1-1-1989

Asymptotically Fast Algorithms for Spherical and Related Transforms

James R. Driscoll
Dartmouth College

Dennis M. Healy
Dartmouth College

Asymptotically Fast Algorithms for Spherical and Related Transforms

James R. Driscoll and Dennis M. Healy

PCS-TR89-141

Asymptotically Fast Algorithms for Spherical and Related Transforms

James R. Driscoll* and Dennis M. Healy
Department of Mathematics and Computer Science
Dartmouth College
Hanover, NH 03755

May 1, 1989

Abstract

This paper considers the problem of computing the harmonic expansion of functions defined on the sphere. We begin by proving convolution theorems that relate the convolution of two functions on the sphere to a "multiplication" in the sprectral domain, as well as the multiplication of two functions on the sphere to a "convolution" in the spectral domain. These convolution theorems are then used to develop a sampling theorem on the sphere. In particular, if a function on the sphere is band-limited (i.e. in the span of the spherical harmonics $Y_l^m, l < L, |m| \leq l$), then a sampling of the function at $O(L^2)$ points can *exactly* recover the function. This sampling uses an asymptotically optimal number of points, and improves a sampling of $O(L^3)$ points for which error estimates were previously known. The sampling theorem also explains why a commonly used method of computing spherical transforms actually gives a poor estimate of the actual harmonic content of the function. Next, preliminary to developing a transform on the sphere, we give an $O(n(\log n)^2)$ time algorithm for computing the Legendre transform of a function on the interval $[-1, 1]$ sampled at n points. This improves the best previously known time bound of $O(n^2)$. This algorithm is then generalized to achieve algorithms for the spherical transform and its inverse transform that take $O(n^{1.5}(\log n)^2)$ time, where n is the number of sample points that are in accordance with the sampling theorem. This improves the naive $O(n^2)$ bound, which is the best previously known. These transforms give an $O(n^{1.5}(\log n)^2)$ algorithm for convolving two functions on the sphere, which would be useful for computer vision and elsewhere. This appears to be the first method for exactly computing the convolution, because it is difficult to determine how compute a spherical convolution in the absence of both an exact sampling theorem and a convolution theorem. The techniques developed are also applicable to computing other transforms for which only naive $O(n^2)$ algorithms are known, such as the Laguerre, Hermite, and Hankel transforms.

1 Introduction

It is difficult to overstate the importance of the $O(n \log n)$ Fourier transform of Cooley and Tukey [3]. The algorithm is widely applicable

*Supported by NSF grant CCR-8809573

and it can been used to develop efficient algorithms for other related transforms, such as the two dimensional Fourier transform. However, despite the fact that the fast Fourier transform is considered an integral part of theoretical computer science, as evidenced by its inclusion in nearly every algorithms text and the continuous attention it has received in the literature (see, for example, [5,17,15,16,1]), there is almost no work on other important transforms. Recently, however, the spherical tranform has been identified as a transform in need of a fast algorithm [17]. It is the purpose of this paper to provide one.

Spherical transforms are useful in a variety of fields including computer vision, statistics, tomography, geomagnetics, seismology, atmospheric science, signal processing, and crystallography. To date, however, no algorithm for computing spherical transforms is known that is asymptotically better than the naive $O(n^2)$ algorithm, where n is the number of sampled points of the function. Moreover, the standard method of approximating the transform from a sampling produces severe aliasing problems, as is explained by our convolution and sampling theorems.

The spherical transform is rather different from the two dimensional fourier transform. When the 2-D Fourier transform is approximated in the discrete case, one must assume the the sampled function is periodic along two orthogonal axes. This is equivalent to defining the discrete function on a torus. Thus, the familiar two-dimensional Fourier transform is a toroidal transform. We are concerned with harmonic analysis on the sphere, which exhibits rather different periodicity as compared to the torus.

The spherical transform, like the Fourier transform, has an interesting convolution property. As is well known, the 2-D convolution, which is efficiently computed by the two-dimensional Fourier transform, is useful in pattern matching. By convolving an image with a pattern, it is possible to locate translated copies of the pattern in the image. If the pattern has been rotated, however, this method will fail. In the case of functions on the sphere, the convolution method can detect translated *and* rotated versions of the pattern. Thus, a fast spherical convolution, which we show to follow from a fast transform and inverse, would be useful in computer vision.

One not uncommon approach to harmonic analysis on the sphere, motivated by the absence of efficient methods for spherical transforms, is to approximate the sphere as a torus by identifying two poles and to proceed as though the problem were framed on the torus. This leaves many planetary scientists in the curious position of having approximated the earth as a torus. A refinement of this approach is then to transform the toroidal harmonics back to the sphere to get an approximation of the spherical harmonic coefficients [6]. In general this approximation can be quite poor. In contrast, we develop *exact* and efficient methods for for computing transforms of band limited signals by first finding an apropriate sampling of the functions for which we can efficiently compute the transform. Of course, in order to justify the claim of exactness we must take as our model of computation the RAM with exact real arithmetic. This model in this context is at least historically justified, since a similar model was used by Gauss for the method of Gauss quadrature that exactly integrates degree bounded polynomials.

There are three immediate obstacles to the efficient computation of spherical harmonics, all of which make the problem interesting and challenging, in addition to explaining why little progress has been made in this area. First, the fast Fourier transform relies on the fact that the roots of unity generate a discrete subgroup on the circle. Unfortunately the discrete subgoups of the sphere consist of cyclic and dihedral groups of rotations about one axis, and the symmetries of the platonic solids [4]. These do not fill up the sphere densely enough to get even a reasonable approximation [12]. The second obstacle is that spherical harmonics are products of exponentials in one coordinate and associated Legendre functions in the other. This would seem to require a fast Legendre transform. Yet there is no known algorithm for transforms with nontrivial polynomial kernels that takes $o(n^2)$ time. Finally, the only exact sampling theorem on the sphere that we are aware of requires an aymptotically large number of sample points, which is related to the lack of useful discrete subgroups.

This paper presents an $O(n^{1.5}(\log n)^2)$ algorithm that, given a data structure of size $O(N \log N)$, computes the spherical harmonics of a discrete function on the sphere that is defined on an equi-angular grid of $n = 2^k \leq N$ points. This improves the naive $O(n^2)$ bound, which is the best previously known. Additionally, we describe methods to invert this transform in $O(n^{1.5})$ time, and using these trasforms, an $O(n^{1.5}(\log n)^2)$ time algorithm to convolve discrete functions on the sphere.

First we develop convolution theorems on the sphere, and using these, prove a sampling theorem for band-limited signals. It is this sampling theorem that allows the exact computation of the the spherical harmonics. As a preliminary step in devoloping a fast spherical transform, we give an $O(n(\log n)^2)$ time algorithm to compute the Legendre transform of a function sampled at the n points $\cos(2\pi i/n)$, for $0 \leq i < n$ and $n = 2^k \leq N$, given a data-structure of size $O(N \log N)$. This improves the best previously known bound of $O(n^2)$ [9]. Moreover, the technique used generalizes to any class of polynomials defined by a simple recurrence. This includes the Jacobi, Chebyshev, generalized Laguerre, and Hermite polynomials. Transforms based on these families of orthogonal polynomials, and related (non polynomial) transforms, such as the Hankel transform, are useful in weather modelling, signal processing, and tomography. No algorithms were known for these transforms that were asymptotically better than the naive $O(n^2)$ time algorithms, where n is the number of sample points.

Finally, we apply these results to develop the spherical transform.

The paper is organized as follows. Section 2 defines the spherical transform and some of its properties. Section 3 proves the convolution theorems. Section 4 discusses sampling on the sphere. Section 5 develops an $O(n(\log n)^2)$ Legendre transform. Section 6 adapts the methods for the Legendre transform to the spherical transform. Section 7 concludes.

2 Preliminaries

Many problems in physics and engineering possess spherical symmetry, and seperation of variables in spherical coordinates reduces these problems in part to the analysis of functions on S^2, the two dimensional unit sphere (surface of the unit ball) comprised of all unit vectors in three space, $\mathbf{R}^3$. The techniques of Fourier analysis, so familiar and useful for many problems in Euclidean space, are also helpful in the analysis in the non-Euclidean setting of the sphere. What follows is a brief review of these ideas.

Analysis on the sphere requires the use of coordinates. A familiar choice is the parameterization of the points of S^2 by angles of colattitude and longitude, θ and ϕ, with θ measured down from the z-axis,varying between 0 and π, and ϕ varying between 0 and 2π, measured from the x-axis. Thus a unit vector ω on S^2 may be parameterized as $(\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta))$ in these coordinates.

Fourier analysis may be characterized as the systematic use of symmetry to simplify certain linear operators. In the case considered in this paper, the unit sphere admits the special orthogonal group in three variables, $SO(3)$ as a transitive group of symmetries. These are the proper rotations of $\mathbf{R}^3$ about the origin, and are characterized as those three by three real matrices of determinant one whose inverses are given by their transposes. For example, we have the rotations about the z-axis:

$$\left\{ k(\phi) = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \middle| \phi \in [0, 2\pi] \right\},$$

or those about the x-axis:

$$\left\{ a(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & \sin\phi \\ 0 & -\sin\phi & \cos\phi \end{pmatrix} \middle| \phi \in [0, 2\pi] \right\}.$$

In fact, any rotation g in $SO(3)$ may be written as a product of matrices of these forms in the well known Euler angle decomposition: $g = u(\phi)a(\theta)u(\psi)$ with $\phi \in [0, 2\pi]$, $\theta \in [0, \pi]$, and $\psi \in [0, 2\pi]$ determined uniquely for almost all g [10,19]. Transitivity in this context means that any point on the sphere may be obtained from any other by a rotation. In particular, the entire sphere is swept out by taking all rotations of the north pole, $\mathbf{n} = (0, 0, 1)$. It should be noted that the rotation $u(\phi)a(\theta)u(\psi)$ with Euler angles ϕ, θ, and ψ takes the north pole to the point

$$\omega(\theta, \phi) = (\cos(\phi)\sin(\theta), \sin(\phi)\sin(\theta), \cos(\theta))$$

with spherical coordinates ϕ and θ. In fact, the sphere is a quotient of the rotation group, and inherits its natural coordinate system from that of the group.

Fourier analysis on the sphere amounts to the decomposition of the space of square integrable functions on S^2 into minimal subspaces invariant under all of the rotations in $SO(3)$, thus simplifying the analysis of rotation invariant operators. The rotations of the sphere induce operators on functions by rotating the graphs over the sphere. Specifically, for each rotation $g \in SO(3)$ we have the operator $\Lambda(g)$ defined on functions on the sphere by:

$$\Lambda(g)f \ \ (\omega) = f(g^{-1}\omega).$$

The presence of the inverse is required in order that the assignment Λ of rotations to their associated operators on functions respects the group law:

$$\Lambda(g_1 g_2) = \Lambda(g_1)\Lambda(g_2).$$

A vector subspace of functions on the sphere is invariant if all of the operators $\Lambda(g)$ for $g \in SO(3)$ take each function in the space back into the space.

The sphere admits an (essentially) unique rotation invariant area element which is very familiar when written out in coordinates:

$$\int_{\omega \in S^2} f(\omega)\, d\omega = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi} f(\omega(\theta, \phi)) \sin\theta d\theta d\phi.$$

One way to see that this is rotation invariant:

$$\int_{\omega \in S^2} f(g\omega)\, d\omega = \int_{\omega \in S^2} f(\omega)\, d\omega \quad g \in SO(3)$$

is to note that the usual volume element of $\mathbb{R}^3$ is rotation invariant, and when written in polar coordinates has angular part agreeing with the expression we have written for the sphere. Another way to see this is pass to the quotient from the invariant volume element on $SO(3)$ itself, which may be written $dg = \sin\theta d\theta d\phi d\psi$ in terms of the Euler angle coordinates. We will use this measure on $SO(3)$ when we discuss convolution.

Like the more familiar case of periodic functions on the line, or equivalently, functions on the circle, the minimal rotationally invariant subspaces are spanned by the restrictions of homogeneous polynomials of a fixed degree which are harmonic in the sense of being annihilated by the Laplace operator. Unlike the circle, the sphere admits a non commutative group of symmetries, and the invariant subspaces spanned by these harmonic homogeneous polynomials of degree $l = 0$ have dimension $2l + 1$ instead of being one dimensional. The invariant subspace of degree l harmonic polynomials restricted to the sphere is called the space of spherical harmonics of degree l. Spherical harmonics of different degrees are orthogonal to one another. Choosing an orthonormal basis of $2l+1$ spherical harmonics Y_l^m, $-l \leq m \leq l$ for each degree $l \geq 0$ gives an orthonormal basis for all of $L^2(S^2)$.

The Fourier decomposition of functions on the sphere into these classical spherical harmonics is given by

$$f(\theta,\phi) = \sum_{l\in\mathbf{N}} \sum_{|m|\leq l} \hat{f}(l,m) Y_l^m(\theta,\phi) \quad (1)$$

$$\hat{f}(l,m) = \int_{S^2} f\, \overline{Y_l^m}\, d\omega, \quad (2)$$

The spherical harmonics Y_l^m are harmonic polynomials of degree l restricted to the sphere. In coordinates $Y_l^m(\theta,\phi)$ is

$$(-1)^m \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi}.$$

The associated Legendre functions are defined as follows.

$$\begin{aligned} P_l^0(x) &= \frac{1}{2^l l!} D^l[(x^2-1)^l] \\ &= \frac{(-1)^l}{2^l l!} D^l[(1-x^2)^l], \quad (3) \end{aligned}$$

where D denotes derivative, and

$$P_l^k(x) = (1-x^2)^{k/2} D^l P_l(x).$$

The Legendre functions satisfy the recurrence

$$(l-m+1)P_{l+1}^m(x) - (2l+1)xP_l^m(x) + (l+m)P_{l-1}^m(x) = 0$$

This will be of use later.

One may single out the Legendre polynomials $P_l = P_l^0$ which arise in the analysis of functions on the sphere symmetric under rotations about the z-axis. They may be obtained by orthonormalizing the monomials $1, x, x^2, \ldots$ on the interval $[-1,1]$ by the Gram-Schmidt procedure. They find many applications in other fields ranging from data compression to optimal quadrature. For this reason, fast Legendre transforms should prove to have a broad utility.

Of all the possible bases for $L^2(S^2)$, the spherical harmonics uniquely exploit the symmetries of the sphere. Under a rotation g, each spherical harmonic of degree l is transformed into a linear combination of only those Y_l^m, $-l \leq m \leq l$ with the same degree:

$$\Lambda(g) Y_l^m(\omega) = \sum_{|m|\leq l} Y_l^m(\omega) D_{m,n}^{(l)}(g)$$

Thus the effect of a rotation on a function expressed in the basis of spherical harmonics is multiplication by a semi-infinite block diagonal matrix, with the $2l+1 \times 2l+1$ blocks for each $l \geq 0$ given by $D^{(l)}(g) = (D_{m,n}^{(l)})(g)$. In technical language, this constitutes the decomposition of the regular representation of $SO(3)$ on $L^2(S^2)$ into irreducible subrepresentations. For future use we note the explict expression for $D^{(l)}(g)$ when $g = k(\phi)a(\theta)k(\psi)$:

$$D_{m,n}^{(l)}(k(\phi)a(\theta)k(\psi)) = e^{-im\psi} d_{m,n}^{(l)}(\cos\theta) e^{-in\phi},$$

where the $d^{(l)}$ are related to Jacobi polynomials. [1,20]. From this, one may derive the relation between spherical harmonics and the matrix elements :

$$Y_l^m(\phi,\theta) = \sqrt{\frac{(2l+1)}{4\pi}} D_{m,0}^{(l)}(k(\phi)a(\theta)k(\psi))^*,$$

where the $*$ denotes transpose conjugate. In particular,

$$d^{(l)}_{m,0}(\cos\theta) = (-1)^m \sqrt{\frac{(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta).$$

The effect of all this is to block diagonalize rotationally invariant operators; namely convolution operators obtained as weighted averages of the rotation operators by (generalized) functions, or kernels. A well known example is the Laplace-Beltrami operator on smooth functions on S^2,

$$\Delta^* = \frac{1}{\sin\theta}(\frac{\partial}{\partial\theta}(\sin\theta\frac{\partial}{\partial\theta} + \frac{1}{\sin\theta}\frac{\partial^2}{\partial\phi^2})$$

which acts diagonally in the spherical harmonic basis, a fact exploited in many problems, including the quantum mechanical analysis of hydrogenic atoms. We now turn to a more detailed analysis of convolution.

3 Convolution on S^2

Each (generalized) function k on the sphere may be used to define a convolution operator. This is accomplished by employing it as a weighting factor for the operators Λ induced by the rotations of the sphere. The operator of convolution by k is then:

$$\begin{aligned} \mathfrak{K}_k f(\omega) &= \left(\int_{g\in SO(3)} dg\ k(g\mathbf{n})\Lambda(g)\right) f\ (\omega) \\ &= (\int_{g\in SO(3)} k(g\mathbf{n}) f(g^{-1}\omega)\ dg \\ &= k * f(\omega). \end{aligned}$$

Here, ω is any point of the sphere, and $\mathbf{n}$ is the north pole. It is immediate to verify that this sort of definition generalizes the usual one for functions on $\mathbb{R}$ up to quite general spaces admitting transitive symmetry groups.

Since the operators $\Lambda(R)$ are simultaneously block diagonalized for all $R \in SO(3)$ in the spherical harmonic basis, it follows that the convolution operators obtained as linear combinations of them must also be. This is the true utility of Fourier representations. Here is an explict statement of this well known fact.

Theorem 1 *For functions f, h in $L^2(S^2)$, the transform of the convolution is a pointwise product of the transforms:*

$$(f * h)\hat{}\,(l,m) = \hat{f}(l,m)\hat{h}(l,0)$$

Proof: The proof follows immediately from interchanging the order of the transform and convolution integrals, using invariance of the integrals under rotations, and the expression for the rotated spherical harmonics.

There is also an important dual convolution result relating pointwise products of functions on S^2 to a convolution of their transforms. This follows from the well known Clebsch-Gordon decomposition of tensor products of representations into irreducible representations [2,18,19]. This may be thought of as giving the description of the algebraic (ring) structure of the transform domain.

As any functions on S^2 may be decomposed into spherical harmonics, the description of the transform of a product of functions follows from the following classically known transform of a product of spherical harmonics.

Theorem 2

$$Y_{l_1}^{m_1} Y_{l_2}^{m_2} = \sum_{L=|l_1-l_2|}^{l_1+l_2} \sum_{|M|\le L} \sqrt{\frac{(2l_1+1)(2l_2+1)}{4\pi(2L+1)}} \times C^{l_1,l_2,L}_{0,0,0} C^{l_1,l_2,L}_{m_1,m_2,M} Y_L^M.$$

The Wigner symbols $C^{l_1,l_2,L}_{m_1,m_2,M}$ are discussed in standard references: [2]. In particular they vanish unless $m_1 + m_2 = M$ and $|l_1 - l_2| \le L \le l_1 + l_2$.

Proof: The proof of this, the many symmetries and orthogonality properties of the Wigner coefficients, and recursive techniques for their calculation may be found in standard references. [2,18,19]

4 Sampling Theorems

It is often desirable to sample a band-limited function is such a way that the original function can be exactly recovered from the samples. In the case of functions on the line the classical sampling theorem was given by Shannon and states that a function of bounded frequency must be sampled at a frequency at least twice the bounding frequency. Despite the fact that no real-life signal is band-limited in a strong sense, this has been enormously successful in practice and it provides the foundation for the now common technology of digital audio.

While we are not aware of any sampling theorems for the sphere in the literature, the investigation of integration formulae has a long history and there are relevant quadrature results on the sphere, though each has its drawbacks. One, [13], has the unfortunate requirement that the function be sampled differently for each value of m that we wish to compute the transform $\hat{f}(l,m)$. This results in the function being significantly oversampled, even in an asymptotic sense. Another, [8], requires that the evaluation of each $\hat{f}(l,m)$ weight the samples differently, thus precluding a more efficient means of computing the spherical transform. In both cases, only error estimates are determined, and no claim is made that either exactly integrates band-limited signals.

We, however, exhibit a sampling for a band-limited function that requires an asymptotically optimal number of samples, which can exectly recover the sampled function. Moreover, we can efficiently and exactly compute the transform from the sampling.

Let $f(\theta,\phi)$ be a band-limited function such that $\hat{f}(l,m) = 0$ for $l \geq b$. We will sample the function at the equiangular grid of points (θ_i,ϕ_j), $i = 0,\ldots 2b-1$, $j = 0,\ldots,2b-1$, where $\theta_i = \pi i/2b$ and $\phi_j = \pi j/b$. Define $a_k^{(b)}$ to be the unique solution to the system of equations

$$a_0^{(b)} P_m(\theta_0) + a_1^{(b)} P_m(\theta_1) + \cdots$$

$$+ a_{2b-1}^{(b)} P_m(\theta_{2b-1}) = c_m$$

for $m = 0,\ldots,2b-1$ and where $c_0 = 1$ and $c_m = 0$ for $m \neq 0$. (That the solution exists and is unique follows from the orthogonality of the Legendre polynomials.) Multiply each sampled point at (θ_i,ϕ_j) by $a_i^{(b)}$.

This can be equivalently thought of as multiplying the function f by a weighted grid of impulses on the sphere which we will denote be s, the sampling function. Thus the sampled function f_s can be thought of as the product $f \cdot s$. We know the spectral content of f is band-limited, and if we knew the spectral content of s, we could determine the spectral content of $f \cdot s = f_s$ from the convolution theorem. We can easily calculate the transform of s, and in fact $\hat{s}(0,0) = 1$ and $\hat{s}(l,m) = 0$ for $0 < l < 2b$. Thus, from the convolution theorem we deduce that $\hat{f}_s(l,m) = \hat{f}(l,m)$ for $l < b$.

In terms of computing the spherical harmonics exactly this is precisely what we desire. However it is possible to frame this in terms similar to the classical sampling theorems. In order to recover $\hat{f}$ from $\hat{f}_s$ we must multiply $\hat{f}_s(l,m)$ by 1 for $l < b$ and by 0 otherwise. We can, by means of the convolution theorem, pull this multipliucation back to the sphere to arrive at the following:

Theorem 3 *If $\hat{f}(l,m) = 0$ for $l < b$ and $f_s = f \cdot s$, where s is the weighted grid of impulses* *as above, then*

$$f = f_s * \sum_{l=0}^{b-1} Y_l^0$$

where $*$ denotes spherical convolution.

5 Legendre Transforms

As a first step toward developing a spherical transform algorithm, we consider the problem of computing the Legendre tranform of a band-limited function over the interval $[-1,1]$. We shall restrict our attention to functions that are sampled at the $n = 2^k$ points $\cos(\pi i/n)$, for $i = 0, \ldots, n-1$.

Lemma 1 *If $f(x)$ over the interval $[-1,1]$ is in the span of $P_i(x)$, $i < n/2$, with $x_j = \cos(\pi j/n)$, and $a_j^{(n)}$ defined as in the sampling theorem, then*

$$\hat{f}(k) = \sum_{j=0}^{n-1} a_j^{(n)} f(x_j) P_k(x_j)$$

for $k < n/2$, and zero otherwise.

Proof: An immediate consequence of the sampling theorem.

The naive method of computing the transform is to evaluate each of the $n/2$ sums by evaluating and adding n terms, which requires $O(n^2)$ time. No asymptotically better algorithm is currently known. Below, we develop an $O(n(\log n)^2)$ algorithm, for n a power of two, which requires a preprocessed data structure. The overall plan is to first project the function onto an exponential basis by the fast Fourier transform. This basis is then transformed into a non-orthogonal polynomial basis. Finally, this polynomial basis is transformed into the Legendre basis. The new techniques required by the algorithm are the use of the intermediate basis as well as the methods to accomplish the change efficiently. (Naively, each change of basis takes $O(n^2)$ time.)

The intermediate transform that we will compute is defined as

$$g(k) = \sum_{j=0}^{n-1} a_j^{(n)} f(\cos j\theta)(\cos j\theta)^k$$

where $\theta = \pi/n$), here and in the following. Let $\omega = e^{-i\theta}$. Then $\omega^k = e^{-ik\theta}$, and $w^k + w^{-k} = 2\cos k\theta$. The transform can then be equivalently written as

$$g(k) = \sum_{j=0}^{n-1} a_j^{(n)} f(\cos j\theta)(\frac{\omega^j}{2} + \frac{\omega^{-j}}{2})^k.$$

Let

$$X(k,l) = \sum_{j=0}^{n-1} a_j^{(n)} f(\cos j\theta)(\omega^j)^k(\frac{\omega^j}{2} + \frac{\omega^{-j}}{2})^l,$$

where $|k| < n$ and $0 \le l < n$. Then $X(0,l) = g(l)$, the intermediate transform we wish to compute. Also,

$$X(k,0) = \sum_{j=1}^{n-1} a_j^{(n)} f(\cos j\theta)(\omega^j)^k,$$

which can be efficiently obtained from a single discrete Fourier transform, as shown below.

Let

$$h(j) = \begin{cases} a_j^{(n)} f(\cos j\theta) & 0 \le j < n \\ 0 & \text{otherwise} \end{cases},$$

for $j = 0, \ldots, 2n-1$. Then the discrete Fourier transform of h, $\mathcal{F}[h]$, gives us the vales of $X(k,0)$ in $O(n \log n)$ time, since

$$X(k,0) = \begin{cases} \mathcal{F}[h](k) & 0 \le k < n \\ \mathcal{F}[h](2n+k) & 2n-1 \le j < 0 \end{cases}.$$

This is because, for $k \ge 0$, $\mathcal{F}[h](k) = \sum_{l=0}^{2n-1} h(l) \exp(-ikl2\pi/2n) =$

$\sum_{l=0}^{n-1} f(cos j\theta)(\omega^l)^k = X(k,0)$. For $k < 0$ note that $\omega^k = \omega^{2n+k}$ where $2n+k$ is positive. Thus $\mathcal{F}[h](2n+k) = X(k,0)$ by the previous reasoning.

Our immediate goal is to compute $X(0,l) = g(l)$, for $l = 0, \ldots, n-1$, from the $X(k,0)$. One way to do this is to use the relation $X(k,l) =$

$$\frac{1}{2}X(k+1,l-1) + \frac{1}{2}X(k-1,l-1) \quad (4)$$

for $|k| < n - l$.

Although this allows us to compute the values of $X(k,l)$, for some fixed l, in terms the $X(k,l-1)$, this is not immediately of much help in that direct computation of the $X(0,l) = g(l)$ by this method requires $O(n^2)$ time.

In what follows it will be convienent to think of $X(k,l)$, for fixed k as a function of l, so we will often write $X_k(l)$ for $X(k,l)$.

Let $U(k) = 1/2$ for $|k| = 1$ and zero otherwise. The the relationship between X_l and X_{l-1} can be rewritten as

$$X_l(k) = [X_{l-1} * U](k)$$

where $*$ denotes discrete convolution. If we call the convolution of U with iteself i times $U^{(i)}$, then by the associativity of convolution,

$$X_{l_1+l_2}(k) = [X_{l_1} * U^{(l_2)}](k).$$

We now show how to use this equation to compute $X_l(0) = g(l)$ efficiently.

Since discrete convolution of two functions on n values can be computed by the fast Fourier transform in $O(n \log n)$ time, $X_{n/2}$ can be computed by convolving X_0, which we have computed, with $U^{(n/2)}$, which we have not (this will be adressed shortly).

The $g(l) = X_l(0)$ for $l < n/2$ can be computed from the values of X_0. However, they will only depend on $X_0(j)$ for $|j| < n/2$ when they are calculated by using relation 4. If we define $X_l^i(k)$ to be the function $X_l(k)$ restricted to the values $|k| < i$, then

$$X_{n/4}^{n/2} = X_0^{n/2} * U^{(n/4)}$$

and similarly

$$X_{3n/4}^{n/2} = X_{n/2}^{n/2} * U^{(n/4)}$$

where $U^{(n/4)}$ is as before, but restricted to $n/2$ points. Thus, in $O(2(n/2)\log(n/2))$ steps we can recover $X_{n/4}(0)$ and $X_{3n/4}(0)$, since we must do two convolutions of discrete functions of $n/2$ sampled points. By continuing in this fashion, we can determine $g(l)$ for all l in time proportional to

$$\sum_{i=0}^{\log n} 2^i \frac{n}{2^i} \log \frac{n}{2^i}$$

which is $O(n(\log)^2)$.

We have omitted the computation of the $U^{(i)}$. These are needed for all powers of 2 less than n, and each $U^{(i)}$ is defined on $2i$ points. If we know that $n < N$, then it suffices to compute all $U^{(2^k)}$, $2^k \leq n$. The total size of the $U^{(i)}$ will then be $O(N)$. To compute these for a single Legendre transform would increase the asymptotic running time, so we assume that we have precomputed these values which are independent of the input.

It remains to transform this polynomial basis to the Legendre basis. This can be accomplished by repeated convolution in a manner similar to the above, by using the fact that

$$(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$$

The details of this transformation are omitted in this abstract.

The preceeding discussion can be summarized by the following:

Lemma 2 *If $n = 2^k < N$ then*

$$\hat{f}(k) = \sum_{j=0}^{n-1} a_j^{(n)} f(x_j) P_k(x_j)$$

can be computed for all $k < n/2$ in $O(n(\log n)^2)$ time given a precomputed data structure of size $N \log N$.

By using the recurrence for the associated Legendre functions, we also obtain the more general result:

Theorem 4 *If $n = 2^k < N$ then*

$$\hat{f}(k) = \sum_{j=0}^{n-1} a_j^{(n)} f(x_j) P_k^m(x_j)$$

can be computed for all $m \leq k < n/2$ in $O(n(\log n)^2)$ time given a precomputed data structure of size $N \log N$.

It is important to note that the sum in the preceeding theorem is not claimed to be the transform of f with respect to any family of associated Legendre functions except for $m = 0$, where the sampling theorem applies. This, however, poses no difficulty for computing spherical transforms.

It remains to consider the problem of inverting the Legendre transform. The inversion is obtained by

$$f(x) = \sum_{k=0}^{n-1} \hat{f}(k) P_k^m(x)$$

. It is easily seen that this sum can be calculated in $O(n)$ time, and hence the function can be recovered at the $2n$ sampled points in $O(n^2)$ time. This will prove to be sufficient for our purposes, and we know of no better method.

6 Spherical Transforms

Suppose $f(\theta, \phi)$ is in the span of $\{Y_l^m | l < b/2, |m| \leq l\}$. Then from the sampling theorem we know that the transform $\hat{f}(l, m) =$

$$\sum_{j=0}^{b-1} \sum_{k=0}^{b-1} a_k^{(b/2)} f(\theta_k, \phi_j) Y_l^m(\theta_k, \phi_j)$$

where $\theta = \pi/b$ and $\phi = 2\pi/b$. Rearranging we get $\hat{f}(l, m) =$

$$c_{l,m} \sum_{j=0}^{b-1} e^{-im\phi_j} \sum_{k=0}^{b-1} a_k^{(b/2)} f(\theta_k, \phi_j) P_l^m(\cos\theta_k)$$

If we take

$$g_{m,\phi_j}(l) = \sum_{k=0}^{b-1} a_k^{(b/2)} f(\theta_k, \phi_j) P_l^m(\cos\theta_k)$$

then for fixed m and j, g_{m,ϕ_j} is in the form of an associated Legendre transform. Each transform can be accomplished in $O(b(\log b)^2)$ time, and since each of m and j range over $O(b^2)$ values, the total time to compute all values of $g_{m,\phi_j}(l)$ is $O(b^3(\log b)^2)$. It remains to compute

$$\hat{f}(l, m) = c_{l,m} \sum_{j=0}^{b-1} e^{-im\phi_j} g_{m,\phi_j}(l)$$

This, however, consists of $O(b^2)$ sums each of $O(b)$ terms, which can be directly computed in $O(b^3)$ time. Since the number of points, n, that the functions is sampled at is $O(b^2)$, the total time required to compute the transform is $O(n^{1.5}(\log n)^2)$. We thus have:

Theorem 5 *If $f(\theta, \phi)$ is in the span of $\{Y_l^m | l < b/2, |m| \leq l\}$, then the spherical transform of f can be computed in $O(n^{1.5}(\log n)^2)$ time from $n = O(b^2)$, $n = 2^k < N$, sampled points, using a preprocessed data structure of size $O(N \log N)$.*

The transform is inverted by the sum

$$f(\theta_k, \phi_j) = \sum_{l=0}^{b-1} \sum_{m=-l}^{l} \hat{f}(l, m) Y_l^m(\theta_k, \phi_j)$$

If we rearrange and exchange sums by extending $\hat{f}$ to be 0 at inappropriate indices, we get

$f(\theta_k, \phi_j) =$

$$\sum_{m=1-b}^{b-1} e^{-im\phi_j} \sum_{l=0}^{b-1} \hat{f}(l,m) c_{l,m} P_l^m(\cos\theta_k)$$

Take $g(\theta_k, m)$ to be the sum

$$\sum_{l=0}^{b-1} \hat{f}(l,m) c_{l,m} P_l^m(\cos\theta_k)$$

The $O(b^2)$ values for g can each be determined by summing the $O(b)$ terms, for a total of $O(b^3)$ time. Then

$$f(\theta_k, \phi_j) = \sum_{m=1-b}^{b-1} e^{-im\phi_j} g(\theta_k, m)$$

can be determined also by a naive summation in $O(b^3)$ time. Therefore we conclude:

Theorem 6 *If $\hat{f}(l,m) = 0$ for $l \geq b$, then the inverse transform of $\hat{f}$ can be computed at $n = O(b^2)$ points, $n = 2^k$, in $O(n^{1.5})$ time.*

Combining the previous theorems and the convolution theorem we recover:

Theorem 7 *If f and g are band-limited functions on the sphere (i.e. $\hat{f}(l,m)$ and $\hat{g}(l,m) = 0$ for $l \geq b$), then $f * g$, the spherical convolution of f and g can be computed in $O(n^{1.5}(\log n)^2)$ time, for $n = O(b^2)$, $n = 2^k < N$ using a precomputed data structure of size $O(N \log N)$.*

7 Conclusion

We have given natural convolution theorems for functions on the sphere and used these to produce an asymptotically optimal sampling theorem for band limited functions. These were then applied to the problem of efficiently computing spherical transforms, and an $O(n^{1.5}(\log n)^2)$ time algorithm was give to compute these transforms for functions sampled at $n = 2^k$ points on an equiangular grid. This improves the naive bound of $O(n^2)$, the best previously known algorithm. This is then used to efficiently compute convolutions of functions on the sphere, a problem for which no exact algorithms were known. The methods used generalize to many other related transforms that are of wide utility.

Many tantalizing questions remain, from which we single out a few:

- Our sampling theorem requires an asymptotically optimal number of sampled points for a given band-limited function, but is it possible to exhibit an "optimal" sampling? That is, if a function has no energy for $l \geq b$, then there are at most $(b-1)(b-2)+b$ non-zero coeffiicents in the transform, and we ask is there a sampling of the function at $(b-1)(b-2)+b$ points that can exactly recover the transform? Such a sampling would make sense of a discrete spherical transform. We suspect that no such sampling is possible.

- It is curious that we were unable to invert the Legendre transform in better than the naive $O(n^2)$ time, yet it was not an obstacle to inverting the spherical transform. Perhaps this could be explained by relating time bounds for these problems, or perhaps an efficient inversion can be found.

- It would, of course, be most interesting to improve the $O(n^{1.5}(\log n)^2)$ time bound. We suspect that it will be difficult to break the $O(n^{1.5})$ barrier, because the harmonics don't naturally decompose into products of functions of the coordinate system (as in the case of the torus). Perhaps a lower bound argument could be made, or alternately, a differenct harmonic representation on another coordinate system might remove the obstacle (though we doubt it).

References

[1] Beth, T., On the Computational Complexity of the General Discrete Fourier Transform, *Theoretical Computer Science*, 51, 331-339 (1987).

[2] L. C. Biedenharn and J. D. Louck *Angular Momentum in Quantum Mechanics*, Addison Wesley, Reading (1981).

[3] J. W. Cooley and J. W. Tukey, An Algorithm for Machine Calculation of Complex Fourier Series, *Math. Comput.*, 19, 297-301 (1965).

[4] H. S. M. Coxeter, *Regular Polytopes*, Dover, New York (1973).

[5] Diaconis, P., Average Running Time of the Fast Fourier Transform, *J. Algorithms*, 1, 187-208 (1980).

[6] G. A. Dilts, Computation of Spherical Harmonic Expansion Coefficients via FFT's, *Journal of Computationsal Physics*, vol 57, no 3,439-453 (1985).

[7] W. Freeden, On integral formulas of the (unit) sphere and their application to numerical computation of integrals, *Computing*, 25, 131-146 (1980).

[8] N. C. Gallagher, G. L. Wise, and J. W. Allen, A Novel Approach for the Computation of Legendre Polynomial Expansions, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, no. 1, 105-106 (1978).

[9] H. Goldstein *Classical Mechanics*, Addison Wesley, Reading (1980).

[10] M. Hammermesh *Group Theory*, Addison Wesley, Reading (1964).

[11] D. M. Healy, *A Relationship Between Harmonic Analysis on $SU(2)$ and on $SL(2,C)/SU(2)$*, Ph.D. Thesis, U.C. San Diego, 1986.

[12] K. Kobayashi, Solution of multi-dimensional neutron transport equation of the spherical harmonics method using the finite Fourier transformation and quadrature formula, *Transport Theory and Stat. Phys.*, 14, 63-81 (1985).

[13] A. Messiah *Quantum Mechanics, volume II*, John Wiley and Sons, New York (1958).

[14] Morgenstern, J., Note on a Lower Bound of the Linear Complexity of the Fast Fourier Transform, *J. Assoc. Comput. Mach.*, 20, 305-306 (1973).

[15] Morgera, S. D., Efficient Synthesis and Implementation of Large Discrete Fourier Transformations, *SIAM J. Comput.*, 9, 251-272 (1980).

[16] Papadimitriou, C. H., Optimality of the Fast Fourier Transform, *J. Assoc. Comput. Mach.*, 26, 95-102 (1979).

[17] J. T. Schwartz, Mathematics Addresses Problems in Computer Vision for Advanced Robotics, *SIAM News*, 18, 3 (1985).

[18] A. A. Terras *Harmonic Analysis on Symmetric Spaces I*, Springer-Verlag, New York (1985).

[19] N. J. Vilenkin *Special Functions and the Theory of Group Representations*, American Mathematical Society, Providence (1968).