Electronic Journal of Linear Algebra

Volume 34 *Volume 34 (2018)* Article 41

2018

Asymptotic results on the condition number of FD matrices approximating semi-elliptic PDEs

Paris Vassalos
Athens University of Economics and Business, pvassal@aueb.gr

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

ASYMPTOTIC RESULTS ON THE CONDITION NUMBER OF FD MATRICES APPROXIMATING SEMI-ELLIPTIC PDES*

PARIS VASSALOS†

Abstract. This work studies the asymptotic behavior of the spectral condition number of the matrices A_{nn} arising from the discretization of semi-elliptic partial differential equations of the form

$$-\left(a(x,y)u_{xx}+b(x,y)u_{yy}\right)=f(x,y),$$

on the square $\Omega=(0,1)^2$, with Dirichlet boundary conditions, where the smooth enough variable coefficients $a(x,y), b(x,y)$ are nonnegative functions on $\overline{\Omega}$ with zeros. In the case of coefficient functions with a single and common zero, it is discovered that apart from the minimum order of the zero also the direction that it occurs is of great importance for the characterization of the growth of the condition number of A_{nn}. On the contrary, when the coefficient functions have non intersecting zeros, it is proved that independently of the order their zeros, and their positions, the condition number of A_{nn} behaves asymptotically exactly as in the case of strictly elliptic differential equations, i.e., it grows asymptotically as n^2. Finally, the more complicated case of coefficient functions having curves of roots is considered, and conjectures for future work are given. In conclusion, several experiments are presented that numerically confirm the developed theoretical analysis.

Key words. Finite differences, GLT sequences, Semi elliptic PDEs, Spectral condition number.

AMS subject classifications. 65N22, 65F10, 15A12.

1. Introduction.

A classical but still pertinent problem, appearing in many applications, is the numerical solution of partial differential equations (PDEs) of the form

$$-\frac{\partial}{\partial x}\left(a(x,y)\frac{\partial}{\partial x}u(x,y)\right)-\frac{\partial}{\partial y}\left(b(x,y)\frac{\partial}{\partial y}u(x,y)\right)=f(x,y), \tag{1.1}$$

with Dirichlet boundary conditions on the domain $\Omega=(0,1)^2$. Important examples of PDEs with such a nonnegative canonical form, which appear in the study of transonic flow, are the families of equations of the so called Keldysh/Tricomi type (see [7, 8]), which are given by

$$y^{2m+1}u_{xx}+u_{yy}=0$$

and

$$u_{xx}+y^{2m+1}u_{yy}=0,$$

respectively. Also the Laplace-Beltrami equation, defined as

$$(1-x^2)u_{xx}+(1-y^2)u_{yy}=0,$$

belongs to the considered class. The latter has applications in differential geometry and specifically to isometric embedding of Riemannian manifolds. We can indicate other research fields where PDEs with a

*Received by the editors on August 11, 2018. Accepted for publication on September 29, 2018. Handling Editor: Panayiotis Psarrakos.

†Department of Informatics, Athens University of Economics and Business, Patision 76, 10434, Athens, Greece (pvassal@aueb.gr). Author greatly acknowledges financial support received from the Research Center of the Athens University of Economics and Business in the framework of the project entitled "Original Scientific Publications" (EP-2790-01).

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

nonnegative characteristic form are involved: mathematical biology, physical models, chemistry and mathematical finance. Usually, the strict ellipticity is lost due to some isolated zeros of the coefficient functions usually located at the boundary $\partial\Omega$ of the definition domain Ω (see [2],[9], and references therein). Using the centered finite difference (FD) formula of precision order two and minimal bandwidth for discretizing (1.1), we obtain a 5-points formula, with respect to the x axis and to the y axis. The resulting linear system is

$$A_{nn}\mathbf{x} = \mathbf{b}, \tag{1.2}$$

where A_{nn} is a symmetric positive definite two-level banded matrix. Although the latter does not have the multilevel Toeplitz structure, the associated matrix sequence belongs to the wider class of Generalized Locally Toeplitz (GLT) sequences of matrices (see [15], [13], and for more details [3]). In the framework of this beautiful and powerful theory, many properties concerning the spectral distribution of these sequences of matrices can be proved.

Despite the block band form of A_{nn}, the associated system (1.1) is not trivial to solve. In the 1D case there are many optimal direct solvers, with Thomas being the best known, but in multidimensional settings things are completely changed. The reason relies in the difficulty of exploiting the inner structure of block band matrices and, as a consequence, to benefit from their inner sparsity. Adding to the above reasoning, the inevitably sequential nature of such methods makes direct methods not an ideal choice for multilevel linear systems, as those appearing in (1.2).

Iterative techniques, like SOR, ADI, Chebyshev or the Conjugate Gradient method, take advantage of the sparsity of A_{nn}, but their convergence features depend mainly on the condition number of the matrix which, as we will show for our problem, grows at least as $O(n^2)$. To alleviate this problem, preconditioning is usually the first option. Popular preconditioners arise from the incomplete Cholesky factorization or matrices belonging to some special class, such as trigonometric matrix algebras (usually circulant or τ), and Toeplitz plus diagonal [12], or band plus algebra matrices [6]. On the other hand, full-multigrid (FMG) techniques are probably the fastest known solver for discretized elliptic PDEs. However, attempts to extend the same techniques to its degenerate cases, for example the Keldysh equation, have met with a more limited success. We mention that multigrid methods can be used also within the PCG method in the "inversion" of the preconditioner in each step, if the latter is chosen to belong to some specific class of matrices [1].

For all the above methods and for any iterative methods that will be developed in the future, the crucial information that is needed in order to establish fast convergence is the knowledge of the spectrum of A_{nn} and, especially, the source of ill conditioning. In this work, we extend the theory developed in [4] concerning the asymptotic behavior of the condition number of A_n, i.e., the analogue of the 1D matrix, in the multidimensional case. We note that in the aforementioned paper, it was mentioned that the main goal of that work was to develop the tools and to set the foundations for the interesting two-dimensional, or, more generally, to the multi-dimensional, case. Our results complete the spectral picture described by the GLT theory. Specifically, adopting a different point of view, we show again that there are two sources of ill conditioning, one coming from the discretization of differential operator via the FD method and the other from the sampling of the functions $a(x, y)$ and $b(x, y)$ near their zeros. Moreover, our analysis leads to the interesting corollary that the two sources of ill-conditioning, i.e., the low frequencies coming from the constant-coefficient Laplacian, and the space spanned by few canonical vectors related to the position of the zeros of the coefficient functions, do not in general interfere, and thus, also the extreme eigenvalues of the aforementioned matrices behave smoothly and according to what we expect from the GLT theory.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

These results can be directly used in the development of efficient preconditioners when the Preconditioned Conjugate Gradient (PCG) method is used and in the construction of optimal multigrid methods.

2. Preliminaries. Throughout our work, we will extensively use the following notations and definitions.

Let f, g real valued functions both defined on some unbounded subset D of $\mathbb{R}^+$.

- $f(x) = O(g(x))$ as $x \to \infty$, if and only if there exists a pure positive constant M such that $|f(x)| \leq Mg(x)$, for all sufficient large values of x.
- $f(x) = o(g(x))$ if and only if for every positive constant ϵ, exists positive constant N, such that $|f(x)| \leq \epsilon g(x)$, for all $x > N$. In other words, if $g(x)$ is nonzero (or at least becomes nonzero beyond a certain point), it holds $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.
- $f(x) = \Omega(g(x))$ is the negation of $f(x) = o(g(x))$, i.e., if and only if $\lim_{x\to\infty} \frac{f(x)}{g(x)} > 0$.
- $f(x) \sim g(x)$ if and only if $f(x) = O(g(x))$ and $g(x) = O(f(x))$.

Definition 2.1. A nonnegative function f has a zero at (x_0, y_0) if for every neighborhood I of (x_0, y_0) we have $\inf_{(x,y)\in I} f(x, y) = 0$. Moreover, a set Z of the zeros of $f \geq 0$ is called *isolated* if there exists a neighborhood J of Z such that for every neighborhood $I \subset J, I \neq J$ of Z we have $\inf_{(x,y)\in J\setminus I} f(x, y) > 0$.

Definition 2.2. A nonnegative function f has a zero of order $\alpha > 0$ at (x_0, y_0), if for some norm $\|\cdot\|$, there exist pure positive constants c, C such that

$$c\,\|(x, y) - (x_0, y_0)\|^\alpha \leq f(x, y) \leq C\,\|(x, y) - (x_0, y_0)\|^\alpha,$$

i.e., $f(x, y) \sim \|(x, y) - (x_0, y_0)\|^\alpha$ in a suitable neighborhood of (x_0, y_0).

We mention that in [5], a more general definition about the order of a zero was given. However, for our needs, Definition 2.2 is sufficient.

We briefly introduce the 1D instance of the problem, i.e.,

$$\begin{cases} -\frac{d}{dx}\left(a(x)\frac{d}{dx}u(x)\right) = f(x) & \text{on } I = (0, 1), \\ \text{Dirichlet B.C.} & \text{on } \partial I, \end{cases} \tag{2.3}$$

since results referring to this case, will be used in our generalization process.

Using centered finite differences, of minimal bandwidth and precision order two, and stepsize $h = (n+1)^{-1}$ on the grid-points $x_0 = 0, x_1, x_2, \ldots, x_n, x_{n+1} = 1$, we arrive to the $n \times n$ linear system $A_n(a)x = b$ whose coefficient matrix has the form

$$A_n(a) = \begin{bmatrix} a_{\frac{1}{2}} + a_{\frac{3}{2}} & -a_{\frac{3}{2}} & & & & & \\ -a_{\frac{3}{2}} & a_{\frac{3}{2}} + a_{\frac{5}{2}} & -a_{\frac{5}{2}} & & & & \\ & -a_{\frac{5}{2}} & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & -a_{n-\frac{1}{2}} & \\ & & & & -a_{n-\frac{1}{2}} & a_{n-\frac{1}{2}} + a_{n+\frac{1}{2}} \end{bmatrix}, \tag{2.4}$$

where $a_{i-\frac{1}{2}}$, $i = 1(1)n + 1$ are the sampling values of the bounded smooth enough function $a(x)$ on the points $x_i - \frac{h}{2}, i = 1(1)n + 1$, of the interval $(0, 1)$. Considering $A(\cdot)$ as an operator from a suitable function space $\mathcal{J}([0, 1])$ into the set of symmetric matrices $\mathrm{Sym}(\mathbb{R}^{n\times n})$, defines a linear positive operator, i.e.,

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

- $A_n(c_1a + c_2b) = c_1A_n(a) + c_2A_n(b)$,
- $A_n(a)$ is nonnegative definite, for every nonnegative function $a \in \mathcal{J}([0,1])$.

An immediate consequence is that $A_n(\cdot)$ is also a monotone operator, that is, $a \geq \tilde{a}$ implies $A_n(a) - A_n(\tilde{a}) > 0$, and thus, $A_n(a) > A_n(\tilde{a})$. Moreover, the j-th eigenvalue of $A_n(a)$ is bounded by the j-th eigenvalue of $A_n(\tilde{a})$, where the eigenvalues of each matrix are ordered non decreasingly (see [11], [14], for a general discussion and several results on matrix-valued linear positive operators).

In [4], the asymptotic behavior of the condition number of $A_n(a)$ was studied, and the main result presented there is summarized in the next theorem.

THEOREM 2.3. *Let $\{A_n(a)\}_n$, $A_n(a) \in \mathbb{R}^{n\times n}$, be the sequence of matrices derived from the discretization of the semi-elliptic differential equation (2.3) with the bounded coefficient function $a(x)$ having a unique root at $x_0 \in I = [0,1]$ of order α, i.e., $a(x) \sim |x - x_0|^\alpha$ on I. Then, for the spectral condition number $\kappa(A_n(a)) \triangleq \|A_n(a)\|_2\|A_n^{-1}(a)\|_2$ of the matrix $A_n(a)$, which coincides in order with the spectral radius $\rho(A_n^{-1}(a))$ of $A_n^{-1}(a)$, it holds that*

$$\kappa(A_n(a)) \sim \rho(A_n^{-1}(a)) \sim \begin{cases} n^2, & 0 \leq \alpha < 2, \\ O(n^2\log(n)) \cap \Omega(n^2), & \alpha = 2, \\ n^\alpha, & \alpha > 2. \end{cases} \tag{2.5}$$

3. Main analysis. We return to the 2D setting, i.e., to the problem described by (1.1), where we have assumed Dirichlet boundary conditions on the domain $\Omega = [0,1]^2$ and both the coefficient functions are bounded, piecewise continuous and nonnegative on it. Discretizing the rectangular boundary of Ω in n and m nodes in the x and y direction, respectively, and using centered FDs with step-size $h = \frac{1}{n+1}$ in the x direction and $k = \frac{1}{m+1}$ in y direction, we arrive at the following set of equations:

$$-b_{i,j-\frac{1}{2}}u_{i,j-1} - a_{i-\frac{1}{2},j}u_{i-1,j} + (a_{i+\frac{1}{2},j} + a_{i-\frac{1}{2},j} + b_{i,j+\frac{1}{2}} + b_{i,j-\frac{1}{2}})u_{i,j} - a_{i+\frac{1}{2},j}u_{i+1,j} - b_{i,j+\frac{1}{2}}u_{i,j+1} = f_{i,j},$$

for $i = 1(1)n$ and $j = 1(1)m$, or, in matrix vector notation, to the system

$$A_{nm}(a,b)\mathbf{u}_{nm} = \mathbf{f}_{nm},$$

where $\mathbf{f}_{nm}$ depends on the nm vector containing the discretized values of f on the grid and $A_{nm}(a,b) = A_{nm}\left(a(x,y), b(x,y)\right)$. Since, in practice we always have $n \sim m$, for simplification in the computations, and without loss of generality, for the rest of the manuscript we will assume $n = m$. Obviously, the coefficient matrix $A_{nn}(a,b)$ is block tridiagonal with the diagonal blocks being tridiagonal matrices and the upper and lower to these blocks being diagonal matrices containing the sampling values $b_{i,j-\frac{1}{2}}$ and $b_{i,j+\frac{1}{2}}$, respectively. Similarly to 1D case, $A_{nn}(\cdot,\cdot)$ defines a linear positive operator from a suitable function space $\mathcal{J}$ into the space $\mathrm{Sym}(\mathbb{R}^{n^2\times n^2})$, i.e., is linear with respect to its arguments and positive in the sense that for every $a, b \in \mathcal{J}(\Omega)$ with $a, b \geq 0$, $A_{nn}(a,b)$ is symmetric nonnegative definite. In addition, the monotonicity implies $A_{nn}(\tilde{a},\tilde{b}) \leq A_{nn}(a,b)$, whenever $a \leq \tilde{a},\ b \leq \tilde{b}$. An immediate consequence of the latter is that

$$\|A_{nn}(a,b)\|_\infty \leq \max\{\|a\|_\infty, \|b\|_\infty\} \cdot \|L_{nn}\|_\infty \leq 8\max\{\|a\|_\infty, \|b\|_\infty\}, \tag{3.6}$$

where L_{nn} is the well known 2D Laplacian matrix that can be written as

$$L_{nn} = I_n \otimes L_n + L_n \otimes I_n,$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

with $L_n = \text{trid}[-1\ 2\ -1]$ being the Laplace matrix, i.e., the tridiagonal Toeplitz matrix having 2 in the main diagonal and -1 in the superdiagonal and subdiagonal, respectively. With I_n we denote the identity matrix of dimension n. Since the functions $a(x,y), b(x,y)$ are assumed to be bounded on the domain Ω, then, from (3.6), also the maximum eigenvalue $\lambda_{\max}(A_{nn}(a,b))$ is asymptotically bounded. Accordingly, the spectral condition number of $A_{nn}(a,b)$, depends only on the behavior of $\lambda_{\min}(A_{nn}(a,b))$, whose study is our main goal.

There exists a specific selection of coefficient functions that, under some assumptions concerning their zeros, significantly simplifies the analysis of the general problem and reveals some counterintuitive situations. For that, we choose

$$\hat{a}(x,y) = |x-x_0|^{\alpha} + |y-y_o|^{\beta}, \quad \hat{b}(x,y) = |x-x_1|^{\gamma} + |y-y_o|^{\delta}, \tag{3.7}$$

where $\alpha, \beta, \gamma, \delta \in \mathbb{R}_0^+$. Taking into account the specific structure of $A_{nn}(\hat{a}, \hat{b})$, it can be shown that for these coefficient functions, the matrix can be decomposed in the following form

$$\begin{aligned} \hat{A}_{nn} = A_{nn}(\hat{a}(x,y), \hat{b}(x,y)) = I_n \otimes A_n(|x-x_0|^{\alpha}) + D_n(|y-y_0|^{\beta}) \otimes L_n \\ + L_n \otimes D_n(|x-x_1|^{\gamma}) + A_n(|y-y_1|^{\delta}) \otimes I_n, \end{aligned} \tag{3.8}$$

where $A_n(\cdot)$ is defined in (2.4), and $D_n(x^{\gamma})$, $D_n(y^{\beta})$ are the diagonal matrices formed by the values of the functions x^{γ} and y^{β} on the points $\frac{i}{n+1}$, $i = 1(1)n$, respectively. In addition, (3.8) shows the exact connection between the 2D case and 1D case, and uncovers the possible influence that each term can have separately. A concrete application is the following lemma.

LEMMA 3.1. *Let $a(x,y), b(x,y)$ be nonnegative, piecewise continuous and bounded functions of Ω. If at least one of them is strictly positive on Ω, then, no matter how many zeros the other has or which are their orders, the minimum eigenvalue of $A_{nn}(a,b)$, behaves as n^{-2}.*

Proof. Assuming that $a(x,y)$ is the strictly positive function, then there exists a universal positive constant c such that $a(x,y) \geq c > 0$. Moreover, from (3.8),

$$A_{nn}(a,b) > I_n \otimes A_n(c) + cI_n \otimes L_n > cI_n \otimes L_n,$$

while if $b(x,y) \geq c > 0$,

$$A_{nn}(a,b) > cL_n \otimes I_n + A_n(c) \otimes I_n > cL_n \otimes I_n.$$

Taking the Rayleigh quotient on the above relationships, and using the well known property connecting the eigenvalues of the Kronecker product with these of each part, we arrive at

$$\lambda_{\min}(A_{nn}(a,b)) \geq c\lambda_{\min}(L_n) = c\left(\sin^2\left(\frac{\pi}{2(n+1)}\right)\right) \sim \frac{1}{n^2}. \tag{3.9}$$

On the other hand,

$$A_{nn}(a,b) \leq A_{nn}(\|a\|_{\infty}, \|b\|_{\infty}) \leq \max\{\|a\|_{\infty}, \|b\|_{\infty}\} A_{nn}, (1,1) = M \cdot L_{nn},$$

where $M = \max\{\|a\|_{\infty}, \|b\|_{\infty}\}$. Using the same reasoning as before, we have

$$\lambda_{\min}(A_{nn}(a,b)) \leq \sin^2\left(\frac{\pi}{2(n+1)}\right). \tag{3.10}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

Thus, from (3.9) and (3.10), whenever at least one of the coefficient functions is uniformly bounded below by a positive constant, then

$$\lambda_{\min}\left(A_{nn}(a,b)\right) \sim \frac{1}{n^2}. \qquad \square$$

COROLLARY 3.2. *A concrete application of Lemma 3.1 is the case of the Keldysh class of PDEs that we presented in the introductory section. As a result, the FD techniques applied to this kind of problem will lead to a coefficient matrix having spectral condition number behaving exactly as the constant Laplace 2D analogue, i.e., as* n^2.

The following theorem concerns the case where both the coefficient functions have a single and common zero at the point $(x_0, y_0) \in \Omega$.

THEOREM 3.3. *Assume that the coefficient functions* $a(x,y)$,$b(x,y)$ *have a single and common zero at the point* (x_0, y_0)*, of orders* α *in the* x *direction and* β *in the* y *direction, respectively, i.e., there exist* $c_i > 0,\ i = 1(1)4$, *such that*

$$c_1\left(|x-x_0|^{\alpha}\right) \le a(x,y) \le c_2\left(|x-x_0|^{\alpha}\right),$$

and

$$c_3\left(|y-y_0|^{\beta}\right) \le b(x,y) \le c_4\left(|y-y_0|^{\beta}\right).$$

By defining $\rho = \min\{\alpha, \beta\}$, *we have*

$$\kappa_2(A_{nn}) \sim \begin{cases} n^2, & 0 \le \rho < 2, \\ \mathcal{O}(n^2\log(n)) \cap \Omega(n^2), & \rho = 2, \\ n^{\rho}, & \rho \ge 2. \end{cases}$$

Proof. The key points of our proof are the monotonicity of the operator $A_{nn}(\cdot,\cdot)$ and the use of proper inequalities on the coefficients $a(x,y)$ and $b(x,y)$. We will study in details the case where $\rho = \alpha$, and the other case is similarly treated. From the monotonicity of the operator and the assumptions about the partial minimum orders of the zero we have that

$$cA_{nn}\left(\hat{a}(x,y), \hat{b}(x,y)\right) \le A_{nn}\left(a(x,y), b(x,y)\right) \le CA_{nn}\left(\hat{a}(x,y), \hat{b}(x,y)\right),$$

where $c = \min\{c_1, c_3\}$ and $C = \max\{c_2, c_4\}$. Thus, taking advantage of the above equivalently, is sufficient to study the behavior of the minimum eigenvalue of $A_{nn}(\hat{a}, \hat{b})$.

Omitting in (3.8) the positive terms contain the matrix L_n, we have

$$A_{nn}(\hat{a}, \hat{b}) > I_n \otimes A_n(|x-x_0|^{\alpha}) + A_n(|y-y_0|^{\beta}) \otimes I_n > I_n \otimes A_n(|x-x_0|^{\alpha}).$$

From [4] we know that the minimum eigenvalue of $A_n(|x-x_0|^{\alpha})$ tends to zero as (2.5) predicts. As a consequence, the minimum eigenvalue of $A_{nn}(a,b)$ cannot tend to zero asymptotical faster than the theorem predicts.

On the other hand, using Rayleigh quotients we can show that the minimum eigenvalue of the matrix $A_{nn}(\hat{a}, \hat{b})$, and so this of $A_{nn}(a,b)$, tends to zero at least as fast as $O(n^{-\alpha})$. We break down the complete analysis in two cases: $\alpha \ge 2$ and $\alpha < 2$.

In the first case, we define

$$\bar{a}(x,y) \equiv |x-x_0|^{\alpha} + |y-y_0|^{\alpha}.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

Obviously $a(x,y) \leq C_1\bar{a}(x,y)$ and $b(x,y) \leq C_2\bar{a}(x,y)$, where C_1, C_2 are properly chosen universal positive constants. Accordingly, we choose the normalized vector $e_{kl} = e_k \otimes e_l$ where e_k, e_l are the k-th and l-th columns of $n \times n$ identity matrix, respectively. The indexes k, l are given by

$$k = \arg\min_i |x_i - x_0|, \quad l = \arg\min_j |y_i - y_0|,$$

with x_i, y_j being the discretization points in x and y direction, respectively. The idea behind this selection is to point at the minimum value of the diagonal of A_{nn}. Hence, using this vector in the Rayleigh quotient and $C = \max\{C_1, C_2\}$, we obtain that

$$\begin{aligned} \lambda_{\min}(A_{nn}(a,b)) &\leq e_{kl}^T A_{nn}(a,b) e_{kl} \leq C\left(e_{kl}^T A_{nn}(\hat{a},\hat{a}) e_{kl}\right) \\ &= 2C\left(\hat{a}(x_{k-\frac{1}{2}}, y_l) + \hat{a}(x_{k+\frac{1}{2}}, y_l)\right). \end{aligned} \tag{3.11}$$

Then, from the way we selected k, l, we obtain that the latter quantity tends to zero as $n^{-\alpha}$.

In the second case, since $\max_{(x,y)\in\Omega} x^\alpha + y^\alpha = 2$,

$$A_{nn}(x^\alpha + y^\alpha, x^\alpha + y^\alpha) \leq 2A_{nn}(1,1) = 2L_{nn}.$$

Since L_{nn} is the 2D Laplacian matrix, the eigenvalues are explicitly known and are given by

$$\lambda_{i,j} = 4\left(\sin^2\left(\frac{\pi i}{2(n+1)}\right) + \sin^2\left(\frac{\pi j}{2(n+1)}\right)\right), \quad i,j = 1(1)n,$$

with the corresponding eigenvectors being

$$l_{i,j} = l_i \otimes l_j, \quad i,j = 1(1)n,$$

where

$$l_k = \sqrt{\frac{2}{n+1}}\left[\sin\left(\frac{k\pi}{n+1}\right)\ \sin\left(\frac{2k\pi}{n+1}\right)\cdots\ \sin\left(\frac{nk\pi}{n+1}\right)\right]^T, \quad k = 1(1)n. \tag{3.12}$$

Defining $c = \max\{\|a\|_\infty, \|\beta\|_\infty\}$, we have

$$\begin{aligned} \lambda_{\min}(A_{nn}(a,b)) &= \min_{z\in\mathbb{R}^{n^2}, \|z\|_2=1} z^T A_{nn}(a,b) z \leq \min_{z\in\mathbb{R}^{n^2}, \|z\|_2=1} c z^T A_{nn}(1,1) z \\ &\leq c\,(l_1 \otimes l_1)^T \cdot A_{nn}(1,1) \cdot l_1 \otimes l_1 = c\lambda_{1,1} \sim \frac{1}{n^2}, \end{aligned}$$

and the proof is completed. ∎

Remark 3.4. The assumptions of Theorem 3.3 exclude cases where the minimum of the order is in y direction of $a(x,y)$ or the x direction of $b(x,y)$. The reason is evident, again, by virtue of (3.8). Specifically, the terms $L_n \otimes D_n(x^\gamma)$ and $D_n(y^\beta) \otimes L_n$ increase the corresponding orders of the zero by 2. Therefore, if the rest of the partial orders are large enough, there exists a possibility that the order of growth of the condition number of $A_{nn}(a,b)$ is greater than the quantity $\min\{\beta, \gamma\}$. An illustrative example describing the above situation is the following: let $a(x,y) = x^{10} + y^1$ and $b(x,y) = x^{10} + y^{10}$, i.e., $\alpha = \gamma = \delta = 10, \beta = 1$. Then, from (3.8) we expect that $\lambda_{\min}(A_{nn}(a,b))$ can tend to zero much faster than the order of the differential operator, which is 2, a speculation numerically confirmed in Section 6. We mention that, in the above example, swapping the value of α with that of β, we can guarantee that the condition number of the new matrix will grow as n^2, since now the assumptions of Theorem 3.3 hold.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

4. $a(x,y)$ **and** $b(x,y)$ **with zeros at different points.** We start this section assuming that the coefficient functions $a(x,y), b(x,y)$ each have a single zero on Ω, but the zeros are different to each other. As we prove in the next theorem, in this case the orders of the zeros of $a(x,y)$ and $b(x,y)$ do not play any role in the asymptotic behavior of the condition number of $A_{nn}(a(x,y),b(x,y))$. In that case, it is the differential operator that always governs the behavior of the minimum eigenvalue of this matrix.

THEOREM 4.1. *Let us assume that the nonnegative coefficient functions $a(x,y)$, $b(x,y)$ each have in Ω a single, but different from each other, zero, at the points (x_0,y_0) and (x_1,y_1), respectively, of any polynomial order. Then, the minimum eigenvalue of $A_{nn}(a,b)$ tends asymptotically to zero as n^{-2}, and thus, the spectral condition number of $A_{nn}(a,b)$ grows as n^2.*

Proof. From the inequality

$$A_{nn}(a,b) \leq \max\{\|a\|_\infty, \|b\|_\infty\} A_{nn}(1,1), \tag{4.13}$$

we immediately have that $\lambda_{\min}(A_{nn}(a,b))$ tends to zero at least as fast as the minimum eigenvalue of 2D Laplacian, i.e., as $O(n^{-2})$.

From the assumptions of the theorem and (3.7), we have that

$$c_1\hat{a}(x,y) \leq a(x,y) \leq C_1\hat{a}(x,y),$$

and

$$c_2\hat{b}(x,y) \leq b(x,y) \leq C_2\hat{b}(x,y).$$

Consequently, from the monotonicity of the operator A_{nn} we have that

$$c\hat{A}_{nn} \equiv cA_{nn}(\hat{a},\hat{b}) \leq A_{nn}(a,b) \leq CA_{nn}(\hat{a},\hat{b}) \equiv C\hat{A}_{nn},$$

where $c = \min\{c_1,c_2\}$, $C = \max\{C_1,C_2\}$. Thus, the behavior of the minimum eigenvalue of $A_{nn}(a,b)$ is similar to that of $A_{nn}(\hat{a},\hat{b})$. Moreover using (3.8), we have

$$\hat{A}_{nn} = I_n \otimes A_n(|x-x^0|^\alpha) + L_n \otimes D_n(|x-x^1|^\gamma) + A_n(|y-y^1|^\delta) \otimes I_n + D_n(|y-y^0|^\beta) \otimes L_n. \tag{4.14}$$

Since the zeros are different from each other, either $x_0 \neq x_1$ or $y_0 \neq y_1$. If $x_0 \neq x_1$, omitting the symmetric and positive definite terms caused by the variable y, we have

$$\hat{A}_{nn} \geq I_n \otimes A_n(|x-x^0|^\alpha) + L_n \otimes D_n(|x-x^1|^\gamma),$$

otherwise, it holds that

$$\hat{A}_{nn} \geq A_n(|y-y^1|^\delta) \otimes I_n + D_n(|y-y^0|^\beta) \otimes L_n.$$

In both cases, the following argument holds unaltered. Thus, we choose to present only the first case, i.e., $x_0 \neq x_1$ and the other case is similarly treated. Without loss of generality, we assume $x^0 < x^1$. Knowing that $\lambda_{\min}(L_n) \sim n^{-2}$, there exist a pure constant $d > 0$ such that

$$L_n \geq \frac{d}{n^2} I_n.$$

Hence,

$$\begin{aligned}
\hat{A}_{nn} &\geq I_n \otimes A_n(|x-x^0|^\alpha) + \frac{d}{n^2} I_n \otimes D_n(|x-x^1|^\gamma) \\
&= I_n \otimes \left(A_n(|x-x^0|^\alpha) + \frac{d}{n^2} D_n(|x-x^1|^\gamma) \right) \\
&\geq I_n \otimes \left(A_n(|x-x^0|^r) + \frac{d}{n^2} D_n(|x-x^1|^r) \right),
\end{aligned}$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

where $r = \max\{\gamma, \alpha\}$. As a result, we concentrate our attention to the study of the asymptotic behavior of the minimum eigenvalue of

$$\hat{A}_n = A_n(|x - x^0|^r) + \frac{1}{n^2} D_n(d|x - x^1|^r) = A_n(a(x)) + \frac{1}{n^2} D_n(b(x)),$$

or, equivalently, to the study of the Rayleigh quotient

$$\min_{z \in \mathbf{R}^n,\ \|z\|_2 = 1} z^T \left(A_n(a(x)) + \frac{1}{n^2} D_n(b(x)) \right) z = \min_{z \in \mathbf{R}^n,\ \|z\|_2 = 1} z^T \hat{A}_n z, \tag{4.15}$$

where

$$\hat{A}_n = \begin{bmatrix} a^{(n)}_{\frac{1}{2}} + a^{(n)}_{\frac{3}{2}} + \frac{1}{n^2} b^{(n)}_1 & -a^{(n)}_{\frac{3}{2}} & & & & & \\ -a^{(n)}_{\frac{3}{2}} & a^{(n)}_{\frac{3}{2}} + a^{(n)}_{\frac{5}{2}} + \frac{1}{n^2} b^{(n)}_2 & -a^{(n)}_{\frac{5}{2}} & & & & \\ & -a^{(n)}_{\frac{5}{2}} & \ddots & \ddots & & & \\ & & \ddots & \ddots & \ddots & & \\ & & & \ddots & \ddots & -a^{(n)}_{n-\frac{1}{2}} & \\ & & & & -a^{(n)}_{n-\frac{1}{2}} & a^{(n)}_{n-\frac{1}{2}} + a^{(n)}_{n+\frac{1}{2}} + \frac{1}{n^2} b^{(n)}_n \end{bmatrix}, \tag{4.16}$$

with $b^n_i = b(x^n_i) = d\left|\frac{i}{n+1} - x^1\right|^r$, $i = 1(1)n$.

Since $a(x^0) = 0$, there exists an index k_n depending on the dimension n, such that

$$i_{k_n} = \underset{i \in \{1,2,\ldots,n+1\}}{\arg\min} |x^{(n)}_{i-1/2} - x^0|. \tag{4.17}$$

Consequently, if we define as I_{x^0} the set containing all the neighboring indexes i's of i_{k_n} for which

$$a^{(n)}_{i-\frac{1}{2}} = a(x^{(n)}_{i-\frac{1}{2}}) = o(1),$$

then that set will have cardinality $o(n)$. Moreover, for all the others $O(n)$ indexes belonging to $I^c_{x^0}$, with $I_{x^0} \cup I^c_{x^0} = \{1, 2, \ldots, n+1\}$, it will hold that

$$a(x^{(n)}_{i-\frac{1}{2}}) \sim O(1) \cap \Omega(1).$$

The same argument, but for a different index i_{l_n}, holds true also for $b(x^1)$. Specifically, there exists an index l_n such that

$$i_{l_n} = \underset{i \in \{1,2,\ldots,n\}}{\arg\min} |x^{(n)}_i - x^1|,$$

and similarly to I_{x^0} and $I^c_{x^0}$, there exist I_{x^1} and $I^c_{x^1}$ with $I_{x^1} \cup I^c_{x^1} = \{1, 2, \ldots, n\}$, such that

$$b(x^{(n)}_i) = o(1), \quad i \in I_{x^1},$$

and

$$b(x^{(n)}_i) \sim O(1) \cap \Omega(1), \quad i \in I^c_{x^1}.$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

In addition, it must hold that $|i_{k_n} - i_{l_n}| \sim n$, since we have assumed that the zeros x^0, x^1 are sufficiently far away from each other. Thus, there exists sequence of indexes i_{q_n} belonging to $I^c_{x^0} \cap I^c_{x^1}$, for example $i_{q_n} = \lfloor \frac{|i_{k_n} - i_{l_n}|}{2} \rfloor$, such that

$$a(x_i^{(n)}) = O(1), \;\; i \geq i_{q_n} \quad \text{and} \quad b(x_i) = O(1), \;\; i \leq i_{q_n}. \tag{4.18}$$

We use this index to split $\hat{A}_n$ in two positive semidefinite terms, $\hat{A}_n^U$ and $\hat{A}_n^L$, such that

$$\hat{A}_n = \hat{A}_n^U + \hat{A}_n^L,$$

where

$$\hat{A}_n^U[ij] = \begin{cases} \hat{A}_n[ij], & i < i_{q_n},\ j = 1(1)n, \\ -a_{i_{q_n}-\frac{1}{2}}, & i = i_{q_n},\ j = i_{q_n} - 1, \\ a_{i_{q_n}-\frac{1}{2}} + \frac{1}{n^2} b_{i_{q_n}}, & i = i_{q_n},\ j = i_{q_n}, \\ 0, & i = i_{q_n},\ j \neq \{i_{q_n} - 1, i_{q_n}\}, \\ 0, & i > i_{q_n},\ j = 1(1)n, \end{cases}$$

and

$$\hat{A}_n^L[ij] = \begin{cases} 0, & i < i_{q_n},\ j = 1(1)n, \\ a_{i_{q_n}+\frac{1}{2}}, & i = i_{q_n},\ j = i_{q_n}, \\ -a_{i_{q_n}+\frac{1}{2}} & i = i_{q_n},\ j = i_{q_n} + 1, \\ 0, & i = i_{q_n},\ j \neq \{i_{q_n}, i_{q_n} + 1\}, \\ \hat{A}_n[ij], & i > i_{q_n},\ j = 1(1)n, \end{cases}$$

Choosing the canonical base $\{e_1, e_2, \ldots, e_n\}$ for writing z in (4.15), we arrive at

$$\min_{z \in \mathbf{R}^n, \|z\|_2 = 1} z^T \hat{A}_n z = \sum_{i=1}^{i_{q_n}} \sum_{j=1}^{i_{q_n}} c_i c_j e_i^T (\hat{A}_n^U)_{ij} e_j + \sum_{i=i_{q_n}}^{n} \sum_{j=i_{q_n}}^{n} c_i c_j e_i^T (\hat{A}_n^L)_{ij} e_j, \tag{4.19}$$

where we have omitted in the first term the zeros obtained by the last $n - i_{q_n}$ zero rows and columns of $\hat{A}_n^U$, and in the second one, the zeros obtained by the first $i_{q_n - 1}$ rows and columns. This is permitted, since we know that the initial matrix $\hat{A}$ is diagonally dominant and as such it has only positive eigenvalues.

The first term in (4.19) is bounded from below by $\lambda_{\min}((\hat{A}^U)_{i_{q_n}})$, i.e., from the minimum eigenvalue of the principal submatrix of order i_{q_n} of $\hat{A}_n^U$. From the monotonicity of $a(x)$ and $b(x)$ we obtain

$$(\hat{A}^U)_{i_{q_n}} \geq 2a(x_{i_{k_n}+\frac{1}{2}}) L_1 + \frac{b(i_{q_n})}{n^2} I_{i_{q_n}} \geq \frac{b(i_{q_n})}{n^2} I_{i_{q_n}}, \tag{4.20}$$

where L_1 is the 1D Laplace matrix of dimension i_{q_n}, with the only difference being that at the position (i_{q_n}, i_{q_n}) it has the value 1 instead of 2. We recall that matrices of this form can be obtained by the discretization of the second derivative using mixed boundary conditions and especially Dirichlet condition on the left and Neumann condition on the right. Since i_{q_n} belongs to $I^c_{x^1}$, from (4.20) we conclude that the first part of the sum in (4.15) can give terms up to an order n^{-2}.

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

Similarly, the second term in (4.19) is bounded from below by the minimum eigenvalue of $(\hat{A}^L)_{n+1-i_{q_n}}$. In addition, omitting the positive values $b(x_i),\ i \geq i_{q_n}$ in its diagonal, we have that

$$(\hat{A}^L)_{n+1-i_{q_n}} \geq a(x_{i_{q_n}-\frac{1}{2}})L^1,$$

where the dimension $n+1-i_{q_n}$ matrix L^1 is like $L_{n+1-i_{q_n}}$ with the only difference being that at the position $(1,1)$ it is equal to 1 instead of 2. Similarly to L_1 this matrix - observe that for the same dimension L_1 and L^1 are similar matrices- can be obtained by the discretization of the second derivative using Neumann condition on the left and Dirichlet condition on the right. Moreover, using difference equations, it can be shown that the minimum eigenvalue $\lambda_1(L^1)$ of L^1 is given by the formula

$$\lambda_1 = 4\sin^2\left(\frac{\pi}{2(2n+1)}\right),$$

and, thus, tends to zero as fast as n^{-2}. Since i_{q_n} belongs to $I^c_{x^0}$, we conclude that also the second term in (4.19) cannot tend to zero faster than n^{-2}, and the proof is completed . ∎

The above rigorous proof can be supplemented with an informal one, that spotlights in-depth our reasoning and shows in a different way the local effect of the possible zeros of the coefficients functions in (1.1), in the ill-conditioning subspace of A_{nn}.

The matrix $\hat{A}_n$ can be expressed (see [10]) as,

$$\begin{aligned}\hat{A}_n = &\sum_{i=2}^{n}\left|\frac{2i+1}{2(n+1)} - x^0\right|^r Q_n(i) + \sum_{i=2}^{n-1}\left|\frac{2i+1}{2(n+1)} - x^1\right|^r E_n(i,i)\\ &+\left(\left|\frac{1}{2(n+1)} - x^0\right|^r + \left|\frac{1}{2(n+1)} - x^1\right|^r\right)E_n(1,1)\\ &+\left(\left|\frac{2n-1}{2(n+1)} - x^0\right|^r + \left|\frac{2n-1}{2(n+1)} - x^1\right|^r\right)E_n(n,n),\end{aligned} \tag{4.21}$$

where the matrices $Q_n(i)$ are symmetric nonnegative definite dyads given by

$$Q_n(i) = \begin{bmatrix} 0 & \cdots & 0 & 0 & \overset{i}{\cdots} & 0\\ \vdots & \ddots & \vdots & \vdots & & \vdots\\ 0 & \cdots & 1 & -1 & \cdots & 0\\ 0 & \cdots & -1 & 1 & \cdots & 0\\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots\\ 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad i = 2,\ldots,n,$$

and $E_n(i,i) = e_i e_i^T$, with e_j, $j = 1,\ldots,n$, representing the j-th column of the identity matrix. Obviously,

$$Q_n(i) = E_n(i,i) + E_n(i-1,i-1) - E_n(i,i-1) - E_n(i-1,i).$$

We recall that the matrix L_n, which is the contribution of the differential operator, can be completely decomposed in terms of dyads. From the above relationship we infer that the influence of the zeros of the coefficient function has a local effect and contributes to a specific elements $E_n(s,t)$ of the basis of the vector space of matrices, i.e., affects only $o(n)$ elements where the generating vectors e_s and e_t are close, in the

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

geometrical sense, to the zero of the function. Thus, due to this "local structure", whenever the zeros x^0, x^1 are sufficiently far away from each other, the relationship (4.21) ensure us that the asymptotically small weights on the correspondent $E_n(i,j)$, caused by the zero of the first coefficient, are canceled by the $O(1)$ weights on the same $E_n(i,j)$ caused by the second one, and vise versa. As a consequence, we expect that the case where the coefficient functions have zeros at different locations, is quite similar in the spectral analysis sense, to the case where the coefficient functions are strictly positive.

REMARK 4.2. Things significantly change if we assume that each of $a(x,y)$, $b(x,y)$ has two -or more- isolated zeros in points of Ω. In this circumstance, in Section 6, we give numerical examples where the condition number of the generated matrix behaves in an unpredictable way. This result is not surprising since it has been observed in the 1D case (see[4]) and the analysis presented in (3.8) exposes the influence of the latter one to the multidimensional one. As a result, the observed anomaly in the asymptotic behavior of the condition number of $A_n(a(x))$ is expected to be inherited by that of $\kappa_2(A_{nn})$, when analogous assumptions hold.

5. Curves of zeros. In this section, we study the case where the coefficient functions $a(x,y)$, $b(x,y)$ have curves of zeros on Ω. Obviously, under these assumptions, the general analysis is much more complicated and perhaps new tools have to be used. Similarly to the case of more than one isolated roots per coefficient, it seems that there might not be a general rule describing the asymptotic behavior of the condition number of A_{nn}. The following statements, which are supported by the numerical experiments of the next section, describe the difficulties. Let $r = \min\{\alpha, \beta, \gamma, \delta\}$. Then:

- There exist $a(x,y)$, $b(x,y)$ such that $r < 2$ and $\kappa_2(A_{nn}) \nsim n^2$.
- There exist $a(x,y)$, $b(x,y)$ such that $r = 2$ and $\kappa_2(A_{nn}) \nsim n^2 \log(n) \cap \Omega(n^2)$.
- There exist $a(x,y)$, $b(x,y)$ such that $r > 2$ and $\kappa_2(A_{nn}) \nsim n^r$.

However, there are concrete occasions where the condition number of A_{nn} behaves according to what Theorem 3.3 predicts. For instance, assuming that one of the coefficient functions, say $a(x,y)$, is of the form $p(x,y) \cdot |x - cy|^\alpha$ where $\alpha > 1$, c positive constant and $p(x,y) > 0$, $\forall (x,y) \in \Omega$, and the other, $b(x,y)$, has a root at the origin of order β, in the y direction, i.e., is of the form $b(x,y) = (x^\gamma + y^\beta) q(x,y)$ with $q(x,y) > 0$, $\forall (x,y) \in \Omega$ and $\beta \leq \gamma$. Then, if $\beta \leq \alpha$,

$$0 \leq |x - cy|^\alpha \leq x^\alpha + c y^\alpha,$$

and

$$c_1(x^\beta + y^\gamma) \leq b(x,y) \leq c_2(x^\beta + y^\gamma),$$

it follows that

$$A_{nn}\left(0, c_1\left(x^\beta + y^\gamma\right)\right) \leq A_{nn}(a,b) \leq A_{nn}\left(x^\alpha + y^\alpha, c_2\left(x^\beta + y^\gamma\right)\right),$$

where $c_1 = \min_{(x,y)\in\Omega} q(x,y)$ and $c_2 = \max_{(x,y)\in\Omega} q(x,y)$. Following the analysis presented in the previous section, we conclude that

$$\lambda_{\min}\left(A_{nn}\left(0, c_1\left(x^\beta + y^\gamma\right)\right)\right) \sim \lambda_{\min}\left(A_{nn}\left(0, x^\beta + y^\beta\right)\right) \sim n^{-r}$$

and

$$\lambda_{\min}\left(A_{nn}\left(x^\alpha + y^\alpha, c_2(x^\beta + y^\gamma)\right)\right) \sim \lambda_{\min}\left(A_{nn}\left(x^\beta + y^\beta, c_2(x^\beta + y^\beta)\right)\right) \sim n^{-r},$$

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

where

$$r = \begin{cases} 2, & 0 \leq \beta < 2, \\ [2,\ 2\log(n)], & \beta = 2, \\ \beta, & \beta > 2. \end{cases}$$

Thus,

$$\lambda_{\min}(A_{nn}(a(x,y), b(x,y))) \sim n^{-r}.$$

Similarly, it can be treated the case where $\beta > \alpha$.

REMARK 5.1. According to Remark 3.4, the assumption about the order of the single zero to occur in a specific direction, i.e., in the x direction in the case of $a(x,y)$ or in the y direction in the case of $b(x,y)$, is necessary in order the Theorem 3.3 to hold, and thus, to have the above analysis.

REMARK 5.2. The restriction about the concrete zero being at the origin can be relaxed. The important requirement is for it to belong on the line of zeros of the other function.

CONJECTURE 5.1. Assuming that the isolated zero of one function and the curve of zeros of the second, or more generally the curves of zeros of both of them, are disjoint sets from each other, the condition number of the generated FD matrix grows as n^2.

The informal proof of the above hypothesis is based on the locality of the influence of the roots of the coefficient functions presented and analyzed in the previous section. Unfortunately, the theory presented there cannot be applied, at least unaltered, under the new assumptions since formula (3.8) does not hold. Numerical experiments concerning that case are presented in the end of the following section.

REMARK 5.3. The analysis we presented can be naturally extended to cover the p-dimensional semielliptical problems with Dirichlet boundary conditions. The straightforward generalization requires only the proper use of tensor arguments and the careful analysis of the resulting Rayleigh quotient.

6. Numerical experiments. We present several tests that numerically confirm the theoretical results presented in the previous sections. For simplification in the notation we assume $n = m$ and $N = n^2$. The quantity that we are focus on and is of great interest in our context is

$$\rho_m = \log_2 \left(\frac{\lambda_{\min}\left(A_{2^m}\right)}{\lambda_{\min}\left(A_{2^{(m+1)}}\right)} \right).$$

Clearly, the sequence $\{\rho_m\}$ reflects the rate of decrease of the minimal eigenvalue of the coefficient matrix A_N. The first set of examples contains functions that fulfill the assumptions of Theorem 3.3 and, for that, we expect the minimum eigenvalue of the corresponding matrix to behave as Theorem 3.3 predicts. In addition, we have considered coefficient functions with different analytical behaviors, and with zeros of order smaller, equal or greater than 2, i.e., the order that the differential operator can contribute. In this direction, we considered the following functions as coefficients in the equation(1.1):

1) $a_1(x,y) = x^1 + y^3$ and $b_1(x,y) = x^{\frac{3}{2}} + y^4$,
2) $a_2(x,y) = x^2 + |y-1|^3$ and $b_2(x,y) = x^3 + |y-1|^2$,
3) $a_3(x,y) = |x - \frac{1}{2}|^3 + y^3$, $b_3(x,y) = |x - \frac{1}{2}|^4 + y^4$.

Specifically, in the first example the minimum order of the zero of the coefficient functions is one, in the second two, while in the last one three. From Theorem 3.3 we expect the condition number to grow as n^2, n^2 or slightly greater, and as n^3, respectively. We remark that in each example we have chosen a different point

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

TABLE 1
Asymptotic behavior of $A_N^i(a_i(x,y), b_i(x,y))$.

n	A_N^1	A_N^2	A_N^3	A_N^4	A_N^5	A_N^6
ρ_4	1.944	2.143	2.967	3.219	2.771	2.885
ρ_5	1.975	2.156	2.983	3.274	2.818	2.938
ρ_6	1.989	2.154	2.996	3.301	2.834	2.967
ρ_7	1.993	2.147	3.025	3.317	2.849	2.983
ρ_8	1.996	2.145	3.031	3.325	2.858	2.991

to be the common zero. The asymptotic behavior of the minimum eigenvalue of A_N is presented through the quantity ρ_m in the columns 2–4 of Table 1.

The examples

4) $a_4(x,y) = x^{10} + y^1$ and $b_4(x,y) = x^{10} + y^{10}$,
5) $a_5(x,y) = x^5 + y^2$ and $b_5(x,y) = x^5 + y^5$,
6) $a_6(x,y) = |x - \frac{1}{2}|^5 + y^3$ and $b_6(x,y) = |x - \frac{1}{2}|^3 + y^5$,

do not fulfill the assumptions of Theorem 3.3 concerning the direction in which the minimum order must occur, i.e., in the x direction for $a(x,y)$ or in the y direction for $b(x,y)$, and thus, fall in the cases mentioned in Remark 3.4. In addition, Example 6 shows the following attribute: in the presence of equal contributions in the minimum order of the zero in the y direction of $a(x,y)$ and in x of $b(x,y)$, the condition number of matrix A_N tends to infinity as exactly the order of the zero if the latter is greater or equal to 2. The results of this group are presented in the last three columns of Table 1.

To illustrate the case of functions with zeros in different points, we used the functions

7) $a_7(x,y) = x^3 + y^3$ and $b_7(x,y) = |x-1|^4 + |y-1|^4$,
8) $a_8(x,y) = x^3 + y^3$ and $b_8(x,y) = |x - 0.05|^3 + y^3$.

The first one, numerically confirms what Theorem 4.1 certifies. The second one attempts to point, from a different perspective, the "local effect" that the zeros of the coefficient function have in the conditioning of the matrix A_N. Obviously, the functions $a_8(x,y)$ and $b_8(x,y)$ have 2 different isolated zeros which are very close to each other. In the third column of Table 2, we present the results. We observe that, as long as the dimension of the matrix is relatively small, i.e., the discretization is coarse enough, then the two zeros are considered by the numerical scheme as the same one and the behavior of the condition number of the matrix is in agreement with what Theorem 3.3 predicts. When n is large enough such that $\frac{1}{n} \ll |x_0 - x_1| = 0.05$, then the condition number tends to infinity as n^2 i.e., according to the Theorem 4.1, which covers the case of functions with a single, but different from each other, zero in the domain Ω.

In agreement with the observations in the 1D case for the coefficient function having two or more concrete zeros, here in the 2D case, when at least one of the coefficient functions have more than one zeros, even if these are common in both functions, the behavior of the condition number could be unpredictable. The following examples

9) $a_9(x,y) = \left(|x - \frac{1}{2}|^4 + y^4\right) \cdot \left(|x-1|^3 + y^3\right)$,
$b_9(x,y) = \left(|x-1|^3 + |y-1|^3\right)\left(|x - \frac{1}{2}|\right)^3 + \left(|y - \frac{1}{2}|\right)^3$,

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

TABLE 2
Asymptotic behavior of $A_N^i(a_i(x,y), b_i(x,y))$.

n	A_N^7	A_N^8	A_N^9	A_N^{10}	A_N^{11}	A_N^{12}	A_N^{13}	A_N^{14}
ρ_4	1.908	2.939	1.897	3.219	2.871	2.871	1.909	4.078
ρ_5	1.955	2.676	1.952	2.219	2.915	2.912	1.955	2.619
ρ_6	1.977	1.775	1.977	3.527	2.929	2.929	1.977	5.118
ρ_7	1.988	1.924	1.988	2.334	2.934	2.937	1.988	1.582
ρ_8	1.994	1.963	1.993	3.608	2.939	2.942	1.994	6.231

10) $a_{10}(x,y) = \left(|x-\frac{1}{2}|^4 + |y-1|^4\right) \cdot \left(|x-\frac{1}{3}|^3 + |y-\frac{1}{2}|^3\right)$,
$b_{10}(x,y) = \left(|x-\frac{1}{2}|^4 + |y-1|^4\right) \cdot \left(|x-\frac{1}{3}|^3 + |y-\frac{1}{3}|^3\right)$,

are indicative, and the results are presented in the correspondent columns of Table 2. If Theorem 3.3 holds, then the condition number in the first example would grow as n^3. Instead, we observe a growth of order 2. In addition, for the second example we notice the unexpected behavior of the condition number of the matrix.

When one of the functions $a(x,y), b(x,y)$ has a curve of zeros and the other has a single zero on this curve, then as we have mentioned in the previous section, under suitable assumptions we can again precisely described the asymptotic behavior of the minimum eigenvalue and so the behavior of the spectral condition number. Representative case is the example 11. On the other hand, if the order of the single zero is not in the appropriate, for the validity of Theorem 3.3, direction, then again we observe the condition number to grow faster than Theorem 3.3 predicts. This case is illustrated by the example 12, where the order of the single zero is 2 but the condition number grows as n^3. The results are presented in the corresponding columns of Table 2.

11) $a_{11}(x,y) = |x-y|^4$, and $b_{11}(x,y) = |x-\frac{1}{2}|^3 + |y-\frac{1}{2}|^3$,
12) $a_{12}(x,y) = |x-y|^5$, and $b_{12}(x,y) = |x-\frac{1}{2}|^2 + |y-\frac{1}{2}|^4$,

For the case of two disjoint curves of zeros we chose the functions

13) $a_{13}(x,y) = |x-y|^3$, and $b_{13}(x,y) = |x-y+\frac{1}{2}|^3$.

The numerical results fully confirm the Conjecture 5.1. Finally, in the last column of Table 2, we give the results of a rather complicate case where both functions

14) $a_{14}(x,y) = |\cos 5x - y|^4$, and $b_{14}(x,y) = |x-y|^3$,

have curves of zeros intersecting each other. As we can observe, the condition number seems to grow in an unpredictable way.

7. Conclusions. In this paper, we have studied the conditioning of the FD sequence of matrices obtained from discretization process of 2D semi-elliptic differential problems. Using elementary and at the same time indicative coefficient functions, we decomposed the matrix to a sum of four simpler terms. Through the asymptotical study of their minimum eigenvalue we were able to estimate the condition number of the matrix for a wide class of coefficient functions having a simple zero. Moreover, using this analysis we were capable to understand the different influence that the order of the zero can offer to the minimum eigenvalue of the matrix, depending on the variable and on the coefficient it happens. Finally, we pointed the

Electronic Journal of Linear Algebra, ISSN 1081-3810
A publication of the International Linear Algebra Society
Volume 34, pp. 566-581, November 2018.
http://repository.uwyo.edu/ela

difficulties that the general analysis will have if the coefficients functions have curves of zeros, we analyzed the case of a combination of curve of zeros and a isolated zero, and we stated a conjecture for the case of disjoint curves of zeros. Concluding, as a future work, we mention the open problem commented in Remark 3.4, the rigorous proof of Conjecture 5.1 and of course the study of the asymptotic behavior of the condition number of the matrix under the hypothesis of more general and complicated cases of zeros.

REFERENCES

[1] A. Arico and M. Donatelli. A V-cycle multigrid for multilevel matrix algebras: proof of optimality. *Numer. Math.*, 105:511–547, 2007.
[2] P. Feehan and C. Pop. Degenerate-elliptic operators in mathematical finance and higher-order regularity for solutions to variational equations. *Adv. Difference Equ.*, 20:361–432, 2015.
[3] C. Garoni and S. Serra Capizzano. *Generalized Locally Toeplitz Sequences: Theory and Applications*. Springer International Publishing, 2017.
[4] D. Noutsos, S. Serra Capizzano, and P. Vassalos. The conditioning of FD matrix sequences coming from semi-elliptic differential equations. *Linear Algebra Appl.*, 428:600–624, 2008.
[5] D. Noutsos, S. Serra Capizzano, and P. Vassalos. Two-level Toeplitz preconditioning: Approximation results for matrices and functions. *SIAM J. Sci. Comput.*, 28:439–458, 2006.
[6] D. Noutsos and P. Vassalos. Band plus algebra preconditioners for two-level Toeplitz systems. *BIT*, 51:695–719, 2011.
[7] O.A. Oleinik and E.V. Radkevich. *Second-Order Equations With Nonnegative Characteristic Form*. American Mathematical Society, Providence, 1973.
[8] T. Otway. *The Dirichlet Problem for Elliptic-Hyperbolic Equations of Keldysh Type*. Springer-Verlag, Berlin, 2012.
[9] E. Radkevich. Equations with nonnegative characteristic form. *J. Math. Sci.*, 158:297–452, 2009.
[10] S. Serra Capizzano. Spectral and structural analysis of high precision finite difference matrices for elliptic operator. *Linear Algebra Appl.*, 293:85–131, 1999.
[11] S. Serra Capizzano. Some theorems on linear positive operators and functionals and their applications. *Comput. Math. Appl.*, 39:139–167, 2008.
[12] S. Serra Capizzano. Spectral behavior of matrix sequences and discretized boundary value problems. *Linear Algebra Appl.*, 337:37–78, 2001.
[13] S. Serra Capizzano. Generalized Locally Toeplitz sequences: spectral analysis and applications to discretized partial differential equations. *Linear Algebra Appl.*, 366:371–402, 2003.
[14] S. Serra Capizzano and C. Tablino Possio. Positive representation formulas for finite difference discretizations of (elliptic) second order pdes. *Comput. Math. Appl.*, 39:139–167, 2000.
[15] P. Tilli. Locally Toeplitz sequences: spectral properties and applications. *Linear Algebra Appl.*, 278:91–120, 1998.