On the Characterization of Dennis, El-Alem, and Maciel's Class of Trust-Region Algorithms

Mahmoud El-Alem

CRPC-TR96656
Revised October 1997

Center for Research on Parallel Computation
Rice University
6100 South Main Street
CRPC - MS 41
Houston, TX 77005

Submitted in October 1996

ON THE CHARACTERIZATION OF DENNIS, EL-ALEM, AND MACIEL'S CLASS OF TRUST-REGION ALGORITHMS

MAHMOUD EL-ALEM

Department of Mathematics, Faculty of Science, Alexandria University, Alexandria, Egypt.

(Received date to be inserted)

In a recent paper, Dennis, El-Alem, and Maciel suggested a class of trust-region-based algorithms for solving the equality constrained optimization problem. They proved global convergence for the class.
In this paper, we characterize this class and present a short, straightforward, and self-contained global convergence theory. The results are stronger than Dennis, El-Alem, and Maciel's results.

KEY WORDS: Constrained optimization, equality constraints, global convergence, trust-region

1 INTRODUCTION

Over the last two decades, trust-region algorithms have proven to be very effective and robust techniques for solving the unconstrained optimization problems. Since mid eighties, many authors have considered extending the trust-region idea to the following equality constrained optimization problem

$$\text{(EQ)} \equiv \begin{cases} \textit{minimize} & f(x) \\ \textit{subject to} & C(x) = 0. \end{cases}$$

The functions $f : \Re^n \to \Re$ and $C : \Re^n \to \Re^m$ are at least twice continuously differentiable where $C(x) = [c_1(x), \ldots, c_m(x)]^T$ and $m < n$.

Most trust-region algorithms for solving problem (EQ) attempt to incorporate the trust-region idea within the successive quadratic programming (SQP) framework. The SQP method iteratively minimizes a quadratic model of the Lagrangian function

$$\ell(x, \lambda) = f(x) + \lambda^T C(x), \tag{1}$$

subject to a linear approximation of the constraints. At each iterate x_k, the SQP method obtains a step s_k^{QP} and an associated Lagrange multiplier step $\Delta\lambda_k^{QP}$ by solving the following quadratic programming subproblem

$$\begin{array}{ll} \textit{minimize} & \nabla_x \ell_k^T s + \frac{1}{2} s^T H_k s, \\ \textit{subject to} & C_k + \nabla C_k^T s = 0, \end{array}$$

where H_k is the Hessian of the Lagrangian function (1) or an approximation to it.

The reduced Hessian technique is one approach to incorporate a trust region into the above subproblem. This approach was suggested by Byrd and Omojokun (1987)[2] (see also Omojokun (1989)[12]). In this approach, the trial step s_k is decomposed into two components; the tangential component s_k^t and the normal component s_k^n. The step s_k^n is computed by solving the following trust-region subproblem

$$\begin{array}{ll} \textit{minimize} & \|C_k + \nabla C_k^T s^n\|_2^2, \\ \textit{subject to} & \|s^n\|_2 \leq \sigma \delta_k, \end{array}$$

for some $\sigma \in (0,1)$, where δ_k is the trust-region radius. The tangential component s_k^t is then obtained by solving another trust-region subproblem. Let Z_k be a matrix that forms an orthonormal basis for the null space of ∇C_k^T and let $s_k^t = Z_k \bar{s}_k^t$. The step $\bar{s}_k^t$ is computed by solving the following trust-region subproblem

$$\begin{array}{ll} \textit{minimize} & [Z_k^T(\nabla_x \ell_k + H_k s_k^n)]^T \bar{s}^t + \frac{1}{2}(\bar{s}^t)^T Z_k^T H_k Z_k \bar{s}^t, \\ \textit{subject to} & \|Z_k \bar{s}^t\|_2^2 \leq \delta_k^2 - \|s_k^n\|_2^2. \end{array}$$

The trial step s_k has the form $s_k = s_k^n + Z_k \bar{s}_k^t$.

This approach has been used by many authors. See, for example, Alexandrov (1993)[1], Dennis and Vicente (1997)[4], El-Alem (1995)[7], and (1996)[8], Lalee (1993) [9], Lalee, Nocedal, and Plantenga (1993)[10], Maciel (1992)[11], Plantenga (1995)[13], Vicente (1996)[16], and Zhang and Zhu (1990)[17].

Dennis, El-Alem, and Maciel (1997)[3] have considered a general class of trust-region based algorithms for solving problem (EQ). In their algorithms, the two components of the trial step are not necessarily orthogonal. We present this class of algorithms in the next section. In Section 3, we state the assumptions under which global convergence is proved. Section 4 is devoted to presenting the global convergence results. Finally, Section 5 contains concluding remarks.

The following notations are used throughout the rest of the paper. The sequence of points generated by the algorithm is denoted $\{x_k\}$. A subscripted function denotes the value of the function evaluated at a particular point. For example, $f_k \equiv f(x_k)$, $C_k \equiv C(x_k)$, $\ell_k \equiv \ell(x_k, \lambda_k)$, and so on. Finally, all the norms are ℓ_2-norms.

2 DENNIS, EL-ALEM, AND MACIEL'S CLASS OF ALGORITHMS

In a generic trust-region algorithm, a trial step s_k that satisfies some conditions is computed. The step is then tested. If the step is accepted the algorithm proceeds by setting $x_{k+1} = x_k + s_k$ and the radius of the trust-region is increased accordingly. If the step is rejected the trust-region radius is decreased and another trial step is computed in the smaller trust region.

2.1 Description of the class

In Dennis, El-Alem, and Maciel's class of algorithms, the trial step s_k is the sum of two components: the tangential component s_k^t and the quasi-normal component s_k^n. Dennis, El-Alem, and Maciel (1997)[3] do not require the two components of the trial step to be orthogonal. Instead, they impose a condition on the quasi-normal component s_k^n to prevent it from being too long when the violation of the constraints is small. They require that the following condition holds for the quasi-normal component of the trial step at every iteration k,

$$||s_k^n|| \leq K_1 ||C_k||,$$

where K_1 is a positive constant. This condition can be viewed as a relaxation to the orthogonality of the two components of the trial step.

The quasi-normal component of the trial step is taken to be any step inside the trust region $||s^n|| \leq \sigma_1 \delta_k$, for some $\sigma_1 \in (0,1)$, and gives at least a fraction of the Cauchy decrease obtained by the Cauchy step s_k^{ncp}. Thus, the step s^n satisfies,

$$||C_k||^2 - ||C_k + \nabla C_k^T s^n||^2 \geq \sigma_2 [||C_k||^2 - ||C_k + \nabla C_k^T s_k^{ncp}||^2],$$

for some $\sigma_2 \in (0,1]$.

Let the quadratic model of the Lagrangian function be $q_k(s) = \ell_k + \nabla_x \ell_k^T s + \frac{1}{2} s^T H_k s$ and let W_k be a matrix whose columns form a basis for the null space of ∇C_k^T. The tangential component s_k^t is then taken to be any step that satisfies the trust-region constraint $||s^t + s_k^n|| \leq \delta_k$ and gives at least a fraction of the decrease obtained by the Cauchy step s_k^{tcp} on the quadratic model of the Lagrangian $q_k(s)$ reduced to the null space of ∇C_k^T. *i. e.*, it satisfies

$$q_k(s_k^n) - q_k(s_k^n + s^t) \geq \sigma_3 [q_k(s_k^n) - q_k(s_k^n + s_k^{tcp})],$$

for some $\sigma_3 \in (0,1)$. See Dennis, El-Alem, and Maciel (1997)[3] for the definitions of the steps s_k^{ncp} and s_k^{tcp} and for more details about the double fraction-of-Cauchy decrease conditions imposed on the two components of the trial step.

Once the trial step is computed, the algorithm requires an estimate for the Lagrange multiplier λ_{k+1} to test the trial step for acceptance. Any approximation to the Lagrange multiplier vector that produces a bounded sequence of multipliers can be used.

We test whether the point $(x_k + s_k, \lambda_{k+1})$ is acceptable as a next iterate. For testing the trial steps, Dennis, El-Alem and Maciel (1997)[3] use, as a merit function, the augmented Lagrangian

$$\Phi(x, \lambda; \rho) = f(x) + \lambda^T C(x) + \rho ||C(x)||^2,$$

where $\rho > 0$ is the penalty parameter.

The actual reduction in the merit function in moving from (x_k, λ_k) to $(x_k + s_k, \lambda_{k+1})$ is

$$Ared_k = \Phi(x_k, \lambda_k; \rho_k) - \Phi(x_k + s_k, \lambda_{k+1}; \rho_k).$$

The predicted reduction that Dennis, El-Alem, and Maciel (1997)[3] use has the form:

$$\begin{aligned} Pred_k &= -\nabla_x \ell_k^T s_k - \frac{1}{2} s_k{}^T H_k s_k - (\lambda_{k+1} - \lambda_k)^T [C_k + \nabla C_k^T s_k] \\ &\quad + \rho_k [\, ||C_k||^2 - ||C_k + \nabla C_k^T s_k||^2]. \end{aligned} \tag{2}$$

The acceptable step must produce a decrease in the merit function Φ. To test for this, the predicted reduction has to be made greater than zero. Thus, if necessary, the value of the penalty parameter is increased before the algorithm tests the trial step.

For updating the penalty parameter, Dennis, El-Alem, and Maciel (1997)[3] use a scheme suggested by the author [5]. This scheme ensures that the predicted decrease in the merit function at each iteration is at least a fraction of the Cauchy decrease in the quadratic model of the linearized constraints. At the current iterate x_k, after choosing the step s_k and the multiplier λ_{k+1}, we tentatively set $\rho_k = \rho_{k-1}$ and if $Pred_k < \frac{\rho_k}{2}[||C_k||^2 - ||C_k + \nabla C_k^T s_k||^2]$ then, we change ρ_k to

$$\rho_k = \frac{2[q_k(s_k) - q_k(0) + (\lambda_{k+1} - \lambda_k)^T (C_k + \nabla C_k^T s_k)]}{||C_k||^2 - ||C_k + \nabla C_k^T s_k||^2} + \beta_o,$$

for some $\beta_o > 0$. This way of updating the penalty parameter ensures that, at the current iterate x_k,

$$Pred_k \geq \frac{\rho_k}{2}[||C_k||^2 - ||C_k + \nabla C_k^T s_k||^2]. \tag{3}$$

After computing a trial step and updating the penalty parameter, we test the step and accept it only if the actual reduction is greater than some fraction of the predicted reduction. That is, we accept the step s_k if $\frac{Ared_k}{Pred_k} \geq \eta_1$ where $\eta_1 \in (0,1)$ is a small fixed constant. Otherwise, we reject the step and decrease the trust-region radius by setting $\delta_k = \alpha||s_k||$, where $\alpha \in (0,1)$.

From the theoretical point of view, a proof of global convergence requires that the trust-region radius be decreased when the trial step is rejected. On the other hand, when the step is accepted, the radius of the trust region must increase or remain the same. However, $\delta_{\min} \leq \delta_{k+1}$ is required once the trial step s_k is accepted. The analysis also requires that, for all k, $\delta_k \leq \delta_{\max}$. See Dennis, El-Alem, and Maciel (1997)[3] for a practical way of updating the trust-region radius.

After accepting the step, the approximate Hessian matrix H_k must be updated. The algorithm requires the sequence $\{H_k\}$ of approximate Hessians be bounded. Thus, the exact Hessians or any approximation scheme that produces a bounded sequence of Hessians can be used.

Since we do not specify a particular way of computing W_k, it is required that $\{||W_k||\}$ be bounded and the smallest singular values of the members of the sequence $\{W_k\}$ be bounded below and away from zero.

Finally, the algorithm is terminated when $||W_k^T \nabla \ell_k|| + ||C_k|| \leq \varepsilon_{tol}$, for some $\varepsilon_{tol} > 0$.

2.2 Summary of the DEM class

We present a summary of Dennis, El-Alem, and Maciel's class of trust-region-based algorithms for solving problem (EQ).

Algorithm 2.1. The DEM Algorithm.

step 0. *Given x_0, λ_0, compute W_0.*

Choose $\delta_{\min}$, $\delta_{\max}$, and δ_0 such that $\delta_{\min} \leq \delta_0 \leq \delta_{\max}$.
Choose $\beta_o > 0$ and $\varepsilon_{tol} > 0$.
Set $\rho_{-1} = 1$ and $k = 0$

step 1. *If $\|W_k^T \nabla \ell_k\| + \|C_k\| \leq \varepsilon_{tol}$ then terminate.*

step 2. *If x_k is feasible then*

find a step s_k^t that satisfies a fraction-of-Cauchy decrease condition on the quadratic model $q_k(s)$ of the Lagrangian function around x_k.
Set $s_k = s_k^t$.
else

a) Compute a quasi-normal step s_k^n that satisfies a fraction-of-Cauchy decrease condition on the quadratic model of the linearized constraints.
b) If $W_k^T \nabla q(s_k^n) = 0$ then set $s_k^t = 0$,
else find s_k^t that satisfies a fraction-of-Cauchy decrease condition on the quadratic model $q_k(s_k^n + s^t)$ from s_k^n.
c) Set $s_k = s_k^n + s_k^t$.

step 3. *Choose an estimate λ_{k+1} of the Lagrange multiplier vector.*

step 4. *Update the penalty parameter ρ_{k-1} to obtain ρ_k.*

step 5. *Evaluate the step and update the trust-region radius.*

If the step is accepted then update H_k, set $k = k + 1$, and go to step 1,
else go to step 2.

2.3 Characterization of the DEM class

In this section, we list a set of algorithmic assumptions that hold for any member of the DEM class of algorithms. In other words, the following assumptions characterize the DEM class.

(A1) The step s_k^n satisfies the fraction-of-Cauchy decrease condition on the quadratic model of the linearized constraints.

(A2) The step s_k^t satisfies the fraction-of-Cauchy decrease condition on the quadratic model $q_k(s_k^n + s^t)$ from s_k^n reduced to the null space of ∇C_k^T.

(A3) The step s_k^n satisfies, for all k, $\|s_k^n\| \leq K_1 \|C_k\|$.

(A4) The sequence of projection matrices $\{W_k\}$ is bounded and the sequence of smallest singular values of the W_k's is bounded below and away from zero.

(A5) The sequence of Lagrange multiplier vectors $\{\lambda_k\}$ is bounded.

(A6) If an approximation to the Hessian of the Lagrangian is used, then the sequence of matrices $\{H_k\}$ is bounded.

(A7) The trial steps are tested using the augmented Lagrangian as a merit function.

(A8) The penalty parameter is updated so that the predicted decrease in the merit function at each iteration is at least ρ_k times a fraction of the Cauchy decrease in the quadratic model of the linearized constraints.

Throughout the rest of the paper we assume that the algorithm we use belongs to the DEM class. In other words, we assume that A1-A8 hold.

In Section 4, we present a global convergence theory for the algorithm. However, for this result to follow, we require that the problem we solve satisfy certain assumptions that we describe in the next section.

3 PROBLEM ASSUMPTIONS

We assume certain continuity and boundedness assumptions on the functions f and C of problem (EQ) and on their derivatives.

Let $\Omega \in \Re^n$ be a convex set that, contains all iterates x_k and $x_k + s_k$, for all trial steps s_k examined in the course of the algorithm. The following assumptions are imposed on problem (EQ).

(P1) f and $C_i \in C^2(\Omega)$ $i = 1, ..., m$.

(P2) $\nabla C(x)$ has full column rank for all $x \in \Omega$.

(P3) $f(x)$, $C(x)$, $\nabla f(x)$, $\nabla C(x)$, $\nabla^2 f(x)$, $(\nabla C(x)^T \nabla C(x))^{-1}$, and each of $\nabla^2 C_i(x)$, for $i = 1, ..., m$ are all uniformly bounded in Ω.

An immediate consequence of the above assumptions and A4-A6 is the existence of positive constants b, b_1, and b_2 such that any iteration k generated by the DEM class of algorithm satisfies

$$||H_k|| \leq b, \quad ||W_k^T H_k W_k|| \leq b, \quad ||W_k^T H_k|| \leq b,$$

$$|l(x_k, \lambda_k)| \leq b_1, \tag{4}$$

and

$$||C_k|| \leq b_2. \tag{5}$$

4 GLOBAL CONVERGENCE

In this section, we prove global convergence. In the first two subsections, we present some technical lemmas and intermediate results. The proof of our main global convergence result builds upon these intermediate results and appears in the third and fourth subsections.

4.1 Technical Lemmas

The following two lemmas express in a manageable form the pair of fraction of Cauchy decrease conditions imposed on the two components, s_k^n and s_k^t, of the trial steps.

Lemma 4.1. *Assume (P1)-(P3). Then there exists a positive constant K_2 independent of the iterates such that the quasi-normal component s_k^n of the trial step s_k satisfies*

$$\|C_k\|^2 - \|C_k + \nabla C_k^T s_k^n\|^2 \geq K_2 \|C_k\| \min\{\|C_k\|, \delta_k\}.$$

Proof. See Powell (1975)[14]. □

From the penalty parameter update formula and the above lemma, we have for all k,

$$Pred_k \geq \frac{1}{2}\rho_k K_2 \|C_k\| \min\{\|C_k\|, \delta_k\}. \tag{6}$$

Lemma 4.2. *Assume (P1)-(P3). Then there exists a positive constant K_3 independent of the iterates such that*

$$q_k(s_k^n) - q_k(s_k) \geq K_3 \|W_k^T \nabla q_k(s_k^n)\| \min\{\|W_k^T \nabla q_k(s_k^n)\|, \delta_k\}.$$

Proof. See Powell and Yuan (1991)[15]. □

The following lemma gives an upper bound on the difference between the actual reduction and the predicted reduction. It shows how accurate our definition of $Pred_k$ is as an approximation to $Ared_k$.

Lemma 4.3. *Assume (P1) and (P3), then there exists a constant $K_4 > 0$ that does not depend on k, such that*

$$|Ared_k - Pred_k| \leq K_4 \rho_k \|s_k\|^2. \tag{7}$$

Proof. See Corollary 6.4 of El-Alem (1991)[6]. □

The following lemma gives a lower bound to the predicted decrease in the merit function produced by the trial step.

Lemma 4.4. *Assume (P1) and (P3). Then the predicted decrease in the merit function satisfies*

$$\begin{aligned} Pred_k \quad \geq \quad & K_3 \|W_k^T \nabla q_k(s_k^n)\| \min\{\|W_k^T \nabla q_k(s_k^n)\| \ , \ \delta_k\} \\ & - K_5 \|C_k\| + \rho_k [\|C_k\|^2 - \|C_k + \nabla C_k^T s_k\|^2], \end{aligned}$$

where K_3 is as in Lemma 4.2 and K_5 is a positive constant independent of k.

Proof. See Lemma 7.6 of Dennis, El-Alem, and Maciel (1997)[3]. □

Lemma 4.5. *Assume (P1)-(P3). Let k be the index of an iteration at which ρ_k is increased. Then there exists a positive constant K_6 that does not depend on k, such that*

$$\rho_k \min\{\|C_k\|, \delta_k\} \leq \ K_6, \tag{8}$$

Proof. The proof follows from the first part of the proof of Lemma 7.10 of Dennis, El-Alem, and Maciel (1997)[3]. ☐

Lemma 4.6. *Assume (P1) and (P3). If at a given iteration k, $\|W_k^T \nabla f_k\| \geq \varepsilon_0$ and $\|C_k\| \leq \beta \delta_k$ where ε_0 is a positive constant and β is given by*

$$0 < \beta \leq \min \left\{ \frac{\varepsilon_0}{2bK_1\delta_{\max}}, \frac{K_3\varepsilon_0}{4K_5} \min\{\frac{\varepsilon_0}{2\delta_{\max}}, 1\} \right\},$$

then there exists a positive constant K_7 that depends on ε_0 but does not depend on k, such that

$$Pred_k \geq K_7\delta_k + \rho_k\{\|C_k\|^2 - \|C_k + \nabla C_k^T s_k\|^2\}. \tag{9}$$

Proof. The proof is similar to the proof of Lemma 7.7 plus the proof of Lemma 7.8 of Dennis, El-Alem, and Maciel (1997)[3]. ☐

4.2 Intermediate Results

This section is devoted to presenting some intermediate lemmas that are needed in the proof of our main results. We start with the following lemma which shows that if at any iteration k, the point x_k is not feasible, then the algorithm can not loop infinitely without finding an acceptable step. To state this result, we need to introduce one more notation. The j^{th} trial of iteration k is denoted by k^j.

Lemma 4.7. *Assume (P1)-(P3). If $\|C_k\| \geq \varepsilon$, where ε is any positive constant, then an acceptable step is found after finitely many trials. i. e., the condition $Ared_{k^j}/Pred_{k^j} \geq \eta_1$ will be satisfied for some finite j.*

Proof. Since $\|C_k\| \geq \varepsilon > 0$, then we have, using (6) and (7),

$$\left|\frac{Ared_k}{Pred_k} - 1\right| = \frac{|Ared_k - Pred_k|}{Pred_k} \leq \frac{2K_4\delta_k^2}{K_2\varepsilon\min\{\varepsilon, \delta_k\}}.$$

Now as the trial step s_{k^j} gets rejected, δ_{k^j} becomes small and eventually we will have

$$\left|\frac{Ared_{k^j}}{Pred_{k^j}} - 1\right| \leq \frac{2K_4\delta_{k^j}}{K_2\varepsilon}.$$

This inequality implies that after finite number of trials (*i. e.*, for j finite), the acceptance rule will be met. This completes the proof. ☐

Lemma 4.8. *Assume (P1)-(P3). If at a given iteration k, the j^{th} trial step satisfies*

$$\|s_{k^j}\| \leq \min\{\frac{(1-\eta_1)K_2}{4K_4}, 1\}\|C_k\|, \tag{10}$$

then it must be accepted.

Proof. The proof is by contradiction. Suppose that the statement of the lemma is not true. *i. e.*, suppose that the step s_{k^j} is rejected. Then we have

$$(1-\eta_1) < \frac{|Ared_{k^j} - Pred_{k^j}|}{Pred_{k^j}}.$$

Substituting from (6) and (7) and using (10), we have

$$(1-\eta_1) < \frac{2K_4\|s_{k^j}\|^2}{K_2\|C_k\|\|s_{k^j}\|} \leq \frac{1}{2}(1-\eta_1).$$

This gives a contradiction and implies that the step must be accepted. This competes the proof of the lemma. □

The following lemma is a consequence of the above lemma.

Lemma 4.9. *Assume (P1)-(P3). Then for all trail iterates j of any iterate k generated by the algorithm, we have*

$$\delta_{k^j} \geq \min\{\frac{\delta_{\min}}{b_2}, \alpha\frac{(1-\eta_1)K_2}{4K_4}, \alpha\}\|C_k\|. \tag{11}$$

Proof. Consider any iterate k^j. If the previous step was accepted; *i. e.*, $j=1$, then $\delta_k \geq \delta_{\min}$. Using (5), we can write

$$\delta_{k^j} \geq \frac{\delta_{\min}}{b_2}\|C_k\|.$$

Therefore, (11) holds in this case.

Now assume that $j>1$. *i. e.*, there exists at least one rejected trial step. For all rejected trial steps we must have

$$\|s_{k^i}\| > \min\{\frac{(1-\eta_1)K_2}{4K_4}, 1\}\|C_k\|,$$

for all $i=1,...,j-1$, otherwise, we get a contradiction to Lemma 4.8. From the way of updating the trust region, we have

$$\delta_{k^j} = \alpha\|s_{k^{j-1}}\| > \alpha\min\{\frac{(1-\eta_1)K_2}{4K_4}, 1\}\|C_k\|,$$

Hence the lemma is proved. □

The following lemma will be used in proving that the algorithm converges to the feasible region. It says that as long as $\|C_k\|$ is bounded away from zero, the trust-region radius is bounded away from zero.

Lemma 4.10. *Assume (P1)-(P3). Then any iterate x_k such that $\|C_k\| \geq \varepsilon_0$, where $\varepsilon_0 > 0$, satisfies*

$$\delta_{k^j} \geq K_8,$$

where K_8 is a positive constant that depends on ε_0 but does not depend on k.

Proof. The proof follows directly from the above lemma. Simply take $K_8 = \varepsilon_0\min\{\frac{\delta_{\min}}{b_2}, \alpha\frac{(1-\eta_1)K_2}{4K_4}, \alpha\}$. □

From (8) and (11), we have for all k^j at which the penalty parameter is increased

$$\rho_{k^j}\|C_k\| \leq K_9, \tag{12}$$

where K_9 is a positive constant that does not depend on k or j. This inequality is used in proving that the sequence of iterates generated by the algorithm converges to the feasible set. This is the subject of the following section.

4.3 Convergence to the Feasible Set

In the following theorem, we prove that the sequence of iterates generated by the algorithm converges to the feasible set.

Theorem 4.11. *Assume (P1)-(P3). Then the sequence of iterates generated by the algorithm satisfies*

$$\lim_{k\to\infty} ||C_k|| = 0.$$

Proof. Suppose that $\limsup_{k\to\infty} ||C_k|| \geq \varepsilon > 0$. This implies the existence of an infinite subsequence of indices $\{k_j\}$ indexing iterates that satisfy $||C_k|| \geq \frac{\varepsilon}{2}$, for all $k \in \{k_j\}$.

From Lemma 4.7, there exists an infinite sequence of acceptable steps. Without loss of generality, we assume that all members of the sequence $\{k_j\}$ are acceptable iterates.

Consider two cases. If $\{\rho_k\}$ is bounded, then there exists an integer $\bar{k}$ such that for all $k \geq \bar{k}$ the value of the penalty parameter remains the same. We have, for all $\hat{k} \geq \bar{k}$ and $\hat{k} \in \{k_j\}$,

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} = Ared_{\hat{k}} \geq \eta_1 Pred_{\hat{k}} \geq \eta_1 K_2 \frac{\varepsilon}{4} \min\{\frac{\varepsilon}{2}, \delta_{\hat{k}}\}.$$

Using Lemma 4.10, we have

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} \geq \eta_1 K_2 \frac{\varepsilon}{4} \min\{\frac{\varepsilon}{2}, \bar{K}_8\} > 0,$$

where $\bar{K}_8$ is as K_8 in Lemma 4.10 with ε_0 is replaced by $\frac{\varepsilon}{2}$. Now, as $\hat{k} \to \infty$, we get a contradiction.

The second case is when $\{\rho_k\}$ is unbounded. This implies the existence of a subsequence of indices $\{k_i\}$ indexing iterates at which the penalty parameter in increased. Because of (12),

$$\lim_{k_i\to\infty} ||C_{k_i}|| = 0. \tag{13}$$

Therefore, for k sufficiently large, there are no common elements between the two sequences $\{k_i\}$ and $\{k_j\}$. We have, for all $\hat{k} \in \{k_j\}$,

$$\frac{Ared_{\hat{k}}}{\rho_{\hat{k}}} \geq \eta_1 \frac{Pred_{\hat{k}}}{\rho_{\hat{k}}} \geq \eta_1 \frac{\varepsilon K_2}{4} \min[\frac{\varepsilon}{2}, \delta_{\hat{k}}] \geq \eta_1 \frac{\varepsilon K_2}{4} \min[\frac{\varepsilon}{2}, \bar{K}_8],$$

where $\bar{K}_8$ is as above. Hence, we have

$$\frac{\ell_{\hat{k}} - \ell_{\hat{k}+1}}{\rho_{\hat{k}}} + ||C_{\hat{k}}||^2 - ||C_{\hat{k}+1}||^2 \geq \eta_1 \frac{\varepsilon K_2}{4} \min[\frac{\varepsilon}{2}, \bar{K}_8] > 0. \tag{14}$$

On the other hand, for all acceptable steps generated by the algorithm, we have

$$\frac{\ell_k - \ell_{k+1}}{\rho_k} + ||C_k||^2 - ||C_{k+1}||^2 \geq 0. \tag{15}$$

Let $k_{\hat{i}}$ and $k_{\hat{i}+1}$ be two consecutive elements of the sequence $\{k_i\}$ such that there exists an iterate $k \in \{k_j\}$ between $k_{\hat{i}}$ and $k_{\hat{i}+1}$. From (14) and (15) , we can write

$$\sum_{k=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\{\ell_k - \ell_{k+1}\}}{\rho_k} + \|C_{k_{\hat{i}}}\|^2 - \|C_{\hat{i}+1}\|^2 \geq \eta_1 \frac{\varepsilon K_2}{4} \min[\frac{\varepsilon}{2}, \bar{K}_8] > 0.$$

Because the value of the penalty parameter is the same for all iterates $k_{\hat{i}}, \ldots, k_{\hat{i}+1} - 1$, we have

$$\frac{\ell_{k_{\hat{i}}} - \ell_{k_{\hat{i}+1}}}{\rho_{k_{\hat{i}}}} + \|C_{k_{\hat{i}}}\|^2 - \|C_{k_{\hat{i}+1}}\|^2 \geq \eta_1 \frac{\varepsilon K_2}{4} \min[\frac{\varepsilon}{2}, \bar{K}_8].$$

But because ℓ_k is bounded and $\rho_k \to \infty$ as $k \to \infty$, we can write, for $k_{\hat{i}}$ sufficiently large

$$\|C_{k_{\hat{i}}}\|^2 - \|C_{k_{\hat{i}+1}}\|^2 \geq \eta_1 \frac{\varepsilon K_2}{8} \min[\frac{\varepsilon}{2}, \bar{K}_8] > 0.$$

This contradicts (13). The supposition is wrong. This proves the theorem. □

4.4 Main Result

The main result is proved in Theorem 4.15. First, we present the following two lemmas which are used in the proof of the theorem.

Lemma 4.12. *Assume (P1)-(P3), then there exists a positive constant K_{10}, independent of k, such that*

$$|Ared_k - Pred_k| \leq K_{10} \, [\ \|s_k\|^2 + \rho_k \|s_k\|^3 + \rho_k \|s_k\|^2 \|C_k\|\ ].$$

Proof. See Lemma 6.3 of El-Alem (1991)[6]. □

Lemma 4.13. *Assume (P1)-(P3). If at a given iteration k^j, $\|W_k^T \nabla f_k\| \geq \varepsilon_0$ and $\|C_k\| \leq \beta \delta_{k^j}$ where ε_0 is a positive constant and β is as in Lemma 4.6 and if the penalty parameter is bounded, then*

$$\delta_{k^j} \geq K_{11},$$

where K_{11} is a positive constant that does not depend on k or j.

Proof. Because ρ_k is bounded, there exist $\bar{k}$ and $\bar{\rho}$ such that for all $k > \bar{k}$, $\rho_k = \bar{\rho}$.

Let $k^j > \bar{k}$. If the previous step was accepted; *i. e.*, $j = 1$, then $\delta_{k^1} \geq \delta_{\min}$. Hence δ_{k^j} is bounded in this case.

Now assume that $j > 1$. *i. e.*, there exists at least one rejected trial step. For all rejected trial steps, using Lemmas 4.3 and 4.6, we must have for $i = 1, ..., j-1$

$$(1 - \eta_1) < \frac{\bar{\rho} K_4 \|s_{k^i}\|^2}{K_7 \delta_{k^i}}.$$

From the way of updating the trust-region radius, we have

$$\delta_{k^j} = \alpha \|s_{k^{j-1}}\| > \frac{\alpha(1-\eta_1) K_7}{\bar{\rho} K_4}.$$

Hence the lemma is proved. □

The following theorem together with Theorem 4.11 prove that the sequence of iterates generated by any member of the DEM class of algorithms satisfies the termination condition of the algorithm.

Theorem 4.14. *Assume (P1)-(P3). Then the sequence of iterates satisfies*

$$\liminf_{k\to\infty} \quad ||W_k^T \nabla f_k|| \;=\; 0.$$

Proof. The proof is by contradiction. Suppose that for all k, $||W_k^T\nabla f_k|| > \varepsilon_0$. Assume that there exists an infinite subsequence $\{k_i\}$ such that $||C_{k_i}|| > \beta\delta_{k_i}$, where β is any positive constant. For later use of β, we take it to be as in Lemma 4.6. Then, because $||C_k|| \to 0$, we have

$$\lim_{k_i\to\infty} \quad \delta_{k_i} = 0.$$

Consider any iterate $k^j \in \{k_i\}$. There are two cases to consider.

First, consider the case where the sequence $\{\rho_k\}$ is unbounded. We have for the rejected trial step $j-1$ of iteration k, $||C_k|| > \beta\delta_{k^j} = \alpha\beta||s_{k^{j-1}}||$. Use Lemma 4.12 and the fact that the trial step $s_{k^{j-1}}$ is rejected,

$$\begin{aligned}
(1-\eta_1) \quad &\leq \quad \frac{|Ared_{k^{j-1}} - Pred_{k^{j-1}}|}{Pred_{k^{j-1}}} \\
&\leq \quad \frac{2K_{10}[||s_{k^{j-1}}|| + \rho_{k^{j-1}}(||s_{k^{j-1}}||^2 + ||s_{k^{j-1}}|| ||C_k||)]}{\rho_{k^{j-1}} K_2 \min(\alpha\beta, 1)||C_k||} \\
&\leq \quad \frac{2K_{10}}{\rho_{k^{j-1}} K_2 \alpha\beta \min(\alpha\beta, 1)} + \frac{2K_{10}(1+\alpha\beta)}{K_2\alpha\beta\min(\alpha\beta,1)}\delta_{k^{j-1}}.
\end{aligned}$$

This gives a contradiction when $\rho_{k^{j-1}}$ is sufficiently large and $\delta_{k^{j-1}}$ is sufficiently small. So δ_{k^j} can not go to zero in this case.

Second, consider the case when the sequence $\{\rho_k\}$ is bounded. There exists an integer $\bar{k}$ such that for all $k \geq \bar{k}$ the value of the penalty parameter is the same. Let $k \geq \bar{k}$ and consider a trial step j of iteration k, such that $||C_k|| > \beta\delta_{k^j}$.

If $j = 1$, then from our way of updating the trust-region radius, we have $\delta_{k^j} \geq \delta_{\min}$. Hence δ_{k^j} is bounded in this case. If $j > 1$, and

$$||C_{k^l}|| > \beta\delta_{k^l}, \tag{16}$$

where $l = 1, ..., j$, then for all rejected trial steps $l = 1, \ldots, j-1$ of iteration k we have

$$(1-\eta_1) \leq \frac{|Ared_{k^l} - Pred_{k^l}|}{Pred_{k^l}} \leq \frac{2K_4||s_{k^l}||}{K_2\min(\beta,1)||C_k||}.$$

Hence,

$$\begin{aligned}
\delta_{k^j} = \alpha||s_{k^{j-1}}|| \quad &\geq \quad \frac{\alpha K_2 \min(\beta,1)(1-\eta_1)||C_k||}{2K_4} \geq \frac{\alpha K_2\min(\beta,1)(1-\eta_1)\beta}{2K_4}\delta_{k^1} \\
&\geq \quad \frac{\alpha K_2\min(\beta,1)(1-\eta_1)\beta}{2K_4}\delta_{\min}.
\end{aligned}$$

Hence δ_{k^j} is bounded in this case too. If $j > 1$ and (16) does not hold for all l, there exists an integer m such that (16) holds for $l = m+1, ..., j$ and

$$\|C_{k^l}\| \leq \beta\delta_{k^l}, \tag{17}$$

for $l = 1, ..., m$. As in the above case, we can write

$$\delta_{k^j} \geq \frac{\alpha K_2 \min(\beta,1)(1-\eta_1)}{2K_4}\|C_k\| \geq \frac{\alpha K_2 \min(\beta,1)(1-\eta_1)\beta}{2K_4}\delta_{k^{m+1}}. \tag{18}$$

But from our way of updating the trust-region radius, we have

$$\delta_{k^{m+1}} \geq \alpha\|s_{k^m}\|. \tag{19}$$

Now using (17), Lemma 4.6, and the fact that the trial steps for all $l = 1, ..., m$ are rejected, we can write

$$(1-\eta_1) \leq \frac{|Ared_{k^l} - Pred_{k^l}|}{Pred_{k^l}} \leq \frac{2K_4\rho_k\|s_{k^l}\|}{K_9}.$$

Notice that ρ_k is bounded. This implies that, $\|s_{k^m}\|$ is bounded. This fact together with (18) and (19) imply that δ_{k^j} is bounded in this case too. Hence δ_{k^j} is bounded in all cases.

This contradiction implies that for k sufficiently large, all the iterates satisfy $\|C_k\| \leq \beta\delta_{k^j}$. This implies using Lemma 4.6 that there is no need to increase the value of the penalty parameter. So, $\{\rho_k\}$ is bounded. Let $k^j \geq \bar{k}$ and using (9), we have

$$\Phi_{k^j} - \Phi_{k^j+1} = Ared_{k^j} \geq \eta_1 Pred_{k^j} \geq \eta_1 K_7 \delta_{k^j}.$$

As k goes to infinity the above inequality implies that $\lim_{k\to\infty}\delta_{k^j} = 0$. This gives a contradiction to Lemma 4.13. This contradiction proves the theorem. □

5 CONCLUDING REMARKS

We have presented a short, straightforward, and self-contained global convergence theory for Dennis, El-Alem, and Maciel's class of algorithms. The result is slightly stronger than the result obtained by Dennis, El-Alem, and Maciel. We have shown that the whole sequence of iterates generated by the algorithm converges to the feasible set and a subsequence converges to a first-order point.

Many practical applications require feasible points that produce a reasonable amount of decrease in the objective function of problem (EQ). It is guaranteed by our theory that, from any starting point, any member of the DEM class of algorithms converges to the feasible set. It is also guaranteed that, for any $\varepsilon_{tol} > 0$, the algorithm terminates at a point that satisfies $\|W_k\nabla f_k\| + \|C_k\| \leq \varepsilon_{tol}$.

6 ACKNOWLEDGMENT

This work was done while the author was visiting the Department of Computational and Applied Mathematics and the Center of Research on Parallel Computation, Rice University. He thanks Rice University for its financial support and for the congenial scientific atmosphere that it provided.

REFERENCES

1. N. ALEXANDROV. *Multi-level Algorithms for Nonlinear Equations and Equality Constrained Optimization*. PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251, 1993.
2. R. H. BYRD and E. OMOJOKUN. Robust trust region methods for nonlinearly constrained optimization. 1987. Present at the 1987 SIAM Conference on Optimization, Houston, TX.
3. J. E. DENNIS Jr., M. M. EL-ALEM, and M. C. MACIEL. A global convergence theory for general trust-region-based algorithms for equality constrained optimization. *SIAM J. Optimization*, 7(1):177–207, Feb., 1997.
4. J. E. DENNIS Jr. and L. VICENTE. On the convergence theory of general trust-region-based algorithms for equality-constrained optimization. *SIAM J. Optimization*, 7(4), Nov., 1997.
5. M. M. EL-ALEM. *A global convergence theory for a class of trust region algorithms for constrained optimization*. PhD thesis, Department of Mathematical Sciences, Rice University, Houston, TX 77251, 1988.
6. M. M. EL-ALEM. A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization. *SIAM J. Numer. Anal.*, 28:266–290, Feb. 1991.
7. M. M. EL-ALEM. A robust trust region algorithm with a non-monotonic penalty parameter scheme for constrained optimization. *SIAM J. Optim.*, 5 No. 2:348–378, May 1995.
8. M. M. EL-ALEM. Convergence to a second order point of a trust-region algorithm with a non-monotonic penalty parameter scheme for constrained optimization. *J. Optim. Theory Appl.*, 91 No. 1:61–79, Oct. 1996.
9. M. LALEE. *Algorithms for nonlinear optimization*. PhD thesis, Department of Industrial Engineering, Northwestern University, Evanston, IL, 1993.
10. M. LALEE, J. NOCEDAL, and PLANTENGA. On the implementation of an algorithm for large-scale equality constrained optimization. Technical Report 93, EECS Department, Northwestern University, Evanston, IL, Oct. 1993.
11. M. C. MACIEL. *A global convergence theory for a general class of trust region algorithm for equality constrained optimization*. PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, 1992.
12. E. O. OMOJOKUN. *Trust region strategies for optimization with nonlinear equality and inequality constraints*. PhD thesis, Department of Computer Science, University of Colorado, Boulder, Colorado, 1989.
13. T. PLANTENGA. *Large-Scale nonlinear constrained optimization using trust regions*. PhD thesis, Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, 1995.
14. M. J. D. POWELL. Convergence properties of a class of minimization algorithms. In O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, editors, *Nonlinear Programming 2*, pages 1–27. Academic Press, New York, 1975.
15. M. J. D. POWELL and Y. YUAN. A trust region algorithm for equality constrained optimization. *Math. Prog.*, 49:189–211, 1991.
16. L. N. VICENTE. *Trust-Region interior-point algorithms for a class of nonlinear programming problems*. PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, 1996.

17. J. Z. ZHANG and D. T. ZHU. Projected quasi-Newton algorithm with trust region for constrained optimization. *J. Opt. Th. and Appl.*, 67:369–393, 1990.