SIAM J. SCI. COMPUT.
Vol. 20, No. 4, pp. 1438–1455

EFFICIENT SPECTRAL-GALERKIN METHODS IV. SPHERICAL GEOMETRIES*

JIE SHEN†

Abstract. Fast spectral-Galerkin algorithms are developed for elliptic equations on the sphere. The algorithms are based on a double Fourier expansion and have quasi-optimal (optimal up to a logarithmic term) computational complexity. Numerical experiments indicate that they are significantly more efficient and/or accurate when compared with the algorithms based on spherical harmonics and on finite difference. Extensions to problems in spherical layers and to vector equations are also discussed.

Key words. spectral-Galerkin, spherical harmonics, double Fourier series, spherical geometry

AMS subject classifications. 65N35, 65N22, 65F05, 35J05

PII. S1064827597317028

1. Introduction.

This paper is the fourth and final part of a sequential work on efficient spectral-Galerkin methods for elliptic problems. In the first and second parts [14, 15], we presented efficient algorithms for elliptic equations in rectangular domains. In the third part [17], we dealt with polar and cylindrical geometries. In this part, we shall develop efficient spectral-Galerkin algorithms for elliptic equations in spherical geometries, which have important applications in many fields of science and engineering, especially in computational fluid dynamics, climate modeling, and numerical weather prediction.

The most natural way to deal with spherical geometries is to use spherical coordinates and spherical harmonic functions. Although the spherical harmonic functions, being the eigenfunctions of the spherical Laplace operator, offer a number of distinct advantages, such as uniform resolution on the sphere and simplicity for linear, constant-coefficient problems, they have a notorious drawback due to the lack of a fast algorithm for the analysis (from the function values at a set of appropriate points to the coefficients of its spherical harmonic expansion) and synthesis (the inverse operation of analysis) in spherical harmonic expansions. Even though this situation may be improved in light of some recent contributions (see, for instance, [6, 10]), it is still unlikely that spherical harmonic transforms can be as efficient as the fast Fourier transform (FFT). An alternative approach is to use double Fourier series whose analysis and synthesis can be done by the FFTs. The advantages and disadvantages of using double Fourier series as opposed to using spherical harmonic functions were discussed in great detail by Orszag [12] (see also Boyd [2, 4]).

There have been a number of attempts at using double Fourier series for solving Poisson equations on the sphere. Orszag [12] (see also the references therein for other earlier and less successful attempts) suggested an efficient algorithm based on the tau method for solving the Poisson equation by using double Fourier series. Later, Yee [20] presented a similar algorithm with all the details needed for implementation. Unfortunately, although very efficient and easy to implement, Yee's algorithm is not entirely

*Received by the editors February 31, 1997; accepted for publication (in revised form) October 31, 1997; published electronically March 30, 1999. This work was supported in part by NSF grant DMS-9623020.

http://www.siam.org/journals/sisc/20-4/31702.html

†Department of Mathematics, Penn State University, University Park, PA 16802 (shen@math.psu.edu).

correct, for the pole condition (10) in [20] is enforced incorrectly by determining u_{0m} from the relation (15) in [20], which may lead to inaccurate results. Some interesting pseudospectral approaches can be found in [7, 8]. However, the double Fourier series approach has not been used much in practice, partly because most computations on the sphere were restricted, by computing resource in the past, to using coarse to moderate resolutions for which not much could be gained by using double Fourier series and perhaps partly because of the lack of fast and reliable algorithms (see also p. 495 in Boyd [4]). With the rapid increase in computing power, researchers are bound to use finer and finer resolutions in their computations to tackle more complex problems. Thus, fast algorithms using double Fourier series may become more and more advantageous with finer and finer resolutions when compared to the algorithms using spherical harmonics.

The purpose of this paper is to develop fast algorithms for elliptic equations on the sphere by using double Fourier series. We will also consider the problems in spherical layers by using the double Fourier series or spherical harmonics for the horizontal variables and Legendre or Chebyshev polynomials for the vertical variable. It is hoped that this paper will help to revive researchers' interests in using double Fourier series for elliptic- or parabolic-type equations in spherical geometries.

The rest of the paper is organized as follows. In the next section, we present two new algorithms by using double Fourier series for the Helmholtz equation and we report on the numerical experiments which compare the new algorithms with the existing algorithms based on the spherical harmonics and finite difference. In section 3, we present various extensions of our algorithms, including, in particular, applications to the elliptic equations in spherical layers and to the vector elliptic equations. Finally, we offer some concluding remarks followed by an appendix detailing the matrices introduced in section 3.

2. Elliptic problems on the sphere. Since the surface of the sphere has no boundary, high-order equations such as biharmonic or Stokes equations can be readily split up into a couple of second-order equations. Therefore, we shall only consider the Helmholtz equation on the surface of a unit sphere

$$\alpha U - \Delta U = F \ \text{ in } S := \{(x, y, z) : \ x^2 + y^2 + z^2 = 1\}. \tag{2.1}$$

We refer to [5] for a classical analysis on the well-posedness of this equation.

Applying the spherical transformation

$$x = \sin\theta\cos\phi, \ y = \sin\theta\sin\phi, \ z = \cos\theta \tag{2.2}$$

to (2.1) and setting $u(\theta, \phi) = U(x, y, z)$, $f(\theta, \phi) = F(x, y, z)$, and $D := (0, \pi) \times [0, 2\pi)$, we obtain

$$\mathcal{L}u := \alpha u - \frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta u) - \frac{1}{\sin^2\theta}\partial_{\phi\phi}u = f, \quad (\theta, \phi) \in D. \tag{2.3}$$

If $\alpha \neq 0$, the above equation has a unique solution, while for $\alpha = 0$, the compatibility condition

$$\int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta f(\theta, \phi) d\theta = 0 \tag{2.4}$$

should be satisfied, and the solution u is only determined up to an additive constant.

Equation (2.3) can be trivially solved by using the spherical harmonics which are in fact the eigenfunctions of the elliptic operator $\mathcal{L}$ (cf. [5]). The main disadvantage of using spherical harmonics is that the transformations between physical and spectral spaces are expensive. An alternative is to expand functions on the sphere by using double Fourier series (cf. [12] and [2]). In order to develop a fast algorithm based on a double Fourier expansion, we need to recall some important properties satisfied by u_m and f_m.

Since u is periodic in ϕ, we may write

$$u(\theta,\phi)=\sum_{|m|=0}^{\infty} u_m(\theta)e^{im\phi},\ \text{with}\ \ u_{-m}(\theta)=\bar{u}_m(\theta)\ \ \text{for all}\ \ m, \tag{2.5}$$

where $\bar{u}_m$ is the complex conjugate of u_m, and likewise for f. Substituting the expansion (2.5) (likewise for f) in (2.3), we find

$$\alpha u_m-\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\,\frac{d}{d\theta}u_m\right)+\frac{m^2}{\sin^2\theta}u_m=f_m,\quad \theta\in(0,\pi). \tag{2.6}$$

On the other hand, we may expand $u(\theta,\phi)$ in spherical harmonics functions,

$$u(\theta,\phi)=\sum_{n=0}^{\infty}\sum_{|m|=0}^{n} a_{nm}P_n^m(\cos\theta)e^{im\phi}, \tag{2.7}$$

where $P_n^m(x)$ is the associated Legendre function. Following Orszag (cf. [12]), $P_n^m(\cos\theta)$ can be expressed as $\sin^{|m|}\theta P_{nm}(\cos\theta)$ (where $P_{nm}(x)$ is a polynomial of degree $n-|m|$ in x), and we find from (2.5) and (2.7) that

$$u_m(\theta)=\sin^{|m|}\theta\sum_{n=|m|}^{\infty} a_{nm}P_{nm}(\cos\theta). \tag{2.8}$$

Hence, $u_m(\theta)$ has the same parity as m and is a periodic function with period 2π; i.e., u_m (and f_m) belongs to the space

$$Y^{(m)}:=\left\{v(\theta)=\sum_{n=0}^{\infty} v_n\phi_n^{(m)}(\theta):\ v\in L^2(0,\pi)\right\}, \tag{2.9}$$

where

$$\phi_n^{(m)}(\theta):=\begin{cases}\sin n\theta & \text{for } m \text{ odd},\\ \cos n\theta & \text{for } m \text{ even}.\end{cases} \tag{2.10}$$

Consequently, u (and f) belongs to the space

$$Y:=\left\{v=\sum_{|m|=0}^{\infty} v_m(\theta)e^{im\phi}\in L^2(D):\ v_m\in Y^{(m)},\ v_{-m}(\theta)=\bar{v}_m(\theta)\right\}. \tag{2.11}$$

Note that $Y^{(m)}$ (and $X^{(m)}$, $Z^{(m)}$ defined later) are complex spaces, while Y (and X, Z defined later) are real spaces. Furthermore, $u_m(\theta)$ has (at least) an $|m|$th order zero at $\theta=0,\ \pi$; i.e., $u_m(\theta)$ satisfies the following pole condition:

$$\frac{d^k u_m}{d\theta^k}|_{\theta=0,\,\pi}=0 \text{ for } m\neq 0,\ k=0,1,2,\ldots,|m|-1. \tag{2.12}$$

As pointed out by Orszag [12] (see also [2]), it is impractical and not necessary for the approximate solution to satisfy all the pole conditions in (2.12). We present below a fast solver for (2.6) (hence for (2.3)) based on a weighted Galerkin method in which the approximate solution, while being spectrally accurate, only satisfies the essential or necessary pole condition(s) to be specified below.

2.1. An interpolation operator. Given a pair of even integers (N, M), we define the approximations of $Y^{(m)}$ and Y as

$$Y_N^{(m)} := \left\{ v(\theta) = \sum_{\mathrm{n=mod}(m,2)}^{N-\mathrm{mod}(m,2)} v_n \phi_n^{(m)}(\theta) \right\} \tag{2.13}$$

and

$$Y_{NM} := \left\{ v = \sum_{|m|=0}^{M} v_m(\theta) e^{im\phi} : \ v_m \in Y_N^{(m)}, \ \begin{array}{c} v_{-m} = \bar{v}_m \text{ for all } m \\ v_0,\ v_{-M},\ v_M \text{ are real} \end{array} \right\}. \tag{2.14}$$

Observing that, for $u \in Y \cap C(D)$, we have

$$u(0,\phi) = u(0,\pi+\phi),\ u(\pi,\phi) = u(\pi,\pi+\phi)\ \forall \phi \in (0,2\pi), \tag{2.15}$$

we define a grid associated to Y_{NM} by

$$\tilde{\Sigma}_{NM} := \left\{ (\theta_i, \phi_j) : \ \begin{array}{c} j = 0,1,\ldots,2M-1 \text{ for } i = 1,2,\ldots,N-1 \\ j = 0,1,\ldots,M-1 \text{ for } i = 0 \text{ and } N \end{array} \right\}, \tag{2.16}$$

where $\theta_i = \frac{i\pi}{N}$ and $\phi_j = \frac{j\pi}{M}$.

One observes that there are $2NM$ points in $\tilde{\Sigma}_{NM}$ and that the degree of freedom of Y_{NM} is also $2NM$. Next, for any continuous, 2π-periodic function $v(\phi)$, we define an interpolation operator based on $\{\phi_j\}_{j=0,1,\ldots,2M-1}$ by

$$I_M^\phi v(\phi) = \sum_{|m|=0}^{M} v_m e^{im\phi} \text{ and } I_M^\phi v(\phi_j) = v(\phi_j),\ j = 0,1,\ldots,2M-1. \tag{2.17}$$

One derives from the discrete Fourier transform that

$$v_m = \frac{1}{c_m M} \sum_{j=0}^{2M-1} v(\phi_j) e^{-im\phi_j},\ v_{-m} = \bar{v}_m,\ m = 0,1,\ldots,M, \tag{2.18}$$

where $c_m = 1$ for $|m| \neq M$, $c_M = c_{-M} = 2$.

Now, for $u \in Y \cap C(D)$ and for $k = 0, 1 \ldots, N$, we have

$$I_M^\phi u(\theta_k,\phi) = \sum_{|m|=0}^{M} u_m(\theta_k) e^{im\phi} \text{ with } u_m(\theta_k) = \frac{1}{c_m M} \sum_{j=0}^{2M-1} u(\theta_k,\phi_j) e^{-im\phi_j}. \tag{2.19}$$

In particular, we have

$$\begin{aligned} u(0,\phi_j) &= I_M^\phi u(0,\phi_j) = \sum_{|m|=0}^{M} u_m(0) e^{im\phi_j}, \\ u(0,\pi+\phi_j) &= I_M^\phi u(0,\pi+\phi_j) = \sum_{|m|=0}^{M} (-1)^m u_m(0) e^{im\phi_j}. \end{aligned} \tag{2.20}$$

Since $u(0,\phi) = u(0,\pi+\phi)$ for $u \in Y$, we derive from the above relations that

$$\sum_{\substack{|m|=0 \\ m \text{ odd}}}^{M} u_m(0) e^{im\phi_j} = 0 \ \text{ for } j = 0, 1, \ldots, 2M-1. \tag{2.21}$$

Consequently, $u_m(0) = 0$ for m odd. Similarly, $u_m(\pi) = 0$ for m odd.
Let $\bar{c}_j = 1$ for $j = 1, \ldots, N-1$ and $\bar{c}_0 = \bar{c}_{_N} = 2$; we set

$$u_{nm} = \frac{2}{N} \sum_{j=1}^{N-1} u_m(\theta_j) \sin(n\theta_j), \ n = 1, \ldots, N-1 \quad \text{for } m \text{ odd}, \tag{2.22}$$

and

$$u_{nm} = \frac{2}{\bar{c}_n N} \sum_{j=0}^{N} \frac{1}{\bar{c}_j} u_m(\theta_j) \cos(n\theta_j), \ n = 0, \ldots, N \quad \text{for } m \text{ even}. \tag{2.23}$$

Then one finds from the discrete sin and cos transforms that

$$u_m(\theta_j) = \sum_{j=\mathrm{mod}(m,2)}^{N-\mathrm{mod}(m,2)} u_{nm} \phi_n^{(m)}(\theta_j), \ \ j = \mathrm{mod}(m,2), \ldots, N - \mathrm{mod}(m,2). \tag{2.24}$$

Now we define an operator $I_{NM} : Y \cap C(D) \longrightarrow Y_{NM}$ by

$$I_{NM} u(\theta,\phi) := \sum_{|m|=0}^{M} \sum_{n=\mathrm{mod}(m,2)}^{N-\mathrm{mod}(m,2)} u_{nm} \phi_n^{(m)}(\theta) e^{im\phi}. \tag{2.25}$$

By construction, we have

$$I_{NM} u(\theta,\phi) = u(\theta,\phi) \ \ \forall (\theta,\phi) \in \tilde{\Sigma}_{NM}, \ u \in Y \cap C(D);$$

i.e., $I_{NM} : Y \cap C(D) \longrightarrow Y_{NM}$ is an interpolation operator associated with the grid $\tilde{\Sigma}_{NM}$.

Note that functions in Y do not satisfy any pole conditions in (2.12), hence, they are not even single valued at the poles in the Cartesian coordinates. We define below a sequence of spaces whose functions satisfy the first pole condition in (2.12):

$$\begin{aligned}
Z^{(m)} &= \left\{ v \in Y^{(m)} \cap H^1(0,\pi) : \ v(0) = v(\pi) = 0 \text{ when } m \neq 0 \right\}, \\
Z &= \left\{ v = \sum_{|m|=0}^{\infty} v_m(\theta) e^{im\phi} \in H^1(D) : \ v_m \in Z^{(m)}, \ v_{-m} = \bar{v}_m \right\}, \\
Z_N^{(m)} &= \left\{ v \in Y_N^{(m)} : \ v(0) = v(\pi) = 0 \text{ when } m \neq 0 \right\}, \\
Z_{NM} &= \left\{ v = \sum_{|m|=0}^{M} v_m(\theta) e^{im\phi} : \ v_m \in Z_N^{(m)}, \ \begin{array}{c} v_{-m} = \bar{v}_m \text{ for all } m \\ v_0, \ v_{-_M}, \ v_{_M} \text{ are real} \end{array} \right\}.
\end{aligned} \tag{2.26}$$

We observe that, for $u \in Z$, we have

$$u(0,\phi) = u(0,0), \ u(\pi,\phi) = u(\pi,0) \ \ \forall \phi \in (0, 2\pi). \tag{2.27}$$

Therefore, functions in Z are single valued at the poles in the Cartesian coordinates. Note that one verifies easily that the spherical gradient $\nabla_s u = (u_\theta, \frac{u_\phi}{\sin\theta})^t \in L^2(D)$ for $u \in Z$. Thus, functions in Z are also differentiable at the poles in the Cartesian coordinates.

2.2. Formulation of the algorithm. Given a positive weight function $\omega(\theta)$ satisfying $\omega(0) = \omega(\pi) = 0$, and an appropriate subspace $X_N^{(m)} \subset Y_N^{(m)}$, the weighted spectral-Galerkin (Petrov–Galerkin) approximation to (2.6) is to find $u_N^{(m)} \in X_N^{(m)}$ (for $m = 0, 1, \ldots, M$) such that

$$\begin{aligned}\alpha \int_o^\pi u_N^{(m)} v \omega(\theta) d\theta - \int_0^\pi \frac{d}{d\theta}\left(\sin\theta \frac{d}{d\theta} u_N^{(m)}\right) v \frac{\omega(\theta)}{\sin\theta} d\theta \\ + m^2 \int_0^\pi u_N^{(m)} v \frac{\omega(\theta)}{\sin^2\theta} d\theta = \int_0^\pi f_N^{(m)} v \omega(\theta) d\theta \quad \forall v \in X_N^{(m)},\end{aligned} \tag{2.28}$$

where $f_N^{(m)}$ is an approximation of f_m defined by

$$I_{NM} f = \sum_{|m|=0}^{M} f_N^{(m)}(\theta) e^{im\phi}, \quad \text{with} \quad f_N^{(m)} \in Y_N^{(m)}. \tag{2.29}$$

Then the approximation of u is defined as

$$u_{NM} = \sum_{|m|=0}^{M} u_N^{(m)}(\theta) e^{im\phi}, \quad u_N^{(-m)}(\theta) = \bar{u}_N^{(m)}(\theta). \tag{2.30}$$

Given a set of basis functions $\{\psi_n^{(m)}\}$ for $X_N^{(m)}$ (the dimension of $X_N^{(m)}$ will be specified later), we may write $u_N^{(m)} = \sum_n x_n^{(m)} \psi_n^{(m)}$. Setting

$$\begin{aligned} a_{kj}^{(m)} &= \int_0^\pi \psi_j^{(m)} \psi_k^{(m)} \omega(\theta) d\theta, \ b_{kj}^{(m)} = -\int_0^\pi \frac{d}{d\theta}(\sin\theta \frac{d}{d\theta}\psi_j^{(m)}) \psi_k^{(m)} \frac{\omega(\theta)}{\sin\theta} d\theta, \\ c_{kj}^{(m)} &= \int_0^\pi \psi_j^{(m)} \psi_k^{(m)} \frac{\omega(\theta)}{\sin^2\theta} d\theta, \ g_j^{(m)} = \int_0^\pi f_N^{(m)} \psi_j^{(m)} \omega(\theta) d\theta, \end{aligned} \tag{2.31}$$

and

$$\begin{aligned} &A^{(m)} = (a_{kj}^{(m)}), \ B^{(m)} = (b_{kj}^{(m)}), \ C^{(m)} = (c_{kj}^{(m)}), \\ &\boldsymbol{x}^{(m)} = (x_j^{(m)}), \ \boldsymbol{g}^{(m)} = (g_j^{(m)}), \end{aligned} \tag{2.32}$$

then (2.28) is equivalent to the following linear system:

$$(\alpha A^{(m)} + B^{(m)} + m^2 C^{(m)}) \boldsymbol{x}^{(m)} = \boldsymbol{g}^{(m)}. \tag{2.33}$$

The efficiency of the method depends on the structure of these matrices, which in turn depend on the choice of $\omega(\theta)$ and $X_N^{(m)}$.

There are numerous ways to choose the weight $\omega(\theta)$ and $X_N^{(m)}$ as long as the linear system (2.28) is well-posed. Note that although the exact solution u_m of (2.6) will *automatically* satisfy all the pole conditions in (2.12), the approximation solution $u_N^{(m)}$ will not unless these pole conditions are built into the space $X_N^{(m)}$. We recall that it is impractical and not necessary for the approximate solution to satisfy all the pole conditions in (2.12). Hence, we shall search a pair of $w(\theta)$ and $X_N^{(m)}$ such that (i) the linear system (2.33) is nonsymmetric but positive definite; and (ii) the matrices $A^{(m)}$, $B^{(m)}$, and $C^{(m)}$ are banded matrices with small bandwidth so that (2.33) can be solved efficiently.

To fix the idea, we assume $\alpha \neq 0$. The case $\alpha = 0$ can be treated similarly with slight modifications (see Remark 2.1).

The form of (2.6) suggests that we choose $\omega(\theta) = \sin^2\theta$. In this case, the system (2.28) is well defined for $X_N^{(m)} = Y_N^{(m)}$; i.e., no pole condition is *essential* — all the pole conditions in (2.12) are *nonessential* (i.e., natural). We consider below two different choices for $X_N^{(m)}$.

Case 1. $\omega(\theta) = \sin^2\theta,\ X_N^{(m)} = Y_N^{(m)}$.

In this case, $\psi_n^{(m)}(\theta)$ is defined in (2.10). By elementary computations, we find that the system (2.33) has only three nonzero diagonals. More precisely, the nonzero entries of $A^{(m)}$, $B^{(m)}$, and $C^{(m)}$ are given below (the superscript $^{(m)}$ is omitted for simplicity).

For m odd and $k, j = 1, 2, \ldots, N-1$:

$$a_{jk} = a_{kj} = \frac{\pi}{2}\begin{cases} -\frac{1}{4}, & k = j+2, \\ \frac{1}{2}, & k = j \neq 1, \end{cases} \qquad a_{11} = \frac{3\pi}{8}; \tag{2.34}$$

$$b_{kj} = \frac{\pi}{2}\begin{cases} -\frac{1}{4}j(j+1), & k = j+2, \\ \frac{1}{2}j^2, & k = j, \\ -\frac{1}{4}j(j-1), & k = j-2; \end{cases} \tag{2.35}$$

$$c_{kj} = \frac{\pi}{2}\delta_{kj}. \tag{2.36}$$

For m even (including $m = 0$) and $k, j = 0, 1, \ldots, N$: For $k, j = 1, 2, \ldots, N$, the a_{kj}, b_{kj}, and c_{kj} are exactly the same as above except that $a_{11} = \frac{\pi}{8}$. The additional nonzero entries (for k or $j = 0$) are

$$a_{00} = \frac{\pi}{2},\ \ a_{02} = a_{20} = -\frac{\pi}{4},\ \ b_{02} = -\frac{\pi}{2},\ \ \text{and}\ \ c_{00} = \pi.$$

The choice $X_N^{(m)} = Y_N^{(m)}$ leads to the simplest linear system. However, the solution $u_{_{NM}}$ (in Y_{NM}) is neither necessarily single valued nor differentiable at the poles in the Cartesian coordinates, although the approximate solution $u_{_{NM}}$ still converges to u exponentially provided that f is smooth in the spherical coordinates (see section 3). In order to apply this algorithm to more complicated situations, such as nonlinear or time dependent problems, we may project, in an appropriate way, the solution $u_{_{NM}}$ to the space Z_{NM}. To this end, we define a simple projector $\pi_{_{NM}} : Y_{NM} \longrightarrow Z_{NM}$ as follows.

Let us first define a projector $\pi_{_N}^{(m)} : Y_N^{(m)} \longrightarrow Z_N^{(m)}$.

- For m odd or $m = 0$, we set $\pi_{_N}^{(m)}$ to be the identity operator.
- For m even and $m \neq 0$, any function u in $Y_N^{(m)}$ can be written as $u = \sum_{k=0}^{N} a_k \cos k\theta$. We set $\pi_{_N}^{(m)} u := \sum_{k=0}^{N} \tilde{a}_k \cos k\theta$, where $\tilde{a}_k = a_k$ for $k = 0, 1, \ldots, N-2$, while $\tilde{a}_{_{N-1}}$ and $\tilde{a}_{_N}$ are uniquely determined by the conditions $(\pi_{_N}^{(m)} u)(0) = (\pi_{_N}^{(m)} u)(\pi) = 0$.

Then, for $u = \sum_{|m|=0}^{M} u_m(\theta) e^{im\phi} \in Y_{NM}$, we set

$$\pi_{_{NM}} u := \sum_{|m|=0}^{M} \pi_{_N}^{(m)} u_m e^{im\phi}. \tag{2.37}$$

Thus, $\pi_{NM}^{(m)} u_{NM} \in Z_{NM}$ and is single valued and differentiable at the poles, and we may use $\pi_N^{(m)} u_{NM}$ as an approximation to u whenever necessary.

Case 2. $\omega(\theta) = \sin^2\theta$, $X_N^{(m)} = Z_N^{(m)}$.

In this case, the first pole condition in (2.12) is imposed so that the solutions are single valued and differentiable at the poles. A convenient set of basis functions for $Z_N^{(m)}$ is

$$\psi_n^{(m)}(\theta) = \begin{cases} \cos(n-1)\theta - \cos(n+1)\theta, & m \text{ even and } m \neq 0, \\ \cos n\theta, & m = 0, \\ \sin n\theta, & m \text{ odd.} \end{cases} \tag{2.38}$$

It can be verified that in this case the system (2.33) has at most five nonzero diagonals. For m odd or $m = 0$, the matrices are exactly the same as in the case of $X_N^{(m)} = Y_N^{(m)}$, while for m even and $m \neq 0$, their nonzero entries are (the superscript $^{(m)}$ is omitted below for simplicity)

$$a_{jk} = a_{kj} = \frac{\pi}{2} \begin{cases} \frac{1}{4}, & k = j+4, \\ -1, & k = j+2 \neq 3, \\ \frac{3}{2}, & k = j \neq 1, 2, \end{cases} \tag{2.39}$$

with $a_{11} = \frac{5\pi}{4}$, $a_{22} = \frac{5\pi}{8}$, $a_{13} = a_{31} = -\frac{5\pi}{8}$;

$$b_{kj} = \frac{\pi}{2} \begin{cases} \frac{1}{4}(j+2)(j+3), & k = j+4, \\ -\frac{1}{2}(2j^2+7j+3), & k = j+2, \\ \frac{1}{2}(3j^2+6j+5), & k = j \neq 1, \\ -\frac{1}{2}(2j^2+j+1), & k = j-2 \neq 1, \\ \frac{1}{4}j(j-1), & k = j-4, \end{cases} \tag{2.40}$$

with $b_{11} = \frac{3\pi}{2}$, $b_{13} = -\frac{\pi}{3}$;

$$c_{jk} = c_{kj} = \frac{\pi}{2} \begin{cases} -1, & k = j+2, \\ 2, & k = j \neq 1, \end{cases} \tag{2.41}$$

with $c_{11} = \frac{3\pi}{2}$.

Note that in both cases the linear system (2.28) (or (2.33)) is nonsymmetric but positive definite for $\alpha \geq 0$. Indeed, for $u \in Y_N^{(m)}$,

$$\begin{aligned} \int_0^\pi \sin\theta u_\theta (\sin\theta u)_\theta \, d\theta &= \int_0^\pi \sin^2\theta u_\theta^2 d\theta + \frac{1}{2}\int_0^\pi \sin 2\theta u_\theta u d\theta \\ &= \int_0^\pi \sin^2\theta u_\theta^2 d\theta + \frac{1}{4}\int_0^\pi \sin 2\theta (u^2)_\theta d\theta \\ &= \int_0^\pi \sin^2\theta u_\theta^2 d\theta - \frac{1}{2}\int_0^\pi \cos 2\theta u^2 d\theta. \end{aligned}$$

Therefore, for $u \in Y_N^{(m)}$, we have

$$\begin{aligned} a_m(u,u) &= \alpha \int_0^\pi \sin^2\theta u^2 \, d\theta + \int_0^\pi \sin\theta u_\theta (\sin\theta u)_\theta \, d\theta + m^2 \int_0^\pi u^2 \, d\theta \\ &\geq \int_0^\pi \sin^2\theta u_\theta^2 d\theta + \int_0^\pi \left(m^2 - \frac{1}{2} + \alpha \sin^2\theta \right) u^2 \, d\theta. \end{aligned} \tag{2.42}$$

TABLE 2.1
Condition number of (2.33) *with* $m = 0$.

$N = M$	16	32	64	128	256
$\alpha = 0$	1.36E+3	9.03E+3	5.74E+4	3.52E+5	2.11E+6
$\alpha = 1,000$	1.51E+2	5.83E+2	3.85E+3	2.89E+4	2.08E+5
$\alpha = 1,000,000$	1.66E+2	6.07E+2	2.32E+3	9.01E+3	3.51E+4

Hence, for $m \neq 0$, the bilinear form is nonsymmetric positive definite. For the case $m = 0$, the same conclusion can be drawn since the matrix B^0 given in (2.35) is irreducibly (column) diagonally dominant with positive diagonal elements. In addition, the above inequality indicates that the conditioning of the linear system (2.28) (or (2.33)) improves as α increases. In Table 2.1, we list the condition numbers, defined as the ratio of the largest and smallest singular values, of the matrices associated to (2.33) with several values of α. Only the case $m = 0$ is listed since it is clear from (2.42) and it is found numerically that the largest condition number occurs at $m = 0$. Since $Y_N^{(0)} = Z_N^{(0)}$, the matrix of (2.33) with $m = 0$ is the same for both cases.

Note that the condition numbers are relatively large, so it is advised to use double precision, especially for large N. On the other hand, for large values of α, which occur often in solving time dependent fluid problems with large Reynolds numbers, the condition number improves significantly, especially for large N. Thus, our algorithms are well-suited for problems with large values of α; see Table 2.3 for more numerical results on this point.

Note also that (2.33) can be split up into two tridiagonal or pentadiagonal subsystems which can be solved very efficiently in $O(N)$ operations. Thus, the overall computational complexity of this Helmholtz solver is $O(NM \log M)$.

REMARK 2.1. *For* $\alpha = 0$, *we can fix the additive constant by removing the constant function, i.e.,* $\phi_0^{(0)}$, *from the expansion. In other words, we look for* $u_N^{(0)} \in$ *span*$\{\cos n\theta : \ n = 1, 2, \ldots, N\}$. *Then, the system* (2.33) *at* $m = 0$ *becomes a well defined* $N \times N$ *system. Hence,* $u_N^{(0)}$ *is uniquely determined without requiring any compatibility condition. However, the corresponding approximate solution* u_{NM} *will converge to* u *only if* f *satisfies the compatibility condition* (2.4).

REMARK 2.2. *A seemingly more natural weight function is* $\omega(\theta) = \sin\theta$ *which is in fact the Jacobian of the spherical transformation and which leads to symmetric positive definite systems* (2.33). *In this case,* (2.28) *makes sense only if*

$$u_m(0) = u_m(\pi) = 0, \quad m \neq 0. \tag{2.43}$$

Hence, we may take $X_N^{(m)} = Z_N^{(m)}$ *and use* $\psi_n^{(m)}$ *defined in* (2.38). *Unfortunately, the resulting linear systems are not sparse.*

2.3. Numerical results. We have solved the Helmholtz equation (2.3) on the sphere with the function f being such that the exact solution is

$$u(\theta, \phi) = \cos(8(\sin\theta\cos\phi + \sin\theta\sin\phi + \cos\theta)) \tag{2.44}$$

by using four different algorithms. Namely, (i) sph_har: using spherical harmonics; (ii) Fourier_I: using double Fourier series in (2.28) with $\omega(\theta) = \sin^2(\theta)$ and $X_N^{(m)} = Y_N^{(m)}$; (iii) Fourier_II: using double Fourier series in (2.28) with $\omega(\theta) = \sin^2(\theta)$ and $X_N^{(m)} = Z_N^{(m)}$; (iv) FISHPACK: using the finite difference code **hwsssp.f** in FISHPACK [19].

TABLE 2.2
Comparison of the l^2 errors for the solution u with $\alpha = 0$.

$N = M$	16	20	24	32	40
sph_har	4.93E-2	1.80E-3	2.63E-5	5.06E-10	6.24E-15
Fourier_I	1.29E-2	1.47E-4	1.20E-6	5.55E-11	4.54E-15
Fourier_II	1.28E-2	1.45E-4	1.99E-6	5.56E-11	4.59E-15
FISHPACK	3.41E-1	2.13E-1	1.41E-1	7.24E-2	4.35E-2

TABLE 2.3
l^2 errors for the solution u with different α.

$N = M$	16	32	64	128
Fourier_I ($\alpha = 0$)	1.29E-2	5.55E-11	8.54E-15	4.76E-15
Fourier_II ($\alpha = 0$)	1.28E-2	5.56E-11	8.52E-15	5.04E-15
Fourier_I ($\alpha = 1,000$)	3.17E-3	2.03E-11	7.72E-15	1.52E-14
Fourier_II ($\alpha = 1,000$)	1.68E-3	2.03E-11	8.74E-15	1.51E-14
Fourier_I ($\alpha = 1,000,000$)	1.37E-5	5.88E-12	2.14E-12	7.66E-13
Fourier_II ($\alpha = 1,000,000$)	2.11E-6	5.88E-12	2.14E-12	7.67E-13
Fourier_I ($\alpha = -1,000$)	1.03E-2	2.91E-10	2.18E-14	1.34E-14
Fourier_II ($\alpha = -1,000$)	3.43E-3	2.91E-10	1.77E-14	1.69E-14

All computations were carried out in double precision on a SUN Ultra-1 workstation Model-140 with the standard optimization option "-O."

In Table 2.2, we list the discrete l^2 errors, based on the collocation points in $\tilde{\Sigma}_{NN}$, of the four algorithms applied to (2.3) with $\alpha = 0$. As expected, all three spectral methods converge exponentially and the finite difference code converges quadratically. Note that the two algorithms using double Fourier series are slightly more accurate than that using spherical harmonics.

Next, we demonstrate the robustness of our algorithms with respect to α. In Table 2.3, we list the l^2 errors of our algorithms with several different values of α. The exact solution is again given in (2.44).

The results in Table 2.3 clearly indicate that our algorithms are robust with respect to large values of α. They even produce spectrally accurate results for negative values of α which are not close to the critical values for which (2.33) is singular.

In Table 2.4, we list the CPU times of the four algorithms for solving the Helmholtz equation (2.3) on the sphere. FFTPACK, developed by Swarztrauber and Sweet, is used to compute the discrete Fourier transforms in the algorithms Fourier_I and Fourier_II, while SPHEREPACK, developed by Swarztrauber and Adams, is used to compute the discrete spherical harmonic transforms. For the spherical harmonic algorithm, ($\approx N^3$) values of the associate Legendre polynomials need to be precomputed and stored. The CPU times for the precomputation are listed in parentheses. The two algorithms using double Fourier series are significantly more efficient than that using spherical harmonics; e.g., a factor of about 30 (taking into account the precomputation) is observed with $N = M = 256$. They are also considerably more efficient than the finite difference code while providing exponential convergence as opposed to the second-order convergence of the finite difference method.

For applications in nonlinear and/or time dependent problems, it is important that the spherical gradient of the solution $\nabla_s u = (u_\theta, \frac{u_\phi}{\sin\theta})^t$ is approximated with a desirable accuracy. Although $\nabla_s u$ is not necessarily defined for $u \in Y_{NM}$, we can use, in the algorithm Fourier_I, $\nabla_s(\pi_{_{NM}} u_{_{NM}})$ (see (2.37)) as the approximation to $\nabla_s u$. In Table 2.5, we list the l^2 errors for $\nabla_s u$ by using Fourier_I and Fourier_II applied

TABLE 2.4
CPU time (in second) for the Helmholtz solvers on the sphere.

$N = M$	32	48	64	96	128	192	256
sph_har	6.2E-3	1.7E-2	3.7E-2	.12	.28	1.19	3.06
(pre_comp)	(2.5E-2)	(7.0E-2)	(.15)	(.45)	(1.18)	(3.85)	(9.02)
Fourier_I	6.6E-3	1.4E-2	2.3E-2	5.3E-2	9.0E-2	.24	..42
Fourier_II	7.1E-3	1.5E-2	2.4E-2	6.0E-2	.11	.27	.46
FISHPACK	6.8E-3	3.1E-2	6.9E-2	.13	.27	.65	1.22

TABLE 2.5
Comparison of the l^2 errors for $\nabla_s u = (u_\theta, \frac{u_\phi}{\sin\theta})^t$.

$N = M$	20	24	32	36	40
Fourier_I	2.81E-2	5.31E-4	1.13E-8	1.44E-11	3.54E-14
Fourier_II	2.82E-2	5.32E-4	1.12E-8	1.40E-11	3.91E-14

to (2.3) with $\alpha = 0$. We find that both algorithms offer essentially the same accuracy for $\nabla_s u$. Thus, by applying the projector $\pi_{_{NM}}$ whenever needed, we believe that the algorithm Fourier_I is as widely applicable as the algorithm Fourier_II. We emphasize that the main purpose of this comparison is not to determine which algorithm is better, but rather to support the view that *nonessential* pole conditions in (2.12) can be safely ignored.

3. Extensions. As indicated previously (cf. [14, 17]), elliptic problems with variable coefficients can be efficiently solved by using a conjugate gradient-type iterative method with a spectrally equivalent operator associated to a constant coefficient problem as a preconditioner. Below, we provide details on how to treat vector equations and problems in spherical layers.

3.1. Vector equations. In many physical applications, one often needs to solve vector equations in spherical geometries. Consider, for instance, the vector Poisson equation on the unit sphere in spherical coordinates

$$\boldsymbol{\nabla^2} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} := \begin{pmatrix} \nabla^2 - \frac{1}{\sin^2\theta}, & -\frac{2\cos\theta}{\sin^2\theta}\frac{\partial}{\partial\phi} \\ \frac{2\cos\theta}{\sin^2\theta}\frac{\partial}{\partial\phi}, & \nabla^2 - \frac{1}{\sin^2\theta} \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}, \tag{3.1}$$

where (u_1, u_2) and (f_1, f_2) are vectors in spherical coordinates and

$$\nabla^2 := \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial}{\partial\theta}) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2} \tag{3.2}$$

is the scalar Laplace operator in spherical coordinates. The difficulty here is that the uncoupled vector Poisson equation in Cartesian coordinates now becomes coupled in spherical coordinates. One may use, of course, the so-called vectorial spherical harmonics (cf. [11, 18]) which are the eigenfunctions of the operator $\boldsymbol{\nabla^2}$. However, in addition to the inevitable expensive transforms, the implementation task with vectorial spherical harmonics can be formidable. In order to be able to use double Fourier series, one can first decouple the system (3.1) by applying the transform

$$\begin{pmatrix} \tilde{u}_1 \\ \tilde{u}_2 \end{pmatrix} = \begin{pmatrix} \sin\phi, & \cos\phi \\ \cos\phi, & -\sin\phi \end{pmatrix} \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \tilde{f}_1 \\ \tilde{f}_2 \end{pmatrix} = \begin{pmatrix} \sin\phi, & \cos\phi \\ \cos\phi, & -\sin\phi \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \end{pmatrix}. \tag{3.3}$$

Then the system (3.1) is reduced to

$$\nabla^2\tilde{u}_1 = \tilde{f}_1 \ \text{ and } \ \nabla^2\tilde{u}_2 = \tilde{f}_2, \tag{3.4}$$

which can be solved by using double Fourier series as in section 2 or by using scalar spherical harmonics.

The three-dimensional (3D) case can be handled in a similar way. We refer to Appendix B in [13] for more details on how to decouple the 3D vector equations in spherical coordinates.

3.2. Elliptic equations in a spherical layer. We consider again the model Helmholtz equation

$$\begin{aligned} &\alpha U - \Delta U = F \ \text{ in } \Omega = \{(x,y,z) :\ R_1 < x^2+y^2+z^2 < R_2\}, \\ &U|_{\partial\Omega} = 0. \end{aligned} \tag{3.5}$$

Applying the spherical transformation

$$x = r\sin\theta\cos\phi,\ y = r\sin\theta\sin\phi,\ z = r\cos\theta \tag{3.6}$$

to (3.5), and setting $u(r,\theta,\phi) = U(x,y,z)$, $f(r,\theta,\phi) = F(x,y,z)$, we obtain

$$\begin{aligned} &\alpha u - \frac{1}{r^2}\partial_r(r^2\partial_r u) - \frac{1}{r^2\sin\theta}\partial_\theta(\sin\theta\partial_\theta u) - \frac{1}{r^2\sin^2\theta}\partial_{\phi\phi}u = f, \\ &(r,\theta,\phi) \in (R_1,R_2)\times(0,\pi)\times[0,2\pi), \\ &u = 0 \ \text{ at } \ r = R_1 \ (\text{if } R_1 \neq 0) \text{ and } r = R_2. \end{aligned} \tag{3.7}$$

Below, we present two approaches; one is based on spherical harmonics and the other on double Fourier series.

3.2.1. Using spherical harmonics. Let $Y_{lm}(\theta,\phi) = P_l^m(\cos\theta)e^{im\phi}$ be the spherical harmonics function of index (l,m); we can write

$$u(r,\theta,\phi) = \sum_{\substack{l=0 \\ |m|\le l}}^{\infty} u_{lm}(r)Y_{lm}(\theta,\phi)$$

(and likewise for f). Then, thanks to the property [5]

$$\mathcal{L}Y_{lm} := -\frac{1}{\sin\theta}\partial_\theta(\sin\theta\partial_\theta Y_{lm}) - \frac{1}{\sin^2\theta}\partial_{\phi\phi}Y_{lm} = l(l+1)Y_{lm}, \tag{3.8}$$

we find, after multiplying (3.7) by r^2 (this is a natural process since $r^2\sin\theta$ is the Jacobian of the spherical transform), that (3.7) reduces to

$$\begin{aligned} &\alpha u r_{lm}^2 - (r^2u'_{lm})' + l(l+1)u_{lm} = r^2 f_{lm}, \ \ 0 \le |m| \le l < \infty, \\ &u_{lm}(R_1) = 0 \text{ if } R_1 \neq 0,\ u_{lm}(R_2) = 0. \end{aligned} \tag{3.9}$$

Since the interval $[R_1,R_2]$ can be mapped to $[-1,1]$ by using the transform

$$r = \frac{R_2-R_1}{2}(t+\beta) \ \text{ with } \beta = \frac{R_2+R_1}{R_2-R_1} \ge 1, \tag{3.10}$$

we have only to consider the following prototypical one-dimensional problem:

$$\begin{aligned} &\alpha(t+\beta)^2u - ((t+\beta)^2u')' + \gamma u = (t+\beta)^2 f, \\ &u(-1) = 0 \text{ if } \beta > 1; \ \ u(1) = 0. \end{aligned} \tag{3.11}$$

Let P_K be the space of polynomials of degree less than or equal to K, and let

$$Z_K = Z_K(\beta) := \{v \in P_K :\ v(-1) = 0 \text{ if } \beta > 1,\ v(1) = 0\}. \tag{3.12}$$

Then the weighted spectral-Galerkin approximation to (3.11) is to find $u_N \in Z_K$ such that

$$\begin{aligned} \alpha((t+\beta)^2 u_N, v\omega) + ((t+\beta)^2 u_N', (v\omega)') + \gamma(u_N, v\omega) \\ = ((t+\beta)^2 J_N f, v\omega) \ \ \forall\, v \in Z_K, \end{aligned} \tag{3.13}$$

where $\omega \equiv 1$ in the Legendre case and $\omega(t) = (1-t^2)^{-\frac{1}{2}}$ in the Chebyshev case, $(u, v) = \int_{-1}^{1} uv dt$, and J_N is the operator of interpolation based on the Legendre or Chebyshev Gauss–Lobatto points.

Let $\phi_k(t),\ k = 0, 1, \ldots, K - I_\beta$ be a set of basis function for Z_K, where $I_\beta = 1$ if $\beta = 1$ while $I_\beta = 2$ if $\beta > 1$. We set

$$\begin{aligned} q_{kj} &= \int_{-1}^{1} (t+\beta)^2\, \phi_j\, \phi_k\, \omega\, dt,\ \ Q = (q_{kj}), \\ r_{kj} &= \int_{-1}^{1} (t+\beta)^2\, \phi_j'\, (\phi_k \omega)'\, dt,\ \ R = (r_{kj}), \\ s_{kj} &= \int_{-1}^{1} \phi_j\, \phi_i \omega\, dt,\ \ S = (s_{kj}), \\ f_j &= \int_{-1}^{1} J_N f\, \phi_j\, \omega\, dt,\ \ \boldsymbol{f} = (f_j), \\ u_N &= \sum_{i=0}^{N-2} x_j \phi_j(t),\ \ \boldsymbol{x} = (x_j). \end{aligned} \tag{3.14}$$

Then (3.13) becomes the matrix equation

$$(\alpha Q + R + \gamma S)\boldsymbol{x} = \boldsymbol{f}. \tag{3.15}$$

The efficiency of the method depends on the choice of the basis functions which, in turn, determine the matrices Q, R, and S.

To simplify the notation, we will only consider the case $R_1 > 0$ (i.e., $\beta > 1$). The case $R_1 = 0$ (i.e., $\beta = 1$) can be treated similarly.

We present below three different ways to implement (3.13).

- *Legendre–Galerkin.* In this case, we set $\omega = 1$ and $\phi_j(t) = L_j(t) - L_{j+2}(t)$. By using the identities

$$\phi_i'(t) = -(2i+3)L_{i+1}(t), \tag{3.16}$$

$$(i+1)L_{i+1}(t) = (2i+1)t\, L_i(t) - i\, L_{i-1}(t), \tag{3.17}$$

$$\phi_i(t) = \frac{2i+3}{(i+1)(i+2)}(1-t^2)L_{i+1}'(t), \tag{3.18}$$

one can readily derive that Q, R, and S are positive definite symmetric matrices with $q_{ij} = 0$ for $|i-j| > 4$, $r_{ij} = 0$ for $|i-j| > 2$, and $s_{ij} = 0$ for $j \neq i, i \pm 2$. Their nonzero entries are given in Appendix A. The drawback of this approach is that no fast Legendre transform is available. The Chebyshev–Legendre Galerkin method presented below is designed to overcome this shortcoming.

- *Chebyshev–Legendre Galerkin.* In this case, we use the *Legendre* formulation with $\omega = 1$ and $\phi_j(t) = L_j(t) - L_{j+2}(t)$, but we define the interpolation operator J_κ based on the *Chebyshev* Gauss–Lobatto points. Then (3.13) leads to a symmetric banded system with small bandwidth, while the transformation between function values at the Chebyshev Gauss–Lobatto points and the coefficients of its Legendre expansion can be done with quasi-optimal computational complexity (cf. [1, 16]).
- *Chebyshev–Galerkin.* We set $\omega = (1-t^2)^{-\frac{1}{2}}$ and $\phi_j(t) = (1-t^2)T_j(t)$. It can be easily shown that Q and S are positive definite symmetric matrices with $q_{ij} = 0$ for $|i-j| > 6$, $s_{ij} = 0$ for $|i-j| > 4$ and $|i-j|$ odd. Although R is nonsymmetric, it can be shown that R is banded with $r_{ij} = 0$ for $i-4 \le j \le i+4$. Indeed, it is easy to see that

$$r_{ij} = -\int_I \big((t+\beta)^2\phi_j'\big)'(1-t^2)T_i\omega dt = 0 \quad \text{for} \quad i > j+4.$$

On the other hand, thanks to identity $\omega'(t) = t\omega^3(t)$ and integration by parts,

$$\begin{aligned} r_{ij} &= \int_I (t+\beta)^2\phi_j'\big((1-t^2)T_i\omega\big)' dt \\ &= \int_I (t+\beta)^2\phi_j'\big(((1-t^2)T_i)' + tT_i\big)\omega dt \\ &= \int_I \phi_j' P_{i+3}\omega dt = -\int_I T_j(1-t^2)(P_{i+3}\omega)' dt \\ &= -\int_I T_j\big((1-t^2)P_{i+3}' + tP_{i+3}\big)\omega dt = \int_I T_j P_{i+4}\omega dt, \end{aligned}$$

where P_{i+3} (resp., P_{i+4}) is a polynomial of degree less than or equal to $i+3$ (resp., $i+4$). Hence, $r_{ij} = 0$ for $j > i+4$.

Although it is very tedious to determine their nonzero entries by hand, one can easily compute them by using appropriate Gaussian quadratures. The details are left to the interested readers. The main advantage of the Chebyshev method is the availability of the fast Chebyshev transform. Beside the tedious process involved in evaluating the nonzero entries of Q, R, and S, the nonsymmetry of R may introduce additional difficulties when iterative methods are needed for problems with variable coefficients.

A performance comparison of the three methods in a different context can be found in [16].

REMARK 3.1. *In case $R_1 = 0$ (i.e., $\beta = 1$), the appropriate basis functions are $\phi_j(t) = L_j(t) - L_{j+1}(t)$ in the Legendre case and $\phi_j(t) = (1-t)T_j(t)$ in the Chebyshev case.*

Higher-order equations can be solved in a similar fashion. For instance, by using the expansion in spherical harmonics, the biharmonic equation would be reduced to a set of one-dimensional fourth-order equations which can be solved efficiently by using a spectral-Galerkin method (see [17] for a similar case).

3.2.2. Using double Fourier series. Since u is periodic in ϕ, we may write

$$u = \sum_{|m|=0}^{\infty} \tilde{u}_m(r,\theta)e^{im\phi},$$

and likewise for f. Then (3.7) becomes

$$\begin{aligned} &\alpha\tilde{u}_m - \frac{1}{r^2}\partial_r(r^2\partial_r\tilde{u}_m) - \frac{1}{r^2\sin\theta}\partial_\theta(\sin\theta\partial_\theta\tilde{u}_m) + \frac{m^2}{r^2\sin^2\theta}\tilde{u}_m = \tilde{f}_m, \\ &\tilde{u}_m|_{\partial\tilde{D}} = 0, \quad \tilde{D} := (R_1, R_2)\times(0,\pi). \end{aligned} \tag{3.19}$$

Applying again the transform (3.10) and multiplying by $r^2\sin^2\theta$, (3.19) becomes

$$\begin{aligned} &\alpha\gamma^2(t+\beta)^2\sin^2\theta u_m - \sin^2\theta\partial_t((t+\beta)^2\partial_t u_m) - \sin\theta\partial_\theta(\sin\theta\partial_\theta u_m) \\ &\qquad + m^2 u_m = \gamma^2(t+\beta)^2\sin^2\theta f_m, \text{ in } D := (-1,1)\times(0,\pi), \\ &u_m|_{\partial D} = 0, \end{aligned} \tag{3.20}$$

where we have set $\beta = \frac{R_2+R_1}{R_2-R_1}$, $\gamma = \frac{R_2-R_1}{2}$, $r = \gamma(t+\beta)$, $u_m(t) = \tilde{u}_m(r)$, and $f_m(t) = \tilde{f}_m(r)$.

As before, we only have to consider the approximation of (3.20), where f_m and u_m are real functions.

We now describe a spectral-Galerkin scheme for (3.20). We shall look for an approximation of u_m in the tensor-product space

$$X_{NK}^{(m)} := Y_N^{(m)} \otimes Z_K \text{ or alternatively } X_{NK}^{(m)} := Z_N^{(m)} \otimes Z_K, \tag{3.21}$$

where $Y_N^{(m)}$, $Z_N^{(m)}$ and Z_K are defined, respectively, in (2.13), (2.26), and (3.12).

Let $\omega = \omega(t)$ be the weight function associated with the Legendre or Chebyshev polynomials and $I_{NMK} := I_{NM}\times J_K$ be the 3D interpolation operator, where I_{NM} is defined in section 2.1 and J_K is the interpolation operator associated with the Legendre or Chebyshev Gauss–Lobatto points in $[-1,1]$. We write

$$I_{NMK}f = \sum_{|m|=0}^{M} f_{NK}^{(m)}(\theta,t)e^{im\phi}, \quad \text{with } f_{NK}^{(m)} \in X_{NK}^{(m)}.$$

Then the weighted spectral-Galerkin method for (3.20) is to find $u_{NK}^{(m)} \in X_{NK}^{(m)}$ such that

$$\begin{aligned} &\alpha\gamma^2\big((t+\beta)^2\sin^2\theta u_{NK}^{(m)}, v\omega\big) - \big(\sin^2\theta\partial_t((t+\beta)^2\partial_t u_{NK}^{(m)}), v\omega\big) \\ &\qquad - \big(\sin\theta\partial_\theta(\sin\theta\partial_\theta u_{NK}^{(m)}), v\omega\big) + m^2\big(u_{NK}^{(m)}, v\omega\big) \\ &\qquad = \gamma^2\big((t+\beta)^2\sin^2\theta f_{NK}^{(m)}, v\omega\big) \quad \forall\, v \in X_{NK}^{(m)}, \end{aligned} \tag{3.22}$$

where $(u,v) = \int_{-1}^{1} dt\int_0^\pi uv d\theta$.

If we denote

$$\begin{aligned} &u_{NK}^{(m)} = \sum_{k,j} u_{kj}^{(m)}\psi_k^{(m)}(\theta)\phi_j(t), \ U^{(m)} = (u_{kj}^{(m)}), \\ &f_{kj}^{(m)} = \big((t+\beta)^2\sin^2\theta f_{NK}^{(m)}, \psi_k^{(m)}(\theta)\phi_j(t)\omega\big), \ F^{(m)} = (f_{kj}^{(m)}), \\ &\text{with } k = \begin{cases} 1,2,\ldots,N-1 & m\neq 0 \\ 0,1,\ldots,N & m=0 \end{cases}, \ j = 0,1,\ldots,K-2, \end{aligned} \tag{3.23}$$

then (3.22) is equivalent to the following matrix system:

$$\alpha\gamma^2 A^{(m)}U^{(m)}Q + A^{(m)}U^{(m)}R^t + B^{(m)}U^{(m)}S + m^2C^{(m)}U^{(m)}S = \gamma^2F^{(m)}, \tag{3.24}$$

where the matrices $A^{(m)}$, $B^{(m)}$, and $C^{(m)}$ are defined in (2.31)–(2.32), and Q, R, and S are defined in (3.14). We recall that $A^{(m)}$, $C^{(m)}$, Q, and S are symmetric, $B^{(m)}$ is nonsymmetric, and R is symmetric in the Legendre case and nonsymmetric in the Chebyshev case.

Since (3.24) is derived from the discretization of the separable elliptic equation (3.19), it can be solved by using the matrix diagonalization method which is, in fact, a discrete counterpart of *separation of variables*. To this end, we first solve the generalized eigenvalue problem

$$(\alpha\gamma^2 Q + R)G = SG\Lambda \text{ (i.e., } G^t(\alpha\gamma^2 Q + R^t) = \Lambda G^t S), \tag{3.25}$$

where $\Lambda = \mathrm{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{K-2})$ and G is formed by the eigenfunctions. Note that the ellipticity of the equation ensures that all the eigenvalues are real and positive even in the Chebyshev case when R is nonsymmetric.

Making the transform $U^{(m)} = V^{(m)}G^t$ in (3.24) and using (3.25), we find

$$A^{(m)}V^{(m)}\Lambda G^t S + (B^{(m)} + m^2 C^{(m)})V^{(m)}G^t S = F^{(m)}. \tag{3.26}$$

Therefore,

$$A^{(m)}V^{(m)}\Lambda + (B^{(m)} + m^2 C^{(m)})V^{(m)} = F^{(m)}S^{-1}G^{-t} := \tilde{F}^{(m)}. \tag{3.27}$$

Let $\boldsymbol{v}_j^{(m)}$ and $\tilde{\boldsymbol{f}}_j^{(m)}$ be, respectively, the jth column of $V^{(m)}$ and $\tilde{F}^{(m)}$. Then, the above equation splits up into a sequence of one-dimensional equations,

$$(\lambda_j A^{(m)} + B^{(m)} + m^2 C^{(m)})\boldsymbol{v}_j^{(m)} = \tilde{\boldsymbol{f}}_j^{(m)}, \quad j = 0, 1, \ldots, K-2, \tag{3.28}$$

which can be solved in $O(NK)$ operations. Hence, neglecting the preprocessing cost for computing G and Λ, the cost of solving (3.24) is dominated by the two matrix multiplications $U^{(m)} = V^{(m)}G^t$ and $(F^{(m)}S^{-1})G^{-t}$, which require about $2K^2N$ flops. Again, one can use three different approaches, namely, Legendre–Galerkin, Chebyshev–Galerkin, and Chebyshev–Legendre Galerkin, for the discretization of the vertical variable.

Although the operation count of this algorithm is not optimal, but since in most applications the vertical scale is much smaller than the horizontal scales so that K can be chosen to be much smaller than N, this algorithm is still very competitive, thanks to the fast Fourier transform, when compared with the algorithm using spherical harmonics. The question of which algorithm is superior will depend on, among others, the size of the problem, computer platform, and availability of machine coded BLAS or FFT.

REMARK 3.2. *Elliptic problems in the unbounded domain* $\Omega = \{(x, y, z) : R_1 < x^2 + y^2 + z^2\}$ *can be handled in a similar fashion by discretizing the radial direction using Laguerre polynomials, if the solution decays exponentially at infinity* (*cf.* [9]), *or mapped Chebyshev or Legendre functions* (*cf.* [3]).

4. Concluding remarks. We have presented several fast algorithms for elliptic equations in spherical geometries by using double Fourier series for the horizontal variables and Chebyshev or Legendre polynomials for the vertical variable. Our algorithms for elliptic problems on the sphere have quasi-optimal (optimal up to a logarithmic term) computational complexity as well as optimal error estimates. We have performed numerical experiments which indicate that our algorithms are significantly more efficient and/or accurate when compared with the algorithms based on

spherical harmonics and on finite difference. The numerical experiments also confirm that *nonessential* pole conditions can be safely ignored, at least for steady problems, in agreement with the arguments in [12] and [2]. For time dependent problems, it is also generally believed that ignoring *nonessential* pole conditions will not lead to unreasonable time step restrictions as long as the principle elliptic operator is treated implicitly [12].

It is hoped that this paper will help to revive researchers' interests in using double Fourier series for elliptic or parabolic type equations in spherical geometries.

Appendix A. The nonzero entries of Q, R, and S. Let Q, R, and S be defined in (3.14) with $\omega \equiv 1$ and $\phi_j(t) = L_j(t) - L_{j+2}(t)$. Then, their nonzero entries are

$$q_{kj} = q_{jk} := \int_I (t+\beta)^2 \phi_j \phi_k dt$$
$$= \begin{cases} -\dfrac{2(j+3)(j+4)}{(2j+5)(2j+7)(2j+9)}\,, & k = j+4\,, \\ -2\beta\dfrac{2(j+3)}{(2j+5)(2j+7)}\,, & k = j+3\,, \\ \dfrac{2(j+2)}{(2j+1)(2j+5)^2} - \dfrac{2(j+3)}{(2j+9)(2j+5)^2} - \beta^2\dfrac{2}{2j+5}\,, & k = j+2\,, \\ 2\beta\left(\dfrac{2}{(2j+1)(2j+5)} + \dfrac{2(j+3)}{(2j+5)(2j+7)}\right), & k = j+1\,, \\ \dfrac{2j^2}{(2j+1)^2(2j-1)} + \dfrac{2(2j+3)}{(2j+1)^2(2j+5)^2}\,, & k = j\,, \\ +\dfrac{2(j+3)^2}{(2j+5)^2(2j+7)} + \beta^2\left(\dfrac{2}{2j+1} + \dfrac{2}{2j+5}\right), & \end{cases}$$

$$r_{kj} = r_{jk} := \int_I (t+\beta)^2 \phi_j' \phi_k' dt = (2j+3)(2k+3)\int_I (t+\beta)^2 L_{j+1} L_{k+1} dt$$
$$= \begin{cases} \dfrac{2(j+2)(j+3)}{2j+5}\,, & k = j+2\,, \\ 4\beta(j+2)\,, & k = j+1\,, \\ \dfrac{2(j+2)^2}{2j+5} + \dfrac{2(j+1)^2}{2j+1} + 2\beta^2(2j+3)\,, & k = j\,, \end{cases}$$

$$s_{kj} = s_{jk} := \int_I \phi_j \phi_k dt = \begin{cases} -\dfrac{2}{2j+5}\,, & k = j+2\,, \\ \dfrac{2}{2j+1} + \dfrac{2}{2j+5}\,, & k = j\,. \end{cases}$$

REFERENCES

[1] B. K. Alpert and V. Rokhlin, *A fast algorithm for the evaluation of Legendre expansions*, SIAM J. Sci. Stat. Comput., 12 (1991), pp. 158–179.

[2] J. P. Boyd, *The choice of spectral functions on a sphere for boundary and eigenvalue problems: A comparison of Chebyshev, Fourier and associated Legendre expansions*, Monthly Weather Rev., 106 (1978), pp. 1184–1191.

[3] J. P. Boyd, *Orthogonal rational functions on a semi-infinite interval*, J. Comput. Phys., 70 (1987), pp. 63–88.

[4] J. P. Boyd, *Chebyshev and Fourier Spectral Methods*, Springer-Verlag, New York, 1989.

[5] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Vol. 1, Interscience Publishers, New York, 1953.

[6] J. R. Driscoll and D. M. Healy, *Computing fourier transforms and convolutions on the 2-sphere*, Adv. in Appl. Math., 15 (1994), pp. 202–250.

[7] B. Fornberg, *A pseudospectral approach for polar and spherical geometries*, SIAM J. Sci. Comput., 16 (1995), pp. 1071–1081.

[8] B. Fornberg, *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, London, 1996.

[9] Y. Maday, B. Pernaud-Thomas, and H. Vandeven, *Reappraisal of Laguerre type spectral methods*, Rech. Aerospat., 6 (1985), pp. 13–35.

[10] M. J. Mohlenkamp, *A Fast Transform for Spherical Harmonics*, Ph.D. thesis, Yale University, New Haven, CT, 1997.

[11] P. M. Morse and H. Feshback, *Methods of Theoretical Physics*, McGraw-Hill, New York, 1953.

[12] S. A. Orszag, *Fourier series on spheres*, Monthly Weather Rev., 102 (1974), pp. 56–75.

[13] L. Quartapelle, *Numerical Solution of the Incompressible Navier-Stokes Equations*, Birkhäuser, Boston, Cambridge, MA, 1993.

[14] J. Shen, *Efficient spectral-Galerkin method* I. *Direct solvers for second- and fourth-order equations by using Legendre polynomials*, SIAM J. Sci. Comput., 15 (1994), pp. 1489–1505.

[15] J. Shen, *Efficient spectral-Galerkin method* II. *Direct solvers for second- and fourth-order equations by using Chebyshev polynomials*, SIAM J. Sci. Comput., 16 (1995), pp. 74–87.

[16] J. Shen, *Efficient Chebyshev-Legendre Galerkin methods for elliptic problems*, in Proceedings of the ICOSAHOM'95, Houston J. Math., (1996), pp. 233–240.

[17] J. Shen, *Efficient spectral-Galerkin methods* III. *Polar and cylindrical geometries*, SIAM J. Sci. Comput., 18 (1997).

[18] P. N. Swarztrauber, *The approximation of vector functions and their derivatives on the sphere*, SIAM J. Numer. Anal., 18 (1981), pp. 191–210.

[19] P. N. Swarztrauber and R. A. Sweet, *Efficient FORTRAN subprograms for the solution of elliptic partial differential equations*, ACM Trans. Math. Software, 5 (1979), pp. 352–364.

[20] S. Y. K. Yee, *Solution of Poisson's equation on a sphere by truncated double Fourier series*, Monthly Weather Rev., 109 (1981), pp. 501–505.