Notes on Basic Ideas of Spherical Harmonics

In the representation of wavefields (solutions of the wave equation) one of the natural considerations that arise along the lines of Huygens Principle is the representation of a wavefield inside a (large) sphere in terms of sources on the boundary—the concept usually associated with the names Helmholtz and Kirchoff (the H-K integral) for general boundary surfaces, too. In this context and in many others, it is useful to have a way of expanding functions on the sphere in series in a way analogous to the expansion of functions on the circle (functions of an angle) in Fourier series.

Approximation by Polynomials and Spherical Harmonics Defined

It is a familiar basic fact of analysis that every (continuous) function on the unit sphere around the origin in Euclidean three-dimensional space can be approximated by a polynomial in the x, y and z coordinates on Euclidean three–dimensional space. The unit sphere in question is exactly the set of points (x,y,z) where the sum of the squares of x ,y and z is equal to 1. Thus, the functions x, y, and z themselves are not independent on the sphere. And if one approximates a continuous function on the sphere by a polynomial in x, y and z-- then the same polynomial can assume many different forms in terms of x, y and z-- just because of this lack of independence. You can see the same thing on the circle in the (x,y) plane defined by $x^2 + y^2 = 1$: the polynomial $x^4 + y^4 + 2x^2y^2$ has the same values on the unit circle as the polynomial $x^2 + y^2$, both of them being everywhere equal to 1!

Similarly, on the unit sphere $x^2 + y^2 + z^2 = 1$, the polynomials $x^4 + y^4 + z^4 + 2x^2y^2 + 2x^2z^2 + 2y^2z^2$ and $x^2 + y^2 + z^2$ are the same functions since they are both equal to 1 everywhere (on the unit sphere) .

To make some order in this chaos of polynomials, one needs to find , for each degree of polynomial, a set of polynomials of that degree such that taken together all the chosen polynomials up to and including a certain degree generate all the polynomials of that degree or less.

Let us illustrate first how this works for the circle. For degree 0, we choose of course 1. For degree one, x and y are independent and suffice. But when we get to degree 2, we have to throw away one of the obvious monomial candidates x^2, y^2 or xy -- because they are not independent in the collection of themselves and the lower degree things we already have: $x^2 + y^2 = 1$, and 1 is something we have already in the lower degree things. A basic observation of Fourier series is that the two polynomials corresponding to the sine and cosine of the doubled angle are a generating set in degree 2 if you throw in the relation that $x^2 + y^2 = 1$.

For example, the sine and cosine of twice the angle are $2xy$ and $x^2 - y^2$ respectively (using that x= cos {of the angle} and y =sin {of the angle} and that for a given angle α, $\sin 2\alpha = 2\sin\alpha\cos\alpha$ and $\cos 2\alpha = \cos^2\alpha - \sin^2\alpha$) . And these two, together with the functions:

$1, \cos\alpha = x, \sin\alpha = y$, do generate all the second degree polynomials on the unit circle. For instance,

$$x^2 = \frac{1}{2}\left[(x^2 + y^2) - (x^2 - y^2)\right] = \frac{1}{2}(1) - \frac{1}{2}(x^2 - y^2).$$

Carrying this further, one can see that the linear combinations (with constant coefficients) of $1, \cos\alpha, \sin\alpha, \cos 2\alpha, \sin 2\alpha, ..., \cos n\alpha, \sin n\alpha$ written in terms of x and y by using trigonometry formulas , give exactly the same set of functions on the unit circle as the polynomials in x and y with terms of degree up to and including degree n. Since the constant 1 and the sines and cosines are all linearly independent, this is a good way to approximate functions on the unit circle.

In these terms, one can understand what we want from spherical harmonics: they should be , for each degree {0,1,2,3,....} , a set of homogeneous polynomials (all terms having the given degree) in x y and z with the properties that if you take all the spherical harmonics up to and including degree n, then they are linearly independent and every polynomial in x y and z of degree n (not necessarily homogeneous: just all terms of degree less than or equal to n) is a linear combination (with constant coefficients) of these spherical harmonics.

An additional condition is usually added, that the degree-n harmonics should be "perpendicular" to the harmonics of degree $\leq$ n-1 in a sense that we shall explain later. This makes the process more orderly and more nearly unique.

In a given degree n, you need 2n+1 spherical harmonics; that is , to go from the set of degree n-1 or less harmonics to the degree n or less harmonics, you need to adjoin exactly 2n+1 new harmonics of degree n. This contrasts with the circle case where one needs one for degree 0 and only two for all higher degrees, namely the sine and cosine of n times the angle, where n=the degree.

Examples for low degrees may help:

Degree 0	One harmonic: the constant 1
Degree 1	Three harmonics: x, y and z
Degree 2	Five harmonics: $x^2 - y^2$, $x^2 - z^2$, xy, xz, yz

Notice that the five degree 2 harmonics , though we picked a set that looks nice and are homogeneous of degree 2, are in fact equal on the sphere to functions which have no power of x greater than 1. For instance, $x^2 - y^2$ is equal on the sphere to $1-2y^2 - z^2$ since $x^2 = 1 - y^2 - z^2$. This observation will help us to understand the general situation!

You can figure out why you need 2n +1 additional harmonics to go from degree n-1 to degree n as follows: If, in a degree n term, there is a power of x that is two or higher, it can be reduced, using the $x^2 + y^2 + z^2 = 1$ relation, to a term of lower degree together with two terms still of total degree n but of degree in x lowered by 2 compared to the original term, i.e., by noting that $x^k y^l z^m = (1 - y^2 - z^2) x^{k-2} y^l z^m$. So everything of degree n will be taken care of provided one adjoins to the degree n-1 or less polynomials enough things to take care of terms of the form $y^l z^{n-l}$ and of the form $xy^l z^{n-l-1}$. There are 2n +1 of these terms, so in terms of dimension we increase the dimension of the polynomial functions on the sphere by no more than 2n +1 as we go from degree n-1 to degree n. It turns out that we need at least 2n+1 so that this is in fact the exact number we have to have. This is because the 2n +1 functions of degree n discussed together with the corresponding ones of lower degree really are an independent set on the sphere.(The details of this and the proof will be given later, to avoid interrupting the continuity).

Expansion in Spherical Harmonics: An Example

We have set things up so that spherical harmonics up to and including degree n generate as linear combinations all the polynomial functions in x, y and z as far as values on the sphere goes. It is worth trying this out in a low degree case just to see it in action: If you take a second degree arbitrary polynomial, you can check using algebra and the fact that $x^2+y^2+z^2=1$ on the sphere that the second degree polynomial is a linear combination of the degree 0, degree 1 and degree 2 spherical harmonics AS FAR AS VALUES ON THE SPHERE ARE CONCERNED.

The point is that a lot of items that are not zero on the space as a whole actually are zero on the unit sphere! E.g., $x^2+y^2+z^2-1$. So you need fewer polynomials on the sphere to generate everything than you do if you look at the polynomials over all of space.

For instance, the function x^2 on the sphere seems not to be a linear combination, and indeed is *not* a linear combination, of the five degree 2 spherical harmonics $x^2-y^2, x^2-z^2, xy, xz, yz$—if you think of x^2 and all the rest of the items as polynomials over the whole of Euclidean x, y, z space. But x^2 is a linear combination AS IT BEHAVES ON THE SPHERE of the spherical harmonics of degree less than or equal to 2, namely using $x^2+y^2+z^2=1$ gives $x^2=\frac{1}{3}\left[1+(x^2-y^2)+(x^2-z^2)\right]$

This always works—that is what we set up our spherical harmonics to do and they do it.

The Spherical Harmonic Expansion in General

Since every continuous function on the sphere can be approximated by polynomials (standard analysis) and since spherical harmonics up to and including degree n have linear combinations that represent any given polynomial on the sphere, any continuous function on the sphere can be approximated by linear combinations of spherical harmonics. You may want to read that sentence twice!!!

But there is actually a more systematic series expansion by spherical harmonics that is a quite exact analogue of the Fourier series on the circle. It is not just some sort of abstractly guaranteed bunch of better and better approximations but is rather a systematic infinite series expansion.

To make this work out well, we need to go back to that condition on spherical harmonics that we mentioned earlier, the one about being "perpendicular." It may seem a bit odd to use such geometric terms about functions ,but this kind of thinking about functions in terms of geometry is actually a really good idea. In fact, it is the basic idea of a whole mathematical subject called "functional analysis".

First we define the idea of an "inner product" $\langle f, g \rangle$ of two functions f and g on the sphere. It is by definition what you get by integrating over the sphere the product of the functions: $\langle f, g \rangle = \int fg$. Now, we define just as in vector calculus, the norm $\| f \| = \sqrt{\langle f, f \rangle}$.

This makes the space of continuous functions look like a linear algebra item, namely a vector space with an inner product . And the norm is like length. Also we can talk about perpendicular functions; f is perpendicular to g if <f,g> =0.

Of course we can describe what it means without the geometric terminology: f is perpendicular to g if the average value on the sphere of $f \cdot g$ is 0!!

It is easy to check that one can make the spherical harmonics all unit length and make every pair of harmonics, whether of the same degree or not, perpendicular. In fact, you can do this on any vector space with an inner product. If you start with a generating set of independent vectors, or in this case, with a generating set of polynomials of degree 0, degree 1 ,degree 2, and so on, you can change each on as needed to make them all perpendicular to each other and all unit length ,too. This is done by what is known as the Gram Schmidt process.

Once we have this in sight, one way or another, the spherical harmonics expansion is just like Fourier series. Namely, an arbitrary (continuous) function f on the sphere is represented by a series $\sum_l \langle f, h_l \rangle h_l$ where h is some numbering off all of the unit length and mutually perpendicular spherical harmonics.

Of course one usually does the sum by doing degree 0, then degree1, then degree 2 etc. So the harmonics are numbered $h_{n,j}$ with n=degree and j being some index that runs from –n to +n so that there are 2n+1 of them as we say earlier.

Why does this series represent the function and in what sense? The sense is easy. While the series need not converge at every point, it converges to the function in the sense that the norm of the difference between the function and the sum of the first big bunch of terms of the series gets arbitrarily close to 0 as the bunch becomes big enough. ("Convergence in L2 norm" is the catch phrase here—it all looks just like Fourier series).

Why does this work? There is a general theorem, that Laplacian eigenfunction expansions are always convergent in our "L2" norm as it is called (on compact manifolds). [This is why I have included an Aside at the end of this article—to explain where the spherical harmonics come from in terms of harmonic polynomials—-it gives one a way to find spherical harmonics explicitly in terms of linear algebra.]

But the general theorem is hard to prove. And it turns out that in the specific case of spherical harmonics, the situation can be understood much more directly. The argument is almost exactly like the one for the corresponding result about Fourier series.

First, let's introduce a convenient notation. Let $S_f(n)$ be the spherical harmonic expansion of f up to and including harmonics of degree n. Now the way we set things up, $f - S_f(n)$ is perpendicular to the set of all polynomial functions on the sphere of degree less than or equal to n. By what amounts to a theorem of Euclidean geometry, this implies that $S_f(n)$ is the unique best approximation of f in the norm sense among all the polynomials that are of degree n or less. (Just remember Euclid's theorem that the closest point on a line to a given point not on the line is the foot of the perpendicular to the line). So $S_f(n)$ is the best L2 approximation of f of a given degree.

On the other hand, we already know that, given any amount of closeness you want, there is SOME polynomial , say of degree n , that is that close to f in the L2 norm—just take a polynomial that is close enough to f at every single point. (This is possible on account of some standard theorem from analysis called Weierstrass's theorem). But then $S_f(n)$ is that close or even closer in L2 norm to f —because $S_f(n)$ is the best approximation in the L2 sense of degree n. A detailed version of this proof outline will be added later. Anyway, the approximation result is true! And you can just take it on faith if you want to.

A partially worked out example: How the Gram Schmidt process works in low degrees, up to and including degree 2.

Working out the first few degrees may help with understanding all of this. Let us try first on the one harmonic of degree 0. We were choosing 1, but actually $\|1\|$ is not =1 because the integral $\int 1^2 = 4\pi$. So we ought to replace 1 by $\frac{1}{\sqrt{4\pi}} = \frac{1}{2\sqrt{\pi}}$ to have unit length! Set $h_{0,0} = \frac{1}{2\sqrt{\pi}}$. This is our one and only degree 0 harmonic.

The natural generating set for degree 1 is x,y, and z-- or so it seems. These are all right on being perpendicular to $h_{0,0}$: $\int x = 0$ by symmetry, so $\int (\frac{1}{2\sqrt{\pi}})x = 0$,too: and similarly for y and z. Furthermore, x and y are perpendicular to each other: $\int xy = 0$ again by symmetry (think that one over!). Similarly, y and z are perpendicular and x and z are perpendicular. But x is not unit length!! Namely $\int x^2 = \frac{4\pi}{3}$.

You can see this without any work. It is apparent by symmetry that $\int x^2 = \int y^2 = \int z^2$ while clearly $\int x^2 + y^2 + z^2 = \int 1 = 4\pi$.

So we ought to take $h_{1,-1} = \frac{1}{2}x\sqrt{\frac{3}{\pi}}$, $h_{1,0} = \frac{1}{2}y\sqrt{\frac{3}{\pi}}$, and $h_{1,1} = \frac{1}{2}z\sqrt{\frac{3}{\pi}}$.

So far, so good. Now our guaranteed generating set for moving from first degree to second degree was xy, xz, yz, y^2, z^2 . The first three turn out to be perpendicular to $1, x, y$ and z , and hence are also perpendicular to $h_{0,0}, h_{1,-1}, h_{1,0}$ and $h_{1,1}$. But there are problems: y^2 and z^2 are not perpendicular to 1, for instance. And of course we want to make everything unit norm, too.

Let us worry about what it takes to make y and z perpendicular to the lower degree items . All we need to do ,it turns out ,is to get them perpendicular to 1 by subtracting something. Namely, $y^2 - \frac{1}{3}$ is perpendicular to 1 since $\int (y^2 - \frac{1}{3}) \cdot 1 = \frac{4\pi}{3} - \frac{4\pi}{3} = 0$. So we should replace y^2 by $y^2 - \frac{1}{3}$. To make this look homogeneous of degree 2 , we could write it as $y^2 - \frac{1}{3}(x^2 + y^2 + z^2)$. Obviously, the algebra of this is going to get messy, but in principle you can make the generating set of five harmonics of degree 2 all of unit norm and perpendicular to each other and to the lower degree harmonics. It turns out to be a lot easier to make it all systematic if one does it in spherical angular coordinates—latitude and longitude in effect. You can read up on this in many sources.

People came up with another way to find spherical harmonics that is the basis for the systematic calculations in the latitude-longitude situation. But you do not really have to know this if all you care about are fairly low orders. If you are curious, you can read the aside that follows on how this other way goes, and then follow up by looking into the standardized representation in terms of what are called Legendre functions, a representation that uses the idea explained in the Aside below. But what you have already gets you about as far as you need to go to start in on the mathematics of Ambisonics, which is all that this summary was intended to do!

An Aside explaining what spherical harmonics really are!

For the circle, there are naturally arising definite items to use as harmonics, namely the sines and cosines. But on the sphere, there is no definite, familiar set of items. For example, for the harmonics of degree 2 , we could have used $x^2 - y^2, y^2 - z^2, xy, xz, yz$. There is ,however, a way to describe not the specific harmonics but at least what the whole set of linear combinations of them for a fixed degree actually is. Again, it is easiest to start with the circle $x^2 + y^2 = 1$. With α being the angle , the functions $r^n \sin n\alpha, r^n \cos n\alpha$ are polynomials in x and y that are harmonic on the x,y plane. (Recall that a harmonic function f is one that satisfies $\Delta f = 0$, where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$.

Moreover, these two harmonic polynomials generate the set of all the harmonic homogeneous degree n polynomials as linear combinations. (This is a standard fact from complex analysis).

So we can try the same thing on three-dimensional space. Namely, we can look at the harmonic homogeneous degree n polynomials and restrict them to the unit sphere to get our degree n spherical harmonics.

Now you can see how to get the number of degree harmonics to compare with the answer we got before, namely 2n +1. Think about the Laplacian acting on homogenous polynomials of degree n. It converts them to homogeneous polynomials of degree n-2. It turns out that this operation generates ALL the homogenous polynomials of degree n-2 (we are supposing n is at least 2 here). So we can find the dimension of the space of harmonic polynomials that are homogeneous of degree n by subtracting:

The space of homogeneous harmonic polynomials has dimension given by: (dimension of homogeneous polynomials of degree n) – (dimension of space of homogeneous polynomial of degree n-2).

The dimension of homogeneous polynomials in x, y and z of degree k can be computed as follows:

If x has degree j, j at least 0 but not more than k, then y can have degree anything from 0 to k-j, and then the degree of z is determined by the fact that the total degree is k.

If x has degree 0	There are k+1 possibilities
If x has degree 1	There are k possibilities
If x has degree 2	There are k-1 possibilities
If x has degree k	There is k-(k-1)=1 possibility , it is x^k.

So the total number of possibilities is the sum

$$(k+1)+(k)+(k-1)+...+(1)=\frac{(k+1)(k+2)}{2}$$

Thus the dimension of harmonic homogeneous polynomials of degree n is

$$\frac{(n+1)(n+2)}{2}-\frac{(n-1)(n)}{2}=\frac{4n+2}{2}=2n+1.$$

It works ! This is the same answer we got earlier. And the fact that we have identified the harmonics of degree n as a certain specific space of polynomials that satisfy a differential equation makes a systematic treatment with explicit formulas possible, instead of our Gram Schmidt process, which is complicated and does not lead easily to systematically labeled results, although for any particular degree you can get a basis that makes the expansion work out correctly—with enough work!

One of the really cool things about this new approach is that the fact that the different degree harmonics are perpendicular to each other is automatic! We do not have to do any Gram Schmidt maneuvering—except to get unit length perpendicular things WITHIN a fixed degree. For example, let us look at the harmonic polynomials of degrees 0,1, and 2 one more time, with a more or less randomly chosen bunch of generators:

1 for degree 0,

x,y, z for degree 1

$x^2 - y^2, x^2 - z^2$, xy, xz, and yz for degree 2.

Now we do not know about unit length nor about perpendicularity within a degree unless we compute. But we know without computing that for example $x^2 - y^2$ is perpendicular to 1 and to x , and that x is perpendicular to 1 and so on. This is a good thing, and it is true in all degrees. A degree k homogeneous harmonic polynomial is always perpendicular on the sphere to a degree n one, as long as k and n are different! There will be another item explaining why this is so later on. But anyway, you can see how much easier it makes like computationally. A lot of the messiness goes away.

----------------(End of Aside)