AN EXACT PENALTY ALGORITHM FOR NONLINEAR EQUALITY CONSTRAINED OPTIMIZATION PROBLEMS

M. EL HALLABI*

Abstract. In this paper we define a trust-region globalization strategies to solve a continuously differentiable nonlinear equality constrained minimization problem. The trust-region approach uses a penalty parameter that is proven to be uniformly bounded.

Under rather weak hypotheses and without the usual regularity assumption that the linearized constraints gradients are linearly independent, we prove that the hybrid algorithm is globally convergent

Moreover, under the standard hypotheses of the SQP method, we prove that the rate of convergence is q-quadratic.

Key Words: Cauchy Decrease, Constrained Optimization, Equality Constrained, Global Convergence, Non Regularity, Quadratic Convergence, Trust-Region.

AMS subject classification: 65K05, 49D37.

1. Introduction

For approximating a solution of a differentiable unconstrained minimization problem, the use of linesearch approaches in globalization strategies of quasi-Newton or Newton's methods stems from the fact that, given a descent direction of the objective function at a current iterate, obtaining an acceptable steplength, and hence a next iterate, costs only few function evaluations. Unfortunately, global convergence properties can only be obtained under quite strong hypotheses, especially on the second order informations. Usually, these hypotheses are that the second order information matrices, say B_k, are uniformly positive definite and uniformly bounded. Otherwise, to obtain a global convergence result, strong hypotheses are made on the objective. For example, a global convergence result for the BFGS variable metric method using the inexact linesearch of Wolfe [52] is given in Powell [42] for a convex objective function. Under the same hypothesis, this results is extended to the entire Broyden class in Ritter [46].

On the other hand, the trust-region approach, derived by Levenberg [32] and Marquardt [34] for least-squares problems and extended to the general unconstrained minimization problem by Powell [40], has proved to be a quite powerful tool for designing globally convergent algorithm (e.g. Powell [40], [41] and Schultz, Schnabel and Byrd [48]). We emphasize that, in the trust-region approach, the second order information matrices B_k do not need to be positive definite. Indeed, a global convergence result is obtained by Powell [43] for a variable metric algorithm where $\|B_k\|_2 \leq \alpha_1 + \alpha_2 k$ where α_1 and α_2 are positive constants.

Also, the trust-region approach has proved to be a very effective tool for designing globalization of the Newton methods for nonlinear equations (see Duff, Nocédal and Reid [15], Eisenstat and Walker [26], El Hallabi and Tapia [25] and Powell [44]).

We believe that the global convergence behavior of the trust-region approach is relevant to the strategy of reducing the trust-region radius δ_k until an acceptable step is obtained. This stands on the property that the trial step points to the steepest descent as δ_k converges to zero (see El Hallabi and Tapia [25, Theorem 4.1]). In other words, the trust-region strategy is reliant on the possibility of going, in the worst case and as far as global convergence is concerned, into the steepest descent direction. In

* Currently at the Institut National des Postes et Télécommunications, B.P. 6201, Charia Allal El Fassi, Rabat- Instituts, Rabat, Morocco.

fact, in the unconstrained optimization case, Powell [41] proved that if the generated iteration sequence satisfies the fraction of Cauchy decrease condition then

$$\liminf_{k\to\infty} \nabla f(x_k) = 0 \tag{1.1}$$

where f is the objective function, i.e. the iteration sequence has an accumulation point that is stationary.

Because of the strong global convergence properties of the trust-region approach for unconstrained optimization and nonlinear equations, considerable attention has been given to extend such a strategy to the equality constrained differentiable optimization problem

$$(EQCP) \equiv \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0 \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m, m \leq n$. Since the quasi-Newton (or Newton's) method to solve the problem $(EQCP)$ is the (local) SQP (sequential quadratic programming) method, the quasi-Newton step being the solution of the quadratic optimization problem

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

where B_k is an approximation of the Hessian of the Lagrangian $\nabla_x l(x_k, \lambda_k)$ at (x_k, λ_k), the master idea in any extension of trust-region approach to the problem $(EQCP)$ was that the resulting algorithm should be a globalization strategy of the SQP method.

The obvious extension was to add a trust-region constraint to (QP) leading to the subproblem

$$(TRQP) \equiv \begin{cases} \text{minimize} & \nabla_x l(x_k, \lambda_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & h(x_k) + \nabla h(x_k)^T s = 0 \\ & \|s\|_2 \leq \delta_k \end{cases}$$

where $0 < \delta_k$ is the trust-region radius. But, unless $h(x_k) = 0$, it is obvious that the feasible region of such a problem can be empty.

The first idea to overcome this difficulty was to translate the hyperplane of the linearized constraints so that it intersects the trust-region ball, leading to the relaxed trust-region subproblem

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

where $\eta_k \in (0, 1]$ is determined so that the feasible region of $(RTRQP)$ is not empty (and not reduced to a singleton). This approach was first suggested by Mièle, Huang and Heideman [35] to obtain a linesearch globalization method to solve $(EQCP)$. Then, it was used in the trust-region globalization framework by Vardi [51], Byrd, Schnabel, and Schultz [6] and El Hallabi [21], [22] and [23]. In El hallabi [21], [22] and [23] the proposed algorithm is shown to be globally convergent in the sense that any accumulation point of the iteration sequence is a Karush-Kuhn-Tucker point of $(EQCP)$. Moreover, this result is obtained without assuming the regularity assumption that the gradients of linearized constraints are linearly independent; it is obtained under the assumption that a positive lower bound of the positive singular values of $\nabla h(x_k)^T$ is known.

It is commonly admitted that the main problem with this approach lies in the difficulty of obtaining a translating parameter η_k. In El Hallabi [21], [22] and [23], a way of obtaining such a parameter is given provided that a positive lower bound of the smallest positive singular value of $\nabla h(x_k)$ (or, in a similar formulation, the smallest positive diagonal entry of the R factor of the QR-decomposition with column pivoting of $\nabla h(x_k)^T$) is known. For the QR-decomposition and the singular value decomposition, we refer to Golub and Van Loan [28]. Because of this, we believe that the major problem with this globalization strategy lies rather in the inconsistency of the linearized constraints that may occur far from a solution.

The second approach to resolve the infeasibility problem in $(RTRQP)$ is due to Célis, Dennis, and Tapia [9]. Instead of the $(TRSQP)$, they use the subproblem

$$(CDT) \equiv \begin{cases} \text{minimize} & \nabla_x l(x_k, \lambda_k)^T s + \frac{1}{2} s^T B_k s \\ \text{subject to} & \left\| h(x_k) + \nabla h(x_k)^T s \right\|_2^2 \leq \theta_k \\ & \|s\|_2 \leq \delta_k \end{cases}$$

where the parameter θ_k is chosen to define a nonempty feasible region and to realize some predicted decrease in the ℓ_2-norm of the linearized constraint at x_k inside the ball of radius δ_k. Different ways of defining such a parameter are given in Célis, Dennis, and Tapia [9], El-Alem [17], and Powell and Yuan [45]. In practice it is hard to solve the (CDT) subproblem (e.g. Zhang [54]). El-Alem [17] and Powell and Yuan [45] proved global convergence for this approach in the sense that

$$\liminf_{k\to\infty} \left\| Z_k^T \nabla f(x_k) \right\|_2 + \|h(x_k)\|_2 = 0 \tag{1.2}$$

where Z_k is a uniformly bounded full rank matrix (e.g. an orthogonal matrix) satisfying $\nabla h(x_k)^T Z_k = 0$, i.e. the iteration sequence has an accumulation point that is a Karush-Kuhn-Tucker point of $(EQCP)$.

The third approach to remedy the infeasibility problem in $(RTRQP)$ was suggested by Byrd [5] and Omojokun [38]. In this approach, the trial step s_k is of the form $u_k + y_k$ where u_k and y_k, referred to as the normal and tangential components, are respectively approximate solutions of the subproblems

$$(N-STEP) \equiv \begin{cases} \text{minimize} & \left\| h(x_k) + \nabla h(x_k)^T u \right\|_2^2 \\ \text{subject to} & \|u\|_2 \leq \tau\delta_k \end{cases}$$

where $\tau \in (0,1)$, and

$$(T-STEP) \equiv \begin{cases} \text{minimize} & \left(\nabla_x l(x_k, \lambda_k) + B_k u_k\right)^T v + \frac{1}{2} v^T B_k v \\ \text{subject to} & \nabla h(x_k)^T v = 0 \\ & \|v\|_2^2 \leq \delta_k^2 - \|u_k\|_2^2 . \end{cases}$$

Usually, the tangential component y_k is obtained as $Z_k v_k$ where v_k is an approximate solution of the reduced subproblem

$$(RT{-}STEP) \equiv \begin{cases} \text{minimize} & \left(Z_k(\nabla_x l(x_k, \lambda_k) + B_k u_k)\right)^T v + \frac{1}{2} v^T Z_k^T B_k Z_k v \\ \text{subject to} & \|Z_k v\|_2 \leq \sqrt{\delta_k^2 - \|u_k\|_2^2}. \end{cases}$$

Both subproblems $(N-STEP)$ and $(RT-STEP)$ can be solved by very effective methods such as those in Moré and Sorensen [36], Sorensen [49] or Steihaug [50].

This last approach is now standard and has been used extensively. We refer to Alexandrov [1], Alexandrov and Dennis [2], Biegler, Nocédal and Schmid [3], Dennis, El-Alem and Maciél [11], Dennis and Vicente [14], El-Alem [18] and [19], Lalee, Nocédal and Plantega [31], and Maciél [33]. Many authors have established global convergence results for the algorithms they proposed to solve the problem $(EQCP)$. Similarly to Powell [41] for the unconstrained minimization, Dennis, El-Alem, and Maciél [11] proved that the limit (1.2) holds. This result was obtained under the regularity assumption that the gradients of linearized constraints are linearly independent. In El-Alem [19], the author reestablished the results in Dennis, El-Alem, and Maciél

[11] without the regularity assumption.

In this research, we aim to combine both trust-region and linesearch globalization strategies in a globally convergent hybrid algorithm to solve the problem $(EQCP)$. This combination has been done in the unconstrained case by Nocédal and Yuan [37].

We will combine any inexact linesearch technique with the trust-region approach of Byrd [5] and Omojokun [38], but we will allow both the ℓ_2 or the ℓ_∞ norms in the trust-region constraint of the subproblems $(N-STEP)$ and $(T-STEP)$.

First, given a current estimation of a solution x_k, an associated Lagrangian λ_k, a second order approximation matrix B_k, and a trust-region radius δ_k, a trial step s_k is obtained as an approximate solution of the trust-region local model along with an associated Lagrange multiplier μ_k. The step s_k is obtained as $u_k + y_k$ where u_k and y_k are respectively, in a sense that will be specified later, approximate solutions of the normal subproblem

$$(1.3) \qquad (N-STEP_p) \equiv \begin{cases} \text{minimize} & q_{1,k}(u) = \frac{1}{2}\left\|h(x_k) + \nabla h(x_k)^T u\right\|_2^2 \\ \text{subject to} & \|u\|_p \le \tau_1 \delta_k \end{cases}$$

and the tangential subproblem

$$(1.4) \quad (T-STEP_p) \equiv \begin{cases} \text{minimize} & q_{2,k}(y) = (\nabla_x l(x_k,\lambda_k) + B_k u_k)^T y + \frac{1}{2} y^T B_k y \\ \text{subject to} & \nabla h(x_k)^T y = 0 \\ & \|y\|_p \le \tau_2 \delta_k \end{cases}$$

where $\tau_1, \tau_2 \in (0,1]$, and where $\|.\|_p$ denotes an arbitrary norm on $\mathbb{R}^n$ (in practice, $\|.\|_p$ can be either the ℓ_2 or the ℓ_∞ norm). Then, we introduce a penalty parameter update implying that the direction $d_k = (s_k, \mu_k)$ is a descent direction of the merit function that is the augmented Lagrangian, i.e.

$$(1.5) \quad \Phi_k(d) = l(x_k + s, \lambda_k + \mu) + r_k \|h(x_k + s)\|_2^2 .$$

This will enable us to use linesearch techniques to obtain an acceptable steplength t_k and hence a next iterate $x_{k+1} = x_k + t_k s_k$ and an associated Lagrange multiplier $\lambda_{k+1} = \lambda_k + t_k \mu_k$. To avoid that the algorithm breaks down at points satisfying

$$(1.6) \qquad h(x_k) \neq 0 \quad \text{and} \quad \nabla h(x_k) h(x_k) = 0$$

a special treatment is given for such iterates.

The trust-region δ_k, updated in a standard form (see Dennis and Schnabel [13]), is forced, at the beginning of each iteration, to be as large as a fixed arbitrary small positive constant $\delta_{\min}$, during the iteration process δ_k can become smaller than $\delta_{\min}$. Finally, the second order information approximation B_k can be either the exact Hessian of the Lagrangian $\nabla_x^2 l(x_k, \lambda_k)$, its finite difference approximation, or a defined by the BFGS updating formula (see Dennis and Schnabel [13]).

Observe that if $\|.\|_p$ is the ℓ_∞-norm, then $(N-STEP_p)$ and $(T-STEP_p)$ can be formulated as quadratic programming problems. On the other hand, if $\|.\|_p$ is the ℓ_2-norm, then $(T-STEP_p)$ is equivalent to the reduced subproblem

$$(1.7) (RT-STEP_2) \equiv \begin{cases} \text{minimize} & (\nabla_x l(x_k,\lambda_k) + B_k u_k)^T Z_k z + \frac{1}{2} z^T Z_k^T B_k Z_k z \\ \text{subject to} & \|Z_k z\|_2 \le \tau_2 \delta_k \end{cases}$$

where Z_k is a uniformly bounded full rank matrix (e.g. an orthogonal matrix) satisfying $\nabla h(x_k)^T Z_k = 0$. Therefore both Subproblems $(N-STEP_2)$ and $(RT-STEP_2)$ can be solved by techniques such as in Moré and Sorensen [36], Sorensen [49] or Steinhaug [50].

In Section 2, we discuss different stationarity characterizations and we investigate the relations between such characterizations, the consistency of linearized constraints, and the Karush-Kuhn-Tucker stationarity. In Section 3, we define our hybrid algorithm to solve $(EQCP)$. In Section 4, we give the hypotheses that will be used in the present work. In Section 5, we give some descent properties of the normal and tangential components and the trial direction $d_k = (s_k, \mu_k)$ with respect, respectively, to the auxiliary functions $q_{1,k}$ and $q_{2,k}$ and the merit function Φ_k. The global convergence is established in Section 6. In Section 7, we proove that, under the standard hypotheses of Newton method, the rate of convergence of the HYBEQ Algorithm is q-quadratic. Some primary numerical results are given in Section 8. Finally, we give some concluding remarks in Section 9.

2. Stationarity. In this section, we discuss different stationarity characterizations and we investigate the relations between such characterizations, the consistency of linearized constraints and the Karush-Kuhn-Tucker stationarity.

First, we recall the usual Karush-Kuhn-Tucker stationarity characterization.

DEFINITION 2.1. *A point $x \in \mathbb{R}^n$ is said to be a Karush-Kuhn-Tucker point of $(EQCP)$ if there exists $\lambda \in \mathbb{R}^m$ such that*

$$\begin{aligned} \nabla_x l(x,\lambda) &= 0 \\ h(x) &= 0. \end{aligned}$$

On the other hand, from an algorithmic point of view, it seems quite natural to consider the following characterization of stationarity.

DEFINITION 2.2. *Let $(x,\lambda) \in \mathbb{R}^{nxm}$, $\tau_1, \tau_2 \in (0,1], \delta > 0$, and $B \in \mathbb{R}^{n\times n}$. Then x is a stationary point of $(EQCP)$ if $u = 0$ and $y = 0$ are respectively local solutions of the Subproblems $(N - STEP_p)$ and $(T - STEP_p)$.*

Observe that in Definition 2.2, we do not assume that $h(x) = 0$.

In the following two lemmas, we investigate equivalent formulations of the later characterization in terms of the data of $(EQCP)$.

LEMMA 2.1. *Let $x \in \mathbb{R}^n, \tau_1 \in (0,1]$, and $\delta > 0$. Then zero solves the normal subproblem*

$$(2.1) \qquad (N - STEP_p) \equiv \begin{cases} \text{minimize} & c(u) = \frac{1}{2}\left\|h(x) + \nabla h(x)^T u\right\|_2^2 \\ \text{subject to} & \|u\|_p \le \tau_1 \delta \end{cases}$$

if and only if

$$(2.2) \qquad \nabla h(x) h(x) = 0.$$

Proof. The steepest descent of c at zero is $u^{sd} = -\nabla h(x) h(x)$. Therefore, the proof is a consequence of the convexity of the function c. □

LEMMA 2.2. *Let $(x,\lambda) \in \mathbb{R}^{nxm}, \tau_2 \in (0,1]$, $\delta > 0$, and $B \in \mathbb{R}^{n\times n}$. Assume that $h(x) = 0$. If zero is a local solution of the tangential subproblem*

$$(2.3) \qquad (T - STEP_p) \equiv \begin{cases} \text{minimize} & \nabla_x l(x,\lambda)^T y + \frac{1}{2} y^T B y \\ \text{subject to} & \nabla h(x)^T y = 0 \\ & \|y\|_p \le \tau_2 \delta_k, \end{cases}$$

then there exists $\mu \in \mathbb{R}^m$ such that

$$(2.4) \qquad \nabla_x \ell(x, \lambda + \mu) = 0 \ .$$

Proof. Because of the constraints structure, at least one constraint qualification holds at the local solution zero. For constraint qualifications, we refer to Petersen [39] and Fletcher [27, chapter 9]. Therefore, the proof follows from the first order necessary optimality conditions. □

From Lemmas 2.1 and 2.2, we have the following weaker characterization of stationarity.

DEFINITION 2.3. *A point $x \in \mathbb{R}^n$ is said to satisfy the algorithmic stationarity if there exists $\lambda \in \mathbb{R}^m$ such that*

$$\left\{ \begin{array}{rcl} \nabla_x \ell(x,\lambda) & = & 0 \\ \nabla h(x) h(x) & = & 0. \end{array} \right. \tag{2.5}$$

Observe that two cases may happen:

1) $\nabla h(x)h(x) = 0$ and $h(x) = 0$, in which case x is a Karush-Kuhn-Tucker point of $(EQCP)$, or

2) $\nabla h(x)h(x) = 0$ and $h(x) \neq 0$, in which case x is referred to as an infeasible first order point (see Bliss [4] and El-Alem [19]).

In the following lemma, we give a necessary and sufficient condition for an algorithmic stationary point x, i.e. in the sense of Definitions 2.2 or 2.3, to be a feasible point of $(EQCP)$ and hence a Karush-Kuhn-Tucker point of $(EQCP)$.

LEMMA 2.3. *Let $x \in \mathbb{R}^n$ satisfy $\nabla h(x)h(x) = 0$. Then $h(x) = 0$ if and only if the system of linear linearized constraints at x,* i.e.

$$h(x) + \nabla h(x)^T u = 0, \tag{2.6}$$

is consistent.

Proof. It is obvious that $h(x) = 0$ implies that the system in (2.6) is consistent. Conversely, let $u \in \mathbb{R}^n$ be a solution of such a system. From $\nabla h(x)h(x) = 0$, we obtain

$$\begin{array}{rcl} 0 & = & u^T \nabla h(x) h(x) \\ & = & \left(\nabla h(x)^T u\right)^T h(x) \\ & = & -\left\|h(x)\right\|_2^2 \end{array} \tag{2.7}$$

and hence $h(x) = 0$. □

Observe that the usual uniform regularity assumption that $\nabla h(x_k)$ has uniformly full rank is a particular case of the sufficient condition of the result above.

COROLLARY 2.1. *Let x be a stationary point in either sense given in the Definitions 2.2 or 2.3. Then x is a Karush-Kuhn-Tucker point of $(EQCP)$ if and only if the system of linearized constraints at x is consistent.*

Proof. The proof is a consequence of Definitions 2.2 or 2.3, and Lemmas 2.1, 2.2, and 2.3. □

From the following lemma, we obtain, as a necessary condition, a hypothesis implying that a stationary point in the sense of Definitions 2.2 or 2.3 is indeed feasible.

LEMMA 2.4. *Let $\{x_j | j \in J\}$ be a sequence converging to x_*. Assume that $\nabla h(x_*) \neq 0$ and that, for sufficiently large j, the systems of linearized constraints at x_j are consistent. Assume further that $\nabla h(x)$ has constant rank in a neighborhood of x_*. Then there exists a positive constant η_* such that*

$$\left\|\nabla h(x_j) h(x_j)\right\|_2 \geq \eta_* \left\|h(x_j)\right\|_2 \tag{2.8}$$

holds for sufficiently large $j \in J$.

Proof. Without loss of generality, we can assume that $h(x_j) \neq 0$ for all $j \in J$. Since we are assuming the asymptotic constant rank, we can consider, without loss of generality, that such a rank is r_*. Let $U_j \Sigma_j V_j^T$ be the singular value decomposition of $\nabla h(x_j)$ where U_j and V_j are orthogonal. We have

$$U_j = (U_{j,1}, U_{j,2}) \quad , \quad \Sigma_j = \begin{pmatrix} \Sigma_{j,1} & 0 \\ 0 & 0 \end{pmatrix} \quad , \quad V_j = (V_{j,1}, V_{j,2}) \tag{2.9}$$

where $\Sigma_{j,1}$ is a diagonal matrix whose r_* entries are the positive singular values of $\nabla h(x)$, $U_{j,1}$ and $V_{j,1}$ are respectively the matrices of the r_* left and right eigenvectors of $\nabla h(x_j)$ (see Golub and Van Loan [28]).

Since the linear system $\nabla h(x_j)^T u + h(x_j) = 0$ is consistent for sufficiently large j, we have

$$V_{j,2}^T h(x_j) = 0 \tag{2.10}$$

(see El Hallabi [21] and [23]). Indeed, let u_j be a solution of such a system. We have

$$V_{j,1}^T h(x_j) + \Sigma_{j,1} U_{j,1}^T u_j = 0 \quad \text{and} \quad V_{j,2}^T h(x_j) = 0. \tag{2.11}$$

Assume that there exists a subsequence for which

$$\lim_{j \to \infty} \frac{\|\nabla h(x_j) h(x_j)\|_2}{\|h(x_j)\|_2} = 0 \tag{2.12}$$

holds. We have

$$\frac{\|\nabla h(x_j) h(x_j)\|_2}{\|h(x_j)\|_2} = \frac{\left\|U_{j,1}\Sigma_{j,1}V_{j,1}^T h(x_j)\right\|_2}{\|h(x_j)\|_2} = \frac{\left\|\Sigma_{j,1}V_{j,1}^T h(x_j)\right\|_2}{\|h(x_j)\|_2}. \tag{2.13}$$

Now, since for sufficiently large j, the rank is identically equal to r_*, we obtain from (2.12) and (2.13) that

$$\lim_{j \to \infty} \frac{\left\|V_{j,1}^T h(x_j)\right\|_2}{\|h(x_j)\|_2} = 0, \tag{2.14}$$

which, together with (2.10), implies that

$$\lim_{j \to \infty} \frac{\left\|V_j^T h(x_j)\right\|_2}{\|h(x_j)\|_2} = 0. \tag{2.15}$$

This contradicts the orthogonality of V_j. Therefore, there exists a positive constant η_* such that (2.8) holds. □

3. A hybrid algorithm for equality constrained optimization. In this section we define our hybrid algorithm ($HYBEQ$) for equality constrained optimization. In our approach, we propose a heuristic that allows the algorithm not to break down at a point x_k satisfying $\nabla h(x_k)h(x_k) = 0$ and $h(x_k) \neq 0$. We will start by detailing the different stage of the algorithm.

3.1. The search direction. Let $\tau_1, \tau_2 \in (0,1]$ and let $\|.\|_p$ denote an arbitrary ℓ_p-norm on $\mathbb{R}^n$ (in practice it canbe either the ℓ_2-norm or the ℓ_∞-norm) . Also let $B_k \in \mathbb{R}^{nxn}$, and $\delta_k > 0$. Our trial search direction is of the form $d_k = (s_k, \mu_k)$ where s_k is the trial step and μ_k an associated Lagrange multiplier. The trial step s_k is of the form $u_k + y_k$ where the normal component u_k and the tangential component y_k are, in a sense that will be specified later, respectively approximate solutions of Subproblems $(N - STEP_p)$ and $(T - STEP_p)$ defined respectively in (1.3) and (1.4).

3.1.a) The normal component. If $\nabla h(x_k)h(x_k) = 0$, we set $u_k = 0$. Let us consider the case where $\nabla h(x_k)h(x_k) \neq 0$. The normal component u_k is a feasible point of $(N - STEP_p)$ that satisfies the usual fraction of Cauchy decrease condition

$$q_{1,k}(u_k) \leq q_{1,k}(0) + \mu_1 \left[q_{1,k}(\alpha_k^{sd} u_k^{sd}) - q_{1,k}(0)\right] \tag{3.1}$$

where

$$u_k^{sd} = -\nabla h(x_k)h(x_k) \tag{3.2}$$

is the steepest descent of $q_{1,k}$ at zero and α_k^{sd} is the minimizer of $q_{1,k}(\alpha u_k^{sd})$ inside the feasible region of $(N - STEP_p)$, i.e.

$$\alpha_k^{sd} = \begin{cases} \frac{(-\nabla h(x_k)h(x_k))^T u_k^{sd}}{\left\|\nabla h(x_k)^T u_k^{sd}\right\|_2^2} & \text{if } \frac{(-\nabla h(x_k)h(x_k))^T u_k^{sd}}{\left\|\nabla h(x_k)^T u_k^{sd}\right\|_2^2} \left\|u_k^{sd}\right\|_p \leq \tau\delta_k \\ \tau \frac{\delta_k}{\left\|u_k^{sd}\right\|_p} & \text{othewise,} \end{cases} \tag{3.3a}$$

or equivalently

$$\alpha_k^{sd} = \min\left(\tau \frac{\delta_k}{\left\|u_k^{sd}\right\|_p}, \frac{(-\nabla h(x_k)h(x_k))^T u_k^{sd}}{\left\|\nabla h(x_k)^T u_k^{sd}\right\|_2^2}\right) \tag{3.3b}$$

(In Section 4, we show that α_k^{sd} is well defined in the sense that the denominators in (3.3a,b) are not zero).

3.1.b) The tangential component. To obtain the tangential component y_k, we will need the (tangential) steepest descent of $(T-STEP_p)$ that is either defined to be

$$y_k^{sd} = -Z_k Z_k^T(\nabla_x l(x_k, \lambda_k) + B_k u_k), \tag{3.4}$$

where Z_k is a uniformly bounded full rank matrix satisfying

$$\nabla h(x_k)^T Z_k = 0. \tag{3.5}$$

We consider two cases:

3.1.b.1) First, we consider the case where $y_k^{sd} \neq 0$. Then y_k is an approximate solution of the local model subproblem $(T - STEP_p)$ in the sense that it satisfies

$$\max\left(\ell_k(y_k); q_{2,k}(y_k)\right) \leq \mu_2 \max\left(\ell_k(\beta_k y_k^{sd}); q_{2,k}(\beta_k y_k^{sd})\right) \tag{3.7}$$

where $\mu_2 \in (0,1)$, and β_k is the minimum of $q_{2,k}(\beta y_k^{sd})$ inside the feasible region of $(T - STEP_p)$, i.e.

$$\beta_k = \begin{cases} \beta_k^* & \text{if } \left(y_k^{sd}\right)^T B_k y_k^{sd} > 0 \text{ and } \left\|\beta_k^* y_k^{sd}\right\|_p \leq \tau_2\delta_k \\ \frac{\tau_2\delta_k}{\left\|y_k^{sd}\right\|_p} & \text{otherwise} \end{cases} \tag{3.8}$$

with

$$\beta_k^* = -\frac{(\nabla_x l(x_k,\lambda_k)+B_k u_k)^T y_k^{sd}}{\left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right)}. \tag{3.9}$$

Observe that if y_k^{sd} is defined by (3.4) then

$$\beta_k^* = \frac{\left\|Z_k^T(\nabla_x l(x_k,\lambda_k)+B_k u_k)\right\|_2^2}{\left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right)}, \tag{3.10}$$

moreover, if Z_k is orthogonal then

$$\beta_k^* = \frac{\left\|y_k^{sd}\right\|_2^2}{\left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right)}.$$

REMARK 3.1. The test in (3.7) can always be satisfied. A y_k can be obtained by approximately minimizing $(T - STEP_p)$, with $\beta_k y_k^{sd}$ as initial point, until

$$q_{2,k}(y_k) \leq \mu_2 \max\left(q_{2,k}(\beta_k y_k^{sd}), \ell_k(\beta_k y_k^{sd})\right) \tag{3.11}$$

is satisfied. Assume that $y_k B_k y_k \geq 0$. Then (3.11) is equivalent to (3.7). On the other hand, assume that $y_k B_k y_k < 0$. Then we accept y_k if

$$\ell_k(y_k) = \max\left(q_{2,k}(y_k), \ell_k(y_k)\right) \leq \mu_2 \max\left(q_{2,k}(\beta_k y_k^{sd}), \ell_k(\beta_k y_k^{sd})\right) \tag{3.12a}$$

is satisfied, otherwise we accept $-y_k$ if it satisfies

$$\ell_k(-y_k) = \max\left(q_{2,k}(-y_k), \ell_k(-y_k)\right) \leq \mu_2 \max\left(q_{2,k}(\beta_k y_k^{sd}), \ell_k(\beta_k y_k^{sd})\right). \tag{3.12b}$$

Otherwise we set $y_k = \beta_k y_k^{sd}$, in which case it is obvious that y_k satisfies (3.7).

3.1.b.2) Second, we consider the case where $y_k^{sd} = 0$. We consider two subcases:

3.1.b.2.a) If $\nabla h(x_k)h(x_k) \neq 0$ (equivalently $u_k \neq 0$) we set $y_k = 0$.

3.1.b.2.b) Assume that $\nabla h(x_k)h(x_k) = 0$ (equivalently $u_k = 0$), which implies that x_k is a stationary point of $(EQCP)$ in the sense of the Definition 2.3. Since x_k is not a Karush-Kuhn-Tucker point of $(EQCP)$, we obtain from Lemma 2.2 and Definition 2.1 that $h(x_k) \neq 0$ necessarily holds. The idea is to force the algorithm to leave such a point which will be actually obtained through the heuristic we propose for choosing the penalty parameter in this case, provided that $y_k \neq 0$. Therefore we choose y_k to be any feasible point of $(T - STEP_p)$ satisfying $\|y_k\|_p = \tau_2\delta_k$.

3.2. Lagrange multiplier. If x_k satisfies

$$\text{(a) } h(x_k) \neq 0 \quad \text{and} \quad \text{(b) } \|\nabla h(x_k)h(x_k)\|_2 < \varepsilon \min\left(1, \|h(x_k)\|_2\right) \tag{3.13a}$$

where ε is an arbitrary small constant in $(0,1)$, we define μ_k by

(3.13b) $$\mu_k = -\varepsilon \frac{h(x_k)}{\|h(x_k)\|_2}.$$

This heuristic is used to force the direction $d_k = (s_k, \mu_k)$ to be a descent direction of the merit function at (x_k, λ_k) as will be established in Section 5, and hence allows the algorithm not to break down at such a point.

Otherwise, the Lagrange multiplier μ_k associated with the trial step s_k is updated as follows:

3.2.1) If $y_k^{sd} \neq 0$, then μ_k is the associated Lagrange multiplier to y_k.

3.2.2) Assume that $y_k^{sd} = 0$, which implies, together with Lemma 2.3 since x_k is not a Karush-Kuhn-Tucker point of $(EQCP)$, that we necessarily have $h(x_k) \neq 0$. Then, we choose μ_k to be the least squares solution of

$$\nabla h(x)\mu_k = \nabla_x l(x_k, \lambda_k) + B_k s_k,$$

equivalently $\lambda_{k+1} = \lambda_k + \mu_k$ to be the least squares solution of

$$\nabla_x l(x_k, \lambda) + B_k s_k = 0.$$

3.3. Merit function and penalty parameter. To accept or reject a trial step, we will use the actual reduction

(3.14) $$ared_k(d) = \Phi_k(d) - \Phi_k(0)$$

and the predicted reduction

(3.15) $$pred_k(d) = \Psi_k(d) - \Psi_k(0)$$

where Φ is the merit function defined by

(3.16) $$\Phi_k(d) = l(x_k + s, \lambda_k + \mu) + r_k \|h(x_k + s)\|_2^2,$$

i.e. the augmented Lagrangian, and where Ψ is the predicted merit function defined by

(3.17) $$\Psi_k(d) = l(x_k, \lambda_k) + \nabla_x l(x_k, \lambda_k)^T d + \tfrac{1}{2} d^T W_k d + r_k \left\| h(x_k) + \nabla h(x_k)^T s \right\|_2^2.$$

In (3.16) and (3.17), r_k denotes the penalty parameter defined by

(3.18) $$r_k = \max(\widehat{r}_k, r_{k-1})$$

where, for a given positive constant ρ and an arbitrary small positive constant ε, $\widehat{r}_k$ is given by

(3.19) $$\widehat{r}_k = \begin{cases} \rho & \text{if } \nabla h(x_k)h(x_k) = 0 \\ 2\max\left(0, \frac{\nabla l(x_k,\lambda_k)^T d_k + \frac{1}{2} d_k^T W_k d_k}{\|h(x_k)\|_2^2 - \|h(x_k) + \nabla h(x_k)^T s_k\|_2^2}\right) + \rho & \text{otherwise.} \end{cases}$$

Observe that s_k is a descent direction of the squared ℓ_2-norm of the constraints (this will be established in Lemma 5.3).

In the following, we give the major reason for the use of (3.19).

3.4. Acceptance-rejection test. In our algorithm, the acceptance-rejection test uses the directional derivative of $\Phi_k(.)$ at zero along with the predicted and actual reductions.

If

(3.23) $$ared_k(d_k) \leq c_1\, pred_k(d_k)$$

holds, then the trial step s_k (and the related Lagrange multiplier μ_k) is accepted, and we set

Otherwise, we choose a trust-region δ_{k+1} such that

(3.27b) $$c_3 \delta_k \leq \delta_{k+1} \leq c_4 \delta_k$$

and resolve the subproblem defined by $(N - STEP_p)$ and $(T - STEP_p)$.

(3.25) $$x_{k+1} = x_k + s_k \quad \text{and} \quad \lambda_{k+1} = \lambda_k + \mu_k.$$

3.5. Trust-region update. To update the trust-region radius we use the following scheme. Let $c_2 \in (c_1, 1), 1 < c_5 < c_6$ and $0 < \delta_{\min}$. If

(3.27a) $$ared_k(d_k) \leq c_2 pred_k(d_k)$$

then we Choose δ_{k+1} such that

(3.27b) $$c_5\delta_k \leq \delta_{k+1} \leq c_6\delta_k$$

otherwise, we set

(3.27c) $$\delta_{k+1} = \|s_k\|_p .$$

Moreover to ensure that the trust-region radius is bounded away from zero at a non-stationary point, we correct the trust-region as follows

(3.27d) $$\delta_{k+1} = \max\left(\delta_{\min}, \min\left(\delta_{\max}, \delta_{k+1}\right)\right) .$$

Such a safeguard strategy is used in Alexandrov [1], Alexandrov and Dennis [2], Dennis, El-Alem, and Maciél [11], and El Hallabi [21] and [23], El Hallabi and Tapia [25], Zhang, Kim, and Lasdon [53], Dennis, Li, and Tapia [12].

3.6. Definition of the trust-region algorithm ($HYBEQ$). Let $c_i, i = 1,..,6$, ρ, $\delta_{\min}$ and $\delta_{\max}$ be constants satisfying

$$0 < c_1 < c_2 < 1 \qquad 0 < c_3 < c_4 < 1 < c_5 < c_6 .$$

Let $x_0 \in \mathbb{R}^n$ be an arbitrary point and λ_0 an associated Lagrange multiplier, $B_0 \in \mathbb{R}^{n\times n}$ be an arbitrary symmetric matrix, $0 < \delta_{\min} \leq \delta_0 < \delta_{\max}$, and $r_0 \geq \rho$.

Let x_k be the iterate, λ_k the associated Lagrange multiplier, δ_k the trust-region radius and B_k the second order information approximation given by the k^{th} iteration. The algorithm generates $x_{k+1}, \lambda_{k+1}, \delta_{k+1}$ and B_{k+1} by the following iterative scheme:

STEP 1 . Obtain a trial step s_k and an associated Lagrange multiplier μ_k and set $d_k = (s_k, \mu_k)$ (see Sections 3.1 and 3.2);

STEP 2 . Update the penalty parameter r_k using (3.18) and (3.19);

STEP 3 . Perform a acceptance-rejection test

STEP 4 . Update the trust-region radius according to (3.27a,b,c,d).

STEP 5 . Choose a symmetric approximation $B_{k+1} \in \mathbb{R}^{n\times n}$ of $\nabla^2 l(x_{k+1}, \lambda_{k+1})$;

4. Hypotheses. In this section, we give the hypotheses that will be used to establish that our hybrid algorithm is globally convergent.

GCH.1) the functions f and h are continuously differentiable;

GCH.2) the iteration sequence $\{x_k | k \in \mathbb{N}\}$ is contained in a bounded convex subset of $\mathbb{R}^n$;

GCH.3) the sequence of Lagrange multipliers $\{\lambda_k | k \in \mathbb{N}\}$ is contained in a bounded convex subset of $\mathbb{R}^m$;

GCH.4) there exists a positive constant ν_*^n such that
$$\|B_k u_k\|_2 \leq \nu_*^n \|u_k\|_2;$$

GCH.5) there exists a positive constant ν_*^t such that
$$\|B_k y_k\|_2 \leq \nu_*^t \|y_k\|_2 ;$$

Moreover, if $\{x_k | k \in N\}$ is a subsequence converging to some x_*, then

GCH.6) the gradient of the constraints at x_* is not zero, i.e. $\nabla h(x_*) \neq 0$;

GCH.7) there exists an arbitrary small positive constant κ_* such that
$$\|u_k\|_2 \leq \kappa_* \|\nabla h(x_k) h(x_k)\|_2 ;$$

REMARK 4.1. It is obvious that Hypothesis GCH.7 is satisfied by u_k^{sd}.

5. Descent properties. In this section we give some descent properties of the normal and tangential components u_k and y_k and the trial direction $d_k = (s_k, \mu_k)$ respectively to the objective functions of $(N - STEP_p)$ and $(\overline{T - STEP_p})$ and the merit function. Such properties are used to show that d_k is a descent direction of the merit function at (x_k, λ_k).

In the following lemma, we show that α_k^{sd} given in (3.3a,b) is well defined.

LEMMA 5.2. *Let $x_k \in \mathbb{R}^n$. Then $\nabla h(x_k)h(x_k) = 0$ if and only if $\nabla h(x_k)^T u_k^{sd} = 0$.*

Proof. We have

$$\nabla h(x_k)^T u_k^{sd} = -\nabla h(x_k)^T \nabla h(x_k) h(x_k). \tag{5.1}$$

It is obvious that $\nabla h(x_k)h(x_k) = 0$ implies that $\nabla h(x_k)^T u_k^{sd} = 0$. Conversely, assume that $\nabla h(x_k)^T u_k^{sd} = 0$. Then, we obtain from (5.1)

$$\|\nabla h(x_k)h(x_k)\|_2^2 = h(x_k)^T \nabla h(x_k)^T \nabla h(x_k) h(x_k) = 0. \;\square$$

In the following lemma, we show that the normal component u_k is a descent direction of the ℓ_2-norm of the linearized constraints at x_k.

LEMMA 5.3. *Assume that $\nabla h(x_k)h(x_k) \neq 0$. Then*

$$q_{1,k}(\alpha_k^{sd} u_k^{sd}) < q_{1,k}(0) \tag{5.2}$$

and hence

$$\text{(a) } q_{1,k}(u_k) < q_{1,k}(0) \qquad \text{(b) } \nabla q_{1,k}(0)^T u_k < 0. \tag{5.3}$$

Proof. Consider the function

$$q_{1,k}(tu_k) = \left\| h(x_k) + t\nabla h(x_k)^T u_k \right\|_2^2 .$$

Since the function $q_{1,k}$ is convex, it satisfies

$$t\nabla q_{1,k}(0)^T u_k \leq q_{1,k}(tu_k) - q_{1,k}(0) \tag{5.4}$$

for all $t \in [0, 1]$ and (5.3b) follows from (5.3a). The definitions of u_k^{sd} and α_k^{sd} imply that (5.2) hold. Finally, from the choice of u_k, i.e. (3.1), we obtain that (5.3a) holds. $\square$

For the tangential component, we have

$$\left(\nabla_x l(x_k, \lambda_k) + B_k u_k\right)^T y_k \leq 0 \tag{5.5}$$

which follows from (3.7). Moreover we have the following lemma.

LEMMA 5.4. *Assume that x_k is not an algorithmic stationary point of $(EQCP)$, i.e. in the sense of Definition 2.3. Assume further that $\nabla h(x_k)h(x_k) = 0$. Then*

$$\nabla_x l(x_k, \lambda_k)^T y_k < 0 \quad \text{and} \quad \nabla_x l(x_k, \lambda_k)^T y_k + \tfrac{1}{2} y_k^T B_k y_k < 0 \tag{5.6a}$$

holds and hence

$$\nabla_x l(x_k, \lambda_k)^T y_k + \tfrac{1}{2} \max\left(0, y_k^T B_k y_k\right) < 0. \tag{5.6b}$$

Proof. We have $u_k = 0$. Assume that (5.6a) does not hold. Therefore, we have

$$\nabla_x l(x_k, \lambda_k)^T y_k = 0 \quad \text{or} \quad \nabla_x l(x_k, \lambda_k)^T y_k + \tfrac{1}{2} y_k^T B_k y_k = 0$$

i.e. $\ell_k(y_k) = 0$ or $q(y_k) = 0$, which, together with (3.7), implies that

$$\max\left(\ell_k(\beta_k y_k^{sd}), q_{2,k}(\beta_k y_k^{sd})\right) = 0. \tag{5.7}$$

This implies that $y_k^{sd} = 0$, or equivalently $Z_k^T \nabla_x l(x_k, \lambda_k) = 0$, which contradicts our hypothesis that x_k is not an algorithmic stationary point of $(EQCP)$. $\square$

COROLLARY 5.2. *Assume that x_k is not a Karush-Kuhn-Tucker point of $(EQCP)$. Assume further that $h(x) = 0$. Then*

$$\nabla_x l(x_k, \lambda_k)^T y_k + \tfrac{1}{2} \max\left(0, y_k B_k y_k\right) < 0.$$

The result of the following lemma is the key point of the heuristic proposed in Subsection 3.2 for an infeasible first order point.

LEMMA 5.5. *Let x_k* satisfy *$\nabla h(x_k)h(x_k) = 0$ and $h(x_k) \neq 0$. Then*

$$\nabla l(x_k, \lambda_k)^T d_k < 0 \quad \text{and} \quad \nabla l(x_k, \lambda_k)^T d_k + \tfrac{1}{2} \max\left(0, d_k^T W_k d_k\right) < 0 \tag{5.8}$$

holds.

Proof. Since $\nabla h(x_k)h(x_k) = 0$, we have $u_k = 0$, and

$$\nabla l(x_k, \lambda_k)^T d_k = \nabla_x l(x_k, \lambda_k)^T y_k + h(x_k)^T \mu_k \tag{5.9}$$

where, because of the heuristic used in Subsection 3.2,

$$h(x_k)^T \mu_k = -\varepsilon \left\| h(x_k) \right\|_2 , \tag{5.10}$$

and

(5.11) $$d_k^T W_k d_k = s_k^T B_k s_k = y_k B_k y_k$$
which implies that
(5.12) $$\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) = \nabla_x l(x_k,\lambda_k)^T y_k + \tfrac{1}{2}\max\left(0, y_k B_k y_k\right) + h(x_k)^T \mu_k.$$
On the other hand, we have
(5.13) $$\nabla_x l(x_k,\lambda_k)^T y_k \le 0 \quad \text{and} \quad \nabla_x l(x_k,\lambda_k)^T y_k + \tfrac{1}{2} y_k^T B_k y_k \le 0.$$
Therefore, we obtain from (5.9), (5.10), (5.11), (5.12) and (5.13) that (5.8) holds. □

Now, we give some properties of the directional derivative of Φ_k at zero.

LEMMA 5.6. *Let* $z_k = (x_k, \lambda_k) \in \mathbb{R}^n \mathrm{x} \mathbb{R}^m$. *We have*
(5.14) $$\nabla\Psi_k(0) = \nabla\Phi_k(0)$$
Moreover it satisfies
(5.15) $$\nabla\Psi_k(0)^T s \le \nabla l(x_k,\lambda_k)^T d + r_k \left[\left\|h(x_k) + \nabla h(x_k)^T s\right\|_2^2 - \|h(x_k)\|_2^2\right].$$

Proof. It is obvious that (5.14) holds. Now, Inequality (5.15) is a consequence of (5.14) and (5.4). □

In the following proposition, we prove that, because of the definition of the penalty parameter r_k, and the heuristic defined in Subsection 3.2, the trial direction d_k is a descent direction of the merit function $\Phi_k(.)$ at the origin.

PROPOSITION 5.1. *Let* $(x_k, \lambda_k) \in \mathbb{R}^{nxm}, \delta_k > 0$, *and* $B_k \in \mathbb{R}^{n\times n}$. *Assume that* x_k *is not a Karush-Kuhn-Tucker point of* $(EQCP)$. *Then*
(5.16) $$pred_k(td_k) \le -t\left\{\left|\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right)\right| + \rho\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T s_k\right\|_2^2\right]\right\}$$
holds for any $t \in (0,1]$. *Moreover we have*
(5.17) $$\nabla\Phi_k(0)^T d_k \le -\left|\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right)\right| - \rho\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T s_k\right\|_2^2\right],$$
and consequently d_k *is a descent direction of* Φ_k *at the origin.*

Proof. Observe that (5.17) follows from (5.16). Let us show (5.16). We have, for $t \in (0,1]$,
$$pred_k(td_k) = t\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}t^2 d_k^T W_k d_k + r_k\left[\left\|h(x_k) + t\nabla h(x_k)^T s_k\right\|_2^2 - \|h(x_k)\|_2^2\right]$$
which implies that
(5.18) $$pred_k(td_k) \le t\left\{\left[\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right)\right] - r_k\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T s_k\right\|_2^2\right]\right\}.$$
Now, assume that
(5.19) $$\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) \ge 0.$$
Then $\nabla h(x_k)h(x_k) \ne 0$ must hold. Indeed, assume that $\nabla h(x_k)h(x_k) = 0$ which, together with Lemma 5.5, implies that $h(x_k) = 0$. Then we obtain from (5.9), (5.11) and Lemma 5.4 that
$$\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) < 0$$
which contradicts (5.19). Therefore, we have
$$r_k \ge 2\frac{\nabla l(x_k,\lambda_k)^T d_k + \frac{1}{2}\max\left(0, d_k^T W_k d_k\right)}{\|h(x_k)\|_2^2 - \|h(x_k) + \nabla h(x)^T s_k\|_2^2} + \rho$$

which together with (5.19), implies that (5.16) holds. Finally, assume that (5.19) does not hold, i.e. we have

$$\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) < 0. \tag{5.20}$$

We have $r_k \geq \rho$. Therefore (5.18) implies that

$$\begin{aligned} pred_k(td_k) \leq t\Big\{ & \left[\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right)\right] - \\ & \rho\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T s_k\right\|_2^2\right]\Big\}. \end{aligned} \tag{5.21}$$

0Using (5.20), we rewrite (5.21) as (5.16).

If $\nabla h(x_k)h(x_k) \neq 0$ we obtain from (5.17), Lemma 5.3, and the choice of u_k that $\nabla\Phi_k(0)^T d_k < 0$. On the other hand, assume that $\nabla h(x_k)h(x_k) = 0$. First, if $h(x_k) \neq 0$, then we obtain from Lemma 5.5 that $\nabla\Phi_k(0)^T d_k < 0$ holds. Second, assume that $h(x_k) = 0$. Then, since x_k is not a Karush-Kuhn-Tucker point of $(EQCP)$, we obtain from Corollary 5.2 that $\nabla\Phi_k(0)^T d_k < 0$ holds. □

We end this section by proving in the following theorem that, unless the iterate x_k is a Karush-Kuhn-Tucker point of $(EQCP)$, the $HYBEQ$ Algorithm finds an acceptable steplength by decreasing the t_k a finite number of times.

THEOREM 5.1. *Assume that x_k is not a Karush-Kuhn-Tucker point of $(EQCP)$. Then the algorithm finds an acceptable steplength t_k by backtracking a finite number from the steplength one.*

Proof. Assume that the algorithm decreases t indefinitely without obtaining an acceptable steplength. This implies that

$$ared_k(t_k d_k) \geq c_1 pred_k(t_k d_k) \tag{5.22}$$

which, by simple calculations and because of (5.14), leads to

$$\nabla\Phi_k(0)^T d_k \geq 0. \tag{5.23}$$

This contradicts the fact that d_k is a descent direction of $\Phi_k(.)$ at zero. □

Theorem 5.1 implies that either the $HYBEQ$ Algorithm generates a sequence $\{x_j | j = 1,..,l\}$ such that x_l 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 $HYBEQ$ Algorithm generates an infinite sequence $\{x_k\}$.

6. Global convergence. In this section, we demonstrate that the $HYBEQ$ Algorithm is globally convergent.

The derivation of some properties of the algorithm near nonstationary points will play a crucial role in our global convergence theory analysis, especially the following two lemmas.

LEMMA 6.1. *Assume that a subsequence $\{x_k | k \in N\}$, where $\nabla h(x_k)h(x_k) \neq 0$, converges to x_*. Also assume that Hypotheses H.6-7 hold. If*

$$\lim_{k\in N\to+\infty}\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T u_k\right\|_2^2\right] = 0 \tag{6.1}$$

holds. Then

$$\text{(a)} \lim_{k\in N\to+\infty} \nabla h(x_k)h(x_k) = 0, \qquad \text{and} \qquad \text{(b)} \lim_{k\in N\to+\infty} u_k = 0. \tag{6.2}$$

Moreover

$$\alpha_k^{sd} = \frac{(-\nabla h(x_k)h(x_k))^T u_k^{sd}}{\left\|\nabla h(x_k)^T u_k^{sd}\right\|_2^2} \tag{6.3}$$

holds for sufficiently large $k \in N$, and hence $\left\{\alpha_k^{sd} | k \in N\right\}$ is bounded away from zero.

Proof. The limit in (6.2b) follows from (6.2a) and hypothesis H.7. Let us prove (6.2a). We obtain from (3.1)

$$\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T u_k\right\|_2^2\right] \geq c_1 \left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \alpha_k^{sd} \nabla h(x_k)^T u_k^{sd}\right\|_2^2\right] \tag{6.4}$$

where

$$\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \alpha_k^{sd} \nabla h(x_k)^T u_k^{sd}\right\|_2^2\right] = \alpha_k^{sd} \left(-\nabla h(x_k) h(x_k)\right)^T u_k^{sd} \left(2 - \alpha_k^{sd} \frac{\left\|\nabla h(x_k)^T u_k^{sd}\right\|_2^2}{(-\nabla h(x_k)h(x_k))^T u_k^{sd}}\right). \tag{6.5}$$

Also, because of the definition of α_k^{sd}, i.e. (3.3), we have

$$0 < \alpha_k^{sd} \frac{\left\|\nabla h(x_k)^T u_k^{sd}\right\|_2^2}{(-\nabla h(x_k)h(x_k))^T u_k^{sd}} \leq 1$$

which, together with (6.5), implies that

$$\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \alpha_k^{sd} \nabla h(x_k)^T u_k^{sd}\right\|_2^2\right] \geq \alpha_k^{sd} \left(-\nabla h(x_k) h(x_k)\right)^T u_k^{sd}$$

or equivalently

$$\left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \alpha_k^{sd} \nabla h(x_k)^T u_k^{sd}\right\|_2^2\right] \geq \alpha_k^{sd} \left\|\nabla h(x_k) h(x_k)\right\|_2^2. \tag{6.6}$$

From (6.1) and (6.6), we obtain

$$\lim_{k \in N \to +\infty} \alpha_k^{sd} \left\|\nabla h(x_k) h(x_k)\right\|_2^2 = 0. \tag{6.7}$$

The definition of α_k^{sd}, together with (6.7), implies that

$$\lim_{k \in N \to +\infty} \min\left(\tau \frac{\|\nabla h(x_k)h(x_k)\|_2^2}{\|\nabla h(x_k)h(x_k)\|_p}, \frac{\|\nabla h(x_k)h(x_k)\|_2^4}{\|\nabla h(x_k)^T \nabla h(x_k) h(x_k)\|_2^2}\right) = 0 \tag{6.8}$$

where

$$\frac{\|\nabla h(x_k)h(x_k)\|_2^2}{\|\nabla h(x_k)^T \nabla h(x_k) h(x_k)\|_2^2} \geq \frac{1}{\|\nabla h(x_k)\|_2^2}. \tag{6.9}$$

Therefore we obtain from (6.8), (6.9), and the norm equivalency that

$$\lim_{k \in N \to +\infty} \|\nabla h(x_k)h(x_k)\|_2 \min\left(\tau, \frac{\|\nabla h(x_k)h(x_k)\|_2}{\|\nabla h(x_k)\|_2^2}\right) = 0 \tag{6.10}$$

which implies that (6.2a) holds. Finally, the boundedness away from zero of $\left\{\alpha_k^{sd} | k \in N\right\}$ follows from the definition of α_k^{sd} in (3.3), (6.2b), (6.9) and Hypothesis H.6. □

LEMMA 6.2. *Assume that the hypotheses H1-7 hold. Moreover assume that* $\{x_k | k \in N\}$, *where* $\nabla h(x_k)h(x_k) \neq 0$, *is a subsequence converging to* x_*. *If*

$$\lim_{k \in N \to +\infty} \left[\|h(x_k)\|_2^2 - \left\|h(x_k) + \nabla h(x_k)^T u_k\right\|_2^2\right] = 0 \tag{6.11}$$

and

$$\liminf_{k \in N \to +\infty} \left[\nabla l(x_k, \lambda_k)^T d_k + \tfrac{1}{2} \max\left(0, d_k^T W_k d_k\right)\right] \geq 0, \tag{6.12}$$

then x_* *is necessarily a Karush-Kuhn-Tucker point of* $(EQCP)$.

Proof. First, we show that $\{h(x_k) | k \in N\}$ converges to zero. We have

$$\begin{aligned} \nabla l(x_k, \lambda_k)^T d_k + \tfrac{1}{2} \max\left(0, d_k^T W_k d_k\right) &= \nabla_x l(x_k, \lambda_k)^T s_k + h(x_k)^T \mu_k \\ &\quad + \tfrac{1}{2} \max\left(0, 2\left(\nabla h(x_k)^T u_k\right)^T \mu_k + s_k^T B_k s_k\right). \end{aligned} \tag{6.13}$$

On the other hand, we obtain from Lemma 6.1 and (6.11) that

$$\text{(a)} \lim_{k \in N \to +\infty} \nabla h(x_k) h(x_k) = 0, \quad \text{and} \quad \text{(b)} \lim_{k \in N \to +\infty} u_k = 0. \tag{6.14}$$

Assume that some subsequence $\{h(x_j) | j \in j \subset N\}$ is bounded away from zero. This, together with (6.14a), implies that

$$\lim_{j \in J \to +\infty} \frac{\|\nabla h(x_j) h(x_j)\|_2}{\|h(x_j)\|_2} = 0$$

and hence

$$\mu_j = -\varepsilon \frac{h(x_j)}{\|h(x_j)\|_2} \tag{6.15}$$

holds for sufficiently large $j \in J$ regarding to the heuristic used in Subsection 3.2. Therefore, we obtain from (6.14b), (6.12), (6.13) and (6.15) that

$$(6.16)\qquad \liminf_{j\in J\to+\infty} \nabla_x l(x_j,\lambda_j)^T y_j + \tfrac{1}{2}\max\left(0; y_j^T B_j y_j\right) \geq \varepsilon \liminf_{j\in J\to+\infty} \|h(x_j)\|_2,$$

and consequently

$$(6.17)\qquad \liminf_{j\in J\to+\infty} \max\left(\ell_j(y_j), q_j(y_j)\right) > 0,$$

which contradicts the choice of y_j in (3.7). Therefore we obtain

$$(6.18)\qquad \liminf_{k\in N\to+\infty} h(x_k) = 0.$$

Now, we show that x_* is a Karush-Kuhn-Tucker point of $(EQCP)$. Let y_*^{sd} be an accumulation point of $\left\{y_k^{sd} | k \in N\right\}$. Without loss of generality, we can assume that y_*^{sd} is a limit.

If $\nabla_x l(x_*,\lambda_*)^T y_*^{sd} = 0$ then we obtain from Lemma 2.2 and (6.18) that x_* is a Karush-Kuhn-Tucker point of $(EQCP)$. Assume that

$$(6.19)\qquad \nabla_x l(x_*,\lambda_*)^T y_*^{sd} < 0.$$

Observe that this implies that β_k is bounded. Let us show that it is uniformly bounded away from zero. Recall that β_k is a minimizer of $q_{2,k}(\beta y_k^{sd})$ over the trust-region of the $(\overline{T-STEP_p})$ Subproblem. Let us consider the auxiliary function

$$\widetilde{q}_k(\beta) = q_{2,k}(\beta y_k^{sd}) = \tfrac{1}{2}\beta^2 \left(y_k^{sd}\right)^T B_k y_k^{sd} + \beta\left(\nabla_x l(x_k,\lambda_k) + B_k u_k\right)^T y_k^{sd}.$$

We have, using (6.14b),

$$\nabla\widetilde{q}_k(0) = \left(\nabla_x l(x_k,\lambda_k) + B_k u_k\right)^T y_k^{sd} \leq \tfrac{1}{2}\nabla_x l(x_*,\lambda_*)^T y_*^{sd} < 0$$

for sufficiently large $k \in N$. Hence, there exist a positive integer, say k_*, and a constant $\beta_{\min} > 0$ such that

$$\nabla\widetilde{q}_k(\beta) \leq \tfrac{1}{4}\nabla_x l(x_*,\lambda_*)^T y_*^{sd} < 0$$

for $k \geq k_*$ and $\beta \in [0, \beta_{\min}]$ which implies that

$$(6.20)\qquad \beta_k \geq \beta_{\min}$$

and hence β_k is uniformly bounded away from zero.

First, for $k \in N$ such that $d_k^T W_k d_k < 0$ we have

$$\begin{aligned}
\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) &= \nabla l(x_k,\lambda_k)^T d_k \\
&= \nabla_x l(x_k,\lambda_k)^T s_k + h(x_k)^T \mu_k \\
&= \left(\nabla_x l(x_k,\lambda_k) + B_k u_k\right)^T y_k + h(x_k)^T \mu_k \\
&\quad +\nabla_x l(x_k,\lambda_k)^T u_k - u_k^T B_k y_k \\
&= \ell_k(y_k) + h(x_k)^T \mu_k + \nabla_x l(x_k,\lambda_k)^T u_k - u_k^T B_k y_k,
\end{aligned}$$

which implies that

$$(6.21)\qquad \begin{aligned}
\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) &\leq \mu_2 \max\left(\ell_k(\beta_k y_k^{sd}), q_{2,k}(\beta_k y_k^{sd})\right) + h(x_k)^T \mu_k \\
&\quad + \nabla_x l(x_k,\lambda_k)^T u_k - u_k^T B_k y_k.
\end{aligned}$$

Second, for $k \in N$ such that $d_k^T W_k d_k > 0$, we have

$$\begin{aligned}
\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) &= \nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2} d_k^T W_k d_k \\
&= \left(\nabla_x l(x_k,\lambda_k) + B_k u_k\right)^T y_k + \tfrac{1}{2} y_k^T B_k y_k + \tfrac{1}{2} u_k^T B_k u_k \\
&\quad +\nabla_x l(x_k,\lambda_k)^T u_k + \left(h(x_k) + \nabla h(x_k)^T u_k\right)^T \mu_k
\end{aligned}$$

or equivalently

$$\begin{aligned}
\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) &= q_{2,k}(y_k) + \nabla_x l(x_k,\lambda_k)^T u_k + \tfrac{1}{2} u_k^T B_k u_k \\
&\quad + \left(h(x_k) + \nabla h(x_k)^T u_k\right)^T \mu_k
\end{aligned}$$

which implies that

$$(6.22)\qquad \begin{aligned}
\nabla l(x_k,\lambda_k)^T d_k + \tfrac{1}{2}\max\left(0, d_k^T W_k d_k\right) \leq{}& \mu_2 \max\left(\ell_k(\beta_k y_k^{sd}), q_{2,k}(\beta_k y_k^{sd})\right) + \tfrac{1}{2} u_k^T B_k u_k \\
&+\nabla_x l(x_k,\lambda_k)^T u_k + \left(h(x_k) + \nabla h(x_k)^T u_k\right)^T \mu_k.
\end{aligned}$$

Therefore, we obtain from (6.12), (6.21), (6.22), Lemma 4.1, and (6.18) that

$$\lim_{k\in N\to\infty} \max\left(\ell_k(\beta_k y_k^{sd}), q_{2,k}(\beta_k y_k^{sd})\right) = 0. \tag{6.23}$$

First assume that there exists $J \subset N$ such that

$$\left(y_j^{sd}\right)^T B_j \left(y_j^{sd}\right) \le 0 \tag{6.24}$$

holds for all $j \in J$, which, together with (6.23), implies that

$$\lim_{j\in J\to\infty} \ell_j(\beta_j y_j^{sd}) = 0. \tag{6.25}$$

Since β_j is bounded away from zero, we obtain, using (6.14b), that $\nabla_x l(x_*, \lambda_*)^T y_*^{sd} = 0$ which contradicts (6.19). Consequently, for sufficiently large $k \in N$ we have

$$\left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right) > 0 \tag{6.26}$$

and hence

$$q_{2,k}(\beta_k y_k^{sd}) > \ell_k(\beta_k y_k^{sd}). \tag{6.27}$$

From (6.26) and the definition of β_k, we obtain

$$0 < \beta_k \left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right) \le -2\left(\nabla_x l(x_k, \lambda_k) + B_k u_k\right)^T y_k^{sd} \tag{6.28}$$

which implies, since β_k is bounded away from zero, that $\left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right)$ is bounded. Without loss of generality, we can assume that such a sequence converges to some θ_* and that β_k converges to β_*. Let us consider for $t \ge 0$ the function

$$\begin{aligned}\varphi_k(t) &= q_{2,k}(t\beta_k y_k^{sd}) \\ &= t\beta_k \left(\nabla_x l(x_k, \lambda_k) + B_k u_k\right)^T y_k^{sd} + \tfrac{1}{2}\beta_k^2 t^2 \left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right)\end{aligned}$$

which is strictly convex. Because of the choice of β_k, such a function has its minimum over $[0,1]$ at $t_k = 1$. Let $t \in [0,1]$. Then for all $k \in N$, we have for $t \in [0,1]$

$$t\beta_k \left(\nabla_x l(x_k, \lambda_k) + B_k u_k\right)^T y_k^{sd} + \tfrac{1}{2}\beta_k^2 t^2 \left(y_k^{sd}\right)^T B_k \left(y_k^{sd}\right) \ge q_{2,k}(\beta_k y_k^{sd}). \tag{6.29}$$

From (6.14b), (6.23), (6.27), and (6.28) we obtain

$$\lim_{k\in N\to+\infty} t\beta_k \left(\nabla_x l(x_k, \lambda_k) + B_k u_k\right)^T y_k^{sd} + \tfrac{1}{2}\beta_k^2 t^2 \theta_* \ge 0,$$

and since t is arbitrary in $[0,1]$ and β_k is bounded away from zero, this implies that $\nabla_x l(x_*, \lambda_*)^T y_*^{sd} = 0$ which contradicts (6.19). Consequently, $\nabla_x l(x_*, \lambda_*)^T y_*^{sd} = 0$ must hold which, together with Lemma 2.2 and (6.18), implies that x_* is a Karush-Kuhn-Tucker point of $(EQCP)$. ∎

As a consequence of Lemma 6.2, we establish in the following lemma that the penalty parameter r_k is uniformly bounded.

LEMMA 6.3. *Assume Hypotheses GCH.1-7. Then the penalty parameter r_k is uniformly bounded.*

Proof. We give a proof by contradiction. Assume that $\{r_k | k \in \mathbb{N}\}$ is unbounded. Then, since it is a nondecreasing sequence, there exists a subsequence $\{x_j | j \in J \subset \mathbb{N}\}$ satisfying

$$\nabla h(x_j) h(x_j) \ne 0 \tag{6.30}$$

for all $j \in J$ and

$$\widehat{r}_j = 2\frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2} d_j^T W_j d_j}{\|h(x_j)\|_2^2 - \|h(x_j) + \nabla h(x_j)^T u_j\|_2^2} + \rho, \tag{6.31}$$

where

$$0 < \nabla l(x_j, \lambda_j)^T d_j + \tfrac{1}{2} d_j^T W_j d_j. \tag{6.32}$$

We have

$$\begin{aligned}0 < \nabla l(x_j, \lambda_j)^T d_j + \tfrac{1}{2} d_j^T W_j d_j &= \nabla_x l(x_j, \lambda_j)^T u_j + \tfrac{1}{2} u_j^T B_j u_j + \tfrac{1}{2} y_j^T B_j y_j \\ &\quad + \left(\nabla_x l(x_j, \lambda_j) + B_j u_j\right)^T y_j \\ &\quad + \left(h(x_j) + \nabla h(x_j)^T u_j\right)^T \mu_j,\end{aligned}$$

and hence

$$(6.33)\quad \begin{aligned} 0 < \nabla l(x_j,\lambda_j)^T d_j + \tfrac{1}{2} d_j^T W_j d_j \quad &\le \nabla_x l(x_j,\lambda_j)^T u_j + \tfrac{1}{2} u_j^T B_j u_j \\ &+ \Big(h(x_j) + \nabla h(x_j)^T u_j \Big)^T \mu_j . \end{aligned}$$

From Hypotheses GCH1-7 and (6.34), we obtain that the numerator in (6.31) is bounded. Consequently, the denominator in (6.31) converges to zero, i.e.

$$(6.34)\quad \lim_{j \in J \to +\infty} \left(\|h(x_j)\|_2^2 - \left\| h(x_j) + \nabla h(x_j)^T u_j \right\|_2^2 \right) = 0$$

and hence α_j^{sd} is bounded away from zero, say

$$(6.34)\quad \alpha_j^{sd} \ge \alpha_{\min}^{sd}.$$

Also, we have

$$(6.6)\quad \left[\|h(x_j)\|_2^2 - \left\| h(x_j) + \alpha_j^{sd} \nabla h(x_j)^T u_j^{sd} \right\|_2^2 \right] \ge \alpha_j^{sd} \, \|\nabla h(x_j) h(x_j)\|_2^2$$

and by Hypothesis this implies that

$$(6.6)\quad \frac{1}{\left[\|h(x_j)\|_2^2 - \left\| h(x_j) + \alpha_j^{sd} \nabla h(x_j)^T u_j^{sd} \right\|_2^2 \right]} \le \frac{1}{\alpha_{\min}^{sd} \|\nabla h(x_k) h(x_k)\|_2^2}$$

$$(6.33)\quad 0 < \quad c_1 \alpha_{\min}^{sd} \frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2} d_j^T W_j d_j}{\|h(x_j)\|_2^2 - \|h(x_j) + \nabla h(x_j)^T u_j\|_2^2} \quad \le \quad \frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2} d_j^T W_j d_j}{\|\nabla h(x_k) h(x_k)\|_2^2}$$

$$(6.33)\quad \begin{aligned} 0 < \quad c_1 \alpha_{\min}^{sd} \frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2} d_j^T W_j d_j}{\|h(x_j)\|_2^2 - \|h(x_j) + \nabla h(x_j)^T u_j\|_2^2} \quad &\le \quad \frac{\nabla_x l(x_j,\lambda_j)^T u_j + \frac{1}{2} u_j^T B_j u_j}{\|\nabla h(x_k) h(x_k)\|_2^2} \\ &+ \quad \frac{\big(h(x_j) + \nabla h(x_j)^T u_j \big)^T \mu_j,}{\|\nabla h(x_k) h(x_k)\|_2^2} \end{aligned}$$

Therefore, we obtain from (6.34), and Lemma 6.1 that $\{\nabla h(x_j) h(x_j) | j \in J\}$ converges to zero which contradicts (6.31).Therefore $\{r_k | k \in \mathbb{N}\}$ is indeed uniformly bounded. □

In the following theorem, we establish that the steplength t_k is bounded away from zero at a nonstationary point.

THEOREM 6.1. *Let $\{x_k | k \in N\}$ converge to x_* where x_k and x_* are not Karush-Kuhn-Tucker points of $(EQCP)$. Also let $\{t_k | k \in N\}$ be the sequence of acceptable steplengths. Then there exists a positive scalar $t(x_*)$ such that*

$$(6.35)\quad t_* \ge t(x_*)$$

holds for any accumulation point t_* of $\{t_k | k \in N\}$.

Proof. We give a proof by contradiction. Assume that there exists a subsequence $\{t_j | j \in J\}$ converging to zero. Since $t_j = 1$ is not acceptable, let $\bar{t}_j$ be the last nonacceptable steplength. We have

$$c_4 t_j \le \bar{t}_j \le c_5 t_j$$

which implies that $\{\bar{t}_j | j \in J\}$ converges to zero. Because of Theorem 5.1, the nonacceptable steplength $\bar{t}_j$ is positive. We have

$$\Phi_j(\bar{t}_j d_j) - \Phi_j(0) > c_1 \max \Big(pred_j(\bar{t}_j d_j), \bar{t}_j \nabla \Psi_j(0)^T d_j \Big).$$

This implies that

$$\Phi_j(\bar{t}_j d_j) - \Phi_j(0) > c_1 \left[\Psi_j(\bar{t}_j d_j) - \Psi_j(0) \right]$$

and hence

$$(6.36)\quad \begin{aligned} \frac{l(x_j + \bar{t}_j s_j, \lambda_j + \bar{t}_j \mu_j) - l(x_j,\lambda_j)}{\bar{t}_j} + r_j \frac{\left\| h(x_j + \bar{t}_j s_j) \right\|_2^2 - \|h(x_j)\|_2^2}{\bar{t}_j} &> c_1 \frac{\frac{1}{2} \bar{t}_j^2 s_j^T B_j s_j + \bar{t}_j \nabla_x l(x_j,\lambda_j)^T s_j}{\bar{t}_j} \\ + c_1 r_j \frac{\left\| h(x_j) + \nabla h(x_j)^T \bar{t}_j s_j \right\|_2^2 - \|h(x_j)\|_2^2}{\bar{t}_j} &+ c_1 \frac{\big(h(x_j) + \nabla h(x_j)^T \bar{t}_j s_j \big)^T \bar{t}_j \mu_j}{\bar{t}_j}. \end{aligned}$$

We have

$$(6.37)\quad \frac{l(x_j + \bar{t}_j s_j, \lambda_j + \bar{t}_j \mu_j) - l(x_j,\lambda_j)}{\bar{t}_j} = \frac{f(x_j + \bar{t}_j s_j) - f(x_j)}{\bar{t}_j} + \frac{\big(h(x_j + \bar{t}_j s_j) - h(x_j) \big)^T \lambda_j}{\bar{t}_j} + \frac{h(x_j + \bar{t}_j s_j)^T \bar{t}_j \mu_j}{\bar{t}_j}.$$

On the other hand, there exists $\bar{z}_j \in (x_j, x_j + \bar{t}_j s_j)$ such that

$$\begin{aligned}\frac{f(x_j+\bar{t}_j s_j)-f(x_j)}{\bar{t}_j} &= \nabla f(\bar{z}_j)^T s_j \\ &= \nabla f(x_j)^T s_j + \left[\nabla f(\bar{z}_j) - \nabla f(x_j)\right]^T s_j\end{aligned}$$

where, since $\{x_j | j \in J\}$ converges to x_* and $\{\bar{t}_j | j \in J \subset N\}$ converges to zero,

$$\lim_{j \to +\infty} \left[\nabla f(\bar{z}_j) - \nabla f(x_j)\right] = 0.$$

Therefore, we have

$$\frac{f(x_j+\bar{t}_j s_j)-f(x_j)}{\bar{t}_j} = \nabla f(x_j)^T s_j + o_1(\tfrac{1}{j}) s_j \tag{6.38a}$$

where

$$\lim_{j \to +\infty} o_1(\tfrac{1}{j}) = 0. \tag{6.38b}$$

Similarly we have for $i = 1 \ldots m$

$$\frac{h_i(x_j+\bar{t}_j s_j)-h_i(x_j)}{\bar{t}_j} = \nabla h_i(x_j)^T s_j + \hat{o}_i(\tfrac{1}{j}) \tag{6.39a}$$

where

$$\lim_{j \to +\infty} \hat{o}_i(\tfrac{1}{j}) = 0, \tag{6.39b}$$

and hence, since λ_j is bounded,

$$\frac{\left(h(x_j+\bar{t}_j s_j)-h(x_j)\right)^T \lambda_j}{\bar{t}_j} = \left(\nabla h(x_j)\lambda_j\right)^T s_j + o_2(\tfrac{1}{j}) \tag{6.39c}$$

where

$$\lim_{j \to +\infty} o_2(\tfrac{1}{j}) = 0. \tag{6.39d}$$

From (6.39a,b), we obtain

$$\frac{\left\|h(x_j+\bar{t}_j s_j)\right\|_2^2-\|h(x_j)\|_2^2}{\bar{t}_j} \leq \frac{\left\|h(x_j)+\nabla h(x_j)^T\ \bar{t}_j s_j\right\|_2^2-\|h(x_j)\|_2^2}{\bar{t}_j} + o_3(\tfrac{1}{j}) \tag{6.40a}$$

where

$$\lim_{j \in J \to +\infty} o_3(\tfrac{1}{j}) = 0, \tag{6.40b}$$

which, together with the convexity of $c_j(t)$, implies that

$$\frac{\left\|h(x_j+\bar{t}_j s_j)\right\|_2^2-\|h(x_j)\|_2^2}{\bar{t}_j} \leq \left[\left\|h(x_j) + \nabla h(x_j)^T\ s_j\right\|_2^2 - \|h(x_j)\|_2^2\right] + o_3(\tfrac{1}{j}). \tag{6.40c}$$

Also, we have, because of the definition of W_j,

$$\left(\nabla h(x_j)^T \bar{t}_j s_j\right)^T \bar{t}_j \mu_j = \tfrac{1}{2}\bar{t}_j^2 d_j^T W_j d_j - \tfrac{1}{2}\bar{t}_j^2 s_j^T B_j s_j$$

which we rewrite as

$$\left(\nabla h(x_j)^T \bar{t}_j s_j\right)^T \bar{t}_j \mu_j = \tfrac{1}{2}\bar{t}_j^2 d_j^T W_j d_j + \bar{t}_j o_4(\tfrac{1}{j}) \tag{6.41a}$$

where

$$\lim_{j \in J \to +\infty} o_4(\tfrac{1}{j}) = 0. \tag{6.41b}$$

Because $\{r_j\}$ is bounded, $0 < 1 - c_1$, and $\{\bar{t}_j | j \in J\}$ converges to zero, we obtain from (6.36), (6.37), (6.38a,b), (6.39a,b), (6.40a,b,c) and (6.41a,b),

$$\frac{\nabla l(x_j,\lambda_j)^T \bar{t}_j d_j + \left(\nabla h(x_j)\bar{t}_j \mu_j\right)^T \bar{t}_j s_j + \frac{1}{2}\bar{t}_j^2 s_j^T B_j s_j}{\bar{t}_j} > r_j \frac{\|h(x_j)\|_2^2 - \left\|\ h(x_j) + \bar{t}_j \nabla h(x_j)^T\ s_j\right\|_2^2}{\bar{t}_j} + o(\tfrac{1}{j}) \tag{6.42a}$$

or equivalently

$$\frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2} d_j^T W_j d_j}{\bar{t}_j} > r_j \frac{\|h(x_j)\|_2^2 - \left\|\ h(x_j) + \nabla h(x_j)^T\ s_j\right\|_2^2}{\bar{t}_j} + o(\tfrac{1}{j}) \tag{6.42b}$$

where

$$\lim_{j \in J \to +\infty} o(\tfrac{1}{j}) = 0, \tag{6.42c}$$

and hence

(6.43) $$\frac{\nabla l(x_j,\lambda_j)^T\bar{t}_j d_j+\frac{1}{2}\max\left(0,\bar{t}_j^2 d_j^T W_j d_j\right)}{\bar{t}_j} > r_j\frac{\|h(x_j)\|_2^2-\left\| h(x_j)+\nabla h(x_j)^T\ \bar{t}_j s_j\right\|_2^2}{\bar{t}_j}+o(\tfrac{1}{j}),$$

which implies that

(6.44) $$\liminf_{j\in J\to+\infty} \nabla l(x_j,\lambda_j)^T d_j+\tfrac{1}{2}\max\left(0,d_j^T W_j d_j\right)\geq 0.$$

From (6.43) we obtain

$$2\frac{\nabla l(x_j,\lambda_j)^T\bar{t}_j d_j+\frac{1}{2}\max\left(0,\bar{t}_j^2 d_j^T W_j d_j\right)}{\bar{t}_j}+r_j\frac{\left\| h(x_j)+\nabla h(x_j)^T\ \bar{t}_j s_j\right\|_2^2-\|h(x_j)\|_2^2}{\bar{t}_j} >$$
$$r_j\frac{\|h(x_j)\|_2^2-\left\| h(x_j)+\nabla h(x_j)^T\bar{t}_j\ s_j\right\|_2^2}{\bar{t}_j}+o(\tfrac{1}{j}),$$

which, together with the convexity of c_j, implies that

$$2\left(\nabla l(x_j,\lambda_j)^T d_j+\tfrac{1}{2}\max\left(0,d_j^T W_j d_j\right)\right)+r_j\left[\left\| h(x_j)+\nabla h(x_j)^T s_j\right\|_2^2-\|h(x_j)\|_2^2\right]$$
$$> \ r_j\frac{\|h(x_j)\|_2^2-\left\| h(x_j)+\nabla h(x_j)^T\bar{t}_j\ s_j\right\|_2^2}{\bar{t}_j}+o(\tfrac{1}{j}),$$

and hence

(6.45) $$\liminf_{j\in J\to+\infty}\ \liminf_{j\in J\to+\infty}\left(2\left(\nabla l(x_j,\lambda_j)^T d_j+\tfrac{1}{2}\max\left(0,d_j^T W_j d_j\right)\right)+ r_j\left[\left\| h(x_j)+\nabla h(x_j)^T s_j\right\|_2^2-\|h(x_j)\|_2^2\right]\right)\geq 0.$$

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

$$2\left(\nabla l(x_j,\lambda_j)^T d_j+\tfrac{1}{2}\max\left(0,d_j^T W_j d_j\right)\right)+(r_j+\rho)\left[\left\| h(x_j)+\nabla h(x_j)^T s_j\right\|_2^2-\|h(x_j)\|_2^2\right]\leq 0$$

which implies, together with (6.45) and the convexity of c_j, that

(6.46) $$\lim_{j\in J\to+\infty}\left[\|h(x_j)\|_2^2-\left\| h(x_j)+\nabla h(x_j)^T\ s_j\right\|_2^2\right]=0.$$

Finally, the limits (6.44) and (6.46), together with Lemma 6.2, imply that x_* is a Karush-Kuhn-Tucker point of $(EQCP)$, which contradicts our hypothesis. Consequently the theorem does hold. □

Theorem 6.2. *The iteration sequence $\{z_k=(x_k,\lambda_k)\}$ generated by the $HYBEQ$ Algorithm is not bounded away from stationarity.*

Proof. Assume that the iteration sequence $\{z_k=(x_k,\lambda_k)\}$ is bounded away from stationarity. First, we obtain from Lemma 6.3 that the penalty parameter r_k is uniformly bounded, and since we are assuming that the iteration sequence is infinite, such a parameter is constant for sufficiently large k, say $r_k=r_*$ for $k\geq k_*$, and the merit function $\Phi_k(.)$ is constant with respect to this parameter. Therefore, we denote $\Phi(x_k+s)$ instead of $\Phi_k(s)$.

Let (x_*,λ_*) be an arbitrary accumulation point of $\{z_k=(x_k,\lambda_k)\}$. Because we are assuming that the iteration sequence is bounded away from stationarity, x_* is not a Karush–Kuhn-Tucker point of $(EQCP)$. Let $\{(x_j,\lambda_j)\,|\,j\in J\}$ be a subsequence converging to (x_*,λ_*). We claim that there exist a positive integer $j_*\geq k_*$, and a positive scalar ω_* such that

(6.47) $$\Phi(z_{j+1})\leq\Phi(z_*)-\omega_*$$

holds for sufficiently large j,say $j\geq j_*$.

Recall that $d_j=(s_j,\mu_j)$. The sequence $\{d_j\mid j\in J\}$ is bounded, without loss of generality, we can assume that it converges to d_*. We have

(6.48) $$\lim_{j\in J\to\infty}\nabla\Phi(z_j)^T d_j=\nabla\Phi^T(z_*)^T d_*.$$

Because x_* is not a stationary point of $(EQCP)$, we obtain from Proposition 5.1, and Theorem 6.1 that there exists $\widehat{\omega} > 0$ such that

$$\limsup_{j \in J} \max\left(pred_j(t_j d_j); t_j \nabla\Phi(z_j)^T d_j\right) \leq -\widehat{\omega} < 0 \tag{6.49}$$

and hence

$$\max\left(pred_j(t_j d_j); t_j \nabla\Phi(z_j)^T d_j\right) \leq -\tfrac{1}{2}\widehat{\omega} \tag{6.50}$$

for sufficiently large $j \in J$. Since $\{\Phi(z_j) | j \in J\}$ converges to $\Phi(z_*)$, for $\epsilon = \frac{c_1}{4}\widehat{\omega}$, there exists a positive integer, say j_*, such that

$$\Phi(z_*) < \Phi(z_j) \leq \Phi(x_*) + \epsilon \tag{6.51}$$

holds for $j \geq j_*$. On the other hand, we have

$$\Phi(z_{j+1}) \leq \Phi(z_j) + c_1 \max\left(pred_j(t_j d_j); t_j \nabla\Phi(z_j)^T d_j\right) \tag{6.52}$$

which, together with (6.49) and (6.50), implies that

$$\Phi(z_{j+1}) \leq \Phi(z_*) + \tfrac{1}{2}\left(\tfrac{c_1}{2} - c_1\right)\widehat{\omega}. \tag{6.53}$$

Let us set $\omega_* = \frac{c_1}{4}\widehat{\omega}$ and rewrite (6.53) as

$$\Phi(z_{j+1}) \leq \Phi(z_*) - \omega_*. \tag{6.54}$$

Finally, since the entire sequence $\{\Phi(z_k) | k \geq j_*\}$ is decreasing, we obtain from (6.54) that (6.47) holds.

But, since the entire sequence $\{\Phi(z_k) | k \geq j_*\}$ is decreasing, we have

$$\Phi(z_*) < \Phi(z_k) \qquad \forall\ k \geq j_*.$$

This contradicts (6.54). Therefore the iteration sequence is indeed not bounded away from stationarity. □

7. Local convergence. In this section, we show, under the standard assumptions of Newton method, that, for sufficiently large k, the algorithm reduces to the SQP method and hence it is q-quadratically convergent.

We make the following local convergence hypotheses:

LCH.1) the function f and h are twice continuously differentiable;

LCH.2) the hypotheses of global convergence hold;

LCH.3) for sufficiently large k, the normal component u_k is the least-squares solution of $(N - STEP_p)$ and y_k solves $(T - STEP_p)$;

LCH.4) the iteration sequence $\{(x_k, \lambda_k) | k \in \mathbb{N}\}$ converges to (x_*, λ_*);

LCH.5) $\nabla h(x_k)$ has full rank and the linear system

$$h(x_k) + \nabla h(x_k)^T u = 0$$

is consistent for sufficiently large k ;

LCH.6) the limit reduced Hessian approximation is positive definite, i.e. sequence $\{B_k | k \in \mathbb{N}\}$ converge to a matrix B_* that is positive definite on

$$H(x_*) = \left\{y \in \mathbb{R}^n \mid \nabla h(x_*)^T y = 0\right\}$$

LCH.7) the limit

$$\lim_{k \in \mathbb{N} \to +\infty} \frac{\left(B_k - \nabla_x^2 l(x_*, \lambda_*)\right)^T s_k}{\|s_k\|} = 0$$

(i.e. $B_k = \nabla_x^2 l(x_k, \lambda_k)$) holds.

REMARK 7.1. The hypotheses LCH.2-4 imply that x_* is a Karush-Kuhn-Tucker point of $(EQCP)$.

REMARK 7.2. The hypotheses LCH.3, LCH.5 and LCH.6 imply that y_* is the exact solution of $(T-STEP_p)$.

LEMMA 7.1. *Assume the Hypotheses LCH.1-7. Then, then exits a positive constant $\widetilde{\rho}$ such that*

$$\frac{|pred_k(d_k)|}{\|d_k\|_2^2} \geq \widetilde{\rho}. \tag{7.1}$$

for all sufficiently large $k \in \mathbb{N}$.

Proof. First, observe that $\{u_k | k \in \mathbb{N}\}$ converges to zero. Therefore, since $\delta_k \geq \delta_{\min}$, we obtain from Hypotheses LCH.3 and LCH.5 that

$$h(x_k) + \nabla h(x_k)^T u_k = 0 \tag{7.2}$$

and hence

$$h(x_k) + \nabla h(x_k)^T s_k = 0. \tag{7.3}$$

From Proposition 5.1 and (7.3), we obtain

$$\frac{|pred_k(d_k)|}{\|d_k\|_2^2} \geq \frac{\left|\nabla l(x_k,\lambda_k)^T d_k + \frac{1}{2}\max\left(0, d_k^T W_k d_k\right)\right|}{\|d_k\|_2^2} + \rho \frac{\|h(x)\|_2^2}{\|d_k\|_2^2}. \tag{7.4}$$

Assume that there exists a subsequence $J \subset \mathbb{N}$ such that the right-hand side of (7.4) converges to zero. This implies that

$$\lim_{j\in J} \frac{h(x_j)}{\|d_j\|_2} = 0 \tag{7.5}$$

and hence

$$\lim_{j\in J} \frac{u_j}{\|d_j\|_2} = 0 \tag{7.6}$$

and

$$\lim_{j\in J} \frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2}\max\left(0, d_j^T W_j d_j\right)}{\|d_j\|_2^2} = 0. \tag{7.7}$$

On the other hand, Hypotheses LCH.6, (7.2) and (7.3) imply that

$$\frac{\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2}\max\left(0, d_j^T W_j d_j\right)}{\|d_j\|_2^2} = \frac{\nabla_x l(x_j,\lambda_j)^T s_j + \frac{1}{2} s_j^T B_j s_j}{\|d_j\|_2^2} \tag{7.8}$$

holds for sufficiently large j.

Let us assume that $\|.\|_p$ is the ℓ_∞-norm (the case where $\|.\|_p = \ell_2$ is handled in the same way). Since y_j is a solution of $(T-STEP_p)$, there exist $\mu_j \in \mathbb{R}^m$ such that

$$\begin{cases} a) & \nabla_x l(x_j,\lambda_j) + B_j u_j + B_j y_j + \nabla h(x_j)\mu_j = 0 \\ b) & \nabla h(x_j)^T y_j = 0 \end{cases} \tag{7.9}$$

We obtain from (7.9b) that

$$\left(\nabla_x l(x_j,\lambda_j) + B_j u_j\right)^T y_j + \tfrac{1}{2} y_j^T B_j y_j = -\tfrac{1}{2} y_j^T B_j y_j \tag{7.10}$$

and hence

$$\begin{aligned} \left(\nabla_x l(x_j,\lambda_j) + B_j u_j\right)^T y_j + \tfrac{1}{2} y_j^T B_j y_j + \nabla_x l(x_j,\lambda_j)^T u_j + \tfrac{1}{2} u_j^T B_j u_j = -\tfrac{1}{2} y_j^T B_j y_j \\ + \nabla_x l(x_j,\lambda_j)^T u_j + \tfrac{1}{2} u_j^T B_j u_j \end{aligned} \tag{7.11}$$

or equivalently

$$\frac{\nabla_x l(x_j,\lambda_j)^T s_j + \frac{1}{2} s_j^T B_j s_j}{\|d_j\|_2^2} = -\frac{1}{2}\frac{y_j^T B_j y_j}{\|d_j\|_2^2} + \frac{\nabla_x l(x_j,\lambda_j)^T u_j}{\|d_j\|_2^2} + \frac{1}{2}\frac{u_j^T}{\|d_j\|_2} B_j \frac{u_j}{\|d_j\|_2}. \tag{7.12}$$

Also, we obtain from (7.9) and (7.3) that

$$\frac{\nabla_x l(x_j,\lambda_j)^T u_k}{\|d_j\|_2^2} + \frac{u_j^T}{\|d_j\|_2} B_j \frac{s_j}{\|d_j\|_2} - \frac{\mu_j^T}{\|d_j\|_2}\frac{h(x_j)}{\|d_j\|_2} = 0 \tag{7.13}$$

where, because of (7.6),

$$\lim_{j\in J} \frac{u_j^T}{\|d_j\|_2} B_j \frac{s_j}{\|d_j\|_2} = 0. \tag{7.14}$$

On the other hand, we have

$$\left|\frac{\mu_j^T}{\|d_j\|_2}\frac{h(x_j)}{\|d_j\|_2}\right| \leq \frac{\|\mu_j\|_2}{\|d_j\|_2}\frac{\|h(x_j)\|_2}{\|d_j\|_2} \leq \frac{\|h(x_j)\|_2}{\|d_j\|_2} \tag{7.15}$$

and hence, because of (7.5),

$$\lim_{j\in J} \frac{\mu_j^T}{\|d_j\|_2}\frac{h(x_j)}{\|d_j\|_2} = 0. \tag{7.16}$$

Therefore, we obtain from (7.13), (7.14), and (7.16) that

$$\lim_{j\in J} \frac{\nabla_x l(x_j,\lambda_j)^T u_k}{\|d_j\|_2^2} = 0 \tag{7.17}$$

which, together with (7.12) and (7.6), implies that

$$\limsup_{j\in J} \frac{\nabla_x l(x_j,\lambda_j)^T s_j + \frac{1}{2} s_j^T B_j s_j}{\|d_j\|_2^2} \le -\liminf_{k\in J} \frac{1}{2} \frac{y_j^T B_j y_j}{\|d_j\|_2^2}. \tag{7.18}$$

and hence, because $\nabla h(x_j)^T y_j = 0$,

$$\liminf_{j\in J} \frac{\left|\nabla_x l(x_j,\lambda_j)^T s_j + \frac{1}{2} s_j^T B_j s_j\right|}{\|d_j\|_2^2} > \frac{1}{2}\zeta_* \liminf_{j\in J} \left(\frac{\|y_j\|_2}{\|d_j\|_2}\right)^2 \tag{7.19}$$

where ζ_* is a lower bound of the reduced Hessian approximation of B_j to the subspaces $H(x_j) = \left\{ y \in \mathbb{R}^n \mid \nabla h(x_j)^T y = 0 \right\}$.

Now, we claim that there exists a constant $\widetilde{\alpha}$ such that

$$\frac{\|y_j\|_2}{\|d_j\|_2} \ge \widetilde{\alpha} \tag{7.20}$$

holds for all sufficiently large $j \in J$.

Assume that this does not hold. Then, there exists a subsequence $J' \subset J$ such that

$$\lim_{j\in J'} \frac{y_j}{\|d_j\|_2} = 0\ . \tag{7.21}$$

This, together with (7.6), implies that

$$\lim_{j\in J'} \frac{s_j}{\|d_j\|_2} = 0. \tag{7.22}$$

We have

$$\frac{\nabla l(x_j,\lambda_j)^T d_j}{\|d_j\|_2^2} = \left(\frac{\nabla_x l(x_j,\lambda_j)}{\|d_j\|_2}\right)^T \frac{s_j}{\|d_j\|_2} + \frac{h(x_j)^T \mu_j}{\|d_j\|_2^2} \tag{7.23}$$

where, because of (7.9),

$$\frac{\|\nabla_x l(x_j,\lambda_j)\|_2}{\|d_j\|_2} \le (\nu_*^n + \nu_*^t) + \|\nabla h(x_j)\|_2\,. \tag{7.24}$$

Therefore, we obtain from (7.22), (7.23) and (7.16) that

$$\lim_{j\in J'} \frac{\nabla l(x_j,\lambda_j)^T d_j}{\|d_j\|_2^2} = 0 \tag{7.25}$$

which, together with (7.7), implies that

$$\lim_{j\in J'} \frac{d_j^T W_j d_j}{\|d_j\|_2^2} = 0. \tag{7.26}$$

This contradicts the hypothesis that W_j is positive definite for sufficiently large j. Therefore our claim in (7.20) holds.

Finally, from (7.20), (7.19) and (7.8) we obtain that

$$\liminf_{j\in J} \frac{\left|\nabla l(x_j,\lambda_j)^T d_j + \frac{1}{2}\max\left(0; d_j^T W_j d_j\right)\right|}{\|d_j\|_2^2} > \frac{1}{2}\zeta_* \widetilde{\alpha}^2 \tag{7.27}$$

which contradicts (7.7). Consequently the right-hand side of (7.4) is bounded away from zero, which ends the proof. □

In the following, recall that z_k and d_k denote $(x_k;\lambda_k)$ and $(s_k;\mu_k)$ respectively.

LEMMA 7.2. *Assume the Hypotheses LCH.1-7. Then, the entire step d_k is acceptable for sufficiently large k. Moreover, the trust-region radius is increased.*

Proof. Since l is twice continuously differentiable, we have

$$\begin{aligned} l(z_k + d_k) &= l(z_k) + \nabla l(z_k)^T d_k + \tfrac{1}{2} d_k^T \nabla^2 l(z_k + \widetilde{t}_k d_k) d_k \\ &= l(z_k) + \nabla l(z_k)^T d_k + \tfrac{1}{2} d_k^T \nabla^2 l(z_k) d_k + \\ &\quad \tfrac{1}{2} d_k^T \left(\nabla^2 l(z_k + \widetilde{t}_k d_k) - \nabla_x^2 l(z_k)\right) d_k. \end{aligned} \tag{7.28}$$

where $\widetilde{t}_k \in (0,1)$, and

$$\limsup_{k\in\mathbb{N}\to+\infty} \left(\nabla^2 l(z_k + \widetilde{t}_k d_k) - \nabla^2 l(z_k)\right) = 0 \tag{7.29}$$

since $\{d_k | k \in \mathbb{N}\}$ converges to zero. Therefore, by using (7.29), we rewrite (7.28) as

$$l(z_k+d_k) = l(z_k)+\nabla l(z_k)^T d_k+\tfrac{1}{2}d_k^T\nabla^2 l(z_k)d_k+o_1\left(\|d_k\|_p^2\right). \tag{7.30}$$

Also, we have

$$\begin{aligned} h_i(x_k+s_k) &= h_i(x_k)+\nabla h_i(x_k+t_k^i s_k)^T s_k \\ &= h_i(x_k)+\nabla h_i(x_k)^T s_k+\left(\nabla h_i(x_k+t_k^i s_k)-\nabla h_i(x_k)\right)^T s_k \end{aligned} \tag{7.31}$$

where, since h_i is twice continuously differentiable and $\{s_k|k\in\mathbb{N}\}$ converges to zero,

$$\limsup_{k\in\mathbb{N}\to+\infty}\ \left(\nabla h_i(x_k+t_k^i s_k)-\nabla h_i(x_k)\right)=0. \tag{7.32}$$

Therefore, we rewrite (7.31) as

$$h_i(x_k+s_k) = h_i(x_k)+\nabla h_i(x_k)^T s_k+\widehat{o}_i\left(\|s_k\|_p\right) \tag{7.33}$$

and hence

$$h(x_k+s_k)=h(x_k)+\nabla h(x_k)^T s_k+o_2\left(\|s_k\|_p\right). \tag{7.34}$$

From (7.30) and (7.34), we obtain

$$\begin{aligned} ared_k(d_k) &= l(z_k+d_k)-l(z_k)+r_k\left[\|h(x_k+s_k)\|_2^2-\|h(x_k)\|_2^2\right] \\ &= \nabla l(z_k)^T d_k+\tfrac{1}{2}d_k^T\nabla^2 l(z_k)d_k+o_1\left(\|d_k\|_p\right)+ \\ &\quad r_k\left[\left\|h(x_k)+\nabla h(x_k)^T s_k+o_2\left(\|s_k\|_p\right)\right\|_p^2-\|h(x_k)\|_p^2\right] \end{aligned} \tag{7.35}$$

and

$$pred_k(d_k) = \nabla l(z_k)^T d_k+\tfrac{1}{2}d_k^T W_k d_k+r_k\left[\left\|h(x_k)+\nabla h(x_k)^T s_k\right\|_p^2-\|h(x_k)\|_p^2\right] \tag{7.36}$$

which implies that

$$\begin{aligned} &ared_k(d_k)-pred_k(d_k)=\tfrac{1}{2}d_k^T\left[\nabla^2 l(z_k)-W_k\right]d_k+o_1\left(\|d_k\|_p^2\right) \\ &\quad r_k\left\|h(x_k)+\nabla h(x_k)^T s_k+o_2\left(\|s_k\|_p\right)\right\|_p^2-\left\|h(x_k)+\nabla h(x_k)^T s_k\right\|_p^2. \end{aligned} \tag{7.37}$$

From (7.37) and (7.3) we obtain

$$ared_k(d_k)-pred_k(d_k)=\tfrac{1}{2}d_k^T\left[\nabla^2 l(z_k)-W_k\right]d_k+o_1\left(\|d_k\|_p^2\right)+r_k o_2\left(\|s_k\|_2^2\right). \tag{7.38}$$

which, together with Hypothesis LCH.6, implies that

$$\lim_{k\in\mathbb{N}\to+\infty}\frac{ared_k(d_k)-pred_k(d_k)}{\|d_k\|_2^2}=0 \tag{7.39}$$

On the other hand,we have

$$\frac{ared_k(d_k)-pred_k(d_k)}{pred_k(d_k)}=\frac{ared_k(d_k)-pred_k(d_k)}{\|d_k\|_2^2}\frac{\|d_k\|_2^2}{pred_k(d_k)}. \tag{7.40}$$

From Lemma 7.1 and (7.40) we obtain that

$$\lim_{k\in\mathbb{N}}\frac{ared_k(s_k)}{pred_k(s_k)}=1 \tag{7.41}$$

and hence

$$ared_k(d_k)\geq c_2\quad pred_k(d_k) \tag{7.42}$$

holds for suficiently large k. This implies, first, that the entire step d_k is acceptable and second that the trust-region radius δ_k is increased. The second point shows that a very small $\delta_{\min}$ does not preclude the rate of convergence. □

THEOREM 7.1. *Assume the Hypotheses LCH.1-7. Then, the rate of convergence is q-quadratic.*

Proof. From Lemma 7.2, we obtain that, for sufficiently large $k\in\mathbb{N}$, the sequence $\{(x_k,\lambda_k)\}$ is generated by the iteration $(x_{k+1},\lambda_{k+1})=(x_k+s_k,\lambda_k+\mu_k)$ where s_k is the solution of

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

i.e. the SQP step and μ_k the associated Lagrange multiplier. Therefore the algorithm reduces to the SQP method and hence the rate of convergence is q-quadratic. □

8. Numerical Tests. In this section we present some primary numerical tests. To solve these subproblems we used Fortran subroutine provided by Pola and Casas [37] of the Cantabria University in Spain . We used the ℓ_∞-norm in the trust-region constraints. We also used $\tau_1 = \tau_2 = 1$, because we feel that the usual choice of $\tau_2 = 1 - \tau_1$ where τ_1 is very small would give more weight to the normal component with respect to tangential component, which would define an algorithm that is directed more toward feasibility that stationarity. The number of internal increases is maxiis, in which case the trust-region is multiplied by inc_del. Moreover, in decreasing the steplength, we multiply the current step by $\beta_1 = 0.25$ if the predicted reduction is negative and by $\beta_2 = 0.75$ otherwise. The other parameters are as follows: $\rho = 1, c_1 = 10^{-4}, c_2 = 0.9, c_3 = 0.5, \delta_0 = 1, \delta_{\min} = 10^{-9}, \delta_{\max} = 10^5$. The algorithm stops if

1) $\|\nabla l(x_k, \lambda_k)\|_2 \le \varepsilon$, or
2) $\|s_k\|_2 \le \varepsilon$.

We run the hybrid algorithm with the associated pure trust-region version on a set of small problems from Hook and Schitowski [23]. The initial approximation of the solution is the same as in Hook and Schitowski [23]. The results are reported in the following Table 1. where:

Hybeq	:	Hybrid Algorithm,
TR	:	Trust-region version,
n0	:	Hook and Schitowski[23] problem number,
n	:	Problem dimesion,
m	:	Constraints number,
it	:	Iteration number,
nf	:	Function evaluation,
back	:	Steplentgh decreases number,
decr	:	Trust-region radius decreases number,
Hy_Tol	:	Tolerance ε for the Hybrid Algorithm,
TR-Tol	:	Tolerance ε for the Trust-region version.
e_n	:	The n-vector of ones.

HS-Problem	Hybeq		TR	
(n0,n,m)	(it,nf,back)	Hy-Tol	(it,bf,decr)	TR-Tol
(6,2,1)	(10,11,1)	1.e-11	(12,13,1)	1.e-11
(7,2,1)	(9,10,0)	1.e-11	(9,10,0)	1.e-11
(8,2,1)	(8,10,0)	1.e-11	(8,10,0)	1.e-11
(26,3,1)	(22,40,15))	1.e-11	(33,57,24)	1.e-11
(27,3,1)	(26,46,7)	1.e-11	(25,26,1)	1.e-11
(40,4,3)	(16,16,0)	1.e-11	(16,16,0)	1.e-11
(46,5,2)	(27,30,3)	1.e-11	(25,26,1)	1.e-11
(47,5,3)	(27,30,2))	1.e-11	(139,167,27)	1.e-11
(61,3,2)	(8,11,1)	1.e-11	(8,11,1)	1.e-11
(77,5,2)	(18,23,2)	1.e-11	(20,35,14)	1.e-09
(78,5,3)	(9,10,0)	1.e-11	(9,10,0)	1.e-11
(79,5,3)	(9,9,0)	1.e-11	(8,10,0)	1.e-11
(216,2,1)	(12,15,2)	1.e-11	(11,15,2)	1.e-11
(219,4,2)	(26,38,5)	1.e-11	(19,32,6)	1.e-11
(254,3,2)	(8,10,1)	1.e-11	(8,13,1)	1.e-11
(316,2,1)	(8,11,0)	1.e-11	(6,7,1)	1.e-11
(317,2,1)	(8,10,0)	1.e-11	(7,8,0)	1.e-11
(318,2,1)	(9,14,0)	1.e-11	(7,9,1)	1.e-11
(319,2,1)	(9,11,0)	1.e-11	(9,11,1)	1.e-11
(320,2,1)	(10,12,0)	1.e-11	(10,12,0)	1.e-11
(321,2,1)	(11,12,0)	1.e-11	(11,12,0)	1.e-11
(322,2,1)	(13,2,3)	1.e-11	(17,38,13)	1.e-11
(335,3,2)	(22,35,5)	1.e-11	(24,36,6)	1.e-11
(336,3,2)	(11,19,2)	1.e-11	(13,25,6)	1.e-11
(338,3,2)	(7,7,0)	1.e-11	(7,9,0)	1.e-08
(344,3,1)	(18,18,0)	1.e-11	(18,18,0)	1.e-11
(373,9,6)	(10,16,3)	1.e-11	(12,20,0)	1.e-11

Table 1

Also, we run a version of the Hybrid algorithm from a remote initial point IP where the normal component

$$u_k \equiv \begin{cases} u_k^{sd} & \text{if } \|h(x)\|_2 \geq \text{nor-tol} \\ u_k^{exact} & \text{otherwise} \end{cases}$$

and

$$y_k \equiv \begin{cases} y_k^{sd} & \text{if } \|\nabla l(x_k, \lambda_k)\|_2 \geq \text{tan-tol} \\ y_k^{exact} & \text{otherwise} \end{cases}$$

where u_k^{exact} and y_k^{exact} are the exact solutions of the $(N-STEP_\infty)$ and $(\overline{T-STEP}_\infty)$ Subproblem given by the routine of Pola and Casas[37]. The results are reported in Table 2.

(n0,n,m)	(it,nf,back)	IP	(nstp,tstp)	nor-tol	tan-tol	Tol
(6,2,1)	(13,23,4)	1e+03 x0	(10,10)	1.e+03	1.e+01	1.e-11
(7,2,1)	(28,46,8)	100 x0	(22,13)	1.e+01	1.e+01	1.e-11
(8,2,1)	(19,19,0)	1 e+04 x0	(13,-)	-	1.e+02	1.e-11
(26,3,1)	(25,45,17)	10 x0	(15,10)	1.e-01	1.e+02	1.e-11
(27,3,1)	(24,46,7)	x0	(0,0)	1.e+01	1.e+02	1.e-11
(40,4,3)	(21,28,0)	4 x0	(0,7)	1.e+02	1.e-01	1.e-11
(46,5,2)	(36,46,0)	4.5 x0	(3,3)	1.e+03	1.e+04	1.e-11
(47,5,3)	(90,104,10)	3 x0	(0,0)	1.e+03	1.e+04	1.e-11
(61,3,2)	(21,32,0)	-1.500 e_n	(10,11)	1.e+02	1.e+02	1.e-11
(77,5,2)	(27,36,1)	10 x0	(9,10)	1.e+03	1.e+02	1.e-11
(78,5,3)	(15,23,3)	6 x0	(1,4)	1.e+03	1.e+03	1.e-11
(79,5,3)	(16,18,0)	6 x0	(2,7)	1.e+03	1.e+03	1.e-11
(216,2,1)	(13,15,0)	-21 x0	(3,9)	1.e+03	1.e+02	1.e-08
(219,4,2)	(42,81,6)	80 x0	(22,22)	1.e+03	1.e+03	1.e-11
(254,3,2)	(32,41,9)	60 x0	(23,23)	1.e+01	1.e+01	1.e-11
(316,2,1)	(24,25,1)	1 e+04 e_n	(16,18)	1.e+01	1.e+01	1.e-11
(317,2,1)	(17,25,1)	1 e+03 e_n	(12,14)	1.	1.	1.e-11
(318,2,1)	(15,21,0)	800 e_n	(8,11)	1.e+01	1.e+01	1.e-11
(319,2,1)	(18,25,2)	800 e_n	(11,13)	1.e+01	1.e+01	1.e-11
(320,2,1)	(18,23,0)	-1 e+04 e_n	(11,14)	1.e+01	1.e+01	1.e-11
(321,2,1)	(22,37,2)	800 e_n	(10,18)	1.e+01	1.e+01	1.e-11
(322,2,1)	(23,32,2)	-800 e_n	(12,19)	1.e+01	1.e+01	1.e-11
(335,3,2)	(31,54,14)	40 x0	(10,21)	1.e+03	1.e+01	1.e-11
(336,3,2)	(26,30,1)	1e+03 e_n	(16,22)	1.e+01	1.e-01	1.e-11
(338,3,2)	(23,27,1)	1 e+04 x0	(16,16)	1.e+01	1.e+01	1.e-11
(344,3,1)	(37,41,1)	125 x0	(18,21)	1.e-01	1.e-01	1.e-08
(373,9,6)	(40,59,11)	x0	(1,32)	1.e+03	1.e-02	1.e-11

Table 2

9. Concluding Remarks. In this paper, we have presented a hybrid algorithm to solve the nonlinear minimization problem

$$(EQCP) \equiv \begin{cases} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \end{cases}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m, m < n$, are continuously differentiable.

Starting from an arbitrary initial approximation to a solution, say x_0, and an associated Lagrange multiplier approximation λ_0, the algorithm generates a sequence $z_k = (x_k, \lambda_k)$ by the iteration scheme $x_{k+1} = x_k + t_k s_k$ and $\lambda_{k+1} = \lambda_k + t_k \mu_k$. It uses a trust-region approach to find a search direction of the form $d_k = (s_k, \mu_k)$, where s_k is the trial step and μ_k an associated Lagrange multiplier, that is a descent direction at (x_k, λ_k) of the merit function defined to be the augmented Lagrangian

(9.1) $$\Phi_k(d) = l(x_k + s, \lambda_k + \mu) + r_k \, \|h(x_k + s)\|_2^2,$$

and then uses linesearch techniques to obtain an acceptable steplength t_k is such a direction. For the trust-region part, we used the approach of Byrd [5] and Omojokun [38]. Given a trust-region $\delta_k, \tau_1 \in (0, 1], \tau_2 \in (0, 1]$, and B_k, an approximation of the

Hessian Lagrangian, the trial step s_k is determined as $u_k + y_k$ where u_k, the normal component, is an approximate solution of

$$(9.2) \qquad (N - STEP_p) \equiv \begin{cases} \text{minimize} & q_{1,k}(u) = \frac{1}{2}\left\|h(x_k) + \nabla h(x_k)^T u\right\|_2^2 \\ \text{subject to} & \|u\|_p \leq \tau_1 \delta_k, \end{cases}$$

and y_k, the tangential component, is an approximate solution of

$$(9.3) \qquad (\overline{T - STEP_p}) \equiv \begin{cases} \text{minimize} & q_{2,k}(y) = (\nabla_x l(x_k, \lambda_k) + B_k u_k)^T y + \frac{1}{2} y^T B_k y \\ \text{subject to} & \nabla h(x_k)^T y = 0 \\ & \|y\|_p \leq \tau_2 \delta_k, \end{cases}$$

where $\|.\|_p$ is in practice either the ℓ_2 or the $\ell\infty$ norm.

The normal component of the trial step direction u_k satisfies the well-known fraction of Cauchy-decrease. The tangential component satisfies a test that takes into account the curvature of the local model.

Moreover, such a curvature is taken into account in the definition of penalty parameter update implying, together with an heuristic dealing with points satisfying

$$(9.6) \qquad h(x_k) \neq 0 \quad \text{and} \quad \nabla h(x_k) h(x_k) = 0,$$

that direction $d_k = (s_k, \mu_k)$ is a descent of the merit function which enabled us to propose the present hybrid approach.

We establish, under rather weak hypotheses, that the $HYBEQ$ Algorithm is globally convergent.

Moreover, under the standard hypotheses of the SQP method, we prove that for sufficiently large $k \in \mathbb{N}$, the $HYBEQ$ Algorithm reduces to the SQP method and therefore its rate of convergence is q-quadratic.

REFERENCES

[1] N. ALEXANDROV, *Multi-Level algorithms for nonlinear equations and constrained optimization*, Ph.D. thesis, Tech. Rep. TR93-20, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1993.

[2] N. ALEXANDROV AND J.E. DENNIS, JR., *A class of trust-region multilevel algorithms for nonlinear constrained optimization: global convergence analysis*, Tech. Rep. TR99786-S, Center for Research on Parallel Computation, Rice University, Houston, TX, 1999.

[3] L.T. BIEGLER, J. NOCEDAL, AND C. SCHMID, *A reduced Hessian method for large-scale constrained optimization*, SIAM J. Optim., 5 (1995), pp. 314-347.

[4] G. A. BLISS, *Normality and abnormality in the calculus of variations*, Trans. Amer. Math. Soc., 43 (1938), pp. 365-376.

[5] R.H. BYRD, *Robust trust-region methods for nonlinearly constrained optimization*, Presented at the 1987 SIAM Conference on Optimization, Houston, TX, 1987.

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

[7] E. CASAS AND C. POLA, *PLCBAS user's guide, Version 1.1*, Departamento e Mathematicas, Estadistica y Computacio, Universidad de Cantabria, Spain.

[8] E. CASAS AND C. POLA, *An algorithm for indefinite quadratic programming based on a partial Cholesky decomplosition*, Recherche Operatinnelle, vol. 27, n0. 4 (1993), pp. 401-426.

[9] M.R. CELIS, J.E. DENNIS, AND R.A. TAPIA, *A trust-region strategy for nonlinear equality constrained optimization*, in Numerical Optimization 1984, SIAM, Philadelphia, PA, 1985.

[10] F.H. CLARKE, *Optimization and nonsmooth analysis*, Canadian Mathematical Society Series on Monographs and Advanced Texts, John Wiley and Sons, 1983.

[11] J.E. DENNIS, M. EL-ALEM, AND M.C. MACIEL, *A global convergence theory for general trust-region-based algorithms for equality constrained optimization*, SIAM J. Optim., 7 (1997), pp. 177-207.

[12] J.E. DENNIS, S. B. Li, AND R.T. TAPIA, *A unified approach to global convergence of trust-region methods for nonsmooth optimization*, Math. Prog. 68 (1995), n0. 3, Ser. A, pp. 319-346.

[13] J.E. DENNIS, AND R.B. SCHNABEL, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, ed. Soc. Indust. App. Math., Philadelphia, 1996.

[14] J.E. DENNIS AND L.N. VICENTE, *On the convergence theory of trust-region-based algorithms for equality-constrained optimization*, SIAM J. Optim., 7 (1997), pp. 927-950.

[15] I.S. DUFF, J. NOCEDAL, AND J.K. REID, *The use of linear programming for the solution of sparse sets of nonlinear equations*, SIAM J. Sci. Stat. Comp., 2 (1987), pp. 99-108.

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

[17] 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.

[18] M. EL-ALEM, *Robust trust-region algorithm with non monotonic penalty parameter scheme for constrained optimization*, SIAM J. Opt., 5 (1995), pp. 348-378.

[19] M. EL-ALEM, *A global convergence theory for Dennis, El-Alem, and Maciel's class of trust-region algorithms for equality constrained optimization without assuming regularity*, SIAM J. Optim., 9 (1999), pp. 965-990.

[20] M. EL-ALEM AND R.A. TAPIA, *Numerical experience with polyhedral norm CDT trust-region algorithm*, J. Optim. Theory Appl., 85 (1985), pp. 575-592.

[21] M. EL HALLABI, *A global convergence theory for arbitrary norm trust-region algorithm for equality constrained optimization*, Tech. Rep. TR93-60, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1993.

[22] M. EL HALLABI, *A globally convergent trust-region algorithm for equality constrained optimization*, Les Annales Maghrébines de l'Ingénieur, Vol. 12, N0. Hors Série, Nov. 1998, pp. 87-92. Actes du Sixième Colloque Maghrébin sur les Modèles Mathématiques de l'Ingénieur, Tunis, 24-26 Nov. 1998.

[23] M. EL HALLABI, *A global convergence theory for sequential linear programming inexact hybrid algorithms*, Tech. Rep. CRPC-TR95543, Center for Research on Parallel Computation, Rice University, Houston, TX, 1993. ftp://softlib.rice.edu/pub/CRPC-TRs/reports/CRPC-TR94371.pdf

[24] M. EL HALLABI, *A hybrid algorithm for equality constrained optimization*, Presented at The Sixth SIAM Conference on Optimization, CP 14, May 10-12, 1999, Atlanta, Gerorgia.

[25] M. EL HALLABI AND R.A. TAPIA, *A global convergence theory for arbitrary norm trust-region methods for nonlinear equations*, Tech. Rep. TR93-41, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1993.

[26] S.C. EISENSTAT AND H.F. WALKER, *Globally convergent inexact newton methods*, SIAM J. Sci. Stat. Comp., 2 (1987), pp. 99-108.

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

[28] G.H. GOLUB, AND VAN LOAN, *Matrix Computation*, John Wiley and Sons, New York, 1987.

[29] M. HEIKENSCHLOSS AND L.N. VICENTE, *Analysis of inexact trust-region algorithms*, Tech. Rep. TR99-18, Department of Computational and Applied Mathematics, Rice University, Houston, TX, 1998.

[30] W. HOCK AND K.SCHITTKOWSKI, *The examples for Nonlinear Programming Codes*, Lectures Notes in Economics and Mathematical systems, 187, Spronger, Berlin, 1981.

[31] M. LALEE, J. NOCEDAL, AND T. PLANTEGA, *On the implementation of an algorithm for large-scale equality optimization*, SIAM J. Opt., 1998,8:682.

[32] K. LEVENBERG, *A method for the solution of certain problems in least squares*, Quart. Appli. Math., 2 (1944), pp. 164-168.

[33] 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, TX, 1992.

[34] D.W. MARQUARDT, *An algorithm for least-squares estimation of nonlinear parameters*, J. SIAM, 11 (1963), pp. 431-441.

[35] A. MIELE, H.Y. HUANG, AND J.C. HEIDEMAN, *Sequential gradient-restoration algorithm for the minimization of constrained functions: Ordinary and conjugate gradient versions*, J. Optim. Theorey Appl., 4 (1969), pp, 231-243.

[36] J.J. MORE AND D.C. SORENSEN, *Computing a trust-region step*, SIAM J. on Scientific and Statistical Computing, 4 (1983), pp, 553-572.

[37] J. NOCEDAL AND Y. YUAN, *Combining trust-region and line search techniques*, in Y.Yuan (ed.), Advances in Nonlinear Programming (Kluwer, 1998), pp. 153-175.

[38] E.O. OMOJOKUN, *Trust-region strategies for optimization with nonlinear equality and inequality constraints*, Ph. D. thesis, University of Colorado, Boulder, 1989.

[39] D. W. PETERSON, *A review of constraint qualifications in the finte dimensional spaces*, SIAM Review, 15 (1973), pp. 639-654.

[40] M.J.D. POWELL, *A new algorithm for unconstrained optimization*, in Nonlinear Programming, O. Mangasarian, R. Meyer, and S. Robinson, eds., Academic Press, New York, 1970, pp. 31-65.

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

[42] M.J.D. POWELL, *Some global convergence properties of a variable metric algorithgm for minimization without exact linesearch*, in Nonlinear Programming, SIAM-AMS Proceedings, Vol. 9, American Mathematical Society, Providence, R.I., 1976.

[43] M.J.D. POWELL, *On the convergence of trust-region algorithms for unconstrained minimization*, Math. Prog. (29) 1984, pp. 297-303.

[44] M.J.D. POWELL, *General algorithms for discrete nonlinear approximation calculations*, in Approximation Theory IV, C.K. Chui, L.L. Schumaker, and J.D. Ward, eds., New York, Academic Press, pp. 187-218, 1983.

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

[46] K. RITTER, *Global and superlinear convergence of a class of variable metric methods*, Technical Summary Report no.1945, Mathematics Research Center, University of Wisconsin-Madison, 1979.

[47] R.T. ROCKAFELLAR, *Generalized directional derivatives and subgradients of nonconvex functions*, Canadian J. Math., 32 (1980), pp. 157-180.

[48] G.S. SHULTZ, R.B. SCHNABBEL, AND R.H. BYRD, *A family of trust-region-based algorithms for unconstrained minimlization with strong global convergences properties*, SIAM J. Numer. Anal., 22 (1985), pp. 47-67.

[49] D.C. SORENSEN, *Minimization of a large scale quadratic function subject to spherical constraint*, SIAM J. Optim., (7) 1997, pp. 141-161.

[50] T. STEINHAUG, *The conjugate gradient method and trust-regions in large scale optimization*, SIAM J. Num. Anal., 20 (1983), pp. 626-637.

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

[52] P. WOLFE, *Convergence conditions for ascent methods*, SIAM Review, 11 (1969), pp. 226-235.

[53] J. ZHANG, N. KIM, AND L. LASDON, *An improved successive linear programming algorithm*, Management Sci. 31 (1985), pp. 1312-1331.

[54] J. ZHANG, *Computing a Celis-Dennis-Tapia trust-region step for equality constrained optimization*, J. Optim. Theory Appl., 67 (1990), pp. 369-393.