Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 14, No. 1, pp. 125–159, 2018

AN ACTIVE-SET INTERIOR-POINT TRUST-REGION ALGORITHM

Bothina El-Sobky

Abstract: In this work, an active-set method is used to transform a general nonlinear programming problem with bounds on the variables to an equality constrained optimization problem with bound on the variables. By using a Coleman-Li strategy the iterates which are generated by the proposed algorithm are strictly feasible. An interior-point Newton method is used and a trust-region globalization strategy is added to the algorithm to insure global convergence. A reduced Hessian technique is used to overcome the difficulty of having an infeasible trust-region subproblem. A global convergence analysis for this algorithm is presented under credible assumptions. Preliminary numerical results are reported.

Key words: *interior-point, Coleman-Li method, active-set, reduced Hessian technique, trust region, global convergence*

Mathematics Subject Classification: *49N35, 49N10, 93D52, 93D22, 65K05*

1 Introduction

In this paper, we consider the following general nonlinear programming problem with bounds on the variables

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & h(x) = 0, \\ & g(x) \leq 0, \\ & \alpha \leq x \leq \beta, \end{array} \tag{1.1}$$

where $\alpha \in \{\Re \bigcup \{-\infty\}\}^n$, $\beta \in \{\Re \bigcup \{+\infty\}\}^n$, and $\alpha < \beta$. The functions $f : \Re^n \to \Re$, $h : \Re^n \to \Re^p$, and $g : \Re^n \to \Re^m$ are twice continuously differentiable. We assume that $p < n$ and no restriction is assumed on m.

In this paper, we use the active-set method in [5] to transform the above problem to an equality constrained optimization problem with bound variables. The characteristic of the proposed active set is that it is identified and updated naturally by the step. See ([9,10,12]).

Motivated by the impressive computational performance of the primal dual interior-point method for linear programming, authors in [6, 22] using the Coleman-Li scaling matrix [2], proposed a primal interior-point algorithm for solving nonlinear programming problems having a special structure. In particular, their algorithm is designed for solving a special nonlinear programming where the vector of primal variables x is naturally divided into a vector of state variables and a vector of control variables. They proved several global and local convergence results for their algorithm. In this paper, we use the Coleman-Li scaling matrix [2] to propose the interior-point trust-region algorithm for solving nonlinear programming Problem (1.1). The Coleman-Li scaling matrix was first introduced by [2] for unconstrained optimization problem and used by [6,8] for equality constrained optimization problem.

As we know a trust-region method is a well-accepted technique in nonlinear optimization to assure global convergence and is more robust when they deal with rounding errors, so we used it in this paper. One of the advantages of trust-region method is that it does not require the objective function of the model to be convex. However, in traditional trust-region method, after solving a trust-region subproblem, we need to use some criterion to check if the trial step is acceptable. If not, the subproblem must be resolved with a reduced trust-region radius. For more details see ([11, 16, 19, 23, 24, 26]).

If the trust-region constraint is simply added to the sequential quadratic subproblem of the equality constrained optimization problem, the resulting trust-region subproblem may be infeasible, because there may be no intersecting points between the trust-region constraint and the hyperplane of the linearized constraints. Even if they intersect, there is no guarantee that this will remain true if the trust-region radius is decreased. For more details see [6].

A reduced Hessian is a successful approach to overcoming the difficulty of having an infeasible trust-region subproblem. The approach was suggested by [1,17] and used by [4,17–19], [27]. 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. One of the advantages of this approaches, the two subproblems are similar to the trust-region subproblem for the unconstrained case. Under credible assumptions, a convergence theory for the proposed interior-point trust-region algorithm is introduced.

In this paper, we use the symbol $f_k = f(x_k)$, $h_k = h(x_k)$, $g_k = g(x_k)$, $\ell_k = \ell(x_k, \mu_k)$, $\nabla_x \ell_k = \nabla_x \ell(x_k, \mu_k)$, and so on. We use the notation $x_k^{(i)}$ to denote the ith component of the vector x_k, and so on. Finally, all the norms used in this paper are ℓ_2-norms.

The paper is organized as follows. Active-set and Newton's method are described in in Section 2. A detailed description of the main steps of the algorithm are presented in Section 3. Section 4 is devoted to analysis of the global convergence of the proposed algorithm. In Section 5, numerical results are reported. Finally, Section 6 contains concluding remarks.

2 An Active-Set and Newton's Method

Following the active set method in [5], we define a 0-1 diagonal matrix $V(x) \in \Re^{m\times m}$, whose diagonal entries are

$$v_i(x) = \begin{cases} 1 & \text{if } g_i(x) \geq 0, \\ 0 & \text{if } g_i(x) < 0. \end{cases} \tag{2.1}$$

Using the above matrix, Problem (1.1) is converted to the following equality constrained optimization problem

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & h(x) = 0, \\ & g(x)^T V(x) g(x) = 0, \\ & \alpha \leq x \leq \beta. \end{aligned}$$

Using penalty method, the above problem can be written as follows

$$\begin{aligned} \text{minimize} \quad & f(x) + \frac{\rho}{2}\|V(x)g(x)\|^2 \\ \text{subject to} \quad & h(x) = 0, \\ & \alpha \leq x \leq \beta, \end{aligned} \tag{2.2}$$

where ρ is a positive parameter. Let

$$\ell(x, \mu) = f(x) + \mu^T h(x), \tag{2.3}$$

and

$$\ell(x, \mu; \rho) = \ell(x, \mu) + \frac{\rho}{2}\|V(x)g(x)\|^2, \tag{2.4}$$

where μ is a Lagrange multiplier vector associated with equality constraint $h(x)$.

The Lagrangian function associated with Problem (2.2) is defined as follows

$$L(x, \mu, \lambda_\alpha, \lambda_\beta) = \ell(x, \mu; \rho) - \lambda_\alpha^T (x - \alpha) - \lambda_\beta^T (\beta - x), \tag{2.5}$$

where the vectors λ_α, and λ_β are Lagrange multiplier vectors associated with inequality constraints $(x - \alpha)$ and $(\beta - x)$ respectively.

The first-order necessary conditions for a point x_* to be a local minimizer of Problem (2.2) are the existence of multipliers $\mu_* \in \Re^p$, $\lambda_{\alpha *} \in \Re^n_+$, and $\lambda_{\beta_*} \in \Re^n_+$, such that $(x_*, \mu_*, \lambda_{\alpha *}, \lambda_{\beta_*})$ satisfies

$$\nabla_x \ell(x_*, \mu_*; \rho_*) - \lambda_{\alpha *} + \lambda_{\beta_*} \;=\; 0, \tag{2.6}$$

$$h(x_*) \;=\; 0, \tag{2.7}$$

$$\alpha \leq x_* \;\leq\; \beta, \tag{2.8}$$

and for all e corresponding to $x^{(e)}$ with finite bound, we have

$$\lambda_{\alpha *}^{(e)} (x_*^{(e)} - \alpha^{(e)}) \;=\; 0, \tag{2.9}$$

$$\lambda_{\beta_*}^{(e)} (\beta^{(e)} - x_*^{(e)}) \;=\; 0, \tag{2.10}$$

where

$$\nabla_x \ell(x_*, \mu_*; \rho_*) = \nabla_x \ell(x_*, \mu_*) + \rho_* \nabla g(x_*) V(x_*) g(x_*), \tag{2.11}$$

and

$$\nabla_x \ell(x_*, \mu_*) = \nabla f(x_*) + \nabla h(x_*) \mu_*. \tag{2.12}$$

Motivated by the Coleman-Li scaling matrix [8], we define a diagonal matrix $A(x)$ whose diagonal elements are

$$a^{(e)}(x) = \begin{cases} \sqrt{(x^{(e)} - \alpha^{(e)})}, & \text{if } (\nabla_x \ell(x, \mu; \rho))^{(e)} \geq 0 \text{ and } \alpha^{(e)} > -\infty, \\ \sqrt{(\beta^{(e)} - x^{(e)})}, & \text{if } (\nabla_x \ell(x, \mu; \rho))^{(e)} < 0 \text{ and } \beta^{(e)} < +\infty\ , \\ 1, & \text{otherwise.} \end{cases} \tag{2.13}$$

Let $\boldsymbol{F} = \{x : \alpha \leq x \leq \beta\}$ and $int(\boldsymbol{F}) = \{x : \alpha < x < \beta\}$. Using the scaling matrix $A(x)$, the first-order necessary conditions for the point x_* to be a local minimizer of Problem (1.1) are that $x_* \in \boldsymbol{F}$ and there exists a Lagrange multiplier vector μ_*, such that (x_*, μ_*) solves the following nonlinear system

$$\begin{aligned} A^2(x)\nabla_x \ell(x,\mu;\rho) &= 0, & (2.14)\\ h(x) &= 0. & (2.15)\end{aligned}$$

Any point (x_*, μ_*) that satisfies the conditions (2.14)-(2.15) is called a Karush-Kuhn-Tucker point or a KKT point. For more details see [13].

System (2.14)-(2.15) is continuous but not everywhere differentiable. The non-differentiability occurs at two cases:

i) If $a^{(e)}(x) = 0$, then these points are avoided by restricting $x \in int\boldsymbol{F}$.

ii) If a variable $x^{(e)}$ has a finite lower bound and an infinite upper bound (or vice-verse) and $(\nabla_x \ell(x,\mu;\rho))^{(e)} = 0$. But these points are not significant. So, we define a vector

$$\eta^{(e)}(x) = \frac{\partial (a^{(e)}(x))^2}{\partial x^{(e)}}, \quad e = 1, \ldots, n,$$

such that $\eta^{(e)}(x) = 0$ when $(\nabla_x \ell(x,\mu;\rho))^{(e)} = 0$. This is equivalent to

$$\eta^{(e)}(x) = \begin{cases} 1, & \text{if} \quad (\nabla_x \ell(x,\mu;\rho))^{(e)} \geq 0 \quad \text{and} \quad \alpha^{(e)} > -\infty, \\ -1, & \text{if} \quad (\nabla_x \ell(x,\mu;\rho))^{(e)} < 0 \quad \text{and} \quad \beta^{(e)} < \infty, \\ 0, & \text{otherwise}. \end{cases} \tag{2.16}$$

Applying Newton's method on the nonlinear system (2.14)-(2.15), then we have

$$\begin{aligned} &[A^2(x)\nabla_x^2 \ell(x,\mu;\rho) + \text{diag}(\nabla_x \ell(x,\mu;\rho))\text{diag}(\eta(x))]\Delta x \\ &\qquad + A^2(x)\nabla h(x)\Delta\mu = -A^2(x)\nabla_x \ell(x,\mu;\rho), & (2.17)\\ &\nabla h(x)^T \Delta x = -h(x), & (2.18)\end{aligned}$$

where

$$\nabla_x^2 \ell(x,\mu;\rho) = H + \rho \nabla g(x) V(x) \nabla g(x)^T, \tag{2.19}$$

and H is the Hessian of the Lagrangian function (2.3) or an approximation to it.

Restricting $x \in int(\boldsymbol{F})$ makes $A(x)$ necessarily nonsingular. Therefore, multiplying both side of Equation (2.17) by $A^{-1}(x)$, and put $\Delta x = A(x)s$ in both Equation (2.17)and (2.18), we have

$$\begin{aligned} &[A(x)\nabla_x^2 \ell(x,\mu;\rho)A(x) + \text{diag}(\nabla_x \ell(x,\mu;\rho))\text{diag}(\eta(x))]s \\ &\qquad + A(x)\nabla h(x)\Delta\mu = -A(x)\nabla_x \ell(x,\mu;\rho), & (2.20)\\ &(A(x)\nabla h(x))^T s = -h(x). & (2.21)\end{aligned}$$

The above system shares the advantages and the disadvantages of Newton's method. From the good side of Newton's method, it converges quadratically to stationary point (x_*, μ_*) under reasonable assumptions. From the bade side of Newton's method, it may not converge at all if the starting point is far away from the solution. Trust-region approach is a very successful approach to ensure global convergence from any starting point. To add trust region constraint we have to rewrite the above system as a minimization problem. An equivalent problem is the following quadratically programming problem

$$\begin{aligned} &\text{minimize} \quad \ell(x,\mu;\rho) + (A(x)\nabla_x \ell(x,\mu;\rho))^T s + \frac{1}{2} s^T B s \\ &\text{subject to} \quad h(x) + (A(x)\nabla h(x))^T s = 0, \end{aligned} \tag{2.22}$$

where

$$B = G(x) + \rho A(x)\nabla g(x) V(x) \nabla g(x)^T A(x), \tag{2.23}$$

and

$$G(x) = A(x)H(x)A(x) + \mathrm{diag}(\nabla_x \ell(x,\mu;\rho))\mathrm{diag}(\eta(x)). \tag{2.24}$$

That is, the point (x_*, μ_*) that satisfies the first order necessary conditions of Problem (2.22) will satisfy the first order necessary conditions of Problem (1.1).

In the following section, we present main steps of the proposed interior-point trust-region algorithm for solving Problem (1.1).

3 Outline of the Proposed Algorithm

This section is devoted to presenting the outline for the main steps of interior-point trust-region algorithm.

3.1 Evaluating s_k

Consider the following trust-region sub problem

$$\begin{aligned} &\text{minimize} \quad \ell(x_k,\mu_k;\rho_k) + (A_k \nabla_x \ell(x_k,\mu_k;\rho_k))^T s + \frac{1}{2} s^T B_k s \\ &\text{subject to} \quad h_k + (A_k \nabla h_k)^T s = 0, \\ &\qquad\qquad \|s\| \le \delta_k, \end{aligned} \tag{3.1}$$

where δ_k is the radius of the trust-region. To evaluate the step s_k, a reduced Hessian method is used for overcoming the difficulty of having an infeasible trust-region subproblem (3.1). In this method, the step s_k is decomposed into two orthogonal components; the normal component s_k^n and the tangential component $s_k^t = Z_k \bar{s}_k^t$ where Z_k is a matrix whose columns form an orthonormal basis for the null space of $(A_k \nabla h_k)^T$. The step s_k has the form $s_k = s_k^n + Z_k \bar{s}_k^t$.

To compute the normal component s_k^n, we solve the following trust-region subproblem

$$\begin{aligned} &\text{minimize} \quad \tfrac{1}{2}\|h_k + (A_k \nabla h_k)^T s^n\|^2 \\ &\text{subject to} \quad \|s^n\| \le \zeta \delta_k, \end{aligned} \tag{3.2}$$

for some $\zeta \in (0,1)$.

It is not necessary to obtain a very accurate approximation to the solution of subproblem (3.2). Instead any approximation to the solution of subproblem (3.2) can be used as long as the normal predicted decrease obtained by the normal component s_k^n is greater than or equal to a fraction of the normal predicted decrease obtained by the normal Cauchy step s_k^{ncp}. That is

$$\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T s_k^n\|^2 \geq \vartheta_1 \{\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T s_k^{ncp}\|^2\}, \tag{3.3}$$

for some $\vartheta_1 \in (0,1]$. The normal Cauchy step s_k^{ncp} is defined as

$$s_k^{ncp} = -t_k^{ncp} A_k \nabla h_k h_k, \tag{3.4}$$

where the parameter t_k^{ncp} is given by

$$t_k^{ncp} = \begin{cases} \dfrac{\|A_k \nabla h_k h_k\|^2}{\|(A_k \nabla h_k)^T A_k \nabla h_k h_k\|^2} & \text{if } \dfrac{\|A_k \nabla h_k h_k\|^3}{\|(A_k \nabla h_k)^T A_k \nabla h_k h_k)\|^2} \leq \delta_k \\ & \text{and } \|(A_k \nabla h_k)^T A_k \nabla h_k h_k)\| > 0, \\ \dfrac{\delta_k}{\|A_k \nabla h_k h_k\|} & \text{otherwise.} \end{cases} \tag{3.5}$$

Let $q(A_k s)$ be the quadratic form of the function (2.4) and defined as follows

$$q(A_k s) = \ell(x_k, \mu_k; \rho_k) + (A_k \nabla_x \ell(x_k, \mu_k; \rho_k))^T s + \frac{1}{2} s^T B_k s. \tag{3.6}$$

Then $\nabla q_k(A_k s_k^n) = A_k \nabla_x \ell(x_k, \mu_k; \rho_k) + B_k s_k^n$.

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

$$\begin{aligned} &\text{minimize} \quad [Z_k^T \nabla q_k(A_k s_k^n) + B_k s_k^n]^T \bar{s}^t + \frac{1}{2} \bar{s}^{t^T} Z_k^T B_k Z_k \bar{s}^t \\ &\text{subject to} \quad \|Z_k \bar{s}^t\| \leq \Delta_k, \end{aligned} \tag{3.7}$$

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

A tangential predicted decrease which is obtained by the tangential component s_k^t is given by

$$Tpred_k(\bar{s}_k^t) = q_k(A_k s_k^n) - q_k(A_k(s_k^n + Z_k \bar{s}_k^t)). \tag{3.8}$$

To solve subproblem (3.7), any method can be used as long as $Tpred_k(\bar{s}_k^t)$ is greater than or equal to a fraction of the tangential predicted decrease $Tpred_k(\bar{s}_k^{tcp})$ which is obtained by the tangential Cauchy step $\bar{s}_k^{tcp}$. That is

$$Tpred_k(\bar{s}_k^t) \geq \vartheta_2 \ \ Tpred_k(\bar{s}_k^{tcp}), \tag{3.9}$$

for some $\vartheta_2 \in (0,1]$. The tangential Cauchy step $\bar{s}_k^{tcp}$ is defined as follows

$$\bar{s}_k^{tcp} = -t_k^{tcp} Z_k^T \nabla q_k(A_k s_k^n), \tag{3.10}$$

where the parameter t_k^{tcp} is given by

$$t_k^{tcp} = \begin{cases} \dfrac{\|Z_k^T \nabla q_k(A_k s_k^n)\|^2}{(Z_k^T \nabla q_k(A_k s_k^n))^T \bar{B}_k Z_k^T \nabla q_k(A_k s_k^n)} & \text{if } \dfrac{\|Z_k^T \nabla q_k(A_k s_k^n)\|^3}{(Z_k^T \nabla q_k(A_k s_k^n))^T \bar{B}_k Z_k^T \nabla q_k(A_k s_k^n)} \leq \Delta_k \\ & \text{and } (Z_k^T \nabla q_k(A_k s_k^n))^T \bar{B}_k Z_k^T \nabla q_k(A_k s_k^n) > 0, \\ \dfrac{\Delta_k}{\|Z_k^T \nabla q_k(A_k s_k^n)\|} & \text{otherwise,} \end{cases} \tag{3.11}$$

such that $\bar{B}_k = Z_k^T B_k Z_k$.

A generalized dogleg algorithm can be used to compute the two components of the trial step. This algorithm produces the double fraction of the Cauchy decrease on the tangential and the normal predicted decrease. A convergence theory of the proposed algorithm is based on the fraction of cauchy decrease condition. For more details see [20] and [21].

Once s_k is computed, we set $x_{k+1} = x_k + A_k s_k$. To ensure that $x_{k+1} \in int\boldsymbol{F}$, we need to compute the damping parameter ψ_k at every iteration k. The damping parameter ψ_k is computed as follows:

$$\psi = \min\{\min_i\{c_k^{(i)}, \omega_k^{(i)}\}, 1\}, \tag{3.12}$$

where

$$c_k^{(i)} = \begin{cases} \dfrac{\alpha^{(i)} - x_k^{(i)}}{A_k^{(i)} s_k^{(i)}}, & \text{if } \alpha^{(i)} > -\infty \text{ and } A_k^{(i)} s_k^{(i)} < 0 \\ 1, & \text{otherwise,} \end{cases}$$

and

$$\omega_k^{(i)} = \begin{cases} \dfrac{\beta^{(i)} - x_k^{(i)}}{A_k^{(i)} s_k^{(i)}}, & \text{if } \beta^{(i)} < \infty \text{ and } A_k^{(i)} s_k^{(i)} > 0 \\ 1, & \text{otherwise.} \end{cases}$$

Another damping parameter σ_k in the step may be needed to satisfy $x_{k+1} \in int\boldsymbol{F}$, where σ_k is defined as follows. If $(x_k + \psi_k A_k s_k) \in int\boldsymbol{F}$, we set $\sigma_k = 1$. Otherwise, we set $x_{k+1} = x_k + \sigma_k \psi_k A_k s_k$, such that $\sigma_k \in [1 - \theta\|A_k s_k\|, 1]$ and $\theta > 0$ is a pre-specified fixed constant. It is easy to see that $1 - \sigma_k = O(\|A_k s_k\|)$.

$\boxed{3.2}$ Accepting the step and Updating δ_k

Once the scaled step $\sigma_k \psi_k A_k s_k$, 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 following augmented Lagrangian function

$$\Phi(x, \mu; \rho; r) = f(x) + \mu^T h(x) + \frac{\rho}{2}\|V(x)g(x)\|^2 + r\|h(x)\|^2, \tag{3.13}$$

as a merit function, where r is the penalty parameter.

To test the scaled step, we need to estimate the Lagrange multiplier μ_{k+1}. To estimate the Lagrange multiplier μ_{k+1} we use the following scheme

$$\text{minimize} \ \ \|\nabla f_{k+1} + \nabla h_{k+1}\mu + \rho_k \nabla g_{k+1} V_{k+1} g_{k+1}\|^2. \tag{3.14}$$

Let μ_{k+1} be an estimate of the Lagrange multiplier vector. We test whether the point (x_{k+1}, μ_{k+1}) will be taken as a next iterate.

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

$$Ared_k = \Phi(x_k, \mu_k, ; \rho_k; r_k) - \Phi(x_k + \tilde{\psi}_k A_k s_k, \mu_{k+1}; \rho_k; r_k),$$

where $\tilde{\psi}_k = \sigma_k \psi_k$. The actual reduction $Ared_k$ can be written as,

$$\begin{aligned} Ared_k \ \ &= \ \ \ell(x_k, \mu_k) - \ell(x_{k+1}, \mu_k) - \Delta\mu_k^T h_{k+1} \\ &\quad + \frac{\rho_k}{2}[g_k^T V_k g_k - g_{k+1}^T V_{k+1} g_{k+1}] + r_k[\|h_k\|^2 - \|h_{k+1}\|^2], \end{aligned} \tag{3.15}$$

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

The predicted reduction in the merit function is defined to be

$$
\begin{aligned}
Pred_k &= -(A_k\nabla_x\ell(x_k,\mu_k))^T\tilde{\psi}_k s_k - \frac{1}{2}\tilde{\psi}_k^2 s_k^T G_k s_k - \Delta\mu_k^T(h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k)\\
&\quad +\frac{\rho_k}{2}[\|V_k g_k\|^2 - \|V_k(g_k+(A_k\nabla g_k)^T\tilde{\psi}_k s_k)\|^2]\\
&\quad +r_k[\|h_k\|^2 - \|h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2]. \qquad (3.16)
\end{aligned}
$$

The predicted reduction can be written as

$$
\begin{aligned}
Pred_k &= q_k(0) - q_k(A_k\tilde{\psi}_k s_k) - \Delta\mu_k^T(h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k)\\
&\quad +r_k[\|h_k\|^2 - \|h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2], \qquad (3.17)
\end{aligned}
$$

where

$$
\begin{aligned}
q_k(A_k\tilde{\psi}_k s_k) &= \ell(x_k,\mu_k) + (A_k\nabla_x\ell(x_k,\mu_k))^T\tilde{\psi}_k s_k\\
&\quad + \frac{1}{2}\tilde{\psi}_k^2 s_k^T G_k s_k + \frac{\rho_k}{2}\|V_k(g_k+(A_k\nabla g_k)^T\tilde{\psi}_k s_k)\|^2. \qquad (3.18)
\end{aligned}
$$

After computing the scaled step and updating the Lagrange multiplier, the penalty parameter is updated to ensure that $Pred_k \geq 0$. To update r_k, we use a scheme prossed [7]. We tentatively set

$$r_{k+1} = \max(r_k, \rho_k^2), \qquad (3.19)$$

and if $Pred_k < \frac{r_k}{2}[\|h_k\|^2 - \|h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2]$, then we set

$$r_k = \frac{2[q_k(A_k\tilde{\psi}_k s_k) - q_k(0) + \Delta\mu_k^T(h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k)]}{\|h_k\|^2 - \|h_k+(A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2} + b_0, \qquad (3.20)$$

where $b_0 > 0$ is a small fixed constant. This scheme is described in Step 7 of Algorithm 3.1 below.

The scaled step is tested by comparing $Pred_k$ against $Ared_k$ to know whether it is accepted. Our way of testing the scaled step and updating the trust-region radius are presented in Step 8 of Algorithm 3.1 below.

To update ρ_k, we use a scheme prossed in [25]. In this scheme, if

$$\frac{1}{2}Tpred_k(\tilde{\psi}_k\bar{s}_k^t) \geq \|A_k\nabla g_k V_k g_k\|\min\{\|A_k\nabla g_k V_k g_k\|, \Delta_k\}, \qquad (3.21)$$

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

Finally, the algorithm is terminated when either $\|Z_k^T A_k\nabla_x\ell(x_k,\mu_k)\| + \|A_k\nabla g_k V_k g_k\| + \|h_k\| \leq \varepsilon_1$ or $\|s_k\| \leq \varepsilon_2$, for some $\varepsilon_1, \varepsilon_2 > 0$.

3.3 The master algorithm

A formal description of our an active-set interior-point trust-region algorithm is presented in the following algorithm.

Algorithm 3.1. (An active-set interior-point trust-region algorithm)

Step 0. Given $x_0 \in int\boldsymbol{F}$. Compute V_0, A_0, η_0, and μ_0. Set $\rho_0 = 1$, $r_0 = 1$, and $b_0 = 0.1$. Choose $\theta > 0$, $\varepsilon_1 > 0$, $\varepsilon_2 > 0$. Choose δ_{min}, δ_{max}, and δ_0 such that $\delta_{min} \leq \delta_0 \leq \delta_{max}$. Choose α_1, α_2, τ_1, and τ_2 such that $0 < \alpha_1 < 1 < \alpha_2$, and $0 < \tau_1 < \tau_2 < 1$. Set $k = 0$.

Step 1. If $\|Z_k^T A_k \nabla_x \ell(x_k, \mu_k)\| + \|A_k \nabla g_k V_k g_k\| + \|h_k\| \leq \varepsilon_1$, then stop the algorithm.

Step 2.

a) Compute the normal component s_k^n by solving subproblem (3.2).

b) Compute $\bar{s}_k^t$ by solving subproblem (3.7).

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

Step 3. If $\|s_k\| \leq \varepsilon_2$, then stop the algorithm.

Step 4.

a) Compute the damping parameter ψ_k using (3.12).

b) Set $x_{k+1} = x_k + A_k \psi_k s_k$.

c) If $x_{k+1} \in int\boldsymbol{F}$, then go to Step 5.

Else, set $x_{k+1} = x_k + \sigma_k \psi_k A_k s_k$, where $\sigma_k \in [1 - \theta\|A_k s_k\|, 1]$.
End if.

Step 5. Compute V_{k+1} given by (2.1).

Step 6. Compute μ_{k+1} by solving

$$\text{minimize } \|\nabla f_{k+1} + \nabla h_{k+1}\mu + \rho_k \nabla g_{k+1} V_{k+1} g_{k+1}\|^2.$$

Step 7.

a) Set $r_{k+1} = \max(r_k, \rho_k^2)$.

b) If $Pred_k \leq \frac{r_k}{2}[\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k\|^2]$, then set

$$r_k = \frac{2[q_k(A_k \tilde{\psi}_k s_k) - q_k(0) + \Delta\mu_k^T(h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k)]}{\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k\|^2} + b_0.$$

End if.

Step 8. If $Ared_k < \tau_1 Pred_k$. Set $\delta_k = \alpha_1\|s_k\|$ and go to Step 2.
Else, if $\tau_1 Pred_k \leq Ared_k < \tau_2 Pred_k$, then accept the step and set $\delta_{k+1} = \max(\delta_k, \delta_{min})$.
Else, accept the step and set $\delta_{k+1} = \min\{\delta_{max}, \max\{\delta_{min}, \alpha_2\delta_k\}\}$.
End if.

Step 9.

a) Set $\rho_{k+1} = \rho_k$.

b) If $\frac{1}{2} Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \leq \|A_k \nabla g_k V_k g_k\| \min\{\|A_k \nabla g_k V_k g_k\|, \Delta_k\}$, then set $\rho_{k+1} = 2\rho_k$.

End if.

Step 10. Compute A_{k+1} given by (2.13) and η_{k+1} given by (2.16).

Step 11. Set $k = k + 1$ and go to Step 1.

$\boxed{4}$ Global Convergence Analysis

Let $\{(x_k, \mu_k)\}$ be the sequence of points generated by Algorithm 3.1 and let Ω be a convex subset of $\Re^n$ that contains all iterates $x_k \in int(\boldsymbol{F})$ and $x_k + \tilde{\psi}_k A_k s_k \in int(\boldsymbol{F})$, for all trial steps s_k. On the set Ω, we state the following problem assumptions under which our global convergence theory is proved.

Problem Assumptions:

PA_1. The functions f, h, and g are twice continuously differentiable for all $x \in \Omega$.

PA_2. The matrix $A(x)\nabla h(x)$ has full column rank.

PA_3. All of $f(x)$, $\nabla f(x)$, $\nabla^2 f(x)$, $h(x)$, $\nabla h(x)$, $\nabla^2 h_i(x)$ for $i = 1, 2, \ldots, p$, $g(x)$, $\nabla g(x)$, $\nabla^2 g_i(x)$ for $i = 1, 2, \ldots, m$, and $(\nabla h(x)^T \nabla h(x))^{-1}$ are uniformly bounded in Ω.

PA_4. The sequence $\{\mu_k\}$ is bounded.

PA_5. If an approximation to the Hessian of the Lagrangian is used, then the sequence of approximated Hessian matrices $\{H_k\}$ is bounded.

In the above problem assumptions, even though we assume that $A\nabla h(x)$ has full column rank for all $x \in \Omega$, we do not require $A\nabla g(x)$ has full column rank for all $x \in \Omega$. So, we may have other kinds of stationary points. They are presented in the following two definitions.

Definition 4.1. (Fritz John Point) A point $x_* \in \Omega$ is called a Fritz John point if there exist γ_*, μ_*, and ν_*, not all zeros, such that

$$\begin{aligned} \gamma_* A(x_*)\nabla f(x_*) + A(x_*)\nabla h(x_*)\mu_* + A(x_*)\nabla g(x_*)\nu_* &= 0, && (4.1)\\ h(x_*) &= 0, && (4.2)\\ V_* g(x_*) &= 0, && (4.3)\\ (\nu_*)_i g_i(x_*) &= 0, \quad i = 1, 2, \ldots, m, && (4.4)\\ \gamma_*,\ (\nu_*)_i &\geq 0, \quad i = 1, 2, \ldots, m. && (4.5) \end{aligned}$$

Equations (4.1)-(4.1) are called a Fritz John conditions. More details see [15].

If $\gamma_* \neq 0$, then the point $(x_*, 1, \frac{\mu_*}{\gamma_*}, \frac{\nu_*}{\gamma_*})$ is called a KKT point and the Fritz John conditions are called the KKT conditions.

Definition 4.2. (Infeasible Fritz John Point) A point $x_* \in \Omega$ is called an infeasible Fritz John point if there exist γ_*, μ_*, and ν_* such that

$$\begin{aligned} \gamma_* A(x_*)\nabla f(x_*) + A(x_*)\nabla h(x_*)\mu_* + A(x_*)\nabla g(x_*)\nu_* &= 0, && (4.6)\\ h(x_*) &= 0, && (4.7)\\ A(x_*)\nabla g(x_*) V_* g(x_*) &= 0 \quad \text{but } \|V_* g(x_*)\| > 0, && (4.8)\\ (\nu_*)_i g_i(x_*) &\geq 0, \quad i = 1, 2, \ldots, m, && (4.9)\\ \gamma_*,\ (\nu_*)_i &\geq 0, \quad i = 1, 2, \ldots, m. && (4.10) \end{aligned}$$

Equations (4.6)-(4.10) are called an infeasible Fritz John conditions.

If $\gamma_* \neq 0$, then the point $(x_*, 1, \frac{\mu_*}{\gamma_*}, \frac{\nu_*}{\gamma_*})$ is called an infeasible KKT point and the infeasible Fritz John conditions are called the infeasible KKT conditions.

Lemma 4.3. *Assume PA_1-PA_5. A subsequence $\{x_{k_i}\}$ 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) $\lim_{k_i\to\infty} \|V_{k_i} g(x_{k_i})\| > 0$.

3) $\lim_{k_i\to\infty} \left\{min_{s\in\Re^{n-p}} \|V_{k_i}(g_{k_i} + (A_k\nabla g_{k_i})^T Z_{k_i}\tilde{\psi}_{k_i}\bar{s}^t)\|^2\right\} = \lim_{k_i\to\infty} \|V_{k_i} g_{k_i}\|^2$.

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. The minimizer $\hat{s}_k$ of $minimize_{\bar{s}^t} \|V_k(g_k + (A_k\nabla g_k)^T Z_k\tilde{\psi}_k\bar{s}^t)\|^2$ satisfies

$$Z_k^T A_k \nabla g_k V_k g_k \tilde{\psi}_k + Z_k^T A_k \nabla g_k V_k \nabla g_k^T A_k Z_k \tilde{\psi}_k^2 \hat{s}_k = 0. \tag{4.11}$$

From Condition 3, we have

$$\lim_{k\to\infty} \{2\tilde{\psi}_k \hat{s}_k{}^T Z_k^T A_k \nabla g_k V_k g_k + \tilde{\psi}_k^2 \hat{s}_k{}^T Z_k^T A_k \nabla g_k V_k \nabla g_k^T A_k Z_k \hat{s}_k\} = 0. \tag{4.12}$$

We will consider two cases:
1) If $\lim_{k\to\infty} \hat{s}_k = 0$, then from (4.11) we have $\lim_{k\to\infty} \tilde{\psi}_k Z_k^T A_k \nabla g_k V_k g_k = 0$.
2) If $\lim_{k\to\infty} \hat{s}_k \neq 0$, multiply Equation (4.11) from the left by $2\hat{s}_k^T$ and subtract from the limit (4.12), we have $\lim_{k\to\infty} \|V_k(A_k\nabla g_k)^T Z_k\tilde{\psi}_k\hat{s}_k\|^2 = 0$. But this implies $\lim_{k\to\infty} \tilde{\psi}_k Z_k^T A_k \nabla g_k V_k g_k = 0$. Hence, in either case, we have

$$\lim_{k\to\infty} Z_k^T A_k \nabla g_k V_k g_k = 0. \tag{4.13}$$

Take $(\nu_k)_i = (V_k g_k)_i$, $i = 1, \ldots, m$. Since $\lim_{k\to\infty} \|V_k g_k\| > 0$, then $\lim_{k\to\infty} (\nu_k)_i \geq 0$, for $i = 1, \ldots, m$ and $\lim_{k\to\infty} (\nu_k)_i > 0$, for some i. Therefore $\lim_{k\to\infty} Z_k^T A_k \nabla g_k \nu_k = 0$. But this implies the existence of a sequence $\{\mu_k\}$ such that $\lim_{k\to\infty} \{A_k \nabla h_k \mu_k + A_k \nabla g_k \nu_k\} = 0$. Thus the infeasible Fritz John conditions are hold in the limit with $\gamma_* = 0$. □

From the following lemma, we can easily see that, for any subsequence of the iteration sequence that asymptotically satisfies the Fritz John conditions, the corresponding subsequence of smallest singular values of $\{Z_k^T A_k \nabla g_k V_k\}$ is not bounded away from zero. This means that asymptotically the gradient of the active constraints are linearly dependent.

Lemma 4.4. *Assume PA_1-PA_5. A subsequence $\{k_i\}$ of the iteration sequence asymptotically satisfies Fritz John conditions if it satisfies:*

1) $\lim_{k_i\to\infty} h(x_{k_i}) = 0$.

2) *For all k_i, $\|V_{k_i} g_{k_i}\| > 0$ and $\lim_{k_i\to\infty} V_{k_i} g_{k_i} = 0$.*

3) $\lim_{k_i\to\infty} \left\{\min_{s\in\Re^{n-p}} \frac{\|V_{k_i}(g_{k_i} + (A_k\nabla g_{k_i})^T Z_{k_i}\tilde{\psi}_{k_i}\bar{s}^t)\|^2}{\|V_{k_i} g_{k_i}\|^2}\right\} = 1$.

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. The limit in Condition 3 is equivalent to

$$\lim_{k\to\infty} \left\{\min_{\hat{d}\in\Re^{n-p}} \left\{\|U_k + V_k(A_k\nabla g_k)^T Z_k\tilde{\psi}_k\hat{d}\|^2\right\}\right\} = 1, \tag{4.14}$$

where U_k is a unit vector in the direction of $V_k g_k$, $\hat{d} = \frac{\bar{s}^t}{\|V_k g_k\|}$. Consider the problem

$$\min_{\hat{d}\in\Re^{n-p}} \left\{\|U_k + V_k(A_k\nabla g_k)^T Z_k\tilde{\psi}_k\hat{d}\|^2\right\}. \tag{4.15}$$

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

$$Z_k^T(A_k\nabla g_k)V_k(A_k\nabla g_k)^T Z_k\bar{d}_k\tilde{\psi}_k^2 + Z_k^T(A_k\nabla g_k)V_kU_k\tilde{\psi}_k = 0. \tag{4.16}$$

We consider two cases:

i) If $\lim_{k\to\infty} Z_k\bar{d}_k = 0$ in the above equation, then we have, $\lim_{k\to\infty} Z_k^T(A_k\nabla g_k)V_kU_k\tilde{\psi}_k = 0$.

ii)If $\lim_{k\to\infty} Z_k\bar{d}_k \neq 0$, then from (4.14) and the fact that $\bar{d}_k$ is a solution to the minimization Problem (4.15), we have

$$\lim_{k\to\infty}\{\bar{d}_k{}^T Z_k^T(A_k\nabla g_k)V_k(A_k\nabla g_k)^T Z_k\bar{d}_k\tilde{\psi}_k^2 + 2U_k^T V_k(A_k\nabla g_k)^T Z_k\bar{d}_k\tilde{\psi}_k\} = 0.$$

Multiplying (4.16) from the left by $2\bar{d}_k{}^T$ and subtract it from the above limit, we have

$$\lim_{k\to\infty} \bar{d}_k Z_k^T(A_k\nabla g_k)V_k(A_k\nabla g_k)^T Z_k\bar{d}_k\tilde{\psi}_k^2 = 0.$$

This implies $\lim_{k\to\infty}\left\{Z_kA_k\nabla g_kV_kU_k\tilde{\psi}_k\right\} = 0$. Hence in both cases, we have

$$\lim_{k\to\infty}\{Z_kA_k\nabla g_kV_kU_k\} = 0, \tag{4.17}$$

where $\lim_{k\to\infty}\tilde{\psi}_k = 1$. The rest of the proof follows using arguments similar to those in the above lemma. □

In the following subsection, we present some important lemmas needed in the proof of our main global convergence results.

$\boxed{4.1}$ Important Lemmas

We present some important lemmas needed in the subsequent proofs.

Lemma 4.5. *Let PA_1-PA_3 hold, then at any iteration k*

$$\|s_k^n\| \leq K_1\|h_k\|, \tag{4.18}$$

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

Proof. See Lemma (7.1) of [4]. □

Lemma 4.6. *Assume PA_1 and PA_3. Then $V(x)g(x)$ is Lipschitz continuous in Ω.*

Proof. See Lemma (4.1) of [5]. □

From the above lemma, we conclude that $g(x)^TV(x)g(x)$ is differentiable and $\nabla g(x)V(x)g(x)$ 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 4.7. *Assume PA_1-PA_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*

$$\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T s_k^n\|^2 \geq K_2\|h_k\|\min\{\delta_k, \|h_k\|\}. \tag{4.19}$$

Proof. See Lemma (4.6) of [9].

From the above lemma and the fact that

$$\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k^n\|^2 \geq \tilde{\psi}_k [\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T s_k^n\|^2],$$

then we have

$$\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k^n\|^2 \geq K_2 \tilde{\psi}_k \|h_k\| \min\{\delta_k, \|h_k\|\}. \tag{4.20}$$

From the way of updating the penalty parameter r_k given in Step 7 of Algorithm 3.1 and Inequality (4.20), we have, for all k,

$$Pred_k \geq \frac{r_k}{2} K_2 \tilde{\psi}_k \|h_k\| \min\{\delta_k, \|h_k\|\}. \tag{4.21}$$

□

The following lemma gives a lower bound to the tangential predicted decrease which is obtained by s_k^t.

Lemma 4.8. *Assume PA_1-PA_5. The tangential predicted decrease obtained by s_k^t satisfies*

$$Tpred_k(\bar{s}_k^t) \geq \frac{1}{2} K_3 \|Z_k^T \nabla q_k(A_k s_k^n)\| \min\Big\{\Delta_k, \frac{\|Z_k^T \nabla q_k(A_k s_k^n)\|}{\|\bar{B}_k\|}\Big\}, \tag{4.22}$$

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

Proof. See Lemma (4.7) of [9].

From (3.8), we have

$$Tpred_k(\tilde{\psi}_k \bar{s}_k^t) = q_k(A_k \tilde{\psi}_k s_k^n) - q_k(A_k \tilde{\psi}_k (s_k^n + Z_k \bar{s}_k^t)). \tag{4.23}$$

From (4.23) and the fact that

$$q_k(A_k \tilde{\psi}_k s_k^n) - q_k(A_k \tilde{\psi}_k (s_k^n + Z_k \bar{s}_k^t)) \geq \tilde{\psi}_k [q_k(A_k s_k^n) - q_k(A_k (s_k^n + Z_k \bar{s}_k^t))],$$

then we have

$$Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \geq \tilde{\psi}_k Tpred_k(\bar{s}_k^t). \tag{4.24}$$

From inequalities (4.22) and (4.24), we have

$$Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \geq \frac{1}{2} K_3 \tilde{\psi}_k \|Z_k^T \nabla q_k(A_k s_k^n)\| \min\Big\{\Delta_k, \frac{\|Z_k^T \nabla q_k(A_k s_k^n)\|}{\|\bar{B}_k\|}\Big\}. \tag{4.25}$$

□

Lemma 4.9. *At any iteration k, let $D(x_k) \in \Re^{m \times m}$ be a diagonal matrix whose diagonal entries are*

$$(d_k)_i = \begin{cases} 1 & \text{if } (g_k)_i < 0 \text{ and } (g_{k+1})_i \geq 0, \\ -1 & \text{if } (g_k)_i \geq 0 \text{ and } (g_{k+1})_i < 0, \\ 0 & \text{otherwise}, \end{cases} \tag{4.26}$$

where $i = 1, 2, \ldots, m$. Then

$$V_{k+1} = V_k + D_k. \tag{4.27}$$

Proof. See Lemma (6.2) of [10]. □

Lemma 4.10. *Assume PA_1 and PA_3. At any iteration k, there exists a positive constant K_4 independent of k, such that*

$$\|D_k g_k\| \leq K_4 \|s_k\|, \tag{4.28}$$

where $D_k \in \Re^{m\times m}$ is the diagonal matrix whose diagonal entries are defined in (4.26).

Proof. The proof is similar to the proof of lemma (6.3) of [10]. □

The following lemma shows how accurate our definition of $Pred_k$ is as an approximation to $Ared_k$.

Lemma 4.11. *Assume PA_1-PA_5, then there exists a constant $K_5 > 0$ that does not depend on k, such that*

$$| \; Ared_k - Pred_k \; | \leq K_5 r_k \tilde{\psi}_k \|s_k\|^2. \tag{4.29}$$

Proof. From (3.15) and using (4.27), we have

$$\begin{aligned} Ared_k \;\; &= \;\; \ell(x_k,\mu_k) - \ell(x_{k+1},\mu_k) - \Delta\mu_k^T h_{k+1} \\ &\quad + \frac{\rho_k}{2}[g_k^T V_k g_k - g_{k+1}^T (V_k + D_k) g_{k+1}] + r_k[\|h_k\|^2 - \|h_{k+1}\|^2]. \end{aligned}$$

From the above equation and (3.16) and using Cauchy-Schwarz inequality, we have

$$\begin{aligned} | \; Ared_k - Pred_k \; | \;\; \leq \;\; & | \; \ell(x_k,\mu_k) + (A_k \nabla \ell(x_k,\mu_k))^T \tilde{\psi}_k s_k - \ell(x_{k+1},\mu_k) \; | \\ & + | \; \Delta\mu_k^T [\; h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k - h_{k+1} \;] \; | \\ & + \frac{\rho_k}{2} \; | \; \|V_k (g_k + (A_k \nabla g_k)^T \tilde{\psi}_k s_k)\|^2 - g_{k+1}^T (V_k + D_k) g_{k+1} \; | \\ & + r_k \; | \; \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k\|^2 - \|h_{k+1}\|^2 \; | \; . \end{aligned}$$

Hence,

$$\begin{aligned} & |Ared_k - Pred_k| \\ \leq \;\; & \frac{\tilde{\psi}_k^2}{2} \; | \; s_k^T A_k (H_k - \nabla^2 \ell(x_k + \xi_1 A_k \tilde{\psi}_k s_k, \mu_k)) A_k s_k \; | \\ & + \frac{\tilde{\psi}_k^2}{2} \; | \; s_k^T \mathrm{diag}(\nabla_x \ell(x_k,\mu_k)) \mathrm{diag}(\eta(x_k)) s_k \; | \\ & + \frac{\tilde{\psi}_k^2}{2} \; | \; s_k^T A_k [\nabla^2 h(x_k + \xi_2 A_k \tilde{\psi}_k s_k) \Delta\mu_k] A_k s_k \; | \\ & + \frac{\rho_k}{2} \tilde{\psi}_k^2 \; | \; s_k^T A_k [\nabla g_k V_k \nabla g_k^T - \nabla g(x_k + \xi_4 A_k \tilde{\psi}_k s_k) V_k \nabla g(x_k + \xi_4 A_k \tilde{\psi}_k s_k)^T] A_k s_k \; | \\ & + \frac{\rho_k}{2} \tilde{\psi}_k^2 \; | \; s_k^T \mathrm{diag}(\nabla g_k V_k g_k) \; \mathrm{diag}(\eta(x_k)) s_k \; | \\ & + \frac{\rho_k}{2} \tilde{\psi}_k^2 \; | \; s_k^T A_k \nabla^2 g(x_k + \xi_4 A_k \tilde{\psi}_k s_k) V_k g(x_k + \xi_4 A_k \tilde{\psi}_k s_k) A_k s_k \; | \\ & + \frac{\rho_k}{2} \tilde{\psi}_k \|D_k [g_k + \nabla g(x_k + \xi_5 A_k \tilde{\psi}_k s_k)^T A_k s_k]\|^2 \\ & + r_k \tilde{\psi}_k^2 \; | \; s_k^T A_k [\nabla h_k \nabla h_k^T - \nabla h(x_k + \xi_6 A_k \tilde{\psi}_k s_k) \nabla h(x_k + \xi_6 A_k \tilde{\psi}_k s_k)^T] A_k s_k \; | \\ & + r_k \tilde{\psi}_k^2 \; | \; s_k^T A_k \nabla^2 h(x_k + \xi_6 A_k \tilde{\psi}_k s_k) h(x_k + \xi_6 A_k \tilde{\psi}_k s_k) A_k s_k \; | \end{aligned}$$

for some ξ_1, ξ_2, ξ_3, ξ_4, ξ_5, and $\xi_6 \in (0,1)$. Hence, by using the problem assumptions, the fact that $\tilde{\psi}_k^2 \leq \tilde{\psi}_k$, $r_k \geq \rho_k$, $r_k \geq 1$, and Inequality (4.28), we have

$$| \, Ared_k - Pred_k \, | \leq \tilde{\psi}_k [K_6 \|s_k\|^2 + K_7 r_k \|s_k\|^2 \|h_k\| + K_8 r_k \|s_k\|^3], \tag{4.30}$$

where K_6, K_7, and K_8 are positive constants independent of k. From the above inequality, the fact that $r_k \geq 1$, and $\|s_k\|$ and $\|h_k\|$ are uniformly bound. The proof follows.

The j^{th} trial iterate of iteration k is denoted by k^j.

The following lemma 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. □

Lemma 4.12. *Assume PA_1-PA_5. If $\|h_k\| \geq \varepsilon$ where ε is any positive constant, then the condition $\frac{Ared_{kj}}{Pred_{kj}} \geq \tau_1$ will be satisfied for some finite j.*

Proof. Since $\|h_k\| \geq \varepsilon$, then from (4.21) and (4.29), we have

$$\left| \frac{Ared_k}{Pred_k} - 1 \right| = \frac{| \, Ared_k - Pred_k \, |}{Pred_k} \leq \frac{2K_5 \tilde{\psi}_k \delta_k^2}{K_2 \tilde{\psi}_k \varepsilon \min\{\varepsilon, \delta_k\}}.$$

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

$$\left| \frac{Ared_{k^j}}{Pred_{k^j}} - 1 \right| \leq \frac{2K_5 \delta_{k^j}}{K_2 \varepsilon}.$$

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

Lemma 4.13. *Assume PA_1-PA_5. If at a given iteration k, the j^{th} trial step satisfies*

$$\|s_{k^j}\| \leq \min\left\{ \frac{(1-\tau_1)K_2}{4K_5}, 1 \right\} \|h_k\|, \tag{4.31}$$

then it must be accepted.

Proof. We prove this lemma by contradiction. Assume that the step s_{k^j} is rejected and Inequality (4.31) holds. Then, from (4.21), (4.29), and (4.31), we have

$$(1-\tau_1) < \frac{| \, Ared_{k^j} - Pred_{k^j} \, |}{Pred_{k^j}} < \frac{2K_5 \tilde{\psi}_{k^j} \|s_{k^j}\|^2}{K_2 \tilde{\psi}_{k^j} \|h_k\| \|s_{k^j}\|} \leq \frac{(1-\tau_1)}{2}.$$

This gives a contradiction and proves the lemma.

Lemma 4.14. *Assume PA_1-PA_5. For all trial steps j of any iteration k generated by the algorithm, δ_{k^j} satisfies*

$$\delta_{k^j} \geq \min\left\{ \frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\tau_1)K_2}{4K_5}, \alpha_1 \right\} \|h_k\|, \tag{4.32}$$

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

$$\delta_k \geq \delta_{min} \geq \frac{\delta_{min}}{b_1} \|h_k\|. \tag{4.33}$$

Therefore, (4.32) 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 (4.13)

$$\|s_{k^i}\| > \min\Big\{\frac{(1-\tau_1)K_2}{4K_5}, 1\Big\}\|h_k\|,$$

for all $i = 1, 2, \dots, j-1$. Since s_{k^i} is a rejected trial step, then from the way of updating the radius of trust region (see Algorithm 3.1) and using the above inequality, we have

$$\delta_{k^j} = \alpha_1\|s_{k^{j-1}}\| > \alpha_1 \min\Big\{\frac{(1-\tau_1)K_2}{4K_5}, 1\Big\}\|h_k\|.$$

From Inequality (4.33) and the above inequality, Inequality (4.32) holds. □

The following lemma says that as long as $\|h_k\|$ is bound away from zero, the trust-region radius is bound away from zero.

Lemma 4.15. *Assume PA_1-PA_5. If $\|h_k\| \geq \varepsilon$, where $\varepsilon > 0$, then there exists a positive constant K_9 that depends on ε but does not depend on k such that*

$$\delta_{k^j} \geq K_9.$$

Proof. Using (4.32) and taking

$$K_9 = \varepsilon \min\Big\{\frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\tau_1)K_2}{4K_5}, \alpha_1\Big\}, \tag{4.34}$$

the proof follows directly. □

$\boxed{4.2}$ Convergence When ρ_k is Unbound

In this section, we study the convergence of the iteration sequence when the parameter ρ_k goes to infinity. Notice that, from the way of updating the penalty parameter r_k, the penalty parameter r_k is increased at a given iteration k, because of the rule (3.19) or (3.20).

Lemma 4.16. *Assume PA_1-PA_5. Let k be the index of an iteration at which the penalty parameter r_k is increased. Then there exists a positive constant K_{10} that does not depend on k, such that*

$$\rho_k\tilde{\psi}_k\|h_k\|^2 \leq K_{10}. \tag{4.35}$$

Proof. Since r_k is increased using the rule (3.20), then

$$\begin{aligned}\frac{r_k}{2}[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2] &= [q_k(A_k\tilde{\psi}_k s_k) - q_k(0) + \Delta\mu_k^T(h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k)]\\ &\quad + \frac{b_0}{2}[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2].\end{aligned}$$

By using (3.18), (4.20), and (4.33) in the above equality, we have

$$\begin{aligned}\frac{r_k}{2}K_2\tilde{\psi}_k\|h_k\|^2 \min\Big\{\frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\tau_1)K_2}{4K_5}, \alpha_1\Big\} &\leq (A_k\nabla_x\ell(x_k,\mu_k))^T\tilde{\psi}_k s_k + \frac{1}{2}\tilde{\psi}_k^2 s_k^T G_k s_k\\ &\quad + \Delta\mu_k^T(h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k)\\ &\quad + \frac{\rho_k}{2}[\|V_k(g_k + (A_k\nabla g_k)^T\tilde{\psi}_k s_k)\|^2 - \|V_k g_k\|^2]\end{aligned}$$

$$+ \frac{b_0}{2}[\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k\|^2].$$

Since $r_k \geq \rho_k^2$, then we can write the above inequality as follows

$$\begin{aligned}\frac{\rho_k^2}{2} K_2 \tilde{\psi}_k \|h_k\|^2 \min\left\{\frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\tau_1)K_2}{4K_5}, \alpha_1\right\} \leq\ & (A_k \nabla_x \ell(x_k, \mu_k))^T \tilde{\psi}_k s_k + \frac{1}{2}\tilde{\psi}_k^2 s_k^T G_k s_k \\ & + \Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k) \\ & + \frac{\rho_k}{2}[\|V_k(g_k + (A_k \nabla g_k)^T \tilde{\psi}_k s_k)\|^2 \\ & + \frac{b_0}{2}\|h_k\|^2.\end{aligned}$$

Hence,

$$\begin{aligned}\frac{\rho_k}{2} K_2 \tilde{\psi}_k \|h_k\|^2 \min\left\{\frac{\delta_{min}}{b_1}, \frac{\alpha_1(1-\tau_1)K_2}{4K_5}, \alpha_1\right\} \leq\ & \frac{1}{\rho_k}\Big[(A_k \nabla_x \ell(x_k, \mu_k))^T \tilde{\psi}_k s_k + \frac{1}{2}\tilde{\psi}_k^2 s_k^T G_k s_k \\ & + \Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k) + \frac{b_0}{2}\|h_k\|^2\Big] \\ & + \frac{1}{2}\|V_k(g_k + (A_k \nabla g_k)^T \tilde{\psi}_k s_k)\|^2 \\ \leq\ & \frac{1}{\rho_k}\Big[|(A_k \nabla_x \ell(x_k, \mu_k))^T s_k| + \frac{1}{2}|s_k^T G_k s_k| \\ & + |\Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k)| + \frac{b_0}{2}\|h_k\|^2\Big] \\ & + \frac{1}{2}\|V_k(g_k + (A_k \nabla g_k)^T \tilde{\psi}_k s_k)\|^2,\end{aligned}$$

where $\tilde{\psi}_k \leq 1$. Using the Cauchy-Schwarz inequality, problem assumptions $PA_3 - PA_5$, and the fact that $\|s_k\| \leq \delta_{max}$ the proof is completed. □

Lemma 4.17. *Assume PA_1-PA_5. If ρ_k is unbound and if there exists an infinite subsequence $\{k_i\}$ of the iteration sequence at which the penalty parameter r_k is increased, then*

$$\lim_{k_i \to \infty} \|h_{k_i}\| = 0. \tag{4.36}$$

Proof. The proof follows directly from Lemma (4.16),$\lim_{k_i \to \infty} \tilde{\psi}_{k_i} = 1$, and the assumption that ρ_k is unbound. □

Theorem 4.18. *Assume PA_1-PA_5. If $\rho_k \to \infty$ as $k \to \infty$, then the sequence of iterates generated by the algorithm satisfies*

$$\lim_{k \to \infty} \|h_k\| = 0. \tag{4.37}$$

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 $\|h_{k_j}\| \geq \frac{\varepsilon}{2}$. We prove contradiction by showing that ε must be zero.

Using (3.19), then $r_k \to \infty$ such that $\rho_k \to \infty$ as $k \to \infty$. We consider two cases.

i) If there exists an infinite subsequence $\{k_i\}$ of the iteration sequence at which the penalty parameter is increased by the rule (3.20), then from Lemma (4.17), we have

$$\lim_{k_i \to \infty} \|h_{k_i}\| = 0. \tag{4.38}$$

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

$$\frac{Ared_{\hat{k}}}{r_{\hat{k}}} \geq \tau_1 \frac{Pred_{\hat{k}}}{r_{\hat{k}}} \geq \tau_1 \tilde{\psi}_{\hat{k}} \frac{\varepsilon K_2}{4} \min\left\{\frac{\varepsilon}{2}, \delta_{\hat{k}}\right\} \geq \tau_1 \tilde{\psi}_{\hat{k}} \frac{\varepsilon K_2}{4} \min\left\{\frac{\varepsilon}{2}, \hat{K_9}\right\},$$

where $\hat{K_9}$ is as K_9 in (4.34), with ε is replaced by $\frac{\varepsilon}{2}$. Hence for all acceptable steps, we have

$$\frac{\ell_{\hat{k}} - \ell_{\hat{k}+1}}{r_{\hat{k}}} + \frac{\rho_k\{\|V_{\hat{k}} g_{\hat{k}}\|^2 - \|V_{\hat{k}+1} g_{\hat{k}+1}\|^2\}}{r_{\hat{k}}} + \|h_{\hat{k}}\|^2 - \|h_{\hat{k}+1}\|^2 \geq \tau_1 \tilde{\psi}_{\hat{k}} \frac{\varepsilon K_2}{4} \min\left\{\frac{\varepsilon}{2}, \hat{K_9}\right\} > 0.$$

Let $k_{\hat{i}}$ and $k_{\hat{i}+1}$ be two elements of the sequence $\{k_i\}$ such that there exists an iterate indexed $k \in \{k_j\}$ between $k_{\hat{i}}$ and $k_{\hat{i}+1}$. From above inequality we have

$$\sum_{\hat{k}=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\{\ell_{\hat{k}} - \ell_{\hat{k}+1}\}}{r_{\hat{k}}} + \sum_{\hat{k}=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\rho_k\{\|V_{\hat{k}} g_{\hat{k}}\|^2 - \|V_{\hat{k}+1} g_{\hat{k}+1}\|^2\}}{r_{\hat{k}}} + \|h_{\hat{k}}\|^2 - \|h_{\hat{k}+1}\|^2 \geq \tau_1 \tilde{\psi}_{\hat{k}} \frac{\varepsilon K_2}{4} \min\left\{\frac{\varepsilon}{2}, \hat{K_9}\right\} > 0.$$

Assume that there exists k_γ iterations between the iterates indexed $k_{\hat{i}}$ and $k_{\hat{i}+1}$ where the penalty parameter is increased. Because there is no iteration between the iterates indexed $k_{\hat{i}}$ and $k_{\hat{i}+1}$ with a penalty parameter that was increased using the rule (3.20), all the penalty parameters of the k_γ iterations are increased using the rule (3.19). If we notice that for all k, $r_k \geq \rho_k^2$ and if the value of ρ_k is increased, it is increased by two times its previous value, we can write

$$\sum_{\hat{k}=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\{\ell_{\hat{k}} - \ell_{\hat{k}+1}\}}{r_{\hat{k}}} + \sum_{\hat{k}=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\rho_k\{\|V_{\hat{k}} g_{\hat{k}}\|^2 - \|V_{\hat{k}+1} g_{\hat{k}+1}\|^2\}}{r_{\hat{k}}} \leq \frac{2sup_{x\in\Omega}\{|f(x)| + \bar{\mu}\|h(x)\|\}}{\rho_{k_{\hat{i}}}^2} \sum_{i=0}^{k_\gamma-1} \frac{1}{2^{2i}} + \frac{2sup_{x\in\Omega}\|g(x)\|^2}{\rho_{k_{\hat{i}}}} \sum_{i=0}^{k_\gamma-1} \frac{1}{2^i},$$

where $\bar{\mu}$ is the upper bound on the sequence $\{\mu_k\}$. Since the two infinite series $\sum_{i=0}^{\infty} \frac{1}{2^{2i}}$ and $\sum_{i=0}^{\infty} \frac{1}{2^i}$ converge, the two series $\sum_{i=0}^{k_\gamma-1} \frac{1}{2^{2i}}$ and $\sum_{i=0}^{k_\gamma-1} \frac{1}{2^i}$ are bound for any k_γ. Since $\rho_k \to \infty$, then for $k_{\hat{i}}$ sufficiently large, we can write

$$\Big| \sum_{\hat{k}=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\{\ell_{\hat{k}} - \ell_{\hat{k}+1}\}}{r_{\hat{k}}} + \sum_{\hat{k}=k_{\hat{i}}}^{k_{\hat{i}+1}-1} \frac{\rho_k\{\|V_{\hat{k}} g_{\hat{k}}\|^2 - \|V_{\hat{k}+1} g_{\hat{k}+1}\|^2\}}{r_{\hat{k}}} \Big| \leq \tau_1 \tilde{\psi}_{\hat{k}} \frac{\varepsilon K_2}{8} \min\left\{\frac{\varepsilon}{2}, \hat{K_9}\right\}.$$

Therefore,

$$\|h_{\hat{k}}\|^2 - \|h_{\hat{k}+1}\|^2 \geq \tau_1 \tilde{\psi}_{\hat{k}} \frac{\varepsilon K_2}{8} \min\left\{\frac{\varepsilon}{2}, \hat{K_9}\right\}.$$

But this contradicts (4.38). Thus the supposition is not correct and the theorem is proved in this case.

ii) Consider the case where there exist finite number of iterates where the penalty parameter

r_k is increased using the rule (3.20). This implies the existence of an integer $\breve{k}$ such that there exists no iterate indexed k for all $k \geq \breve{k}$ where the penalty parameter is increased using the rule (3.20). Hence, for all $k \geq \breve{k}$, if the penalty parameter is increased, then it will be increased because of the rule (3.19). Without loss of generality, we assume that $\breve{k} = 1$.

From Inequality (4.21), we can write for any acceptable iteration index k,

$$\frac{Ared_k}{r_k} \geq \tau_1 \frac{Pred_k}{r_k} \geq \tau_1 \frac{r_k}{2} K_2 \tilde{\psi}_k \|h_k\| \min\{\delta_k, \|h_k\|\}.$$

Let k_{a_1} and k_{a_2} be two consecutive iterates of $\{k_j\}$. Summing over all acceptable iterates indexed k such that $k_{a_1} \leq k < k_{a_2}$, we have

$$\sum_{k=k_{a_1}}^{k_{a_2}-1} \frac{Ared_k}{r_k} \geq \tau_1 \tilde{\psi}_k \frac{\varepsilon K_2}{4} \min\Big\{\frac{\varepsilon}{2}, \hat{K}_9\Big\}.$$

Hence

$$\sum_{k=k_{a_1}}^{k_{a_2}-1} \frac{\{\ell_k - \ell_{k+1}\}}{r_k} + \sum_{k=k_{a_1}}^{k_{a_2}-1} \frac{\rho_k\{\|V_k g_k\|^2 - \|V_{k+1} g_{k+1}\|^2\}}{r_k} + \|h_{k_{a_1}}\|^2 - \|h_{k_{a_2}}\|^2 \geq \tau_1 \tilde{\psi}_k \frac{\varepsilon K_2}{4} \min\Big\{\frac{\varepsilon}{2}, \hat{K}_9\Big\}.$$

Using an argument similar to the first case, we can conclude that, for k_{a_1} sufficiently large

$$\Big|\sum_{k=k_{a_1}}^{k_{a_2}-1} \frac{\{\ell_k - \ell_{k+1}\}}{r_k} + \sum_{k=k_{a_1}}^{k_{a_2}-1} \frac{\rho_k\{\|V_k g_k\|^2 - \|V_{k+1} g_{k+1}\|^2\}}{r_k}\Big| \leq \tau_1 \tilde{\psi}_k \frac{\varepsilon K_2}{8} \min\Big\{\frac{\varepsilon}{2}, \hat{K}_9\Big\}.$$

Therefore,

$$\|h_{a_1}\|^2 - \|h_{a_2}\|^2 \geq \tau_1 \tilde{\psi}_k \frac{\varepsilon K_2}{8} \min\Big\{\frac{\varepsilon}{2}, \hat{K}_9\Big\}.$$

But this implies that the sequence $\{\|h_{k_j}\|\}$ is unbound. This contradicts assumption PA_3. Thus the supposition is not correct and the theorem is proved. □

From the way of updating the parameter ρ, we notice that the sequence $\{\rho_k\}$ is unbounded only when there exist an infinite subsequence of indices $\{k_i\}$, at which

$$\frac{1}{2} Tpred_k(\tilde{\psi}_k \bar{s}_k^t) < \|A_k \nabla g_k V_k g_k\| \min\Big\{\|A_k \nabla g_k V_k g_k\|, \Delta_k\Big\}. \tag{4.39}$$

The following lemma shows that if $\rho_k \to \infty$ and $\limsup_{k\to\infty} \|V_k g_k\| > 0$ as $k \to \infty$, then the iteration sequence generated by the algorithm has a subsequence that satisfies the infeasible Fritz John conditions in the limit.

Lemma 4.19. *Assume PA_1-PA_5. If $\rho_k \to \infty$, as $k \to \infty$ and there exists a subsequence $\{k_j\}$ of indices indexing iterates that satisfy $\|V_k g_k\| \geq \varepsilon > 0$ for all $k \in \{k_j\}$, then a subsequence of the iteration sequence indexed $\{k_j\}$ satisfies the infeasible Fritz John conditions in the limit.*

Proof. Let the subsequence $\{k_i\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. 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 (4.3), we have for all k, $|\|V_k g_k\|^2 - \|V_k(g_k + (A_k\nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2| \geq \varepsilon_1$ for some $\varepsilon_1 > 0$. From (4.13), we have $\|Z_k A_k \nabla g_k V_k g_k\| \geq \varepsilon_2$, for some $\varepsilon_2 > 0$, hence

$$\begin{aligned}
&\|Z_k^T A_k \nabla g_k V_k g_k + Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)] s_k^n\| \\
&\geq \|Z_k^T A_k \nabla g_k V_k g_k\| - \|Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)]\|\|s_k^n\| \\
&\geq \varepsilon_2 - K_1 \|Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)]\|\|h_k\|.
\end{aligned}$$

Since $\{\|h_k\|\}$ convergence to zero and $\|Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)]\|$ is bounded, then we can write $\|Z_k^T A_k \nabla g_k V_k g_k + [Z_k^T A_k \nabla g_k V_k A_k \nabla g_k^T + Z_k^T \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)] s_k^n\| \geq \frac{\varepsilon_2}{2}$. Therefore

$$\begin{aligned}
\|Z_k^T \nabla q_k(A_k s_k^n)\| &\geq \rho_k \|Z_k^T A_k \nabla g_k V_k g_k + Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)] s_k^n\| \\
&\quad - \|Z_k^T A_k \nabla_x \ell_k + Z_k^T [A_k H_k A_k + \operatorname{diag}(\nabla_x \ell_k)\operatorname{diag}(\eta_k)] s_k^n\| \\
&\geq \rho_k \frac{\varepsilon_2}{2} - \|Z_k^T A_k \nabla_x \ell_k + Z_k^T [A_k H_k A_k + \operatorname{diag}(\nabla_x \ell_k)\operatorname{diag}(\eta_k)] s_k^n\| \\
&\geq \rho_k \Big[\frac{\varepsilon_2}{2} - \frac{1}{\rho_k} \|Z_k^T A_k \nabla_x \ell_k + Z_k^T [A_k H_k A_k + \operatorname{diag}(\nabla_x \ell_k)\operatorname{diag}(\eta_k)] s_k^n\|\Big].
\end{aligned}$$

From (4.25), we have

$$\begin{aligned}
&Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \geq \frac{1}{2} K_3 \tilde{\psi}_k \rho_k \Big[\frac{\varepsilon_2}{2} - \frac{1}{\rho_k} \|Z_k^T [A_k \nabla_x \ell_k + \tilde{H}_k s_k^n]\| \\
&\min\Big\{\Delta_k, \frac{\frac{\varepsilon_2}{2} - \frac{1}{\rho_k}\|Z_k^T [A_k \nabla_x \ell_k + \tilde{H}_k s_k^n]\|}{\|Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \ \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)] Z_k\| + \frac{1}{\rho_k}\|Z_k^T \tilde{H}_k Z_k\|}\Big\}\Big],
\end{aligned}$$

where $\tilde{H}_k = A_k H_k A_k + \operatorname{diag}(\nabla_x \ell_k)\operatorname{diag}(\eta_k)$. For k sufficiently large we have

$$\begin{aligned}
&Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \\
&\quad \geq \frac{\varepsilon_2}{4} K_3 \tilde{\psi}_k \rho_k \min\Big\{\Delta_k, \frac{\varepsilon_2}{2\|Z_k^T [A_k \nabla g_k V_k A_k \nabla g_k^T + \operatorname{diag}(\nabla g_k V_k g_k)\operatorname{diag}(\eta_k)] Z_k\|}\Big\}.
\end{aligned}$$

Since $\rho_k \to \infty$ there exists infinite number of acceptable iterates at which (4.39) hold. From the way of updating the parameter ρ_k, as k goes to infinity, $\rho_k \to \infty$. This gives a contradiction unless $\rho_k \Delta_k$ is bounded. Hence $\Delta_k \to 0$ and therefore $\|s_k\| \to 0$. We consider two cases:

i) If $\|V_k g_k\|^2 - \|V_k(g_k + (A_k \nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2 > \varepsilon_1$, we have

$$\rho_k \{\|V_k g_k\|^2 - \|V_k(g_k + (A_k \nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2|\} > \rho_k \varepsilon_1 \to \infty. \tag{4.40}$$

From (4.23), $Tpred_k(\tilde{\psi}_k \bar{s}_k^t)$ can be written as follows

$$\begin{aligned}
Tpred_k(\tilde{\psi}_k \bar{s}_k^t) = -\Big[&Z_k^T(\nabla_x \ell_k + B_k \tilde{\psi}_k s_k^n)]^T \tilde{\psi}_k \bar{s}_k^t - \frac{1}{2} \tilde{\psi}_k^2 \bar{s}_k^{t^T} Z_k^T G_k Z_k \bar{s}_k^t \\
&+ \frac{\rho_k}{2} [\|V_k g_k\|^2 - \|V_k(g_k + (A_k \nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2\Big].
\end{aligned} \tag{4.41}$$

Using (4.41), (4.40), and under problem assumptions $PA_3 - PA_5$, we have $Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \to \infty$. Hence , the left hand side of Inequality (4.39) tends to infinity while the right hand side

goes to zero. This gives a contradiction in this case.
ii) If $\|V_k g_k\|^2 - \|V_k(g_k + (A_k\nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2 < -\varepsilon_1$, then

$$\rho_k\{\|V_k g_k\|^2 - \|V_k(g_k + (A_k\nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2|\} < -\rho_k\varepsilon_1 \to -\infty,$$

where $\rho_k \to \infty$ as $k \to \infty$. Similar to the above case, $Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \to -\infty$. This gives a contradiction with $Tpred_k(\tilde{\psi}_k \bar{s}_k^t) > 0$. This two contradictions prove the lemma. □

The following lemma shows that if $\rho_k \to \infty$ and $\liminf_{k\to\infty} \|V_k g_k\| = 0$ 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 4.20. *Assume PA_1-PA_5. If $\rho_k \to \infty$, as $k \to \infty$, and there exists a subsequence indexed $\{k_j\}$ of iterates that satisfy $\|V_{k_j} g_{k_j}\| \lim_{k\to\infty} \|h_k\| = 0$, and there exists a subsequence $\{k_j\}$ of iterates that satisfies $\|V_k g_k\| > 0$ for all $k \in \{k_j\}$ and $\lim_{k_j\to\infty} \|V_{k_j} g_{k_j}\| = 0$, then a subsequence of the sequence of iterates indexed $\{k_j\}$ satisfies Fritz John conditions in the limit.*

Proof. Let the subsequence $\{k_j\}$ be renamed to $\{k\}$ to simplify the notations avoiding double indices. The proof is by contradiction. Assume there exists no subsequence that satisfies the feasible Fritz John's conditions in the limit. By using Lemma (4.4), there exists a constant ε_3 such that for all k sufficiently large,

$$\frac{|\ \|V_k g_k\|^2 - \|V_k(g_k + (A_k\nabla g_k)^T Z_k \tilde{\psi}_k \bar{s}_k^t)\|^2\ |}{\|V_k g_k\|^2} \geq \varepsilon_3. \tag{4.42}$$

We consider three cases:

i) If $\lim inf_{k\to\infty} \frac{\tilde{\psi}_k \bar{s}_k^t}{\|V_k g_k\|} = 0$, the above inequality gives a contradiction.

ii) If $\lim sup_{k\to\infty} \frac{\tilde{\psi}_k \bar{s}_k^t}{\|V_k g_k\|} = \infty$. From the way of computing the tangential component of the trial step, we have

$$Z_k^T \nabla q_k(A_k s_k^n) = -Z_k^T(B_k + \upsilon_k I) Z_k \bar{s}_k^t,$$

where $\upsilon_k \geq 0$ is the Lagrange multiplier of the trust region constraint. Using the above equation, then Inequality (4.25) can be written in the form

$$\begin{aligned} & Tpred_k(\tilde{\psi}_k \bar{s}_k^t) \\ & \geq \frac{K_3}{2}\tilde{\psi}_k \|Z_k^T \nabla q_k(A_k s_k^n)\| \min\Big\{\Delta_k, \frac{\|Z_k^T[\frac{1}{\rho_k}G_k + (\frac{\upsilon_k}{\rho_k}I + A_k\nabla g_k V_k \nabla g_k^T A_k)]Z_k \bar{s}_k^t\|}{\|Z_k^T(\frac{1}{\rho_k}G_k + A_k\nabla g_k V_k \nabla g_k^T A_k)Z_k\|}\Big\}. \end{aligned} \tag{4.43}$$

Because $\rho_k \to \infty$, as $k \to \infty$, there exists an infinite number of acceptable steps such that Inequality (4.39) holds. But Inequality (4.39) can be written as

$$\frac{1}{2} Tpred_k(\tilde{\psi}_k \bar{s}_k^t) < \|Z_k A_k \nabla g_k\|^2 \|V_k g_k\|^2. \tag{4.44}$$

From Inequalities (4.43) and (4.44), we have

$$\frac{K_3}{2}\tilde{\psi}_k \|Z_k^T \nabla q_k(A_k s_k^n)\| \min\Big\{\Delta_k, \frac{\|Z_k^T[\frac{1}{\rho_k}G_k + (\frac{\upsilon_k}{\rho_k}I + A_k\nabla g_k V_k \nabla g_k^T A_k)]Z_k \bar{s}_k^t\|}{\|Z_k^T(\frac{1}{\rho_k}G_k + A_k\nabla g_k V_k \nabla g_k^T A_k)Z_k\|}\Big\}$$

$$< 2b_2^2\|V_kg_k\|^2,$$

where $b_2 = sup_{x\in\Omega}\|Z_kA_k\nabla g_k\|$. Hence, if we divided the above inequality by $\|V_kg_k\|$, we obtain

$$\frac{K_3}{2}\tilde{\psi}_k\|Z_k^T\nabla q_k(A_ks_k^n)\|\min\left\{\frac{\Delta_k}{\|V_kg_k\|},\frac{\|Z_k^T[\frac{1}{\rho_k}G_k+(\frac{\upsilon_k}{\rho_k}I+A_k\nabla g_kV_k\nabla g_k^TA_k)]Z_k\bar{s}_k^t\|}{\|Z_k^T(\frac{1}{\rho_k}G_k+A_k\nabla g_kV_k\nabla g_k^TA_k)Z_k\|\|V_kg_k\|}\right\}$$
$$< 2b_2^2\|V_kg_k\|. \quad (4.45)$$

The right hand side of the above inequality goes to zero as $k\to\infty$. This implies that along the subsequence $\{k_i\}$ where $\lim_{k_i\to\infty}\frac{\tilde{\psi}_{k_i}\bar{s}_{k_i}^t}{\|V_{k_i}g_{k_i}\|}=\infty$, we have

$$\|Z_{k_i}^T\nabla q_{k_i}(A_{k_i}s_{k_i}^n)\|\frac{\|Z_{k_i}^T[\frac{1}{\rho_{k_i}}G_{k_i}+(\frac{\upsilon_{k_i}}{\rho_{k_i}}I+A_{k_i}\nabla g_{k_i}V_{k_i}\nabla g_{k_i}^TA_{k_i})]Z_{k_i}\tilde{\psi}_{k_i}\bar{s}_{k_i}^t\|}{\|Z_{k_i}^T(\frac{1}{\rho_{k_i}}G_{k_i}+A_{k_i}\nabla g_{k_i}V_{k_i}\nabla g_{k_i}^TA_{k_i})Z_{k_i}\|\|V_{k_i}g_{k_i}\|},$$

is bounded. Therefore, asymptotically, either $\frac{\tilde{\psi}_{k_i}\bar{s}_{k_i}^t}{\|V_{k_i}g_{k_i}\|}$ lies in the null space of $Z_{k_i}^T(\frac{\upsilon_{k_i}}{\rho_{k_i}}I+A_{k_i}\nabla g_{k_i}V_{k_i}\nabla g_{k_i}^TA_{k_i})Z_{k_i}^T$ or $\|Z_{k_i}\nabla q_{k_i}(A_{k_i}s_{k_i}^n)\|\to 0$.

The first possibility occurs only when $\frac{\upsilon_{k_i}}{\rho_{k_i}}\to 0$ as $k_i\to\infty$ and $\frac{\tilde{\psi}_{k_i}\bar{s}_{k_i}^t}{\|V_{k_i}g_{k_i}\|}$ lies in the null space of the matrix $Z_{k_i}^TA_{k_i}\nabla g_{k_i}V_{k_i}\nabla g_{k_i}^TA_{k_i}Z_{k_i}$ which contradicts Assumption (4.42) and implies that a subsequence of the iteration sequence satisfies the Fritz John conditions in the limit. The second possibility implies as $k_i\to\infty$, $\| Z_{k_i}^T(\nabla_x\ell_{k_i}+G_{k_i}s_{k_i}^n+\rho_{k_i}A_{k_i}\nabla g_{k_i}V_{k_i}(g_{k_i}+(A_{k_i}\nabla g_{k_i})^Ts_{k_i}^n))\|\to 0$. Hence as $k_i\to\infty$, $\rho_{k_i}\|Z_{k_i}^TA_{k_i}\nabla g_{k_i}V_{k_i}(g_{k_i}+(A_{k_i}\nabla g_{k_i})^Ts_{k_i}^n)\|$ must be bounded. Also, $\|Z_{k_i}^T\nabla q_{k_i}(A_{k_i}s_{k_i}^n)\|\to 0$, implies that $\|\bar{s}_{k_i}^t\|\to 0$. Using the fact that $\|h_{k_i}\|\to 0$, implies $\|s_{k_i}^n\|\to 0$, we have

$$\nabla f_{k_i}+\nabla h_{k_i}\bar{\mu}_{k_i}+\nabla g_{k_i}\bar{\nu}_{k_i}=0,$$

for some $\bar{\mu}_{k_i}$ and $\bar{\nu}_{k_i}$. This implies that a subsequence of the iteration sequence satisfies the Fritz John conditions in the limit.

iii)If $\lim sup_{k\to\infty}\frac{\tilde{\psi}_k\bar{s}_k^t}{\|V_kg_k\|}<\infty$ and $\lim\inf_{k\to\infty}\frac{\tilde{\psi}_k\bar{s}_k^t}{\|V_kg_k\|}>0$. Therefore $\|\bar{s}_k^t\|\to 0$ and $\|s_k^n\|\to 0$ because $\|h_k\|\to 0$, $s_k\to 0$.

Hence, as in the second case, the right hand side of (4.45) goes to zero as $k\to\infty$. This implies that

$$\|Z_k^T\nabla q_k(A_ks_k^n)\|\frac{\|Z_k^T(\frac{\upsilon_k}{\rho_k}I+A_k\nabla g_kV_k\nabla g_k^TA_k)Z_k\tilde{\psi}_k\bar{s}_k^t\|}{\|Z_k^TA_k\nabla g_kV_k\nabla g_k^TA_k)Z_k\|\|V_kg_k\|}\to 0.$$

But this implies that asymptotically, either $\|Z_k^T\nabla q_k(A_ks_k^n)\|\to 0$, or $\frac{\|Z_k^T(\frac{\upsilon_k}{\rho_k}I+A_k\nabla g_kV_k\nabla g_k^TA_k)Z_k\tilde{\psi}_k\bar{s}_k^t\|}{\|Z_k^TA_k\nabla g_kV_k\nabla g_k^TA_k)Z_k\|\|V_kg_k\|}\to 0$. As the second case, the two possibilities imply that a subsequence of the iteration sequence satisfies the Fritz John conditions in the limit. This completes the proof. □

4.3 Convergence When ρ_k is bounded

We continue our analysis assuming that the parameter ρ_k is bounded. This means that, we assume the existence of an integer $\bar{k}$ such that for all $k\geq\bar{k}$, $\rho_k=\bar{\rho}<\infty$, and

$$\frac{1}{2}Tpred_k(\tilde{\psi}_k\bar{s}_k^t)\geq\|A_k\nabla g_kV_kg_k\|\min\{\|A_k\nabla g_kV_kg_k\|,\Delta_k\}. \quad (4.46)$$

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

From assumptions PA_3, PA_5, and Assumption (4.46), we can say that there exists a positive constant b_3 such that for all k

$$\|B_k\| \leq b_3, \quad \|Z_k^T B_k\| \leq b_3, \ \ and \ \ \|Z_k^T B_k Z_k\| \leq b_3, \tag{4.47}$$

where $B_k = A_k H_k A_k + \bar{\rho} A_k \nabla g_k V_k \nabla g_k^T A_k + \mathrm{diag}(\nabla_x \ell(x_k, \mu_k; \bar{\rho}))\mathrm{diag}(\eta_k)$.

Lemma 4.21. *Assume PA_1-PA_5. Then there exists a constant $K_{11} > 0$ that does not depend on k such that*

$$q_k(0) - q_k(A_k \tilde{\psi}_k s_k^n) - \Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k) \geq -K_{11} \tilde{\psi}_k \|h_k\|. \tag{4.48}$$

Proof. From (3.18), we have

$$\begin{aligned} q_k(0) - q_k(A_k \tilde{\psi}_k s_k^n) &= -(A_k \nabla_x \ell(x_k, \mu_k))^T \tilde{\psi}_k s_k^n - \frac{1}{2} \tilde{\psi}_k^2 {s_k^n}^T G_k s_k^n \\ &\quad + \frac{\bar{\rho}}{2}[\|V_k g_k\|^2 - \|V_k (g_k + (A_k \nabla g_k)^T \tilde{\psi}_k s_k^n)\|^2] \\ &= -(A_k \nabla_x \ell(x_k, \mu_k) + \bar{\rho} A_k \nabla g_k V_k g_k)^T \tilde{\psi}_k s_k^n \\ &\quad - \frac{1}{2} \tilde{\psi}_k^2 s_k^{nT} (G_k + \bar{\rho} A_k \nabla g_k V_k \nabla g_k^T A_k) s_k^n \\ &= -(A_k \nabla_x \ell(x_k, \mu_k) + \bar{\rho} A_k \nabla g_k V_k g_k)^T \tilde{\psi}_k s_k^n \\ &\quad - \frac{1}{2} \tilde{\psi}_k^2 s_k^{nT} B_k s_k^n. \end{aligned}$$

Hence,

$$\begin{aligned} q_k(0) - q_k(A_k \tilde{\psi}_k s_k^n) - \Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k) &= -(A_k \nabla_x \ell(x_k, \mu_k) + \bar{\rho} A_k \nabla g_k V_k g_k)^T \tilde{\psi}_k s_k^n \\ &\quad - \frac{1}{2} \tilde{\psi}_k^2 s_k^{nT} B_k s_k^n - \Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k) \\ &\geq -\tilde{\psi}_k \|A_k \nabla_x \ell(x_k, \mu_k)\| \|s_k^n\| - \bar{\rho} \tilde{\psi}_k \|A_k \nabla g_k V_k g_k\| \|s_k^n\| - \tilde{\psi}_k^2 \|B_k\| \|s_k^n\|^2 \\ &\quad - \|\Delta\mu_k\| \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k\| \\ &\geq -\tilde{\psi}_k [\|A_k \nabla_x \ell(x_k, \mu_k)\| + \bar{\rho} \|A_k \nabla g_k V_k g_k\| + \|B_k\| \|s_k^n\|] \|s_k^n\| \\ &\quad - \tilde{\psi}_k \|\Delta\mu_k\| \|A_k \nabla h_k\| \|s_k^n\|, \end{aligned}$$

where $-\tilde{\psi}_k^2 \geq -\tilde{\psi}_k$. Using Inequality (4.18), we obtain

$$\begin{aligned} & q_k(0) - q_k(A_k \tilde{\psi}_k s_k^n) - \Delta\mu_k^T (h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k) \\ & \quad \geq -\tilde{\psi}_k [(\|A_k \nabla_x \ell(x_k, \mu_k)\| + \bar{\rho} \|A_k \nabla g_k V_k g_k\| \\ & \qquad + \|B_k\| \|s_k^n\| + \|\Delta\mu_k\| \|A_k \nabla h_k\|) K_1] \|h_k\|. \end{aligned}$$

Under Assumptions PA_3, PA_4, and PA_5, the facts that $\|s_k^n\| \leq \delta_{max}$, and using (4.47), there exists $K_{11} > 0$ which is independent of k, such that Inequality (4.48) hold. This completes the proof. □

Lemma 4.22. *Assume PA_1-PA_5, then for all k,*

$$\begin{aligned} Pred_k \quad \geq \quad & \frac{1}{2} K_3 \tilde{\psi}_k \|Z_k^T \nabla q_k(A_k s_k^n)\| \min\left\{\Delta_k, \frac{\|Z_k^T \nabla q_k(A_k s_k^n)\|}{\|\bar{B}_k\|}\right\} \\ & + \|A_k \nabla g_k V_k g_k\| \min\{\|A_k \nabla g_k V_k g_k\|, \Delta_k\} \\ & - K_{11} \tilde{\psi}_k \|h_k\| + r_k [\|h_k\|^2 - \|h_k + (A_k \nabla h_k)^T \tilde{\psi}_k s_k\|^2]. \end{aligned} \tag{4.49}$$

Proof. From (3.17), we have

$$Pred_k = [q_k(A_k\tilde{\psi}_k s_k^n) - q_k(A_k\tilde{\psi}_k s_k)] + [q_k(0) - q_k(A_k\tilde{\psi}_k s_k^n) - \Delta\mu_k^T(h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k)] \\ + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2].$$

Using (4.23), the above equation can be written in the form

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

Using Inequalities (4.25), (4.46), and (4.48), we have

$$\begin{aligned} Pred_k &\geq \frac{1}{2}K_3\tilde{\psi}_k\|Z_k^T\nabla q_k(A_k s_k^n)\|\min\left\{\Delta_k, \frac{\|Z_k^T\nabla q_k(A_k s_k^n)\|}{\|\bar{B}_k\|}\right\} \\ &\quad +\|A_k\nabla g_k V_k g_k\|\min\{\|A_k\nabla g_k V_k g_k\|, \Delta_k\} \\ &\quad -K_{11}\tilde{\psi}_k\|h_k\| + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2]. \end{aligned}$$

This completes the proof. □

Lemma 4.23. *Assume PA_1-PA_5. Let k be the index of an iteration at which r_k is increased. Then there exists a constant $K_{12} > 0$ that does not depend on k, such that*

$$r_k\tilde{\psi}_k\min\{\|h_k\|, \delta_k\} \leq K_{12}. \tag{4.50}$$

Proof. Since r_k is increased at the k^{th} iteration and since $\rho_k = \bar{\rho}$ is bounded, then from (3.20), we can write

$$\begin{aligned} \frac{r_k}{2}[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2] &= [q_k(A_k\tilde{\psi}_k s_k) - q_k(A_k\tilde{\psi}_k s_k^n)] + [q_k(A_k\tilde{\psi}_k s_k^n) - q_k(0)] \\ &\quad + \Delta\mu_k^T(h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k) \\ &\quad + \frac{b_0}{2}[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2] \\ &= -\frac{1}{2}Tpred_k(\tilde{\psi}_k\bar{s}_k^t) - \frac{1}{2}Tpred_k(\tilde{\psi}_k\bar{s}_k^t) \\ &\quad + [q_k(A_k\tilde{\psi}_k s_k^n) - q_k(0) + \Delta\mu_k^T(h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k)] \\ &\quad + \frac{b_0}{2}[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2]. \end{aligned}$$

Applying Inequality (4.20) to the left hand side and Inequalities (4.25), (4.46), and (4.48) to the right hand side, we obtain

$$\begin{aligned} \frac{r_k}{2}K_2\tilde{\psi}_k\|h_k\|\min\{\delta_k, \|h_k\|\} &\leq -\frac{K_3}{2}\tilde{\psi}_k\|Z_k^T\nabla q_k(A_k s_k^n)\|\min\left\{\Delta_k, \frac{\|Z_k^T\nabla q_k(A_k s_k^n)\|}{\|\bar{B}_k\|}\right\} \\ &\quad - \|A_k\nabla g_k V_k g_k\|\min\{\|\nabla g_k V_k g_k\|, \Delta_k\} + K_{11}\tilde{\psi}_k\|h_k\| \\ &\quad + \frac{b_0}{2}\|h_k\|^2 \\ &\leq K_{11}\tilde{\psi}_k\|h_k\| + \frac{b_0}{2}\|h_k\|^2. \end{aligned}$$

The rest of the proof follows using the fact that $\tilde{\psi}_k \leq 1$ and assumption PA_3. □

Lemma 4.24. *Assume PA_1-PA_5. At any given iteration k at which $\|h_k\| \leq \tilde{\phi}\delta_k$ and $\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k)\| + \|A_k\nabla g_k V_k g_k\| \geq \varepsilon$, where $\varepsilon > 0$ and $\tilde{\phi}$ is a positive constant given by*

$$\tilde{\phi} \leq \min\left\{\frac{\varepsilon}{6b_3K_1\delta_{max}}, \frac{\sqrt{3}}{2K_1}, \frac{K_3\varepsilon}{24K_{11}}\min\left\{\frac{2\varepsilon}{3\delta_{max}}, 1\right\}, \frac{\varepsilon}{8K_{11}}\min\left\{\frac{2\varepsilon}{\delta_{max}}, 1\right\}\right\}, \tag{4.51}$$

there exists a positive constant K_{13} that depends on ε but does not depend on k, such that

$$Pred_k \geq K_{13}\tilde{\psi}\delta_k + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}s_k\|^2]. \tag{4.52}$$

Proof. Let $\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k)\| \geq \frac{\varepsilon}{2}$. Using Inequalities (4.18) and (4.47), we have

$$\begin{aligned}\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k + B_k s_k^n)\| &\geq \|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k)\| \\ &\quad - \|Z_k^T B_k s_k^n\| \\ &\geq \|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k)\| \\ &\quad - b_3K_1\|h_k\|.\end{aligned}$$

Since $\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k)\| \geq \frac{\varepsilon}{2}$, $\|h_k\| \leq \tilde{\phi}\delta_k$, and $\tilde{\phi} \leq \frac{\varepsilon}{6b_3K_1\delta_{max}}$, then we have

$$\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k + B_k s_k^n)\| \geq \frac{\varepsilon}{2} - b_3K_1\tilde{\phi}\delta_k \geq \frac{\varepsilon}{3}. \tag{4.53}$$

Since $\Delta_k = \sqrt{{\delta_k}^2 - \|s_k^n\|^2}$ and $\|s_k^n\| \leq K_1\|h_k\| \leq K_1\tilde{\phi}\delta_k \leq K_1\frac{\sqrt{3}}{2K_1}\delta_k = \frac{\sqrt{3}}{2}\delta_k$, 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,

$$\Delta_k \geq \frac{1}{2}\delta_k. \tag{4.54}$$

Since $\|h_k\| \leq \tilde{\phi}\delta_k$, $\tilde{\psi}_k \leq 1$, and using Inequalities (4.49), (4.53), and (4.54), then

$$\begin{aligned}Pred_k \geq\ & \frac{1}{2}K_3\tilde{\psi}_k\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k + B_k s_k^n)\| \\ & \min\left\{\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k + B_k s_k^n)\|, \frac{1}{2}\delta_k\right\} \\ & -K_{11}\tilde{\psi}_k\|h_k\| + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}s_k\|^2] \\ \geq\ & \frac{K_3\tilde{\psi}_k\varepsilon}{12}\delta_k\min\left\{\frac{2\varepsilon}{3\delta_{max}}, 1\right\} - K_{11}\tilde{\phi}\tilde{\psi}_k\delta_k + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2].\end{aligned}$$

Since $\tilde{\phi} \leq \frac{K_3\varepsilon}{24K_{11}}\min\{\frac{2\varepsilon}{3\delta_{max}}, 1\}$, then we have

$$Pred_k \geq \frac{K_3\tilde{\psi}_k\varepsilon}{24}\min\left\{\frac{2\varepsilon}{3\delta_{max}}, 1\right\}\delta_k + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2].$$

Now, consider the case when $\|A_k\nabla g_k V_k g_k\| \geq \frac{\varepsilon}{2}$. Using Inequalities (4.49) and (4.54), we have

$$\begin{aligned}Pred_k \geq\ & \|A_k\nabla g_k V_k g_k\|\min\left\{\|A_k\nabla g_k V_k g_k\|, \frac{1}{2}\delta_k\right\} \\ & -K_{11}\tilde{\psi}\|h_k\| + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2]\end{aligned}$$

$$\begin{aligned}
&\geq \tilde{\psi}\|A_k\nabla g_k V_k g_k\| \min\Big\{\|A_k\nabla g_k V_k g_k\|, \frac{1}{2}\delta_k\Big\} \\
&\quad -K_{11}\tilde{\psi}\|h_k\| + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2] \\
&\geq \frac{\tilde{\psi}\varepsilon}{4}\min\Big\{\frac{2\varepsilon}{\delta_{\max}}, 1\Big\}\delta_k - K_{11}\tilde{\psi}\tilde{\phi}\delta_k + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2].
\end{aligned}$$

Since $\tilde{\phi} \leq \frac{\varepsilon}{8K_{11}}\min\{\frac{2\varepsilon}{\delta_{\max}}, 1\}$, we have

$$Pred_k \geq \frac{\tilde{\psi}\varepsilon}{8}\min\Big\{\frac{2\varepsilon}{\delta_{\max}}, 1\Big\}\delta_k + r_k[\|h_k\|^2 - \|h_k + (A_k\nabla h_k)^T\tilde{\psi}_k s_k\|^2].$$

Take $K_{13} = \min\big\{\frac{K_3\varepsilon}{24}\min\{\frac{2\varepsilon}{3\delta_{max}}, 1\}\ ,\ \frac{\varepsilon}{8}\min\{\frac{2\varepsilon}{\delta_{\max}}, 1\}\big\}$, the result follows. □

From the above lemma, we can easily see that, at any iteration at which either $\|Z_k^T(A_k\nabla_x\ell(x_k,\mu_k) + \bar{\rho}A_k\nabla g_k V_k g_k)\| \geq \frac{\varepsilon}{2} > 0$ or $\|A_k\nabla g_k V_k g_k\| \geq \frac{\varepsilon}{2} > 0$ and $\|h_k\| \leq \tilde{\phi}\delta_k$, where $\tilde{\phi}$ is given by (4.51), there is no need to increase the value of the penalty parameter. *i.e.*, r_k is increased only when $\|h_k\| \geq \tilde{\phi}\delta_k$.

Lemma 4.25. *Assume PA_1-PA_5. If at the j^{th} trial iterate of any iteration indexed k, the penalty parameter r_{k^j} is increased, then there exists a positive constant K_{14} that does not depend on k or j, such that*

$$r_{k^j}\tilde{\psi}_{k^j}\|h_k\| \leq K_{14}. \tag{4.55}$$

Proof. The proof follows directly from Inequalities (4.32) and (4.50). □

Lemma 4.26. *Assume PA_1-PA_5. If $r_k \to \infty$, then*

$$\lim_{k_i\to\infty}\|h_{k_i}\| = 0, \tag{4.56}$$

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

Proof. The proof follows directly from the above lemma and $\lim_{k\to\infty}\tilde{\psi}_k = 1$. □

$\boxed{4.4}$ Main 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 theorem, we prove that the sequence $\{\|h_k\|\}$ converges to zero.

Theorem 4.27. *Assume PA_1-PA_5. Then the sequence of iterates generated by the algorithm satisfies*

$$\lim_{k\to\infty}\|h_k\| = 0. \tag{4.57}$$

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 (4.12), 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:
i) 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 (4.26), for k sufficiently large, the two sequences $\{k_i\}$ and $\{k_j\}$ do not have common elements. Let k_{a_1} and k_{a_2} be two consecutive iterates at which the penalty parameter r_k is increased and $k_{a_1} < k < k_{a_2}$, where $k \in \{k_j\}$.

The penalty parameter r_k is the same for all iterates that lie between k_{a_1} and k_{a_2}. Since all the iterates of $\{k_j\}$ are acceptable, then for all $k \in \{k_j\}$,

$$\Phi_k - \Phi_{k+1} = Ared_k \geq \tau_1 Pred_k.$$

From Inequality (4.21) and the above inequality, we can write

$$\frac{\Phi_k - \Phi_{k+1}}{r_k} \geq \frac{\tau_1 K_2 \tilde{\psi}_k}{2} \|h_k\| \min\{\|h_k\|, \delta_k\}.$$

Summing over all acceptable iterates that lie between k_{a_1} and k_{a_2}, we have

$$\sum_{k=k_{a_1}}^{k_{a_2}-1} \frac{\Phi_k - \Phi_{k+1}}{r_k} \geq \frac{\tau_1 K_2 \tilde{\psi}_k \varepsilon}{4} \min\left\{\hat{K_9}, \frac{\varepsilon}{2}\right\},$$

where $\hat{K_9}$ is as K_9 in (4.34), with ε is replaced by $\frac{\varepsilon}{2}$. Hence,

$$\frac{\ell(x_{k_{a_1}}, \mu_{k_{a_1}}; \bar{\rho}) - \ell(x_{k_{a_2}}, \mu_{k_{a_2}}; \bar{\rho})}{r_{k_{a_1}}} + [\|h_{k_{a_1}}\|^2 - \|h_{k_{a_2}}\|^2] \geq \frac{\tau_1 K_2 \varepsilon}{4} \min\left\{\hat{K_9}, \frac{\varepsilon}{2}\right\}.$$

Since $r_k \to \infty$, then for k_{a_1} sufficiently large, we have

$$\frac{|\ \ell(x_{k_{a_1}}, \mu_{k_{a_1}}; \bar{\rho}) - \ell(x_{k_{a_2}}, \mu_{k_{a_2}}; \bar{\rho})\ |}{r_{k_{a_1}}} < \frac{\tau_1 K_2 \varepsilon}{8} \min\left\{\hat{K_9}, \frac{\varepsilon}{2}\right\}.$$

Therefore,

$$\|h_{k_{a_1}}\|^2 - \|h_{k_{a_2}}\|^2 \geq \frac{\tau_1 K_2 \varepsilon}{8} \min\left\{\hat{K_9}, \frac{\varepsilon}{2}\right\}.$$

But this leads to a contradiction with Lemma (4.26) unless $\varepsilon = 0$.
ii) 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 (4.21), we have for any $\hat{k} \in \{k_j\}$ and $\hat{k} \geq \tilde{k}$

$$Pred_{\hat{k}} \quad \geq \quad \frac{\tilde{r} K_2 \tilde{\psi}_{\hat{k}}}{2} \|h_{\hat{k}}\| \min\{\delta_{\hat{k}}, \|h_{\hat{k}}\|\} \geq \frac{\varepsilon \tilde{r} K_2 \tilde{\psi}_{\hat{k}}}{4} \min\left\{\frac{\varepsilon}{2\delta_{max}}, 1\right\} \delta_{\hat{k}}. \tag{4.58}$$

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

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} = Ared_{\hat{k}} \geq \tau_1 Pred_{\hat{k}}.$$

Hence, from Inequality (4.58) and the above inequality we have

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} \geq \frac{\tau_1 \varepsilon \tilde{r} K_2 \tilde{\psi}_{\hat{k}}}{4} \min\left\{\frac{\varepsilon}{2\delta_{max}}, 1\right\} \delta_{\hat{k}}.$$

Using Lemma (4.15) and the above inequality, we have

$$\Phi_{\hat{k}} - \Phi_{\hat{k}+1} \geq \frac{\tau_1 \varepsilon \tilde{r} K_2 \tilde{\psi}_{\hat{k}}}{4} \min\left\{\frac{\varepsilon}{2\delta_{max}}, 1\right\} \hat{K_9} > 0,$$

where $\hat{K_9}$ 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. $\square$

Theorem 4.28. *Assume PA_1-PA_5. Then the sequence of iterates generated by the algorithm satisfies*

$$\liminf_{k\to\infty} \ [\ \|Z_k^T A_k \nabla_x \ell_k\| + \|A_k \nabla g_k V_k g_k\|\ ] \ = \ 0. \tag{4.59}$$

Proof. First, we prove that

$$\liminf_{k\to\infty} [\|Z_k^T (A_k \nabla_x \ell_k + \bar{\rho} A_k \nabla g_k V_k g_k)\| + \|A_k \nabla g_k V_k g_k \ \|\ ] \ = \ 0. \tag{4.60}$$

We prove (4.60) by contradiction. Suppose that, for all k, $\| \ Z_k^T A_k(\nabla_x \ell_k + \bar{\rho} \nabla g_k V_k g_k) \ \| + \|A_k \nabla g_k V_k g_k\| > \varepsilon$. Assume that there exists an infinite subsequence $\{k_i\}$ such that $\|h_{k_i}\| > \tilde{\phi}\delta_{k_i}$, where $\tilde{\phi}$ satisfy (4.51). 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.
i) If $\{r_k\}$ is unbounded. For the rejected trial step $j-1$ of iteration k, we have $\|h_k\| > \tilde{\phi}\delta_{k^j} = \alpha_1 \tilde{\phi} \|s_{k^{j-1}}\|$. Using Inequalities (4.21) and (4.30) and the fact that the trial step $s_{k^{j-1}}$ was rejected, we have

$$\begin{aligned}
(1-\tau_1) \quad &\leq \quad \frac{|Ared_{k^{j-1}} - Pred_{k^{j-1}}|}{Pred_{k^{j-1}}} \\
&\leq \quad \frac{[2K_6\|s_{k^{j-1}}\| + 2K_7 r_{k^{j-1}} \|s_{k^{j-1}}\| \|h_k\| + 2K_8 r_{k^{j-1}} \|s_{k^{j-1}}\|^2]}{r_{k^{j-1}} K_2 \min(\alpha_1\tilde{\phi}, 1)\|h_k\|} \\
&\leq \quad \frac{2K_6}{r_{k^{j-1}} K_2 \alpha_1 \tilde{\phi} \min(\alpha_1 \tilde{\phi}, 1)} + \frac{2K_7 + 2K_8\alpha_1\tilde{\phi}}{K_2\alpha_1\tilde{\phi}\min(\alpha_1\tilde{\phi},1)} \|s_{k^{j-1}}\|.
\end{aligned}$$

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

$$r_{k^{j-1}} > \frac{4K_6}{K_2\alpha_1\tilde{\phi}\min(\alpha_1\tilde{\phi},1)(1-\tau_1)}.$$

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

$$\|s_{k^{j-1}}\| \geq \frac{K_2\alpha_1\tilde{\phi}\min(\alpha_1\tilde{\phi},1)(1-\tau_1)}{4(K_7 + K_8\alpha_1\tilde{\phi})}.$$

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

$$\delta_{k^j} = \alpha_1 \|s_{k^{j-1}}\| \geq \frac{K_2\alpha_1^2\tilde{\phi}\min(\alpha_1\tilde{\phi},1)(1-\tau_1)}{4(K_7 + K_8\alpha_1\tilde{\phi})}.$$

This gives a contradiction. So δ_{k^j} can not go to zero in this case.
ii) If 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 $\|h_k\| > \tilde{\phi}\delta_{k^j}$.

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

$$\|h_{k^l}\| > \tilde{\phi}\delta_{k^l}, \tag{4.61}$$

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

$$(1-\tau_1) \leq \frac{|Ared_{k^l} - Pred_{k^l}|}{Pred_{k^l}} \leq \frac{2K_5\|s_{k^l}\|}{K_2\min(\tilde{\phi},1)\|h_k\|}.$$

Hence,

$$\begin{aligned}\delta_{k^j} = \alpha_1 \|s_{k^{j-1}}\| &\geq \frac{\alpha_1 K_2 \min(\tilde{\phi},1)(1-\tau_1)\|h_k\|}{2K_5} \geq \frac{\alpha_1 K_2 \min(\tilde{\phi},1)(1-\tau_1)\tilde{\phi}}{2K_5}\delta_{k^1} \\ &\geq \frac{\alpha_1 K_2 \min(\tilde{\phi},1)(1-\tau_1)\tilde{\phi}}{2K_5}\delta_{\min}.\end{aligned}$$

Hence δ_{k^j} is bounded in this case too. If $j > 1$ and (4.61) does not hold for all l, there exists an integer m_1 such that (4.61) holds for $l = m_1 + 1, \ldots, j$ and

$$\|h_{k^l}\| \leq \tilde{\phi}\delta_{k^l}, \tag{4.62}$$

for $l = 1, \ldots, m_1$. As in the above case, we can write

$$\delta_{k^j} \geq \frac{\alpha_1 K_2 \min(\alpha,1)(1-\tau_1)}{2K_5}\|h_k\| \geq \frac{\alpha_1 K_2 \min(\tilde{\phi},1)(1-\tau_1)\tilde{\phi}}{2K_5}\delta_{k^{m_1+1}}. \tag{4.63}$$

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

$$\delta_{k^{m_1+1}} \geq \alpha_1 \|s_{k^{m_1}}\|. \tag{4.64}$$

Now, using (4.62), Lemma (4.24), and the fact that the trial steps $s_{k^{m_1}}$ is rejected, we can write

$$(1-\tau_1) \leq \frac{|Ared_{k^{m_1}} - Pred_{k^{m_1}}|}{Pred_{k^{m_1}}} \leq \frac{2K_5\bar{r}\|s_{k^{m_1}}\|}{K_{13}}.$$

This implies

$$\|s_{k^{m_1}}\| \geq \frac{K_{13}(1-\tau_1)}{2K_5\bar{r}}.$$

This implies that, $\|s_{k^{m_1}}\|$ is bounded. This fact together with (4.63) and (4.64) 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 $\|h_k\| \leq \tilde{\phi}\delta_{k^j}$. This implies using Lemma (4.23) 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 (4.23), we have

$$\Phi_{k^j} - \Phi_{k^j+1} = Ared_{k^j} \geq \tau_1 Pred_{k^j} \geq \tau_1 K_{13}\delta_{k^j}.$$

As k goes to infinity the above inequality implies that

$$\lim_{k\to\infty} \delta_{k^j} = 0. \tag{4.65}$$

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 (4.11) and (4.23), we must have

$$(1-\tau_1) < \frac{\bar{r}K_5\|s_{k^{j-1}}\|^2}{K_{13}\delta_{k^{j-1}}}.$$

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

$$\delta_{k^j} = \alpha_1 \|s_{k^{j-1}}\| > \frac{\alpha_1 K_{13}(1-\tau_1)}{\bar{r}K_5}.$$

Hence δ_{k^j} is bounded. But this contradicts (4.65). The supposition is wrong. Hence,

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

But this also implies (4.59). This completes the proof of the theorem. □

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

5 Numerical Results

In this section, we present the numerical results of the interior-point trust-region Algorithm 3.1 which have been performed on a laptop with Intel Core (TM)i7-2670QM CPU 2.2 GHz and 8 GB RAM. Algorithm 3.1 was implemented as a MATLAB code and run under MATLAB version 7.10.0.499 (R2010a).

Given a starting point $x_0 \in int(\boldsymbol{F})$, we choose the initial trust-region radius $\delta_0 = max(\|s_0^{ncp}\|, \delta_{min})$, where $\delta_{min} = 10^{-3}$. We choose the maximum trust-region radius $\delta_{max} = 10^3\delta_0$. The values of the constants that are needed in Step 0 of Algorithm 3.1 were set $\tau_1 = .25$, $\tau_2 = 0.75$, $\alpha_1 = 0.5$, $\alpha_2 = 2$, $\varepsilon_1 = 10^{-8}$, $\varepsilon_2 = 10^{-10}$ and $\theta = 0.9995$.

Successful termination with respect to our trust-region algorithm means that the termination condition of the algorithm is met with $\varepsilon_1 = 10^{-8}$. On the other hand, unsuccessful termination means that the number of iterations is greater than 300, the number of function evaluations is greater than 500, or the length of the trial step is less than ε_2. A flowchart of Algorithm 3.1 as shown in Figure(6.1)

The results Algorithm 3.1 are reported in Table 1 where the test problems are numbered in the same way as in [14]. For example, HS53 is the problem 53 in [14]. For comparison, we have included the corresponding results obtained by a trust-region algorithm combining line search filter technique for nonlinear constrained optimization in [19] and Lancelot [3]. For all problems, these algorithms achieved the same optimal solution at the same starting points in [14].

In many of the test problems reported in Table 1, the number of iterations (iter) and the number of function evaluations (nfunc) of Algorithm 3.1 are better than those obtained by method [3] or method [19]. This indicates the viability of our approach. However, we believe that our algorithm needs to be refined with efficiency in mined to be suitable for bounded large-scale problems.

6 Concluding Remarks

We described an interior-point active-set trust-region algorithm for solving general nonlinear programming problem with bound on variables. The algorithm handles inequality constraints in a fashion similar to the approach of [5] for treating the active constraints. In this algorithm, an active set strategy is used together with a Coleman-Li strategy and a projected Hessian technique to transform the computation of the trial step at each iteration to two easy trust-region sub-problems similar to the trust region sub-problems of the unconstrained optimization problem.

We proved that the algorithm is globally convergent under mild conditions and a subsequent of the sequence of iterate generated by the algorithm converges to either Fritz John point, or an infeasible Fritz John point or KKT point.

For future work, there are many question should be answered.

- Although we have implemented the algorithm and tested it, we believe that the implementation of the algorithm should be refined with efficiency in mined. In particular, a better way of solving the trust-region subproblems that can handle large-scale bound constrained optimization problems should be used.
- Improving the proposed algorithm to be capable for treating nondifferentiation bound constrained optimization problem with equality and inequality constraints.
- Updating the Lagrange multiplier is another point that needs to be refined. In particular, an inexpensive way for updating the Lagrange multiplier is needed. This indeed will reduce the cost of the computation of the steps.
- A related important question that has to be looked at is how to use a secant approximation of the Hessian of the Lagrangian matrix in order to produce a more efficient algorithm.

Acknowledgments

I am very grateful to the anonymous reviewers for their valuable and insightful comments, which have aided us in improving the quality of this paper. I would like to thank my professor Dr.El-Alem for his helpful comments.

References

[1] R. Byrd, *Robust Trust-Region Methods for Nonlinearly Constrained Optimization*, A talk presented at the Second SIAM Conference on Optimization, Houston, TX 1987.

[2] T. R. Coleman and Y. Li, An interior trust region approach for nonlinear minimization subject to bounds, *SIAM J. Optimization* 6 (1996) 418-445.

[3] A. Conn, N. Could, and L. Toint, *(LANCELOT): A Fortran Package for Large-Scale Nonlinear (Release A) Optimization*, No.17 in springer series in computational mathematics, Springer-Verlag, New York, 1992 .

[4] 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* 7 (1997) 177–207.

[5] J. Dennis, M. El-Alem, and K. Williamson, A trust-region approach to nonlinear systems of equalities and inequalities, *SIAM J Optimization* 9 (1999) 291–315.

[6] J. Dennis, M. Heinkenschloss and L. Vicente, Trust-region interior-point SQP algorithms for a class of nonlinear programming problems, *SIAM J. Control and Optimization* 36 (1998) 1750–1794.

[7] 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.

[8] M. El-Alem, M. El-Sayed, and B. El-Sobky, Local convergence of interior-point Newton method for constrained optimization, *J. Optimization Theory and Applications* 120 (2004) 487–502.

[9] B. El-Sobky, A robust trust-region algorithm for general nonlinear constrained optimization problems, PhD thesis, Department of Mathematics, Alexandria University, Alexandria, Egypt, 1998.

[10] B. El-Sobky, A multiplier active-set trust-region algorithm for solving constrained optimization problem, *Applied Mathematics and Computation* 219 (2012), 127–157.

[11] B. El-Sobky, A new convergence theory for trust-region algorithm for solving constrained optimization problems, *Applied Mathematics Science* 7 (2013), 5469–5489.

[12] B. El-Sobky, A penalty active-set trust-region algorithm for solving general nonlinear programming problem, *Sylwan* 9 (2014), 273–290.

[13] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, John Wiley and Sons, New York, 1968.

[14] W. Hock and K. Schittkowski, *Test Examples for Nonlinear Programming Codes*, volume 187 of Lecture notes in economics and mathematical systems, Springer-Verlag, Berlin, 1981.

[15] O. Mangasarian, *Nonlinear Programming*, McGraw-Hill Book Company, New York, 1969.

[16] L.F. Niu and Y. Yuan, A new trust region algorithm for nonlinear constrained optimization, *J. Computational Mathematics* 28 (2010) 72–86.

[17] 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.

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

[19] Y. Pei and D. Zhu, A trust-region algorithm combining line search filter technique for nonlinear constrained optimization, *Int. J. Computer Mathematics* 91 (2014) 1817–1839.

[20] T. Steihaug, The conjugate gradient method and trust region in large scale optimization, *SIAM J.Numer. Anal.* 20 (1983) 626–637.

[21] P. L. Toint, *Towards an Efficient Sparsity Exploiting Newton's Method for Minimization*, Academic Press, New York, 1981, pp. 57–87.

[22] 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.

[23] X. Wang and Y. Yuan, A trust region method based on a new affine scaling technique for simple bounded optimization, *Optimization Methods and Software* 28 (2013) 871–888.

[24] X. Wang and Y. Yuan, An augmented Lagrangian trust region method for equality constrained optimization, *Optimization Methods and Software* 30 (2015), 559–582.

[25] Y. Yuan, On the convergence of a new trust region algorithm, *Numer. Math.* 70 (1995) 515–539.

[26] Y. Yuan, Recent advances in trust region algorithms. *Math. Program. Ser. B* 151 (2015) 249–281.

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

Manuscript received 21 January 2016
revised 25 September 2016, 23 November 2016
accepted for publication 24 November 2016

Bothina El-Sobky
Department of Mathematics, Faculty of Science
Alexandria University, Alexandria, Egypt
E-mail address: bothinaelsobky@yahoo.com

Problem name	Method in [3]		Method in [19]		Algorithm 3.1	
	iter	nfunc	iter	nfunc	iter	nfunc
HS6	49	56	12	21	4	5
HS7	18	19	9	21	7	8
HS8	-	-	2	2	5	7
HS9	4	5	7	8	6	8
HS10	17	18	-	-	17	20
HS11	15	16	-	-	20	23
HS12	22	23	-	-	12	15
HS14	12	13	-	-	13	14
HS16	15	16	-	-	17	20
HS26	-	-	20	50	12	13
HS28	-	-	5	6	5	6
HS30	7	8	-	-	3	4
HS33	12	12	-	-	13	14
HS34	19	19	-	-	21	23
HS39	20	21	21	66	12	13
HS40	10	11	16	56	5	6
HS41	6	7	-	-	8	9
HS42	12	13	9	12	3	4
HS47	-	-	19	27	5	6
HS48	-	-	7	9	4	5
HS49	-	-	32	42	9	10
HS50	-	-	11	12	15	20
HS51	-	-	4	6	4	5
HS52	-	-	7	7	7	14
HS53	6	7	-	-	4	5
HS60	15	15	-	-	6	7
HS77	22	24	-	-	9	10
HS78	11	11	16	59	4	7
HS79	9	10	10	13	6	7
HS80	14	15	-	-	5	6
HS81	16	17	-	-	5	6

Table 1: Comparison of method in [3] and [19] with Algorithm 3.1 respectively.

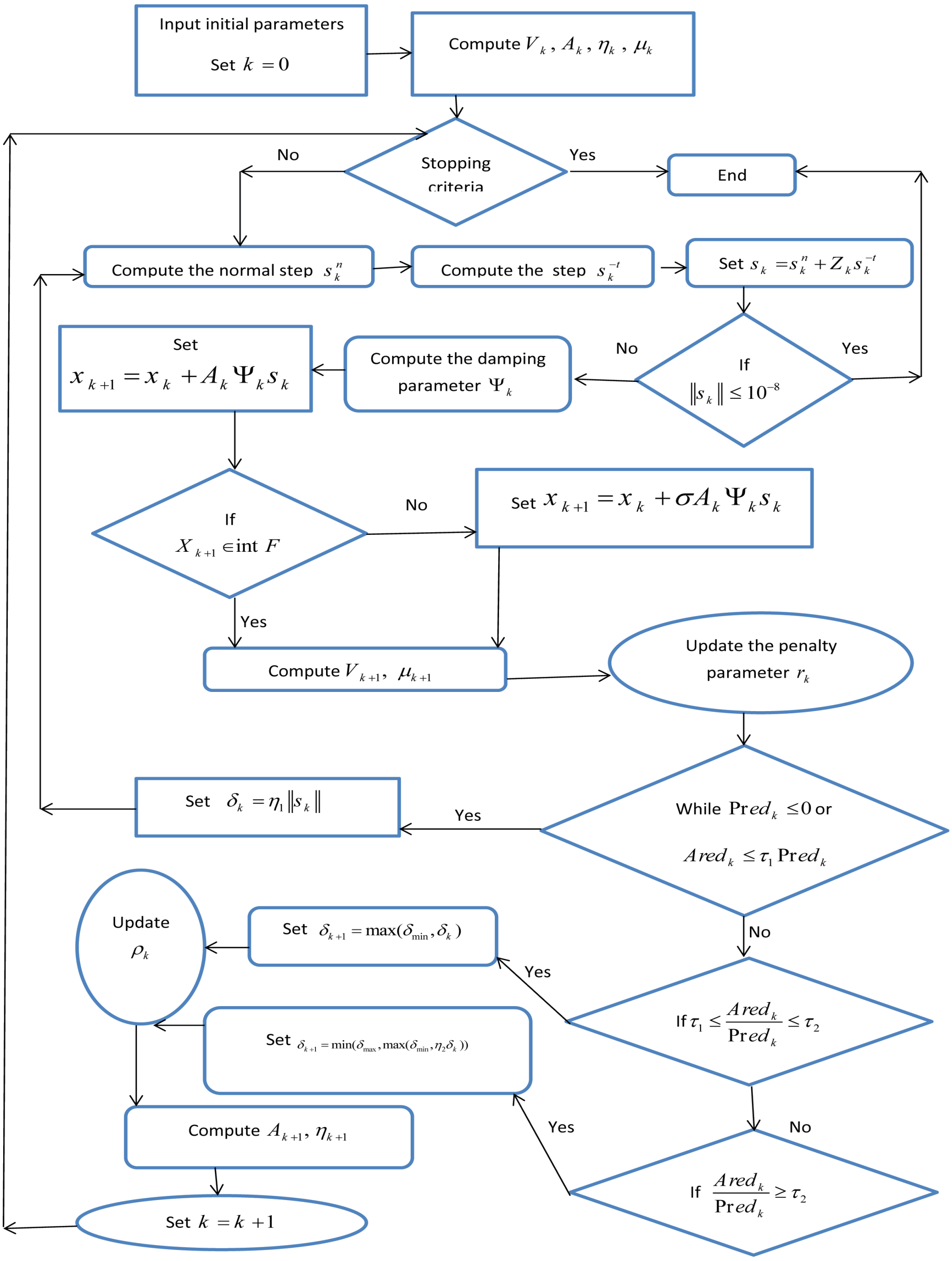

Figure 1: flowchart of Algorithm 3.1