

# SUBSPACE TRUST-REGION METHODS FOR LARGE BOUND-CONSTRAINED NONLINEAR EQUATIONS \*

STEFANIA BELLAVIA, BENEDETTA MORINI †

**Abstract.** Trust-region methods for solving large bound-constrained nonlinear systems are considered. They allow for spherical or elliptical trust-regions where the search of an approximate solution is restricted to a low dimensional space. A general formulation for these methods is introduced and global and superlinear/quadratic convergence is shown under standard assumptions. Viable approaches for implementation in conjunction with Krylov methods are discussed.

**Key words.** bound-constrained nonlinear systems, subspace trust-region methods, inexact Newton step, Krylov subspace methods.

**1. Introduction.** A number of applications arising in chemical engineering [17, 18], power engineering [34], PDE-constrained optimization [4] are naturally stated as large constrained nonlinear systems. In particular, systems where the variables are subjected to lower and upper bounds are fairly general because sets of algebraic equations and inequalities and the Karush-Kuhn-Tucker (KKT) systems can be cast in such form.

This paper is concerned with the development of a trust-region method for solving large bound-constrained nonlinear systems

$$(1.1) \quad F(x) = 0, \quad x \in \Omega.$$

Here  $F : X \rightarrow \mathbb{R}^n$  is a continuously differentiable mapping,  $X \subseteq \mathbb{R}^n$  is an open set containing the feasible region  $\Omega$  and  $\Omega$  is an  $n$ -dimensional box,  $\Omega = \{x \in \mathbb{R}^n : l \leq x \leq u\}$ . These inequalities are meant component-wise and  $l \in (\mathbb{R} \cup -\infty)^n$ ,  $u \in (\mathbb{R} \cup \infty)^n$ .

The development of globally convergent methods for large unconstrained nonlinear systems has received a great attention, see e.g. [2, 6, 7, 13, 23, 24]. The methods proposed suggest that the search directions employed in the global strategy might belong to a low dimensional subspace as such directions may often be computed cheaply. Therefore, they avoid the factorization of the matrices involved and consider the combination of global strategies such as linesearch techniques and model trust-region algorithms with Krylov methods. The resulting procedures belong to the framework of the inexact Newton methods [11].

More insight into trust-region methods, their main computational effort is the solution of the so-called trust-region problem. This is to find the minimizer of some model of the objective function within a region where the model adequately mimics the objective function. For small and medium problems, solving the trust-region problem relies on matrix factorization. When  $n$  is large several authors have suggested to restrict the search of an approximate solution of such problem to a low dimensional subspace. Thus, the full space trust-region problem is replaced with a subspace trust-region problem and high overhead of computing is avoided, see [9] and references therein. Proposed approaches include the truncated Conjugate Gradient method [29,

---

\* Work supported by MIUR, Rome, Italy, through “Cofinanziamenti Programmi di Ricerca Scientifica di Interesse Nazionale” and “Gruppo Nazionale per il Calcolo Scientifico”, Florence, Italy.

† Dipartimento di Energetica “S. Stecco”, Università di Firenze, via C. Lombroso 6/17, 50134 Firenze, Italia, e-mail: stefania.bellavia@unifi.it, benedetta.morini@unifi.it

30], the truncated Lanczos approach [16], the two dimensional subspace minimization [5, 8] and a subspace dogleg method [6, 7].

Numerical methods for problem (1.1) differ from the procedures for unconstrained nonlinear systems in several respects. They are augmented with strategies that enforce feasibility of the iterates. Additionally, major modifications in the globalization techniques are necessary. We are aware of the trust-region methods [1, 3, 15, 31, 32] tailored for small and medium size problems and of the procedures [25, 26] appropriate for large systems. Focusing on the approaches for large problems, Qi, Tong and Li [25] propose an active set projected trust-region algorithm for bound-constrained nonlinear systems. Due to the active set strategy, the trust-region problem may be of reduced dimension which is potentially cheaper when the method is applied to large problems. The method by Qi, Qi and Sun [26] concerns the solution of the KKT systems. The trust-region problem is built around those components of the current iterates which are far from the boundary of the positive orthant and it is solved by the truncated Conjugate Gradient method.

In this paper we introduce a prototype subspace trust-region method for (1.1). Our proposal is to investigate the idea of solving the trust-region problem in a small subspace while still attaining global and locally fast convergence. Both spherical and elliptical trust-regions are allowed. To ensure global convergence properties we use a generalized Cauchy step and ideas from the method recently proposed by the authors in [3]. Fast local convergence relies on mild conditions on the subspace and it is independent from the way of computing an approximate trust-region solution. At each iteration the trial step used to compute the new iterate is a linear combination of the generalized Cauchy step and the approximate trust-region solution.

The general scheme proposed serves as a paradigm for some specific implementations. In particular, the theoretical results obtained suggest ways to implement it by using Krylov solvers [27]. The first proposal is a two dimensional subspace strategy. The second is a dogleg subspace strategy in conjunction with the iterative linear solver GMRES. Both strategies compute an approximate solution of the related subspace trust-region problem with a low computational cost and require matrix-vector products only. At this regard, we remark that the computation of the generalized Cauchy point calls for the product of the transpose of the Jacobian of  $F$  with vectors. Then, the proposed strategies cannot be implemented in a matrix free manner, i.e. without computing the whole Jacobian matrix. On the other hand if the Jacobian of  $F$  is not available, these products can be effectively computed by using software for automatic differentiation [33].

We mention that [6, 7] propose a matrix free Newton-GMRES dogleg strategy for unconstrained nonlinear systems where the Cauchy point is replaced by the steepest descent direction in a space generated by GMRES. But, in our opinion, this approach is not easy to be extended to constrained problems.

In §2 we describe the main features a trust-region method for problem (1.1) and in §3 we propose a prototype method for large problems. In §4 we provide global and local convergence properties. In §5 we discuss ways in which an implementation of our procedure may be developed.

**1.1. Notation.** Throughout the paper we use the following notation. For any mapping  $F : X \rightarrow \mathbb{R}^n$ , differentiable at a point  $x \in X \subset \mathbb{R}^n$ , the Jacobian matrix of  $F$  at  $x$  is denoted by  $F'(x)$  and  $F(x_k)$  is denoted by  $F_k$ . To represent the  $i$ -th component of  $x$  the symbol  $(x)_i$  is used but, when clear from the context, the brackets are omitted. For any vector  $y \in \mathbb{R}^n$ , the 2-norm is denoted by  $\|y\|$  and the open ball

with center  $y$  and radius  $\rho$  is indicated by  $B_\rho(y)$ , i.e.  $B_\rho(y) = \{x : \|x - y\| < \rho\}$ . The identity matrix of dimension  $n$  is denoted by  $I$ .

**2. Preliminaries.** In this section we provide the essential features of a trust-region method for the solution of (1.1). The sequence  $\{x_k\}$  generated is expected to converge to a point which solves the optimization problem

$$(2.1) \quad \min_{x \in \Omega} f(x) = \min_{x \in \Omega} \frac{1}{2} \|F(x)\|^2.$$

In fact, the solutions to (1.1) solve the constrained minimization problem (2.1). A solution  $x^*$  of (2.1) satisfies

$$(2.2) \quad D^{-2}(x^*) \nabla f(x^*) = 0,$$

where  $\nabla f(x) = F'(x)^T F(x)$ ,  $D(x)$  is the diagonal scaling matrix

$$(2.3) \quad D(x) = \text{diag}(|v_1(x)|^{-1/2}, |v_2(x)|^{-1/2}, \dots, |v_n(x)|^{-1/2}),$$

and

$$v_i(x) = \begin{cases} x_i - u_i & \text{if } (\nabla f(x))_i < 0 \quad \text{and } u_i < \infty \\ x_i - l_i & \text{if } (\nabla f(x))_i > 0 \quad \text{and } l_i > -\infty \\ \min\{x_i - l_i, u_i - x_i\} & \text{if } (\nabla f(x))_i = 0 \quad \text{and } l_i > -\infty \text{ or } u_i < \infty \\ 1 & \text{otherwise} \end{cases}$$

for  $i = 1, \dots, n$ , see [10].

Numerical methods for (1.1) need well-angled directions to handle the bounds. In particular, search directions biased towards the interior of  $\Omega$  are required. This way, sufficiently long steps along such directions are allowed before violating the constraints. If  $x_k$  lies in  $\Omega$ , the following scaled gradient of  $f$

$$d_k = -D_k^{-2} \nabla f_k,$$

is well-angled with respect to the bounds. This is due to the fact that  $D_k^{-2}$  penalizes the step  $\nabla f_k$  preventing a step directly toward a boundary point. Moreover, by (2.2)  $d_k$  monitors the progress toward a solution of problem (2.1).

In a framework for (1.1), we consider the trust-region problem

$$(2.4) \quad \min_{p \in \mathbb{R}^n} \{m_k(p) : \|G_k p\| \leq \Delta_k\},$$

where  $\Delta_k$  is the used trust-region radius,  $m_k$  is the quadratic model for  $f$  at  $x_k$

$$m_k(p) = \frac{1}{2} \|F_k + F_k' p\|^2,$$

and  $G_k = G(x_k) \in \mathbb{R}^{n \times n}$  with  $G : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$  is defined as

$$G(x) = I, \quad \text{or} \quad G(x) = D(x).$$

The first choice of  $G$  is used in [25, 31, 32] and it yields the standard spherical trust-region problem. The choice  $G(x) = D(x)$ ,  $x \in \mathbb{R}^n$ , has been considered in [3, 1] and gives rise to an elliptical trust-region. In this case, for decreasing values of  $\Delta_k$ , the solution to problem (2.4) tends to become parallel to  $d_k$ .

For all the proposed methods, the iterates  $x_k$  are forced to belong to  $\Omega$  in order to deal with problems where  $F$  is not defined outside  $\Omega$ . Moreover, since  $D$  is not defined on the boundary of  $\Omega$ , the methods given in [3, 1] generate strictly feasible iterates,  $x_k \in \text{int}(\Omega) = \{x \in \mathbb{R}^n : l < x < u\}$ .

To find the next iterate, a key role is played by a so-called generalized Cauchy step  $p_c(\Delta_k)$  depending on the scaled gradient  $d_k$ . The vector  $p_c(\Delta_k)$  has the form

$$(2.5) \quad p_c(\Delta_k) = \tau_k d_k,$$

and it is such that  $x_k + p_c(\Delta_k) \in \text{int}(\Omega)$ . The value of  $\tau_k$  is fixed as follows. Consider

$$(2.6) \quad \hat{\tau}_k = \min \left\{ \frac{\|D_k^{-1} \nabla f_k\|^2}{\|F_k' D_k^{-2} \nabla f_k\|^2}, \frac{\Delta_k}{\|G_k D_k^{-2} \nabla f_k\|} \right\} = \underset{\|\tau G_k d_k\| \leq \Delta_k}{\operatorname{argmin}} m_k(\tau d_k).$$

If  $x_k + \hat{\tau}_k d_k \in \text{int}(\Omega)$ , we let  $\tau_k = \hat{\tau}_k$  in (2.5). Otherwise we let  $\lambda_k$  be the stepsize along  $d_k$  to the boundary, i.e.

$$\lambda_k = \min_{1 \leq i \leq n} \Lambda_i \quad \text{where} \quad \Lambda_i = \begin{cases} \max \left\{ \frac{l_i - (x_k)_i}{(d_k)_i}, \frac{u_i - (x_k)_i}{(d_k)_i} \right\} & \text{if } (d_k)_i \neq 0 \\ \infty & \text{if } (d_k)_i = 0 \end{cases},$$

and set  $\tau_k$  smaller than  $\lambda_k$ . Summarizing, the parameter  $\tau_k$  is given by

$$(2.7) \quad \tau_k = \begin{cases} \hat{\tau}_k & \text{if } x_k + \hat{\tau}_k d_k \in \text{int}(\Omega) \\ \theta \lambda_k & \theta \in (0, 1), \text{ otherwise.} \end{cases}$$

In [10], it has been shown that global convergence to a first-order stationary point of (2.1) depends on obtaining, at each iteration, at least as much decrease in  $m_k$  as a fixed fraction of the decrease attained by the generalized Cauchy step  $p_c(\Delta_k)$ . In particular, letting  $p(\Delta_k)$  be the step taken to update  $x_k$ ,  $p(\Delta_k)$  must satisfy the following condition

$$(2.8) \quad \rho_c(p(\Delta_k)) = \frac{m_k(0) - m_k(p(\Delta_k))}{m_k(0) - m_k(p_c(\Delta_k))} \geq \beta_1,$$

for a given constant  $\beta_1 \in (0, 1)$ .

Finally, as for the unconstrained problems, the sufficient improvement condition

$$(2.9) \quad \rho_f(p(\Delta_k)) = \frac{f(x_k) - f(x_k + p(\Delta_k))}{m_k(0) - m_k(p(\Delta_k))} \geq \beta_2,$$

is required to hold for a given constant  $\beta_2 \in (0, 1)$ . Namely, if (2.9) is satisfied, then  $p(\Delta_k)$  is accepted, the new iterate  $x_{k+1} = x_k + p(\Delta_k)$  is formed and the trust-region radius may be increased. Otherwise,  $p(\Delta_k)$  is rejected and  $\Delta_k$  is shrunk.

**3. A paradigm method for large-scale problems.** In this section we present a general trust-region scheme for large bound-constrained nonlinear systems. Since the main source of computational effort of a trust-region algorithm is the work for solving problem (2.4), we replace (2.4) by the following subspace trust-region problem

$$(3.1) \quad \min_{p \in S_k} \{m_k(p) : \|G_k p\| \leq \Delta_k\}.$$

In fact, for a small subspace  $S_k$  of  $\mathbb{R}^n$  the solution of (3.1) can be computed cheaply. At each iteration our scheme includes: the choice of the subspace  $S_k$ , the solution of

the subspace trust-region problem (3.1), the construction of a step which combines the generalized Cauchy step and the subspace trust-region solution.

Our subspace model trust-region approach is based upon finding a small dimension subspace  $S_k$  of  $\mathbb{R}^n$  such that the minimum value of  $m_k$  on  $S_k$  is a fraction of  $m_k(0)$ . In particular, we fix  $S_k$  so that

$$(3.2) \quad p_k = \underset{p \in S_k}{\operatorname{argmin}} m_k(p),$$

satisfies

$$(3.3) \quad m_k(p_k) \leq \eta_k^2 m_k(0),$$

with  $\eta_k \in [0, 1)$ . Clearly, by (3.3)

$$(3.4) \quad F'_k p_k = -F_k + r_k, \quad \|r_k\| \leq \eta_k \|F_k\|,$$

i.e.  $p_k$  is an inexact Newton step for the problem  $F(x) = 0$ , [11].

To provide a flexible scheme, here we deliberately not specify how to determine  $S_k$  and the solution  $p_{tr}(\Delta_k)$  to (3.1). In §5 we will show that these tasks can be readily implemented in different ways.

In regard to the problem of finding the actual step from  $x_k$ , our aim is to find a step  $p(\Delta_k)$  producing a strictly feasible iterate  $x_{k+1} = x_k + p(\Delta_k)$  and a sufficient reduction in the values of both the model function  $m_k$  and the objective function  $f$ . We proceed following the ideas of [3].

If the point  $x_k + p_{tr}(\Delta_k)$  is not feasible, we form a vector  $\bar{p}_{tr}(\Delta_k)$  such that  $x_k + \bar{p}_{tr}(\Delta_k) \in \operatorname{int}(\Omega)$ . The vector  $\bar{p}_{tr}(\Delta_k)$  is defined component-wise as follows

$$(3.5) \quad (\bar{p}_{tr}(\Delta_k))_i = \begin{cases} \min\{(1-\alpha)(l-x_k)_i, (2(l-x_k)-p_{tr}(\Delta_k))_i\} & \text{if} \\ (x_k+p_{tr}(\Delta_k))_i \leq l_i \\ (p_{tr}(\Delta_k))_i & \text{if } l_i < (x_k+p_{tr}(\Delta_k))_i < u_i \\ \max\{(1-\alpha)(u-x_k)_i, (2(u-x_k)-p_{tr}(\Delta_k))_i\} & \text{if} \\ (x_k+p_{tr}(\Delta_k))_i \geq u_i \end{cases}$$

for  $i = 1, 2, \dots, n$ ,  $\alpha \in (0, 1)$ . Clearly,  $(p_{tr}(\Delta_k))_i$  and  $(\bar{p}_{tr}(\Delta_k))_i$  have the same sign and  $\bar{p}_{tr}(\Delta_k) = p_{tr}(\Delta_k)$  if  $x_k + \bar{p}_{tr}(\Delta_k)$  lies in the interior of  $\Omega$ . In fact, the point  $x_k + \bar{p}_{tr}(\Delta_k)$  can be interpreted as the projection of  $x_k + p_{tr}(\Delta_k)$  onto  $\Omega$ ,  $P_\Omega(x_k + p_{tr}(\Delta_k)) = \max\{l, \min\{x_k + p_{tr}(\Delta_k), u\}\}$ , followed by a small step toward the interior of  $\Omega$ .

Then, we set

$$(3.6) \quad p(\Delta_k) = t p_c(\Delta_k) + (1-t) \bar{p}_{tr}(\Delta_k) \quad t \in [0, 1),$$

and seek the value of  $t$  such that  $p(\Delta_k)$  satisfies (2.8). Specifically, if  $\rho_c(\bar{p}_{tr}(\Delta_k)) \geq \beta_1$ , we take  $t = 0$ . Otherwise, since  $m_k(t p_c(\Delta_k) + (1-t) \bar{p}_{tr}(\Delta_k))$  is a quadratic function in  $t$  and  $\rho_c(p_c(\Delta_k)) = 1$ , it is easy to see that there exists  $t \in (0, 1)$  such that  $p(\Delta_k)$  satisfies  $\rho_c(p(\Delta_k)) = \beta_1$ . Next, we summarize the process for finding  $p(\Delta_k)$ .

ALGORITHM I. FINDING A STEP THAT SATISFIES THE MODEL DECREASE (2.8).

Input parameters:  $x_k \in \operatorname{int}(\Omega)$ ,  $\Delta_k > 0$ ,  $\bar{p}_{tr}(\Delta_k)$ ,  $p_c(\Delta_k)$ .

1. If  $\rho_c(\bar{p}_{tr}(\Delta_k)) \geq \beta_1$   
Set  $t = 0$ .

Else

$$\text{Compute } u_1 = F'_k p_c(\Delta_k), u_2 = F'_k \bar{p}_{tr}(\Delta_k), u = u_1 - u_2, z = -F_k - u_2,$$

$$w = \left( (z^T u)^2 - 2 \|u\|^2 (F_k^T (u_2 - \beta_1 u_1) + \frac{1}{2} \|u_2\|^2 - \frac{\beta_1}{2} \|u_1\|^2) \right)^{\frac{1}{2}}.$$

$$\text{Set } t = (z^T u - w) / \|u\|^2.$$

2. Compute  $p(\Delta_k)$  by (3.6).

We point out that the inclusion of  $p_c(\Delta_k)$  in (3.6) ensures the existence of a vector  $p(\Delta_k)$  satisfying (2.8). Then, it enables the method to converge globally. Furthermore, the use of  $\bar{p}_{tr}(\Delta_k)$  in (3.6) and suitable choices of the sequence  $\{\eta_k\}$  yield rapid local convergence. The convergence analysis provided in the next section will highlight these features.

Below we summarize the overall procedure named Subspace Interior Affine Trust-Region (SIATR) method. Note that  $\bar{\Delta}_k$  is the initial value of the trust-region radius at  $k$ th iteration. It is larger than or equal to a fixed threshold  $\Delta_{min} > 0$  for all  $k \geq 0$ . This feature will be crucial for the convergence analysis.

#### SIATR METHOD

Input parameters: the starting point  $x_0 \in \text{int}(\Omega)$ , the function  $G$ , the initial trust-region size  $\Delta_0 > 0$ ,  $\Delta_{min} > 0$ ,  $\alpha, \beta_1, \beta_2, \delta, \theta \in (0, 1)$ .

For  $k = 0, 1, \dots$

1. Set  $\bar{\Delta}_k = \max\{\Delta_{min}, \Delta_k\}$ , set  $\Delta_k = \bar{\Delta}_k / \delta$ .
2. Choose  $\eta_k \in [0, 1)$ .
3. Repeat
  - 3.1 Set  $\Delta_k = \delta \Delta_k$ .
  - 3.2 Find  $S_k \subset \mathbb{R}^n$  s.t. (3.3) holds.
  - 3.3 Compute the solution  $p_{tr}(\Delta_k)$  to (3.1).
  - 3.4 Form  $\bar{p}_{tr}(\Delta_k)$  by (3.5).
  - 3.5 Compute  $p_c(\Delta_k)$  by (2.5) and (2.7).
  - 3.6 Find  $p(\Delta_k)$  by Algorithm I.

Until  $\rho_f(p(\Delta_k)) \geq \beta_2$

4. Set  $x_{k+1} = x_k + p(\Delta_k)$ .
5. Choose  $\Delta_{k+1}$ .

**4. Convergence analysis.** In this section we develop a theoretical foundation for the SIATR method. We assume that

Assumption 1:  $F'$  is Lipschitz continuous in  $L = \cup_{k=0}^{\infty} \{x \in X : \|x - x_k\| \leq r\}$ ,  $r > 0$ , with constant  $2\gamma_L$ .

Assumption 2:  $\|F'(x)\|$  is bounded above on  $L$  and  $\chi_J = \sup_{x \in L} \|F'(x)\|$ .

We begin studying the features of  $p(\Delta_k)$ , provided that  $F'_k$  is nonsingular. First, [3, Lemma 3.2] shows that condition (2.9) is met after a finite number of repeats in Step 3. Second, by Algorithm I we have  $\rho_c(p(\Delta_k)) \geq \rho_c(\bar{p}_{tr}(\Delta_k))$  i.e.

$$(4.1) \quad m_k(p(\Delta_k)) \leq m_k(\bar{p}_{tr}(\Delta_k)).$$

Third, the decrease attained in the value of  $m_k$  by  $p(\Delta_k)$  is given in the following result.

LEMMA 4.1. *If  $p(\Delta_k)$  satisfies (2.8) then*

$$m_k(0) - m_k(p(\Delta_k)) \geq \frac{1}{2}\beta_1 \|D_k^{-1}\nabla f_k\| \min\left\{\frac{\Delta_k}{\|G_k D_k^{-1}\|}, \frac{\|D_k^{-1}\nabla f_k\|}{\|F'_k D_k^{-1}\|^2}, \frac{\theta\|D_k^{-1}\nabla f_k\|}{\|\nabla f_k\|_\infty}\right\}.$$

*Proof.* To prove the thesis we provide a lower bound for  $m_k(0) - m_k(p_c(\Delta_k))$ .

If  $p_c(\Delta_k) = \hat{\tau}_k d_k$  and  $\hat{\tau}_k = \|D_k^{-1}\nabla f_k\|^2 / \|F'_k D_k^{-2}\nabla f_k\|^2$  we get

$$\begin{aligned} m_k(0) - m_k(p_c(\Delta_k)) &= \frac{1}{2} \frac{\|D_k^{-1}\nabla f_k\|^4}{\|F'_k D_k^{-2}\nabla f_k\|^2} \\ &= \frac{1}{2} \frac{\|D_k^{-1}\nabla f_k\|^4}{(D_k^{-1}\nabla f_k)^T (D_k^{-1} F_k'^T F'_k D_k^{-1}) (D_k^{-1}\nabla f_k)} \\ (4.2) \quad &\geq \frac{1}{2} \frac{\|D_k^{-1}\nabla f_k\|^2}{\|D_k^{-1} F_k'^T F'_k D_k^{-1}\|}. \end{aligned}$$

If  $p_c(\Delta_k) = \hat{\tau}_k d_k$  with  $\hat{\tau}_k = \Delta_k / \|G_k D_k^{-2}\nabla f_k\|$ , then  $\hat{\tau}_k \leq \|D_k^{-1}\nabla f_k\|^2 / \|F'_k D_k^{-2}\nabla f_k\|^2$ . Therefore, we get

$$\begin{aligned} m_k(0) - m_k(p_c(\Delta_k)) &= \hat{\tau}_k (\|D_k^{-1}\nabla f_k\|^2 - \frac{1}{2}\hat{\tau}_k \|F'_k D_k^{-2}\nabla f_k\|^2) \\ &\geq \frac{1}{2}\hat{\tau}_k \|D_k^{-1}\nabla f_k\|^2 \\ &= \frac{1}{2} \frac{\Delta_k}{\|G_k D_k^{-2}\nabla f_k\|} \|D_k^{-1}\nabla f_k\|^2 \\ (4.3) \quad &\geq \frac{1}{2} \frac{\Delta_k}{\|G_k D_k^{-1}\|} \|D_k^{-1}\nabla f_k\|. \end{aligned}$$

Finally, consider the case where  $p_c(\Delta_k) = \theta\lambda_k d_k$ . By construction  $\lambda_k \leq \hat{\tau}_k \leq \frac{\|D_k^{-1}\nabla f_k\|^2}{\|F'_k D_k^{-2}\nabla f_k\|^2}$ . Thus, we obtain

$$m_k(0) - m_k(p_c(\Delta_k)) = \theta\lambda_k (\|D_k^{-1}\nabla f_k\|^2 - \frac{1}{2}\theta\lambda_k \|F'_k D_k^{-2}\nabla f_k\|^2) \geq \frac{1}{2}\theta\lambda_k \|D_k^{-1}\nabla f_k\|^2.$$

Recalling that  $\lambda_k \geq 1/\|\nabla f_k\|_\infty$ , see [10, Lemma 3.1], we conclude

$$(4.4) \quad m_k(0) - m_k(p_c(\Delta_k)) \geq \frac{1}{2}\theta \frac{\|D_k^{-1}\nabla f_k\|^2}{\|\nabla f_k\|_\infty}.$$

From (2.8), (4.2), (4.3) and (4.4) the thesis follows.  $\square$

Now we can formalize the global convergence properties of the SIATR method. They essentially derive from forcing (2.8) and can be easily proved following the lines of [3, Theorem 3.1] and using Lemma 4.1.

THEOREM 4.1. *If the sequence of iterates  $\{x_k\}$  generated by the SIATR method is bounded and  $F'_k$  is nonsingular for  $k \geq 0$ , then all the limit points of  $\{x_k\}$  are stationary points for the problem (2.1) i.e.*

$$\lim_{k \rightarrow \infty} \|D_k^{-1}\nabla f_k\| = 0.$$

Further, if there exists a limit point  $x^* \in \text{int}(\Omega)$  of  $\{x_k\}$  such that  $F'(x^*)$  is nonsingular, then  $\|F_k\| \rightarrow 0$  and all the accumulation points of  $\{x_k\}$  solve problem (1.1).

Now, we focus on the local convergence behaviour of the SIATR method. We first collect two technical results that will be useful in the sequel.

LEMMA 4.2. *Let  $x^* \in \Omega$  be a limit point of the sequence of iterates  $\{x_k\}$  generated by the SIATR method such that  $F(x^*) = 0$  and  $F'(x^*)$  is nonsingular. Let  $\Gamma > 0$  be given and  $K = \|F'(x^*)^{-1}\|$ . Then, there exists  $\rho > 0$  so that if  $\|x - x^*\| \leq \rho$  then  $x \in L$  and*

$$(4.5) \quad \|x - x^*\| \leq 2K \|F(x)\|, \quad \forall x \in B_\rho(x^*),$$

$$(4.6) \quad \|F(x)\| \leq 2\|F'(x^*)\| \|x - x^*\|, \quad \forall x \in B_\rho(x^*),$$

$$(4.7) \quad \|F'(x)^{-1}\| \leq 2K, \quad \forall x \in B_\rho(x^*),$$

$$(4.8) \quad \|F(x) - F(z) - F'(z)(x - z)\| \leq \Gamma \|x - z\|^2, \quad \forall x, z \in B_\rho(x^*).$$

*Proof.* The existence of  $\rho > 0$  so that if  $\|x - x^*\| \leq \rho$  then  $x \in L$  is shown in [3, Lemma 3.3]. Conditions (4.5)-(4.7) follows from [19, Lemma 4.3.1]. Finally, (4.8) is given in [22, Lemma 3.2.10].  $\square$

LEMMA 4.3. [3, Lemma 2.1] *Let  $x_k \in \text{int}(\Omega)$ ,  $\bar{p}_{tr}(\Delta_k)$  be given in (3.5). Then*

$$(4.9) \quad \|\bar{p}_{tr}(\Delta_k)\| \leq \|p_{tr}(\Delta_k)\|,$$

$$(4.10) \quad \|x_k + \bar{p}_{tr}(\Delta_k) - z\| \leq \|x_k + p_{tr}(\Delta_k) - z\|, \quad \text{for each } z \in \Omega.$$

We investigate the convergence of  $\{x_k\}$  under the assumption that there exists an isolated limit point  $x^*$  such that  $F(x^*) = 0$ ,  $F'(x^*)$  is invertible. However, it is worthy noting that if  $x^* \in \text{int}(\Omega)$ , the assumptions  $F(x^*) = 0$  and  $x^*$  is an isolated limit point are redundant. In fact, Theorem 4.1 ensures  $F(x^*) = 0$ . Moreover, from the invertibility of  $F'(x^*)$  there exists a neighbourhood of  $x^*$  belonging to  $\text{int}(\Omega)$  where  $F'(x)$  is nonsingular and  $\|F(x)\| > 0$  if  $x \neq x^*$  (see [19, Lemma 4.3.1]). Hence, all the vectors  $x \neq x^*$  in such neighbourhood verify  $D(x)^{-1}F'(x)^T F(x) \neq 0$  and from Theorem 4.1 we can conclude that  $x^*$  is an isolated limit point for  $\{x_k\}$ .

THEOREM 4.2. *Assume that the sequence of iterates  $\{x_k\}$  generated by the SIATR method is bounded. If  $x^*$  is an isolated limit point of  $\{x_k\}$  such that  $F(x^*) = 0$  and  $F'(x^*)$  is nonsingular, then  $\{x_k\}$  converges to  $x^*$ .*

*Proof.* Since the sequence  $\{\|F_k\|\}$  is monotone decreasing, it is convergent. The assumption  $F(x^*) = 0$  implies  $\|F_k\| \rightarrow 0$ . Let  $\{x_{k_j}\}$  be a subsequence such that  $x_{k_j} \rightarrow x^*$  and  $j_0$  be the index such that  $x_{k_j} \in B_\rho(x^*) \cap \Omega$  when  $k_j \geq k_{j_0}$ . Assume  $k_j \geq k_{j_0}$ .

To prove the thesis, we will show that  $\lim_{k_j \rightarrow \infty} \|p(\Delta_{k_j})\| = 0$ . Hence, using [21, Lemma 4.10] we conclude that  $\{x_k\}$  converges to  $x^*$ . Since  $p(\Delta_{k_j})$  is a convex combination of  $p_c(\Delta_{k_j})$  and  $\bar{p}_{tr}(\Delta_{k_j})$ , we examine the asymptotic behaviour of these two vectors separately.



We have  $p_c(\Delta_{k_j}) = -\tau_{k_j} D_{k_j}^{-2} F_{k_j}'^T F_{k_j}$  where  $\tau_k$  is defined in (2.7). Then, the following inequality follows

$$\begin{aligned}
\|p_c(\Delta_{k_j})\| &\leq \frac{\|D_{k_j}^{-1} F_{k_j}'^T F_{k_j}\|^2}{\|F_{k_j}' D_{k_j}^{-2} F_{k_j}'^T F_{k_j}\|^2} \|D_{k_j}^{-2} F_{k_j}'^T F_{k_j}\| \\
&\leq \frac{\|D_{k_j}^{-1} F_{k_j}'^T F_{k_j}\|^2}{\|D_{k_j}^{-2} F_{k_j}'^T F_{k_j}\|} \|F_{k_j}'^{-1}\|^2 \\
&= \frac{(F_{k_j}'^T F_{k_j})^T (D_{k_j}^{-2} F_{k_j}'^T F_{k_j})}{\|D_{k_j}^{-2} F_{k_j}'^T F_{k_j}\|} \|F_{k_j}'^{-1}\|^2 \\
(4.11) \quad &\leq \|F_{k_j}'^T F_{k_j}\| \|F_{k_j}'^{-1}\|^2.
\end{aligned}$$

Since  $\|F_{k_j}'^{-1}\| \leq 2K$  and  $F_{k_j}'^T F_{k_j} \rightarrow 0$ , we conclude that  $\|p_c(\Delta_{k_j})\| \rightarrow 0$ .

Now, we concentrate on  $p_{tr}(\Delta_{k_j})$ . By construction,  $m_k(p_{tr}(\Delta_{k_j})) \leq m_k(0)$ , i.e. letting  $\hat{r}_{k_j} = F_{k_j}' p_{tr}(\Delta_{k_j}) + F_{k_j}$  we have  $\|\hat{r}_{k_j}\| \leq \|F_{k_j}\|$ . Then,

$$\|p_{tr}(\Delta_{k_j})\| = \|F_{k_j}'^{-1}(-F_{k_j} + \hat{r}_{k_j})\| \leq 2 \|F_{k_j}'^{-1}\| \|F_{k_j}\| \leq 4K \|F_{k_j}\|.$$

Thus,  $\|F_k\| \rightarrow 0$  and (4.9) yield  $\lim_{k_j \rightarrow \infty} \|\bar{p}_{tr}(\Delta_{k_j})\| = 0$ .

Since  $\|p_c(\Delta_{k_j})\| \rightarrow 0$  and  $\|\bar{p}_{tr}(\Delta_{k_j})\| \rightarrow 0$  as  $k_j \rightarrow \infty$ , the thesis follows.  $\square$

To investigate the asymptotic rate of convergence of  $\{x_k\}$  to  $x^*$ , we make the additional hypothesis  $\|G_k p_k\| \rightarrow 0$  as  $k \rightarrow 0$ . In practice, this condition may fail to hold only when  $G_k = D_k$  and  $x^*$  belongs to the boundary of  $\Omega$ . On the contrary, it is guaranteed when  $G_k = I$  or when  $G_k = D_k$  and  $x^*$  lies in the interior of  $\Omega$ . To show this, note that by (3.4) and (4.7) we get

$$(4.12) \quad \|p_k\| = \|F_k'^{-1}(-F_k + r_k)\| \leq (1 + \eta_k) \|F_k'^{-1}\| \|F_k\| \leq 4K \|F_k\|,$$

and this implies that  $\|p_k\| \rightarrow 0$  as  $k \rightarrow 0$ . Also, it is easy to see that  $\|D_k\| \leq \sqrt{2/\bar{\rho}}$  whenever  $x_k \in B_{\bar{\rho}/2}(x^*)$  with  $x^* \in \text{int}(\Omega)$  and  $\bar{\rho}$  sufficiently small that  $B_{\bar{\rho}}(x^*) \subset \text{int}(\Omega)$ , [1, Corollary 3.1]. Then  $\|D_k p_k\| \rightarrow 0$  as  $k \rightarrow 0$  when  $x^* \in \text{int}(\Omega)$ .

We now prove that eventually, for  $\Delta_k$  equal to the initial trust-region radius  $\bar{\Delta}_k$ , the trust-region constraint in (3.1) becomes inactive, i.e.  $p_{tr}(\bar{\Delta}_k)$  is the minimizer of  $m_k$  on  $S_k$ . Furthermore, the step  $p(\bar{\Delta}_k)$  satisfies the sufficient decrease condition (2.9). In other words:

- $p_{tr}(\bar{\Delta}_k) = p_k$ ;
- $p(\bar{\Delta}_k)$  yields a successful iteration.

Then, we characterize the convergence rate showing that

- $x_k \rightarrow x^*$  superlinearly if  $\eta_k \rightarrow 0$  as  $k \rightarrow \infty$ ;
- $x_k \rightarrow x^*$  quadratically if  $\eta_k = O(\|F_k\|)$  as  $k \rightarrow \infty$ .

The next lemma shows the behaviour of  $p_{tr}(\bar{\Delta}_k)$ ,  $\bar{p}_{tr}(\bar{\Delta}_k)$  and  $p(\bar{\Delta}_k)$  when  $x_k$  is sufficiently near to  $x^*$ . In the remaining we let

$$(4.13) \quad \nu = 8K(K\chi_j + 1)\|F'(x^*)\|,$$

$$(4.14) \quad \epsilon_k = 2\eta_k \|F'(x^*)\|,$$

$$(4.15) \quad \delta_k = 2K(\Gamma\nu^2\|x_k - x^*\| + \epsilon_k),$$

$$(4.16) \quad \psi_k = \gamma_L\nu^2\|x_k - x^*\| + \chi_J\delta_k,$$

$$(4.17) \quad \sigma_k = \max\{\psi_k, 2K(\Gamma\nu^2\|x_k - x^*\| + \psi_k)\}.$$

LEMMA 4.4. *Assume that there exists a solution  $x^*$  of (1.1) such that  $F'(x^*)$  is nonsingular and that the sequence  $\{x_k\}$  generated by the SIATR method converges to  $x^*$ . Suppose that*

- *either  $G_k = I$ ,  $k \geq 0$ , or*
- *$G_k = D_k$ ,  $k \geq 0$ , and  $\|D_k p_k\| \rightarrow 0$  as  $k \rightarrow \infty$ .*

*Then there exists  $\rho_1 \leq \rho$  such that for all  $x_k \in B_{\rho_1}(x^*) \cap \text{int}(\Omega)$ ,*

$$(4.18) \quad p_{tr}(\bar{\Delta}_k) = p_k,$$

*where  $p_k$  is given in (3.2) and  $\bar{\Delta}_k$  is the initial trust-region radius at  $k$ th iteration. Further, when  $x_k \in B_{\rho_1}(x^*) \cap \text{int}(\Omega)$  we have*

$$(4.19) \quad \|F_k + F'_k p_{tr}(\bar{\Delta}_k)\| \leq \epsilon_k \|x_k - x^*\|,$$

$$(4.20) \quad \|\bar{p}_{tr}(\bar{\Delta}_k)\| \leq \|p_{tr}(\bar{\Delta}_k)\| \leq \nu \|x_k - x^*\|,$$

$$(4.21) \quad \|p(\bar{\Delta}_k)\| \leq \nu \|x_k - x^*\|.$$

*Proof.* The relationship (4.18) is proved by using the fact that  $\bar{\Delta}_k \geq \Delta_{min}$ , i.e.  $\bar{\Delta}_k$  is bounded below from zero for each  $k \geq 0$ . Let  $G_k = I$ ,  $\forall k \geq 0$ . By (4.12)  $\lim_{k \rightarrow \infty} \|p_k\| = 0$ . Then, there exists  $\rho_1 \leq \rho$  such that  $\|p_k\| \leq \bar{\Delta}_k$  when  $x_k \in B_{\rho_1}(x^*) \cap \text{int}(\Omega)$ . Since  $p_k$  is feasible for the trust-region problem (3.1), the thesis follows. Now consider the case  $G_k = D_k$ ,  $\forall k \geq 0$ . The assumption  $\lim_{k \rightarrow \infty} \|D_k p_k\| = 0$  implies that there exists  $\rho_1 \leq \rho$  such that  $p_k$  solves the trust-region problem (3.1) whenever  $x_k \in B_{\rho_1}(x^*) \cap \text{int}(\Omega)$ .

The remaining results are proved independently of the form of  $G_k$ . By (3.3), (4.18), (4.6) and (4.14) we obtain (4.19) as follows

$$\|F_k + F'_k p_{tr}(\bar{\Delta}_k)\| \leq 2\eta_k \|F'(x^*)\| \|x_k - x^*\| = \epsilon_k \|x_k - x^*\|.$$

The result (4.20) is derived noting that by (4.9), (4.18), (4.12) and (4.6) we get

$$(4.22) \quad \|\bar{p}_{tr}(\bar{\Delta}_k)\| \leq \|p_{tr}(\bar{\Delta}_k)\| \leq 4K \|F_k\| \leq 8K \|F'(x^*)\| \|x_k - x^*\|.$$

Then, by (4.13) relation (4.20) follows.

Finally, (4.11) and (4.7) yield

$$\|p_c(\bar{\Delta}_k)\| \leq 8K^2 \chi_J \|F'(x^*)\| \|x_k - x^*\|.$$

Hence, by (3.6) and (4.22)

$$\begin{aligned} \|p(\bar{\Delta}_k)\| &\leq t \|p_c(\bar{\Delta}_k)\| + (1-t) \|\bar{p}_{tr}(\bar{\Delta}_k)\| \\ &\leq 8K (tK \chi_J + (1-t)) \|F'(x^*)\| \|x_k - x^*\|. \end{aligned}$$

This, along with (4.13), proves (4.21). □

Next lemma investigate the features of  $p(\bar{\Delta}_k)$ .

LEMMA 4.5. *Assume that there exists a solution  $x^*$  of (1.1) such that  $F'(x^*)$  is nonsingular and that the sequence  $\{x_k\}$  generated by the SIATR method converges to  $x^*$ . Suppose that*

- either  $G_k = I$ ,  $k \geq 0$ , or
- $G_k = D_k$ ,  $k \geq 0$ , and  $\|D_k p_k\| \rightarrow 0$  as  $k \rightarrow \infty$ .

Then for all  $x_k \in B_{\rho_2}(x^*) \cap \text{int}(\Omega)$ ,  $\rho_2 \leq \rho_1/(1+\nu)$  we have

$$(4.23) \quad \|F_k + F'_k p(\bar{\Delta}_k)\| \leq \sigma_k \|x_k - x^*\|,$$

$$(4.24) \quad \|x_k + p(\bar{\Delta}_k) - x^*\| \leq \sigma_k \|x_k - x^*\|,$$

where  $\bar{\Delta}_k$  is the initial trust-region radius at  $k$ th iteration.

*Proof.* To begin, note that from (4.5) and (4.8), any vector  $x_k + q \in B_{\rho_1}(x^*)$  satisfies

$$(4.25) \quad \begin{aligned} \|x_k + q - x^*\| &\leq 2K \|F(x_k + q)\| \\ &\leq 2K (\|F(x_k + q) - F_k - F'_k q\| + \|F_k + F'_k q\|) \\ &\leq 2K (\Gamma \|q\|^2 + \|F_k + F'_k q\|). \end{aligned}$$

Let  $0 < \rho_2 \leq \rho_1/(1+\nu)$  and  $k$  sufficiently large to have  $x_k \in B_{\rho_2}(x^*) \cap \text{int}(\Omega)$ . From (4.10) and (4.20) we get

$$\|x_k + \bar{p}_{tr}(\bar{\Delta}_k) - x^*\| \leq \|x_k + p_{tr}(\bar{\Delta}_k) - x^*\| \leq \|x_k - x^*\| + \|p_{tr}(\bar{\Delta}_k)\| \leq \rho_2 + \nu \rho_2 \leq \rho_1.$$

Hence,  $x_k + p_{tr}(\bar{\Delta}_k)$  and  $x_k + \bar{p}_{tr}(\bar{\Delta}_k)$  belong to  $B_{\rho_1}(x^*)$ . Analogously, we conclude  $x_k + p(\bar{\Delta}_k) \in B_{\rho_1}(x^*)$ . Further, from (4.25), (4.19) and (4.20) we get

$$(4.26) \quad \begin{aligned} \|x_k + p_{tr}(\bar{\Delta}_k) - x^*\| &\leq 2K (\Gamma \|p_{tr}(\bar{\Delta}_k)\|^2 + \epsilon_k \|x_k - x^*\|) \\ &\leq 2K (\Gamma \nu^2 \|x_k - x^*\| + \epsilon_k) \|x_k - x^*\| \\ &= \delta_k \|x_k - x^*\|, \end{aligned}$$

where  $\delta_k$  is given in (4.15).

Next we turn our attention to (4.23). First we need to estimate  $\|F_k + F'_k \bar{p}_{tr}(\bar{\Delta}_k)\|$ . Note that

$$\begin{aligned} F_k + F'_k \bar{p}_{tr}(\bar{\Delta}_k) &= F(x_k + \bar{p}_{tr}(\bar{\Delta}_k)) - F(x^*) + \int_0^1 (F'(x_k) - F'(x_k + t\bar{p}_{tr}(\bar{\Delta}_k))) \bar{p}_{tr}(\bar{\Delta}_k) dt \\ &= \int_0^1 F'(x^* + t(x_k + \bar{p}_{tr}(\bar{\Delta}_k) - x^*)) (x_k + \bar{p}_{tr}(\bar{\Delta}_k) - x^*) dt \\ &\quad + \int_0^1 (F'(x_k) - F'(x_k + t\bar{p}_{tr}(\bar{\Delta}_k))) \bar{p}_{tr}(\bar{\Delta}_k) dt. \end{aligned}$$

Then, by Assumption 1, (4.10), (4.20) and (4.26) we obtain

$$\begin{aligned} \|F_k + F'_k \bar{p}_{tr}(\bar{\Delta}_k)\| &\leq \chi_J \|x_k + p_{tr}(\bar{\Delta}_k) - x^*\| + \gamma_L \|p_{tr}(\bar{\Delta}_k)\|^2 \\ &\leq (\chi_J \delta_k + \gamma_L \nu^2 \|x_k - x^*\|) \|x_k - x^*\|, \end{aligned}$$

and by relationship (4.1)

$$(4.27) \quad \|F_k + F'_k p(\bar{\Delta}_k)\| \leq (\chi_J \delta_k + \gamma_L \nu^2 \|x_k - x^*\|) \|x_k - x^*\|.$$

Thus, (4.17) yields the thesis.

Finally (4.24) is derived by (4.25), (4.21) and (4.27) as follows

$$\|x_k + p(\bar{\Delta}_k) - x^*\| \leq 2K (\Gamma \nu^2 \|x_k - x^*\| + \gamma_L \nu^2 \|x_k - x^*\| + \chi_J \delta_k) \|x_k - x^*\|.$$

□

We close this section proving the main result.

**THEOREM 4.3.** *Assume that there exists a solution  $x^*$  of (1.1) such that  $F'(x^*)$  is nonsingular and that the sequence  $\{x_k\}$  generated by the SIATR method converges to  $x^*$ . Suppose that*

- either  $G_k = I$ ,  $k \geq 0$ , or
- $G_k = D_k$ ,  $k \geq 0$ , and  $\|D_k p_k\| \rightarrow 0$  as  $k \rightarrow \infty$ .

*Then, eventually,  $p(\bar{\Delta}_k)$  satisfies (2.9). Moreover, if  $\eta_k \rightarrow 0$  as  $k \rightarrow \infty$ , the sequence  $\{x_k\}$  converges to  $x^*$  superlinearly. If  $\eta_k = O(\|F_k\|)$  as  $k \rightarrow \infty$ , the convergence rate is quadratic.*

*Proof.* Let  $\zeta$  be such that

$$(4.28) \quad \zeta < \min \left\{ \rho_2, \frac{1}{8\nu^2 K^2 \Gamma}, \frac{3(1-\beta_2)}{4\nu^2 \Gamma} \right\}.$$

Let  $k$  be sufficiently large to have  $x_k \in B_\zeta(x^*) \cap \text{int}(\Omega)$  and

$$(4.29) \quad \sigma_k < \frac{1}{16K^2}.$$

This condition is met for  $k$  sufficiently large since

$$(4.30) \quad \sigma_k = O(\|x_k - x^*\| + \eta_k), \quad k \rightarrow \infty.$$

First, we show that  $\rho_f(p(\bar{\Delta}_k)) \geq \beta_2$ . Note that

$$\frac{\text{ared}(p(\bar{\Delta}_k))}{\text{pred}(p(\bar{\Delta}_k))} = 1 - \frac{\|F(x_k + p(\bar{\Delta}_k))\|^2 - \|F_k + F'_k p(\bar{\Delta}_k)\|^2}{\|F_k\|^2 - \|F_k + F'_k p(\bar{\Delta}_k)\|^2},$$

and

$$\begin{aligned} \|F(x_k + p(\bar{\Delta}_k))\|^2 - \|F_k + F'_k p(\bar{\Delta}_k)\|^2 &= \|F(x_k + p(\bar{\Delta}_k)) - F_k - F'_k p(\bar{\Delta}_k)\|^2 \\ &\quad + 2(F(x_k + p(\bar{\Delta}_k)) - F_k - F'_k p(\bar{\Delta}_k))^T \\ &\quad (F_k + F'_k p(\bar{\Delta}_k)). \end{aligned}$$

Then, by (4.8), (4.23) and (4.21) we get

$$\begin{aligned} \|F(x_k + p(\bar{\Delta}_k))\|^2 - \|F_k + F'_k p(\bar{\Delta}_k)\|^2 &\leq \Gamma^2 \|p(\bar{\Delta}_k)\|^4 + 2\sigma_k \Gamma \|p(\bar{\Delta}_k)\|^2 \|x_k - x^*\| \\ &\leq (\Gamma^2 \nu^4 \|x_k - x^*\| + 2\sigma_k \Gamma \nu^2) \|x_k - x^*\|^3. \end{aligned}$$

Further, from (4.5), (4.23) and (4.29) we get

$$\|F_k\|^2 - \|F_k + F'_k p(\bar{\Delta}_k)\|^2 \geq \left( \frac{1}{(2K)^2} - \sigma_k \right) \|x_k - x^*\|^2 > \frac{3}{16K^2} \|x_k - x^*\|^2.$$

Therefore, since  $x_k \in B_\zeta(x^*) \cap \text{int}(\Omega)$ , by (4.28) and (4.29)

$$\begin{aligned} \frac{\text{ared}(p(\bar{\Delta}_k))}{\text{pred}(p(\bar{\Delta}_k))} &\geq 1 - \frac{16\nu^2 K^2 \Gamma}{3} (\nu^2 \Gamma \|x_k - x^*\| + 2\sigma_k) \|x_k - x^*\| \\ &\geq 1 - \frac{16\nu^2 K^2 \Gamma}{3} (\nu^2 \Gamma \zeta + \frac{1}{8K^2}) \|x_k - x^*\| \\ &\geq 1 - \frac{4\nu^2 \Gamma}{3} \|x_k - x^*\| > \beta_2. \end{aligned}$$

Therefore,  $x_{k+1} = x_k + p(\bar{\Delta}_k)$  and from (4.24) we conclude

$$(4.31) \quad \|x_{k+1} - x^*\| \leq \sigma_k \|x_k - x^*\|.$$

The form of  $\sigma_k$  given in (4.30) ensures superlinear convergence rate if  $\eta_k \rightarrow 0$ . Moreover, if  $\eta_k = O(\|F_k\|) = O(\|x_k - x^*\|)$  we get quadratic convergence rate.  $\square$

**5. Applications.** To develop viable approaches for large scale problems, this section discusses the two issues left unspecified in the description of the SIATR method: the choice of the subspace  $S_k$ , the way of solving the subspace trust-region problem (3.1).

Since there is no finite method of determining the exact solution of (3.1), an approximation to it is used. Remarkably, it is easy to see that the convergence properties of the SIATR method take place using an approximate solution  $p_{tr}(\Delta_k)$  to (3.1) which satisfies the two mild conditions:

- a)  $m_k(p_{tr}(\Delta_k)) \leq m_k(0)$ ;
- b)  $p_{tr}(\Delta_k) = p_k$  when  $p_k$  is feasible for (3.1).

The key is that global convergence is provided by  $p_c(\Delta_k)$  and rapid local convergence is ensured if eventually  $p_k$  is the solution of (3.1).

We outline a subspace dogleg strategy for solving (3.1) approximately. Let  $S_k = \text{span}\{s_1, s_2, \dots, s_r\}$ ,  $S_k^G = \text{span}\{G_k s_1, G_k s_2, \dots, G_k s_r\}$ . Once an orthonormal basis  $W \in \mathbb{R}^{n \times r}$  for  $S_k^G$  has been constructed, a vector  $p \in S_k$  is such that  $G_k p = Wq$  for some  $q \in \mathbb{R}^r$  and instead of (3.1) one can consider the spherical trust-region problem

$$(5.1) \quad \min_{q \in \mathbb{R}^r} \{\psi_k(q) : \|q\| \leq \Delta_k\},$$

where  $\psi_k$  is the quadratic model on  $\mathbb{R}^r$

$$(5.2) \quad \psi_k(q) = \frac{1}{2} \|F_k + F_k' G_k^{-1} Wq\|^2.$$

Let  $q_{tr}(\Delta_k)$  be the dogleg solution to (5.1), [9]. Its evaluation calls for the Cauchy point  $q_c(\Delta_k)$  for (5.1) and the vector

$$(5.3) \quad q_k = \underset{q \in \mathbb{R}^r}{\operatorname{argmin}} \psi_k(q).$$

We remark that  $q_k$  is such that

$$(5.4) \quad q_k = W^T G_k p_k,$$

with  $p_k$  given in (3.2). Therefore, if one of the two vectors  $p_k$  and  $q_k$  is known the other one can be trivially evaluated.

Finally, once  $q_{tr}(\Delta_k)$  has been computed, coming back into the original space, the vector

$$(5.5) \quad p_{tr}(\Delta_k) = G_k^{-1} W q_{tr}(\Delta_k),$$

is built. It approximately solves (3.1) and satisfies both a) and b). In fact, a) is straightforward and b) is due to the fact that  $G_k p_k \in S_k^G$ , i.e.  $W W^T G_k p_k = G_k p_k$ .

This yields  $\|q_k\| = \|W^T G_k p_k\| = \|G_k p_k\|$ . Then,  $q_k$  is feasible for (5.1) if  $p_k$  is feasible for (3.1). Consequently,  $q_{tr}(\Delta_k) = q_k$  whenever  $\|q_k\| \leq \Delta_k$  and this implies  $p_{tr}(\Delta_k) = p_k$  whenever  $\|G_k p_k\| \leq \Delta_k$ .

This discussion leads to the subspace dogleg strategy sketched below.

ALGORITHM II. A SUBSPACE DOGLEG STRATEGY FOR (3.1).

Input parameters  $x_k \in \text{int}(\Omega)$ ,  $\Delta_k > 0$ ,  $\eta_k \in [0, 1)$ ,  $G_k$ ,  $\nabla f_k$ .

1. Choose a subspace  $S_k = \text{span}\{s_1, s_2, \dots, s_r\}$  such that (3.3) holds.
2. Find an orthonormal basis  $W \in \mathbb{R}^{n \times r}$  for  $S_k^G = \text{span}\{G_k s_1, G_k s_2, \dots, G_k s_r\}$ .
3. Compute the vector  $q_k \in \mathbb{R}^r$  satisfying (5.3).
4. Compute the Cauchy step  $q_c(\Delta_k) = -\hat{\mu}_k W^T G_k^{-1} \nabla f_k$  with

$$(5.6) \quad \begin{aligned} \hat{\mu}_k &= \underset{\|\mu W^T G_k^{-1} \nabla f_k\| \leq \Delta_k}{\text{argmin}} \quad \psi_k(-\mu W^T G_k^{-1} \nabla f_k) \\ &= \min \left\{ \frac{\|W^T G_k^{-1} \nabla f_k\|^2}{\|F_k' G_k^{-1} W W^T G_k^{-1} \nabla f_k\|^2}, \frac{\Delta_k}{\|W^T G_k^{-1} \nabla f_k\|} \right\}. \end{aligned}$$

6. Find the dogleg solution  $q_{tr}(\Delta_k)$  to (5.1)

$$q_{tr}(\Delta_k) = \begin{cases} q_k & \text{if } \|q_k\| \leq \Delta_k, \\ q_c(\Delta_k) & \text{if } \|q_c(\Delta_k)\| \leq \Delta_k, \\ sq_k + (1-s)q_c(\Delta_k), s \in (0, 1), \text{ s.t. } \|q_{tr}(\Delta_k)\| = \Delta_k, & \text{otherwise.} \end{cases}$$

7. Compute  $p_{tr}(\Delta_k)$  by (5.5).

We recall that when (3.3) holds, the subspace  $S_k$  contains an inexact Newton step  $p_k^I$  for the problem  $F(x) = 0$  such that

$$(5.7) \quad F_k' p_k^I = -F_k + r_k, \quad \|r_k\| \leq \eta_k \|F_k\|.$$

Our main purpose now is to show that Krylov subspace methods for solving (5.7) provide the way to perform Steps 1-3 of the above algorithm effectively. In this presentation we restrict to unpreconditioned Krylov subspace methods. However, the global strategies discussed work with right preconditioning too, see [6]. The resulting methods belong to the class of trust-region Newton-Krylov methods [6, 7, 13]. Moreover, thanks to our convergence results, the linear system (5.7) can be solved with an accuracy that increases as the solution is approached and oversolving can be avoided by choosing suitable sequences  $\{\eta_k\}$ , see [14].

**Two dimensional subspace minimization.** A possible approach consists in determining  $p_k^I$  by a Krylov method and fixing

$$S_k = \text{span}\{p_k^I, G_k^{-2} \nabla f_k\}.$$

This way, Algorithm II is a two dimensional subspace trust-region strategy. Step 2 requires one step of the Gram-Schmidt procedure to compute  $W$ . The least-square problem  $\min_{q \in \mathbb{R}^2} \|F_k + F_k' G_k^{-1} W q\|^2$  in Step 3 can be solved without much effort either by the  $QR$  factorization of the  $n \times 2$  matrix  $F_k' G_k^{-1} W$  or by solving the normal equations.

In the case  $G_k = D_k$ ,  $k \geq 0$ , the vector  $p_{tr}(\Delta_k)$  produces as much decrease in the quadratic model  $m_k$  as the generalized Cauchy point  $p_c(\Delta_k)$ . To show this fact,

note that by  $d_k \in S_k$  and  $G_k d_k = -D_k^{-1} \nabla f_k \in S_k^G$ , trivially follows  $D_k^{-1} \nabla f_k = WW^T D_k^{-1} \nabla f_k$ . Consequently, it is easy to see that  $G_k^{-1} W q_c(\Delta_k) = \hat{\tau}_k d_k$  where  $\hat{\tau}_k$  is given in (2.6) and

$$m_k(p_{tr}(\Delta_k)) = \psi_k(q_{tr}(\Delta_k)) \leq \psi_k(q_c(\Delta_k)) \leq m_k(p_c(\Delta_k)),$$

i.e.  $p_{tr}(\Delta_k)$  satisfies (2.8).

**GMRES subspace implementation.** Another implementation can be proposed in connection with GMRES method [28]. GMRES shows a certain optimality among all Krylov methods, BICGSTAB, TFQMR etc., commonly used in the solution of general linear systems. In practice, it minimizes the residual norm  $2 m_k(p)^{\frac{1}{2}} = \|F_k + F_k' p\|$  over all corrections in the current Krylov subspace. Due to this property it is possible to restrict the search direction to a subspace of small dimension using information provided by GMRES and to perform the subspace dogleg strategy at a low computational cost.

For sake of clarity, we sketch the application of GMRES to (5.7). For details we refer to [28]. GMRES generates a sequence of iterates  $\{p_k^m\}$ ,  $p_k^m \in \mathbb{R}^n$ ,  $m \geq 0$ , until  $\|F_k' p_k^m + F_k\| \leq \eta_k \|F_k\|$ . Then,  $p_k^I = p_k^m$  is set. Given  $p_k^0$ , at each GMRES iteration  $p_k^m$  solves the least-squares problem

$$(5.8) \quad \min_{p \in p_k^0 + K_m} \|F_k + F_k' p\|,$$

where  $K_m$  is the Krylov subspace  $K_m = \text{span}\{r_k^0, F_k' r_k^0, (F_k')^2 r_k^0, \dots, (F_k')^{m-1} r_k^0\}$  and  $r_k^0 = \|F_k' p_k^0 + F_k\|$ . An orthonormal basis  $V = [v_1, \dots, v_m] \in \mathbb{R}^{n \times m}$  of  $K_m$  is computed by the Arnoldi process and satisfies

$$(5.9) \quad F_k' V_m = V_{m+1} H_m,$$

where  $V_{m+1} = [v_1, \dots, v_{m+1}] \in \mathbb{R}^{n \times (m+1)}$  is the orthonormal basis for  $K_{m+1}$  and  $H_m \in \mathbb{R}^{(m+1) \times m}$  is an Hessenberg matrix. Thus, the least-squares problem (5.8) reduces to

$$(5.10) \quad \min_{y \in \mathbb{R}^m} \|\beta_k e_1 - H_m y\|,$$

where  $\beta_k = \|r_k^0\|$  and  $e_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^n$ . GMRES solves problem (5.10) by the QR factorization of  $H_m$

$$(5.11) \quad H_m = \hat{Q} \hat{R},$$

with  $\hat{Q} \in \mathbb{R}^{m+1 \times m+1}$ ,  $\hat{R} \in \mathbb{R}^{m+1 \times m}$ . Note that  $p_k^m = p_k^0 + V_m y_m$  where  $y_m$  is the solution of (5.10).

Since storing the basis of  $K_m$  may become prohibitive when  $n$  is large, typically a maximum value  $m_M$  of GMRES iterations is fixed. If after  $m_M$  GMRES iterations the required accuracy has not been achieved, GMRES is restarted with initial guess  $p_k^0$  equal to  $p_k^{m_M}$ . The convergence of such a procedure, denoted as GMRES( $m_M$ ), is not always guaranteed but the idea works well in practice.

Suppose the vector  $p_k^I$  is generated using the GMRES( $m_M$ ) method. By construction,  $p_k^I$  minimizes  $m_k(p)$  within the affine subspace  $p_k^0 + K_m$ . Then if  $p_k^0 = 0$ , it is convenient to set

$$S_k = K_m.$$

Trivially, we get  $p_k = p_k^I$ . Moreover,  $W = V_m$  if  $G_k = I$ ,  $k \geq 0$ , while for the other choice of  $G$  the matrix  $W$  has to be computed. Finally, Step 3 of the subspace dogleg strategy is completed by using (5.4).

If  $p_k^0 \neq 0$ , let

$$S_k = \text{span}\{v_1, \dots, v_m, p_k^0\}.$$

If  $G_k = I$ ,  $k \geq 0$ , an orthonormal basis  $W$  for  $S_k$  can be easily obtained adding one column to the matrix  $V_m$ . Such column can be computed with a step of the Gram-Schmidt procedure. Otherwise, the whole matrix  $W$  must be computed. To evaluate  $q_k$  we proceed as follows. Let  $T_k = [V_m, p_k^0] \in \mathbb{R}^{n \times r}$ ,  $r = m + 1$ . Then, the solution  $p_k$  to problem (3.2) has the form  $p_k = T_k y$ ,  $y \in \mathbb{R}^r$ , and

$$\min_{p \in S_k} \|F_k + F_k' p\| = \min_{y \in \mathbb{R}^r} \|F_k + F_k' T_k y\|.$$

Hence, if the columns of  $F_k' T_k$  are linearly independent, i.e. the columns of  $T_k$  are linearly independent, the solution

$$(5.12) \quad y_k = \underset{y \in \mathbb{R}^r}{\operatorname{argmin}} \|F_k + F_k' T_k y\|,$$

is unique. Moreover, in [6] it has been shown that the Cholesky factorization  $R_k^T R_k$  of the matrix  $(F_k' T_k)^T F_k' T_k \in \mathbb{R}^{r \times r}$  can be cheaply computed. In fact, this factorization can be obtained exploiting the  $\hat{Q}\hat{R}$  factorization of the Hessember matrix  $H_m$  provided by GMRES and solving one upper triangular system of dimension  $m$ . A complication to this global strategy arises when  $F_k' T_k$  is ill-conditioned. We can monitor this occurrence estimating the condition number of the small matrix  $R_k$ . If such number is greater than a fixed threshold, we may perturb the quadratic model  $\psi_k$  following the strategy given in [12, p. 151].

#### REFERENCES

- [1] 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.
- [2] S. Bellavia, B. Morini, *A globally convergent Newton-GMRES subspace method for systems of nonlinear equations*, SIAM J. Sci. Comput., 23 (2001), pp. 940-960.
- [3] S. Bellavia, B. Morini, *An interior global method for nonlinear systems with simple bounds*, revised, submitted to Optim. Methods Softw.
- [4] L.T. Biegler, O. Ghattas, M. Heinkenschloss, B. van Bloemen Waanders, *Large-scale PDE-constrained optimization*, Lecture Notes in Computational Science and Engineering, 30, Berlin, Springer, 2003.
- [5] 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.
- [6] P.N. Brown, Y. Saad, *Hybrid Krylov methods for nonlinear systems of equations*, SIAM J. Sci. Stat. Comput., 11 (1990), pp. 450-481.
- [7] P.N. Brown, Y. Saad, *Convergence theory of nonlinear Newton-Krylov algorithms*, SIAM J. Optim., 4 (1994), pp. 297-330.
- [8] R. H. Byrd, R. B. Schnabel and M. H. Schultz *Approximate solution of the trust-region problem by minimization over two-dimensional subspaces*, Math. Program., 40 (1988), pp. 247-263.
- [9] A.R. Conn, N.I.M. Gould, Ph.L. Toint, *Trust-region methods*, SMP/SIAM Series on Optimization, 2000.
- [10] T.F. Coleman, Y. Li, *An interior trust-region approach for nonlinear minimization subject to bounds*, SIAM J. Optim., 6 (1996), pp. 418-445.
- [11] R.S. Dembo, S.C. Eisenstat, T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal., 19 (1982), pp. 400-408.



- [12] J.E. Dennis, R.B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice Hall, Englewood Cliffs, NJ, 1983.
- [13] S.C. Eisenstat, H.F. Walker *Globally convergent Inexact Newton methods*, SIAM J. Optim., 4 (1994), pp. 393-422.
- [14] S.C. Eisenstat and H.F. Walker, *Choosing the forcing term in an inexact Newton method*, SIAM J. Sci. Comput., 17 (1996), pp. 16-32.
- [15] R. Fletcher, S. Leyffer, *Filter-type Algorithms for Solving Systems of Algebraic Equations and Inequalities*, High Performance Algorithms and Software for Nonlinear Optimization, G. Di Pillo and A. Murli, editors, Kluwer Academic Publishers, 2003, pp. 259-278.
- [16] N.I.M. Gould, S. Lucidi, M. Roma, P.L. Toint, *Solving the trust-region subproblem using the Lanczos method*, SIAM J. Optim., 9 (1999), pp. 504-525.
- [17] A. Hasnat, S. Roy, *Microphase-enhanced reactions: simultaneous effects on ion coupling and counterion binding*, Ind. Eng. Chem. Res., 38 (1999), pp. 4571-4578.
- [18] V. A. Juvekar, C. V. Anoop, S. K. Pattanayek, V. M. Naik, *A continuum model for polymer adsorption at the solid-liquid interface*, Macromolecules, 32 (1999), pp. 863-873.
- [19] C.T. Kelley, *Iterative Methods for Linear and Nonlinear Equations*, Frontiers in Applied Mathematics, SIAM, 1995.
- [20] L. Luksan, *Inexact trust-region method for large sparse systems of nonlinear equations*, J. Optim. Theory Appl., 81 (1994), pp. 569-591.
- [21] J.J. Moré, D.C. Sorensen, *Computing a trust-region step*, SIAM J. Sci. Stat. Comput., 4 (1983), pp. 553-572.
- [22] J.M. Ortega, W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, 1970.
- [23] R. P. Pawlowski, J. N. Shadid, J. P. Simonis, H.F. Walker, *Globalization techniques for Newton-Krylov methods and applications to the fully-coupled solution of the Navier-Stokes equations*, Sandia National Laboratories Report SAND2004-1777, submitted to SIAM Rev.
- [24] M. Pernice, H.F. Walker, *NITSOL: a new iterative solver for nonlinear systems*, SIAM Journal Sci Comput., 19 (1998), pp. 302-318.
- [25] L. Qi, X.J. Tong, D.H.Li, *An active-set projected trust-region algorithm for box-constrained nonsmooth equations*, J. Optim. Theory Appl., 120 (2004), pp. 627-649.
- [26] H. Qi, L. Qi, D. Sun, *Solving Karush-Kuhn-Tucker systems via the trust-region and the conjugate gradient methods*, Siam. J. Optim., 14 (2003), pp. 439-463.
- [27] Y. Saad, *Iterative methods for sparse linear systems*, SIAM, 2003.
- [28] Y. Saad and M.H. Schultz, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856-869.
- [29] T. Steihaug, *The conjugate gradient method and trust-regions in large scale optimization*, SIAM J. Numer. Anal., 20 (1983), pp. 626-637.
- [30] Ph. L. Toint, *Towards an efficient sparsity exploiting Newton method for minimization*, Sparse matrices and their uses, I.S. Duff ed., Academic Press, (1981), pp. 57-88.
- [31] X.J. Tong, L.Qi, *On the convergence of a trust-region method for solving constrained nonlinear equations with degenerate solutions*, revised, submitted to J. Optim. Theory Appl.
- [32] M. Ulbrich, *Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems*, SIAM J. Optim., 11 (2000), pp. 889-917.
- [33] A. Walter, A. Griewank, A. Best, *Multiple vector-Jacobian products are cheap*, Appl. Numer. Math., 30 (1999), pp. 367-377.
- [34] A.J. Wood, B.F. Wollenberg, *Power generation, Operation and Control*, John Wiley & Sons, New York, USA, 1996.