# On the Regul rity of Trust Region-CG Algorithm for Nonline r Ill-posed Inverse Problems*

Yan-fei Wang Ya-xiang Yuan
State Key Laboratory of Scientific and Engineering Computing,
Institute of Computational Mathematics and Scientific/Engineering computing,
Academy of Mathematics and System Sciences,
Chinese Academy of Sciences, P.O.BOX 2719, Beijing, 100080, China
Email: wyf@lsec.cc.ac.cn yyx@lsec.cc.ac.cn

**Abstract.** In this paper we consider the regularity of the trust region-cg algorithm, when it is applied to nonlinear ill-posed iverse problems. The trust region algorithm can be viewed as a regularization method, but it differs from the traditional regularization method, because no penalty term is need. Thus, the determing of the so-called regularization parameter in a standard regularization method is avoided. Theoretical analysis of the trust region-cg method is presented, convergency and regularity of the trust region algorithm are proved, and numerical tests are also given.
**Key Words.** Nonlinear ill-posed problems, trust region-cg method, convergence, regularity
**AMS Subject Classifications:** 65J15, 65J20, 65K10

## 1 Introduction

In scientific and engineering computing, we are often encountered with nonlinear inverse problems. An inverse problem consists of a direct problem and some unknown function(s) or parameters. Inverse problems are usually ill-posed in the sense of J. Hadmard, i.e., at leat one term of the *existence, uniqueness, stability* of the solution is vilated. Particularly we are concerned with the stability, since in many applications the solution does not depend continuously on the unknown quantities and the problem is ill-posed. A typical ill-posed problems is to determine these unknowns given measured, or condaminated data.

We can outline the nonlinear ill-posed problems into an abstract operator equations

$$F(x) = y, \tag{1}$$

where $F : D(F) \subset X \longrightarrow Y$ is a nonlinear mapping, $X$ and $Y$ are both seperable Hilbert spaces. We assume that $F$ is continuous and compact for fixed $x \in D(F)$.

Problem (1) is typically ill-posed in the sense that a solution $x^+$ does not depend continuously on the obervation data $y$. Since in practice only approximate data with some error level

*Partially supported by Chinese NSF grant 19731010 and the Knowledge Innovation Program of CAS. The first author also partially supported by Hebei Province NSF grant 101001.

$\delta$, i.e.,

$$\|y_\delta - y\| \le \delta \tag{2}$$

are available, problem (1) has to be regularized (see e.g. [3, 7, 23]). Through out this paper we assume that a solution $x^+$ of (1) exists, i.e.

$$F(x^+) = y. \tag{3}$$

Regularization methods are such kind of methods which replace the ill-posed problem with a stabilized problem whose solution depends on a parameter, named as the regularization parameter. The regularized problem is well-posed in the sense of J. Hadamard. For a complete theoretical analysis of such kind of method, please see some well-written books [23, 3, 13, 7, 14, 17].

Certainly the most well-known and most widely used regularization method for nonlinear ill-posed problems is the method of Tikhonov regularization. In which one solves the unconstrained minimization problem

$$\min_{x\in X} J_\alpha[x,y] := \|F(x) - y_\delta\|^2 + \alpha\theta(x). \tag{4}$$

$\alpha > 0$ is the regularization parameter. $\theta(x)$ serves as the stabilizer, i.e., stablizes the minimization process and provides a priori information about the solution.

Replacing $F(x)$ by first order Taylor's expansion, i.e., (4) turns into

$$\min_{\xi\in X} J_\alpha[\xi,y] := \|y_\delta - F(x_k) - F'(x_k)\xi\|^2 + \alpha\theta(\xi). \tag{5}$$

If an approximate solution $\xi_k$ of (5) is computed, we can let $x_{k+1} = x_k + \xi_k$.

Assume that $x_k$ is some approximation of the solution $x^+$, then

$$F(x^+) - F(x_k) = F'(x_k)(x^+ - x_k) + r(x^+; x_k), \tag{6}$$

where $r(x^+; x_k)$ is the Taylor remainder. Denoting $\xi^+ = x^+ - x_k$ and solving for it leads to

$$F'(x_k)\xi^+ = y - F(x_k) - r(x^+; x_k). \tag{7}$$

The above relation can be rewritten as

$$F'(x_k)\xi^+ = y_\delta - F(x_k) + y - y_\delta - r(x^+; x_k). \tag{8}$$

Thus, the linearized problem

$$F'(x_k)\xi = y_\delta - F(x_k) \tag{9}$$

is an approximation to the original problem up to up to an error $err = y_\delta - y + r(x^+; x_k)$ with

$$\|err\| \le \delta + \|r(x^+; x_k)\|. \tag{10}$$

Clearly (5) is equivalent to applying Tikhonov regularization to problem (9).

**Assumption 1.1** *Assume that after $k$ iterations, $\xi^+ = x^+ - x_k$ satisfies*

$$\|y_\delta - F(x_k) - F'(x_k)\xi^+\| \le \omega\|y_\delta - F(x_k)\|, \ \ 0 < \omega < 1.$$

Apart from the above analysis, we need the following assumption:

**Assumption 1.2** *For a ertain ball $\mathcal{B} \subset D(F)$ around the exa t solution $x^+$ of (1), and some $1 > d > 0$ let*

$$\|F(x) - F(\hat{x}) - F'(\hat{x})(x - \hat{x})\| \le d\|F(x) - F(\hat{x})\| \tag{11}$$

*for all $x, \ \hat{x} \in \mathcal{B}$.*

This assumption is helpful for analyzing the properties of the trust region algorithm which we presented in this paper.

Recently, optimization methods are becoming popular for solving nonlinear ill-posed inverse problems, for example, Gauss-Newton method ([1, 12]), Broyden's method ([11]), and Levenberg-Marquardt method ([9]), which have been well developed in nonlinear programming.

Trust region method has been used in parameter identification problem and image restoration problem (see [25, 26]) and seems promising. This paper will consider trust region method for nonlinear ill-posed inverse problems.

## 2 A Trust Region-CG Algorithm

Considering the unconstrained optimization problem

$$\min_{x \in X} J[x, y_\delta] := \|F(x) - y_\delta\|^2. \tag{12}$$

We denote by $g(x)$ the gradient of the functional $J$, $Hess(x)$ the approximate Hessian of $J$, i.e.,

$$g(x) = F'(x)^T(F(x) - y_\delta), \ \ Hess(x) = F'(x)^T F'(x).$$

At the $k$-th iteration, a trust region subproblem (TRS) for (12) is

$$\min_{x \in R^n} g_k^T \xi + \frac{1}{2}(Hess_k \xi, \xi) := \psi_k(\xi), \tag{13}$$

$$s. \ t. \ \|\xi\| \le \Delta, \tag{14}$$

where $g_k = g(x_k), \ Hess_k = Hess(x_k)$ and $\Delta > 0$ is the trust region bound. (13)–(14) is solved exactly or inexactly to obtain a trial step $\xi_k$. The ratio

$$r_k = \frac{Ared_k}{Pred_k} \tag{15}$$

is used to decide whether the trial step $\xi_k$ is acceptable and to adjust the trust region bound.

$$Ared_k = J[x_k, y_\delta] - J[x_k + \xi_k, y_\delta] \tag{16}$$

is called the actual reduction in the objective model, and

$$Pred_k = \psi_k(0) - \psi_k(\xi_k) \tag{17}$$

is the predicted reduction. We outline the general trust region algorithm for unconstrained optimization as follows.

**Algorithm 2.1** *(Trust region algorithm for nonlinear ill-posed problem)*

*STEP 1 Given the initial guess value $x_1 \in R^n$, $\Delta_1 > 0$, $0 < \tau_3 < \tau_4 < 1 < \tau_1$, $0 \le \tau_0 \le \tau_2 < 1$, $\tau_2 > 0$, $k := 1$;*

*STEP 2 If <u>some stopping rule is satisfied</u> then STOP; Else, solve (13)-(14) giving $\xi_k$;*

*STEP 3 Compute $r_k$;*

$$x_{k+1} = \begin{cases} x_k & if \quad r_k \le \tau_0, \\ x_k + \xi_k & otherwise; \end{cases} \tag{18}$$

*Choose $\Delta_{k+1}$ that satisfies*

$$\Delta_{k+1} = \begin{cases} [\tau_3 \|\xi_k\|, \tau_4 \Delta_k] & if \quad r_k < \tau_2, \\ [\Delta_k, \tau_1 \Delta_k] & otherwise; \end{cases} \tag{19}$$

*STEP 4 Evaluate $g_k$ and $Hess_k$; k:=k+1; GOTO STEP 2.*

The constant $\tau_i$ $(i = 0, \cdots, 4)$ can be chosen by users. Typical values are $\tau_0 = 0$, $\tau_1 = 2$, $\tau_2 = \tau_3 = 0.25$, $\tau_4 = 0.5$. For other choices of those constants, please see [4], [5], [15], [19], etc.. The parameter $\tau_0$ is usually zero (see [4], [20]) or a small positive constant (see [2] and [21]). The advantage of using zero $\tau_0$ is that a trial step is accepted whenever the objective function is reduced. Hence it would not throw away a “good point”, which is a desirable property especially when the function evaluations are very expensive.

In STEP 2, the stopping rule is based on some kind of so-called discrepancy principle, i.e., once the inequality

$$\|F(x_k) - y_\delta\| \le \tilde{\omega}\delta, \text{ with } \tilde{\omega} > 1$$

is satisfied, no further iteration is needed.

The following lemma is well known (for example, see [16] and [6]):

**Lemma 2.2** *A ve tor $\xi^* \in R^n$ is a solution of (13)-(14) if and only if there exists $\lambda^* \ge 0$ su h that*

$$(Hess_k + \lambda^* I)\xi^* = -g_k \tag{20}$$

*and that $Hess_k + \lambda^* I$ is positive semi-definite, $\|\xi^*\| \le \Delta$ and*

$$\lambda^*(\Delta - \|\xi^*\|) = 0. \tag{21}$$

It is shown by Powell [20] that trust region algorithms for (12) is convergent if the trust region step satisfies

$$Pred(\xi) \geq c\|g\| \min\{\Delta, \|g\|/\|Hess\|\} \tag{22}$$

and some other conditions on $Hess$ are satisfied. It is easy to see that

$$\psi(0) - \min_{\|\xi\| \leq \Delta, \xi \in span\{g\}} \psi(s) \geq \frac{1}{2} \quad \|g\| \min\{\Delta, \|g\|/\|Hess\|\}. \tag{23}$$

Therefore it is quite common that in practice the trial step at each iteration of a trust region method is computed by solving the TRS (13)-(14) inexactly. One way to compute an inexact solution of (13)-(14) was the truncated conjugate gradient method proposed by Toint [24] and Steihaug [22] and analyzed by Yuan [30].

The conjugate gradient method for (13) generates a sequence as follows:

$$\xi_{l+1} = \xi_l + \alpha_l d_l, \tag{24}$$

$$d_{l+1} = -g_{l+1}^{\psi} + \beta_l d_l, \tag{25}$$

where $g_l^{\psi} = \nabla\psi_k(\xi_l) = Hess_k\, \xi_l + g_k$ with $g_k = g(x_k) = F'(x_k)^T(F(x_k) - y_\delta)$, $Hess_k = Hess(x_k) = F'(x_k)^T F'(x_k)$ and

$$\alpha_l = -{g_l^{\psi}}^T d_l / d_l^T Hess_k\, d_l, \;\; \beta_l = \|g_{l+1}^{\psi}\|^2/\|g_l^{\psi}\|^2, \tag{26}$$

with the initial values $\xi_1 = 0, \;\; d_1 = -g_1^{\psi} = -g_k$.

Toint [24] and Steihaug [22] were the first to use the conjugate gradient method to solve the general trust region subproblem (13)-(14). Even without assuming the positive definiteness of $Hess$, we can continue the conjugate gradient method provided that $d_l^T Hess\, d_l$ is positive. If the iterate $\xi_l + \alpha_l d_l$ computed is in the trust region ball, it can be accepted, and the conjugate gradient iterates can be continued to the next iteration. Whenever $d_l^T Hess\, d_l$ is not positive or $\xi_l + \alpha_l d_l$ is outside the trust region, we can take the longest step along $d_l$ within the trust region and terminate the calculations.

**Algorithm 2.3** *(Trun ated onjugate gradient method for T S)*

*STEP 1* *Given* $\xi_1 = 0$, $0 < \tau < 1$, $\epsilon$ *(tolerance)* $> 0$ *and ompute* $g_1^{\psi} = \nabla\psi(\xi_1)$, *set* $l := 1$, $d_1 = -g_1^{\psi} = -g_k$;

*STEP 2* *If* $\|A_k \xi_l - \tilde{u}_k\| \leq \tau \|\tilde{u}_k\|$, *stop, output* $\xi^* = \xi_l$;

*Compute* $d_l^T Hess\, d_l$: *if* $d_l^T Hess_k\, d_l \leq 0$ *then goto step 4;*

*Cal ulate* $\alpha_l$ *by (26).*

*STEP 3* *If* $\|\xi_l + \alpha_l d_l\| \geq \Delta_k$ *then goto step 4;*

*Set* $\xi_{l+1}$ *by (24) and* $g_{l+1}^{\psi} = g_l^{\psi} + \alpha_l Hess_k\, d_l$;

*Compute* $\beta_l$ *by (26) and set* $d_{l+1}$ *by (25);*

*l:=l+1, goto step 2.*

*STEP 4* *Compute* $\alpha_l^* \geq 0$ *satisfying* $\|\xi_l + \alpha_l^* d_l\| = \Delta$;

*Set* $\xi^* = \xi_l + \alpha_l^* d_l$, *and stop.*

$$\begin{aligned} d_{l+1} &= -g_{l+1}^{\psi} + \beta_l d_l = -A_k^*(A_k \xi_{l+1} - \tilde{u}_k) + \beta_l d_l \\ &= -A_k^* A_k \xi_l - \alpha_l A_k^* A_k d_l + A_k^* \tilde{u}_k + \beta_l d_l. \end{aligned}$$

If we let $d_l = A_k^* z_l$ provided that such $z_l$ exists, then

$$d_{l+1} = A_k^*(\tilde{u}_k - A_k \xi_l - \alpha_l A_k d_l + \beta_l z_l).$$

Further if we denote $r_l = \tilde{u}_k - A_k \xi_l$, then clearly

$$r_{l+1} = \tilde{u}_k - A_k \xi_{l+1} = r_l - \alpha_l A_k d_l$$

and

$$d_{l+1} = A_k^*(r_l - \alpha_l A_k d_l + \beta_l z_l)$$

hold. We can generate the next search direction by $d_{l+1} = A_k^* z_{l+1}$ with $z_{l+1} = r_l - \alpha_l A_k d_l + \beta_l z_l = r_{l+1} + \beta_l z_l$. Hence, the conjugate gradient iterates can be genereated in the following way:

$$\xi_{l+1} = \xi_l + \alpha_l d_l \tag{30}$$

$$d_l = A_k^* z_l \tag{31}$$

$$z_{l+1} = r_{l+1} + \beta_l z_l \tag{32}$$

$$r_{l+1} = r_l - \alpha_l A_k d_l \tag{33}$$

$$\alpha_l = -\frac{{g_l^{\psi}}^T d_l}{d_l^T Hess\, d_l} = \frac{\|A_k^* r_l\|^2}{\|A_k d_l\|^2} \tag{34}$$

$$\beta_l = \frac{\|A_k^* r_{l+1}\|^2}{\|A_k^* r_l\|^2} \tag{35}$$

$$\sigma_{l+1} = 1 + \beta_l \sigma_l \tag{36}$$

The initial values are $\xi_1 = 0,\ d_1 = -g_k,\ z_1 = r_1,\ r_1 = \tilde{u}_k = y_\delta - F(x_k),\ \sigma_1 = 1$. Here another scalar $\sigma_l$ is added, which will be used for the analysis of the truncated conjugate gradient method.

One tool for the analysis of the truncated conjugate gradient method is the so-called residual polynomials (see [10, 3]). Let $\Pi_l$ be the set of all polynomials of degree $l$ or less, and set

$$\Pi_l^0 := \{p \in \Pi_l : p(0) = 1\}.$$

Then there is an 1-1 relation between elements $\xi \in \mathcal{K}_l(A_k^* \tilde{u}_k; A_k^* A_k)$ and $p \in \Pi_l^0$ via the representation

$$\tilde{u}_k - A_k \xi = p(A_k A_k^*) \tilde{u}_k \tag{37}$$

of the corresponding residual, where $\mathcal{K}_l(A_k^* \tilde{u}_k; A_k^* A_k)$ is the $l$-th Krylov subspace

$$\mathcal{K}_l(A_k^* \tilde{u}_k; A_k^* A_k) = \mathrm{span}\{A_k^* \tilde{u}_k, (A_k^* A_k) A_k^* \tilde{u}_k, \cdots, (A_k^* A_k)^{l-1} A_k^* \tilde{u}_k\}.$$

For simplicity, $p_l \in \Pi_l^0$ denotes the residual polynomial associated with $\xi_l$, the $l$-th CG iterate.

The bilinear form

$$< p, q > := (p(A_k A_k^*)\tilde{u}_k, q(A_k A_k^*)\tilde{u}_k)$$

defines the inner product for $p,\ q \in \Pi_l$. If $q \in \Pi_{l-1}$ is an arbitrary polynomial of degree $l-1$ then the polynomial $p$ given by $p(\lambda) = p_l(\lambda) + t\lambda q(\lambda)$ belongs to $\Pi_l^0$ for every $t \in R$. Noticing that $p_l$ solves the minimization problem

$$< p, p > \longrightarrow \min \ \text{ for } p \in \Pi_l^0.$$

Hence,

$$< p_l, \lambda q > = \frac{1}{2}\frac{d}{dt} < p, p > |_{t=0} = 0, \text{ for all } q \in \Pi_{l-1}. \tag{38}$$

If we define $q = q_{l-1}$ by $p_l = 1 - \lambda q_{l-1}$, then clearly we have that,

$$< p_l, 1 > = < p_l, p_l >, \tag{39}$$

which will be used for later analysis.

From (30)-(36) and the above definitions, we have that

$$z_l = s_l(A_k A_k^*)\tilde{u}_k,\ s_l(\lambda) := \frac{p_l(\lambda) - p_{l+1}(\lambda)}{\alpha_l \lambda} \in \Pi_l. \tag{40}$$

It was pointed out by [8] that in general $s_l$ does not belong to $\Pi_l^0$. Instead, since the vectors $z_l$ are updated by $z_{l+1} = r_{l+1} + \beta_l z_l$ with $r_{l+1} = \tilde{u}_k - A_k\xi_{l+1}$, it follows from (40) that

$$s_{l+1}(\lambda) = p_{l+1}(\lambda) + \beta_l s_l(\lambda), \tag{41}$$

and hence, $s_l(0)$ and $\sigma_l$ of Algorithm 2.3 share the same recurrence relation (this in fact has been observed by [8]), i.e.,

$$s_l(0) = \sigma_l. \tag{42}$$

With the above analysis, we can now present the monotonicity of the iterates for perturbed right-hand side.

**Theorem 3.1** *et $\gamma \geq 2,\ l^* \in N$. If Assumption 1.1 holds and if*

$$\|\tilde{u}_k - A_k\xi_l\|^2 + \|\tilde{u}_k - A_k\xi_{l+1}\|^2 > \omega\gamma\frac{\|z_l\|\|\tilde{u}_k\|}{\sigma_l},\ 0 < \omega < 1,\ l = 1, 2, \cdots, l^*, \tag{43}$$

*then $\|\xi^+ - \xi_l\|$ is stri tly monotoni ally de reasing for $l = 1, 2, \cdots, l^*$, and*

$$\|\xi^+\|^2 - \|\xi^+ - \xi_{l^*+1}\|^2 > (\gamma - 2)\omega\|\tilde{u}_k\|\sum_{l=1}^{l^*}\alpha_l\|z_l\|. \tag{44}$$

*Proof.* By induction, we obtain

$$\begin{aligned} \|\xi^+ - \xi_{l+1}\|^2 &= \|\xi^+ - \xi_l - \alpha_l d_l\|^2 \\ &= \|\xi^+ - \xi_l\|^2 - 2(\xi^+ - \xi_l, \alpha_l A_k^* z_l) + (\alpha_l A_k^* z_l, \alpha_l A_k^* z_l) \\ &= \|\xi^+ - \xi_l\|^2 - \alpha_l(2A_k\xi^+ - 2A_k\xi_l - \alpha_l A_k d_l, z_l) \\ &= \|\xi^+ - \xi_l\|^2 - \alpha_l(\tilde{u}_k - A_k\xi_l, z_l) - \alpha_l(\tilde{u}_k - A_k\xi_{l+1}, z_l) \\ &\quad +2\alpha_l(\tilde{u}_k - A_k\xi^+, z_l). \end{aligned}$$

From the definitions of $p_l$ and $z_l$, we have

$$\|\xi^+ - \xi_l\|^2 - \|\xi^+ - \xi_{l+1}\|^2 = \alpha_l < p_l, s_l > +\alpha_l < p_{l+1}, s_l > -2\alpha_l < \tilde{u}_k - A_k\xi^+, z_l > .$$

By (42), $s_l(\lambda) = \sigma_l + \lambda q(\lambda)$ for some polynomial $q \in \Pi_{l-1}$, and hence from (38) and (39) we find that

$$\begin{aligned} &\|\xi^+ - \xi_l\|^2 - \|\xi^+ - \xi_{l+1}\|^2 = \alpha_l\sigma_l(< p_l, 1 > + < p_{l+1}, 1 >) - 2\alpha_l < \tilde{u}_k - A_k\xi^+, z_l > \\ &= \alpha_l\sigma_l(< p_l, p_l > + < p_{l+1}, p_{l+1} >) - 2\alpha_l < \tilde{u}_k - A_k\xi^+, z_l > . \end{aligned}$$

Since $< p_l, p_l >= \|\tilde{u}_k - A_k\xi_l\|^2$, hence it follows from the above relation and (43) that

$$\|\xi^+ - \xi_l\|^2 - \|\xi^+ - \xi_{l+1}\|^2 > \omega\gamma\alpha_l\|z_l\|\|\tilde{u}_k\| - 2\omega\alpha_l\|\tilde{u}_k\|\|z_l\|, \text{ for all } l = 1, \cdots, l^*. \qquad (45)$$

Thus, the monotonicity of $\|\xi^+ - \xi_l\|$ follows from the above inequality and the assumption $\gamma > 2$. Relation (44) follows by taking the sum of (45) for $l = 1, \cdots, l^*$ and observing the fact that $\xi_1 = 0$. Q.E.D

**Remark 3.2** *We have noted that in general $s_l$ will not belong to $\Pi_l^0$, however, $\hat{s}_l := s_l/\sigma_l \in \Pi_l^0$. Hen e from the minimization property of the trun ated CG we obtain*

$$\|\tilde{u}_k - A_k\xi_{l+1}\| \le \|\tilde{u}_k - A_k\xi_l\| =< p_l, p_l >^{\frac{1}{2}} \le < \hat{s}_l, \hat{s}_l >^{\frac{1}{2}} = \frac{1}{\sigma_l} < s_l, s_l >^{\frac{1}{2}} = \frac{1}{\sigma_l}\|z_l\|. \qquad (46)$$

*This together with (43) yield that*

$$\begin{aligned} 2\|\tilde{u}_k - A_k\xi_{l^*}\|^2 &\ge \|\tilde{u}_k - A_k\xi_{l^*}\|^2 + \|\tilde{u}_k - A_k\xi_{l^*+1}\|^2 \\ &> \omega\gamma\frac{\|z_{l^*}\|\|\tilde{u}_k\|}{\sigma_{l^*}} > \omega\gamma\|\tilde{u}_k\|\|\tilde{u}_k - A_k\xi_{l^*}\|, \end{aligned}$$

*whi h indi ates that*

$$\|\tilde{u}_k - A_k\xi_{l^*}\| \ge \frac{\omega\gamma\|\tilde{u}_k\|}{2} > \omega\|\tilde{u}_k\|.$$

*Sin e $\|\tilde{u}_k - A_k\xi^+\| \le \omega\|\tilde{u}_k\|$ a ording to Assumption 1.1, this shows that $l^*$ an not ex eed $l_k$, the smallest index of the inner iteration.*

# 4 Convergence of Trust Region-CG for Exact Data

Before presenting the proposition in the following paragraph, we first give an assumption:

**Assumption 4.1** *Assume that in ea h inner iteration, $\xi_l$ satisfies*

$$\|\tilde{u}_k - A_k\xi_l\| \geq \omega_1\|\tilde{u}_k\|, \text{ with } \omega_1^2 = \gamma\omega \tag{47}$$

*until onvergen e, where* $0 < \omega < \gamma^{-1}$, $\gamma > 2$.

Assumption 4.1 is closely related with the termination rule of Algorithm 2.3. Where, $\tau$ in STEP 2 serves as the number $\omega_1$ here. Once the opposite inequality of Assumption 4.1 is satisfied, we stop the inner iteration.

**Proposition 4.2** *Suppose that Assumption 4.1 holds. The inequality (47) indi ates that (43) is true if $l > 0$. Furthermore, there are only finitely many $l$ for whi h (47) holds.*

*Proof.* From the Remark 3.2 we know

$$\frac{1}{\sigma_l}\|z_l\| =< \hat{s}_l, \hat{s}_l >^{\frac{1}{2}} .$$

It is proved by [10] that $< \hat{s}_l, \hat{s}_l >$ is strictly monotonically decreasing with $l$, consequently

$$\frac{1}{\sigma_l}\|z_l\| < \frac{1}{\sigma_1}\|z_1\| = \|r_1\| = \|\tilde{u}_k\|.$$

The above inequality together and (47) indicates that (43) is true for $l > 0$. From Remark 3.2, we see that (47) holds for only finitely many indices $l$. Q.E.D

Now assume that $y_\delta = y$, we first prove the monotonicity of the trust region-cg algorithm, i.e., $x_k + \xi_{l_k}$ is a better approximation of $x^+$ than $x_k$. We also assumed that $F(x) = y$ has a solution $x^+ \in \mathcal{B} \subset D(F)$.

**Proposition 4.3** *Assume that Assumptions 1.2 and 4.1 holds, then the iteration error $\|x^+ - x_k\|$ is monotoni ally de reasing.*

*Proof.* According to Assumption 1.2, (11) holds for $x = x^+$, $\hat{x} = x_k$, i.e.

$$\|y - F(x_k) - F'(x_k)(x^+ - x_k)\| \leq d\|y - F(x_k)\|.$$

Note that $y_\delta = y$, the above expression indicates that Assumption 1.1 is fulfilled with $0 < d < 1$.

Due to Assumption 4.1,

$$\|\tilde{u}_k - A_k\xi_l\|^2 \geq \gamma\omega\|\tilde{u}_k\|^2$$

is satisfied. Hence from Proposition 4.2, the requirement of Theorem 3.1 is fulfilled.

From our notations, we have that $x_{k+1} = x_k + \xi_{l_k}$. Thus, from $\xi^+ = x^+ - x_k$, $\xi^+ - \xi_{l_k} = x^+ - x_{k+1}$ and Theorem 3.1, we see that $\|x^+ - x_{k+1}\| < \|x^+ - x_k\|$. Q.E.D

**Remark 4.4** *Propostion 4.3 also implies two inequalities:*

$$\|u_k\|\|w_k\| < \frac{1}{d(\gamma-2)}(\|x^+ - x_k\|^2 - \|x^+ - x_{k+1}\|^2) \tag{48}$$

*and*

$$\|u_k\|^2 < \frac{\|A_k\|^2}{d(\gamma-2)}(\|x^+ - x_k\|^2 - \|x^+ - x_{k+1}\|^2). \tag{49}$$

(48) is straightforward as $\|u_k - A_k\xi^+\| \leq d\|u_k\|$ with $0 < d < 1$. (49) follows from relations

$$\|u_k\|\|w_k\| > \alpha_1\|u_k\|\|z_1\| = \alpha_1\|u_k\|^2$$

and

$$\alpha_1 = \frac{\|A_k^* r_1\|^2}{\|A_k d_1\|^2} = \frac{\|A_k^* u_k\|^2}{\|A_k A_k^* u_k\|^2} \geq \|A_k\|^{-2}.$$

**Theorem 4.5** *Given the exa t data $y_\delta = y$ and suppose that Assumptions 1.2 and 4.1 hold. Then the iterates $\{x_k\}$ generated by Algorithms 2.1 and 2.3 onverge to a solution of (1) as $k \to \infty$.*

*Proof.* First we prove that $\{x_k\}$

Relation (48) and the above expression give that

$$\begin{aligned} |(e_\nu, e_\nu - e_j)| &\leq (1+3d)\sum_{i=j}^{\nu-1}\|w_i\|\|y - F(x_i)\| \\ &\leq \frac{1+3d}{d(\gamma-2)}(\|x^+ - x_j\|^2 - \|x^+ - x_\nu\|^2), \end{aligned}$$

which, together with (50), yields

$$\|e_\nu - e_j\|^2 \leq C(\|x^+ - x_j\|^2 - \|x^+ - x_\nu\|^2)$$

with $C = \frac{2(1+3d)}{d(\gamma-2)} + 1$ independent of $\nu$, $j$, $k$. Similarly one can obtain

$$\|e_j - e_\nu\|^2 \leq C(\|x^+ - x_\nu\|^2 - \|x^+ - x_j\|^2),$$

hence

$$\begin{aligned} \|x_k - x_j\|^2 &= \|e_j - e_k\|^2 \leq 2(\|e_k - e_\nu\|^2 + \|e_\nu - e_j\|^2) \\ &\leq 2C(\|x^+ - x_j\|^2 - \|x^+ - x_k\|^2). \end{aligned} \quad (51)$$

Therefore, $\{x_k\}$ form a Cauchy sequence because the monotonicity of $\{\|x^+ - x_k\|\}$.

Denote the limit of $x_k$ by $x$. From (49) we know $\sum_{k=1}^{\infty}\|u_k\|^2$ converges, and therefore $F(x_k) \to y$ as $k \to \infty$. This indicates that $x$ is a solution of (1). Q.E.D

## 5 Regularity of the Algorithm for Inexact Data

Now we consider the case where inexact date $y_\delta$ instead of $y$. It is assumed that (2) is satisfied.

Our stopping rule is based on the discrepancy principle, i.e., we terminate the calculations at the smallest iteration index $k_D$ such that the discrepancy inequality

$$\|y_\delta - F(x_k^\delta)\| \leq \tilde{\omega}\delta, \text{ with } \tilde{\omega} > 1 \quad (52)$$

holds.

We denote $x_k^\delta$ the corresponding iterates and consider the regularity of the trust region-cg algorithm.

**Theorem 5.1** *Assume that Assumptions 1.2 and 4.1 hold. et $x$ be a solution of (1) with $F$ satisfies (11) for some $1 > d > 0$ in a ball $\mathcal{B} \subset D(F)$ around $x$. et $\tilde{\omega}$ in (52) be hosen that $\tilde{\omega} > \frac{1+d}{1-d}$. Then $\|x - x_k^\delta\|$ is monotoni ally de reasing. Moreover, Algorithm 2.1 terminates after $k_D < \infty$ iterations.*

*Proof.* We prove that

$$\|x - x_{k+1}^\delta\| \leq \|x - x_k^\delta\| \quad (53)$$

with $x$ a solution of (1).

Using Assumption 1.2, we estimate that

$$\begin{aligned}\|y_\delta - F(x_k^\delta - F'(x_k^\delta)(x - x_k^\delta)\| &\le \delta + \|F(x) - F(x_k^\delta - F'(x_k^\delta)(x - x_k^\delta)\| \\ &\le \delta + d\|y - F(x_k^\delta)\| \\ &\le (1+d)\delta + d\|y_\delta - F(x_k^\delta)\|.\end{aligned}$$

According to the discrepancy principle, $\|y_\delta - F(x_k^\delta)\| > \tilde{\omega}\delta$ as $k < k_D$, hence

$$\delta < \frac{1}{\tilde{\omega}}\|y_\delta - F(x_k^\delta)\|$$

and

$$\|y_\delta - F(x_k^\delta - F'(x_k^\delta)(x - x_k^\delta)\| \le \frac{1+d+\tilde{\omega}}{\tilde{\omega}}\|y_\delta - F(x_k^\delta)\|.$$

By assumption, $0 < (1+d+\tilde{\omega})/\tilde{\omega} < 1$, hence Assumption 1.1 is fulfilled. Consequently Propostion 4.3 applies and the monotonicity assertion (53) follows as in the proof of Proposition 4.3.

Next we show that there are only finite number of iterations. In fact as the same as in the proof of (49), we have

$$\|y_\delta - F(x_k^\delta)\|^2 < \frac{L}{d(\gamma-2)}(\|x - x_k^\delta\|^2 - \|x - x_{k+1}^\delta\|^2) \tag{54}$$

with $L = \sup\{\|F^{'}(x_k^\delta)\|^2\}$ for all $k < k_D$.

Assume that (53) holds for $x = x^+$. Now taking the sum of (54) for $k = 1, 2, \cdots, k_D - 1$ we obtain

$$(k_D - 1)\tilde{\omega}^2\delta^2 \le \sum_{k=1}^{k_D-1}\|y_\delta - F(x_k^\delta)\|^2 \le \frac{L}{d(\gamma-2)}\|x^+ - x_1\|^2 < \infty.$$

This indicates that $k_D$ is a finite number. Q.E.D

**Theorem 5.2** *Assume that $F(x_k^\delta) \to F(x_k)$ as $\delta \to 0$. If $k \le k_D$ for all $\delta$ suffi iently small, then $x_k^\delta \to x_k$ for $k \le k_D$ as $\delta \to 0$.*

*Proof.* Given sufficiently small number $\epsilon$, we want to prove $\|x_k^\delta - x_k\| \le \epsilon$ as $\delta \to 0$ for $k \le k_D$. We proceed by induction.

Assume that $x_k^\delta \to x_k$ as $\delta \to 0$, and that $k + 1 \le k_D$. Note that

$$x_{k+1} = x_k + \xi_{l_k},\;\; x_{k+1}^\delta = x_k^\delta + \xi_{l_k}^\delta,$$

$$\xi_{l_k} = F(x_k)^* \sum_{i=1}^{l_k-1}\alpha_i z_i,\;\; \xi_{l_k}^\delta = F(x_k^\delta)^* \sum_{i=1}^{l_k-1}\alpha_i^\delta z_i^\delta,$$

we can estimate that

$$\begin{aligned}\|x_{k+1}^\delta - x_{k+1}\| &\le \|x_k^\delta - x_k\| + \|\xi_{l_k}^\delta - \xi_{l_k}\| \\ &\le \|x_k^\delta - x_k\| + \|F(x_k^\delta)^* \sum_{i=1}^{l_k-1} (\alpha_i^\delta z_i^\delta - \alpha_i z_i)\| \\ &\quad + \|(F(x_k^\delta)^* - F(x_k)^*) \sum_{i=1}^{l_k-1} \alpha_i z_i\| \\ &\le \|x_k^\delta - x_k\| + \|F(x_k^\delta)\|(l_k - 1) \max_i \{\|\alpha_i^\delta z_i^\delta - \alpha_i z_i\|\} \\ &\quad + \|F(x_k^\delta) - F(x_k)\| \|\sum_{i=1}^{l_k-1} \alpha_i z_i\| \end{aligned} \tag{55}$$

By the induction assumption $x_k^\delta \to x_k$, we have that

$$F(x_k^\delta) \to F(x_k).$$

Therefore, it follows that $\alpha_i^\delta \to \alpha_i,\ z_i^\delta \to z_i$. Consequently, it from (55) that $x_{k+1}^\delta \to x_{k+1}$. Q.E.D

**Theorem 5.3** *Assume that $F$ satisfies (11) in some ball $\mathcal{B} \subset D(F)$ and let $y_\delta,\ x_\delta$ as before. Then the iterates $x_k^\delta$ generated by Algorithms 2.1 and 2.3 onverge to a solution of (1) as $k \to \infty$ and $\delta \to 0$.*

*Proof.* For simplicity, We use $k(\delta)$ instead of $k_D$ in the following analysis.

From Theorem 4.5 we know that iterates $x_k$ converge to a solution of (1). Combining this fact with Theorem 5.2, we find that the iterates $x_k^\delta$ converge to a solution of (1) for $k \le k(\delta)$ as $\delta \to 0$.

Now assume that $k(\delta) \to \infty$ as $\delta \to 0$, and denote $x^+$ the limit of the iterates $x_k$, $x^+$ is a solution of (1). It suffieces to consider subsequences $\{k(\delta_n)\}_n$ which are monotinically increasing to infinity as $n \to \infty$ and $\delta_n \to 0$. Without loss of generality, let us consider $k(\delta_m) > k(\delta_n)$ for $m > n$. By the monotonicity of $x_k^\delta$, i.e., Theorem 5.1, we have

$$\|x_{k(\delta_m)}^{\delta_m} - x^+\| \le \|x_{k(\delta_n)}^{\delta_m} - x^+\| \le \|x_{k(\delta(n)}^{\delta(m)} - x_k(\delta_n)\| + \|x_{k(\delta_n)} - x^+\|.$$

Given a sufficiently small number $\epsilon > 0$ and for some sufficiently large number $n$, $\|x_{k(\delta_n)} - x^+\| \le \epsilon/2$ by Theorem 4.5. On the other hand, for sufficiently large number $m$ and fixed $n$, $\|x_{k(\delta(n)}^{\delta(m)} - x_k(\delta_n)\| \le \epsilon/2$ by Theorem 5.1. This proves that $\|x_{k(\delta_m)}^{\delta_m} - x^+\| \le \epsilon$ for all $m$ sufficiently large, and thereafter $x_{k(\delta_m)}^{\delta_m} \to x^+$ as $m \to \infty$. Hence, we see that $x_k^\delta \to x^+$ as $k \to \infty$ and $\delta \to 0$. Q.E.D.

## 6 Numerical Test

In this section, we give an example to test our algorithm. The example is the inverse Gravimetry problem (see [23]). We write it as

$$\begin{aligned} & F(x) : X \longrightarrow Y \\ F(x)(t) &= \int_a^b k(t,s,x(s))ds = y(t),\ t \in [c,d]. \end{aligned} \tag{56}$$

with $k(t,s,x(s)) = \ln \frac{(t-s)^2+H^2}{(t-s)^2+(x(s)-H)^2}$. Clearly the kernel $k$ is defined on the set $\Pi = \{[c,d] \times [a,b] \times R\}$ and $k(t,s,x(s)) \in C^1(\Pi)$. The first derivative $F'(x) : X \longrightarrow Y$ is defined by

$$[F^{'}(x)u](t) = \int_a^b \frac{\partial k}{\partial x}(t,s,x(s))u(s)ds,\ t \in [c,d], \tag{57}$$

where the kernel $\frac{\partial k}{\partial x}(t,s,x(s))$ can be evaluated by

$$\frac{\partial k}{\partial x}(t,s,x(s)) = \frac{2(H-x(s))}{(t-s)^2+(x(s)-H)^2}.$$

$F^{'}(x)$ is compact, since the kernel is square integrable.

Now, we will set up the problem of approximate determination of normal pseudosolution to the equation (56).

For simplicity, two equidistant grids on intervals $[a,b]$ and $[c,d]$ are applied:

$$\Sigma_n(s) = \{s_j : s_j = a + h_s(j-1), j = 1,2,\cdots,n\},\ h_s = \frac{b-a}{n-1},$$
$$\Sigma_m(t) = \{t_i : t_i = c + h_t(i-1), i = 1,2,\cdots,n\},\ h_t = \frac{d-c}{m-1}.$$

In this way, the spaces of all grid functions defined on $\Sigma_n(s)$ and $\Sigma_m(t)$, respectively, are treated as $X_n$ and $Y_m$.

The integral operator $F$ gives rise to an operator $F_{mn} : X_n \longrightarrow Y_m$ by

$$[F_{mn}(x)]_i = \int_a^b k(t_i,s,x(s))ds,\ 1 \le i \le m.$$

Similarly the derivative operator $F^{'}(x)$ yields an $m \times n$ matrix:

$$[F^{'}_{mn}(x)]_{ij} = \int_a^b \frac{\partial k}{\partial x}(t_i,s,x(s))\phi_j(s)ds,\ 1 \le i \le m,\ 1 \le j \le n.$$

Where $\phi_j(s)$ we used is the standard linear basic functions

$$\phi_j(s) = \begin{cases} \frac{s-s_{j-1}}{h},\ if\ s \in [s_{j-1},s_j], \\ \frac{s_{j+1}-s}{h},\ if\ s \in [s_j,s_{j+1}], \\ 0,\ else. \end{cases}$$

In which, $s_j = jh,\ h = \frac{1}{n},\ j = 1,2,\cdots,n.$ The integral can be computed numerically.

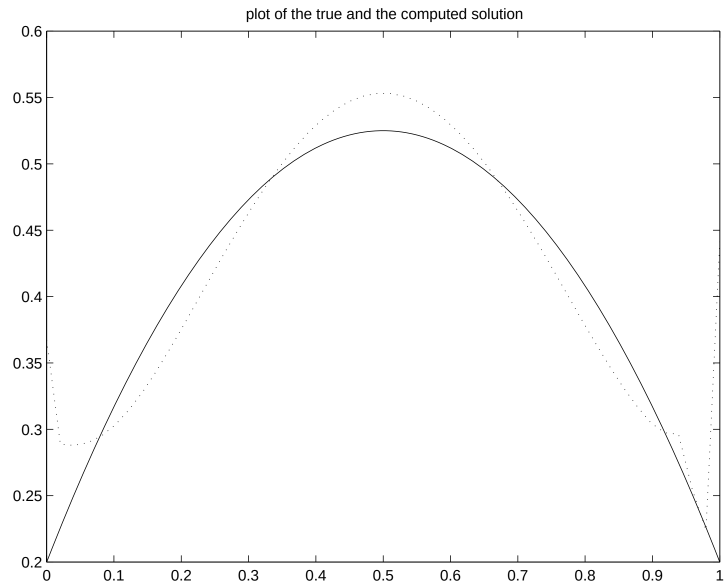


Figure 2: Solution of the inverse problem: nonlinear Fredholm equation

*observation is a poor approximation to the original problem, it will also not enough to give a reasonable results.*

# 7 Conclusion

We have establised the convergence and regularity of the trust region-cg method for nonlinear ill-posed inverse problems. It deserved pointing out that Plato in [18] had establised the regularity property of the conjugate gradient method. Later on, Hanke in [8] had established the regularity of Newon-CG method. All of the methods are stable for solving ill-posed inverse problems.

# References

[1] Bakushinsky A and Goncharsky A 1994, *Ill-Posed P oblems: Theo y and Applications*, Kluwer Academic Publishers.

[2] Duff I S, Nocedal J 1987, Reid J K et al., The use of linear programming for the solution of sparse sets of nonlinear equations. *SIAM J Sci Stat Comput*, **8**, 99-108.

[3] Engl H W, Hanke M and Neubauer 1996, *Regula ization of Inve se P oblems*, Dordrecht: Kluwer.

[4] Fletcher R 1987, *P actical Methods of Optimization*, 2nd ed., Chichester: John Wiley and Sons.

[5] Fletcher R 1982, A model algorithm for composite NDO problem, *Math P og Study*, **17**, 67-76.

[6] Gay D M 1981, Computing optimal local constrained step, *SIAM J Sci Stat Comp*, **2**, 186-197.

[7] Groetsch C W 1984, *The Theo y of Tikhonov Regula ization fo F edholm Equations of The Fi st Kind* (Boston,MA:Pitman)

[8] Hanke M, Regularizing Properties of a Truncated Newton-CG Algorithm for Nonlinear Inverse Problems, *submitted*.

[9] Hanke M 1997, A Regularizing Levenberg-Marquardt Scheme, with Applications to Inverse Groundwater Filteration Problems, *Inve se P oblems* **13**, 79-95.

[10] Hanke M 1995, *Conjugate G adient Type Methods fo Ill-posed P oblems*, Longman Scientific and Technical, Harlow, Essex.

[11] Kaltenbacher B 1998, On Broyden's Method for Nonlinear Ill-posed Problems, *Nume Funct Anal Opt,* **19**.

[12] Kaltenbacher B 1998, A Posteriori Parameter Choice Strategies for Some Newton Type Methods for the Regularization of Nonlinear Ill-posed Problems, *Nume .Math.* **79**.

[13] Kirsch A 1996, *An Int oduction to the Mathematical Theo y of Inve se P oblems*, New York: Springer-Verlag.

[14] Louis A K 1992, *Inve se und Schlecht Gestellte P oblems*, Teubner, Stuttgart.

[15] Moré J J 1983, Recent developments in algorithms and software for trust region methods. In: Bachem A, Grötschel, Korte B, eds. Mathematical Programming: The State of the Art. Berlin: Springer-Verlag, 258-287.

[16] Moré J J and Sorensen D C 1983, Computing a Trust Region Step, *SIAM J.Sci.Stat.Comput.* **4**, 553-572.

[17] Morozov V A 1984 *Methods fo Solving Inco ectly Posed P oblems* (New York: Springer)

[18] Plato R 1990, Optimal Algorithms for Linear Ill-posed Problems Yield Regularization Methods, *Nume .Funct.Anal.and Optim.* **11**, 111-118.

[19] Powell M J D 1984, Nonconvex minimization calculations and the conjugate gradient method. In: Griffiths, ed. Lecture Notes in Mathematics 1066: Numerical Analysis. Berlin: Springer-Verlag, 122-141.

[20] Powell M J D 1975, Convergence properties of a class of minimization algorithms. In: Mangasarian O L, Meyer R R, Robinson S M, eds. Nonlinear Programming. New York: Academic Press, **2**, 1-27.

[21] Sorensen D C 1982, Newton's Method with a Model Trust Region Modification, *SIAM J.Nume .Anal.* **19**, 409-426.

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

[23] Tikhonov A N and Arsenin V Y 1977, *Solutions of Ill-Posed P oblems*, New York: Wiley.

[24] Toint Ph L 1981, Towards an efficient sparsity exploiting Newton method for minimization. In: Duff I ed. Sparse Matrices and Their Uses. Berlin: Academic Press, 57-88.

[25] Yan-fei Wang and Ya-xiang Yuan 2001, A Trust Region Method for Solving Distributed Parameter Identification Problems, *Repo t, AMSS, Chinese Academy of Sciences, 2001.*

[26] Yan-fei Wang, Ya-xiang Yuan and Hong-chao Zhang 2001, A Trust Region-CG Algorithm for Deblurring Problem in Atmospheric Image Reconstruction, to appear in *Science in China*.

[27] Y. Yuan 1994, Nonlinear Programming: Trust Region Algorithms, in: S.T. Xiao and F. Wu, eds., Proceedings of Chinese SIAM annual meeting (Tsinghua University, Beijing) pp. 83 –97.

[28] Y. Yuan 1998, Matrix Computation Problems in Trust Region Algorithms for Optimization, in: Q.C. Zeng, T.Q. Li, Z.S. Xue and Q.S. Cheng, eds. Proceedings of the 5th CSIAM annual meeting, (Tsinghua Unviersity Press, Beijing) pp. 52–64.

[29] Y. Yuan 1999, Problems on Convergence of Unconstrained Optimization Algorithms, in Y. Yuan ed., Numerical Linear Algebra and Optimization, (Science Press, Beijing, New York), pp. 95-107.

[30] Y. Yuan 2000, On the truncated conjugate gradient method. Math Prog, **87**, 561-571.