A Global Convergence Theory for Arbitrary Norm Trust-Region Algorithms for Equality Constrained Optimization

M. El Hallabi

December 1993
(Revised May 1995)

TR93-60

A Global Convergence Theory for Arbitrary Norm Trust-Region Algorithms for Equality Constrained Optimization[1]

M. El Hallabi[2]

May 13, 1995

Abstract. In this paper, we propose a trust-region algorithm to minimize a nonlinear function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ subject to nonlinear equality constraints $h_i(x) = 0, i = 1, \cdots, m$ where $h_i : \mathbb{R}^n \rightarrow \mathbb{R}$. We adopt the approach taken in Vardi (1985). We also replace the ℓ_2-norm in the trust-region constraint by an arbitrary ℓ_p-norm. In particular, if polyhedral norms are used, then the algorithm can be viewed as a sequential quadratic programming or a sequential linear programming method. At each iteration, the local model subproblem is only solved within some tolerance. Instead of the regularity assumption of linear independent gradients, we assume that the system of linearized constraints is consistent at any point of the iteration sequence, and that, at any accumulation point of the iteration sequence, the largest singular value of the constraints gradient is bounded away from zero. Moreover, we assume that the functions f and $h_i, i = 1 \cdots m$, are only continuously differentiable. Furthermore, we do not assume that the second order information matrices are uniformly bounded. We demonstrate that if (x_*, B_*) is an arbitrary accumulation point of the sequence $\{(x_k, B_k)\}$ obtained from an arbitrary starting (x_0, B_0), then x_* is a Karush-Kuhn-Tucker point of the constrained minimization problem.

Key Words: Sequential Linear Programming, Sequential Quadratic Programming, Global Convergence, Constrained Optimization, Consistency , Non Regularity, Equality Constrained , Trust-Region , Local Uniform Decrease.

---

[1]This work was partially performed when the author was visiting the Center for Research on Parallel computation and the Computational and Applied Mathematics Department at Rice University, Houston, Tx 77251-1890

[2]Departement des Sciences de Base, Ecole Hassania des Travaux Publics, B.P 8108, Route d'El Jadida, Km.7, Oasis, Casablanca, Morocco.

Introduction

In this paper we present an algorithm for approximating a solution of the equality constrained optimization problem

$$(1.1) \qquad (EQCP) \quad \equiv \quad \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & h_i(x) = 0, i = 1 \cdots m, \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}$, $\quad i = 1 \cdots m < n$, are continuously differentiable.

To solve (EQCP), SQP algorithms generate sequences $\{x_k\}$ by setting $x_{k+1} = x_k + s_k$, where s_k is obtained as the solution of the local model subproblem

$$(1.2) \qquad (QP) \quad \equiv \quad \begin{cases} \text{minimize} & c_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & h(x_k) + \nabla h(x_k)^T s = 0. \end{cases}$$

In (QP), B_k is an approximation of the Hessian of the Lagrangian, and c_k represents either the gradient of the objective function of (EQCP) or the gradient of the Lagrangian. Local convergence of SQP algorithms is generally well understood (see Fletcher [17], Tapia [34] and[35] , [36], [37]).

The problem of global convergence has been given much consideration recently. Because the trust-region strategy had proven to be a very successful tool for designing globally convergent algorithms for unconstrained optimization (e.g. Powell [26] and [28]) and for nonlinear systems of equations (e.g. El Hallabi and Tapia [15], El Hallabi [13], and Powell [27], it was quite natural to extend this strategy to constrained optimization. Global convergence results are given in Vardi [38], Byrd, Schnabel, and Shultz [4], El-Alem [10], [11], and [12], Powell and Yuan [29], Maciel [22], Dennis, El-Alem, and Maciel [9], and Alexandrov [1]. Except for Vardi [38], all the proposed algorithms are either of the framework of Celis, Dennis, and Tapia [5], or of the framework of of the two level algorithm of Byrd, Omojokun, Schnabel, and Shultz [4]. An implementation of the latest algorithm is described in Marucha, Nocedal, and Plantega [24].

To the best knowledge of the author, most global convergence theories in the literature for trust-region algorithms, that are proposed for solving problem EQCP, give global convergence in the sense that the iteration sequence has an accumulation point that is a Karush-Kuhn-Tucker point of (EQCP). Moreover, these global convergence results are obtained under the uniform regularity assumption that $\left(\nabla h(x)^T \nabla h(x)\right)^{-1}$ is uniformly bounded on a subset of $\mathbb{R}^n$ containing the iteration sequence, and that the functions f and $h_i, i \cdots m$, are twice continuously differentiable. The first hypothesis is very restrictive, especially for large-scale constrained problems.

In this research we propose an arbitrary norm inexact trust-region algorithm that is globally convergent in the sense that any accumulation point of the iteration sequence is a Karush-Kuhn-Tucker point of (EQCP). To obtain this convergence theory, instead of the uniform regularity hypothesis, we assume that the linearized constraints are consistent at all the iterates and that, at any accumulation point of the iteration sequence, the largest singular value of the constraints gradient is bounded away from zero. Furthermore, we assume that the functions f and $h_i, i \cdots m$, are only continuously differentiable.

In our method, we adopt the approach suggested by Vardi [38]. However, we replace the ℓ_2-norm in the trust-region constraint by an arbitrary ℓ_p-norm, which leads to a local model subproblem of the form

$$(1.3) \qquad (RTRQP) \quad \equiv \quad \begin{cases} \text{minimize} & c_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \alpha_k h(x_k) + \nabla h(x_k)^T s = 0, \\ & \qquad\qquad \|s\|_p \le \delta_k. \end{cases}$$

In particular, if polyhedral norms are used, then we obtain either a sequential quadratic programming or a sequential linear programming method.

In Section 2, we give a sufficient condition for the translation parameter α_k to define a nonempty feasible region for (RTRQP). In Section 3, we derive a characterization of stationarity in terms of minimizers of local model. We define the arbitrary norm inexact trust-region algorithm ANITRA in Section 4. We will consider that the algorithm generates a sequence of the form $\{(x_k, B_k, \Delta_k, \beta_k)\}$, where x_k is the iterate, B_k the matrix of second order informations, Δ_k the trust region radius, and β_k is a positive scalar used in the required accuracy to solve the local model subproblem RTRQP. In Section 5, we prove that if $(x_*, B_*, \Delta_*, 0)$ is an arbitrary accumulation point of the sequence $\{(x_k, B_k, \Delta_k, \beta_k)\}$ generated by the ANITRA algorithm from an arbitrary starting point $(x_0, B_0, \Delta_0, \beta_0)$, then x_* is a Karush-Kuhn-Tucker point of (EQCP). We end this paper by giving some concluding remarks in Section 6.

2 Linearized Constraint Translation

In this section we give a sufficient condition for the translation parameter α to define a nonempty feasible region for the subproblem (1.3). This condition is stated in the following proposition from El Hallabi [14].

PROPOSITION 2.1. [El Hallabi [14]] *Assume that $x \in \mathbb{R}^n$ is not feasible for (EQCP), i.e. $h(x) \neq 0$, and that the linear system*

$$(2.1) \qquad h(x) + \nabla h(x)^T s = 0,$$

is consistent. Assume further that $\hat{\delta} > 0$. Let σ_x be the smallest positive singular value of $\nabla h(x)$.

If

$$(2.2) \qquad 0 \le \alpha \le \min\Big(1, \hat{\delta}\, \frac{\sigma_x}{\|h(x)\|_2}\Big),$$

then the subset

$$(2.3) \qquad \mathcal{F}_2(x, \hat{\delta}, \alpha) = \Big\{ s \in \mathbb{R}^n \ \Big| \ \alpha h(x) + \nabla h(x)^T s = 0, \quad \|s\|_2 \le \hat{\delta} \Big\}$$

is not empty.

REMARK 2.1. Observe that we can generalize Inequality (2.2) to the case where $h(x_k) \neq 0$ but $\nabla h(x_k) = 0$, i.e. all its singular values are zero by choosing α equal to zero.

REMARK 2.2. A lower bound of the the singular values of $\nabla h(x)$ can be estimated by using the QR decomposition with column pivoting.

REMARK 2.3. Proposition 2.1 generalizes the result in Byrd, Schnabel, and Scultz [4] where $\nabla h(x)$ is assumed to have full rank.

3 Characterization of stationary points of problem EQCP

In this section we derive a useful notion of stationarity in terms of minimizers of the local model trust region subproblem TRQP.

PROPOSITION 3.1. *Let $B_k \in \mathbb{R}^{n\times n}$ and let $\delta_k > 0$. Consider x_k satisfying $h(x_k) = 0$. If $s_k = 0$ is a local solution of the local model subproblem*

$$(3.1) \qquad (TRQP) \quad \equiv \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \nabla h(x_k)^T s = 0 \\ & \|s\|_p \le \delta_k, \end{cases}$$

then x_k is necessarily a Karush-Kuhn-Tucker point of (EQCP). Moreover if B_k is positive semi-definite on the subspace

$$(3.2) \qquad H(x_k) = \{s \in \mathbb{R}^n \mid \nabla h(x_k)^T s = 0\}$$

then this condition is also sufficient.

Proof. The necessary condition is a consequence of the slater constraint qualification and the first order necessary conditions for zero to be a local minimizer of $(TRQP)$. Now, assume that B_k is positive semi-definite on the subspace $H(x_k)$. Then $(TRQP)$ is a convex problem, and the first order necessary conditions are also sufficient. □

In all what follows, $\| \ \|_p$ will denote an arbitrary ℓ_p-norm.

4 Arbitrary Norm Inexact Trust-Region Algorithm

In this section we propose an arbitrary norm inexact trust-region algorithm (ANITRA) for solving problem (EQCP). We also show that the choice of the penalty parameter fits well with the objective function and the constraints.

Approximate solution of the local model subproblem.
At each iteration, we solve the local model subproblem RTRQP

$$(4.1) \qquad (RTRQP) \quad \equiv \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \alpha_k h(x_k) + \nabla h(x_k)^T s = 0, \\ & \|s\|_p \le \delta_k, \end{cases}$$

for some fixed $(x_k, B_k, \alpha_k, \delta_k)$, and within some tolerance ϵ_k in the sense given in the following definition.

DEFINITION 4.1. *Let $x \in \mathbb{R}$, $B \in \mathbb{R}^{n\times n}$, $0 < \alpha$, and $0 < \delta$. Assume that x is not a Karush-Kuhn-Tucker point of (EQCP). Then we say that s_ϵ is an ϵ-solution of subproblem RTRQP if s_ϵ is feasible,*

$$(4.2) \qquad \nabla f(x)^T s_\epsilon + \frac{1}{2} s_\epsilon^T B s_\epsilon \le \nabla f(x)^T s + \frac{1}{2} s^T B s + \epsilon$$

for any feasible s, and if in addition $h(x) = 0$, we also ask that

$$(4.3) \qquad \nabla f(x)^T s_\epsilon + \frac{1}{2} s_\epsilon^T B s_\epsilon < 0.$$

Our trial step s_k will be any ϵ_k-solution of the subproblem RTRQP for fixed $(x_k, B_k, \alpha_k, \delta_k)$, and with the tolerance

$$\epsilon_k = \eta_k \begin{cases} \alpha_k \|h(x_k)\|) & \text{if } h(x_k) \neq 0 \\ \|s_k\|_p & \text{otherwise} \end{cases} \tag{4.4}$$

for some $0 < \eta_k$ that will be set by the algorithm. Observe that in (4.4)

$$\alpha_k \|h(x_k)\| = \|h(x_k)\| - \|h(x_k) \nabla h(x_k)^T s\|,$$

i.e. the decrease in the constraints for any feasible s for RTRQP subproblem. Also, when $h(x_k) = 0$, the accuracy test (4.4) ensures that any ϵ_k-solution s_k points toward the gradient projection when η_k and δ_k converge to zero (see El Hallabi and Tapia [15]). This will enable the algorithm to never stop at a nonstationary point.

In general the ϵ_k-solution is well defined. If $p = 2$, the exact solution can be computed by either method described in Sorenson [33] or Rendl and Wolkowicz [30]. Moreover, when the ℓ_∞-norm is used, (RTRQP) is a quadratic programming problem. It can be solved using the algorithm in Boggs, Domich, and Rogers [2]. Furthermore, in the following lemma and its corollary, we show that the ϵ_k-solution is well defined for a special class of such quadratic problems.

LEMMA 4.1. *Assume that the local model subproblem RTRQP is convex. Let $p_k(s_j)$ and d_k^j denote respectively the primal and the dual objective function values obtained at the j^{th} iteration for solving the subproblem RTRQP. Let $\epsilon_k > 0$. If*

$$|p_k(s_j) - d_k^j| \leq \epsilon_k \tag{4.5}$$

holds, i.e. the duality gap is less that ϵ_k, then s_j is an ϵ_k-solution of (RTRQP).

Proof. Since (RTRQP) is convex, its dual problem is well defined (see Jahn [21], Rockafellar [31], [32]). Assume that (4.5) holds. Then we have

$$p_k(s_j) \leq d_k^j + \epsilon_k \leq p_k(s_*) + \epsilon_k \tag{4.6}$$

where s_* is an exact solution of (RTRQP). From (4.6) we obtain that

$$p_k(s_j) \leq p_k(s) + \epsilon_k$$

for all feasible s for (RTRQP), i.e. s_j is an ϵ_k-solution of (RTRQP). □

COROLLARY 4.1. *If B_k is positive semi-definite on $\{s \in \mathbb{R}^n \mid \nabla h(x)^T s = 0\}$ then the ϵ_k-solution is well defined.*

Proof. Under the hypothesis of the corollary, the local model subproblem RTRQP is convex, and hence the optimal duality gap is zero (see Jahn [21]). □

Penalty parameter and merit function.

To accept or reject a trial step s_k, we will use the actual reduction

$$Ared_k(s) = \Phi(\mu_k, x_k; s) - \Phi(\mu_k, x_k; 0) \tag{4.7}$$

and the predicted reduction

$$Pred_k(s) = \Psi(\mu_k, x_k; s) - \Psi(\mu_k, x_k; 0) \tag{4.8}$$

where

$$\Phi(\mu_k, x_k; s) = f(x_k + s) + \mu_k \|h(x_k + s)\| \tag{4.9}$$

is the merit function approximated by

$$\Psi(\mu_k, x_k; s) = f(x_k) + \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s + \mu \|h(x_k) + \nabla h(x_k)^T s\|. \tag{4.10}$$

In (4.9) and (4.10), μ_k denotes the penalty parameter, and $\| \quad \|$ denotes an arbitrary (but fixed) norm on $\mathbb{R}^m$. In our converges analysis, this parameter has to satisfy two properties. First, it should force the ϵ_k-solution to be a descent direction of the predicted reduction, and second, it should be constant for large k. To obtain the first property, we choose

$$\mu_k \geq \bar{\mu_k} + \rho$$

where ρ is a positive constant and $\bar{\mu_k}$ is given by

$$\bar{\mu_k} = \begin{cases} 0 & \text{if } h(x_k) = 0 \\ 2 \max(0, \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|}) & \text{otherwise,} \end{cases} \tag{4.11}$$

(see Proposition 4.1); and to obtain the second property, we force μ_k to satisfy

$$\mu_k = \begin{cases} \mu_{k-1} & \text{if } \mu_{k-1} \geq \bar{\mu_k} + \rho \\ \bar{\mu_k} + 2\rho & \text{otherwise,} \end{cases} \tag{4.12}$$

(see Corollary 5.1).

Definition of the algorithm ANITRA.

Let $c_i, i = 1, \cdots, 5, \rho, \beta$, Δ_{min}, and Δ_{max} be constants satisfying

$$\begin{array}{lll} 0 < c_1 < c_2 < 1 & , \; 0 < c_3 < c_4 < 1 & , \; 1 < c_5 \\ 0 < \gamma < 1 & , \; 0 < \rho & , \; 0 < \beta \\ 0 < \Delta_{min} < \Delta_{max}. & & \end{array}$$

Let $x_0 \in \mathbb{R}^n$ be an arbitrary point, $B_0 \in \mathbb{R}^{n \times n}$ be an arbitrary square matrix, $\Delta_{min} \leq \Delta_0 \leq \Delta_{max}$, $0 \leq \beta_0 < \beta$, and $\mu_0 = \rho$.

Let x_k be the iterate given by the k^{th} iteration, and $0 < \beta_k$. The algorithm generates x_{k+1} by the following iterative scheme:

STEP 1. Set $\delta_k = \Delta_k, \eta_k = \beta_k$

STEP 2. If $h(x_k) = 0$ set $\alpha_k = 1$ and go to STEP 4,

STEP 3. Obtain ω_k a lower bound of the positive singular values of $\nabla h(x_k)$ and set

$$\alpha_k = \min\Big(1, \frac{\sqrt{2}}{2} \delta_k \frac{\omega_k}{\|h(x_k)\|_2}\Big),$$

STEP 4. Choose a square matrix $B_k \in \mathbb{R}^{n \times n}$,

STEP 5. Obtain an ϵ_k-solution of the subproblem RTRQP with ϵ_k defined in (4.4),

STEP 6. Update the penalty parameter μ_k using (4.12) and (4.11).

STEP 7. If $Ared_k(s_k) \leq c_1 Pred_k(s_k)\}$

set $x_{k+1} = x_k + s_k$ and go to STEP 8

Else choose δ_k such that $c_3||s_k||_p \leq \delta_k \leq c_4||s_k||_p$,

choose $0 \leq \eta_k \leq \gamma\eta_k$ and go to STEP 3.

STEP 8. If $Ared_k(s_k) \leq c_2 Pred_k(s_k)$

then choose δ_{k+1} such that $\delta_k \leq \delta_{k+1} \leq \max(\delta_k, c_5||s_k||_p)$

Else choose δ_{k+1} such that $c_4||s_k||_p \leq \delta_{k+1} \leq ||s_k||_p$.

Set $\Delta_{k+1} = \min(\Delta_{max}, \max(\delta_{k+1}, \Delta_{min}))$.

STEP 10. Choose $0 \leq \beta_{k+1} < \beta$.

REMARK 4.1. The merit function Φ, defined in (4.9), has the drawback of possessing the Maratos effect (see Maratos [23]). So, to overcome this difficulty, one may use a second order correction before decreasing the trust-region radius δ_k in STEP 7. Since adding a second order correction is irrelevant to obtaining a global convergence result, we will not extend on this technique in the present paper, and refer the interested reader to Coleman and Conn [7], Fletcher [16], Byrd, Schnabel, and Shultz [4], or Chamberlain, Lemarechal, and Pedersen [6].

We start each iteration with the trust-region radius $\Delta_k \geq \Delta_{min}$. But the actual trust-region radius, which we denote δ_k (see STEP 1), might be smaller than Δ_{min}. Throughout the paper we will use the following definition.

DEFINITION 4.2. *If for some (δ_k, η_k) defined in STEP 1, the test in STEP 7 is satisfied, then we say that (δ_k, η_k) determines an acceptable step s_k. Moreover, the iterate $x_{k+1} = x_k + s_k$ will be called a successor of x_k. Furthermore, we refer to $(x_{k+1}, B_{k+1}, \Delta_{k+1}, \beta_{k+1})$ as a successor of $(x_k, B_k, \Delta_k, \beta_k)$.*

In the following proposition, we prove that the penalty parameter μ_k fits well with the objective function and the constraints.

PROPOSITION 4.1. *Assume that x_k is not a Karush-Kuhn-Tucker point of (EQCP). Then we have*

$$Pred_k(s_k) \leq -\left|\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k\right| - \rho\alpha_k||h(x_k)||. \tag{4.13}$$

Moreover if s_k is an acceptable step, then

$$Ared_k(s_k) < 0. \tag{4.14}$$

Proof. We have

$$Pred_k(s_k) = \nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k - \mu_k\alpha_k||h(x_k)||. \tag{4.15}$$

First, we assume that

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k > 0. \tag{4.16}$$

Therefore $h(x_k) \neq 0$ must hold, and hence

$$\mu_k \geq 2\frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k||h(x_k)||} + \rho. \tag{4.17}$$

From (4.16) and (4.17), inequality (4.13) is obvious. Now we assume that

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k \leq 0. \tag{4.18}$$

Then we have $\mu_k \geq \rho$. Therefore, from (4.15) and (4.18) inequality (4.13) is obvious. Finally, assume that s_k is acceptable. Then, when $h(x_k) \neq 0$, (4.13) implies that (4.14) holds; moreover if $h(x_k) = 0$, we obtain from Definition 3.1 that

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k < 0,$$

which, together with (4.10), implies that (4.14) holds.□

5 Global Convergence

In this section, we demonstrate that any accumulation point of the sequence generated by the ANITRA Algorithm is a Karush-Kuhn-Tucker point of (EQCP). In our proofs, we will use extensively the well known Farkas Lemma. The global convergence result is given by Theorem 5.3. The derivation of some properties of the algorithm near non stationary points will play a crucial role in our global convergence theory analysis.

We make the following hypotheses:

H.1) The functions f and $h_i, i \cdots m$, are continuously differentiable,

H.2) The systems of linearized constraints $h(x_k) + \nabla h(x_k)^T s = 0$, are consistent for all k,

H.3) At any accumulation point of the iteration sequence $\{x_k\}$, say x_*, there exists $\nu_* > 0$ such that $\|\nabla h(x_*)\| \geq \nu_*$,

H.4) The sequence $\{\beta_k\}$ converges to zero, and

H.5) The ϵ_k-solution of subproblem RTRQP is well defined.

In the following lemma and its corollary, we analyze the behavior of the penalty parameter μ_k.

LEMMA 5.1. *Let* $\{(x_k, B_k, \Delta_k, \beta_k)\}$ *converge to* $(x_*, B_*, \Delta_*, 0)$. *Then there exist a positive constant* μ_* *such that* $\bar{\mu_k} \leq \mu_*$.

Proof. Because of the definition of $\bar{\mu_k}$, it is sufficient to consider the case where

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k > 0. \tag{5.1}$$

Observe that this excludes the case where $h(x_k) = 0$. Consider the least-square solution of

$$\alpha_k h(x_k) + \nabla h(x_k)^T s = 0, \tag{5.2}$$

i.e.

$$v_k \quad = \quad -\alpha_k U_{k,1} \textstyle\sum_{k,1}^{-1} V_{k,1}^T h(x_k) \tag{5.3}$$

Because of the choice of α_k, the translated hyperplane $\alpha_k h(x_k) + \nabla h(x_k)^T s = 0$ intersects the ℓ_2-ball of radius $\frac{\sqrt{2}}{2}\delta_k$, and hence it intersects any ℓ_p-ball of radius δ_k. Therefore, because of the convexity of such a ball, we have necessarily

$$\|v_k\|_p \leq \delta_k. \tag{5.4}$$

From (5.2) and (5.4), we obtain that v_k is a feasible point for the subproblem RTRQP. Therefore, because s_k is an $\epsilon_k(s_k, \eta_k)$-solution of (RTRQP) we have

$$0 < \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|} \leq \frac{\nabla f(x_k)^T v_k + \frac{1}{2} v_k^T B_k v_k}{\alpha_k \|h(x_k)\|} + \eta_k$$

for some $0 < \eta_k \leq \beta_k$, which implies that

$$(5.5) \qquad 0 < \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|} \leq \frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \left(\|\nabla f(x_k)\|_2 + \frac{1}{2} \|B_k\|_2 \delta_k \right) + \eta_k .$$

Since the sequences $\{\nabla f(x_k)\}$, $\{\delta_k\}$, $\{B_k\}$, and $\{\eta_k\}$ are bounded, we obtain from (5.5) that

$$(5.6) \qquad 0 < \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|} \leq \frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} M_1 + \beta,$$

for some positive constant M_1. From (5.3), we obtain

$$(5.7) \qquad \frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \leq \frac{\|\Sigma_{k,1}^{-1} V_{k,1}^T h(x_k)\|_2}{\|h(x_k)\|},$$

and because of the equivalence of norms on $\mathbb{R}^n$, this implies that

$$(5.8) \qquad \frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \leq \frac{1}{\nu_{min}} \quad \frac{1}{\sigma_{k,r_k}},$$

where ν_{min} is a positive constant such that $\| \; \| \geq \nu_{min} \| \; \|_2$. On the other hand, the singular values of $\nabla h(x)$ are continuous functions of x. Let σ_* be the smallest nonzero singular value of $\nabla h(x_*)$. From Hypothesis H.3, we obtain that $\sigma_* > 0$. Therefore there exists a positive integer k_1, depending on x_*, such that

$$(5.9) \qquad \sigma_{k,r_k} \geq \frac{1}{2} \sigma_* > 0 \quad \forall \quad k \geq k_1 .$$

Now, from (5.9) and (5.8), we obtain

$$(5.10) \qquad \frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \leq \frac{2}{\sigma_* \; \nu_{min}},$$

which, together with (5.6), implies that there exists μ_* such that $\bar{\mu_k} \leq \mu_*$.□

Now we show that the penalty parameter μ_k is constant for sufficiently large k.

COROLLARY 5.1 *Assume the hypotheses of Lemma 5.2. Then there exists an integer k^* such $\mu_k = \mu_{k^*}$ for all k.*

Proof. The sequence $\{\mu_k\}$ is not decreasing, and when it increases, it does it with a positive constant ρ. Therefore it is sufficient to establish that it is bounded. Assume that there exists a subsequence $\{\mu_k, k \in N\}$ such that $\lim_{k \in N \to +\infty} \mu_k = +\infty$. Let μ_* the upper bound of $\{\bar{\mu_k}\}$ and let k_* be the smallest integer in N such that $\mu_{k_*} \geq \mu_* + \rho$. Since $\{\mu_k\}$ is not decreasing, we obtain

$$\mu_{k-1} \geq \bar{\mu_k} + \rho \quad \forall k \geq k_1 + 1,$$

and hence $\mu_k = \mu_{k-1}$ for all $k \geq k_*$. This contradicts the divergence assumption of a subsequence.□

The following lemma will be needed later.

LEMMA 5.2. *Assume that $\{(x_k, \Delta_k, \beta_k)\}$ converges to $(x_*, , \Delta_*, 0)$. Assume further that*

$$(5.11) \qquad \lim_{k \to +\infty} \alpha_k h(x_k) = 0$$

holds. Then

$$(5.12) \qquad h(x_*) = 0$$

holds. Moreover $\alpha_k = 1$ holds for sufficiently large k.

Proof. From the equivalence of norms, the definition of α_k, and (5.11), we obtain

$$\lim_{k \to +\infty} \min\Big(\|h(x_k)\|_2, \frac{\sqrt{2}}{2} \, \Delta_k \, \omega_k \Big) = 0. \tag{5.13}$$

On the other hand, the singular values of $\nabla h(x)$ are continuous functions of x. Let σ_* denote the smallest nonzero singular value of $\nabla h(x_*)$. We obtain form hypothesis H.3 that, necessarily

$$\sigma_{k,r_k} \geq \frac{\sigma_*}{2} > 0 \tag{5.14}$$

holds for sufficiently large k. Therefore we can assume that the lower bound of the singular values of $\nabla h(x_k)$, i.e. ω_k, satisfies

$$\omega_k \geq \tau_* \sigma_* \tag{5.15}$$

where $\tau_* \in (0, \frac{1}{10})$ is an arbitrary small positive constant that depends on x_*. Now, from (5.13) and (5.15), we obtain

$$h(x_*) = 0. \tag{5.16}$$

Finally, we obtain from the definition of α_k, $\Delta_k \geq \Delta_{min}$, and (5.15) that

$$\alpha_k = 1 \tag{5.17}$$

holds for sufficiently large k. □

In the following lemma, we establish an intermediate result needed to prove Proposition 5.1 and Theorem 5.1.

LEMMA 5.3. *Let $\{(x_k, B_k, \delta_k, \eta_k)\}$ converge to $(x_*, B_*, 0, 0)$. Assume that for all k, s_k is feasible for subproblem RTRQP, but it is not an acceptable step. Then we have*

$$a) \lim_{k \to +\infty} \frac{\|h(x_k)\|}{\|s_k\|_p} = 0, \qquad b)\ h(x_*) = 0, \qquad c) \liminf_{k \to +\infty} \nabla f(x_k)^T d_k \geq 0, \tag{5.18}$$

where $d_k = s_k / \|s_k\|_p$.

Proof. The proof of this lemma is quite long, it will be given in Appendix A.

In the following proposition, we establish that, unless the iterate x_k is a Karush-Kuhn-Tucker point of (EQCP), the algorithm ANITRA finds an acceptable step s_k by solving a finite number of times the subproblem RTRQP with decreasing trust-region radii. This is the first important property of the trust-region strategy.

PROPOSITION 5.1. *If x_k is not a Karush-Kuhn-Tucker point of (EQCP), then the algorithm finds an acceptable step s_k after a finite number of loops between STEP 7 and STEP 3.*

Proof. Assume that the algorithm loops indefinitely between STEP 7 and STEP 3 without obtaining an acceptable step s_k. Then the algorithm generates a sequence $\{s_j\}$ of non acceptable steps that converge to zero. Therefore, we obtain from Lemma 5.3 that necessarily $h(x_k) = 0$ and

$$\limsup_{j \to +\infty} \nabla f(x_k)^T d_j \geq 0, \tag{5.19}$$

where $d_j = s_j / \|s_j\|_p$. Consider $s \in \mathbb{R}^n$ such that $\|s\|_p = 1$ and

$$\nabla h(x_k)^T s = 0, \tag{5.20}$$

and let $t_j = \|s_j\|_p$. This implies that $t_j s$ is feasible for the local model subproblem. Hence, because s_j is an ϵ_j-solution, with $\epsilon_j = \eta_j \|s_j\|_p$, we have

$$\nabla f(x_k)^T d_j + \frac{1}{2} s_j^T B_k d_j \leq \nabla f(x_k)^T s + \frac{1}{2} t_j s^T B_k s + \eta_j$$

which, together with (5.19) and the convergence of $\{\eta_j, j \in \mathbb{N}\}$ and $\{s_j\}$ to zero, implies that

$$\nabla f(x_k)^T s \geq 0. \tag{5.21}$$

From the Farkas Lemma, (5.20), and (5.21) we obtain that x is a Karush-Kuhn-Tucker point of (EQCP), which contradicts our hypothesis.□

Proposition 5.1 implies that either the algorithm ANITRA generates a sequence $\{x_i, i = 1, \cdots, s\}$ such that x_s is a Karush-Kuhn-Tucker point of (EQCP), or the iteration sequence is infinite. Therefore, throughout the remaining part of the paper, we assume that ANITRA algorithm generates an infinite sequence $\{x_k\}$.

The second important property of our trust-region framework is established in the following theorem. This property is equivalent to saying that if some subsequence of the iteration sequence $\{x_k\}$ converges to x_* and the corresponding subsequence of trust-region radii that determine acceptable steps converges to zero, then necessarily x_* is a Karush-Kuhn-Tucker point of (EQCP).

THEOREM 5.1. *Let $\{(x_k, B_k, \Delta_k, \beta_k)\}$ converge to $(x_*, B_*, \Delta_*, 0)$ where x_k and x_* are not Karush-Kuhn-Tucker points of (EQCP) and $0 < \Delta_*$. If (δ_k, η_k) determines an acceptable step s_k, then there exists a positive scalar $\delta(x_*, B_*, \Delta_*)$ such that*

$$\delta_* \geq \delta(x_*, B_*, \Delta_*) \tag{5.22}$$

holds for any accumulation point δ_* of $\{\delta_k\}$.

Proof. The proof of this theorem is quite long and will be given in Appendix A.

Before we give our global convergence result, we establish in the following theorem, perhaps the most important property of our trust-region algorithm. This property is called *local uniform decrease.* Since, for all sufficiently large k the penalty parameter $\mu_k = \mu_{k^*}$ (see Corollary 5.1), and since we assume that the iteration sequence is infinite, the merit function $\Phi(\mu_k, x, s)$ is constant with respect to this parameter. Therefore, we denote $\Phi(x+s)$ instead of $\Phi(\mu_k, x, s)$.

THEOREM 5.2 (Local Uniform Decrease). *Let $(x_k, B_k, \Delta_k, \beta_k)$ converges to $(x_*, B_*, \Delta_*, 0)$. If x_* is not a Karush-Kuhn-Tucker point of (EQCP), then there exists a positive integer, say k_*, such that for all $k \geq k_*$*

$$\Phi(x_{k+}) < \Phi(x_*) \tag{5.23}$$

holds for any successor $(x_{k+}, B_{k+}, \Delta_{k+}, \beta_{k+})$ of $(x_k, B_k, \delta_k, \beta_k)$.

Proof. Assume that the theorem does not hold. Then there exists a subsequence $\{(x_k, B_k, \Delta_k, \beta_k\}$, $\beta_k > 0$ converging to $(x_*, B_*, \Delta_*, 0)$, and a subsequence of successors $\{(x_{k+}, B_{k+}, \Delta_{k+}, \beta_{k+})\}$ (see Definition 4.2) such that

$$\Phi(x_{k+}) \geq \Phi(x_*) \tag{5.24}$$

holds for all k. Without loss of generality, we assume that the entire sequence converges. Therefore there exists an ϵ_k-solution s_k of the local model subproblem

$$(RTRQP) \quad \equiv \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \alpha_k h(x_k) + \nabla h(x_k)^T s = 0 \\ & \|s\|_p \leq \delta_k \end{cases} ,$$

for some $0 < \eta_k \leq \beta_k$ and $0 < \delta_k \leq \Delta_k$, such that $x_{k+} = x_k + s_k$ and

$$\Phi(x_k + s_k) \leq \Phi(x_k) + c_1 \{\Psi(x_k + s_k) - \Psi(x_k)\}. \tag{5.25}$$

From inequalities (5.25), (5.24) and Proposition 4.1, we obtain

$$\Phi(x_*) - \Phi(x_k) \leq c_1 [-\rho\alpha_k \|h(x_k)\|] - c_1 |\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k|. \tag{5.26}$$

Since $\{x_k\}$ converges to x_*, we obtain from (5.26)

$$\lim_{k \to +\infty} \alpha_k \|h(x_k)\| = 0, \tag{5.27}$$

which, together with Lemma 5.2, implies that

$$h(x_*) = 0, \tag{5.28}$$

and for sufficiently large k, the translation parameter α_k is identically equal to one. Hence, for all sufficiently large k, s_k is an ϵ_k-solution of the local model subproblem

$$(TRQP) \quad \equiv \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & h(x_k) + \nabla h(x_k)^T s = 0 \\ & \|s\|_p \leq \delta_k . \end{cases} \tag{5.28}$$

The sequences $\{s_k\}$ and $\{\delta_k\}$ are bounded, then without loss of generality, we can assume that they converge respectively to s_*, and δ_*, where, by Theorem 5.1, $0 < \delta_*$. Therefore, by Theorem A15 of Huard [19], we have that s_* solves the subproblem

$$\begin{cases} \text{minimize} & \nabla f(x_*)^T s + \frac{1}{2} s^T B_* s \\ \text{subject to} & \nabla h(x_*)^T s = 0 \\ & \|s\|_p \leq \delta_* . \end{cases} \tag{5.29}$$

(Huard's Theorem establishes, in a more general case, the continuity of an approximate solution (in our case s_k) of a given optimization problem (in our case RTRQP), considered as function of the variables that play the role of parameters (in our case x_k, B_k, and δ_k). On the other hand, we obtain from (5.26) that

$$\nabla f(x_*)^T s_* + \frac{1}{2} s_*^T B_* s_* = 0. \tag{5.31}$$

Consequently $s = 0$ solves the subproblem (5.29) which, by Proposition 3.1, contradicts the hypothesis that x_* is not a Karush-Kuhn-Tucker point of (EQCP).□

Finally, we give our global convergence result which detracts from the matter at hand.

THEOREM 5.3. *Assume hypotheses H.1 - H.5. Then any accumulation point of the sequence* $\{(x_k, B_k, \Delta_k, \beta_k)\}$ *generated by the algorithm ANITRA of Section 3, say* $\{(x_*, B_*, \Delta_*, 0)\}$, *is such that* x_* *is a Karush-Kuhn-Tucker point of (EQCP).*

Proof. Let $\{(x_*, B_*, \Delta_*, 0)\}$ be an arbitrary accumulation point of $\{(x_k, B_k, \Delta_k, \beta_k)\}$. Since we assume that the sequence is infinite, we consider the sequence for $k \geq k^*$, where k^* is defined in Corollary 5.1. Because for $k \geq k^*$ the penalty parameter μ_k is constant, the merit function Φ is constant with respect to this parameter and therefore will be denoted $\Phi(x)$. Let $\{x_j, B_j, \Delta_j, \beta_j, j \geq k^*\}$ be a subsequence that converges to $\{(x_k, B_k, \Delta_k, \beta_k)\}$. Consider $k \geq k^*$. Since the sequence $\{\Phi(x_k), k \geq k^*\}$ is decreasing, we have

$$\Phi(x_*) \leq \Phi(x_k) \quad \forall k \geq k^*. \tag{5.32}$$

Assume that x_* is not a Karush-Kuhn-Tucker point of (EQCP) and consider the neighborhood N_* defined in Theorem 5.2. Since $(x_j, B_j, \Delta_j, \beta_j)$ converges to $(x_*, B_*, \Delta_*, 0)$ there exists $j(x_*) \geq k^*$ such $(x_j, B_j, \Delta_j, \beta_j) \in N_*$ for all $j \geq j(x_*)$, and hence

$$\Phi(x_{j+1}) < \Phi(x_*) \quad \forall j \geq j(x_*).$$

This contradicts (5.32). Therefore x_* is a Karush-Kuhn-Tucker point of (EQCP).□

Actually, Theorem 5.3 can be obtained as an application of Theorem 5.1 and the work of either Huard [20] or Polak [25] 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 completeness of the presentation.

6 Concluding Remarks

In this paper, we have presented an arbitrary norm inexact trust-region algorithm ANITRA for approximating a solution of the equality constrained problem

$$(EQCP) \quad \equiv \quad \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & h_i(x) = 0, i = 1 \cdots m, \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h_i : \mathbb{R}^n \to \mathbb{R}, \quad i = 1 \cdots m < n$, are continuously differentiable.

The local model trust region subproblem has the form

$$(RTRQP) \quad \equiv \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \alpha_k h(x_k) + \nabla h(x_k)^T s = 0 \\ & \qquad ||s||_p \leq \delta_k, \end{cases}$$

where $|| \, ||_p$ is an arbitrary ℓ_p-norm on $\mathbb{R}^n$. In particular if polyhedral norms are use then the subproblem can be formulated as a sequential linear programming or a sequential quadratic programming problem.

We established, under rather weak assumption, that in any accumulation point of the the sequence $\{(x_k, B_k, \Delta_k, \beta_k)\}$ generated by the ANITRA Algorithm, say $(x_*, B_*, \Delta_*, 0)$, the point x_* is a Karush-Kuhn-Tucker point of (EQCP). This convergence result is obtained without assuming that $\{B_k\}$ is bounded; the sequence $\{B_k\}$ may have accumulation points without being bounded. Also, we only assume that f and $h_i, i = 1 \cdots m$, are continuously differentiable, that the system of linearized constraints is consistent, and that the largest singular value of the constraints gradient is bounded away from zero. To the best of our knowledge, the other convergence results establish that there exists some accumulation point of the iteration sequence that is a Karush-Kuhn-Tucker point of (EQCP). Moreover, generally, stronger assumptions such as uniform linear independence of the gradients $\nabla h_i(x), i = 1 \cdots m$, continuity of the second derivatives of f and $h_i, i = 1 \cdots m$, and boundness of the second order informations matrices are needed. We end this paper by pointing out that our approach provides a global convergence theory of the sequential linear programming version obtained by working without the matrices B_k.

7 Appendix A

In this appendix, we give the proofs of Lemma 5.3 and Theorem 5.1.

Proof of Lemma 5.3. Since s_k is not acceptable, we have

$$\Phi(\mu_k, x_k, s_k) - \Phi(\mu_k, x_k, 0) > c_1[\Psi(\mu_k, x_k, s_k) - \Psi(\mu_k, x_k, 0)]$$

or equivalently

$$f(x_k + s_k) - f(x_k) + \mu_k \Big[||h(x_k + s_k)|| - ||h(x_k)||\Big] > \frac{c_1}{2} s_k^T B_k s_k + c_1 \Big[\nabla f(x_k)^T s_k + \mu_k \big[||h(x_k) + \nabla h(x_k)^T s_k|| - ||h(x_k)||\big]\Big]. \tag{7.1}$$

On the other hand, we have

$$\frac{f(x_k + s_k) - f(x_k)}{||s_k||_p} = \nabla f(x_k)^T d_k + \frac{o_1(||s_k||_p)}{||s_k||_p}. \tag{7.2}$$

and

$$\frac{||h(x_k + s_k)|| - ||h(x_k)||}{||s_k||_p} \leq \frac{||h(x_k) + \nabla h(x_k)^T s_k|| - ||h(x_k)||}{||s_k||_p} + \frac{o_2(||s_k||_p)}{s_k||_p}. \tag{7.3}$$

Since $\{x_k\}$ converges to x_* and $\{s_k\}$ converges to zero, we have that

$$\lim_{k\to+\infty} \frac{o_1(||s_k||)}{||s_k||_p} = 0 \quad \text{and} \quad \lim_{k\to+\infty} \frac{o_2(||s_k||)}{||s_k||_p} = 0. \tag{7.4}$$

Because $\{\mu_k\}$ and $\{B_k\}$ are bounded, $0 < 1 - c_1$, and $\{s_k\}$ converges to zero, we obtainfrom (7.1) (7.2), (7.3) and (7.4)

$$\nabla f(x_k)^T d_k + \mu_k \frac{||h(x_k) + \nabla h(x_k)^T s_k|| - ||h(x_k||}{||s_k||_p} > \frac{o(||s_k||)}{||s_k||_p} \tag{7.5}$$

where

$$\lim_{k\to+\infty} \frac{o(||s_k||)}{||s_k||_p} = 0. \tag{7.6}$$

Using $\alpha_k h(x_k) + \nabla h(x_k)^T s_k = 0$, we can rewrite (7.5) as

$$\nabla f(x_k)^T d_k > \mu_k \alpha_k \frac{||h(x_k)||}{||s_k||_p} + \frac{o(||s_k||_p)}{||s_k||_p}, \tag{7.7}$$

which, together with (7.6), implies (5.18c), i.e.

$$\liminf_{k\to+\infty} \nabla f(x_k)^T d_k \geq 0. \tag{7.8}$$

From (7.7), we obtain

$$2\nabla f(x_k)^T d_k - \mu_k\alpha_k \frac{||h(x_k)||}{||s_k||_p} > \mu_k\alpha_k \frac{||h(x_k)||}{||s_k||_p} + \frac{o(||s_k||_p)}{||s_k||_p},$$

which, together with (7.6), implies that

$$\liminf_{k\to+\infty}\Big[2\nabla f(x_k)^T d_k - \mu_k\alpha_k \frac{||h(x_k)||}{||s_k||_p}\Big] \geq 0. \tag{7.9}$$

But from the definition of the penalty parameter μ_k, we obtain

$$2\nabla f(x_k)^T d_k - \mu_k\alpha_k \frac{||h(x_k)||}{||s_k||_p} + d_k^T B_k s_k \leq -\rho\alpha_k \frac{||h(x_k)||}{||s_k||_p},$$

which implies, since $\liminf_{k\to+\infty} d_k^T B_k s_k = \lim_{k\to+\infty} d_k^T B_k s_k = 0$, that

$$\liminf_{k\to+\infty}\Big(2\nabla f(x_k)^T d_k - \mu_k\alpha_k \frac{||h(x_k)||}{||s_k||_p}\Big) \leq \liminf_{k\to+\infty} -\alpha_k \frac{||h(x_k)||}{||s_k||_p}. \tag{7.10}$$

From (7.9) and (7.10) we obtain

$$\lim_{k\to+\infty} \alpha_k \frac{||h(x_k)||_2}{||s_k||_p} = 0. \tag{7.11}$$

Since $\{x_k\}$ converges to zero, we obtain from Lemma 5.2 and (7.11) that $\alpha_k = 1$ for sufficiently large k. Therefore, we rewrite (7.11) as

$$\lim_{k\to+\infty} \frac{||h(x_k)||}{||s_k||_p} = 0,$$

i.e. (5.18a) which implies, since $\{s_k\}$ converges to zero, that $h(x_*) = 0$, i.e. (5.18b). □

Now we give the proof of Theorem 5.1.

Proof of Theorem 5.1. Let δ_* be any accumulation point of $\{\delta_k\}$. Without loss of generality, we can assume that $\{\delta_k\}$ converges to δ_*. We have $\delta_k \leq \Delta_k$. First, assume that there exists a subsequence $\{\delta_k, k \in N \subset \mathbb{N}\}$ such that $\delta_k = \Delta_k$ for all $k \in N$, in which case we have $\delta_* \geq \Delta_{min}$. Consequently (5.22) holds for $\delta(x_*, B_*, \Delta_*) = \frac{1}{2}\Delta_{min}$. Now we assume that $\delta_k < \Delta_k$ for all sufficiently large k, which implies that (Δ_k, β_k) never determines an acceptable step. Let $\bar{s_k} \neq 0$ be the last non acceptable step. Observe that $\bar{s_k}$ is an $\bar{\epsilon_k}(\bar{s_k}, \bar{\eta_k})$-solution of the local subproblem

$$(RTRQP) \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2}s^T B_k s \\ \text{subject to} & \bar{\alpha_k} h(x_k) + \nabla h(x_k)^T s = 0 \\ & ||s||_p \leq \bar{\delta_k} \end{cases}$$

for some $0 < \bar{\eta_k} \leq \beta_k$, $\delta_k < \bar{\delta_k} \leq \Delta_k$ and the corresponding $\bar{\alpha_k}$. We have

$$c_3||\bar{s_k}||_p \leq \delta_k \leq c_4||\bar{s_k}||_p.$$

Suppose that $\delta_* = 0$. Then the sequence $\{\bar{s_k}\}$ converges to zero. From Lemma 5.3, we obtain

$$h(x_*) = 0, \tag{7.12}$$

$$\lim_{k\to+\infty} \frac{||h(x_k)||}{||\bar{s_k}||_p} = 0, \tag{7.13}$$

and

$$\liminf_{k\to+\infty} \nabla f(x_k)^T \frac{\bar{s_k}}{||\bar{s_k}||_p} \geq 0. \tag{7.14}$$

Let $s \in \mathbb{R}^n$ satisfy

$$\nabla h(x_*)^T s = 0 \quad \text{and} \quad ||s||_p = 1 \tag{7.15}$$

To obtain a contradiction to our assumption that x_* is not a Karush-Kuhn-Tucker point of (EQCP), we need to show that

$$\nabla f(x_*)^T s \geq 0$$

holds. Let us construct a feasible point for $(RTRQP)$, say w_k, such that

$$\lim_{k'+\infty} \frac{w_k}{||w_k||_p} = s.$$

The consistency of the linearized constraints, i.e. $\nabla h(x_k) + h(x_k) = 0$, implies that $V_{k,2}^T h(x_k) = 0$ and $V_{k,2}^T \nabla h(x_k)^T = 0$. Therefore we have

$$0 = V_{k,2}^T \Big[-\bar{\alpha_k} h(x_k) - \frac{||\bar{s_k}||_p}{2} \nabla h(x_k)^T s\Big]. \tag{7.16}$$

Let v_k be defined by

$$v_k = U_{k,1} \textstyle\sum_{k,1}^{-1} \left[-\bar{\alpha_k} V_{k,1}^T h(x_k) - \frac{||\bar{s_k}||_p}{2} V_{k,1}^T \nabla h(x_k)^T s\right]. \tag{7.17}$$

From (7.16) and (7.17), we obtain

$$\nabla h(x_k)^T v_k = -\bar{\alpha_k} h(x_k) - \frac{||\bar{s_k}||_p}{2} \nabla h(x_k)^T s. \tag{7.18}$$

Let us define

$$w_k = v_k + \frac{||\bar{s_k}||_p}{2} s, \tag{7.19}$$

and rewrite (7.18) as

$$\bar{\alpha_k} h(x_k) + \nabla h(x_k)^T w_k = 0. \tag{7.20}$$

This implies that w_k satisfies the linear constraints of $(RTRQP)$. Let us establish that it also satisfies the trust-region constraint. From (7.17) and (5.9), we obtain

$$\frac{||v_k||_2}{||\bar{s_k}||_p} \leq \frac{1}{\sigma_*} \Big[\frac{||h(x_k)||_2}{||\bar{s_k}||_p} + \frac{1}{2} ||\nabla h(x_k)^T s||_2\Big], \tag{7.21}$$

and hence, using (7.13) and (7.15),

$$\lim_{k\to+\infty} \frac{||v_k||_p}{||\bar{s_k}||_p} = 0. \tag{7.22}$$

Now let us obtain $lim_{k\to+\infty} w_k/||w_k||_p$. We have

$$\frac{w_k}{||w_k||_p} = \frac{w_k}{||\bar{s_k}||_p} \left(\frac{||w_k||_p}{||\bar{s_k}||_p}\right)^{-1}, \tag{7.23}$$

where

$$\frac{w_k}{||\bar{s_k}||_p} = \frac{v_k}{||\bar{s_k}||_p} + \frac{1}{2} s.$$

which, together with (7.22), implies that

$$\lim_{k\to+\infty} \frac{w_k}{\|\bar{s_k}\|_p} = \frac{1}{2}s. \tag{7.24}$$

Since $\|s\|_p = 1$, from (7.23) and (7.24) we conclude that

$$\lim_{k\to+\infty} \frac{w_k}{\|w_k\|_p} = s. \tag{7.25}$$

From (7.24), we obtain

$$\lim_{k\to+\infty} \frac{\|w_k\|_p}{\|\bar{s_k}\|_p} = \frac{1}{2},$$

which implies that, for sufficiently large k, we have

$$\|w_k\|_p \le \|\bar{s_k}\|_p \le \bar{\delta_k} \tag{7.26}$$

i.e. w_k satisfies the trust-region constraint of the subproblem RTRQP, and

$$\frac{1}{3}\|\bar{s_k}\|_p \le \|w_k\|_p. \tag{7.27}$$

Finally, we are ready to establish that $\nabla f(x)^T s \ge 0$. Because $\bar{s_k}$ is an $\bar{\epsilon_k} = \epsilon_k(\bar{s_k}, \bar{\eta_k})$-solution of the subproblem RTRQP and because w_k is a feasible point for this subproblem, we have

$$\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} \le \nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k + \bar{\epsilon_k} \tag{7.28}$$

First let us assume that for all sufficiently large k, we have

$$\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k \ge 0. \tag{7.29}$$

This implies that

$$\nabla f(x_k)^T \frac{w_k}{\|w_k\|_p} + \frac{1}{2} w_k^T B_k \frac{w_k}{\|w_k\|_p} \ge 0. \tag{7.30}$$

Since $\{w_k\}$ converges to zero, $\{x_k\}$ converge to x_*, and $\{B_k\}$ is bounded, we obtain from (7.25) and (7.30) that

$$\nabla f(x_*)^T s \ge 0. \tag{7.31}$$

Now let assume that there exists a subsequence $\{x_k, k \in N \subset \mathbb{N}\}$ for which

$$\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k < 0. \tag{7.32}$$

From (7.28) and (7.32) we obtain

$$\frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} - \bar{\epsilon_k}}{\|w_k\|_p} \le \frac{\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k}{\|w_k\|_p} < 0, \tag{7.33}$$

which, together with (7.27), implies

$$3\frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} - \bar{\epsilon_k}}{\|\bar{s_k}\|_p} \le \frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} - \bar{\epsilon_k}}{\|w_k\|_p}. \tag{7.34}$$

Inequalities (7.33) and (7.34) imply that

$$(7.35) \qquad 3\frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k}}{||\bar{s_k}||_p} \leq \frac{\nabla f(x_k)^T w_k + \frac{1}{2}w_k^T B_k w_k}{||w_k||_p} + 3\bar{\eta_k}\max\Big(1, \frac{||h(x_k)||}{||\bar{s_k}||_p}\Big)$$

Therefore, because ∇f is continuous, $\{B_k\}$ is bounded, $\{w_k\}$, $\{\bar{s_k}\}$, and $\{\bar{\eta_k}\}$ converge to zero, and $\{x_k\}$ converges to x_*, we obtain form (7.35), (7.25), (7.13), and (7.14) that inequality (7.31) holds also in the case (7.32). Therefore in both cases (7.29) and (7.32), we obtain that

$$\nabla f(x_*)^T s \geq 0.$$

Now, since this inequality holds for any s such that (7.18) holds, i.e. $\nabla h(x_*)^T s = 0$, and because of (7.12), i.e. $h(x_*) = 0$, we conclude from the Farkas Lemma that x_* is a Karush-Kuhn-Tucker point of (EQCP), which contradicts the hypothesis of the theorem. Therefore there exists a positive scalar $\delta(x_*, B_*, \Delta_*)$ such that

$$\delta_* \geq \delta(x_*, B_*, \Delta_*)$$

holds for any accumulation point δ_* of $\{\delta_k\}$, where δ_k determines an acceptable step at the k^{th} iteration.□

Acknowledgment. This work was done partially 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. Finally he thanks the Council for International Exchange of Scholars (CIES) for its administrative assistance.

References

[1] N. ALEXANDROV, *Multilevel Algorithms for Nonlinear Equations and Equality Constrained Optimization*, Ph.D. thesis, Rice University, Tech. Rep. TR93-20, 1993.

[2] P.T. BOGGS, P.D. DOMOCH, and J.E. ROGERS, *An interior-point method for general large-scale quadratic programming problems*, NISTIR 5406 (January 1995), Applied and Computational Mathematics Division, National Institute of Standards and Technology, Gaithersburg, MD 20899.

[3] R. H. BYRD, E. O. OMOJOKUN, R.B.SCHNABEL, G.A. SHULTZ, *Robust trust-region methods for nonlinearly constrained optimization*, Presented at 1987 SIAM Conference on Optimization, Houston,TX, 1987.

[4] R. H. BYRD, R.B.SCHNABEL, G.A. SHULTZ, *A trust-region algorithm for nonlinearly constrained optimization*, SIAM J. Numer. Anal., 24 (1985), pp. 1152-1170.

[5] M. R. CELIS, J. E. DENNIS, and R. A. TAPIA, *A trust-region strategy for nonlinear equality constrained optimization*, in Numerical Optimization 1984, P. Boggs, R. Byrd, and R. Schnabel, eds., Proceedings 20, Society for industrial and Applied Mathematics, Philadelphia, 1985, pp. 71-82.

[6] R. M. CHAMBERLAIN, C. LEMARECHAL, H. C. PEDERSEN, *The watchdog technique for forcing convergence in algorithms for constrained optimization*, in A. G. Buckley and J.-L. Goffin, Algorithms for constrained minimization of smooth nonlinear functions, Mathematical Programming Study 16, North Holland, Amsterdam, 1982, pp.1-17.

[7] T. F. COLEMAN, and A. R. CONN, *Nonlinear programming via an exact penalty function: asymptotic analysis*, Math. Programming, 24 (1982), pp. 123-136. (1982).

[8] A. DAX, *A new Theorem of the alternative*, Math. Programming, 47(1990), pp.297-299.

[9] J. E. DENNIS, M. EL-ALEM, and M. C. MACIEL, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization*, Tech. Rep. TR92-28 (1992), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892.

[10] M. EL-ALEM, *A global convergence theory for a class of trust-region algorithm for constrained optimization*, Ph. D. thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1988.

[11] M. EL-ALEM, *A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization*, SIAM J. Numer. Anal., 28 (1991), pp. 266-290.

[12] M. EL-ALEM, *Robust trust-region algorithm with non monotonic penalty parameter scheme for constrained optimization*, Tech. Rep. CRPC-TR92245, Center for Research on Parallel Computation, Rice University, Houston, Texas.

[13] M. EL HALLABI, *An Inexact trust-region algorithm: Globalization of Newton's method*, Tech. Rep. TR93-41(1993), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892

[14] M. EL HALLABI, *A necessary and sufficient condition for intersecting a translated hyperplane and a ball*, Tech. Rep. TR95-14(1995), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892

[15] M. EL HALLABI, and R. TAPIA, *A Global convergence theory for arbitrary norm trust-region methods for nonlinear equations*,Tech. Rep. TR93-41(1993), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892

[16] R. FLETCHER, *Second order corrections for non differentiable optimization*, in G. A. Watson, Numerical Analysis Proceedings, Dundee 1981, Lectures Notes in Mathematics 912, Springer Verlag, Berlin, 1982, pp. 85-113. 1982.

[17] R. FLETCHER, *Practical Optimization*, John Wiley and Sons, New York, 1987.

[18] G. H. GOLUB, and C. F. VAN LOAN, *Matrix Computation*, John Hopkins University Press, Baltimore, Maryland, 1983.

[19] P. HUARD, *Optimization algorithms and Point-to-set maps*, Math. Programming, 8 (1975), pp. 308-331.

[20] P. HUARD, *Extensions of Zangwill's Theorem*, Mathematical Programming Study 10 , North Holland Publishing, Amsterdam, (1979), pp. 308-331.

[21] J. JAHN *Introduction to the Theory of nonlinear optimization*, Springer-Verlag, Berlin, 1994.

[22] M.C. MACIEL, *A global convergence theory for a general class of trust region algorithm for equality constrained optimization*, Ph. D. thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1992.

[23] N. MARATOS, *Exact penalty function for finite dimensional and control optimization problems*, PhD thesis, University of London, London, UK, 1978.

[24] L. MARUCHA, J. NOCEDAL, and T. PLANTEGA, *On the implementation of an algorithm for large-scale equality optimization*, Tech. Rep. NAM 09-93, Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois, 1993.

[25] 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.

[26] M.J.D. POWELL, *Convergence properties of a class of minimization algorithms*, in O.L. Mangasarian, R.R. Meyer, S.M. Robinson, eds., Nonlinear Programming 2, Academic Press, New York, 1975, pp. 1-27

[27] 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.

[28] M.J.D. POWELL, *On the global convergence of trust-region algorithms for unconstrained minimization*, Math. Programming, 29 (1983), pp. 297-303.

[29] M. J. D. POWELL and Y. YUAN, *A trust-region region algorithm for equality constrained optimization*, Math. Programming, 49 (1991), pp. 189-211.

[30] F. RENDL, and H. WOLKOWICZ, *A semidefinite framework for trust region subproblems with applications to large scale minimization*, CORR Report 94-32, Technische Universitaat Gras, Institut fur Mathematik, Kopernikusgasse 24, A-8010 Gras, Austria.

[31] R.T. ROCKAFELLAR, *Convex analysis.* Princeton University Press, Princeton, 1970.

[32] R.T. ROCKAFELLAR, *The Theory of subgradients and its applications to problems of optimization: convex and nonconvex functions.* Heldermann, Berlin, 1981.

[33] D. SORENSON, *Minimization of a large scale quadratic function subject to ellipsoidal constraint*, Tech. Rep. TR94-27 (1994), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892

[34] R. A. TAPIA, *Newton's method for optimization problems with equality constraints*, SIAM J. on Numer. Anal. 11 (1974), pp. 874-886.

[35] R. A. TAPIA, *Newton's method for problems with equality constraints*, SIAM J. on Numer. Anal. 11 (1974), pp. 174-196.

[36] R. A. TAPIA, J. *Diagonalized multiplier methods and quasi-Newton methods for constrained optimization* J. Optim. Theory Appl., 22 (1977), pp. 135-194.

[37] R. A. TAPIA, *Quasi-Newton methods for equality constrained optimization: equivalence of existing methods and a new implementation.* In O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, editors, Nonlinear Programming 3, Academic Press, New York, 1978, pp.125-164.

[38] A. VARDI, *A trust-region algorithm for equality constrained minimization and implementation.* SIAM J. Numer. Anal., 22 (1985), pp.575-591.