On a Multigrid Method for Ill-Posed Problems [1]

Barbara Kaltenbacher

Abstract. For the stable solution of a linear ill-posed operator equation

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

we project the equation onto finite dimensional spaces and apply a multigrid method closely related to the one proposed by King in [21]. We estimate the condition numbers of the resulting MG preconditioned system matrices and study the regularizing properties of the corresponding full multigrid method with a priori chosen finest level. For the MG preconditiones under consideration, applied to a parameter identification problem from inverse groundwater filtration, numerical results are given.

Key words. ill-posed problems, multigrid methods, parameter identification

AMS(MOS) Subject Classifications: 65J20, 65R30, 65N55, 35R30

1 Introduction

We consider

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

and the problem of finding a best approximate solution $x^\dagger := \mathcal{T}^\dagger y$ (i.e., the element of minimal norm minimizing the residual of $\|\mathcal{T}x - y\|$), where $\mathcal{T} : \mathcal{X} \to \mathcal{Y}$ is a bounded linear operator between Hilbert spaces $\mathcal{X}$ and $\mathcal{Y}$. We are particularly interested in the situation that the range $R(\mathcal{T})$ of $\mathcal{T}$ is non-closed or, equivalently, the Moore-Penrose generalized inverse $\mathcal{T}^\dagger$ is unbounded, so that $x^\dagger$ only exists if $y \in R(\mathcal{T}) + R(\mathcal{T})^\perp$ and does not depend continuously on the (measured) data y^δ (with noise level $\delta := \|y - y^\delta\|$) — think, e.g., of an integral equation of the first kind with weakly singular kernel. In order to cope with this instability one usually applies so-called regularization methods (see e.g., [9], [12], [22], [23], [25], [27], [28]), the most well-known one being Tikhonov regularization, which approximates $x^\dagger$ by a minimizer of the Tikhonov functional $\|\mathcal{T}x - y\|^2 + \alpha\|x\|^2$, where $\alpha > 0$ is a small regularization parameter. It is characteristic for ill-posed problems that any solution method converges in general only arbitrarily slowly and that convergence rates can only be achieved under so-called source or regularity conditions

$$x^\dagger \in R((\mathcal{T}^*\mathcal{T})^\mu), \tag{2}$$

with some real positive exponent μ. We just mention in passing that when taking into account also convergence rates for severely ill-posed problems such as inverse heat conduction or inverse scattering problems, one would have to work with logarithmic instead of Hölder type source conditions (cf. [18]), which we will not do in more detail here, though.

For a practical implementation, a regularization method has to be discretized. On the other hand, discretization, i.e., projection on a finite-dimensional space itself can have a regularizing effect: Finite dimensional problems — though they might be ill-conditioned — are always well-posed in the sense of stable dependence of the best approximate solution on the data. Moreover, this seems to be a very natural and — by the choice of different bases — also general way of introducing a problem-adapted numerical stabilization of some ill-posed problem. As a convergence analysis (see [10], [13], [24]) summarized in [9] shows, there are good reasons for using regularization by projection in image space rather that in preimage space, the most crucial one being the fact that image space projection — unlike preimage space projection — generically yields convergence.

Any solution method for (1) will at some point involve the solution of a possibly large system of linear equations. Multigrid methods (MGM) are well known as extremely efficient solvers for large

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

scale systems of equations resulting from partial differential equations and integral equations of the second kind and have been extensively studied in the last years (see e.g. [1], [3], [4], [6], [14], [15], [16]). For ill-posed problems their behavior is not so well understood, yet (see [17], [21], [26]). The main problem lies in the definition of smoothing operators: The crucial effect of smoothing out the higher frequency part of the error, that the usual iterative schemes such as Jacobi, Gauß-Seidel, or SOR show in discretizations of PDEs or second kind integral equations, gets lost for first kind integral equations, due to their adverse eigensystem structure: here high frequencies correspond to small singular values of the operator $\mathcal{T}$ and are therefore strongly amplified by inversion. Therefore the good results on multigrid methods for well-posed problems such as V-cycle convergence with a level-independent or only moderately level-dependent contraction number, cannot be expected to be completely carried over to ill-posed operator equations.

While the equations whose solution is studied in the papers [17], [26] come from (Tikhonov) regularization of (1) combined with projection in preimage space, we here apply projection in image space for the reasons mentioned above, and put more emphasis on the aspect of regularization *by* projection, still including the possibility of additional regularization (see Section 5 below).

The multigrid approach, which we will study here, comes from a paper by King, [21], (see also [20]), where the smoothers are defined by subdividing each finite-dimensional projection space into a relatively low frequent and a relatively high frequent part. By the smoothers defined in [21], the low frequent part remains unchanged, while the high frequent part is treated by a simple appropriate smoother, that in fact copes with the ill-posedness. In [21] convergence of a resulting W-cycle type MGM with a level-dependent contraction factor is shown.

Going on from [21] we here study the question under which conditions a level-independent contraction factor can be achieved. Moreover we introduce a simple modification of Kings multigrid operator that enables us to give an additional worst case condition number estimate of the MGM preconditioned operators. Finally we investigate the regularizing properties of a full multigrid method.

The paper is organized as follows: In Section 2 we outline the discretization and some corresponding norms and operators that we will make use of in the definitions and analysis of the subsequent Section 3, which, besides conditions for level-independent contraction factors as well as worst case estimates for the V-cycle and an additive multilevel preconditioner, contains algorithmic investigations including data noise propagation. In particular we show that the full multigrid method, with an a priori, and in dependence of the data noise magnitude, defined number of refinement levels, is a regularization method in the sense of [9], i.e. it is stable with respect to data noise and converges to the exact solution as the noise level goes to zero. The theoretical results are illustrated by some numerical experiments in Section 4.

2 Preliminaries

In this section we describe the discretization of (1) preceding the application of the MGM, and define some norms, projections, and operators that will be needed later on; the norms, inner products and adjoints on the Hilbert spaces $\mathcal{X}$, $\mathcal{Y}$ will be labelled by a respective subscript. 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.

The method of regularization by projection in image space (or dual least squares method) is defined as follows: Given a sequence

$$\mathcal{Y}_0 \subset \mathcal{Y}_1 \subset \mathcal{Y}_2 \subset \ldots$$

of finite-dimensional subspaces of $\overline{R(\mathcal{T})} = N(\mathcal{T}^*)^\perp \subset \mathcal{Y}$, whose union is dense in $N(\mathcal{T}^*)^\perp$, x_n is defined as the best approximate solution of the projected equation on the $n-th$ level

$$\mathcal{Q}_n \mathcal{T} x = \mathcal{Q}_n y \ , \tag{3}$$

where $\mathcal{Q}_n$ is the ($\mathcal{Y}$-) orthogonal projection onto $\mathcal{Y}_n$, i.e.,

$$x_n = (\mathcal{Q}_n \mathcal{T})^\dagger \mathcal{Q}_n y \tag{4}$$

It is immediate (see also, e.g. [9]) that the so defined x_n is just the ($\mathcal{X}$-) orthogonal projection $\mathcal{P}_n$ onto

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

of the solution $x^\dagger$:

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

and, since $\mathcal{X}_0 \subset \mathcal{X}_1 \subset \mathcal{X}_2 \subset \ldots$ and $\overline{\bigcup_{n \in \mathbb{N}} \mathcal{X}_n} = N(\mathcal{T})^\perp$, converges to $x^\dagger$ as $n \to \infty$. It has been shown (see e.g. Section 3.3 in[9] and the references therein) that an optimal choice of projection spaces is given by those spanned by singular functions of $\mathcal{T}$. The operator $(\mathcal{Q}_n \mathcal{T})^\dagger \mathcal{Q}_n$ then corresponds to truncated SVD and has diagonal form. We are here interested in more general discretization concepts, such as finite elements, splines, or wavelets.

It is this x_n for some fixed $n \in \mathbb{N}$, which we wish to approximate by some multilevel technique. Here, using again the nested spaces $\mathcal{Y}_k$, with $0 \le k \le n$, the indices $k = 0$ and $k = n$ correspond to the coarsest and finest grid, respectively.

With some bases $\{\psi_1^k, \ldots \psi_{d(k)}^k\}$ of $\mathcal{Y}_k$, where $d(k)$ is the dimension of $\mathcal{Y}_k$, $0 \le k \le n$, x_n is equivalently defined as the unique solution of the $d(n)$-dimensional problem

$$\begin{array}{ll} \langle \mathcal{T} x, \psi_j^n \rangle_{\mathcal{Y}} = \langle y, \psi_j^n \rangle_{\mathcal{Y}} & \quad j \in \{1, \ldots, d(n)\} \\ x \in \operatorname{span}\{\phi_1^n, \ldots, \phi_{d(n)}^n\} & \end{array}$$

with

$$\phi_j^k := \mathcal{T}^* \psi_j^k \ ,$$

i.e., $x_n = \sum_{j=1}^{d(k)} x_j^n \phi_j^n$ where the coefficient vector x^n solves the linear system

$$\tilde{A}^n x^n = \tilde{b}^n \tag{5}$$

with

$$\tilde{A}_{i,j}^k = \langle \mathcal{T}^* \psi_i^k, \mathcal{T}^* \psi_j^k \rangle_{\mathcal{X}} \ ,$$

and $\tilde{b}_i^k = \langle y, \psi_i^k \rangle_{\mathcal{Y}}$, $0 \le k \le n$. In the sequel we will assume that the ψ_j^k are normed (in $\mathcal{Y}$) and fulfill some uniform linear independence property:

$$\forall\ v \in Y_k \ : \quad c_\psi \|v\|_{\mathcal{Y}} \le \sqrt{\sum_{j=1}^{d(k)} \langle v, \psi_j^k \rangle_{\mathcal{Y}}^2} \le C_\psi \|v\|_{\mathcal{Y}} \ , \tag{6}$$

with $0 < c_\psi \le C_\psi$ independent of k and of n. This is equivalent to uniform boundedness of the spectrum of the Gramian matrix G_ψ^k with $(G_\psi^k)_{i,j} := \langle \psi_i^k, \psi_j^k \rangle_{\mathcal{Y}}$:

$$c_\psi^2 \le \lambda_{\min}(G_\psi^k) \le \lambda_{\max}(G_\psi^k) \le C_\psi^2 \ ,$$

which can be seen by inserting the identity

$$v = \sum_{i=1}^{d(k)} \sum_{j=1}^{d(k)} (G_\psi^k)_{i,j}^{-1} \langle v, \psi_j^k \rangle_{\mathcal{Y}} \psi_i^k \ , \quad \forall v \in Y_k$$

into (6).

On the finite dimensional subspaces $\mathcal{X}_k$ of $\mathcal{X}$ we define (in addition to the $\mathcal{X}$-topology) an inner product $\langle \cdot, \cdot \rangle_k$, that is equal or equivalent to the $\mathcal{Y}$-inner product of the respective preimages under $\mathcal{T}^*$:

$$\langle u, v \rangle_k \sim \langle (\mathcal{T}^*)^{-1} u, (\mathcal{T}^*)^{-1} v \rangle_{\mathcal{Y}} \ , \quad u, v \in \mathcal{X}_k \ .$$

Along with $\langle\cdot,\cdot\rangle_k$ we define the corresponding projection Q_k and restriction operator R_k:

$$Q_k : \mathcal{X}_{k+1} \to \mathcal{X}_k, \quad R_k : \mathcal{X}_{k+1} \to \mathcal{X}_k,$$

by

$$\langle Q_k u, v\rangle_{k+1} = \langle u, v\rangle_{k+1} \quad \forall u \in \mathcal{X}_{k+1},\ v \in \mathcal{X}_k$$
$$\langle R_k u, v\rangle_k = \langle u, v\rangle_{k+1} \quad \forall u \in \mathcal{X}_{k+1},\ v \in \mathcal{X}_k\ ,$$

as well as their extensions to all of X_n

$$\begin{array}{ll} \bar{Q}_k : \mathcal{X}_n \to \mathcal{X}_k & \bar{Q}_k := Q_k Q_{k+1} \cdots Q_{n-1} \\ \bar{R}_k : \mathcal{X}_n \to \mathcal{X}_k & \bar{R}_k := R_k R_{k+1} \cdots R_{n-1}\ . \end{array}$$

With the operators A_k defined by

$$\begin{array}{ll} A_k : & \mathcal{X}_k \to \mathcal{X}_k \\ & u = \sum_{j=1}^{d(k)} u_j^k \phi_j^k \mapsto \sum_{j=1}^{d(k)} (A^k u^k)_j \phi_j^k\ , \end{array}$$

where

$$A^k = (G_k^k)^{-1} \tilde{A}_k\ , \tag{7}$$

and with $b_k := \sum_{j=1}^{d(k)} ((G_k^k)^{-1} \tilde{b}^k)_j \phi_j^k$, one obtains the equation

$$A_n x_n = b_n \tag{8}$$

determining $x_n \in \mathcal{X}_n$ equivalently to (4). Here G_k^k is the Gramian matrix corresponding to $\{\phi_1^k, \ldots, \phi_{d(k)}^k\}$ wrt. the $_k$-inner product, i.e., $(G_k^k)_{i,j} := \langle \psi_i^k, \psi_i^k\rangle_k$. Note that

$$\langle u, v\rangle_{\mathcal{X}} = \langle A_k u, v\rangle_k\ , \quad \forall u, v \in \mathcal{X}_k\ ,$$

the operator A_k is symmetric and positive semidefinite both in $\langle\cdot,\cdot\rangle_{\mathcal{X}}$ and in $\langle\cdot,\cdot\rangle_k$, and (see also Lemma 2.1 in [21])

$$R_{k-1} A_k = A_{k-1} \mathcal{P}_{k-1} \quad \forall k \in \mathbb{N}\ .$$

Analogously to A_k, the operators Q_k and R_k can be described by premultiplication of coefficients wrt. the basis $\{\phi_1^k, \ldots, \phi_{d(k)}^k\}$ of $\mathcal{X}_k$ with some matrix Q^k or R^k, respectively, in $Q_k(\sum_{j=1}^{d(k)} u_j^k \phi_j^k) = \sum_{j=1}^{d(k)} (Q^k u^k)_j \phi_j^k$, $R_k(\sum_{j=1}^{d(k)} u_j^k \phi_j^k) = \sum_{j=1}^{d(k)} (R^k u^k)_j \phi_j^k$, where

$$Q^k = (C^k G_k^k C^{k\ T})^{-1} C^k G_{k+1}^{k+1}\ , \quad R^k = G_k^k C^k G_{k+1}^{k+1}\ ,$$

and C^k is some transformation matrix acting between the bases of the nested spaces $\mathcal{Y}_k \subseteq \mathcal{Y}_{k+1}$ such that

$$\psi_i^k = \sum_{j=1}^{d(k+1)} C_{i,j}^k \psi_j^{k+1}\ , \quad i = 1, \ldots d(k)\ .$$

If one just sets

$$\langle u, v\rangle_k := \langle (\mathcal{T}^*)^{-1} u, (\mathcal{T}^*)^{-1} v\rangle_{\mathcal{Y}}\ , \quad u, v \in \mathcal{X}_k\ , \tag{9}$$

i.e., for $u = \sum_{j=1}^{d(k)} u_j^k \phi_j^k,\ v = \sum_{j=1}^{d(k)} v_j^k \phi_j^k \in \mathcal{X}_k$,

$$\langle u, v\rangle_k := u^{k\ T} G_\psi^k v^k\ ,$$

which we will do in the following for simplicity of exposition, then R_k and Q_k coincide and $G_k^k = G_\psi^k$. In order to avoid the inversion of G_ψ^k, other discrete inner products equivalent to (9) may be used, though (cf. [5]).

The norm $||\cdot||_k$ is asymptotically stronger than the $\mathcal{X}$-norm, in the sense that

$$\min_{u\in\mathcal{X}_k}\frac{\|u\|_{\mathcal{X}}}{\|u\|_k}=\min_{v\in\mathcal{X}_k}\frac{\|A_k v\|_k}{\|v\|_{\mathcal{X}}}=:\underline{\underline{\gamma}}_k\sim\sqrt{\lambda_{\min}(\tilde{A}^k)}\to 0 \text{ as } k\to\infty\ .$$

Obviously the amplification factor of the data noise (when replacing y in (4) by y^δ) is just proportional to $1/\underline{\underline{\gamma}}_n$, whose convergence to infinity as $n\to\infty$ reflects the ill-posedness of the underlying problem, or, in other words, the smoothing property of the forward operator $\mathcal{T}$ and its adjoint $\mathcal{T}^*$, whose quantification on the discrete spaces $\mathcal{X}_k$ we will make use of in the following real sequences:

$$\overline{\overline{\gamma}}_k:=\max_{v\in\mathcal{X}_k}\frac{\|[I-Q_{k-1}]A_k v\|_k}{\|v\|_{\mathcal{X}}}$$

$$\underline{\gamma}_k:=\min_{v\in\mathcal{X}_{k-1}^{\perp_{\mathcal{X}}}}\frac{\|A_k v\|_k}{\|v\|_{\mathcal{X}}}\ ,$$

where $\mathcal{X}_{k-1}^{\perp_{\mathcal{X}}}\subseteq\mathcal{X}_k$ is the orthogonal complement of $\mathcal{X}_{k-1}$ in $\mathcal{X}_k$ with respect to the $\mathcal{X}$-topology, $\mathcal{X}_{k-1}^{\perp_{\mathcal{X}}}:=\{x\in\mathcal{X}_k\mid\langle x,v\rangle_{\mathcal{X}}=0\ \forall v\in\mathcal{X}_{k-1}\}$. They are related by

$$\underline{\underline{\gamma}}_k\le\underline{\gamma}_k\le\overline{\overline{\gamma}}_k\quad\forall k\in\mathbb{N}\ ,$$

because $\mathcal{X}_{k-1}^{\perp}\subseteq\mathcal{X}_k$ and since for any $v\in\mathcal{X}_{k-1}^{\perp}$

$$\langle[I-Q_{k-1}]A_k v,A_k v\rangle_k=\langle[I-Q_{k-1}]A_k v,v\rangle_{\mathcal{X}}=\langle A_k v,v\rangle_{\mathcal{X}}=\|A_k v\|_k^2$$

(see also Lemma 2.2 in [21]). Note that

$$\overline{\overline{\gamma}}_k\le\|\mathcal{T}^*[I-\mathcal{Q}_{k-1}]\|_{\mathcal{Y}\to\mathcal{X}}\ ,$$

i.e., $\overline{\overline{\gamma}}_k$ describes the approximation of functions in the "smooth" space $\mathcal{T}^*\mathcal{Y}$ (cf. (2)) by elements of $\mathcal{X}_{k-1}$.

3 Multigrid analysis

The multigrid method for differential or second kind integral equations is based on the combination of smoothing out the high frequency contribution of the residual by appropriate iterative methods, with one or more correction step(s) on a coarser grid. Here the smoothing operators S_k are defined e.g. by Jacobi, Gauss-Seidel, or SOR iterations. In the ill-posed situation the smoothing properties of these methods get lost. Therefore to define a MGM for ill-posed problems some considerations have to be made on how to replace or modify them appropriately. In [21] it is shown that an appropriate way of doing so is to use the operator

$$S_k:\mathcal{X}_k\to\mathcal{X}_k\ ,\quad S_k:=\xi_k[I-Q_{k-1}] \tag{10}$$

with $\xi_k\approx\|A_k^{-1}\|$, more precisely,

$$\xi_k\overline{\overline{\gamma}}_k^2\le 1\ , \tag{11}$$

$$\xi_k\underline{\gamma}_k^2\ge\underline{c}>0\ . \tag{12}$$

This operator is symmetric and positive semidefinite wrt. $\langle\cdot,\cdot\rangle_k$, and acts as a stabilized rough approximation of the inverse of A_k on the "high frequency part" $[I-Q_{k-1}]\mathcal{X}_k$ of $\mathcal{X}_k$, while on the "low frequency part" $Q_{k-1}\mathcal{X}_k=\mathcal{X}_{k-1}$ it is just zero. Therefore one could suppose that some improvement would be possible by, instead of leaving the low frequency part unchanged, applying some approximation of $(A_{k-1})^{-1}$, e.g., the operator $\rho_k R_k$, where

$$0\le\rho_k\|A_{k-1}\|_{\mathcal{X}\to\mathcal{X}}\le 1\ , \tag{13}$$

which results in the smoother

$$S_k : \mathcal{X}_k \to \mathcal{X}_k \ , \quad S_k := \xi_k[I - Q_{k-1}] + \rho_k R_k \tag{14}$$

Due to the ill-posedness of the underlying operator equation, such a modification, even with a more sophisticated approximation of $(A_{k-1})^{-1}$ e.g., by a Gauss-Seidel or a SOR iteration, cannot be expected to improve the order of magnitude of the resulting contraction factor and condition number estimates, though. The reason why we nevertheless propose this modification — that does not incorporate a significant amount of additional computational cost and yields convergence factors at least as good as those obtained with (10) — is the fact that it enables us to guarantee, that the condition of the corresponding multigrid preconditioned operators never gets worse than the condition of the unpreconditioned operators A_n.

It can be easily seen that the difference operator

$$K_k : \mathcal{X}_k \to \mathcal{X}_k \ , \quad K_k := I - S_k A_k = I - \xi_k[I - Q_{k-1}]A_k - \rho_k A_{k-1}\mathcal{P}_{k-1}$$

is symmetric wrt. $\langle\cdot,\cdot\rangle_{\mathcal{X}}$, with spectrum contained in $[0,1]$: For arbitrary $v \in \mathcal{X}_k$ one gets

$$\begin{aligned}\langle K_k v, v\rangle_{\mathcal{X}} &= \|v\|_{\mathcal{X}}^2 - \xi_k\langle [I - Q_{k-1}]A_k v, v\rangle_{\mathcal{X}} - \rho_k\langle A_{k-1}\mathcal{P}_{k-1}v, v\rangle_{\mathcal{X}} \\ &= \|v\|_{\mathcal{X}}^2 - \xi_k \,\|[I - Q_{k-1}]A_k v\|_k^2 - \rho_k \,\|A_{k-1}^{\frac{1}{2}}\mathcal{P}_{k-1}v\|_{\mathcal{X}}^2 \\ &\left\{\begin{array}{ll} \leq & \|v\|_{\mathcal{X}}^2 \\ \geq & (1 - \xi_k\overline{\overline{\gamma}}_k^2 - \rho_k\,\|A_{k-1}\|_{\mathcal{X}\to\mathcal{X}})\,\|v\|_{\mathcal{X}}^2 \end{array}\right.\end{aligned} \tag{15}$$

Moreover, it can be shown like in [21] (Lemma 2.5) that K_k is a contraction on $\mathcal{X}_{k-1}^{\perp_{\mathcal{X}}}$: Since $(I - \mathcal{P}_{k-1})Q_{k-1} = 0$, we have, for any $v \in \mathcal{X}_k$,

$$\begin{aligned}\langle K_k[I - \mathcal{P}_{k-1}]v, [I - \mathcal{P}_{k-1}]v\rangle_{\mathcal{X}} &= \|[I - \mathcal{P}_{k-1}]v\|_{\mathcal{X}}^2 - \xi_k\,\|A_k[I - \mathcal{P}_{k-1}]v\|_k \\ &\leq (1 - \xi_k\underline{\gamma}_k^2)\,\|[I - \mathcal{P}_{k-1}]v\|_{\mathcal{X}}^2\end{aligned} \tag{16}$$

and therefore, with $\tilde{K}_k := I - \xi_k[I - Q_{k-1}]A_k$,

$$\begin{aligned}\|K_k[I - \mathcal{P}_{k-1}]\|_{\mathcal{X}\to\mathcal{X}}^2 &\leq \|\tilde{K}_k\|_{\mathcal{X}\to\mathcal{X}}\,\|\tilde{K}_k^{\frac{1}{2}}[I - \mathcal{P}_{k-1}]\|_{\mathcal{X}\to\mathcal{X}}^2 \\ &= \sup_{v\in\mathcal{X}_k}\frac{\langle K_k[I-\mathcal{P}_{k-1}]v, [I-\mathcal{P}_{k-1}]v\rangle_{\mathcal{X}}}{\|v\|_{\mathcal{X}}^2} \\ &\leq 1 - \xi_k\underline{\gamma}_k^2 \ ,\end{aligned} \tag{17}$$

where we have used the facts that $K_k[I - \mathcal{P}_{k-1}] = \tilde{K}_k[I - \mathcal{P}_{k-1}]$, and that $\tilde{K}_k$ is nonnegative definite (see (15)). The latter estimate is sharp in the sense that, due to the general structure of the operator A_k, there may be some $v_0 \in \mathcal{X}_k$ such that $A_k\tilde{K}_k^{\frac{1}{2}}[I - \mathcal{P}_{k-1}]v_0 \in \mathcal{X}_{k-1}$ and therefore

$$\begin{aligned}\|K_k[I - \mathcal{P}_{k-1}]v_0\|_{\mathcal{X}}^2 &= \langle (I - \xi_k[I - Q_{k-1}]A_k)\tilde{K}_k^{\frac{1}{2}}[I - \mathcal{P}_{k-1}]v_0, \tilde{K}_k^{\frac{1}{2}}[I - \mathcal{P}_{k-1}]v_0\rangle_{\mathcal{X}} \\ &= \langle \tilde{K}_k[I - \mathcal{P}_{k-1}]v_0, [I - \mathcal{P}_{k-1}]v_0\rangle_{\mathcal{X}} \\ &= (1 - \xi_k(\tfrac{\|A_k[I-\mathcal{P}_{k-1})v_0\|_k}{\|[I-\mathcal{P}_{k-1}]v_0\|_{\mathcal{X}}})^2)\,\|[I - \mathcal{P}_{k-1})v_0\|_{\mathcal{X}}^2.\end{aligned}$$

This also makes obvious that in general no reduction of the norm of K_k (and, with that, improvement of the overall error estiates) can be achived by doing more than one pre- and post-smoothing step per level.

We here a consider a multigrid operator B_k defined by the symmetric recursion

Algorithm 1. (MGM)
Set $B_0 := (A_0)^{-1}$. Assume that B_{k-1} has been defined and define B_k for $b_k \in \mathcal{X}_k$ as follows:

pre-smoothing:
Set $x := S_k b_k$.

coarse grid correction:
Set $s := 0$.
For $l = 1 \ldots m(k)$
 Set $s := s + B_{k-1}\Big(R_{k-1}(b_k - A_k x) - A_{k-1}s\Big)$.
Set $x := x + s$.

post-smoothing:
Set $x := x + S_k(b_k - A_k x)$.

Set $B_k b_k := x$.

where $m(k) \geq 1$ describes the cycle regime — $m(k) \equiv 1$ corresponds to the V-cycle, $m(k) \equiv 2$ to the W-cycle and $m(k) = 2 - \mathrm{mod}(k, 2)$ to the alternating cycle.

Rewriting this algorithm in matrix notation, we denote by b^k the vector of coefficients of some element $b_k \in \mathcal{X}_k$ wrt. $\{\phi_1, \ldots, \phi^k_{d(k)}\}$, i.e., $b_k = \sum_{j=1}^{d(k)} b^k_j \phi^k_j$. Then the coefficients of $B_k b_k$ wrt. $\{\phi_1, \ldots, \phi^k_{d(k)}\}$ are given by $B^k b^k$, i.e., $B_k b_k = \sum_{j=1}^{d(k)} (B^k b^k)_j \phi_j$ with B^k defined by

Algorithm 2. (MGM; matrix formulation)
Set $B^0 := (A^0)^{-1}$. Assume that B^{k-1} has been defined and define B^k for $b^k \in \mathbb{R}^{d(k)}$ as follows:
 Set $x^k := \xi_k(I^k - C^{k-1\ T} Q^{k-1}) b^k$.
 Set $s^{k-1} := 0$.
 For $l = 1 \ldots m(k)$
  Set $s^{k-1} := s^{k-1} + B^{k-1}\Big(R^{k-1}(b^k - A^k x^k) - A^{k-1}s^{k-1}\Big)$.
 Set $x^k := x^k + C^{k-1\ T} s^{k-1}$.
 Set $x^k := x^k + \xi_k(I^k - C^{k-1\ T} Q^{k-1})(b^k - A^k x^k)$.
 Set $B^k b^k := x^k$.

where I^k is the $d(k) \times d(k)$ identity matrix and the matrices $A^k, C^{k-1}, Q^{k-1}, R^{k-1}$ are defined in the previous section.

It can be easily checked that the corresponding error operator $E_k := (I - B_k A_k)\mathcal{P}_k$ obeys the recursion

$$E_k = K_k\Big([I - \mathcal{P}_{k-1}] + (E_{k-1})^{m(k)}\mathcal{P}_{k-1}\Big)K_k \tag{18}$$

with

$$E_0 = 0 \ .$$

Therefore, by an inductive argument, E_k is symmetric wrt. $\langle\cdot,\cdot\rangle_{\mathcal{X}}$ and nonnegative definite. Since for any $v \in X$

$$\begin{aligned}
\langle E_k v, v\rangle_{\mathcal{X}} &= \|[I - \mathcal{P}_{k-1}]K_k v\|^2_{\mathcal{X}} + \|(E_{k-1})^{m(k)}\mathcal{P}_{k-1}K_k v\|^2_{\mathcal{X}} \\
&\leq \|[I - \mathcal{P}_{k-1}]K_k v\|^2_{\mathcal{X}} + \|E_{k-1}\|^{m(k)}_{\mathcal{X}\to\mathcal{X}}\|\mathcal{P}_{k-1}K_k v\|^2_{\mathcal{X}} \\
&= (1 - \|E_{k-1}\|^{m(k)}_{\mathcal{X}\to\mathcal{X}})\|[I - \mathcal{P}_{k-1}]K_k v\|^2_{\mathcal{X}} + \|E_{k-1}\|^{m(k)}_{\mathcal{X}\to\mathcal{X}}\|K_k v\|^2_{\mathcal{X}} \ ,
\end{aligned}$$

and by (15), (17), the contraction factors, i.e., norms of the error operators can be estimated by

$$\|E_k\|_{\mathcal{X}\to\mathcal{X}} \leq \epsilon_k \tag{19}$$

where $(\epsilon_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}$ is defined by the recursion

$$\epsilon_k = \xi_k \underline{\gamma}_k^2\ \epsilon_{k-1}^{m(k)} + (1 - \xi_k \underline{\gamma}_k^2) \ . \tag{20}$$

The contraction factor estimation is therefore reduced to the analysis of a real sequence, which, in a quite straightforward manner yields

Proposition 1. *Let $(B_k)_{1\le k\le n}$ be defined by 1 with (11), (13) and assume that (12) holds with some constant $\underline{c}>0$ independent of k and n.*

If $\underline{c} > \sqrt{\frac{1}{2}}$ then the alternating cycle ($m(k) = 2 - mod(k,2)$) and, if $\underline{c} > \frac{1}{2}$ the W-cycle ($m(k) \equiv 2$) converges with a contraction number independent of the number of levels, i.e., for any $n \in \mathbb{N}$

$$\|[I - B_n A_n]\mathcal{P}_n\|_{\mathcal{X}} \le c < 1 \ , \tag{21}$$

with c independent of n.

Else, for any $\underline{c}$ in (12) there exists an $\underline{m}$ independent of n such that setting $m(k) \ge \underline{m}$ in Algorithm 1 yields a level-independent convergence factor c in (21).

For arbitrary cycle regime $m(k) \ge 1$, i.e., especially also for the V-cycle, there holds

$$\|[I - B_n A_n]\mathcal{P}_n\|_{\mathcal{X}} \le 1 - \underline{c}^n \ . \tag{22}$$

*Proof.*First of all one can immediately conclude by induction from (20) and $0 < \xi_k\underline{\gamma}_k^2 \le \xi_k\overline{\overline{\gamma}}_k^2 \le 1$, that the sequence $(\epsilon_k)_{k\in\{1,\dots,n\}}$ is contained in $[0,1)$. Rearranging terms we get

$$\epsilon_k = (1 - \xi_k\underline{\gamma}_k^2)(1 - \epsilon_{k-1}^{m(k)}) + \epsilon_{k-1}^{m(k)} \le (1 - \underline{c})(1 - \epsilon_{k-1}^{m(k)}) + \epsilon_{k-1}^{m(k)} \tag{23}$$

Subtracting from one on both sides yields

$$1 - \epsilon_k \ge \underline{c}(1 - \epsilon_{k-1}^{\underline{m}}) \ , \tag{24}$$

and, with $\overline{\epsilon} := \max_{k\in\{1,\dots,n\}} \epsilon_k$,

$$1 - \overline{\epsilon} \ge \underline{c}(1 - \overline{\epsilon}^{\underline{m}}) \ ,$$

hence,

$$\frac{1}{\underline{c}} \ge \sum_{i=0}^{\underline{m}-1} \overline{\epsilon}^i \ge \underline{m}\,\overline{\epsilon}^{\underline{m}-1}$$

which, if $\underline{m} > \frac{1}{\underline{c}}$ yields an upper bound smaller than one and independent of n for $\overline{\epsilon}$.

For arbitrary $m(k)$, i.e., if we can only set $\underline{m} := 1$ in (24), we get at least

$$1 - \epsilon_k \ge \underline{c}^k(1 - \epsilon_0) = \underline{c}^k$$

and therefore (22).

It remains to consider the alternating cycle. Here we have, by (23)

$$\epsilon_{2l} \le \underline{c}^2\epsilon_{2(l-1)}^2 + 1 - \underline{c}^2$$

for all $l \in \{1\dots,[n/2]\}$ and

$$\epsilon_{2l+1} \le \underline{c}^3\epsilon_{2l-1}^2 + 2\underline{c}^2(1-\underline{c})\epsilon_{2l-1} + \underline{c}(1-\underline{c})^2 + (1-\underline{c})$$

for all $l \in \{0,\dots,[(n-1)/2]\}$. Taking the maximum over l on both sides of these two inequalities we get, for $\overline{\epsilon}_0 := \max\{\epsilon_k \mid k \in \{1,\dots,n\} \wedge \mathrm{mod}(k,2) = 0\}$, $\overline{\epsilon}_1 := \max\{\epsilon_k \mid k \in \{1,\dots,n\} \wedge \mathrm{mod}(k,2) = 1\}$, respectively,

$$0 \le \underline{c}^2\overline{\epsilon}_0^2 - \overline{\epsilon}_0 + 1 - \underline{c}^2 = \underline{c}^2(\overline{\epsilon}_0 - \tfrac{1+|1-2\underline{c}^2|}{2\underline{c}^2})(\overline{\epsilon}_0 - \tfrac{1-|1-2\underline{c}^2|}{2\underline{c}^2})$$

$$\begin{aligned} 0 &\le \underline{c}^3\overline{\epsilon}_1^2 - (1 - 2\underline{c}^2(1-\underline{c}))\overline{\epsilon}_1 + \underline{c}(1-\underline{c})^2 + (1-\underline{c}) \\ &= \underline{c}^3(\overline{\epsilon}_1 - \tfrac{1-\underline{c}^2(1-\underline{c})+|1-2\underline{c}^2|}{2\underline{c}^3})(\overline{\epsilon}_1 - \tfrac{1-\underline{c}^2(1-\underline{c})-|1-2\underline{c}^2|}{2\underline{c}^3}) \end{aligned}$$

and therefore, since $\overline{\epsilon}_0, \overline{\epsilon}_1 < 1$,

$$\overline{\epsilon}_0 \le \min\{\tfrac{1+|1-2\underline{c}^2|}{2\underline{c}^2}, \tfrac{1-|1-2\underline{c}^2|}{2\underline{c}^2}\} = \begin{cases} \frac{1}{\underline{c}^2} - 1 < 1 & \text{if } \underline{c} > \sqrt{\frac{1}{2}} \\ 1 & \text{if } \underline{c} \le \sqrt{\frac{1}{2}} \end{cases} \ ,$$

$$\overline{\epsilon}_1 \leq \min\{\tfrac{1-\underline{c}^2(1-\underline{c})+|1-2\underline{c}^2|}{2\underline{c}^3}, \tfrac{1-\underline{c}^2(1-\underline{c})-|1-2\underline{c}^2|}{2\underline{c}^3}\} = \begin{cases} 1-\frac{2\underline{c}^2-1}{\underline{c}^3} < 1 & \text{if } \underline{c} > \sqrt{\frac{1}{2}} \\ 1 & \text{if } \underline{c} \leq \sqrt{\frac{1}{2}} \end{cases} .$$

□

The contraction factor estimate (21) has, to some extent, the usual consequences in application of the multigrid operators B_k as preconditioners of (conjugate gradient) iterations or in a full multigrid method:

3.1 Multilevel Preconditioning

If the assertion (21) of Proposition 1 holds, the condition number of $B^k A^k$, that is crucial for the speed of convergence of a preconditioned CG or other iterative solution method, can be estimated by

$$\text{cond}(B^n A^n) = |||B^n A^n|||_n |||(B^n A^n)^{-1}|||_n \leq \frac{1}{1-\|I-B_n A_n\|_{\mathcal{X}}} \leq \frac{1}{1-c} ,$$

where we measure convergence of the computed coefficient vector $\tilde{x}^n$ to the exact x^n in the energy type norm $|||\cdot|||_n := \|(\tilde{A}^n)^{1/2}\cdot\|_{\mathbb{R}^{d(n)}}$, since then

$$|||\tilde{x}^n - x^n|||_n = \|\tilde{x}_n - x_n\|_{\mathcal{X}} ,$$

where $\tilde{x}_n := \sum_{j=1}^{d(k)} \tilde{x}_j^n \phi_j^n$ is our approximation for the solution x_n of (8).

By the following worst case estimate we can exclude the pathological case that the operator condition (and, with that, the contraction factors of iterative solution methods) might be asymptotically deteriorated instead of improved by MGM preconditioning; note that it holds for arbitrary ξ_k satisfying (11) and bounded away from zero, hence also for "bad" estimates ξ_k of $(\overline{\overline{\gamma}}_k)^{-2}$.

Lemma 1. *Assume that (11), (13), as well as*

$$\xi_k \overline{\overline{\gamma}}_k^2 + \rho_k \|A_{k-1}\|_{\mathcal{X}\to\mathcal{X}} \leq C_1 < 2$$

and

$$\rho_k \geq \underline{\rho} > 0 , \tag{25}$$

$$\xi_k \geq \underline{\xi} > 0 , \tag{26}$$

holds with constants $C_1, \underline{\xi}, \underline{\rho}$ independent of k and n. Then, for $(B_k)_{k=1,\ldots,n}$ defined by Algorithm 1 with any choice of $m(k)$ — especially for the V-cycle —

$$\mathit{cond}(B^n A^n) \leq C\, \mathit{cond}(\tilde{A}^n) \tag{27}$$

for all $n \in \mathbb{N}$, with $C = \frac{C_\psi^2}{(2-C_1)\min\{\underline{\xi},\underline{\rho}\}}$ independent of n.

Proof. The proof of (27) uses the V-cycle convergence proof techniques of [6] — unfortunately without achieving comparably good results in our ill-posed situation.

Since obviously $\|E_k\|_{\mathcal{X}\to\mathcal{X}} \leq 1$ for all $k \in \mathbb{N}$, by (20) it suffices to consider the case $m(k) \equiv 1$ (V-cycle) for which a simple inductive argument shows

$$I - B_k A_k = (I - B_k^{ns} A_k)^{*_{\mathcal{X}}} (I - B_k^{ns} A_k) ,$$

where B_k^{ns} is the nonsymmetric MG operator defined by setting $m(k) \equiv 1$ and omitting the post smoothing step in Algorithm 1.

We use the notation

$$\begin{array}{llll} \tilde{E}_k : \mathcal{X} \to \mathcal{X} \ , & \tilde{E}_k &=& (I - B_k^{ns} A_k \mathcal{P}_k)^{*\mathcal{X}} \ , \quad k \in \mathbb{N} \\ & \tilde{E}_0 &=& I - \mathcal{P}_0 \\ & \tilde{E}_{-1} &=& I \end{array}$$

$$\begin{array}{llll} T_k : \mathcal{X} \to \mathcal{X}_k \ , & T_k &=& (I - K_k)\mathcal{P}_k = S_k A_k \mathcal{P}_k \ , \quad k \in \mathbb{N} \\ & T_0 &=& \mathcal{P}_0 \end{array}$$

which yields

$$\mathrm{cond}(B^n A^n) = \frac{1}{1 - \|\tilde{E}_n \mathcal{P}_n\|^2_{\mathcal{X}\to\mathcal{X}}}$$

and the recursion

$$\tilde{E}_k = (I - T_k)\tilde{E}_{k-1} \ , \quad k \in \mathbb{N}_0 \ .$$

For any $u \in \mathcal{X}_n$

$$\begin{array}{lll} \|u\|^2_{\mathcal{X}} - \|\tilde{E}_n u\|^2_{\mathcal{X}} &=& \sum_{k=0}^n \|\tilde{E}_{k-1}u\|^2_{\mathcal{X}} - \|\tilde{E}_k u\|^2_{\mathcal{X}} \\ &=& (2 - C_1)\sum_{k=0}^n \langle T_k \tilde{E}_{k-1}u, \tilde{E}_{k-1}u\rangle_{\mathcal{X}} + \langle T_k(C_1 I - T_k)\tilde{E}_{k-1}u, \tilde{E}_{k-1}u\rangle_{\mathcal{X}} \\ &\geq& (2 - C_1)\sum_{k=0}^n \langle T_k \tilde{E}_{k-1}u, \tilde{E}_{k-1}u\rangle_{\mathcal{X}} \end{array} \tag{28}$$

where we have used the fact that, by (15), the operator $T_k(C_1 I - T_k)$ is symmetric and nonnegative definite wrt. $\langle\cdot,\cdot\rangle_{\mathcal{X}}$.

We further estimate, for $k \in \{1, \ldots, n\}$, $v := \mathcal{P}_k \tilde{E}_{k-1}u \in \mathcal{X}_k$

$$\begin{array}{lll} \langle T_k v, v\rangle_{\mathcal{X}} &=& \xi_k \|[I - Q_{k-1}]A_k v\|^2_k + \rho_k \|R_{k-1}A_k v\|^2_{k-1} \\ &\geq& \min\{\xi_k, \rho_k\}\|A_k^{\frac{1}{2}} v\|^2_{\mathcal{X}} \\ &\geq& \min\{\underline{\xi}, \underline{\rho}\}\underline{\underline{\gamma}}_k^2 \|[I - \mathcal{P}_{k-1}]v\|^2_{\mathcal{X}} \ , \end{array} \tag{29}$$

where we have used the identity of $R_{k-1}$with Q_{k-1} and of $\|\cdot\|_{k-1}$ with $\|\cdot\|_k$ (that can be replaced by respective estimates in the case of a discrete norm other than but equivalent to (9)) in the first inequality. Inserting this into (28) and taking into account the identities

$$\sum_{k=1}^n \|[I - \mathcal{P}_{k-1}]\mathcal{P}_k \tilde{E}_{k-1}u\|^2_{\mathcal{X}} = \sum_{k=1}^n \|[I - \mathcal{P}_{k-1}]\mathcal{P}_k u\|^2_{\mathcal{X}} = \|[I - \mathcal{P}_0]u\|^2_{\mathcal{X}} \ ,$$

$$\langle T_0 \tilde{E}_{-1}u, \tilde{E}_{-1}u\rangle_{\mathcal{X}} = \|\mathcal{P}_0 u\|^2_{\mathcal{X}} \ ,$$

we get

$$\|\tilde{E}_n u\|^2_{\mathcal{X}} \leq \Big(1 - (2 - C_1)\min\{\underline{\xi}, \underline{\rho}\}\Big) \min_{1\leq k\leq n} \underline{\underline{\gamma}}_k^2 \|u\|^2_{\mathcal{X}} \ ,$$

which, since $(\underline{\underline{\gamma}}_k)_{k\in\mathbb{N}}$ is monotonically decreasing and $\underline{\underline{\gamma}}_n^2 \leq C_\psi^2 \lambda_{\min}(\tilde{A}^n)$, yields the assertion. □

Remark 1. The crucial point in the proof, why we do not obtain a level-independent contraction factor here, lies in the inequality (29). Here we would need to have an estimate of the form $\langle T_k v, v\rangle_{\mathcal{X}} \geq c\|[I - \mathcal{P}_{k-1}]v\|^2_{\mathcal{X}}$ with c independent of v and k. This, however, does not seem to be possible, since there might be some $v \in \mathcal{X}_k$ with $[I - \mathcal{P}_{k-1}]v \neq 0$ and $A_k v \in \mathcal{X}_{k-1}$, and therefore $\langle T_k v, v\rangle_{\mathcal{X}} = \rho_k \|A_k v\|_{k-1}$, which in general can only be bounded from below by $\rho_k \underline{\underline{\gamma}}_k^2 \|v\|_{\mathcal{X}}$.

Estimates like (27) for multigrid methods like Algorithm 1 in the context of an ill-posed operator equation (1) can analogously be derived for conventional smoothers (such as Jacobi iteration) inducing a strictly positive definite S_k. Note however, that for S_k according to (14) and as soon as $\underline{c}$ is sufficiently close to one, the estimate (22) yields an asymptotically better condition number of $B^n A^n$ than (27), which also becomes obvious in our numerical experiments of Section 4.

Similarly to the analysis of B_n we obtain estimates for the additive multilevel preconditioner

$$B_n^a := \sum_{k=1}^{n} \xi_k [I - Q_{k-1}]\bar{R}_k + \rho_0 \bar{R}_0 \ , \tag{30}$$

cf. [7], whose structure allows a completely parallel and therefore efficient implementation.

Corollary 1. *For B^{an} being the matrix representation of B_n^a as defined in (30) with (11), (26) for all $k \in \{1 \ldots, n\}$, and $0 \le \rho_0 \|A_0\|_{\mathcal{X} \to \mathcal{X}} \le 1$ there holds*

$$cond(B^{an} A^n) \le C(n+1) cond(\tilde{A}^n) \ . \tag{31}$$

If, in addition, (12) holds then

$$cond(B^{an} A^n) \le \tfrac{n+1}{\rho_0 \underline{\underline{\gamma}}_0^2} (\tfrac{2-\underline{c}}{\underline{c}})^{2n} \ . \tag{32}$$

Proof. The operator $B_n^a A_n$ is obviously symmetric wrt. $\langle \cdot, \cdot \rangle_{\mathcal{X}}$ and positive semidefinite with

$$\begin{aligned} \langle B_n^a A_n u, u \rangle_{\mathcal{X}} &= \textstyle\sum_{k=1}^n \xi_k \|[I - Q_{k-1}] A_k \mathcal{P}_k u\|_k^2 + \rho_0 \|A_0 \mathcal{P}_0 u\|_0^2 \\ &= \textstyle\sum_{k=1}^n \xi_k \|[\bar{Q}_k - \bar{Q}_{k-1}] A_n u\|_k^2 + \rho_0 \|\bar{Q}_0 A_n u\|_0^2 \\ &\left\{ \begin{array}{ll} \le & \sum_{k=1}^n \xi_k \overline{\overline{\gamma}}_k^2 \|\mathcal{P}_k u\|_{\mathcal{X}}^2 + \rho_0 \|A_0\| \, \|\mathcal{P}_0 u\|_{\mathcal{X}}^2 \le (n+1) \|u\|_{\mathcal{X}}^2 \\ \ge & \min\{\underline{\xi}, \rho_0\} \|A_n u\|_n^2 \ge \min\{\underline{\xi}, \rho_0\} \underline{\underline{\gamma}}_n^2 \|u\|_{\mathcal{X}}^2 \end{array} \right. \end{aligned} \tag{33}$$

for any $u \in \mathcal{X}_n$.

Moreover, if (12) holds then, abbreviating $c_k := \rho_0 \underline{\underline{\gamma}}_0^2 (\frac{\underline{c}}{2-\underline{c}})^{2k}$, we get $\langle B_0^a A_0 v, v \rangle_{\mathcal{X}} = c_0 \|v\|_{\mathcal{X}}^2$ for all $v \in \mathcal{X}_0$, and, under the induction hypothesis $\langle B_{n-1}^a A_{n-1} v, v \rangle_{\mathcal{X}} \ge c_{n-1} \|v\|_{\mathcal{X}}^2$ for all $v \in \mathcal{X}_{n-1}$, we obtain, for any $u \in \mathcal{X}_n$, $\theta \in [0,1]$ and $q > 0$:

$$\begin{aligned} \langle B_n^a A_n u, u \rangle_{\mathcal{X}} &= \langle B_{n-1}^a A_{n-1} \mathcal{P}_{n-1} u, \mathcal{P}_{n-1} u \rangle_{\mathcal{X}} + \xi_n \|[I - Q_{n-1}] A_n u\|_n^2 \\ &\ge c_{n-1} \|\mathcal{P}_{n-1} u\|_{\mathcal{X}}^2 + \theta \xi_n \|[I - Q_{n-1}] A_n u\|_n^2 \\ &\ge c_{n-1} \|\mathcal{P}_{n-1} u\|_{\mathcal{X}}^2 + \theta \Big(\xi_n \|A_n [I - \mathcal{P}_{n-1}] u\|_n^2 + 2 \underbrace{\xi_n \langle A_n [I - \mathcal{P}_{n-1}] u, \mathcal{P}_{n-1} u \rangle_{\mathcal{X}}}_{\langle K_n [I - \mathcal{P}_{n-1}] u, \mathcal{P}_{n-1} u \rangle_{\mathcal{X}}} \Big) \\ &\ge \theta (\underline{c} - \sqrt{1 - \underline{c}} \, q) \|[I - \mathcal{P}_{n-1}] u\|_{\mathcal{X}}^2 + (c_{n-1} - \theta \sqrt{1 - \underline{c}} \, \tfrac{1}{q}) \|\mathcal{P}_{n-1} u\|_{\mathcal{X}}^2 \ , \end{aligned}$$

where we have used (12), (17), and the simple relation $2ab \le qa^2 + q^{-1} b^2 \ \forall a, b \in \mathbb{R}, q > 0$ in the last inequality. Setting $q := \frac{\underline{c}}{2\sqrt{1-\underline{c}}}$ and $\theta := c_{n-1} \frac{2\underline{c}}{(2-\underline{c})^2}$ yields $\langle B_n^a A_n u, u \rangle_{\mathcal{X}} \ge c_n \|u\|_{\mathcal{X}}^2$ and therefore, together with the upper estimate of (33), the assertion (32).

3.2 The Full Multigrid Method as a Regularization Method

For the full multigrid method (FMGM), defined by the nested iteration

Algorithm 3. (FMGM)
Set $x_0 := A_0^{-1} b_0 =: x_0^{(0)}$, $q(0) := 0$.
For $k = 1 \ldots n$
Set $x_k^{(0)} := x_{k-1}^{(q(k-1))}$.
For $l = 1 \ldots q(k)$
Set $x_k^{(l)} = x_k^{(l-1)} + B_k (b_k - A_k x_k^{(l-1)})$.
Set $\tilde{x}_n := x_n^{(q(n))}$.

where $b_k := \bar{R}_k b_n$ and $q(k) \geq 1$, we get

$$
\begin{array}{rcl}
\tilde{x}_n & = & x_n^{(q(n))} \\
& = & (I - B_n A_n)^{(q(n))} x_{n-1}^{(q(n-1))} + \sum_{l=0}^{q(n)-1} (I - B_n A_n)^l B_n b_n \\
& = & (I - B_n A_n)^{q(n)} (I - B_{n-1} A_{n-1})^{q(n-1)} \cdots (I - B_1 A_1)^{q(1)} A_0^{-1} b_0 \\
& & + \sum_{k=1}^{n} (I - B_n A_n)^{q(n)} \cdots (I - B_{k+1} A_{k+1})^{q(k+1)} \sum_{l=0}^{q(k)-1} (I - B_k A_k)^l B_k b_k \ ,
\end{array} \tag{34}
$$

and, for the difference between $\tilde{x}_n$ and the solution x_n of (8)

$$
\tilde{x}_n - x_n = \sum_{k=1}^{n} (I - B_n A_n)^{q(n)} \cdots (I - B_k A_k)^{q(k)} [I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger \ ,
$$

where $x^\dagger$ is the best approximate solution of (1). Therefore, if assertion (21) of Proposition 1 holds, splitting the sum on the right hand side into two parts and using the Cauchy Schwarz inequality be obtain

$$
\begin{array}{l}
\|\tilde{x}_n - x_n\|_{\mathcal{X}} \leq \sum_{k=1}^{n} c^{q(k)+\ldots+q(n)} \|[I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger\|_{\mathcal{X}} \\
\leq \sum_{k=1}^{n/2} c^{n-k+1} \|[I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger\|_{\mathcal{X}} + \sqrt{\sum_{k=n/2+1}^{n} c^{2(n-k+1)}} \sqrt{\sum_{k=n/2+1}^{n} \|[I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger\|_{\mathcal{X}}^2} \\
\leq c^{n/2} \frac{1}{1-c} \|x^\dagger\| + \frac{1}{1-c} \|[\mathcal{P}_{n/2} - \mathcal{P}_n] x^\dagger\|_{\mathcal{X}} \ ,
\end{array} \tag{35}
$$

since $x_n = \mathcal{P}_0 x^\dagger + \sum_{k=1}^{n} [I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger$ and this sum is orthogonal in $\mathcal{X}$. With

$$
\|x_n - x^\dagger\|_{\mathcal{X}} = \|[I - \mathcal{P}_n] x^\dagger\|_{\mathcal{X}}
$$

this yields the overall estimate

$$
\begin{array}{rcl}
\|\tilde{x}_n - x^\dagger\|_{\mathcal{X}} & \leq & \frac{1}{1-c} (c^{n/2} \|x^\dagger\| + \|[I - \mathcal{P}_{n/2}] x^\dagger\|_{\mathcal{X}} + (2 - c) \|[I - \mathcal{P}_n] x^\dagger\|_{\mathcal{X}}) \\
& & \to 0 \text{ as } n \to \infty \ .
\end{array}
$$

Under a source condition (2) one can use the smoothing properties of $(\mathcal{T}^* \mathcal{T})^\mu$ in combination with approximation properties of the spaces $\mathcal{Y}_k$ to $\mathcal{Y}$ to obtain faster convergence of the projections $[I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger$ to zero: Indeed, if we consider projections onto spaces of piecewise polynomial functions (such as finite elements or splines) on a mesh of size h_n, in Sobolev spaces $\mathcal{X} := H^s$, $\mathcal{Y} = H^t$, where $\mathcal{T}^*$ maps not only H^t- but even H^{t-p}-functions into $\mathcal{X}$, (i.e., $p > 0$ quantifies the degree of smoothing that is done by an application of $\mathcal{T}^*$,) and under a source condition with $\mu = \frac{1}{2}$, i.e., if

$$
x^\dagger = \mathcal{T}^* w \text{ for some } w \in \mathcal{Y} \ ,
$$

we obtain

$$
\begin{array}{rcl}
\|[I - \mathcal{P}_{k-1}] \mathcal{P}_k x^\dagger\|_{\mathcal{X}} & \leq & \|[I - \mathcal{P}_{k-1}] \mathcal{T}^* w\|_{\mathcal{X}} + \|[I - \mathcal{P}_k] \mathcal{T}^* w\|_{\mathcal{X}} \\
& = & \min_{v \in \mathcal{Y}_{k-1}} \|\mathcal{T}^* w - \mathcal{T}^* v\|_{\mathcal{X}} + \min_{v \in \mathcal{Y}_k} \|\mathcal{T}^* w - \mathcal{T}^* v\|_{\mathcal{X}} \\
& \leq & \|\mathcal{T}^*\|_{H^{t-p} \to \mathcal{X}} \left(\|[I - \mathcal{Q}_{k-1}]\|_{H^t \to H^{t-p}} + \|[I - \mathcal{Q}_k]\|_{H^t \to H^{t-p}} \right) \|w\|_{\mathcal{Y}} \\
& \leq & C (h_k)^p \|w\|_{\mathcal{Y}} \ ,
\end{array} \tag{36}
$$

(see also [19]). This together with the first line of (35), and a geometrically decreasing mesh size

$$
h_k \sim \sigma^k
$$

for some $\sigma \in (0, 1)$, yields

$$
\|\tilde{x}_n - x_n\|_{\mathcal{X}} \leq C \sum_{k=1}^{n} c^{n-k+1} \sigma^{pk} \|w\|_{\mathcal{Y}} \leq C n \max\{c, \sigma^p\}^{n+1} \|w\|_{\mathcal{Y}} \ ;
$$

If $c < \sigma^p$, one even gets the order optimal estimate

$$
\|\tilde{x}_n - x_n\|_{\mathcal{X}} \leq C \sigma^{pn} \|w\|_{\mathcal{Y}} = O((h_n)^p) \ .
$$

Noisy data y^δ with noise level δ in

$$\|y^\delta - y\| \leq \delta \tag{37}$$

inserted into (4) instead of y, lead to a perturbed version

$$A_n x_n^\delta = b_n^\delta \tag{38}$$

of (8), with $b_n^\delta = \sum_{j=1}^{d(n)} ((G_n^n)^{-1} \langle y^\delta, \psi^n \rangle_{\mathcal{Y}})_j \phi_j^n$ and

$$\|b_n^\delta - b_n\|_n \leq \delta \ .$$

Denoting by $\tilde{x}_n^\delta$ the FMGM solution of (38), we can estimate

$$\|\tilde{x}_n^\delta - x^\dagger\|_{\mathcal{X}} \leq \|\tilde{x}_n^\delta - \tilde{x}_n\|_{\mathcal{X}} + \|\tilde{x}_n - x^\dagger\|_{\mathcal{X}} \ ,$$

where, by (34)

$$\begin{aligned}
&\tilde{x}_n^\delta - \tilde{x}_n \\
&= \Big((I - B_n A_n)^{q(n)} (I - B_{n-1} A_{n-1})^{q(n-1)} \cdots (I - B_1 A_1)^{q(1)} \mathcal{P}_0 \\
&\quad + \textstyle\sum_{k=1}^n (I - B_n A_n)^{q(n)} \cdots (I - B_{k+1} A_{k+1})^{q(k+1)} \underbrace{\sum_{l=0}^{q(k)-1} (I - B_k A_k)^l B_k A_k \, \mathcal{P}_k}_{= I - (I - B_k A_k)^{q(k)}} \Big) A_n^{-1} [b_n^\delta - b_n] \\
&= \Big(\textstyle\sum_{k=1}^n (I - B_n A_n)^{q(n)} \cdots (I - B_{k+1} A_{k+1})^{q(k+1)} [\mathcal{P}_k - \mathcal{P}_{k-1}] + I \Big) A_n^{-1} [b_n^\delta - b_n] \ ,
\end{aligned}$$

and therefore, estimating similar to (35)

$$\begin{aligned}
\|\tilde{x}_n^\delta - \tilde{x}_n\|_{\mathcal{X}} &\leq (\sqrt{\textstyle\sum_{k=1}^n c^{2(q(k+1)+\ldots+q(n))}} + 1) \|A_n^{-1} [b_n^\delta - b_n]\|_{\mathcal{X}} \\
&\leq (\sqrt{\textstyle\sum_{k=1}^n c^{2(n-k)}} + 1) \|A_n^{-1/2} [b_n^\delta - b_n]\|_n \\
&\leq C \frac{1}{\underline{\underline{\gamma}}_n} \delta \ .
\end{aligned}$$

The overall estimate

$$\|\tilde{x}_n^\delta - x^\dagger\|_{\mathcal{X}} \leq C \Big(\frac{1}{\underline{\underline{\gamma}}_n} \delta \ + \ \max\{ c^{n/2} \|x^\dagger\| \ , \ \|[I - P_{n/2}] x^\dagger\|_{\mathcal{X}} \} \Big) \tag{39}$$

suggests to choose $n := N = N(\delta)$ such that both terms on the right hand side go to zero as $\delta \to 0$. This leads to the a priori choice

$$N \to \infty \text{ and } \frac{\delta}{\underline{\underline{\gamma}}_N} \to 0 \text{ as } \delta \to 0 \ , \tag{40}$$

that, by the considerations above, yields convergence of $\tilde{x}_N^\delta$ to $x^\dagger$ as $\delta \to 0$:

Corollary 2. *Under the assertion (21) of Proposition 1 the full multigrid method defined in Algorithm 3 is a* regularization method *in the sense that, for any fixed finite level n, $\tilde{x}_n^\delta$ continuously depends on the data y^δ and, with the a priori choice (40) of the finest grid level $n = N$, $\tilde{x}_N^\delta$ uniformly converges to $x^\dagger$ as the noise level δ in (37) tends to zero:*

$$\sup\{ \|\tilde{x}_N^\delta - x^\dagger\|_{\mathcal{X}} \mid y^\delta \in \mathcal{Y} \wedge \|y^\delta - y\|_{\mathcal{Y}} \leq \delta \} \to 0 \ \textit{as} \ \delta \to 0 \ .$$

If one can make use of a source condition in the way just sketched, an optimal choice of N should obviously make both terms on the right hand side of (39) of the same order of magnitude and therefore fulfill

$$\underline{\underline{\gamma}}_N \delta \sim \sigma^{pN}$$

if $\mu = \frac{1}{2}$ in (2), which for arbitrary μ would probably become

$$\underline{\underline{\gamma}}_N \delta \sim \sigma^{2\mu p N} \ .$$

Since the exponent μ in the source condition is in practice not known explicitely, it would be of interest to find such an optimal finest level a posteriori, i.e., without using a priori knowledge on μ.

Remark 2. The main assumption of Proposition 1 is uniform boundedness of $\frac{1}{\xi_k \underline{\gamma}_k^2}$, and therefore, with (11), of the quotient $\frac{\overline{\overline{\gamma}}_k}{\underline{\gamma}_k}$, that can be rewritten as

$$\frac{\overline{\overline{\gamma}}_k}{\underline{\gamma}_k} \leq \frac{\max_{x \in \mathcal{X}_k} \frac{\|[I-Q_{k-1}]A_k v\|_k}{\|v\|_\mathcal{X}}}{\min_{x \in \mathcal{X}_{k-1}^{\perp_\mathcal{X}}} \frac{\|[I-Q_{k-1}]A_k v\|_k}{\|v\|_\mathcal{X}}} \ ,$$

because $\|A_k v\|_k \geq \|[I-Q_{k-1}]A_k v\|_k$, or as

$$\frac{\overline{\overline{\gamma}}_k}{\underline{\gamma}_k} \leq \frac{\max_{x \in \mathcal{X}_{k-1}^{\perp_k}} \frac{\|v\|_\mathcal{X}}{\|v\|_k}}{\min_{x \in \mathcal{X}_{k-1}^{\perp_\mathcal{X}}} \frac{\|v\|_\mathcal{X}}{\|v\|_k}}$$

because

$$\overline{\overline{\gamma}}_k = \|[I-Q_{k-1}]A_k^{1/2}\|_k = \|A_k^{1/2}[I-Q_{k-1}]\|_k = \max_{v \in \mathcal{X}_k} \frac{\|[I-Q_{k-1}]v\|_\mathcal{X}}{\|v\|_k} \leq \max_{v \in \mathcal{X}_k} \frac{\|[I-Q_{k-1}]v\|_\mathcal{X}}{\|[I-Q_{k-1}]v\|_k} \ .$$

This shows that (11), (12) has to be regarded as a kind of restriction on the growth of the spaces $\mathcal{X}_k$; the refinement from level $k-1$ to level k must not proceed too fast. Consider, for illustrative purpose, the special choice $\mathcal{X}_k = \text{span}\{u_1 \ldots, u_{d(k)}\}$, $\mathcal{Y}_k = \text{span}\{v_1 \ldots, v_{d(k)}\}$, where $(\sigma_j; u_j, v_j)_{j \in \mathbb{N}}$ is a singular system of a compact operator $\mathcal{T}$. Then

$$\frac{\overline{\overline{\gamma}}_k}{\underline{\gamma}_k} = \frac{\sigma_{d(k-1)+1}}{\sigma_{d(k)}} \ .$$

For mildly ill-posed problems with $\sigma_j \sim j^{-p}$ for some $p > 0$ — note that the exponent p here corresponds to the degree of smoothing p in (36) and below — (11) with (12) therefore means that

$$\frac{d(k)}{d(k-1)} \leq C \ ,$$

so that geometrically growing space dimensions are feasible, while for severely ill-posed problems, where $\mathcal{T}^*$ is infinitely smoothing and $\sigma_j \sim e^{-j}$, (11), (12) is equivalent to

$$e^{d(k)-d(k-1)} \leq C \ ,$$

and therefore the projection spaces are only allowed to be augmented by the span of a fixed number of linear independent functions. These theoretical considerations in the framework of singular systems are, in our context, of course not of immediate practical significance, since then the operators A_k are of diagonal form; still they are relevant also for more general bases, since (cf. Section 3.3 in [9])

$$\overline{\overline{\gamma}}_k \geq \sigma_{d(k-1)+1} \ ,$$

and

$$\underline{\gamma}_k \leq \sigma_{d(k)} \ .$$

When projecting onto finite element spaces $\mathcal{Y}_k$, the approximation properties of piecewise polynomial functions (cf. Bramble-Hilbert Lemma) in combination with the smoothing properties of $\mathcal{T}^*$ can be used to estimate $\overline{\overline{\gamma}}_k$: Let, like in (36), $p > 0$ denote the degree of smoothing of $\mathcal{T}^*$ acting from $\mathcal{Y} := H^t$ to $\mathcal{X} := H^s$, i.e., let $\|\mathcal{T}^*\|_{H^{t-p}\to H^s}$ be finite, and let $\mathcal{Y}_k$ be a space of piecewise polynomial functions of order not less that $p-1$ on a mesh of size h_k, so that $\|[I-\mathcal{Q}_k]\|_{H^t\to H^{t-p}} = O((h_k)^p)$. Then

$$\overline{\overline{\gamma}}_k \leq \|\mathcal{T}^*[I-\mathcal{Q}_{k-1}]\|_{\mathcal{Y}\to\mathcal{X}} \leq \|\mathcal{T}^*\|_{H^{t-p}\to H^s}\,\|[I-\mathcal{Q}_k]\|_{H^t\to H^{t-p}} = O((h_k)^p) \ .$$

To see how $\underline{\underline{\gamma}}_k$ and therefore also $\underline{\gamma}_k$ can be bounded from below in the Sobolev space setting with piecewise polynomial discretization from above, we assume that not only

$$\|\mathcal{T}^*v\|_{H^s} \leq C_1\|v\|_{H^{t-p}} \ , \quad \forall v \in \mathcal{Y}$$

with some $C_1 < \infty$ but also

$$\|\mathcal{T}^*v\|_{H^s} \geq c_2\|v\|_{H^{t-p}} \ , \quad \forall v \in \mathcal{Y}$$

with some $c_2 > 0$. Then an estimate of the square root of the largest eigenvalue of A_k^{-1}

$$\begin{array}{rcl} \frac{1}{\underline{\underline{\gamma}}_k} & = & \max_{v\in\mathcal{X}_k} \frac{\|v\|_k}{\|v\|_{\mathcal{X}}} \\ & \leq & C_\psi \max_{v\in\mathcal{Y}_k} \frac{\|v\|_{\mathcal{Y}}}{\|\mathcal{T}^*v\|_{\mathcal{X}}} \\ & \leq & \frac{C_\psi}{c_2} \max_{v\in\mathcal{Y}_k} \frac{\|v\|_{H^t}}{\|v\|_{H^{t-p}}} \\ & \leq & Ch_k^{-p} \end{array}$$

obviously corresponds to some so-called inverse inequality relating different Sobolev norms on finite-dimensional piecewise polynomial spaces Y_k (cf. Theorem 3.2.6 in [8]).

We close this section with an example of a mildly ill-posed problem, namely the linearization of the inverse groundwater filtration problem of identifying a distributed parameter a from measurements of a solution u to the boundary value problem

$$\begin{array}{rcll} -\nabla(a\nabla u) & = & f & \text{in } \Omega \\ u & = & g & \text{on } \partial\Omega \end{array} \tag{41}$$

on a domain $\Omega \subseteq \mathbb{R}^d$, $d = 1,2$ or 3. (cf., e.g., [19] and the references therein). The formal derivative of the forward operator F, mapping a parameter a to the solution u of (41), at some initial guess $a_0 \in W^{1,\infty}(\Omega)$ with $a_0 \geq \underline{a} > 0$ *a.e*, into some direction s, is given by the solution $v = F'(a_0)s$ of the BVP

$$\begin{array}{rcll} -\nabla(a_0\nabla v) & = & \nabla(s\nabla u_0) & \text{in } \Omega \\ v & = & 0 & \text{on } \partial\Omega \ , \end{array} \tag{42}$$

where $u_0 = F(a_0)$. Therefore

$$a_1 := a_0 + s$$

with

$$\mathcal{T}s := F'(a_0)s = u - F(a_0) \ , \tag{43}$$

i.e., the first Newton iterate, can be expected to be a better approximation to a than a_0.

A natural choice of the preimage space of $\mathcal{T}$ is

$$\mathcal{X} := \{s \in L^2(\Omega) \mid \nabla(s\nabla u_0) \in L^2(\Omega) \ \wedge \ s = 0 \text{ on } \Gamma_-\} \ ,$$

where Γ_- is the instream boundary $\Gamma_- := \{\vec{n}\cdot\nabla u_0 < 0\} \subseteq \partial\Omega$, with the anisotrope norm $\|s\|_{\mathcal{X}} := \|\nabla(s\nabla u_0)\|_{L^2(\Omega)}$. The $\mathcal{Y}$-topology, i.e., the one in which the measured data live, is naturally $L^2(\Omega)$, (only values but not derivative values of u can be measured,) whereas a solution v of (42), with $s \in \mathcal{X}$, is in $H^2(\Omega)$ — this makes it intuitively clear that $\mathcal{T}$, acting between the so defined spaces $\mathcal{X}$ and $\mathcal{Y}$, is just as smoothing as twice integration. Indeed, with a discretization by piecewise linear functions on a regular grid of size $h_k := C\sigma^k$, for some $\sigma \in (0,1)$, and $\langle \cdot, \cdot \rangle_k$ according to (9) we get, denoting by $D(a_0)$ the second order differential operator $D(a_0) : H_0^1(\Omega) \cap H^2(\Omega) \to L^2(\Omega)$, $v \mapsto - \nabla (a_0 \nabla v)$, whose inverse is obviously symmetric and positive definite wrt. L^2,

$$
\begin{array}{rcl}
\overline{\overline{\gamma}}_k & \leq & \|[I - \mathcal{Q}_{k-1}]\mathcal{T}\|_{\mathcal{X}\to\mathcal{Y}} \\
& = & \max_{s\in\mathcal{X}} \min_{w\in\mathcal{Y}_{k-1}} \frac{\|\mathcal{T}s - w\|_{\mathcal{Y}}}{\|s\|_{\mathcal{X}}} \\
& = & \max_{s\in\mathcal{X}} \min_{w\in\mathcal{Y}_{k-1}} \frac{\|D(a_0)^{-1}[\nabla(s\nabla u_0)] - w\|_{L^2(\Omega)}}{\|\nabla(s\nabla u_0)\|_{L^2(\Omega)}} \\
& \leq & \max_{v\in H_0^1(\Omega)\cap H^2(\Omega)} \min_{w\in\mathcal{Y}_{k-1}} \frac{\|v-w\|_{L^2(\Omega)}}{\|D(a_0)v\|_{L^2(\Omega)}} \\
& = & O((h_{k-1})^2) \ ,
\end{array}
$$

since $|v|_{H^2(\Omega)} \leq C\|D(a_0)v\|_{L^2(\Omega)}$ holds for all $v \in H_0^1(\Omega) \cap H^2(\Omega)$ and some constant C depending only on $\underline{a}$ and $\|a_0\|_{W^{1,\infty}(\Omega)}$. On the other hand, if we additionally assume that $\triangle u_0$ is bounded away from zero, we get, with the then existent solution $z \in \mathcal{X}$ of $\nabla(z \nabla u_0) = v$,

$$
\begin{array}{rcl}
\underline{\underline{\gamma}}_k & = & \min_{v\in\mathcal{Y}_k} \frac{\|\mathcal{T}^* v\|_{\mathcal{X}}}{\|v\|_{\mathcal{Y}}} \\
& \geq & \min_{v\in\mathcal{Y}_k} \frac{\langle v, \mathcal{T}z\rangle_{\mathcal{Y}}}{\|v\|_{\mathcal{Y}}\|z\|_{\mathcal{X}}} \\
& = & \min_{v\in\mathcal{Y}_k} \frac{\|D(a_0)^{-1/2}v\|^2_{L^2(\Omega)}}{\|v\|^2_{L^2(\Omega)}} \\
& \geq & \min_{v\in\mathcal{Y}_k} \frac{\|v\|_{L^2(\Omega)}}{\|D(a_0)^{1/2}v\|_{L^2(\Omega)}} \\
& \geq & \frac{1}{\|a_0\|_{L^\infty(\Omega)}} \left(\max_{v\in\mathcal{Y}_k} \frac{|v|_{H_0^1(\Omega)}}{\|v\|_{L^2(\Omega)}} \right)^{-2} \\
& \geq & c(h_k)^{-2}
\end{array}
$$

where we have used Cauchy-Schwarz inequality in the second and fourth line, and an inverse inequality according to Theorem 3.2.6 in [8] in the last line. Therefore, with a choice $\xi_k \sim (h_k)^{-4}$, (12) holds for this example. Moreover, under source condition with $\mu = \frac{1}{2}$, which is here equivalent to

$$\nabla(s^\dagger \nabla u_0) \in H_0^1(\Omega) \cap H^2(\Omega) \ \wedge \ s^\dagger = 0 \text{ on } \Gamma_- \ ,$$

one obtains, with $w \in \mathcal{Y}$ such that $s^\dagger = \mathcal{T}^* w$,

$$
\begin{array}{rcl}
\|[I - \mathcal{P}_k]s^\dagger\|_{\mathcal{X}} & = & \min_{t\in\mathcal{X}_k} \|\mathcal{T}^* w - t\|_{\mathcal{X}} \\
& = & \min_{v\in\mathcal{Y}_k} \max_{z\in\mathcal{X}} \frac{\langle w - v, \mathcal{T}z\rangle_{\mathcal{Y}}}{\|z\|_{\mathcal{X}}} \\
& = & \min_{v\in\mathcal{Y}_k} \max_{z\in\mathcal{X}} \frac{\langle w - v, D(a_0)^{-1}[\nabla(z\nabla u_0]\rangle_{L^2(\Omega)}}{\|\nabla(z\nabla u_0\|_{L^2(\Omega)}} \\
& \leq & C \min_{v\in\mathcal{Y}_k} \|w - v\|_{(H^2(\Omega))^*} \\
& \leq & C(h_k)^2 \|w\|_{L^2(\Omega)} \ ,
\end{array}
$$

and therefore the convergence rates

$$\|\tilde{s}_n - s^\dagger\|_{\mathcal{X}} = O((h_n)^2)$$

in the noise free case and, for noisy data with noise level δ, the optimal rate

$$\|\tilde{s}_N - s^\dagger\|_{\mathcal{X}} = O(\sqrt{\delta})$$

where $s^\dagger$ is the exact solution of (43), $\tilde{s}_n$ its FMGM approximation, and

$$N \sim \frac{1}{4} \log(\frac{1}{\delta}) \ .$$

Figure 3.2, showing the exact solution $s(x) = \sin(\pi x)$ (solid line) of (43) in the one-dimensional setting $\Omega = (0,1)$ of (41) with $g(0) := 0$, $g(1) := 1$, $f :\equiv 0$, $a(x) := \frac{2+\pi}{2+\pi(1-\sin(\pi x))}$, $u(x) :=$

$\frac{1}{\pi}(\cos(\pi x) - 1 + (2+\pi)x)$, as well as its projection-regularized approximation s_n with no data noise (dash-dotted line) and with only 0.01 per cent data noise (dotted line), for the dimensions $d(n) = 2^n - 1 = 7, 15, 31, 63, 127, 255$, is supposed to illustrate the instability of the inverse problem under consideration wrt. data noise as well as the necessity of regularization, i.e., of a proper trade-off between approximation error and data noise propagation, which is here obviously achieved at $d(n) = 63$ or 127.

4 Numerical Tests

For a numerical illustration of our theoretical results we implemented Algorithm 1 in a MatLab program and computed the resulting condition numbers of $B^n A^n$.

First of all we tested it for the probably most simple example of an ill-posed problem, namely numerical differentiation of a real function on the unit interval, which is equivalent to solving the first kind equation (1) with

$$(\mathcal{T}x)(s) = \int_0^s x(t)\ dt\ , \quad s \in [0,1]\ ,$$

where the projection spaces $\mathcal{Y}_k$ were chosen to consist of piecewise linear splines on an equidistant grid of size 2^{-k}. In Table 1 we listed the condition numbers of $B^k A^k$, and for comparison, also the condition numbers of A^k, versus the dimension 2^k of $\mathcal{Y}_k$, for the V-, the alternating, and the W-cycle, respectively. Obviously, cond$(B^k A^k)$ is level independent already for the alternating cycle (and hence also for the W-cycle).

Table 1

d(n)		cond($B^n A^n$)		cond($\tilde{A}^n$)
	V-cycle	alt. cycle	W-cycle	
7	2.6622e+00	2.6622e+00	2.6622e+00	7.2724e+02
15	2.8172e+00	2.8172e+00	2.7518e+00	3.0596e+03
31	3.8830e+00	2.8168e+00	2.7992e+00	1.2397e+04
63	5.2733e+00	3.4738e+00	2.8130e+00	4.9748e+04
127	6.7477e+00	3.1516e+00	2.8201e+00	1.9915e+05
255	8.5577e+00	3.8226e+00	2.8256e+00	7.9677e+05
511	1.0783e+01	3.3688e+00	2.8280e+00	3.1872e+06

The same tests were also done, using, instead of (9), the discrete inner product

$$\langle u, v\rangle_k := (u^k)^T v^k\ , \quad u = \sum_{j=1}^{d(k)} u_j^k \phi_j^k,\ v = \sum_{j=1}^{d(k)} v_j^k \phi_j^k\ \in \mathcal{X}_k\ , \tag{44}$$

i.e., the Euklidean inner product of the corresponding coefficients wrt. the basis $\{\phi_1^k, \ldots, \phi_{d(k)}^k\}$ of $\mathcal{X}_k$, that obviously avoids the inversion of Gramian matrices in the matrix implementations of the operators A_k, Q_k, and R_k. The corresponding results listed in Table 2 show a qualitatively similar behavior of the resulting MGM preconditioner as with (9):

Table 2

d(n)		cond($B^n A^n$)		cond($\tilde{A}^n$)
	V-cycle	alt. cycle	W-cycle	
7	5.7347e+00	5.7347e+00	5.7347e+00	7.2724e+02
15	6.2600e+00	6.2600e+00	6.1168e+00	3.0596e+03
31	1.2612e+01	7.2732e+00	6.3114e+00	1.2397e+04
63	2.1431e+01	1.2847e+01	6.4696e+00	4.9748e+04
127	3.5130e+01	1.3561e+01	6.7018e+00	1.9915e+05
255	5.6940e+01	2.1141e+01	6.9881e+00	7.9677e+05
511	9.3323e+01	2.3492e+01	7.1592e+00	3.1872e+06

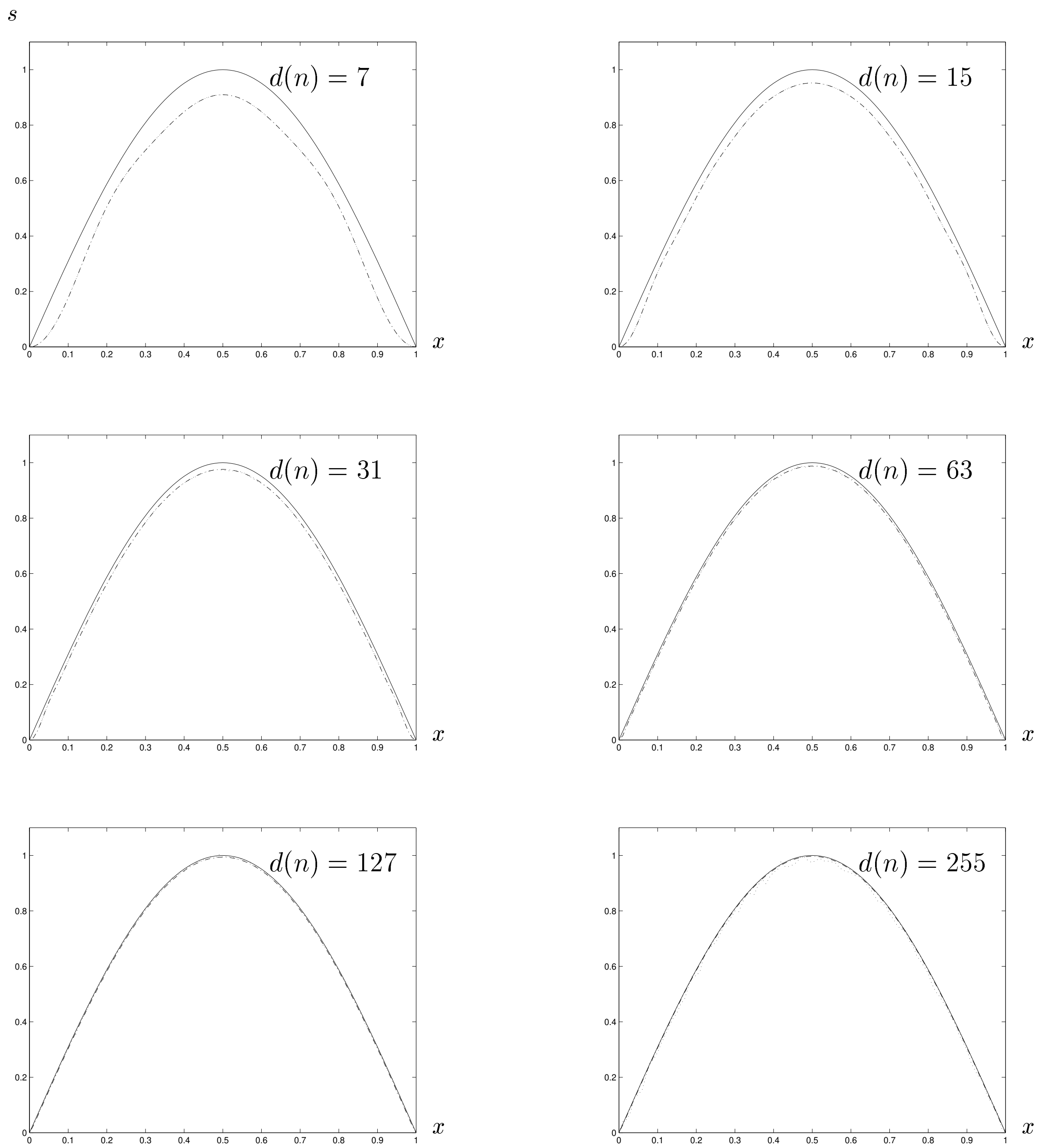

Figure 1: Exact and projection-regularized solutions of (43) with 0 and 0.01 per cent data noise for several projection space dimensions $d(n)$.

To show the numerical performance of the multilevel preconditioners defined by Algorithm 1, (see column PCG a) in Table 3,) and formula (30), (see column PCG b) in Table 3,) respectively, we applied it also to the application described at the end of the preceding section, for the one-dimensional case $\Omega = (0,1)$ with the same finite bases as in the numerical differentiation example, the discrete inner product (44), with a starting parameter $a_0 :\equiv 1$ and the searched-for exact solution $s(x) := \sin(\pi * x)$. The MG operators B_n, B_n^a were used as preconditioners for a PCG method to approximately (with a maximal error of less than 0.1 per cent) solve the linear system (5). A comparison of the total elapsed CPU-time (and the number of CG iterations), with a diagonally preconditioned CG (see column DCG) , an unpreconditioned CG (see column CG), and a direct Cholesky solver, listed in Table 3, shows the superiority of the MG-preconditioned CG already for 2^6 and more unknowns. The CPU time for assembling the additive MG preconditioner could certainly be significantly reduced by a parallel implementation (which we did not do here, yet).

Table 3

d(n)	PCG a) sec/it	PCG b) sec/it	DCG sec/it	CG sec/it	Cholesky sec
7	0.03/3	0.01/3	0.01/3	0.01/3	0.02
15	0.06/6	0.03/7	0.03/10	0.03/10	0.07
31	0.17/13	0.13/21	0.14/52	0.13/44	0.32
63	0.58/28	0.47/56	0.72/373	0.67/322	1.84
127	2.33/58	1.90/137	3.51/1412	4.18/1890	11.97
255	10.37/109	8.40/310	19.84/4367	(failed)	85.30
511	69.19/197	51.92/780	(failed)	(failed)	644.37

5 Conclusions and Remarks

As mentioned in Section 3, it would be of major interest to develop a posteriori strategies for determining the number $n = N(\delta)$ of levels in the FMGM. This could probably be done by a discrepancy type principle, i.e., by a criterion similar to

$$\|\mathcal{T}\tilde{x}_N^\delta - y^\delta\|_\mathcal{Y} \sim \delta \ ,$$

that is known as Morozov's principle in the field of ill-posed problems (cf., e.g., [9] and the literature cited there).

The considerations of this paper not only apply to pure regularization by projection but also to regularized versions of (1) that are discretized by projection in image space. Consider, e.g, Tikhonov regularization, where (1) is replaced by

$$(\mathcal{T}^*\mathcal{T} + \alpha I)x_\alpha = \mathcal{T}^* y \ ,$$

with a small regularization parameter α, and (4) by

$$x_{\alpha,n} := \mathcal{T}^*(\mathcal{Q}_n(\mathcal{T}\mathcal{T}^* + \alpha I))^\dagger \mathcal{Q}_n y \in \mathcal{X}_n \ ,$$

or, equivalently, by the solution $x_{\alpha,n}$ in $\mathcal{X}_n$ of

$$\langle x_{\alpha,n}, v\rangle_\mathcal{X} + \alpha\langle x_{\alpha,n}, v\rangle_n = \langle \mathcal{T}^* y, v\rangle_n \ , \quad \forall v \in \mathcal{X}_n \ ,$$

that leads to an equation (8) with the matrices A^k in (7) replaced by

$$A_\alpha^k = A^k + \alpha I^k \ , \tag{45}$$

which becomes (7) for $\alpha = 0$. Analogously to [21] our results can be extended to the case $\alpha \geq 0$ in (45). To get convergence of the regularized (and discretized) approximations to $x^\dagger$ one will have to let α tend to zero. Note that in the framework of [21] as well as the present paper this is admissible and does not disturb the uniform boundedness away from one of the multigrid convergence factor.

Acknowledgment

The author wishes to thank Dr. Schöberl, University of Linz, for interesting and instructive discussions on the multigrid method.

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] H. Brakhage, *On ill-posed problems and the method of conjugate gradients*, in: H.W. Engl, C.W. Groetsch, eds., *Inverse and Ill-Posed Problems*, Academic Press, Boston - New York - London, 1987, 165-175.

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

[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.

[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] P.C. Ciarlet, *The Finite Element Method for Elliptic Problems*, North Holland, 1978.

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

[10] 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.

[11] C.W. Groetsch, *The Theory of Tikhonov Regularization for Fredholm Integral Equations of the First Kind*, Pitman, Boston, London, Melbourne, 1984.

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

[13] 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.

[14] G. Haase, U. Langer, A. Meyer, S.V. Nepomnyashikh, *Hierarchical extension operators and local multigrid methods in domain decomposition preconditioners*, East-West J. Numer. Math. 2 81994), 179-193.

[15] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer, Berlin–Heidelberg–New York, 1985.

[16] W. Hackbusch, *Integral Equations. Theory and Numerical Treatment*, Birkhäuser, Basel, 1995.

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

[18] T. Hohage, *Iterative Methods in Inverse Obstacle Scattering: Regularization Theory of Linear and Nonlinear Exponentially Ill-Posed Problems*, PhD thesis, University of Linz, 1999.

[19] B. Kaltenbacher, *A projection-regularized Newton method for nonlinear ill-posed problems and its application to parameter identification problems with finite element discretization*, Technical Report 13/1998, Industrial Mathematics Institute, University of Linz, 1998.

[20] J.T. King, *On the construction of preconditioners by subspace decomposition*, J. Comp. Appl. Math. 29 (1990), 195-205.

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

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

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

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

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

[26] A. Rieder, *A wavelet multilevel method for ill-posed problems stabilized by Tikhonov regularization*, Numer. Math. 75 (1997), 501-522.

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

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