Inverse Problems **15** (1999) 309–327. Printed in the UK PII: S0266-5611(99)96240-8

On the regularization of nonlinear ill-posed problems via inexact Newton iterations

Andreas Rieder†
Fachbereich Mathematik, Geb. 38, Universität des Saarlandes, 66041 Saarbrücken, Germany

Received 24 July 1998

Abstract. Inexact Newton methods for the stable solution of nonlinear ill-posed problems are considered. The corresponding inner scheme can be chosen to be any linear regularization with a sufficient modulus of convergence. The regularization property of these Newton-type algorithms is verified, that is, the iterates converge to a solution of the nonlinear problem with exact data when the noise level tends to zero. Moreover, convergence rates are given. Finally, implementation issues are discussed and the algorithm is applied to a parameter identification problem for an elliptic PDE. The numerical results reproduce nicely theoretical predictions and show the efficiency of the proposed method.

1. Introduction

We consider the stable solution of the nonlinear problem

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

where $F : \mathsf{D}(F) \subset X \to Y$ operates between the Hilbert spaces X and Y. Here, $\mathsf{D}(F)$ denotes the domain of definition of F. In (1.1) y^δ is a perturbation of the exact but unknown data $y = F(x^\dagger)$ satisfying

$$\|y - y^\delta\|_Y \leqslant \delta \tag{1.2}$$

with the *a priori* known *noise level* $\delta \geqslant 0$. We call (1.1) *ill posed* if $x^\dagger$, the solution of (1.1) with exact data, does not depend continuously on y. Any algorithm for solving (1.1) has to take care of this instability. Algorithms computing approximations from y^δ to $x^\dagger$ in a stable way are named *regularizations*.

Suppose, throughout the paper, that F is compact and continuous and $\mathsf{D}(F)$ is infinite dimensional. Then (1.1) is ill posed (essentially), see, e.g., [7, proposition 10.1].

The theory of regularization for linear ill-posed problems has reached a certain maturity, if not its final state, see, e.g., [2, 7, 19, 21].

The investigation of regularizations for nonlinear ill-posed problems is still in its infancy though considerable results have already been obtained. Basically, three concepts from the linear theory have been carried over to the nonlinear situation to a certain extent. Those are the Tikhonov–Phillips regularization (see, e.g., [5, 8, 20]), iterative regularizations (see, e.g., [1, 3, 13, 14, 16, 17, 23]) and the approximative inverse approach (see [22]).

Nonlinear ill-posed problems are of growing interest in the applied sciences. For instance, the mathematical modelling of ultrasonic, electrical impedance and microwave tomography

† E-mail address: andreas@num.uni-sb.de

leads to such kinds of problem, see, e.g., the recent proceedings volumes [9, 10] on inverse problems in applications edited by Engl, Louis and Rundell.

The paper is organized as follows. In the next section our algorithm of inexact Newton type is formulated and the mathematical set-up is introduced. An inexact Newton method consists of two components: the outer Newton iteration which updates the current iterate and an inner scheme which provides the update by approximately solving a local linear version of (1.1). As inner scheme we allow any linear regularization method with a sufficient modulus of convergence, for instance, the *Tikhonov–Phillips* regularization, the *truncated singular value decomposition*, the *Landweber* iteration, and the ν*-methods*. The inner scheme is stopped as soon as the relative (linear) residual is less than a given tolerance. We will show in section 3 that this stopping criterion is well defined whenever the tolerance is not too small.

To terminate the outer iteration we rely on a discrepancy principle, that is, we accept the first iterate as an approximation to $x^\dagger$ which yields a (nonlinear) residual having roughly the order of magnitude of the noise level. In section 4 we verify termination of the outer iteration by showing that the (nonlinear) residuals decrease linearly.

Our inexact Newton iteration is a regularization scheme which we will prove in section 5, that is, the iterates converge to a solution of (1.1) as the noise level δ tends to zero. Moreover we determine the rate of convergence which is (almost) optimal under the source condition we use.

The efficiency of our algorithm depends on the choice of the tolerances. We propose a dynamic selection strategy based on our convergence analysis (section 6).

In section 7 we report on numerical experiments with respect to a parameter identification problem for an elliptic PDE. This nonlinear model problem satisfies our theoretical prerequisites. Indeed we are able to reproduce some of our analytical predictions on the performance of the algorithm.

The relation between our inexact Newton method and other iterative techniques to regularize (1.1) will be discussed in the final section.

2. The algorithm and preparatory considerations

One step of the Newton iteration applied to (1.1) consists in solving a linearized version of (1.1). Suppose we have an approximation x_n to $x^\dagger$. Then we get the new approximation $x_{n+1} = x_n + s_n$ where the *Newton correction* s_n is computed as a solution of

$$F'(x_n)\, s = y^\delta - F(x_n) := b_n^\varepsilon. \tag{2.1}$$

Here, $F' : \mathsf{D}(F) \to \mathcal{L}(X, Y)$ is the Fréchet derivative of F which we assume to exist as a continuous mapping.

Unfortunately, (2.1) is a linear ill-posed problem since $F'(v)$ is a compact operator for all $v \in \mathsf{D}(F)$, see, e.g., [24, proposition 7.33]. Furthermore, the *nonlinear defect* b_n^ε is not the exact right-hand side for computing s_n. The *exact Newton update* $s_n^{\mathrm{e}} := x^\dagger - x_n$ is a solution of

$$F'(x_n)\, s_n^{\mathrm{e}} = y - F(x_n) - E(x^\dagger, x_n) =: b_n \tag{2.2}$$

where

$$E(v, w) := F(v) - F(w) - F'(w)\,(v - w)$$

is the remainder term of the first-order Taylor expansion. Hence,

$$\|b_n^\varepsilon - b_n\|_Y \leqslant \delta + \|E(x^\dagger, x_n)\|_Y \qquad \text{and} \qquad b_n \in \mathsf{R}\big(F'(x_n)\big). \tag{2.3}$$

The noise in the right-hand side of (2.1) is twofold. One part comes from the noise in y^δ whereas the other part is introduced by the linearization.

To obtain a useful approximation s_n to s_n^{e} equation (2.1) needs to be regularized. A general regularization scheme applied to (2.1) gives the Newton update

$$s_n = s_{n,r} = g_r\big(A_n^* A_n\big)A_n^* \left(y^\delta - F(x_n)\right)$$

where $A_n := F'(x_n)$ and $g_r : [0, \theta] \to \mathbb{R}$, $\theta = \|A_n\|^2$, is a piecewise continuous function. The parameter $r \geqslant 0$ is called *regularization parameter*.

Example 2.1. Let us look at five examples of regularization schemes.

1. The choice $g_r(t) = 1/(t + 1/r)$ leads to the *Tikhonov–Phillips regularization* where $g_r(A_n^* A_n) = (A_n^* A_n + r^{-1} I)^{-1}$. Here `REGINN` is the variation of the *Levenberg–Marquardt* algorithm investigated by Hanke [13].
2. The *truncated singular value decomposition* is characterized by $g_r(t) = 1/t$, for $t \geqslant 1/r$ and $g_r(t) = 0$, otherwise.
3. If $g_r(t) = \sum_{j=0}^{r-1}(1-t)^j$ and $\|A_n\| \leqslant 1$ then we have the *Landweber* regularization which is an iterative regularization technique.
4. Other iterative regularization schemes are given by the *ν-methods* ($\nu > 0$) due to Brakhage [4], see also Hanke [12]. For scaled A_n, that is, $\|A_n\| \leqslant 1$, the function g_r has the representation $g_r(t) = \big(1 - \widetilde{P}_r^{(\nu)}(t)\big)/t$ where $\widetilde{P}_r^{(\nu)}(t) = P_r^{(2\nu-1/2,-1/2)}(1 - 2t)/P_r^{(2\nu-1/2,-1/2)}(1)$ with $P_r^{(\alpha,\beta)}$ denoting the Jacobi polynomials.
5. The *conjugate gradient* method is a further iterative regularization scheme where g_r is a polynomial of degree $r - 1$. It differs from the first four examples in its nonlinearity, that is, $g_r\big(A_n^* A_n\big)$ is a nonlinear operator.

In all the examples above $g_r(A_n^* A_n)A_n^* y$ provides an approximation to $A_n^\dagger y$ for $y \in \mathsf{D}(A_n^\dagger)$ where $A_n^\dagger$ is the pseudo-inverse of A_n, see, e.g., [7, 21].

Our iterative scheme for solving (1.1) in a stable way now has the form

$$x_{n+1} = x_n + g_{i_n}\big(A_n^* A_n\big)A_n^* \left(y^\delta - F(x_n)\right) \qquad n = 0, 1, 2, \ldots \tag{2.4}$$

with an initial guess $x_0 \in \mathsf{D}(F)$. Here we face two problems.

First, the Newton iteration has to be stopped in time to avoid noise amplification. This will be done by a *discrepancy principle*, that is, we choose an $R > 0$ and accept the iterate x_N as an approximation to $x^\dagger$ for which

$$\|y^\delta - F(x_N)\|_Y \leqslant R\,\delta < \|y^\delta - F(x_k)\|_Y \qquad k = 0, \ldots, N-1 \tag{2.5}$$

holds true.

Second, we have to supply a sequence $\{i_n\}$ of regularization parameters which allow for a good approximation s_{n,i_n} to s_n^{e}. All known *a priori* and *a posteriori* selection strategies for i_n call for a precise knowledge of the noise level $\|b_n^\varepsilon - b_n\|_Y$ in (2.1). In view of (2.3) we realize that this knowledge is not easily at hand even if δ is known. Since we have to rely on computable quantities we determine i_n as the smallest $r \in \mathbb{N}$ such that the relative (linear) residual is smaller than a given tolerance $\mu_n \in]0, 1]$:

$$\|A_n s_{n,r} - b_n^\varepsilon\|_Y / \|b_n^\varepsilon\|_Y < \mu_n. \tag{2.6}$$

Algorithm `REGINN`, see figure 1, realizes our approach and belongs to the class of *inexact Newton iterations*, see, e.g., [18]. The `while` loop implements the outer (Newton) iteration and the `repeat` loop determines the correction step for the outer iteration. In the language of inexact Newton iterations for well-posed problems the tolerances $\{\mu_n\}$ are called *forcing terms*.

```
REGINN(x, R, {μ_n})
n = 0,  x_0 = x
while ‖F(x_n) − y^δ‖_Y > R δ do
{   i_n = 0
    repeat
        i_n = i_n + 1
        s_{n,i_n} = g_{i_n} (F'(x_n)^* F'(x_n)) F'(x_n)^* (y^δ − F(x_n))
    until ‖F'(x_n) s_{n,i_n} + F(x_n) − y^δ‖_Y < μ_n ‖F(x_n) − y^δ‖_Y
    x_{n+1} = x_n + s_{n,i_n}
    n = n + 1
}
x = x_n
```

Figure 1. REGINN: REGularization based on INexact Newton iteration.

There is a difference in quality between the first two and the last three schemes from example 2.1 when it comes to an implementation of REGINN. The first two regularization schemes require an explicit expression of A_n (respectively of a matrix version thereof). If such explicit representations are available at all then only under additional computational effort, see, for instance, the example in section 7. In contrast, the iterative regularizations only need the operator–vector products $A_n v$ and $A_n^* w$ to be implemented.

Primarily we are interested in using iterative regularizations in the inner loop of REGINN. Therefore we assume that F' is scaled such that

$$\|F'(v)\| \leqslant 1 \qquad \text{for all } v \in \mathsf{D}(F). \tag{2.7}$$

3. Termination of the inner loop

In a first step towards an analysis of the algorithm REGINN we verify the termination of the repeat loop provided suitable μ_n's are given.

We recall that the *discrepancy principle* applied to the regularizations $\{g_r\}_{r\in\mathbb{N}}$ from example 2.1 returns a well defined stopping index, see, e.g., [7] or [21]. Thus, if $\tau_n > 1$ then there exists a smallest index $r_\mathrm{s} \in \mathbb{N}$, the stopping index, such that

$$\|A_n\, s_{n,r_\mathrm{s}} - b_n^\varepsilon\|_Y < \tau_n\, \varepsilon \tag{3.1}$$

where $\|b_n^\varepsilon - b_n\|_Y \leqslant \varepsilon$.

In the case when F is a linear mapping, algorithm REGINN coincides with the regularization scheme within its repeat loop. For details see the following lemma. Its straightforward proof is omitted.

Lemma 3.1. *Let $\{g_r\}_{r\in\mathbb{N}}$ satisfy (3.1). Suppose F is a linear operator and let $x_0 \in X$ satisfy $\|F(x_0) - y^\delta\|_Y > R\,\delta$ where $R > 1$ (otherwise accept x_0 as an approximation to $x^\dagger$). Then, for any $\mu_0 = \tau_0\ \delta/\|F(x_0) - y^\delta\|_Y$ with $\tau_0 \in]1, R]$, algorithm* REGINN *stops after the first outer iteration. Moreover,* REGINN *reduces to the regularization method $\{g_r\}_{r\in\mathbb{N}}$ stopped by the discrepancy principle (2.5).*

Next we stipulate the local property (3.2) for the nonlinear function F. Let $Q : X \times X \to \mathcal{L}(Y)$ be a mapping such that

$$F'(v) = Q(v, w)\, F'(w) \qquad \text{and} \qquad \|I - Q(v, w)\| \leqslant C_Q\, \|v - w\|_X \tag{3.2}$$

for all $v, w \in B_\rho(x^\dagger)$, the ball about $x^\dagger$ with radius ρ. We refer to [16] for a discussion of (3.2) and for examples of operators fulfilling (3.2), see also [7]. The Fréchet derivatives of nonlinear operators with property (3.2) have a null space which is invariant in $B_\rho(x^\dagger)$, that is, $\mathsf{N}\big(F'(v)\big) = \mathsf{N}\big(F'(w)\big)$ for all $v, w \in B_\rho(x^\dagger)$.

Hypothesis (3.2) yields the Lipschitz-like estimate (3.3) in $B_\rho(x^\dagger)$,

$$\|\big(F'(v) - F'(w)\big)\,(v - w)\|_Y \leqslant C_Q\,\|v - w\|_X\,\|F'(w)\,(v - w)\|_Y \tag{3.3}$$

which, in turn, implies that

$$\begin{aligned}\|E(v, w)\|_Y &\leqslant \int_0^1 \big\|\,\big[F'\big(w + t\,(v - w)\big) - F'(w)\big]\,(v - w)\,\big\|_Y\,\mathrm{d}t\\ &\leqslant \frac{C_Q}{2}\;\|v - w\|_X\,\|F'(w)\,(v - w)\|_Y\end{aligned}$$

in $B_\rho(x^\dagger)$. Let $C_Q\,\rho < 1$. Then, the latter displayed inequality in combination with the inverse triangle inequality gives

$$\|F(v) - F(w)\|_Y \geqslant (1 - C_Q\,\rho)\;\|F'(w)\,(v - w)\|_Y \tag{3.4}$$

so that finally

$$\|E(v, w)\|_Y \leqslant \omega\;\|F(v) - F(w)\|_Y \qquad \text{for all } v, w \in B_\rho(x^\dagger) \tag{3.5}$$

where $\omega := C_Q\,\rho/(1 - C_Q\,\rho)$. Note that $\omega < 1$ for $C_Q\,\rho < 1/2$.

Employing (3.5) we are able to estimate the data error $\|b_n^\varepsilon - b_n\|_Y$ in terms of δ, ω and the nonlinear defect

$$d_n := \|y^\delta - F(x_n)\|_Y = \|b_n^\varepsilon\|_Y.$$

We have, for $x_n \in B_\rho(x^\dagger)$,

$$\|E(x^\dagger, x_n)\|_Y \leqslant \omega\,\|y - F(x_n)\|_Y \leqslant \omega\,\big(\|y - y^\delta\|_Y + d_n\big) \leqslant \omega\,\delta + \omega\,d_n.$$

Thus

$$\|b_n^\varepsilon - b_n\|_Y \leqslant (1 + \omega)\,\delta + \omega\,d_n := \varepsilon = \varepsilon(x_n, \delta).$$

We are finally in a position to derive *sufficient* conditions on μ_n to stop the `repeat` loop.

Lemma 3.2. *Let $\{g_r\}_{r\in\mathbb{N}}$ satisfy (3.1) and let (3.2) hold true with $C_Q\,\rho < 1/2$. Further assume that $x_n \in B_\rho(x^\dagger)$. If $R \geqslant (1+\omega)/(1-\omega)$ then the* `repeat` *loop of algorithm* `REGINN` *terminates for any*

$$\mu_n \in \Big]\omega + \frac{(1 + \omega)\,\delta}{d_n},\; 1\Big].$$

Proof. We will show that the stopping criterion of algorithm `REGINN` can be rewritten as the discrepancy principle (3.1) with a $\tau_n > 1$. This guarantees termination. We have that

$$\tau_n := \frac{\mu_n\,d_n}{\varepsilon(x_n, \delta)} = \frac{\mu_n}{(1 + \omega)\,\delta/d_n + \omega} > 1. \tag{3.6}$$

Since $\delta < d_n/R$ (otherwise the outer iteration would have been stopped with x_n) our hypothesis on R gives $(1 + \omega)\,\delta/d_n + \omega < (1 + \omega)/R + \omega \leqslant 1$. □

Throughout the paper and without further notification let R and μ_n be chosen such that (3.6) holds true for $x_n \in B_\rho(x^\dagger)$. Also, let $\{g_r\}_{r\in\mathbb{N}}$ satisfy (3.1) always.

4. Termination of the outer iteration

We will show that the nonlinear residuals decrease linearly.

Lemma 4.1. *Suppose that the kth iterate x_k of algorithm* `REGINN` *is well defined and lies in $B_\rho(x^\dagger)$. Further, let (3.5) hold with*

$$\omega < \eta/(2+\eta) \qquad \text{for one} \qquad \eta < 1 \tag{4.1}$$

(this will be true, for instance, if ρ is sufficiently small). If, furthermore,

$$R \geqslant \frac{1+\omega}{\eta-(2+\eta)\,\omega} \qquad \text{and} \qquad \mu_k \in \left]\omega + \frac{(1+\omega)\,\delta}{d_k},\ \eta-(1+\eta)\,\omega\right] \tag{4.2}$$

as well as $x_{k+1} \in B_\rho(x^\dagger)$ then

$$\frac{\|y^\delta - F(x_{k+1})\|_Y}{\|y^\delta - F(x_k)\|_Y} \leqslant \frac{\mu_k+\omega}{1-\omega} \leqslant \eta. \tag{4.3}$$

Proof. Before we start proving (4.3) we discuss the assumptions on ω, R and μ_k. The bound on ω implies that the denominator of the lower bound of R is positive. The lower bound on R guarantees that $\omega + (1+\omega)/R$ is smaller then $\eta - (1+\eta)\,\omega$ which is the upper bound for μ_k yielding $(\mu_k+\omega)/(1-\omega) \leqslant \eta$. All parameters satisfy the requirements of lemma 3.2 so that s_{k,i_k} is well defined.

Since $F(x_{k+1}) - y^\delta = A_k s_{k,i_k} + F(x_k) - y^\delta + E(x_{k+1}, x_k)$ we obtain

$$\begin{aligned}\|F(x_{k+1}) - y^\delta\|_Y &\leqslant \mu_k\,\|F(x_k)-y^\delta\|_Y + \omega\,\|F(x_{k+1})-F(x_k)\|_Y \\ &\leqslant \mu_k\,\|F(x_k)-y^\delta\|_Y + \omega\left(\|F(x_{k+1})-y^\delta\|_Y + \|F(x_k)-y^\delta\|_Y\right)\end{aligned}$$

which readily implies (4.3). □

The key estimate in the proof of the lemma from above can be traced back to Hanke [13, equation (2.10)].

In the setting of lemma 4.1 we have that the residuals decrease η-linearly uniformly in $\delta \in [0, \overline{\delta}]$ for a $\overline{\delta} > 0$ small enough. This implies termination of `REGINN`.

By $\lfloor t \rfloor \in \mathbb{Z}$ for $t \in \mathbb{R}$ we denote the greatest integer: $\lfloor t \rfloor \leqslant t < \lfloor t \rfloor + 1$.

Theorem 4.2. *Adopt the assumptions (4.1) and (4.2) on ω, R, and the μ_k's from lemma 4.1. Suppose further that all iterates $\{x_k\}_k$ stay in $B_\rho(x^\dagger)$.*

If $d_0 = \|y^\delta - F(x_0)\|_Y > R\,\delta$ then `REGINN` *terminates after*

$$N(\delta) \leqslant \lfloor \log_\eta(R\,\delta/d_0) \rfloor + 1 \tag{4.4}$$

outer iteration steps for all $0 < \delta \leqslant \overline{\delta}$. Moreover,

$$\left\|A(x^\dagger - x_{N(\delta)})\right\|_Y \leqslant \frac{R+1}{1-C_Q\,\rho}\ \delta \qquad \text{as } \delta \to 0 \tag{4.5}$$

where $A := F'(x^\dagger)$.

Proof. The bound (4.4) for $N(\delta)$ follows directly from lemma 4.1 since

$$\|F(x_k) - y^\delta\|_Y \leqslant \eta^k\,\|F(x_0) - y^\delta\|_Y. \tag{4.6}$$

We infer from (3.4) and (1.2) that

$$\left\|A(x^\dagger - x_k)\right\|_Y \leqslant \frac{\|y - F(x_k)\|_Y}{1-C_Q\,\rho} \leqslant \frac{1}{1-C_Q\,\rho}\left(\delta + \|y^\delta - F(x_k)\|_Y\right)$$

for $k \in \{0, \ldots N(\delta)\}$. Especially, $k = N(\delta)$ together with (2.5) yields (4.5). □

Let us discuss (4.5). Observe that $\|A\cdot\|_Y$ is a norm on $\mathsf{N}(A)^\perp$ being, in general, *weaker* than the standard norm on X. If we start `REGINN` with $x_0 \in \mathsf{N}(A)^\perp$ then all iterates will stay in $\mathsf{N}(A)^\perp = \overline{\mathsf{R}(A^*)}$ due to (2.4) and (3.2). In the case of $x^\dagger \in \mathsf{N}(A)^\perp$, the estimate (4.5) describes norm convergence. This is a result which carries over from the linear to the nonlinear situation.

For our further analysis of `REGINN` we restrict ourselves to *linear* regularization schemes $\{g_r\}_{r\in\mathbb{N}_0}$, $g_0 := 0$, satisfying the assumptions (4.7) below with $p_r(t) := 1 - t\, g_r(t)$. Let us assume the existence of positive constants C_g, C_p, and α such that

$$\sup_{t\in[0,\theta]} |g_r(t)| \leqslant C_g\, r^\alpha \qquad \sup_{t\in[0,\theta]} |p_r(t)| = 1 \qquad \sup_{t\in[0,\theta]} |t\, p_r(t)| \leqslant C_p\, r^{-\alpha}. \tag{4.7}$$

From now on the conjugate gradient method will not be considered anymore.

Example 4.3. The first four regularization schemes from example 2.1 satisfy (4.7).

1. Tikhonov–Phillips regularization: $C_g = C_p = \alpha = 1$.
2. Truncated singular value decomposition: $C_g = C_p = \alpha = 1$.
3. Landweber regularization: $C_g = \alpha = 1$ and $C_p = \exp(-1)$.
4. ν-methods ($\nu \geqslant 1$): $\alpha = 2$, sharp estimates for C_g and C_p are difficult to obtain.

Next we supply a norm estimate of $s_{k,i_k} = g_{i_k}(A_k^* A_k)\, A_k^* b_k^\varepsilon$ where $A_k = F'(x_k)$. The left and middle relations in (4.7) as well as standard arguments, see, e.g., [7] and [21], lead to the norm bound (4.8) for the operator $R_{i_k} := g_{i_k}(A_k^* A_k)\, A_k^*$,

$$\|R_{i_k}\| \leqslant C_R\, i_k^{\alpha/2} \qquad C_R := \sqrt{2\, C_g}. \tag{4.8}$$

So we get

$$\|s_{k,i_k}\|_X \leqslant C_R\, i_k^{\alpha/2}\, \|y^\delta - F(x_k)\|_Y. \tag{4.9}$$

In the following we bound the stopping index i_k. According to the definition of i_k, cf (3.1) and (3.6), we have, for $i_k \geqslant 2$,

$$\begin{aligned}\tau_k\, \varepsilon(x_k,\delta) &\leqslant \|A_k\, s_{k,i_k-1} - b_k^\varepsilon\|_Y \\ &\leqslant \|p_{i_k-1}(A_k\, A_k^*)\, b_k\|_Y + \|p_{i_k-1}(A_k\, A_k^*)\, (b_k^\varepsilon - b_k)\|_Y \\ &\leqslant \|p_{i_k-1}(A_k\, A_k^*)\, A_k s_k^{\mathrm{e}}\|_Y + \varepsilon(x_k,\delta).\end{aligned}$$

In the last step we used relation (2.2) for the exact Newton update $s_k^{\mathrm{e}} = x^\dagger - x_k$ and we used the standardization of the p_r's in (4.7). To proceed we assume there exists a $w_k \in Y$ so that $s_k^{\mathrm{e}} = A_k^* w_k$ (we will comment on this assumption in lemma 4.5 below). Hence, by the right relation in (4.7),

$$\varepsilon(x_k,\delta) \leqslant C_p\, \frac{\|w_k\|_Y}{\tau_k - 1}\, (i_k - 1)^{-\alpha}. \tag{4.10}$$

Altogether we are able to verify the following lemma.

Lemma 4.4. *Let $\{g_r\}_{r\in\mathbb{N}_0}$ fulfil (4.7) and assume that x_k is well defined. Further, let there be a $w_k \in Y$ so that $s_k^{\mathrm{e}} = F'(x_k)^*\, w_k$. Then, there exists a positive constant $C_I \leqslant \max\{1, 2^\alpha\, C_p\}$ such that*

$$i_k \leqslant \left(C_I\, \frac{\|w_k\|_Y}{\tau_k - 1}\right)^{1/\alpha}\, \varepsilon(x_k,\delta)^{-1/\alpha}. \tag{4.11}$$

Proof. First, we consider the case $i_k \geqslant 2$. From (4.10) we obtain $(i_k-1)^\alpha \leqslant C_p\, \|w_k\|_Y/(\tau_k - 1)/\varepsilon(x_k,\delta)$. Since $i_k^\alpha \leqslant 2^\alpha\, (i_k-1)^\alpha$ the inequality (4.11) is established with $C_I = 2^\alpha\, C_p$. In the case of $i_k = 1$ the trivial estimate $\tau_k\, \varepsilon(x_k,\delta) = \mu_k\, \|b_k^\varepsilon\|_Y \leqslant \|b_k^\varepsilon\|_Y \leqslant \|b_k\|_Y + \varepsilon(x_k,\delta)$ and (2.7) readily imply (4.11) with $C_I = 1$. □

Lemma 4.5. *Suppose that the first n iterates $\{x_1, \dots, x_n\}$ of algorithm* `REGINN` *are well defined and stay in $B_\rho(x^\dagger)$. Moreover, let the initial guess $x_0 \in B_\rho(x^\dagger)$ be such that*

$$s_0^{\mathrm{e}} = F'(x_0)^* w_0 \qquad \text{for one } w_0 \in Y. \tag{4.12}$$

Then,

$$s_k^{\mathrm{e}} = A_k^* w_k \qquad \text{with } w_k = Q(x_0, x_k)^* w_0 - \sum_{j=0}^{k-1} Q(x_j, x_k)^* g_{i_j}\big(A_j A_j^*\big) b_j^\varepsilon \tag{4.13}$$

for $k = 1, \dots, n$, where $A_j = F'(x_j)$ and Q is the mapping from (3.2). Furthermore, if (4.7) applies then

$$\|w_k\|_Y \leqslant \widetilde{C}_Q\ (1+\lambda)\ (1+\widetilde{C}_Q\,\lambda)^{k-1}\ \|w_0\|_Y \qquad k = 1, \dots, n \tag{4.14}$$

where

$$\lambda = \frac{C_g\, C_I}{\tau - 1} + \widetilde{C}_g \qquad \text{with } \tau = \min\{\tau_0, \dots, \tau_{n-1}\} \text{ and } \widetilde{C}_g = \sup_{r\in\mathbb{N}} \sup_{t\in[0,\theta]} t\,|g_r(t)| \leqslant 2.$$

The constant $\widetilde{C}_Q$ is an upper bound of Q: $\|Q(v,z)\| \leqslant \widetilde{C}_Q$ for all $v, z \in B_\rho(x^\dagger)$.

Proof. Rewrite $s_k^{\mathrm{e}} = s_0^{\mathrm{e}} - \sum_{j=0}^{k-1} s_{j,i_j}$ and note that $s_0^{\mathrm{e}} = A_0^* w_0 = A_k^* Q(x_0,x_k)^* w_0$ as well as $s_{j,i_j} = A_j^* g_{i_j}(A_j A_j^*) b_j^\varepsilon = A_k^* Q(x_j,x_k)^* g_{i_j}(A_j A_j^*) b_j^\varepsilon$. The first assertion follows readily.

From the relation on the right of (4.13) together with (4.7) we obtain that

$$\begin{aligned}
\|w_k\|_Y &\leqslant \widetilde{C}_Q \left(\|w_0\|_Y + \sum_{j=0}^{k-1} \left(\|g_{i_j}\big(A_j A_j^*\big)(b_j^\varepsilon - b_j)\|_Y + \|g_{i_j}\big(A_j A_j^*\big) b_j\|_Y\right)\right) \\
&\leqslant \widetilde{C}_Q \left(\|w_0\|_Y + \sum_{j=0}^{k-1} \left(C_g\, i_j^\alpha\, \varepsilon(x_j,\delta) + \|g_{i_j}\big(A_j A_j^*\big) A_j A_j^* w_j\|_Y\right)\right) \\
&\leqslant \widetilde{C}_Q \left(\|w_0\|_Y + \left(\frac{C_g\, C_I}{\tau-1} + \widetilde{C}_g\right) \sum_{j=0}^{k-1} \|w_j\|_Y\right)
\end{aligned}$$

where we used (4.11) for the last inequality. The second assertion of lemma 4.5 follows now inductively. □

For convenience we simplify (4.14) to

$$\|w_k\|_Y \leqslant C_W\ \Lambda^k\ \|w_0\|_Y \qquad \text{with } \Lambda := 1 + \widetilde{C}_Q\,\lambda. \tag{4.15}$$

By (3.2) condition (4.12) can be rewritten as

$$x^\dagger - x_0 \in \mathsf{R}\big(F'(x^\dagger)^*\big) = \mathsf{R}\Big(\big(F'(x^\dagger)^* F'(x^\dagger)\big)^{1/2}\Big) \tag{4.16}$$

and is called *source representation*, see, e.g., [7]. It is an abstract smoothness assumption on $x^\dagger - x_0$.

Under certain assumptions the linear decrease of the nonlinear residuals carries over to the Newton steps.

Lemma 4.6. *Let (3.2) and (4.7) hold true. Let (4.12) apply for the initial guess $x_0 \in B_\rho(x^\dagger)$ and assume that the first n iterates $\{x_1, \dots, x_n\}$ stay in $B_\rho(x^\dagger)$.*

Let $\tau > 1$. Further, let (3.5) hold true for

$$\omega < \frac{\eta}{\eta + (1+\tau)}$$

where $\eta := \sigma/\Lambda$ *with* $\sigma < 1$ *and* Λ *from (4.15). Finally, choose*

$$R \geqslant \frac{\tau\,(1+\omega)}{\eta - \omega\big(\eta + (1+\tau)\big)} \qquad \text{and} \qquad \mu_k \in \Big[\,\tau\,\Big(\omega + \frac{(1+\omega)\,\delta}{d_k}\Big),\ \eta - (1+\eta)\,\omega\,\Big]$$

for $k = 0, \ldots, n$. *Then,*

$$\|s_{k,i_k}\|_X \leqslant C_S\,\|w_0\|_Y^{1/2}\,\|y^\delta - F(x_0)\|_Y^{1/2}\,\sigma^{k/2} \tag{4.17}$$

where $C_S = C_R\,\sqrt{C_I\,C_W\,\tau/(\tau-1)/m}$ *with* $m = \min\{\mu_0, \ldots, \mu_n\} \geqslant \tau\,\big(\omega + \delta\,(1+\omega)/d_0\big)$.

Proof. For a discussion of the restrictions on ω and R see the opening of the proof of lemma 4.1. The lower bound on the μ_k's yields $\tau_k \geqslant \tau > 1$, $k = 0, \ldots, n$, cf (3.6).

Recalling (4.9), (4.11) and (4.15) we find that

$$\|s_{k,i_k}\|_X \leqslant C_R\,\sqrt{\frac{C_I\,C_W\,\|w_0\|_Y}{\tau_k - 1}}\;\Lambda^{k/2}\,\varepsilon(x_k,\delta)^{-1/2}\,\|y^\delta - F(x_k)\|_Y.$$

The proof of (4.17) is established when applying (3.6) and (4.6) to the right hand side of the above inequality. □

Using (4.17) we easily see that the Newton iterates stay in the ball with radius

$$a = a(\delta) := C_S\,\frac{\|w_0\|_Y^{1/2}\,\|y^\delta - F(x_0)\|_Y^{1/2}}{1 - \sqrt{\sigma}}$$

about the centre x_0. Hence, we may abandon the condition that the iterates stay in $B_\rho(x^\dagger)$ (see theorem 4.2 and lemmata 4.5 and 4.6) by the following assumption: $B_{a(\delta)}(x_0) \subset B_\rho(x^\dagger)$. This may be interpreted as a closeness assumption on x_0 which is typical for Newton-type methods where we can expect local convergence only.

Our next result shows that the reduction rate d_{k+1}/d_k for the nonlinear residuals approximates the tolerance μ_k as the iteration progresses.

Corollary 4.7. *Adopt the assumptions of lemma 4.6. Further, choose* x_0 *such that* $B_{a(\delta)}(x_0) \subset B_\rho(x^\dagger)$. *Then, for* $k = 0, \ldots, N(\delta) - 1$,

$$\frac{\|y^\delta - F(x_{k+1})\|_Y}{\|y^\delta - F(x_k)\|_Y} \leqslant \min\Big\{\frac{\mu_k + \omega}{1-\omega},\ \mu_k + \widetilde{C}_S\,\sigma^{k/2}\Big\} \tag{4.18}$$

where $\widetilde{C}_S = C_Q\,C_S\,\|w_0\|_Y^{1/2}\,\|y^\delta - F(x_0)\|_Y^{1/2}$.

Proof. Due to (4.3) it suffices to verify that $d_{k+1}/d_k \leqslant \mu_k + \widetilde{C}_S\,\sigma^{k/2}$. Let $r_k := F'(x_k)\,s_{k,i_k} + F(x_k) - y^\delta$. We have

$$\begin{aligned} F(x_{k+1}) - y^\delta &= F(x_k + s_{k,i_k}) - F(x_k) + F(x_k) - y^\delta \\ &= \int_0^1 \big(F'(x_k + t\,s_{k,i_k}) - F'(x_k)\big)\,s_{k,i_k}\,\mathrm{d}t + r_k. \end{aligned}$$

We apply (3.3) which yields

$$\|y^\delta - F(x_{k+1})\|_Y \leqslant \frac{C_Q}{2}\,\|s_{k,i_k}\|_X\,\|F'(x_k)\,s_{k,i_k}\|_Y + \|r_k\|_Y.$$

Note that $\|r_k\|_Y < \|y^\delta - F(x_k)\|_Y$. Thus,

$$\|F'(x_k)\,s_{k,i_k}\|_Y \leqslant \|r_k\|_Y + \|y^\delta - F(x_k)\|_Y < 2\,\|y^\delta - F(x_k)\|_Y.$$

From both latter inequalities we deduce that

$$\|y^\delta - F(x_{k+1})\|_Y < \big(C_Q\,\|s_{k,i_k}\|_X + \mu_k\big)\,\|y^\delta - F(x_k)\|_Y.$$

Now, the assertion follows from lemma 4.6. □

Our above results yield convergence of the Newton iterates to $x^\dagger$ in the noise-free situation.

Theorem 4.8. *Adopt the assumptions of lemma 4.6 but let $\delta = 0$, that is, the right-hand side of (1.1) is known exactly. If $B_{a(0)}(x_0) \subset B_\rho(x^\dagger)$ then*

$$\|x^\dagger - x_k\| = \mathcal{O}\big(\sigma^{k/2}\big) \qquad \text{as } k \to \infty.$$

Proof. We infer from (4.13) and (3.4) that

$$\begin{aligned}\|x^\dagger - x_k\|_X^2 &= \langle s_k^{\mathrm{e}},\ F'(x_k)^* w_k\rangle_X \leqslant \|F'(x_k)\, s_k^{\mathrm{e}}\|_Y\ \|w_k\|_Y \\ &\leqslant \frac{\|w_k\|_Y}{1 - C_Q\,\rho}\ \ \|y - F(x_k)\|_Y\end{aligned} \tag{4.19}$$

which implies the assertion by (4.6) and (4.15). □

5. Regularization property

We will prove the regularization property of the algorithm REGINN, that is, the convergence of $x_{N(\delta)}$ to $x^\dagger$ as $\delta \to 0$ where $N(\delta)$ is the finite stopping index of the outer iteration according to theorem 4.2.

Theorem 5.1. *Let the assumptions of lemma 4.6 hold true and let $B_{a(0)}(x_0) \subset \mathrm{int}\big(B_\rho(x^\dagger)\big)$. If $d_0 = \|y^\delta - F(x_0)\|_Y > R\,\delta$ (for instance, $F(x_0) \neq y$ and δ sufficiently small) then*

$$\|x^\dagger - x_{N(\delta)}\|_X \ \leqslant\ \sqrt{\frac{C_W\,(R+1)\,\|w_0\|_Y\,\Lambda}{1 - C_Q\,\rho}\ \left(\frac{d_0}{R}\right)^{\log_{1/\eta}\Lambda}}\ \delta^{(1-\log_{1/\eta}\Lambda)/2} \tag{5.1}$$

as $\delta \to 0$ where $0 \leqslant \log_{1/\eta}\Lambda < 1$.

Proof. Note that the elements of the Newton sequence produced by REGINN depend on δ, that is, $x_k = x_k^\delta$, $k = 1, \ldots, N(\delta)$. Since $B_{a(0)}(x_0)$ lies in the interior of $B_\rho(x^\dagger)$ there exists a $\overline{\delta} > 0$ such that $B_{a(\delta)}(x_0) \subset B_\rho(x^\dagger)$ for all $0 < \delta \leqslant \overline{\delta}$. The estimates (4.19) and (1.2) together with (2.5) and (4.15) imply

$$\begin{aligned}\|x^\dagger - x_{N(\delta)}\|_X^2 &\leqslant \frac{\|w_{N(\delta)}\|_Y}{1 - C_Q\,\rho}\ \ \big(\delta + \|y^\delta - F(x_{N(\delta)})\|_Y\big) \\ &\leqslant \frac{C_W\,(R+1)\,\|w_0\|_Y}{1 - C_Q\,\rho}\ \ \Lambda^{N(\delta)}\ \delta.\end{aligned}$$

Since $N(\delta) \leqslant \log_\eta(R\,\delta/d_0) + 1$, see (4.4), we obtain that $\Lambda^{N(\delta)} \leqslant \Lambda\,\Lambda^{\log_\eta(R\,\delta/d_0)} = \Lambda\,(R\,\delta/d_0)^{\log_\eta\Lambda}$. Further, $\log_\eta\Lambda = -\log_{1/\eta}\Lambda$ which verifies (5.1). Finally, $1 \leqslant \Lambda < 1/\eta$, see lemma 4.6, is equivalent to $0 \leqslant \log_{1/\eta}\Lambda < 1$. □

Suppose F is a linear mapping, see lemma 3.1. Then $N(\delta) = 1$ for all $\delta > 0$ and $C_Q\,\rho$ may be considered zero. Further, $\|w_1\|_Y \leqslant (1+\lambda)\,\|w_0\|_Y$, see (4.14). The technique of proof of theorem 5.1 gives now the error bound

$$\|x^\dagger - x_1\|_X \leqslant \sqrt{(R+1)\,(1+\lambda)}\ \|w_0\|_Y^{1/2}\ \delta^{1/2}.$$

The latter error bound reflects exactly the order optimality of the linear regularization schemes $\{g_r\}_{r\in\mathbb{N}_0}$ (4.7) applied to a linear problem under the source conditions (4.12) and (4.16), respectively, see, e.g., [7, 21].

In the nonlinear case, if $\|w_{N(\delta)}\|_Y$ is uniformly bounded in δ, that is $\Lambda = 1$, the iterates $\{x_{N(\delta)}\}_{\delta>0}$ converge with order 1/2: $\|x^\dagger - x_{N(\delta)}\|_X = \mathcal{O}\big(\sqrt{\delta}\big)$ as $\delta \to 0$. Hence, the optimality result carries over from the linear to the nonlinear situation.

6. Choosing the tolerances

In implementing algorithm REGINN we have the freedom to select the sequence of tolerances $\{\mu_k\}$. Our analysis includes non-constant μ_k's within certain limits, see lemma 4.1. We like to choose the tolerances dynamically such that the overall number $\sum_{k=0}^{N(\delta)-1} i_k$ of passes through the `repeat` loop becomes rather 'small'.

To this end we try to minimize $N(\delta)$, the number of Newton steps, by allowing the μ_k's to be small. However, the tolerances should not be too small to avoid noise amplification while solving the linearization (2.1). In the starting phase of algorithm REGINN the nonlinear defect will be relatively large and the `repeat` loop will terminate in spite of a small tolerance.

We therefore start with a small tolerance and increase it during the Newton iteration. This is in accordance with (4.2). An increase of the tolerance will be indicated when the number of passes through the `repeat` loop of two successive Newton steps increases significantly. The tolerances shall be decreased by a constant factor whenever the consecutive numbers of passes through the `repeat` loop drop.

We propose the choice (6.2). Choose $\mu_{\text{start}} \in]0, 1[$, $\gamma \in]0, 1]$, and let $\widetilde{\mu}_0 = \widetilde{\mu}_1 := \mu_{\text{start}}$. For $k = 2, \ldots, N(\delta) - 1$ define

$$\widetilde{\mu}_k := \begin{cases} 1 - \dfrac{i_{k-2}}{i_{k-1}}\,(1 - \mu_{k-1}) & i_{k-1} \geqslant i_{k-2} \\ \gamma \cdot \mu_{k-1} & \text{otherwise} \end{cases} \tag{6.1}$$

and choose

$$\mu_k := \mu_{\max} \cdot \max\left\{R \cdot \delta / \|F(x_{k-1}) - y^\delta\|_Y,\ \widetilde{\mu}_k\right\} \qquad k = 0, 1, \ldots, N(\delta) - 1 \tag{6.2}$$

where $\mu_{\max} \in]\mu_{\text{start}}, 1[$ bounds the μ_k's away from 1 (uniformly in k and δ). The parameter $\mu_{\max}$ should be very close to 1, for instance, $\mu_{\max} = 0.999$ is reasonable. We know that the `repeat` loop may not terminate if the tolerance is too small. A rapid decrease of the tolerances should be avoided therefore. Restricting γ to the interval [0.9, 1] has proved quite satisfactory in our numerical experiments.

In the following section we demonstrate the performance of the algorithm REGINN together with the strategy (6.2) where μ_{start} is as small as 0.1.

In defining the μ_k's from the auxiliary $\widetilde{\mu}_k$'s we incorporated a *safeguarding* technique to prevent oversolving of (2.5) in the final Newton step. The idea is obvious: if the nonlinear defect of $x_{N(\delta)-1}$ is already close to $R \cdot \delta$ it is superfluous to reduce it in the last step possibly far beyond the desired level by the factor $\widetilde{\mu}_{N(\delta)-1}$. Safeguarding is a standard procedure in inexact Newton methods for well-posed problems, see, e.g., [18, section 6.3].

Remark 6.1. *Our tolerance selection scheme (6.2) can be modified in an obvious way. Replace the quotient i_{k-2}/i_{k-1} by $Q(i_{k-2}/i_{k-1})$ in (6.1). The function $Q : [0, 1] \to [0, 1]$ should be strict monotonically increasing with $Q(0) = 0$ and $Q(1) = 1$, for instance, $Q(t) = t^\beta$, $\beta > 0$. For $\beta < 1$ ($\beta > 1$) the respective tolerances will increase slower (faster) compared to our choice (6.1).*

Further, the factor $\gamma = \gamma_k$ may also be determined from the ratio i_{k-1}/i_{k-2}.

7. Numerical experiments: a model problem

We present numerical experiments for a parameter identification model problem from interior measurements. Because our main assumption (3.2) is satisfied the model problem is well suited to study the performance of the algorithm REGINN. Indeed we will see that some of our theoretical assumptions have exactly the impact predicted by our analysis of the former sections.

We would like to reconstruct c in the 2D-elliptic problem

$$\begin{aligned} -\Delta u + c\,u &= f \qquad \text{in } \Omega \\ u &= g \qquad \text{on } \partial\Omega \end{aligned} \tag{7.1}$$

from the knowledge of u in $\Omega =]0,1[^2$. In (7.1), Δ is the Laplacian. Further, $f \in L^2(\Omega)$ and g is the trace of a function in $H^2(\Omega)$. Let $F : \mathsf{D}(F) \to L^2(\Omega)$ be the operator mapping the parameter c to the solution u of (7.1). Here, $\mathsf{D}(F) = \{c \in L^2(\Omega) \mid \|c - \widetilde{c}\|_{L^2} \leqslant \beta$ for some $\widetilde{c} \geqslant 0\}$ for a positive β small enough, see [6, lemma 2.1]. Identifying c thus reduces to solving the nonlinear problem

$$F(c) = u. \tag{7.2}$$

If u has no zeroes in Ω then we can solve (7.2) for c explicitly: $c^\dagger = (f + \Delta u)/u$ thereby showing that (7.2) has a unique solution $c^\dagger$ which does not depend continuously on the data. Hence, the direct inversion formula is useless if only perturbed data u^δ are available.

Hanke, Neubauer and Scherzer have been able to verify (3.2) in the vicinity of any $c \in \mathsf{D}(F)$ such that $F(c) > 0$ a.e., see [16, example 4.2]. Consequently, we should be able to reproduce some of our theoretical results when applying algorithm `REGINN` to the parameter identification problem (7.2).

The Fréchet derivative $F'(c) : L^2(\Omega) \to L^2(\Omega)$ is given by

$$F'(c)v = -\mathcal{L}(c)^{-1}\big(v \cdot F(c)\big) \tag{7.3}$$

where $\mathcal{L}(c) : H^2(\Omega) \cap H^1_0(\Omega) \to L^2(\Omega)$ is the differential operator $\mathcal{L}(c)u = -\Delta u + c\,u$, cf [6, lemma 2.4]. Hence, the abstract smoothness condition (4.16) in the present situation reduces to

$$(c^\dagger - c_0)/F(c_0) \in H^2(\Omega) \cap H^1_0(\Omega), \tag{7.4}$$

especially, $(c^\dagger - c_0)|_{\partial\Omega} = 0$.

For our numerical approach we discretize (7.1) using finite differences, see e.g. [11] for the following notation. We approximate the action of $\mathcal{L}$ on u in (x_i, y_j) by the difference star

$$\mathcal{L}u(x_i, y_j) \approx h^{-2} \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 + h^2\,c(x_i, y_j) & -1 \\ 0 & -1 & 0 \end{bmatrix} u(x_i, y_j).$$

Here, $h = 1/(n+1)$, $n \in \mathbb{N}$, is the discretization step size and the grid points are $(x_i, y_j) = (i\,h, j\,h)$, $1 \leqslant i, j \leqslant n$. Proceeding in the standard way using lexicographical ordering of the grid points and incorporating the boundary constraints into the right hand side yields the $n^2 \times n^2$-linear system

$$(\mathsf{A} + \mathsf{C})\,\boldsymbol{u} = \boldsymbol{f}$$

where A is the matrix belonging to the difference star of $-\Delta$ and $\mathsf{C} = \mathrm{diag}(c_1, \ldots, c_{n^2})$ is the diagonal matrix with entries $c_{\ell(i,j)} = c(x_i, y_j)$. By $\ell : \{1, \ldots, n\}^2 \to \{1, \ldots, n^2\}$ we denote the lexicographical ordering. Please note that $\boldsymbol{u}_{\ell(i,j)} = u(x_i, y_j) + \mathcal{O}(h^2)$ as $h \to 0$ for u sufficiently smooth.

In this discrete setting we would like to recover C from $\boldsymbol{u}$. Again, in the presence of noise, the direct reconstruction formula $c_\ell = (\boldsymbol{f} - \boldsymbol{A}\boldsymbol{u})_\ell / \boldsymbol{u}_\ell$ is useless. Instead we consider the nonlinear equation

$$\boldsymbol{F}(\mathsf{C}) = \boldsymbol{u} \tag{7.5}$$

with $\boldsymbol{F}:\mathbb{R}^{n^2}\to\mathbb{R}^{n^2}$ defined by $\boldsymbol{F}(\mathbf{C})=(\mathbf{A}+\mathbf{C})^{-1}\boldsymbol{f}$. The function $\boldsymbol{F}$ is differentiable with Jacobian

$$\boldsymbol{F}'(\mathbf{C})w=-(\mathbf{A}+\mathbf{C})^{-1}\big(\boldsymbol{F}(\mathbf{C})\odot w\big) \tag{7.6}$$

where $\odot$ denotes the component-wise multiplication of vectors. Similar to the infinite dimensional setting we can verify (3.2) for $\boldsymbol{F}'$.

For our numerical experiments we have chosen the following set-up: the parameter to be identified is $c^{\dagger}(x,y)=1.5\sin(4\pi x)\sin(6\pi y)+3((x-0.5)^2+(y-0.5)^2)+2$, see figure 5 (top). Further, f and g have been selected such that $u(x,y)=16\,x\,(x-1)\,y\,(1-y)+1$ is the solution of (7.1) with respect to $c^{\dagger}$.

As the perturbed right-hand side $\boldsymbol{u}^{\delta}$ of (7.5) we worked with $\boldsymbol{u}^{\delta}=\boldsymbol{u}+\delta\,\boldsymbol{v}$. Here, $\boldsymbol{u}_{\ell(i,j)}=u(x_i,y_j)$ and $\boldsymbol{v}=\boldsymbol{z}/\|\boldsymbol{z}\|_h$ with $\boldsymbol{z}$ being a vector with random entries uniformly distributed in $[-1,1]$. Hence, $\|\boldsymbol{u}-\boldsymbol{u}^{\delta}\|_h=\delta$ measured in the weighted Euclidean norm $\|\cdot\|_h=h\,\|\cdot\|_2$ on $\mathbb{R}^{n^2}$ which approximates the $L^2(\Omega)$-norm.

The eigenvalues of $\mathbf{A}$ are known explicitly, see, e.g., [11]. Thus,

$$\|\boldsymbol{F}'(\mathbf{C})\|_h\leqslant\|(\mathbf{A}+\mathbf{C})^{-1}\|_h^2\,\|\boldsymbol{f}\|_h\leqslant\|\mathbf{A}^{-1}\|_h^2\,\|\boldsymbol{f}\|_h\leqslant\frac{1}{4\pi^4}\left(\frac{\pi h/2}{\sin(\pi h/2)}\right)^4\|\boldsymbol{f}\|_h$$

for $\mathbf{C}\geqslant 0$. The scaling requirement (2.7) will be satisfied automatically in our computations below. We are thus allowed to use the ν-method, $\nu=1$, as inner regularization scheme throughout.

In our first experiment we shall illustrate the regularization property, see theorem 5.1, and the growth behaviour of $N(\delta)$ as $\delta\to 0$, see theorem 4.2. In order not to pollute the asymptotic behaviour by other effects we fix $\mu_k=0.995$. We start our Newton iteration on (7.5) with $\mathbf{C}_0=\mathrm{diag}(c^0_{\ell})$, $c^0_{\ell(i,j)}=c_0(x_i,x_j)$, where

$$c_0(x,y)=3\,((x-0.5)^2+(y-0.5)^2)+2+48\,x\,(x-1)\,y\,(1-y) \tag{7.7}$$

which satisfies (7.4).

The results presented in figures 2, 3 and 4 are based on the parameter $R=3$, see (2.5), and the discretization step size $h=1/100$.

Figure 2 displays the relative error $\|\mathbf{C}_{N(\delta)}-\mathbf{C}^{\dagger}\|_h/\|\mathbf{C}^{\dagger}\|_h$ for $\delta\in\{10^{-(r+1)/2}\,|\,r=1,\ldots,11\}$ where $\mathbf{C}^{\dagger}$ is $c^{\dagger}$ evaluated at the grid points. Since both coordinate axes in figure 2 are scaled logarithmically the linear decrease of the error with a slope of about 1/2 indicates that $\|\mathbf{C}_{N(\delta)}-\mathbf{C}^{\dagger}\|_h=\mathcal{O}\big(\delta^{1/2}\big)$ as $\delta\to 0$. This is the optimal rate according to our theory, see (5.1).

The curve of the semi-logarithmical plot in figure 3 demonstrates the asymptotic relation $N(\delta)=\mathcal{O}(|\log\delta|)$ as $\delta\to 0$, see (4.4), which is, in turn, a confirmation of the linear decrease of the nonlinear residuals. Recall that $\eta\approx\mu_k$ for k large, see (4.18). Thus, the slope of

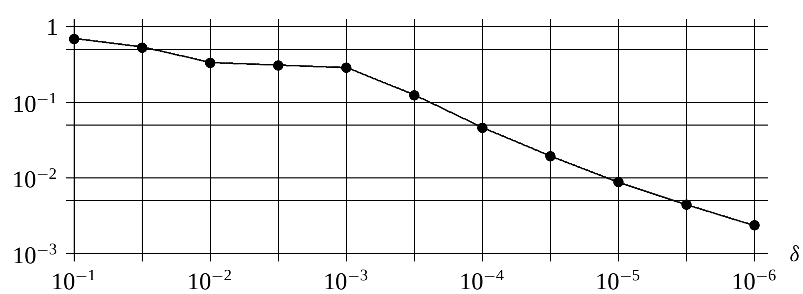

Figure 2. Relative error versus noise level δ ($\mu_k=0.995$ for all k).

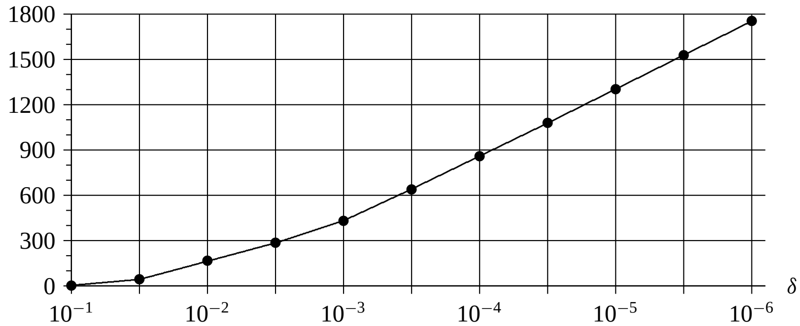

Figure 3. Stopping index $N(\delta)$ versus noise level δ ($\mu_k = 0.995$ for all k).

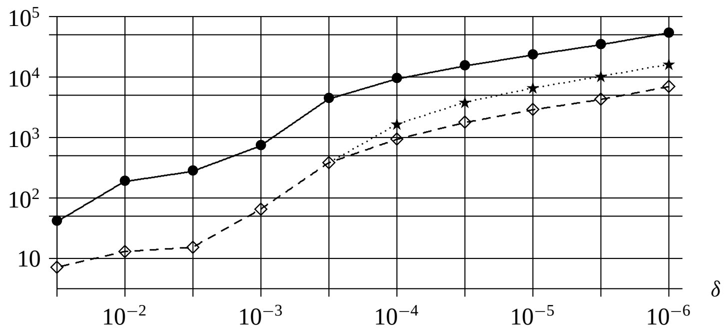

Figure 4. The overall number $\sum_{k=0}^{N(\delta)-1} i_k$ of inner iteration steps versus noise level δ. Solid line: $\mu_k = 0.995$ for all k, dotted and dashed lines: μ_k chosen according to (6.2) with $\mu_{\text{start}} = 0.1$, $\mu_{\max} = 0.999$, where $\gamma = 1.0$ (dotted with $\star$) and $\gamma = 0.9$ (dashed with $\diamond$).

the curve in figure 3 matches perfectly with its theoretical value $-1/\log 0.995 \approx 459$ for $\delta \leqslant 10^{-3}$, i.e., $N(10^{-(r+2)/2}) - N(10^{-(r+1)/2}) \approx 459/2$ for $r = 5, \ldots, 10$.

Replacing the static tolerance strategy $\mu_k = 0.995$ for all k by the dynamic strategy (6.2) improves—as expected—the efficiency of algorithm `REGINN`. In figure 4 we plotted the overall number $S_\delta := \sum_{k=0}^{N(\delta)-1} i_k$ of inner iteration steps versus the noise level δ. The numerical value of S_δ is a reliable measure for the computational effort. Hence, figure 4 shows clearly that the dynamic strategy ($\mu_{\text{start}} = 0.1$, $\mu_{\max} = 0.999$, $\gamma = 1.0$ respectively $\gamma = 0.9$) outperforms the static one. Please note that both tolerance choices lead to the same relative errors displayed in figure 2. The stopping indices $N(\delta)$ relative to the dynamic strategy are rather small, for instance, $N(10^{-6}) = 32$ ($\gamma = 0.9$).

Now we demonstrate the mode of action of our tolerance selection strategy (6.2) more explicitly. The effects of using (6.2) can be studied by looking at tables 1 and 2. For $\delta = 10^{-5/2}$, $R = 1.5$, $\mu_{\text{start}} = 0.6$, $\mu_{\max} = 0.999$, and $\gamma = 0.95$ the convergence history of algorithm `REGINN` is listed in table 1. The discretization step size is $h = 1/64$. By $d_k := \|\boldsymbol{F}(\mathbf{C}_k) - \boldsymbol{u}^\delta\|_h$ and $e_k := \|\mathbf{C}_k - \mathbf{C}^\dagger\|_h / \|\mathbf{C}^\dagger\|_h$ we denote the nonlinear defect and the relative error of the kth Newton iterate, respectively. We observe the following.

- The tolerances decrease by the factor $\gamma \cdot \mu_{\max}$ (respectively $\mu_{\max}$) during the iteration whenever the stopping indices of the inner iteration drop (respectively do not change) for two successive Newton steps.
- The ratio d_k/d_{k-1} tends to μ_{k-1} from below as k goes to $N(\delta)$, cf (4.18).
- The safeguarding technique indeed prevents `REGINN` from unnecessary work in the final

Table 1. Convergence history of REGINN with respect to the tolerance selection (6.2) where $\mu_{\text{start}} = 0.6$, $\mu_{\max} = 0.999$, and $\gamma = 0.95$.

k	μ_{k-1}	i_{k-1}	d_k/d_{k-1}	$d_k/(R \cdot \delta)$	e_k
1	0.6	9	0.5149	12.509	0.4282
2	0.6	8	0.5447	6.814	0.3438
3	0.5694	8	0.5284	3.600	0.3149
4	0.5689	9	0.4968	1.789	0.3056
5	0.6161	50	0.6091	1.090	0.2853
6	0.9300	205	0.9298	1.013	0.1798
7	0.9861	127	0.9861	0.999	0.1596
$7'$	0.9819	150	0.9819	0.995	0.1526

step. The row $k = 7'$ of table 1 shows the final step in the case when safeguarding is turned off (all other iteration steps remain unchanged).

Safeguarding is more vital for a small μ_{start}, see table 2. With the exception of $\mu_{\text{start}} = 0.1$ all parameters are selected as above. Note that the overall number of inner iteration steps S_δ and the error $e_{N(\delta)}$ are smaller than for the choice $\mu_{\text{start}} = 0.6$.

Table 2. Convergence history of REGINN with respect to the tolerance selection (6.2) where $\mu_{\text{start}} = 0.1$, $\mu_{\max} = 0.999$, and $\gamma = 0.95$.

k	μ_{k-1}	i_{k-1}	d_k/d_{k-1}	$d_k/(R \cdot \delta)$	e_k
1	0.1	14	0.0826	2.007	0.3056
2	0.4978	260	0.4974	0.998	0.1410

Finally, we present the graphs of two reconstructions $\boldsymbol{C}_{N(\delta)}$ to $\boldsymbol{C}^\dagger$ respectively $c^\dagger$ for different initial iterates. We ran the algorithm REGINN using $\delta = 0.01$, $R = 1.4$ and the discretization step size $h = 1/64$. The middle part of figure 5 shows the reconstruction with respect to the starting guess c_0 from (7.7) which satisfies (7.4). The bottom part of figure 5 shows the result for the starting guess $c_0 = 2$ which obviously violates (7.4).

Though both starting iterates have about the same distance to the exact solution, the algorithm REGINN started with (7.7) provides the better reconstruction in less than half the run-time compared to algorithm REGINN started with $c_0 = 2$. This observation has the following explanation. Each Newton step s_{i_k} is in the range of $F'(x_k)^*$. In view of (7.3) we conclude that $s_{i_k}|_{\partial\Omega} = 0$. Thus, the starting guess will not be changed on $\partial\Omega$ as the iteration progresses. This behaviour for the present example lies in the nature of *any* algorithm computing the correction step by approximating the minimum norm solution of (2.1).

8. Discussion and conclusion

We compare our algorithm REGINN to other iterative regularization strategies for nonlinear ill-posed problems studied recently in the literature.

The following type of iteration ($A_n = F'(x_n)$)

$$x_{n+1} = x_n + g_{i_n}\big(A_n^* A_n\big)A_n^*\big(y^\delta - F(x_n)\big) + \big(I - g_{i_n}\big(A_n^* A_n\big)A_n^* A_n\big)(x_0 - x_n) \tag{8.1}$$

has been investigated extensively by several authors, see, e.g., [1, 3, 17]. We discuss (8.1) furnished with the *a posteriori* stopping rule (2.5) (*a priori* stopping strategies are also known).

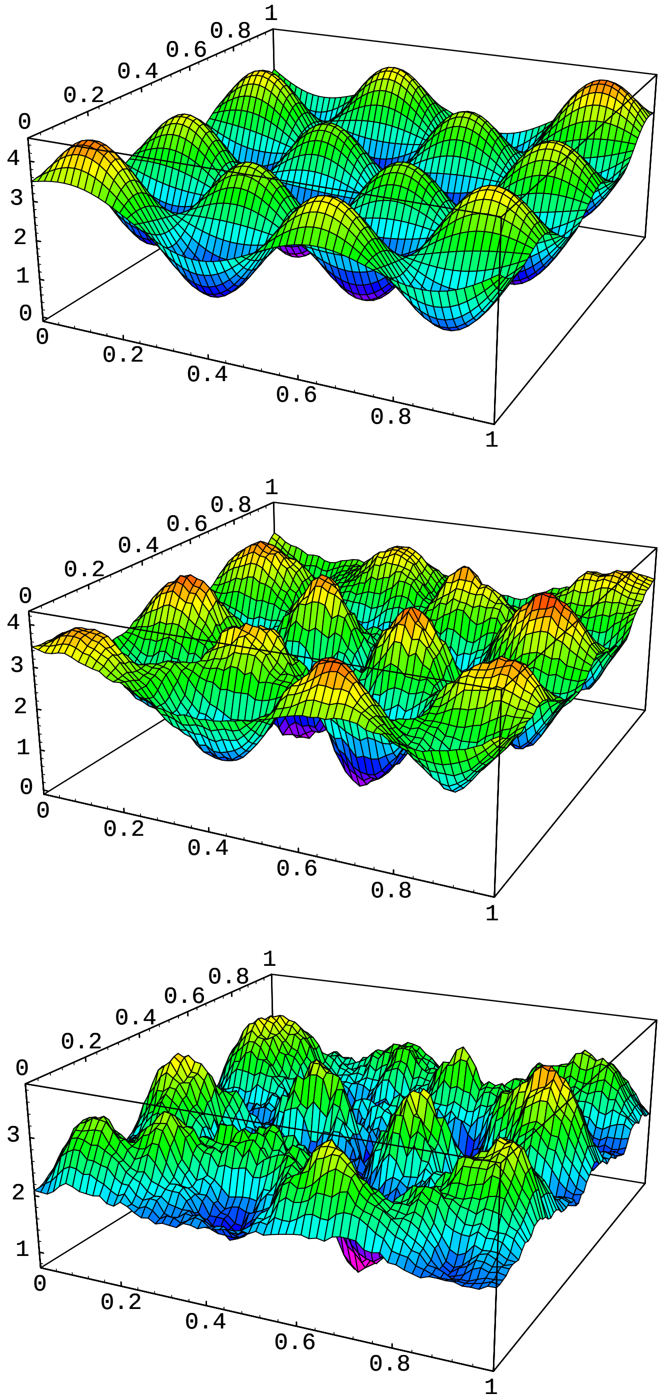

Figure 5. Top: the parameter $c^{\dagger}$ to be identified at the grid points $(i\,h, j\,h)$, $1 \leqslant i, j \leqslant 63$, $h = 1/64$. Middle and bottom: reconstruction $\boldsymbol{C}_{N(\delta)}$ with respect to initial iterates $\boldsymbol{C}_0$ satisfying (middle) and violating (7.4) (bottom).

There are two differences to our algorithm: the appearance of an additional summand, cf (2.4), and the regularization parameters i_n are increased *a priori* by a certain rate.

The additional term is claimed to bring in an extra stability. Indeed, for several linear regularization schemes $\{g_r\}$ (e.g., Tikhonov–Phillips regularization, Landweber iteration) optimal convergence rates have been established under a slightly more general version of (3.2) and under the source condition

$$x^\dagger - x_0 \in \mathsf{R}\Big(\big(F'(x^\dagger)^* F'(x^\dagger)\big)^\kappa\Big), \qquad 0 \leqslant \kappa \leqslant \kappa_{\max} \tag{8.2}$$

where $\kappa_{\max}$ depends on $\{g_r\}$.

For our algorithm we have only been able to prove optimal convergence rates for $\kappa = 1/2$, see (4.16). However, we do not need the extra term which means additional numerical effort in *each* iteration step. For instance, if the Landweber iteration is used in (8.1) an extra operator-vector product with A_n has to be performed in each Landweber step! See [17, method 2.8]. If one considers (8.1) with the Tikhonov–Phillips regularization then an implementation is available requiring only an additional vector subtraction and scalar multiplication per iteration step, see, e.g., [3]. But the latter variation of (8.1) is expensive by itself since a linear system with operator $A_n^* A_n + i_n^{-1} I$ has to be solved exactly in each step. An explicit representation of A_n is not always at hand, see, for instance, our parameter identification problem in the former section, cf (7.6). Therefore, this method is sometimes impossible to use.

The *a priori* determination of the inner regularization parameters i_n in (8.1) makes an oversolving of (2.5) very likely in the final step.

Tautenhahn [23] suggested another theoretically appealing regularization for (1.1) which is closely related to (8.1). He also obtains optimal convergence rates under (8.2). However, the regularizer of Tautenhahn's method is not given explicitly. It is itself the solution of a nonlinear (well-posed) equation. From a practical point of view this nonlinear equation has to be solved approximately by an iteration. This additional approximation process was not incorporated in the convergence analysis.

Our analysis of algorithm `REGINN` does not settle the convergence of $x_{N(\delta)}$ to a solution of (1.1) as $\delta \to 0$ under assumptions weaker than (4.12). Let us consider the source condition (8.2) for $0 < \kappa < 1/2$. We need to find a meaningful upper bound for i_k (cf (4.11) where the source condition (4.16) enters our analysis). Such an estimate is crucial to show that the iterates stay in the ball $B_\rho(x^\dagger)$ by bounding the Newton steps, see (4.17). One, of course, could try to circumvent this problem by some modifications which, however, *must not* affect the numerical performance of `REGINN`. Hence, adding a stabilizing term or choosing the i_k's *a priori* as in (8.1) is ruled out. In our opinion, completely new techniques are necessary to show convergence (and convergence rates) in the general situation (8.2).

Inexact Newton methods for the regularization of nonlinear ill-posed problems have already been suggested by Hanke [13, 14]. In [14] the conjugate gradient (*cg*) iteration serves as inner regularization. Under a slightly weaker form of (3.2) only the convergence of subsequences of $\{x_{N(\delta)}\}_{\delta>0}$ to solutions of (1.1) is shown. We could prove convergence in a weaker norm under (3.5), see (4.5).

One may consider this as a theoretical gap of Newton-*cg*. One may further argue that *cg* will outperform the ν-methods as inner iterations since *cg* reduces the (linear) residual faster. Also, in using *cg* the scaling (2.7) will be superfluous.

The following two aspects weaken the above arguments.

- Hanke and Hanson [15] proposed a method to guarantee (2.7) almost without additional effort.
- The ν-methods have advantages over the *cg*-method concerning the stability of the provided approximative solution with respect to the stopping index. If the *cg*-iteration is

not stopped at the optimal point it diverges more rapidly than semiiterative solvers do, see, e.g., Hanke [12]. Thus, the Newton correction computed by *cg* will be very sensitive to changes in the μ_k's. The performance of Newton-*cg* under a dynamic tolerance selection scheme like (6.2) is unclear therefore.

From a theoretical point of view one cannot—at the present time—pass a judgement whether Newton-*cg* or `REGINN` with the ν-method will perform better. This question has to be answered by extensive numerical experiments.

We end the discussion by commenting on [13]. The regularization property of the Levenberg–Marquardt algorithm (first method in example 2.1) was shown under a slightly weaker form of (3.2) and without a source condition (hence, no rates are given). Hanke's analysis, on the other hand, requires another strong assumption: the regularization parameter has to be chosen from (2.6) with equality! This cannot be realized when allowing *discrete* regularization parameters. Therefore Hanke's approach does not apply to `REGINN` in general.

For the Levenberg–Marquardt algorithm with the weaker condition (2.6) we could prove convergence rates.

In the present paper we gave a regularization analysis for inexact Newton iterations furnished with a rather general class of inner regularizations, see (4.7). For the first time, the linear decrease of the nonlinear residuals could be shown and convergence rates have been established. Moreover, our analysis gave rise to a dynamic selection strategy for the tolerances which greatly improves the performance of our algorithm.

References

[1] Bakushinskii A B 1992 The problem of the convergence of the iteratively regularized Gauss–Newton method *Comput. Math. Phys.* **32** 1353–9

[2] Baumeister J 1987 *Stable Solution of Inverse Problems* (*Advanced Lectures in Mathematics*) (Braunschweig/Wiesbaden: Vieweg)

[3] Blaschke(-Kaltenbacher) B, Neubauer A and Scherzer O 1997 On convergence rates for the iteratively regularized Gauss–Newton method *IMA J. Numer. Anal.* **17** 421–36

[4] Brakhage H 1987 On ill-posed problems and the method of conjugate gradients *Inverse and Ill-Posed Problems* ed H W Engl and C W Groetsch (Boston, MA: Academic)

[5] Chavent G and Kunisch K 1994 Convergence of Tikhonov regularization for constrained ill-posed problems *Inverse Problems* **10** 63–73

[6] Colonius F and Kunisch K 1986 Stability for parameter estimation in two point boundary value problems *J. Reine Angew. Math.* **370** 1–29

[7] Engl H W, Hanke M and Neubauer A 1996 *Regularization of Inverse Problems* (*Mathematics and Its Applications* vol 375) (Dordrecht: Kluwer)

[8] Engl H W, Kunisch K and Neubauer A 1989 Convergence rates for Tikhonov regularization of nonlinear ill-posed problems *Inverse Problems* **5** 63–79

[9] Engl H W, Louis A K and Rundell W (ed) 1997 *Inverse Problems in Geophysical Applications* (Philadelphia: SIAM)

[10] Engl H W, Louis A K and Rundell W (ed) 1997 *Inverse Problems in Medical Imaging and Nondestructive Testing* (Wien: Springer)

[11] Hackbusch W 1992 *Elliptic Differential Equations: Theory and Numerical Treatment* (*Springer Series in Computational Mathematics* vol 18) (Heidelberg: Springer)

[12] Hanke M 1991 Accelerated Landweber iterations for the solution of ill-posed problems *Numer. Math.* **5** 341–73

[13] Hanke M 1997 A regularizing Levenberg–Marquardt scheme, with applications to inverse groundwater filtration problems *Inverse Problems* **13** 79–95

[14] Hanke M 1998 Regularizing properties of a truncated Newton–CG algorithm for non-linear inverse problems *Numer. Funct. Anal. Optim.* **18** 971–93

[15] Hanke M and Hanson P C 1993 Regularization methods for large-scale problems *Surv. Math. Indust.* **3** 253–315

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

[17] Kaltenbacher B 1997 Some Newton-type methods for the regularization of nonlinear ill-posed problems *Inverse Problems* **13** 729–53

[18] Kelley C T 1995 *Iterative Methods for Linear and Nonlinear Equations* (*Frontiers in Applied Mathematics* vol 16) (Philadelphia: SIAM)

[19] Kirsch A 1996 *An Introduction to the Mathematical Theory of Inverse Problems* (*Applied Mathematical Sciences* vol 120) (New York: Springer)

[20] Köhler J and Tautenhahn U 1995 Error bounds for regularized solutions of nonlinear ill-posed problems *J. Inverse Ill-Posed Problems* **3** 47–74

[21] Louis A K 1989 Inverse und schlecht gestellte Probleme *Studienbücher Mathematik* (Stuttgart: Teubner) English translation in preparation

[22] Louis A K 1996 Approximative inverse for linear and some nonlinear problems *Inverse Problems* **12** 175–90

[23] Tautenhahn U 1997 On a general regularization scheme for nonlinear ill-posed problems *Inverse Problems* **13** 1427–37

[24] Zeidler E 1993 *Nonlinear Funtional Analysis and its Applications I: Fixed-Point Theorems* (New York: Springer)