SPHERICAL HARMONIC PROJECTORS

PAUL N. SWARZTRAUBER AND WILLIAM F. SPOTZ

Abstract. The harmonic projection (HP), which is implicit in the numerous harmonic transforms between physical and spectral spaces, is responsible for the reliability of the spectral method for modeling geophysical phenomena. As currently configured, the HP consists of a forward transform from physical to spectral space (harmonic analysis) immediately followed by a harmonic synthesis back to physical space. Unlike its Fourier counterpart in Cartesian coordinates, the HP does not identically reconstruct the original function on the surface of the sphere but rather replaces it with a weighted least-squares approximation. The importance of the HP is that it uniformly resolves waves on the surface of the sphere and therefore eliminates high frequencies that are artificially induced by the clustering of grid points in the neighborhood of the poles. The HP also maintains spectral accuracy when combined with the double Fourier method. Originally the HP required $O(N^3)$ storage where $N$ is the number of latitudinal points. However, this was recently reduced to $O(N^2)$ using an algorithm that also provided a savings of up to 50 percent in compute time. The HP was also generalized to an arbitrary latitudinal distribution of points. However, the HP as a composite of analysis and synthesis can be subject to considerable error depending on the point distribution. Here we define a variant of the traditional HP that is well conditioned, with condition number 1, for any point distribution. In addition, storage requirements are further reduced because the projections corresponding to all longitudinal wave numbers $m$ are expressed in terms of a single orthogonal matrix.

1. Introduction

For some time now, projections in one form or another have played an important role in the modeling of dynamical systems, particularly systems that admit solutions with vastly different time scales. Weather and climate provide such systems, and methods have long been researched for “projecting” equations and/or solutions onto slow manifolds to stabilize computations and make the computation of dominant slow solutions feasible in the presence of small but fast solutions corresponding to gravity and sound waves. Projections form the basis for modal analysis, and they are also used to obtain curl or divergent free approximations to any given functions. The Cartesian method [12] makes extensive use of projections at the level of the equations themselves. The three-dimensional equations in Cartesian coordinates are projected onto the surface of the sphere with an attendant reduction in the computational complexity as well as the elimination of high frequencies induced by the clustering of points near the poles.
Another example is given by the harmonic projections (HPs) that are implicit to the spectral method for solving partial differential equations on the sphere [11]. The stability and accuracy of this popular method for geophysical modeling is largely due to the HPs that are implicit to the numerous transforms between spectral and physical space. Early efforts to develop models on the sphere attempted to take advantage of the fast Fourier transform by duplicating the surface of the sphere and using the resulting two-dimensional periodic Fourier representation to compute the required spatial gradients [3]. These efforts met with limited success because of the high frequencies introduced by the closeness of the points near the poles.

More recently the double Fourier representation, combined with a number of other advances, has been elevated to the accuracy and robustness of the spectral method [4]. Indeed the computations can be made indistinguishable from the spectral method. The key ingredient is the periodic application of the HP. Although the HP still requires \( O(N^3) \) operations, it is more efficient than the traditional spectral method because more efficient formulations of the HP are available [13, 4, 8]. In addition, fewer \( O(N^3) \) computations are required because the fast Fourier transform (FFT) rather than harmonic transforms are used to compute spatial derivatives. It is important to note that the HP can also be computed using efficient iterative methods [6], [14] that have been compared to the direct methods considered here [8].

In previous work [13], the harmonic transforms (analysis and synthesis) were generalized to arbitrary distributions of latitudinal points. This unified the seemingly quite distinct transforms based on (for example) Gaussian and equally spaced latitudinal distributions. The transforms were also generalized to arbitrary rotations of the traditional transforms. In particular, rotations were defined that reduced storage requirements from \( O(N^3) \) to \( O(N^2) \) for applications that do not require the explicit computation of the spectral coefficients such as the harmonic projection that consists of a forward transform (analysis) followed immediately by a backward transform (synthesis). It is known that this combination projects the original function onto a smooth least-squares approximation without the high frequencies induced by the closeness of the points in the neighborhood of the poles [10].

Unlike the discrete Fourier transform the HP can amplify certain functions on the sphere. However, because it is a projection, amplification only occurs on its first application. Therefore one does not observe the exponential error growth that is traditionally associated with an amplification matrix. For a Gaussian or reasonably uniform latitudinal grid distribution, the initial amplification it quite acceptable. However, for arbitrary grid distributions, harmonic analysis can be subject to considerable error. Consequently, the computation of a projection as a composite of these transforms can also be subject to error.

We define a highly stable variant of the traditional harmonic projections with best possible condition number 1 for any distribution of latitudinal points. Storage requirements are further reduced by demonstrating that the projections corresponding to all longitudinal wave numbers can be computed from the columns of a single \( N \times N \) matrix where \( N \) is the number of latitudinal points. This may be significant for multicomputer implementations because it reduces the communication requirements. Unlike the traditional projections, the variants correspond to symmetric matrices with all singular values equal to 1.
2. Harmonic Transforms and Projections

Given the discrete function \( f_{i,j} \), defined at latitudes \( \theta_i \), \( i = 1, \ldots, N \), and longitudes \( \phi_j \), \( j = 1, \ldots, 2N - 2 \), the forward harmonic transform or harmonic analysis consists of determining coefficients \( a_{m,n} \) and \( b_{m,n} \) such that \( f_{i,j} \) has the representation

\[
 f_{i,j} = \sum_{m=0}^{N-1} \sum_{n=0}^{n} \sum_{m=0}^{N} \sum_{n=0}^{N} P_{m,n}^{(\theta_i)}(a_{m,n} \cos m\phi_j + b_{m,n} \sin m\phi_j),
\]

where \( P_{m,n}^{(\theta)} \) are tabulations of the normalized associated Legendre functions

\[
 P_{n}^{(\theta)} = \frac{1}{2^{2n}n!} \left[ \frac{2n+1}{2} \frac{(n-m)!}{(n+m)!} \right]^{1/2} (\cos \theta)^{n} \frac{d^{m+n}}{dx^{n}+m}(x^2-1)^n,
\]

where \( x = \sin \theta \).

The harmonic analysis, \( a_{m,n} \) and \( b_{m,n} \), are determined by first using the FFT to compute

\[
 a_{m}(\theta_i) = \frac{1}{2N-2} \sum_{j=1}^{2N-2} f_{i,j} \cos m\phi_j,
\]

\[
 b_{m}(\theta_i) = \frac{1}{2N-2} \sum_{j=1}^{2N-2} f_{i,j} \sin m\phi_j.
\]

If the \( \theta_i \) are Gauss distributed, then the harmonic analysis is given by

\[
 a_{m,n} = \sum_{i=1}^{N} w_{i} a_{m}(\theta_i) P_{n}^{(\theta_i)}(\theta_i),
\]

\[
 b_{m,n} = \sum_{i=1}^{N} w_{i} b_{m}(\theta_i) P_{n}^{(\theta_i)}(\theta_i),
\]

where the \( w_i \) are the Gaussian weights. Once the \( a_{m,n} \) and \( b_{m,n} \) are determined, the harmonic synthesis or backward transform is given by \( (1) \). As mentioned in the introduction, the forward harmonic transform followed immediately by a backward transform will not, in general, reconstruct the discrete function \( f_{i,j} \) but rather will provide a useful weighted least-squares approximation \( \hat{f}_{i,j} \) \[10\]. We begin with a study of the harmonic projection as a composite of the analysis followed by synthesis and later we develop the projection as a single entity with individual attributes.

In this work the focus is on the transforms \( (4) \) between Fourier and harmonic space because the Fourier transforms \( (3) \) are one-to-one, orthonormal, and well understood. The computation of both \( a_{m,n} \) and \( b_{m,n} \) in \( (4) \) is by application of the matrix operator \( P_m^T W_0 \) where \( W_0 \) is an \( N \times N \) diagonal matrix with Gaussian weights \( w_i \) and \( P_m \) is the \( N \times (N-m) \) matrix

\[
 P_m = \begin{bmatrix}
 P_{n}^{(\theta_1)} & \cdots & P_{n}^{(\theta_{N-1})} \\
 \vdots & \ddots & \vdots \\
 P_{n}^{(\theta_N)} & \cdots & P_{n}^{(\theta_{N-1})}
 \end{bmatrix}.
\]

The harmonic projection combines both the analysis and synthesis,

\[
 F_m = P_m P_m^T W_0,
\]
into a single \( N \times N \) matrix for each longitudinal wave number \( m \). The discrete Legendre functions are orthogonal with respect to Gaussian quadrature. That is, \( P^T_m W_0 P_m = I_{(N-m) \times (N-m)} \). This implies \( P_0 P_0^T W_0 = I_{N \times N} \) and therefore

\[
W_0 = (P_0 P_0^T)^{-1}.
\]

From \([9]\) we obtain \( P_0 = I_{N \times N} \). For a Gauss distribution of latitudinal points, the Christoffel-Darboux formula yields the well-known result that \( W_0 \) is diagonal, e.g., \([13]\).

These results can be generalized to any point distribution starting with the weighted least-squares projection of any vector onto any other set of vectors. For the moment assume \( P_m \) to be an arbitrary matrix with \( N - m \) columns, each with \( N \) elements. By direct substitution it can be established that the columns are weighted orthogonal with \( N \times N \) weight matrix

\[
W_m = (P_m P_m^T)^{-1}.
\]

That is, \( P^T_m W_a P_m = I_{(N-m) \times (N-m)} \). This provides an algorithm for approximating any given vector in terms of any other vectors; namely, the columns of \( P_m \), e.g., \([9]\). Following the earlier development, the “forward” transform of an arbitrary vector \( u \) is given by \( a = P^T_m W_m u \). The “backward” transform then provides a least-squares approximation to \( u \) in terms of the vectors \( P_m \); namely, \( \hat{u} = P_m a \).

The vector \( \hat{u} \) is the projection of \( u \) onto the vectors spanned by \( P_m \) because any subsequent application leaves \( \hat{u} \) unchanged. The projection \( \hat{u} \) is the closest vector in the span of \( P_m \) to the vector \( u \) in the weighted sense where distance is measured by the pseudonorm based on the weighted inner product \( \langle u, v \rangle_{W_m} = u^T W_m v \) \([10]\). Up to this point the presentation has been formal; however, the inverse on the right side of \((8)\) does not exist for \( m > 0 \). We will continue this discussion later. Now we return to the harmonic transforms based on a Gauss distribution of points.

The traditional harmonic transforms for a Gauss distribution of points fixes the weight matrix as \( W_0 = (P_0 P_0^T)^{-1} \) which is nonsingular and provides weighted orthogonality for all \( P_m \). This follows from the existence of an \( N \times (N - m) \) \( l_2 \) orthonormal matrix \( H_m \) such that \( P_m = P_0 H_m \) \([13]\). Therefore

\[
P_m^T W_0 P_m = H_m^T P_0^T (P_0 P_0^T)^{-1} P_0 H_m = I_{(N-m) \times (N-m)}.
\]

For other than a Gaussian distribution of points, two weight matrices are required corresponding to \( P_0 \) and \( P_1 \) \([13]\). Formally this discussion does not depend on the point distribution. That is, for any distribution with nonsingular \( P_0 P_0^T \), its inverse then provides the weight function for all \( P_m \), \( m = 0, \ldots, N - 1 \). Of course \( P_0 P_0^T \) can be quite poorly conditioned depending on the point distribution even though, as we will show, the projection itself is always well conditioned.

Numerical experiments reveal that traditional projections, based on a Gaussian distribution of points, correspond to a nonsymmetric matrix with singular values in excess of 1; i.e., there exist initial fields defined on the sphere whose \( l_2 \) norm is increased by a traditional projection. The maximum singular value for several latitudinal grid sizes is given in Table \([14]\) together with its corresponding longitudinal wave number \( m \).

A projection with a singular value greater than one does not generate the level of concern that it might for an arbitrary matrix because amplification does not
occur on subsequent applications that do not contain components corresponding to singular values that are greater than one. Therefore the exponential or explosive error growth associated with eigenvalues in excess of 1 is replaced by linear growth on the order of that induced by any orthonormal transformation.

3. The variant harmonic projection

In this section we develop variant symmetric projections that do not amplify in the sense that all singular values are equal to 1. We also develop a well-conditioned procedure for computing the projections that does not require the computation of the harmonic transforms or the weight matrix. To this end we use the singular value decomposition (SVD) [5] which is available at [7]. In particular, subroutine DSVDC was used to compute the numerical results reported here.

The SVD of an arbitrary $N \times (N - m)$ matrix $P_m$ is given by

$$P_m = U_m S_m V_m^T$$

(10)

where $U_m$ is an $N \times (N - m)$ column orthogonal matrix; $S_m$ is an $(N - m) \times (N - m)$ diagonal matrix with elements $s_{i,i}$ called singular values, and $V_m$ is an $(N - m) \times (N - m)$ orthogonal matrix. If $P^2_0 = P_0$ is itself a projection, then

$$S_0 = S_0 V_0^T U_0 S_0.$$  

(11)

If there are $l$ nonzero singular values, the first $l$ columns of $V_0$ and $U_0$ are mutually orthogonal (but not orthonormal) because $V_0^T U_0 = S_0^{-1}$. This explains why linear error growth occurs with singular values that are greater than one, unlike the exponential error growth associated with eigenvalues that are greater than one.

The variant projection, defined in this section, has a weight matrix that is based on $P_m$ rather than $P_0$. That is, the weight matrix is given by [8], which takes the form

$$W_m = U_m S_m^{-2} U_m^T.$$  

(12)

The harmonic analysis matrix has the form

$$A_m = V_m S_m^{-1} U_m^T.$$  

(13)

The synthesis matrix is just $P_m$ given by [11]. Finally the projection matrix, as the composite of [10] and [13], is given simply by

$$\hat{F}_m = U_m U_m^T.$$  

(14)
It is therefore evident that with singular values \( s_{i,i}^2 \), the weight matrix (12) is more poorly conditioned than the other matrices (13), (14), and even the synthesis matrix \( P \) itself. However, quite independent of the condition of the harmonic transforms, the projection (14) has best possible condition (condition number = 1) for any distribution of latitudinal points.

However, as noted in the previous section, projections with singular values in excess of 1 are not a major concern because the error growth is linear and in proportion to the singular value. That is, from Table 1 we note that the maximum singular value for \( N=1 \) is about 1.22 and occurs for longitudinal wave number \( m=1 \). The actual error following 10,000 applications of the traditional projection is \( 2 \times 10^{-3} \) compared with \( 1.783 \times 10^{-3} \) for the variant projection.

This completes the development of the variant projections. However, memory-efficient versions like those developed for the traditional projections in [13] still need to be determined. We will outline the representation of the \( N \) projections (14) in terms of a single \( N \times N \) matrix \( T \). The development is similar to that given for the traditional projections in [13] and reduces storage requirements from \( O(N^3) \) to \( O(N^2) \).

The following algorithm is valid for any \( U \) that span nested subspaces, which follows from the fact that the \( P \) span nested subspaces. Let \( u_n^{(m)} \) denote the columns of \( U \). Let first determine a unique vector \( t_0 \) that is \( l_2 \) orthogonal to \( u_0^{(1)} \) for \( n=1,\ldots,N-1 \). Next for \( m=2,\ldots,N-1 \) determine, in sequence, unique vectors \( t_{m-1} \) that are \( l_2 \) orthogonal to \( t_n \) for \( n=0,\ldots,m-2 \) and \( u_n^{(m)} \) for \( n=m,\ldots,N-1 \). Finally, set \( t_{N-1} = u_{N-1}^{(N-1)} \). If we define matrices \( T_m = [t_0,\ldots,t_{m-1}] \) and \( T = [t_m,\ldots,t_{N-1}] \), then all \( N \) projections can be represented in terms of a single matrix \( T \) as

\[
\hat{F}_m = T_m T_m^T \quad \text{or} \quad \hat{F}_m = I - T_m T_m^T.
\]

The first form is more efficient if \( m \leq N/2 \) and the second is more efficient if \( N/2 < m < N \). This algorithm differs significantly from traditional Gram-Schmidt orthogonalization because the vectors \( t_n \) are uniquely determined.

It should be noted that the variant projections alias differently than traditional projection. Variant projections of wave numbers \( n > N \) are not restricted to the set greater than \( N - m \); rather, they may distribute over all latitudinal wave numbers. This has been known to be true for equally spaced grids even with the traditional discrete harmonic transforms but it is true for any point distribution using the variant transforms. This implies that for strict alias control, one must use the one-half rather than the two-thirds rule. However, both rules are now considered rather severe, and more recently, some aliasing has been tolerated in exchange for the increased accuracy that is provided by the additional modes.

We have observed that the variant projections are \( l_2 \) norm reducing (less than or equal) in physical space, which prompts the question: Is the associated variant analysis also norm reducing? That is, is the norm of the resulting harmonic coefficients for the variant projection less than those obtained with the traditional projection? Tables 2 and 3 below imply that the singular values of the variant analysis matrix (13) are less than or equal to the singular values of the traditional analysis matrix. The variant and traditional singular values are the same for \( m = 0 \).
TABLE 2. Singular values of the variant analysis matrix $A_m$ for $N = 16$

<table>
<thead>
<tr>
<th>$m = 1$</th>
<th>$m = 2$</th>
<th>$m = 5$</th>
<th>$m = 6$</th>
<th>$m = 9$</th>
<th>$m = 13$</th>
<th>$m = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.435259</td>
<td>.435578</td>
<td>.435220</td>
<td>.435031</td>
<td>.434983</td>
<td>.434182</td>
<td>.432741</td>
</tr>
<tr>
<td>.431863</td>
<td>.432616</td>
<td>.431567</td>
<td>.432005</td>
<td>.430229</td>
<td>.424699</td>
<td></td>
</tr>
<tr>
<td>.427321</td>
<td>.426577</td>
<td>.426945</td>
<td>.425207</td>
<td>.424691</td>
<td>.416167</td>
<td></td>
</tr>
<tr>
<td>.418603</td>
<td>.421596</td>
<td>.417339</td>
<td>.419100</td>
<td>.411378</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.411286</td>
<td>.409162</td>
<td>.410124</td>
<td>.405021</td>
<td>.403007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.396113</td>
<td>.402809</td>
<td>.392902</td>
<td>.396899</td>
<td>.375868</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.386776</td>
<td>.382413</td>
<td>.384060</td>
<td>.372956</td>
<td>.366038</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.363625</td>
<td>.375549</td>
<td>.356601</td>
<td>.363844</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.353028</td>
<td>.345221</td>
<td>.347118</td>
<td>.324428</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.319526</td>
<td>.338537</td>
<td>.303585</td>
<td>.315399</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.308478</td>
<td>.295083</td>
<td>.294573</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.260035</td>
<td>.289129</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.249507</td>
<td>.224757</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.173186</td>
<td>.219996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.164780</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3. Singular values of the traditional analysis matrix $P_m^T W_m$ for $N = 16$

<table>
<thead>
<tr>
<th>$m = 1$</th>
<th>$m = 2$</th>
<th>$m = 5$</th>
<th>$m = 6$</th>
<th>$m = 9$</th>
<th>$m = 13$</th>
<th>$m = 15$</th>
</tr>
</thead>
<tbody>
<tr>
<td>.435259</td>
<td>.435236</td>
<td>.435242</td>
<td>.435086</td>
<td>.435030</td>
<td>.434250</td>
<td>.432851</td>
</tr>
<tr>
<td>.432122</td>
<td>.433719</td>
<td>.431785</td>
<td>.432249</td>
<td>.430469</td>
<td>.425093</td>
<td></td>
</tr>
<tr>
<td>.427321</td>
<td>.427103</td>
<td>.427154</td>
<td>.425708</td>
<td>.425123</td>
<td>.416825</td>
<td></td>
</tr>
<tr>
<td>.419651</td>
<td>.424899</td>
<td>.418218</td>
<td>.420062</td>
<td>.412355</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.411286</td>
<td>.410644</td>
<td>.410739</td>
<td>.406429</td>
<td>.404237</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.398516</td>
<td>.408324</td>
<td>.394906</td>
<td>.399024</td>
<td>.378185</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.386776</td>
<td>.385384</td>
<td>.385376</td>
<td>.375779</td>
<td>.368584</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.368039</td>
<td>.383141</td>
<td>.360253</td>
<td>.367559</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.353028</td>
<td>.350332</td>
<td>.349579</td>
<td>.329322</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.326818</td>
<td>.348267</td>
<td>.309597</td>
<td>.321213</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.308478</td>
<td>.303273</td>
<td>.298960</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.271667</td>
<td>.301471</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.249507</td>
<td>.237944</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.193087</td>
<td>.236499</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.164780</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Summary

Variant spherical harmonic projections have been defined with the following attributes.

(1) The error associated with the variant projection is marginally less than the traditional projection on a Gauss distributed latitudinal grid but may be substantially less on an arbitrary grid.
The variant projections are symmetric and all $N$ can be expressed as the outer product of orthonormal vectors from a single $N \times N$ matrix.

The variant projection is norm reducing in the $L_2$ sense, unlike the traditional spherical harmonic projection.

The algorithm for computing the variant projections, as well as the projections themselves, are well conditioned for any latitudinal grid distribution.

On a Gaussian grid the singular values of the variant analysis matrix are up to 10 percent less than the traditional analysis matrix.

Unlike the discrete Fourier transform, the harmonic projection has singular values that are larger than 1. However, error growth is only linear rather than exponential because the transform is a projection. Nevertheless, on an arbitrary grid the traditional HP, as a composite of analysis followed by synthesis, can be subject to substantial error unlike the variant HP provided here. The singular values of the variant analysis matrix are evidently also smaller. More importantly the SVD provides a mechanism to limit their size on an arbitrary grid.

REFERENCES


National Center for Atmospheric Research, P.O. Box 3000, Boulder, Colorado 80307-3000
E-mail address: pauls@ucar.edu

Sandia Corporation, P.O. Box 5800, Albuquerque, New Mexico 87123-1110
E-mail address: wfspotz@sandia.gov