Structured Matrices, Multigrid Methods, and the Helmholtz Equation

Rainer Fischer [1], Thomas Huckle [2]

Technical University of Munich, Germany

Abstract

Discretization of the Helmholtz equation with certain boundary conditions results in structured linear systems which are associated with generating functions. Depending on the type of boundary conditions one obtains matrices of a certain class. By solving these systems with normal equations, we have the advantage that the corresponding generating functions are nonnegative, although they have a whole curve of zeros. The multigrid methods we develop are especially designed for structured matrix classes, making heavy use of the associated generating functions. In a previous article [13] , we have extended well-known multigrid methods from the isolated zero case to functions with zero curves. In those methods the whole zero curve is represented on coarser grids. Since this implies that zero curves become larger on coarser levels, the number of levels in our multigrid method is limited. Therefore we propose a different approach in this paper. We split the original problem into a fixed number of coarse grid problems. Corresponding to a generating function with isolated zeros, each of them locally represents one part of the zero curve. We combine this splitting technique with the methods from [13] to construct a faster and more robust multigrid solver.

Key words: Helmholtz equation, multilevel Toeplitz, trigonometric matrix algebras, multigrid methods, iterative methods

1 Introduction

Several articles have been published concerning the solution of the Helmholtz equation with multigrid methods, see for example [3,8,9] . In this article, we wish to present a different approach to the multigrid solution of the Helmholtz equation with constant coefficients. It is primarily based on certain classes of structured matrices and their strong correspondence to generating functions. Depending on the type of boundary conditions, the discretized Helmholtz equation is a linear system of Toeplitz, tau, circulant, or DCT-III type. Starting with [10,11] a specific theory of multigrid methods for these structured matrix classes has been developed which is based on the AMG approach of Ruge and Stüben [15] . We extend theoretical results on multigrid methods for certain matrix algebras, and apply them to the solution of the Helmholtz equation.
This article is organized as follows. In Chapter 2 we describe the discretized Helmholtz equation with different kinds of boundary conditions, and summarize fundamental properties of the resulting matrix classes and corresponding generating functions. In Chapter 3

[1] Institut für Informatik, Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany, email: fischerr@in.tum.de
[2] Institut für Informatik, Technische Universität München, Boltzmannstr. 3, 85748 Garching b. München, Germany, email: huckle@in.tum.de

we review results from [13] on multigrid methods for matrices whose generating functions have whole zero curves. These methods are based on the idea of representing the whole zero curve on all grids. Furthermore, we extend these results to the DCT-III algebra. Chapter 4 introduces a different type of multigrid method, which is based on a splitting technique. Several coarse grid corrections are computed, each of them representing the zero curve locally. We describe how the two techniques can be combined to construct a multigrid solver which has the advantages of both methods. Chapter 5 summarizes the main results and points to some directions for future work.

2 The Helmholtz equation and structured matrix classes

The 2D Helmholtz equation is a boundary value problem given by

$$\begin{aligned} -\Delta u - k^2 u &= g \qquad \text{on} \quad \Omega \subset \mathbb{R}^2 \,, \\ Bu &= h \qquad \text{on} \quad \Gamma \subset \delta\Omega \,, \end{aligned} \tag{1}$$

where k is a constant wavenumber and B an operator on the boundary Γ. Discretization of the Helmholtz equation is done for example with a five-point finite difference scheme. Depending on the boundary condition this results in a sparse and structured matrix A of a certain class. For example, if Dirichlet boundary conditions are imposed on Γ, the matrix belongs to both the two-level Toeplitz and the two-level tau class. For periodic boundary conditions we obtain a two-level circulant matrix, and for certain Neumann boundary conditions a two-level DCT-III matrix. Since all these matrix classes are strongly related to generating functions, we wish to apply the theory of generating function to the solution of the discretized Helmholtz equation.

For the Toeplitz case the correspondence between functions and matrices is described e.g. in [14]. Suppose $f(x,y)$ is a real-valued Lebesgue-integrable function which is defined on $]-\pi,\pi]^2$, and periodically extended on the whole plane. $A_{mn}[f]$ is the corresponding mn-by-mn two-level Toeplitz matrix whose entries are the Fourier coefficients of f. If f_{min} and f_{max} denote the infimum and supremum values of f, and if $f_{min} < f_{max}$, then for all $m,n \geq 1$, the eigenvalues of $A_{mn}[f]$ lie in the interval (f_{min}, f_{max}). For $n,m \to \infty$, the extreme eigenvalues tend to f_{min} and f_{max}. Matrices belonging to a two-level trigonometric algebra are diagonalized by a unitary transform Q_{mn}, i.e. they are of the form

$$A_{mn}[f] = Q_{mn}^H \cdot \Lambda_{mn}[f] \cdot Q_{mn} \quad , \tag{2}$$

where $\Lambda_{mn}[f]$ is the diagonal matrix containing the eigenvalues $\lambda_{k,l}$ of $A_{mn}[f]$. For two-level circulant and DCT-III matrices these are given by $\lambda_{k,l} = f(\frac{2\pi l}{m}, \frac{2\pi k}{n})$ with $0 \leq k \leq n-1, 0 \leq l \leq m-1$, and for two-level tau matrices by $\lambda_{k,l} = f(\frac{\pi l}{m+1}, \frac{\pi k}{n+1})$ with $1 \leq k \leq n, 1 \leq l \leq m$. This implies that unlike Toeplitz matrices, matrices from trigonometric algebras can become singular if f is zero at one of the grid points. If this happens, $A_{mn}[f]$ is usually replaced by the so called Strang correction [16], which was originally defined for circulant matrices, but which can be used for other matrix algebras as well [2]. Since the matrix A of the discrete Helmholtz equation is indefinite, we solve the corresponding linear system via normal equations. The generating function f, which corresponds to the matrix $A_{mn}[f] = A^T A$, is of the form

$$f(x,y) = (\rho - \cos(x) - \cos(y))^2 \qquad (0 < \rho \leq 2)\,, \tag{3}$$

where ρ takes the value $2 - \frac{k^2 h^2}{2}$ with k denoting the wavenumber and h the size of a discretization step, see [9]. If $|\rho| = 2$, $f(x,y)$ has a single isolated zero of order 4, whereas for $|\rho| < 2$, f is zero along a whole curve. This zero curve becomes larger as ρ decreases. Depending on the boundary condition, $A_{mn}[f]$ is the two-level tau, circulant, or DCT-III matrix corresponding to $f(x,y)$.

3 Multigrid methods for functions with whole zero curves

For structured matrices corresponding to generating functions multigrid methods are among the fastest iterative solvers. For large matrix classes they have optimal computational cost, i.e. asymptotically the same as one matrix-vector product. For banded matrices this means that the cost is $O(mn)$ compared to $O(mn \cdot \log(mn))$ of the fastest direct methods. Moreover, multigrid methods can be extended to the case of variable coefficients. Extensive research has been done on the development of multigrid methods for matrices corresponding to functions with isolated zeros. For Toeplitz and tau matrices we refer to [10,11,4,2,7], for circulant matrices to [5], and for DCT-III matrices to [6]. These results are the starting point for the development of our methods, which are designed for functions with whole zero curves such as f from (3). After briefly introducing the methods for the isolated zero case at the beginning of this chapter, we review our results on functions with zero curves [13] and extend them to the DCT-III algebra.

3.1 Galerkin-based multigrid for functions with isolated zeros

Most multigrid methods for structured matrices which have been developed so far are based on the AMG method of Ruge and Stüben [15]. This means that usually a simple smoother S such as the damped Richardson, damped Jacobi, or Gauss-Seidel method is used. After defining the restriction matrix R, the coarse-grid matrix A_C is computed with the Galerkin approach $A_C = RA_{mn}[f]R^H$, and hence the coarse grid correction X with $X = I - R^H A_C^{-1} R A_{mn}[f]$. If ν_1 denotes the number of presmoothing steps and ν_2 the number of postsmoothing steps, the iteration matrix TG of the two grid method (TGM) is $TG = S^{\nu_2} X S^{\nu_1}$. The TGM is extended to a multigrid method (MGM) by recursively using the TGM scheme for A_C instead of inverting A_C exactly. Ruge and Stüben have proved convergence of the TGM if presmoothing condition (4), postsmoothing condition (5), and correcting condition (6) are satisfied.

Theorem 1 (Ruge, Stüben,[15])
Let A be a positive definite mn-by-mn matrix, and let S and R be smoother and restriction operator. Suppose that there exist $\alpha_{pre}, \alpha_{post}, \beta$ such that

$$\|S^{\nu_1}x\|_A^2 \le \|x\|_A^2 - \alpha_{pre}\|x\|^2_{A\cdot diag\cdot(A)^{-1}A}\,, \quad \forall x \in \mathbb{R}^{mn}\,, \tag{4}$$

$$\|S^{\nu_2}x\|_A^2 \le \|x\|_A^2 - \alpha_{post}\|S^{\nu_2}x\|^2_{A\cdot diag\cdot(A)^{-1}A}\,, \quad \forall x \in \mathbb{R}^{mn}\,, \tag{5}$$

$$\min_{y\in\mathbb{R}^{m_C n_C}} \|x - R^H y\|^2_{diag(A)} \le \beta\|x\|_A^2\,, \quad \forall x \in \mathbb{R}^{mn}\,. \tag{6}$$

Then $\beta > \alpha_{post}$, and the convergence factor of the two-level method $\|TG\|_A$ is bounded by $\|TG\|_A \le \sqrt{\frac{1-\alpha_{post}/\beta}{1+\alpha_{pre}/\beta}}$.

TGM and MGM for two-level tau, circulant, and DCT-III matrices can be described in terms of generating functions. The restriction is split into two parts $R = B \cdot E$, where B is in the same class as $A_{mn}[f]$, corresponding to a function $b(x,y)$, and E is the elementary restriction matrix of the class. Computation of the coarse grid matrix A_C is done in two steps, $\hat{A} = B \cdot A_{mn}[f] \cdot B^H$ and $A_C = E \cdot \hat{A} \cdot E^H$. Since for trigonometric matrix algebras $\hat{A}$ is still in the same class, the first product is directly translated to generating functions:

$f(x,y) = f(x,y)\cdot b^2(x,y)$. Elementary restriction represents the spectral link between the space of frequencies on the fine and on the coarse grid. With the choice

$$E_n^{(\mathrm{circ})} = \begin{pmatrix} 1\ 0 & & & \\ & 1\ 0 & & \\ & & \ddots\ \ddots & \\ & & & 1\ 0 \end{pmatrix},\ E_n^{(\mathrm{tau})} = \begin{pmatrix} 0\ 1\ 0 & & & \\ & 0\ 1\ \ 0 & & \\ & & \ddots\ \ddots\ \ddots & \\ & & & 0\ \ 1\ 0 \end{pmatrix},\ E_n^{(\mathrm{DCT\text{-}III})} = \begin{pmatrix} 1\ 1 & & & \\ & 1\ 1 & & \\ & & \ddots\ \ddots & \\ & & & 1\ 1 \end{pmatrix} \quad (7)$$

of the one-dimensional restriction matrices, we obtain $E = E_m \otimes E_n$ with the respective matrices E_m and E_n from (7). In the DCT-III algebra, computation of $A_C = E\cdot\hat{A}\cdot E^H$ is translated to

$$\begin{aligned} f_2(x,y) = 4[&\cos^2(\frac{x}{4})\cos^2(\frac{y}{4})\hat{f}(\frac{x}{2},\frac{y}{2}) + \sin^2(\frac{x}{4})\cos^2(\frac{y}{4})\hat{f}(\pi-\frac{x}{2},\frac{y}{2}) \\ &+\cos^2(\frac{x}{4})\sin^2(\frac{y}{4})\hat{f}(\frac{x}{2},\pi-\frac{y}{2}) + \sin^2(\frac{x}{4})\sin^2(\frac{y}{4})\hat{f}(\pi-\frac{x}{2},\pi-\frac{y}{2})]\ , \end{aligned} \quad (8)$$

whereas for tau and circulant matrices f_2 is computed as in [13]. For the Toeplitz class additional cutting must be applied such that A_C has again Toeplitz structure [4,2].
At the very heart of the TGM and MGM algorithms is the choice of $b(x,y)$. If f is zero at (x_0,y_0), then b is chosen to be zero at the three mirror points, which are

$$M((x_0,y_0)) = \{(\pi-x_0,y_0),(x_0,\pi-y_0),(\pi-x_0,\pi-y_0)\} \quad (9)$$

in the tau and DCT-III case. For circulant matrices $\pi - x_0$ and $\pi - y_0$ are replaced by $x_0+\pi$ and $y_0+\pi$. A typical choice for b in the tau or DCT-III case is

$$b(x,y) = (\cos(x_0)+\cos(x))^k(\cos(y_0)+\cos(y))^k\ . \quad (10)$$

More precisely, the following two conditions must be satisfied for two-grid convergence [2]:

$$\limsup_{(x,y)\to(x_0,y_0)} \frac{b^2((x_i,y_i))}{f((x,y))} < \infty \quad \text{for } (x_i,y_i)\in M((x_0,y_0)) \quad (11)$$

$$0 < \sum_{(x_i,y_i)\in M((x_0,y_0))\cup\{(x_0,y_0)\}} b^2((x_i,y_i))\ . \quad (12)$$

In [4] convergence of the TGM is proved for multilevel tau and Toeplitz matrices, in [5] for multilevel circulant matrices, and in [6] for multilevel DCT-III matrices. MGM convergence proofs for multilevel tau and circulant matrices can be found in [2,1], where the slightly stronger $\left|\frac{b((x_i,y_i))}{f((x,y))}\right|$ must be used in (11). For the DCT-III algebra and for the Toeplitz class such a proof has not yet been given.

3.2 Functions with zero curves: TGM convergence and MGM problems

In [13] we proposed a Galerkin-based TGM which extends the method from Chapter 3.1 to generating functions with whole zero curves, using the same smoother S and the same elementary restriction matrix E. Conditions (11) and (12) must hold for each zero on the curve. This means b is chosen to be zero at three mirror *curves*. For the tau, DCT-III, and Toeplitz class such a function is

$$b(x,y) = f(\pi-x,y)\cdot f(x,\pi-y)\cdot f(\pi-x,\pi-y)\ , \quad (13)$$

whereas for circulant matrices terms such as $\pi - x$ are replaced by $x+\pi$. With this choice of b, f_2 is computed as in the isolated zero case. In [13] we have proved convergence of the TGM for two-level tau and circulant matrices. Here we show that the following theorem also holds for DCT-III matrices.

Theorem 2
Let $A := A_{mn}[f]$ be a two-level matrix from the circulant, tau, or DCT-III algebra. Assume that $f(x,y)$ is a cosine nonnegative polynomial (not identically zero) with a zero curve in $]-\frac{\pi}{2},\frac{\pi}{2}[^2$. Suppose that the smoother is the damped Richardson or Jacobi method. Furthermore, let the restriction be $R = B \cdot E$ with B corresponding to $b(x,y)$ from (13) and with the elementary restriction matrix E of the respective algebra. Then there exist $\alpha_{pre}, \alpha_{post}, \beta > 0$ such that conditions (4)-(6) are satisfied, and the TGM converges.

Proof: In [5,1] the smoothing conditions (4) and (5) are proved for multilevel circulant and tau matrices. For multilevel DCT-III matrices (5) is proved in [6], the postsmoothing condition is proved analogously. The proof of (6) for two-level tau and circulant matrices can be found in [13] . Therefore it remains to show that (6) is satisfied for two-level DCT-III matrices. This is similar to the proof in [13], this time the proof of [6] must be extended to functions with whole zero curves. As in [13] we have to show the 4-by-4 matrix inequality $L(x,y) \leq \frac{\beta}{\hat{a}} I_4$ with

$$L(x,y) = diag(f[x,y]))^{-1/2}(I_4 - \frac{1}{\|b[x,y]\|_2^2} b[x,y](b[x,y])^T)diag(f[x,y]))^{-1/2} ,$$

where $\hat{a} = \max_j A_{jj}$ and $f[x,y] := (f(\bar{\mathbf{x}}_1), f(\bar{\mathbf{x}}_2), f(\bar{\mathbf{x}}_3), f(\bar{\mathbf{x}}_4))$ with $\bar{\mathbf{x}}_1 = (x,y)$ and its mirror points $\bar{\mathbf{x}}_2, \bar{\mathbf{x}}_3, \bar{\mathbf{x}}_4$. However, this time we have

$$b[x,y] := (\cos(\frac{x}{2})\cos(\frac{y}{2})b(\bar{\mathbf{x}}_1), -\sin(\frac{x}{2})\cos(\frac{y}{2})b(\bar{\mathbf{x}}_2), -\cos(\frac{x}{2})\sin(\frac{y}{2})b(\bar{\mathbf{x}}_3), \sin(\frac{x}{2})\sin(\frac{y}{2})b(\bar{\mathbf{x}}_4)) .$$

It remains to show that $L(x,y)$ is uniformly bounded in spectral norm, which is in turn true if each element $L_{i,j}(x,y)$ of the 4-by-4 matrix function $L(x,y)$ is bounded in L^∞ . In [6] this is done with (11) and (12) . With our choice of b with (13) and with the additional requirement that the zero curve must lie inside $]-\pi/2, \pi/2[^2$, it is guaranteed that (11) and (12) hold for each zero of f on the curve. Therefore, each $L_{i,j}(x,y)$ is shown to be bounded, and (6) holds with $\beta > 0$. ∎

For the construction of a Galerkin-based MGM we chose

$$b_2(x,y) = f_2(\pi - x, y) \cdot f_2(x, \pi - y) \cdot f_2(\pi - x, \pi - y) \tag{14}$$

or its circulant equivalent to compute the restriction matrix on the next level. This leads to a function $f_3(x,y)$, which has the same zero curve as $f(x/4, y/4)$. Figure 1 depicts the zero curves of f ,f_2 , and f_3 for the function f from (3) with $\rho = 1.9$, where $f(x,y) = 0$ is the smallest and $f_3(x,y) = 0$ is the largest of the three curves. This MGM has optimal convergence behavior, but it is not practical for the following two reasons:

- Matrices become significantly denser on coarser levels. Since the matrices we are interested in are sparse, only one or two coarsening steps can be performed before matrices become dense and the method is completely inefficient.
- As it is shown in the left picture of Figure 1 zero curves become larger on each level, limiting the number of grids which can be used. Another coarsening step can only be performed if the zero curve on some level does not come close to the boundaries of $]-\pi,\pi[^2$. Otherwise (11) or (12) are not satisfied anymore for some points on the curve, and the MGM breaks down in most numerical experiments.

3.3 Rediscretization on coarser levels

The MGM proposed in [13] for circulant and tau matrices overcomes the first of the two problems. It uses the same smoothing, prolongation, and restriction, but A_C is not computed with the Galerkin approach. Instead, A_C is chosen corresponding to a function

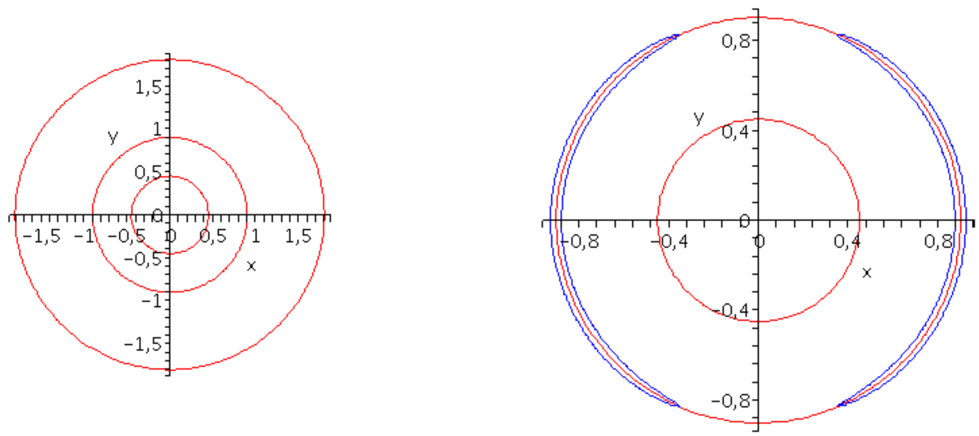

Fig. 1. Left: Zero curves of f , f_2 , f_3 for $\rho = 1.9$ (Section 3.2) ; Right: Zero curves of f , f_2 and level curve $f_2^{(j)} = 0.005$ approximating $f(x/2, y/2) = 0$ in the neighborhood of the points $(\pm x_j, 0)$ (Section 4.1)

f_2 , whose zero curve only approximates $f(x/2, y/2) = 0$. This can be interpreted as a rediscretization on the coarse grid. f_2 is a cosine polynomial of low degree with some free parameters which are determined such that the two zero curves match at some points. Two possible choices for f_2 are

$$\begin{aligned} f_2(x,y) &= [\sigma - \alpha(\cos(x) + \cos(y)) - \beta\cos(x)\cos(y)]^2 \\ f_2(x,y) &= [\sigma - \beta(\cos(x) + \cos(y)) - \alpha\cos(x)\cos(y) - \gamma(\cos(2x) + \cos(2y))]^2 \ , \end{aligned} \qquad (15)$$

Due to the symmetry of f the zero curve of the first function has 8 points in common with $f(x/2, y/2) = 0$, the zero curve of the second function 16 points.
For the construction of an MGM, b_2 is computed with (14) , using f_2 from (15) . The function f_3 on the next grid is chosen such that its zero curve has common points with $f(x/4, y/4) = 0$. This procedure continues on coarser levels, leading to coarse grid matrices which have the same sparsity pattern as A and A_C . Therefore each V-cycle iteration of the MGM is fast to compute. Furthermore, the approximations obtained with (15) are accurate enough to ensure fast convergence of the MGM. For the circulant and tau algebras this is confirmed by the numerical results in [13] . In order to show that this approximation technique works unchanged for the DCT-III algebra we apply a four-grid method to the DCT-III matrices corresponding to f from (3) .

ρ	$n=(2^6)^2$	$n=(2^7)^2$	$n=(2^8)^2$	$n=(2^9)^2$
1.99	15	15	15	15
1.98	15	15	15	15
1.97	16	16	16	16

However, the MGM does not get rid of the second problem described in Chapter 3.2 . Therefore, only $k + 1$ grids can be used in the MGM, where k is the smallest integer such that $f(x/2^k, y/2^k) = 0$ reaches or crosses the boundaries of $]-\pi/2, \pi/2[^2$. For the four-grid method this means that fast convergence is obtained only for ρ not much smaller than 1.97 . In terms of the Helmholtz equation this is equivalent to $k \cdot h < 0.25$.

4 A new splitting technique

In the following we propose a method which also overcomes the second problem. Instead of a single coarse grid correction several coarse grid corrections are computed in every iterative step, each of them representing one part of the zero curve of f. We start with the description of a TGM, explaining the principal ideas of our approach. Then we extend the TGM to an MGM and finally combine the method with the ones from Chapter 3.

4.1 A TGM with splitting

In the Galerkin-based TGM from Chapter 3.2 the coarse grid correction X in each iteration is computed from one coarse grid matrix A_C, which corresponds to the function $f_2(x,y)$. f_2 has the same zero curve as $f(x/2,y/2)$, and therefore represents the whole zero curve of $f(x,y)$ on the coarse grid. Now we compute k coarse grid corrections X_j with restriction matrices R_j and coarse grid matrices A_{C_j} in every iteration of the TGM. Each of the corresponding generating functions $f_2^{(j)}$ is a local coarse grid representation of the zero curve $f(x/2,y/2)=0$ in the neighborhood of at least one of its points (x_j,y_j). The iteration matrix TG of the TGM is of the form

$$TG = S^{\nu_{k+1}} \cdot \prod_{j=1}^{k} (X_j \cdot S^{\nu_j}) \tag{16}$$

with smoother S and coarse grid corrections $X_j = I - R_j^H A_{C_j}^{-1} R_j A_{mn}[f]$. This means one TGM iteration consists of k coarse grid corrections X_j and smoothers between the X_j. It is important that each part of the zero curve is approximated well and that the number of coarse grid corrections per iteration is small. Each coarse grid matrix A_{C_j} is computed as follows. The restriction matrices are of the form $R_j = B_j \cdot E$, where B_j corresponds to a function $b_j(x,y)$ and E is the elementary restriction matrix defined in (7). Since f from (3) is symmetric in x and y, b is chosen such that $f_2^{(j)}$ represents the zero curve of $f(x/2,y/2)$ in two or four point. A zero of f at (x_j,y_j) implies that f is also zero at $(-x_j,y_j),(x_j,-y_j),(-x_j,-y_j)$. If b_j is chosen to be

$$b_j(x,y) = (\cos(x_j)+\cos(x))^2(\cos(y_j)+\cos(y))^2 \quad . \tag{17}$$

and if A_{C_j} is computed with the Galerkin approach $A_{C_j} = R_j A_{mn}[f] R_j^H$, the zero curve of $f(x/2,y/2)$ is approximated very well in the neighborhood of the points $(2x_j,2y_j)$, $(-2x_j,2y_j),(2x_j,-2y_j),(-2x_j,-2y_j)$. For a point $(x_j,0)$ on the x-axis, $f_2^{(j)}$ is zero at $(-2x_j,0),(2x_j,0)$ and very small in the neighborhood of the two points. The right picture of Figure 1 shows the zero curves of $f(x,y)$ and $f(x/2,y/2)$, and the level curve $f_2^{(j)}(x,y) = 0.005$. From this picture we see that large parts of the zero curve of $f(x/2,y/2)$ are approximated well by $f_2^{(j)}$. The following theorem summarizes properties of A_{C_j} and $f_2^{(j)}$.

Theorem 3
Let $f(x,y)$ be the function defined in (3) with $\rho > 1$, and $A_{mn}[f]$ the corresponding matrix of the two-level tau, circulant or DCT-III algebra. Furthermore, let (x_j,y_j) be a point on the zero curve of $f(x,y)$, and assume that $R_j = B_j \cdot E$ with B_j corresponding to b_j from (17). Moreover, assume that A_{C_j} is computed with the Galerkin approach. Then the following holds:

(1) A_{C_j} is symmetric positive definite and its generating function $f_2^{(j)}$ has zeros at $(2x_j,2y_j),(-2x_j,2y_j),(2x_j,-2y_j),(-2x_j,-2y_j)$ and is positive elsewhere in $]-\pi,\pi]^2$.

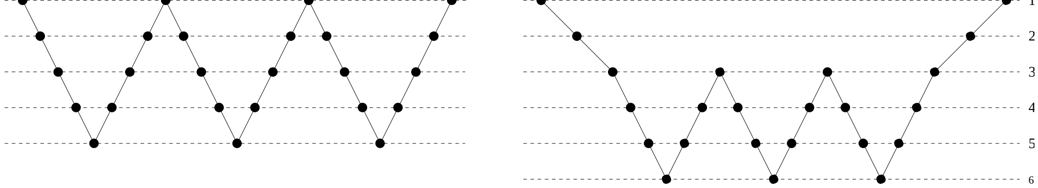

Fig. 2. One iteration of the MGM with the pure splitting method from Section 4.2 (left) and with the combined method from Section 4.3 (right)

(2) In the direction t of the tangent on $f(x/2,y/2)$ in (x_j,y_j) the first directional derivatives of $f(x/2,y/2)$ and $f_2^{(j)}(x,y)$ are both zero.

Proof: From [2] we know that A_{C_j} is the matrix of the respective algebra corresponding to $f_2^{(j)}$. Since $f_2^{(j)}$ is real-valued even and nonnegative, A_{C_j} is symmetric positive definite. Due to $\rho > 1$ the curve $f(x,y) = 0$ is located within $]-\pi/2, \pi/2[$. Then by the results from [2], $f_2^{(j)}$ is zero at the four points. From (8) we can deduce that $f_2^{(j)}$ is strictly positive at all other points.
For zeros on the x- and y-axis, the result stated in (2) follows from direct calculation, for other zeros it is obtained by a simple coordinate transformation. ■

4.2 From two-grid to multigrid

The TGM becomes an MGM if each coarse grid system in (16) with the matrix A_{C_j} is solved recursively with the MGM scheme from Chapter 3.1. This implies that one iteration of the MGM consists of k V-cycles instead of one in the method from Chapter 3.3. The left picture of Figure 2 shows one iteration of the MGM which uses three coarse grid corrections and five grids for each of them. Since A_{C_j} corresponds to a generating function $f_2^{(j)}$ with isolated zeros, the functions $b_2^{(j)}, b_3^{(j)}$, etc. corresponding to the restriction matrices on coarser levels are of the form (10). Each zero $(2x_j, 2y_j)$ of $f_2^{(j)}$ moves to $(4x_j, 4y_j), (8x_j, 8y_j), \ldots$ on the next levels. Thus extra care has to be taken if either the x- or the y-value of the zero approaches $\pi/2 \pmod{\pi}$.
From the form of the level curves of $f_2^{(j)}$ we see that the coarse grid matrices are of slightly anisotropic type. Especially when ρ becomes smaller, some of the coarse grid matrices have significant anisotropies. In these cases we can use the results from [12] and apply one or two semicoarsening steps. This means that (17) is replaced by functions such as $b(x,y) = (\cos(x_0) - \cos(x))^2$. Elementary restriction is done only in one direction, i.e. the matrix size is reduced only by a factor of 2.
For the design of an efficient MGM the number of coarse grid corrections is critical. For zeros of $f(x,y)$ on the x- or y-axis, one coarse grid correction approximates the zero curve of $f(x/2,y/2)$ in the neighborhood of two points, for all other zeros in the neighborhood of four points. Since we ususally start with the zeros on the axes, this means that $\frac{k}{4}+1$ coarse grid corrections X_j are needed for an approximation in k equidistant points. Thus 2 corrections are needed for 4 points, 3 corrections for 8 points, and 5 corrections for 16 points. The number of necessary points depends on two factors, the size of the zero curve and the size of the matrices. To illustrate that indeed a fairly small number of coarse grid corrections is needed, we have tested our method for f from (3) and Dirichlet boundary conditions. The following table contains the number of necessary coarse grid corrections ($\sharp cgc$) and the iteration numbers for a three-level method. Since each iteration contains two or three V-cycles, it roughly corresponds to to or three iterations of a method from Section 3.3.

ρ	$\sharp cgc$	$m{\cdot}n{=}(2^5)^2{-}1$	$m{\cdot}n{=}(2^6)^2{-}1$	$m{\cdot}n{=}(2^7)^2{-}1$	$m{\cdot}n{=}(2^8)^2{-}1$
1.95	2	4	4	4	4
1.9	2	6	6	6	6
1.8	3	4	4	4	4
1.6	3	5	5	5	6

4.3 Combining both strategies

So far we have defined two different kinds of multilevel methods for matrices corresponding to generating functions with whole zero curves. Both strategies have advantages and disadvantages:

(I) The strategy from Section 3.3 uses only one coarse grid matrix on each level. The main advantage of this method is that the zero curve of f is represented very well by generating functions on coarser levels, which leads to fast convergence. The main disadvantage is the increase of the zero curve on each level, limiting the possible number of levels in our multigrid method.

(II) The splitting strategy from Sections 4.1 and 4.2 has the main advantage that zero curves do not grow and therefore a much larger number of levels can be used. The disadvantage of these methods is that for very large matrices an approximation in 4 or 8 points is not accurate enough, and therefore k may become too big for a fast algorithm.

Therefore, we suggest to apply an MGM which combines the advantages of both types. We use the following heuristic: *Start with the first strategy until the zero curve is too large for another coarsening step. Then split the resulting coarse grid problem into k subproblems and apply further levels of restriction to each of them.*

To illustrate this new strategy the right picture of Figure 2 depicts one MGM iteration. It starts with two coarsening steps using the first strategy and then computes three coarse grid corrections on three further levels, leading to an approximation in the neighborhood of 8 points. The main advantage of this method is the following. After a few coarsening steps with the strategy **(I)** we obtain a matrix which is considerably smaller than $A_{mn}[f]$, because in each step the number of rows and columns is reduced by a factor 4 each. Thus a slightly larger number of necessary coarse grid corrections (due to the increased size of the zero curve after a few coarsening steps), is still acceptable.

In many applications discretization of the Helmholtz equation results in a function f from (3) with $1.99 < \rho < 2$. In this case we can apply several coarsening steps with strategy **(I)** before the zero curve reaches the boundaries of $]-\pi/2, \pi/2[$. Then we switch to strategy **(II)**. The following numerical results are obtained for two-level DCT-III matrices corresponding to f from (3). For $\rho = 1.95$ we test two different methods. Method 1 uses two steps of the first strategy followed by splitting into three subproblems and one further level of prolongation. Method 2 uses one step of the first strategy followed by splitting into three subproblems and two more coarsening steps.

ρ	method	$m{\cdot}n{=}(2^5)^2$	$m{\cdot}n{=}(2^6)^2$	$m{\cdot}n{=}(2^7)^2$	$m{\cdot}n{=}(2^8)^2$
1.95	1	13	13	14	14
1.95	2	9	9	10	10

5 Conclusions and future work

In this article we have presented a multigrid method for the solution of the Helmholtz equation with different boundary conditions. The method is based on the strong correspondence between certain classes of structured matrices and generating functions. Fast convergence is obtained by splitting the original problem into a small number of coarse grid problems, each of them representing part of the zero curve of f . The following two aspects have not been discussed in this article for the sake of brevity:

- Anisotropic versions of the Helmholtz equation, whose discretization leads to generating function such as

$$f(x,y) = (\rho - \alpha\cos(x) - (2-\alpha)\cos(y))^2 \qquad (0 < \rho \leq 2 \ , \ 0 \leq \alpha \ll 1)\,, \qquad (18)$$

  are solved by combining the methods described in [12] with the methods from this paper. In the MGM from Chapter 3.3 we replace full coarsening by semicoarsening, and the methods from Chapter 4 are modified to deal with less symmetry in f .
- Multigrid methods are not only used as solvers, but also as preconditioners for Krylov subspace methods by splitting the original problem into a fixed number of subproblems. Different choices for the corresponding generating functions f_j lead to satisfactory convergence results.

The treatment of other boundary value problems will be subject of future research. A straightforward extension should be possible for other Toeplitz plus Hankel matrices such as the DST-III matrices. A more difficult task is an adaptation of our methods to complex boundary conditions such as the Sommerfeld radiation boundary conditions [9] . Furthermore, an extension of our methods for the solution of the Helmholtz equation with variable coefficients seems a challenging task for the future.

References

[1] A. Arico and M. Donatelli. A V-cycle multigrid for multilevel matrix algebras: proof of optimality and applications. *Numer. Math., to appear.*

[2] A. Arico, M. Donatelli, and S. Serra Capizzano. V-cycle optimal convergence for certain (multilevel) structured linear systems. *SIAM J. Mat. Anal. Appl.*, 26:186–214, 2004.

[3] A. Brandt and I. Livshits. Wave-ray multigrid method for standing wave equations. *ETNA*, 6:162–181, 1997.

[4] S. Serra Capizzano. Convergence analysis of two-grid methods for elliptic Toeplitz and PDEs matrix-sequences. *Numer. Math.*, 92:433–465, 2002.

[5] S. Serra Capizzano and C. Tablino-Possio. Multigrid methods for multilevel circulant matrices. *SIAM J. Sci. Comp.*, 26(1):55–85, 2005.

[6] R. Chan, S. Serra Capizzano, and C. Tablino-Possio. Two-grid methods for banded linear systems from DCT III algebra. *Numer. Linear Algebra Appl.*, 12(2–3):241–249, 2005.

[7] R. Chan, Q. Chang, and H. Sun. Multigrid methods for ill-conditioned Toeplitz systems. *SIAM J. Sci. Comp.*, 19:516–529, 1998.

[8] H. Elman, O. Ernst, and D. O'Leary. A multigrid method enhanced by krylov subspace iteration for discrete helmholtz equations. *SIAM J. Sci. Comp.*, 23:1290–1314, 2001.

[9] Y. Erlangga, C. Oosterlee, and C. Vuik. A novel multigrid based preconditioner for heterogeneous Helmholtz problems. *To appear in SIAM J. Sci. Comp.*

[10] G. Fiorentino and S. Serra. Multigrid methods for Toeplitz matrices. *Calcolo*, 28:283–305, 1992.

[11] G. Fiorentino and S. Serra. Multigrid methods for symmetric positive definite block Toeplitz matrices with nonnegative generating functions. *SIAM J. Sci. Comp.*, 17:1068–1081, 1996.

[12] R. Fischer and T. Huckle. Multigrid methods for anisotropic BTTB systems. *Submitted to Lin. Alg. Appl. (Special issue on the 80th birthday of F. L. Bauer).*

[13] R. Fischer and T. Huckle. Multigrid methods for strongly ill-conditioned structured matrices. *Submitted to ECOMAS Proceedings of the 8th European Multigrid Conference.*

[14] U. Grenander and G. Szegö. *Toeplitz forms and their applications.* Chelsea Publishing, New York, second edition, 1984.

[15] J. W. Ruge and K. Stüben. Algebraic multigrid. In *Frontiers in Applied Mathematics: Multigrid Methods, S. McCormick Ed.*, pages 73–130, Philadelphia, 1987. SIAM.

[16] E. Tyrtyshnikov. Circulant preconditioners with unbounded inverses. *Linear Algebra Appl.*, 216:1–23, 1995.