V-cycle convergence of some multigrid methods for ill-posed problems [1]

Barbara Kaltenbacher

Abstract. For ill-posed linear operator equations we consider some V-cycle multigrid approaches, that, in the framework of Bramble, Pasciak, Wang, and Xu (1991), we prove to yield level independent contraction factor estimates. Consequently, we can incorporate these multigrid operators in a full multigrid method, that, together with a discrepancy principle, is shown to act as an iterative regularization method for the underlying infinite dimensional ill-posed problem. Numerical experiments illustrate the theoretical results.

Key words. ill-posed problem, multigrid methods

AMS(MOS) Subject Classifications: 65J20, 65R30, 65N55

1 Introduction

Consider the first kind operator equation

$$\mathcal{T}x = y \tag{1}$$

where $\mathcal{T} : \mathcal{X} \to \mathcal{Y}$ is a compact linear operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$, and $y \in \mathcal{R}(\mathcal{T})(+\mathcal{N}(\mathcal{T})^\perp)$, which we suppose to model some linear or linearized inverse problem. Numerous applications leading to inverse problems appear in science and industry; think, e.g., of the wide field of parameter identification. Their mathematical formulation (1) typically leads to an ill-posed problem in the sense that the range of $\mathcal{T}$ is non-closed, i.e., the (generalized) inverse of $\mathcal{T}$ is unbounded. Therefore, given data y^δ with arbitrary small noise δ in

$$\|y - y^\delta\| \le \delta \ , \tag{2}$$

one can possibly arrive at large deviations in the solution when using conventional numerical methods. Therefore special stable approximation methods for solving (1) — so-called regularization methods (cf. [10], [11], [20], [21], [23], [26], [28]) — have to be applied, the most well-known one certainly being Tikhonov regularization:

$$\tilde{x} := (\mathcal{T}^*\mathcal{T} + \alpha I)^{-1}\mathcal{T}^* y^\delta \ . \tag{3}$$

As a consequence of the ill-posedness, any numerical approximation method for (1) converges — if at all — 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

$$x^\dagger \in \mathcal{R}((\mathcal{T}^*\mathcal{T})^\mu) \tag{4}$$

[1] Supported by the Fonds zur Förderung der wissenschaftlichen Forschung under grant T 7-TEC and project F1308 within Spezialforschungsbereich F013

where $\mu > 0$ is a real exponent and the operator $(\mathcal{T}^*\mathcal{T})^\mu$ is defined in the sense of functional calculus (cf., e.g., Section 2.3 in [10]). The under condition (4) optimal convergence rates for regularized approximations $\tilde{x}$ to $x^\dagger$ are

$$\|\tilde{x} - x^\dagger\| = O(\delta^{\frac{2\mu}{2\mu+1}}) \tag{5}$$

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

While multigrid 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], [4], [6], [7], [13]), the situation is quite different for first kind integral equations, or, generally, ill-posed problems (1), though. 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.

Applications of MGM to (Tikhonov-)regularized ill-posed problems can be found in [14], [25]. While these papers analyze MGM for equations with small but nonvanishing regularization parameter $\alpha > 0$ in (3), we are more interested in the situation that the regularization is *solely due to discreteness* (see [9, 12, 15, 17, 22, 27]), so without introducing any additional (possibly artificial) regularization term. For this purpose, MGM based on a smoother proposed by King in [19] have turned out to be appropriate. Further contraction number estimates for these MGMs as well as an analysis of the full MGM with a priori or a posteriori stopping rule as a regularization method for solving the original infinite-dimensional ill-posed problem (1), including a generalization to nonlinear problems, can be found in [16] and [18]. Since the theory presented there still needs at least alternating or W-cycles, the aim of the present paper is to give V-cycle convergence proofs of MGM for ill-posed problems. To this end, we follow the approach of [3], where a V-cycle convergent multigrid method for pseudodifferential operators of order minus one is presented and analyzed, and show how it can be used for defining and analyzing V-cycle MG algorithms for general equations (1). Here norm equivalence theorems such as those from [24], with their computational implementation as given, e.g., in [5], play an important role. Based on these norm estimates, also the smoother by King can be modified in such a way that V-cycle convergence can be proved, see Section 3.

The level-independence of the contraction factors in the proposed multigrid operators makes it possible to use them in a full multigrid method for the iterative solution of the underlying infinite dimensional problem (1). In combination with an a posteriori stopping rule that finds the optimal balance between approximation error and propagated data noise, this can be shown to define a convergent and order optimal regularization method for the ill-posed operator equation (1), see Section 4.

Section 5 reports on numerical experiments with the proposed methods and is supposed to 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" and can have different values whenever they appear.

2 Discretization

Discretization of the ill-posed problem (1)

Finite-dimensional problems, though they might be ill-conditioned, are always well-posed in the sense of stable data dependence of a solution (as long as its existence and uniqueness can be guaranteed, which can be done, e.g., by using a best approximate solution concept). This fact forms the basis for several finite dimensional projection approaches for the regularization of ill-posed operator equations (1). Note, however, that for ill-posed problems, discretization in pre-image space does not generically yield a convergent approximation, even in the noiseless situation (cf. the counterexample due to Seidman cited in [10]). Therefore we here prefer the approach of discretization in image space ([9, 12, 22, 27]), defined by projecting (1), with possibly noisy data y^δ satisfying (2), onto finite dimensional nested subspaces $\mathcal{Y}_n$ of $\mathcal{Y}$, whose union is dense in $\mathcal{Y}$:

$$\mathcal{Y}_1 \subseteq \mathcal{Y}_2 \subseteq \ldots \subseteq \mathcal{Y} \quad \wedge \quad \overline{\bigcup_{n\in\mathbb{N}} \mathcal{Y}_n} = \mathcal{Y} \ . \tag{6}$$

The best approximate solution of this projected equation lies in the finite dimensional space

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

so with

$$\mathcal{P}_n := Proj_{\mathcal{X}_n}^{\mathcal{X}} \ , \quad \mathcal{Q}_n := Proj_{\mathcal{Y}_n}^{\mathcal{Y}} \ ,$$

we arrive at an appropriate discretization of (1)

$$\mathcal{Q}_n\mathcal{T}\mathcal{P}_n x_n = \mathcal{Q}_n y^\delta \ , \tag{7}$$

or, equivalently

$$\begin{array}{c} x_n = \mathcal{T}^*u \ , \quad u \in \mathcal{Y}_n \\ \langle \mathcal{T}^*u, \mathcal{T}^*v\rangle_{\mathcal{X}} = \langle y^\delta, v\rangle_{\mathcal{Y}} \ , \quad \text{for all } v \in \mathcal{Y}_n \ . \end{array} \tag{8}$$

Indeed, the exact solution of (7) can be seen as an approximation of the exact bestapproximate solution $x^\dagger = \mathcal{T}^\dagger y$ of (1) since, in the noiseless case $\delta = 0$, we have

$$x_n = \mathcal{P}_n x^\dagger \ .$$

Nonvanishing data noise is in the worst case amplified by the factor $\underline{\gamma}_n^{-1}$, where $\underline{\gamma}_n$ is the smallest sigular value of the system matrix in (7), i.e.,

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

that, by the compactness of $\mathcal{T}$, goes to infinity as $n \to \infty$, i.e., with increasing refinement of the discretization we have to face a possibly unboundedly growing propagated data noise contribution to the total error — this reflects the instability of the underlying infinite-dimensional problem. Consequently, an optimal choice $n := N(\delta, y^\delta)$ of the discretization level has to balance between two error terms of different asymptotic behaviour: the approximation error, that goes to zero, and the propagated data noise that in the worst case goes to infinity as $n \to \infty$. This corresponds to the necessity of correctly choosing a regularization parameter (e.g., α in (3)) in regularization methods for ill-posed problems.

Note that the residual $\|\mathcal{T} x_n - y\|$ can, up to an $O(\delta)$ rest term, obviously be estimated by

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

and that, under a source condition (4) the approximation part of the error (i.e., the total error in the noiseless case) goes to zero at some rate

$$\overline{\gamma}_n(\mu) := \|(I - \mathcal{P}_n)(\mathcal{T}^*\mathcal{T})^\mu\| \leq \overline{\overline{\gamma}}_n^{2\mu} \ , \tag{9}$$

for $\mu \leq \frac{1}{2}$ (cf., e.g., [18]). Moreover,

$$\underline{\gamma}_n \leq \overline{\overline{\gamma}}_{n-1} \ , \tag{10}$$

and, in order to obtain optimal convergence rates (5) of $\tilde{x} := x_{N(\delta, y^\delta)}$ under source conditions (4), one also needs

$$\overline{\overline{\gamma}}_{n-1} \leq C \underline{\gamma}_n \tag{11}$$

for some constant $C > 0$, which is basically a condition on the choice of the approximation spaces $\mathcal{Y}_n$ (see [19], [16], [17]). It is satisfied, e.g., by truncated SVD, which means that the finite dimensional spaces $\mathcal{Y}_n$ are spanned by eigenvectors of $\mathcal{T}\mathcal{T}^*$, and which is optimal both with respect to stability and convergence, (see, e.g., Section 3.3 in [10], where a more detailed exposition on regularization by projection can be found). For reasons of practical applicability in cases where an explicite SVD is hard or impossible to compute, we here concentrate on projection onto spaces of piecewise polynomial functions such as splines, finite elements, or wavelets, though.

In this paper, like in [16], [18], we consider instead of the exact solution x_n of (7), its approximation $\tilde{x}_n$ by (iterative) multigrid techniques — again with $n := N(\delta, y^\delta)$ to be appropriately chosen — as our regularized approximate solution of the infinite dimensional problem (1).

The asymptotic behaviour of the condition of our finite dimensional system (7) is characterized by the real sequence $\underline{\gamma}_n$: while the largest singular value of $\mathcal{Q}_n \mathcal{T} \mathcal{P}_n$ is uniformly bounded by the norm of $\mathcal{T}$, its smallest singular value $\underline{\gamma}_n$ goes to zero as $n \to \infty$. This is due to the ill-posedness of (1) or, in other words, the smoothing property of the forward operator $\mathcal{T}$ that, in combination with the approximation property of the finite dimensional spaces $\mathcal{Y}_n$, makes $\overline{\overline{\gamma}}_n$ go to zero as $n \to \infty$, on the other hand.

To concretize the asymptotics, we denote by h_n some discretization parameter (think, e.g., of a mesh size) that geometrically goes to zero as $n \to \infty$:

$$h_n = C_\sigma \sigma^n \tag{12}$$

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

$$\mathcal{Y}^{-p} \supset \mathcal{Y} ,$$

i.e., we think of $\mathcal{Y}^{-p}$ as a less smooth function space than $\mathcal{Y}$, e.g., $\mathcal{Y} = H^t(\Omega)$, $\mathcal{Y}^{-p} = H^{t-p}(\Omega)$, for some $t \in \mathbb{R}$, $p > 0$, and some domain Ω. The real positive number p is supposed to quantify the degree of smooting by $\mathcal{T}$, which is closely related to the concept of degree of ill-posedness as it is frequently used in the literature on ill-posed problems.

Assumption 1. (Smoothing property of $\mathcal{T}^*$:)

$$\forall v \in \mathcal{Y} \; : \quad \|\mathcal{T}^* v\|_{\mathcal{X}} \simeq \|v\|_{\mathcal{Y}^{-p}} \; .$$

For the pseudodifferential operators of order minus one considered in [5] one has $p = 1/2$; below we will consider a parameter identification problem for an elliptic boundary value problem, where $p = 2$.

In order to be able to exploit this smoothing assumption we combine it with a condition on the quality of approximation by functions in $\mathcal{Y}_n$:

Assumption 2. (Approximation property of $\mathcal{Y}_n$)

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

These assumptions (having in mind Sobolev spaces and piecewise polynomial approximation) yield the following crucial estimate

$$\overline{\overline{\gamma}}_n \leq \overline{\overline{C}} h_n^p \; , \tag{13}$$

(and, by (9), $\overline{\gamma}_n(\mu) = O(h_n^{2p\mu})$).

On the other hand, estimating the lowest singular value of the system matrix in (7) corresponds to assuming some kind of inverse inequality on $\mathcal{Y}_n$:

Assumption 3. (Inverse Inequality)

$$\sup_{v \in \mathcal{Y}_n} \frac{\|v\|_{\mathcal{Y}}}{\|v\|_{\mathcal{Y}^{-p}}} \leq C h_n^{-p}$$

which, when thinking of the $\mathcal{Y}_n$ as piecewise polynomial spaces, seems to be realistic, and from which we get

$$\underline{\gamma}_n \geq \underline{\underline{c}} h_n^p \; , \tag{14}$$

(cf. [17]). Note that this implies uniform boundedness of the quotient $\overline{\overline{\gamma}}_n / \underline{\underline{\gamma}}_n$, (i.e., (11),) as well as of the operators $\mathcal{T}(\mathcal{Q}_n \mathcal{T})^\dagger \mathcal{Q}_n$ approximating the projection $\mathcal{T}\mathcal{T}^\dagger = Proj^{\mathcal{Y}}_{\mathcal{R}(\mathcal{T})}|_{\mathcal{D}(\mathcal{T}^\dagger)}$ (cf. [18]).

Discrete norms

The discrete inner products used in the respective multigrid approaches below, will be based on norm equivalences of the form

Assumption 4. (norm equivalence)

$$\forall v \in \mathcal{Y}_n \ : \quad \|v\|^2_{\mathcal{Y}'} \simeq \sum_{j=1}^{n} h_j^{-2q} \|(\mathcal{Q}_j - \mathcal{Q}_{j-1})v\|^2_{\mathcal{Y}} + \|\mathcal{Q}_0 v\|^2_{\mathcal{Y}} \ , \tag{15}$$

where $\mathcal{Y}'$ is some Hilbert space containing $\mathcal{Y}_n$, q is an index that will be specified below, and the equivalence constants are supposed to be independent of n. In the case of $\mathcal{Y}'$ being a Sobolev space on a (sufficiently smooth) domain Ω, $\mathcal{Y}$ the space of square integrable functions on Ω,

$$\mathcal{Y} = L^2(\Omega) \ , \quad \mathcal{Y}' = H^q(\Omega) \ ,$$

(as typical, e.g., in parameter identification) and the $\mathcal{Y}_n$ finite element spaces of piecewise polynomial C^r functions on a regular and quasi-uniform triangulation, this norm equivalence follows for

$$-r - 3/2 \leq q \leq r + 3/2 \tag{16}$$

from Theorem 15 in Oswald [24]. The exact implementation of projections $\mathcal{Q}_j$ involves inversion of Gramian matrices with respect to bases of $\mathcal{Y}_n$. Note that the system matrices in (7) are typically full so that the inversion of a tridiagonal matrix asymptotically does not count as compared to, e.g., a system matrix $\times$ vector multiplication. Nevertheless it is advisable to save computational effort by using efficient numerical approximations $\tilde{\mathcal{Q}}_j$ of the projections $\mathcal{Q}_j$, that avoid Gram matrix inversions. Bramble, Pasciak, and Vassilevski [5] provide an abstract norm equivalence theorem allowing for approximation operators $\tilde{\mathcal{Q}}_j$ to $\mathcal{Q}_j$ in (15). Moreover, a concrete choice of $\tilde{\mathcal{Q}}_j$, based on nodal bases and local L^2-projections is given, that satisfies the assumptions of this theorem in the context of piecewise linear finite elements ($r = 0$), is generalizable to higher order finite elements, and therefore can be used for an efficient inner product computation. For other implementations $\tilde{\mathcal{Q}}_j$ of $\mathcal{Q}_j$, that are based on wavelet-like space decompositions, see the references in [5].

3 Two V-cycle-convergent multigrid approaches

Before proposing concrete multigrid approaches for (1) in Subsections 3.1, 3.2, we introduce a general notational framework together with a fundamental convergence assertion, following [5].

Going out from a finite-dimensional variational equation of finding $u \in \mathcal{Y}_n$ such that

$$A(u, v) = f(v) \text{ for all } v \in \mathcal{Y}_n \ ,$$

where $A(.,.)$ is a positive definite symmetric bilinear form on $\mathcal{Y}_n$, inducing a norm

$$|||u||| = \sqrt{A(u,u)} \ ,$$

and projection operators $P_k : \mathcal{Y}_n \to \mathcal{Y}_k$, $k = 0, \ldots, n$,

$$A(P_k w, v) = A(w, v) \quad \text{for all } v \in \mathcal{Y}_k \ ,$$

and f a linear functional on $\mathcal{Y}_n$, one can, via an additional inner product $\langle ., . \rangle$ on $\mathcal{Y}_n$ with corresponding projectors $Q_k : \mathcal{Y}_n \to \mathcal{Y}_k$

$$\langle Q_k w, v \rangle = \langle w, v \rangle \quad \text{for all } v \in \mathcal{Y}_k \ ,$$

and operators $A_k : \mathcal{Y}_k \to \mathcal{Y}_k$ given by

$$\langle A_k w, v \rangle = A(w, v) \text{ for all } v \in \mathcal{Y}_n \ ,$$

(which are spsd. with respect to both inner products $A(., .)$ and $\langle ., . \rangle$), and smoothing operators $S_k : \mathcal{Y}_k \to \mathcal{Y}_k$, inductively define a symmetric V-cycle multigrid operator $B_k^s : \mathcal{Y}_k \to \mathcal{Y}_k$ by:

Algorithm 1. (SYMMETRIC MG)
SET $B_0^s := (A_0)^{-1}$. ASSUME THAT B_{k-1}^s HAS BEEN DEFINED AND DEFINE $B_k^s b$ FOR $b \in \mathcal{Y}_k$ AS FOLLOWS:
SET $u := S_k b$.
SET $u := u + B_{k-1}^s Q_{k-1}(b - A_k u)$.
SET $u := u + S_k(b - A_k u)$.
SET $B_k^s b := u$.

A non-symmetric V-cycle MG operator B_k^n can be defined analogously by omitting the post-smoothing step $u := u + S_k(b - A_k u)$ in Algorithm 1.

If, on one hand

$$S_k A_k \text{ and } I - S_k A_k \text{ are spsd. with respect to } A(., .) \ , \tag{17}$$

and, on the other hand,

$$|||(P_k - P_{k-1})v||| \leq C \ A(S_k A_k v, v) \quad \text{for all } v \in \mathcal{Y}_k \ , \tag{18}$$

Theorem 1 in [6] yields a level independent contraction number estimate, whose proof in our special situation which is quite short and will therefore be given explicitely for convenience of the reader and since it will be referred to in a remark below.

Corollary 1. *(to* **Theorem 1** *in [6])*
Assume that (17) and (18) hold and define B_n^s and B_n^n by algorithm 1 (omitting the post-smoothing step in the second case). Then

$$A((I - B_n^s A_n)u, u) = |||(I - B_n^n A_n)u|||^2 \leq (1 - \frac{1}{C^2})|||u|||^2 \textit{ for all } u \in \mathcal{Y}_n \tag{19}$$

Remark 1. The constant C in the right hand side of (19) is equal the one in (18), which by (17) must be larger or equal to one.

Proof. It can be easily checked that the sequence of operators $E_k : \mathcal{Y}_n \to \mathcal{Y}_k$ defined by

$$\begin{array}{l} E_{-1} := I \\ E_k := (I - B_k^n A_k P_k)^* \quad k = 0, 1, \dots \ , \end{array}$$

(where the adjoint * is taken wrt. $A(.,.),$) obeys the recursion

$$E_k = (I - S_k A_k P_k) E_{k-1} \quad k = 0, 1, \dots \ ,$$

where we have set $S_0 := A_0^{-1}$ for convenience of notation. Therefore, and since by (17) $S_k A_k - (S_k A_k)^2$ is spsd., we have, for any $u \in \mathcal{Y}_n$,

$$\begin{array}{rcl} |||u|||^2 - |||(I - B_n^n A_n)^* u|||^2 & = & \sum_{k=0}^n |||E_{k-1} u|||^2 - |||E_k u|||^2 \\ & = & \sum_{k=0}^n \Big(A((S_k A_k - (S_k A_k)^2) P_k E_{k-1} u, P_k E_{k-1} u) \\ & & \qquad + A(S_k A_k P_k E_{k-1} u, P_k E_{k-1} u) \Big) \\ & \geq & \sum_{k=0}^n A(S_k A_k P_k E_{k-1} u, P_k E_{k-1} u) \ , \end{array}$$

so that it only remains to show that the latter is greater or equal to $1/C^2 |||u|||^2$. To do so we decompose

$$u = \sum_{k=1}^n (P_k - P_{k-1}) u + P_0 u$$

and get, since $(I - E_{k-1}) u = (B_{k-1}^n A_{k-1} P_{k-1})^* u \in \mathcal{Y}_{k-1}$ is orthogonal to $(P_k - P_{k-1}) u = (I - P_{k-1}) P_{k-1} u$, that

$$\begin{array}{rcl} |||u|||^2 & = & \sum_{k=0}^n A(E_{k-1} u + (I - E_{k-1}) u, (P_k - P_{k-1}) u) + |||P_0 u|||^2 \\ & = & \sum_{k=1}^n A(E_{k-1} u, (P_k - P_{k-1}) u) + |||P_0 u|||^2 \\ & \leq & \sum_{k=1}^n |||(P_k - P_{k-1}) E_{k-1} u|||\ |||(P_k - P_{k-1}) u||| + |||P_0 u|||^2 \\ & \leq & \sum_{k=1}^n \sqrt{C\ A(S_k A_k P_k E_{k-1} u, P_k E_{k-1} u)}\ |||(P_k - P_{k-1}) u||| \\ & & + \sqrt{A(S_0 A_0 P_0 E_{-1} u, P_0 E_{-1} u)}\ |||P_0 u||| \\ & \leq & \sqrt{\max\{C, 1\} \sum_{k=0}^n A(S_k A_k P_k E_{k-1} u, P_k E_{k-1} u)}\ |||u||| \ , \end{array} \tag{20}$$

where we have used (18) with $v := P_k E_{k-1} u$ and the fact that we have set $S_0 A_0 P_0 = P_0$ in the fourth line, and Cauchy-Schwarz in $\mathbb{R}^n$ yielded the last inequality. $\square$

In our situation, the bilinear form A and the projections P_k are, due to (8), given by

$$A(w, v) := \langle \mathcal{T}^* w, \mathcal{T}^* v \rangle_{\mathcal{X}} \ , \quad P_k := (\mathcal{T}^*)^{-1} \mathcal{P}_k \mathcal{T}^* \tag{21}$$

and the right hand side f by $\langle y^\delta, . \rangle_{\mathcal{Y}}$.

The behaviour of the multigrid method obviously heavily depends on the choice of the smoothers S_k and the inner products $\langle ., . \rangle$. In the following we describe two approaches that will lead to level independent contraction factor estimates according to Corollary (1).

3.1 Generalization of Bramble, Pasciak, Vassilievski (2000)

Here, $\langle .,. \rangle$ is chosen to induce a norm weaker than $|||.|||$, namely, in our context,

$$\langle w, v\rangle = \langle w, v\rangle_- = \langle \mathcal{T}\mathcal{T}^* w, \mathcal{T}\mathcal{T}^* v\rangle_{\mathcal{Y}} \ .$$

Its discrete implementation $\langle .,. \rangle_{-,k}$ is based on the norm equivalence in Assumption 4 with

$$q = -p \ , \quad \mathcal{Y}' = \mathcal{Y}^{-p} \ , \tag{22}$$

and the additional assumptions

$$\forall v \in \mathcal{Y} \ : \quad \|v\|_- = \|\mathcal{T}\mathcal{T}^* v\| \geq c\|v\|_{\mathcal{Y}^{-2p}} \tag{23}$$

$$\forall v \in \mathcal{Y}_k \ : \quad \|v\|^2_{\mathcal{Y}^{-2p}} \geq c\Big(\sum_{j=1}^{k} h_j^{4p} \|(\mathcal{Q}_j - \mathcal{Q}_{j-1})v\|^2_{\mathcal{Y}} + \|\mathcal{Q}_0 v\|^2_{\mathcal{Y}}\Big) \tag{24}$$

where $\mathcal{Y}^{-2p}$ is, as indicated by the superscript, a function space with weaker topology than $\mathcal{Y}$ and $\mathcal{Y}^{-p}$:

$$\mathcal{Y}^{-2p} \supset \mathcal{Y}^{-p} \supset \mathcal{Y} \ .$$

With the $\langle .,. \rangle_{\mathcal{Y}}$-spsd. difference operator $D_k : \mathcal{Y}_k \to \mathcal{Y}_k$ defined by

$$D_k v = \sum_{j=1}^{k} h_j^{-4p}(\mathcal{Q}_j - \mathcal{Q}_{j-1})v + \mathcal{Q}_0 v \ ,$$

and the discrete inner product $\langle .,. \rangle_{-,k}$ given by

$$\langle v, w\rangle_{-,k} = \langle D_k^{-1} v, w\rangle_{\mathcal{Y}} \ ,$$

we get, with some constant C_- independent of k:

$$\forall v \in \mathcal{Y}_k \ : \quad \|v\|_{-,k} \leq C_- \|v\|_- \ ; \tag{25}$$

moreover, by Assumption 4 with (22),

$$\forall v \in \mathcal{Y}_k \ : \quad \|v\|_{\mathcal{Y}^{-p}} \simeq \|D_k^{-1/4} v\|_{\mathcal{Y}} \ . \tag{26}$$

Following [5], we define the smoothing operator $S_k : \mathcal{Y}_k \to \mathcal{Y}_k$ by

$$\langle S_k w, v\rangle_{-,k} = \frac{1}{\xi_k}\langle w, v\rangle_- \quad \text{for all } v \in \mathcal{Y}_k \ ,$$

where

$$C \sup_{v\in\mathcal{Y}_k} \frac{|||v|||^2}{\|v\|^2_{-,k}} \geq \xi_k \geq \sup_{v\in\mathcal{Y}_k} \frac{|||v|||^2}{\|v\|^2_{-,k}} \ , \tag{27}$$

for some constant $C > 0$, i.e., $\xi \sim h_k^{-2p}$,

$$S_k = \frac{1}{\xi_k} D_k \mathcal{Q}_k (\mathcal{T}\mathcal{T}^*)^2 \tag{28}$$

so that

$$S_k A_k = \frac{1}{\xi_k} D_k \mathcal{Q}_k \mathcal{T}\mathcal{T}^*$$

is symmetric and (17) holds by

$$A(S_k A_k v, v) = \frac{1}{\xi_k} \langle D_k \mathcal{Q}_k \mathcal{T}\mathcal{T}^* v, \mathcal{T}\mathcal{T}^* v\rangle_{\mathcal{Y}} = \frac{1}{\xi_k} \|D_k \mathcal{Q}_k \mathcal{T}\mathcal{T}^* v\|_{-,k}^2 \geq 0 \ , \tag{29}$$

$$\begin{array}{rcl} A(S_k A_k v, v) & \leq & \frac{1}{\xi_k} |||D_k \mathcal{Q}_k \mathcal{T}\mathcal{T}^* v|||\ |||v||| \\ & \leq & \frac{1}{\sqrt{\xi_k}} ||D_k \mathcal{Q}_k \mathcal{T}\mathcal{T}^* v||_{-,k}\ |||v||| \\ & = & \sqrt{A(S_k A_k v, v)}\ |||v|||\ , \end{array} \tag{30}$$

where we have used the definition of ξ_k according to (27). On the other hand, the upper bound on ξ_k in (27) implies, by the identity $\mathcal{Q}_n \mathcal{T}[I - \mathcal{P}_n] = 0$, (13), Assumption 1, and (26):

$$\xi_k \|\mathcal{T}(I - \mathcal{P}_{k-1})\| \leq \xi_k \overline{\overline{\gamma}}_{k-1}^2 \leq C\ h_k^{2p} \left(\sup_{v \in \mathcal{Y}_k} \frac{\|D_k^{-1/4} v\|}{\|D_k^{-1/2} v\|} \right)^2 = C\ h_k^{2p} \|D_k^{1/4}\|^2 \leq C_{\xi\overline{\overline{\gamma}}} \tag{31}$$

for some constant $C_{\xi\overline{\overline{\gamma}}} > 0$, so that, by Cauchy-Schwarz, (21), and (25), we get

$$\begin{array}{rcl} A(S_k A_k v, v) & = & \frac{1}{\xi_k} \|D_k^{1/2} \mathcal{Q}_k \mathcal{T}\mathcal{T}^* v\|_{\mathcal{Y}}^2 \\ & \geq & \frac{1}{\xi_k} \left(\frac{\langle D_k^{1/2} \mathcal{Q}_k \mathcal{T}\mathcal{T}^* v, D_k^{-1/2} (P_k - P_{k-1}) v\rangle_{\mathcal{Y}}}{\|D_k^{-1/2} (P_k - P_{k-1}) v\|_{\mathcal{Y}}} \right)^2 \\ & = & \frac{1}{\xi_k} \left(\frac{|||(P_k - P_{k-1}) v|||^2}{\|(P_k - P_{k-1}) v\|_{-,k}} \right)^2 \\ & \geq & \frac{1}{C_-\ \xi_k} \left(\frac{\|(I - \mathcal{P}_{k-1}) \mathcal{P}_k \mathcal{T}^* v\|_{\mathcal{X}}}{\|\mathcal{T}(I - \mathcal{P}_{k-1}) \mathcal{P}_k \mathcal{T}^* v\|_{\mathcal{Y}}} \right)^2 |||(P_k - P_{k-1}) v|||^2 \\ & \geq & \frac{1}{C_-\ C_{\xi\overline{\overline{\gamma}}}} |||(P_k - P_{k-1}) v|||^2 \ . \end{array} \tag{32}$$

Taking into account the fact that here

$$Q_k = \Big(\mathcal{Q}_k (\mathcal{T}\mathcal{T}^*)^2 \mathcal{Q}_k \Big)^{-1} \mathcal{Q}_k (\mathcal{T}\mathcal{T}^*)^2$$

$$A_k = \Big(\mathcal{Q}_k (\mathcal{T}\mathcal{T}^*)^2 \mathcal{Q}_k \Big)^{-1} \mathcal{Q}_k \mathcal{T}\mathcal{T}^* \mathcal{Q}_k$$

and setting

$$\mathcal{S}_k := \frac{1}{\xi_k} \mathcal{T}^* D_k$$

we can transform Algorithm 1 (or its non-symmetric version) to define a multigrid operator $\mathcal{B}_n^s$ (or $\mathcal{B}_n^n$) for the iterative solution or preconditioning of equation (7):

Algorithm 2. SET $\mathcal{B}_0^{s/n} := (\mathcal{Q}_0\mathcal{T}\mathcal{P}_0)^{-1}$. ASSUME THAT $\mathcal{B}_{k-1}^{s/n}$ HAS BEEN DEFINED AND DEFINE $\mathcal{B}_k^{s/n} b$ FOR $b \in \mathcal{Y}_k$ AS FOLLOWS:
SET $x := \mathcal{S}_k b$.
SET $x := x + \mathcal{B}_{k-1}^{s/n}\mathcal{Q}_{k-1}(b - \mathcal{Q}_k\mathcal{T}x)$.
(SET $x := x + \mathcal{S}_k(b - \mathcal{Q}_k\mathcal{T}x)$.)
SET $\mathcal{B}_k^{s/n} b := x$.

It can be easily seen by induction that the so defined $\mathcal{B}_k^s$, ($\mathcal{B}_k^n$) is related to B_k^s, (B_k^n) by

$$\mathcal{B}_k^{s/n} = \mathcal{T}^* B_k^{s/n}\Big(\mathcal{Q}_k(\mathcal{T}\mathcal{T}^*)^2\mathcal{Q}_k\Big)^{-1} ,$$

and hence we have

$$\mathcal{B}_k^{s/n}\mathcal{Q}_k\mathcal{T}\mathcal{P}_k = \mathcal{T}^* B_k^{s/n} A_k \ . \tag{33}$$

Corollary 1 together with (29), (30), (31), (32) therefore implies

Corollary 2. *Let Assumptions 1, 2, 3, 4 for (22) and (23), (24) hold and define $\mathcal{B}_n^s$ and $\mathcal{B}_n^s$ by Algorithm 2 with $\mathcal{S}_k$ according to (28). Then*

$$\langle (I - \mathcal{B}_n^s\mathcal{Q}_n\mathcal{T}\mathcal{P}_n)x, x\rangle_{\mathcal{X}} = \|(I - \mathcal{B}_n^n\mathcal{Q}_n\mathcal{T}\mathcal{P}_n)x\|_{\mathcal{X}}^2 \le c\|x\|_{\mathcal{X}}^2 \text{ for all } x \in \mathcal{X}_n \tag{34}$$

for some $c < 1$ independent of n.

With $b^k \in \mathbb{R}^{dim(\mathcal{Y}_k)}$ being the vector of coefficients of $b \in \mathcal{Y}_k$ with respect to some basis $\{\psi_1^k, \dots, \psi_{dim(\mathcal{Y}_k)}^k\}$ of $\mathcal{Y}_k$,

$$b = \sum_{j=1}^{dim(\mathcal{Y}_k)} b_j^k \psi_j^k \ , \tag{35}$$

and with the matrices

$$\bar{D}^k := \Big(\langle D_k\psi_i^k, \psi_j^k\rangle_{\mathcal{Y}}\Big)_{1\le i,j\le dim(\mathcal{Y}_k)} ,$$

$$\bar{G}^k := \Big(\langle \psi_i^k, \psi_j^k\rangle_{\mathcal{X}}\Big)_{1\le i,j\le dim(\mathcal{Y}_k)} , \tag{36}$$

$$\bar{M}^k := \Big(\langle \mathcal{T}^*\psi_i^k, \mathcal{T}^*\psi_j^k\rangle_{\mathcal{X}}\Big)_{1\le i,j\le dim(\mathcal{Y}_k)} , \tag{37}$$

and $\bar{C}^{k-1} \in \mathbb{R}^{dim(\mathcal{Y}_k)}_{dim(\mathcal{Y}_{k-1})}$ such that, by (6)

$$(\psi_1^{k-1}, \dots, \psi_{dim(\mathcal{Y}_{k-1})}^{k-1})^T = \bar{C}^{k-1}(\psi_1^k, \dots, \psi_{dim(\mathcal{Y}_k)}^k)^T \tag{38}$$

we can write

$$\mathcal{B}_k^{s/n} = \mathcal{T}^* \sum_{j=1}^{dim(\mathcal{Y}_k)} (\overline{\mathcal{B}^{s/n}}^k b^k)_j \psi_j^k \ , \tag{39}$$

where

Algorithm 3. (MATRIX FORM OF ALGORITHM 2 WITH (28))
SET $\overline{\mathcal{B}^{s/n}}^0 := (\bar{M}^0)^{-1}G^0$. ASSUME THAT $\overline{\mathcal{B}^{s/n}}^{k-1}$ HAS BEEN DEFINED AND DEFINE $\overline{\mathcal{B}^{s/n}}^k b^k$ FOR $b^k \in \mathbb{R}^{dim(\mathcal{Y}_k)}$ AS FOLLOWS:
SET $x^k := \frac{1}{\xi_k}(\bar{G}^k)^{-1}\bar{D}^k b^k$.
SET $x^k := x^k + (\bar{C}^{k-1})^T\overline{\mathcal{B}^{s/n}}^{k-1}\bar{G}^{k-1}\bar{C}^{k-1}(\bar{G}^k b^k - \bar{M}^k x^k)$.
(SET $x^k := x^k + \frac{1}{\xi_k}(\bar{G}^k)^{-1}\bar{D}^k(b^k - (\bar{G}^k)^{-1}\bar{M}^k x^k)$.)
SET $\overline{\mathcal{B}^{s/n}}^k b^k := x^k$.

3.2 Modification of King (1992)

In our second approach we choose $\langle .,. \rangle$ to be equal to the $\mathcal{Y}$-inner product, i.e.,

$$\langle .,. \rangle = \langle .,. \rangle_{\mathcal{Y}} \ ,$$

hence this time corresponding to a stronger norm than $|||.|||$ on $\mathcal{Y}_k$. The norm equivalence Assumption 4, which we here assume to hold for

$$q = p \ , \quad \mathcal{Y}' = (\mathcal{Y}^{-p})^* \tag{40}$$

in combination with Assumption 1 is here used to define a smoothing operator

$$S_k : \mathcal{Y}_k \to \mathcal{Y}_k$$

by

$$S_k := c_\xi \hat{D}_k \ ,$$

where the difference operator $\hat{D}_k : \mathcal{Y}_k \to \mathcal{Y}_k$ is the square root of D_k from Subsection 3.1:

$$\hat{D}_k v = \sum_{j=1}^{k} h_j^{-2p}(\mathcal{Q}_j - \mathcal{Q}_{j-1})v + \mathcal{Q}_0 v \ ,$$

so that, by our assumptions,

$$c_{\hat{D}} \|v\|_{(\mathcal{Y}^{-p})^*} \le \langle \hat{D}_k v, v \rangle_{\mathcal{Y}} \le C_{\hat{D}} \|v\|_{(\mathcal{Y}^{-p})^*} \ ,$$

and the constant c_ξ is chosen such that

$$0 < c_\xi \le \frac{1}{C_{\hat{D}} \|\mathcal{T}^*\|_{\mathcal{Y}^{-p}\to\mathcal{X}}} \ .$$

The operator $S_k A_k$ is then clearly symmetric with respect to $A(.,.)$ and

$$\begin{aligned} 0 \le A(S_k A_k v, v) &= c_\xi\left(\textstyle\sum_{j=1}^k h_j^{-2p}\|\mathcal{Q}_j - \mathcal{Q}_{j-1})A_k v\|_{\mathcal{Y}}^2 + \|\mathcal{Q}_0 A_k v\|_{\mathcal{Y}}^2\right) \\ &\le c_\xi C_{\hat{D}} \|A_k v\|^2_{(\mathcal{Y}^{-p})^*} \\ &= c_\xi C_{\hat{D}} \left(\sup_{w\in\mathcal{Y}_k} \frac{\langle \mathcal{T}^* v, \mathcal{T}^* w\rangle_{\mathcal{X}}}{\|w\|_{\mathcal{Y}^{-p}}}\right)^2 \\ &\le c_\xi C_{\hat{D}} \|\mathcal{T}^*\|_{\mathcal{Y}^{-p}\to\mathcal{X}} |||v|||^2 \end{aligned} \tag{41}$$

Analogously one sees that

$$A(S_kA_kv,v) \geq c|||v|||^2 \tag{42}$$

for some $c>0$.

By the fact that now

$$Q_k = \mathcal{Q}_k \ , \quad A_k = \mathcal{Q}_k\mathcal{T}\mathcal{T}^*\mathcal{Q}_k$$

we obtain, like in the section above, multigrid operators $\mathcal{B}_n^s$, $\mathcal{B}_n^n$ for (7) from Algorithm 2, setting this time

$$\mathcal{S}_k := c_\xi \mathcal{T}^*\hat{D}_k \ . \tag{43}$$

Note that the so defined $\mathcal{B}_k^s$, $\mathcal{B}_k^n$ satisfy

$$\mathcal{B}_k^{s/n} = \mathcal{T}^* B_k^{s/n} \ ,$$

as well as (33), so that Corollary 1 together with (41), (42) implies

Corollary 3. *Let Assumptions 1, 2, 3, and 4 for (40) hold and define $\mathcal{B}_n^s$ and $\mathcal{B}_n^s$ by algorithm 2 with $\mathcal{S}_k$ according to (43). Then*

$$\langle (I-\mathcal{B}_n^s\mathcal{Q}_n\mathcal{T}\mathcal{P}_n)x,x\rangle_{\mathcal{X}} = \|(I-\mathcal{B}_n^n\mathcal{Q}_n\mathcal{T}\mathcal{P}_n)x\|_{\mathcal{X}}^2 \leq c\|x\|_{\mathcal{X}}^2 \text{ for all } x\in\mathcal{X}_n \tag{44}$$

for some $c<1$ independent of n.

In matrix notation, Algorithm (2) with (43) by (35), (36), (37), (38), (39) and

$$\hat{\bar{D}}^k := \Big(\langle \hat{D}_k\psi_i^k,\psi_j^k\rangle_{\mathcal{Y}}\Big)_{1\leq i,j\leq dim(\mathcal{Y}_k)} \ ,$$

becomes

Algorithm 4. (MATRIX FORM OF ALGORITHM 2 WITH (43))
SET $\overline{\mathcal{B}^{s/n}}^0 := (\bar{M}^0)^{-1}G^0$. ASSUME THAT $\overline{\mathcal{B}^{s/n}}^{k-1}$ HAS BEEN DEFINED AND DEFINE $\overline{\mathcal{B}^{s/n}}^k b^k$ FOR $b^k\in\mathbb{R}^{dim(\mathcal{Y}_k)}$ AS FOLLOWS:
SET $x^k := c_\xi(\bar{G}^k)^{-1}\hat{\bar{D}}^k b^k$.
SET $x^k := x^k + (\bar{C}^{k-1})^T\overline{\mathcal{B}^{s/n}}^{k-1}\bar{G}^{k-1}\bar{C}^{k-1}(\bar{G}^kb^k - \bar{M}^kx^k)$.
(SET $x^k := x^k + c_\xi(\bar{G}^k)^{-1}\hat{\bar{D}}^k(b^k-(\bar{G}^k)^{-1}\bar{M}^kx^k)$.)
SET $\overline{\mathcal{B}^{s/n}}^k b^k := x^k$.

By inequalities (41) and (42), also the additive BPX preconditioner from [7]

$$\mathcal{B}_n^a = c_\xi\mathcal{T}^*\Big(\sum_{k=1}^n h_k^{-2p}(\mathcal{Q}_k-\mathcal{Q}_{k-1})+\mathcal{Q}_0\Big) \tag{45}$$

yields a level independent condition number:

Corollary 4. *Let assumptions 1, 2, 3, and 4 for (40) hold and define $\mathcal{B}_n^a$ by (45). Then*

$$0 \leq \langle (I - \mathcal{B}_n^a \mathcal{Q}_n \mathcal{T} \mathcal{P}_n)x, x\rangle_{\mathcal{X}} \leq c\|x\|_{\mathcal{X}}^2 \text{ for all } x \in \mathcal{X}_n \tag{46}$$

for some $c < 1$ independent of n.

Remark 2. In applications like parameter identification, the evaluation of $\mathcal{T}$ (and of $\mathcal{T}^*$) will often involve the solution of a PDE, which usually has to be done in an approximate way. In that case it will be more appropriate not to explicitely calculate the entries of the matrix $\bar{M}^k$, but to assemble $u_k := \sum_{j=1}^{dim(\mathcal{Y}_k)} x_j^k \psi_j^k$ and (approximately) compute the coefficients of $\mathcal{Q}_n TT^* u_k$, in the coarse grid correction (and in the post smoothing step) of the respective multigrid algoritm. Denoting by ϵ_k^{SA} the energy norm of the additional error that is made in $S_k A_k$ by using numerical approximations $(\tilde{\mathcal{T}}^*\tilde{\mathcal{T}})_k$ of $\mathcal{T}^*\mathcal{T}$,

$$\epsilon_k^{SA} := |||\tilde{S}_k \tilde{A}_k - S_k A_k||| \ ,$$

and denoting by a tilde the respective perturbed operators, we see by inspection of the proof of Corollary 1, that for the perturbed multigrid operators $\tilde{B}_n$, on one hand

$$|||u|||^2 - |||(I - \tilde{E}_n)u|||^2 \geq \sum_{k=0}^{n} A(S_k A_k \mathcal{P}_k \tilde{E}_{k-1} u, \mathcal{P}_k \tilde{E}_{k-1} u) - \sum_{k=0}^{n} (3 + \epsilon_k^{SA})\epsilon_k^{SA} |||\mathcal{P}_k \tilde{E}_{k-1} u|||^2$$

for any $u \in \mathcal{Y}_n$. Under the induction hypothesis that for all $k \leq n-1$

$$|||\mathcal{P}_k \tilde{E}_k|||^2 \leq (1 - \frac{1}{C^2} + \epsilon) \tag{47}$$

for some $\epsilon < \frac{1}{C^2}$, (which obviousy holds for $k = 0$), we can estimates the terms in the second sum by

$$\begin{aligned} |||\mathcal{P}_k \tilde{E}_{k-1} u|||^2 &\leq |||(\mathcal{P}_k - \mathcal{P}_{k-1}) E_{k-1} u|||^2 + (1 - \tfrac{1}{C^2} + \epsilon)|||\mathcal{P}_{k-1} u|||^2 \\ &\leq |||(\mathcal{P}_k - \mathcal{P}_{k-1}) u|||^2 + |||\mathcal{P}_{k-1} u|||^2 = |||u|||^2 \ . \end{aligned}$$

On the other hand, since $(\tilde{E}_{k-1} - E_{k-1}) \in \mathcal{Y}_{k-1}$ and $\tilde{E}_{-1} = E_{-1} = I$, estimate (20) remains valid with E_{k-1} replaced by $\tilde{E}_{k-1}$. Altogether we arrive at (47) for $k = n$, i.e., a level independent contraction factor $1 - \frac{1}{C^2} + \epsilon < 1$ if we assume that

$$\sum_{k=0}^{n} (3 + \epsilon_k^{SA})\epsilon_k^{SA} < \epsilon < \frac{1}{C^2} \ . \tag{48}$$

For Algorithm 2 of Subsection 3.1 we have, for any $v \in \mathcal{Y}_k$, by Assumptions 1, 2, 3, and (27),

$$\begin{aligned} |||(\tilde{S}_k \tilde{A}_k - S_k A_k)v||| &= \tfrac{1}{\xi_k} \|\mathcal{T}^* D_k \mathcal{Q}_k (\tilde{\mathcal{T}}^*\tilde{\mathcal{T}} - \mathcal{T}^*\mathcal{T})v\|_{\mathcal{X}} \\ &\leq C h_k^{-p} \|\mathcal{Q}_k (\tilde{\mathcal{T}}^*\tilde{\mathcal{T}} - \mathcal{T}^*\mathcal{T})v\|_{\mathcal{Y}} \\ &\leq C h_k^{-2p} \|\mathcal{Q}_k (\tilde{\mathcal{T}}^*\tilde{\mathcal{T}} - \mathcal{T}^*\mathcal{T})\| \, |||v||| \end{aligned}$$

for some constant $C > 0$. Due to requirement (48), the approximation $\tilde{\mathcal{T}}^*\tilde{\mathcal{T}}$ to $\mathcal{T}^*\mathcal{T}$ therefore has to become more accurate for growing k, such that the sum

$$\sum_{k=0}^{\infty} h_k^{-2p} \|\mathcal{Q}_k(\tilde{\mathcal{T}}^*\tilde{\mathcal{T}} - \mathcal{T}^*\mathcal{T})\| \tag{49}$$

is finite and sufficiently small. The summands may be further estimated due to Assumptions 1 and 3:

$$\|\mathcal{Q}_k(\tilde{\mathcal{T}}^*\tilde{\mathcal{T}} - \mathcal{T}^*\mathcal{T})\| \leq h_k^{-p} \|\mathcal{Q}_k(\tilde{\mathcal{T}}^*\tilde{\mathcal{T}} - \mathcal{T}^*\mathcal{T})\|_{\mathcal{Y} \to \mathcal{Y}^{-p}}$$

so that (49) can be achived, e.g., by using a sufficiently fine discretization (finer than the one of $\mathcal{Y}_k$) in the numerical approximation of $\mathcal{T}$ and $\mathcal{T}^*$. A similar assumption has to be made when using approximate operator evaluations of $\mathcal{T}$, $\mathcal{T}^*$ in the context of Section 3.2.

4 Full multigrid as an iterative regularization method

The contraction factor estimates in Corollaries 2, 3 and 4 (that obviously also hold for cycles of more than one coarse grid correction) make it possible to apply the respective multigrid operators as preconditioners for (7), yielding level independent condition numbers.

Alternatively, the multigrid operators of Section 3 can be used in in a full multigrid method (cf., e.g., [8]) for the iterative solution of (7):

Algorithm 5. (FULL MULTIGRID METHOD)
SET $\tilde{x}_0 := (\mathcal{Q}_0 T)^\dagger \mathcal{Q}_0 y^\delta$
FOR $k = 1, \ldots, n$
    SET $\tilde{x}_k := \tilde{x}_{k-1} + \mathcal{B}_k(\mathcal{Q}_k y^\delta - \mathcal{Q}_k T \tilde{x}_{k-1})$

This iterative scheme can be easily shown to converge to an exact solution of the infinite-dimensional problem (1), if the data are exact, i.e., $\delta = 0$. In the practically relevant situation of nonvanishing data noise, one has to take special care of the propagated data noise, though, which is amplified by a worst case factor $\sim h_n^{-p}$ at the $n-$th level, hence may explode as $n \to \infty$. It is therefore crucial to find for given data y^δ and noise level δ, a stopping rule $n = N(\delta, y^\delta)$ for the iteration 5, that carries out the trade-off between approximation error and propagated data noise in an optimal way in the sense that the so-defined approximation $\tilde{x}_{N(\delta,y^\delta)}$ converges to the exact bestapproximate solution of (1) as $\delta \to 0$, and additional a priori information (4) yields optimal rates (5) (cf. [16]). It was shown in [18], that a Morozov type discrepancy principle does so even without needing explicit knowledge of the exponent μ in (4).

Corollary 5. *Let the assumptions of Corollary 2 or of Corollary 3 hold and fix $\tau > 1$.*
Then for any $\delta < \frac{\|y\|}{\tau+1}$ and for any data y^δ with (2), a finite $N(\delta)$, with

$$N(\delta, y^\delta) := \min\{n \in \mathbb{N} \mid \|\mathcal{T}\tilde{x}_n - y^\delta\| \leq \tau\delta\} \tag{50}$$

exists.

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

Under an additional source condition (4) for $\mu \leq \frac{1}{2}$, the optimal convergence rate (5) is achieved.

Proof. see the proof of Theorem 3 in [18].

Remark 3. To be able to treat also nonlinear ill-posed problems

$$F(x) = y$$

one can either use the proposed multigrid operators in each linear step of a Newton iteration, or, analogously to [18], generalize the smoothing operators to directly incorporate them into a nonlinear multigrid method.

5 Numerical experiments

For numerically verifying the theoretically predicted condition number estimates using the proposed multigrid operators as preconditioners, we considered two different model problems for which we implemented the proposed multigrid methods in a simple MatLab program on an SGI origin.

As our first example we study the Abel integral equation

$$\int_0^t \frac{x(s)}{\sqrt{t-s}}\, ds = y(t) \ , \tag{51}$$

which represents the rotational symmetric two-dimensional case in X-ray tomography. Here the degree of smoothing of the forward operator is $p = \frac{1}{2}$ and therefore corresponds to the situation considered in [3]. Using continuous piecewise linear splines for defining our discretization spaces $\mathcal{Y}_n$, we fulfill the norm equivalence requirements in both multiplicative multigrid variants of Subsections 3.1, 3.2 as well as the additive one (45). The resulting condition numbers are plotted in Figure 5.

The second example is a simple parameter identification problem, namely the one-dimensional case in the steady state inverse groundwater filtration problem of retrieving a space-dependent filtration coefficient $a \in H^1(0,1)$ in

$$(au_x)_x = 0 \quad \text{in } (0,1) \tag{52}$$

from L^2-data of u in $(0,1)$. Moreover, since this inverse problem is nonlinear, we linearize it around $a_0 :\equiv 1$. The resulting first kind operator equation satisfies Assumptions 1, (23) with $p = 2$ (cf. [16]). Choosing $\mathcal{Y}_n$ as spaces of C^0 (and piecewise linear) splines can therefore, in view of Assumption 4 with (22) or (40), as well as (16), not be expected to suffice for obtaining level independent V-cycle convergence, as is reflected in the condition number plot of Figure 5. Using C^1-splies instead, restores level independence, though (see Figure 5), which for the preconditioner of Section 3.1 is even more than our theory would suggest (cf. (24) and (16)). In Table 1 we list numbers of iterates and CPU times for preconditioned CG-iterations for the latter situation, as they were obtained for a prescribed reduction of the error by a factor of 1.e-4. While the number of unpreconditioned iterates explodes for growing projection space dimension, due to the ill condition of the system matrix, the increase is moderate for the multigrid preconditioned systems. The CPU times (Table 2) for the multigrid preconditioned CG, growing roughly quadratically (as can be expected for a full system), are better than those for the unpreconditioned CG already for relatively small dimensions. The tests made here are within the interesting range of degrees of freedom, as typically chosen by the discrepancy principle (50) for noise levels δ between 0.1 and 10 per cent of the data norm.

Table 1

dim(k)	number of iterates			
	Alg.3 PCG	Alg.4 PCG	BPX (45) PCG	unprecond. CG
34	20	20	30	103
66	40	38	53	343
130	56	56	80	1112
258	74	78	102	3938

Table 2

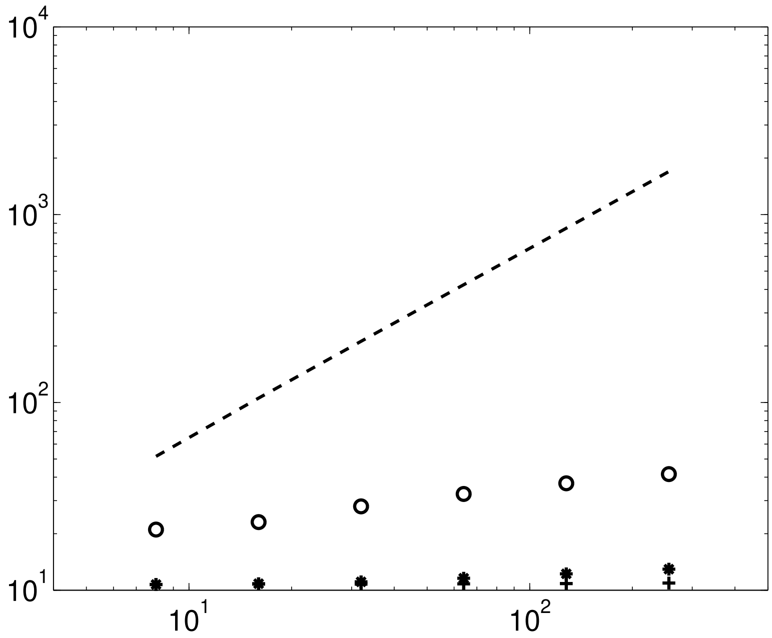
dim(k)	CPU time (in sec.)			
	Alg.3 PCG	Alg.4 PCG	BPX (45) PCG	unprecond. CG
34	6.00e-02	5.00e-02	6.00e-02	8.00e-02
66	1.60e-01	1.70e-01	1.40e-01	3.30e-01
130	6.00e-01	6.60e-01	4.50e-01	1.78e+00
258	3.33e+00	3.89e+00	2.18e+00	1.53e+01

Acknowledgement The author wishes to thank Dr. Schöberl from the SFB F013 in Linz for the hint at the most important papers [3] and [5] as well as for interesting and instructive discussions on multigrid methods. 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 multigrid 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.



$- -$ . . . unpreconditioned
$*$ . . . Alg.3 preconditioner
$+$ . . . Alg.4 preconditioner
$\circ$ . . . BPX preconditioner (45)

Figure 1: condition numbers for Abel integral equation (51), projection onto piecewise linear C^0-splines.

[3] J.H. Bramble, Z. Leyk, J.E. Pasciak, *The analysis of multigrid algorithms for pseudo-differential operators of order minus one*, Math. Comp., 63 (1994), 461-478.

[4] J.H. Bramble, J.E. Pasciak, *New estimates for multigrid algorithms including the V-cycle*, Math. Comp. 60 (1993), 447-471.

[5] J.H. Bramble, J.E. Pasciak, P.S. Vassilievski, *Computational scales of Sobolev norms with application to preconditioning*, Math. Comp. 69 (2000), 463-480.

[6] J.H. Bramble, J.E. Pasciak, J. Wang, J. Xu, *Convergence estimates for multigrid algorithms without regularity assumptions*, Math. Comp. 57 (1991), 23-45.

[7] J.H. Bramble, J.E. Pasciak, J. Xu, *Parallel multilevel preconditioners*, Math. Comp. 55 (1990), 1-22.

[8] A. Brandt, *Multilevel adaptive solutions to boundary value problems* Math. Comp. 31 (1977), 333-390.

[9] 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.

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

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

[12] 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.

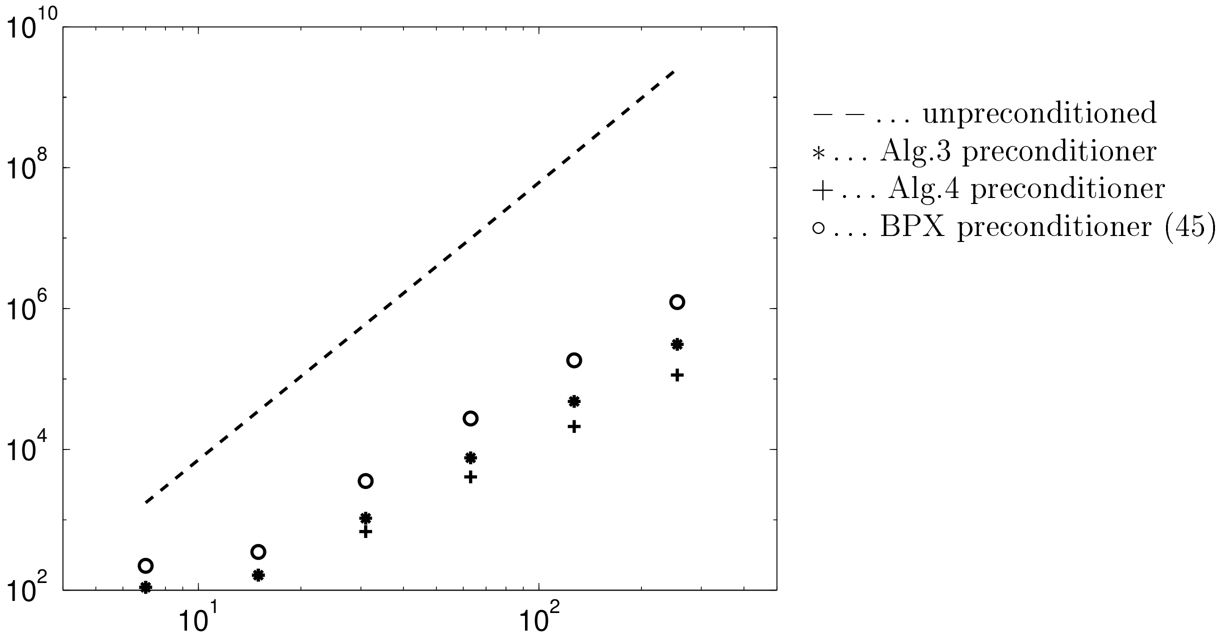


Figure 2: condition numbers for linearized 1-d parameter identification (52), projection onto C^0-splines.

[13] W. HACKBUSCH, Multi-Grid Methods and Applications, Springer, Berlin, 1985.

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

[15] B. KALTENBACHER, *A projection-regularized Newton method for nonlinear ill-posed problems with application to parameter identification problems with finite element discretization*, SIAM J.Numer.Anal. 37 (2000), 1885-1908.

[16] B. KALTENBACHER, *On the regularizing properties of a full multigrid method for ill-posed problems*, submitted.

[17] B. KALTENBACHER, *Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems*, to appear in Inverse Problems.

[18] B. KALTENBACHER, *A Multigrid Method with A Priori and A Posteriori Level Choice for the Regularization of Nonlinear Ill-Posed Problems* , SFB013-report No.00-3, University of Linz, 2000.

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

[20] A.K. LOUIS, *Inverse und schlecht gestellte Probleme*, Teubner, Stuttgart, 1989.

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

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

[23] F. NATTERER, *The Mathematics of Computerized Tomography*, Teubner, Stuttgart, 1986.

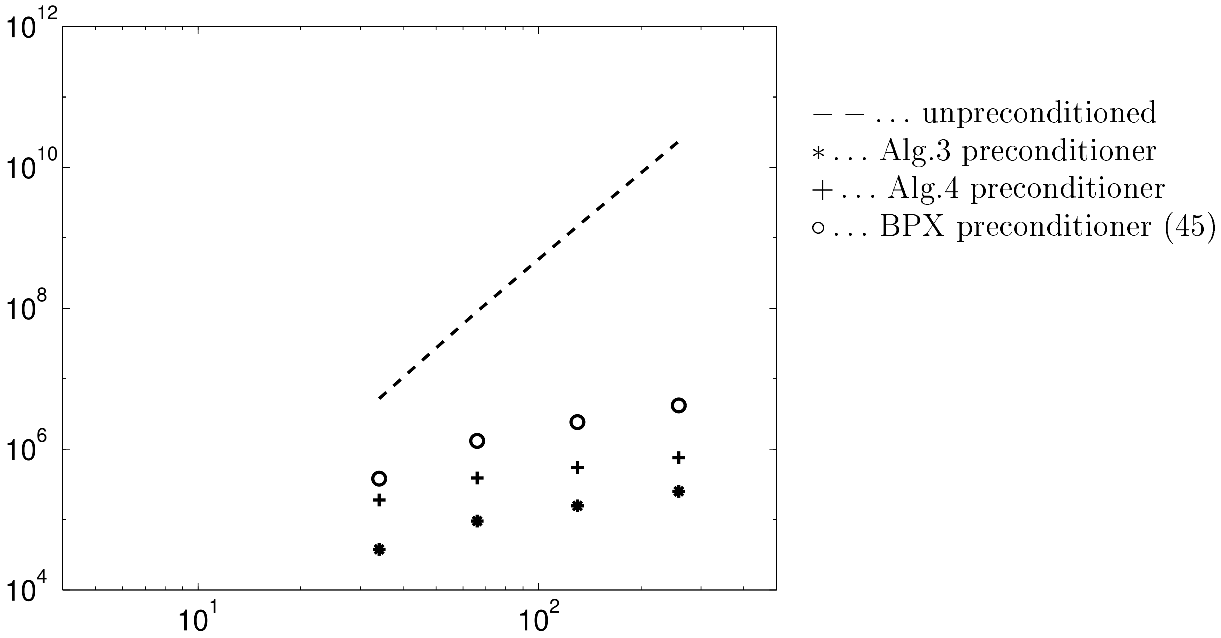


Figure 3: condition numbers for linearized 1-d parameter identification (52), projection onto C^1-splines.

[24] P. Oswald, *Multilevel Finite Element Approximations. Theory and Applications*, Teubner, Stuttgart, 1994.

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

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

[27] G.M. Vainikko, U. Hämarik, *Projection methods and self-regularization in ill-posed problems*, Sov. Math. 29 (1985), 1-20.

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