JOURNAL OF INTEGRAL EQUATIONS
AND APPLICATIONS
Volume 22, Number 2, Summer 2010

THE REGULARIZING LEVENBERG-MARQUARDT SCHEME IS OF OPTIMAL ORDER

MARTIN HANKE

Communicated by Kendall Atkinson

ABSTRACT. We prove that the regularizing Levenberg-Marquardt scheme, introduced for nonlinear ill-posed problems, achieves order optimal accuracy under standard assumptions on the nonlinearity of the underlying operator equation.

1. Introduction. The Levenberg-Marquardt method is a Newton-type method for nonlinear least-squares problems that is treated in many numerical optimization textbooks, cf., e.g., Kelley [**14**]. In each iteration of the Levenberg-Marquardt method the nonlinear operator is linearized around the current approximation, and the original problem is turned into a linear least squares problem with a quadratic inequality constraint. This constraint is derived on the grounds that one can only trust in the linearization within a certain neighborhood of the present approximation (*trust region*). Eventually, this leads to the same linear equation to be solved as in Tikhonov regularization, the corresponding regularization parameter being coupled with the Lagrange parameter associated with the constrained problem.

The Levenberg-Marquardt method is often used as a black box method for parameter identification problems, regardless of whether these are well-posed or not. Its convergence analysis, however, relies on the assumption that the derivative of the nonlinear operator is continuously invertible near the exact solution, cf. [**14**], which irrevocably fails to hold for ill-posed problems. To adjust the method to ill-posed problems we therefore proposed a modification of the Levenberg-Marquardt approach back in 1997 [**6**], to be called the *regularizing Levenberg-Marquardt scheme* below, which uses a different quadratic constraint that assesses the reliability of the right-hand side of the linearized problem rather than the trust region of the linearization. With

2010 AMS *Mathematics subject classification.* Primary 65J15, 65J20.
Keywords and phrases. Levenberg-Marquardt method, nonlinear ill-posed problems, regularization method, discrepancy principle, order optimal convergence rates.
Received by the editors on July 28, 2009.

DOI:10.1216/JIE-2010-22-2-259

this approach the associated Tikhonov regularization parameter corresponds to the well-known discrepancy principle from the ill-posed problems literature, cf., e.g., Groetsch [**4**].

It was shown in [**6**] that this Levenberg-Marquardt variant is a regularization method under mild restrictions on the nonlinearity of the operator equation, but no convergence rates were provided in [**6**]. Later, Rieder [**15, 16**] proved convergence rates (for more general schemes, actually, that include the Levenberg-Marquardt iteration as one special case) which, however, are suboptimal as compared to other methods. One of these competing methods, for which order-optimal convergence rates are known, is the so-called iteratively regularized Gauß-Newton method by Bakushinskii [**1**]. We refer to the monograph by Kaltenbacher, Neubauer and Scherzer [**13**] for a comprehensive compilation of its convergence analysis, and to the paper by Jin and Tautenhahn [**12**] for a more recent relevant contribution.

When compared with the Levenberg-Marquardt scheme, the iteratively regularized Gauß-Newton method suffers from the somewhat counterintuitive stipulation that its iterates stay near the initial guess–in the Levenberg-Marquardt terminology this would correspond to a trust region centered around the initial guess, rather than the current approximation. Accordingly, the analysis of the regularizing Levenberg-Marquardt scheme has received renewed interest recently: Jin [**11**] and Hochbruck and Hönig [**9**] reconsidered the method under additional assumptions that constrain the decay of the regularization parameters of the Tikhonov steps. Under these circumstances they could establish order optimal convergence rates, with similar restrictions on the nonlinearity of the operator as are omnipresent in [**13**], for example. These results, however, do not cover the original algorithm from [**6**], as it is not clear whether the corresponding regularization parameters will actually satisfy these constraints.

It is the purpose of this note to eventually dissolve the uncertainty concerning the order-optimality of the regularizing Levenberg-Marquardt scheme, and to raise the method on comparable solid theoretical grounds as the iteratively regularized Gauß-Newton method; the precise statement of our result will be given at the end of the following section, once the setting of the problem has been completely specified. Our analysis follows to a certain extent the line of argument given in [**9, 11**], which in turn goes back to the treatment of the nonlinear Landweber iteration in [**8**]. The details of our proof, however, are different from [**9**], as we avoid any constraints on the rate of decay of the regularization parameters.

To present the proof we first recollect some basic inequalities in Section 3 which are more or less familiar from related works in this context. Then, in Section 4 we set up the basic recursion of the iteration error and provide a crucial lemma on the size of the accumulated perturbations from the individual iterations. The main induction argument, finally, can be found in Section 5; after that it only remains to gather the pieces of the puzzle to establish order-optimality. We will do so in Section 6.

Last but not least, it is our pleasure to use the occasion to point out that our analysis of the regularizing Levenberg-Marquardt scheme (like the one in [**9, 11**]) is built on earlier results by Chuck Groetsch and this author ([**7**], see also [**5**]) on the nonstationary iterated Tikhonov regularization method for *linear* ill-posed problems. Once again, this reflects the substantial influence that Chuck's work has had on the present shape of this fascinating field.

2. Setting of the problem. We consider the operator equation

$$F(x) = y, \tag{2.1}$$

where $F : \mathcal{D}(F) \subset \mathcal{X} \to \mathcal{Y}$ is a differentiable nonlinear operator between the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, and $\mathcal{D}(F)$ denotes the domain of F. Without loss of generality we assume that problem (2.1) has been scaled such that

$$\|F'(x)\| \le 1 \tag{2.2}$$

within the relevant subset of $\mathcal{D}(F)$. Moreover, as in [**9, 11**], we presume that F satisfies the so-called *strong Scherzer condition*, namely, that for any two elements $x, \widetilde{x} \in \mathcal{D}(F)$ there is a bounded and linear operator $R(x,\widetilde{x}) : \mathcal{X} \to \mathcal{Y}$ such that

$$F'(x) = R(x,\widetilde{x})F'(\widetilde{x}), \tag{2.3}$$

where

$$\|R(x,\widetilde{x}) - I\| \le C_R\,\|x - \widetilde{x}\| \tag{2.4}$$

for some constant $C_R > 0$. This assumption, which appeared in [**8**] for the first time, is now a standard one for proving convergence rates for nonlinear ill-posed problems. It ensures that all derivative operators $F'(x) : \mathcal{X} \to \mathcal{Y}$ can be continuously extended to the same maximal domain $\mathcal{Z} \supset \mathcal{X}$, i.e., have similar smoothing properties, see Proposition 3.1 below. To be precise, in [**8**] assumption (2.3) has only been required for x near $\widetilde{x} = x^\dagger$, which was the particular solution of (2.1) to be approximated, and it is precisely this way that we are going to utilize this assumption here, too.

When the problem is ill-posed, the solution of (2.1) does not depend continuously on the given data, generically. Accordingly, if we are given perturbed data y^δ instead of y in (2.1) satisfying

$$\|y^\delta - y\| \le \delta, \tag{2.5}$$

then we need to regularize the problem in order to compute approximate solutions x^δ that converge to some solution of (2.1) as $\delta \to 0$. To this end the regularizing Levenberg-Marquardt scheme of [**6**] proceeds iteratively by solving regularized and linearized subproblems to update a given iterate:

$$x_{n+1}^\delta = x_n^\delta + \big(F'(x_n^\delta)^* F'(x_n^\delta) + \alpha_n I\big)^{-1} F'(x_n^\delta)^* \big(y^\delta - F(x_n^\delta)\big). \tag{2.6}$$

The recursion starts for $n = 0$ with some initial guess $x_0^\delta = x_0$, independent of δ, and is stopped after $n(y^\delta)$ iterations according to the *discrepancy principle*, i.e., when the residual drops for the first time below the noise level, that is,

$$\|y^\delta - F(x_{n(y^\delta)}^\delta)\| \le \tau\delta < \|y^\delta - F(x_n^\delta)\|, \quad n < n(y^\delta). \tag{2.7}$$

Here, τ is a fudge parameter that was set to be $\tau = 2.5$ in numerical examples presented in [**6**]; below we require that

$$\tau > 2. \tag{2.8}$$

The only essential difference to the classical Levenberg-Marquardt method from nonlinear optimization is in the choice of the regularization parameters α_n in (2.6). Here they are chosen such that

$$\|y^\delta - F(x_n^\delta) - F'(x_n^\delta)(x_{n+1}^\delta - x_n^\delta)\| = \rho\|y^\delta - F(x_n^\delta)\| \tag{2.9}$$

for some preassigned nonnegative value $0 < \rho < 1$, which resembles the spirit of an inexact Newton scheme, cf. [**14**].

It is known, see [**8**], that (2.3), (2.4) imply

$$\|F(x) - F(\widetilde{x}) - F'(\widetilde{x})(x - \widetilde{x})\| \leq \frac{3}{2}\, C_R \|x - \widetilde{x}\|\, \|F'(\widetilde{x})(x - \widetilde{x})\|, \tag{2.10}$$

and hence

$$\|F(x) - F(\widetilde{x}) - F'(\widetilde{x})(x - \widetilde{x})\| \leq 2C_R \|x - \widetilde{x}\|\, \|F(x) - F(\widetilde{x})\|, \tag{2.11}$$

provided that x and $\widetilde{x}$ are sufficiently close. As the latter is the assumption on F that has been used for the convergence analysis of the regularizing Levenberg-Marquardt scheme in [**6**], we conclude that if we constrain the two parameters τ from (2.7) and ρ from (2.9) by

$$\rho\tau > 1 \tag{2.12}$$

as in [**6**], and if the initial guess is sufficiently close to a solution of (2.1) in the interior of $\mathcal{D}(F)$, then the method (2.6), (2.9) is well-defined, the parameters α_n of (2.6) are all strictly positive, and, for $\delta > 0$, the method terminates after a finite number of steps. Moreover, the iteration is a regularization method in the sense of Tikhonov, i.e., the final iterate $x^\delta_{n(y^\delta)}$ corresponding to the data y^δ of (2.5) converges to some solution $x^\dagger$ of (2.1) as $\delta \to 0$. Even more, for a fixed value of δ and associated data y^δ the norm of the error $x^\delta_n - x^\dagger$ is strictly decreasing from $n = 0$ up to $n = n(y^\delta)$.

We note, see [**8**, Proposition 2.1] or the discussion in [**13**, page 10], that (2.11) implies that the set of solutions of (2.1) within a certain ball around x_0 is the intersection of this ball with an affine subspace parallel to the null space $\mathcal{N}$ of the derivative at all these solutions. Moreover, because of our accentuation of the assumptions from [**6**], namely that F satisfies (2.3), we observe that the update of x^δ_n in (2.6) always belongs to $\mathcal{R}(F'(x^\delta_n)^*) = \mathcal{R}(F'(x^\dagger)^*) = \mathcal{N}^\perp$, and hence, $x^\dagger$ is the x_0-*minimum-norm solution* of (2.1), i.e., the solution of (2.1) that has minimal distance to x_0. In other words, $\|x^\dagger - x_0\|$ is the distance between x_0 and the set of solutions of (2.1), the size of which will play a crucial role in our analysis via the quantity

$$\eta = C_R\, \|x_0 - x^\dagger\|. \tag{2.13}$$

We will also use repeatedly that (2.13) implies that

$$C_R\|x_n^\delta - x^\dagger\| \le \eta \quad \text{for all } 0 \le n \le n(y^\delta)$$

by virtue of the monotonicity of the iteration error.

As mentioned in the introduction, much less has been known so far concerning the rate of convergence of x_n^δ to $x^\dagger$ as $\delta \to 0$. Of course, it is known that in general the convergence can be arbitrarily slow when the problem is ill-posed, so that the initial guess is usually constrained to satisfy a so-called *source condition*

$$x_0 - x^\dagger = \big(F'(x^\dagger)^* F'(x^\dagger)\big)^\nu w \tag{2.14}$$

for some $\nu > 0$ and associated $w \in \mathcal{X}$ to achieve a given convergence rate. It is said that a regularization method is of *optimal order* (for a given value of ν), if (2.14) implies that

$$\|x_{n(y^\delta)}^\delta - x^\dagger\| \le C\|w\|^{1/(2\nu+1)}\delta^{2\nu/(2\nu+1)} \tag{2.15}$$

for some constant $C > 0$. If F were a linear operator then (2.15) is the best possible general bound if nothing else than (2.14) is known about $x^\dagger - x_0$, cf., e.g., [**2**].

Convergence rates for the regularizing Levenberg-Marquardt scheme had been established in [**15, 16**], but with exponents of δ that are worse than the one in (2.15). In [**9**], on the other hand, the order optimal bound (2.15) has recently been verified provided that the regularization parameters α_n determined via (2.9) do not decay too fast; unfortunately, it has been left open whether this condition is reasonable, or even generally true. As we will see below (see Proposition 5.3), such an assumption is indeed reasonable, but probably not fully correct. Here we prove the order-optimality of the method without any additional constraints:

Theorem 2.1. *Let F satisfy* (2.2) *and* (2.3), (2.4), *and assume that for a given initial guess x_0 the x_0-minimum-norm solution $x^\dagger$ of $F(x) = y$ satisfies* (2.14) *for some $0 < \nu \le 1/2$. Moreover, let $y^\dagger$ fulfill* (2.5), *and denote by $n(y^\delta)$ of* (2.7) *the well-defined stopping index of the regularizing Levenberg-Marquardt scheme* (2.6), (2.9), *where $\tau > 2$*

and $1 > \rho > 1/\tau$. *Then there is a constant* C *depending only on* τ, *such that*

$$\|x^\delta_{n(y^\delta)} - x^\dagger\| \leq C\|w\|^{1/(2\nu+1)}\delta^{2\nu/(2\nu+1)},$$

provided that $\|w\|$ *is sufficiently small.*

The following three sections of the paper are devoted to prepare for a proof of this result, which will be given in Section 6, eventually.

3. Some useful auxiliary results. For our analysis of the regularizing Levenberg-Marquardt scheme we shall first simplify our notation by setting

$$T_n = F'(x^\delta_n), \quad T = F'(x^\dagger), \quad \text{and} \quad R_n = R(x^\delta_n, x^\dagger).$$

Also, we denote the error of the iteration by

$$e_n = x^\delta_n - x^\dagger.$$

Next we provide the interpretation of the strong Scherzer condition (2.3), (2.4), that we have mentioned above.

Proposition 3.1. *Assume that* (2.3) *holds true for some operators* $R(x, \widetilde{x}) : \mathcal{X} \to \mathcal{Y}$ *that are uniformly bounded by* c_0, *together with their inverses, for* x *within a neighborhood* $\mathcal{B}(\widetilde{x})$ *of some interior point* $\widetilde{x}$ *of* $\mathcal{D}(F)$. *If* $e \in \mathcal{X}$ *satisfies*

$$e = \big(F'(\widetilde{x})^* F'(\widetilde{x})\big)^\nu w$$

for some $0 < \nu \leq 1/2$ *and* $w \in \mathcal{X}$, *then for each* $x \in \mathcal{B}(\widetilde{x})$ *there is a* $w_x \in \mathcal{X}$, *such that*

$$e = \big(F'(x)^* F'(x)\big)^\nu w_x, \quad \textit{with } \|w_x\| \leq c_0^{2\nu}\|w\|.$$

Proof. By virtue of (2.3) and the given assumptions, the operators $F'(x)$ all have the same null space $\mathcal{N}$ as long as x belongs to the given neighborhood $\mathcal{B}(\widetilde{x})$, and this is also the null space of $\big(F'(x)^* F'(x)\big)^\nu$

for every $\nu > 0$. This implies that e belongs to $\mathcal{N}^\perp$, and it suffices to prove the result for $w \in \mathcal{N}^\perp$ only. In other words, we can assume without loss of generality that $\mathcal{N} = \{0\}$ is trivial.

Then, consider any $z \in \mathcal{D}((F'(\widetilde{x})^*F'(\widetilde{x}))^{-1/2}) \subset \mathcal{X}$. From (2.3) it follows that

$$\begin{aligned}\|(F'(x)^*F'(x))^{-1/2}z\|^2 &= \|F'(x)^{-*}z\|^2 \\ &= \|R(x,\widetilde{x})^{-*}F'(\widetilde{x})^{-*}z\|^2 \\ &\le c_0^2\|F'(\widetilde{x})^{-*}z\|^2 \\ &= c_0^2\|(F'(\widetilde{x})^*F'(\widetilde{x}))^{-1/2}z\|^2.\end{aligned}$$

Using the Heinz inequality (cf., e.g., [**2**, Proposition 8.21]) it follows that

$$\mathcal{D}((F'(\widetilde{x})^*F'(\widetilde{x}))^{-\nu}) \subset \mathcal{D}((F'(x)^*F'(x))^{-\nu})$$

for all $0 < \nu \le 1/2$, and that

$$\|(F'(x)^*F'(x))^{-\nu}z\| \le c_0^{2\nu}\|(F'(\widetilde{x})^*F'(\widetilde{x}))^{-\nu}z\|$$

for all $z \in \mathcal{D}((F'(\widetilde{x})^*F'(\widetilde{x}))^{-\nu})$.

By assumption, we can set $z = e$ and have thus shown that e belongs to $\mathcal{D}((F'(x)^*F'(x))^{-\nu})$, and that $w_x = (F'(x)^*F'(x))^{-\nu}e$ satisfies $\|w_x\| \le c_0^{2\nu}\|w\|$. □

For ill-posed problems the inverse (or the Moore-Penrose generalized inverse) of $F'(x)^*F'(x)$ will typically be an unbounded, densely defined operator on $\mathcal{X}$, cf., e.g., Groetsch [**3**]. Accordingly, a source condition of type (2.14) is a restrictive condition (the more, the larger is ν), which can often be interpreted as an a priori smoothness assumption, cf. [**2**] for examples. As shown in Proposition 3.1, the assumption (2.3) on F ensures that the smoothing effect of $(F'(x)^*F'(x))^\nu$ is independent of the particular element $x \in \mathcal{X}$.

Later on, in Proposition 5.3, we will utilize this result to derive a reasonably sharp lower bound for the regularization parameters α_n of (2.6), (2.9). For the proof of Theorem 2.1, however, it suffices to recall a useful and well-known *upper* bound for α_n.

Proposition 3.2. *Let F satisfy* (2.2). *Then $\alpha_n \le \rho/(1-\rho)$ holds for every $n = 0, 1, \dots, n(y^\delta) - 1$.*

Proof. For the (comparatively simple) argument we refer to [**6**, page 81]. □

We also need a relation between the nonlinear residual $y^\delta - F(x_n^\delta)$ and its linearized counterpart Te_n. Results of this type can be found in various works on iterative methods for nonlinear ill-posed problems, differing in the particular schemes that are used, and in the respective assumptions on F.

Proposition 3.3. *If* η *of* (2.13) *satisfies* $\eta \leq 1/2$, *then*

$$\frac{1}{6}\,\|Te_n\| \leq \|y^\delta - F(x_n^\delta)\| \leq c_1\|Te_n\|, \quad n < n(y^\delta), \tag{3.1}$$

where

$$c_1 = \left(1 + \frac{3}{2}\eta\right)\frac{\tau}{\tau - 1} \leq \frac{7}{2}.$$

Proof. For $n < n(y^\delta)$ the nonlinear residual $y^\delta - F(x_n^\delta)$ can be estimated by means of (2.5) and (2.7) as follows:

$$\begin{aligned}\|F(x^\dagger) - F(x_n^\delta)\| &\geq \|y^\delta - F(x_n^\delta)\| - \delta\\ &\geq \|y^\delta - F(x_n^\delta)\| - \frac{1}{\tau}\,\|y^\delta - F(x_n^\delta)\|\\ &= \frac{\tau - 1}{\tau}\,\|y^\delta - F(x_n^\delta)\|.\end{aligned}$$

Similarly, one obtains

$$\|y^\delta - F(x_n^\delta)\| \geq \frac{\tau}{\tau + 1}\,\|F(x^\dagger) - F(x_n^\delta)\| \geq \frac{2}{3}\,\|F(x^\dagger) - F(x_n^\delta)\|$$

by virtue of (2.8). On the other hand we have from (2.10) that

$$\frac{1}{4}\,\|Te_n\| \leq \left(1 - \frac{3}{2}\,\eta\right)\|Te_n\| \leq \|F(x^\dagger) - F(x_n^\delta)\| \leq \left(1 + \frac{3}{2}\,\eta\right)\|Te_n\|,$$

and hence, (3.1) follows. □

Another result relates subsequent linearized residuals.

Lemma 3.4. *Assume that* η *of* (2.13) *satisfies* $\eta \le 1/(4+8\tau)$. *Then the following holds*

$$\|Te_{n-1}\| \le 42\tau\,\|Te_n\|, \quad 0 < n < n(y^\delta).$$

Proof. Because of (2.12) and our assumptions the following holds

$$\eta \le \frac{\rho}{4\rho+8} < \frac{\rho}{8}. \tag{3.2}$$

Also, (2.9), together with (2.11), yield

$$\begin{aligned}
\rho\,\|y^\delta - F(x_{n-1}^\delta)\| \\
&\le \|y^\delta - F(x_n^\delta)\| + \|F(x_n^\delta) - F(x_{n-1}^\delta) - T_{n-1}(x_n^\delta - x_{n-1}^\delta)\| \\
&\le \|y^\delta - F(x_n^\delta)\| + 2C_R\,\|x_n^\delta - x_{n-1}^\delta\|\,\|F(x_n^\delta) - F(x_{n-1}^\delta)\| \\
&\le \|y^\delta - F(x_n^\delta)\| + 4\eta\,\|F(x_n^\delta) - F(x_{n-1}^\delta)\| \\
&\le \|y^\delta - F(x_n^\delta)\| + 4\eta\,\|y^\delta - F(x_n^\delta)\| + 4\eta\|y^\delta - F(x_{n-1}^\delta)\|,
\end{aligned}$$

and hence, as $\rho - 4\eta > 0$ by virtue of (3.2),

$$\|y^\delta - F(x_{n-1}^\delta)\| \le \frac{1+4\eta}{\rho-4\eta}\,\|y^\delta - F(x_n^\delta)\| \le \frac{2}{\rho}\,\|y^\delta - F(x_n^\delta)\|. \tag{3.3}$$

The result now follows immediately from (2.12) and Proposition 3.3. Note that (3.3) and its derivation remain true even for $n = n(y^\delta)$. $\square$

4. Analysis of the regularizing Levenberg-Marquardt scheme. Before we proceed we introduce the spectral filter functions

$$r_j(\lambda) = \frac{\alpha_j}{\lambda+\alpha_j} \quad \text{and} \quad g_j(\lambda) = \frac{1}{\lambda+\alpha_j} \tag{4.1}$$

associated with Tikhonov regularization, cf. [**2**, Chapter 5] or [**4**]. Then we obtain from (2.6) that the error e_n satisfies the recursion

$$\begin{aligned}
e_{n+1} &= e_n + (T^*T+\alpha_n I)^{-1}T^*\bigl(y - F(x_n^\delta)\bigr)\\
&\qquad + (T^*T+\alpha_n I)^{-1}T^*(y^\delta - y)\\
&\qquad + \bigl((T_n^*T_n+\alpha_n I)^{-1}T_n^* - (T^*T+\alpha_n I)^{-1}T^*\bigr)\bigl(y^\delta - F(x_n^\delta)\bigr)\\
&= r_n(T^*T)e_n - g_n(T^*T)T^*\bigl(F(x_n^\delta) - F(x^\dagger) - Te_n\bigr)\\
&\qquad + g_n(T^*T)T^*(y^\delta - y)\\
&\qquad + \bigl((T_n^*T_n+\alpha_n I)^{-1}T_n^* - (T^*T+\alpha_n I)^{-1}T^*\bigr)\bigl(y^\delta - F(x_n^\delta)\bigr).
\end{aligned}$$

By means of (2.3) we can rewrite

$$\begin{aligned}
(T_n^*T_n+\alpha_n I)^{-1}T_n^* &- (T^*T+\alpha_n I)^{-1}T^*\\
&= g_n(T^*T)T^*(R_n^* - I)r_n(T_nT_n^*)\\
&\quad + g_n(T^*T)T^*(R_n^{-1} - I)T_nT_n^*g_n(T_nT_n^*),
\end{aligned}$$

provided that η of (2.13) is less than one, in which case R_n is invertible by virtue of (2.4). Inserting this into the recursion we therefore obtain

$$e_{n+1} = r_n(T^*T)e_n + g_n(T^*T)T^*(y^\delta - y) + g_n(T^*T)T^*z_n \tag{4.2}$$

with

$$\begin{aligned}
z_n &= (R_n^* - I)r_n(T_nT_n^*)\bigl(y^\delta - F(x_n^\delta)\bigr)\\
&\quad + (R_n^{-1} - I)T_nT_n^*g_n(T_nT_n^*)\bigl(y^\delta - F(x_n^\delta)\bigr)\\
&\quad - \bigl(F(x_n^\delta) - F(x^\dagger) - Te_n\bigr).
\end{aligned}\tag{4.3}$$

Lemma 4.1. *Let η of* (2.13) *satisfy $\eta \le 1/2$. Then*

$$\|z_n\| \le 12C_R\,\|e_n\|\|Te_n\|, \quad 0 \le n < n(y^\delta). \tag{4.4}$$

Proof. By virtue of (4.1), both operators $r_n(T_nT_n^*)$ and $T_nT_n^*g_n(T_nT_n^*)$ are bounded by one. Moreover, since $\eta \le 1/2$ it follows from (2.4) that

$$\|R_n^{-1} - I\| \le \frac{C_R\|e_n\|}{1 - C_R\|e_n\|} \le 2C_R\|e_n\|.$$

Inserting these estimates and (2.10) into (4.3) we conclude that

$$\|z_n\| \le 3C_R\,\|e_n\|\|y^\delta - F(x_n^\delta)\| + \frac{3}{2}\,C_R\|e_n\|\|Te_n\|,$$

and the assertion now follows from Proposition 3.3. □

Resolving the recursion (4.2) and inserting (2.14), we obtain the expression

$$\begin{aligned} e_n &= \prod_{j=0}^{n-1} r_j(T^*T)(T^*T)^\nu w + q_n(T^*T)T^*(y^\delta - y) \\ &\quad + \sum_{j=0}^{n-1} T^* g_j(TT^*) \prod_{k=j+1}^{n-1} r_k(TT^*)z_j \end{aligned} \tag{4.5}$$

for $n = 1, \dots, n(y^\delta)$, where

$$q_n(\lambda) = \sum_{j=0}^{n-1} g_j(\lambda) \prod_{k=j+1}^{n-1} r_k(\lambda) = \frac{1}{\lambda}\Big(1 - \prod_{j=0}^{n-1} r_j(\lambda)\Big). \tag{4.6}$$

We mention that the first two terms of the right-hand side of (4.5) correspond to the error of nonstationary Tikhonov regularization for the linear problem $Tx^\dagger = y + u$ with appropriate $u \in \mathcal{Y}$ and given data $y^\delta + u$, and with initial guess x_0 satisfying (2.14). In [**7**] this linear method has been analyzed in detail, and it has been shown that the maxima of the corresponding spectral filter functions depend on the size of

$$s_n = \sum_{j=0}^{n-1} \frac{1}{\alpha_j}. \tag{4.7}$$

More precisely, we have the following bounds.

Lemma 4.2. *Assume that $\alpha_j > 0$ for $j = 0, \dots, n-1$ and some $n \in \mathbb{N}$, and define $s_k = \sum_{j=0}^{k-1} 1/\alpha_j$ for $k = 1, \dots, n$, cf.* (4.7), *and $s_0 = 0$. For $0 \le \mu \le 1$, $-1/2 \le \nu \le 1/2$, and $\lambda \ge 0$ there holds*:

$$0 \le \lambda^\mu \prod_{j=0}^{n-1} r_j(\lambda) \le s_n^{-\mu}, \tag{4.8}$$

$$0 \le \lambda^{1/2-\nu} q_n(\lambda) \le s_n^{\nu+1/2}, \tag{4.9}$$

$$0 \le \lambda^{\mu} g_j(\lambda) \prod_{k=j+1}^{n-1} r_k(\lambda) \le \frac{1}{\alpha_j}(s_n - s_j)^{-\mu}, \quad 0 \le j \le n-1. \tag{4.10}$$

Proof. Estimate (4.8) is taken from [**11**, Lemma 2], see also [**7**] or [**5**, Lemma 4.13]; a slightly stronger bound is given in [**9**, Lemma 2]. Estimate (4.9) with $\nu = 1/2$ corresponds to display (15) in [**7**]; the general case then follows by interpolation: Since $0 \le r_j(\lambda) \le 1$ and $0 \le \lambda q_n(\lambda) \le 1$, we have

$$\lambda^{1/2-\nu} q_n(\lambda) = \big(\lambda q_n(\lambda)\big)^{1/2-\nu} q_n(\lambda)^{\nu+1/2} \le q_n(\lambda)^{\nu+1/2} \le s_n^{\nu+1/2}.$$

Finally, the third estimate, (4.10), follows from the first one after rewriting $g_j(\lambda) = r_j(\lambda)/\alpha_j$. $\square$

The convergence analysis of the regularizing Levenberg-Marquardt scheme in [**9**] stipulates a restriction on the decay of the regularization parameters, namely, that

$$\alpha_n \ge \frac{c}{s_n}, \quad n = 0, \dots, n(y^\delta) - 1, \tag{4.11}$$

for some constant $c > 0$. From this it readily follows that $s_{n+1} \le (1 + 1/c)s_n$, i.e., an at most geometric growth rate of the sequence $\{s_n\}$. Although it is very likely that this assumption holds true for the regularizing Levenberg-Marquardt scheme we have not been able to verify an inequality like (4.11). We can, however, circumvent this assumption in our setting. To this end we require the following lemma, which is the appropriate modification of the corresponding estimate by Hochbruck, Hönig and Ostermann, given in [**10**, Lemma 4.11].

Lemma 4.3. *Let $\nu > 0$ and $0 \le \mu < 1$, and define s_k as in Lemma 4.2. Furthermore, let $\alpha_j \le \rho/(1-\rho)$ for $j = 0, \dots, n-1$. Then there exists a constant $C_s > 0$, depending only on ρ, μ and ν,* *such that for all* $n \in \mathbb{N}$,

$$
\begin{aligned}
&\sum_{j=0}^{n-1} \frac{1}{\alpha_j}(s_n - s_j)^{-\mu} s_{j+1}^{-2\nu-1/2} \\
&\qquad \le C_s \begin{cases} s_n^{1/2-\mu-2\nu} & \nu < 1/4, \\ s_n^{-\mu}\log(1+s_n) & \nu = 1/4, \\ s_n^{-\mu} & \nu > 1/4. \end{cases}
\end{aligned}
\tag{4.12}
$$

In particular, when $\mu = 1/2 - \nu$ *and* ρ *is fixed, then the expression in* (4.12) *is bounded for every* $0 < \nu \le 1/2$.

For $\mu = 1$ *and* $\nu > 0$ *we have instead*

$$
\begin{aligned}
&\sum_{j=0}^{n-1} \frac{1}{\alpha_j}(s_n - s_j)^{-1} s_{j+1}^{-2\nu-1/2} \\
&\qquad \le C_s \begin{cases} s_n^{-1/2-2\nu}\log(1+s_n) & \nu < 1/4, \\ s_n^{-1}\log(1+s_n) & \nu = 1/4, \\ s_n^{-1} & \nu > 1/4, \end{cases}
\end{aligned}
\tag{4.13}
$$

and for $0 < \nu \le 1/2$ *this is always bounded by* $C_s' s_n^{-\nu-1/2}$ *for some* $C_s' > 0$ *depending only on* ρ *and* ν.

Proof. When $n = 1$, the left-hand sides of (4.12), (4.13) equal

$$\frac{1}{\alpha_0}\, s_1^{-2\nu-\mu-1/2} = s_1^{1/2-\mu-2\nu},$$

which is always smaller than the corresponding right-hand side for an appropriate constant C_s, since $s_1 = 1/\alpha_0 \ge (1-\rho)/\rho$.

For $n \ge 2$ we rewrite

$$
\begin{aligned}
&\sum_{j=0}^{n-1} \frac{1}{\alpha_j}(s_n - s_j)^{-\mu} s_{j+1}^{-2\nu-1/2} \\
&\qquad = s_n^{1/2-\mu-2\nu} \sum_{j=0}^{n-1} \frac{1}{\alpha_j s_n}\left(1 - \frac{s_j}{s_n}\right)^{-\mu} \left(\frac{s_{j+1}}{s_n}\right)^{-2\nu-1/2},
\end{aligned}
$$

and observe that $0 \le s_j/s_n < s_{j+1}/s_n \le 1$ for $0 \le j \le n-1$. Accordingly, we can estimate

$$\left(1-\frac{s_j}{s_n}\right)^{-\mu}\left(\frac{s_{j+1}}{s_n}\right)^{-2\nu-1/2} \le \begin{cases} 2^{\mu}(\frac{s_{j+1}}{s_n})^{-2\nu-1/2} & 0 \le s_j/s_n \le 1/2, \\ 2^{2\nu+1/2}(1-\frac{s_j}{s_n})^{-\mu} & 1/2 < s_j/s_n \le 1, \end{cases}$$

and hence, with $\beta = \max\{\mu, 2\nu+1/2\}$,

$$\sum_{j=0}^{n-1}\frac{1}{\alpha_j}(s_n-s_j)^{-\mu}s_{j+1}^{-2\nu-1/2} \le 2^{\beta}s_n^{1/2-\mu-2\nu}\left(\sum_{j=0}^{n-1}\frac{1}{\alpha_j s_n}\left(\frac{s_{j+1}}{s_n}\right)^{-2\nu-1/2} + \sum_{j=0}^{n-1}\frac{1}{\alpha_j s_n}\left(1-\frac{s_j}{s_n}\right)^{-\mu}\right). \tag{4.14}$$

One can interpret the first sum in (4.14) excluding its first term ($j=0$) as a rectangular quadrature rule for the integral

$$\int_{s_1/s_n}^{1} x^{-2\nu-1/2}\mathrm{d}x,$$

whereas its first term can be considered a mid-point quadrature approximation for the integral

$$\int_{s_1/s_n-h}^{s_1/s_n+h} x^{-2\nu-1/2}\mathrm{d}x \quad \text{with} \quad h = \frac{1}{2\alpha_0 s_n}.$$

The integrand being decreasing and convex, these integrals are actually upper bounds of the respective terms of the sum, and hence,

$$\begin{aligned}\sum_{j=0}^{n-1}\frac{1}{\alpha_j s_n}\left(\frac{s_{j+1}}{s_n}\right)^{-2\nu-1/2} &\le \int_{s_1/s_n}^{1} x^{-2\nu-1/2}\mathrm{d}x \\ &\quad + \int_{s_1/s_n-h}^{s_1/s_n+h} x^{-2\nu-1/2}\mathrm{d}x \\ &\le 2\int_{s_1/s_n-h}^{2} x^{-2\nu-1/2}\mathrm{d}x,\end{aligned}$$

where we have used that $h < 1$. Concerning the left endpoint of the last interval of integration we have

$$\frac{s_1}{s_n} - h = \frac{1}{\alpha_0 s_n} - \frac{1}{2\alpha_0 s_n} = \frac{1}{2\alpha_0 s_n},$$

and hence, we finally arrive at

$$\sum_{j=0}^{n-1} \frac{1}{\alpha_j s_n} \left(\frac{s_{j+1}}{s_n} \right)^{-2\nu-1/2} = \begin{cases} O(1) & \nu < 1/4, \\ O(\log(1+s_n)) & \nu = 1/4, \\ O(s_n^{2\nu-1/2}) & \nu > 1/4. \end{cases} \tag{4.15}$$

Since $\alpha_0 \leq \rho/(1-\rho)$ by assumption, the constant in the $O(\,\cdot\,)$ notation only depends on ρ and ν,

Now we turn to the second sum in (4.14). Excluding its last term where $j = n-1$ this sum can be interpreted as a rectangular quadrature rule for the integral

$$\int_0^{s_{n-1}/s_n} (1-x)^{-\mu} \mathrm{d}x,$$

whereas its last term is the mid-point quadrature rule for the integral

$$\int_{s_{n-1}/s_n - h}^{s_{n-1}/s_n + h} (1-x)^{-\mu} \mathrm{d}x \quad \text{with} \quad h = \frac{1}{2\alpha_{n-1} s_n} < 1.$$

Again, the two integrals yield an upper bound for the sum, i.e.,

$$\begin{aligned} \sum_{j=0}^{n-1} \frac{1}{\alpha_j s_n} \left(1 - \frac{s_j}{s_n} \right)^{-\mu} &\leq \int_0^{s_{n-1}/s_n} (1-x)^{-\mu} \mathrm{d}x \\ &\quad + \int_{s_{n-1}/s_n - h}^{s_{n-1}/s_n + h} (1-x)^{-\mu} \mathrm{d}x \\ &\leq 2 \int_{-1}^{s_{n-1}/s_n + h} (1-x)^{-\mu} \mathrm{d}x. \end{aligned}$$

For $0 \leq \mu < 1$ this integral is bounded by some constant depending only on μ. For $\mu = 1$, on the other hand, we can use that

$$\frac{s_{n-1}}{s_n} + h = \frac{s_{n-1}}{s_n} + \frac{1}{2\alpha_{n-1} s_n} = 1 - \frac{1}{2\alpha_{n-1} s_n},$$

and that α_{n-1} is bounded by $\rho/(1-\rho)$, to obtain

$$\sum_{j=0}^{n-1} \frac{1}{\alpha_j s_n}\left(1-\frac{s_j}{s_n}\right)^{-1} \leq 2\log 2 + 2\log(2\alpha_{n-1}s_n) = O\big(\log(1+s_n)\big),$$

the corresponding constant depending only on ρ.

Combining this with (4.15), and inserting both into (4.14), we finally arrive at our claim (4.12), respectively (4.13). $\square$

We note that Proposition 3.2 guarantees that the required bound in Lemma 4.3 for the regularization parameters is always fulfilled for the regularizing Levenberg-Marquardt scheme, provided that the normalization (2.2) holds true.

5. The induction argument. The equation (4.5) for the error e_n differs from the error of nonstationary Tikhonov regularization for *linear problems* by the accumulated error component due to the perturbation terms z_j that enter in each step of the iteration. In [**9, 11**], similar to the approach in [**8**] for the nonlinear Landweber iteration, a major effort of the analysis was to show that the influence of these perturbations is largely negligible, and this required a nontrivial induction argument.

In the sequel we proceed in much the same way; to avoid the constraints on the regularization parameters that have been used in [**9, 10**], cf. (4.11), we will utilize Lemma 3.4 appropriately.

Lemma 5.1. *Let F satisfy* (2.2) *and* (2.3), (2.4), *and let $\tau > 2$ and $0 < \rho < 1$ be chosen subject to the constraint $\rho\tau > 1$. Furthermore assume that $x_0 - x^\dagger$ satisfies the source condition* (2.14) *for some $0 < \nu \leq 1/2$ and $w \in \mathcal{X}$ with $\|w\| = \omega$ being sufficiently small. Then there is a constant C_* such that for every $j = 1, \dots, n(y^\delta) - 1$, the iteration error e_j of the regularizing Levenberg-Marquardt scheme satisfies*

$$e_j = (T^*T)^\nu w_j \quad \textit{for some } w_j \in \mathcal{X} \textit{ with } \|w_j\| \leq C_*\omega\,. \tag{5.1}$$

Moreover, the following holds

$$\|e_j\| \le C_* \omega s_j^{-\nu} \quad and \quad \|Te_j\| \le C_* \omega s_j^{-\nu-1/2} \tag{5.2}$$

for every $j = 1, 2, \ldots, n(y^\delta) - 1$.

Proof. It is easy to see that the three terms on the right-hand side of (4.5) belong to $\mathcal{R}((T^*T)^\nu)$ and $\mathcal{R}(T^*) = \mathcal{R}((T^*T)^{1/2})$, respectively, the latter being in turn a subspace of $\mathcal{R}((T^*T)^\nu)$. Accordingly, the iteration error can be written in the form (5.1) for some $w_j \in \mathcal{X}$. More precisely, cf., e.g., [**2**, Proposition 2.18], we have

$$\begin{gathered} T^*(y^\delta - y) = (T^*T)^{1/2} d, \\ \text{with } d \in \mathcal{X}, \quad \|d\| \le \|y^\delta - y\|, \end{gathered} \tag{5.3}$$

and

$$\begin{gathered} T^* z_j = (T^*T)^{1/2} \widetilde{z}_j, \\ \text{with } \widetilde{z}_j \in \mathcal{X}, \quad \|\widetilde{z}_j\| \le \|z_j\|, \end{gathered} \tag{5.4}$$

and we can thus choose w_n, $1 \le n < n(y^\delta)$, according to (4.5) to be

$$\begin{aligned} w_n = {} & \prod_{j=0}^{n-1} r_j(T^*T) w + q_n(T^*T)(T^*T)^{1/2-\nu} d \\ & + \sum_{j=0}^{n-1} (T^*T)^{1/2-\nu} g_j(T^*T) \prod_{k=j+1}^{n-1} r_k(T^*T) \widetilde{z}_j. \end{aligned} \tag{5.5}$$

Similarly, we can rewrite $Te_0 = (TT^*)^{\nu+1/2} \widetilde{w}$ for some $\widetilde{w} \in \mathcal{Y}$ with $\|\widetilde{w}\| \le \|w\| = \omega$, such that

$$\begin{aligned} Te_n = {} & \prod_{j=0}^{n-1} r_j(TT^*)(TT^*)^{\nu+1/2} \widetilde{w} + q_n(TT^*) TT^*(y^\delta - y) \\ & + \sum_{j=0}^{n-1} TT^* g_j(TT^*) \prod_{k=j+1}^{n-1} r_k(TT^*) z_j \end{aligned} \tag{5.6}$$

for $n = 1, \ldots, n(y^\delta) - 1$.

As mentioned before we will prove the lemma by induction, and we start with the inductive step, i.e., we assume that for some $1 < n < n(y^\delta)$ all inequalities in (5.1) and (5.2) hold true for all indices $1 \leq j < n$, and we are going to establish now these inequalities for $j = n$. For this we assume that ω is so small that η of (2.13) satisfies

$$\eta \leq \frac{1}{4+8\tau} < 1/2.$$

Using (5.1), respectively (2.14), setting $w_0 = w$, and the interpolation inequality, cf., e.g., [**2**, (2.49)], we obtain for $0 \leq j \leq n-2$ that

$$\begin{aligned} \|e_j\| &\leq \|w_j\|^{1/(2\nu+1)} \|Te_j\|^{2\nu/(2\nu+1)} \\ &\leq (C_*\omega)^{1/(2\nu+1)} \|Te_j\|^{2\nu/(2\nu+1)}, \end{aligned}$$

and hence, using Lemma 4.1 and Lemma 3.4,

$$\begin{aligned} \|z_j\| &\leq 12 C_R (C_*\omega)^{1/(2\nu+1)} \|Te_j\|^{1+2\nu/(2\nu+1)} \\ &\leq 12(42\tau)^2 C_R (C_*\omega)^{1/(2\nu+1)} \|Te_{j+1}\|^{1+2\nu/(2\nu+1)}. \end{aligned}$$

Inserting the induction hypothesis (5.2) we thus arrive at

$$\|z_j\| \leq C_1 C_*^2 \omega^2 s_{j+1}^{-2\nu-1/2}, \quad 0 \leq j \leq n-2 \tag{5.7}$$

where $C_1 > 0$ only depends on τ and C_R. z_{n-1}, on the other hand, can be estimated by means of Lemma 4.1 and Lemma 3.4 as

$$\|z_{n-1}\| \leq 12 C_R \|e_0\| \|Te_{n-1}\| \leq \varepsilon \|Te_n\| \tag{5.8}$$

with

$$\varepsilon = \frac{\tau - 2}{2\tau} \tag{5.9}$$

by choosing ω, and thus η, sufficiently small. Note that ε defined by (5.9) is positive; it is here, where the assumption (2.8) is crucial.

Now consider the sum in the second row of (5.6). Applying Lemma 4.2, and using the bounds (5.7), respectively (5.8), for $\|z_j\|$, we obtain

$$\begin{aligned}\Big\| \sum_{j=0}^{n-1} TT^* g_j(TT^*) \prod_{k=j+1}^{n-1} r_k(TT^*) z_j \Big\| & \\ &\le \|TT^* g_{n-1}(TT^*) z_{n-1}\| \\ &\quad + \sum_{j=0}^{n-2} \Big\| TT^* g_j(TT^*) \prod_{k=j+1}^{n-1} r_k(TT^*) z_j \Big\| \\ &\le \|z_{n-1}\| + \sum_{j=0}^{n-2} \frac{1}{\alpha_j} (s_n - s_j)^{-1} \|z_j\| \\ &\le \varepsilon \|Te_n\| + C_1 C_*^2 \omega^2 \sum_{j=0}^{n-2} \frac{1}{\alpha_j} (s_n - s_j)^{-1} s_{j+1}^{-2\nu-1/2},\end{aligned}$$

and hence, by virtue of Lemma 4.3,

$$\begin{aligned}\Big\| \sum_{j=0}^{n-1} TT^* g_j(TT^*) \prod_{k=j+1}^{n-1} r_k(TT^*) z_j \Big\| & \\ &\le \varepsilon \|Te_n\| + C_2 C_*^2 \omega^2 s_n^{-\nu-1/2}\end{aligned} \tag{5.10}$$

for some constant C_2 depending only on ρ, τ, ν, and C_R. Inserting this into (5.6), and making another use of Lemma 4.2, we arrive at

$$\begin{aligned}(1-\varepsilon)\|Te_n\| &\le \omega s_n^{-\nu-1/2} + \delta + C_2 C_*^2 \omega^2 s_n^{-\nu-1/2} \\ &\le (1 + C_2 C_*^2 \omega)\, \omega s_n^{-\nu-1/2} + \delta.\end{aligned} \tag{5.11}$$

Next we estimate δ: Because of (2.7) and Proposition 3.3 we can arrange that

$$\delta < \frac{1}{\tau} \|y^\delta - F(x_n^\delta)\| \le \left(1 + \frac{3}{2}\eta\right) \frac{1}{\tau - 1} \|Te_n\| \le \frac{1+\varepsilon}{\tau-1} \|Te_n\| \tag{5.12}$$

by forcing η to be sufficiently small. Inserting this and (5.9) into (5.11), we obtain

$$\frac{\tau-2}{2(\tau-1)} \|Te_n\| = \left(1 - \varepsilon - \frac{1+\varepsilon}{\tau-1}\right) \|Te_n\| \le (1 + C_2 C_*^2 \omega)\, \omega s_n^{-\nu-1/2},$$

and the required estimate (5.2) for $\|Te_n\|$ now follows for any $C_* > 2(\tau-1)/(\tau-2)$ by choosing ω sufficiently small.

Having established (5.2) for $\|Te_n\|$ we can now extend the validity of the estimate (5.7) to $j = n-1$, and obtain, by using Lemma 4.2, (5.4), and Lemma 4.3 that

$$\begin{aligned}\Big\| \sum_{j=0}^{n-1} (T^*T)^{1/2-\nu} g_j(T^*T) \prod_{k=j+1}^{n-1} r_k(T^*T)\widetilde{z}_j \Big\| &\\ &\le \sum_{j=0}^{n-1} \frac{1}{\alpha_j}(s_n - s_j)^{\nu-1/2}\|z_j\| \\ &\le C_1 C_*^2 \omega^2 \sum_{j=0}^{n-1} \frac{1}{\alpha_j}(s_n - s_j)^{\nu-1/2} s_{j+1}^{-2\nu-1/2} \\ &\le C_2 C_*^2 \omega^2,\end{aligned}$$

where C_2 can be tuned to be the same as above. Inserting this and (5.3) into (5.5), and using Lemma 4.2 again, we conclude that

$$\|w_n\| \le \omega + s_n^{\nu+1/2}\delta + C_2 C_*^2 \omega^2.$$

Estimating δ by (5.12) as before, and subsequently $\|Te_n\|$ by (5.2), we obtain

$$\|w_n\| \le \omega + \frac{1+\varepsilon}{\tau - 1}\, C_* \omega + C_2 C_*^2 \omega^2,$$

and, again, the right-hand side is going to be less than $C_*\omega$ for any $C_* > 2(\tau-1)/(\tau-2)$ because of (5.9), provided that ω is sufficiently small. This proves (5.1), and the estimate (5.2) for e_n now follows from the interpolation inequality.

To complete the proof it remains to establish the base case, i.e., inequalities (5.1) and (5.2) for $j = 1$, given that $1 < n(y^\delta)$. We start with the estimate of Te_1. According to (5.6) we have

$$Te_1 = r_0(TT^*)(TT^*)^{\nu+1/2}\widetilde{w} + q_1(TT^*)TT^*(y^\delta - y + z_0),$$

as q_1 and g_0 coincide. It thus follows from Lemma 4.2 that

$$\|Te_1\| \le \omega s_1^{-\nu-1/2} + \delta + \|z_0\|.$$

We can estimate δ as in (5.12), and z_0 as in (5.8) with ε of (5.9). This then yields

$$\frac{\tau-2}{2(\tau-1)}\,\|Te_1\| \leq \omega s_1^{-\nu-1/2},$$

which verifies the second claim in (5.2) for $j=1$.

Next we turn to (5.1). According to (5.5) w_1 is given by

$$w_1 = r_0(T^*T)w + q_1(T^*T)(T^*T)^{1/2-\nu}(d+\widetilde{z}_0),$$

and Lemma 4.2 therefore yields the upper bound

$$\|w_1\| \leq \omega + s_1^{\nu+1/2}(\delta + \|z_0\|).$$

Having already verified the inequality (5.2) for Te_1, we can now argue as in (5.7) to estimate z_0, and to obtain

$$\|w_1\| \leq \omega + s_1^{\nu+1/2}\delta + C_2C_*^2\omega^2.$$

The desired inequality (5.1) now follows with the same argument as in the inductive step, and likewise we obtain the remaining inequality (5.2) for e_1. □

Remark 5.2. It can be seen from the proof that the formulation of Lemma 5.1 can be made more precise in the following way: For any $C_* > 2(\tau-1)/(\tau-2)$ there exists an $\omega_0 > 0$, depending on C_*, τ, ρ, ν and C_R, such that (5.1) and (5.2) hold true whenever $\omega \leq \omega_0$.

As promised in Section 2, we also provide a lower bound for the regularization parameter α_n determined via (2.9) for the sake of completeness.

Proposition 5.3. *Under the assumptions of Lemma* 5.1 *the regularization parameters α_n, $n=0,\dots,n(y^\delta)-1$, of the regularizing Levenberg-Marquardt scheme* (2.6), (2.9) *satisfy*

$$\alpha_n \geq c_*\left(\|Te_n\|/\omega\right)^{2/(2\nu+1)}$$

for some constant $c_* > 0$, depending only on ρ, τ, and ν.

Proof. According to [**6**] the error e_n satisfies the linear operator equation

$$T_n e = F(x_n^\delta) - y^\delta \tag{5.13}$$

up to an error

$$\|F(x_n^\delta) - y^\delta - T_n e_n\| \le \frac{\rho}{\gamma}\,\|y^\delta - F(x_n^\delta)\|, \tag{5.14}$$

where $\gamma > 1$ is some fixed number, depending only on $\rho\tau$. Moreover, by virtue of (5.1) and Proposition 3.1, e_n satisfies the associated source condition

$$e_n = (T_n^* T_n)^\nu \widehat{w}_n \quad \text{with} \quad \|\widehat{w}_n\| \le c_0^{2\nu}\|w_n\| \le c_0^{2\nu} C_* \omega, \tag{5.15}$$

where $c_0 \le 1/(1-\eta) \le 2$ provided that ω is sufficiently small. The parameter α_n can now be interpreted as the particular regularization parameter for problem (5.13) which corresponds to the discrepancy principle with error bound (5.14) and fudge factor γ, compare (2.9). It thus follows from display (4.71) in [**2**] that

$$\alpha_n \ge c\biggl(\frac{\rho}{\gamma}\,\|y^\delta - F(x_n^\delta)\|/\|\widehat{\omega}_n\|\biggr)^{2/(2\nu+1)}$$

for some constant c depending only on γ and on ν. Inserting (5.15) and applying Proposition 3.3, the assertion now follows readily. □

Note that this result, when combined with estimate (5.2) suggests that

$$\alpha_n \gtrsim s_n^{-1},$$

provided that (5.2) is sharp, and this would then correspond to the assumption (4.11) that has been used in [**9**]. As said before, however, we have not been able to establish such a bound rigorously.

6. Proof of Theorem 2.1. Let us assume first that $n(y^\delta) = 0$. Then we conclude from (2.11) and (2.7) that

$$\begin{aligned}
\|Te_0\| &\le \|F(x_0) - F(x^\dagger) - Te_0\| + \|F(x_0) - F(x^\dagger)\| \\
&\le (2\eta+1)\|F(x_0) - F(x^\dagger)\| \\
&\le (2\eta+1)\bigl(\|F(x_0) - y^\delta\| + \|y^\delta - y\|\bigr) \\
&\le (2\eta+1)(\tau+1)\,\delta.
\end{aligned}$$

The interpolation inequality and (2.14) then provide the desired inequality for $e_{n(y^\delta)} = x_0 - x^\dagger$.

In the case when $n = n(y^\delta) > 0$ we assume that ω is small enough so that we can proceed as in the proof of Lemma 3.4, cf. (3.3), to deduce that

$$\|y^\delta - F(x_{n-1}^\delta)\| \leq \frac{2}{\rho}\,\|y^\delta - F(x_n^\delta)\|,$$

and hence,

$$\|y^\delta - F(x_{n-1}^\delta)\| \leq \frac{2}{\rho}\,\tau\delta < 2\tau^2\delta$$

by virtue of (2.12) and the definition (2.7) of the stopping index $n = n(y^\delta)$. It thus follows from Proposition 3.3 that

$$\|Te_{n-1}\| \leq 12\tau^2\delta,$$

and therefore the interpolation inequality and (5.1) yield

$$\|e_{n-1}\| \leq \|w_{n-1}\|^{1/(2\nu+1)}\|Te_{n-1}\|^{2\nu/(2\nu+1)} \leq C\omega^{1/(2\nu+1)}\delta^{2\nu/(2\nu+1)}$$

for some constant $C > 0$ depending only on τ, provided that ω has been chosen sufficiently small, compare Remark 5.2. The assertion of Theorem 2.1 thus follows from the monotonicity of the iteration error up to the stopping index, cf. display (2.9) in [**6**]. $\square$

ENDNOTES

1. The constants in (2.10), (2.11) are somewhat larger than the optimal ones because we have confined ourselves to use (2.4) only for one particular element $\widetilde{x} = x^\dagger \in \mathcal{X}$. However, if (2.4) only holds for that particular element, then x and $\widetilde{x}$ in (2.10), (2.11), both have to be close to $x^\dagger$.

REFERENCES

1. A.B. Bakushinskii, *The problem of the convergence of the iteratively regularized Gauss-Newton method*, Comput. Maths. Math. Phys. **32** (1992), 1353–1359.

2. H.W. Engl, M. Hanke and A. Neubauer, *Regularization of inverse problems*, Kluwer, Dordrecht, 1996.

3. C.W. Groetsch, *Generalized inverses of linear operators: representation and approximation*, Dekker, New York, 1977.

4. ———, *The theory of Tikhonov regularization for Fredholm integral equations of the first kind*, Pitman, Boston, 1984.

5. ———, *Stable approximate evaluation of unbounded operators*, Springer, Berlin, 2007.

6. M. Hanke, *A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems **13** (1997), 79–95.

7. M. Hanke and C.W. Groetsch, *Nonstationary iterated Tikhonov regularization*, J. Optim. Theor. Appl. **98** (1998), 37–53.

8. M. Hanke, A. Neubauer and O. Scherzer, *A convergence analysis of the Landweber iteration for nonlinear ill-posed problems*, Numer. Math. **72** (1995), 21–37.

9. M. Hochbruck and M. Hönig, *On the convergence of a regularizing Levenberg-Marquardt scheme for nonlinear ill-posed problems*, Numer. Math. **115** (2010), 71–79.

10. M. Hochbruck, M. Hönig and A. Ostermann, *A convergence analysis of the exponential Euler iteration for nonlinear ill-posed problems*, Inverse Problems **25** (2009), 075009 (18pp).

11. Q. Jin, *On a regularized Levenberg-Marquardt method for solving nonlinear inverse problems*, Numer. Math. **115** (2010), 229–259.

12. Q. Jin and U. Tautenhahn, *On the discrepancy principles for some Newton type methods for solving nonlinear inverse problems*, Numer. Math. **111** (2009), 509–558.

13. B. Kaltenbacher, A. Neubauer and O. Scherzer, *Iterative regularization methods for nonlinear ill-posed problems*, de Gruyter, Berlin, 2008.

14. C.T. Kelley, *Iterative methods for optimization*, SIAM, Philadelphia, 1999.

15. A. Rieder, *On the regularization of nonlinear ill-posed problems via inexact Newton iterations*, Inverse Problems **15** (1999), 309–327.

16. ———, *On convergence rates of inexact Newton regularizations*, Numer. Math. **88** (2001), 347–365.

Institut für Mathematik, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany
Email address: hanke@math.uni-mainz.de