JOTA manuscript No.
(will be inserted by the editor)

On an Elliptical Trust-Region Procedure for Ill-Posed Nonlinear Least-Squares Problems

Stefania Bellavia and Elisa Riccietti

Received: date / Accepted: date

Abstract In this paper, we address the stable numerical solution of ill-posed nonlinear least-squares problems with small residual. We propose an elliptical trust-region reformulation of a Levenberg-Marquardt procedure. Thanks to an appropriate choice of the trust-region radius, the proposed procedure guarantees an automatic choice of the free regularization parameters that, together with a suitable stopping criterion, ensures regularizing properties to the method. Specifically, the proposed procedure generates a sequence that even in case of noisy data has the potential to approach a solution of the unperturbed problem. The case of constrained problems is considered, too. The effectiveness of the procedure is shown on several examples of ill-posed least-squares problems.

Keywords Ill-posed Nonlinear Least-Squares Problems · Regularization · Nonlinear Stepsize Control · Levenberg-Marquardt Methods · Trust-Region Methods

1 Introduction

In this paper, we consider Levenberg-Marquardt methods [1,2] for the stable solution of nonlinear ill-posed least-squares problems in Hilbert spaces with noisy data. These methods are also known in this context as non-stationary iterated Tikhonov procedures [3,4]. They depend on the choice of a free regularization parameter λ_k and of a regularizing operator M_k. The regularization parameter's choice is crucial, it is difficult to make, and a bad choice could lead to poor results. Then, it is desirable to obtain a parameter-free method, i.e. a method in which λ_k is automatically set. Many strategies have been proposed in the literature, cf. [3,5–8]. All these methods anyway consider nonlinear operator equations, i.e. least-squares problems in which it is assumed that it exists a solution $x^\dagger$ that attains zero residual. The case in which such a solution does not exist, and the residual at the solution is strictly positive, is considered. These problems arise in many applications. Usually indeed, jointly to observation errors also modelling errors are present, so that it is not realistic to assume that the data are attainable, and one must admit the case $y \notin \mathcal{R}(F)$ [9,10], where $\mathcal{R}(F)$ denotes the range of F. This is the case when a mathematical model approximating a true distribution is fit to given data or in parameter estimation, experimental design or imaging problems [9–16]. These problems are indeed usually formulated as least-squares problems. A possibility is to consider the modelling errors as part of the noise in the data. The same algorithms as for zero-residual problems can be used, with a proper a-posteriori parameter choice, based on an estimate on the noise level that comprises both noise in the data and modelling errors [17]. However, it is generally difficult to estimate this last contribution. Then, in this paper

S. Bellavia
Dipartimento di Ingegneria Industriale, Università di Firenze, viale G.B. Morgagni 40, 50134 Firenze, Italia.

E. Riccietti
IRIT-INPT, rue Charles Camichel, B.P. 7122 31071, Toulouse Cedex 7, France. E-mail: elisa.riccietti@enseeiht.it

an ad-hoc method for ill-posed least-squares problems with nonzero residual is proposed, that does not need the estimation of modelling errors in order to choose the regularization parameter. Regularization strategies for nonzero residual ill-posed problems arising from specific applications have been considered in [12–16] . In these papers, the focus is on the reliable and efficient solution of the specific problem and a theoretical analysis of the regularizing properties of the methods used is not performed. The authors are aware only of [10], where convergence rates of Tikhonov method are considered.

Our target is then to design an implementable approach, that guarantees regularizing properties for the solution of such problems. Specifically, in case of exact data, it is desirable that the sequence of gradients of the objective function goes to zero and that the sequence of generated solution approximations converges to a solution of the unperturbed problem, even in case of noninvertible Fréchet derivative. In case of noisy data, if an initial guess close to the true solution is given, the method should have the potential to approach a solution of the unperturbed problem.

These properties are ensured by the Levenberg-Marquardt approaches in [5, 6] for zero residual problems, thanks to two key ingredients:

1. the employment of a proper stopping criterion to avoid semiconvergence phenomenon, i.e. the method must be stopped before convergence is reached, to be sure that the generated sequence is not approaching a solution of the noisy problem,
2. a mechanism to control the step length, that cannot be too large.

To this aim, in this manuscript, the conditions employed in [3, 5, 6, 8] are extended to the nonzero residual case and provide an adaptive choice of the regularization parameter λ_k and of the operator M_k in order to obtain a regularizing method for this class of problems. A regularized Gauss-Newton model is employed, as the aim is to handle small residual or mildly nonlinear problems. The obtained step p_k turns out to be an elliptical trust-region step, cf. [18, §4.5], whenever the Fréchet derivative is invertible. Also problems with convex constraints are considered, and a suitable extension of the procedure is devised, in order to solve them.

The local properties of the obtained method are investigated. This paper represents also a contribution in the analysis of iterated Tikhonov regularization methods for nonlinear least-squares problems. To the authors' knowledge indeed, analysis of such methods with regularizing operators different from the identity has never been considered. The authors are aware just of [4, 19] in which an analysis of iterated Tikhonov regularization with a fairly general regularization matrix is proposed for linear least-squares in a finite dimensional setting.

The paper is organized as follows. In Section 2, we give conditions that guarantee monotonic decrease of the error between the solution of the unperturbed problem and the current iterate, even when noisy data are considered. Satisfying these conditions motivates the choice of the free parameters of the method. In Section 3, we describe the choice of the operator M_k and we present the elliptical trust-region reformulation of the method, that implicitly provides an automatic choice of the regularizing parameters λ_k. We show that these two choices ensure the desired regularization properties. In Section 4, a convergence analysis of the proposed method is reported, focusing on the noise-free case in Section 4.1 and on noisy case in Section 4.2. In Section 5 the proposed approach is adapted to constrained problems. In Section 6 implementation details are given and the results of the numerical experimentation performed on examples of discrete ill-posed nonlinear least-squares problems ar shown. Numerical evidences of the method's effectiveness and of its regularizing properties are reported. It is worth mentioning that, to implement the proposed procedure, it is necessary to compute the action of the square root of $J(x_k)^T J(x_k)$ on a vector, where x_k is the current iterate and J is the discrete Jacobian. In procedure designed in this paper is tailored for medium scale problems. Therefore, in the numerical results provided, the square root of the matrix is directly computed by the singular value decomposition of J, that is then used also to solve the linear systems to compute the step. In case of large scale problems this is not feasible, and the method should rather rely on suitable iterative approaches that approximate the action of $J(x_k)^T J(x_k)$ on a vector, such as the Lanczos process [20, §13.2].

2 Preliminaries

Let $F : \mathcal{D}(\mathcal{F}) \subseteq \mathcal{X} \to \mathcal{Y}$ be a nonlinear map between Hilbert spaces $\mathcal{X}, \mathcal{Y}$, with inner products $\langle \cdot, \cdot \rangle$ and norms $\| \cdot \|$. The domain of definition of F will be denoted with $\mathcal{D}(F)$ and least squares problems

$$\min_x f(x) = \frac{1}{2}\|F(x) - y\|^2, \tag{1}$$

that are ill-posed, in the sense that the solutions do not depend continuously on the data, will be considered. It is assumed that a solution $x^\dagger$ for (1) exists. If for all the solutions of (1) it holds $\|F(x) - y\| > 0$, then we say that the problem has nonzero residual. It is assumed throughout the paper that $\mathcal{D}(F)$ is infinite dimensional and F has compact Fréchet derivative F'. The adjoint operator will be denoted with F'^*. It is assumed to have only noisy data y^δ at disposal, such that, given $\delta \geq 0$:

$$\|y - y^\delta\| \leq \delta, \tag{2}$$

so that the following noisy problem has to be dealt with:

$$\min_x f_\delta(x) = \frac{1}{2}\|F(x) - y^\delta\|^2. \tag{3}$$

A Levenberg-Marquardt method is considered. Throughout the paper, the iterates are denoted by x_k^δ, if the data are exact x_k may be used in alternative to x_k^δ. By $x_0^\delta = x_0$ an initial guess is denoted, which may incorporate a-priori knowledge of an exact solution. Le also be defined

$$B_k = F'(x_k^\delta)^* F'(x_k^\delta), \qquad f'_k = F'(x_k^\delta)^*(F(x_k^\delta) - y^\delta), \tag{4}$$

and let $B_k^\dagger$ be the Moore-Penrose pseudoinverse of B_k, and with $\mathcal{R}(B_k)$ its range. The open ball of center x and radius ρ will be denoted as $\mathcal{B}_\rho(x)$. At each iteration, given a positive parameter λ_k and the current solution approximation x_k^δ, the step is computed solving the following problem:

$$\min_p \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p\|^2 + \frac{\lambda_k}{2}\|p\|^2.$$

In order to improve the quality of the computed solution approximation, a symmetric and positive definite regularizing operator $M_k : \mathcal{X} \to \mathcal{X}$ can be introduced [19], so that the minimization problem becomes:

$$\min_p \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p\|^2 + \frac{\lambda_k}{2}\|M_k^{\frac{1}{2}} p\|^2. \tag{5}$$

A sequence $\{x_k^\delta\}$ is generated, forming at each iteration the new iterate using the computed step p_k: $x_{k+1}^\delta = x_k^\delta + p_k$.

In order to properly choose parameter λ_k and operator M_k in (5), it has to be taken into account that non-stationary iterated Tikhonov procedures for zero-residual problems [3,6] provide regularizing properties thanks to the fact that the method achieves monotone decrease of the norm of the error $e_k = x^\dagger - x_k^\delta$ between the true solution and the current iterate, even when noisy problems are solved. It will be proved in Lemma 2.1 that this property is joined by the Levenberg-Marquardt procedure (5) applied to nonzero residual problems, whenever the following two conditions hold:

$$B_k^\dagger p_k = -\frac{1}{\lambda_k} m_k(p_k), \tag{6}$$

$$\|m_k(e_k)\| \leq \frac{1}{\theta_k}\|m_k(p_k)\|, \quad \theta_k > 1, \tag{7}$$

where, using notation (4),

$$\begin{aligned} m_k(p) =& F'(x_k^\delta)^* F'(x_k^\delta)p + F'(x_k^\delta)^*(F(x_k^\delta) - y^\delta) \\ =& F'(x_k^\delta)^* F'(x_k^\delta)p + f'_\delta(x_k^\delta) = B_k p + f'_k. \end{aligned} \tag{8}$$

Note that in case f_δ is twice continuously differentiable, for its second derivative it holds:

$$f''_\delta(x) = F'(x)^* F'(x) + S(x) = F'(x)^* F'(x) + F''(x)^*(F(x) - y^\delta). \tag{9}$$

Then the model $m_k(p)$ corresponds to the first order Taylor expansion of f'_δ in x_k^δ, omitting the second order term $S(x_k^\delta)$ in $f''_\delta(x_k^\delta)$. It is well known that if $\|S(x^\dagger)\|$ is too large, the Gauss-Newton method may not be locally convergent, [21, §10.2]. $\|S(x^\dagger)\|$ is a combined measure of the nonlinearity and residual size of the problem. This is the reason why the method can handle only small residual or mildly nonlinear problems. Note that $m_k(p)$ is the gradient of

$$\frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p\|^2,$$

i.e. of the approximation of the function f_δ around the current iterate x_k^δ adopted in (5).

Lemma 2.1 *Assume that $x^\dagger$ is a solution of* (1)*. Let $e_k = x^\dagger - x_k^\delta$ and $m_k(p)$ defined in* (8)*. Assume that* (6) *is satisfied and that there exists $\theta_k > 1$ such that condition* (7) *holds. Let $x_{k+1}^\delta = x_k^\delta + p_k$ with $p_k = p(\lambda_k)$ solution of* (5)*. If $p_k \in \mathcal{R}(B_k)$, it holds*

$$\|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k^\delta - x^\dagger\|^2 \leq \frac{2}{\lambda_k}\left(\frac{1}{\theta_k} - 1\right)\|m_k(p_k)\|^2. \tag{10}$$

Proof Note that,

$$\begin{aligned}\|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k^\delta - x^\dagger\|^2 =& 2\langle x_{k+1}^\delta - x_k^\delta, x_k^\delta - x^\dagger\rangle + \|x_{k+1}^\delta - x_k^\delta\|^2\\ =& 2\langle p_k, -e_k\rangle + \|p_k\|^2.\end{aligned}$$

The fact that $p_k \in \mathcal{R}(B_k)$ yields:

$$\langle B_k^\dagger p_k, B_k p_k\rangle = \|p_k\|^2, \qquad \langle B_k^\dagger p_k, B_k e_k\rangle = \langle p_k, e_k\rangle. \tag{11}$$

From (11) it follows:

$$\begin{aligned}2\langle p_k, -e_k\rangle =& 2\langle B_k^\dagger p_k, -B_k e_k\rangle\\ =& -2\langle B_k^\dagger p_k, f'_k + B_k e_k\rangle + 2\langle B_k^\dagger p_k, f'_k + B_k p_k\rangle - 2\langle B_k^\dagger p_k, B_k p_k\rangle\\ =& -2\langle B_k^\dagger p_k, f'_k + B_k e_k\rangle + 2\langle B_k^\dagger p_k, f'_k + B_k p_k\rangle - 2\|p_k\|^2.\end{aligned}$$

Then, taking into account also (6) and (8) it holds:

$$\begin{aligned}\|e_{k+1}\|^2 - \|e_k\|^2 &= \frac{2}{\lambda_k}\langle m_k(p_k), m_k(e_k)\rangle - \frac{2}{\lambda_k}\langle m_k(p_k), m_k(p_k)\rangle - \|p_k\|^2\\ &\leq \frac{2}{\lambda_k}\|m_k(p_k)\|\|m_k(e_k)\| - \frac{2}{\lambda_k}\|m_k(p_k)\|^2 - \|p_k\|^2.\end{aligned}$$

From (7) it follows

$$\begin{aligned}\|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k^\delta - x^\dagger\|^2 &\leq \frac{2}{\lambda_k}\frac{1}{\theta_k}\|m_k(p_k)\|^2 - \frac{2}{\lambda_k}\|m_k(p_k)\|^2 - \|p_k\|^2\\ &\leq \frac{2}{\lambda_k}\left(\frac{1}{\theta_k} - 1\right)\|m_k(p_k)\|^2,\end{aligned}$$

which yields the thesis. □

Then, in order to obtain the desired monotone decrease of the error, it needs to be ensured $p_k \in \mathcal{R}(B)$, and (6) and (7) to hold.

3 The Method

Motivated by the previous considerations, a non-stationary iterated Tikhonov procedure for least-squares problems employing a step satisfying both conditions (6) and (7) is proposed here.

A step p_k is solution of (5), if and only if it satisfies the following linear operator equation:

$$(F'(x_k^\delta)^* F'(x_k^\delta) + \lambda_k M_k)p = -F'(x_k^\delta)^*(F(x_k^\delta) - y^\delta). \tag{12}$$

Relationship (6) can be obtained with a suitable choice of operator M_k. From (8) and (12) it follows that $-\frac{1}{\lambda_k} m_k(p_k) = M_k p_k$. This suggests us to choose $M_k = B_k^\dagger$ to have condition (6) verified.

To get (7), a reformulation of problem (5) is adopted, which has the advantage of providing us an automatic way for setting parameter λ_k ensuring (7) to hold. Assuming that $F'(x_k^\delta)$ is invertible, given the trust-region radius $\Delta_k > 0$ and the current iterate x_k^δ, at generic iteration k the following elliptic trust-region subproblem is solved, [18, ch. 4], [22, §7.4]:

$$\begin{aligned} &\min_p \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p\|^2, \\ &\text{s.t. } \|B_k^{-1/2}p\| \le \Delta_k, \end{aligned} \tag{13}$$

which has a unique solution, [23, Theorems 9.2.7, 10.2.15, 10.3.4]. The choice of the scaling operator B_k^{-1} is driven by the considerations stated above. KKT conditions for problem (13) are given by

$$(F'(x_k^\delta)^* F'(x_k^\delta) + \lambda B_k^{-1} p(\lambda) = -F'(x_k^\delta)^*(F(x_k^\delta) - y^\delta), \tag{14a}$$

$$\lambda(\|B_k^{-1/2}p(\lambda)\| - \Delta_k) = 0, \tag{14b}$$

$$\lambda \ge 0, \tag{14c}$$

$$\|B_k^{-1/2}p(\lambda)\| \le \Delta_k. \tag{14d}$$

Then, given the pair $(\lambda_k, p(\lambda_k)) \in \mathbb{R}^+ \times \mathcal{X}$ solution of (14), if $\lambda_k > 0$ the step $p_k = p(\lambda_k)$ solves (5) with $M_k = B_k^{-1}$.

With this reformulation of the problem, instead of choosing λ_k in (5), the trust-region radius Δ_k has to be properly selected in order to obtain a pair $(\lambda_k, p(\lambda_k))$ with $\lambda_k > 0$ and $p(\lambda_k)$ satisfying (7). It will be shown in Lemma 3.4 how to choose Δ_k to let these two conditions hold.

Letting $z = B_k^{-1/2}p$, problem (13) reduces to

$$\begin{aligned} &\min_z \frac{1}{2}\langle z, B_k^2 z\rangle + \langle B_k^{1/2} f_k', z\rangle + f_\delta(x_k^\delta), \\ &\text{s.t. } \|z\| \le \Delta_k. \end{aligned} \tag{15}$$

KKT conditions for problem (15) are given by

$$(B_k^2 + \lambda I)z(\lambda) = -B_k^{1/2} f_k', \tag{16a}$$

$$\lambda(\|z(\lambda)\| - \Delta_k) = 0, \tag{16b}$$

$$\lambda \ge 0, \tag{16c}$$

$$\|z(\lambda)\| \le \Delta_k. \tag{16d}$$

Let $(\lambda_k, z(\lambda_k)) \in \mathbb{R}^+ \times \mathcal{X}$ be the solution of (16), with $z(\lambda_k)$ the minimum norm solution of (16a). Let

$$p(\lambda) = B_k^{1/2} z(\lambda), \tag{17}$$

and $p_k = p(\lambda_k)$. Then, the pair (λ_k, p_k) is a KKT point for (13). The solution of (13) can be then found by solving (15), and through relation (17).

It will be proved that, in the proposed approach, λ_k is ensured to be strictly positive. Then, the numerical computation of (λ_k, p_k) after discretization requires the solution of a sequence of linear systems that are regularized by strictly positive λ values. Therefore, even if the discretized counterpart of the operator B_k is squared in (16a), the conditioning of the involved linear system is not deteriorated. Indeed, in the applications considered here, the ill-conditioning of the discrete Jacobian derives from the smallest singular values close to zero and the linear systems are regularized by strictly positive λ_k. It is assumed to have at disposal the singular value decomposition of the discrete Jacobian, in order to compute the square root of the matrix arising from the discretization of operator B_k.

In case $F'(x_k^\delta)$ is not invertible problem (13) is not well defined. However, the pair $(\lambda_k, z(\lambda_k))$, solution of the KKT (16), can still be computed and the step can be defined as

$$p_k = p(\lambda_k) = B_k^{1/2} z(\lambda_k). \tag{18}$$

In this case, step p_k satisfies (6), as it is shown in the next Lemma, but we cannot claim that it solves problem (13). In every case, $p \in \mathcal{R}(B_k)$.

In the following lemma, the singular value expansion of $F'(x_k^\delta)$ will be used, that will be denoted as $(\sigma_n; u_n, v_n)$, $n \in \mathbb{N}$, where $\{u_n\}_{n\in\mathbb{N}}$ and $\{v_n\}_{n\in\mathbb{N}}$ are a complete orthonormal system of eigenvectors for $F'(x_k^\delta)^* F'(x_k^\delta)$ and $F'(x_k^\delta) F'(x_k^\delta)^*$ respectively, and $\sigma_n > 0$ are written down in decreasing order with multiplicity, with 0 being the only accumulating point for the sequence $\{\sigma_n\}_{n\in\mathbb{N}}$ when $\dim \mathcal{R}(F'(x_k^\delta)) = \infty$.

Lemma 3.1 *Suppose* $\|f_k'\| \neq 0$. *Let* $z(\lambda)$ *be the minimum norm solution of* (16a) *with* $\lambda \geq 0$ *and* $p(\lambda)$ *given in* (17). *Then,*

$$B_k^\dagger p(\lambda) = -\frac{1}{\lambda} m_k(p(\lambda)). \tag{19}$$

Proof From the singular value expansion of $F'(x_k^\delta)$ it follows

$$F'(x_k^\delta) h = \sum_{n=1}^{\infty} \sigma_n \langle h, u_n \rangle v_n, \quad h \in \mathcal{X}, \tag{20}$$

$$F'(x_k^\delta)^* h = \sum_{n=1}^{\infty} \sigma_n \langle h, v_n \rangle u_n, \quad h \in \mathcal{Y}, \tag{21}$$

$$B_k h = \sum_{n=1}^{\infty} \sigma_n^2 \langle h, u_n \rangle u_n, \quad h \in \mathcal{X}. \tag{22}$$

From these relations and (16a) it follows

$$z(\lambda) = -\sum_{n=1}^{\infty} \frac{\sigma_n^2}{\sigma_n^4 + \lambda} \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n \tag{23}$$

and (17) yields

$$p(\lambda) = -\sum_{n=1}^{\infty} \frac{\sigma_n^3}{\sigma_n^4 + \lambda} \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n. \tag{24}$$

As $F'(x_k^\delta)$ is compact, for $h \in \mathcal{X}$ the Moore-Penrose pseudoinverse $B_k^\dagger$ of B_k can be defined as [24, §2.1]

$$B_k^\dagger h = \sum_{n=1}^{\infty} \sigma_n^{-2} \langle h, u_n \rangle u_n, \; h \in D(B_k^\dagger), \tag{25}$$

$$D(B_k^\dagger) = \{h \in \mathcal{X} \mid \sum_{n=1}^{\infty} \sigma_n^{-4} |\langle h, u_n \rangle|^2 < \infty\}.$$

From (24) and (25) it follows

$$B_k^\dagger p(\lambda) = \sum_{n=1}^{\infty} \sigma_n^{-2} \langle p(\lambda), u_n \rangle u_n = -\sum_{n=1}^{\infty} \frac{\sigma_n}{\sigma_n^4 + \lambda} \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n. \tag{26}$$

Relations (21), (22) and (24) yield

$$m_k(p(\lambda)) = B_k p(\lambda) + f'_k \tag{27a}$$

$$= \sum_{n=1}^{\infty} \sigma_n^2 \langle p(\lambda), u_n \rangle u_n + \sum_{n=1}^{\infty} \sigma_n \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n \tag{27b}$$

$$= - \sum_{n=1}^{\infty} \frac{\sigma_n^5}{\sigma_n^4 + \lambda} \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n + \sum_{n=1}^{\infty} \sigma_n \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n \tag{27c}$$

$$= \sum_{n=1}^{\infty} \frac{\lambda \sigma_n}{\sigma_n^4 + \lambda} \langle F(x_k^\delta) - y^\delta, v_n \rangle u_n. \tag{27d}$$

Comparing (26) and (27) the thesis is obtained. □

3.1 The q-Condition and the Choice of the Trust-Region Radius

As already discussed, it is crucial to use a step p_k satisfying condition (7). It will be shown in the next section that this can be guaranteed provided that the step $p_k = p(\lambda_k)$ satisfies the following condition, which will be referred to as the q-condition:

$$\|m(p(\lambda_k))\| \geq q\|f'_k\|, \tag{28}$$

where $q \in (0, 1)$ is a given constant. In this section it will be shown how (28) can be enforced by a suitable trust-region radius choice.

Condition (28) controls the value of the norm of the model, that has to be greater than a fixed fraction of the norm of the gradient. It also provides a criterion to choose the free parameter λ_k in (5). Condition (28) is actually a constraint on the length of the step. Its effect on the step-length is illustrated in Figure 1, where $\|m_k(p(\lambda))\|$ (top) and $\|p(\lambda)\|$ (bottom) are plotted, varying λ, for a finite dimensional test case. By imposing (28), the regularization parameter λ is forced to be greater then the value λ_k^q satisfying

$$\|m(p(\lambda_k^q))\| = q\|f'_k\|, \tag{29}$$

avoiding too small values that correspond to large steps, as it is shown at the bottom of Figure 1. Lemmas 4.1 and 4.3 will show that this control on the step length allows to establish relationship

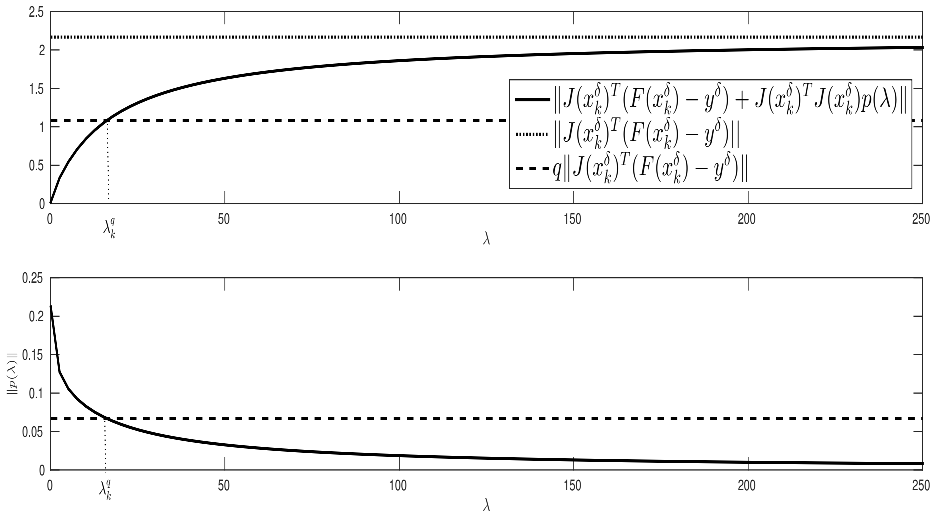


Fig. 1 Effect of q-condition on the step length.

(7) between the model evaluated in the step and that evaluated in the error.

It will be proved that a step p_k of the form (18) satisfying (28) exists and that an appropriate trust-region radius choice provides such a step. To this end the following preliminary results is needed.

Lemma 3.2 *Suppose* $\|f'_k\| \neq 0$. *Let* $z(\lambda)$ *be the minimum norm solution of* (16a) *with* $\lambda \geq 0$ *and* $p(\lambda)$ *given in* (17). *Then, it follows*

$$\|z(\lambda)\|^2 = \sum_{n=1}^{\infty} \left(\frac{\sigma_n^2}{\sigma_n^4 + \lambda} \right)^2 |\langle F(x_k^\delta) - y^\delta, v_n \rangle|^2. \tag{30}$$

Moreover, $\|m_k(p(\lambda))\|$ *is a monotone increasing function for* $\lambda \geq 0$ *and*

$$\lim_{\lambda \to 0} \|m_k(p(\lambda))\| = 0,$$
$$\lim_{\lambda \to \infty} \|m_k(p(\lambda))\| = \|F'(x_k^\delta)^*(F(x_k^\delta) - y^\delta)\|.$$

Proof Note that (23) yields (30). Moreover, from (26) it follows

$$\|B_k^\dagger p(\lambda)\|^2 = \sum_{n=1}^{\infty} \left(\frac{\sigma_n}{\sigma_n^4 + \lambda} \right)^2 |\langle F(x_k^\delta) - y^\delta, v_n \rangle|^2$$

and (27) yields

$$\|m_k(p(\lambda))\|^2 = \sum_{n=1}^{\infty} \left(\frac{\lambda \sigma_n}{\sigma_n^4 + \lambda} \right)^2 |\langle F(x_k^\delta) - y^\delta, v_n \rangle|^2. \tag{31}$$

This implies that the function $\|m_k(p(\lambda))\|$ is monotonic increasing. Then, taking into account that

$$\|F'(x_k^\delta)^*(F(x_k^\delta) - y^\delta)\|^2 = \sum_{n=1}^{\infty} \sigma_n^2 |\langle F(x_k^\delta) - y^\delta, v_n \rangle|^2, \tag{32}$$

the thesis easily follows. □

Now it is possible to prove that condition (28) can be satisfied.

Lemma 3.3 *Let* $z(\lambda)$ *be the minimum norm solution of* (16a) *with* $\lambda \geq 0$ *and* $p(\lambda)$ *be given in* (17). *It exists* $\lambda_k^q > 0$ *such that if* $\lambda_k \geq \lambda_k^q$ *then* $p_k = p(\lambda_k)$ *satisfies condition* (28).

Proof From Lemma 3.2 if follows that $\|m_k(p(\lambda))\|$ is a monotonic increasing function for $\lambda \geq 0$ and that there exists λ_k^q such that (29) holds. Then, condition (28) is satisfied for any $\lambda_k \geq \lambda_k^q$ and $\lambda_k^q = 0$ if and only if it holds $\|F'(x_k^\delta)^*(y^\delta - F(x_k^\delta))\| = 0$. □

We now provide a suitable choice of the trust-region radius, that guarantees that the resulting regularization parameter λ_k is strictly positive and big enough to ensure the step $p_k = p(\lambda_k)$ to satisfy condition (28).

Lemma 3.4 *Let* $z(\lambda)$ *be the minimum norm solution of* (16a) *with* $\lambda \geq 0$ *and* $p(\lambda)$ *be given in* (17). *If*

$$\Delta_k \leq \frac{1-q}{\|B_k\|^2} \|B_k^{1/2} f'_k\| \tag{33}$$

the step $p_k = p(\lambda_k)$ *satisfies* (28) *and* $\lambda_k > 0$.

Proof From (31) and (32) it follows

$$\begin{aligned}\|m_k(p(\lambda))\|^2 =&\lambda^2 \sum_{n=1}^{\infty} \left(\frac{\sigma_n}{\sigma_n^4+\lambda}\right)^2 |\langle F(x_k^\delta)-y^\delta, v_n\rangle|^2 \\ \geq& \frac{\lambda^2 \sum_{n=1}^{\infty} \sigma_n^2 |\langle F(x_k^\delta)-y^\delta, v_n\rangle|^2}{(\|B_k\|^2+\lambda)^2} \\ =& \frac{\lambda^2}{(\|B_k\|^2+\lambda)^2} \|F'(x_k^\delta)^*(F(x_k^\delta)-y^\delta)\|^2.\end{aligned}$$

Then, an upper bound for λ_k^q, defined in (29), can be obtained proceeding as follows:

$$\begin{aligned}q\|F'(x_k^\delta)^*(F(x_k^\delta)-y^\delta)\| &= \|m_k(p(\lambda_k^q))\| \\ &\geq \frac{\lambda_k^q}{\|B_k\|^2+\lambda_k^q} \|F'(x_k^\delta)^*(F(x_k^\delta)-y^\delta)\|,\end{aligned}$$

so

$$\lambda_k^q \leq \frac{q\|B_k\|^2}{1-q}. \tag{34}$$

By (16a) one has

$$\|z(\lambda_k^q)\| \geq \frac{\|B_k^{1/2} f_k'\|}{\|B_k^2+\lambda_k^q I\|}, \tag{35}$$

and by (34) it holds

$$\|B_k^2+\lambda_k^q I\| \leq \frac{\|B_k\|^2}{1-q}.$$

By construction $\|z_k\| \leq \Delta_k$. If (33) holds, from (35) it follows

$$\|z_k\| = \|z(\lambda_k)\| \leq \frac{1-q}{\|B_k\|^2}\|B_k^{1/2} f_k'\| \leq \frac{\|B_k^{1/2} f_k'\|}{\|B_k^2+\lambda_k^q I\|} \leq \|z(\lambda_k^q)\|.$$

Since by (30) it follows that $\|z(\lambda)\|$ is monotonically decreasing, the previous inequality yields $\lambda_k \geq \lambda_k^q > 0$ and by Lemma 3.3 the thesis holds. □

Note that with this choice of Δ_k it is not necessary to check if condition (28) is satisfied. Moreover, from (33) the trust-region radius goes to zero whenever $\|f_k'\|$ converges to zero.

Remark 3.1 Lemma 3.4 shows that λ_k is strictly positive, then from (16b) it follows that $\|z_k\| = \Delta_k$, i.e. the trust-region is active. Standard trust-region approaches, used to solve well-posed problems, are designed to have inactive trust-region in the last stage of the process, in order to recover the fast convergence of the full Newton step. On the contrary, in this context the trust-region is required to be eventually active in order to produce a regularizing effect.

From Lemma 3.1 and Lemma 3.4 it can be concluded that a step p_k of the form (18), satisfying both (6) and (28) exists. In the next section it will be shown that if condition (28) is met, then also condition (7) holds, so the method provides the desired monotone decrease of the norm of the error, as stated in Lemma 2.1.

The result in Lemma 3.4 suggests the trust-region iteration described in Algorithm 3.1. Once p_k has been obtained, the classical ratio:

$$\pi_k(p_k) = \frac{ared(p_k)}{pred(p_k)} = \frac{f_\delta(x_k^\delta) - f_\delta(x_k^\delta + p_k)}{f_\delta(x_k^\delta) - \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\|^2} \tag{36}$$

is computed. As in classical trust-region approaches, if there is a good agreement between the function and the model, the step is accepted. Otherwise the step is rejected and the trust-region is reduced.

Algorithm 3.1: k-th iteration of the regularizing trust-region method for problem (1)

Given x_k^δ, $\eta \in (0,1)$, $\gamma \in (0,1)$, $0 < C_{\min} < C_{\max}$, $q \in (0,1)$.
Exact data: y; Noisy data: y^δ.

1. Choose $\Delta_k \in \left[C_{\min}\|B_k^{1/2} f_k'\|, \min\left\{C_{\max}, \dfrac{1-q}{\|B_k\|^2}\right\} \|B_k^{1/2} f_k'\|\right]$
2. Repeat
   2.1 Find the pair (z_k, λ_k) solution of KKT conditions (16).
   2.2 Set $p_k = B_k^{1/2} z_k$
   2.3 Compute
   $$\pi_k(p_k) = \frac{f_\delta(x_k^\delta) - f_\delta(x_k^\delta + p_k)}{f_\delta(x_k^\delta) - \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta) p_k\|^2}$$
   2.4 If $\pi_k(p_k) < \eta$,set $\Delta_k = \gamma \Delta_k$.
   Until $\pi_k(p_k) \geq \eta$.
3. Set $x_{k+1}^\delta = x_k^\delta + p_k$.

In case of noisy data, the process is stopped at iteration $k^*(\delta)$ satisfying the following discrepancy principle:

$$\|f'_{k^*(\delta)}\| \leq \tau\delta < \|f_k'\|, \quad 0 \leq k < k^*(\delta), \tag{37}$$

where $\tau > 0$ is appropriately chosen. Then, as the aim is to find a zero of f', the process is stopped whenever the norm of the gradient goes under the noise level. In case of zero residual problems, it is common to use the following discrepancy principle:

$$\|F(x_{k^*(\delta)}^\delta) - y^\delta\| \leq \tau\delta < \|F(x_k^\delta) - y^\delta\|, \quad 0 \leq k < k^*(\delta).$$

However, in case of nonzero residual problems this criterion cannot be employed, as the residual is not expected to go to zero, and at the solution it may be greater than the noise level.

4 Convergence Theory

For the convergence analysis the following Assumption is needed.

Assumption 4.1 *Given $x, \tilde{x}$ in a suitable neighbourhood of the solution $x^\dagger$ to* (1)*, the following inequality holds:*

$$\|f'(\tilde{x}) - f'(x) - F'(x)^* F'(x)(\tilde{x} - x)\| \leq (c\|\tilde{x} - x\| + \sigma)\|f'(x) - f'(\tilde{x})\|, \tag{38}$$

for suitable constants $c > 0$ and $\sigma \in (0, q)$.

Inequality (38) is motivated by the following observations. In case of zero residual problems the tangential cone condition:

$$\|F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)\| \leq c\|\tilde{x} - x\| \; \|F(\tilde{x}) - F(x)\|, \quad x, \tilde{x} \in \mathcal{B}_\rho(x^\dagger),$$

with $c > 0$, is widely assumed, see for example ([1, §2.1],[5–8,25]. This condition is actually a requirement on the Taylor reminder of F. Stationary points of (1) are the solutions of $f'(x) = 0$. Moreover it holds

$$f''(x) = F'(x)^* F'(x) + F''(x)^*(F(x) - y). \tag{39}$$

Then, drawing inspiration from the tangential cone condition, (38) is a requirement on the Taylor reminder of $f'(x)$. In fact, the tangential cone condition for problem $f'(x) = 0$ turns out to be

$$\|f'(\tilde{x}) - f'(x) - f''(x)(\tilde{x} - x)\| \leq c\|\tilde{x} - x\| \; \|f'(\tilde{x}) - f'(x)\|, \quad x, \tilde{x} \in \mathcal{B}_\rho(x^\dagger).$$

Since in the proposed approach the term $S(x)$ is dropped, (38) is assumed, where the constant σ appears in the right-hand side. This constant is a combined measure of the nonlinearity and the residual size of the problem, cf. [21, §10.2], as it represents a bound for $\|S(x)\| = \|F''(x)^*(F(x) - y)\|$. The assumption $\sigma < q$ implies that the analysis is focused on small residual problems, due to the fact that second order information are discarded. In Section 6 numerical evidence for condition (38) will be provided, for the considered test problems.

In the analysis also the Lipschitz continuity of F' will be assumed.

Assumption 4.2 *F' is Lipschitz continuous in a neighbourhood of the level set $\mathcal{L} = \{x \in \mathbb{R}^n$ s.t. $f_\delta(x) \le f_\delta(x_0)\}$ with Lipschitz constant L.*

4.1 Noise-Free Case

In this section, noise-free problems are considered, $\delta = 0$ is assumed and the symbol δ is dropped from the generated sequence, the data y and the function. First, note that well-known global convergence properties of trust-region and Levenberg-Marquardt methods yield the following result.

Theorem 4.1 (Theorem 10.3 [18], §8.4 [22])
Suppose that Assumption 4.2 holds and that $\|F'(x)\|$ is bounded above on $\mathcal{L}$. Then, the sequence $\{x_k\}$ generated by Algorithm 3.1 satisfies

$$\lim_{k\to\infty} \|f'_k\| = \lim_{k\to\infty} \|F'(x_k)^*(y - F(x_k))\| = 0.$$

Then, any accumulation point of the sequence is a stationary point of the function f.

The analysis will now be focused on the local convergence properties of the method. Then, we assume that there exists a specific iterate $\bar{k}$ such that $x_{\bar{k}}$ is sufficiently close to a solution of (1) and inequality (38) holds in a neighbourhood of such iterate.

Assumption 4.3 *Let $\delta = 0$ and $x^\dagger$ be a solution of (1). Suppose that for some iteration index $\bar{k}$ there exist $\rho > 0$, $c > 0$ and $\sigma \in (0, q)$ such that inequality (38) holds for any $x, \tilde{x} \in \mathcal{B}_{2\rho}(x_{\bar{k}})$ and*

$$\|x_{\bar{k}} - x^\dagger\| < \min\left\{\frac{q-\sigma}{c}, \rho\right\}. \tag{40}$$

In the following Lemma, we show that, under the previous assumptions, (7) holds for all $k \ge \bar{k}$ and therefore by Lemma 2.1 the error decreases monotonically for $k \ge \bar{k}$.

Lemma 4.1 *Assume that Assumption 4.3 holds. Let $e_k = x^\dagger - x_k$ and p_k computed at Step 2.2 of Algorithm 3.1. Then, it exists $\theta_k > 1$ such that condition (7) holds for all $k \ge \bar{k}$.*

Proof From the choice of Δ_k at Step 2 of Algorithm 3.1 and Lemma 3.4 it follows that the step p_k satisfies condition (28). From (38) and (28)

$$\begin{aligned}\|m_{\bar{k}}(e_{\bar{k}}\| &\le (c\|e_{\bar{k}}\| + \sigma)\,\|F'(x^*_{\bar{k}}(F(x_{\bar{k}} - y)\| \le \\ &\le \left(\frac{c\|e_{\bar{k}}\| + \sigma}{q}\right)\|m_{\bar{k}}(p_{\bar{k}}\|,\end{aligned}$$

so that condition (7) is satisfied for $k = \bar{k}$ with $\theta_{\bar{k}} = \frac{q}{c\|e_{\bar{k}}\|+\sigma} > 1$ from Assumption 4.3. From Lemma 2.1 it follows $\|e_{\bar{k}+1}\| < \|e_{\bar{k}}\|$, so that it holds $x_{\bar{k}+1} \in \mathcal{B}_{2\rho}(x_{\bar{k}}) \cap B_\rho(x^\dagger)$ and Assumption 4.3 is valid also for $k = \bar{k}+1$. Repeating the above arguments, by induction it is possible to prove that condition (7) holds for all $k \ge \bar{k}$, with

$$\theta_k = \frac{q}{c\|e_k\| + \sigma} > 1. \tag{41}$$

□

In next Lemma some important features of the procedure are shown.

Lemma 4.2 *Suppose that Assumption 4.3 holds. Then, Algorithm 3.1 generates a sequence $\{x_k\}$ such that, for $k \ge \bar{k}$,*

(i) x_k belongs to $\mathcal{B}_{2\rho}(x_{\bar{k}})$ and to $\mathcal{B}_\rho(x^\dagger)$ and $\|x_{k+1} - x^\dagger\| < \|x_k - x^\dagger\|$,
(ii) it exists a constant $\bar{\lambda} > 0$ such that $\lambda_k \le \bar{\lambda}$.

Proof (i) From Lemma 4.1, Lemma 2.1 holds for all $k \geq \bar{k}$ and (10) implies that the sequence $\{\|x_k - x^\dagger\|\}_{k=\bar{k}}^{\infty}$ is monotonic decreasing. As a consequence, x_k belongs to $\mathcal{B}_{2\rho}(x_{\bar{k}}) \cap \mathcal{B}_\rho(x^\dagger)$ for all $k \geq \bar{k}$ and $\theta_{k+1} > \theta_k$ for all $k \geq \bar{k}$.

(ii) The proof of the last assertion of the Lemma follows the lines of Lemma 4.2 in [5] and is reported in the Appendix. □

In the following theorem convergence of the sequence $\{x_k\}$ to a point belonging to $\mathcal{S} \cap \bar{\mathcal{B}}_\rho(x^\dagger)$ is proved, where

$$\mathcal{S} = \{x \mid f'(x) = 0\}, \tag{42}$$

and $\bar{\mathcal{B}}_\rho(x^\dagger)$ is the closed ball of center $x^\dagger$ and radius ρ.

Theorem 4.2 *Suppose that Assumption 4.3 holds. Then, the sequence $\{x_k\}$ generated by Algorithm 3.1 converges to $x^* \in \mathcal{S} \cap \bar{\mathcal{B}}_\rho(x^\dagger)$.*

Proof Let $\bar{k}$ as in Assumption 4.3. Lemma 4.1 and 4.2 show that (10) holds for all $k \geq \bar{k}$ with θ_k given in (41). Let $e_k = x^\dagger - x_k$, $\bar{k} \leq j < k$ and l between j and k such that

$$\|F'(x_l)^*(y - F(x_l))\| = \min_{j \leq i < k} \|F'(x_i)^*(y - F(x_i))\|.$$

It holds:

$$\begin{aligned}
\|e_j - e_l\|^2 &= 2\langle e_l - e_j, e_l\rangle + \|e_j\|^2 - \|e_l\|^2, \\
\|e_l - e_k\|^2 &= 2\langle e_l - e_k, e_l\rangle + \|e_k\|^2 - \|e_l\|^2.
\end{aligned}$$

Let $\gamma = c\|x^\dagger - x_{\bar{k}}\| \geq c\|x^\dagger - x_i\|$ for all $i \geq \bar{k}$. Using (38) and the definition of l it holds, for all $j \leq i < k$, that

$$\begin{aligned}
\|B_i e_l\| \leq & \|-f_i' - B_i(x^\dagger - x_i)\| + \|f_l' - f_i' - B_i(x_l - x_i)\| + \|f_l'\| \\
\leq & (c\|x^\dagger - x_i\| + \sigma)\|f_i'\| + (c\|x_l - x_i\| + \sigma)\|f_l' - f_i'\| + \|f_l'\| \\
\leq & (c\|x^\dagger - x_i\| + 2\sigma + c\|x_l - x_i\|)\|f_i'\| + (c\|x_l - x_i\| + \sigma + 1)\|f_l'\| \\
\leq & (3c\|x^\dagger - x_i\| + 2c\|x^\dagger - x_l\| + 3\sigma + 1)\|f_i'\| \\
\leq & (5\gamma + 3\sigma + 1)\|f_i'\| = \tilde{c}\|f_i'\|
\end{aligned}$$

where $\tilde{c} = 5\gamma + 3\sigma + 1$, so that

$$\|B_i e_l\| \leq \tilde{c}\|f_i'\|, \tag{43}$$

for all $j \leq i < k$. Taking into account that p_k belongs to the range space of $B_k^{1/2}$, from (6), (43) and (28) it follows that for $k > j \geq \bar{k}$:

$$\begin{aligned}
|\langle e_l - e_k, e_l\rangle| &= \left|\sum_{i=l}^{k-1} \langle p_i, e_l\rangle\right| = \left|\sum_{i=l}^{k-1} \langle B_i^\dagger p_i, B_i e_l\rangle\right| \\
&= \left|\sum_{i=l}^{k-1} \frac{1}{\lambda_i}\langle m_i(p_i), B_i e_l\rangle\right| \leq \sum_{i=l}^{k-1} \frac{1}{\lambda_i}\|m_i(p_i)\|\|B_i e_l\| \\
&\leq \sum_{i=l}^{k-1} \frac{\tilde{c}}{\lambda_i}\|m_i(p_i)\|\|f_i'\| \leq \sum_{i=l}^{k-1} \frac{\tilde{c}}{q\lambda_i}\|m_i(p_i)\|^2.
\end{aligned}$$

Thus (10) yields

$$|\langle e_l - e_k, e_l\rangle| \leq \sum_{i=l}^{k-1} \frac{\tilde{c}}{2q}\frac{\theta_i}{\theta_i - 1}(\|e_i\|^2 - \|e_{i+1}\|^2) \leq \beta_{\bar{k}}(\|e_l\|^2 - \|e_k\|^2), \tag{44}$$

where $\beta_{\bar{k}} = \dfrac{\tilde{c}}{2q}\dfrac{\theta_{\bar{k}}}{\theta_{\bar{k}} - 1}$ and the fact that $\theta_i/(\theta_i - 1) < \theta_{\bar{k}}/(\theta_{\bar{k}} - 1)$ has been used, since function $\theta/(\theta - 1)$ is monotonic decreasing and sequence θ_k is monotonic increasing (see proof of Lemma 4.2. Similarly, it is possible to show that

$$|\langle e_l - e_j, e_l\rangle| \leq \beta_{\bar{k}}(\|e_j\|^2 - \|e_l\|^2). \tag{45}$$

Then from (44) and (45) it follows

$$\begin{aligned}
\|e_k - e_l\|^2 &= 2\langle e_l - e_k, e_l\rangle + \|e_k\|^2 - \|e_l\|^2 \le (2\beta_{\bar{k}} + 1)(\|e_l\|^2 - \|e_k\|^2),\\
\|e_l - e_j\|^2 &= 2\langle e_l - e_j, e_l\rangle + \|e_j\|^2 - \|e_l\|^2 \le (2\beta_{\bar{k}} + 1)(\|e_j\|^2 - \|e_l\|^2),\\
\|x_k - x_j\|^2 &= \|e_k - e_j\|^2 \le \|e_k - e_l\|^2 + \|e_l - e_j\|^2.
\end{aligned}$$

Since the sequence $\{\|e_k\|\}$ is bounded from below and monotonic decreasing, hence convergent, it follows that $\{x_k\}$ is a Cauchy sequence, i.e. $\{x_k\}$ converges to a limit point x^*. As $x_k \in \mathcal{B}_\rho(x^\dagger)$ for $k \ge \bar{k}$, it follows $\|x^* - x^\dagger\| \le \rho$. Finally, from Lemma 4.2 it is known that $\lambda_k \le \bar{\lambda}$ for $k \ge \bar{k}$. Also, $\frac{\theta_k - 1}{\theta_k} > \frac{\theta_{\bar{k}} - 1}{\theta_{\bar{k}}}$ for $k \ge \bar{k}$ since the function $\frac{\theta-1}{\theta}$ is monotonically increasing. Then, by (10) and (28)

$$\|x_k - x^\dagger\| - \|x_{k+1} - x^\dagger\|^2 \ge 2q^2 \frac{\theta_{\bar{k}} - 1}{\theta_{\bar{k}}\bar{\lambda}} \|f'_k\|^2.$$

Thus, it is possible to conclude that $\|f'_k\|$ tends to zero and the limit x^* of x_k belongs to $\mathcal{S}$. □

4.2 Noisy Case

Here, the assumption $\delta > 0$ is made, and the regularizing properties of the method in case of noisy data are shown. It is assumed that there exists a specific iterate $x^\delta_{\bar{k}}$ sufficiently close to a solution $x^\dagger$ of (1), and that inequality (38) holds in a neighbourhood of such iterate.

Assumption 4.4 *Let $\delta > 0$ and $x^\dagger$ be a solution of* (1). *Suppose that for some iteration index $\bar{k} < k^*(\delta)$, with $k^*(\delta)$ defined in* (37), *there exist $\rho > 0$, $c > 0$ and $\sigma \in (0, q)$ such that inequality* (38) *holds for any $x, \tilde{x} \in \mathcal{B}_{2\rho}(x^\delta_{\bar{k}})$. Moreover assume that it exists a positive constant K such that*

$$\|F'(x)\| \le K$$

for any x belonging to the level set $\mathcal{L} = \{x \in \mathbb{R}^n \text{ s.t. } f_\delta(x) \le f_\delta(x_0)\}$ and that $x^\delta_{\bar{k}}$ satisfies

$$\|x^\delta_{\bar{k}} - x^\dagger\| < \min\left\{\frac{(q-\sigma)\tau - K(\sigma+1)}{c(K+\tau)}, \rho\right\}, \quad \text{with} \quad \tau > \frac{K(\sigma+1)}{q-\sigma}. \tag{46}$$

Note that in problems considered here, bound K is generally not large, typically of the order of the unit. Moreover, in the numerical results section it will be shown that the behaviour of the procedure does not depend strongly on the choice of q. Then, it is possible to ensure $q - \sigma$ to be positive and reasonably far from zero without affecting the method performance.

Lemma 4.3 *Assume that Assumption 4.4 holds and let $e_k = x^\dagger - x^\delta_k$ and p_k computed at Step 2.2 of Algorithm 3.1. Then, it exists $\theta_k > 1$ such that condition* (7) *holds for all $\bar{k} \le k < k^*(\delta)$.*

Proof From the choice of Δ_k at Step 1 of Algorithm 3.1 and Lemma 3.4 it follows that the step p_k satisfies condition (28). By (38) and (2) it follows

$$\begin{aligned}
\|m_{\bar{k}}(e_{\bar{k}})\| &= \|F'(x^\delta_{\bar{k}})^*(F(x^\delta_{\bar{k}}) - y^\delta + F'(x^\delta_{\bar{k}})(x^\dagger - x^\delta_{\bar{k}}))\|\\
&\le \|F'(x^\delta_{\bar{k}})^*(y^\delta - y)\| + \|F'(x^\delta_{\bar{k}})^*(F(x^\delta_{\bar{k}}) - y + F'(x^\delta_{\bar{k}})(x^\dagger - x^\delta_{\bar{k}}))\|\\
&\le K\delta + (c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma)\,\|F'(x^\delta_{\bar{k}})^*(y - F(x^\delta_{\bar{k}}))\|\\
&\le (1 + c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma)K\delta + (c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma)\,\|F'(x^\delta_{\bar{k}})^*(y^\delta - F(x^\delta_{\bar{k}}))\|.
\end{aligned}$$

Then, at iteration $\bar{k}$, conditions (37) and (28) give

$$\begin{aligned}
\|m_{\bar{k}}(e_{\bar{k}})\| &\le \left(K\frac{1 + c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma}{\tau} + (c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma)\right)\|F'(x^\delta_{\bar{k}})^*(y^\delta - F(x^\delta_{\bar{k}}))\|\\
&\le \left(K\frac{1 + c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma}{q\tau} + \frac{c\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma}{q}\right)\|m_{\bar{k}}(p_{\bar{k}}\|,
\end{aligned}$$

and (46) yields (7) at $k = \bar{k}$ with $\theta_{\bar{k}} = \dfrac{q\tau}{K + c(K+\tau)\|x^\dagger - x^\delta_{\bar{k}}\| + \sigma(K+\tau)}$, $\theta_{\bar{k}} > 1$. So Lemma 2.1 holds for $k = \bar{k}$ and $\|x_{\bar{k}+1} - x^\dagger\| < \|x_{\bar{k}} - x^\dagger\|$. Then, (46) holds also for $k = \bar{k}+1$. Repeating the above arguments, by induction it is possible to prove that, for $\bar{k} < k < k^*(\delta)$, condition (7) holds, with $\theta_k = \dfrac{q\tau}{K + c(K+\tau)\|x^\dagger - x^\delta_k\| + \sigma(K+\tau)} > 1$, which yields the thesis. □

Next Lemma shows key properties of Algorithm 3.1. Then, exploiting these results, it will be shown in Theorem 4.3 that, given a sequence $\{\delta_n\}$ of noise levels, under suitable assumptions, the sequence of computed approximations $\{x^{\delta_n}_{k_*(\delta_n)}\}$ goes to a stationary point of (1) whenever δ_n tends to zero. It is also proved that the discrepancy principle (37) is satisfied after a finite number $k^*(\delta)$ of steps.

Lemma 4.4 *Suppose that Assumptions 4.2 and 4.4 hold. Then, Algorithm 3.1 generates a sequence $\{x^\delta_k\}$ such that, for $\bar{k} \le k < k^*(\delta)$,*

(i) x^δ_k belongs to $\mathcal{B}_{2\rho}(x^\delta_{\bar{k}}) \cap \mathcal{B}_\rho(x^\dagger)$ and $\|x^\delta_{k+1} - x^\dagger\| < \|x^\delta_k - x^\dagger\|$;
(ii) there exists a constant $\bar{\lambda} > 0$ such that $\lambda_k \le \bar{\lambda}$.

Proof (i) From Lemmas 3.4 and 4.3, (10) is satisfied for all $\bar{k} \le k < k^*(\delta)$, and this implies that $\|x^\delta_{k+1} - x^\dagger\| < \|x^\delta_k - x^\dagger\|$, x^δ_k belongs to $\mathcal{B}_{2\rho}(x^\delta_{\bar{k}}) \cap \mathcal{B}_\rho(x^\dagger)$, and $\theta_{k+1} > \theta_k$ for all $\bar{k} \le k < k^*(\delta)$.

(ii) The proof of the last assertion of the Lemma follows the lines of Lemma 4.2 in [5] and is reported in the Appendix. □

Theorem 4.3 *Suppose that Assumptions 4.2 and 4.4 hold. Then, the iterates generated by Algorithm 3.1 satisfy the stopping criterion* (37) *after a finite number $k^*(\delta)$ of iterations.*

Moreover, suppose that the sequence $\{x_k\}$ generated with the exact data y satisfies $\pi_k(x_{k+1} - x_k) \neq \eta$, for all k, with $\pi_k(p_k)$ defined in (36). *Then the sequence $\{x^\delta_{k^*(\delta)}\}$ converges to a $x^* \in \mathcal{S} \cap \bar{\mathcal{B}}_\rho(x^\dagger)$, with $\mathcal{S}$ and $\bar{\mathcal{B}}_\rho(x^\dagger)$ defined in* (42), *whenever δ tends to zero.*

The proof of this result follows the lines of Theorem 4.5 in [5] and is reported in the Appendix. In Theorem 4.3 it is assumed that for all $k \ge 0$ it holds $\pi_k(x_{k+1} - x_k) \neq \eta$. Under this assumption the trust-region radius Δ_k selected in Algorithm 3.1, and the scalar λ_k, implicitly defined by the trust-region problem, depend continuously on $\delta > 0$. This feature is crucial for proving the convergence of sequence $\{x^\delta_{k^*(\delta)}\}$ to a stationary point of (1) as δ tends to zero. In fact, in case there exists an index k for which it holds $\pi_k(x_{k+1} - x_k) = \eta$, the sign of $\pi_k(x^{\delta_n}_{k+1} - x^{\delta_n}_k) - \eta$ is not predictable even for small δ_n, and therefore Δ_k does not depend continuously on δ.

5 Constrained Case

In many practical applications, problems with constraints on the variables are to be dealt with. Non-stationary iterated Tikhonov methods for linear least-squares problems with convex constraints have been considered in [4]. Let $\Omega \subset \mathcal{X}$ be a closed and convex set and consider the following problem:

$$\min_{x \in \Omega} f(x) = \frac{1}{2}\|F(x) - y^\delta\|^2. \tag{47}$$

Let $P_\Omega : \mathcal{X} \to \Omega$ be the metric projection of $\mathcal{X}$ on Ω:

$$P_\Omega(x) = \arg\min_{y \in \Omega} \frac{1}{2}\|x - y\|^2,$$

for all x in $\mathcal{X}$. It is assumed that a solution $x^\dagger \in \Omega$ exists and the computation of the projection P_Ω is not computationally expensive.

The procedure described in Section 3 can be modified as outlined in Algorithm 5.1 in order to handle the constraints and preserve its local properties. In what follows the noisy case is considered.

Algorithm 5.1: k-th iteration of the regularizing trust-region method for problem (47)
Given x_k^δ, $\eta \in (0,1)$, $0 < C_{\min} < C_{\max}$, $q \in (0,1)$, y^δ.
1. Choose $\Delta_k \in \left[C_{\min}\|B_k^{1/2} f_k'\|, \min\left\{C_{\max}, \frac{1-q}{\|B_k\|^2}\right\} \|B_k^{1/2} f_k'\|\right]$
2. Find the pair (z_k, λ_k) solution of KKT conditions (16), and set $p_k = B_k^{1/2} z_k$.
3. Set $x_{k+1}^\delta = P_\Omega(x_k^\delta + p_k)$.

In Algorithm 5.1, the decrease of the objective function is not imposed, as the aim here is just to sketch a local procedure for the constrained problem. A global convergent procedure for the noise-free case would require more sophisticated strategies to handle the constraints and at the same time provide regularization properties. Then the role of the trust-region is just that of providing a step satisfying (28). This step is used to compute the updated point $x_k^\delta + p_k$ that is then projected on the feasible set, so that the new solution approximation is computed as $x_{k+1}^\delta = P_\Omega(x_k^\delta + p_k)$. In this way the generated sequence x_k^δ belongs to Ω. All the local properties of the procedure are maintained, in particular the monotone decrease of the error, thanks to the following remark:

Remark 5.1 Since $x^\dagger \in \Omega$, then $\|P_\Omega(x_k^\delta + p_k) - x^\dagger\| \le \|x_k^\delta + p_k - x^\dagger\|$.

Lemma 5.1 *Assume that $x^\dagger$ is a solution of* (47). *Assume that there exists $\theta_k > 1$ such that condition* (7) *holds. Let $x_{k+1}^\delta = P_\Omega(x_k^\delta + p_k)$ with p_k computed at Step 2 of Algorithm 5.1. Then* (10) *holds.*

Proof

$$\begin{aligned}\|x_{k+1}^\delta - x^\dagger\|^2 - \|x_k - x^\dagger\|^2 &= \|P_\Omega(x_k^\delta + p_k) - x^\dagger\|^2 - \|x_k - x^\dagger\|^2 \\ &\le \|x_k^\delta + p_k - x^\dagger\|^2 - \|x_k - x^\dagger\|^2,\end{aligned}$$

and the thesis can be obtained repeating the proof of Lemma 2.1 since the step computed in Step 2 of Algorithm 5.1 satisfies (6). □

Thanks to this key result, proofs of Lemmas 4.3-4.4 and Theorem 4.3 can be repeated. In this regard it is worth underlying that the proof of point (ii) of Lemma 4.4 simplifies, as the upper bound on λ_k is given by inequality (56) as Δ_k is chosen at Step 1 of Algorithm 5.1 and it is not further reduced.

6 Numerical Results

In this section the numerical behaviour of the procedure is studied, that will be addressed as regularizing trust-region, in case of noisy data.

Four nonlinear ill-posed least-squares problems have been selected. **Problem 6.1** and **Problem 6.2** arise from the discretization of two parameter identification problems, while **Problem 6.3** and **Problem 6.4** are originally formulated as discrete problems. In the following, the Euclidean norm will be denoted with $\|\cdot\|$.

- **Problem 6.1**: `A 1D parameter identification problem.` Reconstruct c in the 1D-elliptic problem

$$-au_{xx} + cu = \varphi \quad \text{in } (0,1) \tag{48a}$$
$$u'(0) = 0,\ u'(1) = 0, \tag{48b}$$

given $u, \varphi \in L^2(0,1)$, [26]. Identifying c reduces to solve, for a given approximation $\bar{u}$ of a solution of (48), the following nonlinear least-squares problem:

$$\min_c \|F(x) - \bar{u}\|_{L^2}^2,$$

for F the operator mapping c to the corresponding solution of (48). It is chosen $a=4$, and φ given by (48a) with

$$c(x) = \sqrt{2}cos(2\pi x) + 2, \tag{49}$$

$$u(x) = cos(2\pi x) + 2. \tag{50}$$

The realistic situation in which both the solution u given in (50) of the partial differential equation and the function φ are known just in n points, $\{t_1,\dots,t_n\} \subset [0,1]$, is considered. Then $\tilde{\varphi}$ and $\tilde{u}$ are defined as the piecewise linear functions built interpolating $\{(t_i, \varphi(t_i))\}$ and $\{(t_i, u(t_i))\}$ respectively, for $i = 1,\dots,n$. Then, φ in (48) is replaced with $\tilde{\varphi}$. In the experiments u and φ are assumed to be known in $n = 39$ equispaced points in $[0,1]$. Note that when $\tilde{u}$ is built in this way, it cannot be attainable. In fact, if we define $H^2(0,1) = \{f \in L^2(0,1) \,|\, D^\alpha f \in L^2(0,1)\ \forall\alpha : |\alpha| \le 2\}$, the solution of (48) belongs to $H^2(0,1)$ for all $c \in L^2(0,1)$, while $\tilde{u}$ does not. Then, this test problem results to be a nonzero residual problem. Problem (48) is discretized using finite differences and $N = 113$ equispaced grids points. The matrix arising from the discretization of the differential operator $-au_{xx}$ on the grid $x_i = (i-1)\,h$, $h = 1/(N-1)$, $i = 1,\dots,N$ will be denoted by L. Let $\bar{\varphi}, \bar{u} \in \mathbb{R}^N$ be such that $\bar{\varphi}_i = \tilde{\varphi}(x_i)$, $\bar{u}_i = \tilde{u}(x_i)$, $i = 1,\dots,N$ and define for $c \in \mathbb{R}^N$ $F(c) = (L + \ \mathtt{diag}(c))^{-1}\bar{\varphi}$, with $\mathtt{diag}(c) \in \mathbb{R}^{n\times n}$. Then, identifying c reduces to solving

$$\min_{c\in\mathbb{R}^N} \frac{1}{2}\|F(c) - \bar{u}\|^2. \tag{51}$$

If c^* is the solution approximation found with exact data (i.e. $\delta = 0$), it holds $\|F(c^*) - \bar{u}\| \sim 1.e-3$. For this test problem the exact form of the the Jacobian matrix of F is given by:

$$J(c) = -(L + \mathtt{diag}(c))^{-1}(\mathtt{diag}(F(c))). \tag{52}$$

– **Problem 6.2**: `A 2D parameter identification problem.` The 2D version of **Problem 6.1** with $a = 1$ is considered. Namely, the problem consists of reconstructing c in the 2D-elliptic problem

$$-\Delta u + cu = \varphi \text{ in } \Omega \tag{53a}$$

$$u = \zeta \text{ on } \partial\Omega \tag{53b}$$

from the knowledge of u in $\Omega = (0,1)\times(0,1)$, $\varphi \in L^2(\Omega)$ and ζ the trace of a function in $H^2(\Omega)$. This problem has been widely studied, see for example [27,28].
The discretized version of the arising nonlinear least-squares problem is considered, obtained as described in [27]. Namely problem (53a)-(53b) was discretized using finite differences choosing as grid points $x_i = y_i = \frac{i-1}{n-1}$, for $i = 1,\dots,n$ and $n = 50$, and using lexicographical ordering, denoted by $l : \{1,\dots,n^2\} \to \{1,\dots,n^2\}$. Let A be the matrix arising from the discretization of the Laplacian operator, with $\bar{\varphi} = [\bar{\varphi}_1,\dots,\bar{\varphi}_{n^2}]^T$, where $\bar{\varphi}_{l(i,j)} = \varphi(x_i, y_j)$. Moreover for $c \in \mathbb{R}^{n^2}$ let $F(c) = (A + \ \mathtt{diag}(c))^{-1}\bar{\varphi}$. Then, $F : \mathbb{R}^{n^2} \to \mathbb{R}^{n^2}$, and the resulting discrete problem is a nonlinear least-squares problem of size $n^2 = 2500$:

$$\min_{c\in\mathbb{R}^{n^2}} \frac{1}{2}\|F(c) - \bar{u}\|^2,$$

for a given $\bar{u} \in \mathbb{R}^{n^2}$. For further details see [27]. The tests were conducted choosing $c(x,y) = 1.5\sin(4\pi x)\sin(6\pi y) + 3((x-0.5)^2 + (y-0.5)^2) + 2$ as a parameter to be identified. The solution $u(x,y)$ of (53) corresponding to this choice of $c(x,y)$ is $u(x,y) = 16x(1-x)y(y-1)+1$. Function φ in (53) has been defined from (53a). When the solution u is analytically known, this is a zero residual problem. In order to obtain a nonzero residual problem the data $\bar{u}$ are artificially set as a perturbation of $[u_1,\dots,u_{n^2}]$ with $u_{l(i,j)} = u(x_i, y_j)$, to let $c^\dagger = [c_1^\dagger,\dots,c_{n^2}^\dagger]^T$, where $c^\dagger_{l(i,j)} = c(x_i, y_j)$, be a stationary point with strictly positive residual. Specifically $\|J(c^\dagger)^T(F(c^\dagger) - \bar{u})\| = 0$ and $\|F(c^\dagger) - \bar{u}\| \simeq 0.1$, for J the Jacobian matrix of F.
For this test problem the exact form of the the Jacobian matrix of F is given by:

$$J(c) = -(A + \mathtt{diag}(c))^{-1}(\mathtt{diag}(F(c))). \tag{54}$$

- **Problem 6.3**: `A test problem arising in geophysics [11].` Starting from electromagnetic data collected by a ground conductivity meter, the aim is to reconstruct the electrical conductivity x of the soil with respect to depth z. In [11] a nonlinear model for the apparent conductivity $m_C(x, h)$ at height h above the ground is used, which depends on the value x of the conductivity. The soil is assumed to be divided in n layers, so that x_i is the conductivity in each layer and $x = (x_1, \dots, x_n)^T$. Multiple measurements are needed to recover the distribution of conductivity with respect to depth. It is assumed to record the apparent conductivity at heights $h_i, i = 1, \dots, m$, that generates values $b = (b_1, \dots, b_m)$. The problem of data inversion consists of computing the conductivity x minimizing the error in the model prediction:

$$\min_x \frac{1}{2}\|b - m_C(x)\|^2.$$

It is assumed that the conductivity distribution is a function of the depth, $x = \phi(z)$. In the experiments the piecewise linear function

$$\phi(z) = \begin{cases} \frac{8z+1}{5}, & \text{if } z \leq 0.5, \\ \frac{-2z+6}{5}, & \text{if } z > 0.5, \end{cases}$$

was used, expressed in Siemens/meter, with respect to the depth z, measured in meters. This implies the presence of a strongly conductive material at a given depth. The measurements are assumed to be taken at different heights $h_i = (i-1)\bar{h}$ above the ground, $i = 1, \dots, m$, for a chosen height step $\bar{h}$. The soil is divided into $n = 60$ layers, up to the depth of 2.5 meters, each of thickness $\bar{d} = 2.5/(n-1)$, selecting different depths under the ground level, $[z_1, \dots, z_n]$, where $z_j = (j-1)/\bar{d}$, $j = 1, \dots, n$. The proposed method is applied to synthetic data sets. Synthetic measurements are generated at $m = 40$ equispaced heights up to 1.9 meters to let $x^\dagger = (\phi(z_1), \dots, \phi(z_n))$ be a stationary point such that $\|m_C(x^\dagger) - b\| \simeq 0.48$. Note that the true electrical conductivity is approximated with a mathematical model, so it is reasonable to expect it to fit the data with a nonzero residual, even in the case of exact data. On this test problem also bound constraints are present, as the solution must be positive. Then, the projection strategy described in Section 5 is employed.
- **Problem 6.4**: `A fitting of a sum of two exponentials.` Given the model

$$y(t) = x_1 e^{-x_2 t} + x_3 e^{-x_4 t}, \tag{55}$$

recover the set of parameters $x^\dagger$ solving the following discrete least-squares problem:

$$\min_{x=[x_1,x_2,x_3,x_4]^T} \frac{1}{2}\|F(x) - y\|^2, \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}, \qquad F(x) = \begin{bmatrix} x_1 e^{-x_2 t_1} + x_3 e^{-x_4 t_1} \\ \vdots \\ x_1 e^{-x_2 t_m} + x_3 e^{-x_4 t_m} \end{bmatrix},$$

where the observations (t_i, y_i) $i = 1, \dots, m$, are given as follows.
The points t_i are equispaced in $[0, 10]$ and the data y_i have been chosen to let $x^\dagger = [0.2, -5, 0.4, -100]^T$ be a minimum of the problem with nonlinear residual $\|F(x^\dagger) - y\|$ of the order of 0.54, when the data are fitted with model (55). The experiments were conducted choosing $m = 1000$.

In the following, for uniformity of notation, for all the tests it is assumed that the minimization problem to be solved is

$$\min_x \frac{1}{2}\|F(x) - y\|^2$$

and the solution approached when the minimization problem is solved with exact data will be denoted with x^*.

First of all, the method relies on Assumption 4.1 for the proof of the regularizing properties. In Figure 2, numerical evidence for the assumption is reported, on all the test problems. Specifically, $\|g(\tilde{x}) - g(x_k) - J(x_k)^T J(x_k)(\tilde{x} - x_k)\|$ (dashed line) and $(c\|\tilde{x} - x_k\| + \sigma)\|g(\tilde{x}) - g(x_k)\|$ (solid line) are plotted, where x_k is the current iterate, $\tilde{x} \in \mathcal{B}_\rho(x_k)$, J is the Jacobian of F and g is the discrete gradient $g(x) = J(x)^T(F(x) - y)$. It was chosen $c = 0.1$, $\rho = 0.3$ and σ equal to the residual $\|F(x^*) - y\|$. Note that condition

$$\|g(\tilde{x}) - g(x_k) - J(x_k)^T J(x_k)(\tilde{x} - x_k)\| \leq (c\|\tilde{x} - x_k\| + \sigma)\|g(\tilde{x}) - g(x_k)\|,$$

is satisfied for x_k approaching x^*. These tests were repeated varying $\tilde{x} \in \mathcal{B}_\rho(x_k)$ and qualitatively the same results were obtained. In [29] the authors have proved that Assumption 4.1 holds for **Problem 6.1** in case of finite dimensional Hilbert spaces. That analysis could be extended also to problem **Problem 6.2**.

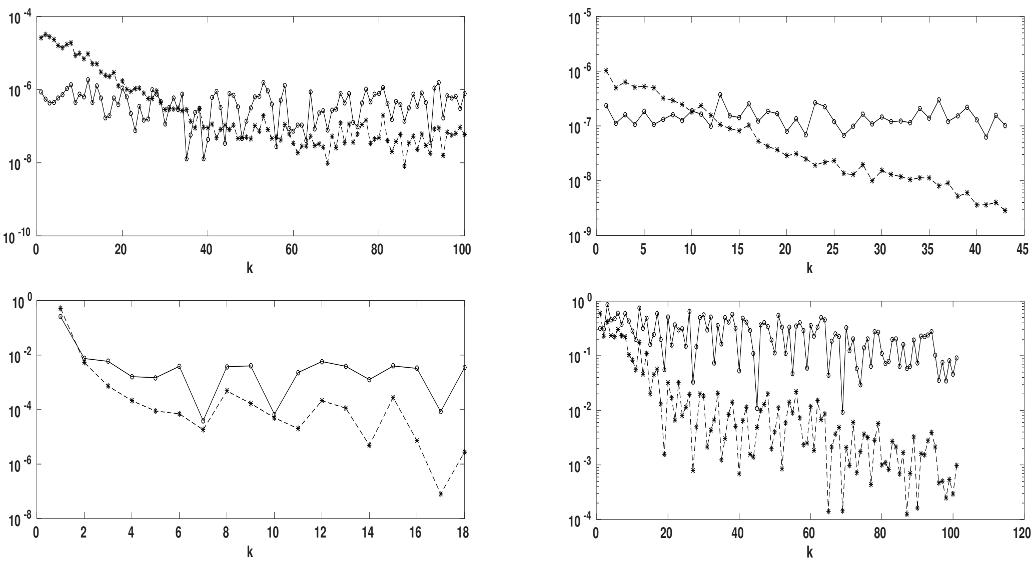


Fig. 2 Numerical evidence for Assumption 4.1: plot of $\|g(\tilde{x}) - g(x_k) - J(x_k)^T J(x_k)(\tilde{x} - x_k)\|$ (dashed line) and $(0.1\|\tilde{x} - x_k\| + \sigma)\|g(\tilde{x}) - g(x_k)\|$ (solid line) for $\tilde{x}$ randomly chosen in $B_\rho(x_k)$, $\rho = 0.3$, for **Problem 6.1** (top left, $\sigma = 4 \cdot 10^{-3}$), **Problem 6.2** (top right, $\sigma = 0.1$), **Problem 6.3** (bottom left, $\sigma = 0.48$), **Problem 6.4** (bottom right, $\sigma = 0.54$).

The practical implementation of the method will be now described. All procedures were implemented in MATLAB and run using MATLAB 2015A on an Intel Core(TM) i5-2467M 1.6 GHz, 4 GB RAM; the machine precision is $\epsilon_m \sim 2 \cdot 10^{-16}$. The trust-region procedure was implemented according to Algorithm 3.1.

The major implementation issues are as follows. Regarding the Jacobian matrix of F, the analytical expression was used for all test problems. Specifically for **Problem 6.1** and **Problem 6.2** the exact Jacobian matrices are given in (54) and (52), for **Problem 6.3** the exact analytical formulae is developed in [11] and for **Problem 6.4** it is easily computable.

To compute the square root of matrix B_k the singular value decomposition of the Jacobian, provided by MATLAB function `svd`, is used.

In case of noisy problems, given the error level δ, the exact data y was perturbed by normally distributed values with mean 0 and variance δ^2 using the MATLAB function `randn`.

To compute the KKT point (z_k, λ_k) at Step 2 (16) has to be solved. Since the trust-region is ensured to be active, this can be accomplished solving the following nonlinear scalar equation: $\psi(\lambda) = \frac{1}{\|z(\lambda)\|} - \frac{1}{\Delta_k}$, [22, §7.3]. Starting from an initial guess greater than the sought solution λ_k the sequence generated converges monotonically to λ_k. Typically high accuracy in the solution of the above scalar equations is not needed, hence the Newton process is terminated as soon as the absolute value of function ψ is below 10^{-2}. Each Newton iteration applied to $\psi(\lambda) = 0$ requires the solution of a linear system with shifted matrix of the form $B_k^2 + \lambda I$. Taking into account that λ_k is always bounded away from zero, as the trust-region is always active, a sequence of regularized linear systems has to be solved, and this allows to overcome the ill-conditioning of B_k^2, since in these problems the ill-conditioning of the Jacobian is due to the presence of small singular values. The linear systems are solved employing the singular value decomposition of the Jacobian, computed to obtain the square root of matrix B_k.

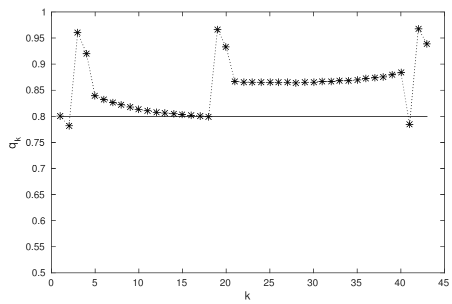
Algorithms 3.1 and 5.1 were run setting $\eta = 10^{-1}$. In Step 1 the trust-region radius was updated as follows

$$\Delta_0 = \mu_0 \|B_0^{1/2} g_0\| \qquad \mu_0 = 10^{-1}$$

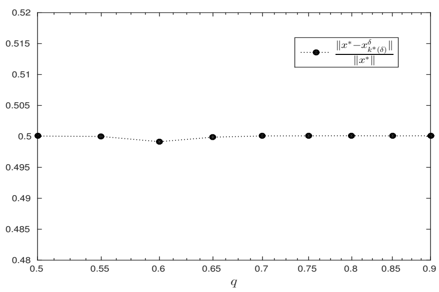
$$\Delta_{k+1} = \mu_{k+1} \|B_{k+1}^{1/2} g_{k+1}\|, \qquad \mu_{k+1} = \begin{cases} \frac{1}{6}\mu_k, & \text{if } q_k < q \text{ or } \rho_k < \eta_2, \\ 2\mu_k, & \text{if } q_k > \nu q \text{ and } \rho_k > \eta_2, \\ \mu_k, & \text{otherwise,} \end{cases}$$

with $q_k = \frac{\|B_k p_k + g_k\|}{\|g_k\|}$, $\nu = 1.1$ and $\eta_2 = 0.25$. The maximum and minimum values for Δ_k were set to $\Delta_{\max} = 10^4$ and $\Delta_{\min} = 10^{-12}$ and the maximum value for μ_k was set to 10^5. This updating strategy is based on the following considerations. The choice of Δ_k suggested by (33) guarantees that the q-condition (28) is satisfied, but gives rise to small values of the trust-region radius and as a consequence to a slow procedure. On the other hand, Δ_k given by the procedure described above preserves the property of converging to zero in case of exact data, as $\|B_k^{1/2} g_k\|$ tends to zero. Further, Δ_k is adjusted taking into account the q-condition by monitoring the value q_k, as it is satisfied whenever $q_k \geq q$. Therefore, if the q-condition was not satisfied at the last computed iterate x_k^δ, it is reasonable to take a smaller radius than in the case where it was fulfilled. This updating strategy turned out to be efficient in practice. As an example in Figure 3 (a) the obtained values of q_k for **Problem 6.1** for $\delta = 10^{-2}$ are reported. For almost all the iterations the values q_k are greater than the chosen value $q = 0.8$, marked by the horizontal solid line, so the q-condition is fulfilled.

The free parameter q was set equal to 0.8, but this choice is not critical. Actually, the behaviour of the procedure does not seem to be deeply affected by the value of q. As an example in Figure 3 (b) it is shown for **Problem 6.4** the value of the relative error $\frac{\|x^* - x_{k^*(\delta)}^\delta\|}{\|x^*\|}$ between the solution x^* approached with exact data and that computed with $\delta = 10^{-2}$, for different values of parameter q. Note that the error does not vary significantly with q.



(a)



(b)

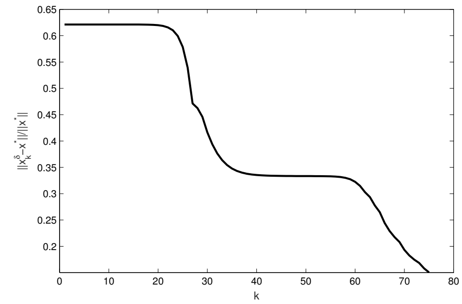
Fig. 3 (a): **Problem 6.1**, $\delta = 10^{-2}$. Obtained values of q_k. (b): **Problem 6.4**, $\delta = 10^{-2}$. Dependence of the relative error between computed solution $x_{k^*(\delta)}^\delta$ and x^* on the parameter q.

The scalar τ in the discrepancy principle (37) was chosen adaptively as
$\tau_k = \bar{\tau}\|J(x_k^\delta)\|$, with $\bar{\tau} = 0.1$. The value of $\bar{\tau}$ is not in agreement with Assumption 4.4, but in practice σ is not known and the numerical tests provide an evidence of the effectiveness of this stopping rule. As τ_k depends on k, the stopping rule changes at each iteration. However, τ_k varies only slightly along the iterations as $\|J(x_k^\delta)\|$ is almost constant. Values of τ_k used in the tests, for $\delta = 10^{-2}$, are as follows:

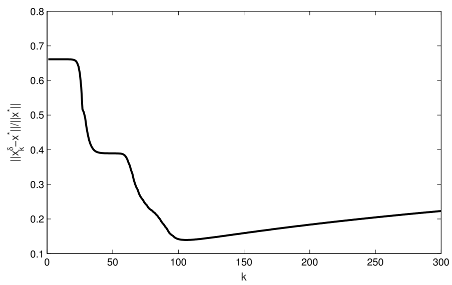
– **Problem 6.1**: $\|J(x_k^\delta)\| \simeq 0.04$ and $\tau_k \simeq 4 \cdot 10^{-3}$,
– **Problem 6.2**: $\|J(x_k^\delta)\| \simeq 10^{-2}$ and $\tau_k \simeq 10^{-3}$,

– **Problem 6.3**: $\|J(x_k^\delta)\| \simeq 10^{-1}$ and $\tau_k \simeq 10^{-2}$,
– **Problem 6.4**: $\|J(x_k^\delta)\| \simeq 5$ and $\tau_k \simeq 0.5$.

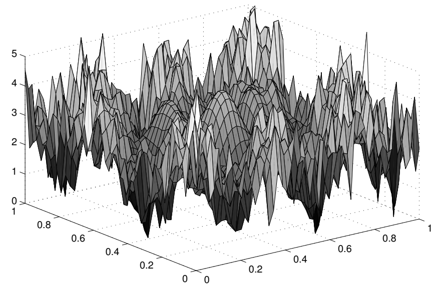
In agreement with the theory, the error is monotonically decreasing as long as the discrepancy principle is not satisfied, as it is shown for example in Figure 4 (a), in which the decrease of the relative error $\frac{\|x^*-x_k^\delta\|}{\|x^*\|}$ between the solution approached with exact data and the current iterate x_k^δ is reported, varying k, for **Problem 6.2** with $\delta = 10^{-2}$. Iterating further is not useful and it can also lead to a new increase in the norm of the error, as it is shown at the bottom of Figure 4. The procedure was not stopped when the discrepancy principle was satisfied and the error started to increase (b), as the sequence approaches a solution of the noisy problem (c).



(a)



(b)



(c)

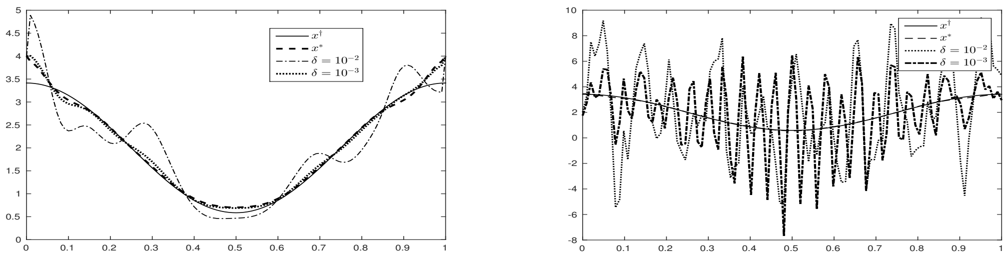
Fig. 4 **Problem 6.2**, $\delta = 10^{-2}$. Upper part: The procedure is stopped when the discrepancy principle is satisfied, the norm of the error decreases monotonically (a). Lower part: The procedure is not stopped when the discrepancy principle is satisfied, the norm of the relative error increases (b) and a solution of the noisy problem is approached (c).

The stopping criterion alone is not sufficient to obtain a regularizing method. To show this, a standard trust-region procedure was implemented, according to technicalities well-known in the literature, [18, §4], [22, §6.1]. Particularly, it was set $\Delta_0 = 1$ and

$$\Delta_{k+1} = \begin{cases} \dfrac{\|p_k\|}{4}, & \text{if } \pi_k(p_k) < \dfrac{1}{4}, \\ \Delta_k, & \text{if } \dfrac{1}{4} \le \pi_k(p_k) \le \dfrac{3}{4}, \\ \min\{2\Delta_k, \Delta_{\max}\}, & \text{otherwise,} \end{cases}$$

This procedure was used also to solve the test problems in case of noisy data. For **Problem 6.1**, **Problem 6.2** and **Problem 6.3** in Figures 5, 6 and 8 it is shown that, while the regularizing trust-region manages to handle the noise in the data, the sequence generated by the standard trust-region converges to a solution of the noisy problem. For **Problem 6.4** the solution found by the standard trust-region is as accurate as the solution found by the regularizing trust-region. This good result is due to the fact that the trust-region happens to be active in all iterations and therefore a regularizing behaviour is implicitly provided.

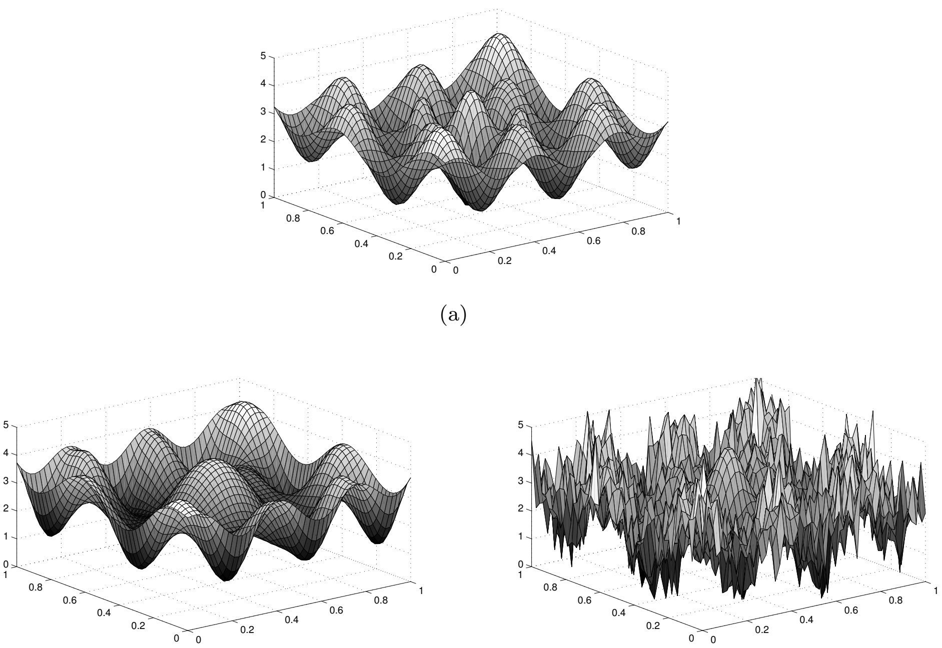
In Figure 5, **Problem 6.1** is considered. The solution approximations computed with $\delta = 0$ (x^*, dashed line), $\delta = 10^{-3}$ (dotted line) and $\delta = 10^{-2}$ (dash-dotted line), are compared to $x^\dagger = [c(x_1), \ldots, c(x_N)]^T$, with c in (49) (solid line). The solution approximation improves with decreasing noise for the regularizing trust-region (a) and not for the standard trust-region (b), and that x^*, which is computed with exact data, is a good approximation to $x^\dagger$.



(a) (b)

Fig. 5 **Problem 6.1**. Comparison of solution approximations computed with noise levels $\delta = 10^{-3}, 10^{-2}$ and solutions x^* and $x^\dagger = [c(x_1), \ldots, c(x_N)]^T$ with c in (49), (a): regularizing trust-region, (b): standard trust-region.

In Figure 6, **Problem 6.2** is considered. The solution x^* approached with exact data (a) is compared to the solution approximations computed by the regularizing trust-region (b) and the standard trust-region (c) with $\delta = 10^{-2}$.



(a)

(b) (c)

Fig. 6 **Problem 6.2**. Upper part: solution approximation x^* obtained with exact data (a). Lower part: solution approximations obtained with regularizing trust-region (b) and standard trust-region (c) for $\delta = 10^{-2}$.

In Figure 7, we analyse the behaviour of the regularizing trust-region method depending on the noise level. We report the contour plots of the computed solution approximations for **Problem 6.2** with $\delta = 0, 10^{-2}, 10^{-4}$. As the noise level decreases, according to the theory, the computed solution approaches a minimum of the problem. In fact, as δ decreases, the contour plots become more and more similar to that obtained for $\delta = 0$.

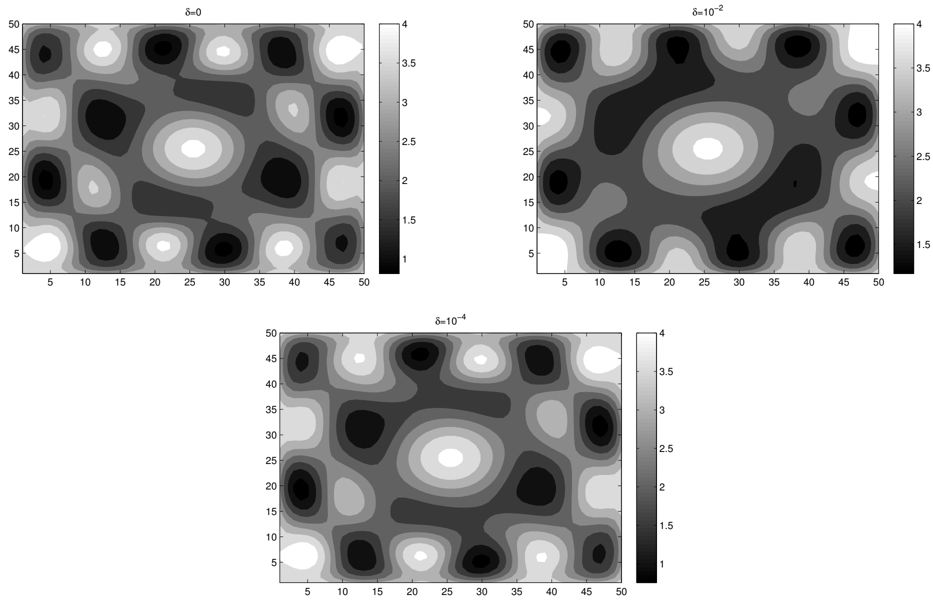
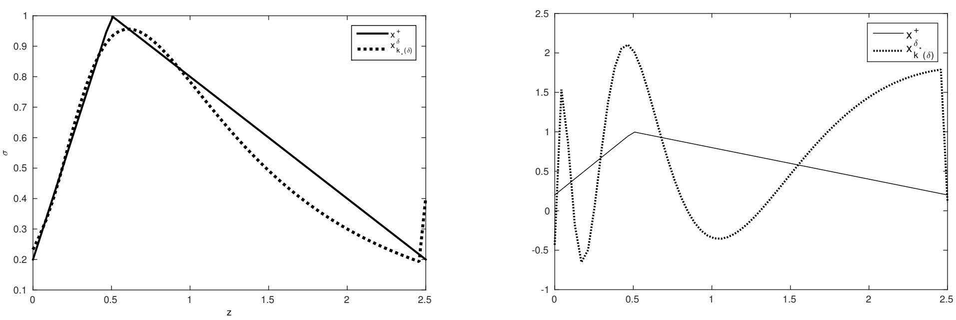


Fig. 7 Contour plots of solution approximations of **Problem 6.2** for different noise levels, $\delta = 0, 10^{-2}, 10^{-4}$.



(a) (b)

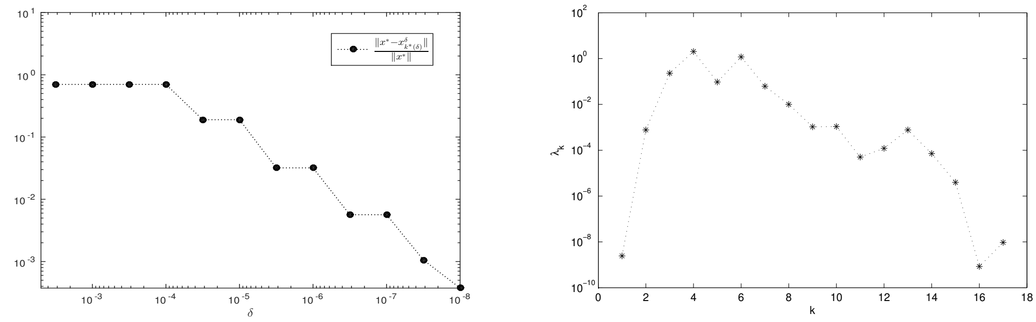
Fig. 8 **Problem 6.3** with $\delta = 10^{-2}$: plot of the true solution $x^\dagger$ and of the computed solution $x^\delta_{k^*(\delta)}$ for regularizing trust-region (a), standard trust-region (b).

In Figure 8, **Problem 6.3** with $\delta = 10^{-2}$ is considered. We report the plots of the true solution $x^\dagger$ and of the computed solution $x^\delta_{k^*(\delta)}$ for $\delta = 10^{-2}$, for the regularizing trust-region (a) and for the standard trust-region (b). Also in this case, the solution $x^\dagger$ has been well-approximated by the regularizing trust-region, while the standard trust-region is not able to provide a good approximation.

Finally, in Figure 9 numerical evidence of properties theoretically proved is reported. In (a), for **Problem 6.4**, the reduction of the relative error $\frac{\|x^* - x^\delta_{k^*(\delta)}\|}{\|x^*\|}$ with the noise level δ is shown. In (b) the regularization parameters λ_k for **Problem 6.3**, $\delta = 10^{-2}$ are plotted. In accordance with the theory, the regularization parameters are strictly positive and bounded above.

7 Conclusions

A non-stationary iterated Tikhonov procedure to solve nonlinear ill-posed least-squares problems with small residual has been proposed. A trust-region reformulation of the method has been proposed, that allows us to set the regularization parameters λ_k in an automatic way. Along with a suitable choice of matrix M_k, this ensures regularizing properties of the method and gives rise to a procedure able to find a stable approximation of a solution of the unperturbed problem,



(a) (b)

Fig. 9 (a): **Problem 6.4**. Reduction of the relative error $\frac{\|x^*-x^\delta_{k^*(\delta)}\|}{\|x^*\|}$ with the noise level δ. (b): **Problem 6.3**, $\delta = 10^{-2}$. Plot of the regularization parameters λ_k.

even in case of noisy data. A theoretical analysis and a reliable implementation of the proposed method, that has been validated on different tests, are provided. The obtained numerical results highlight the effectiveness of the procedure and its regularizing properties. However, the current implementation is not suitable for large scale problems, as the square root of the matrix B_k is computed by the svd decomposition of the current Jacobian. In order to develop a reliable method for large scale problems both the action of the square root of B_k on a vector and the solution of the linear systems should be approximated with a rough accuracy by using an iterative method, such as the Lanczos method. The Krylov subspace generated by the Lanczos method could be used both to compute the right-hand side of the linear systems, and to solve them, projecting the problem on the generated subspace. However, this introduces a source of inexactness, that is not taken into account in the presented theory and will be subject of further study.

Appendix

Proof of item (ii) of Lemma 4.2 and item (ii) of Lemma 4.4

Proof The proof is the same for the noise free and the noisy case, for generality the notation of the noisy case is employed.

Since $\lambda_k > 0$ from Lemma 3.4, the trust-region is active, and from (30) it follows that

$$\Delta_k = \|z(\lambda_k)\| \leq \frac{\|B_k^{1/2} f_k'\|}{\lambda_k}.$$

Then, if Δ_k chosen at Step 1 of Algorithm 3.1 guarantees condition $\pi_k(p_k) \geq \eta$, the thesis follows as

$$\lambda_k \leq \frac{\|B_k^{1/2} f_k'\|}{\Delta_k} \leq \frac{1}{C_{\min}} = \bar{\lambda}. \quad (56)$$

Otherwise, the trust-region radius is progressively reduced, and a bound for the value of Δ_k at termination of Step 2 of Algorithm 3.1 can be provided. First, consider the case

$$f_\delta(x_k^\delta + p_k) > \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\|^2.$$

Trivially,

$$1 - \pi_k(p_k) = \frac{f_\delta(x_k^\delta + p_k) - \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\|^2}{f_\delta(x_k^\delta) - \frac{1}{2}\|F(x_k^\delta) - y^\delta + F'(x_k^\delta)p_k\|^2},$$

and

$$\begin{aligned} f_\delta(x_k^\delta+p_k)-\frac{1}{2}\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p_k\|^2 &= \frac{1}{2}\|F(x_k^\delta+p_k)\pm F(x_k^\delta)\pm F'(x_k^\delta)p_k-y^\delta\|^2 \\ &-\frac{1}{2}\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p_k\|^2 \\ &= \frac{1}{2}\|F(x_k^\delta+p_k)-F(x_k^\delta)-F'(x_k^\delta)p_k\|^2 \\ &+\|F(x_k^\delta+p_k)-F(x_k^\delta)-F'(x_k^\delta)p_k\| \\ &\cdot\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p_k\|. \end{aligned}$$

By the Lipschitz continuity of F' it holds

$$\|F(x_k^\delta+p_k)-F(x_k^\delta)-F'(x_k^\delta)p_k\|\leq\frac{L}{2}\|p_k\|^2.$$

Moreover, using (24)

$$\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p(\lambda)\|<\|F(x_k^\delta)-y^\delta\|$$

for any $\lambda\geq 0$. Consequently, as $\|p_k\|\leq\|B_k^{1/2}\|\Delta_k$ and $\Delta_k\leq C_{\max}\|B_k^{1/2}f_k'\|$,

$$\begin{aligned} f_\delta(x_k^\delta+p_k)-&\frac{1}{2}\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p_k\|^2\leq \\ &\frac{L}{2}K^2\Delta_k^2\|F(x_0)-y\|\left(\frac{L}{4}K^6C_{\max}^2\|F(x_0)-y\|+1\right). \end{aligned}$$

From [22, Theorem 6.3.1 and §8.3] it holds

$$f_\delta(x_k^\delta)-\left(\frac{1}{2}\langle z_k,B_k^2z_k\rangle+\langle B_k^{1/2}f_k',z_k\rangle+f_\delta(x_k^\delta)\right)\geq\frac{1}{2}\|B_k^{1/2}f_k'\|\min\left\{\Delta_k,\frac{\|B_k^{1/2}f_k'\|}{\|B_k^2\|}\right\}.$$

Then, (18) yields

$$f_\delta(x_k^\delta)-\frac{1}{2}\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p_k\|^2\geq\frac{1}{2}\Delta_k\|B_k^{1/2}f_k'\|,$$

whenever $\Delta_k\leq\dfrac{\|B_k^{1/2}f_k'\|}{K^4}$, and this implies

$$1-\pi_k(p_k)\leq\frac{LK^2\Delta_k\|F(x_0)-y\|(\frac{1}{4}LK^6C_{\max}^2\|F(x_0)-y\|+1)}{\|B_k^{1/2}f_k'\|}.$$

Namely, termination of the repeat loop occurs with

$$\Delta_k\leq\omega\|B_k^{1/2}f_k'\|,$$

and

$$\omega=\min\left\{\frac{1}{K^4},\frac{1-\eta}{LK^2\|F(x_0)-y\|(\frac{1}{4}LK^6C_{\max}^2\|F(x_0)-y\|+1)}\right\}.$$

Taking into account Step 1 and the updating rule at Step 2.4, it can be concluded that, at termination of Step 2, the trust-region radius Δ_k satisfies

$$\Delta_k\geq\min\{C_{\min},\gamma\omega\}\|B_k^{1/2}f_k'\|.$$

In fact, in case of a smaller value of Δ_k, it happens $f_\delta(x_k^\delta+p_k)\leq\frac{1}{2}\|F(x_k^\delta)-y^\delta+F'(x_k^\delta)p_k\|^2$, then $\pi_k(p_k)\geq 1>\eta$ and the loop at Step 2 terminates. Then, it terminates for a trust-region radius greater than or equal to the one estimated above.

Then, $\lambda_k\leq\bar{\lambda}$ as

$$\lambda_k\leq\frac{\|B_k^{1/2}f_k'\|}{\Delta_k}\leq\max\left\{\frac{1}{\gamma\omega},\frac{1}{C_{\min}}\right\},$$

and the thesis follows. □

Proof of Theorem 4.3 *Proof* Summing up from $\bar{k}$ to $k^*(\delta)-1$, by (37), (28), (10) and Lemma 4.4, it follows

$$(k^*(\delta)-\bar{k})\tau^2\delta^2\leq\sum_{k=\bar{k}}^{k^*(\delta)-1}\|F'(x_k^\delta)^*(F(x_k^\delta)-y^\delta)\|^2\leq\frac{\theta_{\bar{k}}\bar{\lambda}}{2(\theta_{\bar{k}}-1)q^2}\|x_{\bar{k}}^\delta-x^\dagger\|^2.$$

Thus, $k^*(\delta)$ is finite for $\delta>0$.

Convergence of $x_{k^*(\delta)}^\delta$ to a stationary point of (1) as δ tends to zero is obtained by adapting the proof of [5, Theorem 4.5]. Specifically, let x^* be the limit of the sequence $\{x_k\}$ corresponding to the exact data y and let $\{\delta_n\}$ be a sequence of values of δ converging to zero as $n \to \infty$. Denote by y^{δ_n} a corresponding sequence of perturbed data, and by $k_n = k_*(\delta_n)$ the stopping index determined from the discrepancy principle (37) applied with $\delta = \delta_n$. Assume first that $\tilde{k}$ is a finite accumulation point of $\{k_n\}$. Without loss of generality, possibly considering a subsequence, it can be assumed that $k_n = \tilde{k}$ for all $n \in \mathbb{N}$. Thus, from the definition of k_n, it follows that

$$\|F'(x_{\tilde{k}}^{\delta_n})^*(y^{\delta_n} - F(x_{\tilde{k}}^{\delta_n}))\| \leq \tau\delta_n. \tag{57}$$

As, by assumption, $\pi_k(x_{k+1} - x_k) \neq \eta$, for all k, it follows that for the fixed index $\tilde{k}$, the iterate $x_{\tilde{k}}^{\delta}$ depends continuously on δ. Then

$$x_{\tilde{k}}^{\delta_n} \to x_{\tilde{k}}, \qquad F'(x_{\tilde{k}}^{\delta_n}) \to F'(x_{\tilde{k}}), \qquad F(x_{\tilde{k}}^{\delta_n}) \to F(x_{\tilde{k}}) \qquad \text{as } \delta_n \to 0.$$

Therefore, by (57), it follows that $F'(x_{\tilde{k}})^*(y - F(x_{\tilde{k}})) = 0$, and the $\tilde{k}$-th iterate with exact data y is a stationary point of (1), i.e. $x^* = x_{\tilde{k}}$, and it is possible to conclude that $x_{k_n}^{\delta_n} \to x^*$ as $\delta_n \to 0$.

It remains to consider the case where $k_n \to \infty$ as $n \to \infty$. As $\{x_k\}$ converges to a stationary point x^* of (1) by Theorem 4.2, there exists $\tilde{k} > 0$ such that

$$\|x_k - x^*\| \leq \frac{1}{2}\bar{\rho} \qquad \text{for all} \qquad k \geq \tilde{k},$$

where $\bar{\rho} < \min\left\{\dfrac{(q-\sigma)\tau - K(\sigma+1)}{c(K+\tau)}, \rho\right\}$. Then, as x_k^{δ} depends continuously on δ, δ_n tends to zero and $k_*(\delta_n) \to \infty$, there exists δ_n sufficiently small such that $\tilde{k} \leq k_*(\delta_n)$, and

$$\|x_{\tilde{k}}^{\delta_n} - x_{\tilde{k}}\| \leq \frac{1}{2}\bar{\rho}.$$

Then, for δ_n sufficiently small

$$\|x_{\tilde{k}}^{\delta_n} - x^*\| \leq \|x_{\tilde{k}}^{\delta_n} - x_{\tilde{k}}\| + \|x_{\tilde{k}} - x^*\| \leq \bar{\rho}. \tag{58}$$

Now, from item (i) of Lemma 4.4, it follows $x_{\tilde{k}}^{\delta_n} \in \mathcal{B}_{2\rho}(x_{\tilde{k}}^{\delta_n})$, while from (46) and Theorem 4.2 it holds $x^* \in \mathcal{B}_{2\rho}(x_{\tilde{k}}^{\delta_n})$ as

$$\|x_{\tilde{k}}^{\delta_n} - x^*\| \leq \|x_{\tilde{k}}^{\delta_n} - x^\dagger\| + \|x^\dagger - x^*\| \leq 2\rho.$$

Letting $e_{\tilde{k}}^* = x^* - x_{\tilde{k}}^{\delta_n}$. Repeating arguments from Lemma 4.3, and using (38), (2) it follows

$$\begin{aligned}
\|m_{\tilde{k}}(e_{\tilde{k}}^*)\| &\leq K\delta_n + \|F'(x_{\tilde{k}}^{\delta_n})^*(y - F(x_{\tilde{k}}^{\delta_n}) + F'(x_{\tilde{k}}^{\delta_n})(x^* - x_{\tilde{k}}^{\delta_n}))\| \\
&\leq K\delta_n + (c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma)\,\|F'(x_{\tilde{k}}^{\delta_n})^*(y - F(x_{\tilde{k}}^{\delta_n}))\| \\
&\leq (1 + c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma)K\delta_n + (c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma)\,\|F'(x_{\tilde{k}}^{\delta_n})^*(y^{\delta_n} - F(x_{\tilde{k}}^{\delta_n}))\|.
\end{aligned}$$

Then, at iteration $\tilde{k}$, conditions (37) and (28) give

$$\begin{aligned}
\|m_{\tilde{k}}(e_{\tilde{k}}^*)\| &\leq \left(K\frac{1 + c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma}{\tau} + (c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma)\right)\|F'(x_{\tilde{k}}^{\delta_n})^*(y^{\delta_n} - F(x_{\tilde{k}}^{\delta_n}))\| \\
&\leq \left(K\frac{1 + c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma}{q\tau} + \frac{c\|x^* - x_{\tilde{k}}^{\delta_n}\| + \sigma}{q}\right)\|m_{\tilde{k}}(p_{\tilde{k}})\|.
\end{aligned}$$

Thus, by (58) and $\bar{\rho} < \min\left\{\dfrac{(q-\sigma)\tau - K(\sigma+1)}{c(K+\tau)}, \rho\right\}$, it follows that

$$\|m_{\tilde{k}}(e_{\tilde{k}}^*)\| \leq \frac{1}{\theta_{\tilde{k}}}\|m_{\tilde{k}}(p_{\tilde{k}})\|$$

is satisfied with $\theta_{\tilde{k}} = \dfrac{q\tau}{1 + c(1+\tau)\bar{\rho} + \sigma(1+\tau)} > 1$. Replacing $x^\dagger$ with x^*, (10) gives $\|x_{\tilde{k}+1}^{\delta_n} - x^*\| < \|x_{\tilde{k}}^{\delta_n} - x^*\|$ and repeating the above arguments, by induction monotonicity of the error $\|x_k^{\delta_n} - x^*\|$ for $\tilde{k} \leq k \leq k_n$ is obtained. Then

$$\|x_{k_n}^{\delta_n} - x^*\| < \|x_{\tilde{k}}^{\delta_n} - x^*\| \leq \bar{\rho}. \tag{59}$$

Finally, repeating the previous arguments, it can be shown that, for every $0 \leq \epsilon \leq \bar{\rho}$, it exists $\bar{\delta}_\epsilon$ such that $\|x_{k_n}^{\delta_n} - x^*\| \leq \epsilon$ for all $\delta_n \leq \bar{\delta}_\epsilon$, i.e.

$$x_{k_n}^{\delta_n} \to x^* \qquad \text{as} \qquad \delta_n \to 0,$$

and the thesis is proved. □

Acknowledgements The authors thank Gian Piero Deidda, Caterina Fenu and Giuseppe Rodriguez for providing us the Matlab code for **Problem 6.3**. The authors also thank Sergio Vessella and Elisa Francini for helpful discussion on this manuscript. Work supported by *Gruppo Nazionale per il Calcolo Scientifico* (GNCS-INdAM) of Italy.

References

1. Kaltenbacher, B., Neubauer, A., Scherzer, O.: Iterative regularization methods for nonlinear ill-posed problems, vol. 6. Walter de Gruyter (2008)
2. Moré, J., Sorensen, D.: Computing a trust region step. SIAM Journal on Scientific and Statistical Computing **4**(3), 553–572 (1983)
3. Donatelli, M., Hanke, M.: Fast nonstationary preconditioned iterative methods for ill-posed problems, with application to image deblurring. Inverse Problems **29**(9), 095008 (2013)
4. Buccini, A.: Regularizing preconditioners by non-stationary iterated Tikhonov with general penalty term. Applied Numerical Mathematics **116**, 64 – 81 (2017)
5. Bellavia, S., Morini, B., Riccietti, E.: On an adaptive regularization for ill-posed nonlinear systems and its trust-region implementation. Computational Optimization and Applications **64**(1), 1–30 (2016)
6. Hanke, M.: A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse problems **13**(1), 79 (1997)
7. Wang, Y., Yuan, Y.: On the regularity of trust region-cg algorithm for nonlinear ill-posed inverse problems with application to image deconvolution problem. Science in China Ser.A **46**, 312–325 (2003)
8. Wang, Y., Yuan, Y.: Convergence and regularity of trust region methods for nonlinear ill-posed problems. Inverse Problems **21**, 821–838 (2005)
9. Banks, H., Murphy, K.: Estimation of coefficients and boundary parameters in hyperbolic systems. SIAM journal on control and optimization **24**(5), 926–950 (1986)
10. Binder, A., Engl, H., Neubauer, A., Scherzer, O., Groetsch, C.: Weakly closed nonlinear operators and parameter identification in parabolic equations by Tikhonov regularization. Applicable Analysis **55**(3-4), 215–234 (1994)
11. Deidda, G., Fenu, C., Rodriguez, G.: Regularized solution of a nonlinear problem in electromagnetic sounding. Inverse Problems **30**(12), 125014 (2014)
12. Henn, S.: A Levenberg–Marquardt scheme for nonlinear image registration. BIT Numerical Mathematics **43**(4), 743–759 (2003)
13. Tang, L.: A regularization homotopy iterative method for ill-posed nonlinear least squares problem and its application. In: Advances in Civil Engineering, ICCET 2011, *Applied Mechanics and Materials*, vol. 90, pp. 3268–3273. Trans Tech Publications (2011)
14. Lopez, D., Barz, T., Körkel, S., Wozny, G.: Nonlinear ill-posed problem analysis in model-based parameter estimation and experimental design. Computers & Chemical Engineering **77**(Supplement C), 24 – 42 (2015)
15. Landi, G., Piccolomini, E.L., Nagy, J.G.: A limited memory BFGS method for a nonlinear inverse problem in digital breast tomosynthesis. Inverse Problems **33**(9), 095005 (2017)
16. Cornelio, A.: Regularized nonlinear least squares methods for hit position reconstruction in small gamma cameras. Applied Mathematics and Computation **217**(12), 5589 – 5595 (2011)
17. Neubauer, A.: An a posteriori parameter choice for Tikhonov regularization in the presence of modeling error. Applied Numerical Mathematics **4**(6), 507 – 519 (1988)
18. Nocedal, J., Wright, S.: Numerical optimization. Springer Science & Business Media (2006)
19. Buccini, A., Donatelli, M., Reichel, L.: Iterated Tikhonov regularization with a general penalty term. Numerical Linear Algebra with Applications **24**(4), e2089 (2017)
20. Higham, N.J.: Functions of matrices: theory and computation, vol. 104. Siam (2008)
21. Dennis, J., Schnabel, R.: Numerical methods for unconstrained optimization and nonlinear equations. Siam (1996)
22. Conn, A., Gould, N., Toint, P.: Trust region methods, vol. 1. Siam (2000)
23. Allaire, G.: Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation. Oxford University Press (2007)
24. Engl, H., Hanke, M., Neubauer, A.: Regularization of inverse problems. Kluwer Academic Publishers Group (1996)
25. Hanke, M.: Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. Numerical Functional Analysis and Optimization **18**(9-10), 971–993 (1997)
26. Kunisch, K., White, L.: Parameter estimation, regularity and the penalty method for a class of two point boundary value problems. SIAM Journal on Control and Optimization **25**(1), 100–120 (1987)
27. Rieder, A.: On the regularization of nonlinear ill-posed problems via inexact Newton iterations. Inverse Problems **15**(1), 309 (1999)
28. Scherzer, O., Engl, H.W., Kunisch, K.: Optimal a posteriori parameter choice for Tikhonov regularization for solving nonlinear ill-posed problems. SIAM Journal on Numerical Analysis **30**(6), 1796–1838 (1993)
29. Riccietti, E.: Levenberg-Marquardt methods for the solution of noisy nonlinear least squares problems, PhD Thesis, University of Florence. http://web.math.unifi.it/users/riccietti/publications.html (2018)