Frames, Quadratures and Global Illumination: New Math for Games



Robin Green – Microsoft Corp Manny Ko – PDI/Dreamworks

WARNING

- This talk is **MATH HEAVY**
- We assume you understand the basics of:
 - Linear Algebra, Calculus, 3D Mathematics
 - Spherical Harmonic Lighting, Visibility, BRDF, Cosine Term
 - Monte Carlo Integration, Unbiased Spherical Sampling
 - Precomputed Radiance Transfer, Rendering Equation

- This is bleeding edge research (like new results *last night*)
- There are still a lot of unanswered questions

Some Definitions

- \mathbb{S}^2 is the unit sphere in \mathbb{R}^3
- ξ is a point on the sphere

$$\xi = (\theta, \varphi)$$
 where
 $\theta \in [0, 2\pi[$
 $\varphi \in [0, \pi]$

$$\xi = (x, y, z)$$
 where
 $\sqrt{x^2 + y^2 + z^2} = 1$

 Right-handed coordinate system, + z is up



Spherical Harmonics

• The Real SH functions are a family of orthonormal basis function on the sphere.



Spherical Harmonics

• They are defined on the sphere as a signed function of every direction

$$y_l^m(\theta,\varphi) = \begin{cases} \sqrt{2}K_l^m \cos(m\varphi)P_l^m(\cos\theta), & m > 0\\ \sqrt{2}K_l^m \sin(-m\varphi)P_l^{-m}(\cos\theta), & m < 0\\ K_l^0 P_l^0(\cos\theta), & m = 0 \end{cases}$$

• The functions are orthogonal to each other

$$\int_{\xi \in \mathbb{S}^2} y_i(\xi) y_j(\xi) d\xi = \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$$

SH Deficiencies

- SH produces signed values yet all visibility functions, BRDFs and light probes are strictly positive.
- SH projections are global and smooth, visibility functions are local and sharp.
- SH reproduces a signal *at the limit*. There is no guarantee the result is close to the original at low orders. Even at high orders it "rings" esp when restricted to the hemisphere.



Haar Wavelets

- Haar wavelets are spatially compact and produce a lot of zero coefficients.
- Generating 6 times the coefficients, papers rely on compression and highly conditional code.
- Projecting cube faces onto the sphere introduces distortions, and seams for filtering and rotation.



Radial Basis Functions

- Radial Basis Functions are also used, usually sums of Gaussian lobes.
- Need to solve two variables direction and spread. Leads to conditional code that is not GPU friendly.
- *Zonal Harmonics* are another form of steerable RBF built out of orthogonal parts.



Smoothness vs. Localization

 Haar and SH are two ends of a continuum – one smooth and global, the other highly local and unsmooth. This is Spatial vs. Spectral compactness.



Q: What lives in the middle ground?

Spatial vs. Spectral

- It turns out, the Spatial vs. Spectral problem is exactly *Heisenberg's Uncertainty Principle*.
- You cannot have both spatial compactness and spectral compactness at the same time – e.g. The Fourier transform of a delta function is infinitely spread out spectrally.

• But... thanks to a theorem by David Slepian called the *Spherical Concentration Problem* you can get pretty close.

Fundamental Questions

- 1. Where do these Orthonormal Basis Functions come from?
- 2. How can we loosen the rules so we can define better functions for our own use cases?
- 3. What are the key properties we need to retain for our functions to be useful?

What You Need To Know

• We are going to introduce *Frame Theory* and *Spherical Quadrature,* just enough to understand two key concepts:

Parseval Tight Frames

Spherical t-Designs

Back to Fundamentals

• We choose a vector space, like \mathbb{R}^n or \mathbb{C}^n

$$x = \{x_1, x_2, \dots, x_n\}$$

where $I = \{1, ..., n\}$ is an index set, we say the space has a dimension n

• Using the rules of Arithmetic we can add and subtract vectors, or multiply and rescale them using a Scalar value:

$$x + y = \{x_1 + y_1, x_2 + y_2, \dots x_n + y_n\}$$

$$3x = \{3x_1, 3x_2, \dots 3x_n\}$$

Back to Fundamentals

• When we add an Inner Product and a Norm things get interesting:

$$\langle x, y \rangle = \sum_{i \in I} x_i^* y_i$$

$$|x| = \sqrt{\langle x, x \rangle}$$

• Now we can measure angles, perpendicularity, sizes, distance and similarity:

$$\langle x, y \rangle = 0 \Rightarrow x \perp y$$

• All of Geometry comes from these simple definitions

Hilbert Spaces

• A Hilbert space \mathcal{H} is a vector space with a finite energy

$$\sum_{i\in\mathcal{H}}\langle e_i,e_i\rangle<\infty$$

- These *finite square summable* signals termed L^2 after Lebesgue
- L^2 is the mathematical world of data we see in the real world
 - Photographs
 - Audio streams
 - Motion Capture or GPS data

Hilbert Spaces

- The field \mathbb{C} has the inner product $x\overline{y}$
- The field \mathbb{R}^n has the *dot product* defined $\sum_{i=1}^n x_i y_i$
- The infinite dimensional space of finite sequences $\ell_2(\mathbb{N})$ has the inner product $\sum_{i=1}^{\infty} x_i \overline{y}_i$
- The space of functions on the interval [*a*, *b*] called *L*²(*a*, *b*) has the standard inner product:

$$\langle f,g\rangle = \int_{a}^{b} f(x)\overline{g(x)} \, dx$$

Orthonormal Basis

• An orthonormal basis Φ for Hilbert space $\mathcal H$ is a set of vectors: $\Phi = \{e_i\}_{i\in\mathbb Z}$

where each pair of vectors are mutually orthogonal:

$$\langle e_j, e_k \rangle = \delta_{j,k}$$

$$\operatorname{span}(\Phi) = \mathcal{H}$$

- A span(x) is the set of all finite linear combinations of the elements of x

Orthonormal Bases

- For example
 - the family $\left\{\frac{1}{2\pi}e^{inx}\right\}_{n\in\mathbb{Z}}$ is an orthonormal basis for $L^2(-\pi,\pi)$ called the *standard Fourier basis* from which we get the Fourier transform.



Orthonormal Bases

- For example
 - The family of polynomials $\{1, x, x^2 \frac{1}{3}, x^3 \frac{3}{5}x, ...\}$ are the *Legendre Polynomials*, and form an orthonormal basis on the interval $L^2(-1,1)$



Orthonormal Bases

- For example
 - The family $\{e_n\}_{n=1}^{\infty}$ is an orthonormal basis on $\ell^2(\mathbb{N})$ where

$$e_1 = \{1,0,0,0,0,0,0,\dots\}$$
$$e_2 = \{0,1,0,0,0,0,0,\dots\}$$
$$e_3 = \{0,0,1,0,0,0,0,\dots\}$$

 $-\ell^2(\mathbb{N})$ is the infinite dimensional space of finite, time-related signals like audio, motion capture joints or accelerometer data.

Orthonormal Basis Characteristics

• **Projection:** Given a signal or function $f \in \mathcal{H}$

$$c_i = \langle e_i, f \rangle$$

• If e_i is a vector, this projection is a dot product. If e_i is a function in 1D this is an integral $\int_a^b e_i(x)f(x)dx$ If e_i is a function on the sphere, this integral is over the sphere \mathbb{S} $2\pi \qquad \pi$

$$\int_{\varphi=0}^{\infty} \int_{\theta=0}^{\infty} e_i(\theta,\varphi) f(\theta,\varphi) \sin \varphi \, d\theta \, d\varphi$$

Orthonormal Basis Characteristics

• Perfect reconstruction:

$$f = \sum_{i \in I} \langle e_i, f \rangle e_i \quad \text{ for all } f \in \mathcal{H}$$

• This says we can project then exactly reconstruct our signal from just it's coefficients

Orthonormal Basis Characteristics

• Parseval's Identity:

$$\|f\|^2 = \sum_{i \in I} |\langle e_i, f \rangle|^2$$
 for all $f \in \mathcal{H}$

- Sometimes called *norm preservation*, this says that the total energy in the function is the same as the magnitude of the coefficients.
 - This is a key property for a lot of algorithms. Working on coefficients is a lot quicker than working on functions.

ONB Characteristics

• Successive Approximation:

$$\hat{x}^{(k+1)} = \hat{x}^{(k)} + \langle e_{k+1}, x \rangle e_{k+1}$$

• This is a roundabout way of saying that projecting to a subset of indexes is the best approximation in a *least squares sense*.

General Bases

- We use Orthonormal Bases all the time
- Every rotation matrix in 3D is an Orthonormal Basis



General Bases

• What if you chose vectors that are not orthogonal?



General Base

 We can still represent points, but we need a "helper" basis to get us there.



General Bases

• We can now project the point $f = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

$$f' = \sum_{i=1}^{2} \langle \tilde{e}_i, f \rangle e_i$$

$$= \langle \tilde{e}_1, f \rangle e_1 + \langle \tilde{e}_2, f \rangle e_2$$

= $(1 \cdot 1 + -1 \cdot 1)e_1 + (0 \cdot 1 + \sqrt{2} \cdot 1)e_2$
= $0 \cdot e_1 + \sqrt{2} \cdot e_2$

 $= \begin{bmatrix} 0\\\sqrt{2} \end{bmatrix}$



Biorthogonal Bases

• This second "helper" matrix is called the *dual basis* $\widetilde{\Phi}$

$$\begin{array}{l} \langle e_1, \tilde{e}_1 \rangle = 1 \cdot 1 + 0 \cdot -1 = 1 \\ \langle e_2, \tilde{e}_2 \rangle = \frac{\sqrt{2}}{2} \cdot 0 + \frac{\sqrt{2}}{2} \cdot \sqrt{2} = 1 \\ \langle e_j, \tilde{e}_k \rangle = \delta_{j-k} \quad where \ \delta = \end{array}$$

• Biorthogonal bases are pairwise orthogonal and commute.

$$f = \sum_{i \in I} \langle \tilde{e}_i, f \rangle e_i = \sum_{i \in I} \langle e_i, f \rangle \tilde{e}_i$$

Matrix Notation

- Now we switch to a matrix notation.
- Every basis in $\mathcal H$ can be written as a matrix with basis vectors as columns

$$\Phi = \{e_1, e_2, e_3, \dots\}$$

$$= \begin{bmatrix} e_{1x} & e_{1y} \\ e_{2x} & e_{2y} \\ \vdots & \vdots \end{bmatrix}$$

$$p = \begin{bmatrix} x \\ y \end{bmatrix}$$

• Points are now column vectors.

Matrix Notation

• Our projection and reconstruction now turn into *operators*

$$p = \widetilde{\Phi}f$$
$$f = \Phi^*p$$

(where M^* is the transpose)

• We can now show that orthonormal bases are *self dual*:

$$\widetilde{\Phi} = \Phi$$
$$\widetilde{\Phi}\Phi^* = I$$

Breaking the Rules

• What happens if we add another vector to the basis?

$$\Phi = \{e_1, e_2, e_3\} \qquad \qquad \widetilde{\Phi} = \{\widetilde{e}_1, \widetilde{e}_2, \widetilde{e}_3\} \\ = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{bmatrix} \qquad \qquad = \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ -1 & 0 \end{bmatrix}$$

 Now we have an overcomplete system, and coordinates are now linearly dependent

Breaking the Rules



Breaking the Rules

• We can still project a point and reconstruct it

$$p = \widetilde{\Phi}f \qquad f = \Phi^*p$$

$$= \begin{bmatrix} 2 & 0 \\ -1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad = \begin{bmatrix} 0 & 1 & -1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

General Biorthogonal Bases

• Biorthogonal bases demonstrate *Perfect Reconstruction* but we lose *Norm Preservation* and *Successive Approximation*

$$f = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad ||f|| = \sqrt{(1^2 + 1^2)} = \sqrt{2}$$

$$f' = \begin{bmatrix} 2 \\ 0 \\ -1 \end{bmatrix} \qquad \|f'\| = \sqrt{(2^2 + 0^2 + (-1)^2)} = \sqrt{5}$$

Frames

• This redundant set of vectors $\Phi = \{e_i\}_{i \in I}$ is called a *frame* and the set $\widetilde{\Phi} = \{\widetilde{e}_i\}_{i \in I}$ is the *dual frame*

• Just like biorthogonal bases the frame and it's dual are interchangeable and reversible

$$f = \Phi \widetilde{\Phi}^* f$$
$$= \widetilde{\Phi} \Phi^* f$$
Mercedes Benz Frame

- Certain frames have properties that mimic Orthonormal bases.
- The *Mercedes Benz* frame has unit length elements and produces a norm 3/2 times too large:

$$\sum_{i=1}^{3} |\langle e_i, p \rangle|^2 = \frac{3}{2} ||p||^2$$

• 3/2 is the redundancy in the system.



Parseval Tight Frame

• We can factor out this constant and we end up with a frame that obeys *Parseval's identity*

$$\Phi_{PTF} = \sqrt{\frac{2}{3}} \Phi_{MB}$$

- This is called a *parseval tight frame*, or PTF.
- Parseval tight frames have all the same properties as orthonormal bases, except for *successive approximation*.



PTF-Mercedes Benz is Self Dual

• The PTF-MB basis is self dual and preserves the norm.

$$\Phi_{PTF}f = \begin{bmatrix} 0 & \sqrt{2/3} \\ -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.8165 \\ -1.1154 \\ 0.2989 \end{bmatrix} = f'$$

$$\Phi_{PTF}^* f' = \begin{bmatrix} 0 & -1/\sqrt{2} & \frac{1}{\sqrt{2}} \\ \sqrt{\frac{2}{3}} & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} 0.8165 \\ -1.1154 \\ 0.2989 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} = f$$

 $||f|| = \sqrt{2}$ ||f'|| = 1.4142

Parseval Tight Frame

- PTFs have *exact reconstruction* like orthonormal bases
- PTFs are *self dual*, so we do not need a *dual frame* to project



Frame Bounds

 A family of elements {e_n}_{n∈Z} in a Hilbert space H is a *frame* if there exists positive constants A and B such that:

$$A\|f\|^2 \le \sum_{n \in \mathbb{Z}} |\langle e_n, f \rangle|^2 \le B\|f\|^2$$

- The two values A and B are called the *frame bounds*
- Ensuring A > 0 means that the whole space is spanned
- Ensuring $B < \infty$ means the space is finite

Frame Bounds

• We can categorize frames based on their construction

| $\ e_i\ = 1$ | Unit Frame |
|---------------|----------------------|
| A = B | Tight Frame |
| A = B = 1 | Parseval Tight Frame |

• Any tight frame can be factored into a PTF

Gram Matrix

 One way to check that a frame is a tight frame is to generate the *Gram Matrix* ΦΦ*

$$M_{ij} = \langle e_i, e_j \rangle$$

• If the frame is Parseval Tight, it will have 1 in the leading diagonal and the frame bound A in the off-diagonals

$$\Phi = \{e_1, e_2, e_3, e_4\}$$

$$\mathbf{M} = \Phi \Phi^* = \begin{bmatrix} 1 & a & a & a \\ a & 1 & a & a \\ a & a & 1 & a \\ a & a & a & 1 \end{bmatrix}$$

Spherical Polynomials

- A spherical polynomial is simply an expression in (*x*, *y*, *z*) that is evaluated on the surface of the unit sphere.
- Add the highest power on each axis to find the *order* of the polynomial, e.g.

$$f(x, y, z) = 3x^2 + yz$$

is a $2 + 1 + 1 = 4^{th}$ order spherical polynomial



Integrating on the Sphere

- We have three ways of integrating over a sphere
 - 1. Symbolic integration over \mathbb{S}^2

$$\int_{\varphi=0}^{2\pi} \int_{\theta=0}^{\pi} e_i(\theta,\varphi) f(\theta,\varphi) \sin\theta \, d\theta \, d\varphi$$

2. Numerical integration using unbiased Monte Carlo

$$E(f) \approx \frac{4\pi}{N} \sum_{n=1}^{N} e_i(\xi_n) f(\xi_n)$$

Gaussian Quadrature

• If you are integrating a fixed order polynomial over a closed range, Gaussian quadrature can find the integral using a small number of evaluations



- Trapezium Rule is a quadrature for linear curves.
- Simpson's Rule is a quadrature for quadratic curves.

Spherical Quadrature

• Given a set of points and their weights, quadrature will quickly find you the integral

$$\int_{-1}^{1} f(x) dx = \sum_{j=1}^{N} w_j f(x_j)$$

- To find the integral over [a, b] we scale the range on x_j
- This also applies to integration over the sphere, sometimes termed *spherical cubature*



Spherical t-designs

- A spherical t-design is a special quadrature on the sphere where each point has the same weight $\frac{1}{N}$
- There are designs in 3D for N points from 1 to 100, the full list of known low order designs is on the web.
- A t-design can accurately integrate a spherical polynomial of order t *and below*.



Minimum Order t-designs



The Mission

- We need to find a spherical basis that is
 - Is defined natively on the sphere
 - Retains the norm as a Parseval Tight Frame
 - Allows us to select the number of coefficients
 - Is spectrally and spatially concentrated
 - Is cheap to project
 - Is cheap to rotate
 - Exhibits rotational invariance

Spherical Needlet

• Thanks to Narcowitch et al, 2005 we have the *Spherical Needlet*, a type of third generation Wavelet

$$e_i(\xi) = \sqrt{\lambda_i} \sum_{\ell=0}^d b\left(\frac{\ell}{B^j}\right) \sum_{m=-\ell}^\ell \overline{Y}_{\ell m}(\xi) Y_{\ell m}(\xi_i)$$

Where $Y_{\ell m}(\xi)$ are the complex Spherical Harmonics, *B* is the *bandwidth* and *j* is the polynomial order

Simplifications

• The product-sum of all Complex Spherical Harmonics in one "row" is just a simple Legendre polynomial:

$$\frac{2n+1}{4\pi}P_{\ell}(\xi'\cdot\xi) = \sum_{m=-\ell}^{\ell}Y_{\ell m}^{*}(\xi)Y_{\ell m}(\xi')$$

• So needlets are defined in frequency space from orthonormal parts and are natively embedded on the sphere

Legendre Polynomials

• The Legendre polys are normalized to simplify the definitions.

$$L_{\ell}(\xi' \cdot \xi) = \frac{2n+1}{4\pi} P_{\ell}(\xi' \cdot \xi)$$

 Legendre polys can be quickly generated iteratively using Bonnet's Recursion:

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

where $P_0(x) = 1$
 $P_1(x) = x$

Littlewood-Paley Decomposition

• The key part of the algorithm is the $b\left(\frac{\ell}{B^{j}}\right)$ function.

$$f(t) = \begin{cases} \exp(-\frac{1}{1-t^2}), & -1 \le t \le 1\\ 0, & \text{otherwise} \end{cases}$$
$$w(u) = \frac{\int_{-1}^{u} f(t) dt}{\int_{-1}^{1} f(t) dt}$$
$$p(t) = \begin{cases} 1, & 0 \le t \le \frac{1}{B}\\ w(1 - \frac{2B}{B-1}(t - \frac{1}{B})), & \frac{1}{B} \le t \le 1\\ 0, & t > 1 \end{cases}$$
$$b(t) = \sqrt{p(\frac{t}{B}) - p(t)}$$

• Defined as a continuous function, evaluated at integer points.

Littlewood Paley Decomposition

• LP Decomposition allows us to break down spectral space into chunks of bandwidth *B*.



Spherical Needlet

• For use in signal space, the needlet is defined as:



Spherical Needlet



What does this integrate to?



What does this integrate to?



Needlet B=2.0 and j=1



Needlet B=2.0 and j=2





-1 D





Spherical Basis

• The complete spherical basis is a set of needlets, each pointing in a quadrature direction

$$\Phi = \{e_i\}_{i \in (1,N)}$$

- 1. Needlets are a solution to the *Spherical Concentration Problem*
 - for a given bandwidth it is the most compact spatial support
- 2. The sum of needlet bases over $j = \{2,3,4,...\}$ form a tight frame on the sphere.
- 3. A needlet of order N can exactly reconstruct spherical polynomials of order N and below.



Needlet vs. SH



Monte Carlo Sampling

• Sampling needlets correctly requires non-uniform sampling



Fast Projection

- Needlets are radially symmetric ($\xi \cdot \xi_i$ is a scalar)
- The needlet function is 1D
- Approximate the needlet with a LUT, lerp the values.

Plot error of lerp LUT versus actual function.

Fast Rotation

• The same rotation idea as SH, generate a matrix that reinterprets a needlet as sums of other needlets.

$$M_{ij} = \langle e_i, Re_j \rangle$$

= $\int_{\xi \in \mathbb{S}} e_i(\xi) e_j(R(\xi)) d\xi$

- The bases e_i and Re_j differ only in the quadrature direction.
- Which falls out to be a 1D function...

Fast Rotation

• By calculating each angle offset integral and tabulating it, we can generate a rotation function.



-3

-2

Key Features of a Spherical Basis

- Radially symmetric basis
 - Allows fast projection
 - Allows fast and stable rotation
- Defined from natively embedded atoms
 - No parameterization problems
 - Use *lifting* to construct a more performant basis
 - Spherical concentration shows that localization is possible
- Using Frames
 - Allows simpler definition of the problem
 - Who needs successive approximation anyway?
Future Work

• Littlewood-Paley is just one *partition of unity* optimized for spectral concentration. Other papers have optimized for spatial and other metrics.

Key References

- D. Marinucci et al, "Spherical Needlets for CMB Data Analysis", arxiv.org/pdf/7070.0844.pdf, 2008
- F. Guilloux et al, "Practical Wavelet Design on the Sphere", Applied and Computational Harmonic Analysis, 2008
- J. Kovacevic et al, "*Life Beyond Bases: The Advent of Frames*", Signal Processing Magazine, IEEE, Vol.24, No.4, July 2007
- T. Hines, "An Introduction to Frame Theory", Aug 2009, <u>http://mathpost.asu.edu/~hines/docs/090727IntroFrames.pdf</u>

