An Inexact Trust-Region Globalization of Newton's Method

M. El Hallabi

CRPC-TR94372
January 1994

Center for Research on Parallel Computation
Rice University
P.O. Box 1892
Houston, TX 77251-1892

AN INEXACT TRUST-REGION GLOBALIZATION OF NEWTON'S METHOD[1,2,3]

M. EL HALLABI[4]

Abstract.

In this work we define a trust region algorithm for approximating zeros of the nonlinear system $F(x) = 0$, where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. We are concerned with the fact that n may be large. So we replace the ℓ_2 norm with arbitrary norms in the objective function and in the trust region constraint. In particular, if polyhedral norms are used, then the algorithm can be viewed as a sequential linear programming algorithm. At each iteration, the local trust-region model is only solved within some tolerance. This research is an extension of El Hallabi and Tapia (1993) for nonlinear equations, where an exact solution of the local model was required. We demonstrate that the algorithm under consideration is globally convergent, and that, under mild assumptions, the iteration sequence generated by the algorithm converges to a solution of the nonlinear system. We also demonstrate that, under the standard assumptions for Newton's method theory, the rate of convergence is q-superlinear. Moreover, quadratic convergence can be obtained by requiring sufficient accuracy in the solution of the local model.

Key Words: Nonlinear systems, Trust-Region, Inexact Newton's method, Global convergence, superlinear convergence, quadratic convergence.

AMS subject classifications. 65K05, 49D37

1. Introduction. In this paper we consider the problem of solving the nonlinear system of equations

$$F(x) = 0, \tag{1.1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function. We will be concerned with the fact that the Jacobian of F at x, say $F'(x)$, may be sparse.

Locally, problem (1.1) is often solved by Newton's method, which is known to have fast convergence. We refer the reader to Chapter 10 of Dennis and Schnabel(1983)[3] for details. The most popular trust-region globalization strategy for Newton's method is the Levenberg-Marquardt strategy which, starting from a remote point x_0, attempts to solve the problem

$$\text{minimize}_{x \in \mathbb{R}^n} \ f(x) = \|F(x)\|_2^2.$$

[1] Research sponsored by The Moroccan American Commission for Educational and Cultural Exchanges (MACECE)(3, Rue Tiddas, Rabat, Morocco.)

[2] The author would like to thank The Council for International Exchange of Scholars (CIES) for its administrative assistance.

[3] This work was finished while the author was visiting the Computational and Applied Mathematics Department and the Center of Research on Parallel Computation, Rice University, Houston, Texas.

[4] Currently at Ecole Hassania des Travaux Publics, B.P. 8108 Route d'El Jadida, Km.7, Oasis, Casablanca, Morocco.

At each iteration, given a current estimate x_k and a current trust-region radius δ_k, the trial step s_k is obtained as the solution of the local model subproblem

$$\text{(1.2}a\text{)} \qquad \text{minimize} \quad \|F(x_k) + F'(x_k)s\|_2^2$$
$$\text{(1.2}b\text{)} \qquad \text{subject to} \quad \|s\|_2^2 \le \delta_k.$$

The Karush-Kuhn-Tucker conditions for Problem (1.2) are equivalent to

$$\text{(1.3a)} \qquad s(\mu) = -\left[F'(x_k)^T F'(x_k) + \mu I\right]^{-1} F'(x_k)^T F(x_k)$$

$$\text{(1.3b)} \qquad \mu \ge 0 \quad \text{and} \quad \|s(\mu)\|_2^2 \le \delta_k.$$

Because $s = 0$ is never the solution of (1.2) (unless $F(x) = 0$), these conditions are both necessary and sufficient. In general, the robust Hebden-Moré implementation of the Levenberg-Marquardt algorithm described in Moré (1977)[10] is used to solve (1.3a) and (1.3b) for the unique nonnegative μ_k such that $\|s(\mu_k)\|_2^2 = \delta_k$, unless $\|s(0)\|_2^2 \le \delta_k$, in which case $s(0) = s_k^N$, i.e. the Newton step for (1.1) is the solution.

The Levenberg-Marquardt algorithm is known to be globally convergent, i.e. any accumulation point of the sequence $\{x_k\}$ generated by this algorithm, starting from an arbitrary x_0, is a stationary point of the objective function. However, for larger systems, it has the disadvantage that the matrix $F'(x_k)^T F'(x_k)$ in (1.3) may mask the sparsity present in in $F'(x_k)$ along with the possibility that the computation of $s(\mu_k)$ may require several iterations.

To avoid solving (1.3) at each iteration, the dogleg (Powell (1970)[12]) or the double dogleg (Dennis and Mei (1979)[2]) can be used to obtain a good approximation to the solution of Problem (1.2). However, we cannot expect the dogleg strategies to be as robust as the Levenberg-Marquardt algorithm. In fact, Reid (1973)[13] adapted the dogleg method to the sparse case, and reported finding examples for which the method did not converge, but the standard Levenberg-Marquardt method did converge. This suggests the use of a Levenberg-Marquardt type algorithm that does not hide the sparsity pattern of $F'(x)$. This can be accomplished by using polyhedral norms instead of the ℓ_2 norm.

The use of norms different from the ℓ_2 norm in (1.2a) and in (1.2b) has been suggested and investigated by many authors. Madsen (1975)[9] uses the ℓ_∞-norm.

Powell (1983)[13], and Yuan (1983)[15][16] also considered a trust-region algorithm for minimizing $h(F(x))$, where $F : \mathbb{R}^n \to \mathbb{R}^n$, $n \le m$, is continuously differentiable and h is any coercive continuous convex function. Their local models are

$$\text{(1.4)} \qquad m_k(s) = h(F(x_k) + F'(x_k)s) + \frac{1}{2}s^T B_k s$$

where $\{B_k\}$ is a bounded sequence of symmetric matrices.

El Hallabi and Tapia (1993)[6] use arbitrary norms in $\mathbb{R}^n$ in (1.2) for the trust-region globalization strategy of Newton's method. The same approach is taken in Eisenstst and Walker (1993)[4] where the authors propose a trust-region globalization strategy for an inexact Newton's method.

In the present work, we extend the work of El Hallabi and Tapia (1993)[6] to the inexact minimization of the local model subproblem.

Consider the optimization problem

$$\text{minimize}_{x \in \mathbb{R}^n} \|F(x)\|_a \tag{1.5}$$

where $\|\ \|_a$ denotes an arbitrary, but fixed, norm in $\mathbb{R}^n$. We propose a globally convergent trust-region algorithm for approximating solutions of (1.5). At each iteration, we solve, to within some tolerance ϵ_k, the following (local model) trust-region problem:

$$(LMTR) \quad \equiv \quad \begin{cases} \text{minimize} & m_k(s) = \|F(x_k) + F'(x_k)s\|_a \\ \text{subject to} & \|s\|_b \leq \delta_k \end{cases}$$

where $\|\ \|_b$ is an arbitrary but fixed norm on $\mathbb{R}^n$, and $\|\ \|_a$ is the norm selected in (1.5). In Section 2 we review the optimality conditions for solving (1.5), and recall from El Hallabi and Tapia (1993)[6] that stationary points x_*, for which the linear system $F(x_*) + F'(x_*)s = 0$ is consistent, are solutions of $F(x) = 0$. The inexact trust-region algorithm is described in Section 3. In Section 4 we demonstrate that the inexact trust-region algorithm is globally convergent. In Section 5 we prove that this algorithm converges to a solution of $F(x) = 0$ whenever the iteration sequence has an accumulation point x_* such that $F'(x_*)$ is nonsingular. The q-superlinear convergence of the algorithm is demonstrated in Section 6; so is the fact that if more accuracy is required in the minimization of the local model subproblem, then the rate of convergence is q-quadratic. Finally, in Section 7 we present a summary and some concluding remarks.

2. Optimality Conditions. In this section, we present optimality conditions for problem (1.5) and characterize the solutions of $F(x) = 0$ where $F : \mathbb{R}^n \to \mathbb{R}^n$ is C^1, i.e. continuously differentiable. To this end, we need some subdifferentiability properties of $f = \|F\|_a$.

The composite function $f = \|F\|_a$, is locally Lipschitz (see Clarke (1983)[1]). Therefore, at any x and in any direction s in $\mathbb{R}^n$, its *generalized directional derivative* denoted $f^0(x;s)$ and defined by

$$f^0(x;s) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+ts) - f(y)}{t}. \tag{2.1}$$

Also its *generalized gradient* at x, denoted $\partial f(x)$ is the subset of $\mathbb{R}^n$ defined by

$$\partial f(x) = \{g \in \mathbb{R}^n \mid f^0(x;s) \geq g^T s, \ \ \forall s \in \mathbb{R}^n\}$$

Moreover the usual *one-sided directional derivative* of f exists, and is defined by

$$f'(x;s) = \lim_{t \downarrow 0} \frac{f(x+ts) - f(x)}{t}. \tag{2.2}$$

It is obvious that

$$f'(x;s) \leq f^0(s;s) \qquad \forall s \in \mathbb{R}^n. \tag{2.3}$$

A function for which (2.3) is an equality is said to be *regular at x* ; moreover if (2.3) holds for all x in some $X \subset \mathbb{R}^n$, then f is regular on X.

In our case, since the norm is a regular function and F is a C^1 function, we have that f is regular on $\mathbb{R}^n$.

We should mention that when the $\| \|_a$ is the ℓ_1- norm

$$f'(x;s) = \sum_{i \notin A(x)} \text{sign } F_i(x) \cdot s^T \nabla F_i(x) + \sum_{i \in A(x)} |s^T \nabla F_i(x)|$$

where

$$A(x) = \{i \mid F_i(x) = 0\}.$$

Moreover when $\| \|_a$ is the ℓ_∞-norm

$$f'(x;s) = \max_{i \in M(x)} \left\{ \text{sign } F_i(x) s^T \nabla F_i(x) \right\}$$

where

$$M(x) = \{i \mid |F_i(x)| = \|F(x)\|_\infty \}.$$

The following lemma shows that the local model m_x and the function f have the same descent directions. This is important from an algorithmic point of view.

LEMMA 2.1. [El Hallabi and Tapia (1993)[6]]. *Let x and s be any points in $\mathbb{R}^n$, $F : \mathbb{R}^n \to \mathbb{R}^n$ a continuously differentiable function at x, and $f = \|F\|$. Then*

$$f'(x;s) = m_x'(0;s),$$

where

$$m_x(s) = \|F(x) + F'(x)s\|_a. \tag{2.4}$$

The standard definition of a stationary point x_* of a real-valued function f in unconstrained nonsmooth optimization is that $0 \in \partial f(x_*)$. In our case, the function f is regular, therefore this characterization is equivalent to

$$f'(x_*;s) \geq 0$$

for all s in $\mathbb{R}^n$. The following proposition relates the definition of stationarity to the set of minimizers of the local model.

PROPOSITION 2.1. [El Hallabi and Tapia (1993)[6]]. *Let $f = \|F\|$ where $F : \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable. Then $x_* \in \mathbb{R}^n$ is a stationary point of f if and only if for all $s \in \mathbb{R}^n$*

$$\|F(x_*)\|_a \leq \|F(x_*) + F'(x_*)s\|_a$$

or equivalently $m_{x_*}(0) \leq m_{x_*}(s)$ for all $s \in \mathbb{R}^n$ where m_x is given in (2.4).

¿From Proposition 2.1 it is obvious that any solution of the nonlinear system (1.1) is a stationary point of $f = \|F\|_a$. In the following theorem, we establish a necessary and sufficient condition for a stationary point of f to be a solution of the nonlinear system (1.1).

THEOREM 2.1. *Let x_* be a stationary point of $f = \|F\|$. Then x_* is a solution of the nonlinear system*

$$F(x_*) = 0 \tag{2.6a}$$

if and only if the linear system

$$F(x_*) + F'(x_*)s = 0 \tag{2.6b}$$

is consistent.

Proof. It is obvious that if (2.6a) holds then the linear system (2.6b) has a solution. Now assume that the linear system (2.6b) has a solution s_*, i.e.

$$F(x_*) + F'(x_*)s_* = 0.$$

Since x_* is a stationary point of f, we conclude from Proposition 2.1, that $F(x_*) = 0.\Box$

3. The Inexact Trust-Region Algorithm (ITRA): In this section we define our general trust-region algorithm for approximating a solution of the nondifferentiable optimization problem

$$\text{minimize}_{x \in \mathbb{R}^n}\ f(x) = \|F(x)\|_a$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$, is continuously differentiable and where $\|\ \|_a$ is an arbitrary norm on $\mathbb{R}^n$.

DEFINITION 3.1. *Consider $x \in \mathbb{R}^n$, $\epsilon > 0$, and $\mu > 0$. Also let $\|\ \|_a$ and $\|\ \|_b$ be any two norms on $\mathbb{R}^n$. We say that s_* is an ϵ-solution of the local model trust-region subproblem*

$$\begin{array}{ll} \text{minimize} & m_x(s) = \|F(x) + F'(x)s\|_a \\ \text{subject to} & \|s\|_b \leq \mu. \end{array}$$

if s_* satisfies

$$m_x(s_*) - m_x(0) < 0 \quad \text{and} \quad m_x(s_*) \leq m_x(s) + \epsilon$$

for all s satisfying $\|s\|_b \leq \mu$.

Inexact Trust-Region Algorithm (ITRA).

Let c_i, $i = 0, \ldots, 5$, θ, δ_{min}, δ_0, and β_0 be constants satisfying :

$$\begin{array}{ll} 0 < c_1 < c_2 < 1 \leq c_3 & \quad 0 < c_4 < c_5 < 1 \\ 0 < \theta < 1 & \quad 0 < \delta_{min} \\ 0 < \delta_0 & \quad 0 < \beta_0 \end{array}$$

Let x_0 be any point in $\mathbb{R}^n$, and let $\|\ \|_a$ and $\|\ \|_b$ be any two norms on $\mathbb{R}^n$ and $\mathbb{R}^n$ respectively.

Suppose that x_k and δ_k are the iterate and the trust-region radius determined by the algorithm at the k^{th} iteration. The algorithm determines x_{k+1} and δ_{k+1} in the following manner:

<u>STEP 1</u>. Set $\mu_k = \delta_k$ and $\eta_k = \beta_k$

STEP 2. Obtain an ϵ_k-solution, with

$$\epsilon_k = \eta_k \min \left(\|s_k\|_b, \|F(x_k)\|_a \right),$$

of the local model trust-region subproblem

$$(LMTR) \quad \equiv \quad \begin{cases} \text{minimize} & m_k(s) = \|F(x_k) + F'(x_k)s\|_a \\ \text{subject to} & \|s\|_b \leq \mu_k \end{cases}$$

STEP 3. If $f(x_k + s_k) \leq f(x_k) + c_1[m_k(s_k) - f(x_k)]$
set $x_{k+1} = x_k + s_k$,
go to STEP 4.
Else
choose μ_k such that

$$c_4\|s_k\|_b \leq \mu_k \leq c_5\|s_k\|_b \ ,$$

choose $0 \leq \eta_k \leq \theta\eta_k$,
and go to STEP 2;

STEP 4. If $f(x_k + s_k) \leq f(x_k) + c_2[m_k(s_k) - f(x_k)]$
choose δ_{k+1} so that

$$\mu_k \leq \delta_{k+1} \leq \max(\mu_k, c_3\|s_k\|_b),$$

Else
choose δ_{k+1} such that

$$c_4\|s_k\|_b \leq \delta_{k+1} \leq c_5\|s_k\|_b;$$

STEP 5. Set $\delta_{k+1} = \max(\delta_{k+1}, \delta_{min})$,
and choose $0 \leq \beta_{k+1} \leq \theta\beta_k$.

REMARK 3.1. Observe that, in the process of solving subproblem (LMTR), if for some feasible step s_k, the gap between the primal value and the corresponding dual value is less than ϵ_k, then s_k is an ϵ_k-solution of subproblem (LMTR).

DEFINITION 3.2. *The couple (μ_k, η_k) defined in STEP 1, for which a solution s_k of the local model subproblem in STEP 2 satisfies the test in STEP 3, is said to determine an acceptable step s_k with respect to (x_k, δ_k, β_k). Moreover the iterate $x_{k+1} = x_k + s_k$ will be called a successor of x_k.*

4. A Fundamental Property of Trust Region Algorithms. In this section we will demonstrate that trust region algorithms enjoy the satisfying property that as the radius of the trust region approaches zero the approximate solutions, in the sense of Definition 3.1, of the model trust-region problem approach directions of steepest descent of f. The directions are steepest with respect to the norm used in defining the trust region. This result will play an important role in the convergence analysis developed in this paper.

THEOREM 4.1. *Let $\omega : \mathbb{R}^n \to \mathbb{R}$ be locally Lipschitz and let $x \in \mathbb{R}^n$ be such that the one-sided directional derivative $\omega'(x; s)$ exists for all $s \in \mathbb{R}^n$. Also let $\{\delta_k\}$ and $\{\beta_k\}$ be sequences of real numbers decreasing respectively to $\delta = 0$ and $\beta = 0$. Consider a sequence $\{s_k\}$, where s_k is an ϵ_k-solution, with $\epsilon_k = \beta_k \min(\|s_k\|, \omega(x))$ of the problem*

$$\begin{array}{ll} \text{minimize} & \omega(x + s) \\ \text{subject to} & \|s\| \leq \delta_k. \end{array}$$

If $s_k \neq 0$ for all $k \in \mathbb{N}$, then any accumulation point d_* of $\{d_k = s_k/\|s_k\|\}$ is a steepest descent direction for ω at x with respect to the norm $\| \ \|$.

Proof. Let s be any vector of norm one, and let d_* be any accumulation point of $\{d_k\}$. By choosing a subsequence, if needed, we can assume without loss of generality that $\{d_k\}$ converges to d_*. We have

$$\frac{1}{\|s_k\|}[\omega(x+s_k)-\omega(x)] \le \frac{1}{\|s_k\|}[\omega(x+\|s_k\|s)-\omega(x)]+\beta_k \tag{4.1}$$

By substituting the quantities $d_k = s_k/\|s_k\|$ and $t_k = \|s_k\|$ into (4.1) and rewriting we obtain

$$\frac{\omega(x+t_kd_*)-\omega(x)}{t_k}+\frac{\omega(x+t_kd_k)-\omega(x+t_kd_*)}{t_k} \le \frac{\omega(x+t_ks)-\omega(x)}{t_k}+\beta_k$$

which implies, because ω is locally Lipschitz and $\{\beta_k\}$ converges to zero,

$$\omega'(x,d_*) \le \omega'(x;s)\ .$$

This inequality shows that d_* is a steepest descent direction for ω at x with respect to the norm $\|\ \|$. □

REMARK 4.1. In our application the function ω will be either f or equivalently m_x (see Lemma 2.1).

5. Global Convergence of the Inexact Trust-Region Algorithm. In this section, we will establish global convergence of the Inexact Trust-Region Algorithm defined in Section 3. Throughout this section, unless otherwise mentioned, $\epsilon_k(s_k,\beta_k)$ is defined by

$$\epsilon_k(s_k,\beta_k) = \beta_k min[\|s_k\|_b, \|F(x_k)\|_a]. \tag{5.1}$$

PROPOSITION 5.1. *Consider (x,δ,β) where $\delta > 0$, $\beta > 0$, and x is not a stationary point of f. Then the Inexact Trust-Region Algorithm cannot loop infinitely often between STEP 3 and STEP 2.*

Proof. We prove the contrapositive. Suppose that the algorithm loops indefinitely. Let $\{x_j\}$ be the sequence generated by letting $x_j = x + s_j$ where s_j is a solution within $\epsilon_j(s_j,\eta_j)$ of the following model trust-region problem

$$\begin{array}{ll} \text{minimize} & m_x(s) = \|F(x)+F'(x)s\|_a \\ \text{subject to} & \|s\|_b \le \mu_j. \end{array}$$

Observe that $\|s_{j+1}\| \le \mu_{j+1} \le c_4\|s_j\|$ for all $j \in \mathbb{N}$, $0 \le \eta_{j+1} \le \theta\eta_j$, $0 < \theta < 1$, and that $0 < c_4 < 1$, so the sequences $\{\|s_j\|_b\}$ and $\{\eta_j\}$ are decreasing to zero. Under our hypothesis the test in Step 3 fails for all $j \in \mathbb{N}$, thus, because $m_x(s_j)-f(x) \ge f'(x;s_j)$, we have

$$f(x+s_j) > f(x)+c_1f'(x;s_j). \tag{5.2}$$

Since $s_j \ne 0$, we set

$$d_j = \frac{s_j}{\|s_j\|_b} \quad \text{and} \quad t_j = \|s_j\|_b.$$

Let d_* be any accumulation point of $\{d_j\}$. Without loss of generality, we can assume that the sequence converges to d_*. Therefore, from (5.2) we obtain

$$\frac{f(x+t_jd_*)-f(x)}{t_j} - \frac{f(x+t_jd_*)-f(x+t_jd_j)}{t_j} > c_1 f'(x;d_j),$$

which implies, since f is locally Lipschitz and $f'(x;\,\cdot)$ is continuous, that

$$f'(x;d_*) \geq 0.$$

But from Theorem 4.1 we obtain that d_* is a steepest descent direction for f at x. Consequently, for all $d \in \mathbb{R}^n$ with norm one, we have

$$f'(x;d) \geq 0,$$

which implies that x is a stationary point of f. □

PROPOSITION 5.2. *Consider a sequence $\{(x_k, \delta_k, \beta_k)\}$ converging to some $(x, \delta_*, 0)$ where x is not a stationary point of f and $\delta > 0$. If (μ_k, η_k) determines an acceptable step s_k with respect to (x_k, δ_k, β_k), then there exists a positive scalar $\mu(x,\delta)$ such that any accumulation point of $\{\mu_k\}$, say μ_*, satisfies the inequality*

$$\mu_* > \mu(x,\delta)\,. \tag{5.3}$$

Proof. Let μ_* be any accumulation point of $\{\mu_k\}$. Without loss of generality, we can assume that $\{\mu_k\}$ converges to μ_*. It follows that $\mu_k \leq \delta_k$. We consider two cases:
Case i) We suppose that there exists a subsequence $\{\mu_k, k \in N' \subset \mathbb{N}\}$ such that $\mu_k = \delta_k$ in which case we have $\mu_* = \delta$. Consequently we obtain (5.3) with $\mu(x,\delta) = \delta/2$.
Case ii) Suppose that $\mu_k < \delta_k$ for all sufficiently large $k \in \mathbb{N}$. Therefore, for sufficiently large k, (δ_k, β_k) never gives an acceptable step. Let $\bar{s}_k$ be the last non-acceptable step obtained by decreasing δ_k and β_k. Observe that $\bar{s}_k$ is a solution of the local model subproblem within $\epsilon_k(\bar{s}_k, \eta_k)$ for some $0 \leq \eta_k \leq \beta_k$. Since $\delta_k > 0$ and x_k is not a stationary point of f we have, by Proposition 2.1, that $\bar{s}_k \neq 0$ and $\mu_k > 0$. Also we have for large $k \in \mathbb{N}$

$$c_3\|\bar{s}_k\|_k \leq \mu_k \leq c_4\|\bar{s}_k\|_k. \tag{5.4}$$

Assume that $\mu_* = 0$. Let $s_k^* \in \operatorname{argmin}\{m_{x_*}(s) \mid \|s\| \leq \mu_k\}$, and let d^* be any accumulation point of $\{d_k^* = s_k^*/\|s_k^*\|_b\}$. Without loss of generality we can assume that $\{d_k^*\}$ converges to d^*. Since $\{0 < \mu_k\}$ converges to zero, we obtain from Theorem 4.1 and Lemma 2.1 that d^* is a steepest descent direction of f at x.

Let α_k be a positive scalar such that $\|\alpha_k s_k^*\|_b = \|\bar{s}_k\|_b$. Since $\bar{s}_k$ is an $\epsilon_k(\bar{s}_k, \eta_k)$-solution, we have

$$\frac{m_k(\bar{s}_k) - f(x_k)}{\|\bar{s}_k\|_b} \leq \frac{m_k(\alpha_k s_k^*) - f(x_k)}{\|\alpha_k s_k^*\|_b} + \eta_k. \tag{5.5a}$$

Let us set $t_k = \|\bar{s}_k\|_b = \|\alpha_k s_k^*\|_b$, $y_k^* = \alpha_k s_k^*$ and

$$\bar{d}_k = \frac{\bar{s}_k}{\|\bar{s}_k\|_b}, \quad d_k^* = \frac{y_k^*}{\|y_k^*\|_b} = \frac{s_k^*}{\|s_k^*\|_b}\,.$$

Therefore (5.5a) becomes

$$\text{(5.5b)} \qquad \frac{m_k(t_k\bar{d}_k) - f(x_k)}{t_k} \leq \frac{m_k(t_k d_k^*) - f(x_k)}{t_k} + \eta_k.$$

From inequality (5.4) we obtain that $\{t_k\}$ converges to zero. Since $\bar{s}_k$ is not acceptable, we have

$$\text{(5.6)} \qquad f(x_k + t_k\bar{d}_k) - f(x_k) > c_1[m_k(t_k\bar{d}_k) - f(x_k)].$$

On the other hand, because F if continuously differentiable, we have for all $i = 1 \cdots n$

$$F_i(x_k + t_k\bar{d}_k) = F_i(x_k) + t_k\nabla F_i(x_k)\bar{d}_k + t_k[\nabla F_i(\xi_i) - \nabla F_i(x_k)]^T\bar{d}_k$$

where $\xi_i \in (x_k, x_k + t_k\bar{d}_k)$, which implies that

$$F(x_k + t_k\bar{d}_k) = F(x_k) + t_k\nabla F(x_k)\bar{d}_k + t_k A(x_k, \bar{s}_k)\bar{d}_k$$

where $A(x_k, t_k\bar{d}_k)$ is a matrix whose i^{th} row is $[\nabla F_i(\xi_i) - \nabla F_i(x_k)]$. Since $\{x_k\}$ converges to x_*, and $\{t_k\}$ converges to zero, we have

$$\lim_{k\to+\infty} A(x_k, t_k\bar{d}_k) = 0.$$

Therefore we obtain

$$\text{(5.7a)} \qquad m_k(t_k\bar{d}_k) - f(x_k) - \|o(\|t_k\|_b)\|_a \leq f(x_k + t_k\bar{d}_k) - f(x_k)$$

and

$$\text{(5.7b)} \qquad f(x_k + t_k\bar{d}_k) - f(x_k) \leq m_k(t_k\bar{d}_k) - f(x_k) + \|o(\|t_k\|_b)\|_a.$$

¿From (5.6), (5.7b) we obtain

$$(1 - c_1)f(x_k + t_k\bar{d}_k) - f(x_k) > -\|o(\|t_k\|_b)\|_a$$

and since $0 < c_1 < 1$, this implies that

$$\text{(5.8)} \qquad 0 \leq \limsup_{k\to+\infty} \frac{f(x_k + t_k\bar{d}_k) - f(x_k)}{t_k}.$$

Similarly to (5.7a), we obtain

$$\text{(5.9)} \qquad m_k(t_k d_k^*) - f(x_k) \leq f(x_k + t_k d_k^*) - f(x_k) + \|o(\|t_k\|_b)\|_a.$$

Using (5.7b) and (5.9), we rewrite (5.5b) as

$$\frac{f(x_k + t_k\bar{d}_k) - f(x_k)}{t_k} \leq \frac{f(x_k + t_k d_k^*) - f(x_k)}{t_k} + \frac{o(t_k)}{t_k} + \eta_k$$

or

$$\text{(5.10)} \qquad \frac{f(x_k + t_k\bar{d}_k) - f(x_k)}{t_k} \leq \frac{f(x_k + t_k d^*) - f(x_k)}{t_k} +$$

$$\frac{f(x_k + t_k d_k^*) - f(x_k + t_k d^*)}{t_k} + \frac{o(t_k)}{t_k} + \eta_k.$$

Because $\{d_k^*\}$ converges to d^*, $\{t_k\}$ and $\{\eta_k\}$ converge to zero, and f is Lipschitz continuous, we obtain from (5.8) and (5.10)

$$0 \leq \limsup_{k \to +\infty} \frac{f(x_k + t_k d^*) - f(x_k)}{t_k}.$$

which implies, together with the regularity of f and the definition of f^0, that

$$0 \leq f'(x; d^*),$$

and since d_* is a steepest descent direction of f at x, we conclude that

$$0 \leq f'(x; s) \quad \forall \quad s \in \mathbb{R}^n,$$

which contradicts the hypothesis that x is not a stationary point of f. Therefore there exists a positive scalar $\mu(x, \delta)$ such that any accumulation point of the sequence $\{\mu_k\}$, say μ_*, satisfies (5.3). □

PROPOSITION 5.3. *Let $\{(x_k, \mu_k, \eta_k)\}$ be a sequence converging to $(x_*, \mu_*, 0)$, where x_* and x_k are not stationary points of f, and where μ_k and μ_* are positive. Let s_k be an $\epsilon_k(s_k, \eta_k)$-solution of the local subproblem*

(5.12a) minimize $m_k(s) = \|F(x_k) + F'(x_{x_k})s\|_a$

(5.12b) subject to $\|s\|_b \leq \mu_k$.

Then any accumulation point of $\{s_k\}$, say s_*, is an exact solution of the local subproblem

(5.13a) minimize $m_{x_*}(s) = \|F(x_*) + F'(x_*)s\|_a$

(5.13b) subject to $\|s\|_b \leq \mu_*$

Proof. Since $\{\mu_k\}$ converges to μ_* and $\|s_k\|_b \leq \mu_k$ for all k, the sequence $\{s_k\}$ is bounded. Consider any accumulation point s_* of this sequence. We prove that

(5.14) $$\|F(x_*) + F'(x_*)s_*\|_a \leq \|F(x_*) + F'(x_*)s\|_a$$

holds for all s such that $\|s\|_b \leq \mu_*$, i.e. s_* is an exact solution of (5.16). Let s satisfy $\|s\|_b \leq \mu_*$. We consider two cases:

i) Assume that $\|s\|_b < \mu_*$. Therefore, since $\{\mu_k\}$ converges to μ_* $\|s\| \leq \mu_k$ holds for sufficiently large $k \in \mathbb{N}$. Consequently, because s_k is a solution within $\epsilon_k(s_k, \eta_k)$ of the local subproblem (5.12), we obtain

(5.15) $$\|F(x_k) + F'(x_k)s_k\|_a \leq \|F(x_k) + F'(x_k)s\|_a + \epsilon_k(s_k, \eta_k)$$

which implies that (5.14) is satisfied.

ii) Assume that $\|s\|_b = \mu_*$. Consider for all $k \in \mathbb{N}$ $y_k = \frac{\mu_k}{\|s\|_b}s$, which satisfies

$\|y_k\|_b = \mu_k$. Because s_k is a solution within $\epsilon(s_k, \eta_k)$ of the local subproblem (5.12), we obtain

$$\|F(x_k) + F'(x_k)s_k\|_a \le \|F(x_k) + \frac{\mu_k}{\|s\|_b} F'(x_k)s\|_a + \epsilon_k(s_k, \eta_k). \tag{5.16a}$$

By passing to the limit when $k \to +\infty$, we obtain

$$\|F(x_*) + F'(x_*)s_*\|_a \le \|F(x_*) + \frac{\mu_*}{\|s\|_b} F'(x_*)s\|_a. \tag{5.16b}$$

Since $\|s\|_b = \mu_*$, this implies (5.14).□

THEOREM 5.1. *Consider $(x_*, \delta_*, 0)$ where $\delta_* > 0$ and x_* is not a stationary point of f. Then there exists a neighborhood $N_* = N_*(x_*, \delta_*, 0)$ and a positive scalar $\rho_* = \rho_*(x_*, \delta_*)$ such that for any (x, δ, β)*

$$f(x^+) < f(x_*) - \rho_* \tag{5.17}$$

holds for any successor (x^+, δ^+, β^+) of (x, δ, β).

Proof. We prove the contrapositive. Then there exists a sequence $\{(x_k, \delta_k, \beta_k)\}$ converging to $(x_*, \delta_*, 0)$, a sequence $\{\rho_k\}$ converging to zero, and a sequence $\{(x_k^+, \delta_k^+, \beta_k^+)\}$ where $(x_k^+, \delta_k^+, \beta_k^+)$ is a successor of (x_k, δ_k, β_k) such that

$$f(x_k^+) \ge f(x_*) - \rho_k \tag{5.18a}$$

holds for all $k \in \mathbb{N}$. Therefore, for all $k \in \mathbb{N}$, there exists an ϵ_k-solution s_k, with

$$\epsilon_k = \eta_k min(\|s_k\|_b, \|F(x_k\|_a))$$

of the subproblem

$$\begin{array}{ll} \text{minimize} & \|F(x_k) + F'(x_k)s\|_a \\ \text{subject to} & \|s\|_b \le \mu_k, \end{array}$$

where $0 < \mu_k \le \delta_k$ and $0 \le \eta_k \le \beta_k$ such that

$$x_k^+ = s_k + s_k$$

satisfies (5.18a). Because $x_k^+ = s_k + s_k$ is a successor of x_k, we have

$$f(x_k^+) \le f(x_k) + c_1[m_k(s_k) - f(x_k)].$$

¿From (5.18a) and (5.18b) we obtain

$$f(x_*) - \rho_k \le f(x_k) + c_1[m_k(s_k) - f(x_k)] \tag{5.19}$$

which implies that

$$\limsup_{k \to \infty}[m_k(s_k) - f(x_k)] \ge 0. \tag{5.20}$$

Because $\|s_k\|_b \le \delta_k$ and $\{\delta_k\}$ converges to δ_*, the sequences $\{s_k\}$ is bounded. From (5.20a) we obtain

$$\|F(x_*)\|_a = \|F(x_*) + F'(x_*)s_*\|_a. \tag{5.21}$$

for any accumulatuion point s_* of $\{s_k\}$.

Because $0 \le \mu_k \le \delta_k$ and $\{\delta_k\}$ converges to δ_*, the sequence $\{\mu_k\}$ is bounded. Let μ_* be any accumulation point of $\{\mu_k\}$. Since $\{(x_k, \delta_k, \beta_k)\}$ converges to $(x_*, \delta_*, 0)$ and x_* is not a stationary point of f, we have, by Proposition 5.2, that $\mu_* > 0$. Also, because s_k is a solution within ϵ_k, we obtain from Proposition 5.3, that s_* is an exact solution of the subproblem

$$\begin{array}{ll} \text{minimize} & m_{x_*}(s) = \|F(x_*) + F'(x_*)s\|_a \\ \text{subject to} & \|s\|_b \le \mu_*. \end{array}$$

which, together with (5.21), implies that

$$\|F(x_*)\|_a \le \|F(x_*) + F'(x_*)s\|_a \tag{5.22}$$

for all s such that $\|s\| \le \mu_*$. By Proposition 2.1, (5.22) contradicts the hypothesis that x_* is not a stationary point of f. □

It is standard in trust-region convergence theory to assume that the level set $X_0 = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$ is bounded. The following lemma shows that this hypothesis implies that the trust-region radii are uniformly bounded.

LEMMA 5.1. [El Hallabi and Tapia (1993)[6]]. *Let x_0 be any point $\in \mathbb{R}^n$. If the subset of $\mathbb{R}^n$ $X_0 = \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$ is bounded, then there exists a positive scalar δ_{max} such that the trust-region radius δ_k satisfies*

$$0 < \delta_k \le \delta_{max} \quad \forall k \in \mathbb{N}. \tag{5.23}$$

In the following theorem, we demonstrate that the Inexact Trust-Region Algorithm described in Section 3 is globally convergent.

THEOREM 5.2. *Consider a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}^n$. Let $\| \|_a$ and $\| \|_b$ be arbitrary norms on $\mathbb{R}^n$, let x_0 be an arbitrary point in $\mathbb{R}^n$, and finally let $f(x) = \|F(x)\|_a$. Assume that the level set $X_0 = \{x \in \mathbb{R}^n \| f(x) \le f(x_0)\}$ is bounded. Then any accumulation point of the sequence $\{x_k\}$ generated by the Inexact Trust-Region Algorithm using x_0 as initial point is a stationary point of f.*

Proof. Let x_* be an accumulation point of the sequence $\{x_k\}$ generated by the algorithm. Without loss of generality (by considering a subsequence if necessary), we can assume that the sequence converges to x_*. The sequence $\{(x_k, \delta_k, \beta_k)\}$ is bounded. Let $\{(x_j, \delta_j, \beta_j)\}$ be a subsequence that converges to $(x_*, \delta_*, 0)$. Because the sequence $\{f(x_k)\}$ is strictly decreasing, we have

$$f(x_j) \le f(x_k) \quad \forall j \ge k,$$

which implies that

$$f(x_*) \le f(x_k) \quad \forall k \in \mathbb{N}. \tag{5.24}$$

Suppose that x_* is not a stationary point of f. Since in $(x_*, \delta_*, 0)$, x_* is not a stationary point of f and $\delta_* > 0$, by Theorem 5.1, there exists a neighborhood $N_* = N_*(x_*, \delta_*, 0)$ and a positive scalar ρ_* such that for any $(x, \delta, \beta) \in N_*$

$$f(x^+) < f(x_*) - \rho_*$$

holds for any successor x^+ of x. Now since the sequence $\{(x_j, \delta_j, \beta_j)\}$ conerges to $(x_*, \delta_*, 0)$, there exists an integer j_* such that $(x_j, \delta_j, \beta_j) \in N_*$ for all $j \geq j_*$ and

$$f(x_{j+1}) < f(x_*) - \rho_* \quad \forall \ j \geq j_*, \tag{5.25}$$

which contradicts (5.24). Consequently, any accumulation point of the sequence $\{x_k\}$ generated by the algorithm in Section 3 is a stationary point of $f = \|F\|_a$. □

REMARK 5.1. Actually, Theorem 5.2 is can be obtained as an application of Theorem 5.1 and the work of either Huard (1979)[8] or Polak (1970)[11] dealing with the global convergence of conceptual algorithms. We choose to give a direct proof because that proof is not long and contributes to the cohesiveness of the presentation.

6. Convergence to a Solution of $F(x) = 0$. In this section we demonstrate that under standard assumptions for Newton method theory, the iteration sequence generated by the Inexact Trust-Region Algorithm actually converges to a solution of the nonlinear system (1.1).

THEOREM 6.1. *Assume the hypothesis of Theorem 5.1. Then either the linear systems*

$$F(x_*) + F'(x_*)s = 0 \tag{6.1}$$

are inconsistent for all stationary points x_* of f, or

$$F(x_*) = 0 \tag{6.2}$$

for any accumulation point x_* of f. Moreover, if the sequence $\{x_k\}$ generated by the Inexact Trust-Region Algorithm has an accumulation point, say x_*, such that $F'(x_*)$ is nonsingular, then $F(x_*) = 0$ and $\{x_k\}$ converges to x_*.

Proof. Since the sequence $\{\|F(x_k)\|_a\}$ is decreasing, it is constant on the set of accumulation points of the iteration sequence $\{x_k\}$ (see El Hallabi and Tapia (1993)[6]). Let x_{**} be an arbitrary accumulation point of $\{x_k\}$. By Theorem 5.2, x_{**} is a stationary point of $f = \|F\|_a$. Assume that (6.2) holds for x_{**}. Then $F(x_*) = 0$ for any accumulation point x_* of $\{x_k\}$, and obviously the linear system $F(x_*) + F'(x_*)s = 0$ is consistent. On the other hand, assume that the linear system (6.1), where x_* is an arbitrary accumulation point of $\{x_k\}$, is consistent. Then, by Theorem 2.1, (6.2) holds for x_*, and consequently any accumulation point is a solution of the nonlinear system $F(x) = 0$.

Now assume that $F'(x_*)$ is nonsingular. Then the linear system

$$F(x_*) + F'(x_*)s = 0$$

is consistent, which, by Theorem 2.1,implies that

$$F(x_*) = 0.$$

Finally the convergence of the sequence $\{x_k\}$ to x_* follows from Theorem 3.3 of Eisenstat and Walker (1993)[4].□

7. Convergence Rate of the Inexact Trust-Region Algorithm. In this section, we prove that the Inexact Trust Region Algorithm is q-superlinearly convergent and that it is q-quadratically convergent if either $\beta_k = O(\|F(x_k)\|)$ or $\beta_k = O(\|s_k\|)$.

THEOREM 7.1. *Assume the hypothesis of Theorem 5.1. Also assume that the sequence $\{x_k\}$ generated by the Inexact Trust-Region Algorithm has an accumulation point, say x_*, such that $F'(x_*)$ is nonsingular and F' is Lipschitz near x_*. Then the iteration sequence converges superlinearly to x_*.*

Proof. By Theorem 6.1, the iteration sequence converges to x_*. To prove that the algorithm converges superlinearly to x_*, first we need to establish that the trust-region radius cannot be decreased for sufficiently large k, i.e.

$$f(x_{k+1}) < f(x_k) + c_2[m_k(s_k) - m_k(0)], \tag{7.1}$$

where s_k is an $\epsilon_k(\|s_k\|_b, \|F(x_k\|_a)$-solution, is satisfied for sufficiently large k. Since F is continuously differentiable and $\{x_k\}$ converges to x_*, we have

$$f(x_k + s_k) = \|F(x_k) + F'(x_k)s_k + o(\|s_k\|_b)\|_a$$

(see (5.6)) and therefore

$$f(x_k) - f(x_k + s_k) \geq f(x_k) - (m_k(s_k) - \|o(\|s_k\|_b)\|_a). \tag{7.2}$$

Because $f(x_k) - m_k(s_k) > 0$, this implies

$$\frac{f(x_k) - f(x_k + s_k)}{f(x_k) - m_k(s_k)} \geq 1 - \frac{\|o(\|s_k\|_b)\|_a}{f(x_k) - m_k(s_k)}$$

or equivalently

$$\frac{f(x_k) - f(x_k + s_k)}{f(x_k) - m_k(s_k)} \geq 1 - \frac{\|o(\|s_k\|_b)\|_a}{\|s_k\|_b} \frac{\|s_k\|_b}{f(x_k) - m_k(s_k)}. \tag{7.3}$$

Let us show that the ratio

$$\frac{f(x_k) - m_k(s_k)}{\|s_k\|_b}$$

is bounded away from zero. Since $\{x_k\}$ converges to x_*, $F'(x_*)$ is nonsingular, and F is continuously differentiable, there exists k_* in $\mathbb{N}$ and a positive constant λ_* such that $F'(x_k)$ is nonsingular for all $k \geq k_*$ and

$$\|F'(x_k)d\|_a \geq \lambda_* \|d\|_b \quad \forall d \in \mathbb{R}^n \text{ and } \forall k \geq k_*. \tag{7.4}$$

Consider (x_k, δ_k, β_k) for $k \geq k_*$. Denote by s_k^N the Newton step, i.e. the solution of

$$F'(x_k)s_k^N + F(x_k) = 0.$$

First assume that $\|s_k\| \leq \|s_k^N\|$. Set

$$\alpha_k = \frac{\|s_k\|_b}{\|s_k^N\|_b} \quad \text{and} \quad \hat{s}_k = \alpha_k s_k^N. \tag{7.5}$$

Since $0 < \alpha_k \le 1$, we have

$$
(7.6) \qquad \begin{aligned} m_k(\hat{s}_k) &= (1-\alpha_k)\|F'(x_k)s_k^N\|_a \\ &= (1-\alpha_k)\|F(x_k)\|_a \end{aligned}
$$

Because $\|\hat{s}_k\|_b = \|s_k\|_b$ and s_k is an $\epsilon_k(s_k, \beta_k)$-solution of the local model, we have

$$
(7.7) \qquad f(x_k) - m_k(\hat{s}_k) \le f(x_k) - m_k(s_k) + \epsilon_k(s_k, \beta_k).
$$

¿From (7.6), (7.7) and $\|s_k\|_b = \|\hat{s}_k\|_b$ we obtain

$$
(7.9) \qquad \frac{\|F'(x_k)s_k^N\|_a}{\|s_k^N\|_b} \le \frac{f(x_k) - m_k(s_k)}{\|s_k\|_b} + \beta_k.
$$

Therefore, since $\{\beta_k\}$ converges to zero, we obtain from (7.4) and (7.8)

$$
(7.9) \qquad \frac{\lambda_*}{2} \le \frac{f(x_k) - m_k(s_k)}{\|s_k\|_b}
$$

for all sufficiently large k, say $k \ge k_*$ for convenience.

Now we assume that $\|s_k^N\|_b < \|s_k\|_b$. Since $m_k(s_k^N) = 0$ and s_k is an $\epsilon_k(s_k, \beta_k)$-solution of the local model, we have

$$
\frac{f(x_k) - m_k(s_k)}{\|s_k\|_b} \ge \frac{\|F(x_k)\|_b}{\|s_k\|_b} - \beta_k,
$$

or equivalently

$$
\frac{f(x_k) - m_k(s_k)}{\|s_k\|_b} \ge \frac{\|F(x_k) - F(x_*)\|_b}{\|s_k\|_b} - \beta_k.
$$

Also, since $F'(x_*)$ is nonsingular, we have for sufficiently large k,

$$
(7.10) \qquad \frac{f(x_k) - m_k(s_k)}{\|s_k\|_b} \ge \lambda_1(x_*)\frac{\|x_k - x_*\|_b}{\|s_k\|_b} - \beta_k
$$

for some $0 < \lambda_1(x_*)$. Because $\{f(x_k)\}$ is decreasing, we have

$$
\|F(x_{k+1}) - F(x_*)\|_a \le \|F(x_k) - F(x_*)\|_a.
$$

On the other hand, for all $i = 1, \cdots, m$, we have

$$
F_i(x_{k+1}) - F_i(x_*) = \nabla F_i(\xi_{k+1})^T(x_{k+1} - x_*), \quad \xi_{k+1} \in (x_{k+1}, x_*)
$$

or

$$
F_i(x_{k+1}) - F_i(x_*) = \nabla F_i(x_*)^T(x_{k+1} - x_*) + [\nabla F_i(\xi_{k+1}) - \nabla F_i(x_*)]^T(x_{k+1} - x_*).
$$

This implies that

$$
(7.11) \qquad F(x_{k+1}) - F(x_*) = [F'(x_*) + A(x_{k+1}, x_*)](x_{k+1} - x_*)
$$

where $A(x_{k+1}, x_*)$ is a matrix whose i^{th} row is $[\nabla F_i(\xi_{k+1}) - \nabla F_i(x_*)]^T$. Since F' is continuous and $F'(x_*)$ is nonsingular, we obtain fron (7.11) and a similar equality for x_k that

$$L_1 \|x_{k+1} - x_*\|_b \leq L_2 \|x_k - x_*\|_b \tag{7.12}$$

for some constants $0 < L_1 \leq L_2$ depending ov x_*. We claim that there exists a constant $0 < M_1$ such that

$$\frac{\|x_k - x_*\|_b}{\|s_k\|_b} \geq M_1. \tag{7.13}$$

Assume the contrary. Then

$$\lim_{k \in N \to +\infty} \frac{\|s_k\|_b}{\|x_k - x_*\|_b} = +\infty \tag{7.14}$$

for some $N \subset \mathbb{N}$. But we have

$$\frac{\|s_k\|_b}{\|x_k - x_*\|_b} = \frac{\|(x_{k+1} - x_*) - (x_k - x_*)\|_b}{\|x_k - x_*\|_b} \leq \frac{\|x_{k+1} - x_*\|_b}{\|x_k - x_*\|_b} + 1 \tag{7.15}$$

¿From (7.14), (7.15), and the convergence of $\{\beta_k\}$ to zero, we obtain

$$\lim_{k \in N \to +\infty} \frac{\|x_{k+1} - x_*\|_b}{\|x_k - x_*\|_b} = +\infty.$$

This contradicts (7.12). ¿From (7.10) and (7.13) we obtain

$$\frac{f(x_k - f(x_k + s_k)}{\|s_k\|_b} \geq M_2 \tag{7.16}$$

for all $k \geq k_*$ such that $\|s_k^N\|_a < \|s_k\|_a$. Consequently, we obtain from (7.9) and (7.16) that

$$\frac{f(x_k - f(x_k + s_k)}{\|s_k\|_b} \geq M_* \tag{7.17}$$

for all $k \geq k_*$. Inequalities (7.17) and (7.3) imply that for $k \geq k_*$ we have

$$\frac{f(x_k - f(x_k + s_k)}{f(x_k) - m_k(s)} \geq 1 - \frac{1}{M_*} \frac{\|o(\|s_k\|_b)\|}{\|s_k\|_b}. \tag{7.18a}$$

On the other hand, there exists an integer, say k_* for convenience, such that

$$1 - \frac{1}{M_*} \frac{\|o(\|s_k\|_b)\|}{\|s_k\|_b} > c_2 \tag{7.18b}$$

for all $k > k_*$. From (7,18a) and (7.18b), we obtain that equality (7.1) holds for $k \geq k_*$. Also, since $0 < c_1 < c_2$, we have

$$c_1 \left[m_k(s_k) - f(x_k)\right] \geq c_2 \left[m_k(s_k) - f(x_k)\right],$$

which, together with (7.1), implies that δ_k determines an acceptable step with respect to (x_k, δ_k, β_k), i.e. $\mu_k = \delta_k$ for all $k \geq k_*$. Consequently the trust-region radius δ_k is updated according to the rule

$$\delta_k \leq \delta_{k+1} \leq \max(\delta_k, c_3\|s_k\|_b),$$

which implies that

$$\delta_k \geq \delta_{min} \quad \forall \quad k \geq k_*. \tag{7.19}$$

We now prove that the algorithm converges q-superlinearly. We have

$$\lim_{k \to +\infty} s_k^N = -F'(x_k)^{-1}F(x_k) = 0,$$

which implies that, say for $k \geq k_*$ for convenience, the Newton step s_k^N is feasible for the subproblem (LMTR), i.e.

$$\|s_k^N\|_b \leq \delta_k. \tag{7.20}$$

We have for $k \geq k_*$

$$\begin{aligned} F(x_k) + F'(x_k)s_k &= F(x_k) + F'(x_k)(s_k - s_k^N) + F'(x_k)s_k^N \\ &= F'(x_k)(s_k - s_k^N). \end{aligned} \tag{7.21}$$

Since s_k is an $\epsilon_k(s_k, \beta_k)$-solution of the local model, we obtain from (7.20) and (7.21)

$$\|F'(x_k)(s_k - s_k^N)\|_a \leq \beta_k \min(\|s_k\|_b, \|F(x_k)\|_a) \tag{7.22}$$

and consequently

$$\begin{aligned} \|s_k - s_k^N\|_a &\leq \lambda_*^{-1}\beta_k\|F(x_k)\|_a \\ &\leq \lambda_*^{-1}\beta_k\|F(x_k) - F(x_*)\|_a \\ &\leq L_*\beta_k\|x_k - x_*\|_b \end{aligned} \tag{7.23}$$

for some constant L_* depending on x_*. On the other hand we have

$$\begin{aligned} x_{k+1} - x_* &= x_k - x_* + s_k \\ &= (x_k - x_* + s_k^N) \quad + \quad (s_k - s_k^N) \\ &= (x_k + s_k^N - x_*) \quad + \quad (s_k - s_k^N) \\ &= (x_{k+1}^N - x_*) \qquad + \quad (s_k - s_k^N). \end{aligned}$$

Observe that x_{k+1}^N is the Newton step obtained from x_k, an iterate generated by the ITRA algorithm. Therefore we have

$$\|x_{k+1} - x_*\|_b \leq \|x_{k+1}^N - x_*\|_b + \|s_k - s_k^N\|_b. \tag{7.24}$$

¿From (7.23) and (7.24), we obtain

$$\|x_{k+1} - x_*\|_b \leq \|x_{k+1}^N - x_*\|_b + L_*\beta_k\|x_k - x_*\|_b. \tag{7.25}$$

Consider D_* a convex neighborhood of x_* contained in the domain of the q-quadratic convergence of Newton's method for some constant L_2 (see Dennis and Schnabel

(1983)[3]). Since $\{x_k\}$ converges to x_*, there exists an integer , say k_* for convenience, such that $x_k \in D_*$ for all $k \geq k_*$. Then we have

$$\|x_{k+1}^N - x_*\|_b \leq L_2\|x_k - x_*\|_b^2 \quad \forall \quad k \geq k_*. \tag{7.26}$$

¿From (7.25) and (7.26) we obtain

$$\|x_{k+1} - x_*\|_b \leq L_2\|x_k - x_*\|_b^2 + L_*\beta_k\|x_k - x_*\|_b. \tag{7.27}$$

Therefore, because $\{\beta_k\}$ converges to zero, (7.27) implies that

$$\lim_{k\to+\infty} \frac{\|x_{k+1} - x_*\|_b}{\|x_k - x_*\|_b} = 0, \tag{7.28}$$

i.e. the iteration sequence $\{x_k\}$ generated by the algorithm converges q-superlinearly.□

Theorem 7.2. *Assume the hypothesis of Theorem 7.1. Then*
i)*if $\beta_k = O(\|F(x_k)\|)$ or $\beta_k = O(\|s_k\|)$, the iteration sequence converges q-quadratically to x_*, and*
ii) *if $\beta_k = 0$ for sufficiently large k, x_k is the Newton iterate for the nonlinear equation $F(x) = 0$ and consequently the rate of convergence of $\{x_k\}$ to x_* is q-quadratic.*

Proof. Assume that

$$\beta_k = O(\|F(x_k)\|_a) \tag{7.29a}$$

or

$$\beta_k = O(\|s_k\|_a). \tag{7.29b}$$

¿From (7.28) we obtain

$$\lim_{k\to+\infty} \frac{\|s_k\|_b}{\|x_k - x_*\|_b} = 1. \tag{7.30}$$

Since $F(x_*) = 0$ and F is continously differentiable, we have

$$\begin{aligned} \|F(x_k)\|_a &= \|F(x_k) - F(x_*)\|_a \\ &\leq L_2\|x_k - x_*\|_b. \end{aligned} \tag{7.31}$$

¿From (7.29a) and (7.31) or (7.30) and (7.29b), we obtain

$$\beta_k = O(\|x_k - x_*\|_b). \tag{7.32}$$

Therefore (7.27) becomes

$$\|x_{k+1} - x_*\|_b \leq L^*\|x_k - x_*\|_b^2 \tag{7.33}$$

i.e. the iteration sequence $\{x_k\}$ generated by the ITRA algorithm converges q-quadratically to x_*.

Now asume that $\beta_k = 0$ for $k \geq k_*$, which means that we are solving the local model trust-region (LMTR) exactly. The proof is similar to the one given for Theorem 8.1 of El Hallabi and Tapia (1993)[6].□

8. Summary and Concluding Remarks. A successful trust-region algorithm for approximating the solution of the nonlinear system of equations $F(x) = 0$ is the well-known Levenberg-Marquardt trust-region algorithm. The model trust-region problem in the Levenberg-Marquardt algorithm has the form

$$\text{(8.1a)} \qquad \text{minimize} \quad \|F(x) + F'(x)s\|_2^2$$
$$\text{(8.1b)} \qquad \text{subject to} \quad \|s\|_2^2 \leq \delta.$$

where $\| \ \|_2$ denotes the ℓ_2-norm on $\mathbb{R}^n$ in (8.1a) and (8.1b). But this globalization formulation of the problem is not adequate for large nonlinear problems.

A much better formulation of the problem is to consider the optimization problem

$$\text{(8.2a)} \qquad \text{minimize} \quad \|F(x) + F'(x)s\|_{norm}$$
$$\text{(8.2b)} \qquad \text{subject to} \quad \|s\|_\infty \leq \delta,$$

where *norm* can be the ℓ_1 or ℓ_∞ norm or a convex combination of these norms on $\mathbb{R}^n$. The subproblem (8.2) can be solved using linear programming techniques and allows one to take advantage of sparsity patterns in $F'(x)$. This formulation is a special case of the Inexact Trust-Region Algorithm presented in this paper, where the norms in (8.2a) and (8.2b) can be arbitrary norms in $\mathbb{R}^n$. This special case can be considered as an SLP algorithm for solving large nonlinear equations.

It is satisfying to us that we have been able to demonstrate that our Inexact Trust-Region Algorithm converges q-superlinearly under mild assumptions and that it converges q-quadratically if more accurate minimization, but not exact, of the local model is performed.

Acknowledgment. This work was done while the author was visiting the Department of Computational and Applied Mathematics and the Center for Research on Parallel Computation at Rice University, Houston, Texas. The author would like to thank all the member of this department, especially Richard Tapia for fruitful discussions and a supportive and pleasant environment. Also the author is deeply indebted to his wife and son who took care of themselves during his long absence.

The author also thanks the Moroccan American Commission for Educational and Cultural Exchange (MACECE) for supporting this research, and the administration and his colleagues of the 'Departement des Sciences de Base' of 'Ecole Hassania des Travaux Publics' for allowing him to leave for this visit. Finally he thanks The Council for International Exchange of Scholars (CIES) for its administrative assistance.

REFERENCES

[1] F.H. CLARKE, *Optimization and Nonsmooth Analysis.* Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley and Sons Publications, 1983.

[2] J.E. DENNIS Jr., and H.H. MEI, *Two new unconstrained optimization algorithms which use function and gradient values*, J. Optim. Theory Appl., 28 (1979), pp. 453-482.

[3] J.E. DENNIS Jr., and R.B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice-Hall, Englewood Cliffs, New Jersey, 1983.

[4] S.C. EISENSTAT, and H.F. WALKER, *Globally convergent inexact Newton methods*, Res. Rep. February/91/51, Department of Mathematics and Statistics, Utah State University. Revised August 1992 and March 1993. Siam J. on Optim., to appear.

[5] M. EL HALLABI, *A global convergence theory for arbitrary norm trust-region methods for nonlinear equations*, Ph.D. Thesis, Department of Mathematical Sciences, Rice University, Houston, TX, 1987.

[6] M. EL HALLABI, and R. TAPIA, *A global convergence theory for arbitrary norm trust-region methods for nonlinear equations*, Tech, Rep, 93/41, Computational and Applied Mathematics Department, Rice University, Houston, TX, 1993.

[7] M.D. HEBDEN, *An algorithm for minimization using exact second derivatives*, Tech. Rep. TP515, Atomic Energy Research Establishment, Harwell, England, 1973.

[8] P. HUARD, *Point-to set maps and mathematical programming.* Mathematical Programming Study 10, North Holland, Amsterdam, 1979.

[9] K. MADSEN, *An algorithm for minimax solution of overdetermined systems of non-linear equations* JIMA J. Instit. Math. Appl., 16 (1975), pp.321-328.

[10] J.J. MORÉ, *The Levenberg-Marquardt algorithm: implementation and theory*, in G.A. Watson, ed., Lecture Notes in Mathematics 630, Springer-Verlag, Berlin, Heidelberg and New York, 1977, pp.105-119.

[11] E. POLAK, *On the implementation of conceptual algorithm*, in O.L.Mangasarian, K.Ritter and J.B. Rosen, eds., Nonlinear Programming, Academic Press, New York, 1970, pp.275-291.

[12] M.J.D. POWELL, *A hybrid method for nonlinear equations*, in P. Rabinowitz, ed., Numerical Methods for Nonlinear Algebraic Equations, Gordon and Breach, London, 1970, pp.87-114.

[13] M.J.D. POWELL, *General algorithms for discrete nonlinear approximation calculations*, Rep. DAMTP 1983/NA2, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, England, 1983.

[14] J.K. REID. *Least squares solution of sparse systems of non-linear equations by a modified Marquardt algorithm*, in Proceedings of the NATO Conference at Cambridge, July 1972, North Holland, Amsterdam, 1973, pp.437-445.

[15] Y. YUAN. *Global convergence of trust region algorithms for nonsmooth optimization*, Department of Applied Mathematics and Theoretical Physics, Rep. DAMTP 1983/NA13, University of Cambridge, England, 1983.

[16] Y. YUAN. *Some properties of trust region algorithms for nonsmooth optimization*, Department of Applied Mathematics and Theoretical Physics, Rep. DAMTP 1983/NA4, University of Cambridge, England, 1983.