $k \in \underline{\ell}$, $\nabla g^j(x^*)$, $j \in J^*$, are linearly independent. ¤

Now consider the equalities part of the Kuhn-Tucker conditions (2), viz:

$$\nabla_x L(x,\mu,\lambda) = 0 \; ; \tag{4a}$$

$$h(x) = 0 \; ; \tag{4b}$$

$$\mu^j g^j(x) = 0 \quad \forall j \in \underline{m} \; . \tag{4c}$$

We define our <u>local</u> algorithm as a quasi-Newton method applied to (4), viz., given $z_i \triangleq (x_i, \mu_i, \lambda_i)$,

$$z_{i+1} = z_i + \Delta z_i , \tag{5}$$

where $\Delta z_i = (\Delta x_i, \Delta\mu_i, \Delta\lambda_i)$ is a solution of the linear system

$$\nabla_x L(x_i,\mu_i,\lambda_i) + G(z_i)\Delta x_i + \frac{\partial g}{\partial x}^T (x_i)\Delta\mu_i + \frac{\partial h}{\partial x}^T (x_i)\Delta\lambda_i = 0 \; ; \tag{6a}$$

$$h(x_i) + \frac{\partial h}{\partial x} (x_i)\Delta x_i = 0 \; ; \tag{6b}$$

$$\mu_i^j g^j(x_i) + \mu_i^j \frac{\partial g}{\partial x} (x_i)\Delta x_i + \Delta\mu_i^j g^j(x_i) = 0, \quad \forall j \in \underline{m} \; . \tag{6c}$$

Clearly, when $G(z_i) = \frac{\partial^2 L}{\partial x^2} (z_i)$, (6a-6c) defines the ordinary Newton method for solving (4).

The Jacobian of the system (6) is given by

$$J(z,G) = \begin{pmatrix} G(z) & \frac{\partial g}{\partial x}^T(x) & \frac{\partial h}{\partial x}^T(x) \\ \frac{\partial h}{\partial x}(x) & 0 & 0 \\ \mu^1 \frac{\partial g^1}{\partial x}(x) & g^1(x) & 0 \\ \vdots & \vdots & \vdots \\ \mu^m \frac{\partial g^m}{\partial x}(x) & g^m(x) & 0 \end{pmatrix} \tag{7}$$

It was shown by McCormick [7] that under Assumption 1, $J(z^*, \frac{\partial^2 L}{\partial x^2}(z^*))$ is nonsingular.

We define the norm $\|\cdot\|$ on $\mathbb{R}^{n+m+\ell}$ by

$$\|z\|^2 = \|x\|^2 + \|\mu\|^2 + \|\lambda\|^2, \tag{8}$$

so that $\|(x,0_m,0_\ell)^T\| = \|x\|$. Then, using induced norms for matrices, we get

$$\|J(z,G_1(z)) - J(z,G_2(z))\| = \|G_1(z) - G_2(z)\| \tag{9}$$

<u>Theorem 1</u> (Local convergence): Suppose that for all i,

$$\|G(z_i) - \frac{\partial^2 L}{\partial x^2}(z_i)\| \leq \frac{1}{2\|J^*(z^*)\|}, \tag{10}$$

where $J^*(z^*) = J(z^*, \frac{\partial^2 L}{\partial x^2}(z^*))$. Then there exists a $\delta > 0$ such that if $z_0 \in B(z^*,\delta)$ then

(i) The sequence $\{z_i\}$ constructed according to (5), (6) is well defined;

(ii) $z_i \to z^*$ R-linearly in the norm $\|\cdot\|$.

(iii) If, in addition,

$$\|[G(z_{i-1}) - \frac{\partial^2 L}{\partial x^2}(z_{i-1})] \frac{(x_i - x_{i-1})}{\|z_i - z_{i-1}\|}\| \to 0 \text{ as } i \to \infty, \tag{11}$$

then $z_i \to z^*$ R-superlinearly.

(iv) If for some $k > 0$ and $i = 0,1,2...$,

$$\|[G(z_{i-1}) - \frac{\partial^2 L}{\partial x^2}(z_{i-1})](x_i - x_{i-1})\| < k\|z_i - z_{i-1}\|^2 \tag{12}$$

then $z_i \to z^*$ R-quadratically.

Proof: This theorem follows directly from theorems A1 and A2 in the Appendix and (9). ¤

3. Stablilization of the Local Method

In this section we shall restrict ourselves to the important subclass of problems of the form (1) which have inequality constraints only, viz. to problems of the form

$$\min\{f(x)\,|\,g(x) \leqq 0\} \tag{13}$$

Newton's method is particularly attractive for such problems, because, assuming that at least some inequalities are active, the optimality conditions for a local minimum are quite distinct from those for a local maximum, so that Newton's method cannot, inadvertently, produce a local maximum rather than a local minimum.

Obviously, we can use any globally convergent first order method on problem (13) to obtain an approximation $\tilde{z}$ to z^*, a local minimizer satisfying Assumption 1. The difficulty is in determining whether $\tilde{z}$ is in the domain of convergence of the Newton method (5), (6). We propose to do this adaptively, by monitoring whether $\tilde{\mu}$ is sufficiently "positive", $g(\tilde{z})$ sufficiently "negative" and whether Newton's method is giving signs of at least linear convergence. We shall use the phase I - phase II algorithm described in [9] for stabilization. This algorithm requires the following quantities:

$$\psi(x) \triangleq \max_{j\in\underline{m}} g^j(x) \ , \tag{14}$$

$$\psi(x)_+ \triangleq \{\max 0,\psi(x)\} \ . \tag{15}$$

For $\varepsilon > 0$, $x \in \mathbb{R}^n$ given,

$$I_\varepsilon(x) \triangleq \{j \in m \mid g^j(x) \geq \psi(x)_+ - \varepsilon\} \quad . \tag{16}$$

For $\varepsilon > 0$, $\delta > 0$ and $x \in \mathbb{R}^n$ given,

$$\theta_\varepsilon(x) \triangleq \min_{\bar{\mu}}\{\bar{\mu}^0\gamma\psi(x)_+ + \frac{1}{2}\|\bar{\mu}^0\nabla f(x) + \sum_{j\in I_\varepsilon(x)} \bar{\mu}^j\nabla g^j(x)\|^2 \mid \bar{\mu} \geqq 0,\ \Sigma\bar{\mu}^j = 1\} \tag{17}$$

For $\varepsilon_0 > 0$, $\nu \in (0,1)$ given,

$$E \triangleq \{0, \varepsilon_0, \nu\varepsilon_0, \nu^2\varepsilon_0, \ldots\} \ , \tag{18}$$

$$\varepsilon(x) \triangleq \max\{\varepsilon \in E \mid \theta_\varepsilon(x) \geq \varepsilon\} \ , \tag{19}$$

$$h(x) \triangleq -[\bar{\mu}^0_{\varepsilon(x)}\nabla f^0(x) + \sum_{j\in I_{\varepsilon(x)}(x)} \bar{\mu}^j_{\varepsilon(x)}\nabla g^j(x)] \ , \tag{20}$$

where $\bar{\mu}^k_{(x)}$, $k = 0,1,\ldots,m$, are the solutions of (17) for $\varepsilon = \varepsilon(x)$. We assume that the matrices G_i in the algorithm below will be constructed by one of the quasi-Newton formulas or set equal to $\frac{\partial^2 L}{\partial x^2}(z_i)$. In addition, we need the following standard hypothesis:

<u>Assumption 2</u>: For all $x \in \mathbb{R}^n$ such that $\psi(x) > 0$, $0 \notin \mathrm{co}\{\nabla g^j(x) \mid j \in I_0(x)\}$, where co denotes convex hull. ¤

<u>Algorithm 1</u>:

<u>Parameters</u>: ε_0, K_g, K_μ, $K_z > 0$; α, β, $\gamma \in (0,1)$.

<u>Data</u>: $x_0 \in \mathbb{R}^n$, $\bar{x}_0 = x_0$, $k = 0$, $s = 0$.

<u>Step 0</u>: Compute $\mu_0 \in \mathbb{R}^m$ by solving

$$\mu_0 = \arg\min_{\mu \geqq 0}\{\sum_{j=1}^m \mu^j g^j(x_0) + \frac{1}{2}\|\nabla f(x_0) + \sum_{j=1}^m \mu^j g^j(x_0)\|^2\} \tag{21}$$

and set $i = 0$.

Step 1: If $\min_{j\in\underline{m}} \mu_i^j < -K_\mu\gamma^k$ or $\max_{j\in\underline{m}} g^j(x_i) > K_g\gamma^k$ go to step 3. Else, compute $\Delta z_i = (\Delta x_i, \Delta\mu_i)$ by solving the linear system of equations

$$\nabla_x L(x_i,\mu_i) + G_i\Delta x_i + \frac{\partial g^T}{\partial x}(x_i)\Delta\mu_i = 0 \tag{22}$$

$$\mu_i^j g^j(x_i) + \mu_i^j \frac{\partial g}{\partial x}(x_i)\Delta x_i + \Delta\mu_i^j g^j(x_i) = 0, \ \forall j \in \underline{m} \ . \tag{23}$$

Step 2: If $\|\Delta z_i\| \le K_z\gamma^k$, set $x_{i+1} = x_i + \Delta x_i$, $\mu_{i+1} = \mu_i + \Delta\mu_i$, $i = i + 1$, $k = k + 1$ and go to step 1. Else, set $i = 0$, $k = k + 1$ and go to step 3.

Step 3: Compute $\varepsilon(\overline{x}_s)$, $h(\overline{x}_s)$ according to (18) and (19).

Step 4: If $\varepsilon(\overline{x}_s) \le \varepsilon_0\nu^k$, set $x_0 = \overline{x}_s$ and go to step 0. Else, if $\psi(x_s)_+ > 0$ compute the largest $t_s \in \{1,\beta,\beta^2,\ldots\}$ such that

$$\psi(\overline{x}_s+t_s h(\overline{x}_s)) - \psi(\overline{x}_s) \le -\alpha t_s\varepsilon(\overline{x}_s) \ , \tag{24}$$

if $\psi(\overline{x}_s) \le 0$, compute largest $t_s \in \{1,\beta,\beta^2,\ldots\}$ such that

$$\psi(\overline{x}_s+t_s h(\overline{x}_s)) \le 0 \tag{25a}$$

$$f(\overline{x}_s+t_s h(\overline{x})) - f(\overline{x}_s) \le -\alpha t_s\varepsilon(\overline{x}_s), \tag{25b}$$

set $\overline{x}_{s+1} = x_s + t_s h(\overline{x}_s)$, set $s = s + 1$ and go to step 3. ¤

Theorem 2: Suppose that (10) is satisfied for all i and that the sequence $\{\overline{x}_s\}$ is bounded. (i) If $\{\overline{x}_s\}$ is infinite then, (a) every accumulation point x^* of $\{\overline{x}_s\}$ satisfies $g(\overline{x}_s) \le 0$ and the F. John first order conditions of optimality; (b) let $\{x_s\}_K$ be the subsequence of $\{\overline{x}_s\}$ at which a transfer to step 0 takes place (i.e. $x_0 = \overline{x}_s$), then no accumulation point of $\{x_s\}_{s\in K}$ satisfies Assumption 1. (ii) If $\{\overline{x}_s\}$ is finite, then $z_i \to \hat{z}$ as $i \to \infty$, with $\hat{z} = (\hat{x},\hat{\mu})$ a Kuhn-Tucker pair. Furthermore, if $\hat{z}$ satisfies Assumption 1, then Theorem 1 gives rate of convergence, provided its assumptions are satisfied.

Proof: (i) (a) If $\{\bar{x}_s\}$ is infinite, then every accumulation point of $\{\bar{x}_s\}$ is a feasible F. John point by [9]. Furthermore, $\varepsilon(\bar{x}_s) \to 0$ as $s \to \infty$. (i) (b) Suppose that $\bar{x}_s \overset{K'}{\to} \bar{x}^*$ with $K' \subset K$ and that $\bar{x}^*$, together with the corresponding multiplier μ^* satisfy Assumption 1. We note that because of Assumption 1, μ^* is a unique Kuhn-Tucker multiplier for $\bar{x}^*$. Now, let $\{\mu_{0,s}\}_{s\in K'}$ be the multipliers μ_0 computed in Step 0 for $x_0 = \bar{x}_s$, $s \in K'$. Then, because μ^* is unique and the solutions $\mu_{0,s}$ are u.s.c. in $\bar{x}_s$, it follows that $\mu_{0,s} \overset{K'}{\to} \mu^*$ as $s \to \infty$. Consequently, there must exist an $s' \in K'$ such that the local algorithm converges superlinearly from $\mu_0 = \mu_{0,s'}$, $x_0 = \bar{x}_{s'}$ and satisfies the tests in step 1 and step 2 for all $i \geq 0$. Thus we get a contradiction that $\{\bar{x}_s\}$ is infinite.

(ii) If $\{z_i\}$ is infinite, then, since we must have that $k = i + k_0$, for some k_0, it follows that $g^j(x_i) \leq K_g\gamma^{k_0+i}$, $\forall j \in \underline{m}$ and $\mu_i^j \geq -K_\mu\gamma^{k_0+i}$ $\forall j \in \underline{m}$, for all i, so that $\overline{\lim}\, g^j(x_i) \leq 0$, and $\underline{\lim}\, \mu_i^j \geq 0$, $j \in \underline{m}$. Since $\|\Delta z_i\| \leq K_z\gamma^{k_0+i}$ for all i, it follows that $\{z_i\}$ is Cauchy and hence that $z_i \to \hat{z}$ as $i \to \infty$. It follows then from (14a,b), that $\hat{z} = (\hat{x},\hat{\mu})$ is a Kuhn-Tucker pair. The rate of convergence result follows from Theorem 1.

¤

Acknowledgement

Research sponsored by the National Science Foundation Grants ECS-79-13148 and ENV76-04264 also by the Air Force Office of Scientific Research (AFOSR) United States Air Force under Contract F49620-79-C-0178.

References

[1] Bertsekas, D.P., "Enlarging the region of convergence of Newton's method for constrained optimization," to appear in JOTA.

[2] Dipillo, G. and Grippo, L., "A new class of augmented lagrangians in nonlinear programming," SIAM J. Control, Vol. 17, No. 5, pp. 618-628, 1979.

[3] Garcia-Palomares, U.M. and Mangassarian, O.L., "Superlinear convergent quasi-Newton algorithms for nonlinearly constrained optimization problems," Mathematical Programming, Vol. 11, No. 1, pp. 1-13, 1976.

[4] Han, S.P., "A globally convergent method for nonlinear programming," JOTA, Vol. 22, 1977.

[5] Han, S.P., "Superlinearly convergent variable metric algorithms for general nonlinear programming problems," Math. Programming, Vol. 11, 1976.

[6] Han, S.P., "A hybrid method for nonlinear programming," in Nonlinear Programming 3, R.R. Meyer and S.M. Robinson eds., Academic Press, pp. 65-95, 1978.

[7] McCormick, G.P., "Penalty function versus nonpenalty function methods for constrained nonlinear programming," Mathematical Programming, Vol. 1, pp. 217-238, 1971.

[8] Mayne, D.Q., and Polak, E., "A superlinearly convergent algorithm for constrained optimization problems, " to appear in Mathematical Programming Study on Constrained Minimization. (See also publication No. 78/52, revised 15/1/1980, Department of Computing and Control, Imperial College, London).

[9] Polak, E., Trahan, R., and Mayne, D.Q., "Combined phase I - phase II methods of feasible directions," Mathematical Programming, Vol. 17, No. 1, pp. 32-61, 1979.

[10] Powell, M.J.D., "A fast algorithm for nonlinearly constrained optimization calculations," The Dundee Conference on Numerical Analysis, 1977.

[11] Robinson, S.M., "Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear programming algorithms," Math Programming, Vol. 7, No. 1, pp. 1-16, 1974.

[12] Robinson, S.M., "Extension of Newton's method to mixed system of nonlinear equations and inequalities," University of Wisconsin, Math Research Center, MRC Technical Summary Report #1161, August 1971.

[13] Tapia, R., "Diagonalized multiplier methods and quasi-Newton methods for constrained optimization," JOTA, Vol. 22, No. 2, pp. 135-194, 1977.

[14] Wilson, R.B., "A simplified algorithm for concave programming," Ph.D. Dissertation, Grad. School of Bs. Ad., Harvard Univ., Cambridge, 1963,

APPENDIX.

The following results are somewhat stronger than the ones in the open literature, cf. [3]. Consider the equation

$$f(x) = 0, \tag{A.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. A quasi-Newton method is defined by the recursion

$$G(x_i)(x_{i-1}-x_i) + f(x_i) = 0, \quad i = 1,2,3,\ldots \tag{A.2}$$

We use the notation

$$F(x) \triangleq \frac{\partial f}{\partial x}(x). \tag{A.3}$$

Let x^* be a solution of (A.1).

Assumption A 1:

(i) $F(x^*)$ is non singular.

(ii) There exists an $\varepsilon > 0$ such that

$$\|G(x) - F(x)\| < \frac{1}{2\|F(x^*)^{-1}\|} \quad \forall x \in \overline{B}(x^*,\varepsilon) \tag{A.4}$$

The following result is obvious.

Lemma A1: Let $\varepsilon > 0$ be as in Assumption A 1. Then there exist $\rho \in (0,\varepsilon)$, $M < \infty$, $\beta > 0$, $\alpha < 1/2\beta$ such that $\forall x, x' \in B(x^*,\rho)$, $F(x)$ is nonsingular and

$$\|F(x)^{-1}\| < \beta, \tag{A.5}$$

$$\|F(x) - G(x)\| < \alpha, \tag{A.6}$$

$$\|f(x') - f(x) + G(x)(x'-x)\| \leq M\|x' - x\|^2 + \|F(x)-G(x))(x'-x)\|. \tag{A.7}$$

Furthermore, x^* is the unique solution to (A.1) in $\overline{B}(x^*,\rho)$. ¤

Lemma A2: Let ρ, α, β be as in Lemma A1. Suppose that $\hat{x} \in B(x^*,\rho)$. Then $G(\hat{x})$ is nonsingular and the solution v of

$$G(\hat{x})\, v + f(\hat{x}) = 0 \tag{A.8}$$

satisfies

$$\|v\| \leq 2\beta\|f(\hat{x})\| \tag{A.9}$$

Proof : From (A.5), $F(\hat{x})$ is nonsingular and $\|F(\hat{x})^{-1}\| < \beta$. From (A.6), $\|G(\hat{x}) - F(\hat{x})\| \leq \alpha$. Since $\alpha\beta < 1$, we can apply the perturbation Lemma (see [8a] p. 45) which yields

(i) $G(\hat{x})$ is nonsingular

(ii) $\|G(\hat{x})^{-1}\| \leq \frac{\beta}{1-\alpha\beta} < 2\beta$.

Since $v = -G(\hat{x})^{-1} f(\hat{x})$, the result follows. ¤

Theorem A1: There exists a $\delta > 0$ such that, if $x_0 \in B(x^*,\delta)$, then

(i) the sequence $\{x_i\}$ constructed by (A.2) is well defined and remains in $B(x^*,\rho)$;

(ii) $\{x_i\}$ converges R-linearly to x^* in the norm $\|.\|$;

(iii) for $i = 1,2,3,\ldots$.

$$\|x_{i+1} - x_i\| \leq 2\beta[M\|x_i - x_{i-1}\|^2 + \|(F(x_{i-1})-G(x_{i-1}))(x_i-x_{i-1})\|] \tag{A.10}$$

Proof: Choose $\eta \in (2\alpha\beta,1)$, and $\delta \in (0,\frac{\rho}{2})$ such that, $\forall x \in B(x^*,\delta)$

$$\|f(x)\| \leq \min\left(\frac{\eta-2\alpha\beta}{4\beta^2 M}, \frac{(1-\eta)\rho}{4\beta}\right). \tag{A.11}$$

Now, suppose $x_0 \in B(x^*,\delta)$. We shall show by induction, that, for $i = 0,1,\ldots$

$$\|x_{i+1} - x_i\| \leq \frac{\eta-2\alpha\beta}{2\beta M}, \tag{A.12}$$

$$\|x_{i+1} - x_i\| \leq \eta^i(1-\eta)\frac{\rho}{2}. \tag{A.13}$$

First we show that (A.12) and (A.13) hold for $i = 0$. Since $x_0 \in B(x^*,\delta) \subset B(x^*,\rho)$ we have, from Lemma A2

$$\|x_1 - x_0\| \leq 2\beta\|f(x_0)\| \tag{A.14}$$

and from (A.11)

$$\|x_1 - x_0\| \leq 2\beta\,\frac{\eta-2\alpha\beta}{4\beta^2 M} = \frac{\eta-2\alpha\beta}{2\beta M}. \tag{A.15}$$

and

$$\|x_1 - x_0\| \leq 2\beta\,\frac{(1-\eta)\rho}{4\beta} = (1-\eta)\,\frac{\rho}{2}. \tag{A.16}$$

Now suppose that

$$\|x_k - x_{k-1}\| \leq \frac{\eta-2\alpha\beta}{2\beta M} \qquad \text{for } k = 1,2,\ldots,i \tag{A.17}$$

and

$$\|x_k - x_{k-1}\| \leq \eta^{k-1}(1-\eta)\,\frac{\rho}{2} \qquad \text{for } k = 1,2,\ldots i \tag{A.18}$$

Then, (a)

$$\begin{aligned}\|x_i - x_0\| &\leq \|x_i - x_{i-1}\| + \|x_{i-1} - x_{i-2}\| + \ldots + \|x_1 - x_0\| \\ &\leq (\eta^{i-1}+\eta^{i-2}+\ldots+1)(1-\eta)\,\frac{\rho}{2} \\ &< \frac{1}{1-\eta}\,(1-\eta)\,\frac{\rho}{2} < \frac{\rho}{2} \end{aligned} \tag{A.19}$$

$$\|x_i - x^*\| \leq \|x_i - x_0\| + \|x_0 - x^*\| < \frac{\rho}{2} + \delta < \rho \tag{A.20}$$

and hence $x_i \in B(x^*,\rho)$.

From (A.2),

$$f(x_i) = f(x_i) - f(x_{i-1}) + G(x_i)(x_i - x_{i-1}) \tag{A.21}$$

and, since $x_i \in B(x^*,\rho)$, we obtain from(A.7), (A.6) and (A.17) that

$$\begin{aligned} \|f(x_i)\| &\leq M\|x_i - x_{i-1}\|^2 + \|(F(x_{i-1}) - G(x_{i-1})(x_i - x_{i-1})\| \\ &\leq (M\|x_i - x_{i-1}\| + \alpha)\|x_i - x_{i-1}\| < \frac{\eta}{2\beta}\|x_i - x_{i-1}\| \end{aligned} \tag{A.22}$$

Hence, from Lemma A1, since $x_i \in B(x^*,\rho)$

$$\|x_{i+1} - x_i\| \leq 2\beta\|f(x_i)\| < \eta\|x_i - x_{i-1}\| \tag{A.23}$$

From (A.17)

$$\|x_{i+1} - x_i\| \leq \|x_i - x_{i-1}\| \leq \frac{\eta - 2\alpha\beta}{2\beta M} \tag{A.24}$$

and from (A.18),

$$\|x_{i+1} - x_i\| \leq \eta\|x_i - x_{i-1}\| \leq \eta^i(1-\eta)\frac{\rho}{2}, \tag{A.25}$$

which proves (A.12) and (A.13). Hence (A.10) holds. Also $\{x_i\}$ is well defined for all i.

Now, from (A.13), and for $j > i$

$$\|x_j - x_i\| \leq (\eta^{j-1} + \eta^{j-2} + \ldots + \eta^i)(1-\eta)\frac{\rho}{2} \leq \frac{\eta^i}{1-\eta}(1-\eta)\frac{\rho}{2} = \eta^i\frac{\rho}{2} \tag{A.26}$$

hence $\{x_i\}$ is Cauchy and $x_i \to \bar{x} \in \bar{B}(x^*,\rho)$. But from Lemma A1, this implies that $x_i \to x^*$. From (A.26), with $j \to \infty$, $\|x^* - x_i\| \leq \eta^i \frac{\rho}{2}$. Hence convergence is R-linear and this completes the proof. ¤

Theorem A2: Consider the sequence $\{x_i\}$ satisfying (A.2), with x_0 such that the conclusions of Theorem A1 hold.

(i) If $\|G(x_{i-1}) - F(x_{i-1}) \frac{(x_i-x_{i-1})}{\|x_i-x_{i-1}\|}\| \to 0$ as $i \to \infty$, then $x_i \to x^*$ R-superlinearly.

(ii) If $\|G(x_i) - F(x_i)(x_i-x_{i-1})\| < K \|x_i - x_{i-1}\|^2$

for some $K > 0$ and for $i = 1,2,\ldots$

then $x_i \to x^*$ R-quadratically.

Proof:

(i) From (A.10)

$$\frac{\|x_{i+1}-x_i\|}{\|x_i-x_{i-1}\|} \leq 2\beta[M\|x_i-x_{i-1}\| + \|F(x_{i-1}) - G(x_{i-1}) \frac{(x_i-x_{i-1})}{\|x_i-x_{i-1}\|}\|] \to 0$$

$$\text{as } i \to \infty \qquad \text{(A.27)}$$

which implies R-superlinear convergence.

(ii) From (A.10),

$$\|x_{i+1} - x_i\| \leq \gamma\|x_i - x_{i-1}\|^2 \text{ for some } \gamma > 0 \qquad \text{(A.28)}$$

which implies R-quadratic convergence. ¤