Electronic Transactions on Numerical Analysis.
Volume 16, pp. 1-29, 2003.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

PRECONDITIONING STRATEGIES FOR 2D FINITE DIFFERENCE MATRIX SEQUENCES *

STEFANO SERRA CAPIZZANO† AND CRISTINA TABLINO POSSIO‡

Abstract. In this paper we are concerned with the spectral analysis of the sequence of preconditioned matrices $\{P_n^{-1}A_n(a, m_1, m_2, k)\}_n$, where $n = (n_1, n_2)$, $N(n) = n_1 n_2$ and where $A_n(a, m_1, m_2, k) \in \mathbb{R}^{N(n)\times N(n)}$ is the symmetric two-level matrix coming from a high–order Finite Difference (FD) discretization of the problem

$$\begin{cases} (-1)^k \left(\dfrac{\partial^k}{\partial x^k} \left(a(x,y) \dfrac{\partial^k}{\partial x^k} u(x,y) \right) + \dfrac{\partial^k}{\partial y^k} \left(a(x,y) \dfrac{\partial^k}{\partial y^k} u(x,y) \right) \right) = f(x,y) & \text{on } \Omega = (0,1)^2, \\ \left(\dfrac{\partial^s}{\partial \nu^s} u(x,y) \right)_{|\partial\Omega} = 0 \quad s = 0, \ldots, k-1 & \text{on } \partial\Omega, \end{cases}$$

with ν denoting the unit outward normal direction and where m_1 and m_2 are parameters identifying the precision order of the used FD schemes. We assume that the coefficient $a(x, y)$ is nonnegative and that the set of the possible zeros can be represented by a finite collection of curves. The proposed preconditioning matrix sequences correspond to two different choices: the Toeplitz sequence $\{A_n(1, m_1, m_2, k)\}_n$ and a Toeplitz based sequence that adds to the Toeplitz structure the informative content given by the suitable scaled diagonal part of $A_n(a, m_1, m_2, k)$. The former case gives rise to optimal preconditioning sequences under the assumption of positivity and boundedness of a. With respect to the latter, the main result is the proof of the asymptotic clustering at unity of the eigenvalues of the preconditioned matrices, where the "strength" of the cluster depends on the order k, on the regularity features of $a(x, y)$ and on the presence of zeros of $a(x, y)$.

Key words. finite differences, Toeplitz and Vandermonde matrices, clustering and preconditioning, spectral distribution.

AMS subject classifications. 65F10, 65N22, 65F15.

1. Introduction.

The numerical solution of elliptic boundary value problems is of interest in several classical applications including elasticity problems and nuclear and petroleum engineering [41]. In particular, high order ($k > 1$) elliptic differential equations arise when modelling problems of plane elasticity ($k = 2$), when considering the vibration of a thin beam ($k = 3$), when dealing with shell analysis ($k = 4$), etc. [1]. In these contexts the weight function $a(x, y)$ can be continuous or discontinuous, but it is strictly positive and therefore the ellipticity of the continuous problem is guaranteed. Conversely, when dealing with problems arising in mathematical biology and mathematical finance, or when computing special functions, the strict ellipticity is sometimes lost and indeed the function $a(x, y)$ can have isolated zeros generally located at the boundary of the definition domain. Therefore, in the differential problem we simply assume that $a(x, y) \geq 0$, where the set of the zeros is represented by (at most) a finite collection of curves.

In preceding works we have considered these types of problems by focusing our attention on Finite Element methods [35] and the Finite Differences (FD) of minimal order of accuracy [27, 30] or on the one-dimensional case [33]. The resulting symmetric positive definite linear systems are solved by using Preconditioned Conjugate Gradient (PCG) algorithms, where the chosen preconditioners ensure the "optimality" of the method [3] and even a "clustering" [38] of the preconditioned spectra at unity [27].

This paper is addressed to high–order FD formulae for the approximation of the quoted two-dimensional differential problem. The motivation is given by the increased accuracy

*Received March 13, 2001. Accepted for publication September 4, 2002. Recommended by L. Reichel.
†Dipartimento di Chimica, Fisica e Matematica, Università dell'Insubria - Sede di Como, via Valleggio 11, 22100 Como, Italy. E-mail: `Stefano.Serrac@uninsubria.it`, `Serra@mail.dm.unipi.it`
‡Dipartimento di Matematica e Applicazioni, Università di Milano Bicocca, via Bicocca degli Arcimboldi, 8, 20126 Milano, Italy. E-mail: `Cristina.Tablinopossio@unimib.it`

ETNA
Kent State University
etna@mcs.kent.edu

when the related solution is regular enough. Even though the diminished sparsity of the resulting linear system has been considered a serious drawback for the practical application of these methods, here we propose different ways to overcome this difficulty. Both theoretical and practical comparisons prove that the new proposal is more effective than classical techniques such as matrix algebra preconditioning [7, 19, 20, 21] or incomplete LU factorization preconditioning [15, 2]. More precisely, we study some Toeplitz and Toeplitz based preconditioners for matrices $A_n(a, m_1, m_2, k)$ coming from a large class of high–order FD discretizations of differential problems of the form

$$\left\{\begin{array}{l} (-1)^k\left(\dfrac{\partial^k}{\partial x^k}\left(a(x,y)\dfrac{\partial^k}{\partial x^k}u(x,y)\right)+\dfrac{\partial^k}{\partial y^k}\left(a(x,y)\dfrac{\partial^k}{\partial y^k}u(x,y)\right)\right)=f(x,y) \text{ on } \Omega, \\ \left(\dfrac{\partial^s}{\partial \nu^s}u(x,y)\right)_{|\partial\Omega} = 0 \quad s=0,\dots,k-1 \text{ on } \partial\Omega, \end{array}\right. \tag{1.1}$$

where the parameter m_1 identifies the precision order of the FD scheme used for approximating the operator $\partial^k/\partial x^k$ and m_2 identifies the precision order of the FD scheme used for approximating the operator $\partial^k/\partial y^k$.

The Toeplitz based preconditioner is devised as

$$P_n(a) = D_n^{1/2}(a)A_n(1, m_1, m_2, k)D_n^{1/2}(a),$$

where $D_n(a)$ is a suitably scaled diagonal part of $A_n(a, m_1, m_2, k)$ and $A_n(1, m_1, m_2, k)$ is the symmetric positive definite (SPD) Toeplitz matrix obtained when $a(x, y) \equiv 1$. In this paper, we first derive asymptotic expansions concerning the preconditioned matrices $P_n^{-1}(a)A_n(a, m_1, m_2, k)$ in terms of related Toeplitz structure. Then we analyze in detail the sequence of matrices $\{P_n^{-1}(a)A_n(a, m_1, m_2, k)\}_n$ in order to prove the clustering at unity of the related spectra and to obtain an asymptotic estimate of the number of outliers. These results are used to understand the asymptotic behavior of the PCG techniques when the Toeplitz based preconditioners are applied. In fact, by using the Axelsson and Lindskog Theorems [3], we deduce a somehow accurate upperbound on the number of PCG iterations required in order to reach the solution within a preassigned accuracy η. In many cases the number of iterations is bounded by a constant independent of the size of the involved matrices. In this way, the solution of a system with coefficient matrix given by $A_n(a, m_1, m_2, k)$ is reduced to the solution of a few linear systems of diagonal type and of two-level band-Toeplitz type. The existence of numerical procedures ad hoc for the computation of the solution of two-level band-Toeplitz linear systems (see [12] and [26, 24]) makes the proposed preconditioning techniques very attractive in the context of differential boundary value problems.

The paper is organized as follows. In Section **2** we give preliminary results concerning the FD matrices $A_n(a, m_1, m_2, k)$. Section **3** is devoted to defining and analyzing the Toeplitz preconditioner and the spectral properties of the related preconditioned sequence of matrices. In section **4** we address the clustering analysis of the preconditioned matrix sequence $\{P_n(a)\}_n$ and the derivation of some estimates regarding the number of outliers. In Section **5** we study the spectral distribution of the latter preconditioned matrix sequences and we deal with the irregular case in which $a(x, y)$ is assumed just $L^\infty(\Omega)$. Section **6** is devoted to the study of some numerical experiments concerning the PCG method for $A_n(a, m_1, m_2, k)$ and a specialized multigrid technique for $A_n(1, m_1, m_2, k)$ (see [31]) and to a comparison with the previous literature. Some concluding remarks in Section **7** end the paper.

2. The discretized 2D problem. In this section we analyze the main structural and spectral properties of the two-level band matrices associated with high–order FD formulae. The main result is the dyadic representation theorem that allows one to give a spectral characterization of high–order FD matrix sequences.

ETNA
Kent State University
etna@mcs.kent.edu

2.1. High–order FD matrices. Let us consider a 2D elliptic problem of the form (1.1). The FD discretization is performed over a sequence of equispaced 2D grids $\mathcal{U}_1 \times \mathcal{U}_2$, $\mathcal{U}_i = \{\mathcal{U}_{i,n_i}\}$, where $\mathcal{U}_{i,n_i}$, $i = 1, 2$ are $(n_i + 2)$-dimensional grids on $[0, 1]$. More precisely,

$$\begin{array}{lll} \mathcal{U}_{1,n_1} = \{x_r = rh_1 : & h_1 = (n_1+1)^{-1}, & r = 0, \ldots n_1 + 1\}, \\ \mathcal{U}_{2,n_2} = \{y_t = th_2 : & h_2 = (n_2+1)^{-1}, & t = 0, \ldots n_2 + 1\}. \end{array}$$

In a previous paper [32], we highlighted some general features of high–order FD formulae for the discretization of the differential operator d^k/dx^k by using q $(q \geq k + 1)$ equispaced mesh points. This can be easily generalized to the multivariate case through tensorial arguments. Therefore, we briefly give the essential notation and the key properties necessary to define and to analyze the arising FD matrices. As template we consider the case of the discretization of $(\partial^k u(x, y)/\partial x^k)_{|(x,y)=(x_r,y_t)}$. We assume that this discretization formula involves $m = \lfloor q/2 \rfloor$ mesh points less than x_r, $m = \lfloor q/2 \rfloor$ greater than x_r, plus the point x_r if q is odd. More precisely, if $q = 2m + 1$ the mesh points are defined as $x_j = x_r + jh$, $j = -m, \ldots, m$, while if $q = 2m$ as $x_j = x_r + (j - 1/2)h$, $j = 1, \ldots, m$ and $x_j = x_r + (j + 1/2)h$, $j = -m, \ldots, -1$.

Let $\mathbf{c} \in \mathbb{R}^q$ be the coefficient vector defining an FD formula that has an order of accuracy ν under the assumption of a proper regularity of the function $u(x, y)$, namely

$$\frac{\partial^k}{\partial x^k} u(x, y)_{|(x,y)=(x_r,y_t)} = h^{-k} \sum_j c_j u(x_j, y_t) + O(h^\nu).$$

Such a coefficient vector can be obtained as the solution of a Vandermonde-like linear system [32]. Moreover, as in the univariate case (we refer to Lemma 2.2 in [32]) the maximal order FD formula with respect to q mesh points exhibits some specific features due to the structural properties of Vandermonde matrices.

LEMMA 2.1. *Let $\mathbf{c} \in \mathbb{R}^q$ be the coefficient vector related to a maximal order FD formula discretizing $\partial^k/\partial z^k$, $k \geq 1$ by using q $(q \geq k+1)$ equispaced mesh points. Then $\mathbf{c}$ is unique and its entries are rational, $\mathbf{c}$ is symmetric i.e. $c_j = c_{q+1-j}$ for every j or antisymmetric i.e. $c_j = c_{q+1-j}$ for every j according to whether the quantity $k (\mathrm{mod}\ 2)$ equals 0 or 1 and, finally, the order of accuracy ν equals $q - k + 1$ if $k + q$ is odd and equals $q - k$ if $k + q$ is even.*

Now, in order to deal with symmetric FD matrices we leave the operator in "divergence form" and we discretize the inner and the outer partial derivatives separately. For the sake of computational convenience, here we limit ourselves to the case where both the inner and the outer operator are discretized by means of FD formulae of maximal order of accuracy. On the other hand, we may discretize the operators $\partial^k/\partial x^k$ and $\partial^k/\partial y^k$ by means of two different FD formulae. The reason for such a choice can be found in different regularity properties of the solution with respect to the space variables x and y.

Since the FD discretization of the quoted 2D problem is a trivial generalization of the 1D case, we refer to [32] for any detail concerning the discretization process. Hereafter, we simply report the final expressions of the resulting FD matrix sequence.

DEFINITION 2.2. *Let $h_1 = (n_1 + 1)^{-1}$ and $h_2 = (n_2 + 1)^{-1}$ be the discretization stepsizes with respect to the x and y space variables respectively. The symbol $A_n(a, m_1, m_2, k) \in \mathbb{R}^{N(n)\times N(n)}$, $N(n) = n_1 n_2$ and $n = (n_1, n_2)$, denotes the n-th symmetric two-level matrix discretizing the problem (1.1) through the FD formula of maximal order of accuracy ν_1 related to the coefficient vector $\mathbf{c} \in \mathbb{R}^{q_1}$ $(m_1 = \lfloor q_1/2 \rfloor)$ for both the inner and the outer partial derivative with respect to x and the FD formula of maximal order of accuracy ν_2 related to the coefficient vector $\mathbf{d} \in \mathbb{R}^{q_2}$ $(m_2 = \lfloor q_2/2 \rfloor)$ for both the inner and the outer partial* *derivative with respect to y. Due to the comparison between the computational cost and the order of accuracy, we always consider $q_i,\ i=1,2$ odd when k is even and vice versa. When $m_1=m_2=m$ we write in short $A_n(a,m,k)$.*

It is worthwhile stressing that the considered high–order FD formulae work with some extra points not belonging to Ω. So, for mathematical consistency, we need to define the coefficient $a(x,y)$ over the set $\Omega^*=(-\varepsilon_1,1+\varepsilon_1)\times(-\varepsilon_2,1+\varepsilon_2)$, where $\varepsilon_i,\ i=1,2$ are some positive quantities. Therefore, when we write $a(x,y)\in C^s(\overline{\Omega})$ it is understood that $a(x,y)$ is simply defined in Ω^*, while the regularity is required in $\overline{\Omega}$. The only needed assumption is that $\min_{(x,y)\in\overline{\Omega}} a(x,y)\le a(x,y)\le \max_{(x,y)\in\overline{\Omega}} a(x,y)$ holds for any $(x,y)\in\Omega^*$. More precisely, for any fixed n_1 and n_2, when considering the x-derivatives, the function $a(x,y)$ is sampled on the 2D grid $\tilde{\mathcal{U}}_{1,n_1}\times(\mathcal{U}_{2,n_2}\backslash\{0,1\})$, while, when considering the y-derivatives, $a(x,y)$ is sampled on the 2D grid $(\mathcal{U}_{1,n_1}\backslash\{0,1\})\times\tilde{\mathcal{U}}_{2,n_2}$, where

$$\tilde{\mathcal{U}}_{1,n_1}=\left\{\tilde{x}_i=\left\{\begin{array}{ll} ih_1,\ i=1-m_1,n_1+m_1 & \text{if } q_1=2m_1+1\\ (i+1/2)h_1,\ i=1-m_1,n_1+m_1-1 & \text{if } q_1=2m_1\end{array}\right.\right\}, \tag{2.1}$$

and

$$\tilde{\mathcal{U}}_{2,n_2}=\left\{\tilde{y}_j=\left\{\begin{array}{ll} jh_2,\ j=1-m_2,n_2+m_2 & \text{if } q_2=2m_2+1\\ (j+1/2)h_2,\ j=1-m_2,n_2+m_2-1 & \text{if } q_2=2m_2\end{array}\right.\right\}. \tag{2.2}$$

Finally, let us denote by x_{r+s} the quantity x_r+sh_1 and by y_{t+s} the quantity y_t+sh_2.
By virtue of the symmetric property only the lower triangular entries are reported in the following.
Case k odd ($q_1=2m_1,\ q_2=2m_2$): As a consequence of Lemma 2.1, we are dealing with two antisymmetric coefficient vectors $\mathbf{c}=(-c_{m_1},\dots,-c_1,c_1,\dots,c_{m_1})\in\mathbb{R}^{q_1}$ and $\mathbf{d}=(-d_{m_2},\dots,-d_1,d_1,\dots,d_{m_2})\in\mathbb{R}^{q_2}$. So according to Definition 2.2 we have

$$\begin{aligned}
(A_n)_{s,s} &= \frac{1}{h_1^{2k}}\left[\sum_{j=1}^{m_1}\left(a\left(x_{r-j+\frac12},y_t\right)+a\left(x_{r+j-\frac12},y_t\right)\right)c_j^2\right]\\
&\quad+\frac{1}{h_2^{2k}}\left[\sum_{j=1}^{m_2}\left(a\left(x_r,y_{t-j+\frac12}\right)+a\left(x_r,t_{t+j-\frac12}\right)\right)d_j^2\right],\\
(A_n)_{s,s-p} &= \frac{1}{h_1^{2k}}\left[\sum_{j=1}^{m_1-p}\left(a\left(x_{r-p-j+\frac12},y_t\right)+a\left(x_{r+j-\frac12},y_t\right)\right)c_jc_{j+p}\right.\\
&\quad\left.-\sum_{j=1}^{p}a\left(x_{r-j+\frac12},y_t\right)c_jc_{p+1-j}\right]\ \text{if } p=1,\dots,m_1-1,\\
(A_n)_{s,s-p} &= -\frac{1}{h_1^{2k}}\left[\sum_{j=p+1-m_1}^{m_1}a\left(x_{r-j+\frac12},y_t\right)c_jc_{p+1-j}\right]\ \text{if } p=m_1,\dots,2m_1-1,\\
(A_n)_{s,s-n_1p} &= \frac{1}{h_2^{2k}}\left[\sum_{j=1}^{m_2-p}\left(a\left(x_r,y_{t-p-j+\frac12}\right)+a\left(x_r,y_{t+j-\frac12}\right)\right)d_jd_{j+p}\right.\\
&\quad\left.-\sum_{j=1}^{p}a\left(x_r,y_{t-j+\frac12}\right)d_jd_{p+1-j}\right]\ \text{if } p=1,\dots,m_2-1,
\end{aligned}$$

$$(A_n)_{s,s-n_1p} = -\frac{1}{h_2^{2k}}\left[\sum_{j=p+1-m_2}^{m_2} a\left(x_r, y_{t-j+\frac{1}{2}}\right) d_j d_{p+1-j}\right] \text{ if } p = m_2, \dots, 2m_2 - 1.$$

Case k even ($q_1 = 2m_1+1$, $q_2 = 2m_2+1$): As a consequence of Lemma 2.1, we are dealing with two symmetric coefficient vectors $\mathbf{c} = (c_{m_1}, \dots, c_1, c_0\ c_1, \dots, c_{m_1})$ and $\mathbf{d} = (d_{m_2}, \dots, d_1, d_0\ d_1, \dots, d_{m_2})$. So according to Definition 2.2 we have

$$\begin{aligned}
(A_n)_{s,s} &= \frac{1}{h_1^{2k}}\left[a(x_r, y_t)c_0^2 + \sum_{j=1}^{m_1}(a(x_{r-j}, y_t) + a(x_{r+j}, y_t))c_j^2\right] \\
&\quad + \frac{1}{h_2^{2k}}\left[a(x_r, y_t)d_0^2 + \sum_{j=1}^{m_2}(a(x_r, y_{t-j}) + a(x_r, t_{t+j}))d_j^2\right], \\
(A_n)_{s,s-p} &= \frac{1}{h_1^{2k}}\left[\sum_{j=1}^{m_1-p}(a(x_{r-p-j}, y_t) + a(x_{r+j}, y_t))c_j c_{p+j} + \sum_{j=0}^{p} a(x_{r-j}, y_t)c_j c_{p-j}\right] \\
&\quad \text{if } p = 1, \dots, m_1 - 1, \\
(A_n)_{s,s-p} &= \frac{1}{h_1^{2k}}\left[\sum_{j=p-m_1}^{m_1} a(x_{r-j}, y_t)c_j c_{p-j}\right] \text{ if } p = m_1, \dots, 2m_1, \\
(A_n)_{s,s-n_1p} &= \frac{1}{h_2^{2k}}\left[\sum_{j=1}^{m_2-p}(a(x_r, y_{t-p-j}) + a(x_r, y_{t+j}))d_j d_{p+j} + \sum_{j=0}^{p} a(x_r, y_{t-j})d_j d_{p-j}\right] \\
&\quad \text{if } p = 1, \dots, m_2 - 1, \\
(A_n)_{s,s-n_1p} &= \frac{1}{h_2^{2k}}\left[\sum_{j=p-m_2}^{m_2} a(x_r, y_{t-j})d_j d_{p-j}\right] \text{ if } p = m_2, \dots, 2m_2.
\end{aligned}$$

These defining relations have been used in Appendix **A** in [34] in the evaluation of the asymptotic expansion of the matrices $A_n(a, m_1, m_2, k)$. We recall that these asymptotic expansions are essential for the spectral analysis of the second type of proposed preconditioned matrix sequences.

2.2. The dyadic representation theorem. By using the dyadic representation theorem in the 1D case (Theorem 3.5 of [32]) and the Kronecker structure of the matrices $A_n(a, m_1, m_2, k)$, the following representation theorem clearly follows.

THEOREM 2.3. *Let $A_n(a, m_1, m_2, k)$ be the FD matrix according to Definition 2.2. The following dyadic representation holds true*

$$\begin{aligned}
A_n(a, m_1, m_2, k) &= \frac{1}{h_1^{2k}}\sum_{i,t} a(\tilde{x}_i, y_t)(\mathbf{e}_t \otimes \mathbf{c}[i])(\mathbf{e}_t \otimes \mathbf{c}[i])^T \\
&\quad + \frac{1}{h_2^{2k}}\sum_{r,j} a(x_r, \tilde{y}_j)(\mathbf{d}[j] \otimes \mathbf{e}_r)(\mathbf{d}[j] \otimes \mathbf{e}_r)^T,
\end{aligned}$$

with $(\tilde{x}_i, y_t)$ ranging in $\tilde{\mathcal{U}}_{1,n_1} \times (\mathcal{U}_{2,n_2} \backslash \{0, 1\})$ and $(x_r, \tilde{y}_j)$ ranging in $(\mathcal{U}_{1,n_1} \backslash \{0, 1\}) \times \tilde{\mathcal{U}}_{2,n_2}$ according to Eqs. (2.1) and (2.2). Here, $\mathbf{e}_t$ denotes the t^{th} canonical vector of $\mathbb{R}^{n_2}$, $\mathbf{e}_r$ denotes the r^{th} canonical vector of $\mathbb{R}^{n_1}$. The vector $\mathbf{c}[i] \in \mathbb{R}^{n_1}$ equals $[0, \dots, 0, \mathbf{c}, 0, \dots, 0]^T$, where $\mathbf{c} \in \mathbb{R}^{q_1}$ is the FD formula coefficient vector whose first entry is at position $i - \lceil q_1/2 \rceil +$

ETNA
Kent State University
etna@mcs.kent.edu

1, *according to the univariate case. The vector* $\mathbf{d}[j] \in \mathbb{R}^{n_2}$ *equals* $[0,\ldots,0,\mathbf{d},0,\ldots,0]^T$, *where* $\mathbf{d} \in \mathbb{R}^{q_2}$ *is the FD formula coefficient vector whose first entry is at position* $j - \lceil q_2/2 \rceil + 1$. *If the starting location (or the final location) of the vector* $\mathbf{c}$ *lies outside the vector* $\mathbf{c}[i]$, *then the outgoing entries are simply neglected (see [32] for more details). The same holds for the vector* $\mathbf{d}[j]$.

Theorem 2.3 can be used to provide a link between the zeros of the nonnegative function $a(x,y)$ and the rank of each matrix $A_n(a,m_1,m_2,k)$.

THEOREM 2.4. *Let* $A_n(a,m_1,m_2,k)$ *be the FD matrix according to Definition 2.2 and let* $a(x,y)$ *be a nonnegative function. For any* t *such that* $y_t \in (\mathcal{U}_{2,n_2}\backslash\{0,1\})$ *let* $I^+(a,1)[t] = \{i : a(\tilde{x}_i, y_t) > 0\}$ *and, for any* r *such that* $x_r \in (\mathcal{U}_{1,n_1}\backslash\{0,1\})$ *let* $I^+(a,2)[r] = \{j : a(x_r, \tilde{y}_j) > 0\}$. *Suppose that vectors* $\{\mathbf{c}[i] : i = 1,\ldots,n_1+q_1-1\}$ *strongly generate* $\mathbb{R}^{n_1}$ *in the sense that each subset* $\{\mathbf{c}[i_k] : 1 \le i_1 < i_2 < \cdots < i_{n_1} \le n_1+q_1-1\}$ *is a basis for* $\mathbb{R}^{n_1}$. *Analogously, suppose that the vectors* $\{\mathbf{d}[j] : j = 1,\ldots,n_2+q_2-1\}$ *strongly generate* $\mathbb{R}^{n_2}$. *Then*

$$\mathrm{rank}(A_n(a,m_1,m_2,k)) \ge \max\left\{\sum_{t=1}^{n_2}\min\{n_1,\#I^+(a,1)[t]\},\sum_{r=1}^{n_1}\min\{n_2,\#I^+(a,2)[r]\}\right\}$$

Proof. It is a straightforward consequence of Theorem 2.3. □

REMARK 2.5. *As a consequence of Theorem 2.4 and under its assumptions, the matrices* $A_n(a,m_1,m_2,k)$ *are positive definite if* $a(x,y)$ *is positive. In the case where* $a(x,y)$ *has zeros, we have positive definiteness if the set of the zeros is given by a finite number of curves with a finite number of intersections with horizontal and vertical lines (this finite number is required to be less than some universal constant that can be explicitly calculated using the inequalities given in Theorem 2.4). On the other hand, if there exists an open set where* $a(x,y)$ *vanishes, then the matrices* $A_n(a,m_1,m_2,k)$ *are singular for any* n_1, n_2 *large enough.*

3. The Toeplitz preconditioning sequence. When $a(x,y) \equiv 1$, the matrices of $\{A_n(a,m_1,m_2,k)\}_n$ enjoy a two-level Toeplitz structure. Here, we are interested in applying the results on the preconditioning through the concept of equivalent functions, where, for instance, the constant function $a(x,y) \equiv 1$ is equivalent to any strictly positive and bounded function $a(x,y)$. We indicate the equivalence relation by the symbol $\sim$ with the natural meaning that $g_1 \sim g_2$ if and only if the two functions (or sequences) have the same definition domain $\mathcal{D}$ and there exist two positive constants c and C such that $cg_1(\underline{x}) \le g_2(\underline{x}) \le Cg_1(\underline{x})$ for any $\underline{x} \in \mathcal{D}$.

As a consequence of Theorems 2.3, 2.4 and of the relation $a(x,y) \equiv 1$, we deduce the following optimality result, where, in the context of the preconditioning, a sequence $\{P_n\}_n$ of positive definite matrices is an optimal preconditioning sequence for the sequence $\{A_n\}_n$ if and only if there exists an $\bar{n}$ such that for any $n \ge \bar{n}$ all the eigenvalues of $P_n^{-1}A_n$ belong to a positive bounded universal interval independent of n [4, 3].

THEOREM 3.1. *If* $a(x,y)$ *is strictly positive then the Toeplitz sequence* $\{A_n(1,m_1,m_2,k)\}_n$ *is an optimal sequence of preconditioners for the matrix sequence* $\{A_n(a,m_1,m_2,k)\}_n$ *and* $k_2(A_n(a,m_1,m_2,k)) \sim k_2(A_n(1,m_1,m_2,k))$, *where* k_2 *denotes the spectral condition number. If* $a(x,y)$ *is nonnegative and the matrices* $A_n(a,m_1,m_2,k)$ *are positive definite then, for* n *large enough,* $k_2(A_n(a,m_1,m_2,k)) \ge Ck_2(A_n(1,m_1,m_2,k))$ *with* C *a universal constant independent of* n.

Proof. We refer to [32] for an analogous result. □

The following claim now takes into account the presence of zeros in the function $a(x,y)$ and explains why the matrix sequence $\{A_n(1,m_1,m_2,k)\}_n$ is not an optimal preconditioning sequence. Clearly, if $a(x, y)$ is strictly positive then the matrices $A_n(a, m_1, m_2, k)$ are positive definite, otherwise we refer to Remark 2.5.

THEOREM 3.2. *If the coefficient $a(x, y) \geq 0$ belongs to $\mathbf{C}(\overline{\Omega})$, with $\min a(x, y) = 0$ and the matrices $A_n(a, m_1, m_2, k)$ are positive definite, then the spectra of the preconditioned matrices*

$$\{A_n^{-1}(1, m_1, m_2, k)A_n(a, m_1, m_2, k)\}_n$$

belong to the interval $(0, C]$, C being the maximum of $a(x, y)$ on $\overline{\Omega}$; otherwise the spectra belong to the interval $[0, C]$. Moreover, in the first case the lower bound is sharp in the sense that the smallest eigenvalue of the preconditioned matrix tends to zero as n tends to infinity.

Proof. First, for the sake of simplicity, we assume that the function $a(x, y)$ has a unique zero located at $(x, y) = (0, 0)$. ¿From Theorem 2.3, it follows that for any $\mathbf{x} \in \mathbb{R}^n$

$$\mathbf{x}^T A_n(a, m_1, m_2, k)\mathbf{x} \leq C\mathbf{x}^T A_n(1, m_1, m_2, k)\mathbf{x}, \quad C = \max_{\overline{\Omega}} a(x, y),$$

so that, by virtue of the positive definiteness of the matrices $A_n(1, m_1, m_2, k)$, we infer that

$$\lambda_{\max}(A_n^{-1}(1, m_1, m_2, k)A_n(a, m_1, m_2, k)) \leq C.$$

Now, the limit relation concerning the smallest eigenvalue can be easily proved by considering the Rayleigh quotient for $\mathbf{x}$ equal to the first vector $\mathbf{e}_1$ of the canonical basis of $\mathbb{R}^n$. In fact, we have

$$\lambda_{\min}(A_n^{-1}(1, m_1, m_2, k)A_n(a, m_1, m_2, k)) \leq \frac{\mathbf{e}_1^T A_n(a, m_1, m_2, k)\mathbf{e}_1}{\mathbf{e}_1^T A_n(1, m_1, m_2, k)\mathbf{e}_1},$$

which is $O(\omega_a(h))$, $h = \max\{(n_1 + 1)^{-1}, (n_2 + 1)^{-1}\}$ and so tends to zero as n tends to infinity. More can be said if we know further information on the behavior of $a(x, y)$ in a neighbourhood J of $(0, 0)$. If $a(x, y) \sim \|(x, y)\|_2^t$ for some $t > 0$ in J, then

$$\lambda_{\min}(A_n^{-1}(1, m_1, m_2, k)A_n(a, m_1, m_2, k)) = O(h^t).$$

Clearly, the vector $\mathbf{e}_1$ is chosen according to the assumption that the zero is located at $(x, y) = (0, 0)$; otherwise it is enough to consider a suitable vector within the canonical basis of $\mathbb{R}^n$. In the case of more than one zero, but under the assumption of positive definiteness of the matrices $A_n(a, m_1, m_2, k)$, the proof is unchanged. □

Now, the proposed preconditioning technique is of practical interest if the solution of a linear system whose matrix is given by $A_n(1, m_1, m_2, k)$ can be efficiently computed. We remark that $\{A_n(1, m_1, m_2, k)\}_n$ is a sequence of two-level band Toeplitz matrices with asymptotic ill-conditioning, due to the zero of order $2k$ of the generating function at $(0, 0)$. Under the assumption that $n_1 \sim n_2$, the classical band solvers (based on Gaussian elimination) require $O((N(n))^2)$ arithmetic operations (ops) [15]. The methods based on the "correction" in two-level algebras as circulants, τ, etc. cost $O((N(n))^{3/2})$ ops [13]. Finally, in some cases it has been proved that the cost of the multigrid approach is $O(N(n))$ ops [12]. In particular, the optimality of the (V-cycle) multigrid method has been proved for $k \leq 2$ [17], while the optimality of the two-grid method has been proved recently for any k [31].

Finally, we want to highlight some specific link between the matrix sequence $\{A_n(1, m_1, m_2, k)\}_n$ and the Toeplitz matrix sequences generated by 2π–periodic integrable functions. We recall that the symbol $T_q(f)$, with $f \in L^1((-\pi, \pi], \mathbb{C})$, denotes the unilevel

Toeplitz matrix (generated by f) whose coefficient along the s^{th} diagonal is given by the s^{th} Fourier coefficient of f

$$a_s = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x)e^{-\mathbf{i}sx}dx, \quad \mathbf{i}^2=-1, \ \ |s|\le q-1\in\mathbb{N}.$$

The definition of two-level (and multilevel) Toeplitz matrix is straightforward through the Fourier coefficient of a bivariate (multivariate) function and can be found in [38].
As a preliminary step we need some results related to the univariate version of problem (1.1). Let us denote by $\hat{\Delta}_n(m,k)$ the scaled unilevel Toeplitz matrices obtained in the discretization of the univariate problem by considering the maximal order FD formula with respect to q mesh points for both the inner and the outer derivative.

THEOREM 3.3. [32] *The matrices $\{\hat{\Delta}_n(m,k)\}_n$ are Toeplitz matrices generated by the nonnegative real–valued polynomial $p_{\mathbf{w}}(x) = |p_{\mathbf{c}}(x)|^2, x\in(-\pi,\pi]$, where $p_{\mathbf{c}}$ is the polynomial related to the maximal order FD formula coefficient vector $\mathbf{c}$ for the discretization of the operator d^k/dx^k over q equispaced mesh points ($q=2m$ if k is odd, $q=2m+1$ if k is even), defined as*

$$p_{\mathbf{c}}(x) = \begin{cases} \sum_{j=-m}^{m} c_j e^{\mathbf{i}jx} & \text{if } q=2m+1,\\ \sum_{j=-m}^{-1} c_j e^{\mathbf{i}jx} + \sum_{j=1}^{m} c_j e^{\mathbf{i}(j-1)x} & \text{if } q=2m. \end{cases}$$

Therefore, for any n the matrices $\hat{\Delta}_n(m,k)$ are symmetric positive definite and their spectral condition number is asymptotically greater than Cn^{2k}, with C universal constant. Finally, if the zeros of $|p_{\mathbf{c}}(x)|^2$ not located at $x=0$ are of order at most $2k$, then the spectral condition number of $\{\hat{\Delta}_n(m,k)\}_n$ is asymptotic to n^{2k}.

This statement can be generalized to the bivariate case, by making use of the following notion of *sparsely vanishing* matrix sequence.

DEFINITION 3.4. *A sequence of matrices $\{X_n\}_n$, with $X_n\in\mathbb{C}^{d_n\times d_n}$ and $d_n<d_{n+1}$, is said to be* sparsely vanishing *if there exists a nonnegative function $x(s)$ with $\lim_{s\to 0} x(s)=0$ so that for any $\varepsilon>0$ there exists $n_\varepsilon\in\mathbb{N}$ such that for any $n\ge n_\varepsilon$*

$$\frac{1}{d_n}\#\{i : \sigma_i(X_n)\le\varepsilon\}\le x(\varepsilon),$$

where $\{\sigma_i(X_n)\},\ i=1,\dots,d_n$ denotes the complete set of the singular values of X_n.
It is understood that if the matrices are all Hermitian the definition holds with a more special characterization since $\sigma_i(X_n)=|\lambda_i(X_n)|$, where $\{\lambda_i(X_n)\}$ denotes the complete set of the eigenvalues of X_n.

The quoted definition makes direct reference to the notion, first introduced by Tyrtyshnikov [37], of *sparsely vanishing* Lebesgue-measurable functions as those functions whose set of zeros has zero Lebesgue measure [10]. In fact, a matrix sequence $\{X_n\}_n$ spectrally distributed as a *sparsely vanishing* function is *sparsely vanishing* in the sense of Definition 3.4 (we refer to Proposition **B.1** in [36]).

DEFINITION 3.5. *The symbol $\Delta_n(m_1,m_2,k)\in\mathbb{R}^{N(n)\times N(n)}$ denotes the FD matrix obtained according to Definition 2.2 when $a(x,y)\equiv 1$ and by setting $h_i=(n_i+1)^{-1}=h/\alpha_i$, $i=1,2$ with α_i absolute positive integer constants.*

THEOREM 3.6. *Let $\Delta_n(m_1,m_2,k)$ be the FD matrix according to Definition 3.5. Then*

$$\Delta_n(m_1,m_2,k) = h^{-2k}T_n\left(\alpha_1^{2k}|p_{\mathbf{c}}(x)|^2+\alpha_2^{2k}|p_{\mathbf{d}}(y)|^2\right),$$

where $T_n\left(\alpha_1^{2k}|p_{\mathbf{c}}(x)|^2+\alpha_2^{2k}|p_{\mathbf{d}}(y)|^2\right)$, $n=(n_1,n_2)$, is the two-level band Toeplitz matrix generated by the nonnegative real-valued polynomial $\alpha_1^{2k}|p_{\mathbf{c}}(x)|^2+\alpha_2^{2k}|p_{\mathbf{d}}(y)|^2$ and where

ETNA
Kent State University
etna@mcs.kent.edu

$|p_{\mathbf{c}}(x)|^2$ and $|p_{\mathbf{d}}(y)|^2$ are the polynomials defined as in Theorem 3.3. Therefore, the matrices $\Delta_n(m_1, m_2, k)$ are symmetric positive definite for any value of the dimension $N(n)$ and their spectral condition number is asymptotically greater than $C[N(n)]^k$, with C universal constant. In addition, if the zeros of $|p_{\mathbf{c}}(x)|^2$ and $|p_{\mathbf{d}}(y)|^2$ not located at $x = 0$ nor at $y = 0$ are of order at most $2k$, then the spectral condition number is asymptotic to $[N(n)]^k$. Finally, the matrix sequence $\{\Delta_n(m_1, m_2, k)\}_n$ is sparsely vanishing.

Proof. When $a(x, y) \equiv 1$, by virtue of the dyadic representation Theorem 2.3, we have that

$$\Delta_n(m_1, m_2, k) = h_1^{-2k} \sum_{i,t} (\mathbf{e}_t \mathbf{e}_t^T) \otimes (\mathbf{c}[i]\mathbf{c}^T[i]) + h_2^{-2k} \sum_{r,j} (\mathbf{d}[j]\mathbf{d}^T[j]) \otimes (\mathbf{e}_r \mathbf{e}_r^T),$$

and according to the properties of the Kronecker product we find that

$$\Delta_n(m_1, m_2, k) = h^{-2k} \left(I_{n_2} \otimes T_{n_1} \left(\alpha_1^{2k} |p_{\mathbf{c}}(x)|^2 \right) + T_{n_2} \left(\alpha_2^{2k} |p_{\mathbf{d}}(y)|^2 \right) \otimes I_{n_1} \right),$$

the last being a two-level Toeplitz matrix. A direct comparison with the entries defining the two-level Toeplitz matrix generated by the nonnegative real-valued polynomial $\alpha_1^{2k}|p_{\mathbf{c}}(x)|^2 + \alpha_2^{2k}|p_{\mathbf{d}}(y)|^2$ proves the claim. Now, at $(x, y) = (0, 0)$ the generating bivariate polynomial shows a zero of order $2k$ due to the consistency condition [32]. Since $n_1 \sim n_2$, by invoking the results of [6, 28], it follows that the spectral condition number of the matrices $\Delta_n(m_1, m_2, k)$ is asymptotically greater than $C[N(n)]^k$ for some positive C. With the assumption that the other zeros (if any) of $|p_{\mathbf{c}}(x)|^2$ and $|p_{\mathbf{d}}(y)|^2$ are of order at most $2k$, then the spectral condition number of $\{\Delta_n(m_1, m_2, k)\}_n$ is asymptotic to $[N(n)]^k$, again as a consequence of the analysis given in [6, 28]. Finally, the polynomial $p(x, y) = \alpha_1^{2k}|p_{\mathbf{c}}(x)|^2 + \alpha_2^{2k}|p_{\mathbf{d}}(y)|^2$ is sparsely vanishing since it is not identically zero and consequently the Lebesgue measure of its zeros is zero [10]. Moreover, $\{T_n(p)\}_n$ distributes as p with regard to the eigenvalues by Tyrtyshnikov's theorem [38]. Therefore, by Proposition B.1 of [36], since the matrix sequence distributes as a sparsely vanishing function, we infer that $\{T_n(p)\}_n$ is sparsely vanishing in the matrix sense given in Definition 3.4. □

4. The Toeplitz based preconditioning sequences. The second proposed preconditioning matrix sequence is devised in order to introduce a special improvement with respect to the previous Toeplitz preconditioning sequence, also giving effective results in the case where $a(x, y)$ shows a finite number of zeros. These preconditioning sequences can be constructed by coupling the previously considered Toeplitz sequence with the suitable scaled main diagonal of the matrices $\{A_n(a, m_1, m_2, k)\}_n$, the aim being to introduce more informative content from the original linear system into the preconditioner, while keeping the additional computational cost as low as possible. It is clear that the cost of solving a linear system with $P_n(a, m_1, m_2, k)$ as coefficient matrix is substantially the same as the one of solving a Toeplitz system. The additional cost is in fact just the one of multiplying a constant number of diagonal matrices by a vector.

DEFINITION 4.1. *Assume $h_i = h/\alpha_i$, $i = 1, 2$ with α_i positive integer constants. Let $\hat{A}_n(a, m_1, m_2, k) = h^{2k} A_n(a, m_1, m_2, k)$ with $A_n(a, m_1, m_2, k) \in \mathbb{R}^{N(n)\times N(n)}$ being the symmetric two-level matrix given in accordance with Definition 2.2. Moreover let $\hat{\Delta}_n(m_1, m_2, k) = h^{2k} \Delta_n(m_1, m_2, k)$ with $\Delta_n(m_1, m_2, k) \in \mathbb{R}^{N(n)\times N(n)}$ being the symmetric two-level Toeplitz matrix given in Definition 3.5. Finally, we define $\{\hat{P}_n(a)\}_n$, $\hat{P}_n(a) = \hat{D}_n^{1/2}(a) \hat{\Delta}_n(m_1, m_2, k) \hat{D}_n^{1/2}(a)$ the Toeplitz based preconditioning matrix sequence where $\hat{D}_n(a) = \mathrm{diag}(\hat{A}_n(a, m_1, m_2, k))/\Delta$, with $\Delta > 0$ being the main diagonal entry of the positive definite Toeplitz matrix $\hat{\Delta}_n(m_1, m_2, k)$.*

PROPOSITION 4.2. *If the coefficient $a(x, y)$ is strictly positive, then $\{\hat{P}_n(a)\}_n$ is a sequence of well-defined symmetric positive definite matrices. The same holds true, at least for*

n large enough, in the case where $a(x,y)$ is a nonnegative function whose zeros are given by a finite number of curves intersecting any horizontal or vertical line only a finite number of times.

Proof. For any r and t, the functions $a_1[t](x) = a(x,y_t)$ and $a_2[r](y) = a(x_r,y)$ have at most a finite number of isolated zeros. Since the discretization of the operators $(-1)^k\partial^k/\partial x^k(a(x,y_t)\partial^k(\cdot)/\partial x^k)$ and $(-1)^k\partial^k/\partial y^k(a(x_r,y)\partial^k(\cdot)\ /\partial y^k)$ leads to matrices such that each diagonal entry is a positive linear combination of evaluations of $a_1[t](x)$ and $a_2[r](y)$, respectively, over a constant number of asymptotically close points, it follows that the generic diagonal entry is positive for n_1 and n_2 large enough. Therefore, $\hat{D}_n(a)$ is positive definite and this ends the proof. □

4.1. The strictly elliptic case.

4.1.1. Asymptotic expansion of preconditioned matrices.

First we introduce and analyze an auxiliary matrix sequence $\{\hat{A}_n^*(a,m_1,m_2,k)\}$, that is used to deduce the clustering properties of the preconditioned matrix sequence $\{\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)\}$.

DEFINITION 4.3. *Set $\hat{A}_n^*(a,m_1,m_2,k) = \hat{D}_n^{-1/2}(a)\hat{A}_n(a,m_1,m_2,k)\hat{D}_n^{-1/2}(a)$, where $\hat{A}_n(a,m_1,m_2,k)$ and $\hat{D}_n(a)$ are defined according to Definition 4.1.*

PROPOSITION 4.4. *If the coefficient $a(x,y)$ is strictly positive and belongs to $\mathbf{C}^2(\overline{\Omega})$, then the matrices $\hat{A}_n^*(a,m_1,m_2,k)$ can be expanded as*

$$\hat{A}_n^*(a,m_1,m_2,k) = \hat{\Delta}_n(m_1,m_2,k) + h^2\Theta_n(a,m_1,m_2,k) + o(h^2)E_n(a,m_1,m_2,k),$$

where $\Theta_n(a,m_1,m_2,k)$ and $E_n(a,m_1,m_2,k)$ are symmetric bounded two-level band matrices. If $a(x,y)$ belongs to $\mathbf{C}^1(\overline{\Omega})$, then $\hat{A}_n^*(a,m_1,m_2,\ k) = \hat{\Delta}_n(m_1,m_2,k) + \Theta_n(a,m_1,m_2,k)$, where $\Theta_n(a,m_1,m_2,k)$ is a two-level band matrix whose elements are $O(h\omega_{a_x}(h)) + O(h\omega_{a_y}(h))$. Finally, if $a(x,y)$ belongs to $\mathbf{C}(\overline{\Omega})$, then $\hat{A}_n^*(a,m_1,m_2,k) = \hat{\Delta}_n(m_1,m_2,k) + \Theta_n(a,m_1,m_2,k)$, where $\Theta_n(a,m_1,m_2,k)$ is a two-level band matrix whose elements are $O(\omega_a(h))$. Here the matrices $\Theta_n(a,m_1,m_2,k)$ and $E_n(a,m_1,m_2,k)$ always show the same pattern as $\hat{\Delta}_n(m_1,m_2,k)$ and the symbol $\omega_f(\cdot)$ denotes the modulus of continuity of a function f.

Proof. Due to the symmetry of the matrices $\hat{A}_n^*(a,m_1,m_2,k)$ it is enough to consider the nonzero coefficients $(\hat{A}_n^*)_{s,s-v}$ related to the lower triangular part. Let us denote by $(\hat{A}_n)_{s,s-v}^{[r,t]}$ the entry of the matrix $\hat{A}_n(a,m_1,m_2,k)$ at position $(s,s-v)$ related to the discretization of the continuous operator at the grid point $(x,y)=(x_r,y_t)$. Let us denote by $\hat{\Delta}_s^{[i]} = \hat{\Delta}_s^{[i]}(m_i,k)$ the entry along the main diagonal of the unilevel Toeplitz matrix $\hat{\Delta}(m_i,k)$ $i=1,2$ and as $\hat{\Delta}_{s-v}^{[i]} = \hat{\Delta}_{s-v}^{[i]}(m_i,k)$ the entry along the v^{th} subdiagonal. Clearly, the following equalities are true:

$$\begin{aligned}
\hat{\Delta}_{s,s}(m_1,m_2,k) &= \alpha_1^{2k}\hat{\Delta}_s^{[1]}(m_1,k) + \alpha_1^{2k}\hat{\Delta}_s^{[2]}(m_2,k),\\
\hat{\Delta}_{s,s-v}(m_1,m_2,k) &= \alpha_1^{2k}\hat{\Delta}_{s-v}^{[1]}(m_1,k),\\
\hat{\Delta}_{s,s-n_1v}(m_1,m_2,k) &= \alpha_2^{2k}\hat{\Delta}_{s-v}^{[2]}(m_2,k).
\end{aligned}$$

Let us denote by Δ the main diagonal entry of the two-level Toeplitz matrix, i.e. $\Delta = \hat{\Delta}_{s,s}(m_1,m_2,k)$. Finally, set $f^{[r,t]} = f(x_r,y_t), f_z^{[r,t]} = (\partial f(x,y)/\partial z)_{|(x,y)=(x_r,y_t)}, f_{zz}^{[r,t]} = (\partial^2 f(x,y)/\partial z^2)_{|(x,y)=(x_r,y_t)}$, where $z\in\{x,y\}$. Now, the coefficients of the matrix $\hat{A}_n^* = \hat{A}_n^*(a,m_1,m_2,k)$ are defined as

$$(\hat{A}_n^*)_{s,s-t} = \Delta\ (\hat{A}_n)_{s,s-t} \Big/ \sqrt{(\hat{A}_n)_{s,s}(\hat{A}_n)_{s-t,s-t}}\,.$$

Case k odd:
For $v=0$, we simply have $(\hat{A}_n^*)_{s,s}=\Delta$, so that $(\Theta_n)_{s,s}=(E_n)_{s,s}=0$.
For $v=1,\dots 2m_1-1$, we set $p=v$ and we have

$$(\hat{A}_n^*)_{s,s-p}=\Delta\,(\hat{A}_n)^{[r,t]}_{s,s-p}\Big/\sqrt{(\hat{A}_n)^{[r,t]}_{s,s}(\hat{A}_n)^{[r-p,t]}_{s,s}}.$$

Now, by considering Taylor's expansions centered at $(x^*,y^*)=(x_{r-p/2},y_t)$ according to Proposition A.1 in [34] and under the assumption of strict positiveness of the coefficient $a(x,y)$, we infer that

$$\begin{aligned}
&(\hat{A}_n^*)_{s,s-p}\\
&=\frac{\Delta\left[a^{[r-\frac{p}{2},t]}\alpha_1^{2k}\Delta^{[1]}_{s-p}+h^2a_{xx}^{[r-\frac{p}{2},t]}\alpha_p+o(h^2)\right]}{\sqrt{\left[a^{[r-\frac{p}{2},t]}(\alpha_1^{2k}\Delta^{[1]}_s+\alpha_2^{2k}\Delta^{[2]}_s)+ha_x^{[r-\frac{p}{2},t]}\beta_p+h^2(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p+a_{yy}^{[r-\frac{p}{2},t]}\eta_p\right]+o(h^2)}}\\
&\cdot\frac{1}{\sqrt{\left[a^{[r-\frac{p}{2},t]}(\alpha_1^{2k}\Delta^{[1]}_s+\alpha_2^{2k}\Delta^{[2]}_s)-ha_x^{[r-\frac{p}{2},t]}\beta_p+h^2(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p+a_{yy}^{[r-\frac{p}{2},t]}\eta_p\right]+o(h^2)}}
\end{aligned}$$

$$\begin{aligned}
&=\frac{\left[\alpha_1^{2k}\Delta^{[1]}_{s-p}+h^2a_{xx}^{[r-\frac{p}{2},t]}\alpha_p^*+o(h^2)\right]}{\sqrt{\left[1+ha_x^{[r-\frac{p}{2},t]}\beta_p^*+h^2(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p^*+a_{yy}^{[r-\frac{p}{2},t]}\eta_p^*\right]\left[1-ha_x^{[r-\frac{p}{2},t]}\beta_p^*+h^2(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p^*+a_{yy}^{[r-\frac{p}{2},t]}\eta_p^*\right]+o(h^2)}}\\
&=\frac{\left[\alpha_1^{2k}\Delta^{[1]}_{s-p}+h^2a_{xx}^{[r-\frac{p}{2},t]}\alpha_p^*+o(h^2)\right]}{\sqrt{1+h^2\left(2\left(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p^*+a_{yy}^{[r-\frac{p}{2},t]}\eta_p^*\right)-\left(a_x^{[r-\frac{p}{2},t]}\beta_p^*\right)^2\right)+o(h^2)}}\\
&=\alpha_1^{2k}\Delta^{[1]}_{s-p}+\left(\alpha_p^*a_{xx}^{[r-\frac{p}{2},t]}-\tfrac{1}{2}\Delta^{[1]}_{s-p}\left(2\left(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p^*+a_{yy}^{[r-\frac{p}{2},t]}\eta_p^*\right)\right.\right.\\
&\left.\left.-\left(a_x^{[r-\frac{p}{2},t]}\beta_p^*\right)^2\right)\right)h^2+o(h^2),
\end{aligned}$$

so that

$$(\Theta_n)_{s,s-p}=\alpha_p^*a_{xx}^{[r-\frac{p}{2},t]}-\frac{1}{2}\Delta^{[1]}_{s-p}\left(2\left(a_{xx}^{[r-\frac{p}{2},t]}\gamma_p^*+a_{yy}^{[r-\frac{p}{2},t]}\eta_p^*\right)-\left(a_x^{[r-\frac{p}{2},t]}\beta_p^*\right)^2\right).$$

For $v=n_1,\dots,n_1(2m_2-1)$, we set $p=v/n_1$ and we deduce that

$$(\hat{A}_n^*)_{s,s-n_1p}=\Delta\,(\hat{A}_n)^{[r,t]}_{s,s-n_1p}\Big/\sqrt{(\hat{A}_n)^{[r,t]}_{s,s}(\hat{A}_n)^{[r,t-p]}_{s,s}}.$$

Again by considering Taylor's expansions centered at $(x^*,y^*)=(x_r,y_{t-p/2})$ according to Proposition A.1 in [34] and under the assumption of strict positiveness of the function $a(x,y)$, we prove in the same way that

$$\begin{aligned}
(\hat{A}_n^*)_{s,s-n_1p}=&\frac{\Delta\left[a^{[r,t-\frac{p}{2}]}\alpha_2^{2k}\Delta^{[2]}_{s-p}+h^2a_{yy}^{[r,t-\frac{p}{2}]}\psi_p+o(h^2)\right]}{\sqrt{\left[a^{[r,t-\frac{p}{2}]}(\alpha_1^{2k}\Delta^{[1]}_s+\alpha_2^{2k}\Delta^{[2]}_s)+ha_y^{[r,t-\frac{p}{2}]}\rho_p+h^2(a_{xx}^{[r,t-\frac{p}{2}]}\delta_p+a_{yy}^{[r,t-\frac{p}{2}]}\phi_p\right]+o(h^2)}}\\
&\cdot\frac{1}{\sqrt{\left[a^{[r,t-\frac{p}{2}]}(\alpha_1^{2k}\Delta^{[1]}_s+\alpha_2^{2k}\Delta^{[2]}_s)-ha_y^{[r,t-\frac{p}{2}]}\rho_p+h^2(a_{xx}^{[r,t-\frac{p}{2}]}\delta_p+a_{yy}^{[r,t-\frac{p}{2}]}\phi_p\right]+o(h^2)}}\\
=&\ \alpha_2^{2k}\Delta^{[2]}_{s-p}+\left(\psi_p^*a_{yy}^{[r,t-\frac{p}{2}]}-\tfrac{1}{2}\Delta^{[2]}_{s-p}\left(2\left(a_{xx}^{[r,t-\frac{p}{2}]}\delta_p^*+a_{yy}^{[r,t-\frac{p}{2}]}\phi_p^*\right)\right.\right.\\
&\left.\left.-\left(a_y^{[r,t-\frac{p}{2}]}\rho_p^*\right)^2\right)\right)h^2+o(h^2),
\end{aligned}$$

ETNA
Kent State University
etna@mcs.kent.edu

so that

$$(\Theta_n)_{s,s-n_1p} = \psi_p^* a_{yy}^{[r,t-\frac{p}{2}]} - \frac{1}{2}\Delta_{s-p}^{[2]}\left(2\left(a_{xx}^{[r,t-\frac{p}{2}]}\delta_p^* + a_{yy}^{[r,t-\frac{p}{2}]}\phi_p^*\right) - \left(a_y^{[r,t-\frac{p}{2}]}\rho_p^*\right)^2\right),$$

and the claimed result follows.

Case k even:
Since, by virtue of Proposition A.2 in [34], the same type of asymptotic expansions hold true for the coefficients of the matrix $\hat{A}_n(a,m_1,m_2,k)$, the result follows in the same manner and with analogous defining relations with respect to the entries of the matrix Θ_n. When the function $a(x,y)$ has less regularity, the claimed result is an easy consequence of the preceding steps where the Taylor expansions are stopped according to the regularity of $a(x,y)$. □

4.1.2. Optimality properties of preconditioned matrices. In the special case $k=1$ and when the function $a(x,y)$ is strictly positive and belongs to $\mathbf{C}^2(\overline{\Omega})$, it is possible to prove the optimality of the devised preconditioner sequence according to the fact that the spectra of the preconditioned matrices belong for any n to a positive interval well separated from zero.

THEOREM 4.5. *If $k = 1$, the coefficient $a(x,y)$ is strictly positive and belongs to $\mathbf{C}^2(\overline{\Omega})$, and the order of the zeros of the related Toeplitz generating polynomial $\alpha_1^2|p_{\mathbf{c}}(x)|^2 + \alpha_2^2|p_{\mathbf{d}}(y)|^2$ does not exceed 2, then the spectra of the sequence of preconditioned matrices $\{\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)\}_n$ belong to the interval $[d_1,d_2]$, with d_i positive universal constants well separated from zero.*

Proof. Due to a similarity argument of $\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)$ and $\hat{\Delta}_n^{-1}(m_1,m_2,k)\hat{A}_n^*(a,m_1,m_2,k)$ we analyze the sequence $\{\hat{\Delta}_n^{-1}(m_1,m_2,k)\hat{A}_n^*(a,m_1,m_2,k)\}_n$. By the assumptions and by Proposition 4.4 we have

$$\hat{A}_n^*(a,m_1,m_2,k) = \hat{\Delta}_n(m_1,m_2,k) + h^2\Theta_n(a,m_1,m_2,k) + o(h^2)E_n(a,m_1,m_2,k),$$

so that

$$\begin{aligned}&\hat{\Delta}_n^{-1}(m_1,m_2,k)\hat{A}_n^*(a,m_1,m_2,k) =\\ &\quad = I_n + \hat{\Delta}_n^{-1}(m_1,m_2,k)(h^2\Theta_n(a,m_1,m_2,k) + o(h^2)E_n(a,m_1,m_2,k)).\end{aligned}$$

By the hypothesis on the order of the zeros of $\alpha_1^2|p_{\mathbf{c}}(x)|^2 + \alpha_2^2|p_{\mathbf{d}}(y)|^2$ we infer that there exists a constant C so that $\|\hat{\Delta}_n^{-1}(m_1,m_2,k)\|_2 \le Ch^{-2}$ [6, 28]. Therefore, by standard linear algebra, we know that

$$\begin{aligned}\lambda_{\max}(\hat{\Delta}_n^{-1}(m_1,m_2,k)\hat{A}_n^*(a,m_1,m_2,k)) &\le \|\hat{\Delta}_n^{-1}(m_1,m_2,k)\hat{A}_n^*(a,m_1,m_2,k)\|_2\\ &\le \|I_n\|_2 + \|\hat{\Delta}_n^{-1}(m_1,m_2,k)(h^2\Theta_n(a,m_1,m_2,k)\\ &\quad + o(h^2)E_n(a,m_1,m_2,k))\\ &\le 1 + C\|\Theta_n\|_2 + o(1).\end{aligned}$$

Conversely, for obtaining a bound from below for $\lambda_{\min}(\hat{\Delta}_n^{-1}(m_1,m_2,k)\hat{A}_n^*(a,m_1,m_2,k))$ we consider the inverse matrix $[\hat{A}_n^*(a,m_1,m_2,k)]^{-1}\hat{\Delta}_n(m_1,m_2,k)$ and we apply Proposition 4.4 to get

$$\begin{aligned}&\left[\hat{A}_n^*(a,m_1,m_2,k)\right]^{-1}\hat{\Delta}_n(m_1,m_2,k) =\\ &= I_n - \left[\hat{A}_n^*(a,m_1,m_2,k)\right]^{-1}(h^2\Theta_n(a,m_1,m_2,k) + o(h^2)E_n(a,m_1,m_2,k)).\end{aligned}$$

ETNA
Kent State University
etna@mcs.kent.edu

Since $a(x, y)$ is positive, we deduce that

$$\begin{aligned}\|[\hat{A}_n^*(a, m_1, m_2, k)]^{-1}\|_2 &\le (\max_{\overline{\Omega}} a(x, y))(\min_{\overline{\Omega}} a(x, y))^{-1} \|\hat{\Delta}_n^{-1}(m_1, m_2, k)\|_2 \\ &\le C(\max_{\overline{\Omega}} a(x, y))(\min_{\overline{\Omega}} a(x, y))^{-1} h^{-2},\end{aligned}$$

so that

$$\begin{aligned}&\lambda_{\min}(\hat{\Delta}_n^{-1}(m_1, m_2, k)\hat{A}_n^*(a, m_1, m_2, k)) \ge \\ &\qquad (1 + C(\max_{\overline{\Omega}} a(x, y))(\min_{\overline{\Omega}} a(x, y))^{-1} \|\Theta_n\|_2 + o(1))^{-1}.\end{aligned}$$

□

Observe that the preceding result in the special case where $m_1 = m_2 = 1$ was stated in [27]. However, in the proof of Theorem 4.1 in [27] there is a flaw, so the preceding result also provides a rigorous proof of the quoted theorem.

4.1.3. Clustering properties of preconditioned matrices. First, we give some definitions of clustering properties.

DEFINITION 4.6. *Let $\{P_n\}_n$ be a preconditioning matrix sequence for the sequence $\{A_n\}_n$ with $A_n, P_n \in \mathbb{C}^{d_n \times d_n}$ and $d_n < d_{n+1}$. Suppose that for any fixed n and for any $\varepsilon > 0$ all the singular values of $P_n^{-1}(A_n - P_n)$ belong to $[0, \varepsilon)$ except for $N_o(\varepsilon, n)$ outliers. If $N_o(\varepsilon, n) = o(d_n)$ then the Weak Clustering property holds, while if $N_o(\varepsilon, n) = O(1)$ then the Strong Clustering property holds. We call the Weakest Strong Clustering property the limit case when $\lim_{\varepsilon \to 0} \lim_{n \to \infty} N_o(\varepsilon, n) = 0$. More precisely, this case occurs when for any $C_\varepsilon > 0$ so that $N_o(\varepsilon, n) \le C_\varepsilon$ holds definitely, $\sup_\varepsilon C_\varepsilon = \infty$ is found. If, for any n, the matrices A_n and P_n are Hermitian and the matrices P_n are positive definite, then the previous clustering properties are in the sense of eigenvalues.*

Notice that the case where $\sup_\varepsilon C_\varepsilon = C \in \mathbb{R}^+$ is characterized by a "true superlinear" behavior of the PCG method, in the sense that for n going to infinity the number of the iterations decreases to a value close to $\lceil C \rceil$. When the *Weakest Strong Clustering* is obtained, then the PCG method is optimal [3] in the sense that we generally observe a number of iterations which is constant with respect to n. (This behavior also characterizes the case where all the eigenvalues belong to a fixed interval bounded away from zero.)

A preliminary result on clustering properties makes direct use of the following lemma.

LEMMA 4.7. [30] *Consider two sequences $\{A_n\}_n$ and $\{B_n\}_n$, where $A_n, B_n \in \mathbb{C}^{d_n \times d_n}$ and $d_n < d_{n+1}$. If the sequence $\{B_n\}_n$ is* sparsely vanishing *(with B_n nonsingular at least definitely) and for any $\varepsilon > 0$ there exists a sequence $\{D_n(\varepsilon)\}_n$, where $D_n(\varepsilon) \in \mathbb{C}^{d_n \times d_n}$, so that $\|A_n - B_n - D_n(\varepsilon)\|_2 \le \varepsilon$ with* rank$(D_n(\varepsilon)) \le \varepsilon d_n$*, then the* Weak Clustering property *holds.*

THEOREM 4.8. *If the coefficient $a(x, y)$ is strictly positive and belongs to $\mathbf{C}(\overline{\Omega})$, then for any $\varepsilon > 0$ all the eigenvalues of the preconditioned matrix*

$$\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$$

lie in the open interval $(1 - \varepsilon, 1 + \varepsilon)$ except for $o(N(n))$ outliers [Weak Clustering property].

Proof. Due to the similarity between $\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$ and $\hat{\Delta}_n^{-1}(m_1, m_2, k)$ $\hat{A}_n^*(a, m_1, m_2, k)$, we can analyze the spectra of the latter. By virtue of Theorem 3.6, the sequence $\{\hat{\Delta}_n(m_1, m_2, k)\}_n$ is *sparsely vanishing* and the considered matrices are nonsingular for any n. Therefore, by recalling Proposition 4.4, Lemma 4.7 applies with $D_n \equiv 0$ for any n, so that the claimed result follows. □

In the special case $k = 1$, a stronger result can be obtained, by making use of the following technical lemma.

LEMMA 4.9. *Let $\{\varepsilon_n\}$ be a sequence decreasing to zero (as slowly as we want) and let us assume that the polynomial $p(x, y) = \alpha_1^2|p_{\mathbf{c}}(x)|^2 + \alpha_2^2|p_{\mathbf{d}}(y)|^2$, generating the Toeplitz sequence $\{\hat{\Delta}_n(m_1, m_2, 1)\}_n$, is such that at least one between $|p_{\mathbf{c}}(v)|^2$ and $|p_{\mathbf{d}}(v)|^2$ is strictly positive for $v \neq 0 \pmod{2\pi}$. Then,*

$$\#\left\{i : \lambda_i\left(\hat{\Delta}_n(m_1, m_2, 1)\right) < \left\lceil \varepsilon_n^{-1} \right\rceil h^2\right\} = O\left(\left\lceil \varepsilon_n^{-1} \right\rceil\right).$$

Proof. Under the given assumption we have $p(x, y) = 0$ if and only if $(x, y) = (0, 0)$. Moreover, the consistency condition implies that $p(x, y) \sim x^2 + y^2$ on the whole definition domain. Therefore, by the transitivity of the relation $\sim$, it follows that $p(x, y) \sim 4 - 2\cos(x) - 2\cos(y)$.

Let S_n be the real space of the $N(n) \times N(n)$ Hermitian matrices. Since $T_n : L^1((-\pi, \pi]^2, \mathbb{R}) \to S_n$ is a linear positive operator, it is easy to check that each $\lambda_j(T_n(\cdot))$ is a homogeneous monotone functional [14]. Therefore, if $f \sim g$ over $(-\pi, \pi]^2$ it ensues that

$$\lambda_j(T_n(f)) \sim \lambda_j(T_n(g))), \tag{4.1}$$

with uniform asymptoticity constants independent of j and n. Now, the eigenvalues of $T_n(4 - 2\cos(x) - 2\cos(y))$ are explicitly known [5] and are given by $4 - 2\cos(s\pi/(n_1 + 1)) - 2\cos(t\pi/(n_2 + 1))$ with $1 \leq s \leq n_1$ and $1 \leq t \leq n_2$ so that a straightforward calculation leads to

$$\#\left\{i : \lambda_i\left(T_n(4 - 2\cos(x) - 2\cos(y))\right) < \left\lceil \varepsilon_n^{-1} \right\rceil h^2\right\} = O\left(\left\lceil \varepsilon_n^{-1} \right\rceil\right).$$

The use of relation (4.1) concludes the proof. □

THEOREM 4.10. *Let us consider $k = 1$ and any choice of m_1, m_2 such that the assumptions of Lemma 4.9 hold true for the polynomial generating the Toeplitz sequence $\{\hat{\Delta}_n(m_1, m_2, 1)\}_n$. If the coefficient $a(x, y)$ is strictly positive and belongs to $\mathbf{C}^2(\overline{\Omega})$, then for any sequence $\{\varepsilon_n\}$ decreasing to zero (as slowly as we want) and for any $\varepsilon > 0$, there exists $\bar{n}$ such that for any $n > \bar{n}$ (with respect to the partial ordering of $\mathbb{N}^2$), $N(n) - O\left(\left\lceil \varepsilon_n^{-1} \right\rceil\right)$ eigenvalues of the preconditioned matrix $\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$ belong to the open interval $(1 - \varepsilon, 1 + \varepsilon)$* [Weakest Strong Clustering property].

Proof. It is exactly the same proof of Theorem 5.6 in [33]. The key fact to be used is the statement of Lemma 4.9 in place of Lemma 5.5 in [33]. □

Now, the fact $\hat{A}_n^*(a, m_1, m_2, k) = \hat{\Delta}_n(m_1, m_2, k) + O(h^2)$, proved in Proposition 4.4 in the case $a(x, y) \in C^2(\overline{\Omega})$, is in some sense exceptional because it is produced by the cancellation of $O(h)$ terms in the expression of $\hat{A}_n^*(a, m_1, m_2, k) - \hat{\Delta}_n(m_1, m_2, k)$. Indeed, in the case $k > 1$ or when $k = 1$, but the assumptions of Lemma 4.9 are violated or $a(x, y)$ does not belong to $C^2(\overline{\Omega})$, we loose the *Weakest Strong Clustering property*. Nevertheless, we are able to characterize the asymptotic behavior of the function describing the number of outlying eigenvalues. More precisely, under the assumption that the template problem (1.1) has been discretized so that the error matrix $\hat{A}_n^*(a, m_1, m_2, k) - \hat{\Delta}_n(m_1, m_2, k)$ has a spectral norm bounded by $O(h^t)$, with t a positive real value, we infer an estimate concerning the number of outlying eigenvalues as a function of t, but also depending on the "spectral difficulty" of the problem represented by the parameter k.

THEOREM 4.11. *Assume that $\hat{A}_n^*(a, m_1, m_2, k) = \hat{\Delta}_n(m_1, m_2, k) + O(h^t)$, with t a positive real value and assume that the nonnegative polynomial generating the Toeplitz* *sequence* $\{\hat{\Delta}_n(m_1, m_2, k)\}_n$ *has a unique zero at* $(x, y) = (0, 0)$*. If the coefficient* $a(x, y)$ *is strictly positive and belongs to* $\mathbf{C}(\overline{\Omega})$*, then for any* $\varepsilon_n = o(h^{2-t/k})$ *decreasing to zero and for any* $\varepsilon > 0$ *there exists* $\bar{n}$ *such that for any* $n > \bar{n}$ *at least* $N(n) - O\left(\lceil \varepsilon_n^{-1} \rceil\right)$ *eigenvalues of the preconditioned matrix* $\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$ *lie in the open interval* $(1-\varepsilon, 1+\varepsilon)$*.*

Proof. Let $\tau_n(p)$ be the two-level τ correction of the Toeplitz matrix $\hat{\Delta}_n(m_1, m_2, k)$ generated by the polynomial p (see [5] for details). By [31] it is known that there exists a constant $c > 0$ independent of n such that $\lambda(\tau_n^{-1}(p)\hat{\Delta}_n(m_1, m_2, k)) \geq c$ with $\lambda(X)$ denoting the generic eigenvalue of a square matrix X. Since $\tau_n(p)$ and $\hat{\Delta}_n(m_1, m_2, k)$ are both positive definite, it follows that the preceding eigenvalue inequality can be read in terms of a Rayleigh quotient inequality, namely

$$c\mathbf{x}^H \tau_n(p)\mathbf{x} \leq \mathbf{x}^H \hat{\Delta}_n(m_1, m_2, k)\mathbf{x}, \quad \forall \mathbf{x} \in \mathbb{C}^{N(n)}.$$

This means that the matrix $\hat{\Delta}_n(m_1, m_2, k) - c\tau_n(p)$ is nonnegative definite, so that by the homogeneous monotonicity of the eigenvalues it follows that, for any j, $c\lambda_j(\tau_n(p)) \leq \lambda_j(\hat{\Delta}_n(m_1, m_2, k))$. Therefore, we have

$$\#\left\{i : \lambda_i\left(\hat{\Delta}_n(m_1, m_2, k)\right) < \lceil \varepsilon_n^{-1} \rceil h^t\right\} \leq \#\left\{i : \lambda_i\left(\tau_n(p)\right) < c^{-1}\lceil \varepsilon_n^{-1} \rceil h^t\right\}. \tag{4.2}$$

As in Lemma 4.9, we now exploit the exact knowledge of the eigenvalues of $\tau_n(p)$ which are given by the sampling of p over the points $(s\pi/(n_1+1), t\pi/(n_2+1)), 1 \leq s \leq n_1, 1 \leq t \leq n_2$. The key point is that, by the consistency condition and by the assumption, p has a unique zero of order $2k$ so that the small eigenvalues of $\tau_n(p)$ behave like $h^{2k}(j^2+q^2)^k$. At this point we follow the same lines as in Theorem 5.7 in [33] and we apply inequality (4.2) in order to complete the proof. □

Notice that the growth of the order k of the differential problem leads to a deterioration of the "strength" of the cluster, so that in order to obtain a better clustering for higher order problems, it is necessary to increase the order of approximation of $\hat{\Delta}_n(m_1, m_2, k)$ by $\hat{A}_n^*(a, m_1, m_2, k)$. Moreover, up to positive constants, the number of the outliers in the two-dimensional case is the square of the number of outliers of the one-dimensional case. This is part of a more general fact. Actually, if a d-dimensional elliptic problem is considered, then the number of outliers will grow as the d-th power of the number of outliers of the one-dimensional case. To see this it is enough to follow the proof of Theorem 4.11 in d dimensions, with the natural assumption that the d-variate polynomial p has a unique zero at the origin.

4.2. The degenerate elliptic case.

4.2.1. Clustering properties of preconditioned matrices. The reason of the very fast PCG convergence observed when we consider the preconditioning matrix sequence $\{\hat{P}_n(a)\}_n$ with respect to the case of the basic Toeplitz preconditioning matrix sequence $\{\hat{\Delta}_n(m_1, m_2, k)\}_n$ is explained in the following theorem.

THEOREM 4.12. *If the coefficient* $a(x, y)$ *belongs to* $\mathbf{C}(\overline{\Omega})$ *and is nonnegative with a finite number of zeros, then for any* $\varepsilon > 0$ *all the eigenvalues of the preconditioned matrix* $\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$ *lie in the open interval* $(1-\varepsilon, 1+\varepsilon)$ *except for* $o(N(n))$ *outliers* [Weak Clustering property].

Proof. Due to the similarity between $\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$ and $\hat{\Delta}_n^{-1}(m_1, m_2, k)$ $\hat{A}_n^*(a, m_1, m_2, k)$, we can analyze the spectra of the latter matrix sequence. First, for the sake of simplicity, we consider the case when $a(x, y)$ has a unique zero located at $(x, y) = (0, 0)$. For any mesh point (x_r, y_t) such that the involved sampling of the coefficient $a(x, y)$ are well separated from $(x,y)=(0,0)$, by virtue of Proposition 4.4, we find that

$$(\hat{A}_n^*(a,m_1,m_2,k))_{s,s\pm v}=(\hat{\Delta}_n(m_1,m_2,k))_{s,s\pm v}+(\Theta_n(a,m_1,m_2,k))_{s,s\pm v}, \tag{4.3}$$

where $\|\Theta_n(a,m_1,m_2,k)\|_2=O(\omega_a(h))$. Now, for any positive ϵ, due to the continuity of a and to the assumption on the zeros, it follows that there exists $\delta=\delta_\epsilon$ so that the set $\{(x,y)\in\Omega^*:a(x,y)<\delta\}$ is contained in a finite union of balls $\mathcal{B}$, whose Lebesgue measure is bounded by $\epsilon/4$. Therefore, we consider a correction matrix D_n involving only those rows for which at least one associated mesh point is such that the evaluation of a is less than δ. Since the relationships

$$\begin{aligned}\#\{\{(i,t):a(\tilde{x}_i,y_t)<\delta\}\cup\{(r,j):a(x_r,\tilde{y}_j)<\delta\}\}&\le\#\{\{(i,t):a(\tilde{x}_i,y_t)\in\mathcal{B}\}\cup\\&\qquad\{(r,j):a(x_r,\tilde{y}_j)\in\mathcal{B}\}\}\\&=\frac{2N(n)}{m\{\Omega^*\}}m\{\mathcal{B}\}+O([N(n)]^{1/2})\\&\le N(n)(\varepsilon+o(1))/2\end{aligned}$$

are true, we deduce that $\mathrm{rank}(D_n)\le\varepsilon N(n)$. Therefore, Lemma 4.7 applies as in Theorem 4.8 and the claimed result follows. Notice that the presence of a zero in different position moves the position of the nonzero part of D_n along the diagonal, but does not change its asymptotic rank. In addition, the proof is unchanged in the case of presence of a finite number of zeros (or in the general case where the zeros are a Peano-Jordan measurable set with zero Lebesgue measure). ☐

¿From Theorem 4.12, we deduce that almost all the eigenvalues of the preconditioned matrices are clustered at unity except for $o(N(n))$ outliers. This is not completely satisfactory, but is very good when compared with the results obtained by considering the purely Toeplitz preconditioning sequence.

5. General results on distribution and clustering. The aim of this section is to give general results on the approximation of $\hat{A}_n^*(a,m_1,m_2,k)$ by $\hat{\Delta}_n(m_1,m_2,k)$ in the spirit of the ergodic Theorems proved by Szegö. Let us first recall the following definition

DEFINITION 5.1. *Two sequences $\{A_n\}_n$ and $\{B_n\}_n$, $A_n,\ B_n\in\mathbb{C}^{d_n\times d_n}$, $d_n<d_{n+1}$ are said to be* equally distributed *in the sense of the eigenvalues if and only if, for any real-valued continuous function F with bounded support we have*

$$\lim_{n\to\infty}d_n^{-1}\sum_{i=1}^{d_n}(F(\lambda_i(A_n))-F(\lambda_i(B_n))=0.$$

THEOREM 5.2. *Let $\{\hat{A}_n^*(a,m_1,m_2,k)\}_n$ and $\{\hat{\Delta}_n(m_1,m_2,k)\}_n$ be defined according to Definitions 4.3 and 4.1. If the coefficient $a(x,y)$ is strictly positive and belongs to $\mathbf{C}(\overline{\Omega})$, then*

$$\begin{aligned}\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)\|_F^2&=N(n)\cdot O(\omega_a^2([N(n)]^{-1/2})),\\\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)\|_2&=O(\omega_a([N(n)]^{-1/2})),\\\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)\|_{\mathrm{tr}}&=N(n)\cdot O(\omega_a([N(n)]^{-1/2})),\end{aligned}$$

where ω_a is the modulus of continuity of $a(x,y)$. If the coefficient $a(x,y)$ belongs to $\mathbf{C}(\overline{\Omega})$ and is nonnegative with a finite number of zeros, then there exists a matrix sequence $\{D_n\}_n$, with $\mathrm{rank}(D_n)=o(N(n))$, such that

$$\begin{aligned}\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)-D_n\|_F^2&=o(N(n)),\\\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)-D_n\|_2&=o(1).\end{aligned}$$

ETNA
Kent State University
etna@mcs.kent.edu

The latter result, in view of Theorem 2.1 in [38] and of the Cauchy interlacing Theorem, tell us that the eigenvalues of the two symmetric matrix sequences $\{\hat{A}_n^*(a,m_1,m_2,k)\}_n$ and $\{\hat{\Delta}_n(m_1,m_2,k)\}_n$ are *equally distributed* according to Definition 5.1. But each matrix $\hat{\Delta}_n(m_1,m_2,k)$ is the $N(n)\times N(n)$ Toeplitz matrix generated by $p(x,y)=\alpha_1^{2k}|p_{\mathbf{c}}(x)|^2+\alpha_2^{2k}|p_{\mathbf{d}}(y)|^2$ and therefore, by taking into account the Szegö ergodic formula [38], for any real-valued continuous function F with bounded support we have

$$\lim_{n\to\infty}\frac{1}{N(n)}\sum_{i=1}^{N(n)}F(\lambda_i(\hat{A}_n^*(a,m_1,m_2,k)))=\frac{1}{(2\pi)^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi}F(p(x,y))dxdy,$$

i.e. the matrix sequence $\{\hat{A}_n^*(a,m_1,m_2,k)\}_n$ is spectrally distributed as $p(x,y)$.

Moreover, it is worth pointing out that equation (1.1) requires that the coefficient $a(x,y)$ belongs to $\mathbf{C}^k(\overline{\Omega})$, so that a refined analysis seems to be just an academic exercise. However, when we consider the "weak formulation" [8], the problem (1.1) is transformed into an integral problem. Therefore, in this sense, the given analysis becomes again meaningful. Concerning this fact, in the case where the coefficient $a(x,y)$ belongs to $L^\infty(\Omega)$, the application of the Lusin Theorem [25] allows one to prove the following result as a simple generalization of Theorem 5.3 in [30].

THEOREM 5.3. *Let $\{\hat{\Delta}_n(m_1,m_2,k)\}_n$ and $\{\hat{A}_n^*(a,m_1,m_2,k)\}_n$ be defined according to Definitions 4.1 and 4.3, but where the coefficients $a(x_i,y_j)$ should be replaced by mean value on the domain $I_i\times I_j=[x_i,x_{i+1}]\times[y_j,y_{j+1}]$ in the sense that $a(x_i,y_j)$ means $N(n)\int_{I_i\times I_j}a(x,y)dxdy$. If the coefficient $a(x,y)$ belonging to $L^\infty(\Omega)$ is nonnegative and, at most, sparsely vanishing, then for any $\varepsilon>0$ there exists a matrix sequence $\{D_n(\varepsilon)\}_n$, with $\mathrm{rank}(D_n)\le\varepsilon N(n)$ and n large enough, such that*

$$\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)-D_n\|_F^2\le c_1(\varepsilon)N(n),$$
$$\|\hat{A}_n^*(a,m_1,m_2,k)-\hat{\Delta}_n(m_1,m_2,k)-D_n\|_2\le c_2(\varepsilon),$$

with $\lim_{\varepsilon\to0}\max\{c_1(\varepsilon),c_2(\varepsilon)\}=0$ and n large enough. In addition, the number of outliers of the sequence of preconditioned matrices $\{\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)\}_n$ is generically $o(N(n))$, while if $a(x,y)$ is not sparsely vanishing then the preconditioners $\hat{P}_n(a)$ are not well defined or the preconditioned matrices have not clustered eigenvalues.

6. Numerical experiments and comparison with the literature. This section is divided into three parts. In the first part, Subsection 6.1, we report and discuss some numerical experiments concerning the PCG method with our preconditioners and with classical preconditioners. Moreover, we heuristically extend our technique to the case of more general domains (L and T shaped). The numbers are very encouraging and seem to suggest that the theoretical analysis of the previous sections should be generalizable to these more interesting cases as well. In the second part, Subsection 6.2, we compare the theoretical and the numerical results obtained in this paper with the existing literature. Finally, in Subsection 6.3, we complete the picture of the whole numerical procedure with some numerical evidences concerning the optimality of a multigrid technique for the ill-conditioned multilevel Toeplitz case (see [12, 31]). Indeed our preconditioning reduces the nonconstant semi-elliptic case to the solution of two-level Toeplitz structures for which we propose the use of a specialized multigrid technique.

6.1. PCG numerical results. We first present in Tables 7.1, 7.2, and 7.3 the number of PCG iterations required to obtain $\|r_s\|_2/\|b\|_2\le10^{-7}$ for increasing values of the matrix

ETNA
Kent State University
etna@mcs.kent.edu

dimension $N(n) = n_1 n_2$, where $n_1 = n_2$ and r_s denotes the residual at the s-th step. The data vector $\mathbf{b}$ is made up of all ones. In [34] we also considered the case of random data uniformly distributed on the interval $[0, 1]$ and the results were absolutely similar. The test functions a are listed in the first column, the preconditioners are given in the heading by $\hat{D} = \hat{D}_n(a)$, $\hat{\Delta} = \hat{\Delta}_n(m_1, m_2, k)$, IC the incomplete Choleski factorization, and $\hat{P} = \hat{P}_n(a)$. Finally, the pair (k, m) varies among $(1, 3)$, $(2, 2)$ and $(3, 2)$ with $m_1 = m_2 = m$. Moreover, in Tables 7.4, 7.5, and 7.6 we give the total number of outliers (and the related percentage) of $\hat{P}_n^{-1}(a)\hat{A}_n(a, m_1, m_2, k)$ with respect to a cluster at unity with radius $\delta = 0.1$. In addition, between round brackets, we report also the number of outliers less than $1 - \delta$.

Concerning the preconditioners $\hat{D}$ and IC, Tables 7.1, 7.2, and 7.3 prove that the associated PCG methods are never optimal. In particular, the number of iterations in the case of the diagonal preconditioning grows linearly as the dimension $N(n)$ for all the considered test functions; for the classical incomplete Choleski preconditioning we observe that the number of iterations is proportional to the square root of $N(n)$. The case of the two Toeplitz based preconditioners, $\hat{\Delta}_n(m_1, m_2, k)$ and $\hat{P}_n(a)$, is quite different and needs a more detailed analysis.

In Tables 7.1, 7.2, and 7.3, we observe that the number of PCG iterations is practically constant when the preconditioner is $\hat{\Delta}_n(m, k)$ or $\hat{P}_n(a)$ and the coefficient a is strictly positive. This independence with regard to n fully agrees with the spectral clustering theorems proved in this paper and with spectral analysis of $\{\hat{\Delta}_n^{-1}(m, k)\hat{A}_n(a, m, k)\}$ given in [29].

The presence of jumps or discontinuities of a or of its derivatives does not spoil the performances of the associated PCG methods when $\hat{\Delta}_n(m, k)$ or $\hat{P}_n(a)$ are used as preconditioners (see [27] for more details on this). The case of highly oscillating coefficient a slightly deteriorates the performances of the second preconditioner $\hat{P}_n(a)$ so that it becomes substantially equivalent to the Toeplitz preconditioner. This is obvious since the matrix $\hat{D}_n(a)$ (the diagonal part of $\hat{A}_n(a, m, k)$) is given by equispaced samples of $a(x, y)$. Therefore, $\hat{D}_n(a)$ cannot be in general a faithful representation of a when a oscillates too much with regard to the grid parameter h. This phenomenon was also noticed in [27] with regard to similar problems where $(k, m) = (1, 1)$.

It should be noticed that a certain growth of the number of outliers predicted by Theorem 4.11 for the preconditioned matrix $\hat{P}_n^{-1}(a)\hat{A}_n(a, m, k)$ in the case where $k > 1$ and $a > 0$, is not observed in Tables 7.5 and 7.6. Probably, it is possible to prove something more.

In the case of $k = 1$, the qualitative improvement given by the *Strong Clustering property* is just theoretical and "redundant" from a practical point of view since, for the case where a is positive and regular, the optimality of our PCG iterations follows from the fact that each eigenvalue of $\hat{P}_n^{-1}(a)\hat{A}_n(a, m, k)$ belongs to $[d_1, d_2]$ with d_1 and d_2 universal positive constants (Theorem 4.5).

If a is essentially positive, then the presence of a countably infinite number of jumps of a (as in the last four test functions) does not spoil the performances of the associated PCG methods when $\hat{\Delta}_n(m, k)$ or $\hat{P}_n(a)$ are used as preconditioners according to the results of Theorem 5.3. Finally notice the similarity of these results with respect to the case where a is smooth [27].

When zero belongs to the essential range of a or a is unbounded, it is immediate to observe that the only working preconditioner is $\hat{P}_n(a)$. Also this result agrees with the theoretical expectations of this paper. In this case, as shown in Tables 7.4, 7.5, and 7.6, the number of outlying eigenvalues grows very slowly (it seems only logarithmically with $N(n)$) and this behavior is much better when compared with the theoretical results. Therefore, we think that the analysis presented in Theorem 4.12 can be substantially refined. In particular, the theoretical tools introduced in [37] and [10] could be used in this context. Concerning the preconditioner $\hat{\Delta}_n(m,k)$ it is worthwhile observing that the case where a has zeros is much worse when compared to the case of a unbounded. This is in accordance with the analysis of Axelsson and Lindskog that showed that "small" outliers slowdown the convergence much more than "big" outliers. Other numerical experiments confirming the preceding observations are reported in [34] and in [27] for the case $(k,m)=(1,1)$.

Furthermore, it is worth stressing that the cases $k=1$, $k=2$ and $k=3$ are substantially identical from the point of view of the preconditioner $\hat{P}_n(a)$ and this indicates that high order differential problems can be successfully handled by using the proposed techniques and that $\hat{P}_n(a)$ is a robust preconditioner with regard to the parameter k.

Further extensions. Finally we briefly show that our technique can be extended to the case where the elliptic operator is defined on a more general domain. We consider three cases: a square domain $Q=(0,1)^2$, a basic L shaped domain $L=Q\backslash Q'$ where $Q'=(0,1/2]^2$, and a basic T shaped domain $T=Q\backslash(R_1\cup R_2)$ with $R_1=(0,1/2]\times(0,1/4]$ and $R_1=(0,1/2]\times[3/4,1)$. The considered coefficient functions include elliptic $(a(x,y)=1+x+y)$, elliptic oscillating $a(x,y)=\sin^2(7(x+y))+1$, semielliptic $(a(x,y)=(1-x+y)^p$, $p=1,2)$ and discontinuous $(a(x,y)=\exp(x+y)\mathrm{Ch}_{\{x+y\leq 2/3\}}+(2-(x+y))\mathrm{Ch}_{\{x+y>2/3\}})$ examples. The parameters are the basic ones $k=1$ and $m_1=m_2=1$ and the symbol Ch_X denotes the characteristic function a set X. Looking at Table 7.8, it is interesting to observe that there is no dependence of the iteration count on the domain and, strangely enough, in the semielliptic case with coefficient $a(x,y)=(1-x+y)^2$ we observe that our preconditioning technique leads to just one iteration of the PCG method for $n_1=n_2$ large enough. In addition in these examples we considered much higher dimensions in order to show that our technique (in combination with a multigrid one, see Table 7.7) is really faster than a usual incomplete Choleski factorization (refer to Table 7.9). Moreover, these larger dimensions allow one to appreciate the superlinearity of the proposed preconditioning technique which leads, in many cases, to a decrease of the number of iterations as $n_1=n_2$ increases (elliptic and semielliptic smooth examples).

6.2. Comparison with the literature. The present paper generalizes the analysis of Toeplitz based preconditioners in many senses. Any order k of the differential operators in (1.1) can be included, all precision orders of the FD formulas are allowed, i.e., all parameters m_1 and m_2 can be used.

Therefore, the results of the preceding sections can be viewed as the final point of the research work started in [27] and [33]. Moreover, this analysis has been helpful in order to extend the technique in the case of Finite Element matrices where the presence of various types of geometrical elements makes the study much more intricate (see [35]).

Now, regarding problem (1.1), and in order to make a short, but informative, comparison with well-established techniques in the literature, we analyze the following three cases:
a.1) $a>0$ and $a\in C^2([0,1]^2)$,
a.2) $a>0$ and $\infty>\sup a\geq\inf a>0$,
a.3) $a\geq 0$ with at most isolated zeros and a piecewise continuous.

The PCG methods based on preconditioners from incomplete LU and Choleski factorizations [23, 9, 16] and from the circulant algebra [7, 18, 21] are sublinear, i.e., require a number of iterations $O([N(n)]^\beta)$ with positive β. This is true even in the case **a.1** where a is positive and smooth as we observed in the Numerical Experiments subsection.

On the other hand the PCG methods defined by using separable preconditioners [11] and the multigrid algorithms [17, 22, 40] are optimal in the sense reported in Section 3 in the cases **a.1** and **a.2**, but not in the case **a.3**. On the contrary, our technique is superlinear in the case **a.1** (therefore also optimal) and assure a "weak" clustering in the cases **a.2** and **a.3**. This property does not guarantee theoretically the optimality, but the numerous numerical

ETNA
Kent State University
etna@mcs.kent.edu

experiments performed in the previous subsection and in [30, 32, 33] suggest a convergence rate independent of the size $N(n)$.

Finally, with regard to the cases **a.1** and **a.2**, we want to stress some interesting features of our technique in comparison with a pure multigrid strategy. The preconditioner $\hat{P}_n(a)$ reduces the nonconstant elliptic and semi-elliptic case to the constant elliptic case, i.e., to two-level Toeplitz positive definite structures. For this specific elliptic problem we propose the use of specialized multigrid strategies costing $O(N(n))$ arithmetic operations, where $N(n)$ is the size of algebraic problem (see, e.g., [12, 31] for multigrid techniques applied to positive definite multilevel Toeplitz structures and [39, 40] for multigrid techniques applied to elliptic differential problems). For instance, the algorithm in [31] can be implemented in a very optimized way since at each level of the recursion of a single multigrid iteration we only need $O(1)$ parameters that identify the actual coefficient matrix. In this way the matrix is never explicitly formed except at the lowest level where the dimension is really small (see the Subsection 6.3). In conclusion, the message is that our hybrid PCG-multigrid technique can solve semi-elliptic problems as well and can save both in computation and in memory. Of course, the latter feature is of crucial interest when huge dimensions are required.

6.3. Multigrid numerical results. Here, for the sake of completeness, we report the number of multigrid iterations in order to solve two-level Toeplitz linear systems where the coefficient matrix takes the form

$$A_n(1, m_1, m_2, k),$$

and where the pair (k, m) varies among $(1, 1)$, $(2, 1)$, $(2, 2)$ and $(3, 2)$ with $m_1 = m_2 = m$. As it is clearly shown in Table 7.7, the considered multigrid technique is optimal with respect to the size $N(n)$, since the number of iterations stabilizes to a given constant depending only on m and k. In addition, we observe a negligible dependency on the bandwidth m and a sublinear growth with regard to the order of the differential operator $2k$. We recall that the related systems are quite ill-conditioned and, in actuality, the condition number of $A_n(a, m_1, m_2, k)$ is asymptotic to $[N(n)]^k$ (refer to [6, 28]) so that a certain deterioration with respect to the parameter k should be expected.

7. Conclusive remarks. To conclude, in this paper, we have discussed the asymptotic distributional properties of the spectra of Toeplitz–based preconditioned matrices arising from FD discretization of the differential problems of the form (1.1). We proved that the general clustering of the spectra holds for $a(x, y)$ ranging from the good case in which it is regular and strictly positive to the bad case where $a(x, y)$ is only $L^\infty(\Omega)$ and sparsely vanishing. Moreover, the results indicate that a possible weak deterioration of the convergence properties of the associated PCG methods occurs when the parameter k and/or the order of the zeros of $a(x, y)$ increases. Finally we stress that the discussed results concern $2D$ differential problems. This choice is motivated by a requirement of notational simplicity. However, there is no difficulty in extending the whole analysis to the case of an arbitrary number of dimensions since the basic tools such as the multidimensional Szegö formula [38] or the extremal behavior [6, 28] of multilevel Toeplitz sequences are available in the relevant literature.

REFERENCES

[1] W. Ames, *Numerical Methods for Partial Differential Equations, Third Edition*, Academic press Inc., New York, 1992.

[2] O. Axelsson and V. Barker, *Finite Element Solution of Boundary Value Problems, Theory and Computation*, Academic press Inc., New York, 1984.

[3] O. Axelsson and G. Lindskog, *On the rate of convergence of the preconditioned conjugate gradient method*, Numer. Math., 48 (1986), pp. 499–523.

[4] O. Axelsson and M. Neytcheva, *The algebraic multilevel iteration methods – theory and applications*, Proc. 2nd Int. Coll. on Numerical Analysis, D. Bainov Ed., Plovdiv (Bulgaria), August 1993, pp. 13–23.

[5] D. Bini and M. Capovani, *Spectral and computational properties of band symmetric Toeplitz matrices*, Linear Algebra Appl., 52/53 (1983), pp. 99–126.

[6] A. Böttcher ans S. Grudsky, *On the condition numbers of large semi-definite Toeplitz matrices*, Linear Algebra Appl., 279 (1998), pp. 285–301.

[7] R. Chan and T. Chan, *Circulant preconditioners for elliptic problems*, J. Numer. Linear Algebra Appl., 1 (1992), pp. 77–101.

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

[9] P. Concus, G. Golub, and G. Meurant, *Block preconditioning for the conjugate gradient method*, TR nr. LBL-14856, Lawrence-Berkeley Laboratory, UCLA, USA, (1982).

[10] F. Di Benedetto and S. Serra Capizzano, *A unifying approach to abstract matrix algebra preconditioning*, Numer. Math., 82-1 (1999), pp. 57–90.

[11] H. Elman and M. Shultz, *Preconditioning by fast direct methods for nonself-adjoint nonseparable elliptic equations*, SIAM J. Numer. Anal., 23 (1986), pp. 44–57.

[12] G. Fiorentino and S. Serra Capizzano, *Multigrid methods for symmetric positive definite block Toeplitz matrices with nonnegative generating functions*, SIAM J. Sci. Comput., 17 (1996), pp. 1068–1081.

[13] ———, *Fast parallel solvers for elliptic problems*, Comput. Math. Appl., 32 (1996), pp. 61–68.

[14] A. Frangioni and S. Serra Capizzano, *Matrix-valued linear positive operators and applications to graph optimization*, TR. 04-99, Dept. Informatica, University of Pisa, Italy. Available online from http://www.di.unipi.it.

[15] G. Golub and C. Van Loan, *Matrix Computations*, The Johns Hopkins Univ. Press, Baltimore, 1983.

[16] I. Gustafsson, *Stability and rate of covergence of modified incomplete Cholesky factorization methods*, *TR nr. 79.02R, Chalmers University of Technology, Sweden,* (1979)

[17] W. Hackbusch, *Multigrid Methods and Applications*, Springer Verlag, Berlin, 1985.

[18] S. Holmgren and K. Otto, *Iterative methods for block-tridiagonal systems of equations*, *SIAM J. Matrix Anal. Appl.*, 13 (1992), pp. 863–886.

[19] ———, *A framework for polynomial preconditioners based on fast transform I*: *Theory*, BIT, 38 (1998), pp. 544–559.

[20] ———, *A framework for polynomial preconditioners based on fast transform II*: *PDE applications*, BIT, 38 (1998), pp. 721–736.

[21] I. Lirkov, S. Margenov and P. Vassilevsky, *Circulant block factorization for elliptic problems*, Computing, 53 (1994), pp. 59–74.

[22] J. Mandel, *Some recent advances in Multigrid Methods*, Adv. Electronics Electr. Phys., 82 (1991), pp. 327–377.

[23] J. Meijerink and H. Van Der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comp., 31 (1977), pp. 148–162.

[24] M. Ng, *Band preconditioners for block-Toeplitz–Toeplitz–block systems*, Linear Algebra Appl., 259 (1997), pp. 307–327.

[25] W. Rudin, *Real and Complex Analysis*, McGraw-Hill, New York, 1985.

[26] S. Serra Capizzano, *Preconditioning strategies for asymptotically ill–conditioned block Toeplitz systems*, BIT, 34 (1994), pp. 579–594.

[27] ———, *The rate of convergence of Toeplitz based PCG methods for second order nonlinear boundary value problems*, Numer. Math., 81-3 (1999), pp. 461–495.

[28] ———, *On the extreme eigenvalues of Hermitian* (*block*) *Toeplitz matrices*, Linear Algebra Appl., 270 (1998), pp. 109–129.

[29] ———*A note on the asymptotic spectra of finite difference discretizations of second order elliptic Partial Differential Equations*, Asian J. Math., 4 (2000), pp. 499–514.

[30] ———, *Spectral behavior of matrix sequences and discretized boundary value problems*, Linear Algebra Appl., 337 (2001), pp. 37–78.

[31] ———, *Convergence analysis of two–grid methods for elliptic Toeplitz and PDEs Matrix-sequences*, Numer. Math., **92-3** (2002), pp. 433–465.

[32] S. Serra Capizzano and C. Tablino Possio, *Spectral and structural analysis of high precision Finite Difference matrices for Elliptic Operators*, Linear Algebra Appl., 293 (1999), pp. 85–131.

[33] ———, *High-order Finite Difference schemes and Toeplitz based preconditioners for Elliptic Problems*, Electron. Trans. Numer. Anal., 11 (2000), pp. 55–84. Available on line from http://etna.mcs.kent.edu/vol.11.2000/pp55-84.dir/pp55-84.pdf.

[34] ———, *Preconditioning strategies for 2D Finite Difference matrix sequences*, TR. 2/1999 - Dipartimento di Scienza dei Materiali, Università di Milano-Bicocca, a shorter version to appear in Linear Algebra Appl.

[35] ———, *Finite Elements matrix sequences*: *the case of rectangular domains*, Numer. Algorithms, 28 (2001), pp. 309–327.

[36] ———, *Superlinear preconditioning of optimal preconditioners for collocation linear systems*, TR. 3/1999

$k=1\ m=3$	$n_1+1=n_2+1$											
	10				20				30			
$a(x,y)$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$
$1+x+y$	32	12	12	3	66	21	13	3	100	30	13	3
$\exp(x+y)$	33	12	18	3	68	21	21	4	100	31	22	4
$\sin^2(7(x+y))+1$	33	13	10	10	69	22	10	11	104	32	11	11
$1+\sqrt{x+y}$, 1 if $x+y<0$	32	12	9	3	65	21	10	3	99	30	10	3
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert+\frac{1}{2}$	15	12	10	5	37	20	12	5	59	27	13	5
$x+y$	34	13	21	4	70	22	31	4	106	32	38	4
$(x+y)^2$	35	13	39	4	73	22	88	4	110	32	137	3
$(x+y)^3$	35	13	62	5	74	22	207	5	112	33	400	5
$(x+y)^4$	36	13	86	6	75	23	421	6	113	33	—	6
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert$	15	12	14	6	37	18	23	7	57	27	30	7
$(x-\frac{1}{2})^2+(y-\frac{1}{2})^2$	15	11	15	6	38	19	45	6	60	27	73	7
$\exp(x+y)$ if $x+y\le\frac{2}{3}$ $4+x+y$ if $x+y>\frac{2}{3}$	34	13	16	7	71	22	19	9	106	31	19	11
$\exp(x+y)$ if $x+y\le\frac{2}{3}$ $2-(x+y)$ if $x+y>\frac{2}{3}$	34	13	21	6	71	22	31	7	106	32	38	8
$\left\lceil\frac{1}{\sqrt{x}}\right\rceil+\left\lceil\frac{1}{\sqrt{y}}\right\rceil$ if $x,y>0$ 1 if x or $y\le 0$	31	12	9	6	64	20	12	7	96	30	13	8
$\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}$ if $x,y>0$ 1 if x or $y\le 0$	31	12	7	5	63	21	8	6	96	30	8	6
$(x+y)\left(\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}\right)$ if $x,y>0$, 1 if x or $y\le 0$	34	13	18	5	68	21	26	6	106	31	32	6
$\exp(\lceil\frac{1}{x}\rceil/(1.+\lceil\frac{1}{x}\rceil))\left\lceil\frac{1}{\sqrt{y}}\right\rceil+y$ if $x,y>0$, 1 if x or $y\le 0$	35	13	12	6	73	23	15	8	114	34	17	9

TABLE 7.1
Number of PCG iterations in the case $k=1,\ m_1=m_2=3$.

- Dipartimento di Scienza dei Materiali, Università di Milano-Bicocca, available from the authors by request.
[37] E. TYRTYSHNIKOV, *Circulant preconditioners with unbounded inverses*, Linear Algebra Appl., 216 (1995), pp. 1–23.
[38] ———, *A unifying approach to some old and new theorems on distribution and clustering*, Linear Algebra Appl., 232 (1996), pp. 1–43.
[39] P. VANEK, J. MANDEL, AND M. BREZINA, *Algebraic Multigrid by smoothed aggregation for second and fourth order elliptic problems*, Computing, 56 (1996), pp. 179–196.
[40] P. VANEK, J. MANDEL, AND M. BREZINA, *Two-level algebraic Multigrid for the Helmholtz problem*, Contemp. Math., 218 (1998), pp. 349–356.
[41] R. S. VARGA, *Matrix Iterative Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1962.

ETNA
Kent State University
etna@mcs.kent.edu

$k = 2\; m = 2$	$n_1 + 1 = n_2 + 1$											
	10				20				30			
$a(x, y)$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$
$1 + x + y$	56	23	13	3	206	64	14	4	448	130	14	4
$\exp(x + y)$	55	23	19	3	205	64	23	3	448	131	24	4
$\sin^2(7(x + y)) + 1$	57	23	10	11	207	64	11	14	456	132	12	14
$1 + \sqrt{x + y}$, 1 if $x + y < 0$	56	23	9	4	206	64	10	4	446	130	11	4
$\|x - \frac{1}{2}\| + \|y - \frac{1}{2}\| + \frac{1}{2}$	15	21	10	5	67	62	13	6	147	123	14	7
$x + y$	58	23	23	5	212	65	35	5	462	129	44	5
$(x + y)^2$	58	24	41	5	214	66	96	6	470	134	149	6
$(x + y)^3$	59	24	66	5	217	67	224	5	465	135	434	5
$(x + y)^4$	59	24	90	5	219	68	447	4	479	138	–	4
$\|x - \frac{1}{2}\| + \|y - \frac{1}{2}\|$	15	21	14	7	67	59	25	9	146	121	33	11
$(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2$	15	20	15	7	67	57	47	9	146	118	76	10
$\exp(x + y)$ if $x + y \leq \frac{2}{3}$ $4 + x + y$ if $x + y > \frac{2}{3}$	58	23	18	9	211	65	20	13	469	134	21	17
$\exp(x + y)$ if $x + y \leq \frac{2}{3}$ $2 - (x + y)$ if $x + y > \frac{2}{3}$	57	24	23	8	213	66	35	10	469	133	43	12
$\left\lceil \frac{1}{\sqrt{x}} \right\rceil + \left\lceil \frac{1}{\sqrt{y}} \right\rceil$ if $x, y > 0$ 1 if x or $y \leq 0$	56	25	11	10	214	68	16	15	476	136	18	21
$\frac{\lceil \frac{1}{x} \rceil}{1 + \lceil \frac{1}{x} \rceil} + \frac{\lceil \frac{1}{y} \rceil}{1 + \lceil \frac{1}{y} \rceil}$ if $x, y > 0$ 1 if x or $y \leq 0$	54	23	8	7	207	65	10	10	458	132	10	11
$(x + y) \left(\frac{\lceil \frac{1}{x} \rceil}{1 + \lceil \frac{1}{x} \rceil} + \frac{\lceil \frac{1}{y} \rceil}{1 + \lceil \frac{1}{y} \rceil} \right)$ if $x, y > 0$, 1 if x or $y \leq 0$	56	23	18	7	211	65	25	10	469	130	31	12
$\exp(\lceil \frac{1}{x} \rceil / (1. + \lceil \frac{1}{x} \rceil)) \left\lceil \frac{1}{\sqrt{y}} \right\rceil + y$ if $x, y > 0$, 1 if x or $y \leq 0$	84	31	14	12	324	94	20	19	698	194	25	27

TABLE 7.2
Number of PCG iterations in the case $k = 2,\ m_1 = m_2 = 2$.

ETNA
Kent State University
etna@mcs.kent.edu

$k = 3\ m = 2$	$n_1 + 1 = n_2 + 1$											
	10				20				30			
$a(x, y)$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$	$\hat{D}$	IC	$\hat{\Delta}$	$\hat{P}$
$1 + x + y$	63	31	13	4	280	142	14	4	795	380	15	4
$\exp(x + y)$	62	31	20	3	276	142	24	4	796	377	25	4
$\sin^2(7(x + y)) + 1$	63	31	10	15	284	142	12	22	807	380	12	26
$1 + \sqrt{x + y}$, 1 if $x + y < 0$	63	32	10	4	281	142	11	4	803	381	11	4
$\|x - \frac{1}{2}\| + \|y - \frac{1}{2}\| + \frac{1}{2}$	16	30	11	6	76	131	14	8	203	362	15	10
$x + y$	63	32	24	6	286	145	36	6	827	382	46	6
$(x + y)^2$	64	31	44	6	289	145	99	7	834	383	157	7
$(x + y)^3$	65	31	70	6	293	147	236	7	846	382	465	7
$(x + y)^4$	64	30	92	6	293	146	467	6	847	382	–	7
$\|x - \frac{1}{2}\| + \|y - \frac{1}{2}\|$	16	30	14	8	76	140	26	14	199	375	34	18
$(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2$	16	28	15	8	75	138	48	13	197	372	79	16
$\exp(x + y)$ if $x + y \leq \frac{2}{3}$ $4 + x + y$ if $x + y > \frac{2}{3}$	63	32	18	10	285	150	21	19	823	390	22	32
$\exp(x + y)$ if $x + y \leq \frac{2}{3}$ $2 - (x + y)$ if $x + y > \frac{2}{3}$	63	30	23	8	289	150	36	14	824	390	46	21
$\left\lceil \frac{1}{\sqrt{x}} \right\rceil + \left\lceil \frac{1}{\sqrt{y}} \right\rceil$ if $x, y > 0$ 1 if x or $y \leq 0$	63	33	10	9	284	151	15	18	822	397	18	30
$\frac{\lceil \frac{1}{x} \rceil}{1+\lceil \frac{1}{x} \rceil} + \frac{\lceil \frac{1}{y} \rceil}{1+\lceil \frac{1}{y} \rceil}$ if $x, y > 0$ 1 if x or $y \leq 0$	62	33	8	7	271	151	16	16	804	397	10	16
$(x + y)\left(\frac{\lceil \frac{1}{x} \rceil}{1+\lceil \frac{1}{x} \rceil} + \frac{\lceil \frac{1}{y} \rceil}{1+\lceil \frac{1}{y} \rceil}\right)$ if $x, y > 0$, 1 if x or $y \leq 0$	63	31	18	7	271	145	37	17	823	385	34	17
$\exp(\lceil \frac{1}{x} \rceil / (1. + \lceil \frac{1}{x} \rceil)) \left\lceil \frac{1}{\sqrt{y}} \right\rceil + y$ if $x, y > 0$, 1 if x or $y \leq 0$	108	33	13	11	496	151	26	27	–	397	24	44

TABLE 7.3
Number of PCG iterations in the case $k = 3,\ m_1 = m_2 = 2$.

	$n_1+1=n_2+1$		
$k=1\ m=3$	10	20	30
$1+x+y$	0	0	0
$\exp(x+y)$	0	0	0
$\sin^2(7(x+y))+1$	17 (8) 17%	28 (12) 7%	28 (12) 3.1%
$1+\sqrt{x+y}$, 1 if $x+y<0$	0	0	0
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert+\frac{1}{2}$	1 1%	1 0.25%	1 0.1%
$x+y$	0	0	0
$(x+y)^2$	0	0	0
$(x+y)^3$	1 1%	1 0.25%	1 0.1%
$(x+y)^4$	2 2%	4 1%	4 0.4%
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert$	6 (1) 6%	10 (1) 2.5%	12 (3) 1.3%
$(x-\frac{1}{2})^2+(y-\frac{1}{2})^2$	7 (1) 7%	10 (1) 2.5%	10 (1) 1.1%
$\exp(x+y)$ if $x+y\leq\frac{2}{3}$ $4+x+y$ if $x+y>\frac{2}{3}$	7 (4) 7%	15 (8) 3.75%	22 (11) 2.4%
$\exp(x+y)$ if $x+y\leq\frac{2}{3}$ $2-(x+y)$ if $x+y>\frac{2}{3}$	2 (1) 2%	8 (4) 2%	12 (6) 1.3%
$\left\lceil\frac{1}{\sqrt{x}}\right\rceil+\left\lceil\frac{1}{\sqrt{y}}\right\rceil$ if $x,y>0$ 1 if x or $y\leq 0$	2 2%	10 (2) 2.5%	17 (1) 1.8%
$\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}$ if $x,y>0$ 1 if x or $y\leq 0$	0 0%	3 0.75%	0 0%
$(x+y)\left(\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}\right)$ if $x,y>0$ 1 if x or $y\leq 0$	0 0%	3 0.75%	0 0%
$\exp(\lceil\frac{1}{x}\rceil/(1.+\lceil\frac{1}{x}\rceil)\left\lceil\frac{1}{\sqrt{y}}\right\rceil+y\ x,y>0$ 1 if x or $y\leq 0$	4 4%	14 (4) 3.5%	19 (4) 2.1%

TABLE 7.4
Number of outliers of $\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)$ for $\delta=0.1$ in the case $k=1,\ m_1=m_2=3$.

ETNA
Kent State University
etna@mcs.kent.edu

	$n_1+1=n_2+1$		
$k=2\ m=2$	10	20	30
$1+x+y$	0	0	0
$\exp(x+y)$	0	0	0
$\sin^2(7(x+y))+1$	26 (13) 26%	54 (26) 13.5%	66 (28) 7.3%
$1+\sqrt{x+y}$, 1 if $x+y<0$	0	0	0
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert+\frac{1}{2}$	4 4%	7 (1) 1.75%	7 (1) 0.7%
$x+y$	0	0	0
$(x+y)^2$	1 (1) 1%	1 (1) 0.25%	1 (1) 0.1%
$(x+y)^3$	0	0	0
$(x+y)^4$	0	0	0
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert$	9 (3) 9%	19 (5) 4.75%	26 (8) 2.8%
$(x-\frac{1}{2})^2+(y-\frac{1}{2})^2$	12 (4) 12%	16 (4) 4%	20 (5) 2.2%
$\exp(x+y)$ if $x+y\le\frac{2}{3}$ $4+x+y$ if $x+y>\frac{2}{3}$	10 (7) 10%	23 (13) 5.75%	37 (22) 4.1%
$\exp(x+y)$ if $x+y\le\frac{2}{3}$ $2-(x+y)$ if $x+y>\frac{2}{3}$	5 (3) 5%	12 (7) 3%	17 (9) 1.8%
$\left\lceil\frac{1}{\sqrt{x}}\right\rceil+\left\lceil\frac{1}{\sqrt{y}}\right\rceil$ if $x,y>0$ 1 if x or $y\le 0$	12 (9) 12%	39 (27) 9.75%	60 (36) 6.6%
$\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}$ if $x,y>0$ 1 if x or $y\le 0$	6 (5) 6%	17 (14) 4.25%	28 (22) 3.1%
$(x+y)\left(\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}\right)$ if $x,y>0$ 1 if x or $y\le 0$	4 (1) 4%	14 (6) 3.5%	25 (9) 2.7%
$\exp(\lceil\frac{1}{x}\rceil/(1.+\lceil\frac{1}{x}\rceil)\left\lceil\frac{1}{\sqrt{y}}\right\rceil+y$ if $x,y>0$ 1 if x or $y\le 0$	12 (10) 12%	39 (28) 9.75%	66 (40) 7.3%

TABLE 7.5
Number of outliers of $\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)$ *for* $\delta=0.1$ *in the case* $k=2,\ m_1=m_2=2$.

ETNA
Kent State University
etna@mcs.kent.edu

	$n_1+1=n_2+1$		
$k=3$ $m=2$	10	20	30
$1+x+y$	0	0	0
$\exp(x+y)$	0	0	0
$\sin^2(7(x+y))+1$	38 (10) 38%	81 (39) 20.5%	109 (53) 12.1%
$1+\sqrt{x+y}$, 1 if $x+y<0$	0	0	0
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert+\frac{1}{2}$	7 (1) 7%	13 (3) 3.25%	20 (6) 2.22%
$x+y$	1 (1) 1%	1 (1) 0.25%	1 (1) 0.11%
$(x+y)^2$	2 (2) 2%	3 (3) 0.75%	4 (4) 0.44%
$(x+y)^3$	2 (2) 2%	3 (3) 0.75%	4 (4) 0.44%
$(x+y)^4$	1 (1) 1%	2 (2) 0.5%	3 (3) 0.3%
$\lvert x-\frac{1}{2}\rvert+\lvert y-\frac{1}{2}\rvert$	15 (5) 15%	33 (13) 8.25%	49 (21) 5.44%
$(x-\frac{1}{2})^2+(y-\frac{1}{2})^2$	15 (5) 15%	33 (13) 8.25%	49 (21) 5.44%
$\exp(x+y)$ if $x+y\le\frac{2}{3}$ $4+x+y$ if $x+y>\frac{2}{3}$	13 (8) 13%	32 (17) 8%	50 (26) 5.55%
$\exp(x+y)$ if $x+y\le\frac{2}{3}$ $2-(x+y)$ if $x+y>\frac{2}{3}$	5 (4) 5%	18 (10) 4.5%	31 (16) 3.44%
$\left\lceil\frac{1}{\sqrt{x}}\right\rceil+\left\lceil\frac{1}{\sqrt{y}}\right\rceil$ if $x,y>0$ 1 if x or $y\le 0$	10 (5) 10%	39 (21) 9.75%	86 (42) 9.55%
$\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}$ if $x,y>0$ 1 if x or $y\le 0$	4 (3) 4%	21 (10) 5.2%	35 (21) 3.88%
$(x+y)\left(\frac{\lceil\frac{1}{x}\rceil}{1+\lceil\frac{1}{x}\rceil}+\frac{\lceil\frac{1}{y}\rceil}{1+\lceil\frac{1}{y}\rceil}\right)$ if $x,y>0$ 1 if x or $y\le 0$	3 (1) 3%	23 (5) 5.75%	37 (16) 4.11%
$\exp(\lceil\frac{1}{x}\rceil/(1.+\lceil\frac{1}{x}\rceil)\left\lceil\frac{1}{\sqrt{y}}\right\rceil+y$ if $x,y>0$ 1 if x or $y\le 0$	12 (7) 12%	44 (24) 11%	86 (45) 9.55%

TABLE 7.6

Number of outliers of $\hat{P}_n^{-1}(a)\hat{A}_n(a,m_1,m_2,k)$ for $\delta=0.1$ in the case $k=3,\ m_1=m_2=2$.

	$n_1+1=n_2+1$					
	16	32	64	128	256	512
$k=1, m=1$	1	8	8	8	8	8
$k=2, m=1$	1	13	13	13	13	13
$k=2, m=2$	1	16	16	16	16	16
$k=3, m=2$	1	24	25	25	25	25

TABLE 7.7

Number of multigrid (C-cycle) iterations.

$k=1, m_1=m_2=1$	$xe_i=1$						xe_i random					
	$n_1+1=n_2+1$						$n_1+1=n_2+1$					
	16	32	64	128	256	512	16	32	64	128	256	512
$a(x,y)=1+x+y$												
Q	3	3	3	3	3	3	3	3	2	2	2	2
L	3	3	3	3	3	3	3	3	2	2	2	2
T	3	3	3	3	3	3	3	3	2	2	2	2
$a(x,y)=\sin^2(7(x+y))+1$												
Q	10	10	10	9	9	8	9	9	9	8	7	6
L	9	9	9	8	8	8	8	8	8	7	7	6
T	9	9	9	9	8	8	9	9	8	7	7	6
$a(x,y)=1-x+y$												
Q	4	4	4	4	4	3	4	4	3	3	3	2
L	4	4	4	4	4	3	4	4	3	3	3	2
T	4	4	4	4	4	3	4	4	3	3	3	2
$a(x,y)=(1-x+y)^2$												
Q	2	2	2	2	1	1	2	2	2	1	1	1
L	2	2	2	2	1	1	2	2	2	1	1	1
T	2	2	2	2	1	1	2	2	2	1	1	1
$a(x,y)=\exp(x+y)$ $\cdot\mathrm{Ch}_{\{x+y\le 2/3\}}$ $+(2-(x+y))$ $\cdot\mathrm{Ch}_{\{x+y>2/3\}}$												
Q	7	8	9	10	13	15	7	7	9	10	12	15
L	5	5	7	8	9	10	5	5	7	7	9	10
T	7	7	9	10	12	14	6	7	8	9	10	11

TABLE 7.8

Number of PCG iterations in the case $k=1$, $m_1=m_2=1$, square, L and T shaped domains, with preconditioner $\hat{P}$, exact solution xe of all ones and random.

$k=1, m_1=m_2=1$	$xe_i=1$						xe_i random					
	$n_1+1=n_2+1$						$n_1+1=n_2+1$					
	16	32	64	128	256	512	16	32	64	128	256	512
$a(x,y)=1+x+y$												
Q	16	28	53	100	196	371	16	27	48	85	155	294
L	13	23	42	80	154	300	14	24	43	74	133	251
T	15	27	50	95	186	362	14	26	47	87	158	283
$a(x,y)=\sin^2(7(x+y))+1$												
Q	16	29	54	104	202	391	17	30	51	95	165	300
L	13	23	44	83	159	303	14	26	43	74	139	265
T	15	27	52	99	192	367	15	26	47	89	167	290
$a(x,y)=1-x+y$												
Q	17	28	53	102	199	387	16	27	49	92	165	302
L	14	27	50	96	186	361	14	24	45	84	159	290
T	15	27	51	96	188	366	14	25	47	87	164	300
$a(x,y)=(1-x+y)^2$												
Q	16	28	52	100	182	353	16	26	49	86	161	302
L	14	26	49	96	186	359	13	25	46	86	160	295
T	14	26	50	95	186	363	14	25	47	88	163	307
$a(x,y)=\exp(x+y)$ $\cdot\mathrm{Ch}_{\{x+y\le 2/3\}}$ $+(2-(x+y))$ $\cdot\mathrm{Ch}_{\{x+y>2/3\}}$												
Q	16	28	53	102	199	385	16	29	49	93	171	324
L	13	23	44	84	162	309	14	24	44	76	143	267
T	15	27	51	98	189	370	14	26	47	89	165	298

TABLE 7.9

Number of PCG iterations in the case $k=1$, $m_1=m_2=1$, square, L and T shaped domains, with IC preconditioner, exact solution xe of all ones and random.