

如何看懂这些"该死的"图形学公式



FOXhunt

<https://foxhunted.github.io/>

写在前面

开始学习图形学的人，尤其是计算机专业而不是数学专业，可能都有一个感觉，很多图形学公式让人完全摸不着头脑。这些公式看起来是微积分公式，但和高数里的不太一样，也不能套用高数课本上的方法推导。不少图形学教科书写得过于简略，导致初学者看这些公式如同看天书一样。教课书里莫名其妙写了几个公式，又莫名其妙地推导出了代码，代码里面还有一些完全不知道怎么来的参数。很多人因为这样的原因在图形学的门口就跪倒了。。。

由此还产生了一些自暴自弃的想法

一、我高数是不是白学了？

比如经典的渲染等式

$$L_o = \int_{\Omega} \rho(\omega_o, \omega_i) L_i \cos(\theta_i) d\omega_i$$

虽然大学都学过基本的微积分，但这个积分怎么要算啊

二、积分其实能用乘法代替吧？！

因为这样的积分公式

$$L_{\omega_o} = \frac{1}{\cos(\theta_i)} \int_{\Omega} L(\omega_o, \omega_n) G_1(\omega_o, \omega_n) (\omega_o \cdot \omega_n) D(\omega_n) d\omega_n$$

推倒完成后写在代码里面是这样

```
float Dr = GTR1(NdotH, mix(.1, .001, clearcoatGloss));
float Fr = mix(.04, 1, FH);
float Gr = smithG_GGX(NdotL, .25) * smithG_GGX(NdotV, .25);

return ((1/PI) * mix(Fd, ss, subsurface)*Cdlin + Fsheen)
    * (1-metallic)
    + Gs*Fs*Ds + .25*clearcoat*Gr*Fr*Dr;
```

所以会觉得，都是乘法嘛。。。

三、这些参数绝对都是瞎凑出来的！！！！

$L_o(v) = \pi f(l_c, v) \otimes c_{light}(l_c \cdot n)$ diffuse公式里的 π

$$\rho(\omega_o, \omega_i) = \frac{F(\omega_o, \omega_h) G(\omega_o, \omega_i, \omega_h) D(\omega_h)}{4 |\omega_g \cdot \omega_o| |\omega_g \cdot \omega_i|} \quad \text{微表面公式里的4}$$

有很多教课书上都没写清楚这些参数是怎么来的，让初学者觉得都是瞎凑的。

然而到底有没有直观的方法给人讲解这些公式呢，肯定是有的。因为图形学用的模型大部分还是来自于几何的，只要是几何模型多少还是能通过画图的方法解释的。所以会遇到这些问题，还是因为对图形学背后的数学模型不了解造成的。写本文的目的就是试图用直观的方法解释一些图形学中的基本公式和概念，如渲染公式，微表面模型等。

从光照模型开始

既然是研究渲染公式，光照显然是其中最基本的概念，首先要做的就是对光建立一些模型。从简单的中学物理角度来看，光之所以能照亮物体就是光子从光源发出，击中物体表面又反射到观察者眼中。而物体亮度高低显然和进入眼睛中的光子数量有关系。有个专门

的物理量来描述，叫做光通量。光通量简单理解就是单位时间内某一表面接受到的光子总数量，和磁通量是类似的概念。这里要注意的是光通量不是亮度，我们通常认知的亮度应该是光通量的密度，也就是单位面积的光通量。

了解了光通量和亮度的概念之后，我们可以来观察光是如何辐射到物体表面的。先假设有一个点状光源均匀向四面八方发射光子，它发射出去的光通量记作P。再在点光源外面套一个半径为1的球体（单位球体），球体的面积为 4π ，可以算出球体上的亮度是：

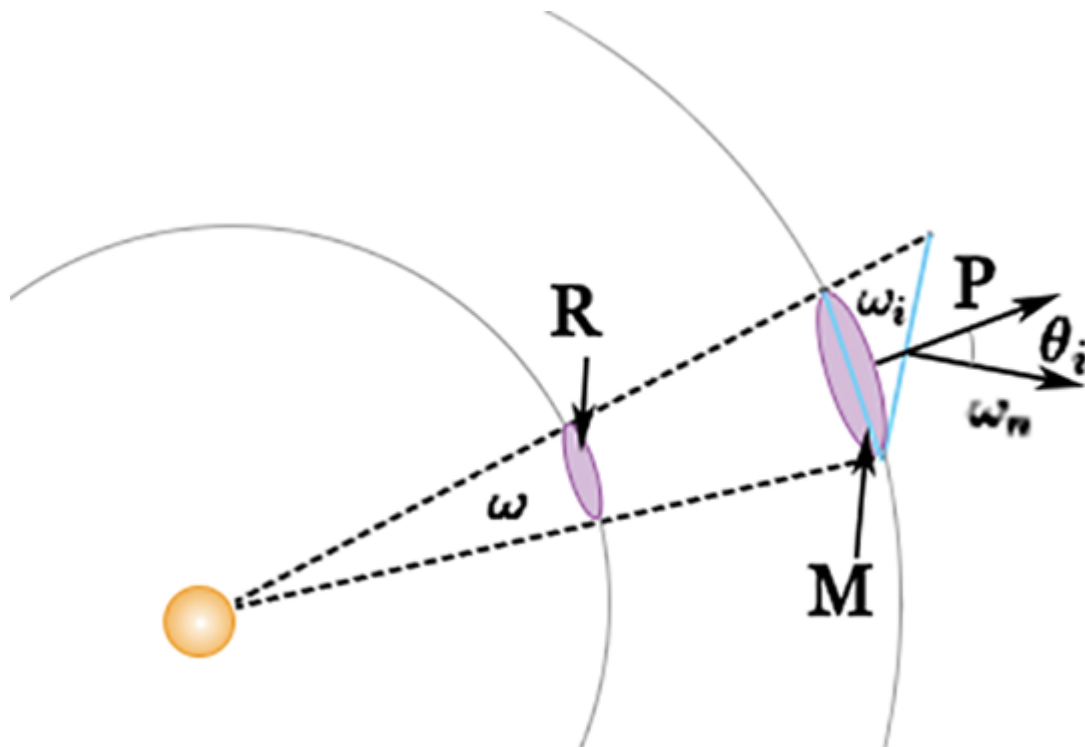
$$L_i = \frac{P}{4\pi}$$

如果在球体之外再放任意圆弧，如何求出圆弧上一点的亮度呢。这里就要作图分解一下了，假设取一个夹角 ω 。如图，只要是在夹角 ω 范围内，平行于球体表面的所有圆弧接受到的光通量应该是一样的，但随着距离增加，圆弧的面积呈平方增加。这样单位面积的光通量反而减少了，这就是中学物理中讲的光亮度在真空中会随着传播距离呈平方反比减少的原因。

另一方面如果要求的是任意平面的亮度呢。所谓任意平面也就是说受光面不一定平行于单位球面。求法是把平面先投影到球面，用投影面积作为受光面积来计算。这里要注意的是夹角 ω 不是真的有个角度，我们要求的是一点上的亮度，所以 ω 是个极小的角，有个数学名词叫做立体角，立体角在球面上覆盖的范围就是球面积分里的微元。微元的投影要用雅可比行列式计算，计算过程放在附录，有兴趣的可以看看。

从计算结果来看其实可以看成直角三角形的投影。如图所示， ω_i 是光线的方向， ω_n 是平面的法线方向， θ_i 是光线和平面法线的夹角，P是受光的平面，设平面长度为L。假设有一个平面M垂直于 ω_i ，求P到平面M的投影。简单的三角几何就能求出投影长度：

$$L_m = \cos(\theta_i)L$$



这也就是很多图形学公式里法线和光方向点乘法的由来，因为 $\cos(\theta_i) = \omega_n \cdot \omega_i$ 。综合上面的公式点光源在任意平面上一点的亮度如下：

$$I = \frac{\cos(\theta_i)}{r^2} L_i$$

另一种情况是无限光，无限光的定义就是光亮度不随距离变化，计算光照亮度时可以把 r^2 去掉。

$$I_{infi} = \cos(\theta_i) L_i$$

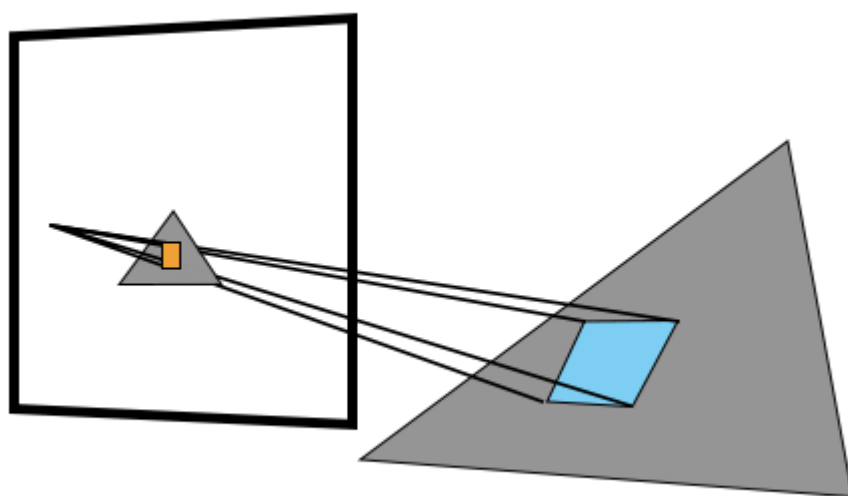
从视线到视锥到立体角

现在我们已经对光照模型有所理解了，这个模型和一般人直觉思维有所不同。中学物理对光和光传播的描述都是点和线，而图形学中讨论的光照模型多了很多和面积有关的结构。正是这些结构带来更复杂的公式。公式当然是复杂的，但我们能不能透过公式对这些结构有更直观的认识呢。一定是有的，因为我一直想绝大多数人实际上是没办法在公式这种抽象形式上思考问题的，只有很少数受过长期专门训练的人才能只通过抽象符号思维。大多数人还是习惯于有一个直观模型来作为思考的基础。

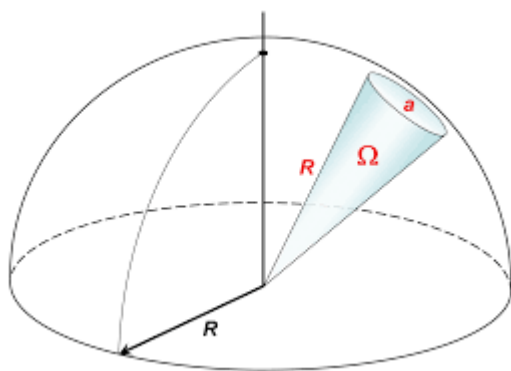
那么如何直观地描述光照模型的结构呢，还是要回到基本的光线追踪上来。之前说过光线追踪就是在屏幕外做一个虚拟的眼睛，描绘屏幕上每一个像素的时候从眼睛做一条射线通

过要绘制的像素点射入屏幕内的三维空间，射线在三维空间击中的那个点就是要绘制的目标，然后根据该点的法线和光照信息计算出屏幕上像素应该是什么样的。

这里要打脸了，射线这种说法只是一个非常简化的模型。为什么这么说呢，因为即使屏幕的像素点再小，它也是有面积的，即使缩小到所谓无穷小，也不会是零。也就是说像素永远只能是一个面而不是一个点。这样的话所谓光线追踪的射线，其实应该是一个锥形，我们称为视锥，如图所示，这个视锥和屏幕的相交面就是像素。这样我们在这个像素上呈现的也不是射线在三维空间击中的一个点，而是视锥在三维物体表面上的投影区域。同样这个投影区域的光照信息不是一个点的光照信息那么简单。另外要注意的是如果保持视锥位置不变，将三维物体沿着光锥中轴线移动，随着移动投影面积的大小也会变化。



可惜的是目前的渲染技术架构没法使用这种模型，因为这种模型的计算要对每个像素的投影区域进行采样，这种计算量在很多情况下是无法达到的。这时候图形学家就要用别的方式在达到接近的效果。现在渲染公式的基本架构就是用立体角来计算，直观的讲就是保留原来射线追踪的方式，以原来射线落点为球心面向法线方向做一个半球。在半球上来捕获面积和光辐射信息。这样做可以等同认为将投影区域面积缩小到无穷小，用半球上的立体角来计算微元投影面积。这样就保留了几何信息，因为微元面积也是有大小方向的，不像点是没有面积的。



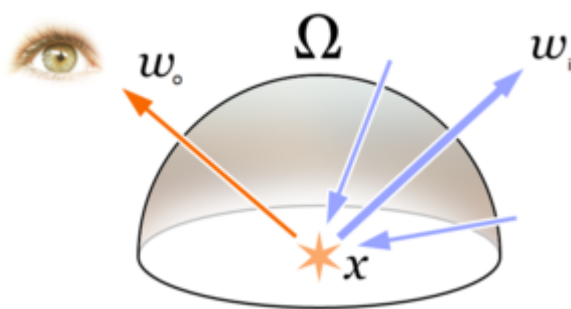
行踪神秘的PI

下面我们回到渲染公式，也是图形学里最重要的公式。

$$L_o = \int_{\Omega} \rho(\omega_o, \omega_i) L_i \cos(\theta_i) d\omega_i$$

这里的符号我们已经认识大半了， ω_i 是光线入射的方向， $d\omega_i$ 是入射光线的立体角微元面积， $\cos(\theta_i)$ 是在入射光线垂直方向上的投影， L_i 是入射光线的单位面积光通量。

剩下的就是 L_o 反射方向上单位面积光通量，也就是我们看到的亮度。 $\rho(\omega_o, \omega_i)$ 是入射光线和出射光线亮度的反射比例系数。整个积分要表达的意思就是在单位半球表面收集所有方向的入射光线亮度和面积信息以及对应的反射系数，用这些参数在半球表面的作积分。



在漫反射的情况下反射系数是个常数 $\rho(\omega_o, \omega_i) = \frac{\text{albedo}}{\pi}$ ，albedo 就是材质的颜色，这个可以理解，但是 π 是怎么来的呢。

要理解这个可以先设想一个最简单的情况，入射光遍布所有的方向，而且每个方向的亮度都是相同的常数。因为这样 ρ 和 L_i 都是常数，可以提到积分外面。公式变成：

$$L_o = \rho L_i \int_{\Omega} \cos(\theta_i) d\omega_i$$

需要的计算积分的部分只剩下 $\int_{\Omega} \cos(\theta_i) d\omega_i$

这个积分很多人都不认得，这里简单介绍一下，这是立体角积分。所谓立体角就是从球心作一个圆锥体，圆锥和球体相交的面就是立体角。 $d\omega_i$ 就是这个面无穷小时的状态，也就是球体表面的微元。如果把上面的 $\cos(\theta_i)$ 去掉， $\int_{\Omega} d\omega_i$ 其实就是求球体表面积。

所以立体角的积分计算的时候可以换成我们熟悉的球体表面积分的方法来算。

$$\int_{\Omega} d\omega_i = \int_{-\pi}^{\pi} \int_0^{\frac{\pi}{2}} \sin(\theta) d\theta d\phi$$

将 $\cos(\theta_i)$ 放回公式 $\int_{-\pi}^{\pi} \int_0^{\frac{\pi}{2}} \sin(\theta) \cos(\theta) d\theta d\phi$

这里简单推导一下

因为

$$2\sin(\theta)\cos(\theta) = \cos(\theta)\sin(\theta) + \cos(\theta)\sin(\theta) = D(\sin(\theta)\sin(\theta))$$

所以里面一层的积分可以写成

$$\int_0^{\frac{\pi}{2}} \sin(\theta)\cos(\theta) d\theta = \frac{1}{2} \left[\sin^2\left(\frac{\pi}{2}\right) - \sin^2(0) \right] = \frac{1}{2}$$

再放入外层的积分

$$\int_{\Omega} \cos(\theta_i) d\omega_i = \int_{-\pi}^{\pi} \int_0^{\frac{\pi}{2}} \sin(\theta)\cos(\theta) d\theta d\phi = \frac{1}{2} \int_{-\pi}^{\pi} d\phi = \pi$$

得到

$$L_o = \pi \rho L_i$$

这里我们考虑一下，如果所有方向都有射向球面的光，而且每个方向的光强度都等于 L_i ，那么穿过球面的所有光通量之和，应该等于光强度乘以半球面积。

$$F_i = 2\pi L_i$$

为了保持能量守恒，这所有的光通量应该全部射出球面，射出的光也应该每个方向强度都相同。上面我们已经按照公式算出一个方向的光强度 L_o ，如果所有射出方向光强度都相同的话同样可以计算射出球面的光通量：

$$F_o = 2\pi L_o = 2\pi * \pi \rho L_i$$

这里要让能量守恒，射入的光通量和射出的光通量相等 $F_i = F_o$ ，显然只有让 $\rho = \frac{1}{\pi}$ 。

又一个谜团揭开了。

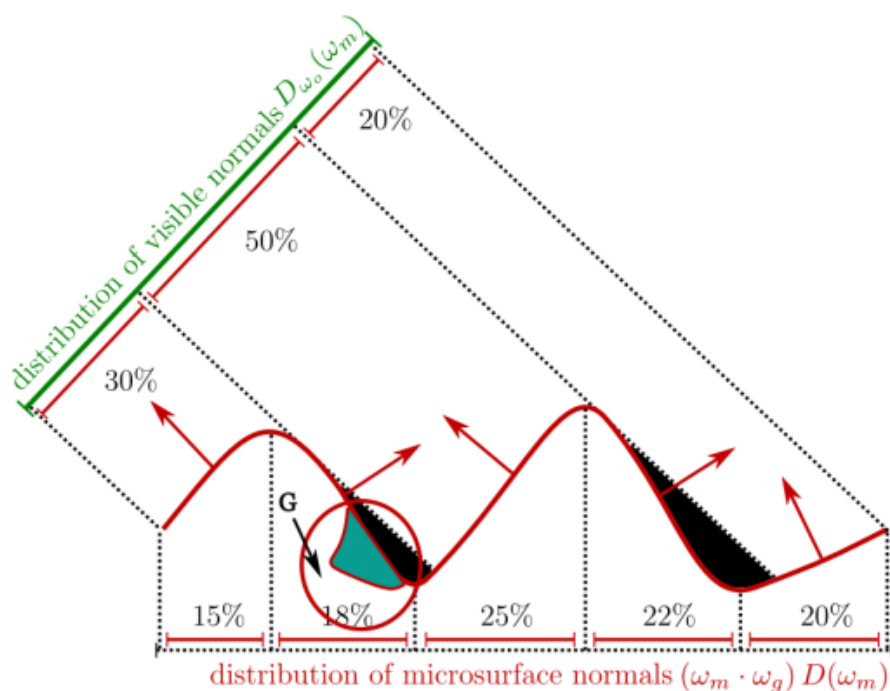
比微小更微小的面

接下来我们再讨论图形学里面另一个著名的模型，就是微表面。微表面结构是在立体角的计算架构上发明的。上面说过微元的投影可以等同于平面的投影。这样我们可以想象视线落点扩大为一个平面，称为几何平面，而垂直视线的立体角微元也扩大为一个平面，称为投影平面。因为几何平面扩大了一个点的几何信息容量，我们可以在上面随意做一些起伏，这就是微表面。微表面要保证一点就是不论如何随即起伏，它在几何平面上的投影面积必须为1，就是归一化。投影平面有一个固定面积 $\cos(\theta_o) = \omega_o \cdot \omega_g$ ，是单位面积1投影在投影平面上的大小。微表面在投影平面的投影大小也应该是 $\cos(\theta_o)$ ，但如图中所示，微表面有一些区域会被自己遮挡住，这部分称为阴影区域，用 $G(\omega_o, \omega_i)$ 表示。

另一个元素D是微表面法线的分布函数，作为函数就是输入一个法线向量，输出一个 $[0, 1]$ 的数值，而且这个函数在所有方向上的积分应该为1。首先能想到的当然是高斯正态分布函数。实际上常用的Beckmann分布就是对高斯正态分布略微修改的结果。

这样我们可以得到一个等式。

$$\text{投影面积} = \int_{\Omega} G(\omega_o, \omega_i)(\omega_o \cdot \omega_n) D(\omega_n) d\omega_n = \cos(\theta_o)$$

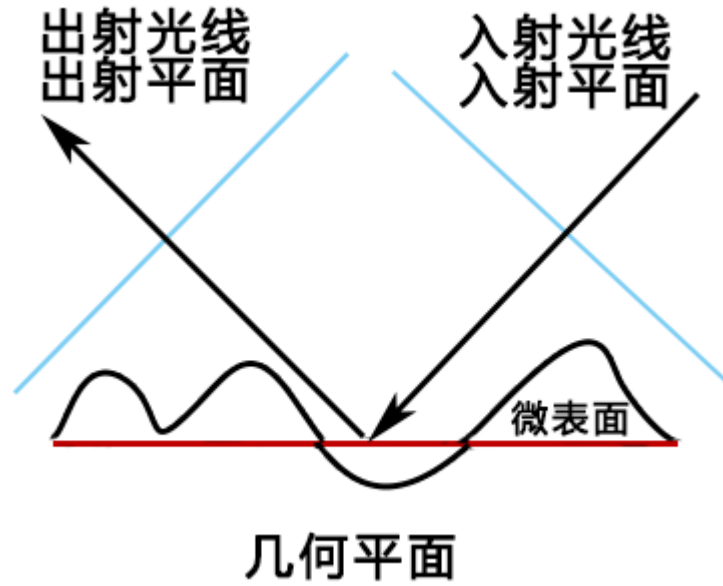


其中求阴影区域的大小是个麻烦的问题，因为G是在积分里的。实际上图形学在计算阴影区域的时候做了一个大胆的假设，认为阴影区域是一个全局几何性质和局部的法线没有关系，也就是说微表面的局部变化不会改变阴影区域的大小。这里完全是一个假设，没有经过数学证明。但可以画个图来说明这个假设是怎么来的。图中可以看到G区域不论如何弯曲都不会影响到阴影区域的大小。借着这个假设我们可以把G项从积分里拿出来，因为它和被积微元 ω_n 没有关系。我们反而可以借助上面的等式来计算G。

$$G(\omega_n) = \frac{\cos(\theta_o)}{\int_{\Omega} (\omega_o \cdot \omega_n) D(\omega_n) d\omega_n}$$

这样就能按照公式推导出不同的分布函数D所对应的阴影函数G。

重新画一个图来说明光照模型从光线入射到出射的路程上都有哪些元素。



图中可以看到，微表面，几何平面，入射平面，出射平面，各自对应的面积是几何面积，入射面积，出射面积，还有入射光通量，出射光通量，入射亮度，出射亮度。

现在我们有了一个很对称的模型，入射亮度 \times 入射面积=入射光通量，出射亮度 \times 出射面积=出射光通量。为了保持能量守恒入射光通量应该等于出射光通量。这里稍微不同的是入射光通量要乘以菲涅尔系数，这是一个光学性质，可以简单认为物体吸收了部分光子。关于菲涅尔系数以后有机会再详细表述，这里先把模型推导下去。 dF_i 是入射光通量， dF_o 是出射光通量， $F(\omega_o)$ 是菲涅耳系数，加上微分符号是因为都是在微元表面计算的。

$$dF_o = F(\omega_o)dF_i$$

入射亮度 L_i ，入射面积 D_i ，出射面积 D_o ，出射亮度 L_o 。要求的就是 L_o 。

$$L_o = L_i \frac{F(\omega_o)D_i}{D_o}$$

出射面积很好计算，就是几何面积在出射平面上的投影面积。

$$D_o = \cos(\theta_o)dA$$

在微表面结构下，我们可以认为只有部分微表面起到了发射作用。因为微表面朝向不定，有可能把光线发射到视线看不到的位置。只有微表面中正好能把光线从入射角度反射到视线角度的那些部分起到了作用。这一部分的法线正好在入射角度和视线角度的中间，称为半向量角：

$$\omega_h = \frac{\omega_i + \omega_o}{2}$$

$dA(\omega_h)$ 是微表面中法线方向为 ω_h 的面积，这个面积投影到入射平面上，就是入射面积。

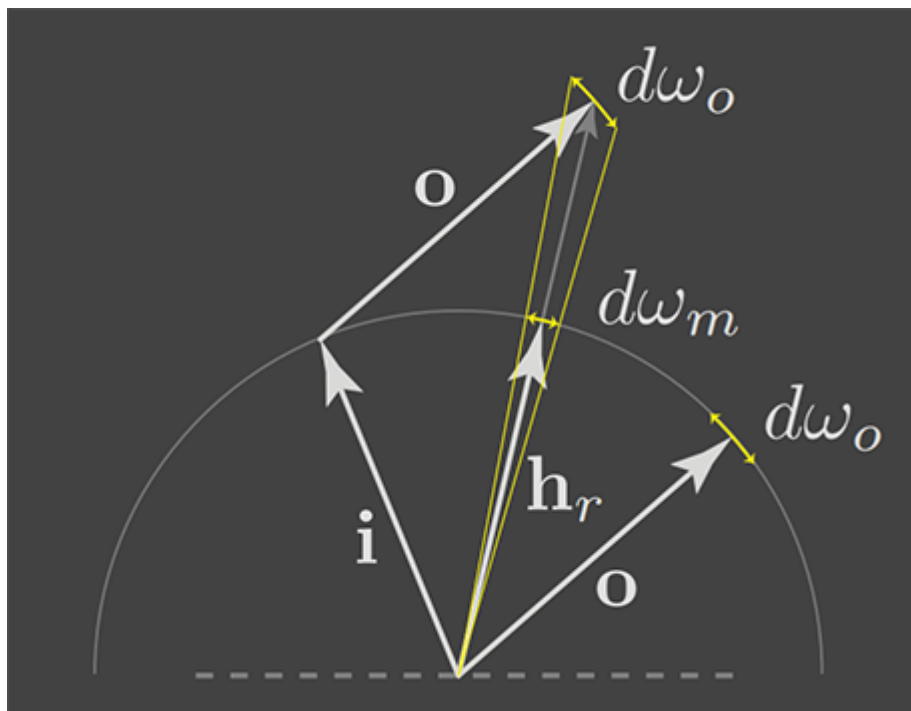
$$D_i = \cos(\theta_h) dA(\omega_h)$$

设几何平面的微元面积为 dA ，微表面中法线方向为 ω_h 的概率为 $D(\omega_h)$ ，在几何平面的投影面积就是 $D(\omega_h) dA$ ，反过来要求微表面中法线方向为 ω_h 的面积大小，就要把 $D(\omega_h) dA$ 投影回垂直于 ω_h 的平面，需要使用雅克比行列式计算投影。

$$dA(\omega_h) = \left\| \frac{\partial \omega_h}{\partial \omega_o} \right\| D(\omega_h) dA$$

$$\left\| \frac{\partial \omega_h}{\partial \omega_o} \right\| = \frac{\omega_i \cdot \omega_h}{|\omega_o + \omega_i|^2} = \frac{1}{4\cos(\theta_h)}$$

为什么要用雅克比行列式。如图所示， θ_o 变动的时候， θ_h 变动的幅度不同。因为微元的大小和这个变动有关，所以并不是感觉上两个面积应该相等。还有一个直观的理解，在 ω_o 上面画一小段圆弧，然后将 ω_o 平移到 ω_i 上，等于做 ω_o 和 ω_i 的向量加法，这时候这一小段圆弧在半球上的投影就是 ω_h 的微元面积。



雅克比行列式的推导详见附录。

公式整理之后得到

$$L_o = Li \frac{F(\omega_o) D_i}{D_o} = Li \frac{F(\omega_o) D_i}{\cos(\theta_o) dA} = \left\| \frac{\partial \omega_h}{\partial \omega_o} \right\| Li \frac{F(\omega_o) D(\omega_h) G(\omega_o, \omega_i) \cos(\theta_h) dA}{\cos(\theta_o) dA}$$

$$L_o = Li \frac{F(\omega_o) D(\omega_h) G(\omega_o, \omega_i)}{4 \cos(\theta_o)}$$

大功告成，希望这篇短文能帮助图形学的初学者建立对光照，渲染模型的直观认识。

附录：雅可比行列式推导

设 $\mu = \cos(\theta_h)$

出射向量可以写成

$$\omega_o = (\sqrt{1 - \mu^2}, 0, \mu)$$

半向量作为笛卡尔坐标函数（也可以看作坐标系），写成

$$\omega_h(x_a, y_a) = (x_a, y_a, \sqrt{1 - x_a^2 - y_a^2})$$

出射向量和半向量的关系为

$$\omega_o = 2(\omega_h \cdot \omega_i) \omega_h - \omega_i$$

出射向量也写为笛卡尔坐标函数，写成

$$x = 2(\sqrt{1 - \mu^2} x_a + \mu \sqrt{1 - x_a^2 - y_a^2}) x_a - \sqrt{1 - \mu^2}$$

$$y = 2(\sqrt{1 - \mu^2} x_a + \mu \sqrt{1 - x_a^2 - y_a^2}) y_a$$

这样可以求行列式，结果为

$$J = 4\mu^2$$

因为笛卡尔坐标和球坐标系还差了一个 μ

最后等式写成

$$\mu d\omega_o = 4\mu^2 d\omega_h$$

所以

$$\left\|\frac{\partial \omega_h}{\partial \omega_o}\right\| = \frac{1}{4\cos(\theta_h)}$$