Bergische Universität Wuppertal

Fachbereich Mathematik und Naturwissenschaften

Institute of Mathematical Modelling, Analysis and Computational Mathematics (IMACM)

Preprint BUW-IMACM 13/11

Matthias Bolten, Marco Donatelli, Thomas Huckle, Christos Kravvaritis

Generalized grid transfer operators for multigrid methods applied on Toeplitz matrices

July 2013

http://www-ai.math.uni-wuppertal.de

BIT manuscript No.
(will be inserted by the editor)

Generalized grid transfer operators for multigrid methods applied on Toeplitz matrices

Matthias Bolten · **Marco Donatelli** ·
Thomas Huckle · **Christos Kravvaritis**

Abstract In this paper we discuss classical sufficient conditions to be satisfied from the grid transfer operators in order to obtain optimal two-grid and V-cycle multigrid methods utilizing the theory for Toeplitz matrices. We derive relaxed conditions that allow for the construction of special grid transfer operators that are computationally less expensive while preserving optimality. Especially we allow to use rank deficient grid transfer operators. In this case the use of an intermediate iteration as a pre-smoother that is lacking the smoothing property is proposed. Such an intermediate iteration is necessary if the used smoother does not remove error components relative to the nullspace of the grid transfer operator. Combining these new rank deficient grid transfer operators with the intermediate iteration we obtain a substantial reduction of the convergence rate compared with the classical choice for Toeplitz matrices.

Using high-order polynomials as generating symbols for the system matrix and/or the grid transfer operators usually destroys the Toeplitz structure on the coarser levels. We discuss some effective and computational cheap coarsening strategies found in the literature. For the case of Toeplitz matrices with a zero of order two (like the Laplacian) we prove the optimality of the V-cycle for these strategies, while for the high-order operators considered in the paper we present numerical results showing near-optimal behavior while keeping the Toeplitz structure on the coarser levels.

M. Bolten
Department of Mathematics and Science, University of Wuppertal, 42097 Wuppertal, Germany
E-mail: bolten@math.uni-wuppertal.de

M. Donatelli
Dipartimento di Scienza e Alta Tecnologia, Università dell'Insubria, Via Valleggio 11, 22100 Como, Italy
E-mail: marco.donatelli@uninsubria.it

T. Huckle
Department of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany
E-mail: huckle@in.tum.de

C. Kravvaritis
Department of Mathematics, University of Athens, Panepistimioupolis, 157 84 Athens, Greece
E-mail: ckrav@math.uoa.gr

Keywords Multigrid methods · Toeplitz matrices · grid transfer operators

Mathematics Subject Classification (2000) 15B05 · 65F10 · 65N22 · 65N55

1 Introduction

Multigrid methods are well-known to be optimal methods for a variety of problems, including, but not limited to the solution of partial differential equations. The convergence of these methods has been studied from the very beginning by Fedorenko [8] and Bakhvalov [3], and later by various authors including Hackbusch [11, 12, 13], Ruge and McCormick [17], McCormick [16], and many others. Different tools exist for the analysis of multigrid methods, one of the most important being the local Fourier analysis (LFA). In [25] a practical guide to the use of the LFA is presented. A good overview over multigrid in general and the available theory is given in [23]. The classical convergence theory is based on the smoothing property that has to be fulfilled by the smoothing iteration and the approximation property that the coarse grid correction has to satisfy. Based on this theory the convergence of a two-grid method (TGM) and also of the W-cycle can be shown easily. In the case of variational problems where the coarse grid operator fulfills the Galerkin-condition, the coarse grid correction has a minimization property and the convergence of the V-cycle can be shown straightforwardly [17].

In this paper, we are dealing with multigrid methods for Toeplitz and circulant matrices. The theory is based on the classical algebraic multigrid (AMG) theory that is presented, e.g., in [19], and can be considered as a generalization of the classical local Fourier analysis (see [7]). Multigrid for Toeplitz matrices goes back to Fiorentino and Serra-Capizzano [9, 10] and was investigated by many others [4, 6, 15, 22]. The convergence of the two-grid method for Toeplitz matrices has been studied in more detail by Serra-Capizzano [21]. The optimality of the V-cycle for certain matrices from matrix algebras was proved in [2] and extended in [1]. The class of Toeplitz matrices is a useful framework that allows to deal also with constant coefficient PDEs, cf. [7].

A classical request on the grid transfer operators is that they have to be full rank when the Galerkin approach is applied (as positivity of the LF orders is required [26]). In this paper we show that also rank deficient projectors can be considered if a proper pre-smoother, or better an intermediate iteration [20], is added that has to remove error components relative to the kernel of the projector. We discuss in detail practical conditions to design effective grid transfer operators. According to the proposed analysis, we provide some examples of rank deficient projectors that reduce the computational cost of classic projectors (e.g. interpolation operators) without spoiling the convergence.

On the other hand, the classical optimality conditions guarantee a constant convergence rate, which could be very large in practice if the projector is not chosen properly. Hence, we add a further simple condition on the projector to ensure fast convergence.

We show by numerical examples that our proposal, which combines a new rank deficient grid transfer operator with an intermediate iteration, halves the convergence rate with respect to the classical choices for Toeplitz matrices. Moreover, the computational cost of each iteration is reduced because the projector is sparser than the common choice. For the specific case of Toeplitz matrices, we prove that when the symbol has a zero of order two (like the Laplacian) the V-cycle is optimal and all the different proposals in the literature [2,4,15] are equivalent. For higher order projectors the Galerkin strategy destroys the Toeplitz structure at the coarser levels, thus we discuss some effective and computational cheap coarsening strategies that maintain the Toeplitz structure. The numerical experiments show that our proposal, the new rank deficient grid transfer operator combined with an intermediate iteration, is very effective for the strategy proposed in [2] for preserving the Toeplitz structure by changing the cutting matrix.

The paper is organized as follows. In Section 2 we describe multilevel Toeplitz and circulant matrices and multigrid methods for these classes of matrices. In Section 3 we study the two-grid and V-cycle convergence obtaining new optimality conditions that are discussed with a 1D example in Section 4. Using the new optimality conditions in Section 5 we define new effective and computationally cheap projectors for elliptic PDEs. In Section 6 we discuss different coarsening strategies for Toeplitz matrices and Section 7 is devoted to concluding remarks.

2 Multigrid methods for Toeplitz matrices

Let $f:\mathbb{R}^d\to\mathbb{R}$ be a continuous function and suppose that f has period 2π with respect to each variable. Let $\langle\cdot|\cdot\rangle$ denote the usual scalar product. The Fourier coefficients of f are

$$a_j=\frac{1}{(2\pi)^d}\int_{[-\pi,\pi]^d} f(x)e^{-\mathrm{i}\langle j|x\rangle}\,dx, \qquad \mathrm{i}^2=-1, \qquad j\in\mathbb{Z}^d,$$

and they enjoy the relation $a_{-j}=\bar{a}_j$ for every $j\in\mathbb{Z}^d$. From the coefficients a_j one can build [24] the sequence $\{T_n(f)\}$, $n\in\mathbb{N}^d$, of multilevel Toeplitz matrices of size $N=N(n)=\prod_{r=1}^{d} n_r$. Every matrix $T_n(f)$ is explicitly written as

$$T_n(f)=\sum_{|j_1|\leqslant n_1-1}\cdots\sum_{|j_d|\leqslant n_d-1} a_j J_{n_1}^{[j_1]}\otimes\cdots\otimes J_{n_d}^{[j_d]},$$

where $\otimes$ denotes the usual tensor product. If $n,j\in\mathbb{Z}$ then $J_n^{[j]}\in\mathbb{R}^{n\times n}$ is the matrix whose entry (s,t) equals 1 if $s-t=j$ and is 0 elsewhere. From the identity $a_{-j}=\bar{a}_j$ for every j, it follows that the matrices $T_n(f)$ are Hermitian for every n. Moreover, if $f\geq 0$ then $T_n(f)$ is positive definite. In the following, we assume that $f\geq 0$ with at least one zero of finite order.

Multilevel circulant matrices are a subset of multilevel Toeplitz matrices that are simultaneously diagonalized by the multidimensional discrete Fourier transform. For $n\in\mathbb{N}$, the Fourier matrix of order n is $F_n=[\exp(-\mathrm{i}jy_k^{(n)})]_{k,j}/\sqrt{n}$, where $y_k^{(n)}=2\pi k/n$, $k=0,\ldots,n-1$. The d-dimensional Fourier matrix is defined by tensor product as $F_n=F_{n_1}\otimes\cdots\otimes F_{n_d}$. Let $D_n(f)=\mathrm{diag}(f(y^{[n]}))$ be the diagonal matrix where

$y^{[n]} = y^{(n_1)} \times \cdots \times y^{(n_d)}$. The multilevel circulant matrix generated by f is

$$C_n(f) = F_n D_n(f) F_n^H.$$

It is immediate to see that $C_n(f)$ is ill-conditioned if f has zeros in $[-\pi,\pi]^d$ and is singular if the zeros contain a grid point of $y^{[n]}$.

In the following, we consider circulant matrices for the theoretical analysis implied by the algebra structure and we consider Toeplitz matrices for practical applications.

Multigrid methods for circulant and Toeplitz matrices are usually based on the Galerkin approach. The grid transfer operator is defined by combining a down-sampling operator with a circulant or Toeplitz matrix that selects the subspace on which the error equation is projected. We set $n = n^{(0)} > n^{(1)} > \cdots > n^{(m)} > 0$, $m \in \mathbb{N}$, such that $n^{(i+1)} = (n^{(i)} - (n^{(i)} \bmod 2))/2$ where the operations and relations are intended componentwise. In the one-dimensional case, we define the down-sampling matrix $K_{n^{(i)}} \in \mathbb{R}^{n^{(i+1)} \times n^{(i)}}$ as

$$[K_{n^{(i)}}]_{j,k} = \begin{cases} 1 \text{ if } j = 2k - (n^{(i)}+1) \bmod 2, \\ 0 \text{ otherwise}, \end{cases} \qquad k = 1,\ldots,n^{(i)}.$$

In the d-dimensional case the down-sampling matrix is defined by the tensor product $K_{n^{(i)}} = K_{n_1^{(i)}} \otimes K_{n_2^{(i)}} \otimes \cdots \otimes K_{n_d^{(i)}}$. For circulant matrices we fix $n = 2^\alpha$, $\alpha \in \mathbb{N}^d$, and the restriction/projection operators are defined as

$$P_{n^{(i)}}(p_i) = K_{n^{(i)}} C_{n^{(i)}}(p_i), \qquad i = 0,\ldots,m-1$$

where p_i is a trigonometric polynomial that will be chosen later. For the Galerkin approach the prolongations are $P_{n^{(i)}}(p_i)^H$ and the coarse matrices are

$$A_{n^{(i+1)}} = P_{n^{(i)}}(p_i) C_{n^{(i)}}(f_i) P_{n^{(i)}}(p_i)^H,$$

for $i = 0,\ldots,m-1$. For Toeplitz matrices a natural choice is $n = 2^\alpha - 1$ and $P_{n^{(i)}}(p_i) = K_{n^{(i)}} T_{n^{(i)}}(p_i)$. Other coarsening strategies will be discussed in Section 6.

To compute the coarse matrices and to give theoretical convergence results, it is usefull to define the set of all *corners* of $x \in \mathbb{R}^d$ as

$$\Omega(x) = \{y \,|\, y_j \in \{x_j, \pi + x_j\},\ j = 1,\ldots,d\},$$

this is the set of all frequencies on the fine grid that correspond to the same frequency on the coarse grid. Moreover, we define the set of the *mirror points* of x as $\mathscr{M}(x) = \Omega(x) \setminus \{x\}$. Now, let $A_n = C_n(f)$ and $f_0 = f$. For $i = 0,\ldots,m-1$ the coarse matrix satisfies

$$A_{n^{(i+1)}} = P_{n^{(i)}}(p_i) C_{n^{(i)}}(f_i) P_{n^{(i)}}(p_i)^H = C_{n^{(i+1)}}(f_{i+1}),$$

where

$$f_{i+1}(x) = \frac{1}{2^d} \sum_{y \in \Omega(x/2)} |p_i|^2 f_i(y), \qquad x \in [-\pi,\pi]^d. \tag{2.1}$$

Here, the p_i should be chosen according to the following theorem for obtaining optimal two-grid or V-cycle methods. By *optimality* we refer to the property that the spectral radius of the iteration matrix is bounded by a constant independent of n and each iteration has computational cost proportional to the matrix-vector product.

Theorem 2.1 *Let the coefficient matrix be $A_n = C_n(f)$ with f having a unique zero at x^0. Further assume that one step of a post-smoother with iteration matrix*

$$S^{\text{post}}_{n^{(i)}} = I - \omega_i^{\text{post}} A_{n^{(i)}}, \qquad \omega_i^{\text{post}} \in \left(0, \frac{2}{\|f_i\|_\infty}\right),$$

is applied, where the optimal choice, which gives the smallest convergence rate, is $\omega_i^{\text{post}} = 1/\|f_i\|_\infty$.

Define for $i = 0, \ldots, m-1$ the restriction $P_{n^{(i)}}(p_i) = K_{n^{(i)}} C_{n^{(i)}}(p_i)$, where p_i is a trigonometric polynomial not vanishing identically and such that for each $x \in [-\pi, \pi]^d$

$$\limsup_{x \to x^0,\, y \in \mathscr{M}(x)} \frac{|p_i(y)|^\gamma}{f_i(x)} = c < +\infty, \tag{2.2a}$$

where

$$\sum_{y \in \Omega(x)} p_i(y)^2 > 0, \tag{2.2b}$$

and $A_{n^{(i+1)}} = P_{n^{(i)}}(p_i) A_{n^{(i)}} P_{n^{(i)}}(p_i)^H$. Then

(i) the TGM ($m = 1$) is optimal if $\gamma = 2$,

(ii) the V-cycle is optimal if $\gamma = 1$.

Proof See [1,2,22].

For elliptic PDEs, in [7] the conditions (2.2) were compared with the classical LFA showing that the TGM condition (2.2a) with $\gamma = 2$ is equivalent to the well-known condition on the order of the grid transfer operator as formulated in [14]. Condition (2.2b) was not present in [14] because this paper used a rediscretization approach at the coarse levels. Nevertheless condition (2.2b) is equivalent to a condition stated in [26] when the Galerkin approach is used. The V-cycle conditions are analogous to the conditions obtained in [18].

3 New optimality conditions

In this section we discuss the conditions (2.2) and we show how they can be relaxed preserving the optimality. Furthermore, we show that a pre-smoother, called also intermediate iteration according to [20], which should depend on the projection can be useful to accelerate the convergence. Since the theoretical optimality could be numerically obtained only for a huge number of iterations, we add a further condition on the projector in order to ensure fast convergence. In conclusion, at the end of the analysis in the present section, we replace the conditions (2.2) with new conditions to derive an effective and efficient multigrid method.

For the sake of notational simplicity, in the following analysis we omit the subscript i from p_i, f_i etc. and we add the subscript c for the next coarser level, e.g., f_c for the symbol f_{i+1} of the coarse grid operator, since we work only with a fixed level i. We refer by $\mathscr{N}(M)$ to the null space of a matrix M.

We use the Galerkin approach and so we discuss in detail the condition (2.2b). If it is not satisfied then $P_n(p)$ is not full-rank and so A_c is not invertible. This implies that at the coarsest level, where the system is solved directly, we have to use a solver for a singular linear system, e.g. the minimum norm least squares solution. But note that the rank deficient projection is also applied on the right hand side and therefore the singular linear system is solvable. Moreover, the smoother has to be especially effective in $\mathscr{N}(P_n(p))$ because the coarse grid correction does not reduce the error in such a subspace.

From the two-grid analysis, the two quantities

$$\left| \frac{p(y_r)p(y_s)}{\sqrt{f(y_r)f(y_s)}\sum_{y\in\Omega(x)} p^2(y)} \right| \qquad \text{for } r \neq s \tag{3.1}$$

and

$$\left| \frac{\sum_{y\in\mathscr{M}(y_s)} p^2(y)}{f(y_s)\sum_{y\in\Omega(x)} p^2(y)} \right| \qquad \text{for } r = s \tag{3.2}$$

have to be bounded for all $x \in \mathbb{R}^d$, where $y_r, y_s \in \Omega(x)$ (see [21]). Therefore, (2.2b) should be satisfied in order to avoid additional zeros in the denominator. Now we relax this condition allowing additional zeros. Given $f : \mathbb{R}^d \to \mathbb{R}$ and $x^0 \in \mathbb{R}^d$ such that $f(x^0) = 0$, we define $\delta(f(x^0)) = r$ where

$$\limsup_{x\to x^0} \frac{f(x^0)}{(x-x^0)^r} = c, \qquad 0 < c < +\infty.$$

If the condition (2.2b) is violated for some $x_1 \notin \Omega(x_0)$, i.e., $p(y) = 0$ for all $y \in \Omega(x_1)$, we fix $\delta_{\min} = \min_{y\in\Omega(x_1)} \delta(p(y))$. Since $x_1 \notin \Omega(x_0)$, we have that $f(y_r) \neq 0$ and $f(y_s) \neq 0$ for $y_r, y_s \in \Omega(x_1)$, otherwise the condition (2.2a) is violated. Moreover, $\delta(\sum_{y\in\Omega(x_1)} p^2(y)) = \delta_{\min}^2$ while $\delta(p(y_r)p(y_s)) = \delta(p(y_r))\delta(p(y_s)) \geq \delta_{\min}^2$ and hence (3.1) is bounded. Similarly, also (3.2) is bounded because $\mathscr{M}(y_s) = \Omega(x_1) \setminus \{y_s\}$ for $y_s \in \Omega(x_1)$. On the other hand, the condition (2.2b) can not be violated for a $x_1 \in \Omega(x_0)$, because $y_s = x^0$ belongs to $\Omega(x_1)$ thus $f(y_s) = 0$ and hence the two quantities in (3.1) and (3.2) could be unbounded.

Nevertheless, the projector does not have full rank and the null space of the projector $\mathscr{N}(P_n(p))$ is spanned by the eigenvectors associated to the points in $\Omega(x_1)$, where x_1 is the point violating (2.2b). For these vectors the coarse grid correction can not be effective at all, so the smoother has to reduce the corresponding error components.

For the V-cycle optimality, in (3.1) the function p has to be replaced with $pf^{1/2}$ (see Proposition 4 in [1]), but the same arguments as above hold true. Further, because of the rank deficient projection the coarse grid operator has an additional zero that has to be taken care of in the recursive application of the multigrid cycle.

For both, the two-grid method and the multigrid methods, it is therefore feasible to nullify the error components collinear to $\mathscr{N}(P_n(p))$. This can be done, e.g., by applying a special smoothing procedure as outlined below. We can summarize the previous discussion as follows:

Remark 3.1 The condition (2.2b) has to be satisfied only at $x = x^0$ and if it is violated at $x_1 \neq x^0$ then f_c has a further zero with respect to f and the smoother has to nullify the error, or at least to be very effective, in the subspace generated by the eigenvectors associated to the eigenvalues $f(y)$ for $y \in \Omega(x_1)$, i.e., in the subspace generated by the corresponding Fourier frequencies. However, if the smoother nullifies the error in the corresponding subspace, at the coarser level we do not have to take care of the ill-conditioned subspace related to the new zero because in this subspace the error has already been removed by the smoother on the finer level. This will be discussed in detail in the Example 5.1 in Section 5.2.

We study explicitly the 1D case. Assume that the condition (2.2b) is not satisfied at a point $\hat{x}_j = \frac{2\pi j}{n}$, assume $j < n/2$ without loss of generality, then

$$p(\hat{x}_j) = p(\hat{x}_{j+n/2}) = 0. \tag{3.3}$$

Since $K_n F_n = \frac{1}{\sqrt{2}} \left[F_{n/2} \mid F_{n/2} \right]$, the restriction is

$$\begin{aligned} P_n(p) &= K_n F_n D_n(p) F_n^H \\ &= \frac{1}{\sqrt{2}} \left[F_{n/2} \mid F_{n/2} \right] \begin{bmatrix} D_1 & \\ & D_2 \end{bmatrix} F_n^H \\ &= \frac{1}{\sqrt{2}} \left[F_{n/2} D_1 \mid F_{n/2} D_2 \right] F_n^H \end{aligned}$$

where $D_1 = \mathrm{diag}[p(\frac{2\pi j}{n})]_{j=0}^{\frac{n}{2}-1}$ and $D_2 = \mathrm{diag}[p(\frac{2\pi j}{n})]_{j=\frac{n}{2}}^{n-1}$. The equation (3.3) implies that the jth columns of D_1 and D_2 are zero and thus the jth and the $(j+n/2)$th columns of $[F_{n/2}D_1 \,|\, F_{n/2}D_2]$ are zero, as well. Therefore, $P_n(p)F_n e_j = P_n(p)F_n e_{j+n/2} = 0$ and thus the jth and the $(j+n/2)$th columns of F_n, i.e., the frequencies corresponding to $\Omega(\hat{x}_j)$ where $\hat{x}_j$ is the point that violates the condition (2.2b), belong to $\mathscr{N}(P_n(p))$.

In the previous discussion and Remark 3.1, we have shown that the condition (2.2b) can be relaxed. In the opposite direction, in the following we show that to define an effective multigrid method the value of p at x^0 is crucial.

Remark 3.2 Condition (2.2a) requires that p vanishes at $\mathscr{M}(x^0)$, while condition (2.2b), also in the form of Remark 3.1, requires that $p(x^0) > 0$. From a practical point of view, the value of p at x^0 is important like its value at the mirror points $\mathscr{M}(x^0)$. Indeed, if $p(x^0) = 0$ the coarser symbol f_c has a zero of higher order at $2x^0$ and the multigrid method can not be optimal; while if $p(x^0) \neq 0$ the conditioning of the coarse problems depends on $f_c^{(r)}(2x^0)$, where r is the order of x^0 and the conditioning is proportional to $1/f_c^{(r)}(2x^0)$. Therefore, maximizing $f_c^{(r)}(2x^0)$ according to (2.1), we should choose p such that it holds

$$|p(x^0)| = \max_{x \in [-\pi,\pi]^d} |p(x)|, \tag{3.4}$$

thus the ill-conditioning of f_c does not increase compared to f. This is confirmed also by the numerical experiments in Section 4. The condition (3.4) implies that the condition (2.2b) is automatically satisfied for $x = x^0$, otherwise p is the zero function.

In conclusion, we suggest to replace the conditions (2.2) with the following:

NEW CONDITIONS:

$$\limsup_{x\to x^0,\, y\in\mathcal{M}(x)} \frac{|p_i(y)|^\gamma}{f_i(x)} = c < +\infty, \quad \gamma = \begin{cases} 2 & \textit{for TGM} \\ 1 & \textit{for V-cycle}\end{cases} \tag{3.5a}$$

$$|p_i(x^0)| = \max_{x\in[-\pi,\pi]^d} |p_i(x)|, \tag{3.5b}$$

If $\exists x \in [0,\pi]^d$, *s.t.* $\sum_{y\in\Omega(x)} p_i(y)^2 = 0$, *then add a pre-smoother* $S^{\text{pre}}_{n^{(i)}}$ *s.t.*

$$\left(\bigotimes_{k=1}^d F_{n_k}\right)\left(\bigotimes_{k=1}^d e_{I_k}\right) \in \mathcal{N}(S^{\text{pre}}_{n^{(i)}}),\ \textit{where } I_k = \operatorname*{argmin}_{j=0,\dots,n_k^{(i)}-1} \left\| x_k - \frac{2\pi j}{n_k^{(i)}} \right\|_\infty . \tag{3.5c}$$

In practice, the condition (3.5c) is not necessary for obtaining an optimal method, because the new condition (3.5b) states that the frequencies from the null space of the prolongation do not belong to the near null space of the system matrix. However, the addition of a pre-smoother that satisfies the condition (3.5c) could greatly speed up the convergence (cf. Example 5.1). Such a pre-smoother iteration does not work exactly as smoother, but according to the terminology of the multi-iterative methods in [20], it is an intermediate (or residual) iteration that allows to ignore the newly introduced zero on the coarse grid. Of course such a pre-smoother has to preserve the convergence of the whole iteration, even if it is not necessary that it is convergent when applied alone (like it happens for the coarse grid correction), even a divergent method can be used, cf. Example 5.1.

4 A set of 1D projectors

In this section we give an analytic and numerical evidence of the importance of the new conditions (3.5), in particular (3.5b), with respect to the classical conditions (2.2). To do that we consider the 1D Laplacian with the symbol $f(x) = 2-2\cos(x)$. It is well-known that the full-weighting and the linear interpolation give an optimal V-cycle. According to condition (2.2a) p must have a zero at π in order to derive an efficient coarse grid correction. On the other side p could have additional zeros. Here, we want to discuss how additional zeros and the shape of p, in particular the new condition (3.5b), affect the convergence of the multigrid algorithm.

We consider the set of projectors

$$p_z(x) = 1 + z\cos(x) + (z-1)\cos(2x), \quad z\in\mathbb{R}, \tag{4.1}$$

that have a zero at π of order at least 2 independent of z. Some special cases are:

1. p_1 has only one zero of order two at π and it generates the full weighting projection which defines an optimal V-cycle as it is well-known. It satisfies the conditions (3.5).

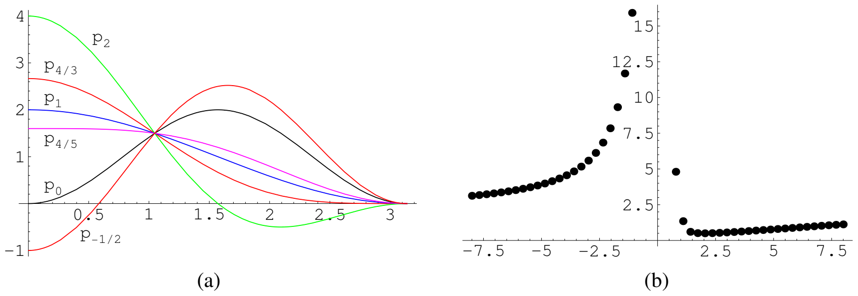


Fig. 4.1 (a) $p_z(x)$ for $x \in [0,\pi]$ and some values of z. (b) $\kappa(f_c) = \|f_c\|_\infty / f_c''(0)$ varying $z \in [-8,8]$ (note that $\kappa(0) = +\infty$).

2. p_0 has two zeros at 0 and π, and so it violates the condition (2.2b) at the origin and according to Remark 3.1 it does not define an optimal TGM. It violates also the condition (3.5b).
3. $p_{4/3}$ has a zero at π of order four and satisfies the conditions (3.5).
4. p_2 vanishes at π and $\frac{\pi}{2}$, thus the coarse function f_c has an additional zero at π and p_2 violates the condition (2.2b) for $x = \frac{\pi}{2}$, but this does not hinder the convergence since the condition (3.5b) is satisfied.

We study p_z varying $z \in \mathbb{R}$ and $x \in [0,\pi]$. Figure 4.1 (a) shows the graph of $p_z(x)$ for $x \in [0,\pi]$ and some choices of z. It holds that $p_z(x^0) = 0$ with $x^0 = \arccos((z-2)/(2z-2))$ for $z < 0$ or $z > 4/3$, while p_z has only one zero at π of order two for $z \in (0,4/3)$. To estimate the maximum modulus of p_z in $[0,\pi]$ we compute

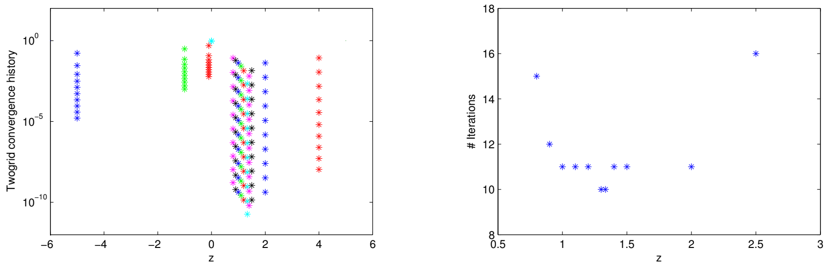
$$\frac{\partial}{\partial x} p_z(x) = -z\sin(x) - 2(z-1)\sin(2x)$$

and we obtain that $|p_z(y^0)| = \max_{x\in[0,\pi]} |p_z(x)|$ for

$$y^0 = \begin{cases} \arccos\left(\frac{z}{4(1-z)}\right), & z < \frac{4}{5}, \\ 0, & \text{otherwise.} \end{cases}$$

It follows that for $z \geq 4/5$ the function p_z has a maximum at the origin and a zero at π of order at least 2. Therefore, p_z with $z \geq 4/5$ satisfies the new conditions (3.5) and it defines an optimal V-cycle. For $z > \frac{4}{3}$ the function p_z is also negative and it has a further zero at $x^0 = \arccos((z-2)/(2(z-1))) > \pi/3$, hence the pre-smoother should be chosen according to condition (3.5c) .

It remains to study the multigrid optimality for $z < 4/5$. As discussed in the special case 2. there is no convergence for $z = 0$. Then, the same also holds for $z \approx 0$. The effectiveness of the projector p_z can be deduced from the ill-conditioning of the coarse problem around the origin. Indeed, in such subspace the smoother at the finer level can not nullify the error because also the finer problem is ill-conditioned in the same subspace, i.e., the function f is close to zero. Note that f_c vanishes at the origin with order 2, except for $z = 0$ where the zero has order 4. Therefore, the coarse problem can be considered more ill-conditioned as f_c is more flat around the origin and



(a) (b)

Fig. 4.2 (a) Two-cycle residual error after $k = 1,\ldots,10$ steps $(*)$ for different values of z. (b) Two-cycle iterations for different values of z to reach residual error less than 10^{-10}. Here, the values for z are chosen as 0.8, 0.9, 1, 1.1, 1.2, 4/3, 1.4, 1.5, 2, 2.5.

the quantity

$$\kappa(f_c) = \|f_c\|_\infty / f_c''(0), \qquad z \neq 0 \tag{4.2}$$

gives a measure of such ill-conditioning. For $z = 0$, it holds $f_c''(0) = 0$ since the zero has order four, giving an unbounded conditioning that agrees with the previous discussion for special case 2. Figure 4.1 (b) shows that for $z < 0$ the ill-conditioning $\kappa(f_c)$ decreases with z. For $z > 0$ the ill-conditioning initially decreases with z and around 2 it slowly increases with $\kappa(f_c) < 2$ for $z > 2$ according to the previous analysis.

Example 4.1 We give now a numerical evidence of the previous discussion. Two steps of damped Jacobi pre-and postsmoother were used for matrix size $n = 2^{10} - 1$ and random right hand side. Figure 4.2 (a) shows the residual error after $k = 1,\ldots,10$ TGM steps for different values of z. We note that the behavior of the residual error agrees with the previous analysis and in particular with the conditioning (4.2) plotted in Figure 4.1 (b). Figure 4.2 (b) gives the number of iterations to reach residual error $\leq 10^{-10}$. Optimal convergence is obtained for z near $4/3$, which gives the more powerful projector according to the previous discussion and special case 3.

5 High order projectors for 2D elliptic PDEs

In this section we use the new conditions (3.5) to construct effective and computational cheap projectors for symbols with a zero at the origin. However, a similar analysis can be performed for a generic zero in $[0,\pi]^d$. The new projectors do not satisfy the condition (2.2b), but they satisfy the new condition (3.5b). Moreover, Example 5.1 will show that the addition of a pre-smoother according to the further condition (3.5c) can noticeably accelerate the convergence.

Let us consider standard finite differences discretization for the following d-dimensional problem

$$\begin{cases} (-1)^q \sum_{i=1}^d \frac{\mathrm{d}^{2q}}{\mathrm{d}x_i^{2q}} u(x) = g(x), & x \in \Omega = (0,1)^d,\ q \geq 1, \\ \text{boundary conditions on } \partial\Omega, \end{cases} \tag{5.1}$$

where $x = (x_1, \dots, x_d)$. For $q = 1$ equation (5.1) is the Poisson equation, while for $q = 2$ it is the biharmonic equation. Using centered finite differences of precision two and minimal bandwidth, in the case of Dirichlet boundary conditions we obtain a linear system with coefficient matrix $A_n = T_n(f^{(q)})$ where

$$f^{(q)}(x) = \sum_{j=1}^{d} (2 - 2\cos(x_i))^q.$$

The functions $f^{(q)}$ are nonnegative and increasing in $[0,\pi]^d$ and they vanish at the origin with order $2q$. We discuss in detail the case $d = 2$, even if the same analysis can be straightforwardly extended to $d > 2$.

5.1 B-spline projectors

Starting from the refinement equation of B-splines, in [7] the following choice for the symbol of the projector has been proposed:

$$\phi_m(x) = \prod_{j=1}^{2} \left(\frac{1 + \mathrm{e}^{-\mathrm{i}x_j}}{2} \right)^m \mathrm{e}^{\mathrm{i}x_j \lfloor \frac{m}{2} \rfloor}. \tag{5.2}$$

We note that ϕ_m vanishes at $\mathscr{M}(0)$ and it has minimum support among the functions that vanish in such set with order at least m. This implies that $P_n(\phi_m)$ and the coarse matrices with the Galerkin approach have minimum bandwidth. Furthermore, $|\phi_m(0)| = \max_{x\in[-\pi,\pi]^2} |\phi_m(x)|$ and so ϕ_m satisfies the condition (3.5b).

For $m = 1$ we have explicitly

$$\phi_1(x) = \frac{1}{4}\left(1 + \mathrm{e}^{-\mathrm{i}x_1} + \mathrm{e}^{-\mathrm{i}x_2} + \mathrm{e}^{-\mathrm{i}(x_1+x_2)}\right),$$

which is zero if and only if $x_1 = \pi$ or $x_2 = \pi$. Hence it satisfies (2.2b). Let us consider $f^{(1)}$ and the condition (2.2a). We have that

$$\begin{aligned}\limsup_{x\to 0} \frac{|\phi_1(x_1, x_2+\pi)|}{f^{(1)}(x)} &= \limsup_{x\to 0} \frac{|1 + \mathrm{e}^{-\mathrm{i}x_1} - \mathrm{e}^{-\mathrm{i}x_2} - \mathrm{e}^{-\mathrm{i}(x_1+x_2)}|}{4 - 2\cos(x_1) - 2\cos(x_2)} \\ &= \limsup_{x\to 0} \frac{|2\mathrm{i}x_2 + O(x_1+x_2)^2|}{O(x_1^2) + O(x_2^2)} = +\infty.\end{aligned}$$

Similarly

$$\limsup_{x\to 0} \frac{|\phi_1(x_1+\pi, x_2)|}{f^{(1)}(x)} = +\infty.$$

On the other hand

$$\begin{aligned}\limsup_{x\to 0} \frac{|\phi_1(x_1+\pi, x_2+\pi)|}{f^{(1)}(x)} &= \limsup_{x\to 0} \frac{|1 - \mathrm{e}^{-\mathrm{i}x_1} - \mathrm{e}^{-\mathrm{i}x_2} + \mathrm{e}^{-\mathrm{i}(x_1+x_2)}|}{4 - 2\cos(x_1) - 2\cos(x_2)} \\ &= \limsup_{x\to 0} \frac{|O(x_1+x_2)^2|}{O(x_1^2) + O(x_2^2)} = c, \qquad 0 < c < +\infty.\end{aligned}$$

Therefore ϕ_1 is very effective at (π,π), but it is not enough to obtain on optimal V-cycle because the same does not hold at $(0,\pi)$ and $(\pi,0)$. However, $\phi_1(x)$ is enough to obtain the TGM optimality since

$$\limsup_{x\to 0} \frac{|\phi_1(y)|^2}{f^{(1)}(x)} = c, \qquad y \in \{(x_1, x_2+\pi), (x_1+\pi, x_2)\}, \qquad 0 < c < +\infty.$$

For $m = 2$ we have

$$\phi_2(x) = \frac{1}{2}(1+\cos(x_1))(1+\cos(x_2)),$$

which is the *bilinear* interpolation. The function ϕ_2 is the square of ϕ_1 up to a shift factor and it is zero if and only if $x_1 = \pi$ or $x_2 = \pi$. For $f^{(1)}$ the projector ϕ_2 satisfies the conditions (2.2) with $\gamma = 1$ and so it defines an optimal V-cycle as it is well-known. On the other hand, if we apply ϕ_2 to the biharmonic $f^{(2)}$, we have a similar behavior as observed before when ϕ_1 is applied to the Laplacian. Indeed, it holds

$$\limsup_{x\to 0} \frac{|\phi_2(y)|^2}{f^{(2)}(x)} = c_1 < +\infty, \qquad y \in \{(x_1, x_2+\pi), (x_1+\pi, x_2)\},$$

$$\limsup_{x\to 0} \frac{|\phi_2(x_1+\pi, x_2+\pi)|}{f^{(2)}(x)} = c_2 < +\infty.$$

In general, for $m = j$, with $j = 1, 2, \ldots$, we have that

$$\limsup_{x\to 0} \frac{|\phi_j(y)|}{x_1^j + x_2^j} = c_1 < +\infty, \qquad y \in \{(x_1, x_2+\pi), (x_1+\pi, x_2)\},$$

$$\limsup_{x\to 0} \frac{|\phi_j(x_1+\pi, x_2+\pi)|}{x_1^{2j} + x_2^{2j}} = c_2 + \infty.$$

Roughly speaking, the function ϕ_j has a zero at (π,π) of order $2j$ and two zeros of order j at $(0,\pi)$ and $(\pi,0)$. Geometrically, this follows from the fact that (π,π) is at the intersection of two zero lines $x_1 = \pi$ and $x_2 = \pi$.

To obtain higher order projectors it is possible to choose a larger m like proposed in [7]. However, the previous analysis shows that the projector ϕ_j needs to be improved only at the mirror points $(0,\pi)$ and $(\pi,0)$, because the point (π,π) already has a zero of double order.

5.2 Projectors with two zeros at $(0,\pi)$ and $(\pi,0)$

According to the analysis in the previous subsection we look for a projector that vanishes at $(0,\pi)$ and $(\pi,0)$ and not at (π,π), such that it can be combined (by multiplication) with ϕ_j.

A first choice is the function $\tilde{q}(x) = (2+\cos(x)-\cos(y))(2-\cos(x)+\cos(y))$, but we do not consider it in the following because it has the stencil

$$\frac{1}{4}\begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ -1 & 0 & 12 & 0 & -1 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix},$$

which is too large and therefore increases the number of the nonzero diagonals at the coarse levels when combined with a Galerkin strategy.

Instead, we propose to use

$$q(x) = \cos(x_1) + \cos(x_2),$$

which has a 3×3 stencil. It holds $q(x) = 0$ if and only if $x_1 = \pi - x_2$, so it vanishes at $(0,\pi)$ and $(\pi,0)$. Since $q(x)$ vanishes along the whole curve $x_1 = \pi - x_2$, it does not satisfy the condition (2.2b) at the point $(\frac{\pi}{2},\frac{\pi}{2})$. Nevertheless, according to Remark 3.1, we obtain again an optimal V-cycle convergence since $f^{(1)}(x)$ and $f^{(2)}(x)$ do not vanish at $(\frac{\pi}{2},\frac{\pi}{2})$ and the condition (3.5b) holds. Moreover, we can define a pre-smoothing iteration that is very effective in the middle frequencies around $(\frac{\pi}{2},\frac{\pi}{2})$ such that the condition (3.5c) holds true. This can be done with a proper choice of the Jacobi or Richardson relaxation parameter. The Richardson iteration matrix $S_{n^{(i)}} = I - \omega_i C_{n^{(i)}}(f_i)$ has eigenvalues $1 - \omega_i f_i(2\pi j_1/n^{(i)}, 2\pi j_2/n^{(i)})$, $j_1, j_2 = 0,\ldots,n^{(i)}-1$. Hence, for $\omega_i \in [0, 2/\|f\|_\infty]$ the method converges and satisfies the *smoothing condition* necessary for the optimality as shown in Theorem 2.1. If we want to nullify the error at the frequency $(F_n \otimes F_n)(e_{n/4} \otimes e_{n/4})$ to satisfy the condition (3.5c), we can apply Richardson with $\omega_i^{\text{pre}} = 1/f_i(\frac{\pi}{2},\frac{\pi}{2})$ as pre-smoother.

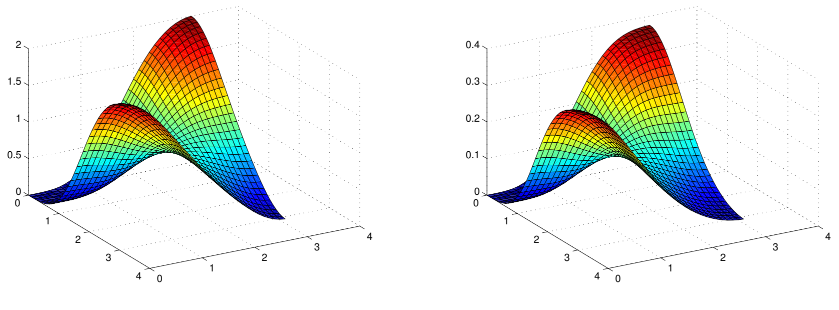
Example 5.1 Let us introduce an example to better clarify the previous discussion. We consider the biharmonic operator having symbol $f^{(2)}$. As it is well-known and according to the previous discussion, the bilinear interpolator ϕ_2 is not enough to obtain the V-cycle optimality. We consider the projector

$$p(x) = q(x)\phi_2(x), \tag{5.3}$$

which satisfies the condition (2.2a) but not the condition (2.2b). Let f_i be the symbol at the level $i = 0,1,\ldots,m$, according to Theorem 2.1, we use Richardson as post-smoother with $\omega_i^{\text{post}} = 1/\|f_i\|_\infty$. Moreover, we add a pre-smoothing step of Richardson with $\omega_i^{\text{pre}} = 1/f_i(\frac{\pi}{2},\frac{\pi}{2})$ in order to have an intermediate iteration that satisfies the condition (3.5c) according to the previous analysis.

We note that at the finest level where $f_0 = f^{(2)}$, it holds $\|f^{(2)}\|_\infty = f^{(2)}(\pi,\pi)$ and $f^{(2)}(\frac{\pi}{2},\frac{\pi}{2}) = \|f^{(2)}\|_\infty/4$, so the pre-smoother alone does not give convergence. Nevertheless, the whole smoothing iteration is convergent because it holds

$$\left| \left(1 - \frac{4f^{(2)}\left(\frac{2\pi j_1}{n}, \frac{2\pi j_2}{n}\right)}{\|f^{(2)}\|_\infty}\right)\left(1 - \frac{f^{(2)}\left(\frac{2\pi j_1}{n}, \frac{2\pi j_2}{n}\right)}{\|f^{(2)}\|_\infty}\right)\right| < 1, \qquad \forall j_1, j_2.$$



f_1 f_2

Fig. 5.1 Symbols at the coarser levels in the half plane $[0,\pi]\times[0,\pi]$, where $f_0(x) = f^{(2)}(x)$ and $p_i(x) = p(x)$ in (5.3).

This does not limit the type of post-smoother to Richardson, but other smoothers (like Gauss-Seidel) could be used, because the subspace where the pre-smoother does diverge is generated by the frequencies $(F_n \otimes F_n)(e_{j_1} \otimes e_{j_2})$ such that

$$f^{(2)}\left(\frac{2\pi j_1}{n}, \frac{2\pi j_2}{n}\right) \geq 2f^{(2)}\left(\frac{\pi}{2}, \frac{\pi}{2}\right),$$

which are high frequencies and hence damped by every smoother.

Computing the coarse symbol by (2.1), we note that it has an additional zero at (π,π) (see Figure 5.1). However the error components around such point were already reduced by the intermediate iteration at the finer level and so the next projector has to take care only of the zero at $(0,0)$. Continuing with the coarsening strategy, the ill-conditioned subspace at (π,π) enlarges, according to (2.1), see also Figure 5.1, but the error components were already reduced from the previous pre-smoothers. Therefore, the same symbol of the projector, defined for only a zero at the origin, can be used at each level.

We give the numerical evidence of such behavior. We consider the Toeplitz linear system $A_n = T_n(f^{(2)})$ and we compare the projector $p(x)$ in (5.3) with $\phi_4(x)$. Moreover, we compare the previous choice of ω_i^{pre} with $\omega_i^{\text{pre}} = 1.5/\|f_i\|_\infty$, which is the optimal damping parameter for the pre-smoother without post-smoothing (see Section 5.2 in [1]), and with $\omega_i^{\text{pre}} = 2/\|f_i\|_\infty$ that is the largest value of ω_i^{pre} that ensures the convergence of the pre-smoother. Using the Galerkin strategy the coarse matrices are not exactly block-Toeplitz-Toeplitz-block but they have a low rank correction, so Jacobi seems to be much more robust and it is used instead of Richardson. The relaxation parameters are multiplied by the Fourier coefficient of index $(0,0)$ (denoted by α_i at the i-th level) to take into account the diagonal scaling. At each level one step of the pre- and post-smoother are applied. The linear system is solved directly on a grid of size 7×7 and the tolerance is 10^{-6}. The true solution is the smooth function $\sin(x_1) + x_1\cos(x_2)/(2\pi)$.

Table 5.1 shows the number of iterations required for the V-cycle convergence. We note that the intermediate iteration with $\omega_i^{\text{pre}} = 1/f_i(\frac{\pi}{2},\frac{\pi}{2})$ is more effective than the pre-smoothing with $\omega_i^{\text{pre}} = 1.5/\|f_i\|_\infty$ or $\omega_i^{\text{pre}} = 2/\|f_i\|_\infty$. Indeed, even if also the

	$p(x)$			$\phi_4(x)$		
$n \times n \setminus \frac{\omega_{pre}}{\alpha_i}$	$f_i(\frac{\pi}{2},\frac{\pi}{2})^{-1}$	$\frac{1.5}{\Vert f_i\Vert_\infty}$	$\frac{2}{\Vert f_i\Vert_\infty}$	$f_i(\frac{\pi}{2},\frac{\pi}{2})^{-1}$	$\frac{1.5}{\Vert f_i\Vert_\infty}$	$\frac{2}{\Vert f_i\Vert_\infty}$
15×15	20	34	27	21	34	27
31×31	19	35	28	20	35	28
63×63	19	36	29	19	36	29
127×127	21	36	29	21	36	29

Table 5.1 Number of iterations vs grid size for the V-cycle with one step of damped Jacobi as pre- and post-smoother.

prolongation	$p(x)$				$\phi_4(x)$	
smoother	Jacobi w. $\omega_i^{\text{pre}} = \alpha_i/f_i(\frac{\pi}{2},\frac{\pi}{2})$		Gauss-Seidel		Gauss-Seidel	
$n \times n$	iterations	time	iterations	time	iterations	time
15×15	20	0.053 s	12	0.035 s	11	0.029 s
31×31	19	0.072 s	12	0.054 s	11	0.057 s
63×63	19	0.137 s	13	0.143 s	13	0.161 s
127×127	21	0.419 s	16	0.931 s	16	0.995 s
255×255	27	2.059 s	19	4.363 s	20	5.251 s
511×511	32	9.515 s	22	20.916 s	23	23.317 s

Table 5.2 Number of iterations and timings vs grid size for the V-cycle.

latter two choices ensure an optimal convergence (i.e., conditions (3.5a) and (3.5b) are enough for obtaining the optimality), they require a number of iterations that is about one third more than our proposed $\omega_i^{\text{pre}} = 1/f_i(\frac{\pi}{2},\frac{\pi}{2})$ for converging to the desired accuracy (i.e., the condition (3.5c) greatly speed up the convergence). Furthermore, we observe that the choice $\omega_i^{\text{pre}} = 2/\Vert f_i\Vert_\infty$ is more effective than the "optimal" choice $\omega_i^{\text{pre}} = 1.5/\Vert f_i\Vert_\infty$. This is due to the fact that the pre-smoother is here combined with the post-smoother and the pre-smoother with $\omega_i^{\text{pre}} = 2/\Vert f_i\Vert_\infty$ is effective at the middle frequencies anyway similarly to an intermediate iteration; especially at the coarser levels, as can be seen in Figure 5.1, where $f_i(\frac{\pi}{2},\frac{\pi}{2}) \approx \Vert f_i\Vert_\infty/2$.

Finally, Table 5.2 shows a comparison of our proposed prolongation operator with symbol $p(x)$ with the traditional choice with symbol $\phi_4(x)$ using Gauss-Seidel as pre-smoother and – in the case of $p(x)$ – Jacobi with the choice $\omega_i^{\text{pre}} = \alpha_i/f_i(\frac{\pi}{2},\frac{\pi}{2})$ as suggested by the previous tests and $\omega_i^{\text{post}} = \alpha_i/\Vert f_i\Vert_\infty$. Besides the pure iteration count we also measured the time to reduce the relative residual below the tolerance 10^{-6}. Timings were obtained in MATLAB R2012a running on a dual core Intel Core i5 at 1.6 GHz. As before we were using one pre- and one post-smoothing step and a coarsest grid of size 7×7. The proposed prolongation works as good as the traditional choice with timings that are better. This is due to the fact that the operator complexity is slightly lower for our proposed prolongation. Further, while the iteration count is much worse when the Jacobi iteration as smoother plus intermediate iteration is chosen, the timings are actually much better. This can be explained by the fact that Jacobi is using a diagonal scaling plus vector addition, only, while in Gauss-Seidel the components depend on the previously updated ones. We like to note that while an intermediate iteration can be combined with Gauss-Seidel, the combination of ones step of Jacobi with $\omega_i^{\text{pre}} = \alpha_i/f_i(\frac{\pi}{2},\frac{\pi}{2})$ as intermediate iteration and one step of Gauss-Seidel as a smoother does not lead to a converging multigrid method, increasing the number of Gauss-Seidel steps again works. For a converging multigrid method we require the combination of intermediate iteration plus smoother to converge, as noted before. The analysis of the combination of intermediate iteration and smoother can be carried out by local Fourier analysis [25].

Comparing the two projectors p and ϕ_4, the previous example shows that they have the same behavior in terms of number of iterations. Hence, we suggest to use p because it is computationally cheaper than ϕ_4. Indeed, each row of $P_{n^{(i)}}(p)$ has less nonzero entries than that of $P_{n^{(i)}}(\phi_4)$, 21 instead of 25. This implies also less nonzero diagonals in the coarse matrices. If we consider $f_0 = f^{(2)}$ and ϕ_4, then the coarse matrices have 5×5 stencils with all nonzero entries except for the stencil of f_1 which has 4 zeros at the corners. If we use p instead of ϕ_4, the coarse matrices have again a 5×5 stencil but the stencils of f_1, f_2, and f_i, $i = 3, \dots$, have 12, 8, and 4 zero entries, respectively.

For the 3D biharmonic problem one could use the projector $\sum_{j=1}^{3}(\cos(x_j))\phi_2(x)$ instead of ϕ_4, with a higher sparsity of the coarser matrices (for the definition of ϕ_m in 3D the production in (5.2) as to be taken until 3 instead of 2). The reduction of the nonzero entries and of the computational costs is also more relevant than in the 2D case.

6 Coarsening strategies for Toeplitz matrices

In this section we discuss how the projector can affect the structure of the coarser matrices when the Galerkin coarsening is applied to a Toeplitz matrix. This is usually the case when higher-order polynomials are needed for the representation of the system matrix and the prolongation, like in the cases considered so far.

In the case of $A_n = T_n(f)$, using the natural projector $P_n(p) = K_n T_n(p)$, the Galerkin coarse grid matrix resulting, i.e., $A_{n^{(1)}} = P_n(p)T_n(f)P_n(p)^H$, will not be Toeplitz in general due to low rank perturbation caused by the multiplication with the projection. For analytic reasons the restriction and the prolongation are usually restricted to a reduced index set such that the low-rank perturbation at the boundary is removed and the coarse system has again Toeplitz structure. Another advantage of this Toeplitzisation is that every occurring matrix can be represented by the Fourier coefficients, only.

Here, we want to analyze the effect of such Toeplitzisation on the convergence. Therefore we consider three different variations:

1. the original non-Toeplitz coarse grid matrices on each level,
2. Toeplitzisation by modifying the original coarse grid matrix to Toeplitz form,
3. Toeplitzisation by cutting.

The coarse matrix for choice 1. is $A_{n^{(1)}} = P_n(p)T_n(f)P_n(p)^H$. The Toeplitzisation in case 2. was introduced in [15] choosing as coarse matrix $T_{n^{(1)}}(f_c)$ similar to a rediscretization. The Toeplitzisation by cutting in 3. was introduced in [21] and further improved in [2].

In the following we prove that the proposal in [2] gives the minimum cut that preserves the Toeplitz structure. Further, for a unique zero of order two all the previous strategies are equal if the projector is chosen with minimum bandwidth and $n = 2^\alpha - 1$, $\alpha \in \mathbb{N}$.

Let us discuss in detail the 1D case. We denote by δ_g the degree of an even trigonometric polynomial g, which has a symmetric stencil of the Fourier coefficients different from zero having length $2\delta_g + 1$ and central coefficient in position $\delta_g + 1$. We define the cutting matrix

$$K_i\{t\} = \left[\,0_t \middle| K_{n^{(i)}-2t} \middle| 0_t\,\right] \in \mathbb{R}^{n^{(i+1)}\times n^{(i)}}, \tag{6.1}$$

where t will be chosen as $t = \delta_p - 1$ when used for approach 3 and 0_t is the null matrix of size $n^{(i+1)} \times t$. To apply the multigrid method recursively when the cutting matrix (6.1) is used in approach 3, the size of the finer grid should be $2^\alpha - 1 - 2t$ (see [2]). The following proposition shows that the class of Toeplitz matrices has a quasi-algebra structure.

Proposition 6.1 *Let p and f be two even trigonometric polynomials, then*

$$[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]T_n(p)T_n(f)T_n(p)[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]^H = T_{n-2\delta_p}(p^2 f).$$

Proof First, we prove that

$$[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]T_n(p)T_n(f) = [0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]T_n(pf). \tag{6.2}$$

It is enough to observe that $[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]T_n(p)$ is the convolution matrix generated from p. We denote by the subscript i and the superscript j the ith row and the jth column of a matrix, respectively. Let g_k be the convolution of p and f centered at the kth coefficient of f, i.e., the kth Fourier coefficient of $g = pf$, then by definition

$$[T_n(p)T_n(f)]_{ij} = T_n(p)_i T_n(f)^j = g_{i-j} = [T_n(pf)]_{ij}, \tag{6.3}$$

for $i = \delta_p + 1, \dots, n - \delta_p$ and $j = 1, \dots, n$. We are not interested in the first and last δ_p rows of $T_n(p)T_n(f)$ since they are removed by the product with $[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]$.

Similarly

$$T_n(f)T_n(p)[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}] = T_n(pf)[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]. \tag{6.4}$$

The assertion follows combining equation (6.2) with equation (6.4) and observing that

$$[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}]T_n(p^2 f)[0_{\delta_p}\,|\,I_{n-2\delta_p}\,|\,0_{\delta_p}] = T_{n-2\delta_p}(p^2 f).$$

Corollary 6.1 *Fix $P_i = K_i\{t\}T_{n^{(i)}}(p_i)$, with $K_i\{t\}$ defined in* (6.1)*, p_i and f_i two even trigonometric polynomials, $t = \delta_{p_i} - 1$, and $n^{(i)} = 2^\alpha - 1 - 2t$, then*

$$P_i T_{n^{(i)}}(f_i) P_i^H = K_i\{t\}T_{n^{(i)}}(p_i^2 f_i)K_i\{t\}^H = T_{n^{(i+1)}}(f_{i+1}).$$

Furthermore, P_i preserves the Toeplitz structure cutting the minimum possible number of rows.

Proof It is enough to observe that $K_{n^{(i)}-2t}$ has the first and the last column equal to the null vector, and then apply the Proposition 6.1. The fact that by cutting one less row and column the Toeplitz structure is lost follows from the proof of Proposition 6.1 (compare equation (6.3)).

Remark 6.1 From the Corollary 6.1, we have that if $n = 2^\alpha - 1$, f has a zero of order two and p is chosen to have smallest bandwidth, then $t = 0$ and so all the different strategies 1–3 are equivalent.

For the multidimensional case, it is enough to extend definition (6.1) using tensor products and to use tensor product arguments in Proposition 6.1 and Corollary 6.1.

We now study the V-cycle optimality for $T_n(f)$ when f has a unique zero of order two. In the literature the V-cycle optimality in [1,2] was proved only in the algebra (circulant, tau, etc.) case. For the Toeplitz case, in [21] the TGM optimality was proved, while in [4] the level independency for zeros of order two has been shown. Note that the level independency does not necessarily imply the V-cycle optimality, but only the W-cycle optimality (see [2]).

For the V-cycle optimality we need to introduce the τ algebra. Multilevel τ matrices are simultaneously diagonalized by the multidimensional discrete sine transform. For $n \in \mathbb{N}$, the sine matrix of order n is

$$S_n = \sqrt{\frac{2}{n+1}} \left[\sin\left(j y_k^{(n)}\right)\right]_{k,j},$$

where $y_k^{(n)} = \pi k/(n+1)$, $k = 1, \dots, n$. The d-dimensional sine matrix is defined using tensor products as $S_n = S_{n_1} \otimes \cdots \otimes S_{n_d}$. The multilevel τ matrix generated by f is $\tau_n(f) = S_n D_n(f) S_n^H$. A τ matrix can be expressed as $\tau_n(f) = T_n(f) + H_n(f)$ where $H_n(f)$ is the centrosymmetric Hankel matrix generated by f. A Hankel matrix has the property that its entries are constant along any lower-left-upper-right diagonal and it can be defined similar to the Toeplitz case:

$$H_n(f) = \sum_{2 \leqslant |j_1| \leqslant n_1 - 1} \cdots \sum_{2 \leqslant |j_d| \leqslant n_d - 1} a_{(j_1,\dots,j_d)} K_{n_1}^{[j_1]} \otimes \cdots \otimes K_{n_d}^{[j_d]}, \tag{6.5}$$

where, for $n, j \in \mathbb{Z}$, $K_n^{[j]}$ denotes the matrix of order n whose entry (s,t) equals 1 if $s + t = j \bmod 2(n-1)$ and equals zero otherwise. From equation (6.5), it follows that if f has degree at most one in each variable then $\tau_n(f) = T_n(f)$.

We can now prove the V-cycle optimality of multigrid methods for Toeplitz matrices having a symbol of order two.

Theorem 6.1 *Let f be an even trigonometric polynomial having a unique zero of order two at x^0 and choose $p_i(x) = \prod_{j=1}^d (1 + \cos(x - [x_i^0]_j))$, where $x_i^0 \in \mathbb{R}^d$ is the zero of f_i, $i = 0, \dots m-1$ and $f_0 = f$. Then*

(i) the cutting strategies $1-3$ coincide,
(ii) the V-cycle is optimal if furthermore $f(x) = \sum_{j=1}^d (1 - \cos(x - [x^0]_j))$.

Proof First, we note that the functions f_i, for $i = 0, \dots m$, have a unique zero of order two which moves in $[0, \pi]^d$ according to Proposition 7.2 in [21] (see [10] for the 2D case).

The symbol p_i of the projector has degree one in each variable and so, according to Remark 6.1, the projection strategies and the coarse matrices in $1-3$ are the same. More in detail, $A_{n^{(i+1)}} = T_{n^{(i+1)}}(f_{i+1}) = P_i T_{n^{(i)}}(f_i) P_i^H$, where f_{i+1} is the coarse function obtained from f_i by applying equation (2.1).

To prove the V-cycle optimality, it is enough to observe that p_i and f_i have degree at most one in each variable for $i = 0, \dots, m-1$, and so $T_n(p_i) = \tau_n(p_i)$ and $T_n(f_i) = \tau_n(f_i)$. Therefore, the V-cycle optimality prove for the τ algebra presented in [1] can be applied.

We note that the assumption $f(x) = \sum_{j=1}^{d}(1 - \cos(x - [x^0]_j))$ in (ii) of Theorem 6.1 to prove the V-cycle optimality is only a structural assumption to work with the τ algebra also at the coarse levels. However, if f is an even trigonometric polynomial having a unique zero of order two at x^0, it is well-known that $\sum_{j=1}^{d}(1 - \cos(x - [x^0]_j))$ is an optimal preconditioner that gives a strong cluster at one (see [5]) and so if the V-cycle is optimal for $\sum_{j=1}^{d}(1 - \cos(x - [x^0]_j))$ the same is true for f.

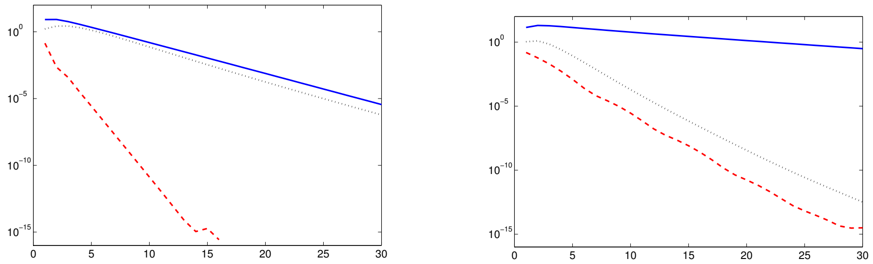
We have proved that for a zero of order 2 the three coarsening strategies are the same and give an optimal V-cycle. If f has a zero of order greater than two higher order projectors must be considered and the three coarsening strategies are no longer the same. We note that the strategies 1 and 2 can be applied for every n, while the strategy 3 requires $n = 2^\alpha + 1 - 2\delta_p$. We now give a numerical comparison for $\delta_p > 1$.

Example 6.1 We consider the fourth-order “long stencil” discretization of the 1D Laplacian leading to the generating symbol

$$f(x) = 30 - 32\cos(x) + 2\cos(2x).$$

We consider two different projectors: $p_{4/3}$ and p_2, where p_z is defined in (4.1). According to the analysis in Section 4, $p_{4/3}$ has a zero of order 4 at π and satisfies both conditions (2.2) and (3.5), while p_2 vanishes at π and $\pi/2$ and as a consequence violates condition (2.2b) but satisfies the new condition (3.5b). The projector $p_2(x) = 2\cos(x)(1 + \cos(x))$ is the 1D version of our 2D proposal in (5.3) analyzed in Example 5.1. One step of Gauss-Seidel as post-smoother is applied. The pre-smoother is one step of Richardson with relaxation parameter $\omega_i^{\text{pre}} = 1.5/\|f_i\|_\infty$ for $p_{4/3}$ and $\omega_i^{\text{pre}} = 1/f_i(\pi/2)$ for p_2 according to condition (3.5c).

Figure 6.1 shows the relative error at each iteration for the three considered coarsening strategies. We observe that Toeplitzisation can slow down the convergence, but such effect is largely reduced using p_2 with the strategy 3 (Toeplitzisation by cutting). In detail, combining the projector p_2 that does not satisfy the classical conditions (2.2) with an intermediate iteration as pre-smoother given by $\omega_i^{\text{pre}} = 1/f_i(\pi/2)$, we obtain that the Toeplitzification approach 3 has about the same effectiveness as the Galerkin approach 1 that destroys the Toeplitz structure at the coarser levels. Like in Example 5.1 the pre-smoother $\omega_i^{\text{pre}} = 1/f_i(\pi/2)$ is not convergent, but its combination with a post smoother – here a Gauss-Seidel smoother – leads to a converging multigrid method.



$p_i = p_{4/3}$, $\omega_i^{\mathrm{pre}} = 1.5/\|f_i\|_\infty$ $p_i = p_2$, $\omega_i^{\mathrm{pre}} = 1/f_i(\pi/2)$

Fig. 6.1 Relative error of Example 6.1 varying the iteration number: the dashed curve depicts the strategy 1. (the original Galerkin matrix), the solid curve depicts the strategy 2. (the modified Toeplitz matrix), the dotted curve depicts the strategy 3. (the Toeplitz matrix via cutting).

7 Conclusions

In this paper we analyzed the multigrid optimality conditions from a practical point of view to construct high order grid transfer operators at lower cost. This has lead to the new optimality conditions where the old condition (2.2b) is now relaxed with its new version (3.5b). This allows to use computationally convenient projections that lead to singular coarse grid problems. Hence a further condition (3.5c) is provided to ensure fast convergence adding a pre-smoother, i.e., an intermediate iteration, linked to the projection when necessary.

The new relaxed condition (3.5b) allows to define new projectors, like in equation (5.3), that do not satisfy the old condition (2.2b) but are computationally cheaper than the classical grid transfer operators used in the literature without increasing the convergence rate. Moreover combining such projectors with the intermediate iteration, the convergence rate is improved by a factor at least $1/3$ with respect to the classical choices for Toeplitz matrices.

Further, coarsening strategies for Toeplitz matrices have been investigated. We have proved that if projectors with order greater than two are not necessary, all the different approaches are equivalent and the V-cycle is optimal. For high-order projectors, like the ones introduced in this paper, we have numerically shown that the new conditions (3.5) are very effective when combined with the coarsening strategy that preserves the Toeplitz structure at each level by cutting.

References

1. Aricò, A., Donatelli, M.: A V-cycle multigrid for multilevel matrix algebras: proof of optimality. Numer. Math. **105**, 511–547 (2007)
2. Aricò, A., Donatelli, M., Serra-Capizzano, S.: V-cycle optimal convergence for certain (multilevel) structured linear systems. SIAM J. Matrix Anal. Appl. **26**(1), 186–214 (2004)
3. Bakhvalov, N.S.: On the convergence of a relaxation method with natural constraints on the elliptic operator. USSR Comp. Math. Math. Phys. **6**, 101–135 (1966)
4. Chan, R.H., Chang, Q.S., Sun, H.W.: Multigrid method for ill-conditioned symmetric Toeplitz systems. SIAM J. Sci. Comput. **19**(2), 516–529 (1998)

5. Chan, R.H., Ng, M.K.: Conjugate gradient methods for Toeplitz systems. SIAM Rev. **38**(3), 427–482 (1996)
6. Cheng, L., Wang, H., Zhang, Z.: The solution of ill-conditioned symmetric toeplitz systems via two-grid and wavelet methods. Comput. Math. Appl. **46**(5–6), 793–804 (2003)
7. Donatelli, M.: An algebraic generalization of local Fourier analysis for grid transfer operators in multigrid based on Toeplitz matrices. Num. Lin. Alg. Appl. **17**, 179–197 (2010)
8. Fedorenko, R.P.: The speed of convergence of one iterative process. USSR Comp. Math. Math. Phys. **4**(3), 227–235 (1964)
9. Fiorentino, G., Serra, S.: Multigrid methods for Toeplitz matrices. Calcolo **28**, 238–305 (1991)
10. Fiorentino, G., Serra, S.: Multigrid methods for symmetric positive definite block Toeplitz matrices with nonnegative generating functions. SIAM J. Sci. Comput. **17**(5), 1068–1081 (1996)
11. Hackbusch, W.: Convergence of multi-grid iterations applied to difference equations. Math. Comp. **34**, 425–440 (1980)
12. Hackbusch, W.: On the convergence of multi-grid iterations. Beiträge Numer. Math. **9**, 213–239 (1981)
13. Hackbusch, W.: Multi-grid convergence theory. In: W. Hackbusch, U. Trottenberg (eds.) Multigrid methods, *Lecture Notes in Mathematics*, vol. 960, pp. 177–219. Springer-Verlag, Berlin (1982)
14. Hemker, P.W.: On the order of prolongations and restrictions in multigrid procedures. J. Comp. Appl. Math. **32**, 423–429 (1990)
15. Huckle, T., Staudacher, J.: Multigrid preconditioning and toeplitz matrices. Electron. Trans. Numer. Anal. **13**, 82–105 (2002)
16. McCormick, S.F.: Multigrid methods for variational problems: General theory for the V-cycle. SIAM J. Numer. Anal. **22**(4), 634–643 (1985)
17. McCormick, S.F., Ruge, J.W.: Multigrid methods for variational problems. SIAM J. Numer. Anal. **19**(5), 924–929 (1982)
18. Napov, A., Notay, Y.: Smoothing factor, order of prolongation and actual multigrid convergence. Numer. Math. **118**, 457–483 (2011)
19. Ruge, J.W., Stüben, K.: Algebraic multigrid. In: S.F. McCormick (ed.) Multigrid methods, *Frontiers Appl. Math.*, vol. 3, pp. 73–130. SIAM, Philadelphia (1987)
20. Serra, S.: Multi-iterative methods. Comput. Math. Appl. **26**(4), 65–87 (1993)
21. Serra-Capizzano, S.: Convergence analysis of two-grid methods for elliptic toeplitz and pdes matrix sequences. Numer. Math. **92**, 433–465 (2002)
22. Serra-Capizzano, S., Tablino-Possio, C.: Multigrid methods for multilevel circulant matrices. SIAM J. Sci. Comput. **26**(1), 55–85 (2004)
23. Trottenberg, U., Oosterlee, C., Schüller, A.: Multigrid. Academic Press, San Diego (2001)
24. Tyrtyshnikov, E.: A unifying approach to some old and new theorems on preconditioning and clustering. Linear Algebra Appl. **232**, 1–43 (1996)
25. Wienands, R., Joppich, W.: Practical Fourier analysis for multigrid methods, *Numerical Insights*, vol. 4. Chapman & Hall/CRC, Boca Raton (2005)
26. Yavneh, I.: Coarse-grid correction for nonelliptic and singular pertubation problems. SIAM J. Sci. Comput. **19**(5), 1682–1699 (1998)