Approximating the Reflection Integral as a Summation: Where did the $d\omega_i$ go?

Aner Ben-Artzi
Columbia University

October 7, 2005

Abstract

In this note, I explore why the the common approximation of the reflection integral is not written with a delta omega-in ($\Delta\omega_i$) to replace the differential omega-in ($d\omega_i$). After that, I go on to discover what really happens when the sum over all directions is reduced to a sum over a small number of directions. In the final section, I make recommendations for correctly approximating the reflection sum, and briefly suggest a possible framework for multiple importance sampling on both lighting and brdf.

1. Introduction

The reflection equation gives the outgoing radiance at a point, $I(\vec{x})$, as an integration of the incoming radiance over all the directions of the upper hemisphere of a point:

$$I(\vec{x}) = \int_{\Omega_{2\pi}} R_{\vec{x}}(\omega_i) F_r(\omega_o(\vec{x}), \omega_i) d\omega_i \tag{1}$$

The radiance of the upper hemisphere is different, depending on the normal of the surface at $\vec{x}$. Similarly, the outgoing (view) direction is dependant on the normal at $\vec{x}$.

We can rewrite this integral so that the directions are in world-coordinates, making the brdf, a 6-D function, and F_r at a particular view-direction and surface location is a 2D cross-section of it:

$$I(\vec{x}, \omega_o, \vec{n}) = \int_{\Omega_{4\pi}} R(\omega_i) F_r(\omega_o, \omega_i, \vec{n}) d\omega_i \tag{2}$$

where we have folded the cosine falloff into F_r and ignored visibility[1]. Most rendering approaches will next turn this integral into the following sum:

$$I(\vec{x}, \omega_o, \vec{n}) = \sum_i R_i F_r(\omega_o, \omega(R_i), \vec{n}) \tag{3}$$

Here, $\{R_i\}$ is a set of point lights that are chosen as a replacement for the incoming radiance of the upper hemisphere. Keep in mind that while this seems like environment map approximations, we have not made any assumptions about the continuity of lighting within the scene. So this is fully general for any complex lighting. Where did $d\omega_i$ go? When an integral is converted into a sum, the differential quantity

[1] If you really want to include visibility in this treatment, you may assume that F_r is also a function of $\vec{x}$, and that it includes the visibility of the scene.

should be replaced with a delta of a finite width, such that the sum of the deltas is equal to the integral of the differential:

$$d\omega_i \Rightarrow \Delta\omega_i \mid \sum_i \Delta\omega_i = \int_\Omega d\omega_i \qquad 4$$

Clearly, some assumption has been made to eliminate the need for such a $\Delta\omega_i$. In the next section, we will show what this assumption is, and that it is not always valid.

2. Looking for $\Delta\omega_i$

We start by expanding the integral into its two dimensions:

$$I(\vec{x},\omega_o,\vec{n}) = \int_0^\pi \int_0^{2\pi} R(\theta_i,\phi_i)\, F_r(\omega_o,\theta_i,\phi_i,\vec{n})\, d\phi_i \sin\theta_i d\theta_i \qquad 5$$

now we partition each dimension into equal slices, and convert the above integral into a proper sum:

$$I(\vec{x},\omega_o,\vec{n}) = \sum_i \sum_j R\big(\theta_i,\phi_j\big)\, F_r\big(\omega_o,\theta_i,\phi_j,\vec{n}\big)\Delta\phi_j \sin\theta_i \Delta\theta_i \qquad 6$$

Just to be explicit about this conversion, we recall that it is only valid if both $R(\theta_i,\phi_i)$ and F_r are approximately piecewise constant over the domains $\Delta\theta_i \times \Delta\phi_j$. (A point light is one example where this assumption only holds as i and j approach infinity.)

We now rewrite the above double-summation as a single summation by replacing the two indices i and j with a single index, $k_{i,j}$ and the two deltas with a single width, $\Delta\omega_{k_{i,j}}$, where $\Delta\omega_{k_{i,j}} = \Delta\phi_j \sin\theta_i \Delta\theta_i$. For a cleaner presentation, we will no longer use the sub-indices when referring to k. Other 2D domains may be chosen besides the standard theta-phi partitioning. In those cases, each partition, ω_k, would be accompanied by the appropriate weight, $\Delta\omega_k$, indicating the solid-angle it subtends. If an equal-area partitioning is chosen[2], all of the weights would be $\frac{4\pi}{|k|}$.

$$I(\vec{x},\omega_o,\vec{n}) = \sum_k R(\omega_k)\, F_r(\omega_o,\omega_k,\vec{n})\,\Delta\omega_k \qquad 7$$

At this point, we have to be careful about the units of each element in the summation, so we write them out explicitly:

$$I(\vec{x},\omega_o,\vec{n})\Big[\tfrac{W}{m^2 sr}\Big] = \sum_k R(\omega_k)\Big[\tfrac{W}{m^2 sr}\Big] F_r(\omega_o,\omega_k,\vec{n})\Big[\tfrac{1}{sr}\Big]\Delta\omega_k\,[sr] \qquad 8$$

When we compare this to the common two-element dot-product, it becomes clear that the lighting term must be an irradiance, not radiance. We will come back to this shortly.

[2] equal-area partitions of the sphere are impossible, but many approximations exist

The above sum can be rewritten as the following matrix triple-product (where a dot has been inserted to improve readability, but should be taken as the standard matrix multiplication):

$$R \cdot diag(\Delta\omega_k) \cdot {F_r}^T \qquad , 9$$

where vectors are row-vectors, and $diag(\vec{v})$ is the square, diagonal matrix defined as:

$$diag(\vec{v})_{s,t} = \begin{Bmatrix} if\, s = t, \vec{v}_s \\ otherwise, 0 \end{Bmatrix} \qquad 10$$

In pictorial form, this look like:

$$\begin{bmatrix} \cdots & R & \cdots \end{bmatrix} \begin{bmatrix} \ddots & 0 & 0 \\ 0 & \Delta\omega_k & 0 \\ 0 & 0 & \ddots \end{bmatrix} \begin{bmatrix} \vdots \\ F_r \\ \vdots \end{bmatrix} \qquad 11$$

Recall that the common two-element dot-product uses irradiance, instead of radiance. This can be achieved by multiplying the weights into the lighting term to yield:

$$(R \cdot diag(\Delta\omega_k)) \cdot {F_r}^T \equiv R^\Delta {F_r}^T = \sum_k R_k^\Delta F_{r,k} \qquad 12$$

Result of omitting the weights: So in order to omit the weights without loss of accuracy, one of two conditions must hold:

1. All of the weights are equal, as a result of equal-area partitioning of the sphere.
2. The lookup for the lighting term is not into the original incoming radiance, but rather into a constructed irradiance map that has pre-integrated the area of each partition, to weight it accordingly.

All we have done so far is rewritten the original integration as a sum. At this point $|k|$ is still equal to $|i|$ times $|j|$, and the computation required to evaluate the above sum is quite substantial. In many cases, condition 2 holds, but not for the original k partitions.

3. Approximating the sum with fewer terms

Now we can finally attempt to reduce the size of this dot-product by selecting a set of 'lights', $\{L_\ell^\delta\}$, so that we only have to evaluate F_r at a small number of points. In order to do this, we introduce a non-square selection matrix[3] that reduces the $|k|$-element dot-product to an $|\ell|$-element dot-product, where we assume that $\ell \ll k$. Let's take a look at how the selection matrix fits into our equations:

---

[3] These are similar to incidence matrices, except that here we use them to describe the relationship between two sets of possibly equal dimensionality.

$$R^{\Delta}{F_r}^T = R^{\Delta}I_k{F_r}^T \;\tilde{\Leftrightarrow}\; R^{\Delta}SS^{\overset{*}{-}1}{F_r}^T = \left(R^{\Delta}S\right)\left(F_rS^{\overset{*}{-}T}\right)^T \equiv \overline{R^{\Delta}}\,\overline{F_r}^T \qquad 13$$

The bar over R^{Δ} and F_r indicate a reduced version of size $|\,\ell\,|$. Since S is not a square matrix, we approximate it's inverse with an appropriate pseudo-inverse.

3.1 Delta functions for point lights

For the case of a ray-tracer that simply takes point-samples of the a chosen set of directions, we can define the entries of the selection matrix as follows:

$$S_{\delta} = S_{k,\ell} = \int_{\sigma\in\Omega_{4\pi}} \omega_k L_{\ell}^{\delta} d\sigma \qquad 14$$

Let us state explicitly here that ω_k represents a box function over the solid angle subtended by the ith partition of θ cross the jth partition of ϕ. The value within the box is 1, and 0 elsewhere. L_{ℓ}^{δ} is a Dirac delta function, so that $\left\{L_{\ell}^{\delta}\right\}$ is a set of point samples. This makes the selection matrix 1 when ω_k contains L_{ℓ}^{δ}, and 0 otherwise.

It turns out that for the types of matrices that arise from point-light approximations, the pseudo-inverse of S_{δ} is its transpose. (Later, we will examine selection matrices for which this is not the case.) Pictorially, this look like:

$$\begin{bmatrix}\cdots & \cdots & R^{\Delta} & \cdots & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ \vdots\\ F_r\\ \vdots\\ \vdots\end{bmatrix} \simeq \begin{bmatrix}\cdots & \cdots & R^{\Delta} & \cdots & \cdots\end{bmatrix}\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & \ddots\\ 0 & S_{\delta} & 0\\ \ddots & 0 & 0\\ 0 & \ddots & 0\end{bmatrix}\begin{bmatrix}0 & 0 & 0 & \ddots & 0\\ 0 & 0 & S_{\delta}^{\overset{*}{-}1} & 0 & \ddots\\ 0 & \ddots & 0 & 0 & 0\end{bmatrix}\begin{bmatrix}\vdots\\ \vdots\\ F_r\\ \vdots\\ \vdots\end{bmatrix} = \begin{bmatrix}\cdots & \overline{R^{\Delta}} & \cdots\end{bmatrix}\begin{bmatrix}\vdots\\ \overline{F_r}\\ \vdots\end{bmatrix} \qquad 15$$

Rewritten as a summatin over ℓ, this becomes:

$$I(\vec{x},\omega_o,\vec{n}) = \overline{R^{\Delta}}\,\overline{F_r} = \sum_{\ell} {R^{\Delta}}_{\ell} F_{r,\ell} \qquad 16$$

A simple ray-trace into the scene would ignore the weights that were rolled into R^{Δ}, and just use the unweighted values of the radiance from the direction of R_{ℓ}. If no appropriate pre-integrated irradiance map is used, ignoring the weights assumes that the hemisphere of incoming radiance was partitioned into regions of equal solid-angles, so that each light is off by a constant factor. This constant is usually compensated for by multiplying the final result by an arbitrary constant so that the result 'looks' right. For area light-sources that are approximated by a set of equally-spaced point lights, the constant factor can be calculated as the reciprocal of the number of samples. The error of such an approximation grows as the partitions of the lights occupy increasingly different solid angles of the surface's upper hemisphere.

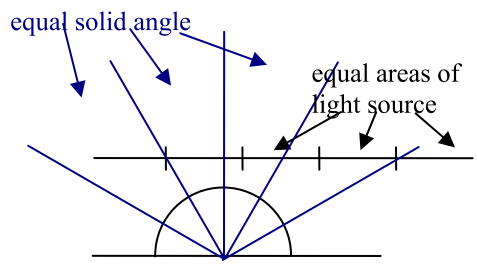

Figure 1: Area light source partitioned into equal areas, but not equal solid angles

Let's take note of some properties of S_δ, the selection matrix produced by choosing delta functions for the L_ℓ^δ's:

1. It only contains 1's and 0's.
2. It is very sparse.
3. Every row contains at most a single 1. Some rows are all 0's.
4. Every column contains exactly a single 1. (Every column sums to 1.)
5. $(S_\delta)\left(S_\delta{}^T\right) \neq [I_k]$, the identity matrix of size $|k| \times |k|$

In 15 above, we show that S_δ does not have all of it's non-zero elements near the diagonal. This can be fixed by choosing the indices, k and ℓ in a different order. However, we will soon see that it is not always possible to make the selection matrix pseudo-diagonal, so we don't bother to do that here.

By looking at the result of multiplying S_δ with it's transpose (also the pseudo-inverse), we see that the approximation to the identity alluded to in Property 5 is an identity matrix with some of the 1's zeroed out. This means that the delta L_ℓ^δ's do not preserve energy, simply ignoring parts of the incoming radiance, and the corresponding parts of the brdf. Note that even though some of the lights may have had low-energy, and ignoring them seems valid, in general, there is no guarantee that the brdf isn't very large to compensate.

Result of using Dirac delta functions: The error is directly related to the zeros on the diagonal in the approximation to the identity matrix that results from multiplying S_δ with it's transpose. More zeros indicates that more energy is lost.

3.2 Box functions for point light sources

A more sophisticated method adds up the contributions of several ω_k, and evaluates the brdf for each aggregate only once. This leads to the following definition for the selection matrix:

$$S_{box} = S_{k,l} = \tfrac{1}{\Delta\omega_k} \int_{\sigma \in \Omega_{4\pi}} \omega_k L_\ell^{box} d\sigma \qquad 17$$

L_ℓ^{box} is a binary box function. While many methods use a set of non-overlapping, domain-filling box functions to approximate lighting for the selection matrix, they still use a delta function where the pseudo-inverse should be used. This leads to a delta-style masking of the brdf.

Here is the common, but flawed formulation:

$$I(x) = R^{\Delta} S_{box} {S_\delta}^T F_r \tag{18}$$

Usually, the L_ℓ^{box}'s are pure aggregates of the ω_k's. While it is possible to imagine random overlaps between these two box functions, in practice, any given ω_k falls entirely within a L_ℓ^{box}. For example, the ω_k's may represent pixels in a cubemap, and the L_ℓ^{box}'s are contiguous groups of these pixels.

A schematic of S_{box} is shown below:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 1 & S_{box} & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{19}$$

Some properties of S_{box} are:

1. It only contains 1's and 0's
2. It is very sparse
3. Every row contains exactly a single 1. No row is all 0's.
4. Columns contain one or more 1's.
5. The transpose of S_{box} is not its pseudo-inverse

When we multiply S_{box} by the transpose of S_δ, the result is an identity matrix for which some of the 1's have been moved off the diagonal by permuting the row:

$$\text{sample } S_{box} {S_\delta}^T\text{:} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{20}$$

The error introduced by this matrix's lack of being a true identity is that some directions of the brdf are not accounted for. This error will be small when the brdf is approximately piecewise-constant over the regions defined by $\{L_\ell^{box}\}$.

4. Recommendations for future approximations

Based on the result in the previous section, we would like to recommend that any time a selection matrix is chosen to reduce the number of light directions, it's pseudo-inverse be used as a selection on the brdf.

Specifically, for S_{box}, the pseudo-inverse ends up being a low-pass filter on the brdf in regions that are aggregated. This weakens the requirement that the brdf be piecewise-constant, allowing the brdf to only be piecewise-linear within an aggregated region.

At the beginning of these derivations, we converted all directions to a global frame-of-reference. This is because the lighting is often approximated once for the whole scene, and shared among all the surface points to be shaded. An alternative would be to use a local frame-of-reference and a global view-direction. I that case, there is a global brdf that is shared among all the surfaces, and the same approximations can be made by aggregating directions based on the brdf, instead of the lighting. This is, in fact, the basis for a work we are currently producing.

Instead of creating the selection matrix for either the lighting, xor the brdf, and then using it's pseudo-inverse, one may wish to look at the product that approximates the identity matrix. Choices on deviations from the true identity matrix can be made based on both the brdf and lighting simultaneously.

5. Conclusions

In section 2, we learned that the omission of a delta term in the sum that approximates the reflection equation can lead to error. Specifically, the partitioning of the sphere of incoming radiance is left ambiguous without it. When the partitions are made explicit, as in cube-map pixels, the weights must be carefully calculated to account for variation in the solid-angle subtended by different partitions (cube-map) pixels. When this is done, a silent conversion from radiance to irradiance is made, which may become important, and should be made explicit.

In section 3, we learned that when delta functions approximate the incoming radiance, they ignore parts of the sphere. This can lead to error if the product of that radiance and the corresponding portion of the brdf is not negligible. We then discovered that even a proper box function on the lighting will place harsh restrictions on the brdf. A proper treatment of the conversion from all directions, to a smaller basis can lead to a more principled use of a smoothing filter on the brdf.