An affine-scaling trust-region method with interior backtracking technique for bound-constrained nonlinear equations

SUN Dong-mei, ZHU De-tong

(College of Mathematics and Sciences, Shanghai Normal University, Shanghai 200234, China)

Abstract: We develop an affine scaling trust region algorithm in association with the nonmonotone interior backtracking line technique for solving smooth nonlinear equations subject to bounds on variables. The trust region subproblem is defined by minimizing a squared Euclidean norm of linear model with a new affine matrix called minimum-scaling. Under a reasonable assumption of this new affine-scaling matrix, we stress that the minimum-scaling has some additional properties that allow us to prove stronger global convergence results without nondegenerate property than those about the Coleman-Li-scaling. The nonmonotonic criterion is used to speed up the convergence progress in the contours of objective function with large curvature.

Key words: bound constraint; trust region; affine scaling; nonmonotone technique

CLC number: O221.2 **Document code:** A **Article ID:** 1000-5137(2007)06-0001-07

1 Introduction

In this paper, we propose and analyze a trust-region method for the solution of a simply constrainted system of smooth nonlinear equations

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

Here, the function $F: \mathbf{R}^n \supset U \to \mathbf{R}^n$ is defined on the open set U containing the feasible set $\Omega = [l, u] = \{ x \in \mathbf{R}^n \mid l_i \leq x_i \leq u_i, i = 1, \cdots, n \}$. The bound $l_i \in \mathbf{R} \cup \{-\infty\}$ and $u_i \in \mathbf{R} \cup \{+\infty\}$ are assumed to satisfy $l_i < u_i$, $i = 1, \cdots, n$, (otherwise the variable $x_i = l_i = u_i$ could be eliminated).

Recently, Bellavia et al in [1] further extended the ideas and presented an affine scaling trust-region approach for solving the bound-constrained smooth nonlinear systems (1.1). More recently, Zhu in [3] proposed the affine scaling trust region Newton methods which switch to strict interior feasibility by line search backtracking technique. In this paper, we introduce another more general affine scaling interior point projective, called the minimum-scaling, to generate the affine scaling trust region methods which switch to strict interior feasibility by line search backtracking technique with some reasonable conditions. We stress that under a

Received date: 2007-09-10

Foundation item: Support of the National Science Foundation Grant of China (10471094).

Biography: Dongmei Sun, (1983 -), female, Graduate Student, College of Mathematics and Sciences, Shanghai Normal University; Detong Zhu, (1954 -), male, Professor, College of Mathematics and Sciences, Shanghai Normal University.

reasonable assumption of this new affine-scaling matrix, the minimum-scaling has some additional properties that allow us to prove stronger global convergence results without nondegenerate property than for the Coleman - Li-scaling. The nonmonotone idea also motivates the study of trust region methods in association with nonmonotone backtracking line search technique for approximating zeros of the smooth equations (1. 1) in [4].

The paper is organized as follows: we derive our affine-scaling trust - region method with nonmonotone interior backtracking technique for the solution of (1. 1) in Section 2. The global convergence properties of this method is investigated in Section 3.

2 Algorithms

In this section, we propose and analyze a trust - region method for the solution of problem (1. 1). Bellavia et al in [1] presented the affine scaling trust - region approach scheme. Note that (1. 1) is closely related to the box constrained optimization problem

$$\min\ f(x),\quad \text{s. t.}\quad x \in \Omega, \tag{2.1}$$

where $f: U \to \mathbf{R}$, $f(x) = \|F(x)\|^2/2$. Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm. Obviously, (1. 1) and (2. 1) are equivalent if (1. 1) possesses a solution.

Now we exploit the relation between the two problems (1. 1) and (2. 1) and follow an idea by Coleman and Li [5] who observed that the first order optimality conditions of (2. 1) are equivalent to the nonlinear system of equations

$$D(x)\nabla f(x) = 0,\quad x \in \Omega, \tag{2.2}$$

with a suitable scaling matrix

$$D(x) = \mathrm{diag}(d_1(x), \cdots, d_n(x)). \tag{2.3}$$

Originally, Coleman and Li [5] consider only one particular choice of the scaling matrix $D(x)$. In this work, we allow the scaling matrix taken from a rather general class, defined by the minimum-scaling

$$d_i^{\min}(x) \overset{\text{def}}{=} \begin{cases} 1 & \text{if } l_i = -\infty \text{ and } u_i = +\infty \\ \min\{x_i - l_i + \gamma\max\{0, -g_i\}, u_i - x_i + \gamma\max\{0, g_i\}\} & \text{otherwise} \end{cases} \tag{2.4}$$

where $\gamma > 0$ is a given constant, $g(x) = \nabla f(x)$, and g_i is the ith component of $g(x)$ (cf. [2]). The scaling matrix may be used to prove suitable global and local convergence results. Nevertheless, we stress that the minimum-scaling has some additional properties that allow us to prove global convergence results without nondegenerate property. But the following assumption is needed in [2].

Assumption 1 The scaling matrix $D(x) \overset{\text{def}}{=} \mathrm{diag}\{d_i^{\min}(x)\}$ satisfies (2. 4) and is bounded on Ω. Furthermore, there exists a constant $\alpha > 0$ such that

$$\alpha\, d_i^{\min}(x) \le \begin{cases} x_i - l_i & \text{if } g_i > 0 \text{ and } l_i > -\infty \\ u_i - x_i & \text{if } g_i < 0 \text{ and } u_i > +\infty \end{cases}, \tag{2.5}$$

for all $i = 1, \cdots, n$ and all $x \in \mathrm{int}(\Omega)$.

Note that the conditions hold automatically if Ω itself is bounded, i. e. , if all lower and upper bounds l_i and u_i are finite. This assumption is quite realistic in many cases since otherwise one may replace infinite bounds by sufficiently large bounds. Furthermore, if the scalar α_k given in (2. 8) of step 3 denotes the step size along the direction p_k to the boundary on the variables $l \le x_k + \alpha_k p_k \le u$, and the Assumption 1 holds, it is clear that $x_{k+1} = x_k + \alpha_k p_k \in \mathrm{int}(\Omega)$. Moreover, it also ensures the nondegenerate property of the system (1. 1) at any limit point.

In order to construct a suitable method for the solution of problem (1. 1), we follow an interior - point trust - region approach for (2. 1) similar to those in [3]. The basic idea is based on the following elliptical trust region subproblem at the kth iteration

$$\min m_k(p) = g_k^T p + \frac{1}{2} p^T (F_k'^T F_k') p$$

$$\text{s. t. } \{p \in \mathbf{R}^n \mid \| D(x_k)^{-1/2} p \| \le \Delta_k\}, \tag{2. 6}$$

where $\Delta_k > 0$ denotes the trust - region radius. To get a feasible solution and relax the acceptability conditions on the trial step p_k, we suggest to use the nonmonotone technique instead of the monotone technique.

Next, we describe an affine scaling trust region algorithm with nonmonotone line search technique for solving the bound - constrained system (1. 1).

Algorithm 2. 1

Initialization step Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0, 1)$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < 1 < \gamma_2$, $\varepsilon > 0$ and positive integer C as nonmonotonic parameter. Let $m(0) = 0$. Select an initial trust region radius $\Delta_0 > 0$ and a maximal trust region radius $\Delta_{\max} \ge \Delta_0$. Give a starting strict feasibility interior point $x_0 \in \mathrm{int}(\Omega) \subseteq \mathbf{R}^n$. Set $k = 0$, and go to the main step.

Main step

(1) If $\| D_k^{1/2} g_k \| = \| D_k^{1/2} (F_k')^T F_k \| \le \varepsilon$, stop with the approximate solution x_k.

(2) Compute p_k by solving the trust region subproblem (2. 6).

(3) Choose $\alpha_k = 1, \omega, \omega^2, \cdots$ until the following inequality is satisfied

$$f(x_k + \alpha_k p_k) \le f(x_{l(k)}) + \alpha_k \beta g_k^T p_k, \tag{2. 7}$$

$$\text{with } x_k + \alpha_k p_k \in \Omega, \tag{2. 8}$$

where $f(x_{l(k)}) = \max_{0 \le j \le m(k)} \{f(x_{k-j})\}$, with the nonmonotone index function $0 \le m(k) \le \min\{m(k-1) + 1, C\}$, $k \ge 1$. Set

$$x_{k+1} = x_k + \alpha_k p_k. \tag{2. 9}$$

(4) Calculate

$$\mathrm{ared}_k(\alpha_k p_k) = f(x_{l(k)}) - f(x_k + \alpha_k p_k),$$

$$\mathrm{pred}_k(\alpha_k p_k) = -m_k(\alpha_k p_k) = f(x_k) - f(x_k + \alpha_k p_k),$$

$$\rho_k \overset{\text{def}}{=} \rho_k(\alpha_k p_k) = \frac{\mathrm{ared}_k(\alpha_k p_k)}{\mathrm{pred}_k(\alpha_k p_k)}. \tag{2. 10}$$

(5) Update the trust - region radius according to the following rules:

$$\Delta_{k+1} = \begin{cases} \gamma_1 \Delta_k, & \rho_k < \eta_1 \\ \Delta_k, & \rho_k \in [\eta_1, \eta_2) \\ \gamma_2 \Delta_k, & \rho_k \ge \eta_2 \end{cases}.$$

(6) Take the nonmonotone control parameter $m(k+1) = \min\{m(k) + 1, C\}$. Then set $k \leftarrow k + 1$ and go to step 1.

3 Global convergence analysis

In our analysis, the level set of f is denoted by

$$\mathbf{L}(x_0) = \{ x \in \mathbf{R}^n \mid f(x) \leq f(x_0), \ l \leq x \leq u \}.$$

The following assumption is used in our convergence analysis.

Assumption 2 The sequence $\{x_k\}$ generated by the algorithm is contained in a compact set $\mathbf{L}(x_0)$ on $\mathbf{R}^n$ and there exist some positive constants $\chi_D, \chi_g, \chi_F > 0$, $\| D(x) \| \leq \chi_D$, $\| g(x) \| \leq \chi_g$, $\forall x \in \mathbf{L}(x_0)$ such that $\| D_k^{1/2} (F_k')^T (F_k') D_k^{1/2} \| \leq \chi_F$, $\forall k$.

First, we state some lemmas, which are essentially the same as that in [2] and [3], that ensures the global convergence of the trust region algorithm.

Lemma 3.1 Suppose that Assumptions (1) and (2) hold, and let p_k be a solution to the subproblem (2.6). Then there exists a constant $\tau > 0$ such that

$$\mathrm{pred}_k(p_k) \geq \tau \| D_k^{1/2} g_k \|^2 \min\{\Delta_k, 1\}. \tag{3.1}$$

Lemma 3.2 At the iteration, let p_k be generated in trust region subproblem (2.6), then

$$g_k^T p_k \leq -\tau \| D_k^{1/2} g_k \|^2 \min\{\Delta_k, 1\}, \tag{3.2}$$

where the constant τ is given in (3.1).

Proof of this lemma is similar to that in [3].

Lemma 3.3 Suppose that Assumption 2 holds, and assume that the gradient of f is such that

$$\| \nabla f(x) - \nabla f(y) \|_2 \leq \gamma \| x - y \|_2, \ x, y \in \mathbf{R}^n, \tag{3.3}$$

where γ is the Lipschitz constant. Let $\beta \in (0, 1)$ and p_k be proposed by the subproblem (2.6). If $\| D_k^{1/2} g_k \| \neq 0$, then Algorithm 2.1 will produce an iterate $x_{k+1} = x_k + \alpha_k p_k$ in a finite number of backtracking steps in (2.7).

Proof of this lemma is similar to that in [3]. The following lemma establishes the necessary and sufficient conditions concerning the pair λ_k, p_k when p_k solves the subproblem (2.6).

Lemma 3.4 p_k is a solution to the subproblem (1.8) if and only if p_k is a solution to the following equations of the forms

$$[D_k^{1/2} F_k'^T F_k' D_k^{1/2} + \lambda_k I] D_k^{-1/2} p_k = - D_k^{1/2} g_k, \tag{3.4}$$

$$\lambda_k (\| D_k^{-1/2} p_k \| - \Delta_k) = 0, \ \lambda_k \geq 0, \tag{3.5}$$

and $[D_k^{1/2} F_k'^T F_k' D_k^{1/2} + \lambda_k I]$ is positive semidefinite.

Theorem 3.1 Assume that Assumptions 1 and 2 hold. Let $\{x_k\} \subset \mathbf{R}^n$ be a sequence generated by the algorithm. Then

$$\liminf_{k \to \infty} \| D_k^{1/2} g_k \| = 0. \tag{3.6}$$

Proof According to the acceptance rule in step 3, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k p_k) \geq -\alpha_k \beta g_k^T p_k = -\alpha_k \beta (D_k^{1/2} g_k)^T (D_k^{-1/2} p_k). \tag{3.7}$$

Taking into account that $m(k+1) \leq m(k) + 1$, and $f(x_{k+1}) \leq f(x_{l(k)})$, we have $f(x_{l(k+1)}) \leq \max_{0 \leq j \leq m(k)+1} \{ f(x_{k+1-j}) \} = f(x_{l(k)})$. This means that the sequence $\{ f(x_{l(k)}) \}$ is nonincreasing for all k and hence $\{ f(x_{l(k)}) \}$ is convergent.

By (2.7) and (3.2), for all $k > C$, we get

$$f(x_{l(k)}) \le \max_{0\le j\le m(l(k)-1)} \{f(x_{l(k)-j-1})\} + \alpha_{l(k)-1}\beta \nabla f^T_{l(k)-1} p_{l(k)-1} \le \tag{3.8}$$

$$\max_{0\le j\le m(l(k)-1)} \{f(x_{l(k)-j-1})\} - \alpha_{l(k)-1}\beta\tau \| D^{1/2}_{l(k)-1} g_{l(k)-1} \|^2 \min\{\Delta_{l(k)-1}, 1\}.$$

If the conclusion of the theorem is not true, there exists some $\varepsilon > 0$ such that

$$\| D^{1/2}_k g_k \| \ge \varepsilon, \quad k = 1, 2, \cdots. \tag{3.9}$$

As $\{f(x_{l(k)})\}$ is convergent, it follows from (3.9) that

$$\lim_{k\to\infty} \alpha_{l(k)-1} \Delta_{l(k)-1} = 0$$

which implies that either

$$\lim_{k\to\infty} \alpha_{l(k)-1} = 0 \tag{3.10}$$

or

$$\lim_{k\to\infty} \inf \Delta_{l(k)-1} = 0. \tag{3.11}$$

By the updating formula of Δ_k, we have $\gamma_1^{C+1}\Delta_{l(k)-1} \le \Delta_k \le \gamma_2^{C+1}\Delta_{l(k)-1}$ which means that from (3.11)

$$\lim_{k\to\infty} \inf \Delta_k = 0. \tag{3.12}$$

Assume that α_k given in step 3 is the stepsize to the boundary of box constraints along p_k. We have

$$\alpha_k \overset{\text{def}}{=} \min\{\max\{\frac{l_i - x_{k,i}}{p_{k,i}}, \frac{u_i - x_{k,i}}{p_{k,i}}\}, i = 1, 2, \cdots, n\}.$$

Combined with (3.12), it means that

$$p_{k,i} \to 0, \quad \text{for all } i$$

Hence, we get that there exists a subset $\kappa \subset \mathbf{K}$ such that

$$\lim_{k\to\infty,\ k\in \mathbf{K}} \alpha_k = 0$$

and hence, without loss of generality, assume $x_{*,i} = l_i$ for some i. Since $\{(F_k^{'T} F_k^{'}) p_k\}$ converges to zero, $\lambda_k I$ is a positive semidefinite diagonal matrix, and x_* is nondegenerate with $D^{1/2}_* g_* = 0$ since Assumption 1 holds, for any i with $(v_*)_i = 0$, and using (3.4), we have that $(p_k)_i$ and $(g_k)_i$ have the same sign for k sufficiently large. Hence, if α_k is defined by some $(v_*)_j = 0$ and hence $(g_*)_j \ne 0$, then $\alpha_k = \frac{|(v_k)_j|}{|(p_k)_j|}$ for k sufficiently large.

Taking norm in the equation (3.4), we can obtain

$$\lambda_k \Delta_k = \lambda_k \| D_k p_k \| \ge \| D^{1/2}_k g_k \| - \| D^{1/2}_k F_k^{'T} F_k^{'} D^{1/2}_k \| \cdot \| D^{-1/2}_k p_k \|. \tag{3.13}$$

Dividing (3.13) by Δ_k and noticing $\| D_k p_k \| \le \Delta_k$, we get

$$\lambda_k \ge \frac{\| D^{1/2}_k g_k \|}{\Delta_k} - \| D^{1/2}_k F_k^{'T} F_k^{'} D^{1/2}_k \|. \tag{3.14}$$

It is clear that from (3.14) and $\| D^{1/2}_k g(x_k) \| \ne 0$, $\lim_{k\to\infty} \lambda_k = +\infty$, as $\Delta_k \to 0$. Using $\alpha_k = \frac{|(v_k)_j|}{|(p_k)_j|}$, (3.4), the assumptions of the boundedness of $\{g_k + F_k^{'T} F_k^{'} p_k\}$ and $\{\lambda_k\}$ converging to $+\infty$, we conclude that

$$\lim_{k\to\infty} \alpha_k = +\infty, \tag{3.15}$$

where α_k given is the step size to the boundary of box constraints along p_k.

Similar to the proof in [3], we can also prove that if

$$\Delta_k \le \frac{\tau\varepsilon^2 (1-\beta)}{\gamma\chi_0}, \tag{3.16}$$

then $\alpha_k = 1$ must satisfy the accepted condition (2.7) in step 3, that is,

$$f(x_k + p_k) \le f(x_{l(k)}) + \beta g_k^T p_k.$$

From the above, we see that if (3.15) holds, then the step size will be determined by (2.7). So, for large k, $\alpha_k = 1$. Since Assumption 1 holds, it is clear that $x_{k+1} = x_k + p_k \in \mathrm{int}(\Omega)$.
Furthermore, for some $\xi_k \in [0, 1]$, we have

$$\mathrm{ared}_k(p_k) = f(x_k) - f(x_k + p_k) = - p\nabla f(x_k + \xi_k p_k)^T p_k =$$

$$\mathrm{pred}_k(p_k) + [\nabla f(x_k) - \nabla f(x_k + \xi_k p_k)]^T p_k + \frac{1}{2} p_k^T (F_k')^T (F_k') p_k. \qquad (3.17)$$

Because $f(x)$ is Lipschitz continuously differentiable with constant γ, we obtain that

$$|[\nabla f(x_k) - \nabla f(x_k + \xi_k p_k)]^T p_k| \le \gamma\chi_D \|D_k^{-1/2} p_k\|^2 \le \gamma\chi_D \Delta_k^2.$$

Using Assumption 2, (3.17) implies that

$$|\mathrm{ared}_k(p_k) - \mathrm{pred}_k(p_k)| \le (\gamma\chi_D + \frac{1}{2}\chi_F)\Delta_k^2.$$

Since Lemma 3.1 holds, we readily obtain that for large k, $\mathrm{pred}(p_k) \ge \tau\varepsilon^2\Delta_k$, if setting

$$\rho_k = \frac{\mathrm{ared}_k(f_k)}{\mathrm{pred}_k(f_k)}, \qquad (3.18)$$

then $\{\rho_k - 1\}$ converges to zero, as $\Delta_k \to 0$. $\rho_k \ge \rho_k \ge \eta_2$ implies that $\{\Delta_k\}$ is not decreased for sufficiently large k and hence bounded away from zero. Thus $\{\Delta_k\}$ cannot converge to zero, contradicting (3.12).

From the above, we conclude that (3.11) is not true. So, (3.10) must hold. Similar to the proof of theorem in [4], we have that if (3.10) holds, then

$$\lim_{k\to\infty}\alpha_k = 0.$$

Now, assume that α_k given in step 3 is the step size to the boundary of box constraints along p_k. From

$$\alpha_k \overset{\mathrm{def}}{=} \min\{\max\{\frac{l_i - x_{k,i}}{p_{k,i}}, \frac{u_i - x_{k,i}}{p_{k,i}}\}, i = 1, 2, \cdots, n\}$$

and Assumption 1

$$\alpha\, d_i^{\min}(x) \le \begin{cases} x_i - l_i, & \text{if } g_i > 0, \text{ and } l_i > -\infty, \\ u_i - x_i, & \text{if } g_i < 0, \text{ and } u_i > +\infty, \end{cases}$$

we get

$$0 < \alpha \le \alpha_k,$$

and then

$$\alpha_k \nrightarrow 0, \qquad (3.19)$$

where α_k given is the step size to the boundary of box constraints along p_k.

Furthermore, the acceptance rule (2.7) means that, for large k,

$$f(x_k + \frac{\alpha_k}{\omega}p_k) - f_k \ge f(x_k + \frac{\alpha_k}{\omega}p_k) - f(x_{l(k)}) > \beta\frac{\alpha_k}{\omega}g_k^T p_k.$$

Since

$$f(x_k + \frac{\alpha_k}{\omega}p_k) - f_k = \frac{\alpha_k}{\omega}g_k^T p_k + o(\frac{\alpha_k}{\omega}\|p_k\|)$$

and Assumption 1 holds, we have, for large k, $\|p_k\| \le \chi_D\Delta_k$, and

$$0 < (1 - \beta)\frac{\alpha_k}{\omega} g_k^T p_k + o(\frac{\alpha_k}{\omega}\| p_k \|) \leq - \tau\varepsilon^2 (1 - \beta)\frac{\alpha_k}{\omega}\Delta_k + o(\frac{\alpha_k}{\omega}\| p_k \|) < 0, \quad (3.20)$$

and hence the conclusion of the theorem is true. □

Theorem 3.1 indicates that at least one limit point of $\{x_k\}$ is a stationary point. Next, we extend this theorem to a stronger result. The proof is essentially similar to that in [3].

Theorem 3.2 Assume that Assumptions 1 and 2 hold, Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Then

$$\lim_{k\to+\infty} \| D_k^{1/2} g_k \| = 0. \quad (3.21)$$

Since the length of the paper is limited, we will study the local convergent rate and numerical experiments in our further research.

References:

[1] BELLAVIA S, MACCONI M, MORINI B. An affine scaling trust - region approach to bound - constrained nonlinear systems[J]. Applied Numerical Mathematics, 2003, 44: 257 - 280.

[2] KANZOW C, KLUG A. An interior-point affine-scaling trust-region method for semismooth equations with box constraints [J]. Computational Optimization and Applications, 2007, 37(3): 329 - 353.

[3] ZHU D T. An affine scaling trust region algorithm with interior backtracking technique for solving bound-constrained nonlinear systems[J]. Journal of Computational and Applied Mathematics, 2005, 184(2): 343 - 361.

[4] GRIPP L, LAMPARIELLO F, LUCIDI. A nonmonotone line search technique for Newton's methods[J]. SIAM J Numer Anal, 1986, 23: 707 - 716.

[5] COLEMAN T F, LI Y. An interior trust region approach for minimization subject to bounds[J]. SIAM J Optimization, 1996, 6(3): 418 - 445.

有界约束非线性方程组的仿射尺度内点信赖域方法

孙冬梅 ,朱德通

(上海师范大学 数理信息学院 , 上海 200234)

摘要 : 提供了仿射信赖域策略结合非单调线搜索算法解有界约束非线性方程组. 基于简单有界约束的非线性优化问题构建信赖域子问题 ,但所用的最小仿射尺度比 Coleman和 Li所用的仿射尺度更为一般. 在合理的条件下 ,文中提供的最小仿射尺度 ,在没有严格互补假设条件下 ,可给出更强的全局收敛性结果. 引入非单调技术能克服高度非线性的病态问题.

关键词 : 有界约束 ;信赖域 ;仿射尺度 ;非单调线搜索技术

(责任编辑 :冯珍珍)