A New Trust-region Algorithm for General Nonlinear Programming

Mahmoud El-Alem

Bothina El-Sobky

October 1997

TR97-25

A NEW TRUST-REGION ALGORITHM FOR GENERAL NONLINEAR PROGRAMMING

MAHMOUD EL-ALEM* AND BOTHINA EL-SOBKY*

Abstract. A new trust-region algorithm for solving the general nonlinear programming problem is introduced. In this algorithm, an active set strategy is used together with a projected Hessian technique to convert the computation of the trial step to two easy trust-region subproblems similar to those for the unconstrained case. To force global convergence, the augmented Lagrangian for general nonlinear programming is used as a merit function. A convergence theory for this algorithm is presented. Under reasonable assumptions, it is shown that the algorithm is globally convergent.

Key Words : Constrained optimization, nonlinear programming, global convergence, active set, trust region, stationary points, KKT conditions, Fritz John points.

AMS subject classifications. 65K05, 49D37.

1. Introduction. In this paper, we introduce a new trust-region algorithm for solving the general nonlinear programming problem

$$\begin{array}{lrcl} \textit{minimize} & f(x) & & \\ \textit{subject to} & h(x) & = & 0, \\ & g(x) & \leq & 0, \end{array} \tag{1.1}$$

where $h = (h_1, h_2, ..., h_m)^T$ and $g = (g_1, g_2, ..., g_p)^T$. The functions $f : \Re^n \rightarrow \Re$, $h : \Re^n \rightarrow \Re^m$, and $g : \Re^n \rightarrow \Re^p$ are twice continuously differentiable. We assume that $m < n$ and no restriction is assumed on p.

The Lagrangian function $l : \Re^n \times \Re^m \times \Re^p \rightarrow \Re$ associated with Problem (1.1) is the function

$$l(x, \lambda, \mu) = f(x) + \lambda^T h(x) + \mu^T g(x), \tag{1.2}$$

where $\lambda = (\lambda_1, ..., \lambda_m)^T$ and $\mu = (\mu_1, ..., \mu_p)^T$ are the Lagrange multiplier vectors.

Over the last two decades, trust-region algorithms have proven to be robust techniques for solving optimization problems. Their high regard is due to the strong global convergence properties that they possess [20] and due to the existence of reliable, well developed, and efficient software [17].

Since mid eighties, many authors have considered trust-region algorithms for solving the equality constrained optimization problem

$$\begin{array}{ll} \textit{minimize} & f(x) \\ \textit{subject to} & h(x) = 0. \end{array}$$

Most trust-region algorithms for solving this problem try to combine the trust-region idea with the successive quadratic programming (SQP) method [24]. The SQP method iteratively solves a quadratic programming subproblem that consists of minimizing a quadratic model of the Lagrangian function $\ell(x, \lambda) = f(x) + \lambda^T h(x)$ subject to satisfying a linear approximation of the constraints.

If a trust-region constraint is simply added to the quadratic programming subproblem the resulting trust-region subproblem may be infeasible because there may be no intersecting points between the trust-region constraint and the hyperplane of

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

the linearized constraints [4]. Even if they intersect, there is no guarantee that this will remain true if the trust-region radius is decreased.

The reduced Hessian is a successful approach to overcoming the difficulty of having an infeasible trust-region subproblem. The approach was suggested by Byrd (1987)[2] and Omojokun (1989)[18]. In this approach, the trial step is decomposed into two orthogonal components; the tangential component and the normal component. Each component is computed by solving a trust-region subproblem. The two subproblems are similar to the trust-region subproblem for the unconstrained case.

In this paper, we propose an algorithm for solving Problem (1.1). The trial steps are computed using the projected Hessian technique in the tradition of numerous works on equality constrained optimization (see, for example, Alexandrov and Dennis (1994)[1], Dennis, El-Alem, and Maciel (1997)[5], El-Alem (1996)[11] Omojokun (1989)[18], Plantenga (1995)[19], and Zhang and Zhu (1990)[27]).

The proposed algorithm uses active-set strategy. The chief feature of the proposed active set is that the active set is identified and updated naturally by the trial step.

Many authors, including Gill, Murray, Saunders, and Wright (1986)[14], Han (1977)[15], Powell (1978)[21], and Schittkowski (1983)[22], have considered active set techniques for extending the SQP method to handle Problem (1.1). By using a backtracking line-search they were able to prove global convergence. None of these theories are based on a trust-region globalization strategy. Omojokun (1989)[18] proposed a trust-region active-set algorithm for solving Problem (1.1). Unfortunately, there is no global convergence theory for that algorithm. Furthermore, the active set strategy proposed here is much simpler than that proposed by Omojokun. Yuan (1995)[26] proposed a trust-region algorithm that uses an active set technique similar to the one used by our algorithm. However, he employed a non-differentiable merit function.

Motivated by the impressive computational performance of primal-dual interior point algorithms for linear programming [25], El-Bakry, Tapia, Tsuchiya, and Zhang (1996)[12] proposed an interior-point method for solving Problem (1.1). By adding slack variables, Problem (1.1) is transformed to equality constraints problem with non-negativity condition on the slack variables. By using a backtracking line-search they proved several global and local convergence results.

Using a barrier technique, Byrd, Gilbert, and Nocedal (1996)[3] proposed a trust-region algorithm for solving Problem (1.1). Their algorithm can be viewed as a trust-region version of the above algorithm. They proved global convergence for their algorithm.

Dennis, Heinkenschloss, and Vicente (1995)[7] proposed a trust-region interior-point algorithm for a class of Problem (1.1) having a special structure. They proved several global and local convergence results for their algorithm.

Since in interior point methods, the strict feasibility with respect to the bounds must be maintained every iteration, the trial step may need to be truncated. However, if some component of the iterate is too close to a bound that is not active at the solution and if the corresponding component of the step violates that bound, it could happen that the truncating parameter be very small, resulting in very short steps, taking too many steps from the algorithm to move away from wrong bounds.

The algorithm proposed in this paper, however, has no feasibility restriction to be satisfied at every iteration and therefore has no truncating parameter. Hence, the full step is taken once it produces a sufficient decrease.

Let $I(x)$ be the set of indices of violated or binding inequality constraints at a point x. *i.e.*, $I(x) = \{i : g_i(x) \geq 0\}$.

A point x_* is a regular point for Problem (1.1) if the vectors in the set $\{\nabla h_i(x_*), i = 1, 2, ..., m\} \bigcup \{\nabla g_i(x_*), i \in I(x_*)\}$ are linearly independent.

Let x_* be a regular point. The first-order necessary conditions (or the KKT conditions) for the point x_* to be a stationary point of Problem (1.1) are the existence of Lagrange multipliers $\lambda_* \in \Re^m$ and $\mu_* \in \Re^p$ such that

$$\begin{aligned} &(1.3) & \nabla f(x_*) + \nabla h(x_*)\lambda_* + \nabla g(x_*)\mu_* &= 0, \\ &(1.4) & h(x_*) &= 0, \\ &(1.5) & g(x_*) &\leq 0, \\ &(1.6) & (\mu_*)_i g_i(x_*) &= 0, \quad i = 1, ..., p, \\ &(1.7) & (\mu_*)_i &\geq 0, \quad i = 1, ..., p. \end{aligned}$$

Here we used the notations $\nabla h(x)$ and $\nabla g(x)$ for the matrices whose columns are $\nabla h_i(x)$, $i = 1, 2, ..., m$ and $\nabla g_i(x)$, $i = 1, 2, ..., p$, respectively. For a detailed discussion of optimality conditions, see Fiacco and McCormick (1968)[13].

Following Dennis, El-Alem, and Williamson (1996)[6], we define a 0-1 diagonal indicator matrix $W(x) \in \Re^{p \times p}$, whose diagonal entries are

$$(1.8) \qquad w_i(x) = \begin{cases} 1 & \text{if } g_i(x) \geq 0, \\ 0 & \text{if } g_i(x) < 0. \end{cases}$$

Our approach is to transform the general nonlinear programming problem (1.1) to the following equality constrained optimization problem

$$(1.9) \qquad \begin{array}{ll} \textit{minimize} & f(x) + \mu^T g(x) + \frac{\rho}{2}\|W(x)g(x)\|_2^2, \\ \textit{subject to} & h(x) = 0, \end{array}$$

where $\mu \in \Re^p$ is the Lagrange multiplier vector corresponding to $g(x)$ and ρ is a positive parameter.

The Lagrangian function associated with Problem (1.9) is given by

$$(1.10) \qquad L(x, \lambda, \mu; \rho) = f(x) + \lambda^T h(x) + \mu^T g(x) + \frac{\rho}{2}\|W(x)g(x)\|_2^2,$$

and the augmented Lagrangian is the function

$$(1.11) \quad \Phi(x, \lambda, \mu; \rho; r) = f(x) + \lambda^T h(x) + \mu^T g(x) + \frac{\rho}{2} g(x)^T W(x) g(x) + r h(x)^T h(x),$$

where $r > 0$ is a parameter usually called the penalty parameter.

Consider Problem (1.9) and let x_* be such that the matrix $\nabla h(x_*)$ has full column rank. The first-order necessary condition for the point x_* to be a stationary point is the existence of a Lagrange multiplier vector $\lambda_* \in \Re^m$ such that

$$\begin{aligned} &(1.12) & \nabla f(x_*) + \nabla h(x_*)\lambda_* + \nabla g(x_*)\mu_* + \rho \nabla g(x_*) W(x_*) g(x_*) &= 0, \\ &(1.13) & h(x_*) &= 0. \end{aligned}$$

It is easy to see that Equations (1.3)-(1.7) imply (1.12)-(1.13) but the converse is not true in general. Observe that when we transformed Problem (1.1) to Problem (1.9) two extra parameters appeared in the objective function. We design our trust-region algorithm such that it chooses μ and ρ in such a way that, if a point (x_*, λ_*, μ_*) satisfies (1.12)-(1.13) and x_* is regular, then it also satisfies (1.3)-(1.7).

The rest of this section introduces some notations. In Section 2, we present our trust-region algorithm. In Section 3, we present the assumptions under which we prove global convergence. In Section 4, we present definitions of other types of stationary points and study their properties. Sections 5-8 are devoted to presenting our global convergence theory. Section 9 contains concluding remarks.

Subscripted functions denote function values at particular points; for example, $f_k = f(x_k)$, $l_{k+1} = l(x_{k+1}, \lambda_{k+1}, \mu_{k+1})$, $\nabla_x l_k = \nabla_x l(x_k, \lambda_k, \mu_k)$, and so on. However, the arguments of the functions are not abbreviated when emphasizing the dependence of the functions on their arguments. We use the same symbol 0 to denote the real number zero, the zero vector, and the zero matrix. The matrix H_k denotes the Hessian of the Lagrangian function (1.2) at the point (x_k, λ_k, μ_k) or an approximation to it. The i^{th} component of a vector $V(x_k)$ is denoted by either $(V_k)_i$ or $V_i(x_k)$. Finally, all norms are l_2-norms.

2. Algorithm Outline. This section is devoted to presenting the details description of our trust-region algorithm for solving Problem (1.1). The algorithm combines ideas from Byrd (1987)[2], Omojokun (1989)[18], El-Alem (1995)[10], Dennis, El-Alem and Williamson (1994)[6], and Yuan (1995) [26].

2.1. Computing a Trial Step. The reduced Hessian approach is used to compute a trial step s_k. In this approach, the trial step s_k is decomposed into two orthogonal components; the normal component s_k^n and the tangential component s_k^t. The trial step s_k has the form $s_k = s_k^n + Z_k \bar{s}_k^t$, where Z_k is a matrix whose columns form an orthonormal basis for the null space of ∇h_k^T.

We obtain the normal component s_k^n by solving the following trust-region subproblem

$$\begin{array}{ll} \textit{minimize} & \frac{1}{2}\|\nabla h_k^T s^n + h_k\|^2 \\ \textit{subject to} & \|s^n\| \leq \zeta \delta_k, \end{array}$$

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

Let the quadratic model of the Lagrangian function (1.10) be

$$q_k(s) = l_k + \nabla_x l_k^T s + \frac{1}{2} s^T H_k s + \frac{\rho_k}{2}\|W_k(g_k + \nabla g_k^T s)\|^2. \tag{2.1}$$

Given the normal component s_k^n, we compute the tangential component $s_k^t = Z_k \bar{s}_k^t$ by solving the following trust-region subproblem

$$\begin{array}{ll} \textit{minimize} & (Z_k^T \nabla q_k(s_k^n))^T \bar{s}^t + \frac{1}{2}\bar{s}^{t^T} Z_k^T B_k Z_k \bar{s}^t \\ \textit{subject to} & \|Z_k \bar{s}^t\| \leq \Delta_k, \end{array}$$

where $\Delta_k = \sqrt{\delta_k^2 - \|s_k^n\|^2}$, $B_k = H_k + \rho_k \nabla g_k W_k \nabla g_k^T$, and $q_k(s)$ is given by (2.1).

2.2. Testing the Step and Updating δ_k. Once the trial step is computed, it needs to be tested to determine whether it will be accepted. To do that, a merit function is needed. We use the augmented Lagrangian (1.11) as a merit function.

To test the step, estimates for the two Lagrange multipliers λ_{k+1} and μ_{k+1} are needed. Our way of updating λ_k and μ_k is presented in Algorithm 2.2. Our theory requires that the sequences $\{\lambda_k\}$ and $\{\mu_k\}$ be bounded. We assume that the computed Lagrange multipliers by Algorithm 2.2 are bounded.

Let λ_{k+1} and μ_{k+1} be estimates of the two Lagrange multiplier vectors. We test whether the point $(x_k + s_k, \lambda_{k+1}, \mu_{k+1})$ will be taken as a next iterate.

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

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

This can be written as,

$$\begin{aligned} (2.2) \qquad Ared_k &= l(x_k, \lambda_k, \mu_k) - l(x_{k+1}, \lambda_k, \mu_k) - \Delta\lambda_k^T h_{k+1} - \Delta\mu_k^T g_{k+1} \\ &\quad + \frac{\rho_k}{2}[g_k^T W_k g_k - g_{k+1}^T W_{k+1} g_{k+1}] + r_k[\|h_k\|^2 - \|h_{k+1}\|^2], \end{aligned}$$

where $\Delta\lambda_k = (\lambda_{k+1} - \lambda_k)$ and $\Delta\mu_k = (\mu_{k+1} - \mu_k)$.

The predicted reduction in the merit function is defined to be

$$\begin{aligned} (2.3) \quad Pred_k &= -\nabla_x l(x_k, \lambda_k, \mu_k)^T s_k - \frac{1}{2} s_k^T H_k s_k - \Delta\lambda_k^T (h_k + \nabla h_k^T s_k) \\ &\quad -\Delta\mu_k^T (g_k + \nabla g_k^T s_k) + \frac{\rho_k}{2}[\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T s_k)\|^2] \\ &\quad + r_k[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

Using (2.1), the predicted reduction can be written as

$$\begin{aligned} (2.4) \qquad Pred_k &= q_k(0) - q_k(s_k) - \Delta\lambda_k^T (h_k + \nabla h_k^T s_k) \\ &\quad -\Delta\mu_k^T (g_k + \nabla g_k^T s_k) + r_k[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

For later reference, we define the tangential predicted decrease $Tpred_k$ to be

$$\begin{aligned} (2.5) \qquad Tpred_k &= -(Z_k^T(\nabla_x l_k + H_k s_k^n))^T \bar{s}_k^t - \frac{1}{2} \bar{s}_k^{t^T} Z_k^T H_k Z_k \bar{s}_k^t \\ &\quad + \frac{\rho_k}{2}[\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2]. \end{aligned}$$

The normal predicted decrease $Npred_k$ is defined to be the decrease at the k^{th} iteration in the linearized model of the equality constraints by the step s_k^n. That is,

$$(2.6) \qquad Npred_k = \|h_k\|^2 - \|h_k + \nabla h_k^T s_k^n\|^2.$$

After computing a trial step and updating the Lagrange multipliers, the penalty parameter is updated to ensure that $Pred_k \geq 0$. To update r_k, we use the scheme proposed by El-Alem (1988)[8]. We tentatively set $r_k = r_{k-1}$ and if $Pred_k < \frac{r_k}{2}[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]$, then we set

$$(2.7) \quad r_k = \frac{2[q_k(s_k) - q_k(0) + \Delta\lambda_k^T(h_k + \nabla h_k^T s_k) + \Delta\mu_k^T(g_k + \nabla g_k^T s_k)]}{\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2} + \beta,$$

where $\beta > 0$ is a small fixed constant.

After that, the step is tested to know whether it is accepted. This is done by comparing $Pred_k$ against $Ared_k$. If $\frac{Ared_k}{Pred_k} < \eta_1$ where $\eta_1 \in (0,1)$ is a small fixed constant, then the step is rejected. In this case, the radius of the trust region δ_k is decreased by setting $\delta_k = \alpha_1 \|s_k\|$, where $\alpha_1 \in (0,1)$, and another trial step is computed using the new trust-region radius.

If $\frac{Ared_k}{Pred_k} \geq \eta_1$, then the step is accepted. Our theory requires that at the beginning of the next iteration, δ_{k+1} must be greater than or equal to δ_{min}, where δ_{min} is a positive constant chosen at the beginning of the algorithm. That is, δ_k can be reduced below δ_{min} while finding an acceptable step. But, $\delta_{k+1} \geq \delta_{min}$ is required at the beginning of the next iteration after accepting the step s_k. Our theory also requires that, for all k, $\delta_k \leq \delta_{max}$, where $\delta_{max} > \delta_{min}$ is another positive constant chosen at the beginning of the algorithm. Our way of evaluating the trial steps and updating the trust-region radius is presented in Step 6 of Algorithm 2.1 below.

After accepting the step, we update the parameter ρ_k and the Hessian matrix H_k. To update ρ_k, we use a scheme suggested by Yuan (1995)[26]. In this scheme, another parameter σ_k has to be updated with ρ_k. We set $\rho_{k+1} = \rho_k$ and $\sigma_{k+1} = \sigma_k$, if

$$\frac{1}{2}Tpred_k - \Delta\mu_k^T(g_k + \nabla g_k^T s_k) \geq \sigma_k \|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\}. \tag{2.8}$$

Otherwise, we set $\rho_{k+1} = 2\rho_k$ and $\sigma_{k+1} = \frac{1}{2}\sigma_k$.

Our theory requires that the sequence $\{H_k\}$ of Hessian matrices be bounded. Thus, we can use the exact Hessians or any approximation to the Hessians that produces a bounded sequence of matrices.

Finally, the algorithm is terminated when either $\|s_k\| \leq \varepsilon_1$ or

$$\|Z_k^T \nabla_x l_k\| + \|\nabla g_k W_k g_k\| + \|h_k\| \leq \varepsilon_2,$$

for some $\varepsilon_1, \varepsilon_2 > 0$,

2.3. Main Algorithm. A formal description of our trust-region algorithm for solving Problem (1.1) is presented in the following algorithm.

ALGORITHM 2.1. *(The Main Algorithm)*

Step 0. (Initialization)

Given $x_1 \in \Re^n$. Compute W_1.

Evaluate μ_1 and λ_1 (see Step 4 with $k = 0$ and $\lambda_0 = (0, 0, ..., 0)^T$).

Set $\rho_1 = 1$, $r_0 = 1$, $\sigma_1 = 1$, and $\beta = 0.1$.

Choose ε, ε_1, ε_2, α_1, α_2, η_1, and η_2 such that $\varepsilon > 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$, $0 < \alpha_1 < 1 < \alpha_2$, and $0 < \eta_1 < \eta_2 < 1$.

Choose δ_{min}, δ_{max}, and δ_1 such that $\delta_{min} \leq \delta_1 \leq \delta_{max}$.

Set $k = 1$.

Step 1. (Test for convergence)

If $\|Z_k^T \nabla_x l_k\| + \|\nabla g_k W_k g_k\| + \|h_k\| \leq \varepsilon_2$, then terminate the algorithm.

Step 2. (Compute a trial step)

If $\|h_k\| = 0$, then

a) Set $s_k^n = 0$.

b) Compute the step $\bar{s}_k^t$ by solving

$$\begin{array}{ll} \text{minimize} & (Z_k{}^T(\nabla_x l_k + \rho_k \nabla g_k W_k g_k))^T \bar{s}^t + \frac{1}{2}\bar{s}^{t^T} Z_k^T B_k Z_k \bar{s}^t \\ \text{subject to} & \|Z_k \bar{s}^t\| \leq \delta_k. \end{array}$$

c) Set $s_k = Z_k \bar{s}_k^t$.

Else

a) Compute the normal component s_k^n by solving

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\|\nabla h_k^T s^n + h_k\|^2 \\ \text{subject to} & \|s^n\| \leq \zeta\delta_k. \end{array}$$

b) If $\|Z_k^T(\nabla_x l_k + \rho_k \nabla g_k W_k g_k + B_k s_k^n)\| = 0$,
then set $\bar{s}_k^t = 0$.
Else, compute $\bar{s}_k^t$ *by solving*

$$\begin{array}{ll} \textit{minimize} & ({Z_k}^T \nabla q_k(s_k^n))^T \bar{s}^t + \frac{1}{2} \bar{s}^{t^T} {Z_k}^T B_k Z_k \bar{s}^t \\ \textit{subject to} & \|Z_k \bar{s}^t\| \leq \Delta_k. \end{array}$$

End if.
c) Set $s_k = s_k^n + Z_k \bar{s}_k^t$ *and* $x_{k+1} = x_k + s_k$.
End if.
If $\|s_k\| \leq \varepsilon_1$, *then terminate the algorithm.*

Step 3. (Update the active set)
Compute W_{k+1}.

Step 4. (Compute the Lagrange multipliers μ_{k+1} *and* λ_{k+1}*)*
a) Compute μ_{k+1} *by solving*

$$\begin{array}{ll} \textit{minimize} & \|Z_{k+1}^T(\nabla f_{k+1} + \nabla g_{k+1} W_{k+1} \mu)\|^2 \\ \textit{subject to} & W_{k+1}\mu \geq 0 \end{array} \tag{2.9}$$

and set the rest of the components of μ_{k+1} *to zero.*
b) If $\|\nabla f_{k+1} + \nabla h_{k+1} \lambda_k + \nabla g_{k+1} W_{k+1} \mu_{k+1}\| \leq \varepsilon$, *then set* $\lambda_{k+1} = \lambda_k$.
Else, compute λ_{k+1} *by solving*

$$\textit{minimize} \ \|\nabla f_{k+1} + \nabla g_{k+1} \mu_{k+1} + \nabla h_{k+1} \lambda\|^2.$$

End if.

Step 5. (Update the penalty parameter r_k*)*
a) Set $r_k = r_{k-1}$.
b) If $Pred_k \leq \frac{r_k}{2}[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]$,
then set

$$r_k = \frac{2[q_k(s_k) - q_k(0) + \Delta\lambda_k^T(h_k + \nabla h_k^T s_k) + \Delta\mu_k^T(g_k + \nabla g_k^T s_k)]}{\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2} + \beta,$$

End if.

Step 6. (Test the step and update the trust-region radius)
If $\frac{Ared_k}{Pred_k} < \eta_1$
Reduce the trust-region radius by setting $\delta_k = \alpha_1 \|s_k\|$.
Go to step 2.
Else 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: $\delta_{k+1} = \max(\delta_k, \delta_{min})$.
Else
Accept the step: $x_{k+1} = x_k + s_k$.
Set the trust-region radius: $\delta_{k+1} = \min\{\delta_{max}, \max\{\delta_{min}, \alpha_2 \delta_k\}\}$.
End if.

Step 7. (Update the parameters ρ_k *and* σ_k*)*
a) Set $\rho_{k+1} = \rho_k$ *and* $\sigma_{k+1} = \sigma_k$.
b) If

$$\frac{1}{2} Tpred_k - \Delta\mu_k^T(g_k + \nabla g_k^T s) \leq \sigma_k \|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\},$$

then set $\rho_{k+1} = 2\rho_k$ *and* $\sigma_{k+1} = \frac{1}{2}\sigma_k$.

End if

Step 8. Set $k = k + 1$ and go to Step 1.

In Sections 5-8, we present our global convergence theory. However, for the global convergence results to follow, we require some assumptions to be satisfied by the problem. The following section is devoted to this subject.

3. Assumptions. We state the assumptions under which our global convergence theory is proved. Let $\{(x_k, \lambda_k, \mu_k)\}$ be the sequence of points generated by the algorithm and let Ω be a convex subset of $\Re^n$ that contains all iterates x_k and $x_k + s_k$, for all trial steps s_k examined in the course of the algorithm. On the set Ω, the following assumptions are imposed.

Problem Assumptions:

A_1. The functions f, h, and g are twice continuously differentiable for all $x \in \Omega$.

A_2. The matrix $\nabla h(x)$ has full column rank.

A_3. All of $f(x)$, $\nabla f(x)$, $\nabla^2 f(x)$, $h(x)$, $\nabla h(x)$, $\nabla^2 h_i(x)$ for $i = 1, 2, ..., m$, $g(x)$, $\nabla g(x)$, $\nabla^2 g_i(x)$ for $i = 1, 2, ..., p$, and $(\nabla h(x)^T \nabla h(x))^{-1}$ are uniformly bounded in Ω.

A_4. The sequence $\{\mu_k\}$ is bounded.

A_5. If an approximation to the Hessian of the Lagrangian is used, then the sequence of approximated Hessian matrices $\{H_k\}$ is bounded.

The boundedness of the sequence $\{\lambda_k\}$ follows from Assumptions A_2 and A_4. The boundedness of the sequence $\{\mu_k\}$ holds, if we solve (2.9) for the minimum norm solution, as long as the sequence of smallest singular values of the elements of $\{Z_k^T \nabla g_k W_k\}$ is bounded away from zero. This always hold unless a subsequence of the iteration sequence asymptotically satisfies some stationarity conditions that are not KKT conditions. This is takes us to the following section.

4. Stationary Points. In the above assumptions, even though we assume that $\nabla h(x)$ has full column rank for all $x \in \Omega$, we do not require the regularity assumption to hold. So, we may have other kinds of stationary points. They are presented in the following three definitions.

DEFINITION 4.1. *(Infeasible KKT Point) A point x_* is called an infeasible KKT point if there exist λ_* and μ_* such that*

$$\begin{aligned} \nabla f(x_*) + \nabla h(x_*)\lambda_* + \nabla g(x_*)\mu_* &= 0, \\ h(x_*) &= 0, \\ \nabla g(x_*) W_* g(x_*) &= 0 \quad \text{but } \|W_* g(x_*)\| > 0, \\ (\mu_*)_i g_i(x_*) &\geq 0, \quad i = 1, 2, ..., p, \\ (\mu_*)_i &\geq 0, \quad i = 1, 2, ..., p. \end{aligned}$$

DEFINITION 4.2. *(Feasible Fritz John Point) A point x_* is called a feasible Fritz John point if there exist γ_*, λ_*, and μ_* such that*

$$\begin{aligned} \gamma_* \nabla f(x_*) + \nabla h(x_*)\lambda_* + \nabla g(x_*)\mu_* &= 0, \\ h(x_*) &= 0, \\ W_* g(x_*) &= 0, \\ (\mu_*)_i g_i(x_*) &= 0, \quad i = 1, 2, ..., p, \\ \gamma_*, \ (\mu_*)_i &\geq 0, \quad i = 1, 2, ..., p. \end{aligned}$$

In the above definition, the conditions that the point $(x_*, \gamma_*, \lambda_*, \mu_*)$ must satisfy to be a feasible Fritz John point are called the feasible Fritz John conditions. More details about Fritz John conditions can be found in Mangasarian (1969)[16].

DEFINITION 4.3. *(Infeasible Fritz John Point) A point x_* is called an infeasible Fritz John point if there exist γ_*, λ_*, and μ_* such that*

$$\begin{aligned} \gamma_* \nabla f(x_*) + \nabla h(x_*)\lambda_* + \nabla g(x_*)\mu_* &= 0, \\ h(x_*) &= 0, \\ \nabla g(x_*) W_* g(x_*) &= 0 \quad \textit{but } \|W_* g(x_*)\| > 0, \\ (\mu_*)_i g_i(x_*) &\geq 0, \quad i = 1, 2, ..., p, \\ \gamma_*,\ (\mu_*)_i &\geq 0, \quad i = 1, 2, ..., p. \end{aligned}$$

In the above definition, the conditions that the point $(x_*, \gamma_*, \lambda_*, \mu_*)$ must satisfy to be an infeasible Fritz John point are called the infeasible Fritz John conditions.

The following two lemmas provide conditions such that if they are satisfied by the sequence of iterates generated by the algorithm, then the Fritz John conditions are satisfied in the limit. Another two similar lemmas for a different algorithm were given by Yuan (1995)[26].

LEMMA 4.1. *Assume A_1-A_5. A subsequence $\{k_i\}$ of the iteration sequence asymptotically satisfies the infeasible Fritz John conditions if it satisfies:*
1) $\lim_{k_i \to \infty} h(x_{k_i}) = 0$.
2) $\lim_{k_i \to \infty} \nabla g(x_{k_i}) W_{k_i} g(x_{k_i}) = 0$ *but* $\lim_{k_i \to \infty} \|W_{k_i} g(x_{k_i})\| > 0$.
3) $\lim_{k_i \to \infty} \left\{ min_{s \in \Re^{n-m}} \|W_k(g_k + \nabla g_k^T Z_k s)\|^2 \right\} = \lim_{k_i \to \infty} \|W_{k_i} g_{k_i}\|^2$.

Proof. Without loss of generality and to simplify the notation, we take $\{k_i\}$ to be the whole sequence $\{k\}$.

The minimizer $\bar{s}_k$ of $minimize_s \|W_k(g_k + \nabla g_k^T Z_k s)\|^2$ satisfies

$$Z_k^T \nabla g_k W_k g_k + Z_k^T \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k = 0. \tag{4.1}$$

From Condition 3,

$$\lim_{k \to \infty} \{2\bar{s}_k{}^T Z_k^T \nabla g_k W_k g_k + \bar{s}_k{}^T Z_k^T \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k\} = 0.$$

If $\lim_{k \to \infty} \bar{s}_k = 0$, then from (4.1) $\lim_{k \to \infty} Z_k^T \nabla g_k W_k g_k = 0$. Otherwise, multiply equation (4.1) from the left by $2\bar{s}_k{}^T$ and subtract from the above limit, we have $\lim_{k \to \infty} \|W_k \nabla g_k^T Z_k \bar{s}_k\| = 0$. But this implies $\lim_{k \to \infty} Z_k^T \nabla g_k W_k g_k = 0$. Hence, in either case, we have

$$\lim_{k \to \infty} Z_k^T \nabla g_k W_k g_k = 0.$$

Take $(\mu_k)_i = (W_k g_k)_i$, $i = 1, ..., p$. Since $\lim_{k \to \infty} \|W_k g_k\| > 0$, then $\lim_{k \to \infty} (\mu_k)_i \geq 0$, for $i = 1, ..., p$ and $\lim_{k \to \infty} (\mu_k)_i > 0$ for some i. Therefore, $\lim_{k \to \infty} Z_k^T \nabla g_k \mu_k = 0$. But this implies the existence of a sequence $\{\lambda_k\}$ such that

$$\lim_{k \to \infty} \{\nabla h_k \lambda_k + \nabla g_k \mu_k\} = 0.$$

Thus the conditions of Definition 4.3 hold in the limit with $\gamma = 0$. □

The above lemma shows a situation where in the limit the step $Z_{k_i} s_{k_i}$ lies in the null space of $\nabla g_{k_i} W_{k_i}$ and therefore was unable to decrease $\|\nabla g_{k_i} W_{k_i}\|$. Note that there is no condition on ∇l_{k_i}. Hence, if $\lim_{k \to \infty} \nabla l_{k_i} = 0$, then the infeasible KKT conditions are satisfied in the limit. Otherwise, the infeasible Fritz John conditions are satisfied. In the latter case, the termination condition inside the "while loop" of Algorithm 2.1 will eventually be satisfied and the algorithm terminates.

LEMMA 4.2. *Assume A_1-A_5. A subsequence $\{k_j\}$ of the iteration sequence asymptotically satisfies the feasible Fritz John conditions if it satisfies:*
1) $\lim_{k_j\to\infty} h(x_{k_j}) = 0$.
2) For all k_j, $\|W_{k_j}g_{k_j}\| > 0$ *and* $\lim_{k_j\to\infty} W_k g_{k_j}(x_k) = 0$.
3) $\lim_{k_j\to\infty}\left\{\min_{s\in\Re^{n-m}} \frac{\|W_{k_j}(g_{k_j}+\nabla g_{k_j}^T Z_{k_j}s)\|^2}{\|W_{k_j}g_{k_j}\|^2} : \|Z_{k_j}s + s_k^n\| \le \|W_{k_j}g_{k_j}\|\right\} = 1$.

Proof. Without loss of generality and to simplify the notation, we take $\{k_j\}$ to be the whole sequence $\{k\}$. The limit in Condition 3 is equivalent to

$$\lim_{k\to\infty}\left\{\min_{\|Z_k d+v_k\|\le 1} \|U_k + W_k\nabla g_k^T Z_k d\|^2\right\} = 1, \tag{4.2}$$

where U_k is a unit vector in the direction of $W_k g_k$, $d = \frac{s}{\|W_k g_k\|}$, and $v_k = \frac{s_k^n}{\|W_k g_k\|}$.

Consider the problem

$$\min_{\|Z_k d+v_k\|\le 1} \|U_k + W_k\nabla g_k^T Z_k d\|^2.$$

The Lagrangian function associated with the above problem is the function

$$l(d_k,\gamma_k) = d_k^T Z_k^T \nabla g_k W_k \nabla g_k^T Z_k d_k + 2U_k^T W_k \nabla g_k^T Z_k d_k + 1 + \nu_k(d_k^T Z_k^T Z_k d_k + v_k^T v_k - 1),$$

where $\nu_k \ge 0$ is the Lagrange multiplier of the constraint $\|Z_k d_k\|^2 + \|v_k\|^2 \le 1$.

Let $\bar{d}_k$ be a minimizer to the above problem. Then, from the optimality conditions

$$\nabla_x l_k = Z_k^T \nabla g_k W_k \nabla g_k^T Z_k \bar{d}_k + Z_k^T \nabla g_k W_k U_k + \gamma_k Z_k^T Z_k \bar{d}_k = 0, \tag{4.3}$$

If $\lim_{k\to\infty} Z_k\bar{d}_k = 0$, then $\lim_{k\to\infty} Z_k^T\nabla g_k W U_k = 0$. Otherwise, using the fact that $\bar{d}_k$ is a solution to the minimization problem in (4.2), we have

$$\lim_{k\to\infty}\{\bar{d}_k{}^T Z_k^T \nabla g_k W_k \nabla g_k^T Z_k \bar{d}_k + 2U_k^T W_k \nabla g_k^T Z_k \bar{d}_k\} = 0.$$

Multiplying (4.3) from the left by $2\bar{d}_k{}^T$ and using the above limit, we have $\lim_{k\to\infty}\nu_k = 0$ and $\lim_{k\to\infty}\bar{d}_k Z_k^T\nabla g_k W_k \nabla g_k^T Z_k \bar{d}_k = 0$. This implies $\lim_{k\to\infty} Z_k^T \nabla g_k W_k U_k = 0$. Hence in both cases, we have

$$\lim_{k\to\infty} Z_k^T \nabla g_k W_k U_k = 0.$$

The rest of the proof follows using arguments similar to those in the above lemma. □

Eventhough it looks like there is no condition on $\{\nabla l_{k_j}\}$, the condition is hidden inside the constraint $\|Z_{k_j}s + s_k^n\| \le \|W_{k_j}g_{k_j}\|$. In particular, if $\lim_{k_j\to\infty}\nabla l_{k_j} = 0$, the sequence $\{k_j\}$ satisfies the KKT conditions in the limit. Otherwise, from the way of computing the step, $\|\nabla l_{k_j}^T s_{k_j}^t\| = o(\rho_{k_j}\|W_{k_j}g_{k_j}\|^2)$. This means that the angle between ∇l_{k_j} and $s_{k_j}^t$ approaches $\frac{\pi}{2}$ as $k\to\infty$. For this reason the step $s_{k_j}^t$ was unable to decrease ∇l_{k_j}. In the latter case, the termination condition inside the "while loop" of Algorithm 2.1 will eventually be satisfied and the algorithm terminates.

From the above two lemmas, we can easily see that, for any subsequence of the iteration sequence that asymptotically satisfies the infeasible or the feasible Fritz-John conditions, the corresponding subsequence of smallest singular values of $Z_k^T \nabla g_k W_k$ is not bounded away from zero.

The following two lemmas show that if $\rho_k \to \infty$ as $k \to \infty$, then the iteration sequence generated by the algorithm has a subsequence that satisfies the Fritz John conditions in the limit.

LEMMA 4.3. *Assume A_1-A_5. If $\rho_k \to \infty$, as $k \to \infty$ and $\limsup_{k\to\infty} \|W_k g_k\| > 0$, then the sequence of iterates has a subsequence that satisfies the infeasible Fritz John conditions in the limit.*

Proof. The proof is by contradiction. Suppose there exists no subsequence of the sequence of iterates that satisfies the infeasible Fritz John conditions in the limit. Using Lemma 4.1, we have for all k, $\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2 \geq \varepsilon > 0$. Since $\{\rho_k\}$ is unbounded, then as $k \to \infty$,

$$\rho_k\{\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2\} \to \infty. \tag{4.4}$$

From the way of updating the two parameters ρ_k and σ_k we have, $\sigma_k \to 0$. This implies the existence of infinite number of iterates at which

$$\frac{1}{2} Tpred_k - \Delta\mu_k^T(g_k + \nabla g_k^T s_k) \leq \sigma_k \|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\}. \tag{4.5}$$

Under Assumptions A_3-A_5, and using (4.4), we have $Tpred_k \to \infty$. Hence, as $k \to \infty$, the left hand side of Inequality (4.5) tends to infinity while the right hand side goes to zero. This contradiction proves the lemma. □

LEMMA 4.4. *Assume A_1-A_5. If $\rho_k \to \infty$, as $k \to \infty$, $\lim_{k\to\infty} \|h_k\| = 0$, and there exists a subsequence $\{k_j\}$ of iterates that satisfies $\|W_k g_k\| > 0$ for all $k \in \{k_j\}$ and $\lim_{k_j \to\infty} \|W_{k_j} g_{k_j}\| = 0$, then the sequence $\{k_j\}$ satisfies the feasible Fritz John conditions in the limit.*

Proof. The proof is by contradiction. Assume that for all k sufficiently large $\|h_k\| \leq \varepsilon_0$, $\|W_k g_k\| \leq \varepsilon_1$, and there exists no subsequence that satisfies the feasible Fritz John conditions in the limit. Hence, the sequence of smallest singular values of $\{Z_k^T \nabla g_k W_k\}$ is bounded away from zero. Using Lemma 4.2, there exists a constant $\varepsilon_2 > 0$ such that, for all k sufficiently large,

$$\frac{\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2}{\|W_k g_k\|^2} \geq \varepsilon_2.$$

Hence, $\varepsilon_2 \|W_k g_k\|^2 \leq \|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2$.

Since $\rho_k \to \infty$ and $\sigma_k \to 0$, we can write for k sufficiently large

$$\begin{aligned}
\sigma_k \|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\} &\leq \varepsilon_2 \|W_k g_k\|^2 \\
&\leq \frac{\rho_k}{4}[\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2] \\
&\leq \frac{1}{2} Tpred_k - \Delta\mu_k^T(g_k + \nabla g_k^T \bar{s}_k^t)) \\
&\quad + O(\|W_k g_k\|^2) + o(\rho_k \|W_k g_k\|^2).
\end{aligned}$$

But this implies that, for k sufficiently large, Condition (2.8) is satisfied. This contradicts the fact that $\rho_k \to \infty$ as $k \to \infty$ and proves the lemma. □

We continue our analysis assuming that the algorithm generates a sequence of iterates $\{k\}$ that has no subsequence that satisfies in the limit the Fritz John conditions that are not KKT conditions. This means that, we assume the existence of an integer $\bar{k}$ such that for all $k \geq \bar{k}$, $\rho_k = \bar{\rho}$, $\sigma_k = \bar{\sigma} > 0$, and

$$\frac{1}{2} Tpred_k - \Delta\mu_k^T (g_k + \nabla g_k^T s_k) \geq \bar{\sigma} \|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\}. \tag{4.6}$$

Without loss of generality we take $\bar{k} = 1$.

From Assumptions A_3 and A_5 and the above assumption, we can say that there exists a positive constant b such that for all k

$$\|B_k\| \leq b, \quad \|Z_k^T B_k\| \leq b, \quad and \quad \|Z_k^T B_k Z_k\| \leq b, \tag{4.7}$$

where $B_k = H_k + \bar{\rho} \nabla g_k W_k \nabla g_k^T$.

5. Technical Lemmas. We present some important lemmas needed in the subsequent proofs.

LEMMA 5.1. *Let A_1-A_3 hold, then at any iteration k*

$$\|s_k^n\| \leq K_1 \|h_k\|, \tag{5.1}$$

where $K_1 > 0$ is a constant independent of k.

Proof. The proof is similar to the proof of Lemma 7.1 of Dennis, El-Alem, and Maciel (1997)[5]. □

LEMMA 5.2. *Assume A_1 and A_3. Then $W(x)g(x)$ is Lipschitz continuous in Ω.*

Proof. See Lemma 4.1 of Dennis, El-Alem, and Williamson (1994)[6]. □

From the above lemma, we conclude that $g(x)^T W(x) g(x)$ is differentiable and $\nabla g(x) W(x) g(x)$ is Lipschitz continuous in Ω.

The following lemma shows that, at any iteration k, the normal predicted reduction is at least equal to the decrease in the 2-norm of the linearized constraints obtained by the Cauchy step.

LEMMA 5.3. *Assume A_1-A_5. Then there exists a positive constant K_2 independent of the iterates such that the predicted decrease obtained by the normal component s_k^n of the trial step satisfies*

$$Npred_k \geq K_2 \|h_k\| \min\{\delta_k, \|h_k\|\}. \tag{5.2}$$

Proof. See El-Alem (1991)[9] □

From the way of updating the penalty parameter r_k and the above lemma, we have, for all k,

$$Pred_k \geq \frac{r_k}{2} K_2 \|h_k\| \min\{\delta_k, \|h_k\|\}. \tag{5.3}$$

The following lemma gives a relation between the tangential predicted decrease and the decrease in the quadratic model of the Lagrangian function obtained by the Cauchy step.

LEMMA 5.4. *Assume A_1-A_5. Then there exists a positive constant K_3 independent of the iterates such that*

$$\begin{aligned} Tpred_k - \bar{\rho} {s_k^n}^T \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \quad \geq \quad & K_3 \|Z_k^T (\nabla_x l_k + \bar{\rho} \nabla g_k W_k g_k + B_k s_k^n)\| \\ & \min\{\Delta_k, \|Z_k^T (\nabla_x l_k + \bar{\rho} \nabla g_k W_k g_k + B_k s_k^n)\|\}, \end{aligned}$$

where $\Delta_k = \sqrt{\delta_k^2 - \|s_k^n\|^2}$.

Proof. The component $\bar{s}_k^t$ satisfies the fraction-of-Cauchy decrease condition; *i.e.*,

$$q_k(s_k^n) - q_k(s_k^n + Z_k\bar{s}_k^t) \geq \vartheta[q_k(s_k^n) - q_k(s_k^n + Z_k s_k^{tcp})], \tag{5.4}$$

for some $\vartheta \in (0,1]$, where s_k^{tcp} is given by $s_k^{tcp} = -t_k^{cp} Z_k^T \nabla q_k(s_k^n)$ and the parameter t_k^{cp} is defined by

$$t_k^{cp} = \begin{cases} \frac{\|Z_k^T\nabla q_k(s_k^n)\|^2}{(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n)} & \text{if } \frac{\|Z_k^T\nabla q_k(s_k^n)\|^3}{(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n)} \leq \Delta_k \\ & \text{and } (Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n) > 0, \\ \frac{\Delta_k}{\|Z_k^T\nabla q_k(s_k^n)\|} & \text{otherwise,} \end{cases}$$

where $\bar{B}_k = Z_k^T B_k Z_k$. If $s_k^{tcp} = -\frac{\Delta_k}{\|Z_k^T\nabla q_k(s_k^n)\|} Z_k^T\nabla q_k(s_k^n)$ and $\|Z_k^T\nabla q_k(s_k^n)\|^3 \geq \Delta_k(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n)$, then

$$\begin{aligned} q_k(s_k^n) - q_k(s_k^n + Z_k s_k^{tcp}) &= [l_k + \nabla_x l_k^T s_k^n + \frac{1}{2}s_k^{n^T} H_k s_k^n + \frac{\bar{\rho}}{2}\|W_k(g_k + \nabla g_k^T s_k^n)\|^2] \\ &\quad -[l_k + \nabla_x l_k^T(s_k^n + Z_k s_k^{tcp}) \\ &\quad +\frac{1}{2}(s_k^n + Z_k s_k^{tcp})^T H_k (s_k^n + Z_k s_k^{tcp}) \\ &\quad +\frac{\bar{\rho}}{2}\|W_k(g_k + \nabla g_k^T(s_k^n + Z_k s_k^{tcp}))\|^2] \\ &= -(Z_k^T\nabla q_k(s_k^n))^T s_k^{tcp} - \frac{1}{2}s_k^{tcp^T}\bar{B}_k s_k^{tcp} \\ &= \Delta_k\|Z_k^T\nabla q_k(s_k^n)\| \\ &\quad -\frac{\Delta_k^2}{2\|Z_k^T\nabla q_k(s_k^n)\|^2}((Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n)) \\ &\geq \frac{1}{2}\Delta_k\|Z_k^T\nabla q_k(s_k^n)\|. \end{aligned}$$

On the other hand, if $s_k^{tcp} = -\frac{\|Z_k^T\nabla q_k(s_k^n)\|^2}{(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n)} Z_k^T\nabla q_k(s_k^n)$ and $\|Z_k^T\nabla q_k(s_k^n)\|^3 \leq \Delta_k(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k Z_k^T\nabla q_k(s_k^n)$, then

$$\begin{aligned} q_k(s_k^n) - q_k(s_k^n + Z_k s_k^{tcp}) &= -(Z_k^T\nabla q_k(s_k^n))^T s_k^{tcp} - \frac{1}{2}s_k^{tcp^T}\bar{B}_k s_k^{tcp} \\ &= \frac{\|(Z_k^T\nabla q_k(s_k^n))\|^4}{(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k(Z_k^T\nabla q_k(s_k^n))} \\ &\quad -\frac{\|Z_k^T\nabla q_k(s_k^n)\|^4}{2(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k(Z_k^T\nabla q_k(s_k^n))} \\ &= \frac{\|Z_k^T\nabla q_k(s_k^n)\|^4}{2(Z_k^T\nabla q_k(s_k^n))^T\bar{B}_k(Z_k^T\nabla q_k(s_k^n))} \\ &\geq \frac{\|Z_k^T\nabla q_k(s_k^n)\|^2}{2\|\bar{B}_k\|}. \end{aligned}$$

From the above two cases and using the problem assumptions, we have

$$\begin{aligned} q_k(s_k^n) - q_k(s_k^n + Z_k s_k^{tcp}) &\geq \bar{K}_3\|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\| \\ &\quad \min\{\Delta_k, \|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\|\}, \end{aligned} \tag{5.5}$$

where $\bar{K}_3$ is a positive constant that does not depend on k. On the other hand, from (2.1) and (2.5), we have

$$(5.6)\quad \begin{aligned} q_k(s_k^n) - q_k(s_k^n + Z_k\bar{s}_k^t) &= -[Z_k^T(\nabla_x l_k + H_k s_k^n)]^T \bar{s}_k^t - \frac{1}{2}\bar{s}_k^{t^T} Z_k^T H_k Z_k \bar{s}_k^t \\ &\quad + \frac{\bar{\rho}}{2}[\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T Z_k \bar{s}_k^t)\|^2] \\ &\quad - \bar{\rho} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t. \\ &= Tpred_k - \bar{\rho} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t. \end{aligned}$$

From (5.4), (5.5), and (5.6), we can write

$$\begin{aligned} Tpred_k - \bar{\rho} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \geq\ & \vartheta \bar{K}_3 \|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\| \\ & \min\{\Delta_k, \|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\|\}. \end{aligned}$$

Taking $K_3 = \vartheta \bar{K}_3$, we obtain the desired result. □

LEMMA 5.5. *At any iteration k, let $D(x_k) \in \Re^{p\times p}$ be a diagonal matrix whose diagonal entries are*

$$(5.7)\qquad (d_k)_i = \begin{cases} 1 & \text{if } (g_k)_i < 0 \text{ and } (g_{k+1})_i \geq 0, \\ -1 & \text{if } (g_k)_i \geq 0 \text{ and } (g_{k+1})_i < 0, \\ 0 & \text{otherwise}, \end{cases}$$

where $i = 1, 2, ..., p$. Then

$$(5.8)\qquad W_{k+1} = W_k + D_k.$$

Proof. We consider four cases: first, if $(g_k)_i \geq 0$ and $(g_{k+1})_i \geq 0$, then $(w_k)_i = 1$, $(w_{k+1})_i = 1$. Hence $(d_k)_i = 0$ and (5.8) holds. Second, if $(g_k)_i < 0$ and $(g_{k+1})_i < 0$, then $(w_k)_i = 0$, $(w_{k+1})_i = 0$. Hence $(d_k)_i = 0$ and (5.8) holds. Third, if $(g_k)_i < 0$ and $(g_{k+1})_i \geq 0$, then $(w_k)_i = 0$, $(w_{k+1})_i = 1$ and hence $(d_k)_i = 1$. Therefore (5.8) holds. Finally, if $(g_k)_i \geq 0$ and $(g_{k+1})_i < 0$, then $(w_k)_i = 1$, $(w_{k+1})_i = 0$ and hence $(d_k)_i = -1$ and therefore (5.8) holds. This completes the proof of the lemma. □

LEMMA 5.6. *Assume A_1 and A_3. At any iteration k, there exists a positive constant K_4 independent of k, such that*

$$(5.9)\qquad \|D_k g_k\| \leq K_4 \|s_k\|,$$

where $D_k \in \Re^{p\times p}$ is the diagonal matrix whose diagonal entries are defined in (5.7).

Proof. We consider three cases. First, if $(d_k)_i = 0$, then $(d_k)_i(g_k)_i = 0$. If $(d_k)_i = 1$, then $(g_k)_i < 0$ and $(g_{k+1})_i \geq 0$. From Assumption A_1, we can write $(g_{k+1})_i = (g_k)_i + \nabla g_i(x_k + \xi_1 s_k)^T s_k \geq 0$ where $\xi_1 \in (0,1)$. Since $(g_{k+1})_i \geq 0$ and $(g_k)_i < 0$, then $|(g_k)_i| \leq \nabla g_i(x_k + \xi_1 s_k)^T s_k$ and hence $|(g_k)_i| \leq \|\nabla g_i(x_k + \xi_1 s_k)\| \|s_k\|$. Finally, if $(d_k)_i = -1$, then $(g_k)_i \geq 0$ and $(g_{k+1})_i < 0$. Again using Assumption A_1, we can write $(g_{k+1})_i = (g_k)_i + \nabla g_i(x_k + \xi_1 s_k)^T s_k < 0$ and since $(g_k)_i \geq 0$, then $(g_k)_i \leq |\nabla g_i(x_k + \xi_1 s_k)^T s_k|$ and hence $(g_k)_i \leq \|\nabla g_i(x_k + \xi_1 s_k)\| \|s_k\|$. From the above and using Assumption A_3 there exists a positive constant K_4 such that

$$\|D_k g_k\| = \sqrt{\sum_{i=1}^{p} ((d_k)_i (g_k)_i)^2} \leq K_4 \|s_k\|.$$

Hence Inequality (5.9) holds. □

The following two lemmas give upper bounds 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 5.7. *Assume A_1-A_5, then there exists a constant $K_5 > 0$ that does not depend on k, such that*

$$(5.10) \qquad | Ared_k - Pred_k | \leq K_5[\|s_k\|^2 + r_k\|s_k\|^2\|h_k\| + r_k\|s_k\|^3].$$

Proof. From (2.2) and using (5.8), we have

$$\begin{aligned} Ared_k &= l(x_k, \lambda_k, \mu_k) - l(x_{k+1}, \lambda_k, \mu_k) - \Delta\lambda_k^T h_{k+1} - \Delta\mu_k^T g_{k+1} \\ &+ \frac{\bar{\rho}}{2}[g_k^T W_k g_k - g_{k+1}^T(W_k + D_k)g_{k+1}] + r_k[\|h_k\|^2 - \|h_{k+1}\|^2]. \end{aligned}$$

From the above equation and (2.3) and using Cauchy-Schwarz inequality, we have

$$\begin{aligned} | Ared_k - Pred_k | \;\leq\; & | l(x_k, \lambda_k, \mu_k) + \nabla l(x_k, \lambda_k, \mu_k)^T s_k - l(x_{k+1}, \lambda_k, \mu_k) | \\ & + | \Delta\lambda_k^T[h_k + \nabla h_k^T s_k - h_{k+1}] | \\ & + | \Delta\mu_k^T[g_k + \nabla g_k^T s_k - g_{k+1}] | \\ & + \frac{\bar{\rho}}{2} | \|W_k(g_k + \nabla g_k^T s_k)\|^2 - g_{k+1}^T(W_k + D_k)g_{k+1} | \\ & + r_k | \|h_k + \nabla h_k^T s_k\|^2 - \|h_{k+1}\|^2 | . \end{aligned}$$

Hence,

$$\begin{aligned} |Ared_k - Pred_k| \;\leq\; & | \frac{1}{2} s_k^T(H_k - \nabla^2 l(x_k + \xi_1 s_k, \lambda_k, \mu_k))s_k | \\ & + \frac{1}{2} | s_k^T[\nabla^2 h(x_k + \xi_2 s_k)\Delta\lambda_k]s_k | + \frac{1}{2} | s_k^T[\nabla^2 g(x_k + \xi_3 s_k)\Delta\mu_k]s_k | \\ & + \frac{\bar{\rho}}{2} | s_k^T[\nabla g_k W_k \nabla g_k^T - \nabla g(x_k + \xi_4 s_k) W_k \nabla g(x_k + \xi_4 s_k)^T]s_k | \\ & + \frac{\bar{\rho}}{2} | s_k^T \nabla^2 g(x_k + \xi_4 s_k) W_k g(x_k + \xi_4 s_k) s_k | \\ & + \frac{\bar{\rho}}{2}\|D_k[g_k + \nabla g(x_k + \xi_5 s_k)^T s_k]\|^2 \\ & + r_k | s_k^T[\nabla h_k \nabla h_k^T - \nabla h(x_k + \xi_6 s_k)\nabla h(x_k + \xi_6 s_k)^T]s_k | \\ & + r_k | s_k^T \nabla^2 h(x_k + \xi_6 s_k) h(x_k + \xi_6 s_k) s_k |, \end{aligned}$$

for some ξ_1, ξ_2, ξ_3, ξ_4, ξ_5, and $\xi_6 \in (0, 1)$. Hence, by using the problem assumptions and Inequality (5.9), the proof follows. □

LEMMA 5.8. *Assume A_1-A_5, then there exists a constant $K_6 > 0$ that does not depend on k, such that*

$$(5.11) \qquad | Ared_k - Pred_k | \leq K_6 r_k \|s_k\|^2.$$

Proof. The proof follows directly from the above lemma, the fact that $r_k \geq 1$, and the problem assumptions. □

The following inequality is needed in many forthcoming lemmas.

LEMMA 5.9. *Assume A_1-A_5. Then there exists a constant $K_7 > 0$ that does not depend on k such that*

$$q_k(0) - q_k(s_k^n) - \Delta\lambda_k^T(h_k + \nabla h_k^T s_k) - \frac{\bar{\rho}}{2} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \geq -K_7 \|h_k\|.$$

Proof. From (2.1), we have

$$\begin{aligned} q_k(0) - q_k(s_k^n) &= -\nabla_x l_k^T s_k^n - \frac{1}{2} s_k^{n^T} H_k s_k^n + \frac{\bar{\rho}}{2}[\|W_k g_k\|^2 - \|W_k(g_k + \nabla g_k^T s_k^n)\|^2] \\ &= -(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k)^T s_k^n - \frac{1}{2} s_k^{n^T}(H_k + \bar{\rho}\nabla g_k W_k \nabla g_k^T) s_k^n \\ &= -(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k)^T s_k^n - \frac{1}{2} s_k^{n^T} B_k s_k^n. \end{aligned}$$

Hence,

$$\begin{aligned} q_k(0) - q_k(s_k^n) &- \Delta\lambda_k^T(h_k + \nabla h_k^T s_k) - \frac{\bar{\rho}}{2} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \\ &= -(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k)^T s_k^n - \frac{1}{2} s_k^{n^T} B_k s_k^n - \Delta\lambda_k^T(h_k + \nabla h_k^T s_k^n) \\ &\quad -\frac{\bar{\rho}}{2} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \\ &\geq -\|\nabla_x l_k\|\|s_k^n\| - \bar{\rho}\|\nabla g_k W_k g_k\|\|s_k^n\| - \|B_k\|\|s_k^n\|^2 \\ &\quad -\|\Delta\lambda_k\|\|h_k + \nabla h_k^T s_k^n\| - \frac{\bar{\rho}}{2}\|s_k^n\|\|\nabla g_k W_k \nabla g_k^T\|\|Z_k \bar{s}_k^t\| \\ &\geq -[\ \|\nabla_x l_k\| + \bar{\rho}\|\nabla g_k W_k g_k\| + \|B_k\|\|s_k^n\| \\ &\quad -\frac{\bar{\rho}}{2}\|\nabla g_k W_k \nabla g_k^T\|\|Z_k \bar{s}_k^t\|\ ]\ \|s_k^n\| - \|\Delta\lambda_k\|\|h_k + \nabla h_k^T s_k^n\|. \end{aligned}$$

Using Inequality (5.1), the fact that $\|h_k + \nabla h_k^T s_k^n\| \leq \|h_k\|$, we obtain

$$\begin{aligned} q_k(0) - q_k(s_k^n) &- \Delta\lambda_k^T(h_k + \nabla h_k^T s_k^n) - \bar{\rho} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \\ &\geq -[\ (\ \|\nabla_x l_k\| + \bar{\rho}\|\nabla g_k W_k g_k\| + \|B_k\|\|s_k^n\| \\ &\quad +\bar{\rho}\|\nabla g_k W_k \nabla g_k^T\|\|Z_k \bar{s}_k^t\|\ )\ K_1 - \|\Delta\lambda_k\|\ ]\ \|h_k\|. \end{aligned}$$

Under Assumptions A_3, A_4, and A_5, $\{\bar{\rho}\}$ is bounded, the facts that $\|s_k^n\| \leq \delta_{max}$ and $\|Z_k \bar{s}_k^t\| \leq \Delta_k \leq \delta_k \leq \delta_{max}$, and using (4.7), there exists $K_7 > 0$ which is independent of k, such that

$$q_k(0) - q_k(s_k^n) - \Delta\lambda_k^T(h_k + \nabla h_k^T s_k^n) - \frac{\bar{\rho}}{2} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t \geq -K_7 \|h_k\|.$$

This completes the proof. □

6. Sufficient Decrease in the Model. In this section, we deal with the predicted decrease in the augmented Lagrangian function produced by the trial step. We start with this lemma.

LEMMA 6.1. *Assume A_1-A_5, then for all k,*

$$\begin{aligned} Pred_k \geq\ & \frac{K_3}{2}\|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\| \\ & \min\{\|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\|, \Delta_k\} \\ & +\bar{\sigma}\|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\} \\ & -K_7\|h_k\| + r_k[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned} \tag{6.1}$$

Proof. From (2.4), we have

$$\begin{aligned} Pred_k &= [q_k(s_k^n) - q_k(s_k)] + [q_k(0) - q_k(s_k^n)] - \Delta\lambda_k^T(h_k + \nabla h_k^T s_k^n) \\ &\quad -\Delta\mu_k^T(g_k + \nabla g_k^T s_k) + r_k[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

The above equation can be written in the form

$$\begin{aligned} Pred_k &= \frac{1}{2}[Tpred_k - \bar{\rho} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t] \\ &\quad +[q_k(0) - q_k(s_k^n) - \Delta\lambda_k^T(h_k + \nabla h_k^T s_k^n) - \frac{\bar{\rho}}{2} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t] \\ &\quad +[\frac{1}{2}Tpred_k - \Delta\mu_k^T(g_k + \nabla g_k^T s_k)] + r_k[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

Using Inequality (4.6) and Lemmas 5.4 and 5.9, we have

$$\begin{aligned} Pred_k &\geq \frac{K_3}{2}\|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\| \\ &\quad \min\{\|Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k + B_k s_k^n)\|, \Delta_k\} \\ &\quad +\bar{\sigma}\|\nabla g_k W_k g_k\| \min\{\|\nabla g_k W_k g_k\|, \Delta_k\} \\ &\quad -K_7\|h_k\| + r_k[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

This completes the proof. □

LEMMA 6.2. *Assume A_1-A_5. Let k be the index of an iteration at which r_k is increased. Then there exists a positive constant K_8 that does not depend on k, such that*

$$r_k \min\{\|h_k\|, \delta_k\} \leq K_8. \tag{6.2}$$

Proof. Since r_k is increased at the k^{th} iteration then from (2.7) (see also Step 5 of Algorithm 2.1), we can write

$$\begin{aligned} \frac{r_k}{2}[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2] &= [q_k(s_k) - q_k(s_k^n)] + [q_k(s_k^n) - q_k(0)] \\ &\quad +\Delta\lambda_k^T(h_k + \nabla h_k^T s_k) + \Delta\mu_k^T(g_k + \nabla g_k^T s_k) \\ &\quad +\frac{\beta}{2}[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

By substituting (5.6) in the above equality, we have

$$\begin{aligned} \frac{r_k}{2}[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2] &= -\frac{1}{2}[Tpred_k - \bar{\rho} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t] \\ &\quad -[\frac{1}{2}Tpred_k - \Delta\mu_k^T(g_k + \nabla g_k^T s_k)] \\ &\quad +[q_k(s_k^n) - q_k(0) + \Delta\lambda_k^T(h_k + \nabla h_k^T s_k) \\ &\quad\quad +\frac{\bar{\rho}}{2} s_k^n \nabla g_k W_k \nabla g_k^T Z_k \bar{s}_k^t] \\ &\quad +\frac{\beta}{2}[\|h_k\|^2 - \|h_k + \nabla h_k^T s_k\|^2]. \end{aligned}$$

Applying Inequality (5.2) to the left hand side and Inequality (4.6) and Lemmas 5.4 and 5.9 to the right hand side, we obtain

$$\begin{aligned}\frac{r_k}{2}K_2\|h_k\|\min\{\delta_k,\|h_k\|\} &\leq -\frac{K_3}{2}\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\|\\ &\quad \min\{\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\|,\Delta_k\}\\ &\quad -\bar{\sigma}\|\nabla g_k W_k g_k\|\min\{\|\nabla g_k W_k g_k\|,\Delta_k\}\\ &\quad +K_7\|h_k\|+\frac{\beta}{2}[-2(\nabla h_k h_k)^T s_k^n-\|\nabla h_k^T s_k^n\|^2]\\ &\leq K_7\|h_k\|+\beta\|h_k\|\|\nabla h_k\|\|s_k^n\|.\end{aligned}$$

The rest of the proof follows using Assumption A_3 and the fact that $\|s_k^n\|\leq\delta_{max}$. □

LEMMA 6.3. *Assume A_1-A_5. At any given iteration k at which $\|h_k\|\leq\alpha\delta_k$ and $\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k)\|+\|\nabla g_k W_k g_k\|\geq\varepsilon$, where $\varepsilon>0$ and α is a positive constant given by*

$$\alpha\leq\min\left\{\frac{\varepsilon}{6bK_1\delta_{max}},\frac{\sqrt{3}}{2K_1},\frac{K_3\varepsilon}{24K_7}\min\{\frac{2\varepsilon}{3\delta_{max}},1\},\frac{\bar{\sigma}\varepsilon}{8K_7}\min\{\frac{2\varepsilon}{\delta_{max}},1\},\right\}, \tag{6.3}$$

there exists a positive constant K_9 that depends on ε but does not depend on k, such that

$$Pred_k\geq K_9\delta_k+r_k[\|h_k\|^2-\|h_k+\nabla h_k^T s_k\|^2]. \tag{6.4}$$

Proof. Let $\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k)\|\geq\frac{\varepsilon}{2}$. Since, $\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\|\geq\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k)\|-\|Z_k^T B_k s_k^n\|$, then using Inequalities (4.7) and (5.1), we have

$$\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\| \geq \|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k)\|-bK_1\|h_k\|.$$

Since $\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k)\|\geq\frac{\varepsilon}{2}$, $\|h_k\|\leq\alpha\delta_k$, and $\alpha\leq\frac{\varepsilon}{6bK_1\delta_{max}}$, then we have

$$\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\|\geq\frac{\varepsilon}{2}-bK_1\alpha\delta_k\geq\frac{\varepsilon}{3}. \tag{6.5}$$

Since $\Delta_k=\sqrt{{\delta_k}^2-\|s_k^n\|^2}$ and $\|s_k^n\|\leq K_1\|h_k\|\leq K_1\alpha\delta_k\leq K_1\frac{\sqrt{3}}{2K_1}\delta_k=\frac{\sqrt{3}}{2}\delta_k$, then we obtain $\Delta_k^2=\delta_k^2-\|s_k^n\|^2\geq\delta_k^2-\frac{3}{4}\delta_k^2=\frac{1}{4}\delta_k^2$. Hence,

$$\Delta_k\geq\frac{1}{2}\delta_k. \tag{6.6}$$

Since $\|h_k\|\leq\alpha\delta_k$, and using Inequalities (6.1), (6.5), and (6.6), then

$$\begin{aligned}Pred_k &\geq \frac{K_3}{2}\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\|\\ &\quad \min\{\|Z_k^T(\nabla_x l_k+\bar{\rho}\nabla g_k W_k g_k+B_k s_k^n)\|,\frac{1}{2}\delta_k\}\\ &\quad -K_7\|h_k\|+r_k[\|h_k\|^2-\|h_k+\nabla h_k^T s_k\|^2]\\ &\geq \frac{K_3\varepsilon}{12}\delta_k\min\{\frac{2\varepsilon}{3\delta_{max}},1\}-K_7\alpha\delta_k+r_k[\|h_k\|^2-\|h_k+\nabla h_k^T s_k\|^2].\end{aligned}$$

Since $\alpha \leq \frac{K_3\varepsilon}{24K_7}\min\{\frac{2\varepsilon}{3\delta_{max}}, 1\}$, then we have

$$Pred_k \geq \frac{K_3\varepsilon}{24}\min\{\frac{2\varepsilon}{3\delta_{max}}, 1\}\delta_k + r_k[||h_k||^2 - ||h_k + \nabla h_k^T s_k||^2].$$

Now, consider the case when $||\nabla g_k W_k g_k|| \geq \frac{\varepsilon}{2}$. Using Inequalities (6.1) and (6.6), we have

$$\begin{aligned} Pred_k &\geq \bar{\sigma}||\nabla g_k W_k g_k|| \min\{||\nabla g_k W_k g_k||, \frac{1}{2}\delta_k\} \\ &\quad -K_7||h_k|| + r_k[||h_k||^2 - ||h_k + \nabla h_k^T s_k||^2] \\ &\geq \frac{\bar{\sigma}\varepsilon}{4}\min\{\frac{2\varepsilon}{\delta_{\max}}, 1\}\delta_k - K_7\alpha\delta_k + r_k[||h_k||^2 - ||h_k + \nabla h_k^T s_k||^2]. \end{aligned}$$

Since $\alpha \leq \frac{\bar{\sigma}\varepsilon}{8K_7}\min\{\frac{2\varepsilon}{\delta_{\max}}, 1\}$, we have

$$Pred_k \geq \frac{\bar{\sigma}\varepsilon}{8}\min\{\frac{2\varepsilon}{\delta_{\max}}, 1\}\delta_k + r_k[||h_k||^2 - ||h_k + \nabla h_k^T s_k||^2].$$

Take $K_9 = \min\left\{\frac{K_3\varepsilon}{24}\min\{\frac{2\varepsilon}{3\delta_{max}}, 1\}\ ,\ \frac{\bar{\sigma}\varepsilon}{8}\min\{\frac{2\varepsilon}{\delta_{\max}}, 1\}\right\}$, the result follows. □

From the above lemma, we can easily see that, at any iteration at which either $||Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k)|| \geq \frac{\varepsilon}{2} > 0$ or $||\nabla g_k W_k g_k|| \geq \frac{\varepsilon}{2} > 0$ and $||h_k|| \leq \alpha\delta_k$, where α is given by (6.3), there is no need to increase the value of the penalty parameter. *i.e.*, r_k is increased only when $||h_k|| \geq \alpha\delta_k$.

7. 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. Equivalently, the condition $\frac{Ared_k}{Pred_k} \geq \eta_1$ will eventually be satisfied. To state this result, we need to introduce one more notation. The j^{th} trial iterate of iteration k is denoted by k^j.

LEMMA 7.1. *Assume A_1-A_5. If $||h_k|| \geq \varepsilon$ where ε is any positive constant, then an acceptable step is found after finitely many trials. i.e., the condition $\frac{Ared_{k^j}}{Pred_{k^j}} \geq \eta_1$ will be satisfied for some finite j.*

Proof. Since $||h_k|| \geq \varepsilon$, then from (5.3) and (5.11), we have

$$\left|\frac{Ared_k}{Pred_k} - 1\right| = \frac{|\ Ared_k - Pred_k\ |}{Pred_k} \leq \frac{2K_6\delta_k^2}{K_2\varepsilon\min\{\varepsilon, \delta_k\}}.$$

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

$$\left|\frac{Ared_{k^j}}{Pred_{k^j}} - 1\right| \leq \frac{2K_6\delta_{k^j}}{K_2\varepsilon}.$$

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 7.2. *Assume A_1-A_5. If at a given iteration k, the j^{th} trial step satisfies*

$$||s_{k^j}|| \leq \min\{\frac{(1-\eta_1)K_2}{4K_6}, 1\}||h_k||, \tag{7.1}$$

then it must be accepted.

Proof. We prove this lemma by contradiction. Assume that the step s_{k^j} is rejected and Inequality (7.1) holds. Then, from (5.3), (5.11), and (7.1), we have

$$(1-\eta_1) < \frac{|\, Ared_{k^j} - Pred_{k^j} \,|}{Pred_{k^j}} < \frac{2K_6\|s_{k^j}\|^2}{K_2\|h_k\|\|s_{k^j}\|} \leq \frac{(1-\eta_1)}{2}.$$

This gives a contradiction and proves the lemma. □

LEMMA 7.3. *Assume A_1-A_5. For all trial steps j of any iteration k generated by the algorithm, δ_{k^j} satisfies*

$$\delta_{k^j} \geq \min\{\frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\eta_1)K_2}{4K_6}, \alpha_1\}\|h_k\|, \tag{7.2}$$

where b_1 is a positive constant independent of k or j.

Proof. Consider any iterate k^j. If the previous step was accepted; *i.e.* if $j = 1$, then $\delta_k \geq \delta_{min}$. Take $b_1 = \sup_{x\in\Omega}\|h_k\|$, we can write

$$\delta_k \geq \delta_{min} \geq \frac{\delta_{min}}{b_1}\|h_k\|, \tag{7.3}$$

Therefore, (7.2) holds in this case.

Now assume that $j > 1$, then there exists at least one rejected trial step. For all rejected trial steps, we have from Lemma 7.2

$$\|s_{k^i}\| > \min\{\frac{(1-\eta_1)K_2}{4K_6}, 1\}\|h_k\|,$$

for all $i = 1, 2, ...j-1$. Since s_{k^i} is a rejected trial step, then from the way of updating the radius of trust region (see Algorithm 2.1) and using the above inequality, we have

$$\delta_{k^j} = \alpha_1\|s_{k^{j-1}}\| > \alpha_1 \min\{\frac{(1-\eta_1)K_2}{4K_6}, 1\}\|h_k\|.$$

Inequality (7.3) and the above inequality prove the lemma. □

LEMMA 7.4. *Assume A_1-A_5. For all k^j at which the penalty parameter r_{k^j} is increased, there exists a positive constant K_{10} that does not depend on k or j, such that*

$$r_{k^j}\|h_k\| \leq K_{10}. \tag{7.4}$$

Proof. The proof follows from Lemma 6.2 and inequality (7.2). □

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

LEMMA 7.5. *Assume A_1-A_5. If $\|h_k\| \geq \varepsilon$, where $\varepsilon > 0$, then there exists a positive constant K_{11} that depends on ε but does not depend on k such that*

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

Proof. Using (7.2) and taking

$$K_{11} = \varepsilon \min\{\frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\eta_1)K_2}{4K_6}, \alpha_1\}, \tag{7.5}$$

the proof follows directly. □

LEMMA 7.6. *Assume A_1-A_5. If $r_k \to \infty$, then*

$$\lim_{k_i \to \infty} ||h_{k_i}|| = 0, \tag{7.6}$$

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

Proof. The proof follows directly from Lemma 7.4. □

8. Global Convergence Theory. In this section, we prove our main global convergence results for our trust-region algorithm for solving Problem (1.1). In the following lemma, we prove that the sequence $\{||h_k||\}$ converges to zero.

THEOREM 8.1. *Assume A_1-A_5. Then the sequence of iterates generated by the algorithm satisfies*

$$\lim_{k \to \infty} ||h_k|| = 0. \tag{8.1}$$

Proof. Assume that $\limsup_{k\to\infty} ||h_k|| \geq \varepsilon > 0$. This implies the existence of an infinite subsequence of indices $\{k_j\}$ indexing iterates that satisfy $||h_{k_j}|| \geq \frac{\varepsilon}{2}$. From Lemma 7.1, 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.

We consider two cases: first, if $\{r_k\}$ is unbounded, then there exists an infinite number of iterates $\{k_i\}$ at which the penalty parameter r_k is increased. From Lemma 7.6, for k sufficiently large, the two sequences $\{k_i\}$ and $\{k_j\}$ do not have common elements. Let k_α and k_β be two consecutive iterates at which the penalty parameter r_k is increased and $k_\alpha < k < k_\beta$, where $k \in \{k_j\}$. The penalty parameter r_k is the same for all iterates that lie between k_α and k_β. Since all the iterates of $\{k_j\}$ are acceptable, then for all $k \in \{k_j\}$,

$$\Phi_k - \Phi_{k+1} = Ared_k \geq \eta_1 Pred_k.$$

From Inequality (5.3) and the above inequality, we can write

$$\frac{\Phi_k - \Phi_{k+1}}{r_k} \geq \eta_1 \frac{K_2}{2} ||h_k|| \min\{||h_k||, \delta_k\}.$$

Summing over all acceptable iterates that lie between k_α and k_β, we have

$$\sum_{k=k_\alpha}^{k_\beta - 1} \frac{\Phi_k - \Phi_{k+1}}{r_k} \geq \frac{\eta_1 K_2 \varepsilon}{4} \min\{\bar{K}_{11}, \frac{\varepsilon}{2}\},$$

where $\bar{K}_{11}$ is as K_{11} in (7.5), with ε is replaced by $\frac{\varepsilon}{2}$. Hence,

$$\frac{L_{k_\alpha} - L_{k_\beta}}{r_{k_\alpha}} + [||h_{k_\alpha}||^2 - ||h_{k_\beta}||^2] \geq \frac{\eta_1 K_2 \varepsilon}{4} \min\{\bar{K}_{11}, \frac{\varepsilon}{2}\},$$

where L is given by (1.10). Since $r_k \to \infty$, then for k_α sufficiently large, we have $\frac{|L_{k_\alpha} - L_{k_\beta}|}{r_{k_\alpha}} < \frac{\eta_1 K_2 \varepsilon}{8} \min\{\bar{K}_{11}, \frac{\varepsilon}{2}\}$. Therefore,

$$||h_{k_\alpha}||^2 - ||h_{k_\beta}||^2 \geq \frac{\eta_1 K_2 \varepsilon}{8} \min\{\bar{K}_{11}, \frac{\varepsilon}{2}\}.$$

But this leads to a contradiction with Lemma 7.6 unless $\varepsilon = 0$.

Now, consider the second case. If $\{r_k\}$ is bounded, then there exists an integer $\tilde{k}$ such that for all $k \geq \tilde{k}$, $r_k = \tilde{r}$. Hence from Inequality (5.3), we have for any $\hat{k} \in \{k_j\}$ and $\hat{k} \geq \tilde{k}$

$$\begin{aligned} Pred_{\hat{k}} &\geq \frac{\tilde{r}}{2} K_2 \|h_{\hat{k}}\| \min\{\delta_{\hat{k}}, \|h_{\hat{k}}\|\} \\ &\geq \frac{\varepsilon}{4} \tilde{r} K_2 \min\{\frac{\varepsilon}{2\delta_{max}}, 1\}\delta_{\hat{k}} \end{aligned} \tag{8.2}$$

Since all the iterates of $\{k_j\}$ are acceptable, then for any $\hat{k} \in \{k_j\}$, we have

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} = Ared_{\hat{k}} \geq \eta_1 Pred_{\hat{k}}.$$

Hence, from Inequality (8.2) and the above inequality we have

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} \geq \frac{\eta_1 \varepsilon \tilde{r} K_2}{4} \min\{\frac{\varepsilon}{2\delta_{max}}, 1\}\delta_{\hat{k}}.$$

Using Lemma 7.5 and the above inequality, we have

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} \geq \frac{\eta_1 \varepsilon \tilde{r} K_2}{4} \min\{\frac{\varepsilon}{2\delta_{max}}, 1\}\bar{K}_{11} > 0,$$

where $\bar{K}_{11}$ is as above. This gives a contradiction with the fact that $\{\Phi_k\}$ is bounded when $\{r_k\}$ is bounded. Hence, in both cases, we have a contradiction. Thus the supposition is not correct and the theorem is proved. □

THEOREM 8.2. *Assume A_1-A_5. Then the sequence of iterates generated by the algorithm satisfies*

$$\liminf_{k\to\infty} [\ \|Z_k^T \nabla_x l_k\| + \|\nabla g_k W_k g_k\|\ ] = 0. \tag{8.3}$$

Proof. First, we prove that

$$\liminf_{k\to\infty} [\ \|\ Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k\ )\ \| + \|\ \nabla g_k W_k g_k\ \|\ ] = 0. \tag{8.4}$$

We prove (8.4) by contradiction. Suppose that, for all k, $\|\ Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k)\ \| + \|\nabla g_k W_k g_k\| > \varepsilon$.

Assume that there exists an infinite subsequence $\{k_i\}$ such that $\|h_{k_i}\| > \alpha\delta_{k_i}$, where α is any positive constant. For later use of α, we require it to satisfy (6.3). Since $\|h_k\| \to 0$, we have

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

Consider any iterate $k^j \in \{k_i\}$. There are two cases to consider. First, consider the case where the sequence $\{r_k\}$ is unbounded. For the rejected trial step $j-1$ of iteration k, we have $\|h_k\| > \alpha\delta_{k^j} = \alpha_1\alpha\|s_{k^{j-1}}\|$. Using Inequalities (5.3) and (5.10) and the fact that the trial step $s_{k^{j-1}}$ was rejected, we have

$$\begin{aligned} (1-\eta_1) &\leq \frac{|Ared_{k^{j-1}} - Pred_{k^{j-1}}|}{Pred_{k^{j-1}}} \\ &\leq \frac{2K_5[\|s_{k^{j-1}}\| + r_{k^{j-1}}(\|s_{k^{j-1}}\|^2 + \|s_{k^{j-1}}\|\|h_k\|)]}{r_{k^{j-1}} K_2 \min(\alpha_1\alpha, 1)\|h_k\|} \\ &\leq \frac{2K_5}{r_{k^{j-1}} K_2 \alpha_1\alpha \min(\alpha_1\alpha, 1)} + \frac{2K_5(1+\alpha_1\alpha)}{K_2\alpha_1\alpha\min(\alpha_1\alpha, 1)}\|s_{k^{j-1}}\|. \end{aligned}$$

Because $\{r_k\}$ is unbounded, there exists an iterate $\hat{k}$ sufficiently large such that for all $k \geq \hat{k}$, we have

$$r_{k^{j-1}} > \frac{4K_5}{K_2\alpha_1\alpha\min(\alpha_1\alpha, 1)(1-\eta_1)}.$$

This implies that for all $k \geq \hat{k}$,

$$\|s_{k^{j-1}}\| \geq \frac{K_2\alpha_1\alpha\min(\alpha_1\alpha, 1)(1-\eta_1)}{4K_5(1+\alpha_1\alpha)}.$$

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

$$\delta_{k^j} = \alpha_1\|s_{k^{j-1}}\| \geq \frac{K_2\alpha_1^2\alpha\min(\alpha_1\alpha, 1)(1-\eta_1)}{4K_5(1+\alpha_1\alpha)}.$$

This gives a contradiction. So δ_{k^j} can not go to zero in this case.

Second, consider the case when the sequence $\{r_k\}$ is bounded. There exists an integer $\bar{k}$ and a constant $\bar{r}$ such that for all $k \geq \bar{k}$, $r_k = \bar{r}$. Let $k \geq \bar{k}$ and consider a trial step j of iteration k, such that $\|h_k\| > \alpha\delta_{k^j}$.

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

$$\|h_{k^l}\| > \alpha\delta_{k^l}, \tag{8.5}$$

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

$$(1-\eta_1) \leq \frac{|Ared_{k^l} - Pred_{k^l}|}{Pred_{k^l}} \leq \frac{2K_6\|s_{k^l}\|}{K_2\min(\alpha, 1)\|h_k\|}.$$

Hence,

$$\begin{aligned}\delta_{k^j} = \alpha_1\|s_{k^{j-1}}\| &\geq \frac{\alpha_1K_2\min(\alpha, 1)(1-\eta_1)\|h_k\|}{2K_6} \geq \frac{\alpha_1K_2\min(\alpha, 1)(1-\eta_1)\alpha}{2K_6}\delta_{k^1} \\ &\geq \frac{\alpha_1K_2\min(\alpha, 1)(1-\eta_1)\alpha}{2K_6}\delta_{\min}.\end{aligned}$$

Hence δ_{k^j} is bounded in this case too. If $j > 1$ and (8.5) does not hold for all l, there exists an integer m such that (8.5) holds for $l = m+1, ..., j$ and

$$\|h_{k^l}\| \leq \alpha\delta_{k^l}, \tag{8.6}$$

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

$$\delta_{k^j} \geq \frac{\alpha_1K_2\min(\alpha, 1)(1-\eta_1)}{2K_6}\|h_k\| \geq \frac{\alpha_1K_2\min(\alpha, 1)(1-\eta_1)\alpha}{2K_6}\delta_{k^{m+1}}. \tag{8.7}$$

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

$$\delta_{k^{m+1}} \geq \alpha_1\|s_{k^m}\|. \tag{8.8}$$

Now, using (8.6), Lemma 6.3, and the fact that the trial steps s_{k^m} is rejected, we can write

$$(1-\eta_1) \leq \frac{|Ared_{k^m} - Pred_{k^m}|}{Pred_{k^m}} \leq \frac{2K_6\bar{r}\|s_{k^m}\|}{K_9}.$$

This implies

$$||s_{k^m}|| \geq \frac{K_9(1-\eta_1)}{2K_6\bar{r}}.$$

This implies that, $||s_{k^m}||$ is bounded. This fact together with (8.7) and (8.8) 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^j sufficiently large, all the iterates satisfy $||h_k|| \leq \alpha\delta_{k^j}$. This implies using Lemma 6.3 that there is no need to increase the value of the penalty parameter. So, $\{r_k\}$ is bounded. Letting $k^j \geq \bar{k}$ and using Lemma 6.3, we have

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

As k goes to infinity the above inequality implies that

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

This implies that the radius of the trust region is not bounded below. But this leads to a contradiction because if we consider an iteration $k^j > \bar{k}$ and 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 the rejected trial step $s_{k^{j-1}}$, using Lemmas 5.8 and 6.3, we must have

$$(1-\eta_1) < \frac{\bar{r}K_6||s_{k^{j-1}}||^2}{K_9\delta_{k^{j-1}}}.$$

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

$$\delta_{k^j} = \alpha_1||s_{k^{j-1}}|| > \frac{\alpha_1 K_9(1-\eta_1)}{\bar{r}K_6}.$$

Hence δ_{k^j} is bounded. But this contradicts (8.9). The supposition is wrong. Hence,

$$\liminf_{k\to\infty} [\ ||Z_k^T(\nabla_x l_k + \bar{\rho}\nabla g_k W_k g_k)|| + ||\nabla g_k W_k g_k||\ ] \ = \ 0.$$

But this also implies (8.3). This completes the proof of the theorem. □

From the above two theorems, we conclude that, given any $\varepsilon > 0$, the algorithm terminates because $||Z_k^T\nabla l|| + ||\nabla g_k W_k g_k|| + ||h_k|| < \varepsilon$, for some finite k.

9. Concluding Remarks. We introduced a new trust-region algorithm for solving the general nonlinear programming problem. This algorithm can be viewed as an extension of Byrd and Omojokun's trust-region algorithm for solving the equality constrained optimization problem. The algorithm handles inequality constraints in a fashion similar to the approach of Dennis, El-Alem, and Williamson for treating the active constraints. At every iteration, the step is computed by solving two simple trust-region subproblems similar to those for unconstrained optimization.

We proved that the algorithm is globally convergent in the sense that, in the limit, a subsequence of the iteration sequence generated by the algorithm satisfies either the Fritz John conditions or the KKT conditions.

REFERENCES

[1] N. Alexandrov and J. Dennis. Multi-level algorithms for nonlinear optimization. Technical Report 94-24, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251, June (1994).

[2] R. Byrd. Robust trust-region methods for nonlinearly constrained optimization. A talk presented at the Second SIAM Conference on Optimization, Houston, TX (1987).

[3] R. Byrd, J. Gilbert, and J. Nocedal. A trust-region method based on interior point techniques for nonlinear programming. Technical Report, INRIA, 2896, (1988).

[4] M. Celis, J. Dennis, and R. Tapia. A trust-region strategy for nonlinear equality constrained optimization. In Numerical Optimization 1984. SIAM Philadelphia, Pennsylvania, (1985).

[5] J. Dennis, M. El-Alem, and M. Maciel. A global convergence theory for general trust-region-based algorithms for equality constrained optimization. SIAM J. Optimization, Vol. 7, No. 1:177-207 (1997).

[6] J. Dennis, M. El-Alem, and K. Williamson. A trust-region approach to nonlinear systems of equalities and inequalities. Technical Report TR94-38, Dept. of Computational and Applied Mathematics, Rice University, Houston, TX 77251 (1994).

[7] J. Dennis, M. Heinkenschloss, and L. Vicente. Trust-region interior-point SQP algorithms for a class of nonlinear programming problems. Tech. Rep. TR94-45, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77251 (1994).

[8] 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, Texas, (1988).

[9] 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).

[10] 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).

[11] M. El-Alem. Convergence to a second order point of a trust-region algorithm with a a non-monotonic penalty parameter scheme for constrained optimization.. J. Optim. Th. and Appl., 91(1):61-79, October (1996).

[12] A. El-Bakry, R. Tapia, T. Tsuchiya, and Y. Zahng. On the formulation and theory of the Newton interior-point method for nonlinear programming. J. Optim. Theory Appl., Vol 89, No. 3: 507-541, June (1996).

[13] A. Fiacco and G. McCormick. Nonlinear programming: Sequential unconstrained minimization techniques. John Wiley and Sons, New York, (1968).

[14] P. Gill, W. Murray, M. Saunders, and M. Wright. Some theoretical properties of an augmented Lagrangian merit function. Report SOL 86-6, Stanford University, (1986).

[15] S. Han. A globally convergent method for nonlinear programming. JOTA, 22:297-309, (1977).

[16] O. Mangasarian Nonlinear Programming. McGraw-Hill Book Company, New York (1969)

[17] J. Moré Recent developments in algorithms and software for trust-region methods. In A. Bachem, M. Grotschel, B. Korte, editors, Mathematical Programming, 258-287, Springer-Verlag, New York, (1983).

[18] E. Omojokun. Trust-region strategies for optimization with nonlinear equality and inequality constraints. PhD thesis, Department of Computer Science, University of Colorado, Boulder, Colorado, (1989).

[19] T. Plantenga. Large-Scale nonlinear constrained optimization using trust region. PhD thesis, Department of Electrical and Comp. Eng., Northwestern University, Evanston, IL, (1995).

[20] M. Powell. Convergence properties of a class of minimization algorithms. In O. L.Mangasarian, R. R. Meyer, and S. M. Robinson, editors, Nonlinear Programming 2, 1:27, Academic Press, New York, (1975).

[21] M. Powell. Algorithms for nonlinear constraints that used Lagrangian functions. Mathematical Programming, North-Holland Publishing Company, 14:224-248, (1978).

[22] K. Schittkowski. On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function. Math. Operations Forchung U.Statistic Ser., Optimization, 14:197-216, (1983).

[23] L. 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, (1996).

[24] R. Wilson. A simplicial algorithm concave programming. PhD thesis, Graduate School of Business Administration, Harvard University, Boston, (1963).

[25] S. Wright. Primal-Dual Interior-Point Methods. SIAM publications, Philadelphia, PA, (1997).

[26] Y. Yuan. On the convergence of a new trust region algorithm. Numer. Math. 70:515-539 (1995).

[27] J. Zhang and D. Zhu. Projected quasi-Newton algorithm with trust-region for constrained optimization. J. Opt. Th. and Appl., 67:369-393, (1990).