Numer. Math. Theor. Meth. Appl.
doi: 10.4208/nmtma.2011.42s.1

Vol. **4**, No. 2, pp. 113-141
May 2011

Spectral Matrix Conditioning in Cylindrical and Spherical Elliptic Equations

F. Auteri* and L. Quartapelle

Politecnico di Milano, Dipartimento di Ingegneria Aerospaziale,
Via La Masa 34, 20156 Milano, Italy.

Received 31 July 2010; Accepted (in revised version) 16 November 2010

Available online 6 April 2011

Abstract. In the spectral solution of 3-D Poisson equations in cylindrical and spherical coordinates including the axis or the center, it is convenient to employ radial basis functions that depend on the Fourier wavenumber or on the latitudinal mode. This idea has been adopted by Matsushima and Marcus and by Verkley for planar problems and pursued by the present authors for spherical ones. For the Dirichlet boundary value problem in both geometries, original bases have been introduced built upon Jacobi polynomials which lead to a purely diagonal representation of the radial second-order differential operator of all spectral modes. This note details the origin of such a diagonalization which extends to cylindrical and spherical regions the properties of the Legendre basis introduced by Jie Shen for Cartesian domains. Closed form expressions are derived for the diagonal elements of the stiffness matrices as well as for the elements of the tridiagonal mass matrices occurring in evolutionary problems. Furthermore, the bound on the condition number of the spectral matrices associated with the Helmholtz equation are determined, proving in a rigorous way one of the main advantages of the proposed radial bases.

AMS subject classifications: 65M10, 78A48

Key words: Poisson equation, cylindrical and spherical coordinates, Fourier expansion, spherical harmonics, Jacobi polynomials, spectral methods, condition number.

1. Introduction

The spectral method is the par excellence approach for solving elliptic problems in geometrically simple domains. For instance, for the simplest plane domain—a rectangle—the solution can be approximated by a double expansion of product functions defined conveniently so that the 2-D Laplace operator is reduced to a pair of ordinary differential equations, see, *e.g.*, Gustafson [6, p.144]. In this way, variable separation is recognized to be the leading lady of the play of the solution procedure, whenever the partial differential equation is homogeneous.

*Corresponding author. *Email address:* auteri@aero.polimi.it (F. Auteri)

For nonhomogeneous equations, variable separation changes its character substantially on the stage of the numerical dance but it remains starring. The expansion functions, that in the homogeneous case are the product of analytical solutions to ODEs, are now replaced by basis functions that can be chosen more freely. For instance, always with reference to the Poisson equation in a rectangle, polynomials in the Cartesian coordinate variables can be employed. This leads to a matrix representation of the terms of the elliptic equation usually written by the direct product notation. Alternatively, the terms of the constant coefficient discrete equation can be read very conveniently as the pre- and post-multiplication of the rectangular array of unknown coefficients by the operator matrices associated with the two spatial directions. Such an interpretation lends itself to direct solution algorithms which are based on the diagonalization of the pre- and/or post-multiplying matrices, as it was proposed in the classical paper of Haidvogel and Zang [7]. The diagonalization procedure acts independently on each spatial direction and represents therefore a numerical counterpart of the analytical method of separation of variables, as pointed out by Boyd [5, p. 314].

On the other hand, the relative arbitrariness in the selection of the spectral basis functions for nonhomogeneous problems allows one to search for appropriate bases that give the most convenient matrices from the viewpoint of sparsity and conditioning. For the Cartesian Laplacian, Jie Shen has introduced a basis of Legendre polynomials [13] which is optimal for the solution of Dirichlet problems for second-order elliptic equations. In fact, Shen's functions constitute a hierarchical basis and lead to the simplest representation of the second derivative operator: the identity matrix. Furthermore, the spectral components of the unknown in a given direction are coupled only very weakly, as revealed by the tridiagonal profile of the mass matrix, when even-odd mode reordering is exploited. The good properties of this basis are also revealed by the condition number of the basic matrix of the Galerkin spectral solver which goes as N^2, where N is the number of the basis functions in one spatial direction, instead of N^4 as in pure Legendre or Chebyshev polynomial approximations. Thus, a double diagonalization to build a direct solver *à la* Haidvogel and Zang can be efficiently employed and provides an optimally stable solution procedure.

But what happens to the diagonalization and the direct solution strategy for fully 3-D problems in a cylindrical or spherical domain which includes the axis or the centre? Almost invariably, the first step consists in a Fourier expansion of the angular dependence around the cylindrical or polar axis. In this way, the 3-D elliptic equation is transformed into a sequence of 2-D elliptic equations for the Fourier expansion coefficients of the unknown. Then, the dependence on the radial variable must be tackled, having in mind that the equation becomes singular for $r \to 0$. The singularity is actually only a mathematical artifact of the coordinate system employed, while the 3-D solution to any physical problem should not suffer any loss of differentiability there. As a consequence, the expansion coefficients of an infinitely differentiable function expressed in cylindrical or spherical coordinates by a direct product basis must satisfy suitable regularity conditions as $r \to 0$.

In the cylindrical case, a double expansion in direct product form can be employed to represent the dependence on the radial and axial variable, see, *e.g.*, [9, 14]. This approach is simple but it presents the drawback of retaining more degrees of freedom than strictly

necessary since the regularity conditions are disregarded, with an implied wasteful over-resolution near the axis.

An alternative approach is to exploit the regularity conditions [10] and to build a representation of the radial variable employing a number of basis functions which depends on the Fourier wavenumber m. Matsushima and Marcus [12] and independently Verkley [16, 17] proposed one-sided Jacobi polynomials to represent the radial dependence and used different radial functions for each m. These polynomials are one-sided in the sense that there is a transformation $x = 2r^2 - 1$ relating the radial interval $0 \leq r \leq 1$ for the unit circle to the standard unit interval $|x| \leq 1$ and that they incorporate a factor r^m. A triangular truncation scheme is then adopted so that linear systems of order decreasing with m are obtained to solve the second derivative operator in the radial variable. By means of recurrence relations, the Laplacian operator in 2-D polar coordinates is inverted by solving pentadiagonal matrix problems.

The same idea of exploiting the regularity conditions was followed by the present authors for the solution of the 3-D Dirichlet boundary value problem in a finite cylinder [3]. A different basis of one-sided polynomials was introduced, endowed with additional favourable properties: the essential fulfillment of homogeneous Dirichlet conditions and a maximal reduction of the algebraic complexity of the stiffness problem in a circle and in a cylinder. In fact, a term proportional to m^2 is made to appear in the Laplacian term by the Fourier expansion. Apparently, this term has nothing to do with that representing the second-order radial derivative. As a matter of fact, thanks to a judicious, and quite natural, choice of the Jacobi polynomials, the m^2 term cancels out with part of the derivative term. As a result, the remaining part of the second derivative is found to be represented quite simply by a diagonal matrix or, more precisely, by a sequence of diagonal matrices of dimension decreasing with m. At the same time, the remaining z term of the Laplacian requires the introduction of mass matrices which are found to be tridiagonal. Thus, the cylindrical Laplacian reduces to a sequence of 2-D elliptic equations which can be solved directly by double diagonalization, similarly to the 2-D Cartesian Laplacian. However, since the diagonal matrix representing the radial stiffness is not the identity nor a constant, a generalized eigenvector decomposition is required for dealing with the matrices for the radial variable. Consequently, the separation of variables is still possible in polar/cylindrical geometry, through the generalized eigendecomposition of matrices with dimension decreasing with m. In conclusion, by construction, the new Jacobi basis avoids the so-called pole problem encountered in evolutionary problems by spectral methods of direct-product type and leads to a solution algorithm free from any time-step over-restriction.

Coming to the Poisson equation inside a sphere, most spectral methods rely upon spherical harmonics which, being the eigenfunction of the surface Laplacian operator, reduce the 3-D equation to a set of independent ordinary differential equations in the radial variable. Different choices for the approximation over the sphere are possible, see [15] and the references therein. An extended discussion about the spectral approximation of the radial dependence is presented in [11]. Considering the complete equation for the expansion coefficients in spherical harmonics, a term proportional to $\ell(\ell+1)$ occurs in it, similarly to the cylindrical case. Luckily, also in the spherical case, a basis of functions of the radial variable

can be conceived so that the $\ell(\ell+1)$ term is absorbed in a single final term associated with the second order radial derivative [2]. Moreover, the matrix associated with such a term turns out to be diagonal, as in the cylindrical case. Thus, combining spherical harmonics with the new Jacobi radial basis for the spherical domain achieves the remarkable result of a complete diagonalization of the 3-D Poisson equation inside a sphere: the basis functions of the two sets combined together are indeed the eigenfunctions of the three-dimensional Laplacian in a spherical region. Only when solving evolutionary problems, a second term is present in the ordinary differential equation, that involves the mass matrices, found to be tridiagonal. In this cases the highest variable separation achievable in the solution of the Dirichlet problem inside a sphere requires to invert only tridiagonal matrix problems. The same applies also to the Neumann problem which can be solved by the same basis [2].

Once the most natural bases have been formulated to solve Dirichlet problems in cylindrical and spherical domains by fully spectral approximations, it remains to assess their numerical properties, embodied in the condition numbers of the matrices associated to the considered elliptic operator. This is precisely the subject of the present paper that focuses on the determination of theoretical estimates and bounds for the condition numbers of the spectral matrices defined by the new Jacobi polynomials, after the explicit expressions of all their nonzero elements have been derived.

The paper is structured as follows. Section 2 discusses the case of cylindrical coordinates while Section 3 is devoted to spherical coordinates. In both sections, we first define the spectral approximation of the Dirichlet problem for the Laplace operator, then give the formula defining all matrix elements of the diagonal stiffness and the tridiagonal mass matrices. Moreover, we establish bounds on matrix elements and provide estimates and bounds for the eigenvalues and the condition numbers of the relevant spectral solution matrices. The last section is devoted to the concluding remarks.

2. Spectral approximation in cylindrical coordinates

2.1. Dirichlet problem

Let us consider the Poisson equation in cylindrical coordinates for a scalar unknown $u = u(r,z,\phi)$

$$-\frac{1}{r}\,\partial_r\big(r\,\partial_r u\big) - \frac{1}{r^2}\,\partial_\phi^2 u - \partial_z^2 u = f(r,z,\phi), \tag{2.1}$$

where $f(r,z,\phi)$ is a known source term defined in a cylindrical domain of finite axial extent $\Omega \equiv (0,c] \times [-h,h] \times [0,2\pi)$, including part of the z axis. The elliptic equation (2.1) is assumed to be supplemented by the homogeneous Dirichlet boundary condition $u = 0$ on the entire boundary $\partial\Omega$ of the cylinder.

Thanks to the periodic character of the ϕ variable, the right-hand side and the unknown can be expanded by means of a real Fourier series. In order to discretize the problem, the series is truncated at a suitable integer $N > 0$, so that $-N+1 \le m \le N$, and the Fourier expansions are approximated by finite summations; for instance, the truncated

expansion of the unknown is

$$u(r,z,\phi) = u^0(r,z) + u^N(r,z)\cos(N\phi) + 2\sum_{m=1}^{N-1}\big(u^m(r,z)\cos(m\phi) - u^{-m}(r,z)\sin(m\phi)\big), \tag{2.2}$$

where the coefficients $u^m(r,z)$ and $u^{-m}(r,z)$, for $m = 0,1,2,\cdots$, are defined by

$$u^{\pm m}(r,z) = \frac{1}{2\pi}\int_0^{2\pi} u(r,z,\phi)\begin{matrix}\cos(m\phi)\\ \sin(m\phi)\end{matrix} d\phi. \tag{2.3}$$

The absence of the coefficient 2 in front of last term in (2.2) should be noticed. A similar Fourier expansion is used for the right-hand side f. To ensure infinite differentiability of the solution on the axis, the Fourier components $u^m(r,z)$ will have to satisfy the regularity conditions for $r \to 0$ reported in [10]:

$$u^m(r,z) \sim r^{|m|}U^m(r^2,z), \tag{2.4}$$

where U^m is a regular function of both variables.

The expansion (2.2) is now introduced in the original elliptic equation (2.1). Equating similar terms and simplifying, we obtain a system of uncoupled equations for the modal unknowns $u^m(r,z)$

$$-\frac{1}{r}\,\partial_r\big(r\,\partial_r u^m\big) + \frac{m^2u^m}{r^2} - \partial_z^2 u^m = f^m(r,z). \tag{2.5}$$

Let us introduce, for $m \geq 0$, the polynomials

$$P_i^{\star\, m}(s) = \frac{1-s}{2}\,P_{i-1}^{(1,m)}(s), \qquad i = 1,2,\cdots, \tag{2.6}$$

where $P_i^{(\alpha,\beta)}(s)$, $-1 \leq s \leq 1$, denotes the Jacobi polynomials [1]. Then, we introduce the expansion functions

$$Q_i^m(s) \equiv \Big(\frac{1+s}{2}\Big)^{m/2} P_i^{\star\, m}(s) = \frac{1-s}{2}\Big(\frac{1+s}{2}\Big)^{m/2} P_{i-1}^{(1,m)}(s), \qquad i = 1,2,\cdots, \tag{2.7}$$

as well as their counterparts with the dimensionless radial variable $\rho = r/c$, related to s by $s = 2\rho^2 - 1$, as independent variable

$$\begin{aligned} B_i^m(\rho) \equiv & Q_i^m\big(2\rho^2-1\big) = \rho^m\, P_i^{\star\, m}\big(2\rho^2-1\big) \\ = & (1-\rho^2)\rho^m\, P_{i-1}^{(1,m)}\big(2\rho^2-1\big), \qquad i \geq 1. \end{aligned} \tag{2.8}$$

Basis functions $B_i^m(\rho)\sin(m\phi)$ and $B_i^m(\rho)\cos(m\phi)$ for the spectral approximation in the circle are shown in Fig. 1 for $N = 4$. The plot clearly illustrates the triangular truncation of the proposed approximation: the higher the wavenumber the lower the number of basis functions in the radial direction. Moreover, the plot shows the dependence on the Fourier

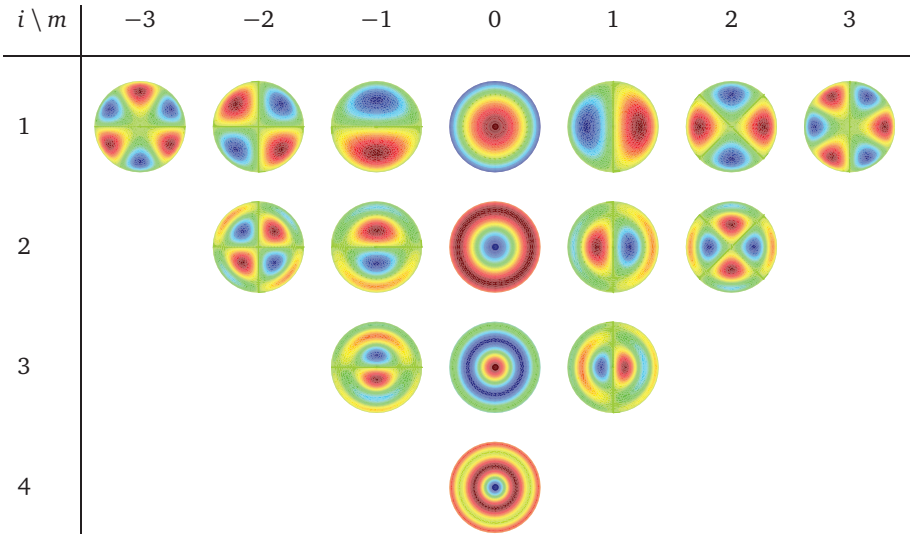

Figure 1: Basis functions for the Dirichlet problem inside a circle $B_i^m(\rho)\sin(m\phi)$, left, and $B_i^m(\rho)\cos(m\phi)$, right, for $N=4$.

wavenumber of the radial resolution near the centre: the higher the wavenumber, the lower the resolution.

The approximate solution $u^m(\rho,\zeta)$, where $\zeta = z/h$, with $|\zeta| \le 1$, is expanded in the double series

$$u^m(\rho,\zeta) = \sum_{i=1}^{N-|m|} B_i^{|m|}(\rho)\, u_{i;j}^m\, L_j^*(\zeta) \sum_{j=2}^{J}, \tag{2.9}$$

where the inverted summation symbol is used for the sum over the second index.

On the other side, the basis $\{L_j^*(\zeta)\}$, $|\zeta| \le 1$, is defined as

$$L_j^*(\zeta) = \frac{L_{j-2}(\zeta) - L_j(\zeta)}{\sqrt{2(2j-1)}}, \qquad j \ge 2. \tag{2.10}$$

This basis was introduced by Shen [13] and contains linear combinations of two Legendre polynomials in order to satisfy homogeneous Dirichlet conditions at both extremes $\zeta = \pm 1$.

The complete spectral expansion considered here for the unknown $u(\rho,\zeta,\phi)$ of the Dirichlet boundary value problem with homogeneous conditions in the cylindrical domain is therefore given by

$$\begin{aligned} u(\rho,\zeta,\phi) = &\sum_{i=1}^{N} B_i^0(\rho)\, u_{i;j}^0\, L_j^*(\zeta) \sum_{j=2}^{J} \\ &+ 2\sum_{m=1}^{N-1} \left(\sum_{i=1}^{N-m} B_i^m(\rho)\, u_{i;j}^{\pm m}\, L_j^*(\zeta) \sum_{j=2}^{J} \right) \begin{matrix} \cos(m\phi) \\ -\sin(m\phi) \end{matrix}. \end{aligned} \tag{2.11}$$

The presence of superposed cosine and sine functions means that two distinct series are involved by the Fourier summation. The nested dependence of the upper extreme of the summation over i on the index m of the Fourier expansion must be noticed.

The standard way of obtaining the fully discrete spectral equations of the Dirichlet-Poisson problem is to start from the modal equation (2.5) and to multiply it by r to obtain the weak formulation in the proper weighted Sobolev space [4]. Then, having introduced the dimensionless variables $\rho = r/c$ and $\zeta = z/h$, $0 < \rho \leq 1$ and $|\zeta| \leq 1$, equation (2.5) is recast in weak form by the Galerkin method. The equation for the transformed modal unknown $\tilde{u}^m(\rho,\zeta) = u^m(r,z)$, which will still be indicated by the same letter u^m, as $u^m(\rho,\zeta)$, is obtained in the form

$$\int_0^1 \int_{-1}^1 \left(\frac{\rho}{c^2} (\partial_\rho v)(\partial_\rho u^m) + \frac{m^2}{c^2} \frac{v\, u^m}{\rho} + \rho\, (\partial_\zeta v)(\partial_\zeta u^m) \frac{1}{h^2} \right) d\rho\, d\zeta$$
$$= \int_0^1 \int_{-1}^1 \rho\, v(\rho,\zeta) f^m(c\rho, h\zeta)\, d\rho\, d\zeta, \tag{2.12}$$

where $v(\rho,\zeta)$ denotes the weighting function vanishing on the domain boundary. Introducing the expansion (2.9) of $u^m(\rho,\zeta)$ into the weak equation for this unknown, and choosing as weighting functions $v(\rho,\zeta)$ the same basis functions used to expand the solution, the weak equation leads to the following system of equations

$$c^{-2} D^{\boxed{m}}\ \mathsf{U}^m M + M^{\boxed{m}}\ \mathsf{U}^m D h^{-2} = \mathsf{F}^m, \tag{2.13}$$

for $-N+1 \leq m \leq N$. Here the matrices $D^{\boxed{m}}$ and $M^{\boxed{m}}$ have elements defined as follows, with $m \geq 0$,

$$d^{\boxed{m}}_{i,i'} = \int_{-1}^1 \left\{ 4 \left(\frac{1+s}{2} \right) [D_s Q_i^m(s)][D_s Q_{i'}^m(s)] + \frac{m^2}{4} \left(\frac{1+s}{2} \right)^{-1} Q_i^m(s) Q_{i'}^m(s) \right\} ds,$$
$$\mu^{\boxed{m}}_{i,i'} = \frac{1}{4} \int_{-1}^1 Q_i^m(s) Q_{i'}^m(s)\, ds, \tag{2.14}$$

where $D_s = d/ds$. The order of a matrix with superscript $\boxed{m}$ is $N-m$, and is therefore $N, (N-1), \cdots, 2, 1$ for $m = 0, 1, 2, \cdots, N-1$, respectively. On the other side, matrix M in the axial direction have elements defined by

$$\mu_{j,j'} = \int_{-1}^1 L_j^*(\zeta)\, L_{j'}^*(\zeta)\, d\zeta$$

and is found to be pentadiagonal, while the stiffness matrix D has elements defined by

$$d_{j,j'} = \int_{-1}^1 [D_\zeta L_j^*(\zeta)][D_\zeta L_{j'}^*(\zeta)]\, d\zeta$$

and is found to be coincident with the identity matrix.

2.2. Tridiagonal mass matrices

The profile of the symmetric mass matrix $M^{\boxed{m}}$ and its elements are established by the following

Proposition 2.1. *For any $m \geq 0$, the mass matrix $M^{\boxed{m}}$ is tridiagonal and its nonzero elements are defined by*

$$\mu^{\boxed{m}}_{i,i} = \frac{i^2}{(2i+m-1)(2i+m)(2i+m+1)}, \qquad \text{for } 1 \leq i \leq N-m, \tag{2.15}$$

$$\mu^{\boxed{m}}_{i,i+1} = \frac{-i(i+1)}{2(2i+m)(2i+m+1)(2i+m+2)}, \qquad \text{for } 1 \leq i \leq N-m-1, \tag{2.16}$$

and $\mu^{\boxed{m}}_{i,i-1} = \mu^{\boxed{m}}_{i-1,i}$, for $2 \leq i \leq N-m$.

Proof. The tridiagonal character of $M^{\boxed{m}}$ is a consequence of the orthogonality relation of the Jacobi polynomials $P_i^{(\alpha,\beta)}(s)$, with $|s| \leq 1$, $i = 0, 1, 2, \cdots$, and $\alpha > 0$ and $\beta > 0$, which reads

$$\begin{aligned}&\int_{-1}^{1} (1-s)^\alpha (1+s)^\beta P_i^{(\alpha,\beta)}(s)\, P_k^{(\alpha,\beta)}(s)\, ds \\ &= \frac{2^{\alpha+\beta+1}}{\alpha+\beta+2i+1} \frac{\Gamma(\alpha+i+1)\Gamma(\beta+i+1)}{i!\,\Gamma(\alpha+\beta+i+1)}\, \delta_{i,k}.\end{aligned} \tag{2.17}$$

In the particular case $\alpha = 1$ this relation reduces to

$$\int_{-1}^{1} (1-s)(1+s)^\beta P_i^{(1,\beta)}(s)\, P_k^{(1,\beta)}(s)\, ds = \frac{2^{\beta+2}(i+1)}{(\beta+2i+2)(\beta+i+1)}\, \delta_{i,k}, \tag{2.18}$$

and for $\alpha = 0$

$$\int_{-1}^{1} (1+s)^\beta P_i^{(0,\beta)}(s)\, P_k^{(0,\beta)}(s)\, ds = \frac{2^{\beta+1}}{\beta+2i+1}\, \delta_{i,k}. \tag{2.19}$$

By the definition (2.7) the elements of the mass matrix are

$$\mu^{\boxed{m}}_{i,i'} = \frac{1}{4} \int_{-1}^{1} \left(\frac{1-s}{2}\right)^2 \left(\frac{1+s}{2}\right)^m P_{i-1}^{(1,m)}(s)\, P_{i'-1}^{(1,m)}(s)\, ds. \tag{2.20}$$

Thanks to the recurrence relation for Jacobi polynomials, for $i \geq 1$,

$$\frac{1-s}{2} P_{i-1}^{(1,m)}(s) = \frac{i}{2i+m} \left(P_{i-1}^{(0,m)}(s) - P_i^{(0,m)}(s) \right), \tag{2.21}$$

we obtain

$$\begin{aligned}\mu^{\boxed{m}}_{i,i'} =& \frac{ii'}{4(2i+m)(2i'+m)} \int_{-1}^{1} \left(\frac{1+s}{2}\right)^m \\ &\times \left(P_{i-1}^{(0,m)}(s) - P_i^{(0,m)}(s) \right) \left(P_{i'-1}^{(0,m)}(s) - P_{i'}^{(0,m)}(s) \right) ds.\end{aligned}$$

Developing the product, the following four contributions

$$P_{i-1}^{(0,m)} P_{i'-1}^{(0,m)}, \quad -P_{i-1}^{(0,m)} P_{i'}^{(0,m)}, \quad -P_{i}^{(0,m)} P_{i'-1}^{(0,m)}, \quad P_{i}^{(0,m)} P_{i'}^{(0,m)}$$

are obtained in the integrand. However, thanks to the orthogonality relation (2.19), each of them can integrate to a nonzero quantity only provided the $i' = i$ or $i' = i \pm 1$. Thus, the symmetric mass matrix $M^{\boxed{m}}$ is tridiagonal. A direct calculation provides the values stated in the proposition. □

2.3. Diagonal stiffness matrices

The standard variational procedure leads to the definition of the elements $d_{i,i'}^{\boxed{m}}$ of the stiffness matrix $D^{\boxed{m}}$ given by (2.14). However, this definition hides the fundamental property of the stiffness matrix $D^{\boxed{m}}$, namely, that of being diagonal and of not depending on the term $\propto m^2$. These properties can be established starting from the integral form of the equation but with the differential operator retained in strong form before the integration by parts and employing the Jacoby equation, as suggested by P. W. Livermore in a private communication (2008) about the spherical case (see below). We have in fact the

Proposition 2.2. *For any $m \geq 0$, the stiffness matrix $D^{\boxed{m}}$ is diagonal and its elements are given by*

$$d_i^{\boxed{m}} = \frac{2i^2}{2i+m} \qquad \text{with } 1 \leq i \leq N-m. \tag{2.22}$$

Proof. Instead of the stiffness element definition (2.14), let us consider the modal elliptic equation (2.5) in strong form and then consider the expansion component $Q_i^m(s)$ of the first two terms of the equation. After introducing the change of variables, they become

$$-4D_s\left\{\frac{1+s}{2}D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s)\right]\right\} + \frac{m^2}{4}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}-1}P_{i-1}^{(1,m)}(s).$$

Let us focus on the first contribution, by developing the internal derivative, we obtain

$$\begin{aligned} D_s\left\{\frac{1+s}{2}D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s)\right]\right\} &= D_s\left\{\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}+1}D_sP_{i-1}^{(1,m)}(s)p\right. \\ &\left. + \frac{m}{4}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s) - \frac{1}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}+1}P_{i-1}^{(1,m)}(s)\right\}. \end{aligned}$$

By evaluating the external derivative we obtain

$$\begin{aligned} &D_s\left\{\frac{1+s}{2}D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s)\right]\right\} \\ =&\frac{m^2}{16}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}-1}P_{i-1}^{(1,m)}(s) + \frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}+1}D_s^2P_{i-1}^{(1,m)}(s) \\ &-\frac{m+1}{4}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s) + \left(\frac{(m+1)(1-s)}{4} - \frac{1+s}{2}\right)\left(\frac{1+s}{2}\right)^{\frac{m}{2}}D_sP_{i-1}^{(1,m)}(s). \end{aligned}$$

By taking into account the coefficients -4 in front of the original term, the term $\propto m^2$ in the third line above is found to cancel with the second contribution of the expression we started from. This result is remarkable and leads to the following replacement, after reordering the remaining terms:

$$D_s\left\{\frac{1+s}{2}D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s)\right]\right\} \longrightarrow \frac{1}{4}\left(\frac{1+s}{2}\right)^{\frac{m}{2}} \times\Big\{(1-s^2)D_s^2P_{i-1}^{(1,m)}(s)+[m-1-(m+3)s]D_sP_{i-1}^{(1,m)}(s)-(m+1)P_{i-1}^{(1,m)}(s)\Big\}.$$

By the Jacobi equation

$$(1-s^2)D_s^2P_i^{(1,m)}+[m-1-(m+3)s]D_sP_i^{(1,m)}+i(i+m+2)P_i^{(1,m)}=0, \tag{2.23}$$

with $i=0,1,2\cdots$, the replacement above simplifies to

$$D_s\left\{\frac{1+s}{2}D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s)\right]\right\} \longrightarrow -\frac{i(i+m)}{4}\left(\frac{1+s}{2}\right)^{\frac{m}{2}}P_{i-1}^{(1,m)}(s).$$

As a consequence, the stiffness matrix elements can be characterized more simply as

$$d_{i,i'}^{\boxed{m}}=i'(i'+m)\int_{-1}^{1}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{m}P_{i-1}^{(1,m)}(s)\,P_{i'-1}^{(1,m)}(s)\,ds,$$

which, by virtue of the orthogonality condition (2.18) with $\beta=m$, gives

$$d_{i,i'}^{\boxed{m}}=\frac{2i^2}{2i+m}\,\delta_{i,i'}.$$

For simplicity the diagonal elements will be denoted by $d_i^{\boxed{m}}$ instead of $d_{i,i}^{\boxed{m}}$. □

2.4. Matrix properties

Once the stiffness matrix $D^{\boxed{m}}$ and the mass matrix $M^{\boxed{m}}$ have been computed, bounds on their elements can be obtained to be used for estimating the condition numbers.

2.4.1. Stiffness matrix

First, let us examine the behaviour of the matrix elements with respect to i.

Proposition 2.3. *For fixed m and for $i\geq 1$, the diagonal elements $d_i^{\boxed{m}}$ of $D^{\boxed{m}}$, are increasing with i, hence*

$$\min_{i\geq 1}d_i^{\boxed{m}}=d_1^{\boxed{m}}\equiv\frac{2}{2+m},\qquad m\geq 0. \tag{2.24}$$

Proof. The order property is stated by the inequality:

$$\frac{(i+1)^2}{2(i+1)+m} \geq \frac{i^2}{2i+m} \quad \Longleftrightarrow \quad 2i^2+2(m+1)\,i+m \geq 0.$$

which is satisfied for

$$i \leq -\tfrac{1}{2}\big(m+1+\sqrt{m^2+1}\,\big) < 0 \quad \text{or} \quad i \geq -\tfrac{1}{2}\big(m+1-\sqrt{m^2+1}\,\big) \leq 0.$$

The thesis follows from the second inequality since $i \geq 1$. □

The opposite trend is obtained with respect to m. It is indeed trivial to prove the following

Proposition 2.4. *The diagonal elements $d_i^{\boxed{m}}$ of $D^{\boxed{m}}$ are decreasing with $m \geq 0$, for fixed $i \geq 1$ and as $m \to \infty$, $d_i^{\boxed{m}}$ behave as m^{-1}.*

2.4.2. Mass matrix

Let us analyse the elements of $M^{\boxed{m}}$. Let us start with the diagonal elements $\mu_{i,i}^{\boxed{m}}$. First, we prove that $\mu_{i,i}^{\boxed{m}}$ are bounded from above.

Proposition 2.5. *$\forall m \geq 0$ and $i \geq 1$, $\mu_{i,i}^{\boxed{m}} \leq \frac{1}{6}$.*

Proof. The diagonal elements $\mu_{i,i}^{\boxed{m}}$ of $M^{\boxed{m}}$ are decreasing as m is increased at fixed i. So, we can simply analyse the case $m = 0$. In this case, $\mu_{i,i}^{\boxed{\cdot}}$ are decreasing with i, provided the following inequality

$$\frac{i^2}{(2i-1)\,2i\,(2i+1)} \geq \frac{(i+1)^2}{(2i+1)(2i+2)(2i+3)}$$

is satisfied, which occurs for $i < -\frac{3}{2}$ and $i > \frac{1}{2}$. Recalling that $i \geq 1$, the maximum is therefore found to be $\mu_{1,1}^{\boxed{\cdot}} = \frac{1}{6}$. □

Much in the same manner we prove the following:

Proposition 2.6. *$\forall m \geq 0$ and $i \geq 1$, $\left|\mu_{i,i+1}^{\boxed{m}}\right| \leq \frac{1}{24}$.*

It will be useful also the following:

Proposition 2.7. *For any row $i \geq 1$ and for $m \geq 0$, the diagonal element of the mass matrix is greater than the sum of the modulus of the offdiagonal elements:*

$$\mu_{1,1}^{\boxed{m}} > \left|\mu_{1,2}^{\boxed{m}}\right|, \qquad \mu_{i,i}^{\boxed{m}} > \left|\mu_{i,i-1}^{\boxed{m}}\right| + \left|\mu_{i,i+1}^{\boxed{m}}\right|, \qquad \textit{for} \;\; 2 \leq i \leq N-m-1,$$

and

$$\mu_{N-m,N-m}^{\boxed{m}} > \left|\mu_{N-m,N-m-1}^{\boxed{m}}\right|.$$

Proof. The first row $i=1$ has only two nonzero elements,

$$\mu^{\boxed{m}}_{1,1} = \frac{1}{(m+1)(m+2)(m+3)}, \qquad \mu^{\boxed{m}}_{1,2} = -\frac{1}{(m+2)(m+3)(m+4)},$$

thus the inequality $\mu^{\boxed{m}}_{1,1} > \left|\mu^{\boxed{m}}_{1,2}\right|$ is satisfied, as could be inferred also from the positive definiteness of the mass matrix. For $i>1$ the thesis means that

$$\frac{i^2}{(2i+m-1)(2i+m)(2i+m+1)} > \frac{i(i+1)}{2(2i+m)(2i+m+1)(2i+m+2)} + \frac{(i-1)\,i}{2(2i+m-2)(2i+m-1)(2i+m)}.$$

Since $m>0$ and $i>1$, by direct calculation it is readily shown that this inequality is always satisfied for $2 \le i \le N-m-1$. Finally, for the last row we have

$$\mu^{\boxed{m}}_{N,N} = \frac{N^2}{(2N+m-1)(2N+m)(2N+m+1)}, \quad \mu^{\boxed{m}}_{N,N-1} = \frac{-(N-1)N}{2(2N+m-2)(2N+m-1)(2N+m)},$$

thus the inequality $\mu^{\boxed{m}}_{N,N} > |\mu^{\boxed{m}}_{N,N-1}|$ simplifies to

$$\frac{N}{2N+m+1} > \frac{N-1}{2(2N+m-2)},$$

which is identically satisfied for any $N \ge 1$, as could be also predicted, as before, from the positive definiteness of the mass matrix. □

As will be shown later, these estimates are useful to describe how the condition number depends on the parameter N. However, to investigate also how the condition number depends on m for fixed N, the sharper estimates reported in the following propositions will be needed.

Proposition 2.8. *$\forall m \ge 0$ and $i \ge 1$,*

$$\mu^{\boxed{m}}_{i,i} \le \min\left(\frac{1}{6}, \frac{(N-m)^2}{(m+1)(2N-m)(2N-m+1)}\right). \tag{2.25}$$

Proof. Let us first exploit the fact that $i \ge 1$ to rewrite the diagonal elements as

$$\mu^{\boxed{m}}_{i,i} = \frac{1}{(2i+m-1)(2+m/i)(2+(m+1)/i)} \qquad \text{for } 1 \le i \le N-m. \tag{2.26}$$

To obtain an upper bound with respect to i of the diagonal elements, we need to consider the minimum of the fraction denominator since its numerator is constant. Let us consider separately the first term in the denominator from the second and third terms. For the first one, it is simple to show that it has a minimum for fixed m and $i \ge 1$ by verifying the inequalities

$$2i+m+1 > 2i+m-1 \ge m+1.$$

The other two terms in the denominator can be written as $2+k/i$, $k \geq 0$. They are both decreasing with i since the inequality

$$2+m/i \geq 2+m/(i+1)$$

is identically satisfied provided $k \geq 0$ which is true for both terms for all $m \geq 0$. Then the minimum denominator corresponds to $(m+1)[2+m/(N-m)][2+(m+1)/(N-m)]$ and rearranging we obtain immediately

$$\mu_{i,i}^{\boxed{m}} \leq \frac{(N-m)^2}{(m+1)(2N-m)(2N-m+1)}.$$

Now, for $m=0$ and N sufficiently high this bound can be less sharp than the one provided in Proposition 2.5, in fact the limit

$$\lim_{N\to\infty} \frac{(N-m)^2}{(m+1)(2N-m)(2N-m+1)} = \frac{1}{4(m+1)}$$

is greater than $1/6$ if $m=0$. The thesis follows considering the minimum between the two. □

Proposition 2.9. *$\forall m \geq 0$ and $i \geq 1$,*

$$\left|\mu_{i,i+1}^{\boxed{m}}\right| \leq \min\left(\frac{1}{24}, \frac{(N-m)(N-m-1)}{2(2N-m)(2N-m-2)(m+3)}\right). \tag{2.27}$$

Proof. The estimate can be obtained by an approach similar to the one adopted to prove Proposition 2.6. Starting from the expression of the subdiagonal element $\left|\mu_{i,i+1}^{\boxed{m}}\right|$, we rewrite the expression as

$$\frac{1}{2(2+m/i)(2i+m+1)[2+m/(i+1)]} \tag{2.28}$$

it is possible since $i \geq 1$. Then observe that this expression is maximum if the minimum value of $i=1$ is taken in the second term in the denominator and the maximum value of $i=N-m-1$ is taken in the other two terms. □

2.5. Eigenvalue estimates

To estimate the condition numbers, we first investigate the bounds of the eigenvalues of the spectral matrices representing the radial operator associated with the Helmholtz operator $(-\nabla^2+\gamma)$, with $\gamma \geq 0$, which occurs in evolutionary problems, where typically $\gamma \propto 1/\Delta t$, Δt denoting a time step. This operator leads to matrices $A^{\boxed{m}}$ which are a linear combination of the stiffness matrix $D^{\boxed{m}}$ and the mass matrix $M^{\boxed{m}}$:

$$A^{\boxed{m}} = D^{\boxed{m}} + \alpha M^{\boxed{m}}, \tag{2.29}$$

where $\alpha = c^2\gamma \geq 0$. Therefore the diagonal elements of the $A^{\boxed{m}}$ matrices are

$$a_{i,i}^{\boxed{m}} = \left(2 + \frac{\alpha}{(2i+m)^2-1}\right)\frac{i^2}{2i+m}, \qquad 1 \leq i \leq N-m-1, \tag{2.30}$$

while $a_{i,i+1}^{\boxed{m}}$ are simply the extradiagonal elements of $M^{\boxed{m}}$ multiplied by α, namely:

$$a_{i,i+1}^{\boxed{m}} = \frac{-\alpha\, i(i+1)}{2(2i+m)(2i+m+1)(2i+m+2)}, \qquad 1 \leq i \leq N-m-1, \tag{2.31}$$

and $a_{i,i-1}^{\boxed{m}} = a_{i-1,i}^{\boxed{m}}$, for $2 \leq i \leq N-m$.

Owing to their definition, matrices $A^{\boxed{m}}$ are tridiagonal symmetric (and simply diagonal when $\alpha = 0$). They are also positive definite since the stiffness matrix is diagonal with strictly positive elements and the mass matrices come from the discretization of the L^2 norm. As a consequence, the eigenvalues of $A^{\boxed{m}}$ lie on the strictly positive real axis.

To estimate the eigenvalues, the Gershgorin-Hadamard theorem can be employed. It states that matrix eigenvalues are contained in the union of the disks whose center is each diagonal element and whose radius is the absolute sum of all other elements of the corresponding matrix row. Since matrix $A^{\boxed{m}}$ is real and symmetric any its eigenvalue λ is real and since the matrix is tridiagonal all the eigenvalues are such that

$$\begin{cases} \left|\lambda - a_{1,1}^{\boxed{m}}\right| \leq \left|a_{1,2}^{\boxed{m}}\right|, \\ \left|\lambda - a_{i,i}^{\boxed{m}}\right| \leq \left|a_{i,i-1}^{\boxed{m}}\right| + \left|a_{i,i+1}^{\boxed{m}}\right|, & \text{for } 2 \leq i \leq N-m-1. \\ \left|\lambda - a_{N-m,N-m}^{\boxed{m}}\right| \leq \left|a_{N-m,N-m-1}^{\boxed{m}}\right|, \end{cases} \tag{2.32}$$

2.5.1. Largest and smallest eigenvalue estimate

We can now proceed to estimate the maximum eigenvalue of matrix $A^{\boxed{m}}$, namely its spectral radius $\rho(A^{\boxed{m}})$, and the minimum eigenvalue

$$\rho(A^{\boxed{m}}) = \max_k \left|\lambda_k(A^{\boxed{m}})\right| \qquad \text{and} \qquad \nu(A^{\boxed{m}}) = \min_k \left|\lambda_k(A^{\boxed{m}})\right|, \tag{2.33}$$

where $\lambda_k(A^{\boxed{m}})$ denotes the k-th eigenvalue of matrix $A^{\boxed{m}}$.

Proposition 2.10. *The following bound applies to the spectral radius of matrix $A^{\boxed{m}}$:*

$$\rho(A^{\boxed{m}}) \leq \frac{2(N-m)^2}{2N-m} + \frac{\alpha}{4}. \tag{2.34}$$

Proof. The proof follows directly from the Gershgorin-Hadamard theorem thanks to the definition of $A^{\boxed{m}}$. Let $R_{\max}^{\boxed{m}}$ denote the maximum radius of the circles centered at $|a_{i,i}^{\boxed{m}}|$, $1 \leq i \leq N-m$, namely

$$R_{\max}^{\boxed{m}} = \max\left\{\left|a_{1,2}^{\boxed{m}}\right|,\ \left|a_{i,i-1}^{\boxed{m}}\right| + \left|a_{i,i+1}^{\boxed{m}}\right|,\ 2 \leq i \leq N-m-1,\ \left|a_{N-m,N-m-1}^{\boxed{m}}\right|\right\}.$$

The largest eigenvalue of $A^{\boxed{m}}$ must lie to the left of the right end of the interval of size $R^{\boxed{m}}_{\max}$ and centered at the point given by the largest diagonal element of $A^{\boxed{m}}$:

$$\rho(A^{\boxed{m}}) \leq R^{\boxed{m}}_{\max} + \max_{1\leq i\leq N-m} \left|a^{\boxed{m}}_{i,i}\right|.$$

Thanks to the estimate of off-diagonal elements of matrix $M^{\boxed{m}}$ given in Proposition 2.6, $R^{\boxed{m}}_{\max} \leq \alpha/12$, so that

$$\rho(A^{\boxed{m}}) \leq \max_{1\leq i\leq N-m} \left|d^{\boxed{m}}_{i}\right| + \alpha \max_{1\leq i\leq N-m} \left|\mu^{\boxed{m}}_{i,i}\right| + \frac{\alpha}{12}.$$

The estimates of the diagonal elements of $D^{\boxed{m}}$ given in Proposition 2.3 and of the diagonal elements of $M^{\boxed{m}}$ given in Proposition 2.5 complete the proof. □

The estimate given in the previous proposition is valuable to obtain an overall bound for the growth with respect to N of the condition number of the matrices representing the discrete elliptic operators. Unfortunately, it is quite coarse and thus prevents a finer description of how the condition number depends on m for fixed N. For this purpose, in the following proposition we estimate sharper bounds depending on m for the largest eigenvalues.

Proposition 2.11. *The following bound applies to the spectral radius of matrix* $A^{\boxed{m}}$ *:*

$$\begin{aligned}\rho(A^{\boxed{m}}) \leq \frac{2(N-m)^2}{2N-m} + \alpha \min\Bigg(\frac{1}{4}, \frac{(N-m)^2}{(2N-m)(2N-m+1)(m+1)} \\ + \frac{(N-m)(N-m-1)}{(2N-m)(2N-m-2)(m+3)}\Bigg).\end{aligned} \tag{2.35}$$

Proof. The proof is identical to the previous one but it exploits the estimates provided in Propositions 2.8 and 2.9 instead of those in Propositions 2.5 and 2.6. □

This new estimate is finer since it describes the rapid decay of the mass matrix contribution as m grows and it will allow to better estimate how the condition number of $A^{\boxed{m}}$ depends on m for fixed N.

In estimating the eigenvalue lower bound care must be taken to avoid finding zero or negative estimates which would be in fact totally useless. Gershgorin-Hadamard theorem states that the minimum eigenvalue satisfies the following relationship:

$$\begin{aligned}\nu(A^{\boxed{m}}) &\geq \min_{1\leq i\leq N-m} \{a^{\boxed{m}}_{i,i} - \left|a^{\boxed{m}}_{i,i-1}\right| - \left|a^{\boxed{m}}_{i,i+1}\right|\} \\ &= \min_{1\leq i\leq N-m} \{d^{\boxed{m}}_{i} + \alpha(\mu^{\boxed{m}}_{i,i} - \left|\mu^{\boxed{m}}_{i,i-1}\right| - \left|\mu^{\boxed{m}}_{i,i+1}\right|)\}.\end{aligned} \tag{2.36}$$

This inequality provides us with the following estimate on the smallest eigenvalue of $A^{\boxed{m}}$.

Proposition 2.12. *For* $m \geq 0$, $\nu(A^{\boxed{m}}) \geq d^{\boxed{m}}_{1} = \dfrac{2}{2+m}$.

Proof. The proof follows directly from Eq. (2.36), and from Propositions 2.7 and 2.3. □

2.6. Condition number estimates

We are now in the position to estimate the behaviour of the maximum condition number associated with the matrices $A^{\boxed{m}}$ by the following

Lemma 2.1. *The maximum condition number of the matrices $A^{\boxed{m}}$, $m \geq 0$, is bounded from above by CN^2 as $N \to \infty$, where $C > 1/2$ is a constant independent of α.*

Proof. Thanks to Propositions 2.10 and 2.12, the condition number χ of matrix $A^{\boxed{m}}$ is such that

$$\chi(A^{\boxed{m}}) = \frac{\rho(A^{\boxed{m}})}{\nu(A^{\boxed{m}})} \leq \frac{\frac{2(N-m)^2}{2N-m} + \frac{\alpha}{4}}{\frac{2}{2+m}}.$$

Since $\frac{2(N-m)^2}{2N-m} \leq N$ and $m \leq N$, rearranging we can write

$$\chi(A^{\boxed{m}}) \leq \left(N + \frac{\alpha}{4}\right)\left(1 + \frac{N}{2}\right).$$

The thesis follows observing that the inequality

$$\left(N + \frac{\alpha}{4}\right)\left(\frac{N}{2} + 1\right) \leq CN^2$$

is satisfied for N sufficiently large if, and only if, $C > 1/2$.

It is now interesting to estimate the bound of the condition number as a function of m for fixed N. This is the aim of the following

Lemma 2.2. *For fixed N, the condition number of the matrices $A^{\boxed{m}}$ is bounded from above by a function of m as*

$$\chi(A^{\boxed{m}}) \leq \frac{\frac{2(N-m)^2}{2N-m} + \alpha \min\left(\frac{1}{4}, \frac{(N-m)^2}{(2N-m)(2N-m+1)(m+1)} + \frac{(N-m)(N-m-1)}{(2N-m)(2N-m-2)(m+3)}\right)}{\frac{2}{2+m}}. \tag{2.37}$$

Proof. The result follows by exploiting the finer bound for the maximum eigenvalue of $A^{\boxed{m}}$ stated by Proposition 2.11 and dividing it by the smallest eigenvalue in Proposition 2.12. □

The result of the previous Lemma is quite sharp and provides a good estimate of how the condition number varies as m is increased. The left plot in Fig. 2 provides a comparison between the estimate and the computed values of the condition numbers of different modes for $N = 128$ and $\alpha = 1000$. The right plot shows the values of the maximum condition number for different truncations N up to $N = 256$ and confirms the theoretically predicted bound $\propto N^2$.

The following Lemma completes the information about the behaviour of the maximum condition number of the matrices $A^{\boxed{m}}$ as N increases provided in the previous results.

Lemma 2.3. *There exists $N_C > 0$ such that, for $N > N_C$, the condition number of the matrices $A^{\boxplus}$, for $m = \epsilon N$, $0 < \epsilon < 1$ is bounded by $C N^2$, where C is a strictly positive constant satisfying $C > \epsilon(1-\epsilon)^2/(2-\epsilon)$.*

Proof. By introducing the assumption $m = \epsilon N$ in the bound provided by the previous Lemma we obtain

$$\chi(A^{\boxplus}) \leq \left(\frac{2(N-\epsilon N)^2}{2N-\epsilon N} + \frac{\alpha}{4}\right)\left(\frac{2+\epsilon N}{2}\right) = \left(\frac{(1-\epsilon)^2}{2-\epsilon} N + \frac{\alpha}{8}\right)(2+\epsilon N).$$

Since the coefficient of the dominant term N^2 in the right hand side of this inequality is $\epsilon(1-\epsilon)^2/(2-\epsilon) > 0$ the thesis easily follows. □

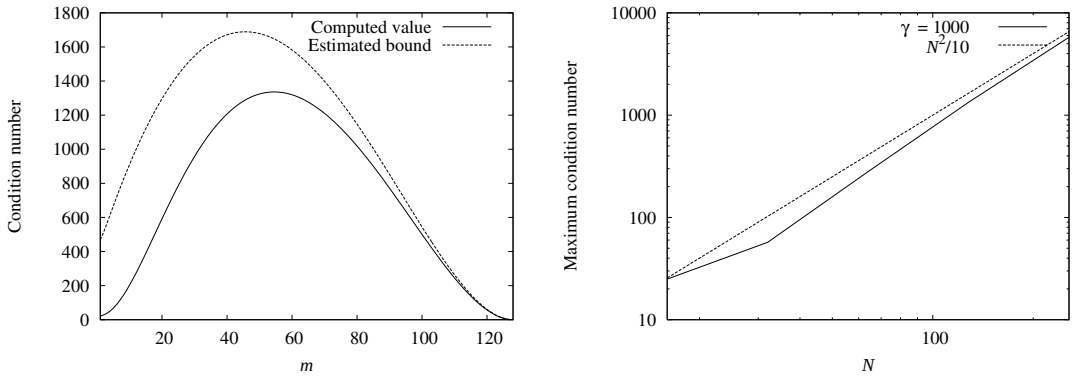

Figure 2: Left: Condition numbers of different modes for $N = 128$ and $\alpha = 1000$; comparison between the estimate provided in Lemma 2.2 and the computed condition numbers. Right: Maximum condition number for different truncations N, up to $N = 256$.

3. Spectral approximation in spherical coordinates

3.1. Dirichlet problem

Consider the scalar Poisson equation $-\nabla^2 u = f$ in spherical coordinates (r, θ, ϕ)

$$-\frac{1}{r^2}\partial_r\left(r^2\,\partial_r u\right) - \frac{\partial_\theta\left(\sin\theta\,\partial_\theta u\right)}{r^2\sin\theta} - \frac{\partial_\phi^2 u}{r^2\sin^2\theta} = f(r,\theta,\phi) \tag{3.1}$$

to be solved inside the unit sphere $r \leq 1$, supplemented by the homogeneous Dirichlet condition $u = 0$ over the spherical surface $r = 1$, since the radial coordinate r has been made already dimensionless. The solution is expanded in the Fourier series

$$\begin{aligned} u(r,\theta,\phi) =& u^0(r,\theta) + u^N(r,\theta)\cos(N\phi) \\ &+ 2\sum_{m=1}^{N-1}\left(u^m(r,\theta)\cos(m\phi) - u^{-m}(r,\theta)\sin(m\phi)\right). \end{aligned} \tag{3.2}$$

In the space of the Fourier coefficients, the three-dimensional equation reduces to a set of uncoupled two-dimensional elliptic equations

$$-\partial_m^2 u^m = f^m(r,\theta), \tag{3.3}$$

where

$$\partial_m^2 \equiv \frac{1}{r^2}\partial_r\big(r^2\partial_r\ \big) + \frac{\partial_\theta\big(\sin\theta\,\partial_\theta\ \big)}{r^2\sin\theta} - \frac{m^2}{r^2\sin^2\theta}, \tag{3.4}$$

for the Fourier expansion coefficients, $u^m(r,\theta)$ with $-N+1 \le m \le N$, also supplemented by homogeneous Dirichlet conditions $u^m = 0$, for $r = 1$.

The unknown variable $u^m(r,\theta)$ is then expanded in terms of the normalized associated Legendre functions $\hat{P}_\ell^m(\cos\theta)$

$$u^m(r,\theta) = \sum_{\ell=|m|}^{N} u_\ell^m(r)\,\hat{P}_\ell^{|m|}(\cos\theta), \qquad -N+1 \le m \le N, \tag{3.5}$$

with the basis functions $\hat{P}_\ell^m(z)$, $z = \cos\theta$, related to the standard associated Legendre functions $P_\ell^m(z)$ by

$$\hat{P}_\ell^m(z) \equiv \sqrt{\frac{2\ell+1}{2}\,\frac{(\ell-m)!}{(\ell+m)!}}\,P_\ell^m(z). \tag{3.6}$$

The 2-D elliptic equation (3.3), once expressed in weak form, reduces to the following set of equations for the modal unknowns $u_\ell^m(r)$:

$$\int_0^1 v\big[-D_r\big(r^2 D_r u_\ell^m\big) + \ell(\ell+1)\,u_\ell^m\big]\,dr = \int_0^1 r^2\,v(r)\,f_\ell^m(r)\,dr, \tag{3.7}$$

where $D_r = d/dr$ while $v(r)$ is a suitable weighting function and

$$f_\ell^m(r) = \int_0^\pi f^m(r,\theta)\,\hat{P}_\ell^{|m|}(\cos\theta)\,\sin\theta\,d\theta. \tag{3.8}$$

Eq. (3.7) is supplemented by the homogeneous boundary condition $u_\ell^m(1) = 0$, so that $v(r)$ must be chosen vanishing for $r = 1$. Thus the integration by parts yields

$$\int_0^1 \big[r^2(D_r v)(D_r u_\ell^m) + \ell(\ell+1)\,v u_\ell^m\big]\,dr = \int_0^1 r^2\,v(r)\,f_\ell^m(r)\,dr. \tag{3.9}$$

Finally, the modal unknown $u_\ell^m(r)$ is expanded according to

$$u_\ell^m(r) = \sum_{i=1}^{N-\ell} u_i^{m;\ell}\,B_i^\ell(r), \tag{3.10}$$

where the basis functions $B_i^\ell(r) \equiv R_i^\ell(s)$, with $s = 2r^2 - 1$, are defined in terms of Jacobi polynomials $P_{i-1}^{(1,\ell+\frac{1}{2})}(s)$ as follows [2]

$$R_i^\ell(s) = \frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s), \qquad i = 1, 2, \cdots. \tag{3.11}$$

Thus the complete spectral expansion of the Fourier component $u^m(r,\theta)$ of the unknown is given by the relation

$$u^m(r,\theta) = \sum_{\ell=|m|}^{N} \left[\sum_{i=1}^{N-\ell} u_i^{m;\ell}\, B_i^\ell(r)\right] \hat{P}_\ell^{|m|}(\cos\theta). \tag{3.12}$$

The corresponding spectral version of the weak equation (3.9) for the modal unknowns $u_\ell^m(r)$ consists in a linear system of algebraic equations of the form

$$D^{\mathbb{Z}}\, u^{\mathbb{Z},m} = f^{\mathbb{Z},m}, \tag{3.13}$$

where $u^{\mathbb{Z},m} = \{u_i^{m;\ell},\ i = 1, 2, \cdots, N-\ell\}$ and the matrix $D^{\mathbb{Z}}$ has elements defined by (using the mapped variable $s = 2r^2 - 1$)

$$d_{i,i'}^{\mathbb{Z}} = \int_{-1}^{1} \left\{4\Big(\frac{1+s}{2}\Big)^{\frac{3}{2}} \big[D_s R_i^\ell(s)\big]\big[D_s R_{i'}^\ell(s)\big] + \frac{\ell(\ell+1)}{4}\Big(\frac{1+s}{2}\Big)^{-\frac{1}{2}} R_i^\ell(s) R_{i'}^\ell(s)\right\} ds, \tag{3.14}$$

for $1 \le (i, i') \le N - \ell$. The order of a matrix with superscript $^{\mathbb{Z}}$ is $N - \ell$, and is therefore $N, (N-1), \cdots, 2, 1$ for $\ell = 0, 1, 2, \cdots, N-1$, respectively.

Fig. 3 shows basis functions $B_i^\ell(r)\hat{P}_\ell^m(\cos\theta)$ plotted for $\phi = 0$ and $N = 4$.

3.2. Diagonal stiffness matrices

The stiffness matrix $D^{\mathbb{Z}}$ is diagonal and this can be proved as suggested by P. W. Livermore in a private communication (2008) starting from Eq. (3.7) and using the orthogonality relation and differential equation of Jacobi polynomials. In fact, we have:

Proposition 3.1. *For any $\ell \ge 0$, the stiffness matrix $D^{\mathbb{Z}}$ is diagonal and its nonzero elements are given by*

$$d_i^{\mathbb{Z}} = \frac{2i^2}{2i + \ell + \frac{1}{2}}, \qquad \text{for } 1 \le i \le N - \ell. \tag{3.15}$$

Proof. Consider the expansion component $R_i^\ell(s)$ of the two terms inside the square brackets:

$$-4D_s\left\{\left(\frac{1+s}{2}\right)^{\frac{3}{2}} D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right]\right\} + \frac{\ell(\ell+1)}{4}\,\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell-1}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)$$

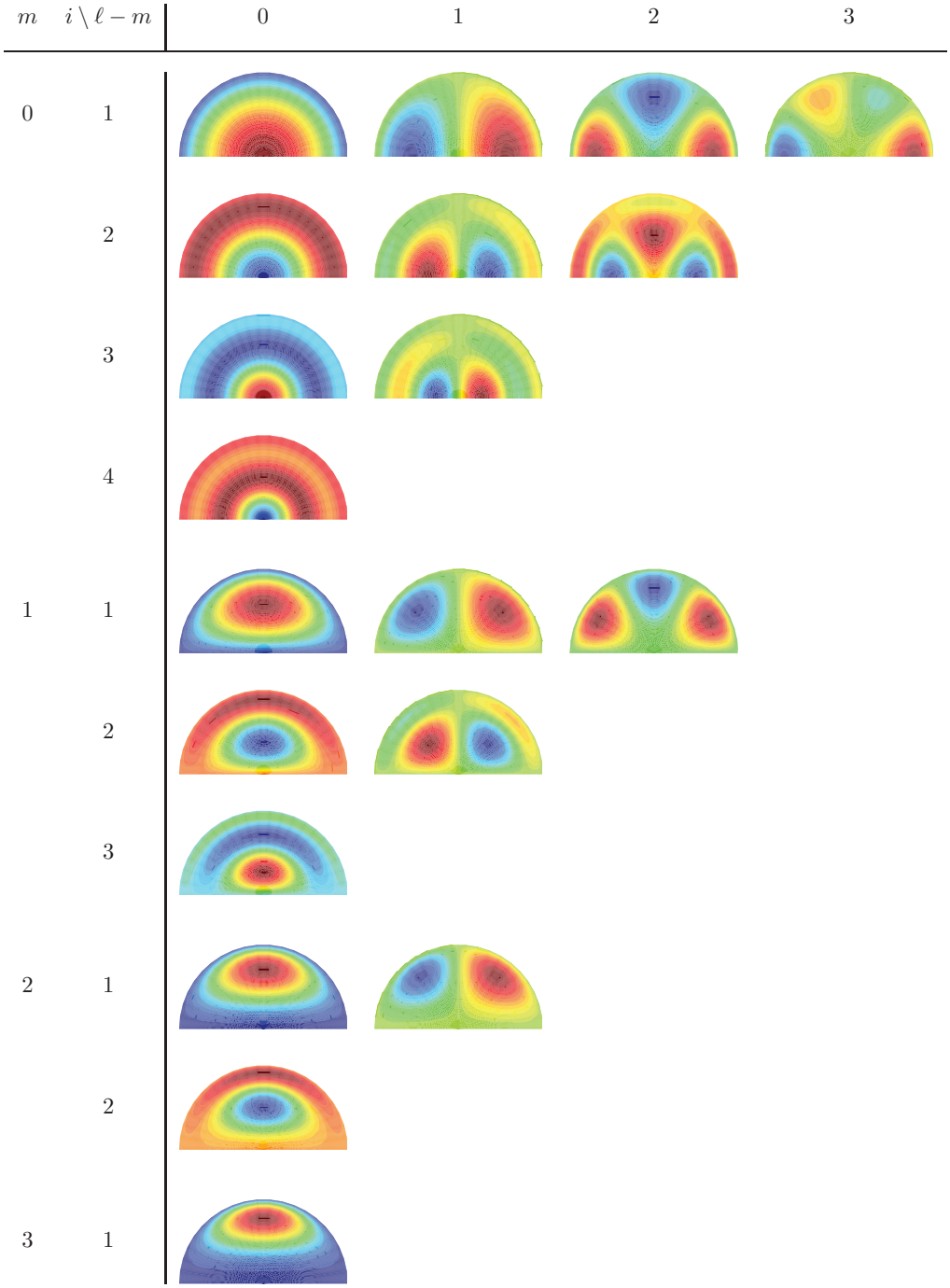

Figure 3: Basis functions for the Dirichlet problem inside a sphere $B_i^\ell(r)\hat{P}_\ell^m(\cos\theta)$, for $N = 4$.

and focus on the first contribution. By developing the internal derivative, we obtain

$$\begin{aligned}
&D_s\left\{\left(\frac{1+s}{2}\right)^{\frac{3}{2}} D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right]\right\}\\
=&D_s\left\{\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell+3}{2}} D_s P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right.\\
&\qquad\left.+\frac{\ell}{4}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell+1}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)-\frac{1}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell+3}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right\}.
\end{aligned}$$

By evaluating the external derivative we obtain

$$\begin{aligned}
&D_s\left\{\left(\frac{1+s}{2}\right)^{\frac{3}{2}} D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right]\right\}\\
=&\frac{1-s^2}{4}\left(\frac{1+s}{2}\right)^{\frac{\ell+1}{2}} D_s^2 P_{i-1}^{(1,\ell+\frac{1}{2})}(s)+\frac{\ell(\ell+1)}{16}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell-1}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\\
&\quad+\left[\frac{2\ell+3}{4}\frac{1-s}{2}-\frac{1+s}{2}\right]\left(\frac{1+s}{2}\right)^{\frac{\ell+1}{2}} D_s P_{i-1}^{(1,\ell+\frac{1}{2})}(s)-\frac{2\ell+3}{8}\left(\frac{1+s}{2}\right)^{\ell+\frac{1}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s).
\end{aligned}$$

By taking into account the coefficients -4 in front of the original term, the term $\propto \ell(\ell+1)$ in the third line above is found to cancel with the second contribution of the starting expression. This result leads to the following replacement, after reordering the remaining terms:

$$\begin{aligned}
&D_s\left\{\left(\frac{1+s}{2}\right)^{\frac{3}{2}} D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right]\right\} \longrightarrow \frac{1}{4}\left(\frac{1+s}{2}\right)^{\frac{\ell+1}{2}}\\
&\quad\times\left\{(1-s^2)D_s^2 P_{i-1}^{(1,\ell+\frac{1}{2})}(s)+\left[\ell-\tfrac{1}{2}-(\ell+\tfrac{7}{2})s\right]D_s P_{i-1}^{(1,\ell+\frac{1}{2})}(s)-(\ell+\tfrac{3}{2})P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right\}.
\end{aligned}$$

By the Jacobi equation (2.23) with $m=\ell+\frac{1}{2}$, the replacement above simplifies to

$$\begin{aligned}
&D_s\left\{\left(\frac{1+s}{2}\right)^{\frac{3}{2}} D_s\left[\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\frac{\ell}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\right]\right\}\\
&\qquad\longrightarrow -\frac{i(i+\ell+\frac{1}{2})}{4}\left(\frac{1+s}{2}\right)^{\frac{\ell+1}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s).
\end{aligned}$$

As a consequence, the stiffness matrix elements can be characterized more simply as

$$d_{i,i'}^{\boxed{\ell}} = i'(i'+\ell+\tfrac{1}{2})\int_{-1}^{1}\frac{1-s}{2}\left(\frac{1+s}{2}\right)^{\ell+\frac{1}{2}} P_{i-1}^{(1,\ell+\frac{1}{2})}(s)\,P_{i'-1}^{(1,\ell+\frac{1}{2})}(s)\,ds,$$

for $i\geq 1$, which, by virtue of the orthogonality condition (2.18) with $\beta=\ell+\frac{1}{2}$, gives

$$d_{i,i'}^{\boxed{\ell}} = \frac{2i^2}{2i+\ell+\frac{1}{2}}\,\delta_{i,i'}, \qquad \text{for } 1\leq i\leq N-\ell.$$

□

3.3. Tridiagonal mass matrices

In the solution of evolutionary problems, such as the heat equation, one has to introduce the Helmholtz operator and this leads to the radial differential operator

$$A^{[\ell]} = D^{[\ell]} + \alpha M^{[\ell]}, \tag{3.16}$$

where $\alpha \propto 1/\Delta t$ and $M^{[\ell]}$ denotes the mass matrix, with elements defined by

$$\mu^{[\ell]}_{i,i'} = \frac{1}{4}\int_{-1}^{1} \left(\frac{1+s}{2}\right)^{\frac{1}{2}} R^{\ell}_{i}(s) R^{\ell}_{i'}(s)\, ds, \qquad 1 \le (i,i') \le N-\ell. \tag{3.17}$$

Proposition 3.2. *For any $\ell \ge 0$, the mass matrix $M^{[\ell]}$ is tridiagonal and the values of its nonzero elements are*

$$\begin{aligned} \mu^{[\ell]}_{i,i} &= \frac{i^2}{\left(2i+\ell-\frac{1}{2}\right)\left(2i+\ell+\frac{1}{2}\right)\left(2i+\ell+\frac{3}{2}\right)}, && \text{for } 1 \le i \le N-\ell, \\ \mu^{[\ell]}_{i,i+1} &= \frac{-i(i+1)}{2\left(2i+\ell+\frac{1}{2}\right)\left(2i+\ell+\frac{3}{2}\right)\left(2i+\ell+\frac{5}{2}\right)}, && \text{for } 1 \le i \le N-\ell-1, \end{aligned} \tag{3.18}$$

and $\mu^{[\ell]}_{i,i-1} = \mu^{[\ell]}_{i-1,i}$, for $2 \le i \le N-\ell$.

Proof. Thanks to the definition of the basis functions $R^{\ell}_{i}(s)$ we have

$$\mu^{[\ell]}_{i,i'} = \frac{1}{4}\int_{-1}^{1} \left(\frac{1+s}{2}\right)^{\ell+\frac{1}{2}} \left(\frac{1-s}{2}\right)^{2} P^{(1,\ell+\frac{1}{2})}_{i-1}(s)\, P^{(1,\ell+\frac{1}{2})}_{i'-1}(s)\, ds,$$

for $1 \le (i,i') \le N-\ell$. The tridiagonal character of the mass matrix $M^{[\ell]}$ follows from the recurrence relation (2.21) with $m = \ell + \frac{1}{2}$, and the orthogonality of the Jacobi polynomials $P^{(0,\ell+\frac{1}{2})}_{i}(s)$. The nonzero elements of the mass matrix $M^{[\ell]}$ are obtained by means of a direct evaluation. □

3.4. Matrix properties

The properties of the stiffness and mass matrices for the spherical geometry closely resemble those of their cylindrical counterpart.

3.4.1. Stiffness matrix

The properties that characterize the stiffness matrices for the cylindrical coordinate system hold also for the spherical one.

Proposition 3.3. *The diagonal elements $d^{[\ell]}_{i}$ of $D^{[\ell]}$, for fixed ℓ are increasing with i for $i \ge 1$, hence*

$$\min_{1 \le i \le N-\ell} d^{[\ell]}_{i} = d^{[\ell]}_{1} \equiv \frac{4}{2\ell+5}, \qquad \ell \ge 0. \tag{3.19}$$

Proof. The proposition is a trivial consequence of the definition (3.15). □

It is trivial also in this case to prove the following

Proposition 3.4. *The diagonal elements $d_i^{\square}$ of $D^{\square}$ are decreasing with $\ell \geq 0$, for fixed $i \geq 1$ and, asymptotically, $d_i^{\square}$ behave as ℓ^{-1} as $\ell \to \infty$.*

3.4.2. Mass matrix

Also in this case we investigate the ordering of the elements. Starting from diagonal elements we can prove that they are bounded from above, in fact the following proposition holds:

Proposition 3.5. $\forall \ell \geq 0$ *and* $i \geq 1$, $\mu_{i,i}^{\square} \leq \frac{8}{105}$.

Proof. The diagonal elements $\mu_{i,i}^{\square}$ of $M^{\square}$ are decreasing as ℓ is increased at fixed i. So, we can simply analyse the case $\ell = 0$. In this case, $\mu_{i,i}^{\square}$ are decreasing with i, provided the following inequality

$$\frac{i^2}{(2i-\frac{1}{2})(2i+\frac{1}{2})(2i+\frac{3}{2})} \geq \frac{(i+1)^2}{(2i+\frac{3}{2})(2i+\frac{5}{2})(2i+\frac{7}{2})}$$

is satisfied. But this reduces to $16i^3 + 20i^2 + 2i + 1 \geq 0$. Since the polynomial on the left hand side has no positive real roots thanks to Descartes sign rule, the inequality is satisfied for all $i \geq 1$. A direct calculation for $i = 1$ and $\ell = 0$ shows that the maximum is $\mu_{1,1}^{\square} = \frac{8}{105}$, *q.e.d.*. □

Similarly the following proposition can be proved.

Proposition 3.6. $\forall \ell \geq 0$ *and* $i \geq 1$, $\left|\mu_{i,i+1}^{\square}\right| \leq \frac{8}{315}$.

Finally, the following proposition holds, as in the cylindrical case,

Proposition 3.7. *For any row $i \geq 1$ and for $\ell \geq 0$, the diagonal element of the mass matrix is greater then the sum of the modulus of the offdiagonal elements:*

$$\mu_{1,1}^{\square} > \left|\mu_{1,2}^{\square}\right|, \mu_{i,i}^{\square} > \left|\mu_{i,i-1}^{\square}\right| + \left|\mu_{i,i+1}^{\square}\right|, \qquad \textit{for} \quad 2 \leq i \leq N - \ell - 1,$$

and

$$\mu_{N-\ell,N-\ell}^{\square} > \left|\mu_{N-\ell,N-\ell-1}^{\square}\right|.$$

Proof. The proof is simply obtained by substituting $m = l + \frac{1}{2}$ in the proof of Proposition 2.7. □

As will be shown later, these estimates are useful to describe how the condition number depends on the parameter N. However, to investigate also how the condition number depends on ℓ for fixed N, the sharper estimates reported in the following propositions will be needed.

Proposition 3.8. *For $\ell \geq 0$ and $i \geq 1$, the diagonal elements of the mass matrix are bounded according to*

$$\mu_{i,i}^{\boxplus} \leq \min\left(\frac{8}{105}, \frac{(N-\ell)^2}{\left(2N-\ell+\frac{1}{2}\right)\left(2N-\ell+\frac{3}{2}\right)\left(\ell+\frac{3}{2}\right)}\right). \tag{3.20}$$

Proof. The proof follows closely the one provided in Proposition 2.8 and exploiting the result of Proposition 3.5. □

Proposition 3.9. *For $\ell \geq 0$ and $i \geq 1$, the extra-diagonal elements of the mass matrix are bounded according to*

$$\left|\mu_{i,i+1}^{\boxplus}\right| \leq \min\left(\frac{8}{315}, \frac{(N-\ell-1)(N-\ell)}{2\left(2N-\ell-\frac{3}{2}\right)\left(2N-\ell+\frac{1}{2}\right)\left(\ell+\frac{7}{2}\right)}\right). \tag{3.21}$$

Proof. The proof follows closely the one provided in Proposition 2.9, rewriting the matrix element expression as

$$\frac{1}{2\left[2+\left(\ell+\frac{1}{2}\right)/i\right]\left(2i+\ell+\frac{3}{2}\right)\left[2+\left(\ell+\frac{1}{2}\right)/(i+1)\right]}, \tag{3.22}$$

and it exploits the result of Proposition 3.6. □

3.5. Eigenvalue estimates

We can follow the same procedure adopted for the cylindrical geometry to prove the following propositions which provide estimates on the matrix elements. Also in this case we proceed by exploiting Gershgorin-Hadamard theorem, which applied to the matrices $A^{\boxplus}$ leads to

$$\begin{cases} \left|\lambda - a_{1,1}^{\boxplus}\right| \leq \left|a_{1,2}^{\boxplus}\right|, \\ \left|\lambda - a_{i,i}^{\boxplus}\right| \leq \left|a_{i,i-1}^{\boxplus}\right| + \left|a_{i,i+1}^{\boxplus}\right|, \qquad \text{for } 2 \leq i \leq N-\ell-1, \\ \left|\lambda - a_{N-\ell,N-\ell}^{\boxplus}\right| \leq \left|a_{N-\ell,N-\ell-1}^{\boxplus}\right|, \end{cases} \tag{3.23}$$

where

$$a_{i,i}^{\boxplus} = \left[2 + \frac{\alpha}{\left(2i+\ell+\frac{1}{2}\right)^2 - 1}\right] \frac{i^2}{2i+\ell+\frac{1}{2}}, \qquad 1 \leq i \leq N-\ell, \tag{3.24}$$

while $a_{i,i+1}^{\boxplus}$ are simply the extradiagonal elements of $M^{\boxplus}$ multiplied by α:

$$a_{i,i+1}^{\boxplus} = \frac{-\alpha\, i(i+1)}{2\left(2i+\ell+\frac{1}{2}\right)\left(2i+\ell+\frac{3}{2}\right)\left(2i+\ell+\frac{5}{2}\right)}, \qquad 1 \leq i \leq N-\ell-1, \tag{3.25}$$

and $a_{i,i-1}^{\boxplus} = a_{i-1,i}^{\boxplus}$, for $2 \leq i \leq N-\ell$.

Then the following propositions can be proved.

3.5.1. Largest and smallest eigenvalue estimate

Proposition 3.10. *The following bound applies to the spectral radius $\rho(A^{\boxplus})$ of matrix $A^{\boxplus}$:*

$$\rho(A^{\boxplus}) \leq \frac{2(N-\ell)^2}{2N-\ell+\frac{1}{2}} + \frac{8\alpha}{63}. \tag{3.26}$$

Proof. The proof closely follows the one for Proposition 2.10. □

The estimate given in the previous proposition is valuable to estimate an overall bound for the growth with respect to N of the condition number of the matrices representing the discrete elliptic operators. Unfortunately, it is quite coarse and thus prevents a finer description of how the condition number depends on ℓ for fixed N. For this purpose, in the following proposition, we estimate sharper bounds for the largest eigenvalue depending on ℓ.

Proposition 3.11. *The following bound applies to the spectral radius of matrix $A^{\boxplus}$:*

$$\begin{aligned}\rho(A^{\boxplus}) \leq & \frac{2(N-\ell)^2}{2N-\ell+\frac{1}{2}} \\ & + \alpha \min\left(\frac{8}{63}, \frac{N-\ell}{2N-\ell+\frac{1}{2}}\left[\frac{N-\ell}{\left(2N-\ell+\frac{3}{2}\right)\left(\ell+\frac{3}{2}\right)} + \frac{N-\ell-1}{\left(2N-\ell-\frac{3}{2}\right)\left(\ell+\frac{7}{2}\right)}\right]\right).\end{aligned} \tag{3.27}$$

Proof. The proof is identical to the previous one but it exploits the estimates provided in Propositions 3.8 and 3.9 instead of those in Propositions 3.5 and 3.6. □

This new estimate is finer since it describes the rapid decay of the mass matrix contribution as ℓ grows and it will allow to better estimate how the condition number of $A^{\boxplus}$ depends on ℓ for fixed N.

To estimate a lower bound for the eigenvalues we resort again to Gershgorin-Hadamard theorem which insures that the minimum eigenvalue satisfies the following relationship:

$$\begin{aligned}\nu(A^{\boxplus}) &\geq \min_{1\leq i\leq N-\ell}\left\{a_{i,i}^{\boxplus} - \left|a_{i,i-1}^{\boxplus}\right| - \left|a_{i,i+1}^{\boxplus}\right|\right\} \\ &= \min_{1\leq i\leq N-\ell}\left\{d_i^{\boxplus} + \alpha\left(\mu_{i,i}^{\boxplus} - \left|\mu_{i,i-1}^{\boxplus}\right| - \left|\mu_{i,i+1}^{\boxplus}\right|\right)\right\}.\end{aligned} \tag{3.28}$$

The following inequality provides us with an estimate of the smallest eigenvalue of $A^{\boxplus}$.

Proposition 3.12. *For $\ell \geq 0$, $\nu(A^{\boxplus}) \geq d_1^{\boxplus} = \frac{4}{2\ell+5}$.*

Proof. The proof follows directly from Eq. (3.28), and from Propositions 3.7 and 3.3. □

3.6. Condition number estimates

As done for the cylindrical coordinates we can now provide estimates on the maximum condition number of the discrete second-order operators by the two following propositions.

Lemma 3.1. *The maximum condition number of the matrices $A^{\boxplus}$, $\ell \geq 0$ is bounded by CN^2 as $N \to \infty$, where $C > 1/2$ is a constant independent of α.*

Proof. We proceed exactly as in the proof of Lemma 2.1. Thanks to Propositions 3.10 and 3.12, the condition number χ of matrix $A^{\boxplus}$ is such that

$$\chi(A^{\boxplus}) \leq \left(\frac{2(N-\ell)^2}{2N-\ell+\frac{1}{2}} + \frac{8\alpha}{63} \right) \Bigg/ \frac{4}{2\ell+5}.$$

Since $\frac{2(N-\ell)^2}{2N-\ell+1/2} \leq N$ and $\ell \leq N$, rearranging we can write

$$\chi(A^{\boxplus}) \leq \left(N + \frac{8\alpha}{63} \right) \left(\frac{N}{2} + \frac{5}{4} \right).$$

The thesis follows observing that the inequality

$$\left(N + \frac{8\alpha}{63} \right) \left(\frac{N}{2} + \frac{5}{4} \right) \leq CN^2$$

is satisfied for N sufficiently large if, and only if, $C > 1/2$. □

By exploiting the finer bound for the maximum eigenvalue of $A^{\boxplus}$ stated by Proposition 3.11, it can be estimated now how the condition number $\chi(A^{\boxplus})$ depends on ℓ.

Lemma 3.2. *For fixed N, the condition number of the matrices $A^{\boxplus}$ is bounded from above by the curve*

$$\chi(A^{\boxplus}) \leq \frac{\frac{2(N-\ell)^2}{2N-\ell+\frac{1}{2}} + \alpha \min\left(\frac{8}{63} \frac{N-\ell}{2N-\ell+\frac{1}{2}} \left[\frac{N-\ell}{\left(2N-\ell+\frac{3}{2}\right)\left(\ell+\frac{3}{2}\right)} + \frac{N-\ell-1}{\left(2N-\ell-\frac{3}{2}\right)\left(\ell+\frac{7}{2}\right)} \right] \right)}{\frac{4}{2\ell+5}}. \quad (3.29)$$

Proof. It is sufficient to divide the upper bound for the maximum eigenvalue in Proposition 3.11 by the lower bound for the minimum eigenvalue in Proposition 3.12. □

As for cylindrical coordinates, also in this case the previous Lemma provides a good estimate of how the condition number varies as ℓ is increased. The left plot in Fig. 4 provides the comparison between the estimate and the computed values of the condition numbers of the different modes for $N = 128$ and $\alpha = 1000$. The right plot shows the values of the maximum condition number for different truncations N up to $N = 256$ and confirms the theoretically predicted bound $\propto N^2$.

Lemma 3.3. *There exists $N_C > 0$ such that, for $N > N_C$, the condition number of the matrices $A^{\boxplus}$, for $\ell = \epsilon N$, $0 < \epsilon < 1$ is bounded by $C N^2$, where C is a strictly positive constant satisfying $C > \epsilon(1-\epsilon)^2/(2-\epsilon)$.*

Proof. Note that

$$\chi(A^{\mathbb{Z}}) \leq \left(\frac{2(N-\epsilon N)^2}{2N-\epsilon N+\frac{1}{2}}+\frac{8\alpha}{63}\right)\left(\frac{5+2\epsilon N}{4}\right)$$
$$\leq \left(\frac{2(N-\epsilon N)^2}{2N-\epsilon N}+\frac{8\alpha}{63}\right)\left(\frac{5+2\epsilon N}{4}\right)=\left(\frac{(1-\epsilon)^2}{4-2\epsilon}N+\frac{2\alpha}{63}\right)(5+2\epsilon N).$$

Since the coefficient of the dominant term N^2 in the right hand side of this inequality is $\epsilon(1-\epsilon)^2/(2-\epsilon)>0$ the thesis easily follows. □

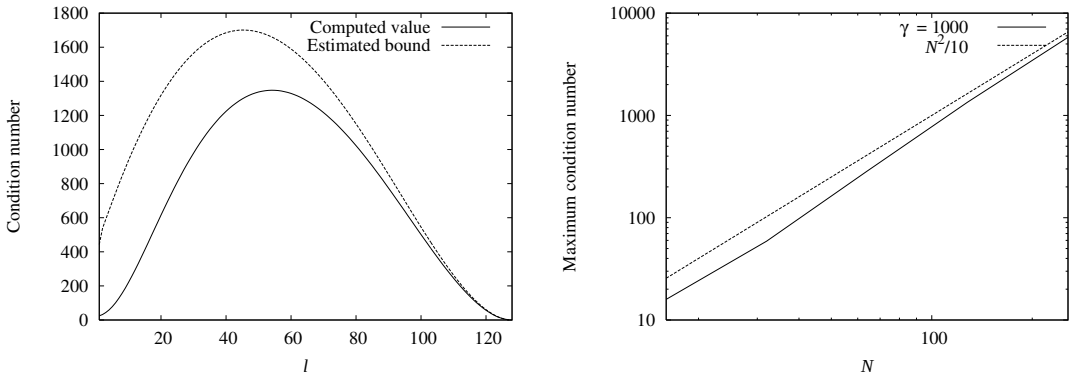

Figure 4: Left: Condition numbers of different modes for $N=128$ and $\alpha=1000$; comparison between the estimate provided in Lemma 3.2 and the computed condition numbers. Right: Maximum condition number for different truncations N, up to $N=256$.

4. Conclusion

This paper has focused on the radial discretization in the solution of the Dirichlet problem for second order elliptic operators in cylindrical and spherical domains by means of spectral methods. While the spectral representation of the other spatial variables has been solved definitely since a long time by harmonic analysis and by the associated Legendre functions over the sphere, a proper treatment for the radial variable when it reaches the cylindrical axis or the sphere centre has been achieved only in the last years. A serious difficulty is encountered in fact when trying to formulate appropriate bases capable of dealing with the singularity of the cylindrical and spherical coordinates in a mathematically satisfactory way.

A few bases have been proposed which seem to solve the axis and centre problem, but no rigorous results have been provided so far concerning the issue of the actual stability of elliptic operators discretized exploiting such bases. This property is fundamental from the standpoint of the approximation theory and it has been investigated in the present paper

in a rigorous way by establishing the optimal conditioning of the basis proposed by the authors in [2] for spherical coordinates and in [3] for cylindrical ones.

The reported results show that the condition number of the spectral operators arising from the radial discretization of second order elliptic problems (2.1) in cylindrical coordinates and (3.1) in spherical coordinates is bounded by the square of the truncation number when the solution is expanded by means of the basis functions (2.8) and (3.11), respectively. This result is optimal in the context of a spectral approximation to second order elliptic problems, and extends the results reported in [8] for Cartesian coordinates and a Chebyshev basis to the new bases built upon Jacobi polynomials. Moreover, a sharp estimate of how the condition number depends on the wavenumber of the spherical harmonics is also provided and assessed by comparison with numerical results.

By virtue of the sparsity and favourable conditioning of the resulting discrete operators, the basis functions proposed in [2] and [3], seem the best choice to discretize second order elliptic problems in cylindrical and spherical regions, respectively, especially as far as Dirichlet boundary conditions are concerned.

Acknowledgments We wish to thank an anonymous referee whose criticisms helped us improve the content of the paper.

References

[1] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, New York, 1965.

[2] F. Auteri and L. Quartapelle, Spectral solvers for spherical elliptic problems, *J. Comput. Phys.*, **227** (2007), pp. 36–54.

[3] F. Auteri and L. Quartapelle, Spectral elliptic solvers in a finite cylinder, *Commun. Comput. Phys.*, **5** (2009), pp. 426–441.

[4] C. Bernardi, M. Dauge and Y. Maday, *Spectral Methods for Axisymmetric Domains*, Elsevier, Paris, 1999.

[5] J. B. Boyd, *Chebyshev and Fourier Spectral Methods*, Dover Publications, Second Edition, Mineola, New York, 2001.

[6] K. E. Gustafson, *Introduction to Partial Differential Equations and Hilbert Space Methods*, Dover Publications, Third Edition, Calcutta and Charleston, Illinois, 1993.

[7] D. B. Haidvogel and T. A. Zang, The accurate solution of Poisson's equation by expansion in Chebuyshev polynomials, *J. Comput. Phys.*, **30** (1979), pp. 167–180.

[8] W. Heinrichs, Improved condition number for spectral methods, *Math. Comp.*, **53** (1989), pp. 103–119.

[9] P. Le Quéré and J. Pécheux, A three-dimensional pseudo-spectral algorithm for the computation of convection in a rotating annulus, *Spectral and High Order Methods for Partial Differential Equations*, Proceedings of the ICOSAHOM '89 Conference, Villa Olmo, Como, Italy, 26–29 June 1989, pp. 261–271.

[10] H. R. Lewis and P. M. Bellan, Physical constraints on the coefficients of Fourier expansions in cylindrical coordinates, *Journal of Mathematical Physics*, **31** (1990), pp. 2592–2596.

[11] P. W. Livermore, C. A. Jones and S. J. Worland, Spectral radial basis functions for full sphere computations, *J. Comput. Phys.*, **227** (2007), pp. 1209–1224.

[12] T. MATSUSHIMA AND P. S. MARCUS, A spectral method for polar coordinates, *J. Comput. Phys.*, **120** (1995), pp. 365–374.
[13] J. SHEN, Efficient spectral–Galerkin method. I. Direct solvers of second- and fourth-order equations using Legendre polynomials, *SIAM J. Sci. Comput.*, **15**, (1994) pp. 1489–1505.
[14] J. SHEN, Efficient spectral–Galerkin methods. III. Polar and cylindrical geometries, *SIAM J. Sci. Comput.*, **18** (1997), pp. 74–87.
[15] J. SHEN, Efficient spectral–Galerkin methods. IV. Spherical geometries, *SIAM J. Sci. Comput.*, **20** (1999), pp. 1438–1455.
[16] W. T. M. VERKLEY, A pseudo-spectral model for two-dimensional incompressible flow in a circular basin. I. Mathematical formulation, *J. Comput. Phys.*, **136** (1997), pp. 100–114.
[17] W. T. M. VERKLEY, A pseudo-spectral model for two-dimensional incompressible flow in a circular basin. II. Numerical examples, *J. Comput. Phys.*, **136** (1997), pp. 115–131.