Deringing Spherical Harmonics

Peter-Pike Sloan[1]
[1]Activision

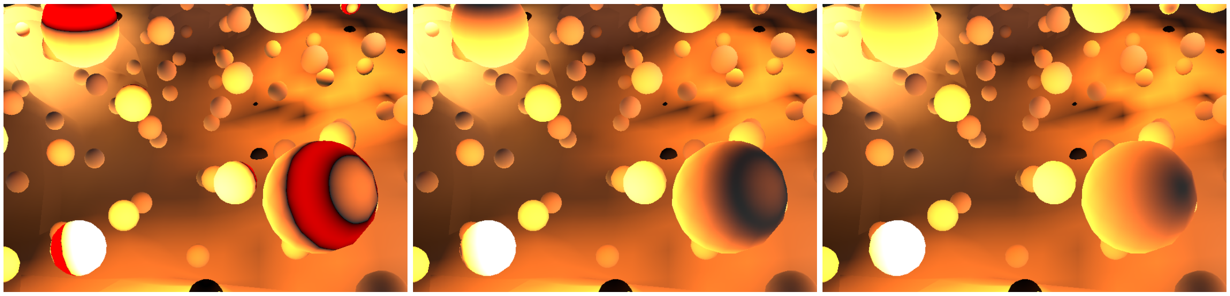

Figure 1: Light grid near a baked source without any deringing, our algorithm, and adding extra windowing from a punctual source. Negative values drawn in red.

ABSTRACT

Spherical Harmonics (SH) are a convenient basis for representing various signals in computer graphics, with lighting and visibility being the most common. While the inputs tend to be strictly positive, after projection the reconstructed function can be negative. The projection can also exhibit "ringing" artifacts, oscillations that are common with least squares. This paper presents an algorithm to efficiently and conservatively solve for a windowing function that results in a strictly positive function by minimizing a univariate polynomial that works well for irradiance signals in video games.

CCS CONCEPTS

• **Computing methodologies** → **Rendering**;

ACM Reference Format:
Peter-Pike Sloan[1], [1]Activision, . 2017. Deringing Spherical Harmonics. In *Proceedings of SIGGRAPH Asia Technical Briefs.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

Spherical harmonics (SH) are an orthonormal basis on the unit sphere. They are used extensively in interactive computer graphics[7, 8] and games [1] to encode lighting, visibility and more general light transport[10]. For lighting in video games, there are other bases that are used as well[4–6], but even then SH are often used as an intermediate basis[3]. Even when projecting strictly positive functions like lighting and visibility, the resulting projection can end up having negative values or oscillations. Oscillations are exacerbated with the move to physically based lighting that has high dynamic range. The common solution is to "window" the function[2, 8, 10], trading off blurring with aliasing. What the prior work is largely missing, and this paper addresses, is how to determine the amount of windowing to apply in a given scenario. It is often simply tweaked by hand.

In this paper we focus on the most common case in games, which is irradiance encoded as quadratic spherical harmonics. We do this by efficiently computing a conservative bound on the smallest value of the function, and search for a windowing parameter that makes this non-negative.

SIGGRAPH Asia Technical Briefs, November 2017, Bangkok Thailand

2 SPHERICAL HARMONICS

The real spherical harmonics can be expressed in spherical coordinates:

$$y_l^m = \begin{cases} \sqrt{2}K_l^m \cos(m\phi)P_l^m(\cos\theta) & m > 0 \\ \sqrt{2}K_l^m \sin(|m|\phi)P_l^{|m|}(\cos\theta) & m < 0 \\ K_l^m P_l^m(\cos\theta) & m = 0, \end{cases} \tag{1}$$

where P_l^m are the associated Legendre polynomials and K_l^m are the normalization constants

$$K_l^m = \sqrt{\frac{(2l+1)(l-|m|)!}{4\pi(l+|m|)!}}.$$

They are indexed by band l and function in a band m, where l is a non-negative integer, and m is an integer in $[-l, l]$ in band l. An order O SH consists of all the bands between 0 and $O-1$, which has O^2 basis functions. A function $f(s)$ has projection coefficients f_l^m, or using a single index f_i where $i = l(l+1)+m$. This form is convenient for symbolic computations and evaluating analytic integrals, but is expensive to evaluate at run-time. SH can also be evaluated efficiently as polynomials on the unit sphere[9]. The basis is orthogonal, closed under rotations, and can accurately represent smooth functions using a small number of bands.

2.1 Zonal Harmonics

Any function that has circular symmetry in Z projects into only the Zonal Harmonics (ZH), with one basis function per band. A SH

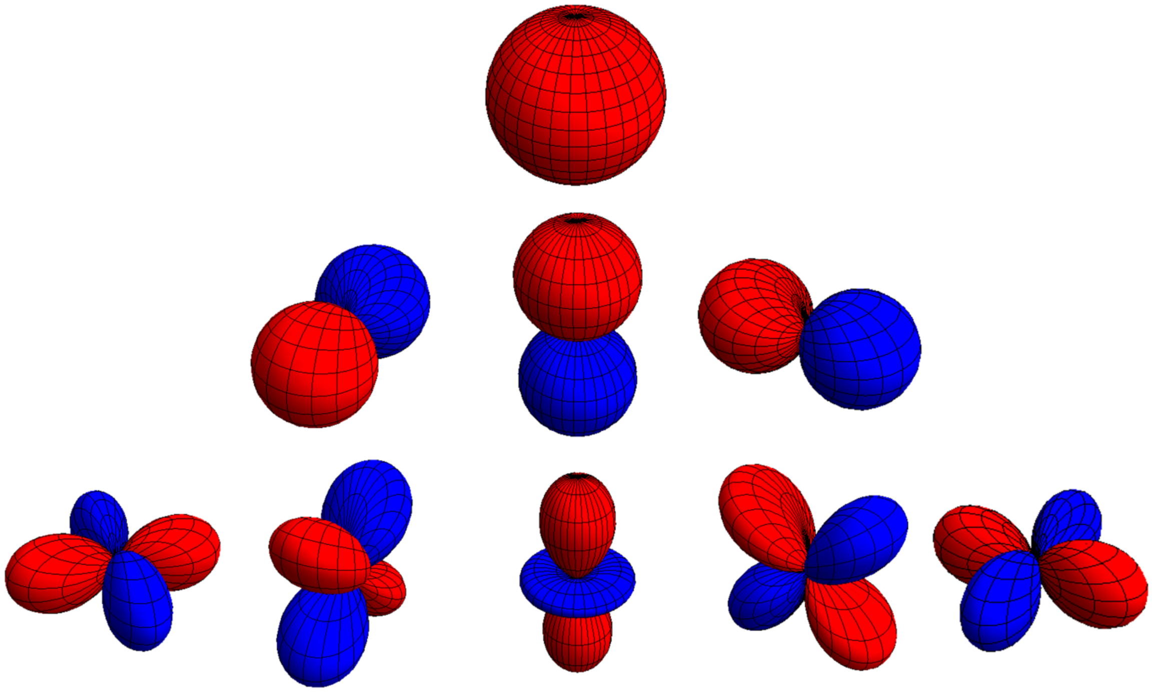

Figure 2: Rows are bands l, columns are m values going from $-l$ to l. Red is positive, blue is negative, radius is the absolute value. The central column are the Zonal Harmonics, notice how the pair of $\pm m$ basis functions are rotated copies, and the non-zonal functions have symmetric signs for all local maxima.

function f and a ZH function h can be convolved in closed form using the following equation:

$$(f * h)_l^m = \sqrt{\frac{4\pi}{2l+1}} f_l^m h_l^0 = \frac{f_l^m h_l^0}{K_l^0}. \quad (2)$$

When discussing windowing functions, we refer to the coefficient h_l^0/K_l^0 which is the per-band scaling coefficient. All of the kernels we look at have DC integrate to one, preserving the average value over the sphere.

2.2 Relevant Properties of Spherical Harmonics

The "optimal linear direction" [11] is trivial to compute from the linear SH coefficients: $(-f_3, -f_1, f_2)$. If the SH are rotated so that the "optimal linear direction" aligns with the Z axis, the non-Z linear basis functions f_1^{-1} and f_1^1 vanish.

Any pair of $\pm m$ basis functions is a rotation in Z of a single function, with a phase shift of $90/m$. When $l > 1$ and $m \neq 0$ this function also has symmetry so that for any local maxima, there is rotation around Z where the same local maxima have opposite sign. See Figure 2.

For a function f, the reconstruction of a pair of m coefficients always results in a scaled rotation of the basic function. ie: $f_2^{-2} Y_2^{-2} + f_2^2 Y_2^2 = a Y_2^{2'}$ where $Y_2^{2'}$ is a rotation in Z of the basis function. This comes from the trigonometric addition theorems, the phase shift (rotation) is not used, but the new amplitude is $a = \sqrt{f_2^{-2\,2} + f_2^{2\,2}}$.

This is just like how any sinusoid of a fixed frequency ω can always be expressed as a linear combination of a sin and a cosine of the same frequency.

2.3 Finding the minimum value of a spherical harmonic

The exact location and value of the minimum of a spherical harmonic function involves finding the roots of a quadratic polynomial in $\mathbb{R}^3$ restricted to the unit sphere. This can be done using multiple starting points and a non-linear solver[8], but is more than is needed. If you want to just know conservatively if a SH function has a negative or not, a bound can be computed accurately.

First rotate the SH into a coordinate system with the "optimal linear direction" along the Z axis. This is done once, and does not need to be inverted. This aligns Z with the linear gradient. The computation will exactly evaluate the ZH basis functions in this coordinate systems, the only remaining basis functions are the two pairs of $|m| > 0$ basis functions that will be handled using a conservative bound. For the $|m| > 0$ basis functions, any pair of $|m| = n$ are simple rotations of the same functions in Z. This means that there is some phase angle ϕ that has a maximal positive value, and due to symmetry a negative copy exists, and this contour can be used to determine the most negative location as a function of Z.

2.3.1 *Finding the minimum value of a zonal harmonic.*

A zonal harmonic of a function f is the polynomial:

$$\frac{f_0}{2\sqrt{\pi}} + \frac{f_2\sqrt{3}z}{2\sqrt{\pi}} + \frac{f_6\sqrt{5}(3z^2-1)}{4\sqrt{\pi}}. \tag{3}$$

This is a quadratic polynomial in z, and the minimum can be trivially found. Putting it in the canonical form $az^2 + bz + c$, the coefficents are: $a = \frac{f_6\sqrt{5}3z^2}{4\sqrt{\pi}}$, $b = \frac{f_2\sqrt{3}z}{2\sqrt{\pi}}$ and $c = \frac{f_0}{2\sqrt{\pi}} - \frac{f_6\sqrt{5}}{4\sqrt{\pi}}$. The minimum value of this polynomial with $1 \geq z \geq -1$ is simple to compute. If $a > 0$ the location of the minimum is at $z_{min} = \frac{-b}{2a}$, if $1 > z_{min} > -1$ simply evaluate the quadratic at that location. Otherwise the minimum is at one of the end points, which have values $a + b + c$ and $a - b + c$.

2.3.2 $|m| = 2$. The outer most $|m|$ can also be computed analytically. The polynomial form of the basis function is:

$$Y_2^2 = \frac{\sqrt{15}(x^2-y^2)}{4\sqrt{\pi}}. \tag{4}$$

The $|m| = 2$ pair of basis functions sum to a rotated version of either basis function with amplitude $q_2 = \sqrt{f_4^2 + f_8^2}$. The Y_2^2 basis function is at a maximum when $y = 0$, and due to symmetry the negative of this will be the minimum. The polynomials are on the sphere, $x^2 + y^2 + z^2 = 1$. The maximum contour can be expressed as a function of z when $y = 0$ since $x^2 = (1 - z^2)$, $\frac{\sqrt{15}(1-z^2)}{4\sqrt{\pi}}$, and the minimum is simply the negative. This is added to the ZH quadratic polynomial, where $a = a_{zh} + \frac{q_2\sqrt{15}}{4\sqrt{\pi}}$ and $c = c_{zh} - \frac{q_2\sqrt{15}}{4\sqrt{\pi}}$, and the minimum includes everything but the $|m| = 1$ pair of basis functions.

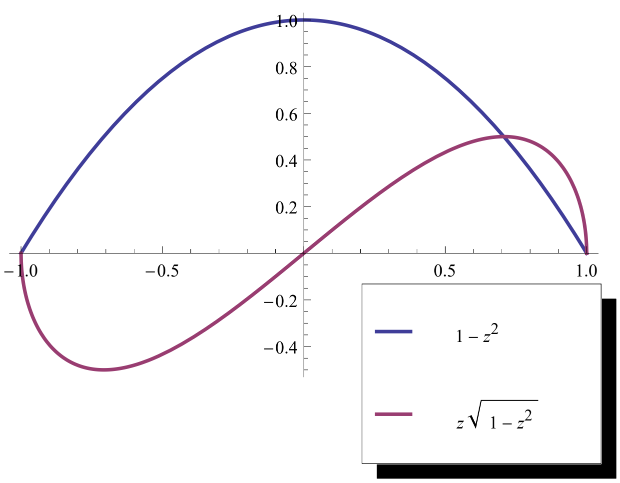

Figure 3: Basis functions for $|m| = 2$ and $|m| = 1$. $z\sqrt{1-z^2}$ is not a polynomial, but is well behaved.

2.3.3 $|m| = 1$. Up to this point, the z and minimum are exact, while the phase is unknown, but could be computed. The reason the algorithm is a conservative bound is how $|m| = 1$ is dealt with, assuming it is in the worst possible phase with respect to $|m| = 2$.

$$Y_2^1 = -\frac{\sqrt{15}xz}{2\sqrt{\pi}}. \tag{5}$$

The xz basis function is at a maximum when $y = 0$. In that case $x = \sqrt{1-z^2}$, so as a function of z it is $z\sqrt{1-z^2}$. The coefficient for this basis function is $\frac{\sqrt{15}\sqrt{f_5^2+f_7^2}}{2\sqrt{\pi}}$. This function is smooth, has a minimum value of -0.5 at $-1/\sqrt{2}$ and the minimum can be found using Newtons method, with an initial guess at the minimum of this basis function.

Given the lower bound of the $|m| = 1$ basis function of -0.5, an early out exists that can skip using Newtons method. The $|m| = 1$ basis function is positive in the upper hemisphere, which in practice is not a problem since the negative values tend to be below the "optimal linear direction". The optimization could split between the upper and lower hemispheres, forcing the function to be negative on both sides if this was an issue, or run twice with the sign of the coefficient negated.

2.4 Deringing

A thorough discussion of windowing with SH is presented in [8]. At a high level you need the coefficients per band to decrease, smoothly to zero at some cut-off frequency. For this paper, we are not explicitly evaluating a windowing function but instead using a table of windowing coefficients. We perform a binary search with linear interpolation to find the least amount of windowing that makes the resulting function non-negative. One end of the table is a delta function, so no windowing, and at the other end we have a windowing function that makes the projection of a delta function strictly positive after convolution with the clamped cosine. We window by $sinc^4$, including an extra attenuation of the linear band to make the function non-negative.

where l is the SH band index and w is the band the windowing function becomes zero, and should be greater than the cutoff band being used. The explicit values used are:

w	inf	16.7	11.3	10	9	7	5.6
l1	1	0.9941	0.9872	0.9634	0.9602	0.9184	0.8915
l2	1	0.9766	0.9493	0.9355	0.9207	0.8710	0.8030
l3	1	0.9478	0.8880	0.8584	0.8270	0.7241	0.5904

Table 1: Window size and per band scaling coefficients.

The most aggressive window makes the irradiance of a delta function have a strictly positive projection, as can be seen in Figure 4. Any positive signal can always be interpreted as a linear combination of delta functions, if those are windowed to be strictly positive, the sum will be positive, this is our most conservative filter.

$$\left(\frac{\sin\frac{\pi l}{w}}{\frac{\pi l}{w}}\right)^4, \tag{6}$$

3 DISCUSSION

Windowing values are computed separately for the RGB coefficients, and the minimum is chosen and applied to all three color

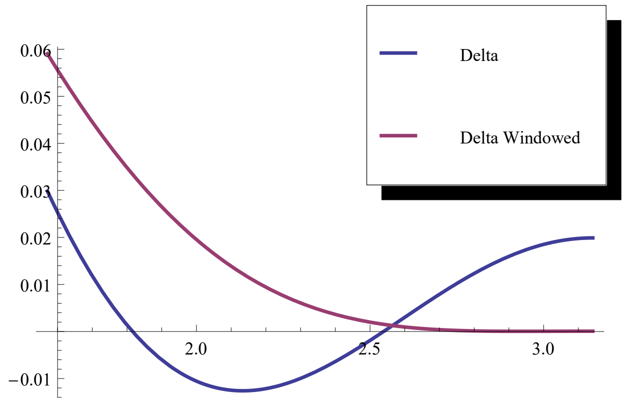

Figure 4: The irradiance of a delta function from $[\pi/2, \pi]$ has a negative and positive lobe. The windowing shown here is our most aggressive, making the projection non-negative.

channels. Filtering the color channels differently could result in color shifts otherwise. We tend to store radiance functions, but apply windowing values to them so that after convolution with a clamped cosine kernel they are non-negative.

Punctual lights can have a small positive ring, a hand tuned convolution is added to direct lights before adding them to SH, the per-band scaling coefficients for linear/quadratic are 0.92 and 0.73, while indirect light and skylight only runs the standard algorithm. Splitting the source this way and consistently windowing them separately is useful in practice.

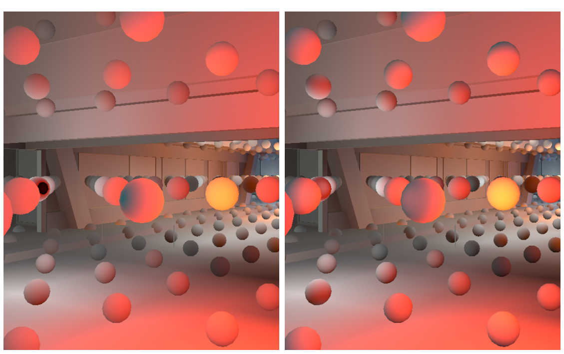

Figure 5: An earlier deringer used the ratio of linear to dc energy, but this over smoothed the lighting and left some spots negative, rendered as black. While the new deringer has strictly positive irradiance functions and preserves more angular detail.

We have gone through several algorithms to dering spherical harmonics. Using the ratio of linear to quadratic energy and a similar binary search worked well until we added linear light sources, that push a lot of energy outside of the zonal harmonics, leading us to the algorithm presented in the paper. We also started to do more multiplies with visibility functions at run-time, where having negatives in the lighting could cause more serious problems compared to just looking at raw irradiance.

4 CONCLUSIONS AND FUTURE WORK

Deringing spherical harmonics can be difficult, and we went down several blind alleys before coming up with the solution presented here. The algorithm is fast, simple to implement and proved to be robust in practice on multiple games.

These ideas could be extended to higher order spherical harmonics, where the functions with $0 < |m| < l$ would always be a conservative bound. One could try and reason about the phase differences between the $|m| = 1$ and $|m| = 2$ basis functions, generating a tighter bound. It is unclear how useful this would be in practice, since we already do some extra windowing for punctual sources.

This work only applies to spherical functions, extending these ideas to better handle hemispherical functions is an interesting avenue for future work. Finally windowing separate components (direct, indirect, possibly by intensity, etc.) more carefully would result in a non-linear filter that better respects directional features. Our extra windowing for direct lights is a crude version of this.

5 ACKNOWLEDGMENTS

Michał Iwanicki, Josiah Manson, Michał Drobot and Derek Nowrouzezahrai gave feedback on earlier versions of this document. Shan Ling found a typo in the minimum value of the $|m| = 1$ basis function.

REFERENCES

[1] Hao Chen. 2008. Lighting and Materials of Halo3. In *Game Developers Conference*.
[2] Ralf Habel, Bogdan Mustata, and Michael Wimmer. 2008. Efficient Spherical Harmonics Lighting with the Preetham Skylight Model. In *Eurographics 2008 - Short Papers*.
[3] JT Hooker. 2016. Volumetric Global Illumination at Treyarch. In *SIGGRAPH 2016 Course: Advances in Real-Time Rendering in 3D Graphics and Games*.
[4] Michał Iwanicki and Peter-Pike Sloan. 2017. Ambient Dice. In *"Eurographics Symposium on Rendering - Experimental Ideas & Implementations"*, Matthias Zwicker and Pedro Sander (Eds.). The Eurographics Association.
[5] Gary McTaggart. 2004. Half-Life 2 source shading. In *Game Developers Conference*.
[6] David Neubelt and Matt Pettineo. 2015. Advanced Lighting R&D at Ready At Dawn Studios. In *SIGGRAPH 2015 Course: Physically Based Shading in Theory and Practice*.
[7] Ravi Ramamoorthi and Pat Hanrahan. 2001. An efficient representation for irradiance environment maps. In *SIGGRAPH 2001 Conference Proceedings, August 12–17, 2001, Los Angeles, CA*, ACM (Ed.). ACM Press, pub-ACM:adr, 497–500.
[8] Peter-Pike Sloan. 2008. Stupid Spherical Harmonics (SH) Tricks. In *Game Developers Conference*.
[9] Peter-Pike Sloan. 2013. Efficient Spherical Harmonic Evaluation. *Journal of Computer Graphics Techniques (JCGT)* 2, 2 (8 September 2013).
[10] Peter-Pike Sloan, Jan Kautz, and John Snyder. 2002. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Transactions on Graphics* 21, 3 (July 2002).
[11] Peter-Pike Sloan, Ben Luna, and John Snyder. 2005. Local, Deformable Precomputed Radiance Transfer. *ACM Trans. Graph.* 24, 3 (July 2005), 9.