

# REGULARIZING NEWTON-KACZMARZ METHODS FOR NONLINEAR ILL-POSED PROBLEMS

MARTIN BURGER \* AND BARBARA KALTENBACHER †

**Abstract.** We introduce a class of stabilizing Newton-Kaczmarz methods for nonlinear ill-posed problems and analyze their convergence and regularization behaviour. As usual for iterative methods for solving nonlinear ill-posed problems, conditions on the nonlinearity (or the derivatives) have to be imposed in order to obtain convergence. As we shall discuss in general and in some specific examples, the nonlinearity conditions obtained for the Newton-Kaczmarz methods are less restrictive than those for previously existing iteration methods and can be verified for several practical applications.

We also discuss the discretization and efficient numerical solution of the linear problems arising in each step of a Newton-Kaczmarz method, and carry out numerical experiments for two model problems.

**Key words.** Newton-Kaczmarz Methods, Ill-Posed Problems, Tomography.

**AMS subject classifications.** 65J20, 65J15, 65N21, 47J06.

**1. Introduction.** The aim of this paper is to develop and analyze *Newton-Kaczmarz methods* for nonlinear inverse problems, focusing in particular on the important class of identification problems with multiple boundary data. The main idea of the Kaczmarz method is to split the inverse problem into a finite number of subproblems and to approximate its solution by performing a cyclic iteration over the subproblems.

As a *regularized Newton-Kaczmarz method* we understand the cyclic iteration where at each step one iteration of a regularized Newton method is applied to a subproblem. As we shall discuss in detail in this paper, the benefit from this approach is twofold:

1. Instead of solving one large problem in each iteration step, we can solve several smaller subproblems, which might lead to a reduction of the overall computational effort.
2. Due to the ill-posedness of the problem, conditions on the nonlinearity of the problem have to be imposed in order to ensure convergence of iterative methods (cf. [6] for an overview). These conditions are rather restrictive and cannot be verified for many practical problems, in particular for parameter identification problems using boundary data related to the solutions of partial differential equations. As we shall show below for several applications, the nonlinearity conditions for the Newton-Kaczmarz method are less restrictive and can be verified in more realistic cases.

The price which one has to pay is that at least theoretically it turns out that more a priori information has to be contained in the initial values.

Another motivation for the analysis in this paper is that Kaczmarz-type methods (also called *algebraic reconstruction technique*) have been used already in several

---

\*Department of Mathematics, UCLA, 520 Portola Plaza, Los Angeles, CA 90095, USA. e-mail: martin.burger@jku.at. On leave from: Industrial Mathematics Institute, Johannes Kepler Universität Linz, Austria. Supported by the Austrian National Science Foundation FWF through project SFB F 013 / 08

†junior research group “Inverse Problem in Piezoelectricity”, Department of Sensor Technology, Universität Erlangen-Nürnberg, Paul-Gordan-Strasse 3/5, D 91053 Erlangen, Germany. e-mail: barbara.kaltenbacher@lse.eei.uni-erlangen.de, supported by the German Science Foundation DFG under grant Ka 1778/1-1

applications with multiple boundary data (cf. [3, 8, 9, 24, 30]) and performed better than standard iterative methods. This paper, together with the results of Kowar and Scherzer [24] on the Landweber-Kaczmarz method, might serve to provide a theoretical basis.

Many inverse problems can be formulated as nonlinear operator equations

$$F(x) = y, \quad (1.1)$$

or as collections of  $p$  coupled operator equations

$$F_i(x) = y_i, \quad i = 0, \dots, p-1 \quad (1.2)$$

with nonlinear operators  $F_i$  mapping between Hilbert spaces  $X$  and  $Y_i$ . We will here assume that a solution  $x^\dagger$  of (1.2) exists, but need not necessarily be unique.

Note that (1.1) can be seen as a special case of (1.2) with  $p = 1$ ; on the other hand defining

$$F := (F_0, \dots, F_{p-1}), \quad y := (y_0, \dots, y_{p-1}), \quad (1.3)$$

one can reduce (1.2) to (1.1). However, one potential advantage of (1.2) over (1.1) can be, that it might better reflect the structure of the underlying information  $(y_0, \dots, y_{p-1})$  leading to the coupled system, than a plain concatenation into one single data element  $y$  could. The most important feature that we have in mind, though, is that it enables the definition of Newton type solution methods and to proof their convergence for certain relevant problems, for which Newton type methods applied to the single equation formulation (1.1) cannot be shown to converge.

In general we assume that we only have noisy data  $y_i^\delta$  with some noise level  $\delta$  bounding the noise of every measurement by

$$\|y_i^\delta - y_i\| \leq \delta, \quad (1.4)$$

Note that for  $p > 1$ , this assumption on the noise is more restrictive than the frequently used noise bound

$$\|y^\delta - y\| \leq \delta,$$

but it reflects the case of multiple measurements, where an individual noise bound is available for each. If the noise level for each measurement is different, we can make it equal by using a relative scaling between the operators  $F_i$ .

Since we are interested in the situation that (1.2) is ill-posed in the sense that small perturbations in the data can lead to large deviations in the solution, and since in practice only noisy data are available, we have to apply suitable regularization techniques (see, e.g., [11, 13, 23, 25, 28, 29, 33]). Typically, the instability in nonlinear inverse problems (1.1) corresponds to a smoothing property of the forward operator  $F$  and its linearization  $F'(x)$ . In particular, for an ill-posed problem, we cannot expect that  $F'(x)$  is continuously invertible, and consequently a standard Newton or Gauss-Newton cannot be used. Modified Newton-type method for solving (1.1) have been studied and analyzed in several recent publications, see, e.g. [1, 6, 14, 15, 22, 31]. Regularization is here achieved by replacing the in general unbounded inverse of  $F'(x)$  in the definition of the Newton step by a bounded approximation, defined via a regularizing operator

$$G_\alpha(F'(x)) \approx F'(x)^\dagger.$$

Here,  $K^\dagger$  denotes the pseudo-inverse of a linear operator  $K$ ,  $\alpha > 0$  is a small regularization parameter, and  $G_\alpha$  satisfies

$$G_\alpha(K)y \rightarrow K^\dagger y \quad \text{as } \alpha \rightarrow 0 \quad \forall y \in \mathcal{R}(K), \quad (1.5)$$

and

$$\|G_\alpha(K)\| \leq \Phi(\alpha), \quad (1.6)$$

for any linear operator  $K$  within some uniformly bounded set. Note that, especially in view of operators  $K$  with unbounded inverses, the constant  $\Phi(\alpha)$  has to tend to infinity as  $\alpha$  goes to zero; we assume w.l.o.g. that  $\Phi(\alpha)$  is strictly monotonically decreasing.

Choosing a sequence  $(\alpha_n)$  of regularization parameters and applying the bounded operators  $G_{\alpha_n}(F'(x_n))$  in place of  $F'(x_n)^{-1}$  in Newton's method results in the iteration

$$x_{n+1} = x_n - G_{\alpha_n}(F'(x_n))(F(x_n) - y^\delta). \quad (1.7)$$

If  $G_\alpha$  is defined by Tikhonov regularization

$$G_\alpha(K) = (K^*K + \alpha I)^{-1}K^*, \quad (1.8)$$

one arrives at the Levenberg-Marquardt method (see [15]; for  $G_\alpha$  given by a conjugate gradient iteration, see [14], further work on this class of methods can be found in [32]).

A different class of regularized Newton methods emerged from the iteratively regularized Gauss-Newton method (IRGNM),

$$x_{n+1} = x_0 - G_{\alpha_n}(F'(x_n))(F(x_n) - y^\delta - F'(x_n)(x_n - x_0)),$$

with (1.8), which was first proposed and analyzed by Bakushinskii in [1] and later extended to regularization with general regularization operators  $G_{\alpha_n}$ , cf. [2], see also [19], [22]. Here,  $\alpha_n \xrightarrow{n \rightarrow \infty} 0$  is an a priori chosen monotonically decreasing sequence of regularization parameters. One observes that in the limiting case  $\alpha_n \rightarrow 0$  (i.e.,  $G_{\alpha_n}(F'(x_n)) \rightarrow F'(x)^\dagger$ ) also this formulation is equivalent to the usual Newton method.

In order to make these Newton-type methods applicable to multiple equations (1.2), we combine them with a Kaczmarz approach (similar to [24]). Starting from an initial guess  $x_{0,i}$ , we perform a Newton step for the equation  $F_i(x) = y_i$ , for  $i$  from 0 to  $p-1$ , and repeat this procedure in a cyclic manner. Incorporating the possibility of different regularization methods  $G^i$  for each equation in (1.2), and using the “overloading” notation

$$x_{0,n} := x_{0, \text{mod}(n,p)}, \quad F_n := F_{\text{mod}(n,p)}, \quad y_n := y_{\text{mod}(n,p)}, \quad G_\alpha^n := G_\alpha^{\text{mod}(n,p)} \quad (1.9)$$

this can be written as

$$x_{n+1} = x_{0,n} - G_{\alpha_n}^n(F'_n(x_n))(F_n(x_n) - y_n^\delta - F'_n(x_n)(x_n - x_{0,n})). \quad (1.10)$$

A combination of the Levenberg-Marquardt method with a Kaczmarz approach will be shortly discussed in Section 3 below.

Our convergence analysis will be a local one, i.e., we will work in a neighborhood  $\mathcal{B}_\rho(x^\dagger)$  of the solution, which we assume to be a subset of the domains of the operators  $F_i$

$$\mathcal{B}_\rho(x^\dagger) \subseteq \mathcal{D}(F_i), \quad i = 0, \dots, p-1.$$

The remainder of the paper is organized as follows: In Section 2 we discuss conditions on the nonlinearity of the problem and so-called *source conditions*, which are abstract smoothness assumptions on the solution. Section 3 contains a convergence analysis of (1.10) including the case of noisy data and convergence rates under additional regularity assumptions. In Section 4, we derive some approaches for the efficient implementation of the proposed methods, and Section 5 provides numerical results.

**2. Nonlinearity and Source Conditions.** In the following we shall discuss the basic conditions needed for the subsequent analysis in this paper. In particular we shall introduce conditions on the nonlinearity of the involved operators  $F_i$  and investigate their applicability to tomography-type problems.

**2.1. Nonlinearity Conditions.** To make these methods well-defined, we assume the forward operators  $F_i$  to be Fréchet differentiable with derivatives being uniformly bounded in a neighborhood of the solution. This uniform bound has to be such that applicability of the respective regularization method can be guaranteed

$$\|F'_i(x)\| \leq C_S^i \quad \forall x \in \mathcal{B}_\rho(x^\dagger) \quad (2.1)$$

which can always be achieved by a proper scaling. In order to prove convergence of regularization methods for nonlinear ill-posed problems, one usually needs assumptions not only on the smoothness of the forward operator  $F$  but also on the type of nonlinearity it contains, though. Here we shall mainly consider the condition

$$F'_i(\bar{x}) = F'_i(x)R_i(\bar{x}, x) \quad \forall \bar{x}, x \in \mathcal{B}_\rho(x^\dagger) \quad (2.2)$$

which means that the range of the Fréchet derivative of each forward operator  $F_i$  is locally invariant around the solution. The linear operators  $R_i(\bar{x}, x)$  (that by the way need not be known explicitly) should satisfy a Lipschitz type estimate

$$\|R_i(\bar{x}, x) - I\| = \|R_i(\bar{x}, x) - R_i(x, x)\| \leq C_R \|\bar{x} - x\|. \quad (2.3)$$

This corresponds to an analogous assumption in the context of  $p = 1$ , i.e., (1.1),

$$F'(\bar{x}) = F'(x)R(\bar{x}, x), \quad \forall \bar{x}, x \in \mathcal{B}_\rho(x^\dagger) \quad (2.4)$$

as it was used, e.g., in the convergence analysis of [22], and is closely related to the so-called affine covariant Lipschitz condition in [7]. Condition (2.2) seems to be natural especially in the context of parameter identification in PDEs from boundary measurements where the forward operator consists of a (typically invertible) solution operator for the PDE, composed with a linear operator mapping the PDE solution to the measured boundary values. In fact, by the additional freedom arising from the possibility of having different operators  $R_i$  for each  $i$ , it can be verified for important applications of parameter identification, like ultrasound tomography (see below) and impedance tomography, for which other nonlinearity conditions used in literature cannot be proven to hold.

An alternative nonlinearity condition that can be found in the literature on regularization methods for nonlinear inverse problems (1.1) is

$$F'(\bar{x}) = R(\bar{x}, x)F'(x) \quad \forall \bar{x}, x \in \mathcal{B}_\rho(x^\dagger) \quad (2.5)$$

with regular operators  $R(\bar{x}, x)$ , i.e., range invariance of the adjoints of  $F'(x)$ , which is closely related to the tangential cone condition used e.g. in [14, 15, 17, 21, 22], and to the Newton-Mysovskii conditions discussed in [6].

We want to mention that the nonlinearity condition (2.2) is less restrictive than the corresponding nonlinearity condition (2.4) for the operator  $F$  defined by (1.3). If (2.4) holds, we can easily deduce (2.2) by choosing  $R_i = R$  for all  $i$ . For the alternative condition

$$F'_i(\bar{x}) = R_i(\bar{x}, x)F'_i(x) \quad \forall \bar{x}, x \in \mathcal{B}_\rho(x^\dagger) \quad (2.6)$$

and the corresponding condition (2.5) for  $F$  defined by (1.3), we obtain sufficiency in the other direction, since we can choose  $R$  to be the diagonal operator consisting of all  $R_i$  to obtain the range invariance of  $F'^*$  from (2.6).

Finally, we examine a special case of a decomposition of  $F_i$  in a linear singular and a nonlinear regular operator. As we shall see below in several examples, the operators  $F_i$  can often be written as the composition of linear trace-type operators with nonlinear parameter-to-solution maps for partial differential equations. Thus we start with a simple observation that allows to verify the nonlinearity condition for the parameter-to-solution map only. In this context we wish to refer to Section 5 in [18] where a class of operators satisfying the nonlinearity condition (2.5) is derived.

**LEMMA 2.1.** *Let  $X, Y, Z$  be Hilbert spaces, and let  $L_i \in \mathcal{L}(Z, Y)$ . Moreover, let  $H_i : X \rightarrow Z$ ,  $i = 0, \dots, p-1$  be continuously Fréchet differentiable operators. Then,*

$$F_i = L_i \circ H_i \quad (2.7)$$

*satisfies (2.2), (2.3) if  $H_i$  satisfies (2.2), (2.3).*

*Moreover, if  $H'_i(x)$  is regular for all  $x \in \mathcal{B}_\rho(x^\dagger)$  with uniformly bounded inverse, and the map  $x \mapsto H'_i(x)$  is Lipschitz continuous, then the condition (2.2), (2.3) is satisfied by  $H_i$ .*

*Proof.* The first assertion follows from

$$F'_i(\bar{x}) = L_i \circ H'_i(\bar{x}) = L_i \circ H'_i(x) \circ R_i(\bar{x}, x) = F'_i(x) \circ R_i(\bar{x}, x).$$

Moreover, if  $H'_i$  is regular, we may define

$$R_i(\bar{x}, x) := H'_i(x)^{-1}H'_i(\bar{x})$$

which implies (2.2). Due to the regularity of  $H'_i(x)^{-1}$  and the Lipschitz-continuity of  $x \mapsto H'_i(x)$ , we obtain

$$\|R_i(\bar{x}, x) - I\| = \|H'_i(x)^{-1}(H'_i(\bar{x}) - H'_i(x))\| \leq C_0 \|H'_i(\bar{x}) - H'_i(x)\| \leq C_R \|\bar{x} - x\|,$$

i.e., (2.3) holds.  $\square$

**2.2. Examples.** In the following we discuss several examples of inverse problems satisfying the nonlinearity condition including tomography-type problems for

partial differential equations in the above framework, and show that they satisfy the nonlinearity condition (2.2).

EXAMPLE 1 (Reconstruction from Dirichlet-Neumann Map). We start with a rather simple model problem, namely the estimation of the coefficient  $q \geq 0$  in the partial differential equation

$$-\Delta u + qu = 0, \quad \text{in } \Omega \subset \mathbb{R}^d$$

from measurements of the Neumann value  $g = \frac{\partial u}{\partial \nu}$  on  $\partial\Omega$  for several different Dirichlet values  $u = f$  on  $\partial\Omega$ .

If we denote the different Dirichlet values by  $f_i$ ,  $i = 0, \dots, p-1$ , and the corresponding measurements by  $g_i$ , we may rewrite the problem as

$$F_i(q) = g_i, \quad i = 0, \dots, p-1,$$

where  $F_i : \mathcal{D}(F_i) \subseteq L^2(\Omega) \rightarrow H^{-\frac{1}{2}}(\partial\Omega)$  is the nonlinear operator mapping  $q$  to  $\frac{\partial u_i}{\partial \nu}$ , where  $u_i \in H^1(\Omega)$  is the weak solution of

$$\begin{aligned} -\Delta u_i + qu_i &= 0, & \text{in } \Omega, \\ u_i &= f_i & \text{on } \partial\Omega, \end{aligned}$$

and  $\mathcal{D}(F_i)$  is to be specified below.

The decomposition (2.7) is obtained with  $L : H^1(\Omega) \mapsto H^{-\frac{1}{2}}(\partial\Omega)$  being the trace operator that maps a function to its normal derivative on the boundary, and  $H_i : q \mapsto u_i$  is the parameter-to-solution map.

The derivative  $v_i = H'_i(q)s$  is given as the unique weak solution of

$$\begin{aligned} -\Delta v_i + qv_i + su_i &= 0, & \text{in } \Omega, \\ v_i &= 0 & \text{on } \partial\Omega. \end{aligned}$$

Formally, we can write  $H'_i(q) = -(-\Delta + q)^{-1}(u_i \cdot)$ . It can be shown easily, that this operator is regular between  $L^2(\Omega)$  and  $H^1(\Omega)$ , if  $u_i > 0$ . Due to a standard maximum principle for second order elliptic differential equations, this is the case if  $q \geq 0$  and  $f_i > 0$ . Moreover, since embedding operators are continuous and regular, the operator  $H'_i(q)$  is also regular between a Sobolev space  $H^\beta(\Omega)$ ,  $\beta \geq 0$ , and  $H^1(\Omega)$ . Thus, if  $\beta > \frac{d}{2}$  (i.e.,  $H^\beta(\Omega) \hookrightarrow C(\bar{\Omega})$ ) and there exists a minimum norm solution  $q^\dagger \in H^\beta(\Omega)$ , which is positive in  $\bar{\Omega}$ , then  $q \in \mathcal{B}_\rho(q^\dagger)$  is nonnegative for  $\rho$  sufficiently small and due to the above reasoning Lemma 2.1 implies that the nonlinearity condition (2.2), (2.3) is satisfied for  $f_i > 0$ , if we consider  $F_i$  as an operator from  $\mathcal{D}(F_i) := \mathcal{B}_\rho(q^\dagger) \subseteq H^\beta(\Omega) =: X$  to  $H^{-\frac{1}{2}}(\partial\Omega) =: Y_i$ .

EXAMPLE 2 (Reconstruction from Multiple Sources). In some examples, one rather tries to estimate coefficients in partial differential equations from boundary measurements for different interior sources rather than from different boundary values. We consider the estimation of the coefficient  $q \geq 0$  in

$$-\Delta u + qu = h, \quad \text{in } \Omega \subset \mathbb{R}^d$$

subject to a homogeneous Neumann boundary condition  $\frac{\partial u}{\partial \nu} = 0$  on  $\partial\Omega$ , and measurements of the Dirichlet values  $u = f$  on  $\partial\Omega$  for different sources  $h \in H^{-1}(\Omega)$ . Problems of this kind have been discussed by Lowe and Rundell [26, 27] and in an application to semiconductor devices by Fang and Ito [12].

Again, we can decompose the corresponding operators  $F_i$  into the trace operator  $L : H^1(\Omega) \rightarrow L^2(\partial\Omega)$  concatenated with the parameter-to-solution maps  $H_i : q \mapsto u_i$  defined by the solution of

$$\begin{aligned} -\Delta u_i + q u_i &= h_i, & \text{in } \Omega, \\ \frac{\partial u_i}{\partial \nu} &= 0 & \text{on } \partial\Omega. \end{aligned}$$

The derivative  $H_i'(q)$  is almost the same as in the previous example, except for a change from Dirichlet to Neumann boundary conditions. One can verify the regularity of  $u_i$  in the same way as above for  $h_i > 0$  (which allows to apply a maximum principle for  $u_i$ ), and consequently show that the nonlinearity condition (2.2), (2.3) holds.

EXAMPLE 3 (SPECT). In the application of *Single Photon Emission Computed Tomography (SPECT)* one wants to compute the source  $f$  and the coefficient  $a \geq 0$  from

$$\theta_i \cdot \nabla u_i + a u_i = f \quad \text{in } \Omega \subset \mathbb{R}^d,$$

for different values  $\theta_i$  on the unit sphere, and the boundary values

$$\begin{aligned} u_i &= 0 & \text{on } \partial\Omega_i^- := \{ x \in \partial\Omega \mid \nu(x) \cdot \theta_i \leq 0 \}, \\ u_i &= g_i & \text{on } \partial\Omega_i^+ := \{ x \in \partial\Omega \mid \nu(x) \cdot \theta_i \geq 0 \}. \end{aligned}$$

Here, the condition on  $\partial\Omega_i^-$  has to be understood as the boundary condition, while the values  $g_i$  on  $\partial\Omega_i^+$  are the measurements. Thus, the operators  $F_i$  map  $(a, f)$  to  $g_i$ . They can be decomposed into the trace operators  $L_i : L^2(\Omega) \rightarrow H^{-\frac{1}{2}}(\partial\Omega_i^+)$  and the parameter-to-solution maps  $H_i : \mathcal{D}(F_i) \subseteq L^2(\Omega)^2 \rightarrow L^2(\Omega)$ ,  $(a, f) \mapsto u_i$ .

It can be shown (cf. [30]) that the derivative  $v_i = H_i'(a, f)(\hat{a}, \hat{f})$  can be determined as the unique solution of

$$\theta_i \cdot \nabla v_i + a v_i = \hat{f} - \hat{a} u_i \quad \text{in } \Omega \subset \mathbb{R}^d,$$

subject to  $v_i = 0$  on  $\partial\Omega_i^-$ . If  $a > 0$ ,  $f > 0$ , a maximum principle applies also to the first-order equation and one may conclude  $u_i > 0$ , which subsequently can be used to verify the nonlinearity condition (2.2), (2.3) in the same way as for the above examples.

EXAMPLE 4 (Ultrasound Tomography). The inverse problem in ultrasound tomography consists in finding  $f \in L^2(\Omega)$  from boundary measurements  $g_i = u_i$  on  $\partial\Omega$  for complex-valued waves  $u_i = e^{ikx \cdot \theta_i} + v_i$ , where  $v_i$  solves the Helmholtz equations

$$\begin{aligned} \Delta v_i + k^2(1 - f)v_i &= k^2 f e^{ikx \cdot \theta_i} & \text{in } \Omega, \\ \frac{\partial v_i}{\partial \nu} &= B v_i & \text{on } \partial\Omega, \end{aligned}$$

with  $B$  being an appropriate operator representing the radiation condition, and  $k$  a real parameter controlling the spatial resolution. Again we can decompose the operator  $F_i : f \mapsto g_i$  into the trace operator  $L : H^1(\Omega) \rightarrow L^2(\partial\Omega)$  and the parameter-to-solution map  $H_i : \mathcal{D}(F_i) \subseteq L^2(\Omega) \rightarrow H^1(\Omega)$ ,  $f \mapsto u_i$ .

One can show (cf. [30]) that the derivative  $w_i = H_i'(f)$  is defined by the solution of

$$\begin{aligned} \Delta w_i + k^2(1 - f)w_i &= k^2 f w_i & \text{in } \Omega, \\ \frac{\partial w_i}{\partial \nu} &= B w_i & \text{on } \partial\Omega. \end{aligned}$$

If  $f, k$  are such that the operator  $\Delta + k^2(1 - f)$  is regular, and if  $|u_i| \neq 0$ , then one can easily verify the nonlinearity condition in the same way as for the examples above.

EXAMPLE 5 (Nonlinear Moment Estimation). We finally consider a nonlinear moment estimation problem, which consists in finding  $u \in L^2(\Omega)$ ,  $\Omega \subset \mathbb{R}^d$  a bounded domain, given

$$g_i := \int_{\Omega} k_i(x, u(x)) \, dx \in \mathbb{R}^m$$

for given smooth kernel functions  $k_i : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^m$  (which could e.g. arise from the discretization of an integral kernel, i.e.,  $k_i(x, u(x)) = K(x, u(x), y_i)$ ). Here the operator  $F_i : L^2(\Omega) \rightarrow \mathbb{R}^m$  is the concatenation of the linear integration operator  $L : L^2(\Omega)^m \rightarrow \mathbb{R}^m$ ,  $w \mapsto \int_{\Omega} w \, dx$  and the Nemitskij-type operator  $H_i : L^2(\Omega) \rightarrow L^2(\Omega)$ ,  $u \mapsto k_i(\cdot, u)$ . The derivative of the nonlinear operator  $H_i$  is given by

$$H'_i(u)v = \frac{\partial k_i}{\partial u}(\cdot, u)v.$$

If  $k_i \in C(\Omega, C_b^{1,1}(\mathbb{R}))$  and  $\frac{\partial k_i}{\partial u} \neq 0$ , then  $H'_i(u)$  is regular and the map  $u \mapsto H'_i(u)$  is Lipschitz continuous, which implies the nonlinearity condition (2.2), (2.3).

**2.3. Source Conditions.** Convergence of regularization methods for ill-posed problems is, as a direct consequence of the instability, in general arbitrarily slow. In order to obtain convergence rates, additional regularity assumptions on the difference between an exact solution  $x^\dagger$  and some initial guess  $x_0$  used in the regularization method, have to be made. These have the form of so-called source wise representation conditions and in our context read as

$$x^\dagger - x_{0,i} = f(F'_i(x^\dagger)^* F'_i(x^\dagger))w_i \quad i = 0, \dots, p-1 \quad (2.8)$$

for some  $w_i$ , where  $f$  is some real function and for the positive semidefinite operator  $F'_i(x^\dagger)^* F'_i(x^\dagger)$ ,  $f(F'_i(x^\dagger)^* F'_i(x^\dagger))$  is defined via functional calculus (cf. e.g. [11]). Condition (2.8) expresses the assumed regularity of  $x^\dagger - x_{0,i}$  in terms of the smoothing property of  $F'(x^\dagger)$  mentioned above. Typical functions  $f$  used here are

$$f(\lambda) := f_\nu^H(\lambda) := \lambda^\nu \quad (2.9)$$

for some Hölder exponent  $\nu$ , or the weaker, but for exponentially ill-posed problems more appropriate logarithmic functions

$$f(\lambda) := f_\mu^L(\lambda) = (-\ln(\lambda))^{-\mu} . \quad (2.10)$$

REMARK 1. Note that under sufficiently strong source conditions, namely (2.8) with  $f(\lambda) = O(\sqrt{\lambda})$ , which corresponds to  $\nu \geq \frac{1}{2}$  in (2.9), the nonlinearity assumptions on  $F$  (see the previous subsection) can be considerably relaxed in case  $p = 1$ . In place of (2.4) or (2.5) one only needs Lipschitz continuity of  $F'$ . This was already observed by Bakushinsky in [1] for the case  $\nu \geq 1$  and IRGNM (see [21] for the case  $\nu \geq \frac{1}{2}$ ) and later extended to several instances of a general  $G_\alpha$  (see the monograph by Bakushinsky and Kokurin [2], as well as [22]) including those methods  $G_\alpha$  that are considered in this paper (see Section 3.3 below). Actually, it can be shown that the same holds true for the Newton-Kaczmarz method (1.10) for multiple equations.

**3. Convergence Analysis.** In this section we will state a quite general convergence theorem. Its proof is closely related to convergence proofs in [6, 19, 20, 21, 22]. Therefore we shall provide the proof in a somewhat compressed form, but highlight the important ideas for convenience of the reader. We aim at giving the statements in a general and comprehensive way so that they might be of interest even for the special case  $p = 1$ , i.e., for (1.1). Especially, according to the authors' knowledge, the convergence result with logarithmic source conditions under nonlinearity assumptions of the type considered here is new also for  $p = 1$ .

**3.1. Preliminaries and Assumptions.** In order to be able to carry out the estimates in the proof of Theorem 3.1, we have to make some assumptions on the regularization methods  $G^\alpha$ , additional to (1.5), (1.6) in the introduction, i.e.,

$$G_\alpha^i(K)y \rightarrow K^\dagger y \quad \text{as } \alpha \rightarrow 0 \quad \forall y \in \mathcal{R}(K), \quad (3.1)$$

and

$$\|G_\alpha^i(K)\| \leq \Phi(\alpha), \quad (3.2)$$

for all  $K \in \mathcal{L}(X, Y_i)$ , with  $\|K\| \leq C_S^i$ : In view of the nonlinearity condition (2.2), we assume that

$$\|G_\alpha^i(KR)KR - G_\alpha^i(K)K\| \leq \bar{C}_G \|R - I\| \quad (3.3)$$

for all  $K \in \mathcal{L}(X, Y_i)$ ,  $R \in \mathcal{L}(X, X)$  with  $\|K\|, \|KR\| \leq C_S^i$ ,  $\|R - I\| \leq c < 1$ , as well as

$$\|G_\alpha^i(K)K\| \leq C_G, \quad \forall K \in \mathcal{L}(X, Y_i) : \|K\| \leq C_S^i, \quad (3.4)$$

with positive real constants  $\bar{C}_G$ ,  $c$ , and  $C_S^i$  as in (2.1). To yield convergence rates under additional regularity conditions (2.8), the regularizing operators  $G_\alpha^i$  have to converge to the inverse of  $K$  at some rate on the set of solutions satisfying (2.8), i.e., a condition of the form

$$\|(I - G_\alpha^i(K)K)f(K^*K)\| \leq \psi(\alpha) \quad \forall K \in \mathcal{L}(X, Y_i) : \|K\| \leq C_S^i \quad (3.5)$$

is needed, with a strictly monotone function  $\psi$  that decreases to zero as  $\alpha \rightarrow 0$ . Moreover, the sequence  $\psi(\alpha_n)$  must not tend to zero too fast, in the sense that

$$\frac{\psi(\alpha_n)}{\psi(\alpha_{n+1})} \leq C_\psi \quad \forall n \in \mathbb{N}, \quad (3.6)$$

for some constant  $C_\psi \in \mathbb{R}^+$ .

In the situation of noisy data, convergence of the reconstructions as the noise level  $\delta$  tends to zero is only obtained for appropriate choices of the stopping index  $N = N(\delta)$  in dependence of the noise level  $\delta$ . In the general case, convergence can be achieved if  $N(\delta)$  is chosen such that

$$N(\delta) \rightarrow \infty \quad \text{and} \quad \Phi(\alpha_{N(\delta)}) \cdot \delta \rightarrow 0 \quad \text{as } \delta \rightarrow 0 \quad (3.7)$$

and

$$\Phi(\alpha_n) \cdot \delta \leq \tau \quad \forall n \leq N(\delta), \quad (3.8)$$

for some  $\tau > 0$  sufficiently small. If additional source conditions (2.8) hold, an appropriate choice is such that

$$\phi(\alpha_{N(\delta)}) \leq \delta < \phi(\alpha_n) \quad \forall n \leq N(\delta) \quad (3.9)$$

where

$$\phi(\alpha) = \frac{\tau\psi(\alpha)}{\Phi(\alpha)}$$

for  $\tau$  defined in (3.8).

**3.2. Main Result.** Now we shall state and prove the main convergence result of this paper, a comprehensive convergence theorem for Newton-Kaczmarz methods:

**THEOREM 3.1.** *Let  $x_n$  be defined by the sequence (1.10) with Fréchet differentiable operators  $F_i$  satisfying (2.1), (2.2) with (2.3), data  $y^\delta$  satisfying (1.4), the regularization methods  $G_\alpha^i$  fulfilling (3.1), (3.2), (3.3), (3.4), for all  $K \in \mathcal{L}(X, Y_i)$ ,  $R \in \mathcal{L}(X, X)$  with  $\|K\|, \|KR\| \leq C_S^i$ ,  $\|R - I\| \leq c < 1$ , and (3.5), as well as a sequence  $\alpha_n$  tending to zero and satisfying (3.6). Moreover, let  $\tau$  and  $\|x_{0,i} - x^\dagger\|$  be sufficiently small and  $x_{0,i} - x^\dagger \in \mathcal{N}(F_i'(x^\dagger)^\perp)$ ,  $i = 0, \dots, p-1$ .*

*Then, in the noise free case ( $\delta = 0$ ), the sequence  $x_n$  converges to  $x^\dagger$  as  $n \rightarrow \infty$ . In case of noisy data and with the choice (3.7), (3.8),  $x_{N(\delta)}$  converges to  $x^\dagger$  as  $\delta \rightarrow 0$ .*

*If the source conditions (2.8) and (3.5), (3.6) hold, with  $\|w^i\|$  sufficiently small, then the convergence rates*

$$\|x_n - x^\dagger\| = O(\psi(\alpha_n))$$

*in the noise free situation and, with (3.9),*

$$\|x_{N(\delta)} - x^\dagger\| = O(\psi(\phi^{-1}(\delta)))$$

*in the noisy situation, respectively, hold.*

*Proof.* We will make use of the following Lemma, whose proof can be found in [21]:

**LEMMA 3.2.** *Let  $\{a_n\}$  be a sequence satisfying*

$$0 \leq a_n \leq a \quad \text{and} \quad \lim_{n \rightarrow \infty} a_n = \tilde{a} \leq a .$$

*Moreover, we assume that  $\{\gamma_n\}$  is a sequence for which the estimate*

$$0 \leq \gamma_{n+1} \leq a_n + b\gamma_n + c\gamma_n^2, \quad n \in \mathbb{N}_0, \gamma_0 \geq 0$$

*holds for some  $b, c \geq 0$ . Let  $\underline{\gamma}$  and  $\overline{\gamma}$  be defined as*

$$\underline{\gamma} := \frac{2a}{1 - b + \sqrt{(1-b)^2 - 4ac}}, \quad \overline{\gamma} := \frac{1 - b + \sqrt{(1-b)^2 - 4ac}}{2c} .$$

*If  $b + 2\sqrt{ac} < 1$  and if  $\gamma_0 \leq \overline{\gamma}$ , then*

$$\gamma_n \leq \max(\gamma_0, \underline{\gamma}), \quad n \in \mathbb{N}_0$$

*and if  $\tilde{a} < a$ , then*

$$\limsup_{n \rightarrow \infty} \gamma_n \leq \frac{2\tilde{a}}{1 - b + \sqrt{(1-b)^2 - 4\tilde{a}c}} .$$

To derive a recursive error estimate, we assume that the current iterate  $x_n$  is in  $\mathcal{B}_\rho(x^\dagger)$  and that  $n < N(\delta)$  ( $= \infty$  if  $\delta = 0$ ). Then

$$\begin{aligned} x_{n+1} - x^\dagger &= \left( I - G_{\alpha_n}^n(F'_n(x^\dagger))F'_n(x^\dagger) \right) (x_{0,i} - x^\dagger) \\ &\quad + \left( G_{\alpha_n}^n(F'_n(x^\dagger))F'_n(x^\dagger) - G_{\alpha_n}^n(F'_n(x_n))F'_n(x_n) \right) (x_{0,i} - x^\dagger) \\ &\quad - G_{\alpha_n}^n(F'_n(x_n))(F_n(x_n) - F_n(x^\dagger) - F'_n(x_n)(x_n - x^\dagger)) \\ &\quad - G_{\alpha_n}^n(F'_n(x_n))(y_n - y_n^\delta) \end{aligned} \quad (3.10)$$

The third term on the right hand side can be rewritten as

$$G_{\alpha_n}^n(F'_n(x_n))F'_n(x_n) \int_0^1 \left( R^i(x^\dagger + \theta(x_n - x^\dagger), x_n) - I \right) d\theta(x_n - x^\dagger),$$

with  $i = \text{mod}(n, p)$ , so

$$\begin{aligned} \|x_{n+1} - x^\dagger\| &\leq \xi_n \\ &\quad + \bar{C}_G C_R \|x_{0,i} - x^\dagger\| \|x_n - x^\dagger\| \\ &\quad + \frac{1}{2} C_G C_R \|x_n - x^\dagger\|^2 \\ &\quad + \Phi(\alpha_n) \delta, \end{aligned} \quad (3.11)$$

where

$$\xi_n := \left\| \left( I - G_{\alpha_n}^n(F'_n(x^\dagger))F'_n(x^\dagger) \right) (x_{0,i} - x^\dagger) \right\| \leq \psi(\alpha_n) \|w^i\|,$$

if (2.8) holds and

$$\xi_n \rightarrow 0 \quad \text{as } n \rightarrow \infty \quad (3.12)$$

also without (2.8). The latter can be seen by (3.1) together with the following subsequence-subsequence argument:

Let  $(\xi_{n_m})_{m \in \mathbb{N}}$  be an arbitrary subsequence of  $(\xi_n)_{n \in \mathbb{N}}$ . Then there exists an  $i \in \{0, \dots, p-1\}$  such that the set  $\{m \in \mathbb{N} \mid \text{mod}(n_m, p) = i\}$  has infinite cardinality. Define by  $(m_l)_{l \in \mathbb{N}}$  a numbering of this set in ascending order, then for  $(\xi_{n_{m_l}})_{l \in \mathbb{N}}$  we get

$$\xi_{n_{m_l}} = \left\| \left( I - G_{\alpha_{n_{m_l}}}^i(F'_i(x^\dagger))F'_i(x^\dagger) \right) (x_{0,i} - x^\dagger) \right\| \rightarrow 0 \text{ for } l \rightarrow \infty,$$

since  $\alpha_{n_{m_l}} \rightarrow 0$  for  $l \rightarrow \infty$ .

Now we can apply induction, together with Lemma 3.2 to the sequence

$$\gamma_n := \|x_n - x^\dagger\|.$$

The boundedness (3.8) in the stopping rule and our assumption on closeness of  $x_{0,i}$  to  $x^\dagger$  and on smallness of  $\tau$  permit to make the constants  $a$  and  $b$  sufficiently small so that the assumptions of the Lemma are satisfied, and the bound  $\max\{\gamma_0, \underline{\gamma}\}$  is smaller than  $\rho$ , so that we can guarantee that the iterates remain in  $\mathcal{B}_\rho(x^\dagger)$  for all  $n \leq N(\delta)$ . Moreover, by (3.12) as well as the asymptotics (3.7) in the stopping rule, we can set

$\tilde{a} = 0$  and conclude and that  $x_n$  converges to  $x^\dagger$  as  $n \rightarrow \infty$  in the noise free case, and as  $\delta \rightarrow 0$  in the noisy case, respectively.

To prove convergence rates under source conditions, we consider the sequence

$$\gamma_n := \frac{\|x_n - x^\dagger\|}{\psi(\alpha_n)},$$

that satisfies

$$\gamma_{n+1} \leq C_\psi \left( \|w^i\| + \bar{C}_G C_R \|x_{0,i} - x^\dagger\| \gamma_n + \frac{1}{2} C_G C_R \psi(\alpha_n) \gamma_n^2 + \frac{\Phi(\alpha_n)}{\psi(\alpha_n)} \delta \right).$$

Hence, Lemma 3.2 together with the stopping rule (3.9) implies that  $x_n$  remains in  $\mathcal{B}_\rho(x^\dagger)$  for all  $n \leq N(\delta)$ , and that  $\gamma_n$  is uniformly bounded, i.e.,

$$\|x_n - x^\dagger\| \leq C \psi(\alpha_n), \quad (3.13)$$

for some constant  $C$ . This immediately yields the convergence rate result in the noiseless case. To obtain the error estimate in terms of  $\delta$  in the noisy case, we make use of the fact that by (3.9)

$$\delta \geq \phi(\alpha_{N(\delta)}),$$

which, since  $\psi$  and  $\phi$  are strictly monotonically increasing, by (3.13) implies

$$\psi(\phi^{-1}(\delta)) \geq \psi(\alpha_{N(\delta)}) \geq \frac{1}{C} \|x_{N(\delta)} - x^\dagger\|.$$

□

The assumption

$$x_{0,i} - x^\dagger \in \mathcal{N}(F'_i(x^\dagger))^\perp, \quad i = 0, \dots, p-1, \quad (3.14)$$

is rather limiting, since the dimensionality of  $x_0 - x^\dagger$  is related to the "smaller" space  $\mathcal{N}(F'_i(x^\dagger))^\perp$ . In the special case  $p = 1$ , the difference between  $x_0$  and an  $x_0$ -minimum-norm-solution  $x^\dagger$  will automatically lie within  $\mathcal{N}(F'(x^\dagger))^\perp$  under certain nonlinearity conditions (see Proposition 2.1 in [21]). However, for general  $p > 1$  one only gets  $x_0 - x^\dagger \in \left( \bigcap_{i=0}^{p-1} \mathcal{N}(F'_i(x^\dagger)) \right)^\perp$  and not (3.14) with  $x_{0,i} := x_0$ . Thus, condition (3.14) requires the choice of appropriate initial guesses  $x_{0,i}$ . To see necessity of condition (3.14) for convergence, consider the linear case

$$F_i x = y_i, \quad i = 0, \dots, p-1, \quad (3.15)$$

with  $F_i \in \mathcal{L}(X, Y_i)$ ,  $y_i \in Y_i$ ,  $i = 0, \dots, p-1$ , where the sequence  $x_n$  is defined by

$$x_{n+1} = x_{0,n} - G_{\alpha_n}^n(F_n)(F_n x_{0,n} - y_n^\delta). \quad (3.16)$$

In case of exact data, the error can be written as

$$x_{kp+i+1} - x^\dagger = (I - G_{\alpha_{kp+i}}^i(F_i)F_i)(x_{0,i} - x^\dagger), \quad (3.17)$$

for  $n = kp + i$ ,  $k \in \mathbb{N}$ , so by (3.1) and  $\alpha_n \rightarrow \infty$  as  $n \rightarrow \infty$ ,

$$x_{kp+i+1} - x^\dagger \rightarrow \mathcal{P}_{\mathcal{N}(F_i)^\perp}(x_{0,i} - x^\dagger) \text{ as } k \rightarrow \infty,$$

whence convergence of  $x_n$  to  $x^\dagger$  as  $n \rightarrow \infty$  implies (3.14).

In this sense, Theorem 3.1 means that the regularized Newton Kaczmarz method is as least as good as application of Newton's method separately to each of the  $p$  equations, which might a priori not be evident due to the mixing up of the equations during the iteration (1.10). Since it takes into account more information, it should intuitively be even better, which is also reflected in our numerical tests, that showed convergence without any specific choice of the initial guesses.

Note that in the linear case, subsequent iterates completely decouple, i.e., subsequences  $(x_{kp+i_1})_{k \in \mathbb{N}}$ ,  $(x_{kp+i_2})_{k \in \mathbb{N}}$  are independent of each other for  $i_1 \neq i_2$ . Thus it suffices to have

$$x_{0,i} - x^\dagger \in \mathcal{N}(F_i)^\perp \quad (3.18)$$

for one  $i = \bar{i} \in \{0, \dots, p-1\}$ , to obtain convergence of the respective subsequence  $x_{k_{p+\bar{i}+1}}$  from standard results for linear regularization methods. The same holds true for convergence rates. Consequently, in order to get convergence (and convergence rates) with noisy data, it suffices to have (3.18) (and  $x_{0,i} - x^\dagger \in \mathcal{R}(f(F_i^* F_i))$ ) for one  $i = \bar{i} \in \{0, \dots, p-1\}$  only, and to stop the iteration at an index from the respective subsequence  $k_{p+\bar{i}+1}$  with  $k_* = k_*(\delta)$  being determined a priori from (3.7), (3.8), (3.9) or, alternatively, a posteriori from a discrepancy principle

$$\|F_i x_{k_{p+\bar{i}+1}} - y_i\| \leq \tau \delta < \|F_i x_{k_{p+\bar{i}+1}} - y_i\|, \quad 0 \leq k < k_* .$$

Unfortunately this complete decoupling gets lost as soon as the operators  $F_i$  are nonlinear. Moreover, we have to remark that already (3.18) for one  $i = \bar{i}$  is a very strong condition in case  $p > 1$ , since it means the other equations for  $i \neq \bar{i}$  are not really required for determining  $x^\dagger$ .

**3.3. Standard Regularizing Operators.** Now we shall apply Theorem 3.1 to some regularization methods  $G^i$  of particular interest. Moreover, in the abstract source condition (2.8), we insert the most relevant special cases of a Hölder function  $f$  in (2.9) or a logarithmic function  $f$  in (2.10).

As important examples from a larger class of regularization methods defined by real functions  $g_\alpha : \mathbb{R}^+ \mapsto \mathbb{R}^+$  approximating  $\lambda \mapsto \frac{1}{\lambda}$  and

$$G_\alpha(K) := g_\alpha(K^* K) K^* \quad (3.19)$$

via functional calculus (cf., e.g., [11, 25]), we consider

- Tikhonov-Philips regularization:

$$G_\alpha(K) = (K^* K + \alpha I)^{-1} K^*, \quad I - G_\alpha(K) K = \alpha (K^* K + \alpha I)^{-1} \quad (3.20)$$

In this case, we shall call the arising iterative method *iteratively regularized Gauss-Newton-Kaczmarz (IRGNK) method*.

- iterated Tikhonov regularization:

$$\begin{aligned} G_\alpha(K) &= \sum_{l=0}^k \prod_{j=l}^k \alpha_j (K^* K + \alpha_j I)^{-1} \frac{1}{\alpha_l} K^*, \\ I - G_\alpha(K) K &= \prod_{l=0}^k \alpha_l (K^* K + \alpha_l I)^{-1}, \end{aligned} \quad (3.21)$$

with the effective regularization parameter

$$\alpha := \frac{1}{\sum_{l=0}^k \frac{1}{\alpha_l}}.$$

We shall call the arising iterative method *k-iteratively regularized Gauss-Newton-Kaczmarz (IRGNK<sub>k</sub>) method*. Here we distinguish between the special stationary case

$$\alpha_l \equiv 1, \quad (3.22)$$

i.e., Lardy's method, and the (due to its faster convergence more attractive, cf. [16]) nonstationary case of, e.g., geometrically decaying  $\alpha_l$

$$\alpha_l := Cq^l, \quad (3.23)$$

with  $q \in (0, 1)$ .

- Landweber iteration:

$$G_\alpha(K) = \sum_{l=0}^k (I - K^*K)^l K^*, \quad I - G_\alpha(K)K = (I - K^*K)^{k+1}, \quad (3.24)$$

$$\alpha := \frac{1}{k+1},$$

where the scaling is assumed to be done such that  $\|I - K^*K\| \leq 1$ , i.e.,  $C_S^i = \sqrt{2}$  in (2.1)). For obvious reasons, this method shall be called *Newton-Landweber-Kaczmarz (NLK) method* here and below.

These methods are well known to satisfy (3.1), (3.2) with

$$\Phi(\alpha) = C \frac{1}{\sqrt{\alpha}},$$

as well as (3.4) (cf. e.g., [11, 25, 16]). Moreover, for the Hölder type source representation functions (2.9), they satisfy (3.5) with

$$\psi(\alpha) = C\alpha^\nu \quad (3.25)$$

(where  $\nu$  is restricted to the interval  $[0, 1]$  in Tikhonov regularization and to the interval  $[0, k]$  in iterated Tikhonov regularization). from which one can conclude by Lemma 4 in [20], that they also satisfy (3.5) for the logarithmic functions (2.10) with

$$\psi(\alpha) = C(-\ln(\alpha))^{-\mu}$$

where w.l.o.g., both  $\|K\|^2$  and  $\alpha$  are restricted to the interval  $(0, \exp(-1)]$  (i.e.,  $C_S^i = \exp(-1/2)$  in (2.1)) in order to avoid the singularity of  $f_\mu^L$  at zero. Therewith, a decay restriction

$$\frac{\alpha_n}{\alpha_{n+1}} \leq C_\alpha \quad \forall n \in \mathbb{N} \quad (3.26)$$

is sufficient for (3.6).

COROLLARY 3.3. Let  $x_n$  be defined by the sequence (1.10) with Fréchet differentiable operators  $F_i$  satisfying (2.1), (2.2) with (2.3), data  $y^\delta$  satisfying (1.4) and the regularization methods  $G_\alpha^i$  defined by Tikhonov-Philips regularization, nonstationary iterated Tikhonov regularization, or Landweber iteration, as well as a sequence  $\alpha_n$  tending to zero and satisfying (3.26). Moreover, let  $\tau$  and  $\|x_{0,i} - x^\dagger\|$  be sufficiently small and  $x_{0,i} - x^\dagger \in \mathcal{N}(F_i'(x^\dagger)^\perp)$ ,  $i = 0, \dots, p-1$

Then, the assertions of Theorem (3.1) hold. In particular, under a Hölder type source condition (2.8) with (2.9), we obtain

$$\|x_{N(\delta)} - x^\dagger\| = O(\delta^{\frac{2\nu}{2\nu+1}}),$$

(where  $\nu$  is restricted to  $[0, 1]$  in case of Tikhonov regularization), and under a logarithmic type source condition (2.8) with (2.10)

$$\|x_{N(\delta)} - x^\dagger\| = O((-\ln(\delta^2))^{-\mu}).$$

Note that the saturation of iterated Tikhonov regularization at  $\nu = k$  does not take effect here, since we do not consider  $k$  but  $(\sum_{l=0}^k \alpha_l^{-1})^{-1}$  as the regularization parameter.

*Proof.* It remains to show that the differences between applications of the regularization methods to two different operators can be estimated according to (3.3). For Tikhonov regularization can make use of estimates already presented in [21], as well as in Hohage's thesis [19], namely, for arbitrary  $K \in \mathcal{L}(X, Y_i)$ ,  $R \in \mathcal{L}(X, X)$  with  $\|R - I\| \leq c < 1$ ,

$$\begin{aligned} & \|G_\alpha^i(KR)KR - G_\alpha^i(K)K\| \\ &= \alpha \|(K^*K + \alpha I)^{-1} - ((KR)^*KR + \alpha I)^{-1}\| \\ &= \alpha \|((KR)^*KR + \alpha I)^{-1} \left( (KR)^*KR(I - R^{-1}) + (R - I)^*K^*K \right) (K^*K + \alpha I)^{-1}\| \\ &\leq (1 + \frac{1}{1-c}) \|R - I\| \end{aligned}$$

for  $f$  according to (2.9) with  $\nu \leq \frac{1}{2}$  or  $f$  according to (2.10).

For the iterative methods — iterated Tikhonov regularization and for Landweber iteration — we make use of the identity

$$\prod_{l=0}^k A_l - \prod_{l=0}^k B_l = \sum_{l=0}^k \prod_{j=0}^{l-1} A_j (A_l - B_l) \prod_{j=l+1}^k B_j \quad (3.27)$$

for linear operators  $A_l, B_l$ , with the notation  $\prod_{l=0}^{-1} A_l = I = \prod_{l=k+1}^k B_j$ , and first of all consider case a):

To obtain (3.3) for Landweber iteration, we set

$$A_l := (I - (KR)^*KR), \quad B_l := (I - K^*K), \quad (3.28)$$

and use the fact that

$$A_l - B_l = (KR)^*KR(R^{-1} - I) + (I - R^*)K^*K \quad (3.29)$$

to derive

$$\begin{aligned} G_\alpha^i(K)K - G_\alpha^i(KR)KR &= \sum_{l=0}^k \prod_{j=0}^{l-1} A_j (KR)^*KR(R^{-1} - I) \prod_{j=l+1}^k B_j \\ &+ \sum_{l=0}^k \prod_{j=0}^{l-1} A_j (I - R^*)K^*K \prod_{j=l+1}^k B_j \end{aligned} \quad (3.30)$$

To estimate the sums from 0 to  $k$  we decompose them into sums from 0 to  $\lfloor \frac{k}{2} \rfloor$  and from  $\lfloor \frac{k}{2} \rfloor + 1$  to  $k$  and use the fact that

$$\left\| \prod_{j=0}^{l-1} A_j(KR)^*KR \right\| \leq \frac{1}{l+1}, \quad \left\| \prod_{j=l+1}^k K^*KB_j \right\| \leq \frac{1}{k-l+1} \quad (3.31)$$

as well as

$$(KR)^*KR = I - A_l, \quad K^*K = I - B_l \quad (3.32)$$

and a telescope sum trick to obtain, for the first sum on the right hand side of (3.30)

$$\left\| \sum_{l=\lfloor \frac{k}{2} \rfloor + 1}^k \prod_{j=0}^{l-1} A_j(KR)^*KR(R^{-1} - I) \prod_{j=l+1}^k B_j \right\| \leq \sum_{l=\lfloor \frac{k}{2} \rfloor + 1}^k \frac{1}{l+1} \|R^{-1} - I\| \leq \|R^{-1} - I\|$$

and

$$\begin{aligned} & \left\| \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \prod_{j=0}^{l-1} A_j(KR)^*KR(R^{-1} - I) \prod_{j=l+1}^k B_j \right\| \\ &= \left\| \sum_{l=0}^{\lfloor \frac{k}{2} \rfloor} \prod_{j=0}^{l-1} A_j(I - A_l)(R^{-1} - I) \prod_{j=l+1}^k B_j \right\| \\ &= \left\| \sum_{l=1}^{\lfloor \frac{k}{2} \rfloor} \prod_{j=0}^{l-1} A_j(R^{-1} - I)(I - B_l) \prod_{j=l+1}^k B_j + \right. \\ & \quad \left. (R^{-1} - I) \prod_{j=1}^k B_j + \prod_{j=0}^{\lfloor \frac{k}{2} \rfloor} A_j(R^{-1} - I) \prod_{j=\lfloor \frac{k}{2} \rfloor + 1}^k B_j \right\| \\ &\leq \left( \sum_{l=1}^{\lfloor \frac{k}{2} \rfloor} \frac{1}{k-l+1} + 2 \right) \|R^{-1} - I\| \leq 3 \|R^{-1} - I\|. \end{aligned}$$

and analogously for the second sum on the right hand side of (3.30).

For iterated Tikhonov regularization, the estimates can be obtained analogously, with this time

$$A_l := \alpha_l((KR)^*KR + \alpha_l I)^{-1}, \quad B_l := \alpha_l(K^*K + \alpha_l I)^{-1},$$

$$A_l - B_l = A_l \frac{1}{\alpha_l} \left( (KR)^*KR(R^{-1} - I) + (I - R^*)K^*K \right) B_l$$

$$\begin{aligned} G_\alpha^i(K)K - G_\alpha^i(KR)KR &= \sum_{l=0}^k \prod_{j=0}^l A_j(KR)^*KR(R^{-1} - I) \prod_{j=l}^k B_j \\ & \quad + \sum_{l=0}^k \prod_{j=0}^l A_j(I - R^*)K^*K \prod_{j=l}^k B_j \end{aligned} \quad (3.33)$$

$$\left\| \prod_{j=0}^l A_j (KR)^* KR \right\| \leq \left( \sum_{j=0}^l \frac{1}{\alpha_j} \right)^{-1}, \quad \left\| \prod_{j=l}^k K^* K B_l \right\| \leq \left( \sum_{j=l}^k \frac{1}{\alpha_j} \right)^{-1}$$

and

$$\frac{1}{\alpha_l} (KR)^* KR A_l = I - A_l, \quad \frac{1}{\alpha_l} B_l K^* K = I - B_l$$

in place of (3.28), (3.29), (3.30), (3.31), (3.32), respectively. In the stationary case (3.22) again cutting the sum at  $\lfloor \frac{k}{2} \rfloor$  and the telescope trick have to be used, and in the nonstationary case (3.23), we apply the telescope sum trick to the whole first sum in (3.33), and leave the second sum unchanged.

□

We finally want to mention that these results can be extended to the situation where discretization is applied to any of the standard regularization methods.

**3.4. Levenberg-Marquardt-Kaczmarz.** An alternative to considering the regularized Newton-Kaczmarz approach (1.10) is the generalization of a Levenberg-Marquardt method (cf. [15] as well as (1.7) in the introduction) to the situation of multiple equations in the following form:

$$x_{n+1} = x_n - (F'_n(x_n)^* F'_n(x_n) + \alpha_n I)^{-1} F'_n(x_n)^* (F_n(x_n) - y_n^\delta).$$

Note that this formally corresponds to the (intuitively optimal) formal choice of  $x_{0,n} = x_n$  in (1.10), which however is not admissible in view of the convergence analysis given here, that requires a cyclic repetition of the starting guesses according to  $x_{0,n} = x_{0, \text{mod}(n,p)}$ .

Under a nonlinearity condition of the type (2.6) and with an appropriate a posteriori choice of the sequence  $\alpha_n$ , along the lines of the proofs in [15], and similarly to [24], one can show that the error  $\|x_n - x^\dagger\|$  is monotonically decreasing up to an index  $n = N(\delta)$  determined by the discrepancy principle, without having to make assumptions of the type (3.14). Moreover the norms of the residuals are squared summable in case of exact data and therewith

$$F_n(x_n) - y_n \rightarrow 0 \text{ as } n \rightarrow \infty. \quad (3.34)$$

This implies that there exists a weakly convergent subsequence of  $x_n$ . However the limit of a weakly convergent subsequence  $(x_{n_l})_{l \in \mathbb{N}}$  of  $(x_n)_{n \in \mathbb{N}}$  need not necessarily be a solution to (1.2), even if the  $F_i$  are (weakly) sequentially closed, i.e., for any sequence  $(x_k)_{k \in \mathbb{N}} \subseteq \mathcal{D}(F_i)$  and  $f_i \in Y_i$ ,

$$\left( x_k \rightharpoonup x \wedge F_i(x_k) \rightarrow f_i \right) \Rightarrow \left( x \in \mathcal{D}(F_i) \wedge F_i(x) = f_i \right). \quad (3.35)$$

Namely, if, e.g.,  $(x_{n_l})_{l \in \mathbb{N}} \subseteq (x_{mp+\bar{i}})_{m \in \mathbb{N}}$  for some  $\bar{i} \in \{0, \dots, p-1\}$ , then (3.34) and (3.35) imply that the weak limit of  $(x_{n_l})_{l \in \mathbb{N}}$  is a solution of  $F_{\bar{i}}(x) = y_{\bar{i}}$  only but not necessarily of  $F_i(x) = y_i$  with  $i \neq \bar{i}$ . Also, strong convergence to  $x^\dagger$  of  $x_n$  as  $n \rightarrow \infty$  in the case of exact data or of  $x_{N(\delta)}$  in the noisy situation, cannot be proved by methods like those used in [15], [24], even in the linear case. Still, necessary convergence conditions on the initial guess can be expected to be less restrictive for (1.7) than for (1.10) as the linear case with bounded generalized inverses indicates: Setting all regularization parameters  $\alpha_n$  to zero we arrive at the error recursion

$$x_{n+1} - x^\dagger = \mathcal{P}_{N(F_n)}(x_n - x^\dagger) = \mathcal{P}_{N(F_n)} \mathcal{P}_{N(F_{n-1})} \cdots \mathcal{P}_{N(F_0)}(x_0 - x^\dagger),$$

so that one even obtains termination of the iteration with  $x_{n+1} = x^\dagger$  as soon as  $\mathcal{P}_{N(F_{n-1})} \cdots \mathcal{P}_{N(F_0)}(x_0 - x^\dagger) \in \mathcal{N}(F_n)^\perp$  for some  $n$ .

**4. Numerical Solution Methods.** In the following we discuss some possible discretization strategies and methods for the solution of the arising finite-dimensional problems.

**4.1. Primal Method.** For all of the optimization approaches discussed above, one can use a standard Galerkin discretization strategy, by choosing a finite-dimensional subspace  $X^h \subset X$  and solving a weak form of the discretized Newton equation for  $x_{n+1}^h$ . For the IRGNK method, we have  $G_{\alpha_n}^n = M_n^{-1} F_n'(x_n^h)^*$  with the positive definite operator  $M_n := F_n'(x_n^h)^* F_n'(x_n^h) + \alpha_n I$ . Using this special form, we can discretize a step of the IRGNK method via

$$\langle M_n(x_{n+1}^h - x_{0,n}^h), \varphi \rangle = -\langle (F_n(x_n^h) - y_n^\delta - F_n'(x_n^h)(x_n^h - x_{0,n}^h)), F_n'(x_n^h)\varphi \rangle \quad \forall \varphi \in X^h.$$

By iterating this discretization procedure  $k$  times, one obtains a discrete form of the IRGNK $_k$  method. Due to the positive definiteness of  $M_n$ , one can solve this problem iteratively by a preconditioned conjugate gradient method, where all standard preconditioners for the Tikhonov regularization can be used (cf. [34] for an overview).

In the case of the Newton-Landweber iteration, we obtain the same equation for each Landweber step finally leading to  $x_{n+1}^h$ , but now with  $M_n = I$ , which gives a quasi-explicit form for the next iteration (one only has to invert a mass matrix corresponding to the identity operator, which does not even change during the iteration).

**4.2. Dual Method.** In the following we shall consider a dual method for the *iteratively regularized Gauss-Newton-Kaczmarz method (IRGNK)*, i.e., the Newton-Kaczmarz method with the choice  $G_\alpha(K) = (K^*K + \alpha I)^{-1}K^*$ . We shall now derive a dual method, which is particularly suitable for the important case that the output spaces  $Y_i$  are of lower dimensionality than the parameter space  $X$  (which is the case for the examples considered above).

A first observation is that each iteration step of the IRGNK method is equivalent to the minimization problem

$$\frac{1}{2} \|F_n(x_n) + F_n'(x_n)(x - x_n) - y_n\|^2 + \frac{\alpha_n}{2} \|x - x_{0,n}\|^2 \rightarrow \min_{x \in X}. \quad (4.1)$$

By defining the right-hand side  $z := y_n - F_n(x_n) - F_n'(x_n)x_n$ , and the linear operator  $K := F_n'(x_n)$ , this optimization problem is of the form

$$J_1(Kx) + J_2(x) \rightarrow \min_{x \in X}, \quad (4.2)$$

with (omitting the index  $n$  in the regularization parameter  $\alpha_n$ )

$$J_1(y) = \frac{1}{2} \|y - z\|^2, \quad J_2(x) = \frac{\alpha}{2} \|x - x_{0,n}\|^2.$$

Both the functionals  $J_1$  and  $J_2$  are strictly convex, and therefore standard Fenchel duality (cf. [10]) implies that the primal problem (4.2) is equivalent to the dual problem

$$J_1^*(-v) + J_2^*(K^*v) \rightarrow \min_{v \in Y_n}, \quad (4.3)$$

where  $J_1^*$  and  $J_2^*$  are the conjugate functionals, which are obtained as

$$\begin{aligned} J_1^*(v) &= \sup_{y \in Y_n} \langle v, y \rangle - J_1(y) = \frac{1}{2} \|v + z\|^2 - \frac{1}{2} \|z\|^2 \\ J_2^*(w) &= \sup_{x \in X} \langle w, x \rangle - J_2(x) = \frac{1}{2\alpha} \|w + \alpha x_{0,n}\|^2 - \frac{\alpha}{2} \|x_{0,n}\|^2. \end{aligned}$$

Moreover, the solution  $v$  of the dual problem (4.3) and the solution  $x$  of the primal problem are connected by the optimality condition

$$K^*v = J_2'(x) = \alpha(x - x_{0,n}).$$

Thus, we may compute  $x = x_{0,n} + \frac{1}{\alpha} K^*v$  once we have solved the dual problem.

By ignoring the constant terms in the conjugate functionals, we may equivalently state the dual problem as

$$\frac{1}{2} \|-v + z\|^2 + \frac{1}{2\alpha} \|K^*v + \alpha x_{0,n}\|^2 \rightarrow \min_{v \in Y_n}, \quad (4.4)$$

which can be discretized e.g. by the Ritz-method on a subspace of  $Y_n$ , i.e., by minimizing the functional in (4.4) on a finite-dimensional subspace  $Y_n^h \subset Y_n$ . This automatically yields a discretization of the update in the primal space via  $x_n^h - x_{0,n} = \frac{1}{\alpha} K^*v^h$ , where  $v^h$  is the discrete solution of the dual problem.

The main advantage of a dual strategy is the (possible) lower dimensionality of the spaces  $Y_n$ , which yields smaller discrete problems and consequently a faster solution. In many important cases such as the examples presented above, the spaces  $Y_n$  do not depend on the iteration index, but are the same for each step, such that one does not have to change the basis over the Kaczmarz sweep.

**4.3. Primal-Dual Methods for PDE-Constrained Problems.** As we have seen in the examples above, the operator  $F_i$  is defined implicitly via the solution of partial differential equations in many applications. We formally write the partial differential equation as a nonlinear operator equation of the form

$$E_i(u_i, q) = 0,$$

where  $E_i : \mathcal{U} \times X \rightarrow \mathcal{V}$  is a continuously differentiable nonlinear operator such that  $\frac{\partial E_i}{\partial u}$  is regular for each  $u \in \mathcal{U}$ . The operator  $F_i$  is typically obtained as  $F_i := L_i \circ H_i$ , where  $H_i(q) = u_i$ . We shall derive a primal-dual solution method in this case.

One step of the IRGNK method can be rewritten as the constrained problem

$$\frac{1}{2} \|L_n v + L_n u_n - y_n\|^2 + \frac{\alpha_n}{2} \|s + q_n - q_{0,n}\|^2 \rightarrow \min_{(v,s)}$$

subject to the constraint that  $v = H_n'(q_n)s$ , which can be expressed using the implicit function theorem as

$$\frac{\partial E_{n+1}}{\partial u}(u_n, q_n)v + \frac{\partial E_{n+1}}{\partial q}(u_n, q_n)s = 0,$$

where  $u_n = H_n(q_n)$ . Deriving the KKT-conditions for this constrained problem, we obtain an indefinite system for the primal variables  $v$ ,  $s$ , and a dual variable  $w$ , given by

$$\begin{pmatrix} L_n^* L_n & 0 & A_n^* \\ 0 & \alpha_n I & B_n^* \\ A_n & B_n & 0 \end{pmatrix} \begin{pmatrix} v \\ s \\ w \end{pmatrix} = \begin{pmatrix} L_n^* y_n - L_n^* L_n u_n \\ \alpha_n (q_{0,n} - q_n) \\ 0 \end{pmatrix}$$

with the linear operators  $A_n := \frac{\partial E_{n+1}}{\partial u}(u_n, q_n)$  and  $B_n := \frac{\partial E_{n+1}}{\partial q}(u_n, q_n)$ .

This indefinite system can be discretized using a mixed approach, i.e., we look for a solution  $(v^h, s^h, w^h)$  in the finite-dimensional subspaces  $\mathcal{U}^h \times X^h \times \mathcal{V}^h$  satisfying

$$\begin{aligned} \langle L_n v, L_n \varphi \rangle + \langle A_n \varphi, w \rangle &= \langle y_n - L_n u_n, L_n \varphi \rangle \\ \alpha_n \langle s, \sigma \rangle + \langle B_n \sigma, w \rangle &= \alpha_n \langle q_{0,n} - q_n, \sigma \rangle \\ \langle A_n v, \psi \rangle + \langle B_n s, \psi \rangle &= 0 \end{aligned}$$

for all  $(\varphi, \sigma, \psi) \in \mathcal{V}^h \times X^h \times \mathcal{U}^h$ .

The resulting indefinite system can be solved by a preconditioned conjugate gradient method for the Schur complement, or directly by a preconditioned Krylov subspace method for indefinite systems like GMRES, QMR, or MINRES. We refer to [4, 5] for the discussion of solution methods for indefinite systems arising from primal-dual formulations in parameter identification.

**5. Numerical Examples.** In the following we shall present numerical results for two of the examples introduced above.

**5.1. Reconstruction with Multiple Sources.** We start with numerical results for Example 2 in the one-dimensional domain  $\Omega = (0, 1)$ , using  $p = 20$  localized sources of the form

$$h_i(x) = 10e^{-10(x - \frac{i+1}{p+1})^2}$$

The data correspond to the "exact solution"  $q^*(x) = 5 + 5x(1 - x)$  and the initial value is  $q_0 \equiv 5$ . Note that in general we cannot expect the least-squares minimum norm solution  $q^\dagger$  to be equal to  $q^*$ , since we only use a finite number of measurements. However, we shall see below that the resulting limit  $q^\dagger$  is close to  $q^*$ , with a difference probably caused due to the limited numerical resolution only.

For the numerical solution we use the iteratively regularized Gauss-Newton-Kaczmarz method, i.e., Tikhonov regularization in  $H^1(\Omega)$  as the linear regularization method. The iteration is discretized using a primal-dual method as described above, with piecewise linear finite elements on a uniform grid of size  $h = 0.01$ .

We first test the convergence behavior in the noise free case. To this end, we generate the data on the same grid as we later solve the inverse problem and choose the regularization parameters as

$$\alpha_n = \alpha_0 \zeta^{-n} \tag{5.1}$$

with  $\zeta = 1.1$  and  $\alpha_0 = 10^{-5}$ . The convergence behaviour is illustrated in Figures 5.1 and 5.2 by the iterates at several different steps. The behaviour during the first Kaczmarz sweep is illustrated in Figure 5.1. In the iterations 1 and 4, for which we use sources localized close to the left boundary  $x = 0$ , the convergence is more pronounced close to the left boundary. Vice versa, for iterations 16 and 20, with sources localized close to the right boundary  $x = 1$ , the reconstruction is better close to the right boundary. In the medium stage of a Kaczmarz sweep, at iterates 8 and 12 with sources localized in the middle of the interval  $(0, 1)$ , the iterate appears almost symmetric. In the later stage of the iteration we plot the iterates  $q_n$  at  $n = 30, 40, 50, 60$  (i.e., those in the middle and at the end of a Kaczmarz sweep) in Figure 5.2. One observes convergence of the algorithm, which turns out to be slightly faster for the iterates in the middle of the Kaczmarz sweep. The reason for this behaviour is mainly the ordering of the sources, one will of course obtain a different behaviour for different

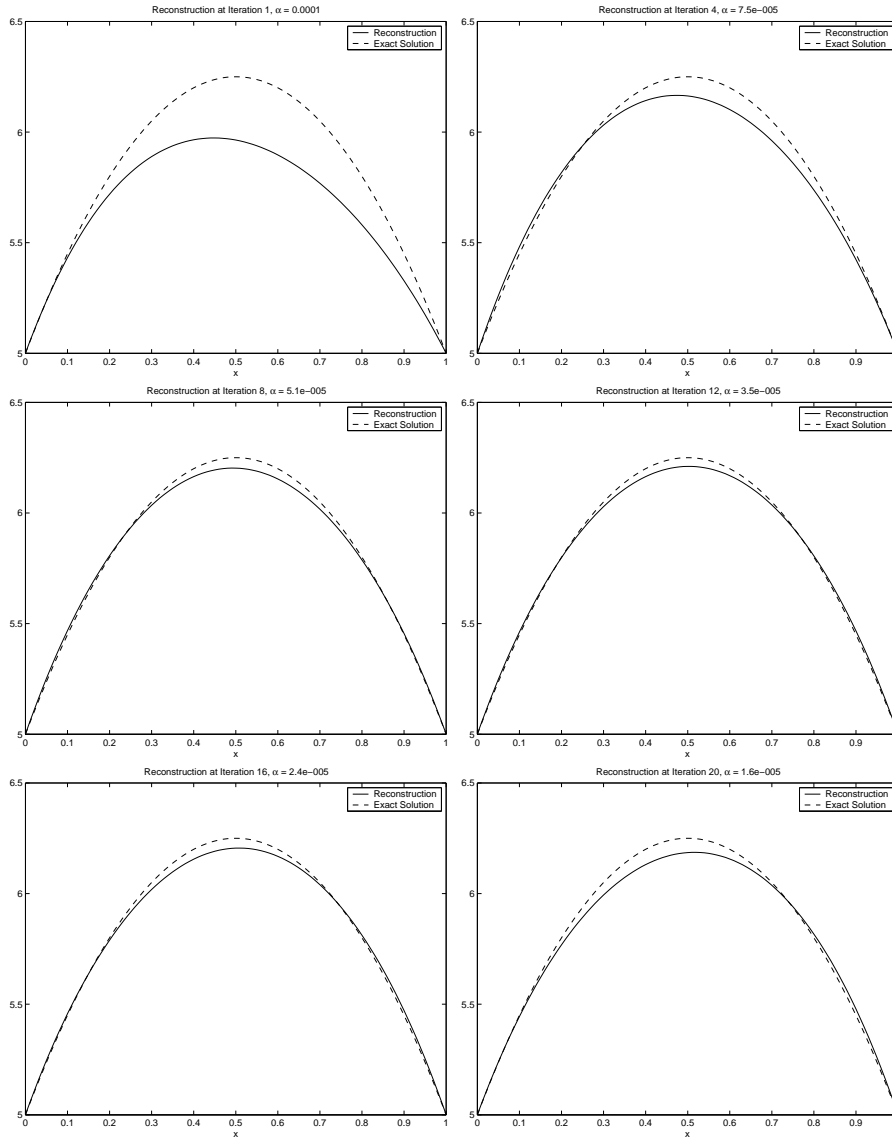


FIG. 5.1. Reconstructions in the first example,  $\delta = 0$ , at iterates 1, 4, 8, 12, 16, and 20

ordering. We finally provide a quantitative basis for the above observations on the behaviour of the iterates in Figure 5.3, where we plot the development of the error  $\|q^* - q_n\|$  (dashed, on the left) and of the residual  $\|F_n(q_n) - y_n\|$  (on the right). In the left plot we also plot the error at the end of each Kaczmarz sweep  $\|q^* - q_{np}\|$  (solid) and in the middle of the Kaczmarz sweep  $\|q^* - q_{np+n/2}\|$  (dotted). In this example it turned out that the total error is not always decreasing, but the error at the same stage of the Kaczmarz sweep  $\|q^* - q_{np+j}\|$  (for  $0 \leq j \leq p-1$ ) is decreasing with  $n$ . In particular, it seems that the error at the beginning and end of the sweep is always the maximum one in the sweep, while the one in the middle of the sweep is always the minimum one. Since all of them decrease towards zero, we obtain the expected worst-

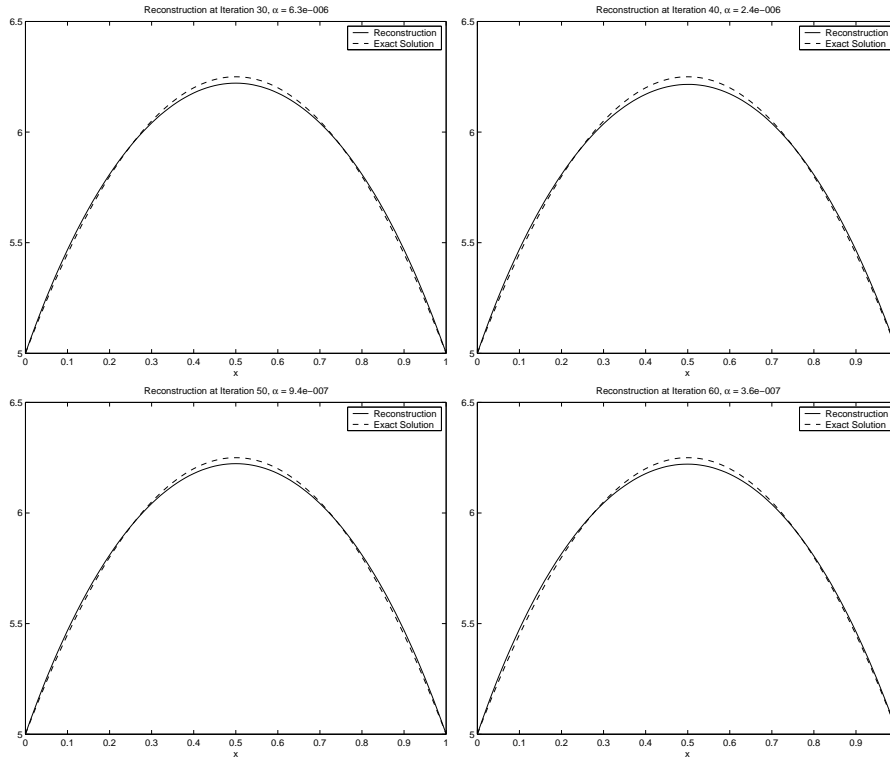


FIG. 5.2. Reconstructions in the first example,  $\delta = 0$ , at iterates 30, 40, 50, and 60

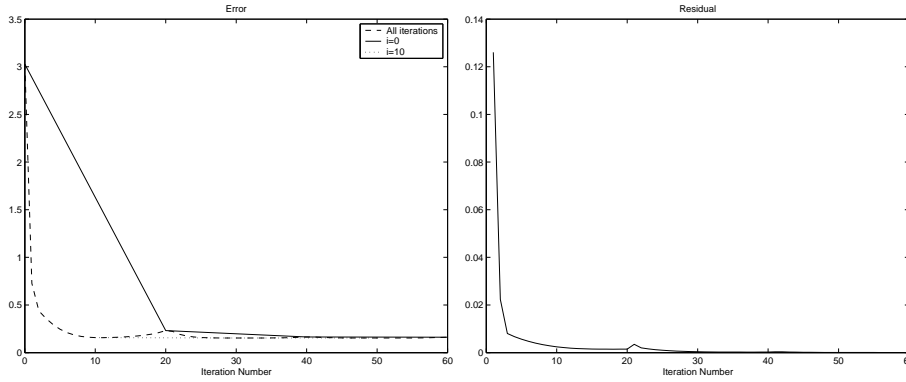


FIG. 5.3. Plot of error (left) and residual (right) vs. iteration number in the first example,  $\delta = 0$

case convergence, but of course in practice one should consider suitable orderings of the data  $y_i$ . The comparison of the residual at different iterates is even more difficult, since the operators and data are different in each step. However, we also obtain that  $\|F_{np+j}(q_{np+j}) - y_{np+j}\|$  (for  $0 \leq j \leq p-1$ ) is decreasing to zero with  $n$ .

For the noisy case we generated data on a finer grid of size  $h = \frac{1}{347}$  in order to avoid inverse crimes. The resulting data are then perturbed using uniform random noise in the interval  $[-\delta, \delta]$ . The regularization parameters are chosen again via (5.1)

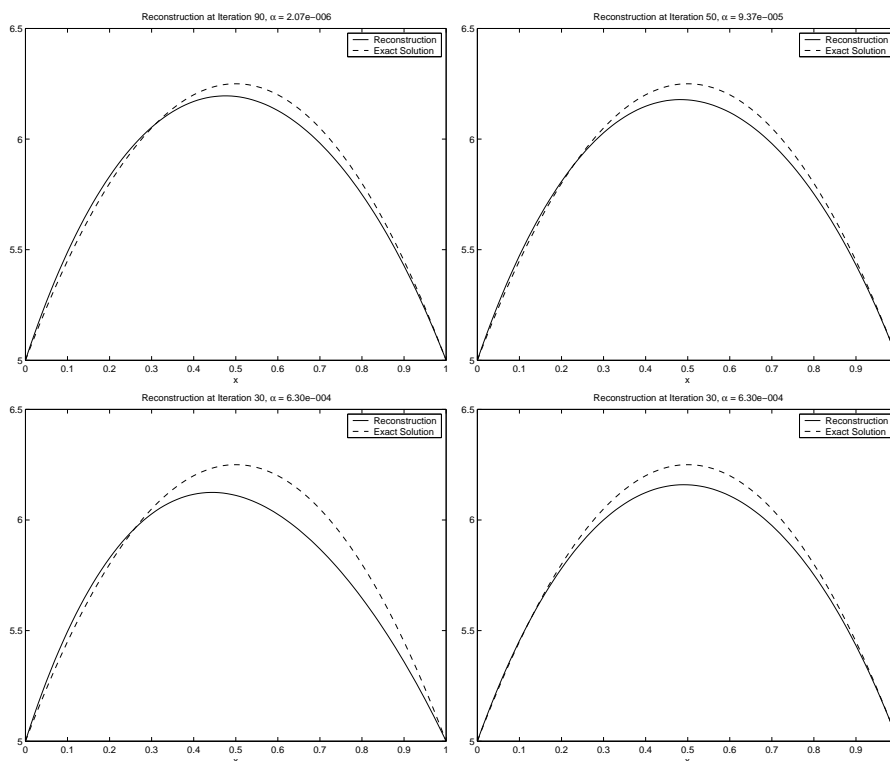


FIG. 5.4. Reconstructions in the first example, for noise levels  $\delta = 0.5\%$  (top left),  $\delta = 1\%$  (top right),  $\delta = 3\%$  (bottom right),  $\delta = 5\%$  (bottom left)

with  $\zeta = 1.1$  and  $\alpha_0 = 10^{-2}$ .

We illustrate the reconstructions obtained for different noise levels (close to the minimum of the error during the iteration) in Figure 5.4. In clockwise order the plots show the reconstruction for noise level  $\delta = 0.5\%$  (at iteration 90),  $\delta = 1\%$  (at iteration 50),  $\delta = 3\%$  (at iteration 30),  $\delta = 5\%$  (at iteration 30). One observes that the quality of the reconstruction improves with decreasing  $\delta$ , i.e., the error of the iterate at the stopping index decreases with  $\delta$ , thus confirming the convergence result for the noisy case. A quantitative monitoring of error and residual vs. the iteration number is presented in Figure 5.5, for  $\delta = 1\%$  (top),  $\delta = 3\%$  (middle), and  $\delta = 5\%$  (bottom). One also sees that the minimal error and residual obtained during the iteration decreases with  $\delta$  as expected. As usual for ill-posed problems the error decreases only until some iteration step and then increases again though the residual is still decreasing. Note that this statement has to be interpreted in a different sense, namely for the subsequences  $np + i$ ,  $0 \leq i \leq p - 1$ . Moreover, the variation in the error and residual during a sweep over the different sources increases with the noise level, which obviously makes the choice of the stopping index more difficult.

**5.2. Reconstruction from Dirichlet-Neumann Data.** Our second numerical experiment is the solution of Example 1, i.e., the reconstruction of the coefficient  $q$  in

$$-\Delta u + qu = 0, \quad \text{in } \Omega \subset \mathbb{R}^d$$

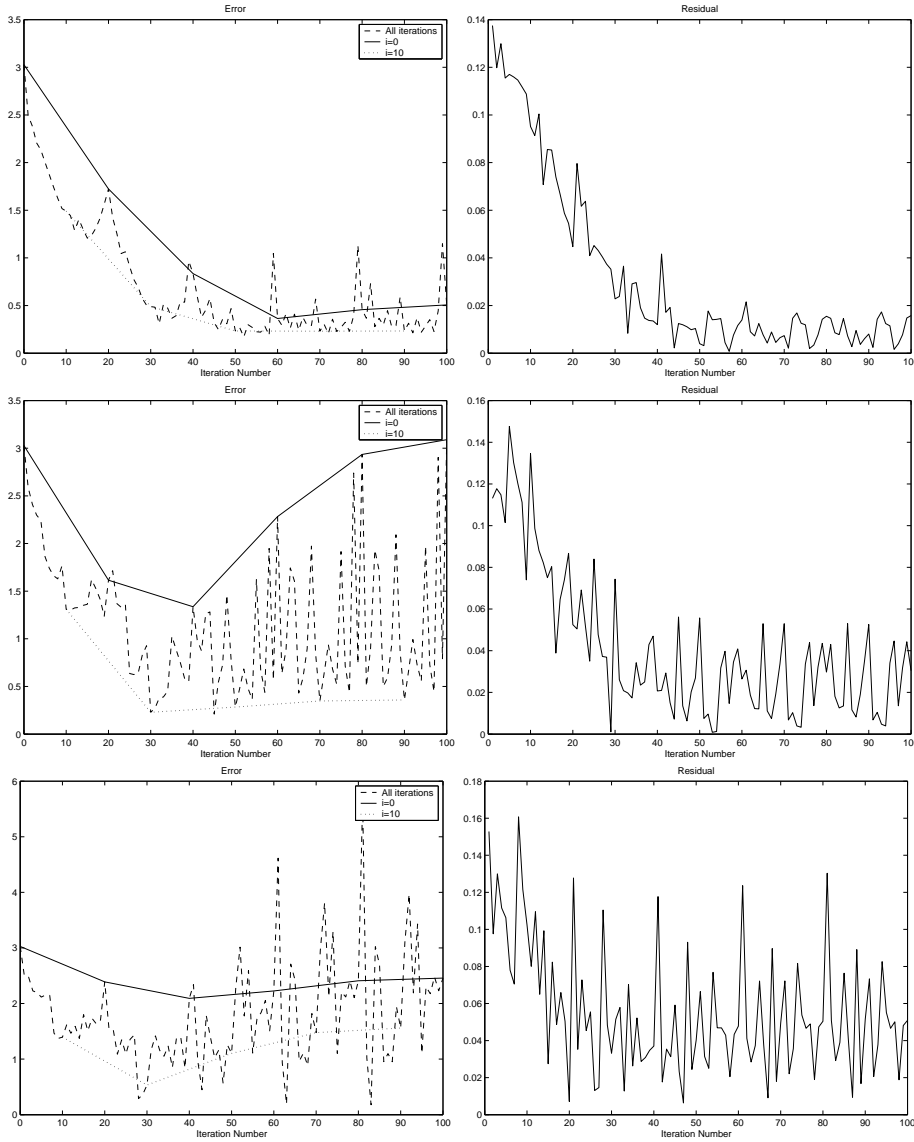


FIG. 5.5. Plot of error (left) and residual (right) vs. iteration number in the first example, for noise levels  $\delta = 1\%$  (top),  $\delta = 3\%$  (middle),  $\delta = 5\%$  (bottom)

from  $p = 20$  values of the Dirichlet-to-Neumann map. In our numerical example, the two-dimensional domain is  $\Omega = (0, 1)^2$ , on which the differential equation is discretized by finite differences on a uniform grid of size  $h = 0.025$ .

The applied Dirichlet sources  $f_j$  are identically zero on three of the boundary segments and of the form

$$f_j(x_1, x_2) = \begin{cases} 10^3 e^{-50((x_1-j)/6)^2} & \text{for } j = 1, \dots, 5, x_2 = 0 \\ 10^3 e^{-50((x_1-(j-5))/6)^2} & \text{for } j = 6, \dots, 10, x_2 = 1 \\ 10^3 e^{-50((x_2-(j-10))/6)^2} & \text{for } j = 11, \dots, 15, x_1 = 0 \\ 10^3 e^{-50((x_2-(j-15))/6)^2} & \text{for } j = 16, \dots, 20, x_1 = 1 \end{cases}$$

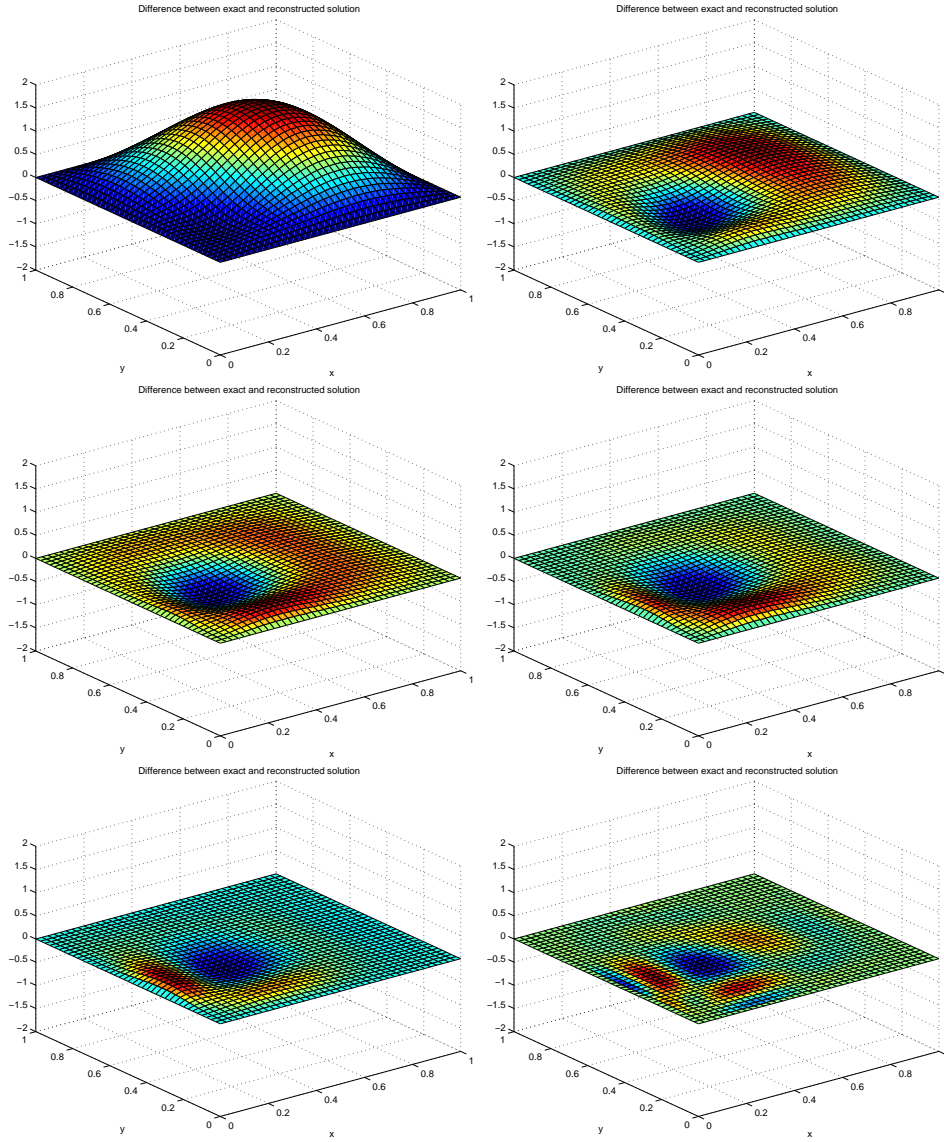


FIG. 5.6. *Difference  $\hat{q} - q_n$  in the second example at iterates 1, 2, 3, 5, 10, and 100.*

on the fourth segment, i.e., they approximate Dirac-delta impulses equally distributed over the boundary.

In this case we use the Levenberg-Marquardt-Kaczmarz method, i.e., a Tikhonov type stabilization in the  $H^1$ -norm in each step with prior  $q_n$ . This means that in each step of the method, the update  $s = q_{n+1} - q_n$  is obtained by solving the minimization problem

$$\frac{1}{2} \left\| \frac{\partial v_n}{\partial \nu} - g_n \right\|_{H^{-1/2}(\partial\Omega)}^2 + \frac{\alpha_n}{2} \|s\|_{H^1(\Omega)}^2$$

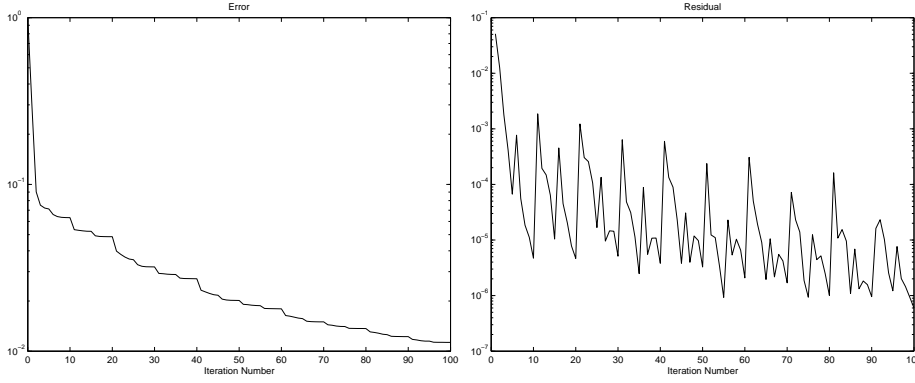


FIG. 5.7. Semi-logarithmic plot of error (left) and residual (right) vs. iteration number in the second example,  $\delta = 0$

subject to the linear equation

$$-\Delta v_n + q_n v_n + s u_n = 0 \quad \text{in } \Omega$$

for  $v_n$  with homogeneous Dirichlet boundary values on  $\partial\Omega$ . The norm in  $H^{-1/2}(\partial\Omega)$  of an element  $g$  is realized by

$$\|g\|_{H^{-1/2}(\partial\Omega)} := \|\phi_g\|_{H^1(\Omega)},$$

where  $\phi_g \in H^1(\Omega)$  is the unique weak solution of

$$\int_{\Omega} (\nabla \phi_g \cdot \nabla \psi + \phi_g \psi) \, dx = \int_{\partial\Omega} g \psi \, d\sigma \quad \forall \psi \in H^1(\Omega).$$

This means we have to solve an additional Neumann problem to evaluate the norm.

We use a primal-dual approach to discretize this problem, which means that we have to find two Lagrange multipliers corresponding to the partial differential equations for  $v_n$  and the function  $\phi_g$  used to evaluate the norm. A careful investigation of the optimality system shows that  $\phi_g$  can be eliminated in favour of one of the Lagrange multipliers, and the optimality system in each steps becomes after straightforward transformations

$$\begin{aligned} -\Delta v_n + q v_n + s u_n &= 0 \\ -\Delta \lambda + q \lambda &= 0 \\ -\Delta \mu + \mu + (1 - q) v_n - s u_n &= 0 \\ -\Delta s + s + \frac{1}{\alpha} u_n \lambda &= -\Delta(q_0 - q_n) + q_0 - q_n \end{aligned}$$

in  $\Omega$ , supplemented by the boundary conditions

$$\begin{aligned} v_n &= 0 \\ \lambda - \mu &= \phi_n - \phi_{g_n} \\ \frac{\partial \mu}{\partial \nu} &= 0 \\ s &= 0 \end{aligned}$$

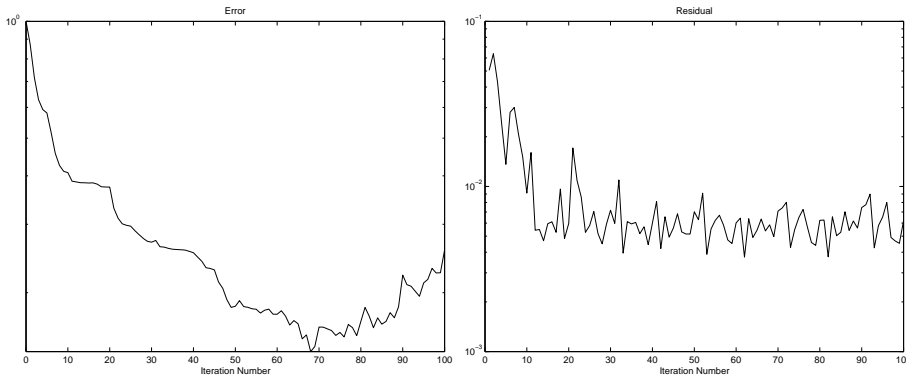


FIG. 5.8. Semi-logarithmic plot of error (left) and residual (right) vs. iteration number in the second example,  $\delta = 1\%$

on  $\partial\Omega$ . The functions  $\phi_n$  and  $\phi_{g_n}$  are the functions used to evaluate the  $H^{-1/2}$ -norm of  $\frac{\partial u_n}{\partial \nu}$  and  $g_n$ , respectively, defined in the same way as  $\phi_g$  above.

We start with some examples using data generated from the parameter

$$\hat{q} = 3 + 5 \sin(\pi x) \sin(\pi y)$$

and the starting value  $q_0 \equiv 3$ . Note that again  $\hat{q}$  is not necessarily the minimum norm solution of the inverse problem with the above measurements, but since we expect that a successful reconstruction algorithm should at least approximate  $\hat{q}$  and since we do not know the minimum norm solution, we measure the error as the difference between  $\hat{q}$  and  $q_n$ . In order to test the convergence of exact data, we generate data on the same grid as the one used for solving the inverse problem and then perform the IRGNK algorithm with  $\alpha_n$  chosen according to (5.1) with  $\zeta = 1.05$  and  $\alpha_0 = 10^{-8}$ .

The difference between  $\hat{q}$  and  $q_n$  is shown in Figure 5.6, at the iterates  $n = 1, 2$  (top),  $n = 3, 5$  (middle), and  $n = 10, 100$  (bottom). One observes that the error is reduced very fast globally, but one also observes a certain local influence of the sources, i.e., the convergence seems faster closer to the support of the boundary sources. The quantitative development of the error  $\|\hat{q} - q_n\|$  (left) and the residual  $\|F(q_n) - g_n\|$  (right) are shown in a semi-logarithmic scale in Figure 5.7.

Moreover, we test the behaviour of the algorithm with respect to noise by using Gaussian random noise of variance  $\delta = 1\%$  and  $\delta = 0.5\%$ . We plot the development of the error (left) and the residual (right) in a semi-logarithmic scale in Figure 5.8 for  $\delta = 1\%$ , and in Figure 5.9 for  $\delta = 0.5\%$ . One observes the expected semi-convergence in both cases, i.e., the error reaches a minimum around which one should stop the iteration, and then starts to increase again. As expected, the minimal error appearing during the iteration decreases with the noise level, one obtains a minimal relative error 0.14 for  $\delta = 1\%$  and 0.11 for  $\delta = 0.5\%$ .

We finally test the behaviour for a more complicated exact parameter value

$$\hat{q} = 3 + 2 \sin(3\pi x_1) \sin(2\pi x_2).$$

In this case we change the initial value  $\alpha_0$  to  $10^{-12}$  due to the lower sensitivity of the data with respect to this parameter. The development of error and residual are shown in semi-logarithmic scale in Figure 5.10. One observes that the method seems to converge in this case, too, although slower than in the above example, which is also caused by the lower sensitivity.

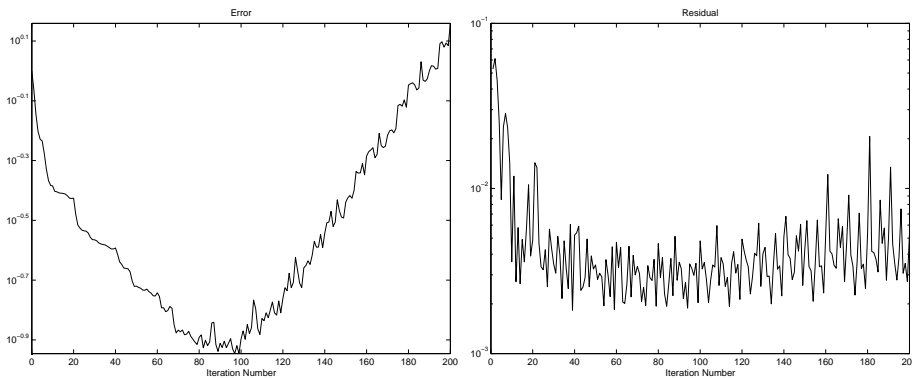


FIG. 5.9. *Semi-logarithmic plot of error (left) and residual (right) vs. iteration number in the second example,  $\delta = 0.5\%$*

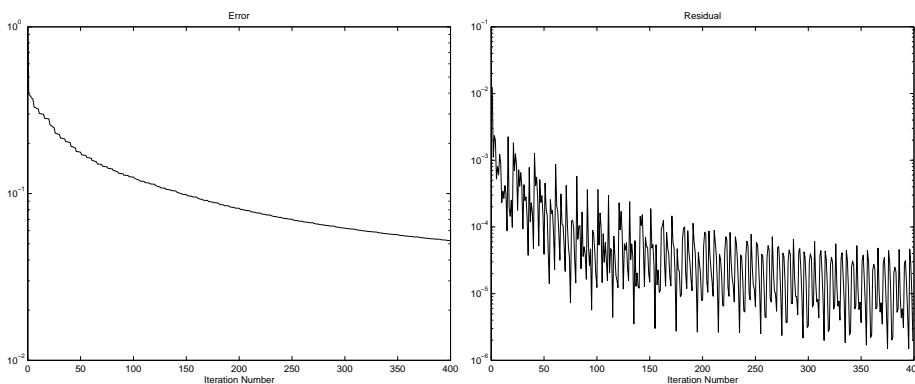


FIG. 5.10. *Semi-logarithmic plot of error (left) and residual (right) vs. iteration number in the second example for different exact solution,  $\delta = 0$*

**6. Conclusions and Open Problems.** We have derived a detailed convergence analysis of regularized Newton-Kaczmarz methods for nonlinear ill-posed problems, which - as usual for ill-posed problems - can be carried out under certain conditions on the nonlinearity of the operators involved. As we have demonstrated in several examples from practice, these conditions seem not to be too restrictive in the case of Newton-Kaczmarz methods. Moreover, we have discussed the numerical solution of the linear problems arising in each step of the iteration method by three different approaches. The numerical experiments we carried out confirm the theoretical predictions.

So far, we have discussed a-priori stopping rules (in the sense of [11]) only, whereas in practice it seems to be more important to have a-posteriori stopping rules, which do not only depend on the noise level  $\delta$ , but also on the actual data  $y^\delta$ . In [24], a stopping rule is proposed for the Landweber Kaczmarz Method that is based on Morozov's discrepancy principle, i.e., a comparison of the residual norm with the noise level. Such an approach would probably be appropriate also in our Newton-Kaczmarz context. However, we expect that a rigorous convergence analysis with such a residual type stopping criterion would have to be based on a nonlinearity assumption similar to (2.6) (as it is done in [24]) rather than the condition (2.2) that

we have verified for our application examples here.

As mentioned in Section 3.2, the condition (3.14) on the initial values poses a severe theoretical restriction that seems to be inevitable for Newton-Kaczmarz Methods of the type (1.10) as the linear case shows. A possible way out might be to define the iteration by (1.7). Here the methods of proof considered so far for  $p = 1$  (cf. [15]) rely on nonlinearity conditions of the type (2.5), but not on (2.4), in whose extension to  $p > 1$ , (2.2) we are interested here, though. Thus, new ideas would be necessary for proving convergence, maybe based on a sweep wise instead of a step wise analysis.

**Acknowledgment.** The authors wish to thank Prof. Andreas Neubauer for valuable hints concerning this paper. Moreover, useful comments by the referees are gratefully acknowledged.

#### REFERENCES

- [1] A.B. BAKUSHINSKII, *The Problem of the convergence of the iteratively regularized Gauss–Newton method*, Comput.Math.Math.Phys. 32 (1992), 1353–1359.
- [2] A.B. BAKUSHINSKII, M.YU. KOKURIN, *Iterative Methods for Approximate Solution of Inverse Problems*, Springer, Dordrecht, 2004.
- [3] S. BIEDENSTEIN, *Numerische Verfahren zur Impedanztomographie*, Diploma Thesis, University Münster, 1997.
- [4] M. BURGER, W. MÜHLHUBER, *Iterative regularization of parameter identification problems by SQP methods*, Inverse Problems 18 (2002), 943–970.
- [5] M. BURGER, W. MÜHLHUBER, *Numerical approximation of an SQP-type method for parameter identification*, SIAM J. Numer. Anal. 40 (2002), 1775 - 1797.
- [6] P. DEUFLHARD, H.W. ENGL, O. SCHERZER, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinity invariant conditions*, Inverse Problems 14 (1998), 1081–1106.
- [7] P. DEUFLHARD, G. HEINDL, *Affine invariant convergence theorems for Newton’s method and extensions to related methods*, SIAM J. Numer. Anal. 16 (1979), 1–10.
- [8] T. DIERKES, *Rekonstruktionsverfahren zur optischen Tomographie*, PhD-Thesis, Universität Münster, 2000.
- [9] O. DORN, H. BERTETE-AGUIRRE, J.G. BERRYMAN, G.C. PAPANICOLAOU, *A nonlinear inversion method for 3D-electromagnetic imaging using adjoint fields*, Inverse Problems 15 (1999), 1523–1558.
- [10] I. EKELAND, R. TEMAM, *Convex Analysis and Variational Problems*, North Holland Publishers, Amsterdam, 1976.
- [11] H.W. ENGL, M. HANKE, A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
- [12] W. FANG, K. ITO, *Reconstruction of semiconductor doping profile from LBIC image*, SIAM J. Appl. Math. 54 (1994), 1067–1082.
- [13] C.W. GROETSCH, *Inverse Problems in Mathematical Sciences*, Vieweg, Braunschweig, 1993.
- [14] M. HANKE, *Regularizing properties of a truncated Newton–CG algorithm for nonlinear inverse problems*, Numer. Funct. Anal. Optim. 18 (1997), 971–993.
- [15] M. HANKE, *A regularization Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems*, Inverse Problems 13 (1997), 79–95.
- [16] M. HANKE-BOURGEOIS, C.W. GROETSCH *Nonstationary iterated Tikhonov regularization*, J. Optim. Th. Appl. 98 (1998), 37–53.
- [17] 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.
- [18] B. HOFMANN, O. SCHERZER, *Influence factors of ill-posedness for nonlinear ill-posed problems*, Inverse Problems 10 (1994), 1277–1297.
- [19] T. HOHAGE, *Iterative Methods in Inverse Obstacle Scattering: Regularization Theory of Linear and Nonlinear Exponentially Ill-Posed Problems*, PhD thesis, University of Linz, 1999.
- [20] T. HOHAGE, *Regularization of exponentially ill-posed problems*, Numer. Funct. Anal. Optim. 21 (2000) 439–464.
- [21] B. BLASCHKE(-KALTENBACHER), A. NEUBAUER, AND O. SCHERZER, *On convergence rates for the iteratively regularized Gauß–Newton method*, IMA J. Numer. Anal. 17 (1997), 421–436.

- [22] B. KALTENBACHER, *Some Newton type methods for the regularization of nonlinear ill-posed problems*, Inverse Problems 13 (1997), 729–753.
- [23] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer, New York, 1996.
- [24] R. KOWAR, O. SCHERZER, *Convergence analysis of a Landweber-Kaczmarz method for solving nonlinear ill-posed problems*, in: S. Romanov, S.I. Kabanikhin, Y.E. Anikonov, A.L. Bukhgein, *Ill-Posed and Inverse Problems*, VSP Publishers, Zeist, 2002.
- [25] A.K. LOUIS, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.
- [26] B. LOWE, W. RUNDELL, *The determination of multiple coefficients in a second order differential equation from input sources*, Inverse Problems 9 (1993), 469-482.
- [27] B. LOWE, W. RUNDELL, *The determination of a coefficient in a parabolic equation from input sources*, IMA J. Appld. Math. 52 (1994), 31-50.
- [28] V.A. MOROZOV, *Regularization Methods for Ill-Posed Problems*, CRC Press, Boca Raton, 1993.
- [29] F. NATTERER, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, 1986.
- [30] F. NATTERER, *Numerical solution of bilinear inverse problems*, Technical Report, Universität Münster, 1996.
- [31] A.G. RAMM, A.B. SMIRNOVA, *A numerical method for solving nonlinear ill-posed problems*, Nonlinear Funct. Anal. and Optimiz. 20 (1999), 317-332.
- [32] A. RIEDER, *On convergence rates of inexact Newton regularizations*, Numer. Math. 88(2001), 347-365.
- [33] A.N. TIKHONOV, V.A. ARSEININ, *Methods for Solving Ill-Posed Problems*, Nauka, Moskau, 1979.
- [34] C.R. VOGEL, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.