Fundamentals of Computer Graphics

FOURTH EDITION

Steve Marschner
Peter Shirley
with
Michael Ashikhmin
Michael Gleicher
Naty Hoffman
Garrett Johnson
Tamara Munzner
Erik Reinhard
William B. Thompson
Peter Willemsen
Brian Wyvill

CRC Press
Taylor & Francis Group
AN A K PETERS BOOK

Fundamentals of Computer Graphics

FOURTH EDITION

2.6 Linear Interpolation

Perhaps the most common mathematical operation in graphics is *linear interpolation*. We have already seen an example of linear interpolation of position to form line segments in 2D and 3D, where two points $\mathbf{a}$ and $\mathbf{b}$ are associated with a parameter t to form the line $\mathbf{p} = (1-t)\mathbf{a} + t\mathbf{b}$. This is *interpolation* because $\mathbf{p}$ goes through $\mathbf{a}$ and $\mathbf{b}$ exactly at $t = 0$ and $t = 1$. It is *linear* interpolation because the weighting terms t and $1 - t$ are linear polynomials of t.

Another common linear interpolation is among a set of positions on the x-axis: $x_0, x_1, \ldots, x_n$, and for each x_i we have an associated height, y_i. We want to create a continuous function $y = f(x)$ that interpolates these positions, so that f goes through every data point, i.e., $f(x_i) = y_i$. For linear interpolation, the points (x_i, y_i) are connected by straight line segments. It is natural to use parametric line equations for these segments. The parameter t is just the fractional distance between x_i and x_{i+1}:

$$f(x) = y_i + \frac{x - x_i}{x_{i+1} - x_i}(y_{i+1} - y_i). \tag{2.26}$$

Because the weighting functions are linear polynomials of x, this is linear interpolation.

The two examples above have the common form of linear interpolation. We create a variable t that varies from 0 to 1 as we move from data item A to data item B. Intermediate values are just the function $(1-t)A + tB$. Notice that Equation (2.26) has this form with

$$t = \frac{x - x_i}{x_{i+1} - x_i}.$$

2.7 Triangles

Triangles in both 2D and 3D are the fundamental modeling primitive in many graphics programs. Often information such as color is tagged onto triangle vertices, and this information is interpolated across the triangle. The coordinate system that makes such interpolation straightforward is called *barycentric coordinates*; we will develop these from scratch. We will also discuss 2D triangles, which must be understood before we can draw their pictures on 2D screens.

2.7.1 2D Triangles

If we have a 2D triangle defined by 2D points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$, we can first find its area:

$$\begin{aligned} \text{area} &= \frac{1}{2}\begin{vmatrix} x_b - x_a & x_c - x_a \\ y_b - y_a & y_c - y_a \end{vmatrix} \\ &= \frac{1}{2}\left(x_a y_b + x_b y_c + x_c y_a - x_a y_c - x_b y_a - x_c y_b\right). \end{aligned} \tag{2.27}$$

The derivation of this formula can be found in Section 5.3. This area will have a positive sign if the points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are in counterclockwise order and a negative sign, otherwise.

Often in graphics, we wish to assign a property, such as color, at each triangle vertex and smoothly interpolate the value of that property across the triangle. There are a variety of ways to do this, but the simplest is to use *barycentric* coordinates. One way to think of barycentric coordinates is as a nonorthogonal coordinate system as was discussed briefly in Section 2.4.2. Such a coordinate system is shown in Figure 2.36, where the coordinate origin is $\mathbf{a}$ and the vectors from $\mathbf{a}$ to $\mathbf{b}$ and $\mathbf{c}$ are the basis vectors. With that origin and those basis vectors,

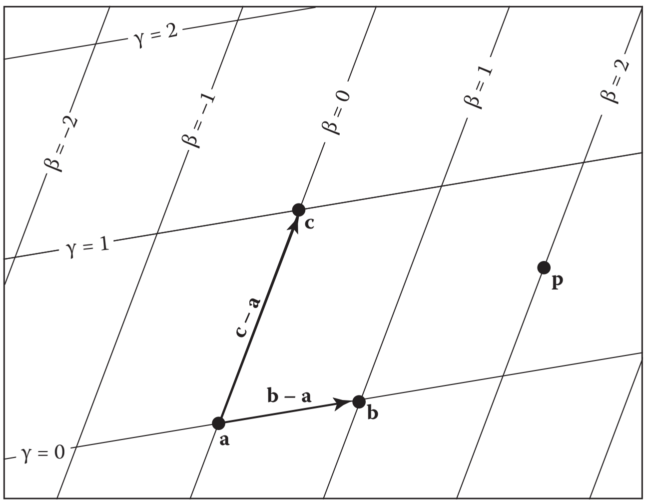

Figure 2.36. A 2D triangle with vertices **a**, **b**, **c** can be used to set up a nonorthogonal coordinate system with origin **a** and basis vectors (**b** – **a**) and (**c** – **a**). A point is then represented by an ordered pair (β, γ). For example, the point **p** = (2.0, 0.5), i.e., **p** = **a** + 2.0 (**b** – **a**) + 0.5 (**c** – **a**).

any point $\mathbf{p}$ can be written as

$$\mathbf{p} = \mathbf{a} + \beta(\mathbf{b} - \mathbf{a}) + \gamma(\mathbf{c} - \mathbf{a}). \tag{2.28}$$

Note that we can reorder the terms in Equation (2.28) to get

$$\mathbf{p} = (1 - \beta - \gamma)\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}.$$

Often people define a new variable α to improve the symmetry of the equations:

$$\alpha \equiv 1 - \beta - \gamma,$$

which yields the equation

$$\mathbf{p}(\alpha, \beta, \gamma) = \alpha\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}, \tag{2.29}$$

with the constraint that

$$\alpha + \beta + \gamma = 1. \tag{2.30}$$

Barycentric coordinates seem like an abstract and unintuitive construct at first, but they turn out to be powerful and convenient. You may find it useful to think of how street addresses would work in a city where there are two sets of parallel streets, but where those sets are not at right angles. The natural system would essentially be barycentric coordinates, and you would quickly get used to them. Barycentric coordinates are defined for all points on the plane. A particularly nice feature of barycentric coordinates is that a point $\mathbf{p}$ is inside the triangle formed by $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ if and only if

$$\begin{aligned} 0 < \alpha < 1, \\ 0 < \beta < 1, \\ 0 < \gamma < 1. \end{aligned}$$

If one of the coordinates is zero and the other two are between zero and one, then you are on an edge. If two of the coordinates are zero, then the other is one, and you are at a vertex. Another nice property of barycentric coordinates is that Equation (2.29) in effect mixes the coordinates of the three vertices in a smooth way. The same mixing coefficients (α, β, γ) can be used to mix other properties, such as color, as we will see in the next chapter.

Given a point $\mathbf{p}$, how do we compute its barycentric coordinates? One way is to write Equation (2.28) as a linear system with unknowns β and γ, solve, and set $\alpha = 1 - \beta - \gamma$. That linear system is

$$\begin{bmatrix} x_b - x_a & x_c - x_a \\ y_b - y_a & y_c - y_a \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x_p - x_a \\ y_p - y_a \end{bmatrix}. \tag{2.31}$$

Although it is straightforward to solve Equation (2.31) algebraically, it is often fruitful to compute a direct geometric solution.

One geometric property of barycentric coordinates is that they are the signed scaled distance from the lines through the triangle sides, as is shown for β in Figure 2.37. Recall from Section 2.5.2 that evaluating the equation $f(x, y)$ for the line $f(x, y) = 0$ returns the scaled signed distance from (x, y) to the line. Also recall that if $f(x, y) = 0$ is the equation for a particular line, so is $kf(x, y) = 0$ for any nonzero k. Changing k scales the distance and controls which side of the line has positive signed distance, and which negative. We would like to choose k such that, for example, $kf(x, y) = \beta$. Since k is only one unknown, we can force this with one constraint, namely that at point $\mathbf{b}$ we know $\beta = 1$. So if the line $f_{ac}(x, y) = 0$ goes through both $\mathbf{a}$ and $\mathbf{c}$, then we can compute β for a point (x, y) as follows:

$$\beta = \frac{f_{ac}(x, y)}{f_{ac}(x_b, y_b)}, \tag{2.32}$$

and we can compute γ and α in a similar fashion. For efficiency, it is usually wise to compute only two of the barycentric coordinates directly and to compute the third using Equation (2.30).

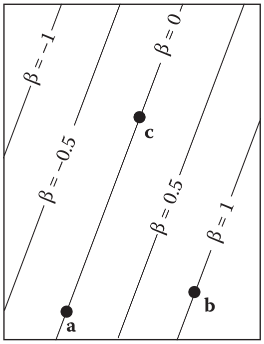

Figure 2.37. The barycentric coordinate β is the signed scaled distance from the line through $\mathbf{a}$ and $\mathbf{c}$.

To find this "ideal" form for the line through $\mathbf{p}_0$ and $\mathbf{p}_1$, we can first use the technique of Section 2.5.2 to find *some* valid implicit lines through the vertices. Equation (2.17) gives us

$$f_{ab}(x, y) \equiv (y_a - y_b)x + (x_b - x_a)y + x_a y_b - x_b y_a = 0.$$

Note that $f_{ab}(x_c, y_c)$ probably does not equal one, so it is probably not the ideal form we seek. By dividing through by $f_{ab}(x_c, y_c)$ we get

$$\gamma = \frac{(y_a - y_b)x + (x_b - x_a)y + x_a y_b - x_b y_a}{(y_a - y_b)x_c + (x_b - x_a)y_c + x_a y_b - x_b y_a}.$$

The presence of the division might worry us because it introduces the possibility of divide-by-zero, but this cannot occur for triangles with areas that are not near zero. There are analogous formulas for α and β, but typically only one is needed:

$$\beta = \frac{(y_a - y_c)x + (x_c - x_a)y + x_a y_c - x_c y_a}{(y_a - y_c)x_b + (x_c - x_a)y_b + x_a y_c - x_c y_a},$$
$$\alpha = 1 - \beta - \gamma.$$

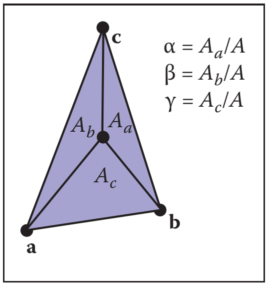

Figure 2.38. The barycentric coordinates are proportional to the areas of the three subtriangles shown.

Another way to compute barycentric coordinates is to compute the areas A_a, A_b, and A_c, of subtriangles as shown in Figure 2.38. Barycentric coordinates obey the rule

$$\alpha = A_a/A,$$
$$\beta = A_b/A, \quad (2.33)$$
$$\gamma = A_c/A,$$

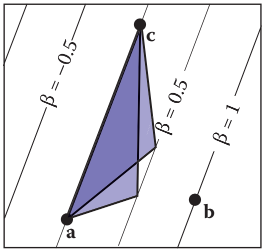

Figure 2.39. The area of the two triangles shown is base times height and are thus the same, as is any triangle with a vertex on the $\beta = 0.5$ line. The height and thus the area is proportional to β.

where A is the area of the triangle. Note that $A = A_a + A_b + A_c$, so it can be computed with two additions rather than a full area formula. This rule still holds for points outside the triangle if the areas are allowed to be signed. The reason for this is shown in Figure 2.39. Note that these are signed areas and will be computed correctly as long as the same signed area computation is used for both A and the subtriangles A_a, A_b, and A_c.

2.7.2 3D Triangles

One wonderful thing about barycentric coordinates is that they extend almost transparently to 3D. If we assume the points $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are 3D, then we can still use the representation

$$\mathbf{p} = (1 - \beta - \gamma)\mathbf{a} + \beta\mathbf{b} + \gamma\mathbf{c}.$$

Now, as we vary β and γ, we sweep out a plane.

The normal vector to a triangle can be found by taking the cross product of any two vectors in the plane of the triangle (Figure 2.40). It is easiest to use two of the three edges as these vectors, for example,

$$\mathbf{n} = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a}). \quad (2.34)$$

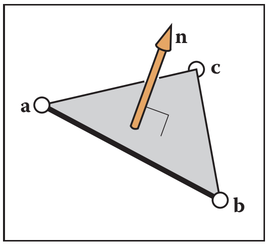

Figure 2.40. The normal vector of the triangle is perpendicular to all vectors in the plane of the triangle, and thus perpendicular to the edges of the triangle.

Note that this normal vector is not necessarily of unit length, and it obeys the right-hand rule of cross products.

The area of the triangle can be found by taking the length of the cross product:

$$\text{area} = \frac{1}{2}\|(\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})\|. \quad (2.35)$$

Note that this is *not* a signed area, so it cannot be used directly to evaluate barycentric coordinates. However, we can observe that a triangle with a "clockwise" vertex order will have a normal vector that points in the opposite direction to the normal of a triangle in the same plane with a "counterclockwise" vertex order. Recall that

$$\mathbf{a} \cdot \mathbf{b} = \|\mathbf{a}\| \ \|\mathbf{b}\| \ \cos\phi,$$

where ϕ is the angle between the vectors. If $\mathbf{a}$ and $\mathbf{b}$ are parallel, then $\cos\phi = \pm 1$, and this gives a test of whether the vectors point in the same or opposite directions. This, along with Equations (2.33), (2.34), and (2.35), suggest the formulas

$$\alpha = \frac{\mathbf{n} \cdot \mathbf{n}_a}{\|\mathbf{n}\|^2},$$

$$\beta = \frac{\mathbf{n} \cdot \mathbf{n}_b}{\|\mathbf{n}\|^2},$$

$$\gamma = \frac{\mathbf{n} \cdot \mathbf{n}_c}{\|\mathbf{n}\|^2},$$

where $\mathbf{n}$ is Equation (2.34) evaluated with vertices $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$; $\mathbf{n}_a$ is Equation (2.34) evaluated with vertices $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{p}$, and so on, i.e.,

$$\begin{aligned} \mathbf{n}_a &= (\mathbf{c} - \mathbf{b}) \times (\mathbf{p} - \mathbf{b}), \\ \mathbf{n}_b &= (\mathbf{a} - \mathbf{c}) \times (\mathbf{p} - \mathbf{c}), \\ \mathbf{n}_c &= (\mathbf{b} - \mathbf{a}) \times (\mathbf{p} - \mathbf{a}). \end{aligned} \tag{2.36}$$

Frequently Asked Questions

• Why isn't there vector division?

It turns out that there is no "nice" analogy of division for vectors. However, it is possible to motivate the quaternions by examining this question in detail (see Hoffmann's book referenced in the chapter notes).

• Is there something as clean as barycentric coordinates for polygons with more than three sides?

Unfortunately there is not. Even convex quadrilaterals are much more complicated. This is one reason triangles are such a common geometric primitive in graphics.

• Is there an implicit form for 3D lines?

No. However, the intersection of two 3D planes defines a 3D line, so a 3D line can be described by two simultaneous implicit 3D equations.