

Research Article

Newton-Krylov Type Algorithm for Solving Nonlinear Least Squares Problems

Mohammedi R. Abdel-Aziz¹ and Mahmoud M. El-Alem^{1,2}

¹ Department of Mathematics and Computer Science, Faculty of Science, Kuwait University, P.O. 5969, Safat 13060, Kuwait City, Kuwait

² Department of Mathematics, Faculty of Science, Alexandria University, Alexandria, Egypt

Correspondence should be addressed to Mohammedi R. Abdel-Aziz, mr_azizmona@yahoo.com

Received 15 December 2008; Accepted 2 February 2009

Recommended by Irena Lasiecka

The minimization of a quadratic function within an ellipsoidal trust region is an important subproblem for many nonlinear programming algorithms. When the number of variables is large, one of the most widely used strategies is to project the original problem into a small dimensional subspace. In this paper, we introduce an algorithm for solving nonlinear least squares problems. This algorithm is based on constructing a basis for the Krylov subspace in conjunction with a model trust region technique to choose the step. The computational step on the small dimensional subspace lies inside the trust region. The Krylov subspace is terminated such that the termination condition allows the gradient to be decreased on it. A convergence theory of this algorithm is presented. It is shown that this algorithm is globally convergent.

Copyright © 2009 M. R. Abdel-Aziz and M. M. El-Alem. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Nonlinear least squares (NLS) problems are unconstrained optimization problems with special structures. These problems arise in many aspects such as the solution of overdetermined systems of nonlinear equations, some scientific experiments, pattern recognition, and maximum likelihood estimation. For more details about these problems [1]. The general formulation of the NLS problem is to determine the solution $x^* \in \mathfrak{R}^n$ that minimizes the function

$$f(x) = \|R(x)\|_2^2 = \sum_{i=1}^m (R_i(x))^2, \quad (1.1)$$

where $R(x) = (R_1(x), R_2(x), \dots, R_m(x))^t$, $R_i : \mathfrak{R}^n \rightarrow \mathfrak{R}$, $1 \leq i \leq m$ and $m \geq n$.

There are two general types of algorithms for solving NLS problem (1.1). The first type is most closely related to solving systems of nonlinear equations [2] and it leads to

Gauss-Newton and Levenberg-Marquart methods [3]. The second type is closely related to unconstrained minimization techniques [4]. Of course the two types of algorithms are closely related to each other [5].

The presented algorithm is a Newton-Krylov type algorithm. It requires a fixed-size limited storage proportional to the size of the problem and relies only upon matrix vector product. It is based on the implicitly restarted Arnoldi method (IRAM) to construct a basis for the Krylov subspace and to reduce the Jacobian into a Hessenberg matrix in conjunction with a trust region strategy to control the step on that subspace [6].

Trust region methods for unconstrained minimization are blessed with both strong theoretical convergence properties and a good accurate results in practice. The trial computational step in these methods is to find an approximate minimizer of some model of the true objective function within a trust region for which a suitable norm of the correction lies inside a given bound. This restriction is known as the trust region constraint, and the bound on the norm is its radius. The radius is adjusted so that successive model problems minimized the true objective function within the trust region [7].

The trust region subproblem is the problem of finding $s(\Delta)$ so that

$$\phi(s(\Delta)) = \min\{\phi(s) : \|s\| \leq \Delta\}, \quad (1.2)$$

where Δ is some positive constant, $\|\cdot\|$ is the Euclidean norm in \mathfrak{R}^n , and

$$\phi(s) = g^t s + \frac{1}{2} s^t H s, \quad (1.3)$$

where $g \in \mathfrak{R}^n$ and $H \in \mathfrak{R}^{n \times n}$ are, respectively, the gradient vector and the Hessian matrix or their approximations.

There are two different approaches to solve (1.2). These approaches are based on either solving several linear systems [8] or approximating the curve $s(\Delta)$ by a piecewise linear approximation “dogleg strategies” [9]. In large-scale optimization, solving linear systems is computationally expensive. Moreover, it is not clear how to define the dogleg curves when the matrix H is singular [10].

Several authors have studied inexact Newton’s methods for solving NLS problems [11]. Xiaofang et al. have introduced stable factorized quasi-Newton methods for solving large-scale NLS [12]. Dennis et al. proposed a convergence theory for structured BFGS secant method with an application for NLS [13]. The Newton-Krylov method is an example of inexact Newton methods. Krylov techniques inside a Newton iteration in the context of system of equations have been proposed in [14]. The recent work of Sorensen, provides an algorithm which is based on recasting the trust region subproblem into a parameterized eigenvalue problem. This algorithm provides a super linearly convergent scheme to adjust the parameter and find the optimal solution from the eigenvector of the parameterized problem, as long as the hard case does not occur [15].

This contribution is organized as follows. In Section 2, we introduce the structure of the NLS problem and a general view about the theory of the trust region strategies and their convergence. The statement of the algorithm and its properties is developed in Section 3. Section 4 states the assumptions and presents the role of restarting mechanism to control the dimension of the Krylov subspace. The global convergence analysis is presented in Section 5. Concluding remarks and future ideas are given in the last section.

2. Structure of the Problem

The subspace technique plays an important role in solving the NLS problem (1.1). We assume that the current iterate x_k and a Krylov subspace S_k . Denote the dimension of S_k to be k_j and $\{v_1^{(k)}, v_2^{(k)}, \dots, v_{k_j}^{(k)}\}$ be a set of linearly independent vectors in S_k . The next iterate x_{k+1} is such that the increment $x_{k+1} - x_k \in S_k$. Thus, we have $\|R_j(x_k + V_k y)\| = 0$, for $j = 1, 2, \dots, m$ and $y \in \mathfrak{R}^{k_j}$, where the matrix $V_k = [v_1^{(k)}, v_2^{(k)}, \dots, v_{k_j}^{(k)}]$. The second-order Taylor expansion of $\|R(x)\|_2^2$ at x_k is

$$\|R(x_k) + J_k s\|_2^2 + s^t Q_k s, \quad (2.1)$$

where $Q_k \in \mathfrak{R}^{n \times n}$ is defined by $Q_k = \sum_{j=1}^m R_j(x_k) \nabla^2 R_j(x_k)$ and the jacobian matrix is $J_k = J(x_k) = [\nabla R_1(x_k), \nabla R_2(x_k), \dots, \nabla R_m(x_k)]^t$. If we consider $s \in S_k$, we get the quadratic model

$$q_k(y) = \|R(x_k) + J_k V_k y\|_2^2 + y^t A_k y, \quad (2.2)$$

where $A_k \in \mathfrak{R}^{k_j, k_j}$ approximates the reduced matrix $V_k^t Q_k V_k = \sum_{j=1}^m R_j(x_k) V_k^t \nabla^2 R_j(x_k) V_k$. The first order necessary conditions of (2.1) is

$$(J_k^t J_k + Q_k) s + R_k^t J_k = 0. \quad (2.3)$$

Thus, a solution of the quadratic model is a solution to an equation of the form

$$(J_k^t J_k + Q_k) s = -R_k^t J_k. \quad (2.4)$$

The model trust region algorithm generates a sequence of points x_k , and at the k th stage of the iteration the quadratic model of problem (1.2) has the form

$$\phi_k(s) = J_k^t R_k s + \frac{1}{2} s^t (J_k^t J_k + Q_k) s. \quad (2.5)$$

At this stage, an initial value for the trust region radius Δ_k is also available. An inner iteration is then performed which consists of using the current trust region radius, Δ_k , and the information contained in the quadratic model to compute a step, $s(\Delta_k)$. Then a comparison of the actual reduction of the objective function

$$\text{ared}_k(\Delta_k) = \|R(x_k)\|_2^2 - \|R(x_k + s_k(\Delta_k))\|_2^2, \quad (2.6)$$

and the reduction predicted by the quadratic model

$$\text{pred}_k(\Delta_k) = \|R(x_k)\|_2^2 - q_k(y) \quad (2.7)$$

is performed. If there is a satisfactory reduction, then the step can be taken, or a possibly larger trust region used. If not, then the trust region is reduced and the inner iteration is repeated. For now, we leave unspecified what algorithm is used to form the step $s(\Delta)$, and how the trust region radius Δ is changed. We also leave unspecified the selection of $(J_k^t J_k + Q_k)$ except to restrict it to be symmetric. Details on these points will be addressed in our forth coming paper.

The solution s of the quadratic model (2.2) is a solution to an equation of the form

$$[(J^t J + Q) + \mu I]s = -R^t J, \quad (2.8)$$

with $\mu \geq 0$, $\mu(s^t s - \Delta^2) = 0$ and $[(J^t J + Q) + \mu I]$ is positive semidefinite. This represents the first-order necessary conditions concerning the pair μ and s , where s is a solution to model (2.2) and μ is the Lagrange multiplier associated with the constraint $s^t s \leq \Delta^2$. The sufficient conditions that will ensure s to be a solution to (2.2) can be established in the following lemma which has a proof in [16].

Lemma 2.1. *Let $\mu \in \Re$ and $s \in \Re^n$ satisfy $[(J^t J + Q) + \mu I]s = -R^t J$ with $[(J^t J + Q) + \mu I]$ positive semidefinite, if*

$$\begin{aligned} \mu = 0, \quad \|s\| \leq \Delta \text{ then } s \text{ solves (2.2),} \\ \|s\| = \Delta, \text{ then } s \text{ solves } \min\{\phi(s) : \|s\| = \Delta\}, \\ \mu \geq 0 \text{ with } \|s\| = \Delta, \text{ then } s \text{ solves (2.2).} \end{aligned} \quad (2.9)$$

Of course, if $[(J^t J + Q) + \mu I]$ is positive definite then s is a unique solution.

The main result which is used to prove that the sequence of gradients tends to zero for modified Newton methods is

$$f - \phi(s) \geq \frac{1}{2} \|R^t J\| \min \left\{ \Delta, \frac{\|R^t J\|}{\|J^t J + Q\|} \right\}, \quad (2.10)$$

where s is a solution to (2.2). The geometric interpretation of this inequality is that for a quadratic function, any solution s to (2.2) produces a decrease $f - \phi(s)$ that is at least as much as the decrease along the steepest descent direction $-R^t J$. A proof of this result may be found in [17].

3. Algorithmic Framework

The algorithm we will discuss here requires that f is twice differentiable at any point x in the domain of f . This algorithm involves two levels. In the first level we use IRA method to reduce the Hessian to a tridiagonal matrix and construct an orthonormal basis for the invariant subspace of the Hessian matrix. The second level is used to compute the step and update the trust region radius for the reduced local model (a model defined on the subspace).

The desired properties of this algorithm are:

- (1) the algorithm should be well defined for a sufficiently general class of functions and it is globally convergent;
- (2) the algorithm should be invariant under linear affine scalings of the variables, that is, if we replace $f(x)$ by $\hat{f}(y) = f(s + Vy)$, where $V \in \mathfrak{R}^{n \times k}$ is an orthonormal matrix, $x \in \mathfrak{R}^n$ and $y \in \mathfrak{R}^k$, then applying the iteration to \hat{f} with initial guess y_0 satisfying $x_0 = s + Vy_0$ should produce a sequence $\{y_l\}$ related to the sequence $\{x_l\}$ by $x_l = Vy_l + s$, where x_l is produced by applying the algorithm to f with initial guess x_0 ;
- (3) the algorithm should provide a decrease that is at least as large as a given multiple of the minimum decrease that would be provided by a quadratic search along the steepest descent direction;
- (4) the algorithm should give as good a decrease of the quadratic model as a direction of the negative gradient when the Hessian, $(J^t J + Q)$, is indefinite and should force the direction to be equal to the Newton direction when $(J^t J + Q)$ is symmetric positive definite.

The following describes a full iteration of a truncated Newton type method. Some of the previous characteristics will be obvious and the other ones will be proved in the next section.

Algorithm 3.1.

- (1) Step 0 (initialization).

Given $x_0 \in \mathfrak{R}^n$, compute $(R^t J)(x_0)$, $(J^t J + Q)(x_0)$.

Choose $\xi_1 < \xi_2$, $\eta_1 \in (0, 1)$, $\eta_2 > 1$, $\Delta_1 > 0$, $k = 1$.
- (2) Step 1 (construction a basis for the Krylov subspace).
 - (a) Choose $\epsilon \in (0, 1)$, an initial guess s_0 which can be chosen to be zero, form $r_0 = -R^t J - (J^t J + Q)s_0$, where $R^t J = (R^t J)(x_n)$, $(J^t J + Q) = (J^t J + Q)(x_n)$, and compute $\tau = \|r_0\|$ and $v_1 = r_0/\tau$.
 - (b) For $j = 1, 2, \dots$ until convergence, do.
 - (i) Form $(J^t J + Q)v_j$ and orthogonalize it against the previous v_1, v_2, \dots, v_j then

$$h(i, j) = ((J^t J + Q)v_j, v_i), \quad i = 1, 2, \dots, j,$$

$$\hat{v}_{j+1} = (J^t J + Q)v_j - \sum_{i=1}^j h(i, j)v_i,$$

$$h(i+1, j) = \|\hat{v}_{j+1}\|.$$

$$v_{j+1} = \hat{v}_{j+1}/h(i+1, j).$$
 - (ii) Compute the residual norm $\beta_j = \|R^t J + (J^t J + Q)s_j\|$ of the solution s_j that would be obtained if we stopped.
 - (iii) If $\beta_j \leq \epsilon_j \|R^t J\|$, where $\epsilon_j \leq (1/\mathfrak{D}_j) \in (0, \epsilon]$, and \mathfrak{D}_j is an approximation to the condition number of $(J^t J + Q)_j$, then set $k = j$ and go to 3, else continue.

(3) Step 2 (compute the trial step).

(i) Construct the local model,

$$\psi_k(\mathbf{y}) = f_k + h_k^t \mathbf{y} + \frac{1}{2} \mathbf{y}^t H_k \mathbf{y}, \quad (3.1)$$

where $h_k = V_k^t (R^t J)_k \in \mathfrak{R}^{k_j}$ and $H_k = V_k^t (J^t J + Q)_k V_k \in \mathfrak{R}^{k_j \times k_j}$.

(ii) Compute the solution \mathbf{y} to the problem $\min\{\psi_k(\mathbf{y}) : \|\mathbf{y}\| \leq \Delta_k\}$ and set $s_k = V_k \mathbf{y}$.

(4) Step 3 (test the step and update Δ_k).

(i) Evaluate $f_{k+1} = f(x_k + V_k \mathbf{y})$, $\text{ared}_k = f_k - f(x_k + V_k \mathbf{y})$ and $\text{pred}_k = f_k - \psi(V_k \mathbf{y})$.

(ii) If $(\text{ared}_k / \text{pred}_k) < \xi_1$ then

begin $\Delta_k = \eta_1 \Delta_k$ go to 3(i)

Comment: Do not accept the step " $\|\mathbf{y}\|$ ", reduce the trust region radius " Δ_k " and compute a new step

Else if $\xi_1 \leq (\text{ared}_k / \text{pred}_k) \leq \xi_2$ then

$x_{k+1} = x_k + V_k \mathbf{y}$

Comment: Accept the step and keep the previous trust region radius the same

Else $(\text{ared}_k / \text{pred}_k) > \xi_2$ then

$\Delta_k = \eta_2 \Delta_k$

Comment: Accept the step and increase the trust region radius

end if

(5) Step 4.

Set $k + 1 \leftarrow k$ and go to Step 1.

4. The Restarting Mechanism

In this section, we discuss the important role of the restarting mechanism to control the possibility of the failure of nonsingularity of the Hessian matrix, and introduce the assumptions under which we prove the global convergence.

Let the sequence of the iterates generated by Algorithm 3.1 be $\{x_k\}$; for such sequence we assume

(G1) for all k , x_k and $x_k + s_k \in \Omega$ where $\Omega \subset \mathfrak{R}^n$ is a convex set;

(G2) $f \in C^2(\Omega)$ and $(R^t J) = \nabla f(x) \neq 0$ for all $x \in \Omega$;

(G3) $f(x)$ is bounded in norm in Ω and $(J^t J + Q)$ is nonsingular for all $x \in \Omega$;

(G4) for all s_k such that $x_k + s_k \in \Omega$, the termination condition, $\|(R^t J)_k + (J^t J + Q)_{s_k}\| \leq \epsilon_j \|(R^t J)_k\|$, $\epsilon_k \in (0, 1)$ is satisfied;

An immediate consequence of the global assumptions is that the Hessian matrix $(J^t J + Q)(x)$ is bounded, that is, there exists a constant $\beta_1 > 0$ such that $\|(J^t J + Q)(x)\| < \beta_1$. Therefore, the condition numbers of the Hessian, $\mathfrak{D}_k = \text{cond}_2((J^t J + Q)_k)$ are bounded.

Assumption G3 that $(J^t J + Q)$ is nonsingular is necessary since the IRA iteration is only guaranteed to converge for nonsingular systems. We first discuss the two ways in which the IRA method can breakdown. This can happen if either $\hat{v}_{k+1} = 0$ so that v_{k+1} cannot be formed, or if $H_{k_{\max}}$ is singular which means that the maximum number of IRA steps has been taken, but the final iterate cannot be formed. The first situation has often been referred to as a soft breakdown, since $\hat{v}_{k+1} = 0$ implies $(J^t J + Q)_k$ is nonsingular and x_k is the solution [18]. The second case is more serious in that it causes a convergence failure of the Arnoldi process. One possible recourse is to hope that $(J^t J + Q)_k$ is nonsingular for some k among $1, 2, \dots, k_{\max}-1$. If such k exists, we can compute x_j and then restart the algorithm using x_k as the new initial guess x_0 . It may not be possible to do this.

To handle the problem of singular H_k , we can use a technique similar to that done in the full dimensional space. If H_k is singular, or its condition number is greater than $1/\sqrt{\nu}$, where ν is the machine epsilon, then we perturb the quadratic model $\psi(y)$ to

$$\hat{\psi}(y) = \psi(y) + \frac{1}{2}\gamma y^t y = f + h^t y + \frac{1}{2}y^t (H_k + \gamma I)y, \quad (4.1)$$

where $\gamma = \sqrt{n\nu}\|H_k\|_1$. The condition number of $(H_k + \gamma I)$ is roughly $1/\sqrt{\nu}$. For more details about this technique see reference [4].

We have discussed the possibility of stagnation in the linear IRA method which results in a break down in the nonlinear iteration. Sufficient conditions under which stagnation of the linear iteration never occurs are

- (1) we have to ensure that the steepest descent direction belongs to the subspace, \mathcal{S}_k . Indeed in Algorithm 3.1, the Krylov subspace will contain the steepest descent direction because $v_1 = -R^t J / \|R^t J\|$ and the Hessian is nonsingular;
- (2) there is no difficulty, if one required the Hessian matrix, $(J^t J + Q)(x)$ to be positive definite for all $x \in \Omega$, then the termination condition can be satisfied, and so convergence of the sequence of iterates will follow. This will be the case for some problems, but, requiring $(J^t J + Q)(x)$ to be positive definite every where is very restrictive.

One of the main restrictions of most of the Newton-Krylov schemes is that the subspace onto which a given Newton step is projected must solve the Newton equations with a certain accuracy which is monitored by the termination condition (assumption G4). This condition is enough to essentially guarantee convergence of the trust region algorithm. Practically, the main difficulty is that one does not know in advance if the subspace chosen for projection will be good enough to guarantee this condition. Thus, k_j can be chosen as large as the termination condition will eventually be satisfied, but, when k_j is too large the computational cost and the storage become too high. An alternative to that is to restart the algorithm keeping $(J^t J + Q)$ nonsingular and $k_j < n$. Moreover, preconditioning and scaling are essential for the successful application of these schemes [19].

5. Global Convergence Analysis

In this section, we are going to establish some convergence properties which are possessed by Algorithm 3.1. The major differences between the proposed results and the preexisting

ones arise from the fact that a lower dimensional quadratic model is used, rather than the full dimension model that is used in the preexisting results [10].

Lemma 5.1. *Let the global assumptions hold. Then for any $s \in \Omega$, one has*

$$\frac{|(R^t J)_k^t s|}{\|s\|_2} \geq \frac{\mathfrak{D}_k - \epsilon_k \|(J^t J + Q)_k\|}{(1 + \epsilon_k)\mathfrak{D}_k} \|(R^t J)_k\| > 0. \quad (5.1)$$

Proof. Suppose r_k be the residual associated with s so that $r_k = (R^t J)_k + (J^t J + Q)_k s$ and $(R^t J)_k \neq 0$. Then $\|r_k\| \leq \epsilon_k \|(R^t J)_k\|$, and $s = -(J^t J + Q)_k^{-1} ((R^t J)_k - r_k)$. So,

$$(R^t J)_k^t s = (R^t J)_k^t [- (J^t J + Q)_k^{-1} ((R^t J)_k - r_k)] = - (R^t J)_k^t (J^t J + Q)_k^{-1} (R^t J)_k + (R^t J)_k^t r_k. \quad (5.2)$$

Hence,

$$\begin{aligned} \frac{|(R^t J)_k^t s|}{\|s\|} &= \frac{|(R^t J)_k^t (J^t J + Q)_k^{-1} (R^t J)_k - (R^t J)_k^t r_k|}{\|(J^t J + Q)_k^{-1} ((R^t J)_k - r_k)\|} \\ &\geq \frac{(R^t J)_k^t (J^t J + Q)_k^{-1} (R^t J)_k - |(R^t J)_k^t r_k|}{\|(J^t J + Q)_k^{-1} ((R^t J)_k - r_k)\|}. \end{aligned} \quad (5.3)$$

Next, $\|r_k\| \leq \epsilon \|(R^t J)_k\|$ implies $|(R^t J)_k^t r_k| \leq \epsilon_k \|(R^t J)_k\|^2$ which gives

$$\begin{aligned} (R^t J)_k^t (J^t J + Q)_k^{-1} (R^t J)_k - |(R^t J)_k^t r_k| &\geq (\|(J^t J + Q)_k^{-1}\| - \epsilon_k) \|(R^t J)_k\|^2, \\ \|(J^t J + Q)_k^{-1} ((R^t J)_k - r_k)\| &\leq \|(J^t J + Q)_k^{-1}\| \|(R^t J)_k\| + \|(J^t J + Q)_k^{-1} r_k\| \\ &\leq (1 + \epsilon_k) \|(J^t J + Q)_k^{-1}\| \|(R^t J)_k\|. \end{aligned} \quad (5.4)$$

Introduce (5.4) into (5.3), we obtain the following:

$$\begin{aligned} \frac{|(R^t J)_k^t s|}{\|s\|} &\geq \frac{(\|(J^t J + Q)_k^{-1}\| - \epsilon_k) \|(R^t J)_k\|^2}{(1 + \epsilon_k) \|(J^t J + Q)_k^{-1}\| \|(R^t J)_k\|} \\ &\geq \frac{(\|(J^t J + Q)_k^{-1}\| - \epsilon_k)}{(1 + \epsilon_k) \|(J^t J + Q)_k^{-1}\|} \|(R^t J)_k\| \\ &= \left(\frac{\mathfrak{D}_k - \epsilon_k \|(J^t J + Q)_k\|}{(1 + \epsilon_k)\mathfrak{D}_k} \right) \|(R^t J)_k\|. \end{aligned} \quad (5.5)$$

□

Inequality (5.1) can be written as

$$\frac{|(R^t J)_k^t s|}{\|s\| \|(R^t J)_k\|} \geq \frac{\mathfrak{D}_k - \epsilon_k \|(J^t J + Q)_k\|}{(1 + \epsilon_k) \mathfrak{D}_k}. \quad (5.6)$$

This condition (5.6) tells us at each iteration of Algorithm 3.1, we require that the termination condition holds. Since the condition numbers, \mathfrak{D}_k , are bounded from above, condition (5.6) gives

$$\cos \theta \geq \frac{\mathfrak{D} - \epsilon \|(J^t J + Q)_k\|}{(1 + \epsilon) \mathfrak{D}}, \quad (5.7)$$

where θ is the angle between $(R^t J)_k$ and s and \mathfrak{D} are the upper bound of the condition numbers \mathfrak{D}_k . Inequality (5.7) shows that the acute angle θ is bounded away from $\pi/2$.

The following lemma shows that the termination norm assumption G4 implies that the cosine of the angle between the gradient and the Krylov subspace is bounded below.

Lemma 5.2. *Let the global assumptions hold. Then*

$$\|V_k^t (R^t J)_k\| \geq \frac{\mathfrak{D}_k - \epsilon_k \|(J^t J + Q)_k\|}{(1 + \epsilon_k) \mathfrak{D}_k} \|(R^t J)_k\|. \quad (5.8)$$

Proof. Suppose $s = V_k y$ be a vector of \mathcal{S}_k satisfying the termination condition, $\|(R^t J)_k + (J^t J + Q)_k s\| \leq \epsilon_k \|(R^t J)_k\|$, $\epsilon_k \in (0, 1)$. From Lemma 5.1 and the fact that $\|s\| = \|V y\| = \|y\|$, we obtain

$$\frac{|(R^t J)_k^t V_k y|}{\|y\|} \geq \frac{\mathfrak{D}_k - \epsilon_k \|(J^t J + Q)_k\|}{(1 + \epsilon_k) \mathfrak{D}_k} \|(R^t J)_k\|. \quad (5.9)$$

From Cauchy-Schwartz inequality, we obtain

$$|(R^t J)_k^t V_k y| \leq \|V_k^t (R^t J)_k\| \|y\|. \quad (5.10)$$

Combining formulas (5.9) and (5.10), we obtain the result. \square

The following Lemma establishes that Algorithm 3.1 converges with a decrease in the quadratic model on the lower dimensional space at least equal to the decrease in the quadratic model on the full dimensional space.

Lemma 5.3. *Let the global assumptions hold and let \mathbf{y} be a solution of the minimization problem, $\min_{\|\mathbf{y}\| \leq \Delta} \varphi(\mathbf{y})$. Then*

$$\|h_k\| \min \left\{ \Delta_k, \frac{\|h_k\|}{\|H_k\|} \right\} \geq \alpha_k \|(R^t J)_k\| \min \left\{ \Delta_k, \frac{\alpha_k \|(R^t J)_k\|}{\|(J^t J + Q)_k\|} \right\}, \quad (5.11)$$

where $\alpha_k = (\mathfrak{D}_k - \epsilon_k \|(J^t J + Q)_k\|) / (1 + \epsilon_k) \mathfrak{D}_k$.

Proof. Since, $\varphi(\mathbf{y}) = \phi(V_k \mathbf{y}) = f_k + (R^t J)_k^t V_k \mathbf{y} + (1/2)(V_k \mathbf{y})^t (J^t J + Q)(V_k \mathbf{y})$. Thus $\nabla \varphi(0) = V_k^t (R^t J)_k$. From Step 3 of Algorithm 3.1, we obtain

$$f_k - f_{k+1} \geq \xi_1 (f_k - \phi_k(\mathbf{y})) \quad (5.12)$$

$$\geq \frac{\xi_1}{2} \|V_k^t (R^t J)_k\| \min \left\{ \Delta_k, \frac{\|V_k^t (R^t J)_k\|}{\|V_k^t (J^t J + Q)_k V_k\|} \right\} \quad (5.13)$$

$$= \frac{\xi_1}{2} \|h_k\| \min \left\{ \Delta_k, \frac{\|h_k\|}{\|H_k\|} \right\}. \quad (5.14)$$

We have to convert the lower bound of this inequality to one involving $\|(R^t J)_k\|$ rather than $\|V_k^t (R^t J)_k\|$. Using Lemma 5.2, we obtain $\|h_k\| \geq \alpha_k \|(R^t J)_k\|$, but, $\|H_k\| = \|V_k^t (J^t J + Q)_k V_k\| \leq \|(J^t J + Q)_k\|$. Substituting in (5.14), we obtain

$$\frac{\|h_k\|}{2} \min \left\{ \Delta_k, \frac{\|h_k\|}{\|H_k\|} \right\} \geq \frac{\alpha_k}{2} \|(R^t J)_k\| \min \left\{ \Delta_k, \frac{\alpha_k \|(R^t J)_k\|}{\|(J^t J + Q)_k\|} \right\}, \quad (5.15)$$

which completes the proof. \square

The following two facts will be used in the remainder of the proof.

Fact 1.

By Taylor's theorem for any k and $\Delta > 0$, we have

$$\begin{aligned} |\text{ared}_k(\Delta) - \text{pred}_k(\Delta)| &= \left| (f_k - f_{k+1}) - \left(-h_k^t \mathbf{y}_k(\Delta) - \frac{1}{2} \mathbf{y}_k^t(\Delta) V_k^t (J^t J + Q)_k V_k \mathbf{y}_k(\Delta) \right) \right| \\ &= \left| \frac{1}{2} \mathbf{y}_k^t(\Delta) H_k \mathbf{y}_k(\Delta) - \int_0^1 \mathbf{y}_k^t(\Delta) H_k(\eta) \mathbf{y}_k(\Delta) (1 - \eta) d\eta \right| \\ &\leq \|\mathbf{y}_k(\Delta)\|^2 \int_0^1 \|H_k - H_k(\eta)\| (1 - \eta) d\eta. \end{aligned} \quad (5.16)$$

where, $H_k(\eta) = V_k^t(J^t J + Q)(x_k + \eta s_k)V_k$ with $s_k = x_{k+1} - x_k = V y_k$. Thus,

$$\left| \frac{\text{ared}_k(\Delta)}{\text{pred}_k(\Delta)} - 1 \right| \leq \frac{\|y_k(\Delta)\|^2}{|\text{pred}_k(\Delta)|} \int_0^1 \|H_k - H_k(\eta)\| (1 - \eta) d\eta. \tag{5.17}$$

Fact 2.

For any sequence $\{x_k\}$ generated by an algorithm satisfying the global assumptions, the related sequence $\{f_k\}$ is monotonically decreasing and bounded from below, that is, $f_k \rightarrow f_*$ as $k \rightarrow \infty$.

The next result establishes that every limit point of the sequence $\{x_k\}$ satisfies the first-order necessary conditions for a minimum.

Lemma 5.4. *Let the global assumptions hold and Algorithm 3.1 is applied to $f(x)$, generating a sequence $\{x_k\}$, $x_k \in \Omega$, and $k = 1, 2, \dots$, then $h_k \rightarrow 0$.*

Proof. Since $e_k \leq \epsilon < 1$ for all k , we have

$$\alpha_k \geq \frac{\mathfrak{D} - \epsilon \|(J^t J + Q)_k\|}{(1 + \epsilon)\mathfrak{D}} = \alpha > 0 \quad \forall k. \tag{5.18}$$

Consider any x_l with $\|(R^t J)_l\| \neq 0$, $\|(R^t J)(x) - (R^t J)_l\| \leq \beta_1 \|x - x_l\|$. Thus, if $\|x - x_k\| \leq (\|(R^t J)_k\|/2\beta_1)$, then $\|(R^t J)(x)\| \geq \|(R^t J)_k\| - \|(R^t J)(x) - (R^t J)_k\| \geq (\|(R^t J)_k\|/2)$. Let, $\rho = \|(R^t J)_k\|/2$ and $\mathcal{B}_\rho = \{x : \|x - x_k\| < \rho\}$. There are two cases, either for all $k \geq l$, $x_l \in \mathcal{B}_\rho$ or eventually $\{x_l\}$ leaves the ball \mathcal{B}_ρ . Suppose $x_l \in \mathcal{B}_\rho$ for all $k \geq l$, then $\|g_l\| \geq \|g_k\|/2 = \bar{\sigma}$ for all $l > k$. Thus by Lemma 5.3, we have

$$\text{pred}_k(\Delta) \geq \frac{\xi_1 \alpha \bar{\sigma}}{2} \min \left\{ \Delta_k, \frac{\alpha \bar{\sigma}}{\|H_k\|} \right\} \quad \forall k \geq l. \tag{5.19}$$

Put $\alpha \bar{\sigma} = \sigma$, introduce inequality (5.19) into (5.17) and since $\|H_k\| \leq \|(J^t J + Q)_k\| \leq \beta_1$, we obtain

$$\begin{aligned} \left| \frac{\text{ared}_k(\Delta)}{\text{pred}_k(\Delta)} - 1 \right| &\leq \frac{\|y_k\|^2 \int_0^1 \|H_k - H_k(\eta)\| (1 - \eta) d\eta}{\xi_1 \sigma \min \{ \Delta_k, \sigma / \beta_1 \}} \\ &\leq \frac{\Delta_k^2 (2\beta_1)}{\xi_1 \sigma \min \{ \Delta_k, \sigma / \beta_1 \}} \\ &\leq \frac{2\Delta_k \beta_1}{\xi_1 \sigma}, \quad \Delta_k \leq \frac{\sigma}{\beta_1} \quad \forall k \geq l. \end{aligned} \tag{5.20}$$

This gives for Δ_k sufficiently small and $k \geq l$ that

$$\frac{\text{ared}_k(\Delta)}{\text{pred}_k(\Delta)} > \xi_2. \tag{5.21}$$

In addition, we have

$$\|(J^t J + Q)_k^{-1} h_k\| \geq \frac{\|h_k\|}{\|H_k\|} \geq \frac{\alpha \|(R^t J)_k\|}{\beta_1} \geq \frac{\sigma}{\beta_1}. \quad (5.22)$$

Inequalities (5.21) and (5.22) tell us that, for Δ_k sufficiently small, we cannot get a decrease in Δ_k in Step 3(ii) of Algorithm 3.1. It follows that Δ_k is bounded away from 0, but, since

$$f_k - f_{k+1} = \text{ared}_k(\Delta) \geq \xi_1 \text{pred}_k(\Delta) \geq \xi_1 \sigma \min \left\{ \Delta_k, \frac{\sigma}{\beta_1} \right\}, \quad (5.23)$$

and since f is bounded from below, we must have $\Delta_k \rightarrow 0$ which is a contradiction. Therefore, $\{x_k\}$ must be outside \mathcal{B}_ρ for some $k > l$.

Let x_{j+1} be the first term after x_l that does not belong to \mathcal{B}_ρ . From inequality (5.23), we get

$$\begin{aligned} f(x_l) - f(x_{j+1}) &= \sum_{k=l}^j (f_k - f_{k+1}) \\ &\geq \sum_{k=l}^j \xi_1 \text{pred}_k(\Delta) \\ &\geq \sum_{k=l}^j \xi_1 \sigma \min \left\{ \Delta_k, \frac{\sigma}{\beta_1} \right\}. \end{aligned} \quad (5.24)$$

If $\Delta_k \leq \sigma/\beta_1$ for all $l \leq k \leq j$, we have

$$f(x_l) - f(x_{j+1}) \geq \xi_1 \sigma \sum_{k=l}^j \Delta_k \geq \xi_1 \sigma \rho. \quad (5.25)$$

If there exists at least one k with $\Delta_k > \sigma/\beta_1$, we get

$$f(x_m) - f(x_{m+1}) \geq \xi_1 \frac{\sigma^2}{\beta_1}. \quad (5.26)$$

In either case, we have

$$\begin{aligned} f(x_m) - f(x_{j+1}) &\geq \xi_1 \sigma \min \left\{ \rho, \frac{\sigma}{\beta_1} \right\} \\ &= \xi_1 \alpha \frac{\|(R^t J)_k\|}{2} \min \left\{ \frac{\alpha \|(R^t J)_k\|}{2\beta_1}, \frac{\alpha \|(R^t J)_k\|}{2\beta_1} \right\} \\ &\geq \frac{\xi \alpha^2}{4\beta_1} \|(R^t J)_k\|^2. \end{aligned} \quad (5.27)$$

Since, f is bounded below, $\{f_k\}$ is a monotonically decreasing sequence (i.e., $f_k \rightarrow f_*$). Hence,

$$\|(R^t J)_k\|^2 \leq \frac{4\beta_1}{\xi_1 \alpha^2} (f_k - f_*), \quad (5.28)$$

(i.e., $(R^t J)_k \rightarrow 0$ as $k \rightarrow \infty$). The proof is completed by using Lemma 5.2. \square

The following lemma proves that under the global assumptions, if each member of the sequence of iterates generated by Algorithm 3.1 satisfies the termination condition of the algorithm, then this sequence converges to one of its limit points.

Lemma 5.5. *Let the global assumptions hold, $H_k = V_k^t (J^t J + Q)(x_k) V_k$ for all k , $(J^t J + Q)(x)$ is Lipschitz continuous with L , and x_* is a limit point of the sequence $\{x_k\}$ with $(J^t J + Q)(x_*)$ be positive definite. Then $\{x_k\}$ converges q -Superlinearly to x_* .*

Proof. Let $\{x_{k_m}\}$ be a subsequence of the sequence $\{x_k\}$ that converges to x_* , where, x_* is a limit of $\{x_k\}$. From Lemma 5.4, $(R^t J)(x_*) = 0$. Since $(J^t J + Q)(x_*)$ is positive definite and $(J^t J + Q)(x)$ is continuous there exists $\delta_1 > 0$ such that if $\|x - x_*\| < \delta_1$, then $(J^t J + Q)(x)$ is positive definite, and if $x \neq x_*$ then $(R^t J)(x_*) \neq 0$. Let, $B_1 = \{x : \|x - x_*\| < \delta_1\}$.

Since $(R^t J)(x_*) = 0$, we can find $\delta_2 > 0$ with $\delta_2 < \delta_1/4$ and $\|(J^t J + Q)^{-1}\| \|(R^t J)\| \leq \delta_1/2$ for all $x \in B_2 = \{x : \|x - x_*\| < \delta_2\}$.

Find l_0 such that $f(x_{l_0}) < \inf\{f(x) : x \in B_1 - B_2\}$ and $x_{l_0} \in B_2$. Consider any x_i with $i \geq k_l$, $x_i \in B_2$.

We claim that $x_{i+1} \in B_2$, which implies that the entire sequence beyond $x_{k_{l_0}} \in B_2$. Suppose x_{i+1} does not belong to B_2 . Since $f_{i+1} < f_{k_{l_0}}$, x_{i+1} does not belong to B_1 either. So

$$\begin{aligned} \Delta_i &\geq \|x_{i+1} - x_i\| \\ &\geq \|x_{i+1} - x_*\| - \|x_i - x_*\| \\ &\geq \delta_1 - \frac{\delta_1}{4} \\ &= \frac{3\delta_1}{4} \\ &> \frac{\delta_1}{2} \\ &\geq \|(J^t J + Q)_i^{-1}\| \|(R^t J)_i\| \\ &\geq \|H_i^{-1}\| \|h_i\|. \end{aligned} \quad (5.29)$$

So, the truncated-Newton step is within the trust region, we obtain

$$y_i(\Delta_i) = -(V_i^t (J^t J + Q)_i V_i)^{-1} V_i^t (R^t J)_i, \quad (5.30)$$

Since $\|y_i(\Delta)\| < \delta_1/2$, it follows that $x_{i+1} \in B_1$, which is a contradiction. Hence, for all $k \geq k_0$, $x_k \in B_2$, and so since $f(x_k)$ is strictly decreasing sequence and x_* is the unique minimizer of f in B_2 , we have that $\{x_k\} \rightarrow x_*$. \square

The following lemma establishes the rate of convergence of the sequence of iterates generated by Algorithm 3.1 to be q -super linear.

Lemma 5.6. *Let the assumptions of Lemma 5.5 hold. Then the rate of convergence of the sequence $\{x_k\}$ is q -Super linear.*

Proof. To show that the convergence rate is super linear, we will show eventually that $\|H_k^{-1}h_k\|$ will always be less than Δ_j and hence the truncated-Newton step will always be taken. Since $(J^t J + Q)(x_*)$ is positive definite, it follows that the convergence rate of $\{x_k\}$ to x_* is super linear.

To show that eventually the truncated-Newton step is always shorter than the trust region radius, we need a particular lower bound on $\text{pred}_k(\Delta)$. By assumptions, for all k large enough, H_k is positive definite. Hence, either the truncated-Newton step is longer than the trust region radius, or $y_k(\Delta)$ is the truncated-Newton step. In either case

$$\|y_k(\Delta)\| \leq \|(V_k^t(J^t J + Q)V_j)^{-1}V_k^t(R^t J)_k\| \leq \|(V_k^t(J^t J + Q)_k V_k)^{-1}\| \|V_k^t(R^t J)_k\|, \quad (5.31)$$

and so it follows that $\|V_k^t(R^t J)_k\| \geq \|s_k(\Delta)\| / \|(V_k^t(J^t J + Q)_k V_k)^{-1}\|$. By Lemma 5.3, for all k large enough, we have

$$\text{pred}_k(\Delta) \geq \frac{\xi_1 \alpha}{2} \frac{\|y_k(\Delta)\|}{\|(J^t J + Q)_k^{-1}\|} \min \left\{ \Delta_k, \frac{\alpha \|(R^t J)_k\|}{\|(J^t J + Q)_k\|} \right\}. \quad (5.32)$$

Thus,

$$\begin{aligned} \text{pred}_k(\Delta) &\geq \frac{\xi_1 \alpha}{2} \frac{\|y_k(\Delta)\|}{\|(J^t J + Q)_k^{-1}\|} \min \left\{ \|y_k(\Delta)\|, \frac{\alpha \|y_k(\Delta)\|}{\|(J^t J + Q)_k\| \|(J^t J + Q)_k^{-1}\|} \right\} \\ &\geq \frac{\xi_1 \alpha^2}{2} \frac{\|y_k(\Delta)\|^2}{\mathfrak{D} \|(J^t J + Q)_k^{-1}\|}. \end{aligned} \quad (5.33)$$

So, by the continuity of $(J^t J + Q)(x)$, for all k large enough, we obtain

$$\text{pred}_k(\Delta) \geq \frac{\xi_1 \alpha^2}{4} \frac{\|y_k(\Delta)\|^2}{\|(J^t J + Q)^{-1}(x_*)\|}. \quad (5.34)$$

Finally, by the argument leading up to Fact 1 and the Lipschitz continuity assumption, we have

$$|\text{ared}_k(\Delta) - \text{pred}_k(\Delta)| \leq \|y_k(\Delta)\|^3 \frac{L}{2}. \quad (5.35)$$

Thus, for any $\Delta_k > 0$ and k large enough, we obtain

$$\begin{aligned} \left| \frac{\text{ared}_k(\Delta)}{\text{pred}_k(\Delta)} - 1 \right| &\leq \|s_k(\Delta)\| \frac{3L}{2} \frac{4\|(J^t J + Q)^{-1}(x_*)\|}{\xi_1 \alpha^2 \|y_k(\Delta)\|^2} \\ &= \frac{2L}{\xi_1 \alpha^2} \|(J^t J + Q)^{-1}(x_*)\| \|y_k(\Delta)\| \\ &\leq \frac{2L\|(J^t J + Q)^{-1}(x_*)\|}{\xi_1 \alpha^2} \Delta_k. \end{aligned} \tag{5.36}$$

Thus by Step 3(ii) of Algorithm 3.1, there is a $\bar{\Delta}$ such that if $\Delta_{k-1} < \bar{\Delta}$, then Δ_k will be less than Δ_{k-1} only if $\Delta_k \geq \|(V_{k-1}^t(J^t J + Q)_{k-1} V_{k-1})^{-1} V_{k-1}^t(R^t J)_{k-1}\|$. It follows from the superlinear convergence of the truncated-Newton method that for x_{k-1} close enough to x_* and k large enough,

$$\|(V_k^t(J^t J + Q)_k V_k)^{-1} V_k^t(R^t J)_k\| < \|(V_{k-1}^t(J^t J + Q)_{k-1} V_{k-1})^{-1} V_{k-1}^t(R^t J)_{k-1}\|. \tag{5.37}$$

Now, if Δ_k is bounded away from 0 for all large k , then we are done. Otherwise, if for an arbitrarily large k , Δ_k is reduced, that is, $\Delta_k < \Delta_{k-1}$, then we have

$$\Delta_k \geq \|(V_{k-1}^t(J^t J + Q)_{k-1} V_{k-1})^{-1} V_{k-1}^t(R^t J)_{k-1}\| > \|(V_k^t(J^t J + Q)_k V_k)^{-1} V_k^t(R^t J)_k\|, \tag{5.38}$$

and so the full truncated-Newton step is taken. Inductively, this occurs for all subsequence iterates and super linear convergence follows. \square

The next result shows that under the global assumptions every limit point of the sequence $\{x_k\}$ satisfies the second-order necessary conditions for a minimum.

Lemma 5.7. *Let the global assumptions hold, and assume $\text{pred}_k(\Delta) \geq \bar{\alpha}_2(-\lambda_1(H))\Delta^2$ for all the symmetric matrix $H \in \mathfrak{R}^{k_j \times k_j}$, $h \in \mathfrak{R}^{k_j}$, $\Delta > 0$ and some $\bar{\alpha}_2 > 0$. If $\{x_k\} \rightarrow x_*$, then $H(x_*)$ is positive semidefinite, where $\lambda_1(H)$ is the smallest eigenvalue of the matrix H .*

Proof. Suppose to the contrary that $\lambda_1(H(x_*)) < 0$. There exists K such that if $k \geq K$, then $\lambda_1(H_k) < (1/2)\lambda_1(H(x_*))$. For all $k \geq K$ and for all $\Delta_k > 0$, we obtain

$$\text{pred}_k(\Delta) \geq \bar{\alpha}_2(-\lambda_1(H_k))\Delta_k^2 \geq \frac{\bar{\alpha}_2}{2}(-\lambda_1(H(x_*)))\Delta_k^2. \tag{5.39}$$

Using inequality (5.17), we obtain

$$\begin{aligned} \left| \frac{\text{ared}_k(\Delta)}{\text{pred}_k(\Delta)} - 1 \right| &\leq \frac{\|y_k(\Delta)\|^2}{\|\text{pred}_k(\Delta)\|} \int_0^1 \|H_k(\eta) - H_k\|(1-\eta)d\eta \\ &\leq \frac{2\|y_k(\Delta)\|^2}{\Delta_k^2} \int_0^1 \frac{\|H_k(\eta) - H_k\|(1-\eta)d\eta}{\bar{\alpha}_2(-\lambda_1(H(x_*)))}. \end{aligned} \tag{5.40}$$

Since the last quantity goes to zero as $\Delta_k \rightarrow 0$ and since a Newton step is never taken for $k > K$, it follows from Step 3(ii) of Algorithm 3.1 that for $k \geq K$ Δ_k sufficiently small. Δ_k cannot be decreased further. Thus, Δ is bounded below, but

$$\text{ared}_k(\Delta) \geq \xi \text{pred}_k(\Delta) \geq \frac{\xi \bar{\alpha}_2}{2} (-\lambda_1(H(x_*))) \Delta_k^2. \quad (5.41)$$

Since f is bounded below, $\text{ared}_k(\Delta) \rightarrow 0$, so $\Delta_k \rightarrow 0$ which is a contradiction. Hence $\lambda_1(H(x_*)) \geq 0$. This completes the proof of the lemma. \square

6. Concluding Remarks

In this paper, we have shown that the implicitly restarted Arnoldi method can be combined with the Newton iteration and trust region strategy to obtain a globally convergent algorithm for solving large-scale NLS problems.

The main restriction of this scheme is that the subspace onto which a given Newton step is projected must solve the Newton equations with a certain accuracy which is monitored by the termination condition. This is enough to guarantee convergence.

This theory is sufficiently general that it holds for any algorithm that projects the problem on a lower dimensional subspace. The convergence results indicate promise for this research to solve large-scale NLS problems. Our next step is to investigate the performance of this algorithm on some NLS problems. The results will be reported in our forthcoming paper.

Acknowledgment

The authors would like to thank the editor-in-chief and the anonymous referees for their comments and useful suggestions in the improvement of the final form of this paper.

References

- [1] G. Golub and V. Pereyra, "Separable nonlinear least squares: the variable projection method and its applications," *Inverse Problems*, vol. 19, no. 2, pp. R1–R26, 2003.
- [2] J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, NY, USA, 1970.
- [3] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory," in *Numerical Analysis (Proc. 7th Biennial Conf., Univ. Dundee, 1977)*, G. A. Watson, Ed., vol. 630 of *Lecture Notes in Mathematics*, pp. 105–116, Springer, Berlin, Germany, 1978.
- [4] J. E. Dennis Jr. and R. B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall Series in Computational Mathematics, Prentice-Hall, Englewood Cliffs, NJ, USA, 1983.
- [5] M. Gulliksson, I. Söderkvist, and P.-Å. Wedin, "Algorithms for constrained and weighted nonlinear least squares," *SIAM Journal on Optimization*, vol. 7, no. 1, pp. 208–224, 1997.
- [6] M. R. Abdel-Aziz, "Globally convergent algorithm for solving large nonlinear systems of equations," *Numerical Functional Analysis and Optimization*, vol. 25, no. 3-4, pp. 199–219, 2004.
- [7] M. J. D. Powell, "Convergence properties of a class of minimization algorithms," in *Nonlinear Programming, 2 (Proc. Sympos. Special Interest Group on Math. Programming, Univ. Wisconsin, Madison, Wis., 1974)*, O. Mangasarian, R. Meyer, and S. Robinson, Eds., pp. 1–27, Academic Press, New York, NY, USA, 1975.
- [8] J. J. Moré and D. C. Sorensen, "Computing a trust region step," *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 3, pp. 553–572, 1983.

- [9] J. J. Moré, "Recent developments in algorithms and software for trust region methods," in *Mathematical Programming: The State of the Art (Bonn, 1982)*, A. Bacham, M. Grotschel, and B. Korte, Eds., pp. 258–287, Springer, Berlin, Germany, 1983.
- [10] G. A. Shultz, R. B. Schnabel, and R. H. Byrd, "A family of trust-region-based algorithms for unconstrained minimization with strong global convergence properties," *SIAM Journal on Numerical Analysis*, vol. 22, no. 1, pp. 47–67, 1985.
- [11] M. R. Osborne, "Some special nonlinear least squares problems," *SIAM Journal on Numerical Analysis*, vol. 12, no. 4, pp. 571–592, 1975.
- [12] M. Xiaofang, F. R. Ying Kit, and X. Chengxian, "Stable factorized quasi-Newton methods for nonlinear least-squares problems," *Journal of Computational and Applied Mathematics*, vol. 129, no. 1-2, pp. 1–14, 2001.
- [13] J. E. Dennis Jr., H. J. Martínez, and R. A. Tapia, "Convergence theory for the structured BFGS secant method with an application to nonlinear least squares," *Journal of Optimization Theory and Applications*, vol. 61, no. 2, pp. 161–178, 1989.
- [14] P. N. Brown and Y. Saad, "Hybrid Krylov methods for nonlinear systems of equations," *SIAM Journal on Scientific and Statistical Computing*, vol. 11, no. 3, pp. 450–481, 1990.
- [15] D. C. Sorensen, "Minimization of a large scale quadratic function subject to an ellipsoidal constraint," Tech. Rep. TR94-27, Department of Computational and Applied Mathematics, Rice University, Houston, Tex, USA, 1994.
- [16] D. C. Sorensen, "Newton's method with a model trust region modification," *SIAM Journal on Numerical Analysis*, vol. 19, no. 2, pp. 409–426, 1982.
- [17] H. Mukai and E. Polak, "A second-order method for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 26, no. 4, pp. 501–513, 1978.
- [18] Y. Saad, "Krylov subspace methods for solving large unsymmetric linear systems," *Mathematics of Computation*, vol. 37, no. 155, pp. 105–126, 1981.
- [19] S. G. Nash, "Preconditioning of truncated-Newton methods," *SIAM Journal on Scientific and Statistical Computing*, vol. 6, no. 3, pp. 599–616, 1985.