Theoretical Foundations

We are now ready to begin discussing the radiometric theory that underpins ray tracing. It is important to note that this section is going to be making use of differential and integral calculus, including multi-dimensional integrals. The main reason behind this is because calculus is the only way in which we can fully define and discuss radiometry and the rendering equation.

Before we begin our discussion, it is important to keep in mind the objective that we are trying to attain. Up until this point, we have been performing *flat shading*. We have been shooting rays into the scene and checking if they intersect the objects. If they do, then the colour of the corresponding pixel is determined by the colour of the nearest object. This results in only being able to produce solid colours in our final renders. What we want to be able to do is to accurately render the way light interacts with surfaces in order to produce a better final render. What we need is therefore a way of modelling the way light bounces off a surface to be able to compute the final colour.

Radiometric Quantities

Radiometry deals with the measurement of radiation throughout the electromagnetic spectrum. In simple terms, it means the measurement of light on any portion of the electromagnetic spectrum. Specifically, we are concerned with visible light. We will now discuss some radiometric quantities, which will build up to the one we are ultimately interested in: radiance.

- *Radiant Energy* Q: The basic unit of electromagnetic energy, measured in joules. Specifically, it indicates the amount of energy carried by each photon.
- *Radiant Flux* Φ: The amount of radiant energy that passes through a given surface per second, measured in joules per second or *watts*. Radiant flux can also refer to the amount of energy that a light source emits per second and is also called *radiant power*.
- *Radiant flux density:* The radiant flux per unit surface area and is measured in watts per square metre. As the definition isn't restricted to real surfaces, it can be applied to any imaginary surface in space where it specifies the total amount of radiant energy that passes through a unit area per second.
- *Irradiance* E: Irradiance is the flux density that *arrives* at a surface. It's denoted by:

$$E = \frac{d\Phi}{dA} \tag{1}$$

- *Radiant Exitance* M: The flux density that *leaves* the surface. It is also called the *radiosity*.
- *Radiant Intensity* I: Its the flux density per solid angle. In simple terms, this is the amount of energy that is coming from a given direction, and it only makes sense for point light sources.

- *Radiance L*: Its the flux per unit projected area per unit solid angle. Radiance measures the flux at an arbitrary point in space, coming from a specific direction and measured per unit area on an imaginary surface that is perpendicular to the direction. Radiance is defined as:

$$L = \frac{d^2\Phi}{dA^{\perp} d\omega} \tag{2}$$

Since rays are infinitely thin lines, radiance is the natural quantity to compute for them. Radiance also several properties that are useful for ray tracing:

- Radiance is constant along rays that travel through empty space and is also the same in both directions along the ray.
- As radiance can be defined at any point in space, not just on surfaces, it can be defined at the eye point of a pinhole camera or a point on a pixel.
- If the point is on a surface, the radiance doesn't depend on whether the flux is arriving at or leaving the surface. In fact, it doesn't matter if the flux is reflected, transmitted, emitted, or scattered.

Radiance that hits a surface is called *incident radiance*, while radiance that leaves a surface is called *exitant radiance*.

Angular Dependence of Irradiance

When a point p is on a real surface, it's more convenient to represent radiance using an area element on the surface instead of $dA^{\perp}$. Consider irradiance E that hits a surface with normal incidence. If the irradiance is in a beam with cross-section $dA^{\perp}$, then the surface that receives it will also have area $dA^{\perp}$. If we tilt the beam so that it hits the surface with an incidence angle θ measured from the normal, then the beam will hit a larger area $dA^{\perp}/\cos\theta$. Irradiance is therefore proportional to the cosine of the incidence angle. This property is known as *Lambert's law*. As the beam becomes parallel to the surface, then the irradiance goes to zero.

We therefore have the following relation between projected area nd surface area:

$$dA^{\perp} = \cos\theta dA \tag{3}$$

We can therefore substitute this into equation eq. 1 which yields:

$$L = \frac{d^2\Phi}{dA \cos\theta d\omega} \tag{4}$$

Here $\cos\theta d\omega$ is called the *projected solid angle* because it's the projection of the differential solid angle $d\omega$ onto the (x, z) plane.

Notation and Directions

What we're ultimately interested in is specifying how light is reflected at a surface point p, which requires the directions of incoming and reflected light. We are going to use ω_i for the incoming light and ω_o for the reflected direction. By convention, both vectors are unit, are pointing away from p and are on the same side of the surface as the normal. We will also specify the angle from the normal that each vector produces by θ_i and θ_o. For shading purposes, ω_o will be the direction from which the ray hits the surface at p. Therefore, the incoming ray direction will be $-\omega_o$.

We want to be able to compute the reflected radiance along ω_o as a function of the total incoming radiance at p from all directions in the hemisphere above p. The thing is that there are actually two hemispheres, one on each side of the surface, though this isn't much of a concern as all we have to do is reflect the normal and the two vectors across the surface and everything will work the same.

Radiance and Irradiance

We are now ready to establish the relationship between incident radiance and irradiance. It follows from eq. 1 and eq. 4 that

$$dE_i(p, \omega_i) = L_i(p, \omega_i) \cos\theta_i d\omega_i \tag{5}$$

where $L_i(p, \omega_i)$ is the incident radiance at p from direction ω_i and $dE_i(p, \omega_i)$ is the irradiance in a cone with differential solid angle $d\omega_i$ centred on ω_i. The irradiance at p from a finite solid angle Ω_i is obtained by integrating dE over Ω_i:

$$E_i(p) = \int_{\Omega_i} L_i(p, \omega_i) \cos\theta_i d\omega_i \tag{6}$$

It is worth noting that there is an alternative representation of this formula that takes into account the particular wavelength called the *spectral representation*, but we will not be using this in our ray tracers.

BRDFs

Definition

It's a common practice in ray tracing to represent the reflective properties of surfaces with materials, but in order to do this, we require a way of precisely describing how the light is reflected at the surfaces. This is called the *bidirectional reflectance distribution function* (or *BRDF*). So, consider a differential amount of irradiance $dE_i(p, \omega_i)$ at a point p that's arriving in an element of solid angle $d\omega_i$ centred on the direction ω_i. The BRDF then specifies the contribution of this irradiance to the reflected radiance in the direction ω_o. This is a differential amount of radiance $dL_o(p, \omega_o)$.

In ray tracing, a ray will hit p from the direction ω_o, and what we want to compute is the reflected radiance along this ray. Because ω_o points away from p, the incoming direction is opposite of ω_o.

It so happens that due to the fact that the optical properties of materials are linear, the irradiance and radiance elements are proportional to each other, and the BRDF $f_r(p, \omega_i, \omega_o)$ is simply the constant of proportionality. We therefore have:

$$dL_o(p, \omega_o) = f_r(p, \omega_i, \omega_o) dE_i(p, \omega_i)$$

We can therefore express $dL_o(p, \omega_o)$ in terms of the incoming radiance by using eq. 5:

$$dL_o(p\omega_o) = f_r(p, \omega_i, \omega_o) L_i(p, \omega_i) \cos\theta d\omega_i \tag{7}$$

Solving eq. 7 for $f_r(p, \omega_i, \omega_o)$ yields:

$$f_r(p, \omega_i, \omega_o) = \frac{dL_o(p, \omega_o)}{L_i(p, \omega_i) \cos\theta_i d\omega_i} \tag{8}$$

Note that the BRDF is only a function of p and the two directions and can range from zero to infinity. An example of an infinite BRDF occurs with mirror reflection.

BRDFs that vary over the surface of objects are known as *spatially variant*, whereas those that do not are called *spatially invariant*.

Reflected Radiance

The reflected radiance in the ω_o direction that results from irradiance in a finite solid angle Ω_i is obtained by integrating over Ω_i:

$$L_o(p, \omega_o) = \int_{\Omega_i} f_r(p, \omega_i, \omega_o) L_i(p, \omega_i) \cos\theta_i d\omega_i \tag{9}$$

Since we're only interested in the hemisphere above p, then we get the following expression for the total reflected radiance in the ω_o direction:

$$L_o(p, \omega_o) = \int_{2\pi^+} f_r(p, \omega_i, \omega_o) L_i(p, \omega_i) \cos\theta_i d\omega_i \tag{10}$$

This is known as the *reflection equation.* It is extremely important in ray tracing because the value of the integral on the right-hand side needs to be computed by ray tracers at each hit point, at least for purely reflective materials that are not light sources.

Properties

BRDFs have the following properties:

- *Reciprocity:* Swapping the values of ω_o and ω_i has no effect on the value of the BRDF. If we therefore change the direction in which the light travels, the reflected radiance stays the same. This is important for *bidirectional ray tracing* where some rays are traced from the light sources.
- *Linearity:* Materials often require multiple BRDFs to fully model their reflective properties. The total reflected radiance is therefore the sum of the reflected radiance from each BRDF. A common example involves diffuse and specular reflection, which will be modelled by different BRDFs.
- *Conservation of energy:* This is specified in terms of reflectance, which we will discuss next.

Reflectance

Reflectance is defined as the ratio of reflected flux to incident flux (or equivalently reflected power to incident power). The radiant flux that's incident on a differential surface element dA through the solid angle Ω_i is:

$$d\Phi_i = dA \int_{\Omega_i} L_i(p, \omega_i) \cos\theta_i d\omega_i \tag{11}$$

The reflected flux from the same surface element, in the solid angle Ω_o is

$$d\Phi_o = dA \int_{\Omega_o} L_o(p, \omega_o) \cos\theta_o d\omega_o \tag{12}$$

We can then substitute eq. 9 into eq. 12 to obtain

$$d\Phi_o = dA \int_{\Omega_o} \int_{\Omega_i} f_r(p, \omega_i, \omega_o) L_i(p, \omega_i) \cos\theta_i \cos\theta_o d\omega_i d\omega_o \tag{13}$$

The reflectance $\rho(p, \Omega_i, \Omega_o)$ is defined by

$$\rho(p, \Omega_i, \Omega_o) = \frac{d\Phi_o}{d\Phi_i} \tag{14}$$

In this expression, dAs have canceled, which leads to the conclusion that the reflectance is dimensionless. Notice also that this expression makes no assumptions about the solid angles, the BRDF, or the angular distribution of incidence radiance in the hemisphere above p.

- *Conservation of energy:* No real materials reflect all of the light that hits them; some is absorbed and then re-radiated, often as heat. Therefore ρ satisfies $\rho(p, 2\pi, 2\pi) < 1$. This states that the surface element dA reflects less light in all directions than it receives from all directions.

The Perfect Diffuse BRDF

Now that we have covered all of this theory, let us look at an example to clarify some of the topics we've been discussing. We will discuss one of the simplest of BRDFs that represents perfect diffuse reflection, where incident light is scattered equally in all directions. Perfect diffuse reflection is also called *Lambertian reflection.* Note that although no real materials behave in this way, it is a good approximation for dull, matte materials, such as paper and completely flat paint.

For Lambertian surfaces, the reflected radiance $L_o(p, \omega_o) = L_{r,d}(p)$ is independent of ω_o. From eq. 9, this is only possible when the BRDF is independent of ω_i and ω_o, so we will denote this as $f_{r,d}(p)$ and take it out of the integral. This yields, with eq. 6:

$$L_{r,d}(p) = f_{r,d}(p) = \int_{\Omega_i} L_i(p, \omega_i) \cos\theta_i d\omega_i = f_{r,d}(p) E_i(p) \tag{15}$$

From eq. 5, we have

$$f_{r,d}(p) = \frac{L_{r,d}(p)}{E_i(p)} \tag{16}$$

We want to now express $f_{r,d}$ in terms of *perfect diffuse reflectance* ρ_d, defined as the fraction of the total incident flux that's reflected into the full hemisphere above the surface element dA when the BRDF is independent of ω_i and ω_o. In this case, eq. 12 becomes

$$d\Phi_o = dAL_{r,d}(p) \int_{2\pi^+} \cos\theta_o d\omega_o = dAL_{r,d}(p)\pi$$

From eq. 11:

$$d\Phi_i = dA \int_{2\pi^+} L_i(p, \omega_i) \cos\theta_i d\omega_i = dAE_i(p)$$

where now $E_i(p)$ is the total irradiance from the hemisphere above dA. Dividing these expressions and using eq. 16, the reflectance eq. 14 becomes

$$\rho_d(p) = \frac{d\Phi_o}{d\Phi_i} = \frac{L_r, d(p)\pi}{E_i(p)} = f_{r,d}(p)\pi$$

It then follows that:

$$f_{r,d}(p) = \frac{\rho_d}{\pi} \tag{17}$$

To model ambient illumination for perfectly diffuse surfaces, we need to use the *bihemispherical reflectance*, denoted by ρ_{hh}. In simple terms, it means that the incoming radiance is the same from all directions and doesn't vary with position. Therefore $L_i(p, \omega_i) = L_i$.

With this in mind, we can take $L_i(p)$ out of both integrals in eq. 14 and cancel it, as well as taking $f_{r,d}(p)$ out of the top integral. This yields:

$$\rho_{hh} = \frac{f_{r,d}(p)}{\pi} = \int_{2\pi^+} \int_{2\pi^+} \cos\theta_i \cos\theta_o d\omega_i d\omega_o = \pi f_{r,d}(p) = \rho(p) \tag{18}$$

Now that we have this, we can explain the functionality in practical terms for our Lambertian BRDF. In essence, we need to compute:

- ρ_d,
- $f_{r,d}(p)$, and
- ρ_{hh}

Since we are expressing everything in RGB colours, then ρ_d becomes $k_d c_d$, where k_d is called the *diffuse reflection coefficient* and c_d is the diffuse colour. With this in hand we can now derive the remaining components of our BRDF:

- $f_{r,d} = \frac{k_d c_d}{\pi}$
- $\rho hh = k_d c_d$

The Rendering Equation

The *rendering equation* expresses the steady-state radiative energy balance in a scene. The reflection equation is part of the rendering equation, as it expresses the reflective energy balance between surfaces. All we have to do is add the light sources to this equation to get the rendering equation for reflective surfaces.

The surface of a light source emits light and is therefore known as an *emissive* surface. All that we need to do to complete the rendering equation is to take eq. 10 and add a term for the *emitted radiance* in the direction ω_o to obtain:

$$L_o(p, \omega_o) = L_e(p, \omega_o) + \int_{2\pi^+} f_r(p, \omega_i, \omega_o) L_i(p, \omega_i) \cos\theta_i d\omega_i \tag{19}$$

This is the *hemisphere form* of the rendering equation. Ultimately, all that rendering is trying to do is to solve (or approximate) the value of this integral. Now there is another formulation called the *area form*, but we will not discuss it in class. The question now is, how do we go about solving this equation? Well it turns out that the incident radiance $L_i(p, \omega_i)$ can be computed by tracing a ray from the point p. Therefore, we can substitute $L_i(p, \omega_i)$ with $L_o(r_c(p, \omega_i), -\omega_i)$ where r_c is the ray-casting operator which is the nearest hit point along the ray in the direction ω_i.

This last substitution resolves in a recursive integral, as L_o appears on both sides of the equation. Now while there is a way of approximating the solution to this integral, we will not discuss it in this course. What we will focus instead are on the consequences of the solution. In particular:

- Each time we approximate the equation, we are in fact bouncing a ray on each surface. While the true solution requires infinite bounces, the discrete nature of the computer and the representation of data limits this to a finite number, which we will call the *depth*.
- For shading with point and directional lights and simulating perfect specular relfection and transmission, the integrals reduce to small sums of simple expressions that we can calculate exactly.
- For shasing with area lights and simulating diffuse-diffuse light transport, we can use Monte Carlo techniques to numerically evaluate the multi-dimensional integrals.

The second point is of particular importance, since these simple sums will be the basis for the shading we will do for real-time graphics. The third point yields entirely new rendering techniques for physically-based rendering, which include photon mapping, Metropolis light transport, path tracing, among others.