A Global Convergence Theory for SLP and SQP Trust-Region Algorithms

M. El Hallabi

CRPC-TR94371
January 1994

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

A GLOBAL CONVERGENCE THEORY FOR SLP AND SQP TRUST-REGION ALGORITHMS[1,2,3]

M. EL HALLABI[4]

Abstract. In this paper, we propose a trust-region algorithm to minimize a nonlinear function $f : \mathbb{R}^n \to \mathbb{R}$ subject to nonlinear equality constraints $h_i(x) = 0, i = 1, \cdots, m$ where $h_i : \mathbb{R}^n \to \mathbb{R}$. We are concerned with the fact that n and m may be large. We adopt the approach taken in Vardi (1985). We also replace the ℓ_2-norm in the trust-region constraint by either a polyhedral norm ℓ_1 or ℓ_∞, an arbitrary ℓ_p-norm with $p \geq 2$, or an arbitrary convex combination of these norms. In particular, when polyhedral norms are used, the algorithm can be viewed as a sequential quadratic programming method or a sequential linear programming method regarding on whether or not we use second order information in the local model subproblem. 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 systems of linearized constraints are consistent. Also, we assume that the functions f and $h_i, i = 1 \cdots m$, are only continuously differentiable. We demonstrate that any accumulation point of the iteration sequence, obtained from a remote starting point, is a Karush-Kuhn-Tucker point of the constrained minimization problem. This convergence theory follows from very powerful and natural properties of the trust-region strategy, for example a property we call *local uniform decrease*.

Key Words: SLP, SQP, Global Convergence, Constrained Optimization, Consistency , Non Regularity, Equality Constrained , Trust-Region , Local Uniform Decrease.

AMS subject classifications. 65K05, 49D37

1. Introduction. In this paper we present an algorithm for approximating a solution of the equality constrained optimization 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 Lagrangian function $l : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ associated with problem (EQCP) is defined by

$$l(x, \lambda) = f(x) + \lambda^T h(x), \tag{1.1}$$

where λ is the vector of Lagrange multipliers.

[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.

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

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

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 (1987)[19], Tapia (1974)[32] and [33], (1977)[34], (1978)[35]).

The problem of global convergence has been given much consideration recently. Global convergence results are given in Vardi(1985)[36], Byrd, Schnabel, and Shultz (1987)[4], El-Alem (1988)[12], (1991)[13], and (1992)[14], Powell and Yuan (1991)[30], Maciel (1992)[23], Dennis, El-Alem, and Maciel (1992)[10], and Alexandrov (1993)[1]. Except for Vardi (1985)[36], all the proposed algorithms are either of the framework of Celis, Dennis, and Tapia (1985)[5], or of the framework of Byrd , Omojokun, Schnabel, and Shultz (1987)[3].

Because the trust-region strategy had proven to be a very successful tool for designing globally convergent algorithms for unconstrained optimization (e.g. Powell (1975)[27] and (1983)[29]) and for nonlinear systems of equations (e.g. El Hallabi and Tapia (1993)[16], El Hallabi (1993)[17], and Powell (1983)[28]), it was quite natural to extend this strategy to constrained optimization. The obvious extension is to add a trust-region constraint to the subproblem (QP) to obtain

$$(TRQP) \quad \equiv \quad \begin{cases} \text{minimize}_{s \in \mathbb{R}^n} & c_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & h(x_k) + \nabla h(x_k)^T s = 0, \\ & \|s\|_2^2 \le \delta_k^2, \end{cases}$$

where $0 < \delta_k$ is the trust-region radius. But, unless $h(x_k) = 0$, problem (TRQP) may have inconsistent constraints. To overcome this difficulty, Vardi (1985)[36] proposed shifting the linearized equality constraints, which led to the following Relaxed Trust-Region Quadratic Programming subproblem

$$(RTRQP) \quad \equiv \quad \begin{cases} \text{minimize} & c_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \alpha_k h(x_k) + \nabla h(x_k)^T s = 0, \\ & \|s\|_2^2 \le \delta_k^2, \end{cases}$$

where $0 < \alpha_k \le 1$, the relaxation parameter, is chosen such that the feasible region of (RTRQP) is not empty. Because there was not a straightforward way of choosing the relaxation parameter α_k, Celis, Dennis, and Tapia (1985)[5] considered obtaining a trial step s_k as a solution of the subproblem

$$(CDT) \quad \equiv \quad \begin{cases} \text{minimize}_{s \in \mathbb{R}^n} & c_k^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \|h(x_k) + \nabla h(x_k)^T s\|_2^2 \le \theta_k, \\ & \|s\|_2^2 \le \delta_k, \end{cases}$$

where $0 < \theta_k$ is chosen to realize some predicted decrease in the ℓ_2-norm of the linearized constraint $h(x_k) + \nabla h(x_k)^T s$ inside the ball of radius δ_k . In practice it is hard to solve the (CDT) subproblem.

Byrd, Omojokun, Schnabel, and Shultz (1987)[3] proposed a two-level algorithm where the trial step is of the form $s_k = s_k^n + s_k^t$ where s_k^n, the normal component of

s_k, is a solution of

$$(N-STEP) \quad \equiv \quad \begin{cases} \text{minimize}_{s\in\mathbb{R}^n} & \|h(x_k)+\nabla h(x_k)^T s\|_2^2 \\ \text{subject to} & \|s\|_2^2 \leq \tau\delta_k^2, \end{cases}$$

where $0<\tau<1$, and s_k^t, the tangent component of s_k, is of the form $s_k^t = Z_k u_k$, where Z_k ia a basis for the null space of $\nabla h(x_k)$, and u_k is a solution of

$$(T-STEP) \quad \equiv \quad \begin{cases} \text{minimize}_{u\in\mathbb{R}^{n-m}} & (\nabla f(x_k)+Z_k s_k^n)^T Z_k u + \frac{1}{2}u^T Z_k^T B_k Z_k u \\ \text{subject to} & \|u\|_2^2 \leq \delta_k^2 - \|s_k^n\|_2^2. \end{cases}$$

A very recent implementation of the two level algorithm is described in Marucha, Nocedal, and Plantega (1993)[25]

To the best knowledge of the author, all 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 only that the linearized constraints are consistent. We also 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 (1985)[36], i.e. we use subproblem (RTRQP) as our local model subproblem. However, we replace the ℓ_2-norm in the trust-region constraint by a polyhedral norm, an arbitrary ℓ_p-norm with $p\geq 2$, or a convex combination of these norms. In particular, when polyhedral norms are used, our method can be considered as a sequential quadratic programming method or a sequential linear programming method depending on whether or not we use second order informations in the local model subproblem.

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

$$Ared_k(s) = \Phi(\mu_k, x_k; s) - \Phi(\mu_k, x_k; 0) \tag{1.1}$$

and the predicted reduction

$$Pred_k(s) = \Psi(\mu_k, x_k; s) - \Psi(\mu_k, x_k; 0) \tag{1.2}$$

where

$$\Phi(\mu, x; s) = f(x+s) + \mu\|h(x+s)\| \tag{1.3}$$

is the merit function approximated by

$$\Psi(\mu, x; s) = f(x) + \nabla f(x)^T s + \frac{1}{2}s^T Bs + \mu\|h(x)+\nabla h(x)^T s\|. \tag{1.4}$$

In (1.3) and (1.4), μ denotes the penalty parameter, and $\| \quad \|$ denotes an arbitrary (but fixed) norm on $\mathbb{R}^m$.

In Section 2, we give a sufficient condition for the relaxation parameter α_k to define a nonempty feasible region for (RTRQP). In Section 3, we extend to problem (EQCP) the characterization of stationarity given in terms of minimizers of local models in El Hallabi and Tapia (1993)[16] for unconstrained minimization. In Section 4, we define the Arbitrary Norm Inexact Trust-Region Algorithm (ANITRA), and we show that the penalty parameter fits well with the objective function and the constraints. In Section 5, we prove, under rather weak assumptions, and by establishing some powerful properties of the trust-region strategy such as the one we call *local uniform decrease*, that any accumulation point of the sequence generated by the ANITRA algorithm, from a remote starting point x_0, is a Karush-Kuhn-Tucker point of (EQCP). Extensive use of the well known Farkas Lemma is made throughout this section. We end this paper by giving some concluding remarks in Section 6.

2. Linearized Constraint Relaxation. In this section we give a sufficient condition for the relaxation parameter α to define a nonempty feasible region for the subproblem

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

where $\| \ \|_p$ is any ℓ_p-norm on $\mathbb{R}^n$. This condition is stated in the following proposition.

PROPOSITION 2.1. *Consider x in $\mathbb{R}^n$ such that*

$$h(x) \neq 0 \tag{2.1a}$$

and such that the linear system

$$h(x) + \nabla h(x)^T s = 0, \tag{2.1b}$$

is consistent. Let $r_x \leq m$ be the rank of the matrix $\nabla h(x)$, and let ω_x be a positive lower bound of the positive generalized eigenvalues of $\nabla h(x)$. Also let p and q be extended reals satisfying

$$\frac{1}{p} + \frac{1}{q} = 1 \tag{2.2}$$

i.e. $(p = 1, q = +\infty)$ and $(p = +\infty, q = 1)$ are allowed, and let ν_q satisfy

$$\| \quad \|_q \geq \nu_q \| \quad \|_2 \tag{2.3}$$

where $\| \quad \|_q$ denotes the ℓ_q vector norm, and $\| \quad \|_2$ denotes the ℓ_2 vector norm. If

$$0 < \alpha \leq \min\Big(1, \nu_q \frac{\delta\, \omega_x}{\|h(x)\|_2}\Big), \tag{2.4}$$

then the subset

$$\mathcal{F}_p(x) = \Big\{ s \in \mathbb{R}^n \ \Big| \ \alpha h(x) + \nabla h(x)^T s = 0, \|s\|_p \leq \delta \Big\} \tag{2.5}$$

is not empty.

The proof of this proposition requires the following two lemmas, whose proofs detract from the matter at hand and will be given immediately following the proof of Proposition 2.1.

LEMMA 2.1. *Assume the hypotheses of Proposition 2.1. If*

$$0 \leq \alpha \leq \nu_q \frac{\delta\, \omega_x}{\|h(x)\|_2}, \tag{2.6}$$

then the symmetric matrix

$$M(x) = \nabla h(x)^T \nabla h(x) - \left(\frac{\alpha}{\delta\, \nu_q}\right)^2 h(x)h(x)^T \tag{2.7}$$

is positive semi-definite.

LEMMA 2.2. *Assume the hypotheses of Proposition 2.1. If the symmetric matrix $M(x)$ in (2.7) is positive semi-definite, then the subset in (2.5), i.e.*

$$\mathcal{F}_p(x) = \left\{ s \in \mathbb{R}^n \;\middle|\; \alpha h(x) + \nabla h(x)^T s = 0, \|s\|_p \leq \delta \right\}$$

is not empty.

Proof of Proposition 2.1. Assume that (2.4) holds. Then obviously (2.6) also holds. Consequently, from Lemma 2.1, we obtain that the symmetric matrix $M(x)$ defined in (2.7) is positive semi-definite, which, by Lemma 2.2, implies that the subset $\mathcal{F}_p(x)$ defined in (2.5) is not empty.□

Now we give the proofs of Lemmas 2.1 and 2.2.

Proof of Lemma 2.1. Let $\nabla h(x) = U_x \sum_x V_x^T$ be the singular value decomposition of $\nabla h(x)$, (see Golub and Van Loan (1983)[20]), with

$$\sum_x = \begin{pmatrix} \sum_{x,1} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \sum_{x,1} = diag(\sigma_{x,1}, \cdots, \sigma_{x,r_x}),$$

where $\sigma_{x,i}, i = 1, \cdots, r_x$ are the positive generalized eigenvalues of $\nabla h(x)$,

$$U_x = (\, U_{x,1} \quad U_{x,2}\,), \quad \text{with} \quad U_{x,1} \in \mathbb{R}^{n\times r(x)} \quad , \quad U_{x,2} \in \mathbb{R}^{n\times(n-r(x))},$$

and

$$V_x = (\, V_{x,1} \quad V_{x,2}\,), \quad \text{with} \quad V_{x,1} \in \mathbb{R}^{m\times r(x)} \quad , \quad V_{x,2} \in \mathbb{R}^{m\times(m-r(x))}.$$

First we show that

$$V_{x,2}^T h(x) = 0. \tag{2.8}$$

Let s be any solution of the linear system (2.1b). We have

$$V_x^T h(x) + \sum_x^T U_x^T s = 0,$$

or equivalently

$$V_{x,1}^T h(x) + \sum_{x,1}^T U_{x,1}^T s = 0, \tag{2.9}$$

and

$$V_{x,2}^T h(x) = 0$$

which is (2.8). Now, we prove that the symmetric matrix in (2.7) is positive semi-definite. We have

$$\begin{aligned}\nabla h(x)^T \nabla h(x) - \left(\tfrac{\alpha}{\delta\,\nu_q}\right)^2 h(x)h(x)^T &= V_x \textstyle\sum_x^T \sum_x V_x^T - \left(\tfrac{\alpha}{\delta\,\nu_q}\right)^2 h(x)h(x)^T \\ &= V_x \left[\textstyle\sum_x^T \sum_x - \left(\tfrac{\alpha}{\delta\,\nu_q}\right)^2 V_x^T h(x)(V_x^T h(x))^T\right] V_x^T.\end{aligned}$$

On the other hand, using (2.8), we obtain

$$\begin{aligned}\textstyle\sum_x^T \sum_x - \left(\tfrac{\alpha}{\delta\,\nu_q}\right)^2 V_x^T h(x)(V_x^T h(x))^T &= \begin{pmatrix} (\sum_{x,1})^2 & 0 \\ 0 & 0 \end{pmatrix} \\ &\quad - \left(\tfrac{\alpha}{\delta\,\nu_q}\right)^2 \begin{pmatrix} V_{x,1}^T h(x) \\ 0 \end{pmatrix} \begin{pmatrix} V_{x,1}^T h(x) \\ 0 \end{pmatrix}^T \\ &= \begin{pmatrix} H_{x,1} & 0 \\ 0 & 0 \end{pmatrix},\end{aligned}$$

where

$$H_{x,1} = \left(\textstyle\sum_{x,1}\right)^2 - \left(\frac{\alpha}{\delta\,\nu_q}\right)^2 \left(V_{x,1}^T h(x)\right)\left(V_{x,1}^T h(x)\right)^T. \tag{2.10}$$

Therefore, to prove that the symmetric matrix $M(x)$ is positive semi-definite, it is sufficient to prove that the symmetric matrix $H_{x,1}$ in (2.10) is positive semi-definite.

The matrix $H_{x,1}$ results from the shifting of the matrix $\left(\sum_{x,1}\right)^2$ by the rank one matrix $\left(\frac{\alpha}{\delta\,\nu_q}\right)^2\left(V_x^T h(x)\right)\left(V_x^T h(x)\right)^T$. Therefore, if the ℓ_2-norm of this rank one matrix is smaller that ω_x^2, i.e.

$$0 \le \left(\frac{\alpha}{\delta\,\nu_q} \|V_{x,1}^T h(x)\|_2\right)^2 \le \omega_x^2,$$

which is, by (2.8), equivalent to (2.6), then the matrix $H_{x,1}$ is symmetric positive definite.□

Now we prove Lemma 2.2.

Proof of Lemma 2.2. By Theorem 1 of the alternative of Dax (1990)[9] (see Appendix B), the subset $\mathcal{F}_p(x)$ is not empty if and only if

$$\delta \|\nabla h(x) y\|_q \ge \alpha h(x)^T y \tag{2.13}$$

holds for all $y \in \mathbb{R}^m$.

Assume that the symmetric matrix $M(x)$ in (2.7) is positive semi-definite. Then

$$y^T \left[\nabla h(x)^T \nabla h(x) - \left(\frac{\alpha}{\delta\,\nu_q}\right)^2 h(x)h(x)^T\right] y \ge 0$$

or equivalently

$$(\delta\,\nu_q)^2 y^T \nabla h(x)^T \nabla h(x) y - \alpha^2 y^T h(x)h(x)^T y \ge 0$$

holds for all $y \in \mathbb{R}^m$. Therefore we have

$$\delta\, \nu_q \|\nabla h(x) y\|_2 \geq |\alpha h(x)^T y|, \quad \forall y \in \mathbb{R}^m,$$

which implies that

$$\delta\, \nu_q \|\nabla h(x) y\|_2 \geq \alpha h(x)^T y, \quad \forall y \in \mathbb{R}^m. \tag{2.14}$$

Now, from (2.3), i.e. $\|\ \|_q \geq \nu_q \|\ \|_2$, and (2.14), we obtain (2.13), which implies that the subset $\mathcal{F}_p(x)$ is not empty.□

In some applications (see El-Alem and Tapia (1993)[15]), the trust-region constraint is a convex combination of ℓ_p-norms. To include this case, we give the following corollary.

COROLLARY 2.1. *Assume the hypotheses of Proposition 2.1. Let p_i and q_i be extended reals and ν_i be a positive scalar such that*

$$\|\ \|_{q_i} \geq \nu_i \|\ \|_2 \quad \text{and} \quad \frac{1}{p_i} + \frac{1}{q_i} = 1,$$

where $\|\ \|_{q_i}$ is the ℓ_{q_i} vector norm for $i = 1 \cdots l$. Let $\gamma_i, i = 1 \cdots l$ be positive reals satisfying $\sum_{i=1}^{l} \gamma_i = 1$. Also let $\nu = \min_{1 \leq i \leq l} \nu_i$. If

$$0 < \alpha \leq \nu \frac{\delta\, \omega_x}{\|h(x)\|_2}$$

then the subset of $\mathbb{R}^n$

$$\mathcal{F} = \{s \in \mathbb{R}^n \mid \alpha h(x) + \nabla h(x)^T s = 0, \quad \sum_{i=1}^{l} \gamma_i \|s\|_{p_i} \leq \delta\}$$

is not empty.

Proof. By Proposition 2.1, $\mathcal{F}_{p_i}(x)$ is not empty, for $i = 1, \cdots, l$. Let $p_j = \min_{1 \leq i \leq l} p_i$. It is obvious that the p_j-ball of radius δ is contained in the p_i-ball of the same radius for $i = 1 \cdots l$. Let z_j be in $\mathcal{F}_{p_j}(x)$. We have

$$\|z_j\|_{p_i} \leq \|z_j\|_{p_j} \leq \delta \quad \forall i = 1 \cdots l.$$

Therefore we obtain

$$\sum_{1 \leq i \leq l} \gamma_i \|z_j\|_{p_i} \leq \sum_{1 \leq i \leq l} \gamma_i \delta = \delta,$$

and hence the subset

$$\{s \in \mathbb{R}^n \mid \alpha \nabla h(x)^T s + h(x) = 0, \quad \sum_{1 \leq i \leq l} \gamma_i \|s\|_{p_i} \leq \delta\}$$

is not empty. □

Proposition 2.1 implies that the choice of the equality constraint relaxation parameter α is practical only if we have a way of computing a lower bound ω_k of the positive generalized eigenvalues of $\nabla h(x_k)$ for the current iterate x_k. It is obvious that this is exactly the rank determination problem, and is a hard problem. But the QR-decomposition with column pivoting is an acceptable solution. In Appendix A, we propose a way for obtaining ω_k by using the QR-decomposition with column pivoting.

3. Characterization of stationary points of problem (EQCP). In unconstrained optimization, the notion of stationarity can be defined in terms of minimizers of the local model subproblem. This was done in El Hallabi and Tapia (1993)[16], where the local model subproblem was convex. That notion was that a given point x is stationary for the objective function if zero solves the local model subproblem. In El Hallabi and Tapia (1993)[16], this notion was shown to be equivalent to saying that there is no descent direction of the objective function at x. In the present research, the second order approximation matrices may not be positive semi-definite; hence the subproblem may not be convex and the El-hallabi-Tapia theory does not apply. Therefore, we give a characterization of stationarity for problem (EQCP) in terms of local minimizers of subproblem $(RTRQP)$.

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

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

then x_k is necessarily a Karush-Kuhn-Tucker point of (EQCP).

Proof. Assume that $s_k = 0$ is a local solution of $(RTRQP)$. Let $s \in \mathbb{R}^n$ be any point such that

$$\nabla h(x_k)^T s = 0. \tag{3.2}$$

Then for sufficiently small positive t, the point ts is feasible for subproblem $(RTRQP)$ and sufficiently close to the local solution $s_k = 0$. Therefore we have

$$\nabla f(x_k)^T s + \frac{1}{2} t s^T B_k s \ge 0,$$

for all sufficiently small positive t, which implies that

$$\nabla f(x_k)^T s \ge 0. \tag{3.3}$$

¿From (3.2), (3.3), and the well known Farkas Lemma, we conclude that x_k is a Karush-Kuhn-Tucker point of (EQCP). □

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 in the sense that the predicted decrease in the merit function is bounded above by the sum of the predicted decreases in the objective function and in the constraints considered separately.

At each iteration, we solve a local model subproblem of the form

$$(RTRQP) \quad \equiv \begin{cases} \text{minimize} & \nabla f(x)^T s + \frac{1}{2} s^T B s \\ \text{subject to} & \alpha h(x) + \nabla h(x)^T s = 0 \\ & \|s\|_p \le \delta \end{cases},$$

for some fixed (x, B, α, δ), and within some tolerance ϵ 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,*

$$\nabla f(x)^T s_\epsilon + \frac{1}{2} s_\epsilon^T B s_\epsilon \leq \nabla f(x)^T s + \frac{1}{2} s^T B s + \epsilon \tag{4.1}$$

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

$$\nabla f(x)^T s_\epsilon + \frac{1}{2} s_\epsilon^T B s_\epsilon < 0. \tag{4.2}$$

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

$$\epsilon_k = \eta_k \begin{cases} \min(\|s_k\|_p, \alpha_k \|h(x_k)\|) & \text{if } h(x_k) \neq 0 \\ \|s_k\|_p & \text{otherwise} \end{cases} \tag{4.3}$$

for some $0 < \eta_k$ that will set by the algorithm.

To accept or reject the trial step s_k, we will use the actual reduction $Ared_k(s_k)$ and the predicted reduction $Pred_k(s_k)$ defined in (1.1) and (1.2) respectively. The penalty parameter will be given by the following update scheme.

To define our Arbitrary Norm Inexact Trust-Region Algorithm (ANITRA), it remains to describe our way of updating the penalty parameter μ_k. Usually, in constrained optimization, the use of a penalty parameter answers to concerns. First, the penalty parameter should be set so that the predicted decrease is negative. In our case this first concern is answered by choosing

$$\mu_k \geq \bar{\mu_k} + \rho \quad \forall\ k, \tag{4.4}$$

where ρ is an arbitrary positive constant, and

$$\bar{\mu_k} = \begin{cases} 0 & \text{if } h(x_k) = 0 \\ 2\max(0, \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|}) & \text{otherwise.} \end{cases} \tag{4.5}$$

This property will be demonstrated in Proposition 4.1. Second, for sufficiently large k, the penalty parameter should force the penalty function to become an exact penalty function. This is usually obtained by choosing a penalty parameter that is constant for sufficiently large k. To answer this second concern, we begin by allowing the parameter to only satisfy (3.5) for a maximum of k_{max} iterations, where k_{max} is an arbitrary large integer. If all iterations $k \leq k_{max}$ fail to locate a Karush-Kuhn-Tucker point of (EQCP), then we will force μ_k to become constant for sufficiently large k. In summary, given an arbitrary large integer k_{max},
i) for $k \leq k_{max}$, we choose $\mu_k \geq \bar{\mu_k} + \rho$, and
ii) for $k > k_{max}$, we set

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

where $\bar{\mu_k}$ is defined by (4.5). Observe that the sequence $\{\mu_k, k > k_{max}\}$ is not decreasing.

Definition of the algorithm ANITRA.
Let $c_i, i = 1, \cdots, 5, \tau, \rho, \epsilon_1, \epsilon_2, \beta, and \Delta_{min}$ be constants satisfying

$$\begin{array}{lll} 0 < c_1 < c_2 < 1 & , \ 0 < c_3 < c_4 < 1 & , \ 1 < c_5 \\ 0 < \tau < 1, & , \ 0 < \rho & , \ 0 < \Delta_{min} \\ 0 < \epsilon_1 \ll 1, & , \ 0 < \epsilon_2 \ll 1 & , \ 0 < \gamma < 1, \quad \beta. \end{array}$$

Let p and q be extended reals such that

$$\frac{1}{p} + \frac{1}{q} = 1, \quad p = 1 \text{ or } p \geq 2,$$

i.e. $(p, q) = (+\infty, 1)$ and $(p, q) = (1, +\infty)$ are allowed .

Let $x_0 \in \mathbb{R}^n$ be an arbitrary point, $B_0 \in \mathbb{R}^{n \times n}$ be an arbitrary square matrix, $0 \leq \Delta_0$, $\beta_0 = \beta$, and $\mu_0 = \rho$. Also let k_{max} be a very large integer.

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

<u>STEP 1</u>. Set $\delta_k = \Delta_k, \eta_k = \beta_k$
<u>STEP 2</u>. If $h(x_k) = 0$ set $\alpha_k = 1$ and go to STEP 5,
<u>STEP 3</u>. Choose a positive lower bound $\omega_k = \omega_{x_k}$ of the positive generalized eigenvalues of $\nabla h(x_k)$,
<u>STEP 4</u>. Set

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

<u>STEP 5</u>. If $\delta_k = \Delta_k$, choose a square matrix $B_k \in \mathbb{R}^{n \times n}$,

<u>STEP 6</u>. Obtain an ϵ_k-solution of the subproblem (RTRQP) with $\epsilon_k = \epsilon_k(\eta_k, s_k)$

<u>STEP 7</u>. Update the penalty parameter μ_k

<u>STEP 8</u>. If $Ared_k(s_k) \leq c_1 Pred_k(s_k)$
set $x_{k+1} = x_k + s_k$ and go to STEP 9
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 4.

<u>STEP 9</u>. Choice of Δ_{k+1}
If $Ared_k(s_k) \leq c_2 Pred_k(s_k)$
then
choose Δ_{k+1} satisfying
$\delta_k \leq \Delta_{k+1} \leq \max(\delta_k, c_5||s_k||_p)$
Else
choose Δ_{k+1} satisfying
$c_4||s_k||_p \leq \Delta_{k+1} \leq ||s_k||_p$.
Set $\Delta_{k+1} = \max(\Delta_{k+1}, \Delta_{min})$.

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

REMARK 4.1.
i) The merit function Φ, defined in (1.3), has the drawback of possessing the Maratos effect (see Maratos (1978)[24]). So, to overcome this difficulty, one may use a second order correction before decreasing the trust-region radius δ_k in STEP 8. 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 (1982)[7], Fletcher (1982)[18], or Byrd, Schnabel, and Schultz (1985)[4].
ii) We could update B_k in STEP 10 instead of STEP 5 that appears quite unusual. This organization will be cleared in Section 6 (see An equivalent subproblem). STEP 5 means that, if at the iteration k, the trust-region has been decreased because of a non acceptable step, and since all what has been changed in the local model subproblem are the trust-region and the relaxation parameter α_k, we do not update B_k.
iii) The parameter τ in STEP 4 plays the same role that the parameter τ plays in the subproblem (N - STEP) that approximates the normal component of the solution in the approach of Byrd, Omojokun, Byrd, and Shultz (1987)[3].

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, might be smaller than Δ_{min}. Throughout the paper we will use the following definition.

DEFINITION 4.2. *If for some couple (δ_k, η_k) defined in STEP 1, the test in STEP 8 is satisfied, then we say that (δ_k, η_k) (or δ_k) determines an acceptable step s_k with respect to $(x_k, B_k, \Delta_k, \beta_k)$. Moreover, the iterate $x_{k+1} = x_k + s_k$ will be called a successor of x_k.*

The penalty parameter μ_k fits well with the objective function and the constraints. Indeed, as the following proposition shows, the predicted decrease in the merit function is less than or equal to (can be equal if $k \leq k_{max}$ and μ_k is at its lower bound) the sum of the predicted decrease in the objective function and the predicted decrease in the constraints if they were considered separately.

PROPOSITION 4.1. *Assume that x_k is not a Karush-Kuhn-Tucker point of (EQCP). Then the approximation Ψ of the merit function Φ satisfies*

$$Pred_k(s_k) \leq -\left|\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k\right| - \rho \alpha_k \|h(x_k)\|, \tag{4.6}$$

and consequently

$$Pred_k(s_k) < 0. \tag{4.7}$$

Moreover if s_k is an acceptable step, then

$$Ared_k(s_k) < 0. \tag{4.8}$$

Proof. We have

$$\begin{aligned} \Psi(\mu_k, x_k, s_k) - \Psi(\mu_k, x_k, 0) &= \nabla f(x_k)^T s_k + \tfrac{1}{2} s_k^T B_k s_k \\ &\quad \mu_k [\|h(x_k) + \nabla h(x_k)^T s_k\| - \|h(x_k)\|] \end{aligned}$$

or equivalently

$$Pred_k(s_k) = \nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k - \mu_k \alpha_k \|h(x_k)\|. \tag{4.9}$$

First, we assume that

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k > 0. \tag{4.10}$$

Therefore $h(x_k) \neq 0$ must hold. We have

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

which, together with (4.9), implies that

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

which can be written as (4.6). From (4.6) and (4.10), we obtain (4.7).

Now we assume that

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k \leq 0. \tag{4.11}$$

We have $\mu_k \geq \rho$. Therefore (4.9) implies that

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

or equivalently

$$Pred_k(s_k) \leq -\left| \nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k \right| - \rho \alpha_k \|h(x_k)\|, \tag{4.12}$$

which is (4.6). When $h(x_k) \neq 0$, (4.12) implies that (4.7) holds. On the other hand 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 (4.12), implies that(4.7) holds also in this case . Finally, from (4.7) it is obvious that (4.8) holds whenever s_k is an acceptable step.□

5. Global Convergence. In this section, we demonstrate under rather weak hypotheses that any accumulation point of the sequence generated by the ANITRA Algorithm is a Karush-Kuhn-Tucker point of (EQCP). To obtain this result, we establish useful and important properties of the trust-region strategy, mainly the property we call *local uniform decrease*. In our proofs, we will use extensively the well known Farkas Lemma.

We make the following hypotheses:

H.1) The functions f and $h_i, i \cdots m$, are continuously differentiable,
H.2) The iteration sequence $\{x_k\}$ is bounded,
H.3) The sequence $\{B_k\}$ is bounded,
H.4) The systems of linearized constraints $h(x_k) + \nabla h(x_k)^T s = 0$, are consistent for all k, and
H.5) The sequence $\{\beta_k\}$ used to obtain an approximate solution of the local model converges to zero.

The global convergence result is given by Theorem 5.3.

Generally, to obtain a global convergence result for solving (EQCP), the uniform regularity assumption, i.e. $(\nabla h(x)^T \nabla h(x))^{-1}$ is uniformly bounded on a subset of $\mathbb{R}^n$ containing the iteration sequence $\{x_k\}$ is used. This assumption implicitely provides a unifrorm lower bound for the generalized eigenvalues of $\nabla h(x_k)$. In the present paper, we do not use the regulatity assumption, but we need some uniform lower bound for the generalized eigenvalues of $\nabla h(x_k)$. Therefore, in the following lemma, we show that the lower bound ω_k of the generalized eigenvalues of $\nabla h(x_k)$ that is used in the definition of the equality constraints relaxation parameter α_k can be chosen so that it is bounded away from zero.

LEMMA 5.1. *Assume hypothesis H4. Then there exists a positive constant* ω *such that*

$$\omega \leq \omega_k \tag{5.1}$$

holds for all $k \in \mathbb{N}$.

Proof. The proof of this lemma depends on the choice we made to obtain ω_k in Appendix A, i.e. the QR-decomposition with column pivoting. Therefore, it will be also give in that appendix.

In the following lemma and its corollary, we demonstrate that the penalty parameter μ_k is constant for sufficiently large k. First we prove that $\bar{\mu_k}$ is uniformly bounded.

LEMMA 5.2. *Assume hypotheses H1 and H2. Then the sequence* $\{\bar{\mu_k}\}$ *defined by (4.5) is bounded.*

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

$$\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k > 0. \tag{5.2}$$

Observe that this excludes the case where $h(x_k) = 0$. Consequently, we have

$$0 < \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|}. \tag{5.3}$$

Consider

$$v_k \quad = \quad -\alpha_k U_{k,1} \textstyle\sum_{k,1}^{-1} V_{k,1}^T h(x_k) \tag{5.4}$$

where

$$U_{k,1} \in \mathbb{R}^{n \times r_k}, \quad V_{k,1} \in \mathbb{R}^{m \times r_k}, \quad \text{and} \quad \textstyle\sum_{k,1} = diag(\sigma_{k,1}, \cdots, \sigma_{k,r_k}),$$

with r_k indicating the rank of $\nabla h(x_k)$. ¿From (2.8), i.e. $V_{k,2}^T h(x_k) = 0$, and equality (5.4), we obtain

$$\alpha_k h(x_k) + \nabla h(x_k)^T v_k = 0. \tag{5.5.}$$

In the case $p \geq 2$, the definition of α_k in STEP 4 and equality (5.4) imply that $\|v_k\|_2 \leq \delta_k$, which implies, since $\| \ \|_p \leq \| \ \|_2$, that

$$\|v_k\|_p \leq \delta_k. \tag{5.6a.}$$

Now we consider the case $p = 1$. We know that $\| \quad \|_1 \leq \sqrt{n} \, \| \quad \|_2$. We also have $\sqrt{n} \, \| \quad \|_\infty \geq \| \quad \|_2$, which according to (2.3), implies that $\nu_\infty = 1/\sqrt{n}$. Therefore, we obtain from the definition of α_k and the definition of ν_q in (2.3) that

$$\|v_k\|_1 \leq \delta_k. \tag{5.6b}$$

¿From (5.5), (5.6a), and (5.6b), we obtain that v_k is a feasible point for the subproblem (RTRQP) for $p = 1$ or any extended real $p \geq 2$. 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

$$0 < \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|} \leq \frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \; \sqrt{n} \tag{5.7}$$

$$\left(\|\nabla f(x_k)\|_p + \frac{1}{2} \|B_k\|_p \delta_k \right) + \eta_k.$$

Since ∇f is continuous, the sequences $\{x_k\}$, $\{\delta_k\}$, $\{B_k\}$ are bounded, and $0 \leq \eta_k \leq \beta$ for all k , we obtain from (5.7) that

$$0 < \frac{\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k}{\alpha_k \|h(x_k)\|} \leq \frac{\sqrt{n} \, \|v_k\|_2}{\alpha_k \|h(x_k)\|} M_1 + \beta_0, \tag{5.8}$$

for some positive constant M_1. ¿From (5.4), we obtain

$$\frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} = \frac{\|U_{k,1} \sum_{k,1}^{-1} V_{k,1}^T h(x_k)\|_2}{\|h(x_k)\|},$$

or, because $\|U_{k,1}\|_2 \leq 1$,

$$\frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \leq \frac{\|\sum_{k,1}^{-1} V_{k,1}^T h(x_k)\|_2}{\|h(x_k)\|}. \tag{5.9}$$

On the other hand, there exists a positive constant ν_{min} such that

$$\| \quad \| \geq \|\nu_{min} \quad \|_2. \tag{5.10}$$

Therefore, from Lemma 5.1, (5.9), (5.10), and $\|V_{k,1}\|_2 \leq 1$, we obtain

$$\frac{\|v_k\|_2}{\alpha_k \|h(x_k)\|} \leq \frac{1}{\omega \; \nu_{min}}, \tag{5.11}$$

which, together with (5.8), implies that

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

for some constant M. Consequently the sequence $\{\bar{\mu}_k\}$ is bounded.□

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 $\{\mu_k, k \geq k_{max}\}$ is not decreasing; let us show that it is bounded. Assume that there exists a subsequence $\{\mu_k, k \in N\}$ such that

$$\lim_{k \in N \to +\infty} \mu_k = +\infty.$$

Denote by M the upper bound of $\{\bar{\mu_k}\}$. Let $k_1 \geq k_{max}$ be the smallest integer in N such that $\mu_{k_1} \geq M + \rho$. Since $\{\mu_k, k \geq k_{max}\}$ 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 $k \geq k_1$. This contradicts the divergence hypothesis of the subsequence $\{\mu_k, k \in N\}$. Consequently $\{\mu_k\}$ is bounded. Now, for $k \geq k_{max}$, μ_k is either equal to μ_{k-1} or it is increased by at least ρ. Therefore, in the entire process of updating the penalty parameter, the increase can happen at most $p+1$ times, where p is the smallest integer greater or equal to $(M - \mu_{k_{max}})/\rho$. Consequently there exists k^* such that $\mu_k = \mu_{k^*}$ for all $k \geq k^*$.□

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

LEMMA 5.3. *Let $\{(x_k, B_k, \delta_k, \beta_k)\}$ 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 with respect to $(x_k, B_k, \delta_k, \beta_k)$. Then we have*

$$\lim_{k \to +\infty} \frac{\|h(x_k)\|}{\|s_k\|_p} = 0, \tag{5.12a}$$

$$h(x_*) = 0, \tag{5.12b}$$

and

$$\liminf_{k \to +\infty} \nabla f(x_k)^T d_k \geq 0, \tag{5.13}$$

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

Proof. 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

$$f(x_k + s_k) - f(x_k) + \mu_k \Big[\|h(x_k + s_k)\| - \|h(x_k)\|\Big] > \frac{c_1}{2} s_k^T B_k s_k + \tag{5.14}$$

$$c_1 \Big[\nabla f(x_k)^T s_k + \mu_k \big[\|h(x_k) + \nabla h(x_k)^T s_k\| - \|h(x_k)\|\big]\Big].$$

On the other hand, the exists $\xi_k \in (x_k, x_k + s_k)$ such that

$$\begin{aligned} f(x_k + s_k) - f(x_k) &= \nabla f(\xi_k)^T s_k, \\ &= \nabla f(x_k)^T s_k + [\nabla f(\xi_k) - \nabla f(x_k)]^T s_k, \\ &= \nabla f(x_k)^T s_k + o_{m+1}(\|s_k\|_p), \end{aligned}$$

which implies that

$$\text{(5.15)} \qquad \frac{f(x_k+s_k)-f(x_k)}{\|s_k\|_p} = \nabla f(x_k)^T d_k + \frac{o_{m+1}(\|s_k\|_p)}{\|s_k\|_p}.$$

Similarly we have for $j = 1, \cdots, m$

$$h_j(x_k+s_k) = h_j(x_k) + \nabla h_j(x_k)^T s_k + o_j(\|s_k\|_p),$$

and therefore

$$\text{(5.16)} \qquad \frac{\|h(x_k+s_k)\| - \|h(x_k)\|}{\|s_k\|_p} \le \frac{\|h(x_k)+\nabla h(x_k)^T s_k\| - \|h(x_k)\|}{\|s_k\|_p} + \|\frac{o(\|s_k\|_p)}{\|s_k\|_p}\|.$$

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

$$\text{(5.17)} \qquad \lim_{k\to+\infty} \frac{o(\|s_k\|)}{\|s_k\|_p} = 0 \quad \text{and} \quad \lim_{k\to+\infty} \frac{o_{m+1}(\|s_k\|)}{\|s_k\|_p} = 0.$$

Because $\{\mu_k\}$ and $\{B_k\}$ are bounded, $0 < 1 - c_1$, and $\{s_k\}$ converges to zero, we obtain from (5.14), (5.15), (5.16) and (5.17)

$$\text{(5.18a)} \qquad \nabla f(x_k)^T d_k + \mu_k \frac{\|h(x_k)+\nabla h(x_k)^T s_k\| - \|h(x_k\|}{\|s_k\|_p} > \frac{o(\|s_k\|)}{\|s_k\|_p}$$

where

$$\text{(5.18b)} \qquad \lim_{k\to+\infty} \frac{o(\|s_k\|)}{\|s_k\|_p} = 0.$$

Using $\alpha_k h(x_k) + \nabla h(x_k)^T s_k = 0$, we can rewrite (5.18a) as

$$\text{(5.18c)} \qquad \nabla f(x_k)^T d_k - \mu_k \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} > \frac{o(\|s_k\|_p)}{\|s_k\|_p},$$

or

$$\text{(5.18d)} \qquad \nabla f(x_k)^T d_k > \mu_k \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} + \frac{o(\|s_k\|_p)}{\|s_k\|_p},$$

which, together with (5.18b), implies (5.13), i.e.

$$\text{(5.19)} \qquad \liminf_{k\to+\infty} \nabla f(x_k)^T d_k \ge 0.$$

Adding inequalities (5.18c) and (5.18d), we obtain

$$\text{(5.20)} \qquad 2\nabla f(x_k)^T d_k - \mu_k \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} > \mu_k \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} + \frac{o(\|s_k\|_p)}{\|s_k\|_p},$$

which, together with (5.18b), implies that

$$\text{(5.21)} \qquad \liminf_{k\to+\infty} \Big[2\nabla f(x_k)^T d_k - \mu_k \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} \Big] \ge 0.$$

But we have

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

or equivalently

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

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

$$\liminf_{k\to+\infty}\Big(2\nabla f(x_k)^T d_k - \mu_k \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p}\Big) \leq \liminf_{k\to+\infty} -\alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p}. \tag{5.22}$$

¿From (5.21) and (5.22) we obtain

$$\liminf_{k\to+\infty} -\alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} \geq 0, \tag{5.23}$$

which implies that

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

and therefore

$$\lim_{k\to+\infty} \alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} = 0. \tag{5.24}$$

¿From the definition of the equality constraint relaxation parameter we obtain

$$\alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} = \min\Big(\frac{\|h(x_k)\|}{\|s_k\|_p}, \tau \frac{\delta_k \omega_k}{\|s_k\|_p} \frac{\|h(x_k)\|}{\|h(x_k)\|_2}\Big). \tag{5.25}$$

Suppose that there exists $N \subset \mathbb{N}$ such that

$$\tau \frac{\delta_k \omega_k}{\|s_k\|_p} \frac{\|h(x_k)\|}{\|h(x_k)\|_2} \leq \frac{\|h(x_k)\|}{\|s_k\|_p}$$

holds for all sufficiently large $k \in N$. This implies that for large $k \in N$ we have

$$\alpha_k \frac{\|h(x_k)\|}{\|s_k\|_p} = \tau \frac{\delta_k \omega_k}{\|s_k\|_p} \frac{\|h(x_k)\|}{\|h(x_k)\|_2}. \tag{5.26}$$

Therefore, we obtaib from (5.24), (5.25), (5.10) and Lemma 5.1, that

$$\lim_{k\in N\to+\infty} \frac{\delta_k}{\|s_k\|_p} = 0$$

which contradicts the inequality $\|s_k\|_p \leq \delta_k$ for all k. Consequently, for all large k we have

$$\frac{\|h(x_k)\|}{\|s_k\|_p} < \tau \frac{\delta_k \omega_k}{\|s_k\|_p} \frac{\|h(x_k)\|}{\|h(x_k)\|_2}. \tag{5.27}$$

Now from (5.24), (5.25), and (5.27), we obtain(5.12a), i.e.

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

which in turn implies, since $\{s_k\}$ converges to zero, that $h(x_*) = 0$, i.e. (5.12b). □

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 8 and STEP 4.*

Proof. Assume that the algorithm loops indefinitely between STEP 8 and STEP 4 without obtaining an acceptable step s_k. 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

$$\limsup_{j\to+\infty} \nabla f(x_k)^T d_j \geq 0, \tag{5.28}$$

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

$$\nabla h(x_k)^T s = 0. \tag{5.29}$$

Observe that s_j is an ϵ_j-solution (RTRQP) with $\epsilon_j = \eta_j\|s_j\|_p$ where $0 < \eta_j$ converges to zero. Let $t_j > 0$ be such that $\|t_j s\|_p = \|s_j\|_p$, i.e. $t_j s$ is feasible for the local model subproblem. 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 (5.28) and the convergence of $\{\eta_j, j \in \mathbb{N}\}$ and $\{s_j\}$ to zero, implies that

$$\nabla f(x_k)^T s \geq 0. \tag{5.30}$$

¿From the Farkas Lemma, (5.29), and (5.30) 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 ia 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 $\{x_k\}$ and hence, for convenience, we set $k_{max} = 0$, which implies that the sequence $\{\mu_k\}$ is not decreasing.

The way we update the trust-region radius follows from El Hallabi and Tapia (1993)[16]. It implies, since we assume that the iteration sequence is bounded, that the trust-region radii are uniformly bounded.

LEMMA 5.3 [El Hallabi and Tapia](1993)[16]. *Assume that the iteration sequence $\{x_k\}$ is bounded. Then the sequence $\{\Delta_k\}$ is bounded.*

The second important property of our trust-region framework is established in the following theorem in the form it will be used later. This property is equivalent to saying that if the iteration sequence $\{x_k\}$ converges to x_* and the sequence of trust-region radii that detemine acceptable steps converges to zero, then necessarily x_* is a Karush-Kuhn-Tucker point of (EQCP). (see El Hallabi and Tapia (1993)[16] or El Hallabi (1993)[17]).

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 with respect to $(x_k, B_k, \Delta_k, \beta_k)$, then there exists a positive scalar $\delta(x_*, B_*, \Delta_*)$ such that*

$$\delta_* \geq \delta(x_*, B_*, \Delta_*) \tag{5.31}$$

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

Proof. 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 (5.36) 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 with respect to $\{(x_k, B_k, \Delta_k, \beta_k)\}$. 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 5.2, we obtain

$$h(x_*) = 0, \tag{5.32}$$

$$\lim_{k \to +\infty} \frac{\|h(x_k)\|}{\|\bar{s_k}\|_p} = 0, \tag{5.33}$$

and

$$\limsup_{k \to +\infty} \nabla f(x_k)^T \frac{\bar{s_k}}{\|\bar{s_k}\|_p} \geq 0. \tag{5.34}$$

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

$$\nabla h(x_*)^T s = 0 \quad \text{and} \quad \|s\|_p = 1 \tag{5.35.}$$

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

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

Since $\bar{s_k}$ tends to be an exact minimizer of $(RTRQP)$, (5.34) suggests to us to construct a feasible point for $(RTRQP)$, say w_k, such that

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

Let us construct w_k. We have seen in (2.8) that $V_{k,2}^T h(x_k) = 0$. Now, we show that

$$V_{k,2}^T \nabla h(x_k)^T = 0. \tag{5.36}$$

We have

$$V_{k,2}^T \nabla h(x_k)^T = V_{k,2}^T V_k \textstyle\sum_k^T U_{x_k}^T,$$

or

$$V_{k,2}^T \nabla h(x_k)^T = \begin{pmatrix} 0 \\ I_{m-r_k} \end{pmatrix}^T \begin{pmatrix} \sum_{k,1} U_{k,1} \\ 0 \end{pmatrix},$$

therefore (5.36) holds. Let v_k be defined by

$$v_k = U_{k,1} \textstyle\sum_{k,1}^{-1} \left[-\bar{\alpha_k} V_{k,1}^T h(x_k) - \frac{\|\bar{s_k}\|_p}{2} V_{k,1}^T \nabla h(x_k)^T s\right] \tag{5.37),}$$

which implies that

$$\textstyle\sum_{k,1} U_{k,1}^T v_k = V_{k,1}^T \left[-\bar{\alpha_k} h(x_k) - \frac{\|\bar{s_k}\|_p}{2} \nabla h(x_k)^T s\right]. \tag{5.38}$$

On the other hand, from (5.36) and (2.8), we obtain

$$0 = V_{k,2}^T \left[-\bar{\alpha_k} h(x_k) - \frac{\|\bar{s_k}\|_p}{2} \nabla h(x_k)^T s\right]. \tag{5.39}$$

Consequently both equalities (5.38) and (5.39) can be gathered into

$$\textstyle\sum_k U_k^T v_k = V_k^T \left[-\bar{\alpha_k} h(x_k) - \frac{\|\bar{s_k}\|_p}{2} \nabla h(x_k)^T s\right],$$

or equivalently

$$V_k \textstyle\sum_k U_k^T v_k = \left[-\bar{\alpha_k} h(x_k) - \frac{\|\bar{s_k}\|_p}{2} \nabla h(x_k)^T s\right],$$

and therefore

$$\nabla h(x_k)^T v_k = -\bar{\alpha_k} h(x_k) - \frac{\|\bar{s_k}\|_p}{2} \nabla h(x_k)^T s. \tag{5.40}$$

Let us set

$$w_k = v_k + \frac{\|\bar{s_k}\|_p}{2} s. \tag{5.41}$$

From (5.40), we obtain

$$\bar{\alpha_k} h(x_k) + \nabla h(x_k)^T w_k = 0, \tag{5.42}$$

i.e. w_k satisfies the linear constraints of $(RTRQP)$. Let us establish that it also satisfies the trust-region constraint. ¿From (5.37) and Lemma 5.1, we obtain

$$\frac{\|v_k\|_2}{\|\bar{s_k}\|_p} \leq \frac{1}{\omega}\Big[\frac{\|h(x_k)\|_2}{\|\bar{s_k}\|_p} + \frac{1}{2}\|\nabla h(x_k)^T s\|_2\Big]. \tag{5.43}$$

Form (5.43), (5.33), (5.35), we obtain

$$\lim_{k\to+\infty} \frac{\|v_k\|_p}{\|\bar{s_k}\|_p} = 0. \tag{5.44}$$

Since $\|s\|_p = 1$, the definition of w_k in (5.41) implies that

$$\left|\frac{\|v_k\|_p}{\|\bar{s_k}\|_p} - \frac{1}{2}\right| \leq \frac{\|w_k\|_p}{\|\bar{s_k}\|_p} \leq \frac{\|v_k\|_p}{\|\bar{s_k}\|_p} + \frac{1}{2}.$$

¿From this double inequality and limit (5.44) 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

$$\|w_k\|_p \leq \|\bar{s_k}\|_p \leq \bar{\delta_k} \tag{5.45}$$

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

$$\frac{1}{3}\|\bar{s_k}\|_p \leq \|w_k\|_p. \tag{5.46}$$

¿From (5.42) and (5.45), we obtain that w_k is a feasible point for (RTRQP). Now let us obtain $lim_{k\to+\infty} w_k/\|w_k\|_p$. We have

$$\frac{w_k}{\|w_k\|_p} = \frac{w_k}{\|\bar{s_k}\|_p}\left(\frac{\|w_k\|_p}{\|\bar{s_k}\|_p}\right)^{-1}, \tag{5.47}$$

where

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

which, together with (5.44), implies that

$$\lim_{k\to+\infty} \frac{w_k}{\|\bar{s_k}\|_p} = \frac{1}{2}s. \tag{5.48}$$

Since $\|s\|_p = 1$, from (5.47) and (5.48) we conclude that

$$\lim_{k\to+\infty} \frac{w_k}{\|w_k\|_p} = s. \tag{5.49}$$

Finally, we are ready to establish that $\nabla f(x)^T s \geq 0$. 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

$$\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} \leq \nabla f(x_k)^T w_k + \frac{1}{2}w_k^T B_k w_k + \bar{\epsilon_k} \tag{5.50}$$

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

$$\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k \geq 0. \tag{5.51}$$

This implies that

$$\nabla f(x_k)^T \frac{w_k}{\|w_k\|_p} + \frac{1}{2} w_k^T B_k \frac{w_k}{\|w_k\|_p} \geq 0. \tag{5.52}$$

Since $\{w_k\}$ converges to zero, $\{x_k\}$ converge to x_*, and $\{B_k\}$ is bounded, we obtain from (5.49) and (5.52) that

$$\nabla f(x_*)^T s \geq 0. \tag{5.53}$$

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

$$\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k < 0. \tag{5.54}$$

¿From (5.50) and (5.54) we obtain

$$\frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} - \bar{\epsilon_k}}{\|w_k\|_p} \leq \frac{\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k}{\|w_k\|_p} < 0, \tag{5.55}$$

which, together with (5.45), implies

$$3\frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} - \bar{\epsilon_k}}{\|\bar{s_k}\|_p} \leq \frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k} - \bar{\epsilon_k}}{\|w_k\|_p}. \tag{5.54}$$

Inequalities (5.55) and (5.56) imply that

$$3\frac{\nabla f(x_k)^T \bar{s_k} + \frac{1}{2}\bar{s_k}^T B_k \bar{s_k}}{\|\bar{s_k}\|_p} \leq \frac{\nabla f(x_k)^T w_k + \frac{1}{2} w_k^T B_k w_k}{\|w_k\|_p} + 3\bar{\eta_k}. \tag{5.57}$$

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 (5.57), (5.49), and (5.34) that inequality (5.53) holds also in the case (5.54). Therefore in both cases (5.51) and (5.54), we obtain that

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

Now, since this inequality holds for any s such that (5.40) holds, i.e. $\nabla h(x_*)^T s = 0$, and because of (5.32), 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.□

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.* We emphasize that this property played a pivotal role in El Hallabi (1993)[16]. 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). *Consider $(x_*, B_*, \Delta_*, 0)$, where B_* is an arbitrary matrix, and $0 < \Delta_*$. If x_* is not a Karush-Kuhn-Tucker point of (EQCP), then there exists a neighborhood $N_* = N_*(x_*, B_*, \Delta_*, 0)$ and a positive scalar ρ_* such that for any $(x, B, \Delta, \beta) \in N_*$*

$$\Phi(x_+) < \Phi(x_*) - \rho_* \tag{5.58}$$

holds for any successor $(x_+, B_+, \Delta_+, \beta_+)$ of (x, B, δ, β).

Proof. We prove the contrapositive. Then there exists a sequence $\{(x_k, B_k, \Delta_k, \beta_k)\}$ converging to $(x_*, B_*, \Delta_*, 0)$, a sequence $\{\rho_k\}$ converging to zero, and a sequence $\{(x_{k+}, B_{k+}, \Delta_{k+}, \beta_{k+})\}$ where $(x_{k+}, B_{k+}, \Delta_{k+}, \beta_{k+})$ is a successor of $(x_k, B_k, \delta_k, \beta_k)$ (see Definition 4.2) such that

$$\Phi(x_{k+}) \geq \Phi(x_*) - \rho_k \tag{5.59}$$

holds fo all k. Therefore there exists an $\epsilon_k = \epsilon(x_k, \eta_k)$-solution 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

$$\Phi(x_k + s_k) \leq \Phi(x_k) + c_1 \big[\Psi(x_k + s_k) - \Psi(x_k)\big]. \tag{5.60}$$

Inequalities (5.60) and (5.59) imply that

$$\Phi(x_*) - \Phi(x_k) \leq c_1 \big[\Psi(x_k + s_k) - \Psi(x_k)\big] + \rho_k, \tag{5.61a}$$

or, by Proposition 4.1,

$$\Phi(x_*) - \Phi(x_k) \leq c_1 \big[-\rho \alpha_k \|h(x_k)\|\big] - c_1 |\nabla f(x_k)^T s_k + \frac{1}{2} s_k^T B_k s_k| + \rho_k. \tag{5.61b}$$

Since $\{x_k\}$ converges to x_*, and $\{\rho_k\}$ converges to zero, we obtain from (5.61b)

$$\lim_{k \to +\infty} \alpha_k \|h(x_k)\| = 0. \tag{5.62a}$$

But we have

$$\alpha_k \|h(x_k)\| = \min(\|h(x_k)\|, \tau \delta_k \omega_k \frac{\|h(x_k)\|}{\|h(x_k)\|_2}),$$

which, by Lemma 5.1, Theorem 5.1, and (5.10), implies that

$$\alpha_k \|h(x_k)\| \geq \min\left(\|h(x_k)\|, \ \tau \ \omega \ \nu_{min} \ \frac{\delta(x_*, B_*, \Delta_*)}{2} \right) \tag{5.62b}$$

holds for sufficiently large k. Therefore, from limit (5.62a) and (5.62b), we obtain

$$h(x_*) = 0, \tag{5.63}$$

which, together with Lemma 5.1, Theorem 5.1, and the definition of α_k, implies that, for all sufficiently large k, the constraint relaxation 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

$$(RTRQP) \equiv \begin{cases} \text{minimize} & \nabla f(x_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & h(x_k) + \nabla h(x_k)^T s = 0 \\ & \|s\|_p \leq \delta_k. \end{cases} \tag{5.64a}$$

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 (1975)[21], we have that s_* solves the subproblem

$$\begin{cases} \text{minimize} & \nabla f(x_*)^T s + \frac{1}{2} s^T B_* s \\ \text{subject to} & \nabla h(x_*)^T s = 0 \\ & \|s\|_q \leq \delta_*. \end{cases} \tag{5.64b}$$

(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 (5.61b) that

$$\nabla f(x_*)^T s_* + \frac{1}{2} s_*^T B_* s_* = 0. \tag{5.65}$$

Consequently $s = 0$ solves the subproblem (5.64b) 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. *Let $\{x_k\}$ be a sequence generated by the algorithm ANITRA of Section 3, and let $\{B_k\}$ be the sequence of matrices used by the algorithm. Assume that*

1) $\{x_k\}$ is bounded,
2) $\{B_k\}$ is bounded,
3) for all k, the linearized constraints are consistent,
4) the functions f and $h_i, i = 1 \cdots m$, are continuously differentiable, and
5) the sequence $\{\beta_k\}$ that is used to obtain ϵ_k-solutions of the local model subproblem converges to zero.

Then any accumulation point of $\{x_k\}$ is a Karush-Kuhn-Tucker point of (EQCP).

Proof. Let x_* be an arbitrary accumulation point x_* of $\{x_k\}$. Consider the sequence $\{x_k, k \geq k^*\}$ 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 $\Phi(x)$. Since, for all $k \geq k^*$, x_k is not a Karush-Kuhn-Tucker point of (EQCP), we have

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

Let $\{x_j, j \geq k^*\}$ be a subsequence that converges to x_*. Consider $k \geq k^*$. There exists $j(k) > k$ such that

$$\Phi(x_{j(k)}) < \Phi(x_k),$$

and consequently

$$\Phi(x_j) < \Phi(x_k),$$

holds for all $j \geq j(k)$. Therefore, we obtain

$$\Phi(x_*) \leq \Phi(x_k) \quad \forall k \geq k^*. \tag{5.66}$$

Assume that x_* is not a Karush-Kuhn-Tucker point of (EQCP). Since $\{x_j, j \geq k^*\}$ converges to x_*, there exists $j(x_*) \geq k^*$ such $x_j \in N_*$ for all $j \geq j(x_*)$, where N_* is defined in Theorem 5.2, which implies that

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

This contradicts (5.66). 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 (1979)[22] or Polak (1970)[26] 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. 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 has the form

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

where the fixed $\| \, \|_p$ can be a polyhedral norm or an arbitrary ℓ_p norm with $p \geq 2$.

When $\| \, \|_p$ is a polyhedral norm, i.e. ℓ_1 or ℓ_∞ norm, or a convex combination of polyhedral norms, the subproblem can be formulated as an SLP or SQP depending on whether or not we use second order information matrices B_k. Because we only assume first order differentiability of the functions f and $h_i, i = 1 \cdots m$, our theory fits well both SLP and SQP formulations. Moreover, since at each iteration, only an approximate solution is required, we believe that our theory would apply to the case where the first derivatives are approximated by some finite difference scheme.

We established, under rather weak assumptions, that any accumulation point of the iteration sequence is a Karush-Kuhn-Tucker point of (EQCP). 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).

Also, we only assume that f and $h_i, i = 1 \cdots m$, are continuously differentiable, and that the system of linearized constraints is consistent, whereas generally stronger assumptions such as uniform linear independence of the gradients $\nabla h_i(x), i = 1 \cdots m$, and continuity of the second derivatives of f and $h_i, i = 1 \cdots m$.

Also observe that to obtain a trial step s_k, one subproblem per trust-region radius needs to be solved within some tolerance, while in the two level algorithm in Byrd, Omojokun, Byrd, and Shultz (1987)[3], two subproblems per trust-region radius need to be solved approximately.

An equivalent subproblem to RTRQP. Before ending this section we present an equivalent local model subproblem to (RTRQP) that, we believe, deserves further consideration. This equivalent subproblem is

$$(RTRQP') \quad \equiv \quad \begin{cases} \text{minimize} & \nabla f(x_k)^T u + \frac{1}{2} u^T (\alpha_k B_k) u \\ \text{subject to} & h(x_k) + \nabla h(x_k)^T u = 0 \\ & \|u\|_p \le \delta_k' \end{cases},$$

where

$$\delta_k' = \frac{\delta_k}{\alpha_k} = \max(\delta_k, \frac{\|h(x_k)\|_2}{\omega_k}).$$

Observe that (RTRQP) and (RTRQP') are equal when $h(x_k) = 0$. Also, when $h(x_k) \neq 0$, it can be shown that s_k is an ϵ_k-solution of subproblem (RTRQP) with $\epsilon_k = \eta_k \min(\|s_k\|_p, \alpha_k \|h(x_k)\|)$ if and only if u_k is an ϵ_k'-solution of subproblem $(RTRQP')$ with $\epsilon_k' = \eta_k \min(\|u_k\|_p, \|h(x_k)\|)$ where $s_k = \alpha_k u_k$.

In this equivalent subproblem formulation, the equality constraints relaxation parameter α_k also seems to behave like a scale for the second order approximation B_k. In STEP 5, we could scale before updating the matrix B_k. In our theory, we ask for α_k to satisfy

$$0 < \alpha_k = \min(1, \alpha_k^*) \quad \text{where} \quad \alpha_k^* = \tau \nu_q \frac{\delta_k \omega_k}{\|h(x_k)\|_2}, \tag{6.1}$$

and we expect to have $\alpha_k = 1$ for sufficiently large k. On the other hand, in quasi-Newton's methods, it is common to scale the matrices B_k (see Contreras and Tapia (1993)[8] or Dennis and Wolkowicz (1990)[10]). When these matrices are symmetric positive definite, a scaling factor is

$$\theta_k = \frac{y_k^T s_k}{s_k^T B_k s_k} \tag{6.2}$$

which is known to converge to one, under suitable assumptions (see Yabé, Martinez, and Tapia (1993)[37]). Therefore, our theory applies for any relaxation parameter $\widehat{\alpha_k}$ satisfying

$$0 \le \widehat{\alpha_k} = \min(1, \alpha_k^*, \theta_k), \tag{6.3}$$

where α_k^* is defined in (6.1) and θ_k is given by (6.2) or could be any other scaling parameter for B_k's that is bounded away from zero at non stationary points of (EQCP). This would shift the linearized constraints and scale the second order approximation matrices in the same time away from a solution of (EQCP) and become inactive in some neighborhood of a solution of (EQCP).

7. Appendix A. In this appendix, we show that the use the QR-decomposition with column pivoting algorithm provides an ω_x that may be taken as the lower bound sought for in Proposition 2.1. Also, since it is a by-product of the determination of α_k, we use the QR-decomposition to obtain the least square estimate of the vector of Lagrange multipliers. Finally, we give the proof of Lemma 5.1, showing that ω_k can be obtained such that it is bounded away from zero. When we apply the QR-decomposition to $\nabla h(x)$, we stop at the iteration number r_x, where $r_x \leq m$ is the smallest integer such that

$$\|{}^{(r_x)}R_{22}\|_F \leq \epsilon_1 \max(\epsilon_2, \|{}^{(r_x)}R\|_F),$$

or by orthogonality

$$\|{}^{(r_x)}R_{22}\|_F \leq \epsilon_1 \max(\epsilon_2, \|\nabla h(x)\|_F). \tag{7.1}$$

where $\|\ \|_F$ denotes the Frobenius norm, $0 < \epsilon_1 \ll 1$, $0 < \epsilon_2 \ll 1$ and

$${}^{(r_x)}R = \begin{pmatrix} {}^{(r_x)}R_{11} & {}^{(r_x)}R_{12} \\ 0 & {}^{(r_x)}R_{22} \end{pmatrix},$$

with the upper-left index (r_x) indicating the QR-decomposition iteration number. We refer to (r_x) for which (7.1) holds as the *estimated rank* of $\nabla h(x)$. Observe that this definition depends on the method used to obtain (r_x) (in our case QR-decomposition with column pivoting).

The following proposition and its corollary give a positive lower bound ω_x.

PROPOSITION 7.1. *Let r_x be the estimated rank of $\nabla h(x)$ where x is a given point in $\mathbb{R}^n$. Consider the partial QR-decomposition*

$$\nabla h(x)\Pi_1 \cdots \Pi_j = Q_j \cdots Q_1 \quad {}^{(j)}R \tag{7.2}$$

where Π_i and $Q_i, i = 1, \cdots j \leq r_x$, are permutation matrices, and

$${}^{(j)}R = \begin{pmatrix} {}^{(j)}R_{11} & {}^{(j)}R_{12} \\ 0 & {}^{(j)}R_{22} \end{pmatrix},$$

where ${}^{(j)}R_{11}$ is a nonsingular upper triangular matrix of rank j. Then

$$\left|\left({}^{(j)}R_{11}\right)_{j,j}\right| \leq \sigma_j(\nabla h(x)). \tag{7.3}$$

Proof. Let ${}^{(j)}Q = Q_j \cdots Q_1$ and ${}^{(j)}\Pi = \Pi_1 \cdots \Pi_j$. Then equality (7.2) can be written as

$$\nabla h(x) \quad {}^{(j)}\Pi = ({}^{(j)}A_1 \quad , \quad {}^{(j)}A_2) \tag{7.4}$$

where

$${}^{(j)}A_1 = \quad {}^{(j)}Q \begin{pmatrix} {}^{(j)}R_{11} \\ 0 \end{pmatrix} \quad \text{and} \quad {}^{(j)}A_2 = \quad {}^{(j)}Q \begin{pmatrix} {}^{(j)}R_{12} \\ {}^{(j)}R_{22} \end{pmatrix}. \tag{7.5}$$

By Theorem 12.12-1 of Golub and Van Loan (1983)[19], (7.4) implies that

$$\sigma_j({}^{(j)}A_1) \leq \sigma_j(\nabla h(x)). \tag{7.6}$$

On the other hand, the matrix ${}^{(j)}A_1$ is full rank (i.e. j) and its QR-decomposition is given in (7.5). Therefore

$$\left|\left({}^{(j)}R_{11}\right)_{j,j}\right| \le \sigma_j\left({}^{(j)}A_1\right)$$

which, together with (7.6), implies (7.3). □

COROLLARY 7.1. *Let r_x be the estimated rank of $\nabla h(x)$ where x is a given point in $\mathbb{R}^n$. Then*

$$\omega_x = \left|\left({}^{(r_x)}R_{11}\right)_{r_x,r_x}\right| \le \sigma_{r_x}(\nabla h(x))$$

where ${}^{(r_x)}R_{11}$ is the nonsingular upper triangular component of

$${}^{(r_x)}R = \begin{pmatrix} {}^{(r_x)}R_{11} & {}^{(r_x)}R_{12} \\ 0 & {}^{(r_x)}R_{22} \end{pmatrix}$$

given by the QR-decomposition with column pivoting satisfying

$$\left\|{}^{(r_x)}R_{22}\right\|_F \le \epsilon_1 \max(\epsilon_2, \|\nabla h(x)\|_F).$$

Usually, the matrices B_k are approximation to the Hessian of the Lagrangian. So to perform STEP 5 of the ANITRA Algorithm, we will need the vector of Lagrange multipliers λ_k, which may be determined as follows. Consider the QR-decomposition with column pivoting $\nabla h(x_k)\ \Pi = \ Q\ R^{(k)}$, where the subscript on Π and Q are omited. Let A_k be the matrix of the first r_k columns of the matrix $\nabla h(x_k)\Pi$, and b_k the vector of the first r_k components of $\Pi^T h(x_k)$. Therefore A_k is a full rank $n \times r_k$ matrix whose QR-decomposition is given by

$$A_k = Q \begin{pmatrix} R_{11}^{(k)} \\ \\ 0 \end{pmatrix}. \tag{7.7}$$

We define λ_k as

$$\lambda_k = \Pi^T y_k \tag{7.8a}$$

where the first r_k components of y_k satisfy

$$(y_{k,1}, \cdots, y_{k,r_k})^T = -\left[A_k^T A_k\right]^{-1} A_k^T\left(\nabla f(x_k) + B_k s_k\right), \tag{7.8b}$$

and the last $m - r_k$ components satisfy

$$y_{k,j} = 0 \qquad j = r_k + 1, \cdots, m. \tag{7.8c}$$

To end this appendix, we give the proof of Lemma 5.1.

Proof of Lemma 5.1. Let r_k be the estimated rank of $\nabla h(x_k)$ by performing the QR-decomposition with column pivoting

$$\nabla h(x_k)\Pi_1 \cdots \Pi_{r_k} = Q_{r_k} \cdots Q_1 \ \ {}^{(r_k)}R^{(k)}, \tag{7.9a}$$

where

$$
{}^{(r_k)}R^{(k)} = \begin{pmatrix} {}^{(r_k)}R_{11}^{(k)} & {}^{(r_k)}R_{12}^{(k)} \\ 0 & {}^{(r_k)}R_{22}^{(k)} \end{pmatrix}, \tag{7.9b}
$$

along with the stopping criterion

$$
\|{}^{(r_k)}R_{22}^{(k)}\|_F \le \epsilon_1 \max(\epsilon_2, \|\nabla h(x_k)\|_F), \tag{7.10}
$$

where the upper-left index (r_k) indicates the QR-decomposition iteration number, and the upper-right (k) the iteration number of the algorithm ANITRA of section 3. Let $\epsilon = \epsilon_1 \max(\epsilon_2, \|\nabla h(x_k)\|_F)$. Inequality (7.10) implies that

$$
\left\| \left({}^{(r_k)}R_{22}^{(k)} \right)_{.j} \right\|_2 \le \epsilon \tag{7.11}
$$

for any column j of the matrix $({}^{(r_k)}R_{22}^{(k)})$, $j = r_k + 1, \cdots, m$. We also have

$$
\| \; {}^{(r_k-1)}R_{22}^{(k)} \; \|_2 > \epsilon. \tag{7.12}
$$

But

$$
\max_{r_k-1\le j\le m} \left\| \left({}^{(r_k-1)}R_{22}^{(k)} \right)_{.j} \right\|_2 = \left| \left({}^{(r_k)}R_{11}^{(k)} \right)_{r_k,r_k} \right|. \tag{7.13}
$$

On the other hand

$$
\left\| \; {}^{(r_k-1)}R_{22}^{(k)} \; \right\|_F \le (m - r_k + 1) \max_{r_k-1\le j\le m} \left\| \left({}^{(r_k-1)}R_{22}^{(k)} \right)_{.j} \right\|_2.
$$

Therefore, from (7.11), (7.12), and (7.13), we obtain

$$
\frac{\epsilon}{\sqrt{m - r_k + 1}} < \left| \left({}^{(r_k)}R_{11}^{(k)} \right)_{r_k,r_k} \right|,
$$

which implies that

$$
\frac{\epsilon}{\sqrt{m}} < \left| \left({}^{(r_k)}R_{11}^{(k)} \right)_{r_k,r_k} \right|.
$$

Consequently, we obtain

$$
\frac{\epsilon_1\epsilon_2}{\sqrt{m}} < \left| \left({}^{(r_k)}R_{11}^{(k)} \right)_{r_k,r_k} \right|. \tag{7.14}
$$

Let us set

$$
\omega = \frac{\epsilon_1\epsilon_2}{\sqrt{m}}. \tag{7.15}
$$

Therefore, from Corollary 7.1, (7.14), and (7.15), we obtain (5.1), i.e $\omega \le \omega_k$ for all k.□

8. Appendix B. We give in this appendix the Theorem of the alternative of Dax (1990)[9] used in the proof of Lemma 2.2. According to the dimensions in (EQCP), we cite the original Dax's Theorem with inverted dimension n and m, since there are no conditions on these dimensions. Also, in the present work, the extended reals p and q are in reverse order with respect to [9]. So, we invert the roles of p and q in the original version of Dax's Theorem.

THEOREM OF THE ALTERNATIVE [Dax] (1990) [9]. *Let p and q be extended reals satisfying $q > 1$ and $p = q/(q-1)$. Let A be an $n \times m$ matrix, let g be an non-zero m-vector, and let W be an $n \times n$ matrix with positive diagonal elements ω_i, $i = 1, \cdots, n$. Then either the inequality*

$$g^T y + \|WAy\|_q < 0 \tag{8.1}$$

has a solution $y \in \mathbb{R}^m$, or the system

$$A^T s = g \quad \text{and} \quad \|W^{-1}s\|_p \tag{8.2}$$

has a solution $s \in \mathbb{R}^m$, but never both.

In our application, for a given $x \in \mathbb{R}^n$ satisfying $h(x) \neq 0$, we set:

1. $g = -\alpha h(x) \in \mathbb{R}^m$

2. $A = \nabla h(x) \in \mathbb{R}^{n \times m}$ and

3. $W = \delta I_n$, where I_n denotes the identity matrix in $\mathbb{R}^{n \times n}$.

Therefore, the system (8.2), which becomes

$$\nabla h(x)^T s + \alpha h(x) = 0 \quad \text{and} \quad \|s\|_p \leq \delta, \tag{8.3}$$

has a solution $s \in \mathbb{R}^n$, if and only if the inequality (8.1), which becomes

$$\delta \|\nabla h(x) y\|_q \geq \alpha h(x)^T y,$$

for all $y \in \mathbb{R}^m$.

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

[2] R. BENVENISTE, *A quadratic programming algorithm using conjugate search directions*, Math. Programming, 16 (1979), pp. 63-80.

[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] M. CONTRERAS, and R. A. TAPIA, *Sizing the BFGS and the DFP updates: A numerical study*, J. Optim. Theory Appl., 78 (1993), pp. 93-108.

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

[10] 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.

[11] J. E. DENNIS, H. WOLKOWICZ, *Sizing and least change secant methods*, Tech. Rep. TR90-5 (1990), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892.

[12] 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.

[13] 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.

[14] 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.

[15] R. A. TAPIA, M. EL-Alem, *Numerical experience with a polyhedral norm CDT trust-region algorithm*, Tech. Rep. TR93-26 (1993), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892

[16] 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

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

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

[19] R. FLETCHER, *Practical Optimization*, John Wiley and Sons, NewYork, 1987.

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

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

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

[23] 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.

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

[25] 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.

[26] 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.

[27] 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

[28] 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.

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

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

[31] D. F. SHANNO, and R. E. MARSTEN, *Conjugate gradient for linearly constrained nonlinear programming*, in A.G. Buckley and J.-L. Goffin, Algorithms for constrained minimization of smooth nonlinear functions, Mathematical Programming Study 16, North Holland, Amsterdam, 1982.

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

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

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

[35] 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.

[36] R. A. TAPIA, M. EL-Alem, *Numerical experience with a polyhedral norm CDT trust-region algorithm*, Tech. Rep. TR93-26 (1993), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892

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

[38] H. YABÉ, H.J. MARTINEZ, R.A. TAPIA, *On sizing and shifting the BFGS update within the sized Broyden family of secant update*, Tech. Rep. TR93-58 (1993), Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 77251-1892