A nonmonotone trust-region algorithm with nonmonotone penalty parameters for constrained optimization

Zhongwen Chen\textsuperscript{a,*}\textsuperscript{,1}, Xiangsun Zhang\textsuperscript{b,2}

\textsuperscript{a}Department of Mathematics, Suzhou University, Suzhou 215006, PR China
\textsuperscript{b}Institute of Applied Mathematics, AMSS, CAS, Beijing 100080, PR China

Received 4 January 2002; received in revised form 29 December 2003

Abstract

In this paper, we present a nonmonotone trust-region algorithm with nonmonotone penalty parameters for the solution of optimization problems, with nonlinear equality constraints and bound constraints. The proposed algorithm combines an SQP approach with a trust-region strategy to globalize the process. Each step is obtained through the computation of a normal step (to reduce infeasibility) and a tangential step (to decrease some merit function). The algorithm makes use of an augmented Lagrangian function as merit function, and allows the value of this merit function and the penalty parameter involved in it to decrease non-monotonically. The global convergence theory for the proposed algorithm is developed without regularity assumption, and shows that any limit point of the sequence generated by the algorithm is a $\varphi$-stationary point, while at least one limit point, under the suitable assumptions, is a substationary point (and a stationary point if it is feasible). Some preliminary numerical experiments are also reported.

\textcopyright{} 2003 Elsevier B.V. All rights reserved.

\textbf{MSC:} 65K05

\textbf{Keywords:} Trust-region method; Nonmonotone; Constrained optimization

---

\* Corresponding author.
E-mail address: wenchen@pub.sz.jsinfo.net (Z. Chen).

\textsuperscript{1}Supported by Chinese NSF Grant 10201026 and Jiangsu NSF BK2002037.

\textsuperscript{2}Supported by Chinese NSF Grant 39830070.
1. Introduction

Every minimization problem with nonlinear equality and inequality constraints can be reduced, by means of the introduction of slack variables, to the standard form

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad c(x) = 0, \\
& \quad l \leq x \leq u,
\end{align*}
\]

where \(x \in \mathbb{R}^n\), \(c(x) = (c_1(x), c_2(x), \ldots, c_m(x))^T\), \(m < n\), \(f(x)\) and \(c_i(x) (i=1,2,\ldots,m)\) are real functions defined in \(\mathcal{X} = \{x \in \mathbb{R}^n \mid l \leq x \leq u\}\) and \(-\infty \leq l_i < u_i \leq +\infty (i=1,2,\ldots,n)\). All along the paper, we denote by \(g(x)\) the gradient of \(f(x)\) and \(A(x) = (\nabla c_1(x), \ldots, \nabla c_m(x))\). Let \(\|\cdot\|\) be the \(l_2\)-norm on \(\mathbb{R}^n\) and also use \(f_k\) for \(f(x_k)\), etc. We define a \(\varphi\)-stationary point and a substationary point of (1.1) as follows.

**Definition 1.1** (Gomes et al. [8]). For some \(x^* \in \mathcal{X}\), if there exist \(\mu_l \geq 0, \mu_u \geq 0, \mu_l, \mu_u \in \mathbb{R}^n\), such that

\[
A(x^*)c(x^*) - \mu_l + \mu_u = 0, \\
\mu_l^T(x^* - l) = 0, \quad \mu_u^T(u - x^*) = 0,
\]

then \(x^*\) is called a \(\varphi\)-stationary point of (1.1).

**Definition 1.2** (Chen et al. [3]). For some \(x^* \in \mathcal{X}\), if there exist \(\mu_l \geq 0, \mu_u \geq 0\) and \(\lambda \in \mathbb{R}^m\) such that

\[
g(x^*) + A(x^*)\lambda - \mu_l + \mu_u = 0, \\
\mu_l^T(x^* - l) = 0, \quad \mu_u^T(u - x^*) = 0,
\]

then \(x^*\) is called a substationary point of problem (1.1). Moreover, if \(c(x^*) = 0\), then \(x^*\) is called a stationary point of problem (1.1).

Moreover, a point \(x\) is said to be feasible if \(x \in \mathcal{X}\) and \(c(x) = 0\). A feasible point \(x\) is said to be regular if the gradients of the active constraints at \(x\) are linearly independent. It is obvious that a regular substationary point is a stationary point.

For a general problem such as (1.1), we do not know whether the feasible set \(\mathcal{F} = \{x \in \mathbb{R}^n \mid c(x)=0, l \leq x \leq u\}\) is a nonempty set. If \(\mathcal{F}=\emptyset\), that is, (1.1) has no feasible solution, then it has no stationary point. In other words, we can only obtain a substationary point or a \(\varphi\)-stationary point of (1.1). For example, we consider

\[
\begin{align*}
\min & \quad f(x) = (x_1 - 1)^2 + x_2^2 \\
\text{s.t.} & \quad 2x_1 + (x_2 - 2)^2 = 0, \\
& \quad 0 \leq x_1, x_2 \leq 1.
\end{align*}
\]

Take \(x^0=(0,1)^T\), we can check that \(x^0\) is a \(\varphi\)-stationary point, furthermore, it is also a substationary point. But \(c(x^0) \neq 0\) and the example is infeasible.
Formulation (1.1) is used in many successful practical algorithms for nonlinear programming, like those based on the generalized reduced gradient (see [11]) and on the augmented Lagrangian approach (see [4]).

Given \( x_k \in \mathcal{X} \) an estimate of the solution, problem (1.1) is also often solved by sequential quadratic programming (SQP) methods, and it is assumed that a search direction \( d_k \) can be computed by solving the following quadratic programming subproblem

\[
\begin{align*}
\min & \quad g_k^T d + \frac{1}{2} d^T B_k d \\
\text{s.t.} & \quad c_k + A_k^T d = 0, \\
& \quad l \leq x_k + d \leq u,
\end{align*}
\]  

where \( B_k \in \mathbb{R}^{n \times n} \) is a symmetric matrix. The new iterate is then set to \( x_{k+1} = x_k + \alpha_k d_k \), where \( \alpha_k > 0 \) is a step length and \( \alpha_k \) depends on some line search techniques [1].

In this paper, we use the concepts of the substationary point and the \( \varphi \)-stationary point and introduce a general algorithm, based on SQP and trust regions for solving (1.1). In particular, most trust-region algorithms proposed up to now are descent methods, in that they only accept the trial point as next iterate if its merit function value is strictly lower than that at the current iterate. This monotonicity property ensures that each “successful iteration” produces a point that is better than any other point found so far, a property which is heavily used in the theoretical justifications for such algorithms. Toint [16] pointed out that abandoning this algorithmic restriction allows the sequence of iterates to follow the bottom of curved narrow valleys (a common occurrence in difficult nonlinear problems) much more loosely, which hopefully results in longer and more efficient steps. References [6,10], etc., also discuss nonmonotone trust-region methods.

In this work, we use the following augmented Lagrangian function as merit function as [8]

\[
P(x, \lambda, \theta) = \theta \ell(x, \lambda) + (1 - \theta) \varphi(x), \quad \theta \in [0, 1],
\]  

(1.3)

where

\[
\ell(x, \lambda) = f(x) + \lambda^T c(x), \quad \varphi(x) = \frac{1}{2} \|c(x)\|^2,
\]

\( \lambda \in \mathbb{R}^m \) is the Lagrange multiplier. In the convergence theory, the monotone decrease of \( \theta \) is not necessary. Gomes et al. [8] define a nonmonotone strategy that ensures convergence and allows one to test different “degrees of nonmonotonicity”. Their numerical results show that the nonmonotone strategy algorithm spends less time and takes less iterations than the monotone strategy one in many occasions.

This work generalizes and modifies, in many aspects, the approach given in [2,3,7,8,13,14,16], for constrained optimization. Our algorithms are nonmonotone methods, i.e., the value of the merit function does not decrease monotonically, no monotone property of the penalty parameter, either. At each iterate, two subproblems are solved. Under no regularity assumption, we prove that any limit point of the sequence generated by our algorithm is a \( \varphi \)-stationary point or one of the limit points is a substationary point of (1.1). Furthermore, under the suitable conditions, the substationary point is just a stationary point.

This paper is organized as follows: Section 2 introduces the main model algorithm. In Section 3, it is proved that, under mild conditions, this algorithm is well defined and some global convergence results are given. In Section 4, we report some numerical experiments. Some conclusions are stated in Section 5.
2. Description of the main algorithm

Let $x_k \in \mathcal{X}$ and $\lambda_k \in \mathbb{R}^m$ be an approximate solution of (1.1) and Lagrange multiplier estimates at the $k$th iteration, respectively. Then many sequential quadratic programming (SQP) methods for solving problem (1.1) obtain a search direction $d_k$ by solving subproblem (1.2). However, (1.2) may be infeasible, especially when a trust-region constraint is added. One way to overcome the infeasibility is the known Byrd–Omojokun strategy, which splits a step into its normal and tangential components. The purpose of a normal step is to reduce infeasibility while one of a tangential step is to decrease the merit function values. Therefore, we first consider solving the following subproblem:

$$\min \Phi_k(v) = \frac{1}{2} \| c_k + A_k^T v \|^2 + \frac{1}{2} \| v \|^2$$

s.t. $l \leq x_k + v \leq u,
\| v \|_\infty \leq 0.8 \Delta,$

(2.1)

where $\Delta > 0$ is a trust region radius and the regularizing term $\| v \|^2/2$ guarantees that (2.1) is a strictly convex programming problem, which implies from $v_k(\Delta) = 0$ that $x_k$ is a $\varphi$-stationary point of (1.1) (see Lemma 3.1), where $v_k(\Delta)$ is the solution to subproblem (2.1). We then solve the following subproblem:

$$\min \Psi_k(d) = \nabla f_k^T d + \frac{1}{2} d^T B_k d$$

s.t. $A_k^T d = A_k^T v_k(\Delta),
\| d \|_\infty \leq \Delta,$

(2.2)

where $B_k \in \mathbb{R}^{n \times n}$ is an approximate Hessian matrix of the Lagrangian function for problem (1.1) at $x_k$. Since $v_k(\Delta)$ is feasible for (2.2), there exists an optimal solution to (2.2). Let $d_k(\Delta)$ be a solution to (2.2). Having determined the trial step $d_k(\Delta)$, We define the actual reduction of the merit function (1.3) from $x_k$ to $x_k + d_k(\Delta)$ as

$$\text{ared}_k(\Delta) = P(x_k, \lambda_k, \theta_k) - P(x_k + d_k(\Delta), \lambda_k + \delta \lambda_k, \theta_k)$$

and the predicted reduction as

$$\text{pred}_k(\Delta) = \theta_k \left[ - \nabla f_k^T d_k(\Delta) - \frac{1}{2} d_k(\Delta)^T B_k d_k(\Delta) - (c_k + A_k^T d_k(\Delta))^T \delta \lambda_k \right] + \frac{1}{2} (1 - \theta_k) (\| c_k \|^2 - \| c_k + A_k^T d_k(\Delta) \|^2),$$

where $\lambda_k + \delta \lambda_k$ is a new Lagrange multiplier estimate at the trial point $x_k + d_k(\Delta)$. In order to allow the nonmonotonicity, we let

$$P_{i(k)} = P(x_{i(k)}, \lambda_{i(k)}, \theta_{i(k)}) = \max_{0 \leq j \leq m(k)} \{ P(x_{k-j}, \lambda_{k-j}, \theta_{k-j}) \},$$

(2.3)

where $m(k) = \min \{ m(k-1) + 1, M, M_k \}$, $m(0) := 0$, $M \geq 0$ is an integer constant and $M_k \geq 0$ is an integer variable. $P_{i(k)}$ denotes the maximum among $P_{k-m(k)}, P_{k-m(k)+1}, \ldots, P_k$ and $i(k)$ denotes the
where the index corresponding to the maximum. Obviously, \( k - m(k) \leq i(k) \leq k \). Let
\[
\rho_{1,k}(\Delta) = \frac{P_{i(k)} - P(x_k + d_k(\Delta), \lambda_k + \delta \lambda_k, \theta_k)}{\sum_{j=i(k)}^{k} \text{pred}_j(\Delta)}, \tag{2.4}
\]
\[
\rho_{2,k}(\Delta) = \frac{P_k - P(x_k + d_k(\Delta), \lambda_k + \delta \lambda_k, \theta_k)}{\text{pred}_k(\Delta)}, \tag{2.5}
\]
\[
\rho_k(\Delta) = \max\{\rho_{1,k}(\Delta), \rho_{2,k}(\Delta)\}, \tag{2.6}
\]
where \( \rho_{2,k}(\Delta) \) is the usual ratio in the standard trust-region method and \( \rho_{1,k}(\Delta) \) is the measurement in the nonmonotone sense. If \( \rho_k(\Delta) \geq \eta_1 \in (0, 1) \), then the next iteration point is \( x_{k+1} = x_k + d_k(\Delta) \), otherwise, a new smaller trust-region radius \( \Delta \) is chosen, then resolve (2.1) and (2.2) until \( \rho_k(\Delta) \geq \eta_1 \). The \( k \)th iteration is completed. If \( M = 0 \), it is obvious that the algorithm is a monotone trust-region one in the usual sense. The nonmonotone strategy for penalty parameter is given in the following algorithm and the details are discussed in Section 3.

**Algorithm 2.1.** Step 0: Given \( x_0 \in \mathcal{X}, \lambda_0 \in \mathbb{R}^m \) is an estimate of the Lagrange multiplier at \( x_0 \), a symmetric matrix \( B_0 \in \mathbb{R}^{m \times n}, \Delta_{\min} > 0, \eta_1, \eta_2 \in (0, 1) \), an integer constant \( M \geq 0, m(0) := 0, M_0 := M \) and \( k := 0 \).

**Step 1:** \( \Delta^{(0)}_k := \Delta_{\min}, i := 0 \).

**Step 2:** Compute \( \delta \lambda_k \in \mathbb{R}^m \) such that \( \| \lambda_k + \delta \lambda_k \| \leq L_1 \). (In a practical implementation, \( \delta \lambda_k \) depends on \( B_k, x_k, \lambda_k, \Delta_k \)).

**Step 4:** Choose \( \theta_{k,i} \in [0, 1] \) such that
\[
\text{pred}_k(\Delta^{(i)}_k) \geq \frac{1}{4} (\| c_k \|^2 - \| c_k + A_k^T d_k(\Delta^{(i)}_k) \|^2). \tag{2.7}
\]

Especially, if \( c_k = 0 \), then set \( \theta_{k,i} \in [\theta_{k,0}, 1] \), where \( \theta_{k,0} \geq \min\{1, \theta_0, \ldots, \theta_{i-1}\} \).

**Step 5:** Calculate \( \rho_k(\Delta^{(i)}_k) \).

(i) If \( \rho_k(\Delta^{(i)}_k) < \eta_1 \), then \( \Delta^{(i+1)}_k := \eta_2 \Delta^{(i)}_k, M_k := M_k + 1, i := i + 1 \), go to Step 2;

(ii) If \( \rho_k(\Delta^{(i)}_k) \geq \eta_1 \), then
\[
d_k = d_k(\Delta^{(i)}_k), \quad v_k = v_k(\Delta^{(i)}_k), \quad \Delta_k = \Delta^{(i)}_k,
\]
\[
x_{k+1} = x_k + d_k, \quad M_{k+1} = M, \quad \theta_k = \theta_{k,i},
\]
\[
m(k + 1) = \min\{m(k) + 1, M, M_k\}.
\]

**Step 6:** Generate \( B_{k+1}, k := k + 1 \), and go to Step 1.

**Remark.** In Step 3, \( L_1 \) is a large constant, e.g., \( L_1 = 10^4 \). In the convergence analysis of Algorithm 2.1, \( L_1 \) is only a given positive constant. In the implementation of the algorithm, if some Lagrange multiplier estimate \( \lambda \) is such that \( \| \lambda \| > L_1 \), it can be projected on the box \( \| \lambda \| \leq L_1 / \sqrt{n} \). In other words, the estimates of the Lagrangian multipliers are arbitrary in the convergence analysis of Algorithm 2.1, where we only assume that it is bounded (although we use the Lagrangian multiplier
obtained when we solve subproblem (2.2)). In Step 4, the details about the choice of $\theta_{k,i}$ are given in Section 3. In Step 5, if $\rho_k(\Delta^{(i)}_k) < \eta_1$, then we decrease the trust-region radius, enlarge the measurement of nonmonotonicity, and go to step 2, whose process is called the inner cycle. Otherwise, it is called the outer cycle. $\Delta_k$ is the trust-region radius for which the trial step $d_k$ is accepted, whose iteration is called the successful one. $\Delta^{(i)}_k$ means the trust-region radius inside the $i$th inner cycle. The index $i$ represents the iteration number in the inner cycle. Moreover, at the beginning of each iteration, we always set $\Delta^{(0)}_k = \Delta_{\text{min}}$ in Step 1, which will avoid too small trust-region radii. The details can be found in [12].

3. Global convergence

We introduce the following assumptions for the global convergence analysis in this paper:

(AS1) $f(x)$ and $c_i(x)$ ($i = 1, 2, \ldots, m$) are twice continuously differentiable for all $x \in \mathcal{X}$.
(AS2) There exists a bounded convex closed set $\Omega$ such that $x_k$ are all in $\Omega$ for all $k$.
(AS3) The matrices $\{B_k\}$ are uniformly bounded.

At first, we have the following lemma:

**Lemma 3.1.** $x_k$ is a $\varphi$-stationary point if and only if $v_k(\Delta) = 0$ for all $\Delta > 0$.

**Proof.** Note that (2.1) is a strictly convex programming problem, which implies that $v_k(\Delta) = 0$ is equivalent to 0 being a Karush–Kuhn–Tucker (KKT) point of (2.1). The result is obtained by the definition of the $\varphi$-stationary point. □

**Theorem 3.2.** If $v_k(\Delta) = 0$ and 0 is a stationary point of (2.2), then $x_k$ is a substationary point of (1.1).

**Proof.** The conclusion is easily deduced from the definition of the substationary point and the fact that 0 is a stationary point of (2.2). □

**Lemma 3.3.** Algorithm 2.1 is well defined. That is, if the process does not terminate at $x_k$, then the inner cycle stops after a finite number of iterations.

**Proof.** Suppose that Algorithm 2.1 does not stop at $k$th iteration such that the inner cycle at the $k$th iteration is looped. It follows that

$$
\rho_k(\Delta^{(i)}_k) < \eta_1, \quad i = 0, 1, 2, \ldots,
$$

where $\Delta^{(i)}_k = \eta_1^i \Delta_{\text{min}}$, the super index $i$ denotes the $i$th inner iteration. By the Taylor’s expansion, we have that

$$
c(x_k + d_k(\Delta^{(i)}_k)) = c(x_k) + A(x_k)^Td_k(\Delta^{(i)}_k) + O(\|d_k(\Delta^{(i)}_k)\|^2),
$$

$$
f(x_k + d_k(\Delta^{(i)}_k)) = f(x_k) + \nabla f(x_k)^Td_k(\Delta^{(i)}_k) + \frac{1}{2} d_k(\Delta^{(i)}_k)^T\nabla^2 f(x_k + \zeta_k)d_k(\Delta^{(i)}_k)d_k(\Delta^{(i)}_k),
$$
\[ P(x_k + d_k(\Delta_k^{(i)}), \lambda_k + \delta \lambda_k, \theta_{k,i}) \]
\[ = \theta_{k,i} f(x_k + d_k(\Delta_k^{(i)}), \lambda_k + \delta \lambda_k) + (1 - \theta_{k,i})\phi(x_k + d_k(\Delta_k^{(i)})) \]
\[ = \theta_{k,i} [f(x_k + d_k(\Delta_k^{(i)})) + (\lambda_k + \delta \lambda_k)^T c(x_k + d_k(\Delta_k^{(i)})) + \frac{1}{2} (1 - \theta_{k,i}) \| c(x_k + d_k(\Delta_k^{(i)})) \|^2 \]
\[ = \theta_{k,i} f(x_k, \lambda_k) + \theta_{k,i} \| \nabla \phi(x_k, \lambda_k) \| d_k(\Delta_k^{(i)}) + \frac{1}{2} d_k(\Delta_k^{(i)})^T \nabla^2 f(x_k + \xi_{k,i} d_k(\Delta_k^{(i)})) d_k(\Delta_k^{(i)}) \]
\[ + (c(x_k) + A(x_k)^T d_k(\Delta_k^{(i)})^T \delta \lambda_k) + \| \lambda_k + \delta \lambda_k \| O(\| d_k(\Delta_k^{(i)}) \|^2) \]
\[ + \frac{1}{2} (1 - \theta_{k,i}) \| c(x_k) + A(x_k)^T d_k(\Delta_k^{(i)}) \|^2 + O(\| d_k(\Delta_k^{(i)}) \|^2), \]
where \( \xi_{k,i} \in (0, 1) \). By Step 3 of Algorithm 2.1, \( \| \lambda_k + \delta \lambda_k \| \leq L_1 \), we have that
\[ P(x_k, \lambda_k, \theta_{k,i}) - P(x_k + d_k(\Delta_k^{(i)}), \lambda_k + \delta \lambda_k, \theta_{k,i}) - \text{pred}_k(\Delta_k^{(i)}) \]
\[ = \frac{1}{2} \theta_{k,i} d_k(\Delta_k^{(i)})^T (B_k - \nabla^2 f(x_k + \xi_{k,i} d_k(\Delta_k^{(i)}))) d_k(\Delta_k^{(i)}) + O(\| d_k(\Delta_k^{(i)}) \|^2). \]

By (AS1)–(AS3), there exist positive constants \( a_1 \) and \( a_2 \) such that
\[ |P(x_k, \lambda_k, \theta_{k,i}) - P(x_k + d_k(\Delta_k^{(i)}), \lambda_k + \delta \lambda_k, \theta_{k,i}) - \text{pred}_k(\Delta_k^{(i)})| \leq (a_1 \theta_{k,i} + a_2) \| d_k(\Delta_k^{(i)}) \|^2. \]

(3.2)

Now we consider the following two cases:

1. \( \nu_k(\Delta_k^{(i)}) = 0 \)
2. \( \nu_k(\Delta_k^{(i)}) \neq 0 \).

For (1), we have that \( c_k = 0, A_k^T d_k(\Delta_k^{(i)}) = 0 \). So
\[ \text{pred}_k(\Delta_k^{(i)}) = \theta_{k,i} (-\nabla^T \phi_k d_k(\Delta_k^{(i)}) - \frac{1}{2} d_k(\Delta_k^{(i)})^T B_k d_k(\Delta_k^{(i)})) = -\theta_{k,i} \Psi_k(d_k(\Delta_k^{(i)})). \]

(3.3)

Since the algorithm does not stop, 0 is not a stationary point of (2.2), which implies that \( \Psi_k(d_k(\Delta_k^{(i)})) < 0 \). By Step 4 in Algorithm 2.1, \( \theta_{k,i} \geq \theta_{k,0} \) under case (1). Moreover, there exists \( s_k \neq 0 \) such that \( s_k \) is a feasible descent direction of (2.2), which implies that there exists a scalar \( \tilde{\delta}_k > 0 \) such that \( A_k^T s_k = A_k^T \nu_k(\Delta_k^{(i)}) = 0, l \leq x_k + s_k \leq u \) for any \( x \in [0, \tilde{\delta}_k] \) and \( \nabla^T \phi_k s_k = \nabla^T \Psi_k(0)^T s_k < 0 \).
Let \( s_k^{(i)} = (\Delta_k^{(i)}/\| s_k \|_\infty) s_k \). Then for \( \Delta_k^{(i)} > 0 \) small enough, \( s_k^{(i)} \) is a feasible point of (2.2). Hence, by the boundedness of \( \{B_k\} \),
\[ \Psi_k(d_k(\Delta_k^{(i)})) \leq \Psi_k(s_k^{(i)}) = \nabla^T \phi_k s_k^{(i)} + \frac{1}{2} s_k^{(i)^T} B_k s_k^{(i)} \]
\[ = \frac{\Delta_k^{(i)}}{\| s_k \|_\infty} \nabla^T \phi_k s_k + \frac{(\Delta_k^{(i)})^2}{2 \| s_k \|_\infty^2} s_k^{(i)^T} B_k s_k \]
\[ \leq \frac{\Delta_k^{(i)}}{2 \| s_k \|_\infty} \nabla^T \phi_k s_k \]

(3.4)
holds for all sufficiently large $i$. From (3.2)–(3.4), we have that

\[
\left| P(x_k, \mu_k, \theta_{k,i}) - P(x_k + d_k(\Delta_k^{(i)}), \mu_k + \delta \mu_k, \theta_{k,i}) \right|_{\text{pred}_k(\Delta_k^{(i)})} - 1 \\
\leq \frac{2(a_1 \theta_{k,i} + a_2)\|s_k\|_\infty \|d_k(\Delta_k^{(i)})\|^2}{\theta_{k,i} |\nabla f_k^T s_k| \Delta_k^{(i)}} \\
\leq \frac{2n(a_1 \theta_{k,i} + a_2)\|s_k\|_\infty (\Delta_k^{(i)})^2}{\theta_{k,0} |\nabla f_k^T s_k| \Delta_k^{(i)}} 
\to 0 \quad \text{if} \quad i \to \infty.
\]

(3.5)

The last inequality follows from $\|d_k(\Delta_k^{(i)})\| \leq \sqrt{n}\|d_k(\Delta_k^{(i)})\|_\infty$ and $\theta_{k,i} \geq \theta_{k,0}$. It yields a contradiction from (3.1) and (3.5).

For (2), we consider the following problem:

\[
\begin{align*}
\min & \quad (A_k c_k)^T y \\
\text{s.t.} & \quad l \leq x_k + y \leq u, \\
& \quad \|y\|_\infty \leq 1.
\end{align*}
\]

Let $y^*$ be a solution to the subproblem above. Since 0 is not a solution to (2.1), $y^*^T A_k c_k < 0$ and $y^*$ is a feasible descent direction of subproblem (2.1) at $v = 0$. Let $v = \alpha y^*$, $\alpha \geq 0$ and let $\bar{z}_k$ be the solution of $\min_{x \geq 0} \Phi_k(\alpha y^*)$, we then have that

\[
\bar{z}_k = -\frac{y^*^T A_k c_k}{\|y^*\|^2 + \|A_k^T y^*\|^2}.
\]

On the other hand, if $0 \leq \alpha \leq \min\{1, 0.8\Delta_k^{(i)}\} \text{ def } \alpha_k^{(i)}$, then

\[
l \leq x_k + \alpha y^* \leq u, \quad \alpha \|y\|_\infty \leq 0.8\Delta_k^{(i)}.
\]

If $\bar{z}_k = \alpha_k^{(i)}$, then

\[
\Phi_k(v_k(\Delta_k^{(i)})) \leq \Phi_k(\bar{z}_k y^*) = \frac{1}{2} \|c_k\|^2 - \frac{1}{2} \|y^*\|^2 + \|A_k^T y^*\|^2
\]

which implies that

\[
\|c_k\|^2 - \|c_k + A_k^T v_k(\Delta_k^{(i)})\|^2 \geq \frac{(c_k^T A_k^T y^*)^2}{\|y^*\|^2 + \|A_k^T y^*\|^2} + \|v_k(\Delta_k^{(i)})\|^2 \\
\geq \frac{(c_k^T A_k^T y^*)^2}{\|y^*\|^2 + \|A_k^T y^*\|^2} = -\bar{z}_k y^*^T A_k c_k.
\]

If $\bar{z}_k > \alpha_k^{(i)}$, then

\[
\Phi_k(v_k(\Delta_k^{(i)})) \leq \Phi_k(\alpha_k^{(i)} y^*) = \frac{1}{2} \|c_k\|^2 + \alpha_k^{(i)} c_k^T A_k^T y^* + \frac{(\alpha_k^{(i)})^2}{2} (\|y^*\|^2 + \|A_k^T y^*\|^2).
\]
Note that $\tilde{\alpha}_k > t_k^{(i)}$ implies that
\[ t_k^{(i)}(\|y^*\|^2 + \|A_k^Ty^*\|^2) < -c_k^TA_k^Ty^*. \]
By the two relations above,
\[ \|c_k\|^2 - \|c_k + A_k^Tv_k(\Delta_k^{(i)})\|^2 \geq -t_k^{(i)}y^TA_kc_k. \]
So, we have that
\[ \|c_k\|^2 - \|c_k + A_k^Tv_k(\Delta_k^{(i)})\|^2 \geq |y^TA_kc_k| \min\{\tilde{\alpha}_k, t_k^{(i)}\}. \]
For all sufficiently large $i$, $\min\{\tilde{\alpha}_k, t_k^{(i)}\} = 0.8\Delta_k^{(i)}$. By $A_k^Td_k(\Delta_k^{(i)}) = A_k^Tv_k(\Delta_k^{(i)})$ and (2.7),
\[ \pred_k(\Delta_k^{(i)}) \geq \frac{1}{4}(\|c_k\|^2 - \|c_k + A_k^Tv_k(\Delta_k^{(i)})\|^2) \geq a_3\Delta_k^{(i)}, \quad (3.6) \]
where $a_3 = 0.2\|y^TA_kc_k\|$. Similar to (3.5), we have, from (3.2) and (3.6), that
\[ |\rho_{2,k}(\Delta_k^{(i)}) - 1| \leq \frac{(a_1\theta_{k,i} + a_2)d_k(\Delta_k^{(i)})}{a_3\Delta_k^{(i)}} \leq \frac{n(a_1\theta_{k,i} + a_2)(\Delta_k^{(i)})^2}{a_3\Delta_k^{(i)}} \rightarrow 0 \quad \text{if} \quad i \rightarrow \infty, \]
which is also a contradiction with (3.1).
So the result is true. □

The parameter $\theta$ that satisfies (2.7) is chosen according to F.A.M. Gomes et al. (see [8]). Let us first define
\[ \theta_{0}^{\text{min}} = 1, \]
\[ \theta_{k}^{\text{min}} = \min\{1, \theta_0, \theta_1, \ldots, \theta_{k-1}\}, \quad k \geq 1, \]
\[ \theta_{k}^{\text{large}} = \left(1 + \frac{N^r}{(k + 1)^{1.1}}\right)\theta_{k}^{\text{min}}, \quad (3.7) \]
where $N^r \geq 0$ is a number that reflects the "degree of nonmonotonicity" desired for the penalty parameter. At the $i$th inner cycle of $x_k$, we define
\[ \theta_{k,i}^{\text{sup}} = \theta^{\text{sup}}(x_k, \Delta_k^{(i)}) = \sup\{\theta \in [0, 1]|\pred_k(\Delta_k^{(i)}) \geq 0.25(\|c_k\|^2 - \|c_k + A_k^Td(\Delta_k^{(i)})\|^2)\}. \quad (3.8) \]
Thus the value of $\theta_{k,i}$ that satisfies (2.7) is given by
\[ \theta_{k,i} = \theta(x_k, \Delta_k^{(i)}) = \min\{\theta_{k,i}^{\text{sup}}, \theta_{k,i}'\}, \quad (3.9) \]
where $\theta_{k,i}'$ is given by
\[ \theta_{k,i}' = \begin{cases} \theta_{k}^{\text{large}}, & \text{if} \quad i = 0, \\ \theta'(x_k, \Delta_k^{(i)}) \text{ def } \theta_{k,i-1} = \theta(x_k, \Delta_k^{(i-1)}), & \text{if} \quad i > 0. \end{cases} \quad (3.10) \]
Lemma 3.4. For the penalty parameter sequence \( \{ \theta_k \} \), we have that
\[
\theta_k - \theta_{k+1} + \frac{\theta_{k+N'}}{(k + 1)^{1.1}} \geq 0
\]
holds for all \( k \).

Proof. We assume that
\[
\rho_{k+1}(\Delta_{k+1}^{(i_{k+1})}) \geq \eta_1, \quad \rho_{k+1}(\Delta_{k+1}^{(i_{k+1}-1)}) < \eta_1,
\]
where \( i_{k+1} \) is a nonnegative integer. By the choice of the penalty parameter \( \theta \),
\[
\theta_{k+1} = \theta_{k+1,i_{k+1}} = \min\{ \theta_{k+1,i_{k+1}}, \theta'_{k+1,i_{k+1}} \}
\]
\[
\leq \theta'_{k+1,i_{k+1}} = \theta_{k+1,i_{k+1}-1} = \min\{ \theta'_{k+1,i_{k+1}-1}, \theta'_{k+1,i_{k+1}-2} \}
\]
\[
\leq \theta'_{k+1,i_{k+1}-1} = \theta_{k+1,i_{k+1}-2} \leq \cdots \leq \theta'_{k+1,0} = \theta_{k+1}^{large}.
\]
Moreover,
\[
\theta_{k+1}^{large} = \left( 1 + \frac{\mathcal{N}'}{(k + 1)^{1.1}} \right) \theta_{k+1}^{min}
\]
\[
\leq \left( 1 + \frac{\mathcal{N}'}{(k + 1)^{1.1}} \right) \theta_{k+1}^{min}
\]
\[
\leq \left( 1 + \frac{\mathcal{N}'}{(k + 1)^{1.1}} \right) \theta_{k},
\]
where the last inequality above is deduced by \( \theta_{k+1}^{min} = \min\{1, \theta_0, \theta_1, \ldots, \theta_k \} \). So the result holds. \( \square \)

The asymptotic stability of \( \theta_k \), which corresponds to the successful iteration, is given in the following lemma (also see [8, Lemma 6]).

Lemma 3.5 (Gomes et al. [8]). The penalty parameter sequence \( \{ \theta_k \} \) is convergent.

By (AS1), (AS2) and the boundedness of \( \lambda_k \), it follows that \( \ell(x_k, \lambda_k), \varphi(x_k) \) and \( P(x_k, \lambda_k, \theta_k) \) are bounded, which also implies that there exists a positive constant \( L_2 \) such that
\[
|\ell(x_k, \lambda_k) - \varphi(x_k)| \leq L_2 \quad \text{for all } k. \tag{3.11}
\]

Define the nonmonotone index set
\[
S^{NM} = \{ k | \rho_k(\Delta_k) = \rho_{1,k}(\Delta_k) \text{ and } \rho_{1,k}(\Delta_k) \neq \rho_{2,k}(\Delta_k) \}.
\]
In the discussion below, \( P_k, \text{pred}_k, v_k, d_k \) and \( \theta_k \) are all the values corresponding to the successful iterations. We have the following properties on \( S^{NM} \).
Lemma 3.6. Assume that $S^{NM}$ is an infinite set and \( \liminf_{k \to \infty} \text{pred}_k = a_4 > 0 \). Then there exists $k_1$ such that for any $j_u \in S^{NM}$, $j_u \geq \max\{2M + 3, M + k_1 + 1\}$, we have an index $t$: $j_u - 2 - 2M \leq t < j_u$ such that

$$P_{i(j_u)} \leq P_{t+1} - \frac{1}{4} \eta_1 a_4.$$ 

Proof. It follows from Lemma 3.5 that \( \lim_{k \to \infty} (\theta_k - \theta_{k-1}) = 0 \), which implies that, for $\eta_1 a_4/(2L_2) > 0$, there exists $k_1$ such that

$$|\theta_k - \theta_{k-1}| < \frac{\eta_1 a_4}{2L_2} \quad \forall k \geq k_1,$$ 

(3.12)

where $L_2$ is defined in (3.11). Suppose, by contradiction, that

$$P_{i(j_u)} > P_{j_u-j} - \frac{1}{4} \eta_1 a_4, \quad j = 0, 1, \ldots, 2M + 1$$

holds for some $j_u \in S^{NM}$ and $j_u \geq \max\{2M + 3, M + k_1 + 1\}$. By the definition of $P_{i(j_u)}$, we have that

$$P_{i(j_u)} = \max\{P_{j_u-m(j_u)}, \ldots, P_{j_u-1}, P_{j_u}\} = P_{j_u-j_0}$$

holds for some $j_0: 0 \leq j_0 \leq m(j_u)$. It follows from $j_u \in S^{NM}$ and the definition of $S^{NM}$ that $j_0 > 0$.

If $j_u - j_0 - 1 \in S^{NM}$, which implies that $\rho_{j_u-j_0-1} = \rho_1, j_u-j_0-1$ and $i(j_u-j_0-1) < j_u-j_0-1$, then, since we consider the successful iterations,

$$P_{i(j_u-j_0-1)} - P(x_{j_u-j_0}, \theta_{j_u-j_0}, \theta_{j_u-j_0-1}) \geq \eta_1 \sum_{l=(j_u-j_0-1)}^{j_u-j_0-1} \text{pred}_l.$$ 

(3.13)

Note that

$$P(x_{j_u-j_0}, \theta_{j_u-j_0}, \theta_{j_u-j_0-1}) = P_{j_u-j_0} + (\theta_{j_u-j_0-1} - \theta_{j_u-j_0})(\ell_{j_u-j_0} - \phi_{j_u-j_0}).$$ 

(3.14)

So,

$$P_{i(j_u-j_0-1)} > P(x_{j_u-j_0}, \theta_{j_u-j_0}, \theta_{j_u-j_0-1}) + \eta_1 \text{pred}_{j_u-j_0-1}$$

(by (3.13) and $i(j_u - j_0 - 1) < j_u - j_0 - 1$)

$$= P_{j_u-j_0} + (\theta_{j_u-j_0-1} - \theta_{j_u-j_0})(\ell_{j_u-j_0} - \phi_{j_u-j_0}) + \eta_1 \text{pred}_{j_u-j_0-1}$$

(by (3.14))

$$> P_{j_u-j_0} - L_2|\theta_{j_u-j_0-1} - \theta_{j_u-j_0}| + \eta_1 a_4$$

(by (3.11) and $\text{pred}_l \geq a_4, \forall l$)

$$> P_{j_u-j_0} + \frac{1}{2} \eta_1 a_4$$

(by (3.12))

$$= P_{i(j_u)} + \frac{1}{2} \eta_1 a_4$$

(by the definition of $i(j_u)$)

$$> P_{j_u-j} + \frac{1}{4} \eta_1 a_4, \quad j = 0, 1, \ldots, 2M + 1$$

(by the supposition).

(3.15)

On the other hand,

$$P_{i(j_u-j_0-1)} = \max\{P_{j_u-j_0-1-m(j_u-j_0-1)}, \ldots, P_{j_u-j_0-2}, P_{j_u-j_0-1}\} = P_{j_u-j_0-1-j_1}$$

holds for some $j_1: 0 < j_1 \leq m(j_u - j_0 - 1)$ and $j_0 + j_1 + 1 \leq 2M + 1$, which yields a contradiction from (3.15).
If \( j_u - j_0 - 1 \not\in S^{NM} \), which implies that \( \rho_{j_u - j_0 - 1} = \rho_{j_u - j_0 - 1} \), then

\[
P_{j_u - j_0 - 1} - P(x_{j_u - j_0}, \lambda_{j_u - j_0}, \theta_{j_u - j_0 - 1}) \geq \eta_1 \text{ pred}_{j_u - j_0 - 1} \geq \eta_1 a_4. \tag{3.16}
\]

So,

\[
P_{j_u - j_0 - 1} \geq P_{j_u - j_0} + (\theta_{j_u - j_0 - 1} - \theta_{j_u - j_0})(\ell_{j_u - j_0} - \varphi_{j_u - j_0}) + \eta_1 a_4 \quad \text{(by (3.14))}
\]

\[
> P_{j_u - j_0} + \frac{1}{2} \eta_1 a_4 \quad \text{(by (3.11) and (3.12))}
\]

\[
= P_{i(j_u)} + \frac{1}{2} \eta_1 a_4 \quad \text{(by the definition of } i(j_u))
\]

\[
> P_{j_u - j} + \frac{1}{4} \eta_1 a_4, \quad j = 0, 1, \ldots, 2M + 1 \quad \text{(by the supposition)},
\]

which also yields a contradiction. Thus, the result is proved. \( \square \)

Let \( S^{NM} = \{j_1, j_2, \ldots, j_u, \ldots\} \), \( \hat{S} = \{j_v \in S^{NM} | j_v - j_0 \geq M + 2\} \), then we have the following lemma.

**Lemma 3.7.** Assume that the hypotheses of Lemma 3.6 hold, then \( \hat{S} \) is a finite set.

**Proof.** We assume, by contradiction, that \( \hat{S} \) is an infinite index subset. Let \( \hat{S} = \{j_{v_1}, j_{v_2}, \ldots, j_{v_t}, \ldots\} \) and \( j_{v_{t+1}} - j_{v_t} \geq M + 2, \ t = 1, 2, \ldots \).

At first, We prove that

\[
P_{j_{v_1} + 2} > P_{j_{v_2} + 1}. \tag{3.17}
\]

By the definition of \( \hat{S} \), \( j_{v_2} \in S^{NM} \), so there exists \( u_1: 0 < u_1 \leq m(j_{v_2}) \) such that

\[
P_{i(j_{v_2})} = \max_{0 \leq u \leq m(j_{v_2})} \{P_{j_{v_2} - u}\} = P_{j_{v_2} - u_1}
\]

\[
\geq P(x_{j_{v_2} + 1}, \lambda_{j_{v_2} + 1}, \theta_{j_{v_2}}) + \eta_1 \sum_{t = i(j_{v_2})}^{j_{v_2}} \text{ pred}_t \quad \text{(the successful iteration)}
\]

\[
> P_{j_{v_2} + 1} + (\theta_{j_{v_2}} - \theta_{j_{v_2} + 1})(\ell_{j_{v_2} + 1} - \varphi_{j_{v_2} + 1}) + \eta_1 a_4 \quad \text{(by (3.14) and } u_1 > 0)
\]

\[
\geq P_{j_{v_2} + 1} + \frac{1}{2} \eta_1 a_4 > P_{j_{v_2} + 1} \quad \text{(by (3.11) and (3.12))}. \tag{3.18}
\]

We now consider two cases.

**Case 1:** \( v_1 + 1 = v_2 \).

We have, by the definition of \( S^{NM} \), that

\[
k \not\in S^{NM}, \quad \text{for all } k: j_{v_1} + 1 \leq k \leq j_{v_1 + 1} - 1,
\]

which implies that

\[
P_k > P_{k+1} \quad \text{for all } k: j_{v_1} + 1 \leq k \leq j_{v_1 + 1} - 1. \tag{3.19}
\]
Moreover, it follows from \(0 < u_1 \leq M\) and \(j_{v_1+1} - j_{v_1} = j_{v_2} - j_{v_1} \geq M + 2\) that
\[
j_{v_2} - u_1 = j_{v_1+1} - u_1 \leq j_{v_1+1} - 1,
\]
\[
j_{v_2} - u_1 \geq j_{v_2} - M = j_{v_1} + (j_{v_2} - j_{v_1}) - M \geq j_{v_1} + 2.
\]
That is, \(j_{v_1} + 1 < j_{v_1} + 2 \leq j_{v_2} - u_1 \leq j_{v_1+1} - 1\), by (3.19),
\[
P_{j_{v_1}+2} > P_{j_{v_2} - u_1} \geq P_{j_{v_2}+1}.
\]
The equality in the first relation above holds only when \(j_{v_2} - u_1 = j_{v_1} + 2\). So (3.17) holds.

**Case 2:** \(v_1 + 1 < v_2\).

It follows from \(v_1 + 1 < v_2\) that \(j_{v_2} \geq j_{v_1+2} \geq j_{v_1+1} + 1\). So, by \(0 < u_1 \leq M\) and \(j_{v_1+1} - j_{v_1} \geq M+2\),
\[
j_{v_2} - u_1 \geq j_{v_1+1} + 1 - u_1 = j_{v_1} + 1 + (j_{v_1+1} - j_{v_1}) - u_1 \geq j_{v_1} + 3.
\] (3.20)

**Case 2.1:** \(j_{v_2} - u_1 \leq j_{v_1+1}\).

It follows, by (3.19) and (3.20), that (3.17) also holds.

**Case 2.2:** \(j_{v_2} - u_1 = j_{v_1+1} + 1\).

Since \(j_{v_2} - u_1 - 1 = j_{v_1+1} \in S^{NM}\), there exists \(u': 0 < u' \leq m(j_{v_2} - u_1 - 1)\) such that
\[
P_{j_{v_1+1} - u_1} = P_{j_{v_2} - u_1 - u'} > P_{j_{v_2} - u_1} \geq P_{j_{v_2}+1}.
\] (3.21)

Note that \(j_{v_2} - u_1 - 1 - u' = j_{v_1+1} - u' < j_{v_1+1}\) and
\[
j_{v_2} - u_1 - 1 - u' = j_{v_1+1} - u'
\]
\[
= j_{v_1} + (j_{v_1+1} - j_{v_1}) - u'
\]
\[
\geq j_{v_1} + 2 \quad \text{(by } j_{v_1} \in \hat{S} \text{ and } u' \leq M).\]

It follows from (3.19) that \(P_{j_{v_1}+2} > P_{j_{v_2} - u_1 - u'}\). Combining (3.21), we get (3.17).

**Case 2.3:** \(j_{v_2} - u_1 > j_{v_1+1} + 1\).

If \(j_{v_2} - u_1 - 1 \not\in S^{NM}\), we have, from \(\rho_{2, j_{v_2} - u_1 - 1} \geq \eta_1\) and (3.18), that
\[
P_{j_{v_2} - u_1 - 1} > P_{j_{v_2} - u_1} \geq P_{j_{v_2}+1}.
\]

If \(j_{v_2} - u_1 - 1 \in S^{NM}\), then there exists \(u'_2: 0 < u'_2 \leq m(j_{v_2} - u_1 - 1)\) such that
\[
P_{j_{v_1+1} - u_1} = P_{j_{v_2} - u_1 - u'_2} > P_{j_{v_2} - u_1} \geq P_{j_{v_2}+1}.
\]

Let
\[
u_2 = \begin{cases} 
1 & \text{if } j_{v_2} - u_1 - 1 \not\in S^{NM}, \\
1 + u'_2 & \text{if } j_{v_2} - u_1 - 1 \in S^{NM}.
\end{cases}
\]

So, under case 2.3, there exists \(u_2: 1 \leq u_2 \leq 1 + M\) (by \(u'_2 \leq M\)) such that
\[
P_{j_{v_2} - u_1 - u_2} > P_{j_{v_2}+1}\]
\[ j_{v_2} - u_1 - u_2 \geq j_{v_1+1} + 2 - u_2 = j_{v_1} + (j_{v_1+1} - j_{v_1}) + 2 - u_2 \geq j_{v_1} + 3. \]

If \( j_{v_2} - u_1 - u_2 \leq j_{v_1+1} + 1 \), then, we have, similarly to cases 2.1 and 2.2, that (3.17) holds.

If \( j_{v_2} - u_1 - u_2 > j_{v_1+1} + 1 \), then, we can find an integer number \( u_3 \) such that
\[ P_{j_{v_2} - u_1 - u_2 - u_3} > P_{j_{v_2}+1} \quad \text{and} \quad j_{v_2} - u_1 - u_2 - u_3 \geq j_{v_1} + 3. \]

Repeating the process above, since \( v_1 \) and \( v_2 \) are two given indices, there exists \( u_l \):
\[ 1 \leq u_l \leq 1 + M \]
and
\[ j_{v_1} + 3 \leq j_{v_2} - u_1 - u_2 - \cdots - u_l \leq j_{v_1+1} + 1. \]

So, (3.17) holds.

Next, for any \( j_{v_t} \in \hat{S} \), we have, similar to the proof of (3.17), that
\[ P_{j_{v_t+2}} > P_{j_{v_{t+1}+1}}, \quad t = 1, 2, \ldots. \]  \hspace{1cm} (3.22)

Finally, we have that
\[ P_{j_{v_t+1} - P_{j_{v_t+2}}} \geq \frac{1}{2} \eta_1 a_4, \quad t = 1, 2, \ldots. \]  \hspace{1cm} (3.23)

In fact, it follows, from \( j_{v_t} \in \hat{S} \) and \( j_{v_{t+1} - j_{v_t}} \geq M + 2 \), that \( j_{v_t} + 1 \notin S^{NM} \). So \( \rho_{j_{v_t+1}} = \rho_{j_{v_{t+1}+1}} \geq \eta_1 \), which implies that
\[ P_{j_{v_t+1} - P_{j_{v_{t+1}+1}}} \geq P(x_{j_{v_t+2}}, \lambda_{j_{v_t+2}}, \theta_{j_{v_{t+1}+1}}) + \eta_1 \text{ pred}_{j_{v_{t+1}+1}} \]
\[ = \frac{1}{2} \rho_{j_{v_{t+1}+1}} - \rho_{j_{v_{t+2}+1}}(\lambda_{j_{v_{t+2}+1}} - \phi_{j_{v_{t+2}+1}}) + \eta_1 \text{ pred}_{j_{v_{t+2}+1}} \]
\[ \geq P_{j_{v_{t+2}}} + \frac{1}{2} \eta_1 a_4 \quad \text{(by (3.11) and (3.12)).} \]

Adding those inequalities in (3.23) from \( t = 1 \) to \( q \), where \( q \) is an arbitrary positive integer, and using the relation \( P_{j_{v_{t+2}}} > P_{j_{v_{t+1}+1}}, \quad t = 1, 2, \ldots \), we have that
\[ P_{j_{v_t+1} - P_{j_{v_{t+2}}} + q \geq \frac{1}{2} \eta_1 a_4 \]
which, by the arbitrariness of \( q \), (AS1) and (AS2), is a contradiction with the boundedness of \( \{P_k\} \).
Thus the lemma is proved. \( \square \)

In the sequel, it is assumed that the distance between any two successive indices in \( S^{NM} \) is not greater than \( M + 2 \).

**Lemma 3.8.** Assume that the hypotheses of Lemma 3.6 hold and, without loss of generality, assume that the smallest index of \( S^{NM} \) is not less than \( k_1 \) (defined in the proof of Lemma 3.6). If \( \rho_t = \rho_{2,t} \) at \( x_t \) and \( \hat{S} = \{ j \in S^{NM} | j < t \} \neq \emptyset \), then there exists \( j_v \in S^{NM} \) such that \( P_{j_v+1} > P_{t+1} \) and
\[ 0 < t - j_v \leq M + 1. \]
Proof. Suppose that \( j_v \) is the largest index in \( \tilde{S} \). By the definition of the index \( j_v \) and the set \( \tilde{S} \), for \( j: j_v + 1 \leq j \leq t \), \( j \notin S^{NM} \). So it follows from the definition of the successful iteration that

\[
P_{j_v+1} > P(x_{j_v+2}, \lambda_{j_v+2}, \theta_{j_v+1}) + \eta_1 a_4
\]

\[
= P_{j_v+2} + (\theta_{j_v+2} - \theta_{j_v+1})(\ell_{j_v+2} - \phi_{j_v+2}) + \eta_1 a_4 \quad \text{(by (3.14))}
\]

\[
> P_{j_v+2} + \frac{1}{2} \eta_1 a_4 \quad \text{(by (3.11) and (3.12))}
\]

\[
> P_{j_v+2} > \cdots > P_t > P_{t+1}.
\]

Let \( j_v, j_{v+1} \) be two successive indices in \( S^{NM} \). Then we have \( 0 < t - j_v < j_{v+1} - j_v \leq M + 1 \). So the result is proved. \( \square \)

Lemma 3.9. Assume that the hypotheses of Lemma 3.6 hold, let \( S^{NM} = \{j_1, j_2, \ldots, \} \), and assume, without loss of generality, that \( j_1 \geq \max\{2M+3, M+k_1 + 1\} \), where \( k_1 \) is defined in the proof of Lemma 3.6. If \( u \geq 3M + 3 + j_2 \), then there exists \( j_v \in S^{NM} \) such that \( P_{i(j_u)} < P_{j_v+1} \) and \( 0 < j_u - j_v \leq 4 + 4M \).

Proof. Since \( j_1 \geq \max\{2M+3, M+k_1 + 1\} \), \( j_u > u \). For \( j_u \in S^{NM} \), it follows from \( u \geq 3M + 3 + j_2 > j_1 \) that \( j_u > \max\{2M + 3, M + k_1 + 1\} \). By Lemma 3.6, there exists \( t \) satisfying \( j_u - 2 - 2M \leq t < j_u \) such that

\[
P_{i(j_u)} \leq P_{t+1} - 0.25 \eta_1 a_4.
\]

(3.24)

Consider the following two cases.

Case 1: \( t \in S^{NM} \).

We have, by (2.3), that

\[
P_{i(t)} = \max_{0 \leq l \leq m(t)} \{P_{i-1}\} = P_{t-l_0}, \quad i(t) = t - l_0, \quad 0 < l_0 \leq m(t) \leq M.
\]

(3.25)

Moreover, \( t \in S^{NM} \) implies that \( \rho_t = \rho_{1,t} \). So, by the definition of the successful iteration,

\[
P_{i(t)} \geq P(x_{t+1}, \lambda_{t+1}, \theta_t) + \eta_1 \sum_{l=i(t)}^{t} \text{pred}_l
\]

\[
> P_{t+1} + (\theta_t - \theta_{t+1})(\ell_{t+1} - \phi_{t+1}) + \eta_1 a_4 \quad \text{(by (3.14) and } l_0 > 0)
\]

\[
> P_{t+1} + 0.5 \eta_1 a_4 \quad \text{(by (3.11) and (3.12))}.
\]

(3.26)

Note that

\[
i(t) - 1 = t - l_0 - 1 \quad \text{(by (3.25))}
\]

\[
\geq t - M - 1 \quad \text{(by } l_0 \leq M)
\]

\[
\geq j_u - 3M - 3 \quad \text{(by } t \geq j_u - 2 - 2M)
\]

\[
> u - 3M - 3 \quad \text{(by } j_u > u)
\]

\[
\geq j_2,
\]

(3.27)

that is, \( j_2 < i(t) - 1 \).
Case 1.1: \( j_2 < i(t) - 1 \), where \( i(t) - 1 \in S^{NM} \).

We take \( j_v = i(t) - 1 \), which yields

\[
P_{j_v+1} = P_{i(t)} > P_{t+1} > P_{i(j_v)} \quad \text{(by (3.26) and (3.24))}
\]

and

\[
0 < j_u - j_v = j_u - i(t) + 1
= (j_u - t) + (t - i(t)) + 1
\leq (2 + 2M) + l_0 + 1 \quad \text{(by \( j_u - 2 - 2M \leq t < j_u \) and (3.25))}
\leq 3 + 3M \quad \text{(by (3.25)).}
\]

It deduces the lemma.

Case 1.2: \( j_2 < i(t) - 1 \), where \( i(t) - 1 \notin S^{NM} \).

\( j_2 \in S^{NM} \) implies that \( \{i \in S^{NM} | i < i(t) - 1\} \neq \emptyset \). Moreover, \( i(t) - 1 \notin S^{NM} \) implies that \( \rho_{i(t) - 1} = \rho_{2, i(t) - 1} \). By Lemma 3.8, there exists \( j_v \in S^{NM} \) such that \( P_{j_v+1} > P_{i(t)} \) and \( 0 < i(t) - 1 - j_v \leq M + 1 \). So,

\[
j_u - j_v \leq (2 + 2M + t) - j_v \quad \text{(by \( j_u - 2 - 2M \leq t < j_u \))}
= 2 + 2M + (t - i(t)) + (i(t) - j_v)
\leq 2 + 2M + l_0 + (M + 2) \leq 4 + 4M,
\]

\[
 j_u - j_v > t - j_v = (t - i(t)) + (i(t) - j_v) > l_0 + 1 > 0
\]

and

\[
P_{j_v+1} > P_{i(t)} > P_{t+1} + 0.5\eta_1 a_4 \quad \text{(by (3.26))}
> (P_{i(j_v)} + 0.25\eta_1 a_4) + 0.5\eta_1 a_4 \quad \text{(by (3.24))}
> P_{i(j_v)},
\]

which shows the lemma.

Case 2: \( t \notin S^{NM} \).

Since \( j_u - 2 - 2M \leq t < j_u \) and \( u \geq 3M + 3 + j_2 \),

\[
t \geq j_u - 2M - 2 > u - 2M - 2 > M + 1 + j_2 > j_2,
\]

which implies that \( j_2 \in \{i \in S^{NM} | i < t\} \neq \emptyset \). Since \( t \notin S^{NM} \), \( \rho_t = \rho_{2,t} \). By Lemma 3.8, there exists \( j_v \in S^{NM} \) such that \( P_{j_v+1} > P_{t+1} \) and \( 0 < t - j_v \leq M + 1 \). It follows from (3.24) that \( P_{j_v+1} > P_{i(j_v)} \).

Moreover,

\[
j_u - j_v > t - j_v > 0 \quad \text{(by \( j_u - 2 - 2M \leq t < j_u \))},
\]

\[
j_u - j_v \leq (2 + 2M + t) - j_v \quad \text{(by \( j_u - 2 - 2M \leq t < j_u \))}
= (2 + 2M) + (t - j_v) \leq 3 + 3M \quad \text{(by \( t - j_v \leq M + 1 \)).}
\]

So the result is true. \( \square \)
Lemma 3.10. The following two conditions are inconsistent:

(i) \( S_{NM} \) is an infinite set;
(ii) \( \liminf_{k \to \infty} \text{pred}_k = a_4 > 0. \)

Proof. Suppose, by contradiction, that (i) and (ii) all hold. Let \( S_{NM} = \{j_1, j_2, \ldots\} \) and, without loss of generality, \( j_1 \geq \max\{2M + 3, M + k_1 + 1\} \), where the definition of \( k_1 \) is given by (3.12). By (AS1) and (AS2), we assume that \(|P(x_k)| \leq P_{\max}, \forall k\). We take

\[
u = \max \left\{ 2, \left\lceil \frac{4P_{\max}}{\eta_1 a_4} \right\rceil + 1 \right\} (3M + 3 + j_2),
\] (3.28)

where \( \lceil x \rceil \) means the largest integer not greater than \( x \).

By (3.28), \( \nu > 3M + 3 + j_2 \). For \( j_u \in S_{NM} \), it follows from Lemma 3.9 that there exists \( j_{u_1} \in S_{NM} \) such that \( P_{i(j_u)} < P_{j_{u_1} + 1} \) and \( 0 < j_u - j_{u_1} \leq 4 + 4M \).

If \( u - u_1 \geq 5 + 4M \), then \( j_u \geq j_{u_1} + 5 + 4M \geq j_{u_1} + 5 + 4M \), which is a contradiction with \( j_u - j_{u_1} \leq 4 + 4M \). So \( u - u_1 \leq 4 + 4M \), which implies that

\[
u - (4 + 4M) \geq 2(3M + 3 + j_2) - (4 + 4M) \quad \text{(by (3.28))}
= j_2 - (M + 1) + (3M + 3 + j_2)
> j_1 - (M + 1) + (3M + 3 + j_2)
> 3M + 3 + j_2 \quad \text{(the assumption of \( j_1 \)).}
\]

By Lemma 3.9, there exists \( j_{u_2} \in S_{NM} \) such that \( P_{i(j_{uj})} < P_{j_{u_2} + 1} \) and \( 0 < j_{uj} - j_{u_2} \leq 4 + 4M \).

Similarly, \( u_1 - u_2 \leq 4 + 4M \). So, \( u_2 \geq u_1 - (4 + 4M) \geq u - 2(4 + 4M) \). By (3.28), if \( \frac{4P_{\max}}{\eta_1 a_4} \geq 2 \), then \( u \geq 3(3M + 3 + j_2) > (3M + 3 + j_2) + 2(4 + 4M) \), that is, \( u_2 > 3M + 3 + j_2 \). It follows from Lemma 3.9 that there exists \( j_{uj} \in S_{NM} \) such that \( P_{i(j_{uj})} < P_{j_{uj} + 1} \) and \( 0 < j_{uj} - j_{uj-1} \leq 4 + 4M \).

We repeat the process above, until the index \( q = \max \{ 2, \lceil 4P_{\max}/(\eta_1 a_4) \rceil + 1 \} \), where \( P_{i(j_{uj-1})} < P_{j_{uj} + 1} \) and \( 0 < j_{uj-1} - j_{uj} \leq 4 + 4M. \)

On the other hand, since \( j_u \in S_{NM} \) and \( \rho_{j_u} \geq \eta_1 \), we have that

\[
P_{i(j_u)} - P(x_{j_u} + d_{j_u}, \lambda_{j_u} + \delta \lambda_{j_u}, \theta_{j_u}) \geq \eta_1 \sum_{l = (j_u)}^{j_u} \text{pred}_l.
\]

By (3.14),

\[
P(x_{j_u} + d_{j_u}, \lambda_{j_u} + \delta \lambda_{j_u}, \theta_{j_u}) = P_{j_{u+1}} + (\theta_{j_u} - \theta_{j_{u+1}})(\ell_{j_{u+1}} - \varphi_{j_{u+1}}).
\]

By (3.11) and (3.12), we have that

\[
P_{i(j_u)} - P_{j_{u+1}} > (\theta_{j_u} - \theta_{j_{u+1}})(\ell_{j_{u+1}} - \varphi_{j_{u+1}}) + \eta_1 a_4 \geq 0.5\eta_1 a_4.
\]
Similarly, we have that
\[ P_{i(j_q)} - P_{j_q+1} > 0.5\eta_1a_4, \]
\[ P_{i(j_{q+1})} - P_{j_{q+1}+1} > 0.5\eta_1a_4, \]
\[ \vdots \]
\[ P_{i(j_q)} - P_{j_q+1} > 0.5\eta_1a_4. \]
Adding all inequalities above, we obtain that
\[ 2P_{\text{max}} \geq P_{i(j_q)} - P_{j_q+1} > \frac{q + 1}{2} \eta_1a_4, \]
which yields that
\[ q \leq \frac{4P_{\text{max}}}{\eta_1a_4} - 1, \]
which is a contradiction with the definition of \( q \). So the Lemma is proved. \( \Box \)

**Theorem 3.11.** Any limit point of \( \{x_k\} \) is a \( \varphi \)-stationary point of (1.1).

**Proof.** Suppose, by contradiction, that there exists one: \( \{x_k\}_K \to x^* \), where \( x^* \) is not a \( \varphi \)-stationary point of (1.1). Let \( I(x^*) = \{i | x^*_i = l_i \text{ or } x^*_i = u_i\} \). Since \( x^* \) is not a \( \varphi \)-stationary point of (1.1), we have that

1. there exists at least an index \( i_0 \notin I(x^*) \) such that \( (A(x^*)c(x^*))_{i_0} \neq 0 \); or
2. for all \( i \notin I(x^*) \), \( (A(x^*)c(x^*))_i = 0 \) but there exists at least an index \( i_0 \in I(x^*) \) such that \( (A(x^*)c(x^*))_{i_0} < 0 \) if \( x^*_i = l_i \) or \( (A(x^*)c(x^*))_{i_0} > 0 \) if \( x^*_i = u_i \).

At first, we show that
\[ \liminf_{k \in K} \Delta_k = \bar{\Delta} > 0. \quad (3.29) \]
Otherwise, there exists an infinite subset \( K_1 \subseteq K \) such that \( \lim_{k \in K_1} \Delta_k = 0 \).

For case (1), we take
\[ (s_k)_i = \begin{cases} -(A(x_k)c(x_k))_{i_0} & \text{if } i = i_0, \\ 0 & \text{if } i \neq i_0. \end{cases} \quad (3.30) \]
By (AS1) and \( \lim_{k \in K_1} x_k = x^* \), there exists \( k_2 \) such that, for \( k \in K_1, k \geq k_2 \), we have that
\[ x^*_i - \frac{1}{2} \min\{x^*_i - l_i, u_i - x^*_i\} \leq (x_k)_{i_0} \leq x^*_i + \frac{1}{2} \min\{x^*_i - l_i, u_i - x^*_i\}, \]
\[ \frac{1}{2} \|c(x^*)\| \leq \|c(x_k)\| \leq \frac{3}{2} \|c(x^*)\|, \]
\[ \frac{1}{2} |(A(x^*)c(x^*))_{i_0}| \leq |(A(x_k)c(x_k))_{i_0}| \leq \frac{3}{2} |(A(x^*)c(x^*))_{i_0}|. \]
Let \( a_5 = \frac{1}{3 \|(A(x^*)c(x^*))_0\|} \min\{x^*_0 - l_0, u_0 - x^*_0\} \).

Then, for all \( k \in K_1, k \geq k_2 \), we have that
\[
l \leq x_k + \varepsilon_k \leq u \quad \forall \varepsilon \in [0, a_5].
\]

Let
\[
\varepsilon_k = \frac{1.6 \Delta_k}{3 \|(A(x^*)c(x^*))_0\|}.
\]

Since \( \lim_{k \in K_1} \Delta_k = 0 \), we can assume, without loss of generality, that \( k_2 \) is large enough such that \( \varepsilon_k < a_5 \) for all \( k \geq k_2 \). Therefore, \( \varepsilon_k s_k \) is a feasible solution of (2.1) for all \( k \in K_1, k \geq k_2 \), which implies that
\[
\Phi_k(v_k) \leq \Phi_k(\varepsilon_k s_k) = \frac{1}{2} \|c_k + \varepsilon_k A_k^T s_k\|^2 + \frac{1}{2} \|\varepsilon_k s_k\|^2
\]
\[
\leq \frac{1}{2} \|c_k\|^2 + \frac{1}{2} \|\varepsilon_k^2 A_k^T s_k\|^2 + \frac{\Delta_k}{3 \|(A(x^*)c(x^*))_0\|} \left( 1 + \sum_{i=1}^m \left( \frac{\hat{c}_i(x_k)}{\hat{x}_i_0} \right)^2 \right)
\]
holds for all \( k \in K_1, k \geq k_2 \). By \( \lim_{k \in K_1} \Delta_k = 0 \) and \( \lim_{k \in K_1} x_k = x^* \), there exists \( k_3 \geq k_2 \) such that
\[
\frac{0.8 \Delta_k}{3 \|(A(x^*)c(x^*))_0\|} \left( 1 + \sum_{i=1}^m \left( \frac{\hat{c}_i(x_k)}{\hat{x}_i_0} \right)^2 \right) < \frac{1}{2} \quad \forall k \in K_1, k \geq k_3.
\]

So,
\[
\Phi_k(v_k) \leq \frac{1}{2} \|c_k\|^2 - \frac{0.8 \Delta_k}{3 \|(A(x^*)c(x^*))_0\|} ((A_k c_k)_0)^2
\]
\[
\leq \frac{1}{2} \|c_k\|^2 - \frac{0.8 \Delta_k}{3 \|(A(x^*)c(x^*))_0\|} \frac{1}{4} ((A(x^*)c(x^*))_0)^2
\]
\[
= \frac{1}{2} \|c_k\|^2 - a_6 \Delta_k \quad \forall k \in K_1, k \geq k_3,
\]
where \( a_6 = 0.2 \|(A(x^*)c(x^*))_0\|/3 \). By (3.31),
\[
\|c_k\|^2 - \|c_k + A_k^T v_k\|^2 \geq 2a_6 \Delta_k \quad k \in K_1, k \geq k_3.
\]

From (2.7) and \( A_k^T v_k = A_k^T d_k \),
\[
\text{pred}_k(\Delta_k) \geq \frac{1}{4} (\|c_k\|^2 - \|c_k + A_k^T d_k\|^2) \geq \frac{1}{2} a_6 \Delta_k, \quad k \in K_1, k \geq k_3.
\]
By (3.2) and (3.32), we have that there exists an index $i_k$ such that
\[
\lim_{k \to \infty} \| \times_k - x_k^* \| = 0.
\]
Otherwise, we assume that there exists an infinite set $K_2 \subseteq K$ such that $\{\times_k\}_{k \in K_2}$ converges to $x_k^*$. Therefore, (3.29) is true.

Next, we prove that
\[
\lim \inf_{k \in K} \times_k > 0.
\]
Otherwise, we assume that there exists an infinite set $K_2 \subseteq K$ such that $\{\times_k\}_{k \in K_2}$ converges to $x_k^*$. Therefore, (3.29) is true.
For case (1), since \( i_0 \not\in I(x^*) \), \( l_{i_0} < x^*_{i_0} < u_{i_0} \). Taking \( s_k \) as (3.30). By (AS1) and \( \lim_{k \to \infty} x_k = x^* \), there exists \( k_2 \) such that, for \( k \in K_2, k \geq k_2 \), we have that
\[
x^*_{i_0} - \frac{1}{2} \min \{ x^*_{i_0} - l_{i_0}, u_{i_0} - x^*_{i_0} \} \leq (x_k)_{i_0} \leq x^*_{i_0} + \frac{1}{2} \min \{ x^*_{i_0} - l_{i_0}, u_{i_0} - x^*_{i_0} \},
\]
\[
\frac{1}{2} |(A(x^*)c(x^*))_{i_0}| \leq |(A(x_k)c(x_k))_{i_0}| \leq \frac{3}{2} |(A(x^*)c(x^*))_{i_0}|,
\]
\[
\frac{1}{2} \sum_{i=1}^{m} \left( \frac{\partial c_i(x^*)}{\partial x_{i_0}} \right)^2 \leq \sum_{i=1}^{m} \left( \frac{\partial c_i(x_k)}{\partial x_{i_0}} \right)^2 \leq \frac{3}{2} \sum_{i=1}^{m} \left( \frac{\partial c_i(x^*)}{\partial x_{i_0}} \right)^2.
\]
Take \( \alpha'_k = a_7 \Delta_k \), where
\[
a_7 = \min \left\{ \frac{1}{3|(A(x^*)c(x^*))_{i_0}|} \min \left\{ 0.8, \frac{x^*_{i_0} - l_{i_0}}{\Delta}, \frac{u_{i_0} - x^*_{i_0}}{\Delta} \right\}, \frac{1}{(1 + \frac{3}{2} \sum_{i=1}^{m} (\frac{\partial c_i(x^*)}{\partial x_{i_0}})^2 \Delta_k) \Delta} \right\},
\]
we then have that
\[
l \leq x_k + \alpha'_k s_k \leq u, \quad \|\alpha'_k s_k\|_\infty \leq 0.8 \Delta_k.
\]
So \( \alpha'_k s_k \) is a feasible solution of (2.1) for all \( k \in K_2, k \geq k_2 \). Therefore,
\[
\Phi_k(v_k) \leq \Phi_k(\alpha'_k s_k) = \frac{1}{2} \|c_k + \alpha'_k A_k s_k\|^2 + \frac{1}{2} \|\alpha'_k s_k\|^2
\]
\[
= \frac{1}{2} \|c_k\|^2 + \alpha'_k ((A_k c_k)_{i_0})^2 \left( 1 + \frac{\alpha'_k}{2} \left( 1 + \sum_{i=1}^{m} \left( \frac{\partial c_i(x_k)}{\partial x_{i_0}} \right)^2 \right) \right)
\]
\[
\leq \frac{1}{2} \|c_k\|^2 - \frac{a_7 \Delta_k}{2} ((A_k c_k)_{i_0})^2
\]
\[
\leq \frac{1}{2} \|c_k\|^2 - \frac{a_7}{4} ((A(x^*)c(x^*))_{i_0})^2 \Delta_k,
\]
which implies that
\[
\|c_k\|^2 - \|c_k + A_k d_k\|^2 \geq \frac{a_7}{2} ((A(x^*)c(x^*))_{i_0})^2 \Delta_k.
\]
So, by (2.7) and (3.29) and the relation above, we have that
\[
pred_k(\Delta_k) \geq \frac{a_7}{8} ((A(x^*)c(x^*))_{i_0})^2 \Delta > 0, \quad \forall k \in K_2, k \geq k_2,
\]
which is a contradiction with \( \lim_{k \to \infty} \pred_k = 0 \).
For case (2), we also deduce a contradiction analogously.
So (3.33) holds.
Finally, we prove that
\[
S^c = \{ k \in K | \rho_k = \rho_{1,k} \text{ and } \rho_{1,k} \neq \rho_{2,k} \} \subseteq S^{NM}
\] (3.34)
is an infinite set. Otherwise, there exists $k_4$ such that $\rho_k(\Delta_k) = \rho_{2,k}(\Delta_k)$ for all $k \in K$, $k \geq k_4$. Then, for all $k \geq k_4$, we have that

$$P_{k+1} = \theta_{k+1}\ell_{k+1} + (1 - \theta_{k+1})\varphi_{k+1}$$

$$= \theta_{k+1}\ell_{k+1} + (1 - \theta_{k+1})\varphi_{k+1} - \left[\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}\right]$$

$$+ \left[\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}\right]$$

$$= (\theta_{k+1} - \theta_k)\ell_{k+1} + (\theta_k - \theta_{k+1})\varphi_{k+1} + \left[\theta_k\ell_{k+1} + (1 - \theta_k)\varphi_{k+1}\right]$$

$$= (\theta_k - \theta_{k+1})(\varphi_{k+1} - \ell_{k+1}) + P_k - \text{Ared}_k,$$

where $\text{Ared}_k = P_k - P(x_{k+1}, \ell_{k+1}, \theta_k)$. By (3.11) and Lemma 3.4,

$$P_{k+1} = \left(\theta_k - \theta_{k+1} + \frac{\theta_kN}{(k + 1)^{1.1}}\right)(\varphi_{k+1} - \ell_{k+1})$$

$$+ P_k - \text{Ared}_k - \frac{\theta_kN}{(k + 1)^{1.1}}(\varphi_{k+1} - \ell_{k+1})$$

$$\leq \left(\theta_k - \theta_{k+1} + \frac{\theta_kN}{(k + 1)^{1.1}}\right)L_2 + P_k - \text{Ared}_k + \frac{\theta_kN}{(k + 1)^{1.1}}L_2$$

$$= (\theta_k - \theta_{k+1})L_2 + P_k - \text{Ared}_k + \frac{2\theta_kN}{(k + 1)^{1.1}}L_2$$

holds for all $k \geq k_4$. Writing the inequality above for $k = k_4, k_4 + 1, \ldots, \bar{k}$, we have that

$$P_{k_4+1} \leq (\theta_{k_4} - \theta_{k_4+1})L_2 + P_{k_4} - \text{Ared}_{k_4} + 2N\ell_2 \frac{\theta_{k_4}}{(k_4 + 1)^{1.1}}$$

$$P_{k_4+2} \leq (\theta_{k_4+1} - \theta_{k_4+2})L_2 + P_{k_4+1} - \text{Ared}_{k_4+1} + 2N\ell_2 \frac{\theta_{k_4+1}}{(k_4 + 2)^{1.1}},$$

$$\ldots \ldots \ldots \ldots$$

$$P_{\bar{k}+1} \leq (\theta_{\bar{k}} - \theta_{\bar{k}+1})L_2 + P_{\bar{k}} - \text{Ared}_{\bar{k}} + 2N\ell_2 \frac{\theta_{\bar{k}}}{(\bar{k} + 1)^{1.1}}.$$

Adding all these inequalities above, we obtain, after simplification, that

$$P_{\bar{k}+1} \leq (\theta_{k_4} - \theta_{\bar{k}+1})L_2 + P_{k_4} + 2N\ell_2 \sum_{j=k_4}^{\bar{k}} \frac{\theta_j}{(j + 1)^{1.1}} - \sum_{j=k_4}^{\bar{k}} \text{Ared}_j$$

$$\leq 2L_2 + P_{k_4} + 2N\ell_2 \sum_{j=k_4}^{\bar{k}} \frac{1}{(j + 1)^{1.1}} - \sum_{j=k_4}^{\bar{k}} \text{Ared}_j$$

(3.35)
holds for all \( k > k_4 \). From the convergence of the series \( \sum_{j=0}^{\infty} 1/(j+1)^{1.1} \) and from the boundedness of \( \{ P_k \} \), it follows that

\[
\sum_{j=k_4}^{\infty} A_{red,j} < +\infty,
\]

which implies that \( \lim_{k \to \infty} A_{red,k} = 0 \). On the other hand, for all \( k \in K, k \geq k_4, \rho_k = \rho_{2,k} \). So it follows, from \( A_{red,k} \geq \eta_1 \text{ pred}_k \geq 0, k \in K, k \geq k_4 \), that \( \lim_{k \in K} \text{ pred}_k = 0 \), which is a contradiction with (3.33). So \( S^\infty \) is an infinite set, which implies that \( S^{NM} \) is also an infinite set.

By Lemma 3.10, we prove the result. \( \Box \)

Now, it will be proved that, under the suitable assumptions, there exists at least a limit point of the model algorithm which is a substationary point of problem (1.1). Assume that the infinite sequence \( \{ x_k \} \) is generated by Algorithm 2.1. We first give the following assumption:

(AS4) For any limit point \( x^* \) of \( \{ x_k \} \), there exists a neighbourhood \( U(x^*) \) at \( x^* \) such that \( \text{rank} (A(x^*)) = \text{rank}(A(x^*)) \) for all \( x \in U(x^*) \).

**Lemma 3.12.** There exists an infinite set \( K \) such that \( \lim_{k \in K} \text{ pred}_k = \lim_{k \in K} \| v_k \| = 0 \).

**Proof.** Suppose, by contradiction, that there exists \( a_8 > 0 \) such that \( \text{ pred}_k \geq a_8, \forall k \). By Lemma 3.10, \( S^{NM} \) is a finite set. So we assume that there exists a positive integer \( k_5 \) such that \( CSUB_k(x^*_{OW}) = CSUB_2; k(x^*_{OW}) \) for all \( k \geq k_5 \). By (3.35), the series \( \sum_{j=k_5}^{\infty} A_{red,j} \) is convergent, which yields \( \lim_{k \to \infty} \text{ pred}_k = 0 \). This deduces a contradiction. So \( \lim \inf_{k \in K} \text{ pred}_k = 0 \), which implies that there exists an infinite subset \( K \) such that \( \lim_{k \in K} \text{ pred}_k = 0 \). On the other hand, for any \( k, v_k \) is a solution to (2.1), which implies \( \Phi_k(v_k) \leq \Phi_k(0) \). So, by \( A_k^T d_k = A_k^T v_k \),

\[
\| c_k \|^2 - \| c_k + A_k^T d_k \|^2 \geq \| v_k \|^2 \quad \forall k.
\]

By (2.7), \( \text{ pred}_k \geq 0.25 \| v_k \|^2 \), \( \forall k \), so \( \lim_{k \in K} \| v_k \| = 0 \). Thus the lemma is proved. \( \Box \)

We need a perturbation result, Theorem 4.2, in systems of linear inequalities in [5]. For convenience, we list it as follows:

Let \( A, A' \in R_{m_1 \times n}, B, B' \in R_{m_2 \times n}, C, C' \in R_{m_3 \times n}, a, a' \in R_{m_1}, b, b' \in R_{m_2}, c, c' \in R_{m_3}, \)

\[
D = \{ x | Ax \leq a, Bx \leq b, Cx = c \},
\]

\[
D' = \{ x | A'x \leq a', B'x \leq b', C'x = c' \},
\]

\[
e' = \max \{ \| A' - A \|, \| B' - B \|, \| C' - C \|, \| (a' - a)^+ \|, \| b' - b \|, \| c' - c \| \},
\]

where for any vector \( x \), we denote by \( x^+ \) that vector, of the same dimension, whose \( i \)th component equals the maximum of zero and the \( i \)th component of \( x \).
Lemma 3.13 (Daniel [5]). Let nonempty sets \( \mathcal{D} \) and \( \mathcal{D}' \) be defined by (3.36) and (3.37), respectively, and define \( \varepsilon' \) by (3.38). Suppose that

\[ (i) \text{ for all } x \in \mathcal{D}, \text{ we have } Bx = b, \text{ and either } A \text{ is vacuous or there exists } \hat{x} \in \mathcal{D} \text{ with } A\hat{x} \leq a - h, \]
where \( h > 0 \).
\[ (ii) \text{ rank}\left( \frac{b^T}{c} \right) = \text{rank}\left( \frac{b'_{\varepsilon}}{c} \right). \]

Then there exist positive constants \( \beta \) and \( \varepsilon_0 \) depending on \( \mathcal{D} \) such that to every \( x \in \mathcal{D} \) satisfying \( \varepsilon'(1 + \|x\|) \leq \varepsilon_0 \), there corresponds an \( x' \) in \( \mathcal{D}' \) satisfying \( \|x - x'\| \leq \beta \varepsilon'(1 + \|x\|) \).

Lemma 3.14. Assume that (AS4) holds, \( \liminf_k \Delta_k = \Delta' > 0 \) and all limit points of \( \{x_k\} \) are not substationary points of (1.1). Then there exists \( a_0 > 0 \) such that \( \liminf_k \Psi(d_k) = -a_0 \).

**Proof.** By Lemma 3.12, we can assume, without loss of generality, that there exists an infinite index subset \( K \) such that \( \lim_{x_k \in X} x_k = x^* \), \( \lim_{k \in K} \pred_k = 0 \), \( \lim_{k \in K} v_k = v^* = 0 \), \( \lim_{k \in K} \lambda_k = \lambda^* \). Since \( x^* \) is not a substationary point of problem (1.1), 0 is not a solution to the following subproblem

\[
\min \Psi_*(d) = \nabla \ell_*^T d + \frac{1}{2} L_3 d^T d
\]
\[ \text{s.t. } A_k^T d = 0, \]
\[ l \leq x^* + d \leq u, \]
\[ \|d\|_\infty \leq \Delta', \]
where \( \|B_k\| \leq L_3 \) for all \( k \). Let \( d^* \) be a solution to (3.39). Then, \( \Psi_*(d^*) < 0 \). Let \( \mathcal{D}^* = \{d \in R^n|A_k^T d = 0, l \leq x^* + d \leq u, \|d\|_\infty \leq \Delta'\} \), \( \mathcal{D}'_k = \{d \in R^n|A_k^T d = A_k^T v_k, l \leq x_k + d \leq u, \|d\|_\infty \leq \Delta'\} \).

Let
\[
\hat{x}_i = \begin{cases} 
\min\{0.5(u_i - x_i^*), 0.5\Delta'\} & \text{if } x_i^* = l_i, \\
\max\{0.5(l_i - x_i^*), -0.5\Delta'\} & \text{if } x_i^* = u_i, \\
0 & \text{if } l_i < x_i^* < u_i.
\end{cases}
\]

Then, \( \hat{x} \in \{d \in R^n| l - x^* < d < u - x^*, \|d\|_\infty < \Delta'\} \neq \emptyset \). By (AS4), \( \text{rank}(A(x_k)) = \text{rank}(A(x^*)) \) for all sufficiently large \( k \in K \). By Lemma 3.13, there exist positive constants \( \beta \) and \( \varepsilon_0 \) depending on \( \mathcal{D}^* \) such that for \( d^* \in \mathcal{D}^* \) satisfying \( \varepsilon_k'(1 + \|d^*\|) \leq \varepsilon_0 \), there corresponds a \( d'_k \in \mathcal{D}'_k \) satisfying \( \|d^* - d'_k\| \leq \beta \varepsilon'_k(1 + \|d^*\|) \), where \( \varepsilon'_k = \max\{|\|A_k - A_*\|, |A_k^T v_k|, |x_k - x^*|\} \). Noting that \( d_k \) is a solution to (2.2). Therefore, there exists \( k_1 \) such that for all \( k \in K, k \geq k_1 \),

\[
\Psi_k(d_k) \leq \Psi_k(d'_k) = \nabla \ell_k^T d'_k + \frac{1}{2} d_k^T B_k d'_k
\]
\[ \leq \nabla \ell_k^T d'_k + \frac{1}{2} L_3 \|d'_k\|^2
\]
\[ \leq \Psi_*(d^*) + \Delta'\|\nabla \ell_k - \nabla \ell_*\| + O(\varepsilon'_k)
\]
\[ \leq \frac{1}{2} \Psi_*(d^*) < 0
\]
exists a \in \mathbb{R} whenever \|a\| \not= 0. It follows from the definition of \text{pred} that \lim \inf_k \text{pred}_k (a_k) = -a_9. Thus the proof is concluded. \(\square\)

To prove the convergence, an additional assumption on the decrease \(\|c(x)\|^2 - \|c(x) + A(x)^T d\|^2\) is necessary, which is similar to the second algorithm assumption in [8].

(AS5) For a solution, \(d_k(\Delta), (2.2)\), there exists \(a_{10} > 0\) such that
\[
\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta)\|^2 \geq a_{10} \|c_k\| \Delta,
\]
whenever \(\|c_k\| > - \Psi_k(d_k(\Delta))/(4L_1)\).

Assumption (AS5) means that the algorithm will decrease mainly \(\|c(x_k + d)\|\) when \(\|c_k\| > - \Psi_k(d_k(\Delta))/(4L_1)\) and the algorithm does not terminate at \(x_k\).

**Lemma 3.15.** Assume that the hypotheses of Lemma 3.14 and (AS5) hold, then \(\lim_{k \to \infty} \theta_k = 0\).

**Proof.** By Lemma 3.12, there exists an infinite subset \(K\) such that \(\lim_{k \in K} \text{pred}_k = 0\), \(\lim_{k \in K} v_k = 0\). We can assume, without loss of generality, that \(\lim_{k \in K} x_k = x^*\). By the proof of Lemma 3.14, there exist \(a_9 > 0\), \(K_1 \subseteq K\) and \(k_1\) such that \(k \in K_1\), \(k \geq k_1\), \(\Psi_k(d_k) \leq -0.5a_9\). If \(\|c_k\| \not= - \Psi_k(d_k)/(4L_1)\), \(k \in K_1\), \(k \geq k_1\), then
\[
\text{pred}_k(\Delta_k) \geq 0.25(\|c_k\|^2 - \|c_k + A_k^T d_k\|^2) \quad \text{(by (2.7))}
\geq 0.25a_{10} \|c_k\| \Delta_k \quad \text{(by (AS5))}
\geq \frac{a_9a_{10}}{2L_1} \Delta' > 0 \quad \left(\lim \inf_k \Delta_k = \Delta' > 0\right),
\]
which implies from \(\lim_{k \in K} \text{pred}_k = 0\) that there exists \(k_2 \geq k_1\) such that \(\|c_k\| \leq - \Psi_k(d_k)/(4L_1)\) holds for all \(k \in K\), \(k \geq k_2\). Since \(\Phi_k(\|v_k\|) \leq \Phi_k(0), \|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k)\|^2 \geq 0\) and \(\|c_k\| - \|c_k + A_k^T d_k(\Delta_k)\| \geq 0\). By \(\|\lambda_k\| \leq L_1\), we have that
\[
\|\delta \lambda_k\| \leq \|\lambda_k + \delta \lambda_k\| + \|\lambda_k\| \leq 2L_1. \quad (3.40)
\]
It follows from the definition of \(\text{pred}_k\) that
\[
\text{pred}_k(\Delta_k) = \theta_k\left[ - \nabla c_k^T d_k(\Delta_k) - \frac{1}{2} d_k(\Delta_k)^T B_k d_k(\Delta_k) - (c_k + A_k^T d_k(\Delta_k))^T \delta \lambda_k \right]
\geq \frac{1}{2} \left(1 - \theta_k\right)(\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k)\|^2)
\geq \theta_k(\Psi_k(d_k) - (c_k + A_k^T d_k(\Delta_k))^T \delta \lambda_k)
\geq \theta_k(\Psi_k(d_k) - \|c_k + A_k^T d_k(\Delta_k)\| \|\delta \lambda_k\|)
\geq \theta_k(\Psi_k(d_k) - 2\|c_k\| L_1) \quad \text{(by (3.40))}
\geq -0.5\theta_k \Psi_k(d_k) \quad \text{(by \(\|c_k\| \leq - \Psi_k(d_k)/(4L_1)\))}
\geq 0.25a_9 \theta_k, \quad \forall k \in K_1, \; k \geq k_2.
\]
Since \(\lim_{k \in K} \text{pred}_k = 0\), we have that \(\lim_{k \in K_1} \theta_k = 0\). By Lemma 3.5, \(\{\theta_k\}\) is convergent, so the lemma is true. \(\square\)
Theorem 3.16. Assume that (AS4) and (AS5) hold, and \(\liminf_{k} \Delta_k = \tilde{\Delta} > 0\). Then there exists at least a limit point of \(\{x_k\}\) which is a substationary point of (1.1).

Proof. Suppose, by contradiction, that all limit points of \(\{x_k\}\) are not substationary point of the problem (1.1).

First, we have, by Lemma 3.15, that \(\lim_{k \to \infty} \theta_k = 0\).

Next, at any iterate point \(x_k\), we have, by Algorithm 2.1, that

\[
\rho_k(\Delta_k^{(i_k)}) \geq \eta_1, \quad \rho_k(\Delta_k^{(i-k)}) < \eta_1, \quad j = 1, 2, \ldots, i, \quad \theta_k = \theta_{k,i},
\]

where \(i_k\) is the iteration number in the inner cycle. By \(\liminf_{k} \Delta_k = \tilde{\Delta} > 0\) and \(\Delta_k^{(i_k)} = \eta_2 \Delta_{\min}\),

\[
\eta_2^{(i_k)} \geq \frac{\tilde{\Delta}}{\Delta_{\min}},
\]

which implies that

\[
i_k \leq \left[ \frac{\ln \tilde{\Delta} - \ln \Delta_{\min}}{\ln \eta_2} \right] \stackrel{\text{def}}{=} i_{\max}.
\]  

Let

\[
K^{(0)} = \{k \mid i_k = 0\},
\]

\[
K^{(1)} = \{k \mid i_k = 1\},
\]

\[
\ldots
\]

\[
K^{(i_{\max})} = \{k \mid i_k = i_{\max}\}.
\]

We assume, without loss of generality, that each set \(K^{(j)}(0 \leq j \leq i_{\max})\) is an infinite set.

Now, we prove that there exists a positive constant \(\theta' > 0\) such that

\[
\theta_{k,i}^{\sup} = \theta^{\sup}(x_k, \Delta_k^{(i)}) \geq \theta' > 0
\]

holds for any \(k\) and any \(i: 0 \leq i \leq i_k\).

For any \(k\) and \(0 \leq i \leq i_k\), we consider two cases:

Case 1: \(\|c_k\| \leq -\Psi_k(d_k(\Delta_k^{(i)}))/(4L_1)\)

By Algorithm 2.1, we have that,

\[
\|c_k + A_k^T d_k(\Delta_k^{(i)})\| \leq \|c_k\|.
\]

So, we have, from (3.40) and (3.44), that

\[
-\Psi_k(d_k(\Delta_k^{(i)})) - (c_k + A_k^T d_k(\Delta_k^{(i)}))^T \delta i_k
\]

\[
\geq -\Psi_k(d_k(\Delta_k^{(i)})) - 2\|c_k\| L_1
\]

\[
\geq 2L_1 \|c_k\|.
\]

Let

\[
K^L = \{k \mid \|c_k\| > L_1\}, \quad \tilde{K}^L = \{k \mid \|c_k\| \leq L_1\}.
\]
For \( k \in K^L \), it follows from (3.45) that

\[
-\nabla \Psi_k(d_k(\Delta_k^{(i)})) - (c_k + A_k^T d_k(\Delta_k^{(i)})) \delta \lambda_k \geq 2\|c_k\|^2,
\]

which implies from (2.7) that \( \theta \sup(x_k, \Delta_k^{(i)}) \equiv 1 \).

For \( k \in K^L \), we assume, without loss of generality, that \( K^L \) is an infinite set. Then we can prove that

\[
\lim_{k \to \infty} \|c_k\| \to \infty.
\]

(3.46)

holds for \( \forall i: 0 \leq i \leq i_{\max} \).

In fact, suppose, by contradiction, that (3.46) does not hold for some infinite subset \( K' \subseteq K^L \) and some \( i_0: 0 \leq i_0 \leq i_{\max} \). By (AS2) and the boundedness of \( \{d_k(\Delta^{(i_0)})\} \), we can assume that \( \lim_{k \in K'} x_k = x^* \) and \( \lim_{k \in K'} d_k(\Delta^{(i_0)}) = d^* \). It follows from (AS1) and the definition of \( K^L \) that \( \lim_{k \in K'} \|c(x_k)\| = \|c(x^*)\| \geq L_1 \). By Theorem 3.11, \( x^* \) is a \( \varphi \)-stationary point, which implies that

\[
(A \varphi c_\ast)_i = \begin{cases} 
0 & \text{if } x^*_i = l_i, \\
< 0 & \text{if } x^*_i = u_i, \\
0 & \text{if } l_i < x^*_i < u_i.
\end{cases}
\]

(3.47)

On the other hand,

\[
0 \leq \|c_k\| - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|
\]

\[
= \|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2 / \|c_k\| + \|c_k + A_k^T d_k(\Delta_k^{(i)})\|
\]

\[
\leq 1 / \|c_k\| (-2(A_k c_k)^T d_k(\Delta_k^{(i)}) - \|A_k^T d_k(\Delta_k^{(i)})\|^2).
\]

Taking \( k \in K' \), \( k \to \infty \), we get

\[
0 \leq -2 \sum_{x^*_i = l_i} (A \varphi c_\ast)_i d^*_i - 2 \sum_{x^*_i = u_i} (A \varphi c_\ast)_i d^*_i - \|A_k^T d^*\|^2 \leq 0.
\]

(3.48)

Note that \( l \leq x^* + d^* \leq u \), it follows from (3.47) and (3.48) that

\[
\lim_{k \in K'} (\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2) = 0
\]

which yields a contradiction. So (3.46) holds.

By (3.46), there exists \( k_1 \) such that

\[
4L_1 \geq \|c_k\| - \|c_k + A_k^T d_k(\Delta_k^{(i)})\| \quad \text{for } k \in K^L, \ k \geq k_1 \quad \text{and} \quad \forall i: 0 \leq i \leq i_{\max},
\]

which implies that

\[
2L_1\|c_k\| \geq 0.5\|c_k\| (\|c_k\| - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|)
\]

\[
\geq 0.25(\|c_k\| + \|c_k\|) (\|c_k\| - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|)
\]

\[
\geq 0.25(\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2)
\]
holds for $k \in K^L$, $k \geq k_1$ and $\forall i: 0 \leq i \leq i_{\text{max}}$. Therefore, (2.7) holds for sufficiently large $k \in K^L$ and $\forall i: 0 \leq i \leq i_{\text{max}}$, which implies that $\theta_{\text{sup}}(x_k, \Delta_k^{(i)}) \equiv 1$.

**Case 2:** $\|c_k\| > -\Psi_k(d_k(\Delta_k^{(i)}))/(4L_1)$

A trivial calculation shows that

$$\theta_{\text{sup}}(x_k, \Delta_k^{(i)}) = \min\{1, \tilde{\theta}_{k,i}\},$$

where

$$\tilde{\theta}_{k,i} = \frac{\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2}{2(\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2) + 4\Psi_k(d_k(\Delta_k^{(i)})) + 4(c_k + A_k^T d_k(\Delta_k^{(i)}))^T \delta \lambda_k}.$$

It follows that

$$\frac{1}{2\tilde{\theta}_{k,i}} = 1 + \frac{2\Psi_k(d_k(\Delta_k^{(i)}))}{\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2} + \frac{2(c_k + A_k^T d_k(\Delta_k^{(i)}))^T \delta \lambda_k}{\|c_k\|^2 - \|c_k + A_k^T d_k(\Delta_k^{(i)})\|^2}.$$

So by (AS5),

$$\frac{1}{2\tilde{\theta}_{k,i}} \leq 1 + \frac{4L_1}{a_{10} \Delta_k^{(i)}} + \max\left\{0, \frac{2\Psi_k(d_k(\Delta_k^{(i)}))}{a_{10}\|c_k\|^2}\right\}.$$  

(3.49)

Now we can prove that

$$\lim \inf_{k \in K^{(i)}} \tilde{\theta}_{k,i} > 0$$  

(3.50)

holds for $\forall j: 0 \leq j \leq i_{\text{max}}$, $\forall i: 0 \leq i \leq j$.

In fact, suppose, by contradiction, that (3.50) does not hold for some $j: 0 \leq j \leq i_{\text{max}}$ and some $i: 0 \leq i \leq j$, which implies that there exists an infinite subset $K \subseteq K^{(i)}$ such that

$$\lim \inf_{k \in K} \tilde{\theta}_{k,i} = 0.$$  

(3.51)

By (AS2), we can assume that $\lim_{k \in K} x_k = x^\ast$. It follows from (AS1) that $\lim_{k \in K} c(x_k) = c(x^\ast)$. If $c(x^\ast) \neq 0$, then it follows, by (3.49) and $\Delta_k^{(i)} \geq \tilde{\Delta}$, that $\tilde{\theta}_{k,i}$ is bounded away from zero, which implies a contradiction with (3.51). If $c(x^\ast) = 0$, it follows from $\Phi_k(v_k(\Delta_k^{(i)})) \leq \Phi_k(0)$ that $\lim_{k \in K} \|v_k(\Delta_k^{(i)})\| = 0$. Noting that $x^\ast$ is not a substationary point of (1.1), it follows, from the proof of Lemma 3.14, that there exists $k_1$ such that

$$\Psi_k < 0 \quad \forall k \in K, \ k \geq k_1.$$  

From (3.49) and $\Delta_k^{(i)} \geq \tilde{\Delta}$, we have that

$$\frac{1}{2\tilde{\theta}_{k,i}} \leq 1 + \frac{4L_1}{a_{10} \Delta} \quad \forall k \in K, \ k \geq k_1,$$

which also deduces a contradiction with (3.51). So (3.50) holds. Therefore, (3.43) holds.

By (3.9) and Algorithm 2.1,

$$\theta_k = \theta_{k,i} = \min\{\theta_{k,i}^{\sup}, \theta_{k,i-1}^{\sup}, \ldots, \theta_{k,0}^{\sup}, \theta_k^{\text{large}}\}.$$
It follows from \( \lim_{k \to \infty} \theta_k = 0 \) and (3.43) that \( \lim_{k \to \infty} \theta_k^{\text{large}} = 0 \). So there exists \( k_2 \) such that \( \theta_k = \theta_k^{\text{large}} \) for all \( k \geq k_2 \). By the definition of \( \theta_k^{\text{min}} \) and \( \theta_k^{\text{large}} \), \( \theta_{k_2} = \theta_{k_2}^{\text{large}} > \theta_{k_2}^{\text{min}} \). So \( \theta_{k_2 + 1}^{\text{min}} = \min\{\theta_{k_2}^{\text{min}}, \theta_{k_2}^{\text{large}}\} = \theta_{k_2}^{\text{min}} \).

Analogously, we have that
\[
\theta_k > \theta_{k_2}^{\text{min}}, \quad \theta_k^{\text{min}} = \theta_{k_2}^{\text{min}}, \quad \forall k \geq k_2,
\]
which deduces a contradiction with \( \lim_{k \to \infty} \theta_k = 0 \).

Thus the result is proved. \( \Box \)

We now summarize all the previous results as follows:

**Theorem 3.17.**

(i) All limit points are \( \varphi \)-stationary point.

(ii) Under the assumption of Theorem 3.16, at least one limit point is a substationary point of (1.1).

(iii) A feasible substationary limit point is a stationary point of (1.1).

### 4. Numerical experiments

A FORTRAN subroutine is written to test Algorithm 2.1. The trust region subproblems (2.1) and (2.2) have the general form
\[
\min \ g^T x + \frac{1}{2} x^T B x \\
\text{s.t.} \quad a(\cdot, j)x = b_j, \quad j = 1, 2, \ldots, m_e, \\
\quad a(\cdot, j)x \geq b_j, \quad j = m_e + 1, 2, \ldots, m,
\]
which can be solved by Fletcher’s Harwell subroutine VE02AD.

The intermediate penalty parameter \( \theta' \) is chosen using (3.8) when \( \varphi \omega \geq 10^{-3} \), and using (3.9) when \( \varphi \omega \approx 10^{-3} \). This apparent modification of Algorithm 2.1 was motivated by preliminary numerical experiments (see [8]). However, it is easy to observe that it does not represent a real alteration of the model algorithm, since it can always be interpreted that the first trust region radius tried at iteration \( k \) is the last one which is greater than \( 10^{-3} \). So, only when the trust region radius is less than \( 10^{-3} \), it is necessary to decrease the penalty parameter, as required by the convergence theory.

The test examples that we have run are from [9,15]. For each problem, we choose initial parameters \( \Delta_{\min} = 10, \quad \eta_1 = 0.001, \quad \eta_2 = 0.25, \quad m(0) = 0 \). The stopping condition is either
\[
\|c(x_k)\|_{\infty} < 10^{-6} \quad \text{and} \quad \|g_k + A_k \hat{\lambda}_k\|_{\infty} < 10^{-6}
\]
or
\[
\|v_k(\Delta)\| < 10^{-8} \quad \text{and} \quad \|d_k(\Delta)\| < 10^{-8},
\]
where \( \hat{\lambda}_k \) is Lagrange multiplier of (2.2). \( B_k \) is updated by means of the Powell’s safeguarded BFGS update formula.

The results are listed in Tables 1 and 2. In the tables, the problems are numbered in the same way as in [9]. For example, “HS26” means problem 26 in Hock and Schittkowski (1981). \( n \),
\[ \|c(x_k)\|_\infty \]

<table>
<thead>
<tr>
<th>Problem</th>
<th>(n)</th>
<th>(m_c)</th>
<th>(M)</th>
<th>NI-NF-NG</th>
<th>(|c(x_k)|_\infty)</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>31-41-31</td>
<td>8.3437259E−07</td>
<td>6.7468402E−08</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>52-82-52</td>
<td>9.0890782E−07</td>
<td>2.3665413E−07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>4</td>
<td>49-56-49</td>
<td>6.9652771E−07</td>
<td>2.0747980E−07</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>6</td>
<td>26-30-26</td>
<td>6.6273322E−07</td>
<td>4.7283880E−07</td>
</tr>
<tr>
<td>HS28</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>9-10-9</td>
<td>0.0000000E+00</td>
<td>7.7644788E−08</td>
</tr>
<tr>
<td>HS41</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>13-19-13</td>
<td>1.4375057E−11</td>
<td>2.9782904E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>11-11-11</td>
<td>2.9999991E−10</td>
<td>5.2134752E−07</td>
</tr>
<tr>
<td>HS46</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>33-48-33</td>
<td>7.8495870E−08</td>
<td>3.7340467E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>44-67-44</td>
<td>1.7102660E−07</td>
<td>3.5398342E−07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>4</td>
<td>44-67-44</td>
<td>1.7102660E−07</td>
<td>3.5398342E−07</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>6</td>
<td>40-58-40</td>
<td>1.3056817E−07</td>
<td>9.8062288E−07</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td>12</td>
<td>59-69-59</td>
<td>2.4199293E−07</td>
<td>2.4532796E−07</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
<td>14</td>
<td>45-51-45</td>
<td>6.9962143E−07</td>
<td>7.3214576E−08</td>
</tr>
<tr>
<td>HS53</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>49-94-49</td>
<td>8.5989074E−07</td>
<td>4.2947415E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>49-49-49</td>
<td>8.5989074E−07</td>
<td>4.2947210E−07</td>
</tr>
<tr>
<td>HS60</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>12-14-12</td>
<td>3.2825878E−11</td>
<td>1.9841890E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>11-11-11</td>
<td>2.2505605E−10</td>
<td>1.9853851E−07</td>
</tr>
<tr>
<td>HS61</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>12-39-12</td>
<td>1.6952203E−10</td>
<td>4.7684127E−06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>18-26-18</td>
<td>5.8242068E−11</td>
<td>4.0706472E−05</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>4</td>
<td>22-25-22</td>
<td>7.8565450E−11</td>
<td>4.7277187E−05</td>
</tr>
<tr>
<td>HS62</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>62-277-62</td>
<td>2.7755576E−17</td>
<td>9.9906107E−04</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>63-109-63</td>
<td>2.7755576E−17</td>
<td>1.9207302E−04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>4</td>
<td>40-48-40</td>
<td>0.0000000E+00</td>
<td>9.1504717E−07</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>6</td>
<td>35-44-35</td>
<td>2.7755576E−17</td>
<td>2.3949161E−05</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td>8</td>
<td>44-48-44</td>
<td>2.7755576E−17</td>
<td>9.3875042E−06</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
<td>14</td>
<td>43-45-43</td>
<td>5.5511151E−17</td>
<td>1.9912503E−04</td>
</tr>
<tr>
<td>HS112</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>47-187-47</td>
<td>2.9165038E−17</td>
<td>3.4063407E−06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>51-153-51</td>
<td>7.0256301E−17</td>
<td>2.6073900E−07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td>4</td>
<td>60-115-60</td>
<td>7.1774184E−17</td>
<td>2.6038647E−06</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td>6</td>
<td>54-82-54</td>
<td>5.8980598E−17</td>
<td>9.2592519E−07</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td>8</td>
<td>48-62-48</td>
<td>8.5868812E−17</td>
<td>8.9716074E−07</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td>10</td>
<td>44-64-44</td>
<td>8.4134089E−17</td>
<td>2.0437891E−07</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td>12</td>
<td>53-53-53</td>
<td>8.2182525E−17</td>
<td>7.0874668E−07</td>
</tr>
<tr>
<td>HS216</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>38-63-38</td>
<td>2.9456213E−09</td>
<td>9.4536557E−06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td>2</td>
<td>8-9-8</td>
<td>5.8803989E−08</td>
<td>1.5929800E−09</td>
</tr>
<tr>
<td>HS46*</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>fails</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

\(m\) means numbers of the variables and equality constraints, respectively. NI, NF and NG means numbers of iterations, function evaluations and gradient evaluations, respectively. If NI \(> 300\) or NF \(> 1000\), then we regard the method as failing. \(\mathcal{N}\) denotes the degree of nonmonotonicity of the penalty parameter \(\theta\), according to (3.7). In the tables, only the results for \(\mathcal{N} = 0\) and \(\mathcal{N} = 10^6\) are shown. Moreover, each code runs from \(M = 0, 2, \ldots, 16\), where \(M = 0\) means that we use monotone trust region method. We omit the results if the results are the same as that of the last value of \(M\).
According to the results in the tables, the nonmonotone method is superior to the monotone one for some problems. And the preliminary numerical results show that the computational results for each test problem will not change when $M$ is added to some value, which is uniform with the phenomenon in [16]. Especially, when “HS46*” is solved, the method all fails for $N = 0$, not only for monotone trust region method but also for nonmonotone trust region method. But for $N = 10^6$, the method only fails when $M = 2$.

According to the results in the tables, the nonmonotone method is superior to the monotone one for some problems. And the preliminary numerical results show that the computational results for each test problem will not change when $M$ is added to some value, which is uniform with the phenomenon in [16]. Especially, when “HS46*” is solved, the method all fails for $N = 0$, not only for monotone trust region method but also for nonmonotone trust region method. But for $N = 10^6$, the method only fails when $M = 2$. 

### Table 2

<table>
<thead>
<tr>
<th>Problem</th>
<th>$n$</th>
<th>$m_e$</th>
<th>$M$</th>
<th>NI-NF-NG</th>
<th>$|c(x_k)|_\infty$</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS26</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>41-67-41</td>
<td>8.1274813E−07</td>
<td>3.2291963E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>41-48-41</td>
<td>8.3675325E−07</td>
<td>5.2195262E−08</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>65-74-65</td>
<td>5.4520843E−07</td>
<td>9.7710942E−07</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>44-47-44</td>
<td>3.3132538E−07</td>
<td>2.7295370E−07</td>
</tr>
<tr>
<td>HS28</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>9-10-9</td>
<td>0.0000000E+00</td>
<td>7.7644788E−08</td>
</tr>
<tr>
<td>HS41</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>13-19-13</td>
<td>1.4375057E−11</td>
<td>2.9782904E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>11-11-11</td>
<td>2.9999991E−10</td>
<td>5.2134752E−07</td>
</tr>
<tr>
<td>HS46</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>43-93-43</td>
<td>4.6895245E−08</td>
<td>4.2469256E−07</td>
</tr>
<tr>
<td>HS53</td>
<td>5</td>
<td>3</td>
<td>0</td>
<td>49-64-94</td>
<td>8.5989074E−07</td>
<td>4.2947415E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>49-50-49</td>
<td>8.5989074E−07</td>
<td>4.2947210E−07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>49-49-49</td>
<td>8.5989074E−07</td>
<td>4.2947210E−07</td>
</tr>
<tr>
<td>HS60</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>12-15-12</td>
<td>3.2825878E−11</td>
<td>1.9841980E−07</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>12-39-12</td>
<td>2.2505605E−10</td>
<td>1.9853851E−07</td>
</tr>
<tr>
<td>HS61</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>12-39-12</td>
<td>1.6952203E−10</td>
<td>4.7684127E−06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>12-39-18</td>
<td>1.6952203E−10</td>
<td>4.7684127E−06</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>12-38-12</td>
<td>1.6952203E−10</td>
<td>4.7684127E−06</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>12-36-12</td>
<td>1.6952203E−10</td>
<td>4.7684127E−06</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>13-25-13</td>
<td>5.2045208E−11</td>
<td>4.0168446E−05</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td>12-13-12</td>
<td>9.5323215E−11</td>
<td>5.2046465E−05</td>
</tr>
<tr>
<td>HS62</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>62-277-62</td>
<td>2.7755576E−17</td>
<td>9.9906107E−04</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>63-109-63</td>
<td>2.7755576E−17</td>
<td>1.9207302E−04</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>40-48-40</td>
<td>0.0000000E+00</td>
<td>9.1504717E−07</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>35-44-35</td>
<td>2.7755576E−17</td>
<td>2.3949161E−05</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>44-48-44</td>
<td>2.7755576E−17</td>
<td>9.3875042E−06</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td></td>
<td></td>
<td>43-45-43</td>
<td>5.5511151E−17</td>
<td>1.9912503E−04</td>
</tr>
<tr>
<td>HS112</td>
<td>10</td>
<td>3</td>
<td>0</td>
<td>47-187-47</td>
<td>2.9165038E−17</td>
<td>3.4063407E−06</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>51-153-51</td>
<td>7.0256301E−17</td>
<td>2.6073900E−07</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>60-117-60</td>
<td>6.6570013E−17</td>
<td>4.9286462E−07</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
<td></td>
<td>54-82-54</td>
<td>5.8980598E−17</td>
<td>9.2592519E−07</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td></td>
<td></td>
<td>48-62-48</td>
<td>8.5868812E−17</td>
<td>8.9716074E−07</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td></td>
<td></td>
<td>44-64-44</td>
<td>8.4134089E−17</td>
<td>2.0437891E−07</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td>53-53-53</td>
<td>8.2182525E−17</td>
<td>7.0874668E−07</td>
</tr>
<tr>
<td>HS216</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>fails</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>8-9-8</td>
<td>5.8800398E−08</td>
<td>1.5929800E−09</td>
</tr>
<tr>
<td>HS46*</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>34-101-34</td>
<td>1.5630135E−07</td>
<td>8.1489818E−08</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>fails</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td>39-42-39</td>
<td>7.7298518E−07</td>
<td>1.2084644E−09</td>
</tr>
</tbody>
</table>
Remark. In the tables, if some problem has no simple bounds in \([9,15]\), we all add the simple bounds \(-10 \leq x_i \leq 10, \ i = 1, 2, \ldots, n\). Moreover, if some initial point \(x_0 \not\in D = \{x \in \mathbb{R}^n | l \leq x \leq u\}\), we take the projection of \(x_0\) on \(D\). The problem “HS46” is adapted as the following “HS46∗”:

\[
\begin{align*}
\min & \quad f(x) = (x_1 - x_2)^2 + (x_3 - 1)^2 \\
\text{s.t.} & \quad x_1^2 x_4 + \sin(x_4 - x_5) - 1 = 0, \\
& \quad x_2 + x_3^4 x_4^2 - 2 = 0.
\end{align*}
\]

5. Discussions

In this paper, we present a nonmonotone trust region method with nonmonotone penalty parameters for minimizing differentiable functions with general equality constraints and simple bounds. We believe that the results in this paper are of significance.

The method presented in this paper has the following characteristics: the method is a globalized sequential quadratic programming algorithm; it uses trust region as a globalization strategy; the merit function is an augmented Lagrangian (thus, it is differentiable); the estimates of the Lagrangian multipliers are arbitrary (although we use the Lagrangian multiplier obtained when we solve subproblem (2.2)); the behavior of the penalty parameter is not monotone, which avoids the overflow resulted from too large penalty in the implementation of the algorithm. Furthermore, the sequence of the merit function values is not monotone, either, which relax the restriction of accepting new iterate point. At each iterate, we only solve two quadratic subproblems. All these features are valuable for the development of practical algorithms. Sequential quadratic programming is the most natural extension of Newton’s method to constrained optimization and the trust region approach allows one to deal consistently with infeasibility of quadratic subproblems. The nonmonotonicity feature tends to avoid the inheritance of unnecessary extreme values of penalty parameters from the first few iterations and the iterate sequence can follow the bottom of curved narrow valleys much more loosely, and maybe, results in longer and more efficient steps. Gomes et al. [8] pointed out that the method with nonmonotone penalty parameter is very robust and effective especially for some nonlinear optimization problems. On the other hand, Toint [16] also pointed out that the nonmonotone strategy hopefully results in longer and more efficient steps. The convergent results and the preliminary numerical tests in this paper shows that the method combining these two ideas is very interesting and of significance. It is necessary to unravel further the more exact behavior of such methods and all their characteristics.

Acknowledgements

The authors are very grateful for the valuable and detailed comments and suggestions of two anonymous referees, which greatly improved the presentation of the paper.

References


