Comput Optim Appl (2016) 64:1–30
DOI 10.1007/s10589-015-9806-9

On an adaptive regularization for ill-posed nonlinear systems and its trust-region implementation

S. Bellavia[1] · B. Morini[1] · E. Riccietti[2]

Received: 2 April 2015 / Published online: 24 November 2015

Abstract In this paper we address the stable numerical solution of nonlinear ill-posed systems by a trust-region method. We show that an appropriate choice of the trust-region radius gives rise to a procedure that has the potential to approach a solution of the unperturbed system. This regularizing property is shown theoretically and validated numerically.

Keywords Ill-posed systems of nonlinear equations · Regularization · Nonlinear stepsize control · Trust-region methods

1 Introduction

Nonlinear systems modeling inverse problems are typically ill-posed, in the sense that their solutions do not depend continuously on the data and their data are affected by noise [6,16,26]. In this work we focus on the stable approximation of a solution of these problems. Procedures in the classes of Levenberg–Marquardt and trust-region methods are discussed, and a suitable version of trust-region algorithm is shown to have

✉ S. Bellavia
stefania.bellavia@unifi.it

B. Morini
benedetta.morini@unifi.it

E. Riccietti
elisa.riccietti@unifi.it

1 Dipartimento di Ingegneria Industriale, Università di Firenze, viale G.B. Morgagni 40, 50134 Firenze, Italy

2 Dipartimento di Matematica e Informatica "Ulisse Dini", Università di Firenze, viale G.B. Morgagni 67a, 50134 Firenze, Italy

regularizing properties both theoretically and numerically. The underlying motivation for our study is twofold: most of the practical methods in the literature have been designed for well-posed systems, see e.g., [5,23], and thus are unsuited in the context of inverse problems; adaptation of existing procedures for handling ill-posed problems, carried out in the seminal papers [10,12,13,15,25,27], deserves further theoretical and numerical insights.

Let

$$F(x) = y, \tag{1.1}$$

with $F : \mathbb{R}^n \to \mathbb{R}^n$ continuously differentiable, be obtained from the discretization of a problem modeling an inverse problem. It is realistic to have noisy data y^δ at disposal, satisfying

$$\|y - y^\delta\|_2 \le \delta, \tag{1.2}$$

for some positive δ. Thus, in practice it is necessary to solve a problem of the form

$$F(x) = y^\delta, \tag{1.3}$$

and, due to ill-posedeness, possible solutions may be arbitrarily far from those of the original problem. To approximate solutions of the unperturbed problem (1.1), iterative *regularizing* methods can be applied where both the construction of the iterates x_k^δ and the stopping criterion act as a regularization, see e.g., [16]. Such methods are expected to have the following properties: if iterations are stopped at index $k^*(\delta)$, then

- $x^\delta_{k^*(\delta)}$ is an approximation to a solution of (1.1);
- $x^\delta_{k^*(\delta)}$ converges to a solution of (1.1) as δ tends to zero;
- in the noise-free case, convergence to a solution of (1.1) occurs.

These properties are supposed to hold even if there are no finite bounds on the inverse of the Jacobian of F around a solution of (1.1).

In [12,13], Hanke supposed that an initial guess, close enough to some solution $x^\dagger$ of (1.1), is available. Then, he proposed a regularizing Levenberg–Marquardt procedure which is able to compute a stable approximation $x^\delta_{k_*(\delta)}$ to $x^\dagger$ or to some other solution of the unperturbed problem (1.1) close to $x^\dagger$. This task is achieved through an implicit stepsize control in the Levenberg–Marquardt procedure and the discrepancy principle as the stopping criterion, so that the iterative process is stopped at the iteration $k_*(\delta)$ satisfying

$$\left\| y^\delta - F\left(x^\delta_{k_*(\delta)}\right)\right\|_2 \le \tau\delta < \left\|y^\delta - F\left(x_k^\delta\right)\right\|_2, \quad 0 \le k < k_*(\delta), \tag{1.4}$$

with $\tau > 1$ appropriately chosen [22]. Remarkably the procedure satisfies the regularizing properties listed above and local convergence properties are established under conditions weaker than the so-called local *error-bound condition* used in the literature when the Jacobian J of F is singular at the solution approached, see e.g. [1,3,17].

Further regularizing iterative methods have been proposed, including first-order methods and Newton-type methods. Analogously to the Levenberg–Marquardt procedure proposed by Hanke, instead of promoting convergence to a solution of (1.3), they form approximations of increasing accuracy to some solution of the unperturbed problem (1.1) until the discrepancy principle (1.4) is met. We refer to [6,16] for the description and analysis of such methods.

The above mentioned regularizing Levenberg–Marquardt method belongs to the unifying framework of nonlinear stepsize control algorithms for unconstrained optimization developed by Toint [24] and including trust-region methods [5]. Therefore, elaborating on original ideas by Hanke, we introduce and analyze a regularizing variant of the trust-region method based on a specific rule for selecting the trust-region radius. The resulting method shares the same regularizing properties as the method by Hanke and, as for standard trust-region procedures, it enforces a monotonic decrease of the value of the function

$$\Phi(x) = \frac{1}{2}\|y^\delta - F(x)\|_2^2, \tag{1.5}$$

at the iterates x_k^δ. Convergence properties are enhanced with respect to the regularizing Levenberg–Marquardt procedure in the following respects. With exact data, if there exists an accumulation point of the iterates which solves (1.1), then any accumulation point of the sequence solves (1.1). With noisy data, the method has the potential to satisfy the discrepancy principle (1.4). As for standard trust-region methods, these properties can be enhanced independently of the closeness of the initial guess to a solution of (1.1).

Our contribution covers theoretical and practical aspects of the method proposed. From a theoretical point of view, we propose the use of a trust-region radius converging to zero as δ tends to zero. Trust-region methods with this distinguishing feature have been proposed in several papers, see [7–9,29], but none of such works was either devised for ill-posed problems or applied to them; thus, our study offers new insights on the potential of this choice for the trust-region radius. Moreover, we have made an attempt toward globally convergent methods for ill-posed problems; to our knowledge, this topic has been considered only in a multilevel approach proposed by Kaltenbacher [15]. Finally, local convergence analysis has been carried out without making two common assumptions in the literature: neither the invertibility of the Jacobian J of F and boundness of the inverse, nor the fulfillment of the local error-bound condition (see e.g., [7–9,19,29]) have been used. In fact, such conditions may not be satisfied in the presence of ill-posedeness. Therefore, our results may represent a progress in the theoretical investigation of convergence. Concerning numerical aspects, we discuss an implementation of the regularizing trust-region method, and test its ability to approximate a solution of (1.1) in the presence of noise. Comparison with a standard trust-region scheme highlights the impact of the proposed trust-region radius choice on regularization.

The paper is organized as follows. In Sect. 2 we describe the main features of the regularizing Levenberg–Marquardt method proposed by Hanke. In Sect. 3 we introduce our regularizing version of trust-region methods and in Sect. 4 we study the local convergence properties. A comparative numerical analysis of all the procedures studied is done in Sect. 5.

Notations We indicate the iterates of the procedures analyzed as x_k^δ; if the data are exact, x_k may be used as an alternative to x_k^δ. By $x_0^\delta = x_0$ we denote an initial guess which may incorporate a-priori knowledge of an exact solution. The symbol $\|\cdot\|$ indicates the Euclidean norm. A closed ball of radius ρ around a vector x is denoted as $B_\rho(x)$. The Jacobian matrix of F is denoted as J.

2 Regularizing Levenberg–Marquardt method for ill-posed problems

We describe the regularizing version of the Levenberg–Marquardt method proposed in [12] for solving (1.3), and analyze some issues for its practical implementation.

At k-th iteration of the Levenberg–Marquardt, given $x_k^\delta \in \mathbb{R}^n$ and $\lambda_k > 0$, let

$$m_k^{\mathrm{LM}}(p) = \frac{1}{2}\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\right\|^2 + \frac{1}{2}\lambda_k\|p\|^2, \tag{2.1}$$

be a quadratic model around x_k^δ for the function Φ in (1.5), see [18,20]. The step p_k taken minimizes m_k^{LM}, and $x_{k+1}^\delta = x_k^\delta + p_k$. We observe that, if $p(\lambda)$ is the solution of

$$(B_k + \lambda I)p(\lambda) = -g_k, \tag{2.2}$$

with $B_k = J(x_k^\delta)^T J(x_k^\delta)$ and $g_k = J(x_k^\delta)^T(F(x_k^\delta) - y^\delta)$, then

$$p_k = p(\lambda_k) = -\left(J(x_k^\delta)^T J(x_k^\delta) + \lambda_k I\right)^{-1}\left(J(x_k^\delta)^T\left(F(x_k^\delta) - y^\delta\right)\right). \tag{2.3}$$

If problem (1.3) is ill-posed, and the scalars λ_k are limited to promote convergence of the procedure, see [20], then the solution of (1.1) may be significantly misinterpreted [11,16,26]. The regularizing Levenberg–Marquardt method [12] attempts to approximate solutions of (1.1) by choosing λ_k as the solution λ_k^q of the nonlinear scalar equation

$$\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\right\| = q\left\|F(x_k^\delta) - y^\delta\right\|, \tag{2.4}$$

for some fixed $q \in (0,1)$. Under suitable assumptions discussed below, λ_k^q is uniquely determined from (2.4).

To analyze (2.4), it is useful to establish relations between λ, $\|p(\lambda)\|$ and $\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\|$.

Lemma 2.1 [2, Lemma 4.2] *Suppose $\|g_k\| \neq 0$ and let $p(\lambda)$ be the minimum norm solution of (2.2) with $\lambda \geq 0$. Suppose furthermore that $J(x_k^\delta)$ is of rank ℓ and its singular-value decomposition is given by $U_k\Sigma_k V_k^T$ where Σ_k is the diagonal matrix with entries $\zeta_1, \ldots, \zeta_n$ on the diagonal. Then, denoting $r = (r_1, r_2, \ldots, r_n)^T = U_k^T(F(x_k^\delta) - y^\delta)$, we have that*

$$\|p(\lambda)\|^2 = \sum_{i=1}^{\ell} \frac{\zeta_i^2 r_i^2}{\left(\zeta_i^2+\lambda\right)^2}, \tag{2.5}$$

$$\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\right\|^2 = \sum_{i=1}^{\ell} \frac{\lambda^2 r_i^2}{\left(\zeta_i^2+\lambda\right)^2} + \sum_{i=\ell+1}^{n} r_i^2. \tag{2.6}$$

Using this result, the solution of (2.4) is characterized as follows.

Lemma 2.2 *Suppose* $\|g_k\| \neq 0$*. Let* $p(\lambda)$ *be the minimum norm solution of* (2.2) *with* $\lambda \geq 0$*,* $\mathcal{R}(J(x_k^\delta))^\perp$ *be the orthogonal complement of the range* $\mathcal{R}(J(x_k^\delta))$ *of* $J(x_k^\delta)$*, and* P_k^δ *be the orthogonal projector onto* $\mathcal{R}(J(x_k^\delta))^\perp$*. Then*

(i) *Equation* (2.4) *is not solvable if* $\|P_k^\delta(F(x_k^\delta) - y^\delta)\| > q\|F(x_k^\delta) - y^\delta\|$.
(ii) *If*

$$\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)\left(x^\dagger - x_k^\delta\right)\right\| \leq \frac{q}{\theta_k}\left\|F(x_k^\delta) - y^\delta\right\|, \tag{2.7}$$

for some $\theta_k > 1$*, and* $x^\dagger$ *is a solution of* (1.1)*, then Eq.* (2.4) *has a unique solution* λ_k^q *such that*

$$\lambda_k^q \in \left(0, \frac{q}{1-q}\|B_k\|\right]. \tag{2.8}$$

Proof (i) Equation (2.6) implies

$$\lim_{\lambda\to 0}\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\right\| = \left\|P_k^\delta\left(F(x_k^\delta) - y^\delta\right)\right\|,$$
$$\lim_{\lambda\to \infty}\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\right\| = \|F(x_k^\delta) - y^\delta\|.$$

Thus, since $\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\|$ is monotonically increasing as a function of λ, we conclude that (2.4) does not admit solution if $\|P_k^\delta(F(x_k^\delta) - y^\delta)\| > q\|F(x_k^\delta) - y^\delta\|$.

(ii) Trivially $\|P_k^\delta(F(x_k^\delta) - y^\delta)\| \leq \|F(x_k^\delta) - y^\delta + J(x_k^\delta)(x - x_k^\delta)\|$, for any x. Hence, for the monotonicity of $\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\|$, if (2.7) holds, then Eq. (2.4) admits a solution which is positive and unique. Finally, observing that for a positive λ it holds $(J(x_k^\delta)^T J(x_k^\delta) + \lambda I)^{-1} J(x_k^\delta)^T = J(x_k^\delta)^T (J(x_k^\delta) J(x_k^\delta)^T + \lambda I)^{-1}$, Eq. (2.3) can be written as

$$p_k = p(\lambda_k) = -J(x_k^\delta)^T (J(x_k^\delta) J(x_k^\delta)^T + \lambda_k I)^{-1} (F(x_k^\delta) - y^\delta), \tag{2.9}$$

and consequently

$$\begin{aligned} &F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda_k) \\ &= \left(I - J(x_k^\delta)J(x_k^\delta)^T\left(J(x_k^\delta)J(x_k^\delta)^T + \lambda_k I\right)^{-1}\right)\left(F(x_k^\delta) - y^\delta\right) \\ &= \lambda_k\left(J(x_k^\delta)J(x_k^\delta)^T + \lambda_k I\right)^{-1}\left(F(x_k^\delta) - y^\delta\right). \end{aligned} \tag{2.10}$$

Then (2.4) gives

$$q \left\| F(x_k^\delta) - y^\delta \right\| = \lambda_k^q \left\| \left(J(x_k^\delta) J(x_k^\delta)^T + \lambda_k^q I \right)^{-1} \left(F(x_k^\delta) - y^\delta \right) \right\|$$
$$\geq \frac{\lambda_k^q}{\|B_k\| + \lambda_k^q} \left\| F(x_k^\delta) - y^\delta \right\|,$$

which yields (2.8). □

In [12], the analysis of the regularizing properties of the Levenberg–Marquardt method was made under the subsequent assumptions on the solvability of problem (1.1), the Taylor remainder of F, and the vicinity of the initial guess x_0 to some solution $x^\dagger$ of (1.1).

Assumption 2.1 Given an initial guess x_0, there exist positive ρ and c such that system (1.1) is solvable in $B_\rho(x_0)$, and

$$\|F(x) - F(\tilde{x}) - J(x)(x - \tilde{x})\| \leq c\|x - \tilde{x}\| \, \|F(x) - F(\tilde{x})\|, \quad x, \tilde{x} \in B_{2\rho}(x_0). \tag{2.11}$$

Assumption 2.2 Let x_0, c and ρ as in Assumption 2.1, $x^\dagger$ be a solution of (1.1) and x_0 satisfy

$$\|x_0 - x^\dagger\| < \min\left\{\frac{q}{c}, \rho\right\}, \qquad \text{if} \quad \delta = 0, \tag{2.12}$$

$$\|x_0 - x^\dagger\| < \min\left\{\frac{q\tau - 1}{c(1+\tau)}, \rho\right\}, \qquad \text{if} \quad \delta > 0, \tag{2.13}$$

where $\tau > 1/q$.

Note that, whenever x_k^δ belongs to $B_{2\rho}(x_0)$ and $\|x_k^\delta - x^\dagger\| < \|x_0 - x^\dagger\|$, Assumption 2.1 implies that inequality (2.7) is satisfied for some $\theta_k > 1$, and consequently there exists a solution to (2.4).

Under Assumptions 2.1 and 2.2, the Levenberg–Marquardt method generates an approximation $x_{k^*(\delta)}^\delta$ satisfying (1.4) and the sequence $\{x_{k_*(\delta)}^\delta\}$ converges to a solution of (1.1) as δ tends to zero.

Theorem 2.3 *Let Assumptions 2.1 and 2.2 hold and x_k^δ be the Levenberg–Marquardt iterates determined by using (2.4). For noisy data, suppose $k < k_*(\delta)$ where $k_*(\delta)$ is defined in (1.4). Then, any iterate x_k^δ belongs to $B_{2\rho}(x_0)$. With exact data, the sequence $\{x_k\}$ converges to a solution of (1.1). With noisy data, the stopping criterion (1.4) is satisfied after a finite number $k_*(\delta)$ of iterations and $\{x_{k_*(\delta)}^\delta\}$ converges to a solution of (1.1) as δ tends to zero.*

Proof See [12], Theorems 2.2 and 2.3. □

Let us focus on a specific issue concerning the implementation of the method which, to our knowledge, has not been addressed either in [12] or in related papers. The numerical solution of (2.4) requires the application of a root-finder method and Newton's method is the most efficient procedure, though in general it requires the knowledge of an accurate approximation to the solution. On the other hand, nonlinear equations which are monotone and convex (or concave) on some interval containing the root are particularly suited to an application of Newton's method, see e.g. [14, Theorem 4.8]. Equation (2.4) does not have such properties but it can be replaced by an equivalent equation with strictly decreasing and concave function in $[\lambda_k^q, \infty)$; thus, Newton's method applied to the reformulated equation converges globally to λ_k^q whenever the initial guess overestimates such a root.

Lemma 2.4 *Suppose* $\|F(x_k^\delta) - y^\delta\| \neq 0$, *and that* (2.4) *has positive solution* λ_k^q. *Let*

$$\psi(\lambda) = \frac{\lambda}{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\|} - \frac{\lambda}{q\,\|F(x_k^\delta) - y^\delta\|} = 0. \tag{2.14}$$

Then, Newton's method applied to (2.14) *converges monotonically and globally to the root* λ_k^q *of (2.4) for any initial guess in the interval* $[\lambda_k^q, \infty)$.

Proof Trivially, solving (2.4) is equivalent to finding the positive root of Eq. (2.14). We now show that $\psi(\lambda)$ is strictly decreasing in $[\lambda_k^q, \infty)$ and concave on $(0, \infty)$. By (2.6),

$$\frac{\lambda}{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda)\|} = \left(\sqrt{\sum_{i=1}^{l}\left(\frac{r_i}{\varsigma_i^2+\lambda}\right)^2 + \sum_{i=l+1}^{n}\left(\frac{r_i}{\lambda}\right)^2}\,\right)^{-1}, \tag{2.15}$$

and this function is concave on $(0, \infty)$, cfr. [4, Lemma 2.1]. Thus, ψ is concave on $(0, \infty)$ and trivially $\psi'(\lambda)$ is strictly decreasing.

Now we show that $\psi'(\lambda_k^q)$ is negative; thus, using the monotonicity of $\psi'(\lambda)$, we get that $\psi(\lambda)$ is strictly decreasing in $[\lambda_k^q, \infty)$. Differentiation of $\psi(\lambda)$ and (2.4) give

$$\begin{aligned}
\psi'(\lambda_k^q) &= \frac{\left(\lambda_k^q\right)^3}{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\left(\lambda_k^q\right)\|^3} \\
&\quad\times\left(\sum_{i=1}^{l}\frac{r_i^2}{\left(\varsigma_i^2+\lambda_k^q\right)^3} + \sum_{i=l+1}^{n}\frac{r_i^2}{\left(\lambda_k^q\right)^3}\right) - \frac{1}{q\,\|F(x_k^\delta) - y^\delta\|} \\
&= \frac{\left(\lambda_k^q\right)^2}{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\left(\lambda_k^q\right)\|^3} \\
&\quad\times\left(\sum_{i=1}^{l}\frac{r_i^2\lambda_k^q}{\left(\varsigma_i^2+\lambda_k^q\right)^3} + \sum_{i=l+1}^{n}\left(\frac{r_i}{\lambda_k^q}\right)^2 - \frac{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\left(\lambda_k^q\right)\|^2}{\left(\lambda_k^q\right)^2}\right).
\end{aligned}$$

Moreover, using (2.15), it holds

$$\psi'\left(\lambda_k^q\right) = \frac{\left(\lambda_k^q\right)^2}{\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\left(\lambda_k^q\right)\right\|^3}\left(\sum_{i=1}^{l}\frac{r_i^2\lambda_k^q}{\left(\zeta_i^2+\lambda_k^q\right)^3} - \sum_{i=1}^{l}\left(\frac{r_i}{\zeta_i^2+\lambda_k^q}\right)^2\right)$$
$$= -\frac{\left(\lambda_k^q\right)^2}{\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p(\lambda_k^q)\right\|^3}\sum_{i=1}^{l}\frac{r_i^2\zeta_i^2}{\left(\zeta_i^2+\lambda_k^q\right)^3},$$

i.e. $\psi'(\lambda_k^q)$ is negative.

The claimed convergence of Newton's method follows from results on univariate concave functions given in [14, Theorem 4.8]. □

For the practical evaluation of $\psi(\lambda)$ and $\psi'(\lambda)$ we refer to [5,21].

In [12, Remark p. 6] Hanke observed that (2.4) may be replaced with

$$\left\|F\left(x_k^\delta\right) - y^\delta + J\left(x_k^\delta\right)p_k\right\| \geq q\left\|F\left(x_k^\delta\right) - y^\delta\right\|, \tag{2.16}$$

later denoted as the *q-condition*, but this criterion was not analyzed or employed in numerical computation. Since (2.4) may not have a solution and our aim is to tune λ_k in view of global convergence, while preserving regularizing properties, in the next section we allow more flexibility in its selection and design a trust-region method based on condition (2.16).

3 A regularizing trust-region method

Trust-region methods are globally convergent approaches where the stepsize between two successive iterates is determined via a nonlinear stepsize control mechanism [5]. At a generic iteration k of a trust-region method, the step p_k solves

$$\begin{aligned}&\min_p m_k^{\mathrm{TR}}(p) = \frac{1}{2}\left\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p\right\|^2,\\&\text{s.t. } \|p\| \leq \Delta_k,\end{aligned} \tag{3.1}$$

where Δ_k is a given positive trust-region radius. If $\|g_k\| \neq 0$ then p_k solves (3.1) if and only if it satisfies (2.2) for some nonnegative λ_k such that

$$\lambda_k(\|p_k\| - \Delta_k) = 0. \tag{3.2}$$

Therefore, whenever the minimum norm solution p^+ of

$$B_k p = -g_k,$$

satisfies $\|p^+\| \leq \Delta_k$, the scalar λ_k is null and $p_k = p(0)$ solves (3.1). Otherwise, the step p_k takes the form (2.3), and therefore it is a Levenberg–Marquardt step. If $\|p_k\| = \Delta_k$, then the trust-region is said to be active.

Starting from an arbitrary initial guess, trust-region methods generate a sequence of iterates such that the value of Φ in (1.5) is monotonically decreasing and this feature is enforced by an adaptive choice of the radius Δ_k. Specifically, let p_k be the trust-region step and

$$\pi_k(p_k) = \frac{ared(p_k)}{pred(p_k)}, \tag{3.3}$$

be the ratio between the achieved $ared(p_k)$ and predicted $pred(p_k)$ reductions given by

$$ared(p_k) = \Phi\left(x_k^\delta\right) - \Phi\left(x_k^\delta + p_k\right), \tag{3.4}$$

$$pred(p_k) = \Phi\left(x_k^\delta\right) - m_k^{\mathrm{TR}}(p_k). \tag{3.5}$$

Then, the trust region radius is reduced if $\pi_k(p_k)$ is below some small positive threshold; otherwise it is left unchanged or enlarged [5].

Since trust-region steps and Levenberg–Marquardt steps have the same form (2.2), trust-region and Levenberg–Marquardt methods fall into a single unifying framework which can be used for their description and theoretical analysis [4,21,24]. Due to such a strict connection, we elaborate on the regularizing Levenberg–Marquardt described in the previous section, and introduce a regularizing variant of trust-region methods for solving ill-posed problems.

The standard trust-region strategy is modified so that the nonlinear stepsize control enforces both the monotonic reduction of Φ and the q-condition (2.16). To this end, we first characterize the parameters λ such that $p(\lambda)$ satisfies (2.16).

Lemma 3.1 *Assume $\|g_k\| \neq 0$. Let $p(\lambda)$ be the minimum norm solution of (2.2) with $\lambda \geq 0$ and P_k^δ be the orthogonal projector onto $\mathcal{R}(J(x_k^\delta))^\perp$. Then, Eq. (2.16) is satisfied for any $\lambda \geq 0$ whenever*

$$\left\| P_k^\delta \left(F(x_k^\delta) - y^\delta\right)\right\| \geq q \left\| F(x_k^\delta) - y^\delta \right\|. \tag{3.6}$$

Otherwise, it is satisfied for any $\lambda \geq \lambda_k^q$ where λ_k^q satisfies (2.8).

Proof The claims easily follow from Lemma 2.2. □

Now we are ready to characterize the size of the trust-region radius guaranteeing (2.16).

Lemma 3.2 *Let p_k solve the trust-region problem (3.1). If*

$$\Delta_k \leq \frac{1-q}{\|B_k\|}\|g_k\|, \tag{3.7}$$

then p_k satisfies the q-condition (2.16).

Proof By Lemma 3.1 we know that the q-condition is satisfied either for $\lambda \geq 0$, or for any $\lambda \geq \lambda_k^q$. In the former case, the claim trivially holds. In the latter case, by (2.2) it follows

$$\left\| p\left(\lambda_k^q\right) \right\| \geq \frac{\|g_k\|}{\left\| B_k + \lambda_k^q I \right\|},$$

and by (2.8) it holds

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

By construction $\|p_k\| \leq \Delta_k$, and if (3.7) holds then we obtain

$$\|p_k\| = \|p(\lambda_k)\| \leq \frac{1-q}{\|B_k\|}\|g_k\| \leq \frac{\|g_k\|}{\left\| B_k + \lambda_k^q I \right\|} \leq \left\| p\left(\lambda_k^q\right) \right\|.$$

Since $\|p(\lambda)\|$ is monotonically decreasing, it follows $\lambda_k \geq \lambda_k^q$ and condition (2.16) is satisfied. □

We stress that the bound (3.7) provides a practical rule for choosing the trust-region radius and enforcing the q-condition (2.16). Conversely, in papers [27,29], where trust-region methods for ill-posed problems are studied, such a condition is respectively assumed to be satisfied, and explicitly enforced rejecting the step whenever it does not hold.

The result in Lemma 3.2 suggests the trust-region iteration described in Algorithm 3.1. We distinguish between the parameters needed in the case of exact data and the parameters required with noisy data.

Algorithm 3.1: k-th iteration of the regularizing Trust-Region method for problem (1.3)

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

1. Compute $B_k = J(x_k^\delta)^T J(x_k^\delta)$ and $g_k = J(x_k^\delta)^T (F(x_k^\delta) - y^\delta)$.

2. Choose $\Delta_k \in \left[C_{\min}\|g_k\|,\ \min\left\{ C_{\max}, \frac{1-q}{\|B_k\|} \right\} \|g_k\| \right]$.

3. Repeat
   - 3.1 Compute the solution p_k of the trust-region problem (3.1).
   - 3.2 Compute $\pi_k(p_k)$ given in (3.3)–(3.5).
   - 3.3 If $\pi_k(p_k) < \eta$, then set $\Delta_k = \gamma \Delta_k$.

   Until $\pi_k(p_k) \geq \eta$.

4. Set $x_{k+1}^\delta = x_k^\delta + p_k$.

Algorithm 3.1 is well-defined, provided that the following assumption is met.

Assumption 3.1 There exists a positive constant κ_J such that

$$\|J(x)\| \leq \kappa_J,$$

for any x belonging to the level set $\mathcal{L} = \{x \in \mathbb{R}^n \text{ s.t. } \Phi(x) \leq \Phi(x_0)\}$.

First, Step 2 is well defined for suitable choices of $C_{\min}$; in fact, as long as $C_{\min} < \dfrac{1-q}{\kappa_J^2}$, it holds $C_{\min} < \dfrac{1-q}{\|B_k\|}$ for all k. Second, due to well-known properties of trust-region methods, Assumption 3.1 guarantees that the step p_k is found within a finite number of attempts, whenever $\|g_k\| \neq 0$ [5].

Global convergence of the trust-region method is stated in the following theorem; we refer to [23, Theorem 11.9] for the proof.

Theorem 3.3 *Suppose that Assumption 3.1 holds and J is Lipschitz continuous on $\mathbb{R}^n$. Then, the sequence $\{x_k^\delta\}$ generated by Algorithm 3.1 satisfies*

$$\lim_{k\to\infty} \nabla\Phi\left(x_k^\delta\right) = \lim_{k\to\infty} \left\| J(x_k^\delta)^T \left(F(x_k^\delta) - y^\delta\right)\right\| = 0. \tag{3.8}$$

We observe that assumption on Lipschitz continuity of J is made in [15], too.

By construction, the sequence $\{\|F(x_k^\delta) - y^\delta\|\}$ is monotonically decreasing and bounded below by zero; hence it is convergent. Equation (3.8) implies that any accumulation point of the sequence $\{x_k^\delta\}$ is a stationary point of Φ. As for exact data, we conclude that if there exists an accumulation point of $\{x_k\}$ solving (1.1), then any accumulation point of the sequence solves (1.1). In the case of noisy data, if the value of Φ at some accumulation point of $\{x_k^\delta\}$ is below the scalar $\tau\delta$, then there exists an iterate $x_{k_*(\delta)}^\delta$ such that the discrepancy principle is met.

It remains to show the behaviour of the iterates generated by Algorithm 3.1 when, for some k, x_k^δ is sufficiently close to a solution $x^\dagger$ of (1.1). For instance, this occurs with exact data when the accumulation points of $\{x_k\}$ solve (1.1) and k is sufficiently large. In the next section we show that the trust-region method described in Algorithm 3.1 shares the same local regularizing properties as the regularizing Levenberg–Marquardt method.

4 Local behaviour of the trust-region method

We analyze the local properties of the trust-region method under the same assumptions made for the Levenberg–Marquardt method. Hence, we suppose that there exists an iteration index $\bar{k}$ such that the iterate $x_{\bar{k}}^\delta$ satisfies the following assumptions that are the counterpart of Assumptions 2.1 and 2.2 for the Levenberg–Marquardt method.

Assumption 4.1 Suppose that for some iteration index $\bar{k}$ there exist positive ρ and c such that system (1.1) is solvable in $B_\rho(x_{\bar{k}}^\delta)$, and

$$\|F(x) - F(\tilde{x}) - J(x)(x-\tilde{x})\| \leq c\|x-\tilde{x}\|\,\|F(x) - F(\tilde{x})\|, \quad x, \tilde{x} \in B_{2\rho}\left(x_{\bar{k}}^\delta\right), \tag{4.1}$$

with $\bar{k} < k_*(\delta)$ if the data are noisy, where $k_*(\delta)$ is defined in (1.4). Moreover, letting $x^\dagger$ be a solution of (1.1), suppose that $x^\delta_{\bar{k}}$ satisfies

$$\|x_{\bar{k}} - x^\dagger\| < \min\left\{\frac{q}{c}, \rho\right\}, \qquad \text{if} \quad \delta = 0, \tag{4.2}$$

$$\|x^\delta_{\bar{k}} - x^\dagger\| < \min\left\{\frac{q\tau - 1}{c(1+\tau)}, \rho\right\}, \qquad \text{if} \quad \delta > 0. \tag{4.3}$$

where $\tau > 1/q$.

Typically in the literature assumptions stronger than (4.1) have been made. To our knowledge, except for papers [7–9,27–29], local convergence properties of trust-region strategies have been analyzed under assumptions which involve the inverse of J and its upper bound in a neighbourhood of a solution. In papers [7–9,29] the convergence analysis is carried out assuming a local error-bound condition and a Lipschitz condition on the Jacobian in a neighbourhood of $x^\dagger$.

The following theorems show the local behaviour of the regularizing trust-region method. We prove that locally the trust-region is active, the iterates x^δ_k with $k > \bar{k}$ remain into the ball $B_\rho(x^\dagger)$ and the resulting algorithm is regularizing. We remark that in standard trust-region methods, the trust-region becomes eventually inactive. On the other hand, regularization requires strictly positive scalars λ_k, and consequently an active trust-region in all iterations. First, we give a technical result that will be useful in the subsequent analysis. Then, we focus on the noise-free case and we show that the error $\|x_k - x^\dagger\|$ decreases in a monotonic way for $k \geq \bar{k}$, and the sequence $\{x_k\}$ converges to a solution of (1.1).

Lemma 4.1 *Assume that Eq.* (2.7) *is fulfilled for some* $\theta_k > 1$ *and* $x^\dagger$ *being a solution of* (1.1). *Let* $x_{k+1} = x_k + p_k$ *with* $p_k = p(\lambda_k)$ *satisfying* (2.2) *and* (2.16). *Then it holds*

$$\left\|x^\delta_k - x^\dagger\right\|^2 - \left\|x^\delta_{k+1} - x^\dagger\right\|^2 > \frac{2(\theta_k - 1)}{\theta_k \lambda_k} \left\|F\left(x^\delta_k\right) - y^\delta + J\left(x^\delta_k\right) p_k\right\|^2. \tag{4.4}$$

Proof The proof parallels that of [16, Proposition 4.1], in which it is shown that

$$\begin{aligned}
&\left\|x^\delta_{k+1} - x^\dagger\right\|^2 - \left\|x^\delta_k - x^\dagger\right\|^2 \\
&\quad < \frac{2}{\lambda_k} \left\|F\left(x^\delta_k\right) - y^\delta + J\left(x^\delta_k\right) p_k\right\| \Bigg(\left\|F\left(x^\delta_k\right) - y^\delta + J\left(x^\delta_k\right)\left(x^\dagger - x^\delta_k\right)\right\| \\
&\qquad - \left\|F\left(x^\delta_k\right) - y^\delta + J\left(x^\delta_k\right) p_k\right\| \Bigg).
\end{aligned}$$

From (2.7) and (2.16) it follows that

$$\left\|F\left(x^\delta_k\right) - y^\delta + J\left(x^\delta_k\right)\left(x^\dagger - x^\delta_k\right)\right\| \leq \frac{1}{\theta_k} \left\|F\left(x^\delta_k\right) - y^\delta + J\left(x^\delta_k\right) p_k\right\|,$$

which yields the thesis. □

Lemma 4.2 *Suppose that Assumptions 3.1 and 4.1 hold and $\delta = 0$. Then, Algorithm 3.1 generates a sequence $\{x_k\}$ such that, for $k \geq \bar{k}$,*

(i) *the trust-region is active, i.e. $\lambda_k > 0$, and x_k belongs to $B_{2\rho}(x_{\bar{k}})$ and to $B_\rho(x^\dagger)$;*
(ii) $\|x_{k+1} - x^\dagger\| < \|x_k - x^\dagger\|$;
(iii) *there exists a constant $\bar{\lambda} > 0$ such that $\lambda_k \leq \bar{\lambda}$.*

Proof (i)–(ii) From the choice of Δ_k at Step 2 of Algorithm 3.1 and Lemma 3.2 it follows that the step p_k computed at Step 3 satisfies condition (2.16). Moreover, from Assumption 4.1, it follows that condition (2.7) is satisfied at $k = \bar{k}$ with $\theta_{\bar{k}} = \dfrac{q}{c\|x_{\bar{k}} - x^\dagger\|} > 1$. Consequently, Lemma 2.2 gives that $\lambda^q_{\bar{k}}$ is strictly positive, while Lemma 3.1 yields that the trust-region is active as $\lambda_{\bar{k}} \geq \lambda^q_{\bar{k}}$. Since Lemma 4.1 holds for $k = \bar{k}$, (4.4) implies $\|x_{\bar{k}+1} - x^\dagger\| < \|x_{\bar{k}} - x^\dagger\|$ and, as a consequence, $x_{\bar{k}+1}$ belongs to $B_{2\rho}(x_{\bar{k}})$ and to $B_\rho(x^\dagger)$. Repeating the above arguments, by induction we can prove that condition (2.7) holds for $k > \bar{k}$, with

$$\theta_k = \frac{q}{c\|x^\dagger - x_k\|} > 1, \tag{4.5}$$

and this implies that λ_k is strictly positive. Thus, Lemma 4.1 holds for all $k \geq \bar{k}$ and by induction, the sequence $\{\|x_k - x^\dagger\|\}_{k=\bar{k}}^\infty$ is monotonic decreasing and the sequence $\{\theta_k\}_{k=\bar{k}}^\infty$ is monotonic increasing.

(iii) Since the trust-region is active, by (2.2)

$$\Delta_k = \|p_k\| = \|(B_k + \lambda_k I)^{-1} g_k\| \leq \frac{\|g_k\|}{\lambda_k}. \tag{4.6}$$

Thus our claim follows if $\Delta_k/\|g_k\|$ is larger than a suitable threshold, independent from k. Let us provide such a bound by estimating the value of Δ_k which guarantees condition $\pi_k(p_k) \geq \eta$. If this condition is fulfilled for the value of Δ_k fixed in Step 2 of Algorithm 3.1, then $\Delta_k/\|g_k\| \geq C_{\min}$; otherwise, the trust-region radius is progressively reduced, and we provide a bound for the value of Δ_k at termination of Step 3 of Algorithm 3.1 in the case where $\Phi(x_k + p_k) > m_k^{\mathrm{TR}}(p_k)$. This occurrence represents the most adverse case; in fact if $\Phi(x_k + p_k) \leq m_k^{\mathrm{TR}}(p_k)$ then $\pi_k(p_k) \geq 1 > \eta$ and the repeat loop terminates for a trust-region radius greater than or equal to the one estimated below. Trivially,

$$1 - \pi_k(p_k) = \frac{\Phi(x_k + p_k) - m_k^{\mathrm{TR}}(p_k)}{\Phi(x_k) - m_k^{\mathrm{TR}}(p_k)}, \tag{4.7}$$

and

$$\begin{aligned}\Phi(x_k + p_k) - m_k^{\mathrm{TR}}(p_k) \leq\ & \frac{1}{2}\|F(x_k + p_k) - F(x_k) - J(x_k)p_k\|^2 \\ & + \|F(x_k + p_k) - F(x_k) - J(x_k)p_k\| \\ & \|F(x_k) - y + J(x_k)p_k\|\end{aligned} \tag{4.8}$$

By (4.1) and the mean value Theorem [23, Theorem 11.1], it holds

$$\begin{aligned}\|F(x_k+p_k)-F(x_k)-J(x_k)p_k\| &\le c\|p_k\|\,\|F(x_k+p_k)-F(x_k)\| \\ &\le c\kappa_J\|p_k\|^2. \end{aligned} \tag{4.9}$$

Consequently, as $\Delta_k \le C_{\max}\|g_k\|$,

$$\Phi(x_k+p_k)-m_k^{\mathrm{TR}}(p_k) \le \frac{1}{2}c\kappa_J\Delta_k^2\|F(x_0)-y\|\left(c\kappa_J^3C_{\max}^2\|F(x_0)-y\|+2\right).$$

Theorem 6.3.1 in [5] shows that

$$\Phi(x_k)-m_k^{\mathrm{TR}}(p_k) \ge \frac{1}{2}\|g_k\|\min\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}.$$

Then,

$$\Phi(x_k)-m_k^{\mathrm{TR}}(p_k) \ge \frac{1}{2}\Delta_k\|g_k\|, \tag{4.10}$$

whenever $\Delta_k \le \dfrac{\|g_k\|}{\kappa_J^2}$ and this implies

$$1-\pi_k(p_k) \le \frac{c\kappa_J\Delta_k\|F(x_0)-y\|\left(c\kappa_J^3C_{\max}^2\|F(x_0)-y\|+2\right)}{\|g_k\|}.$$

Namely, termination of the repeat loop occurs with

$$\Delta_k \le \|g_k\|\omega,$$

and

$$\omega = \min\left\{\frac{1}{\kappa_J^2}, \frac{1-\eta}{c\kappa_J\|F(x_0)-y\|\left(c\kappa_J^3C_{\max}^2\|F(x_0)-y\|+2\right)}\right\}. \tag{4.11}$$

Taking into account Step 2 and the updating rule at Step 3.3, we can conclude that, at termination of Step 3, the trust-region radius Δ_k satisfies

$$\Delta_k \ge \min\{C_{\min}, \gamma\omega\}\|g_k\|.$$

Finally, by (4.6) $\lambda_k \le \bar{\lambda}$ as

$$\lambda_k \le \frac{\|g_k\|}{\Delta_k} \le \max\left\{\frac{1}{\gamma\omega}, \frac{1}{C_{\min}}\right\}. \tag{4.12}$$

□

Theorem 4.3 *Suppose that Assumptions 3.1 and 4.1 hold and $\delta = 0$. Then, the sequence $\{x_k\}$ generated by Algorithm 3.1 converges to a solution x^* of (1.1) such that $\|x^* - x^\dagger\| \le \rho$.*

Proof Let $\bar{k}$ as in Assumption 4.1 and $k \ge \bar{k}$. In Lemma 4.2 we showed that (4.4) holds with θ_k given in (4.5) and monotonically increasing. Then, an adaptation of the proof of Theorem 4.2 in [16] gives that $\{x_k\}$ is convergent; the proof is repeated for sake of clarity. Set $\sigma = c\|x_{\bar{k}} - x^\dagger\|$. Clearly, from Lemma 4.2 we have $\sigma \ge c\|x_i - x^\dagger\|$ for all $i \ge \bar{k}$. Moreover, using (4.1) we obtain

$$\|J(x_i)(x_k - x^\dagger)\| \le (1 + 5\sigma)\|F(x_i) - y\|, \tag{4.13}$$

for all $k \ge i \ge \bar{k}$. Letting $e_k = x_k - x^\dagger$, from (2.9), (2.10) and (4.13) we obtain that for $k > j \ge \bar{k}$:

$$\begin{aligned} |\langle e_j - e_k, e_k\rangle| &= \left|\sum_{i=j}^{k-1}\left\langle (J(x_i)J(x_i)^T + \lambda_i I)^{-1})(y - F(x_i)), J(x_i)e_k\right\rangle\right| \\ &\le \sum_{i=j}^{k-1} \|(J(x_i)J(x_i)^T + \lambda_i I)^{-1}(y - F(x_i))\|\|J(x_i)e_k\| \\ &\le (1 + 5\sigma)\sum_{i=j}^{k-1}\frac{1}{\lambda_i}\|F(x_i) - y + J(x_i)(x_{i+1} - x_i)\|\|F(x_i) - y\|. \end{aligned}$$

Thus, (2.16) and (4.4) yield

$$\begin{aligned} | < e_j - e_k, e_k > | &\le (1 + 5\sigma)\sum_{i=j}^{k-1}\frac{1}{\lambda_i q}\|F(x_i) - y + J(x_i)(x_{i+1} - x_i)\|^2 \\ &\le \alpha_{\bar{k}}(\|e_j\|^2 - \|e_k\|^2), \end{aligned} \tag{4.14}$$

where $\alpha_{\bar{k}} = \dfrac{(1 + 5\sigma)\theta_{\bar{k}}}{2q(\theta_{\bar{k}} - 1)}$ and we have used $\theta_k/(\theta_k - 1) < \theta_{\bar{k}}/(\theta_{\bar{k}} - 1)$ since the function $\theta/(\theta - 1)$ is monotonic decreasing. Then

$$\|x_k - x_j\|^2 = 2 < e_k - e_j, e_k > + \|e_j\|^2 - \|e_k\|^2 \le (2\alpha_{\bar{k}} + 1)(\|e_j\|^2 - \|e_k\|^2).$$

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^*. By $x_k \in B_\rho(x^\dagger)$ for $k \ge \bar{k}$, it follows $\|x^* - x^\dagger\| \le \rho$.

Finally, from Lemma 4.2 we know that $\lambda_k \le \bar{\lambda}$ and $(\theta_k - 1)/\theta_k \ge (\theta_{\bar{k}} - 1)/\theta_{\bar{k}}$, for $k \ge \bar{k}$ since the function $(\theta - 1)/\theta$ is monotonic increasing. Then, by (4.4) and (2.16)

$$\|x_k - x^\dagger\|^2 - \|x_{k+1} - x^\dagger\|^2 \ge \frac{2(\theta_{\bar{k}} - 1)q^2}{\theta_{\bar{k}}\bar{\lambda}}\|F(x_k) - y\|^2.$$

Thus we conclude that $\|F(x_k) - y\|$ tends to zero and the limit x^* of x_k solves (1.1). □

A similar result can be given for the noisy case. In the following lemma we prove that for $\bar{k} \le k < k_*(\delta)$, where $k_*(\delta)$ is defined in (1.4), the trust region is active and the scalars $\lambda_k > 0$ are bounded above. Successively, we prove that the stopping criterion (1.4) is satisfied after a finite number of iterations and the method is regularizing as the error decreases monotonically and the sequence $\{x^\delta_{k_*(\delta)}\}$ converges to a solution of (1.1) as δ tends to zero.

Lemma 4.4 *Suppose that* $\delta > 0$ *and Assumptions* 3.1 *and* 4.1 *hold. Then, Algorithm* 3.1 *generates a sequence* x^δ_k *such that, for* $\bar{k} \le k < k_*(\delta)$,

- (i) *the trust-region is active, i.e.* $\lambda_k > 0$ *and* x^δ_k *belongs to* $B_{2\rho}(x^\delta_{\bar{k}})$ *and to* $B_\rho(x^\dagger)$*;*
- (ii) $\|x^\delta_{k+1} - x^\dagger\| < \|x^\delta_k - x^\dagger\|$*;*
- (iii) *there exists a constant* $\bar{\lambda} > 0$ *such that* $\lambda_k \le \bar{\lambda}$.

Proof (i)–(ii) By (4.1) and (1.2) we get

$$\begin{aligned}\left\|y^\delta - F\left(x^\delta_{\bar{k}}\right) - J\left(x^\delta_{\bar{k}}\right)\left(x^\dagger - x^\delta_{\bar{k}}\right)\right\| &\le \delta + \left\|y - F\left(x^\delta_{\bar{k}}\right) - J\left(x^\delta_{\bar{k}}\right)\left(x^\dagger - x^\delta_{\bar{k}}\right)\right\| \\ &\le \delta + c\left\|x^\dagger - x^\delta_{\bar{k}}\right\|\left\|y - F(x^\delta_{\bar{k}})\right\| \\ &\le \left(1 + c\left\|x^\dagger - x^\delta_{\bar{k}}\right\|\right)\delta \\ &\quad + c\left\|x^\dagger - x^\delta_{\bar{k}}\right\|\left\|y^\delta - F\left(x^\delta_{\bar{k}}\right)\right\|.\end{aligned}$$

Then, at iteration $\bar{k}$, condition (1.4) gives

$$\begin{aligned}&\left\|y^\delta - F\left(x^\delta_{\bar{k}}\right) - J\left(x^\delta_{\bar{k}}\right)\left(x^\dagger - x^\delta_{\bar{k}}\right)\right\| \\ &\le \left(\frac{1 + c\left\|x^\dagger - x^\delta_{\bar{k}}\right\|}{\tau} + c\left\|x^\dagger - x^\delta_{\bar{k}}\right\|\right)\left\|y^\delta - F\left(x^\delta_{\bar{k}}\right)\right\|,\end{aligned}$$

and (4.3) yields (2.7) at $k = \bar{k}$ with $\theta_{\bar{k}} = \dfrac{q\tau}{1 + c(1+\tau)\|x^\dagger - x^\delta_{\bar{k}}\|} > 1$. Then, Lemma 2.2 and Lemma 3.2 yield $\lambda_{\bar{k}} \ge \lambda^q_{\bar{k}}$ with $\lambda^q_{\bar{k}} > 0$ strictly positive. Further, by Lemma 4.1 condition (4.4) is satisfied with $\theta_k = \theta_{\bar{k}}$, and this implies $\|x^\delta_{\bar{k}+1} - x^\dagger\| < \|x^\delta_{\bar{k}} - x^\dagger\|$ and consequently $x^\delta_{\bar{k}+1}$ belongs to $B_{2\rho}(x^\delta_{\bar{k}})$ and to $B_\rho(x^\dagger)$. Repeating the above arguments, by induction we can prove that, for $\bar{k} < k < k_*(\delta)$, condition (2.7) holds, $\lambda_k > 0$, and (4.4) is satisfied with $\theta_k = \dfrac{q\tau}{1 + c(1+\tau)\|x^\dagger - x^\delta_k\|}$. Thus $\|x^\delta_{k+1} - x^\dagger\| < \|x^\delta_k - x^\dagger\|$ and $\theta_{k+1} > \theta_k$ for $\bar{k} \le k < k_*(\delta)$.

(iii) Proceeding as in the proof of point (iii) of Theorem 4.2, just replacing x_k with x_k^δ, we get that for $\bar{k} \leq k < k_*(\delta)$, $\lambda_k < \bar{\lambda}$ with

$$\bar{\lambda} \leq \max\left\{\frac{1}{\gamma\omega}, \frac{1}{C_{\min}}\right\}.$$

where ω is obtained replacing y with y^δ in (4.11). □

Theorem 4.5 *Suppose that Assumptions 3.1 and 4.1 hold for $\delta \geq 0$. Then, for $\delta > 0$, the iterates generated by Algorithm 3.1 satisfy the stopping criterion (1.4) 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. Then the sequence $\{x^\delta_{k_*(\delta)}\}$ converges to a solution of (1.1) whenever δ tends to zero.

Proof Summing up from $\bar{k}$ to $k_*(\delta) - 1$, by (2.16) and (4.4) it follows

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

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

Convergence of $x^\delta_{k_*(\delta)}$ to a solution of (1.1) as δ tends to zero is obtained by adapting the proof of [12, Theorem 2.3]. 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 (1.4) applied with $\delta = \delta_n$. Assume first that $\tilde{k}$ is a finite accumulation point of $\{k_n\}$. Without loss of generality, for the monotonicity of (1.5), we can assume that $k_n = \tilde{k}$ for all $n \in \mathbb{N}$. Thus, from the definition of k_n it follows that

$$\left\| y^{\delta_n} - F\left(x_{\tilde{k}}^{\delta_n}\right) \right\| \leq \tau\delta_n. \tag{4.15}$$

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

$$x_{\tilde{k}}^{\delta_n} \to x_{\tilde{k}}, \qquad F\left(x_{\tilde{k}}^{\delta_n}\right) \to F(x_{\tilde{k}}) \qquad \text{as } \delta_n \to 0. \tag{4.16}$$

Therefore, by (4.15), it follows that the $\tilde{k}$-th iterate with exact data y is a solution of $F(x) = y$, i.e. $x^* = x_{\tilde{k}}$, and we can 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 solution x^* of (1.1) by Theorem 4.3, 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 $\bar{\rho} < \min\left\{\dfrac{q\tau-1}{c(1+\tau)}, \rho\right\}$. 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 $\tilde{k} \le k_*(\delta_n)$ and

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

Then, for δ_n sufficiently small

$$\left\|x_{\tilde{k}}^{\delta_n} - x^*\right\| \le \left\|x_{\tilde{k}}^{\delta_n} - x_{\tilde{k}}\right\| + \|x_{\tilde{k}} - x^*\| \le \bar{\rho}. \tag{4.17}$$

Now, from item (i) of Lemma 4.4, it holds $x_{\tilde{k}}^{\delta_n} \in B_{2\rho}(x_{\tilde{k}}^{\delta_n})$, while from (4.3) and Theorem 4.3 it holds $x^* \in B_{2\rho}(x_{\tilde{k}}^{\delta_n})$ as

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

Repeating arguments in Lemma 4.4, we use (4.1), (1.2) and (1.4) and get

$$\begin{aligned}
&\left\|y^{\delta_n} - F\left(x_{\tilde{k}}^{\delta_n}\right) - J\left(x_{\tilde{k}}^{\delta_n}\right)\left(x^* - x_{\tilde{k}}^{\delta_n}\right)\right\| \\
&\le \delta_n + \left\|y - F\left(x_{\tilde{k}}^{\delta_n}\right) - J\left(x_{\tilde{k}}^{\delta_n}\right)\left(x^* - x_{\tilde{k}}^{\delta_n}\right)\right\| \\
&\le \delta_n + c\|x^* - x_{\tilde{k}}^{\delta_n}\| \left\|y - F\left(x_{\tilde{k}}^{\delta_n}\right)\right\| \\
&\le \left(1 + c\left\|x^* - x_{\tilde{k}}^{\delta_n}\right\|\right)\delta + c\left\|x^* - x_{\tilde{k}}^{\delta_n}\right\| \left\|y^{\delta_n} - F\left(x_{\tilde{k}}^{\delta_n}\right)\right\| \\
&\le \left(\frac{1 + c\left\|x^* - x_{\tilde{k}}^{\delta_n}\right\|}{\tau} + c\left\|x^* - x_{\tilde{k}}^{\delta_n}\right\|\right)\left\|y^{\delta_n} - F\left(x_{\tilde{k}}^{\delta_n}\right)\right\|.
\end{aligned}$$

Thus, by (4.17) and $\bar{\rho} < \min\left\{\dfrac{q\tau-1}{c(1+\tau)}, \rho\right\}$, it follows that the following counterpart of (2.7)

$$\|F(x_k^\delta) - y^\delta + J(x_k^\delta)\left(x^* - x_k^\delta\right)\| \le \frac{q}{\theta_k}\|F(x_k^\delta) - y^\delta\|$$

is satisfied at $k = \tilde{k}$ with $\theta_{\tilde{k}} = \dfrac{q\tau}{1 + c(1+\tau)\bar{\rho}} > 1$. Replacing $x^\dagger$ with x^*, (4.4) gives $\|x_{\tilde{k}+1}^{\delta_n} - x^*\| < \|x_{\tilde{k}}^{\delta_n} - x^*\|$ and repeating the above arguments, by induction we obtain monotonicity of the error $\|x_k^{\delta_n} - x^*\|$ for $\tilde{k} \le k \le k_n$. Then

$$\left\|x_{k_n}^{\delta_n} - x^*\right\| < \left\|x_{\tilde{k}}^{\delta_n} - x^*\right\| \le \bar{\rho}. \tag{4.18}$$

Finally, since the previous arguments can be repeated for any positive $\epsilon \le \bar{\rho}$, provided that δ_n is small enough, we obtain that

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

□

We underline that the trust-region radius Δ_k selected in Algorithm 3.1 depends continuously on δ in a right interval of the origin whenever $\pi_k(x_{k+1} - x_k) \neq \eta$, for all $k \ge 0$. Under this assumption, the scalar λ_k, implicitly defined by the trust-region problem, depends continuously on δ and this feature is crucial for proving that the sequence $\{x_{k_*(\delta)}^{\delta}\}$ tends to a solution of (1.1) as δ tends to zero. In the following corollary, we show that, whenever the initial guess x_0 is sufficiently close to a solution of (1.1), it holds $\pi_k(x_{k+1} - x_k) > \eta$ and therefore the regularizing properties of our trust-region method are valid under Assumptions 2.1 and 2.2. Then, the proposed trust-region approach shows the same local regularizing properties of the regularizing Levenberg–Marquardt method.

Corollary 4.6 *Suppose that Assumptions 2.1 and 2.2 hold and $\delta \ge 0$. For $\delta > 0$, let $k_*(\delta)$ be defined in (1.4).*

If x_0 is sufficiently close to a solution of (1.1), then the sequence $\{x_{k_*(\delta)}^{\delta}\}$ converges to a solution of (1.1) whenever δ tends to zero.

Proof Theorem 4.3 implies that $\{x_k\}$ converges to a solution of (1.1). Using (4.7)–(4.10) and $\|p_k\| \le \Delta_k$, it follows

$$1 - \pi_k(p_k) \le \frac{\frac{1}{2} c\kappa_J \Delta_k^2 \left(c\kappa_J \Delta_k^2 + \|F(x_k) - y\|\right)}{\frac{1}{2}\Delta_k \|g_k\|} = \frac{c\kappa_J \Delta_k \left(c\kappa_J \Delta_k^2 + \|F(x_k) - y\|\right)}{\|g_k\|},$$

while $\Delta_k \le C_{\max}\|g_k\|$ implies

$$1 - \pi_k(p_k) \le c\kappa_J C_{\max} \left(c\kappa_J \Delta_k^2 + \|F(x_k) - y\|\right).$$

By the convergence of $\{x_k\}$ to a solution of (1.1), the right-hand side of the above inequality tends to zero. Hence, if x_0 is close enough to a solution of (1.1) to ensure $1 - \pi_k(p_k) > \eta$, for $k \ge 0$, Theorem 4.5 gives the thesis. □

5 Numerical results

In this section we report on the performance of the regularizing trust-region method and make comparisons with the regularizing Levenberg–Marquardt method and a standard version of the trust-region method. The test problems are ill-posed and with noisy data, and arise from the discretization of nonlinear Fredholm integral equations of the first kind

$$\int_0^1 k(t, s, x(s))ds = y(t), \quad t \in [0, 1]. \tag{5.1}$$

The integral equations considered model inverse problems from groundwater hydrology and geophysics. Their kernel is of the form

$$k(t, s, x(s)) = \log\left(\frac{(t-s)^2 + H^2}{(t-s)^2 + (H - x(s))^2}\right), \tag{5.2}$$

see [25, Sect. 3], or

$$k(t, s, x(s)) = \frac{1}{\sqrt{1 + (t-s)^2 + x(s)^2}}, \tag{5.3}$$

see [15, Sect. 6]. The interval [0, 1] was discretized with $n = 64$ equidistant grid points $t_i = (i-1)h$, $h = 1/(n-1)$, $i = 1, \dots, n$. Function $x(s)$ was approximated from the n-dimensional subspace of $H_0^1(0, 1)$ spanned by standard piecewise linear functions. Specifically, we let $s_j = (j-1)h$, $h = 1/(n-1)$, $j = 1, \dots, n$, and looked for an approximation $\hat{x}(s) = \sum_{j=1}^n \hat{x}_j \phi_j(s)$ where

$$\phi_1(s) = \begin{cases} \dfrac{s_2 - s}{h} & \text{if} \quad s_1 \le s \le s_2 \\ 0 & \text{otherwise} \end{cases}, \qquad \phi_n(s) = \begin{cases} \dfrac{s - s_{n-1}}{h} & \text{if} \quad s_{n-1} \le s \le s_n \\ 0 & \text{otherwise} \end{cases},$$

and

$$\phi_j(s) = \begin{cases} \dfrac{s - s_{j-1}}{h} & \text{if} \quad s_{j-1} \le s \le s_j, \\ \dfrac{s_{j+1} - s}{h} & \text{if} \quad s_j \le s \le s_{j+1}, \\ 0 & \text{otherwise} \end{cases} \qquad j = 2, \dots n-1.$$

Finally, the integrals $\int_0^1 k(t_i, s, \hat{x}(s))ds$, $1 \le i \le n$, were approximated by the composite trapezoidal rule on the points s_j, $1 \le j \le n$. The resulting discrete problems are square nonlinear systems (1.1) with unknown $x = (\hat{x}_1, \dots, \hat{x}_n)^T$. We observe that $\hat{x}(s_j) = \hat{x}_j$; thus, the j-th component of x approximates a solution of (5.1) at s_j.

Two problems with kernel (5.2) and two problems with kernel (5.3) were considered and built so that solutions (later denoted as true solutions) are known. Concerning kernel (5.2), the first problem is given in [25, p. 46]; it admits as true continuous solutions the functions $x_{true}(s) = c_1 e^{d_1(s+p_1)^2} + c_2 e^{d_2(s-p_2)^2} + c_3 + c_4$ and $x_{true}(s) = 2H - c_1 e^{d_1(s+p_1)^2} - c_2 e^{d_2(s-p_2)^2} - c_3 - c_4$ where $H = 0.2, c_1 = -0.1, c_2 = -0.075, d_1 = -40, d_2 = -60, p_1 = 0.4, p_2 = 0.67$, c_3 and c_4 are chosen such that $x_{true}(0) = x_{true}(1) = 0$. The second problem was given in [27, p. 835] and it has true continuous solutions $x_{true}(s) = 1.3s(1-s) + 0.2$ and $x_{true}(s) = 1.3s(s-1)$.

The third and fourth problems have kernel (5.3); the former has solutions $x_{true}(s) = 1$ and $x_{true}(s) = -1$, $s \in [0, 1]$, see [15, p. 660], while the latter has the discontinuous functions

$$x_{true}(s)=\begin{cases}1 & \text{if } 0\le s\le \frac{1}{2}\\ 0 & \text{if } \frac{1}{2}< s\le 1\end{cases},\quad x_{true}(s)=\begin{cases}-1 & \text{if } 0\le s\le \frac{1}{2}\\ 0 & \text{if } \frac{1}{2}< s\le 1\end{cases} \tag{5.4}$$

as the true solutions, [15, p. 662].

The nonlinear systems arising from the discretization of the four problems are denoted as `P1`, `P2`, `P3` and `P4` respectively, while $x^\dagger \in \mathbb{R}^n$ denotes a solution of the discretized problems. Given the error level δ, the exact data y was perturbed by normally distributed values with mean 0 and variance δ using the MATLAB function `randn`.

All procedures were implemented in MATLAB and run using MATLAB 2014b on an Intel Core(TM) i7-4510U 2.6 GHz, 8 GB RAM; the machine precision is $\epsilon_m \approx 2\times 10^{-16}$. The Jacobian of the nonlinear function F was computed by finite differences. The parameter q used in (2.4) and in (2.16) was set equal to $1.1/\tau$ and the discrepancy principle (1.4) with $\tau = 1.5$ was used as the stopping criterion. A maximum number of 300 iterations was allowed and a failure was declared when this limit was exceeded.

In the implementation of the regularizing trust-region method, Step 3 in Algorithm 3.1 was performed setting $\eta = \frac{1}{4}$, $\gamma = \frac{1}{6}$. Then, in Step 2 the trust-region radius was updated as follows

$$\Delta_0 = \mu_0 \left\| F(x_0) - y^\delta \right\|, \qquad \mu_0 = 10^{-1}, \tag{5.5}$$

$$\Delta_{k+1} = \mu_{k+1} \left\| F\left(x_{k+1}^\delta\right) - y^\delta \right\|, \qquad \mu_{k+1} = \begin{cases}\frac{1}{6}\mu_k & \text{if } q_k < q\\ 2\mu_k & \text{if } q_k > \nu q\\ \mu_k & \text{otherwise}\end{cases} \tag{5.6}$$

with $q_k = \dfrac{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p_k\|}{\|F(x_k^\delta) - y^\delta\|}$, and $\nu = 1.1$. The maximum and minimum values for Δ_k were set to $\Delta_{\max} = 10^4$ and $\Delta_{\min} = 10^{-12}$. This updating strategy turned out to be efficient in practice and was based on the following considerations. Clearly, Δ_k is cheaper to compute than the upper bound in (3.7) and preserves convergence to zero as δ tends to zero and a solution of problem (1.3) is approached. Further, Δ_k is adjusted taking into account the q-condition and by monitoring the value q_k; 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 the q-condition was fulfilled.

The computation of the parameter λ_k was performed applying Newton's method to the equation

$$\psi(\lambda) = \frac{1}{\|p(\lambda)\|} - \frac{1}{\Delta_k} = 0, \tag{5.7}$$

and each Newton's iteration requires the Cholesky factorization of a shifted matrix of the form $B_k + \lambda I$ [5]. Typically high accuracy in the solution of the above scalar equations is not needed [2,5] and this fact was experimentally verified also for our test problems. Hence, after extensive numerical experience, we decided to terminate the Newton's process as soon as $|\Delta_k - \|p(\lambda)\|| \leq 10^{-2}\Delta_k$.

In our implementation of the standard trust-region method, we chose the trust-region radius accordingly to technicalities well-known in the literature, see e.g. [5, Sect. 6.1] and [23, Sect. 11.2]. In particular, we set $\Delta_0 = 1$,

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

with $\Delta_{\max} = 10^4$ and chose $\Delta_{\min} = 10^{-12}$ as the minimum values for Δ_k.

Finally the Levenberg–Marquardt approach was implemented imposing condition (2.4) and solving (2.14) by Newton's method. In order to find an accurate solution for (2.4) it was necessary to use a tighter tolerance, equal to 10^{-5}, than that used in the trust-region algorithm.

Our experiments were made varying the noise level δ on the data y^δ. Tables 1 and 2 display the results obtained by the regularizing trust-region algorithm with noise $\delta = 10^{-4}$ and $\delta = 10^{-2}$ respectively. Runs for four different initial guesses x_0 are reported in the tables. For problems P1 and P2 the initial guesses are $x_0 = 0e, -0.5e, -e, -2e$ and $x_0 = 0e, 0.5e, e, 2e$ respectively, where e denotes

Table 1 Results obtained by the regularizing trust-region method and the regularizing Levenberg–Marquardt method with noise $\delta = 10^{-4}$ and varying initial guesses

Problem	x_0	RTR						RLM	
		it	$\|F - y\|$	nf	cf	e_I	e_T	e_I	e_T
P1	$0\,e$	43	1.3e−4	44	5	5.5e−3	5.5e−3	4.5e−3	4.5e−3
	$-0.5\,e$	63	1.2e−4	71	5	3.2e−2	7.9e−2	3.0e−2	7.1e−2
	$-1\,e$	82	1.4e−4	94	4	3.4e−2	8.4e−2	4.0e−2	7.2e−2
	$-2\,e$	115	1.5e−4	138	4	3.4e−2	8.6e−2	2.9e−2	6.1e−2
P2	$0\,e$	54	1.2e−4	55	5	7.4e−3	7.4e−3	*	*
	$0.5\,e$	56	1.4e−4	59	5	1.1e−2	1.3e−2	*	*
	$1\,e$	73	1.4e−4	84	4	1.0e−2	1.3e−2	7.3e−3	8.3e−3
	$2\,e$	118	1.4e−4	138	4	9.3e−3	1.1e−2	4.8e−3	4.8e−3
P3	$x_0(1.25)$	35	1.4e−4	36	3	1.2e−2	1.2e−2	3.1e−3	3.1e−3
	$x_0(1.5)$	43	1.4e−4	44	3	5.1e−2	5.1e−2	6.2e−2	6.2e−2
	$x_0(1.75)$	45	1.3e−4	46	3	3.2e−1	3.2e−1	3.1e−1	3.1e−1
	$x_0(2)$	65	1.4e−4	71	3	4.6e−1	4.6e−1	3.8e−1	3.8e−1
P4	$x_0(1, 1)$	68	1.5e−4	82	3	4.8e−1	4.8e−1	*	*
	$x_0(0.5, 0)$	64	1.5e−4	75	3	4.9e−1	4.9e−1	4.7e−1	4.7e−1
	$x_0(1.5, 1)$	69	1.5e−4	78	3	5.1e−1	5.1e−1	4.8e−1	4.8e−1
	$x_0(1.5, 0)$	68	1.5e−4	78	4	5.2e−1	7.1e−1	5.1e−1	6.3e−1

Table 2 Results obtained by the regularizing trust-region method and the regularizing Levenberg–Marquardt method with noise $\delta = 10^{-2}$ and varying initial guesses

Problem	x_0	RTR						RLM	
		it	$\|F-y\|$	nf	cf	e_I	e_T	e_I	e_T
P1	$0\,e$	20	1.5e−2	21	6	1.9e−2	1.9e−2	1.8e−2	1.8e−2
	$-0.5\,e$	29	1.0e−2	30	6	2.2e−2	3.1e−1	2.1e−2	3.1e−1
	$-1\,e$	35	1.4e−2	36	5	3.6e−2	6.1e−1	3.3e−2	6.1e−1
	$-2\,e$	40	1.3e−2	41	5	4.9e−2	1.2e+0	4.5e−2	1.2e+0
P2	$0\,e$	30	1.4e−2	31	5	6.9e−3	1.3e−2	*	*
	$0.5\,e$	25	1.4e−2	26	5	1.7e−2	2.1e−1	*	*
	$1\,e$	29	1.4e−2	30	5	3.8e−2	5.4e−1	1.3e−1	5.2e−1
	$2\,e$	37	1.4e−2	39	5	5.5e−2	1.2e+0	2.2e−1	1.1e+0
P3	$x_0(1.25)$	15	1.2e−2	16	4	1.5e−1	1.5e−1	1.5e−1	1.5e−1
	$x_0(1.5)$	17	1.4e−2	18	4	3.2e−1	3.2e−1	3.2e−1	3.2e−1
	$x_0(1.75)$	19	1.4e−2	20	4	5.0e−1	5.0e−1	5.1e−1	5.1e−1
	$x_0(2)$	22	1.5e−2	23	4	6.9e−1	6.9e−1	7.0e−1	7.0e−1
P4	$x_0(1,1)$	17	1.4e−2	18	5	5.7e−1	5.7e−1	5.4e−1	5.4e−1
	$x_0(0.5,0)$	20	1.3e−2	21	4	5.5e−1	5.5e−1	*	*
	$x_0(1.5,1)$	22	1.4e−2	23	4	5.1e−1	5.1e−1	5.0e−1	5.0e−1
	$x_0(1.5,0)$	26	1.5e−2	27	4	5.2e−1	8.8e−1	*	*

the vector $e = (1, \ldots, 1)^T$. For problem P3 the initial guess was chosen as the vector $x_0(\alpha)$ with j-th component given by $(x_0(\alpha))_j = g_\alpha(s_j)$ for $j = 1, \ldots, n$, where $g_\alpha(s) = (-4\alpha + 4)s^2 + (4\alpha - 4)s + 1$, and s_j being the grid points in $[0, 1]$. We have used the following values of α, $\alpha = 1.25, 1.5, 1.75, 2$. For problem P4 the initial guess $x_0(\beta, \chi)$ has components $(x_0(\beta, \chi))_j = g_{\beta,\chi}(s_j)$ for $j = 1, \ldots, n$ with $g_{\beta,\chi} = \beta - \chi s$ and $(\beta, \chi) = (1, 1), (0.5, 0), (1.5, 1), (1.5, 0)$. In the tables we report: the initial guesses (for increasing distance from the true solutions); the number of iterations `it` performed; the final nonlinear residual; the number of function evaluations `nf` performed; the rounded average number `cf` of Cholesky factorizations per iteration. To assess the quality of the results obtained, we measured the distance between the final iterate $x^\delta_{k^*(\delta)}$ and the true solution approached; in particular $\mathtt{e_I} = \max_{2\le j\le n-1} |x_{true}(s_j) - (x^\delta_{k^*(\delta)})_j|$ is the maximum absolute value of the difference between the components associated to internal points $s_j \in (0, 1)$, while $\mathtt{e_T} = \max_{1\le j\le n} |x_{true}(s_j) - (x^\delta_{k^*(\delta)})_j|$ is the maximum absolute value of the difference between the components associated to points s_j including the end-points of the interval $[0, 1]$. The symbol "$*$" indicates that either the procedure failed to satisfy the discrepancy principle within the prescribed maximum number of iteration, or the final $x^\delta_{k^*(\delta)}$ was not an approximation of one of the true solutions described above.

Tables 1 and 2 show that the regularizing trust-region method solves all the tests. By Step 3 of our Algorithm 3.1, the difference between the number of function evaluations and the number of trust-region iterations, if greater than one, indicates the number of

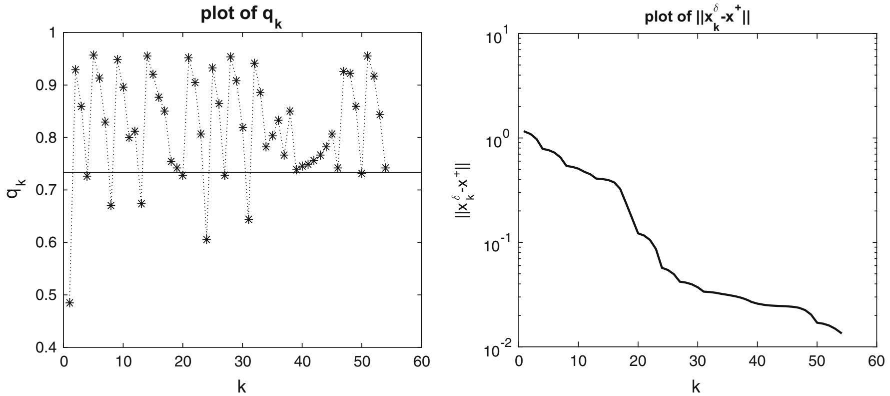

Fig. 1 Regularizing trust-region applied to P2, $x_0 = 0e$, $\delta = 10^{-4}$: values $q_k = \frac{\|F(x_k^\delta)-y^\delta+J(x_k^\delta)p_k\|}{\|F(x_k^\delta)-y^\delta\|}$ (marked by an *asterisk*) and value of $q = 1.1/\tau$ (*solid line*) versus the iterations (*on the left*); semilog plot of the error $\|x_k^\delta - x^\dagger\|$ versus the iterations (*on the right*)

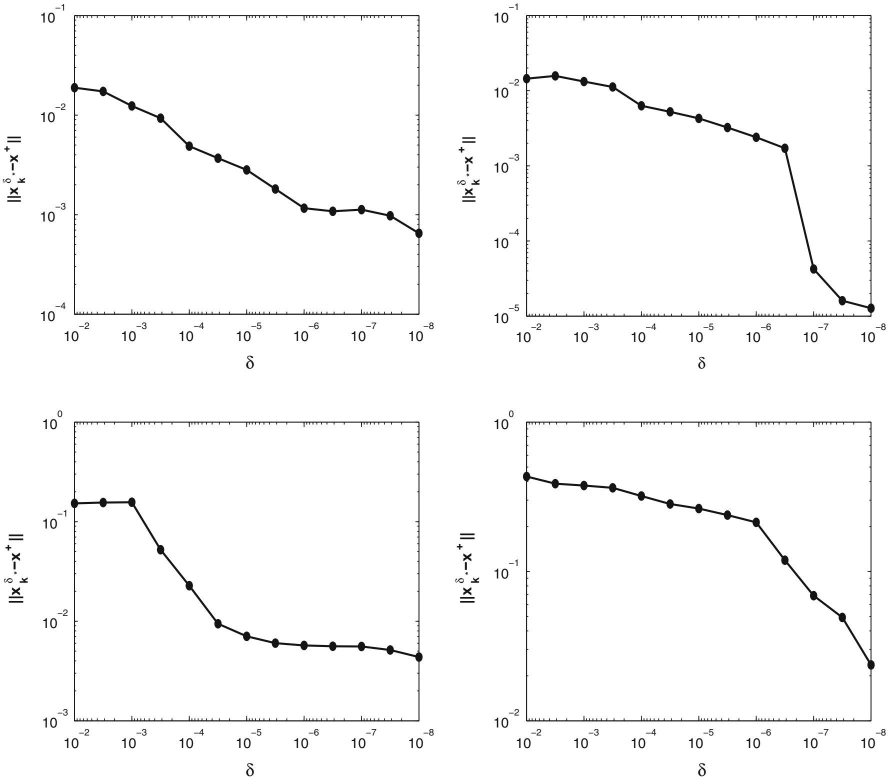

Fig. 2 Regularizing trust-region applied to P1, $x_0 = 0e$ (*top left*), P2, $x_0 = 0e$ (*top right*), P3, $x_0 = x_0(\alpha) = x_0(1.25)$ (*lower left*) and to P4, $x_0 = x_0(\beta, \chi) = x_0(0.5, 0)$ (*lower right*) log plot of the error $\|x_{k^*(\delta)}^\delta - x^\dagger\|$ versus the noise δ

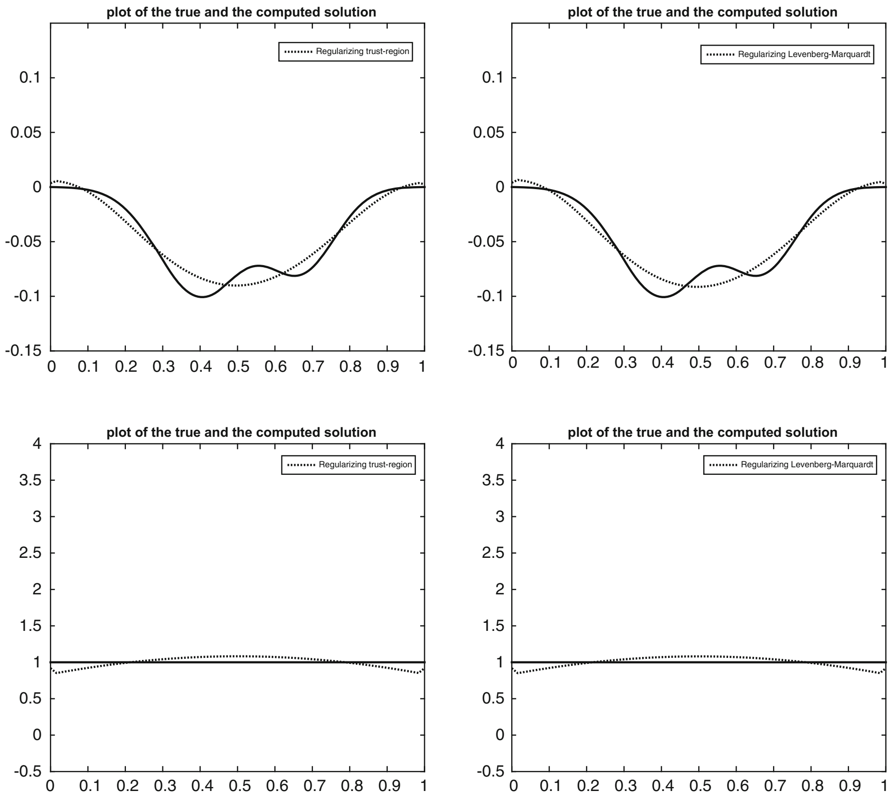

Fig. 3 Regularizing trust-region (*left*) and regularizing Levenberg–Marquardt (*right*), true solution (*solid line*) and approximate solutions (*dotted line*). *Upper part* P1, $\delta = 10^{-2}$, $x_0 = 0e$; *lower part* P3, $\delta = 10^{-2}$, $x_0 = x_0(\alpha) = x_0(1.25)$

trial iterates that were rejected because a sufficient reduction on Φ was not achieved. We observe that in 20 out of 32 runs, all the iterates generated were accepted; this occurrence seems to indicate that the trust-region updating rule works well in practice.

Further insight on the trust-region updating rule (5.5) and (5.6) can be gained analyzing the regularizing properties of the implemented trust-region strategy. First, we verified numerically that, though not explicitly enforced, the q-condition is satisfied in most of the iterations. As an illustrative example, we consider problem P2 with $\delta = 10^{-4}$ and $x_0 = 0e$ and, in the left plot in Fig. 1, we display the values $q_k = \frac{\|F(x_k^\delta) - y^\delta + J(x_k^\delta)p_k\|}{\|F(x_k^\delta) - y^\delta\|}$ at the trust-region iterations, marked by an asterisk, and the value $q = 1.1/\tau \approx 0.733$ fixed in our experiments, depicted by a solid line. We observe that, even if we have not imposed the q-condition, it is satisfied at most of the iterations. The plot on the right of Fig. 1 shows a monotone decay of the error between x_k^δ and $x^\dagger$ through the iterations, which results to be in accordance with the theoretical results in Theorem 4.4. The regularizing properties of the implemented trust-region scheme are also shown in Fig. 2 where, for each test problem we plot the

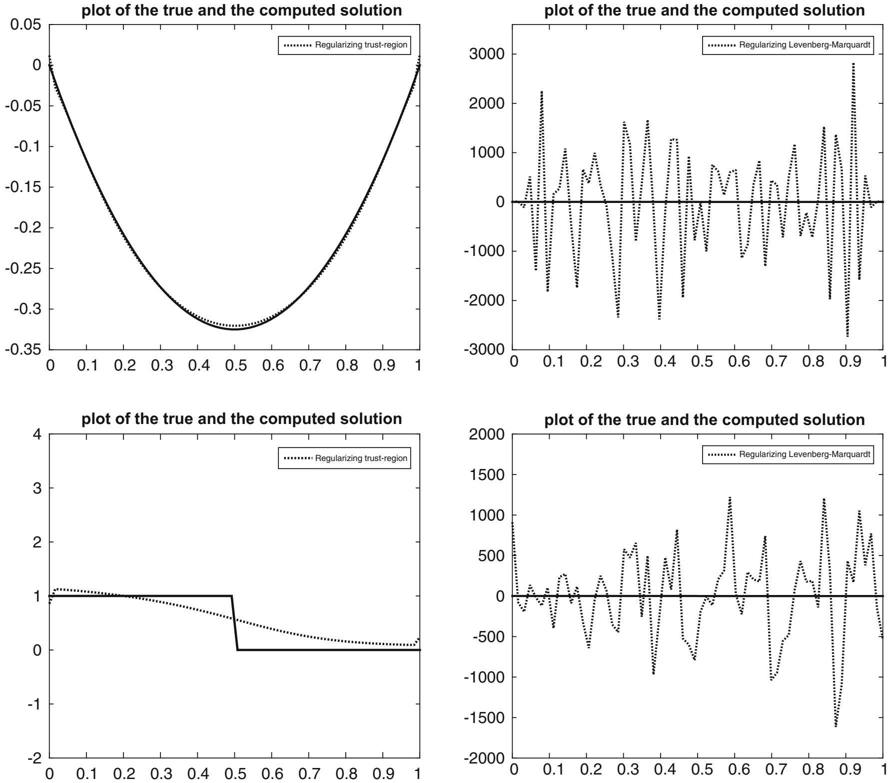

Fig. 4 True solution (*solid line*) and approximate solutions (*dotted line*) computed by the regularizing trust-region method (*on the left*) and the regularizing Levenberg–Marquardt method (*on the right*). *Upper part* problem P2, $\delta = 10^{-2}$, $x_0 = 0e$; *lower part* problem P4, $\delta = 10^{-2}$, $x_0 = x_0(\beta, \chi) = x_0(0.5, 0)$

error $\|x^{\delta}_{k^*(\delta)} - x^{\dagger}\|$ for decreasing noise levels; it is evident that, in accordance with the theory, the error decays as the noise level decreases.

Let now compare the regularizing trust-region and Levenberg–Marquardt procedures. On successful runs for both methods, the two methods provide solutions of similar accuracy and such an accuracy increases with the vicinity of the initial guess to the true solution; as an example Fig. 3 shows the solutions computed by the two methods for problems P1 and P3 for $\delta = 10^{-2}$. On the other hand, for large noise δ and initial guesses farther from the true solution, for both methods the accuracy at the endpoints of the interval [0, 1] may deteriorate; for this occurrence we refer to Table 2 and runs on problems P1 and P2. Concerning failures, in 7 runs out of 32 the Levenberg–Marquardt algorithm does not act as a regularizing method as the generated sequence approaches a solution of the noisy problem. In Fig. 4 we illustrate two unsuccessful runs of the Levenberg–Marquardt method; approximated solution computed by the regularizing trust-region and Levenberg–Marquardt procedures are shown for runs on problems P2 and P4.

The overall experience on the Levenberg–Marquardt algorithm seems to indicate that the use of the q-condition is more flexible than condition (2.4) and provides

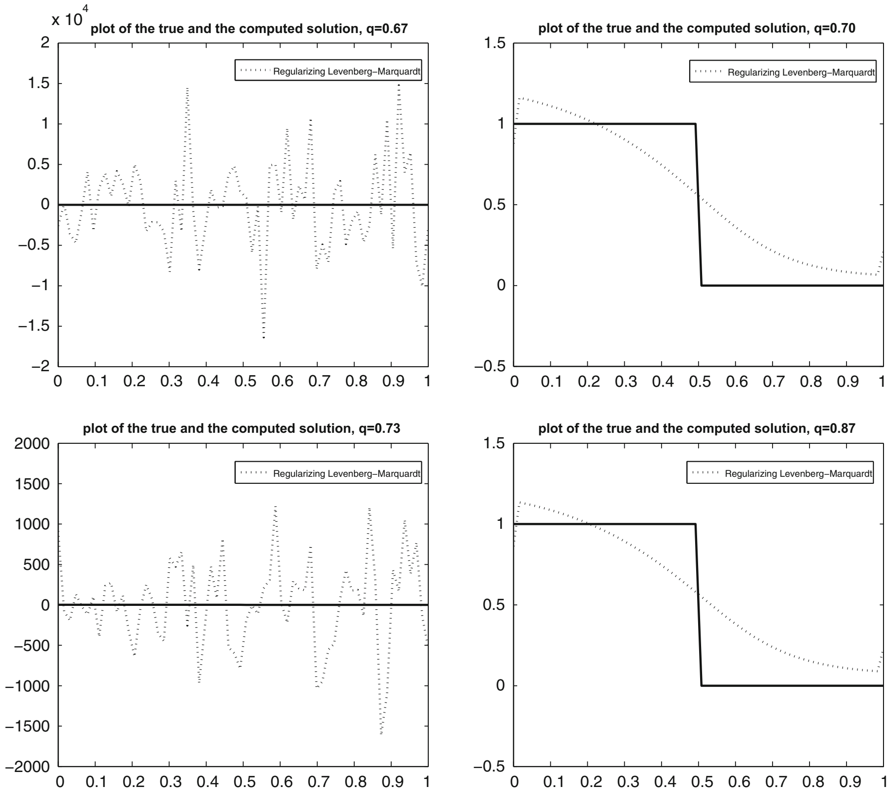

Fig. 5 Problem P4, $\delta = 10^{-2}$, $x_0 = x_0(\beta, \chi) = x_0(1.5, 0)$: true solution (*solid line*) and approximate solution (*dotted line*) computed by the regularizing Levenberg–Marquardt method for values of $q = 0.67,\ 0.70,\ 0.73,\ 0.87$

stronger regularizing properties. In order to support this claim, in Fig. 5 we report four solution approximations computed by the Levenberg–Marquardt algorithm for varying values of q, i.e. $q = 0.67, 0.70, 0.73, 0.87$. It is evident that the method is highly sensitive to the choice of the parameter q and the quality of the solution approximation does not steadily improves as q increases.

We conclude this section considering the standard trust-region strategy. It is well-known that the standard updating rule promotes the use of inactive trust-regions, at least in the late stage of the procedure. Clearly, this can adversely affect the solution of our test problems and our experiments confirmed this fact. In particular, for $\delta = 10^{-2}$ and problems P1 and P2, the sequences computed by the standard trust-region method approach solutions of the noisy problem. The same behaviour occurs in most of the runs with P1 and P2 and noise level $\delta = 10^{-4}$. Conversely, the approximations provided by the regularizing trust-region procedure are accurate approximations of true solutions in all the tests. The approximations computed by the standard trust-region applied to problems P3 and P4 are less accurate than those computed by the regularizing trust-region although they do not show the strong oscillatory behaviour arising in problems

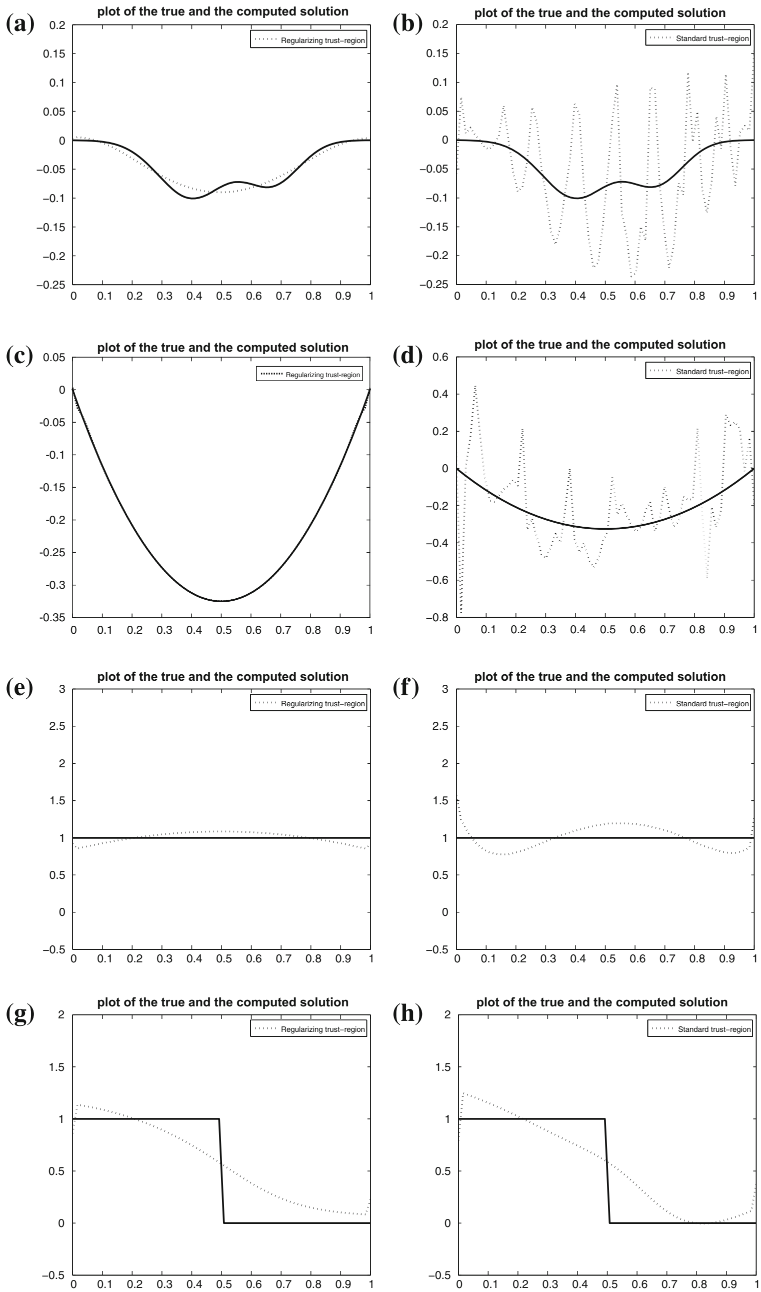

Fig. 6 True solution (*solid line*) and approximate solutions (*dotted line*) computed by the regularizing trust-region method (*on the left*) and the standard trust-region method (*on the right*). **a**, **b** problem P1, $\delta = 10^{-2}$, $x_0 = 0e$; **c**, **d** problem P2, $\delta = 10^{-2}$, $x_0 = 0e$; **e**, **f** problem P3, $\delta = 10^{-2}$, $x_0 = x_0(1.25)$; **g**, **h** problem P4, $\delta = 10^{-2}$, $x_0 = x_0(0.5, 0)$

P1 and P2. In problem P4, this behaviour is evident when the second, third and fourth starting guesses are used, while the approximation computed starting from the first initial guess is as accurate as the one computed by the regularizing trust-region. This good result of the standard trust-region on problem P4 with $x_0 = x_0(1, 1)$ is due to the fact that the trust-region is active in all iterations and therefore a regularizing behaviour is implicitly provided. As an example in Fig. 6 we compare some solution approximations computed by the regularizing trust-region (left) and by the standard trust-region (right) with $\delta = 10^{-2}$ applied to problem P1 (figures (a), (b)), P2 (figures (c), (d)), P3 (figures (e), (f)) and P4 (figures (g), (h)).

6 Conclusions

We have presented a trust-region method for nonlinear ill-posed systems, possibly with noisy data, where the regularizing behaviour is guaranteed by a suitable choice of the trust-region radius. The proposed approach shares the same local convergence properties as the regularizing Levenberg–Marquardt method proposed by Hanke in [12] but it is more likely to satisfy the discrepancy principle irrespective of the closeness of the initial guess to a solution of (1.1). The numerical experience presented confirms the effectiveness of the trust-region radius adopted and the regularizing properties of the resulting trust-region method. It also enlights that the new approach is less sensitive than the regularizing Levenberg–Marquardt method to the choice of the parameter q involved in the regularizations (2.4) and (2.16). Finally, numerical experience confirms that the solution of the noisy problems may be misinterpreted by the standard trust-region method.

Acknowledgments Work partially supported by INdAM-GNCS, under the 2015 Project “Metodi di regolarizzazione per problemi di ottimizzazione e applicazioni”.

References

1. Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim. Lett. **6**, 927–940 (2012)
2. Bellavia, S., Cartis, C., Gould, N.I.M., Morini, B., Toint, Ph.L.: Convergence of a regularized euclidean residual algorithm for nonlinear least-squares. SIAM J. Numer. Anal. **48**, 1–29 (2010)
3. Bellavia, S., Morini, B.: Strong local convergence properties of adaptive regularized methods for nonlinear least-squares. IMA J. Numer. Anal. **35**, 947–968 (2015)
4. Cartis, C., Gould, N.I.M., Toint, Ph.L.: Trust-region and other regularizations of linear least-squares problems. BIT **49**, 21–53 (2009)
5. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. SMPS/SIAM Series on Optimization. SIAM, Philadelphia (2000)
6. de Sturler, E., Kilmer, M.: A regularized Gauss–Newton trust-region approach to imaging in diffuse optical tomography. SIAM J. Sci. Comput. **34**, 3057–3086 (2011)
7. Fan, J.Y.: Convergence rate of the trust-region method for nonlinear equations under local error bound condition. Comput. Optim. Appl. **34**, 215–227 (2005)
8. Fan, J.Y., Pan, J.Y.: A modified trust-region algorithm for nonlinear equations with updating rule for trust-region radius. Int. J. Comput. Math. **87**, 3186–3195 (2010)
9. Fan, J.Y., Pan, J.Y.: An improved trust-region algorithm for nonlinear equations. Comput. Optim. Appl. **48**, 59–70 (2011)

10. Gasparo, M.G., Papini, A., Pasquali, A.: A two-stage method for nonlinear inverse problems. J. Comput. Appl. Math. **198**, 471–482 (2007)
11. Groetsch, C.V.: The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. Pitman Advanced Publishing Program, Boston (1984)
12. Hanke, M.: Regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems. Inverse Prob. **13**, 79–95 (1997)
13. Hanke, M.: The regularizing Levenberg–Marquardt scheme is of optimal order. J. Integral Equ. Appl. **22**, 259–283 (2010)
14. Henrici, P.: Elements of Numerical Analysis. Wiley, Chicester (1964)
15. Kaltenbacher, B.: Toward global convergence for strongly nonlinear ill-posed problems via a regularizing multilevel approach. Numer. Funct. Anal. Optim. **27**, 637–665 (2006)
16. Kaltenbacher, B., Neubauer, A., Scherzer, O.: Iterative Regularization Methods for Nonlinear Ill-Posed Problems. Walter de Gruyter, Berlin (2008)
17. Kanzow, C., Yamashita, N., Fukushima, M.: Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints. J. Comput. Appl. Math. **172**, 375–397 (2004)
18. Levenberg, K.: A method for the solution of certain nonlinear problems in least-squares. Q. Appl. Math. **2**, 164–168 (1944)
19. Macconi, M., Morini, B., Porcelli, M.: Trust-region quadratic methods for nonlinear systems of mixed equalities and inequalities. Appl. Numer. Math. **59**(5), 859–876 (2009)
20. Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters. SIAM J. Appl. Math. **11**, 431–441 (1963)
21. Moré, J.J.: The Levenberg–Marquardt Algorithm: Implementation and Theory. In: Proceedings of 7th Biennial Conference, University of Dundee, Dundee, 1977, Lecture Notes in Mathematics, vol. 630, pp. 105–116. Springer, Berlin (1978)
22. Morozov, V.A.: On the solution of functional equations by the method of regularization. Sov. Math. Dokl. **7**, 414–417 (1996)
23. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer Series in Operations Research. Springer, New York (1999)
24. Toint, Ph.L.: Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. Optim. Methods Softw. **28**, 82–95 (2013)
25. Vogel, C.R.: A constrained least squares regularization method for nonlinear ill-posed problems. SIAM J. Control Optim. **28**, 34–49 (1990)
26. Vogel, C.R.: Computational Methods for Inverse Problems. Frontiers in Applied Mathematics. SIAM, Providence, RI (2002)
27. Wang, Y., Yuan, Y.: Convergence and regularity of trust region methods for nonlinear ill-posed problems. Inverse Prob. **21**, 821–838 (2005)
28. Wang, Y., Yuan, Y.: On the regularity of trust region-CG algorithm for nonlinear ill-posed inverse problems with application to image deconvolution problem. Sci. China Ser. A **46**, 312–325 (2003)
29. Zhang, J.L., Wang, Y.: A new trust region method for nonlinear equations. Math. Methods Oper. Res. **58**, 283–298 (2003)