Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 7999–8016
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

A new method for fast transforms in parity-mixed PDEs: Part I. Numerical techniques and analysis

Geoffrey M. Vasil [a,*], Nicholas H. Brummell [b,c], Keith Julien [d]

[a] Department of Atmospheric and Oceanic Sciences and JILA, University of Colorado, Boulder, CO 80309, United States
[b] Department of Astrophysical and Planetary Sciences and JILA, University of Colorado, Boulder, CO 80309, United States
[c] Department of Applied Mathematics, University of California, Santa Cruz, CA 95064, United States
[d] Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, United States

Received 16 October 2007; received in revised form 22 April 2008; accepted 22 April 2008
Available online 6 May 2008

Abstract

We address the problem of parity mixing, where the projection of a variable expressed as a finite series of half-period cosine (sine) functions onto a half-period sine (cosine) function basis is not finite. We propose new fast methods for computing these complicated projections exactly up to some arbitrary degree using fast Fourier transforms. This method has immediate applications for pseudospectral solutions of many systems of partial differential equations.
Published by Elsevier Inc.

Keywords: Numerical methods; Computational fluid dynamics; Pseudospectral methods; Dealiasing; Parity mixing

1. Introduction

Fourier spectral and pseudospectral methods are important tools in the numerical solution of many nonlinear partial differential equations (PDEs). These methods are appealing because of their many advantages such as their efficiency and good convergence properties (see [1,2] for detailed review). In Fourier spectral/pseudospectral methods, there are three basis function sets in common use: (i) the "full" Fourier series, (ii) the Fourier sine series, and (iii) the Fourier cosine series. The full Fourier series requires both sine and cosine functions for completeness and is the method of choice when periodic geometries are involved. For compact domains with nontrivial boundaries and one or more rectilinear dimensions (i.e., confined domains) one can employ either a sine series or a cosine series. For these second two cases (Fourier sine series and Fourier cosine series), both sets are individually complete bases for representing square-integrable functions on a compact domain. While both basis sets will typically converge for wide classes of functions (independent of the type of boundary conditions), the particular choice of basis function set is typically strongly suggested by the requirements of the individual problem under consideration [3]. Usually boundary conditions make this choice

* Corresponding author. Tel.: +1 303 492 7851; fax: +1 303 907 4604.
E-mail address: geoffrey.vasil@colorado.edu (G.M. Vasil).

0021-9991/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jcp.2008.04.020

somewhat apparent, e.g., Dirichlet (sines), or Neumann (cosines). In particular, while the sine and cosine representations often do not converge geometrically, for problems with algebraic convergence, selecting the appropriate basis can optimize the rate of algebraic convergence (see, e.g., [1], Section 2.9).

In many problems, one often relies heavily on the fast Fourier transform (FFT) to compute nonlinearities efficiently. For all three basis sets (full Fourier series, Fourier sine series, and Fourier cosine series), the FFT can be used to transform rapidly (i.e., in $\mathcal{O}(N \log N)$ instead of $\mathcal{O}(N^2)$ operations) between a grid space where multiplication is a local operation and Fourier space where derivatives are local operations. One key requirement for the accurate calculation of the FFT is that one uses a sufficient number of discrete points (determined by the number of modes contained in the spectrum) so that the transform is not contaminated by aliasing errors (unwanted power in the resolved spectrum from unresolved modes).

Nonlinearities will tend to broaden the spectrum of a dynamic variable over time and aliasing errors can accumulate in the solution of nonlinear PDEs. Dealiasing techniques are able to account for spectrum broadening for a quadratic nonlinearity; see [4]. This procedure is commonly known as Orszag's 2/3-rule (see [1] Chap. 11, Section 5). This rule can easily be generalized to any polynomial nonlinearity of finite-degree. However, dealiasing with a straightforward application of Orszag's rule is *guaranteed* possible only when using the *full* Fourier series on periodic domains. On a full periodic domain, a solution that is approximated with band-limited functions (i.e., is expanded in a finite number of elements) will remain band-limited as long as the governing system only contains finite-degree nonlinearities, (i.e., polynomial nonlinearities as opposed to more general analytic functions, such as in the sine-Gordon equation for example). Specifically, the finite-degree property allows for effective dealiasing. The reason for this is ultimately tied to the fact that on a full periodic domain, all of the basis elements are mutually orthogonal to each other. In particular, any sine basis function is orthogonal to any cosine basis function.

Mutual orthogonality between sine and cosine functions is lost when one moves away from periodic domains to confined domains. In the latter case, even a polynomial nonlinearity can act like a nonlinearity of infinite-degree. Since the Fourier sine series and the Fourier cosine series are each complete orthogonal sets, any element from one set must be expressible as a linear combination of elements from the other. In other words, the Fourier cosine series and Fourier sine series cannot be orthogonal to one another. It is even more problematic that any particular sine basis element is infinitely broadband when represented as a series of cosine functions, and vice versa. A difficulty can arise when a finite-degree (e.g., quadratic or even linear) term in a dynamical equation causes a term of one *parity* (sine or cosine series) to be *mixed* (linearly combined) with a term of the opposite parity (cosine or sine series). When one attempts to project this term onto the appropriate basis, the result is that a finite-degree term has infinitely broadened the spectrum. This broadening cannot be controlled with any standard generalization of Orszag's rule. We call this problem "parity mixing." Likewise, we call the solution to this problem "parity filtering." We are, therefore, prompted to provide the following two definitions.

Definition 1.1. Parity mixing – Any process that linearly combines a function represented by one type of trigonometric series with another function represented by a complementary trigonometric series and thereby producing infinite spectrum broadening.

Definition 1.2. Parity filtering – Any method for extracting Fourier coefficients of a desired type from a function that is naturally represented with a trigonometric series of a complementary type.

One of the most severe drawbacks to Fourier spectral methods is that parity mixing can ostensibly limit the geometry to periodic domains. There are very few natural problems that are genuinely periodic in more than one direction. For many years, researchers who wished to study problems in confined domains have often chosen one of three alternatives when faced with this fact. They have either used periodic Fourier methods and then made justifications as to the reasonableness of a periodic domain (e.g., by arguing that internal dynamics are unaffected by boundary conditions); they have resorted to expensive spectral-Galerkin (non-pseudospectral) methods; or, they have simply abandoned traditional Fourier methods. This paper addresses the parity mixing problem for the case of sine or cosine basis function sets and shows that the introduction of a special function, Id, a truncated spectral expansion of unity, can be used to rectify the problem and enable a fast, stable, and mathematically exact calculation of the problematic terms.

1.1. Physical examples of problems containing parity mixing

We offer the following two examples as concrete illustrations of the concept of parity mixing. The second of these examples is significantly expanded in Part II of this paper, where we employ new parity-filtering methods to solve for rotating thermal convection in a confined box [5].

Example 1.1. Suppose we wish to solve the Boussinesq fluid equations (see [6]) in a rectilinear rigid box. For simplicity, we set all three dimensions of the box equal, $L_x = L_y = L_z = \pi$. For the purpose of illustration, we focus only on the first few terms on the left-hand side of the first component of the Boussinesq equations

$$(\partial_t - \nu\Delta)u + u\partial_x u = \cdots \tag{1}$$

The terms in Eq. (1) are common in many advection-diffusion systems of PDE's and are sufficient to illustrate the issues associated with parity mixing.

Impenetrable and no-slip (fully homogenous) boundary conditions suggest that we should represent the flow as a triple-sine series. From this, we can define the wavevector, $\boldsymbol{k} = (l, m, n)$, and the normalized basis function

$$\psi_{\boldsymbol{k}}(x, y, z) = (2/\pi)^{3/2} \sin(lx)\sin(my)\sin(nz), \tag{2}$$

and then

$$u(x, y, z, t) = \sum_{|\boldsymbol{k}| \leqslant K} A_{\boldsymbol{k}}(t)\psi_{\boldsymbol{k}}(x, y, z), \tag{3}$$

where in practice we can only use a finite number of modes, i.e., $|\boldsymbol{k}| \leqslant K$. We now need only to solve for the amplitudes, $A_{\boldsymbol{k}}(t)$, *viz*.,

$$(\partial_t + \nu k^2)A_{\boldsymbol{k}} + \int_V \psi_{\boldsymbol{k}} u\partial_x u \, \mathrm{d}^3\mathbf{x} = \cdots, \tag{4}$$

where $V = [0, \pi]^3$ is the three-dimensional domain volume. Recall that $u(x, y, z)$ is represented with a triple-sine series

$$u \sim \sin(lx)\sin(my)\sin(nz). \tag{5}$$

This means that the derivative, $\partial_x u$, is a cosine-sine-sine series

$$\partial_x u \sim \cos(lx)\sin(my)\sin(nz). \tag{6}$$

Thus we have

$$u\partial_x u \sim \sin(l'x)\cos(m'y)\cos(n'z), \tag{7}$$

where the triplet, $\{l', m', n'\}$, is used to distinguish a typical wavenumber of the product from the original typical wavenumber, $\{l, m, n\}$. Therefore, if u is a band-limited triple-sine series, then $u\partial_x u$ is a band-limited sine series in the x-direction and a band-limited cosine series in the y and z-directions.

The fact that $u\partial_x u$ is a band-limited sine series in the x-direction is fortunate since we can then use a fast sine transform to project the x-direction of $u\partial_x u$ onto the x-direction of $\psi_{\boldsymbol{k}}$. The y and z directions are not quite so simple since a cosine function cannot be written as a band-limited sine series. Indeed, the projection of a cosine (sine) onto a sine (cosine) expansion requires an infinite number of inner products of an arbitrary sine function and an arbitrary cosine function. For example, in the z-direction

$$\mathcal{P}(n_1, n_2) = \int_0^\pi \sin(n_1 z)\cos(n_2 z)\mathrm{d}z = \begin{cases} \frac{2n_1}{n_1^2 - n_2^2} & \text{if } n_1 + n_2 \text{ is odd} \\ 0 & \text{if } n_1 + n_2 \text{ is even.} \end{cases} \tag{8}$$

Eq. (8) implies that if we were to feed $u\partial_x u$ naively into a (finite) fast sine transform, the y and z-directions would come out incorrectly as a result of the band-unlimited nature of a cosine function when represented by sine functions. The manifestation of this occurs as aliasing errors from the truncated part of the sine series.

Example 1.2. Parity mixing does not only stem from nonlinearities. Consider a rotating fluid occupying the domain $V = [0,\pi]^3$. In this example, the walls of the domain are considered impenetrable and stress-free rather than impenetrable and no-slip. Impenetrability implies that the normal component of velocity vanish at a boundary, while the stress-free condition implies that the normal derivatives of the tangential components vanish at a boundary. If the rotation axis is in the vertical direction, with a rotation rate Ω, then the first few terms in the horizontal components of the momentum equations are

$$\partial_t u - 2\Omega v = -\partial_x P + \cdots \tag{9}$$

$$\partial_t v + 2\Omega u = -\partial_y P + \cdots. \tag{10}$$

Again, this is not an unusual situation; see [7]. In this example, the boundary conditions suggest that each velocity component is a sine series in its own direction and a cosine series in the other two directions; $u \sim \sin(lx)\cos(my)\cos(nz)$ and $v \sim \cos(lx)\sin(my)\cos(nz)$. It is clear that the Coriolis accelerations in a stress-free box produce parity mixing in the horizontal directions, just as nonlinear advection does in a rigid box, i.e., u and v have different parities. In this rotating example, parity mixing arises even for linear dynamics.

While the above two examples focus on rotating and non-rotating hydrodynamics, parity filtering easily arises in many situations where one or more additional interacting physical quantities are included. The details of course depend on the specific boundary conditions imposed on each given variable. Examples of other relevant quantities include scalar fields, such as temperature and solutal concentration, and vectors such as magnetic fields.

1.2. Parity filtering

We must reconcile the parity mixing in the above two examples. For instance, in Example 1.2, ignoring the z-direction, suppose we have the velocity component

$$v = \sum_{l=0}^{N'} \sum_{m=1}^{N} a_{l,m} \cos(lx)\sin(my), \tag{11}$$

where the coefficients, $a_{l,m}$, are known. For notational convenience, we employ the prime-summation notation throughout this paper, i.e., for any sequence, c_l with $l \geqslant 0$

$$\sum_{l=0}^{N'} c_l = \frac{c_0}{2} + \sum_{l=1}^{N} c_l. \tag{12}$$

To solve Eqs. (9) and (10), we need to know how the series in Eq. (11) for v projects onto the cosine basis for u up to order N so that we may compute the Coriolis accelerations in Eq. (9). That is, for a function (such as v) of one *parity*, we need to *filter* out a complementary function of the opposite parity

$$\hat{v} = \sum_{l=1}^{N} \sum_{m=0}^{N'} b_{l,m} \sin(lx)\cos(my). \tag{13}$$

Furthermore, we want to choose this function such that the normed difference, $\|v - \hat{v}\|$, is minimized. If $\|\circ\|$ represents the L^2 norm on $[0,\pi]^2$, then for $\{l,m\} \subset \{0,\ldots,N\}$ we should choose the standard Fourier coefficients

$$b_{l,m} = \frac{4}{\pi^2} \sum_{l_0=0}^{N'} \sum_{m_0=1}^{N} \mathcal{P}(l,l_0)\mathcal{P}(m_0,m) a_{l_0,m_0}, \tag{14}$$

where $\mathcal{P}(n',n)$ is given by Eq. (8). Alternatively, if $\|\circ\|$ represents the H^1 norm on $[0,\pi]^2$, then a detailed analysis produces the coefficients

$$b_{l,m} = \frac{4}{\pi^2} \sum_{l_0=0}^{N'} \sum_{m_0=1}^{N} \frac{1 + l_0^2 + m^2}{1 + l^2 + m^2} \mathcal{P}(l,l_0)\mathcal{P}(m_0,m) a_{l_0,m_0}. \tag{15}$$

Eqs. (14) and (15) are examples of what we call parity filters. In either scenario, a filter is applied through a multiplication with a succession of $\mathcal{P}$ matrices or their adjoints under a suitable metric (inner product and associated norm).

The matrix multiplications in Eqs. (14) and (15) represent extremely expensive methods for parity filtering, i.e., $\mathcal{O}(N^2)$ operations in each direction. The coefficients, $b_{l,m}$, could also be computed using a quadrature formula directly from the function, $v(x,y)$, on a chosen grid. Typically, this would also involve the computation of a sum over each dimension (l_0 and m_0) for each value of l and m. From now on, we refer to such sums (e.g., Eqs. (14) and (15), or quadrature sums) as *slow* parity filters. The purpose of this paper is to derive a more efficient method for computing such operations. From now on, we typically refer to any technique for *fast* parity filtering as simply parity filtering.

For the two examples from Section 1.1 (and many other PDEs), we posit a new effective procedure that will exactly rectify apparent parity mixing difficulties. Our method of fast parity filtering is equivalent to choosing a particular set of quadrature weights so that we can employ fast transforms to compute a collection of spectral coefficients without aliasing errors. In general, any type of nonlinear (or linear) operation on a band-limited function that can be dealiased when the domain is periodic can also be parity filtered. This is useful since there are many optimized pseudospectral codes in existence that utilize a full Fourier series to solve problems in periodic geometries. We show that with a reasonable amount of effort, these codes could be generalized to solve problems in periodic channels or confined boxes.

2. Mathematical analysis of the spectral expansion of unity

The aliasing errors caused by the product of two band-limited sine series being a *band-unlimited sine series* (i.e., a band-limited cosine series) can be corrected. Consider the Fourier sine series of the number 1 on an interval of length, say π. This is the same as representing $\cos(nz)$ for $n = 0$ with a sine series, where $z \in [0,\pi]$. We can, therefore, use Eq. (8) and define the truncated expansion of unity for $z \in [0,\pi]$ as

$$\mathrm{Id}_m(z) = \frac{4}{\pi}\sum_{n=1}^{m}\frac{\sin((2n-1)z)}{2n-1}. \tag{16}$$

The above spectral expansion of unity is advantageous since it can be used to flip the parity of trigonometric functions without changing their summed value. This property allows the transformation of a cosine series into an equivalent sine series and vice versa. This is valuable because the coefficients of this equivalent sine series can be computed via a fast Fourier sine transform.

Using the identity function, $\mathrm{Id}_m(z)$, we can now consider integrals of the type in Eq. (8).

Proposition 2.1. *For all* $m \geqslant \lceil (n_1+n_2)/2 \rceil$, *we have*

$$\int_0^\pi \sin(n_1 z)\cos(n_2 z)\mathrm{Id}_m(z)\,\mathrm{d}z = \int_0^\pi \sin(n_1 z)\cos(n_2 z)\,\mathrm{d}z. \tag{17}$$

Proposition 2.1 says that for large enough m the integrals on the left and right-hand sides of Eq. (17) are totally equivalent, i.e., the inclusion or exclusion of Id_m from the integrand makes no difference.

We will prove Proposition 2.1 shortly. However, the main reason Eq. (17) is helpful is that the product, $\cos(n_2 z)\mathrm{Id}_m(z)$, is a band-limited sine series, and $\sin(n_1 z)\mathrm{Id}_m(z)$ is a band-limited cosine series. This means that for a range for either n_1 or n_2, the projection on the right-hand side of Eq. (17) can be computed exactly via a fast Fourier sine transform (FFST) or a fast Fourier cosine transform (FFCT). That is, we have the sampling result for parity-mixed functions.

Corollary 2.2. *For* $N = 2m$ *and the collocation points,* $z_i = \pi(2i-1)/2N$ *with* $i \in \{1,\ldots,N\}$, *we have*

$$\int_0^\pi \sin(n_1 z)\cos(n_2 z)\,\mathrm{d}z = \frac{\pi}{N}\sum_{i=1}^{N}\sin(n_1 z_i)\cos(n_2 z_i)\mathrm{Id}_m(z_i). \tag{18}$$

Eq. (18) follows directly from Eq. (17) and the sampling theorem for trigonometric functions with even or odd parity (see [8], Section 12.1). Note that, unlike in the standard sampling theorem, the number of possible

wavenumbers in the spectrum is equal to, rather than half, the number of sampling points. Furthermore, if we fix (say) n_2 and vary n_1 over a range of size N, then the sums on the right-hand side of Eq. (18) can be computed in $\mathcal{O}(N \log N)$ operations (see [8], Section 12.2). More specifically, if we have a function

$$f(z) = \sum_{n_2=0}^{n'} a_{n_2} \cos(n_2 z), \tag{19}$$

then we can compute all n integrals (equivalently quadrature sums)

$$\int_0^\pi \sin(z) f(z)\,\mathrm{d}z, \quad \int_0^\pi \sin(2z) f(z)\,\mathrm{d}z, \quad \ldots \quad \int_0^\pi \sin(nz) f(z)\,\mathrm{d}z,$$

in $\mathcal{O}(n \log n)$ operations, rather than directly computing n individual sums each with $\mathcal{O}(n)$ complexity for a total $\mathcal{O}(n^2)$ complexity. This $\mathcal{O}(n \log n)$ method is a pseudospectral approach to parity filtering.

To prove Proposition 2.1 and hence, Corollary 2.2, we must establish two essential properties of this expansion of unity, Id_m, namely that it is almost everywhere convergent and uniformly bounded; independent of the degree of truncation, m. These two properties allow invocation of the Lebesgue dominated-convergence theorem and, therefore, the interchange of limits and integration. In this section, we will briefly outline the proof of Proposition 2.1. More detailed analysis can be found in Appendix A.

Claim 2.3. For almost every $z \in [0, \pi]$, $\lim_{m\to\infty} \mathrm{Id}_m(z) = 1$.

Dirichlet's Fourier series conditions or Dini's test guarantee Claim 2.3; see [3]. The exception to convergence is on the isolated set of measure zero, $z \in \{0, \pi\}$, where $\mathrm{Id}(z) = 0$. More importantly, however, we have $\mathrm{Id}_m(z) \to 1$ for almost all $z \in [0, \pi]$. Also, see Appendix A for a direct proof of the convergence of Id_m.

While $\mathrm{Id}_m(z)$ converges everywhere in the interval $0 < z < \pi$, it does not converge uniformly. It cannot, or else it would converge at the same rate arbitrarily close to the boundary as it does in the interior. This would not allow the function to vanish eventually at the boundary. Therefore, even though $\mathrm{Id}_m(z)$ converges, we know its derivatives are unbounded. However, the rate of divergence can be controlled.

Claim 2.4. *There exists a constant, I_0, (independent of m and z) such that for all $m \geqslant 1$ and all $z \in [0, \pi]$, we have* $|\ \mathrm{Id}_m(z)\ | \leqslant I_0$.

A proof of Claim 2.4 can be found in Appendix A.

Having shown that $\mathrm{Id}_m(z)$ is uniformly bounded and converges to unity for almost all $z \in [0, \pi]$, we can use these properties to our benefit.

Corollary 2.5. *For any integrable function $\varphi(z)$ on the interval $[0, \pi]$, we have*

$$\lim_{m\to\infty} \int_0^\pi \varphi(z) \mathrm{Id}_m(z)\,\mathrm{d}z = \int_0^\pi \varphi(z)\,\mathrm{d}z. \tag{20}$$

Proof. (Corollary 2.5) From Claims 2.3 and 2.4 we know that the sequence of functions, $\varphi(z)\mathrm{Id}_m(z)$, converges almost everywhere to $\varphi(z)$ and that this sequence is also uniformly bounded by $|\ \varphi(z)\mathrm{Id}_m(z)\ | \leqslant I_0\ |\ \varphi(z)\ |$. These facts allow the invocation of the Lebesgue dominated-convergence theorem, and we can interchange the limit and integration in Eq. (20); see [9], Section 1.8. □

Proof. (Proposition 2.1) Eq. (20) implies

$$\lim_{m\to\infty} \int_0^\pi \sin(n_1 z)\cos(n_2 z)\mathrm{Id}_m(z)\,\mathrm{d}z = \int_0^\pi \sin(n_1 z)\cos(n_2 z)\,\mathrm{d}z. \tag{21}$$

However, since the function $\sin(n_1 z)\cos(n_2 z)$ only has power up to $\sin((n_1 + n_2)z)$, it is orthogonal to all the terms in the series representation of $\mathrm{Id}_m(z)$ that are higher order than $n_1 + n_2$. Recall that $\mathrm{Id}_m(z)$ has power up to order $2m - 1$. We, therefore, do not need to take the full limit in Eq. (21); we only need m to be sufficiently large. If $n_1 + n_2$ is even, then by symmetry both sides of Eq. (17) vanish identically. If $n_1 + n_2$ is odd, then we must have $m \geqslant (n_1 + n_2)/2 + 1/2 = \lceil (n_1 + n_2)/2 \rceil$. □

While it is difficult to imagine that Proposition 2.1 is entirely new, we have not found any example of such a result employed in the numerical/spectral analysis literature. Proposition 2.1 can be interpreted as the first half of a periodically extended Fourier projection integral. However, when considering the periodic extension of, say, a sine function, integrated against a cosine function, it is necessary to think of the sine function not as its normal periodic self, but rather as an *even* extension of a naturally *odd* function. A similar argument follows if we were to consider a cosine function integrated against a sine function. While Eq. (17) may seem obvious in the context of periodic extension, we find it novel that equality is achieved in Eq. (17) with only a finite number of terms in the expansion, Id_m. This property is very useful in designing a fast numerical scheme for the pseudospectral solution of PDEs. In the next section we discuss the detail behind implementing such a scheme.

3. Application of the spectral expansion of unity and dealiasing

We use the spectral expansion of unity to preserve the applicability (accuracy) of fast transforms under parity violating operations (e.g., Examples 1.1 and 1.2). The costs we incur to compute correctly the projections of cosines onto sines and vice versa are larger transforms. To see this, consider two band-limited sine series,

$$f(z) = \sum_{j=1}^{n} a_j \sin(jz); \quad g(z) = \sum_{j=1}^{n} b_j \sin(jz). \tag{22}$$

We could just as easily choose to examine one sine series and one cosine series but Eq. (22) will suffice for illustration. Also, there is no reason to examine the product of two cosine series since this produces a cosine series and normal dealiasing can account for this case. We wish to compute how the product of $f(z)$ and $g(z)$ projects onto a sine basis up to order n, i.e.,

$$f(z)g(z) = \sum_{j=0}^{2n} \tilde{c}_j \cos(jz) = \sum_{j=1}^{\infty} c_j \sin(jz). \tag{23}$$

That is, we want a fast error free pseudospectral way to compute the coefficients, c_j for $j \in \{1, \ldots, n\}$, without doing any cumbersome [i.e., $\mathcal{O}(n^2)$] calculations which require explicit use of Eq. (8).

If we know the coefficients, a_j and b_j for $j \in \{1, \ldots, n\}$, we can perform a FFST to obtain $f(z)$ and $g(z)$ in grid space and multiply the functions on collocation points. However, this product is now effectively a cosine series with power up to order $2n$, and so to transform this into an equivalent sine series, we must multiply $f(z)g(z)$ by $\mathrm{Id}_m(z)$ for some m. We can then take the inverse FFST of the result and obtain the correct values for the coefficients, c_j.

A remaining question is: how big must m be? To answer this, recall that we are implicitly computing integrals of the form

$$\int_0^{\pi} \sin(kz)\cos(jz)\mathrm{Id}_m(z)\,\mathrm{d}z, \tag{24}$$

with $1 \leqslant k \leqslant n$ and $0 \leqslant j \leqslant 2n$. The index j goes up to $2n$ because the cosine term in the integrand represents one component of the product of two sine functions each of which can have power up to order n. Therefore, we may take $m = \lceil (n + 2n)/2 \rceil = \lceil 3n/2 \rceil$ (or greater with no extra benefit by Proposition 2.1), so that the maximum order in Id_m, (i.e., $2m - 1$) is the largest odd integer which is less than or equal to the transform size, $N = 3n$. This fact implies that we must compute multiplications on a grid that is three times larger than the grid we would use if there were no quadratic nonlinearities and a grid that is two times larger than we would use if there were no parity-mixed quadratic nonlinearities. However, the factor of two cost over the non-parity-mixed quadratic nonlinearities (as in the case of a periodic box) is mitigated because we need only use real-to-real transforms rather than complex-to-real and real-to-complex transforms; see [10]. Finally, we must check that the choice of grid size, $N = 3n$, does not cause any additional aliasing error. In doing this, we see that there is a slight difference between the cases for when n is even and for when n is odd.

First, assume that n is an even integer. Since $2m - 1$ is the largest odd integer which is less than or equal to $3n$, we have $2m - 1 = 3n - 1$. Therefore, the product, $f(z)g(z)\mathrm{Id}_m(z)$, has power up to order $n + n + (3n - 1) = 5n - 1$. Since, for some arbitrary k, the basic aliasing rule is $N + k \to N - k$, this power

gets aliased to order $5n-1 = 3n + (2n-1) \to 3n - (2n-1) = n+1$ which is out of the range of interest (i.e., $\{0,\ldots,n\}$) and gets deleted. Therefore, when n is even, the choice of grid size, $N = 3n$, does not cause any additional aliasing error.

Next, assume that n is an odd integer and that $2m - 1 = 3n$. Then, the product, $f(z)g(z)\mathrm{Id}_m(z)$, has power up to order $n + n + 3n = 5n$. This power gets aliased to order $3n - 2n = n$. Since this is in the range of interest, as it stands now, aliasing to order n might cause a concern when n is odd. Given the general desire for transforms sizes that are combinations of small prime numbers (such as powers of two), cases when n is odd are not often encountered in practice. Nevertheless, if the situation of odd n were to arise, then we could always pad our transforms by one extra element, i.e., we could make the replacement $N = 3n \to 3n+1$. This replacement (while incurring a minor computational cost) would render the transform size even and the above results would apply. However, in the case when n is odd, aliasing errors are naturally avoided without additionally increasing N. Showing this result is somewhat lengthy and therefore the analysis can be found in Appendix B.

The above analysis, and Appendix B, carefully examines the aliasing for a particular nonlinearity, i.e., a quadratic nonlinearity. In general, we can easily determine what happens in the face of some general nonlinearity of a given degree, say p. For a usual pth order nonlinearity that does not contain any parity mixing, we know from a generalization of Orszag's rule that we must pad our transforms by a factor of $\lfloor (p+1)/2 \rfloor$. However, for a parity-mixed nonlinearity, we know that we must include a multiplication by the identity function, Id, which must have power up to order $p+1$ times the degree of the functions we are trying to represent spectrally. This makes a pth order term look like a $(2p+1)$th order term. Therefore, a pth order parity-mixed term must be padded by a factor of $\lfloor ((2p+1)+1)/2 \rfloor = p+1$. This is also the same factor of bandwidth that the identity function, Id, must contain over our dynamic field variables.

4. Consideration of errors

We shall now examine the general numerical application and the errors associated with the formulae discussed above. Specifically, we wish to examine the errors which are incurred when we use various methods to compute the coefficients, $c_{j,k}$, in the series

$$\cos(jz) = \sum_{k=1}^{\infty} c_{j,k} \sin(kz). \tag{25}$$

We outline three methods for computing the coefficients, $c_{j,k}$. We denote the various methods with a superscript, either 0,1,2. The first way to compute the coefficients is with the explicit use of Eq. (8). Therefore, define the exact coefficient values

$$c_{j,k}^{(0)} = \begin{cases} \frac{4k}{\pi(k^2-j^2)} & \text{if } j+k \text{ is odd} \\ 0 & \text{if } j+k \text{ is even.} \end{cases} \tag{26}$$

The first alternate method for attempting to compute the coefficients in Eq. (25) is by naively taking the discrete sine transform of $\cos(jz_i)$ on the grid $z_i = \pi(2i-1)/2N$, with $i \in \{1,\ldots,N\}$. We must choose these collocation points since they are the points for which both a sine and a cosine transform are mutually defined. Whence,

$$c_{j,k}^{(1)} = \frac{2}{N} \sum_{i=1}^{N} \cos(jz_i) \sin(kz_i). \tag{27}$$

For each j and for all k, these sums can be computed using a FFT. Furthermore, these first alternate coefficients, $c_{j,k}^{(1)}$, can be calculated by analyzing the way the exact coefficients, $c_{j,k}^{(0)}$ for $k \geqslant N$, alias back to the range $1 \leqslant k < N$. By computing these coefficients, one can clearly see that $c_{j,k}^{(0)} \neq c_{j,k}^{(1)}$ for any range of j and/or k. For the interested reader, this calculation can be found in Appendix C.

Finally, we can also compute the coefficients by multiplying $\cos(jz)$ by $\mathrm{Id}_m(z)$ in grid space and then take the discrete sine transform of the result, i.e., we apply the parity filtering method. That is, we compute

$$c^{(2)}_{j,k} = \frac{2}{N}\sum_{i=1}^{N} \cos(jz_i)\mathrm{Id}_m(z_i)\sin(kz_i), \tag{28}$$

where we compute the coefficients of each j and for all k using a FFT. The theory described earlier in this paper demonstrates that $c^{(2)}_{j,k} = c^{(0)}_{j,k}$ for all $j + k \leqslant N$.

To illustrate the difference between calculating the coefficients, $c_{j,k}$, using Eq. (28) versus Eq. (27), we consider the following error definitions

$$\mathcal{E}^{(1)}_{j,k_{\max}} = \sqrt{\frac{\sum_{k=1}^{k_{\max}}\left(c^{(1)}_{j,k} - c^{(0)}_{j,k}\right)^2}{\sum_{k=1}^{k_{\max}}\left(c^{(0)}_{j,k}\right)^2}}; \quad \mathcal{E}^{(2)}_{j,k_{\max}} = \sqrt{\frac{\sum_{k=1}^{k_{\max}}\left(c^{(2)}_{j,k} - c^{(0)}_{j,k}\right)^2}{\sum_{k=1}^{k_{\max}}\left(c^{(0)}_{j,k}\right)^2}}. \tag{29}$$

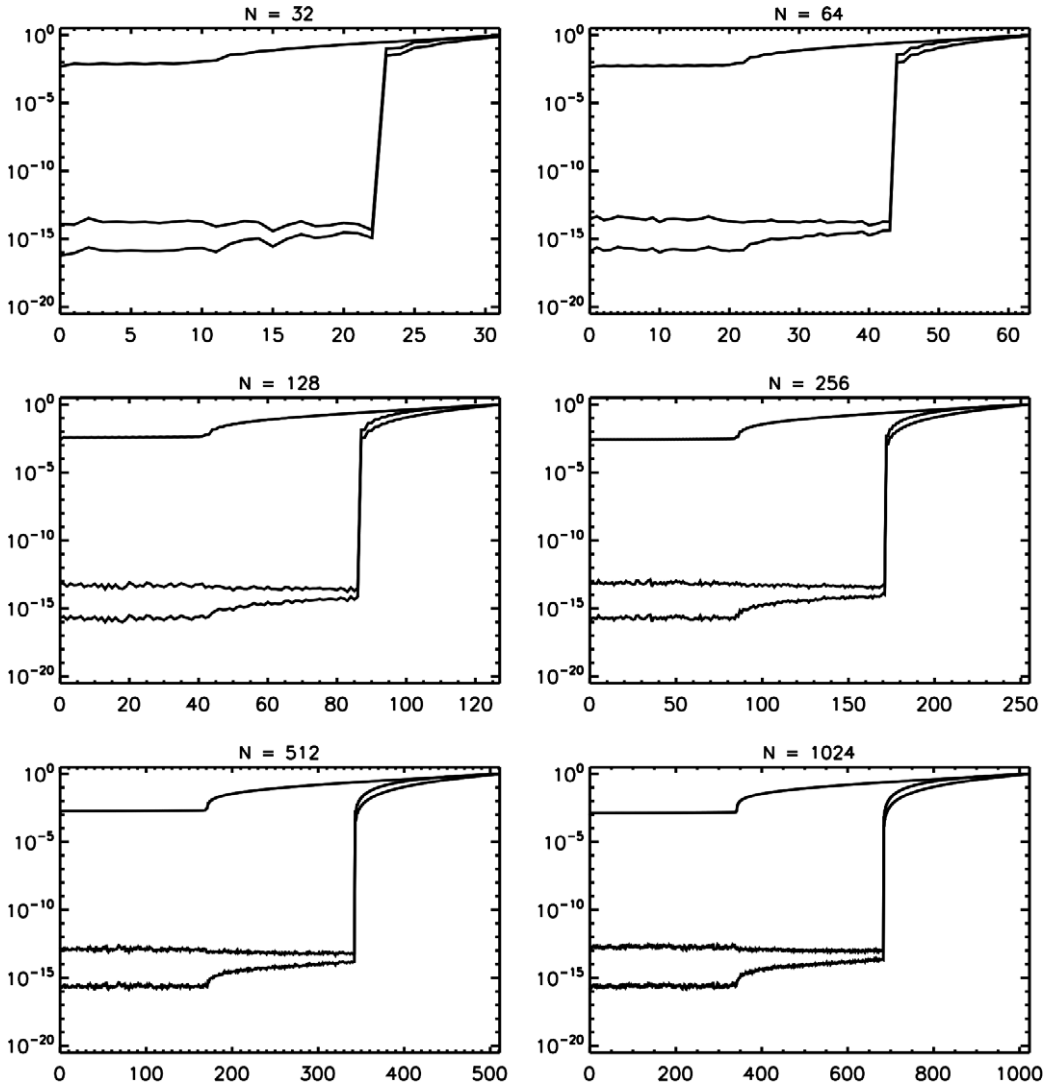

Fig. 1. Plots of $\mathcal{E}^{(1)}_{j,\lfloor N/3\rfloor}$, $\mathcal{E}^{(2)}_{j,\lfloor N/3\rfloor}$, and $\mathcal{E}^{(2,1)}_{j,\lfloor N/3\rfloor}$ for the sine transform of a cosine function for $N \in \{32, 64, 128, 256, 512, 1024\}$ and $j \leqslant N$. In each plot, the top line shows $\mathcal{E}^{(1)}_{j,\lfloor N/3\rfloor}$, the bottom line shows $\mathcal{E}^{(2)}_{j,\lfloor N/3\rfloor}$ and the intermediate line shows the ratios, $\mathcal{E}^{(2,1)}_{j,\lfloor N/3\rfloor}$.

The meaning of these error definitions becomes clear if we define the truncated expansions

$$\mathrm{C}^{(i)}_{j,k_{\max}}(z) = \sum_{k=1}^{k_{\max}} c^{(i)}_{j,k} \sin(kz), \tag{30}$$

for $i \in \{0,1,2\}$. Using the standard L^2 norm on $[0,\pi]$ and Parseval's identity (see [3], p. 37), the definitions in Eqs. (29) become simply

$$\mathcal{E}^{(1)}_{j,k_{\max}} = \frac{\|\mathrm{C}^{(1)}_{j,k_{\max}} - \mathrm{C}^{(0)}_{j,k_{\max}}\|_{L^2}}{\|\mathrm{C}^{(0)}_{j,k_{\max}}\|_{L^2}}, \quad \mathcal{E}^{(2)}_{j,k_{\max}} = \frac{\|\mathrm{C}^{(2)}_{j,k_{\max}} - \mathrm{C}^{(0)}_{j,k_{\max}}\|_{L^2}}{\|\mathrm{C}^{(0)}_{j,k_{\max}}\|_{L^2}}. \tag{31}$$

It is also helpful to define the ratio of these two errors, *viz.*,

$$\mathcal{E}^{(2,1)}_{j,k_{\max}} = \frac{\mathcal{E}^{(2)}_{j,k_{\max}}}{\mathcal{E}^{(1)}_{j,k_{\max}}}. \tag{32}$$

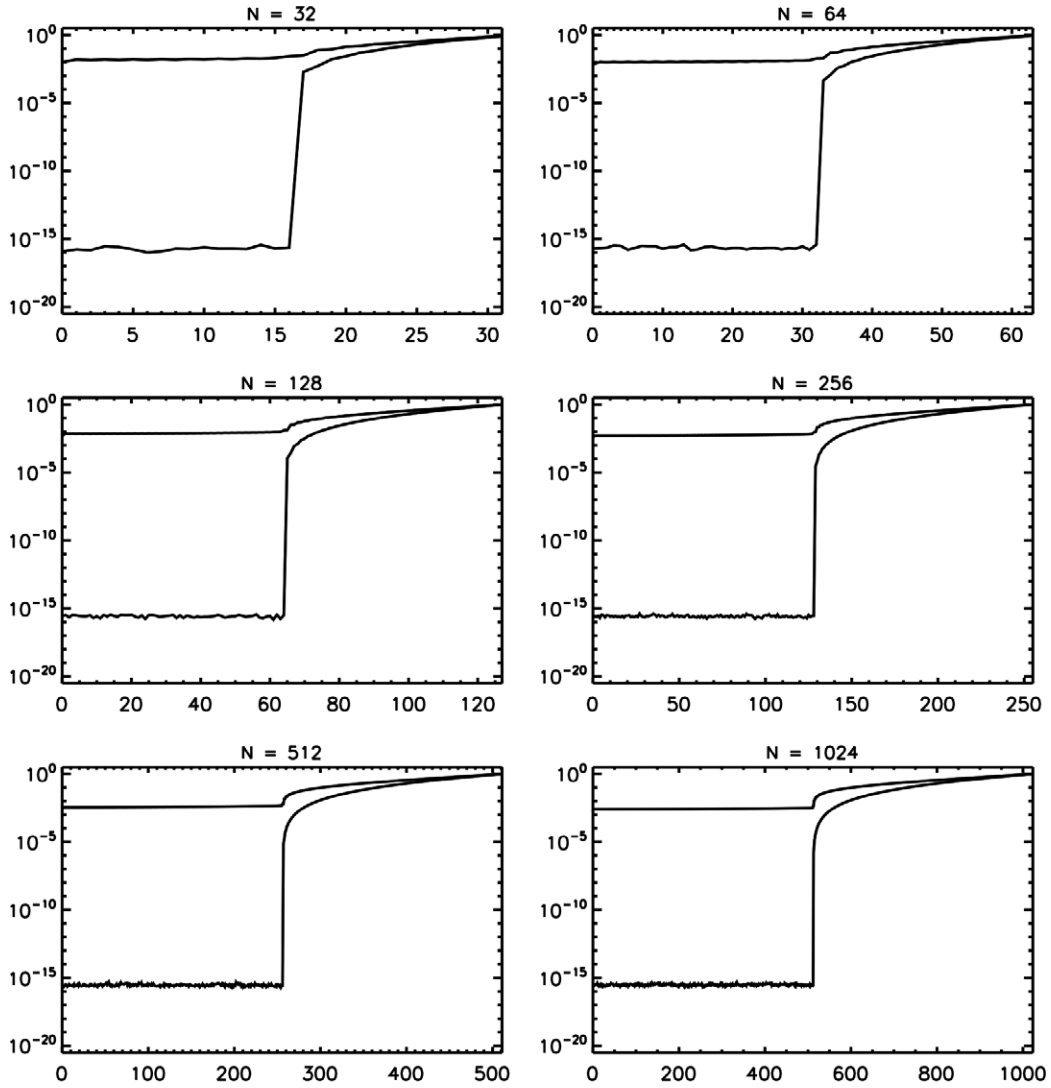

Fig. 2. Plots of $\mathcal{E}^{(1)}_{j,N/2}$, and $\mathcal{E}^{(2)}_{j,N/2}$ for the sine transform of a cosine function for $N \in \{32, 64, 128, 256, 512, 1024\}$ and $j \leqslant N$. In each plot, the top line shows $\mathcal{E}^{(1)}_{j,N/2}$ and the bottom line shows $\mathcal{E}^{(2)}_{j,N/2}$. For clarity, the ratios, $\mathcal{E}^{(2,1)}_{j,N/2}$, are not shown.

In general, Corollary 2.2 asserts that $c^{(2)}_{j,k} = c^{(0)}_{j,k}$ for $j + k_{\max} \leqslant N$, and thus we should expect that $\mathcal{E}^{(2)}_{j,k_{\max}} \simeq 0$ (to numerical precision) in this same range. As a test, the errors, $\mathcal{E}^{(1)}_{j,k_{\max}}$, $\mathcal{E}^{(2)}_{j,k_{\max}}$, and $\mathcal{E}^{(2,1)}_{j,k_{\max}}$, were computed for various values of $k_{\max}$ and N using FFTW (see [10]) with double-precision arithmetic. The coefficients, $c^{(0)}_{j,k}$, were calculated using Eq. (26). The other coefficients, $c^{(1)}_{j,k}$ and $c^{(2)}_{j,k}$, were calculated by first taking the inverse fast cosine transform of the identity matrix, $\delta_{i,j}$, and then feeding the grid values into a fast sine transform [of course multiplying by $\mathrm{Id}_m(z_i)$ for $c^{(2)}_{j,k}$].

Figs. 1–3 show the errors in computing $c^{(2)}_{j,k}$ using a fast version of Eq. (28) for $k_{\max} = \lfloor N/3 \rfloor$, $k_{\max} = N/2$, and $k_{\max} = \lfloor 2N/3 \rfloor$, respectively. For each value of $k_{\max}$, we compute errors for $N \in \{32, 64, 128, 256, 512, 1024\}$. $\mathcal{E}^{(2)}_{j,k_{\max}}$ is zero to within machine-roundoff error $\left(\mathcal{E}^{(2)}_{j,k_{\max}} \simeq 10^{-15} - 10^{-14}\right)$ for all $j \leqslant N - k_{\max}$, and the magnitude of this error only depends extremely weakly on N, if at all.

Likewise, Figs. 4–6 show the errors associated with representing a sine function with a cosine series. In these cases, $\mathcal{E}^{(2)}_{j,k_{\max}}$ is also zero to within machine-roundoff error. However, for small j, the uncorrected error, $\mathcal{E}^{(1)}_{j,k_{\max}}$ is smaller than it is when representing a cosine function with a sine series. This should be expected based on the behavior of sine and cosine functions near boundaries. Since individual cosine functions never vanish at a

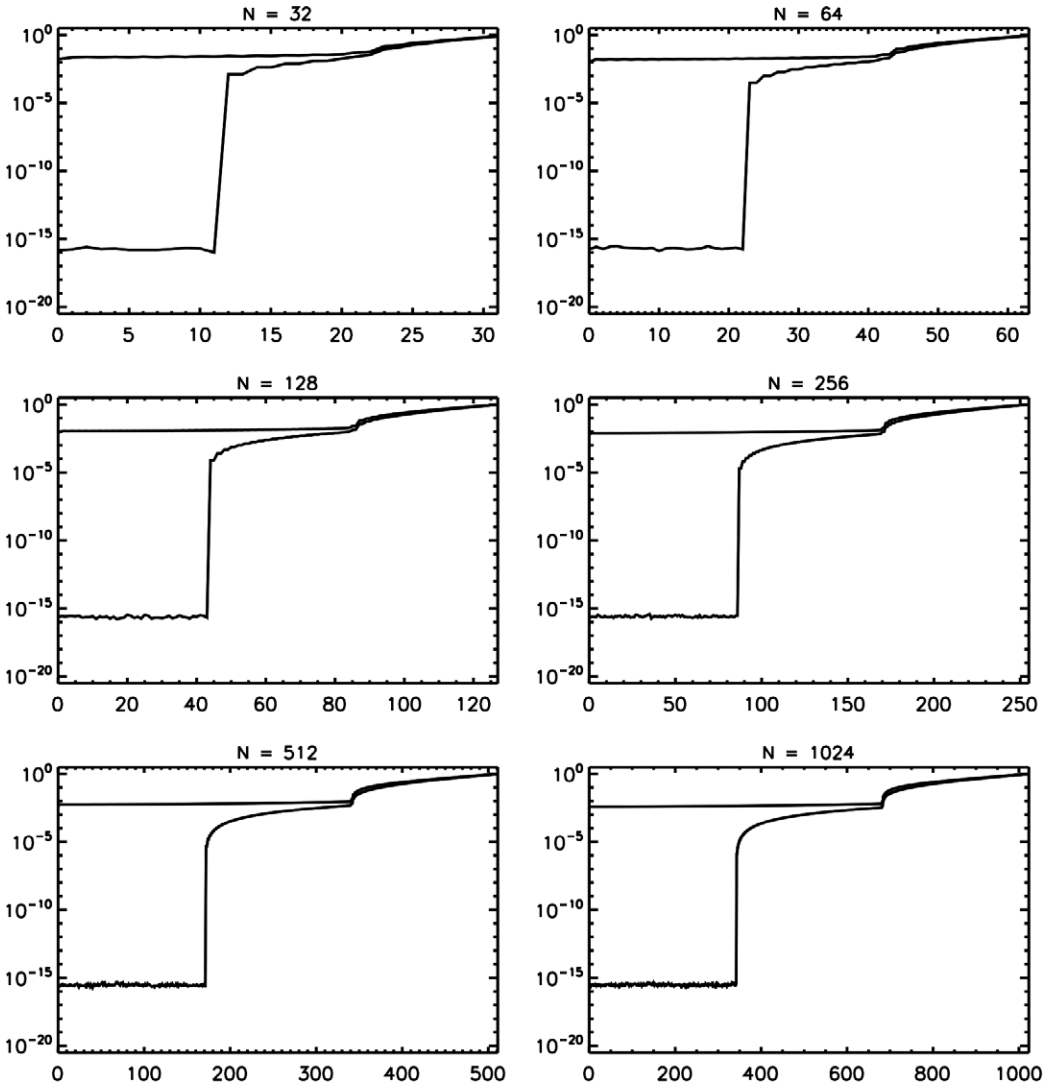

Fig. 3. Plots of $\mathcal{E}^{(1)}_{j,\lfloor 2N/3 \rfloor}$, and $\mathcal{E}^{(2)}_{j,\lfloor 2N/3 \rfloor}$ for the sine transform of a cosine function for $N \in \{32, 64, 128, 256, 512, 1024\}$ and $j \leqslant N$. In each plot, the top line shows $\mathcal{E}^{(1)}_{j,\lfloor 2N/3 \rfloor}$ and the bottom line shows $\mathcal{E}^{(2)}_{j,\lfloor 2N/3 \rfloor}$. For clarity, the ratios, $\mathcal{E}^{(2,1)}_{j,\lfloor 2N/3 \rfloor}$, are not shown.

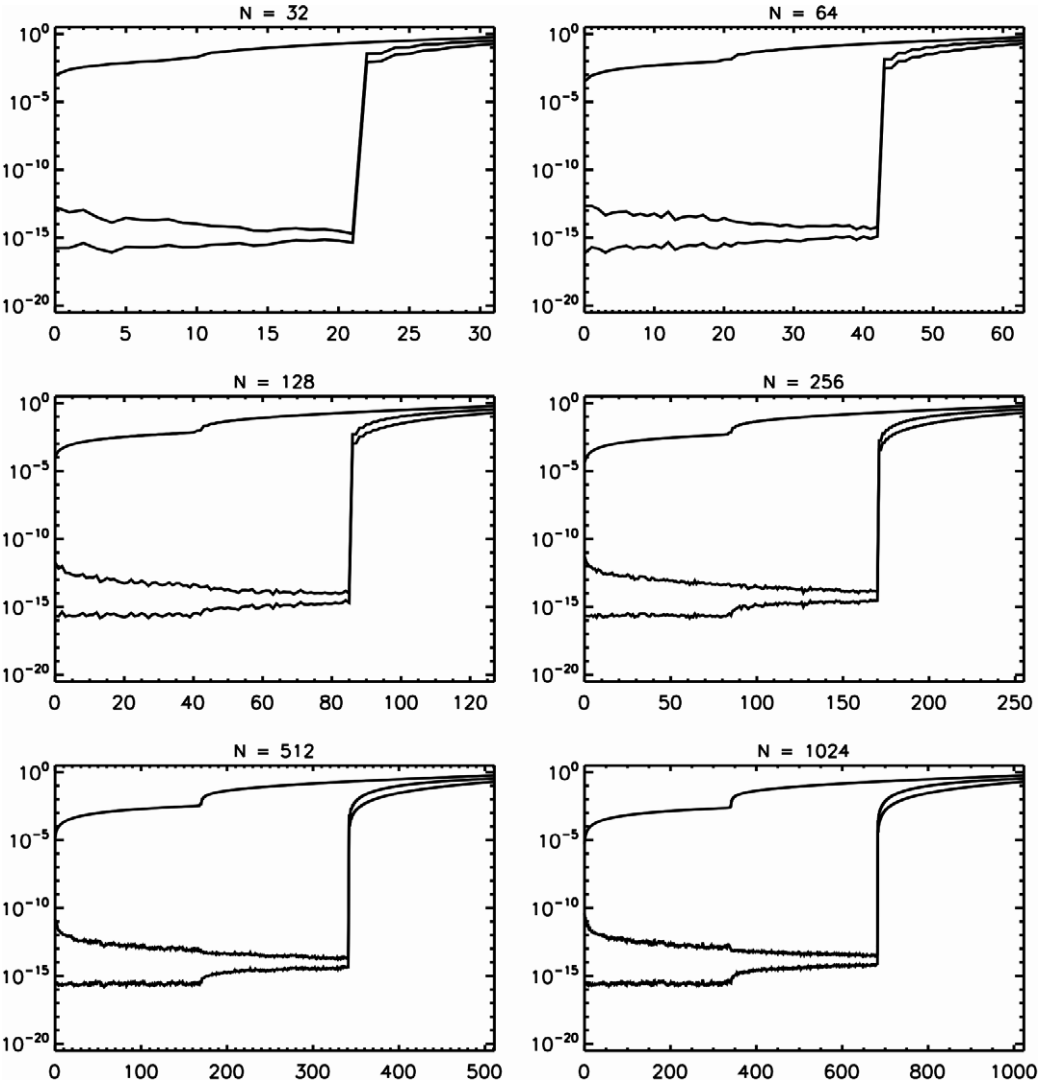

Fig. 4. Plots of $\mathcal{E}^{(1)}_{j,\lfloor N/3\rfloor}$, $\mathcal{E}^{(2)}_{j,\lfloor N/3\rfloor}$, and $\mathcal{E}^{(2,1)}_{j,\lfloor N/3\rfloor}$ for the cosine transform of a sine function for $N \in \{32, 64, 128, 256, 512, 1024\}$ and $j \leqslant N$. In each plot, the top line shows $\mathcal{E}^{(1)}_{j,\lfloor N/3\rfloor}$, the bottom line shows $\mathcal{E}^{(2)}_{j,\lfloor N/3\rfloor}$ and the intermediate line shows the ratios, $\mathcal{E}^{(2,1)}_{j,\lfloor N/3\rfloor}$.

boundary, it is difficult to approximate their behavior by functions that always vanish at a boundary. Obviously, the converse is not true. It is relatively easy to add cosine functions in such a way that they cancel at the boundaries. However, the same cannot be said for their derivatives. Thus, if we were considering parity filters under the H^1 norm on $[0, \pi]$ (as opposed to the L^2 norm), $\mathcal{E}^{(1)}_{j,k_{\max}}$ would be substantially larger when representing sine functions with a cosine series, while $\mathcal{E}^{(2)}_{j,k_{\max}}$ would still be zero to within machine precision.

With regard to all of the errors shown in Figs. 1–6, $k_{\max} = \lfloor 2N/3\rfloor$ corresponds to a value that one would use for the standard 2/3-dealiasing rule for a non-parity-mixed quadratic nonlinearity. However, the spectrum is only error free up to $j_{\max} = \lfloor N/3\rfloor$ which is not sufficient to remain error free up to $k_{\max}$. Nevertheless, $k_{\max} = \lfloor N/3\rfloor$, seen in Fig. 3, corresponds to a value one would use for parity filtering a quadratic nonlinearity (e.g., Example 1.1). In this case, the spectrum going up to $k_{\max}$ would broaden up to $2k_{\max} = \lfloor 2N/3\rfloor$ which is satisfactory since there are no substantial errors up to this value. Finally, $k_{\max} = N/2$ corresponds to a value that one would use for parity filtering a linear term (e.g., Example 1.2 and Part II of this paper). In this case, we have $j_{\max} = k_{\max} = N/2$. This implies that any sine or cosine function up to this degree can be accurately represented on a cosine or sine basis of the same size. Finally, in every case, the following holds

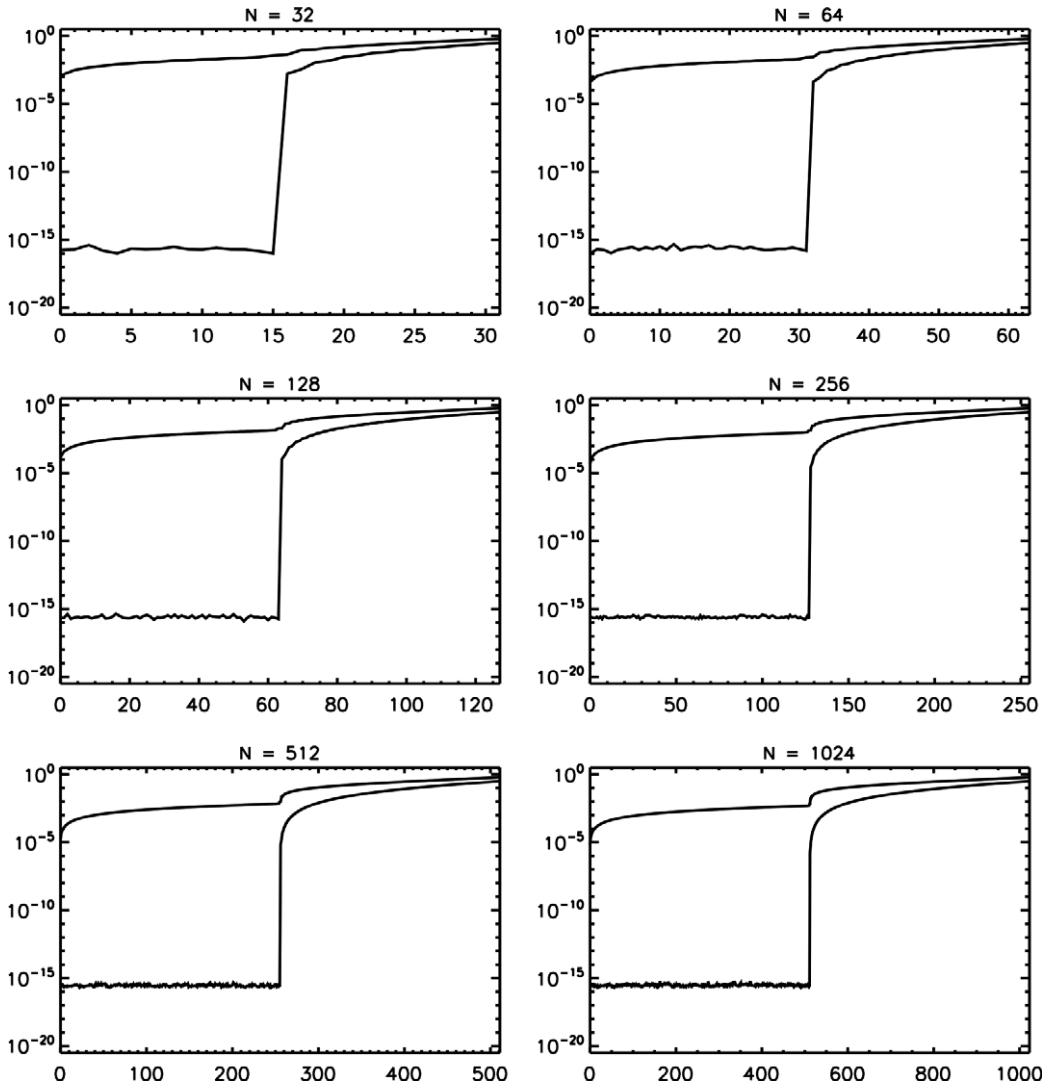

Fig. 5. Plots of $\mathcal{E}^{(1)}_{j,N/2}$, and $\mathcal{E}^{(2)}_{j,N/2}$ for the cosine transform of a sine function for $N \in \{32, 64, 128, 256, 512, 1024\}$ and $j \leqslant N$. In each plot, the top line shows $\mathcal{E}^{(1)}_{j,N/2}$ and the bottom line shows $\mathcal{E}^{(2)}_{j,N/2}$. For clarity, the ratios, $\mathcal{E}^{(2,1)}_{j,N/2}$, are not shown.

$$\mathcal{E}^{(2)}_{j,k_{\max}} < \mathcal{E}^{(1)}_{j,k_{\max}} \quad \text{for all } j \leqslant N, \tag{33}$$

even for j larger than the critical value, i.e., $j > N - k_{\max}$. Therefore, independent of the degree of dealiasing, one can always improve aliasing error by judiciously multiplying by the identity function, $\mathrm{Id}_m(z_i)$, prior to performing phase-mixed discrete transforms.

At this point, we make a final comment regarding the growth of the various parity filtered errors, $\mathcal{E}^{(2)}_{j,k_{\max}}$, as a function of j. While $\mathcal{E}^{(2)}_{j,k_{\max}}$ is formally zero up to some critical $j_{\max} = N - k_{\max}$, machine precision is actually the best that we can expect in reality. After $j_{\max}$, Figs. 1–6 clearly show that the errors climb dramatically until they are $\mathcal{O}(1)$ as $j \simeq N$, (depending on the value of $k_{\max}$). However, machine precision leaves open the possibility of a much wider class of collocation point weights, as opposed to simply $\mathrm{Id}_m(z)$. That is, we could potentially replace the spectral expansion of unity with some other function, $w^{\epsilon}_m(z)$. Then, if we computed the spectral coefficients, $c_{j,k}$, via a third alternative

$$c^{(3)}_{j,k} = \frac{2}{N} \sum_{i=1}^{N} \cos(jz_i) w^{\epsilon}_m(z_i) \sin(kz_i), \tag{34}$$

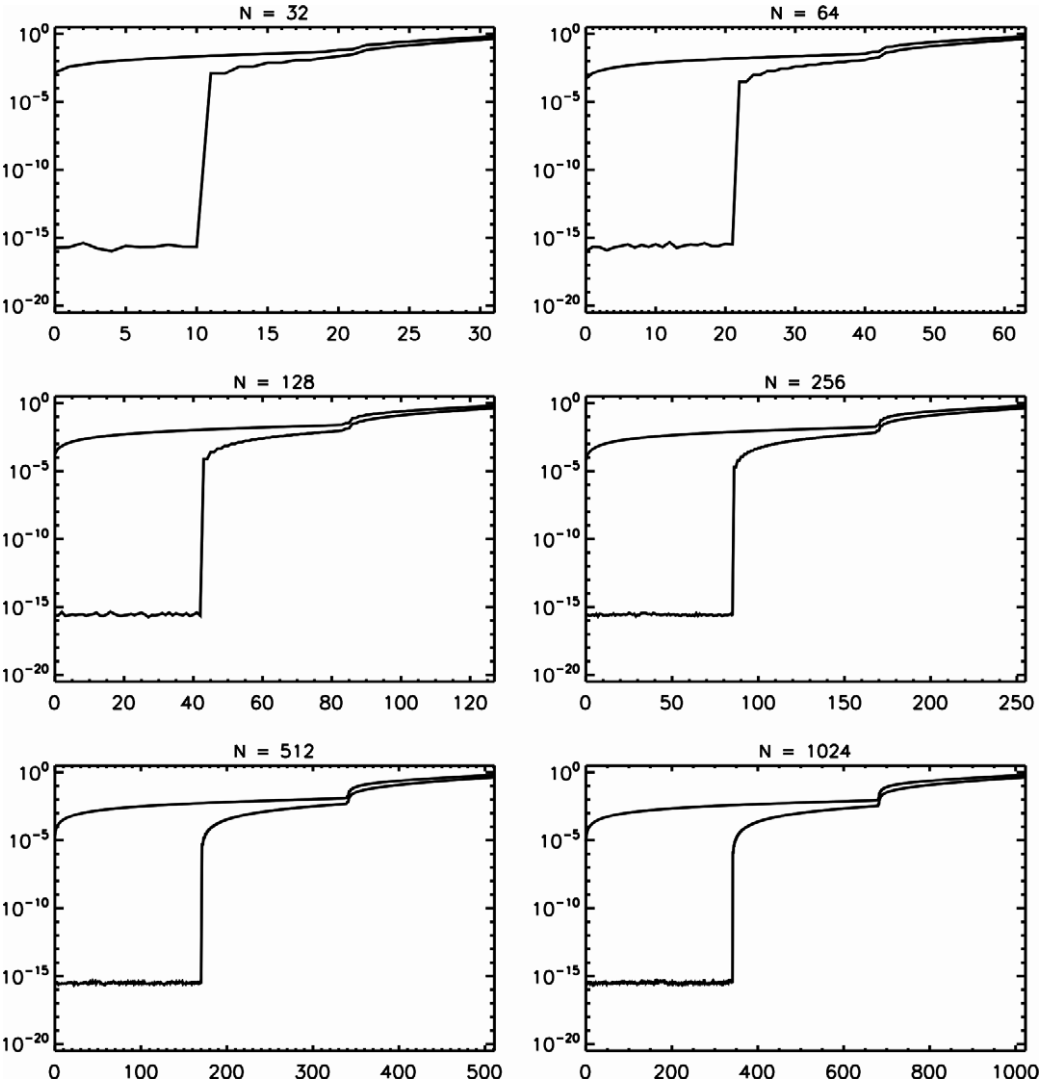

Fig. 6. Plots of $\mathcal{E}^{(1)}_{j,\lfloor 2N/3\rfloor}$, and $\mathcal{E}^{(2)}_{j,\lfloor 2N/3\rfloor}$ for the cosine transform of a sine function for $N \in \{32, 64, 128, 256, 512, 1024\}$ and $j \leqslant N$. In each plot, the top line shows $\mathcal{E}^{(1)}_{j,\lfloor 2N/3\rfloor}$ and the bottom line shows $\mathcal{E}^{(2)}_{j,\lfloor 2N/3\rfloor}$. For clarity, the ratios, $\mathcal{E}^{(2,1)}_{j,\lfloor 2N/3\rfloor}$, are not shown.

then we could require that the associated error was formally bounded by a small parameter of our choosing, i.e.,

$$\mathcal{E}^{(3)}_{j,k_{\max}} \leqslant \epsilon, \tag{35}$$

for $j \leqslant N - k_{\max}$. This type of idea has been successfully carried out on the approximation of functions; see [11]. A finite (but perhaps machine precision) value for ϵ leaves a potentially large class of weights, $w^{\epsilon}_{m}(z)$. The possible advantage to this is we may be able to find a class of weights that gives, practically speaking, no significant error ($\epsilon \simeq 10^{-15}$) for $j \leqslant N - k_{\max}$, but where the error grows much more slowly for $j > N - k_{\max}$ than in Figs. 1–6.

5. Conclusion

The above theory and error tests (Figs. 1–6) demonstrate that we have developed a stable procedure for computing the projection (up to a given degree) of any finite combination of trigonometric functions in either

the Fourier sine series or the Fourier cosine series basis sets. The implementation merely involves the multiplication of Id_m into the relevant terms. The efficiency of this procedure is inherent in the speed of the fast Fourier transform. There are many highly optimized FFT libraries available for virtually all hardware platforms, and there are many numerical codes that employ these libraries to solve a variety of physical problems in periodic domains. We believe that this new and easily implemented method allows for a wider range of use of many of these codes. They should be able to solve problems in a wider range of geometries with only a relatively modest amount of modification.

In Part II of this paper, we implement parity filtering techniques in a numerical code designed to solve for rapidly rotating and high-Rayleigh-number convection in a confined box. In doing so, we find that there exist dramatic errors and physically spurious solutions if parity mixing is not properly accounted for. However, when parity mixing is properly accounted for, we find physically consistent solutions with novel dynamics.

Acknowledgment

GMV and NHB were supported by NASA Sun-Earth Connections Division Grant No. NNG04GB86G. KJ was supported by NASA award NNG05GD37G, and University of Colorado SEED Grant. We would also like to thank Lucas Monzón and Greg Beylkin for helpful comments and pointing out alternative interpretations of various formulae. The authors also gratefully acknowledge an anonymous referee for very helpful comments.

Appendix A. Convergence and boundedness of Id_m

Proof. (Claim 2.3) We consider the modified series

$$\mathrm{Id}_m(z;\xi) = \frac{4}{\pi}\sum_{n=1}^{m} \xi^{2n-1}\frac{\sin((2n-1)z)}{2n-1}. \tag{A.1}$$

Series (A.1) converges geometrically for all $|\,\xi\,| < 1$. Therefore, we take the limit as $m \to \infty$ and consider

$$\mathrm{Id}(z;\xi) = \frac{4}{\pi}\sum_{n=1}^{\infty} \xi^{2n-1}\frac{\sin((2n-1)z)}{2n-1}. \tag{A.2}$$

By virtue that for any complex number, ζ, with $|\,\zeta\,| < 1$

$$\operatorname{arctanh}(\zeta) = \sum_{n=1}^{\infty}\frac{\zeta^{2n-1}}{2n-1}, \tag{A.3}$$

we find (after some standard manipulation) that

$$\mathrm{Id}(z;\xi) = \frac{2}{\pi}\arctan\left(\frac{2\xi\sin(z)}{1-\xi^2}\right). \tag{A.4}$$

Taking the limit as $\xi \to 1$ from the left, we obtain $\lim_{\xi\to 1^-}\mathrm{Id}(z;\xi) = 1$ for all $0 < z < \pi$. □

Lemma 5.1. *For all $m \geqslant 1$ we have the bound*

$$\frac{\pi^2}{8} - \sum_{n=1}^{m}\frac{1}{(2n-1)^2} \leqslant \frac{1}{4m}. \tag{A.5}$$

Proof. (Lemma 5.1) It is a classical result that

$$\sum_{n=1}^{\infty}\frac{1}{(2n-1)^2} = \frac{3\zeta(2)}{4} = \frac{\pi^2}{8}, \tag{A.6}$$

where $\zeta(s)$ is the Riemann ζ-function (see [12], Section 23.2). Therefore, estimate the difference

$$\frac{\pi^2}{8} - \sum_{n=1}^{m} \frac{1}{(2n-1)^2} = \sum_{n=m}^{\infty} \frac{1}{(2n+1)^2}. \tag{A.7}$$

To do this, consider the telescoping sum

$$\sum_{n=m}^{\infty} \frac{1}{4n(n+1)} = \sum_{n=m}^{\infty} \int_{n-\frac{1}{2}}^{n+\frac{1}{2}} \frac{\mathrm{d}t}{(2t+1)^2} = \int_{m-\frac{1}{2}}^{\infty} \frac{\mathrm{d}t}{(2t+1)^2} = \frac{1}{4m}, \tag{A.8}$$

and

$$\frac{1}{4n(n+1)} = \frac{1}{(2n+1)^2} + \frac{1}{4n(n+1)(2n+1)^2}. \tag{A.9}$$

Since $4n(n+1)(2n+1)^2 > 0$, we have Ineq. (A.5) if we substitute Eq. (A.9) into the left-hand side of Eq. (A.8). □

Proof. (Claim 2.4) Consider

$$(\mathrm{Id}_m(z) - 1)^2 = 1 + \int_0^z (\mathrm{Id}_m(z_0) - 1)\mathrm{Id}_m'(z_0)\,\mathrm{d}z_0 - \int_z^{\pi} (\mathrm{Id}_m(z_0) - 1)\mathrm{Id}_m'(z_0)\,\mathrm{d}z_0. \tag{A.10}$$

Expression (A.10) is valid by virtue that $\mathrm{Id}_m(0) = \mathrm{Id}_m(\pi) = 0$ and that for any finite value of $m \geqslant 1$, the derivative of $\mathrm{Id}_m(z)$ is a well-defined trigonometric polynomial

$$\mathrm{Id}_m'(z) = \frac{4}{\pi} \sum_{n=1}^{m} \cos((2n-1)z). \tag{A.11}$$

We use the triangle inequality to obtain the estimate

$$(\mathrm{Id}_m(z) - 1)^2 \leqslant 1 + \int_0^{\pi} |(\mathrm{Id}_m(z_0) - 1)\mathrm{Id}_m'(z_0)|\,\mathrm{d}z_0, \tag{A.12}$$

and then use the Schwartz inequality to obtain

$$(\mathrm{Id}_m(z) - 1)^2 \leqslant 1 + \left(\int_0^{\pi} |\mathrm{Id}_m(z_0) - 1|^2\,\mathrm{d}z_0\right)^{1/2} \times \left(\int_0^{\pi} |\mathrm{Id}_m'(z_0)|^2\,\mathrm{d}z_0\right)^{1/2}. \tag{A.13}$$

The integrals on the right-hand side of Ineq. (A.13) can be computed exactly because of the orthogonality of trigonometric functions. That is

$$\int_0^{\pi} |\mathrm{Id}_m(z_0) - 1|^2\,\mathrm{d}z_0 = \frac{8}{\pi}\left(\frac{\pi^2}{8} - \sum_{n=1}^{m} \frac{\pi}{(2n-1)^2}\right) \tag{A.14}$$

$$\int_0^{\pi} |\mathrm{Id}_m'(z_0)|^2\,\mathrm{d}z_0 = \frac{8m}{\pi}. \tag{A.15}$$

From Ineq. (A.5), we now have

$$(\mathrm{Id}_m(z) - 1)^2 \leqslant 1 + \frac{4}{\pi}, \tag{A.16}$$

and we can take $I_0 = 1 + (1 + 4/\pi)^{1/2} \simeq 2.51$. □

Note that although Ineq. (A.16) may not be the best possible bound on the magnitude of $\mathrm{Id}_m(z)$, it suffices for the purposes of Section 2; the best possible uniform bound is given by $0 \leqslant \mathrm{Id}_m(z) \leqslant 4/\pi$.

Appendix B. Non-aliasing for odd spectrum sizes

To solve the aliasing problem when n is odd, we must examine how the high wavenumber power is transferred by multiplication and eventually aliased. Since the aliasing problem only occurs because of the highest

wavenumbers, we can examine what happens at these scales alone and determine how to fix the aliasing problem. Therefore, without loss of generality, assume that

$$f(x) = g(x) = \sin(nz), \tag{B.1}$$

and

$$\mathrm{Id}_m(z) = \mathrm{Id}_{m-1}(z) + \mathrm{e}_{3n}\sin(3nz). \tag{B.2}$$

From Eq. (16), we should use $\mathrm{e}_{3n} = 4/3n\pi$. However, this is not the best choice in the current situation. Consider

$$\begin{aligned} f(z)g(z)\mathrm{Id}_m(z) &= \sin(nz)^2\mathrm{Id}_{m-1}(z) + \mathrm{e}_{3n}\sin(3nz)\sin(nz)^2 \\ &= \sin(nz)^2\mathrm{Id}_{m-1}(z) - \mathrm{e}_{3n}\frac{\sin(nz)}{4} + \mathrm{e}_{3n}\frac{\sin(3nz)}{4} - \mathrm{e}_{3n}\frac{\sin(5nz)}{4}. \end{aligned} \tag{B.3}$$

After the application of a discrete sine transform, the $\sin(nz)^2\mathrm{Id}_{m-1}(z)$ term in this expression comes out correctly up to order n since there is no power in this term greater than order $2n + 2(m-1) - 1 = 5n - 3$. If we are using a transform size of order $3n$, recall that only terms with a degree greater than or equal to $5n$ can potentially contaminate our spectrum at or below order n. Also, the $\sin(3nz)$ term is computed correctly and then deleted since it is greater than order n. The $\sin(nz)$ term is computed correctly and kept since it is in the range of interest. However, the $\sin(5nz)$ term is aliased to order n and thus is the potential source of error. Therefore, after performing a discrete sine transform on $f(z)g(z)\mathrm{Id}_m(z)$, we obtain a result as though we had performed an exact sine transform on

$$f(z)g(z)\mathrm{Id}_m(z) \to \sin(nz)^2\mathrm{Id}_{m-1}(z) - 2\mathrm{e}_{3n}\frac{\sin(nz)}{4}, \tag{B.4}$$

rather than the correct version

$$f(z)g(z)\mathrm{Id}_m(z) \to \sin(nz)^2\mathrm{Id}_{m-1}(z) - \mathrm{e}_{3n}\frac{\sin(nz)}{4}. \tag{B.5}$$

Notice that because of aliasing the $\mathrm{e}_{3n}\sin(3nz)$ term on the right-hand side of (B.4) plays double the role that it would if it were computed on a grid that was one point larger (i.e., if n were even). However, by taking the value of $\mathrm{e}_{3n} = 2/3n\pi$, rather than $4/3n\pi$, we can let the order $5n$ aliasing occur with impunity. This turns out to be quite serendipitous. If we were to tabulate the values of $\mathrm{Id}_m(z)$ on the grid $z_i = \pi(i/N - 1/2N)$ for $i \in \{1, \ldots, N\}$ and then supply these values to the discrete sine transform

$$\mathrm{e}_k = \frac{2}{N}\sum_{i=1}^{N}\mathrm{Id}_m(z_i)\sin(kz_i), \tag{B.6}$$

the last value, e_{3n}, comes out two times as large as it should, just as occurs with the first value of a discrete cosine transform. Therefore, if we define e_k to be the kth element of the spectrum

$$\left\{\frac{4}{\pi}, 0, \frac{4}{3\pi}, \ldots, 0, \frac{4}{3n\pi}\right\}, \tag{B.7}$$

then the values of $\mathrm{Id}_m(z)$ will come out conveniently so that we can use it to compute the projection correctly of $f(z)g(z)$ on a sine basis up to order n using an FFST on $N = 3n$ points, when n is an odd integer. Likewise, we can use the spectrum

$$\left\{\frac{4}{\pi}, 0, \frac{4}{3\pi}, \ldots, 0, \frac{4}{(3n-1)\pi}, 0\right\}, \tag{B.8}$$

when n is an even integer to compute the projection using an FFST on $N = 3n$ points as well.

Appendix C. Computation of aliased coefficients

In this appendix, we show analytically how to compute the coefficients, $c^{(1)}_{j,k}$ from Section 4. These are the coefficients that result from naively feeding a tabulated sine function (in grid space) into a fast cosine transform. The result produces aliasing and hence, we do not obtain the correct result as required.

For the analytical computation of $c^{(1)}_{j,k}$ from Section 4, recall that the basic aliasing rule is $N + k \to N - k$. Therefore, it follows that

$$2N - k = N + (N - k) \to N - (N - k) = k \tag{C.1}$$

$$2N + k = N + (N + k) \to N - (N + k) = -k \tag{C.2}$$

$$4N - k = N + (3N - k) \to N - (3N - k) = -(2N - k) \tag{C.3}$$

$$4N + k = N + (3N - k) \to N - (3N - k) = -(2N + k). \tag{C.4}$$

From these relations, we can recursively produce the aliasing rule for any positive integer. Therefore, for any integer, $p \geqslant 1$, we have the aliasing rules

$$(4p - 2)N \pm k \to \mp k \tag{C.5}$$

$$4pN \pm k \to \pm k. \tag{C.6}$$

The rules in (C.5) and (C.6) are sufficient to cover any positive integer by using different combinations of p and k. The negative signs on the right-hand sides of (C.5) and (C.6) simply denote $c^{(0)}_{j,-k} = -c^{(0)}_{j,k}$. Therefore, we have

$$c^{(1)}_{j,k} = c^{(0)}_{j,k} + \sum_{p=1}^{\infty} \left(c^{(0)}_{j,(4p-2)N-k} - c^{(0)}_{j,(4p-2)N+k} + c^{(0)}_{j,4pN+k} - c^{(0)}_{j,4pN-k} \right). \tag{C.7}$$

The computation of the series in Eq. (C.7) is awkward, but a straightforward task if we use the identity (see [12], Section 4.3)

$$\pi \cot(x) = \frac{1}{x} + \sum_{p=1}^{\infty} \left(\frac{1}{x - p} + \frac{1}{x + p} \right). \tag{C.8}$$

With this, we obtain

$$c^{(1)}_{j,k} = \begin{cases} \frac{4\cos\left(\frac{j\pi}{2N}\right)\sin\left(\frac{k\pi}{2N}\right)}{N\left(\cos\left(\frac{j\pi}{N}\right) - \cos\left(\frac{k\pi}{N}\right)\right)} & \text{if } j + k \text{ is odd} \\ 0 & \text{if } j + k \text{ is even.} \end{cases} \tag{C.9}$$

These coefficients show clearly that aliasing causes $c^{(1)}_{j,k} \neq c^{(0)}_{j,k}$ (compare Eqs. (C.9) and (26)). While some special values of j and k in the range $j + k \leqslant N$ do provide only a small difference between $c^{(1)}_{j,k}$ and $c^{(0)}_{j,k}$, for high wavenumbers, *i.e*, $j, k \sim N$, it turns out that the difference between the aliased coefficients and the correct coefficients is of order unity. This large error in the high range of the spectrum can have dramatic consequences in certain applications, in particular, in the numerical solution of time-dependent PDE's as discussed in Examples 1 and 2.

References

[1] J. Boyd, Chebyshev and Fourier Spectral Methods, second ed. (Revised)., Dover Publications, New York, 2001.
[2] D. Gottlieb, S. Orszag, Numerical analysis of spectral methods: theory and applications, SIAM, Philadelphia, 1977.
[3] Y. Katznelson, An Introduction to Harmonic Analysis, Cambridge University Press, Cambridge, 2004.
[4] S. Orszag, On the elimination of aliasing errors in finite-difference schemes by filtering high-wavenumber components, J. Atmos. Sci. 28 (6) (1971) 1074.
[5] G. Vasil, N. Brummell, K. Julien, A new method for fast transforms in parity-mixed PDEs: Part II. Application to confined rotating convection, J. Comp. Phys. 227 (17) (2008) 8017–8034.
[6] E.A. Spiegel, G. Veronis, On the Boussinesq approximation for a compressible fluid, Astrophys. J. 131 (2) (1960) 442–447.
[7] H. Greenspan, The Theory of Rotating Fluids, Cambridge, London, 1969.
[8] W. Press, S. Teukolsky, W. Vetterling, B. Flannery, Numerical Recipes in FORTRAN: The Art of Scientific Computing, second ed, Cambridge, 1992.
[9] E. Lieb, M. Loss, Analysis, American Mathematical Society Providence, 1997.
[10] M. Frigo, S. Johnson, The design and implementation of FFTW3, Proceedings of the IEEE: Special Issue on Program Generation, Optimization, and Platform Adaption 92 (2) (2005) 216–231.
[11] G. Beylkin, L. Monzón, On approximation of functions by exponential sums, Appl. Comput. Harmon. Anal. 19 (1) (2005) 17–48.
[12] M. Abramowitz, I. Stegun, Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables, Dover Publications, New York, 1972.