Approximate importance sampling of functions reconstructed from spherical harmonics

Martin Berger
Charles University
Prague, Czech Republic
maca.berger@gmail.com

ABSTRACT

The ability to generate random samples that match a spherical PDF given in terms of spherical harmonic coefficients is very important in many fields of computer graphics. Recent work has shown that generating such samples can be done efficiently, but the published methods are not robust in the presence of reconstruction errors which manifest themselves as negative values of the PDF. In our paper, we extend the approach so that it can handle such errors, and generates uniform distribution of samples in the negative parts of the sampled function while preserving a distribution that matches the original function elsewhere. The overall distribution approximates the original function and guarantees that there are no parts of the spherical domain which remain unsampled. This property makes the scheme suitable for use in unbiased Monte Carlo rendering.

Keywords: Monte Carlo rendering, importance sampling, spherical harmonics.

1 INTRODUCTION

Spherical harmonics are a set of functions $y_l^m(\theta,\phi)$ which form a basis of square-integrable functions defined over the spherical domain. Thus, any such function can be represented as a series of coefficients in this basis. In addition, spherical harmonics have some interesting properties, such as support for rotations and convolutions, which may favor them over other similar bases. This lends to many applications in computer graphics, where functions defined over the sphere or hemisphere are very common. BRDFs ([4]), precomputed radiance transfer ([8]) and irradiance environment maps ([6]) are examples of such functions.

Recently, an efficient strategy for importance sampling of functions given as spherical harmonics coefficients has been introduced in [3]. The ability to effectively produce high quality sample distributions broadens the scope of applications of spherical harmonics to other fields of computer graphics such as unbiased Monte Carlo rendering.

Spherical harmonics are not without limitations, though. The projection of a band-unlimited function to spherical harmonics will yield an infinite sequence of non-zero coefficients, which for practical purposes needs to be truncated. This step introduces errors to the reconstructed function, which manifest themselves as the so-called ringing artifacts. Specially, for a strictly positive function f, its reconstruction $\hat{f}$ can have parts with negative values. The importance sampling scheme of Jarosz et al. is particularly sensitive to this kind of problem, because the hierarchical warping process used to generate the samples is undefined for negative values (negative values can't be used to construct a valid PDF[1]). Simple clamping of the negative values to zero will lead to bias as there will be parts of the function's domain which won't receive any samples. The authors recommend adding a positive offset to the function, but it is not clear how to find a suitable value for the offset. If the offset is set too high, it will prevent negative reconstruction issues but at the same time it will degrade the quality of the resulting distribution (it will tend towards globally uniform distribution).

On the other hand, some applications might not need a sample distribution that exactly matches the reconstructed function. During our work on unbiased Monte Carlo rendering, we faced the problem of importance sampling a local radiance estimate stored as a set of spherical harmonic coefficients. Here, the function we are trying to sample is inaccurate anyway, so an approximate sampling strategy is sufficient.

The contribution of this paper is a modification of the sampling scheme of Jarosz et al., which overcomes the reconstruction problems without function offsetting. Our method doesn't always generate samples that match the sampled function closely, but avoids bias from negative reconstruction values.

2 RELATED WORK

2.1 Background

Real spherical harmonic basis functions are defined by:

[1] Probability density function

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

where K_l^m are constants and P_l^m are the associated Legendre polynomials. For a detailed description of spherical harmonics and their properties, see [7] or [2].

Due to orthogonality, the coefficients of a function f projected onto the spherical harmonic basis can be obtained from:

$$c_l^m = \iint_{4\pi} f(\theta,\phi) y_l^m(\theta,\phi)\, d\theta\, d\phi \tag{2}$$

For practical purposes, we truncate the series by setting $y_l^m = 0$ for $l > N$, where N is a pre-determined maximum band. During reconstruction, we approximate the original function by summing the basis functions weighted by the coefficients c_l^m:

$$f(\theta,\phi) \approx \hat{f}(\theta,\phi) = \sum_{l=0}^{N} \sum_{m=-l}^{l} c_l^m y_l^m(\theta,\phi). \tag{3}$$

Mathematically, truncated spherical harmonics expansion can be shown to be the minimizer of the least squares error functional:

$$\int_{4\pi} (f(\Omega) - \sum_{l=0}^{N} \sum_{m=-l}^{l} c_l^m y_l^m(\Omega))^2\, \mathrm{d}\Omega. \tag{4}$$

Minimizing the square of the error allows the result to oscillate about the original function which gives rise to the so called ringing artifacts. There are a number of techniques how to reduce this effect, for example filtering the resulting coefficients, using constrained least squares projection or offsetting the function before its projection. A survey of these techniques along with a rigorous mathematical description of the problem can be found in [5], [1] and [7].

However, none of these techniques can guarantee a non-negative reconstruction $\hat{f}$ for an arbitrary non-negative function f and arbitrary maximum band N in general.

2.2 Hierarchical sampling

Here, we give a brief overview of the sampling scheme introduced in [3]. The process starts with a uniform sample distribution over the whole surface of the sphere. In the second step, we split the domain into four quadrants and compute the integrals of the function over these sub-domains. The four computed values serve as an importance function, which is used to warp to sample set. This step is then recursively repeated on the four quadrants.

Technically, the warping step is accomplished by doing a warp along the vertical axis first and then along the horizontal axis. For a domain T and its quadrants A, B, C, D (see Figure 1), this means we compute the integrals $I_1 = I_A + I_B$ and $I_2 = I_C + I_D$ of the reconstructed function and warp the set of the samples according to probabilities $p_{AB} = \frac{I_1}{I_T}$ and $1 - p_{AB}$. Warping along the horizontal axis is analogous. The effect of the warping step is that more samples are placed in areas with large values of $\hat{f}$.

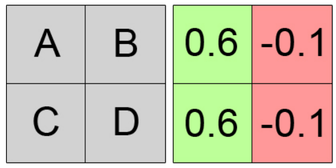


Figure 1: Left: definition of quadrants and integrals of the corresponding domains used in the text. For visualization purposes, we have mapped the spherical surface domain to a square. T denotes the union of all A, B, C, D. Right: one of the possible scenarios where some of the integrals are negative.

Warping continues in this fashion recursively up to a predefined maximum warping depth. The PDF of each sample is then computed from the ratio of the integral over the node containing the sample and the integral over the whole sphere.

This method generates samples that are distributed exactly proportionally to values of the reconstructed function $\hat{f}$ as long as the reconstruction is positive. However, once we encounter negative values for the integrals, we cannot perform the warping step and the scheme breaks. The authors propose adding a positive offset to the function before projection, but finding a suitable value for this parameter automatically is an open problem.

3 OUR APPROACH

Instead of trying to avoid negative reconstructed values completely, we use different rules during the warping process so that it can handle them in an unbiased way.

3.1 Warping step

The basic warping step is similar to [3]. First, the samples are warped along the vertical axis and then along the horizontal axis. As opposed to the original approach, we don't use the values of the integrals I_1, I_2, and corresponding probabilities $p_{AB} = \frac{I_1}{I_T}$, $p_{CD} = 1 - p_{AB}$ directly, but rather we use the values

$$\hat{p}_{AB},\ \hat{p}_{CD} = 1 - \hat{p}_{AB} \tag{5}$$

, where

$$\hat{p}_{AB} \text{ is } p_{AB} \text{ clamped to the } [\varepsilon, 1-\varepsilon] \text{ range} \tag{6}$$

for $0 < \varepsilon \leq \frac{1}{2}$. Warping along the second axis is analogous.

Our observation is that this enables us to continue warping even if some of the integrals I_A, I_B, I_C, I_D are negative, but only as long as the total integral I_T is positive. In effect, we modify the function we are trying to sample so that it has positive values of the respective integrals. If the total integral I_T is negative, we terminate the recursion immediately, which leaves the sample uniformly distributed in the domain of T as we have no suitable definition of corresponding sample distribution in this case.

Our scheme guarantees that we always get valid sample distributions and that there are no areas completely without any samples. This follows from the fact that at each warping level, the probability of each quadrant is at least ε^2, so for K levels of recursion, we have $p_X \geq \varepsilon^{2K} > 0$ for all respective sub-regions X of $\hat{f}$. This along with the fact that we can compute the PDF of each sample exactly means that the importance function is nonzero over the whole domain and the Monte Carlo estimator remains unbiased for any $\varepsilon \in (0, \frac{1}{2}]$.

3.2 Sample PDF

The PDF of each sample after the warping step can no longer be computed simply as the integral of the containing node divided by the total integral. This is because our clamping rule diverts the PDF of generated samples from the original function. Instead of the original calculation, we compute the final PDF incrementally during the recursion. Each warping step modifies the probability of a given quadrant from the original $\frac{1}{4}$ to $\hat{p}_h \hat{p}_v$ for the respective horizontal and vertical probabilities computed from $\hat{f}$. Therefore, we need to scale the sample PDF by the factor $\frac{\hat{p}_h \hat{p}_v}{\frac{1}{4}}$ for each warping level.

If we start with a PDF of a uniform distribution over the whole spherical domain, the final PDF of the sample (after k levels of warping) will be:

$$\frac{1}{4\pi} \prod_{l=1}^{k} \frac{\hat{p}_h \hat{p}_v}{\frac{1}{4}} = \frac{4^k}{4\pi} \prod_{l=1}^{k} \hat{p}_h \hat{p}_v \tag{7}$$

3.3 The role of ε

The value of ε generally affects the uniformity of the resulting distribution.

Setting ε near zero will yield a distribution, whose PDF matches the original function very closely, but very few samples will be in the regions of negative reconstruction. In the limit case of $\varepsilon = 0$, our method will return the same sample distribution as the original method of Jarosz et al. for functions which do not exhibit negative reconstruction issues.

On the other hand, setting $\varepsilon = \frac{1}{2}$ will yield globally uniform distribution, as the probabilities will be equal in each warping step.

In our rendering system, where we sample functions that approximate local radiance estimates, we use a value of $\varepsilon = 0.01$ so that the sample distributions match the functions closely.

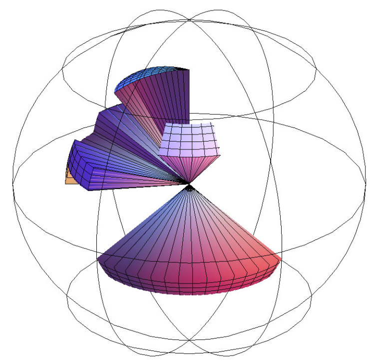


Figure 2: The original non-negative function (before projection) used for evaluation of our method. The blocky behavior and discontinuities are particularly difficult for spherical harmonics and severe ringing artifacts can be expected upon projection and reconstruction of this function.

4 RESULTS

Figure 3 shows distributions obtained with our method and with the original method from [3] with offsetting. The same number of generated samples is shown for both methods. After reconstruction, our function from Figure 2 exhibits ringing artifacts and has parts with negative values. Note that function offsetting causes the distribution to be much more uniform than the distribution from our method.

In our case, where we used the proposed method for importance sampling of local radiance estimates, the distribution generated with our method resulted in faster convergence, because fewer samples were sent to insignificant directions.

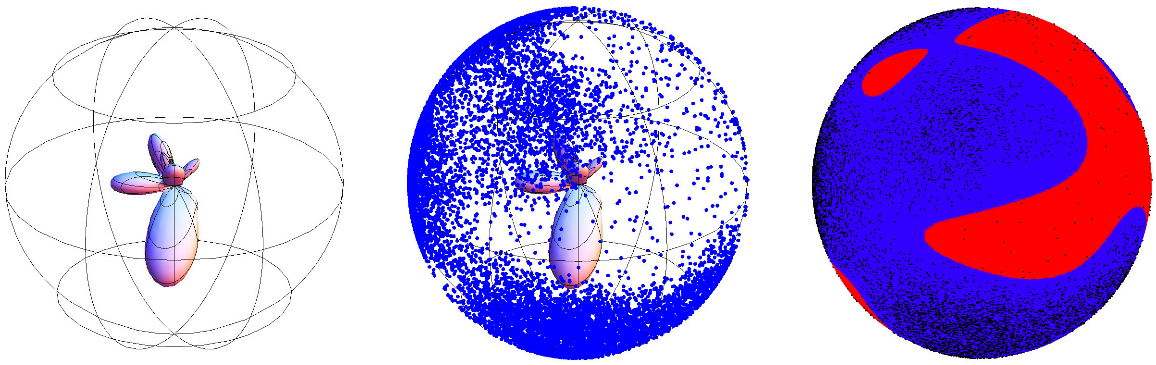
5 CONCLUSION

In our paper, we introduced a method for sampling functions given in terms of spherical harmonic coefficients, which, unlike previous methods, is robust in the presence of reconstruction errors.

The distribution generated with our method will be warped according to the sampled function in its regions of positivity, and will be uniform in its negative regions. Also, there is virtually no memory requirements or performance penalty associated with our modifications.

ACKNOWLEDGEMENTS

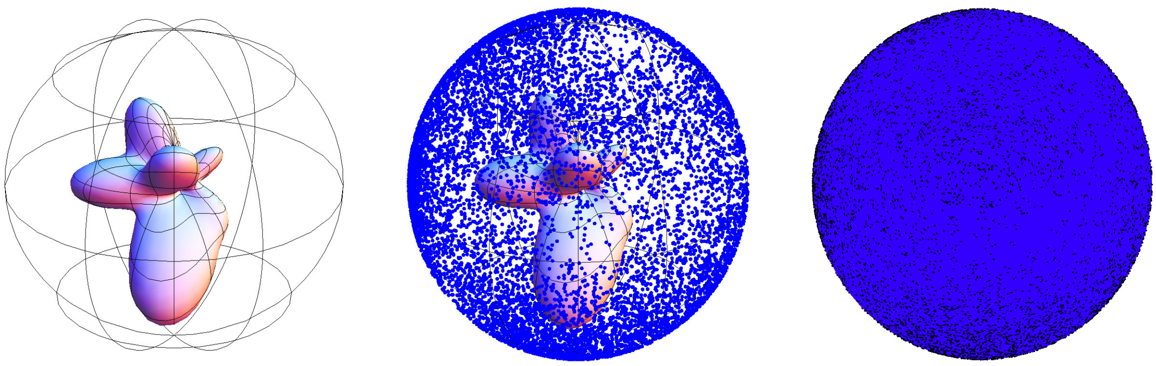
We would like to thank Alexander Wilkie, who provided valuable discussion and insights.



(a) The function reconstructed from projection to spherical harmonics using six bands.

(b) Reconstructed function along with samples generated by our scheme with $\varepsilon = 0.1$.

(c) Negative (red) and positive (blue) parts of the reconstructed function.



(d) Reconstructed function with offsetting. The minimum offset required to make the reconstruction positive across the whole spherical domain was determined by trial and error.

(e) Reconstructed function with offsetting along with samples generated by the original method of Jarosz et al.

(f) Negative (red) and positive (blue) parts of the reconstructed function.

Figure 3: A comparison of our method and the original method of Jarosz et al. The first row shows results obtained with our method. Note that the reconstructed function has large parts with negative values and that these regions do receive a fraction of the samples. On the contrary, to achieve non-negativity of the reconstructed function with the original method (the second row), a comparatively large offset value was needed, and the resulting distribution is much more uniform as a result.

REFERENCES

[1] J. P. Boyd. *Chebyshev and Fourier Spectral Methods*. Dover Publications, New York, 2001.

[2] R. Green. Spherical harmonic lighting: The gritty details. *Archives of the Game Developers Conference*, March 2003.

[3] W. Jarosz, N. A. Carr, and H. W. Jensen. Importance Sampling Spherical Harmonics. *Computer Graphics Forum (Proc. Eurographics EG'09)*, 28(2):577–586, 4 2009.

[4] J. Kautz, P. P. Sloan, and J. Snyder. Fast, arbitrary brdf shading for low-frequency lighting using spherical harmonics. In *EGRW '02: Proceedings of the 13th Eurographics workshop on Rendering*, pages 291–296, Aire-la-Ville, Switzerland, Switzerland, 2002. Eurographics Association.

[5] R. G. McClarren, C. D. Hauck, and R. B. Lowrie. Filtered spherical harmonics methods for transport problems. In *Proceedings of the International Conference on Mathematics and Computational Methods and Reactor Physics, American Nuclear Society*, 2009.

[6] R. Ramamoorthi and P. Hanrahan. An efficient representation for irradiance environment maps. In *SIGGRAPH '01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 497–500, New York, NY, USA, 2001. ACM.

[7] P. P. Sloan. Stupid spherical harmonics (sh) tricks. *Game Developers Conference*, February 2008.

[8] P. P. Sloan, J. Kautz, and J. Snyder. Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. *ACM Trans. Graph.*, 21(3):527–536, 2002.