Differential Geometry and Design of Shape and Motion

Jens Gravesen
Department of Mathematics
Technical University of Denmark

Differential Geometry and Design of Shape and Motion

Jens Gravesen
Department of Mathematics
Technical University of Denmark

Lecture notes for 01243
November 8, 2002

Differential Geometry and Design of Shape and Motion

The picture on the front was drawn on a white board by Nana Brül, except for the numbers which was added by Poul Hjorth. The drawning was electronical photographed by Helle Lægring and converted to vector graphics by Jens Gravesen.

Preface

These note are written for the course *Differential Geometry and Design of Shape and Motion* at Technical University of Denmark. The aim of the course is twofold. Firstly we describe the basic algorithms for handling Bézier and B-spline curves and surfaces with their rational variants, (eg. NURBS), which are widely used as a modelling tool in many scientific and engineering applications.

Secondly we give an introduction to the differential geometry of curves and surfaces in the plane and in space.

Contents

Chapter 1

Polynomial Curves

1.1 Introduction

The acronym CAGD stands for *Computer Aided Geometric Design* and the field is concerned with specifying and analyzing classes of curves (and surfaces) which can be used to model free form shapes, e.g. in CAD-systems. For a short historical introduction, see [8, Chapter 1]

Suppose we are to decide on a class of curves to be used in CAGD. Which requirements would we want such a class to fulfill? An obvious list is the following:

- The curves should be able to produce any shape to any given precision.
- It should be easy to evaluate points, derivatives, etc. of the curves.
- The curves should be easy and intuitive to manipulate.

We have already stated that we want the curves to model any shape we like so the first point is obvious. We want computers to handle the curves and we often want the process to be interactive which means that all calculations have to be very fast, hence the second requirement is necessary. Finally if we have an interactive process, the designer should not be required to know any mathematics (this should be handled by the computer) and the program should give the designer some intuitive "handles" which can be used to manipulate the curves.

If we think a little about the Taylor expansions of a curve we see that we can always approximate any (suitably differentiable[1]) curve locally by a polynomial,

[1]This is actually not required. Due to Weierstrass' approximation theorem, any continuous curve defined on a closed interval can be approximated by a polynomial curve, see Problem 1.3.6.

so the class of piecewise polynomial curves satisfies the first requirement. It is likewise easy to evaluate polynomials, so the second requirement is also fulfilled. In a short while we will see that the last requirement is satisfied as well, and this is the reason this class of curves is so popular. If a complicated shape is modeled by a single polynomial curve then the degree of this curve can be very high. In order to avoid that the curve is divided into small simple segments, and then each segment can be modeled by a polynomial curve of low degree. The industry standard is in fact piecewise rational curves[2] which offer a bit more flexibility, but more importantly they give the possibility to represent e.g. circles exactly.

1.2 Polynomial curves

Let us look at the following example of a polynomial curve of degree 3 in the plane:

$$\mathbf{r}(t) = \left(3t + 6t^2 - 3t^3, 6t - 6t^2\right), \qquad t \in [0, 1], \tag{1.1}$$

In order to tell a computer about this curve we have to use some numbers which characterize the curve. The first thing which comes to mind is to give the coefficients with respect to the *power basis*, $\{1, t, t^2, \dots\}$. I.e., we write the curve as

$$\mathbf{r}(t) = 1 \cdot (0, 0) + t \cdot (3, 6) + t^2 \cdot (6, -6) + t^3(-3, 0), \tag{1.2}$$

and use the pairs $(0, 0)$, $(3, 6)$, $(6, -6)$, and $(-3, 0)$ as input to the computer. The geometric interpretation of these coefficients is the set of the derivatives of the curve up to order 3 at the parameter value $t = 0$, see Figure 1.1. These derivatives are obviously *not* good intuitive handles for a designer. They do provide control in one end of the curve, but only the first few derivatives give immediately predictible control of the curve, and it is impossible to guess what happens at the other end. This is not because there is something wrong with polynomial curves, but because it is a bad idea to use the power basis to represent polynomial curves. We need to come up with another basis for the polynomials (of degree 3 in this case). One other choice could be the so called *Hermite polynomials* $H_{kl}(t)$, $k, l = 0, 1$ which are uniquely defined by the following equations:

$$H_{kl}^{(i)}(t_j) = \delta_{ki}\delta_{lj} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j, \\ 0 & \text{otherwise,} \end{cases} \tag{1.3}$$

where $i = 0, 1$ denotes the order of the derivative, $t_0 = 0$ and $t_1 = 1$ are the endpoints of the curve, and $k, l = 0, 1$. The Hermite polynomials are explicitly

[2] Called NURBS for nonuniform rational B-splines

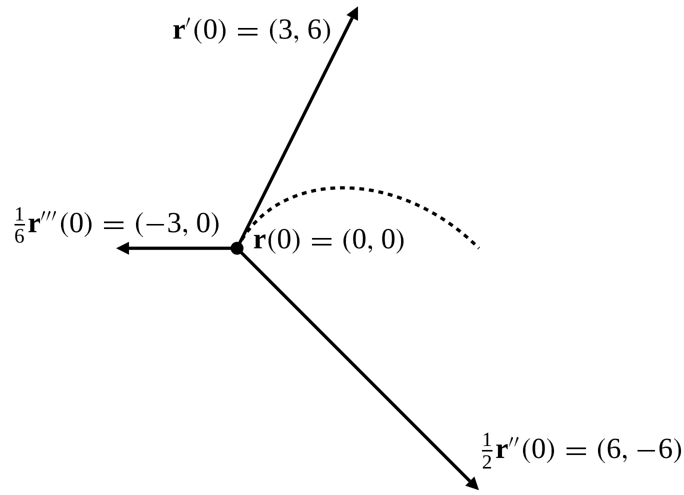


Figure 1.1: The data defining the curve **r**, when we use the power basis. The curve is given by $\mathbf{r}(t) = \sum_k t^k \frac{1}{k!}\mathbf{r}^{(k)}(0)$, $t \in [0, 1]$.

given by

$$\begin{aligned} H_{00}(t) &= 2t^3 - 3t^2 + 1, & H_{10}(t) &= t^3 - 2t^2 + t, \\ H_{01}(t) &= -2t^3 + 3t^2, & H_{11}(t) &= t^3 - t^2. \end{aligned} \tag{1.4}$$

So we write the curve $\mathbf{r}(t)$ as

$$\begin{aligned} \mathbf{r}(t) &= H_{00}(t)\cdot\mathbf{r}(0) + H_{10}(t)\cdot\mathbf{r}'(0) + H_{01}(t)\cdot\mathbf{r}(1) + H_{11}(t)\cdot\mathbf{r}'(1) \\ &= \left(2t^3 - 3t^2 + 1\right)\cdot(0,0) + \left(t^3 - 2t^2 + t\right)\cdot(3,6) \\ &\qquad + \left(-2t^3 + 3t^2\right)\cdot(6,0) + \left(t^3 - t^2\right)\cdot(6,-3). \end{aligned} \tag{1.5}$$

The input to the computer would now be the coordinates with respect to the Hermite basis. The geometric interpretation of these coordinates is that they give the value and the first derivative respectively at the two endpoints of the curve, see Figure 1.2. These coordinates are much more intuitive. We know the value and tangent at both ends so the shape of the curve is more easy to predict. This representation is indeed in use in industry, and cubic polynomial curves in the Hermite representation was introduced by James Ferguson at Boeing, published in 1964 [9], and goes under the name *Ferguson curve*. One drawback is that the generalization to curves of higher degrees gives less intuitive control over the curve.

Finally we have the *Bernstein representation* of a polynomial curve. As the basis for the polynomials of degree 3 we use the *Bernstein polynomials* of degree 3, which are given by

$$B_0^3(t) = (1-t)^3, \quad B_1^3(t) = 3t(1-t)^2, \quad B_2^3(t) = 3t^2(1-t), \quad B_3^3(t) = t^3.$$

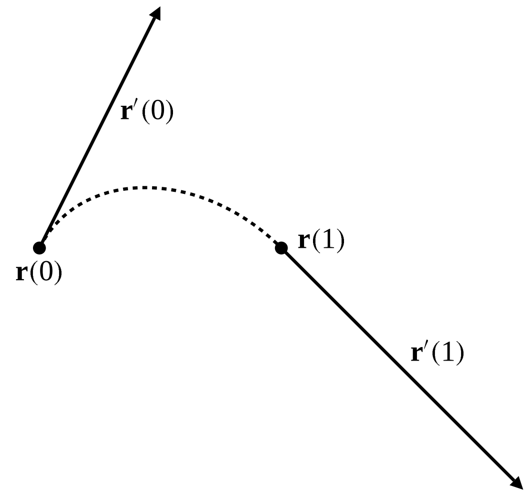


Figure 1.2: The data defining the curve **r**, when we use the Hermite basis. The curve is called a *Ferguson curve* and is given by $\mathbf{r}(t) = \sum_{k,l} H_{kl}(t)\mathbf{r}^{(k)}(l)$, $t \in [0, 1]$.

We write the curve $\mathbf{r}(t)$ as

$$\begin{aligned}\mathbf{r}(t) &= B_0^3(t)\cdot\mathbf{b}_0 + B_1^3(t)\cdot\mathbf{b}_1 + B_2^3(t)\cdot\mathbf{b}_2 + B_3^3(t)\cdot\mathbf{b}_3 \\ &= (1-t)^3\cdot(0,0) + 3t(1-t)^2\cdot(1,2) + 3t^2(1-t)\cdot(4,2) + t^3\cdot(6,0).\end{aligned} \tag{1.6}$$

The geometric interpretation of the *control points* or *Bézier points* $\mathbf{b}_0$, $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$ can be seen in Figure 1.3. The control points form the *control polygon* and the

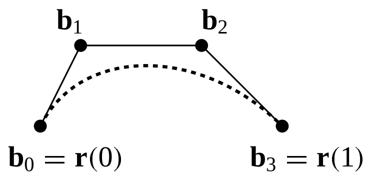


Figure 1.3: The data defining the curve **r**, when we use the Bernstein basis. The curve is called a Bézier curve and is given by $\mathbf{r}(t) = \sum_{k=0}^{n} B_k^n(t)\mathbf{b}_k$.

curve is in some sense a “smoothened” version of the control polygon. Hence the control points provide good intuitive control over the curve.

Polynomial curves in the Bernstein representation was introduced by P. Bézier at Renault in the sixties, and his work was published in 1966 [2, 3], and are called *Bézier curves*. Bézier curves are widely used, e.g., most characters, including the

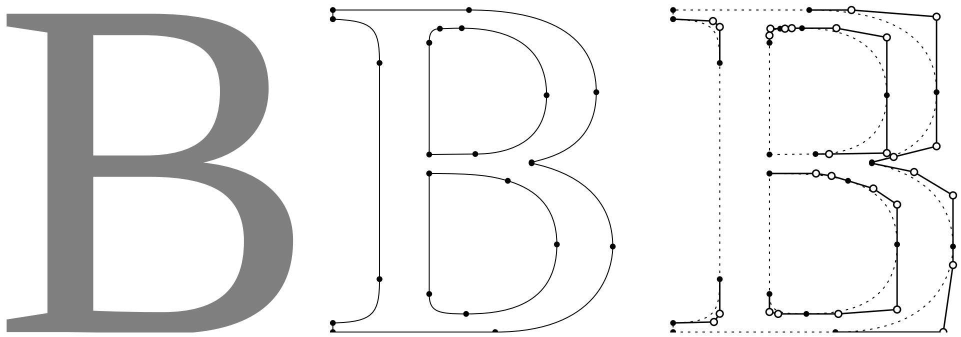


Figure 1.4: The letter ‘B’is described by 8 line segments and 14 cubic Bézier curves. In the middle we have drawn the outline and marked the endpoints of all line segments and Bézier curves with massive circles. At the right we have drawn the control polygons for the Bézier curves and marked the middle control points with open circles.

ones you are reading right now, are described by Bézier curves, see Figure 1.4. Bézier curves are also a standard tool in many programs for drawing, see Figure 1.5.

Problems

1.2.1 Show that the formulas (1.1), (1.2), (1.5), and (1.6) give the same curve.

1.2.2 Determine polynomials $H_{kl}(t)$, $k = 0, 1, 2$ and $l = 0, 1$ of degree at most 5 such that

$$H_{kl}^{(i)}(j) = \delta_{ki}\delta_{lj} = \begin{cases} 1 & \text{if } k = i \text{ and } l = j, \\ 0 & \text{otherwise.} \end{cases}$$

1.2.3 Suppose $\phi_1, \dots, \phi_n$ are linearly independent functions, $\phi_i : [a, b] \to \mathbb{R}$. If $P_1, \dots, P_n$ are points then we define a curve by $\mathbf{r}(t) = \sum_{i=1}^{n} P_i \phi_i(t)$. Show that the construction is affine invariant[3] if and only if $\sum_{i=1}^{n} \phi_i(t) = 1$.

Exercises

1.2.1 Write a program that draws a polynomial curve of degree at most 3 given in:

(a) The power basis.

(b) The Hermite basis.

(c) The Bernstein basis.

[3] i.e., applying an affine transformation to a Bézier curve is the same as applying the transformation to the control points.

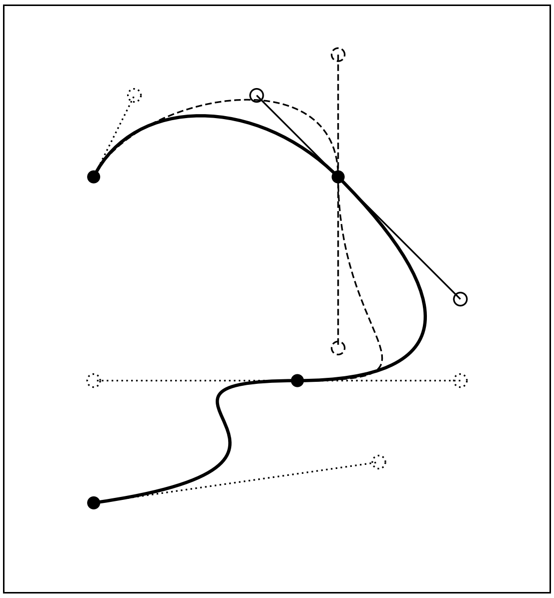


Figure 1.5: Control of a curve in a typical program for drawing. The segments between the solid circles are cubic Bézier curves and the two remaining control points are the open circles. By (1.25) the lines are the tangents to the curves at the endpoints. As can be seen, the program ensures that the three control points round an endpoint is on a straight line. Hereby it is ensured that the two consecutive curves have a common tangent at the common endpoint.

1.2.2 Write a program that draws the Hermite polynomials of degree 3, cf. (1.4) and of degree 5, cf. Problem 1.2.2.

1.3 Bézier curves in the Bernstein representation

Definition 1.1. The *Bernstein polynomials* of degree n are given by

$$B_k^n(t) = \binom{n}{k} t^k (1-t)^{n-k}, \qquad k = 0, \dots, n. \tag{1.7}$$

where

$$\binom{n}{k} = \frac{n!}{(n-k)!\,k!}, \qquad k = 0, \dots, n, \qquad \text{and} \qquad 0! = 1.$$

are the binomial coefficients, see Figure 1.6.

When the Bernstein polynomials are known, we can define a Bézier curve:

Definition 1.2. A *Bézier curve* of degree n with *control points* $\mathbf{b}_0, \ldots, \mathbf{b}_n$ is a curve of the form

$$\mathbf{r}(t) = \sum_{k=0}^{n} B_k^n(t)\mathbf{b}_k, \qquad t \in [0, 1].$$

In Figure 1.7 we have plotted Bézier curves of various degrees.

The properties of a Bézier curve can of course be derived from the properties of the Bernstein polynomials. It is not hard to show that:

$$B_k^n(0) = \begin{cases} 1 & \text{if } k = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{1.8}$$

$$B_k^n(1) = \begin{cases} 1 & \text{if } k = n, \\ 0 & \text{otherwise,} \end{cases} \tag{1.9}$$

$$B_k^n(t) \leq B_k^n\left(\frac{k}{n}\right), \quad t \in [0, 1], \tag{1.10}$$

$$B_l^n\left(\frac{k}{n}\right) < B_k^n\left(\frac{k}{n}\right), \quad l \neq k, \tag{1.11}$$

$$\sum_{k=0}^{n} B_k^n(t) = 1, \tag{1.12}$$

see Figure 1.6. We leave the proof of these properties, and others, as Problem 1.3.1–1.3.6. In the next section we will prove the corresponding properties for Bézier curves without the use of Bernstein polynomials. But given this information about Bernstein polynomials, it is not hard to see that a Bézier curve has the following properties:

- The curve interpolates between the first and the last control point.
- The curve is contained in the convex hull of the control points, see Figure 1.15.
- At the parameter value $\frac{k}{n}$ the control point $\mathbf{b}_k$ has the highest weight. That is the Bézier curve "tries to follow" the control polygon.
- The construction is *affine invariant*, i.e.., Applying an affine transformation to a Bézier curve is the same as applying the transformation to the control points.

Now we could go on and investigate the Bernstein polynomials in greater details and thus obtain information about Bézier curves. We will *not* do this, but instead

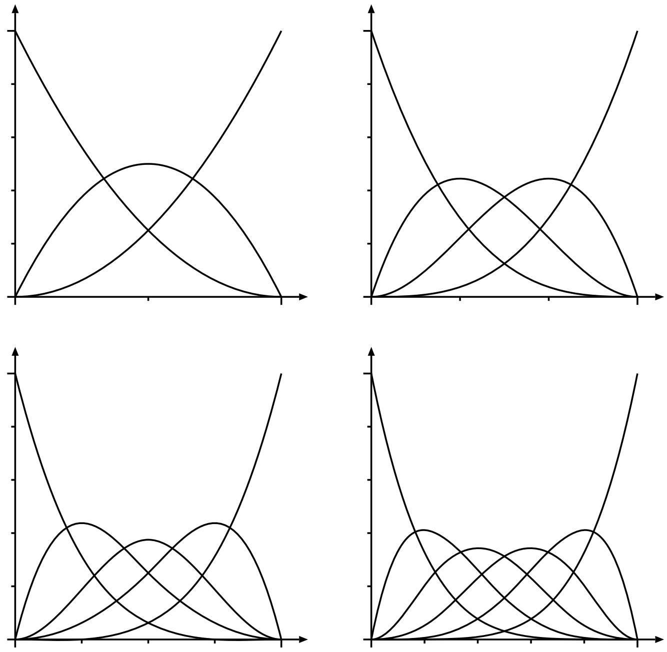


Figure 1.6: The Bernstein polynomials of degree 2, 3, 4, and 5.

give a new definition of a Bézier curve which are more geometric and in my opinion makes the analysis easier. We will in particular prove that Bézier curves have the properties listed above.

Problems

1.3.1 Prove (1.8), (1.9), (1.10), (1.11), and (1.12).

1.3.2 Prove the recurrence relation for the Bernstein polynomials:

$$B_i^n(t) = (1-t)B_i^{n-1}(t) + tB_{i-1}^{n-1}(t), \quad i = 0, \ldots, n. \tag{1.13}$$

1.3.3 Show that the derivative of the Bernstein polynomials is given by

$$\frac{\mathrm{d}}{\mathrm{d}t} B_i^n(t) = n\big(B_{i-1}^{n-1}(t) - B_i^{n-1}(t)\big). \tag{1.14}$$

1.3.4 Prove the identities

$$\int_0^t B_i^n(t)\,\mathrm{d}t = \frac{1}{n+1}\sum_{j=i+1}^{n+1} B_j^{n+1}(t), \qquad \int_0^1 B_i^n(t)\,\mathrm{d}t = \frac{1}{n+1}. \tag{1.15}$$

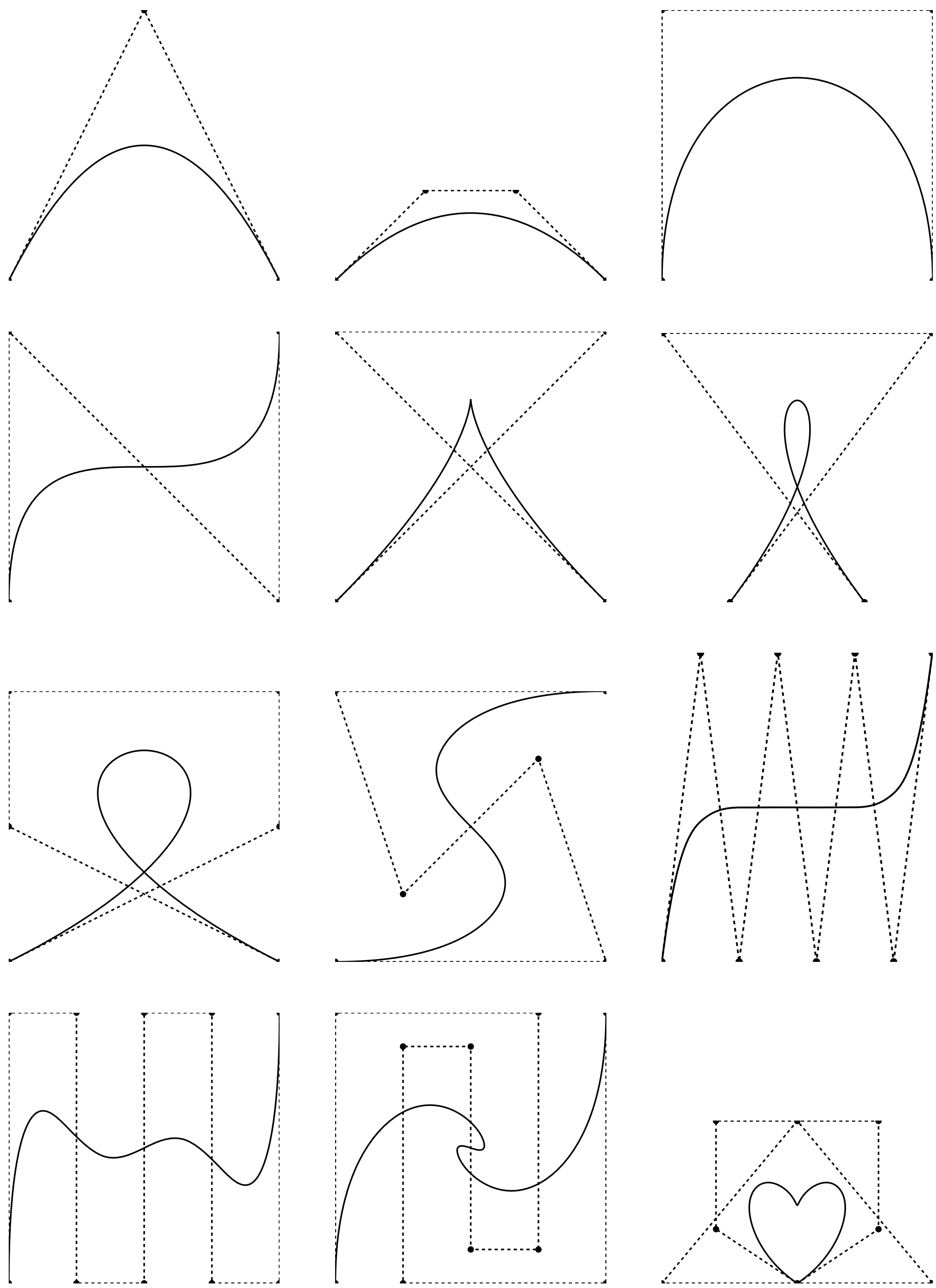


Figure 1.7: Bézier curves. In the first row the curves have degree 2, 3, and 3, in row number two all three curves have degree 3, in row number three the curves have degree 5, 5, and 7, and in the last row all three curves have degee 9.

1.3.5 Prove the identities

$$t = \sum_{i=0}^{n} \frac{i}{n} B_i^n(t), \qquad t^2 = \sum_{i=0}^{n} \frac{i(i-1)}{n(n-1)} B_i^n(t). \tag{1.16}$$

1.3.6 (*Weierstrass' approximation theorem*). Let $f : [0,1] \to \mathbb{R}$ be a continuous function. Show that the sequence of polynomials

$$(B^n f)(t) = \sum_{i=0}^{n} f\left(\frac{i}{n}\right) B_i^n(t)$$

converges uniformly to f in $[0,1]$ as $n \to \infty$.

Exercises

1.3.1 Write a program that use the recurence relation (1.13) to calculates all the Bernstein polynomials of a given degree.

1.3.2 Write a program that uses Exercise 1.3.1 to plot a Bézier curve.

1.4 Bézier curves and de Casteljau's algorithm

At Citroen Paul de Casteljau introduced Bézier curves by repeated linear interpolation, The work was done slightly before P. Bézier's at Renault, but Citroen was more secretive and Paul de Casteljau work was never published. As we shall see this approach is not only geometric in nature, but it even offers simple proofs for the basic properties of Bézier curves. With few exceptions we follow the paper [13], the proof of Theorem 1.14 is taken from [32].

1.4.1 The de Casteljau operator

De Casteljau algorithm is described in Figure 1.8. Formally it can be written as

$$\begin{aligned} \mathbf{b}_k^0(t) &= \mathbf{b}_k, & k &= 0, \ldots, n, \\ \mathbf{b}_k^l(t) &= (1-t)\mathbf{b}_k^{l-1}(t) + t\mathbf{b}_{k+1}^{l-1}(t), & k &= 0, \ldots, n-l, \\ & & l &= 1, \ldots, n. \end{aligned} \tag{1.17}$$

The de Casteljau algorithm (1.17) will be the basis for our development of the Bézier curve theory. As it stands, the algorithm is a bit hard to manipulate, so we will look a litle closer on the algorithm and we will introduce operators or

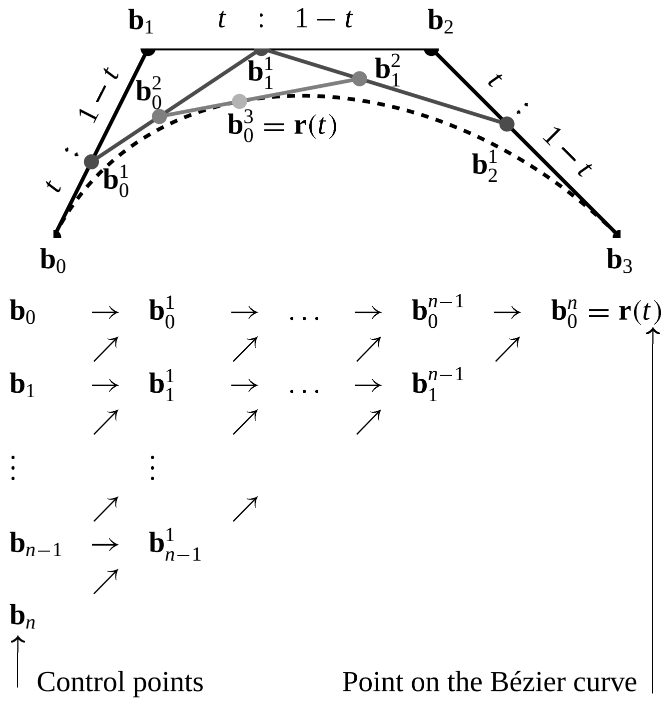


Figure 1.8: De Casteljau's algorithm starts with a polygon with $n+1$ points and in n steps the polygon is reduced to a single point, which is the desired point on the Bézier curve. In each step the points in the new polygon are obtained by dividing the legs of the previous polygon in the proportion $t : 1-t$. In the scheme this corresponds to multiplication of the point from the horizontal arrow by $1-t$ and multiplication of the point from the diagonal arrow by t. The two weighted points are then added and gives the new point.

matrices by which we can describe the algorithm. As we can see in the scheme in Figure 1.8 there are n steps in the algorithm and each step gives one point less. One step (going from one column to the next) is described by:

$$\begin{bmatrix} \mathbf{b}_0^l \\ \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l-1}^l \\ \mathbf{b}_{n-l}^l \end{bmatrix} \mapsto \begin{bmatrix} \mathbf{b}_0^l \\ \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l-1}^l \end{bmatrix}, \begin{bmatrix} \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l-1}^l \\ \mathbf{b}_{n-l}^l \end{bmatrix} = \begin{bmatrix} \mathbf{b}_0^l \\ \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l-1}^l \end{bmatrix}, \begin{bmatrix} \mathbf{b}_1^l \\ \mathbf{b}_2^l \\ \vdots \\ \mathbf{b}_{n-l}^l \end{bmatrix}$$

$$\mapsto (1-t) \begin{bmatrix} \mathbf{b}_0^l \\ \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l-1}^l \end{bmatrix} + t \begin{bmatrix} \mathbf{b}_1^l \\ \mathbf{b}_2^l \\ \vdots \\ \mathbf{b}_{n-l}^l \end{bmatrix} = \begin{bmatrix} \mathbf{b}_0^l \\ \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l-1}^l \end{bmatrix} + t \begin{bmatrix} \mathbf{b}_1^l - \mathbf{b}_0^l \\ \mathbf{b}_2^l - \mathbf{b}_1^l \\ \vdots \\ \mathbf{b}_{n-l}^l - \mathbf{b}_{n-l-1}^l \end{bmatrix}$$

Using matrix notation we can write one step in the algorithm as

$$\begin{bmatrix} \mathbf{b}_0^{l+1} \\ \mathbf{b}_1^{l+1} \\ \vdots \\ \mathbf{b}_{n-l-1}^{l+1} \end{bmatrix} = \begin{bmatrix} 1-t & t & 0 & \dots & 0 \\ 0 & 1-t & t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1-t & t \end{bmatrix} \begin{bmatrix} \mathbf{b}_0^{l} \\ \mathbf{b}_1^{l} \\ \vdots \\ \mathbf{b}_{n-l}^{l} \end{bmatrix}$$

$$= \left((1-t) \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} + t \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix} \right) \begin{bmatrix} \mathbf{b}_0^{l} \\ \mathbf{b}_1^{l} \\ \vdots \\ \mathbf{b}_{n-l}^{l} \end{bmatrix}$$

$$= \left(\begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix} + t \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix} \right) \begin{bmatrix} \mathbf{b}_0^{l} \\ \mathbf{b}_1^{l} \\ \vdots \\ \mathbf{b}_{n-l}^{l} \end{bmatrix}.$$

We now give names to the two matrices in the middle line:

$$R = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 & 0 \end{bmatrix}, \qquad L = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & 1 \end{bmatrix},$$

or equivalently we define two basic operators R and L, which act on a finite sequence of points producing a sequence with one point less. They simply remove the last, respectively the first point, from the sequence:

$$R\colon \big(\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_k\big) \mapsto \big(\mathbf{b}_0, \dots, \mathbf{b}_{k-1}\big), \tag{1.18}$$

$$L\colon \big(\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_k\big) \mapsto \big(\mathbf{b}_1, \dots, \mathbf{b}_k\big). \tag{1.19}$$

From R and L we define two new operators. The *forward difference* operator:

$$\Delta = L - R, \tag{1.20}$$

and for a $t \in \mathbb{R}$ the *de Casteljau* operator:

$$C(t) = (1-t)R + tL = R + t\Delta. \tag{1.21}$$

In matrix notation we have

$$\Delta = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ 0 & -1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -1 & 1 \end{bmatrix}, \qquad C(t) = \begin{bmatrix} 1-t & t & 0 & \dots & 0 \\ 0 & 1-t & t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1-t & t \end{bmatrix}.$$

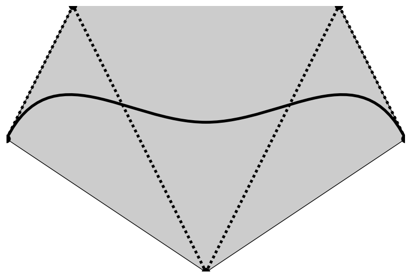


Figure 1.9: A Bézier curve is contained in the convex hull of the control points.

As $C(t)$ describes one step in de Casteljau's algorithm, and there are n steps all in all, we have the following definition:

Definition 1.3. A *Bézier curve* of degree n with *control points* $\mathbf{b}_0, \ldots, \mathbf{b}_n$ is given by

$$\mathbf{r}(t) = C(t)^n \big(\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_n\big), \qquad t \in [0, 1].$$

The basic operators R and L are obviously affinely invariant, so the $C(t)$ is affinely invariant too. This shows the *affine invariance* of Bézier curves. We observe that a point on a Bézier curve is found by performing repeated interpolation between the control points so it is clear that such a point is a convex combination of the control points, so we immediately have the *convex hull property*, see Figure 1.9.

Theorem 1.4. *If* $\mathbf{r}(t)$ *is a Bézier curve with control points* $\mathbf{b}_0, \ldots, \mathbf{b}_n$ *then*

$$\mathbf{r}(t) \in \textit{convex hull of } \big\{\mathbf{b}_0, \ldots, \mathbf{b}_n\big\}.$$

The following *fundamental property* is crucial for our analysis of Bézier curves.

Theorem 1.5. *The basic operators "commute":*

$$LR = RL.$$

We immediately have the following

Corollary 1.6. *The operators* L, R, Δ, *and* $C(t)$ *"commute".*

We have put quotation marks around the word commute, because the equation in Theorem 1.5 should really read

$$L_{k-1} R_k = R_{k-1} L_k, \tag{1.22}$$

where the subscript indicates the length of the sequences on which the operators act, i.e. the R's on the two sides of the equation are different and similarly with the L's. On the other hand, it will always be clear what the length of the sequence is. In order to keep the notation simple we won't decorate the operators and we will use the word commute without any further comments.

We better show that the two definitions of a Bézier curve agree:

Theorem 1.7. *If $\mathbf{r}(t)$ is a Bézier curve with control points $\mathbf{b}_0, \ldots, \mathbf{b}_n$ given by Definition 1.3, then*

$$\mathbf{r}(t) = \sum_{k=0}^{n} B_k^n(t)\mathbf{b}_k,$$

in accordance with Definition 1.2.

Proof. The control point with index k can be found by

$$\mathbf{b}_k = L^k R^{n-k}\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big). \tag{1.23}$$

As $RL = LR$ we can use the binomial formula and obtain

$$\begin{aligned} C(t)^n &= \big(tL + (1-t)R\big)^n \\ &= \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} L^k R^{n-k}. \end{aligned}$$

Thus

$$\begin{aligned} \mathbf{r}(t) &= C(t)^n\big(\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_n\big) \\ &= \sum_{k=0}^{n} \binom{n}{k} t^k (1-t)^{n-k} L^k R^{n-k}\big(\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_n\big) \\ &= \sum_{k=0}^{n} B_k^n(t)\mathbf{b}_k. \end{aligned}$$

□

We can immediately generalize (1.23) to the following

Lemma 1.8. *The intermediate points in de Casteljau's algorithm is given by*

$$\mathbf{b}_k^l = C^l(t) R^{n-l-k} L^k\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big).$$

1.4.2 Differentiation of a Bézier curve

Theorem 1.9. *The derivative of a Bézier curve* $\mathbf{r}(t)$ *of degree n with control points* $\mathbf{b}_0, \ldots, \mathbf{b}_n$ *can be written as*

$$\frac{1}{n}\mathbf{r}'(t) = \mathbf{b}_1^{n-1}(t) - \mathbf{b}_0^{n-1}(t),$$

where $\mathbf{b}_1^{n-1}$ *and* $\mathbf{b}_0^{n-1}$ *are the two second last intermediate points in de Casteljau's algorithm. Alternatively: the derivative* $\frac{1}{n}\mathbf{r}'(t)$ *is a Bézier curve of degree* $n-1$ *with control points*

$$\Delta\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big) = \big(\mathbf{b}_1 - \mathbf{b}_0, \ldots, \mathbf{b}_n - \mathbf{b}_{n-1}\big),$$

see Figure 1.10.

Proof. The de Casteljau operator (or matrix) is a function of t, $C(t) = R + t\Delta$, and the derivative is of course

$$\frac{d}{dt}C(t) = \Delta.$$

Thus, the derivative of $C(t)^n$ is

$$\begin{aligned}\frac{d}{dt}\big(C(t)^n\big) &= \sum_{k=1}^{n} C(t)^{k-1}\left(\frac{d}{dt}C(t)\right)C(t)^{n-k} \\ &= \sum_{k=1}^{n} C(t)^{k-1}\Delta C(t)^{n-k} \\ &= nC(t)^{n-1}\Delta = n\Delta C(t)^{n-1}.\end{aligned} \tag{1.24}$$

If we apply the two expressions in the last line to the sequence $\mathbf{b}_0, \ldots, \mathbf{b}_n$ we obtain the two descriptions of the derivative. □

At the endpoints we have in particular that

$$\begin{aligned}\mathbf{r}'(0) &= n\big(\mathbf{b}_1 - \mathbf{b}_0\big), \\ \mathbf{r}'(1) &= n\big(\mathbf{b}_n - \mathbf{b}_{n-1}\big),\end{aligned} \tag{1.25}$$

i.e., the tangent at an endpoint is the line through the endpoint and the neighboring control point. It is equally easy to find the higher order derivatives:

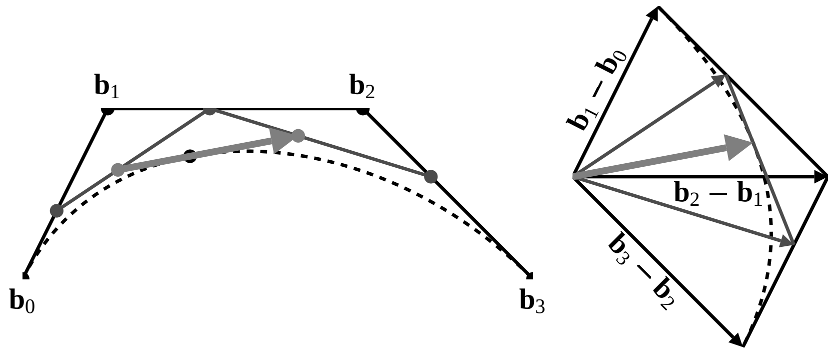


Figure 1.10: To the left: The derivative as the difference of the penultimate points in de Casteljau's algorithm. To the right: The derivative as a Bézier curve with control points equal to the differences of the original control points.

Theorem 1.10. *The k'th derivative of a Bézier curve* $\mathbf{r}(t)$ *of degree n with control points* $\mathbf{b}_0, \ldots, \mathbf{b}_n$ *is given by*

$$\begin{aligned}\mathbf{r}^{(k)}(t) &= \frac{n!}{(n-k)!}\Delta^k C(t)^{n-k}\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big) \\ &= \frac{n!}{(n-k)!}C(t)^{n-k}\Delta^k\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big).\end{aligned}$$

The first expression says that the k'th derivative can be found by performing $n-k$ steps in de Casteljau's algorithm and then perform k times repeated differences. The second expression says that the k'th derivative is a Bézier curve of degree $n-k$ and its control points is found by performing k times repeated differences in the original control polygon.

1.4.3 Linear precision and degree elevation

If the points $\mathbf{b}_1, \ldots, \mathbf{b}_{n-1}$ all lie on the line segment $\mathbf{b}_0\mathbf{b}_n$, then the convex hull property implies that the image of the curve is the line segment, but the parametrization needs not be the usual. The curve might oscillate back and forth, see e.g., the y-coordinate of the first two curves in the last row in Figure 1.7. The next theorem tells us that the control points should be equally spaced on the line segment in order to get the usual parametrization. This is called *linear precision*, see Figure 1.11.

Theorem 1.11. *Let* $\mathbf{b}_0 \ldots, \mathbf{b}_n$ *be the control points for a Bézier curve* $\mathbf{r}(t)$. *Then*

$$\mathbf{r}(t) = (1-t)\mathbf{b}_0 + t\mathbf{b}_n$$

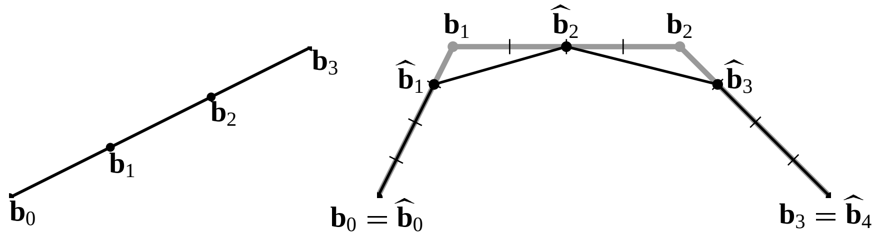


Figure 1.11: To the left we have linear precision: The control points should be equally spaced in order to get the usual parametrized line segment.
To the right we have degree elevation: If the degree of the original curve is n then each leg of the control polygon is divided into n equally sized segments. Besides the first and the last control point, we pick one of the points on each leg of the polygon. The last point on the last leg, the penultimate point on the second leg, and so on, until the first point on the last leg is chosen. This gives $n+1$ points all in all, These points are exactly the control points for the same curve considered as a Bézier curve of one degree more.

if and only if

$$\mathbf{b}_k = \frac{n-k}{n}\mathbf{b}_0 + \frac{k}{n}\mathbf{b}_n, \quad k = 1, \ldots, n-1.$$

Proof. As $\mathbf{r}(0) = \mathbf{b}_0$ we have

$$\mathbf{r}(t) = (1-t)\mathbf{b}_0 + t\mathbf{b}_n \iff \mathbf{r}'(t) = \overrightarrow{\mathbf{b}_0\mathbf{b}_n} \quad \text{for all } t.$$

As $\mathbf{r}'(t)$ is a Bézier curve with control points $n\overrightarrow{\mathbf{b}_0\mathbf{b}_1}, \ldots, n\overrightarrow{\mathbf{b}_{n-1}\mathbf{b}_n}$, this happens if and only if

$$n\overrightarrow{\mathbf{b}_{k-1}\mathbf{b}_k} = \overrightarrow{\mathbf{b}_0\mathbf{b}_n} \quad \text{for } k = 1, \ldots n.$$

Finally this is obviously equivalent to

$$\mathbf{b}_k = \frac{n-k}{n}\mathbf{b}_0 + \frac{k}{n}\mathbf{b}_n, \quad k = 1, \ldots, n-1.$$

□

The above is an example of a Bézier curve (in this case of degree 1) which can be written as a curve of higher degree. The following *degree elevation theorem* tells us how to raise the degree of an arbitrary Bézier curve, see Figure 1.11.

Theorem 1.12. *Let* $\mathbf{b}_0 \ldots, \mathbf{b}_n$ *be the control points for a Bézier curve* $\mathbf{r}(t)$ *of degree* n*. Considered as a curve of degree* $n+1$*,* $\mathbf{r}(t)$ *has control points* $\widehat{\mathbf{b}}_0 \ldots, \widehat{\mathbf{b}}_{n+1}$*, where*

$$\begin{aligned}\widehat{\mathbf{b}}_0 &= \mathbf{b}_0,\\ \widehat{\mathbf{b}}_k &= \frac{n+1-k}{n+1}\mathbf{b}_k + \frac{k}{n+1}\mathbf{b}_{k-1}, \quad k = 1, \ldots, n\\ \widehat{\mathbf{b}}_{n+1} &= \mathbf{b}_n.\end{aligned}$$

Proof. Even though it is possible to prove this theorem using operators, see [13], this is the one case where it is easier to use the Bernstein representation. As

$$\begin{aligned}(1-t)B_i^n(t) &= \frac{n!}{(n-i)!\,i!}(1-t)^{n+1-i}t^i\\ &= \frac{n+1-i}{n+1}\frac{(n+1)!}{(n+1-i)!\,i!}(1-t)^{n+1-i}t^i\\ &= \frac{n+1-i}{n+1}B_i^{n+1}(t),\end{aligned}$$

and

$$\begin{aligned}tB_i^n(t) &= \frac{n!}{(n-i)!\,i!}(1-t)^{n-i}t^{i+1}\\ &= \frac{i+1}{n+1}\frac{(n+1)!}{\big((n+1)-(i+1)\big)!\,(i+1)!}(1-t)^{(n+1)-(i+1)}t^{i+1}\\ &= \frac{i+1}{n+1}B_{i+1}^{n+1}(t),\end{aligned}$$

we get

$$\begin{aligned}\mathbf{r}(t) &= \big((1-t)+t\big)\mathbf{r}(t) = \sum_{i=0}^{n}\big((1-t)B_i^n(t)+tB_i^n(t)\big)\mathbf{b}_i\\ &= \sum_{i=0}^{n}\left(\frac{n+1-i}{n+1}B_i^{n+1}(t)+\frac{i+1}{n+1}B_{i+1}^{n+1}(t)\right)\mathbf{b}_i\\ &= B_0^{n+1}(t)\mathbf{b}_0+\sum_{i=1}^{n}\frac{n+1-i}{n+1}B_i^{n+1}(t)\mathbf{b}_i\\ &\qquad +\sum_{i=0}^{n-1}\frac{i+1}{n+1}B_{i+1}^{n+1}(t)\mathbf{b}_i + B_{n+1}^{n+1}(t)\mathbf{b}_n\\ &= B_0^{n+1}(t)\mathbf{b}_0+\sum_{i=1}^{n}B_i^{n+1}(t)\left(\frac{n+1-i}{n+1}\mathbf{b}_i+\frac{i}{n+1}\mathbf{b}_{i-1}\right)+B_{n+1}^{n+1}(t)\mathbf{b}_n\\ &= \sum_{i=0}^{n+1}B_i^{n+1}(t)\widehat{\mathbf{b}}_i\end{aligned}$$

as claimed. □

1.4.4 Subdivision and the variation diminishing property

For the proof of the next theorem we need the following lemma, which also have some interest in its own right.

Lemma 1.13. *De Casteljau's operator $C(t)$ is invariant under an affine change of parameter, that is for $a, b, t \in \mathbb{R}$:*

$$C\big((1-t)a + tb\big) = (1-t)C(a) + tC(b).$$

Proof. Using the definition of $C(t)$ we immediatly get:

$$\begin{aligned} C\big((1-t)a + tb\big) &= \big(1 - \big((1-t)a + tb\big)\big)R + \big((1-t)a + tb\big)L \\ &= (1 - a + ta - tb)R + (a - ta + tb)L, \end{aligned}$$

and

$$\begin{aligned} (1-t)C(a) + tC(b) &= (1-t)\big((1-a)R + aL\big) + t\big((1-b)R + bL\big) \\ &= (1 - a - t + ta)R + (a - ta)L + (t - tb)R + tbL \\ &= (1 - a + ta - tb)R + (a - ta + tb)L. \end{aligned}$$

The two expressions are equal and we have proven the lemma. □

If $\mathbf{r}(t)$ is a Bézier curve of degree n, then the curve $t \mapsto \mathbf{r}\big((1-t)a+tb\big)$, $t \in [0, 1]$, is obviously also a polynomial curve of degree n, and it is a reparametrization of the restriction of $\mathbf{r}$ to the interval $[a, b]$. As it is polynomial it can be considered as a Bézier curve and its control points can be found by de Casteljau's algorithm:

Theorem 1.14. *If $\mathbf{r}(t)$ is a Bézier curve with control points $\mathbf{b}_0, \dots, \mathbf{b}_n$, then*

$$t \mapsto \mathbf{r}\big((1-t)a + tb\big)$$

is a Bézier curve with control points

$$\widehat{\mathbf{b}}_k = C(a)^{n-k}C(b)^k\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big), \quad k = 0, \dots, n.$$

Proof. (From [32]). We have to prove that

$$C\big((1-t)a + tb\big)^n\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big) = C(t)^n\big(\widehat{\mathbf{b}}_0, \dots, \widehat{\mathbf{b}}_n\big)$$

According to lemma 1.13 and the binomial formula we have

$$\begin{aligned} C\big((1-t)a + tb\big)^n\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big) &= \big((1-t)C(a) + tC(b)\big)^n\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big) \\ &= \sum_{i=0}^{n} \binom{n}{i}\big((1-t)C(a)\big)^{n-i}\big(tC(b)\big)^i\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big) \\ &= \sum_{i=0}^{n} \binom{n}{i}(1-t)^{n-i}t^iC(a)^{n-i}C(b)^i\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big) \\ &= \sum_{i=0}^{n} \binom{n}{i}(1-t)^{n-i}t^i\widehat{\mathbf{b}}_i = C(t)^n\big(\widehat{\mathbf{b}}_0, \dots, \widehat{\mathbf{b}}_n\big), \end{aligned}$$

and the proof is complete. □

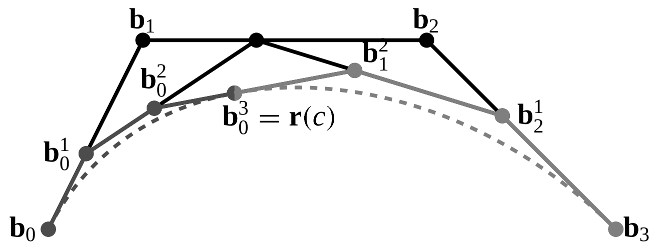


Figure 1.12: Subdivision of a Bézier curve at the parameter value c. The first part of the curve $t \mapsto \mathbf{r}(ct)$ has control points $\mathbf{b}_0$, $\mathbf{b}_0^1(c)$, $\mathbf{b}_0^2(c)$, and $\mathbf{b}_0^3(c)$, the last part of the curve $t \mapsto \mathbf{r}\big(c + (1-c)t\big)$ has control points $\mathbf{b}_0^3(c)$, $\mathbf{b}_1^2(c)$, $\mathbf{b}_2^1(c)$, and $\mathbf{b}_3$.

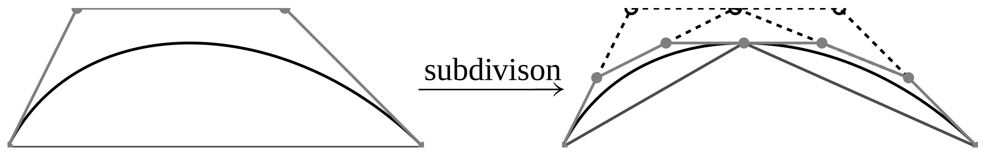


Figure 1.13: Drawing a Bézier curve by drawing the subdivided control polygons.

This process is called *subdivision*, and the two cases $[a, b] = [0, c]$ and $[a, b] = [c, 1]$ are particular simple and are illustrated in Figure 1.12.

Subdivision forms the basis for a powerful method by which we can treat Bézier curves, see [12]. Suppose we have some geometric quantity we want to determine, you may think of just the graph, i.e, the image of the curve, but it could be the length of the curve, the total curvature of the curve, the total curvature variation of the curve, etc. Suppose this quantity is easy to determine for the control polygon, then we can determine the quantity for the curve by the following principle:

- We determine the quantity for the control polygon.
- We estimate the error. If it is small, then we use this quantity.
- Otherwise we subdivide the curve and start over again with each half.

In Figure 1.13 this method is illustrated. The two control polygons on the right gives a much better approximation to the curve than the control polygon on the left.

This method works in a lot of instances and also imply the *variation diminishing property*, see Figure 1.15. Consider once more the scheme in Figure 1.8 on page 11. If we keep the points in the top horizontal row and the points in the lower diagonal row, then each step in de Casteljau's algorithm increases the number of points by one. After k steps in the algorithm we have the polygon consisting of the points:

$$\mathbf{b}_0, \mathbf{b}_0^1, \ldots, \mathbf{b}_0^l, \ldots, \mathbf{b}_{n-l}^l, \ldots, \mathbf{b}_{n-1}^1, \mathbf{b}_n.$$

After n steps we have the control polygons for both halfs of the curve. With this interpretation of de Casteljau's algorithm we have the following lemma:

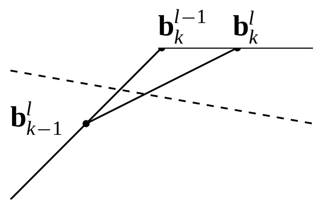


Figure 1.14: De Casteljau's algorithm can not increase the number of intersections.

Lemma 1.15. *A step in de Casteljau's algorithm can not increase the number of intersections with a hyperplane.*

Proof. Consider Figure 1.14, it is obvious that if the line segment $\mathbf{b}_{k-1}^l\mathbf{b}_k^l$ intersect a hyperplane, then the hyperplane must also intersect either the line segment $\mathbf{b}_{k-1}^l\mathbf{b}_k^{l-1}$ or the line segment $\mathbf{b}_k^{l-1}\mathbf{b}_k^l$. I.e., for each intersection in the new polygon exists a corresponding intersection in the old. □

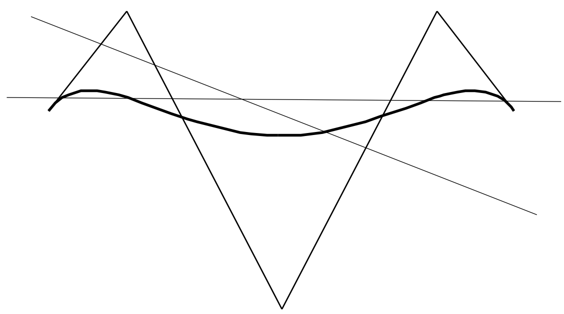


Figure 1.15: The variation diminishing property: the lines intersect the control polygon at least as many times as they intersect the Bézier curve.

Theorem 1.16. *Let $\mathbf{r}(t)$ be a Bézier curve with control polygon $\mathbf{P}$. Let furthermore α be a hyperplane, then*

$$\#\big(\alpha \cap \mathbf{r}\big) \leq \#\big(\alpha \cap \mathbf{P}\big),$$

i.e., the number of intersections between the curve and the hyperplane is less than the number of intersections between the control polygon and the hyperplane.

Proof. Let us consider a hyperplane which intersects a Bézier curve in a number of points. We now subdivide the curve in all those points, and hereby obtain a polygon containing all these points. The hyperplane then intersects this polygon at least as many times as it intersects the curve. According to Lemma 1.15 it must intersect the original control polygon at least that many times. □

Problems

1.4.1 Prove (1.23).

1.4.2 Prove Lemma 1.8.

Exercises

1.4.1 Write a program that determines a point on a Bézier curve using de Casteljau's algorithm.

1.4.2 Write a program that determines a point and the derivatives to order k on a Bézier curve using de Casteljau's algorithm.

1.4.3 Write a program that finds the derivative of a Bézier curve.

1.4.4 Write a program that degree elevate a Bézier curve.

1.4.5 Write a program that subdivides a Bézier curve.

1.5 The polar form of a polynomial curve

Recall that, for each quadratic form $F : V \to \mathbb{R}$ on a linear space V, there is a unique symmetric, bilinear form $f : V \times V \to \mathbb{R}$ that satisfies the identity $F(v) = f(v, v)$. The polar form is to a Bézier curve what the symmetric, bilinear form is to a quadratic form.

Definition 1.17. The *polar form* of a Bézier curve $F(t) = C(t)^n\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big)$ is the map $f(t_1, \dots, t_n) = C(t_1) \dots C(t_n)\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big)$.

Remark 1.18. We also say that f is the *polarization* of F. Lately the word *blossom* has been used instead of the polar form.

Remark 1.19. As any polynomial curve can be written in the Bernstein representation, i.e., as a Bézier curve, we can determine a polar form of a polynomial. It is, on the other hand, possible to write the polynomial as a Bézier curve of arbitrary large degree and if the (formal) degree of the Bernstein representation is n, then the corresponding polar form is called the *n-polar form*.

Theorem 1.20. *The n-polar form $f(t_1, \ldots, t_n)$ of a polynomial $F(t)$ of degree at most n satisfies the following properties:*

1. *f is symmetric.*
2. *f is n-affine (i.e., f is affine with respect to each variable).*
3. *f restricted to the diagonal of $\mathbb{R}^n$ gives F, i.e., $f(t, \ldots, t) = F(t)$.*

Conversely, if f satisfies 1–3, then f is the polar form of F.

Proof. First assume that f is the polar form of F. As $C(t_i)$ and $C(t_j)$ commute, f is symmetric. By Lemma 1.13 f is n-affine, and we clearly have $F(t) = f(t, \ldots, t)$. Now assume f satisfies 1–3, by Lemma 1.21 below we can write

$$f(t_1, \ldots, t_n) = C(t_1) \ldots C(t_n)\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big).$$

By 3 we have

$$F(t) = f(t, \ldots, t) = C(t)^n\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big).$$

This shows that $\mathbf{b}_0, \ldots, \mathbf{b}_n$ are the control points for $F(t)$ and hence that f is the polar form of F. □

In the proof above we have used the following lemma, which also has independent interest.

Lemma 1.21. *If $f : \mathbb{R}^n \to \mathbb{R}$ is symmetric and multi affine, i.e., satisfies 1 and 2 in Theorem 1.20 then*

$$\begin{aligned} f(t_1, \ldots, t_n) &= C(t_1) \ldots C(t_n)\big(\mathbf{b}_0, \ldots, \mathbf{b}_n\big) \\ \textit{where } \mathbf{b}_k &= f(\underbrace{0, \ldots, 0}_{n-k}, \underbrace{1, \ldots, 1}_{k}). \end{aligned} \tag{1.26}$$

Proof. We use induction on n. If $n = 1$, then by the affine invariance we have

$$f(t_1) = f\big((1 - t_1)0 + t_1 1\big) = (1 - t_1) f(0) + t_1 f(1) = C(t_1)\big(f(0), f(1)\big)$$

Now assume the lemma holds for an $n \geq 1$. Using the affine invariance again we have

$$f(t_1, \ldots, t_{n+1}) = (1 - t_{n+1}) f(t_1, \ldots, t_n, 0) + t_{n+1} f(t_1, \ldots, t_n, 1)$$

Letting $t_{n+1} = 0, 1$ respectively shows that the two functions

$$\begin{aligned} f_0(t_1, \ldots, t_n) &= f(t_1, \ldots, t_n, 0) \\ f_1(t_1, \ldots, t_n) &= f(t_1, \ldots, t_n, 1) \end{aligned}$$

are symmetric and n-affine. By the induction hypothesis we can write

$$f_i(t_1, \dots, t_n) = C(t_1) \dots C(t_n)\big(\mathbf{b}_0^i, \dots, \mathbf{b}_n^i\big).$$

Using the symmetry we have

$$\begin{aligned}
\mathbf{b}_k^0 = f_0(\underbrace{0, \dots, 0}_{n-k}, \underbrace{1, \dots, 1}_{k}) &= f(\underbrace{0, \dots, 0}_{n-k}, \underbrace{1, \dots, 1}_{k}, 0) \\
&= f(\underbrace{0, \dots, 0}_{(n+1)-k}, \underbrace{1, \dots, 1}_{k}) = \mathbf{b}_k,
\end{aligned}$$

$$\begin{aligned}
\mathbf{b}_k^1 = f_1(\underbrace{0, \dots, 0}_{n-k}, \underbrace{1, \dots, 1}_{k}) &= f(\underbrace{0, \dots, 0}_{n-k}, \underbrace{1, \dots, 1}_{k}, 1) \\
&= f(\underbrace{0, \dots, 0}_{(n+1)-(k+1)}, \underbrace{1, \dots, 1}_{k+1}) = \mathbf{b}_{k+1}.
\end{aligned}$$

From this we have

$$\begin{aligned}
f(t_1, \dots, t_{n+1}) &= (1 - t_{n+1})C(t_1) \dots C(t_n)\big(\mathbf{b}_0^0, \dots, \mathbf{b}_n^0\big) \\
&\quad + t_{n+1}C(t_1) \dots C(t_n)\big(\mathbf{b}_0^1, \dots, \mathbf{b}_n^1\big) \\
&= (1 - t_{n+1})C(t_1) \dots C(t_n)\big(\mathbf{b}_0, \dots, \mathbf{b}_n\big) \\
&\quad + t_{n+1}C(t_1) \dots C(t_n)\big(\mathbf{b}_1, \dots, \mathbf{b}_{n+1}\big) \\
&= (1 - t_{n+1})C(t_1) \dots C(t_n)R\big(\mathbf{b}_0, \dots, \mathbf{b}_{n+1}\big) \\
&\quad + t_{n+1}C(t_1) \dots C(t_n)L\big(\mathbf{b}_0, \dots, \mathbf{b}_{n+1}\big) \\
&= \big((1 - t_{n+1})R + t_{n+1}L\big)C(t_1) \dots C(t_n)\big(\mathbf{b}_0, \dots, \mathbf{b}_{n+1}\big),
\end{aligned}$$

as should be proved. □

The properties of Bézier curves that we found in Section 1.4 can now be reformulated in the language of polar forms:

Theorem 1.22. *Let f be the polar form of a Bézier curve $F(t)$ with control points $\mathbf{b}_0, \dots, \mathbf{b}_n$. The intermediate points in de Casteljau's algorithm are give by*

$$\mathbf{b}_k^\ell = f(\underbrace{t, \dots, t}_{\ell}, \underbrace{0, \dots, 0}_{n-\ell-k}, \underbrace{1, \dots, 1}_{k}). \tag{1.27}$$

The derivative is given by

$$F'(t) = \frac{n}{b-a}\big(f(t, \dots, t, b) - f(t, \dots, t, a)\big) \tag{1.28}$$

and the higher order derivatives by

$$F^{(k)}(t) = \frac{n!}{(b-a)^k (n-k)!} \sum_{i=0}^{k} (-1)^i \binom{k}{i} f(\underbrace{t, \dots, t}_{n-k}, \underbrace{b, \dots, b}_{k-i}, \underbrace{a, \dots, a}_{i}) \tag{1.29}$$

The $(n+1)$*-polar form of* F *is given by*

$$f^*(t_1, \dots, t_{n+1}) = \frac{1}{n+1} \sum_{k=1}^{n+1} f(t_1, \dots, t_{k-1}, t_{k+1}, \dots, t_n). \tag{1.30}$$

Proof. As $C(0) = R$ and $C(1) = L$, Lemma 1.8 implies (1.27). For the derivative we notice that $\Delta = \frac{C(b)-C(a)}{b-a}$ and hence that

$$\Delta^k = \sum_{i=0}^{k} \frac{(-1)^i}{(b-a)^k} \binom{k}{i} C(b)^{k-i} C(a)^i.$$

Now Theorem 1.9 implies (1.28) and Theorem 1.10 implies (1.29). If we define f^* by (1.30) then f^* is symmetric, $(n+1)$-affine and the restriction to the diagonal is clearly F. By Theorem 1.20 f^* is the $(n+1)$-polar form of F. □

It is well known that a polynomial of degree n is determined uniquely by its values in $n+1$ different points. Therefore the polar form is determined by its values in $n+1$ different points on the diagonal. We also know that the polar form is determined by the $n+1$ values $f(\underbrace{0, \dots, 0}_{k}, \underbrace{1, \dots, 1}_{n-k})$, $k = 0, \dots, n$. We will need yet another set of $n+1$ points that determine the polar form uniquely.

Theorem 1.23. *Let* $f : \mathbb{R}^n \to \mathbb{R}$ *be symmetric and n-affine, and let* $s_1 \le \cdots \le s_{2n}$ *be a sequence of points with* $s_n < s_{n+1}$*. Then* f *is determined uniquely by the values* $f(s_{i+1}, \dots, s_{i+n})$*,* $i = 0, \dots, n$*.*

Proof. Let $t_1, \dots, t_n$ be given. If $1 \le k \le n$ and $k \le i \le n$, then

$$\begin{aligned} t_k &= \frac{(s_{i+n-k+1} - t_k)s_i + (t_k - s_i)s_{i+n-k+1}}{s_{i+n-k+1} - s_i} \\ &= \frac{(s_{i+n-k+1} - s_i - t_k + s_i)s_i + (t_k - s_i)s_{i+n-k+1}}{s_{i+n-k+1} - s_i} \\ &= \left(1 - \frac{t_k - s_i}{s_{i+n-k+1} - s_i}\right) s_i + \frac{t_k - s_i}{s_{i+n-k+1} - s_i} s_{i+n-k+1} \end{aligned}$$

If we put $\alpha_i^k = \dfrac{t_k - s_i}{s_{i+n-k+1} - s_i}$, then $t_k = \big(1-\alpha_i^k\big)s_i + \alpha_i^k s_{i+n-k+1}$ so symmetry and affine invariance yields

$$\begin{aligned} f(t_1, \dots, t_k, s_{i+1}, \dots, s_{i+n-k}) \\ = (t_1, \dots, t_{k-1}, (1-\alpha_i^k)s_i + \alpha_i^k s_{i+n-k+1}, s_{i+1}, \dots, s_{i+n-k}) \\ = (1-\alpha_i^k)\, f(t_1, \dots, t_{k-1}, s_i, \dots, s_{i+n-k}) \\ + \alpha_i^k\, f(t_1, \dots, t_{k-1}, s_{i+1}, \dots, s_{i+n-k+1}). \end{aligned}$$

When k runs from 1 to n we start with the values $f(s_{i+1}, \dots, s_{i+n})$ and we end with $f(t_1, \dots, t_n)$, see Figure 1.16. □

If we put $t_k = t$ in the recursive algorithm described in the proof above then we get the *de Boor's algorithm* which evaluate the polynomial from the values $f(s_{1+i}, \dots, s_{n+i})$ of the polar form.

However it is not every set of $n+1$ values that determines an n-polar form, cf. Problem 1.5.2.

Theorem 1.24. *Let F, G be two polynomials of degree at most n and let f, g be the corresponding n-polar forms. Furthermore let $t \in \mathbb{R}$ and $r \in \mathbb{N}_0$ be given. Then*

$$F^{(k)}(t) = G^{(k)}(t), \quad k = 0, \dots, r,$$

if and only if

$$f(t_1, \dots, t_r, t, \dots, t) = g(t_1, \dots, t_r, t, \dots, t) \quad \textit{for all } t_1, \dots, t_r \in \mathbb{R}.$$

Proof. If we put $(t_1, \dots, t_r) = (\underbrace{t, \dots, t}_{r-k}, \underbrace{b, \dots, b}_{k-i}, \underbrace{a, \dots, a}_{i})$ the 'if' part follows immediately from (1.29). For the other implication we first show by induction on k that

$$f(\underbrace{t, \dots, t}_{n-k}, \underbrace{a, \dots, a}_{k}) = g(\underbrace{t, \dots, t}_{n-k}, \underbrace{a, \dots, a}_{k}) \quad \text{for } k = 0, \dots, r.$$

As $F(t) = G(t)$ the equation holds for $k = 0$. Now assume that the equation holds up to $k-1 \geq 0$. In (1.29) a and b are arbitrary so if we put $b = t$ then we have

$$\sum_{i=0}^{k} (-1)^i \binom{k}{i} f(\underbrace{t, \dots, t}_{n-i}, \underbrace{a, \dots, a}_{i}) = \sum_{i=0}^{k} (-1)^i \binom{k}{i} g(\underbrace{t, \dots, t}_{n-i}, \underbrace{a, \dots, a}_{i})$$

$$
\begin{array}{ccccccccc}
f(s_1,\ldots,s_n) & & f(s_2,\ldots,s_{n+1}) & & \cdots & & f(s_n,\ldots,s_{2n-1}) & & f(s_{n+1},\ldots,s_{2n}) \\
& \searrow (1-\alpha_1^1) \quad \alpha_1^1 \swarrow & & \searrow (1-\alpha_2^1) & & & \alpha_{n-1}^1 \swarrow & \searrow (1-\alpha_n^1) \quad \alpha_n^1 \swarrow & \\
& f(s_2,\ldots,s_n,t_1) & & f(s_3,\ldots,t_1,s_{n+1}) & \cdots & f(s_n,t_1,\ldots,s_{2n-2}) & & f(t_1,s_{n+1}\ldots,s_{2n-1}) & \\
& & \searrow (1-\alpha_2^2) \quad \alpha_2^2 \swarrow & & \searrow (1-\alpha_3^2) \quad \alpha_{n-1}^2 \swarrow & & \searrow (1-\alpha_n^2) \quad \alpha_n^2 \swarrow & & \\
& & f(s_3,\ldots,s_{n-1},t_1,t_2) & & \cdots & & f(t_1,t_2,s_{n-1}\ldots,s_{2n-2}) & & \\
& & & \searrow (1-\alpha_3^3) & & \alpha_n^3 \swarrow & & & \\
& & & & \vdots & & & & \\
& & & \searrow (1-\alpha_{n-1}^{n-1}) \quad \alpha_{n-1}^{n-1} \swarrow & & \searrow (1-\alpha_n^{n-1}) \quad \alpha_n^{n-1} \swarrow & & & \\
& & & f(s_n,t_1,\ldots,t_{n-1}) & & f(t_1,\ldots,t_{n-1},s_{n+1}) & & & \\
& & & & \searrow (1-\alpha_n^n) \quad \alpha_n^n \swarrow & & & & \\
& & & & f(t_1,\ldots,t_n) & & & &
\end{array}
$$

Figure 1.16: de Boor's algorithm for polar forms. The entries of each pair of neighbours become the same if the first entry of the left and the last entry of the right is deleted. We can determine the value when the deleted entry is replaced by $t_k = \left(1-\alpha_i^k\right)s_i + \alpha_i^k s_{i+n-k+1}$ by multiplying on the left with $(1-\alpha_i^k)$, on the right with α_i^k and adding the results. If $t_1 = \cdots = t_n = t$ and $s_n \le t \le s_{n+1}$ then we have the ordinary de Boor's algorithm.

By the induction hypothesis we have

$$\sum_{i=0}^{k-1}(-1)^i\binom{k}{i} f(\underbrace{t,\ldots,t}_{n-i},\underbrace{a,\ldots,a}_{i}) = \sum_{i=0}^{k-1}(-1)^i\binom{k}{i} g(\underbrace{t,\ldots,t}_{n-i},\underbrace{a,\ldots,a}_{i})$$

so subtracting these two equations gives us

$$f(\underbrace{t,\ldots,t}_{n-k},\underbrace{a,\ldots,a}_{k}) = g(\underbrace{t,\ldots,t}_{n-k},\underbrace{a,\ldots,a}_{k}),$$

as wanted. Now note that the two functions

$$f^*(t_1,\ldots,t_r) = f(t_1,\ldots,t_r,t,\ldots,t)$$

and

$$g^*(t_1,\ldots,t_r) = g(t_1,\ldots,t_r,t,\ldots,t)$$

are symmetric and n-affine and we have just shown that

$$f^*(\underbrace{t,\ldots,t}_{k},\underbrace{a,\ldots,a}_{r-k}) = g^*(\underbrace{t,\ldots,t}_{k},\underbrace{a,\ldots,a}_{r-k}) \quad \text{for } k = 0,\ldots,r.$$

Theorem 1.23 then shows that $f^* = g^*$. □

Theorem 1.25. *Let $F(t)$ and $G(t)$ be two polynomials of degree at most n with corresponding n-polar forms f, g, and let there be given numbers*

$$s_1 \le \cdots \le s_{n-r} < s_{n-r+1} = \cdots = s_n < s_{n+1} \le \cdots \le s_{2n-r},$$

where $1 \le r \le n$. Then

$$F^{(k)}(s_n) = G^{(k)}(s_n), \quad k = 0,\ldots,n-r,$$

if and only if

$$f(s_i,\ldots,s_{n+i}) = g(s_i,\ldots,s_{n+i}), \quad i = 1,\ldots,n-r+1.$$

Proof. Let $t = s_n$, $u_i = s_i$ for $i \le n-r$, and $u_i = s_{i+r}$ for $i > n-r$, i.e., we have

$$u_1 \le \cdots \le u_{n-r} < \underbrace{t = \cdots = t}_{r} < u_{n-r+1} \le \cdots \le u_{2n-2r}.$$

If we put

$$f^*(t_1, \ldots, t_{n-r}) = f(t_1, \ldots, t_{n-r}, \underbrace{t, \ldots, t}_{r})$$

and

$$g^*(t_1, \ldots, t_{n-r}) = g(t_1, \ldots, t_{n-r}, \underbrace{t, \ldots, t}_{r}),$$

then Theorem 1.23 shows that $f^* = g^*$ if and only if

$$f^*(u_{1+i}, \ldots, u_{i+n-r}) = g^*(u_{1+i}, \ldots, u_{i+n-r}), \quad \text{for all } i = 0, \ldots, n-r.$$

The result is now a consequence of Theorem 1.24. □

Problems

1.5.1 Determine the 3-polar forms for:

(a) The polynomials $1, t, t^2, t^3$.

(b) The Hermite polynomials of degree 3, cf. (1.4), p. 3.

(c) The Bernstein polynomials of degree 3.

1.5.2 Let $f(t_1, t_2)$ be a symmetric bi-affine function. Show that the three values $f(0, 0)$, $f(1, 0)$, and $f(2, 0)$ are not independent and don't determine f uniquely.

1.5.3 Construct the full triangular scheme in Figure 1.16 in the cubic case, $n = 3$.

Exercises

1.5.1 Write a program that calculates the polar form of a Bézier curve.

1.6 B-spline curves

If we want a large degree of flexibility of a polynomial then the degree has to be large. The evaluation becomes more expensive and the control polygon no longer has to resemble the curve, cf. Figure 1.7. Furthermore, if a control point is changed, then the whole curve is changed. The solution to these problems is to use several polynomials defined on different intervals and meeting with a certain degree of differentiability. This will be secured using Theorem 1.25 of the previous section.

Definition 1.26. A *knot sequence* or *knot vector* in degree n is a sequence

$$\underbrace{t_0 \le \cdots \le t_n}_{\text{boundary knots}} < \underbrace{t_{n+1} \le \cdots \le t_{n+N-1}}_{\text{inner knots}} < \underbrace{t_{n+N} \le \cdots \le t_{2n+N}}_{\text{boundary knots}}$$

The first and last $n+1$ knots are called *boundary knots*, the others are called *inner knots*. If $t_{r-1} < t_r = \cdots = t_{r+\nu-1} < t_{r+\nu}$ then we say that the knot t_r has *multiplicity* ν. We very often have $t_0 = \cdots = t_n$ and $t_{n+N} = \cdots = t_{2n+N}$, i.e., the boundary knots have multiplicity $n+1$.

Definition 1.27. A *B-spline curve* of degree n with knot sequence $t_0, \dots, t_{2n+N}$ and *control points* or *de Boor points* $\mathbf{d}_1 \dots, \mathbf{d}_{N+n}$ is a piecewise polynomial curve of degree n defined on the interval $[t_n, t_{n+N}]$. The polynomial segments are defined on intervals of the form $[t_r, t_{r+1}]$, with $t_r < t_{r+1}$ and the corresponding polar form f_r is given by

$$f_r(t_{r-n+i}, \dots, t_{r+i-1}) = \mathbf{d}_{r-n+i}, \quad i = 1, \dots, n+1$$

cf. Theorem 1.23. I.e., the polar form evaluated on all $n+1$ sets of n consecutive knots in the subsequence $t_{r-n+1}, \dots, t_r, t_{r+1}, \dots, t_{r+n}$.

Remark 1.28. The outermost knots t_0 and t_{2n+N} do not enter into the definition. Their presence is formally needed when we introduce the *basis spline functions* (B-splines), cf. Remark 1.34.

In Figure 1.17 we have plotted B-spline curves of various degrees. If a knot has multiplicity n, so $t_r = \cdots = t_{r+n-1}$, say, then one of the control points are of the form $f(t_r, \dots, t_r)$, i.e. it is a point on the curve, and we say that the knot has *full multiplicity*. If the multiplicity is $n+1$ (or higher) then the curve may be discontinuous. As an immediate consequence of Theorem 1.25 we have

Theorem 1.29. *If an inner knot t_r has multiplicity ν then a B-spline curve of degree n is $C^{n-\nu}$ at t_r.*

Conversely we have

Theorem 1.30. *Suppose we have numbers $u_0, \dots, u_m$, integers $\nu_1, \dots, \nu_{m-1}$, and a curve defined on $[u_0, u_m]$ such that $u_{j-1} < u_j$ and*

1. *The restriction to each subinterval $[u_{j-1}, u_j]$ is polynomial of degree n, $j = 1, \dots, m$.*

2. *It is $C^{n-\nu_j}$ at t_j, $j = 1, \dots, m-1$.*

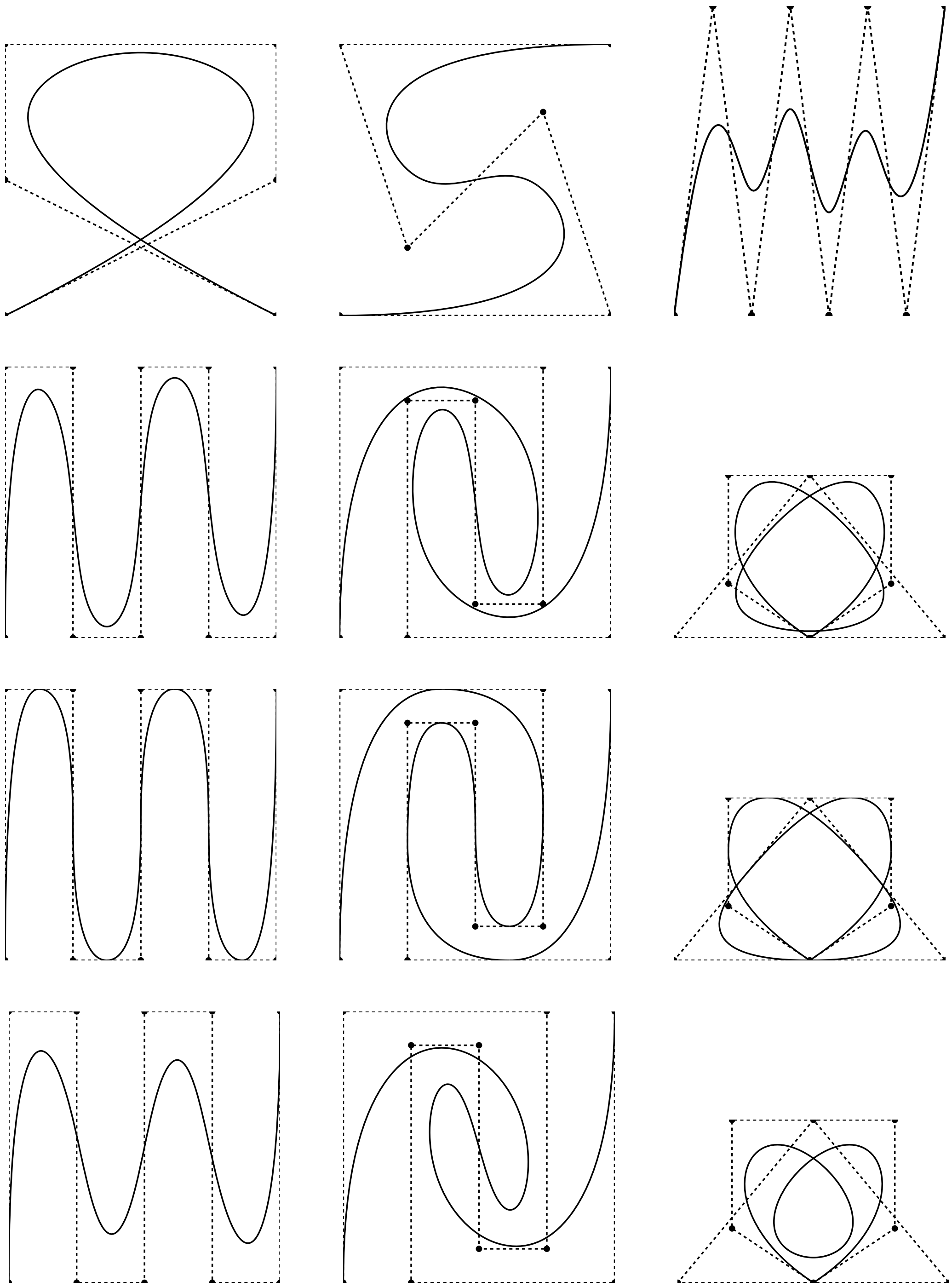


Figure 1.17: B-spline curves. In the first two rows the degree is 3, in the third row the degree is 2, and in the last row the degree is 5. In all cases the boundary knots have full multiplicity and the inner knots are uniform. Compare with Figure 1.7.

Then the curve can be written as a B-spline curve with knot sequence

$$t_0, \dots, t_{2n+N} = \underbrace{u_0, \dots, u_0}_{n+1}, \underbrace{u_1, \dots, u_1}_{\nu_1}, \dots, \underbrace{u_{m-1}, \dots, u_{m-1}}_{\nu_{m-1}}, \underbrace{u_m, \dots, u_m}_{n+1}$$

where $N = 1 + \sum_{j=1}^{m-1} \nu_j$. *Furthermore, if we put* $m_r = n + \sum_{j=1}^{r} \nu_j$, *for* $r = 0, \dots, m$, *then the control points are given by*

$$\mathbf{d}_i = f_r(t_i, \dots, t_{i+n-1}) \quad \textit{for } i = m_{r-1} - n + 1, \dots, m_{r-1} + 1,$$

where f_r *is the polar form of the polynomial segment defined on* $[u_{r-1}, u_r] = [t_{m_{r-1}}, t_{m_{r-1}+1}]$, $r = 1, \dots, m$.

Proof. We only have to show that the control points are well defined. I.e., if $m_{r-1} - n + 1 \le i \le m_{r+l-1} - n + 1$ then we need to show that

$$f_r(t_i, \dots, t_{i+n-1}) = f_{r+l}(t_i, \dots, t_{i+n-1}).$$

If $l = 1$ this follows from Theorem 1.25, and induction gives the result for a general l. □

This theorem gives in particular that a B-spline defined on some knot sequence can be considered as a B-spline on a refined knot sequence and it also tells us what the new control points are. The process of refining a knot sequence is called *knot insertion*. Inserting a single knot is described in the following theorem.

Theorem 1.31. *Suppose* $\mathbf{d}_1, \dots, \mathbf{d}_{n+N}$ *are the control points for a B-spline curve of degree* n *on the knot sequence* $t_0, \dots, t_{2n+N}$. *Let* $t^* \in [t_{r-1}, t_r]$, *such that* $t_n \le t_{r-1} < t_r \le t_{n+N}$. *If we insert* t^* *in the knot sequence then the new control points are given by*

$$\mathbf{d}_i^* = \begin{cases} \mathbf{d}_i & i = 1, \dots, r-n \\ (1-\alpha_i)\mathbf{d}_{i-1} + \alpha_i \mathbf{d}_i & i = r-n+1, \dots, r \\ \mathbf{d}_{i-1} & i = r+1, \dots, n+N+1 \end{cases}$$

where $\alpha_i = \dfrac{t^* - t_{i-1}}{t_{i+n-1} - t_{i-1}}$.

Proof. Let $t_1^*, \dots, t_{2n+m}^*$ be the new knot sequence, i.e.,

$$t_i^* = \begin{cases} t_i & i = 0, \dots, r-1 \\ t^* & i = r \\ t_{i-1} & i = r+1, \dots, 2n+N \end{cases}$$

If $i \leq r - n$, then $i + n - 1 \leq r - 1$ and

$$\mathbf{d}_i^* = g(t_i^*, \dots, t_{i+n-1}^*) = g(t_i, \dots, t_{i+n-1}) = \mathbf{d}_i$$

where g is the polar form of some polynomial segment. Likewise, if $i \geq r + 1$, then

$$\mathbf{d}_i^* = g(t_i^*, \dots, t_{i+n-1}^*) = g(t_{i-1}, \dots, t_{i+n-2}) = \mathbf{d}_{i-1}$$

where g is the polar form of some polynomial segment.

We now let f be the polar form of the polynomial segment defined on the interval $[t_{r-1}, t_r]$ and consider $r - n < i < r + 1$. The new control points are given by

$$\mathbf{d}_i^* = f(t_i^*, \dots, t_{i+n-1}^*) = f(t_i, \dots, t_{r-1}, t^*, t_r, \dots, t_{i+n-2}).$$

If we in the proof of Theorem 1.23 let $t_1 = t^*$ and $s_i = t_{i+r-n}$ then we get

$$\begin{aligned}
&f(t_i, \dots, t_{r-1}, t^*, t_r, \dots, t_{i+n-2})\\
&\qquad = f(s_{i-r+n}, \dots, s_{n-1}, t_1, s_n, \dots, s_{i+2n-r-2})\\
&\qquad = (1 - \alpha_{i-r+n-1}^1) f(s_{i-r+n-1}, \dots, s_{i+2n-r-2})\\
&\qquad\quad + \alpha_{i-r+n-1}^1 f(s_{i-r+n}, \dots, s_{i+2n-r-1})\\
&\qquad = (1 - \alpha_i) f(t_{i-1}, \dots, t_{i+n-2}) + \alpha_i f(t_i, \dots, t_{i+n-1})\\
&\qquad = (1 - \alpha_i)\mathbf{d}_{i-1} + \alpha_i \mathbf{d}_i
\end{aligned}$$

where

$$\alpha_i = \alpha_{i-r+n-1}^1 = \frac{t_1 - s_{i-r+n-1}}{s_{i-r+2n-1} - s_{i-r+n-1}} = \frac{t^* - t_{i-1}}{t_{i+n-1} - t_{i-1}}.$$

See the first two rows in Figure 1.16 □

The *de Boor's algorithm* is the process of repeated insertion of a knot until we get full multiplicity and have obtained a point on the curve. We can formulate it as a theorem:

Theorem 1.32. *Let* $\mathbf{r}$ *be a B-spline curve of degree* n *with knots* $t_0, \dots, t_{2n+N}$ *and control points* $\mathbf{d}_1, \dots, \mathbf{d}_{n+N}$*. If* $t \in [t_{r-1}, t_r]$ *and* $t_n \leq t_{r-1} < t_r \leq t_{n+N}$ *then we can determine the point* $\mathbf{r}(t)$ *on the curve by de Boor's algorithm. First initialize:*

$$\begin{aligned}
s_i &= t_{r-n+i} & i &= 1, \dots, 2n\\
\mathbf{d}_i^0(t) &= \mathbf{d}_{r+1-n+i} & i &= 0, \dots, n
\end{aligned}$$

Then for $k = 1, \dots, n$ *do:*

$$\left.\begin{aligned}
\alpha_i^k &= \frac{t - s_i}{s_{n+1+i-k} - s_i}\\
\mathbf{d}_i^k(t) &= \left(1 - \alpha_i^k\right)\mathbf{d}_{i-1}^{k-1}(t) + \alpha_i^k \mathbf{d}_i^{k-1}(t)
\end{aligned}\right\} \qquad i = k, \dots, n$$

Finally, the point on the curve is $\mathbf{r}(t) = \mathbf{d}_n^n(t)$.

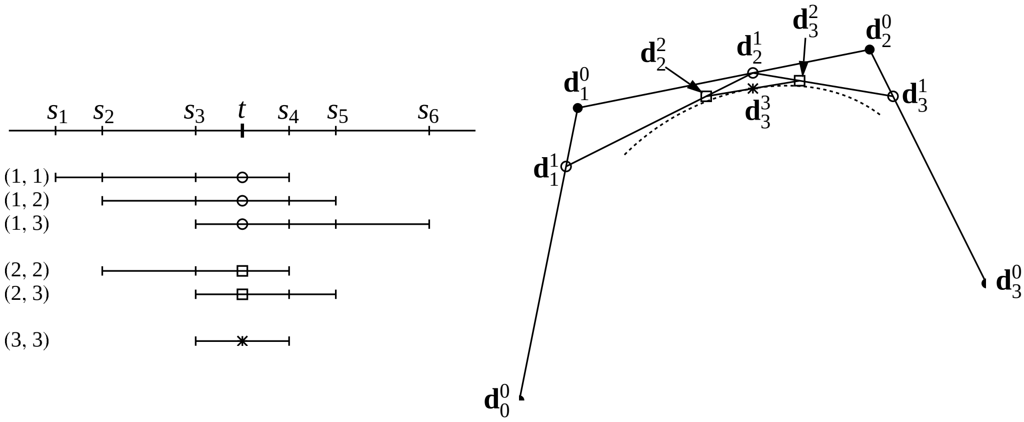


Figure 1.18: de Boor's algorithm for a cubic spline. To the left we have the parameter line. When the interval labeled (k, i) is mapped affinely to the edge $\mathbf{d}_{i-1}^{k-1}\mathbf{d}_i^{k-1}$, then t is mapped to the point $\mathbf{d}_i^k$.

This is the same algorithm that is described in Figure 1.16. We just have to put $t_1 = \cdots = t_n = t$ and

$$\mathbf{d}_i^k(t) = f(\underbrace{s_{1+k+i}, \ldots, s_n}_{n-k-i}, \underbrace{t, \ldots, t}_{k}, \underbrace{s_{n+1}, \ldots, s_{n+i}}_{i}).$$

See Figure 1.18 for a geometric picture. As we repeatedly uses convex combinations we have the convex hull property, but if $t \in [t_{r-1}, t_r]$, then we only uses the control points $\mathbf{d}_{r-n}, \ldots, \mathbf{d}_{r+1}$ so we have *the strong convex hull property*: $\mathbf{r}\big([t_{r-1}, t_r]\big)$ is contained in the convex hull of $\{\mathbf{d}_{r-n}, \ldots, \mathbf{d}_{r+1}\}$, see Figure 1.19. The derivative of a B-spline curve can be found by formulas analogous to Theo-

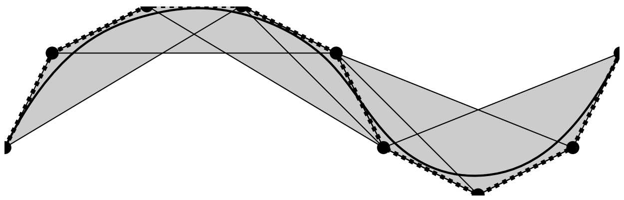


Figure 1.19: A B-spline curve of degree n is contained in the union of the convex hull of any set of $n + 1$ consecutive knots.

rem 1.9

Theorem 1.33. *Let* $\mathbf{r}$ *be a B-spline curve of degree* n*, with knots* $t_0, \ldots, t_{2n+N}$*, and control points* $\mathbf{d}_1, \ldots, \mathbf{d}_{n+N}$*. The derivative* $\mathbf{r}'$ *is a B-spline curve of degree*

$n-1$, *with knot sequence* $t_1, \ldots, t_{2n+N-1}$, *(i.e., with the same inner knots) and control points*

$$\mathbf{d}_i' = \frac{n\big(\mathbf{d}_i - \mathbf{d}_{i-1}\big)}{t_{i+n-1} - t_{i-1}}. \tag{1.31}$$

Alternatively, if $t \in [t_r, t_{r+1}]$, *where* $t_r < t_{r+1}$, *then run de Boor's algorithm to the second last step,* (*i.e., insert* t *as a knot to multiplicity* $n-1$). *The derivative is then given by*

$$\mathbf{r}'(t) = \frac{n\big(\mathbf{d}_n^{n-1}(t) - \mathbf{d}_{n-1}^{n-1}(t)\big)}{s_{n+1} - s_n} = \frac{n\big(\mathbf{d}_n^{n-1}(t) - \mathbf{d}_{n-1}^{n-1}(t)\big)}{t_{r+1} - t_r} \tag{1.32}$$

Proof. It is clear that $\mathbf{r}'$ is piecewise polynomial of degree $n-1$ and if $\mathbf{r}$ is $C^{n-\nu}$ at some inner knot, then $\mathbf{r}'$ is $C^{n-\nu-1}$ at the same knot. Theorem 1.30 now implies that the derivative $\mathbf{r}'$ has the same inner knots as the original curve $\mathbf{r}$. Let f be the n-polar form for $\mathbf{r}$ in the interval $[t_{i+n-1}, t_{i+n}]$. By (1.28), p. 24, we see that the $(n-1)$-polar form for $\mathbf{r}'$ in the same interval is given by

$$(u_1, \ldots, u_{n-1}) \mapsto \frac{n}{t_{i+n-1} - t_{i-1}}\big(f(u_1, \ldots, u_{n-1}, t_{i+n-1}) - f(u_1, \ldots, u_{n-1}, t_{i-1})\big).$$

Using Theorem 1.30 again we see that the control points are given by

$$\begin{aligned}\mathbf{d}_i' = \frac{n}{t_{i+n-1} - t_{i-1}}\big(f(t_i, \ldots, t_{i+n-2}, t_{i+n-1}) - f(t_i, \ldots, t_{i+n-2}, t_{i-1})\big)\\ = \frac{n}{t_{i+n-1} - t_{i-1}}\big(\mathbf{d}_i - \mathbf{d}_{i-1}\big).\end{aligned}$$

This establishes (1.31). Let f be the n-polar form for $\mathbf{r}$ in the interval $[t_r, t_{r+1}]$. The two points in the second last step of de Boor's algorithm are given by

$$\begin{aligned}\mathbf{d}_{n-1}^{n-1}(t) &= f(t, \ldots, t, s_n) = f(t, \ldots, t, t_r)\\ \mathbf{d}_n^{n-1}(t) &= f(t, \ldots, t, s_{n+1}) = f(t, \ldots, t, t_{r+1})\end{aligned}$$

cf. Figure 1.16. The expression (1.32) is now a consequence of (1.28). □

We have a *closed* B-spline if we have a (bi-infinite) knot sequence with periodic differences, and a (bi-infinite) periodic control polygon. Alternatively we have a B-spline defined on $[t_n, t_{n+N}]$ with the knot sequence $t_0, \ldots, t_{2n+N}$ and control polygon $\mathbf{d}_1, \ldots, \mathbf{d}_{n+N}$ such that

$$t_{i+N} - t_{i+N-1} = t_i - t_{i-1}, \qquad i = 1, \ldots, 2n, \tag{1.33}$$

$$\mathbf{d}_{i+N} = \mathbf{d}_i, \qquad i = 1, \ldots, n+1, \tag{1.34}$$

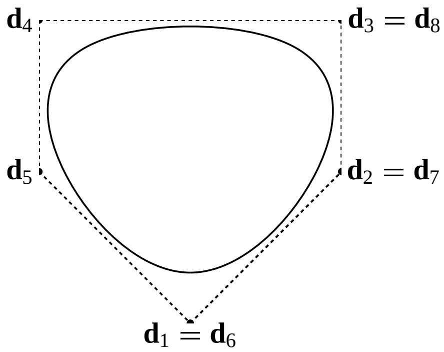


Figure 1.20: A closed cubic B-spline with a uniform knot sequence (0, 1, . . . , 11).

see Figure 1.20.

If we are given a knot sequence and for $i = 1, \dots, N + n$ apply de Boor's algorithm to the scalar coefficients (control points)

$$d_j = \delta_{ij} = \begin{cases} 1, & i = j \\ 0 & i \neq j \end{cases} \qquad j = 1, \dots, N + n,$$

then we obtain $N + n$ functions $N_i^n(t)$ called the *basis-* or *B-splines*, see Figure 1.21. Sometimes we will write $N_i^n(t|\mathbf{t})$ to emphasize the dependence on the knot sequence $\mathbf{t} = t_0, \dots, t_{2n+N}$. The B-splines can be used to parameterize a B-spline curve explicitly, cf. Problem 1.6.1:

$$\mathbf{r}(t) = \sum_{i=1}^{n+N} \mathbf{d}_i N_i^n(t) \tag{1.35}$$

As the de Boor algorithm only uses the coefficients $d_{r+1-n}, \dots, d_{r+1}$ in order to evaluate the B-spline $N_i^n(t)$ for $t \in [t_r, t_{r+1}]$, it follows that

$$N_i^n(t) = 0 \quad \text{if } t \notin [t_{i-1}, t_{i+n}].$$

On the other hand, when we run de Boor's algorithm for a $t \in]t_r, t_{r+1}[$, then the coefficients α_j^k and $1 - \alpha_j^k$ are strictly positive so

$$N_i^n(t) > 0 \quad \text{if } t \in]t_{i-1}, t_{i+n}[.$$

All in all we have that the *support* of a B-spline is

$$\operatorname{supp}\big(N_i^n\big) = [t_{i-1}, t_{i+n}]. \tag{1.36}$$

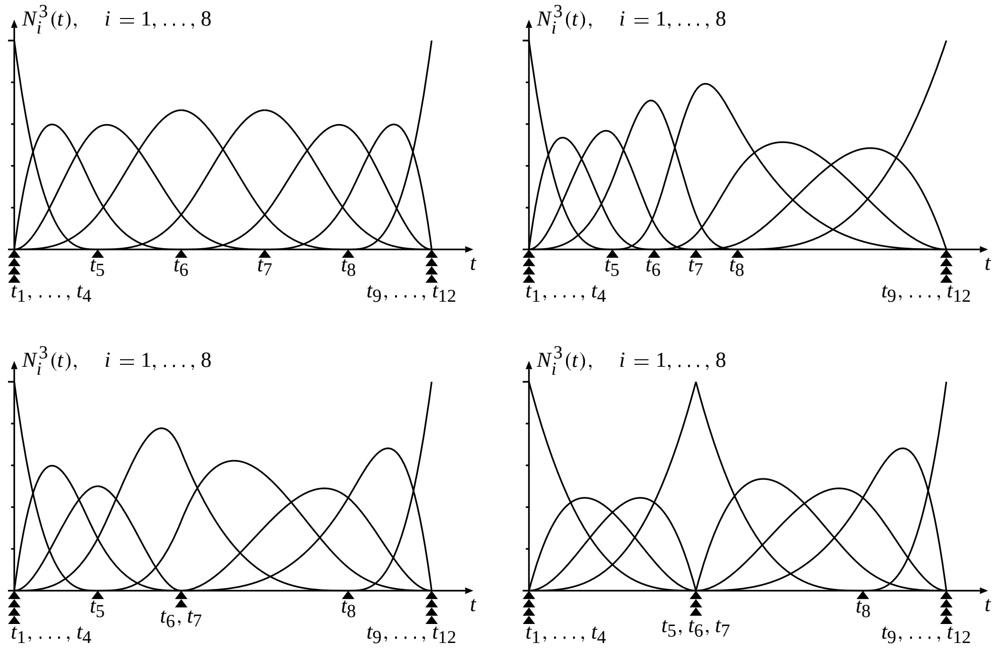


Figure 1.21: Cubic B-splines with different knot sequences

Remark 1.34. If $i = 1$ then we need the knot t_0 and if $i = N - n$ then we need the knot t_{2n+N}. On the other hand, these two knots only have an effect on the B-splines outside the interval $[t_n, t_{n+N}]$, so they have no influence on a B-spline curve, cf. Remark 1.28.

If we choose control points $d_i = 1$ for all i, then the resulting function is constant 1, so the B-splines form a *partition of unity* on the interval $[t_n, \dots, t_{n+N}]$:

$$\begin{aligned} N_i^n(t) &\geq 0, \quad i = 1 \dots, N + n, \\ \sum_{i=1}^{N+n} N_i^n(t) &= 1 \end{aligned} \tag{1.37}$$

The B-splines can be found by a recurrence analogous to (1.13),

$$\begin{aligned} N_i^0(t) &= \begin{cases} 1 & \text{if } t \in [t_{i-1}, t_i[\\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \dots, 2n + N. \\ N_i^r(t) &= \frac{t - t_{i-1}}{t_{i+r-1} - t_{i-1}} N_i^{r-1}(t) + \frac{t_{i+r} - t}{t_{i+r} - t_i} N_{i+1}^{r-1}(t), \\ & i = 1, \dots, 2n + N - r, \quad r = 1, \dots, n. \end{aligned} \tag{1.38}$$

Once more the outermost knots t_0 and t_{2n+N} are needed. The derivative of a B-spline is

$$\frac{\mathrm{d}}{\mathrm{d}t} N_i^n(t) = \frac{n}{t_{n+i-1} - t_{i-1}} N_i^{n-1}(t) - \frac{n}{t_{n+i} - t_i} N_{i+1}^{n-1}(t). \qquad (1.39)$$

The higher order derivative can be found by recursive use of this equation:

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k} N_i^n(t) = \frac{n}{t_{n+i-1} - t_{i-1}} \frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}} N_i^{n-1}(t) - \frac{n}{t_{n+i} - t_i} \frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}} N_{i+1}^{n-1}(t), \qquad (1.40)$$

or we can extend the recursion (1.38) to include the derivatives to any desired order:

$$\begin{aligned} \frac{\mathrm{d}^k}{\mathrm{d}t^k} N_i^r(t) = {} & \frac{t - t_{i-1}}{t_{i+r-1} - t_{i-1}} \frac{\mathrm{d}^k}{\mathrm{d}t^k} N_i^{r-1}(t) + \frac{t_{i+r} - t}{t_{i+r} - t_i} \frac{\mathrm{d}^k}{\mathrm{d}t^k} N_{i+1}^{r-1}(t) \\ & + \frac{k}{t_{i+r-1} - t_{i-1}} \frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}} N_i^{r-1}(t) - \frac{k}{t_{i+r} - t_i} \frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}} N_{i+1}^{r-1}(t). \qquad (1.41) \end{aligned}$$

To start the recursion we have of course that $\frac{\mathrm{d}^k}{\mathrm{d}t^k} N_i^0(t) = 0$ for all i and $k > 0$.

Let $f(t) = \sum_{i=1}^{n+N} d_i N_i^n(t)$. Obviously the graph is a B-spline curve and now we want to find the control points. The ordinates are by definition d_i, but what are the abscissas? The graph is a curve parametrized by $\big(t, f(t)\big)$ so we want to express $t \mapsto t$ as a B-spline function. The de Boor control points of this function are called the *Greville abscissas* and we only need to find the n-polar form and insert n consecutive knots. This is easy, the n-polar form of the polynomial t is $\frac{1}{n}(t_1 + \cdots + t_n)$ so the Greville abscissas are

$$\xi_i = \frac{1}{n}\big(t_i + \cdots + t_{i+n-1}\big), \qquad (1.42)$$

and the graph can be parametrized as

$$\begin{pmatrix} t \\ f(t) \end{pmatrix} = \sum_{i=1}^{n+N} \begin{pmatrix} \xi_i \\ d_i \end{pmatrix} N_i^n(t).$$

Problems

1.6.1 Prove that (1.35) is a parametrization of the B-spline curve.

1.6.2 Prove the recurence formula (1.38). Hint: Consider the de Boor algorithm for a $t \in [t_{r-1}, t_r[$,

$$
\begin{array}{ccccccccc}
\dots & \mathbf{d}_{r-n} & & \mathbf{d}_{r-n+1} & \dots & \mathbf{d}_{r-1} & & \mathbf{d}_r & \dots \\
 & \searrow & & \swarrow & & \searrow & & \swarrow & \\
 & & \mathbf{d}_{r-n+1}^{(1)}(t) & & \dots & & \mathbf{d}_r^{(1)}(t) & & \\
 & & \searrow & & & & \swarrow & & \\
 & & & & \vdots & & & & \\
 & & & \searrow & & \swarrow & & & \\
 & & & & \mathbf{d}_r^{(n)}(t) & & & &
\end{array}
$$

and prove by induction that

$$\mathbf{r}(t) = \sum_{i=r-n+m}^{r} \mathbf{d}_i^m(t) N_i^{n-m}(t),$$

where we now let $N_i^{n-m}(t)$, $m = n, \dots, 0$ be given by (1.38). First show that the equation is true for $m = n$. Then assume it's true for some m (with $0 < m \leq n$). Write $\mathbf{d}_i^m$ in terms of $\mathbf{d}_i^{m-1}$, collect coeficients of $\mathbf{d}_i^{m-1}$ and compare with (1.38).

1.6.3 Let $N_i^n(t)$ be the B-splines of degree n on the bi-infinite uniform knot sequence $\mathbb{Z}$.

(a) Show that $N_i^n(t) = N_0^n(t - i)$ for all $i \in \mathbb{Z}$.

(b) Prove the *refinement* equation:

$$N_1^n(t) = \sum_{i=1}^{n} N_i^n(2t) = \sum_{i=1}^{n} N_1^n(2t - i + 1).$$

1.6.4 Prove (1.39), (1.40), and (1.41).

Exercises

1.6.1 Implement the knot insertion procedure.

1.6.2 Implement de Boor's algorithm.

1.6.3 Write a program that splits a B-spline curve into its Bézier segments.

1.6.4 Write a program that finds the derivative of a B-spline curve.

1.6.5 Write a program that calculates all B-splines of degree n on a given knot sequence, using the recurence relation (1.38).

1.6.6 Write a program that calculates all B-splines of degree n and their derivatives to order $k \leq n$ on a given knot sequence.

Chapter 2

Differential Geometry of Curves

2.1 Introduction

Intuitively a curve is a one-dimensional object, i.e., an object that can be described by a single parameter. A curve with a particular choice of parameter is called a *parameterized curve* and we have in the previous chapter already seen many examples of this concept.

In this chapter we will study *local* properties of abstract curves. The main result is that a plane curve is completely determined by a single real valued function, the *curvature*, and a space curve is completely determined by two real valued functions, the *curvature* and *torsion*. A curve in $\mathbb{R}^n$ is completely determined by $n - 1$ functions, called the curvatures.

2.2 Parameterized Curves

Our study of curves will be restricted to a certain class of curves. First of all we want to use calculus in the analysis so a curve has to be described by a differential function[1]. If the derivative of a map $\mathbf{r} : \mathbb{R} \to \mathbb{R}^n$ vanishes at some point then the image can have sharp corner or a cusp, see Problems 2.2.1 and 2.2.2, and we want to avoid that too. So we will only work with *regular curves*. Our main interest are plane curves or space curves so in the following you may think of $\mathbb{R}^n$ as $\mathbb{R}^2$ or $\mathbb{R}^3$.

Definition 2.1. A *regular parametrization* of class C^k, with $k \geq 1$, of a curve in $\mathbb{R}^n$ is a vector function $\mathbf{r} : I \to \mathbb{R}^n$ defined on an interval I which satisfies

[1] Besides the convenience of being able to use calculus there is a more severe reasons for insisting on differentiable functions. There exists continuous maps $[0, 1] \to [0, 1]^n$ whose image is all of $[0, 1]^n$ and we do not want to call them curves.

1. $\mathbf{r}$ is of class C^k.
2. $\mathbf{r}'(t) \neq \mathbf{0}$ for all $t \in I$.

The variable t is called the *parameter* and I is called the *parameter interval*.

Example 2.1 The function $\mathbf{r}(t) = (t, t^2 - 1)$, $t \in \mathbb{R}$, is a regular parametrization because $\mathbf{r}$ is of class C^∞ and $\mathbf{r}'(t) = (1, 2t) \neq (0, 0)$ for all $t \in \mathbb{R}$. The image of $\mathbf{r}$ is the parabola shown in Figure 2.1

Example 2.2 The function $\mathbf{r}(t) = (r \cos t, r \sin t, ht)$, where $r, h > 0$, is a regular parametrization because $\mathbf{r}$ is of class C^∞ and $|\mathbf{r}'(t)|^2 = r^2 \sin^2 t + r^2 \cos^2 t + h^2 = r^2 + h^2 \neq 0$ for all $t \in \mathbb{R}$. The image of $\mathbf{r}$ is the *right circular helix* shown in Figure 2.1.

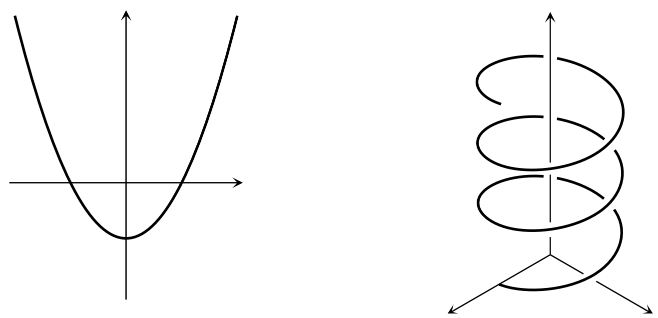


Figure 2.1: To the left a parabola to the right a circular helix.

Definition 2.2. An *allowable change of parameter* of class C^k is a real function $f : I_1 \to I$ such that

1. f is of class C^k.
2. $f'(t) \neq 0$ all $t \in I_1$.

As I is an interval we have either $f'(t) > 0$ for all $t \in I$, in which case we call f *orientation preserving*, or $f'(t) < 0$ for all $t \in I$, in which case we call f *orientation reversing*. If $f : I_1 \to I$ is an allowable change of parameter of class C^k then the condition $f'(t) \neq 0$ implies that the inverse exists and is an allowable change of parameter of class C^k. If $\mathbf{r} : I \to \mathbb{R}^n$ is a regular parameterization of a curve and $f : I_1 \to I$ is an allowable change of parameter and both are of class C^k then $\mathbf{r}_1 = \mathbf{r} \circ f : I_1 \to \mathbb{R}^n$ is of class C^k too, and it satisfies $\mathbf{r}_1'(t) = \mathbf{r}'(f(t)) f'(t) \neq \mathbf{0}$, i.e., it is a regular parametrization, see Figure 2.2. We say that $\mathbf{r}_1$ is a *reparameterization* of $\mathbf{r}$, and this defines an equivalence relation on the set of parametrizations, cf. Problem 2.2.5. We will consider a regular parametrization of class C^k and any reparametrization as defining the same *curve*, that is

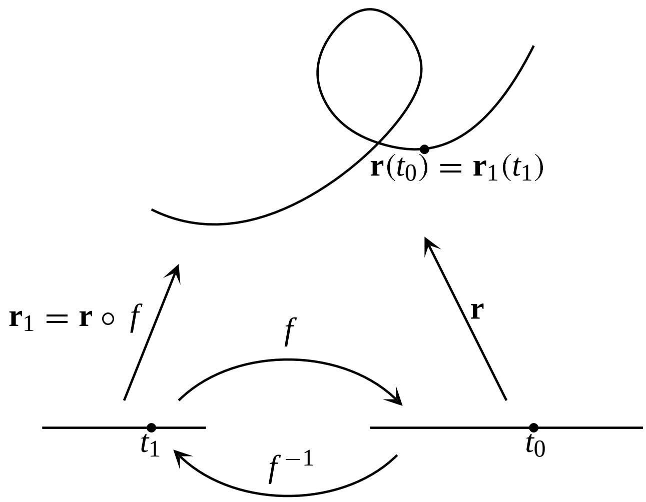


Figure 2.2: Reparametrization of a curve

Definition 2.3. A *regular curve* in $\mathbb{R}^n$ is a collection of regular parametrizations $\mathbf{r} : I \to \mathbb{R}^n$ of class C^k any two of which are reparametrizations of each other.

An *oriented regular curve* in $\mathbb{R}^n$ is a collection of regular parametrizations $\mathbf{r} : I \to \mathbb{R}^n$ of class C^k any two of which are orientation preserving reparametrizations of each other.

A regular parametrization $\mathbf{r} : I \to \mathbb{R}^n$ uniquely determines a curve and all other parametrizations are related to it by an allowable change of parameter. Thus we may say "the curve given by $\mathbf{r}(t)$...". However, a property of or a concept associated with the parametrization $\mathbf{r} : I \to \mathbb{R}^n$ need not be a property of the underlying curve. Any property of or concept associated with the curve must be common to all representations or, as we say, "independent of the parameter".

A regular curve given by $\mathbf{r} : I \to \mathbb{R}^n$ is said to be *simple* if there are no multiple points; that is, if $t_1 \neq t_2$ implies $\mathbf{r}(t_1) \neq \mathbf{r}(t_2)$. This is clearly a property of the curve, not of the parametrization. Locally, a regular curve is always simple, cf. Problem 2.2.6.

If we think of the curve as the path of a moving particle then $\mathbf{r}'(t_0)$ is the velocity of the particle at time $t = t_0$.

Definition 2.4. The *velocity vector* of a regular parametrization $\mathbf{r} : I \to \mathbb{R}^n$ at $t = t_0$ is the derivative $\mathbf{r}'(t_0)$. The *velocity vector field* is the vector valued function $\mathbf{r}' : I \to \mathbb{R}^n$. The *speed* of $\mathbf{r}$ at $t = t_0$ is the length of the velocity vector at $t = t_0$, $\left|\mathbf{r}'(t_0)\right|$. The *tangent vector* is the unit vector $\mathbf{t}(t_0) = \mathbf{r}'(t_0)/|\mathbf{r}'(t_0)|$, and the *tangent vector field* is the vector valued function $t \mapsto \mathbf{t}(t)$.

Observe that the regularity condition ensures that the instantaneous speed always is different from zero so we are able to divide by $|\mathbf{r}'|$ and define $\mathbf{t}$. When we have a vector field $\mathbf{v} : (a, b) \to \mathbb{R}^n$ along a curve $\mathbf{r}$ then we should think of the vector $\mathbf{v}(t)$ to be attached to the point $\alpha(t)$, see Figure 2.3. If $\mathbf{r}$ and $\mathbf{r}_1 = \mathbf{r} \circ f$

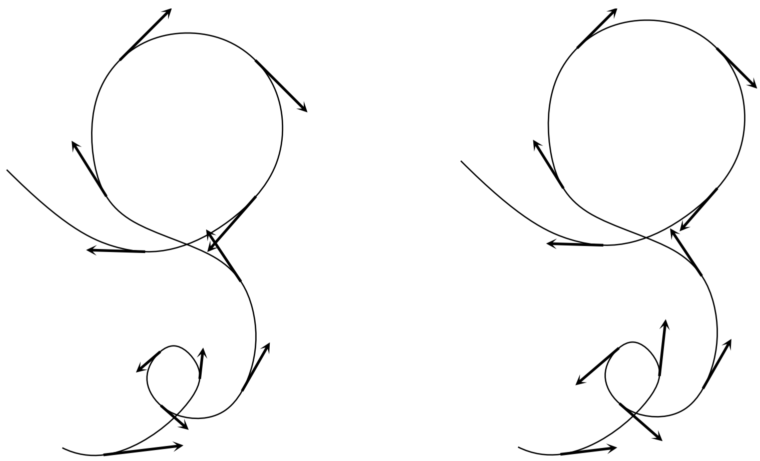


Figure 2.3: To the left the velocity vector field and to the right the tangent vector field.

are reparametrizations of each other then $\mathbf{r}_1'(t) = f'(t)\mathbf{r}'(f(t))$ so the velocity vector depends on the parametrization, but the tangent vectors satisfies $\mathbf{t}_1(t) = f'(t)/|f(t)|\,\mathbf{t}(f(t)) = \pm\mathbf{t}(f(t))$ so the tangent vector is a well defined property of an oriented curve, but is in general only defined up to a sign.

Definition 2.5. The straight line through a point $\mathbf{r}(t)$ on a regular curve parallel to the tangent vector is called the *tangent line* to the curve at $\mathbf{r}(t)$.

A more geometric way of defining the tangent line at a point $\mathbf{x}_0$ on a curve is as the limit position of a *secant*, i.e., a straight line through two points $\mathbf{x}_1 \neq \mathbf{x}_2$ on the curve when $\mathbf{x}_1, \mathbf{x}_2 \to \mathbf{x}_0$, see Figure 2.4.

That the limit position of such a secant indeed is the tangent is shown in Problem 2.2.7.

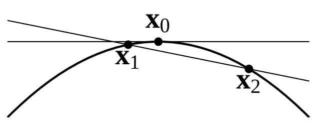


Figure 2.4: As $\mathbf{x}_1$ and $\mathbf{x}_2$ approaches $\mathbf{x}_0$ the secant approaches the tangent line.

The tangent to a regular curve given by $\mathbf{r} : I \to \mathbb{R}^n$ at the point $\mathbf{r}(t_0)$ can be parameterized as

$$u \mapsto \mathbf{r}(t_0) + u\mathbf{t}(t_0) \qquad \text{or} \qquad u \mapsto \mathbf{r}(t_0) + u\mathbf{r}'(t_0) \tag{2.1}$$

2.2.1 Length of curves

An *arc* of a curve given by $\mathbf{r} : I \to \mathbb{R}^n$ is a curve given by the restriction of $\mathbf{r}$ to a *closed* interval $[a, b] \subseteq I$. The points $\mathbf{r}(a)$ and $\mathbf{r}(b)$ are called the *end points* of the arc.

Definition 2.6. If $\mathbf{r} : I \to \mathbb{R}^n$ is a regular parametrization of a curve and $[a, b] \in I$ then the *length* of the arc $\mathbf{r}_{|[a,b]}$ is

$$\int_a^b |\mathbf{r}'(t)| \,\mathrm{d}t.$$

The following proposition shows that the length of an arc is independent of the parametrization.

Proposition 2.7. *Let $f : I_1 \to I$ be a reparametrization of a curve $\mathbf{r} : I \to \mathbb{R}^n$, and let $\mathbf{r}_1 = \mathbf{r} \circ f$. If $f([a_1, b_1]) = [a, b]$, then*

$$\int_a^b |\mathbf{r}'(t)| \,\mathrm{d}t = \int_{a_1}^{b_1} |\mathbf{r}_1'(u)| \,\mathrm{d}u.$$

The proof is left as Problem 2.2.8.

Definition 2.8. If $\mathbf{r} : I \to \mathbb{R}^n$ is a regular parametrization of a curve and $t_0 \in I$ then the *arc length* measured from t_0 is the function

$$s(t) = \int_{t_0}^{t} |\mathbf{r}'(\tau)| \,\mathrm{d}\tau, \quad t \in I. \tag{2.2}$$

If $t \geq t_0$, then $s \geq 0$ and is equal to the length of the arc between $\mathbf{r}(t_0)$ and $\mathbf{r}(t)$. If $t \leq t_0$, then $s \leq 0$ and is equal to minus the length of the arc between $\mathbf{r}(t_0)$ and $\mathbf{r}(t)$.

If $\mathbf{r}$ is of class C^k then the velocity $\mathbf{r}'$ is of class C^{k-1} and as the velocity never vanishes the speed $|\mathbf{r}'|$ is of class C^{k-1} too. It now follows that the arc length s is of class C^k and that $s'(t) = |\mathbf{r}'(t)| > 0$ for all $t \in I$. Hence $s = s(t)$ is an allowable change of parameter and we can use s as a parameter on the curve. This, of course, is an abuse of notation, s denotes both the function defined by (2.2) and a parameter, i.e, a real number. Similarly, we will denote the inverse function of $t \mapsto s(t)$ by $s \mapsto t(s)$ so t will also denotes both a function and a parameter. The reparametrization of $t \mapsto \mathbf{r}(t)$ by arc length, i.e., $s \mapsto \mathbf{r}(t(s))$ will be denoted by the same symbol $\mathbf{r}$, the advantage of this abuse of notation is that we now can write identities like

$$\left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right| = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right| \left|\frac{\mathrm{d}t}{\mathrm{d}s}\right| = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right| \Big/ \left|\frac{\mathrm{d}s}{\mathrm{d}t}\right| = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right| \Big/ \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right| = 1. \tag{2.3}$$

A parametrization by arc length is called a *natural parametrization*, or more precise

Definition 2.9. A parametrization $\mathbf{r} : I \to \mathbb{R}^n$ is called a *natural parametrization* if $|\mathbf{r}'(s)| = 1$ for all $s \in I$.

We now have

Proposition 2.10. *If* $\mathbf{r} : I \to \mathbb{R}^n$ *is a natural parametrization, then*

1. *The length of the arc between* $\mathbf{r}(s_1)$ *and* $\mathbf{r}(s_2)$ *is* $|s_2 - s_1|$.
2. *If* $s^* \mapsto \mathbf{r}^*(s^*)$ *is another natural parametrization, then* $s = \pm s^* + \text{constant}$.
3. *If* t *is an arbitrary parameter, then* $|\mathrm{d}s/\mathrm{d}t| = |\mathrm{d}\mathbf{r}/\mathrm{d}t|$.
4. *The tangent vector is* $\mathbf{t} = \mathrm{d}\mathbf{r}/\mathrm{d}s$.

Proof. The proof of 1 and 2 is left as Problems 2.2.9 and 2.2.10. Now 3 follows from (2.3). Finally $\mathbf{t} = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} \Big/ \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}\right| = \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s}$, which proves 4. □

A more geometric definition of arc length is in terms of approximating polygons. Let an arc be given by a parametrization $\mathbf{r}(t)$ with $t \in [a, b]$ and consider a partition

$$a = t_0 < t_1 < \cdots < t_m = b \tag{2.4}$$

of the interval $[a, b]$. This determines a sequence of points in $\mathbb{R}^n$

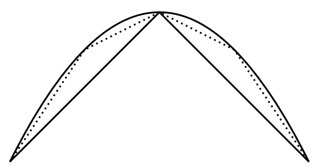


Figure 2.5: An approximating polygon P and a refinement P'.

$$\mathbf{x}_0 = \mathbf{r}(t_0), \quad \mathbf{x}_1 = \mathbf{r}(t_1), \quad \ldots \quad \mathbf{x}_m = \mathbf{r}(t_m).$$

The points form an *approximating polygon* P as shown in Figure 2.5 The length of P is clearly

$$\ell(P) = \sum_{i=1}^{m} |\mathbf{x}_i - \mathbf{x}_{i-1}| = \sum_{i=1}^{n} |\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})|$$

If the partition is refined to give a better polygonal approximating P', see Figure 2.5 then we clearly have $\ell(P') \geq \ell(P)$ so we are lead to consider the quantity

$$\ell = \sup\left\{ \sum_{i=1}^{m} |\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})| \;\middle|\; a = t_0 < t_1 < \cdots < t_m = b, m \in \mathbb{N} \right\}. \tag{2.5}$$

Observe that this makes sense even if $\mathbf{r}$ is only continuous, but we may have $\ell = \infty$.

Definition 2.11. The image of a map $\mathbf{r} : [a, b] \to \mathbb{R}^n$ is called *rectifiable* with length ℓ if

$$\ell = \sup\Big\{ \sum_{i=1}^{m} |\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})| \;\Big|\; a = t_0 < t_1 < \cdots < t_m = b, m \in \mathbb{N} \Big\} < \infty.$$

The following theorem shows that the two notions of arc length coincide.

Theorem 2.12. *Let* $\mathbf{r} : [a, b] \to \mathbb{R}^n$ *be of class* C^1*, then the image is a rectifiable arc with length*

$$\ell = \int_a^b |\mathbf{r}'(t)| \, \mathrm{d}t.$$

Proof. Let $\mathbf{r}(t) = \big(x_1(t), \dots, x_n(t)\big)$ and put $M = \max\big\{|\mathbf{r}'(t)| \;\big|\; t \in [a, b]\big\}$. If we have a partition (2.4) then

$$\begin{aligned}
\sum_{i=1}^{m} |\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})| &\le \sum_{i=1}^{m}\sum_{j=1}^{n} |x_j(t_i) - x_j(t_{i-1})| \\
&= \sum_{i=1}^{m}\sum_{j=1}^{n} |x_j'(\xi_{i,j})(t_i - t_{i-1})| \le \sum_{i=1}^{m}\sum_{j=1}^{n} |x_j'(\xi_{i,j})|(t_i - t_{i-1}) \\
&\le \sum_{i=1}^{m}\sum_{j=1}^{n} M(t_i - t_{i-1}) = \sum_{i=1}^{m} nM(t_i - t_{i-1}) = nM(b-a),
\end{aligned}$$

where $t_{i-1} < \xi_{i,j} < t_i$. So ℓ is finite and the arc is rectifiable.

Now consider an arbitrary $\epsilon > 0$. As $\mathbf{r}$ is of class C^1 we can find $\delta_1 > 0$ such that $|x_j'(t) - x_j'(t')| < \epsilon/\big(3n(b-a)\big)$, $j = 1, \dots, n$, if $|t - t'| < \delta_1$. Furthermore, we can find $\delta_2 > 0$ such that for a partition (2.4) with $t_i - t_{i-1} < \delta_2$ we have $\left|\int_a^b |\mathbf{r}'(t)| \, \mathrm{d}t - \sum_{i=1}^m |\mathbf{r}'(t_i)|(t_i - t_{i-1})\right| < \epsilon/3$. Now let $\delta = \min\{\delta_1, \delta_2\}$, and choose a partition such that the corresponding approximating polygon has a length that satisfies $0 \le \ell - \ell(P) < \epsilon/3$. If we refine the partition then the inequalities are still satisfied so we may assume that the partition has $t_i - t_{i-1} < \delta$. For such a partition we have

$$\begin{aligned}
\left|\ell - \int_a^b |\mathbf{r}'(t)| \, \mathrm{d}t\right| &\le |\ell - \ell(P)| + \left|\ell(P) - \int_a^b |\mathbf{r}'(t)| \, \mathrm{d}t\right| \\
&\le \frac{\epsilon}{3} + \left|\sum_{i=1}^{m} \big|\mathbf{r}(t_i) - \mathbf{r}(t_{i-1})\big| - \int_a^b |\mathbf{r}'(t)| \, \mathrm{d}t\right| \\
&= \frac{\epsilon}{3} + \left|\sum_{i=1}^{m} \left|\sum_{j=1}^{n} (x_j(t_i) - x_j(t_{i-1}))\mathbf{e}_j\right| - \int_a^b |\mathbf{r}'(t)| \, \mathrm{d}t\right|
\end{aligned}$$

where $\mathbf{e}_1, \dots, \mathbf{e}_n$ is the standard basis in $\mathbb{R}^n$, by the mean value theorem we get

$$= \frac{\epsilon}{3} + \left| \sum_{i=1}^{m} \left| \sum_{j=1}^{n} x_j'(\xi_{i,j})\mathbf{e}_j \right| (t_i - t_{i-1}) - \int_a^b |\mathbf{r}'(t)|\, \mathrm{d}t \right|$$

by adding and subtracting $\sum_i |\mathbf{r}(t_i)|(t_i - t_{i-1})$, we obtain

$$\le \frac{\epsilon}{3} + \left| \sum_{i=1}^{m} \left(\left| \sum_{j=1}^{n} x_j'(\xi_{i,j})\mathbf{e}_j \right| - \left| \sum_{j=1}^{n} x_j'(t_i)\mathbf{e}_j \right| \right) (t_i - t_{i-1}) \right|$$
$$+ \left| \sum_{i=1}^{n} |\mathbf{r}(t_i)|(t_i - t_{i-1}) - \int_a^b |\mathbf{r}'(t)|\, \mathrm{d}t \right|$$

as $\big||p| - |q|\big| \le |p - q|$ we have

$$\le \frac{\epsilon}{3} + \left| \sum_{i=1}^{m} \left| \sum_{j=1}^{n} \big(x_j'(\xi_{i,j}) - x_j'(t_i)\big)\mathbf{e}_j \right| (t_i - t_{i-1}) \right| + \frac{\epsilon}{3}$$
$$\le \frac{2\epsilon}{3} + \sum_{i=1}^{m} \sum_{j=1}^{n} \big|x_j'(\xi_{i,j}) - x_j'(t_i)\big|(t_i - t_{i-1})$$
$$< \frac{2\epsilon}{3} + \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\epsilon}{3n(b-a)}(t_i - t_{i-1}) = \epsilon$$

as ϵ is arbitrary we see that $\left|\ell - \int_a^b |\mathbf{r}'(t)|\, \mathrm{d}t\right| = 0.$ □

2.2.2 Contact

An important notion in geometry is the concept of contact between objects. First we need the distance between points and subsets.

Definition 2.13. Let p be a point in $\mathbb{R}^n$ and let A be a nonempty subset of $\mathbb{R}^n$. The *distance* between p and A is

$$d(p, A) = \inf\big\{|p - q| \;\big|\; q \in A\big\}.$$

Normally A will be a nice geometric object like a line, a plane, a circle, etc., but the definition makes sense for any subset of $\mathbb{R}^n$.

Example 2.3 Let L be the line parametrized as $t \mapsto \mathbf{x}_0 + t\mathbf{e}$ where $\mathbf{e}$ is a unit vector. Then

$$d(\mathbf{x}, L) = \sqrt{\mathbf{r}\cdot\mathbf{r} - (\mathbf{r}\cdot\mathbf{e})^2}, \quad \text{where } \mathbf{r} = \mathbf{x} - \mathbf{x}_0. \tag{2.6}$$

Let C be the circle parametrized as $t \mapsto \mathbf{x}_0 + r\cos t\mathbf{e}_1 + r\sin t\mathbf{e}_2$, where $\mathbf{e}_1, \mathbf{e}_2$ is an orthonormal pair of vectors. Then

$$d(\mathbf{x}, C) = \sqrt{\mathbf{r}\cdot\mathbf{r} + r^2 - 2r\sqrt{(\mathbf{r}\cdot\mathbf{e}_1)^2 + (\mathbf{r}\cdot\mathbf{e}_2)^2}} \quad \text{where } \mathbf{r} = \mathbf{x} - \mathbf{x}_0. \tag{2.7}$$

Definition 2.14. Let $\mathbf{r} : I \to \mathbb{R}^n$ be a regular parametrization of a curve and let A be a subset of $\mathbb{R}^n$. We say that the curve has *contact* with A of order k at $\mathbf{r}(t_0)$ if

$$\frac{d\big(\mathbf{r}(t), A\big)}{(t-t_0)^k} \to 0 \quad \text{for} \quad t \to t_0.$$

If $t = f(u)$ is a reparametrization with $f(u_0) = t_0$ then we have

$$\begin{aligned}\frac{d\big(\mathbf{r}(f(u)), A\big)}{(u-u_0)^k} &= \frac{d\big(\mathbf{r}(t), A\big)}{\big(f(u) - f(u_0)\big)^k}\frac{\big(f(u) - f(u_0)\big)^k}{(u-u_0)^k} \\ &= \frac{d\big(\mathbf{r}(t), A\big)}{(t-t_0)^k}\left(\frac{f(u) - f(u_0)}{u-u_0}\right)^k\end{aligned}$$

and as $\big(f(u) - f(u_0)\big) \big/ (u-u_0) \to f'(u_0) \neq 0$ for $u \to u_0$ we see that the notion of contact of order k is a property of the curve. It can in general be difficult to determine the order of contact, but when it comes to contact between curves the following theorem is helpful

Theorem 2.15. *Let $\mathbf{r}_1$ and $\mathbf{r}_2$ be natural parametrizations of two regular curves of class C^k. Suppose $\mathbf{r}_2(s) \neq \mathbf{r}_1(s_0)$ if $s \neq s_0$, then $\mathbf{r}_1$ has contact of order k with $\mathbf{r}_2$ at $\mathbf{r}_1(s_0)$ if and only if $\mathbf{r}_1^{(l)}(s_0) = \mathbf{r}_2^{(l)}(s_0)$ for all $l = 0, \ldots, k$.*

Proof. We only prove the 'if' part. As the Taylor expansions of $\mathbf{r}_1$ and $\mathbf{r}_2$ agree up to order k we have $\mathbf{r}_1(s-s_0) - \mathbf{r}_2(s-s_0) = \mathbf{o}\big((s-s_0)^k\big)$ and hence

$$\begin{aligned}\frac{d\big(\mathbf{r}_1(s), \mathbf{r}_2\big)}{(s-s_0)^k} &= \frac{\inf_{s_1}\big|\mathbf{r}_1(s) - \mathbf{r}_2(s_1)\big|}{(s-s_0)^k} \\ &\leq \frac{\big|\mathbf{r}_1(s) - \mathbf{r}_2(s)\big|}{(s-s_0)^k} = \frac{o\big((s-s_0)^k\big)}{(s-s_0)^k} \to 0 \quad \text{for } s \to s_0.\end{aligned}$$

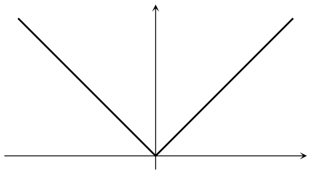
The 'only if' part is considerable more difficult, and we will return to a special case in Chapter 4 ☐

It can now be shown that the only straight line which has contact of order 1 with a curve at some point is the tangent line at that point, cf. Problem 2.2.16.

Problems

2.2.1 Show, that the vector function

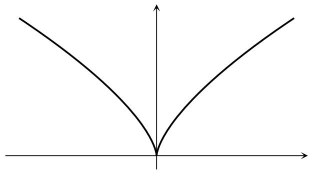
$$\mathbf{r}(t) = \begin{cases} (-e^{-1/t^2}, e^{-1/t^2}) & \text{for } t < 0 \\ (0,0) & \text{for } t = 0 \\ (e^{-1/t^2}, e^{-1/t^2}) & \text{for } t > 0 \end{cases}$$



is of class C^∞, and that $\mathbf{r}'(0) = 0$.

2.2.2 Show that the vector function

$$\mathbf{r}(t) = \left(t^3, t^2\right), \quad t \in \mathbb{R}$$



is of class C^∞, and that $\mathbf{r}'(0) = 0$.

2.2.3 Prove that a Bézier curve is either constant or piecewise a regular curve.

2.2.4 Prove that a B-spline curve is piecewise a constant or a regular curve.

2.2.5 We say that two vector functions $\mathbf{r}_i : I_i \to \mathbb{R}^n$, $i = 1, 2$, of class C^k are *equivalent* and write $\mathbf{r}_1 \sim \mathbf{r}_2$ if there exists an allowable change of parameter $f : I_2 \to I_1$ of class C^k such that $\mathbf{r}_2 = \mathbf{r}_1 \circ f$. Show that $\sim$ is an *equivalence relation*, i.e., that

(a) $\mathbf{r} \sim \mathbf{r}$.

(b) $\mathbf{r}_1 \sim \mathbf{r}_2 \Rightarrow \mathbf{r}_2 \sim \mathbf{r}_1$.

(c) $\mathbf{r}_1 \sim \mathbf{r}_2 \wedge \mathbf{r}_2 \sim \mathbf{r}_3 \Rightarrow \mathbf{r}_1 \sim \mathbf{r}_3$.

2.2.6 Show that if $\mathbf{r} : I \to \mathbb{R}^n$ is of class C^1 and $\mathbf{r}'(t) \neq \mathbf{0}$ for a $t \in I$ then there exists an $\epsilon > 0$ such that $\mathbf{r}_{|(t-\epsilon,t+\epsilon)\cap I}$ is injective.

2.2.7 Let $\mathbf{r} : I \to \mathbb{R}^n$ be a regular parametrization, and let $t_0 \in I$. Show that if $t_1, t_2 \in I$ are different and sufficiently close to t_0 then there is a well defined *secant* through $\mathbf{r}(t_1)$ and $\mathbf{r}(t_2)$. Show that if $t_1 < t_2$ and $t_1, t_2 \to t_0$ then the unit vector in the direction $\mathbf{r}(t_2) - \mathbf{r}(t_1)$ converges to the tangent vector $\mathbf{t}(t_0)$.

2.2.8 Prove Proposition 2.7, p. 45.

2.2.9 Prove 1 in Proposition 2.10, p. 46.

2.2.10 Prove 2 in Proposition 2.10, p 46.

2.2.11 Find the arc length of the *helix* in Example 2.2, and determine a natural parametrization.

2.2.12 Determine a parametrization of the tangent line to the parabola in Example 2.1 at an arbitrary point.

2.2.13 Determine a parametrization of the tangent line to the helix in Example 2.2 at an arbitrary point.

2.2.14 Prove (2.6), p. 49.

2.2.15 Prove (2.7), p. 49.

2.2.16 Prove that if a curve has contact of order 1 with a straight line then the line is the tangent line.

2.2.17 Let a regular curve be given by a parametrization $\mathbf{r}(t)$ defined on the interval $[a, b]$, let $s(t)$ be the arc length function, let $a = t_0 < t_1 < \cdots < t_m = b$ be a sequence of parameter values, and put $s_i = s(t_i)$ and $v_i = s'(t_i) = |\mathbf{r}'(t_i)|$.

(a) Show that the inverse function $t(s)$ satisfies $t_i = t(s_i)$ and $t'(s_i) = w_i = 1/v_i$.

(b) Show that there is a unique B-spline function $f : [s_0, s_m] \to [a, b]$ of degree 3 with knot vector

$$s_0, s_0, s_0, s_0, s_1, s_1, s_2, s_2, \ldots, s_{m-1}, s_{m-1}, s_m, s_m, s_m, s_m$$

such that $f(s_i) = t_i$, and $f'(s_i) = w_i$.

(c) Determine the control points for f (in this situation we consider a map into $\mathbb{R}$ so a control point is just a real number).

Exercises

2.2.1 Write a program that uses numerical integration to determine the arc length of a Bézier curve.

2.2.2 Write a program that for a given Bézier curve of degree n finds

(a) The length ℓ_p of the control polygon.

(b) The distance ℓ_c between the end points.

(c) The weighted average $\big(2\ell_c + (n-1)\ell_p\big)/(n+1)$.

Compare with the result of Exercise 2.2.1 and investigate what happens under repeated subdivision.

2.2.3 Write a program that uses numerical integration to determine the arc length of a B-spline curve.

2.2.4 Write a program that finds the length of a B-spline curve by writing it as a sequence of Bézier curves and finding the length of each of these.

2.2.5 Write a program that finds a C^1 approximation f to the inverse of the arc-length function for a Bézier curve by implementing the procedure outlined in Problem 2.2.17. Experiment with the program and investigate how many knots f needs in order to give a good approximation.

2.2.6 Write a program that finds a C^1 approximation f to the inverse of the arc-length function for a B-spline curve by implementing the procedure outlined in Problem 2.2.17. Experiment with the program and investigate how many knots f needs in order to give a good approximation.

2.3 Plane Curves

We now specialize to curves in the plane, i.e., we consider a regular parametrization $\mathbf{r} : I \to \mathbb{R}^2$. If $\mathbf{t}$ is the tangent vector at some point then the *normal vector* is the vector $\mathbf{n}$ such that $(\mathbf{t}, \mathbf{n})$ is a positively oriented orthonormal basis of $\mathbb{R}^2$, see Figure 2.6. Just like the tangent vector, the normal vector is an invariant concept associated with an oriented curve. It changes sign if the orientation is reversed.

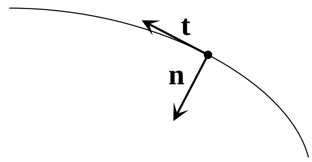


Figure 2.6: The tangent vector $\mathbf{t}$ and the normal vector $\mathbf{n}$ of a plane curve.

Now consider a natural parametrization $s \mapsto \mathbf{r}(s)$, then $\mathbf{t} = \mathbf{r}'$ and as the inner product $\mathbf{t} \cdot \mathbf{t} = 1$ is constant we have

$$0 = \frac{\mathrm{d}(\mathbf{t} \cdot \mathbf{t})}{\mathrm{d}s} = \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} \cdot \mathbf{t} + \mathbf{t} \cdot \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = 2\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} \cdot \mathbf{t} \tag{2.8}$$

i.e., $\mathrm{d}\mathbf{t}/\mathrm{d}s$ and $\mathbf{t}$ are orthogonal so $\mathrm{d}\mathbf{t}/\mathrm{d}s$ is proportional to the normal vector $\mathbf{n}$.

Definition 2.16. Let $s \mapsto \mathbf{r}(s)$ be a natural parametrization of class C^2 of a plane curve. The *plane curvature* of the curve at a point $\mathbf{r}(s_0)$ is

$$\kappa(s_0) = \mathbf{t}'(s_0) \cdot \mathbf{n}(s_0) \quad \text{and the } \textit{curvature vector} \text{ is} \quad \boldsymbol{\kappa} = \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \kappa\mathbf{n}.$$

If $\mathbf{t}'(s_0) \neq \mathbf{0}$ then the *radius of curvature* is $\rho(s_0) = 1/\kappa(s_0)$, the *center of curvature* is the point $\mathbf{c}(s_0) = \mathbf{r}(s_0) + \rho(s_0)\mathbf{n}(s_0) = \mathbf{r}(s_0) + \mathbf{t}'(s_0)/|\mathbf{t}'(s_0)|^2$, and the *circle of curvature* is the circle with center $\mathbf{c}(s_0)$ and radius $|\rho(s_0)|$.

If the orientation on a curve is reversed then both the tangent vector and the arc length changes sign, so the derivative $\mathrm{d}\mathbf{t}/\mathrm{d}s = \kappa\mathbf{n}$ is left unchanged and is thus a property of the curve. On the other hand $\mathbf{n}$ changes sign so κ and ρ changes sign too. All in all we have

Proposition 2.17. *For a regular plane curve we have that* $|\kappa|$, $|\rho|$, $\boldsymbol{\kappa} = \kappa\mathbf{n}$, $\rho\mathbf{n}$, *and the circle of curvature are invariant concepts associated with the curve. While* $\mathbf{t}$, $\mathbf{n}$, κ, *and* ρ *are concepts associated with an oriented curve and changes sign if the orientation is reversed. We furthermore have the* Frenet-Serret equations *for a plane curve:*

$$\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \kappa\mathbf{n}, \qquad \frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s} = -\kappa\mathbf{t}. \tag{2.9}$$

Proof. The only thing left to prove is the Frenet-Serret equations and we leave that as Problem 2.3.7. □

As the normal vector is to the left we easily see that if $\kappa > 0$ then the curve turns left, if $\kappa < 0$ then the curve turns right, and that if $\kappa = 0$ at some point and has different sign on each side of the point then the curve has an inflection point.

Example 2.4 Consider the circle with radius $r > 0$ given by the parametrization

$$\mathbf{r}(t) = (x_0 + r\cos t, y_0 + r\sin t).$$

We easily see that $\mathbf{r}'(t) = (-r\sin t, r\cos t)$, and $|\mathbf{r}'(t)| = r$, so the arc length measured from $t = 0$ is $s = \int_0^t r\,\mathrm{d}\tau = rt$. I.e., $t = s/r$ and we obtain a natural parametrization by $s \mapsto \mathbf{r}(s/r) = \big(x_0 + r\cos(s/r), y_0 + r\sin(s/r)\big)$. The tangent vector is $\mathbf{t} = \mathrm{d}\mathbf{r}/\mathrm{d}s = \big(-\sin(s/r), \cos(s/r)\big)$ and the normal vector is $\mathbf{n} = \big(-\cos(s/r), -\sin(s/r)\big)$ and $\kappa\mathbf{n} = \mathrm{d}\mathbf{t}/\mathrm{d}s = 1/r\big(-\cos(s/r), -\sin(s/r)\big) = \mathbf{n}/r$. From this we see that the curvature is constant $\kappa = 1/r$, the radius of curvature is $\rho = r$ and the circle of curvature is the circle itself.

A more geometric way of defining the circle of curvature at a point $\mathbf{x}_0$ on a curve is as the limit position of a circle through three distinct points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ on the curve as $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \to \mathbf{x}_0$, see Figure 2.7, and Problem 2.3.1.

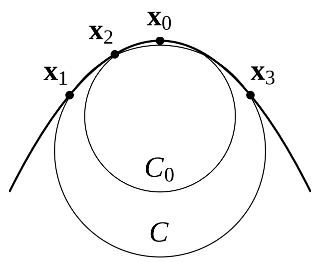


Figure 2.7: If $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \to \mathbf{x}_0$ then $C \to C_0$.

It can also be shown that the circle of curvature is the only circle that has contact of order 2 with the curve, cf. Problem 2.3.2

Given a natural parametrization it is a simple matter to determine the curvature. It is in general impossible to determine a natural parametrization, but the following theorem tells how to calculate the curvature from an arbitrary regular parametrization.

Theorem 2.18. *Let $t \mapsto \mathbf{r}(t) = \big(x(t), y(t)\big)$ be a regular parametrization of class C^2. The curvature is then given by*

$$\kappa(t) = \frac{\big[\mathbf{r}'(t), \mathbf{r}''(t)\big]}{\big|\mathbf{r}'(t)\big|^3} = \frac{\begin{vmatrix} x'(t) & x''(t) \\ y'(t) & y''(t) \end{vmatrix}}{\big(x'(t)^2 + y'(t)^2\big)^{3/2}},$$

Proof. Let s denotes the arc length of the curve. We then have

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \frac{\mathrm{d}s}{\mathrm{d}t}\mathbf{t},$$
$$\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\mathbf{t} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\mathbf{t} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2\kappa\mathbf{n}.$$

Hence

$$\left[\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}, \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right] = \left[\frac{\mathrm{d}s}{\mathrm{d}t}\mathbf{t}, \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\mathbf{t} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \kappa\mathbf{n}\right]$$
$$= \frac{\mathrm{d}s}{\mathrm{d}t}\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 [\mathbf{t},\mathbf{t}] + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \kappa[\mathbf{t},\mathbf{n}] = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|^3 \kappa. \qquad \square$$

The curvature is often used to assess the quality of a curve, either in the form of a *curvature plot* or a *porcupine plot*, see Figure 2.8. In a porcupine plot the curve

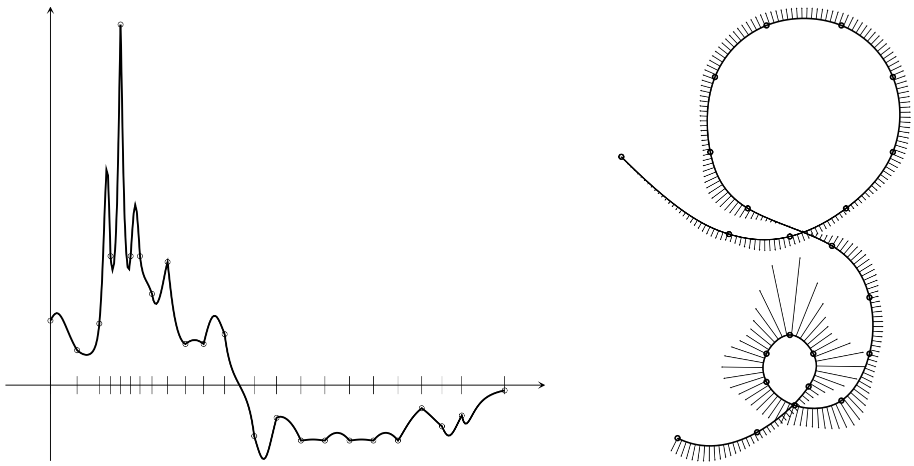


Figure 2.8: To the left a curvature plot and to the right a porcupine plot of a cubic B-spline curve. The endpoints of Bézier segments are indicated on both plots.

is plotted along with the vector field '$-$scale $\times$ $\kappa\mathbf{n}$'. A designer normally wants a slowly varying curvature plot without unnecessary undulations, so the curve above would not be satisfactory. The designer would then change the curve slightly either by changing the control points manually, or by an automatic procedure, eg. by minimizing $\int (\mathrm{d}\kappa/\mathrm{d}s)^2\,\mathrm{d}s$, under the side condition that the control points are only allowed to move a certain distance. This process is called *fairing* and the goal is to obtain a *fair* curve.

As the tangent vector is a unit vector we can write it as $\mathbf{t} = (\cos\phi, \sin\phi)$ where ϕ is an angle determined up to a multiple of 2π, see Figure 2.9. If the tangent vector field is a continuous vector function $u \mapsto \mathbf{t}(u)$, then it is not hard to see that it is possible to make a continuous choice $u \mapsto \phi(u)$ of this angle. Such a choice is unique up to a (constant) multiple of 2π and it has the same degree of differentiability as $\mathbf{t}$.

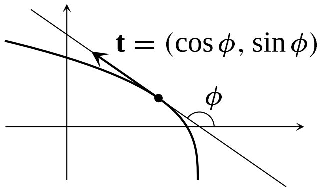


Figure 2.9: The tangent direction is the angle ϕ between the tangent and the x-axis.

Definition 2.19. Let $\mathbf{r} : I \to \mathbb{R}^2$ be a regular parametrization of class C^k. The *tangent direction* is a continuous choice of ϕ such that $\mathbf{t}(u) = \big(\cos\phi(u), \sin\phi(u)\big)$.

We immediately have the following results.

Proposition 2.20. *The tangent direction ϕ is a property of an oriented plane curve, but if the orientation is reversed then $\phi \mapsto \phi + \pi$. If κ is the curvature then $\mathrm{d}\phi/\mathrm{d}s = \kappa$. Furthermore, if $\kappa \neq 0$, then ϕ can be used as a parameter and $\mathrm{d}s/\mathrm{d}\phi = \rho$, where $\rho = 1/\kappa$ is the radius of curvature.*

Proof. As the tangent direction is a property of an oriented curve the same is true for the tangent direction. If the orientation is reversed $\mathbf{t}$ changes sign and that corresponds to adding π to the tangent direction. On one hand we have $\mathrm{d}\mathbf{t}/\mathrm{d}s = \kappa\,\mathbf{n}$ and on the other hand we have $\mathrm{d}(\cos\phi, \sin\phi)/\mathrm{d}s = \mathrm{d}\phi/\mathrm{d}s(-\sin\phi, \cos\phi) = \mathrm{d}\phi/\mathrm{d}s\,\mathbf{n}$, so $\mathrm{d}\phi/\mathrm{d}s = \kappa$. If $\kappa \neq 0$, then ϕ is a monotone function of s, so the inverse function exists and is differentiable with derivative $\mathrm{d}s/\mathrm{d}\phi = 1/\big(\mathrm{d}\phi/\mathrm{d}s\big) = 1/\kappa = \rho$. ☐

We can now prove that the curvature determines a plane curve uniquely up to a Euclidean motion, i.e., up to a rotation and a translation. We formulate it as the following theorem.

Theorem 2.21. *Let $\kappa : I \to \mathbb{R}$ be a continuous function. Then there exists a natural parametrization $\mathbf{r} : I \to \mathbb{R}^2$ of class C^2 such that κ is the curvature of $\mathbf{r}$. Furthermore, the curve is determined uniquely up to a Euclidean motion.*

Proof. Assume $\mathbf{r}$ is a curve with curvature κ and tangent direction ϕ. Then $\mathrm{d}\phi/\mathrm{d}s = \kappa$ so $\phi(s) = \phi_0 + \int_{s_0}^{s} \kappa(\tau)\,\mathrm{d}\tau$. The tangent vector is now $\mathbf{t}(s) = \big(\cos\phi(s), \sin\phi(s)\big)$ and as $\mathrm{d}\mathbf{r}/\mathrm{d}s = \mathbf{t}$ we must have $\mathbf{x}(s) = \mathbf{x}_0 + \int_{s_0}^{s} \mathbf{t}(\tau)\,\mathrm{d}\tau$. Different choices of ϕ_0 corresponds to rotations and different choices of $\mathbf{x}_0$ corresponds to translations. All that remains is to show that the curvature of $\mathbf{r}$ is κ, but by construction we have $\mathbf{t} = \mathrm{d}\mathbf{r}/\mathrm{d}s$ and $\mathrm{d}\mathbf{t}/\mathrm{d}s = \kappa\mathbf{n}$ which shows that κ indeed is the curvature of $\mathbf{r}$. ☐

By inspection of the proof above we realize that the function $s \mapsto \phi(s)$ determines the curve up to a translation. The equation $\phi = \phi(s)$ is called an *intrinsic equation* of the curve, but the equations $s = s(\phi)$ or $\mathrm{d}\phi/\mathrm{d}s = \kappa$ also determine the curve up to a translation and a Euclidean motion respectively. In fact any equation, including differential equations, that links the arc length and the tangent direction is called an *intrinsic equation* of the curve. In [14] the intrinsic equation $\mathrm{d}s/\mathrm{d}\phi = \rho$ was instrumental for the design of scroll compressors.

If $\mathbf{r}$ is a natural parametrization of a plane curve, and we put $\mathbf{r}_0 = \mathbf{r}(s_0)$, and let $\mathbf{t}$, $\mathbf{n}$, and κ be the tangent vector, the normal vector, and the curvature at s_0, respectively, then the Taylor expansion of $\mathbf{r}$ to second order at s_0 is

$$\mathbf{r}(s) = \mathbf{r}_0 + (s - s_0)\mathbf{t} + \frac{1}{2}(s - s_0)^2\kappa\mathbf{n} + \mathbf{o}\big((s - s_0)^2\big). \tag{2.10}$$

This expression is called the *canonical form* of a plane curve. It follows from Theorem 2.15 that every plane curve at a point with non vanishing curvature has second order contact with a unique parabola, cf. Problem 2.3.4. We also see that any plane curve is locally the graph of a function from the tangent line to the normal line, see Figure 2.10, and Problem 2.3.10.

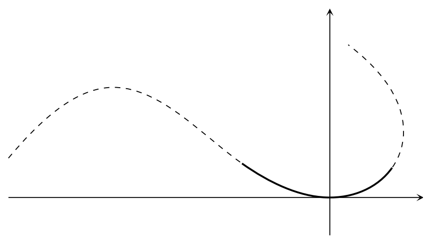


Figure 2.10: A plane curve is locally the graph of function "from" the tangent line "to" the normal line.

We end this section by stating some properties of *involutes* and *evolutes*, all proofs are left as exercises.

Definition 2.22. Let $\mathbf{r} : I \to \mathbb{R}^2$ be a regular parametrization of a regular curve with tangent $\mathbf{t}$, arc length s, and radius of curvature ρ. An *involute* of $\mathbf{r}$ is a curve given by

$$\mathbf{r}^\sharp(t) = \mathbf{r}(t) + (c - s(t))\mathbf{t}(t), \tag{2.11}$$

for a $c \in \mathbb{R}$. The *evolute* of $\mathbf{r}$ is the curve given by

$$\mathbf{r}^\flat(t) = \mathbf{r}(t) + \rho(t)\mathbf{n}(t), \tag{2.12}$$

see Figure 2.11

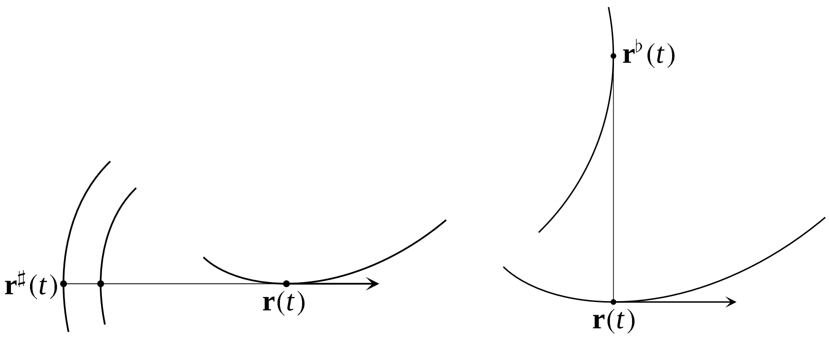


Figure 2.11: To the left two involutes of a curve and to the right the evolute.

Different choices of the constant c in (2.11) lead to *parallel curves* and the distance between the curves is exactly the difference between the two constants, see Figure 2.11 and Problem 2.3.12. Furthermore the two constructions are the inverse of each other in the sense that the evolute of one of the involutes gives the original curve back, cf. Problem 2.3.14, while a curve is itself one of the involutes of its evolute, cf. Problem 2.3.15. An involute to an evolute is parallel to the original curve and any parallel curve is obtained this way. For other properties of involutes and evolutes, cf. Problems 2.3.11–2.3.16.

Problems

2.3.1 Let $\mathbf{r} : I \to \mathbb{R}^2$ be a natural parametrization of a regular curve. Let $\mathbf{t} : I \to \mathbb{R}^2$ be the tangent vector field and assume that $\mathbf{t}(s_0) \neq \mathbf{0}$. Show that if $s_1 < s_2 < s_3$ are sufficiently close to s_0 then there is a well defined circle through $\mathbf{r}(s_1)$, $\mathbf{r}(s_2)$, $\mathbf{r}(s_3)$. Show that if $s_1, s_2, s_3 \to s_0$ then the centre and radius converges to the centre of curvature and the absolute value of the radius of curvature respectively.

2.3.2 Show that if $\kappa(s_0) \neq 0$, then the curve has contact of order 2 with the circle of curvature at the point $\mathbf{r}(s_0)$. Show that the contact with any other circle is of lower order.

2.3.3 Show that if $\kappa(s_0) = 0$ then the curve has contact of order 2 with the tangent line at the point $\mathbf{r}(s_0)$.

2.3.4 Show that if $\kappa(s_0) \neq 0$ then the curve has contact of order 2 with a unique parabola.

2.3.5 Let $t \mapsto \mathbf{r}(t) = \big(x(t), y(t)\big)$ be a regular parametrization of class C^2. Show that the curvature vector is given by

$$\boldsymbol{\kappa}(t) = \kappa(t)\mathbf{n}(t) = \frac{\begin{vmatrix} x'(t) & x''(t) \\ y'(t) & y''(t) \end{vmatrix}}{\big(x'(t)^2 + y'(t)^2\big)^2}\big(-y'(t), x'(t)\big).$$

2.3.6 Find the curvature of the parabola in Example 2.1, p. 42 at an arbitrary point.

2.3.7 Prove the Frenet-Serret equations for a plane curve, cf. (2.9), p. 52.

2.3.8 Show that if the curvature of a plane regular curve is zero then the curve is a straight line.

2.3.9 Show that if a plane regular curve has constant curvature different from zero then the curve is a circle.

2.3.10 Show that a regular plane curve locally is the graph of a function "from" the tangent line "to" the normal line, cf. Figure 2.10, p. 56.

2.3.11 Let $\mathbf{r} : I \to \mathbb{R}^2$ be a natural parametrization of a regular curve and consider the involute $\mathbf{r}^\sharp(s) = \mathbf{r}(s) + (c - s)\mathbf{t}(s)$. Determine $\mathbf{r}^{\sharp\prime}$. For which values of s is $\mathbf{r}^{\sharp\prime}(s) \neq \mathbf{0}$? Determine the tangent vector, the normal vector, and the curvature of $\mathbf{r}^\sharp$.

2.3.12 Let $\mathbf{r} : I \to \mathbb{R}^2$ be a natural parametrization of a regular curve. Let $\mathbf{r}_i^\sharp(s) = \mathbf{r}(s) + (c_i - s)\mathbf{t}(s)$, $i = 1, 2$, be two different involutes of $\mathbf{r}$. Show that a tangent line of $\mathbf{r}$ is a normal line of both $\mathbf{r}_1^\sharp$ and $\mathbf{r}_2^\sharp$, and that $\mathbf{r}_1^\sharp - \mathbf{r}_2^\sharp = (c_2 - c_1)\mathbf{n}^\sharp$, where $\mathbf{n}^\sharp$ is the *common* normal of $\mathbf{r}_1^\sharp$ and $\mathbf{r}_2^\sharp$.

2.3.13 Let $\mathbf{r} : I \to \mathbb{R}^2$ be a natural parametrization of a regular curve and let $\mathbf{t}$, $\mathbf{n}$, and ρ be the tangent vector, the normal vector, and the radius of curvature respectively. Consider the involute $\mathbf{r}^\flat(s) = \mathbf{r}(s) + \rho(s)\mathbf{n}(s)$. Determine $\mathbf{r}^{\flat\prime}$. For which values of s is $\mathbf{r}^{\flat\prime}(s) \neq \mathbf{0}$? Determine the tangent vector, the normal vector, and the curvature of $\mathbf{r}^\flat$. Show that a normal line of $\mathbf{r}$ is a tangent line of $\mathbf{r}^\flat$.

2.3.14 Show that a curve is the evolute of any one of it's involutes.

2.3.15 Show that the a curve is an involute of it's evolute.

2.3.16 Consider a regular curve with non vanishing curvature and let ϕ be the tangent direction. Consider the intrinsic equation $\mathrm{d}s/\mathrm{d}\phi = \rho(\phi)$ where ρ is the radius of curvature. Show that the radius of curvature for the evolute is $\mathrm{d}\rho/\mathrm{d}\phi$. What is the radius of curvature for an involute? Find the intrinsic equation for the evolute and the involutes.

Exercises

2.3.1 Write a program that finds the curvature and/or the curvature vector at an arbitrary point of a plane Bézier curve.

2.3.2 Write a program that finds the curvature and/or the curvature vector at an arbitrary point of a plane B-spline curve.

2.3.3 Write a program that plots the curvature as a function of arc length for a plane Bézier curve.

2.3.4 Write a program that plots the curvature as a function of arc length for a plane B-spline curve.

2.3.5 Write a program that makes a porcupine plot of a plane Bézier curve.

2.3.6 Write a program that makes a porcupine plot of a plane B-spline curve.

2.3.7 Write a program that plots a Bézier curve and it's evolute.

2.3.8 Write a program that plots a Bézier curve and an arbitrary one of it's involutes.

2.3.9 Write a program that plots a B-spline curve and it's evolute.

2.3.10 Write a program that plots a B-spline curve and an arbitrary one of it's involutes.

2.3.11 Write a program that plots a curve with the intrinsic equation $\mathrm{d}s/\mathrm{d}\phi = \rho(\phi)$ in the case where $\rho(\phi)$ is a polynomial.

2.4 Space Curves

We now consider curves in space, i.e., we have a regular parametrization $\mathbf{r} : I \to \mathbb{R}^3$. The curvature vector for a plane curve is $\kappa\mathbf{n} = \mathrm{d}\mathbf{t}/\mathrm{d}s$ so the first part of the following definition is natural.

Definition 2.23. Let $s \to \mathbf{r}(s)$ be a natural parametrization of class C^3 of a space curve. The *normal plane* of the curve at a point $\mathbf{r}(s)$ is the plane through $\mathbf{r}(s)$ orthogonal to the tangent vector. The *curvature vector* is $\boldsymbol{\kappa}(s) = \mathbf{t}'(s)$, and the *curvature* is $\kappa(s) = |\boldsymbol{\kappa}(s)| = |\mathbf{t}'(s)|$.

If $\kappa(s) \neq 0$, then the *radius of curvature* is $\rho(s) = 1/\kappa(s)$, the *principal normal vector* is $\mathbf{n}(s) = \boldsymbol{\kappa}(s)/\kappa(s) = \mathbf{t}'(s)/|\mathbf{t}'(s)|$, the centre of curvature is $\mathbf{c}(s) = \mathbf{r}(s) + \rho(s)\mathbf{n}(s)$, and the *circle of curvature* is the circle with centre $\mathbf{c}(s)$ and radius $\rho(s)$. The *binormal vector* is $\mathbf{b}(s) = \mathbf{t}(s) \times \mathbf{n}(s)$, and the *torsion* is $\tau(s) = -\mathbf{b}'(s) \cdot \mathbf{n}(s)$. The *osculating plane* is the plane through $\mathbf{r}(s)$ orthogonal to the binormal vector and the *rectifying plane* is the plane through $\mathbf{r}(s)$ orthogonal to the principal normal vector.

Notice that the curvature of a space curve is nonnegative, and that the two normal vectors only are defined if $\kappa \neq 0$. The calculation (2.8) on page 52 also holds in $\mathbb{R}^3$ so $\mathbf{t}$ and $\mathbf{n}$ are orthogonal and $\mathbf{t}$, $\mathbf{n}$, $\mathbf{b}$ is a positively oriented orthonormal frame called the *Frenet-Serret frame*. If the orientation on a curve is reversed then both the tangent vector and the arc length changes sign, so the derivative $\mathrm{d}\mathbf{t}/\mathrm{d}s = \kappa\mathbf{n}$ is left unchanged and is an invariant property of the curve, as is κ, ρ, and $\mathbf{n}$. On the other hand $\mathbf{t}$ changes sign so $\mathbf{b}$ and τ change sign too. All in all we have

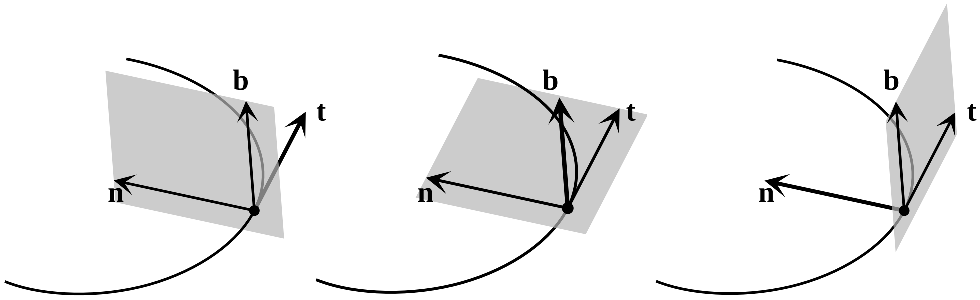


Figure 2.12: The Frenet-Serret frame at a point of a curve. The normal plane is spanned by **n** and **b**, the osculating plane is spanned by **t** and **n**, and the rectifying plane is spanned by **t** and **b**.

Proposition 2.24. *For a regular space curve we have that κ, ρ, $\boldsymbol{\kappa} = \kappa\mathbf{n}$, $\mathbf{n}$, and the circle of curvature are invariant concepts associated with the curve. And* **t**, **b**, *and τ are invariant concepts associated with the oriented curve. They change sign if the orientation is reversed.*

Just as for plane curves the circle of curvature at a point $\mathbf{x}_0$ on a space curve can be defined as the limit of a circle through three distinct points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ on the curve as $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \to \mathbf{x}_0$, see Figure 2.7. Likewise, the osculating plane at a point $\mathbf{x}_0$ on a space curve can be defined as the limit position of a plane through three distinct points $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ on the curve as $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3 \to \mathbf{x}_0$.

It can also be shown that the circle of curvature is the only circle that has contact of order 2 with the curve, and the osculating plane is the only plane that has contact of order 2 with the curve.

The derivative of **t**, **n**, and **b** are given by the *Frenet-Serret equations* (2.13) in the following theorem.

Theorem 2.25. *Let $s \mapsto \mathbf{r}(s)$ be a natural parametrization of a space curve with non vanishing curvature $\kappa(s) \neq 0$, torsion $\tau(s)$ and Frenet-Serret frame $\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)$. Then*

$$\begin{bmatrix} \mathbf{t}'(s) \\ \mathbf{n}'(s) \\ \mathbf{b}'(s) \end{bmatrix} = \begin{bmatrix} 0 & \kappa(s) & 0 \\ -\kappa(s) & 0 & \tau(s) \\ 0 & -\tau(s) & 0 \end{bmatrix} \begin{bmatrix} \mathbf{t}(s) \\ \mathbf{n}(s) \\ \mathbf{b}(s) \end{bmatrix} \tag{2.13}$$

Proof. The equation $\mathbf{t}'(s) = \kappa(s)\,\mathbf{n}(s)$ is the equation that defines κ and **n**. As **t**, **n**, **b** is an orthonormal frame we have

$$\mathbf{b}' = (\mathbf{b}' \cdot \mathbf{t})\mathbf{t} + (\mathbf{b}' \cdot \mathbf{n})\mathbf{n} + (\mathbf{b}' \cdot \mathbf{b})\mathbf{b}$$

As $|\mathbf{b}(s)|$ is constant a calculation like (2.8), p. 52 shows that $\mathbf{b}' \cdot \mathbf{b} = 0$. Similar, $\mathbf{t}(s) \cdot \mathbf{b}(s) = 0$ is constant too, so

$$0 = \frac{\mathrm{d}(\mathbf{t} \cdot \mathbf{b})}{\mathrm{d}s} = \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} \cdot \mathbf{b} + \mathbf{t} \cdot \frac{\mathrm{d}\mathbf{b}}{\mathrm{d}s} \tag{2.14}$$

and we see that $\mathbf{b}' \cdot \mathbf{t} = -\mathbf{b} \cdot \mathbf{t}' = -\kappa\, \mathbf{b} \cdot \mathbf{n} = 0$. The definition of τ tells us that $\mathbf{b}' \cdot \mathbf{n} = -\tau$ so all in all we have the equation $\mathbf{b}'(s) = -\tau(s)\, \mathbf{n}(s)$. Finally, calculations like (2.8) and (2.14) shows that

$$\begin{aligned} \mathbf{n}'(s) &= \big(\mathbf{n}'(s) \cdot \mathbf{t}(s)\big)\mathbf{t}(s) + \big(\mathbf{n}'(s) \cdot \mathbf{n}(s)\big)\mathbf{n}(s) + \big(\mathbf{n}'(s) \cdot \mathbf{b}(s)\big)\mathbf{b}(s) \\ &= -\big(\mathbf{t}'(s) \cdot \mathbf{n}(s)\big)\mathbf{t}(s) + 0\,\mathbf{n}(s) - \big(\mathbf{b}'(s) \cdot \mathbf{n}(s)\big)\mathbf{b}(s) \\ &= -\kappa(s)\,\mathbf{t}(s) + \tau(s)\mathbf{b}(s). \end{aligned}$$

□

As $\mathbf{t}' = \kappa\mathbf{n}$ and $\mathbf{b}' = -\tau\mathbf{n}$ we see that the curvature is a measure for how fast the tangent line turns around $\mathbf{b}$, and the torsion is a measure for how fast the osculating plane turns around $\mathbf{t}$.

For practical calculations we need a formula that expresses the curvature, torsion and Frenet-Serret frame in terms of an arbitrary parametrization, this is the content of the next theorem.

Theorem 2.26. *Let $t \mapsto \mathbf{r}(t) = \big(x(t), y(t), z(t)\big)$ be a regular parametrization of class C^3. The curvature is then given by*

$$\kappa(t) = \frac{\big|\mathbf{r}'(t) \times \mathbf{r}''(t)\big|}{\big|\mathbf{r}'(t)\big|^3}. \tag{2.15}$$

The torsion is given by

$$\tau(t) = \frac{\big[\mathbf{r}'(t), \mathbf{r}''(t), \mathbf{r}'''(t)\big]}{\big|\mathbf{r}'(t) \times \mathbf{r}''(t)\big|^2}. \tag{2.16}$$

The binormal vector is given by

$$\mathbf{b}(t) = \frac{\mathbf{r}'(t) \times \mathbf{r}''(t)}{\big|\mathbf{r}'(t) \times \mathbf{r}''(t)\big|} \tag{2.17}$$

The principal normal vector is given by

$$\mathbf{n}(t) = \mathbf{b}(t) \times \mathbf{t}(t). \tag{2.18}$$

Proof. Let s denotes the arc length of the curve. We then have

$$\begin{aligned}
\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} &= \frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}s} = \frac{\mathrm{d}s}{\mathrm{d}t}\mathbf{t}, \\
\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} &= \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\mathbf{t} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\mathbf{t} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \kappa\mathbf{n}. \\
\frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}t^3} &= \frac{\mathrm{d}^3 s}{\mathrm{d}t^3}\mathbf{t} + \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\frac{\mathrm{d}s}{\mathrm{d}t}\frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} + 2\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\frac{\mathrm{d}s}{\mathrm{d}t}\kappa\mathbf{n} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \frac{\mathrm{d}\kappa}{\mathrm{d}s}\mathbf{n} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \kappa\frac{\mathrm{d}\mathbf{n}}{\mathrm{d}s} \\
&= \left(\frac{\mathrm{d}^3 s}{\mathrm{d}t^3} - \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \kappa^2\right)\mathbf{t} + \left(3\frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\frac{\mathrm{d}s}{\mathrm{d}t}\kappa + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \frac{\mathrm{d}\kappa}{\mathrm{d}s}\right)\mathbf{n} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \kappa\tau\mathbf{b}.
\end{aligned}$$

Hence

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \times \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2} = \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \kappa\mathbf{t} \times \mathbf{n} = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}\right|^3 \kappa\mathbf{b}$$

which implies (2.15) and (2.17). We also have

$$\begin{aligned}
\left[\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t}, \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}, \frac{\mathrm{d}^3\mathbf{r}}{\mathrm{d}t^3}\right] &= \left[\frac{\mathrm{d}s}{\mathrm{d}t}\mathbf{t}, \frac{\mathrm{d}^2 s}{\mathrm{d}t^2}\mathbf{t} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \kappa\mathbf{n}, A\,\mathbf{t} + B\,\mathbf{n} + \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^3 \kappa\tau\mathbf{b}\right] \\
&= \left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^6 \kappa^2\tau[\mathbf{t}, \mathbf{n}, \mathbf{b}] = \left|\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} \times \frac{\mathrm{d}^2\mathbf{r}}{\mathrm{d}t^2}\right|^2 \tau
\end{aligned}$$

which implies (2.16). Finally, (2.18) simply follows from the fact that $\mathbf{t}, \mathbf{n}, \mathbf{b}$ is a positively oriented orthonormal basis. □

Just as in the case of a plane curve a designer will often use a curvature plot or a porcupine plot, see Figure 2.8 to asses the quality of a curve. And in an automatic fairing procedure it is again usually the integral $\int (\mathrm{d}\kappa/\mathrm{d}s)^2\,\mathrm{d}s$ that is minimized, under some suitable side conditions. One may (and should?) take the torsion into acount too, but there is no universally accepted way of doing this.

Just as the plane curvature determines a plane curve up to a Euclidean motion, the curvature and torsion determine a space curve up to a Euclidean motion.

Theorem 2.27. *Let I be an interval, let $\kappa : I \to \mathbb{R}$ be a strictly positive C^1 function and let $\tau : I \to \mathbb{R}$ be a C^0 function. Let furthermore $s_0 \in I$, let $\mathbf{x}_0$ be a fixed point of $\mathbb{R}^3$ and let $\mathbf{t}_0, \mathbf{n}_0, \mathbf{b}_0$ be fixed positively oriented orthonormal basis of $\mathbb{R}^3$. Then there exists a unique regular natural parametrization $\mathbf{r} : I \to \mathbb{R}^3$ of class C^3 such that the curvature is κ, the torsion is τ, $\mathbf{r}(s_0) = \mathbf{x}_0$, and the Frenet-Serret frame $(\mathbf{t}, \mathbf{n}, \mathbf{b})$ satisfies $\mathbf{t}(s_0) = \mathbf{t}_0$, $\mathbf{n}(s_0) = \mathbf{n}_0$, and $\mathbf{b}(s_0) = \mathbf{b}_0$.*

Proof. The Frenet-Serret equations (2.13) is a *linear* system of ordinary differential euations (in $\mathbb{R}^9$). It follows that there is unique solution $\mathbf{t}, \mathbf{n}, \mathbf{b}$ defined on all of I, with $\mathbf{t}(s_0) = \mathbf{t}_0$, $\mathbf{n}(s_0) = \mathbf{n}_0$, and $\mathbf{b}(s_0) = \mathbf{b}_0$. The set $\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)$ is a positively oriented orthonormal frame for $s = s_0$, we want to show that it is a positively oriented orthonormal frame for all $s \in I$. To that end we define six functions $f_i : I \to \mathbb{R}$ by

$$\begin{aligned} f_1(s) &= \mathbf{t}(s)\cdot\mathbf{t}(s) & f_2(s) &= \mathbf{t}(s)\cdot\mathbf{n}(s) & f_3(s) &= \mathbf{t}(s)\cdot\mathbf{b}(s) \\ f_4(s) &= \mathbf{n}(s)\cdot\mathbf{n}(s) & f_5(s) &= \mathbf{n}(s)\cdot\mathbf{b}(s) & f_6(s) &= \mathbf{b}(s)\cdot\mathbf{b}(s) \end{aligned}$$

As $\mathbf{t}, \mathbf{n}, \mathbf{b}$ is a solution to the Frenet-Serret equations (2.13) we have

$$\begin{aligned} f_1' &= 2\mathbf{t}'\cdot\mathbf{t} = 2\kappa\,\mathbf{n}\cdot\mathbf{t} = 2\kappa\ f_2 \\ f_2' &= \mathbf{t}'\cdot\mathbf{n} + \mathbf{t}\cdot\mathbf{n}' = \kappa\,\mathbf{n}\cdot\mathbf{n} - \kappa\,\mathbf{t}\cdot\mathbf{t} + \tau\,\mathbf{t}\cdot\mathbf{b} = -\kappa\ f_1 + \tau\ f_3 + \kappa\ f_4 \\ f_3' &= \mathbf{t}'\cdot\mathbf{b} + \mathbf{t}\cdot\mathbf{b}' = \kappa\,\mathbf{n}\cdot\mathbf{b} - \tau\,\mathbf{t}\cdot\mathbf{n} = -\tau\ f_2 + \kappa\ f_5 \\ f_4' &= 2\mathbf{n}\cdot\mathbf{n}' = -2\kappa\,\mathbf{n}\cdot\mathbf{t} + 2\tau\,\mathbf{n}\cdot\mathbf{b} = -2\kappa\ f_2 + 2\tau\ f_5 \\ f_5' &= \mathbf{n}'\cdot\mathbf{b} + \mathbf{n}\cdot\mathbf{b}' = -\kappa\,\mathbf{t}\cdot\mathbf{b} + \tau\,\mathbf{b}\cdot\mathbf{b} - \tau\,\mathbf{n}\cdot\mathbf{n} = -\kappa\ f_3 - \tau\ f_4 + \tau\ f_6 \\ f_6' &= 2\mathbf{b}\cdot\mathbf{b}' = -2\tau\,\mathbf{b}\cdot\mathbf{n} = -2\tau\ f_5 \end{aligned}$$

We see that $(f_1, \dots, f_6)$ is a solution to the following linear system of ordinary differential equations

$$\begin{bmatrix} f_1' \\ f_2' \\ f_3' \\ f_4' \\ f_5' \\ f_6' \end{bmatrix} = \begin{bmatrix} 0 & 2\kappa & 0 & 0 & 0 & 0 \\ -\kappa & 0 & \tau & \kappa & 0 & 0 \\ 0 & -\tau & 0 & 0 & \kappa & 0 \\ 0 & -2\kappa & 0 & 0 & 2\tau & 0 \\ 0 & 0 & -\kappa & -\tau & 0 & \tau \\ 0 & 0 & 0 & 0 & -2\tau & 0 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \end{bmatrix}, \qquad \begin{bmatrix} f_1(s_0) \\ f_2(s_0) \\ f_3(s_0) \\ f_4(s_0) \\ f_5(s_0) \\ f_6(s_0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}.$$

We can immediately see that the constant function $s \mapsto (1, 0, 0, 1, 0, 1)$ also is a solution. By uniqueness we have that $\big(f_1(s), \dots, f_6(s)\big) = (1, 0, 0, 1, 0, 1)$ for all $s \in I$, i.e., $\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)$ is an orthonormal frame for all $s \in I$. Then we have $[\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)] = \pm 1$ for all $s \in I$ and as $[\mathbf{t}(s_0), \mathbf{n}(s_0), \mathbf{b}(s_0)] = 1$ continuity shows that $\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)$ is positively oriented for all $s \in I$.

We have in particular that $\mathbf{t}(s)$ is a unit vector for all $s \in I$ so if we put

$$\mathbf{r}(s) = \mathbf{x}_0 + \int_{s_0}^{s} \mathbf{t}(u)\,\mathrm{d}u, \qquad \text{for } s \in I$$

then $\mathbf{r} : I \to \mathbb{R}^3$ is a natural parametrization with $\mathbf{r}(s_0) = \mathbf{x}_0$. As $\mathbf{r}' = \mathbf{t}$ and $\mathbf{t}, \mathbf{n}, \mathbf{b}$ is a solution to the Frenet-Serret equations (2.13) we see that κ and τ is the curvature and torsion respectively of $\mathbf{r}$ and that $\mathbf{t}, \mathbf{n}, \mathbf{b}$ is the Frenet-Serret frame.

We have now established the existence of $\mathbf{r}$, but the definition of $\mathbf{r}(s)$ was forced if $\mathbf{r}$ was to solve the problem. □

If $\mathbf{r}$ is a natural parametrization of a space curve, and we put $\mathbf{r}_0 = \mathbf{r}(s_0)$, and let $\mathbf{t}$, $\mathbf{n}$, $\mathbf{b}$, κ, and τ be the tangent vector, the principal normal vector, the binormal vector, the curvature, and the torsion at s_0, respectively, then the Taylor expansion of $\mathbf{r}$ to third order at s_0 is

$$\mathbf{r}(s) = \mathbf{r}_0 + (s - s_0)\mathbf{t} + \frac{1}{2}(s - s_0)^2\kappa\mathbf{n} + \frac{1}{6}(s - s_0)^3\big(-\kappa^2\mathbf{t} + \kappa'\mathbf{n} + \kappa\tau\mathbf{b}\big) + \mathbf{o}\big((s - s_0)^3\big). \tag{2.19}$$

This expression is called the *canonical form* of a space curve. In a neighbourhood of $\mathbf{r}_0$ the projection into the osculating plane looks like the parabola $x_2 = \frac{1}{2}\kappa\, x_1^2$, the projection into the rectifying plane looks like the cubic $x_3 = \frac{1}{6}\kappa\tau\, x_1^3$, and projection into the normal plane looks like the curve $x_3^2 = \frac{2}{9}\big(\tau^2/\kappa\big)\, x_2^3$, see Figure 2.13.

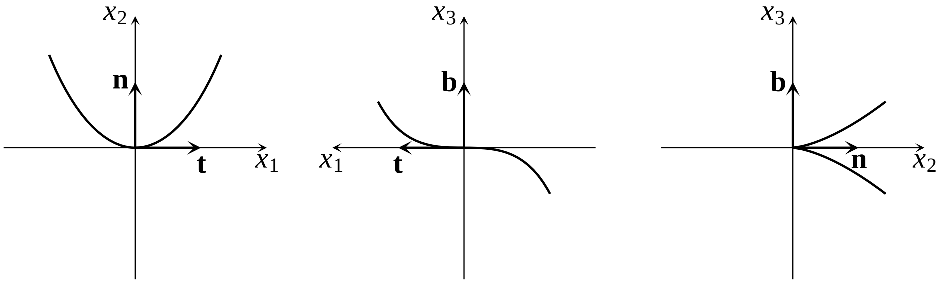


Figure 2.13: The projection of a curve into the osculating plane, the rectifying plane, and the normal plane.

Problems

2.4.1 Find the curvature, the torsion and the Frenet-Serret frame of the *helix* in Example 2.2, p. 42.

2.4.2 Show that if the curvature of a regular space curve is zero then the curve is a straight line.

2.4.3 Show that if a regular space curve has non vanishing curvature and constant torsion equal to zero then the curve is contained in a plane. Hint: first show that the binormal $\mathbf{b}$ is constant. Then consider the quantity $\mathbf{r}(t) \cdot \mathbf{b}$, where $\mathbf{r} : I \to \mathbb{R}^3$ is a regular parametrization of the curve.

2.4.4 Show that if a regular space curve has constant curvature different from zero and constant torsion equal to zero then the curve is a circle.

2.4.5 Consider the curve given by the parametrization

$$\mathbf{r}(t) = \begin{cases} (t, t^4, 0) & \text{for } t < 0 \\ (0, 0, 0) & \text{for } t = 0 \\ (t, 0, t^4) & \text{for } t > 0 \end{cases}$$

Show that this is a regular parametrization of class C^3. Let κ be the curvature and show that $\kappa(t) = 0$ if and only if $t = 0$. Show that the torsion is zero for all $t \neq 0$.

2.4.6 Show that if a regular space curve has constant curvature and torsion, both different from zero, then the curve is a circular helix, cf. Problem 2.4.1.

2.4.7 Let $\mathbf{r} : I \to \mathbb{R}^3$ be a regular parametrization and assume that $\mathbf{r}'(t)$ and $\mathbf{r}''(t)$ are linearly independent. Show that the Gram-Schmidt orthonormalization procedure of $\big(\mathbf{r}'(t), \mathbf{r}''(t)\big)$ gives $\big(\mathbf{t}(t), \mathbf{n}(t)\big)$.

Exercises

2.4.1 Write a program that finds the curvature and torsion at an arbitrary point of a Bézier curve in $\mathbb{R}^3$.

2.4.2 Write a program that finds the curvature and torsion at an arbitrary point of a B-spline curve in $\mathbb{R}^3$.

2.4.3 Write a program that plots the curvature and torsion as a function of arc length for a Bézier curve in $\mathbb{R}^3$.

2.4.4 Write a program that plots the curvature and torsion as a function of arc length for a B-spline curve in $\mathbb{R}^3$.

2.4.5 Write a program that finds the Frenet-Serret frame at an arbitrary point of a Bézier curve in $\mathbb{R}^3$.

2.4.6 Write a program that finds the Frenet-Serret frame at an arbitrary point of a B-spline curve in $\mathbb{R}^3$.

2.5 Curves in higher dimensional spaces

In this section we introduce the generalization of the Frenet-Serret frame and the Frenet-Serret equations to higher dimensions.

Theorem 2.28. *Let $\mathbf{r} : I \to \mathbb{R}^n$ be a natural parametrization of a regular curve in $\mathbb{R}^n$ of class C^n with tangent vector $\mathbf{t}(s)$. If the first $n-1$ derivatives*

$\mathbf{r}'(s), \mathbf{r}''(s), \ldots, \mathbf{r}^{(n-1)}(s)$ *are linearly independent then there exists* $n-1$ normal vectors $\mathbf{n}_1(s), \ldots, \mathbf{n}_{n-1}(s)$ *and* $n-1$ curvatures $\kappa_1, \ldots, \kappa_{n-1}$ *such that* $\kappa_i(s) > 0$ *for* $i = 1, \ldots, n-2$ *and* $\mathbf{t}(s), \mathbf{n}_1(s), \ldots, \mathbf{n}_{n-1}(s)$ *is a positively oriented orthonormal frame that satisfies the* Frenet-Serret equations:

$$\begin{bmatrix} \mathbf{t}' \\ \mathbf{n}_1' \\ \mathbf{n}_2' \\ \vdots \\ \mathbf{n}_{n-2}' \\ \mathbf{n}_{n-1}' \end{bmatrix} = \begin{bmatrix} 0 & \kappa_1 & 0 & 0 & \ldots & 0 \\ -\kappa_1 & 0 & \kappa_2 & 0 & \ldots & 0 \\ 0 & -\kappa_2 & 0 & \kappa_3 & & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & -\kappa_{n-2} & 0 & \kappa_{n-1} \\ 0 & \ldots & 0 & 0 & -\kappa_{n-1} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{n}_1 \\ \mathbf{n}_2 \\ \vdots \\ \mathbf{n}_{n-2} \\ \mathbf{n}_{n-1} \end{bmatrix} \tag{2.20}$$

Furthermore, for $m = 1, \ldots, n-1$

$$\operatorname{span}\{\mathbf{r}'(s), \mathbf{r}''(s), \ldots, \mathbf{r}^{(m)}(s)\} = \operatorname{span}\{\mathbf{t}(s), \mathbf{n}_1(s), \ldots, \mathbf{n}_m(s)\} \tag{2.21}$$

Proof. To ease notation a bit we put $\mathbf{n}_0 = \mathbf{t}$. We now use induction to prove that for $k = 1, \ldots, n-2$ we can find $\mathbf{n}_1, \ldots, \mathbf{n}_k$ and $\kappa_1, \ldots, \kappa_k$ such that

1. $\mathbf{n}_0, \ldots, \mathbf{n}_k$ are orthonormal.
2. (2.21) holds for $m = 1, \ldots, k$.
3. $\mathbf{n}_0' = \kappa_1 \mathbf{n}_1$ and $\mathbf{n}_m' = -\kappa_m \mathbf{n}_{m-1} + \kappa_{m+1} \mathbf{n}_{m+1}$ for $m = 1, \ldots, k-1$

As $\mathbf{r}'$ and $\mathbf{r}''$ are linearly independent we have in particular that $\mathbf{t}' = \mathbf{r}'' \neq \mathbf{0}$. So we can put $\kappa_1 = |\mathbf{t}'|$ and $\mathbf{n}_1 = \mathbf{t}'/\kappa_1$. Then we have

$$\mathbf{t}' = \kappa_1 \mathbf{n}_1, \qquad \text{and} \qquad \operatorname{span}\{\mathbf{r}', \mathbf{r}''\} = \operatorname{span}\{\mathbf{t}, \mathbf{n}_1\}.$$

Furthermore $\mathbf{t}$ is a unit vector so (2.8) shows that $\mathbf{t} \cdot \mathbf{n}_1 = 0$, i.e., $\mathbf{n}_0, \mathbf{n}_1$ are orthonormal. This proves the case $k = 1$.

Now assume we have proved the statement for some k. By hypothesis 1, a calculation like (2.14), and hypothesis 3 we have that

$$\mathbf{n}_m \cdot \mathbf{n}_k' = -\mathbf{n}_m' \cdot \mathbf{n}_k = \begin{cases} -\kappa_k & \text{for } m = k-1 \\ 0 & \text{otherwise} \end{cases}$$

By hypothesis 2 we have that

$$\mathbf{r}^{(k+1)} = \sum_{m=0}^{k} a_m \mathbf{n}_m.$$

Differentiation then gives

$$\begin{aligned}\mathbf{r}^{(k+2)} &= \sum_{m=0}^{k} a_m' \mathbf{n}_m + \sum_{m=0}^{k} a_m \mathbf{n}_m' \\ &= \sum_{m=0}^{k} a_m' \mathbf{n}_m + \sum_{m=0}^{k} a_m (-\kappa_m \mathbf{n}_{m-1} + \kappa_{m+1}\mathbf{n}_{m+1}) + a_k \mathbf{n}_k'\end{aligned}$$

As $\mathbf{r}^{(k+2)} \notin \operatorname{span}\{\mathbf{r}', \dots, \mathbf{r}^{(k+1)}\}$ we see that $\mathbf{n}_k' \notin \operatorname{span}\{\mathbf{n}_0, \dots, \mathbf{n}_k\}$. Furthermore the orthogonal projection of $\mathbf{n}_k'$ on $\operatorname{span}\{\mathbf{n}_0, \dots, \mathbf{n}_k\}$ is

$$\sum_{m=0}^{k} (\mathbf{n}_k' \cdot \mathbf{n}_m)\mathbf{n}_m = -\sum_{m=0}^{k} (\mathbf{n}_k \cdot \mathbf{n}_m')\mathbf{n}_m = -\kappa_{k-1}\mathbf{n}_{k-1}.$$

If we now put $\kappa_{k+1} = \left|\mathbf{n}_k' + \kappa_{k-1}\mathbf{n}_{k-1}\right|$ then $\kappa_{k+1} > 0$ so we can define $\mathbf{n}_{k+1} = \left(\mathbf{n}_k' + \kappa_{k-1}\mathbf{n}_{k-1}\right)/\kappa_{k+1}$. By construction we have $|\mathbf{n}_{k+1}| = 1$ and $\mathbf{n}_{k+1} \cdot \mathbf{n}_m = 0$ all $m = 0, \dots, k$, so $\mathbf{n}_0, \dots, \mathbf{n}_{k+1}$ are orthonormal. We also have that $\mathbf{n}_k' = -\kappa_k \mathbf{n}_{k-1} + \kappa_{k+1}\mathbf{n}_{k+1}$ by construction. Finally

$$\operatorname{span}\{\mathbf{r}', \dots, \mathbf{r}^{(k+2)}\} = \operatorname{span}\{\mathbf{n}_0, \dots, \mathbf{n}_k, \mathbf{n}_k'\} = \operatorname{span}\{\mathbf{n}_0, \dots, \mathbf{n}_k, \mathbf{n}_{k+1}\}.$$

This completes the induction.

We have now found $\mathbf{n}_0, \dots, \mathbf{n}_{n-2}$ and there is a unique unit vector $\mathbf{n}_{n-1}$ such $\mathbf{n}_0, \dots, \mathbf{n}_{n-1}$ is a positively oriented frame. We put $\kappa_{n-1} = \mathbf{n}_{n-2}' \cdot \mathbf{n}_{n-1}$ and then we have

$$\begin{aligned}\mathbf{n}_{n-2}' &= \sum_{k=0}^{n-1} \left(\mathbf{n}_{n-2}' \cdot \mathbf{n}_k\right)\mathbf{n}_k = -\sum_{k=0}^{n-3} \left(\mathbf{n}_{n-2} \cdot \mathbf{n}_k'\right)\mathbf{n}_k + \left(\mathbf{n}_{n-2}' \cdot \mathbf{n}_{n-1}\right)\mathbf{n}_{n-1} \\ &= -\kappa_{n-2}\mathbf{n}_{n-2} + \kappa_{n-1}\mathbf{n}_{n-1}\end{aligned}$$

and

$$\mathbf{n}_{n-1}' = \sum_{k=0}^{n-1} \left(\mathbf{n}_{n-1}' \cdot \mathbf{n}_k\right)\mathbf{n}_k = -\sum_{k=0}^{n-1} \left(\mathbf{n}_{n-1} \cdot \mathbf{n}_k'\right)\mathbf{n}_k = -\kappa_{n-1}\mathbf{n}_{n-2}$$

This completes the proof. □

The proof of Theorem 2.27 generalizes to curves in $\mathbb{R}^n$ and give us the following theorem.

Theorem 2.29. *Let I be an interval, let for $k = 1, \dots, n-2$, $\kappa_k : I \to \mathbb{R}$ be a strictly positiv C^{n-k-1} function and let $\kappa_{n-1} : I \to \mathbb{R}$ be a C^0 function. Let* *furthermore* $s_0 \in I$*, let* $\mathbf{x}_0$ *be a fixed point of* $\mathbb{R}^3$ *and let* $\mathbf{t}_0, \mathbf{n}_{1,0}, \dots, \mathbf{n}_{n-1,0}$ *be a fixed positively oriented orthonormal basis of* $\mathbb{R}^n$*. Then there exists a unique natural parametrization* $\mathbf{r} : I \to \mathbb{R}^n$ *of class* C^n *such that the curvatures are* $\kappa_1, \dots, \kappa_{n-1}$ *and the Frenet-Serret frame* $(\mathbf{t}, \mathbf{n}_1, \dots, \mathbf{n}_{n-1})$ *satisfies* $\mathbf{t}(s_0) = \mathbf{t}_0$ *and* $\mathbf{n}_k(s_0) = \mathbf{n}_{k,0}$ *for* $k = 1, \dots, n-1$.

The following theorem tells us how to find the normals and curvatures from an arbitrary parametrization of a curve in $\mathbb{R}^n$.

Theorem 2.30. *Let* $\mathbf{r} : I \to \mathbb{R}^n$ *be a parametrization of class* C^n *such that the first* $n-1$ *derivatives are linearly independent. The normals* $\mathbf{n}_1, \dots, \mathbf{n}_{n-2}$ *can be found by the Gram-Schmidt orthonormalization procedure.*

Step 1:

$$\mathbf{v}_0 = \mathbf{r}', \qquad \mathbf{n}_0 = \mathbf{t} = \frac{\mathbf{r}'}{|\mathbf{r}'|}.$$

Loop: for $m = 1, \dots, n-2$ *do*

$$\mathbf{v}_m = \mathbf{r}^{(m+1)} - \sum_{k=0}^{m-1} \big(\mathbf{r}^{(m+1)} \cdot \mathbf{n}_k\big)\mathbf{n}_k \qquad \mathbf{n}_m = \frac{\mathbf{v}_m}{|\mathbf{v}_m|}.$$

Let $\mathbf{e}_1, \dots, \mathbf{e}_n$ *be the standard basis in* $\mathbb{R}^n$*.*

For $m = 1, \dots, n$ *we put*

$$\mathbf{w}_m = \mathbf{e}_m - \sum_{k=0}^{n-2} \big(\mathbf{e}_m \cdot \mathbf{n}_k\big)\mathbf{n}_k \quad \text{and if } \mathbf{w}_m \neq \mathbf{0} \text{ then} \quad \mathbf{n}_{n-1} = \pm\frac{\mathbf{w}_m}{|\mathbf{w}_m|}$$

where "+" is used if $[\mathbf{n}_0, \dots, \mathbf{n}_{n-2}, \mathbf{e}_m] < 0$ *otherwise "−" is used.*

The curvatures $\kappa_1, \dots, \kappa_{n-1}$ *are now given by*

$$\kappa_m = \frac{\mathbf{r}^{(m+1)} \cdot \mathbf{n}_m}{|\mathbf{r}'||\mathbf{v}_{m-1}|}, \quad m = 1, \dots, n-1. \tag{2.22}$$

Proof. By Problem 2.5.2 the Gram-Schmidt orthonormalization procedure gives the first $n-2$ normals. For at least one m we have that $\mathbf{n}_0, \dots, \mathbf{n}_{n-2}, \mathbf{e}_m$ is a basis. For such a m $\mathbf{n}_0, \dots, \mathbf{n}_{n-2}, \mathbf{w}_m/|\mathbf{w}_m|$ is an orthonormal basis so $\mathbf{n}_{n-1} = \pm\mathbf{w}_m/|\mathbf{w}_m|$. The sign is determined by the requirement that $\mathbf{n}_0, \dots, \mathbf{n}_{n-1}$ is positively oriented, i.e., by the requirement that $[\mathbf{n}_0, \dots, \mathbf{n}_{n-1}] = 1$. So the sign is the same as the sign of $\big[\mathbf{n}_0, \dots \mathbf{n}_{n-1}, \mathbf{w}_m/|\mathbf{w}_m|\big] = \big[\mathbf{n}_0, \dots \mathbf{n}_{n-1}, \mathbf{w}_m\big]/|\mathbf{w}_m|$ which has the same sign as $\big[\mathbf{n}_0, \dots \mathbf{n}_{n-1}, \mathbf{w}_m\big] = \big[\mathbf{n}_0, \dots \mathbf{n}_{n-1}, \mathbf{e}_m\big]$.

For $m = 1, \ldots, n-1$ we now have

$$|\mathbf{v}_{m-1}|\mathbf{n}_{m-1} = \mathbf{r}^{(m)} - \sum_{k=0}^{m-2} \big(\mathbf{r}^{(m)} \cdot \mathbf{n}_k\big)\mathbf{n}_k$$

so differentiation with respect to s gives

$$\begin{aligned}\frac{\mathrm{d}|\mathbf{v}_{m-1}|}{\mathrm{d}s}\mathbf{n}_{m-1} + |\mathbf{v}_{m-1}|\frac{\mathrm{d}\mathbf{n}_{m-1}}{\mathrm{d}s} \\ = \frac{\mathrm{d}\mathbf{r}^{(m)}}{\mathrm{d}s} - \sum_{k=0}^{m-2} \frac{\mathrm{d}\big(\mathbf{r}^{(m)} \cdot \mathbf{n}_k\big)}{\mathrm{d}s}\mathbf{n}_k - \sum_{k=0}^{m-2}\big(\mathbf{r}^{(m)} \cdot \mathbf{n}_k\big)\frac{\mathrm{d}\mathbf{n}_k}{\mathrm{d}s} \\ = \frac{\mathbf{r}^{(m+1)}}{|\mathbf{r}'|} - \sum_{k=0}^{m-2} \frac{\mathrm{d}\big(\mathbf{r}^{(m)} \cdot \mathbf{n}_k\big)}{\mathrm{d}s}\mathbf{n}_k - \sum_{k=0}^{m-2}\big(\mathbf{r}^{(m)} \cdot \mathbf{n}_k\big)\frac{\mathrm{d}\mathbf{n}_k}{\mathrm{d}s}.\end{aligned}$$

If we take the inner product with $\mathbf{n}_m$ then we obtain

$$|\mathbf{v}_{m-1}|\kappa_m = \frac{\mathbf{r}^{(m+1)} \cdot \mathbf{n}_m}{|\mathbf{r}'|}$$

This is the same as (2.22). □

Problems

2.5.1 Let $\mathbf{r} : I \to \mathbb{R}^n$ be a natural parametrization of a regular curve of class C^n and assume that the derivatives $\mathbf{r}'(s), \ldots, \mathbf{r}^{(n-1)}(s)$ are linearly independent. Show that if the Gram-Schmidt orthonormalization procedure is used on $\big(\mathbf{r}'(s), \ldots, \mathbf{r}^{(n-1)}(s)\big)$ then we get $\big(\mathbf{n}_0(s), \ldots, \mathbf{n}_{n-2}(s)\big)$.

2.5.2 Let $\mathbf{r} : I \to \mathbb{R}^n$ be a regular parametrization of class C^n and assume that the derivatives $\mathbf{r}'(t), \ldots, \mathbf{r}^{(n-1)}(t)$ are linearly independent. Show that the Gram-Smidth orthonormalization procedure of $\big(\mathbf{r}'(t), \ldots, \mathbf{r}^{(n-1)}(t)\big)$ gives $\big(\mathbf{n}_0(t), \ldots, \mathbf{n}_{n-2}(t)\big)$.

2.5.3 Check that the procedure in Theorem 2.30 for $n = 2$ gives the same result as Theorem 2.18.

2.5.4 Check that the procedure in Theorem 2.30 for $n = 3$ gives the same result as Theorem 2.26.

2.5.5 What simplifications can be made to the procedure in Theorem 2.26 if we only want $\kappa_1, \ldots, \kappa_{n-12}, |\kappa_{n-1}|$, i.e., if we ignore the sign of κ_{n-1}.

Exercises

2.5.1 Write a program that for a Bézier curve implements the procedure in Theorem 2.30 with the simplification from Problem 2.5.5.

2.5.2 Write a program that for a Bézier curve implements the procedure in Theorem 2.30.

2.5.3 Write a program that for a B-spline curve implements the procedure in Theorem 2.30 with the simplification from Problem 2.5.5.

2.5.4 Write a program that for a B-spline curve implements the procedure in Theorem 2.30.

Chapter 3

Polynomial Surfaces

3.1 Introduction

We will introduce two surface types, tensor product (Bézier and B-spline) surfaces and triangular Bézier surfaces. The latter is a direct generalization of the de Casteljau constrution to surfaces or any other higher dimension.

The tensor product constrution works for any pair of curve schemes. Suppose we have two curve schemes $t \mapsto \sum_{i=1}^{n} \mathbf{a}_i \phi_i(t)$ and $t \mapsto \sum_{j=1}^{m} \mathbf{b}_j \psi_j(t)$ where $\phi_1(t), \ldots, \phi_n(t)$ and $\psi_1(t), \ldots, \psi_m(t)$ are two sets of linearly independent real functions defined on $[a, b]$ and $[c, d]$ respectively. We can now define a surface scheme, i.e., a space of surfaces defined on $[a, b] \times [c, d]$ by

$$(u, v) \mapsto \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{a}_{i,j} \phi_i(u) \psi_j(v), \qquad \mathbf{a}_{i,j} \in \mathbb{R}^d.$$

From the point of view of linear algebra the two set of functions are bases for two vector spaces V and W respectively. If we form the *tensor product* $V \otimes W$ then the products $\phi_i \otimes \psi_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ is a basis for $V \otimes W$. The map

$$\sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j} \phi_i \otimes \psi_j \mapsto \sum_{i=1}^{n} \sum_{j=1}^{m} a_{i,j} \phi_i(u) \psi_j(v), \qquad a_{i,j} \in \mathbb{R},$$

from $V \otimes W$ to $C([a, b] \times [c, d], \mathbb{R})$ is linear and injective, so the basis functions $\phi_i(u)\psi_j(v)$ in our surface scheme can be considered as a basis in the tensor product of the two curve scheme, hence the name tensor product surface.

3.2 Tensor product Bézier surfaces.

Recall that the Bernstein polynomials $B_i^n(t)$, $i = 0, \ldots, n$ are a basis for the polynomials of degree at most n. Hence the products $B_i^n(u)B_j^m(v)$, $i = 0, \ldots, n$, $j = 0, \ldots, m$ are a basis for the polynomials in (u, v) of degree at most n in u and m in v. Thus, a surface parametrisation $\mathbf{r}(u, v)$ which is polynomial of degree at most n in u and m in v, can be written as a so called *tensor product Bézier surface*:

$$\begin{aligned}\mathbf{r}(u, v) &= \sum_{i=0}^{n}\sum_{j=0}^{m} \mathbf{b}_{i,j} B_i^n(u) B_j^m(v) \\ &= \sum_{j=0}^{m} \underbrace{\left(\sum_{i=0}^{n} \mathbf{b}_{i,j} B_i^n(u)\right)}_{\mathbf{c}_j(u)} B_j^m(v) = \sum_{i=0}^{n} \underbrace{\left(\sum_{j=0}^{m} \mathbf{b}_{i,j} B_j^m(v)\right)}_{\mathbf{d}_i(v)} B_i^n(u). \qquad (3.1)\end{aligned}$$

The coefficients $\mathbf{b}_{i,j}$ are called *control points*: or *Bézier points* and form the *control net*. The last line shows that we can consider a tensor product Bézier surface of degree $n \times m$ as a Bézier curve of degree m in the space of Bézier curves of degree n (with "control points" $\mathbf{c}_j(u)$) or as a Bézier curve of degree n in the space of Bézier curves of degree m (with "control points" $\mathbf{d}_i(v)$). Imagine the surface as being swept out by moving a curve through space while at the same time changing the curve, see Figure 3.1. To evaluate a point on the surface we can take n steps

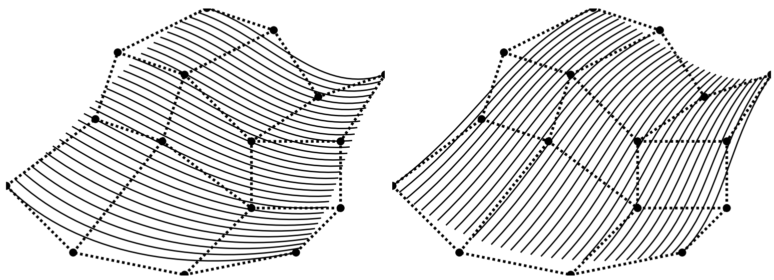


Figure 3.1: A bi-cubic tensor product Bézier surface. To the left the control net with u-parameter curves, and to the right with v-parameter curves.:

of *de Casteljau's algorithm* in the columns of the control net:

$$\mathbf{b}_{i,j}^{r,s}(u, v) = (1-u)\mathbf{b}_{i,j}^{r-1,s}(u, v) + u\mathbf{b}_{i+1,j}^{r-1,s}(u, v), \qquad (3.2)$$

and m steps in the rows of the control net:

$$\mathbf{b}_{i,j}^{r,s}(u, v) = (1-v)\mathbf{b}_{i,j}^{r,s-1}(u, v) + v\mathbf{b}_{i,j+1}^{r,s-1}(u, v). \qquad (3.3)$$

The order in which these $n+m$ steps are performed doesn't matter (de Casteljau's algorithm in one direction permutes with de Casteljau's algorithm in the other direction). If we alternate between the two directions, then we can combine two steps in different directions into one step (called a *direct de Casteljau step*).

$$\mathbf{b}_{i,j}^{r,s}(u,v) = \begin{bmatrix} 1-u & u \end{bmatrix} \begin{bmatrix} \mathbf{b}_{i,j}^{r-1,s-1}(u,v) & \mathbf{b}_{i,j+1}^{r-1,s-1}(u,v) \\ \mathbf{b}_{i+1,j}^{r-1,s-1}(u,v) & \mathbf{b}_{i+1,j+1}^{r-1,s-1}(u,v) \end{bmatrix} \begin{bmatrix} 1-v \\ v \end{bmatrix}. \qquad (3.4)$$

This is *bilinear interpolation* in each cell of the control net. We can also use the de Casteljau operator to define these surfaces. Let R_1 remove the last row in the control net and let L_1 remove the first row in the control net. Similarly let R_2 and L_2 remove the last and first column respectively. We also have two *difference operators* $\Delta_i = L_i - R_i$ and two *de Casteljau operators*

$$C_i(t) = (1-t)R_i + tL_i = R_i + t\Delta_i, \qquad i = 1, 2.$$

A tensor product Bézier surface can now be written as

$$\mathbf{r}(u,v) = C_1(u)^n C_2(v)^m \big(\mathbf{b}_{i,j}\big). \qquad (3.5)$$

A parameter curve is the restriction of the map $\mathbf{r}(u,v)$ to a horizontal or vertical line and they are Bézier curves of degree n and m respectively. The restriction to any other line in the parameter plane is still a Bézier curve, but the degree is in general $n+m$. The bilinear interpolation described above is given by $C(u,v) = C_1(u)C_2(v)$, and if $n \geq m$ then we can write

$$\mathbf{r}(u,v) = C_1(u)^{n-m} C(u,v)^m \big(\mathbf{b}_{i,j}\big), \qquad (3.6)$$

see Figure 3.2.

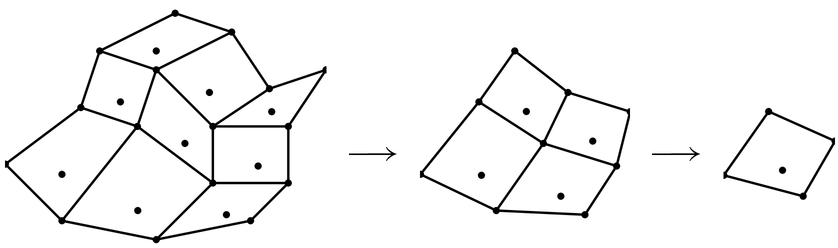


Figure 3.2: Evaluating a point on a bi-cubic Bézier surface by repeated bilinear interpolation, (three direct de Casteljau steps).

It follows from the construction that we have *affine invariance*, cf. Problem 3.2.1, and the *convex hull property*, cf. Problem 3.2.2. The four *boundary curves* of

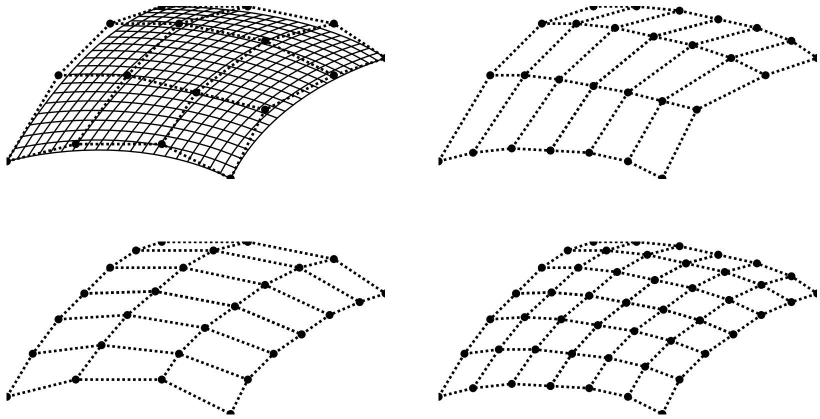


Figure 3.3: Subdivision of tensor product Bézier surfaces. At the top left the surface is shown with the original control net and to the right the control net subdivided at $u = 0.5$. At the bottom to the left the control net is subdivided at $v = 0.5$ and to the right it is subdivided in both directions.

$\mathbf{r}(u, v)$ are Bézier curves whose control polygons are the boundary polygons of the control net.

Just as in the curve case we can perform *subdivision* on a tensor product Bézier surface. Performing n steps of de Casteljau's algorithm in the columns of the control net yields a v-parameter curve (in Bézier form) on the surface and as a byproduct the control net for the two pieces on each side of this curve. Similarly, m steps of de Casteljau's algorithm in the rows of the control net yields a u-parameter curve on the surface and the control net for the two pieces on each side of the curve. Performing all $n + m$ steps yields a point on the surface and the control net for four pieces of the surface, see Figure 3.3.

A polynomial surface can be considered as a polynomial surface of higher degree in u and/or v, and we can perform *degree elevation*. We raise the degree in u by raising the degree of each column $\mathbf{b}_{0j}, \ldots, \mathbf{b}_{nj}$ in the control net and we raise the degree in v by raising the degree of each row $\mathbf{b}_{i0}, \ldots, \mathbf{b}_{im}$.

3.2.1 Differentation of a tensor product Bézier surface

The partial derivatives of a tensor product Bézier surface is a tensor product Bézier surface with control points essentially given by differences in the control net.

$$\frac{\partial^{r+s}\mathbf{r}}{\partial u^r \partial v^s}(u, v) = \frac{n!m!}{(n-r)!(m-s)!} C_1(u)^{n-r} C_2(v)^{m-s} \Delta_1^r \Delta_2^s \big(\mathbf{b}_{i,j}\big) \qquad (3.7)$$

$$= \frac{n!m!}{(n-r)!(m-s)!} \Delta_1^r \Delta_2^s C_1(u)^{n-r} C_2(v)^{m-s} \big(\mathbf{b}_{i,j}\big) \tag{3.8}$$

We have in particular

$$\frac{\partial \mathbf{r}}{\partial u}(u,v) = nC_1(u)^{n-1} C_2(v)^m \Delta_1\big(\mathbf{b}_{i,j}\big), \tag{3.9}$$

$$\frac{\partial \mathbf{r}}{\partial v}(u,v) = mC_1(u)^{n} C_2(v)^{m-1} \Delta_2\big(\mathbf{b}_{i,j}\big), \tag{3.10}$$

$$\frac{\partial^2 \mathbf{r}}{\partial u \partial v}(u,v) = nmC_1(u)^{n-1} C_2(v)^{m-1} \Delta_1 \Delta_2\big(\mathbf{b}_{i,j}\big) \tag{3.11}$$

The mixed derivative is called *the twist* and its control points $\Delta_1\Delta_2\big(\mathbf{b}_{i,j}\big)$ measure the deviation of the cells in the control net from being parallelograms, see Figure 3.4. Of particular interest is the *cross boundary derivative*, they are given

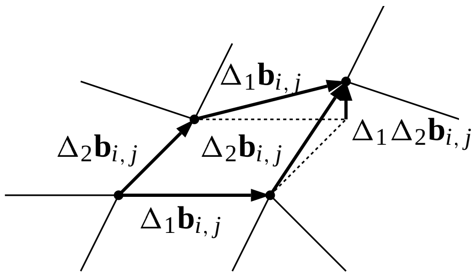


Figure 3.4: The twist vectors $\Delta_1\Delta_2\mathbf{b}_{i,j}$ measure how far the cells are from being parallelograms.

by

$$\frac{\partial \mathbf{r}}{\partial u}(0,v) = nC^m(v)\big(\mathbf{b}_{1,0} - \mathbf{b}_{0,0}, \ldots, \mathbf{b}_{1,m} - \mathbf{b}_{0,m}\big) \tag{3.12}$$

$$\frac{\partial \mathbf{r}}{\partial u}(1,v) = nC^m(v)\big(\mathbf{b}_{n,0} - \mathbf{b}_{n-1,0}, \ldots, \mathbf{b}_{n,m} - \mathbf{b}_{n-1,m}\big) \tag{3.13}$$

$$\frac{\partial \mathbf{r}}{\partial v}(u,0) = mC^n(u)\big(\mathbf{b}_{0,1} - \mathbf{b}_{0,0}, \ldots, \mathbf{b}_{n,1} - \mathbf{b}_{n,0}\big) \tag{3.14}$$

$$\frac{\partial \mathbf{r}}{\partial v}(u,1) = mC^n(v)\big(\mathbf{b}_{0,m} - \mathbf{b}_{0,m-1}, \ldots, \mathbf{b}_{n,m} - \mathbf{b}_{n,m-1}\big) \tag{3.15}$$

Problems

3.2.1 Show that if the control points of a tensor product Bézier surface is subjected to an affine transformation, then the new surface is the image of the original surface under the same affine transformation.

3.2.2 Show that a tensor product Bézier surface is contained in the convex hull of the control points.

3.2.3 Show that the corner twist measures the diviation of the corner cell in the control net from being a parallelogram, cf. Figure 3.4.

3.2.4 Given a surface defined on some rectangle. Show that if you know the cross boundary derivative on an edge of the rectangle, then the corner twists (the mixed 2nd order partial derivative) are determined at the two vertices of that edge.

3.2.5 Now assume that the cross boundary derivatives are known on two edges meeting at a corner. According to the previous exercise the twist at this corner is determined in two ways, can you be sure that the two results agree?

Exercises

3.2.1 Implement de Casteljau's algorithm for tensor product Bézier surfaces. (Use procedures/functions/sub routines for Bézier curves.)

3.2.2 Implement subdivision for tensor product Bézier surfaces, split it into four pieces. (Use procedures/functions/sub routines for Bézier curves.)

3.2.3 Write a program that determines a point and all partial derivatives to order k on a tensor product Bézier surfaces.

3.3 Tensor product B-spline surfaces.

Let two knot vectors $\mathbf{u} = u_0, \ldots, u_{2n+N}$ and $\mathbf{v} = v_0, \ldots, v_{2m+M}$ be given and consider the uv-plane (the parameter plane). The vertical lines $u = u_i$ and the horizontal lines $v = v_j$ are called *knot lines* and they are assigned the same multiplicity as the corresponding knots. If $N_i^n(u|\mathbf{u})$ denote the B-spline functions of degree n on the knot vector $\mathbf{u}$ and $N_j^m(v|\mathbf{v})$ denote the B-spline functions of degree m on the knot vector $\mathbf{v}$, then a *tensor product B-spline surface* with knot vectors $\mathbf{u}$ and $\mathbf{v}$ is written as

$$\begin{aligned}\mathbf{r}(u,v) &= \sum_{i=1}^{N+n}\sum_{j=1}^{M+m}\mathbf{d}_{i,j}N_i^n(u|\mathbf{u})N_j^m(v|\mathbf{v})\\ &= \sum_{j=1}^{M+m}\underbrace{\left(\sum_{i=1}^{N+n}\mathbf{d}_{i,j}N_i^n(u|\mathbf{u})\right)}_{\mathbf{c}_j(u)}N_j^m(v|\mathbf{v}) = \sum_{i=1}^{N+n}\underbrace{\left(\sum_{j=1}^{M+m}\mathbf{d}_{i,j}N_j^m(v|\mathbf{v})\right)}_{\mathbf{a}_i(v)}N_i^n(u|\mathbf{u}).\end{aligned} \tag{3.16}$$

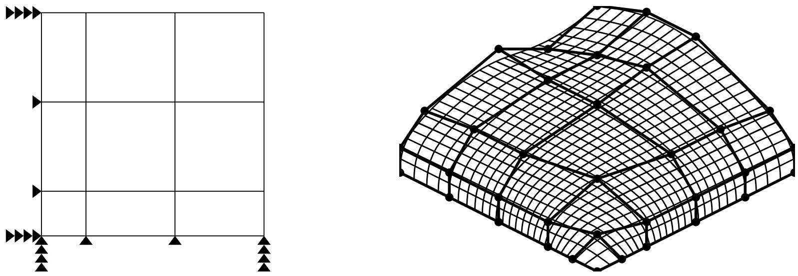


Figure 3.5: To the left the knot lines and to the right the control net and the tensor product B-spline surface.

The *control points* or *de Boor points* $\mathbf{d}_{i,j}$ form the *control net*, see Figure 3.5.

Just as in the previous case we can regard the surface as a B-spline curve of degree m with knot vector $\mathbf{v}$ and "control points" $\mathbf{c}_j(u)$, i.e., a curve in the spline space of degree n on the knot vector $\mathbf{u}$, or as a B-spline curve of degree n with knot vector $\mathbf{u}$ and "control points" $\mathbf{a}_i(v)$, i.e., a curve in the spline space of degree m on the knot vector $\mathbf{v}$.

To evaluate a point on the surface we can take n steps of *de Boor algorithm* in the columns of the control net, and m steps in the rows of the control net.

Theorem 3.1. *Let* $\mathbf{r}$ *be a tensor product B-spline surface of degree* n, m *with knots* $u_0, \ldots, u_{2n+N}$ *and* $v_0, \ldots, v_{2m+M}$ *and control points* $\mathbf{d}_{1,1}, \ldots, \mathbf{d}_{n+N,m+M}$*. If* $u \in [u_{k-1}, u_k]$ *and* $v \in [v_{l-1}, v_l]$ *with* $u_{k-1} < u_k$ *and* $v_{l-1} < v_l$*, then we can determine the point* $\mathbf{r}(u, v)$ *on the curve by de Boor's algorithm. First initialize:*

$$
\begin{aligned}
s_i &= u_{k-n+i} && i = 1, \ldots, 2n \\
t_j &= v_{l-m+j} && j = 1, \ldots, 2m \\
\mathbf{d}_{i,j}^{0,0}(u, v) &= \mathbf{d}_{k+1-n+i,l+1-m+j} && i = 0, \ldots, n, \quad j = 0, \ldots, m
\end{aligned}
$$

Then for $p = 1, \ldots, n$ *do:*

$$
\left.
\begin{aligned}
\alpha_i^p &= \frac{u - s_i}{s_{n+1+i-p} - s_i} \\
\mathbf{d}_{i,j}^{p,0}(u, v) &= \left(1 - \alpha_i^p\right)\mathbf{d}_{i-1,j}^{p-1,0}(u, v) + \alpha_i^r \mathbf{d}_{i,j}^{p-1,0}(u, v)
\end{aligned}
\right\}
\qquad
\begin{cases}
i = p, \ldots, n \\
j = 0, \ldots, m
\end{cases}
$$

and for $q = 1, \ldots, m$ *do:*

$$
\left.
\begin{aligned}
\beta_j^q &= \frac{v - t_j}{t_{m+1+j-q} - t_j} \\
\mathbf{d}_{n,j}^{n,q}(u, v) &= \left(1 - \beta_j^q\right)\mathbf{d}_{n,j-1}^{n,q-1}(u, v) + \beta_j^q \mathbf{d}_{n,j}^{n,q-1}(u, v)
\end{aligned}
\right\}
\qquad j = k, \ldots, m
$$

Finally, the point on the curve is $\mathbf{r}(u, v) = \mathbf{d}_{n,m}^{n,m}(u, v)$.

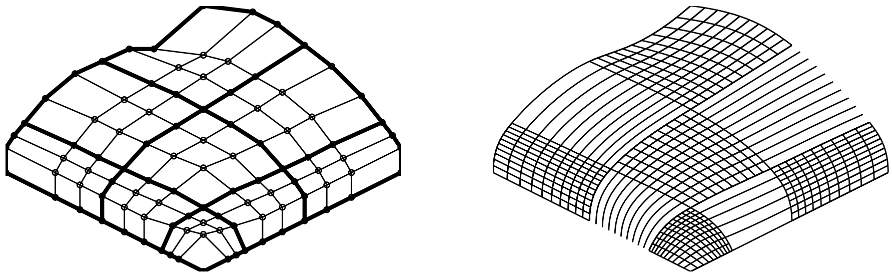


Figure 3.6: The Bézier patches of a bi-cubic B-spline surface. To the left we show the control points.

The order in which these $n + m$ steps are performed doesn't matter (de Boor's algorithm in one direction permutes with de Boor's algorithm in the other direction). So we can extend the recursion to give points

$$\begin{aligned}\mathbf{d}_{i,j}^{p,q}(u,v) &= \left(1-\alpha_i^p\right)\mathbf{d}_{i-1,j}^{p-1,q}(u,v) + \alpha_i^p\mathbf{d}_{i,j}^{p-1,q}(u,v)\\ &= \left(1-\beta_j^q\right)\mathbf{d}_{i,j-1}^{p,q-1}(u,v) + \beta_j^q\mathbf{d}_{i,j}^{i,q-1}(u,v) \qquad (3.17)\end{aligned}$$

It follows from the construction that we have *affine invariance*, cf. Problem 3.3.1, and *the strong convex hull property*, cf. Problem 3.3.2. If the outer knots have full multiplicity then the four *boundary curves* of $\mathbf{r}(u, v)$ are B-spline curves who's control polygons are the boundary polygons of the control net.

One step of de Boor's algorithm in the columns or the rows or the control net gives the control net for the surface with an extra knot line inserted, this is called *knot line insertion*, and as in the curve case de Boor's algorithm is the same as repeated knot line insertion. On each rectangle $[u_i, u_{i+1}] \times [v_j, v_{j+1}]$, bounded by consecutive knot lines, the surface is polynomial, and can of course be considered as a Bézier surface, called a *Bézier patch*. The tensor product Bézier control points are found by inserting each of the four knot lines to full multiplicity, see Figure 3.6.

3.3.1 Differentation of a tensor product B-spline surface

As in the curve case the partial derivatives can be found from de Boor's algorithm (3.17)

$$\frac{\partial \mathbf{r}}{\partial u}(u,v) = \frac{\mathbf{d}_{n,m}^{n-1,m}(u,v) - \mathbf{d}_{n-1,m}^{n-1,m}(u,v)}{s_{n+1} - s_n}, \tag{3.18}$$

$$\frac{\partial \mathbf{r}}{\partial v}(u,v) = \frac{\mathbf{d}_{n,m}^{n,m-1}(u,v) - \mathbf{d}_{n,m-1}^{n,m-1}(u,v)}{t_{m+1} - t_m}. \tag{3.19}$$

Alternatively $\frac{\partial}{\partial u}\mathbf{r}$ and $\frac{\partial}{\partial v}\mathbf{r}$ are B-spline surfaces of lower degree, with the same knot lines as $\mathbf{r}$, except that two of the outermost knot lines are removed, and with control points

$$\frac{\mathbf{d}_{i,j} - \mathbf{d}_{i-1,j}}{u_{i+n-1} - u_{i-1}} \quad \text{and} \quad \frac{\mathbf{d}_{i,j} - \mathbf{d}_{i,j-1}}{v_{j+m-1} - v_{j-1}} \tag{3.20}$$

respectively. If the outer knots have full multiplicity, then the first two rows of control points are Bézier control points for the first rows of Bézier patches. Thus, the cross boundary derivatives are easy to express as B-spline curves, e.g.:

$$\frac{\partial \mathbf{r}}{\partial u}(u_m, v) = \frac{n}{u_{m+1} - u_m} \sum_{j=0}^{m} (\mathbf{d}_{2,j} - \mathbf{d}_{1,j}) M_j^m(v).$$

There are similar formulae for the other three cross boundary derivative.

The restriction to the horizonal and vertical lines are B-spline curves of degree n and with knots $u_0, \ldots, u_{2n+N}$ and B-spline curve of degree m and with knots $v_0, \ldots, v_{2m+M}$ respectively. The restriction to any other line is still a B-spline curve, but the degree is $n + m$ and it has a knot each time the line crosses a knot line. If the multiplicity of the knot line is ν, then the multiplicity of the knot on the slanted line is $m + \nu$ if the knot line is horizontal and it is $n + \nu$ if the knot line is vertical.

Problems

3.3.1 Show that if the control points of a tensor product B-spline surface is subjected to an affine transformation, then the new surface is the image of the original surface under the same affine transformation.

3.3.2 Show that a tensor product B-spline surface of degree n, m with control points $\mathbf{d}_{i,j}$ is contained in

$$\bigcup_{r,s} \text{convex hull of} \big\{ \mathbf{d}_{r+i,s+i} \mid i = 0, \ldots, n, \; j = 0, \ldots, m \big\}.$$

Exercises

3.3.1 Implement the knot line insertion procedure for tensor product B-spline surfaces. (Use procedures/functions/sub routines for a B-spline curve)

3.3.2 Implement de Boor's algorithm for tensor product B-spline surfaces. (Use procedures/functions/sub routines for a B-spline curve)

3.3.3 Write a program that determines a point and all partial derivatives to order k on a tensor product B-spline surface.

3.4 Triangular Bézier surfaces

Let **A**, **B**, and **C** be the corners of a triangle in the plane. For any point **P** in the plane there is a unique *barycentric combination* that gives **P**:

$$\mathbf{P} = u\mathbf{A} + v\mathbf{B} + w\mathbf{C}, \quad u + v + w = 1,$$

the triple (u, v, w) is called *barycentric coordinates* for the point **P**, see Figure 3.7. Points in the interior of the triangle corresponds to positive barycentric coordi-

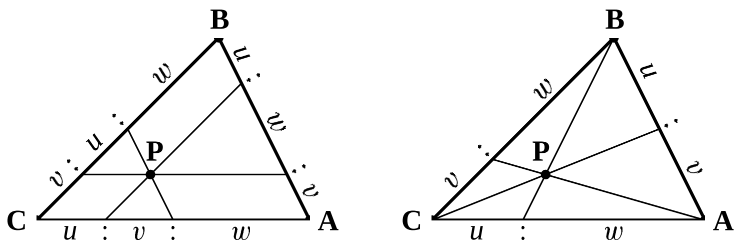


Figure 3.7: Barycentric coordinates in the plane.

nates: $u, v, w \geq 0$, and consequently $u, v, w \leq 1$. By letting $w = 1 - u - v$ we obtain ordinary coordinates (u, v) in the plane. (It corresponds to choosing **C** as the origin and $\overrightarrow{\mathbf{CA}}$ and $\overrightarrow{\mathbf{CB}}$ as basis vectors in a coordiante system). *Points* in the plane have barycentric coordinates that sums to one. A *vector* in the plane can be considered as the difference between two points $\overrightarrow{\mathbf{PQ}} = \mathbf{Q} - \mathbf{P}$ and have barycentric coordinates that sum to zero.

The control net for a triangular Bézier surface of degree n consists of $\frac{(n+1)(n+2)}{2}$ points arranged in a triangular grid, see Figure 3.8. If we let $\mathbf{i} = (i, j, k)$ or $\mathbf{i} = (i_1, i_2, i_3)$ denote a multi-index, then a control point can be denoted $\mathbf{b_i}$. We furthermore put $|\mathbf{i}| = i + j + k = \sum_\ell i_\ell$ and let $\mathbf{e}_1 = (1, 0, 0)$, $\mathbf{e}_2 = (0, 1, 0)$, and $\mathbf{e}_3 = (0, 0, 1)$. Now let $\mathbf{u} = (u, v, w)$ be barycentric coordinates for some

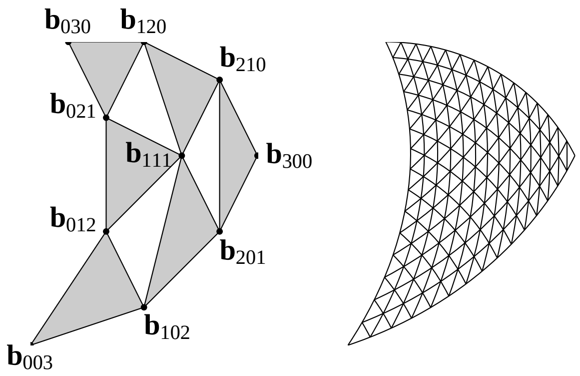


Figure 3.8: To the left the control net for a cubic triangular Bézier surface. The shaded triangles correspond to the edges of the control polygon of a Bézier curve — this is where the interpolation takes place. The surface is shown to the right.

point in the plane (normally the point will be inside the triangle), *de Casteljau's algorithm* is the following. Put

$$\mathbf{b}_{\mathbf{i}}^0(\mathbf{u}) = \mathbf{b}_{\mathbf{i}}, \qquad \text{all } \mathbf{i} \text{ with } |\mathbf{i}| = n$$

and for $r = 1, \ldots, n$:

$$\mathbf{b}_{\mathbf{i}}^r(\mathbf{u}) = u\mathbf{b}_{\mathbf{i}+\mathbf{e}_1}^{r-1}(\mathbf{u}) + v\mathbf{b}_{\mathbf{i}+\mathbf{e}_2}^{r-1}(\mathbf{u}) + w\mathbf{b}_{\mathbf{i}+\mathbf{e}_3}^{r-1}(\mathbf{u}), \qquad \text{all } \mathbf{i} \text{ with } |\mathbf{i}| = n - r$$

the point on the surface is the final point:

$$\mathbf{r}(\mathbf{u}) = \mathbf{b}_{000}^n(\mathbf{u}).$$

This is the direct generalisation of de Casteljau's algorithm for Bézier curves (the numbers $1 - t$ and t are exactly the barycentric coordinates for t with respect to the points $0, 1 \in \mathbb{R}$). In the cubic case we have the following diagrams:

$$\begin{array}{c} \mathbf{b}_{030}^0 \\ \mathbf{b}_{021}^0\mathbf{b}_{120}^0 \\ \mathbf{b}_{012}^0\mathbf{b}_{111}^0\mathbf{b}_{210}^0 \\ \mathbf{b}_{003}^0\mathbf{b}_{102}^0\mathbf{b}_{201}^0\mathbf{b}_{300}^0 \end{array} \xrightarrow{\mathcal{C}(\mathbf{u})} \begin{array}{c} \mathbf{b}_{020}^1 \\ \mathbf{b}_{011}^1\mathbf{b}_{110}^1 \\ \mathbf{b}_{002}^1\mathbf{b}_{101}^1\mathbf{b}_{200}^1 \end{array} \xrightarrow{\mathcal{C}(\mathbf{u})} \begin{array}{c} \mathbf{b}_{010}^2 \\ \mathbf{b}_{001}^2\mathbf{b}_{100}^2 \end{array} \xrightarrow{\mathcal{C}(\mathbf{u})} \mathbf{b}_{000}^3, \tag{3.21}$$

see Figure 3.9, where one step in de Casteljau's algorithm (going from triangular net to the next) is considered as an operator $\mathcal{C}(\mathbf{u})$ taking a triangular net and producing a smaller net by linear interpolation. With this notation we have

$$\mathbf{r}(\mathbf{u}) = \mathcal{C}(\mathbf{u})^n\big(\mathbf{b}_{\mathbf{i}}\big). \tag{3.22}$$

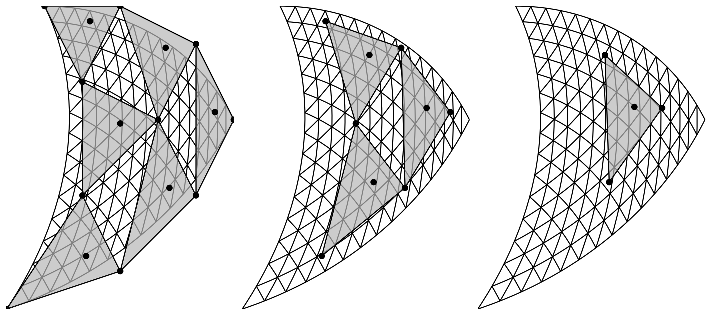


Figure 3.9: de Casteljau's algorithm for a triangular Bézier surface.

If we for $i = 1, 2, 3$ introduce the shorthand notation $\mathcal{C}_i = \mathcal{C}(\mathbf{e}_i)$, then we can write $\mathcal{C}(\mathbf{u}) = u\mathcal{C}_1 + v\mathcal{C}_2 + w\mathcal{C}_3$. The operator $\mathcal{C}_\ell$ simply deletes the row $i_\ell = 0$ (and subtracts 1 from i_ℓ). The operators obviously *commutes*, and this implies that all operators built from these three basic ones commute too. In other words we can calculate with them as with ordinary numbers.

As the surface is defined by repeated interpolation we immediately have *affine invariance* and the *convex hull property*. Putting one of the barycentric coordinates equal to one in de Casteljau's algorithm gives $\mathbf{b}^r_\mathbf{i} = \mathbf{b}^{r-1}_{\mathbf{i}+\mathbf{e}_j}$, i.e., the row opposite the corner $\mathbf{b}_{(n-r)\mathbf{e}_j}$ is deleted. We end with

$$\mathbf{r}(1,0,0) = \mathbf{b}_{n00}, \quad \mathbf{r}(0,1,0) = \mathbf{b}_{0n0}, \quad \mathbf{r}(0,0,1) = \mathbf{b}_{00n},$$

so we have *corner point interpolation*. We get the *boundary curves* by putting one of the barycentric coordinates equal to zero and then de Casteljau's algorithm reduces to the curve algorithm, e.g.:

$$\mathbf{b}^r_{i,0,r-i}(t, 0, 1-t) = (1-t)\mathbf{b}_{i,0,r-i+1} + t\mathbf{b}_{i+1,0,r-i}.$$

So the boundary curves are Bézier curves of degree n and the control polygons are the boundary polygons of the control net.

As $\mathbf{b}_{ijk} = \mathcal{C}_1^i\mathcal{C}_2^j\mathcal{C}_3^k\big(\mathbf{b}_\mathbf{i}\big)$ a point on the surface is given by

$$\begin{aligned}\mathbf{p}(\mathbf{u}) &= \mathcal{C}^n(\mathbf{u})\big(\mathbf{b}_\mathbf{i}\big) = \big(u\mathcal{C}_1 + v\mathcal{C}_2 + w\mathcal{C}_3\big)^n\big(\mathbf{b}_{ijk}\big)\\ &= \sum_{i+j+k=n} \frac{n!}{i!j!k!} u^i v^j w^k \mathcal{C}_1^i\mathcal{C}_2^j\mathcal{C}_3^k\big(\mathbf{b}_{ijk}\big)\\ &= \sum_{i+j+k=n} \frac{n!}{i!j!k!} u^i v^j w^k \mathbf{b}_{ijk} = \sum_{|\mathbf{i}|=n} \mathbf{b}_{ijk} B^n_\mathbf{i}(\mathbf{u})\end{aligned}$$

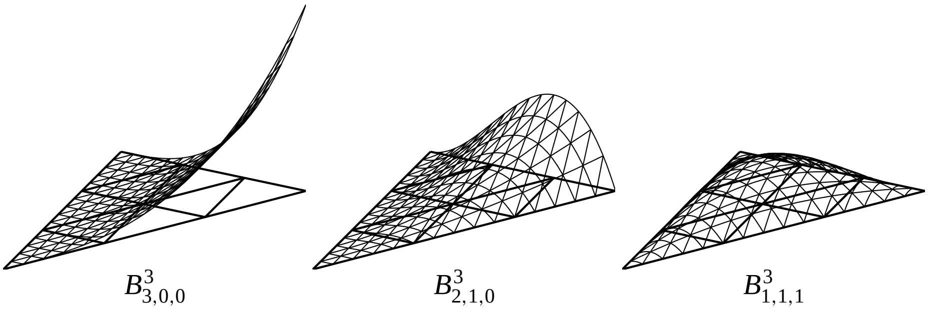


Figure 3.10: There are essentially three different Bernstein polynomials of degree 3.

where we have introduced the *Bernstein polynomials*

$$B^n_{\mathbf{i}}(\mathbf{u}) = \binom{n}{\mathbf{i}}\mathbf{u}^{\mathbf{i}} = \frac{n!}{i!j!k!}u^i v^j w^k,$$

see Figure 3.10.

3.4.1 Subdivision of a triangular Bézier surface

Let **a**, **b**, and **c** be three sets of barycentric coordinates corresponding to some triangle in the plane. We could use barycentric coordinates with respect to this new triangle instead and ask for the control points $\mathbf{b}^*_{\mathbf{i}}$ with respect to this triangle. If a point in the plane has barycentric coordinates $\mathbf{u} = (u, v, w)$ with respect to the new triangle then the barycentric coordinates with respect to the old triangle is $u\mathbf{a} + v\mathbf{b} + w\mathbf{c}$ so the point on the surface is

$$\begin{aligned}
\mathbf{r}(\mathbf{u}) &= \mathcal{C}^n(u\mathbf{a} + v\mathbf{b} + w\mathbf{c})\big(\mathbf{b}_{\mathbf{i}}\big) = \big(u\mathcal{C}(\mathbf{a}) + v\mathcal{C}(\mathbf{b}) + w\mathcal{C}(\mathbf{c})\big)^n\big(\mathbf{b}_{\mathbf{i}}\big) \\
&= \sum_{i+j+k=n} \frac{n!}{i!j!k!}u^i v^j w^k \mathcal{C}^i(\mathbf{a})\mathcal{C}^j(\mathbf{b})\mathcal{C}^k(\mathbf{c})\big(\mathbf{b}_{ijk}\big) \\
&= \sum_{|\mathbf{i}|=n} B^n_{\mathbf{i}}(\mathbf{u})\mathcal{C}^i(\mathbf{a})\mathcal{C}^j(\mathbf{b})\mathcal{C}^k(\mathbf{c})\big(\mathbf{b}_{\mathbf{i}}\big),
\end{aligned}$$

and we can read off the new control points

$$\mathbf{b}^*_{ijk} = \mathcal{C}^i(\mathbf{a})\mathcal{C}^j(\mathbf{b})\mathcal{C}^k(\mathbf{c})\big(\mathbf{b}_{\mathbf{i}}\big). \tag{3.23}$$

This is the direct generalisation of the *subdivision formula* for Bézier curves, see Figure 3.11.

When we run de Casteljau's algorithm for a point **P**, then we can pick up the control points for the restriction to the three triangles **ABP**, **BCP**, and **CAP**. They are

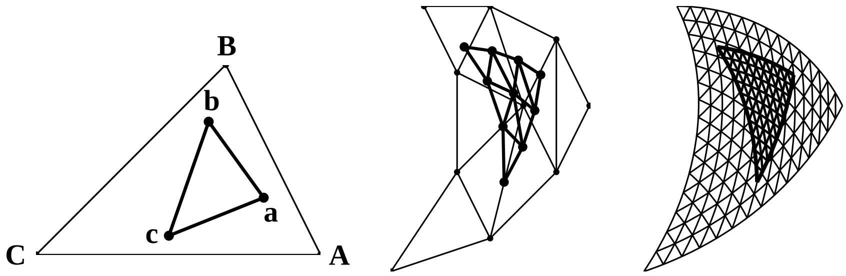


Figure 3.11: Subdivision of a triangular Bézier surface. To the left the parameter plane. We identify a set of barycentric coordinates $\mathbf{a} = (a_1, a_2, a_3)$ with the corresponding point $\mathbf{a} = a_1\mathbf{A} + a_2\mathbf{B} + a_3\mathbf{C}$. In the middle we have the control points, and to the right the surface.

$\mathbf{b}^k_{ij0}$, $\mathbf{b}^k_{0ij}$, and $\mathbf{b}^k_{j0i}$ respectively. If we arrange all the points in (3.21) in a tetrahedron, then one of the sides contains the old control points and the other three each contain the control points for one of the above mentioned restrictions, see Figure 3.12. If we recursively use this kind of subdivision then we never subdi-

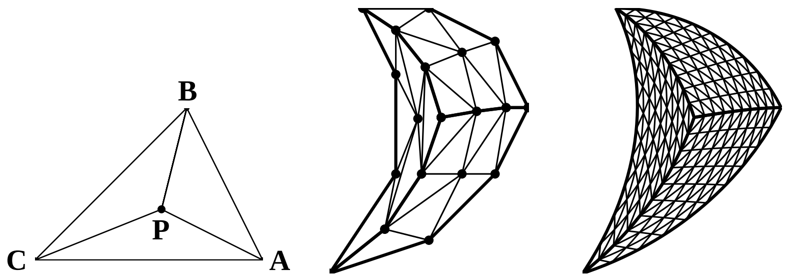


Figure 3.12: Subdividing a triangular Bézier surface by de Casteljau's algorithm.

vide the edges of the domain triangles so the domain triangles becomes thinner and thinner. This is not desirable, we prefer that all the domain triangles have the same shape. This can be achieved if we divide the triangle in four pieces as in Figure 3.13. The three patches at the corners can each be obtained by two succesive applications of de Casteljau's algorithm, and the middle patch can be obtained by three succesive applications of de Casteljau's algorithm, see Figure 3.14. The problem with middle patch is that we use extrapolation in the last of the three steps. If we look at (3.23) we see that it is possible to obtain the control points for the middle patch by using only interpolation, but then it's considerably harder to avoid repeating a calculation.

The straight line spanned by two points with barycentric cooordinates **a** and **b** can be considered as the boundary of a new domain triangle so *straight lines*

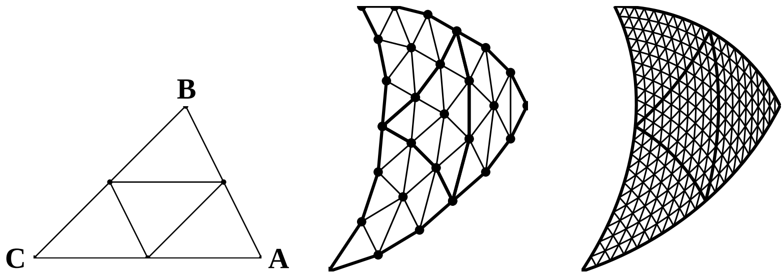


Figure 3.13: Subdividing a triangular Bézier surfaces in four equally sized pieces.

are mappped to Bézier curves of degree n. Still considering the straight line as a boundary (3.23) gives that the control points are $\mathcal{C}(\mathbf{a})^i\mathcal{C}^j(\mathbf{b})\big(\mathbf{b_i}\big)$, where $i + j = n$.

3.4.2 Differentation of a triangular Bézier surface

One way of looking at these surfaces and in particular at the apparently strange parametrization is to consider $\mathbf{r}(\mathbf{u})$ as the restriction of a tri-variate function to the plane $u + v + w = 1$, see Figure 3.15. The assignment of barycentric coordinates to a point is an isomorphism from our parameter plane to this plane in $\mathbb{R}^3$. It is now obvious that the normal partial derivatives $\partial_i\mathbf{r}$, e.g., $\partial_1\mathbf{r} = \frac{\partial\mathbf{r}}{\partial u}$, doesn't tell us anything about the behaviour of the function (or surface). Instead we will for any given vector $\overrightarrow{\mathbf{v}}$ in the plane determine the *directional* derivative

$$\partial_{\overrightarrow{\mathbf{v}}}\mathbf{r}(\mathbf{u}) = \frac{\mathrm{d}\mathbf{r}(\mathbf{u} + t\overrightarrow{\mathbf{v}})}{\mathrm{d}u}\bigg|_{t=0}.$$

The chain rule now yields the following formula:

$$\partial_{\overrightarrow{\mathbf{v}}}\mathbf{r}(\mathbf{u}) = \alpha\frac{\partial\mathbf{r}(\mathbf{u})}{\partial u} + \beta\frac{\partial\mathbf{r}(\mathbf{u})}{\partial v} + \gamma\frac{\partial\mathbf{r}(\mathbf{u})}{\partial w}.$$

As $\mathcal{C}(\mathbf{u}) = u\mathcal{C}_1 + v\mathcal{C}_2 + w\mathcal{C}_3$ we have

$$\frac{\partial\mathcal{C}(\mathbf{u})}{\partial u} = \mathcal{C}_1, \qquad \frac{\partial\mathcal{C}(\mathbf{u})}{\partial v} = \mathcal{C}_2, \qquad \frac{\partial\mathcal{C}(\mathbf{u})}{\partial w} = \mathcal{C}_3.$$

and as $\mathbf{r}(\mathbf{u}) = \mathcal{C}^n(\mathbf{u})\big(\mathbf{b_i}\big)$ and everything commute the usual rules for differentiating a product yields

$$\begin{aligned}\partial_{\overrightarrow{\mathbf{v}}}\mathbf{r}(\mathbf{u}) &= \alpha\frac{\partial\mathbf{r}(\mathbf{u})}{\partial u} + \beta\frac{\partial\mathbf{r}(\mathbf{u})}{\partial v} + \gamma\frac{\partial\mathbf{r}(\mathbf{u})}{\partial w}\\ &= n\big(\alpha\mathcal{C}_1 + \beta\mathcal{C}_2 + \gamma\mathcal{C}_3\big)\mathcal{C}^{n-1}(\mathbf{u})\big(\mathbf{b_i}\big)\\ &= n\mathcal{C}(\overrightarrow{\mathbf{v}})\mathcal{C}^{n-1}(\mathbf{u})\big(\mathbf{b_i}\big) = n\mathcal{C}^{n-1}(\mathbf{u})\mathcal{C}(\overrightarrow{\mathbf{v}})\big(\mathbf{b_i}\big).\end{aligned} \tag{3.24}$$

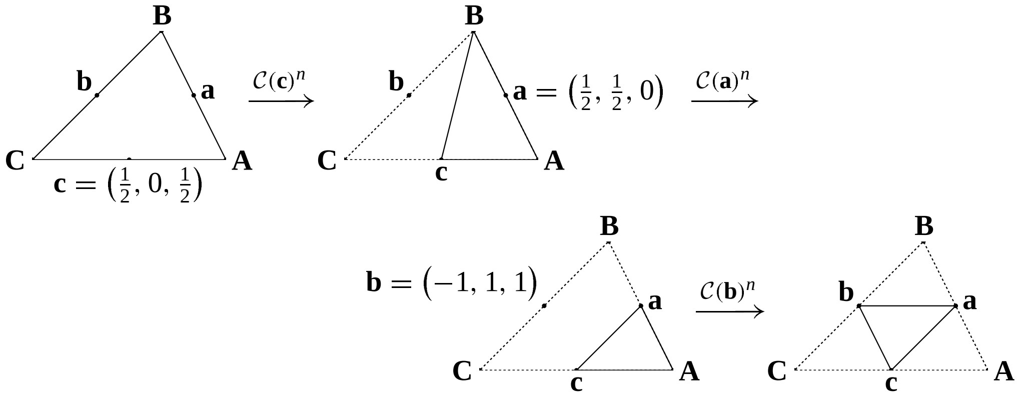


Figure 3.14: Running de Casteljau's algorithm for **c** (that has barycentric coordinates $\left(\frac{1}{2}, 0, \frac{1}{2}\right)$ with respect to **A**, **B**, **C**) gives as a byproduct the control points with respect to the domain triangle **A**, **B**, **c**. Running de Casteljau's algorithm once more for the point **a** gives the control points with respect to the domain triangle **A**, **a**, **c**. In a similar manner we can obtain the control points with respect to the domain triangles **B**, **b**, **a** and **C**, **c**, **b**. To get the control points with respect to the domain triangle **a**, **b**, **c** (the middle patch) we can take the control points with respect to the triangle **A**, **a**, **c** and run de Casteljau's algorithm for **b**, but as **b** has barycentric coordinates $(-1, 1, 1)$ with respect to **A**, **a**, **c**, we extrapolate.

Let us first see how $\mathcal{C}(\vec{\mathbf{v}}) = \alpha\mathcal{C}_1 + \beta\mathcal{C}_2 + \gamma\mathcal{C}_3$ acts on the triangular net. If we put $\left(\mathbf{d}_{ijk}\right) = \mathcal{C}(\vec{\mathbf{v}})\left(\mathbf{b_i}\right)$ then $\alpha + \beta + \gamma = 0$ and we have

$$\begin{aligned} \mathbf{d}_{ijk} &= \alpha\mathbf{b}_{i+1,j,k} + \beta\mathbf{b}_{i,j+1,k} + \gamma\mathbf{b}_{i,j,k+1} \\ &= \alpha(\mathbf{b}_{i+1,j,k} - \mathbf{b}_{i,j,k+1}) + \beta(\mathbf{b}_{i,j+1,k} - \mathbf{b}_{i,j,k+1}) \end{aligned} \tag{3.25}$$

I.e., $\mathcal{C}(\vec{\mathbf{v}})$ acts as a difference operator in the control net. Once again we are in the same situation as in the curve case. We can stop de Casteljau's algorithm at the second last stage and then obtain the directional derivative as a suitable difference in the triangle. Alternatively we can find the differences in the control net and obtain a new triangular Bézier surface that represents the directional derivative.

Combining the above and the result about the boundary curves we have that the directional derivative along an edge (*the "cross" boundary derivative*) is a Bézier

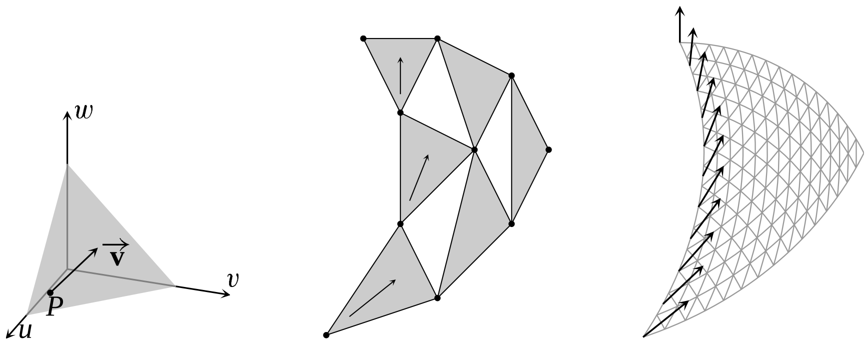


Figure 3.15: A point P and a vector $\vec{\mathbf{v}}$ in the plane $u + v + w = 1$. The vector $\vec{\mathbf{v}}$ is mapped linearly to the triangles along the edge and defines the the "cross" boundary derivative in the given direction.

curve of degree $n - 1$ and that the control points are given by

$$\begin{aligned}
\mathbf{d}_{0jk} &= \alpha \mathbf{b}_{1,j,k} + \beta \mathbf{b}_{0,j+1,k} + \gamma \mathbf{b}_{0,j,k+1} \\
&= \beta(\mathbf{b}_{0,j+1,k} - \mathbf{b}_{1,j,k}) + \gamma(\mathbf{b}_{0,j,k+1} - \mathbf{b}_{1,j,k}), \quad j + k = n - 1, \\
\mathbf{d}_{i0k} &= \alpha \mathbf{b}_{i+1,0,k} + \beta \mathbf{b}_{i,1,k} + \gamma \mathbf{b}_{i,j,k+1} \\
&= \alpha(\mathbf{b}_{i+1,0,k} - \mathbf{b}_{i,1,k}) + \gamma(\mathbf{b}_{i,j,k+1} - \mathbf{b}_{i,1,k}), \quad i + k = n - 1, \\
\mathbf{d}_{ij0} &= \alpha \mathbf{b}_{i+1,j,0} + \beta \mathbf{b}_{i,j+1,0} + \gamma \mathbf{b}_{i,j,1} \\
&= \alpha(\mathbf{b}_{i+1,j,0} - \mathbf{b}_{i,j,1}) + \beta(\mathbf{b}_{i,j+1,0} - \mathbf{b}_{i,j,1}), \quad i + j = n - 1.
\end{aligned}$$

Exercises

3.4.1 Implement de Casteljau's algorithm for triangular Bézier surfaces.

3.4.2 Implement a function that can find the Bézier curve corresponding to the "parameter curves" $u =$ constant, $v =$ constant, and $w =$ constant.

3.4.3 Write a program that determines a point and all partial derivatives to order k on a triangular Bézier surface.

3.4.4 Write a program that splits a triangular Bézier surface into three pieces as in Figure 3.12.

3.4.5 Write a program that splits a triangular Bézier surface into four pieces as in Figure 3.13.

Chapter 4

Differential Geometry of Surfaces

4.1 Introduction

Intuitively a surface is a two-dimensional object, i.e., an object that can be described by two parameters. One complication compared to curves is that a given pair of parameters in general only can be used locally in a part of the surface. Think of a world atlas, it takes more than one chart to describe the globe.

A particular local choice of coordinates is called a *coordinate patch* and it is simply a map from a portion of $\mathbb{R}^2$ into $\mathbb{R}^3$. We have in the previous chapter seen many examples of this.

4.2 Regular coordinate patches and the tangent plane

Just as for a curve we want parametrization with a suitable degree of differentiability and a regularity condition. For curves we wanted the derivatives to be nonzero, for surfaces the corresponding requirement is that the two partial derivatives are linearly independent. This can be formulated as the following definition.

Definition 4.1. A *coordinate patch* (or *chart* or *local parametrization*) of class C^k is a one-to-one C^k map $\mathbf{r} : U \to \mathbb{R}^3$, where U is an open subset of $\mathbb{R}^2$ with coordinates (u, v) and $\big(\partial \mathbf{r}/\partial u\big) \times \big(\partial \mathbf{r}/\partial v\big) \neq \mathbf{0}$ on U. The pair (u, v) are called *local parameters* or *local coordinates*, and the image of the lines $v =$ constant or $u =$ constant are called the first and second *parameter lines* respectively, see Figure 4.1.

Observe that if $\mathbf{r} : U \to \mathbb{R}^3$ is a coordinate patch of class C^k, and $V \subseteq U$ is and open subset of U, then $\mathbf{r}_{|V} : V \to \mathbb{R}^3$ is a coordinate patch of class C^k too.

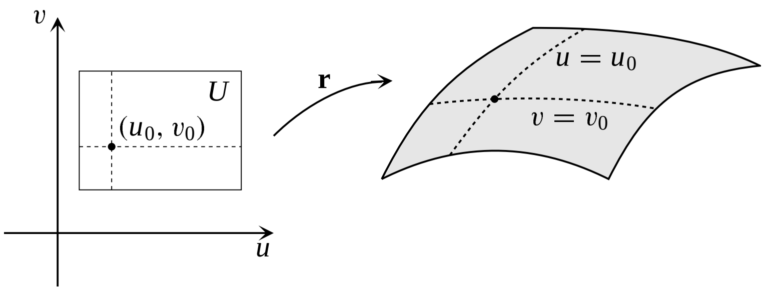


Figure 4.1: A coordinate patch and two parameter lines.

Example 4.1 The simplest example of a coordinate patch is the parametrization of a plane $\mathbf{r}(u, v) = (u, v, 0)$, $(u, v) \in \mathbb{R}^2$. The partial derivatives are $r_u(u, v) = (1, 0, 0)$ and $r_v(u, v) = (0, 1, 0)$, so the cross product is $\mathbf{r}_u \times \mathbf{r}_v = (0, 0, 1) \neq \mathbf{0}$ for all $(u, v) \in \mathbb{R}^2$.

Example 4.2 Consider the unit sphere $S^2 \subseteq \mathbb{R}^3$, a line through the north pole $(0, 0, 1)$, and a point $(u, v, 0)$ in the xy-plane. The line can be parametrized as $t \mapsto (tu, tv, 1 - t)$ and the square of the distance to the origin is $t^2\big(u^2+v^2+1\big)-2t+1$. The line intersects the unit sphere in points where the distance to the origin is 1, i.e., when $t^2\big(u^2+v^2+1\big)-2t+1 = 1$ or equivalently when $2t = t^2\big(u^2+v^2+1\big)$. One solution to this quadratic equation is $t = 0$ which corresponds to the north pole, and the other is $t = 2/\big(u^2+v^2+1\big)$ which corresponds to the point $(x, y, z) = \big(2u, 2v, u^2+v^2-1\big)/\big(u^2+v^2+1\big)$, see Figure 4.2. The opposite map $(x, y, x) \mapsto (u, v)$ is called *stereographic projection* from the north pole and defines a map from $\mathbb{R}^2$ to S^2:

$$\mathbf{r}^+(u, v) = \frac{\big(2u, 2v, u^2+v^2-1\big)}{u^2+v^2+1}, \qquad (u, v) \in \mathbb{R}^2. \tag{4.1}$$

We find

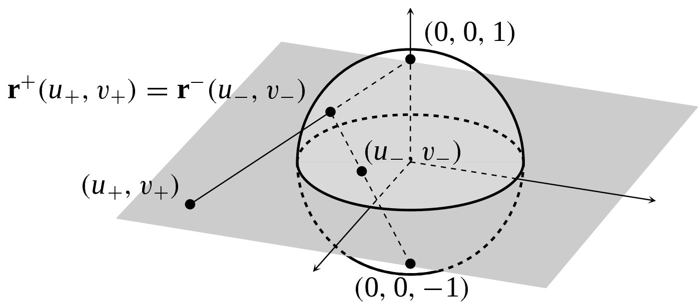


Figure 4.2: Stereographic projection from the north and the south pole.

$$\mathbf{r}^+_u = 2\frac{\big(1-u^2+v^2, -2uv, 2u\big)}{\big(1+u^2+v^2\big)^2},$$
$$\mathbf{r}^+_v = 2\frac{\big(-2uv, 1+u^2-v^2, 2v\big)}{\big(1+u^2+v^2\big)^2},$$

and

$$\mathbf{r}^+_u \times \mathbf{r}^+_v = 4\frac{\big(-2u, -2v, 1-u^2-v^2\big)}{\big(1+u^2+v^2\big)^3} \neq \mathbf{0}, \quad \text{for all } (u,v) \in \mathbb{R}^2,$$

so $\mathbf{r}^+$ is a coordinate patch. In a similar way we define stereographic projection from the south pole and the opposite map gives us another coordinate patch

$$\mathbf{r}^-(u,v) = \frac{\big(2u, 2v, 1-u^2-v^2\big)}{u^2+v^2+1}, \qquad (u,v) \in \mathbb{R}^2, \tag{4.2}$$

cf. Problem 4.2.3. The image of $\mathbf{r}^+$ is $V^+ = S^2 \setminus \{(0,0,1)\}$ and the image of $\mathbf{r}^-$ is $V^- = S^2 \setminus \{(0,0,-1)\}$. So the region $S^2 \setminus \{(0,0,\pm 1)\} = V^+ \cap V^-$ is parametrized in two different ways. If $\mathbf{r}^+(u_+, v_+) = \mathbf{r}^-(u_-, v_-)$ then

$$(u_+, v_+) = \frac{(u_-, v_-)}{u_-^2 + v_-^2}, \qquad (u_-, v_-) = \frac{(u_+, v_+)}{u_+^2 + v_+^2} \tag{4.3}$$

cf. Problem 4.2.5. This is an example of a change of coordinates, cf. Definition 4.2

Definition 4.2. An *allowable change of coordinates* of class C^k is bijective map $f : U \to V$ of open sets in $\mathbb{R}^2$ such that both f and its inverse is of class C^k.

If $f : U \to V$ is an allowable change of coordinates and $(s,t) = f(u,v)$ then we will often write $(s,t) = \big(s(u,v), t(u,v)\big)$ for f, and $(u,v) = \big(u(s,t), v(s,t)\big)$ for the opposite map. So with an abuse of notation we let (u,v) denote both a point in $U \subseteq \mathbb{R}^2$ and a pair of functions. This is the same terminology we used for reparametrization of curves.

If (u,v) and (s,t) are coordinates related by an allowable change of coordinates then by the chain rule we have

$$\begin{bmatrix} \frac{\partial u}{\partial s} & \frac{\partial u}{\partial t} \\ \frac{\partial v}{\partial s} & \frac{\partial v}{\partial t} \end{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial s}\frac{\partial s}{\partial u} + \frac{\partial u}{\partial t}\frac{\partial t}{\partial u} & \frac{\partial u}{\partial s}\frac{\partial s}{\partial v} + \frac{\partial u}{\partial t}\frac{\partial t}{\partial v} \\ \frac{\partial v}{\partial s}\frac{\partial s}{\partial u} + \frac{\partial v}{\partial t}\frac{\partial t}{\partial u} & \frac{\partial v}{\partial s}\frac{\partial s}{\partial v} + \frac{\partial v}{\partial t}\frac{\partial t}{\partial v} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial u} & \frac{\partial u}{\partial v} \\ \frac{\partial v}{\partial u} & \frac{\partial v}{\partial v} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We have in particular that

$$\det \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v} \end{bmatrix} \neq 0.$$

If the determinant is positive then the change of coordinates is called *orientation preserving* otherwise it is called *orientation reversing*.

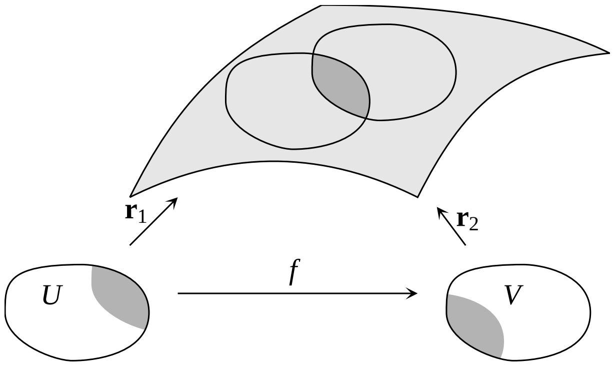


Figure 4.3: Two smoothly overlapping coordinate patches.

Definition 4.3. Two coordinate patches $\mathbf{r}_1 : U \to \mathbb{R}^3$ and $\mathbf{r}_2 : V \to \mathbb{R}^3$ of class C^k *overlap smoothly* if $\mathbf{r}_1^{-1}\big(\mathbf{r}_2(V)\big)$ and $\mathbf{r}_2^{-1}\big(\mathbf{r}_1(U)\big)$ are open subsets of $\mathbb{R}^2$, and there exists an allowable change of coordinates $f : \mathbf{r}_1^{-1}\big(\mathbf{r}_2(V)\big) \to \mathbf{r}_2^{-1}\big(\mathbf{r}_1(U)\big)$ such that $\mathbf{r}_1(u, v) = \mathbf{r}_2\big(f(u, v)\big)$ for all $(u, v) \in \mathbf{r}_1^{-1}\big(\mathbf{r}_2(V)\big)$. Furthermore, if the change of coordinates is orientation preserving then we say that the two coordinate patches defines the same orientation in the overlap.

There does not exits a coordinate patch that covers all of the sphere, so a reasonable definition of a surface needs to involve more than one coordinate patch.

Definition 4.4. A subset $M \subseteq \mathbb{R}^3$ is a *regular surface* of class C^k, if there for each $P \in M$ exits a open set $V \subseteq \mathbb{R}^3$, and a coordinate patch $\mathbf{r} : U \to \mathbb{R}^3$ of class C^k such that $M \cap V = \mathbf{r}(U)$; and if $\mathbf{r}_1 : U \to \mathbb{R}^3$ and $\mathbf{r}_2 : V \to \mathbb{R}^3$ are two such coordinate patches with $\mathbf{r}_1(U) \cap \mathbf{r}_2(V) \neq \emptyset$, then they overlap smoothly.

Definition 4.5. A regular surface is called *orientable* if the coordinate patches can be chosen such that they pairwise defines the same orientation in the overlap.

A particular choice of such coordinate patches is called an *orientation* of the surface.

Example 4.3 The coordinate patches $\mathbf{r}^+$ or $\mathbf{r}^-$ from Example 4.2 overlap smoothly and as any point on the unit sphere S^2 is in the image of $\mathbf{r}^+$ or $\mathbf{r}^-$ S^2 is a regular surface. From (4.3) we see that

$$\det\begin{bmatrix} \frac{\partial u_+}{\partial u_-} & \frac{\partial u_+}{\partial v_-} \\ \frac{\partial v_+}{\partial u_-} & \frac{\partial v_+}{\partial v_-} \end{bmatrix} = \det\begin{bmatrix} \frac{v_-^2-u_-^2}{(u_-^2+v_-^2)^2} & \frac{-2u_-v_-}{(u_-^2+v_-^2)^2} \\ \frac{-2u_-v_-}{(u_-^2+v_-^2)^2} & \frac{u_-^2-v_-^2}{(u_-^2+v_-^2)^2} \end{bmatrix} = \frac{-(u_-^2-v_-^2)^2-4u_-^2v_-^2}{(u_-^2+v_-^2)^4} = \frac{-1}{(u_-^2+v_-^2)^2}.$$

The change of coordinates is not orientation preserving, but if we define a new coordinate patch $\widetilde{\mathbf{r}}$ by $\widetilde{\mathbf{r}}(u, v) = \mathbf{r}^-(v, u)$, then we easily see that $\mathbf{r}^+$ and $\widetilde{\mathbf{r}}$ overlaps smoothly and defines the same orientation in the overlap, so S^2 is orientable.

4.2.1 The tangent plane

Definition 4.6. Let $M \subseteq \mathbb{R}^3$ be a regular surface of class C^k. A *regular curve of class C^ℓ on M* is a space curve $\mathbf{x} : I \to M \subseteq \mathbb{R}^3$ such that there for each $t_0 \in I$ exists an open subinterval $t_0 \in J \subseteq I$, a coordinate patch $\mathbf{r} : U \to \mathbb{R}^3$ on M, and a regular curve $t \mapsto \big(u(t), v(t)\big) \in U$ with $\mathbf{x}(t) = \mathbf{r}\big(u(t), v(t)\big)$ for all $t \in J$.

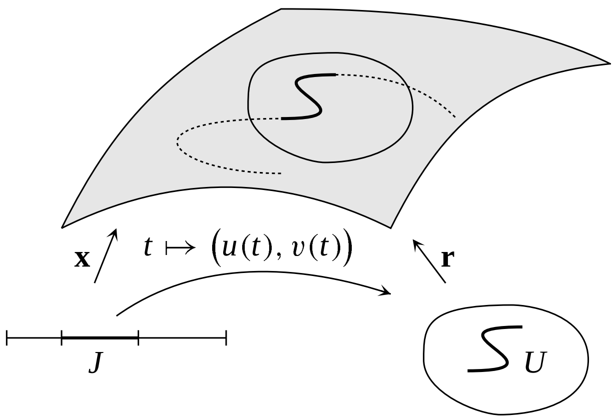


Figure 4.4: A regular curve on a surface

The pair of functions $t \mapsto \big(u(t), v(t)\big)$ is called a *local representation* of the curve. If a curve on M is smooth in the sense above then clearly it is also smooth as a space curve. The opposite is also true, cf. Lemma 4.11.

Let $\mathbf{x}(t) = \mathbf{r}\big(u(t), v(t)\big)$ be a parametrization of a smooth curve on a regular surface M. The velocity vector is then given by

$$\mathbf{x}'(t) = u'(t)\mathbf{r}_u\big(u(t), v(t)\big) + v'(t)\mathbf{r}_v\big(u(t), v(t)\big), \tag{4.4}$$

where we have used the notation

$$\mathbf{r}_u = \frac{\partial \mathbf{r}}{\partial u} \qquad \text{and} \qquad \mathbf{r}_v = \frac{\partial \mathbf{r}}{\partial v}.$$

We see in particular that two smooth curves on a surface have the same velocity vector at a common point if and only the local representations have the same velocity vector.

Definition 4.7. Let M be a regular surface and let $P \in M$. The *tangent space* $T_P M$ of M at P consist of all velocity vectors at P to curves on M through P. If $\mathbf{r} : U \to M$ is a coordinate patch and $P = \mathbf{r}(u_0, v_0)$ then (4.4) shows that

$$T_P M = \big\{ t_1 \mathbf{r}_u(u_0, v_0) + t_2 \mathbf{r}_v(u_0, v_0) \;\big|\; t_1, t_2 \in \mathbb{R} \big\}.$$

The *tangent plane* to M at P consist of all points $Q \in \mathbb{R}^3$ such that $\overrightarrow{PQ} \in T_P M$, i.e., it is the plane $\left\{P + t_1\mathbf{r}_u(u_0, v_0) + t_2\mathbf{r}_v(u_0, v_0) \;\middle|\; t_1, t_2 \in \mathbb{R}\right\}$.

As $\mathbf{r}_u$ and $\mathbf{r}_v$ are linearly independent for a regular surface we see that $T_P M$ is a two dimensional vector space with basis $\mathbf{r}_u, \mathbf{r}_v$.

If (u, v) and (s, t) are two set of local coordinates related by an allowable change of coordinates then we see that

$$\begin{bmatrix}\mathbf{r}_u\\ \mathbf{r}_v\end{bmatrix} = \begin{bmatrix}\frac{\partial s}{\partial u} & \frac{\partial t}{\partial u}\\ \frac{\partial s}{\partial v} & \frac{\partial t}{\partial v}\end{bmatrix}\begin{bmatrix}\mathbf{r}_s\\ \mathbf{r}_t\end{bmatrix} \tag{4.5}$$

and if $\mathbf{v} = a\mathbf{r}_u + b\mathbf{r}_v = c\mathbf{r}_s + d\mathbf{r}_t$ are expansions of a tangent vector $\mathbf{v}$ with respect to the two bases of $T_P M$, then

$$\begin{bmatrix}c\\ d\end{bmatrix} = \begin{bmatrix}\frac{\partial s}{\partial u} & \frac{\partial s}{\partial v}\\ \frac{\partial t}{\partial u} & \frac{\partial t}{\partial v}\end{bmatrix}\begin{bmatrix}a\\ b\end{bmatrix} \tag{4.6}$$

Definition 4.8. Let M be a regular surface and let $P \in M$. A *normal vector* to at P is a vector that is orthogonal to $T_P M$. If $\mathbf{r} : U \to M$ is a coordinate patch and $P = \mathbf{r}(u_0, v_0)$ then $\mathbf{r}_u(u_0, v_0) \times \mathbf{r}_v(u_0, v_0)$ is a normal vector, and all others are proportional to this. By normalization we get a *unit normal vector*:

$$\mathbf{N} = \frac{\mathbf{r}_u \times \mathbf{r}_v}{\left|\mathbf{r}_u \times \mathbf{r}_v\right|} \tag{4.7}$$

The unit normal in (4.7) are often referred to as *the* normal vector, but it is only defined up to a sign, see Problem 4.2.11.

Example 4.4 Consider the coordinate patch (4.1) on the unit sphere. If we normalize $\mathbf{r}_u^+ \times \mathbf{r}_v^+$ we obtain

$$\mathbf{N}^+ = -\frac{\left(2u, 2v, u^2 + v^2 - 1\right)}{1 + u^2 + v^2} = -\mathbf{r}^+(u, v),$$

(the inward normal). Similar (4.2) yields the unit normal vector

$$\mathbf{N}^- = \frac{\left(2u, 2v, 1 - u^2 - v^2\right)}{1 + u^2 + v^2} = \mathbf{r}^-(u, v),$$

(the outward normal on the sphere).

The above example illustrates Problem 4.2.11, an allowable change of coordinates is orientation preserving if $\mathbf{N}$ is preserves and it is orientation reversing if $\mathbf{N}$ changes sign. This obvious gives the following result, cf. Problem 4.2.12

Theorem 4.9. *For a regular surface the following three statements are equivalent.*

1. *The surface is orientable.*
2. *The surface posses a continuous unit normal vectorfield.*
3. *The surface posses a continuous non vanishing normal vectorfield.*

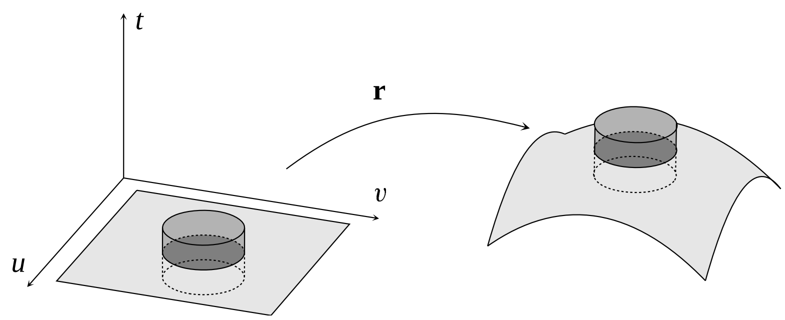


Figure 4.5: A coordinate patch $\mathbf{r} : U \to M$ can be extended to a local diffeomorphism $\widetilde{\mathbf{r}} : \widetilde{V} \to \mathbb{R}^3$.

Lemma 4.10. *Let M be a regular surface of class C^k, let $\mathbf{r} : U \to M$ be a coordinate patch and let $(u_0, v_0) \in U$. Then there exists an open set $V \subseteq U$ and an injective C^k map $\widetilde{\mathbf{r}} : \widetilde{V} = V \times]-\varepsilon, \varepsilon[\to \mathbb{R}^3$ whos inverse is C^k too, such that $\mathbf{r}(u, v) = \widetilde{\mathbf{r}}(u, v, 0)$ for all $(u, v) \in V$, see Figure 4.5.*

Proof. Let $\mathbf{N}(u, v) = \mathbf{r}_u(u, v) \times \mathbf{r}_v(u, v)/|\mathbf{r}_u(u, v) \times \mathbf{r}_v(u, v)|$ be the unit normal and define $\widetilde{\mathbf{r}}(u, v, t) = \mathbf{r}(u, v) + t\mathbf{N}(u_0, v_0)$. This is clearly a C^k map. Furthermore, the partial derivatives at $(u_0, v_0, 0)$ are $\mathbf{r}_u(u_0, v_0)$, $\mathbf{r}_u(u_0, v_0)$, and $\mathbf{N}(u_0, v_0)$ respectively and they are obviously are linearly independent. The lemma is now a consequence of the inverse function theorem. □

Lemma 4.11. *Consider a regular surface $M \subseteq \mathbb{R}^3$ of class C^k, and let $\ell \leq k$. If f is a map $W \to M$ where $W \subseteq \mathbb{R}^n$ is an open set, then the following three statements are equivalent.*

1. *Considered as map $W \to \mathbb{R}^3$, f is C^ℓ.*
2. *For each coordinate patch $\mathbf{r} : U \to M$ the map $\mathbf{r}^{-1} \circ f_{|f^{-1}(\mathbf{r}(U))}$ is C^ℓ.*
3. *For each point $P \in M$ exists a coordinate patch $\mathbf{r} : U \to M$ with $P \in \mathbf{r}(U)$ such that the map $\mathbf{r}^{-1} \circ f_{|f^{-1}(\mathbf{r}(U))}$ is C^ℓ.*

Proof. 1 $\implies$ 2: Let $\mathbf{r} : U \to M$ be a coordinate patch, and consider a point $P \in f^{-1}(\mathbf{r}(U))$. By Lemma 4.10 we can extend $\mathbf{r}$ to a map $\widetilde{\mathbf{r}} : \widetilde{V} \to \mathbb{R}^3$, where $\widetilde{V}$ is an open set in $\mathbb{R}^3$ and $P \in \widetilde{\mathbf{r}}(\widetilde{V})$. The inverse $\widetilde{\mathbf{r}}^{-1} : \widetilde{\mathbf{r}}(\widetilde{V}) \to \mathbb{R}^3$ is of class C^k and as $\widetilde{\mathbf{r}}^{-1} \circ f_{|f^{-1}(\mathbf{r}(V))} = \mathbf{r}^{-1} \circ f_{|f^{-1}(\mathbf{r}(U))}$ the latter is of class C^ℓ.

2 $\implies$ 3: Is trivial.

3 $\implies$ 1: Let $P \in W$ and choose a coordinate patch $\mathbf{r} : U \to M$ with $f(P) \in \mathbf{r}(U)$ such that the map $\mathbf{r}^{-1} \circ f_{|f^{-1}(\mathbf{r}(U))}$ is C^ℓ. We have $P \in f^{-1}(\mathbf{r}(U))$ and as $f_{|f^{-1}(\mathbf{r}(U))} = \mathbf{r} \circ \mathbf{r}^{-1} \circ f_{|f^{-1}(\mathbf{r}(U))}$ we can conclude that f is C^ℓ in $f^{-1}(\mathbf{r}(U))$. As P was arbitray we are done. □

Definition 4.12. A map $f : W \to M$ that satisfies one the three equivalent conditions in Lemma 4.11 is called a C^ℓ map. If $\mathbf{r} : U \to M$ is a coordinate patch and $f(\mathbf{u}) = \mathbf{r}(f_1(\mathbf{u}), f_2(\mathbf{u}))$ for all $\mathbf{u} \in f^{-1}(\mathbf{r}(U))$ then $(f_1(\mathbf{u}), f_2(\mathbf{u}))$ is called the *local expression* of f.

Similar we have

Lemma 4.13. *Consider a regular surface $M \subseteq \mathbb{R}^3$ of class C^k, and let $\ell \le k$. If f is a map $M \to \mathbb{R}^n$, then the following two statements are equivalent.*

1. *For each coordinate patch $\mathbf{r} : U \to M$ the map $f \circ \mathbf{r} : U \to \mathbb{R}^n$ is C^ℓ.*
2. *For each point $P \in M$ exists a coordinate patch $\mathbf{r} : U \to M$ with $P \in \mathbf{r}(U)$ such that the map $f \circ \mathbf{r} : U \to \mathbb{R}^n$ is C^ℓ.*

Proof. 1 $\implies$ 2 is trivial, so we need only consider 2 $\implies$ 3. Let $\mathbf{r} : U \to M$ be a coordinate patch and let $(u_0, v_0) \in U$. Now choose a coordinate patch $\mathbf{r}_1 : U_1 \to M$ with $\mathbf{r}(u_0, v_0) \in \mathbf{r}_1(U_1)$ such that the map $f \circ \mathbf{r}_1 : U_1 \to \mathbb{R}^n$ is C^ℓ. Let $g : \mathbf{r}^{-1}(\mathbf{r}_1(U_1)) \to U_1$ be an allowable change of coordinates such that $\mathbf{r}(u, v) = \mathbf{r}_1(g(u, v))$ for all $(u, v) \in \mathbf{r}^{-1}(\mathbf{r}_1(U_1))$. Then $(u_0, v_0) \in \mathbf{r}^{-1}(\mathbf{r}_1(U_1))$ and $f \circ \mathbf{r}(u, v) = f \circ \mathbf{r}_1 \circ g(u, v)$ for all $(u, v) \in \mathbf{r}^{-1}(\mathbf{r}_1(U_1))$, and we see that $f \circ \mathbf{r}$ is C^ℓ in $\mathbf{r}^{-1}(\mathbf{r}_1(U_1))$. □

Definition 4.14. A map $f : M \to \mathbb{R}^n$ that satisfies one the two equivalent conditions in Lemma 4.13 is called a C^ℓ map. If $\mathbf{r} : U \to M$ is a coordinate patch then $(u, v) \mapsto f(\mathbf{r}(u, v))$ is called the *local expression* of f.

Finally we consider maps between two surfaces.

Lemma 4.15. *Consider two regular sufaces $M_1, M_2 \subseteq \mathbb{R}^3$ of class C^k and let $\ell \le k$. If $f : M_1 \to M_2$, then the following three statements are equivalent.*

1. *Considered as map $M_1 \to \mathbb{R}^3$, f is C^ℓ.*

2. *For each pair of coordinate patches* $\mathbf{r}_i : U_i \to M_i$, $i = 1, 2$, *the map* $\mathbf{r}_2^{-1} \circ f \circ \mathbf{r}_{1|(f\circ\mathbf{r}_1)^{-1}(\mathbf{r}_2(U_2))}$ *is* C^ℓ.
3. *For each* $P \in M_1$ *exists coordinate patches* $\mathbf{r}_i : U_i \to M_i$, $i = 1, 2$, *such that* $P \in \mathbf{r}_1(U_1)$, $f(P) \in \mathbf{r}_2(U_2)$ *and* $\mathbf{r}_2^{-1} \circ f \circ \mathbf{r}_{1|(f\circ\mathbf{r}_1)^{-1}(\mathbf{r}_2(U_2))}$ *is* C^ℓ.

Definition 4.16. A map $f : M_1 \to M_2$ that satisfies one the three equivalent conditions in Lemma 4.15 is called a C^ℓ map. If $\mathbf{r}_i : U_i \to M_i$, $i = 1, 2$, are coordinate patches and $f(\mathbf{r}_1(u, v)) = \mathbf{r}_2(f_1(u, v), f_2(u, v))$ for all $(u, v) \in (f \circ \mathbf{r}_1)^{-1}(\mathbf{r}_2(U_2))$, then $(f_1(u, v), f_2(u, v))$ is called the *local expression* of f.

If f is bijective and the inverse map is a C^k map too, then f is called a *diffeomorphism*.

Let $\mathbf{r}_i : U_i \to M_i$, $i = 1, 2$ be coordinate patches for two regular surfaces and let $f : M_1 \to M_2$ be a smooth map with the local expression $(f_1(u, v), f_2(u, v))$. If $\gamma : I \to M_1$ is a smooth curve in M_1 with the local expression $(\gamma_1(t), \gamma_2(t))$, then $f \circ \gamma : I \to M_2$ is a smooth curve in M_2 and the local expression is $(f_1(\gamma_1(t), \gamma_2(t)), f_2(\gamma_1(t), \gamma_2(t)))$. Differentiating one of the coordinates yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\big(f_i(\gamma_1(t), \gamma_2(t))\big) = \frac{\partial f_i}{\partial u}(\gamma_1(t), \gamma_2(t))\,\gamma_1'(t) + \frac{\partial f_i}{\partial v}(\gamma_1(t), \gamma_2(t))\,\gamma_2'(t)$$

In other words, with respect to the basis $\mathbf{r}_{1u}, \mathbf{r}_{1v}$ of $T_{\gamma(t)}M_1$ and the basis $\mathbf{r}_{2u}, \mathbf{r}_{2v}$ of $T_{f(\gamma(t))}M_2$ the tangent vectors $\gamma'(t) \in T_{\gamma(t)}M_1$ and $(f \circ \gamma)'(t) \in T_{f(\gamma(t))}M_2$ has the coordinates

$$\begin{bmatrix}\gamma_1'(t)\\ \gamma_2'(t)\end{bmatrix} \quad\text{and}\quad \begin{bmatrix}\frac{\partial f_1}{\partial u}(\gamma_1(t), \gamma_2(t)) & \frac{\partial f_1}{\partial v}(\gamma_1(t), \gamma_2(t))\\ \frac{\partial f_2}{\partial u}(\gamma_1(t), \gamma_2(t)) & \frac{\partial f_2}{\partial v}(\gamma_1(t), \gamma_2(t))\end{bmatrix}\begin{bmatrix}\gamma_1'(t)\\ \gamma_2'(t)\end{bmatrix}$$

respectively. If we put $t = 0$, then the last expression only depends on $\gamma(0)$ and $\gamma'(0)$, but not on any higher order derivatives. I.e., if $\mu : I \to M_1$ is an other smooth curve with $\mu(0) = \gamma(0)$ and $\mu'(0) = \gamma'(0)$, then $(f\circ\mu)'(0) = (f\circ\gamma)'(0)$.

Definition 4.17. Let $M_1, M_2 \subseteq \mathbb{R}^3$ be two regular surfaces and let $P \in M_1$. The *differential* of a smooth map $f : M_1 \to M_2$ is a linear map

$$\mathrm{d}f_P : T_P M_1 \to T_{f(P)}M_2 : \gamma'(0) \mapsto (f \circ \gamma)'(0)$$

where $\gamma : I \to M_1$ is a smooth curve with $\gamma(0) = P$.

Furthermore, let $\mathbf{r}_i : U_i \to M_i$, $i = 1, 2$ be coordinate patches and suppose the local expression for f is $(f_1(u, v), f_2(u, v))$. If we for a tangent vector $\mathbf{v} \in T_{\mathbf{r}_1(u,v)}M_1$ and the image $\mathrm{d}f_{\mathbf{r}_1(u,v)}(\mathbf{v}) \in T_{f(\mathbf{r}_1(u,v))}M_2$ has the expansions

$$\mathbf{v} = v^1\mathbf{r}_{1u} + v^2\mathbf{r}_{1v} \quad\text{and}\quad \mathrm{d}f_{\mathbf{r}_1(u,v)}(\mathbf{v}) = w^1\mathbf{r}_{2u} + w^2\mathbf{r}_{2v}$$

respectively, then we have the following matrix expression

$$\begin{bmatrix} w^1 \\ w^2 \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial u} & \frac{\partial f_1}{\partial v} \\ \frac{\partial f_2}{\partial u} & \frac{\partial f_2}{\partial v} \end{bmatrix} \begin{bmatrix} v^1 \\ v^2 \end{bmatrix}.$$

Theorem 4.18 (The inverse function theorem). *Let $f : M_1 \to M_2$ be a C^ℓ map between regular C^k surfaces. If, for a point $P \in M_1$, the differential $\mathrm{d} f_P : T_P M_1 \to T_{f(P)} M_2$ is regular, then there exists open neighbourhoods $P \in V_1 \subseteq M_1$ and $f(P) \in V_2 \subseteq M_2$ of P and $f(P)$ respectively, such that $f_{|V}$ is a diffeomorphism $V \to W$ of class C^ℓ, i.e., f is a* local diffeomorphism *of class C^ℓ.*

Proof. Choose coordinate patches $\mathbf{r}_i : U_i \to M_i$, $i = 1, 2$, with $P \in \mathbf{r}_1(U_1)$ and $f(P) \in \mathbf{r}_2(U_2)$. Let (f_1, f_2) be the local expression for f, such that we have the following picture

$$\begin{array}{ccc} M_1 & \xrightarrow{\ f\ } & M_2 \\ \uparrow \mathbf{r}_1 & & \uparrow \mathbf{r}_2 \\ U_1 & \xrightarrow{(f_1, f_2)} & U_2 \end{array}$$

The regularity od $\mathrm{d} f_P$ implies the regularity of the Jacobian of (f_1, f_2) so by the usual inverse function theorem we can choose a smaller U_1 (and U_2) such that (f_1, f_2) is a diffeomorphism class C^ℓ. If we put $V_i = \mathbf{r}_i(U_i)$, $i = 1, 2$, then $f_{|V_1} : V_1 \to V_2$ is a diffeomorphism class C^ℓ. □

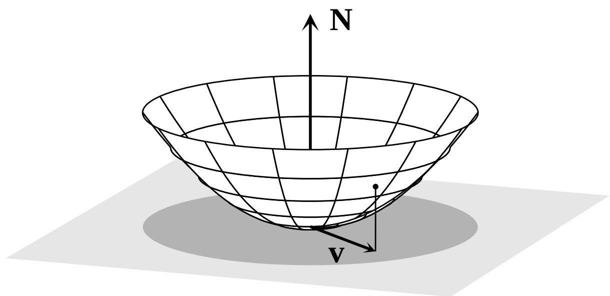


Figure 4.6: A regular surface is locally the graph of a smooth function "from" the tangent plane "to" the normal line.

Theorem 4.19. *Let P be a point om a regular surface M of class C^k, let $\mathbf{v}_1, \mathbf{v}_2$ be a basis for the tangent space $T_P M$ and let $\mathbf{N}$ be the unit normal vector at P. Then there exists a neighbourhood U of $\mathbf{0} \in \mathbb{R}^2$ and a C^k-function $f : U \to \mathbb{R}$ such that the map*

$$U \to \mathbb{R}^3 : (u, v) \mapsto P + u\mathbf{v}_1 + v\mathbf{v}_2 + f(u, v)\mathbf{N}$$

is a coordinate patch of class C^k on M. Furthermore, $f(\mathbf{0}) = \frac{\partial f}{\partial s}(\mathbf{0}) = \frac{\partial f}{\partial t}(\mathbf{0}) = 0$.

Proof. The orthogonal projection $\mathbb{R}^3 \to T_P M$ is a C^∞ map so the restricion $M \to T_P M$ is a C^k map. The differential at P is clearly the identity map $T_P M \to T_{\mathbf{0}}(T_P M) \cong T_P M$, in particular regular. BY the inverse function theorem it locally has a C^k inverse and if we compose this inverse with the map $\mathbb{R}^2 \to T_P M : (u, v) \mapsto u\mathbf{v}_1 + v\mathbf{v}_2$ we obtain a map $U \to M \subseteq \mathbb{R}^3$ of the required form. □

Observe that f measures the distance from the given point on the surface to the tangent plane.

Problems

4.2.1 Let $(x, y, x) \mapsto (u, v)$ be the stereographic projection from the north pole, cf. Example 4.2 and Figure 4.2. Determine (u, v) as a function of (x, y, z).

4.2.2 Let $(x, y, x) \mapsto (u, v)$ be the stereographic projection from the south pole, cf. Example 4.2 and Figure 4.2. Determine (u, v) as a function of (x, y, z).

4.2.3 Find the opposite map, $\mathbf{r}^- : \mathbb{R}^2 \to S^2 \subseteq \mathbb{R}^3$, of stereographic projection from the south pole, cf. Example 4.2, Figure 4.2, and Problem 4.2.2. Show that $\mathbf{r}^-$ is a coordinate patch of class C^∞.

4.2.4 Find the parameter lines for the two coordinate patches (4.1) and (4.2).

4.2.5 Let $\mathbf{r}^+$ and $\mathbf{r}^-$ be the two coordinate patches (4.1) and (4.1). Show that $\mathbf{r}^+(u^+, v^+) = \mathbf{r}^-(u^-, v^-)$ if and only if (4.3) holds and that (4.3) defines an allowable change of coordinates.

4.2.6 Show that *spherical coordinates*

$$\mathbf{r}(u, v) = (\sin u \ \cos v, \sin u \ \sin v, \cos u), \quad (u, v) \in (0, \pi) \times (-\pi, \pi)$$

is a coordinate patch on the unit sphere, see Figure 4.7. What are the parameter lines?

4.2.7 Show that

$$\mathbf{r}(u, v) = (r \cos v, r \sin v, u), \quad (u, v) \in \mathbb{R} \times (-\pi, \pi)$$

is a coordinate patch on a cylinder, see Figure 4.7. What are the parameter lines?

4.2.8 Show that

$$\mathbf{r}(u, v) = (v \cos u, v \sin u, hu), \quad (u, v) \in \mathbb{R}^2 \tag{4.8}$$

is a coordinate patch. This surface is called a *helicoid*, see Figure 4.7. What are the parameter lines?

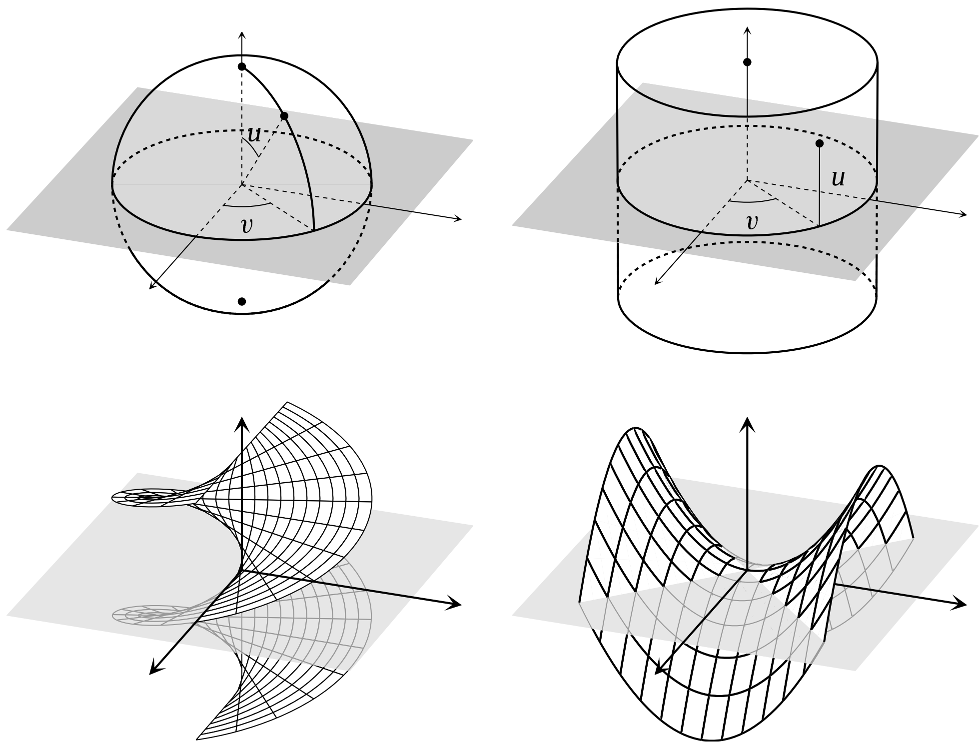


Figure 4.7: Spherical coordinates on the sphere, see Problem 4.2.6. Cylinder coordinates on the cylinder, see Problem 4.2.7. Parameter lines on the helicoid, see Problem 4.2.8 and on the graph of the function $(x, y) \mapsto y^2 - x^2$, see Problem 4.2.9.

4.2.9 Let $U \subseteq \mathbb{R}^2$ be an open subset and let $f : U \to \mathbb{R}$ be a C^k function. Show that

$$\mathbf{r} : (u, v) \mapsto \big(u, v, f(u, v)\big) \tag{4.9}$$

is a C^k-coordinate patch. A patch of this form is often called a *Monge patch*.

4.2.10 Show that the cylinder and the helicoid are regular surfaces.

4.2.11 Show that the unit normal $\mathbf{N}$ in (4.7) is invariant under orientation preserving change of coordinates, but changes sign under an orientation reversing change of coordinates.

4.2.12 Prove Theorem 4.9.

4.2.13 Find the unit normal vector of the cylinder, the helicoid, and the graph of a function $f : U \to \mathbb{R}$.

4.2.14 Consider a plane curve given by the parametrization $T \to \mathbb{R}^2 : u \mapsto \big(r(u), z(u)\big)$ with $r(u) > 0$ for all $u \in I$. If this curve is rotated around the z-axis, we obtain a *surface of revolution*.We can parameterize it as follows:

$$\mathbf{r}(u, v) = \big(r(u)\cos v, r(u)\sin v, z(u)\big). \tag{4.10}$$

Now assume that the original curve $\big(r(u), z(u)\big)$ is regular and one-to-one.

(a) Show that (4.10) is a coordinate patch if $v \in (v_1, v_2)$ with $v_2 - v_1 < 2\pi$.

(b) Show that $M = \{\mathbf{r}(u, v) \mid u \in I, v \in \mathbb{R}\}$ is a regular surface.

(c) Find the unit normal vector of the surface.

4.2.15 Let $\big(r(u), z(u)\big) = \big(2 + \cos u, \sin u\big)$. Show that the corresponding surface of revolution is a regular surface, (called a *torus*), and determine the unit normal vector.

4.2.16 A *ruled surface* is a surface generated by a one parameter family of lines. Let $\mathbf{x} : I \to \mathbb{R}^3$ be a regular curve of class C^k and let $\mathbf{q} : I \to \mathbb{R}^3$ be a non vanishing vector function of class C^k. We get a ruled surface by the following parametrization:

$$\mathbf{r}(u, v) = \mathbf{x}(u) + v\mathbf{q}(u) \tag{4.11}$$

The parameter lines $u =$ constant are called the *rulings* of the surface, and a parametrization of this form is called a parametrization in *ruled form*. It is in general a non trivial task to determine whether the map (4.11) is one-to-one, and it can only be done by a case by case study. The regularity condition on the other hand is easier to handle. Find the cross product $\mathbf{r}_u \times \mathbf{r}_v$.

4.2.17 Show that the hyperbolic paraboloid $z = y^2 - x^2$ is a doubly ruled surface; that is, it can be generated by two different families of lines. Find parametrization of the surface in ruled form.

4.2.18 Let $\mathbf{x}(u) = \big(\cos u, \sin u, 0\big)$, let $\mathbf{q}(u) = \big(\sin \tfrac{1}{2}u \, \cos u, \sin \tfrac{1}{2}u \, \sin u, \cos \tfrac{1}{2}u\big)$, and let $-\pi < u < \pi$. Consider the ruled surface (4.11) for $-\tfrac{1}{4} < v < \tfrac{1}{4}$. Compute the unit normal $\mathbf{N}(u, v)$ and show that

$$\lim_{u \to -\pi} \mathbf{x}(u, v) = \lim_{u \to -\pi} \mathbf{x}(u, -v) \quad \text{and} \quad \lim_{u \to -\pi} \mathbf{N}(u, 0) = - \lim_{u \to -\pi} \mathbf{N}(u, 0),$$

see Figure 4.8. This is called a *Möbius band*.

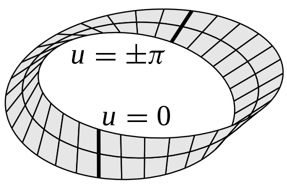


Figure 4.8: A Möbius band.

Exercises

4.2.1 Write a program that calculates the normal vector of a tensor product Bézier surface.

4.2.2 Write a program that calculates the normal vector of a tensor product B-spline surface.

4.2.3 Write a program that calculates the normal vector of a triangular Bézier surface.

4.3 First fundamental form

In this section we will study how to measure length, angles, and area on a surface. As we shall see this is all determined by the *inner product* in the tangent spaces $T_P M$, $P \in M$. If we are given a coordinate patch $\mathbf{r} : U \to M$ then we have a "natural basis" $\mathbf{r}_u, \mathbf{r}_v$ for $T_P M$, but it is in general not orthonormal. Indeed, if $\mathbf{v} = a\mathbf{r}_u + b\mathbf{r}_v$ and $\mathbf{w} = c\mathbf{r}_u + d\mathbf{r}_v$ are two tangent vectors then

$$\begin{aligned} \mathbf{v} \cdot \mathbf{w} &= \big(a\mathbf{r}_u + b\mathbf{r}_v\big) \cdot \big(c\mathbf{r}_u + d\mathbf{r}_v\big) \\ &= ac\mathbf{r}_u \cdot \mathbf{r}_u + (ad + bc)\mathbf{r}_u \cdot \mathbf{r}_v + bd\mathbf{r}_v \cdot \mathbf{r}_v \\ &= acE + (ad + bc)F + bdG, \end{aligned}$$

where we have introduced the three functions

$$\begin{aligned} E(u, v) &= \mathbf{r}_u(u, v) \cdot \mathbf{r}_u(u, v) \\ F(u, v) &= \mathbf{r}_u(u, v) \cdot \mathbf{r}_v(u, v) \\ G(u, v) &= \mathbf{r}_v(u, v) \cdot \mathbf{r}_v(u, v) \end{aligned} \tag{4.12}$$

In the literature you can also find the notation $g_{11} = \mathbf{r}_u \cdot \mathbf{r}_u = E$, $g_{12} = \mathbf{r}_u \cdot \mathbf{r}_v = F$, $g_{21} = \mathbf{r}_v \cdot \mathbf{r}_u = F$, and $g_{22} = \mathbf{r}_v \cdot \mathbf{r}_v = G$. In matrix notation the formula for the inner product reads

$$\mathbf{v} \cdot \mathbf{w} = \begin{bmatrix} a & b \end{bmatrix} \begin{bmatrix} E & F \\ F & G \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}. \tag{4.13}$$

Let $P = \mathbf{r}(u_0, v_0)$ be a point on a regular surface. As a function of the coordinates (a, b) of a tangent vector $\mathbf{v} = a\mathbf{r}_u(u_0, v_0) + b\mathbf{r}_v(u_0, v_0) \in T_P M$, the length squared is given by

$$\mathrm{I}(\mathbf{v}) = |\mathbf{v}|^2 = Ea^2 + 2Fab + Gb^2. \tag{4.14}$$

The function I is a quadratic form on $T_P M$ called the *first fundamental form* and E, F, G are the coefficients of this form with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$.

It is important to realize that the first fundamental form I is invariant under change of coordinates, but the coefficients E, F, G do depend on the parametrization.

If $\mathbf{x}(t) = \mathbf{r}\big(u(t), v(t)\big)$ is a parametrization of a smooth curve on M, then the arc-length s satisfies

$$s'(t)^2 = u'(t)^2 E\big(u(t), v(t)\big) + 2u'(t)v'(t)F\big(u(t), v(t)\big) + v'(t)^2 G\big(u(t), v(t)\big).$$

In short notation we have

$$\frac{\mathrm{d}s}{\mathrm{d}t} = E\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)^2 + 2F\frac{\mathrm{d}u}{\mathrm{d}t}\frac{\mathrm{d}v}{\mathrm{d}t} + G\left(\frac{\mathrm{d}v}{\mathrm{d}t}\right)^2.$$

In the literature you will often find the the following form:

$$\mathrm{d}s^2 = E\,\mathrm{d}u^2 + 2F\,\mathrm{d}u\,\mathrm{d}v + G\,\mathrm{d}v^2.$$

If we interpret ds as the length of a tangent vector and $(\mathrm{d}u, \mathrm{d}v)$ as coordinates for the tangent vector then this is simply the formula for the first fundamental form.

We see immediately that $E = 1$ if an only if u is arc length on the first set of parameter curves, $G = 1$ if and only if v is arc length on the second set of parameter curves, and $F = 0$ if and only if the two set of parameter curves intersects each other orthogonal.

The length of a segment of a curve $t \mapsto \mathbf{r}\big(u(t), v(t)\big)$ is given by

$$s(t_1) = \int_{t_0}^{t_1} s'\,\mathrm{d}t = \int_{t_0}^{t_1} \sqrt{E(u')^2 + 2Fu'v' + G(v')^2}\,\mathrm{d}t.$$

The *angle* α between two tangent vectors $\mathbf{v} = a\mathbf{r}_u + b\mathbf{r}_v$ and $\mathbf{w} = c\mathbf{r}_u + d\mathbf{r}_v$ is given by

$$\cos\alpha = \frac{\mathbf{v}\cdot\mathbf{w}}{|\mathbf{v}||\mathbf{w}|} = \frac{acE + (ad+bc)F + bdG}{\sqrt{\big(Ea^2 + 2Fab + Gb^2\big)\big(Ec^2 + 2Fcd + Gd^2\big)}}.$$

Let (u, v) and (s, t) be two set of local coordinates related by an allowable change of coordinates and let E_1, F_1, G_1 and E_2, F_2, G_2 be the corresponding coefficients of the first fundamental form. In Problem 4.3.7 it is shown that

$$\begin{bmatrix} E_1 & F_1 \\ F_1 & G_1 \end{bmatrix} = \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial t}{\partial u} \\ \frac{\partial s}{\partial v} & \frac{\partial t}{\partial v} \end{bmatrix} \begin{bmatrix} E_2 & F_2 \\ F_2 & G_2 \end{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial s}{\partial u} & \frac{\partial t}{\partial v}, \end{bmatrix} \tag{4.15}$$

It is simply the transformation rule for a quadratic form on a vector space.

Example 4.5 Consider the coordinate patch $\mathbf{r}^+$ on the sphere given by (4.1). We find that

$$E = \mathbf{r}_u^+ \cdot \mathbf{r}_u^+ = \frac{4}{\big(1+u^2+v^2\big)^2}, \quad F = \mathbf{r}_u^+ \cdot \mathbf{r}_v^+ = 0, \quad G = \mathbf{r}_v^+ \cdot \mathbf{r}_v^+ = \frac{4}{\big(1+u^2+v^2\big)^2}.$$

Observe that $\mathbf{r}_u, \mathbf{r}_v$ is an orthogonal basis for the tangent space, it is an orthonormal basis scaled by a factor of $2/\big(1+u^2+v^2\big)$. The parameter lines are in particular orthogonal.

4.3.1 Area

If $\mathbf{v}$ and $\mathbf{w}$ are two vectors in $\mathbb{R}^3$ and α is the angle between them then we have

$$|\mathbf{v}\times\mathbf{w}|^2 = \big(|\mathbf{v}||\mathbf{w}|\sin\alpha\big)^2 = |\mathbf{v}|^2|\mathbf{w}|^2\big(1-\cos^2\alpha\big) = |\mathbf{v}|^2|\mathbf{w}|^2 - (\mathbf{v}\cdot\mathbf{w})^2.$$

Now let $\mathbf{r} : U \to M$ be a coordinate patch on a regular surface M. The vectors $\mathbf{r}_u$ and $\mathbf{r}_v$ span a parallelogram with area

$$|\mathbf{r}_u\times\mathbf{r}_v| = \sqrt{|\mathbf{r}_u|^2|\mathbf{r}_v|^2-(\mathbf{r}_u\cdot\mathbf{r}_v)^2} = \sqrt{EG-F^2}.$$

So it is plausible (and in fact true) that if we have region D on M with $D \subseteq \mathbf{r}(U)$, then the *area* of D is given by a double integral:

$$\text{area}(D) = \iint_{\mathbf{r}^{-1}(D)} |\mathbf{r}_u\times\mathbf{r}_v|\,\mathrm{d}u\,\mathrm{d}v = \iint_{\mathbf{r}^{-1}(D)} \sqrt{EG-F^2}\,\mathrm{d}u\,\mathrm{d}v. \qquad (4.16)$$

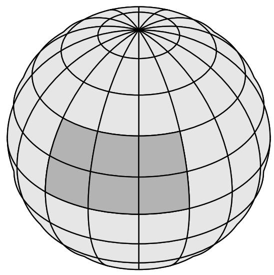


Figure 4.9: A spherical region on S^2.

Example 4.6 Consider the region D on S^2 with spherical coordinates

$$(u, v) \in [u_0, u_1] \times [v_0, v_1],$$

see Figure 4.9. If we use the parametrization

$$\mathbf{r}(u, v) = \big(\sin u\ \cos v, \sin u\ \sin v, \cos u\big),$$

then we have $E = 1$, $F = 0$, and $G = \sin^2 u$, cf. Problem 4.3.2. So $EG - F^2 = \sin^4 u$, and the area of the region becomes

$$\text{area}(D) = \iint_{\mathbf{r}^{-1}(D)} \sqrt{EG-F^2}\,\mathrm{d}u\,\mathrm{d}v = \int_{v_0}^{v_1}\int_{u_0}^{u_1} \sin u\,\mathrm{d}u\mathrm{d}v = (v_1-v_0)(\cos u_0-\cos u_1).$$

If we in particular have $u_0 = 0$, $u_1 = \pi$, $v_0 = 0$, and $v_1 = 2\pi$, then D is the whole sphere and we see that

$$\text{area}\big(S^2\big) = (2\pi - 0)(\cos 0 - \cos\pi) = 4\pi.$$

Problems

4.3.1 Calculate E, F, and G for the coordinate patch $\mathbf{r}^-$ on the unit sphere.

4.3.2 Calculate E, F, and G for the sphere parametrized by spherical coordinates cf. Problem 4.2.6.

4.3.3 Calculate E, F, and G for the cylinder and the helicoid, cf. Problem 4.2.7 and 4.2.8.

4.3.4 Calculate E, F, and G for the graph of a function $f : U \to \mathbb{R}$, cf. Problem 4.2.9.

4.3.5 Calculate E, F, and G for a surface of revolution, cf. Problem 4.2.14.

4.3.6 Calculate E, F, and G for a ruled surface, cf. Problem 4.2.16.

4.3.7 Prove (4.15), hint: use (4.5) or (4.6) and (4.13).

4.3.8 Find the area of the region on the helicoid (4.8) given by $(u, v) \in [0, 2\pi] \times [-1, 1]$.

Exercises

4.3.1 Write a program that calculates E, F, and G for a tensor product Bézier surface.

4.3.2 Write a program that calculates E, F, and G for a tensor product B-spline surface.

4.3.3 Write a program that calculates E, F, and G for a triangular Bézier surface.

4.4 Second fundamental form

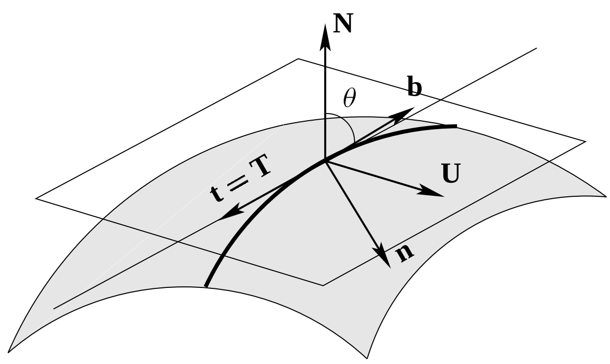


Figure 4.10: The Frenet-Serret frame **t**, **n**, **b** and the Darboux frame **T**, **U**, **N**.

Let M be a regular surface with unit normal $\mathbf{N}$. Let $\mathbf{t}$, $\mathbf{n}$, $\mathbf{b}$ and κ be the Frenet-Serret frame and the curvature of a curve on M respectively. As $\mathbf{t}$ is a tangent vector to the surface it is orthogonal to $\mathbf{N}$ so if we put $\mathbf{T} = \mathbf{t}$ and $\mathbf{U} = \mathbf{N} \times \mathbf{T}$, then

T, **U**, **N** is another positively oriented orthonormal frame along the curve, called the *Darboux frame*, see Figure 4.10.

The curvature vector is orthogonal to **t** so we can write

$$\boldsymbol{\kappa} = \kappa\mathbf{n} = \kappa_g\mathbf{U} + \kappa_n\mathbf{N}, \tag{4.17}$$

where κ_n is called the *normal curvature* and κ_g is called the *geodesic curvature*. We have of course that $\kappa_g = \kappa\mathbf{n}\cdot\mathbf{U}$ and $\kappa_n = \kappa\mathbf{n}\cdot\mathbf{N}$ and we will now find a formula for the latter.

Let the arc length parametrization of the curve be given on the form $\mathbf{x}(s) = \mathbf{r}\big(u(s), v(s)\big)$, where $\mathbf{r} : U \to M$ is a coordinate patch on M. By the chain rule we have

$$\begin{aligned}
\mathbf{t} &= \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}s} = \frac{\mathrm{d}u}{\mathrm{d}s}\mathbf{r}_u + \frac{\mathrm{d}v}{\mathrm{d}s}\mathbf{r}_v \\
\kappa\mathbf{n} &= \frac{\mathrm{d}\mathbf{t}}{\mathrm{d}s} = \frac{\mathrm{d}^2\mathbf{x}}{\mathrm{d}s^2} = \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\mathrm{d}u}{\mathrm{d}s}\mathbf{r}_u\right) + \frac{\mathrm{d}}{\mathrm{d}s}\left(\frac{\mathrm{d}v}{\mathrm{d}s}\mathbf{r}_v\right) \\
&= \left(\frac{\mathrm{d}^2u}{\mathrm{d}s^2}\mathbf{r}_u + \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2\mathbf{r}_{uu} + \frac{\mathrm{d}u}{\mathrm{d}s}\frac{\mathrm{d}v}{\mathrm{d}s}\mathbf{r}_{vu}\right) \\
&\quad + \left(\frac{\mathrm{d}^2v}{\mathrm{d}s^2}\mathbf{r}_u + \frac{\mathrm{d}v}{\mathrm{d}s}\frac{\mathrm{d}u}{\mathrm{d}s}\mathbf{r}_{uv} + \left(\frac{\mathrm{d}v}{\mathrm{d}s}\right)^2\mathbf{r}_{vv}\right) \\
&= \frac{\mathrm{d}^2u}{\mathrm{d}s^2}\mathbf{r}_u + \frac{\mathrm{d}^2v}{\mathrm{d}s^2}\mathbf{r}_u + \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2\mathbf{r}_{uu} + 2\frac{\mathrm{d}u}{\mathrm{d}s}\frac{\mathrm{d}v}{\mathrm{d}s}\mathbf{r}_{uv} + \left(\frac{\mathrm{d}v}{\mathrm{d}s}\right)^2\mathbf{r}_{vv}.
\end{aligned}$$

As $\mathbf{r}_u$ and $\mathbf{r}_v$ are orthogonal to **N** we obtain

$$\kappa_n = \kappa\mathbf{n}\cdot\mathbf{N} = \left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2\mathbf{r}_{uu}\cdot\mathbf{N} + 2\frac{\mathrm{d}u}{\mathrm{d}s}\frac{\mathrm{d}v}{\mathrm{d}s}\mathbf{r}_{uv}\cdot\mathbf{N} + \left(\frac{\mathrm{d}v}{\mathrm{d}s}\right)^2\mathbf{r}_{vv}\cdot\mathbf{N}.$$

If we put

$$L = \mathbf{r}_{uu}\cdot\mathbf{N}, \qquad M = \mathbf{r}_{uv}\cdot\mathbf{N}, \qquad \text{and} \qquad N = \mathbf{r}_{vv}\cdot\mathbf{N} \tag{4.18}$$

then we can write

$$\kappa_n = L\left(\frac{\mathrm{d}u}{\mathrm{d}s}\right)^2 + 2M\frac{\mathrm{d}u}{\mathrm{d}s}\frac{\mathrm{d}v}{\mathrm{d}s} + N\left(\frac{\mathrm{d}v}{\mathrm{d}s}\right)^2.$$

The tangent vector **t** has coordinates $\left(\frac{\mathrm{d}u}{\mathrm{d}s}, \frac{\mathrm{d}v}{\mathrm{d}s}\right)$ with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ for T_PM. So we see that the normal curvature κ_n depends on the tangent only. So if

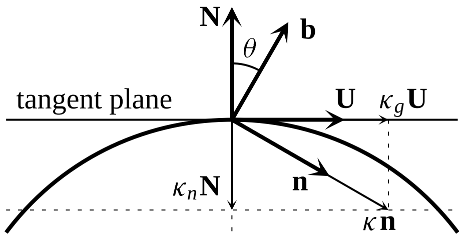


Figure 4.11: The curvature, the normal curvature and the geodesic curvature.

two curve on M have the same tangent at some common point, then they have the same normal curvature at that point.

If we look in the normal plane of the curve then we have the picture in Figure 4.11. If θ is the angle between the tangent plane of the surface and the osculating plane of the curve on the surface, then we can see that the geodesic curvature and the normal curvature is given by

$$\kappa_n = -\kappa \sin\theta \qquad \text{and} \qquad \kappa_g = \kappa \cos\theta. \tag{4.19}$$

The normal curvature is determined by the tangent vector which in turn is determined by the line of intersection between the tangent plane and the osculating plane, so we may write (4.19) as

$$\kappa = \frac{-\kappa_n}{\sin\theta} \qquad \text{and} \qquad \kappa_g = \frac{-\kappa_n}{\tan\theta}$$

A particular example of the situation in Figure 4.11 is obtained by intersecting the surface with a plane that contains the tangent line and have the angle θ with the tangent plane at $P \in M$. If $\theta = \pi/2$ then we intersect the surface with a plane containing the surface normal, we call this a *normal section*. The geodesic curvature at P is then zero and the normal curvature satisfies $\kappa_n = \pm\kappa$.

Definition 4.20. The *second fundamental form* is a quadratic form on the tangent space $T_P M$. If $\mathbf{v} = a\mathbf{r}_u + b\mathbf{r}_v$ is a tangent vector then it is given by

$$\mathrm{I\!I}(\mathbf{v}) = La^2 + 2Mab + Nb^2. \tag{4.20}$$

So L, M, and N are the coefficients of the second fundamental form with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$.

As we saw above it makes sense to talk about the normal curvature in a direction at a given point on a regular surface.

Theorem 4.21. *Let* $\mathbf{v} \in T_P M$ *be a tangent vector of a regular surface. The normal curvature at* $P \in M$ *in the direction* $\mathbf{v}$ *is the ratio between the second and first fundamental form:*

$$\kappa_n = \frac{\mathrm{I\!I}(\mathbf{v})}{\mathrm{I}(\mathbf{v})} \tag{4.21}$$

Proof. Let $\mathbf{x}(t) = \mathbf{r}\big(u(t), v(t)\big)$ be a curve on M. The velocity is $\mathbf{v} = \frac{\mathrm{d}u}{\mathrm{d}t}\mathbf{r}_u + \frac{\mathrm{d}v}{\mathrm{d}t}\mathbf{r}_v$, and if s denotes the arc length the unit tangent vector is

$$\mathbf{t} = \frac{\mathrm{d}u}{\mathrm{d}s}\mathbf{r}_u + \frac{\mathrm{d}v}{\mathrm{d}s}\mathbf{r}_v = \frac{\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)}{\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)}\mathbf{r}_u + \frac{\left(\frac{\mathrm{d}v}{\mathrm{d}t}\right)}{\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)}\mathbf{r}_v.$$

The normal curvature is

$$\begin{aligned}
\kappa_n &= L\left(\frac{\mathrm{d}u}{\mathrm{d}t}\Big/\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 + 2M\left(\frac{\mathrm{d}u}{\mathrm{d}t}\Big/\frac{\mathrm{d}s}{\mathrm{d}t}\right)\left(\frac{\mathrm{d}v}{\mathrm{d}t}\Big/\frac{\mathrm{d}s}{\mathrm{d}t}\right) + N\left(\frac{\mathrm{d}v}{\mathrm{d}t}\Big/\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2 \\
&= \frac{L\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)^2 + 2M\frac{\mathrm{d}u}{\mathrm{d}t}\frac{\mathrm{d}v}{\mathrm{d}t} + N\left(\frac{\mathrm{d}v}{\mathrm{d}t}\right)^2}{\left(\frac{\mathrm{d}s}{\mathrm{d}t}\right)^2} \\
&= \frac{L\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)^2 + 2M\frac{\mathrm{d}u}{\mathrm{d}t}\frac{\mathrm{d}v}{\mathrm{d}t} + N\left(\frac{\mathrm{d}v}{\mathrm{d}t}\right)^2}{E\left(\frac{\mathrm{d}u}{\mathrm{d}t}\right)^2 + 2F\frac{\mathrm{d}u}{\mathrm{d}t}\frac{\mathrm{d}v}{\mathrm{d}t} + G\left(\frac{\mathrm{d}v}{\mathrm{d}t}\right)^2} = \frac{\mathrm{I\!I}(\mathbf{v})}{\mathrm{I}(\mathbf{v})},
\end{aligned}$$

where we have used that $\frac{\mathrm{d}s}{\mathrm{d}t}$ is the length of the velocity vector which squared is given by the first fundamental form. □

In the literature you will often find the the following form:

$$\kappa_n\,\mathrm{d}s^2 = L\,\mathrm{d}u^2 + 2M\,\mathrm{d}u\,\mathrm{d}v + N\,\mathrm{d}v^2.$$

If we as before interpret $\mathrm{d}s$ as the length of a tangent vector and $(\mathrm{d}u, \mathrm{d}v)$ as coordinates for the tangent vector then this equivalent to (4.21).

Example 4.7 Consider the coordinate patch $\mathbf{r}^+$ on the sphere given by (4.1). We find that

$$\begin{aligned}
\mathbf{r}^+_{uu} &= \frac{\big(4u(u^2-3v^2-3), 4v(3u^2-v^2-1), 4(1+v^2-3u^2)\big)}{\big(1+u^2+v^2\big)^3}, \\
\mathbf{r}^+_{uv} &= \frac{\big(4v(3u^2-v^2-1), 4u(3v^2-u^2-1), -16uv\big)}{\big(1+u^2+v^2\big)^3}, \\
\mathbf{r}^+_{vv} &= \frac{\big(4u(3v^2-u^2-1), 4v(v^2-3u^2-3), 4(1+u^2-3v^2)\big)}{\big(1+u^2+v^2\big)^3}.
\end{aligned}$$

By taking the inner product with **N**, see Example 4.4, we obtain

$$L = \mathbf{r}_{uu}^{+} \cdot \mathbf{N} = \frac{4}{\left(1+u^2+v^2\right)^2}, \quad M = \mathbf{r}_{uv}^{+} \cdot \mathbf{N} = 0, \quad N = \mathbf{r}_{vv}^{+} \cdot \mathbf{N} = \frac{4}{\left(1+u^2+v^2\right)^2}.$$

If we compare with Example 4.5 we see that $\mathrm{II} = \mathrm{I}$ and hence $\kappa_n = \mathrm{II}(\mathbf{v})/\mathrm{I}(\mathbf{v}) = 1$ at any point and in any direction $\mathbf{v}$ on the sphere.

As the first and second fundamental form both are quadratic forms on the tangent space they behave similar under a change of local coordinates. If (u, v) and (s, t) are two set of local coordinates related by an allowable change of coordinates and L_1, M_1, N_1 and L_2, M_2, N_2 are the corresponding coefficients of the second fundamental form then

$$\begin{bmatrix} L_1 & M_1 \\ M_1 & N_1 \end{bmatrix} = \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial t}{\partial u} \\ \frac{\partial s}{\partial v} & \frac{\partial t}{\partial v} \end{bmatrix} \begin{bmatrix} L_2 & M_2 \\ M_2 & N_2 \end{bmatrix} \begin{bmatrix} \frac{\partial s}{\partial u} & \frac{\partial s}{\partial v} \\ \frac{\partial s}{\partial u} & \frac{\partial t}{\partial v}. \end{bmatrix} \tag{4.22}$$

From linear algebra we know that there exists an orthonormal basis that diagonalize a given quadratic form. We formulate it as follows:

Theorem 4.22 (Euler). *Let M be a regular surface and let $P \in M$. There exists numbers κ_1 and κ_2, and an orthonormal basis $\mathbf{e}_1, \mathbf{e}_2$ for the tangent space $T_P M$ such that the second fundamental form is given by*

$$\mathrm{II}(a\mathbf{e}_1 + b\mathbf{e}_2) = \kappa_1 a^2 + \kappa_2 b^2.$$

Equivalently, the normal curvature in the direction $\cos\theta\, \mathbf{e}_1 + \sin\theta\, \mathbf{e}_2$ *is*

$$\kappa_n = \kappa_1 \cos^2\theta + \kappa_2 \sin^2\theta \tag{4.23}$$

The numbers κ_1 and κ_2 are called the *principal curvatures* and the directions given by $\mathbf{e}_1$ and $\mathbf{e}_2$ are called the *principal directions*. Before we describe how to determine the principal curvatures and the principal directions we note some immediate consequences of Theorem 4.22.

Corollary 4.23. *The two principal curvatures are the minimum value and maximum value of the normal curvature. Any value between the two principal curvatures is the normal curvature for some direction.*

If $\kappa_1 = \kappa_2$ then the normal curvature is the same in all directions, any orthonormal basis will diagonalize the second fundamental form and all directions are principal.

The following theorem tells how to determine the principal curvatures and the principal directions.

Theorem 4.24. *Let $P \in M$ be a point on a regular surface, let κ_1 and κ_2 be the principal curvatures at P, and let $\mathbf{e}_i = a_i\mathbf{r}_u + b_i\mathbf{r}_v$, $i = 1, 2$, be the corresponding principal directions. They are determined by the following matrix equations:*

$$\begin{bmatrix} L & M \\ M & N \end{bmatrix}\begin{bmatrix} a_i \\ b_i \end{bmatrix} = \kappa_i \begin{bmatrix} E & F \\ F & G \end{bmatrix}\begin{bmatrix} a_i \\ b_i \end{bmatrix}, \qquad i = 1, 2.$$

The principal curvatures κ_1 and κ_2 are in particular the roots of the quadratic equation

$$\det\left(\begin{bmatrix} L & M \\ M & N \end{bmatrix} - \lambda \begin{bmatrix} E & F \\ F & G \end{bmatrix}\right) = 0.$$

When κ_i is known the corresponding principal direction $\mathbf{e}_i = a_i\mathbf{r}_u + b_i\mathbf{r}_v$ is determined by the matrix equation

$$\left(\begin{bmatrix} L & M \\ M & N \end{bmatrix} - \kappa_i \begin{bmatrix} E & F \\ F & G \end{bmatrix}\right)\begin{bmatrix} a_i \\ b_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Proof. The first and the last matrix equation are equivalent, and there is a non trivial solution to the latter if and only if κ_i is a solution to the middle equation. I.e., we only have to prove the first statement.

There exists a unique symmetric linear map $f : T_P M \to T_P M$ such that $\mathrm{II}(\mathbf{v}) = \mathbf{v} \cdot f(\mathbf{v})$ for all $\mathbf{v} \in T_P M$, and the quadratic form II is diagonalized by diagonalizing f. So the principal curvatures and principal directions are the eigenvalues and eigenvectors for f, i.e., they are determined by the equations $f(\mathbf{e}_i) = \kappa_i \mathbf{e}_i$.

Define the matrices

$$\underline{\underline{\mathrm{I}}} = \begin{bmatrix} E & F \\ F & G \end{bmatrix}, \qquad \underline{\underline{\mathrm{II}}} = \begin{bmatrix} L & M \\ M & N \end{bmatrix}, \qquad \text{and} \qquad \underline{\underline{e_i}} = \begin{bmatrix} a_i \\ b_i \end{bmatrix}, \quad i = 1, 2,$$

and recall that if $\underline{\underline{v}}$ and $\underline{\underline{w}}$ are the coordinate matrices for two vectors $\mathbf{v}, \mathbf{w} \in T_P M$ with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ then the inner product is given by $\mathbf{v} \cdot \mathbf{w} = \underline{\underline{v}}^T \underline{\underline{\mathrm{I}}}\, \underline{\underline{w}}$. If $\underline{\underline{F}}$ denotes the matrix for f with respect to the basis $\mathbf{r}_u$ and $\mathbf{r}_v$ then we have $\underline{\underline{v}}^T \underline{\underline{\mathrm{II}}}\, \underline{\underline{v}} = \underline{\underline{v}}^T \underline{\underline{\mathrm{I}}}\, \underline{\underline{F}}\, \underline{\underline{v}}$ for all 2×1 matrices $\underline{\underline{v}}$. Thus $\underline{\underline{\mathrm{II}}} = \underline{\underline{\mathrm{I}}}\, \underline{\underline{F}}$ and $\underline{\underline{F}} = \underline{\underline{\mathrm{I}}}^{-1} \underline{\underline{\mathrm{II}}}$. Hence

$$f(\mathbf{e}_i) = \kappa_i \mathbf{e}_i \iff \underline{\underline{F}}\, \underline{\underline{e_i}} = \kappa_i \underline{\underline{e_i}} \iff \underline{\underline{\mathrm{I}}}^{-1} \underline{\underline{\mathrm{II}}}\, \underline{\underline{e_i}} = \kappa_i \underline{\underline{e_i}} \iff \underline{\underline{\mathrm{II}}}\, \underline{\underline{e_i}} = \kappa_i \underline{\underline{\mathrm{I}}}\, \underline{\underline{e_i}}. \quad \square$$

Corollary 4.25. *At a point P with $F = M = 0$, the principal curvatures and principal directions are given by*

$$\kappa_1 = \frac{L}{E}, \qquad \mathbf{e}_1 = \frac{\mathbf{r}_u}{\sqrt{E}}, \qquad \kappa_2 = \frac{N}{G}, \qquad \mathbf{e}_2 = \frac{\mathbf{r}_v}{\sqrt{G}}.$$

Proof. When $F = M = 0$ the matrix equation for the principal curvatures and principal directions becomes

$$\begin{bmatrix} L - \kappa_i E & 0 \\ 0 & N - \kappa_i N \end{bmatrix} \begin{bmatrix} a_i \\ b_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and we see immediately that the solution is as claimed. □

Definition 4.26. Let $P \in M$ be a point on a regular surface and let κ_1 and κ_2 be the principal curvatures at P. The *Gaussian curvature* $K(P)$ of M at P is the product of the principal curvatures:

$$K(P) = \kappa_1 \kappa_2. \tag{4.24}$$

The *mean curvature* $H(P)$ of M at P is the mean value of the principal curvatures:

$$H(P) = \frac{\kappa_1 + \kappa_2}{2}. \tag{4.25}$$

The sign of the Gaussian curvature is used to classify point on a surface.

Definition 4.27. Let $P \in M$ be a point on a regular surface, and let $K(P)$ and $H(P)$ be the Gaussian and mean curvature at P. The point is called

- *elliptic* if $K(P) > 0$, i.e., the principal curvatures have the same sign.
- *hyperbolic* if $K(P) < 0$, i.e., the principal curvatures have different signs.
- *parabolic* if $K(P) = 0$ and $H(P) \neq 0$, i.e., exactly one principal curvature is zero.
- *planar* if $K(P) = H(P) = 0$, i.e., both principal curvatures are zero.

At an elliptic points the normal curvature has the same sign in all directions so the surface curves in all directions either towards **N** or away from **N**, locally it looks like a bowl and is on one side of the tangent plane. At a hyperbolic points one principal curvature is negative and one is positive. In the direction of the negative principal curvature the surface curves away from **N** and orthogonal to this direction it curves towards **N**, locally it looks like a saddle. Furthermore there are two directions where the normal curvature is zero, such a direction is called an *asymptotic direction*.

An *umbilical point* is a point with $\kappa_1 = \kappa_2$. As $(\kappa_1 - \kappa_2)^2 = (\kappa_1 + \kappa_2)^2 - 4\kappa_1\kappa_2 = 4(H^2 - K)$ a point is an umbilical point if and only if $H^2 - K = 0$.

Example 4.8 Consider the torus

$$\mathbf{r}(u, v) = \big((2 + \cos u)\cos v, (2 + \cos u)\sin v, \sin u\big),$$

we calculate

$$\begin{aligned}
\mathbf{r}_u(u, v) &= \big(-\sin u\ \cos v, -\sin u\ \sin v, \cos u\big)\\
\mathbf{r}_v(u, v) &= \big(-(2 + \cos u)\sin v, (2 + \cos u)\cos v, 0\big),\\
E(u, v) &= \mathbf{r}_u \cdot \mathbf{r}_u = 1,\\
F(u, v) &= \mathbf{r}_u \cdot \mathbf{r}_v = 0,\\
G(u, v) &= \mathbf{r}_v \cdot \mathbf{r}_v = (2 + \cos u)^2,\\
\mathbf{r}_u \times \mathbf{r}_v &= \big(-(2 + \cos u)\cos u\ \cos v, -(2 + \cos u)\cos u\ \sin v, -(2 + \cos u)\sin v\big),\\
|\mathbf{r}_u \times \mathbf{r}_v| &= \sqrt{EG - F^2} = 2 + \cos u,\\
\mathbf{N}(u, v) &= \frac{\mathbf{r}_u \times \mathbf{r}_v}{|\mathbf{r}_u \times \mathbf{r}_v|} = \big(-\cos u\ \cos v, -\cos u\ \sin v, -\sin v\big),\\
\mathbf{r}_{uu}(u, v) &= \big(-\cos u\ \cos v, -\cos u\ \sin v, -\sin v\big),\\
\mathbf{r}_{uv}(u, v) &= \big(\sin u\ \sin v, -\sin u\ \cos v, 0\big),\\
\mathbf{r}_{vv}(u, v) &= \big(-(2 + \cos u)\cos v, -(2 + \cos u)\sin v, 0\big),\\
L(u, v) &= \mathbf{r}_{uu} \cdot \mathbf{N} = 1,\\
M(u, v) &= \mathbf{r}_{uv} \cdot \mathbf{N} = 0,\\
N(u, v) &= \mathbf{r}_{vv} \cdot \mathbf{N} = (2 + \cos u)\cos u.
\end{aligned}$$

We see that we are in the situation of Corollary 4.25, with $F = M = 0$. So the principal directions are the parameter directions and principal curvatures are $\kappa_1 = L/E = 1$ and $\kappa_2 = N/G = \cos u/(2 + \cos u)$. The Gauss curvature is $K(u, v) = \cos u/(2 + \cos u)$ and the mean curvature is $H(u, v) = (1 + \cos u)/(2 + \cos u)$. We have elliptic points for $-\pi/2 < u < \pi/2$, hyperbolic points for $\pi/2 < u < 3\pi/2$, and parabolic points for $u = \pm\pi/2$, see Figure 4.12.

We can determine the Gaussian and mean curvature without calculating the principal curvatures

Proposition 4.28. *Let P be a point on a regular surface M and let E, F, G and L, M, N be the coefficients of the first and second fundamental form at P in some coordinate patch. The Gaussian and mean curvature are given by*

$$K(P) = \frac{LN - M^2}{EG - F^2}$$
$$H(P) = \frac{GL - 2FM + EN}{2(EG - F^2)}$$

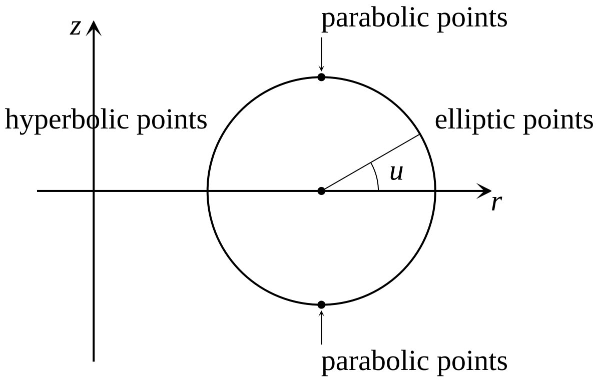


Figure 4.12: The classification of points on the torus.

furthermore, the principal curvatures are

$$\kappa_i = H(P) \pm \sqrt{H(P)^2 - K(P)}$$

Proof. With the notation from the proof of Theorem 4.24 we have

$$K(P) = \kappa_1\kappa_2 = \det(\underline{\underline{F}}) = \det(\underline{\underline{I}}^{-1}\underline{\underline{II}}) = \frac{\det(\underline{\underline{II}})}{\det(\underline{\underline{I}})} = \frac{LN - M^2}{EG - F^2}.$$

To find $H(P)$ we need the matrix $\underline{\underline{F}}$.

$$\begin{aligned}\underline{\underline{F}} = \underline{\underline{I}}^{-1}\underline{\underline{II}} &= \frac{1}{EG - F^2}\begin{bmatrix} G & -F \\ -F & E \end{bmatrix}\begin{bmatrix} L & M \\ M & N \end{bmatrix} \\ &= \frac{1}{EG - F^2}\begin{bmatrix} GL - FM & GM - FN \\ -FL + EM & -FM + EN \end{bmatrix}.\end{aligned}$$

Thus

$$\begin{aligned} H(P) = \frac{\kappa_1 + \kappa_2}{2} &= \frac{1}{2}\operatorname{trace}(\underline{\underline{F}}) \\ &= \frac{1}{2}\operatorname{trace}\left(\frac{1}{EG - F^2}\begin{bmatrix} GL - FM & GM - FN \\ -FL + EM & -FM + EN \end{bmatrix}\right) \\ &= \frac{GL - 2FM + EN}{2(EG - F^2)}. \end{aligned}$$

Finally,

$$(\kappa_1 - \lambda)(\kappa_2 - \lambda) = \lambda^2 - (\kappa_1 + \kappa_2)\lambda + \kappa_1\kappa_2 = \lambda^2 - 2H(P)\lambda + K(P),$$

and the roots of $\lambda^2 - 2H(P)\lambda + K(P)$ are as stated. □

We saw in Theorem 4.19 that we locally can write the surface as a graph of a function defined on the tangent space. We also saw that the first order term of this function vanishes. As we are about to see the second order term is essentially the second fundamental form of the surface.

Proposition 4.29. *Let $P \in M$ be a point on a regular surface, let $\mathbf{r} : U \to M$ be a coordinate patch around P and let L, M, N be the coefficients of the second fundamental form calculated in this coordinate patch. Let $\mathbf{v}_1, \mathbf{v}_2$ be a basis for $T_P M$ and parametrize the surface as*

$$\widetilde{\mathbf{r}}(s,t) = P + s\mathbf{v}_1 + t\mathbf{v}_2 + f(s,t)\mathbf{N}(P).$$

If $\mathbf{v}_1 = \mathbf{r}_u$ and $\mathbf{v}_2 = \mathbf{r}_v$ then

$$f(s,t) = L(P)s^2 + 2M(P)st + N(P)t^2 + o\big(s^2+t^2\big).$$

If $\mathbf{v}_1, \mathbf{v}_2$ is an orthonormal basis in the principal directions then

$$f(s,t) = \kappa_1(P)s^2 + \kappa_2(P)t^2 + o\big(s^2+t^2\big),$$

where κ_1 and κ_2 are the principal curvatures.

Proof. For $(s,t) = (0,0)$, i.e., at the point P we have

$$\widetilde{\mathbf{r}}_s(0,0) = \mathbf{v}_1 + f_s(0,0)\mathbf{N} = \mathbf{v}_1, \qquad \widetilde{\mathbf{r}}_t(0,0) = \mathbf{v}_2 + f_t(0,0)\mathbf{N} = \mathbf{v}_2.$$

and

$$\widetilde{\mathbf{r}}_{ss}(0,0) = f_{ss}(0,0)\mathbf{N}, \quad \widetilde{\mathbf{r}}_{st}(0,0) = f_{st}(0,0)\mathbf{N}, \quad \widetilde{\mathbf{r}}_{tt}(0,0) = f_{tt}(0,0)\mathbf{N}.$$

We see that if use the coordinate patch $\widetilde{\mathbf{r}}$ then the basis for $T_P M$ is $\mathbf{v}_1, \mathbf{v}_2$ and the coefficients of the second fundamental form on $T_P M$ with respect to this basis is

$$\widetilde{L} = \widetilde{\mathbf{r}}_{ss}\cdot\mathbf{N} = f_{ss}(0,0), \qquad \widetilde{M} = \widetilde{\mathbf{r}}_{st}\cdot\mathbf{N} = f_{st}(0,0), \qquad \widetilde{N} = \widetilde{\mathbf{r}}_{tt}\cdot\mathbf{N} = f_{tt}(0,0).$$

As $f(0,0) = f_s(0,0) = f_t(0,0) = 0$ the Taylor expansion of f is

$$\begin{aligned} f(s,t) &= f_{ss}(0,0)t^2 + 2f_{st}(0,0)st + f_{tt}(0,0)t^2 + o\big(s^2+t^2\big) \\ &= \widetilde{L}t^2 + 2\widetilde{M}st + \widetilde{N}t^2 + o\big(s^2+t^2\big). \end{aligned}$$

Now we only have to note that the coefficients of the second fundamental form on $T_P M$ with respect to the basis $\mathbf{r}_u, \mathbf{r}_v$ are $L(P), M(P), N(P)$ and that the coefficients of the second fundamental form on $T_P M$ with respect to an orthonormal basis in the principal directions are $\kappa_1(P), 0, \kappa_2(P)$. $\square$

If we replace the function f above with second order term of f then we obtain a paraboloid, called the *osculating paraboloid*, which has second order contact with M at P. If we intersect the osculating paraboloid with a plane parallel to the tangent plane then we obtain a conic section which

- at an elliptic point is an ellipse, a point (P), or empty.
- at a hyperbolic point is a hyperbola (with asymptotes in the asymptotic direction) or two intersecting lines (through P in the tangent plane in the asymptotic directions).
- at a parabolic point is two lines (parallel with the asymptotic direction), one line (through P in the tangent plane in the asymptotic direction), or empty.
- at a planer point is the tangent plane or empty.

The *Dupin indicatrix* is the union of the intersection with the two planes that have distance one to the tangent plane.

Problems

4.4.1 Find the coefficients of the second fundamental form for the cylinder, cf. Problem 4.2.7. Determine the Gaussian curvature, the mean curvature and the principal curvatures and directions.

4.4.2 Find the coefficients of the second fundamental form for the helicoid, cf. Problem 4.2.8. Determine the Gaussian curvature, the mean curvature and the principal curvatures and directions.

4.4.3 Find the coefficients of the second fundamental form for the graph of a function, cf. Problem 4.2.9. Determine the Gaussian curvature, the mean curvature and the principal curvatures and directions.

4.4.4 Find the coefficients of the second fundamental form for a surface of revolution, cf. Problem 4.2.14. Determine the Gaussian curvature, the mean curvature and the principal curvatures and directions.

4.4.5 Find the coefficients of the second fundamental form for a ruled surface, cf. Problem 4.2.16. Determine the Gaussian curvature, the mean curvature and the principal curvatures and directions.

Exercises

4.4.1 Write a program that calculates L, M, and N for a tensor product Bézier surface.

4.4.2 Write a program that calculates L, M, and N for a tensor product B-spline surface.

4.4.3 Write a program that calculates L, M, and N for a triangular Bézier surface.

4.4.4 Write a program that calculates the Gaussian and mean curvature for a tensor product Bézier surface.

4.4.5 Write a program that calculates the Gaussian and mean curvature for a tensor product B-spline surface.

4.4.6 Write a program that calculates the Gaussian and mean curvature for a triangular Bézier surface.

4.4.7 Write a program that calculates the principal curvatures and the principal directions for a tensor product Bézier surface.

4.4.8 Write a program that calculates the principal curvatures and the principal directions for a tensor product B-spline surface.

4.4.9 Write a program that calculates the principal curvatures and the principal directions for a triangular Bézier surface.

Chapter 5

Rational Curves and Surfaces

5.1 Projective geometry

The fastest way of defining a rational Bézier or B-spline curve in $\mathbb{R}^d$ is probably as a *central projection* of a polynomial curve in $\mathbb{R}^{d+1}$. The study of central projections started in the Renaissance when the artists started to use true perspective, see Figure 5.1. In a central projection any line through the centre (the eye point) is mapped to a single point in the image plane, every plane through the centre is mapped to a single line, and we arrive in a natural way to the concept of a *projective space*.

The d-dimensional *projective space* $\mathbb{RP}^d$ consists of all 1-dimensional subspaces of $\mathbb{R}^{d+1}$, i.e., if $[x_1, \ldots, x_{d+1}]$ denotes the 1-dimensional subspace of $\mathbb{R}^{d+1}$ spanned by $(x_1, \ldots, x_{d+1}) \neq (0, \ldots, 0)$, then

$$\mathbb{RP}^d = \left\{[x_1, \ldots, x_{d+1}] \;\middle|\; (x_1, \ldots, x_{d+1}) \neq \mathbf{0}\right\}. \tag{5.1}$$

In other words, an element (a point) of the projective space is a line through $\mathbf{0}$ in $\mathbb{R}^{d+1}$, and we can specify such a point by its *homogeneous coordinates* $[x_1, \ldots, x_{d+1}]$. Homogeneous coordinates are not unique, if $\lambda \neq 0$ then the homogeneous coordinates $[x_1, \ldots, x_{d+1}]$ and $[\lambda x_1, \ldots, \lambda x_{d+1}]$ represents the same point in projective space.

Consider a plane in $\mathbb{R}^{d+1}$ *not* through $\mathbf{0}$. A line through $\mathbf{0}$ is either parallel to the plane or it intersects it in a unique point. By choosing a suitable basis we can assume the plane is given by $x_{d+1} = 1$, and we have the standard embedding $\mathbb{R}^d \hookrightarrow \mathbb{RP}^d$:

$$(x_1, \ldots, x_d) \mapsto [x_1, \ldots, x_d, 1] \tag{5.2}$$

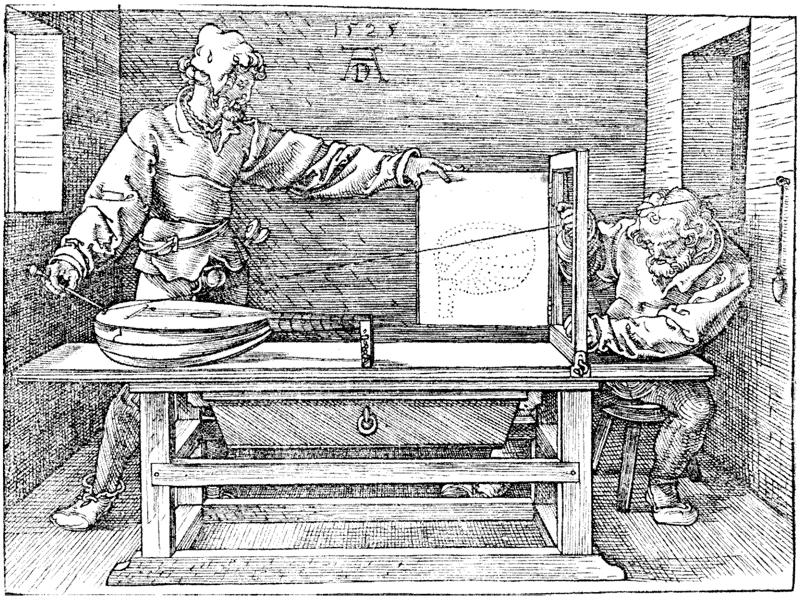


Figure 5.1: Albrecht Dürer, *Artist Drawing a Lute*, woodcut from *Unterweysung der Messung mit dem Zyrkel und Rychtscheyd*, 1525 (The Metropolitan Museum of Art, Harris Brisbane Dick Fund, 1941).

and the opposite map $\{[x_1, \ldots, x_{d+1}] \in \mathbb{RP}^d \mid x_{d+1} \neq 0\} \to \mathbb{R}^d$:

$$[x_1, \ldots, x_{d+1}] \mapsto \left(\frac{x_1}{x_{d+1}}, \ldots, \frac{x_d}{x_{d+1}}\right) \tag{5.3}$$

Points in $\mathbb{RP}^d$ with $x_{d+1} = 0$ can be considered as *points at infinity* in $\mathbb{R}^d$ and they can be identified with a non-oriented direction in $\mathbb{R}^d$ (they are lines in $\mathbb{R}^{d+1}$ parallel to $\mathbb{R}^d$). Considered projectively, two parallel lines in $\mathbb{R}^d$ are two planes in $\mathbb{R}^{d+1}$ that intersect in a line parallel to the two given lines, i.e., two parallel lines intersects at the point at infinity given by the common direction of the two lines.

A d-dimensional subspace in $\mathbb{R}^{d+1}$ is given by an equation

$$a_1x_1 + \cdots + a_{d+1}x_{d+1} = 0,$$

with $\mathbf{a} = (a_1, \ldots, a_{d+1}) \neq \mathbf{0}$. Another set of coefficients $\mathbf{b} = (b_1, \ldots, b_{d+1}) \neq \mathbf{0}$ gives the same subspace if and only if there is a $\lambda \neq 0$ such that $\mathbf{b} = \lambda\mathbf{a}$. So the space of d-dimensional subspaces "is the same" as the space of 1-dimensional subspaces. In other words we can consider an element of $\mathbb{RP}^2$ as either a point or a line, and we can consider an element of $\mathbb{RP}^3$ as either a point or a plane, this is referred to as *duality*. Thus, any statement in projective geometry has two interpretations, see Table 5.1 and 5.2.

Concept/result	Dual concept/result
Point	Line
Line	Point
All points on a line	All lines through a point
All lines through a point	All points on a line
Two distinct points lie on exactly one line	Two distinct lines intersects in exactly one point
Two distinct lines intersects in exactly one point	Two distinct points lie on exactly one line
Two points with homogeneous coordinates [**a**] and [**b**] span a line with coordinates [$\mathbf{a} \times \mathbf{b}$].	Two lines with homogeneous coordinates [**A**] and [**B**] intersect in a point with coordinates [$\mathbf{A} \times \mathbf{B}$].

Table 5.1: The duality of the projective plane.

Concept/result	Dual concept/result
Point	Plane
Line	Line
Plane	Point
All points on a line	All planes through a line
All lines through a point	All lines in a plane
All points on a plane	All planes through a point
All planes through a point	All points on a plane
Two distinct points lie on exactly one line	Two distinct planes intersects in exactly one line
Two distinct planes intersects in exactly one line	Two distinct points lie on exactly one line
A line and a point not on the line lie on exactly one plane.	A line and a plane not containing the line intersects in exactly one point
A line and a plane not containing the line intersects in exactly one point	A line and a point not on the line lie on exactly one plane.

Table 5.2: The duality of the projective three space.

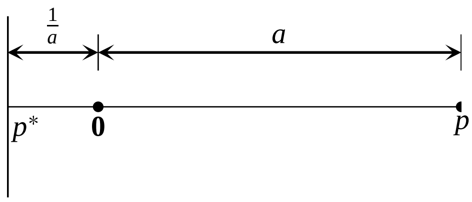


Figure 5.2: A point p and its dual line p^*.

More precisely we have the following description

Theorem 5.1. *In $\mathbb{RP}^2$ we have the following relation between a point p and its dual line p^*, see Figure 5.2*

1. *If a point $p \in \mathbb{R}^2 \subset \mathbb{RP}^2$ has distance $a \neq 0$ to $\mathbf{0} = (0, 0)$, then the dual line p^* has distance $\frac{1}{a}$ to $\mathbf{0}$, the line through $\mathbf{0}$ and p intersects p^* orthogonally and $\mathbf{0}$ is between p and p^*.*
2. *In particular, if $p = (a, 0)$, then p^* is the line $x_1 = -\frac{1}{a}$.*
3. *If $p = \mathbf{0}$, then p^* is the line at infinity.*
4. *If p is a point at infinity, then p^* is the line through $\mathbf{0}$ orthogonal to the direction corresponding to p.*

Proof. If $p = (a, 0)$, then it has homogeneous coordinates $[a, 0, 1]$ and the corresponding line is the intersection between the plane $ax_1 + x_3 = 0$ with the plane $x_3 = 1$, i.e., it's the line with the equation $ax_1 + 1 = 0$. The general case can by a rotation in $\mathbb{R}^2$ always been brought to the previous situation.

If $p = (0, 0)$ then the dual line has the equation $x_3 = 0$, i.e., it's the line at infinity.

If p is a point at infinity corresponding to the direction (a_1, a_2) in $\mathbb{R}^2$ then it has homogeneous coordinates $[a_1, a_2, 0]$, and the dual line has the equation $a_1x_1 + a_2x_2 = 0$, i.e., it's a line through $\mathbf{0}$ orthogonal to (a_1, a_2). $\square$

Theorem 5.2. *In $\mathbb{RP}^3$ we have the following relation between a point p and its dual plane p^*, see Figure 5.3.*

1. *If a point $p \in \mathbb{R}^3 \subset \mathbb{RP}^3$ has distance $a \neq 0$ to $\mathbf{0} = (0, 0, 0)$, then the dual plane p^* has distance $\frac{1}{a}$ to $\mathbf{0}$, the line through $\mathbf{0}$ and p intersects p^* orthogonally and $\mathbf{0}$ is between p and p^*.*
2. *In particular, if $p = (a, 0, 0)$, then p^* is the plane $x_1 = -\frac{1}{a}$.*

3. *If* $p = \mathbf{0}$*, then* p^* *is the plane at infinity.*

4. *If* p *is a point at infinity, then* p^* *is the plane through* $\mathbf{0}$ *orthogonal to the direction corresponding to* p*.*

We have the following relation between a line l *and its dual line* l^**:*

1. *If a line* l *in* $\mathbb{R}^3 \subset \mathbb{RP}^3$ *has distance* $a \neq 0$ *to* $\mathbf{0} = (0, 0, 0)$*, then the dual line* l^* *has distance* $\frac{1}{a}$ *to* $\mathbf{0}$*, the line through* $\mathbf{0}$ *orthogonal to* l *is also orthogonal to* l^* *and* $\mathbf{0}$ *is between* l *and* l^**.*

2. *In particular, if* l *is the line* $t \mapsto (a, 0, 0) + t(0, 1, 0)$*, then* l^* *is the line* $t \mapsto (-\frac{1}{a}, 0, 0) + t(0, 0, 1)$*.*

3. *If* l *is a line through* $\mathbf{0}$ *then* l^* *is the line at infinity consisting of all directions orthogonal to* l*.*

Proof. The relations between a point and its dual plane is proven just as in the case of points and lines in $\mathbb{RP}^2$.

If l can be parametrized as $t \mapsto (a, 0, 0) + t(0, 1, 0)$ then l is spanned by points with homogeneous coordinates $[a, 0, 0, 1]$ and $[0, 1, 0, 0]$ the dual planes have equations $ax_1 + 1 = 0$ and $x_2 = 0$ respectively and their intersection which is l^* can be parametrized as $t \mapsto (-\frac{1}{a}, 0, 0) + t(0, 0, 1)$. The general case can always bee brought to the previous situation by a suitable rotation in $\mathbb{R}^3$.

If l is a line through $\mathbf{0}$, then we assume that it can be parametrized as $t \mapsto (0, 0, t)$ and hence that it is spanned by points with homogeneous coordinates $[0, 0, 0, 1]$ and $[0, 0, 1, 0]$, the dual planes have equations $x_4 = 0$ and $x_3 = 0$ respectively and their intersection which is l^* is the intersection of the x_1, x_2-plane with the plane at infinity. It is the line at infinity in the x_1, x_2-plane, i.e., it consists of all directions in the x_1, x_2-plane. □

A regular linear transformation of $\mathbb{R}^{d+1}$ maps a subspace to a subspace of the same dimension, so it induces in particular a transformation of the projective space $\mathbb{RP}^d$, such a transformation is called a *projective transformation*. In other words a projective transformation is a linear transformation of the homogeneous coordinates. On $\mathbb{R}^2 \subset \mathbb{RP}^2$ we have

$$
\begin{aligned}
(x, y) = [x, y, 1] &\mapsto [a_{11}x + a_{12}y + a_{13}, a_{21}x + a_{22}y + a_{23}, a_{31}x + a_{32}y + a_{33}] \\
&= \left[\frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}, 1\right] \\
&= \left(\frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + a_{33}}, \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + a_{33}}\right)
\end{aligned}
$$

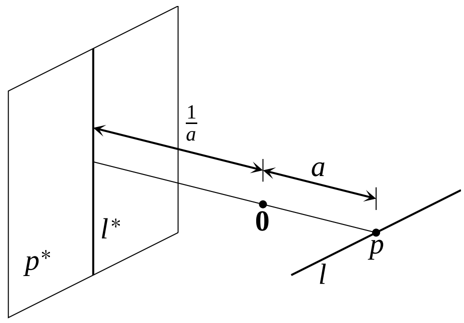


Figure 5.3: A point p and the dual plane p^*, and a line l and the dual line l^*.

Note that if $a_{31} = a_{32} = 0$ and $a_{33} = 1$ then we have an affine transformation of $\mathbb{R}^2$. The case of $\mathbb{RP}^3$ or in general $\mathbb{RP}^d$ is similar. So just as an affine transformation is given by polynomials of degree one in the Cartesian coordinates, a projective transformation is given by rational functions of degree one in the Cartesian coordinates. Similarly, if we have two different bases in $\mathbb{R}^{d+1}$, and hence two different standard embeddings of $\mathbb{R}^d$ in $\mathbb{RP}^d$, then the transformation between the two sets of Cartesian coordinates of a point in $\mathbb{RP}^d$ is given by rational functions of degree one.

Affine transformations don't preserves length, but they do preserve ratios on a line. Projective transformations preserve neither lengths nor ratios, see Figure 5.4.

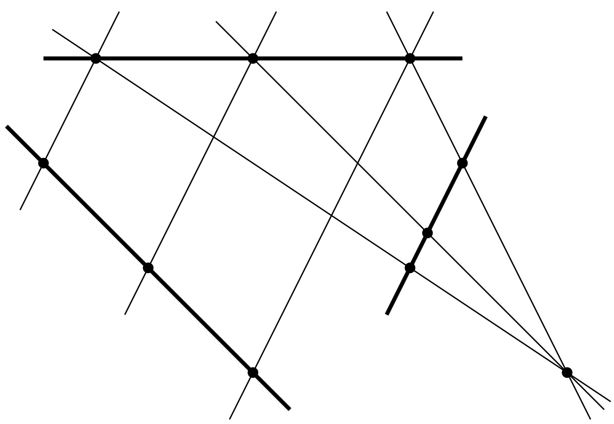


Figure 5.4: Ratios are preserved by a parallel projection, but in general *not* by a central projection.

The ratio of three points a, p, b on a line is given by

$$\text{ratio}(a, p, b) = \frac{d(a, p)}{d(p, b)} \tag{5.4}$$

where $d(a, p)$ denotes the *directed* Euclidean distance a to p, $d(p, b)$ denotes the directed Euclidean distance p to b, and we have used the *same* Euclidean structure and the *same* direction on the line in the numerator and the denominator.

In projective geometry the so called *cross ratio* is a well defined concept. The cross ratio of four points a, p, q, b on a line is defined as

$$\text{cr}(a, b, c, d) = \frac{\text{ratio}(a, p, b)}{\text{ratio}(a, q, b)} = \frac{d(a, p)}{d(p, b)} \Big/ \frac{d(a, q)}{d(q, b)} = \frac{d(a, p)d(q, b)}{d(p, b)d(a, q)} \tag{5.5}$$

If c is the centre of a projection and if the area of the triangle with vertices

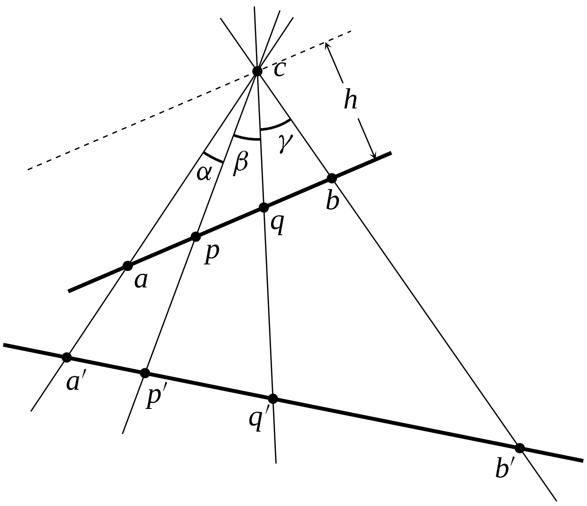


Figure 5.5: The cross ratio of four points depends only on the angles α, β, γ, see (5.6), so it is preserved by a central projection.

A, B, C is denoted $\triangle(A, B, C)$ then with the notation of Figure 5.5 we have eg. $2\triangle(a, p, c) = h \cdot d(a, p)$, and hence

$$\begin{aligned}\text{cr}(a, b, c, d) &= \frac{d(a, p)}{d(p, b)} \Big/ \frac{d(a, q)}{d(q, b)} = \frac{\triangle(a, p, c)}{\triangle(p, b, c)} \Big/ \frac{\triangle(a, q, c)}{\triangle(p, q, c)} \\ &= \frac{l_a l_p \sin\alpha}{l_p l_b \sin(\beta + \gamma)} \Big/ \frac{l_a l_q \sin(\alpha + \beta)}{l_q l_b \sin\gamma} = \frac{\sin\alpha \sin\gamma}{\sin(\beta + \gamma)\sin(\alpha + \beta)}.\end{aligned} \tag{5.6}$$

So just as an *affine embedding* $\mathbb{R} \hookrightarrow \mathbb{R}^d$ is determined by its value on two points, a *projective embedding* $\mathbb{RP}^1 \hookrightarrow \mathbb{RP}^d$ is determined uniquely by its value on three points. Suppose we have three points a, q, b on a line in $\mathbb{RP}^d$, and seek a rational function $f : \mathbb{R} \hookrightarrow \mathbb{RP}^d$ with $f(0) = a$, $f\left(\frac{1}{2}\right) = q$, and $f(1) = b$, then $p = f(t)$ is determined by the invariance of the cross ratio, i.e., we have $\mathrm{cr}(a, p, q, b) = \mathrm{cr}\left(0, t, \frac{1}{2}, 1\right)$, or

$$\frac{d(a, p)d(q, b)}{d(p, b)d(a, q)} = \frac{(t-0)\left(1-\frac{1}{2}\right)}{(1-t)\left(\frac{1}{2}-0\right)} = \frac{t}{1-t}.$$

We can solve for p and obtain

$$f(t) = p = \frac{(1-t)d(q, b)a + td(a, q)b}{(1-t)d(q, b) + td(a, q)}. \tag{5.7}$$

Observe that this is an affine combination of a and b and if $t \in [0, 1]$ and q is between a and b (with respect to Cartesian coordinates) then we have a convex combination of a and b. As a special case we can consider rational functions $f : \mathbb{R} \hookrightarrow \mathbb{RP}^1$ with $f(0) = 0$, $f\left(\frac{1}{2}\right) = \rho$, and $f(1) = 1$ then we have

$$f_\rho(t) = \frac{(1-t)(1-\rho)0 + t(\rho-0)1}{(1-t)(1-\rho) + t(\rho-0)} = \frac{t\rho}{(1-t)(1-\rho) + t\rho}. \tag{5.8}$$

For each $\rho \in]0, 1[$ this gives a rational reparametrization of the interval $[0, 1]$.

Problems

5.1.1 Consider the standard embedding $\mathbb{R}^2 \subset \mathbb{RP}^2$, and find the dual lines of the points $(1, 0)$, $(2, 0)$, $(0, 1/2)$, $(1, 1)$, and $(1, 2)$.

5.1.2 Consider the standard embedding $\mathbb{R}^3 \subset \mathbb{RP}^3$, and find the dual planes of the points $(1, 0, 0)$, $(0, 2, 0)$, $(0, 0, 1/2)$, $(1, 1, 1)$, and $(1, 2, 0)$.

5.1.3 Consider the standard embedding $\mathbb{R}^3 \subset \mathbb{RP}^3$, and find the dual lines of the lines $t \mapsto (1, 0, t)$, $t \mapsto (t, 2, t)$, $t \mapsto (t, (1-t)/2, t)$, $t \mapsto (1+t, 1, 1+t)$, and $t \mapsto (t, 2, 2)$.

5.1.4 Prove (5.7). Hint: introduce a coordinate on the line through a and b.

5.2 Rational Bézier and B-spline curves

Let us first look at the example in Figure 5.6. The line through the points $(0, 1)$ and $(t, 0)$ can be parametrized as $(x, y) = (u\,t, 1-u)$, if we insert this in the equation

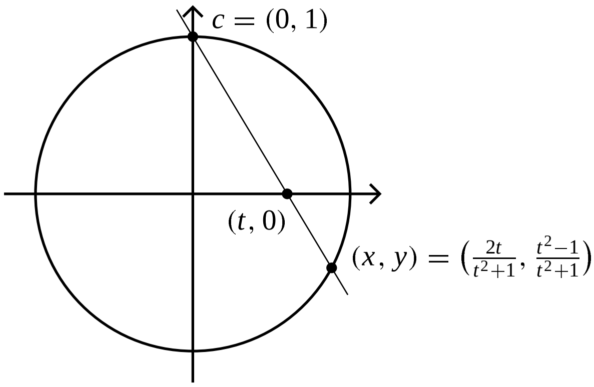


Figure 5.6: Stereographic projections from the unit-circle to the real axis.

of the unit circle $x^2 + y^2 = 1$ then we obtain the equation $u^2(1 + t^2) - 2u = 0$ the two solutions to this equation is $u = 0$ (the centre c of the projection) and $u = \frac{2t}{1+t^2}$ substituted back into the parametrization of the line we obtain the point of intersection

$$(x, y) = \left(\frac{2t}{t^2 + 1}, \frac{t^2 - 1}{t^2 + 1}\right). \tag{5.9}$$

This is a rational parametrization of (part of) the circle. If we consider it as a curve in the projective space then we can specify homogeneous coordinates that are polynomials in the parameter t:

$$[x, y, z] = [2t, t^2 - 1, t^2 + 1]. \tag{5.10}$$

We will now find the three Bézier control points for this quadratic curve in $\mathbb{R}^{2+1}$.

Polynomial	Polar form	evaluated in		
		(0,0)	(0,1)	(1,1)
x=$2t$	$t_1 + t_2$	0	1	2
y=$t^2 - 1$	$t_1 t_2 - 1$	-1	-1	0
z=$t^2 + 1$	$t_1 t_2 + 1$	1	1	2

We see that it has control points that in homogeneous coordinates are $[0, -1, 1]$, $[1, -1, 1]$, and $[2, 0, 2]$. If we project into the plane $z = 1$ then we obtain the Cartesian control points $\mathbf{b}_0 = (0, -1)$, $\mathbf{b}_1 = (1, -1)$, and $\mathbf{b}_3 = (1, 0)$. The z-coordinates of the homogeneous control points are called the *weights* so in the example the weights are $\omega_0 = 1$, $\omega_1 = 1$, and $\omega_2 = 2$.

Definition 5.3. A rational Bézier curve with control points $\mathbf{b}_0, \ldots, \mathbf{b}_n \in \mathbb{R}^d$ and weights $\omega_0, \ldots, \omega_n \in \mathbb{R}$ is given by

$$\mathbf{r}(t) = \frac{\sum_{i=0}^{n} \omega_i \mathbf{b}_i B_i^n(t)}{\sum_{i=0}^{n} \omega_i B_i^n(t)}, \quad t \in [0, 1]. \tag{5.11}$$

I.e., it is the central projection of the Bézier curve in $\mathbb{R}^{d+1}$ with control points $\omega_0(\mathbf{b}_0, 1), \ldots, \omega_n(\mathbf{b}_n, 1)$.

In exactly the same manner we have

Definition 5.4. A rational B-spline curve of degree n with knot sequence $\mathbf{t}$, control points $\mathbf{d}_1, \ldots, \mathbf{d}_{n+N} \in \mathbb{R}^d$ and weights $\omega_1, \ldots, \omega_{n+N} \in \mathbb{R}$ is given by

$$\mathbf{r}(t) = \frac{\sum_{i=1}^{n+N} \omega_i \mathbf{d}_i N_i^n(t|\mathbf{t})}{\sum_{i=1}^{n+N} \omega_i N_i^n(t|\mathbf{t})}, \quad t \in [t_n, t_{n+N}]. \tag{5.12}$$

I.e., it is the central projection of the B-spline curve in $\mathbb{R}^{d+1}$ of degree n with knot sequence $\mathbf{t}$ and control points $\omega_1(\mathbf{d}_1, 1), \ldots, \omega_{n+N}(\mathbf{d}_{n+N}, 1)$.

A rational Bézier (or a B-spline) curve can be evaluated by evaluating the numerator and the denominator separately and dividing through. In [8] it is claimed that it is more stable to find the projection of the intermediate control points. In the case of de Casteljau's algorithm we obtain the following generalized de Casteljau algorithm: For $k = 0, \ldots, n$ do

$$\begin{aligned} \omega_k^0(t) &= \omega_k, \\ \mathbf{b}_k^0(t) &= \mathbf{b}_k, \end{aligned}$$

for $l = 1, \ldots, n$ and $k = 0, \ldots, n - l$ do

$$\begin{aligned} \omega_k^l(t) &= (1-t)\omega_k^{l-1}(t) + t\omega_{k+1}^{l-1}(t), \\ \mathbf{b}_k^l(t) &= (1-t)\frac{\omega_k^{l-1}(t)}{\omega_k^l(t)}\mathbf{b}_k^{l-1}(t) + t\frac{\omega_{k+1}^{l-1}(t)}{\omega_k^l(t)}\mathbf{b}_{k+1}^{l-1}(t). \end{aligned}$$

If all the weights are positive, then we see that we keep taking convex combinations so rational Bézier curves with positive weights have the convex hull property. By considering the intermediate control points in the de Boor algorithm it is seen that rational B-spline curves with positive weights have the strong convex hull property.It can also be shown that the variation diminishing property holds for rational Bézier and B-spline curves with positive weights.

Multiplying all weights with a common factor obviously has no effect at all on the curve, the common factor becomes a factor in both the numerator and the denominator and it cancels out. There are other changes of the weights that leaves the abstract curve unchanged. They stem from *rational linear parameter changes* (*Möbius transformations*). The latter are given by (5.8) and if we substitute this into $(1-t)^{n-i}t^i$ then we obtain

$$\begin{aligned}\big(1-f_\rho(t)\big)^{n-i}\big(f_\rho(t)\big)^i &= \frac{\big((1-t)(1-\rho)\big)^{n-i}(t\rho)^i}{\big((1-t)(1-\rho)+t\rho\big)^n}\\ &= \left(\frac{1-\rho}{(1-t)(1-\rho)+t\rho}\right)^n\left(\frac{\rho}{1-\rho}\right)^i(1-t)^{n-i}t^i.\end{aligned}$$

So if we substitute (5.8) into (5.11) we get

$$\mathbf{r}\big(f_\rho(t)\big) = \frac{\sum_{i=0}^n \omega_i \mathbf{b}_i \big(\frac{\rho}{1-\rho}\big)^i B_i^n(t)}{\sum_{i=0}^n \omega_i \big(\frac{\rho}{1-\rho}\big)^i B_i^n(t)} = \frac{\sum_{i=0}^n \widehat{\omega}_i \mathbf{b}_i B_i^n(t)}{\sum_{i=0}^n \widehat{\omega}_i B_i^n(t)},$$

where $\widehat{\omega}_i = \big(\frac{\rho}{1-\rho}\big)^i \omega_i$. If $\omega_0 \neq 0$ then we can multiply all weights $1/\omega_0$ and obtain $\omega_0 = 1$. We can then reparametrize as above with $\big(\frac{\rho}{1-\rho}\big)^n = 1/\omega_n$, i.e., with $\rho = \frac{1}{1+\sqrt[n]{\omega_n}}$, and obtain $\omega_0 = \omega_n = 1$. In a similar manner we can always obtain $\omega_1 = \omega_{n+N} = 1$ for a rational B-spline curve. When the first and last weight is one we say that the rational curve is in *standard form*.

Both de Casteljau's and de Boor's algorithm repeatedly interpolate between two points, the end points of the line segment, and the weights on the two points determine how the interpolation takes place. The weights are necessary because we use ratios in the interpolation. Alternatively we could specify the image of a certain parameter value, say $\frac{1}{2}$, on the line, and then use cross ratios to calculate the image of any other parameter value, see Figure 5.7. The mid point (line) is spanned by $[\omega_1\mathbf{b}_1 + \omega_2\mathbf{b}_2, \omega_1 + \omega_2]$, i.e., its Cartesian coordinates are given by

$$\mathbf{f} = \frac{\omega_1\mathbf{b}_1 + \omega_2\mathbf{b}_2}{\omega_1 + \omega_2}, \tag{5.13}$$

observe that if the weights are positive then the "mid point" is between the end-points. These auxiliary points are also called *weight points* or *Farin points*, and they can be used as control handles for the weights. Instead of specifying the weights as numbers the designer can slide the weight points back and forth on the legs of the control polygon. They can also be used to define the *projectively invariant de Casteljau algorithm*. Consider Figure 5.8, we are given control points $\mathbf{b}_1$, $\mathbf{b}_2$, and $\mathbf{b}_3$, and weight points $\mathbf{f}_2$ and $\mathbf{f}_3$. There exist weights ω_1, ω_2, and ω_3

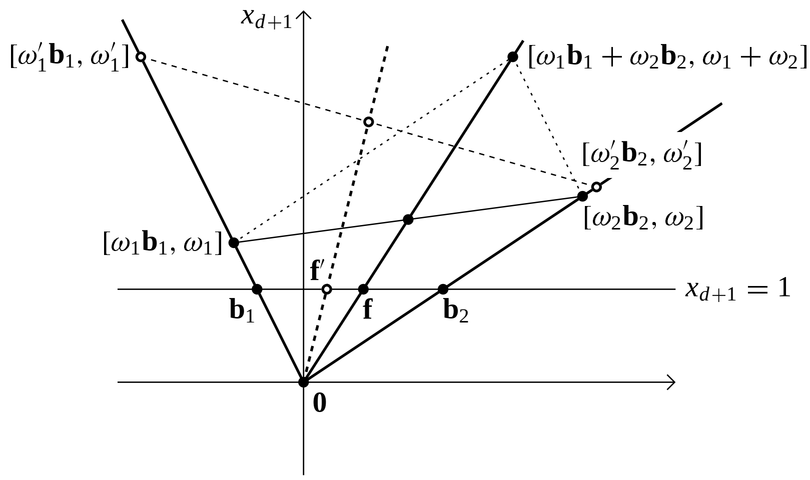


Figure 5.7: We interpolate between two projective points $\mathbf{b}_1$ and $\mathbf{b}_2$, i.e., lines, by choosing a point (not $\mathbf{0}$) on each line and then interpolate between these two points. The weight tells us what point to chose. Alternatively we can specify the weight point (the "mid point") $\mathbf{f}$ of the interpolation. The weight point determines the weights up to a common factor.

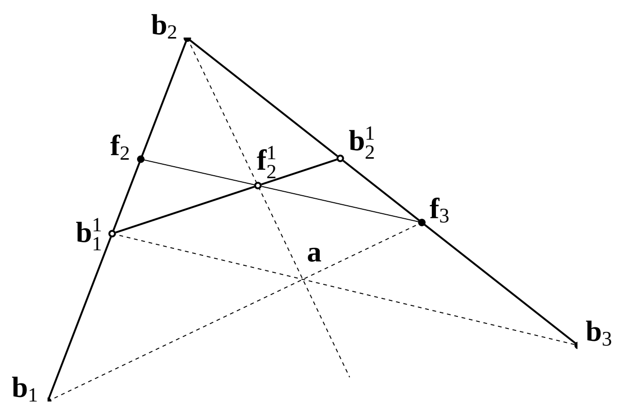


Figure 5.8: Let $\mathbf{a}$ be the intersection between the lines $\mathbf{b}_1\mathbf{f}_3$ and $\mathbf{b}_1^1\mathbf{b}_3$. As $\mathrm{cr}(\mathbf{b}_1, \mathbf{b}_1^1, \mathbf{f}_2, \mathbf{b}_2) = \mathrm{cr}(\mathbf{b}_2, \mathbf{b}_2^1, \mathbf{f}_3, \mathbf{b}_3)$ the points $\mathbf{b}_2$, $\mathbf{f}_2^1$, and $\mathbf{a}$ are collinear.

such that $\mathbf{f}_i = \frac{\omega_{i-1}\mathbf{b}_{i-1}+\omega_i\mathbf{b}_i}{\omega_{i-1}+\omega_i}$. A step in the de Casteljau's algorithm gives us new control points

$$\mathbf{b}_1^1(t) = \frac{(1-t)\omega_1\mathbf{b}_1 + t\omega_2\mathbf{b}_2}{(1-t)\omega_1 + t\omega_2},$$
$$\mathbf{b}_2^1(t) = \frac{(1-t)\omega_2\mathbf{b}_2 + t\omega_3\mathbf{b}_3}{(1-t)\omega_2 + t\omega_3},$$

and new weights

$$\omega_1^1(t) = (1-t)\omega_1 + t\omega_2,$$
$$\omega_2^1(t) = (1-t)\omega_2 + t\omega_3.$$

The new weight point is given by

$$\begin{aligned}
\mathbf{f}_2^1(t) &= \frac{\omega_1^1(t)\mathbf{b}_1^1(t) + \omega_2^1(t)\mathbf{b}_2^1(t)}{\omega_1^1(t) + \omega_2^1(t)} \\
&= \frac{(1-t)\omega_1\mathbf{b}_1 + t\omega_2\mathbf{b}_2 + (1-t)\omega_2\mathbf{b}_2 + t\omega_3\mathbf{b}_3}{(1-t)\omega_1 + t\omega_2 + (1-t)\omega_2 + t\omega_3} \\
&= \frac{(1-t)(\omega_1\mathbf{b}_1 + \omega_2\mathbf{b}_2) + t(\omega_2\mathbf{b}_2 + \omega_3\mathbf{b}_3)}{(1-t)(\omega_1 + \omega_2) + t(\omega_2 + \omega_3)} \\
&= \frac{(1-t)(\omega_1 + \omega_2)\mathbf{f}_2 + t(\omega_2 + \omega_3)\mathbf{f}_3}{(1-t)(\omega_1 + \omega_2) + t(\omega_2 + \omega_3)}.
\end{aligned}$$

And we see that it is the intersection between the lines spanned by the old weight points and the new control points respectively. If $t = \frac{1}{2}$ then the new control points are the weight points and we can't do the intersection. Instead we can find the new weight point as another intersection, cf. Figure 5.8. We now have a purely geometric construction. The new control points are found by the invariance of cross ratios, and the new weight points are found by the intersection of two lines. So the whole construction is projectively invariant.

Problems

5.2.1 Find the standard form of the curve (5.9).

5.2.2 Let $\mathbf{b}(t) = \sum_{i=0}^{n} \omega_i \mathbf{b}_i B_i^n(t)$ and $\omega(t) = \sum_{i=0}^{n} \omega_i B_i^n(t)$, and consider the rational Bézier curve $\mathbf{r}(t) = \mathbf{b}(t)/\omega(t)$. Prove that

$$\mathbf{r}' = \frac{\mathbf{b}' - \omega'\mathbf{r}}{\omega} \quad \text{and} \quad \mathbf{r}'' = \frac{\mathbf{b}'' - 2\omega'\mathbf{r}' - \omega''\mathbf{r}}{\omega}.$$

5.2.3 Consider Figure 5.8. Prove that the points $\mathbf{b}_2$, $\mathbf{f}_2^1$, and $\mathbf{a}$ are collinear.

Exercises

5.2.1 Write a program that draws a rational Bézier curve.

5.2.2 Write a program that finds the first and second derivative of a rational Bézier curve.

5.2.3 Write a program that draws a rational B-spline curve.

5.2.4 Write a program that finds the first and second derivative of a rational B-spline curve.

5.3 Dual curves

A curve $\mathbf{r}(t)$ in $\mathbb{RP}^2$ can be considered as a normal curve giving points in the plane, but by duality equally well as a one-parameter family of lines in the plane. The *envelope* of such a family of lines in the plane is a curve such that every line in the family is a tangent to the curve. The point where the line $\mathbf{r}(t)$ touches the curve is the intersection between the lines $\mathbf{r}(t)$ and $\mathbf{r}'(t)$, i.e., it is given as $\mathbf{c}(t) = \mathbf{r}(t) \times \mathbf{r}'(t)$ in homogeneous coordinates.

As an example we can take the curve (5.10) (a quarter circle), see Figure 5.9

$$\mathbf{r}(t) = [2t, t^2 - 1, t^2 + 1],$$
$$\mathbf{r}'(t) = [2, 2t, 2t],$$
$$\mathbf{c}(t) = \mathbf{r}(t) \times \mathbf{r}'(t) = [-4t, 2 - 2t^2, 2 + 2t^2] = 2[-2t, 1 - t^2, 1 + t^2].$$

Observe that even though $\mathbf{r}$ has degree 2 and $\mathbf{r}'$ has degree 1, the cross product has

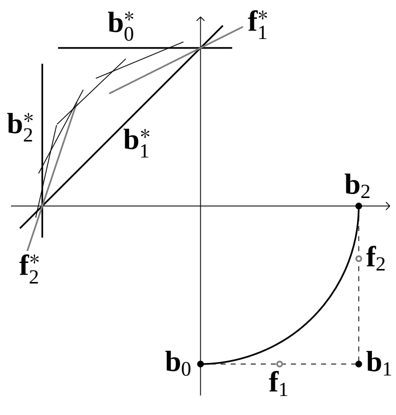


Figure 5.9: Bottom right we have a rational Bézier curve with its control points and weight points. Top left we have the dual curve with its control lines and weight lines.

only degree 2. This is not a coincident, cf. Problem 5.3.1.

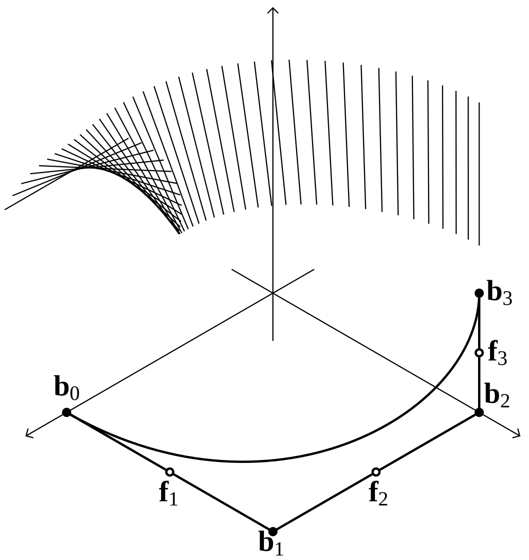


Figure 5.10: A curve in $\mathbb{R}^3 \subset \mathbb{RP}^3$ and the envelope of its dual curve of planes.

For a curve we know that the two tangent lines at the end points are spanned by the two pairs of outermost control points. By duality we have that the two end points of the envelope of a dual curve are the intersection of the two pairs of outermost control lines.

Observe that if a point has Cartesian coordinates $(0, 0)$ then it has homogeneous coordinates $[0, 0, 1]$ and the orthogonal plane is given by $x_3 = 0$ which is parallel to the plane $x_3 = 1$, i.e., the line dual to the point $(0, 0)$ is the line at infinity.

Similarly, a curve $\mathbf{r}(t)$ in $\mathbb{RP}^3$ can by duality be considered as a one parameter family of planes in space. The *envelope* of such a family is a surface such that every plane in the family is a tangent plane of the envelope. The plane $\mathbf{r}(t)$ is tangent to the envelope along the intersection line between $\mathbf{r}(t)$ and $\mathbf{r}'(t)$. This implies in particular that the envelope is a ruled surface. As the tangent plane is constant along the rulings it is a *developable* surface. This gives us a method to construct and design developable surfaces, see Figure 5.10. Above we got the rulings as the intersection between the planes dual to $\mathbf{r}(t)$ and $\mathbf{r}'(t)$. Alternatively we can take the line spanned by $\mathbf{r}(t)$ and $\mathbf{r}'(t)$ (the tangent line), then the dual line is precisely one of the rulings. According to Theorem 5.2 we can find the line dual to the tangent line in the following way. First we find the point on the tangent closest to $\mathbf{0}$, which is

$$\mathbf{c}(t) = \mathbf{r}(t) - \frac{\mathbf{r}(t) \cdot \mathbf{r}'(t)}{|\mathbf{r}'(t)|^2}\mathbf{r}'(t)$$

We can now determine a point on the dual line by

$$\mathbf{c}^*(t) = \frac{-\mathbf{c}(t)}{|\mathbf{c}(t)|^2} = \frac{(\mathbf{r}(t)\cdot\mathbf{r}'(t))\mathbf{r}'(t) - |\mathbf{r}'(t)|^2\mathbf{r}(t)}{|\mathbf{r}(t)|^2|\mathbf{r}'(t)|^2 - (\mathbf{r}(t)\cdot\mathbf{r}'(t))^2}$$

As the ruling is orthogonal to both the tangent line and the line from $\mathbf{c}(t)$ through $\mathbf{0}$ we can find the direction of the rulings as the cross product

$$\mathbf{r}'(t) \times \frac{(\mathbf{r}(t)\cdot\mathbf{r}'(t))\mathbf{r}'(t) - |\mathbf{r}'(t)|^2\mathbf{r}(t)}{|\mathbf{r}(t)|^2|\mathbf{r}'(t)|^2 - (\mathbf{r}(t)\cdot\mathbf{r}'(t))^2} = \frac{|\mathbf{r}'(t)|^2}{|\mathbf{r}(t)|^2|\mathbf{r}'(t)|^2 - (\mathbf{r}(t)\cdot\mathbf{r}'(t))^2}\mathbf{r}(t)\times\mathbf{r}'(t)$$

Observe that the denominator is the length of $\mathbf{r}(u)\times\mathbf{r}'(u)$. So

$$\mathbf{q}(t) = \frac{\mathbf{r}(t)\times\mathbf{r}'(t)}{|\mathbf{r}(t)\times\mathbf{r}'(t)|} = \frac{\mathbf{r}(t)\times\mathbf{r}'(t)}{|\mathbf{r}(t)|^2|\mathbf{r}'(t)|^2 - (\mathbf{r}(t)\cdot\mathbf{r}'(t))^2}$$

is a unit vector in the direction of the rulings. All in all we get the following parametrization of the envelope of the dual curve of planes:

$$\mathbf{r}^*(u,v) = \mathbf{c}^*(u) + v\mathbf{q}(u) = \frac{(\mathbf{r}(u)\cdot\mathbf{r}'(u))\mathbf{r}'(u) - |\mathbf{r}'(u)|^2\mathbf{r}(u) + v\mathbf{r}(u)\times\mathbf{r}'(u)}{|\mathbf{r}(u)|^2|\mathbf{r}'(u)|^2 - (\mathbf{r}(u)\cdot\mathbf{r}'(u))^2}.$$

As an example take the cubic curve with control points

$$\mathbf{b}_0 = (1,0,0), \quad \mathbf{b}_1 = (1,1,0), \quad \mathbf{b}_2 = (0,1,0), \quad \mathbf{b}_3 = \left(0,1,\tfrac{1}{2}\right),$$

and weights $\omega_0 = \cdots = \omega_3 = 1$, or equivalently with weight points

$$\mathbf{f}_1 = \left(1,\tfrac{1}{2},0\right), \quad \mathbf{f}_2 = \left(\tfrac{1}{2},1,0\right), \quad \mathbf{f}_3 = \left(0,1,\tfrac{1}{4}\right).$$

So we have a polynomial curve given by

$$\begin{aligned}\mathbf{r}(t) &= \left(2t^3-3t^2+1,\, t^3-3t^2+3t+\tfrac{1}{2}t^3\right)\\ \mathbf{r}'(t) &= \left(6t^2-6t,\, 3t^2-6t+3,\, \tfrac{3}{2}t^2\right)\end{aligned}$$

We now find

$$c^*(t) = \begin{bmatrix} \frac{-3t^6+24t^5-53t^4+56t^3-36t^2+16t-4}{6t^8-36t^7+98t^6-144t^5+137t^4-96t^3+48t^3-16t+4} \\ \frac{-9t^6+42t^5-64t^4+48t^3-24t^2+8t}{6t^8-36t^7+98t^6-144t^5+137t^4-96t^3+48t^3-16t+4} \\ \frac{-6t^6+12t^5-6t^4}{6t^8-36t^7+98t^6-144t^5+137t^4-96t^3+48t^3-16t+4} \end{bmatrix}$$

and

$$\mathbf{r}(t) \times \mathbf{r}'(t) = \frac{1}{2}\big(-3t^4 + 6t^3, 3t^4 - 6t^2, 6t^4 - 24t^3 + 24t^2 - 12t + 6\big).$$

Finally we have the following parametrization of the envelope of the dual curve of planes:

$$\mathbf{r}^*(u, v) = \frac{1}{6u^8 - 36u^7 + 98u^6 - 144u^5 + 137u^4 - 96u^3 + 48u^3 - 16u + 4}$$
$$\times \begin{bmatrix} -3u^6 + 24u^5 - 53u^4 + 56u^3 - 36u^2 + 16u - 4 + v\frac{3}{2}(-u^4 + 2u^3) \\ -9u^6 + 42u^5 - 64u^4 + 48u^3 - 24u^2 + 8u + v\frac{3}{2}(u^4 - 2u^2) \\ -6u^6 + 12u^5 - 6u^4 + v3(u^4 - 4u^3 + 4u^2 - 2u + 1) \end{bmatrix}$$

Problems

5.3.1 Show that if $\mathbf{r}(t)$ is a polynomial curve in $\mathbb{R}^3$ of degree n, then the cross product $\mathbf{r}(t) \times \mathbf{r}'(t)$ has at most degree $2n - 2$.

Exercises

5.3.1 Write a program that determines the envelope of the dual to a Bézier curve in $\mathbb{R}^2$.

5.3.2 Write a program that determines the envelope of the dual to a Bézier curve in $\mathbb{R}^3$.

5.4 Rational Bézier and B-spline surfaces

The definition of rational surfaces is the same as the definition of rational curves

Definition 5.5. A rational tensor product Bézier surface with control points $\mathbf{b}_{i,j} \in \mathbb{R}^d$ and weights $\omega_{i,j} \in \mathbb{R}$, $i = 0, \ldots, n$, $j = 0, \ldots, m$ is given by

$$\mathbf{r}(u, v) = \frac{\sum_{i=0}^{n} \sum_{j=0}^{m} \omega_{i,j} \mathbf{b}_{i,j} B_i^n(u) B_j^m(v)}{\sum_{i=0}^{n} \sum_{j=0}^{m} \omega_{i,j} B_i^n(u) B_j^m(v)} \quad (u, v) \in [0, 1]^2. \tag{5.14}$$

I.e., it is the central projection of the tensor product Bézier surface in $\mathbb{R}^{d+1}$ with control points $(\omega_{i,j}\mathbf{b}_{i,j}, \omega_{i,j})$.

Definition 5.6. A rational tensor product B-spline surface of degree n, m with knot sequences $\mathbf{u}$ and $\mathbf{v}$, control points $\mathbf{d}_{i,j} \in \mathbb{R}^d$ and weights $\omega_{i,j} \in \mathbb{R}$, $i = 1, \ldots, n+N$, $j = 1, \ldots, m+M$ is given by

$$\mathbf{r}(u, v) = \frac{\sum_{i=1}^{n+N} \sum_{j=1}^{m+M} \omega_{i,j} \mathbf{d}_{i,j} N_i^n(u|\mathbf{u}) N_j^m(v|\mathbf{v})}{\sum_{i=1}^{n+N} \sum_{j=1}^{m+M} \omega_{i,j} N_i^n(u|\mathbf{u}) N_j^m(v|\mathbf{v})}$$

$$(u, v) \in [u_n, u_{n+N}] \times [v_m, v_{m+M}]. \quad (5.15)$$

I.e., it is the central projection of the tensor product B-spline surface in $\mathbb{R}^{d+1}$ of degree n, m with knot sequences $\mathbf{u}$ and $\mathbf{v}$ and control points $(\omega_{i,j}\mathbf{d}_{i,j}, \omega_{i,j})$.

Definition 5.7. A rational triangular Bézier surface of degree n with control points $\mathbf{b}_{ijk} \in \mathbb{R}^d$ and weights $\omega_{ijk} \in \mathbb{R}$, $i + j + k = n$ is given by

$$\mathbf{r}(\mathbf{u}) = \frac{\sum_{|\mathbf{i}|=n} \omega_{\mathbf{i}} \mathbf{b}_{\mathbf{i}} B_{\mathbf{i}}^n(\mathbf{u})}{\sum_{|\mathbf{i}|=n} \omega_{\mathbf{i}} B_{\mathbf{i}}^n(\mathbf{u})} \quad (5.16)$$

I.e., it is the central projection of the triangular Bézier surface in $\mathbb{R}^{d+1}$ of degree n with control points $(\omega_{\mathbf{i}}\mathbf{b}_{\mathbf{i}}, \omega_{\mathbf{i}})$.

The easiest way of treating these surfaces is to evaluate them in $\mathbb{R}^{d+1}$ and then project them to $\mathbb{R}^d$. As in the case of rational curves, it is possible to have a projectively invariant construction, using weight points. But from the practical point of view the importance of weight points is that they can provide a more intuitive access to the weights. In case of the tensor product surfaces we can place a weight point on each edge of the control net, see Figure 5.11. But the weight

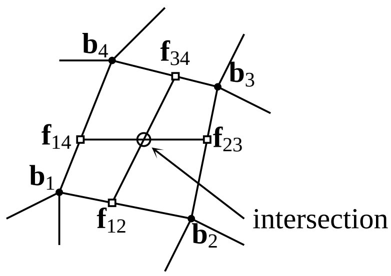


Figure 5.11: The weight points of a tensor product surface satisfies a compatibility condition: The lines between opposite weight points intersect.

points are not independent. With the notation in Figure 5.11 we have

$$\mathbf{f}_{ij} = \frac{\omega_i \mathbf{b}_i + \omega_j \mathbf{b}_j}{\omega_i + \omega_j}$$

and hence

$$\frac{(\omega_1+\omega_2)\mathbf{f}_{12}+(\omega_3+\omega_4)\mathbf{f}_{34}}{\omega_1+\omega_2+\omega_3+\omega_4}=\frac{\sum_{i=1}^4\omega_i\mathbf{b}_i}{\sum_{i=1}^4\omega_i}=\frac{(\omega_1+\omega_4)\mathbf{f}_{14}+(\omega_2+\omega_3)\mathbf{f}_{23}}{\omega_1+\omega_2+\omega_3+\omega_4},$$

so the lines between opposite weight points intersect. The case of the rational triangular Bézier surfaces is more satisfying. To each triangle $\mathbf{b}_{\mathbf{i}+\mathbf{e}_1}\mathbf{b}_{\mathbf{i}+\mathbf{e}_2}\mathbf{b}_{\mathbf{i}+\mathbf{e}_3}$ in the control net we have an independent weight point

$$\mathbf{f}_{\mathbf{i}}=\frac{\omega_{\mathbf{i}+\mathbf{e}_1}\mathbf{b}_{\mathbf{i}+\mathbf{e}_1}+\omega_{\mathbf{i}+\mathbf{e}_2}\mathbf{b}_{\mathbf{i}+\mathbf{e}_2}+\omega_{\mathbf{i}+\mathbf{e}_3}\mathbf{b}_{\mathbf{i}+\mathbf{e}_3}}{\omega_{\mathbf{i}+\mathbf{e}_1}+\omega_{\mathbf{i}+\mathbf{e}_2}+\omega_{\mathbf{i}+\mathbf{e}_3}},$$

see Figure 5.12. These weight points determines the ratio between neighbouring weight points, and hence determines all the weight points up to a common factor.

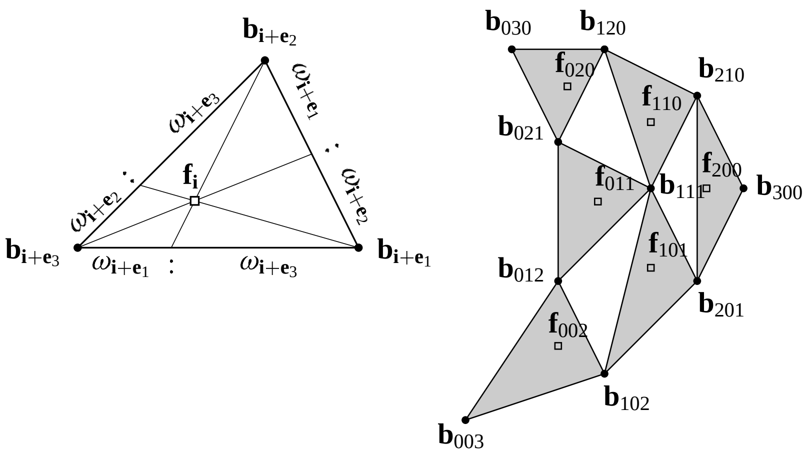


Figure 5.12: Weight points for a triangular Bézier surface.

Exercises

5.4.1 Write a program that draws a rational tensor product Bézier surface.

5.4.2 Write a program that draws a rational tensor product B-spline surface.

5.4.3 Write a program that draws a rational triangular Bézier surface.

Chapter 6

Motion Design

6.1 Introduction

In this chapter we will describe the motion of rigid bodies through space. It can be the motion of a real physical object like (part of) a robot, but it can also be the motion of a virtual object like a character in a computer game, or the camera in a virtual world.

A Euclidean motion is composed of a translation and a rotation, i.e., it's a map $\mathbf{x} \mapsto \underline{\underline{U}}\mathbf{x} + \mathbf{v}$ where $\underline{\underline{U}}$ is a special orthogonal matrix. Using homogeneous coordinates it can be written

$$\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} \underline{\underline{U}} & \mathbf{v} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix},$$

and then the composition of Euclidean motions corresponds to matrix multiplication. When we want to design, or just describe, a smooth motion, i.e., to specify a curve in the space of Euclidean motions, then we have to specify the rotation $\underline{\underline{U}}(t)$ and the translation $\mathbf{v}(t)$ as a function of time t. The latter poses no problem it is just a parametrized curve in $\mathbb{R}^3$, but we need to deal with the space of rotations or equivalently: with the space of special orthogonal 3×3 matrices.

It turns out that the *quaternions* provide a convenient way of dealing with exactly the space of rotations.

6.2 Quaternions and rotations

Just as we get the complex numbers $\mathbb{C}$ as numbers $x + iy$ where $x, y \in \mathbb{R}$ and i is a new number with $i^2 = -1$, we get the quaternions[1] $\mathbb{H}$ as numbers

$$q_0 + iq_1 + jq_2 + kq_3,$$

where $q_0, q_1, q_2, q_3 \in \mathbb{R}$, and i, j, k are three new numbers such that

$$i^2 = j^2 = k^2 = -1,$$
$$ij = -ji = k, \quad jk = -kj = i, \quad ki = -ik = j.$$

For a quaternion $q = q_0 + iq_1 + jq_2 + kq_3$, the *real part* of q is $\Re(q) = q_0$ and the *imaginary part* of q is $\Im(q) = iq_1 + jq_2 + kq_3$. Just as for the complex numbers, *conjugation* is defined by changing the sign of the imaginary part:

$$\overline{q} = \overline{q_0 + iq_1 + jq_2 + kq_3} = q_0 - iq_1 - jq_2 - kq_3, \tag{6.1}$$

and the *modulus* or the *length* of a quaternion is defined as

$$|q| = \sqrt{q\,\overline{q}} = \sqrt{q_0^2 + q_1^2 + q_2^2 + q_3^2}. \tag{6.2}$$

It is not hard to see that we have the following rules for quaternions p, q, and r:

$p + q = q + p,$	the commutative law for addition	(6.3)
$(p + q) + r = p + (q + r),$	the associative law for addition	(6.4)
$(pq)r = p(qr),$	the associative law for multiplication	(6.5)
$p\,(q + r) = p\,q + p\,r,$	the left distributive law	(6.6)
$(p + q)\,r = p\,r + q\,r,$	the right distributive law	(6.7)
$\overline{q\,r} = \overline{r}\,\overline{q},$		(6.8)
$\lvert q\,r\rvert = \lvert r\rvert\,\lvert q\rvert,$		(6.9)
$q^{-1} = \dfrac{\overline{q}}{\lvert q\rvert^2},$		(6.10)
$\Re(q) = \dfrac{q + \overline{q}}{2},$		(6.11)
$\Im(q) = \dfrac{q - \overline{q}}{2},$		(6.12)

The commutative law does not hold for multiplication, in general $qr \neq rq$! But a real number commutes with any other quaternion.

[1] Hamilton 1843

The rules (6.3)–(6.8) can be shown by direct calculation, but it's also a consequence of Problem 6.2.5 and the corresponding rules for matrices. The rules (6.10)–(6.12) is shown by direct calculation, cf. Problem 6.2.4.

If we forget about the multiplication then $\mathbb{H}$ is a real 4-dimensional vector space with basis $1, i, j, k$ and the purely imaginary quaternions $\Im(\mathbb{H})$ form a real 3-dimensional vector space with basis i, j, k. Furthermore (6.2) defines a norm with the corresponding inner product given by $\langle q, r\rangle = \Re(q\,\overline{r})$, cf. Problem 6.2.2. This makes $\mathbb{H}$ into a Euclidean vector space and the abovementioned basis (i, j, k) becomes orthonormal.

Consider for a fixed quaternion $q \in \mathbb{H} \setminus \{0\}$ the so called *adjoint map*

$$\operatorname{Ad}(q) : \mathbb{H} \to \mathbb{H} : r \mapsto q\,r\,q^{-1} = \frac{q\,r\,\overline{q}}{|q|^2}. \tag{6.13}$$

If we consider $\mathbb{H}$ as a real vector space then $\operatorname{Ad}(q)$ is linear, cf. Problem 6.2.7. We have

$$\overline{\operatorname{Ad}(q)\,r} = \frac{\overline{q\,r\,\overline{q}}}{|q|^2} = \frac{\overline{\overline{q}}\,\overline{r}\,\overline{q}}{|q|^2} = \frac{q\,\overline{r}\,\overline{q}}{|q|^2} = \operatorname{Ad}(q)\,\overline{r},$$

so

$$\begin{aligned}\left|\operatorname{Ad}(q)\,r\right|^2 = (\operatorname{Ad}(q)\,r)(\overline{\operatorname{Ad}(q)\,r}) &= \frac{q\,r\,\overline{q}}{|q|^2}\,\frac{q\,\overline{r}\,\overline{q}}{|q|^2} = \frac{q\,r\,\overline{q}\,q\,\overline{r}\,\overline{q}}{|q|^4}\\ &= \frac{q\,r\,|q|^2\,\overline{r}\,\overline{q}}{|q|^4} = \frac{q\,r\,\overline{r}\,\overline{q}}{|q|^2} = \frac{q\,|r|^2\,\overline{q}}{|q|^2} = \frac{|r|^2\,q\,\overline{q}}{|q|^2} = |r|^2\end{aligned}$$

and we see that $\operatorname{Ad}(q)$ acts as an *isometry* on $\mathbb{H}$. As a consequence of Problem 6.2.7 we have

$$\Re(\operatorname{Ad}(q)\,r) = \operatorname{Ad}(q)\,\Re(r) \tag{6.14}$$

$$\Im(\operatorname{Ad}(q)\,r) = \operatorname{Ad}(q)\,\Im(r) \tag{6.15}$$

In particular, $\operatorname{Ad}(q)$ maps $\Im(\mathbb{H})$ into itself and Problem 6.2.7 shows that $\operatorname{Ad}(q)_{|\Im(\mathbb{H})}$ has the matrix

$$\frac{1}{q_0^2+q_2^2+q_1^2+q_3^2} \times \begin{bmatrix} q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2)\\ 2(q_1q_2+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_0q_1)\\ 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{bmatrix} \tag{6.16}$$

with respect to the basis i, j, k. As $\operatorname{Ad}(q)$ is an isometry and the basis i, j, k is orthonormal, the matrix is orthogonal. The determinant of any orthonormal basis is ± 1 and as the Ad(1) is the identity, which has determinant 1, a continuity argument shows that (6.16) has determinant 1. All this can of course also be checked directly.

We now want a geometrical description of the rotation in $\Im(\mathbb{H}) \cong \mathbb{R}^3$ given by $\mathrm{Ad}(q)$. We know that the eigenvalues are 1 and $e^{\pm i\theta}$ where θ is the angle of rotation, and that an eigenvector for the eigenvalue 1 lies on the axis of rotation.

We immediately have

$$\mathrm{Ad}(q)\Im(q) = \Im(\mathrm{ad}(q)q) = \Im(q^{-1}\,q\,q) = \Im(q),$$

so $\Im(q)$ is an eigenvector with eigenvalue 1. Furthermore, the trace of (6.16) is the sum of the eigenvalues which is $1 + 2\cos\theta = 1 + 2\big(\cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}\big)$, on the other hand the trace is also the sum of the diagonal

$$\frac{3q_0^2 - q_1^2 - q_2^2 - q_3^2}{q_0^2 + q_1^2 + q_2^2 + q_3^2} = 1 + 2\frac{q_0^2 - q_1^2 - q_2^2 - q_3^2}{q_0^2 + q_1^2 + q_2^2 + q_3^2}.$$

So we have

$$\frac{q_0^2 - (q_1^2 + q_2^2 + q_3^2)}{q_0^2 + q_1^2 + q_2^2 + q_3^2} = \cos^2\frac{\theta}{2} - \sin^2\frac{\theta}{2}.$$

As

$$\frac{q_0^2 + (q_1^2 + q_2^2 + q_3^2)}{q_0^2 + q_1^2 + q_2^2 + q_3^2} = 1 = \cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2},$$

we can conclude that

$$\frac{q_0^2}{q_0^2 + q_1^2 + q_2^2 + q_3^2} = \cos^2\frac{\theta}{2} \quad \text{and} \quad \frac{q_1^2 + q_2^2 + q_3^2}{q_0^2 + q_1^2 + q_2^2 + q_3^2} = \sin^2\frac{\theta}{2}.$$

Thus

$$\Re(q) = |q|\cos\left(\frac{\theta}{2}\right) \qquad \text{and} \qquad \Im(q) = |q|\sin\left(\frac{\theta}{2}\right)\mathbf{r},$$

or

$$\Re(q) = |q|\cos\left(\frac{\theta}{2} + \pi\right) \qquad \text{and} \qquad \Im(q) = |q|\sin\left(\frac{\theta}{2} + \pi\right)\mathbf{r},$$

where $\mathbf{r}$ is a unit vector in the direction of the axis of rotation. This shows in particular that any rotation can be obtained in this manner, and that $\mathrm{Ad}(q) = \mathrm{Ad}(q')$ if and only if $q = \lambda q'$ for some $\lambda \in \mathbb{R} \setminus \{0\}$. In other words, the group of special orthogonal matrices $SO(3)$ is homeomorphic to the real projective space

$\mathbb{RP}^3$ and (6.16) gives us a rational parametrization (of degree 2) of $SO(3)$. This was found by L. Euler in 1770 and the numbers q_0, q_1, q_2, and q_3 are therefore called *Euler parameters*. Euler made his discovery 70 years before Hamilton introduced the quaternions so it is of course possible to find (6.16) without the quaternions, see eg. [23, Chapter 6].

If we have a curve $t \mapsto q(t)$ in $\mathbb{H}$ then differentiation of the equation $q\,q^{-1} = 1$, yields

$$\left(q^{-1}\right)' = -q^{-1}\,\dot{q}\,q^{-1}. \tag{6.17}$$

If we have $r(t) = \mathrm{Ad}(q(t))r^1$, then

$$\begin{aligned} r'(t) &= \dot{q}\,r^1\,q^{-1} - q\,r^1\,q^{-1}\,\dot{q}\,q^{-1} \\ &= \left(\dot{q}\,q^{-1}\right)\left(q\,r^1\,q^{-1}\right) - \left(q\,r^1\,q^{-1}\right)\left(\dot{q}\,q^{-1}\right) \\ &= \mathrm{ad}\left(\dot{q}\,q^{-1}\right)\,\mathrm{Ad}(q)\,r^1 = \mathrm{ad}\left(\dot{q}\,q^{-1}\right)r(t) = 2\Im\left(\dot{q}\,q^{-1}\right) \times \Im\left(r(t)\right) \end{aligned} \tag{6.18}$$

where we for an $a \in \mathbb{H}$ have introduced the *adjoint map*,

$$\mathrm{ad}(a) : \mathbb{H} \to \mathbb{H} : r \mapsto a\,r - r\,a = 2\Im(a) \times \Im(r). \tag{6.19}$$

The last equality is a consequence of Problem 6.2.8 and shows that the *angular velocity* is $2\Im\left(\dot{q}\,q^{-1}\right)$, this gives in particular the *instantaneous axis of rotation*. We can now use this to analyse the motion of a freely rotating rigid body.

Example 6.1 Suppose we have a robot or some other body rotating in space. We will consider two coordinate systems. The *world system* which is fixed in space and the *body system* which is fixed in the body. A vector can now be given by world coordinates (x^0, y^0, z^0) or by body coordinates (x^1, y^1, z^1), or equivalent by a world quaternion r^0 or a body quaternion r^1. We now assume that the transformation from body coordinates to world coordinates is given by $r^0 = \mathrm{Ad}(q(t))r^1$. Equation (6.18) shows that the angular velocity are given by $\omega^0 = 2\Im\left(\dot{q}q^{-1}\right)$ in world coordinates. In body coordinates we have

$$\omega^1 = \mathrm{Ad}\left(q^{-1}\right)\omega^0 = \mathrm{Ad}\left(q^{-1}\right)2\Im\left(\dot{q}q^{-1}\right) = 2\Im\left(q^{-1}\dot{q}\right).$$

On the other hand, if $\mathbf{I}^1$ is the tensor of inertia and L^0 and L^1 are the angular moments in the world and body system respectively, then $L^1 = \mathbf{I}^1\omega^1$, and hence

$$\omega^1 = \mathbf{I}^{1^{-1}}L^1 = \mathbf{I}^{1^{-1}}\,\mathrm{Ad}\left(q^{-1}\right)L^0.$$

Furthermore, if $|q(t)| = 1$ for all t, then

$$\dot{q} = \frac{1}{2}q\omega^1 = \frac{1}{2}q\mathbf{I}^{1^{-1}}\,\mathrm{Ad}\left(q^{-1}\right)L^0.$$

Conversely assume $q(t)$ is a solution to this differential equation, then

$$\frac{\mathrm{d}}{\mathrm{d}t}|q(t)|^2 = \frac{\mathrm{d}}{\mathrm{d}t}\Re(\overline{q}q) = \Re(\dot{\overline{q}}q + \overline{q}\dot{q}) = 2\Re(\overline{q}\dot{q}) = \Re\left(|q|^2\mathbf{I}^{1^{-1}}\operatorname{Ad}\big(q^{-1}\big)L^0\right) = 0.$$

I.e., $|q(t)|$ is constant. If we choose the axis of the body system in the principle directions, then $\mathbf{I}^1$ is diagonal, with elements I_1, I_2, I_3 say. If the angular momentum L^0 has components α, β, γ then

$$\begin{bmatrix}\dot{q}_0\\ \dot{q}_1\\ \dot{q}_2\\ \dot{q}_3\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-q_1 & -q_2 & -q_3\\ q_0 & -q_3 & q_2\\ q_3 & q_0 & -q_1\\ -q_2 & q_1 & q_0\end{bmatrix}\begin{bmatrix}I_1^{-1} & 0 & 0\\ 0 & I_2^{-1} & 0\\ 0 & 0 & I_3^{-1}\end{bmatrix}$$
$$\times\begin{bmatrix}q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2+q_0q_3) & 2(q_1q_3-q_0q_2)\\ 2(q_1q_2-q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3+q_0q_1)\\ 2(q_1q_3+q_0q_2) & 2(q_2q_3-q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix} \tag{6.20}$$

In the absence of exterior moments the angular momentum L^0 is constant and we can choose the world system with the z-axis in the direction of L^0. I.e., we may assume that $\alpha = \beta = 0$, and $\gamma = L = |L^0|$. In this case we obtain

$$\begin{bmatrix}\dot{q}_0\\ \dot{q}_1\\ \dot{q}_2\\ \dot{q}_3\end{bmatrix} = L\begin{bmatrix}-q_1 & -q_2 & -q_3\\ q_0 & -q_3 & q_2\\ q_3 & q_0 & -q_1\\ -q_2 & q_1 & q_0\end{bmatrix}\begin{bmatrix}\frac{q_1q_3-q_0q_2}{I_1}\\ \frac{q_2q_3+q_0q_1}{I_2}\\ \frac{q_0^2-q_1^2-q_2^2+q_3^2}{2I_3}\end{bmatrix}. \tag{6.21}$$

Alternatively, the components $\omega_1, \omega_2, \omega_3$ of the angular velocity in the body system, are determined by the *Euler equations*

$$\begin{aligned} I_1\dot{\omega}_1 &= (I_2 - I_3)\omega_2\omega_3,\\ I_2\dot{\omega}_2 &= (I_3 - I_1)\omega_3\omega_1,\\ I_3\dot{\omega}_3 &= (I_1 - I_2)\omega_1\omega_2, \end{aligned} \tag{6.22}$$

see [1]. Now $q(t)$ is the solution to the differental equation $\dot{q} = \frac{1}{2}q\omega^1$, i.e.,

$$\begin{bmatrix}\dot{q}_0\\ \dot{q}_1\\ \dot{q}_2\\ \dot{q}_3\end{bmatrix} = \frac{1}{2}\begin{bmatrix}-q_1 & -q_2 & -q_3\\ q_0 & -q_3 & q_2\\ q_3 & q_0 & -q_1\\ -q_2 & q_1 & q_0\end{bmatrix}\begin{bmatrix}\omega_1\\ \omega_2\\ \omega_3\end{bmatrix} \tag{6.23}$$

If $I_1 > I_2 > I_3$, then the angular velocity can be expressed in terms of elliptic functions,

$$\begin{aligned} \omega_1 &= \sqrt{\frac{L^2 - 2EI_3}{I_1(I_1 - I_3)}}\,\operatorname{cn}(\tau - \tau_0, k),\\ \omega_2 &= -\sqrt{\frac{L^2 - 2EI_3}{I_2(I_2 - I_3)}}\,\operatorname{sn}(\tau - \tau_0, k),\\ \omega_3 &= \sqrt{\frac{2EI_1 - E^2}{I_3(I_1 - I_3)}}\,\operatorname{dn}(\tau - \tau_0, k), \end{aligned}$$

where τ_0 is determined by the initial conditions, and

$$\tau = \sqrt{\frac{(I_2 - I_3)(2EI_1 - L^2)}{I_1 I_2 I_3}}\, t, \qquad k = \sqrt{\frac{(I_1 - I_2)(L^2 - 2EI_3)}{(I_2 - I_3)(2EI_1 - L^2)}},$$

and $2E = \sqrt{I_1\omega_1^2 + I_2\omega_2^2 + I_3\omega_3^2}$ is twice the (constant) kinetic energy, see [21]. Alternatively, we can take the equations $\mathrm{I}^1\omega^1 = \mathrm{Ad}(q^{-1})L^0$ and $|q| = 1$ where we as before can assume that $L^0 = (0, 0, L)$ and obtain

$$q_1q_3 - q_0q_2 = \frac{I_1\omega_1}{2L}, \qquad q_0^2 - q_1^2 - q_2^2 + q_3^2 = \frac{I_3\omega_3}{L},$$
$$q_2q_3 + q_0q_1 = \frac{I_2\omega_2}{2L}, \qquad q_0^2 + q_1^2 + q_2^2 + q_3^2 = 1.$$

The equations to the right yields

$$q_0^2 + q_3^2 = \frac{L + I_3\omega_3}{2L}, \qquad q_1^2 + q_2^2 = \frac{L - I_3\omega_3}{2L},$$

and the equations to the left yields

$$q_3(q_1^2 + q_2^2) = \frac{q_1I_1\omega_1 + q_2I_2\omega_2}{2L}, \qquad q_0(q_1^2 + q_2^2) = \frac{q_1I_2\omega_2 - q_2I_1\omega_1}{2L}.$$

Thus

$$q_3 = \frac{q_1I_1\omega_1 + q_2I_2\omega_2}{L - I_3\omega_3}, \qquad q_0 = \frac{q_1I_2\omega_2 - q_2I_1\omega_1}{L - I_3\omega_3}, \tag{6.24}$$

Inserting this in (6.23) yields

$$\dot{q}_1 = \frac{(I_2 - I_1)\omega_1\omega_2}{2(L - I_3\omega_3)}q_1 - \frac{2E^2 - L\omega_3}{2(L - I_3\omega_3)}q_2,$$
$$\dot{q}_2 = \frac{2E^2 - L\omega_3}{2(L - I_3\omega_3)}q_1 + \frac{(I_2 - I_1)\omega_1\omega_2}{2(L - I_3\omega_3)}q_2.$$

We have in particular that

$$q_1\dot{q}_2 - q_2\dot{q}_1 = \frac{(I_2 - I_1)\omega_1\omega_2}{2(L - I_3\omega_3)}(q_1^2 - q_2^2) + \frac{2E^2 - L\omega_3}{L - I_3\omega_3}q_1q_2.$$

We can write

$$(q_1, q_2) = \sqrt{\frac{L - I_3\omega_3}{2L}}(\cos\theta, \sin\theta), \tag{6.25}$$

and hence

$$q_1\dot{q}_2 - q_2\dot{q}_1 = \frac{L - I_3\omega_3}{2L}\dot{\theta}.$$

All in all we have

$$2\dot{\theta} = \frac{(I_2 - I_1)\omega_1\omega_2}{L - I_3\omega_3}(\cos^2\theta - \sin^2\theta) + \frac{2E^2 - L\omega_3}{L - I_3\omega_3}2\cos\theta\sin\theta$$
$$= \frac{(I_2 - I_1)\omega_1\omega_2}{L - I_3\omega_3}\cos 2\theta + \frac{2E^2 - L\omega_3}{L - I_3\omega_3}\sin 2\theta. \tag{6.26}$$

Problems

6.2.1 Show that if we consider the real part of a quaternion as a scalar $\Re(q) \in \mathbb{R}$, and the imaginary part as a vector $\Im(q) \in \mathbb{R}^3$, then multiplication of quaternions is given by

$$\Re(q\,r) = \Re(q)\,\Re(r) - \Im(q) \cdot \Im(r) \tag{6.27}$$

$$\Im(q\,r) = \Re(q)\,\Im(r) + \Re(r)\,\Im(q) + \Im(q) \times \Im(r) \tag{6.28}$$

Where we use the basis i, j, k to identify $\Im(\mathbb{H})$ with $\mathbb{R}^3$ so '$\cdot$' and '$\times$' makes sense.

6.2.2 Show that

$$\langle q, r\rangle = \Re(q\,\bar{r}) \tag{6.29}$$

defines an inner product on $\mathbb{H}$. Show that $1, i, j, k$ are orthonormal with respect to this inner product and that the modulus is given by $|q|^2 = \langle q, q\rangle$.

6.2.3 Let $z_1, z_2, w_1, w_2 \in \mathbb{C}$, then $z_0 + kz_1, w_0 + kw_1 \in \mathbb{H}$. Show that

(a) $(w_1 + kw_2)(z_1 + kz_2) = w_1\,z_1 - \overline{w}_2\,z_2 + k(w_2\,z_1 + \overline{w}_1\,z_2)$

(b) $\overline{z_1 + kz_2} = \bar{z}_1 - kz_2$

6.2.4 Prove (6.10)–(6.12).

6.2.5 Consider the four 2×2 complex matrices

$$\sigma_0 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}.$$

The matrices $\sigma_1, \sigma_2, \sigma_3$ are called the *Pauli spin matrices*. Show that

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = -\sigma_0$$
$$\sigma_1\sigma_2 = -\sigma_2\sigma_1 = \sigma_3, \quad \sigma_2\sigma_3 = -\sigma_3\sigma_2 = \sigma_1, \quad \sigma_3\sigma_1 = -\sigma_1\sigma_3 = \sigma_2.$$

Consider the map $\sigma : \mathbb{H} \to \mathrm{span}_{\mathbb{R}}\{\sigma_0, \sigma_1, \sigma_2, \sigma_3\} \subset \mathbb{C}^{2\times 2}$ given by

$$\sigma : q_0 + iq_1 + jq_2 + kq_3 \mapsto q_0\,\sigma_0 + q_1\,\sigma_1 + q_2\,\sigma_2 + q_3\,\sigma_3.$$

Show that

$$\sigma(q + p) = \sigma(q) + \sigma(p)$$
$$\sigma(q\,p) = \sigma(q)\,\sigma(p)$$
$$\sigma(\overline{q}) = \overline{\sigma(q)}^T$$
$$|q|^2 = \det(\sigma(q))$$

where $\underline{\underline{A}}^T$ denotes the transpose of the matrix $\underline{\underline{A}}$.

6.2.6 Let $L_q : r \mapsto qr$ and $R_q : r \mapsto rq$ be left and right multiplication respectively. Show that the maps are linear over the reals and that their matrices with respect to the basis $1, i, j, k$ are

$$\underline{\underline{L_q}} = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & -q_3 & q_2 \\ q_2 & q_3 & q_0 & -q_1 \\ q_3 & -q_2 & q_1 & q_0 \end{bmatrix}, \qquad \underline{\underline{R_q}} = \begin{bmatrix} q_0 & -q_1 & -q_2 & -q_3 \\ q_1 & q_0 & q_3 & -q_2 \\ q_2 & -q_3 & q_0 & q_1 \\ q_3 & q_2 & -q_1 & q_0 \end{bmatrix},$$

6.2.7 Let $q_0, q_1, q_2, q_3 \in \mathbb{R}$ and put $q = q_0 + iq_1 + jq_2 + kq_3 \in \mathbb{H}$. Determine $q\,i\,\overline{q}$, $q\,j\,\overline{q}$, and $q\,k\,\overline{q}$. Prove that the map $\mathbb{H} \to \mathbb{H} : r \mapsto q\,r\,\overline{q}$ is linear over the reals and has the matrix

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & q_0^2+q_1^2-q_2^2-q_3^2 & 2(q_1q_2-q_0q_3) & 2(q_1q_3+q_0q_2) \\ 0 & 2(q_1q_2+q_0q_3) & q_0^2-q_1^2+q_2^2-q_3^2 & 2(q_2q_3-q_0q_1) \\ 0 & 2(q_1q_3-q_0q_2) & 2(q_2q_3+q_0q_1) & q_0^2-q_1^2-q_2^2+q_3^2 \end{bmatrix}$$

with respect to the basis $1, i, j, k$.

6.2.8 Let $q_0, q_1, q_2, q_3 \in \mathbb{R}$ and put $q = q_0 + iq_1 + jq_2 + kq_3 \in \mathbb{H}$. Determine $q\,i - i\,q$, $q\,j - j\,q$, and $q\,k - k\,q$. Prove that the map $\mathbb{H} \to \mathbb{H} : r \mapsto q\,r - r\,q$ is linear over the reals and has the matrix

$$2\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -q_3 & q_2 \\ 0 & q_3 & 0 & -q_1 \\ 0 & -q_2 & q_1 & 0 \end{bmatrix}$$

with respect to the basis $1, i, j, k$.

6.3 Rational curves in the rotation group

If we want to specify a continuous motion through $\mathbb{R}^3$ then the translation is specified by a parametrized curve in $\mathbb{R}^3$ and the rotation by a parametrized curve in $SO(3)$. As we have just seen the latter can be given by homogeneous coordinates, i.e., we just need a parametrized curve in $\mathbb{R}^4 \setminus \{\mathbf{0}\}$. Recall that a rational curve in $\mathbb{R}^3$ is given by a polynomial curve in $\mathbb{R}^4 \setminus \{\mathbf{0}\}$, and that we normally avoid $\mathbf{0}$ by having *positive* weights $\omega_i > 0$.

A (piecewise) polynomial curve of degree n in $\mathbb{R}^4 \setminus \{\mathbf{0}\}$ gives us a (piecewise) rational curve of degree $2n$ in $SO(3)$. Conversely, in [18] it is shown that an (piecewise) irreducible rational curve of degree $2n$ in $SO(3)$ can be obtained as the image of a (piecewise) polynomial curve of degree n in $\mathbb{R}^4 \setminus \{\mathbf{0}\}$. We don't want to restrict ourselves to curves where all the control points have a positive fourth coordinate, so we have to exercise some care in order to avoid $\mathbf{0}$. Using the convex hull property it is sufficient to demand that $\mathbf{0}$ is outside the convex hull of the control points, and in the case of a B-spline curve of degree n the strong convex hull property shows that it is sufficient that $\mathbf{0}$ is outside the convex hull of any $n+1$ consecutive control points.

A continuous rotation is often obtained by interpolation between certain fixed rotations (like 'key frames' in a computer animation or 'taught positions' in robotics). First we find *unit* quaternions given these rotations. They are determined up to a sign and we choose the signs such that neighbouring quaternions are as close to each other as possible. We can perform a normal interpolation in $\mathbb{R}^4$ so as to obtain a B-spline curve. Normally the interpolation points are so close together that the resulting curve stays close to the unit sphere in $\mathbb{R}^4$, i.e., it avoids $\mathbf{0}$ and thus can be used to define a curve of rotations that interpolates the given rotations.

The image of a *linear* Bézier curve $(1-t)q + tr$, $t \in [0, 1]$ in $\mathbb{H} \setminus \{\mathbf{0}\} \cong \mathbb{R}^4 \setminus \{\mathbf{0}\}$ is essentially a *rotation* around a fixed axis. More precisely

$$(1-t)q + tr = q\big((1-t) + tq^{-1}r\big),$$

and we see that the continuous motion is composed by a *fixed* rotation given by $q \in \mathbb{H}$ and a continuous rotation $(1-t) + tq^{-1}r$ around a fixed axis $\Im(q^{-1}r)$. The trajectories are of course circles around the axis $\operatorname{Ad}(q)\Im\big(q^{-1}r\big) = \Im(rq^{-1})$.

Problems

6.3.1 Let $q(t)$ be a rational Bézier curve in $\mathbb{H} \setminus \{0\}$ with control points $q^0, \dots, q^n$ and weights $\omega^0, \dots, \omega^n$, and let $ix_1 + jx_2 + kx_3 \in \Im(\mathbb{H}) \cong \mathbb{R}^3$. Find the control points and the weights for the Bézier curve $\operatorname{Ad}\big(q(t)\big)x$ in $\Im(\mathbb{H})$ in the case where the degree of q is 1 and in the case where the degree is 2.

Bibliography

[1] V.I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, Berlin, Heidelberg, New York, 2 edition, 1980.

[2] P. Bézier. Définition numérique des courbes et surfaces i. *Automatisme*, 11:625–632, 1966.

[3] P. Bézier. Définition numérique des courbes et surfaces ii. *Automatisme*, 12:17–21, 1967.

[4] W. Boehm, G. Farin, and J. Kahmann. A survey of curve and surface methods in CAGD. *Computer Aided Geomatric Design*, 1:1–60, 1984.

[5] Wolfgang Boehm and Hartmut Prautzsch. *Numerical methods*. A. K. Peters Ltd., Wellesley, MA, 1993.

[6] Wolfgang Boehm and Hartmut Prautzsch. *Geometric concepts for geometric design*. A. K. Peters Ltd., Wellesley, MA, 1994.

[7] Carl de Boor. *A Practical Guide to Splines*. Applied Mathematical Sciences. Springer-Verlag, New York, Berlin, 1978.

[8] Gerald Farin. *Curves and Surfaces for Computer Aided Geometric Design. A Practical Guide*. Academic Press, London, 1988.

[9] J. Ferguson. Multivariable curve interpolation. *JACM*, 11(2):221–228, 1964.

[10] Jean Gallier. *Geometric methods and applications. For computer science and engineering*. Springer-Verlag, New York, 2001.

[11] William C. Graustein. *Differential geometry*. Dover Publications, New York, 1966.

[12] Jens Gravesen. Adaptive subdivision and the length and energy of Bézier curves. *Computational Geometry*, 8:13–31, 1997.

[13] Jens Gravesen. de Casteljau's algorithm revisited. In Morten Dæhlen, Tom Lyche, and Larry L. Schumaker, editors, *Mathematical Methods for Curves and Surfaces II*, pages 221–228, Nashville & London, 1998. Vanderbilt University Press.

[14] Jens Gravesen and Christian Henriksen. The geometry of the scroll compressor. *SIAM Review*, 43:113–126, 2001.

[15] Alfred Gray. *Modern Differential Geometry of Curves and Surfaces with Mathematica®*. CRC Press, Boca Raton, 1998.

[16] Heinrich W. Guggenheimer. *Differential Geometry*. McGraw-Hill Book Company, Inc., New York, 1963.

[17] Josef Hoschek and Dieter Lasser. *Fundamentals of Computer Aided Geometric Design*. A.K. Peters, Wellesley, 1993.

[18] Bert Jüttler. Über zwangläufige rationale Bewegungsvorgänge. *Sb. Österr. Akad. Wiss., Abt. II*, 202:117–132, 1993.

[19] Jeppe Oskar Meyer Larsen and Christian Krog Madsen. Subdivision surfaces. "Midway project", Technical University of Denmark, 2001. `http://www.mat.dtu.dk/student-projects/2001-subdivision`

[20] Martin Lipschutz. *Theory and Problems of Differential Geometry*. Schaum's outline series. McGraw-Hill, New York, 1969.

[21] A.P. Markeev. *Theoretical Mechanics*. Nauka, Moscow, 1990. in Russian.

[22] Richard S. Millman and George D. Parker. *Elements of Differential Geometry*. Prentice Hall, Englewood Cliffs, New Jersey, 1977.

[23] Parviz E. Nikravesh. *Computer-Aided Analysis of Mechanical Systems*. Prentice-Hall International, Inc., Englwood Cliffs, New Jersey, 1988.

[24] Nicholas M. Patrikalakis and Takashi Maekawa. *Shape interrogation for computer aided design and manufacturing*. Springer-Verlag, Berlin, 2002.

[25] Michael A. Penna and Richard R. Patterson. *Projective Geometry and its Applications to Computer Graphics*. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[26] Les Piegl and Wayne Tiller. *The NURBS Book*. Monographs in visual communication. Springer-Verlag, Berlin, 1996.

[27] Lyle Ramshaw. Blossoms are polar form. *Computer Aided Geometric Design*, 6:323–358, 1989.

[28] Martin Raussen. *Elementary Differential Geometry: Curves and Surfaces*. Department of Mathematical Sciences, Aalborg University, Frederik Bajersvej 7G, DK-9000 Aalborg, 2001. `http://www.math.auc.dk/~raussen/MS/01/index.html`

[29] Hans-Peter Seidel. A new multiaffine approach to B-splines. *Computer Aided Geometric Design*, 6:23–32, 1989.

[30] Dirk J. Struik. *Lectures on Classical Differential Geometry*. Addison-Wesley Publishing Company, London, 1961.

[31] Bu Qing Su and Ding Yuan Liu. *Computational geometry*. Academic Press Inc., Boston, MA, 1989. Curve and surface modeling, Translated from the Chinese by Geng Zhe Chang.

[32] Michael Ungstrup. Fairing in computer aided geometric design. Master's thesis, Department of Mathematics, Technical University of Denmark, 1998.

[33] Fujio Yamaguchi. *Curves and Surfaces in Computer Aided Geometric Design.* Springer-Verlag, Berlin, Heidelberg, New York, London, Tokyo, 1988.

[34] Dennis Zorin, Peter Schröder, Tony DeRose, Leif Kobbelt, Adi Levin, and Wim Sweldens. Subdivision for modeling and animation. SIGGRAPH 2000 Course Notes, 2000. `http://mrl.nyu.edu/~dzorin/sig00course/`

Index

DEPARTMENT OF MATHEMATICS
DTU . MATEMATIKTORVET BUILDING 303
DK-2800 KGS. LYNGBY
DENMARK