GJI Geodesy, potential field and applied geophysics

Spherical Slepian functions and the polar gap in geodesy

Frederik J. Simons[1] and F. A. Dahlen[2]

[1]*Department of Earth Sciences, University College London, Gower Street, London* WC1E 6BT, *UK. E-mail: fjsimons@alum.mit.edu*
[2]*Department of Geosciences, Princeton University, Guyot Hall, Princeton, NJ* 08544, *USA*

SUMMARY
The estimation of potential fields such as the gravitational or magnetic potential at the surface of a spherical planet from noisy observations taken at an altitude over an incomplete portion of the globe is a classic example of an ill-posed inverse problem. We show that this potential-field estimation problem has deep-seated connections to Slepian's spatiospectral localization problem which seeks bandlimited spherical functions whose energy is optimally concentrated in some closed portion of the unit sphere. This allows us to formulate an alternative solution to the traditional damped least-squares spherical harmonic approach in geodesy, whereby the source field is now expanded in a truncated Slepian function basis set. We discuss the relative performance of both methods with regard to standard statistical measures such as bias, variance and mean squared error, and pay special attention to the algorithmic efficiency of computing the Slepian functions on the region complementary to the axisymmetric polar gap characteristic of satellite surveys. The ease, speed, and accuracy of our method make the use of spherical Slepian functions in earth and planetary geodesy practical.

Key words: geodesy, inverse theory, satellite geodesy, spectral analysis, spherical harmonics, statistical methods.

1 INTRODUCTION

Satellites mapping out the spatial variations of the gravitational or magnetic fields of the Earth or other planets ideally fly on polar orbits, uniformly covering the entire globe. Thus potential fields on the sphere are usually expressed in spherical harmonics, basis functions with global support. For various, among others, engineering, reasons, however, inclined orbits are favourable. These leave a 'polar gap': an antipodal pair of axisymmetric polar caps without any data coverage, typically smaller than $10°$ in diameter for terrestrial gravitational problems (e.g. Lemoine *et al.* 1998), but $20°$ or more in some planetary magnetic configurations (e.g. Santo *et al.* 2001). Furthermore, in geomagnetism, it is sometimes advantageous to exclude data from locations closer than $30°$ to either pole due to their being inherently far noisier than more equatorial data, including much noise that is systematic rather than random. The estimation of spherical harmonic field coefficients from an incompletely sampled sphere is prone to error, since the spherical harmonics are not orthogonal over the partial domain of the cut sphere.

Constructing local spherical harmonic bases that are orthogonal over limited domains and still behave well under the action of upward and downward continuation operators has become a major research goal in geomagnetism (e.g. Thébault *et al.* 2006; Lesur 2006), but in geodesy the polar gap problem has historically been somewhat neglected. It was revived by, among others, Sneeuw & van Gelderen (1997), and recently, Albertella *et al.* (1999), who constructed a new basis of so-called Slepian functions (after Slepian 1983) on the

sphere. These bandlimited functions are designed to have the majority of their energy optimally concentrated inside the latitudinal belt composed of the entire globe minus the polar gap, that is, the region covered by satellites. Slepian functions are orthogonal on both the entire as well as the cut sphere, a property that can be exploited to our advantage. Here, we study the inverse problem of retrieving a potential field on the unit sphere from noisy and incomplete but continuously available observations made at an altitude above their source. We derive exact expressions for the estimation error due to the traditional method of damped least-squares spherical harmonic analysis as well as that arising from a new approach using a truncated set of Slepian basis functions.

We cast the geodetic estimation problem in the much wider context of spatiospectral localization, whereby bandlimited functions are spatially concentrated to regions of arbitrary shape on the sphere (Wieczorek & Simons 2005; Simons *et al.* 2006), and derive a new semi-analytical numerical method to calculate the spherical Slepian functions on a latitudinal belt symmetric about the equator, or its complement, the double polar cap. This approach requires no numerical integration and avoids the construction of matrices other than a tridiagonal matrix whose elements are prescribed analytically. Finding spherical harmonic expressions for bandlimited functions concentrated to polar caps or latitudinal belts, as in Fig. 1, thus becomes so effortless as to be achievable by a handful of lines of computer code, and the problems with numerical stability that are known to plague alternative approaches (Albertella *et al.* 1999; Pail *et al.* 2001) are avoided altogether.
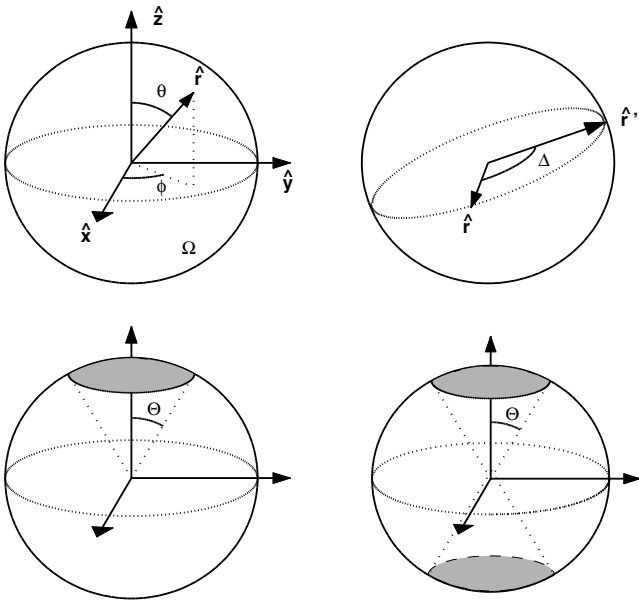
Figure 1. Geometry of the geodetic estimation problem, and some symbols used in this paper. In the lower left, an axisymmetric polar cap, shaded, of colatitudinal radius Θ. In the lower right, an antipodal pair of polar caps, shaded, representing the geodetic polar gap.

The key to this 'magic' lay hidden in two little-known studies published several decades ago: the work by Gilbert & Slepian (1977) on doubly orthogonal polynomials, and that on commuting differential operators by Grünbaum *et al.* (1982). It must be remembered that one of Slepian's main discoveries (see, e.g. Slepian 1983) was the existence of a second-order differential operator that commutes with the spatiospectral localization kernel concentrating to intervals on the real line. Cast in matrix form, finding the prolate spheroidal functions amounts to the diagonalization of a simple tridiagonal matrix (see, e.g. Percival & Walden 1993). In their study, Gilbert & Slepian (1977) presented two additional commuting differential operators, which are applicable to the concentration of Legendre polynomials to one- and two-sided domains. Grünbaum *et al.* (1982) proved that the matrix accompanying the localization to the single polar cap is, once again, tridiagonal. Here, we show this is also the case for the antipodal double polar cap and its complement, the latitudinal belt. The tridiagonal matrix elements coding for the single polar cap, and their solutions, were published by us elsewhere (Simons *et al.* 2006). The expressions applicable to the double polar cap appear here for the first time.

In practice, the geodetic and geomagnetic inverse problems are always ill-conditioned, even in the absence of a polar gap, due the peculiarities of orbital data coverage and the distribution of noise sources (Xu 1992a,b; Holme & Bloxham 1995, 1996). In the standard method of damped spherical harmonic inversion, the ill-conditioning is alleviated by the addition of a small damping parameter to the normal equation matrix prior to inversion. Often the value of this parameter is ad hoc and chosen primarily for numerical stability, but more sophisticated methods use a priori statistical information about the set of model parameters. In this paper we derive the exact structure of the model parameter sensitivity matrix arising from the presence of a contiguous data gap, assuming the data are known continuously everywhere but inside it. The Slepian functions are revealed to be the very eigenfunctions of this matrix. Assuming a particular covariance structure for the model parame-

ters and the observational noise, this knowledge allows us to write analytical expressions for the optimal regularization terms for the damped spherical harmonic method. Such an approach optimally filters out the small eigenvalues, and thus reduces the ill-conditioning of the sensitivity matrix (Jordan & Minster 1972; Wiggins 1972; Aki & Richards 1980; Wingham 1992; Mallat 1998; Xu 1998). Our second method applies a hard truncation to the singular values of the sensitivity matrix in an approach based directly on the Slepian expansion of the model. We show that this is only marginally less successful in minimizing the mean squared estimation error, as well as being computationally advantageous and more intuitively appealing.

The problems posed and solved in this paper are not limited to geodesy and observations made from a satellite. In geomagnetism, our observation level may be the Earth's surface, and the source level at or near the core–mantle boundary (Lowes 1974; Gubbins 1983). In cosmology, the unit sphere constituting the sky is observed from the inside out, and the galactic plane masking spacecraft measurements has the shape of a latitudinal belt (Tegmark 1996; Hinshaw *et al.* 2003). Ground-based astronomical measurements may be confined to a small circular patch of the sky (Peebles 1973; Tegmark 1995). Finally, in planetary science, knowledge of the estimation statistics of properties observed over mere portions of the planetary surface is important in the absence of groundtruthing observations.

2 STATEMENT OF THE PROBLEM

We are concerned with estimating source-level potential fields from noise-contaminated satellite observations at an altitude over an incomplete portion of the unit sphere. The geometry of this problem is illustrated in Fig. 1. The unit sphere Ω on which the unknown signal is defined is parametrized as usual in terms of spherical coordinates, colatitude θ and longitude ϕ. The angular distance between two position coordinates $\hat{\mathbf{r}} = (\theta, \phi)$ and $\hat{\mathbf{r}}' = (\theta', \phi')$ is denoted by Δ. In the lower right, the domain over which satellite observations are available is left unshaded, whereas the area in which measurements are missing is shaded grey. We denote the white region covered by satellite tracks by R, and the shaded, uncovered region by $\bar{R}$. Although our treatment will start out quite general, without restrictions on the shape of R or $\bar{R}$, as long as they are complementary closed regions on the surface of the unit sphere, the lower right panel of Fig. 1 illustrates the case in which the region $\bar{R}$ is a double polar cap symmetric about the polar axis $\hat{\mathbf{z}}$. The angular radius of the polar caps is denoted by Θ. The double polar cap is representative of the geodetic case in which $\bar{R}$ is the so-called polar gap of missing observations; its complement R is a latitudinal belt of angular width $\pi - 2\Theta$ around the equator, as shown. In the following, for brevity, we will shorten all double summations to a notation requiring only a single sum:

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l} \to \sum_{lm}^{\infty}, \quad \sum_{l=0}^{L}\sum_{m=-l}^{l} \to \sum_{lm}^{L}, \quad \sum_{l>L}^{\infty}\sum_{m=-l}^{l} \to \sum_{lm>L}^{\infty}.$$

2.1 Preliminary considerations on the source signal

We model the geophysical signal $s(\hat{\mathbf{r}})$ on the surface of the unit sphere $\Omega = (\theta, \phi)$ as a broadband, square-integrable, real-valued function defined by the transform pair

$$s(\hat{\mathbf{r}}) = \sum_{lm}^{\infty} s_{lm} Y_{lm}(\hat{\mathbf{r}}), \qquad s_{lm} = \int_{\Omega} s(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}}) \, d\Omega. \tag{1}$$

The integers l and m are the degree and order of the real spherical harmonics $Y_{lm}(\hat{\mathbf{r}})$. These are defined by

$$Y_{lm}(\theta, \phi) = \begin{cases} \sqrt{2} X_{lm}(\theta) \cos m\phi & \text{if} \quad -l \le m < 0 \\ X_{l0}(\theta) & \text{if} \quad m = 0 \\ \sqrt{2} X_{lm}(\theta) \sin m\phi & \text{if} \quad 0 < m \le l, \end{cases} \tag{2}$$

$$X_{lm}(\theta) = (-1)^m \sqrt{\frac{C_{lm}}{4\pi}} P_{lm}(\cos\theta), \tag{3}$$

$$P_{lm}(\mu) = \frac{1}{2^l l!} (1 - \mu^2)^{m/2} \left(\frac{d}{d\mu}\right)^{l+m} (\mu^2 - 1)^l, \tag{4}$$

where $P_{lm}(\mu)$ is the associated Legendre function, and the normalization constant (e.g. Edmonds 1996; Dahlen & Tromp 1998)

$$C_{lm} = (2l + 1)\frac{(l - m)!}{(l + m)!}. \tag{5}$$

With these choices the harmonics $Y_{lm}(\theta, \phi)$ are orthonormalized on the unit sphere:

$$\int_{\Omega} Y_{lm}(\hat{\mathbf{r}}) Y_{l'm'}(\hat{\mathbf{r}}) \, d\Omega = \delta_{ll'} \delta_{mm'}. \tag{6}$$

The fixed-order orthogonality relation for $X_{lm}(\theta)$ is

$$\int_0^\pi X_{lm} X_{l'm} \sin\theta \, d\theta = \frac{1}{2\pi} \delta_{ll'}. \tag{7}$$

The addition theorem expresses the sum over all orders of spherical harmonics at different positions in terms of the angular distance $\Delta = \arccos(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}')$ between them as

$$\sum_{m=-l}^{l} Y_{lm}(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}}') = \left(\frac{2l + 1}{4\pi}\right) P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}'), \tag{8}$$

where we note that $P_{l0} = P_l$ and $P_l(1) = 1$. The delta function $\delta(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = (\sin\theta)^{-1}\delta(\theta - \theta')\delta(\phi - \phi')$ defined by

$$\delta(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \sum_{l=0}^{\infty} \left(\frac{2l + 1}{4\pi}\right) P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}') \tag{9}$$

has the usual sifting property

$$\int_{\Omega} \delta(\hat{\mathbf{r}}, \hat{\mathbf{r}}') f(\hat{\mathbf{r}}) \, d\Omega = f(\hat{\mathbf{r}}'). \tag{10}$$

The sum over all degrees to infinity of the fixed-order colatitudinal functions at different arguments results in the colatitudinal delta function:

$$(\sin\theta)^{-1}\delta(\theta - \theta') = 2\pi \sum_{l=m}^{\infty} X_{lm}(\theta) X_{lm}(\theta'). \tag{11}$$

2.2 Noisy measurements at satellite altitude

For convenience we separate the signal into a bandlimited portion restricted to the degrees $l = 0 \to L$ and a portion over the degrees $l = L + 1 \to \infty$ that complement it:

$$s(\hat{\mathbf{r}}) = \sum_{lm}^{L} s_{lm} Y_{lm}(\hat{\mathbf{r}}) + \sum_{lm > L}^{\infty} s_{lm} Y_{lm}(\hat{\mathbf{r}}), \tag{12}$$

where we define L as the spherical harmonic bandwidth. At a satellite altitude a above the unit sphere the analytic signal is given by

$$s_\uparrow(\hat{\mathbf{r}}) = \sum_{lm}^{L} s_{\uparrow lm} Y_{lm}(\hat{\mathbf{r}}) + \sum_{lm > L}^{\infty} s_{\uparrow lm} Y_{lm}(\hat{\mathbf{r}}), \tag{13}$$

where the upward continued signal coefficients are given in terms of the source-level terms by (Stacey 1992; Blakely 1995)

$$s_{\uparrow lm} = (1 + a)^{-l-1} s_{lm}. \tag{14}$$

The data over the region of coverage R are given by eq. (13) but they are contaminated by noise; in the uncovered areas $\bar{R}$, no measurements are available. A satellite thus observes

$$d(\hat{\mathbf{r}}) = \begin{cases} s_\uparrow(\hat{\mathbf{r}}) + n(\hat{\mathbf{r}}) & \text{if} \quad \hat{\mathbf{r}} \in R \\ \text{unknown} & \text{if} \quad \hat{\mathbf{r}} \in \bar{R}. \end{cases} \tag{15}$$

We will restrict attention to the case in which the measurement noise $n(\hat{\mathbf{r}})$ is additive and given by a zero-mean stochastic process, which we assume to be white:

$$\langle n(\hat{\mathbf{r}}) \rangle = 0, \tag{16}$$

$$\langle n(\hat{\mathbf{r}}) n(\hat{\mathbf{r}}') \rangle = N\delta(\hat{\mathbf{r}}, \hat{\mathbf{r}}'). \tag{17}$$

Thus, the power of the noise is denoted by N, and we use angular brackets to denote the ensemble averaging over all possible realizations required to define the process mean and its spatial (co)variance.

Combining eqs (13)–(14) and (1) with the definition (8), we can write the signal observed at orbital level as a convolution of the surface-level signal in the form

$$s_\uparrow(\hat{\mathbf{r}}) = \int_{\Omega} \Gamma(\hat{\mathbf{r}}, \hat{\mathbf{r}}') s(\hat{\mathbf{r}}') \, d\Omega', \tag{18}$$

where we have defined a 'point spread function'

$$\Gamma(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \sum_{l=0}^{\infty} (1 + a)^{-l-1} \left(\frac{2l + 1}{4\pi}\right) P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}'). \tag{19}$$

Thus, the value of the potential field that is observed at a point $\hat{\mathbf{r}}$ outside the unit sphere is a weighted mixture of the function values at $\hat{\mathbf{r}}$ and distant other points $\hat{\mathbf{r}}'$ on the unit sphere. Measurements taken by a satellite at $a > 0$ are affected by regions it does not fly over directly. A satellite thus does probe into the uncovered regions; conversely, in regions of coverage, it may be affected by uncovered areas. As the fractional altitude a increases, the convolution kernel $\Gamma(\hat{\mathbf{r}}, \hat{\mathbf{r}}')$ is increasingly supported globally. On the other hand, when $a = 0$, eq. (19) returns the delta function, eq. (9); and eq. (18) merely illustrates its sifting property (10).

2.3 A new basis for bandlimited field estimators

We seek an estimate $\hat{s}(\hat{\mathbf{r}})$ of the signal in eq. (1), at the level of the source, from the data $d(\hat{\mathbf{r}})$, given by eq. (15), at altitude a. It is crucial to realize that, although any real physical signal $s(\hat{\mathbf{r}})$ will in general be infinite-band, our estimate $\hat{s}(\hat{\mathbf{r}})$ must always be bandlimited. We are thus at liberty to define a new, bandlimited set of basis functions, to replace the spherical harmonics. In this manner, the broadband source field can be expressed as

$$s(\hat{\mathbf{r}}) = \sum_{\alpha=1}^{(L+1)^2} s_\alpha g_\alpha(\hat{\mathbf{r}}) + \sum_{lm > L}^{\infty} s_{lm} Y_{lm}(\hat{\mathbf{r}}). \tag{20}$$

whereas the bandlimited estimated field is given by

$$\hat{s}(\hat{\mathbf{r}}) = \sum_{lm}^{L} \hat{s}_{lm} Y_{lm}(\hat{\mathbf{r}}) = \sum_{\alpha=1}^{(L+1)^2} \hat{s}_\alpha g_\alpha(\hat{\mathbf{r}}). \tag{21}$$

These new bandlimited basis functions $g_\alpha(\hat{\mathbf{r}})$, $\alpha = 1, \ldots, (L + 1)^2$ will themselves be combinations of spherical harmonics, inasmuch

as they are defined by the transform pair

$$g_\alpha(\hat{\mathbf{r}}) = \sum_{lm}^{L} g_{\alpha,lm} Y_{lm}(\hat{\mathbf{r}}), \tag{22a}$$

$$g_{\alpha,lm} = \int_\Omega g_\alpha(\hat{\mathbf{r}}) Y_{lm}(\hat{\mathbf{r}}) \, d\Omega, \tag{22b}$$

$$Y_{lm}(\hat{\mathbf{r}}) = \sum_{\alpha=1}^{(L+1)^2} g_{\alpha,lm} g_\alpha(\hat{\mathbf{r}}). \tag{22c}$$

The new basis is rendered orthonormal by requiring that

$$\int_\Omega g_\alpha(\hat{\mathbf{r}}) g_\beta(\hat{\mathbf{r}}) \, d\Omega = \delta_{\alpha\beta}, \tag{23a}$$

$$\sum_{lm}^{L} g_{\alpha,lm} g_{\beta,lm} = \delta_{\alpha\beta}, \tag{23b}$$

$$\sum_{\alpha=1}^{(L+1)^2} g_{\alpha,lm} g_{\alpha,l'm'} = \delta_{ll'}\delta_{mm'}. \tag{23c}$$

The transformation of the spherical harmonic basis coefficients $\hat{s}_{lm}$ of the estimate to the expansion coefficients $\hat{s}_\alpha$ in the new basis is achieved by

$$\hat{s}_{lm} = \sum_{\alpha=1}^{(L+1)^2} g_{\alpha,lm}\hat{s}_\alpha \quad \text{and} \quad \hat{s}_\alpha = \sum_{lm}^{L} g_{\alpha,lm}\hat{s}_{lm}, \tag{24}$$

as can be easily deduced by combining eq. (21) with the orthonormality conditions, eqs (6) and (23), and using eq. (22).

Upward continued to the satellite altitude, the estimate in either basis is

$$\hat{s}_\uparrow(\hat{\mathbf{r}}) = \sum_{lm}^{L} \hat{s}_{\uparrow lm} Y_{lm}(\hat{\mathbf{r}}) = \sum_{\alpha=1}^{(L+1)^2} \hat{s}_{\uparrow\alpha} g_\alpha(\hat{\mathbf{r}}), \tag{25}$$

where the spherical harmonic coefficients are naturally given by the analogue of eq. (14), namely

$$\hat{s}_{\uparrow lm} = (1+a)^{-l-1}\hat{s}_{lm}. \tag{26}$$

The upward continued expansion coefficients of the new basis are

$$\hat{s}_{\uparrow\alpha} = \sum_{lm}^{L} g_{\alpha,lm}(1+a)^{-l-1}\hat{s}_{lm} \tag{27a}$$

$$= \sum_{\beta=1}^{(L+1)^2} \left(\sum_{lm}^{L} g_{\alpha,lm}(1+a)^{-l-1}g_{\beta,lm} \right) \hat{s}_\beta, \tag{27b}$$

as is verified by combining eqs (23)–(26). As expected, eq. (27) reduces to a trivial identity at the surface of the unit sphere, that is, when $a = 0$, by virtue of eqs (23)–(24).

Given that the measurements made by the satellite are restricted to the domain R on the unit sphere Ω, we consider it natural to require of the new basis functions that they be optimally concentrated on this domain. We will seek a basis of functions $g(\hat{\mathbf{r}})$ whose energy is maximally concentrated inside of the domain R by maximizing the spatial energy ratio,

$$\lambda = \frac{\int_R g^2 \, d\Omega}{\int_\Omega g^2 \, d\Omega} = \text{maximum}. \tag{28}$$

Eq. (28) is a statement of Slepian's problem, a classic in 1-D timeseries analysis (Slepian 1983; Percival & Walden 1993), on the

2-D sphere, which we have recently studied in detail (Simons *et al.* 2006). In the next section, we review the main properties of the general solution of eq. (28) for concentration domains of arbitrary geometry, which we subsequently specialize to the geodetic context by imposing the circular and equatorial symmetry of the double-cap polar gap.

3 SLEPIAN'S SPHERICAL PROBLEM

In Slepian's problem, the spatial concentration of a bandlimited function $g(\hat{\mathbf{r}})$ given by

$$g = \sum_{lm}^{L} g_{lm} Y_{lm}, \qquad g_{lm} = \int_\Omega g \, Y_{lm} \, d\Omega, \tag{29}$$

to a region R of area A on the unit sphere Ω is expressed as the norm ratio, eq. (28). Maximization of this concentration criterion can be achieved in the spectral domain by solving the algebraic eigenvalue problem

$$\mathsf{D}\mathbf{g} = \lambda\mathbf{g}, \tag{30}$$

where $\mathbf{g}$ is the $(L+1)^2$–dimensional spherical harmonic coefficient column vector

$$\mathbf{g} = (g_{00} \cdots g_{lm} \cdots g_{LL})^{\mathsf{T}} \tag{31}$$

and D is the $(L+1)^2 \times (L+1)^2$–dimensional matrix

$$\mathsf{D} = \begin{pmatrix} D_{00,00} & \cdots & D_{00,LL} \\ \vdots & & \vdots \\ D_{LL,00} & \cdots & D_{LL,LL} \end{pmatrix}, \tag{32}$$

whose elements are given by

$$D_{lm,l'm'} = \int_R Y_{lm} Y_{l'm'} \, d\Omega. \tag{33}$$

The obvious symmetry $\mathsf{D}^{\mathsf{T}} = \mathsf{D}$ guarantees that the eigenvectors $\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_{(L+1)^2}$ are mutually orthogonal. We choose them to be orthonormal, following eqs (23) and (30),

$$\mathbf{g}_\alpha^{\mathsf{T}}\mathbf{g}_\beta = \delta_{\alpha\beta} \quad \text{and} \quad \mathbf{g}_\alpha^{\mathsf{T}}\mathsf{D}\mathbf{g}_\beta = \lambda_\alpha\delta_{\alpha\beta}. \tag{34}$$

The corresponding spatial Slepian functions g are orthonormal over the whole sphere Ω and orthogonal over the region R, as follows from eqs (22)–(23), (33) and (30),

$$\int_\Omega g_\alpha g_\beta \, d\Omega = \delta_{\alpha\beta} \quad \text{and} \quad \int_R g_\alpha g_\beta \, d\Omega = \lambda_\alpha\delta_{\alpha\beta}. \tag{35}$$

The leftmost equations in eqs (34)–(35) correspond to the conditions of eq. (23) and guarantee that the solution indeed forms a valid orthonormal basis. The rightmost equations illustrate the so-called double orthogonality of the Slepian basis (Gilbert & Slepian 1977), which, as we will see in a later section, is a central feature of its utility for the geodetic estimation problem.

An approach equivalent to the maximization of eq. (28) is to find broadband functions $h(\hat{\mathbf{r}})$ that are spacelimited to the domain R, but spectrally concentrated in a bandwidth interval $0 \le l \le L$. The concentration measure in this case,

$$\lambda = \frac{\sum_{lm}^{L} h_{lm}^2}{\sum_{lm}^{\infty} h_{lm}^2} = \text{maximum}, \tag{36}$$

is satisfied by the eigenfunctions of a Fredholm integral eigenvalue equation in the spatial domain,

$$\int_R D(\hat{\mathbf{r}}, \hat{\mathbf{r}}') h(\hat{\mathbf{r}}') \, d\Omega' = \lambda h(\hat{\mathbf{r}}), \qquad \hat{\mathbf{r}} \in R. \tag{37}$$

The symmetric kernel of eq. (37) depends only on the geodesic angular distance, Δ, between $\hat{\mathbf{r}}$ and $\hat{\mathbf{r}}'$,

$$D(\hat{\mathbf{r}}, \hat{\mathbf{r}}') = \sum_{l=0}^{L} \left(\frac{2l+1}{4\pi}\right) P_l(\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}'). \tag{38}$$

The problems of finding bandlimited functions g concentrated to a spatial interval or spacelimited functions h concentrated in a spectral interval are completely equivalent. The domain of eq. (37) can be extended to the entire sphere in which case it applies to the band-limited functions g and can be reduced to eq. (30) using eqs (38), (8), (29) and (33). We normalize such that the eigenfunctions g that maximize the spatial energy ratio (28) are identical, within the region R, to the eigenfunctions h maximizing the spectral ratio (36),

$$h(\hat{\mathbf{r}}) = \begin{cases} g(\hat{\mathbf{r}}) & \text{if } \hat{\mathbf{r}} \in R \\ 0 & \text{otherwise.} \end{cases} \tag{39}$$

A straightforward consequence of the eqs (29), (33) and (39) is that

$$h_{lm} = \sum_{l'm'}^{L} D_{lm,l'm'} g_{l'm'}, \quad 0 \le l \le \infty, \quad -l \le m \le l. \tag{40}$$

This expresses the coefficients h_{lm} at all degrees to infinity in terms of the coefficients g_{lm} whose degree range is limited to the bandwidth L. By eq. (30), it amounts to $h_{lm} = \lambda g_{lm}$ when $0 \le l \le L$. The eigenvalues of eqs (30) or (37),

$$1 > \lambda_1 \ge \lambda_2 \cdots \ge \lambda_{(L+1)^2} > 0, \tag{41}$$

measure the quality of the spatiospectral concentration: the band-limited function that is most concentrated inside R is g_1, with λ_1 the largest associated eigenvalue, and so on. The trace of D defines a diagnostic area-bandwidth product termed the Shannon number,

$$K = \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha = \int_R D(\hat{\mathbf{r}}, \hat{\mathbf{r}}) \, d\Omega = (L+1)^2 \frac{A}{4\pi}. \tag{42}$$

Spherical Slepian functions of equal Shannon number are scaled versions of each other in the asymptotic limit $A \to 0$ and $L \to \infty$ with K held fixed (Simons *et al.* 2006). Whenever the area A of the region R is a small fraction of the area of the sphere, $A \ll 4\pi$, that is, when $K \ll (L+1)^2$, there will be many more well excluded eigenfunctions with insignificant eigenvalues ($\lambda \approx 0$) than well concentrated eigenfunctions with significant eigenvalues ($\lambda \approx 1$). If on the other hand, R covers most of the sphere so that $A \approx 4\pi$ and $K \approx (L+1)^2$, there will be many more well concentrated eigenfunctions than well excluded ones.

The sum of the squares of the $(L+1)^2$ bandlimited eigenfunctions $g(\hat{\mathbf{r}})$ is independent of position on the sphere,

$$\sum_{\alpha=1}^{(L+1)^2} g_\alpha^2(\hat{\mathbf{r}}) = \frac{(L+1)^2}{4\pi} = \frac{K}{A}. \tag{43}$$

As the first K eigenfunctions $g_1, g_2, \ldots, g_K$ have near-unity eigenvalues and lie mostly in R, and the $g_{K+1}, g_{K+2}, \ldots, g_{(L+1)^2}$ remaining have eigenvalues near zero and lie mostly in the complementary region $\bar{R} = \Omega - R$, the eigenvalue-weighted sum of squares is well approximated by

$$\sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha g_\alpha^2(\hat{\mathbf{r}}) \approx \begin{cases} K/A & \text{if } \hat{\mathbf{r}} \in R \\ 0 & \text{otherwise.} \end{cases} \tag{44}$$

Taken together, the first K eigenfunctions g_α, $\alpha = 1, 2, \ldots, K$, with significant eigenvalues $\lambda_\alpha \approx 1$, provide an essentially uniform coverage of the region R. Rather than requiring $(L+1)^2$ basis functions to represent an arbitrary spatially concentrated bandlimited function, the first $K = (L+1)^2 A/(4\pi)$ members of the Slepian basis provide a very reasonable approximation.

We shall denote the operator localizing to the complementary region $\bar{R}$ by $\bar{D}$, its eigenfunctions by $\bar{g}$ and its eigenvalues by $\bar{\lambda}$. It follows from the orthogonality (6) that the elements of $\bar{D}$ are

$$\bar{D}_{lm,l'm'} = \int_{\bar{R}} Y_{lm} Y_{l'm'} \, d\Omega \tag{45a}$$

$$= \delta_{ll'} \delta_{mm'} - D_{lm,l'm'}. \tag{45b}$$

The eigenfunctions of $\bar{D}$ are identical to those of D, but their ordering indices are reversed. The bandlimited function that is most concentrated within $\bar{R}$ is most excluded from R, that is, $\bar{g}_1 = g_{(L+1)^2}$, with an associated eigenvalue $\bar{\lambda}_1 = 1 - \lambda_{(L+1)^2}$, and so on.

The localization operator D has an inverse satisfying

$$\sum_{l''m''}^{L} D_{lm,l''m''}^{-1} D_{l''m'',l'm'} = \delta_{ll'} \delta_{mm'}, \tag{46}$$

and for which $D^{-1} g = \lambda^{-1} g$. For future reference, the inverse of a weighted sum of the localization matrix and its complement obeys

$$(D + \eta \bar{D})^{-1} g = [\lambda + \eta(1 - \lambda)]^{-1} g, \tag{47a}$$

$$g_\alpha^{\mathsf{T}} (D + \eta \bar{D})^{-1} g_\beta = [\lambda_\alpha + \eta(1 - \lambda_\alpha)]^{-1} \delta_{\alpha\beta}, \tag{47b}$$

for any η. Finally, we may use eqs (33) and (8)–(10) to prove that

$$\sum_{l''m''}^{\infty} D_{lm,l''m''} D_{l''m'',l'm'} = D_{lm,l'm'}. \tag{48}$$

4 AXISYMMETRIC DOMAINS

In the previous section we showed that the optimally concentrated bandlimited basis functions that are the solutions to the Slepian problem are found by diagonalization of the operator in eq. (33). That this is in general possible for arbitrary geometries was shown by Simons *et al.* (2006). However, the particular geometry of data acquisition on the sphere in the geodetic estimation problem (Fig. 1) allows for substantial simplifications of this general result. We discuss the special case of finding concentrated basis functions on the latitudinal belt, the domain over which satellite measurements are made, via the concentration within a single and a double polar cap. As we have seen, the eigenfunctions on a domain R are identical to those on a complementary spherical domain $\bar{R}$, but with their ordering indices reversed. Identifying R with the polar caps, rather than their complement, the belt, as we do—*in this section and the one that follows only*—greatly simplifies the equations.

4.1 Concentration within an axisymmetric polar cap

When the region of concentration is a circularly symmetric cap of colatitudinal radius Θ, centred on the north pole, that is,

$$R = \{\theta : 0 \le \theta \le \Theta\}, \tag{49}$$

of area $A = 2\pi(1 - \cos \Theta)$, the matrix elements of eq. (33) are

$$D_{lm,l'm'} = 2\pi \delta_{mm'} \int_0^\Theta X_{lm} X_{l'm} \sin \theta \, d\theta. \tag{50}$$

The Kronecker delta δ_{mm} renders the matrix D of eq. (32) block-diagonal,

$$D = \text{diag}\left(D^0, D^1, D^1, \ldots, D^L, D^L\right),\tag{51}$$

where every submatrix $D^m \neq D^0$ occurs twice due to the doublet degeneracy of $\pm m$. Rather than solving the complete $(L+1)^2 \times (L+1)^2$ eigenvalue equation (30), we may instead solve a series of $(L-m+1) \times (L-m+1)$ spectral-domain eigenvalue problems, one for each non-negative order m,

$$Dg = \lambda g,\tag{52}$$

where we have dropped the superscript identifying the order. The eigenvalues belonging to every non-zero order, $m > 0$, occur twice. In eq. (52) the column vector g collects the spherical harmonic coefficients of order m,

$$g = (g_m \cdots g_l \cdots g_L)^{\mathsf{T}},\tag{53}$$

and the fixed-order matrix D is of the form

$$D = \begin{pmatrix} D_{mm} & \cdots & D_{mL} \\ \vdots & & \vdots \\ D_{Lm} & \cdots & D_{LL} \end{pmatrix},\tag{54}$$

where, for a particular order $0 \leq m \leq L$,

$$D_{ll'} = 2\pi \int_0^\Theta X_{lm} X_{l'm} \sin\theta \, d\theta.\tag{55}$$

Various methods exist to evaluate the elements of eq. (55) (Wieczorek & Simons 2005; Simons *et al.* 2006). The important point is that, while symmetric, and banded, the matrix D in eq. (54) is never sparse. Its numerical computation thus requires on the order of $(L-m+1)^2/2$ integrals each. The fixed-order Shannon number is the trace of D,

$$K_m = \sum_{\alpha=1}^{L-m+1} \lambda_\alpha.\tag{56}$$

We rank the $L - m + 1$ eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_{L-m+1}$ obtained by solving the fixed-order problem (52) so that

$$1 > \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{L-m+1} > 0,\tag{57}$$

and orthonormalize the eigenvectors $g_1, g_2, \ldots, g_{L-m+1}$ as in eq. (34). The associated bandlimited eigenfunctions $g_1(\theta)$, $g_2(\theta), \ldots, g_{L-m+1}(\theta)$ are given by

$$g = \sum_{l=m}^{L} g_l X_{lm}, \qquad g_l = 2\pi \int_0^\pi g \, X_{lm} \sin\theta \, d\theta,\tag{58}$$

and satisfy the colatitudinal orthogonality relations

$$2\pi \int_0^\pi g_\alpha g_\beta \sin\theta \, d\theta = \delta_{\alpha\beta},\tag{59a}$$

$$2\pi \int_0^\Theta g_\alpha g_\beta \sin\theta \, d\theta = \lambda_\alpha \delta_{\alpha\beta}.\tag{59b}$$

The optimally concentrated spatial eigenfunctions $g(\hat{\mathbf{r}})$ for a given order $-L \leq m \leq L$ are expressed in terms of the fixed-order colatitudinal eigenfunctions (58) by

$$g(\theta, \phi) = \begin{cases} \sqrt{2} \, g(\theta) \cos m\phi & \text{if } -L \leq m < 0 \\ g(\theta) & \text{if } m = 0 \\ \sqrt{2} \, g(\theta) \sin m\phi & \text{if } 0 < m \leq L. \end{cases}\tag{60}$$

The complementary fixed-order matrices are given by

$$\bar{D}_{ll'} = \delta_{ll'} - D_{ll'}.\tag{61}$$

The eigenfunctions of the fixed-order matrix $\bar{D}$ are identical to those of D but appear in reverse order, and their eigenvalues sum to one. The axisymmetric inversion formula analogous to eq. (46) is

$$\sum_{l''=0}^{L} D_{ll''}^{-1} D_{l''l'} = \delta_{ll'},\tag{62}$$

and the axisymmetric version of eq. (48), from eqs (11) and (55), is

$$\sum_{l''=m}^{\infty} D_{ll''}^m D_{l''l'}^m = D_{ll'}^m.\tag{63}$$

4.2 Concentration within a double polar cap

When the region of concentration is a pair of axisymmetric antipodal caps of colatitudinal radius Θ, that is, when

$$R = \{\theta : 0 \leq \theta \leq \Theta\} \cup \{\theta : \pi - \Theta \leq \theta \leq \pi\},\tag{64}$$

of area $A = 4\pi(1 - \cos\Theta)$, the reflection symmetry

$$X_{lm}(\pi - \theta) = (-1)^{l+m} X_{lm}(\theta)\tag{65}$$

checkers the fixed-order matrices D with zeroes, following

$$D_{ll'} = 2\pi[1 + (-1)^{l+l'}] \int_0^\Theta X_{lm} X_{l'm} \sin\theta \, d\theta.\tag{66}$$

Comparison of eqs (55) and (66) reveals that the eigenfunctions of the double-cap problem can be trivially obtained from the kernels belonging to the single polar cap.

The spherical Slepian functions resulting from the diagonalization of the double-cap kernel in eq. (66) are either even or odd across the equator. Indexing their parity p as even ($p = e$) or odd ($p = o$), we modify eq. (58) to explicitly skip every other degree by using a primed summation symbol,

$$g_p = \sum_{l=m_p}^{L_p} {}' g_l X_{lm},\tag{67}$$

where the lower limit m_p is given by

$$m_e = m \quad \text{and} \quad m_o = m + 1,\tag{68}$$

and the upper limit L_p is

$$L_e = \begin{cases} L & \text{if } m \text{ and } L \text{ have the same parity} \\ L - 1 & \text{opposite,} \end{cases}\tag{69a}$$

$$L_o = \begin{cases} L - 1 & \text{if } m \text{ and } L \text{ have the same parity} \\ L & \text{opposite.} \end{cases}\tag{69b}$$

In this formalism, the coefficients that are required for $g_e(\hat{\mathbf{r}})$ are $g_m, g_{m+2}, \ldots, g_L$ if m and L are both even or both odd, and g_m, $g_{m+2}, \ldots, g_{L-1}$ if m and L are of opposite parity. Likewise, the coefficients of $g_o(\hat{\mathbf{r}})$ are $g_{m+1}, g_{m+3}, \ldots, g_{L-1}$ if m and L are both even or both odd, and $g_{m+1}, g_{m+3}, \ldots, g_L$ if m and L have opposite parity. Eqs (65) and (67) then confirm that

$$g_e(\theta) = g_e(\pi - \theta) \quad \text{and} \quad g_o(\theta) = -g_o(\pi - \theta).\tag{70}$$

While an equation of the form (52) returns an alternation of even and odd functions with decreasing eigenvalues λ, the indices of the matrix D may be permuted to form a block-diagonal form

$$D' = \text{diag}(D_e, D_o),\tag{71}$$

for which the roughly half-size separate eigenvalue equations

$$\mathsf{D}_e \mathsf{g}_e = \lambda_e \mathsf{g}_e \quad \text{and} \quad \mathsf{D}_o \mathsf{g}_o = \lambda_o \mathsf{g}_o \tag{72}$$

return exclusively even or odd solutions. This avoids round-off problems and speeds up the diagonalization. The slight perversity of our notation is that, in an all-even or all-odd approach as in eq. (72), writing the fixed-order coefficient g_l or indeed any expression involving the spherical harmonic degree l, always has to be accompanied by the set of allowable values, since, depending on the parity, $l = m_p$, $m_p + 2, \ldots, L_p$.

5 THE MAGIC OF COMMUTATION

While conceptually simple, the formalism presented in the previous section suffers from two important difficulties. First, assembling the matrices of eqs (55) and (66) requires the calculation of $\mathcal{O}(L^2)$ matrix elements, by numerical integration or other means (Wieczorek & Simons 2005). Second, and more importantly, when a large number of near-zero eigenvalues is present, e.g. when $\Theta \to 0$ and the complementary solutions are sought on $\bar{R}$, the diagonalization is rarely stable, as discussed by Albertella *et al.* (1999). In principle, any orthogonal set of solutions might suffice to solve the problem at hand, but those solutions will vary depending on the method of computation. The method outlined below always produces stable, unique, solutions, and it does so at a speed which requires only $\mathcal{O}(L)$ algebraic evaluations to construct the kernels.

5.1 A commuting operator for the single polar cap

In the case of the single symmetric polar cap, eq. (37) can be rewritten as a series of fixed-order integral equations

$$\int_0^\Theta D(\theta, \theta') h(\theta') \sin \theta' \, d\theta' = \lambda h(\theta), \quad 0 \le \theta \le \Theta, \tag{73}$$

each with an m-dependent, separable, symmetric kernel

$$D(\theta, \theta') = 2\pi \sum_{l=m}^L X_{lm}(\theta) X_{lm}(\theta'). \tag{74}$$

Building on the results derived by Gilbert & Slepian (1977), Grünbaum *et al.* (1982) found a second-order differential operator that commutes with the convolutional integral operator of eq. (73). For any $0 \le m \le L$, it is of the form

$$\mathcal{T} = (\cos \Theta - \cos \theta)\nabla_m^2 + \sin \theta \frac{d}{d\theta} - L(L+2)\cos\theta, \tag{75}$$

where $\nabla_m^2 = d^2/d\theta^2 + \cot\theta \, (d/d\theta) - m^2(\sin\theta)^{-2}$ is the fixed-order Laplace–Beltrami operator. The proof of the commutation relation is sketched in Simons *et al.* (2006). Since commuting operators have identical eigenfunctions, the spacelimited, fixed-order eigenfunctions $h(\theta)$ can be found by solving the differential eigenvalue equation

$$\mathcal{T} h(\theta) = \chi h(\theta), \quad 0 \le \theta \le \Theta, \tag{76}$$

where $\chi \ne \lambda$ is the associated Grünbaum eigenvalue.

Grünbaum's operator is a Sturm-Liouville operator (Simons *et al.* 2006). Thus, eq. (76) has a simple and easily sorted spectrum, with an infinite number of distinct eigenvalues $\chi_1 < \chi_2 < \cdots$ having an accumulation point at infinity. The rank orderings of the eigenvalues $\chi_1, \chi_2, \ldots$ and the spatiospectral concentration factors λ_1, $\lambda_2, \ldots, \lambda_{L-m+1}$ are reversed, so that the eigenfunction $h_1(\theta)$ associated with the numerically smallest eigenvalue χ_1, which has no

nodes in the polar cap $0 \le \theta \le \Theta$, is the best concentrated fixed-order eigenfunction; $h_2(\theta)$, which has exactly one node, is the next best concentrated, and so on.

Extending the domain of eq. (76) to the entire domain $0 \le \theta \le \pi$ transforms the unknown functions from the spacelimited functions $h(\theta)$ again into the bandlimited functions $g(\theta)$. Eq. (76) is then equivalent to the algebraic eigenvalue equation

$$\mathsf{T} \mathsf{g} = \chi \mathsf{g}, \tag{77}$$

where T is the $(L - m + 1) \times (L - m + 1)$ matrix with coefficients

$$T_{ll'} = 2\pi \int_0^\pi X_{lm}(\mathcal{T} X_{l'm}) \sin\theta \, d\theta. \tag{78}$$

Eqs (77)–(78) are completely equivalent to eqs (52) and (55). Both matrices D and T are symmetric, $\mathsf{D} = \mathsf{D}^\mathsf{T}$ and $\mathsf{T} = \mathsf{T}^\mathsf{T}$. In addition, they commute, $\mathsf{DT} = \mathsf{TD}$, so they have identical eigenvectors. There are a number of ways to evaluate the elements of the Grünbaum matrix in eq. (78), but the important result is that T is tridiagonal (Simons *et al.* 2006),

$$T_{ll} = -l(l+1)\cos\Theta, \tag{79a}$$

$$T_{ll+1} = \left[l(l+2) - L(L+2)\right]\sqrt{\frac{(l+1)^2 - m^2}{(2l+1)(2l+3)}}, \tag{79b}$$

$$T_{ll'} = 0 \quad \text{otherwise.} \tag{79c}$$

Eq. (77) can be used to find the $(L - m + 1)$–dimensional eigenvectors g and thus the optimally concentrated polar cap eigenfunctions $g(\theta)$ by numerical diagonalization of a tridiagonal matrix T with analytically prescribed elements and a spectrum of eigenvalues χ that is guaranteed to be regular. Unlike the diagonalization of the original matrix D in eq. (52), this procedure enables the stable computation of bandlimited functions that are optimally concentrated in a large rather than a small region of the unit sphere, as may be the case in geodesy.

5.2 A commuting operator for the double polar cap

Knowing that the solutions to the concentration problem for the double polar cap are either even or odd across the equator, we may write the integral equation (37), by analogy with eq. (73), as follows. Indicating the parity of the solutions by the subscript p, which takes the values $p = e$ for the even solutions and $p = o$ for the odd solutions, it can be seen that

$$\int_0^\Theta D_p(\theta, \theta') h_p(\theta') \sin\theta' \, d\theta' = \lambda h_p(\theta), \tag{80}$$

which is valid inside the double antipodal polar cap

$$\{\theta : 0 \le \theta \le \Theta\} \cup \{\theta : \pi - \Theta \le \theta \le \pi\}, \tag{81}$$

and where the m-dependent kernel, analogous to eq. (74), is

$$D_p(\theta, \theta') = 4\pi \sum_{l=m_p}^{L_p} {}' X_{lm}(\theta) X_{lm}(\theta'). \tag{82}$$

As in eq. (67), the primed summation skips every second entry, and the lower and upper limits are as in eqs (68)–(69).

Again basing ourselves on the results of Gilbert & Slepian (1977) and Grünbaum *et al.* (1982), we show in Appendix A that a Sturm-Liouville second-order differential operator that commutes with the

convolutional integral operators of eq. (80) is of the form

$$\mathcal{T}_p = (\cos^2 \Theta - \cos^2 \theta)\nabla_m^2 + 2\cos\theta \sin\theta \frac{d}{d\theta}$$
$$- L_p(L_p + 3)\cos^2\theta. \tag{83}$$

The individual matrix operators T_p are once again tridiagonal and symmetric and commute with the even or odd D_p of eq. (72). The bandwidth L_p is the same as in (69). The elements of the double-cap Grünbaum matrices are (see Appendix B)

$$T_{ll}^p = -l(l+1)\cos^2\Theta + \frac{2}{2l+3}[(l+1)^2 - m^2]$$
$$+ [(l-2)(l+1) - L_p(L_p+3)]$$
$$\times \left[\frac{1}{3} - \frac{2}{3}\frac{3m^2 - l(l+1)}{(2l+3)(2l-1)} \right], \tag{84a}$$

$$T_{ll+2}^p = \frac{[l(l+3) - L_p(L_p+3)]}{2l+3}$$
$$\times \sqrt{\frac{[(l+2)^2 - m^2][(l+1)^2 - m^2]}{(2l+5)(2l+1)}}, \tag{84b}$$

$$T_{ll'}^p = 0 \quad \text{otherwise.} \tag{84c}$$

We again emphasize that, since we focus our attention separately on the kernels returning even or odd eigenfunctions g_e or g_o, the degrees involved are restricted to $l, l' = m_p, m_p + 2, \ldots, L_p$. Since every other degree in the matrix described by eq. (84) is skipped, both T_e and T_o are tridiagonal as in the single-cap case. As in eq. (72) we compute the even and odd eigenfunctions by separately solving

$$\mathsf{T}_e \mathsf{g}_e = \chi_e \mathsf{g}_e \quad \text{and} \quad \mathsf{T}_o \mathsf{g}_o = \chi_o \mathsf{g}_o. \tag{85}$$

Subsequently, we establish a single rank order of decreasing spatiospectral concentration: either per order, as in eq. (57), or across all orders, as in eq. (41).

6 A SLEPIAN BASIS ON THE BELT

Concentration within a single polar cap was treated extensively by Wieczorek & Simons (2005) and Simons *et al.* (2006). We refer to their figures for illustrations. In this section we illustrate the solutions to the concentration problem when the concentration region contains all but an antipodal pair of polar caps of radius Θ. Reverting to our notational convention in Section 2, we again use R to denote an equatorial strip or latitudinal belt extending $\pi/2 - \Theta$ north and south of the equator. Consequently, the antipodal pair of polar caps themselves is again defined to be the excluded region $\bar{R}$, in line with their role as the geodetic polar gap in which no satellite observations are available.

6.1 Spatial-domain solutions

The fixed-order eigenfunctions $g_\alpha(\theta), \alpha = 1 \to 6, 0 \le m \le 4$, that are most optimally concentrated in the latitudinal belt complementing a $\Theta = 30°$ double polar cap are plotted in Fig. 2. Their associated eigenvalues λ_α are listed to six-figure accuracy. The latitudinal belt ranges from $60°$ north to $60°$ south symmetrically about the equator. With the chosen bandwidth $L = 18$, the Shannon number defined in eq. (42) is $K = (L+1)^2 \cos\Theta \approx 313$, which approximates the number of well concentrated eigenfunctions with $\lambda \approx 1$. The best concentrated eigensolution of every order is a bell-shaped even function with no nodes in the belt. In keeping with the Sturm-Liouville character of the Grünbaum operator, every subsequent
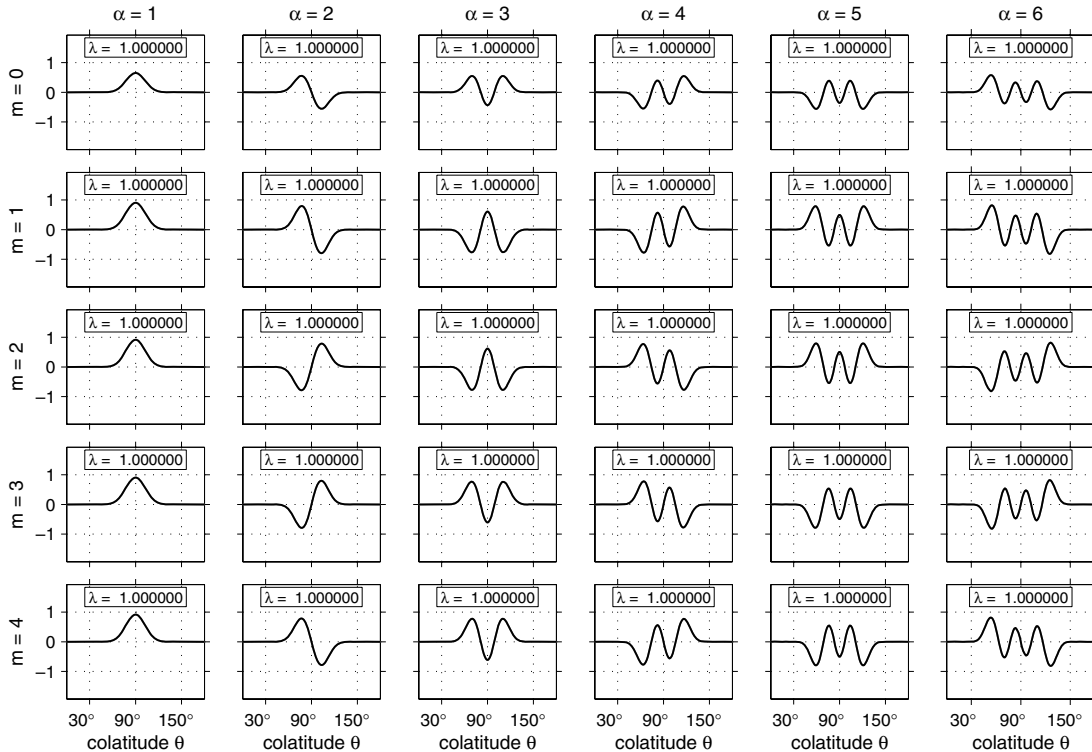
Figure 2. Colatitudinal dependence of the first six fixed-order, $m = 0 \to 4$, eigenfunctions $g_\alpha(\theta), \alpha = 1 \to 6$, bandlimited to $L = 18$, that are well concentrated in the latitudinal belt extending $\pm 60°$ on either side of the equator. The quality of the spatial concentration is expressed by the labelled eigenvalues λ_α. None of the plotted functions show appreciable energy inside the complementary pair of antipodal polar caps of radius $\Theta = 30°$.
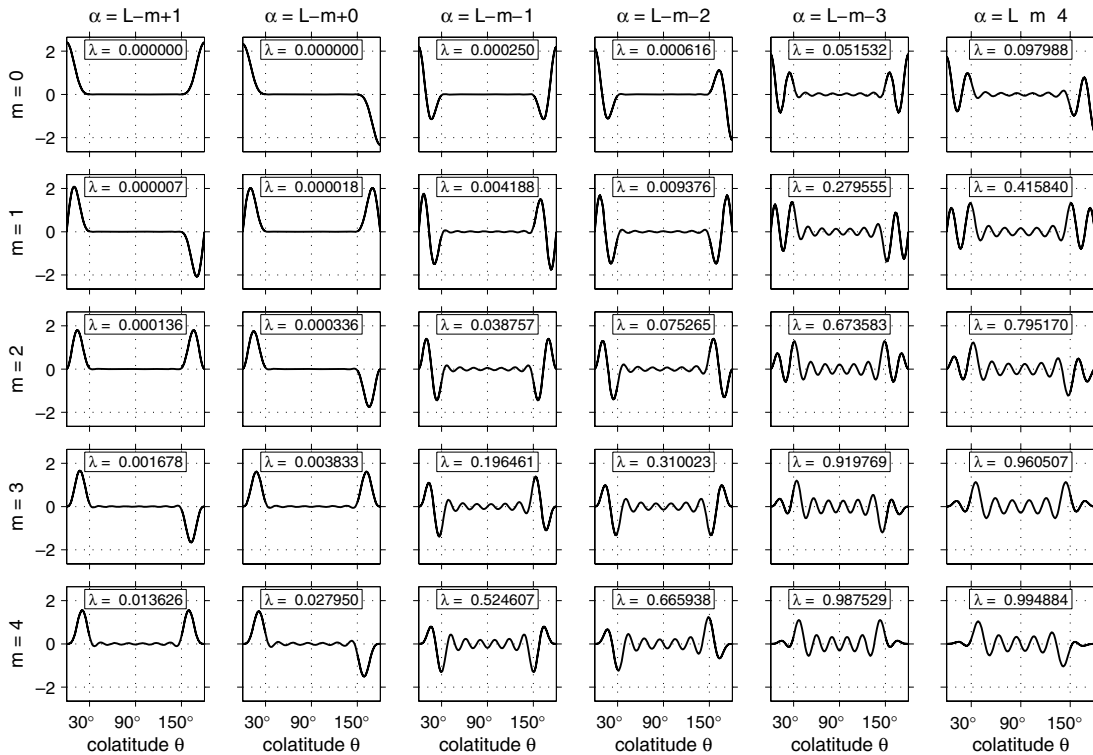
Figure 3. Colatitudinal dependence of the last six fixed-order, $m = 0 \rightarrow 4$, eigenfunctions $g_\alpha(\theta)$, $\alpha = L - m + 1 \rightarrow L - m - 4$, bandlimited to $L = 18$. These are generally poorly concentrated in the latitudinal belt $\pm 60°$ about the equator, except where the rank α exceeds the fixed-order Shannon number K_m (examples in lower right). The functions that have the least energy inside of the equatorial belt, as shown by their low eigenvalues λ_α, are best concentrated inside the complementary polar caps of colatitudinal radius $\Theta = 30°$.

solution acquires one more node, so that the second best of every order is an odd function, the third is even, and so on. All of the eigenvalues shown in Fig. 2, calculated by numerically integrating eq. (28), are equal to one within six-figure accuracy, indicating that the concentration to the belt is nearly perfect, while both poles are almost completely excluded. Since the concentration region is very large, the calculation of these functions by any means other than the Grünbaum procedure described above will fail.

With the parameters unchanged from Fig. 2, Fig. 3 shows the six most poorly concentrated eigenfunctions on the belt, $g_\alpha(\theta)$, for which $\alpha = L - m + 1 \rightarrow L - m - 4$. These now naturally have almost all of their energy inside of the antipodal pair of polar caps of radius $\Theta = 30°$. For the zonal functions of order $m = 0$, the even-odd alternation starting at $\alpha = 1$ with an even function in Fig. 2 ends at $\alpha = L + 1$ with an even function, since L itself is even. At $m = 1$, the sequence starts with an even function but ends at $\alpha = L$ with an odd function, at $m = 2$ with an even function at $\alpha = L - 1$, and so on. Thus, the worst concentrated $m = 0$ eigenfunction is even about the equator, the worst $m = 1$ eigenfunction is odd, and so on, in a pattern that alternates with increasing order. Had L itself been odd, the worst concentrated zonal function would have been odd, the worst $m = 1$ function even, and so on, reversing the pattern.

Three-dimensional perspective views of the first four of the fixed-order $m = 0 \rightarrow 2$ functions whose colatitudinal dependence we plotted in Fig. 2 are shown in Fig. 4. In accordance with eq. (60) the zonal $m = 0$ eigenfunctions do not display any longitudinal zero crossings, since the number of longitudinal nodes follows the order m. Similarly, in Fig. 5 we plot a 3-D rendering of twelve of the worst concentrated eigenfunctions of Fig. 3.

In Fig. 6 we show the eigenvalue-weighted pointwise sums of squares $\sum_\alpha \lambda_\alpha g_\alpha^2(\theta, \phi)$ for latitudinal belts complementary to dou-

ble polar caps of radii $\Theta = 5°, 10°, 15°, 20°$, of bandwidth $L = 18$. The cumulative sums are concentrated inside of the latitudinal belt; solid lines in grey and black distinguish the sums carried up to the first K (the Shannon number) or all $(L + 1)^2$ possible terms. In contrast, the cumulative sums of the cap eigenfunctions, shown in Fig. 7, are concentrated within the double polar cap. The full unweighted sums $\sum_\alpha g_\alpha^2(\theta, \phi)$ of all $(L + 1)^2$ terms (dashed black lines) are exactly $K/A = (L + 1)^2/(4\pi)$ over the entire sphere in accordance with eq. (43), and the expectation in eq. (44) is confirmed: inside of the concentration domain, the weighted sums approach K/A.

6.2 Eigenvalue spectra

In Fig. 8 we show the reordered, mixed-order eigenvalue spectra for the concentration problem within the latitudinal belt between polar caps of colatitudinal radii $\Theta = 5°, 10°, 15°, 20°$. Once again the maximal spherical harmonic degree is $L = 18$. The rounded Shannon numbers $K = 360, 356, 349, 339$ lie in the middle of the steep, transitional part of the spectra, roughly separating the reasonably well concentrated eigensolutions ($\lambda \geq 0.5$) from the more poorly concentrated ones ($\lambda < 0.5$) in all four cases. There are many more functions that are well concentrated in the equatorial strip than there are that are concentrated inside of the double polar cap, as shown by the break at $\alpha = 10$ in the abscissas.

The corresponding Grünbaum eigenvalue spectra are shown in Fig. 9. The ranked eigenvalues χ for every order $0 \leq m \leq L$ are connected by lines, with each sequence offset horizontally by its order, and vertically by an arbitrary 50 units, to facilitate inspection. Thus, $L + 1$ eigenvalues $\chi_1, \chi_2, \ldots, \chi_{L+1}$ are plotted for $m = 0$,
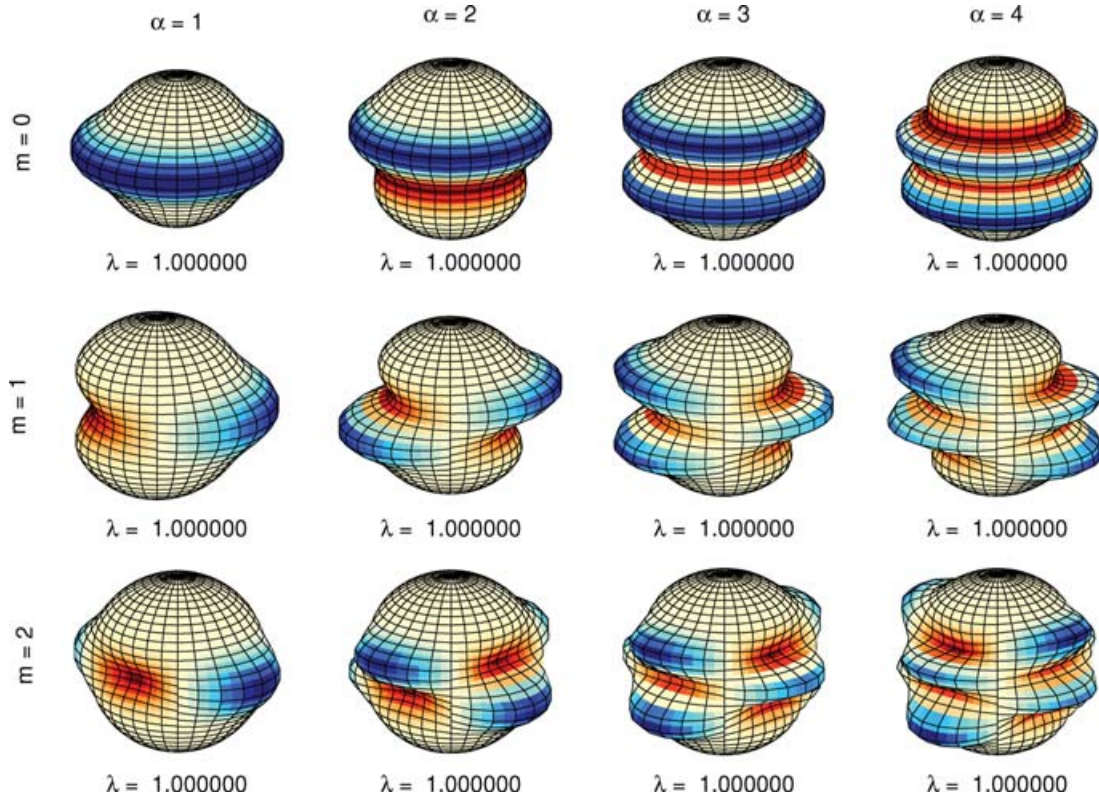
Figure 4. 3-D spatial dependence of the first four fixed-order, $m = 0 \rightarrow 2$, eigenfunctions $g_\alpha(\theta)$, $\alpha = 1 \rightarrow 4$, bandlimited to $L = 18$, well concentrated in the latitudinal belt extending $\pm 60°$ on either side of the equator, as expressed by their eigenvalues λ_α. Plot arrangement is as in Fig. 2.

whereas a single eigenvalue χ_1 is plotted for $m = L$. The spacing between adjacent fixed-order eigenvalues is roughly equant, without the numerically troublesome plateaus of nearly equal values apparent in Fig. 8. This regularity is guaranteed by the Sturm-Liouville character of the Grünbaum operator $\mathcal{T}_p$ in eq. (83).

6.3 Analytic continuation

The Slepian functions $g_1(\hat{\mathbf{r}}), \ldots, g_{(L+1)^2}(\hat{\mathbf{r}})$ are defined on the surface of the unit sphere Ω. Together they form a natural basis set for the expansion of potential fields and, in particular, estimates of these fields, on a sphere of radius $\|\hat{\mathbf{r}}\| = 1$, as in eqs (20)–(21). This new basis is localized: the support of the first Shannon number K basis functions lies mostly in the concentration region R, whereas the remainder are concentrated outside of this area of interest, in $\bar{R}$. With satellite observations we are of course mostly interested in the signal at some height above the surface of the unit sphere. We have previously derived an expression for the expansion of a field estimate at satellite altitude a, in eq. (27). It is immediately obvious from this equation that, even if we were only interested in the first K upward continued Slepian expansion coefficients of the estimate, we would still need to know and calculate the full set of $(L + 1)^2$ coefficients at zero altitude. The full impact of this statement will not become clear until later in this paper, but to anticipate it we derive here a set of Slepian basis functions that are designed specifically to represent signals at an altitude. We can do this by interpreting the Slepian functions we have just constructed as potential functions themselves. In that case their upward harmonic continuation onto a sphere larger radius $\|\mathbf{r}\| = 1 + a$, where $a > 0$, yields functions $g_\uparrow$

for which

$$g_\uparrow = \sum_{lm}^{L} g_{\uparrow lm} Y_{lm}, \qquad g_{\uparrow lm} = (1 + a)^{-l-1} g_{lm}. \tag{86}$$

Had we instead defined the Slepian functions on the larger sphere to begin with, their analogues downward continued onto the unit sphere would be obtained as

$$g_\downarrow = \sum_{lm}^{L} g_{\downarrow lm} Y_{lm}, \qquad g_{\downarrow lm} = (1 + a)^{l+1} g_{lm}. \tag{87}$$

It is thus useful to define a symmetric, $(L + 1)^2 \times (L + 1)^2$, downward continuation matrix A, whose elements are

$$A_{lm,l'm'} = (1 + a)^{l+1} \delta_{ll'} \delta_{mm'}, \tag{88}$$

which allows us to restate the equations relating the upward and downward continued coefficients to each other concisely as

$$g_\uparrow = A^{-1} g, \qquad g = A g_\uparrow, \tag{89a}$$

$$g_\downarrow = A g, \qquad g = A^{-1} g_\downarrow. \tag{89b}$$

The orthogonality relations of eqs (23) and (34) can be rewritten in terms of $g_\uparrow$ and $g_\downarrow$ in the form

$$g_{\uparrow \alpha}^{\mathsf{T}} A^2 g_{\uparrow \beta} = \delta_{\alpha\beta}, \qquad g_{\uparrow \alpha}^{\mathsf{T}} ADA g_{\uparrow \beta} = \lambda_\alpha \delta_{\alpha\beta}, \tag{90a}$$

$$g_{\downarrow \alpha}^{\mathsf{T}} A^{-2} g_{\downarrow \beta} = \delta_{\alpha\beta}, \quad g_{\downarrow \alpha}^{\mathsf{T}} A^{-1} DA^{-1} g_{\downarrow \beta} = \lambda_\alpha \delta_{\alpha\beta}, \tag{90b}$$

and also

$$g_{\uparrow \alpha}^{\mathsf{T}} g_{\downarrow \beta} = \delta_{\alpha\beta}, \qquad g_{\downarrow \alpha}^{\mathsf{T}} g_{\uparrow \beta} = \delta_{\alpha\beta}. \tag{91}$$

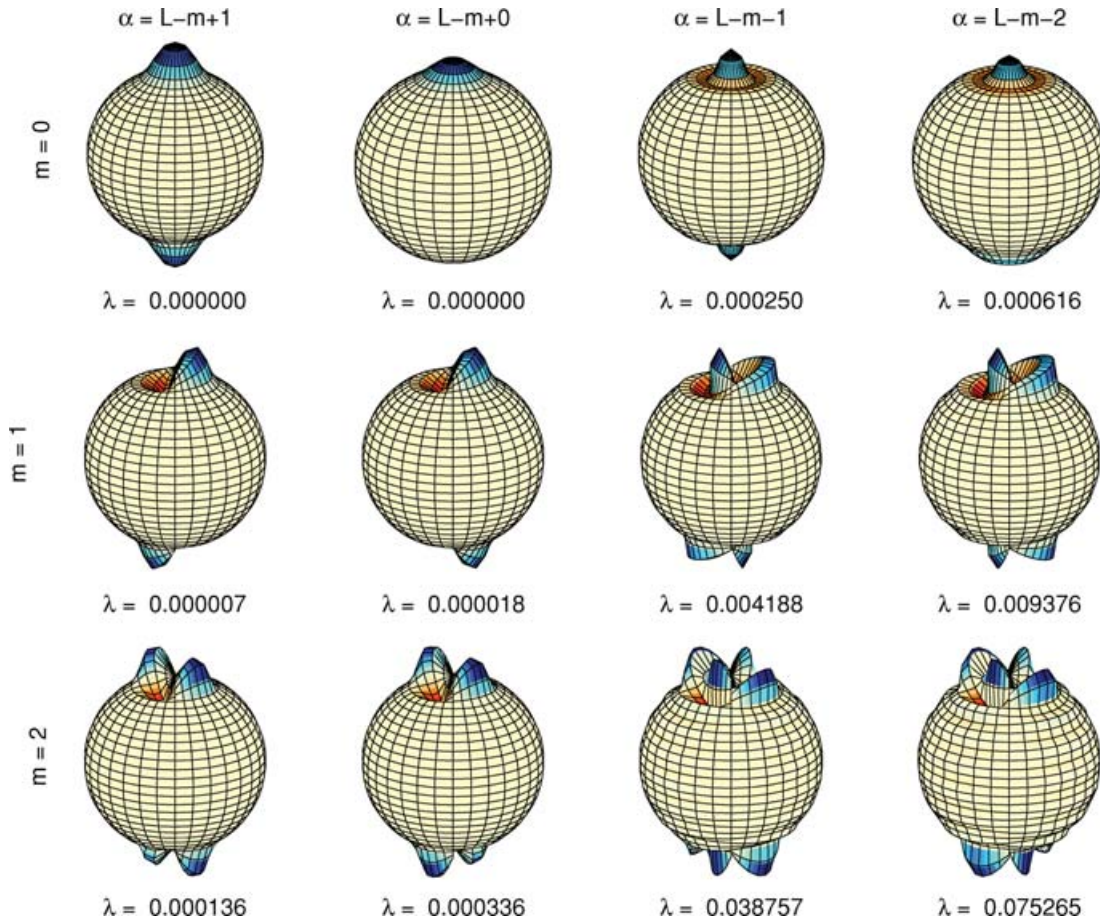**Figure 5.** 3-D spatial dependence of the last four fixed-order, $m = 0 \to 2$, eigenfunctions $g_\alpha(\theta)$, $\alpha = L - m + 1 \to L - m - 2$, bandlimited to $L = 18$, poorly concentrated in the belt $\pm 60°$ about the equator, as expressed by their eigenvalues λ_α. Plot arrangement is as in Fig. 3.
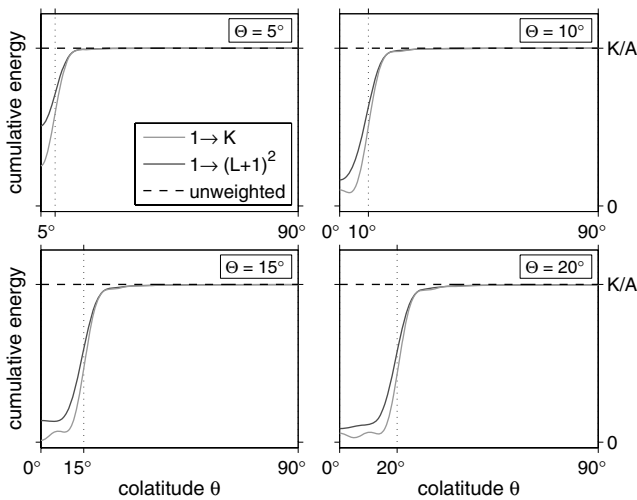

Figure 6. Cumulative energy of $L = 18$ bandlimited eigenfunctions concentrated inside of belts complementary to antipodal polar caps of radius $\Theta = 5°$, $10°$, $15°$ and $20°$. The Shannon numbers are $K = 360$, 356, 349 and 339. The sums of squares $g_1^2(\theta, \phi) + g_2^2(\theta, \phi) + \cdots$ and $\lambda_1 g_1^2(\theta, \phi) + \lambda_2 g_2^2(\theta, \phi) + \cdots$ are plotted versus colatitude θ, along a fixed arbitrary meridian ϕ. Dashed lines show the full unweighted sums of $(L + 1)^2$ terms. Solid lines show the eigenvalue-weighted partial sums of K terms and the full sums of $(L + 1)^2$ terms. All curves are symmetric about the equator.
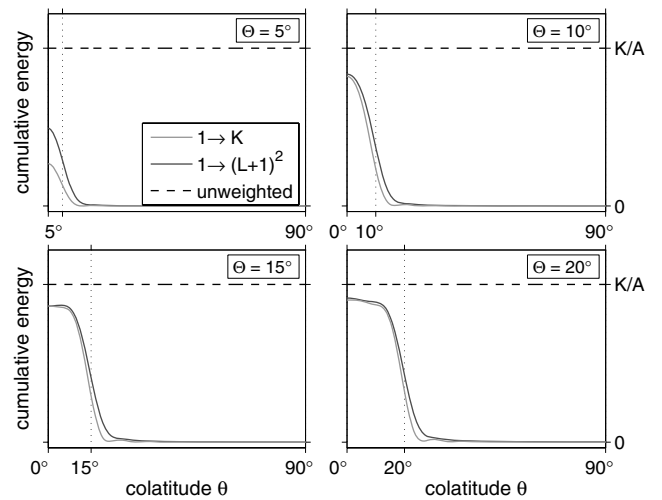
Figure 7. Cumulative energy of $L = 18$ bandlimited eigenfunctions concentrated within circularly symmetric polar caps of colatitudinal radius $\Theta = 5°$, $10°$, $15°$ and $20°$. The Shannon numbers are $K = 1$, 5, 12, 22. The symbols used are identical to those of Fig. 6. The solid lines showing the eigenvalue-weighted partial sums of K terms and full sums of $(L + 1)^2$ terms are very nearly equal, and concentrated uniformly within the pair of antipodal caps $0° \le \theta \le \Theta$ and $180° - \Theta \le \theta \le 180°$. All curves are symmetric about the equator.
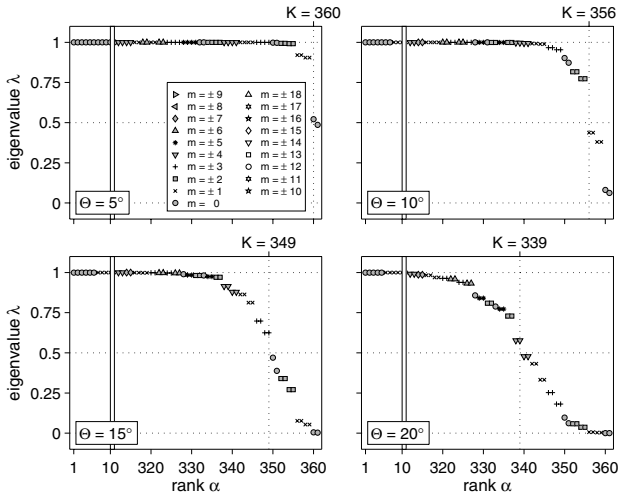
Figure 8. Eigenvalue spectra of the mixed-order $L = 18$ bandlimited operators concentrating within equatorial belts that complement antipodal pairs of axisymmetric caps of radius $\Theta = 5°$, $10°$, $15°$, $20°$. The total number of eigenvalues is $(L + 1)^2 = 361$; only $\lambda_1 \to \lambda_{10}$ and $\lambda_{312} \to \lambda_{361}$ are shown. Different symbols are used for the various orders $-18 \leq m \leq 18$; juxtaposed identical symbols are $\pm m$ doublets. Top labels specify the rounded Shannon numbers $K = 360$, 356, 349 and 339.
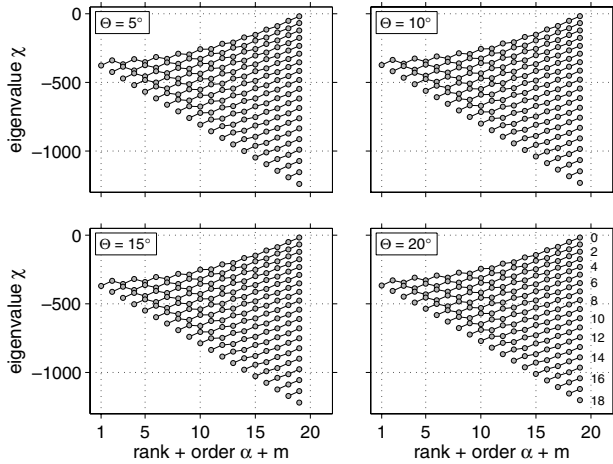
Figure 9. Eigenvalue spectra of the fixed-order $L = 18$ bandlimited Grünbaum operators commuting with the operators whose eigenvalues are shown in Fig. 8. Separate sequences of eigenvalues $\chi_1, \chi_2, \ldots, \chi_{L-m+1}$ for each angular order $0 \leq m \leq L$ are connected by lines. Each sequence is offset horizontally by its order m, and vertically by 50 units per order.

In this matrix notation we repeat eq. (27) as

$$\hat{s}_{\uparrow\alpha} = \sum_{\beta=1}^{(L+1)^2} \left(g_\alpha^\mathsf{T} A^{-1} g_\beta\right) \hat{s}_\beta. \tag{92}$$

We note for future reference that, although the transformation matrix $g_\alpha^\mathsf{T} A^{-1} g_\beta$ may be banded, it is not in general possible to truncate it to circumvent having to calculate the full set of $\hat{s}_\beta$ even if we are only interested in a truncated set of coefficients $\hat{s}_{\uparrow\alpha}$.

Finally, eq. (23) and eqs (86)–(92) can now be combined to prove the equivalent results:

$$\sum_{\alpha=1}^{(L+1)^2} s_\alpha g_\alpha = \sum_{\alpha=1}^{(L+1)^2} s_{\uparrow\alpha} g_{\downarrow\alpha} = \sum_{\alpha=1}^{(L+1)^2} s_{\downarrow\alpha} g_{\uparrow\alpha}, \tag{93}$$

which we use extensively in subsequent sections.

7 POTENTIAL FIELD ESTIMATION

We return to solving the geodetic problem stated in Section 2. We are given noisy data, d, taken by a satellite at an altitude, a, over an incomplete sampling domain, R, and attempting to estimate the potential field, s, that gives rise to these observations, at its source level on the unit sphere, Ω. Although the source field has an infinite bandwidth, we will practically only be able to make bandlimited estimates of it, which we denote by $\hat{s}$. The spectral limitation to the bandwidth L as well as the spatial restriction of the observation domain to the region R motivates our seeking an estimate in terms of a set of basis functions that are spatiospectrally concentrated, rather than using the non-localized spherical harmonics Y_{lm} of more conventional approaches. This new function set is the Slepian basis, g_α, constructed in Sections 3–6 in a variety of geometries, but most notably for the axisymmetric case of a latitudinal belt around the equator, and its complement the double polar cap, representative of the polar gap in geodesy.

That the geodetic estimation problem is essentially a problem of spatiospectral localization can be understood by considering a naive—and in practice unsuitable—estimation scheme. Suppose we construct an estimate in the form of eq. (21),

$$\hat{s} = \sum_{lm}^L \hat{s}_{lm} Y_{lm}, \tag{94}$$

by minimizing its aggregate squared misfit with the data, given by eq. (15), over the sphere. This amounts to solving the variational problem

$$\Phi = \int_R (\hat{s}_\uparrow - d)^2 \, d\Omega = \text{minimum}, \tag{95}$$

where the integration domain is the region R in which observations are available. Substituting eqs (15) and (25) into eq. (95) and requiring the partial derivatives $\partial\Phi/\partial\hat{s}_{lm}$ to vanish yields the condition

$$\int_R \hat{s}_\uparrow Y_{lm} \, d\Omega = \int_R d \, Y_{lm} \, d\Omega, \tag{96}$$

while the result $\partial^2\Phi/\partial\hat{s}_{lm}^2 > 0$ as long as $R \neq 0$ guarantees the convexity of the penalty function Φ. Inserting the representation (25)–(26) into eq. (96) and using the definition of the localization kernel (33) and its inverse (46), the estimate of the field coefficients at source level is given by:

$$\hat{s}_{lm} = (1 + a)^{l+1} \sum_{l'm'}^L D_{lm,l'm'}^{-1} \int_R d \, Y_{l'm'} \, d\Omega. \tag{97}$$

Thus, the estimate depends on the inverse of the localization kernel D. It is, therefore, directly influenced by the size and the shape of the region of missing data, as well as by the chosen bandwidth. Since D tends to have a very low condition number (see, e.g. Fig. 8), finding a stable inverse D^{-1} is problematic: the geodetic inverse problem is ill-conditioned, as is widely advertised even without reference to the localization nature of the problem (Xu 1992a,b).

In the following sections we will derive alternative solutions whose quality we will judge using standard statistical measures (e.g. Cox & Hinkley 1974; Bendat & Piersol 2000). The first will be the average of the squared difference between a single estimate and the mean of all estimates over a set of realizations of the data, the estimation variance:

$$v = \langle(\hat{s} - \langle\hat{s}\rangle)^2\rangle = \langle\hat{s}^2\rangle - \langle\hat{s}\rangle^2. \tag{98}$$

The angular brackets denote averaging over the ensemble of repeated observations, each observation being influenced by a different realization of the random noise. Similarly, we compute the difference between the mean of the estimators and the unknown signal, the estimation bias:

$$b = \langle \hat{s} \rangle - s. \tag{99}$$

We refer to the difference between an estimate and the unknown signal as the estimation error:

$$\epsilon = \hat{s} - s. \tag{100}$$

Finally, the sum of the variance and the squared bias term gives the mean squared error, or mse:

$$\langle \epsilon^2 \rangle = \langle (\hat{s} - s)^2 \rangle = v + b^2. \tag{101}$$

For the moment we regard the unknown source signal s as the unique 'truth', that is, we consider s to be non-stochastic, although the data derived from it are contaminated by stochastic noise, see eqs (15)–(17).

8 SPHERICAL HARMONIC SOLUTION

We have seen that a naive least-squares solution to the geodetic inverse problem in the spherical harmonic basis yields a solution (97) that is dependent on the inverse of the localization matrix and, therefore, in general impossible to stably compute. One of the many approaches to circumvent this difficulty is by adding a model norm to the penalty function (e.g. Hoerl & Kennard 1970a,b; Marquardt 1970; Wiggins 1972; Jackson 1979); eq. (95) only minimized the norm of the data misfit. In this section we discuss the solution to this so-called damped least-squares approach.

8.1 Damped least-squares approach

To stabilize the solution we amend the variational problem of eq. (95) by including a weighted norm of the model outside the observation domain:

$$\int_R (\hat{s}_\uparrow - d)^2 \, d\Omega + \eta \int_{\bar{R}} \hat{s}_\uparrow^2 \, d\Omega = \text{minimum}, \tag{102}$$

where $\eta \geq 0$ is a damping parameter. Retaining the spherical harmonic basis, once again we consider the bandlimited estimate (94) and minimize (102) with respect to the unknown coefficients $\hat{s}_{lm}$. After minimal algebra, involving eqs (25)–(26), (33) and (45)–(46), we obtain the spectral-domain solution,

$$\hat{s}_{lm} = (1 + a)^{l+1} \sum_{l'm'}^{L} (D_{lm,l'm'} + \eta \bar{D}_{lm,l'm'})^{-1}$$
$$\times \int_R d \, Y_{l'm'} \, d\Omega, \tag{103}$$

which only holds at the degrees $l \leq L$, since, when $l > L$, no estimate is available, $\hat{s}_{lm} = 0$. The case where $a = 0$ and $\eta = 1$, for which, from eq. (45), $D + \bar{D} = I$, the identity matrix, was treated in some detail by Sneeuw & van Gelderen (1997). The integral over the data in eq. (103) is made explicit by substituting eq. (15) and using eqs (25) and (33) once again:

$$\int_R d \, Y_{lm} \, d\Omega = \sum_{l'm'}^{\infty} D_{lm,l'm'} s_{\uparrow l'm'} + \int_R n \, Y_{lm} \, d\Omega. \tag{104}$$

Comparing eq. (103) to eq. (97), we now require the inverse of the weighted sum of the operator localizing to R and the complementary operator localizing to the region of missing data $\bar{R}$. The addition of the small, non-diagonal, quantity $\eta \bar{D}$ to the original matrix D improves its condition number. We postpone a discussion on determining the ideal value of the weighting parameter η but it is clear that the estimate of the field coefficients $\hat{s}_{lm}$ in the form of eq. (103) is now computable.

In order to ascertain the statistical properties (eqs 98–101) of the new estimate (eqs 103–104) we first calculate the average of this estimate over all realizations of the noise. From eq. (16), this ensemble averaging of eqs (103)–(104) annihilates the random noise term, and we obtain

$$\langle \hat{s}_{lm} \rangle = (1 + a)^{l+1} \sum_{l'm'}^{L} \left(D_{lm,l'm'} + \eta \bar{D}_{lm,l'm'} \right)^{-1}$$
$$\times \sum_{l''m''}^{\infty} D_{lm',l''m''} s_{\uparrow l''m''}. \tag{105}$$

Again, the coefficients $\hat{s}_{lm}$ are defined only for the degrees $l \leq L$. We note that, were the source signal to be similarly bandlimited, the averages $\langle \hat{s}_{lm} \rangle$ obtained by undamped ($\eta = 0$) estimation would be equal to the true source coefficients s_{lm}. This follows directly from substituting the leftmost term of eq. (13) into eq. (105) and using eqs (14) and (46). The addition of the damping term ($\eta > 0$) biases the estimate away from the truth, hence the name 'biased estimation' (Hoerl & Kennard 1970b). It is the price we pay to be able to calculate the estimate at all.

There are other benefits as well. These are most easily seen by computing a spatial-domain representation of the estimate using the Slepian basis, as in eq. (21). Making use of the equivalence (93), we write for the (bandlimited) estimate

$$\hat{s} = \sum_{\alpha=1}^{(L+1)^2} \hat{s}_{\uparrow \alpha} g_{\downarrow \alpha}, \tag{106}$$

noting that the upward continued coefficients $\hat{s}_{\uparrow \alpha}$ in the Slepian basis are calculated according to eq. (27), and the downward continued Slepian basis functions according to eq. (87). A Slepian basis expansion of the (broadband) observations, combining eqs (13), (15) and (25), is given by

$$d = \sum_{\alpha=1}^{(L+1)^2} s_{\uparrow \alpha} g_\alpha + \sum_{lm>L}^{\infty} s_{\uparrow lm} Y_{lm} + n. \tag{107}$$

This equation allows us to find an alternative expression for the data integral (104), for which we also use eqs (22) and (33), and eq. (30) or the double orthogonality of the Slepian functions (35), namely

$$\int_R d \, Y_{lm} \, d\Omega = \sum_{\alpha=1}^{(L+1)^2} g_{\alpha,lm} \left(\lambda_\alpha s_{\uparrow \alpha} + \int_R n \, g_\alpha \, d\Omega \right)$$
$$+ \sum_{l'm'>L}^{\infty} D_{lm,l'm'} s_{\uparrow l'm'}. \tag{108}$$

Inserting eqs (27), (103) and (108) into eq. (106) and using the expressions (23), (40) and (47) yields the spatial-domain estimate of the field as

$$\hat{s}(\hat{\mathbf{r}}) = \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha^*(\eta) g_{\downarrow \alpha}(\hat{\mathbf{r}})$$
$$\times \left(\lambda_\alpha s_{\uparrow \alpha} + \int_R n \, g_\alpha \, d\Omega + \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm} \right), \tag{109a}$$

$$\lambda_\alpha^*(\eta) = [\lambda_\alpha + \eta(1 - \lambda_\alpha)]^{-1}. \tag{109b}$$

We have introduced the symbol $\lambda_\alpha^*(\eta)$ for notational convenience. In the absence of damping, $\lambda_\alpha^*(0) = \lambda_\alpha^{-1}$, that is, the inverse of the concentration eigenvalue. Here, too, the necessity of damping is readily apparent: as the eigenvalues of the concentration operator, λ_α, become vanishingly small, their inverse grows explosively, inflating the noise term and the term containing the signal at the unmodelled degrees $l > L$, and rendering the stable computation of the estimate (109) impossible. Adding the damping factor to filter the expansion coefficients is an effective way to prevent this.

The vanishing mean of the stochastic noise, eq. (16), guarantees that the ensemble average of the spatial estimate over all realizations of the noise is given by

$$\langle \hat{s}(\hat{\mathbf{r}}) \rangle = \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha^*(\eta) g_{\downarrow\alpha}(\hat{\mathbf{r}})$$
$$\times \left(\lambda_\alpha s_{\uparrow\alpha} + \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm} \right). \tag{110}$$

This equation can be combined with eqs (93) and (20) to confirm that for bandlimited source fields and in the absence of damping, the mean of the bandlimited spatial estimate $\langle \hat{s}(\hat{\mathbf{r}}) \rangle$ is identical to the source field $s(\hat{\mathbf{r}})$, even if the estimate $\hat{s}(\hat{\mathbf{r}})$ is impossible to compute stably without some form of damping.

The introduction of the damping term stabilizes the solution at the cost of added bias. Following eq. (99) the latter is calculated by subtracting the full representation of the signal (20) from eq. (110), making use of eq. (93) and the identity $\lambda^*\lambda - 1 = -\eta(1 - \lambda)\lambda^*$. The spatial estimation bias is then

$$b(\hat{\mathbf{r}}) = -\eta \sum_{\alpha=1}^{(L+1)^2} (1 - \lambda_\alpha)\lambda_\alpha^*(\eta) s_\alpha g_\alpha(\hat{\mathbf{r}}) - \sum_{lm>L}^{\infty} s_{lm} Y_{lm}(\hat{\mathbf{r}})$$
$$+ \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha^*(\eta) g_{\downarrow\alpha}(\hat{\mathbf{r}}) \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm}. \tag{111}$$

In the absence of damping ($\eta = 0$), the first term in this equation vanishes, leaving us with the unavoidable bias due to making bandlimited estimates of broadband fields (the second term) and the broadband leakage (the third term).

An expression for the estimation variance from (98) is obtained by squaring eq. (109) and averaging the result, using the properties of the noise (16)–(17) and eq. (35), and subtracting from the result the square of eq. (110). The spatial estimation variance is

$$v(\hat{\mathbf{r}}) = N \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha \left[\lambda_\alpha^*(\eta) \right]^2 g_{\downarrow\alpha}^2(\hat{\mathbf{r}}). \tag{112}$$

We note that eq. (112) is the only one thus far to assume that the power spectrum of the noise is white, of magnitude N. And one more time, the necessity of the damping is apparent: in its absence, the estimation variance strongly amplifies the measurement noise. At the price of introducing additional bias, damping prevents this.

8.2 A bandlimited white stochastic source

In the previous section we have derived expressions for the average estimate of the spherical harmonic field coefficients, $\langle \hat{s}_{lm} \rangle$, in eq. (105), and for the average of spatial expansions of the estimated field, $\langle s(\hat{\mathbf{r}}) \rangle$, in eq. (110). The averaging was over the different realizations of the stochastic noise process. Both expressions are valid in the most general sense; the only condition being that the average

over all realizations of the noise, $\langle n(\hat{\mathbf{r}}) \rangle$, is zero. No further assumptions are necessary. We have drawn attention to the fact that without the damping term, both estimates are nearly impossible to calculate. However, in that case, they are unbiased when the source signal itself is strictly bandlimited to within a bandwidth L identical to that of the estimate.

We can make this explicit by postulating that the geophysical signal expressed as eq. (12) or eq. (20) has spherical harmonic expansion coefficients that vanish outside of this bandwidth:

$$s_{lm} = 0 \quad \text{for} \quad L < l \leq \infty. \tag{113}$$

We will work with this contrived geophysical signal for the simple reason that no amount of sophistication can cure the fact that forming harmonically truncated estimates leads to multiple bias terms, as can be seen from eq. (111). Under the condition (113), eq. (105) becomes

$$\langle \hat{s}_{lm} \rangle = (1 + a)^{l+1} \sum_{l'm'}^{L} (D_{lm,l'm'} + \eta \bar{D}_{lm,l'm'})^{-1}$$
$$\times \sum_{l''m''}^{L} D_{l'm',l''m''} s_{\uparrow l''m''}, \tag{114}$$

from which, using eqs (14) and (46), we derive immediately that the undamped estimate, given by eq. (103) with $\eta = 0$, which is identical to eq. (97), of the coefficients of a bandlimited source is unbiased:

$$\langle \hat{s}_{lm} \rangle = s_{lm} \quad \text{if} \quad \eta = 0. \tag{115}$$

Similarly, using eq. (93), eq. (110) can be transformed under the same condition (113) into

$$\langle \hat{s}(\hat{\mathbf{r}}) \rangle = \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha \lambda_\alpha^*(\eta) s_\alpha g_\alpha(\hat{\mathbf{r}}), \tag{116}$$

from which, with $\lambda_\alpha^*(0) = \lambda_\alpha^{-1}$, the undamped spatial estimate (94) of such a field is unbiased, as noted earlier:

$$\langle \hat{s}(\hat{\mathbf{r}}) \rangle = s(\hat{\mathbf{r}}) \quad \text{if} \quad \eta = 0. \tag{117}$$

Indeed, under the condition (113), the only term left in the bias equation (111) is directly, though not linearly, dependent on the damping coefficient η: it is

$$b(\hat{\mathbf{r}}) = -\eta \sum_{\alpha=1}^{(L+1)^2} (1 - \lambda_\alpha)\lambda_\alpha^*(\eta) s_\alpha g_\alpha(\hat{\mathbf{r}}). \tag{118}$$

Although we can calculate the mean squared estimation error (101) exactly from eqs (112) and (118), we will gain additional insight when we cease to consider the unknown signal as a non-stochastic signal. The source signal $s(\hat{\mathbf{r}})$, until now, has been considered to be 'given': we have simply assumed it is of the form (12) and attempted to estimate its true unknown coefficients s_{lm} from incomplete and noisy observations. All averaging in the construction of the bias and variance terms was carried out over the different realizations of the noise $n(\hat{\mathbf{r}})$, which we took to be a white stochastic process. By now considering the geophysical signal, as well, to be a stochastic process, we shall calculate the mse after an additional round of averaging, this time over the various realizations of $s(\hat{\mathbf{r}})$, should they be available. Instead of eq. (101) we thus write

$$\langle \epsilon^2 \rangle = v + \langle b^2 \rangle, \tag{119}$$

where the angular brackets now denote an average over the ensemble of signals. Strictly speaking we should write $\langle\langle \epsilon^2 \rangle\rangle$ but we eschew the double brackets in the interest of notational simplicity.

We notice from eq. (118) that to compute $\langle b^2 \rangle$ we shall require the covariance $\langle s_\alpha s_\beta \rangle$ of the expansion coefficients of the field in the Slepian basis. To facilitate the treatment and for easy comparison with the assumed white power spectrum of the noise process, we shall consider a bandlimited source signal that is 'whitish', that is, white within the band $l \leq L$, such that its covariances in the spherical harmonic and Slepian bases, respectively, are given by

$$\langle s_{lm} s_{l'm'} \rangle = S \delta_{ll'} \delta_{mm'}, \tag{120a}$$

$$\langle s_\alpha s_\beta \rangle = S \delta_{\alpha\beta}, \tag{120b}$$

consistently with eqs (23)–(24), while noting that the spatial covariance of this signal is

$$\langle s(\hat{\mathbf{r}}) s(\hat{\mathbf{r}}') \rangle = S D(\hat{\mathbf{r}}, \hat{\mathbf{r}}') \neq S \delta(\hat{\mathbf{r}}, \hat{\mathbf{r}}'), \tag{121}$$

as can be deduced by combining eq. (12) with eq. (120) and using eqs (8)–(9) and (38). A last assumption introduced here is that the noise is wholly uncorrelated with the signal:

$$\langle s(\hat{\mathbf{r}}) n(\hat{\mathbf{r}}') \rangle = 0. \tag{122}$$

The average of the squared bias term (118) under these idealized assumptions is

$$\langle b^2(\hat{\mathbf{r}}) \rangle = \eta^2 S \sum_{\alpha=1}^{(L+1)^2} (1 - \lambda_\alpha)^2 \left[\lambda_\alpha^*(\eta) \right]^2 g_\alpha^2(\hat{\mathbf{r}}), \tag{123}$$

and the mean squared estimation error, following eq. (119), is formed by combining this result with the expression for the variance, eq. (112). The latter expression has remained unchanged even for a stochastic source signal since the noise is uncorrelated with the signal, eq. (122). Thus, the mse of the bandlimited estimation of a bandlimited white source field from incomplete observations at an altitude in the presence of white noise is

$$\langle \epsilon^2(\hat{\mathbf{r}}) \rangle = N \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha \left[\lambda_\alpha^*(\eta) \right]^2 g_{\downarrow\alpha}^2(\hat{\mathbf{r}})$$
$$+ \eta^2 S \sum_{\alpha=1}^{(L+1)^2} (1 - \lambda_\alpha)^2 \left[\lambda_\alpha^*(\eta) \right]^2 g_\alpha^2(\hat{\mathbf{r}}). \tag{124}$$

All $(L + 1)^2$ basis functions are required to form the mse. The first term in the expression for the mse is the variance: it is the only term that depends on the noise. We have seen that in the absence of damping ($\eta = 0$) this term becomes unmanageably large: the addition of damping counteracts this. In addition, the estimation variance also varies with the observation height a above the unit sphere: as a grows, so do the downward continued Slepian basis functions $g_\downarrow(\hat{\mathbf{r}})$, and with them, the noise. The second term in the mse is due to bias. This is the only term that depends on the characteristics of the signal. It is independent of the satellite altitude at which the measurements are taken.

8.3 Optimal damping level

To illustrate the behaviour of the mse (124) we focus on the case where the measurement altitude is $a = 0$, and, from eqs (86)–(87), $g_\alpha = g_{\downarrow\alpha} = g_{\uparrow\alpha}$. This simplifies the expression (124) to:

$$\langle \epsilon^2(\hat{\mathbf{r}}) \rangle = \sum_{\alpha=1}^{(L+1)^2} \mathcal{R}_\alpha(\eta) g_\alpha^2(\hat{\mathbf{r}}), \tag{125a}$$

$$\mathcal{R}_\alpha(\eta) = \left[\lambda_\alpha^*(\eta) \right]^2 \left[N \lambda_\alpha + \eta^2 S (1 - \lambda_\alpha)^2 \right]. \tag{125b}$$

The function $\mathcal{R}_\alpha(\eta)$ combines the effects of data noise, damping, signal strength, and measurement geometry. We will compare the mse with the mean squared signal strength over all realizations, which is given by

$$\langle s^2(\hat{\mathbf{r}}) \rangle = S \frac{(L + 1)^2}{4\pi}. \tag{126}$$

The result (126) is obtained by combining eq. (121) with the definition (38) at $\hat{\mathbf{r}} \cdot \hat{\mathbf{r}} = 1$. We calculate the following two quantities. First, a normalized spatial average of the mse given by the ratio of the mse (125) to the mean squared signal strength (126), both averaged over the entire sphere Ω. Using the orthogonality conditions (35) this 'Ω-average mse' is given by

$$\frac{\int_\Omega \langle \epsilon^2(\hat{\mathbf{r}}) \rangle \, d\Omega}{\int_\Omega \langle s^2(\hat{\mathbf{r}}) \rangle \, d\Omega} = \frac{1}{S(L+1)^2} \sum_{\alpha=1}^{(L+1)^2} \mathcal{R}_\alpha(\eta). \tag{127}$$

Second, a scaled 'R-average mse' is given by the ratio of the same quantities, averaged over the covered region R. Using eq. (35) and the definition (42) of the Shannon number K, it is

$$\frac{\int_R \langle \epsilon^2(\hat{\mathbf{r}}) \rangle \, d\Omega}{\int_R \langle s^2(\hat{\mathbf{r}}) \rangle \, d\Omega} = \frac{1}{KS} \sum_{\alpha=1}^{(L+1)^2} \lambda_\alpha \mathcal{R}_\alpha(\eta). \tag{128}$$

Both quantities are shown in Fig. 10, for a double-cap polar gap of $\Theta = 10°$ and a bandwidth $L = 45$. They are plotted in different panels for different signal-to-noise ratios $S/N = 4, 6, 8$ and 10 as functions of the damping parameter $\eta = 0 \rightarrow 1$. We show eq. (127) in black, with the scale on the left of the panels, and eq. (128) in grey, with the scale on the right hand side. The range of Ω-average mse values shown is much larger (5 per cent in all four panels) than the
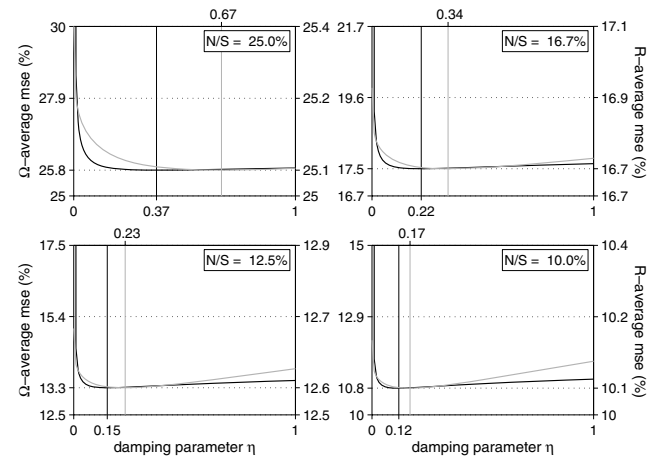
Figure 10. Spatially averaged mse (as a percentage of the average mean squared signal strength) of the damped-least-squares spherical harmonic solution to the geodetic estimation problem for a bandlimited white signal and white noise. The Ω-average mse (black curves and left ordinate) is the average over the entire sphere; the R-average mse (grey curves and right ordinate) is the average over the region of observation, the equatorial belt complementary to the polar gap of radius $\Theta = 10°$. The bandwidth of the signal and its estimate is $L = 45$. The measurement altitude is $a = 0$. The signal-to-noise levels shown are $S/N = 4, 6, 8$ and 10. We plot the normalized average mse values as a function of the damping parameter η, and indicate by vertical lines the values of η that minimize them. The range of the R-average is much reduced compared to the Ω-average values. Both ordinates are truncated below at N/S, the mse value when the observation region is the entire sphere, $R = \Omega$.

equivalent range in R-average mse values (0.4 per cent in all panels): the effects of damping on the overall mse over the entire globe are much more pronounced than its effects on the mse averaged over the region in which data were collected. The ordinate is truncated to aid the visualization. The maximum R-average mse is $(N/S)/\cos\Theta$ which is attained when $\eta = 0$. This can be verified by noting that $\mathcal{R}_\alpha(0) = N\lambda_a^{-1}$, using the definition of the Shannon number (42), and noting that the area of the covered region is $A = 4\pi \cos \Theta$. Thus, at a given signal-to-noise ratio only the size of the polar gap controls the upper bound on the R-average mse. A lower bound for all damping levels is found at full coverage, $R = \Omega$. It thus applies to both measures of the average mse. Indeed without a polar gap, $\Theta = 0°$, $K = (L+1)^2$, $\lambda_\alpha^*(\eta) = \lambda_\alpha = 1$, $\mathcal{R}_\alpha(\eta) = N$, and the scaled average mse curves never drop below N/S, which we use as a lower cut-off for the vertical axes.

A statistically desirable estimator (e.g. Cox & Hinkley 1974; Bendat & Piersol 2000) is one that is unbiased and efficient, that is, it minimizes the mean squared estimation error. We have seen that sacrificing the unbiasedness by introducing damping removes the obstacles in computing the estimate in the first place, and reduces the estimation variance. We can calculate the damping level that is overall optimal by minimizing the mse (125) with respect to the damping parameter η. However, minimization of the R-average and Ω-average mse will yield slightly different optima. Minimizing eq. (127), the normalized mse over the entire sphere, we obtain an optimal damping coefficient η_Ω given by

$$\eta_\Omega = \frac{N}{S} \frac{\sum_{\alpha=1}^{(L+1)^2} \left[\lambda_\alpha^*(\eta_\Omega)\right]^3 \lambda_\alpha(1-\lambda_\alpha)}{\sum_{\alpha=1}^{(L+1)^2} \left[\lambda_\alpha^*(\eta_\Omega)\right]^3 \lambda_\alpha(1-\lambda_\alpha)^2}. \qquad (129)$$

Likewise, minimization of eq. (128), the normalized mse over the region of coverage, yields an optimal damping level η_R given by

$$\eta_R = \frac{N}{S} \frac{\sum_{\alpha=1}^{(L+1)^2} \left[\lambda_\alpha^*(\eta_R)\right]^3 \lambda_\alpha^2(1-\lambda_\alpha)}{\sum_{\alpha=1}^{(L+1)^2} \left[\lambda_\alpha^*(\eta_R)\right]^3 \lambda_\alpha^2(1-\lambda_\alpha)^2}. \qquad (130)$$

Although the unknown optimal damping levels η_Ω and η_R appear on both sides of eqs (129) and (130), their values can be easily computed by iteration. They depend on the measurement geometry, the damping, and the signal-to-noise ratio. In Fig. 10, η_Ω and η_R are shown as black and grey vertical lines, respectively. At high signal-to-noise ratios both coefficients can be approximated as $\eta_R = \eta_\Omega \approx N/S \ll 1$.

When the coverage region is axisymmetric the mse (125) is independent of the longitude, as can be deduced from eq. (60). Thus, in Fig. 11 we plot $\langle\epsilon^2(\theta)\rangle/\langle s^2(\theta)\rangle$, in per cent, for different signal-to-noise ratios, as a function of colatitude and for various damping levels: that is, in the undamped ($\eta = 0$), heavily damped ($\eta = 1$) and optimally damped case ($\eta = \eta_\Omega$). The vertical axes are truncated at 100 per cent as the undamped values exceed this value by many orders of magnitude.

9 SLEPIAN BASIS SOLUTION

In the previous section we expanded the estimate of the signal into a bandlimited spherical harmonic basis and performed a damped least-squares inversion for the unknown coefficients. This estimation procedure resulted in a biased estimate, but the damping prevented the detrimental amplification of the measurement noise. We derived expressions for the optimal level of damping required for 'whitish' signals measured at zero altitude. Adding a small amount of bias made the estimate computable and reduced its variance. We used
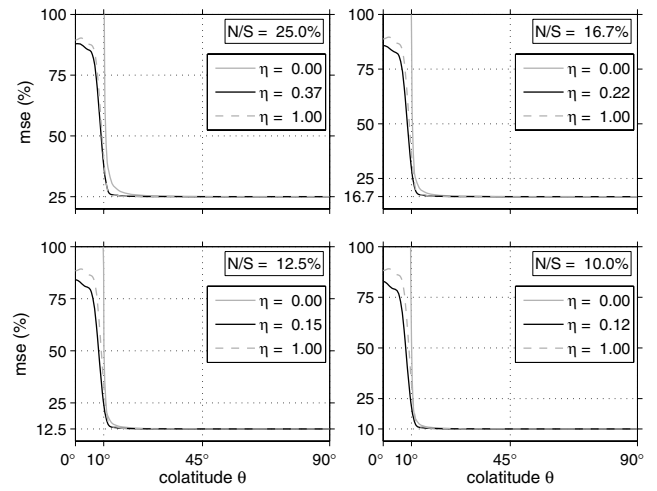
Figure 11. Mean squared error (as a percentage of the mean squared signal strength) of the damped-least-squares spherical harmonic solution to the geodetic estimation problem for a bandlimited white signal and white noise. As in Fig. 10, the signal-to-noise levels shown are $S/N = 4, 6, 8$ and 10, the measurement altitude $a = 0$, the polar gap consists of caps of radius $\Theta = 10°$, and the bandwidth of the signal and its estimate is $L = 45$. As a function of colatitude, for an arbitrary longitude, we plot the mse of the undamped solution ($\eta = 0$), of a heavily damped solution ($\eta = 1$), and of the solution at the damping level which minimizes the normalized average mse over the unit sphere, that is, for the values $\eta_\Omega = 0.37, 0.22, 0.15,$ and $0.12,$ that were marked by black vertical lines in Fig. 10. The mse is symmetric about the equator. The ordinate is truncated at 100 per cent; the mse of the undamped solution in the region of the polar gap exceeds this value by several orders of magnitude.

the (downward continued) Slepian basis to find expressions for the resultant damped spherical harmonic estimate in the spatial domain and to find its bias, variance, and mse. Using the Slepian basis greatly simplified the expressions because of the fact that, as opposed to the spherical harmonics, the Slepian functions are orthogonal over both the entire sphere and the closed subdomains over which, by design, their energy is optimally concentrated.

Alternatively, we might have sought an estimate that is expressed in the spherical Slepian basis at the outset. As we have seen, the first K Slepian eigenfunctions, where K is the Shannon number (eq. 42), provide an excellent coverage of the region of observation. This implies that their associated eigenvalues λ are close to unity, avoiding any problems with their inversion. In this section we will explore the effect on the geodetic solution of using a truncated Slepian basis, consisting of the J basis functions that are best concentrated over the region of satellite observation. Even if $J = K$ appears to be a natural choice, we will determine the truncation level J by optimization of the mean squared estimation error, as we did to find the optimal damping parameter in the damped least-squares spherical harmonic approach.

9.1 Truncated Slepian function approach

The original undamped problem posed in eq. (95),

$$\int_R (\hat{s}_\uparrow - d)^2 \, d\Omega = \text{minimum}, \qquad (131)$$

is now solved by expanding the estimate in the downward continued truncated Slepian basis

$$\hat{s} = \sum_{\alpha=1}^{J} \hat{s}_{\uparrow\alpha} g_{\downarrow\alpha}, \tag{132}$$

and minimizing eq. (131) with respect to the expansion coefficient $\hat{s}_{\uparrow\alpha}$. The second derivative of eq. (131) is always positive. After minimal algebra, using eq. (25) and the double orthogonality (35), the expansion coefficients in eq. (132) are obtained as

$$\hat{s}_{\uparrow\alpha} = \lambda_{\alpha}^{-1} \int_{R} d\, g_{\alpha}\, d\Omega. \tag{133}$$

This result can alternatively be derived by substituting eq. (97) of the undamped spherical harmonic approach into eq. (27) and using eqs (22)–(23) and (47).

We purposely chose an estimate of the form (132) to find the truncated expansion coefficients $\hat{s}_{\uparrow\alpha}$ in their upward continued form and multiplying the downward continued Slepian functions $g_{\downarrow\alpha}$, rather than simply expressing eq. (97) in the Slepian basis g_{α}. In the latter case, as can be readily verified by combining eq. (97) with eqs (24), (22) and (47), every one of the expansion coefficients $\hat{s}_{\alpha}$ would depend on a linear combination of all $(L+1)^2$ terms $\lambda_{\beta}^{-1}\int_{R} d\, g_{\beta}\, d\Omega$ through a matrix term $g_{\alpha}^{\mathsf{T}} A\, g_{\beta}$ whose kind we have encountered in eq. (92). This would, therefore, invalidate the method of truncation as a means to avoid the difficult-to-compute and unnecessarily influential large inverse eigenvalues. By choosing the representation (132) instead, we take advantage of eq. (93) to juxtapose upward and downward continuation, thereby cancelling their effect altogether: eq. (133) shows that every coefficient $\hat{s}_{\uparrow\alpha}$ only depends on the inverse eigenvalue at the same rank α. The effect of the measurement at altitude has not disappeared: it is now contained in eq. (132), per the basis $g_{\downarrow\alpha}$, of which only the first J functions are required. These are calculated via eq. (87) and ultimately, by the stable Grünbaum algorithm central to our analysis.

The data integral in eq. (133) can be calculated by substituting into it eqs (107), (35), (22), (33) and (40), to yield

$$\int_{R} d\, g_{\alpha}\, d\Omega = \lambda_{\alpha} s_{\uparrow\alpha} + \int_{R} n\, g_{\alpha}\, d\Omega + \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm}. \tag{134}$$

Averaging the expressions (133)–(134) over many estimates annihilates the influence of the random noise by virtue of eq. (16),

$$\langle \hat{s}_{\uparrow\alpha} \rangle = s_{\uparrow\alpha} + \lambda_{\alpha}^{-1} \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm}. \tag{135}$$

Combining eqs (132)–(134) yields the space-domain estimate,

$$\hat{s}(\hat{\mathbf{r}}) = \sum_{\alpha=1}^{J} \lambda_{\alpha}^{-1} g_{\downarrow\alpha}(\hat{\mathbf{r}})$$
$$\times \left(\lambda_{\alpha} s_{\uparrow\alpha} + \int_{R} n\, g_{\alpha}\, d\Omega + \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm} \right), \tag{136}$$

which, reassuringly, amounts to the truncated but undamped version of eq. (109), with $\eta = 0$. As before we can eliminate the noise term by averaging over many realizations, to obtain

$$\langle \hat{s}(\hat{\mathbf{r}}) \rangle = \sum_{\alpha=1}^{J} g_{\downarrow\alpha}(\hat{\mathbf{r}}) \left(s_{\uparrow\alpha} + \lambda_{\alpha}^{-1} \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm} \right). \tag{137}$$

The estimation bias, following eq. (99), is obtained by subtracting from eq. (137) the representation of the signal (20) and using the

equivalence (93),

$$b(\hat{\mathbf{r}}) = -\sum_{\alpha>J}^{(L+1)^2} s_{\alpha} g_{\alpha}(\hat{\mathbf{r}}) - \sum_{lm>L}^{\infty} s_{lm} Y_{lm}(\hat{\mathbf{r}})$$
$$+ \sum_{\alpha=1}^{J} \lambda_{\alpha}^{-1} g_{\downarrow\alpha}(\hat{\mathbf{r}}) \sum_{lm>L}^{\infty} h_{\alpha,lm} s_{\uparrow lm}. \tag{138}$$

Without truncation of the Slepian basis function set, that is, when $J = (L+1)^2$, the first term in this equation vanishes. The remaining contributions arise due to forming bandlimited estimates of broadband signals, leading to unavoidable broadband bias (the second term) and leakage (the third term).

Comparing eqs (111) and (138) we discover the parallel roles of damping and truncation. The introduction of the damping parameter η added an extra bias term to the expression (111), and reduced the size of the leakage term by which the coefficients h_{lm} of eq. (40) make the influence of the signal outside the bandwidth, $l > L$, felt, but it was powerless against the bias due to the bandlimited approximation of the broadband signal, which is simply that portion of the signal that is outside the bandwidth L. Similarly, increasing the Slepian truncation level by the reduction of J from $(L+1)^2$ in eq. (138) introduces a new term in the expression for the estimation bias, and reduces the effect of the leakage term containing the coefficients h_{lm}, but it is again no match for the remaining broadband bias from the bandlimitation of the estimate.

An expression for the estimation variance, eq. (98), is obtained by squaring and averaging eq. (136), using the noise properties (16)–(17) and the orthogonality of the Slepian basis functions (35), and subtracting the square of (137). The resulting variance is

$$v(\hat{\mathbf{r}}) = N \sum_{\alpha=1}^{J} \lambda_{\alpha}^{-1} g_{\downarrow\alpha}^{2}(\hat{\mathbf{r}}). \tag{139}$$

This expression is again the first in this section in which we have used the white noise assumption, and once again it will be valid even if the source signal is considered stochastic as long as eq. (122) holds. Comparison of the variance expression in this truncated Slepian basis approach with eq. (112) obtained via the damped spherical harmonics method validates our approach. Without damping, when $\eta = 0$ in eq. (112), or without truncation, $J = (L+1)^2$ in eq. (139), both expressions are identical. Much like the damping term, the truncation of the basis set to its first $J \leq (L+1)^2$ elements reduces the estimation variance by checking the growth of the terms λ_{α}^{-1}. The more severe the truncation, the lower J, and the lower the variance becomes.

9.2 A bandlimited white stochastic source

We again focus on geophysical signals that are bandlimited as in eq. (113). This transforms eq. (135) into

$$\langle \hat{s}_{\uparrow\alpha} \rangle = s_{\uparrow\alpha}, \tag{140}$$

illustrating the fact that an estimate of the form (133) is spectrally unbiased. Just as our analysis of the damped spherical harmonic method showed that for bandlimited source fields, the undamped estimate of eq. (97) is incomputable due to the ill-conditioning of D^{-1}, but unbiased, as shown by eq. (115), we have now shown that the untruncated, that is, $\alpha = 1 \rightarrow (L+1)^2$, Slepian basis estimate of eq. (133) is incomputable due to the growth of the eigenvalues λ^{-1}, although it, too, is unbiased, as shown by eq. (140). The damping term in eq. (103) made the estimate computable but biased, just as

the truncation of the eigenvalues in eq. (132) prevents the blow-up of their inverse, at the cost of added bias.

In the spatial domain, using eqs (137) and (93), the average over all estimates is then

$$\langle \hat{s}(\hat{\mathbf{r}}) \rangle = \sum_{\alpha=1}^{J} s_\alpha g_\alpha(\hat{\mathbf{r}}). \tag{141}$$

In the absence of truncation the spatial estimate of the form (132) is unbiased for bandlimited source fields:

$$\langle \hat{s}(\hat{\mathbf{r}}) \rangle = s(\hat{\mathbf{r}}), \quad \text{if} \quad J = (L+1)^2, \tag{142}$$

which we may again compare to the unbiasedness (117) of the un-damped estimate (94). Explicitly, under the condition (113), the only contributing term in eq. (138) is given by

$$b(\hat{\mathbf{r}}) = - \sum_{\alpha>J}^{(L+1)^2} s_\alpha g_\alpha(\hat{\mathbf{r}}). \tag{143}$$

Its magnitude decreases with increasing J, and vanishes altogether when $J = (L+1)^2$. It can be compared to eq. (118), which vanishes for $\eta = 0$. The average over all realizations of the signal of the squared bias, for a 'whitish' signal with covariance (120), is given by

$$\langle b^2(\hat{\mathbf{r}}) \rangle = S \sum_{\alpha>J}^{(L+1)^2} g_\alpha^2(\hat{\mathbf{r}}), \tag{144}$$

which should be compared with the corresponding eq. (123) in the damped spherical harmonic case. From this and eq. (139) we can calculate the mean squared estimation error following eq. (119),

$$\langle \epsilon^2(\hat{\mathbf{r}}) \rangle = N \sum_{\alpha=1}^{J} \lambda_\alpha^{-1} g_{\downarrow\alpha}^2(\hat{\mathbf{r}}) + S \sum_{\alpha>J}^{(L+1)^2} g_\alpha^2(\hat{\mathbf{r}}). \tag{145}$$

The mse of the untruncated Slepian basis approach and that of the undamped spherical harmonic estimation method (124) are identical. This of course is a direct consequence of the fact that both bases are related to each other by the orthonormal transformation eqs (22)–(23). Of note is the very different form of the damped and truncated expressions, eqs (124) and (145), for the mse. Whereas eq. (124) consists of a weighted sum of all $\alpha = 1 \rightarrow (L+1)^2$ basis functions in a manner that appears to mix the influence of the noise, the damping, and the signal, the truncated expression (145) has dis-entangled the effects of the noise and the signal by distributing the influence of the variance over the basis functions $\alpha = 1 \rightarrow J$ that are well concentrated inside the measurement area, and the effect of the bias over those $\alpha = J + 1 \rightarrow (L+1)^2$ that are confined to the region of missing data. To the one missing piece, the decision on where to truncate the data by the choice of J, we now turn.

9.3 Optimal truncation level

To illustrate the behaviour of the mse in eq. (145) we again focus on the zero-altitude case, for which

$$\langle \epsilon^2(\hat{\mathbf{r}}) \rangle = N \sum_{\alpha=1}^{J} \lambda_\alpha^{-1} g_\alpha^2(\hat{\mathbf{r}}) + S \sum_{\alpha>J}^{(L+1)^2} g_\alpha^2(\hat{\mathbf{r}}). \tag{146}$$

In order to find the optimal truncation level, we consider the full-sphere and coverage-domain average mse (normalized by the cor-responding quadratic signal averages) as in the damped spherical

harmonic approach. Using eqs (126) and (35) we find from eq. (146) that the Ω-average mse in the truncated Slepian case is

$$\frac{\int_\Omega \langle \epsilon^2(\hat{\mathbf{r}}) \rangle \, d\Omega}{\int_\Omega \langle s^2(\hat{\mathbf{r}}) \rangle \, d\Omega} = 1 - \frac{J}{(L+1)^2} + \frac{N}{S} \sum_{\alpha=1}^{J} \frac{\lambda_\alpha^{-1}}{(L+1)^2}. \tag{147}$$

Using the definition (42) of the Shannon number, we likewise find the R-average mse,

$$\frac{\int_R \langle \epsilon^2(\hat{\mathbf{r}}) \rangle \, d\Omega}{\int_R \langle s^2(\hat{\mathbf{r}}) \rangle \, d\Omega} = \frac{N}{S} \frac{J}{K} + \frac{1}{K} \sum_{\alpha>J}^{(L+1)^2} \lambda_\alpha. \tag{148}$$

Both quantities are plotted in Fig. 12 for different signal-to-noise ra-tios $S/N = 4, 6, 8$ and 10 and with the other parameters unchanged from those of Fig. 10: a double-cap polar gap of radius $\Theta = 10°$ and a bandwidth $L = 45$. In black, with the scale on the left of the panel, we show eq. (147) as a function of the truncation level ranging over $J = K - 50 \rightarrow (L+1)^2$. The abscissa is inverted since $J = (L+1)^2$ corresponds to a situation without Slepian truncation, and as J decreases, the degree of truncation increases. In grey, we plot eq. (148), with a much reduced scale on the right hand side. The range of Ω-average mse values shown is much larger (5 per cent in all four panels) than the equivalent range in R-average mse val-ues (which varies from panel to panel but is always smaller than 0.8 per cent): the effects of truncation on the overall mse over the entire globe are much more outspoken than its effects on the mse av-eraged over the region in which data were collected. This behaviour mimics the one seen in Fig. 10 for the damped spherical harmonic case. The ordinate is again truncated for clarity. The value of the untruncated R-average mse, when $J = (L+1)^2$, is $(N/S)/\cos\Theta$.
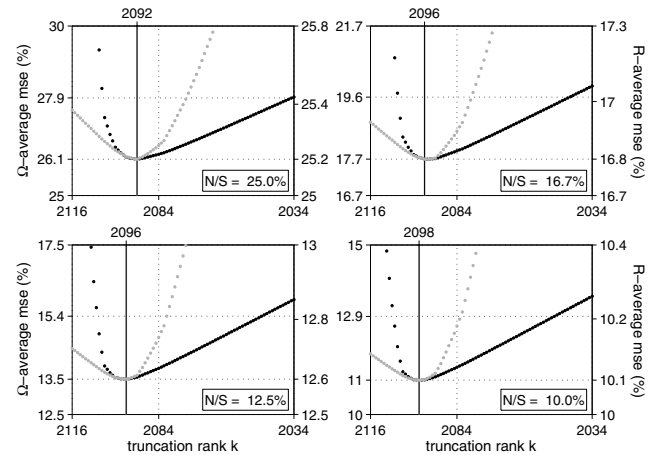
Figure 12. Spatially averaged mean squared error (as a percentage of the average mean squared signal strength) of the truncated Slepian function so-lution to the geodetic estimation problem for a bandlimited white signal and white noise. The Ω-average mse (black curves, left ordinate) is the average over the entire sphere; the R-average mse (grey curves, right ordinate) is the average over the observation region, the equatorial belt complementary to the polar gap of radius $\Theta = 10°$. The bandwidth of the signal and its estimate is $L = 45$. The measurement altitude is $a = 0$. The signal-to-noise levels shown are $S/N = 4, 6, 8$ and 10. We plot the average mse values as a func-tion of the truncation rank J, and indicate by solid vertical lines the values $J_\Omega = J_R$ that minimize them, and the Shannon number $K = 2084$ by the dotted vertical line. The range of the R-average is much reduced compared to the Ω-average values. Both ordinates are truncated below at N/S, the value when the observation region is the entire sphere, $R = \Omega$. The abscissa shows truncation levels from $J = K - 50$ up to the untruncated case, $J = (L+1)^2$. In that case the mse values are identical to those of the undamped ($\eta = 0$) spherical harmonic case shown in Fig. 10.

This follows from the definition of the Shannon number (42) and the area of the covered region, $A = 4\pi \cos \Theta$, and is identical to the corresponding value in the undamped spherical harmonic case. A lower bound is found at full coverage, when $R = \Omega$, $\Theta = 0°$, $K = (L + 1)^2$ and $\lambda_\alpha = 1$. In that case the minimal scaled mse, attained when $J = (L + 1)^2$, equals N/S, as it does in the damped spherical harmonic case. We use this value as a lower cut-off of the vertical axes on the left and the right.

We may obtain the truncation level that minimizes the Ω-average mse by minimizing eq. (147) with respect to J. This will yield an optimal truncation level J_Ω. Likewise, minimizing eq. (148) returns the truncation value J_R at which the R-average mse is minimal. Both minimization problems result in identical constraints on the eigenvalue of the Jth eigenfunction beyond which we truncate,

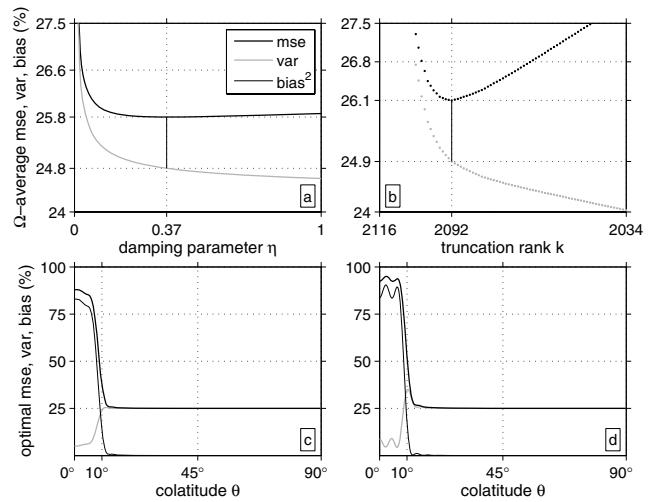$$\lambda_{J_\Omega} = \lambda_{J_R} \approx \frac{N}{S}, \tag{149}$$

which is implicit but solvable. In Fig. 12, the values $J_\Omega = J_R$ identified by the top labels, are shown as a single solid black vertical line; the Shannon number, $K = 2048$, is shown by the dotted black line and the bottom labels.

The mse (146) for axisymmetric coverage regions is independent of the longitude, as can be understood from eq. (60). Thus, in Fig. 13 we plot $\langle \epsilon^2(\theta) \rangle / \langle s^2(\theta) \rangle$, in per cent, as a function of colatitude for various truncation levels: the untruncated, $J = (L + 1)^2$, and optimally truncated cases, $J = J_\Omega = J_R$, and the case truncated at the Shannon number $J = K$. The vertical axes are truncated at 100 per cent since the untruncated values exceed this value by many orders of magnitude.

Finally, in Fig. 14, we plot the relative contributions of variance and bias for both the damped spherical harmonic and the truncated Slepian case. The top panels show the mse, the variance and the

Figure 14. Mean squared error (mse), variance and bias for the damped least-squares solution and the truncated Slepian approach. The antipodal polar gap has a radius $\Theta = 10°$; the bandwidth is $L = 45$; signal-to-noise ratio $S/N = 4$. Panels a & b show the values averaged over the unit sphere, as in Figs 10 and 12; the squared bias is the difference between the mse and variance curves, indicated by the thin black vertical line. Panels c & d show the values at the optimal damping and truncation levels. The minimization of the mse reflects the trade-off between variance, which dominates the covered regions of the polar gap, and bias, dominant in the uncovered equatorial belt region.

squared bias, according to the relation (119), as a function of the damping level (top left) or the truncation rank (top right). The bottom panels show the breakdown of mse, variance and bias in the spatial domain. For both estimation methods, the bias is predominantly concentrated in the areas over which no measurements are available, where it is generated by the power of the 'missing signal'. The variance, on the other hand, arises in the areas of coverage and is influenced by the power of the noise. Both estimation methods show very similar results. Over the covered area, the mse is nearly identical, and in the uncovered region, the mse of the truncated Slepian case approaches that of the damped spherical harmonic case, but it is slightly higher. It is remarkable that eqs (125) and (146), despite their different form, both give rise to a nearly complete spatial separation of bias and variance. Only in eq. (146) is this separation immediately obvious by inspection: signal strength, and thus bias, affect the low-ranking Slepian functions, whose power is mostly concentrated inside of the polar gap, whereas noise, and thus variance, affect the high-ranking Slepian functions whose power is localized to the area of satellite coverage.

10 CONCLUSIONS

Spherical Slepian functions provide a natural solution to the problem of having a polar gap in the satellite coverage of planetary gravitational or magnetic fields. Indeed, the ill-posed estimation problem of finding a source-level potential from noisy observations taken at an altitude over an incomplete region of coverage has natural connections to Slepian's spherical problem of spatiospectral localization. We have proposed a new method that expands the source field in terms of a truncated basis set of spherical Slepian functions, and compared its statistical performance with the traditional damped least-squares method in the spherical harmonic basis. The optimally truncated Slepian method performs nearly as well as the optimally

Figure 13. Mean squared error (as a percentage of the mean squared signal strength) of the truncated Slepian function solution to the geodetic estimation problem for a white signal and white noise. As in Fig. 12, the signal-to-noise levels shown are $S/N = 4$, 6, 8 and 10, the measurement altitude $a = 0$, the polar gap consists of caps with radius $\Theta = 10°$, and the bandwidth of the signal and its estimate is $L = 45$. As a function of colatitude, for an arbitrary longitude, we plot the mse of the untruncated solution, when $J = (L + 1)^2$, of truncation at the Shannon number, $J = K$, and of the truncation levels which minimize the mse, that is, $J_\Omega = J_R = 2092$, 2096, 2096, and 2098, that were marked by black vertical lines in Fig. 12. The mse is symmetric about the equator. The ordinate is truncated at 100 per cent; the mse of the untruncated ($J = 2116$) solution in the region of the polar gap exceeds this value by several orders of magnitude.

damped spherical harmonic method, but it has the significant advantage of an intuitive separation of the estimation bias and variance over those Slepian functions sensitive to the uncovered and covered regions, respectively. The construction of Slepian functions over axisymmetric domains such as the latitudinal belt or its complement, the polar gap, previously dismissed as computationally unstable, has been shown to be eminently tractable. We have shown that the operator that bandlimits a field on the unit sphere and projects it onto the polar caps commutes with a Sturm-Liouville operator. Its eigenfunctions, the Slepian functions, can be computed extremely accurately and efficiently by diagonalizing a tridiagonal matrix with analytically prescribed elements. The gains in ease, speed, and accuracy thus achieved makes the use of spherical Slepian functions in earth and planetary potential-field estimation practical, as our examples have shown.

ACKNOWLEDGMENTS

REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology,* 1st edn, Freeman, San Francisco, California.

Albertella, A., Sansò, F. & Sneeuw, N., 1999. Band-limited functions on a bounded spherical domain: the Slepian problem on the sphere, *J. Geodesy,* **73,** 436–447.

Bendat, J.S. & Piersol, A.G., 2000. *Random Data: Analysis and Measurement Procedures,* 3rd edn, John Wiley, New York.

Blakely, R.J., 1995. *Potential Theory in Gravity and Magnetic Applications,* Cambridge Univ. Press, New York.

Cox, D.R. & Hinkley, D.V., 1974. *Theoretical Statistics,* Chapman and Hall, London, UK.

Dahlen, F.A. & Tromp, J., 1998. *Theoretical Global Seismology,* Princeton Univ. Press, Princeton, New Jersey.

Edmonds, A.R., 1996. *Angular Momentum in Quantum Mechanics,* Princeton Univ. Press, Princeton, New Jersey.

Gilbert, E.N. & Slepian, D., 1977. Doubly orthogonal concentrated polynomials, *SIAM J. Math. Anal.,* **8**(2), 290–319.

Grünbaum, F.A., Longhi, L. & Perlstadt, M., 1982. Differential operators commuting with finite convolution integral operators: some non-Abelian examples, *SIAM J. Appl. Math.,* **42**(5), 941–955.

Gubbins, D., 1983. Geomagnetic field analysis—I. stochastic inversion, *Geophys. J. R. astr. Soc.,* **73**(3), 641–652.

Hinshaw, G. *et al.*, 2003. First-year Wilkinson microwave anisotropy probe (WMAP) observations: the angular power spectrum, *Astroph. J. Supp. Ser.,* **148,** 135–159.

Hoerl, A.E. & Kennard, R.W., 1970a. Ridge regression: biased estimation for nonorthogonal problems, *Technom.,* **12**(1), 55–67.

Hoerl, A.E. & Kennard, R.W., 1970b. Ridge regression: applications to nonorthogonal problems, *Technom.,* **12**(1), 69–82.

Holme, R. & Bloxham, J., 1995. Alleviation of the Backus effect in geomagnetic field modelling, *Geophys. Res. Lett.,* **22**(13), 1641–1644.

Holme, R. & Bloxham, J., 1996. The treatment of attitude errors in satellite geomagnetic data, *Phys. Earth planet. Inter.,* **98**(3–4), 221–233.

Jackson, D.D., 1979. The use of a priori data to resolve non-uniqueness in linear inversion, *Geophys. J. R. astr. Soc.,* **57,** 137–157.

Jordan, T.H. & Minster, J.-B., 1972, Application of a stochastic inverse to the geophysical inverse problem, in *Mathematics of Profile Inversion,* pp. 736–747, ed. Colin, L., no. X-62150 in NASA Tech. Mem., NASA Ames Research Center, Moffett Field.

Lemoine, F.G., Pavlis, N.K., Kenyon, S.C., Rapp, R.H., Pavlis, E.C. & Chao, B.F., 1998. New high-resolution model developed for Earth's gravitational field, *EOS, Trans. Am. geophys. Un.*, **79**(9), 113–118.

Lesur, V., 2006. Introducing localized constraints in global geomagnetic field modelling, *Earth Planets Space,* **58**(4), 477–483.

Lowes, F.J., 1974. Spatial power spectrum of the main geomagnetic field and extrapolation to core, *Geophys. J. R. astr. Soc.,* **36**(3), 717–730.

Mallat, S., 1998. *A Wavelet Tour of Signal Processing,* Academic Press, San Diego, California.

Marquardt, D.W., 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation, *Technom.,* **12,** 591–612.

Pail, R., Plank, G. & Schuh, W.-D., 2001. Spatially restricted data distributions on the sphere: the method of orthonormalized functions and applications, *J. Geodesy,* **75,** 44–56.

Peebles, P.J.E., 1973. Statistical analysis of catalogs of extragalactic objects. I. Theory, *Astroph. J.,* **185,** 413–440.

Percival, D.B. & Walden, A.T., 1993. *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques,* Cambridge Univ. Press, New York.

Santo, A.G. *et al.*, 2001. The MESSENGER mission to Mercury: spacecraft and mission design, *Planet. Space Sc.,* **49**(14–15), 1481–1500, doi:10.1016/S0032–0633(01)00087–3.

Simons, F.J., Dahlen, F.A. & Wieczorek, M.A., 2006. Spatiospectral concentration on a sphere, *SIAM Rev.,* **48**(3), 504–536.

Slepian, D., 1983. Some comments on Fourier-analysis, uncertainty and modeling, *SIAM Rev.,* **25**(3), 379–393.

Sneeuw, N. & van Gelderen, M., 1997, The polar gap, in *Geodetic Boundary Value Problems in View of the One Centimeter Geoid,* pp. 559–568, eds Sansò, F. & Rummel, R., no. 65 in Lecture Notes in Earth Sciences, Springer, Berlin.

Stacey, F.D., 1992. *Physics of the Earth,* 3rd edn, Brookfield Press, Brisbane, Australia.

Szegö, G., 1975. *Orthogonal Polynomials,* 4th edn, American Mathematical Society, Providence, Rhode Island.

Tegmark, M., 1995. A method for extracting maximum resolution power spectra from galaxy surveys, *Astroph. J.,* **455,** 429–438.

Tegmark, M., 1996. A method for extracting maximum resolution power spectra from microwave sky maps, *Mon. Not. R. Astron. Soc,* **280,** 299–308.

Thébault, E., Schott, J.J. & Mandea, M., 2006. Revised spherical cap harmonic analysis (R-SCHA): validation and properties, *J. geophys. Res.,* **111**(B1), B01102, doi:10.1029/2005JB003836.

Wieczorek, M.A. & Simons, F.J., 2005. Localized spectral analysis on the sphere, *Geophys. J. Int.,* **162**(3), 655–675, doi:10.1111/j.1365–246X.2005.02687.x.

Wiggins, R.A., 1972. The general linear inverse problem: implication of surface waves and free oscillations for Earth structure, *Rev. Geophys.,* **10,** 251–285.

Wingham, D.J., 1992. The reconstruction of a band-limited function and its Fourier transform from a finite number of samples at arbitrary locations by Singular Value Decomposition, *IEEE Trans. Signal Process.,* **40**(3), 559–570, doi:10.1109/78.120799.

Xu, P., 1992a. Determination of surface gravity anomalies using gradiometric observables, *Geophys. J. Int.,* **110,** 321–332.

Xu, P., 1992b. The value of minimum norm estimation of geopotential fields, *Geophys. J. Int.,* **111,** 170–178.

Xu, P., 1998. Truncated SVD methods for discrete linear ill-posed problems, *Geophys. J. Int.,* **135**(2), 505–541.

APPENDIX A: GRÜNBAUM COMMUTATION

In this section we prove that the differential operator of eq. (83), rewritten for $\mu = \cos\theta$ and $b = \cos\Theta$,

$$
\begin{aligned}
\mathcal{T}_p &= (b^2 - \mu^2)\nabla_m^2 - 2\mu(1 - \mu^2)\frac{d}{d\mu} - L_p(L_p + 3)\mu^2 \\
&= \frac{d}{d\mu}\left[(b^2 - \mu^2)(1 - \mu^2)\frac{d}{d\mu}\right] - L_p(L_p + 3)\mu^2 \\
&\quad - \frac{m^2(b^2 - \mu^2)}{1 - \mu^2},
\end{aligned}
\tag{A1}
$$

commutes with the integral operator acting on $h_p(\mu')$ in eq. (80),

$$
\int_b^1 D_p(\mu, \mu')\, h_p(\mu')\, d\mu' = \lambda h_p(\mu),
\tag{A2}
$$

whose symmetric kernel, $D_p(\mu, \mu')$, is given by eq. (82), rewritten here for convenience in the form

$$
D_p(\mu, \mu') = \sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu) P_{lm}(\mu').
\tag{A3}
$$

The primed summation skips every second entry, and the lower and upper limits are as in eqs (68)–(69). The domain of eq. (A2) is the interval (81) of the double cap

$$
\{\mu : b \leq \mu \leq 1\} \cup \{\mu : -1 \leq \mu \leq -b\}.
\tag{A4}
$$

As in eq. (4), P_{lm} is the associated Legendre polynomial of degree l and order m, and C_{lm} is the normalization constant, eq. (5). We remind the reader of our notation: $h_p(\mu)$ is a colatitudinally dependent function that is limited in space to the antipodal polar caps of radius $\Theta = \cos^{-1} b$. It is either odd or even about the equator, as indicated by the subscript p,

$$
h_p(\mu) = 0, \quad -b \geq \mu \geq b,
\tag{A5a}
$$

$$
h_e(\mu) = h_e(-\mu),
\tag{A5b}
$$

$$
h_o(\mu) = -h_o(-\mu).
\tag{A5c}
$$

The solutions to eq. (A2) are functions $h_p(\mu)$ that are spectrally concentrated in a spherical harmonic degree interval $0 \leq l \leq L$; the eigenvalue λ is the quadratic measure of this concentration (36). The primed summation symbol skips every other term in the interval from m_p to L_p, which are both of the same parity, either even or odd. Depending on the requested order m and concentration bandwidth L of the solutions, m_p is either m or $m + 1$, and L_p is either L or $L - 1$, following eq. (69). We further distinguish $\mathcal{T}$ acting on μ from $\mathcal{T}'$ that acts on μ'.

To confirm commutativity we are required to show that

$$
\int_b^1 D_p(\mu, \mu')\, \mathcal{T}_p' h_p(\mu')\, d\mu' =
$$
$$
\int_b^1 \mathcal{T}_p D_p(\mu, \mu')\, h_p(\mu')\, d\mu'.
\tag{A6}
$$

We first show that the left side of eq. (A6) can be rewritten as

$$
\int_b^1 D_p(\mu, \mu')\, \mathcal{T}_p' h_p(\mu')\, d\mu' =
$$
$$
\int_b^1 \mathcal{T}_p' D_p(\mu, \mu')\, h_p(\mu')\, d\mu',
\tag{A7}
$$

and then we verify that

$$
\mathcal{T}_p D_p(\mu, \mu') = \mathcal{T}_p' D_p(\mu, \mu').
\tag{A8}
$$

The first result (A7) is easily verified by integration by parts: for any two functions $\zeta(\mu)$ and $\eta(\mu)$, it may be shown that, whether $b = \cos\Theta$ or $b = -1$,

$$
\begin{aligned}
\int_b^1 \zeta\,(\mathcal{T}_p\eta)\, d\mu = -\int_b^1 &\left[(b^2 - \mu^2)(1 - \mu^2)\frac{d\zeta}{d\mu}\frac{d\eta}{d\mu} \right.\\
&+ L_p(L_p + 3)\mu^2\,\zeta\eta \\
&\left.+ m^2(b^2 - \mu^2)(1 - \mu^2)^{-1}\zeta\eta \right] d\mu \\
= \int_b^1 &(\mathcal{T}_p\zeta)\,\eta\, d\mu.
\end{aligned}
\tag{A9}
$$

To verify the second result (A8) we use the Laplace–Beltrami identity $\nabla_m^2 P_{lm} = -l(l+1)P_{lm}$ (Dahlen & Tromp 1998) to write

$$
(\mathcal{T}_p - \mathcal{T}_p') D_p(\mu, \mu') =
$$
$$
(\mu^2 - \mu'^2)\sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu) P_{lm}(\mu')
$$
$$
\times [l(l+1) - L_p(L_p+3)]
$$
$$
- 2\mu(1 - \mu^2)\sum_{l=m_p}^{L_p}{}' C_{lm}\frac{d}{d\mu}P_{lm}(\mu) P_{lm}(\mu')
$$
$$
+ 2\mu'(1 - \mu'^2)\sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu)\frac{d}{d\mu'}P_{lm}(\mu').
\tag{A10}
$$

Two well-known Legendre identities help us simplify the above; a derivative identity and a recursion relation (Dahlen & Tromp 1998),

$$
\frac{dP_{lm}}{d\mu} = \frac{(l+1)\mu P_{lm} - (l-m+1)P_{l+1\,m}}{1 - \mu^2}
\tag{A11a}
$$

$$
\mu P_{lm} = \frac{(l-m+1)P_{l+1\,m} + (l+m)P_{l-1\,m}}{2l+1}.
\tag{A11b}
$$

Applying eq. (A11a), then eq. (A11b), transforms eq. (A10) into

$$
(\mathcal{T}_p - \mathcal{T}_p') D_p(\mu, \mu') =
$$
$$
(\mu^2 - \mu'^2)\sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu) P_{lm}(\mu')
$$
$$
\times [(l-2)(l+1) - L_p(L_p+3)]
$$
$$
+ \sum_{l=m_p}^{L_p}{}' C_{lm}[P_{l+2\,m}(\mu)P_{lm}(\mu') - P_{lm}(\mu)P_{l+2\,m}(\mu')]
$$
$$
\times 2\frac{(l-m+1)(l-m+2)}{2l+3}.
\tag{A12}
$$

The Legendre derivative identity of eq. (A11a) can be manipulated to yield a formula of the Christoffel–Darboux type (Szegö 1975),

$$
(\mu^2 - \mu'^2)\sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu) P_{lm}(\mu') =
$$
$$
\left[P_{L_p+2\,m}(\mu)P_{L_p m}(\mu') - P_{L_p m}(\mu)P_{L_p+2\,m}(\mu')\right]
$$
$$
\times \frac{(L_p - m + 2)!}{(2L_p + 3)(L_p + m)!}.
\tag{A13}
$$

Inserting eq. (A13) into eq. (A12) yields

$$(\mathcal{T}_p - \mathcal{T}_p')D_p(\mu, \mu') =$$

$$(\mu^2 - \mu'^2) \sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu) P_{lm}(\mu')$$

$$\times [(l-2)(l+1) - L_p(L_p+3)]$$

$$+ (\mu^2 - \mu'^2) \sum_{l=m_p}^{L_p}{}' 2(2l+1) \sum_{n=m_p}^{l}{}' C_{nm} P_{nm}(\mu) P_{nm}(\mu'). \qquad (A14)$$

Interchanging the order of summation and relabeling the sums,

$$(\mathcal{T}_p - \mathcal{T}_p')D_p(\mu, \mu') =$$

$$(\mu^2 - \mu'^2) \sum_{l=m_p}^{L_p}{}' C_{lm} P_{lm}(\mu) P_{lm}(\mu')$$

$$\times \left[(l-2)(l+1) - L_p(L_p+3) + \sum_{n=l}^{L_p}{}' 2(2n+1) \right]. \qquad (A15)$$

The term in square brackets always vanishes; in other words, $(\mathcal{T}_p - \mathcal{T}_p')D_p(\mu, \mu') = 0$ and the commutation relation of eq. (A6) is confirmed. Since commuting operators have the same eigenfunctions, we can find the spacelimited, fixed-order eigenfunctions $h(\theta)$ or $h(\mu)$ by solving the integral equation eq. (A2), or, equivalently, by solving the differential equation

$$\mathcal{T}_p h_p(\mu) = \chi h_p(\mu), \quad b \leq \mu \leq 1, \qquad (A16)$$

on the domain of the double polar cap, where $\chi \neq \lambda$ is the associated eigenvalue. The operator $\mathcal{T}_p$ of eq. (A1) is Sturm-Liouville, that is, when acting on h as in eq. (A16) it is of the form

$$(ph')' - qh + \chi \rho h = 0, \qquad (A17)$$

where $p(\mu) = (\mu^2 - b^2)(1 - \mu^2)$, $\rho(\mu) = 1$, $q(\mu) = m^2(1-\mu^2)^{-1}(\mu^2 - b^2) - L_p(L_p+3)\mu^2$, and the prime denotes differentiation with respect to μ. Since $\mathcal{T}_p$ is a Sturm-Liouville operator, the eigenvalue spectrum of eq. (A16) is simple: the spacing between adjacent fixed-order eigenvalues is roughly equant, as we illustrated in Fig. 9.

APPENDIX B: THE GRÜNBAUM MATRIX

The domains of eqs (A2) or (A16), originally restricted to the area contained within the double polar caps, may be extended to the entire sphere, $0 \leq \theta \leq \pi$, by writing

$$\mathcal{T}_p g_p(\mu) = \chi g_p(\mu), \quad -1 \leq \mu \leq 1. \qquad (B1)$$

The unknown fixed-order functions $g(\mu)$ or $g(\theta)$, given by eq. (67), must now be bandlimited rather than spacelimited:

$$g_p(\theta) = \sum_{l=m_p}^{L_p}{}' g_{lm} X_{lm}(\theta), \qquad (B2)$$

where, depending on the requested order m and concentration bandwidth L of the solutions, m_p is either m or $m+1$, L_p is either L or $L-1$, as in eqs (68)–(69), and the primed sum skips every second integer. As a result, g_{lm} requires no further identification.

Inserting the representation of eq. (B2) into eq. (B1), multiplying both sides by $2\pi \sin\theta\, X_{lm}(\theta)$, integrating over colatitudes $0 \leq \theta \leq \pi$, and invoking the orthogonality eq. (7) easily transforms eq. (B1) into

an algebraic eigenvalue equation for the vector of coefficients of the unknown functions g:

$$\mathsf{T}_p \mathbf{g}_p = \chi \mathbf{g}_p, \qquad (B3)$$

where we define

$$T_{ll'}^p = 2\pi \int_0^\pi X_{lm}(\mathcal{T}_p X_{l'm}) \sin\theta\, d\theta. \qquad (B4)$$

To avoid clutter, we have changed the subscript p indicating the parity of the solutions into a superscripted p on the matrix elements $T_{ll'}^p$. These are indexed by the integer degrees l and l'. Since in T_p the only degrees involved range from m_p to L_p, with every second degree skipped, when $m=0$, the first element of the matrix T_e is thus T_{00}^e, the second T_{20}^e, and so on, whereas the first and second elements of T_o are T_{11}^o and T_{30}^o, and so on. The matrix whose diagonalization leads to the even functions g_e is thus given by

$$\mathsf{T}_e = \begin{pmatrix} T_{mm}^e & T_{mm+2}^e & \cdots & T_{mL_e}^e \\ T_{m+2\,m}^e & T_{m+2\,m+2}^e & \cdots & T_{m+2\,L_e}^e \\ \vdots & \vdots & & \vdots \\ T_{L_e m}^e & T_{L_e\,m+2}^e & \cdots & T_{L_e L_e}^e \end{pmatrix}, \qquad (B5)$$

$[(L_e - m)/2 + 1]$. The $[(L_o - m - 1)/2 + 1]$–dimensional square matrix yielding the odd functions g_o is

$$\mathsf{T}_o = \begin{pmatrix} T_{m+1\,m+1}^o & T_{m+1\,m+3}^o & \cdots & T_{m+1\,L_o}^o \\ T_{m+3\,m+1}^o & T_{m+3\,m+3}^o & \cdots & T_{m+3\,L_o}^o \\ \vdots & \vdots & & \vdots \\ T_{L_o m+1}^o & T_{L_o\,m+3}^o & \cdots & T_{L_o L_o}^o \end{pmatrix}. \qquad (B6)$$

The dimension of the combined matrix

$$\mathsf{T}' = \mathrm{diag}\,(\mathsf{T}_e, \mathsf{T}_o) \qquad (B7)$$

is thus $(L - m + 1) \times (L - m + 1)$, as expected. Such a matrix can be thought of as a permutation of a 'full form'

$$\mathsf{T} = \begin{pmatrix} T_{mm} & \cdots & T_{mL} \\ \vdots & & \vdots \\ T_{Lm} & \cdots & T_{LL} \end{pmatrix}, \qquad (B8)$$

where the degrees are arranged in the correct order, without skipping entries. It is this matrix T that commutes with a matrix D that is in the form of eq. (54) and whose elements are given by eq. (66),

Deriving the form of the matrix entries $T_{ll'}^p$ of eq. (B4) requires the operator $\mathcal{T}_p$ of eq. (A1) as a function of colatitude, that is,

$$\mathcal{T}_p = (b^2 - \cos^2\theta)\nabla_m^2 + 2\cos\theta \sin\theta \frac{d}{d\theta} - L_p(L_p+3)\cos^2\theta. \qquad (B9)$$

Evaluating eq. (B4) is perhaps not as pedestrian as by 'simply reading off directly the inner products' as proclaimed by Grünbaum *et al.* (1982), but since only the result is of any practical consequence here, we will simply state it:

$$T_{ll}^p = -l(l+1)b^2 + \frac{2}{2l+3}\left[(l+1)^2 - m^2\right]$$

$$+ [(l-2)(l+1) - L_p(L_p+3)]$$

$$\times \left[\frac{1}{3} - \frac{2}{3}\frac{3m^2 - l(l+1)}{(2l+3)(2l-1)} \right], \qquad (B10a)$$

$$T_{l\,l+2}^{p} = \frac{\left[l(l+3) - L_p(L_p+3)\right]}{2l+3}$$
$$\times \sqrt{\frac{[(l+2)^2 - m^2]\,[(l+1)^2 - m^2]}{(2l+5)(2l+1)}}, \tag{B10b}$$

$$T_{l\,l'}^{p} = 0 \quad \text{otherwise.} \tag{B10c}$$

This is the result given as eq. (84) in the main text. The elements again are specified in terms of the degree and not by the matrix index. Both T_e and T_o are thus not only real and symmetric, but also tridiagonal. The important result is that the coefficients of the even or odd optimally concentrated antipodal polar cap eigenfunctions $g_e(\theta)$ or $g_o(\theta)$ both only require the numerical diagonalization of a tridiagonal matrix. These have analytically prescribed elements, and a spectrum of eigenvalues that is guaranteed to be simple.