MA 323 Geometric Modelling
Course Notes: Day 11
Barycentric Coordinates
and de Casteljau's algorithm

David L. Finn

December 16th, 2004

Today, we introduce barycentric coordinates as an alternate to using a rigid coordinate system, and immediately use barycentric coordinates to define de Casteljau's algorithm in a pure geometric manner without recourse to analytic geometry. Barycentric coordinates will be one of the main tools that we will use in the remainder of the course, as they provide a geometric manner to construct coordinate system and avoid the explicit use of coordinate systems. The constructions that we use are the basis for affine geometry which is the geometry of vector spaces, and is the natural generalization of the geometry of the Euclidean plane without the use of angles and distances.

11.1 Oblique Coordinate Systems

We start with the abstract description of a coordinate system in a plane, and develop our description of a point in the plane from the vector description of a coordinate system. Recall a coordinate system (an oblique coordinate system) is given by choosing two intersecting lines l and m. Let O be the point of intersection, A be a different point on line l than O and B be a different point on line m than O. The coordinates of a point P in the plane described by O, A and B are given by letting X be the intersection of a line parallel to m through P with the line l and Y be the intersection of a line parallel to l through P with the line m. See diagram below. We then let x be number representing the proportion OX/OA if X and A are on the same side of O, and we let $x = -OX/OA$ if X and A are on opposite sides of O (a negative number represents the other side of O than A). Likewise, we define y through the proportion OY/OB.

In terms of vectors (where a vector is defined as the difference of two points and the ray from one point to another point, and multiples of a vector are obtained by considering the line through the two points and constructing a line segment that is a multiple of the original line segment), we thus have $\overrightarrow{OX} = x\,\overrightarrow{OA}$ and $\overrightarrow{OY} = y\,\overrightarrow{OB}$. Recalling that two vectors $\overrightarrow{PQ}$ and $\overrightarrow{RS}$ are equal if a line through P to QS intersects the extended line RS at the point R (the pair of line segments PQ, RS are opposite sides of a parallelogram and the pair of line segments PR, QS are opposite sides of a parallelogram), we can write

$$\overrightarrow{OP} = x\overrightarrow{OA} + y\overrightarrow{OB}$$

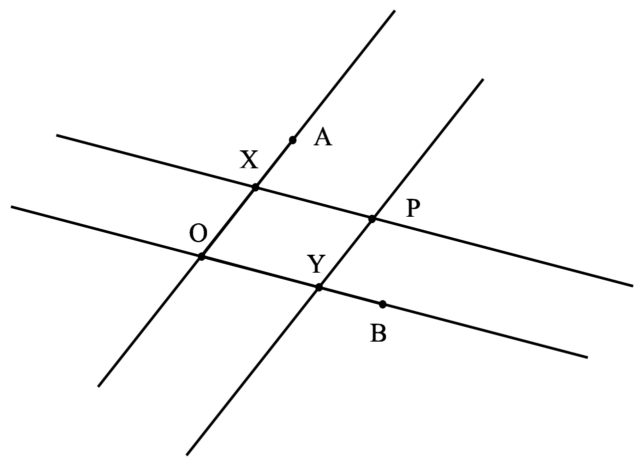

Figure 1: Construction of an Oblique Coordinate System

because $x\overrightarrow{OA} = \overrightarrow{OX}$, $y\overrightarrow{OB} = \overrightarrow{OY} = \overrightarrow{XP}$). Writing $\overrightarrow{OP} = P - O$, $\overrightarrow{OA} = A - O$, and $\overrightarrow{OB} = B - O$, we have

$$\begin{aligned} P &= O + x\,\overrightarrow{OA} + y\,\overrightarrow{OB} \\ &= O + x\ (A - 0) + y\ (B - 0) \\ &= (1 - x - y)\,O + x\,A + y\,B. \end{aligned}$$

The second description gives us a unique description of every point in the plane defined by three non-collinear points O, A, B directly in terms of the points rather than through defining vectors.

We can extend this description of a point in a plane to a point in space by using four non planar points to define three (linearly independent) vectors with a common base point. Thus, we can write a point P in space in terms of four non-planar points O, A, B, C as

$$\begin{aligned} P &= O + x\,\overrightarrow{OA} + y\,\overrightarrow{OB} + z\,\overrightarrow{OC} \\ P &= (1 - x - y - z)\,O + x\,A + y\,B + z\,C. \end{aligned}$$

Notice that both the description of a point in a plane or a point in space can be written in the form

$$P = \alpha_0\,P_0 + \alpha_1\,P_1 + \cdots + \alpha_n\,P_n$$

where $\{\alpha_i\}$ is a sequence of numbers with $\sum_i \alpha_i = 1$ and $\{P_i\}$ is a sequence of points and $n = 2$ or $n = 3$.

This generalizes further to the space of n-dimensional points $\mathbb{E}^n$, where $\mathbb{E}^n$ consists of the space of points that can be represented as n-tuples of real numbers. We are reserving the notation $\mathbb{R}^n$ for the n-dimensional vector space of real numbers. This is a notational convenience to distinguish between points and vectors. To state this generalization in full detail, we define an affine frame as a point O and a set of n linearly independent vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n\}$ in $\mathbb{R}^n$. Since every vector $\mathbf{u}$ in $\mathbb{R}^n$ can be written uniquely as

$$\mathbf{u} = x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_n\,\mathbf{v}_n,$$

setting $\vec{OP} = \mathbf{u}$ we get

$$P = O + x_1\,\mathbf{v}_1 + x_2\,\mathbf{v}_2 + \cdots + x_n\,\mathbf{v}_n.$$

Alternatively, an affine frame for $\mathbb{E}^n$ can be described as collection of $n+1$ points in general position $\{P_0, P_1, \cdots, P_n\}$. The term in general position means that the vectors $\mathbf{v}_1 = P_1 - P_0$, $\mathbf{v}_2 = P_2 - P_0, \cdots, \mathbf{v}_n = P_n - P_0$ are linearly independent. The term general position is meant to reflect that if you choose $n+1$ at random you should have an affine frame. Rewriting this description, we have the point P in the affine frame $\{P_i\}$ to be

$$P = \sum_{i=0}^{n} \alpha_i P_i$$

where $\sum_{i=0}^{n} \alpha_i = 1$. The way to view an affine frame is as the point P_0 as the origin of a coordinate system and the vectors $\mathbf{v}_i = P_i - P_0 = \overrightarrow{P_0 P_i}$ as describing the coordinate axes.

11.2 Barycentric Coordinates

We have just shown that every point in a plane can be described as the addition of three points provided the sum of the coefficients of the points add up to 1. This is an important fact that we will be using throughout the course. This generalizes to higher dimensions, but for now we concentrate on constructions in the plane.

> Every point P in the plane can be written as a combination of three given non-collinear points P_0, P_1 and P_2 as follows
>
> $$P = \alpha_0 P_0 + \alpha_1 P_1 + \alpha_2 P_2$$
>
> with $\alpha_0 + \alpha_1 + \alpha_2 = 1$. The triple $(\alpha_0, \alpha_1, \alpha_2)$ is called the barycentric coordinates of the point P with respect to the points P_0, P_1, and P_2.

The coefficients α_0, α_1, and α_2 of the points P_0, P_1, P_2 represent the weight or influence of the point P_0, P_1 and P_2 in determining the position of the point P.

The notion of barycentric coordinates of a point can be extended to an arbitrary number of points (irrevelant of the dimension of the space). In particular, given points $\{P_i\}$ and coefficients $\{\alpha_i\}$ with $\sum \alpha_i = 1$ we can describe a point P as

$$P = \alpha_0 P_0 + \alpha_1 P_1 + \cdots + \alpha_n P_n.$$

This means that given four points in a plane, we can describe a point as $\alpha_0 P_0 + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3$. The description in this manner is not unique, but it is possible. The uniqueness of the description is a property of the dimension of the space, and whether or not the points $\{P_i\}$ form an affine frame.

It is worth noting that in affine geometry, the following basic result specifies the manner in which points $\{P_0, P_1, P_2, \cdots, P_n\}$ can be combined. An affine combination is defined to be a sum

$$\sum \alpha_i P_i$$

for numbers α_i and points P_i. Affine combinations make sense in two situations either $\sum \alpha_i = 0$ or $\sum \alpha_i = 1$. In the case $\sum \alpha_i = 0$, the combination $\sum \alpha_i P_i$ is a vector and in the case $\sum \alpha_i = 1$, the combination is a point.

To examine in detail the meaning of the coefficients α_i, let's consider the barycentric coordinates of a point P on a line AB. Any point on the line AB can be written as

$$P = (1-t)\, A + t\, B.$$

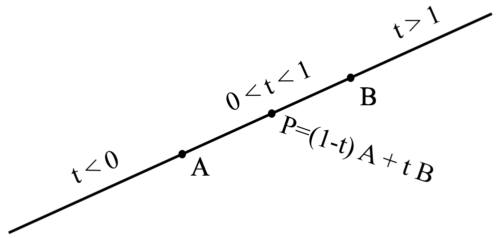

Figure 2: Barycentric Coordinates of a Line

When $t = 0$, we get the point A and when $t = 1$ we get the point B. A third special point on the line AB is the midpoint and this is obtained when $t = 1/2$. Using barycentric coordinates, any point on the line AB can be written as

$$\alpha\, A + \beta\, B$$

with $\alpha + \beta = 1$. When α and β are both greater than zero, the point $P = \alpha\, A + \beta\, B$ is between the points A and B. When coefficients α is close to one, the point P is close to the point A. Similarly when the coefficient β is close to one, then the point P is close to B.

When one of the coefficients is greater than one then the other must be less than zero, which will imply that the point is outside the segment AB, and closer to the point with a positive coefficient. This may be seen by the example $P = \frac{5}{3}\, A - \frac{2}{3}\, B$. Write $P = A + \frac{2}{3}\,(A - B)$ and then P must be on other side of A than B.

In general, the points on the line segment are obtained by comparing the line AB with a number line, and associating the point A with the number 0, the point B with the number 1. Let t be an arbitrary number on the number line. If the point P on the line AB is described by $P = (1 - t)A + tB$, this means that proportion of the lengths of the segments $AP :: BP$ is given by the fraction $t/(1 - t)$. The number $t/(1 - t)$ is called the ratio of the points A, P, B and denoted by $\text{ratio}(A, P, B)$. In the diagram below, the line OL represents the number line with the segment OL being of unit length and the point T on the line OL represents the length t (i.e. OT has length t). The proportion $AP :: BP$ representing the fraction $t/(1 - t)$ means that the proportions $AP :: BP$ equals the proportion $OT :: LT$.

An alternative method for interpreting the coefficients α and β in representing the point P on the line AB is to let $\alpha = PB :: AB$ where the ratio is positive if P is on the same side of B as A and $\beta = AP :: AB$ where the ratio is positive if P is on the same side of A as B.

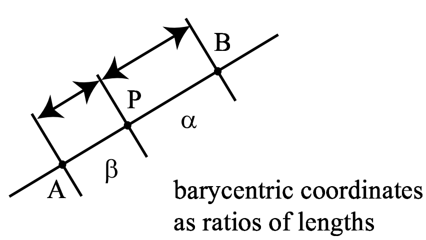

Figure 3: Alternate view of barycentric coordinates

For a point P in the plane, the coefficients α_0, α_1 and α_2 represent the contribution of the points P_0, P_1 and P_2 in representing the point P, and again if α_i is close to 1 the point

P is close to P_i. An important fact about barycentric coordinates is that the point P is inside the triangle $P_0P_1P_2$ if all the barycentric coordinates are positive. One important point inside a triangle is the center or centroid of the triangle. This point is described by the barycentric coordinates $(\alpha_0, \alpha_1, \alpha_2) = (1/3, 1/3, 1/3)$. This point makes the area of the triangles P_0P_1P, P_1P_2P and P_0P_2P all equal, and thus the triangle is balanced at the point P assuming the triangle is made out of a homogeneous material; a material with constant density.

To calculate the barycentric coordinates of a point P in a plane with respect to an affine frame P_0, P_1, P_2, one can proceed as follows. First, determine the intersection of the lines PP_0 and P_1P_2. Call this point Q. The point Q has barycentric coordinates $Q = \beta_1 P_1 + \beta_2 P_2$. With respect to P_0 and Q the point P has barycentric coordinates $\gamma_0 P_0 + \gamma_1 Q$, which yields the coordinates

$$\begin{aligned} P &= \gamma_0 P_0 + \gamma_1 (\beta_1 P_1 + \beta_2 P_2) \\ &= \gamma_0 P_0 + (\gamma_1\beta_1) P_1 + (\gamma_1\beta_2) P_2. \end{aligned}$$

An illustration of this calculation is in the diagram below.

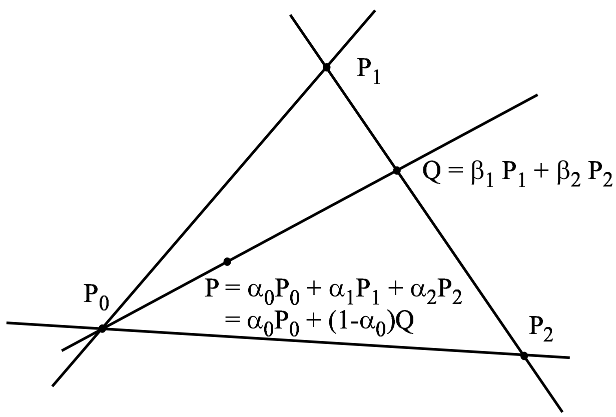

Figure 4: Barycentric coordinates in a plane

Determining a point with barycentric coordinates $(\alpha_0, \alpha_1, \alpha_2)$ for an affine frame $\{P_0, P_1, P_2\}$ may be done by reversing the above procedure. Define $\beta_1 = \alpha_1/(1-\alpha_0)$ and $\beta_2 = \alpha_2/(1-\alpha_0)$, then determine the point $Q = \beta_1 P_1 + \beta_2 P_2$ on the line P_1P_2. The point P is determined by finding the point $P = \alpha_0 P_0 + (1-\alpha_0) Q$. Notice that in this construction, one does not need to know the coordinates of the points $\{P_0, P_1, P_2\}$ and we have only used repeated linear interpolation.

The above procedure is just decomposing barycentric coordinates into repeated linear interpolation. This is a useful way of viewing barycentric coordinates. In fact, the following quote provides a fundamental viewpoint of linear interpolation.

> *Most of the computations that we use in CAGD may be broken down into seemingly trivial steps - sequences of linear interpolation. It is therefore important to understand the properties of these basic building blocks.* (Opening of Chapter Three: Linear Interpolation in **CAGD: Curves and Surfaces in Geometric Design** by G. Farin)

This viewpoint will be justified shortly when we start building curves from barycentric coordinates.

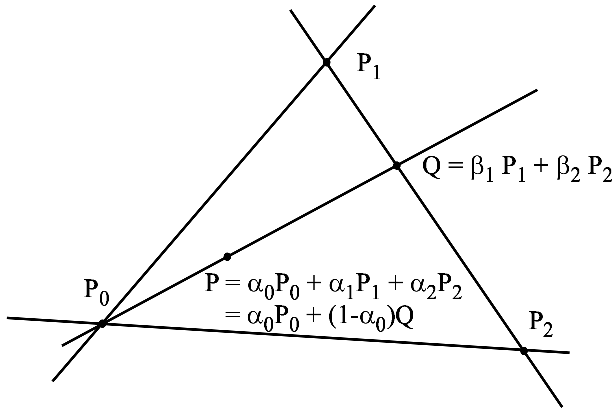

Figure 5: Barycentric coordinates by repeated linear interpolation

There is an alternate view of barycentric coordinates of a point in a plane P with respect to the non-collinear points P_0, P_1, P_2 can be interpreted as a ratio of areas. The coefficient $\alpha_0 = |PP_1P_2|/|P_0P_1P_2|$, the coefficient $\alpha_1 = |PP_0P_2|/|P_0P_1P_2|$, and the coefficient $\alpha_2 = |PP_0P_1|/|P_0P_1P_2|$, where $|ABC|$ stands for the area of the triangle with vertices A, B, C. This interpretation makes sense when P is inside the triangle formed by P_0P1_P2, and outside the triangle once one defines a signed area. [Area is a signed quantity, in the right mathematical notation.] Generalizing this viewpoint, in space, one can interpret the coefficients with respect to four non-planar points as a ratio of volumes.

11.3 Affine Construction of a Parabola

The geometric construction of a parabola where a parabolic arc from the two endpoints of the parabolic arc and the intersection of the tangent lines at the endpoints (see diagram below) is actually an affine construction

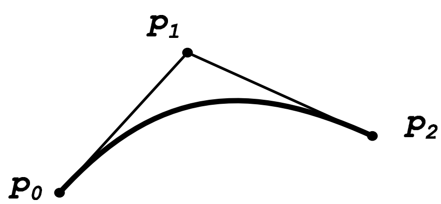

Figure 6: Affine construction of a parabola

Rather than just recall the formula, we will derive it here in another way. A parabolic arc passing through p_0 at time $t = 0$ and through p_1 at $t = 1$ can be given by

$$c(t) = (1-t)^2\, p_0 + a(t)\, p_1 + t^2\, p_2$$

where $a(t)$ is the basis function for the other given point p_1. This basis function must satisfy $a(0) = 0$ and $a(1) = 0$ so $a(t) = b\, t(1-t)$. We want $c'(0) = \lambda\, (p_1 - p_0)$ and $c'(1) = \mu\, (p_2 - p_1)$ so that p_1 is the intersection of the tangent lines. Differentiating, we have

$$c'(t) = -2(1-t)\, p_0 + b(1-t)\, p_1 - bt\, p_1 + 2t\, p_2$$

from which it follows that $c'(0) = -2\,p_0 + b\,p_1$ and $c'(1) = 2\,p_2 - b\,p_1$. Taking $b = \mu = \lambda = 2$. We thus have the parabolic arc

$$c(t) = (1-t)^2\,p_0 + 2t(1-t)\,p_1 + t^2\,p_2.$$

We note that this is an affine construction because the coefficients of the points $(1-t)^2$, $2t(1-t)$ and t^2 sum to one,

$$(1-t)^2 + 2t(1-t) + t^2 = 1 - 2t + t^2 + 2t - 2t + t^2 = 1.$$

Using the viewpoint of barycentric coordinates as repeated linear interpolation, the parabolic arc may be viewed as

$$c(t) = (1-t)((1-t)p_0 + tp_1) + t((1-t)p_1 + tp_2)$$

where we find points on the line segment p_0p_1 and p_1p_2 and then interpolate between these two points. This construction geometrically can be done by noting that the tangent line of c at $t = \tau$ (with $0 \le \tau \le 1$),

$$\begin{aligned} L &= c(\tau) + c'(\tau)\,s \\ &= (1-\tau)^2\,p_0 + 2\tau(1-\tau)\,p_1 + \tau^2\,p_2 + s\,(2(1-\tau)\,(p_1 - p_0) + 2\tau(p_2 - p_1)), \\ &= (1-\tau)(1-\tau-2s)\,p_0 + 2(\tau(1-\tau) + s(1-2\tau))\,p_1 + \tau(\tau+2s)\,p_2 \end{aligned}$$

intersects the segment p_0p_1 at

$$q_0 = (1-\tau)\,p_0 + \tau\,p_1$$

and the segment p_1p_2 at

$$q_1 = (1-\tau)\,p_1 + \tau\,p_2.$$

This follows by noting that using barycentric coordinates the intersection of L with the line segment p_0p_1 occurs when the coefficient of p_2 equals 0, which implies $s = -\tau/2$ and substituting this into L yields $(1-\tau)\,p_0 + \tau\,p_1$. Similarly the intersection of L with p_1p_2 occurs when the coefficient of p_0 equals 0, which implies $s = -(1-\tau)/2$ and substituting this into L yields $(1-\tau)\,p_1 + \tau\,p_2$. Thus, the point $c(\tau) = (1-\tau)\,q_0 + \tau\,q_1$. The parabola can then be obtained by repeated linear interpolation, see diagram below.

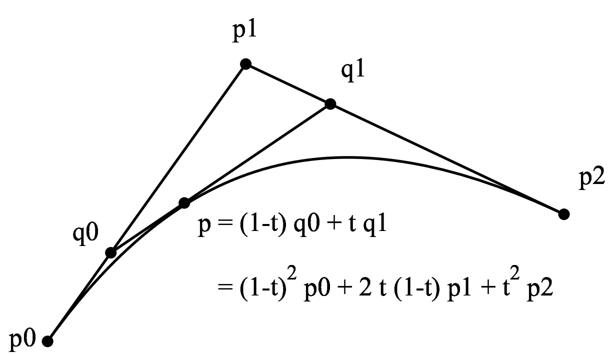

Figure 7: Repeated linear interpolation construction of a parabola

This idea of repeated linear interpolation to create a curve is the crux of de Casteljau's algorithm. This idea also motivates the majority of the the methods used in the remainder of this course, as most of them are extension and generalization of de Casteljau's algorithm. It is for this reason that the quote near the end of the previous subsection is pertinent and poignant.

11.4 de Casteljau's algorithm

Given a sequence of points $p_0,\ p_1,\ \cdots,\ p_n$, de Casteljau's algorithm provides a method for constructing a curve from these points. The curve will not necessarily pass through these points. [The algorithm will not construct an interpolant to these points, solve the interpolation problem given the points.] We thus call the points $p_0,\ p_1,\ \cdots,\ p_n$ control points, as they will control the shape of the curve, through the algorithm. The linear interpolant (piecewise linear curve) through the points $p_0,\ p_1,\ \cdots,\ p_n$ is called the control polyline of the points. The control polyline will be important in later section, when we want to deduce geometric properties about the curve.

From the control points $p_0,\ p_1,\ \cdots,\ p_n$, we define points on the line segments p_ip_{i+1} by choosing a value t and defining $p_i^1 = (1-t)\,p_i + t\,p_{i+1}$. Notice that we defined n points in this manner as $i = 0, 1, 2, \cdots, n-1$. From the n points p_i^1, we can repeat the process and define the $n-1$ points $p_i^2 = (1-t)\,p_i^1 + t\,p_{i+1}^1$ on the line segments $p_i^1p_{i+1}^1$. This process can be repeated defining

$$p_i^{j+1} = (1-t)\,p_i^j + t\,p_{i+1}^j$$

for $i = 0, 1, 2, \cdots, n-j-1$ for $j = 0, 1, 2, \cdots, n-1$ where $p_i^0 = p_i$. The end of this process produces a point p_0^n on a polynomial curve of degree n. An example of this procedure is shown in the diagram below constructing a cubic curve.

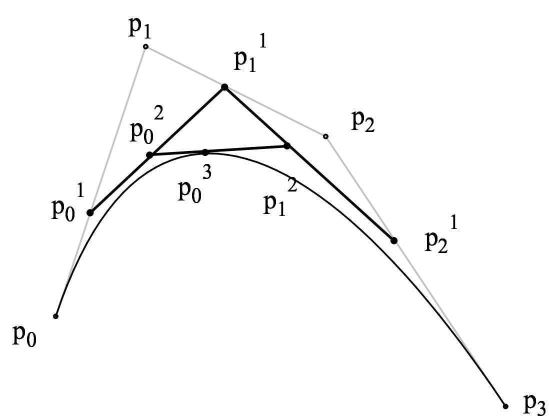

Figure 8: Construction of a cubic curve by repeated linear interpolation

It is worth noting that in the diagram, we obtain two approximations to the curve from using linear interpolation. The first linear interpolant uses the *control polyline* the points $p_0,\ p_1,\ p_2,\ p_3$. The second linear interpolant uses the points $p_0^0 = p_0,\ p_0^1,\ p_0^2,\ p_0^3\ ,\ p_1^2,\ p_2^1,\ p_3^0 = p_3$. The second interpolant is better, that is closer to the curve.

In general, one can use de Casteljau's algorithm to produce points on the curve, and then approximate the curve by using interpolation on the points created. In the diagram below, the points $q_0,\ q_1,\ q_2,\ q_3,\ q_4$ are created by applying de Casteljau's algorithm with $t = 1/4$, $t = 1/2$, and $t = 3/4$.

Illustrating de Casteljau's algorithm

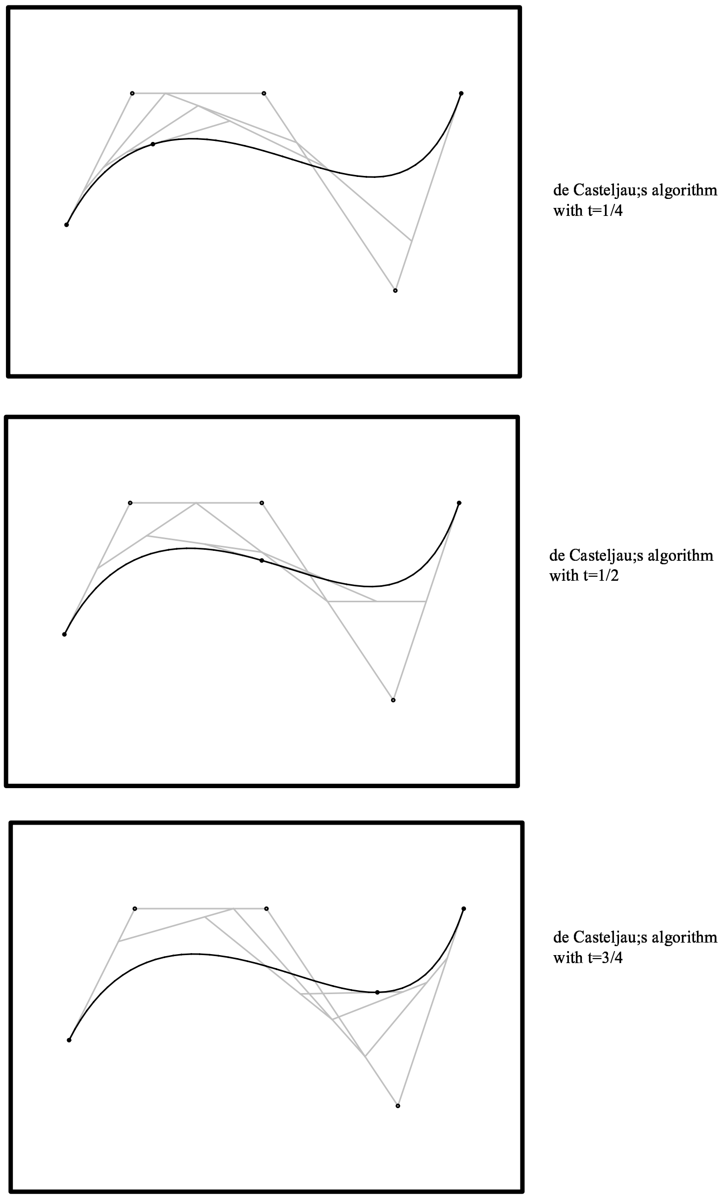

Figure 9: Illustration of de Casteljau's algorithm

We can write the curve generated by de Casteljau's algorithm as an affine combination of the control points by expanding the iterative expressions. We thus obtain a basis function representation of the curve $c(t)$ as the affine combination,

$$c(t) = \alpha_0(t)\, p_0 + \alpha_1(t)\, p_1 + \cdots + \alpha_n(t)\, p_n$$

where $\alpha_0(t) + \alpha_1(t) + \cdots + \alpha_n(t) = 1$. Determining the functions $\alpha_i(t)$ is standardly done by induction (see the Section 4-4). Here, we compute coefficients for the case of cubic curves. We note from the parabolic case (p_i^2), we have

$$\begin{aligned} p_i^2 &= (1-t)\, p_i^1 + t\, p_{i+1}^1 \\ &= (1-t)\,((1-t)\, p_i + t\, p_{i+1}) + t\,((1-t)\, p_{i+1} + t\, p_{i+2}) \\ &= (1-t)^2\, p_i + 2t(1-t)\, p_{i+1} + t^2\, p_{i+2}, \end{aligned}$$

which is the case for quadratic curves on p_i, p_{i+1}, p_{i+2}. For cubic curves, we thus have

$$\begin{aligned} p_0^3 &= (1-t)\, p_0^2 + t\, p_1^2 \\ &= (1-t)\,((1-t)^2\, p_0 + 2t(1-t)\, p_1 + t^2\, p_2) + t\,((1-t)^2\, p_1 + 2t(1-t)\, p_2 + t^3\, p_3) \\ &= (1-t)^3\, p_0 + 3t(1-t)^2\, p_1 + 3t^2(1-t)\, p_2 + t^3\, p_3 \end{aligned}$$

We note that $(1-t)^3 + 3t(1-t)^2 + 3t^2(1-t) + t^3 = (1-t+t)^3 = 1^3 = 1$ by the binomial theorem. The functions $(1-t)^3$, $3t(1-t)^2$, $3t^2(1-t)$, t^3 are plotted below.

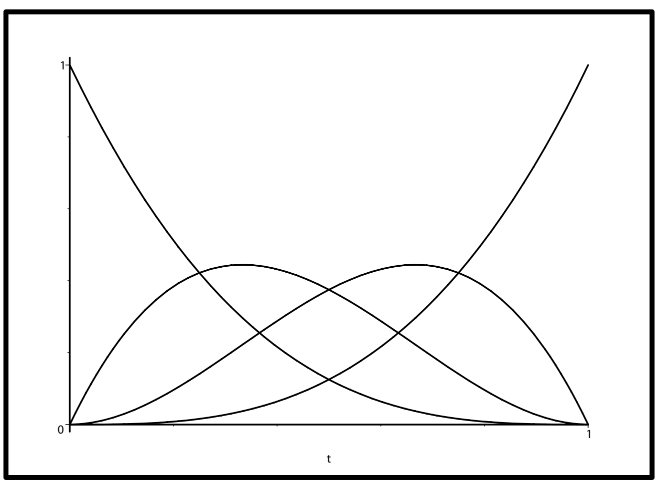

Figure 10: Basis functions for cubic functions from de Casteljau's algorithm

11.5 Exercises

1. (BARYCENTRIC COMPUTATIONS) Consider the triangle ABC below. Determine the position of the the points described by
   (a) $P = \frac{1}{4}A + \frac{1}{2}B + \frac{1}{4}C$
   (b) $P = \frac{1}{3}A + \frac{1}{2}B + \frac{1}{6}C$
   (c) $P = -\frac{1}{6}A + \frac{2}{3}B + \frac{1}{2}C$
   (d) $P = -A + \frac{3}{2}B + \frac{1}{2}C$

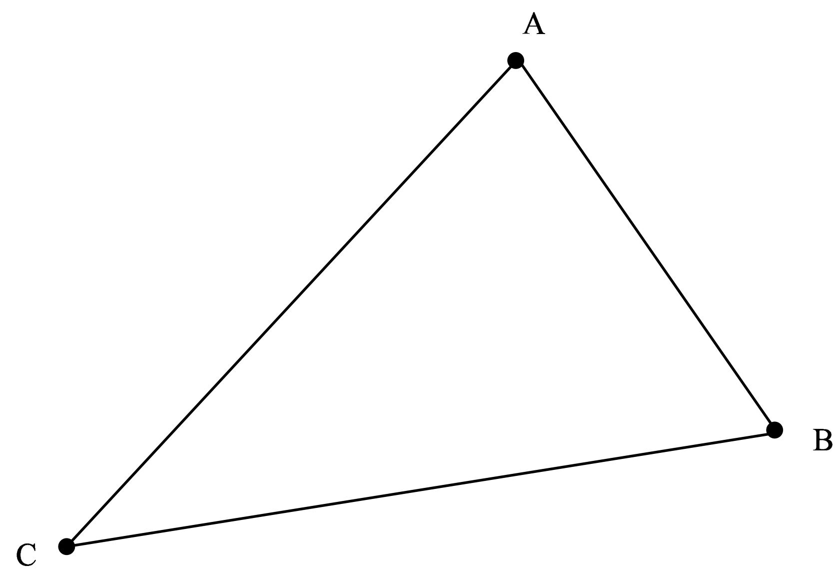

Figure 11: Compute barycentric coordinates of points from triangle ABC

2. (BARYCENTRIC COMPUTATIONS) Determine the barycentric coordinates of the center of the circumscribed and inscribed circles of the points $A = [1, 2]$, $B = [2, 5]$, $C = [6, 1]$. [Circumscribed circle is the circle with each vertex of the triangle is on the circle and the inscribed circle is the circle with the sides of the triangle as tangents to the circle.]

3. (THOUGHT PROBLEM) Are the barycentric coordinates of the center of the circumscribed circle of a triangle the same for each triangle or do the coordinates vary with the triangle? What about the barycentric coordinates of the inscribed circle of a triangle?

4. Complete the interactive exercises for de Casteljau's algorithm.

5. Apply de Casteljau's algorithm on the control points below when $t = 1/3$ and $t = 1/2$.

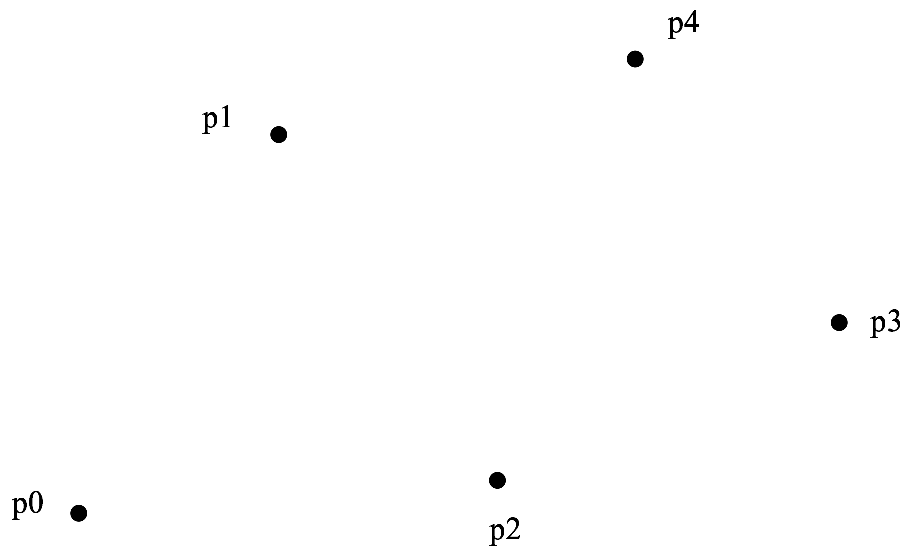

Figure 12: Apply de Casteljau's algorithm with these control points

6. COMPUTATIONAL PROBLEM: Show that the tangent line of the curve $c(t)$ generated by de Casteljau's algorithm when $t = 0$ is given by the line through p_1 and p_0, and likewise the tangent line when $t = 1$ is given by the line through p_{n-1} and p_n.

7. COMPUTATIONAL PROBLEM: Determine the derivative of $P(t)$ generated by de Casteljau's algorithm in terms of the intermediary points $P_i^j(t)$. The tangent line should be obvious from the interactive exercises. [Show analytically that this is the correct tangent line, by computing the derivative.]

8. THOUGHT PROBLEM: Can you define a polygon (from the control points) that necessarily contains the curve generated by de Casteljau's algorithm? Why or why not?