A Multi-grid Method with A Priori and A Posteriori Level Choice for the Regularization of Nonlinear Ill-Posed Problems [1]

Barbara Kaltenbacher

Abstract. In this paper we study a multi-grid method for the numerical solution of nonlinear systems of equations arising from the discretization of ill-posed problems, where the special eigensystem structure of the underlying operator equation makes it necessary to use special smoothers. We provide uniform contraction factor estimates and show that a nested multigrid iteration together with an a priori or a posteriori chosen stopping index defines a regularization method for the ill-posed problem, i.e., a stable solution method, that converges to an exact solution as the data noise level goes to zero, with optimal rates under additional regularity conditions.

Key words. Ill-posed problems, regularization, multi-grid methods, stopping rule, convergence rates

AMS(MOS) Subject Classifications: 65J20, 65R30

1 Introduction

Inverse problems appear in many applications in science and industry; think, e.g., of calculation of the evolution of a system backwards in time or identification of physical parameters. Their mathematical formulation as (presumably solvable) nonlinear operator equations

$$F(x) = y \tag{1}$$

with a continuous operator $F : \mathcal{X} \to \mathcal{Y}$ acting between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ typically leads to an ill-posed problem in the sense that given data y^δ with arbitrary small noise δ in

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

one can possibly arrive at large deviations in the solution. Moreover, since we deal with a nonlinear problem (and will not touch the wide field of globalization strategies in this paper) we have to do a local analysis, assuming that we have an initial guess x_{ig} that is sufficiently close to an exact solution $x^\dagger$ of (1), and working in a neighborhood $\mathcal{B}_\rho(x_{ig})$.

Conditions on the nonlinearity of the operator F, (such as a Lipschitz condition on its Fréchet derivative, implying an estimate

$$\|F(x) - F(\bar{x}) - T(x - \bar{x})\| \leq \tfrac{L}{2}\|x - \bar{x}\|^2$$

[1] Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under grant T 7-TEC and project F1308 within Spezialforschungsbereich 13

of the first order Taylor remainder,) as they are usual in the analysis of well-posed problems, have turned out not to be appropriate for ill-posed problems, though. Here, Taylor remainder estimates like

$$\|F(x) - F(\bar{x}) - T(x - \bar{x})\| \le c_{tc} \|T(x - \bar{x})\| \quad \forall x, \bar{x} \in \mathcal{B}_\rho(x_{ig}) \tag{3}$$

(the so-called tangential cone condition, cf. [23]) or, alternatively, range invariance conditions of the type

$$F(x) - F(\bar{x}) = TR(x, \bar{x}) \qquad \|R(x, \bar{x}) - I\| \le c_{ri} \ , \quad \forall x, \bar{x} \in \mathcal{B}_\rho(x_{ig}) \ , \tag{4}$$

which means that the range of divided difference operators $DF(x, \bar{x})$ in

$$F(x) - F(\bar{x}) = DF(x, \bar{x})(x - \bar{x}) \tag{5}$$

remains unchanged in $\mathcal{B}_\rho(x_{ig})$, are more useful and have been proven to hold for several examples of inverse problems; see, e.g., [3], [10], [12], and note that (4) is closely related to the affine covariant Lipschitz condition in [4] (cf. [13]). Here T in (3) or (4) is a known bounded linear operator that can be seen as some approximation to $F'(x)$ in $\mathcal{B}_\rho(x_{ig})$, which we must be able to make use of in our computations; note that in contrast to this, the linear operators $R(x, \bar{x})$ in (4) need not be known explicitely. The divided difference operators $DF(x, \bar{x})$ are not uniquely defined by (5); if F is Fréchet differentiable in $\mathcal{B}_\rho(x_{ig})$, then one may set $DF(x, \bar{x}) = \int_0^1 F'(\bar{x} + \theta(x - \bar{x}))d\theta$.

As a consequence of the ill-posedness, stable approximation methods for a solution (1) — so-called regularization methods (cf. [6], [7], [18], [19], [21], [24], [25]) — converge in general arbitrarily slowly as the noise level δ goes to zero, and convergence rates can only be obtained under additional regularity or source wise representation conditions

$$x_{ig} - x^\dagger \in \mathcal{R}((F'(x^\dagger)^* F'(x^\dagger))^\mu) \tag{6}$$

where $\mu > 0$ is a real exponent and the operator $(F'(x^\dagger) * F'(x^\dagger))^\mu$ is defined in the sense of functional calculus (cf., e.g., Section 2.3 in [6]). The under condition (6) optimal convergence rates for regularized approximations $\tilde{x}$ to $x^\dagger$ are

$$\|\tilde{x} - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \tag{7}$$

and therefore always slower than the rate $O(\delta)$ that is typical for well-posed problems.

Regularization methods for infinite dimensional problems (1) are usually defined in an infinite dimensional setting and have to be discretized for an implementation. On the other hand, discretization itself can have a stabilizing effect: Finite-dimensional problems (as soon as existence and uniqueness of a solution is guaranteed) are always well-posed in the sense of stable data dependence. An analysis of regularization by discretization for linear ill-posed problems (see, e.g., Section 3.3 in [6]) shows, that, unlike discretization in in pre-image space, image space projection in fact yields a generically convergent regularization method (see also Subsection 2.1 below). The convergence of such methods was studied, e.g., in [5], [8], [20] for linear and in [14], [16] for nonlinear problems. Here we are going to consider the question of efficiently solving the resulting linear or nonlinear systems of equations by multi-grid methods.

Multi-grid Methods (MGM) are already well established as extremely efficient solvers for large scale systems of equations originating from the discretization of partial differential or second kind integral equations (see, e.g, [1], [2], [9]). For systems of equations arising from regularization by projection (or any other discretized regularization method with small regularization parameter) of first kind integral equations or, more generally, of ill-posed problems in the sense described above, the situation is different, though (cf. [11], [22]). This is due to the adverse structure of the singular systems of compact operators — high frequent eigenfunctions correspond to small singular values — that foils the smoothing properties of schemes such as Gauß-Seidel iteration, used as smoothers in MGM for well-posed problems. In the context of ill-posed problems, King [17] has proposed an alternative smoother that indeed turns out to be appropriate for this class of problems, (see [17], [15],) and that we will also make use of here. The fact that, with this special smoother, the nonlinear multi-grid method of Hackbusch (cf. Chapter 9 in [9]) can be successfully applied to nonlinear ill-posed problems, is supposed to be one of the major new contributions of the present paper (see Section 3).

Often the exponent μ in the source condition is not known explicitely. An a posteriori selection of the discretization level is therefore of considerable practical interest, but was still an open question in the context of MGM for ill-posed (even linear) problems. The a posteriori choice we propose here, is a discrepancy type principle, that stops the nested iteration as soon as the residual is of the order of magnitude of the data noise. This strategy is shown to be an order optimal regularization method in the sense of convergence of the approximants $\tilde{x}$ to the exact solution $x^\dagger$ as δ in (2) goes to zero, with rates (7) under corresponding source conditions (see Section 4).

A closing Section 5 on numerical experiments should illustrate the foregoing theoretical results.

In the following we will use the notation c or C, for positive constants that are typically "small" or "large" (in the sense of "not necessarily small") respectively; they can have different values whenever they appear in the text, but they are introduced in such a way that no confusion should occur.

2 Regularization by Projection

2.1 Linear Problems

For a linear ill-posed problem

$$Tx = y \tag{8}$$

with a compact linear operator $T : \mathcal{X} \to \mathcal{Y}$ and given noisy data y^δ satisfying (2) with $y \in \mathcal{R}(T)$, regularization by projection ([5, 8, 20]) is defined by projecting (8) onto finite dimensional nested subspaces $\mathcal{Y}_n$ of $\mathcal{Y}$, whose union is dense in $\mathcal{Y}$:

$$\mathcal{Y}_1 \subseteq \mathcal{Y}_2 \subseteq \ldots \subseteq \mathcal{Y} \quad \wedge \quad \overline{\bigcup_{n\in\mathbb{N}} \mathcal{Y}_n} = \mathcal{Y} \ , \tag{9}$$

and solving it in a best approximate sense

$$x_n := (\mathcal{Q}_n T)^\dagger \mathcal{Q}_n y^\delta \ , \tag{10}$$

where

$$\mathcal{Q}_n := Proj_{\mathcal{Y}_n}^{\mathcal{Y}}$$

The so defined approximation x_n to the exact solution $x^\dagger := T^\dagger y$ lies in the finite dimensional space

$$\mathcal{X}_n := T^* \mathcal{Y}_n$$

and obeys the error estimate

$$\|x_n - x^\dagger\| \leq \|[I - \mathcal{P}_n]x^\dagger\| + \frac{\delta}{\underline{\gamma}_n} \ , \tag{11}$$

where

$$\underline{\gamma}_n := \frac{1}{\|(\mathcal{Q}_n T)^\dagger \mathcal{Q}_n\|} = \inf_{v\in\mathcal{X}_n} \frac{\|\mathcal{Q}_n T v\|}{\|v\|}$$

and

$$\mathcal{P}_n := Proj_{\mathcal{X}_n}^{\mathcal{X}} \ .$$

In the noiseless case $\delta = 0$, x_n is just the orthogonal projection of $x^\dagger$ onto $\mathcal{X}_n$ and therefore, by (9) converges to $x^\dagger$. Non-vanishing data noise is amplified by the factor $\underline{\gamma}_n^{-1}$ that, by the compactness of T, goes to infinity as $n \to \infty$. Therefore, an optimal choice of the discretization parameter n, which here acts as a regularization parameter, has to find an optimal balance between the two different error terms on the right hand side of (11). In general, convergence as $\delta \to 0$ is obviously obtained by choosing

$$N(\delta) \to \infty \ \wedge \ \frac{\delta}{\underline{\gamma}_{N(\delta)}} \to 0 \text{ as } \delta \to 0 \ . \tag{12}$$

Under a source condition

$$x^\dagger \in \mathcal{R}((T^*T)^\mu) \tag{13}$$

the first term in (11) goes to zero at some rate

$$\overline{\gamma}_n(\mu) := \|(I - \mathcal{P}_n)(T^*T)^\mu\| \ , \tag{14}$$

so that an optimal choice of N should satisfy not only (12) but, more precisely,

$$\underline{\gamma}_{N(\delta)} \overline{\gamma}_{N(\delta)}(\mu) \sim \delta \ . \tag{15}$$

For a more detailed exposition of regularization by projection see, e.g., Section 3.3 in [6], where it is also shown that truncated SVD, (which means that the finite dimensional spaces $\mathcal{Y}_n$ are spanned by eigenvectors of TT^*,) is optimal both with respect to stability and convergence; since eigensystems are often hard to obtain in practice, though, we are here more interested into projection on spaces of piecewise polynomial functions such as splines or finite elements.

The fact that the data noise amplification factor $\underline{\gamma}_n^{-1}$ goes to infinity reflects the ill-posedness of the underlying infinite dimensional problem (8) or, in other words, the smoothing property of the forward operator T that, on its part, makes $\overline{\gamma}_n(\mu)$ as defined in (14) go to zero as $n \to \infty$. In the analysis of projection methods for ill-posed problems some more real sequences with similar behavior turn up (cf. e.g., [8], [16], [17]) and will also be used here, namely

$$\underline{\underline{\gamma}}_n := \inf_{v \in \mathcal{Y}_n} \frac{\|T^* v\|}{\|v\|} \;,$$

$$\gamma_n := \inf_{v \in \mathcal{X}_{n-1}^{\perp} \cap \mathcal{X}_n} \frac{\|\mathcal{Q}_n T v\|}{\|v\|} \;,$$

$$\overline{\overline{\gamma}}_n := \|(I - \mathcal{Q}_n) T\| \;,$$

for which, by the interpolation inequality we have

$$\overline{\gamma}_n(\mu) \le (\|T[I - \mathcal{P}_n]\|)^{2\mu} = (\|[I - \mathcal{Q}_n]T[I - \mathcal{P}_n]\|)^{2\mu} \le \overline{\overline{\gamma}}_n^{2\mu} \tag{16}$$

for $\mu \le \frac{1}{2}$, where we have made use of the identity

$$\mathcal{Q}_n T [I - \mathcal{P}_n] = 0 \;. \tag{17}$$

These real sequences are related by

$$\underline{\underline{\gamma}}_n \le \underline{\gamma}_n \le \gamma_n \le \overline{\overline{\gamma}}_{n-1} \tag{18}$$

(see [15], [16], [17]) and each of them goes to zero at a speed, that is determined by the smoothing property of T, in combination with the approximation property of the finite dimensional spaces $\mathcal{Y}_n$.

In order to quantify this asymptotic behavior, we denote by h_n some discretization parameter (think, e.g., of a mesh size) that geometrically goes to zero as $n \to \infty$:

$$h_n = C_\sigma \sigma^n \tag{19}$$

for some $\sigma \in (0, 1)$, and, by H^{t-p} a Hilbert space that contains $\mathcal{Y}$:

$$H^s := \mathcal{X} \quad H^{t-p} \supseteq H^t := \mathcal{Y} \;.$$

Assumption 1. (Smoothing property of T^*:)
Let, for some $p > 0$,

$$\|T^*\|_{H^{t-p} \to H^s} \le C \;,$$

Assumption 2. (Approximation property of $\mathcal{Y}_n$)
Let, for p as in Assumption 1

$$\|I - \mathcal{Q}_n\|_{H^t \to H^{t-p}} \leq C h_n^p \ ,$$

These assumptions (whose notation obviously suggests Sobolev spaces and piecewise polynomial approximation) yields the following crucial estimates.

$$\overline{\overline{\gamma}}_n \leq \|T^*\|_{H^{t-p}\to H^s} \|I - \mathcal{Q}_n\|_{H^t\to H^{t-p}} \leq \overline{\overline{C}} h_n^p \ , \tag{20}$$

and, by (16),

$$\overline{\gamma}_n(\mu) = O(h_n^{2p\mu}) \ . \tag{21}$$

In the multi-grid analysis in image space, which we will carry out under the nonlinearity condition (3), we will need another quantity of the type of $\overline{\overline{\gamma}}_n$, namely

$$\tilde{\overline{\gamma}}_n := \sqrt{\|[I - \mathcal{Q}_n]TT^*\|}$$

which under smoothing and approximation properties

$$H^{t-2p} \supseteq H^t \ , \quad \|TT^*\|_{H^{t-2p}\to H^t} \leq C \ , \quad \|I - \mathcal{Q}_n\|_{H^t\to H^{t-2p}} \leq C h_n^{2p} \tag{22}$$

satisfies

$$\tilde{\overline{\gamma}}_n \leq \tilde{C} h_n^p \ . \tag{23}$$

On the other hand, in order to bound the data noise propagation from above, we need some kind of inverse inequality on $\mathcal{Y}_n$:

Assumption 3. (Inverse inequality)
Let, for some $\tilde{p} > 0$,

$$\sup_{v\in\mathcal{Y}_n} \frac{\|v\|_Y}{\|T^*v\|_X} \leq C h_n^{-\tilde{p}}$$

from which we get

$$\gamma_n \geq \underline{\gamma}_n \geq \underline{\underline{\gamma}}_n \geq \underline{\underline{c}} h_n^{\tilde{p}} \ . \tag{24}$$

(see [16]).

Due to the estimate (18) it follows from these asymptotics, that $\tilde{p}$ is always smaller than or equal to p. We will here assume that

$$\tilde{p} = p$$

which, when thinking of the $\mathcal{Y}_n$ as piecewise polynomial spaces, seems to be realistic, and implies uniform boundedness of the quotient

$$\frac{\overline{\overline{\gamma}}_n}{\underline{\underline{\gamma}}_n} \leq \frac{\overline{\overline{C}}}{\underline{\underline{c}}}$$

with $\overline{\overline{C}}$, $\underline{\underline{c}}$ as in (20), (24). Moreover, this yields boundedness of the operator $T(\mathcal{Q}_nT)^\dagger\mathcal{Q}_n$,

$$\begin{aligned}\|T(\mathcal{Q}_nT)^\dagger\mathcal{Q}_n\| &= \left(\|\mathcal{Q}_nT(\mathcal{Q}_nT)^\dagger\mathcal{Q}_n\|^2 + \|[I-\mathcal{Q}_nT](\mathcal{Q}_nT)^\dagger\mathcal{Q}_n\|^2\right)^{\frac{1}{2}} \\ &\leq \sqrt{1+(\tfrac{\overline{\overline{C}}}{\underline{\underline{c}}})^2} =: C_{T\mathcal{Q}} \ ,\end{aligned} \tag{25}$$

which makes sense, since $T(\mathcal{Q}_nT)^\dagger\mathcal{Q}_n$ should in some sense approximate the orthogonal projection $TT^\dagger = Proj^{\mathcal{Y}}_{\mathcal{R}(T)}|_{\mathcal{D}(T^\dagger)}$; note that, on the other hand,

$$(\mathcal{Q}_nT)^\dagger\mathcal{Q}_nT = \mathcal{P}_n \tag{26}$$

corresponds to $T^\dagger T = Proj^{\mathcal{X}}_{\mathcal{N}(T)^\perp}$.

Under these conditions, (12), (15) give convergence of $x_{N(\delta)}$ to $x^\dagger$ at optimal rates $O(\delta^{\frac{2\mu}{2\mu+1}})$. This is also the case for the a posteriori discretization choice

$$N(\delta) := \min\{n \mid \|Tx_n - y^\delta\| \leq \tau\delta\} \in \mathbb{N}\cup\{\infty\} \tag{27}$$

for some fixed constant $\tau > 1$ (see [16]).

2.2 Nonlinear Problems

The method of regularizing by finite-dimensional projection can, under nonlinearity conditions such as (3) or (4), be extended to the nonlinear case (1), by setting x_n equal to a solution of

$$\mathcal{Q}_nF(x_n) = \mathcal{Q}_ny^\delta \ , \quad x_n \in x_{ig} + \mathcal{X}_n := x_{ig} + T^*\mathcal{Y}_n \ , \tag{28}$$

which is well defined and lies in $\mathcal{B}_\rho(x_{ig})$ for all n up to

$$\bar{n}_{pr} :=:= \min\{n \in \mathbb{N} \mid C_{1pr}\|x_{ig} - x^\dagger\| + C_{2pr}\sigma^{-np}\delta > \rho\} - 1 \ , \tag{29}$$

with constants C_{1pr}, C_{2pr} depending only on the quantities ρ, c_{ri} or c_{tc}, respectively, in the nonlinearity conditions; here $x^\dagger$ is the unique (Proposition 2.1 in [10], Lemma 1 in [16]) solution of (1) with $x_{ig} - x^\dagger \in \mathcal{N}(T)^\perp$. Moreover

$$\|x_n - x^\dagger\| \leq C_{pr}(\|[I-\mathcal{P}_n](x_{ig}-x^\dagger)\| + \frac{\delta}{\underline{\gamma}_n}) \tag{30}$$

holds for some constant C_{pr} independent of n and of δ; under condition (3) one additionally has

$$\|T(x_n - x^\dagger)\| \leq C_{pr}(\overline{\overline{\gamma}}_n\|[I-\mathcal{P}_n](x_{ig}-x^\dagger)\| + \delta) \ , \tag{31}$$

see [16].

Like in the linear case one obtains convergence as $\delta \to 0$, with optimal rates by an a priori ((12), (15)), or a discrepancy type a posteriori ((74) below) discretization choice, under source conditions

$$x_{ig} - x^\dagger \in \mathcal{R}((T^*T)^\mu) \tag{32}$$

(cf. [16]).

3 Multi-grid analysis

The nonlinear multi-grid method (NMGM, cf. [9]) for the approximate solution of the projected equation (1) with noisy data y^δ

$$\mathcal{Q}_n F(x) = \mathcal{Q}_n y^\delta =: f \in \mathcal{Y}_n$$

in $x_{ig} + \mathcal{X}_n$ can be written as:

Algorithm 1. (NMGM)
$n \in \mathbb{N}$
Given $\tilde{x}_k \in x_{ig} + \mathcal{X}_k \cap \mathcal{B}_\rho(x_{ig})$ and $\tilde{f}_k = \mathcal{Q}_k F(\tilde{x}_k)$ for $k = 0, \ldots, n-1$;
$x_{start} \in \mathcal{X}_n$, $f \in \mathcal{Y}_n$
If $n = 0$ then
    $x := \Phi_0(x_{start}, f)$
else
    $x := S_n(x_{start}, f)$
    $d_{n-1} := \mathcal{Q}_{n-1}(F(x) - f)$
    choose $\lambda \in \mathbb{R}^+$
    $d_{n-1} := \tilde{f}_{n-1} - \lambda d_{n-1}$
    $s_{n-1} := \tilde{x}_{n-1}$
    for $i = 1 : 1 : m$
        $s_{n-1} = NMGM(n-1, s_{n-1}, d_{n-1})$
    endfor
    $x = x + \frac{1}{\lambda}(s_{n-1} - \tilde{x}_{n-1})$
endif
$NMGM(n, x_{start}, f) := x.$

Here $\Phi_0(x_{start}, f)$ is a (possibly iterative) solution method on the coarsest grid that is supposed to approximate a solution of $\mathcal{Q}_0 F(x) = \mathcal{Q}_0 f$ in $x_{ig} + \mathcal{X}_0$ in a sense to be specified below (see (44), (62)).

As already mentioned in the introduction, we have to take care of the special eigensystem structure of compact operators, when choosing the smoother S_n in a multi-grid algorithm for ill-posed problems. Generalizing the smoothing operator proposed in [17] to the nonlinear situation, we here set

$$S_n(x_{start}, f) = x_{start} - \xi_n T^*[I - \mathcal{Q}_{n-1}](\mathcal{Q}_n F(x_{start}) - f) \ , \quad x_{start} \in \mathcal{X}_n, f \in \mathcal{Y}_n \ , \tag{33}$$

with $\xi_n \approx \gamma_n^{-2}$, more precisely, we demand on one hand

$$\xi_n \overline{\overline{\gamma}}_{n-1}^2 \leq 1 \tag{34}$$

and, on the other hand either

$$\xi_n \gamma_n^2 \geq c_{\xi\gamma} > 0 \ , \tag{35}$$

or

$$\overline{\overline{\gamma}}_{n-1}\underline{\gamma}_n^{-1}\sqrt{1-\xi_n\gamma_n^2} \leq \overline{c}_{\xi\gamma} < 1 \tag{36}$$

i.e.,

$$\xi_n \sim h_n^{-2p} \ ,$$

so that, furthermore, if (23) holds,

$$\xi_n \overline{\tilde{\gamma}}_{n-1}^2 \leq C_{\xi\overline{\tilde{\gamma}}} \ , \tag{37}$$

for some constant $C_{\xi\overline{\tilde{\gamma}}}$; note that, by (18) $c_{\xi\gamma}$ is always smaller than one.

This operator S_n acts as a stabilized rough approximation of the inverse of $\mathcal{Q}_nF$ on the "high frequent" part $[I-\mathcal{Q}_{n-1}]\mathcal{Y}_n$ of $\mathcal{Y}_n$ while leaving the "low frequent" part $\mathcal{Q}_{n-1}\mathcal{Y}_n = \mathcal{Y}_{n-1}$ unchanged. Indeed, in the linear case $F(x) = Tx$, we have, with $x_{start} := 0$ and

$$S_n^{lin} := \xi_n T^*[I-\mathcal{Q}_{n-1}]\mathcal{Q}_n \ ,$$

that $(I - S_n^{lin}T)$ is symmetric with spectrum contained in $[0,1]$, since

$$\langle (I-S_n^{lin}T)x, x\rangle = \|x\|^2 - \xi_n \|\mathcal{Q}_n[I-\mathcal{Q}_{n-1}]Tx\|^2$$

and by (34); therefore, for the "high frequent" part $[I-\mathcal{P}_{n-1}](I-S_n^{lin}T)$ we get (see also [17])

$$\begin{array}{rcl} \|[I-\mathcal{P}_{n-1}](I-S_n^{lin}T)\| & \leq & \sqrt{\|(I-S_n^{lin}T)\|}\,\|(I-S_n^{lin}T)^{1/2}[I-\mathcal{P}_{n-1}]\| \\ & \leq & \left(\sup_{x\in\mathcal{X}_n} \frac{\langle (I-S_n^{lin}T)[I-\mathcal{P}_{n-1}]x,[I-\mathcal{P}_{n-1}]x\rangle}{\|x\|^2}\right)^{1/2} \\ & \leq & \left(\sup_{v\in\mathcal{X}_{n-1}^\perp\cap\mathcal{X}_n} \frac{\langle v-\xi_nT^*\mathcal{Q}_nTv,v\rangle}{\|v\|^2}\right)^{1/2} \\ & \leq & \sqrt{1-\xi_n\gamma_n^2} \leq \sqrt{1-c_{\xi\gamma}} \ . \end{array} \tag{38}$$

$$\begin{array}{rcl} \sup\limits_{v\in\mathcal{X}_n} \dfrac{\|T[I-\mathcal{P}_{n-1}](I-S_n^{lin}T)\|}{\|Tv\|} & = & \sup\limits_{v\in\mathcal{X}_n} \dfrac{\|[I-\mathcal{Q}_{n-1}]T[I-\mathcal{P}_{n-1}](I-S_n^{lin}T)\|}{\|Tv\|} \\ & \leq & \overline{\overline{\gamma}}_{n-1}\underline{\gamma}_n^{-1}\sqrt{1-\xi_n\gamma_n^2} \leq \overline{c}_{\xi\gamma} \ , \end{array} \tag{39}$$

where we have used (17) in the third inequality of (38) and in the first equality of (39).

This enables us to prove contraction factor estimates for $NMGM$.

For doing so, the following Lemma will be useful:

Lemma 1. *Consider the inequality*

$$c + C\varepsilon^m < \varepsilon \ , \tag{40}$$

with $c > 0$, $C > 0$, $\varepsilon \in [0,\infty)$, $m \geq 2$, and assume that

$$Cm > 1 \ . \tag{41}$$

Then for all $c \leq c_0$ with

$$c_0 := (1-\tfrac{1}{m})(\tfrac{1}{Cm})^{\frac{1}{m-1}} \ ,$$

(40) has a solution

$$\bar{\varepsilon} := (\tfrac{1}{Cm})^{\frac{1}{m-1}} \in (0,1) \ .$$

Proof. Looking for the global minimum on $[0, \infty)$ of the real function ϕ defined by

$$\phi(\varepsilon) := C\varepsilon^m - \varepsilon \ ,$$

we get, by

$$\phi(0) = 0 \ , \quad \lim_{\varepsilon \to \infty} \phi(\varepsilon) = \infty \ ,$$

$$\phi'(\varepsilon) = 0 \Leftrightarrow \varepsilon = (\tfrac{1}{Cm})^{\frac{1}{m-1}} = \bar{\varepsilon} \ , \quad \phi(\bar{\varepsilon}) = -(1 - \tfrac{1}{m})(\tfrac{1}{Cm})^{\frac{1}{m-1}} < 0 \ ,$$

that this global minimum is equal to $-c_0$ and is assumed at the global minimizer $\bar{\varepsilon} \in (0, 1)$. Therefore, for all $c < c_0$ we have:

$$c + C\bar{\varepsilon}^m = \phi(\bar{\varepsilon}) + c + \bar{\varepsilon} < \bar{\varepsilon} \ .$$

□

Considering first nonlinearity condition (4), we obtain

Theorem 1. *Let Assumptions 1, 2, 3, the range invariance condition (4) with sufficiently small* c_{ri}*, (34), (35) hold, let the cycle regime* m *in algorithm 1 be chosen such that*

$$c_{\xi\gamma} m > 1 \tag{42}$$

and λ *such that*

$$\tfrac{1}{1-c_{ri}} \lambda \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}(f - \mathcal{Q}_n F(S_n(x_{start}, f)))\| + \|\tilde{x}_{n-1} - x_{ig}\| < \rho \tag{43}$$

Assume, moreover, that Φ_0 *has a contraction factor* ϵ_0 *not larger than* $(c_{\xi\gamma} m)^{-\frac{1}{2(m-1)}}$ *in*

$$\forall x_{start}, x \in x_{ig} + \mathcal{X}_0 \cap \mathcal{B}_\rho(x_{ig}) \ : \quad \|\Phi_0(x_{start}, \mathcal{Q}_0 F(x)) - x\| \le \epsilon_0 \|x_{start} - x\| \ . \tag{44}$$

Then the contraction factor of $NMGM$ *is uniformly smaller than one, i.e., there exists an* $\epsilon \in (0, 1)$ *such that*

$$\forall n \in \mathbb{N} \ \forall x_{start}, x \in x_{ig} + \mathcal{X}_n \cap \mathcal{B}_\rho(x_{ig}) \ : \quad \|NMGM(n, x_{start}, \mathcal{Q}_n F(x)) - x\| \le \epsilon \|x_{start} - x\| \ . \tag{45}$$

Proof. For $n = 0$, (45) trivially holds by (44). Assume now, that for all $k \in \{0, \ldots, n-1\}$ (45) with n replaced by k and

$$\epsilon := (c_{\xi\gamma} m)^{-\frac{1}{2(m-1)}} \ , \tag{46}$$

independent of k and n holds, and let $x_{start}, x \in x_{ig} + \mathcal{X}_n \cap \mathcal{B}_\rho(x_{ig})$ be arbitrary but fixed, $f := \mathcal{Q}_n F(x)$.

Denoting by

$$x_{sm} := S_n(x_{start}, \mathcal{Q}_n F(x))$$

the result of the smoothing step, we can decompose the error into its relatively high and relatively low frequent contributions as follows

$$\begin{array}{ll} NMGM(n, x_{start}, \mathcal{Q}_n F(x)) - x & \\ = [I - \mathcal{P}_{n-1}](x_{sm} - x) & \in \mathcal{X}_{n-1}^{\perp} \\ \quad + \frac{1}{\lambda}(s_{n-1} - \tilde{x}_{n-1}) + \mathcal{P}_{n-1}(x_{sm} - x) & \in \mathcal{X}_{n-1} \ . \end{array} \tag{47}$$

The difference between x_{sm} and x can, by (4), be written as

$$\begin{array}{rcl} x_{sm} - x & = & (I - \xi_n T^*[I - \mathcal{Q}_{n-1}]\mathcal{Q}_n T R(x_{start}, x))(x_{start} - x) \\ & = & (I - S_n^{lin} T)(x_{start} - x) + S_n^{lin} T (I - R(x_{start}, x))(x_{start} - x) \ , \end{array}$$

and therefore, by the properties of S_n^{lin} given above, we can estimate

$$\|x_{sm} - x\| \leq (1 + c_{ri}) \|x_{start} - x\| \ , \tag{48}$$

$$\|[I - \mathcal{P}_{n-1}](x_{sm} - x)\| \leq (\sqrt{1 - c_{\xi\gamma}} + c_{ri}) \|x_{start} - x\| \ . \tag{49}$$

We now show that the coarse grid equation

$$\mathcal{Q}_{n-1} F(s) = d_{n-1} \tag{50}$$

with

$$d_{n-1} = \mathcal{Q}_{n-1}\Big(F(\tilde{x}_{n-1}) - \lambda(F(x_{sm}) - F(x))\Big)$$

has a solution $\bar{s}_{n-1} \in \mathcal{X}_{n-1}$ (that is supposed to be approximated by s_{n-1} as given in the algorithm). To do so, consider the sequence

$$s^0 := \tilde{x}_{n-1} \ , \quad s^{l+1} = s^l - (\mathcal{Q}_{n-1}T)^{\dagger}(\mathcal{Q}_{n-1}F(s^l) - d_{n-1}) \tag{51}$$

which obviously remains in $x_{ig} + \mathcal{X}_{n-1}$. If, for all $i \leq l$, $s^i \in \mathcal{B}_{\rho_i}(\tilde{x}_{n-1}) \subseteq \mathcal{B}_{\rho}(x_{ig})$ with $\rho_i := \frac{1-c_{ri}^i}{1-c_{ri}} \lambda \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}(f - \mathcal{Q}_n F(x_{sm}))\|$, then, by (4),

$$\begin{array}{rcl} \|s^{l+1} - s^l\| & = & \|s^l - s^{l-1} - (\mathcal{Q}_{n-1}T)^{\dagger}\mathcal{Q}_{n-1}(F(s^l) - F(s^{l-1}))\| \\ & = & \|(\mathcal{Q}_{n-1}T)^{\dagger}\mathcal{Q}_{n-1}T(I - R(s^l, s^{l-1}))(s^l - s^{l-1})\| \\ & \leq & c_{ri}\|s^l - s^{l-1}\| \\ & \leq & c_{ri}^l \lambda \|(\mathcal{Q}_{n-1}T)^{\dagger}\mathcal{Q}_{n-1}(F(x) - F(x_{sm})\| \\ & \leq & c_{ri}^l \lambda \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}(f - \mathcal{Q}_n F(x_{sm}))\| \end{array}$$

and therefore, by the triangle inequality, $s^{l+1} \in \mathcal{B}_{\rho_{l+1}}(\tilde{x}_{n-1}) \subseteq \mathcal{B}_{\rho}(x_{ig})$. Moreover, summing up differences between indices l_1 and l_2, we see that $(s^l)_{l \in \mathbb{N}}$ is a Cauchy sequence and therefore converges to some $\bar{s}_{n-1}$ in the closed set $x_{ig} + \mathcal{X}_{n-1} \cap \mathcal{B}_{\rho}(x_{ig})$ that, by taking limits $l \to \infty$ in (51), solves (50), i.e.,

$$\mathcal{Q}_{n-1}\Big(\tfrac{1}{\lambda}(F(\bar{s}_{n-1}) - F(\tilde{x}_{n-1})) + F(x_{sm}) - F(x)\Big) = 0 \ . \tag{52}$$

Now, by our induction hypothesis, we can assume that

$$\|s_{n-1} - \bar{s}_{n-1}\| \le \epsilon^m \|\tilde{x}_{n-1} - \bar{s}_{n-1}\| \ . \tag{53}$$

Due to the nonlinearity of the problem, there is a non-vanishing discrepancy

$$\partial s := \tfrac{1}{\lambda}(\bar{s}_{n-1} - \tilde{x}_{n-1}) + \mathcal{P}_{n-1}(x_{sm} - x) \ , \tag{54}$$

that also enters into the total error, and that can be estimated as follows. By (4), applying $(\mathcal{Q}_{n-1}T)^\dagger$ to (52), and (26), we can write

$$0 = \mathcal{P}_{n-1}\Big(\tfrac{1}{\lambda}R(\bar{s}_{n-1}, \tilde{x}_{n-1})(\bar{s}_{n-1} - \tilde{x}_{n-1}) + R(x_{sm}, x)(x_{sm} - x)\Big)$$

and therefore

$$\begin{aligned} \partial s &= \mathcal{P}_{n-1}\partial s \\ &= \mathcal{P}_{n-1}R(\bar{s}_{n-1}, \tilde{x}_{n-1})\partial s + \mathcal{P}_{n-1}(I - R(\bar{s}_{n-1}, \tilde{x}_{n-1}))\partial s \\ &= -\mathcal{P}_{n-1}R(x_{sm}, x)(x_{sm} - x) + \mathcal{P}_{n-1}R(\bar{s}_{n-1}, \tilde{x}_{n-1})(x_{sm} - x) \\ &\quad + \mathcal{P}_{n-1}(I - R(\bar{s}_{n-1}, \tilde{x}_{n-1}))\partial s \ , \end{aligned}$$

which yields

$$\begin{aligned} \|\partial s\| &= \|\Big(I + \mathcal{P}_{n-1}(R(\bar{s}_{n-1}, \tilde{x}_{n-1}) - I)\Big)^{-1}\mathcal{P}_{n-1}(R(\bar{s}_{n-1}, \tilde{x}_{n-1}) - R(x_{sm}, x))(x_{sm} - x)\| \\ &\le \tfrac{2c_{ri}}{1-c_{ri}}\|x_{sm} - x\|, \end{aligned} \tag{55}$$

and, by the triangle inequality, also

$$\tfrac{1}{\lambda}\|\tilde{x}_{n-1} - \bar{s}_{n-1}\| = \|\partial s - \mathcal{P}_{n-1}(x_{sm} - x)\| \le \tfrac{2c_{ri}}{1-c_{ri}}\|x_{sm} - x\| + \|\mathcal{P}_{n-1}(x_{sm} - x)\| \ , \tag{56}$$

which we will use for further estimating the right hand side of (53).

Collecting estimates (48), (49), (53), (55), (56), and by (47) we arrive at

$$\begin{aligned} &\Big(\frac{\|NMGM(n, x_{start}, \mathcal{Q}_nF(x)) - x\|}{\|x_{start} - x\|}\Big)^2 \\ &= \Big(\frac{\|[I - \mathcal{P}_{n-1}](x_{sm} - x)\|}{\|x_{start} - x\|}\Big)^2 + \Big(\frac{\|\frac{1}{\lambda}(s_{n-1} - \bar{s}_{n-1}) + \partial s\|}{\|x_{start} - x\|}\Big)^2 \\ &\le b + (\epsilon^m\sqrt{a - b} + \alpha\sqrt{a})^2 =: \varphi(a, b; c_{ri}) \ , \end{aligned} \tag{57}$$

where

$$\begin{aligned} a &:= \Big(\tfrac{\|x_{sm} - x\|}{\|x_{start} - x\|}\Big)^2 &&\le (1 + c_{ri})^2 =: \bar{a} \\ b &:= \Big(\tfrac{\|[I - \mathcal{P}_{n-1}](x_{sm} - x)\|}{\|x_{start} - x\|}\Big)^2 &&\le (\sqrt{1 - c_{\xi\gamma}} + c_{ri})^2 =: \bar{b} \le \bar{a} - c_{\xi\gamma} \\ \alpha &:= \tfrac{2c_{ri}}{1-c_{ri}}(1 + \epsilon^m) &&= O(c_{ri}) \ , \end{aligned} \tag{58}$$

and $0 \le b \le a$. In order to be able to use the upper bounds in (58) on a and b, we have to check monotonicity of φ with respect to a and b, respectively: obviously $\frac{\partial\varphi}{\partial a}$ is nonnegative for all $0 \le b \le a$ and

$$\begin{aligned} \tfrac{\partial\varphi}{\partial b}(a, b; c_{ri}) &= 1 - (\epsilon^m\sqrt{a - b} + \alpha\sqrt{a})\epsilon^m(a - b)^{-1/2} \\ &= (1 + \epsilon^m)(1 - (1 + \tfrac{2c_{ri}}{1-c_{ri}}\tfrac{\sqrt{a}}{\sqrt{a-b}})\epsilon^m) \ , \end{aligned}$$

so that, due to (58)

$$\forall \tilde{b} \in [b, \bar{b}] \ : \quad \frac{\partial \varphi}{\partial b}(\bar{a}, \tilde{b}; c_{ri}) \geq \frac{\partial \varphi}{\partial b}(\bar{a}, \bar{b}; c_{ri}) \geq 0$$

for all $c_{ri} \leq \frac{(1-\epsilon^m)\sqrt{c_{\xi\gamma}}}{2\epsilon^m+(1-\epsilon^m)\sqrt{c_{\xi\gamma}}}$. Therefore we can estimate

$$\varphi(a, b; c_{ri}) \leq \varphi(\bar{a}, b; c_{ri}) \leq \varphi(\bar{a}, \bar{b}; c_{ri}) = 1 - c_{\xi\gamma} + \epsilon^{2m} c_{\xi\gamma} + O(c_{ri}) \ .$$

An application of Lemma 1, with (42), and (46), as well as the fact that

$$1 - c_{\xi\gamma} - (1 - \tfrac{1}{m})(\tfrac{1}{c_{\xi\gamma}})^{\frac{1}{m-1}} =: \psi(c_{\xi\gamma}) > 0 \ ,$$

(since $c_{\xi\gamma} > \frac{1}{m}$, $\psi(\frac{1}{m}) = 0$, and $\psi'(c) = -1 + (\frac{1}{mc})^{\frac{m}{m-1}} < 0$ for all $c > \frac{1}{m}$,) yields

$$\varphi(a, b; c_{ri}) \leq \epsilon^2 \ ,$$

for all sufficiently (but independently of n) small c_{ri}, and therefore, by (57), as well as induction, the assertion. □

Remark 1. Theorem 1 implies a level-independent contraction factor for the W-cycle ($m = 2$) in Algorithm 1 as soon as $c_{\xi\gamma} > \frac{1}{2}$. Analogously to Proposition 1 in [15] one also obtains a level-independent contraction factor for the alternating cycle ($m = m(n) = 2 - mod(n, 2)$) if $c_{\xi\gamma} > \frac{1}{\sqrt{2}}$.

Under the alternative nonlinearity condition (3), it turns out to be more appropriate do the convergence analysis in image space, i.e., to consider the images of errors under T. Note, however that we still obtain convergence results in pre-image space; see Theorem 3 in Section 4 below.

Theorem 2. *Let Assumptions 1, 2, 3, (22), the tangential cone condition (3) with sufficiently small c_{tc}, (34), (36) hold, let the cycle regime m in algorithm 1 be chosen such that*

$$C_{TQ}(1 + C_{\xi\tilde{\tilde{\gamma}}})m > 1 \ , \tag{59}$$

$$\bar{c}_{\xi\gamma} < (1 - \tfrac{1}{m})(\tfrac{1}{C_{TQ}(1+C_{\xi\tilde{\tilde{\gamma}}})m})^{\frac{1}{m-1}} \tag{60}$$

and λ such that

$$\tfrac{1}{1-c_{tc}C_{TQ}} \lambda \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}(f - \mathcal{Q}_n F(S_n(x_{start}, f)))\| + \|\tilde{x}_{n-1} - x_{ig}\| < \rho \ , \tag{61}$$

where C_{TQ} is as in (25) and $C_{\xi\tilde{\tilde{\gamma}}}$ as in (37). Assume, moreover, that the image of Φ_0 under T has a contraction factor $\epsilon_0 \leq (C_{TQ}(1 + C_{\xi\tilde{\tilde{\gamma}}})m)^{-\frac{1}{m-1}}$:

$$\forall x_{start}, x \in x_{ig} + \mathcal{X}_0 \cap \mathcal{B}_\rho(x_{ig}) \ : \quad \|T(\Phi_0(x_{start}, \mathcal{Q}_0 F(x)) - x)\| \leq \epsilon_0 \|T(x_{start} - x)\| \ . \tag{62}$$

Then the contraction factor of the image of NMGM under T is uniformly smaller than one, i.e., there exists an $\epsilon \in (0,1)$ such that

$$\forall n \in \mathbb{N}\ \forall x_{start}, x \in \mathcal{X}_n \cap \mathcal{B}_\rho(x_{ig}) \ : \quad \|T(NMGM(n, x_{start}, \mathcal{Q}_n F(x)) - x)\| \leq \epsilon \|T(x_{start} - x)\| \ . \tag{63}$$

Proof. Assume, like in the proof is of Theorem 1, that (63) with

$$\epsilon := (C_{T\mathcal{Q}}(1 + C_{\xi\tilde{\overline{\gamma}}})m)^{-\frac{1}{m-1}} \ , \tag{64}$$

and n replaced by k holds for all $k \leq n-1$, and fix two arbitrary elements x_{start}, x of $\mathcal{X}_n \cap \mathcal{B}_\rho(x_n)$.

For the error after the smoothing step we have, by (37),

$$\begin{array}{rcl} \|T(x_{sm} - x)\| & = & \|T(x_{start} - x) - \xi_k TT^*[I - \mathcal{Q}_{n-1}]\mathcal{Q}_n(F(x_{start}) - F(x))\| \\ & \leq & (1 + C_{\xi\tilde{\overline{\gamma}}}(1 + c_{tc}))\, \|T(x_{start} - x)\| \ . \end{array} \tag{65}$$

Moreover, by (17), (34), and (39),

$$\begin{array}{rcl} \|T[I - \mathcal{P}_{n-1}](x_{sm} - x)\| & \leq & \|T[I - \mathcal{P}_{n-1}](I - S_n^{lin}T)(x_{start} - x)\| \\ & & + \overline{\overline{\gamma}}_{n-1} \|S_n^{lin}(F(x_{start}) - F(x) - T(x_{start} - x))\| \\ & \leq & (\overline{c}_{\xi\gamma} + c_{tc})\, \|T(x_{start} - x)\| \ . \end{array} \tag{66}$$

To show solvability of the coarse grid equation (which, although working in image space, we need for applying the induction hypothesis), we again consider the fixed point iteration (51) and assume that for $i = 0, \ldots, l$, $s^i \in \mathcal{B}_{\rho_i}(\tilde{x}_{n-1}) \subseteq \mathcal{B}_\rho(x_{ig})$, $\rho_i := \frac{1-(c_{tc}C_{T\mathcal{Q}})^i}{1-c_{tc}C_{T\mathcal{Q}}} \lambda \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}(f - F(x_{sm}))\|$ (which trivially holds for $i = 0$). Then, by (26), (3), and (25)

$$\begin{array}{rcl} \|s^{l+1} - s^l\| & \leq & \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}T(s^{l+1} - s^l)\| \\ & = & \underline{\gamma}_{n-1}^{-1} \|\mathcal{Q}_{n-1}(F(s^l) - F(s^{l-1}) - T(s^l - s^{l-1})\| \\ & \leq & \underline{\gamma}_{n-1}^{-1} c_{tc} \|T(s^l - s^{l-1})\| \\ & \leq & \underline{\gamma}_{n-1}^{-1} c_{tc} C_{T\mathcal{Q}} \|\mathcal{Q}_{n-1}T(s^l - s^{l-1})\| \\ & \leq & \underline{\gamma}_{n-1}^{-1} (c_{tc}C_{T\mathcal{Q}})^l \lambda \|\mathcal{Q}_{n-1}(F(x_{sm}) - F(x))\| \end{array}$$

and therefore $s^{l+1} \in \mathcal{B}_{\rho_{l+1}}(\tilde{x}_{n-1}) \subseteq \mathcal{B}_\rho(x_{ig})$, so that, like in the proof of Theorem 1, we obtain convergence of $(s^l)_{l\in\mathbb{N}}$ to a solution $\bar{s}_{n-1} \in x_{ig} + \mathcal{X}_{n-1} \cap \mathcal{B}_\rho(x_{ig})$ of the coarse grid equation (52). For ∂s as defined in (54) we get, by (26),

$$\begin{array}{rcl} \|T\partial s\| & \leq & C_{T\mathcal{Q}} \|\mathcal{Q}_{n-1}T\partial s\| \\ & = & C_{T\mathcal{Q}} \|\mathcal{Q}_{n-1}(\frac{1}{\lambda}T(\bar{s}_{n-1} - \tilde{x}_{n-1}) + T(x_{sm} - x))\| \\ & \leq & c_{tc}C_{T\mathcal{Q}}(\frac{1}{\lambda}\|T(\bar{s}_{n-1} - \tilde{x}_{n-1})\| + \|T(x_{sm} - x)\|) \\ & \leq & c_{tc}C_{T\mathcal{Q}}(\|T\partial s\| + \|T\mathcal{P}_{n-1}(x_{sm} - x)\| + \|T(x_{sm} - x)\|) \ , \end{array}$$

where we have used (52) in the inequality of the third line, so that

$$\|T\partial s\| \leq \tfrac{c_{tc}C_{T\mathcal{Q}}}{1-c_{tc}C_{T\mathcal{Q}}}(1+C_{T\mathcal{Q}})\,\|T(x_{sm}-x)\|)\ . \tag{67}$$

By induction hypothesis we have

$$\begin{array}{rcl} \tfrac{1}{\lambda}\|T(s_{n-1}-\bar{s}_{n-1})\| & \leq & \tfrac{1}{\lambda}\epsilon^m\|T(\tilde{x}_{n-1}-\bar{s}_{n-1})\| \\ & \leq & \epsilon^m(\|T\partial s\|+\|T\mathcal{P}_{n-1}(x_{sm}-x)\|)\quad , \\ & \leq & \epsilon^m\tfrac{1+c_{tc}C_{T\mathcal{Q}}}{1-c_{tc}C_{T\mathcal{Q}}}C_{T\mathcal{Q}}\|T(x_{sm}-x)\|\ , \end{array}$$

which, in combination with the estimates (65), (66), (67), and with (47) yields

$$\frac{\|T(NMGM(n,x_{start},\mathcal{Q}_nF(x))-x)\|}{\|T(x_{start}-x)\|} \leq \bar{c}_{\xi\gamma}+C_{T\mathcal{Q}}(1+C_{\xi\tilde{\bar{\gamma}}})\epsilon^m-O(c_{tc})$$

with the factor in $O(c_{tc})$ being independent of n. Lemma 1, together with (59), (60), (64) implies that

$$\frac{\|T(NMGM(n,x_{start},\mathcal{Q}_nF(x))-x)\|}{\|T(x_{start}-x)\|}\leq\epsilon$$

for c_{tc} sufficiently small, so that the assertion follows by induction.

$\square$

Remark 2. It follows from the proof of Theorem 2 that, if

$$\bar{c}_{\xi\gamma} < \frac{1}{4\sqrt{1+(\frac{\overline{\overline{C}}}{\underline{\underline{c}}})^2}(1+C_{\xi\tilde{\bar{\gamma}}})}$$

where $\overline{\overline{C}}$, $\underline{\underline{c}}$, $C_{\xi\tilde{\bar{\gamma}}}$ are as in (20), (24), (37), then the W-cycle yields a uniform contraction factor; similarly one can again derive an upper bound on $\bar{c}_{\xi\gamma}$ for obtaining a level-independent contraction factor for the alternating cycle — however it must be mentioned that such bounds might be hard to check since they contain constants that might not be explicitely available. For arbitrary $\bar{c}_{\xi\gamma}<1$, (59), (60) (note that an m satisfying these requirements always exists) yield a (theoretical) cycle regime m, that implies a level independent contraction factor.

4 Nested iteration with a priori and a posteriori level choice

In this Section we consider the application of the nonlinear multi-grid operator of Algorithm 1 in a nested iteration (cf. [9]) for finding an approximation $\tilde{x}_n$ to a solution x_n of (28):

Algorithm 2. (Nested iteration)
$x_{start} := x_{ig}$
$\tilde{x}_0 := \Phi_0(x_{start}, \mathcal{Q}_0 y^\delta)$
for $k = 1, \ldots, n$
    $x_{start} := \tilde{x}_{k-1}$
    $\tilde{f}_{k-1} := \mathcal{Q}_{k-1} F(\tilde{x}_{k-1})$
    $\tilde{x}_k := NMGM(k, x_{start}, \mathcal{Q}_k y^\delta)$
endfor

If $\tilde{x}_k \in \mathcal{B}_\rho(x_{ig})$ for all $k \leq n-1$, the range invariance condition (4) holds and NMGM has a level-independent contraction factor $\epsilon < 1$ in (45), then, by (30),

$$\begin{aligned}
&\|\tilde{x}_n - x^\dagger\| \\
&\leq \|\tilde{x}_n - x_n\| + \|x_n - x^\dagger\| \\
&\leq \epsilon(\|\tilde{x}_{n-1} - x_{n-1}\| + \|x_{n-1} - x_n\|) + \|x_n - x^\dagger\| \\
&\leq \epsilon^{n+1} \|x_{ig} - x_0\| + \textstyle\sum_{k=1}^n \epsilon^k \|x_{n-k} - x_{n-k+1}\| + \|x_n - x^\dagger\| \\
&\leq \epsilon^{n+1} \|\mathcal{P}_0(x_{ig} - x^\dagger)\| + (1+\epsilon) \textstyle\sum_{k=0}^n \epsilon^k \|x_{n-k} - x^\dagger\| \\
&\leq \epsilon^{n+1} \|\mathcal{P}_0(x_{ig} - x^\dagger)\| + (1+\epsilon) C_{pr} \textstyle\sum_{k=0}^n \epsilon^k \|[I - \mathcal{P}_{n-k}](x_{ig} - x^\dagger)\| \\
&\quad + \tfrac{1+\epsilon}{C_\sigma \underline{\underline{c}}} C_{pr} \textstyle\sum_{k=0}^n (\epsilon\sigma^p)^k \, \sigma^{-np}\delta \,,
\end{aligned}$$

where we have used (19) together with (24) in the last inequality.

Alternatively, the tangential cone condition (3) with uniform contraction factor $\epsilon < 1$ in (63) yields, by (31),

$$\begin{aligned}
&\|T(\tilde{x}_n - x^\dagger)\| \\
&\leq \epsilon^{n+1} \|T\mathcal{P}_0(x_{ig} - x^\dagger)\| + (1+\epsilon) C_{pr} \textstyle\sum_{k=0}^n \epsilon^k \overline{\overline{\gamma}}_{n-k} \|[I - \mathcal{P}_{n-k}](x_{ig} - x^\dagger)\| + \tfrac{1+\epsilon}{1-\epsilon} C_{pr}\delta \,,
\end{aligned} \tag{68}$$

as well as, by (26), and since $[I - \mathcal{P}_n](\tilde{x}_n - x_{ig}) = 0$,

$$\begin{aligned}
&\|\tilde{x}_n - x^\dagger\| \\
&\leq \|\mathcal{P}_n(\tilde{x}_n - x^\dagger)\| + \|[I - \mathcal{P}_n](\tilde{x}_n - x^\dagger)\| \\
&\leq \underline{\gamma}_n^{-1} \|T(\tilde{x}_n - x^\dagger)\| + \|[I - \mathcal{P}_n](x_{ig} - x^\dagger)\| \\
&\leq \tfrac{\sigma^p}{C_\sigma \underline{\underline{c}}} (\tfrac{\epsilon}{\sigma^p})^{n+1} \|T\mathcal{P}_0(x_{ig} - x^\dagger)\| + ((1+\epsilon) C_{pr} \tfrac{\overline{\overline{C}}}{\underline{\underline{c}}} + 1) \textstyle\sum_{k=0}^n (\tfrac{\epsilon}{\sigma^p})^k \|[I - \mathcal{P}_{n-k}](x_{ig} - x^\dagger)\| \\
&\quad + \tfrac{1+\epsilon}{1-\epsilon} \tfrac{C_{pr}}{\underline{\underline{c}}} \sigma^{-np}\delta \,.
\end{aligned} \tag{69}$$

By these estimates we immediately obtain an (optimal) a priori level choice:

Proposition 1. *Let (2), (19), Assumptions 1, 2, 3 hold, assume that*

$$\max\{C_{1pr}, C_1\} \|x_{ig} - x^\dagger\| < \rho$$

and set

$$\bar{n} := \min\{n \in \mathbb{N} \mid C_1 \|x_{ig} - x^\dagger\| + C_2 \sigma^{-np}\delta > \rho\} - 1$$

Assume that either

(a) (4) with (45) and $C_1 := \max\{1, (1+\epsilon)C_{pr}\}\frac{1}{\sqrt{1-\epsilon}}$, $C_2 := \frac{(1+\epsilon)C_{pr}}{C_\sigma \underline{\underline{c}}(1-\epsilon\sigma^p)}$, *or*

(b) (3) with (63),

$$\frac{\epsilon}{\sigma^p} < 1 \tag{70}$$

and $C_1 := \max\{\frac{\sigma^p}{C_\sigma \underline{\underline{c}}}, (1+\epsilon)C_{pr}\frac{\overline{\overline{C}}}{\underline{\underline{c}}} + 1\}\frac{1}{\sqrt{1-\epsilon/\sigma^p}}$, $C_2 := \frac{1+\epsilon}{1-\epsilon}\frac{C_{pr}}{\underline{\underline{c}}}$

holds. Here C_{1pr} *and* $\bar{n}_{pr}$ *are as in Subsection 2.2.*

Then for all $n \leq \min\{\bar{n}, \bar{n}_{pr}\}$, $\tilde{x}_n$ *remains in* $\mathcal{B}_\rho(x_{ig})$ *and, in the noiseless case* $\delta = 0$, *where* $\bar{n} = \bar{n}_{pr} = \infty$, *converges to* $x^\dagger$; *under a source condition (32) with* $\mu \leq \frac{1}{2}$, *a rate*

$$\|\tilde{x}_n - x^\dagger\| = O(q^{2n\mu})$$

is obtained, where

$$q \begin{cases} = \epsilon^{\frac{1}{2\mu}} & \text{if } \epsilon > \sigma^{2p\mu} \\ > \sigma^p & \text{if } \epsilon = \sigma^{2p\mu} \\ = \sigma^p & \text{if } \epsilon < \sigma^{2p\mu} \end{cases} \text{ in case (a)} \quad \text{or} \quad q \begin{cases} = (\frac{\epsilon}{\sigma^p})^{\frac{1}{2\mu}} & \text{if } \epsilon > \sigma^{(1+2\mu)p} \\ > \sigma^p & \text{if } \epsilon = \sigma^{(1+2\mu)p} \\ = \sigma^p & \text{if } \epsilon < \sigma^{(1+2\mu)p} \end{cases} \text{ in case (b)} \tag{71}$$

Moreover, with noisy data satisfying (2) and the a priori level choice $n := N(\delta)$ *according to*

$$N(\delta) \to \infty \ \wedge \ \frac{\delta}{\sigma^{N(\delta)p}} \to 0 \text{ as } \delta \to 0 \quad \text{if } \mu = 0 \text{ in (32)} \tag{72}$$

$$\sigma^{N(\delta)p} q^{2N(\delta)\mu} \sim \delta \quad \text{if } 0 < \mu \leq \frac{1}{2} \text{ in (32)} \tag{73}$$

yields convergence of $\tilde{x}_{N(\delta)}$ *to* $x^\dagger$ *as* $\delta \to 0$ *with, in the case* $0 < \mu \leq \frac{1}{2}$ *in (32), rates that are of the optimal order* $O(\delta^{\frac{2\mu}{2\mu+1}})$ *if* $q = \sigma^p$ *in (71).*

By the proofs of Theorems 1, 2 it can be seen that the condition $q = \sigma^p$ in (71), that is required for optimality, can be achieved by increasing σ such that $\sigma^\nu > \epsilon$ with $\nu = 2p\mu$ or $\nu = (1+2\mu)p$, respectively. Since this needs to hold only in an asymptotic sense, one can also (as it is practically relevant) work with powers of $\frac{1}{2}$ as mesh sizes, i.e., $h_n := C2^{-l(n)}$ with $l(n)$ such that $\limsup_{n\to\infty} \epsilon^n 2^{\nu l(n)} < 1$.

As mentioned in the introduction, it is desirable to have the possibility of an a posteriori level choice that does not need explicit knowledge of μ in (6) and still yields optimal convergence rates. Under the tangential cone condition (3) (which includes the linear case) this can be achieved by the use of a discrepancy type principle:

$$N(\delta) := \min\{n \in \mathbb{N} \mid \|F(x_n) - y^\delta\| \leq \tau\delta\} \in \mathbb{N} \cup \{\infty\} , \tag{74}$$

for some $\tau > 1$, i.e. the refinement is stopped as soon as the residual falls below (some fixed multiple of) the data noise.

Theorem 3. *Let the assumptions of Proposition 1 with (b) be satisfied, fix* $\tau > 1 + \frac{1+\epsilon}{1-\epsilon} C_{pr}$ *and assume that* x_{ig} *is sufficiently close to* $x^\dagger$.

Then for any $\delta \leq \frac{\|F(x_{ig})-y\|}{\tau+1}$ *and for any data* y^δ *with (2), a finite* $N(\delta) \leq \bar{n}$, *with (74) exists, where* x_n *is defined by (28), and* $\bar{n}$ *by (29).*

For any family of data y^δ *with (2),* $x_{N(\delta)}$ *converges to* $x^\dagger$ *as* $\delta \to 0$.

Under an additional source condition (32) for $\mu \leq \frac{1}{2}$ *and if*

$$\epsilon < \sigma^{(1+2\mu)p} \ , \tag{75}$$

the optimal convergence rate (7) is achieved.

Proof. The idea of proof is similar to the one of Theorem 3 in [16].

First of all, to make sure that $N(\delta)$ is not too large (which would result in too much instability) we estimate, for any $n \leq \min\{\bar{n}_{pr}, \bar{n}\}$,

$$\|F(\tilde{x}_n) - y^\delta\| \leq (1 + c_{tc})(1 + \tfrac{1+\epsilon}{1-\epsilon} C_{pr})\delta + \bar{C}\sigma^{np} \sum_{k=0}^{n+1} (\tfrac{\epsilon}{\sigma^p})^k \beta_{n-k} \tag{76}$$

for some constant $\bar{C} > 0$, where we have used (2), (3), (68), and where

$$\beta_{-1} := \|\mathcal{P}_0(x_{ig} - x^\dagger)\| \ , \qquad \beta_k := \|[I - \mathcal{P}_k](x_{ig} - x^\dagger)\| \ , \quad k = 0, 1, 2 \ldots \ .$$

Note that $\beta_k \to 0$ as $k \to \infty$ and therefore, by (70), also

$$\sum_{k=0}^{n+1} (\tfrac{\epsilon}{\sigma^p})^k \beta_{n-k} \to 0 \text{ as } n \to \infty \tag{77}$$

For the a priori defined level

$$N_{apr}(\delta) := \min\{n \mid \bar{C}\sigma^{np} \sum_{k=0}^{n+1} (\tfrac{\epsilon}{\sigma^p})^k \beta_{n-k} \leq \tilde{\tau}\delta\} \tag{78}$$

with $\tilde{\tau} := \tau - (1 + c_{tc})(1 + \frac{1+\epsilon}{1-\epsilon} C_{pr})$, we have, by (78),

$$\begin{aligned}
&\max\{C_{1pr}, C_1\} \|x_{ig} - x^\dagger\| + \max\{C_{2pr}, C_2\}\sigma^{-N_{apr}(\delta)p}\delta \\
&= \max\{C_{1pr}, C_1\} \|x_{ig} - x^\dagger\| + \max\{C_{2pr}, C_2\}\sigma^{-p}\sigma^{-(N_{apr}(\delta)-1)p}\delta \\
&\leq (\max\{C_{1pr}, C_1\} + \max\{C_{2pr}, C_2\}\sigma^{-p}\tfrac{C}{\tilde{\tau}}\tfrac{1}{1-\epsilon/\sigma^p}) \|x_{ig} - x^\dagger\|
\end{aligned}$$

which is smaller than ρ for sufficiently small $\|x_{ig} - x^\dagger\|$, so that $N_{apr}(\delta) \leq \min\{\bar{n}_{pr}, \bar{n}\}$. On the other hand, for any $n \leq \min\{\bar{n}_{pr}, \bar{n}, N(\delta) - 1\}$, (74) together with (76) and (78) implies that $n \leq N_{apr}(\delta) - 1$ and therefore

$$N(\delta) \leq N_{apr}(\delta) \leq \min\{\bar{n}_{pr}, \bar{n}\} \ .$$

This allows us to use the inverse inequality (24) on $\mathcal{Y}_{N_{apr}(\delta)} \supset \mathcal{Y}_{N(\delta)}$ and to insert $n := N_{apr}(\delta)$ into (68) and (69) to obtain

$$\begin{aligned}
\|x_{N(\delta)} - x^\dagger\| &\leq \|x_{N(\delta)} - x_{N_{apr}(\delta)}\| + \|x_{N_{apr}(\delta)} - x^\dagger\| \\
&\leq \tfrac{1}{\underline{\underline{c}}}\sigma^{-N_{apr}(\delta)p}\|T(x_{N(\delta)} - x_{N_{apr}(\delta)})\| + \|x_{N_{apr}(\delta)} - x^\dagger\| \\
&\leq \tfrac{1}{\underline{\underline{c}}}\sigma^{-N_{apr}(\delta)p}\Big(\tfrac{1}{1-c_{tc}}\|F(x_{N(\delta)}) - y\| + \|T(x_{N_{apr}(\delta)} - x^\dagger)\|\Big) + \|x_{N_{apr}(\delta)} - x^\dagger\| \\
&\leq C(\textstyle\sum_{k=0}^{N_{apr}(\delta)+1}(\tfrac{\epsilon}{\sigma^p})^k\beta_{N_{apr}(\delta)-k} + \sigma^{-N_{apr}(\delta)p}\delta) \\
&\leq C\sigma^{-N_{apr}(\delta)p}\delta \qquad (79)\\
&\leq C\textstyle\sum_{k=0}^{N_{apr}(\delta)}(\tfrac{\epsilon}{\sigma^p})^k\beta_{N_{apr}(\delta)-1-k}\ , \qquad (80)
\end{aligned}$$

for some constant $C > 0$, where we have used (3) in the third and (2), (74) in the fourth inequality, and inserted (78) to obtain (79); inequality (80) follows from the fact that, by (78),

$$\tau\delta < \bar{C}\sigma^{(N_{apr}(\delta)-1)p}\sum_{k=0}^{N_{apr}(\delta)}(\frac{\epsilon}{\sigma^p})^k\beta_{N_{apr}(\delta)-1-k}\ . \qquad (81)$$

Going on from these estimates our convergence and convergence rates proof can be completed by standard arguments (cf., e.g., [10]):

Considering first the general situation without source condition, we fix an arbitrary sequence $(y_i^\delta)_{i\in\mathbb{N}}$ satisfying (2) with $\delta := \delta_i \to 0$ as $i \to \infty$, and set $\tilde{x}^i := \tilde{x}_{N(\delta_i)}$ according to Algorithm 2 with (74), which we have to show to converge to $x^\dagger$ as $i \to \infty$. To do so, we prove that any subsequence $\tilde{x}^l$ of $\tilde{x}^i$ has a subsequence converging to $x^\dagger$: Namely, in case $N_{apr}(\delta) \to \infty$ as $l \to \infty$, this follows from (80), (77); otherwise, there exists a finite $\bar{N} \in \mathbb{N}$ and a subsequence $\tilde{x}^m$ of $\tilde{x}^l$, such that for all $m \in \mathbb{N}$: $N_{apr}(\delta_m) = \bar{N}$ and therefore, by (79),

$$\|x^m - x^\dagger\| \leq C\sigma^{-\bar{N}}\delta_m \to 0 \text{ as } m \to \infty\ .$$

Under a source condition (32) with $0 < \mu \leq \frac{1}{2}$ we have, by (21),

$$\beta_k = O(\sigma^{2\mu pk})\ ,$$

and therefore, by (75) and (81)

$$\tau\delta < C\sigma^{(1+2\mu)p(N_{apr}(\delta)-1)}\ ,$$

so that (79) implies

$$\|x_{N(\delta)} - x^\dagger\| \leq C\sigma^{-N_{apr}(\delta)p}\delta \leq C\delta^{-\frac{1}{(1+2\mu)}}\delta\ ,$$

hence the assertion. □

Remark 3. The reason why in the case of nonlinearity condition (4) a convergence proof with a discrepancy principle as the one above does not seem to go through is, that we do not have such an estimate as (76), that could provide us with an upper bound on $\sigma^{-(N(\delta)-1)p}\delta$, in that situation.

Remark 4. The exact projection $\mathcal{Q}_n$ in line 8 of Algorithm 1 and lines 5, 6 of Algorithm 2, that would need inversion of the Gramian matrix corresponding to a basis of $\mathcal{Y}_n$, could as usual be replaced by some sufficiently good approximation based on a discrete inner product on $\mathcal{Y}_n$. The same holds true for the (here invisible) prolongation on line 15 of Algorithm 1 and line 4 of Algorithm 2.

5 Numerical Experiments

For a numerical verification of our theoretical results we implemented the proposed multigrid nested iteration 2 in a Matlab program and applied it to a model problem, namely the Hammerstein integral equation

$$\int_0^t (x(s) + s\frac{x^3(s)}{3})\ ds = y(t)\ , \quad t \in (0,1)\ . \tag{82}$$

Here F is the Volterra Integral operator

$$\begin{array}{rcl} F: \ \mathcal{X} & \to & \mathcal{Y} \\ x & \mapsto & \int_0^\cdot (x(s) + s\frac{x^3(s)}{3})\ ds\ , \end{array}$$

where the Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$ are both equal to the space of square integrable functions on the unit interval:

$$\mathcal{X} = \mathcal{Y} = L^2(0,1)\ .$$

With $(R(x,\bar{x})v)(s) := (1 + s(x^2(s) + x(s)\bar{x}(s) + \bar{x}^2(s))v(s)$, $(Tv)(t) := \int_0^t v(s)\ ds$, F satisfies the nonlinearity condition (4). A source condition with $\mu = \frac{1}{2}$, i.e., $x^\dagger - x_{ig} = T^*\tilde{w}$ for some $\tilde{w} \in L^2(0,1)$ is here equivalent to

$$x_{ig} - x^\dagger \in H^1(0,1)\ \wedge\ x_{ig}(1) = x^\dagger(1)\ ; \tag{83}$$

the smoothing quantifier p in Assumptions 1, 2, 3 is obviously equal to one. We project onto the spaces $\mathcal{Y}_k$ of piecewise linear splines on an equidistant grid of size $h_k = 2^{-k}$ with the hat functions ψ_j^k as a basis. As a starting value x_{ig} we just use the constant zero function; the solution of the nonlinear problem on the coarsest grid is done by a frozen Newton method.

The computations were done with two different functions y, corresponding to an exact solution satisfying no source condition:

$$x^\dagger(s) = \cos(4\pi s)\ , \quad s \in (0,1)\ , \tag{84}$$

and a source condition with $\mu = \frac{1}{2}$:

$$x^\dagger(s) = \sin(4\pi s)\ , \quad s \in (0,1)\ , \tag{85}$$

respectively.

The respective results of computations with exact data and $n = 128$ are shown in Table 1 and Figure 1; as one would expect, the source condition (83) especially improves the behavior at $s = 1$. The total elapsed CPU times are given in the first line of Table 2.

To compare the efficiency of the nonlinear multi-grid method, we alternatively applied a Newton iteration to the equation projected on the finest grid, i.e., to (28), and stopped it as soon as a residual reduction comparable to the one reported in Table 1 was achieved. As the CPU times in Table 2 show, the multi-grid approach seems to save a considerable amount of computation time already for a relatively small number of unknowns.

Table 1

	example (84)		example (85)	
$dim(k)$	$\|F(\tilde{x}_k) - y^\delta\|$	$\|\tilde{x}_k - x^\dagger\|$	$\|F(\tilde{x}_k) - y^\delta\|$	$\|\tilde{x}_k - x^\dagger\|$
16	2.6422e-03	1.0239e-01	1.0030e-03	2.907e-02
32	1.7789e-03	7.2333e-02	2.0294e-04	9.706e-03
64	1.2470e-03	5.2027e-02	6.7656e-05	3.430e-03
128	1.0136e-03	3.7577e-03	3.1095e-05	1.269e-03

Table 2

	example (84)	example (85)
CPU time MGM	59.52 sec	60.56 sec
CPU time Newton	156.29 sec	476.23 sec

Finally we tested the behavior of the nonlinear multi-grid method in the presence of data noise with a priori chosen level according to (72), (73). In five test runs each, we added 0.1, 1, and 10 per cent random noise to the exact data and plotted the resulting error norms $\|\tilde{x}_N - x^\dagger\|$ in Figure 2 and the left plot of Figure 3. To see also the predicted convergence rate $O(\sqrt{\delta})$ for test example (85), we additionally plotted the quotients $\frac{\|\tilde{x}_N - x^\dagger\|}{\sqrt{\delta}}$ in the right plot of Figure 3.

Acknowledgement

The author wishes to thank Dr. Schicho from the Research Institute for Symbolic Computation for proving Lemma 1 and therewith simplifying the proofs of Theorems 1 and 2. Moreover we wish to acknowledge support by the Fonds zur Förderung der wissenschaftlichen Forschung under grant T 7-TEC and project F1308 within the Spezialforschungsbereich "Numerical and Symbolic Scientific Computing".

References

[1] D. Braess, W. Hackbusch, *A new convergence proof for the multi-grid method including the V-cycle*, SIAM J. Numer. Anal. 20 (1983), 967-975.

[2] J. Bramble, *Multigrid methods*, Pitman Research Notes in Mathematics Series 294, Longman Scientific& Technical, Harlow, UK, 1993.

[3] P. Deuflhard, H.W. Engl, O. Scherzer, *A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinely invariant conditions*, Inverse Problems 14 (1998), 1081-1106.

[4] 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.

[5] H.W. Engl, A. Neubauer, *On projection methods for solving linear ill–posed problems*, in: A. Vogel, ed., Model Optimization in Exploration Geophysics, Vieweg, Braunschweig, 1987, 73–92.

[6] H.W. Engl, M. Hanke, A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996

[7] C.W. Groetsch, *Inverse Problems in Mathematical Sciences*, Vieweg, Braunschweig, 1993.

[8] C.W. Groetsch, A. Neubauer, *Convergence of a general projection method for an operator equation of the first kind*, Houston J. Mathem. 14 (1988), 201–207.

[9] W. Hackbusch, Multi-Grid Methods and Applications, Springer, Berlin, 1985.

[10] 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.

[11] M. Hanke, C.R. Vogel, *Two-level preconditioners for regularized inverse problems I: Theory*, Numer. Math. 83 (1999), 385-402.

[12] sc B. Blaschke–Kaltenbacher, H.W. Engl, *Regularization methods for nonlinear ill–posed problems with applications to phase reconstruction*, H.W. Engl, A.K. Louis, W.Rundell, eds., Inverse Problems in Medical Imaging and Nondestructive Testing, Springer, Vienna–New York, 1997.

[13] B. Kaltenbacher, *On Broyden's method for nonlinear ill–posed problems*, Numerical Functional Analysis and Optimization 19 (1998), 807-833.

[14] B. Kaltenbacher, *A projection-regularized Newton method for nonlinear ill-posed problems with application to parameter identification problems with finite element discretization*, to appear in SIAM J.Numer.Anal.

[15] B. Kaltenbacher, *On a multigrid method for ill-posed problems*, SFB013-report No.99-20, University of Linz, 1999.

[16] B. Kaltenbacher, *Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems*, SFB013-report No.00-??, University of Linz, 2000.

[17] J.T. King, *Multilevel algorithms for ill-posed problems*, Numer. Math. 61 (1992), 311–334.

[18] A.K. Louis, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.

[19] V.A. Morozov, *Regularization Methods for Ill-Posed Problems*, CRC Press, Boca Raton, 1993.

[20] F. Natterer, *Regularisierung schecht gestellter Probleme durch Projektionsverfahren*, Numer. Math. 28 (1977), 329–341.

[21] F. Natterer, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, 1986.

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

[23] O. Scherzer, *Convergence criteria of iterative methods based on Landweber iteration for nonlinear problems*, J. Math. Anal. Appl. 194 (1995), 911-933.

[24] A.N. Tikhonov, V.A. Arsenin, *Methods for Solving Ill-Posed Problems*, Nauka, Moskau, 1979.

[25] G.M. Vainikko, A.Y. Veterennikov, *Iteration Procedures in Ill–Posed Problems*, Nauka, Moscow, 1986. In Russian.

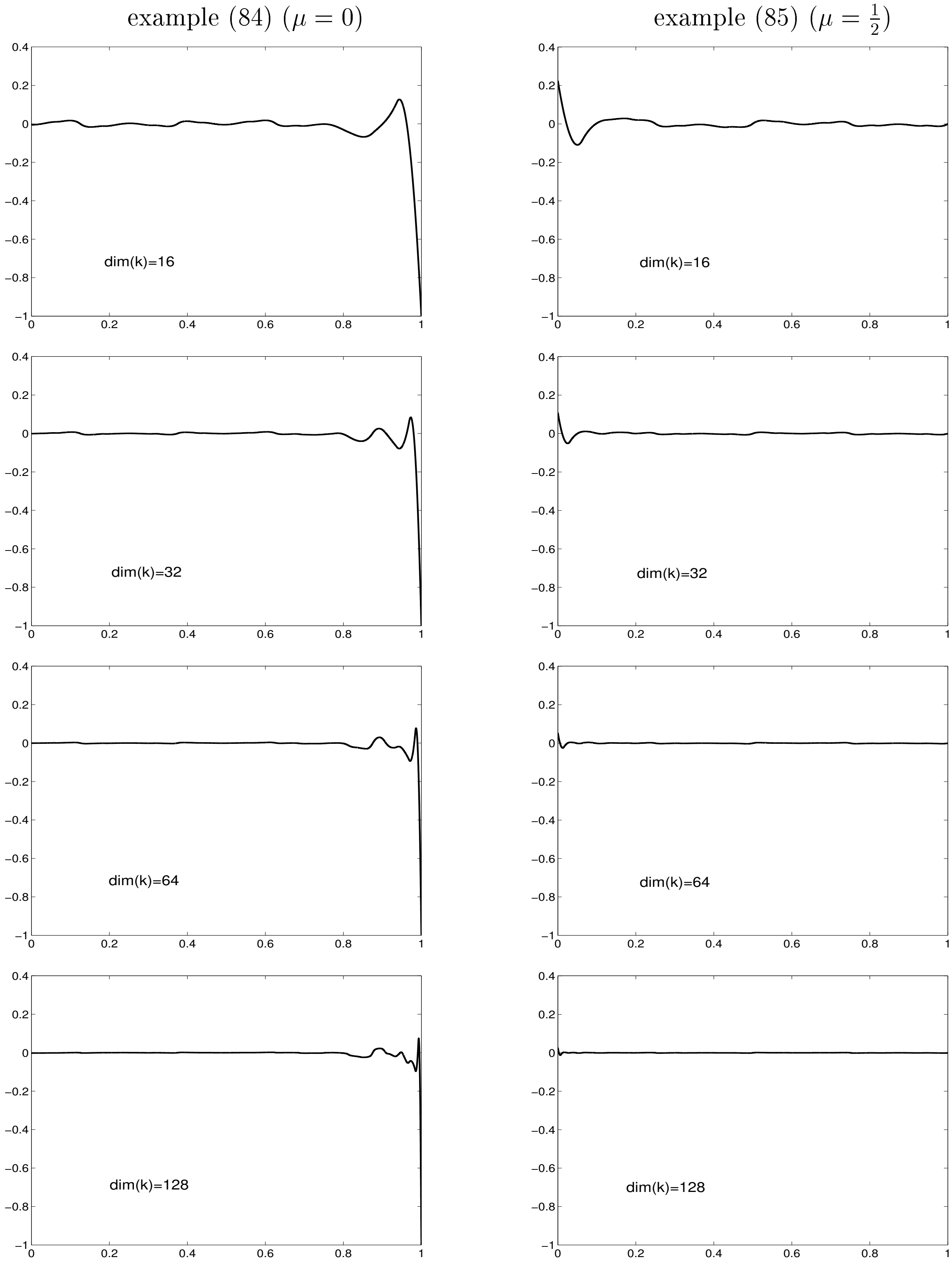

Figure 1: Error $\tilde{x}_k - x^\dagger$ for $k = 16, 32, 64$ and 128 in the noiseless case $\delta = 0$.

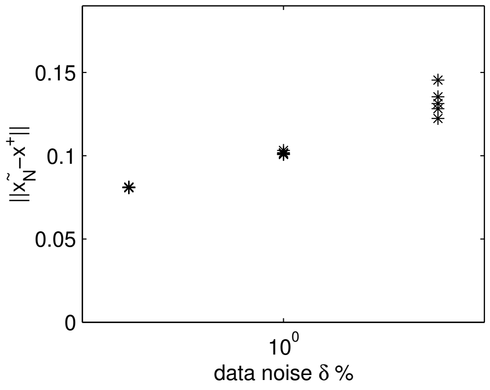

Figure 2: Error norm for $\frac{\delta}{\|y\|} * 100 = 0.1$, 1, and 10 per cent data noise in example (84) ($\mu = 0$)

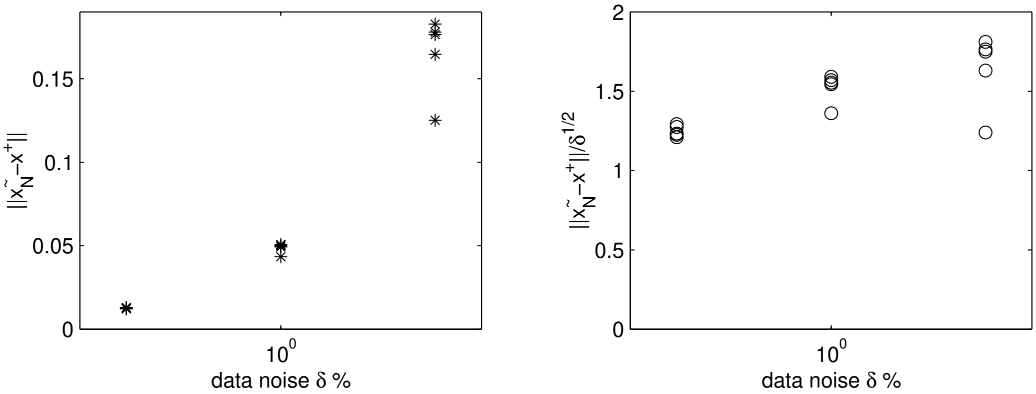

Figure 3: Error norm and quotient $\frac{\|\tilde{x}_N - x^\dagger\|}{\sqrt{\delta}}$ for $\frac{\delta}{\|y\|} * 100 = 0.1$, 1, and 10 per cent data noise in example (85) ($\mu = \frac{1}{2}$)