A Global Convergence Theory for a General Class of Trust-Region-Based Algorithms for Constrained Optimization Without Assuming Regularity

Mahmoud El-Alem

CRPC-TR96655
Revised December 1997

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

A Global Convergence Theory for a General Class of Trust-Region-Based Algorithms for Constrained Optimization Without Assuming Regularity *

MAHMOUD EL-ALEM †

December 8, 1997

Abstract

This work presents a convergence theory for a general class of trust-region-based algorithms for solving the smooth nonlinear programming problem with equality constraints. The results are proved under very mild conditions on the quasi-normal and tangential components of the trial steps. The Lagrange multiplier estimates and the Hessian estimates are assumed to be bounded. In addition, the regularity assumption is not made. In particular, the linear independence of the gradients of the constraints is not assumed. The theory proves global convergence for the class. In particular, it shows that a subsequence of the iteration sequence satisfies one of four types of Mayer-Bliss stationary conditions in the limit. This theory holds for Dennis, El-Alem, and Maciel's class of trust-region-based algorithms.

Key Words: Nonlinear programming, equality constrained problems, constrained optimization, global convergence, regularity assumption, augmented Lagrangian, Mayer-Bliss points, stationary points, quasi-normal step, trust region.
Abbreviated Title: Convergence Theory Without Assuming Regularity.
AMS(MOS) subject classification: 65K05, 49D37.

1 Introduction

Over the last two decades, trust-region algorithms have enjoyed a good reputation on the basis of their remarkable numerical reliability in conjunction with a sound and complete convergence theory. They have proven to be very effective and robust techniques for solving the unconstrained and the equality constrained optimization problems.

*This research was done while the author was visiting the Center for Research on Parallel Computations. The author was supported by the NSF Cooperative Agreement No. CCR-9120008 as a visiting member of the Center for Research on Parallel Computation, Rice University, Houston, Texas 77251-1892. He thanks Rice University for the congenial scientific atmosphere that it provided.
The author is also greatly indebted to Richard Byrd and two anonymous referees and the associate editor for their careful and insightful reading of an earlier version of this paper.
†Department of Mathematics, Faculty of Science, Alexandria University, Alexandria, Egypt.

The first trust-region algorithm was given by Levenberg (1944)[24] and later was re-derived by Marquardt (1963)[27]. The algorithm was designed for solving nonlinear least-squares problems. Powell (1970)[36] derived from the Levenberg-Marquardt method the first trust-region algorithm for solving the unconstrained minimization problem. Detailed discussion about the Levenberg-Marquardt method can be found in Moré (1977)[32] and about the trust-region method for solving the unconstrained optimization problem can be found in Dennis and Schnabel (1983)[13], Fletcher (1987)[21], and Shultz, Schnabel, and Byrd (1985)[39].

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 $m < n$.

Most trust-region algorithms for solving problem (EQ) try to combine the trust-region idea with the successive quadratic programming (SQP) method. In general, the SQP method iteratively minimizes a quadratic model of the Lagrangian function

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

where λ is the Lagrange multiplier vector, subject to a linear approximation of the constraints. At each iteration 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

$$\begin{array}{ll} \textit{minimize} & \nabla_x \ell(x_k, \lambda_k)^T s + \frac{1}{2} s^T H_k s, \\ \textit{subject to} & C(x_k) + \nabla C(x_k)^T s = 0, \end{array}$$

where H_k is the Hessian of the Lagrangian function (1.1) at (x_k, λ_k) or an approximation to it.

If a trust-region constraint is simply added to the quadratic programming subproblem the resulting trust-region subproblem may be infeasible because the trust-region constraint and the hyperplane $C(x_k) + \nabla C(x_k)^T s = 0$ may have no intersecting points. In other words, the two constrained sets may be disjoint. Even if they intersect, there is no guarantee that when the trust-region radius δ_k is decreased, the above subproblem remains feasible. Note that, the global convergence of the trust-region methods is based on being able to reduce δ_k until the model trust-region subproblem accurately represents the actual problem.

To avoid possible infeasibility in the subproblem, different approaches have been proposed. The first approach is to relax the linear constraints in such a way that the resulting feasible set is non-empty. In particular, the hyperplane $C(x_k) + \nabla C(x_k)^T s = 0$ is replaced by the relaxed hyperplane $\nu_k C(x_k) + \nabla C(x_k)^T s = 0$, where $\nu_k \in [0, 1]$. This approach was first suggested by Miele, Huang, and Heideman (1969)[30] in the context of a line-search globalization strategy for solving problem (EQ) (see also Miele, Cragg, and Levy (1971)[29] and Miele, Levy, and Cragg (1971)[31]). It was later used to obtain a feasible trust-region subproblem by Vardi (1985)[40], Byrd, Schnabel, and Schultz (1987)[9], and El-Hallabi (1993)[20].

A major difficulty with this approach lies in the problem of choosing ν_k so that a feasible trust-region subproblem is guaranteed. This difficulty makes this approach impractical.

The second approach for resolving this infeasibility was proposed by Celis, Dennis, and Tapia (1985)[11]. They replaced the linear constraints by the quadratic constraint $\|C(x_k) + \nabla C(x_k)^T s\|_2^2 \le \theta_k$, where θ_k is a

given parameter chosen to ensure that the resulting trust-region subproblem is always feasible. This approach was used by El-Alem (1991)[16] and Powell and Yuan (1991)[38]. The parameter θ_k is also chosen to ensure a sufficient decrease in the quadratic model of the linearized constraints. This decrease is at least a fraction of the decrease obtained by the Cauchy step, which is defined to be the minimizer of $||C(x_k) + \nabla C(x_k)^T s||_2^2$ inside the trust region in the steepest descent direction.

In Celis, Dennis, and Tapia (1985)[11] and El-Alem (1991)[16], the parameter θ_k was taken to be

$$\theta_k = (1 - \hat{\nu})||C(x_k)||_2^2 + \hat{\nu}||C(x_k) + \nabla C(x_k)^T s_k^{cp}||_2^2,$$

for some fixed $\hat{\nu} \in (0, 1)$, where s_k^{cp} is the Cauchy step. In Powell and Yuan (1991)[38], the choice of θ_k was

$$\theta_k = \{\min ||C(x_k) + \nabla C(x_k)^T s||_2^2 \ : \ \underline{\nu}\, \delta_k \leq ||s||_2 \leq \bar{\nu}\delta_k\},$$

where $0 < \underline{\nu} \leq \bar{\nu} \leq 1$ and δ_k is the trust region radius.

A major disadvantage with this approach lies in the fact that the resulting trust-region subproblem has two quadratic constraints, so that there is no efficient algorithm for finding a good approximation to the solution of this subproblem. Although, Williamson (1990)[42] has attempted to produce an efficient algorithm by computing an inexact solution of the subproblem and others have suggested algorithms to solve special cases of this subproblem (see El-Alem and Tapia (1995)[19], Yuan (1988)[43] and (1990)[44], and Zhang (1992)[47]), the results are not in general satisfactory. This approach will remain impractical until an efficient way of solving the trust-region subproblem is discovered.

The reduced Hessian technique is another approach to overcoming the difficulty of having an infeasible trust-region subproblem. The approach was suggested by Byrd (1987)[8] and Omojokun (1989)[34]. 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

$$\begin{array}{ll} \textit{minimize} & ||C(x_k) + \nabla C(x_k)^T s^n||_2^2, \\ \textit{subject to} & ||s^n||_2 \leq \nu\delta_k, \end{array}$$

for some $\nu \in (0, 1)$. 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 $\nabla C(x_k)^T$ 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.

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

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] and [2], Alexandrov and Dennis (1994)[3], Biegler, Nocedal, and Schmid (1995)[4], Dennis and Vicente (1994)[14], El-Alem (1995)[17], and (1996)[18], Lalee (1993)[22], Lalee, Nocedal, and Plantenga (1993)[23], Maciel (1992)[25], Plantenga (1995)[35], Vicente (1996)[41], and Zhang and Zhu (1990)[46].

One of the advantages of this approach is that the two trust-region subproblems are similar to the trust-region subproblem for the unconstrained case.

Dennis, El-Alem, and Maciel (1997)[12] 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 unconstrained optimization, the use of a trust-region has made it possible to make strong guarantees of convergence. In particular, to guarantee global convergence, it is not necessary to require that the Hessian approximation be positive definite or even well conditioned, but only that it be uniformly bounded. To ensure global convergence, the step is required only to satisfy the fraction-of-Cauchy decrease condition; that is, the step must produce at least a fraction of the decrease obtained by the Cauchy step.

Powell (1975)[37] proved a powerful theorem. He showed that if the sequence of iterates generated by the algorithm satisfies the fraction-of-Cauchy decrease condition and if the sequence of Hessian approximations is bounded, then

$$\liminf_{k\to\infty} \|\nabla f(x_k)\|_2 = 0.$$

Powell's theorem does not prove convergence to a solution of the unconstrained problem. It only proves that a subsequence of the sequence of gradients of the objective function converges to zero. The strength of this result, however, comes from the weak assumptions imposed on the sequence of local models. Detailed discussion about the convergence results of trust-region algorithms for unconstrained optimization can be found in Carter (1986)[10], Moré (1983)[33], and Shultz, Schnabel, and Byrd (1985)[39].

Many authors have established global convergence results for algorithms that have been suggested for solving problem (EQ). The author in (1991)[16] and Powell and Yuan (1991)[38] have proved global convergence for variants of the Celis, Dennis, and Tapia trust-region algorithm by showing that

$$\liminf_{k\to\infty} \{\|Z_k^T \nabla f(x_k)\|_2 + \|C(x_k)\|_2\} = 0.$$

Analogous to Powell's theorem for the unconstrained case, Dennis, El-Alem, and Maciel (1992)[12] proved for their class of algorithms that,

$$\liminf_{k\to\infty} \{\|W_k^T \nabla f(x_k)\|_2 + \|C(x_k)\|_2\} = 0,$$

where W_k is a matrix that forms a basis (not necessarily orthogonal) for the null space of $\nabla C(x_k)^T$.

In Dennis, El-Alem, and Maciel's class of algorithms, the local model of the problem is generally taken to be a linear model of the constraints and a quadratic model of the Lagrangian function. The information in the local model depends on the Lagrange multiplier estimates as well as the second order information. Analogous to Powell's theorem, Dennis, El-Alem and Maciel only require the boundedness of the sequences of model Lagrange multipliers and Hessians. The results of Dennis, El-Alem, and Maciel were proved under very mild conditions on the quasi-normal and tangential components of the trial steps. However, their results were proved under the linear independence assumption.

In this paper, we reduce Dennis, El-Alem, and Maciel's assumptions even further and yet obtain similar global convergence results. In our theory, the linear independence assumption on the gradients of the constraints is not made. Our theory is so general that it holds for any algorithm that uses the augmented Lagrangian as a merit function, the El-Alem scheme for updating the penalty parameter [16], and bounded Lagrange multiplier and Hessian estimates.

The following notations are used throughout the rest of the paper. The sequence of points generated by the algorithm is denoted by $\{(x_k, \lambda_k)\}$. We abbreviate $f(x_k)$ as f_k, $\ell(x_k, \lambda_k)$ as ℓ_k, and so on. However, $f(x)$ is not abbreviated when emphasizing the dependence of f on x. We use the same symbol 0 to denote the real number zero, the zero vector, and the zero matrix. Finally, all norms used in this paper are l_2-norms.

The organization of the paper is as follows. In Section 2, we present in detail all the components of the general class of trust-region-based algorithms suggested by Dennis, El-Alem, and Maciel (1997)[12]. An overall summary of the class is presented at the end of this section. In Section 3, we state the assumptions under which we prove global convergence. The main results of this paper show that the algorithm generates a sequence of iterates that has a subsequence that asymptotically satisfies one of four types of stationary conditions for problem (EQ). In Section 4, we identify these conditions, state their definitions, and demonstrate some of their properties. In Section 5, we state our main global convergence results. Our convergence theory is presented in Sections 6-8. Finally, Section 9 contains concluding remarks.

2 General Trust-Region-Based Algorithms

In this section, we present the class of algorithms suggested by Dennis, El-Alem, and Maciel (1992)[12] for solving problem (EQ). This is a general class of trust-region-based algorithms. The basic idea of the trust-region algorithms is as follows. Approximate the problem by a model trust-region subproblem. The trial step is obtained by solving this subproblem. Test for accepting or rejecting the trial step and update the trust-region radius accordingly. If the step is rejected, decrease the radius of the trust region and compute another one using the new value of the trust-region radius. To test the trial steps, a merit function must be employed. Such a merit function often involves a parameter, usually called the penalty parameter. This parameter is updated using an updating scheme. More details about the trust-region method for constrained optimization can be found in Dennis, El-Alem, and Maciel (1992)[12].

In any trust-region algorithm for solving problem (EQ), there are four important issues to be considered. At each iteration k, we must first compute a trial step, and we address this issue in Section 2.1. Once the step is computed, we will need a criteria for accepting the trial step. Section 2.2 is devoted for this subject. To test the step, the penalty parameter needs to be updated. We address this issue in Section 2.3. Finally, we need a procedure for updating the trust-region radius and it is presented in Section 2.4. An overall summary of the algorithm is presented in Section 2.5.

2.1 Computing a Trial Step

We do not present a particular way for computing the trial steps. Instead, we present some conditions the steps must satisfy. Let s_k be decomposed into two components; the tangential component s_k^t and the quasi-normal component s_k^n. The trial step will then have the form $s_k = s_k^t + s_k^n$. Observe that the two components of the trial step are not necessarily orthogonal.

Dennis, El-Alem, and Maciel require that the quasi-normal component s_k^n of the trial step satisfy, at every iteration k,

$$\|s_k^n\| \leq K\|C_k\|, \tag{2.1}$$

where K is a positive constant. This condition is needed to obtain Dennis, El-Alem, and Maciel's global convergence results. It can be viewed as a relaxation to the orthogonality condition of s_k^n and s_k^t.

Because we do not assume linear independence of the gradients of the constraints, $\nabla C_k C_k = 0$ does not necessarily imply that $C_k = 0$. For this reason, we modify condition (2.1) to be

$$\|s_k^n\| \leq K\|s_k^{mn}\|, \tag{2.2}$$

where s_k^{mn} is the minimum-norm solution of

$$\text{(2.3)} \qquad \begin{array}{ll} \textit{minimize} & \|\nabla C_k^T s + C_k\|^2 \\ \textit{subject to} & \|s\| \le \tau\delta_k, \end{array}$$

for some $\tau \in (0,1)$, where δ_k is the trust-region radius.

As stated in Section 5.1 of [12], we do not suggest choosing K and enforcing condition (2.2). Rather, we suggest that (2.2) results naturally from any reasonable algorithm for computing a step s_k^n.

If $C_k \neq 0$, then the quasi-normal component s_k^n of the trial step is required to produce at least a fraction of the decrease in the quadratic model of the linearized constraints obtained by the Cauchy step. The Cauchy step s_k^{cp} is the step that solves the following problem:

$$\begin{array}{ll} \textit{minimize} & \frac{1}{2}\|\nabla C_k^T s + C_k\|^2 \\ \textit{subject to} & \|s\| \le \tau\delta_k \\ & s = -n_k^{cp}\nabla C_k C_k, \qquad n_k^{cp} > 0. \end{array}$$

So, the quasi-normal component s_k^n is chosen such that it satisfies for some $r_1 \in (0,1]$,

$$\|C_k\|^2 - \|C_k + \nabla C_k^T s^n\|^2 \ge r_1\{\|C_k\|^2 - \|C_k + \nabla C_k^T s_k^{cp}\|^2\}.$$

We note here that the Cauchy step defined above satisfies condition (2.2) for some $K > 0$.

Now we use the quasi-normal component to choose a linear manifold $\mathcal{M}_k$, parallel to the null-space of the constraints. Let $\mathcal{M}_k = \{s : \nabla C_k^T s = \nabla C_k^T s_k^n\}$. We select the tangential component from $\mathcal{M}_k$. Observe that, the intersection of $\mathcal{M}_k$ and the set $\{s = s^t + s_k^n : \|s\| \le \delta_k\}$ is not empty.

On the manifold $\mathcal{M}_k$, we consider the quadratic model $q_k(s)$ of the Lagrangian function (1.1) given by

$$\text{(2.4)} \qquad q_k(s) = \ell_k + \nabla_x \ell_k^T s + \frac{1}{2} s^T H_k s.$$

Let W_k be a matrix whose columns form a basis for the null space of ∇C_k^T. Then, when $W_k^T \nabla q_k(s_k^n) \neq 0$, the tangential component s_k^t is taken to be any step that satisfies the fraction of Cauchy decrease condition from s_k^n on $q_k(s)$ reduced to $\mathcal{M}_k$. That is, the trial step $s_k = s_k^t + s_k^n \in \mathcal{G}_k \cap \mathcal{M}_k$, where

$$\mathcal{G}_k = \{s = s^t + s_k^n : \ \|s\| \le \delta_k, \ q_k(s_k^n) - q_k(s) \ge r_2[q_k(s_k^n) - q_k(s_k^n - \mathrm{t}_k^{cp} W_k W_k^T \nabla q_k(s_k^n))]\}.$$

The constant $r_2 \in (0,1]$ and t_k^{cp} is given by

$$\mathrm{t}_k^{cp} = \begin{cases} \dfrac{\|W_k^T \nabla q_k(s_k^n)\|^2}{\nabla q_k(s_k^n)^T W_k H_k W_k^T \nabla q_k(s_k^n)} & \text{if } \dfrac{\|W_k^T \nabla q_k(s_k^n)\|^2 \|W_k W_k^T \nabla q_k(s_k^n)\|}{\nabla q_k(s_k^n)^T W_k H_k W_k^T \nabla q_k(s_k^n)} \le \bar{\delta}_k \\ & \text{and } \nabla q_k(s_k^n)^T W_k \bar{H}_k W_k^T \nabla q_k(s_k^n) > 0 \\ \dfrac{\bar{\delta}_k}{\|W_k W_k^T \nabla q_k(s_k^n)\|} & \text{otherwise,} \end{cases}$$

where $\bar{H}_k = W_k^T H_k W_k$ is the reduced Hessian matrix and $\bar{\delta}_k$ is the maximum length of the step allowed inside the set $\mathcal{M}_k \cap \{s = s^t + s_k^n : \|s\| \le \delta_k\}$ in the negative reduced gradient direction $-W_k^T \nabla q_k(s_k^n)$.

Once the trial step is computed, an estimate for the Lagrange multiplier λ_{k+1} is needed to determine whether the computed trial step will be accepted. Again, we will not present a particular way for computing the Lagrange multiplier. Instead, we impose a condition on the estimates of the Lagrange multiplier that is needed to prove global convergence. The sequence $\{\lambda_k\}$ of Lagrange multiplier estimates is required to be bounded. So, any approximation to the Lagrange multiplier vector that produces a bounded sequence can be used. For example, setting λ_k to a fixed vector (or even the zero vector) for all k is valid.

2.2 Testing the Trial Steps

Let s_k be a trial step computed by the algorithm and let λ_{k+1} be an estimate of the Lagrange multiplier vector. We test whether the point $(x_k + s_k, \lambda_{k+1})$ will be taken as a next iterate. In order to do this, a merit function is needed. We use, as a merit function, the augmented Lagrangian

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

where ρ 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 defined to be

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

This can be written as

$$Ared_k = \ell(x_k, \lambda_k) - \ell(x_k + s_k, \lambda_k) - \Delta\lambda_k^T C(x_k + s_k) + \rho_k [\ ||C_k||^2 - ||C(x_k + s_k)||^2],$$

where $\Delta\lambda_k = \lambda_{k+1} - \lambda_k$. The predicted reduction has the form:

$$Pred_k = -\nabla_x \ell_k^T s_k - \frac{1}{2} {s_k}^T H_k s_k - \Delta\lambda_k^T [C_k + \nabla C_k^T s_k] + \rho_k [\ ||C_k||^2 - ||C_k + \nabla C_k^T s_k||^2].$$

The acceptable step should be the step that produces a decrease in the merit function Φ. To test for this, the predicted reduction has to be forced to be greater than zero by increasing the value of the penalty parameter if necessary. This takes us to the following section.

2.3 Updating the Penalty Parameter

For updating the penalty parameter, Dennis, El-Alem, and Maciel (1992)[12] used a scheme proposed by the author [16]. This scheme ensures that the merit function is predicted to be decreased at each iteration by at least a fraction of the Cauchy decrease in the quadratic model of the linearized constraint. This indicates compatibility with the fraction of Cauchy decrease condition imposed on the quasi-normal component of the trial steps.

It is noteworthy that, since no regularity is assumed, there is no guarantee that when $||C_k||^2 - ||\nabla C_k^T s_k + C_k||^2 = 0$, we have $Pred_k \geq 0$. Therefore, it could happen that $||C_k||^2 - ||\nabla C_k^T s_k + C_k||^2 = 0$ and $Pred_k < 0$. In this case, the algorithm should be terminated because it is an infeasible stationary point of the constraints as we will show in Section 4. We write our way of updating the penalty parameter in algorithmic form as follows.

Algorithm 2.1 Updating the Penalty Parameter

1. Initialization

Set $\rho_{-1} = 1$ *and choose a small constant* $\hat{\rho} > 0$*.*

2. At the current iterate x_k, after s_k has been chosen:

set $\rho_k = \rho_{k-1}$.

a) If $Pred_k \leq 0$ *and* $||C_k||^2 - ||\nabla C_k^T s_k + C_k||^2 = 0$ **then** *terminate.*

b) If $Pred_k < \frac{\rho_k}{2}[||C_k||^2 - ||\nabla C_k^T s_k + C_k||^2]$, **then** *set*

$$\rho_k = \frac{2[q_k(s_k) - q_k(0) + \Delta\lambda_k^T(C_k + \nabla C_k^T s_k)]}{||C_k||^2 - ||\nabla C_k^T s_k + C_k||^2} + \hat{\rho}. \tag{2.5}$$

The initial choice of the penalty parameter ρ_{-1} is arbitrary. However, it should be chosen such that it is consistent with the scale of the problem. Here, for convenience, we take $\rho_{-1} = 1$.

An immediate consequence of the above algorithm is that, at the current iteration, either the point x_k is an infeasible stationary point of the constraints (see Section 4) and the algorithm terminates from Step 2-(a) of the above algorithm or

$$Pred_k \geq \frac{\rho_k}{2}[||C_k||^2 - ||C_k + \nabla C_k^T s_k||^2]. \tag{2.6}$$

2.4 Updating the Trust-Region Radius

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 $\eta_1 \in (0,1)$. Otherwise, we reject the step and decrease the radius of the trust region by setting $\delta_k = \alpha_1||s_k||$, where $\alpha_1 \in (0,1)$.

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, where $\delta_{\min}$ is a pre-specified constant. In short, δ_k can be reduced below $\delta_{\min}$ while seeking an acceptable step, but $\delta_{\min} \leq \delta_{k+1}$ must hold at the beginning of the next iteration after finding an acceptable step. The analysis also requires that, for all k, $\delta_k \leq \delta_{\max}$, where $\delta_{\max}$ is another pre-specified constant such that $\delta_{\min} \leq \delta_{\max}$. We present the way of updating the trust-region radius used by Dennis, El-Alem, and Maciel (1992)[12] in this algorithm.

Algorithm 2.2 Evaluating the Step and Updating the Trust-Region Radius

Given the constants: $0 < \alpha_1 < 1 < \alpha_2$, $0 < \eta_1 < \eta_2 < 1$, *and* $\delta_{\max} \geq \delta_k \geq \delta_{\min} > 0$.

While $\frac{Ared_k}{Pred_k} < \eta_1$

Reduce the trust-region radius: $\delta_k \leftarrow \alpha_1||s_k||$, *and compute a new trial step* s_k.

End while

If $\eta_1 \leq \frac{Ared_k}{Pred_k} < \eta_2$ **then**

Accept the step: $x_{k+1} = x_k + s_k$.

Set the trust-region radius: $\delta_{k+1} = \max\{\delta_k, \delta_{\min}\}$.

End if

If $\frac{Ared_k}{Pred_k} \geq \eta_2$ **then**

Accept the step: $x_{k+1} = x_k + s_k$.

Increase the trust-region radius: $\delta_{k+1} = \min\{\delta_{\max}, \max\{\delta_{\min}, \alpha_2\delta_k\}\}$.

End if

After accepting the step and updating the trust region radius, the approximate Hessian matrix H_k must be updated. Our theory 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. For instant, setting $H_k = 0$ for all k, is valid.

Since we do not specify a particular way for computing W_k, it is required that $\{||W_k||\}$ be bounded and the sequence of smallest singular values of the matrices W_k , $k = 1, 2, ...$ be bounded away from zero.

Finally, the algorithm is terminated when $||W_k^T \nabla \ell_k|| + ||\nabla C_k C_k|| \le \varepsilon_{tol}$, for some $\varepsilon_{tol} > 0$. Observe that, we use $||\nabla C_k C_k||$ instead of $||C_k||$ because the columns of ∇C_k may be linearly dependent. This implies that the algorithm may terminate at a point x_k that satisfies $\nabla C_k C_k = 0$ but does not satisfy $C_k = 0$.

2.5 Summary of the Algorithm

We present a summary of the DEM class of trust-region-based algorithms for solving problem (EQ).

Algorithm 2.3 The Trust-Region Algorithm.

Step 0. *(Initialization)*

Given x_0, λ_0, compute W_0.

Choose α_1, α_2, η_1, η_2, ε_{tol}, $\hat{\rho}$, $\delta_{\min}$, δ_0, and $\delta_{\max}$, such that $0 < \alpha_1 < 1 < \alpha_2$, $0 < \eta_1 < \eta_2 < 1$, $\varepsilon_{tol} > 0$, $\hat{\rho} > 0$, and $\delta_{\min} \le \delta_0 \le \delta_{\max}$.

Set $\rho_{-1} = 1$ and $k = 0$.

Step 1. *(Test for convergence)*

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

Step 2. *(Compute a trial step)*

If *x_k is feasible* **then**

a) find a step s_k^t that satisfies a fraction of Cauchy decrease condition on the quadratic model $q_k(s)$ of the Lagrangian around x_k. (See Section 2.1.)

b) Set $s_k = s_k^t$.

else *(* $C(x_k) \neq 0$ *)*

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. (See Section 2.1.)

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 $q_k(s_k^n + s)$ from s_k^n. (See Section 2.1.)*

End if

c) Set $s_k = s_k^n + s_k^t$.

End if

Step 3. *(Update λ_k)*

Choose an estimate λ_{k+1} of the Lagrange multiplier vector. Set $\Delta\lambda_k = \lambda_{k+1} - \lambda_k$.

Step 4. *(Update the penalty parameter)*

Update ρ_{k-1} to obtain ρ_k by using Algorithm 2.1.

Step 5. *(Evaluate the step)*

Evaluate the step and update the trust-region radius by using Algorithm 2.2.

If *the step is accepted*

then *update H_k, set $k = k + 1$, and go to* **Step 1**.

else *go to* **Step 2**.

End if

Because we do not assume that the columns of ∇C_k are linearly independent, another condition for terminating the algorithm should be added. For example, the following condition can be tested at the end of Step 2 of the above algorithm: if $||s_k|| \le \varepsilon_{tol}$ then terminate. The reason for that will be clear when we proceed with the paper. More details are given in Section 9.

3 General Assumptions

Let Ω be a convex subset of $\Re^n$ that contains all of x_k and $x_k + s_k$ for all trial steps s_k examined in the course of the algorithm. On the set Ω, we assume:

A1) f and C are twice continuously differentiable for all $x \in \Omega$.

A2) $f(x), \nabla f(x),\ \nabla^2 f(x),\ C(x),\ \nabla C(x),$ and $\nabla^2 C_i(x)$ for $i = 1, \cdots, m$ are uniformly bounded in Ω.

A3) The sequence of Lagrange multiplier vectors $\{\lambda_k\}$ is bounded.

A4) If approximations to the Hessian matrices are used, then we require that the matrices $H_k, k = 1, 2, ...$ be uniformly bounded in norms.

A5) The sequence $\{||W_k||\}$ is bounded and the sequence of smallest singular values of the matrices W_k , $k = 1, 2, ...$ is bounded away from zero.

The above are the assumptions under which we prove global convergence. Observe that they do not include the assumption of the linearly independence of the gradients of the constraints, a commonly used assumption by many researchers.

An immediate consequence of the above assumptions is the existence of positive constants b and b_1, such that for all k,

$$||\nabla C_k C_k|| \le b \tag{3.1}$$

and

$$||W_k^T H_k|| \le b_1. \tag{3.2}$$

4 Stationary Points

In this section, we give definitions to four types of stationary points, show some of their properties, and show some relations between them. The terminology used in this section follows Burke (1991)[6] and (1992)[7] and Yuan (1995)[45].

Definition 4.1 *First-order point*
A point $x_\star \in \Re^n$ is called a first-order point of problem (EQ), if it satisfies

$$W(x)^T \nabla f(x) = 0, \tag{4.1}$$

$$C(x) = 0. \tag{4.2}$$

Equations (4.1) and (4.2) are called the first-order conditions. If $x_\star$ solves (4.1), then this implies the existence of $\lambda_\star$ such that $x_\star$ and $\lambda_\star$ satisfy $\nabla f(x) + \nabla C(x)\lambda = 0$.

Definition 4.2 *Mayer-Bliss point*
A point $x_\star \in \Re^n$ is called a feasible Mayer-Bliss point or simply a Mayer-Bliss point of problem (EQ), if there exist a constant $\gamma_\star \in \Re$ and a Lagrange multiplier vector $\lambda_\star \in \Re^m$ such that $(\gamma_\star, \lambda_\star) \neq (0,0)$ and $x_\star$, $\gamma_\star$, and $\lambda_\star$ satisfy the following conditions

$$\gamma \nabla f(x) + \nabla C(x)\lambda = 0, \tag{4.3}$$

$$C(x) = 0. \tag{4.4}$$

Equations (4.3) and (4.4) are called the feasible Mayer-Bliss conditions. See Mayer (1886) [28] and Bliss (1938) [5].

The feasible Mayer-Bliss conditions is the same as the well known Fritz John's conditions for general nonlinear programming. See Mangasarian (1969)[26].

If $(x_\star, \gamma_\star, \lambda_\star)$ is a feasible Mayer-Bliss point and $\gamma_\star \neq 0$ then $(x_\star, \frac{\lambda_\star}{\gamma_\star})$ is a first-order point. Conversely, if $(x_\star, \lambda_\star)$ is a first-order point then it is a feasible Mayer-Bliss point with $\gamma_\star = 1$.

Definition 4.3 *Infeasible first-order point*
An infeasible point $x_\star \in \Re^n$ is called an infeasible first-order point of problem (EQ), if it satisfies

$$W(x)^T \nabla f(x) = 0, \tag{4.5}$$

$$\nabla C(x) C(x) = 0. \tag{4.6}$$

Equations (4.5) and (4.6) are called the infeasible first-order conditions for problem (EQ).

Definition 4.4 *Infeasible Mayer-Bliss point*
A point $x_\star \in \Re^n$ is called an infeasible Mayer-Bliss point if $x_\star$ satisfies the following conditions

$$\nabla C(x) C(x) = 0, \tag{4.7}$$

$$C(x) \neq 0. \tag{4.8}$$

Equations (4.7) and (4.8) are called the infeasible Mayer-Bliss conditions.

If $x_\star$ is an infeasible Mayer-Bliss point, then there exist a constant $\gamma_\star \in \Re$ and a Lagrange multiplier vector $\lambda_\star \in \Re^m$ such that $(\gamma_\star, \lambda_\star) \neq (0,0)$ and $x_\star$, $\gamma_\star$, and $\lambda_\star$ satisfy the following conditions

$$\begin{aligned} \gamma \nabla f(x) + \nabla C(x)\lambda &= 0, \\ \nabla C(x) C(x) &= 0. \end{aligned}$$

If in addition $\gamma_\star \neq 0$ then $(x_\star, \frac{\lambda_\star}{\gamma_\star})$ is an infeasible first-order point.

Definition 4.5 *Stationary conditions*
The conditions stated in any of Definitions 4.1-4.4 are called stationary conditions of problem (EQ) and the point that satisfies any of the stationary conditions is called a stationary point.

The following Lemma gives a condition for an infeasible iterate x_k generated by the algorithm to be a Mayer-Bliss point.

Lemma 4.6 *If at a point x_k generated by the algorithm $||C_k|| \neq 0$ and*

$$\text{minimize}_{||s|| \leq \delta_k} \quad ||C_k + \nabla C_k^T s||^2 = ||C_k||^2, \tag{4.9}$$

then it is an infeasible Mayer-Bliss point.

Proof: Let $\bar{q}(s,\mu) = ||C_k + \nabla C_k^T s||^2 + \mu ||s||^2$, where $\mu \geq 0$ is the multiplier associated with the trust-region constraint. Because $\bar{q}(s,\mu)$ is convex, the local minimizer is a global one. Also, $\nabla_s \bar{q}(s_k, \mu_k) = 0$ implies that the minimizer satisfies $\nabla C_k C_k = 0$. Hence (4.7) holds. This completes the proof. □

In the above Lemma, it is easy to see that at the point x_k, the matrix ∇C_k does not have full column rank.

It is noteworthy that if the algorithm generates a point x_k which is an infeasible Mayer-Bliss point and $Pred_k < 0$, then the algorithm may not be able to move away from this point. In this case, the algorithm terminates from Step 2(a) of Algorithm 2.1. However, we will proceed with the analysis assuming that this case will not happen.

The following lemma gives conditions for the sequence of iterates generated by the algorithm to have a subsequence that satisfies the feasible Mayer-Bliss conditions in the limit. A similar lemma for a different algorithm was given by Yuan (1995) [45].

Lemma 4.7 *If there exists a subsequence of infeasible iterates $\{k_j\}$ such that $\lim_{k_j \to \infty} ||C_{k_j}|| = 0$ and*

$$\lim_{k_j \to \infty} \left\{ \text{minimize}_{s \in \Re^n} \quad \frac{||C_{k_j} + \nabla C_{k_j}^T s||^2}{||C_{k_j}||^2} \quad : \quad ||s|| \leq ||C_{k_j}|| \right\} = 1, \tag{4.10}$$

then it satisfies the feasible Mayer-Bliss conditions in the limit.

Proof: The above limit is equivalent to

$$\lim_{k_j \to \infty} \left\{ \text{minimize}_{||d|| \leq 1} \quad ||U_{k_j} + \nabla C_{k_j}^T d||^2 \right\} = 1, \tag{4.11}$$

where U_{k_j} is a unit vector in the direction of C_{k_j} and $d = \frac{s}{\|C_{k_j}\|}$. Let $\bar{d}_{k_j}$ be a solution to the minimization problem inside the above limit. Then there exists a nonnegative parameter μ_{k_j} such that

$$\nabla C_{k_j} U_{k_j} + \nabla C_{k_j} \nabla C_{k_j}^T \bar{d}_{k_j} + \mu_{k_j} \bar{d}_{k_j} = 0 \tag{4.12}$$

and

$$\mu_{k_j}(\|\bar{d}_{k_j}\|^2 - 1) = 0. \tag{4.13}$$

From the optimality of $\bar{d}_{k_j}$, we have

$$\lim_{k_j \to \infty} \left\{ 2\bar{d}_{k_j}{}^T \nabla C_{k_j} U_{k_j} + \bar{d}_{k_j}{}^T \nabla C_{k_j} \nabla C_{k_j}^T \bar{d}_{k_j} \right\} = 0. \tag{4.14}$$

If $\lim_{k_j \to \infty} \bar{d}_{k_j} = 0$, then from (4.12), we have $\lim_{k_j \to \infty} \nabla C_{k_j} U_{k_j} = 0$. Otherwise, multiply (4.12) from the right by $2d_{k_j}^T$, subtract from (4.14), and use (4.13), we obtain as $k_j \to \infty$, $\nabla C_{k_j}^T \bar{d}_{k_j} \to 0$ and $\mu_{k_j} \to 0$. This yields

$$\lim_{k_j \to \infty} \nabla C_{k_j} U_{k_j} = 0.$$

Hence, in both cases, (4.3) holds in the limit with $\gamma = 0$. This shows that the lemma is true. □

From the above lemma, we can easily see that, for any sequence $\{k_i\}$ that asymptotically satisfies the Mayer-Bliss conditions that are not first-order conditions, the matrix ∇C_{k_i} does not have full column rank in the limit. So, for any subsequence $\{k_i\}$ of the iteration sequence, if $\{\|C_{k_i}\|\}$ converges to zero and the corresponding sequence of smallest singular values of $\{\nabla C_{k_i}\}$ converges to zero, then it satisfies the Mayer-Bliss conditions in the limit.

Throughout the rest of the paper, we use $\{\sigma_{k_i}\}$ to denote the sequence of smallest singular values of ∇C_k for all $k \in \{k_i\}$.

In the rest of the paper, we present our convergence results. We start with the following section which summarizes our global convergence results.

5 Main Result

In this section, we state the main result of our convergence analysis in order to understand the motivation for the lemmas presented in the next two sections.

Theorem 5.1 *Assume A1-A5, then the sequence of iterates generated by Algorithm 2.3 has a subsequence that satisfies one of the stationary conditions of problem (EQ) in the limit. In particular, it asymptotically satisfies either the infeasible Mayer-Bliss conditions, the feasible Mayer-Bliss conditions, the infeasible first-order conditions, or the first- order conditions.*

The above theorem summarizes the main results of this paper. The proof of this theorem is presented in Section 8. The proof needs some intermediate lemmas. They are presented in the following two sections.

6 Intermediate Results

In this section, we present some technical lemmas needed in the proof of our main global convergence results.

The following two lemmas use the fact that the steps s_k^n and s_k^t satisfy the fraction of Cauchy decrease condition. They express in a manageable form the pair of fraction of Cauchy decrease conditions imposed on the trial steps.

Lemma 6.1 *Assume A1-A2. Then there exists a positive constant K_1 independent of the iterates such that the quasi-normal component of the trial step s_k^n satisfies*

$$\|C_k\|^2 - \|C_k + \nabla C_k^T s_k^n\|^2 \geq K_1 \|\nabla C_k C_k\| \min\{\|\nabla C_k C_k\|, \delta_k\}. \tag{6.1}$$

Proof: If $\nabla C_k C_k = 0$, then $\nabla C_k^T s_k^n = 0$ and the lemma is valid a fortiori.

Assume that $\|\nabla C_k C_k\| > 0$. In this case, the proof follows from the fact that the step s_k^n satisfies the fraction of Cauchy decrease condition and using Assumption A2. For a proof see Powell (1975)[37] or Moré (1983)[33]. □

From (2.6) and the above lemma, we have, for all k

$$Pred_k \geq \frac{K_1 \rho_k}{2} \|\nabla C_k C_k\| \min\{\|\nabla C_k C_k\|, \delta_k\}. \tag{6.2}$$

Lemma 6.2 *Assume A1-A5. Then there exists a positive constant K_2 independent of the iterates such that*

$$q_k(s_k^n) - q_k(s_k) \geq K_2 \|W_k^T \nabla q_k(s_k^n)\| \min\{\|W_k^T \nabla q_k(s_k^n)\|, \delta_k\}. \tag{6.3}$$

Proof: The proof is similar to the proof of the above lemma. □

The following two lemmas give upper bounds on the difference between the actual reduction and the predicted reduction.

Lemma 6.3 *Assume A1-A4, then there exists a positive constant K_3, independent of k, such that*

$$|Ared_k - Pred_k| \leq K_3 \; [\; \|s_k\|^2 + \rho_k \|s_k\|^3 + \rho_k \|s_k\|^2 \|C_k\| \;]. \tag{6.4}$$

Proof: See Lemma 6.3 of El-Alem [16]. □

Lemma 6.4 *Assume A1-A4, then there exists a positive constant K_4 independent of k, such that*

$$|Ared_k - Pred_k| \leq K_4 \rho_k \|s_k\|^2. \tag{6.5}$$

Proof: The proof follows directly from the above lemma and the fact that $\rho_k \geq 1$ and $\|s_k\|$ and $\|C_k\|$ are uniformly bounded. □

The following lemma shows that if at any iteration k, the point x_k is not a stationary point of the constraints, then the algorithm can not loop infinitely without finding an acceptable step. To state the lemma, we need to introduce one more notation. The i^{th} trial iterate of iteration k is denoted by k^i.

Lemma 6.5 *Assume A1-A4. If $\|\nabla C_k C_k\| \geq \varepsilon$, 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 $\|\nabla C_k C_k\| \geq \varepsilon > 0$, then using (6.2) and (6.5), we have

$$\left| \frac{Ared_k}{Pred_k} - 1 \right| = \frac{|Ared_k - Pred_k|}{Pred_k} \leq \frac{2K_4 \delta_k^2}{K_1 \varepsilon \min\{\varepsilon, \delta_k\}}.$$

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

$$\left|\frac{Ared_{k^j}}{Pred_{k^j}} - 1\right| \leq \frac{2K_4\delta_{k^j}}{K_1\varepsilon}.$$

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

Lemma 6.6 *Assume A1-A5. Let j and k be any pair of indices such that ρ_{k^j} is increased at the j^{th} trial iterate of the k^{th} iteration, then there exists $K_5 > 0$ that does not depend on j or k, such that*

$$\rho_{k^j}\{\|C_k\| - \|C_k + \nabla C_k^T s_{k^j}\|\} \leq K_5 \max\{\|C_k\|, \|s_{k^j}^{mn}\|\},$$

where $s_{k^j}^{mn}$ is the minimum-norm solution of (2.3) with $\delta_k = \delta_{k^j}$.

Proof: If ρ_{k^j} is increased at the j^{th} trial step of the k^{th} iteration, then it is updated by (2.5). Hence,

$$\begin{aligned}
\frac{\rho_{k^j}}{2}[\|C_k\|^2 - \|C_k + \nabla C_k^T s_{k^j}\|^2] &= [q_k(s_{k^j}) - q_k(0)] + \Delta\lambda_{k^j}{}^T(C_k + \nabla C_k^T s_{k^j}) \\
&\quad + \frac{\hat{\rho}}{2}[\|C_k\|^2 - \|C_k + \nabla C_k^T s_{k^j}\|^2] \\
&= [q_k(s_{k^j}) - q_k(s_{k^j}^n)] + [q_k(s_{k^j}^n) - q_k(0)] + \Delta\lambda_{k^j}^T(C_k + \nabla C_k^T s_{k^j}^n) \\
&\quad + \frac{\hat{\rho}}{2}[-2(\nabla C_k C_k)^T s_{k^j}^n - \|\nabla C_k^T s_{k^j}^n\|^2] \\
&\leq [q_k(s_{k^j}) - q_k(s_{k^j}^n)] + \|\nabla \ell_k\|\|s_{k^j}^n\| + \frac{1}{2}\|H_k\|\|s_{k^j}^n\|^2 \\
&\quad + \|\Delta\lambda_{k^j}\|\|C_k + \nabla C_k^T s_{k^j}^n\| + \hat{\rho}[\|\nabla C_k C_k\|\|s_{k^j}^n\| + \|\nabla C_k^T\|^2\|s_{k^j}^n\|^2].
\end{aligned}$$

The rest of the proof follows by applying (6.3) and (2.2) to the right-hand side, followed by the use of the general assumptions. □

From (6.1) and the above lemma, we can write, at any iteration k^j at which the penalty parameter is increased,

(6.6)
$$\rho_{k^j}\|\nabla C_k C_k\| \min\{\|\nabla C_k C_k\|, \delta_{k^j}\} \leq \frac{K_5}{K_1} \max\{\|C_k\|, \|s_{k^j}^{mn}\|\}.$$

Lemma 6.7 *Assume A1-A4. If the j^{th} trial step of a given iteration k satisfies*

(6.7)
$$\|s_{k^j}\| \leq \min\{\frac{(1-\eta_1)K_1}{4K_4}, 1\}\|\nabla C_k C_k\|,$$

then the step must be accepted.

Proof: The proof is by contradiction. Suppose that (6.7) holds but 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 (6.2) and (6.5) and using (6.7), we have

$$(1-\eta_1) < \frac{2K_4\|s_{k^j}\|^2}{K_1\|\nabla C_k C_k\|\|s_{k^j}\|} \leq \frac{1}{2}(1-\eta_1).$$

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 6.8 *Assume A1-A4. All trail iterates j of any iteration k generated by the algorithm satisfy*

$$\delta_{k^j} \geq \min\{\frac{\delta_{\min}}{b}, \alpha_1 \frac{(1-\eta_1)K_1}{4K_4}, \alpha_1\}||\nabla C_k C_k||, \tag{6.8}$$

where b is as in (3.1).

Proof: Consider any iterate k^j. If the previous step was accepted; *i. e.*, $j = 1$, then $\delta_{k^j} = \delta_{k^1} \geq \delta_{\min}$. Using (3.1), we can write

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

Therefore, (6.8) holds in this case.

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

$$||s_{k^{j-1}}|| > \min\{\frac{(1-\eta_1)K_1}{4K_4}, 1\}||\nabla C_k C_k||,$$

otherwise, we get a contradiction with Lemma 6.7. From the way of updating the trust region, we have

$$\delta_{k^j} = \alpha_1 ||s_{k^{j-1}}|| > \alpha_1 \min\{\frac{(1-\eta_1)K_1}{4K_4}, 1\}||\nabla C_k C_k||,$$

Hence the lemma is proved. □

The following lemma is used in proving that the sequence $\{||\nabla C_k C_k||\}$ converges to zero. It says that as long as $\{||\nabla C_k C_k||\}$ is bounded away from zero, the sequence of trust-region radii $\{\delta_k\}$ is bounded away from zero.

Lemma 6.9 *Assume A1-A4. Then all trial iterates j of any iteration k such that $||\nabla C_k C_k|| \geq \varepsilon_0$, where $\varepsilon_0 > 0$, satisfies*

$$\delta_{k^j} > K_6,$$

where K_6 is a positive constant that depends on ε_0 but does not depend on k.

Proof: Taking

$$K_6 = \varepsilon_0 \min\{\frac{\delta_{\min}}{b}, \alpha_1 \frac{(1-\eta_1)K_1}{4K_4}, \alpha_1\}, \tag{6.9}$$

the proof follows directly from the above lemma. □

From (6.6) and Lemma 6.8 and using the general assumptions, we have for all k^j at which the penalty parameter is increased

$$\rho_{k^j}||\nabla C_k C_k||^2 \leq K_7, \tag{6.10}$$

where K_7 is a positive constant that does not depend on j or k. This inequality is used in studying the convergence of the sequence $\{||\nabla C_k C_k||\}$. This is the subject of the following section.

7 Stationary Points of the Constraints

The following lemma proves that for the iteration sequence generated by Algorithm 2.3, if $\{\rho_k\}$ is unbounded, then the sequence $\{||\nabla C_k C_k||\}$ is not bounded away from zero.

Lemma 7.1 *Assume A1-A5. If $\{\rho_k\}$ is unbounded, then the sequence of iterates generated by the algorithm satisfies*

$$\lim_{k_i \to \infty} \|\nabla C_{k_i} C_{k_i}\| = 0, \tag{7.1}$$

where $\{k_i\}$ is the sequence of iterates at which the penalty parameter is increased.

Proof: The proof follows directly from the assumption that $\{\rho_k\}$ is unbounded and (6.10). □

If in addition to the assumptions of the above lemma, we have $\limsup_{k_i \to \infty} \|C_{k_i}\| > 0$, then the sequence $\{k_i\}$ has a subsequence that satisfies the infeasible Mayer-Bliss conditions in the limit.

The following Lemma proves a stronger result when $\lim_{k_i \to \infty} \|C_{k_i}\| = 0$.

Lemma 7.2 *Assume A1-A5. If $\{\rho_k\}$ is unbounded and $\lim_{k_i \to \infty} \|C_{k_i}\| = 0$, where $\{k_i\}$ is the sequence of iterates at which the penalty parameter is increased, then the iteration sequence satisfies*

$$\lim_{k \to \infty} \|\nabla C_k C_k\| = 0. \tag{7.2}$$

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

From Lemma 6.5, 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.

From Lemma 7.1, $\lim_{k_i \to \infty} \|\nabla C_{k_i} C_{k_i}\| = 0$, where $\{k_i\}$ is the subsequence of the iteration sequence at which the penalty parameter is increased. Therefore, for k sufficiently large, there are no common elements between the two sequences $\{k_i\}$ and $\{k_j\}$. For all $\hat{k} \in \{k_j\}$, using (6.2) and Lemma 6.9, we have

$$\frac{Ared_{\hat{k}}}{\rho_{\hat{k}}} \geq \eta_1 \frac{Pred_{\hat{k}}}{\rho_{\hat{k}}} \geq \eta_1 \frac{\varepsilon K_1}{4} \min[\frac{\varepsilon}{2}, \delta_{\hat{k}}] \geq \eta_1 \frac{\varepsilon K_1}{4} \min[\frac{\varepsilon}{2}, \bar{K}_6],$$

where $\bar{K}_6$ is as K_6 in (6.9) with ε_0 is replaced by $\frac{\varepsilon}{2}$. Hence, we have

$$\frac{\ell_{\hat{k}} - \ell_{\hat{k}+1}}{\rho_{\hat{k}}} + \|C_{\hat{k}}\|^2 - \|C_{\hat{k}+1}\|^2 \geq \eta_1 \frac{\varepsilon K_1}{4} \min[\frac{\varepsilon}{2}, \bar{K}_6] > 0. \tag{7.3}$$

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

$$\frac{\ell_k - \ell_{k+1}}{\rho_k} + \|C_k\|^2 - \|C_{k+1}\|^2 \geq 0. \tag{7.4}$$

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 (7.3) and (7.4) , we can write

$$\sum_{k=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\{\ell_k - \ell_{k+1}\}}{\rho_k} + \|C_{k_{\hat{i}}}\|^2 - \|C_{\hat{i}+1}\|^2 \geq \eta_1 \frac{\varepsilon K_1}{4} \min[\frac{\varepsilon}{2}, \bar{K}_6] > 0.$$

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

$$\frac{\ell_{k_{\hat{i}}} - \ell_{k_{\hat{i}+1}}}{\rho_{k_{\hat{i}}}} + \|C_{k_{\hat{i}}}\|^2 - \|C_{k_{\hat{i}+1}}\|^2 \geq \eta_1 \frac{\varepsilon K_1}{4} \min[\frac{\varepsilon}{2}, \bar{K}_6].$$

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

$$\|C_{k_{\hat{i}}}\|^2 - \|C_{k_{\hat{i}+1}}\|^2 \geq \eta_1 \frac{\varepsilon K_1}{8} \min[\frac{\varepsilon}{2}, \bar{K}_6] > 0.$$

This contradicts the assumption that $\lim_{k_i \to \infty} \|C_{k_i}\| = 0$. The supposition is wrong. This proves the lemma.
□

When $\{\rho_k\}$ is bounded, we have the following result.

Lemma 7.3 *Assume A1-A4. If $\{\rho_k\}$ is bounded then the sequence of iterates generated by the algorithm satisfies*

$$\lim_{k \to \infty} \|\nabla C_k C_k\| = 0. \tag{7.5}$$

Proof: The proof is by contradiction. Suppose that $\limsup_{k \to \infty} \|\nabla C_k C_k\| \geq \varepsilon_0 > 0$. This implies the existence of an infinite subsequence of indices $\{k_j\}$ indexing iterates that satisfy $\|\nabla C_k C_k\| \geq \frac{\varepsilon_0}{2}$, for all $k \in \{k_j\}$.

From Lemma 6.5, there exists an infinite sequence of acceptable steps. Without loss of generality, we let all elements of the sequence $\{k_j\}$ be acceptable iterates.

Since $\{\rho_k\}$ is bounded, there exist an integer $\bar{k}$ and a positive constant $\bar{\rho}$ such that for all $k \geq \bar{k}$, $\rho_k = \bar{\rho}$. Using the general assumptions, this fact implies that $\{\Phi_k\}$ is bounded.

From (6.2) and Lemma 6.9, we have for all $k_j \geq \bar{k}$

$$Pred_{k_j} \geq \frac{K_1 \bar{\rho} \varepsilon_0}{4} \min\{\frac{\varepsilon_0}{2}, \hat{K}_6\} > 0, \tag{7.6}$$

where $\hat{K}_6$ is as K_6 in (6.9) with ε_0 is replaced by $\frac{\varepsilon_0}{2}$. Using the fact that the steps indexed by any member of the sequence $\{k_j\}$ are acceptable, we have

$$\Phi_{k_j} - \Phi_{k_j+1} = Ared_{k_j} \geq \eta_1 Pred_{k_j} \geq \eta_1 \frac{K_1 \bar{\rho} \varepsilon_0}{4} \min\{\frac{\varepsilon_0}{2}, \hat{K}_6\} > 0. \tag{7.7}$$

Since $\{\Phi_k\}$ is bounded below, a contradiction arises if we let k_j go to infinity. This proves the lemma. □

The following theorem proves that if $\limsup_{k \to \infty} \|C_k\| > 0$, then the iteration sequence has a subsequence that satisfies the infeasible Mayer-Bliss conditions in the limit.

Theorem 7.4 *Assume A1-A4. If $\limsup_{k \to \infty} \|C_k\| > 0$ then the iteration sequence has a subsequence that satisfies the infeasible Mayer-Bliss conditions in the limit.*

Proof: Consider first the case when $\{\rho_k\}$ is unbounded. From Lemma 7.1, we have $\lim_{k_i \to \infty} \|\nabla C_{k_i} C_{k_i}\| = 0$, where $\{k_i\}$ is the sequence of iterates at which the penalty parameter is increased.

If $\limsup_{k_i \to \infty} \|C_{k_i}\| > 0$, then there exists a subsequence of the sequence $\{k_i\}$ that satisfies the infeasible Mayer-Bliss conditions in the limit.

Now assume that $\lim_{k_i \to \infty} \|C_{k_i}\| = 0$. Then from Lemma 7.2, we have $\lim_{k \to \infty} \|\nabla C_k C_k\| = 0$. On the other hand, because $\limsup_{k \to \infty} \|C_k\| > 0$, there exists a subsequence of the iteration sequence that satisfies the infeasible Mayer-Bliss conditions in the limit.

Now, consider the case when $\{\rho_k\}$ is bounded. From Lemma 7.3, we have $\lim_{k \to \infty} \|\nabla C_k C_k\| = 0$. This limit and the assumption that $\limsup_{k \to \infty} \|C_k\| > 0$ imply the existence of a subsequence of the iteration sequence that satisfies the infeasible Mayer-Bliss conditions in the limit. This completes the proof. □

8 Stationary Conditions

In this section, we answer the following questions. Does the iteration sequence have a subsequence that satisfies the Mayer-Bliss conditions in the limit? If yes, can we identify it? Does the iteration sequence have a subsequence that satisfies the first-order conditions in the limit? If yes, can we identify it? To answer these questions, we need the following three technical lemmas.

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

Lemma 8.1 *Assume A1-A5. Then the predicted decrease in the merit function satisfies*

$$\begin{aligned} Pred_k \quad \geq \quad & K_2 \|W_k^T \nabla q_k(s_k^n)\| \min\{\|W_k^T \nabla q_k(s_k^n)\| \ , \ \delta_k\} \\ & - K_8 \max\{\|C_k\|, \|s_k^{mn}\|\} + \rho_k [\|C_k\|^2 - \|C_k + \nabla C_k^T s_k\|^2], \end{aligned}$$

where K_2 is as in Lemma 6.2 and K_8 is a positive constant independent of k.

Proof: The proof is similar to the proof of Lemma 7.6 of Dennis, El-Alem, and Maciel (1997)[12] with (2.2) is used instead of (2.1). □

Lemma 8.2 *Assume A1-A5. If at a given iteration k^i, $\|W_k^T \nabla f_k\| \geq \varepsilon_0$ and $\max\{\|C_k\|, \|s_{k^i}^{mn}\|\} \leq \beta \delta_{k^i}$ where ε_0 is a positive constant and β is given by*

$$0 < \beta \leq \min\left\{ \frac{\varepsilon_0}{2 b_1 K \delta_{\max}}, \frac{K_2 \varepsilon_0}{4 K_8} \min\{\frac{\varepsilon_0}{2\delta_{\max}}, 1\} \right\},$$

where K is as in (2.2), b_1 is as in (3.2), K_2 is as in (6.3), and K_8 is as in Lemma 8.1, then there exists a positive constant K_9 that depends on ε_0 but does not depend on k or i, such that

$$Pred_{k^i} \geq K_9 \delta_{k^i} + \rho_{k^i} \{\|C_k\|^2 - \|C_k + \nabla C_k^T s_{k^i}\|^2\}. \tag{8.1}$$

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)[12]. □

The above lemma shows that at any iteration k^i with $\|W_k^T \nabla f_k\| \geq \varepsilon_0$, if $\max\{\|C_k\|, \|s_{k^i}^{mn}\|\} \leq \beta \delta_{k^i}$, then the penalty parameter is not increased.

The following lemma bounds $\|s_k^{mn}\|$ by $\|C_k\|$ and $\|C_k\|$ by $\|\nabla C_k C_k\|$, for any iteration where the smallest singular value of ∇C_k is not zero.

Lemma 8.3 *Assume A1 and A2. If there exists a subsequence $\{k_i\}$ of the iteration sequence such that $\{\sigma_{k_i}\}$ is bounded away from zero, then all trial iterates j of any iteration $k \in \{k_i\}$ satisfy*

$$\|s_{kj}^{mn}\| \leq K_{10} \|C_k\|, \tag{8.2}$$

and for any $k \in \{k_i\}$

$$\|C_k\| \leq K_{11} \|\nabla C_k C_k\|, \tag{8.3}$$

where K_{10} and K_{11} are two positive constants that do not depend on k or j.

Proof: The proof of (8.2) is similar to the proof of Lemma 7.1 of Dennis, El-Alem, and Maciel (1997)[12]. The proof of (8.2) follows from the fact that for all $k \in \{k_i\}$, $\|C_k\| \leq \|(\nabla C_k^T \nabla C_k)^{-1} \nabla C_k^T\| \|\nabla C_k C_k\|$, followed by the use of the assumptions. $\square$

From the above two lemmas, if for the subsequence $\{k_i\}$ of the iteration sequence at which the penalty parameter is increased, $\{\sigma_{k_i}\}$ is bounded away from zero, and $\|W_{k_i}^T \nabla f_{k_i}\| \geq \varepsilon_0$ for all $k \in \{k_i\}$, then

$$\|C_k\| > \beta_1 \delta_k \tag{8.4}$$

holds for all $k \in \{k_i\}$, where $\beta_1 = \frac{\beta}{\max\{1, K_{10}\}}$, β is as in Lemma 8.2, and K_{10} is as in (8.2).

From (6.6), (8.2), and (8.3), if $\{k_i\}$ is the sequence of iterates at which the penalty parameter is increased and $\{\sigma_{k_i}\}$ is bounded away from zero, then we have for all $k \in \{k_i\}$,

$$\rho_k \|C_k\| \leq K_{12}, \tag{8.5}$$

where K_{12} is a positive constant independent of k.

The following theorem studies the behavior of the iteration sequence when $\{\|C_k\|\}$ converges to zero and $\{\rho_k\}$ is unbounded.

Theorem 8.4 *Assume A1-A5. Assume also that $\{\rho_k\}$ is unbounded and $\{\|C_k\|\}$ converges to zero. The iteration sequence at which ρ_k is increased has a subsequence that satisfies either the feasible Mayer-Bliss conditions or the first-order conditions in the limit.*

Proof: Let $\{k_j\}$ be the iteration sequence at which ρ_k is increased. Since $\lim_{k_j \to \infty} \|C_{k_j}\| = 0$, then if there exists a subsequence of the sequence $\{k_j\}$ where the sequence of smallest singular values of ∇C_{k_j} converges to zero. Then it satisfies the feasible Mayer-Bliss conditions in the limit and the proof ends here.

Consider the case where $\{\sigma_{k_j}\}$ is bounded away from zero. Suppose that, for all $k \in \{k_j\}$,

$$\|W_k \nabla f_k\| \geq \varepsilon_0. \tag{8.6}$$

From (8.4), we have $\|C_k\| > \beta_1 \delta_k$, for all $k \in \{k_j\}$. But because $\lim_{k_j \to \infty} \|C_{k_j}\| = 0$, we have

$$\lim_{k_j \to \infty} \delta_{k_j^i} = 0. \tag{8.7}$$

The rest of the proof is by contradiction. From the way of updating the trust-region radius, $\delta_{k_j^1} \geq \delta_{\min}$. Therefore, the superscript $i \neq 1$ in (8.7). Because $\delta_{k_j^1} \geq \delta_{\min}$ and both of $\delta_{k_j^i}$ and C_{k_j} are converging to zero, then for k_j sufficiently large, there must be an $m > 1$ such that $\|C_{k_j}\| > \beta_1 \delta_{k_j^m}$ and $\|C_{k_j}\| \leq \beta_1 \delta_{k_j^{m-1}}$, where β_1 is as in (8.4). Using $\delta_{k_j^m} = \alpha_1 \|s_{k_j^{m-1}}\|$ and (8.5), we have

$$\rho_{k_j^{m-1}} \|s_{k_j^{m-1}}\| \leq \rho_{k_j^m} \frac{\delta_{k_j^m}}{\alpha_1} \leq \rho_{k_j^m} \frac{\|C_{k_j}\|}{\alpha_1 \beta_1} \leq \frac{K_{12}}{\alpha_1 \beta_1}.$$

From Lemma 6.3 and the above inequality, we have

$$\begin{aligned} \left| Ared_{k_j^{m-1}} - Pred_{k_j^{m-1}} \right| &\leq K_3 [1 + (1 + \beta_1)\ \rho_{k_j^{m-1}}\ \|s_{k_j^{m-1}}\|\ ]\ \|s_{k_j^{m-1}}\|\ \delta_{k_j^{m-1}}, \\ &\leq K_3\ [\ 1 + (1 + \beta_1) \frac{K_{12}}{\alpha_1 \beta_1}\ ]\ \|s_{k_j^{m-1}}\|\ \delta_{k_j^{m-1}}. \end{aligned}$$

Also $||C_{k_j}|| \le \beta_1 \delta_{k_j^{m-1}}$ implies that $\max\{||C_{k_j}||, ||s^{mn}_{k_j^{m-1}}||\} \le \beta \delta_{k_j^{m-1}}$. Hence, from Lemma 8.2, we have

$$Pred_{k_j^{m-1}} \ge K_9 \delta_{k_j^{m-1}}.$$

Therefore, since $s_{k_j^{m-1}}$ was a rejected step,

$$(1-\eta_1) < \frac{|Ared_{k_j^{m-1}} - Pred_{k_j^{m-1}}|}{Pred_{k_j^{m-1}}} \quad \le \quad \frac{K_3[1+(1+\beta_1)\frac{K_{12}}{\alpha_1\beta_1}]||s_{k_j^{m-1}}||}{K_9}.$$

Hence, $||s_{k_j^{m-1}}|| > \frac{K_9(1-\eta_1)}{K_3[1+(1+\beta_1)\frac{K_{12}}{\alpha_1\beta_1}]}$ and we obtain

$$\delta_{k_j^m} \ge \alpha_1 ||s_{k_j^{m-1}}|| \ge \frac{\alpha_1 K_9(1-\eta_1)}{K_3[1+(1+\beta_1)\frac{K_{12}}{\alpha_1\beta_1}]}.$$

This means that $\delta_{k_j^m}$ is bounded below. Hence $\{||C_{k_j}||\}$ is bounded away from zero. This contradicts the assumption that $\{||C_k||\}$ converges to zero and means that for k_j sufficiently large there is no m such that $||C_{k_j}|| > \beta_1 \delta_{k_j^m}$ holds. Hence, all trial iterates i of k_j satisfy $||C_{k_j}|| \le \beta_1 \delta_{k_j^i}$. But this contradicts the fact that k_j is an iterate at which $\rho_{k_j^i}$, for some trial i, is increased. This contradiction implies that the supposition (8.6) was wrong and completes the proof of the theorem. □

From the above lemma, we conclude that, if along the subsequence of the iteration sequence at which ρ_k is increased, the corresponding subsequence of σ_k converges to zero, then it has a subsequence that asymptotically satisfies the feasible Mayer-Bliss conditions. Otherwise, it has a subsequence that satisfies the first-order conditions in the limit.

When $\{\rho_k\}$ is bounded, there must exist a positive integer $\bar{k}$ and a positive constant $\bar{\rho}$ such that for all $k \ge \bar{k}$, $\rho_k = \bar{\rho}$. Without loss of generality we will take $\rho_k = \bar{\rho}$ for all k, whenever we assume that $\{\rho_k\}$ is bounded.

To study the case when $\{\rho_k\}$ is bounded, we need the following lemma which is similar to Lemma 6.7.

Lemma 8.5 *Assume A1-A4. Let $\{\rho_k\}$ be bounded. If at the j^{th} trial iterate of iteration k, there exists a positive constant K_{13} independent of k and j such that*

$$Pred_{k^j} \ge K_{13} \delta_{k^j} \tag{8.8}$$

and if

$$||s_{k^j}|| \le \frac{(1-\eta_1)K_{13}}{2K_4\bar{\rho}},$$

where K_4 is as in (6.5), then the step s_{k^j} must be accepted.

Proof: The proof is similar to the proof of Lemma 6.7. □

From the above lemma we can easily conclude that if $\{\rho_k\}$ is bounded and if at a given iteration k all its trial iterates satisfy (8.8), then an acceptable step must be found after a finite number of trials.

The following theorem studies the asymptotic behavior of the iteration sequence when $\{\rho_k\}$ is bounded.

Theorem 8.6 *Assume A1-A5. Assume also that $\{\rho_k\}$ is bounded and $\lim_{k\to\infty} ||C_k|| = 0$. Then the iteration sequence has a subsequence that satisfies either the feasible Mayer-Bliss conditions or the first-order conditions in the limit.*

Proof: Assume that there exists an infinite subsequence $\{k_j\}$ such that $\max\{\|C_{k_j}\|, \|s^{mn}_{k^i_j}\|\} \geq \beta\delta_{k^i_j}$, for some trial iterate i of k_j and some $\beta > 0$. We take β as in Lemma 8.2. If $\{\sigma_{k_j}\}$ is not bounded away from zero, then there exists a subsequence that satisfies the feasible Mayer-Bliss conditions in the limit.

Let us assume that $\{\sigma_{k_j}\}$ is bounded away from zero and suppose that $\|W_k \nabla f_k\| \geq \varepsilon_0 > 0$ for all $k \in \{k_j\}$.

Because $\delta_{k^1_j} \geq \delta_{\min}$ and as $k_j \to \infty$, both of $\|s^{mn}_{k^i_j}\|$ and $\|C_{k_j}\|$ converge to zero, then for k sufficiently large, there must be an $m > 1$ such that $\max\{\|C_{k_j}\|, \|s^{mn}_{k^m_j}\|\} > \beta\delta_{k^m_j}$ and $\max\{\|C_{k_j}\|, \|s^{mn}_{k^{m-1}_j}\|\} \leq \beta\delta_{k^{m-1}_j}$, where β is as in Lemmas 8.2. Note that $\|s^{mn}_{k^i_j}\| \leq K_{10}\|C_{k_j}\|$, for all trial iterates i of iteration k_j, where $k_j \in \{k_j\}$.

From Lemma 8.2, we have $Pred_{k^{m-1}_j} \geq K_9\delta_{k^{m-1}_j}$. Because $s_{k^{m-1}_j}$ is a rejected step, we have using (6.5),

$$(1-\eta_1) < \frac{|Ared_{k^{m-1}_j} - Pred_{k^{m-1}_j}|}{Pred_{k^{m-1}_j}} \quad \leq \quad \frac{K_4\bar{\rho}\|s_{k^{m-1}_j}\|}{K_9}.$$

Hence, $\|s_{k^{m-1}_j}\| > \frac{K_9(1-\eta_1)}{K_4\bar{\rho}}$. Therefore, $\delta_{k^m_j} = \alpha_1\|s_{k^{m-1}_j}\| > \frac{\alpha_1 K_9(1-\eta_1)}{K_4\bar{\rho}}$.

This means that $\delta_{k^m_j}$ is bounded away from zero. This contradicts the assumption that $\{\|C_{k_j}\|\}$ converges to zero and means that there is no m such that $\max\{\|C_{k_j}\|, \|s^{mn}_{k^m_j}\|\} > \beta\delta_{k^m_j}$.

Hence, all trial iterates i of k_j satisfy $\max\{\|C_{k_j}\|, \|s^{mn}_{k^i_j}\|\} \leq \beta\delta_{k^i_j}$. But this implies, using Lemma 8.2 that $Pred_{k^i} \geq K_9\delta_{k^i}$, for all trial iterates i of any iteration $k \in \{k_j\}$. From the above lemma, there must be an infinite sequence of acceptable trial iterates $\{k^l_j\}$. For all acceptable iterates of $\{k^l_j\}$, we have

$$\Phi_{k^l_j} - \Phi_{k^l_j+1} = Ared_{k^l_j} \geq \eta_1 Pred_{k^l_j} \geq \eta_1 K_9 \delta_{k^l_j}.$$

If we take the limit as $k \to \infty$, we obtain

$$\lim_{k\to\infty} \delta_{k^l_j} = 0. \tag{8.9}$$

We show a contradiction by proving that $\{\delta_{k^l_j}\}$ is bounded. Using argument similar to that of Lemma 8.5, we conclude that, the trial iterate k^{l-1}_j must satisfy $\|s_{k^{l-1}_j}\| > \frac{(1-\eta_1)K_9}{2K_4\bar{\rho}}$. From the way of updating the trust region radius, we have

$$\delta_{k^l_j} = \alpha_1\|s_{k^{l-1}_j}\| > \frac{\alpha_1(1-\eta_1)K_9}{2K_4\bar{\rho}}.$$

This implies that $\{\delta_{k^l_j}\}$ is bounded. Therefore, the supposition is wrong and we have

$$\lim_{k_j\to\infty} \|W^T_{k_j}\nabla f_{k_j}\| \;=\; 0.$$

This completes the proof of the theorem. □

Let us again state and then prove our main global convergence result, Theorem 5.1.

Theorem 5.1 *Under assumptions A1-A5, the algorithm produces a sequence of iterates that has a subsequence that satisfies either the infeasible Mayer-Bliss conditions, the feasible Mayer-Bliss conditions, the infeasible first-order conditions, or the first-order conditions.*

Proof: The proof follows immediately from Theorems 7.4, 8.4, and 8.6. □

9 Summary and Concluding Remarks

We have established a global convergence theory for the class of trust-region-based algorithms suggested by Dennis, El-Alem, and Maciel [12]. This class of algorithm is characterized by generating steps such that their quasi-normal components satisfy a fraction of Cauchy decrease condition on the quadratic model of the linearized constraints. Furthermore, their tangential components satisfy a fraction of Cauchy decrease condition on the quadratic model of the Lagrangian function associated with the problem, reduced to the tangent space of the constraints. The augmented Lagrangian is used as a merit function. For updating the penalty parameter, a scheme proposed by the author [15] was used.

Because the two components of the trial step are not necessarily orthogonal, an additional condition on the length of the normal component is needed to prove global convergence. Dennis , El-Alem, and Maciel (1997)[12] suggested the condition $||s_k^n|| \leq K||C_k||$. In this paper, we used $||s_k^n|| \leq K||s_k^{mn}||$, where s_k^{mn} is the minimum-norm solution that minimizes $||C_k + \nabla C_k^T s||$ inside the trust region δ_k. This condition is equivalent to the above condition whenever ∇C_k has full column rank and allows the full SQP step to be taken when it is inside the trust region.

As pointed out in Section 2.3, if at a given iteration k the algorithm generates an infeasible point with $||C_k||^2 - ||C_k + \nabla C_k s_k||^2 = 0$, then it may not be able to move away from that point. We pointed out in Lemma 4.6 that in this case the point is necessarily an infeasible Mayer-Bliss point. Probably, if a good estimate of the Lagrange multiplier vector is used every iteration, or at least at this point, then the algorithm moves away from such points. Avoiding Mayer-Bliss points that are not first-order points is an important issue for algorithms that are designed to handle the lack of linear independence in the gradients of the constraints. This issue indeed deserves to be studied.

Because we do not assume that the columns of ∇C_k are linearly independent, the iteration sequence may have no subsequence that asymptotically satisfies the first order conditions (see Section 4). In other words, it may be the case that no iterate k generated by the algorithm satisfies $||W_k \nabla f_k|| + ||\nabla C_k C_k|| \leq \varepsilon$. Therefore, another condition for terminating the algorithm should be added. For example, the following condition can be tested at the end of Step 2 of the algorithm. If $||s_k|| \leq \varepsilon_{tol}$ then terminate.

The main feature of the global convergence theory presented in this paper is that the gradients of the constraints are allowed to be linearly dependent. We showed that under the general assumptions of Section 3 and without the regularity assumption the iteration sequence has a subsequence that asymptotically satisfies one of four types of stationary conditions. In particular, it asymptotically satisfies either the infeasible Mayer-Bliss conditions, the feasible Mayer-Bliss conditions, the infeasible first-order conditions, or the first-order conditions.

In our theory, we used the assumption that the Lagrange multiplier vector is bounded. Such a reasonable assumption is justified as follows. In practice, if a scheme like the projection formula, for instance, is used for computing the Lagrange multiplier and if the matrix ∇C_k is poorly conditioned, the vector λ_k may contain very large numbers. In order to avoid large numbers in certain calculations, the vector λ should be normalized using a scale factor $\omega = \frac{1}{\max\{1, ||\lambda||\}}$ and ω scales f and ∇f in all expressions where λ appears. Of course, the theory will require that the sequence of scale factors $\{\omega_k\}$ be bounded away from zero. Fortunately, this is the case unless a subsequence of the iteration sequence asymptotically satisfies either the feasible or the infeasible Mayer-Bliss conditions.

References

[1] N. ALEXANDROV. Multi-level algorithms for nonlinear equations and equality constrained optimization. Technical Report 93-20, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251-1892, May 1993.

[2] N. ALEXANDROV. *Multi-level Algorithms for Nonlinear Equations and Equality Constrained Optimization.* PhD thesis, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251-1892, 1993.

[3] N. ALEXANDROV and J. E. DENNIS Jr. Multi-level algorithms for nonlinear optimization. Technical Report 94-24, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251-1892, June 1994.

[4] L. T. BIEGLER, J. NOCEDAL, and C. SCHMID. A reduced Hessian method for large-scale constrained optimization. *SIAM J. Optim.*, 5 No. 2:314–347, May 1995.

[5] G. A. BLISS. Normality and abnormality in the calculus of variations. *Trans. Amer. Maht. Soc.*, 43:365–376, 1938.

[6] J. V. BURKE. An exact penelization viewpoint of constrained optimization. *SIAM J. Control and Opt.*, 29:968–998, 1991.

[7] J. V. BURKE. A robust trust region method for constrained nonlinear programming problems. *SIAM J. Optimization*, 2:325–347, 1992.

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

[9] R. H. BYRD, R. B. SCHNABEL, and G. A. SHULTZ. A trust region algorithm for nonlinearly constrained optimization. *SIAM J. Numer. Anal.*, 24:1152–1170, 1987.

[10] R. CARTER. *Multy-model algorithms for optimization.* PhD thesis, Department of Mathematical Sciences, Rice University, Houston, TX 77251-1892, 1986.

[11] M. R. CELIS, J. E. DENNIS Jr., and R. A. TAPIA. A trust region strategy for nonlinear equality constrained optimization. In *Numerical Optimization 1984*. SIAM, Philadelphia, Pennsylvania, 1985.

[12] 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. Optim.*, 7 No. 1:177–207, February 1997.

[13] J. E. DENNIS Jr. and R. B. SCHNABEL. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, New Jersey, 1983. Russian edition, Mir Publishing Office, Moscow, 1988, O. Burdakov, translator.

[14] J. E. DENNIS Jr. and L. VICENTE. On the convergence theory of general trust-region-based algorithms for equality-constrained optimization. Technical Report TR94-36, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, 1994. Revised September 1995.

[15] 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 77251-1892, 1988.

[16] 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, 1991.

[17] M. M. EL-ALEM. A robust trust region algorithm with a nonmonotonic penalty parameter scheme for constrained optimization. *SIAM J. Optim.*, 5 No. 2:348–378, May 1995.

[18] M. M. EL-ALEM. Convergence to a second order point of a trust-region algorithm with a nonmonotonic penalty parameter scheme for constrained optimization. *J. Optim. Theory Appl.*, 91 No. 1:61–79, Oct. 1996.

[19] M. M. EL-ALEM and R. A. TAPIA. Numerical experience with a polyhedral norm CDT trust-region algorithm. *J. Optim. Theory Appl.*, 85(1):575–592, June 1995.

[20] M. EL-HALLABI. A global convergence theory for arbitrary norm trust-region algorithms for equality constrained optimization. Technical Report TR93-60, Department of Computational and Applied Mathematics, Rice University, Houston, Texas 77251-1892, 1993.

[21] R. FLETCHER. *Practical Methods of Optimization.* John Wiley & Sons, New York, 1987.

[22] M. LALEE. *Algorithms for nonlinear optimization.* PhD thesis, Department of Industrial Engineering, Northwestern University, Evanston, IL, 1993.

[23] 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.

[24] K. LEVENBERG. A method for the solution of certain problems in least squares. *Quart. Appl. Math.*, 2:164–168, 1944.

[25] 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, TX 77251-1892, 1992.

[26] O. L. MANGASARIAN. *Nonlinear Programming.* McGraw-Hill Book Company, New York, 1969.

[27] D. MARQUARDT. An algorithm for least-squares estimation of nonlinear parameters. *SIAM J. Appl. Math.*, 11:431–441, 1963.

[28] A. MAYER. Begrudung der lagrange'schen multiplicatorenmethode in der variationsrechnung. *Math. Annal.*, 26:74–82, 1886.

[29] A. MIELE, E. E. CRAGG, and A. V. LEVY. Use of the augmented penalty function in mathematical programming problems: Part 2. *J. Optim. Theory Appl.*, 8(2):131–153, 1971.

[30] 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. Theory Appl.*, 4(4):213–243, 1969.

[31] A. MIELE, A. V. LEVY, and E. E. CRAGG. Modifications and extensions of the conjugate gradient-restoration algorithm for mathematical programming problems. *J. Optim. Theory Appl.*, 7(6), 1971.

[32] J. J. MORE. The Levenberg-Marquardt algorithm: implementation and theory. In G. A. Watson, editor, *Numerical Analysis. Lecture notes in Mathematics 630*, pages 105–116. Springer-Verlag, Berlin, New York, 1977.

[33] J. J. MORE. Recent developments in algorithms and software for trust region methods. In A. Bachem, M. Grotschel, and B. Korte, editors, *Mathematical programming. The state of the art*, pages 258–287. Springer-Verlag, New York, 1983.

[34] 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.

[35] T. PLANTENGA. *Large-Scale nonlinear constrained optimization using trust regions.* PhD thesis, Department of Electrical and Computer Engineering, Northwestern University, Evanston, IL, 1995.

[36] M. J. D. POWELL. A new algorithm for unconstrained optimization. In J.B. Rosen, O.L. Mangasarian, and K. Ritter, editors, *Nonlinear Programming*, pages 31–65. Academic Press, New York, 1970.

[37] 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.

[38] M. J. D. POWELL and Y. YUAN. A trust region algorithm for equality constrained optimization. *Math. Prog.*, 49:189–211, 1991.

[39] G. A. SHULTZ, R. B. SCHNABEL, and R. H. BYRD. A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties. *SIAM J. Numer. Anal.*, 22:47–67, 1985.

[40] A. VARDI. A trust region algorithm for equality constrained minimization: convergence properties and implementation. *SIAM J. Numer. Anal.*, 22:575–591, 1985.

[41] 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.

[42] K. A. WILLIAMSON. *A robust trust region algorithm for nonlinear programming.* PhD thesis, Department of Mathematical Sciences, Rice University, Houston, Texas 77251-1892, 1990.

[43] Y. YUAN. A dual algorithm for minimizing a quadratic function with two quadratic constraints. Report DAMTP-NA3, University of Cambridge, Cambridge, England, 1988.

[44] Y. YUAN. On a subproblem of trust-region algorithm for constrained optimization. *Math. Prog.*, 47:53–63, 1990.

[45] Y. YUAN. On the convergence of a new trust region algorithm. *Numer. Math.*, 70:515–539, 1995.

[46] 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.

[47] Y. ZHANG. Computing a Celis-Dennis-Tapia trust-region step for equality constrained optimization. *Math. Prog.*, 55:109–124, 1992.