On the convergence of an inexact Gauss-Newton trust-region method for nonlinear least-squares problems with simple bounds

by M. Porcelli

Report naXys-01-2011 3 January 2011

**Namur Center for Complex Systems**
University of Namur
8, rempart de la vierge, B5000 Namur (Belgium)

http://www.naxys.be

# On the convergence of an inexact Gauss-Newton trust-region method for nonlinear least-squares problems with simple bounds

M. Porcelli

3 January 2011

**Abstract**

We introduce an inexact Gauss-Newton trust-region method for solving bound-constrained nonlinear least-squares problems where, at each iteration, a trust-region subproblem is approximately solved by the Conjugate Gradient method. Provided a suitable control on the accuracy to which we attempt to solve the subproblems, we prove that the method has global and asymptotic fast convergence properties.

**Keywords:** bound-constrained nonlinear least-squares, simple bounds, trust-region methods, convergence theory, affine scaling.

## 1 Introduction

We address the solution of the nonlinear least-squares problem with simple bounds

$$\min_{x\in\Omega} \theta(x) = \frac{1}{2}\|\Theta(x)\|^2, \qquad \text{(BCLS)}$$

where $\theta : \mathbb{R}^n \to \mathbb{R}$, $\Theta : \mathbb{R}^n \to \mathbb{R}^m$ is a given continuously differentiable mapping and $\Omega$ is the $n$-dimensional box $\Omega = \{x \in \mathbb{R}^n \mid l \le x \le u\}$, $l \in (\mathbb{R} \cup -\infty)^n$, $u \in (\mathbb{R} \cup \infty)^n$, $l < u$. Remarkably, we do not make any assumption on the relationship between the dimensions $m$ and $n$. If $m > n$ we say the (BCLS) problem is overdetermined. If $m = n$ we have a square problem; if $m < n$ the problem is underdetermined.

In the last few years, a number of different globally convergent methods for the solution of (BCLS) have been proposed. Many of them, e.g. [2, 3, 4, 9, 14, 15, 16, 17, 19, 22], are based on the so-called Affine Scaling approach introduced in [7] by Coleman and Li. This approach relies on the observation that the first-order optimality conditions (KKT conditions) for (BCLS) can be written as a nonlinear system of equations

$$D(x)J(x)^T\Theta(x) = 0,$$

for $x \in \Omega$, where $J$ denotes the Jacobian matrix of $\Theta$ and $D$ is the following diagonal scaling matrix

$$D(x) = diag(|v_1(x)|, \ldots, |v_n(x)|), \qquad (1)$$

with

$$v_i(x) = \begin{cases} x_i - u_i & \text{if } (\nabla\theta(x))_i < 0,\ u_i < \infty, \\ x_i - l_i & \text{if } (\nabla\theta(x))_i \ge 0,\ l_i > -\infty, \\ 1 & \text{if } (\nabla\theta(x))_i \ge 0,\ l_i = -\infty \ \text{ or } \ (\nabla\theta(x))_i < 0,\ u_i = \infty, \end{cases} \qquad (2)$$

for $i = 1, \ldots, n$.

Trust-region methods for square (BCLS) problems were studied both theoretically and computationally in [2, 3, 4, 14]. Francisco et al. [9] designed a trust-region Gauss-Newton method for underdetermined (BCLS) problems. Under full rank assumptions on $J$, local quadratic convergence is proved to interior zero residual points, i.e. points $x^*$ such that $l < x^* < u$ and $\Theta(x^*) = 0$. Kanzow and Petra [15] and Zhu [22] proposed global Levenberg-Marquardt methods for semismooth problems of the form (BCLS). In particular, the paper [15] is focused on overdetermined problems arising from a suitable reformulation of mixed complementarity problems; global convergence is shown and local convergence to zero residual solutions is proved under an error bound assumption. In [22], a nonmonotone interior backtracking line search technique is proposed to ensure the global convergence and, under a local error bound condition, local fast convergence to nondegenerate zero residual solutions is achieved. In [16, 17, 19], trust-region methods have been proposed for problem (BCLS) regardless of its dimensions. These procedures differ in the quadratic model used during the iterations: one method employs a Gauss-Newton model while the other one is based on a regularized Gauss-Newton model and is a Levenberg-Marquardt method. They offer global convergence properties combined with potentially fast local convergence both to degenerate and nondegenerate zero residual solutions. In particular, the local convergence analysis is carried out under full rank assumptions on $J$ for the Gauss-Newton method and under an error bound assumption in the Levenberg-Marquardt case.

It is important to note that all the methods mentioned (except [4]) rely on direct methods for linear systems and linear least-squares problems. These direct methods may be preferable to iterative methods when the cost of a matrix factorization is not excessive, e.g. if the dimension of the problem is sufficiently small or the Jacobian matrix is structured. Otherwise, it becomes necessary to use iterative methods for the solution of the linear systems and linear least-squares problems arising at each iteration.

The purpose of this work is twofold. Firstly, we shall design an inexact version of the Gauss-Newton trust-region method proposed in [16, 17, 19] based on the use of iterative methods for the linear algebra phase. Secondly, we shall analyze the impact of the use of the iterative linear algebra on the convergence properties of the inexact method.

Our study is motivated by the good practical performance of the Gauss-Newton method shown in [19] and by the promising use of the method in different contexts, see [1, 5, 13]. Since the current implementation [19] is limited to the use of direct linear algebra techniques, it is suitable for medium size problems; therefore the inexact procedure presented in this paper widens the applicability of our approach.

Throughout the paper, the subscript $k$ will denote an iteration counter and for any function $h$, $h_k$ will be the shorthand for $h(x_k)$. The 2-norm is denoted by $\|\cdot\|$ and $B_\rho(y) = \{x : \|x - y\| < \rho\}$ with $\rho > 0$. The symbol $I_p$ represents the identity matrix of dimension $p$. Let $A^+$ denote the Moore-Penrose pseudoinverse of the matrix $A$. Finally, we let $P_\Omega(x)$ be the projection of $x$ onto $\Omega$, i.e. $(P_\Omega(x))_i = \max\{l_i, \min\{x_i, u_i\}\}$, $i = 1, \dots, n$.

## 2 Description of the method

In this section we describe the $k$th iteration of an inexact Gauss-Newton trust-region method, called `ITREBO` (Inexact Tust-REgion method for BOund-constrained least-squares problems), for solving the problem (BCLS). The sequence $\{x_k\}$ generated by the method consists of feasible points for (BCLS), i.e. $x_k \in \Omega, k > 0$.

Given $x_k \in \Omega$ and $\Delta_k > 0$, we consider the following trust region problem

$$\min \{m_k(p) = \frac{1}{2}\|J_k\, p + \Theta_k\|^2 \;:\; \|p\| \le \Delta_k\}. \tag{3}$$

Correspondingly to problem (3), we consider a generalized Cauchy step $p_k^C$ along the scaled steepest descent direction $d_k$ defined as

$$d_k = -D_k \nabla\theta_k,$$

where $D$ is the scaling matrix defined in (1) and (2), that has the favorable property to be well-angled with respect to the simple bounds [6]. The step $p_k^C$ minimizes $m_k$ along the direction $d_k$ within the feasible trust-region, i.e.

$$p_k^C \quad = \quad \underset{p \in span\{d_k\}}{\text{argmin}}\; m_k(p) \quad \text{subject to} \quad \|p\| \le \Delta_k, \quad x_k + p \in \Omega. \tag{4}$$

The $k$th iteration of the `ITREBO` procedure is outlined in Algorithm 2.1 and the main steps are now sketched. In Step 1, the trust-region problem (3) is solved approximately ("inexactly") by the Conjugate Gradient (CG) method. The description of this step constitutes the main contribution of this section

**Algorithm 2.1:** `ITREBO`: $k$-TH ITERATION

Input: $x_k \in \Omega$, $0 < \Delta_{min} \le \Delta_k$, $0 \le \eta_k \le \eta_{max} < 1$, $\beta_1$, $\beta_2, \delta \in (0,1)$.

1. Compute the inexact trust-region step $p_{tr}$ for (3) by the CG method.
2. Let $\bar{p}_{tr} = P_\Omega(x_k + p_{tr}) - x_k$.
3. Compute the generalized Cauchy step $p_k^C$ based on (4).
4. If

$$\rho_c(\bar{p}_{tr}) = \frac{m_k(0) - m_k(\bar{p}_{tr})}{m_k(0) - m_k(p_k^C)} \ge \beta_1, \tag{5}$$

Set $p_k = \bar{p}_{tr}$;

Else find $t \in (0,1]$ such that

$$p_k = t\, p_k^C + (1-t)\bar{p}_{tr}, \tag{6}$$

satisfies $\rho_c(p_k) \ge \beta_1$.

5. If

$$\rho_\theta(p_k) = \frac{\theta(x_k) - \theta(x_k + p_k)}{m_k(0) - m_k(p_k)} \ge \beta_2, \tag{7}$$

Set $x_{k+1} = x_k + p_k$, choose $\Delta_{k+1} \ge \Delta_{min}$ and $\eta_{k+1} \in [0, \eta_{max}]$;

Else reduce $\Delta_k$, $\Delta_k = \delta\Delta_k$ and go to Step 1.

and details shall be given further on. To generate a new feasible iterate, in Step 2 we project $x_k + p_{tr}$ onto the box $\Omega$ and define a possibly modified trust-region step $\bar{p}_{tr}$ such that $x_k + \bar{p}_{tr}$ is feasible. Steps 3–5 attempt to find a feasible iterate $x_{k+1} = x_k + p_k$ which provides a sufficient decrease in the value of $\theta$ with respect to $x_k$. In Step 4, we impose a sufficient decrease of $m_k$ in comparison to the generalized Cauchy step $p_k^C$; this is crucial to make the method globally convergent. We let $p_k = \bar{p}_{tr}$ if $\bar{p}_{tr}$ satisfies (5), otherwise, we look for a convex combination of the steps $p_k^C$ and $\bar{p}_{tr}$ such that the condition (5) is satisfied. In Step 5, we measure the quality of the quadratic model $m_k$ as an approximation to $\theta$ around $x_k$. If the sufficient improvement condition (7) is satisfied, the new iterate is $x_{k+1} = x_k + p_k$; otherwise, $p_k$ is rejected and the trust-region size $\Delta_k$ is reduced. We refer the reader to [19] for further details on the Steps 2-5 of the algorithm.

Now, we give an insight to Step 1 of Algorithm 2.1, i.e. to the computation of an inexact trust-region step $p_{tr}$. The basic idea of the methods proposed in [16, 17, 19] was to use a dogleg strategy in order to take the, possibly projected, minimum norm step

$$p_k^N = -J_k^+ \Theta_k, \tag{8}$$

as frequently as possible, taking advantage of its good properties in a neighbourhood of a zero-residual solution. The aim of the `ITREBO` method is to retain this property and at the same time to avoid the high computational cost in forming the minimum norm step by direct factorization methods.

Given $x_k \in \Omega$, we consider the trust-region problem (3). Let $p_k^{(0)} = 0$ and $\{p_k^{(j)}\}$ be the sequence of iterates generated by the CG method applied to the normal equations

$$J_k^T J_k p = -J_k^T \Theta_k. \tag{9}$$

We know that for $j \ge 1$

$$p_k^{(j)} = \text{argmin}\,\{m_k(p) \;:\; p \in \mathcal{K}_j\}, \tag{10}$$

where $\mathcal{K}_j$ is the $j$th Krylov subspace $\mathcal{K}_j = span(\{(J_k^T J_k)^i J_k^T \Theta_k\}_{i=0}^{j-1})$. Let $p_k^I$ be the first CG iterate producing a prescribed reduction of the value of $\nabla m_k$, i.e.

$$p_k^I = \text{argmin}\,\{m_k(p) \;:\; p \in K_j\}, \qquad \|\nabla m_k(p_k^I)\| \le \eta_k \|\nabla m_k(0)\|, \tag{11}$$

where $\eta_k \in [0, 1)$ is the so-called forcing term. We note that by $\nabla m_k(p) = J_k^T(J_k\, p + \Theta_k)$ and (11) it follows

$$\begin{aligned} J_k^T J_k p_k^I &= -J_k^T \Theta_k + r_k, \\ \|r_k\| &\leq \eta_k \|J_k^T \Theta_k\|, \end{aligned} \tag{12}$$

i.e. $p_k^I$ is an inexact Gauss-Newton step for the problem $\nabla\theta(x) = 0$ and the forcing term $\eta_k$ is used to control the accuracy in the solution of the system (9) [8].

Since we initialize $p_k^{(0)}$ to zero, each iterate $p_k^{(j)}$ is larger in norm than its predecessor and CG terminates in a finite number of iterations computing the minimum norm step $p_k^N$ (8), see [11]. This implies that

$$\|p_k^I\| \leq \|p_k^N\|. \tag{13}$$

Therefore, we stop the CG iterations as soon as either the specified accuracy (12) is achieved or the trust-region boundary is reached. In the former case the approximate trust-region solution is the low-dimensional unconstrained minimizer $p_k^I$ of $m_k$. In the latter case, no further iterates giving lower value of $m_k$ will be inside the trust-region. If $p_k^{(j)}$ is such that $\|p_k^{(j-1)}\| < \Delta_k \leq \|p_k^{(j)}\|$ then we take the Steihaug-Toint point $p_k^{ST}$ [20, 21], that is

$$p_k^{ST} = p_k^{(j-1)} + \tau(p_k^{(j)} - p_k^{(j-1)}),\ \tau \in (0, 1] \ \text{ such that } \ \|p_k^{ST}\| = \Delta_k.$$

The Steihaug-Toint point has the favorable property that the optimal decrease of $m_k$ at the exact solution of the trust-region problem (3), i.e. $(m_k(0) - m_k(p_{tr}))$, is no more than twice that achieved at $p_k^{ST}$ [23]. This process is described in Algorithm 2.2.

**Algorithm 2.2:** COMPUTING THE INEXACT TRUST-REGION STEP $p_{tr}$ BY CG

Input: $x_k$, $0 \leq \eta_k < 1$, $\Delta_k > 0$.

1. Set $p_k^{(0)} = 0$ and $j = 1$;
2. Compute the $j$-th CG iterate $p_k^{(j)}$ given in (10);
3. If $\|p_k^{(j)}\| \leq \Delta_k$ and $p_k^{(j)}$ satisfies (12),
   then set $p_k^I = p_k^{(j)}$, return $p_{tr} = p_k^I$.
4. If $\|p_k^{(j)}\| > \Delta_k$, find $\tau \in (0, 1]$ such that
   $p_k^{ST} = p_k^{(j-1)} + \tau(p_k^{(j)} - p_k^{(j-1)})$ satisfies $\|p_k^{ST}\| = \Delta_k$,
   return $p_{tr} = p_k^{ST}$;
5. Set $j = j + 1$ and go to Step 2.

We conclude this section making some comments on Algorithm 2.1. We set an upper bound $\eta_{max} < 1$ on the forcing term $\eta_k$ so that the sequence $\{\eta_k\}$ is uniformly less than 1. This is a basic requirement for the inexact Newton framework [8]. Moreover, we note that the condition (5) is not necessary in trust-region CG framework for unconstrained optimization while it is needed here to handle the simple bounds.

# 3 Convergence analysis

In this section we establish global and local convergence properties of the `ITREBO` method.

Throughout the section we let $\{x_k\}$ be the sequence generated by the `ITREBO` method and, without loss of generality, we assume that for all $k$, $x_k$ is not a stationary point for the least-squares problem (BCLS), i.e.

$$\|D_k \nabla\theta_k\| \neq 0.$$

Moreover we make the following basic assumptions on the function $\Theta$ in (BCLS).

**Assumption 1** *There exists an open, bounded and convex set $L$ containing the whole sequence $\{x_k\}$ such that $L \supset \{x \in \mathbb{R}^n : \exists\, x_k \;\; s.t. \;\; \|x - x_k\| \leq r\}$, for some $r > 0$, and the Jacobian matrix $J$ is Lipschitz continuous in $L$ with Lipschitz constant $2\gamma_D$, i.e. for all $x, z \in L$*

$$\|J(x) - J(z)\| \leq 2\gamma_D \|x - z\|. \tag{14}$$

**Assumption 2** $\|\Theta'\|$ *is bounded above on $L$ and $\chi_L = \sup_{x \in L} \|J(x)\|$.*

Trivially, Assumption 1 implies that the sequence $\{x_k\}$ is bounded. Moreover the following lemma can be easily proved.

**Lemma 3.1** *Let Assumptions 1 and 2 hold. Then for all $x, z \in L$*

$$\|\Theta(x) - \Theta(z)\| \leq \chi_L \|x - z\|, \tag{15}$$
$$\|\Theta(x) - \Theta(z) - J(z)(x - z)\| \leq \gamma_D \|x - z\|^2. \tag{16}$$

Following the lines of the proof of [3, Lemma 3.2] it is easy to show that the `ITREBO` method is well-defined i.e. the $k$-th iteration of the method terminates in a finite number of trials.

The global convergence properties of the method are guaranteed by imposing the condition (5) in Step 4 of Algorithm 2.1 and they are summarized in the following theorem whose proof follows the lines of that of [16, Theorem 4.1].

**Theorem 3.1** *Let Assumptions 1 and 2 hold and $\{x_k\}$ be the sequence generated by the* `ITREBO` *method.*

*i) Every limit point of the sequence $\{x_k\}$ is a first-order stationary point for the problem (BCLS).*

*ii) If $x^*$ is a limit point of $\{x_k\}$ and $\|\Theta(x^*)\| = 0$, then all the limit points of $\{x_k\}$ are zero-residual solutions to problem (BCLS).*

*iii) If $x^*$ is a limit point of $\{x_k\}$ such that $x^* \in int(\Omega)$ and $J(x^*)$ has full row rank, then $x^*$ is such that $\|\Theta(x^*)\| = 0$.*

Under further assumptions, we are able to carry out the local convergence analysis. Let us assume that the sequence $\{x_k\}$ generated by the `ITREBO` method has a limit point $x^*$ such that $\|\Theta(x^*)\| = 0$ and that the Jacobian $J(x^*)$ is full rank. Then, under suitable choices of the sequence $\{\eta_k\}$ of the forcing terms, the use of the inexact step $p_k^I$ as an approximation to the minimum norm step $p_k^N$ yields to fast asymptotic convergence of the sequence $\{x_k\}$.

Let $\mathcal{S}$ be the set of zero-residual solutions to problem (BCLS), $d(x, \mathcal{S})$ denote the distance from the point $x$ to the set $\mathcal{S}$ and $[x]_{\mathcal{S}} \in \mathcal{S}$ be such that $\|x - [x]_{\mathcal{S}}\| = d(x, \mathcal{S})$, i.e.

$$\mathcal{S} = \{y \in \Omega : \|\Theta(y)\| = 0\}, \qquad d(x, \mathcal{S}) = \inf\{\|x - y\|, y \in \mathcal{S}\},$$
$$[x]_{\mathcal{S}} = \operatorname*{argmin}_{y \in \mathcal{S}} \; \|x - y\|. \tag{17}$$

We make the following assumption.

**Assumption 3** *The zero-residual solution set $\mathcal{S}$ of problem (BCLS) is nonempty. The sequence $\{x_k\}$ generated by the* `ITREBO` *method has a limit point $x^* \in \mathcal{S}$ and $rank(J(x^*)) = \min\{n, m\}$, i.e. $J(x^*)$ is full rank.*

It is important to remark that due to Assumption 3 and Theorem 3.1, we know that $\lim_{k \to \infty} \|\Theta_k\| = 0$.

The next lemma provides useful properties that are a direct consequence of Assumptions 1-3. In particular, local properties of the Jacobian $J(x)$ and its pseudoinverse $J(x)^+$, if $x$ is sufficiently close to $x^*$, are stated.

**Lemma 3.2** *Let Assumptions 1-3 hold. Then there exists a positive constant $\tau > 0$ such that if $x \in B_\tau(x^*)$*

$$x \in L \;\; and \;\; [x]_{\mathcal{S}} \in L, \tag{18}$$
$$\|\Theta(x)\| \leq \chi_L \; d(x, \mathcal{S}). \tag{19}$$

*Moreover, there exists a constant $\nu > 0$ such that if $x \in B_\tau(x^*)$ then*

$$rank(J(x)) = \min\{n, m\}, \tag{20}$$
$$\|J(x)^+\| \le \nu. \tag{21}$$

*Proof.* See [17, Lemma 3.1]. □

The following lemma shows an important feature of overdetermined (BCLS) problems. If $J(x^*)$ is full rank, then $\|\Theta\|$ is guaranteed to provide a local error bound on some neighbourhood of $x^*$ and $x^*$ is an isolated zero-residual solution to (BCLS).

**Lemma 3.3** *Let Assumptions 1-3 hold. If $m \ge n$, then there exist positive constants $\alpha_0$ and $\omega$ such that if $x \in B_\omega(x^*)$ then*

$$\frac{1}{\alpha_0}\, d(x, S) = \frac{1}{\alpha_0}\, \|x - x^*\| \le \|\Theta(x)\|. \tag{22}$$

*Proof.* Let $x \in B_\tau(x^*)$ where $\tau$ is the scalar given in Lemma 3.2. Since $J(x^*)$ is full column rank, then $J(x^*)^+ = (J(x^*)^T J(x^*))^{-1} J(x^*)^T$ and $J(x^*)^+\, J(x^*) = I_n$. Also, by (14) we get

$$\|I_n - J(x^*)^+ J(x)\| \le \|J(x^*)^+\|\, \|J(x^*) - J(x)\| \le 2\gamma_D \|J(x^*)^+\|\, \|x - x^*\|.$$

Choosing $\omega < \min\{\tau,\ 1/(4\gamma_D\|J(x^*)^+\|)\}$, we have $\|I_n - J(x^*)^+ J(x)\| \le 1/2$ for $x \in B_\omega(x^*)$. Then, by using the Mean Value Theorem we obtain

$$\begin{aligned}\|J(x^*)^+\Theta(x)\| &= \|(x - x^*) - \int_0^1 \left(I_n - J(x^*)^+ J(x^* + t(x - x^*))\right)(x - x^*)\, dt\| \\ &\ge \left(1 - \frac{1}{2}\right)\|x - x^*\|.\end{aligned}$$

Hence

$$\|\Theta(x)\| \ge \frac{\|J(x^*)^+\Theta(x)\|}{\|J(x^*)^+\|} \ge \frac{1}{2\|J(x^*)^+\|}\|x - x^*\|.$$

This inequality implies that $x^*$ is an isolated zero-residual solution to (BCLS) and reducing $\omega$ if necessary, we get $d(x, \mathcal{S}) = \|x - x^*\|$ for $x \in B_\omega(x^*)$. Thus (22) is obtained with $\alpha_0 = 2\|J(x^*)^+\|$. □

## 3.1 Properties of the inexact trust-region step

In this section we will show that eventually there is a simple transition from the global method with a direction $p_k$ of the form (6) to the projected step

$$\bar{p}_k^I = P_\Omega(x_k + p_k^I) - x_k.$$

In fact, eventually the trust-region constraint becomes inactive and $\bar{p}_k^I$ satisfies (5) and (7). As a consequence the sequence $\{x_k\}$ converges to $x^*$ and, for appropriate choices of $\{\eta_k\}$, the rate of convergence is q-quadratic, see the next Theorems 3.2 and 3.3.

The steps $p_k^I$ and $\bar{p}_k^I$ play a central role in the asymptotic behaviour of the sequence $\{x_k\}$. A result on the distance of a point from the zero-residual solution set $\mathcal{S}$ follows from the contractivity of the projection map $P_\Omega$. From the definition of the steps $p_k^I$ and $\bar{p}_k^I$ we have

$$\|\bar{p}_k^I\| \le \|p_k^I\| \tag{23}$$
$$\|x + \bar{p}_k^I - z\| \le \|x_k + p_k^I - z\|,\ \ z \in \Omega \tag{24}$$

and from (17) we have $d(x_k + \bar{p}_k^I, \mathcal{S}) \le \|x_k + \bar{p}_k^I - [x_k + p_k^I]_\mathcal{S}\| \le \|x_k + p_k^I - [x_k + p_k^I]_\mathcal{S}\|$, i.e.

$$d(x_k + \bar{p}_k^I, \mathcal{S}) \le d(x_k + p_k^I, \mathcal{S}). \tag{25}$$

In the following study we will use the results proved in Lemmas 3.1, 3.2 and 3.3. In particular, for subsequent references, we let $\alpha_1$ and $\alpha_2$ be constants defined as

$$\alpha_1 = \nu \chi_L, \quad \alpha_2 = \nu \gamma_D, \tag{26}$$

where $\chi_L$ and $\gamma_D$ are in Assumptions 1 and 2 and reduce the constant $\tau$ of Lemma 3.2 if necessary so that

$$\alpha_1 \alpha_2 \tau \leq 1/4. \tag{27}$$

It follows two technical lemmas.

**Lemma 3.4** *Let Assumptions 1-3 hold. Let $\tau$ be the constant given in Lemma 3.2. If $x_k \in B_\tau(x^*)$ then*

$$\|p_k^N\| \leq \nu \, \|\Theta_k\| \leq \alpha_1 \, d(x_k, \mathcal{S}). \tag{28}$$

*Proof.* From (8), (21), (19) and (26), the bound (28) easily follows. □

**Lemma 3.5** *Let Assumptions 1-3 hold. Let $\tau_1 \leq \tau$ where $\tau$ is given in Lemma 3.2 and $\tau_2 \leq \tau_1/(1+\alpha_1)$. Then if $x_k \in B_{\tau_2}(x^*)$ then*

$$x_k + p_k^I \in B_{\tau_1}(x^*), \quad x_k + p_k^I \in L, \quad [x_k + p_k^I]_{\mathcal{S}} \in L, \tag{29}$$

*and*

$$x_k + \bar{p}_k^I \in B_{\tau_1}(x^*), \quad x_k + \bar{p}_k^I \in L. \tag{30}$$

*Proof.* Note that (13) yields

$$\|x_k + p_k^I - x^*\| \leq \|x_k - x^*\| + \|p_k^I\| \leq \|x_k - x^*\| + \|p_k^N\|.$$

Then, by (28) we have $\|x_k + p_k^I - x^*\| \leq (1+\alpha_1)\tau_2 \leq \tau_1$. Hence, by Lemma 3.2 the statements in (29) are proved. Similarly, using the contractivity (23), (30) holds. □

The next lemma establishes that if $x_k$ is sufficiently close to $x^*$, then the trust-region is inactive and the inexact step $p_k^I$ is taken as the approximate trust-region step.

**Lemma 3.6** *Let Assumptions 1-3 hold. Then there exists $\varsigma > 0$ such that if $x_k \in B_\varsigma(x^*)$ then the trust-region solution $p_{tr}$ computed by Algorithm 2.2 is the step $p_k^I$ given in (11).*

*Proof.* Let $\tau > 0$ be given in Lemma 3.2 and let $x_k \in B_\tau(x^*)$. Since $x^* \in \mathcal{S}$ and (28) holds, there exists a scalar $\varsigma \leq \tau$ sufficiently small so that if $x_k \in B_\varsigma(x^*)$ then $\|p_k^N\| \leq \Delta_{min}$. Namely, the unconstrained minimum norm minimizer of the quadratic model $m_k$ lies in the trust-region. Therefore, by the property (13), Algorithm 2.2 returns the step $p_k^I$. □

The analysis of the steps $p_k^I$ and $\bar{p}_k^I$ is the subject of the next results. Lemma 3.7 concerns the case $m \geq n$, while Lemma 3.8 refers to the case $m \leq n$.

**Lemma 3.7** *Let $m \geq n$ and let Assumption 1-3 hold. Let $\alpha_1$ and $\alpha_2$ be the constants defined in (26). Then, there exists a positive constant $\rho^o$, such that if $x_k \in B_{\rho^o}(x^*)$*

$$\|x_k + \bar{p}_k^I - x^*\| \leq \|x_k + p_k^I - x^*\| \leq \phi_k^o \, \|x_k - x^*\|, \tag{31}$$

*where*

$$\phi_k^o = \alpha_0(\gamma_D(\alpha_1^2 + 1)\|x_k - x^*\| + \alpha_1 \chi_L \eta_k). \tag{32}$$

*Proof.* Let $\omega$ and $\tau$ as in Lemma 3.3 and Lemma 3.2 respectively. Fix $x_k \in B_{\rho^o}(x^*)$, where

$$\rho^o < \min\{\omega, \tau\}/(1+\alpha_1).$$

Then by Lemma 3.5 and using (29) with $\tau_1 = \min\{\omega, \tau\}$ we obtain

$$x_k + p_k^I \in B_\tau(x^*), \quad x_k + p_k^I \in B_\omega(x^*), \quad x_k + p_k^I \in L, \quad [x_k + p_k^I]_{\mathcal{S}} \in L. \tag{33}$$

By condition (22) we get

$$\|x_k + p_k^I - x^*\| \le \alpha_0 \|\Theta(x_k + p_k^I)\|, \tag{34}$$

so we need to estimate $\|\Theta(x_k + p_k^I)\|$ to prove (31). By (16) we get

$$\begin{aligned} \|\Theta(x_k + p_k^I)\| &\le \|\Theta(x_k + p_k^I) - \Theta_k - J_k p_k^I\| + \|\Theta_k + J_k p_k^I\| \\ &\le \gamma_D \|p_k^I\|^2 + \|\Theta_k + J_k p_k^I\|. \end{aligned} \tag{35}$$

Consider the SVD decomposition of $J_k$. By Lemma 3.2, $J_k$ is full rank. Hence, let $J_k = U_k \Sigma_k V_k^T = (U_{k,1}, U_{k,2})\Sigma_k V_k^T$ where $U_{k,1} \in \mathbb{R}^{m\times n}$, $U_{k,2} \in \mathbb{R}^{m\times(m-n)}$, $V_k \in \mathbb{R}^{n\times n}$ $\Sigma_k \in \mathbb{R}^{m\times n}$, $\Sigma_k = \operatorname{diag}(\varsigma_1, \dots, \varsigma_n)$, $\varsigma_i > 0$ for all $i = 1, \dots, n$. Then we have that

$$U_{k,1}^T = U_{k,1}^T (J_k^T)^+ J_k^T,$$

because $(J_k^T)^+ J_k^T$ is the orthogonal projection onto the range of $J_k$, [23]. As a consequence we may write that

$$\|U_{k,1}^T(\Theta_k + J_k p_k^I)\| = \|U_{k,1}^T (J_k^T)^+ J_k^T(\Theta_k + J_k p_k^I)\|.$$

If we use (21), (12) and Assumption 2 we obtain

$$\begin{aligned} \|U_{k,1}^T(\Theta_k + J_k p_k^I)\| &\le \|J_k^+\| \|J_k^T(\Theta_k + J_k p_k^I)\| \\ &\le \alpha_1 \ \eta_k \|\Theta_k\|. \end{aligned} \tag{36}$$

Moreover we verify easily that $U_{k,2}^T J_k = 0$ and so

$$\|U_{k,2}^T(\Theta_k + J_k p_k^I)\| = \|U_{k,2}^T \Theta_k\| = \|U_{k,2} U_{k,2}^T \Theta_k\|,$$

where the last equality follows from $U_{k,2}^T U_{k,2} = I_{m-n}$. Moreover the equality $I_m = U_k U_k^T$ yields

$$U_{k,2} U_{k,2}^T \Theta_k = (I_m - U_{k,1} U_{k,1}^T)\Theta_k$$

and by $J_k p_k^N = -J_k J_k^+ \Theta_k = -U_{k,1} U_{k,1}^T \Theta_k$ we get

$$U_{k,2} U_{k,2}^T \Theta_k = \Theta_k + J_k p_k^N. \tag{37}$$

Since $p_k^N$ is the global minimizer of $\|\Theta_k + J_k p\|$ we obtain from (37) that

$$\|U_{k,2}^T(\Theta_k + J_k p_k^I)\| = \|\Theta_k + J_k p_k^N\| \le \|\Theta_k + J_k(x_k - x^*)\|.$$

From (16) we get

$$\|U_{k,2}^T(\Theta_k + J_k p_k^I)\| \le \gamma_D \|x_k - x^*\|^2. \tag{38}$$

Combining together $\|\Theta_k + J_k p_k^I\| = \|U_k^T(\Theta_k + J_k p_k^I)\|$, (36) and (38) we find that

$$\|\Theta_k + J_k p_k^I\| \le \|U_{k,1}^T(\Theta_k + J_k p_k^I)\| + \|U_{k,2}^T(\Theta_k + J_k p_k^I)\| \le \alpha_1 \ \eta_k \|\Theta_k\| + \gamma_D \|x_k - x^*\|^2. \tag{39}$$

By (35), (13), (28), (39), (19) and (22) we obtain

$$\|\Theta(x_k + p_k^I)\| \le (\gamma_D(\alpha_1^2 + 1)\|x_k - x^*\| + \alpha_1 \chi_L \eta_k) \ \|x_k - x^*\|.$$

Hence (34) and (24) give (31). □

The same result holds if $m \le n$ and the proof is essentially the one of [17, Lemma 3.2].

**Lemma 3.8** *Let $m \le n$ and let Assumption 1-3 hold. Let $\alpha_1, \alpha_2$ and $\nu$ be the constants defined in (26) and (21) respectively. Then there exists a positive constant $\rho^u$, such that if $x_k \in B_{\rho^u}(x^*)$ and if $\eta_k \le \min\{\eta_{max}, 1/(4\alpha_1)\}$, then*

$$d(x_k + p_k^I, S) \le 2\nu(\nu\alpha_2\|\Theta_k\| + \alpha_1\eta_k)\ \|\Theta_k\|, \tag{40}$$

*and*

$$d(x_k + \bar{p}_k^I, S) \le \phi_k^u\ d(x_k, \mathcal{S}), \tag{41}$$

*where*

$$\phi_k^u = 2\alpha_1^2(\alpha_2 d(x_k, \mathcal{S}) + \eta_k). \tag{42}$$

*Proof.* Let $\tau$ as in Lemma 3.2 and such that (27) holds. Fix $x_k \in B_{\rho^u}(x^*)$, where

$$\rho^u < \tau/(1 + 2\alpha_1). \tag{43}$$

Then by Lemma 3.5, we get

$$x_k + p_k^I \in B_\tau(x^*), \quad x_k + p_k^I \in L, \quad [x_k + p_k^I]_{\mathcal{S}} \in L. \tag{44}$$

To prove the thesis we need intermediate results. Consider the sequence $\{w_{k+l}\}_l$, $l \ge 0$, of the form

$$w_k = x_k, \quad w_{k+l+1} = w_{k+l} + s_{k+l}^I, \quad l \ge 0, \tag{45}$$

where $s_{k+l}^I$ is computed by applying CG method to the linear system

$$J(w_{k+l})^T J(w_{k+l}) s = -J(w_{k+l})^T\, \Theta(w_{k+l}).$$

Specifically, starting from the null initial guess, the step $s_{k+l}^I$ is the first CG iterate such that

$$\|\tilde{r}_{k+l}\| \le \tilde{\eta}_{k+l}\|J(w_{k+l})^T \Theta(w_{k+l})\|,\ l \ge 0,$$

where $\tilde{r}_{k+l}$ is given by

$$\tilde{r}_{k+l} = J(w_{k+l})^T J(w_{k+l}) s_{k+l}^I + J(w_{k+l})^T\, \Theta(w_{k+l}),\ l \ge 0, \tag{46}$$

and $\{\tilde{\eta}_{k+l}\}_{l \ge 0}$ is a sequence of positive scalars such that $\tilde{\eta}_k = \eta_k$ and $\sup_{j \ge 0} \tilde{\eta}_{k+j} \le 1/(4\alpha_1)$. Note that for $l = 0$, we get $s_k^I = p_k^I$. Letting

$$s_{k+l}^N = -J(w_{k+l})^+\, \Theta(w_{k+l}),\ l \ge 0. \tag{47}$$

we have

$$\|s_{k+l}^I\| \le \|s_{k+l}^N\|,\ l \ge 0. \tag{48}$$

First, we show that $\{w_{k+l}\}_{l \ge 0} \subseteq B_\tau(x^*)$. Second, we prove that $\{w_{k+l}\}_{l \ge 0}$ has limit point in $\mathcal{S}$. We begin proving that $\{w_{k+l}\} \subseteq B_\tau(x^*)$ by induction. The thesis trivially holds for $w_k = x_k$. Then, we suppose that $w_{k+j} \in B_\tau(x^*)$ for $j \le l$ and show that $w_{k+l+1} \in B_\tau(x^*)$. By (47), (48), (21), (16) and Lemma 3.2 we get

$$\begin{aligned} \|s_{k+j}^I\| \le \|s_{k+j}^N\| &\le \nu\|\Theta(w_{k+j})\|, & (49)\\ \|\Theta(w_{k+j}) - \Theta(w_{k+j-1}) - J(w_{k+j-1}) s_{k+j-1}^I\| &\le \gamma_D \|s_{k+j-1}^I\|^2, & (50)\\ \|\tilde{r}_{k+j}\| &\le \chi_L\, \tilde{\eta}_{k+j}\|\Theta(w_{k+j})\|, & (51) \end{aligned}$$

for $j = 1, \ldots, l$. Moreover, from (46), (47), (21) and $(J_k^T)^+ J_k^T = J_k {J_k}^+ = I_m$ it follows

$$\|J(w_{k+j-1})(s_{k+j-1}^N - s_{k+j-1}^I)\| \le \nu\|\tilde{r}_{k+j-1}\|, \quad 1 \le j \le l.$$

Hence from (51) and (19)

$$\begin{aligned}
\|\Theta(w_{k+j})\| &= \|\Theta(w_{k+j}) - \Theta(w_{k+j-1}) - J(w_{k+j-1})s^N_{k+j-1} \pm J(w_{k+j-1})s^I_{k+j-1}\| \\
&\le \gamma_D \|s^I_{k+j-1}\|^2 + \nu\|\tilde{r}_{k+j-1}\| \\
&\le (\nu\alpha_2\|\Theta(w_{k+j-1})\| + \alpha_1\tilde{\eta}_{k+j-1})\ \|\Theta(w_{k+j-1})\| \\
&\le (\alpha_1\alpha_2\, d(w_{k+j-1},\mathcal{S}) + \alpha_1\tilde{\eta}_{k+j-1})\ \|\Theta(w_{k+j-1})\| \\
&\le (\alpha_1\alpha_2\tau + \alpha_1\tilde{\eta}_{k+j-1})\ \|\Theta(w_{k+j-1})\| \\
&\le \frac{1}{2}\ \|\Theta(w_{k+j-1})\|,
\end{aligned} \tag{52}$$

for $j = 1,\dots,l$, since $\alpha_1\alpha_2\tau \le \frac{1}{4}$ and $\sup_{j\ge 0}\tilde{\eta}_{k+j} \le 1/(4\alpha_1)$. Then it follows that

$$\|\Theta(w_{k+j})\| \le \left(\frac{1}{2}\right)^j \|\Theta_k\|, \quad 1 \le j \le l,$$

and by (49)

$$\|s^I_{k+j}\| \le \nu\left(\frac{1}{2}\right)^j \|\Theta_k\|, \quad 1 \le j \le l. \tag{53}$$

It then follows from (53) that

$$\begin{aligned}
\|w_{k+l+1} - x^*\| &\le \sum_{j=0}^{l} \|w_{k+j+1} - w_{k+j}\| + \|x_k - x^*\| \\
&\le \sum_{j=0}^{l} \|s^I_{k+j}\| + \rho^u \le \nu\|\Theta_k\| \sum_{j=0}^{\infty}\left(\frac{1}{2}\right)^j + \rho^u,
\end{aligned}$$

and (19) and (43) yield to

$$\|w_{k+l+1} - x^*\| \quad \le \quad 2\nu\|\Theta_k\| + \rho^u \le 2\alpha_1\, d(x_k,\mathcal{S}) + \rho^u \le (2\alpha_1+1)\rho^u \le \tau.$$

As a consequence, $\{w_{k+l}\} \subset B_\tau(x^*)$ and $w_{k+l}$ satisfies Lemma 3.2 and Lemma 3.2 for all $l \ge 0$. Further, the conditions (49), (50) and (51) hold for $j \ge 1$.

Second, we prove that $\{w_{k+l}\}$ is a Cauchy sequence with limit point $\bar{x} \in \mathcal{S}$. In fact, letting $p > q \ge 0$ and proceeding as above we obtain

$$\|w_{k+p} - w_{k+q}\| \le \sum_{j=q}^{p-1}\|s^I_{k+j}\| \le \sum_{j=q}^{p-1}\|s^N_{k+j}\| \le \sum_{j=0}^{\infty}\|s^N_{k+j}\| \le 2\alpha_1\rho^u.$$

Thus, $\{w_{k+l}\}$ is a Cauchy sequence and the limit is denoted as $\bar{x}$. To show that $\bar{x} \in \mathcal{S}$ note that (46), (16), (15), (45), (51) and the property $(J(w_{k+l})^T)^+ J(w_{k+l})^T = I_m$ yield

$$\begin{aligned}
\|\Theta(w_{k+l+1})\| &= \|(J(w_{k+l})^T)^+ J(w_{k+l})^T\Theta(w_{k+l+1})\| \\
&\le \|J(w_{k+l})^+\|\|J(w_{k+l})^T(\Theta(w_{k+l+1}) - J(w_{k+l})s^I_{k+l} - \Theta(w_{k+l})) + \tilde{r}_{k+l}\| \\
&\le \nu\left(\chi_L\gamma_D\|s^I_{k+l}\|^2 + \|\tilde{r}_{k+l}\|\right) \\
&\le \nu\left(\chi_L\gamma_D\|s^I_{k+l}\|^2 + \chi_L\tilde{\eta}_{k+l}\|\Theta(w_{k+l})\|\right) \\
&\le \alpha_1\gamma_D\|w_{k+l+1} - w_{k+l}\|^2 + \alpha_1\tilde{\eta}_{k+l}(\|\Theta(w_{k+l+1}) - \Theta(w_{k+l})\| + \|\Theta(w_{k+l+1})\|) \\
&\le \alpha_1\gamma_D\|w_{k+l+1} - w_{k+l}\|^2 + 1/4(\chi_L\|w_{k+l+1} - w_{k+l}\| + \|\Theta(w_{k+l+1})\|),
\end{aligned}$$

for $l \ge 0$. Hence

$$\|\Theta(w_{k+l+1})\| \quad \le \quad \frac{4}{3}\left(\alpha_1\gamma_D\|w_{k+l+1} - w_{k+l}\| + \chi_L/4\right)\|w_{k+l+1} - w_{k+l}\|,$$

for $l \ge 0$. Since $\lim_{l\to\infty}\|w_{k+l+1} - w_{k+l}\| = 0$, it follows $\|\Theta(\bar{x})\| = \lim_{l\to\infty}\|\Theta(w_{k+l+1})\| = 0$.

Now we can prove the thesis of the lemma. Note that $\|x_k + p_k^I - \bar{x}\| = \|w_{k+1} - \bar{x}\| \leq \sum_{j=1}^{\infty} \|s_{k+j}^I\|$, since

$$w_{k+l} = w_{k+1} + \sum_{j=1}^{l-1} s_{k+j}^I, \quad \lim_{l\to\infty} w_{k+l} = \bar{x},$$

and

$$\begin{aligned}\|w_{k+1} - \bar{x}\| &= \left\|w_{k+1} - \lim_{l\to\infty} w_{k+l}\right\| = \left\|\lim_{l\to\infty} \sum_{j=1}^{l-1} s_{k+j}^I\right\| \\ &\leq \lim_{l\to\infty} \left\|\sum_{j=1}^{l-1} s_{k+j}^I\right\| \leq \lim_{l\to\infty} \sum_{j=1}^{l-1} \|s_{k+j}^I\| = \sum_{j=1}^{\infty} \|s_{k+j}^I\|.\end{aligned}$$

From (49) and (53) we get

$$\|x_k + p_k^I - \bar{x}\| \leq \sum_{j=1}^{\infty} \|s_{k+j}^I\| \leq \sum_{j=1}^{\infty} \nu \left(\frac{1}{2}\right)^{j-1} \|\Theta(w_{k+1})\| = 2\nu\|\Theta(w_{k+1})\|.$$

Then, using equation (52) with $j = 1$ we obtain

$$\|x_k + p_k^I - \bar{x}\| \leq 2\nu(\nu\alpha_2\|\Theta_k\| + \alpha_1\eta_k)\|\Theta_k\|. \tag{54}$$

Since $d(x_k + p_k^I, \mathcal{S}) \leq \|x_k + p_k^I - \bar{x}\|$, (40) holds. Finally, applying (25), (40) and (19) we easily obtain condition (41). $\square$

The next lemma gives useful asymptotic bounds on quantities that are used in the proofs of Lemma 3.10 and Lemma 3.11.

**Lemma 3.9** *Let Assumptions 1-3 hold. Let $\alpha_1, \alpha_2$ and $\nu$ be the constants defined in (26) and (21) respectively. Then there exists a constant $\hat{\tau} > 0$ such that if $x_k \in B_{\hat{\tau}}(x^*)$ then*

$$\|J_k\bar{p}_k^I + \Theta_k\| \leq \chi_L d(x_k + p_k^I, \mathcal{S}) + \nu\alpha_2\|\Theta_k\|^2, \tag{55}$$

$$\|\Theta(x_k + \bar{p}_k^I)\|^2 - \|J_k\bar{p}_k^I + \Theta_k\|^2 \leq \left(\nu^2\alpha_2^2\|\Theta_k\|^2 + 2\nu\alpha_2\|J_k\bar{p}_k^I + \Theta_k\|\right)\|\Theta_k\|^2. \tag{56}$$

*Proof.* Let $\tau$ as in Lemma 3.2. Fix $x_k \in B_{\hat{\tau}}(x^*)$, where $\hat{\tau} < \tau/(1+\alpha_1)$. Then by Lemma 3.5, (29) and (30) hold with $\tau_1 = \tau$.

Consider the equality

$$J_k\bar{p}_k^I + \Theta_k = \Theta(x_k + \bar{p}_k^I) - \Theta([x_k + p_k^I]_{\mathcal{S}}) + J_k\bar{p}_k^I - (\Theta(x_k + \bar{p}_k^I) - \Theta_k).$$

Then by (15), (16), (24), (23), (13) and (28) we obtain

$$\begin{aligned}\|J_k\bar{p}_k^I + \Theta_k\| &\leq \chi_L\|x_k + \bar{p}_k^I - [x_k + p_k^I]_{\mathcal{S}}\| + \gamma_D\|\bar{p}_k^I\|^2 \\ &\leq \chi_L\|x_k + p_k^I - [x_k + p_k^I]_{\mathcal{S}}\| + \gamma_D\|p_k^N\|^2 \\ &\leq \chi_L d(x_k + p_k^I, \mathcal{S}) + \nu\alpha_2\|\Theta_k\|^2,\end{aligned}$$

and (55) is proved.

To prove (56) we use the Mean Value Theorem to get the statement

$$\Theta(x_k + \bar{p}_k^I) = \Theta_k + \int_0^1 J(x_k + t\bar{p}_k^I)\,\bar{p}_k^I\,dt + J_k\bar{p}_k^I - J_k\bar{p}_k^I.$$

Hence,

$$\begin{aligned}\|\Theta(x_k + \bar{p}_k^I)\|^2 &= \|J_k\bar{p}_k^I + \Theta_k\|^2 + \|\int_0^1 (J(x_k + t\bar{p}_k^I) - J_k)\,\bar{p}_k^I\,dt\|^2 \\ &\quad +2\left(\int_0^1 (J(x_k + t\bar{p}_k^I) - J_k)\,\bar{p}_k^I\,dt\right)^T \left(J_k\bar{p}_k^I + \Theta_k\right),\end{aligned}$$

and consequently by (28)

$$\begin{aligned}\|\Theta(x_k+\bar{p}_k^I)\|^2-\|J_k\bar{p}_k^I+\Theta_k\|^2 &\leq \gamma_D^2\|\bar{p}_k^I\|^4+2\gamma_D\|J_k\bar{p}_k^I+\Theta_k\|\,\|\bar{p}_k^I\|^2\\ &\leq \nu^2\alpha_2^2\|\Theta_k\|^4+2\nu\alpha_2\|J_k\bar{p}_k^I+\Theta_k\|\,\|\Theta_k\|^2.\end{aligned}$$

□

## 3.2 Behaviour of the sequence $\{x_k\}$ and convergence rate

In this section we first show that, under proper assumptions on the forcing sequence $\{\eta_k\}$, the step $p_k^I$ satisfies conditions (5) and (7) whenever $x_k$ is sufficiently close to $x^*$ and $k$ is sufficiently large. Then, we prove that the whole sequence $\{x_k\}$ generated by the `ITREBO` method converges to $x^*\in\mathcal{S}$ and the convergence rate is q-quadratic.

The following Lemma 3.10 and Theorem 3.2 concern the overdetermined case; Lemma 3.11 and Theorem 3.3 refer to the underdetermined case.

**Lemma 3.10** *Let $m\geq n$ and let Assumptions 1-3 hold. Assume that $\lim_{k\to\infty}\eta_k=0$. Then $\bar{p}_k^I$ satisfies conditions (5) and (7) whenever $x_k$ is sufficiently close to $x^*$ and $k$ is sufficiently large.*

*Proof.* Let $\psi^o\leq\min\{\rho^o,\hat{\tau}\}$, where $\rho^o$ and $\hat{\tau}$ are given in Lemma 3.7 and Lemma 3.9 respectively and fix $x_k\in B_{\psi^o}(x^*)$.

Note that $m_k(0)-m_k(p_k^C)\leq m_k(0)$ and

$$\rho_c(\bar{p}_k^I)\geq 1-\frac{\|J_k\bar{p}_k^I+\Theta_k\|^2}{\|\Theta_k\|^2}. \tag{57}$$

Since $d(x_k+p_k^I,\mathcal{S})\leq\|x_k+p_k^I-x^*\|$, (55), (31), (19) and (22) yield

$$\begin{aligned}\|J_k\bar{p}_k^I+\Theta_k\| &\leq \chi_L\phi_k^o\|x_k-x^*\|+\nu\alpha_2\|\Theta_k\|^2\\ &\leq \sigma_k^o\|x_k-x^*\|,\end{aligned} \tag{58}$$

where $\phi_k^o$ is given in (32) and $\sigma_k^o$ is defined as

$$\sigma_k^o=\chi_L\phi_k^o+\alpha_1\alpha_2\|x_k-x^*\|. \tag{59}$$

Thus, (57), (58) and (22) give

$$\rho_c(\bar{p}_k^I)\geq 1-\left(\frac{\sigma_k^o}{\alpha_0}\right)^2.$$

Since $\lim_{k\to\infty}\eta_k=0$, if $x_k$ is sufficiently close to $x^*$ and $k$ is sufficiently large we obtain that $\bar{p}_k^I$ satisfies condition (5).

Let $p_k^I$ satisfy (5). To prove that $\bar{p}_k^I$ satisfies (7), observe that $m_k(0)=\|\Theta_k\|^2/2$, $m_k(p_k^I)<m_k(0)$ and

$$\rho_\theta(\bar{p}_k^I)=1-\frac{\|\Theta(x_k+\bar{p}_k^I)\|^2-\|J_k\bar{p}_k^I+\Theta_k\|^2}{\|\Theta_k\|^2-\|J_k\bar{p}_k^I+\Theta_k\|^2}. \tag{60}$$

From (22) and (58) we have

$$\|\Theta_k\|^2-\|J_k\bar{p}_k^I+\Theta_k\|^2 \geq \left(\frac{1}{\alpha_0^2}-(\sigma_k^o)^2\right)\|x_k-x^*\|^2. \tag{61}$$

Using (56), (61), (58), (19), (22) we have

$$\begin{aligned}\rho_\theta(\bar{p}_k^I) &\geq 1-\frac{(\nu^2\alpha_2^2\|\Theta_k\|^2+2\nu\alpha_2\|J_k\bar{p}_k^I+\Theta_k\|)\,\|\Theta_k\|^2}{\left(\frac{1}{\alpha_0^2}-(\sigma_k^o)^2\right)\|x_k-x^*\|^2},\\ &\geq 1-\chi_L^2\frac{\alpha_1^2\alpha_2^2\|x_k-x^*\|^2+2\nu\alpha_2\sigma_k^o\|x_k-x^*\|}{\frac{1}{\alpha_0^2}-(\sigma_k^o)^2}.\end{aligned} \tag{62}$$

Then, if $x_k$ is sufficiently close to the solution $x^*$ and $k$ sufficiently large, the second term in (62) can be made less than $(1-\beta_2)$ and $\bar{p}_k^I$ satisfies condition (7). □

Next theorem provides the main result on the convergence rate of the sequence generated by the `ITREBO` method for the case $m \geq n$.

**Theorem 3.2** *Let $m \geq n$ and let Assumptions 1-3 hold. Then the sequence $\{x_k\}$ generated by the* `ITREBO` *method converges to $x^*$ q-superlinearly if $\lim_{k\to\infty} \eta_k = 0$. Moreover, the convergence rate is q-quadratic if $\eta_k = O(\|\Theta_k\|)$.*

*Proof.* Let $\{x_{k_j}\}$ be a subsequence of $\{x_k\}$ converging to $x^*$. By Lemmas 3.6 and 3.10 if $x_{k_j}$ is sufficiently close to $x^*$ and for $k_j$ sufficiently large, then the step taken is equal to $\bar{p}_{k_j}^I$ and by (23), (13) and (28) $\lim_{j\to\infty} \bar{p}_{k_j}^I = 0$. Then, since $x^*$ is an isolated limit point of $\{x_k\}$, using [18, Lemma 4.10], we conclude that $\lim_{k\to\infty} x_k = x^*$.

To establish the convergence rate of $\{x_k\}$, let $x_k$ sufficiently near to $x^*$ and $k$ sufficiently large so that $x_{k+1} = x_k + \bar{p}_k^I$ and Lemma 3.7 holds. Then from (31)

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \leq \phi_k^o,$$

where $\phi_k^o$ is defined in (32). Since $\lim_{k+\infty} \phi_k^o = 0$, $x_k$ converges to $x^*$ q-superlinearly. Moreover, if $\eta_k = O(\|\Theta_k\|)$, then $\eta_k = O(\|x_k - x^*\|)$ and it follows $\|x_{k+1} - x^*\| = O(\|x_k - x^*\|^2)$, i.e. the q-quadratic rate is guaranteed. □

The main local convergence properties of the `ITREBO` method for the underdetermined case are stated below. Firstly, using the findings in Lemma 3.9 and proceeding as in Lemma 3.10, we get the subsequent lemma.

**Lemma 3.11** *Let $m \leq n$ and let Assumptions 1-3 hold. Assume that $\lim_{k\to\infty} \eta_k = 0$. Then $\bar{p}_k^I$ satisfies conditions (5) and (7) whenever $x_k$ is sufficiently close to $x^*$ and $k$ is sufficiently large.*

*Proof.* Let $\psi^u \leq \min\{\rho^u, \hat{\tau}\}$, where $\rho^u$ and $\hat{\tau}$ are given in Lemma 3.8 and Lemma 3.9 respectively and let $k$ sufficiently large that $\eta_k \leq \min\{\eta_{max}, 1/(4\alpha_1)\}$. Fix $x_k \in B_{\psi^u}(x^*)$. Note that since $x^*$ is a limit point of $\{x_k\}$ and $\lim_{k\to\infty} \eta_k = 0$, there exists an iterate $x_k$ and a forcing term $\eta_k$ satisfying the above conditions.

First we prove that $\bar{p}_k^I$ satisfies (5). By (55), (40) and (19) we obtain

$$\|J_k \bar{p}_k^I + \Theta_k\| \leq \sigma_k^u \|\Theta_k\|, \tag{63}$$

where

$$\sigma_k^u = \alpha_1\alpha_2(2\alpha_1 + 1)d(x_k, \mathcal{S}) + 2\alpha_1^2\eta_k. \tag{64}$$

Thus, (57), (63) and (19) give

$$\rho_c(\bar{p}_k^I) \geq 1 - (\sigma_k^u)^2,$$

Then, $\bar{p}_k^I$ satisfies condition (5) if $x_k$ is sufficiently close to $x^*$ and $k$ is sufficiently large so that $1-(\sigma_k^u)^2 > \beta_1$.

Second we prove that $\bar{p}_k^I$ satisfies (7). Let $\bar{p}_k^I$ satisfy (5). From (63), (19) we have

$$\|\Theta_k\|^2 - \|J_k\bar{p}_k^I + \Theta_k\|^2 \quad \geq \quad \left(1 - (\sigma_k^u)^2\right)\|\Theta_k\|^2. \tag{65}$$

Using (60), (56) and (65) we obtain

$$\begin{aligned} \rho_\theta(\bar{p}_k^I) \quad &\geq \quad 1 - \frac{\nu^2\alpha_2^2\|\Theta_k\|^2 + 2\nu\alpha_2\|J_k\bar{p}_k^I + \Theta_k\|}{1-(\sigma_k^u)^2}, \\ &\geq \quad 1 - \frac{\nu^2\alpha_2^2\|\Theta_k\|^2 + 2\nu\alpha_2\sigma_k^u\|\Theta_k\|}{1-(\sigma_k^u)^2}, \\ &\geq \quad 1 - \frac{\alpha_1\alpha_2^2 d(x_k,\mathcal{S})^2 + 2\alpha_1\alpha_2\sigma_k^u d(x_k,\mathcal{S})}{1-(\sigma_k^u)^2}. \end{aligned} \tag{66}$$

Then, if $x_k$ is sufficiently close to the solution $x^*$ and $k$ is sufficiently large, the second term in (66) can be made less than $(1-\beta_2)$, hence $\bar{p}_k^I$ satisfies condition (7). □

Secondly, the behaviour of the sequence $\{x_k\}$ generated by the `ITREBO` method is given in the next theorem whose proof parallels that of [17, Theorem 3.1].

**Theorem 3.3** *Let $m \le n$ and let Assumptions 1-3 hold. Then the sequence $\{x_k\}$ generated by the `ITREBO` method converges to $x^*$ q-superlinearly if $\lim_{k\to\infty} \eta_k = 0$. Moreover, the convergence rate is q-quadratic if $\eta_k = O(\|\Theta_k\|)$.*

*Proof.* Let $\hat{k}$ be sufficiently large so that $\phi^u$ in (42) satisfies $\phi_k^u \le \frac{1}{2}$ and $\eta_k \le \min\{\eta_{max}, 1/(4\alpha_1)\}$ for $k \ge \hat{k}$. Let $\psi_2 \le \min\{\varsigma, \rho^u\}$, where $\varsigma$ and $\rho^u$ are given in Lemma 3.6 and Lemma 3.8 respectively, and $\bar{k} \ge \hat{k}$ be such that if $x_k \in B_{\psi_2}(x^*)$ and $k \ge \bar{k}$ then $\bar{p}_k^I$ satisfies (5) and (7). Finally, let $\zeta < \frac{\psi_2}{1+2\alpha_1}$. Fix $k \ge \bar{k}$ and $x_k \in B_\zeta(x^*)$.

We begin showing that if $x_k \in B_\zeta(x^*)$ then $x_l \in B_{\psi_2}(x^*)$ for $l > k$. We proceed by induction. First, we show that $x_{k+1} \in B_{\psi_2}(x^*)$. In fact, by (23) we have $\|x_{k+1} - x^*\| = \|x_k + \bar{p}_k^I - x^*\| \le \zeta + \|p_k^I\|$. Thus by (13) and (28) and the definition of $\zeta$, we get $\|x_{k+1} - x^*\| \le (1+\alpha_1)\zeta \le \psi_2$. Second, we assume $x_{k+1}, \ldots, x_{k+m-1} \in B_{\psi_2}(x^*)$, and show that $x_{k+m} \in B_{\psi_2}(x^*)$. From (41) it follows

$$d(x_{k+l}, \mathcal{S}) \le \phi_{k+l-1}^u \; d(x_{k+l-1}, \mathcal{S}) \le \cdots \le \left(\frac{1}{2}\right)^{l-1} \phi_k^u d(x_k, \mathcal{S}) \le \; \zeta\left(\frac{1}{2}\right)^l,$$

for $l = 1, \ldots, m$. Thus,

$$\begin{aligned} \|x_{k+m} - x^*\| \;&\le\; \|x_{k+m} - x_{k+m-1}\| + \cdots + \|x_k - x^*\| \\ &\le\; \sum_{l=0}^{m-1} \|\bar{p}_k^I\| + \zeta \le \alpha_1 \sum_{l=0}^{m-1} d(x_{k+l}, \mathcal{S}) + \zeta, \end{aligned}$$

where the last inequality follows from (13) and (28), and

$$\|x_{k+m} - x^*\| \;\le\; (\alpha_1 \sum_{l=0}^{m-1} \left(\frac{1}{2}\right)^l + 1)\zeta \le (\alpha_1 \sum_{l=0}^{\infty} \left(\frac{1}{2}\right)^l + 1)\zeta = (2\,\alpha_1 + 1)\zeta \le \psi_2,$$

where the last inequality is due to the choice of $\zeta$. Note that we have $x_{k+l} = x_{k+l-1} + \bar{p}_{k+l-1}^I$ for $l > 0$. Moreover, letting $p > q \ge k$ we have

$$\|x_p - x_q\| \;\le\; \sum_{l=q}^{p-1} \|\bar{p}_l^I\| \le \alpha_1 \sum_{l=0}^{\infty} \left(\frac{1}{2}\right)^l \zeta = 2\,\alpha_1\zeta.$$

This means that $\{x_k\}$ is a Cauchy sequence and hence it converges. Since $x^*$ is a limit point we conclude $\lim_{k\to\infty} x_k = x^*$. To establish the convergence rate of $\{x_k\}$, let $k \ge \bar{k}$ sufficiently large so that $x_{k+j+1} \in B_{\psi_2}(x^*)$ for $j \ge 0$. By (28) and (41) we obtain

$$\|\bar{p}_{k+j+1}^I\| \le \|\bar{p}_{k+j+1}^N\| \le \alpha_1 \; d(x_{k+j+1}, \mathcal{S}) \le \alpha_1 \phi_{k+j}^u \; d(x_{k+j}, \mathcal{S}) \le \frac{1}{2}\alpha_1 \; d(x_{k+j}, \mathcal{S}).$$

Then, we proceed as above and using $\|x_{k+1} - x^*\| \le \sum_{j=0}^{\infty} \|\bar{p}_{k+j+1}^I\|$, and (41) we get

$$\begin{aligned} \|x_{k+1} - x^*\| \;&\le\; \alpha_1 \sum_{j=0}^{\infty} d(x_{k+j+1}, \mathcal{S}) \le \alpha_1 \sum_{j=0}^{\infty} \left(\frac{1}{2}\right)^j d(x_{k+1}, \mathcal{S}) \\ &\le\; 2\alpha_1 \phi_k^u \, d(x_k, \mathcal{S}). \end{aligned}$$

Hence

$$\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} \le \frac{2\alpha_1 \phi_k^u \, d(x_k, \mathcal{S})}{d(x_k, \mathcal{S})} = 2\alpha_1 \phi_k^u.$$

Since $\lim_{k+\infty} \phi_k^u = 0$, $x_k$ converges to $x^*$ q-superlinearly. If moreover $\eta_k = O(\|\Theta_k\|)$, then $\eta_k = O(d(x_k, \mathcal{S}))$ by (19) and hence for $k$ sufficiently large there exists $\bar{\phi} > 0$ such that $\phi_k^u \le \bar{\phi}\, d(x_k, \mathcal{S})$, i.e.

$$\|x_{k+1} - x^*\| \le 2\alpha_1 \phi_k^u\, d(x_k, \mathcal{S}) \le 2\alpha_1 \bar{\phi}\, d(x_k, \mathcal{S})^2 \le 2\alpha_1 \bar{\phi}\, \|x_k - x^*\|^2,$$

and then the q-quadratic rate is guaranteed. □

## Acknowledgements

The author is grateful to Benedetta Morini for the several helpful discussions and the many suggestions on the manuscript.

# References

[1] N.M. Alsaifi, P. Englezos, Prediction of multiphase equilibrium using the PC-SAFT equation of state and simultaneous testing of phase stability, Fluid Phase Equilibr. (2010), doi:10.1016/j.fluid.2010.09.002.

[2] S. Bellavia, M.Macconi, B. Morini, An affine scaling trust-region approach to bound-constrained nonlinear systems, Appl. Numer. Math., 44 (2003), pp. 257-280.

[3] S. Bellavia, B. Morini, An interior global method for nonlinear systems with simple bounds, Optim. Methods Softw., 20 (2005), pp. 1-22.

[4] S. Bellavia, B. Morini, Subspace trust-region methods for large bound-constrained nonlinear equations, SIAM J. Numer. Anal., 44 (2006), pp. 1535-1555.

[5] L. Bencini, R. Fantacci, L. Maccari, Analytical model for performance analysis of IEEE 802.11 DCF mechanism in multi-radio wireless networks, in Proc. of ICC 2010, pp. 1-5, Cape Town, South Africa, May 2010.

[6] M.A. Branch, T.F. Coleman, Y. Li, A subspace, interior and conjugate gradient method for large-scale bound-constrained minimization problems, SIAM J. Sci. Comput., 21 (1999), pp. 1-23.

[7] T.F. Coleman, Y. Li, An interior trust-region approach for nonlinear minimization subject to bounds, SIAM J. Optim., 6 (1996), pp. 418-445.

[8] R.S. Dembo, S.C. Eisenstat, T. Steihaug, Inexact Newton methods, SIAM J. Numer. Anal., 19 (1982), pp. 400-408.

[9] J.B. Francisco, N. Krejić, J.M. Martínez, An interior-point method for solving box-constrained underdetermined nonlinear systems, J. Comput. Appl. Math., 177 (2005), pp. 67-88.

[10] M.R. Hestenes, E. Stiefel, Methods of Conjugate Gradients for solving linear systems, J. Res. Natl. Bur. Stand., 49 (1952), pp. 409-436.

[11] M.R. Hestenes, Pseudoinverses and conjugate gradients, Communications of the ACM, 18 (1975), pp. 40-43.

[12] R.A. Horn, C.R. Johnson, Matrix analysis, The Cambridge University Press, 1985.

[13] M. Kaiser, K. Klamroth, A. Thekale and Ph. L. Toint, Solving structured nonlinear least-squares and nonlinear feasibility problems with expensive functions, Report NAXYS-07-2010, Dept of Mathematics, FUNDP, Namur (B), 2010.

[14] C. Kanzow, A. Klug, An interior-point affine-scaling trust-region method for semismooth equations with box constraints, Comput. Optim. Appl., 37 (2007), pp. 329-353.

[15] C. Kanzow, S. Petra, Projected filter trust region methods for a semismooth least-squares formulation of mixed complementarity problems, Optim. Method Soft., 22 (2007), pp. 713-735.

[16] M. Macconi, B. Morini, M. Porcelli, Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities, Applied Numer. Math., 59 (2009), pp. 859-876.

[17] M. Macconi, B. Morini, M. Porcelli, A Gauss-Newton method for solving bound-constrained underdetermined nonlinear systems, Optim. Methods Softw., 24 (2009), pp. 219-235.

[18] J.J. Moré, D.C. Sorensen, Computing a trust-region step, SIAM J. Sci. Stat. Comput., 4 (1983), pp. 553-572.

[19] B. Morini, M. Porcelli, TRESNEI, a Matlab trust-region solver for systems of nonlinear equalities and inequalities, Comput. Optim. Appl. (2010) DOI: 10.1007/s10589-010-9327-5.

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

[21] Ph.L. Toint, Towards an efficient sparsity exploiting Newton method for minimization, in Sparse Matrices and Their Uses (I. S. Duff, ed.), Academic Press, London (1981), pp. 57-88.

[22] D. Zhu, Affine scaling interior LevenbergMarquardt method for bound-constrained semismooth equations under local error bound conditions, J. Comput. Appl. Math., 219 (2008), pp. 198 -215.

[23] Y. Yuan, On the truncated conjugate-gradient method. Math. Program., Ser. A, 87(3):561573, 2000.