Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities

Maria Macconi, Benedetta Morini, Margherita Porcelli

Abstract

Two trust-region methods for systems of mixed nonlinear equalities, general inequalities and simple bounds are proposed. The first method is based on a Gauss–Newton model, the second one is based on a regularized Gauss–Newton model and results to be a Levenberg–Marquardt method. The globalization strategy uses affine scaling matrices arising in bound-constrained optimization. Global convergence results are established and quadratic rate is achieved under an error bound assumption. The numerical efficiency of the new methods is experimentally studied.

Keywords: Systems of nonlinear equalities and inequalities; Trust-region methods; Bound-constrained optimization; Error bound

1. Introduction

We address the problem of solving systems of nonlinear equalities and inequalities where the inequalities are a mixture of general inequalities and simple bounds. The form of our problem is

\[ \begin{align*}
  & c_i(v) = 0, \quad i \in E, \\
  & c_i(v) \leq 0, \quad i \in I, \\
  & v_l \leq v \leq v_u,
\end{align*} \tag{1.1} \]

where the functions \( c_i : \mathbb{R}^p \to \mathbb{R} \) are continuously differentiable, \( E \cup I = \{1, \ldots, m\}, E \cap I = \emptyset \). The vectors \( v_l, v_u \) are \( p \)-dimensional and may have components equal to minus or plus infinity respectively. We assume the presence of simple constraints as it is fairly common to restrict variables to certain meaningful intervals [6,16]. Further, we assume that there exists a point satisfying (1.1).

Problems taking the form (1.1) are general and appear frequently in practice. They occur in many contexts such as the model formulation design, the detection of approximately feasible points in nonlinear programming and the “restoration” phase in filter methods for nonlinear programming problems, see e.g. [10,12,14].

© 2008 IMACS. Published by Elsevier B.V. All rights reserved.

Keywords: Systems of nonlinear equalities and inequalities; Trust-region methods; Bound-constrained optimization; Error bound

\[ \text{doi:10.1016/j.apnum.2008.03.028} \]
Taking into account the variety of applications yielding the problem (1.1), we allow any relationship between $m$ and $p$.

The problem (1.1) reduces to a bound-constrained and possibly nonsquare system of nonlinear equations when the set $I$ is empty. Recently, the solution of this problem has been the subject of research. Francisco et al. [14] designed a trust-region interior point method for underdetermined nonlinear systems which generalizes the method [1] for square systems while Kanzow et al. [24,25] proposed global projected Levenberg–Marquardt methods. In particular, the numerical solution of general nonsquare systems of bound-constrained nonlinear equations was addressed in [25] while the paper [24] is focused on overdetermined nonsmooth systems of equations. To make the methods globally convergent, a simple globalization strategy based on the projected gradient and a trust-region strategy were used in [24] and [25] respectively.

Concerning the solution of the general problem (1.1), several methods have been proposed. We recall extensions of Newton method to systems of mixed equalities and inequalities, see e.g. [8], global quadratic algorithms based on backtracking linesearch, see e.g. [5,15], and the recent trust-region methods [10,12,18]. These trust-region approaches are based on suitable transformations of the problem (1.1) and vary widely from a computational point of view.

Specifically, Fletcher and Leyffer [12] transformed the problem (1.1) into a bi-objective nonlinear programming problem and presented a trust-region filter SQP method. Alternatively, all the inequalities in (1.1) can be replaced by equalities and the problem takes the form of a least-squares problem

$$\min_{v \in \mathbb{R}^p} \phi(v) = \frac{1}{2} \| \Phi(v) \|_2^2,$$

where $\phi : \mathbb{R}^p \to \mathbb{R}$, and $\Phi : \mathbb{R}^p \to \mathbb{R}^{m+2p}$ is not continuously differentiable. Dennis et al. [10] followed this approach and proposed a single-model and a multimodel trust-region method for problem (1.2). Finally, Gould and Toint [18] used the transformation (1.2) and developed a Fortran package that combines the basic trust-region strategy with a filter technique.

Concerning the theoretical features of the above trust-region methods, it is important to note that the methods [10,12] are globally convergent under appropriate assumptions while the technique used in [18] to handle the inequality constraints is heuristic and no theoretical guarantee of convergence can be provided for problems involving inequality constraints. None of these methods are supported by local convergence analysis.

We propose trust-region methods for (1.1) that are globally convergent as well as quadratically convergent locally under an error bound condition. Our work has some ideas in common with the papers [10,18] but actually differs from such approaches in the transformation of problem (1.1) and in the features of local fast convergence. Specifically, to transform the problem (1.1) we distinguish between general inequalities, $c_i(v) \leq 0$, and simple bounds, $v_l \leq v \leq v_u$. We replace the inequalities $c_i(v) \leq 0$ by equalities and pose the problem as a bound-constrained nonlinear system where the residual function is continuously differentiable.

Two alternative transformations of (1.1) are considered. In both cases, the general form of the equivalent bound-constrained least-squares problem is

$$\min_{x \in \Omega} \theta(x) = \frac{1}{2} \| \Theta(x) \|_2^2,$$

where $\theta : \mathbb{R}^n \to \mathbb{R}$ with $n \geq p$, $\Theta : \mathbb{R}^n \to \mathbb{R}^m$ is continuously differentiable and $\Omega$ is an $n$-dimensional box. The motivation to consider (1.3) is twofold. First, the dimension of the vector function $\Theta$ is lower than the dimension of vector function $\phi$ in (1.2). Second, since (1.3) is a bound-constrained optimization problem, we can take advantage of recent developments in the affine scaling methods for nonlinear minimization problems and nonlinear systems with simple bounds (see e.g. [3,7,19,22,23,25]) to obtain theoretically well-founded and efficient algorithms.

We provide two new affine scaling trust-region methods. They differ in the choice of the quadratic model used throughout the iterations. One method employs a Gauss–Newton model while the other is a Levenberg–Marquardt method. Under reasonable assumptions, these methods globally converge to solutions of (1.1) or to first-order stationary points for the problem (1.3). Further, locally quadratic convergence properties are established assuming a local error bound condition on a neighbourhood of a solution of (1.1). The algorithmic options of the methods are discussed and numerical results on several test problems are presented.

The paper is organized as follows. In Section 2 we show two ways to convert the problem (1.1) into the least-squares problem (1.3) and provide details on the latter problem. In Section 3 we present the new methods along with

Please cite this article in press as: M. Macconi et al., Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities, Applied Numerical Mathematics (2008), doi:10.1016/j.apnum.2008.03.028
the algorithmic options while in Section 4 we carry out the theoretical study. In Section 5 we present comparative numerical results involving the two methods. Finally, conclusions and final perspectives are discussed in Section 6.

Notations. Throughout the paper we use the following notations. The \( i \)th component of a vector \( x \) is represented either by \( (x)_i \) or \( x_i \); the \( i \)th row of a matrix \( A \) is represented by \( (A)_i \). The 2-norm is denoted by \( \| \cdot \| \) and \( B_\rho(y) = \{ x : \| x - y \| < \rho \} \). The symbols \( I_n \) and \( I_m \) represent the identity matrix of dimension \( n \) and \( m \) respectively. \( A^+ \) denotes the Moore–Penrose pseudoinverse of the matrix \( A \). Given a sequence of vectors \( \{x_k\} \), for any function \( f \) we let \( f_k = f(x_k) \). Finally, we let \( \Omega = \{ x \in \mathbb{R}^n \mid l \leq x \leq u \} \) and \( P_\Omega(x) \) be the projection of \( x \) onto \( \Omega \), i.e. \( (P_\Omega(x))_i = \max\{l_i, \min\{x_i, u_i\}\}, i = 1, \ldots, n \).

2. Preliminaries

Consider the problem (1.1) and suppose that some components of either \( v_l \) or \( v_u \) are finite and that \( v_l < v_u \). Let \( C_E : \mathbb{R}^p \to \mathbb{R}^{m_E} \) be the vector function whose components are \( c_i \) for \( i \in E \) and \( C_I : \mathbb{R}^p \to \mathbb{R}^{m_I} \) be the vector function whose components are \( c_i \) for \( i \in I \). In the previous section we had set \( m = m_E + m_I \).

We convert the general inequalities, \( C_I(v) < 0 \), into equalities and transform the problem (1.1) into a bound-constrained least-squares problem of the form (1.3). Thus, the solutions to the original problem are zero-residual solutions for the least-squares problem.

In this paper we will restrict to transformations where \( \Theta \) is continuously differentiable. Thinking of the nonlinear least-squares problem (1.3) as a system of equation \( \Theta(x) = 0 \), \( x \in \Omega \), we say the problem is square if the number \( n \) of unknowns is equal to the number of components \( m \) of the function \( \Theta \), overdetermined if \( m > n \), underdetermined otherwise.

A possible transformation is to add a slack variable \( s \in \mathbb{R}^{m_I} \) whose non-positivity is imposed by means of simple bounds. The problem then becomes the least-squares problem (1.3) where \( x = (v^T , s^T)^T \), and \( \Theta : \mathbb{R}^{p+m_I} \to \mathbb{R}^m \) is given by

\[
\Theta(x) = \left( \begin{array}{c} C_E(v) \\ s - C_I(v) \end{array} \right), \quad v_l \leq v \leq v_u, \quad s \leq 0.
\] (2.1)

Another transformation is to use the function \( [t]_+ = \max\{t, 0\}^2/2 \) which is continuously differentiable and the problem then takes the form (1.3) where \( x = v \), \( \Theta : \mathbb{R}^p \to \mathbb{R}^m \),

\[
\Theta(x) = \left( \begin{array}{c} C_E(x) \\ [C_I(x)]_+ \end{array} \right), \quad v_l \leq x \leq v_u.
\] (2.2)

Note that \( [C_I(x)]_+ \) denotes the vector of infeasibilities at \( x \).

In both transformations, the number of components of the vector function \( \Theta \) is \( m \) while the number of variables differs. The transformation (2.1) adds \( m_I \) extra variables and the problem (1.3) is underdetermined if the system of equalities is underdetermined. On the contrary, the transformation (2.2) leaves the number of unknowns unchanged and \( \Theta \) inherits the dimensions of the system of equations and general inequalities.

Concerning the Jacobian matrix \( \Theta' \) of \( \Theta \), if the function \( \Theta \) has the form (2.2), then the \( i \)th row \( (\Theta'(x))_i \) is given by

\[
(\Theta'(x))_i = \begin{cases} \nabla c_i(x)^T & \text{if } i \in E, \\ \max\{c_i(x), 0\}\nabla c_i(x)^T & \text{if } i \in I, \end{cases}
\] (2.3)

for \( i = 1, \ldots, m \). Letting \( x^* \) be a solution to (1.1), the matrix \( \Theta'(x^*) \) has full rank if the system of equalities is square or overdetermined and \( C_E'(x^*) \) has full rank. Otherwise, the Jacobian does not have full rank at a solution and the rate of convergence of a minimization method for smooth problems may be inhibited.

We conclude this section recalling that the first-order optimality conditions for the problem (1.3) are equivalent to the system of nonlinear equations

\[
D(x)\nabla \theta(x) = D(x)\Theta'(x)^T \Theta(x) = 0,
\] (2.4)

where \( D(x) \) is a diagonal matrix,

\[
D(x) = \text{diag}(\left|v_1(x)\right|, \ldots, \left|v_n(x)\right|),
\] (2.5)
Choosing $\mu$ is well-angled with respect to the bounds. In fact, for the components $x_k(i)$ which are approaching the correct bounds, $d_k$ becomes increasingly tangential to the bounds. Hence, the bounds will not prevent a large stepsize along $d_k$. For the components $x_k(i)$ which are approaching the incorrect bounds, $d_k$ points away from these bounds in relatively large angles [4].

3. The method and its algorithmic options

We present a trust-region method for the bound-constrained least-squares problem (1.3). The sequence \( \{x_k\} \) generated by the method consists of feasible points for (1.3), i.e. \( x_k \in \Omega, \ k \geq 0 \). Without loss of generality we assume that for all \( k \), \( x_k \) is not a stationary point for the least-squares problem (1.3).

At the \( k \)th iteration, given \( \Delta_k > 0 \), we consider a quadratic model \( m_k(p) \) for \( \theta \) at \( x_k \) and replace the problem (1.3) by the trust-region problem

\[
\min_p \{ m_k(p) : \| p \| \leq \Delta_k \}.
\]

The quadratic model \( m_k \) has the form

\[
m_k(p) = \frac{1}{2} \| \Theta_k p + \Theta_k \|_2^2 + \frac{1}{2} \mu_k \| p \|_2^2,
\]

where \( \mu_k \) is a nonnegative scalar.

We allow for two definitions of the sequence \( \{\mu_k\} \) in (3.2). The first option is to let \( \mu_k = 0 \) for all \( k \) so that the Gauss–Newton model is used throughout the iterations. In this case, the global minimizer of \( m_k \) is unique if \( \Theta_k \) has full column rank. The second option is to choose a sequence of strictly positive scalars \( \{\mu_k\} \) such that \( \mu_k = O(\| \Theta_k \|^2) \). This way, \( m_k \) is a strictly convex function and can be interpreted as a regularized Gauss–Newton model.

The two options for the sequence \( \{\mu_k\} \) can be defined as follows. Fixed \( \mu \geq 0 \) and a small positive scalar \( \hat{\mu} \), we let

\[
\mu_k = \begin{cases} 
\min\{\hat{\mu}, \mu\} \| \Theta_0 \|_2^2 & \text{if } k = 0, \\
\min\{\mu_{k-1}, \mu \| \Theta_k \|_2^2\} & \text{if } k > 0.
\end{cases}
\]

Choosing \( \mu = 0 \) we get \( \mu_k = 0 \) for all \( k \); on the other hand if \( \mu > 0 \) then \( \mu_k > 0 \) for all \( k \). In the latter case, as \( \hat{\mu} \) is assumed to be small, the model \( m_k \) does not drift too far from the Gauss–Newton model when \( \| \Theta_k \| \) is large.

Any solution \( p_{tu} \) to the trust-region problem (3.1) satisfies the equation

\[
(\Theta_k^T \Theta_k + (\mu_k + \lambda) I_n) p_{tu} = -\Theta_k^T \Theta_k 
\]

where \( \lambda \geq 0 \) and \( \lambda(\| p_{tu} \| - \Delta_k) = 0 \). If \( \Theta_k^T \Theta_k + (\mu_k + \lambda) I_n \) is positive definite, then \( p_{tu} \) is unique [6, Corollary 7.2.2].

In the case where \( \Theta_k^T \Theta_k \) is rank deficient and \( \mu_k = \lambda = 0 \), if \( \Theta_k^T \Theta_k \leq \Delta_k \) then \( p_{tu} = -\Theta_k^T \Theta_k \) is the solution to (3.1) for which \( \| p_{tu} \| \) is least, see [27].

Correspondingly to the problem (3.1), we consider a general Cauchy step \( p_k^C \) along the scaled steepest descent direction \( d_k \) given in (2.7). The step \( p_k^C \) minimizes \( m_k \) along the direction \( d_k \) within the feasible trust-region, i.e.

\[
p_k^C = \arg\min_{p \in \text{span}(d_k)} m_k(p) \quad \text{subject to } \| p \| \leq \Delta_k, \quad x_k + p \in \Omega.
\]

Now we outline the general form of our method which will be denoted as TREI method (Trust-Region method for systems of nonlinear Equalities and Inequalities). The \( k \)th iteration is given below; we remark that the parameters \( \Delta_{\min}, \beta_1, \beta_2, \hat{\mu}, \mu \) are independent of \( k \).
TREI method – kth iteration.

Input: \( x_k \in \Omega, \Delta_k \geq \Delta_{\min} > 0, \beta_1, \beta_2, \hat{\mu} \in (0, 1), \mu \geq 0. \)

1. Define \( \mu_k \) according to (3.3).
2. Compute the minimum-length solution \( p_N^k \) to the problem \( \min_p m_k(p) \).
3. Compute the generalized Cauchy step \( p_C^k \) based on (3.5).
4. If \( \| p_N^k \| \leq \Delta_k \) set \( p_tr = p_N^k \);
   Else find the dogleg step \( p_tr \) for (3.1).
5. Let \( \bar{p}_{tr} = P_\Omega(x_k + p_tr) - x_k \).
6. If \( \rho_c(\bar{p}_{tr}) = \frac{m_k(0) - m_k(\bar{p}_{tr})}{m_k(0) - m_k(p_C^k)} \geq \beta_1 \), (3.6)
   Set \( p_k = \bar{p}_{tr} \);
   Else find \( p_k = t p_C^k + (1 - t) \bar{p}_{tr}, t \in (0, 1] \), such that (3.6) holds.
7. If \( \rho_\theta(p_k) = \frac{\theta(x_k) - \theta(x_k + p_k)}{m_k(0) - m_k(p_k)} \geq \beta_2 \), (3.7)
   Set \( x_{k+1} = x_k + p_k \), choose \( \Delta_{k+1} \geq \Delta_{\min} \);
   Else reduce \( \Delta_k \) and go to Step 3.

The TREI method is the paradigm for our procedure and two different implementations are allowed. The implementation with \( \mu = 0 \) is based on the Gauss–Newton model and will be denoted as TREI-GN. The core of this method is similar to the Gauss–Newton methods given in [10,14,31]. The other implementation is obtained setting \( \mu > 0 \) and gives rise to a Levenberg–Marquardt method; therefore it will be denoted TREI-LM method.

The model \( m_k \) used in the TREI-LM method has been considered in several papers, see e.g. [24,25]. Our method actually differs from the Projected Levenberg–Marquardt method [25] in the globalization strategy. In fact, we adopt a trust-region strategy with affine scaling matrices which is known to be more efficient than the projected gradient strategy used in [25]. With respect to the methods in [24], our procedure does not require at each iteration the minimization of a box constrained quadratic program.

We now analyze in more details the TREI method. First we describe the steps of the method and specify the tasks that are not shared by the two implementations; second we discuss the most relevant algorithmic options.

The parameter \( \mu_k \) is evaluated in Step 1. Then, in Step 2 we find a minimum norm minimizer \( p_N^k \) of \( m_k \). Clearly, in the TREI-GN method the step \( p_N^k \) has the form

\[
p_N^k = -\Theta_k^T \Theta_k,
\]

while in the TREI-LM method \( p_N^k \) solves the system of equations

\[
(\Theta_k^T \Theta_k' + \mu_k I_n) p_N^k = -\Theta_k^T \Theta_k.
\]

We remark that in (3.9) the coefficient matrix is positive definite and \( p_N^k \) can be viewed as the step obtained by using a Levenberg–Marquardt method.

Taking into account (2.7) and (3.5), Step 3 is performed letting

\[
p_C^k = c_k d_k, \quad c_k = \begin{cases} \omega_1 & \text{if } x_k + \omega_1 d_k \in \Omega, \\ \omega_2 & \text{otherwise,} \end{cases}
\]

where

\[
\omega_1 = \min \left\{ \frac{\| D_k^\frac{1}{2} \nabla \theta_k \|}{\| \Theta_k' d_k \| + \mu_k \| d_k \|}, \Delta_k \right\}, \quad \omega_2 = \arg \max \{ \omega > 0, x_k + \omega d_k \in \Omega \}.
\]

In Step 4, the trust-region problem (3.1) is solved. Since the minimum-length step \( p_N^k \) solves (3.4) with \( \lambda = 0 \), there are two possibilities: either \( p_N^k \) is inside the trust-region, i.e., it solves the problem (3.1) or there are no unconstrained
minimizers of $m_k$ inside the trust-region. Therefore, in Step 4 we set $p_\text{tr} = p_k^N$ if $p_k^N$ is inside the trust-region; otherwise, we find an approximate solution $p_\text{tr}$ to (3.1) on the boundary of the trust-region by the dogleg strategy [9]. The reason for using the dogleg strategy instead of the Moré and Sorensen algorithm [28] is that in the hard case the latter strategy steps to the boundary of the trust-region even when an unconstrained minimizer of $m_k$ is safely inside.

To generate a new feasible iterate, in Step 5 we project $x_k + p_\text{tr}$ onto the box $\Omega$ and define a possibly modified trust-region step $\tilde{p}_\text{tr}$ such that $x_k + \tilde{p}_\text{tr}$ is feasible. Steps 6–7 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 6 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 [1,7]. We let $p_k = \tilde{p}_\text{tr}$ if $\tilde{p}_\text{tr}$ satisfies (3.6), otherwise, we look for a step of the form $p_k = tp_k^C + (1-t)\tilde{p}_\text{tr}, t \in (0, 1]$, satisfying the required condition. In Step 7, we measure the quality of the quadratic model $m_k$ as an approximation to $\theta$ around $x_k$. If the sufficient improvement condition (3.7) 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.

It is important to note that the initial trust-region radius satisfies $\Delta_k \geq \Delta_{\text{min}}$ where $\Delta_{\text{min}}$ is a fixed positive scalar. Moreover, in Step 7 after $x_{k+1}$ is formed, the trust-region radius $\Delta_{k+1}$ is fixed so that $\Delta_{k+1} \geq \Delta_{\text{min}}$. Therefore, in each iteration of the method the initial radius is more than or equal to $\Delta_{\text{min}}$ while on termination of the iteration the trust-region radius may be smaller than $\Delta_{\text{min}}$.

We conclude this section making some observation on the algorithmic options. From a computational point of view, the TREI-GN and TREI-LM methods differ in the computation of the step $p_k^N$. The TREI-GN method uses the step $p_k^N$ in (3.8) and such step can be computed in a numerically reliable way from either the complete orthogonal factorization of $\Theta_k'$ or the singular value decomposition of $\Theta_k'$, [16, 6.6.4, 6.6.5]. The TREI-LM method uses the step $p_k^N$ given in (3.9). In this case the simplest way to obtain $p_k^N$ is to use the Cholesky decomposition on the system (3.9). Alternatively, $p_k^N$ can be computed applying the QR decomposition to the least-squares problem

$$\min_m m_k(p) = \min_p \frac{1}{2} \left\| \left( \Theta_k' \sqrt{\mu_k} I_n \right) p + \left( \Theta_k 0 \right) \right\|^2.$$  \hspace{1cm} (3.12)

The computation of the vector $p_k = tp_k^C + (1-t)\tilde{p}_\text{tr}, t \in (0, 1]$, in Step 6 can be accomplished easily in two different ways. The first possibility is to simply set $t = 1$, i.e. $p_k = p_k^C$. The second one is to find the scalar $t \in (0, 1)$ such that $\rho_c(p_k) = \beta_1$; this is equivalent to find the smallest positive root of the scalar quadratic equation in $t$ of the form $\rho_c(tp_k^C + (1-t)\tilde{p}_\text{tr}) - \beta_1 = 0$, see [2].

Finally, it is worth noting that it is possible to achieve economies in the calculations when $\Theta$ is of the form (2.2). In fact, by (2.2) and (2.3) it is easy to see that a zero component of $[C_F(x_k)]_+$ gives rise to a zero component in $\Theta$ and to a null row in $\Theta'$. Then, in the practical algorithm it is convenient to consider the reduced model

$$\tilde{m}_k(p) = \frac{1}{2} \left\| \tilde{\Theta}_k p + \tilde{\Theta}_k \right\|^2 + \frac{1}{2} \mu_k \|p\|^2.$$  \hspace{1cm} (3.13)

where $\tilde{\Theta}_k$ is the vector formed by $C_F(x_k)$ and the nonzero components of $[C_F(x_k)]_+$, and $\tilde{\Theta}_k'$ is the Jacobian of $\tilde{\Theta}$ at $x_k$, see [18]. In the following section, without loss of generality we will refer to (3.2).

4. Convergence results

In this section we establish the global and local convergence properties of the methods presented. Throughout the section we let $\{x_k\}$ be the sequence generated by any implementation of the TREI method and make the following basic assumption on the function $\Theta$ in (1.3).

Assumption 4.1.

- There exists an open, bounded and convex set $L$ containing the whole sequence $\{x_k\}$ such that $L \supset \{x \in \mathbb{R}^n: \exists x_k \text{ s.t. } \|x - x_k\| \leq r\}$, for some $r > 0$, and for all $x, z \in L$, the Jacobian matrix $\Theta'$ satisfies

$$\|\Theta'(x) - \Theta'(z)\| \leq 2\gamma_D \|x - z\|.$$ \hspace{1cm} (4.1)

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

Please cite this article in press as: M. Macconi et al., Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities, Applied Numerical Mathematics (2008), doi:10.1016/j.apnum.2008.03.028
It is easy to see that if \( \Theta \) has the form (2.1) then these assumptions are equivalent to suppose that \( C_E'(x), C'_I(x) \) are bounded in norm for \( x \in L \) and that \( C_E'(x) \) and \( C'_I(x) \) are Lipschitz continuous at every point of the set \( L \). If \( \Theta \) has the form (2.2) it can be shown following the lines of [10, Lemma 4.1] that Assumption 4.1 is equivalent to suppose that \( C_I(x), C'_E(x), C'_I(x) \) are bounded in norm for \( x \in L \) and that \( C'_E(x) \) and \( C'_I(x) \) are Lipschitz continuous at every point of the set \( L \).

Trivially, Assumption 4.1 implies that the sequence \( \{x_k\} \) is bounded and for all \( x, z \in L \)
\[
\|\Theta(x) - \Theta(z)\| \leq \gamma_L \|x - z\|, \tag{4.2}
\]
\[
\|\Theta(x) - \Theta(z) - \Theta'(z)(x - z)\| \leq \gamma_D \|x - z\|^2. \tag{4.3}
\]

Under Assumption 4.1, the TREI-GN and TREI-LM methods are global convergent. The convergence analysis is provided in the following theorem and it is carried out considering the general form of the TREI method. We prove that independently of the choice of the initial guess \( x_0 \), the limit points of the sequence \( \{x_k\} \) are stationary points for the problem (1.3). However, since a limit point \( x^* \) may be a minimum for (1.3) such that \( \|\Theta(x^*)\| > 0 \), we list the cases where the limit points solve (1.1).

**Theorem 4.1.** Let Assumption 4.1 hold and \( \{x_k\} \) be the sequence generated by the TREI method. 

(i) Every limit point of the sequence \( \{x_k\} \) is a first-order stationary point for the problem (1.3).

(ii) If \( x^* \) is a limit point of \( \{x_k\} \) and \( \|\Theta(x^*)\| = 0 \), then all the limit points of \( \{x_k\} \) solve the problem (1.1).

(iii) If the problem is either square and underdetermined and \( x^* \) is a limit point of \( \{x_k\} \) such that \( x^* \in \text{int}(\Omega) \) and \( \Theta'(x^*) \) has full rank, then \( x^* \) solves the problem (1.1).

**Proof.** (i) If \( D_k \nabla \theta_k \neq 0 \), then the step \( p_k \) taken to form the iterate \( x_{k+1} \) satisfies
\[
m_k(0) - m_k(p_k) \geq \frac{\beta_1}{2} \|D_k \nabla \theta_k\| \min \left\{ \frac{\Delta_k}{\|D_k \nabla \theta_k\|}, \frac{\|D_k^\frac{1}{2} \nabla \theta_k\|}{\|D_k^\frac{1}{2} (\Theta'_k \Theta_k' + \mu_k I_h) D_k^\frac{1}{2}\|}, \frac{\|D_k^\frac{1}{2} \nabla \theta_k\|}{\|\nabla \theta_k\|} \right\}. \tag{4.4}
\]
This inequality can be proved following the lines of [3, Lemma 4.1].

Since \( \{x_k\} \) is bounded, there exists a constant \( \chi_1 > 0 \) such that \( \|D_k^\frac{1}{2}\| \leq \chi_1 \) for all \( k \). From Assumption 4.1, \( \nabla \theta \) is Lipschitz continuous in \( L \) [29] and there exists a constant \( \chi_2 > 0 \) such that \( \|\nabla \theta_k\|_{\infty} \leq \chi_2 \) for all \( k \). Further, by construction \( \mu_k \leq \mu_0 \) for all \( k \). We prove by contradiction that \( \lim inf_{k \to \infty} \|D_k^\frac{1}{2} \nabla \theta_k\| = 0 \). In particular, we assume that there exists \( \epsilon > 0 \) such that \( \lim inf_{k \to \infty} \|D_k^\frac{1}{2} \nabla \theta_k\| > \epsilon \). This implies that there exists \( \tilde{k} \) such that \( \|D_k^\frac{1}{2} \nabla \theta_k\| > \epsilon \) whenever \( k > \tilde{k} \). Assume \( k > \tilde{k} \). Then we have
\[
m_k(0) - m_k(p_k) \geq \frac{\beta_1 \epsilon}{2} \min \left\{ \frac{\Delta_k}{\chi_1}, \frac{\epsilon}{\chi_2}, \frac{\epsilon}{\chi_2} \right\},
\]
with \( \chi_L \) given in Assumption 4.1 and proceeding as in [29, Theorem 4.7] we obtain that at termination of the iteration, the trust region radius \( \Delta_k \) is bounded away from zero for all \( k > \tilde{k} \).

On the other hand, at each iteration (3.7) is satisfied, i.e.
\[
\Theta_k - \Theta_{k+1} \geq 2 m_k(0) - m_k(p_k) \geq \frac{\beta_1 \beta_2 \epsilon}{2} \min \left\{ \frac{\Delta_k}{\chi_1}, \frac{\epsilon}{\chi_1}, \frac{\epsilon}{\chi_2} \right\}.
\]
Since the sequence \( \{\theta_k\} \) is monotone decreasing and bounded from below it is convergent. Hence, \( \lim_{k \to \infty} (\theta_k - \theta_{k+1}) = 0 \) implies \( \lim_{k \to \infty} \Delta_k = 0 \). This is a contradiction and we must have \( \lim inf_{k \to \infty} \|D_k^\frac{1}{2} \nabla \theta_k\| = 0 \). Finally, following the proof of [7, Theorem 3.5] the hypothesis follows.

(ii) The sequence \( \{\theta_k\} \) is monotone decreasing and bounded from below; hence it converges. Since \( \|\Theta(x^*)\| = 0 \), then \( \lim_{k \to \infty} \theta_k = 0 \).

(iii) Since \( x^* \in \text{int}(\Omega) \), by \( \|D(x^*)^\frac{1}{2} \nabla \theta(x^*)\| = 0 \) it follows \( \Theta'(x^*)^T \Theta(x^*) = 0 \). Then, \( \text{rank}(\Theta'(x^*)) = m \) yields \( \|\Theta(x^*)\| = 0 \). \( \square \)
Under further assumptions, we are able to carry out the local convergence analysis. We let $S$ be the set of solutions to problem (1.1), $d(x, S)$ denote the distance from the point $x$ to the solution set $S$ and $[x]_S \in S$ be such that $\|x - [x]_S\| = d(x, S)$, i.e.

\[
S = \{ y \in \Omega : \|\Theta(y)\| = 0 \}, \quad d(x, S) = \inf\{\|x - y\|, y \in S\}, \quad [x]_S = \arg\min_{y \in S} \|x - y\|. \tag{4.4}
\]

Then, we make the following assumptions.

**Assumption 4.2.**

- The solution set $S$ of problem (1.1) is nonempty;
- The sequence $\{x_k\}$ generated by the TREI method has a limit point $x^* \in S$;
- There exists positive constants $\rho$ and $\alpha_0$ such that

\[
\frac{1}{\alpha_0} d(x, S) \leq \|\Theta(x)\| \quad \forall x \in B_\rho(x^*). \tag{4.6}
\]

The condition (4.6) deserves some consideration. When it is satisfied the function $\|\Theta\|$ provides a local error bound for the problem (1.3) near $x^* \in S$, see [26,30]. If the problem is either overdetermined or square and $\Theta'(x^*)$ is full rank, then the matrix $\Theta'(x^*)^T \Theta'(x^*)$ is nonsingular and $\|\Theta\|$ provides a local error bound on some neighbourhood of $x^*$. This fact will be shown in the forthcoming Lemma 4.2. On the other hand, the converse is not true: the local error bound is a weaker condition than the nonsingularity of $\Theta'(x^*)^T \Theta'(x^*)$ [25].

Typically, the condition (4.6) is assumed to hold in a region of the form $B_\rho(x^*) \cap \Omega$ [30]. However, in our convergence analysis we need to apply the error bound condition on points that may lie outside $\Omega$. For this reason, we drop the restriction on $\Omega$ and consider the assumption (4.6). Although the condition (4.6) is more restrictive when $x \in B_\rho(x^*)$ than in the case where $x \in B_\rho(x^*) \cap \Omega$, it has been shown that such condition is still significantly weaker than the nonsingularity of $\Theta'(x^*)^T \Theta'(x^*)$, see [25, §3].

The next lemma provides three useful properties. Two of them are direct consequences of Assumptions 4.1 and 4.2; the third one holds under the following additional condition

\[
\Theta'(x^*) \text{ is full rank, i.e. } \operatorname{rank}(\Theta'(x^*)) = \min\{m, n\}. \tag{4.7}
\]

**Lemma 4.1.** Let Assumptions 4.1 and 4.2 hold. Then there exists a constant $\tau > 0$ such that for $x \in B_\tau(x^*)$

\[
x \in L \quad \text{and} \quad [x]_S \in L, \tag{4.8}
\]

\[
\|\Theta(x)\| \leq \chi_L d(x, S). \tag{4.9}
\]

Further, if $\Theta'(x^*)$ is full rank then there exists a constant $v > 0$ such that for $x \in B_\tau(x^*)$

\[
\Theta'(x) \text{ is full rank and } \|\Theta'(x)^+\| \leq v. \tag{4.10}
\]

**Proof.** Let $r$ be the scalar given in Assumption 4.1. Choose $\tau$ so that $\tau < \min\{\rho, r/2\}$ where $\rho$ is the scalar in Assumption 4.2.

Since $x^*$ is a limit of $\{x_k\}$ there exists $x_k$ such that $\|x_k - x^*\| \leq r - 2\tau$. Then, if $x \in B_\tau(x^*)$ we have $\|x - x_k\| \leq \|x - x^*\| + \|x^* - x_k\| \leq r - \tau$, i.e. $x \in L$. Further, let $[x]_S$ as in (4.5). Then, $\|[x]_S - x_k\| \leq \|[x]_S - x\| + \|x - x_k\| \leq \|x^* - x\| + \|x - x_k\| \leq r$ i.e. $[x]_S \in L$.

The second part of the hypothesis follows from $\Theta(x) = \Theta(x) - \Theta([x]_S)$, (4.2) and (4.5).

To show (4.10), let $q = \min\{m, n\}$ and $\sigma_q(\Theta'(x))$ be the smallest singular value of $\Theta'(x)$. By (4.7) we know that $\sigma_q(\Theta'(x^*)) > 0$. Using [21, Corollary 7.3.8] and Assumption 4.1 we get

\[
|\sigma_q(\Theta'(x)) - \sigma_q(\Theta'(x^*))| \leq \|\Theta'(x) - \Theta'(x^*)\| \leq 2\gamma \rho \|x - x^*\|,
\]

and consequently
\begin{equation}
\sigma_q(\Theta'(x)) \geq \sigma_q(\Theta'(x^*)) - 2\gamma_D \|x - x^*\|.
\end{equation}

Then, setting \( \tau \leq \min\{\rho, r/2, \sigma_q(\Theta'(x^*)/(2\gamma_D))\} \) and \( v = 1/(\sigma_q(\Theta'(x^*)) - 2\gamma_D \tau) \) completes the proof. \( \square \)

The following lemma shows that if \( \Theta'(x^*)^T \Theta(x^*) \) is nonsingular, then \( \|\Theta\| \) is guaranteed to provide a local error bound on some neighbourhood of \( x^* \).

**Lemma 4.2.** Suppose \( m \geq n \). If \( x^* \in S \) and \( \Theta'(x^*) \) is full rank then (4.6) holds.

**Proof.** Let \( x \in B_{\tau}(x^*) \) where \( \tau \) is the scalar given in Lemma 4.1. Since \( \Theta'(x^*) \) is full rank, then \( \Theta'(x^*)^+ = (\Theta'(x^*)^T \Theta'(x^*))^{-1} \Theta'(x^*)^T \) and \( \Theta'(x^*)^+ \Theta'(x^*) = I_n \). Also, by Assumption 4.1 we get

\[
\|I_n - \Theta'(x^*)^+ \Theta'(x)\| \leq \|\Theta'(x^*)^+\| \|\Theta'(x^*) - \Theta'(x)\| \leq 2\gamma_D \|\Theta'(x^*)^+\| \|x - x^*\|.
\]

Choosing \( \rho < \min\{\tau, 1/(4\gamma_D \|\Theta'(x^*)^+\|)\} \), we have \( \|I_n - \Theta'(x^*)^+ \Theta'(x)\| \leq 1/2 \) for \( x \in B_{\rho}(x^*) \). Then, by using the mean value theorem we obtain

\[
\|\Theta'(x^*)^+ \Theta'(x)\| = \left\| (x - x^*) - \int_0^1 (I_n - \Theta'(x^*)^+ \Theta'(x^* + t(x - x^*)))(x - x^*) \, dt \right\|
\geq \left( 1 - \frac{1}{2} \right) \|x - x^*\|.
\]

Hence
\[
\|\Theta(x)\| \geq \frac{\|\Theta'(x^*)^+ \Theta'(x)\|}{\|\Theta'(x^*)^+\|} \geq \frac{1}{2 \|\Theta'(x^*)^+\|} \|x - x^*\|,
\]

and recalling (4.5), the result trivially follows. \( \square \)

Clearly, (4.12) implies that \( x^* \) is an isolated zero-residual solution to (1.3) in the neighbourhood \( B_{\rho}(x^*) \).

To prove that the sequence \( \{x_k\} \) is quadratically convergent, it will be necessary to distinguish between the TREI-GN and TREI-LM method. The assumptions required by the TREI-GN method are stronger as we assume that Assumptions 4.1, 4.2 and condition (4.7) are satisfied. On the contrary, the condition (4.7) is not needed for the TREI-LM method.

Since the analysis for the two methods follows the same main steps, the proofs are presented simultaneously. The first step of our analysis establishes results on the step \( p_k \) selected to form \( x_{k+1} \) when \( x_k \) is sufficiently close to \( x^* \).

We start providing an intermediate lemma.

**Lemma 4.3.** Let Assumptions 4.1 and 4.2 hold. Then for the TREI-LM method there exist positive constants \( \epsilon \) and \( \eta \) such that if \( x_k \in B_{\tau}(x^*) \) then
\[
\|\Theta(x_k + p^N_k)\| \leq \eta d(x_k, S)^2.
\]

For the TREI-GN method, (4.13) is true under the additional assumption that \( \Theta'(x^*) \) is full rank.

**Proof.** Due to Assumption 4.2 and Theorem 4.1 we know that \( \lim_{k \to \infty} \|\Theta_k\| = 0 \). Moreover, by the definition (3.3) of \( \mu_k \), it follows
\[
\mu_k = \mu \|\Theta_k\|^2,
\]

for \( x_k \) sufficiently close to \( x^* \). Let \( \tau \) as in Lemma 4.1 and \( \epsilon \leq \tau \) small enough so that if \( x_k \in B_{\tau}(x^*) \) then (4.14) holds. Fix \( x_k \in B_{\tau}(x^*) \).

First, we provide an upper bound on the norm of the step \( p^N_k \), showing that
\[
\|p^N_k\| \leq \alpha_1 d(x_k, S),
\]
for some positive scalar $\alpha$. In the TREI-LM method we prove this fact in the same way as in [25, Lemma 2.3]. In particular, by (4.14), (4.6) and (4.9)

$$
\frac{\mu}{\alpha_0} d(x_k, S)^2 \leq \mu_k \leq \mu \chi_L^2 d(x_k, S)^2.
$$

(4.16)

Also, since $p_k^N$ is the global minimum of the model $m_k$ we have

$$
m_k(p_k^N) \leq m_k(x_k, [x_k]S),
$$

where $[x_k]S$ is the closest solution to $x_k$, see (4.5). Then, by (4.3)

$$
2m_k(p_k^N) \leq \|\Theta'_k(x_k, [x_k]S) + \Theta_k\|^2 + \mu_k \|x_k, [x_k]S\|^2
\leq \gamma_D^2 d(x_k, S)^4 + \mu_k d(x_k, S)^2,
$$

(4.17)

and

$$
\|p_k^N\|^2 \leq \frac{2}{\mu_k} m_k(p_k^N) \leq \left(1 + \frac{\alpha_0^2 \gamma_D^2}{\mu}\right) d(x_k, S)^2.
$$

Thus, we obtain (4.15) setting $\alpha_1 = \sqrt{1 + \alpha_0^2 \gamma_D^2/\mu}$. Moreover, note that from (4.15) and (4.16) we get the inequality

$$
\mu_k \|p_k^N\|^2 \leq \mu \alpha_0^2 \chi_L^2 d(x_k, S)^4.
$$

(4.18)

Focusing on the TREI-GN method, by (3.8), (4.10) and (4.9) we get $\|p_k^N\| \leq v\|\Theta_k\| \leq v\chi_L d(x_k, S)$ and (4.15) holds with $\alpha_1 = \sqrt{v\chi_L}$.

Second, we show that if $\epsilon$ is sufficiently small then

$$
x_k + p_k^N \in B_\epsilon(x^*), \quad x_k + p_k^N \in L, \quad [x_k + p_k^N]S \in L.
$$

(4.19)

To this end, let $\epsilon \leq \tau/(1 + \alpha_1)$ and note that (4.15) yields

$$
\|x_k + p_k^N - x^*\| \leq \|x_k - x^*\| + \|p_k^N\| \leq (1 + \alpha_1)\epsilon \leq \tau.
$$

Then, the last two statements in (4.19) derive from Lemma 4.1.

Finally, to prove (4.13), note that

$$
\left\|\begin{array}{c}
\Theta(x_k + p_k^N) \\
0
\end{array}\right\| \leq \left\|\begin{array}{c}
\Theta(x_k + p_k^N) - \Theta'_k p_k^N - \Theta_k \\
-\sqrt{\mu_k} p_k^N
\end{array}\right\| + \left\|\begin{array}{c}
\Theta'_k p_k^N + \Theta_k \\
\sqrt{\mu_k} p_k^N
\end{array}\right\|
\leq \left\|\begin{array}{c}
\Theta(x_k + p_k^N) - \Theta'_k p_k^N - \Theta_k \\
\sqrt{\mu_k} p_k^N
\end{array}\right\|^2 + \frac{1}{2} + \frac{1}{2} + \left(2m_k(p_k^N)\right)^{1/2}.
$$

Hence, using (4.3), (4.15), (4.18) and (4.17) completes the proof with

$$
\eta = \sqrt{\alpha_1^4 \gamma_D^2 + \mu \alpha_0^2 \chi_L^2 + \gamma_D^2 + \mu \chi_L^2}.
$$

The following lemma shows that if $x_k$ is sufficiently close to $x^*$ then the step

$$
p_k^N = P_Q(x_k + p_k^N) - x_k,
$$

(4.20)

is taken to form the new iterate.

**Lemma 4.4.** Let Assumptions 4.1 and 4.2 hold. Then, for the TREI-LM method there exist positive constants $\psi$ and $\alpha_2$ such that if $x_k \in B_\psi(x^*)$ the iterate $x_{k+1}$ has the form $x_{k+1} = x_k + p_k^N$ and

$$
d(x_{k+1}, S) \leq \alpha_2 d(x_k, S)^2.
$$

(4.21)

For the TREI-GN method, (4.21) is true under the additional assumption that $\Theta'(x^*)$ is full rank.
Proof. Since \( x^* \in S \) is a limit point of the sequence \( \{x_k\} \), it follows \( \lim_{k \to \infty} \|\Theta_k\| = 0 \). As a consequence (4.6) and (4.15) give \( \lim_{k \to \infty} \|\Theta_k\| = 0 \).

Let \( \epsilon \) as in Lemma 4.3 and suppose \( \psi_1 \leq \epsilon \) be sufficiently small so that \( \|P_k^N\| \leq \Delta_{\min} \) if \( x_k \in B_{\psi_1}(x^*) \). Fix \( x_k \in B_{\psi_1}(x^*) \). Then, the step \( P_k^N \) is the solution to the trust-region problem (3.1) and (4.19) holds.

To show the hypothesis we will prove that \( \bar{\Theta}_k \) defined in (4.20) satisfies both condition (3.6) and (3.7). First consider condition (3.6). If \( \bar{P}_k^N = P_k^N \) then (3.6) trivially follows by \( m_k(P_k^N) < m_k(P_k^C) \). If \( \bar{P}_k^N \neq P_k^N \) note that by (4.6) we have

\[
\rho_c(\bar{P}_k^N) = \frac{m_k(0) - m_k(\bar{P}_k^N)}{m_k(0)} \geq 1 - \alpha_0^2 \frac{\|\Theta_k'\bar{P}_k^N + \Theta_k\|^2 + \mu_k\|\bar{P}_k^N\|^2}{d(x_k, S)^2}.
\]

Thus, to investigate the condition (3.6) we need to estimate \( \|\Theta_k'\bar{P}_k^N + \Theta_k\| \). We begin to note that

\[
\Theta_k'\bar{P}_k^N + \Theta_k = \Theta(x_k + \bar{P}_k^N) - \Theta(x_k + P_k^N) + \Theta(x_k + \bar{P}_k^N) - (\Theta(x_k + \bar{P}_k^N) - \Theta_k),
\]

and by the contractivity of the projection map \( P_\Omega \)

\[
\|\bar{P}_k^N\| \leq \| P_k^N \|, \quad \| x_k + \bar{P}_k^N - z \| \leq \| x_k + P_k^N - z \|, \quad z \in \Omega.
\]

Then, (4.19) and (4.23) yield \( x_k + P_k^N \in L \) and by Assumption 4.1 we obtain

\[
\|\Theta_k'\bar{P}_k^N + \Theta_k\| \leq \chi_L \| x_k + P_k^N - [x_k + P_k^N]_S \| + \gamma D \| P_k^N \| \leq \chi_L \| x_k + P_k^N - [x_k + P_k^N]_S \| + \gamma D \| P_k^N \| \leq \chi_L d(x_k, P_k^N, S) + \gamma D \| P_k^N \|.
\]

Moreover, by (4.6), (4.13) and (4.15)

\[
\|\Theta_k'\bar{P}_k^N + \Theta_k\| \leq \alpha_0 \chi_L \| \Theta(x_k + P_k^N) \| + \gamma D \| P_k^N \| \leq \psi \ d(x_k, S)^2,
\]

where \( \psi = (\alpha_0 \chi_L + \gamma D \alpha_1^2) \). Thus, combining (4.22), (4.24) and (4.18)

\[
\rho_c(\bar{P}_k^N) \geq 1 - \alpha_0^2 (\psi^2 + \mu_k^2 \chi_L^2) d(x_k, S)^2 \geq 1 - \alpha_0^2 (\psi^2 + \mu_k^2 \chi_L^2) d(x_k, S)^2,
\]

i.e. \( \bar{P}_k^N \) satisfies condition (3.6) if \( x_k \) is sufficiently close to \( x^* \). Second, we focus on condition (3.7). From the mean value theorem we get

\[
\Theta(x_k + \bar{P}_k^N) = \Theta_k + \int_0^1 \Theta'(x_k + t \bar{P}_k^N) \bar{P}_k^N \, dt + \Theta_k' \bar{P}_k^N - \Theta_k' \bar{P}_k^N.
\]

and consequently

\[
\|\Theta(x_k + \bar{P}_k^N)\|^2 = \|\Theta_k' \bar{P}_k^N + \Theta_k\|^2 + \left\| \int_0^1 (\Theta'(x_k + t \bar{P}_k^N) - \Theta_k') \bar{P}_k^N \, dt \right\|^2
\]

\[
\quad + 2 \left( \int_0^1 (\Theta'(x_k + t \bar{P}_k^N) - \Theta_k') \bar{P}_k^N \, dt \right)^T (\Theta_k' \bar{P}_k^N + \Theta_k).
\]

So (4.24) gives rise to

\[
\|\Theta(x_k + \bar{P}_k^N)\|^2 - \|\Theta_k' \bar{P}_k^N + \Theta_k\|^2 \leq \gamma D \| \bar{P}_k^N \|^2 + 2 \psi d(x_k, S)^2 \|P_k^N\|^2.
\]

Furthermore, using (4.6), (4.24) and the fact that \( x_k \in B_{\psi_1}(x^*) \) we have

\[
\|\Theta_k\|^2 - \|\Theta_k' \bar{P}_k^N + \Theta_k\|^2 \geq \left( \frac{1}{\alpha_0^2} - \psi^2 d(x_k, S)^2 \right) d(x_k, S)^2 \geq \left( \frac{1}{\alpha_0^2} - \psi_1^2 \right) d(x_k, S)^2.
\]
Reduce $\psi_1$ if needed so that $1/\alpha_0^2 - \varphi^2\psi_1^2 \geq 1/(2\alpha_0^2)$. This fact, (4.25) and (4.15) yield
\[
\rho_\mathcal{G}(\tilde{p}_k^N) = 1 - \frac{\|\Theta(x_k + \tilde{p}_k^N)\|^2 - \|\Theta\tilde{p}_k^N + \Theta_k\|^2}{\|\Theta\|^2 - \|\Theta\tilde{p}_k^N + \Theta_k\|^2} \\
\geq 1 - \frac{\frac{2\alpha_0^2\varphi\alpha(\alpha_0^2\varphi)\|\tilde{p}_k^N\|^2}{2\varphi_d(x_k, S)^2}\|\tilde{p}_k^N\|^2}{d(x_k, S)^2} \\
\geq 1 - 2\alpha_0^2\varphi\alpha(\alpha_0^2\varphi + 2\varphi)d(x_k, S)^2.
\]

Finally, by (4.5)
\[
\rho_\mathcal{G}(\tilde{p}_k^N) \geq 1 - 2\alpha_0^2\varphi\alpha(\alpha_0^2\varphi + 2\varphi)\|x_k - x^*\|^2.
\]
Hence, $\tilde{p}_k^N$ satisfies both the conditions (3.6) and (3.7) if $x_k$ is sufficiently close to $x^*$, i.e. if $x_k \in B_\psi(x^*)$ for some $\psi \leq \psi_1$. Concerning the value of $d(x_k, S)$, by (4.5) and (4.23), we have $d(x_k, S) \leq \|x_k + \tilde{p}_k^N - [x_k + \tilde{p}_k^N]S\| \leq \|x_k + p_N^k - [x_k + p_N^k]S\|$. Thus, from (4.6) and (4.13)
\[
d(x_k, S) \leq d(x_k + p_N^k, S) \leq \alpha_0\|\Theta(x_k + p_N^k)\| \leq \alpha_0\varphi_d(x_k, S)^2,
\]
and (4.21) holds with $\alpha_2 = \alpha_0\varphi$. \hfill \Box

Now we provide the main result on the behavior of the sequence $\{x_k\}$.

**Theorem 4.2.** Let Assumptions 4.1 and 4.2 hold. Then, the sequence $\{x_k\}$ generated by the TREI-LM method converges to $x^*$ quadratically.

*For the TREI-GN method, the result holds under the additional assumption that $\Theta'(x^*)$ is full rank.*

**Proof.** Let $\psi$ as in Lemma 4.4 and $\zeta \leq \min\{\psi/(1 + 2\alpha_1), 1/(2\alpha_2)\}$. Since $x^*$ is a limit point of $\{x_k\}$, there exists $x_k$ such that $x_k \in B_\zeta(x^*)$.

We begin showing that if $x_k \in B_\zeta(x^*)$ then $x_l \in B_\zeta(x^*)$ for $l > k$. We proceed by induction. First, we show that $x_{k+1} \in B_\zeta(x^*)$. In fact, by (4.23) we have $\|x_{k+1} - x^*\| = \|x_k + \tilde{p}_k^N - x^*\| \leq \zeta + \|p_k^N\|$. Thus by (4.15) and the definition of $\zeta$, we get
\[
\|x_{k+1} - x^*\| \leq \zeta + \alpha_1\|x_k - x^*\| \leq (1 + \alpha_1)\zeta \leq \psi.
\]

Second, we assume $x_{k+1}, \ldots, x_{k+m-1} \in B_\psi(x^*)$, and show that $x_{k+m} \in B_\psi(x^*)$. Note that (4.21) implies
\[
d(x_{k+l}, S) \leq \alpha_2 d(x_{k+l-1}, S)^2 \leq \cdots \leq (\alpha_2)^{2^{l-1}}d(x_k, S)^2 \leq (\alpha_2)^{2^{l-1}}\zeta^2 \leq \zeta\left(\frac{1}{2}\right)^{2^{l-1}},
\]
where the last inequality is due to the choice of $\zeta$. Further, by Lemma 4.4 we have $x_{k+l} = x_{k+l-1} + \tilde{p}_k^{N_{k+l-1}}$ for $l = 1, \ldots, m$.

Thus
\[
\|x_{k+m} - x^*\| \leq \|x_{k+m} - x_{k+m-1}\| + \cdots + \|x_k - x^*\| \\
\leq \sum_{l=0}^{m-1} \|\tilde{p}_k^{N_{k+l}}\| + \zeta \\
\leq \alpha_1\sum_{l=0}^{m-1} d(x_{k+l}, S) + \zeta.
\]

where the last inequality follows from (4.15). Hence
\[
\|x_{k+m} - x^*\| \leq \left(\alpha_1\sum_{l=0}^{m-1} \left(\frac{1}{2}\right)^{2^{l-1}} + 1\right)\zeta \leq \left(\alpha_1\sum_{l=0}^{\infty} \left(\frac{1}{2}\right)^l + 1\right)\zeta = (2\alpha_1 + 1)\zeta \leq \psi.
\]

Now, letting $p > q \geq k$ and proceeding as above we have.
\[ \|x_p - x_q\| \leq \sum_{l=q}^{p-1} \|P_l^N\| \leq \left( \alpha_1 \sum_{l=0}^{\infty} \left( \frac{1}{2} \right) \right) \xi = 2\alpha_1 \xi. \]

This means that \{x_k\} is a Cauchy sequence and hence it converges. Moreover, since \(x^*\) is a limit point we conclude \(\lim_{k \to \infty} x_k = x^*\).

We conclude discussing the convergence rate of \{x_k\}. Let \(k\) sufficiently large so that \(x_{k+j+1} \in B_\varphi(x^*)\) for \(j \geq 0\).

By (4.15) and (4.21) we obtain
\[
\|P_{k+j+1}^N\| \leq \alpha_1 d(x_{k+j+1}, S) \leq \alpha_1 \alpha_2 d(x_{k+j}, S)^2.
\]

Since \(\|x_{k+1} - x^*\| \leq \sum_{j=0}^{\infty} \|P_{k+j+1}^N\|\) then by (4.21) we get
\[
\|x_{k+1} - x^*\| \leq \alpha_1 \alpha_2 \sum_{j=0}^{\infty} d(x_{k+j}, S)^2
\leq \alpha_1 \alpha_2 \left( 1 + \sum_{j=1}^{\infty} (\alpha_2 d(x_k, S))^{2(j+1)-2} \right) d(x_k, S)^2. \tag{4.26}
\]

Finally, (4.5) and (4.26) yield
\[
\frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|^2} \leq \frac{\|x_{k+1} - x^*\|}{d(x_k, S)^2} \leq \alpha_1 \alpha_2 \left( 1 + \sum_{j=1}^{\infty} (\alpha_2 d(x_k, S))^{2(j+1)-2} \right),
\]
and consequently \(\lim sup_{k \to \infty} \|x_{k+1} - x^*\|/\|x_k - x^*\|^2 < \infty\). This shows that \(\{x_k\}\) converges quadratically to \(x^*\). \(\square\)

5. Numerical experiments

This section provides the description of the numerical experience with the TREI-GN and TREI-LM methods. The two algorithms were written in Matlab and run under Matlab 7.0 with machine precision \(\epsilon_m \simeq 2 \times 10^{-16}\). The numerical experiments reported were all run on an Intel Pentium M Processor 740 1700 MHz.

5.1. Test problems

The test problems are the constraint sets of 30 nonlinear programming problems from the handbook of tests [13] by Floudas et al. and the Hock–Schittkowski test collection [20].

The problems were written as the bound-constrained least-squares problem (1.3) where the residual function \(\Theta\) has the form (2.2). Thus, the dimensions of the problems remain unchanged and the reduced model \(\hat{m}_k\) given in (3.13) can be used.

The test problems used are listed in Table 1 and will be referred to be problem number. The first 10 problems are square, problems 11–20 are underdetermined, the remaining problems are overdetermined. Except for problems 7 and 8, the test problems contain simple bounds on the variables. In problems 7 and 8 we added simple bounds on the variables and let the box \(\Omega\) be \(\Omega = \{x \in \mathbb{R}^n, x \geq 0\}\).

5.2. Implementation issues

We describe the implementation of the methods referring to the algorithm sketched in Section 3. The trust-region parameters used in our runs are: \(\Delta_0 = 1\), \(\Delta_{\text{min}} = 10^{-12}\), \(\beta_1 = 0.1\), \(\beta_2 = 0.25\). The reduced model (3.13) is adopted and the Jacobian matrix \(\hat{\Theta}'\) in (3.13) is formed by using finite differences.

In Step 1 the value of \(\mu_k\) is assigned. According to the rule (3.3), the TREI-GN method is free of such parameter. In fact, the method is obtained setting \(\mu = 0\) and consequently \(\mu_k = 0\) for all \(k\). On the contrary, for the TREI-LM method we chose \(\mu_k\) as in the paper [25], i.e. we fixed the scalars \(\mu_1 = 1\), \(\hat{\mu} = 10^{-8}\). Thus we have \(\mu_0 = 10^{-8}\|\Theta_0\|^2\), \(\mu_k = \min\{\mu_{k-1}, \|\Theta_k\|^2\}, k > 0\), and the scalar \(\mu_k\) is positive unless \(\|x_k\|\) is a solution to problem (1.3). In Step 2, the computation of the vector \(p_k^N\) in (3.8) is performed using the Matlab backslash operator if \(m = n\) and \(\Theta_k^t\) is
nonsingular; otherwise it is done using the singular value decomposition of \( \Theta' \). The vector \( p^N_k \) defined in (3.9) is computed by the QR decomposition applied to the least-squares problem (3.12).

Regarding Steps 6–7, if \( \tilde{p}_{tr} \) does not satisfy condition (3.6) we find the scalar \( t \in (0, 1) \) such that \( \rho_c(p_k^C + (1 - t) \tilde{p}_{tr}) = \beta_1 \) and set \( p_k = tp_k^C + (1 - t) \tilde{p}_{tr} \). Then, the trust-region update is performed as follows: if the step \( p_k \) fails to satisfy (3.7) the trust-region radius is reduced setting \( \Delta_k = \min\{\Delta_k/4, \|p_k\|/2\} \); if the step \( p_k \) satisfies (3.7) and \( \rho_\theta(p_k) \geq 0.75 \) we set \( \Delta_{k+1} = \max\{\Delta_k, 2\|p_k\|, \Delta_{\min}\} \), otherwise we let \( \Delta_{k+1} = \max\{\Delta_k, \Delta_{\min}\} \).

Successful termination of the algorithms means that they return an approximation to a zero residual solution for the problem (1.3). In practice we stop the algorithms when \( \|\Theta_k\| \leq 10^{-6} \). A failure is declared when a stationary nonzero residual point for the problem (1.3) is found, i.e. \( \|\Theta_k\| > 10^{-6} \) whereas \( D_k \nabla \Theta_k \leq 100 \epsilon_m \) or \( \|\Theta_{k+1} - \Theta_k\| \leq 100 \epsilon_m \|\Theta_k\| \). Moreover, the algorithms fail when the trust-region radius is less than \( \Delta_{\min} \) or the number of iterations is greater than 300.

### 5.3. Numerical results

All problems were run starting from three different initial guesses \( x_0 \) whose \( i \)th component, \( i = 1, \ldots, n \), has the form

\[
(x_0)_i = \begin{cases} 
  l_i + c(u_i - l_i)/2 & \text{if } l_i > -\infty \text{ and } u_i < \infty, \\
  l_i + c10^\epsilon & \text{if } l_i > -\infty \text{ and } u_i = \infty, \\
  u_i - c10^\epsilon & \text{if } l_i = -\infty \text{ and } u_i < \infty, \\
  10(c-1) & \text{if } l_i = -\infty \text{ and } u_i = \infty,
\end{cases}
\] (5.1)

for \( c = 0, 1, 2 \). If the vector \( x_0 \) is a solution to (1.3), we perturb the scalar \( c \).

The results of the successful runs are summarized in Table 2. The first column of this table gives the problem number, the second column gives the number of variables of the problem, the third and fourth columns give the number of equalities and inequalities respectively. Furthermore, for both algorithms we show the number of failures occurred starting from the three different initial guesses (5.1), the average number of iterations and the average number of \( \Theta \)-evaluations needed to solve the problems from different starting points. We remark that the number of \( \Theta \)-evaluations does not include \( \Theta \)-evaluations needed to compute the Jacobian matrices by finite differences.

Both methods showed to be reliable and quite insensitive to the choice of the starting point. In fact, on a total of 90 runs the methods TREI-GN and TREI-LM solved 80 and 77 tests respectively. The ability of the methods to handle bounds is supported by the fact that an active solution of (1.3) was computed in 20 runs for the TREI-GN method and in 21 runs for the TREI-LM method.

Most problems were solved with a low number of function evaluations and this number is, on average, favorable for the TREI-GN method. In particular, Table 2 shows that \( \text{Afe} \) is larger than 25 only 7 times when the TREI-GN
method is used and 12 times when the TREI-LM method is used. Moreover, the maximum value of the difference between the numbers $\text{Ait}$ and $\text{Afe}$ is 5 and this occurs for two problems only.

To give more detailed indication on the behavior of our approaches, we focus on the problems where failures occurred or very expensive runs were performed. For these problems, in Table 3 we report the scalar $c$ used in (5.1) to compute the starting point, the number $\text{it}$ of iterations and the number $\text{fe}$ of $\Theta$-evaluations performed to compute an approximation to a zero residual solution. Clearly, if a run was not successful the values $\text{it}$, $\text{fe}$ are omitted.

The results showed in Table 3 confirm that typically the methods require a small number of function evaluations. Also, the high average numbers $\text{Afe}$ in the Table 2 are due to a single expensive run. We remark that the cost of the methods may depend on a slow convergence rate; this is the case when the conditions required for quadratic convergence are not satisfied. However, it is worthy noting that in the very expensive runs of problems 9, 18, 19 and 20 most of iterations were performed in the globalization phase.

Finally, to compare the overall computational effort of our methods we plot the performance profile proposed by Dolan and Moré [11]. We considered the 90 tests performed by each algorithm. For each test $T$ and algorithm $A$, we let $\text{fe}_{T,A}$ denote the number of $\Theta$-evaluations required to solve test $T$ by the algorithm $A$ and $\text{fe}_T$ be the lowest number of $\Theta$-evaluations required by the two algorithms to solve test $T$. Then, for the algorithm $A$ the performance profile is defined as

$$
\pi(\tau) = \frac{\text{number of tests } s.t. \text{ fe}_{T,A}/\text{fe}_T \leq \tau}{\text{number of tests}}, \quad \tau \geq 1.
$$

Please cite this article in press as: M. Macconi et al., Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities, Applied Numerical Mathematics (2008), doi:10.1016/j.apnum.2008.03.028
### Table 3
Computational results for some problems

<table>
<thead>
<tr>
<th>Prob.#</th>
<th>c</th>
<th>it</th>
<th>fe</th>
<th>Prob.#</th>
<th>c</th>
<th>it</th>
<th>fe</th>
<th>Prob.#</th>
<th>c</th>
<th>it</th>
<th>fe</th>
<th>Prob.#</th>
<th>c</th>
<th>it</th>
<th>fe</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>17</td>
<td>0</td>
<td>37</td>
<td>38</td>
<td>40</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>7</td>
<td>25</td>
<td>29</td>
<td>1</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>196</td>
<td>197</td>
<td>196</td>
<td>197</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.5</td>
<td>–</td>
<td>–</td>
<td>18</td>
<td>0</td>
<td>82</td>
<td>94</td>
<td>144</td>
<td>145</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>10</td>
<td>12</td>
<td>2</td>
<td>15</td>
<td>16</td>
<td>16</td>
<td>15</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>21</td>
<td>–</td>
<td>2</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0</td>
<td>9</td>
<td>10</td>
<td>19</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>21</td>
<td>20</td>
<td>21</td>
<td>1</td>
<td>24</td>
<td>25</td>
<td>37</td>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>35</td>
<td>–</td>
<td>2</td>
<td>117</td>
<td>118</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>13</td>
<td>14</td>
<td>20</td>
<td>0</td>
<td>12</td>
<td>13</td>
<td>12</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>15</td>
<td>1</td>
<td>21</td>
<td>22</td>
<td>130</td>
<td>131</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>24</td>
<td>25</td>
<td>222</td>
<td>2</td>
<td>41</td>
<td>42</td>
<td></td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>22</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>17</td>
<td>18</td>
<td>17</td>
<td>18</td>
<td>1</td>
<td>27</td>
<td>33</td>
<td>36</td>
<td>42</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>–</td>
<td>–</td>
<td>2</td>
<td>20</td>
<td>21</td>
<td>39</td>
<td>41</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>40</td>
<td>49</td>
<td>24</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>18</td>
<td>19</td>
<td>68</td>
<td>69</td>
<td>0.5</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>23</td>
<td>–</td>
<td>1</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td>31</td>
<td>26</td>
<td>0.5</td>
<td>9</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>13</td>
<td>14</td>
<td>13</td>
<td>14</td>
<td>1</td>
<td>16</td>
<td>17</td>
<td>16</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>196</td>
<td>197</td>
<td>196</td>
<td>197</td>
<td>2</td>
<td>25</td>
<td>26</td>
<td>104</td>
<td>105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 1 shows the function evaluations performance profile for the successful runs. The difference between the performance profiles of the two methods is modest but on the whole the TREI-GN method can be considered more efficient than the TREI-LM method. In particular, the TREI-GN method is the most efficient algorithm for about 84% of the runs and the TREI-LM method is within a factor 2 of the TREI-GN method for about 80% of the runs.

So far we have provided a description of the successful runs, i.e. the runs where the methods converge to a zero-residual solution of the least-squares problem (1.3). Concerning the failures, it is important to note that there are occurrences where the methods converged to a first-order stationary point for (1.3) that is not a solution for problem (1.1). These failures occurred 3 and 5 times for TREI-GN and TREI-LM method respectively. Therefore,
considering the overall ability of the methods to compute a first-order stationary point for (1.3), the TREI-GN and TREI-LM methods solved 92% and 91% of the tests respectively.

Summarizing the results presented we conclude that the methods are reliable in solving the problem (1.3) and that in most cases a solution for problem (1.1) is computed.

6. Conclusions and perspectives

In this paper we have presented two new trust-region methods for solving systems of equalities and inequalities. They are inspired by recent developments in trust-region methods for nonlinear optimization with simple bounds. The new methods are shown to be globally convergent and to have strong local convergence properties under an error bound condition. The preliminary numerical experience presented shows that the methods work quite well in practice.

The implementations of our methods have been discussed in the above sections but other implementations can be made on the base of the problems considered. To this end, we remark that the trust-region Gauss–Newton method requires the minimum norm solution of a linear least-squares solution at each iteration. For large-scale problems the use of iterative methods can reduce the cost of this task. On the other hand, the trust-region Levenberg–Marquardt method does not require computing the minimum norm solution but its reliability and efficiency depend on the choice of the sequence of scalars (3.3).

We plan more research in the design of the globalization technique. In fact, it has been recently shown that filter techniques produce significant gains in reliability and efficiency compared to classical trust-region approaches [17,18]. Therefore, the combination of the filter techniques with our trust-region approach will be the subject of a follow-up paper.

Acknowledgements

The authors would like to thank the two anonymous referees for their detailed suggestions and comments.

References