SH in Baking

If you can not render Mathematical formula, please read this image_SH_in_Baking_md_to_png

简介

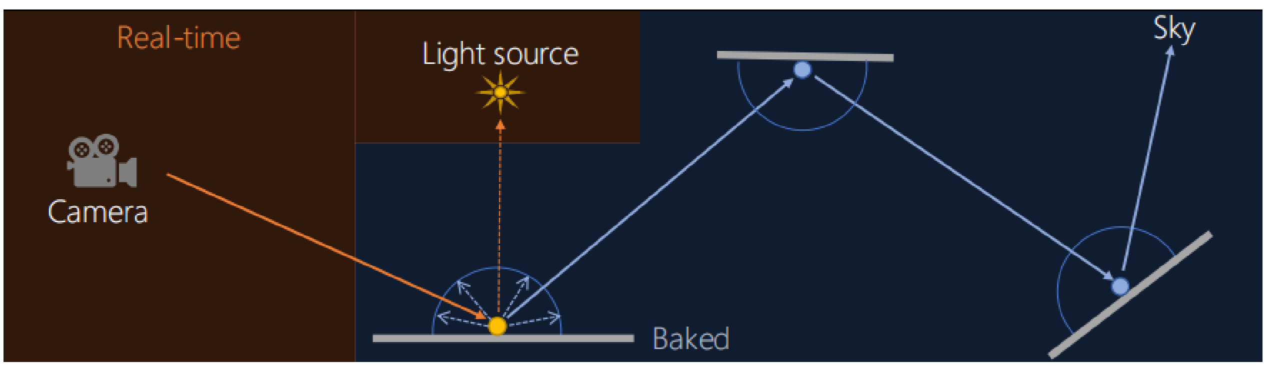
球谐函数是一种非常好的简化光照的数学模型，本文主要介绍我们为什么要在烘焙中使用球谐函数，以及我们在烘焙过程中如何计算球谐、如何简化计算过程。
简略了解球谐基础请点击光照探针中的球谐基础

目录

为什么选用 SH

路径追踪表示的 Global Illumination

先看一下全局光照（Global Illumination，后文简称 GI）在路径追踪里的表示：



我们在镜头里看见某一个位置 x 的光照就像图中显示的一样，有许多来源，并且某些光源的路径比较复杂。
我们尝试表示图中这个路径的光线：$L_o(x,\omega_o) = L_e(x) + \int_\Omega L_i(x,\omega_i) f_r(x,\omega_o,\omega_i)(\omega_i \cdot n) d\omega_i$
有这样两个问题：

- ω_o 是我们的相机位置向量，但是烘焙阶段我们并不知道具体的值
- n 是着色阶段的法向量，我们可能不知道（法线贴图和不同的 LOD 让我们不能确定法线的具体方向）

我们应该需要存储 GI数据，方便 runtime 计算得到。一个比较常见的方法就是用把 GI数据 烘焙成 辐射函数（radiance function）或者辐照度函数（irradiance function），然后使用球谐函数（spherical harmonic，后文简称 SH）模拟后，存储得到系数。
本文使用 1阶球谐函数（L1 SH）来存储辐照度数据（irradiance data）。

为什么使用 SH

寒霜的Flux Baker评估了不同的存储预计算 GI数据的方法，基于以下原因最终选择了 SH

- 不需要切线的相关计算
  - RNM or H-Basis 需要
- RGB 分离的方向光计算（direction lighting）
  - ambient + highlight direction(AHD) 做不到

- 高对比度的漫反射光
- 可以计算近似的镜面反射光
- L1 SH 在存储占比、还原消耗和还原质量表现好

球谐辐照度函数

烘焙球谐系数

使用蒙特卡洛采样（Monte Carlo Sampling）在固定点采样得到辐射值的球谐表示（SH representation of the radiance）

- 在球面（lightmaps 一般用半球）生成随机光线
- 把辐射值（radiance）在光线方向上投影成 SH基（球谐基）
- 我们得到 SH系数和球辐射函数（spherical radiance function）

光线在单位球上的投影过程： `radianceSH += 4pi/N * shEvaluateL1(rayDirection) * rayRadiance`

```
SHL1 shEvaluateL1(vec3 p)
{
    float Y0 = 0.282095f; // sqrt(1/(4pi))
    float Y1 = 0.488603f; // sqrt(3/(4pi))
    SHL1 sh;
    sh[0] = Y0;
    sh[1] = Y1 * p.y;
    sh[2] = Y1 * p.z;
    sh[3] = Y1 * p.x;
    return sh;
}
```

球谐辐射值在 L1 阶的表示：

$$L_{00} = \sqrt{\frac{1}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)$$

$$L_{1-1} = \sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)y_i$$

$$L_{10} = \sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)z_i$$

$$L_{11} = \sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)x_i$$

但是，对于实际的计算来说，我们需要计算固定法向（点法向量或者面法向量）的来自半球各个方向的光量，也就是说我们需要在几何表面固定点的辐照度（irradiance）
使用 clamped cosine lobe 卷积辐射值（radiance）得到需要的辐照度（irradiance）： $E_{lm} = A_l L_{lm}$

- 在时域做卷积，就是在频域做乘积
- 球谐的正交不变性

由这两点可以得到，radianceSH 在时域做卷积，我们转换成了在频域做乘积，直接将球谐系数乘上卷积的对象即可

```
irradianceSH = shApplyDiffuseConvolutionL1(radianceSH);

void shApplyDiffuseConvolutionL1(SHL1& sh)
{
    float A0 = 0.886227f; // pi/sqrt(4pi)
    float A1 = 1.023326f; // sqrt(pi/3)
    sh[0] *= A0;
    sh[1] *= A1;
    sh[2] *= A1;
    sh[3] *= A1;
}
```

clamped cosine lobe 是 Lambertian BRDF中 lambert漫反射余弦值的球谐系数，L1阶表示为$A_0 = \frac{\pi}{\sqrt{4\pi}}$ $A_1 = \sqrt{\frac{\pi}{3}}$

这样所有的步骤我们都清晰了，现在可以得到辐照度的 L1阶球谐表示。

简化计算

我们可以把 `shEvaluateL1(vec3)` 和 `shApplyDiffuseConvolutionL1(SHL1&)` 做一个合并简化，得到：

```
SHL1 shEvaluateDiffuseL1(vec3 p)
{
    float AY0 = 0.25f;
    float AY1 = 0.50f;
    SHL1 sh;
    sh[0] = AY0;
    sh[1] = AY1 * p.y;
    sh[2] = AY1 * p.z;
    sh[3] = AY1 * p.x;
    return sh;
}
```

我们可以看到，合并之后做乘积的因子变成了比较特殊的值 `AY0 = 0.25` 、 `AY1 = 0.50` ，这是巧合么？
显然不是，让我们以 L1阶球谐为例，在代数来解读合并的过程

球谐函数 $L(\theta,\varphi) = \sum_{l,m} L_{lm} Y_{lm}(\theta,\varphi)$

在球面上积分有 $L_{lm} = \int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi} L(\theta,\varphi) Y_{lm}(\theta,\varphi)\sin\theta, d\theta, d\varphi$

用蒙特卡洛求解可得

$$L_{lm} = \frac{4\pi}{N}\sum_{i=1}^{N} L(\theta_i,\varphi_i) Y_{lm}(\theta_i,\varphi_i) \tag{1}$$

我们展开 L1阶球谐基函数

$$Y_{00} = \sqrt{\frac{1}{4\pi}} \tag{2}$$

$$Y_{1m} = \sqrt{\frac{3}{4\pi}}(x;y;z) \tag{3}$$

将 (2)(3) 带入 (1) 可得

$$L_{00} = \frac{4\pi}{N}\sum_{i=1}^{N} L(x_i,y_i,z_i)\sqrt{\frac{1}{4\pi}}$$

$$L_{1m} = \frac{4\pi}{N}\sum_{i=1}^{N} L(x_i,y_i,z_i)\sqrt{\frac{3}{4\pi}}(x_i;y_i;z_i)$$

把 (1) 带入 $E_{lm} = A_l L_{lm}$ 可以得到辐照度的蒙特卡洛解

$$E(\theta,\varphi) = \sum_{l,m}\sqrt{\frac{4\pi}{2l+1}} A_l L_{lm} Y_{lm}(\theta,\varphi) \tag{4}$$

Lambertian BRDF 中 $A(\theta) = max[\cos\theta, 0]$ 的球谐表示为 $A(\theta) = \sum_l A_l Y_{l0}(\theta)$ ，我们可以直接写出 L1阶 SH系数

$$A_0 = \frac{\pi}{\sqrt{4\pi}} \tag{5}$$

$$A_1 = \sqrt{\frac{\pi}{3}} \tag{6}$$

写出 (4) 的 L1阶展开，得到

$$E(\theta,\varphi)=\sqrt{\frac{4\pi}{2\times 0+1}}A_0L_{00}Y_{00}+\sqrt{\frac{4\pi}{2\times 1+1}}A_1L_{1-1}Y_{1-1}(\theta,\varphi)+\sqrt{\frac{4\pi}{2\times 1+1}}A_1L_{10}Y_{10}(\theta,\varphi)+\sqrt{\frac{4\pi}{2\times 1+1}}A_1L_{11}Y_{11}(\theta,\varphi) \quad (7)$$

将 (2)(3)(5)(6) 带入 (7) 可得

$$\begin{aligned}E(x,y,z)&=\sqrt{\frac{4\pi}{1}}\frac{\pi}{\sqrt{4\pi}}\sqrt{\frac{1}{4\pi}}L_{00}+\sqrt{\frac{4\pi}{3}}\sqrt{\frac{\pi}{3}}\sqrt{\frac{3}{4\pi}}L_{1-1}y+\sqrt{\frac{4\pi}{3}}\sqrt{\frac{\pi}{3}}\sqrt{\frac{3}{4\pi}}L_{10}z+\sqrt{\frac{4\pi}{3}}\sqrt{\frac{\pi}{3}}\sqrt{\frac{3}{4\pi}}L_{11}x\\&=0.25\sqrt{4\pi}L_{00}+0.5\sqrt{\frac{4\pi}{3}}L_{1-1}y+0.5\sqrt{\frac{4\pi}{3}}L_{10}z+0.5\sqrt{\frac{4\pi}{3}}L_{11}x\\&=0.886227L_{00}+1.0233L_{1-1}y+1.0233L_{10}z+1.0233L_{11}x\end{aligned}$$

把它记作

$$E(x,y,z)=0.25\sqrt{4\pi}L_{00}+0.5\sqrt{\frac{4\pi}{3}}L_{1-1}y+0.5\sqrt{\frac{4\pi}{3}}L_{10}z+0.5\sqrt{\frac{4\pi}{3}}L_{11}x \quad (8)$$

同时，我们把辐照度函数 L1阶球谐展开得到

$$E(x,y,z)=E_{00}+E_{1-1}y+E_{10}z+E_{11}x \quad (9)$$

我们根据 (1)(2)(3) 写出 L1阶的辐射球谐函数系数的蒙特卡洛积分解

$$L_{00}=\sqrt{\frac{1}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)\quad \begin{aligned}L_{1-1}&=\sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)y_i\\L_{10}&=\sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)z_i\\L_{11}&=\sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)x_i\end{aligned}$$

将得到的结果带入 (8)(9) 就解出了辐照度系数

$$E_{00}=0.25\sqrt{4\pi}\sqrt{\frac{1}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)\quad \begin{aligned}E_{1-1}&=0.5\sqrt{\frac{4\pi}{3}}\sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)y_i\\E_{10}&=0.5\sqrt{\frac{4\pi}{3}}\sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)z_i\\E_{11}&=0.5\sqrt{\frac{4\pi}{3}}\sqrt{\frac{3}{4\pi}}\frac{4\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)x_i\end{aligned}$$

整理一下可以得到 $E(x,y,z)=E_{00}+E_{1-1}y+E_{10}z+E_{11}x$ 球谐系数的蒙特卡洛积分解

$$\begin{aligned}E_{00}&=\frac{\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)\\E_{1-1}&=\frac{2\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)y_i\\E_{10}&=\frac{2\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)z_i\\E_{11}&=\frac{2\pi}{N}\sum_{i=1}^{N}L(x_i,y_i,z_i)x_i\end{aligned}$$

前面都是基于球面做的推导，这里我们给出半球面的结果

$$E_{00} = \frac{\pi}{2N}\sum_{i=1}^{N} L(x_i, y_i, z_i)$$
$$E_{1-1} = \frac{\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)y_i$$
$$E_{10} = \frac{\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)z_i$$
$$E_{11} = \frac{\pi}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)x_i$$

数学推导过程到此结束

现在我们来处理这个得到的结果：

- 我们可以在其他地方约掉一些常量
  - BRDF 中的 $1/\pi$
  - 蒙特卡洛积分里的 4π

将这些条件带入，得到我们最终简化版的 L1 SH irradiance 蒙特卡洛积分

$$\color{green}{E_{00}} = \frac{1}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)\ \color{green}{E_{1-1}} = \frac{2}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)y_i\ \color{green}{E_{10}} = \frac{2}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)z_i\ \color{green}{E_{11}} = \frac{2}{N}\sum_{i=1}^{N} L(x_i, y_i, z_i)x_i$$

那么在单位球上的光线表示为： `irradianceSH += SHL1rgb(rayRadiance, 2 * rayRadiance * rayDirection) / N`
这就让我们发现了以下两点：

- L0阶的球谐系数包含简单的平均辐射值
- L1阶的球谐系数包含加权平均的辐射方向

为了提高存储效率，可以将 L1 SH系数中公共的 L0 SH系数和常数2给提取出来，然后把 L1 存储的数据范围限制在 $[0, 1]$

- `irradianceSH += SHL1rgb(rayRadiance, rayRadiance * rayDirection) / N`，把2给提出来
- `irradianceSH.L1 /= irradiance.L0`，将 L1约掉 L0之后再存储

那么我们在着色阶段重建 radiance 的时候也会很简单，直接把面法向量带进去即可：
`result = (0.5 + dot(irradiance.L1, normal)) * irradiance.L0 * 2.0`
有几个注意点：

- 结果可能为负，因为 L1的值很是 L0的两倍
- BRDF 里的 $\frac{1}{\pi}$ 我们已经在前面简化约分的时候用掉了
- 也输出了缺少反射因子的反射光辐射值（radiance of reflected light）