Curves and Surfaces for CAGD

A Practical Guide

Fifth Edition

Gerald Farin
Arizona State University

MORGAN KAUFMANN PUBLISHERS
AN IMPRINT OF ACADEMIC PRESS
A Harcourt Science and Technology Company
SAN FRANCISCO SAN DIEGO NEW YORK BOSTON
LONDON SYDNEY TOKYO

Linear Interpolation

3

Most of the computations that we use in CAGD may be broken down into seemingly trivial steps—sequences of linear interpolations. It is therefore important to understand the properties of these basic building blocks. This chapter explores those properties and introduces a related concept, called blossoms.

3.1 Linear Interpolation

Let $\mathbf{a}, \mathbf{b}$ be two distinct points in $\mathbb{E}^3$. The set of all points $\mathbf{x} \in \mathbb{E}^3$ of the form

$$\mathbf{x} = \mathbf{x}(t) = (1-t)\mathbf{a} + t\mathbf{b}; \quad t \in \mathbb{R} \tag{3.1}$$

is called the *straight line* through $\mathbf{a}$ and $\mathbf{b}$. Any three (or more) points on a straight line are said to be *collinear*.

For $t = 0$, the straight line passes through $\mathbf{a}$ and for $t = 1$, it passes through $\mathbf{b}$. For $0 \le t \le 1$, the point $\mathbf{x}$ is between $\mathbf{a}$ and $\mathbf{b}$, whereas for all other values of t it is outside; see Figure 3.1.

Equation (3.1) represents $\mathbf{x}$ as a barycentric combination of two points in $\mathbb{E}^3$. The same barycentric combination holds for the three points $0, t, 1$ in $\mathbb{E}^1$: $t = (1-t) \cdot 0 + t \cdot 1$. So t is related to 0 and 1 by the same barycentric combination that relates $\mathbf{x}$ to $\mathbf{a}$ and $\mathbf{b}$. Hence, by the definition of affine maps, the three points $\mathbf{a}, \mathbf{x}, \mathbf{b}$ are an affine map of the three 1D points $0, t, 1$! *Thus linear interpolation is an affine map of the real line onto a straight line in* $\mathbb{E}^3$.[1]

1 Strictly speaking, we should therefore use the term *affine interpolation* instead of *linear interpolation*. We use linear interpolation because its use is so widespread.

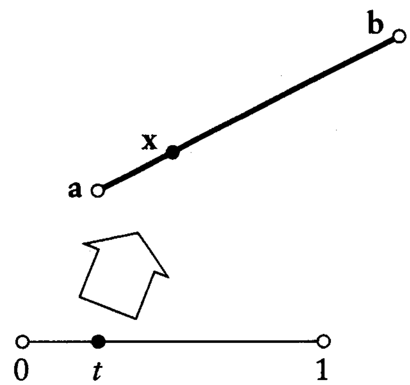

Figure 3.1 Linear interpolation: two points **a**, **b** define a straight line through them. The point t in the domain is mapped to the point **x** in the range.

It is now almost a tautology when we state: *linear interpolation is affinely invariant.* Written as a formula: if Φ is an affine map of $\mathbb{E}^3$ onto itself, and (3.1) holds, then also

$$\Phi\mathbf{x} = \Phi\big((1-t)\mathbf{a} + t\mathbf{b}\big) = (1-t)\Phi\mathbf{a} + t\Phi\mathbf{b}. \tag{3.2}$$

Since affine maps may be applied to vectors as well as to points, it makes sense to ask what linear interpolation will do to vector arguments. These vectors "live" in 1D domain space, and will be denoted by $\vec{v}$.

If c and d are two 1D points in the domain, they define a vector $\vec{v}$ by setting $\vec{v} = d - c$. The corresponding vector $\mathbf{l}(\vec{v})$ in the range is then defined as

$$\mathbf{l}(\vec{v}) = \mathbf{l}(d) - \mathbf{l}(c). \tag{3.3}$$

Figure 3.2 illustrates. For the special case of $\vec{v}$ being the 1D zero vector $\vec{v} = \vec{0}$, we have

$$\mathbf{l}(\vec{0}) = \mathbf{0}.^2 \tag{3.4}$$

Closely related to linear interpolation is the concept of *barycentric coordinates*, due to Moebius [429]. Let **a**, **x**, **b** be three collinear points in $\mathbb{E}^3$:

$$\mathbf{x} = \alpha\mathbf{a} + \beta\mathbf{b}; \quad \alpha + \beta = 1. \tag{3.5}$$

2 Here, **0** denotes the zero vector.

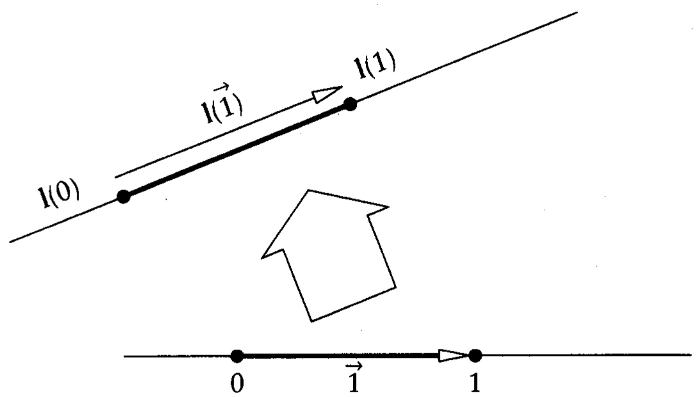

Figure 3.2 Linear interpolation: the vector $\vec{1}$ in the domain is mapped to the vector $\mathbf{l}(\vec{1})$ in the range.

Then α and β are called *barycentric coordinates* of $\mathbf{x}$ with respect to $\mathbf{a}$ and $\mathbf{b}$. Note that by our previous definitions, $\mathbf{x}$ is a barycentric combination of $\mathbf{a}$ and $\mathbf{b}$.

The connection between barycentric coordinates and linear interpolation is obvious: we have $\alpha = 1 - t$ and $\beta = t$. This shows, by the way, that barycentric coordinates do not always have to be positive: for $t \notin [0, 1]$, either α or β is negative. For any three collinear points $\mathbf{a}, \mathbf{b}, \mathbf{c}$, the barycentric coordinates of $\mathbf{b}$ with respect to $\mathbf{a}$ and $\mathbf{c}$ are given by

$$\alpha = \frac{\mathrm{vol}_1(\mathbf{b}, \mathbf{c})}{\mathrm{vol}_1(\mathbf{a}, \mathbf{c})},$$

$$\beta = \frac{\mathrm{vol}_1(\mathbf{a}, \mathbf{b})}{\mathrm{vol}_1(\mathbf{a}, \mathbf{c})},$$

where vol_1 denotes the one-dimensional volume, which is the signed distance between two points. Barycentric coordinates are not only defined on a straight line, but also on a plane. Section 3.5 has more details.

Another important concept in this context is that of *ratios*. The ratio of three collinear points $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is defined by

$$\mathrm{ratio}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \frac{\mathrm{vol}_1(\mathbf{a}, \mathbf{b})}{\mathrm{vol}_1(\mathbf{b}, \mathbf{c})}. \tag{3.6}$$

If α and β are barycentric coordinates of $\mathbf{b}$ with respect to $\mathbf{a}$ and $\mathbf{c}$, it follows that

$$\mathrm{ratio}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \frac{\beta}{\alpha}. \tag{3.7}$$

The barycentric coordinates of a point do not change under affine maps, and neither does their quotient. Thus the ratio of three collinear points is not affected by affine transformations. So if (3.7) holds, then also

$$\text{ratio}(\Phi\mathbf{a}, \Phi\mathbf{b}, \Phi\mathbf{c}) = \frac{\beta}{\alpha}, \tag{3.8}$$

where Φ is an affine map. This property may be used to *compute* ratios efficiently. Instead of using square roots to compute the distances between points **a**, **x**, and **b**, we would project them onto one of the coordinate axes and then use simple differences of their x- or y-coordinates.[3] This shortcut works since parallel projection is an affine map!

Equation (3.8) states that *affine maps are ratio preserving*. This property may be used to define affine maps. Every map that takes straight lines to straight lines and is ratio preserving is an affine map.

The concept of ratio preservation may be used to derive another useful property of linear interpolation. We have defined the straight line segment [**a**, **b**] to be the affine image of the *unit interval* [0, 1], but we can also view that straight line segment as the affine image of any interval $[a, b]$. The interval $[a, b]$ may itself be obtained by an affine map from the interval [0, 1] or vice versa. With $t \in [0, 1]$ and $u \in [a, b]$, that map is given by $t = (u - a)/(b - a)$. The interpolated point on the straight line is now given by both

$$\mathbf{x}(t) = (1 - t)\mathbf{a} + t\mathbf{b}$$

and

$$\mathbf{x}(u) = \frac{b - u}{b - a}\mathbf{a} + \frac{u - a}{b - a}\mathbf{b}. \tag{3.9}$$

Since a, u, b and $0, t, 1$ are in the same ratio as the triple **a**, **x**, **b**, we have shown that *linear interpolation is invariant under affine domain transformations*. By affine domain transformation, we simply mean an affine map of the real line onto itself. The parameter t is sometimes called a *local parameter* of the interval $[a, b]$.

A more general way to express this is by saying that any barycentric combination of three domain points r, s, t (not necessarily involving any interval endpoints) carries over to the corresponding range points:

$$s = (1 - \alpha)r + \alpha t \Rightarrow \mathbf{x}(s) = (1 - \alpha)\mathbf{x}(r) + \alpha\mathbf{x}(t). \tag{3.10}$$

3 But be sure to avoid projection onto the x-axis if the three points are parallel to the y-axis!

A concluding remark: we have demonstrated the interplay between the two concepts of linear interpolation and ratios. In this book, we will often describe methods by saying that points have to be collinear and must be in a given ratio. This is the geometric (descriptive) equivalent of the algebraic (algorithmic) statement that one of the three points may be obtained by linear interpolation from the other two.

3.2 Piecewise Linear Interpolation

Let $\mathbf{b}_0, \ldots, \mathbf{b}_n \in \mathbb{E}^3$ form a *polygon* **B**. This polygon consists of a sequence of straight line segments, each interpolating to a pair of points $\mathbf{b}_i, \mathbf{b}_{i+1}$. It is therefore also called the *piecewise linear interpolant* $\mathcal{PL}$ to the points $\mathbf{b}_i$. If the points $\mathbf{b}_i$ lie on a curve **c**, then **B** is said to be a piecewise linear interpolant to **c**, and we write

$$\mathbf{B} = \mathcal{PL}\,\mathbf{c}. \tag{3.11}$$

One of the important properties of piecewise linear interpolation is *affine invariance*. If the curve **c** is mapped onto a curve $\Phi\mathbf{c}$ by an affine map Φ, then the piecewise linear interpolant to $\Phi\mathbf{c}$ is the affine map of the original piecewise linear interpolant:

$$\mathcal{PL}\,\Phi\mathbf{c} = \Phi\;\mathcal{PL}\,\mathbf{c}. \tag{3.12}$$

Another property is the *variation diminishing property*. Consider a continuous curve **c**, a piecewise linear interpolant $\mathcal{PL}\,\mathbf{c}$, and an arbitrary plane. Let cross **c** be the number of crossings that the curve **c** has with this plane, and let cross($\mathcal{PL}\,\mathbf{c}$) be the number of crossings that the piecewise linear interpolant has with this plane. (Special cases may arise; see Section 3.8.) Then we always have

$$\text{cross}(\mathcal{PL}\,\mathbf{c}) \le \text{cross}\,\mathbf{c}. \tag{3.13}$$

This property follows from a simple observation: consider two points $\mathbf{b}_i, \mathbf{b}_{i+1}$. The straight line segment through them can cross a given plane at one point at most, whereas the curve segment from **c** that connects them may cross the same plane in many arbitrary points. The variation diminishing property is illustrated in Figure 3.3.

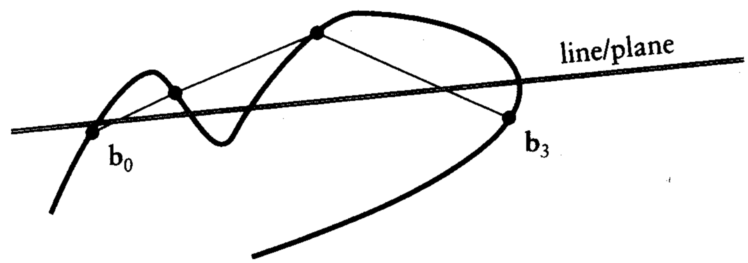

Figure 3.3 The variation diminishing property: a piecewise linear interpolant to a curve has no more intersections with any plane than the curve itself.

3.3 Menelaos' Theorem

We use the concept of piecewise linear interpolation to prove one of the most important geometric theorems for the theory of CAGD: Menelaos' theorem. This theorem can be used for the proof of many constructive algorithms, and its importance was already realized by de Casteljau [146] and W. Boehm [67].

Referring to Figure 3.4, we define

$$\begin{aligned} \mathbf{b}[0,t] &= (1-t)\mathbf{b}_0 + t\mathbf{b}_1, \\ \mathbf{b}[s,0] &= (1-s)\mathbf{b}_0 + s\mathbf{b}_1, \\ \mathbf{b}[1,t] &= (1-t)\mathbf{b}_1 + t\mathbf{b}_2, \\ \mathbf{b}[s,1] &= (1-s)\mathbf{b}_1 + s\mathbf{b}_2. \end{aligned}$$

Let us further define two points

$$\begin{aligned} \mathbf{b}[s,t] &= (1-t)\mathbf{b}[s,0] + t\mathbf{b}[s,1] \quad \text{and} \\ \mathbf{b}[t,s] &= (1-s)\mathbf{b}[0,t] + s\mathbf{b}[t,1]. \end{aligned} \tag{3.14}$$

Menelaos' theorem now states that these points are identical:

$$\mathbf{b}[s,t] = \mathbf{b}[t,s]. \tag{3.15}$$

For a proof, we simply verify that

$$\mathbf{b}[s,t] = \mathbf{b}[t,s] = (1-t)(1-s)\mathbf{b}_0 + [(1-t)s + t(1-s)]\mathbf{b}_1 + st\mathbf{b}_2. \tag{3.16}$$

Some interesting special cases are given by $\mathbf{b}[0,0] = \mathbf{b}_0$ or by $\mathbf{b}[0,1] = \mathbf{b}_1$.

Equation (3.15) is a "CAGD version" of the original Menelaos' theorem, which may be stated as (see Coxeter [130]):

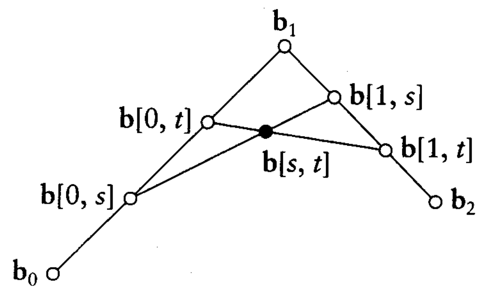

Figure 3.4 Menelaos' theorem: the point $\mathbf{b}[s, t]$ may be obtained from linear interpolation at t or at s.

$$\text{ratio}(\mathbf{b}[s, 1], \mathbf{b}[1, t], \mathbf{b}_1) \cdot \text{ratio}(\mathbf{b}_1, \mathbf{b}[0, t], \mathbf{b}[s, 0]) \cdot$$
$$\text{ratio}(\mathbf{b}[s, 0], \mathbf{b}[s, t], \mathbf{b}[s, 1]) = -1. \tag{3.17}$$

The proof of (3.17) is a direct consequence of (3.15). Note the ordering of points in the second ratio! Menelaos' theorem is closely related to Ceva's, which is given in Section 3.5.

3.4 **Blossoms**

The bivariate function $\mathbf{b}[t_1, t_2]$ from (3.16) will be very important for the remainder of this book. Functions of that type are called *blossoms*. Before we introduce the general concept, we will further explore properties of (3.16).

The first property is called *symmetry*. It states that the order of the blossom arguments does not matter—which is exactly Menelaos' theorem.

In Section 3.1, we saw that linear interpolation carries domain relationships over to corresponding range relationships; see (3.10). Since blossoms are evaluated using linear interpolations, we now have: if the first argument t_1 of a blossom is a barycentric combination of two (or more) 1D points r and s, we may compute the blossom values for each argument and then form their barycentric combination:

$$\mathbf{b}[\alpha r + \beta s, t_2] = \alpha \mathbf{b}[r, t_2] + \beta \mathbf{b}[s, t_2]; \quad \alpha + \beta = 1. \tag{3.18}$$

Equation (3.18) states that the blossom $\mathbf{b}$ is affine with respect to its first argument, but it is affine for the second one as well because of the symmetry property. This is the reason the blossom is called *multiaffine*—the second of its main properties.

For a third property, we study what happens if both blossom arguments are equal: $t_1 = t_2 = t$. Then the expression $\mathbf{b}[t, t]$ denotes a point that depends on one variable t—thus it traces out a polynomial curve.[4] This property is called the *diagonal property.*

Our special blossom $\mathbf{b}[t_1, t_2]$ has two arguments. Blossoms with an arbitrary number n of arguments are easily defined by the preceding three properties. A blossom is an n-variate function $\mathbf{b}[t_1, \ldots, t_n]$ from $\mathbb{R}^n$ into $\mathbb{E}^2$ or $\mathbb{E}^3$. It is defined by three properties:

Symmetry:

$$\mathbf{b}[t_1, \ldots, t_n] = \mathbf{b}[\pi(t_1, \ldots, t_n)] \tag{3.19}$$

where $\pi(t_1, \ldots, t_n)$ denotes a permutation of the arguments $t_1, \ldots, t_n$. Thus, for example $\mathbf{b}[t_1, t_2, t_3] = \mathbf{b}[t_2, t_3, t_1]$.

Multiaffinity:

$$\mathbf{b}[(\alpha r + \beta s), *] = \alpha \mathbf{b}[r, *] + \beta \mathbf{b}[s, *]; \;\; \alpha + \beta = 1. \tag{3.20}$$

Here, the symbol $*$ indicates that there are the same arguments on both sides of the equation, but their exact meaning is not of interest. Because of symmetry, this property holds for all arguments, not just the first one.

Diagonality:
If all arguments of the blossom are the same: $t = t_1, \ldots, t_n$, then we obtain a polynomial curve (to be discussed later). We will use the notation

$$\mathbf{b}[t, \ldots, t] = \mathbf{b}[t^{<n>}]$$

if the argument t is repeated n times.

We defined vector arguments for linear interpolation in Section 3.1. Blossoms may also have vector arguments, resulting in expressions such as $\mathbf{b}[\vec{h}, r, s]$. If we assume (without loss of generality) that the first argument of a blossom is a vector $\vec{h} = b - a$, then the multiaffine property becomes

$$\mathbf{b}[b - a, *] = \mathbf{b}[b, *] - \mathbf{b}[a, *]. \tag{3.21}$$

Thus if (at least) one of the blossom arguments is a vector, then the blossom value is a vector. For example, if we denote by $\vec{1}$ the 1D unit vector, then $\mathbf{b}[\vec{1}, r, s] = \mathbf{b}[1, r, s] - \mathbf{b}[0, r, s]$ or $\mathbf{b}[\vec{1}, r, s] = \mathbf{b}[3, r, s] - \mathbf{b}[2, r, s]$.

4 This kind of curve will later be called a Bézier curve.

As an application of the blossom properties, let us derive a formula that will be used later. We consider the special case when a blossom argument is of the form $(\alpha r + \beta s)^{<n>}$. For this, we get

$$\mathbf{b}[(\alpha r + \beta s)^{<n>}] = \sum_{i=0}^{n} \binom{n}{i} \alpha^i \beta^{n-i} \mathbf{b}[r^{<i>}, s^{<n-i>}]. \tag{3.22}$$

We refer to this equation as the *Leibniz formula.*[5]

The proof is by induction. The case $n = 1$ is a trivial start. The inductive step proceeds as follows (keeping in mind that $\binom{n}{n+1} = \binom{n}{-1} = 0$):

$$\begin{aligned} \mathbf{b}[(\alpha r + \beta s)^{<n+1>}] &= \sum_{i=0}^{n} \binom{n}{i} \alpha^i \beta^{n-i} \mathbf{b}[(\alpha r + \beta s), r^{<i>}, s^{<n-i>}] \\ &= \sum_{i=0}^{n} \binom{n}{i} \alpha^{i+1} \beta^{n-i} \mathbf{b}[r^{<i+1>}, s^{<n-i>}] \\ &\quad + \sum_{i=0}^{n} \binom{n}{i} \alpha^i \beta^{n+1-i} \mathbf{b}[(r^{<i>}, s^{<n+1-i>}] \end{aligned}$$

Now we transform the index of the first sum and let the second sum run to $n + 1$:

$$\begin{aligned} &\sum_{i=1}^{n+1} \binom{n}{i-1} \alpha^i \beta^{n+1-i} \mathbf{b}[r^{<i>}, s^{<n+1-i>}] \\ &+ \sum_{i=0}^{n+1} \binom{n}{i} \alpha^i \beta^{n+1-i} \mathbf{b}[r^{<i>}, s^{<n+1-i>}] \end{aligned}$$

The first sum may start with $i = 0$. Keeping in mind the recursion

$$\binom{n+1}{i} = \binom{n}{i-1} + \binom{n}{i}$$

we can combine the last two sums and get

$$\mathbf{b}[(\alpha r + \beta s)^{<n+1>}] = \sum_{i=0}^{n+1} \binom{n+1}{i} \alpha^i \beta^{n+1-i} \mathbf{b}[r^{<i>}, s^{<n+1-i>}],$$

which concludes our proof. This result will be used several times later on.

---

5 It has the structure of Leibniz's rule for higher-order derivatives of a product of functions.

A different form of (3.22) is sometimes useful:

$$\mathbf{b}[(\alpha r + \beta s)^{<n>}] = \sum_{\substack{i+j=n \\ i,j\geq 0}} \binom{n}{i,j} \alpha^i \beta^j \mathbf{b}[r^{<i>}, s^{<j>}] \tag{3.23}$$

where

$$\binom{n}{i,j} = \frac{n!}{i!j!}.$$

3.5 Barycentric Coordinates in the Plane

Barycentric coordinates were discussed in Section 3.1, where they were used in connection with straight lines. Now we will use them as coordinate systems when dealing with the plane. Planar barycentric coordinates are at the origin of affine geometry—they were introduced by F. Moebius in 1827; see his collected works [429].

Consider a triangle with vertices **a**, **b**, **c** and a fourth point **p**, all in $\mathbb{E}^2$. It is always possible to write **p** as a barycentric combination of **a**, **b**, **c**:

$$\mathbf{p} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}. \tag{3.24}$$

A reminder: if (3.24) is to be a barycentric combination (and hence geometrically meaningful), we require that

$$u + v + w = 1. \tag{3.25}$$

The coefficients $\mathbf{u} := (u, v, w)$ are called *barycentric coordinates* of **p** with respect to **a**, **b**, **c**. We will often drop the distinction between the barycentric coordinates of a point and the point itself; we then speak of "the point **u**."

If the four points **a**, **b**, **c**, and **p** are given, we can always determine **p**'s barycentric coordinates u, v, w: Equations (3.24) and (3.25) can be viewed as a linear system of three equations[6] in three unknowns u, v, w. The solution is obtained by an application of Cramer's rule:

$$u = \frac{\text{area}(\mathbf{p}, \mathbf{b}, \mathbf{c})}{\text{area}(\mathbf{a}, \mathbf{b}, \mathbf{c})}, \quad v = \frac{\text{area}(\mathbf{a}, \mathbf{p}, \mathbf{c})}{\text{area}(\mathbf{a}, \mathbf{b}, \mathbf{c})}, \quad w = \frac{\text{area}(\mathbf{a}, \mathbf{b}, \mathbf{p})}{\text{area}(\mathbf{a}, \mathbf{b}, \mathbf{c})}. \tag{3.26}$$

6 Recall that (3.24) is shorthand for two scalar equations.

Actually, Cramer's rule makes use of determinants; they are related to areas by the identity

$$\text{area}(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \frac{1}{2} \begin{vmatrix} a_x & b_x & c_x \\ a_y & b_y & c_y \\ 1 & 1 & 1 \end{vmatrix}. \tag{3.27}$$

We note that in order for (3.26) to be well defined, we require $\text{area}(\mathbf{a}, \mathbf{b}, \mathbf{c}) \neq 0$, which means that **a**, **b**, **c** must not lie on a straight line.

Because of their connection with barycentric combinations, barycentric coordinates are *affinely invariant*: let **p** have barycentric coordinates u, v, w with respect to **a**, **b**, **c**. Now map all four points to another set of four points by an affine map Φ. Then $\Phi\mathbf{p}$ has the same barycentric coordinates u, v, w with respect to $\Phi\mathbf{a}$, $\Phi\mathbf{b}$, $\Phi\mathbf{c}$.

Figure 3.5 illustrates more of the geometric properties of barycentric coordinates.

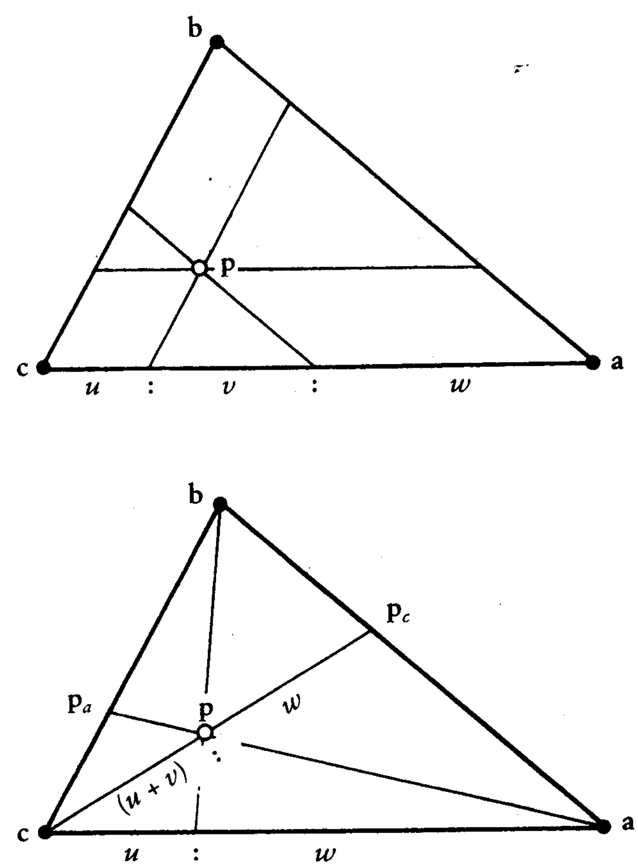

Figure 3.5 Barycentric coordinates: let $\mathbf{p} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$. The two figures show some of the ratios generated by certain straight lines through **p**.

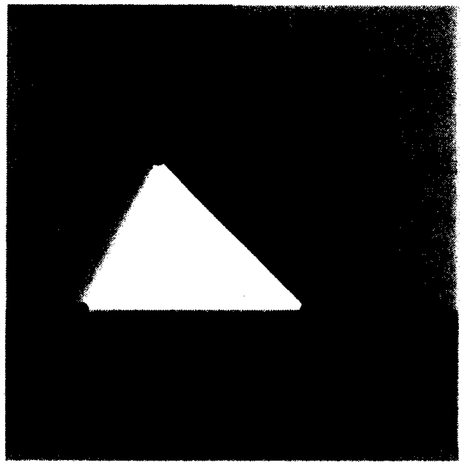

Figure 3.6 Barycentric coordinates: a triangle defines a coordinate system in the plane. Points with three positive barycentric coordinates: white. With one negative barycentric coordinate: light gray. With two negative barycentric coordinates: dark gray.

An immediate consequence of Figure 3.5 is known as *Ceva's theorem:*

$$\text{ratio}(\mathbf{a}, \mathbf{p}_c, \mathbf{b}) \cdot \text{ratio}(\mathbf{b}, \mathbf{p}_a, \mathbf{c}) \cdot \text{ratio}(\mathbf{c}, \mathbf{p}_b, \mathbf{a}) = 1.$$

More details on this and related theorems can be found in most geometry books (e.g., Gans [253] or Berger [52], or Boehm and Prautzsch [85]).

Any three noncollinear points $\mathbf{a}, \mathbf{b}, \mathbf{c}$ define a barycentric coordinate system in the plane. The points inside the triangle $\mathbf{a}, \mathbf{b}, \mathbf{c}$ have positive barycentric coordinates, whereas the remaining ones have (some) negative barycentric coordinates. Figure 3.6 shows more.

We may use barycentric coordinates to define *bivariate linear interpolation.* Suppose we are given three points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \mathbb{E}^3$. Then any point of the form

$$\mathbf{p} = \mathbf{p}(\mathbf{u}) = \mathbf{p}(u, v, w) = u\mathbf{p}_1 + v\mathbf{p}_2 + w\mathbf{p}_3 \tag{3.28}$$

with $u + v + w = 1$ lies in the plane spanned by $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$. This map from $\mathbb{E}^2$ to $\mathbb{E}^3$ is called *linear interpolation.* Since $u + v + w = 1$, we may interpret u, v, w as barycentric coordinates of $\mathbf{p}$ relative to $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$. We may also interpret u, v, w as barycentric coordinates of a point in $\mathbb{E}^2$ relative to some triangle $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{E}^2$. Then (3.28) may be interpreted as a map of the triangle $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{E}^2$ onto the triangle $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3 \in \mathbb{E}^3$. We call the triangle $\mathbf{a}, \mathbf{b}, \mathbf{c}$ the *domain triangle.* Note that the actual location or shape of the domain triangle is totally irrelevant to the definition of linear interpolation. (Of course, we must demand that it be nondegenerate.) Since we can interpret u, v, w as barycentric coordinates in both two and three dimensions, it follows that linear interpolation (3.28) is an affine map.

Barycentric coordinates are not restricted to one and two dimensions; they are defined for spaces of higher dimensions as well. For example, in 3D, any nondegenerate tetrahedron with vertices $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4$ may be used to write any point $\mathbf{p}$ as $\mathbf{p} = u_1\mathbf{p}_1 + u_2\mathbf{p}_2 + u_3\mathbf{p}_3 + u_4\mathbf{p}_4$.

3.6 Tessellations

When dealing with sequences of straight line segments, we were in the context of piecewise linear interpolation. We may also consider more than one triangle, thus introducing bivariate piecewise linear interpolation. Although straight line segments are combined into polygons in a straightforward way, the corresponding concepts for triangles are not so obvious; they are the subject of this section.

We will first introduce the concept of a *Dirichlet tessellation*; this will lead to an efficient way to deal with triangles. So consider a collection of points $\mathbf{p}_i$ in the plane. We are going to construct influence regions around each point in the following way: suppose each point is a transmitter for a cellular phone network. As a car moves through the points $\mathbf{p}_i$, its phone should always be using the closest transmitter. We may think of each transmitter as having an area of influence around it: whenever a car is in a given transmitter's area, its phone switches to that transmitter. More technically speaking, we associate with each point $\mathbf{p}_k$ a *tile* T_k consisting of all points $\mathbf{p}$ that are closer to $\mathbf{p}_k$ than to any other point $\mathbf{p}_i$. The collection of all these tiles is called the *Dirichlet tessellation* of the given point set.[7] Two points are called *neighbors* if their tiles share a common edge. See Figure 3.7.

It is intuitively clear that the tile edges should consist of segments taken from perpendicular bisectors of neighboring points. This observation directly leads to a recursive construction that is due to R. Sibson [576]: suppose that we already constructed the Dirichlet tessellation for a set of points, and we now want to add one more point $\mathbf{p}_L$. First, we determine which of the previously constructed tiles is occupied by $\mathbf{p}_L$; referring to Figure 3.8, let us assume it is T_k. We now draw all perpendicular bisectors between $\mathbf{p}_L$ and its neighbors, thus forming T_L. Continuing in this manner, we can construct the tessellation for an arbitrary number of points. Each point is thus in the "center" of a tile, most of them finite, but some infinite. It is not hard to see that all points with infinite tiles determine the convex hull of the data points; see Section 2.1 for a definition.

Although the preceding method may not be the most efficient one to construct the Dirichlet tessellation for a set of points, it is very intuitive, and also forms the basis of the following fundamental theorem. The tile T_L is formed by cutting

---

7 This structure is also known as a *Voronoi diagram* or *Thiessen regions*.

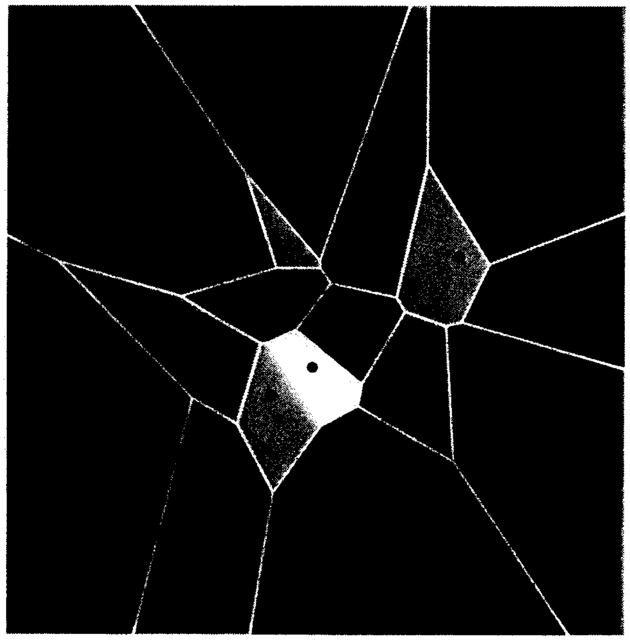

Figure 3.7 Dirichlet tessellations: a point set and its tile edges.

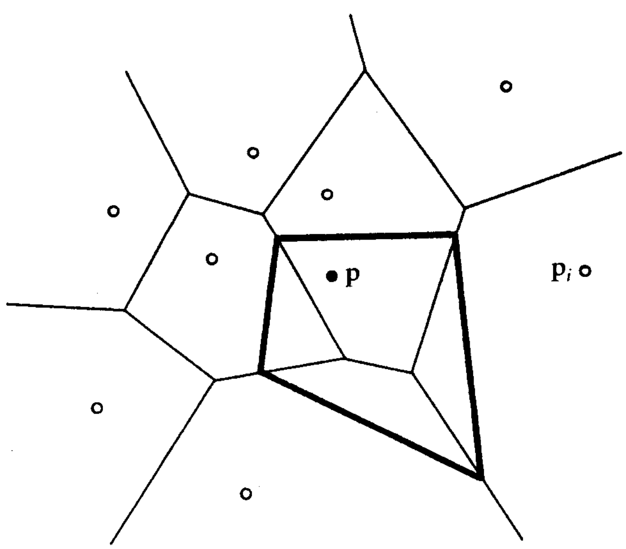

Figure 3.8 Dirichlet tessellations: a new point is inserted into an existing tessellation; its tile is outlined.

out parts of $\mathbf{p}_L$'s neighboring tiles. Let $\mathcal{A}_i$ be the area cut of $\mathbf{T}_i$, and let $\mathcal{A}$ be the area of $\mathbf{T}_L$. Then we can write $\mathbf{p}_L$ as a barycentric combination of its neighbors (note that $\sum \mathcal{A}_i = \mathcal{A}$):

$$\mathbf{p}_L = \sum_i \frac{\mathcal{A}_i}{\mathcal{A}} \mathbf{p}_i. \tag{3.29}$$

This identity is also due to R. Sibson [576]; in case the summation is over only three neighbors, it reduces to the barycentric coordinates of Section 3.5.

3.7 Triangulations

The Dirichlet tessellation of a set of points determines another fundamental structure that is connected with the point set: its *Delaunay triangulation*. If we connect all neighboring points, we have created a set of triangles that cover the convex hull of the point set and that have the given points as their vertices. Figure 3.9 was created in this way from the configuration of Figure 3.7. The points with infinite tiles are now connected; they are called *boundary points* of the triangulation.

We should mention one problem: although the Dirichlet tessellation is unique, the Delaunay triangulation may not be. As an example, consider four points forming a square: either diagonal produces a valid Delaunay triangulation. Four points that have no unique Delaunay triangulation are called *neutral sets*; such points are always cocircular.

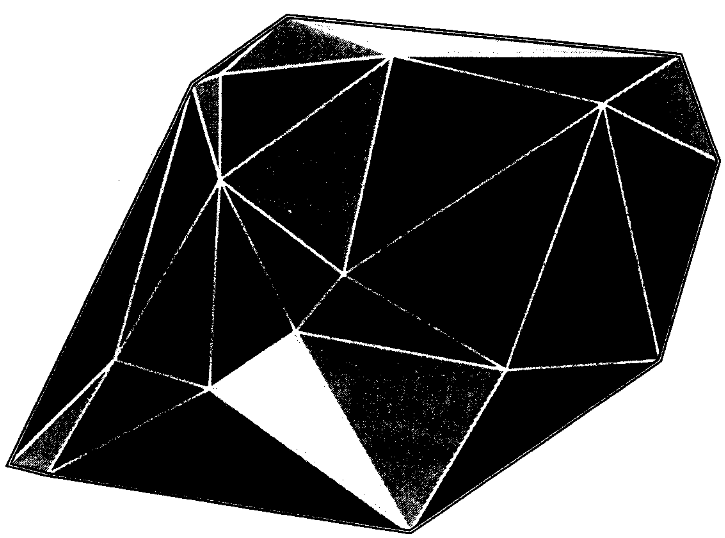

Figure 3.9 Delaunay triangulations: a point set and its Delaunay triangulation.

Clearly, there are many valid triangulations of a given point set. For example, every convex set of four points allows two different triangulations. It is now time to introduce the concept of a triangulation of a point set that is more general than the Delaunay triangulation. A triangulation $\mathcal{T}$ of a set of 2D points $\{\mathbf{p}_i\}$ is a collection of triangles such that

- The vertices of the triangles consist of the $\mathbf{p}_i$
- The interiors of any two triangles do not intersect
- If two triangles are not disjoint, then they share either a vertex or an edge

An important implementation aspect is the type of data structure to be used for triangulations. Data sets with several million points are not unheard of, and for those, an intelligent structure is crucial. Such a structure should have the following elements:

1. A point collection of x, y-coordinate pairs
2. A collection of triangles, each pointing to three elements in the point list and also to three elements in the triangle collection, namely, those that designate a triangle's three neighbors[8]

These collections are best realized in the form of linked lists, for ease of inserting and deleting points. This data structure goes back to F. Little, who implemented it in 1978 at the University of Utah.

As it turns out, the Delaunay triangulation is one of the "nicer" triangulations. Intuitively, we might say that a triangulation is "nice" if it consists of triangles that are close to being equilateral. If we compare two different triangulations of a point set, we might then compute the minimal angle of each triangle. The triangulation that has the *largest* minimal angle would be labeled the better one. Of all possible triangulations, the Delaunay triangulation is the one that is guaranteed to produce the largest minimal angle; for a proof, see Lawson [375]. The Delaunay triangulation is thus said to satisfy the maxmin criterion.

One might also consider the triangulation that satisfies the minmax criterion: the triangulation whose maximal angle is minimal. These triangulations are not easy to compute; one reason is that their neutral point sets are fairly complex, see Hansford [312].

A major use of triangulations is in *piecewise linear interpolation*: suppose that at each data point $\mathbf{p}_k$ we are given a function value z_k. Then we may construct a linear interpolant—using linear interpolation from Section 3.5—over each of the

8 Boundary triangles may have only one or two neighbors.

triangles. We obtain a faceted, continuous surface that interpolates to all given data. This surface is not smooth, but it will give a decent idea of the shape of the given data. One application is in cartography: here, the given data points might be coordinates obtained from satellite readings and the function values might be their elevations. Our piecewise linear surface is an approximation to the landscape being surveyed.

Once function values are involved, it may be advantageous to construct a triangulation that reflects this information. Such triangulations are called *data dependent*; see Dyn, Levin, and Rippa [180] or Brown [92]. Here, one does not just consider triangles in the plane, but rather the 3D triangles generated by the data points (x_k, y_k, z_k).

3.8 **Problems**

1 In the definition of the variation diminishing property, we counted the crossings of a polygon with a plane. Discuss the case when the plane contains a whole polygon leg.

***2** We defined the convex hull of a point set to be the set of all convex combinations formed by the elements of that set. Another definition is the following: the convex hull of a point set is the intersection of all convex sets that contain the given set. Show that both definitions are equivalent.

***3** Our definition of barycentric combinations gives the impression that it needs the involved points expressed in terms of some coordinate system. Show that this is not necessary: draw five points on a piece of paper, assign a weight to each one, and *construct* the barycenter of your points using a ruler (or compass and straightedge if you are more classically inclined).

Remark: For this construction, it is not necessary for the weights to sum to one. This is so because the geometric construction remains the same if we multiplied all weights by a common factor. In fact, one may replace the concept of points (having mass one and requiring barycentric combinations as the basic point operation) by that of *mass points*, having arbitrary weights and yielding their barycenter (with the combined mass of all points) as the basic operation. In such a setting, vectors would also be mass points, but with mass zero.[9]

***4** Let a triangulation consist of b boundary points and of i interior points. Show that the number of triangles is $2i + b - 2$.

---

9 I was introduced to this concept by A. Swimmer. It was developed by H. Grassmann in 1844.

P1 Let three points be given by

$$\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 40 \\ 20 \end{bmatrix}, \begin{bmatrix} 50 \\ 10 \end{bmatrix}.$$

For $s = 0, 0.05, 0.1, \ldots, 1$ and $t = 0, 0.05, 0.1, \ldots, 1$, plot the points $\mathbf{b}[s, t]$ as defined by (3.16). Mark each point by a circle with radius 0.4.

P2 There is a 2D triangulation data set on this book's web site. Plot that triangulation using gray shades or colors such that no two neighboring triangles have the same color.

P3 Use the recursive algorithm from Section 3.6 to implement Dirichlet tessellations.