Aspects of Intersection Algorithms and Approximation

Tor Dokken

Revised January 2000

Submitted
16. December 1996
to
the University of Oslo
for the doctor philosophiae degree.
Defended July 1997

.

Contents

.

Preface

My work on intersection algorithms started in 1978 when I was employed at the Central Institute for Industrial Research (SI) in Oslo, Norway. SI later merged with SINTEF, where I am currently employed. Funding of this work has been through governmental and industrial projects aiming at supplying geometric modeling tools to industry and CAD-vendors. For many years I intended to write a doctoral thesis. However, as geometric modeling activity was steadily growing, I could not find time for the work.

Then in 1992, after being a manager for five years, my colleague Dr. Morten Dæhlen proposed that he could take over management of the geometric modeling department. This suited me well as I had just become a father and wanted to spend less time on administration and traveling.

My intent, when work on the thesis started, was to get a better understanding of the intersection problem in general, and possibly extend the theoretical basis for intersection algorithms by moving into higher dimensions than $\mathbb{R}^3$. In addition, I wanted to take a closer look at the combination of algebraic and parametric representations in intersection algorithms. I started pursuing extensions of my earlier work experimenting with Mathematica. Then gradually the idea of approximative implicitization emerged, and the main bulk of the thesis was gradually directed towards approximation theory.

During the autumn of 1995, a first version of this work was compiled from documents produced during the research process. Then with very valuable advice from Professor Tom Lyche at the Institute for Informatics at the University of Oslo, this document has been molded into its final shape. During the process, some material has been removed and some material has been extended. Very late in the process the results on parametric approximation of algebraically represented manifolds emerged.

I would like to thank SINTEF Applied Mathematics for giving me sufficient time to pursue my ideas on intersection algorithms. During the final process my father, Brynjolf Dokken, has used a lot of time proof reading and simplifying the language.

I hope that my children, Jørgen and Julie, and my wife Mona have not suffered too much when my mind has been occupied with problems related to intersection algorithms. I want to thank them for giving me time to complete this work.

<div align="right">Oslo, December 1996.</div>

In December 1999 and January 2000 corrections and comments to these thesis from Prof. T.W. Sederberg and others were analyzed in detail and corrections done to the original manuscript.   Oslo, January 2000.

.

Chapter 1

Introduction

Since the early days of CAD/CAM, intersection algorithms have played a central role in making geometric modeling systems work. The first geometric modelers were curve based, and efficient algorithms were developed giving satisfactory intersection results. One example of such a system was the AUTOKON system, a complete batch oriented CAD/CAM-system for ship building developed in Norway in the sixties and seventies. However, when the development of surface and volume based systems started, intersection algorithms became more complex.

In curve based systems the result of an intersection is points or segments of the curves being intersected. Thus, describing the intersection is fairly straight forward. In surface based systems the result of an intersection is points, curves, and regions of the surfaces being intersected. The surfaces being intersected are often rational parametric piecewise polynomial surfaces. It is well know that, in most cases, the curves resulting from the intersection of two such parametric surfaces cannot be represented by rational or nonrational parametric piecewise polynomial curves, e.g. see [Bajaj:93]. To harmonize with the formats already used in the systems, the intersection curves are expected to be described by such curves. Since in most cases no exact parametric representation exists, approximations are introduced.

The nature of the approximations employed in a specific system is dependent on the strategies used for designing the geometric modeling system as well as the requirement for geometric quality in the system. Intersection has thus become an important issue in the competition between vendors of CAD/CAM-systems and between vendors of geometric modeling technology. Thus, the commercial players in the CAD/CAM and geometric modeling market are very restrictive with what they publish on their intersection algorithms. As a consequence, most papers published on such algorithms originate from academic institutions.

1.1 Intersection Algorithms Developed in Industrial Projects

The author of this thesis is employed at a noncommercial research institute in Norway called SINTEF where he has been working with intersection algorithms since 1979. The projects financing the research and development of intersection algorithms have been funded both by government and industry. As a consequence, some results have been published, while others have not.

The project initiating this work was an Inter-Nordic project named GPM (Geometric Product Models 1978-1981). The goal of the project was to develop FORTRAN subroutine packages for geometric modeling. The work at SINTEF in Oslo was focused on sculptured surfaces. Among the results was a detailed specification [GPM-16:80] of a sculptured surface system and a partial implementation of the specification. The most important part relevant to this thesis was a B-spline library [GPM-26:83]. The library contained both algorithms for intersection of two B-spline represented curves and for intersecting B-spline represented curves with first and second degree algebraic curves and surfaces. Recursive subdivision was performed by subdivision using the [Oslo-algorithm].

The work from the GPM-project was carried on in a German-Norwegian project named APS (Advanced Production System 1981-1987). A result relevant to this thesis was a subroutine package named APS B-spline library [APS:87]. APS B-spline library consists of a wide range of intersection algorithms including the intersection of two B-spline represented surfaces and the intersection of B-spline represented surfaces and first and second degree algebraic surfaces. A first publication of results from this work can be found in [Dokken:85]. The main topic of this paper was the combination of B-spline curves and surfaces with algebraic surfaces. In addition examples of recursive subdivision techniques used for loop detection were presented. This work on loop detection was further elaborated in [Dokken:89]. The paper also addressed strategies for where to subdivide a surface to get what is denoted a "simple intersection case", i.e. identifying intersection situations with no internal loops.

In 1988 the development of a new spline library, SISL (SINTEF Spline Library) was started at SINTEF. The new properties in this library were:

- Programmed in C.

- Double precision instead of single precision.

- Combine recursive subdivision and iteration to speed up calculations.

- New marching algorithms to get better curve tracks with less data.

The marched curves in SISL are represented by piecewise cubic Hermite polynomials, and represented in a nonuniform B-spline basis. The tangent lengths used in the Hermite interpolant are based on the circle approximation methods described in [Dokken:90].

SISL was delivered to Hewlett-Packard in 1989 and is the basis of NURBS technology used in the solid modeling system Precision Engineer: Solid Designer. This system is marketed by the Hewlett-Packard company CoCreate. In 1996 the development of SISL is still continuing in close cooperation with CoCreate.

In 1992 I decided that the thesis should approach the intersection problem from a theoretical angle. The idea was to combine my long practical experience in the development of intersection algorithms with a generic approach to the intersection problem. Thus, the title of the thesis is "Aspects of Intersection Algorithms and Approximation". No new and better intersection algorithm is presented. However, some central issues in the development of intersection algorithms are addressed. The most important of these are:

- Intersection within a given tolerance.

- Loop elimination.

- Approximation of parametrically represented manifolds by algebraic hypersurfaces.

- Approximation of algebraically represented manifolds by parametrically represented manifolds.

The approximation problems above are given the most attention. This is because better approximation methods are important in both loop elimination and the representation of intersection results.

The thesis is not limited to the intersection of curves and surfaces in $\mathbb{R}^2$ and $\mathbb{R}^3$, but addresses problems related to:

- Intersection of manifolds of possibly different dimensions in $\mathbb{R}^l$, $l \geq 1$.

- Intersection of a manifold and a number of hypersurfaces in $\mathbb{R}^l$, $l \geq 1$.

1.2   Relation to Trends

In [Hohmeyer:92] intersection algorithms were classified into two groups:

- **Re-Approximation Techniques.** The geometries to be intersected are approximated with a large number of simpler geometric elements, thus transforming the problem to a simpler but more voluminous geometric problem. When high accuracy is necessary or the intersections are singular, e.g. the intersection between two surfaces is a curve where the normals of the two surfaces are parallel, the re-approximation methods have limited applicability.

- **Direct Decomposition Methods.** Among these *Loop Detection Decomposition* is the most general.

In this thesis the basic philosophy is *Loop Detection Decomposition*.

1.3   Why is Research on Intersection Algorithms Important?

Design and production are steadily growing more computer based. Thus, the design and production process is influenced by the capabilities of the computer systems supporting these processes.

Much effort has been put into making faster and more stable hardware. An other equally important issue is the development of better software. The software for modeling of complex geometric shape must, to be of practical use, satisfy the following criteria:

- Easy to use.

- Have a sufficient response to the problem posed.

- Produce geometric models of high quality, i.e., to produce geometric models that have sufficient accuracy with a description that is as compact as possible.

If the intersection algorithms in such a system are inaccurate or too slow, the user will have to reduce his design ambition and make geometric models with a lower quality than originally intended. Thus, robust, fast and accurate algorithms are important to enable better solutions in a product development process.

1.4 Summary

In Chapter 2 we present problems related to the intersection of two compact manifolds of possibly different dimensions in $\mathbb{R}^l$, $l \geq 1$. Topics addressed are:

- A generic structure for intersection algorithms.

- The intersection of two compact sets within a tolerance.

- Representations of intersections within a tolerance.

- Separating objects to determine that no intersection exists.

- Where to subdivide objects being intersected to decide if an intersection result is a single object or two or more objects.

- Identification of situations where the result of the intersection is objects touching boundaries of one of the manifolds intersected.

In Chapter 3 we look at necessary properties for representing hypersurfaces and parametric manifolds to enable stable numeric intersection calculations. The much used tensor product Bernstein and B-splines basis functions, as well as the Bernstein basis over a simplex, satisfy these requirements. Chapter 3 ends with the description of the representation of algebraic hypersurfaces in barycentric coordinates.

Barycentric coordinates are also paramount in Chapter 4 where we look at combining a parametrically represented manifold and an algebraically represented hypersurface to a function over the parameter domain of the manifold. Such combinations of the two representations can be used for:

- Reformulating the intersections between a parametrically represented manifold and algebraic hypersurface(s) to the problem of finding the zeroes of one or more functions.

- Approximating a parametrically represented manifold with an algebraic hypersurface. This type of approximation can be used for changing an intersection of parametrically represented manifolds to the intersection of algebraically represented hypersurfaces and one parametrically represented manifold.

The behavior of these combinations is presented in Section 4.1. In the subsequent sections different aspects of such combinations are addressed. In

Section 4.9 an algorithm, named "Approximative Implicitization" for finding an algebraic approximation to a parametrically represented manifold, is presented.

In Chapter 5 we look at how to represent intersection results by parametric or algebraic methods. In addition, a method for approximation of algebraically represented manifolds by parametrically represented manifolds is presented.

As the intersection of parametric surfaces in $\mathbb{R}^3$ can be represented as curves in the parameter domain of the surfaces, we devote Chapter 6 to the problem of approximation of curves in $\mathbb{R}^2$. The topic addressed is high accuracy cubic Hermite approximation of:

- Curves where both the algebraic equation and parametric representation is known.

- Curves where only the algebraic equation is known.

- Curves where only the parametric representation is known.

By high accuracy cubic Hermite approximation we here mean methods that are $O(h^6)$.

In Chapter 7 these results are applied to cubic Hermite ellipse approximation, with circle approximation as a special case. To illustrate the applicability of the results in Chapter 5, $O(h^8)$ circle interpolants of degree 4 are also included in Chapter 7.

Two appendices are added to relate the work to the functionality needed in geometric modeling systems:

- The concept of PosProd Basis Functions defined in Chapter 3 is in Appendix A discussed in relation to the B-spline and Bernstein bases.

- In Appendix B we address how a number of geometric interrogation problems can be reformulated to manifold intersection.

1.5   Problems in Need of More Research

As stated at the start of the introduction, the topic "Intersection Algorithms and Approximation" is so complicated that much research is still needed. As computers grow more powerful, more complex problems can be solved, and solutions can be more accurate. Within the topic "Intersection Algorithms and Approximation" two of the issues still needing research attention are:

1. "Loop elimination". Better results are needed for predicting when the result of an intersection cannot contain internal loops. One possible direction to follow is bounding the partial derivatives of manifolds, as described in [**?**]. Another direction is acquiring better understanding of loop elimination when intersecting manifolds in higher dimensional spaces than $\mathrm{I\!R}^3$.

2. "High accuracy approximation". The main bulk of this thesis is devoted to:

   - High accuracy approximation of parametrically represented manifolds by algebraic hypersurfaces.

   - High accuracy approximation of algebraically represented manifolds by parametrically represented manifolds.

   Within both of these topics there are still many unsolved problems. My feeling is that the work in this thesis is only touching upon this rich field of research.

Chapter 2

A Generic Intersection Algorithm

A great challenge for developers of intersection algorithms is to solve the discrepancy between our natural conception of an intersection, and the mathematical model behind the intersection algorithms in CAGD systems.

When making an intersection algorithm for curves and surfaces, it is easy to base the implementation on well known concepts from set theory instead of trying to model the more inaccurate human conception of what an intersection is. The approach based on intersection of sets, gives the expected results when the geometries are transversal, i.e. when the objects are not parallel in the region around the expected intersection occurrence. However, in regions where the geometries are nonintersecting, but close and near parallel, we want that an intersection is identified. It is easy to say that these situations, also denoted singular and near singular, are very special and rarely occur. This is true in CAGD systems with a user interface that forces the user away from such situations. However, looking at shapes that seem natural to man, they are full of surfaces touching in a singular or near singular manner. Thus, to be able to develop more natural user interfaces, the intersection algorithms have to support the human affection for singularities and near singularities. It is thus natural to introduce intersection tolerances as a basic concept in a model of the intersection problem. In Section 2.2 we try to formulate such a model supporting tolerances. The model requires a number of concepts from set theory. These are listed in Section 2.1 together with concepts from differential geometry needed in later discussions.

Although intersection algorithms have to be tailored to the specific requirements of an application and the dimensionality of the intersection problem at hand, a generic structure can be made for intersection algorithms based on recursive subdivision. The only provision is that the objects to be

intersected are closed and bounded manifolds. The structure is independent of the dimension of the space in which the manifolds lie. The discussions in the current chapter and the chapters following thus address, when possible, manifolds in $\mathbb{R}^l$, $l \geq 1$. We impose, only when necessary, additional properties on the manifolds or require that the space in which the manifolds lie has a certain dimension.

The basic structure of the generic intersection algorithm for two closed and bounded manifolds A and B is as follows:

Algorithm Intersect(A, B)

Intersect(Boundary(A),B)
Intersect(A,Boundary(B))

IF | *All objects resulting from the intersection touch the boundary of* A *or* B. | THEN

Describe the intersection results within an appropriate tolerance, and connect to boundary intersections.

ELSE

Subdivide A and B in an appropriate way into smaller pieces.

For *All subpieces sub_i(A) of* A
    For *All subpieces sub_j(B) of* B
        Intersect(sub_i(A), sub_j(B))

Connect the results of the subproblem intersections.

The algorithm above structures the intersection problem as a number of subproblems:

- What is an intersection? What is the consequence of introducing intersection tolerance? This is discussed in Section 2.2.

- In Section 2.5.3, we look at the intersection of a manifold with the boundary of another manifold, and conclude that subdivision of the

boundary is necessary in most cases. The only exception is when the boundary consists of only single points.

- One of the most challenging problems in intersection algorithms is loop detection; or phrased differently: Loop elimination. By this we mean finding situations where the results of an intersection only consist of objects touching the boundary of one of the manifolds intersected. This is discussed in Section 2.5. We also address the possibility to reduce the complexity of an intersection problem by combining parametric and algebraic descriptions of manifolds.

- The representation of the results of an intersection is also an important issue. What is represented is critical to the performance of the intersection algorithm. The introduction of intersection tolerances introduce more complexity in the representation. This is discussed in Section 2.3. How accurately the intersection results are represented is another problem area. This is addressed in Chapter 5.

- Strategies for subdividing an intersection problem into subproblems are addressed in Section 2.4. The aim is to find subproblems that have intersection results with a high probability of touching the boundary of one of the manifolds intersected.

- Collection of the intersection results from subproblems into a complete intersection result is mainly a software engineering problem and is not addressed further here. However, it should be noted that what is expected to be identical results from two related subproblems, will often be affected with different rounding errors. Thus, great care has to be taken when designing an intersection algorithm.

In this chapter we do not address the representation of the manifolds. The only exception is that we use the fact that both parametric and algebraic descriptions exist for manifolds. Other issues related to geometry representation are addressed in Chapter 3. The major bulk of this chapter is published in [Dokken:97].

2.1   Basic Concepts Used

In the discussions in this chapter, we use basic concepts from set theory and differential geometry. We do this to show that the results are valid for a wide range of representation methods for geometric objects. Currently

most implementations of intersection algorithms are developed for a specific geometry representation.

We will address geometric objects with the following properties:

- **Compact sets.** A set is compact if it closed and bounded. A set A is closed if and only if
$$\neg int(\neg A) = A,$$
where $\neg A$ is the compliment of the set A, and $int(\neg A)$ is the internal of the set $\neg A$.

- **Manifold.** A g dimensional *manifold* $M^g \subset \mathbb{R}^l$, where $l \geq g$, is a subspace that is *locally* homeomorphic to $\mathbb{R}^g$. That is, for every point $\mathbf{p}$ of M^g, there exists a neighborhood U of $\mathbf{p}$ that is homeomorphic to $\mathbb{R}^g$, i.e. that there is locally continuous one to one correspondence between the M^g and $\mathbb{R}^g$. See [Hoffmann:89] page 49.

  We also refer to a g dimensional manifold as a g-manifold. Thus, a 1-manifold in $\mathbb{R}^2$ is a curve, and a 2-manifold in $\mathbb{R}^3$ is a surface.

- **Manifold with boundary.** A g dimensional *manifold* $M^g \subset \mathbb{R}^l$, where $l \geq g$, *with boundary* is a subspace whose boundary points have neigborhoods that are homeomorphic to $\mathbb{R}^{g+}$, the positive half-space,
$$\mathbb{R}^{g+} = \left\{ (x_1, \ldots, x_g) \in \mathbb{R}^g \mid x_1 \geq 0 \right\},$$
and whose interior points have neighborhoods that are homeomorphic to $\mathbb{R}^g$. See [Hoffmann:89] page 49.

- **Differentiable manifold**. A *differentiable manifold* of dimension g is a set M^g and a family of injective mappings $\mathbf{x}_\alpha : U_{\alpha:} \to M^g$ of open sets U_α of $\mathbb{R}^g$ such that:

  1. $\bigcup_\alpha \mathbf{x}_\alpha(U_\alpha) = M^g$ ($\bigcup_\alpha$ is the union of the subsets $\mathbf{x}_\alpha(U_\alpha)$ for all α).
  2. For any pair α, β with $\mathbf{x}_\alpha(U_\alpha) \bigcap \mathbf{x}_\beta(U_\beta) = W \neq \emptyset$, the set $\mathbf{x}_\alpha^{-1}(W)$ and $\mathbf{x}_\beta^{-1}(W)$ are open sets in $\mathbb{R}^g$ and the mapping $\mathbf{x}_\beta^{-1} \circ \mathbf{x}_\alpha$ is differentiable.

  The pair $(U_\alpha, \mathbf{x}_\alpha)$ (or the mapping $\mathbf{x}_\alpha$) with $\mathbf{p} \in \mathbf{x}_\alpha(U_\alpha)$ is called a parametrization (or system of coordinates) of M^g at $\mathbf{p}$. See [do Carmo:92] page 2.

- C^r **continuous manifold.** In [Spivak:79] the concept C^r manifolds are introduced by requiring that the functions $\mathbf{x}_\alpha$ and $\mathbf{x}_\beta$ have partial derivatives that are continuous up to an order r.

We used the following concepts in these definitions:

- Homeomorphic function: A function that is one-to-one and has a continuous inverse.

- Injective function/mapping: $x \neq y$ implies $f(x) \neq f(y)$.

In geometry representations used in CAD/CAM and CAGD, most parametric representation of curves, surfaces and higher dimensional geometric objects are both:

- g-manifolds with boundaries.

- C^r continuous manifolds.

To easily address manifolds with these properties, we make the following definition.

Definition 1 (C^r manifolds with boundary) *By "C^r manifolds with boundary" we mean manifolds that:*

- *Satisfy the conditions of being a "g dimensional manifold with boundary".*

- *Satisfy the conditions for being a "C^r manifold", for $r \geq 1$.*

Definition 2 (Smooth manifold) *We call a "C^1 manifold" a smooth manifold.*

2.2   Geometric Tolerance and Intersection

What do we mean when we say that two objects touch within a tolerance $\epsilon \geq 0$? Although we want to give an answer to the question for the intersection of bounded C^r manifolds with boundaries, we use in this section only properties of closed sets. We do this to avoid introducing unnecessary details in the discussions. The closed sets include manifolds with boundaries. Thus, the discussion covers geometric objects that are just points, NURBS curves with positive weights, NURBS surfaces with positive weights, bounded algebraically represented curves and surfaces.

The following definition is an attempt to formally define how we intuitively interpret an intersection.
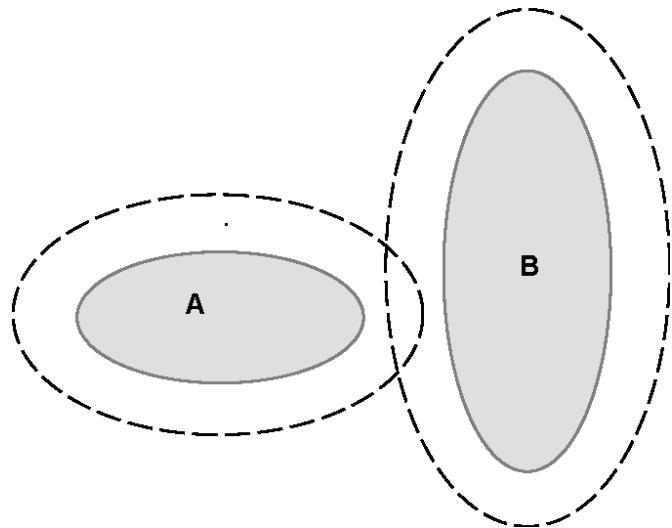
Figure 2.1: The ϵ-intersection of two sets **A** and **B**, when the ϵ-intersection is empty. Around each point in both sets we make a hypersphere with radius ϵ. If this hypersphere touches the other set, an ϵ-intersection exists.
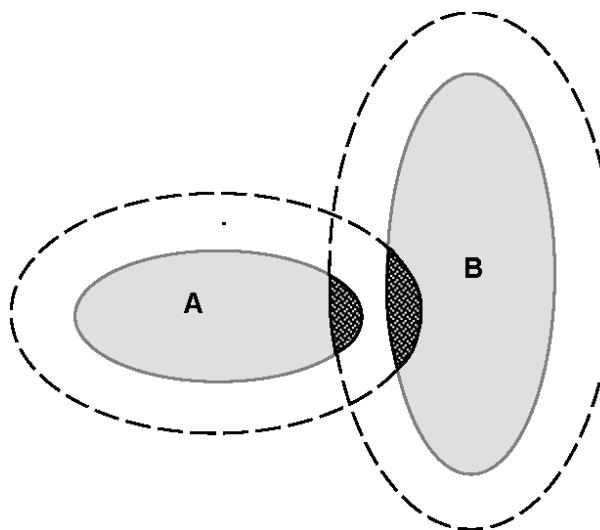
Figure 2.2: The ϵ-intersection of two sets **A** and **B**, when the sets nearly touch, but have no true intersection. If the correspondence between points in the two sets is to be represented as part of the intersection result, then the dimensionality of the intersection result will be higher than the dimensionality of the sets being intersected.
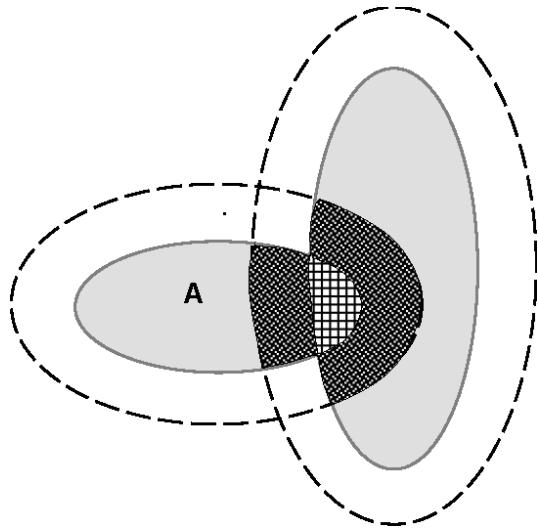
Figure 2.3: The ϵ-intersection of two sets $\mathbf{A}$ and $\mathbf{B}$, when the sets have a true intersection. In this case the true intersection is of more interest than the ϵ-intersection.

Definition 3 (ϵ-intersection) *The ϵ-intersection of two closed sets A and B in $\mathbb{R}^l$ is defined by*

$$A \cap_\epsilon B = \left\{ (\mathbf{p}, \mathbf{q}) \in \mathbb{R}^l \times \mathbb{R}^l \mid \mathbf{p} \in A \wedge \mathbf{q} \in B \wedge \|\mathbf{p} - \mathbf{q}\|_2 \leq \epsilon \right\}.$$

Remark 1 *For $\epsilon < 0$, $A \cap_\epsilon B$ is empty, which should be expected. Further $A \cap_0 B = A \cap B$.*

In figure 2.1, figure 2.2 and figure 2.3 we visualize the ϵ-intersection in different situations.

To simplify the representation of intersections it is advantageous that the result of an ϵ-intersection is closed. The following lemma proves such a property for the ϵ-intersection.

Lemma 4 *If A and B are closed sets, then $A \cap_\epsilon B$ is a closed set.*

Proof. We have to prove that $\neg int(\neg (A \cap_\epsilon B)) = A \cap_\epsilon B$ to show that $A \cap_\epsilon B$ is closed. By $\neg int(X)$ we mean the internal of the set X.

$$\begin{aligned}
\neg int(\neg\, & (A \cap_\epsilon B)) = \\
& \neg int(\neg \left\{ (\mathbf{p}, \mathbf{q}) \in \mathbb{R}^l \times \mathbb{R}^l \mid \mathbf{p} \in A \wedge \mathbf{q} \in B \wedge \|\mathbf{p} - \mathbf{q}\|_2 \leq \epsilon \right\}) \\
& = \neg int(\left\{ (\mathbf{p}, \mathbf{q}) \in \mathbb{R}^l \times \mathbb{R}^l \mid \mathbf{p} \in \neg A \vee \mathbf{q} \in \neg B \vee \|\mathbf{p} - \mathbf{q}\|_2 > \epsilon \right\}) \\
& = \neg \left\{ (\mathbf{p}, \mathbf{q}) \in \mathbb{R}^l \times \mathbb{R}^l \mid \mathbf{p} \in \neg A \vee \mathbf{q} \in \neg B \vee \|\mathbf{p} - \mathbf{q}\|_2 > \epsilon \right\} \\
& = A \cap_\epsilon B.
\end{aligned}$$
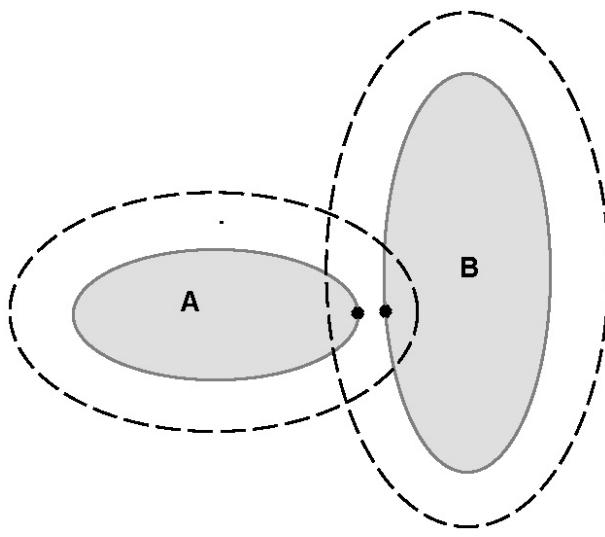
Figure 2.4: The ϵ-intersection represented by the closest point between the sets **A** and **B**.

The $int()$ can be removed since the conditions $\mathbf{p} \in \neg A$, $\mathbf{q} \in \neg B$ and $\|\mathbf{p} - \mathbf{q}\|_2 > \epsilon$ are all conditions giving open sets, and the interior of an open set is the open set itself. ∎

A closed set can contain a compound of points, curves and surfaces. By restricting the sets being intersected to contain one manifold with boundaries, we cover the needs in CAGD.

2.3 Representation of the ϵ-Intersection

Can the definition of the ϵ-intersection be of practical use in intersection algorithms? An ϵ-intersection can consist of disjoint regions, and contain manifolds of higher dimension than the manifolds being intersected. To address these issues we introduce:

- The separated ϵ-intersection to split the intersection results into disjoint subsets.

- The reduced separated ϵ-intersection. A first alternative to reduce the dimension of the manifolds representing the separated ϵ-intersection.

- The projected separated ϵ-intersection. A second alternative for reducing the dimension of the manifolds representing the separated ϵ-intersection.

As stated above, we shall first address the representation of the intersection $A \cap_\epsilon B$. Let the set A consist of one closed and bounded manifold of dimension g_A, and the set B consist of one closed and bounded manifold of dimension g_B. The set $A \cap_\epsilon B$ consists of one or more possibly touching closed and bounded manifolds of dimension g (the g can be different for each of the manifolds describing the intersection), where $g = 0, \ldots, g_A + g_B$. In non-manifold solid modeling, see [Weiler:86], topological data structures are built representing the connectivity between manifolds of different dimensions, as well as data structures for representing each manifold.

To separate non-connected manifolds we want to split the intersection set into disjoint intersection subsets.

Definition 5 (Separated ϵ-Intersection) *Let $A \cap_\epsilon B$ be the ϵ-intersection of A and B. The separated ϵ-intersections of A and B are set $(A \cap_\epsilon B)_i$, $i = 1, \ldots, N_\epsilon$ that satisfy*

$$A \cap_\epsilon B = \bigcup_{i=1}^{N_\epsilon} (A \cap_\epsilon B)_i \quad \wedge \quad \bigcup_{\substack{i,j=1 \\ i \neq j}}^{N_\epsilon} \left((A \cap_\epsilon B)_i \cap (A \cap_\epsilon B)_j \right) = \emptyset,$$

where $(A \cap_\epsilon B)_i$, $i = 1, \ldots, N_\epsilon$ cannot be separated into disjoint subsets.

Now we face the practical problem of the representation of each subset in the separated ϵ-intersection.

Example 6 *Let A and B be two nondegenerate C^1-continuous NURBS surfaces in $\mathbb{R}^3$, each of the NURBS surfaces is a bounded 2-manifold with boundaries. The ϵ-intersection of A and B, and the separated ϵ-intersection contain one or more of possibly intersecting 0-manifolds, 1-manifolds, 2-manifolds, 3-manifolds and 4-manifolds.*

The reason for introducing the tolerances was to treat as intersections cases where the objects do not intersect, but where the distance between the objects is so small that for practical purposes we should handle the case as an intersection. Thus, we are not interested in the complete description of the ϵ-intersection, but interested in the fact that we have an intersection or a near intersection. We want to find a practical way of representing the surface intersection in example 6 with manifolds with dimension $g = 0, 1, 2$. In a general case, when we intersect a manifold of dimension g_A with a manifold of dimension g_B, we want the intersection to be represented by a set of manifolds where the dimension of the manifolds is less than $\min(g_A, g_B)$.

As stated in the start of the section two alternatives for a reduced representation are:
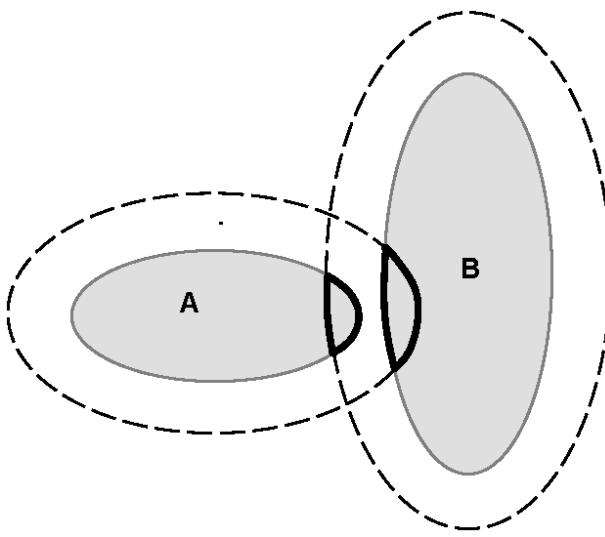
Figure 2.5: The boundary of the ϵ-intersection projected onto the sets $\mathbf{A}$ and $\mathbf{B}$, assuming that the sets are in $\mathbb{R}^2$. Note that the pointwise correspondence between the points is lost by this projection. If the sets $\mathbf{A}$ and $\mathbf{B}$ are 2-manifolds in $\mathbb{R}^3$, then the projection of the intersection set onto the objects will be 2-manifolds and the boundary of the projected intersection sets will be the intersection sets themselves. However, the boundary of the two manifolds representing the projected intersection set will be the curves depicted.

1. If there is an intersection when $\epsilon = 0$ in the disjoint subset, use the intersection as the representation of the disjoint subset. If no such intersection exists use the closest points.

2. Use the border of the projection of the disjoint subset onto object A, and the border of the projection of the disjoint subset onto object B.

Which alternative to chose for the implementation of an intersection algorithm is dependent on the actual requirements to the intersection problems being solved. To cater for the needs in the first alternative, we need a set of closest points within each of the subsets of the separated ϵ-intersection.

Definition 7 (Closest points in separated ϵ-intersection) *We define the set of closest points in the disjoint subsets of the separated ϵ-intersection by*

$$C_{\epsilon,i} = \{(\mathbf{a}, \mathbf{b}) \in (A \cap_\epsilon B)_i \mid \|\mathbf{a} - \mathbf{b}\|_2 = \min_{(\mathbf{x},\mathbf{y}) \in (A \cap_\epsilon B)_i} \|\mathbf{x} - \mathbf{y}\|_2\}.$$

Now we are able to define a reduced ϵ-intersection, thus covering the first alternative in the list.

Definition 8 (Reduced Separated ϵ-Intersection) *We define the reduced version of the disjoint subsets of the separated ϵ-intersection by*

$$\mathcal{R}_{\epsilon,i} = \begin{cases} (A \cap B) \cap (A \cap_{\epsilon} B)_i & if \quad (A \cap B) \cap (A \cap_{\epsilon} B)_i \neq \emptyset \\ C_{\epsilon,i} & if \quad (A \cap B) \cap (A \cap_{\epsilon} B)_i = \emptyset \end{cases}$$

for

$$i = 1, \ldots, N_{\epsilon}.$$

As already mentioned, the reduced separated ϵ-intersection can have a complex topological structure, e.g. in surface intersection a number of intersection curves can meet in a point, or surface regions can meet other regions or curves in single points.

To cover the second item in the list we first make a projection of the separated ϵ-intersection $(A \cap_{\epsilon} B)_i$, $i = 1, \ldots, N_{\epsilon}$ onto set A and set B. Then we address the boundaries of these projections.

Definition 9 (Projected Separated ϵ-Intersection) *We define the projection of the disjoint subsets of the separated ϵ-intersection by*

$$\begin{aligned} (A \cap_{\epsilon} B)_i^A &= \left\{ \mathbf{p} \in A \mid (\mathbf{p}, \mathbf{q}) \in (A \cap_{\epsilon} B)_i \right\}, \quad i = 1, \ldots, N_{\epsilon} \\ (A \cap_{\epsilon} B)_i^B &= \left\{ \mathbf{q} \in B \mid (\mathbf{p}, \mathbf{q}) \in (A \cap_{\epsilon} B)_i \right\}, \quad i = 1, \ldots, N_{\epsilon}. \end{aligned}$$

Remark 2 *By the projection we loose the correspondence between the points in set A and B taking part in the intersection. The correspondence is in general not one to one, thus the storage and handling of such a correspondence is not a trivial task.*

Since $(A \cap_{\epsilon} B)_i$ are closed sets, the projected sets must be closed sets. Thus, the intersection can be represented by the boundary of these sets.

Definition 10 *The Boundary of the Projected Separated ϵ-Intersection is defined by*

$$\begin{aligned} \mathcal{B}\left(A \cap_{\epsilon} B\right)_i^A &= (A \cap_{\epsilon} B)_i^A - int((A \cap_{\epsilon} B)_i^A), \quad i = 1, \ldots, N_{\epsilon} \\ \mathcal{B}\left(A \cap_{\epsilon} B\right)_i^B &= (A \cap_{\epsilon} B)_i^B - int((A \cap_{\epsilon} B)_i^B), \quad i = 1, \ldots, N_{\epsilon}. \end{aligned}$$

Remark 3 *Here it is important to distinguish the boundary of the projected separated ϵ-intersection in $\mathbb{R}^l$ from the boundary of the manifolds describing each projected separated ϵ-intersection. E.g. a projected separated ϵ-intersection in $\mathbb{R}^3$ can consist of a complete circle. The boundary of the circle in $\mathbb{R}^3$ is the circle itself. However, the circle as a manifold has no boundary since no point on the circle is locally homeomorphic to $\mathbb{R}^+$.*

2.4   Finding all Intersection Occurrences

In definition 5 we introduced the disjoint separated ϵ-intersection. However, we did not discuss any method for singling out the disjoint subsets. The main strategy employed to solve such problems is divide and conquer. However, to make a successful divide and conquer strategy we have to solve the following problems:

- Decide if two objects have a guaranteed intersection, a possible intersection or no intersection. This is addressed in sections 2.4.1 and 2.4.3.

- Decide how to split an ϵ-intersection into possible disjoint subsets. This is addressed in Section 2.4.2 and Section 2.4.5.

- Identification of a disjoint subset in the separated ϵ-intersection. This is addressed in Section 2.4.4.

- To decide when a disjoint subset in the separated ϵ-intersection consists of only one bounded manifold, or of touching bounded manifolds of possibly different dimensions. This is crucial for the representation of the intersection, and is addressed in Section 2.4.6.

- To decide if all disjoint subsets in the separated ϵ-intersection set touch the boundary of one of the objects being intersected. This can reduce the problem to intersecting the boundaries of each of the sets with the other set. This is addressed in Section 2.5.

When we introduced the reduced separated ϵ-intersection in definition 8, and the boundary of the disjoint subsets of projected separated ϵ-intersection in definition 10, we did not address the internal topology of these sets. This can in the case of surface intersection consist of one or more points, curves and surfaces regions. A divide and conquer strategy can be employed when sorting out the topology of each disjoint subset of a projected separated ϵ-intersection. The divide and conquer strategy is in general successful since it is possible to position a subdivision border between the curves/regions. However, when curves and/or regions meet, the subdivision has to go through the points where the curve(s) and/or regions meet. The orientation of the subdivision border is crucial to the success of the divide and conquer strategy.

An other challenge is that between a situations where it is obvious that the ϵ-intersection can be split into disjoint subsets, and situations where the ϵ-intersection cannot be split, there are cases where a minor pertubation makes the subsets touch. Some algorithmic approaches detect in such cases a separation, while other algorithmic approaches detect no separation in

the same situations. The reason for such differences can be the nature of propagation of rounding errors in different algorithms.

The concepts of absolute and relative rounding errors, error propagation, algorithms minimizing rounding errors and the best representation formats with respect to rounding error, play a central role to minimize the size of the regions where decisions are difficult to make. The divide and conquer strategy is not efficient in such regions.

2.4.1   Spatial Separation with Respect to a Tolerance

The detection of just one intersection point between two sets establish the fact that the two sets intersect. Thus, strategies than can detect some intersection point are of great value in the intersection process. By intersecting the boundary of a compact set A with a compact set B, and the boundary of B with A, we detect a subset of the intersection. When we later on want to represent the intersection result, these boundary intersections would be natural to use as candidates in the intersection representation. The boundary points are also important since we can use a recursive subdivision strategy to isolate regions where the intersection consists of manifolds that touch the boundary of a closed and bounded smooth manifold, as described in Section 2.5.

A number of different techniques exists for deciding if two objects are spatially separated. Two main approaches exist:

- Bounding each of the objects by a simpler geometry and checking if these bounding geometries intersect.

- Separating the objects by a geometric object.

In general a coarse bounding geometry is faster to calculate than a fine bounding geometry. When comparing two bounding geometries the complexity of the bounding objects has significant influence on the performance of the test.

The most commonly used bounding geometries are:

- **Axis-Parallel Boxes**. In $\mathbb{R}$ these are intervals; in $\mathbb{R}^2$ axis parallel rectangles; in $\mathbb{R}^3$ axis parallel rectangular boxes; and in $\mathbb{R}^l$ l-dimensional axis parallel boxes.

  Axis parallel boxes are easy to calculate for NURBS represented geometries with positive weights, as the geometry is limited by the projection

of the vertices from the projective space to the affine space. Algebraically represented geometries are in the general case infinite. However, if they are represented in barycentric coordinates, the corners of the barycentric coordinate system describe a triangle in $\mathbb{R}^2$, a tetrahedral in $\mathbb{R}^3$, and a l-simplex in $\mathbb{R}^l$. By applying the convention that the part of algebraic geometry of interest is inside the l-simplex, the problem is reduced to checking:

- If two simplices are closer than the tolerance ϵ.

- If a simplex and a box is closer than the tolerance ϵ.

- If two axis parallel boxes are closer than the tolerance ϵ.

- **Boxes with Fixed Rotation**. Naturally the axis parallel boxes are well behaved for geometries that are axis parallel. However, if the geometry is non-axis parallel, then these boxes are a very coarse bounding geometry. Making boxes from all possible rotations of the coordinate main axis by 45 degrees is computationally inexpensive, and handles many of the non-axis parallel situations well.

- **Convex Hulls**. An alternative bounding geometry for a NURBS curve or NURBS surface with positive weights is to calculate a convex hull of the control polygon. The calculation requires more operations than calculating the bounding boxes, and the comparison of two convex hulls is more computational demanding than comparison of boxes.

We can use algebraically represented geometries to separate two manifolds of possibly different dimensions.

- **Separation by Algebraic Hypersurface.** Let two manifolds be A and B, and let an algebraically represented hypersurface be $q(\mathbf{x}) = 0$. If we have that:

$$\forall(\mathbf{p}_A, \mathbf{p}_B) \in A \times B : q(\mathbf{p}_A) \geq 0 \wedge q(\mathbf{p}_B) < 0,$$

then the objects are non-overlapping. Taking then an overlap within a tolerance $\epsilon_q > 0$ into consideration. We have to satisfy the equation

$$\forall(\mathbf{p}_A, \mathbf{p}_B) \in A \times B : q(\mathbf{p}_A) \geq \epsilon_q \wedge q(\mathbf{p}_B) \leq -\epsilon_q$$

to guarantee separation. The value of ϵ_q has to be calculated from the behavior of $q(\mathbf{x})$ in the region around set A and set B. The hypersurface can be a plane; another hypersurface; a hypersurface containing the

27

manifold A; a hypersurface approximating the manifold A. In Chapter 4 we enter into details of such approximative implicitization. In Section 4.11 we look at more details on using an algebraic surface for separating two manifolds.

In [Hohmeyer:92] the separation of objects was divided into spatial separability and spherical separability, the former consisting of bounding boxes, bounding planes, separating planes; the latter including spherical bounding boxes and separating circles. We see that all these concepts are included in the list above.

2.4.2 Subdividing ϵ-Intersections into Disjoint Subsets

Depending on the representation format of the manifolds, more or less numeric stable and efficient subdivision algorithms exist. For subdividing objects represented with the Bernstein basis, the de Casteljau algorithm is the obvious choice [Farin:92]. For subdividing B-spline represented objects the "Oslo algorithm" is to be preferred based on the numerical stability of the knot insertion scheme [Fuggeli:93].

For algebraically represented geometry, a de Casteljau algorithm can be used for subdividing, provided that a Bernstein basis in barycentric coordinates is used. The part of the algebraic surface of interest is supposed to be inside the simplex defining the barycentric coordinate system.

The main challenge in the subdivision process is to decide where to split the objects. The simplest strategies are:

- For a parametric (tensor product) NURBS represented geometry with internal knot(line)s, split at an internal knot(line).

- For a parametric rational (tensor product) Bernstein (B-spline with no internal knot(line)s) represented geometry, split at the midpoint.

- For a barycentric Bernstein basis represented geometry, split at the middle of the barycentric coordinate system. E.g. for a barycentric coordinate system in $\mathbb{R}^l$ split at the midpoint: The point with barycentric coordinates $(\frac{1}{l+1}, \ldots, \frac{1}{l+1})$.

- After reading these thesis Prof. T. W. Sederberg pointed out that a better strategy for subdividing a barycentric Bernstein basis represented triangle is to split the triangle into four triangles.

These strategies do not take into consideration the actual geometric problem to be solved. By doing this more optimal points for subdivision can be found.

The identification of a disjoint subset of a separated ϵ-intersection is a non-trivial task. The set can have a complicated geometric description. It can be necessary to break the intersection result into smaller pieces before we can identify a piece as a piece of a disjoint subset of a separated ϵ-intersection.

A strategy that can be employed is to subdivide the sets A and B into subsets having a "monotonous" behavior with respect to each other. Thus, we have to identify how to subdivide the sets to get subsets that are more "monotonous" with respect to each other than the original sets A and B. One approach is to subdivide sets A and B at local minimal points with respect to the distance function. Another approach is to subdivide the set A in its internal local maximal points with respect to B and the set B in its internal local maximal points with respect to A. These issues are addressed in respectively subsections 2.4.3 and 2.4.4 following.

2.4.3   Subdivision at Points with Minimal Distance

By subdividing at local minimum points of the distance function, we can get different effects.

- In theorem 14 and corollary 15 in Section 2.5.2 we develop a result that identifies intersection situations, where all objects in the intersection touch a boundary of one of the closed and bounded manifolds being intersected. The theorem allows singular intersection points at the boundaries of the manifolds being intersected. It thus supports subdivision in local closest points. Subdivision in these points is, provided the manifolds being intersected are smooth, a strategy for bringing singular points or near singular points to the boundary of submanifolds of the manifolds being intersected. By singular points in an intersection we mean points in the intersection with parallel normals. By near singular points we mean point pairs in the ϵ-intersection with parallel normals.

- By subdividing at local closest points that are separated more than the geometric tolerance ϵ, we can exclude regions of the manifolds that are further apart than ϵ.
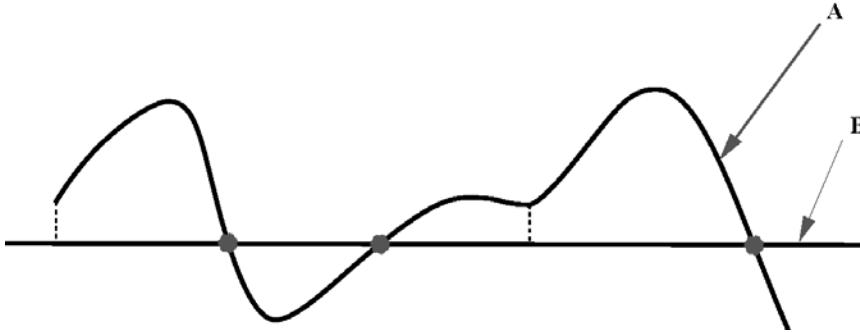
Figure 2.6: Example of local closest points between two curves. Closest points where the curves intersect are marked with a gray disk. Noncoincident local closest points are connected by a dashed line. Splitting of the curves at corresponding closest points can be used to exclude regions form the ϵ-intersection. Splitting at coincident closest points can bring intersection points to the boundary of submanifolds, then those can be analyzed to find the structure of the intersection set.

The local closest points are defined by the set $S_{\min}(\epsilon)$ where the tolerance ϵ is a tool for eliminating points lying inside the user defined tolerance.

$$
\begin{aligned}
S_{\min}(\epsilon) \quad = \quad & \{(\mathbf{p}, \mathbf{q}) \in A \times B \mid \exists \delta > 0 : \\
& \epsilon < \|\mathbf{p} - \mathbf{q}\|_2 \leq \min_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p},\delta) \wedge \mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q},\delta)} \|\mathbf{p}' - \mathbf{q}'\|_2\}.
\end{aligned}
$$

The open ball $\mathcal{B}(\mathbf{p}, r)$ is defined by $\mathcal{B}(\mathbf{p}, r) = \{\mathbf{y} \mid \|\mathbf{p} - \mathbf{y}\|_2 < r\}$.

2.4.4 Subdivision at Points with Maximal Distance

The Hausdorff distance $d(A, B)$ between two compact sets $A, B \subset \mathbb{R}^l$ defines the global maximal points between the two sets. The distance is in [Degen:92] defined by

$$
d(A, B) := \max(\max_{\mathbf{p} \in A}(\min_{\mathbf{q} \in B} \|\mathbf{p} - \mathbf{q}\|_2), \max_{\mathbf{q} \in B}(\min_{\mathbf{p} \in A} \|\mathbf{p} - \mathbf{q}\|_2)). \tag{2.1}
$$

We now intend to find points in the two sets that can be called local maximal points. The set of these points will be denoted $\mathcal{I}_{\max}(A, B)$. Looking at (2.1) we see that the outer maximum is over two expressions where the sets A and B change roles. We take the first of these,

$$
\max_{\mathbf{p} \in A}(\min_{\mathbf{q} \in B} \|\mathbf{p} - \mathbf{q}\|_2), \tag{2.2}
$$

and get a simpler expression. However, the roles of A and B are no longer symmetric. We are still finding the "max min" values over the entire sets A and B. To make the "max min" values local, we can limit the part of A and B used. A way of doing this is to use the following way for defining subsets:

- A subset of A defined by $A \cap \mathcal{B}(\mathbf{p}, \delta)$ where $\mathbf{p} \in A$ and $\delta > 0$.

- A subset of B defined by $B \cap \mathcal{B}(\mathbf{q}, \delta)$ where $\mathbf{q} \in B$ and $\delta > 0$.

Here $\mathcal{B}(\mathbf{x}, r)$ is an open ball defined by

$$\mathcal{B}(\mathbf{x}, r) = \left\{ \mathbf{y} \mid \|\mathbf{x} - \mathbf{y}\|_2 < r \right\}.$$

We require that an element in $\mathcal{I}_{\max}(A, B)$ is related to the expression

$$\max_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p}, \delta)} \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p}' - \mathbf{q}'\|_2 . \tag{2.3}$$

Since we use open balls it is easy to find combinations of $\mathbf{p}$, $\mathbf{q}$ and δ that make (2.3) define no value. However, when (2.3) defines a value we have found a local maximal point in set A with respect to B. The actual location is so far not identified. The next step is thus to link the local maximal points to $\mathbf{p}$ and $\mathbf{q}$. First we link $\mathbf{p}$ to the maximum in (2.3) by requiring

$$\min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p} - \mathbf{q}'\|_2 = \max_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p}, \delta)} \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p}' - \mathbf{q}'\|_2\}.$$

When (2.3) defines a value, the expression above must be true for some $\mathbf{p} \in A$. Now we link $\mathbf{q}$ to the maximum by requiring

$$\|\mathbf{p} - \mathbf{q}\|_2 = \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p} - \mathbf{q}'\|_2 = \max_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p}, \delta)} \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p}' - \mathbf{q}'\|_2\}.$$

When (2.3) defines a value, the expression above must be true for some $\mathbf{q} \in B$.

Thus, we have an expression that describes the properties of what we want to denote "Local maximal points of a set A with respect to a set B".

Definition 11 (Local Maximal Points.) *Let A and B be two compact sets. The set of Local Maximal Points of A with respect to B is defined by*

$$\mathcal{I}_{\max}(A, B) = \quad \{(\mathbf{p}, \mathbf{q}) \in A \times B \mid \exists \delta > 0 : \|\mathbf{p} - \mathbf{q}\|_2 =$$
$$\min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p} - \mathbf{q}'\|_2 = \max_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p}, \delta)} \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p}' - \mathbf{q}'\|_2\}.$$
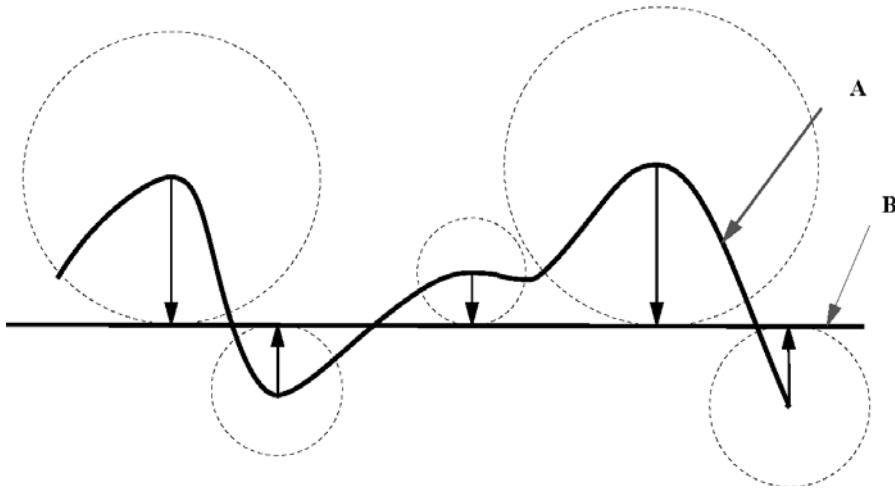
Figure 2.7: The local maximal points $I_{\max}(A, B)$ of set (curve) A with respect to set (curve) B. The arrows point at points of curve B. Note that only one end point of A is included. Let a point move along curve A from left to right. At the left end, where the point moves away from curve B, the endpoint is excluded. Then a point moves along curve A from right to left. At the right end, where the point move closer to curve B, the end point is included. The radius of curvature of curve B is infinite, and is thus greater than the distance to curve A. Hence for every point on curve A, only one point on curve B is included.
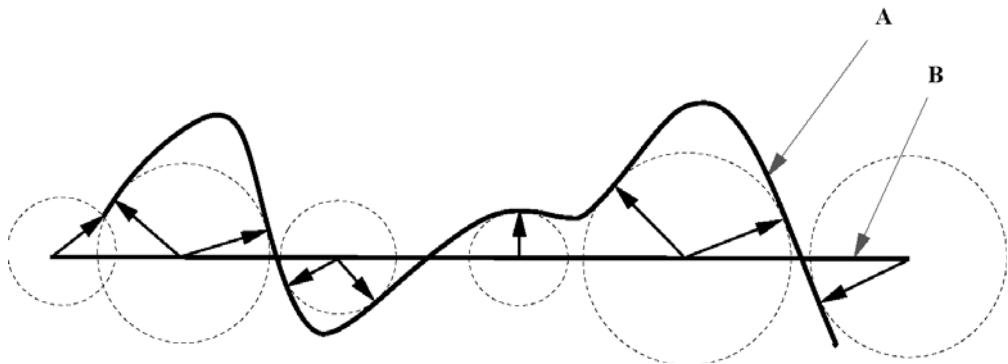
Figure 2.8: The local maximal points $I_{\max}(B, A)$ of set (curve) B with respect to set (curve) A. The arrows point at points of curve A. Note that both end points of B are included, one corresponding to an end of A the other corresponding to an internal point on A. Also note that to some of the points on B two points correspond on curve A.

32

Remark 4 *It is important to note than in general* $\mathcal{I}_{\max}(A, B) \neq \mathcal{I}_{\max}(B, A)$.

Remark 5 *If the set B is a smooth manifold, and $(\mathbf{p}, \mathbf{q}) \in \mathcal{I}_{\max}(A, B)$, then the vector $\mathbf{p} - \mathbf{q}$ is orthogonal to the tangent bundle of B at the point $\mathbf{q}$, provided $\mathbf{q}$ is an internal point in B. This is used in the examples in figure 2.7 and in figure 2.8.*

Example 12 *Let A have the shape of a $\cup$ and let B be a bar $^-$ above the $\cup$. The intersection problem looks like: $\bar{\cup}$. Now $\mathcal{I}_{\max}(A, B)$ contains the bottom of the $\cup$ paired with the middle point of the bar, while $\mathcal{I}_{\max}(B, A)$ contains the middle point of the bar paired with the end points of $\cup$.*

There are two possible applications for the local maximal points between the two compact sets.

- Let $(\mathbf{p}, \mathbf{q}) \in \mathcal{I}_{\max}(A, B)$, and let $\|\mathbf{p} - \mathbf{q}\|_2 \leq \epsilon$, where $\epsilon \geq 0$ is the intersection tolerance, then we can find an open ball $\mathcal{B}(\mathbf{p}, \delta)$ around $\mathbf{p}$ with radius δ and a ball $\mathcal{B}(\mathbf{q}, \delta)$ around $\mathbf{q}$ with radius δ such that

$$\max_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p}, \delta)} \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p}' - \mathbf{q}'\|_2 \leq \epsilon,$$

  i.e. we have detected a part of an ϵ-intersection.

- Let $(\mathbf{p}, \mathbf{q}) \in \mathcal{I}_{\max}(A, B)$, and let $\|\mathbf{p} - \mathbf{q}\|_2 > \epsilon$, where $\epsilon \geq 0$ is the intersection tolerance, then we can find an open ball $\mathcal{B}(\mathbf{p}, \delta)$ around $\mathbf{p}$ with radius δ and a ball $\mathcal{B}(\mathbf{q}, \delta)$ around $\mathbf{q}$ with radius δ such that

$$\max_{\mathbf{p}' \in A \cap \mathcal{B}(\mathbf{p}, \delta)} \min_{\mathbf{q}' \in B \cap \mathcal{B}(\mathbf{q}, \delta)} \|\mathbf{p}' - \mathbf{q}'\|_2 > \epsilon,$$

  i.e. we have detected a region that is not part of the ϵ-intersection. Subdividing the set A through $\mathbf{p}$ and the set B through $\mathbf{q}$ may divide the intersection result into two disjoint sets.

2.4.5 Orientation of the Subdivision Borders

Some of the candidate points for locating the subdivision borders can be on the boundary of a manifold. When subdividing in those points, we have to select subdivisions that produce subsets.

An aspect to take into consideration when subdividing objects is the orientation of the subdivision borders. Using the standard recursive subdivision

techniques the shape of the subdivision borders are hyperplanes in the parameter domain of the parametric geometries, or hyperplanes in $\mathbb{R}^l$ for the algebraically represented geometries. The actual orientation of these hyperplanes, plays a central role in making a successful subdivision. In the tensor product parametric case it is efficient, in some cases, to convert to a parametric barycentric representation. In the barycentric representation it is easier to make an optimal orientation of subdivision borders.

2.4.6 Separating Connected Intersection Regions

Now assume that we have a separated ϵ-intersection that consists of a set of connected intersection regions that cannot be described by a single manifold. We want to describe each intersection region by a bounded manifold of dimension g, $g > 0$, each region can have different values for g.

Thus, we search for boundary points in respectively $(A \cap B)_i^A$, $(A \cap B)_i^B$ and $\mathcal{R}_{\epsilon,i}$, $i = 1, \ldots, N_\epsilon$ that are not homeomorphic to $\mathbb{R}^{g+}$, $g \in \mathbb{N}$. We subdivide the problem in such points to make intersection objects that only contain manifold geometries.

Since we have introduced two reduced representations for the separated ϵ-intersection, we must expect that the intersection geometry and topology for the two alternatives in many cases are different.

2.5 Loop Elimination

In the previous section we proposed different strategies for deciding where to subdivide two objects being intersected to get smaller and possibly simpler intersection problems. We proposed to subdivide in singular or near singular points and in points that would give the subsets a more monotonic behavior with respect to each other. Now we analyze intersection (sub)problems. The intention is to find situations when all objects in the intersection result touch the boundary of one of the objects being intersected. When this is the case we find points on all intersection objects by just intersecting the boundary of each set with the other set. It is assumed in the rest of the section that the sets intersected are smooth bounded manifolds with a boundary.

The smoothness requirement is necessary since the relative orientation of the geometries is essential to the results presented. The section is structured as follows:

- The normal set, which is used to describe the orientation of a manifold, is addressed in Subsection 2.5.1.

- In Subsection 2.5.2, theorem 14, we establish conditions identifying when n smooth closed and bounded $(l-1)$-manifolds in $\mathbb{R}^l$ intersect in objects that touch the boundary of one of the $(l-1)$-manifolds. Then we look at uses of the theorem.

- In Subsection 2.5.3, we show that the boundaries of a manifold have to be split into pieces, to find the number of intersection objects along the boundary.

- Then, in Subsection 2.5.4, we find sufficient conditions for identifying nonsingular points in the intersection of l smooth $(l-1)$-manifolds in $\mathbb{R}^l$.

- The potential of combining algebraically represented hypersurfaces with a manifold is discussed in Subsection 2.5.5.

2.5.1   The Set of Normals

In order to discuss the relative orientation of manifolds, we have to be able to describe the orientation of each of the manifolds involved in the intersection. For curve segments in $\mathbb{R}^3$ the orientation at a point is related to the curve tangent at that point. The orientation of the curve is described by the possible tangent directions of the curve segment. The orientation of higher dimensional manifolds is in a similar way described by tangent bundles.

Let A be a smooth manifold, then to every point $\mathbf{p} \in A$, a tangent bundle $\mathbf{t_p}$, is associated. The tangent bundle is spanned by the first order partial derivatives of the parametrization of A at $\mathbf{p}$. However, we prefer to use normals of the manifold to describe the orientation and thus introduce the concept of the *Normal Set*.

Definition 13 (Normal Set) *The set of normals $\mathcal{N}(A)$ of A, a smooth manifold of dimension g in $\mathbb{R}^l$, is defined by*

$$\mathcal{N}(A) \;=\; \{\mathbf{q} \in \mathbb{R}^l \mid \|\mathbf{q}\|_2 = 1 \;\wedge\; \mathbf{q} \perp \mathbf{t_p}, \; \mathbf{p} \in A\}.$$

where $\mathbf{t_p}$ is the tangent bundle of A at $\mathbf{p}$.

What is the consequence of this definition? For a bounded l-manifold in $\mathbb{R}^l$, the normal set is empty. For a plane in $\mathbb{R}^3$ the normal set consists of the positive and negative unit normal vectors to the plane. For a sphere the normal set is the entire Gaussian Sphere.

2.5.2 Intersections Touching the Boundaries

In [Hohmeyer:92] a theorem was presented for singling out situations where the intersection between two surfaces in $\mathbb{R}^3$ has no internal closed curves. The theorem was a significant improvement over the previous work available. In [Hohmeyer:92] an extensive survey of previous work was given including [Dokken:90], [Kriezes:90-1], [Kriezes:90-2], [Kriezes:91], [Mountaudouin:89], [Sederberg:88], [Sederberg:89-1] and [Sinha:85]. Recent work can be found in [**?**], [Krishnan:96] and [**?**]. The following theorem addresses the situation when n manifolds in $\mathbb{R}^l$ of dimension $(l-1)$ are intersected. The theorem allows singular intersection points on the boundary of the manifolds being intersected. In [Hohmeyer:92] all intersection points on the boundaries had to be nonsingular.

Theorem 14 *Let $\mathcal{A}_n = \{A_i\}_{i=1}^n$ be a set of bounded smooth $(l-1)$-manifolds with boundary in $\mathbb{R}^l$, $l > n > 1$. If $\mathbf{n}_1, \dots, \mathbf{n}_n$ are linearly independent, and linearly independent from a vector $\mathbf{v}$ for any $\mathbf{n}_i \in \mathcal{N}(int(A_i)), i = 1, \dots, n$, then*

- *All r-manifolds $r \neq l-n$ and $r < l-1$ in the intersection between the manifolds in $\mathcal{A}_n$ are at the boundaries of one of A_i, $i = 1, \dots, n$.*

- *The intersection geometries in the internal of the intersection between the manifolds in $\mathcal{A}_n$ are $(l-n)$-manifolds that intersect the boundary of one of A_i, $i = 1, \dots, n$. I.e. the $(l-n)$-manifolds do not form internal loops.*

If $\mathbf{n}_1, \dots, \mathbf{n}_n$ are linearly independent, and linearly independent from a vector $\mathbf{v}$ for any $\mathbf{n}_i \in \mathcal{N}(A_i), i = 1, \dots, n$, then all intersections are r-manifolds $r \leq l-n$.

Proof. If the smooth $(l-1)$-manifolds intersect in a r-manifold, $r > l-n$ and the r-manifold is interior to all the smooth $(l-1)$-manifolds, then at each point of the intersection the normals of at least two of the $l-1$ manifolds have to be linearly dependent. This contradicts the assumption that $\mathbf{n}_1, \dots, \mathbf{n}_n$ are linearly independent for any $\mathbf{n}_i \in \mathcal{N}(int(A_i)), i = 1, \dots, n$.

Assume that a r-manifold $r < l-n$ is an intersection and is interior to all the smooth $l-1$ manifolds. In a neighborhood of an intersection point $\mathbf{p}$ all the $(l-1)$-manifolds A_i, $i = 1, \dots, n$ can be approximated with hyperplanes through the point $\mathbf{p}$ with normals taken from the respective $(l-1)$ manifolds. These hyperplanes intersect in a $(l-n)$-manifold since the normals are linearly independent. Since all $(l-1)$-manifolds are smooth, there is a small region

around $\mathbf{p}$ where the intersection is a $l - n$ manifold. Thus, the intersection cannot be an isolated r-manifold, $r \leq l - n - 1$.

Assume that the intersection consists of a bounded $(l - n)$-manifold not touching the boundaries of any A_i, $i = 1, \ldots, n$. The $(l - n)$-manifold is smooth since all A_i, $i = 1, \ldots, n$ are smooth and the normal sets are linearly independent. Now choose the direction vector $\mathbf{v}$ that is supposed to be linearly dependent from any $\mathbf{n}_i \in \mathcal{N}(A_i), i = 1, \ldots, n$. Since the $(l - n)$-manifold is smooth we can find a hyperplane that is normal to $\mathbf{v}$ and tangent to the $(l - n)$ manifold at some tangent point $\mathbf{p}_T$.

To analyze the tangent bundle at $\mathbf{p}_T$, we make a hyperplane $h_{\mathbf{v}}$ through the origin with normal vector $\mathbf{v}$. The tangent(bundle) at $\mathbf{p}_T$ has to lie in this hyperplane. Looking at which normals of A_i, $i = 1, \ldots, n$ that can make a tangent(bundle) in the hyperplane $h_{\mathbf{v}}$, we see that the tangent bundle is spanned by $(l - n)$ linearly independent vectors, the vectors normal to the tangent bundle have to lie in a linear n-manifold. We know that $\mathbf{v}$ is one vector in this n-manifold. The linear n-manifold can thus only interpolate $(n - 1)$ of the n-normal vectors, since the n normal vectors are linearly independent from $\mathbf{v}$. This implies that the tangent bundle cannot be defined from the normal vectors $\mathbf{n}_i \in \mathcal{N}(A_i), i = 1, \ldots, n$, and contradicts that $\mathbf{p}_T$ lies on the $(l - n)$-manifold in the intersection. Thus, it is impossible for the $(l - n)$-manifold to not touch the boundary.

If the normals set are linearly independent also on the boundaries of the manifolds, then all intersections of the manifolds are r-manifolds with $r \leq l - n$. ∎

Remark 6 *Since a g-manifold $g \leq l - 2$ can be described as the intersection of g manifolds of dimension $(l - 1)$, the theorem is also relevant for the intersection of manifolds with lower dimension than $l - 1$.*

Remark 7 *The use of theorem 14 for values of $n > 2$ requires efficient algorithms for determining the linear independence of $(l - n)$ normal sets, and to decide if there exists a direction $\mathbf{v}$ not spanned by the normal sets.*

In the case $n = 2$ the corollary following is dealing with the intersection of two $(l - 1)$-manifolds in $\mathbb{R}^l$. It states that if the normal sets in the interior of the manifolds can be separated by two planes, then all intersection curves touch the boundary of one of the manifolds. A remark then follows concerning the intersection of two surfaces (2-manifolds) in $\mathbb{R}^3$.

Corollary 15 *Let two hyperplanes in $\mathbb{R}^l$ through the origin be defined by the normal vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. Let A_1 and A_2 be bounded smooth $(l-1)$-manifolds*

with boundary in $\mathbb{R}^l$, $l > 2$. If the normal sets of $int(A_1)$ and $int(A_2)$ can be separated by two hyperplanes i.e.

$$\forall \mathbf{n}_1 \in \mathcal{N}(int(A_1)) \land \forall \mathbf{n}_2 \in \mathcal{N}(int(A_2)) :$$
$$(\mathbf{b}_1 \cdot \mathbf{n}_1)(\mathbf{b}_2 \cdot \mathbf{n}_1) > 0 \land (\mathbf{b}_1 \cdot \mathbf{n}_2)(\mathbf{b}_2 \cdot \mathbf{n}_2) < 0,$$

then

- All r-manifolds $r \leq l - 3$ in the intersection set between A_1 and A_2 are at the boundary of A_1 or A_2.

- The intersection geometries in the internal of A_1 and A_2 are $(l-2)$-manifolds that intersect the boundary of A_1 or A_2. E.g. the $(l-2)$-manifolds do not form internal loops.

Proof. Unit vectors lying in the intersection of the two hyperplanes are linearly independent of the normal sets. Thus, the requirements of theorem 14 are satisfied. ∎

Remark 8 *The theorem in [Hohmeyer:92] used two planes going through the origin to separate the normal sets of two surfaces in $\mathbb{R}^3$, and thus to assure that they are linearly independent, this is corollary 15 for $l = 3$ and $n = 2$. Corollary 15 is a more general result than the theorem in [Hohmeyer:92] since corollary 15 allows for parallel normal vectors, if they are at the boundary of a manifold. The distinction is important since corollary 15 gives the possibility for subdividing manifolds at points where the normals of the manifolds are parallel. Thus, the probability of creating subproblems of the intersection, where all objects in the intersection result touch the boundary, is increased.*

The next corollary addresses situations where we can find linear independence between a vector $\mathbf{v}$ and only a subset of the normals sets of the $(l-1)$-manifolds being intersected.

Corollary 16 *Let $\mathcal{A}_n = \{A_i\}_{i=1}^n$ be a set of closed and bounded smooth $(l-1)$-manifolds in $\mathbb{R}^l$, $l > 2$. Let $\mathcal{A}_{n,r} = \{A_{i(j)}\}_{j=1}^r$ be a subset of $\mathcal{A}_n$ satisfying: $\mathbf{n}_1, \ldots, \mathbf{n}_r$ are linearly independent, and linearly independent from a vector $\mathbf{v}$ for $\forall \mathbf{n}_j \in \mathcal{N}(int(A_{i(j)}))$, $j = 1, \ldots, r$, then the internal intersections lie on $(l-r)$-manifolds that touch the boundaries, and result from the intersection of the manifolds in $\mathcal{A}_{n,r}$.*
Proof. *The intersection of the manifolds in $\mathcal{A}_n$ is a subset of the intersection of the manifolds of $\mathcal{A}_{n,r}$. For $\mathcal{A}_{n,r}$ theorem 14 can be applied.* ∎

Remark 9 *The corollary can be used for identifying the possible location of intersection objects, when the linear independence of the normal vectors and the direction $\mathbf{v}$ is not established for the complete intersection problem, but can be established for a subset of the set of manifolds. Thus, regions where no intersection objects exist can be excluded by analyzing the intersection of these subsets.*

Example 17 *Let $\mathbf{p}(s,t)$, $(s,t) \in \Omega$ be a smooth surface with positive weights in $\mathbb{R}^3$, where Ω is a rectangle in $\mathbb{R}^2$, and let a straight line be described by the intersection of the two planes $q_1(\mathbf{x}) = 0$ and $q_2(\mathbf{x}) = 0$. Define three 2-manifolds in $\mathbb{R}^3$ by*

$$
\begin{aligned}
A_1 &= \{(s,t,q_1(\mathbf{p}(s,t))),\ (s,t) \in \Omega\} \\
A_2 &= \{(s,t,q_2(\mathbf{p}(s,t))),(s,t) \in \Omega\} \\
A_3 &= \{(s,t,0),(s,t) \in \Omega\}.
\end{aligned}
\tag{2.4}
$$

The intersection of the three manifolds A_1, A_2 and A_3 is equivalent to the intersection of the surface and the straight line. In theorem 23 in Subsection 2.5.4 conditions for when an intersection results in nonsingular points are given. If a possible singular situation arise, corollary 16 can be used for looking at the three intersection problems $A_1 \cap A_2$, $A_1 \cap A_3$ and $A_2 \cap A_3$ to establish the possible location of the intersection result.

In the following example we create an intersection between a straight line and a ruled surface resulting in a singular situation.

Example 18 *Let the surface $\mathbf{p}(s,t)$ be a ruled surface in $\mathbb{R}^3$ defined by the two curves $\mathbf{p}(s,0)$ and $\mathbf{p}(s,1)$ i.e.*

$$
\mathbf{p}(s,t) = (1-t)\mathbf{p}(s,0) + t\mathbf{p}(s,1), (s,t) \in [s_{\min}, s_{\max}] \times [0,1].
$$

Let $\mathbf{p}(c,t)$, $c \in [s_{\min}, s_{\max}]$ denote a straight line. Let $q_1(\mathbf{x}) = 0$ and $q_2(\mathbf{x}) = 0$ be the algebraic description of two noncoincident planes intersecting along $\mathbf{p}(c,t)$. The intersection problem can be formulated as (2.4) and is singular. Provided the two curves $\mathbf{p}(s,0)$ and $\mathbf{p}(s,1)$ do not intersect, the intersection curve $\mathbf{p}(c,t)$, $c \in [s_{\min}, s_{\max}]$ is nonsingular in at least two of the three intersection problems $A_1 \cap A_2$, $A_1 \cap A_3$ and $A_2 \cap A_3$.

In the following corollary we apply theorem 14 to a smooth function in g-variables, defined over a compact domain Ω, to establish conditions when the zeroes of the function lie on a $(g-1)$-manifold that touch the boundary of Ω.

Corollary 19 *Let $\Omega \in \mathbb{R}^g$, $g > 1$ denote a compact set, and let f denote a C^1 continuous function $f : \Omega \to \mathbb{R}$ satisfying for some $\mathbf{d} \in \mathbb{R}^g$*

$$\mathbf{d} \cdot (\frac{\partial f}{\partial s_1}, \ldots, \frac{\partial f}{\partial s_g}) > 0, \ (s_1, \ldots, s_g) \in int(\Omega). \qquad (2.5)$$

Then the set of zeroes of f

$$\mathcal{I}_\Omega = \{(s_1, \ldots, s_g) \in \Omega \mid f(s_1, \ldots, s_g) = 0\}$$

satisfies

- *All manifolds of dimension $p < g - 1$ in $\mathcal{I}_\Omega$ are at the boundary of Ω.*

- *The manifolds in $\mathcal{I}_\Omega$ in the internal of Ω have dimension $(g - 1)$, and intersect the boundary of Ω. I.e. the $(g - 1)$-manifolds do not form internal loops.*

Proof. We convert the problem to a $(g+1)$ dimensional problem by defining two g-manifolds for $(s_1, \ldots, s_g) \in \Omega$

$$\begin{aligned}
\mathbf{f}_0(s_1, \ldots, s_g) &= (s_1, \ldots, s_g, 0) \\
\mathbf{f}_1(s_1, \ldots, s_g) &= (s_1, \ldots, s_g, f(s_1, \ldots, s_g)).
\end{aligned}$$

The internal normal set of $\mathbf{f}_0$ and $\mathbf{f}_1$ are:

$$\begin{aligned}
\mathcal{N}(int(\mathbf{f}_0)) &= (\mathbf{0}, 1) \\
\mathcal{N}(int(\mathbf{f}_1)) &= \{(\nabla f(s_1, \ldots, s_l), 1) \mid (s_1, \ldots, s_l) \in int(\Omega)\}.
\end{aligned}$$

By assumption (2.5), vectors from the two normal sets, are linearly independent. Further vectors in the intersection of:

- The hyperplane normal to $(\mathbf{d}, 0)$ through the origin,

- The hyperplane normal to $(0, \ldots, 0, 1)$ through the origin,

cannot be made by a linear combination of the two normal sets. Thus, a vector $\mathbf{v}$ exists linearly independent from $\mathcal{N}(int(\mathbf{f}_0))$ and $\mathcal{N}(int(\mathbf{f}_1))$ and the conditions of theorem 14 are met. ∎

Now the question is the practical construction of the vector $\mathbf{d}$ in the corollary above.

Example 20 *Assume that the function f is described as a NURBS function. $\frac{\partial f(s,t)}{\partial s}$ and $\frac{\partial f(s,t)}{\partial t}$ are then NURBS functions. Thus, $(\frac{\partial f(s,t)}{\partial s}, \frac{\partial f(s,t)}{\partial t})$ is a convex combinations (provide positive weights) of a set of 2D vectors. Adding these, or their normalized version, gives a dominant direction of $(\frac{\partial f(s,t)}{\partial s}, \frac{\partial f(s,t)}{\partial t})$. In case the sum is zero, the vectors cannot be located at one side of a line through the origin, and the separation is impossible. Given the dominant direction we can calculate a "direction" cone of $(\frac{\partial f(s,t)}{\partial s}, \frac{\partial f(s,t)}{\partial t})$. If the opening of the direction cone is less than π, we have obtained the required separation.*

Example 21 *Let $\mathbf{p}(s,t)$, $(s,t) \in \Omega$ denote a NURBS surface with positive weights in $\mathbb{R}^3$, that is to be intersected with an algebraic surface $q(\mathbf{x}) = 0$. The function $q(\mathbf{p}(s,t))$ is of the type described in corollary 19. Corollary 19 can thus be used for identifying situations where all intersections of $\mathbf{p}(s,t)$, and the algebraic surface $q(\mathbf{x}) = 0$, in the internal of Ω, are curves (1-manifolds) that touch the boundary of Ω.*

In appendix B other problems where corollary 19 can be used are discussed.

2.5.3 Boundary Subdivision Necessary

In theorem 14 conditions were established for identifying situations when intersection results touch the boundary of one of the closed and bounded smooth manifold intersected. Thus, in such situations, we can look at the behavior of the boundary of a manifold, to describe the intersection results.

Theorem 22 *Let A be a closed and bounded smooth manifold with a smooth boundary denoted B. Then the Normal Set from (see definition 13) of the boundary B denoted, $\mathcal{N}(B)$ spans the entire Gaussian sphere.*

Proof. Assume that there is a direction $\mathbf{v} \notin \mathcal{N}(B)$, i.e. $\mathbf{v}$ is not in the normal set of the boundary B. Make a hyperplane with normal vector $\mathbf{v}$ that do not touch B. By translation of the hyperplane we can bring it to a position where the hyperplane just touch the boundary B. Since B is smooth the hyperplane is tangential to B and $\mathbf{v}$ must be in the $\mathcal{N}(B)$. Thus, contradicting that $\mathbf{v} \notin \mathcal{N}(B)$. ■

Remark 10 *Often the boundary of a closed and bounded smooth manifold is not smooth. By smoothing out regions on the boundary, where the tangent bundle has breaks, we get a manifold with a smooth boundary, and theorem 22 can be used. Thus, it is natural to define the normal set of the boundary*

*of a closed and bounded smooth manifold as the entire Gaussian Sphere. The
consequence is that the normal set contains no information on the orientation
of the boundary. For use in intersection algorithms, the boundary has to be
split into pieces, for the normal set to be of practical use.*

A tensor product surface has a rectangular parameter domain. Thus, a
first choice is to split the boundary into the four curves corresponding to the
straight lines defining the rectangular parameter domain. Triangular surfaces
have a triangular parameter domain. The primary choice is thus to subdivide
into three boundary curves, each corresponding to one of the straight lines
describing the triangular parameter domain.

2.5.4   Finding Nonsingular Intersection Points

Figure 2.9 illustrates a helix that, when intersected with certain axis parallel
straight lines, gives a number of intersection points. This happens although
the normal vectors of the helix, and the tangent direction of the straight
line are never orthogonal. Since the straight line can be described as the
intersection of two planar surfaces, the intersection can also be viewed as the
intersection of 3 surfaces in $\mathbb{R}^3$. In this section we look at the intersection of
l smooth manifolds of dimension $(l-1)$ in $\mathbb{R}^l$.

Theorem 23 *Give $\mathcal{A} = \{A_i\}_{i=1}^l$ a set of closed and bounded smooth $(l-1)$-
manifolds, with boundaries, in $\mathbb{R}^l$, $l > 1$. Let $\mathbf{n}_1, \ldots, \mathbf{n}_n$ be linearly inde-
pendent for any $\mathbf{n}_i \in \mathcal{N}(int(A_i))$, $i = 1, \ldots, n$. Let for $i = 1, \ldots, l$, $\mathcal{D}_i$ be
$\mathcal{A}$ with manifold A_i removed. Then*

- *All objects in the intersection between the manifolds in $\mathcal{A}$, not lying on
  the boundary of any of the manifolds in $\mathcal{A}$, are nonsingular points.*

- *The intersection points lie on 1-manifolds touching the boundary of one
  of the manifolds in $\mathcal{D}_i$, for $i = 1, \ldots, l$.*

Proof. For an internal intersection point to be singular, at least two man-
ifolds must have the same tangent plane. Since the normal sets are linearly
independent, this is impossible.

By removing one manifold theorem 14 is applicable, and the internal
intersections are 1-manifolds touching the boundary. ■

Now going back to figure 2.9, we see that the normal sets of the planar
surfaces intersecting in the straight line and the helix are linearly indepen-
dent. Thus, all the intersection points are nonsingular, and we can generate
curves on which the intersection points lie.
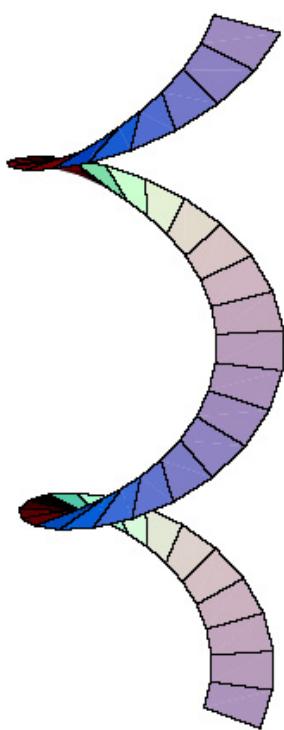
Figure 2.9: The intersection of a helix and a straight line parallel to the helix axis can produce a number of intersection points although the tangent of the straight line and the possible normals of the helix are never orthogonal.

2.5.5 Combining Parametric and Algebraic Representations

A common technique in intersection algorithms is to combine algebraic and parametric descriptions as described in Chapter 4. For certain geometric objects, e.g. the second degree algebraic curves and surfaces, we know both descriptions.

Let $\mathbf{h}(\mathbf{s})$ be a parametric represented manifold of dimension g in $\mathbb{R}^l$, $g < l$, and let $\mathbf{f}_i(\mathbf{t}_i)$, $i = 1, \ldots, n$ be manifolds in $\mathbb{R}^l$ that lie respectively in the hypersurfaces $f_i(\mathbf{x}) = 0$, $i = 1, \ldots, n$. Then instead of intersecting the manifolds $\mathbf{h}(\mathbf{s})$ and $\mathbf{f}_i(\mathbf{t}_i)$, $i = 1, \ldots, n$, we can look at the problem

$$f_1(\mathbf{h}(\mathbf{s})) = \ldots = f_n(\mathbf{h}(\mathbf{s})) = 0. \tag{2.6}$$

This problem can further be reformulated to the intersection of $(n+1)$ manifolds of dimension g in $\mathbb{R}^{g+1}$. These manifolds are

$$\begin{aligned} \mathbf{g}_0(\mathbf{s}) &= (\mathbf{s}, 0) \\ \mathbf{g}_i(\mathbf{s}) &= (\mathbf{s}, f_i(\mathbf{h}(\mathbf{s}))), \quad i = 1, \ldots, n. \end{aligned} \tag{2.7}$$

Since $f_i(\mathbf{f}_i) \equiv 0$, $i = 1, \ldots, n$, the solution of (2.7) contains the solutions of the original intersection problem in $\mathbb{R}^l$. The dimension of the reformulated problem (2.7) is smaller than that of the original intersection problem.

If we can find a sufficiently good approximation $\tilde{f}_i(\mathbf{x}) = 0$ to $f_i(\mathbf{x}) = 0$ close to $\mathbf{f}_i$ for $i = 1, \ldots, n$, then we have a method for finding approximative solutions to the original intersection problem. In Chapter 4 we look at such approximative implicitization.

Chapter 3

Representation of Geometric Objects

3.1   Parametric Representation of Closed and Bounded Manifolds.

Parametric representation of manifolds in $\mathbb{R}^2$ and $\mathbb{R}^3$ by piecewise polynomials is much used in CAGD-systems. It is well known that Bernstein and B-spline basis functions have a better numerical behavior, when used for such representation, than the power basis, see e.g. [Farouki:87]. In the approximative implicitization method presented in Section 4.3, stable and accurate products of powers of the coordinate functions of the manifolds are central. The approximative implicitization method is not limited to manifolds in $\mathbb{R}^2$ and $\mathbb{R}^3$, but addresses manifolds of dimension g in $\mathbb{R}^l$ with $1 \leq g < l$. The properties needed in the basis functions, used for the parametric description of the manifolds to be approximated, are assembled in definition 24. This definition includes the following basis functions:

- Tensor product Bernstein basis.

- Bernstein basis defined over a simplex.

- Tensor product B-spline basis.

- Possibly certain subclasses of linear independent Box-spline and simplex splines bases. However, we do not address these possibilities in this thesis. For more information on multivariate splines see e.g. [Seidel:92].

   When making products of the coordinate functions of a parametric represented manifold, we can without much loss of generality assume that the

coordinate functions are represented with the same basis functions. For our purpose it is convenient to let the description of any of these bases consist of:

- **A Description of the Polynomial Degree or Degrees** of the basis functions, in the following described by the multi index $\mathbf{n}$ or $\mathbf{m}$.

- **A Geometric Description of the Piecewise Polynomial Structure**, i.e. of how the domain of the basis functions is split into piecewise polynomial segments. We denote this description by the multi index $\mathbf{t}$.

- **A Description of the Continuity** between adjacent piecewise polynomial segments. This is denoted by the multi index $\mathbf{c}$.

We refer to this as the *PosProd* basis description. The PosProd basis functions are denoted

$$B_{\mathbf{i},\mathbf{n}}(\mathbf{s}) = B_{\mathbf{i},\mathbf{n},\mathbf{t}}^{\mathbf{c}}(\mathbf{s}),$$

here $\mathbf{i} \in \mathcal{I}_{\mathbf{n}} = \mathcal{I}_{\mathbf{i},\mathbf{n},\mathbf{t}}^{\mathbf{c}}$ is a set of multi indices for the basis functions. $\mathbf{s}$ is a vector with a dimension corresponding to the dimension of the parameter domain. We denote the number of multi indices $\tilde{N}_{\mathbf{n}}$. The following definition only use the form $B_{\mathbf{i},\mathbf{n}}(\mathbf{s})$, because $\mathbf{t}$ and $\mathbf{c}$ are constants for the multiplication operations of interest.

Definition 24 (PosProd Basis Function) *Let $S_{\mathbf{n},\mathbf{t}}^{\mathbf{c}}$ be a space of piecewise polynomials of degree $\mathbf{n}$ and continuity $\mathbf{c}$ with a geometric description $\mathbf{t}$ defined over the compact domain Ω. In addition let $S_{\mathbf{n},\mathbf{t}}^{\mathbf{c}}$ contain the polynomials of degree $\mathbf{n}$. A basis $\{B_{\mathbf{i},\mathbf{n}}\}_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}}$ for $S_{\mathbf{n},\mathbf{t}}^{\mathbf{c}}$ is called a **PosProd basis** if*

1. *It is a partition of unity and the basis functions are nonnegative*

$$\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} B_{\mathbf{i},\mathbf{n}}(\mathbf{s}) = 1, \ \mathbf{s} \in \Omega$$

   with

$$B_{\mathbf{i},\mathbf{n}}(\mathbf{s}) \geq 0, \ \mathbf{s} \in \Omega, \ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}.$$

2. *The product $h(s) \in S_{\mathbf{m}+\mathbf{n},\mathbf{t}}^{\mathbf{c}}$ of the two functions $p(\mathbf{s}) \in S_{\mathbf{n},\mathbf{t}}^{\mathbf{c}}$ and $q(\mathbf{s}) \in S_{\mathbf{m},\mathbf{t}}^{\mathbf{c}}$ with nonnegative coefficients has nonnegative coefficients. If*

$$\begin{aligned}
p(\mathbf{s}) &= \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} p_{\mathbf{i}} B_{\mathbf{i},\mathbf{n}}(\mathbf{s}), \ p_{\mathbf{i}} \geq 0, \quad \mathbf{i} \in \mathcal{I}_{\mathbf{n}} \\
q(\mathbf{s}) &= \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} q_{\mathbf{i}} B_{\mathbf{i},\mathbf{m}}(\mathbf{s}), \ q_{\mathbf{i}} \geq 0, \quad \mathbf{i} \in \mathcal{I}_{\mathbf{m}}.
\end{aligned}$$

Then

$$h(\mathbf{s}) = p(\mathbf{s})q(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m+n}}} h_{\mathbf{i}} B_{\mathbf{i,m+n}}(\mathbf{s})$$

with

$$h(s) \in S^{\mathbf{c}}_{\mathbf{m+n,t}},$$

and

$$h_{\mathbf{i}} \geq 0, \ \mathbf{i} \in \mathcal{I}_{\mathbf{m+n}}.$$

The name *PosProd* reflects that the product of two *PosProd* represented functions p and q with nonnegative cofficients results in a *PosProd* represented function h with with nonnegative cofficients.

Definition 25 (PosProd Algorithm) *Given the product of two positive functions, which both are represented by PosProd Basis functions. A **PosProd Algorithm** ensures that the relative rounding errors in the coefficients of the product are limited by the relative rounding errors of the functions multiplied. If $p_\epsilon(\mathbf{s})$ and $q_\epsilon(\mathbf{s})$ are respectively $p(\mathbf{s})$ and $q(\mathbf{s})$ from condition 2 in definition 24 with imposed relative rounding errors*

$$p_\epsilon(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} p_{\mathbf{i}}(1 + \epsilon_{\mathbf{i}}^p) B_{\mathbf{i,n}}(\mathbf{s}), \quad p_{\mathbf{i}} \geq 0, \ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}$$

$$q_\epsilon(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m}}} q_{\mathbf{i}}(1 + \epsilon_{\mathbf{i}}^q) B_{\mathbf{i,m}}(\mathbf{s}), \quad q_{\mathbf{i}} \geq 0, \ \mathbf{i} \in \mathcal{I}_{\mathbf{m}},$$

where

$$|\epsilon_{\mathbf{i}}^p| \leq \epsilon_{\max}^p \wedge \quad (p_i = 0 \implies \epsilon_i^p = 0), \quad \mathbf{i} \in \mathcal{I}_{\mathbf{n}}$$
$$|\epsilon_{\mathbf{i}}^q| \leq \epsilon_{\max}^q \wedge \quad (q_i = 0 \implies \epsilon_i^q = 0), \quad \mathbf{i} \in \mathcal{I}_{\mathbf{m}}.$$

Then the product of $p_\epsilon(\mathbf{s})$ and $q_\epsilon(\mathbf{s})$ must satisfy

$$h_\epsilon(\mathbf{s}) = p_\epsilon(\mathbf{s})q_\epsilon(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{m+n}}} h_{\mathbf{i}}(1 + \epsilon_{\mathbf{i}}^h) B_{\mathbf{i,n+m}}(\mathbf{s})$$

with

$$\left|\epsilon_{\mathbf{i}}^h\right| \leq \epsilon_{\max}^p + \epsilon_{\max}^q + \epsilon_{\max}^p \epsilon_{\max}^q, \ \mathbf{i} \in \mathcal{I}_{\mathbf{m+n}}.$$

Remark 11 *It should be observed that the description of the PosProd basis functions representing the product of two PosProd represented functions are constructed by keeping the geometry description $\mathbf{t}$ and the continuity description $\mathbf{c}$ fixed, and just adding the polynomial degrees $\mathbf{m}$ and $\mathbf{n}$.*

Example 26 (Univariate B-spline) *Let the PosProd description of the B-spline basis be:*

- *Description of polynomial degree n.*

- *Description of piecewise polynomial structure by strictly increasing knot values*

$$\mathbf{t} = \{t_i\}_{i=1}^{\eta} .$$

- *Description of continuity at each knot value*

$$\mathbf{c} = \{c_i\}_{i=1}^{\eta} = \{-1, c_2, \ldots, c_{\eta-1}, -1\}$$

with $0 \leq c_i \leq n-1$, $i = 2, \ldots, \eta - 1$.

Here η is the number of distinct knot values. The number of basis functions in a B-spline basis is (see Theorem IX.1 page 113 in [de Boor:78]) :

$$\text{"Polynomial order"} + \text{"Number of internal knots"} = n + 1 + \sum_{i=2}^{\eta-1} (n - c_i).$$

Products of m functions represented in this B-spline basis have the same distinct knot values and continuity. However, the polynomial degree is raised to mn. Thus, the number of basis functions of degree mn is

$$\tilde{N}_{mn} = mn + 1 + \sum_{i=2}^{\eta-1} (mn - c_i) = (\eta - 1)mn - \left(\sum_{i=2}^{\eta-1} c_i \right) + 1 \leq (\eta - 1)mn + 1.$$

In the case of a piecewise Bernstein basis representation the number of basis functions of degree mn is

$$\hat{N}_{mn} = (\eta - 1)(mn + 1).$$

The amount of basis functions save by using a B-spline representation instead of a Bernstein representation is

$$\hat{N}_{mn} - \tilde{N}_{mn} = (\eta - 1) + \left(\sum_{i=2}^{\eta-1} c_i \right) - 1 = \eta + \left(\sum_{i=2}^{\eta-1} c_i \right) - 2.$$

This number is independent of the number m of functions in the product . Thus, if a piecewise Bernstein basis representation is used instead of a B-spline representation the relative overhead with respect to the number of basis functions decrease as m increase.

Example 27 (Univariate B-spline) *Let the knot vector defining a B-spline basis be*

$$t = \{1, 1, 1, 1, 2, 2, 2, 3, 3, 4, 5, 5, 5, 5\}.$$

The PosProd basis description is

$$
\begin{aligned}
\mathbf{n} &= & 3 \\
\mathbf{t} &= & \{1, 2, 3, 4, 5\} \\
\mathbf{c} &= & \{-1, 0, 1, 2, -1\}.
\end{aligned}
$$

The product of two such 3rd degree bases yields the PosProd basis description

$$
\begin{aligned}
\mathbf{n} &= & 6 \\
\mathbf{t} &= & \{1, 2, 3, 4, 5\} \\
\mathbf{c} &= & \{-1, 0, 1, 2, -1\}.
\end{aligned}
$$

The product of the 3rd degree basis and 6th degree basis gives the PosProd basis description

$$
\begin{aligned}
\mathbf{n} &= & 9 \\
\mathbf{t} &= & \{1, 2, 3, 4, 5\} \\
\mathbf{c} &= & \{-1, 0, 1, 2, -1\}.
\end{aligned}
$$

Example 28 (Bi-variate B-spline basis.) *The polynomial degree is $\mathbf{n} = (k_1 - 1, k_2 - 1)$, where k_1 and k_2 are the polynomial orders in the two parameter directions of the tensor product B-spline basis. The description of the piecewise polynomial structure $\mathbf{t}$ is two lists of knot values, one for each parameter direction. The continuity $\mathbf{c}$ is two lists of numbers, one list for each knot value list. The lists give as, in the univariate case, the continuity at each knot value.*

Next we deal with a theorem showing that product algorithms based on convex combinations are PosProd Algorithms. Then two corollaries relating the result to the Bernstein basis and to the B-spline basis follow.

Theorem 29 *Let $p(s)$ and $q(s)$ be two functions represented with PosProd basis functions of respectively degree $\mathbf{m}$ and $\mathbf{n}$ with the same piecewise polynomial structure $\mathbf{t}$ and the same continuity description $\mathbf{c}$. Further let the product algorithm used for generating the coefficients $h_{\mathbf{r}} \in \mathcal{I}_{\mathbf{m+n}}$ of $h(s) = p(s)q(s)$ be based on sums of convex combinations of products of the coefficients of $p(s)$ and $q(s)$:*

$$h_{\mathbf{r}} = \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i,j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}}, \ \mathbf{r} \in \mathcal{I}_{\mathbf{m+n}},$$

49

where

$$\sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} = 1, \ \mathbf{r} \in \mathcal{I}_{\mathbf{m+n}}$$

with

$$\gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} \geq 0, \ \mathbf{r} \in \mathcal{I}_{\mathbf{m+n}}, \ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}, \ \mathbf{j} \in \mathcal{I}_{\mathbf{m}}.$$

Then the product algorithm is a PosProd Algorithm.

Proof. Let as in definition 25 $p_\epsilon(\mathbf{s})$ and $q_\epsilon(\mathbf{s})$ be respectively $p(\mathbf{s})$ and $q(\mathbf{s})$ with imposed rounding errors. Then the coefficients $h_{\mathbf{r}}^\epsilon$, $\mathbf{r} \in \mathcal{I}_{\mathbf{m+n}}$ of $h_\epsilon(\mathbf{s}) = p_\epsilon(\mathbf{s}) q_\epsilon(\mathbf{s})$ satisfy

$$\begin{aligned}
h_{\mathbf{r}}^\epsilon &= \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} \left(1 + \epsilon_{\mathbf{i}}^p\right) q_{\mathbf{j}} \left(1 + \epsilon_{\mathbf{j}}^q\right) \\
&= \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} + \sum_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} (\epsilon_{\mathbf{i}}^p + \epsilon_{\mathbf{j}}^q + \epsilon_{\mathbf{i}}^p \epsilon_{\mathbf{j}}^q) \\
&= h_{\mathbf{r}}(1 + \epsilon_{\mathbf{r}}^h),
\end{aligned}$$

where

$$\epsilon_{\mathbf{r}}^h = \begin{cases} \dfrac{\sum\limits_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum\limits_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} (\epsilon_{\mathbf{i}}^p + \epsilon_{\mathbf{j}}^q + \epsilon_{\mathbf{i}}^p \epsilon_{\mathbf{j}}^q)}{\sum\limits_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}} \sum\limits_{\mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}}} & \text{if} \ \ \exists \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} \neq 0 \\[4ex] 0 & \text{if} \ \ \forall \gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} = 0. \end{cases}$$

If there exist $\gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} \neq 0$ we have, since $\gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}}$, $p_{\mathbf{i}}$ and $q_{\mathbf{j}}$ all are nonnegative, that $\epsilon_{\mathbf{r}}^h$ is a convex combination of $(\epsilon_{\mathbf{i}}^p + \epsilon_{\mathbf{j}}^q + \epsilon_{\mathbf{i}}^p \epsilon_{\mathbf{j}}^q)$, $\mathbf{i} \in \mathcal{I}_{\mathbf{n}}$, $\mathbf{j} \in \mathcal{I}_{\mathbf{m}}$ thus

$$\left|\epsilon_{\mathbf{r}}^{\mathbf{h}}\right| \leq \max_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}, \mathbf{j} \in \mathcal{I}_{\mathbf{m}}} \left|\epsilon_{\mathbf{i}}^p + \epsilon_{\mathbf{j}}^q + \epsilon_{\mathbf{i}}^p \epsilon_{\mathbf{j}}^q\right| \leq \epsilon_{\max}^p + \epsilon_{\max}^q + \epsilon_{\max}^p \epsilon_{\max}^q,$$

which is the requirement in definition 25 of a product algorithm to be a PosProd Algorithm. If all $\gamma_{\mathbf{i},\mathbf{j}}^{\mathbf{r}} p_{\mathbf{i}} q_{\mathbf{j}} = 0$ then $\left|\epsilon_{\mathbf{r}}^h\right| = 0$, which also satisfies definition 25. ∎

Corollary 30 *In [de Boor:87-1] and [Farin:86] multiplication formulas are given for multiplication of functions represented in the tensor product Bernstein basis and for multiplication of functions represented in a Bernstein basis over a simplex. These multiplication formulas are based on sums of convex combinations of products of the coefficients of the functions being multiplied, and are thus by theorem 29 PosProd Algorithms.*

Error analysis of the B-spline basis is more difficult than the error analysis of the Bernstein bases. In [Cox:72] the first backwards error analysis on the B-spline recursion formulas was given. Here we show that there exist PosProd Algorithms for tensor product B-splines.

Corollary 31 *In [Moerken:91-1] a product algorithm for univariate B-spline represented functions is given. This product algorithm is based on sums of convex combinations of products of the coefficients of the functions being multiplied. The multiplication algorithm is thus by theorem 29 a PosProd Algorithm.*

For functions represented in a tensor product B-spline basis, the products can be performed one variable at a time. Thus, performing multiplication one variable at the time using the algorithm in [Moerken:91-1] for tensor product B-splines, gives us a PosProd Algorithm.

Remark 12 *The method for the multiplication of B-spline represented functions is so resource demanding, that a good alternative is to use a knot insertion algorithm to represent the function on a Bernstein knot vector. I.e. a knot vector where all knots have multiplicity corresponding to the polynomial order. Then the method for the multiplication of tensor product Bernstein basis represented functions can be used on each polynomial segment in the tensor product B-spline basis.*

The concept of PosProd Basis Functions can possibly also be applicable for certain subclasses of Box-splines and for simplex splines. However, for the PosProd basis concept to be of practical use, there have to be PosProd Algorithms for these types of multivariate splines. As far as the author of the thesis knows, no such algorithms have been developed.

More details on PosProd bases are given in Appendix A. We also consider rational functions and manifolds defined from PosProd basis functions.

Definition 32 (R-Positive Manifold) *Let l and g be integers with $g < l$. Given a compact set $\Omega \subset \mathbb{R}^g$ and a PosProd basis $B_{\mathbf{i},\mathbf{n}}(\mathbf{s})$, $\mathbf{i} \in \mathcal{I}_\mathbf{n}$ defined over Ω. A function $\mathbf{p} : \Omega \to \mathbb{R}^l$ which for some real numbers $c_{\mathbf{i}}^1, \ldots, c_{\mathbf{i}}^l$ and $c_{\mathbf{i}}^h$ can be written in the form*

$$\mathbf{p}(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{n}} \frac{(c_{\mathbf{i}}^1, \ldots, c_{\mathbf{i}}^l)}{c_{\mathbf{i}}^h} \alpha_{\mathbf{i},\mathbf{n}}(\mathbf{s}), \ \mathbf{s} \in \Omega \tag{3.1}$$

with

$$\alpha_{\mathbf{i},\mathbf{n}}(\mathbf{s}) = \frac{c_{\mathbf{i}}^h B_{\mathbf{i},\mathbf{n}}(\mathbf{s})}{\displaystyle\sum_{\mathbf{j} \in \mathcal{I}_\mathbf{n}} c_{\mathbf{j}}^h B_{\mathbf{j},\mathbf{n}}(\mathbf{s})}, \ \mathbf{i} \in \mathcal{I}_\mathbf{n},$$

and

$$c_{\mathbf{j}}^h > 0, \ \mathbf{j} \in \mathcal{I}_\mathbf{n},$$

*is called a **R-positive manifold in** $\mathbb{R}^l$ of degree $\mathbf{n}$. The numbers $c_{\mathbf{j}}^h$, $\mathbf{j} \in \mathcal{I}_\mathbf{n}$ are called the weights of the R-positive manifold.*

Remark 13 *The set of R-positive manifolds includes:*

- *Rational tensor product Bernstein represented manifolds with positive weights.*

- *Manifolds represented by a Rational Bernstein basis with positive weights defined over a simplex.*

- *Tensor product NURBS represented with positive weights.*

The nonrational versions of these are included in the definition, because setting $c_{\mathbf{j}}^{h} = 1$, $\mathbf{j} \in \mathcal{I}_{\mathbf{n}}$, makes the denominator equal to 1 and the manifolds nonrational.

3.2   The Algebraic Hypersurface

Suppose that $q \in P_m(\mathbb{R}^l)$, i.e. that q is a polynomial of total degree m in l variables. Then

$$\left\{ \mathbf{x} \in \mathbb{R}^l \mid q(\mathbf{x}) = 0 \right\} \tag{3.2}$$

is called an algebraic hypersurface.

Traditionally the polynomial q defining the algebraic hypersurfaces is represented in the power basis

$$q(\mathbf{x}) = \sum_{\substack{i_1 + \ldots + i_{l+1} = m \\ i_j \geq 0,\, j=1,\ldots,l+1}} a_{i_1,\ldots,i_l} \prod_{j=1}^{l} x_j^{i_j}.$$

However, describing the algebraic hypersurface using a simplex S in $\mathbb{R}^l$ and representing the polynomial in a Bernstein basis in barycentric coordinates defined over S, has desirable numerical properties in the algebraic implicitization process to be described. The ealiest us of this approach known to the author of this thesis can be found in [Sederberg:84-2]. The Bernstein basis in barycentric coordinates over a simplex S is described in appendix A.2. In this representation we have

$$q(\mathbf{x}) = \sum_{\mathbf{i} \in I(m)} b_{\mathbf{i}} \frac{m!}{\mathbf{i}!} \boldsymbol{\beta}^{\mathbf{i}}(\mathbf{x}) \tag{3.3}$$

with

$$\mathbf{i}! = i_1! \ldots i_{l+1}!$$

$$\boldsymbol{\beta}^{\mathbf{i}}(\mathbf{x}) = \prod_{j=1}^{l+1} \left(\beta_j(\mathbf{x}) \right)^{i_j},$$

and

$$I(m) = \{(i_1, \ldots, i_l, i_{l+1}) \in \mathbb{N}_0^{l+1} \mid 0 \leq i_j \leq m - \sum_{p=1}^{j-1} i_p, \; j = 1, \ldots, l+1\}.$$

Here $\beta_1(\mathbf{x}), \ldots, \beta_{l+1}(\mathbf{x})$ are barycentric coordinates with respect to S, and thus they satisfy

$$\sum_{i=1}^{l+1} \beta_i(\mathbf{x}) = 1$$

with

$$\beta_i(\mathbf{x}) \geq 0, \; \mathbf{x} \in S.$$

Chapter 4

Approximative Implicitization

In Sections 2.4.1 and 2.5.5 the advantage of combining algebraic and parametric representations of the manifolds to be intersected was discussed. However, for most parametric surfaces it is difficult to find the algebraic representation as discussed in Chapter 12 in [Hoschek:93]. The conclusions in this reference are:

- To find the exact algebraic representation of a 2D polynomial parametric curve can be done by building the Bezout resultant. The performance and accuracy of such an implicitization is good enough to be used in CAGD applications. The polynomial degree of the algebraic representation of a parametric polynomial curve is the same as the polynomial degree of the parametric representation.

- For parametric surfaces the situation is more complex. A triangular parametric surface of total degree n has an implicit formula of degree n^2. A parametric tensor product surface of degree (m, n) has an implicit formula of degree $2mn$. A bicubic surface patch thus results in an algebraic representation of degree 18 with 1330 terms. Thus, the computational complexity of finding algebraic representations for polynomial surfaces is high. In [Hoschek:93] references to more details on this topic are [Chionh:92] and [Sederberg:84-1].

In [Sederberg:95] the practical problems of exact implicitization of curves and surfaces are addressed. The main problem is that huge expressions are involved when performing exact implicitization of surfaces. Another problem is that in the case of base points, points where both numerator and denominator of a parametric curve or surface vanish simultaneously, the earlier standard methods fail. The method described in [Sederberg:95] is a major

step in the direction of more applicable exact implicitization methods. Instead of base points being a problem, these methods use the existence of base points to reduce the degree of the algebraic curve or surface found. For a survey of earlier methods [Hoffmann:93] can be consulted. Details on one of these methods, Groebner Bases, can be found in the book [**?**]. Details on resultant based methods can be found in the book [Cox:98]. In certain cases direct implicitization methods exist, see e.g. [Floater:95-2] where rational cubic implicitization is addressed.

The problem of finding the algebraic representation of a R-positive manifold of dimension $(l-1)$ in $\mathbb{R}^l$ can be formulated as follows.

Definition 33 (Exact Implicitization) *Let $l > 1$ and $\mathbf{p}(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^{(l-1)}$ be a R-positive $(l-1)$-manifold in $\mathbb{R}^l$ described as in definition 32. The process of exact implicitization of $\mathbf{p}(\mathbf{s})$ is to find a nontrivial polynomial $q \in P_m(\mathbb{R}^l)$ of minimal degree m such that*

$$q(\mathbf{p}(\mathbf{s})) = 0, \quad \mathbf{s} \in \mathbb{R}^{(l-1)}.$$

We restrict the definition to R-positive manifolds to avoid introducing additional concepts. All geometry representations that can be converted to a R-positive manifold of dimension $(l-1)$ are thus covered. In remark 13 in Section 3.1 we dealt with manifolds represented with a rational tensor product Bernstein basis or a rational Bernstein basis over a simplex. We stated that those have a R-positive description provided the weights are positive. Power basis represented manifolds can be converted to these descriptions and are thus covered provided the resulting weights are positive.

Exact implicitization is as stated in [Sederberg:95] and [Hoschek:93] a difficult problem. In addition if an exact implicit representation is found, the practical use of the implicit representation can be difficult, because:

- The polynomial degree is often high.

- Singular points near or on the implicit representation can give undesired properties if finite precision arithmetic is used.

Since CAGD systems are based on floating point representation, all represented geometries are affected by a small relative rounding error. Thus, when we find the "exact" algebraic representation, it is influenced by a rounding error.

Example 34 *Let a straight line segment be represented as a Bezier curve of degree 3. If the representation is exact, then the algebraic degree of the*

straight line is 1. *Represented in a float format on a computer small rounding errors are assigned to the vertices representing the straight line. Thus, the "straight line" probably has algebraic degree 3 when converted to infinite precision. Using methods that find the exact algebraic representation can thus result in an algebraic representation with a higher polynomial degree than necessary.*

In this chapter we address approximative implicitization of R-positive manifolds in $\mathbb{R}^l$. We allow for the approximation of a manifold of dimension g by hypersurfaces in $\mathbb{R}^l$ also for $g < l-1$. Thus, we are able to approximate a manifold of dimension g by the intersection of $l - g$ hypersurfaces.

Definition 35 (Approximative Implicitization) *Let l and g be integers with $1 \leq g < l$, and let $\mathbf{p}(\mathbf{s})$, $\mathbf{s} \in \Omega \subset \mathbb{R}^g$ be a R-positive manifold of dimension g in $\mathbb{R}^l$. The nontrivial algebraic hypersurface $q(\mathbf{x}) = 0$, $q \in P_m(\mathbb{R}^l)$, is an approximative implicitization of $\mathbf{p}(\mathbf{s})$ within the tolerance $\epsilon \geq 0$ if we can find a continuous direction function $\mathbf{g}(\mathbf{s})$ and a continuous error function $\eta(\mathbf{s})$ such that*

$$q(\mathbf{p}(\mathbf{s}) + \eta(\mathbf{s})\mathbf{g}(\mathbf{s})) = 0, \quad \mathbf{s} \in \Omega,$$

where

$$\|\mathbf{g}(\mathbf{s})\|_2 = 1, \quad \mathbf{s} \in \Omega,$$

and

$$|\eta(\mathbf{s})| \leq \epsilon, \quad \mathbf{s} \in \Omega.$$

$\mathbf{g}(\mathbf{s})$ *is called the direction for error measurement.*

Sometimes it is convenient, to simplify the use of the algebraic approximation $q(\mathbf{x}) = 0$, to require that the gradient of $q(\mathbf{x})$ is nonvanishing in a sufficiently large region around $\mathbf{p}(\mathbf{s})$, $\mathbf{s} \in \Omega$. This additional condition is, however, not part of the definition. If the gradient of $q(\mathbf{x})$ vanish on $\mathbf{p}(\mathbf{s})$ or too close to $\mathbf{p}(\mathbf{s})$, estimates for the error function $\eta(\mathbf{s})$, cannot be based on the gradient of $q(\mathbf{x})$.

An alternative to the approximative implicitization method described in this chapter, is as studied in [Sederberg:90], by minor modifications to insert base points in the parametric description of $\mathbf{p}(\mathbf{s})$. Thus, the degree of the algebraic description is reduced and the algorithms in [Sederberg:95] that take benefit from the existence of base points, can be used.

We focus in this chapter on building the theoretical foundation for an algorithm for approximative implicitization of $\mathbf{p}(\mathbf{s})$. The structure of the chapter is as follows:

- In Section 4.1 we show that the composition of an algebraic hyper-surface $q(\mathbf{x}) = 0$, with $q(\mathbf{x})$ described as in (3.3) and a R-positive g-manifold $\mathbf{p}(\mathbf{s})$, described as in definition 32, can be expressed

$$q(\mathbf{p}(\mathbf{s})) = (\mathbf{D}\mathbf{b})^T \boldsymbol{\alpha}(\mathbf{s}),$$

where $\mathbf{D}$ is a matrix, $\mathbf{b}$ contains the coefficients of q and $\boldsymbol{\alpha}(\mathbf{s})$ contains the basis functions used for representing products of the coordinate functions of $\mathbf{p}$. This means that if $\mathbf{b}$ is in the null space of $\mathbf{D}$, then $q(\mathbf{p}(\mathbf{s})) = 0$. Moreover

$$|q(\mathbf{p}(\mathbf{s}))| \leq \|\mathbf{D}\mathbf{b}\|_2 \,.$$

The matrix $\mathbf{D}$ has desirable numeric properties if the coefficients of $\mathbf{p}(\mathbf{s})$ are contained in a simplex S, and this simplex is used for the description of q in barycentric coordinates.

- In Section 4.2 we show that the relative rounding errors of the entries in $\mathbf{D}$ are well behaved, when PosProd Algorithms are used for building the matrix.

- In Section 4.3 we show that

$$\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s}\in\Omega} | q(\mathbf{p}(\mathbf{s})) | \leq \sigma_1.$$

Where $\sigma_1 \geq 0$ is the smallest singular value of $\mathbf{D}$ when barycentric coordinates are used for describing the algebraic hypersurface, and $\mathbf{p}(\mathbf{s})$ is contained in the simplex defining the barycentric coordinate system. This result shows that singular value decomposition can be used for finding an algebraic approximation of a R-positive manifold.

- In Section 4.4 we show how constraints can be added to the algebraic approximation to control the behavior of the approximation. Then in Subsection 4.4.5 we show that rows in the matrix $\mathbf{D}$ can be used as constraints in a direct elimination process as an alternative to singular value decomposition.

- In Section 4.5 we establish the convergence rate of the algebraic approximation and show that this is high especially for the approximation of parametric curves by algebraic curves (approximation of 1-manifolds in $\mathbb{R}^2$).

- In Section 4.6 we look at how to measure the error of an algebraic approximation to a manifold.

- In Section 4.7 we introduce the concept of admissable algebraic approximation, as an attempt to establish a tool for evaluating the quality of an algebraic approximation.

- In Section 4.8 we look at a method for choosing an approximation, with a possibly well behaved gradient, from the approximative nullspace of **D**.

- In Section 4.9 we give a more detailed version of the following algorithm:

Algorithm Approximative Implicitization

1. Choose the algebraic degree m of q.

2. Choose the barycentric coordinate system to be used in the approximation such that $\mathbf{p}(\Omega)$ is inside the simplex defining the barycentric coordinate system.

3. Build the matrix **D** of coefficients combining the characteristics of q and $\mathbf{p}$.

4. Find an approximative null space of **D**.

5. Choose as coefficients of q a vector from the approximative null space with a "good" gradient behavior.

6. Check if the hypersurface is admissible. If not admissable increase the number of vectors in the approximative null-space and go to 4.

7. Check if the hypersurface is within the prescribed tolerance ϵ. If not within the tolerance, try with a higher algebraic degree or perform the approximation on a smaller part of $\mathbf{p}$.

- In Section 4.10 we look at how a number of manifolds of possibly different dimensions can be approximated by one algebraic hypersurface.

- In Section 4.11 we describe how to find a hypersurface intended for spatially separating two manifolds. Then we look at how to estimate the distance between the manifolds by using such a hypersurface.

- The chapter is concluded in Section 4.12 by examples of approximative implicitization of a degree 7 parametric Bezier curve.

In sections 4.1 through 4.5 the results presented are based on the manifolds being R-positive. However, we only refer to these as manifolds or g-manifolds inside these sections to simplify the presentation.

4.1 Combining Parametric Manifolds and Algebraic Hypersurfaces

In this section we look at the properties of the matrix $\mathbf{D}$ mentioned in the introduction. It is assumed that the algebraic hypersurface is of total degree m and represented in barycentric coordinates. The combination $q(\mathbf{p}(\mathbf{s}))$ is a function of degree mn in the variable(s) $\mathbf{s}$.

The combination $q(\mathbf{p}(\mathbf{s}))$ can be used:

1. To build the matrix/vector structure used in the approximative implicitization process.

2. For transforming an intersection problem to a zero value problem as described in Section 2.5.5.

The results in this section and in the section following are useful to keep the relative rounding errors in $q(\mathbf{p}(\mathbf{s}))$ at a minimal level.

The first step is theorem 36 reformulating $q(\mathbf{p}(\mathbf{s}))$ to an expression using rational basis functions. Based on this, corollary 37 reformulates the sum to an expression of vectors and matrices

$$q(\mathbf{p}(\mathbf{s})) = (\mathbf{Db})^T \boldsymbol{\alpha}(\mathbf{s}).$$

The theorem then following limits $|q(\mathbf{p}(\mathbf{s}))|$ by $\|\mathbf{Db}\|_2$, thus getting rid of $\boldsymbol{\alpha}(\mathbf{s})$. The remainder of the section is devoted to analyze the numerical properties of the matrix $\mathbf{D}$.

Important steps in this section are:

- The isolation of the coefficients of the algebraic hypersurface in one vector $\mathbf{b}$.
$$\mathbf{b} = \left(\tilde{b}_j\right)_{j=1}^{\tilde{M}}.$$

  The entries in the vector $\mathbf{b}$ are related to the coefficients, $b_{\mathbf{j}}$, $\mathbf{j} \in I(m)$, of the algebraic hypersurface $q(\mathbf{x}) = 0$ by a lexicographical ordering ν

$$\tilde{b}_{\nu(\mathbf{j})} = b_{\mathbf{j}}, \ \mathbf{j} \in I(m).$$

  We also use the inverse ν^{-1} of the lexicographical ordering giving

$$b_{\nu^{-1}(j)} = \tilde{b}_j, \ j = 1, \ldots, \tilde{M}.$$

- The isolation of the variables used for the parametrization of the manifold in a vector $\boldsymbol{\alpha}(\mathbf{s})$, also denoted the α-basis.

$$\boldsymbol{\alpha}(\mathbf{s}) = (\alpha_i(\mathbf{s}))_{i=1}^{\tilde{N}_{m\mathbf{n}}}.$$

The entries in the vector $\boldsymbol{\alpha}(\mathbf{s})$ are related to the rational basis functions, $\alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s})$, $\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}$, of a manifold by a lexicographical ordering γ

$$\alpha_{\gamma(\mathbf{i})}(\mathbf{s}) = \alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s}), \ \mathbf{i} \in \mathcal{I}_{m\mathbf{n}}.$$

We also use the inverse γ^{-1} of the lexicographical ordering γ giving

$$\alpha_i(\mathbf{s}) = \alpha_{\gamma^{-1}(i),m\mathbf{n}}(\mathbf{s}), \ i = 1, \ldots, \tilde{N}_{m\mathbf{n}}.$$

- The isolation of expressions dependent of the coefficients of the manifold in a matrix $\mathbf{D}$. The entries in $\mathbf{D}$ are functions of the coefficients of the coordinate functions of $\mathbf{p}(\mathbf{s})$.

$$\mathbf{D} = \left(\tilde{d}_{i,j}\right)_{i,j=1,1}^{\tilde{N}_{m\mathbf{n}},\tilde{M}}. \tag{4.1}$$

We also use the lexicographical ordering ν and γ and their inverse for the ordering of the entries of the $\mathbf{D}$ matrix.

In the proof a transformation from the Bernstein basis to the α-basis plays a central role. Thus, the first step is to represent $q(\mathbf{p}(\mathbf{s}))$ in this basis.

Theorem 36 *Let* $\mathbf{p}(\mathbf{s})$ *be a R-positive manifold defined as in definition 32, and let* $q(\mathbf{x}) = 0$ *be a hypersurface described in barycentric coordinates as in (3.3). Then*

$$q(\mathbf{p}(\mathbf{s})) = \sum_{\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}} \left(\sum_{\mathbf{j} \in I(m)} d_{\mathbf{i}}^{\mathbf{j}} b_{\mathbf{j}}\right) \alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s}), \tag{4.2}$$

where $d_{\mathbf{i}}^{\mathbf{j}}$ *satisfy*

$$\frac{m!}{\mathbf{j}!} \beta^{\mathbf{j}}(\mathbf{p}(\mathbf{s})) = \sum_{\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}} d_{\mathbf{i}}^{\mathbf{j}} \alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s}), \ \mathbf{j} \in I(m).$$

Proof. We have from (3.3) that

$$q(\mathbf{p}(\mathbf{s})) = \sum_{\mathbf{j} \in I(m)} b_{\mathbf{j}} \frac{m!}{\mathbf{j}!} \beta^{\mathbf{j}}(\mathbf{p}(\mathbf{s})).$$

60

Remembering that $\boldsymbol{\beta}^{\mathbf{j}}(\mathbf{p}(\mathbf{s}))$ is a product of m of the coordinate functions of $\mathbf{p}(\mathbf{s})$, we have that the polynomial degree of $\boldsymbol{\beta}^{\mathbf{j}}(\mathbf{p}(\mathbf{s}))$ is $m\mathbf{n}$. Thus, $\frac{m!}{\mathbf{j}!}\boldsymbol{\beta}^{\mathbf{j}}(\mathbf{p}(\mathbf{s}))$ can be expressed in the α-basis of degree $m\mathbf{n}$. We denote these coefficients $d_{\mathbf{i}}^{\mathbf{j}}$ and get

$$\frac{m!}{\mathbf{j}!}\boldsymbol{\beta}^{\mathbf{j}}(\mathbf{p}(\mathbf{s})) = \sum_{\mathbf{i}\in\mathcal{I}_{m\mathbf{n}}} d_{\mathbf{i}}^{\mathbf{j}}\alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s}), \ \mathbf{j}\in I(m)$$

giving

$$q(\mathbf{p}(\mathbf{s})) = \sum_{\mathbf{j}\in I(m)} b_{\mathbf{j}} \sum_{\mathbf{i}\in\mathcal{I}_{m\mathbf{n}}} d_{\mathbf{i}}^{\mathbf{j}}\alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s}).$$

Now rearranging the sum we get (4.2). ∎

We are, at this stage, not addressing how this basis conversion is achieved only that it is possible. We show in Section 4.2 that using PosProd Algorithms for this transformation, limits the growth of relative rounding errors.

Corollary 37 *Let l and g be integers with $1 \le g < l$. Given $q(\mathbf{x}) = 0$, an algebraic hypersurface in $\mathbb{R}^l$ of degree m represented in barycentric coordinates as defined in (3.3), and $\mathbf{p}(\mathbf{s}) \in \mathbb{R}^l$, a g dimensional manifold as defined in definition 32. Then*

$$q(\mathbf{p}(\mathbf{s})) = (\mathbf{Db})^T \boldsymbol{\alpha}(\mathbf{s}). \tag{4.3}$$

Proof. From theorem 36 we have

$$\begin{aligned}
q(\mathbf{p}(\mathbf{s})) &= \sum_{\mathbf{i}\in\mathcal{I}_{m\mathbf{n}}} \left(\sum_{\mathbf{j}\in I(m)} d_{\mathbf{i}}^{\mathbf{j}} b_{\mathbf{j}} \right) \alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s}) \\
&= \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} \left(\sum_{j=1}^{\tilde{M}} d_{\gamma^{-1}(i)}^{\nu^{-1}(j)} b_{\nu^{-1}(j)} \right) \alpha_{\gamma^{-1}(i),m\mathbf{n}}(\mathbf{s}) \tag{4.4} \\
&= \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} (\sum_{j=1}^{\tilde{M}} \tilde{d}_{i,j}\, \tilde{b}_j)\alpha_i(\mathbf{s}) \\
&= (\mathbf{Db})^T \boldsymbol{\alpha}(\mathbf{s}),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{d}_{i,j} &= d_{\gamma^{-1}(i)}^{\nu^{-1}(j)} \\
\tilde{b}_j &= b_{\nu^{-1}(j)} \\
\alpha_i(\mathbf{s}) &= \alpha_{\gamma^{-1}(i),m\mathbf{n}}(\mathbf{s}).
\end{aligned}$$

∎

In the following theorem we utilize the fact that PosProd basis functions are the partition of unity, to remove the vector $\boldsymbol{\alpha}(\mathbf{s})$ containing the variable(s) $\mathbf{s}$ used in the parametrization of the manifold.

Theorem 38 *Let the algebraic hypersurface and the manifold be as in corollary 37 and definition 32. Let $\mathbf{s} \in \boldsymbol{\Omega}$, e.g. be in the support of the basis of the manifold. Then*

$$|q(\mathbf{p}(\mathbf{s}))| \leq \|\mathbf{Db}\|_2. \tag{4.5}$$

Proof. For any $\mathbf{s} \in \Omega$ we have by definition $\sum\limits_{i=1}^{\tilde{N}_{m\mathbf{n}}} \alpha_i(\mathbf{s}) = 1$ and $0 \leq \alpha_i(\mathbf{s}) \leq 1$, $i = 1, \ldots, \tilde{N}_{m\mathbf{n}}$. Thus, $\|\boldsymbol{\alpha}(\mathbf{s})\|_2 \leq 1$, giving

$$
\begin{aligned}
|q(\mathbf{p}(\mathbf{s}))| &= \left|(\mathbf{Db})^T \boldsymbol{\alpha}(\mathbf{s})\right| \\
&\leq \|\mathbf{Db}\|_2 \|\boldsymbol{\alpha}(\mathbf{s})\|_2 \leq \|\mathbf{Db}\|_2.
\end{aligned}
$$

$\blacksquare$

In the following theorem we show that if the algebraic hypersurface is defined by a barycentric Bernstein basis over a simplex, and that the simplex contains the manifold, then the matrix $\mathbf{D}$ has desirable numeric properties:

- All entries are none negative.

- The sum of all entries in a row is one.

- The Frobenius norm of $\mathbf{D}$ is limited by the number of rows.

Theorem 39 *Let $\mathbf{D}$ be defined as in (4.1). Then*

$$\sum_{j=1}^{\tilde{M}} \tilde{d}_{i,j} = 1, \quad i = 1, \ldots, \tilde{N}_{m\mathbf{n}}. \tag{4.6}$$

If the coefficients of the manifold is contained in the simplex defining the barycentric coordinate system in which $q(\mathbf{x})$ is described, then

$$\tilde{d}_{i,j} \geq 0, \quad i = 1, \ldots, \tilde{N}_{m\mathbf{n}}, \quad j = 1, \ldots, \tilde{M}, \tag{4.7}$$

and

$$\|\mathbf{D}\|_F^2 = \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} \sum_{j=1}^{\tilde{M}} \left(\tilde{d}_{i,j}\right)^2 = \sum_{j=1}^{\tilde{M}} \sigma_j^2 \leq \tilde{N}_{m\mathbf{n}}. \tag{4.8}$$

Here $\| \ \|_F$ is the Frobenius norm, and σ_i, $i = 1, \ldots, \tilde{M}$ are the singular values of $\mathbf{D}$.

Proof. Choose $\mathbf{b} = (1, \ldots, 1)^T$, then $q(\mathbf{x}) = 1$ giving

$$q(\mathbf{x}) = \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} \left(\sum_{j=1}^{\tilde{M}} \tilde{d}_{i,j} \right) \alpha_i(\mathbf{s}) = 1.$$

Since

$$\sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} \alpha_i(\mathbf{s}) = 1,$$

and $\alpha_i(\mathbf{s})$ are linearly independent, we have

$$\sum_{j=1}^{\tilde{M}} \tilde{d}_{i,j} = 1 \ .$$

Thus, (4.6) is proved.

Since the coefficients of the manifold are contained in the simplex defining the barycentric coordinate system, the coefficients of the barycentric representation of the manifold are all nonnegative. In addition, because the manifold is R-positive the weights of the manifold are by definition positive. Thus, all the elements in $\mathbf{D}$ are greater than or equal to zero, as all functions involved in the multiplication process making $\mathbf{D}$ have none negative coefficients, and we use a PosProd basis. Thus, proving (4.7).

Because $0 \leq \tilde{d}_{i,j} \leq 1$, we have that $0 \leq (\tilde{d}_{i,j})^2 \leq \tilde{d}_{i,j} \leq 1$, resulting in

$$\|\mathbf{D}\|_F^2 = \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} \sum_{j=1}^{\tilde{M}} \left(\tilde{d}_{i,j} \right)^2 \leq \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} \sum_{j=1}^{\tilde{M}} \tilde{d}_{i,j} = \sum_{i=1}^{\tilde{N}_{m\mathbf{n}}} 1 = \tilde{N}_{m\mathbf{n}}$$

proving (4.8). The connection between the Frobenius norm and the sum of the squares of the singular values can be found in standard books on linear algebra e.g. in [Stewart:73]. ■

In CAGD applications we are most often interested in 1-manifolds in $\mathbb{R}^2$, 1-manifolds in $\mathbb{R}^3$ or 2-manifolds in $\mathbb{R}^3$. The results in theorem 39 are independent of the dimension of the space in which the manifold lies. In the example following we assume that the conditions of theorem 39 are satisfied, and look at the size of $\tilde{N}_{m\mathbf{n}}$ for 1-manifolds represented by nonuniform rational B-splines (NURBS).

Example 40 *Let the PosProd description of the B-spline basis used for defining the 1-manifold be as in example 26 on page 47. Then*

$$\|\mathbf{D}\|_F^2 \leq (\eta - 1)mn + 1,$$

where η is the number of distinct knot values, m is the algebraic degree and n the polynomial degree of the B-spline basis. In the case of a Bernstein basis $\eta = 2$ and

$$\|\mathbf{D}\|_F^2 \leq mn + 1.$$

4.2 Stable Building of the D Matrix

When the matrix $\mathbf{D}$, defined in (4.1), is constructed from a R-positive g-manifold $\mathbf{p}(\mathbf{s})$, the entries result from products of powers of the coefficients of the coordinate functions of $\mathbf{p}(\mathbf{s})$. The number of coordinate functions involved in such a product is the total degree m of the algebraic hypersurface $q(\mathbf{x}) = 0$. In this section we limit the relative rounding error of these products

$$(\mathbf{p}(\mathbf{s}))^{\mathbf{j}} = \prod_{i=1}^{l+1} (p_i(\mathbf{s}))^{j_i},$$

where $\mathbf{j} = (j_1, \ldots, j_{l+1})$ satisfying

$$\sum_{i=1}^{l+1} j_i = m \geq 1,$$

and

$$j_i \geq 0, \ i = 1, \ldots, l+1.$$

We require that PosProd Algorithms are used when building the $\mathbf{D}$ matrix.

Theorem 41 *Let $p_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, g$ be functions represented with the same PosProd basis functions with nonnegative coefficients. Let $p_{\epsilon,i}(\mathbf{s})$ $i = 1, \ldots, g$ be respectively $p_i(\mathbf{s})$, $i = 1, \ldots, g$ with imposed relative rounding errors, and let the relative rounding errors of the coefficients satisfy*

$$\mid \epsilon_{\mathbf{i}}^{p_i} \mid \leq \epsilon_{\max}, \ \mathbf{i} \in \mathcal{I}_{\mathbf{n}}, \ i = 1, \ldots, g.$$

Given $j_i \in \mathbb{N}_0$, $i = 1, \ldots, l+1$ that satisfy

$$m = \sum_{i=1}^{l+1} j_i. \tag{4.9}$$

Then provided a PosProd Algorithm is used for evaluating the coefficients $h_{\mathbf{i}}^{\mathbf{j}}$, $\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}$ of

$$(\mathbf{p}(\mathbf{s}))^{\mathbf{j}} = \sum_{\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}} h_{\mathbf{i}}^{\mathbf{j}} \alpha_{\mathbf{i},m\mathbf{n}}(\mathbf{s})$$

the relative rounding errors $\epsilon_{\mathbf{i}}^{(\mathbf{p^j})}$, $\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}$ of the coefficients $h_{\mathbf{i}}^{\mathbf{j}}$, $\mathbf{i} \in \mathcal{I}_{m\mathbf{n}}$ satisfy

$$\left| \epsilon_{\mathbf{i}}^{(\mathbf{p^j})} \right| \leq \sum_{i=1}^{m} \binom{m}{i} \epsilon_{\max}^{i}. \tag{4.10}$$

Proof. For $m = 1$, only one of j_i, $i = 1, \ldots, l+1$ are different from zero. Assume that this is j_t. By (4.9) $j_t = 1$. Now $\epsilon_{\mathbf{i}}^{(\mathbf{p^j})}$ reduce to

$$\left| \epsilon_{\mathbf{i}}^{(\mathbf{p^j})} \right| = |\epsilon_{\mathbf{i}}^{p_t}| \leq \epsilon_{max},$$

which is (4.10) for $m = 1$.

Assume (4.10) is true for $m = k = \sum_{i=1}^{l+1} j_i$. Now multiplying by p_t, $1 \leq t \leq l+1$, using the properties of a PosProd Algorithms on page 47 we get

$$\begin{aligned}
\left| \epsilon_{i}^{p_t(\mathbf{p^j})} \right| & \leq & \epsilon_{\max}^{p_t} + \epsilon_{\max}^{(\mathbf{p^j})} + \epsilon_{\max}^{p_t}\epsilon_{\max}^{(\mathbf{p^j})} \\
& \leq & \epsilon_{\max} + \sum_{i=1}^{k} \binom{k}{i} \epsilon_{\max}^{i} + \epsilon_{\max} \sum_{i=1}^{k} \binom{k}{i} \epsilon_{\max}^{i} \\
& \leq & \sum_{i=1}^{k} \binom{k}{i} \epsilon_{\max}^{i} + \sum_{i=1}^{k+1} \binom{k}{i-1} \epsilon_{\max}^{i} \\
& \leq & \sum_{i=1}^{k} \left(\binom{k}{i} + \binom{k}{i-1} \right) \epsilon_{\max}^{i} + \epsilon_{\max}^{k+1} \\
& \leq & \sum_{i=1}^{k+1} \binom{k+1}{i} \epsilon_{\max}^{i}.
\end{aligned}$$

Which is (4.10) for $m = k + 1$. ∎

Remark 14 *Let the coefficients of a R-positive manifold $\mathbf{p}(\mathbf{s}) \in \mathbb{R}^l$ be contained in a simplex S used for the definition of a hypersurface in barycentric coordinates. Since the weights of a R-positive manifold are required to be positive the barycentric representation of $\mathbf{p}(\mathbf{s})$ has only non-negative coordinate functions. Thus, theorem 41 can be used for limiting the growth of the rounding errors of $\mathbf{D}$. The consequence of the theorem is that the relative rounding error is growing with a rate related to the algebraic degree m, since we can expect that $m\epsilon_{\max} \ll 1$, and thus $m\epsilon_{\max}$ is the dominant term in the sum $\sum_{i=1}^{m} \binom{m}{i} \epsilon^i$. This reasonably stable combination of algebraic and parametric surfaces can be achieved by:*

Algebraic degree m	Relative rounding error	Comment
2	2.0×10^{-15}	Conic sections
3	3.0×10^{-15}	Algebraic degree of cubic parametric curve
18	1.8×10^{-14}	Algebraic degree of bicubic parametric surface
100	1.0×10^{-13}	
m	$m \times 10^{-15}$	

Table 4.1: Relative rounding errors for different algebraic degrees of the $\mathbf{D}$ matrix when the R-positive manifold $\mathbf{p}(\mathbf{s})$ is defined in a barycentric coordinate system containing the coefficients of $\mathbf{p}(\mathbf{s})$. The coefficients of $\mathbf{p}(\mathbf{s})$ are assumed to have relative rounding errors limited by 10^{-15}.

- *Finding a simplex S containing the coefficients of a manifold.*

- *Represent the manifold in the barycentric coordinates defined by S.*

- *Represent the algebraic hypersurface by Bernstein basis functions defined over S.*

Example 42 *Assume that we have represented the R-positive manifold in barycentric coordinates such that the coefficients are inside a simplex S. Using double precision the value 10^{-15} is typically of the relative rounding errors. Now building the matrix $\mathbf{D}$ for different algebraic degrees, we get table 4.1 as a limit for the relative rounding errors of the elements in $\mathbf{D}$.*

4.3 Implicitization by Singular Values

How can the results from Section 4.1 be used when approximating a manifold in $\mathbb{R}^l$? In (4.5) we established a relationship

$$|q(\mathbf{p}(\mathbf{s}))| \leq \|\mathbf{D}\mathbf{b}\|_2 .$$

We now relate this expression to the singular values of $\mathbf{D}$.

Theorem 43 *Let as in (4.3) $\mathbf{p}(\mathbf{s})$ be a R-positive manifold in $\mathbb{R}^l$. Let $q(\mathbf{x}) = 0$, be an algebraic hypersurface in $\mathbb{R}^l$ with coefficient vector $\mathbf{b}$. Then*

$$\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s}\in\Omega} | q(\mathbf{p}(\mathbf{s})) | \leq \sigma_1. \tag{4.11}$$

Where $\sigma_1 \geq 0$ is the smallest singular value of $\mathbf{D}$, where $\mathbf{D}$ is defined in (4.1).

Proof. From (4.5) we have $|q(\mathbf{p}(\mathbf{s}))| \leq \|\mathbf{Db}\|_2$ when $\mathbf{s} \in \Omega$. Thus, we can reformulate to

$$\max_{\mathbf{s} \in \Omega} |q(\mathbf{p}(\mathbf{s}))|^2 \leq \mathbf{b}^T \mathbf{D}^T \mathbf{Db}.$$

Now restricting the coefficient vector to have length 1, e.g. $\|\mathbf{b}\|_2 = 1$, and taking the minimum we get

$$\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s} \in \Omega} |\, q(\mathbf{p}(\mathbf{s}))\,|^2 \leq \min_{\|\mathbf{b}\|_2 \neq 0} \frac{\mathbf{b}^T \mathbf{D}^T \mathbf{Db}}{\mathbf{b}^T \mathbf{b}} = \lambda_{\min}. \qquad (4.12)$$

Where $\lambda_{\min} \geq 0$ is the smallest eigenvalue of $\mathbf{D}^T \mathbf{D}$. The last equal sign is based on the fact that $\mathbf{D}^T \mathbf{D}$ is a Hermitian matrix (symmetric matrix) and that for a Hermitian matrix $\mathbf{A}$ the Raleigh quotient $\frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}}$ satisfies

$$\lambda_{\min} \leq \frac{\mathbf{x}^T \mathbf{Ax}}{\mathbf{x}^T \mathbf{x}} \leq \lambda_{\max},$$

where $\lambda_{\min}$ is the smallest and $\lambda_{\max}$ the largest eigenvalue of $\mathbf{A}$. The smallest singular value of $\mathbf{D}$ is $\sigma_1 = \sqrt{\lambda_{\min}}$, giving (4.11). ∎

Remark 15 *It is important to observe the number of singular values in $\mathbf{D}$ that are identical zero. If the number exceeds 1, then it is possible that we have used a higher algebraic degree than necessary in the approximation. This also is an indication that the parametric description of the manifold contains base points.*

In the general case when $\mathbf{p}(\mathbf{s}) \in \mathbb{R}^l$ is a R-positive manifold, it is impossible to get an exact algebraic representation as a single algebraic equation. This because the R-positive manifold, if it consists of more than one polynomial segment, in most cases has some discontinuity of some derivative at some point. In this case we see from (4.11) that the eigenvector corresponding to the smallest eigenvalue, is a solution of a least squares approximation of an algebraic hypersurface to $\mathbf{p}(\mathbf{s})$. Although we find a least squares approximation, there is no guarantee that we find the best solution. E.g. the choice of barycentric coordinate system influences $\mathbf{D}$. In addition the behavior of the gradient of the algebraic hypersurface determines how accurate the approximation is. How to find an approximation with a well behaved gradient is addressed in Section 4.8.

4.4 Constraining the Algebraic Equation

To better integrate approximations of R-positive g-manifolds into geometric modeling systems, we introduce in this section different ways to constrain the algebraic approximation to adhere to different sorts of constraints. We use manifolds of dimension l to define the constraints. The dimension of the manifolds being applied as interpolation constraints, is in most cases less than the dimension of the manifold being approximated. The constraints are:

- Interpolation of a manifold is addressed in subsection 4.4.1. Often it is desirable that the approximation interpolates given points. For 1-manifolds (curves) the points are often the start and end points, for 2-manifolds (surface) the points are often break points on the boundary of the 2-manifold. For 2-manifolds the constraining manifolds can also be 1-manifolds (curves). Typical curves applied are the boundary curves of the 2-manifold. However, curves internal to the manifold or outside the 2-manifold can also be used. Internal curves can be used to ensure exact fit to the boundary of another 2-manifold, while external curves can be used to control the shape of the algebraic approximation outside the manifold being approximated.

- Forcing the gradient of the polynomial $q(\mathbf{x})$ used for describing the hypersurface to be orthonogal to a tangent direction varying along a manifold is addressed in subsection 4.4.2. In combination with an interpolation constraint this can be used to induce interpolation of a manifold and tangent direction varying along the manifold at the same time. If the tangent constraint is not accompanied by a position constraint, then we model the behavior of the gradient of $q(\mathbf{x})$ in a region defined by a manifold. If the manifold being approximated is a 1-manifold the typical constraints are tangent at start and end points. If the manifold being approximated is a 2-manifold, then the constraints are typically the two tangent directions at break points in the boundary of the 2-manifold or the tangent behavior along the boundary of the 2-manifold.

- Forcing the gradient of $q(\mathbf{x})$ to take on a specific behavior along a manifold is addressed in subsection 4.4.3. This can be used to induce singular regions in the approximation, either inside or outside the manifold approximated.

In subsection 4.4.4 we consider how the constraints can be organized in a linear equation system, which can be applied for modifying the matrix vec-

tor expression $\|\mathbf{Db}\|_2$ describing the unconstrained approximation problem. Choosing a selection of rows from the $\mathbf{D}$ matrix as constraints is addressed in subsection 4.4.5. These can either be the most dominant rows, or rows that induce interpolation of position, tangent or higher order derivatives along the boundary of the manifold. This approach can also be used for a direct search for an approximative null space of $\mathbf{D}$.

4.4.1 Interpolation of Position along a Manifold

To impose interpolation of the position over a constraining manifold $\mathbf{r}(\mathbf{t})$, we can use the expressions from equation (4.3) giving

$$q(\mathbf{r}(\mathbf{t})) = (\mathbf{D_k}\mathbf{b})^T \boldsymbol{\alpha}_{\mathbf{r}}(\mathbf{t}) = 0.$$

Remembering that the basis functions in $\boldsymbol{\alpha}_{\mathbf{r}}(\mathbf{t})$ are linearly independent we can reformulate the expression, giving

$$\mathbf{D_k}\mathbf{b} = \mathbf{0}. \tag{4.13}$$

4.4.2 Interpolation of Tangent along a Manifold

The normal of the algebraic hypersurfaces is, provided the gradient is nonvanishing, parallel with the gradient of the hypersurface. Thus, to ensure that the normal is orthogonal to a given direction $\mathbf{d_r}(\mathbf{t})$ along the manifold $\mathbf{r}(\mathbf{t})$ we establish

$$\nabla q(\mathbf{r}(\mathbf{t})) \cdot \mathbf{d_r}(\mathbf{t}) = 0.$$

Provided both $\mathbf{r}(\mathbf{t})$ and $\mathbf{d_r}(\mathbf{t})$ can be expressed with the same PosProd basis functions, we rearrange this expression on the form

$$\nabla q(\mathbf{r}(\mathbf{t})) \cdot \mathbf{d_r}(\mathbf{t}) = \left(\mathbf{D'_k}\mathbf{b}\right)^T \tilde{\boldsymbol{\alpha}}_{\mathbf{r}}(\mathbf{t}) = 0.$$

Further remembering that the basis functions in $\tilde{\boldsymbol{\alpha}}_{\mathbf{r}}(\mathbf{t})$ are linearly independent, we get the constraints

$$\mathbf{D'_k}\mathbf{b} = \mathbf{0}. \tag{4.14}$$

4.4.3 Interpolation of Normal along a Manifold

Another possibility is to impose that the gradient takes on certain behavior $\mathbf{n_r}(\mathbf{t})$ along a given manifold $\mathbf{r}(\mathbf{t})$

$$\nabla q(\mathbf{r}(\mathbf{t})) = \mathbf{n_r}(\mathbf{t}) = \begin{pmatrix} n_1(\mathbf{t}) \\ \vdots \\ n_l(\mathbf{t}) \end{pmatrix},$$

or specifying each coordinate function separately

$$\frac{\partial q(\mathbf{r}(\mathbf{t}))}{\partial x_i} = n_i(\mathbf{t}), \ \ i = 1, \ldots, l.$$

Provided $\mathbf{r}(\mathbf{t})$ and $\mathbf{n_r}(\mathbf{t})$ can be express by the same PosProd basis functions, we get

$$\frac{\partial q(\mathbf{r}(\mathbf{t}))}{\partial x_i} = (\mathbf{D_{n},}_i \mathbf{b})^T \, \tilde{\boldsymbol{\alpha}}_\mathbf{r}(\mathbf{t}) = \mathbf{n}_i \cdot \tilde{\boldsymbol{\alpha}}_\mathbf{r}(\mathbf{t}), \ \ i = 1, \ldots, l.$$

Further remembering that the basis functions in $\tilde{\boldsymbol{\alpha}}_\mathbf{r}(\mathbf{t})$ are linearly independent, we get the constraints

$$(\mathbf{D_{n},}_i \mathbf{b})^T = \mathbf{n}_i, \ \ i = 1, \ldots, l. \tag{4.15}$$

4.4.4 The Constraint Equation

Combining (4.13), (4.14), (4.15) and possibly rows from the matrix $\mathbf{D}$ we can influence the result of the algebraic approximation process. The constraints can be organized as r linearly independent constraints on the algebraic hypersurface in a set of linear constraint equations:

$$\mathbf{K}\mathbf{b} = \mathbf{h}. \tag{4.16}$$

In the following theorem we assume that the matrix $\mathbf{K}$ and $\mathbf{b}$ can be split in two parts

$$\mathbf{K} = (\mathbf{K}_1, \mathbf{K}_2) \tag{4.17}$$

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix}. \tag{4.18}$$

where

- $\mathbf{K}_1$ is a nonsingular $r \times r$ matrix.

- $\mathbf{K}_2$ is a $r \times (\tilde{M} - r)$ matrix.

- $\mathbf{b}_1$ is a vector of length r.

- $\mathbf{b}_2$ is a vector of length $(\tilde{M} - r)$.

We thus assume that the constraints are linearly independent and that the number of constraints r is less than the number of coefficients in the algebraic hypersurface.

Theorem 44 *Let the combination of a hypersurface and a manifold satisfy the constraint equation defined in (4.16) which is organized as in (4.17) and (4.18) then*

$$|q(\mathbf{p}(\mathbf{s}))| \leq \|\mathbf{DK'b}_2 + \mathbf{h'}\|_2 \,,$$

where

$$\mathbf{K'} = \begin{pmatrix} -\mathbf{K}_1^{-1}\mathbf{K}_2 \\ \mathbf{I} \end{pmatrix},$$

and

$$\mathbf{h'} = \mathbf{D} \begin{pmatrix} \mathbf{K}_1^{-1}\mathbf{h} \\ \mathbf{0} \end{pmatrix}.$$

Proof. Using (4.17) and (4.18), we get

$$(\mathbf{K}_1, \mathbf{K}_2) \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \mathbf{h},$$

giving

$$\mathbf{b}_1 = \mathbf{K}_1^{-1}\mathbf{h} - \mathbf{K}_1^{-1}\mathbf{K}_2\mathbf{b}_2.$$

Resulting in

$$\mathbf{b} = \begin{pmatrix} -\mathbf{K}_1^{-1}\mathbf{K}_2 \\ \mathbf{I} \end{pmatrix} \mathbf{b}_2 + \begin{pmatrix} \mathbf{K}_1^{-1}\mathbf{h} \\ \mathbf{0} \end{pmatrix}.$$

Substituting this into (4.5) in lemma 38, we get

$$
\begin{aligned}
|q(\mathbf{p}(\mathbf{s}))| \quad &\leq \quad \|\mathbf{Db}\|_2 \qquad\qquad\qquad\qquad\qquad (4.19) \\
&\leq \quad \left\| \mathbf{D} \begin{pmatrix} -\mathbf{K}_1^{-1}\mathbf{K}_2 \\ \mathbf{I} \end{pmatrix} \mathbf{b}_2 + \mathbf{D} \begin{pmatrix} \mathbf{K}_1^{-1}\mathbf{h} \\ \mathbf{0} \end{pmatrix} \right\|_2 \\
&= \quad \|\mathbf{DK'b}_2 + \mathbf{h'}\|_2.
\end{aligned}
$$

∎

Remark 16 *By pivoting and removing redundant constraints, we can bring any constraint equation to the structure described in (4.17).*

Remark 17 *When we want to find an algebraic hypersurface approximating a manifold, a first choice of an algorithm is singular value decomposition of* $\mathbf{DK'}$. *This can be used in the case of both when* $\mathbf{h'} = \mathbf{0}$ *and* $\mathbf{h'} \neq \mathbf{0}$.

- *For* $\mathbf{h'} = \mathbf{0}$ *the problem reduce to* $|q(\mathbf{p}(\mathbf{s}))| \leq \|\mathbf{DK'b}_2\|_2$. *Thus, the coefficient vectors corresponding to the smallest singular values of* $\mathbf{DK'}$ *are solutions to the approximation problem.*

- *For* $\mathbf{h'} \neq \mathbf{0}$, *using e.g. theorem 6.8 in [Stewart:73] the singular value decomposition of* $\mathbf{DK'b}_2$ *can be used for finding the vector* $\mathbf{b}_2$, *giving the minimum of* $\|\mathbf{DK'b}_2 + \mathbf{h'}\|_2$.

4.4.5 Direct Search for an Approximative Null Space

We use rows in the $\mathbf{D}$ matrix as constraints, described as in Section 4.4.4, and here assume that the matrix $\mathbf{D}$ can be given the structure

$$\mathbf{D} = \left(\begin{array}{cc} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} \end{array} \right), \tag{4.20}$$

where

- $\mathbf{D}_{1,1}$ is a $r \times r$ nonsingular matrix.

- $\mathbf{D}_{2,1}$ is a $(N - r) \times r$.

- $\mathbf{D}_{1,2}$ is a $r \times (\tilde{M} - r)$ matrix.

- $\mathbf{D}_{2,2}$ is a $(N - r) \times (\tilde{M} - r)$ matrix.

Further let $\mathbf{b}$ be split as in (4.18)

$$\mathbf{b} = \left(\begin{array}{c} \mathbf{b}_1 \\ \mathbf{b}_2 \end{array} \right),$$

where

- $\mathbf{b}_1$ is a vector of length r.

- $\mathbf{b}_2$ is a vector of length $(\tilde{M} - r)$.

Lemma 45 *Let $\mathbf{D}$ be built up by the submatrices as described in (4.20) and let the constraints be described by*

$$\mathbf{D}_{1,1}\mathbf{b}_1 + \mathbf{D}_{1,2}\mathbf{b}_2 = \mathbf{0}. \tag{4.21}$$

Then

$$\mathbf{Db} = \left(\begin{array}{c} \mathbf{0} \\ -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) \mathbf{b}_2. \tag{4.22}$$

Proof. Using the constraints we get $\mathbf{b}_1 = -\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2}\mathbf{b}_2$, giving

$$\begin{aligned}
\mathbf{Db} &= \left(\begin{array}{cc} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} \end{array} \right) \left(\begin{array}{c} \mathbf{b}_1 \\ \mathbf{b}_2 \end{array} \right) \\[2mm]
&= \left(\begin{array}{cc} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} \end{array} \right) \left(\begin{array}{c} -\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{I} \end{array} \right) \mathbf{b}_2 \\[2mm]
&= \left(\begin{array}{c} \mathbf{0} \\ -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) \mathbf{b}_2.
\end{aligned}$$

■

In an elimination process it is convenient to maintain the size of the matrices during the process. Thus, we combine the constraint equation into the result of lemma 45 to maintain the size of $\mathbf{D}$ in the lemma following.

Theorem 46 *Let $\mathbf{D}$ be split as in (4.20) and use (4.21) as constraints. Then*

$$\mathbf{Db} = \left(\begin{array}{cc} \mathbf{I} & \mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{0} & -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) \left(\begin{array}{c} \mathbf{b}_1 \\ \mathbf{b}_2 \end{array} \right) \qquad (4.23)$$

with

$$\left(\begin{array}{cc} \mathbf{I} & \mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{0} & -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) = \left(\begin{array}{cc} \mathbf{D}_{1,1}^{-1} & \mathbf{0} \\ -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1} & \mathbf{I} \end{array} \right) \left(\begin{array}{cc} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} \end{array} \right).$$
$$(4.24)$$

Proof. Rewriting the constraint equation

$$\mathbf{D}_{1,1}\mathbf{b}_1 + \mathbf{D}_{1,2}\mathbf{b}_2 = \mathbf{0}$$

we get

$$\left(\begin{array}{cc} \mathbf{I} & \mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{0} & \mathbf{0} \end{array} \right) \mathbf{b} = \mathbf{0}.$$

From (4.22) we have that

$$\mathbf{Db} = \left(\begin{array}{c} \mathbf{0} \\ -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) \mathbf{b}_2 = \left(\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) \mathbf{b}.$$

Adding the latter equations, we obtain

$$\mathbf{Db} = \left(\begin{array}{cc} \mathbf{I} & \mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{0} & -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{array} \right) \mathbf{b}.$$

The matrix used for premultiplication in (4.24) can be shown to be correct by multiplying the two matrices. ■

If the entries in the block $\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2}$ in (4.22) have a sufficiently small value, then the results in lemma 45 can be used as an elimination process to find an approximative null space of $\mathbf{D}$. In the theorem following we introduce conditions that have to be satisfied to control such an elimination process.

Theorem 47 *Let $\mathbf{D}$ be built by submatrices as described in (4.20), let*

$$\mathbf{D}_{1,1}\mathbf{b}_1 + \mathbf{D}_{1,2}\mathbf{b}_2 = \mathbf{0},$$

73

and

$$\left\|-\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2}\right\|_\infty \leq \epsilon.$$

Then

$$\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s}\in\mathbf{\Omega}} |q(\mathbf{p}(\mathbf{s}))| \leq \epsilon\sqrt{(N-r)(\tilde{M}-r)}.$$

Proof. Using lemma 45 we get

$$\|\mathbf{D}\mathbf{b}\|_2 \leq \left\|-\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2}\right\|_\infty \|\mathbf{b}\|_2 \leq \epsilon\sqrt{(N-r)(\tilde{M}-r)}\,\|\mathbf{b}\|_2\,.$$

Now using (4.12) we get

$$\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s}\in\Omega} |\, q(\mathbf{p}(\mathbf{s}))\,|^2 \leq \min_{\|\mathbf{b}\|_2\neq 0} \frac{\mathbf{b}^T\mathbf{D}^T\mathbf{D}\mathbf{b}}{\mathbf{b}^T\mathbf{b}} \leq \epsilon^2(N-r)(\tilde{M}-r)$$

proving the theorem. ∎

The following elimination method is inspired by Gaussian elimination with pivotation. It starts with the line containing the element with the largest absolute values as constraint equation and perform the elimination. The elimination is repeated until all lines except one has been used for elimination, or the absolute value of the largest remaining element is less than a specified tolerance. The elimination sequence is remembered as well as the value of the largest element at each step in the process. Since the elimination sequence is determined during the elimination process, pointer arrays are used to introduce pivotation. The algorithms is based on theorem 46 with added pivotation, and repeated use with the submatrix $\mathbf{D}_{1,1}$ having dimension 1×1. The stop condition is based on theorem 47.

Algorithm

1. $\tilde{\mathbf{D}} = \mathbf{D}$.

2. Mark all rows and columns in $\mathbf{D}$ as untreated.

3. While the maximal untreated element in $\mathbf{D}$ is greater than ϵ in absolute value and more than one column untreated:

   (a) Let $d_{i,j}$ be the untrated element with maximal absolute value.

   (b) Divide row $\mathbf{d}_i$ by $d_{i,j}$.

   (c) Eliminate the entries with index j from all other rows of $\mathbf{D}$, i.e. $\mathbf{d}_l := \mathbf{d}_l - d_{l,j}\mathbf{d}_i,\ l \neq i$.

(d) Mark row i and column j as treated and remember that coefficient j was eliminated in row i, remember the absolute value of the largest element before the elimination step, as well as the number of the elimination step.

4. Now all untreated elements have absolute value less than or equal to ϵ, or only one column remains.

Remark 18 *At the end of this process the matrix has the structure that is a row and column pivotation of*

$$\begin{pmatrix} \mathbf{I} & \mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{0} & -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \end{pmatrix}.$$

If $\left\| -\mathbf{D}_{2,1}\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} + \mathbf{D}_{2,2} \right\|_{\infty} \leq \epsilon$ then independent of choice of $\mathbf{b}_2$ the value of

$$\min_{\|\mathbf{b}\|_2 = 1} \max_{\mathbf{s}\in\Omega} \mid q(\mathbf{p}(\mathbf{s})) \mid^2 \leq \epsilon^2 (N - r)(\tilde{M} - r),$$

because $\|\mathbf{b}_2\|_2 \leq \|\mathbf{b}\|_2$. Thus, we have found an approximative null space. Since

$$\begin{pmatrix} \mathbf{I} & \mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \end{pmatrix} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \mathbf{0},$$

we can find $\mathbf{b}_1$ by

$$\mathbf{b}_1 = -\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2}\mathbf{b}_2$$

and thus $\mathbf{b}$ by

$$\mathbf{b} = \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \begin{pmatrix} -\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2} \\ \mathbf{I} \end{pmatrix} \mathbf{b}_2.$$

The algorithm just described remembers $-\mathbf{D}_{1,1}^{-1}\mathbf{D}_{1,2}$, thus the approximative null space can be calculated from the reduced approximative null space.

This elimination algorithm performs more slowly than simple implementations of Singular Value Decomposition (SVD). However, it is faster than SVD implementations requiring $\mathbf{D}^T\mathbf{D}$ to be calculated. The remark, in the original version of these thesis, that the direct search for the approximative null space was slower than the use of SVD was based on the use of a fast SVD-algorithm that do not always converge. For more information on Singular Value Decomposition the book [Golub & Loan:83] can be consulted.

The tried SVD algorithms sometimes fail when the algebraic degree m is higher than 10 and the calculation is performed in double precision. In these cases the smallest singular values of $\mathbf{D}$ are separated less than the rounding errors in the $\mathbf{D}$ matrix. Thus, a minor change in $\mathbf{D}$ can induce a

major change in the singular subspace of $\mathbf{D}$ as stated in theorem 8.3-6. in [Golub & Loan:83]. The SVD implementations report lack of convergence. The reason for the failure is probably that there is an iteration process in the SVD implementation to speed up the calculation, and this fails when the separation of the singular values is small compared to the rounding errors.

4.5 Convergence Rate of Approximative Implicitization

For approximative implicitization to be useful we need to establish results on how the approximation error behaves, as the part of the manifold being approximated is reduced. The convergence rates established are best for low dimensional manifolds.

- Approximation of 1-manifolds in $\mathbb{R}^2$ with convergence rate

$$O(h^{\frac{(m+1)(m+2)}{2}-1}).$$

- Approximation of 1-manifolds in $\mathbb{R}^3$ with convergence rate

$$O(h^{\frac{(m+1)(m+2)(m+3)}{2}-1}).$$

- Approximation of 2-manifolds in $\mathbb{R}^3$ with convergence rate

$$O(h^{\left\lfloor \frac{1}{6}\sqrt{(9+12m^3+72m^2+132m)}-\frac{1}{2}\right\rfloor}).$$

In this section we at first present a theorem on the convergence rate of the approximation of a hypersurface of degree m to a manifold of dimension g in $\mathbb{R}^l$. Then we analyze the above-referenced convergence rates. However, the convergence rate can be misleading if the gradient of the hypersurface found, is vanishing or near vanishing, close to the manifold being approximated.

The proof of the theorem is based on power expansion around a point $\mathbf{v} \in \Omega$, where Ω is the parameter domain of the manifold $\mathbf{p}$ approximated. The value of h is used to define a closed hyperbox in Ω around $\mathbf{v}$ in the following way:

$$\Omega(\mathbf{v}, h) = \Omega \cap \left\{ \mathbf{s} \in \mathbb{R}^g \mid \|\mathbf{v} - \mathbf{s}\|_\infty \leq h \right\}.$$

It should be noted that the matrix $\mathbf{D}$ from (4.3) now is dependent on the value of $\mathbf{v}$ and h.

Theorem 48 *Given $q(\mathbf{x}) = 0$, an algebraic hypersurface in $\mathbb{R}^l$ of degree m with coefficient vector $\mathbf{b}$, and $\mathbf{p}(\mathbf{s})$, a C^{n+1}-continuous manifold of dimension g in $\mathbb{R}^l$. Then for $\mathbf{v} \in \Omega$ and $h > 0$*

$$\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s} \in \Omega(\mathbf{v},h)} |q(\mathbf{p}(\mathbf{s})| \leq \min_{\|\mathbf{b}\|_2=1} \|\mathbf{Db}\|_2 \leq O(h^{n+1}),$$

where n satisfies

$$\binom{n+g+1}{g} \geq \binom{m+l}{l} > \binom{n+g}{g}.$$

Proof. Use the description of the algebraic hypersurface in (3.3)

$$q(\mathbf{x}) = \sum_{\mathbf{i} \in I(m)} b_{\mathbf{i}} \frac{m!}{\mathbf{i}!} \beta^{\mathbf{i}}(\mathbf{x}),$$

and let

$$r_{\mathbf{i}}(\mathbf{s}) = \frac{m!}{\mathbf{i}!} \beta^{\mathbf{i}}(\mathbf{p}(\mathbf{s})).$$

Then (4.4) can be written

$$q(\mathbf{p}(\mathbf{s})) = \sum_{\mathbf{i} \in I(m)} b_{\mathbf{i}}\, r_{\mathbf{i}}(\mathbf{s}).$$

Now assume that $\mathbf{s} \in \Omega(\mathbf{v}, h)$, then with $\mathbf{s} = \mathbf{v} + \boldsymbol{\delta}$, we have that $\|\boldsymbol{\delta}\|_\infty \leq h$. Taylor expansion of $r_{\mathbf{i}}(\mathbf{s})$ around $\mathbf{v}$ with a polynomial of total degree n gives

$$r_{\mathbf{i}}(\mathbf{v} + \boldsymbol{\delta}) = t_{\mathbf{i}}(\boldsymbol{\delta}) + e_{\mathbf{i}}(\boldsymbol{\delta}),$$

where $t_{\mathbf{i}}(\boldsymbol{\delta})$ has total degree n, and $e_{\mathbf{i}}(\boldsymbol{\delta}) = O\left((\|\boldsymbol{\delta}\|_\infty)^{n+1}\right) \leq O(h^{n+1})$. The number of terms in a polynomial of total degree n in g variables is

$$\binom{n+g}{g}.$$

The number of coefficients in the algebraic hypersurface is

$$\binom{m+l}{l}.$$

Provided the number of coefficients in the algebraic hypersurface is greater than number of coefficients in the polynomial of total degree n, then the polynomials $t_{\mathbf{i}}(\boldsymbol{\delta})$ are linearly dependent. I.e. if

$$\binom{m+l}{l} > \binom{n+g}{g}$$

we can find coefficients $\mathbf{b}'$ with $\|\mathbf{b}'\|_2 = 1$, such that

$$\sum_{\mathbf{i} \in I(m)} b'_{\mathbf{i}} \, t_{\mathbf{i}}(\boldsymbol{\delta}) = 0. \tag{4.25}$$

We are interested in the largest natural number n with this property, thus we state

$$\binom{n+1+g}{g} \geq \binom{m+l}{l} > \binom{n+g}{g}.$$

Using the above Taylor expansion, the associated error functions, and that $\boldsymbol{\delta} = \mathbf{s} - \mathbf{v}$, we get

$$q(\mathbf{p}(\mathbf{s})) = (\mathbf{Tb})^T \boldsymbol{\alpha}(\mathbf{s}) + (\mathbf{Eb})^T \boldsymbol{\alpha}(\mathbf{s})$$

where

$$(\mathbf{Tb})^T \boldsymbol{\alpha}(\mathbf{s}) = \sum_{\mathbf{i} \in I(m)} b_{\mathbf{i}} \, t_{\mathbf{i}}(\mathbf{s} - \mathbf{v})$$

and

$$(\mathbf{Eb})^T \boldsymbol{\alpha}(\mathbf{s}) = \sum_{\mathbf{i} \in I(m)} b_{\mathbf{i}} \, e_{\mathbf{i}}(\mathbf{s} - \mathbf{v}).$$

Now remember that $e_{\mathbf{i}}(\mathbf{s} - \mathbf{v}) = e_{\mathbf{i}}(\boldsymbol{\delta}) = O(h^{n+1})$, we have that

$$(\mathbf{Eb})^T \boldsymbol{\alpha}(\mathbf{s}) = O(h^{n+1}).$$

All entries in $\boldsymbol{\alpha}(\mathbf{s})$ are linearly independent, thus all entries in $\mathbf{Eb}'$ must be $O(h^{n+1})$, giving $\|\mathbf{Eb}'\|_2 = O(h^{n+1})$.

Now remembering (4.25) and that $\mathbf{Tb}' = \mathbf{0}$, we get

$$\begin{aligned}
\min_{\|\mathbf{b}\|_2=1} \max_{\mathbf{s} \in \Omega(\mathbf{v},h)} |q(\mathbf{p}(\mathbf{s}))| \;\; &\leq \;\; \min_{\|\mathbf{b}\|_2=1} \|\mathbf{Db}\|_2 \leq \frac{\|\mathbf{Db}'\|_2}{\|\mathbf{b}'\|_2} \\
&= \;\; \frac{\|(\mathbf{T}+\mathbf{E})\mathbf{b}'\|_2}{\|\mathbf{b}'\|_2} = \frac{\|\mathbf{Eb}'\|_2}{\|\mathbf{b}'\|_2} = O(h^{n+1}).
\end{aligned}$$

$\blacksquare$

We now look at curves in $\mathbb{R}^2$ to illustrate the properties of approximative implicitization.

Corollary 49 *Let $q(\mathbf{x}) = 0$ be an algebraic curve of degree m with coefficient vector $\mathbf{b}$, and let $\mathbf{p}(s)$ be a sufficiently smooth 1-manifold in $\mathbb{R}^2$. Then*

$$\min_{\|\mathbf{b}\|_2=1} \max_{s \in \Omega(v,h)} |q(\mathbf{p}(s)| \leq \min_{\|\mathbf{b}\|_2=1} \|\mathbf{Db}\|_2 \leq O(h^{\frac{(m+1)(m+2)}{2}-1}).$$

Algebraic degree m	1	2	3	4	5	6	7	8	9	10
Convergence rate	2	5	9	14	20	27	35	44	54	65

Table 4.2: Convergence rate of the approximative implicitization of a sufficiently smooth parametric curve in $\mathbb{R}^2$ with an algebraic curve of degree m, $1 \le m \le 10$.

Proof. Let $l = 2$ and $g = 1$ in theorem 48 giving

$$\binom{n+2}{1} \ge \binom{m+2}{2} > \binom{n+1}{1}.$$

Rewriting we get

$$n + 2 \ge \frac{(m+1)(m+2)}{2} > n + 1,$$

or

$$n = \frac{(m+1)(m+2)}{2} - 2.$$

We add 1 to get the convergence rate. ∎

Example 50 *In table 4.2 we show the convergence rate from corollary 49. The convergence rate is given for algebraic degrees in the range, $1 \le m \le 10$. Note that the convergence rate is high even for low algebraic degrees.*

The corollary following shows that the convergence rate increases as the dimension l of the space in which the 1-manifold lies grow.

Corollary 51 *Let $q(\mathbf{x}) = 0$ be an algebraic surface of degree m with co-efficient vector $\mathbf{b}$, and let $\mathbf{p}(s)$ be a sufficiently smooth 1-manifold in $\mathbb{R}^3$. Then*

$$\min_{\|\mathbf{b}\|_2=1} \max_{s \in \Omega(v,h)} |q(\mathbf{p}(s)| \le \min_{\|\mathbf{b}\|_2=1} \|\mathbf{Db}\|_2 \le O(h^{\frac{(m+1)(m+2)(m+3)}{6}-1}).$$

Proof. Let $l = 3$ and $g = 1$ in theorem 48 giving

$$\binom{n+2}{1} \ge \binom{m+3}{3} > \binom{n+1}{1}.$$

Rewriting we get

$$n + 2 \ge \frac{(m+1)(m+2)(m+3)}{6} > n + 1,$$

or

$$n = \frac{(m+1)(m+2)(m+3)}{2} - 2.$$

We add 1 to get the convergence rate. ∎

Algebraic degree m	1	2	3	4	5	6	7	8	9	10
Convergence rate	3	9	19	34	55	83	119	164	219	285

Table 4.3: Convergence rate of the approximative implicitization of a sufficiently smooth parametric curve in $\mathbb{R}^3$ with an algebraic surface of degree m, $1 \leq m \leq 10$.

Example 52 *In table 4.3 we show the convergence rate from corollary 51. The convergence rate is given for algebraic degrees m in the range, $1 \leq m \leq 10$. Note that the extremely high convergence rate is due to the high number of coefficients in an algebraic surface.*

The corollary following show that the convergence rate is lower for 2-manifolds in $\mathbb{R}^3$, than for 1-manifolds in $\mathbb{R}^2$.

Corollary 53 *Let $q(\mathbf{x}) = 0$ be an algebraic surface of degree m with coefficient vector $\mathbf{b}$, and let $\mathbf{p}(s,t)$ be a sufficiently smooth 2-manifold in $\mathbb{R}^3$. Then*

$$\min_{\|\mathbf{b}\|_2=1} \max_{(s,t)\in\Omega(\mathbf{v},h)} |q(\mathbf{p}(s,t))| \leq \min_{\|\mathbf{b}\|_2=1} \|\mathbf{Db}\|_2 \leq O(h^{\left\lfloor \frac{1}{6}\sqrt{(9+12m^3+72m^2+132m)}-\frac{1}{2}\right\rfloor}).$$

Proof. Let $l = 3$ and $g = 2$ in theorem 48 giving

$$\binom{n+3}{2} \geq \binom{m+3}{3} > \binom{n+2}{2}.$$

Now finding n such that

$$\binom{m+3}{3} - 1 = \binom{n+2}{2},$$

and using the largest natural number less than the calculated n we get

$$n(m) = \left\lfloor \frac{1}{6}\sqrt{(9+12m^3+72m^2+132m)} - \frac{3}{2}\right\rfloor.$$

We add 1 to get the convergence rate. ■

Example 54 *In table 4.4 we show the convergence rate from corollary 53. The convergence rate is given for algebraic degrees m in the range, $1 \leq m \leq 10$. Note that the convergence rate is relatively high even for low algebraic degrees.*

Algebraic degree m	1	2	3	4	5	6	7	8	9	10
Convergence rate	2	3	5	7	10	12	14	17	20	23

Table 4.4: Convergence rate of the approximative implicitization of a sufficiently smooth parametric surface in $\mathbb{R}^3$ with an algebraic surface of degree m, $1 \leq m \leq 10$.

4.6 Accuracy of Approximative Implicitization

In many cases we want an approximation to be within a certain tolerance $\epsilon > 0$ of the object approximated. When we measure the distance from a manifold $\mathbf{p}(\mathbf{s})$ to a manifold $\mathbf{q}(\mathbf{t})$ a natural choice for distance measure is

$$d(\mathbf{p}, \mathbf{q}) = \max_{\mathbf{s}} \inf_{\mathbf{t}} \|\mathbf{p}(\mathbf{s}) - \mathbf{q}(\mathbf{t})\|_2 .$$

In this definition inf is used since the manifolds are not required to be closed. The distance measure can be interpreted:

- For all points on $\mathbf{p}$ find the closest point on $\mathbf{q}$.

- Use the maximum distance found as distance measure.

Note that this expression is similar to (2.2), that we used in the discussion concluding with the definition of Local Maximal Points of one set with respect to another set.

To use the distance function $d(\mathbf{p}, \mathbf{q})$ we require a parametric description of both manifolds. In the case we shall discuss, this is not the case. We have an algebraic hypersurface $q(\mathbf{x}) = 0$ and a parametric represented manifold $\mathbf{p}(\mathbf{s})$. To measure the error we now introduce a direction for error measurement $\mathbf{g}(\mathbf{s})$ satisfying $\|\mathbf{g}(\mathbf{s})\|_2 = 1$. To be a useful direction for error measurement we assume that there exists an error function $\rho(\mathbf{s}) \in C^0$ such that

$$q(\mathbf{p}(\mathbf{s}) - \rho(\mathbf{s})\mathbf{g}(\mathbf{s})) = 0.$$

Not all choices for direction for error measurement will satisfy this requirement. To simplify the notion we now let $\mathbf{p} = \mathbf{p}(\mathbf{s})$, $\mathbf{g} = \mathbf{g}(\mathbf{s})$ and $\rho = \rho(\mathbf{s})$. The above equation can then be expressed

$$q(\mathbf{p} - \rho\mathbf{g}) = 0.$$

Taylor expansion with respect to ρ now gives

$$q(\mathbf{p}) - \nabla q(\mathbf{p} - \theta\mathbf{g}) \cdot \rho\mathbf{g} = 0$$

with $\theta\rho \geq 0$ and $|\theta| \leq |\rho|$. Rearranging we get

$$\rho\left(\nabla q(\mathbf{p} - \theta\mathbf{g}) \cdot \mathbf{g}\right) = q(\mathbf{p}).$$

Now setting

$$\rho_{\max} = \max_{\mathbf{s}\in\Omega} |\rho(\mathbf{s})|,$$

and assuming that

$$\nabla q(\mathbf{p} - \theta\mathbf{g}) \cdot \mathbf{g} \neq 0, \ \mathbf{s} \in \Omega, \ |\theta| \leq \rho_{\max}$$

we get, by reintroducing $\mathbf{p}(\mathbf{s})$, $\mathbf{g}(\mathbf{s})$ and $\rho(\mathbf{s})$

$$\rho(\mathbf{s}) = \frac{q(\mathbf{p}(\mathbf{s}))}{\nabla q(\mathbf{p}(\mathbf{s}) - \theta\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})} \tag{4.26}$$

with $|\theta| \leq |\rho(\mathbf{s})|$.

We see that the relative orientation of the direction for error measurement, and the gradient of q plays a central role in estimating the error. Thus, an optimal solution is to find a direction for error measurement that gives the maximum value of the denominator in the expression above. This requirement can be formulated as finding $\mathbf{g}(\mathbf{s})$ with $\|\mathbf{g}(\mathbf{s})\|_2 = 1$ that satisfies

$$\min_{\substack{s \in \Omega, \\ |\theta| \leq \rho_{\max}}} |\nabla q(\mathbf{p}(\mathbf{s}) - \theta\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})| = \max_{\|\hat{\mathbf{g}}(\mathbf{s})\|_2 = 1} \min_{\substack{s \in \Omega, \\ |\theta| \leq \rho_{\max}}} |\nabla q(\mathbf{p}(\mathbf{s}) - \theta\hat{\mathbf{g}}(\mathbf{s})) \cdot \hat{\mathbf{g}}(\mathbf{s})|.$$

However, to find such a maximum is not simple. Thus, for practical purposes we want to make a simpler choice of direction for error measurement. From (4.26) we see that if the gradient gets small the error is growing, we also see that if the gradient and direction for error measurement are near normal the error will be growing. Thus, to make the direction for error measurement dependent on the gradient direction seems natural. At the same time for curves in $\mathbb{R}^2$ or surfaces in $\mathbb{R}^3$ it is natural to base the direction for error measurement on the normal. This is detailed in the following subsection.

4.6.1 Simple Direction for Error Measurement

We want to choose a direction for error measurement that is in accordance with [Degen:92], where for curves in $\mathbb{R}^2$ the direction chosen is the curve normal. Further we want that for $(l-1)$-manifolds in $\mathbb{R}^l$ the direction coincides, if possible, with the normals of the smooth parameterized bounded manifolds. Besides we want to relate the direction for error measurement to the gradient of the polynomial q defining the hypersurface $q(\mathbf{x}) = 0$.

The approach is based on Gram-Schmidt orthonormalization, and has 3 steps:

1. For the point $\mathbf{p}(\mathbf{s})$ on the bounded g-manifold, for which we want to measure the distance to the approximating hypersurface, we find an orthonormal basis $\mathbf{t}_i(\mathbf{s})$, $i = 1, \ldots, g$ spanning the tangent bundle.

$$\mathbf{t}_i(\mathbf{s}) = \frac{\frac{\partial \mathbf{p}(\mathbf{s})}{\partial s_i} - \sum_{j=1}^{i-1} \left(\frac{\partial \mathbf{p}(\mathbf{s})}{\partial s_i} \cdot \mathbf{t}_j(\mathbf{s}) \right) \mathbf{t}_j(\mathbf{s})}{\left\| \frac{\partial \mathbf{p}(\mathbf{s})}{\partial s_i} - \sum_{j=1}^{i-1} \left(\frac{\partial \mathbf{p}(\mathbf{s})}{\partial s_i} \cdot \mathbf{t}_j(\mathbf{s}) \right) \mathbf{t}_j(\mathbf{s}) \right\|_2}, \quad i = 1, \ldots, g.$$

This orthonormal basis exists since we assume that $\mathbf{p}(\mathbf{s})$ is properly parameterized.

2. We remove from the gradient the components spanned by the tangent bundle

$$\hat{\mathbf{n}}(\mathbf{s}) = \nabla q(\mathbf{p}(\mathbf{s})) - \sum_{i=1}^{g} \left(\nabla q(\mathbf{p}(\mathbf{s})) \cdot \mathbf{t}_i(\mathbf{s}) \right) \mathbf{t}_i(\mathbf{s}).$$

3. Then we normalize $\hat{\mathbf{n}}(\mathbf{s})$ if $\hat{\mathbf{n}}(\mathbf{s}) \neq \mathbf{0}$

$$\tilde{\mathbf{n}}(\mathbf{s}) = \begin{cases} \dfrac{\hat{\mathbf{n}}(\mathbf{s})}{\|\hat{\mathbf{n}}(\mathbf{s})\|_2} & \text{if} \quad \|\hat{\mathbf{n}}(\mathbf{s})\|_2 \neq 0 \\ \mathbf{0} & \text{if} \quad \|\hat{\mathbf{n}}(\mathbf{s})\|_2 = 0 \end{cases}. \tag{4.27}$$

Definition 55 (Simple Direction for Error Measurement) *The direction $\tilde{\mathbf{n}}(\mathbf{s})$ defined in (4.27), is denoted **Simple Direction for Error Measurement**. This direction is used for measuring the distance from a point $\mathbf{p}(\mathbf{s})$, on a regularly parameterized smooth bounded g-manifold, to a hypersurface $q(\mathbf{x}) = 0$.*

When $\tilde{\mathbf{n}}(\mathbf{s}) = \mathbf{0}$, the gradient of the algebraic hypersurface at $\mathbf{p}(\mathbf{s})$ is either $\mathbf{0}$, or coincide with the tangent bundle of the manifold at $\mathbf{p}(\mathbf{s})$. In both cases the algebraic surface is not an optimal approximation to the manifold.

We now look at the consequence of this definition for curves (1-manifolds) in $\mathbb{R}^2$ and $\mathbb{R}^3$, and for surfaces (2-manifolds) in $\mathbb{R}^3$.

Example 56 *For 1-manifolds in $\mathbb{R}^2$, $\tilde{\mathbf{n}}(s)$ is the unit normal vector at $\mathbf{p}(s)$, provided the gradient of the polynomial describing the algebraic curve is nonvanishing at $\mathbf{p}(s)$ and does not coincide with the tangent at $\mathbf{p}(s)$. This is in accordance with [Degen:92] when $\tilde{\mathbf{n}}(s) \neq \mathbf{0}$.*

Example 57 *For* 2-*manifolds in* $\mathbb{R}^3$, $\tilde{\mathbf{n}}(\mathbf{s})$ *is the unit normal vector at* $\mathbf{p}(\mathbf{s})$, *provided the gradient of the algebraic surface is nonvanishing at* $\mathbf{p}(s)$ *and does not coincide with the tangent bundle at* $\mathbf{p}(\mathbf{s})$.

Example 58 *For* 1-*manifolds in* $\mathbb{R}^3$, $\tilde{\mathbf{n}}(s)$ *is the normalized version of the gradient of the algebraic surface at* $\mathbf{p}(s)$ *with the component of the gradient in the direction of the tangent of* $\mathbf{p}(s)$ *removed.* $\tilde{\mathbf{n}}(s) = \mathbf{0}$ *if the gradient and the tangent coincide at* $\mathbf{p}(\mathbf{s})$.

Prof. T. W. Sederberg pointed out to me that in the case there exists a direction where the first derivative of q is monotonic, then the error bound in [Sederberg:89-3] can be used. This is a easier than approach described above.

4.7   Admissable Algebraic Approximations

The contents of this section is not essential for the use of the approximative implicitization and is only included to try to make a tool for evaluating if an algebraic approximation is "good".

In [Degen:92] a set of *admissible parametric curves* $\mathcal{A}(C)$ for the approximation of a parametric curve $C \in C^2$ and the concept of the *normal distance* from the original curve to the approximation were introduced. We here introduce similar concepts for judging if an algebraic hypersurface approximating a manifold is within a specified approximation tolerance and "admissible". For certain purposes e.g. to separate geometries, it is sufficient that the hypersurface is within a certain distance of the manifold. For judging the approximation of a manifold with a hypersurface or with an intersection of a number of hypersurfaces we have to define a concept of admissible hypersurfaces. The concept of *admissible curves* was in [Degen:92] introduced to avoid that the approximation turns with respect to the curve approximated, and to ensure that the tangents at corresponding points on the two curves have a positive scalar product.

Another useful concept from [Degen:92] is the *normal field* $\mathcal{N}(C)$ of a curve $C \in C^2$. With each point $\mathbf{p}(t) \in C$ it was associated the part of the normal $\{\mathbf{p}(t) + \rho\mathbf{n}(t) \mid \rho \in \mathbb{R}\}$ for which $\rho\kappa(t) < 1$, (ρ unrestricted if $\kappa(t) = 0$), $\kappa(t)$ being the curvature of C and $\mathbf{n}(t)$ the normal unit vector. The mapping $(t, \rho) \to \mathbf{p}(t) + \rho\mathbf{n}(t)$ from $\{(t, \rho) \in \mathbb{R}^2 \mid t \in [a, b], \rho\kappa(t) < 1\}$ is locally one-to-one. I.e., this mapping is a regular parametrization of $\mathcal{N}(C)$.

The problem of approximating a regularly parameterized manifold with an algebraic hypersurface, is that there is no reparametrization connecting points on the manifold with points on the algebraic hypersurface. In addition

the algebraic hypersurfaces are in most cases infinite, and our interest is related to the part of the algebraic hypersurface approximating the manifold. A concept of admissible algebraic hypersurfaces is useful for focusing on the part of the hypersurface that approximate the manifold. We further want to avoid the turning of the hypersurface and manifold with respect to each other in the area of interest. In addition we want to avoid asymptotes, and possibly fix interpolation of the boundary of the bounded smooth manifold.

In subsection 4.6.1 we defined a *simple direction for error measurement* to be used in the approximative implicitization process. The direction defined is normal to the parametric manifold and based on the gradient of the polynomial q defining the hypersurface representing the algebraic approximation. This section is structured as follows:

- Based on the simple direction for error measurement a normal field of the manifold is defined in subsection 4.7.1. This definition requires only C^1 continuity of the manifolds being approximated, not C^2 as required in [Degen:92].

- Then in subsection 4.7.2 we define when a hypersurface resulting from the approximative implicitization process is admissable.

4.7.1   Normal Field of a Manifold

We are now in the position to define the normal field of a smooth bounded manifold in $\mathrm{I\!R}^l$ with respect to an algebraic hypersurface $q(\mathbf{x}) = 0$.

In [Degen:92] the signed curvature $\kappa(s)$ of a C^2 curve $\mathbf{p}(s)$, $s \in [a, b]$ with regular parametrization was used to define the normal field by

$$\mathcal{N} = \{\mathbf{p}(s) + \rho\mathbf{n}(s) \mid s \in [a, b], \ \rho\kappa(s) < 1\} \tag{4.28}$$

with $\mathbf{n}(s)$ the normal of $\mathbf{p}$ at $\mathbf{p}(s)$. The term $\rho\kappa(s)$ is less than 1 to avoid including the centre of the radius of curvature, and points further apart from the curve than the centre of radius of curvature. When addressing higher dimensional manifolds, it is desirable not to handle curvatures. We want to replace the role of curvature by hyperballs, and thus in the following discussion reformulate (4.28). The resulting equation contains open balls instead of curvatures.

All $(s, \rho) \in \mathcal{N}$ satisfy

$$\exists \delta > 0 : \mathcal{B}(\mathbf{p}(s) + \rho\mathbf{n}(s), |\rho|) \cap \mathcal{P}(s, \delta) = \emptyset. \tag{4.29}$$

In this expression the following concepts are used:
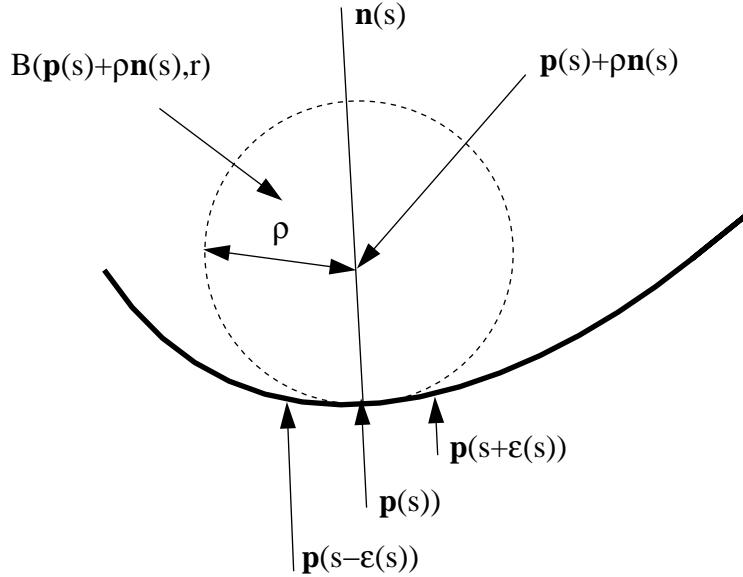
Figure 4.1: Geometric building blocks used for defining the normal set of a 1-manifold in $\mathbb{R}^2$. For 1-manifolds in $\mathbb{R}^2$ the radius of curvature could be used directly as in [Degen:92]. However, for higher dimensional hypersurfaces this definition of the normal field allows us not to be concerned with the definition of curvature.

- An open ball with radius r.

$$\mathcal{B}(\mathbf{p}, r) = \{\mathbf{y} \mid \|\mathbf{p} - \mathbf{y}\|_2 < r\}.$$

- A segment of the manifold $\mathbf{p}$ defined by an open ball of radius δ around the point $\mathbf{s}$ in the parametrization of $\mathbf{p}$.

$$\mathcal{P}(\mathbf{s}, \delta) = \{\mathbf{p}(\mathbf{t}) \mid \mathbf{t} \in \mathcal{B}(\mathbf{s}, \delta) \cap \Omega\}.$$

Equation (4.29) is based on two geometric components:

- Points $\mathbf{p}(s)$ on the curve $\mathbf{p}$.

- Open balls that are tangential to $\mathbf{p}$ at $\mathbf{p}(s)$ with radius $|\rho|$ and center $\mathbf{p}(s) + \rho\mathbf{n}(s)$.

The equation expresses that all values of ρ are allowed that make the intersection of the internal of the ball, and a segment of the curve around $\mathbf{p}(s)$ empty. Figure 4.1 illustrates this for smooth curves in $\mathbb{R}^2$. However, (4.29) also allows for ρ to be equal to the signed radius of curvature of $\mathbf{p}(s)$, this was not allowed in [Degen:92]. Now the normal field in [Degen:92] can be reformulated, giving

$$\mathcal{N} = \{\mathbf{p}(s) + \rho\,\mathbf{n}(s) \in \mathbb{R}^l \mid (s,\rho) \in \mathcal{S}(\mathbf{p})\}$$

with

$$\mathcal{S}(\mathbf{p}) = \left\{(s,\rho) \in \mathbb{R}^2 \mid s \in [a,b] \wedge \rho \in \mathcal{I}(\mathbf{p},s)\right\},$$

where

$$\mathcal{I}(\mathbf{p},s) = int\left\{\rho \in \mathbb{R} \mid \exists \delta > 0 : \mathcal{B}(\mathbf{p}(s) + \rho\mathbf{n}(s), |\rho|) \cap \mathcal{P}(s,\delta) = \emptyset\right\}.$$

The $int()$ is introduced to remove values of ρ equal to the signed radius of curvature from the set $\mathcal{I}(\mathbf{p},s)$. It should be observed that for $\rho = 0$ the open ball is empty. Thus, $\rho = 0$ is included for all values of s independent of the behavior of the curve.

The next step is to generalize this expression to:

- Smooth manifolds $\mathbf{p}$ of dimension g in $\mathbb{R}^l$.

- The *simple direction for error measurement* $\tilde{\mathbf{n}}(s)$ given in definition 55 instead of the normal vector $\mathbf{n}(s)$.

According to the premisses we now define the normal field.

Definition 59 (Normal Field) *Let $\mathbf{p}$ be a smooth manifold of dimension g in $\mathbb{R}^l$, with parametrization defined over a compact set $\Omega \subset \mathbb{R}^g$. The normal field of $\mathbf{p}(\mathbf{s})$ with respect to an algebraic hypersurface $q(\mathbf{x})=0$ is defined by*

$$\mathcal{N}_q(\mathbf{p}) = \{\mathbf{p}(\mathbf{s}) + \rho\,\tilde{\mathbf{n}}(\mathbf{s}) \in \mathbb{R}^l \mid (\mathbf{s},\rho) \in \mathcal{S}_q(\mathbf{p})\}. \qquad (4.30)$$

Here $\tilde{\mathbf{n}}(\mathbf{s})$ is the simple direction for error measurement with respect to the hypersurface $q(\mathbf{x})=0$ given in definition 55. The set $\mathcal{S}_q(\mathbf{p})$, denoted the para-metrization of the normal field, is given by

$$\mathcal{S}_q(\mathbf{p}) = \left\{(\mathbf{s},\rho) \in \mathbb{R}^{g+1} \mid \mathbf{s} \in \Omega \wedge \rho \in \mathcal{I}_q(\mathbf{p},\mathbf{s})\right\} \qquad (4.31)$$

with

$$\mathcal{I}_q(\mathbf{p},\mathbf{s}) = int\left\{\rho \in \mathbb{R} \mid \exists \delta > 0 : \mathcal{B}(\mathbf{p}(\mathbf{s}) + \rho\,\tilde{\mathbf{n}}(\mathbf{s}), |\rho|) \cap \mathcal{P}(\mathbf{s},\delta) = \emptyset\right\}.$$

Since $\tilde{\mathbf{n}}(\mathbf{s})$ is defined relative to the gradient of the polynomial $q(\mathbf{x})$ de-scribing an algebraic hypersurface, $\tilde{\mathbf{n}}(\mathbf{s})$ is sensitive to the relative behavior of $q(\mathbf{x})$ and $\mathbf{p}(\mathbf{s})$. For $(l-1)$-manifolds in $\mathbb{R}^l$, for which the normal can be used directly, this sensitivity is unnecessary and a negative aspect of the definition. However, the consequences are not severe, when the polynomial $q(\mathbf{x})$ used for defining $\tilde{\mathbf{n}}(\mathbf{s})$, describes the algebraic hypersurface we want to analyze. Then the vanishing of $\tilde{\mathbf{n}}(\mathbf{s})$ indicates that $q(\mathbf{x}) = 0$ is not a good approximation to $\mathbf{p}(\mathbf{s})$. This, because $\tilde{\mathbf{n}}(\mathbf{s}) = \mathbf{0}$ if either:

- The gradient of $q(\mathbf{x})$ vanish at some point of $\mathbf{p}(\mathbf{s})$.

- Some tangent in the tangent bundle of the manifold and the gradient of $q(\mathbf{x})$ are parallel at the same point on the manifold.

In the examples following we assume that the *simple direction for error measurement* is nonvanishing for all $\mathbf{s} \in \Omega$, thus we have that $\tilde{\mathbf{n}}(\mathbf{s}) = \mathbf{n}(\mathbf{s})$ for $(l-1)$ manifolds in $\mathbb{R}^l$.

Example 60 *Let $\mathbf{p}$ be a surface in $\mathbb{R}^3$, thus $g = l - 1 = 2$.*

- *Assume first that $\mathbf{p}(\mathbf{s})$ is a saddle point on $\mathbf{p}$. Then we know that the Gaussian curvature at $\mathbf{p}(\mathbf{s})$ is negative, and thus $\rho \in \langle \rho_{\min}, \rho_{\max} \rangle$, with $\rho_{\min} < 0 < \rho_{\max}$.*

- *Assume then that $\mathbf{p}(\mathbf{s})$ is a point with positive Gaussian curvature. Then either*
$$\rho_{\min} < \rho_{\max} < 0 \ \wedge \ \rho \in \langle \rho_{\max}, \infty \rangle,$$
or
$$0 < \rho_{\min} < \rho_{\max} \ \wedge \ \rho \in \langle -\infty, \rho_{\min} \rangle.$$

Example 61 *Let $\mathbf{p}(s)$, $s \in [a, b]$ be a segment of a circle of radius 1 in $\mathbb{R}^2$, with the normal $\mathbf{n}(\mathbf{s})$ pointing outward. Then for all $s \in [a, b]$, ρ satisfy $\rho \in \langle -1, \infty \rangle$. Thus, with exception of the circle centre all points in $\mathbb{R}^2$ are part of the normal field. This also corresponds to the requirement in [Degen:92] where $\kappa(s)\rho < 1$. For the circle in question $\kappa(s) = -1$ and $\rho > -1$ satisfy $\kappa(s)\rho < 1$.*

In the example following the dependency of the normal filed on the algebraic surface it is related to is illustrated.

Example 62 *Let $\mathbf{p}$ be a circle in $\mathbb{R}^3$. First assume that the algebraic surface $q(x, y) = 0$ is the plane in which the circle lies. In this case the simple direction for error measurement is constant and parallel to the normal of the plane, and ρ can take on any value.*

Then assume that the algebraic surface $q(x, y) = 0$ is a cone in which the circle lies. In this case the simple direction for error measurement is parallel to the normal of the cone at any given point.

In the latter case assume that the algebraic surface $q(x, y) = 0$ is cylinder. The simple direction for error measurement is now parallel to the curvature of the circle at any given point on the circle.

4.7.2 Definition of Admissible Hypersurfaces

For practical use of approximative implicitization we must be able to evaluate the quality of the approximation. To do this we introduce the following concepts for determining when an approximation has a certain quality.

- Admissible approximation. This is a natural concept when we want to separate two manifolds of possibly different dimensions.

- Boundary interpolating admissable approximation. This is a natural concept when the approximative implicitization has to take adjacent manifolds into consideration.

Definition 63 (Admissible Hypersurface) *An algebraic hypersurface* $q(\mathbf{x}) = 0$ *is called **admissible** with respect to a regular parameterized smooth manifold with boundary in* $\mathbb{R}^l$, *if the following conditions hold:*

1. *The region of an algebraic hypersurface being used is contained in the normal field* $\mathcal{N}_q(\mathbf{p})$ *of the manifold being approximated. E.g. there exists* $\rho(\mathbf{s})$ *such that for* $\mathbf{s} \in \Omega$ *the following is valid*

   - *The* simple direction for error measurement *is nonvanishing*

     $$\|\tilde{\mathbf{n}}(\mathbf{s})\|_2 \neq 0.$$

   - *There are points in the algebraic hypersurface in the direction of error measurement*

     $$q(\mathbf{p}(\mathbf{s}) + \rho(\mathbf{s})\tilde{\mathbf{n}}(\mathbf{s})) = 0.$$

   - $(\mathbf{s}, \rho(\mathbf{s})) \in S_q(\mathbf{p})$, *i.e. that* $(\mathbf{s}, \rho(\mathbf{s}))$ *is contained in the parametrization of the normal field.*

2. *The gradient of* $q(\mathbf{x})$ *is nonvanishing and not normal to the direction of error measurement, i.e. for* $\mathbf{s} \in \Omega$ *and* $\theta(\mathbf{s}) \in [-|\rho(\mathbf{s})|, |\rho(\mathbf{s})|]$ *we have that*
   $$\nabla q(\mathbf{p}(\mathbf{s}) + \theta(\mathbf{s})\tilde{\mathbf{n}}(\mathbf{s})) \cdot \tilde{\mathbf{n}}(\mathbf{s}) \neq 0.$$

3. *The branch of the algebraic approximation chosen is the branch closest to the manifold. I.e.*

   $$\nexists \mathbf{r} \in \mathbb{R}^l : \quad q(\mathbf{r}) = 0 \quad \wedge \quad (\|\mathbf{r} - \mathbf{p}(\mathbf{s})\|_2 \leq \rho(\mathbf{s}))$$

   $$\wedge \quad \left\{ (\mathbf{r} - \mathbf{p}(\mathbf{s})) \cdot \frac{\partial \mathbf{p}(\mathbf{s})}{\partial s_i} = 0, \ i = 1, \dots, g \right\}.$$

   $$\wedge \quad \mathbf{r} \neq \mathbf{p}(\mathbf{s}).$$

Note that $\dfrac{\partial \mathbf{p}(\mathbf{s})}{\partial s_i} \neq \mathbf{0}$, because $\mathbf{p}(\mathbf{s})$ has a regular parametrization.

We denote all admissible hypersurfaces with respect to a manifold $\mathbf{p}$, $\mathcal{A}(\mathbf{p})$.

For a number of applications, when it is desirable to replace a manifold by algebraic approximations, it is necessary to interpolate the boundaries of the manifold. This is so often desirable that we introduce a separate name for these approximations.

Definition 64 (Boundary Admissible Hypersurface) *Let the algebraic hypersurface $q(\mathbf{x}) = 0$ be admissible with respect to a regular parameterized closed and bounded smooth manifold $\mathbf{p}$ in $\mathbb{R}^l$. The hypersurface is called* **boundary admissible** *if the boundary of the manifold lies on the approximation.*

Remark 19 *We make a distinction between the notion admissible and the notion boundary admissible, because the number of conditions for satisfying interpolation of the boundaries, in many cases, prohibits the use of low polynomial degrees in the algebraic hypersurface in the approximations. In [Degen:92] curves in $\mathbb{R}^2$ were addressed and the boundary interpolation only involved the start and end point. For manifolds of high dimension the number of conditions to satisfy boundary interpolation is high.*

With these definitions we exclude asymptotes, since an asymptote would introduce asymptotes or breaks in $\rho(\mathbf{s})$ and this is impossible since ρ is C^0-continuous. Singular points in the region where an acceptable hypersurface can lie, are excluded by condition 2. in the definition. We also avoid ridges on the polynomial $q(\mathbf{x})$ between $\mathbf{p}$ and the intersection of the normal field of $\mathbf{p}$ and $q(\mathbf{x}) = 0$, since the gradient is nonvanishing. If there are more than one alternative, we take the closest. We do not accept a branch of a hypersurface as an approximative implicitization if there is another branch of the algebraic hypersurface that is closer with respect to normal distance to $\mathbf{p}$ at some point.

4.8 Selecting an Approximative Implicitization

If the algebraic degree is sufficiently high, we can find the exact algebraic representation of any curve and surface that is represented by a rational

polynomial parameterization, see e.g. [Bajaj:93]. However, when we deal with R-positive manifolds, there is in general internal discontinuities that prohibit a representation by a single algebraic hypersurface. Generally we use algebraic representations of low degree for efficiency and stability reasons. Thus, approximation methods that give a "good" enough solution is of great interest. By "good" solutions we here mean hypersurfaces that do not have singularities in the area of interest and are within a given tolerance. It would be desirable to find approximations that satisfy the conditions for admissible hypersurfaces given in the previous section.

The approach employed to find a "good" approximation described in this section has the following steps:

1. Select an approximative null-space for the matrix $\mathbf{D}$, given in (4.1). This can be done by either performing singular value decomposition of $\mathbf{D}$ and selecting a number of coefficient vectors belonging to the smallest singular values, or use a direct search for an approximative null-space, then describe this by orthonormal coefficient vectors.

2. By using a property function $\omega(\mathbf{b})$ on the different orthonormal coefficient vectors, assign property values to the coefficient vectors.

3. Find a coefficient vector $\mathbf{b}'$ with $\|\mathbf{b}'\| = 1$ that is a linear combination of the orthonogal vectors spanning the approximative null space and with a maximal value of ω, the property function.

This approach is inspired by the error given in equation (4.26)

$$\rho(\mathbf{s}) = \frac{\mathbf{q}(\mathbf{p}(s))}{\bigtriangledown\mathbf{q}(\mathbf{p}(\mathbf{s}))-\theta\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})}.$$

Here $\mathbf{g}(\mathbf{s})$ is the direction for error measurement. By maximizing the value of

$$\min_{(\mathbf{s},\theta)\in\Omega\times[-\epsilon,\epsilon]} |\bigtriangledown\mathbf{q}(\mathbf{p}(\mathbf{s}))-\theta\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|$$

for some chosen $\epsilon > 0$, we tend to reduce the total error. However, the error problem is easier facilitated by maximizing the integral of the expression above in the region in question. To measure the behavior of the denominator in this expression, the following integral can be used

$$\int_{\Omega} \int_{-\epsilon}^{\epsilon} (\bigtriangledown\mathbf{q}(\mathbf{p}(\mathbf{s}))-\theta\mathbf{g}(\mathbf{s}))) \cdot\mathbf{g}(\mathbf{s})d\theta d\mathbf{s}. \tag{4.32}$$

This integral has some attractive properties:

91

- If $\nabla\mathbf{q}(\mathbf{p}(\mathbf{s}))-\theta\mathbf{g}(\mathbf{s})$)·$\mathbf{g}(\mathbf{s})$ change sign in the region of interest, the sign change tends to reduce the absolute value of the integral. As ridges and sinks introduce a near vanishing gradient, maximizing (4.32) tends to minimize the existence of near vanishing gradients.

- If $\nabla\mathbf{q}(\mathbf{p}(\mathbf{s}))-\theta\mathbf{g}(\mathbf{s})$)·$\mathbf{g}(\mathbf{s})$ has no sign changes, then this tends to increase the absolute value of the integral.

- If $\nabla\mathbf{q}(\mathbf{p}(\mathbf{s}))-\theta\mathbf{g}(\mathbf{s})$) and $\mathbf{g}(\mathbf{s})$ are nearly parallel, the contribution to the total value of the integral is significant. Thus, the maximization favors solutions where the gradient of q and the direction for error measurement are parallel.

To evaluate these integrals exactly (if possible) is resource consuming. Thus it is practical to employ numeric integration when the property functions are defined.

The first step is to define the features we require in the property functions.

Definition 65 (Gradient property function) *Let* $\omega : \mathbb{R}^{\tilde{M}} \longrightarrow \mathbb{R}$ *satisfy*

$$
\begin{aligned}
\omega(\mathbf{a}+\mathbf{b}) &= \omega(\mathbf{a})+\omega(\mathbf{b}) \\
\omega(c\,\mathbf{a}) &= c\,\omega(\mathbf{a})
\end{aligned}
$$

then ω *is a gradient property function.*

We make in the lemma following all possible combinations of r orthonormal vectors $\mathbf{b}_i$, $i = 1,\ldots,r$ by the function

$$
\mathbf{b}(\boldsymbol{\alpha}) = \frac{\displaystyle\sum_{j=1}^{r} \alpha_j \mathbf{b}_j}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}}
\tag{4.33}
$$

with

$$
\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_r) \neq \mathbf{0}.
$$

Lemma 66 *Let* $\mathbf{b}(\boldsymbol{\alpha})$ *be defined as in (4.33) then*

$$
\|\mathbf{b}(\boldsymbol{\alpha})\|_2 = 1.
$$

Proof.

$$\mathbf{b}(\boldsymbol{\alpha}) \cdot \mathbf{b}(\boldsymbol{\alpha}) = \frac{\displaystyle\sum_{j=1}^{r} \alpha_j \mathbf{b}_j}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}} \cdot \frac{\displaystyle\sum_{j=1}^{r} \alpha_j \mathbf{b}_j}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}} = \frac{\displaystyle\sum_{j=1}^{r} \alpha_j^2}{\displaystyle\sum_{j=1}^{r} \alpha_j^2} = 1.$$

∎

Now we find the maximum of the gradient property function for a selection of orthonormal coefficient vectors.

Lemma 67 *Let let* $\mathbf{b}(\alpha)$ *be defined as in (4.33), let* ω *be a gradient property function, let the vectors* $\mathbf{b}_1, \ldots, \mathbf{b}_r$ *be orthonormal and assigned property values*

$$\omega_i = \omega(\mathbf{b}_i), \ i = 1, \ldots, r$$

with some $\omega_j \neq 0$, $j = 1, \ldots, r$. *Then*

$$\max_{\boldsymbol{\alpha} \neq \mathbf{0}} |\omega(\mathbf{b}(\boldsymbol{\alpha}))| = |\omega(\mathbf{b}(\boldsymbol{\alpha}'))|,$$

where

$$\boldsymbol{\alpha}' = (\alpha_1', \ldots, \alpha_r') = \frac{(\omega_1, \ldots, \omega_r)}{\sqrt{\displaystyle\sum_{j=1}^{r} \omega_j^2}}. \tag{4.34}$$

Proof. By definition 65 we have

$$\omega(\mathbf{b}(\boldsymbol{\alpha})) = \omega\left(\frac{\displaystyle\sum_{j=1}^{r} \alpha_j \mathbf{b}_j}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}}\right) = \frac{\displaystyle\sum_{j=1}^{r} \alpha_j \omega(\mathbf{b}_j)}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}} = \frac{\displaystyle\sum_{j=1}^{r} \alpha_j \omega_j}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}}.$$

The maximum value of $\omega(\mathbf{b}(\boldsymbol{\alpha}))$ has to satisfy

$$0 = \frac{\partial \omega(\mathbf{b}(\boldsymbol{\alpha}))}{\partial \alpha_i} = \frac{\omega_i \sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2} - \displaystyle\sum_{j=1}^{r} \alpha_j \omega_j \frac{\alpha_i}{\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}}}{\left(\sqrt{\displaystyle\sum_{j=1}^{r} \alpha_j^2}\right)^2} = \frac{\omega_i \displaystyle\sum_{j=1}^{r} \alpha_j^2 - \alpha_i \displaystyle\sum_{j=1}^{r} \alpha_j \omega_j}{\left(\displaystyle\sum_{j=1}^{r} \omega_j^2\right)^{5/2}}.$$

The values in (4.34) satisfy this requirement by setting $\boldsymbol{\alpha} = \boldsymbol{\alpha}'$ in the equation above. Further we have

$$
\omega(\mathbf{b}(\boldsymbol{\alpha}')) = \frac{\sum\limits_{j=1}^{r} \alpha'_j \omega(\mathbf{b}_j)}{\sqrt{\sum\limits_{j=1}^{r} \left(\alpha'_j\right)^2}}
$$

$$
= \sum\limits_{j=1}^{r} \alpha'_j \omega_j \frac{1}{\sqrt{\sum\limits_{j=1}^{r} \left(\alpha'_j\right)^2}} = \sqrt{\sum\limits_{j=1}^{r} \omega_j^2}.
$$

Removing the components of $\mathbf{b}(\boldsymbol{\alpha}')$ from $\mathbf{b}_1, \ldots, \mathbf{b}_r$ we get

$$
\mathbf{b}'_i = \mathbf{b}_i - \left[\mathbf{b}_i \cdot \mathbf{b}(\boldsymbol{\alpha}')\right]\mathbf{b}(\boldsymbol{\alpha}') = \mathbf{b}_i - \alpha_i \mathbf{b}(\boldsymbol{\alpha}'), \; i = 1, \ldots, r.
$$

The value of ω for choosing $\mathbf{b}'_i, \; i = 1, \ldots, r$ as arguments is

$$
\omega(\mathbf{b}'_i) = \omega(\mathbf{b}_i - \alpha_i \mathbf{b}(\boldsymbol{\alpha}')) = \omega_i - \alpha_i \omega(\mathbf{b}(\boldsymbol{\alpha}'))
$$

$$
= \omega_i - \frac{\omega_i}{\sqrt{\sum\limits_{j=1}^{r} \omega_j^2}} \sqrt{\sum\limits_{j=1}^{r} \omega_j^2} = 0.
$$

We can thus make $r - 1$ orthonogal coefficient vectors $\mathbf{b}''_i, \; i = 1, \ldots, r - 1$ from $\mathbf{b}'_i, \; i = 1, \ldots, r, \;$, with

$$
\omega''_j = \omega(\mathbf{b}''_j) = 0, \; j = 1, \ldots, r - 1.
$$

Now adding $\mathbf{b}''_r = \mathbf{b}(\boldsymbol{\alpha}')$ we can make any combination $\mathbf{b}''(\boldsymbol{\alpha})$ of the type (4.33) based on $\mathbf{b}''_i, \; i = 1, \ldots, r$. Now

$$
|\omega(\mathbf{b}''(\boldsymbol{\alpha}))| = \left| \frac{\sum\limits_{i=1}^{r} \alpha_i \omega''_i}{\sqrt{\sum\limits_{i=1}^{r} \alpha_i^2}} \right| = \left| \frac{\alpha_r \omega''_r}{\sqrt{\sum\limits_{i=1}^{r} \alpha_i^2}} \right| \leq |\omega(\mathbf{b}(\boldsymbol{\alpha}'))|.
$$

Thus, the choice of $\boldsymbol{\alpha} = \boldsymbol{\alpha}'$ in (4.34) gives the gradient property function an maximal absolute value.

$\blacksquare$

The next step is to show that integration of $\bigtriangledown\mathbf{q_b}(\mathbf{p(s)}) - \theta\mathbf{g(s)} \;)\cdot\mathbf{g(s)}$ satisfies the requirements of a gradient property function, and thus can be used to find a "good" solution.

Lemma 68 *Let the function* $\omega : \mathbb{R}^{\tilde{M}} \longrightarrow \mathbb{R}$ *be defined by*

$$\omega(\mathbf{b}) = \int_{\Omega} \int_{-\epsilon}^{\epsilon} \left(\nabla \mathbf{q_b}(\mathbf{p}(\mathbf{s})) - \theta \mathbf{g}(\mathbf{s}) \right) \cdot \mathbf{g}(\mathbf{s}) d\theta d\mathbf{s}.$$

Here $q_{\mathbf{b}}(\mathbf{x})$ *is an algebraic hypersurface with coefficients*

$$\mathbf{b} = (b_1, \ldots, b_{\tilde{M}})^T.$$

Then $\omega(\mathbf{b})$ *satisfies:*

$$\begin{aligned} \omega(\mathbf{a} + \mathbf{b}) &= \omega(\mathbf{a}) + \omega(\mathbf{b}) \\ \omega(c \, \mathbf{a}) &= c \, \omega(\mathbf{a}). \end{aligned}$$

Proof. *Using the definition of the algebraic hypersurface in (3.3), we get*

$$\begin{aligned} \omega(\mathbf{a} + \mathbf{b}) &= \int_{\Omega} \int_{-\epsilon}^{\epsilon} \nabla \left(q_{\mathbf{a}+\mathbf{b}}(\mathbf{p}(\mathbf{s}) - \theta \mathbf{g}(\mathbf{s})) \right) \cdot \mathbf{g}(\mathbf{s}) d\theta d\mathbf{s} \\ &= \sum_{\mathbf{i} \in I(m)} (a_{\mathbf{i}} + b_{\mathbf{i}}) \frac{m!}{\mathbf{i}!} \int_{\Omega} \int_{-\epsilon}^{\epsilon} \left(\nabla \left(\beta^{\mathbf{i}} \right) (\mathbf{p}(\mathbf{s}) - \theta \mathbf{g}(\mathbf{s})) \right) \cdot \mathbf{g}(\mathbf{s}) d\theta d\mathbf{s} \\ &= \omega(\mathbf{a}) + \omega(\mathbf{b}). \end{aligned}$$

Further

$$\begin{aligned} \omega(c \, \mathbf{a}) &= \int_{\Omega} \int_{-\epsilon}^{\epsilon} \left(\nabla \mathbf{q}_{c \, \mathbf{a}}(\mathbf{p}(\mathbf{s}) - \theta \mathbf{g}(\mathbf{s})) \right) \cdot \mathbf{g}(\mathbf{s}) d\theta d\mathbf{s} \\ &= \sum_{\mathbf{i} \in I(m)} c \, a_{\mathbf{i}} \frac{m!}{\mathbf{i}!} \int_{\Omega} \int_{-\epsilon}^{\epsilon} \left(\nabla \left(\beta^{\mathbf{i}} \right) (\mathbf{p}(\mathbf{s}) - \theta \mathbf{g}(\mathbf{s})) \right) \cdot \mathbf{g}(\mathbf{s}) d\theta d\mathbf{s} \\ &= c \, \omega(\mathbf{a}). \end{aligned}$$

$\blacksquare$

The following theorem connects the results in lemma 68 and lemma 67 to approximative implicitization, and establish a method for choosing a "good" coefficient vector from an approximative null space of the matrix $\mathbf{D}$.

Theorem 69 *Let the gradient property function of the combination of an algebraic hypersurface combined with a properly parameterized C^1-continuous manifold* $\mathbf{p}(\mathbf{s})$ *be defined by*

$$\omega(\mathbf{b}) = \int_{\Omega} \int_{-\epsilon}^{\epsilon} \left(\nabla \mathbf{q_b}(\mathbf{p}(\mathbf{s}) - \theta \mathbf{g}(\mathbf{s})) \right) \cdot \mathbf{g}(\mathbf{s}) d\theta d\mathbf{s}, \tag{4.35}$$

and let $\mathbf{b}_1, \ldots, \mathbf{b}_r$ *satisfying* $\mathbf{b}_i \cdot \mathbf{b}_j = \delta_{i,j}$, $i, j = 1, \ldots, r$, *be* r *approximations to the manifold. Then the hypersurface with coefficient vector*

$$\mathbf{b}' = \sum_{i=1}^{r} \frac{\omega_i \mathbf{b}_i}{\sqrt{\sum_{j=1}^{r} \omega_j^2}} \tag{4.36}$$

satisfies

$$\max_{\boldsymbol{\alpha} \neq 0} |\omega(\mathbf{b}(\boldsymbol{\alpha}))| = \omega(\mathbf{b}'),$$

where $\mathbf{b}(\boldsymbol{\alpha})$ *given in (4.33) is*

$$\mathbf{b}(\boldsymbol{\alpha}) = \frac{\sum_{j=1}^{r} \alpha_j \mathbf{b}_j}{\sqrt{\sum_{j=1}^{r} \alpha_j^2}}.$$

Proof. Combining lemma (68) and lemma (67) proves the theorem. ∎

Remark 20 *Any numeric approximation* $\widetilde{\omega}(\mathbf{b})$ *to (4.35) satisfying*

$$\begin{aligned} \widetilde{\omega}(\mathbf{a} + \mathbf{b}) &= \widetilde{\omega}(\mathbf{a}) + \widetilde{\omega}(\mathbf{b}) \\ \widetilde{\omega}(c\,\mathbf{a}) &= c\,\widetilde{\omega}(\mathbf{a}) \end{aligned}$$

is a gradient property function and can be used for choosing a "good" coefficient vectors from an approximative null space of $\mathbf{D}$.

4.9 Algorithm Approximative Implicitization

In the previous sections we have addressed different aspects of the approximative implicitization algorithm sketched in the start of the chapter. We now give a more detailed version of this algorithm related to the findings in these sections.

Algorithm Approximative Implicitization

1. Choose algebraic degree m of the approximative implicitization.

2. Choose barycentric coordinate system to be used in the approximation such that $\mathbf{p}(\mathbf{s})$ is inside the simplex defining the barycentric coordinate system.

3. Build the **D** matrix combining the algebraic hypersurface and the R-positive manifold, by repeated use of a multiplication algorithm satisfying the requirements to a PosProd Algorithm in definition 25.

4. Find an approximative null space of **D** by either:

   (a) Singular value decomposition, see e.g. [Numres:88] or [Stewart:73].

   (b) Direct search for the null space, see Section 4.4.5.

   (c) Other methods for finding approximative null spaces.

5. Pick the coefficient vector from the approximative null space that either have the maximal absolute value of

$$\omega(\mathbf{b}) = \int\limits_{\Omega} \int\limits_{-\epsilon}^{\epsilon} \bigtriangledown \mathbf{q_b}(\mathbf{p(s)} - \theta\mathbf{g(s)}) \cdot \mathbf{g(s)} d\theta d\mathbf{s},$$

   or a numeric approximation thereof, see theorem 69 and remark 20.

6. Check if the hypersurface is admissible, see definition 63. If not admissible increase the number of vectors in the approximative null-space and go to 5.

7. Check if the hypersurface is within an acceptable distance from **p(s)**, e.g. using (4.26).

4.10 Approximative Implicitization of a Number of Manifolds

There are a set of applications for a simultaneous approximation of more than one manifold. Among these are:

- Controlling the shape of the algebraic approximation in a region larger than covered by one manifold.

- Model an algebraic hypersurface approximating or interpolating a set of manifolds. E.g. in $\mathbb{R}^3$ a mesh of curves can be approximated by such a technique.

- Make a rough approximation of a set of manifolds to simplify a complex structure of manifolds.

- Combine a manifold with additional problem dependent manifolds to direct the hypersurface to behave in a specific way. E.g. approximate the boundary and a translate of the boundary to make a hypersurface containing the boundary but being near normal to the manifold along the boundary. The procedure applies to a segment of the boundary as well.

The dimension of the manifolds approximated is insignificant in the theorem following. They can touch in a smooth way or they can be separated by some distance. Some of the manifolds can be the target of the approximation. Other manifolds can be used to control the behavior of the approximation in a given region. The following theorem shows that we can combine the $\mathbf{D}$ matrices of the different manifolds to build a matrix for approximating a number of manifolds.

Theorem 70 *Let $\mathbf{p}_i(\mathbf{s}_i)$, $i = 1, \ldots, r$ be R-positive manifolds in $\mathbb{R}^l$ of respectively dimension g_i, $i = 1, \ldots, r$ with a regular parametrization $\mathbf{s}_i \in \Omega_i$. The combination of these manifolds with an algebraic hypersurface q of total degree m satisfies*

$$\sum_{i=1}^{r} \left(q(\mathbf{p}_i(\mathbf{s}_i)) \right)^2 \leq \left\| \begin{pmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_r \end{pmatrix} \mathbf{b} \right\|_2^2, \tag{4.37}$$

where $q(\mathbf{p}_i(\mathbf{s}_i)) = (\mathbf{D}_i \mathbf{b})^T \boldsymbol{\alpha}_i(\mathbf{s}_i)$, $i = 1, \ldots, r$.

Proof. Remembering (4.3) we get:

$$
\begin{aligned}
q(\mathbf{p}_i(s))^2 &= \sum_{i=1}^{r} ((\mathbf{D}_i \mathbf{b})^T \boldsymbol{\alpha}_i(\mathbf{s}_i))^2 \\
&\leq \sum_{i=1}^{r} \|\mathbf{D}_i \mathbf{b}\|_2^2 \\
&= \sum_{i=1}^{r} \mathbf{b}^T \mathbf{D}_i^T \mathbf{D}_i \mathbf{b} \\
&= \mathbf{b}^T \begin{pmatrix} \mathbf{D}_1^T & \cdots & \mathbf{D}_r^T \end{pmatrix} \begin{pmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_r \end{pmatrix} \mathbf{b} \\
&= \left\| \begin{pmatrix} \mathbf{D}_1 \\ \vdots \\ \mathbf{D}_r \end{pmatrix} \mathbf{b} \right\|_2^2 .
\end{aligned}
$$

The result is in correspondence with (4.3) when we break a NURBS represented manifold into rational Bernstein basis represented manifolds by introducing the Bernstein knot vector. However, the result is more general, because the manifolds can be of different dimension with no correspondence of the parametrizations.

Remark 21 *We can try to find a hypersurface separating two manifolds by building a matrix of the type (4.37) and to use this to make an approximative implicitization. The result will be a hypersurface that approximates both the manifolds. If the number of degrees of freedom is not large enough to do this, the result is a near singular region close to the manifolds. If the algebraic degree is high enough, the hypersurface interpolates both manifolds. Thus, the resulting hypersurface is not suited for the separation of the two manifolds. In the next section we give a more convenient approach for separation of manifolds by a hypersurface.*

4.11 Separating Manifolds Using Approximative Implicitization

Let $\mathbf{p}_1(\mathbf{s}_1)$ and $\mathbf{p}_2(\mathbf{s}_2)$ be two R-positive manifolds of possibly different dimensions. Let $\mathbf{b}$ be the coefficient vector of a hypersurface $q(\mathbf{x})=0$ in $\mathbb{R}^l$ that is admissible with respect to either manifolds. The combinations of the bounded manifolds and the hypersurface are

$$\begin{aligned}
q(\mathbf{p}_1(\mathbf{s}_1)) &= (\mathbf{D}_1\mathbf{b})^T\boldsymbol{\alpha}_1(\mathbf{s}_1) = \mathbf{f}_1\boldsymbol{\alpha}_1(\mathbf{s}_1) \\
q(\mathbf{p}_2(\mathbf{s}_2)) &= (\mathbf{D}_2\mathbf{b})^T\boldsymbol{\alpha}_2(\mathbf{s}_2) = \mathbf{f}_2\boldsymbol{\alpha}_2(\mathbf{s}_2).
\end{aligned}$$

If either

$$(\mathbf{f}_1)_i < a,\ i = 1, \ldots, \tilde{N}_1 \quad \wedge \quad (\mathbf{f}_2)_i > a,\ i = 1, \ldots, \tilde{N}_2,$$

or

$$(\mathbf{f}_1)_i > a,\ i = 1, \ldots, \tilde{N}_1 \quad \wedge \quad (\mathbf{f}_2)_i < a,\ i = 1, \ldots, \tilde{N}_2,$$

then the two bounded manifolds $\mathbf{p}_1(\mathbf{s}_1)$ and $\mathbf{p}_2(\mathbf{s}_2)$ do not intersect. When we can establish separation between two manifolds the next question is how far the manifolds are apart. In (4.26) we established that the distance from a point on a manifold with regular parametrization to the algebraic hypersurface can be expressed by

$$\rho(\mathbf{s}) = \frac{q(\mathbf{p}(\mathbf{s}))}{\bigtriangledown q(\mathbf{p}(\mathbf{s}) - \theta(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})},$$
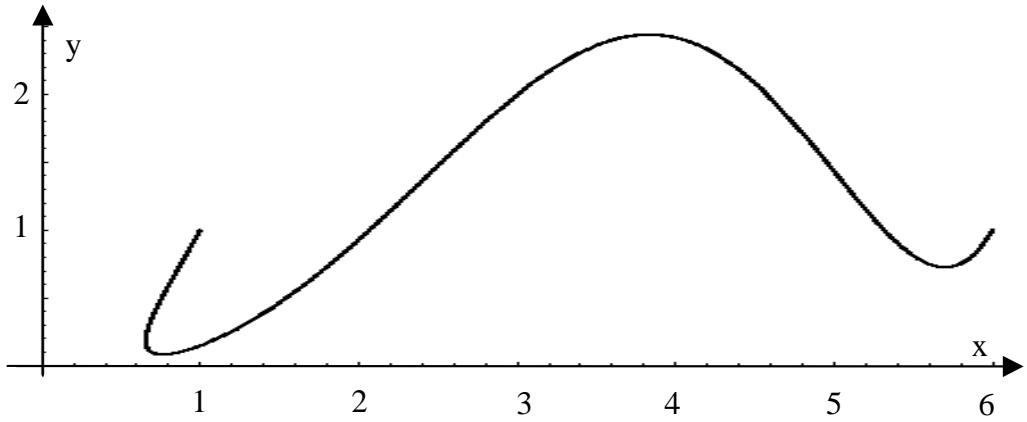
Figure 4.2: The Bezier curve used as example for approximation with 5, 6 and 7 degree algebraic curves. The vertices of this curve are $(1,1)$, $(0,-1)$, $(1,1)$, $(2,-1)$, $(3,2)$, $(4,7)$, $(5,-1)$, and $(6,1)$.

where $\mid \theta(\mathbf{s}) \mid \leq \mid \rho(\mathbf{s}) \mid$ and $\theta(\mathbf{s})\rho(\mathbf{s}) \geq 0$. We now want to find the intervals limiting the distance $[v_{1,\min}, v_{1,\max}]$ from the algebraic hypersurface to the manifold $\mathbf{p}_1(\mathbf{s}_1)$

$$
v_{1,\min} = \min_{\mathbf{s}_1 \in \Omega \wedge |\theta| \leq \rho(\mathbf{s}_1)} \frac{q(\mathbf{p}_1(\mathbf{s}_1))}{\bigtriangledown q(\mathbf{p}_1(\mathbf{s}_1) - \theta(\mathbf{s}_1)\mathbf{g}_1(\mathbf{s}_1)) \cdot \mathbf{g}_1(\mathbf{s}_1)}
$$
$$
v_{1,\max} = \max_{\mathbf{s}_1 \in \Omega \wedge |\theta| \leq \rho(\mathbf{s}_1)} \frac{q(\mathbf{p}(\mathbf{s}))}{\bigtriangledown q(\mathbf{p}_1(\mathbf{s}_1) - \theta(\mathbf{s}_1)\mathbf{g}_1(\mathbf{s}_1)) \cdot \mathbf{g}_1(\mathbf{s}_1)}.
$$

The interval $[v_{2,\min}, v_{2,\max}]$ related to the distance from the hypersurface to $\mathbf{p}_2(\mathbf{s}_2)$, is found in a similar way. These intervals can then be used as an estimate of the distance between the manifolds.

Candidates for choosing surfaces for testing separation are hypersurfaces approximating one of the manifolds. If we know that certain regions are close, we can force the approximation to be very accurate in these regions by adding constraints.

Remark 22 *This approach for separation of surfaces has been employed in the SISL library when boxing techniques are not adequate. This is especially interesting if the surfaces are nearly coincident. It is then possible to determine whether there is a region of coincidence or not.*

4.12 Examples of Approximative Implicitization

In this section we deal with some examples on approximative implicitization. The curve we have selected to use as an example is shown in figure 4.2. This is a degree seven Bezier curve with vertices:

$$(1,1),\ (0,-1),\ (1,1),\ (2,-1),\ (3,2),\ (4,7),\ (5,-1),\ (6,1).$$

Since we can find straight lines that intersect this curve four times, the algebraic degree of the curve is at least 4, thus to try with approximations with lower algebraic degree than four, do not succeed. However, when trying with a degree 4 approximative implicitization, the result was an approximation with asymptotes, and thus no curve that is admissable was found. However, degrees 5, 6 and 7 give admissable approximations. Approximation of subparts of the curve with degree 4 and lower can also succeed when the intervals are sufficiently small. In the examples we have imposed interpolation of start and end of the Bezier curve, e.g. that the point $(1,1)$ and $(6,1)$ are interpolated. The barycentric coordinate system used when building the **D** matrix have the corners $(-3,-1)$, $(10,-1)$ and $(2,10)$. For finding the singular values of **D** the standard singular value decomposition in Mathematica has been used. Mathematica has also been used for generating the plots in this section. The direct elimination method in Section 4.4.5 is slower and in general gives similar approximations, but with somewhat larger errors.

Remembering that the number of coefficients in the algebraic curve of degree m is $\frac{(m+1)(m+2)}{2}$. Thus, **D** has the dimension

$$(7m+1) \times \frac{(m+1)(m+2)}{2}.$$

When imposing interpolation of start and end of $\mathbf{p}(s)$, two degrees of freedom are used. Thus, the resulting dimension of **D** is respectively 36×19, 43×26 and 50×34 for the approximative implicitization of degrees 5, 6 and 7. In the examples we get respectively 19, 26 and 34 singular values.

4.12.1 Example of 5th Degree Approximation

The singular values of the **D** are shown in the following table:

1		17	18	19
1.12	. . .	0.000831	0.000280	0.0000396
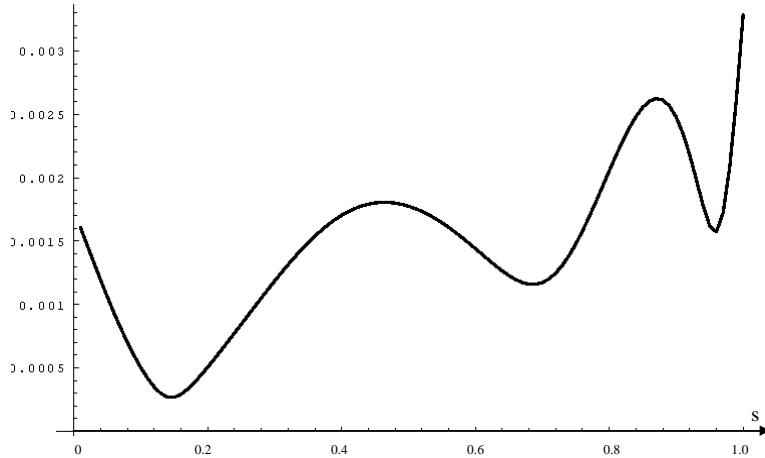
Figure 4.3: A plot for the fifth degree approximative implicitization of the term $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ along the curve approximated, when choosing the coefficients corresponding to the smallest singular value of $\mathbf{D}$ to define the approximative implicitization. $\mathbf{n}(s)$ is the function describing the normals to $\mathbf{p}(s)$. If the expression $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ is zero for some value of $s \in [0,1]$, then the approximation is not admissable, either because the gradient of the approximation is perpendicular to the normal of $\mathbf{p}(s)$ or because $\nabla q(\mathbf{p}(s)) \cdot 0$ giving a singular ornear singular approximation. The plot shows that the approximation is well behaved.

Note that we have a gradual decrease from left to right. Further there are no singular values close to the relative accuracy of 16 digits used in the calculation. When choosing the coefficients corresponding to the smallest singular value of $\mathbf{D}$ to define the approximative implicitization, we get an admissable approximation to $\mathbf{p}(s)$. A plot of the expression

$$\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$$

along $\mathbf{p}(s)$ is shown in figure 4.3. The plot indicates that the approximation is well behaved although the expression have many oscillations. Figure 4.4 shows the error of this approximation. Note that the error is largest in the interval where $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ has the smallest values.

In figure 4.5 we show a contour plot of the 5th degree approximative implicitization to $\mathbf{p}(s)$.
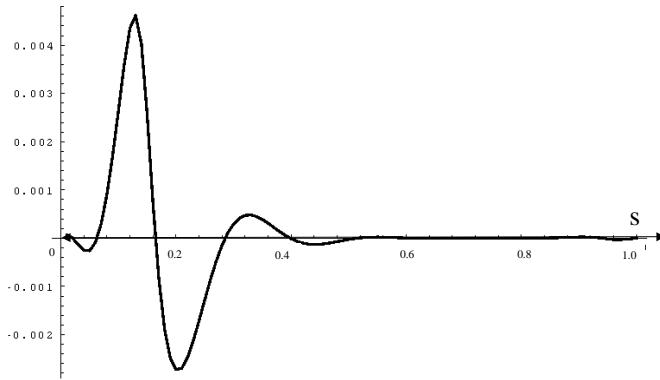
Figure 4.4: The error of the approximative implicitization of degree 5 to $\mathbf{p}(s)$. Note that the error is large where the expression $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ plotted in the previous figure is small.
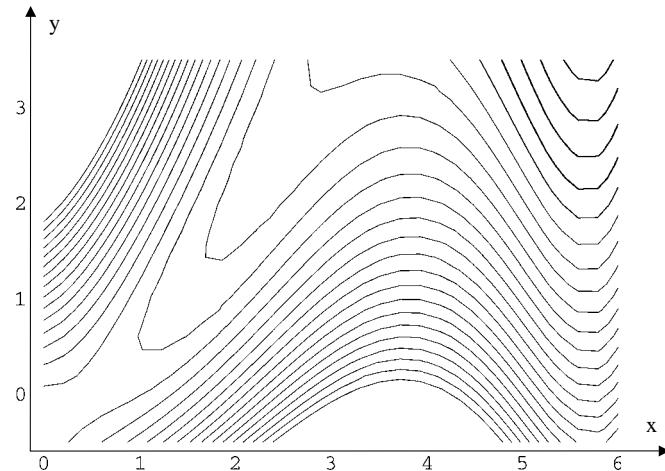
Figure 4.5: Contour plot of the 5th degree approximative implicitization to $\mathbf{p}(s)$. By identifying the point $(1, 1)$ in the plot and following the adjacent curves to the point $(6, 1)$, we recognize the shape of $\mathbf{p}(s)$. The plot shows that regions where the expression $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s) = 0$ are not too close to $\mathbf{p}(s)$.

103
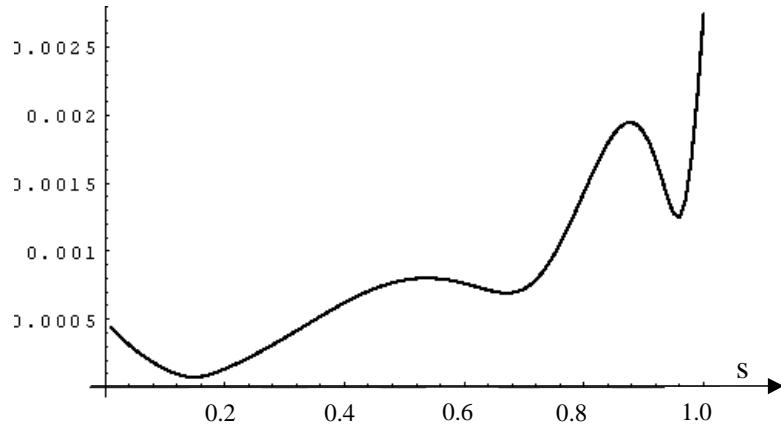
Figure 4.6: A plot of the term $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ along the curve $\mathbf{p}(s)$ for the approximative implicitization of 6th degree. The approximative implicitization corresponds to the smallest singular value.

4.12.2 Example of 6th Degree Approximation

The singular values of $\mathbf{D}$ are shown in the following table:

1		24	25	26
1.23	...	1.147×10^{-5}	5.82×10^{-6}	$4.246464737776847 \times 10^{-7}$

Note that we have a gradual decrease from left to right. Further there are no singular values close to the relative accuracy of 16 digits used in the calculation. Once more we choose the approximations belonging to the smallest singular values. The behavior of $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ is shown in figure 4.6. The error function is shown in figure 4.7, and the contour plot in figure 4.8.

4.12.3 Example of 7th Degree Approximation

The singular values of $\mathbf{D}$ when the algebraic degree is chosen to be 7 are shown in table

1		32	33	34
1.30	...	1.09×10^{-7}	4.23×10^{-15}	2.92×10^{-17}

Note that we have a significant jump in singular values between value 32 and value 33, thus indicating that a very close fit to the exact algebraic representation is found. We have chosen to make an optimal solution of the approximations belonging to the two smallest singular values according to
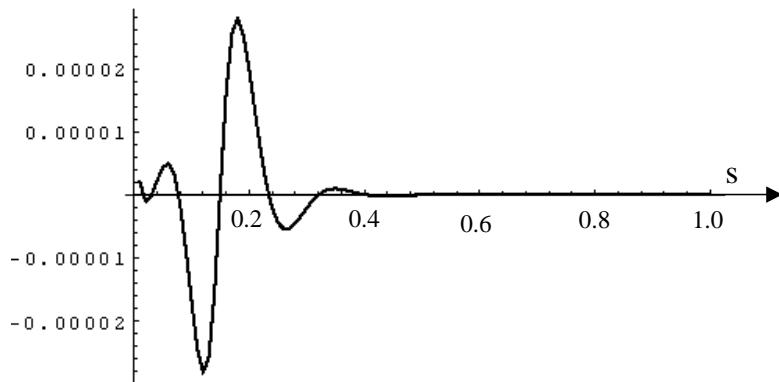
Figure 4.7: Plot of error function for the **6**th degree approximative implicit-ization of $\mathbf{p}(s)$.
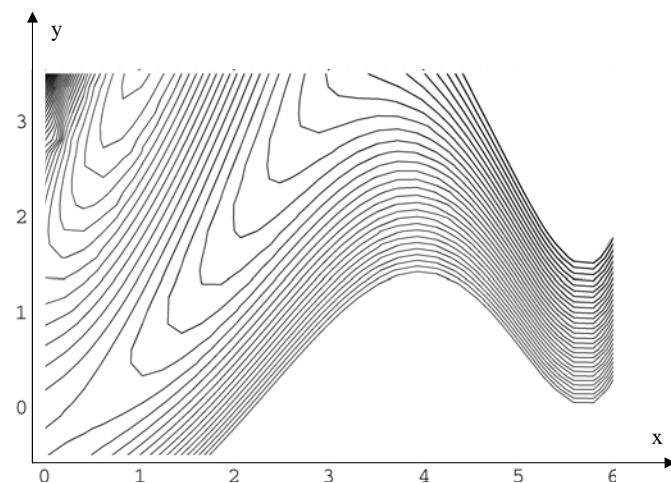
Figure 4.8: Contour plot of the 6th degree approximative implicitization of $\mathbf{p}(s)$. By identifying the point $(1, 1)$ in the plot and following the adjacent curves to the point $(6, 1)$ we recognize the shape of $\mathbf{p}(s)$. The plot shows that regions where the expression $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s) = 0$, regions where either the curve normal is orthogonal to the gradient or where the gradient vanish are not too close to $\mathbf{p}(s)$.
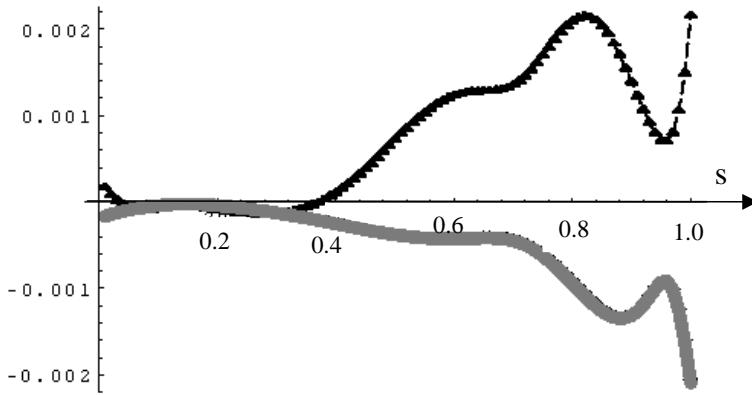
Figure 4.9: A plot of the term $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ along the curve approximated when choosing the coefficients corresponding to the two smallest singular values of $\mathbf{D}$ to define approximative implicitization of degree 7 to $\mathbf{p}(s)$. We see that one of the approximations has value zero of $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ at two places and is thus not an admissable solution. We have combined the two approximations to try to find an even better solution, however, the differences between the acceptable solution and the combined solution do not show on this plot, they are both represented by the gray curve.

method in section 4.8 by approximating the integral of $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ by sampling 100 points on the curve and evaluating the sum of these values. The plots of the approximation of $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s)$ for both solutions as well as their "optimal" combination is shown in figure 4.9. The error function is shown in figure 4.10, and the contour plot is shown in figure 4.11.

106
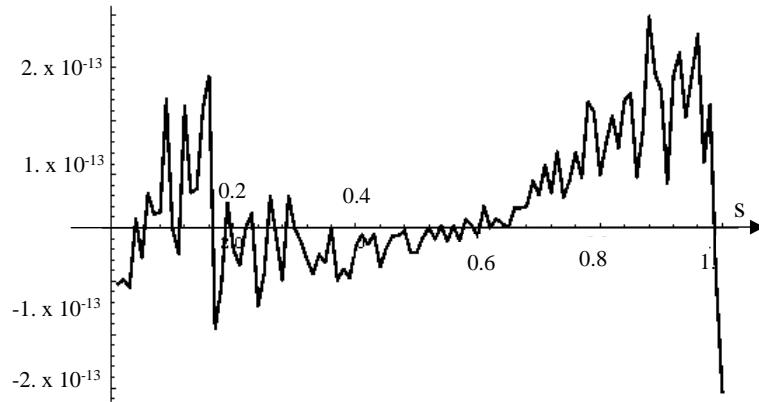
Figure 4.10: The error of the approximation with a 7 degree algebraic curve to $\mathbf{p}(s)$ is varying-between -1.54×10^{-13} and 1.98×10^{-13} and is thus very close to the noise level of the digital representation with 16 digits accuracy used in the example. Because of this the plot has a ragged nature and the first and last points, which should be exactly at the zero line, are also affected with noise.
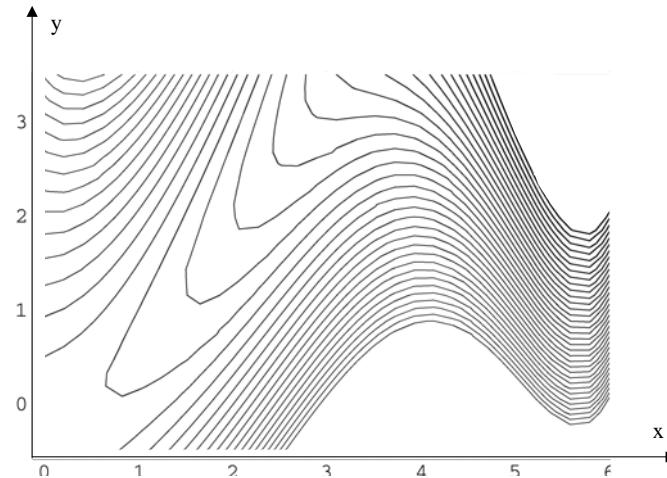
Figure 4.11: Contour plot of the 7th degree approximative implicitization of $\mathbf{p}(s)$. By identifying the point $(1, 1)$ in the plot and following the adjacent curves to the point $(6, 1)$ we recognize the shape of $\mathbf{p}(s)$. The plot shows that regions where the expression $\nabla q(\mathbf{p}(s)) \cdot \mathbf{n}(s) = 0$, e.g. regions where either the curve normal is orthogonal to the gradient or where the gradient vanish are not too close to $\mathbf{p}(s)$.

107

Chapter 5

Approximation of Intersection Results

A major challenge in the implementation of intersection algorithms is to represent the intersections found in a proper way. In 3D CAD-systems boundary structures are used to connect surfaces that describe the outer shell (and possibly inner shells) of a volume. When making the union or intersection of two volumes, the first step is to intersect the shells of the volumes. The intersection of the shells is performed by intersecting the surfaces composing the shells. The curves from the surface intersections are then assembled in a proper way to trim away pieces of the surfaces no longer needed. Then new shells are built to describe the resulting volume object(s).

The results of a surface surface intersection are points, curves and regions of surfaces. The surface regions can be represented by a curve describing the boundary of the intersection region. The methods for the representation of intersection curves are to a great extent based on piecewise cubic Hermite interpolation with GC^1 or GC^2 continuity between adjacent curve segments. Cubic Hermite interpolation is both used for 2D curves in the parameter domain of the surfaces and for 3D curves. The most common intersection representations in CAD-systems are addressed in Section 5.1.

When intersecting manifolds of higher dimension than two, we can expect to get intersection results that must be described by manifolds of dimension 2 or higher. A first choice is to resort to cubic Hermite interpolation defined over simplical domains of dimension $d \geq 2$. However, the discussion in Section 5.2 shows that a piecewise local cubic Hermite interpolation over simplical domains of dimension $d \geq 2$ only has GC^0 continuity between adjacent polynomial pieces. Thus, either higher degree Hermite interpolation has to be employed or other approaches have to be found. One such approach, piecewise algebraic approximation in the parameter domain of the manifolds,

is discussed in Section 5.3. This approach can be based on the approximative implicitization, described in chapter 4.

In Section 5.4 we address methods for approximation of algebraic represented manifolds, and give a general method that has the same convergence rate as the best approximation satisfying the same interpolation constraints. To model the shape of the approximation the method makes use of the degrees of freedom remaining in the polynomial interpolant, when the interpolation constraints are satisfied. Special attention is given to Hermite approximation of 2D curves in Chapter 6. These results are further elaborated on the approximation of circle and ellipse segments in Chapter 7.

5.1 Intersection Representation within CAD

The mostly used intersections in 3D CAD-systems are:

- Curve curve intersection in $\mathbb{R}^2$ and $\mathbb{R}^3$.

- Curve surface intersection in $\mathbb{R}^3$.

- Surface surface intersection in $\mathbb{R}^3$.

Results of such intersections can be described as objects of the following categories:

- **Intersection points** (0-manifolds). These are represented as points in respectively $\mathbb{R}^2$ and $\mathbb{R}^3$, and as a points in the parameter domain of the manifolds (curves/surfaces) being intersected.

- **Intersection curves** (1-manifolds). These are traditionally approximated by parametric curves in $\mathbb{R}^3$. In the parameter domains of the surfaces being intersected, the intersection curves are also approximated by parametric curves. In certain cases exact representation of the intersection curve is present:

  - When the intersection of a curve and another manifold (curve, surface or volume) coincide over an interval of the curve.

  - When the intersection is along a segment of a constant parameter line in the parameter domain of a surface. In these cases a segment of the constant parameter line, is an exact representation. The image of the segment of the constant parameter line is an exact 3D representation of the intersection curve.

– For certain intersection constellations of surfaces the intersection curve is a first or second degree algebraic curve. In geometric modelers based on algebraic surface representation, much energy have been put into developing algorithms identifying such situations.

Geometric Hermite interpolation is the standard method for approximating intersection curves. [Sederberg:91], [Höllig:95] and [Höllig:96] can be consulted for the theoretical background of geometric Hermite approximation of intersection curves. In [Sederberg:91] it is pointed out that the algebraic degree of the intersection curve between two bicubic patches is a curve with implicit equation of degree 324. It is also pointed out that a representation of such a parametric intersection curve, as an exact parametric equation, is non-existent.

- **Intersection regions on surfaces** (2-manifolds). Intersection regions are traditionally not handled in CAD-systems, because traditional intersection algorithms only to a limited degree handle coincident regions in two surfaces. In Chapter 2 we discussed the introduction of intersection tolerances, and proposed an approach for representing intersection regions.

The cubic Hermite interpolant is often represented in a Bernstein basis, and is thus a R-positive 1-manifold, see remark 13 on page 52 in Chapter 3. Often more than one cubic Hermite segment is required to approximate an intersection curve within proper tolerances. Different strategies for determining the segmentation of the Hermite approximation exist. Two of these are:

- Bisection. Two points on the intersection curve are known. A Hermite approximation of the intersection curve is made between these points. If the approximation satisfies the accuracy and shape quality requirements the segment is accepted. However, if the approximation does not satisfy these requirements, the approximation problem is split into two subproblems by finding a point on the intersection curve lying between the two end points. This new point is used to define two subsegments. The bisection process is then performed on each of the subsegments. More information on the bisection process can be found in [Sederberg:91].

- Tracing/Marching. From a point $\mathbf{p}_0$ on the intersection curve determine a step length based on e.g. curvature estimates or a fixed step length s. Find a point on the intersection curve by first moving a distance

equal to the step length s in the direction of the curve tangent $\mathbf{t}_0$ at $\mathbf{p}_0$. Then iterate onto the intersection curve by constraining the interation to:

- A plane that has $\mathbf{t}_0$ as normal vector and is located at a distance s in the direction $\mathbf{t}_0$ from $\mathbf{p}_0$.
- A sphere with center at $\mathbf{p}_0$ and radius s.

If the iteration fails, reduce the step length until the iteration succeeds. Approximate the curve segment found. If the approximation does not satisfy accuracy and shape requirements resort to bisection. In the SISL library, see [SISL:94], such a combined tracing and bisection strategy is employed. The efficiency of the combined tracing and bisection strategy is dependent on finding good initial step lengths s to avoid bisection. However, too small initial step lengths increase the numbers of segments that are required for approximating the intersection curve within a given tolerance. A survey of tracing algorithms can be found in [Krishnan:96] where also the main problems of tracing/marching algorithms are addressed:

1. Converging back to the curve.
2. Component jumping.
3. Inability to handle singularities and multiple branches.

The approximation methods to be used, are not addressed in detail in this chapter. In the next chapter, Chapter 6, we address the approximation of curves in $\mathbb{R}^2$ and show that algebraic curves in the parameter domain of a surface can be approximated with $O(h^6)$ accuracy by controlling the tangent lengths of the cubic Hermite interpolant.

5.2 Cubic Hermite Approximation of Intersection Results

To ensure local GC^r continuity of triangular polynomial patches in $\mathbb{R}^3$, theorem 9.1 on page 410 in [Hoschek:93] states that the total polynomial degree n has to satisfy $n \geq 4r + 1$. Thus, to achieve GC^1-continuity between two patches we have to resort to polynomials of total degree 5. Thus, cubic triangular polynomial patches can only be joined with GC^0 continuity. Cubic interpolation of manifolds of higher dimensions than two defined over simplical domains, can similarly only be joined by GC^0 continuity.

Another description of these limitations is:

- In the case of cubic interpolation of curves, the number of vertices are four. Only two vertices are allocated to ensure GC^0 continuity. The two other vertices can be employed to ensure GC^1 continuity.

- For surfaces in $\mathbb{R}^3$ nine out of ten vertices are used for ensuring GC^0 continuity. Thus, only one vertex remains and these remaining degrees of freedom are too few to ensure GC^1 continuity.

- For volumes and higher dimensional manifolds, all vertices are located on the boundary. They are thus all dedicated to ensure GC^0 continuity.

5.3   Algebraic Approximation of Intersection Results

In CAGD-systems using parameterized surfaces, the main use of curves in the parameter domain of surfaces is to describe the location of intersection curves in the surfaces. The curves in the parameter domain are further used as building blocks in the process assembling trimming structures dedicated to removing parts of the surfaces. One of the interrogation functions in CAGD systems utilizing the trimming information is the function for deciding if a point is on a trimmed surface or not.

As we have mentioned earlier, curves in the parameter domain of a surface in $\mathbb{R}^3$ are in general approximated by 2D parametric piecewise polynomial curves. Such approximations require, in the general case, huge amounts of coefficients to be sufficiently accurate. It is also well known that the transfer of trimmed surfaces between systems from different CAD-vendors is a complex issue. Comments of this problem can e.g. be found in [IVF:94] appendix 4, page 6. The main reason for this is that the CAD-vendors currently (in 1996) use somewhat different strategies for solving approximation problems related to trimming, and somewhat different strategies for trying to compensate for problems resulting from the approximation of intersection curves.

We analyze in this section the consequences of introducing algebraic representation of curves in the parameter domain to perform trimming of parametric represented surfaces. The current method for representing trimming information is used in most CAD-systems, and is part of the ISO-standard ISO 10303 (STEP), "Product data representation and exchange". Thus we can not expect a new method to immediately replace the current method. However, it can be a supplement and possibly an alternative for internal

representation in the 3D CAD-systems. Another possibility is to use the algebraic representation as an intermediate stage for making better approximations to curves in the parameter domain of surfaces.

Definition 71 *Let* $\mathbf{p} : \Omega \to \mathbb{R}^3$ *be a parametric surface (2-manifold) in* $\mathbb{R}^3$ *defined over the compact domain* $\Omega \subset \mathbb{R}^2$ *and let the trimming function associated with* $\mathbf{p}$ *be denoted* τ. *Where* $\tau : \Omega \to \mathbb{R}$ *and* τ *is* C^0*–continuous.* $\mathbf{p}$ *is trimmed according to the following rules:*

$$\tau(s,t) < 0 \Longrightarrow (s,t) \quad \text{is outside the domain of the trimmed surface.}$$
$$\tau(s,t) > 0 \Longrightarrow (s,t) \quad \text{is inside the domain of the trimmed surface.}$$
$$\tau(s,t) = 0 \Longrightarrow (s,t) \quad \text{is on the boundary or inside the domain}$$
$$\text{the trimmed surface.}$$

Provided that trimming functions can be defined in an efficient and accurate way, just one evaluation of a trimming function is needed to determine if a point with known location in the parameter domain is outside a trimmed version of the surface.

Lemma 72 *Given* $\mathbf{p} : \Omega \to \mathbb{R}^3$ *a parametric surface defined over the compact domain* $\Omega \subset \mathbb{R}^2$ *with no parts trimmed away. Then any trimming function associated with* $\mathbf{p}$ *satisfy*

$$\tau(s,t) \geq 0, \ (s,t) \in \Omega.$$

Proof. Definition 71 assures this result. ■

The trimming function of a nontrimmed surface can be assigned a constant positive value.

Example 73 *In figure 5.1 we show an example of a typical trimming curve in the parameter domain of a (rational) parametric NURBS surface* $\mathbf{p}$ *of polynomial orders* (k_1, k_2). *Assuming that a trimming curve is a result of the intersection of a parametric surface of orders* (k_1, k_2) *and an algebraic surface of total degree* m, *then the trimming curve can be described by a piecewise algebraic curve of degrees* $(m(k_1 - 1), m(k_2 - 1))$ *in the parameter domain of* $\mathbf{p}$.

Thus if $m = 2$ *and* $k_1 = k_2 = 4$ *the algebraic curves are described by a tensor product B-spline function of degrees at most* $(6,6)$. *This results in a total algebraic degree of* 12 *for the trimming curves.*

From Section 4.5 table 4.4 on page 81 we remember that for $m = 3$ *the convergence rate of the algebraic approximation to a rational Bernstein basis represented parametric surface is* $O(h^5)$. *We can expect that approximating*
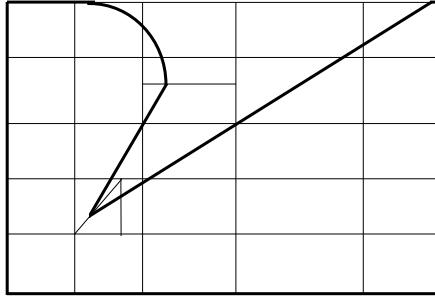
Figure 5.1: An example of a trimming curve and the segmentation of the parameter domain of the surface into a piecewise algebraic trimming function. Assuming that the surface to be trimmed is a NURBS surface, we can base the initial grid on the knot vectors of the NURBS surface and subdivide each mesh as required by the representation into triangular and rectangular parts.

a subset of a NURBS surface with degree 3 algebraic surfaces and then intersecting this approximation with another NURBS surface, results in fairly good approximations of the trace of the intersection curve. For $m = 3$ and $k_1 = k_2 = 4$ we get algebraic curves of degrees $(9, 9)$ given a total degree of 18.

Lemma 74 *Given $\mathbf{p} : \Omega \to \mathbb{R}^3$ a parametric surface defined over the compact domain $\Omega \subset \mathbb{R}^2$. Let $\mathbf{p}_{\tau_1}$ be $\mathbf{p}$ with the trimming function $\tau_1(s,t)$ assigned and let $\mathbf{p}_{\tau_2}$ be $\mathbf{p}$ with the trimming function $\tau_2(s,t)$ assigned. Then*

$$\mathbf{p}_{\tau_1} \cap \mathbf{p}_{\tau_2} \subseteq \mathbf{p}_{\tau_1\tau_2}.$$

Proof.

$$
\begin{aligned}
\mathbf{p}_{\tau_1} \cap \mathbf{p}_{\tau_2} \ &= \ \{\mathbf{p}(s,t) \mid \tau_1(s,t) \geq 0 \wedge \tau_2(s,t) \geq 0\} \\
&\subseteq \ \{\mathbf{p}(s,t) \mid \tau_1(s,t)\tau_2(s,t) \geq 0\} = \mathbf{p}_{\tau_1\tau_2}.
\end{aligned}
$$

$\blacksquare$

The consequence of the lemma is that to perform a stepwise trimming operation just by multiplication of trimming functions, is not straight forward. Phantom regions can appear in addition to a significant increase in the polynomial degree of the trimming functions. In addition singular points are

114

Classification	Inside τ_1	Partly inside τ_1	Outside τ_1
Inside τ_2	Inside	Use τ_1	Outside
Partly inside τ_2	Use τ_2	Detailed analyses	Outside
Outside τ_2	Outside	Outside	Outside

Table 5.1: Classification of the possible trimming result when two trimming functions $\tau_1(s,t)$ and $\tau_2(s,t)$ are applied to a subset of the parameter domain of a parametric surface.

introduced at points where algebraic curves from different trimming functions meet. Thus to base the trimming on products has a number of disadvantages.

Since we deal with NURBS surfaces or rational Bezier surfaces it is natural to use a piecewise polynomial description of the trimming functions over the parameter domain Ω. By doing so, we can subdivide a region to enable description of a trimming function combining two existing trimming functions. The first level of piecewise subdivision can be to use the subdivision resulting from the piecewise description of the surface. In figure 5.1 the rectangular grid can typically originate from the original piecewise description of a NURBS surface. When making the trimming functions over the subregions of Ω, we come up with the cases described in table 5.1. We see from this table that only the case where a subregion is partly inside both the trimmed surfaces described by respectively $\tau_1(s,t)$ and $\tau_2(s,t)$, a detailed analyzes has to be performed.

In this case we have basically three situations:

- The algebraic curves described by $\tau_1(s,t) = 0$ and by $\tau_2(s,t) = 0$ don't intersect in the subregion. Subdivide the subregion such that the curve $\tau_1(s,t) = 0$ is in one part, and $\tau_2(s,t) = 0$ is in another part.

- The algebraic curves described by $\tau_1(s,t) = 0$ and by $\tau_2(s,t) = 0$ intersect in the subregion in a single point. Make a subdivision through the intersection point such that the piece of $\tau_1(s,t) = 0$ to be used in the combined trimming is in one part, and the piece of $\tau_2(s,t) = 0$ to be used in the combined trimming is in another part.

- More than one intersection point exist between $\tau_1(s,t) = 0$ and $\tau_2(s,t) = 0$ in the subregion. Subdivide to separate single intersection points and possibly coincidence between the curves into separate parts.

To make the continuity requirements on the trimming functions less restrictive, by requiring that the trimming functions are C^0 inside the subregions, but to allow jumps between subregions, we get an efficient way of
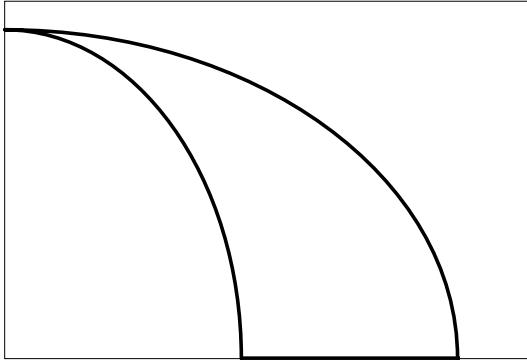
Figure 5.2: Example of a trimmed region where the trimming curves can not be separated by straight line segments. The two elliptic trimming curves intersect tangential, thus they can not be separated by straight line segments.

constructing the combined trimming function by subdividing the parent trimming functions $\tau_1(s,t)$ and $\tau_2(s,t)$. Cases exist when the trimming curves resulting from the combination of two trimming functions cannot be separated by straight line segments. One such example is shown in figure 5.2. To handle these cases, curved boundaries of trimming subregions have to be introduced.

5.4 Approximation of Algebraic Represented Manifolds

In this section we present a method for approximating manifolds that either are contained in one algebraic represented hypersurface, or in the intersection of a number of algebraic represented hypersurfaces. The manifolds used for the approximation are polynomial nonrational and R-positive, see definition 32 on page 51 in Chapter 3. The method has a number of interesting features:

- The approximation is required to satisfy a number of interpolation constraints. Interpolation constraints include but are not restricted to points and tangent directions.

- When the interpolation constraints are met, a number of degrees of freedom remains in the polynomial interpolant. The method use these

116

remaining degrees of freedom to model the shape of the approximation. This is done in such a way that the convergence rate is the same as for the best approximation method using the same polynomial basis and satisfying the same interpolation constraints.

- By doing numeric experiments with the method, the approximation rate of the best approximation with a given polynomial basis and given interpolation constraints can be found.

- The method can be used on manifolds contained in a single algebraic hypersurface, or manifolds contained in the intersection of a number of algebraic hypersurfaces.

- The method is based on finding the minimum of the square sum of the coefficients of the combination $q(\mathbf{p_v}(\mathbf{s}))$ of $\mathbf{p_v}(\mathbf{s})$, the manifold used for the approximation, and the algebraic hypersurface $q(\mathbf{x}) = 0$ being approximated. If the algebraic description of the manifold being approximated is contained in the intersection of a number of algebraic hypersurfaces $q_i(\mathbf{x}) = 0$, $i = 1, \ldots, g$, then we minimize the sum of the square sum of all the coefficients of all combinations $q_i(\mathbf{p}(\mathbf{s}))$. We assume that all combinations $q_i(\mathbf{p}(\mathbf{s}))$ are expressed in a PosProd basis.

The next step is to describe the manifold used in the approximation. Since we shall require interpolation of several points, and allow for the interpolation of tangents, we call this manifold an interpolant.

Description of Interpolant. In definition 32 on page 51 in Chapter 3 we described the R-positive manifold of dimension g. We here address the nonrational variant, and thus describe the interpolant by the R-positive manifold

$$\mathbf{p}(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{n}} \mathbf{c_i} B_{\mathbf{i},\mathbf{n}}(\mathbf{s}), \ \mathbf{s} \in \Omega.$$

Here $B_{\mathbf{i},\mathbf{n}}(\mathbf{s})$, $\mathbf{i} \in \mathcal{I}_\mathbf{n}$ are polynomial PosProd basis functions of degree $\mathbf{n}$.

To approximate a manifold of dimension g, we require the interpolation of at least $g + 1$ linearly independent points. In addition to the points being interpolated, we can allow for the interpolation of a number of tangent directions and expressions dependent on higher order partial derivatives. The remaining degrees of freedom, when the interpolation constraints are satisfied, we denote $\mathbf{v}$. Thus the interpolant can be expressed as

$$\mathbf{p_v}(\mathbf{s}) = \sum_{\mathbf{i} \in \mathcal{I}_\mathbf{n}} \mathbf{c_i}(\mathbf{v}) B_{\mathbf{i},\mathbf{n}}(\mathbf{s}), \ \mathbf{s} \in \Omega. \tag{5.1}$$

The combination of an algebraic hypersurface $q(\mathbf{x}) = 0$ (of total degree m) and $\mathbf{p_v}(\mathbf{s})$ can, because we only address nonrational approximation, have a simpler expression than in (4.3). Adapted to the notation of this section we get the description

$$q(\mathbf{p_v}(\mathbf{s})) = (\mathbf{D_v}\mathbf{b})^T \mathbf{B}(\mathbf{s}), \tag{5.2}$$

where $\mathbf{B}(\mathbf{s})$ contains all PosProd basis functions resulting from making the m-th power of the PosProd basis described by $B_{\mathbf{i},\mathbf{n}}(\mathbf{s})$, $\mathbf{i} \in \mathcal{I}_{\mathbf{n}}$. We have introduced $\mathbf{D_v}$ to express the dependence of the matrix $\mathbf{D}$ on the remaining degrees of freedom in the interpolant. The number of basis function in $\mathbf{B}(\mathbf{s})$ is as before denoted $\tilde{N}_{m\mathbf{n}}$.

Since we know the algebraic hypersurface(s), the coefficient vector $\mathbf{b}$ is constant. However, we want to combine $\mathbf{D_v}$ and $\mathbf{b}$ to $\mathbf{d}(\mathbf{v}) = \mathbf{D_v}\mathbf{b}$. Thus we reformulate (5.2) to

$$q(\mathbf{p_v}(\mathbf{s})) = \mathbf{d}(\mathbf{v}) \cdot \mathbf{B}(\mathbf{s}). \tag{5.3}$$

We divide the further discussion in two parts addressed in separate subsections.

- In subsection 5.4.1 we address the approximation of a region of a manifold of dimension $g \leq (l-1)$ in $\mathbb{R}^l$ lying in one algebraic represented hypersurface $q(\mathbf{x}) = 0$.

  The main result is a theorem stating that the convergence rate is the same as for best approximation based on $\mathbf{p_v}(\mathbf{s})$ when the approximation is based on minimizing $\|\mathbf{d}(\mathbf{v})\|_2$ with respect to $\mathbf{v}$.

  We concluded with a corollary addressing the case $g = l - 1$. Thus addressing the approximation of a $(l-1)$ dimensional manifold in $\mathbb{R}^l$ lying in an algebraic represented hypersurface. For $l = 2$ this is the approximation of algebraic represented curves in $\mathbb{R}^2$. This is further detailed in Chapter 6. For $l = 3$ this is the approximation of algebraic represented surfaces in $\mathbb{R}^3$.

- In subsection 5.4.2 we address the approximation of a part of a manifold of dimension $g < (l-1)$ lying in the intersection of $(l-g)$ algebraic represented hypersurfaces in $\mathbb{R}^l$. Setting $g = 1$ and $l = 3$ we have the approximation of an algebraic curve in $\mathbb{R}^3$.

We have split the discussion into these two parts, because the handling of a single hypersurface is simpler than the handling the intersection of multiple hypersurfaces. Thus the discussion in Section 5.4.1 is more easy going than the discussion in Section 5.4.2.

5.4.1 Approximating Manifolds Contained in an Algebraic Represented Hypersurface

The approximation method to be analyzed is based on finding the value(s) of $\mathbf{v}$ that gives $\|\mathbf{d}(\mathbf{v})\|_2$, defined in (5.3), its minimum value. The proof of the convergence rate for this method is divided in the following steps:

- First we give lemma 75 stating that when given:

    - A manifold $\mathbf{f}$ of dimension g in $\mathbb{R}^l$ with $g < l$.
    - An algebraic represented hypersurface $q(\mathbf{x}) = 0$ containing $\mathbf{f}$.
    - An approximation method $\mathbf{p}_{\tilde{\mathbf{v}}}$ to $\mathbf{f}$ with convergence rate $O(h^r)$.

    Then the combination $q(\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{s}))$ has coefficients $\mathbf{d}(\tilde{\mathbf{v}}) = O(h^r)$.

- Then lemma 76 states that coefficients $\mathbf{d}(\mathbf{v})$ satisfy

$$|q(\mathbf{p}_{\mathbf{v}}(\mathbf{s}))| \leq \|\mathbf{d}(\mathbf{v})\|_2 .$$

    This indicates that minimizing $\|\mathbf{d}(\mathbf{v})\|_2$ can give an approximation to a part of the algebraic hypersurface $q(\mathbf{x}) = 0$.

- Theorem 77 states that, given the above mentioned manifold $\mathbf{f}$ and the existence of the $O(h^r)$ approximation, then the minimization of $\|\mathbf{d}(\mathbf{v})\|_2$ is an $O(h^r)$ approximation provided:

    - The gradient of q is nonvanishing close to $\mathbf{f}$.
    - The gradient of q is not orthogonal to the directions for error measurement close to $\mathbf{f}$.

    The direction for error measurement depends on the actual approximation $\mathbf{p}_{\hat{\mathbf{v}}}$. Thus it is necessary to check if these conditions are satisfied when an approximation has been found.

- This subsection is concluded by corollary 78 giving additional conditions simplifying the use of the approximation method in certain cases. These simplifications are:

    - $g = l - 1$, i.e. approximating a part of a hypersurface.
    - Limit the variation of ∇q to be less than $\frac{\pi}{2}$ close to $\mathbf{f}$.
    - Use $\nabla q(\mathbf{f}(\mathbf{s}))$ as direction for error measurement.

Now we take the first step in the above mentioned process, and prove the convergence rate for the coefficient of the combination of the algebraic represented hypersurface, and the existing approximation method.

Lemma 75 *Given a manifold $\mathbf{f}(\mathbf{t})$ of dimension $g < l$ in $\mathrm{I\!R}^l$, an algebraic hypersurface $q(\mathbf{x}) = 0$ containing $\mathbf{f}$ and an approximation method $\mathbf{p_v}(\mathbf{s})$, that for $\mathbf{v} = \tilde{\mathbf{v}}$, has an $O(h^r)$ convergence to $\mathbf{f}(\mathbf{t})$ for h small enough. Then for h small enough*

$$\|\mathbf{d}(\tilde{\mathbf{v}})\|_\infty = O(h^r)$$

with

$$\|(x_1, \ldots, x_n)\|_\infty = \max_{i=1,\ldots,n} |x_i|,$$

where $\mathbf{d}(\mathbf{v})$ is given in (5.3).

Proof. Since we know we have an approximation $\mathbf{p}_{\tilde{\mathbf{v}}}$ with convergence rate $O(h^r)$ when $\mathbf{v} = \tilde{\mathbf{v}}$, we can decompose $\mathbf{p}_{\tilde{\mathbf{v}}}$ as follows

$$\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{s}) = \mathbf{f}(\varphi(\mathbf{s})) + \eta(\mathbf{s})\mathbf{g}(\mathbf{s}),$$

where $\eta(\mathbf{s}) = O(h^r)$ is the error function, the direction for error measurement $\mathbf{g}(\mathbf{s})$ satisfy $\|\mathbf{g}(\mathbf{s})\|_2 = 1$ and $\varphi(\mathbf{s})$ a reparametrization.. By the assumption in the lemma, we know that $q(\mathbf{f}) \equiv 0$, and thus Taylor expansion with respect to $\eta(\mathbf{s})$ gives ($\phi(\mathbf{s})$ replaces $\eta(\mathbf{s})$ in the error term)

$$
\begin{aligned}
q(\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{t})) &= q(\mathbf{f}(\varphi(\mathbf{s})) + \eta(\mathbf{s})\mathbf{g}(\mathbf{s})) &(5.4)\\
&= q(\mathbf{f}(\varphi(\mathbf{s})) + \nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})\eta(\mathbf{s})\\
&= \nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})\eta(\mathbf{s})
\end{aligned}
$$

with $|\phi(\mathbf{s})| \leq |\eta(\mathbf{s})|$. This Taylor expansion is the same as the one used on page 82 to make (4.26). Combining this with (5.3), we get

$$\mathbf{d}(\tilde{\mathbf{v}}) \cdot \mathbf{B}(\mathbf{s}) = \nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})\eta(\mathbf{s}),$$

and introducing absolute values we get

$$|\mathbf{d}(\tilde{\mathbf{v}}) \cdot \mathbf{B}(\mathbf{s})| \leq \max_{\mathbf{s} \in \Omega} |\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s}))| \, |\eta(\mathbf{s})| = O(h^r).$$

Since the entries in $\mathbf{B}(\mathbf{s})$ are linearly independent and independent of h, entries in $\mathbf{d}(\tilde{\mathbf{v}})$ that have a convergence less than $O(h^r)$, can not be canceled by another entry. Thus for $|\mathbf{d}(\tilde{\mathbf{v}}) \cdot \mathbf{B}(\mathbf{s})|$ to be $O(h^r)$ all entries in $\mathbf{d}(\tilde{\mathbf{v}})$ have to be $O(h^r)$, giving

$$\|\mathbf{d}(\tilde{\mathbf{v}})\|_\infty = O(h^r).$$

■

The step following is to limit the value of $|q(\mathbf{p_v}(\mathbf{t}))|$ by $\|\mathbf{d}(\mathbf{v})\|_2$. Thus indicating the possibility to approximate a part of $q(\mathbf{x}) = 0$ by minimizing $\|\mathbf{d}(\mathbf{v})\|_2$.

Lemma 76 *Given* $\mathbf{f}$, $q(\mathbf{x}) = 0$ *and* $\mathbf{p_v}(\mathbf{s})$ *as described in lemma 75, then*

$$|q(\mathbf{p_v}(\mathbf{t}))| \leq \|\mathbf{d}(\mathbf{v})\|_2.$$

Proof. From (5.3) we have that $q(\mathbf{p_v}(\mathbf{s})) = \mathbf{d}(\mathbf{v}) \cdot \mathbf{B}(\mathbf{s})$ with $\mathbf{B}(\mathbf{s})$ a PosProd basis. Now we have that

$$|q(\mathbf{p_v}(\mathbf{s}))| = |\mathbf{d}(\mathbf{v}) \cdot \mathbf{B}(\mathbf{s})| \leq \|\mathbf{d}(\mathbf{v})\|_\infty \leq \|\mathbf{d}(\mathbf{v})\|_2.$$

■

Now the foundation for proving the convergence rate is established.

Theorem 77 *Given* $\mathbf{f}$, $q(\mathbf{x}) = 0$ *and* $\mathbf{p_{\tilde{v}}}(\mathbf{s})$ *as described in lemma 75 and* $\mathbf{p_{\hat{v}}}(\mathbf{s})$ *defined by*

$$\|\mathbf{d}(\hat{\mathbf{v}})\|_2 = \min_{\mathbf{v}} \|\mathbf{d}(\mathbf{v})\|_2. \tag{5.5}$$

If the gradient of q is nonvanishing and not normal to the directions for error measurement $\mathbf{g}(\mathbf{s})$ *close to* $\mathbf{f}(\mathbf{t})$, *then for h small enough, the approximation method* $\mathbf{p_{\hat{v}}}(\mathbf{s})$ *has convergence rate* $O(h^r)$. *The approximation error is limited by*

$$|\eta(\mathbf{s})| \leq \min_{\substack{s \in \Omega \\ |\rho| \leq |\eta(s)|}} \frac{\|\mathbf{d}(\hat{\mathbf{v}})\|_2}{|\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \rho\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|}. \tag{5.6}$$

Proof. We can decompose $\mathbf{p_{\hat{v}}}$ in the same way as $\mathbf{p_{\tilde{v}}}$ was decomposed in (5.4)

$$\mathbf{p_{\hat{v}}}(\mathbf{s}) = \mathbf{f}(\varphi(\mathbf{s})) + \eta(\mathbf{s})\mathbf{g}(\mathbf{s}),$$

giving

$$q(\mathbf{p_{\hat{v}}}(\mathbf{s})) = \nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})\ \eta(\mathbf{s})$$

with $\eta(\mathbf{s})$ being the approximation error. Since the gradient of q is nonvanishing and not normal to the direction for error measurement $\mathbf{g}(\mathbf{s})$ close to $\mathbf{f}$, we have for h small enough a region with $\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s}) \neq 0$. Thus we have using lemma 76 and (5.5) that

$$
\begin{aligned}
|\eta(\mathbf{s})| &= \frac{|q(\mathbf{p}_{\hat{\mathbf{v}}}(\mathbf{s}))|}{|\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|} \\[2ex]
&\leq \frac{\|\mathbf{d}(\hat{\mathbf{v}})\|_2}{\displaystyle\min_{\substack{s \in \Omega \\ |\rho| \leq |\eta(s)|}} |\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \rho\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|} \\[2ex]
&\leq \frac{\|\mathbf{d}(\tilde{\mathbf{v}})\|_2}{\displaystyle\min_{\substack{s \in \Omega \\ |\rho| \leq |\eta(s)|}} |\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \rho\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|} \\[2ex]
&= O(h^r).
\end{aligned}
$$

$\blacksquare$

We now give a corollary, where the dimension g of the manifold, is set to $(l-1)$, thus covering:

- Approximation of algebraic represented curves in $\mathbb{R}^2$.

- Algebraic represented surfaces in $\mathbb{R}^3$.

Corollary 78 *Given* $\mathbf{f}$, $q(\mathbf{x}) = 0$, $\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{s})$ *and* $\mathbf{p}_{\hat{\mathbf{v}}}(\mathbf{s})$ *as described in theorem 77, and let the dimension of the manifold* $\mathbf{f}$ *be* $(l-1)$. *Let the direction for error measurement at* $\mathbf{g}(\mathbf{s})$ *be parallel to* $\nabla q(\mathbf{f}(\mathbf{s}))$ *and let the gradient of* q *vary less that* $\frac{\pi}{2}$ *close to* $\mathbf{f}$. *Then for* h *small enough* $\mathbf{p}_{\hat{\mathbf{v}}}(\mathbf{s})$ *is an* $O(h^r)$ *approximation to* $\mathbf{f}$.

Proof. With the choice of $\mathbf{g}(\mathbf{s})$ parallel to the gradient, and the limitation of the variation of the gradient direction we have that close to $\mathbf{f}$

$$
\min |\nabla q(\mathbf{f}(\varphi(\mathbf{s})) + \phi(\mathbf{s})\mathbf{g}(\mathbf{s}) \cdot \mathbf{g}(\mathbf{s})| \neq 0,
$$

and theorem 77 gives $|\eta(\mathbf{s})| = O(h^r)$. $\blacksquare$

5.4.2 Approximation of Manifolds in the Intersection of Algebraic Represented Hypersurfaces

The approximation problem to be addressed now is closely related to the problem addressed in the previous subsection. The main difference is that the manifold $\mathbf{f} \in \mathbb{R}^l$ now lies in the intersection of several hypersurfaces.

The notation is thus slightly more complicated. The minimization has to be performed over a sum of the square sums of coefficients.

We still let the dimension of $\mathbf{f}$ be denoted g, and denote the manifolds intersecting along $\mathbf{f}$ by

$$q_i(\mathbf{x}) = 0, i = 1, \ldots, l - g$$

with $g < l - 1$. The expression we want to minimize is

$$\sum_{i=1}^{l-g} \|\mathbf{d}_i(\mathbf{v})\|_2^2 \tag{5.7}$$

with $\mathbf{d}_i(\mathbf{v})$ given by

$$q_i(\mathbf{p_v}(\mathbf{s})) = \mathbf{d}_i(\mathbf{v}) \cdot \mathbf{B}(\mathbf{s}), \ i = 1, \ldots, l - g. \tag{5.8}$$

Here $\mathbf{p_v}(\mathbf{s})$ is the interpolant given in (5.3).

The main challenge is to find conditions that ensure that the minimization of (5.7) results in an approximation with the required convergence rate. The proof of the convergence rate in theorem 77 did not use the dimension of the manifold being approximated, thus the results are also valid for this approximation problem. For manifolds lying in the intersection of two or more hypersurfaces it is not difficult to construct examples where the direction of error measurement and the gradient of one of the manifolds, are normal. Thus the direction of error measurement can not be used with respect to all algebraic hypersurfaces.

To discuss the convergence rate of this approximation method we use a similar structure of lemmas and theorems as in the previous section.

- First we give lemma 79, stating that when given:

  - A manifold $\mathbf{f}$ of dimension g in $\mathbb{R}^l$ with $g < l - 1$.
  - Algebraic represented hypersurface $q_i(\mathbf{x}) = 0$, $i = 1, \ldots, g$ intersecting along $\mathbf{f}$.
  - An approximation method $\mathbf{p_{\tilde{v}}}$ to $\mathbf{f}$ with convergence rate $O(h^r)$.

  Then the coefficients $\mathbf{d}_i(\tilde{\mathbf{v}})$ in (5.8) satisfy

$$\sum_{i=1}^{g} \|\mathbf{d}_i(\tilde{\mathbf{v}})\|_2^2 = O(h^r).$$

- Theorem 80 then following states that the minimization of $\sum_{i=1}^{g} \|\mathbf{d}_i(\mathbf{v})\|_2^2$ gives an $O(h^r)$ approximation to $\mathbf{f}$ provided that:

    - The gradient of one of the hypersurfaces $q_i(\mathbf{x}) = 0$, , $i = 1, \ldots, g$ is nonvanishing and not orthogonal to the direction for error measurement close to $\mathbf{f}$.

    The direction for error measurement is dependent on the actual approximation $\mathbf{p}_{\hat{\mathbf{v}}}$. Thus, as in the case of one algebraic hypersurface, it is necessary to check the direction for error measurement after an approximation has been found.

    By adding the two following requirements to the hypersurfaces, we get a simpler approximation problem to implement:

    - The gradient of q_i, $i = 1, \ldots, g$ is nonvanishing close to $\mathbf{f}$ and varying less than $\frac{\pi}{2}$ close to $\mathbf{f}$, for $i = 1, \ldots, g$.
    - The intersection of the hypersurfaces $q_i(\mathbf{x}) = 0$, $i = 1, \ldots, g$ is nonsingular along $\mathbf{f}$.

Lemma 79 *Given a manifold $\mathbf{f}(\mathbf{t})$ of dimension $g < l - 1$ in $\mathbb{R}^l$, and a set of algebraic hypersurfaces $q_i(\mathbf{x}) = 0$, $i = 1, \ldots, l - g$ intersecting along $\mathbf{f}$. In addition assume that there exists an approximation method $\mathbf{p}_{\mathbf{v}}(\mathbf{s})$ that for $\mathbf{v} = \tilde{\mathbf{v}}$ has an $O(h^r)$ convergence to $\mathbf{f}(\mathbf{t})$ for h small enough. Then for h small enough we have*

$$\max \|\mathbf{d}_i(\tilde{\mathbf{v}})\|_{\infty} = O(h^r), \ i = 1, \ldots, l - g,$$

where $\mathbf{d}_i(\tilde{\mathbf{v}})$ is given in (5.8).

Proof. Since $\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{s})$ is an approximation to the intersection of all the hypersurfaces and have convergence rate $O(h^r)$, $\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{s})$ it is also and approximation to each of the hypersurfaces. Thus lemma 75 can be used on each hypersurface thus proving this lemma. ∎

Now we connect the convergence of the coefficients of the existing approximation method to coefficients of the method based on the minimum square sum.

Theorem 80 *Given $\mathbf{f}$, q and $\mathbf{p}_{\tilde{\mathbf{v}}}(\mathbf{s})$ as described in lemma 79. Let $\mathbf{p}_{\hat{\mathbf{v}}}(\mathbf{s})$ be defined by*

$$\sum_{i=1}^{l-g} \|\mathbf{d}_i(\hat{\mathbf{v}})\|_2^2 \leq \min_{\mathbf{v}} \sum_{i=1}^{l-g} \|\mathbf{d}_i(\mathbf{v})\|_2^2 .$$

If for at least one $j \in \{1, \ldots, g\}$ the gradient of q_j is nonvanishing and not normal to the directions for error measurement $\mathbf{g}(\mathbf{t})$ close to $\mathbf{f}(\mathbf{t})$, then for h small enough the approximation method $\mathbf{p}_{\hat{\mathbf{v}}}(\mathbf{s})$ has convergence rate $O(h^r)$. The approximation error is limited by

$$\left|\eta_j(\mathbf{s})\right| \leq \frac{\|\mathbf{d}_j(\hat{\mathbf{v}})\|_2}{\displaystyle\min_{\substack{s \in \Omega \\ |\rho| \leq |\eta_j(\mathbf{s})|}} |\nabla q_j(\mathbf{f}(\varphi(\mathbf{s})) + \rho\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|}.$$

Proof.   Since the gradient of q_j is nonvanishing and not normal to the directions for error measurement $\mathbf{g}(\mathbf{t})$ close to $\mathbf{f}(\mathbf{t})$, the condition of theorem 77 are satisfied giving

$$\left|\eta_j(\mathbf{s})\right| \leq \frac{\|\mathbf{d}_j(\hat{\mathbf{v}})\|_2}{\displaystyle\min_{\mathbf{s}\in\Omega, |\rho|\leq|\eta_j(\mathbf{s})|} |\nabla q_j(\mathbf{f}(\varphi(\mathbf{s})) + \rho\mathbf{g}(\mathbf{s})) \cdot \mathbf{g}(\mathbf{s})|}.$$

Further the definition of $\hat{\mathbf{v}}$ and lemma 79 give

$$\|\mathbf{d}_j(\hat{\mathbf{v}})\|_2 \leq \sqrt{\sum_{i=1}^{l-g} \|\mathbf{d}_i(\hat{\mathbf{v}})\|_2^2} \leq \sqrt{\sum_{i=1}^{l-g} \|\mathbf{d}_i(\tilde{\mathbf{v}})\|_2^2} = O(h^r).$$

Thus the theorem is proved.   ∎

Remark 23 *This theorem is difficult to put in practice, because the hypersurfaces intersecting along $\mathbf{f}$ can have a singular intersection. Another problem is that the gradients of the hypersurfaces can vanish, or vary more than desirable. Thus to get a method that is simpler to implement, we add the following conditions on hypersurfaces intersecting along $\mathbf{f}$:*

- *For all $i = 1, \ldots, l - g$ the gradient of q_i is required to be nonvanishing and vary less than $\frac{\pi}{2}$ close to $\mathbf{f}$.*

- *The intersection of the hypersurfaces q_i, $i = 1, \ldots, g$ is nonsingular along $\mathbf{f}$.*

Example 81 *In [Höllig:95] theorem 3 states that: "For a smooth curve with nonvanishing curvature and torsion the error of the cubic geometric interpolant is of order $O(h^5)$, where h is the length of the curves segment between s_0 and s_1." Here s_0 and s_1 are references to the parameter interval on the curve $\mathbf{f}$ to be interpolated. The interpolation constraints imposed are interpolation of position and tangent direction at s_0 and s_1. Thus the interpolant can*

be described as $\mathbf{p}_{(l_0,l_1)}(s)$. The remaining degrees of freedom are the tangent lengths at the start and end of the cubic Hermite segment l_0 and l_1.

Now assume that $\mathbf{f}$ lies in the nonsingular intersection of two algebraic surfaces $q_1(\mathbf{x}) = 0$ and $q_2(\mathbf{x}) = 0$. Let $\mathbf{d}_1(l_0, l_1)$ and $\mathbf{d}_2(l_0, l_1)$ be defined as in (5.8). Then minimizing

$$\|\mathbf{d}_1(l_0, l_1)\|_2^2 + \|\mathbf{d}_2(l_0, l_1)\|_2^2$$

gives an $O(h^5)$ approximation, provided the resulting direction for error measurement satisfies the conditions in theorem 80.

Thus using the algebraic implicitization with algebraic degree at least 2 in Chapter 4, and the approximation method of this section, enable us to approximate $\mathbf{f}$ with $O(h^5)$ convergence, provided $\mathbf{f}$ has nonvanishing curvature and torsion.

If the torsion is vanishing all along the curve, then the curve is planar, and the results from approximation of curves in $\mathbb{R}^2$ can be used.

Chapter 6

Cubic Hermite Approximation of 2D Curves

When approximating the trace of an intersection curve in the parameter domain of a parametric surface, it is desirable to find an approximation that use as few polynomial segments as possible. Constraints imposed on such approximations, are often a given polynomial degree and a maximum error allowed.

In this chapter we look at cubic Hermite approximation of curves in $\mathbb{R}^2$ with $O(h^6)$ convergence. The basis for the methods presented is the theorem in [de Boor:87-2] where it is stated:

"If f is a smooth curve with nonvanishing curvature and

$$h := \sup |t_{i+1} - t_i|$$

is sufficiently small, then the positive solution of the system (Q) exist and the corresponding interpolant(s) p_f satisfy $dist(f, p_f) = O(h^6)$."

In this theorem the system (Q) describes the interpolation of position, tangent and curvature from a planar curve $f : \mathbb{R} \to \mathbb{R}^2$ at parameter values t_i and t_{i+1}. The interpolant p_f is a cubic Hermite interpolant. Tangent lengths at the start and end of the Hermite interpolant are controlled to achieve the required interpolation of curvature.

Based on this we see that, provided the curvature is nonvanishing, and the approximation is performed by a proper method, we can expect an $O(h^6)$ approximation to the curves in the parameter domain of a parametric surface by cubic Hermite interpolation.

In Section 5.3 we pointed out that the trace of an intersection curve in the parameter domain of a surface, is an algebraic curve. Thus, there is in general no parametric description of such 2D curves. However, we have the algebraic representation of the curve $q(x, y) = 0$ either as an exact representation or

as an approximation. The algebraic description gives the advantage that we have a description independent of the parametrization of the curve. There is, however, a difficulty as well. From the algebraic representation and two points lying on the algebraic curve, we have no guarantee that the points lie on the same branch of the algebraic curve.

The discussion in this chapter is structured as follows:

- In Section 6.1 we address approximation when both the parametric and algebraic description of the curve is known. We use the convergence rate cited above for cubic Hermite interpolation of parametric curves in $\mathrm{I\!R}^2$ in combination with the algebraic representation of the curve. This combination enables us to find Taylor expansions of the free tangents lengths of the cubic Hermite interpolant that give $O(h^6)$ convergence. The expansion is based on the parametrization of the curve to be approximated.

- In Section 6.2 we address approximation when only the algebraic representation of the curve is known. We give a result being a consequence of theorem 77 on page 121 in Section 5.3. Minimizing the square sum of the coefficients of the combination $q(\mathbf{p}(t))$ is shown to give an $O(h^6)$ approximation. When using this method, the explicit parametric representation of the segment to be approximated, can be unknown.

- Finally in Section 6.3, we combine the approximative implicitization in Chapter 4 and the above mentioned method for approximation of algebraic curves, to describe a two step approach to the approximation of parametric curves in $\mathrm{I\!R}^2$.

6.1 Power Expansion Based on Parametric Representation

Let an algebraic curve be defined by

$$q(x, y) = 0, \tag{6.1}$$

where

$$q(x, y) = \sum_{\substack{i_1+i_2+i_3=m \\ i_1,i_2,i_3\geq 0}} a_{i_1,i_2,i_3} x^{i_1} y^{i_2},$$

and let

$$\mathbf{f}(\theta) = (f_x(\theta), f_y(\theta)), \theta \in [-\frac{\alpha}{2}, \frac{\alpha}{2}] \tag{6.2}$$

128

be an exact parametric representation of a segment with nonvanishing curvature of the algebraic curve. We want to make a cubic Hermite interpolant $\mathbf{p}$ to $q(x, y) = 0$ satisfying

$$
\begin{aligned}
\mathbf{p}(0) &= \mathbf{f}(-\tfrac{\alpha}{2}) \\
\mathbf{p}'(0) &\parallel \mathbf{f}'(-\tfrac{\alpha}{2}) \\
\mathbf{p}(1) &= \mathbf{f}(\tfrac{\alpha}{2}) \\
\mathbf{p}'(1) &\parallel \mathbf{f}'(\tfrac{\alpha}{2}),
\end{aligned}
\tag{6.3}
$$

where $\parallel$ means parallel to.

Description of interpolant. The Hermite interpolant $\mathbf{p}(t)$ is described in a cubic Bernstein basis

$$
\mathbf{p}(t) = (x(t), y(t)) = \sum_{i=0}^{3} \mathbf{p}_i \binom{3}{i} (1-t)^{3-i} t^i,
\tag{6.4}
$$

where

$$
\mathbf{p}_0 = \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} f_x(-\frac{\alpha}{2}) \\ f_y(-\frac{\alpha}{2}) \end{pmatrix}
$$

$$
\mathbf{p}_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} f_x(-\frac{\alpha}{2}) \\ f_y(-\frac{\alpha}{2}) \end{pmatrix} + L_0(\alpha) \begin{pmatrix} f_x'(-\frac{\alpha}{2}) \\ f_y'(-\frac{\alpha}{2}) \end{pmatrix}
$$

$$
\mathbf{p}_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} f_x(\frac{\alpha}{2}) \\ f_y(\frac{\alpha}{2}) \end{pmatrix} - L_1(\alpha) \begin{pmatrix} f_x'(\frac{\alpha}{2}) \\ f_y'(\frac{\alpha}{2}) \end{pmatrix}
$$

$$
\mathbf{p}_3 = \begin{pmatrix} x_3 \\ y_3 \end{pmatrix} = \begin{pmatrix} f_x(\frac{\alpha}{2}) \\ f_y(\frac{\alpha}{2}) \end{pmatrix}.
$$

Since

$$
\begin{aligned}
\mathbf{p}(0) &= \mathbf{p}_0 &&= \mathbf{f}(-\tfrac{\alpha}{2}) \\
\mathbf{p}'(0) &= 3(\mathbf{p}_1 - \mathbf{p}_0) &&= 3L_0(\alpha)\mathbf{f}'(-\tfrac{\alpha}{2}) \\
\mathbf{p}'(1) &= 3(\mathbf{p}_3 - \mathbf{p}_2) &&= 3L_1(\alpha)\mathbf{f}'(\tfrac{\alpha}{2}) \\
\mathbf{p}(1) &= \mathbf{p}_3 &&= \mathbf{f}(\tfrac{\alpha}{2}),
\end{aligned}
$$

this interpolant satisfies the interpolation requirements in (6.3).

The problem is finding $L_0(\alpha)$ and $L_1(\alpha)$ such that $\mathbf{p}$ approximates $f(\theta)$ with accuracy $O(\alpha^6)$. The procedure is divided in the following four steps:

129

1. In lemma 82 we establish an equation combining the algebraic curve $q(\mathbf{x}) = 0$ with the Hermite interpolant $\mathbf{p}(t)$. This combination is a polynomial $q(\mathbf{p}(t))$ of degree $3m$, and is represented in a Bernstein basis.

2. In lemma 83 we show that there exists interpolants $\mathbf{p}$ such that the coefficients of $q(\mathbf{p}(t))$ have $O(\alpha^6)$ convergence provided the curvature of the segment approximated is nonvanishing.

3. Then in lemma 84 we establish a relation between the coefficients with convergence rate $O(\alpha^6)$ and the error of the Hermite interpolant. Here it is required that the gradient of q is nonvanishing and not orthogonal to the direction for error measurement close to the curve segment approximated.

4. In theorem 85 we prove that $O(\alpha^6)$ approximations can be found by Taylor expanding the coefficients found in lemma 82 and forcing all terms of degree less than 6 to be zero.

Lemma 82 *Let the algebraic curve $q(\mathbf{x}) = 0$, be defined as in (6.1) and let the cubic Hermite interpolant $\mathbf{p}(t)$ be defined as in (6.4). Then the combination of $q(x, y)$ and $\mathbf{p}(t)$ can be written in the form*

$$q(\mathbf{p}(t)) = \sum_{k=2}^{3m-2} g_k(\alpha) \binom{3m}{k} (1-t)^{3m-k} t^k, \qquad (6.5)$$

for certain coefficients $g_k(\alpha)$.

Proof. Since $\mathbf{p}(t)$ is a parametric cubic curve and $q(\mathbf{x}) = 0$ is of degree m, then $q(\mathbf{p}(t))$ is of degree $3m$ and thus can be described as in (6.5). But with the index k in the sum going from 0 to $3m$. The zero values of $g_k(\alpha)$ for $k = 0, 1, 3m - 1, 3m$ follow from (6.3). ∎

Lemma 83 *Let $g_k(\alpha)$, $k = 2, \ldots, 3m - 2$ be defined as in lemma 82, let as in (6.2) $\mathbf{f}(\theta)$, $-\frac{\alpha}{2} \leq \theta \leq \frac{\alpha}{2}$ be a segment with nonvanishing curvature of the algebraic curve $q(x, y) = 0$ that we want to approximate. Then for α sufficiently small there exists $\mathbf{p}$ given in (6.4) such that the coefficients $g_k(\alpha)$ satisfy*
$$g_k(\alpha) = O(\alpha^6), \quad k = 2, \ldots, 3m - 2.$$

Proof. In [de Boor:87-2] it was shown that a cubic Hermite interpolant $\mathbf{p}(t)$ with $O(\alpha^6)$ accuracy to a 2D curve segment $\mathbf{f}(\theta)$ exists when the curve

segment has nonvanishing curvature. Thus, there exist $L_1(\alpha)$ and $L_2(\alpha)$ such that the approximation error $\rho(t) = O(\alpha^6)$. Now we decompose $\mathbf{p}(t)$ as follows

$$\mathbf{p}(t) = \mathbf{f}(\varphi(t)) + \rho(t)\mathbf{g}_\varphi(t), \tag{6.6}$$

where $\varphi(t)$, is a reparametrization. and the unit vector $\mathbf{g}_\varphi(t)$ is the direction for error measurement at $\mathbf{f}(\varphi(t))$. The next step is to Taylor expand the expression $q(\mathbf{p}(t))$ with respect to $\rho(t)$. (Note that in addition to the functions $\rho(t)$ and $\varphi(t)$ we also introduce the function $\phi(t)$ in the error term of the Taylor expansion).

$$
\begin{aligned}
q(\mathbf{p}(t)) &= q(\mathbf{f}(\varphi(t)) + \rho(t)\mathbf{g}_\varphi(t)) \\
&= q(\mathbf{f}(\varphi(t))) + \nabla q(\mathbf{f}(\varphi(t)) + \phi(t)\mathbf{g}_\varphi(t)) \cdot \mathbf{g}_\varphi(t)\rho(t) \quad (6.7) \\
&= \nabla q(\mathbf{f}(\varphi(t)) + \phi(t)\mathbf{g}_\varphi(t)) \cdot \mathbf{g}_\varphi(t)\rho(t)
\end{aligned}
$$

with $|\phi(t)| \leq |\rho(t)|$, $\rho(t)\phi(t) \geq 0$. Thus, we can limit $|q(\mathbf{p}(t))|$ as follows

$$|q(\mathbf{p}(t))| \leq \|\nabla q(\mathbf{f}(\varphi(t)) + \phi(t)\mathbf{g}_\varphi(t))\|_2 \, |\rho(t)| \,.$$

Since $\rho(t)$ is $O(\alpha^6)$ for α small enough and $L_1(\alpha)$ and $L_2(\alpha)$ chosen properly, we get

$$|q(\mathbf{p}(t))| \leq \max_{|\phi(t)| \leq |\rho(t)|} \|\nabla q(\mathbf{f}(\varphi(t)) + \phi(t)\mathbf{n}_\varphi(t))\|_2 \, |\rho(t)| = O(\alpha^6).$$

By the definition of g_k, $k = 2, \ldots, 3m - 2$ in (6.5) we get $g_k = O(\alpha^6)$, $k = 2, \ldots, 3m - 2$. $\blacksquare$

Lemma 84 *Let $g_k(\alpha)$, $k = 2, \ldots, 3m - 2$ be defined as in lemma 83, let as in (6.2) $\mathbf{f}(\theta)$, $-\frac{\alpha}{2} \leq \theta \leq -\frac{\alpha}{2}$ be a segment with nonvanishing curvature of the algebraic curve $q(x, y) = 0$ that we want to approximate. If the gradient of q is nonvanishing and not orthogonal to the direction of error measurement close to $\mathbf{f}(\theta)$, $\theta \in [-\frac{\alpha}{2}, \frac{\alpha}{2}]$ then*

$$|\rho(t)| \leq K \, \|(g_2(\alpha), \ldots, g_{3m-2}(\alpha))\|_\infty$$

with K a constant.

Proof. From (6.5) we have that

$$|q(\mathbf{p}(t))| \leq \|(g_2(\alpha), \ldots, g_{3m-2}(\alpha))\|_\infty \,.$$

In (6.7) we showed that

$$q(\mathbf{p}(t)) = \nabla q(\mathbf{f}(\varphi(t)) + \phi(t)\mathbf{g}_\varphi(t)) \cdot \mathbf{g}_\varphi(t)\rho(t).$$

Since for α sufficiently small, $\min\limits_{|\sigma|\leq\rho_{\max}} |\nabla q(\mathbf{f}(\varphi(t)) + \sigma\mathbf{g}_\varphi(t)) \cdot \mathbf{g}_\varphi(t)| \neq 0$ we get

$$
\begin{aligned}
|\rho(t)| &\leq \frac{|q(\mathbf{p}(t))|}{\min\limits_{|\sigma|\leq\rho_{\max}} |\nabla q(\mathbf{f}(\varphi(t)) + \sigma\mathbf{g}_\varphi(t)) \cdot \mathbf{g}_\varphi(t)|} \\
&\leq \frac{\|(g_2(\alpha),\ldots,g_{3m-2}(\alpha))\|_\infty}{\min\limits_{|\sigma|\leq\rho_{\max}} |\nabla q(\mathbf{f}(\varphi(t)) + \sigma\mathbf{g}_\varphi(t)) \cdot \mathbf{g}_\varphi(t)|} \\
&= K\,\|(g_2(\alpha),\ldots,g_{3m-2}(\alpha))\|_\infty\,.
\end{aligned}
$$

∎

Theorem 85 *Let the algebraic curve be defined as in (6.1) by $q(\mathbf{x}) = 0$, and let $\mathbf{f}(\varphi)$ as in (6.2) be an exact parametric representation of a segment with nonvanishing curvature of the algebraic curve. Then for α small enough an $O(\alpha^6)$ approximation to the algebraic segment of the type (6.4) can be found by Taylor expanding*

$$g_k(\alpha),\ k = 2,\ldots,3m-2$$

defined in (6.5), and forcing the terms of degree less than 6 to be zero.

Proof. Lemma 83 ensures that $g_k(\alpha),\ k = 2,\ldots,3m-2$ are $O(\alpha^6)$. Thus, a Taylor expansion around $\alpha = 0$ gives us conditions for determining the first terms in the Taylor expansion of respectively $L_0(\alpha)$ and $L_1(\alpha)$. Lemma 84 connects the $O(\alpha^6)$ behavior to the approximation error. ∎

In Chapter 7 this approach is used to describe a family of cubic Hermite ellipse and circle approximations methods that has $O(\alpha^6)$ convergence and has a specific Taylor expansion of the free tangent lengths $L_0(\alpha)$ and $L_1(\alpha)$.

6.2 Minimizing the Square Sum of the Coefficients

In the previous section we established how to make an $O(\alpha^6)$ cubic Hermite approximation by expansion of the coefficients of $q(\mathbf{p}(t))$ with respect to α. We now show that an $O(\alpha^6)$ interpolant can be found by minimizing the sum of the square of the coefficients of $q(\mathbf{p}(t))$. This is a specialization of theorem 77 on page 121 in Section 5.3.

The first step is a lemma limiting the value of $|q(\mathbf{p}(t))|$ by an expression including the square sum of the coefficients $q(\mathbf{p}(t))$. Then a theorem is given, proving that the convergence rate of this approximation is $O(\alpha^6)$.

Lemma 86 *Let $q(\mathbf{x})$ and $\mathbf{p}(t)$ be respectively defined as in (6.1) and (6.4) then*

$$|q(\mathbf{p}(t))| \leq \sqrt{\sum_{k=2}^{3m-2} (g_k(\alpha))^2}. \tag{6.8}$$

Proof. From (6.5) we have that

$$|q(\mathbf{p}(t))| \leq \max_{0 \leq k \leq 3m} |g_k(\alpha)| \leq \sqrt{\sum_{k=2}^{3m-2} (g_k(\alpha))^2}$$

proving (6.8). ∎

The theorem following is minimizing (6.8) and proving that the resulting Hermite interpolant is $O(\alpha^6)$. We now, instead of using $g_k(\alpha)$, $k = 2, \ldots, 3m - 2$ to describe the coefficients of the combination $q(\mathbf{p}(t))$, use $g_k(L_1, L_2)$, $k = 2, \ldots, 3m - 2$. We do this to focus on the tangent lengths in stead of the parametrization in the theorem to follow.

Theorem 87 *Let $q(\mathbf{x})$ and $\mathbf{f}(\theta)$ be defined respectively as in (6.1) and (6.2) having nonvanishing curvature. Then the cubic Hermite curve $\hat{\mathbf{p}}(t)$ given in (6.4) with tangent lengths given by $\hat{L}_1$ and $\hat{L}_2$ satisfying*

$$\sum_{k=2}^{3m-2} \left(g_k(\hat{L}_1, \hat{L}_2) \right)^2 = \min_{L_1, L_2} \sum_{k=2}^{3m-2} (g_k(L_1, L_2))^2 \tag{6.9}$$

is an $O(\alpha^6)$ approximation to $\mathbf{f}(\varphi)$ for α sufficiently small, provided that the gradient of q is nonvanishing close to $\mathbf{f}$.

Proof. For α small enough we know that an $O(\alpha^6)$ Hermite approximation $\tilde{\mathbf{p}}(t)$ exists. Let the tangent length in such an $O(\alpha^6)$ approximation be respectively $\tilde{L}_1$ and $\tilde{L}_2$. By (6.9) we have

$$\sum_{k=2}^{3m-2} \left(g_k(\hat{L}_1, \hat{L}_2) \right)^2 \leq \sum_{k=2}^{3m-2} \left(g_k(\tilde{L}_1, \tilde{L}_2) \right)^2.$$

We know from lemma 83

$$\left| q(\hat{\mathbf{p}}_{\hat{L}_1, \hat{L}_2}(t)) \right| \leq \sqrt{\sum_{k=2}^{3m-2} \left(g_k(\hat{L}_1, \hat{L}_2) \right)^2}.$$

133

Thus

$$\left| q(\hat{\mathbf{p}}_{\hat{L}_1, \hat{L}_2}(t)) \right| \leq \sqrt{\sum_{k=2}^{3m-2} \left(g_k(\tilde{L}_1, \tilde{L}_2) \right)^2} = O(\alpha^6).$$

Now we choose the direction for error measurement parallel to the gradient of q at $\hat{\mathbf{p}}(t)$. By lemma 84 on page 131, we have since the gradient of q is nonvanishing close to $\mathbf{f}$, that $\hat{\mathbf{p}}(t)$ is an $O(\alpha^6)$ approximation of $\mathbf{f}(\varphi)$. ∎

Remark 24 *Assuming that $\tilde{\mathbf{p}}(t)$ is the best $O(\alpha^6)$ approximation, by this we mean the approximation with the smallest maximum error, we see that the co-efficient of $q(\hat{\mathbf{p}}(t))$ resulting from approximation established in theorem 87 is limited by the 2-norm of the coefficients of $q(\tilde{\mathbf{p}}(t))$ of this best approximation. Thus, the approximation in theorem 87 is close to the best approximation.*

Remark 25 *Since the actual value of α is not involved in the minimization process, the method can be used, if the segment of the algebraic curve has a parametric description. This parametrization does not have to be explicit. We only have to be able to assign a unique parameter value to each point of the segment being approximated. We can find such a parametrization by the following approach, assuming that the tangent direction of the curve segment $\mathbf{f}$ to be approximated, is varying less than π.*

- *Let $\mathbf{l}$ be a straight line going through the start and the end of the curve segment $\mathbf{f}$ to be approximated.*

- *Let $\mathbf{q}$ lie on the side of $\mathbf{l}$ opposite to $\mathbf{f}$.*

- *Let $\mathbf{q}$ lie between the lines that are tangential to respectively the start and end of $\mathbf{f}$.*

By choosing $\mathbf{q}$ according to the constraints above, straight lines through $\mathbf{q}$ only intersect $\mathbf{f}$ once, because we assume that $\mathbf{f}$ has nonvanishing curvature. The parametrization of a point on $\mathbf{f}$ can be chosen to be the rotation angle of the straight line through $\mathbf{q}$, intersecting $\mathbf{f}$ in the given point.

The consequence of this observation is that we can use the algorithm from theorem 87 to make $O(\alpha^6)$ approximations of segments of algebraic curves.

6.3 Cubic Hermite Approximation of Parametric Curves

In corollary 49 on page 78 we established conditions for an $O(h^{\frac{(m+1)(m+2)}{2}-1})$ approximation of an algebraic curve $q(x, y) = 0$ to a parametric curve segment $\mathbf{g}(s)$, $s \in [0, 1]$. Then in theorem 85 on page 132 and remark 25 above

we established conditions for a $O(h^6)$ cubic Hermite approximation $\mathbf{p}(t)$ to an algebraic curve $q(x, y) = 0$. Combining these results we obtain an algorithm for making an $O(h^6)$ cubic Hermite approximation of a segment of parametric curve, by introducing an intermediate algebraic approximation of degree $m \geq 3$.

Let $\mathbf{r}(\alpha)$, $\alpha \in [0, 1]$ be the exact representation of the segment of the algebraic curve used in the approximation. We then get the following separation of the errors of the two approximation stages

$$
\begin{aligned}
\|\mathbf{p}(t) - \mathbf{g}(\phi_1(t))\|_2 &\leq \|\mathbf{p}(t) - \mathbf{r}(\phi_2(t)) + \mathbf{r}(\phi_2(t)) - \mathbf{g}(\phi_1(t))\|_2 \\
&\leq \|\mathbf{p}(t) - \mathbf{r}(\phi_2(t))\|_2 + \|\mathbf{r}(\phi_2(t)) - \mathbf{g}(\phi_1(t))\|_2 \\
&\leq O(h^{\frac{(m+1)(m+2)}{2}-1}) + O(h^6)
\end{aligned}
$$

with $\phi_2(t)$ and $\phi_1(t)$ being reparametrization.

For $m = 1, 2, \ldots$ we get the following total convergence rates

Algebraic degree	1	2	≥ 3
Convergence rate	$O(h^2)$	$O(h^5)$	$O(h^6)$

provided that conditions used for establishing these convergence rates in corollary 49 and theorem 85 are fulfilled. These conditions are:

- The gradient of $\mathbf{q}$ close to $\mathbf{g}(s)$ must be nonvanishing.

- The curvature of the segment of the intermediate algebraic curve used must be nonvanishing.

Example 88 *In the first version of SINTEF Spline Library SISL tangent lengths for a Hermite segment approximating an intersection curve was estimated by using position, tangent and radius of curvature at both start and end point.*

The tangent length L_0 for the start point was calculated by assuming that the curve was a circle with radius r_{start} and opening angle α using the tangent length estimate given in example 96. Similarly the tangent length for the end L_1 was estimated based on r_{end} and α. The opening angle α was set to the angle between the start and end tangent.

This method is in the general case at most an $O(h^5)$ approximation. However, if the curve happened to be a circle segment the approximation rate is $O(h^6)$. Some informal measurements that were performed on the convergence rate, showed that in many practical cases from 3D CAD an $O(h^5)$ behavior was observed. An unexpected consequence of the higher approximation rate

for circles was a bug report from a user stating that too few Hermite segments were produced when marching intersection curves that were true circle segments.

Chapter 7

Ellipse and Circle Approximation

In Chapter 6 we looked at cubic Hermite approximation of curves in $\mathrm{I\!R}^2$ and cited [de Boor:87-2] to establish $O(h^6)$ convergence for cubic Hermite interpolation of curves with nonvanishing curvature. Ellipses are curves satisfying this curvature requirement. The paper introduced the first $O(h^6)$ circle interpolants. Then the paper [Dokken:90] introduced possibilities to model the shape of the error function by combining:

- The algebraic description of the circle being approximated.

- The parametric description of the cubic Hermite interpolant.

In the papers [Floater:95-1], [Floater:96] and [Floater:97] the following aspects of conic approximation are respectively addressed:

- High order approximation by quadratic splines.

- Cubic approximation schemes for conic sections.

- $O(h^{2n})$ Hermite approximation for conic sections.

In this chapter we use the idea from [Dokken:90] and combine it with the results of theorem 85 in Chapter 6 applied to ellipse approximation. Theorem 85 showed how we can utilize the algebraic and parametric description of a curve segment being approximated to control the cubic Hermite interpolation in such a way that $O(h^6)$ convergence is achieved.

The first step in the analyzes of ellipse approximation is to give the description of the ellipse segment and the cubic Hermite interpolant. When

discussing circle approximation the natural circle to use is the unit circle. The area of a unit circle is π; the area of an ellipse is πab where a and b are the half axis. By choosing the half axis to be b and $\frac{1}{b}$, we have an ellipse with the area of a unit circle. By scaling, translation and rotation any ellipses can be derived from this "unit ellipse". In addition for $b \equiv 1$ the "unit ellipse" is the unit circle.

The Ellipse Segment. The ellipse segment is described by

$$\mathbf{c}(\theta) = (b\cos\theta, \frac{1}{b}\sin\theta), \ \tau \le \theta \le \tau + \alpha \tag{7.1}$$

with $b > 0$. The algebraic description of the ellipse segment $\mathbf{c}(\theta)$ is $q(x, y) = 0$ with

$$q(x, y) = \left(\frac{x}{b}\right)^2 + (b\,y)^2 - 1. \tag{7.2}$$

Hermite Interpolant. The Hermite interpolant $\mathbf{p}(t)$ is required to match position and tangent of (7.1) for $\theta = \tau$ and $\theta = \tau + \alpha$. The description of $\mathbf{p}(t)$ in a cubic Bernstein basis is:

$$\mathbf{p}(t) = (x(t), y(t)) = \sum_{i=0}^{3} \mathbf{p}_i \binom{3}{i}(1-t)^{3-i}t^i, \tag{7.3}$$

where

$$\mathbf{p}_0 = \mathbf{c}(\tau) = \begin{pmatrix} b\cos(\tau) \\ \frac{1}{b}\sin\tau \end{pmatrix}$$

$$\mathbf{p}_1 = \mathbf{c}(\tau) + L_0(\alpha)\mathbf{c}'(\tau)$$
$$= \begin{pmatrix} b\cos(\tau) \\ \frac{1}{b}\sin(\tau) \end{pmatrix} + L_0(\alpha)\begin{pmatrix} -b\sin(\tau) \\ \frac{1}{b}\cos\tau \end{pmatrix}$$

$$\mathbf{p}_2 = \mathbf{c}(\tau+\alpha) - L_1(\alpha)\mathbf{c}'(\tau+\alpha)$$
$$= \begin{pmatrix} b\cos(\tau+\alpha) \\ \frac{1}{b}\sin(\tau+\alpha) \end{pmatrix} - L_1(\alpha)\begin{pmatrix} -b\sin(\tau+\alpha) \\ \frac{1}{b}\cos(\tau+\alpha) \end{pmatrix}$$

$$\mathbf{p}_3 = \mathbf{c}(\tau+\alpha) = \begin{pmatrix} b\cos(\tau+\alpha) \\ \frac{1}{b}\sin(\tau+\alpha) \end{pmatrix}.$$

The only unknown quantities in $\mathbf{p}(t)$ are the tangent length descriptions $L_0(\alpha)$ and $L_1(\alpha)$. The problem is thus to control $L_0(\alpha)$ and $L_1(\alpha)$ to get a best possible approximation.

The parametric and algebraic description of the ellipse segment satisfy the conditions of theorem 85. Thus, we can power expand the coefficient of

$q(\mathbf{p}(t))$ with respect to α. We know that the terms of degree lower than 6 vanish for proper choices of $L_0(\alpha)$ and $L_1(\alpha)$.

We structure this chapter as follows.

- In Section 7.1 we give a theorem that states when $L_0(\alpha)$ and $L_1(\alpha)$ satisfy

$$L_0(\alpha) = \frac{\alpha}{3} + \frac{\left(\frac{1}{24} + c_1\right)\alpha^3}{6} + \frac{c_2\alpha^4}{24} + O(\alpha^5) \qquad (7.4)$$

$$L_1(\alpha) = \frac{\alpha}{3} + \frac{\left(\frac{1}{24} - c_1\right)\alpha^3}{6} - \frac{c_2\alpha^4}{24} + O(\alpha^5), \qquad (7.5)$$

then the Hermite ellipse interpolant is $O(\alpha^6)$. The theorem also gives two alternative Taylor expansions with $O(\alpha^6)$ convergence. However, these have a much larger approximation error.

For circle approximation it is natural to require $L_0(\alpha) \equiv L_1(\alpha)$ to get a symmetric approximation. The section is concluded with a corollary stating that forcing $L_0(\alpha) \equiv L_1(\alpha)$ results in

$$L_0(\alpha) = L_1(\alpha) = \frac{\alpha}{3} + \frac{\alpha^3}{144} + O(\alpha^5). \qquad (7.6)$$

- In Section 7.2 we give examples of circle approximation methods satisfying (7.6).

- Examples of methods for ellipse approximation that satisfies (7.4) and (7.5) are then given in Section 7.3. The constants c_1 and c_2 are here used to model the approximation to give a desired shape. Also higher degree terms are added to (7.4) and (7.5) to give better shape control.

- We conclude the chapter in Section 7.4 by three $O(h^8)$ circle approximations based on a fourth degree interpolant that match position and tangent direction at the segment ends.

7.1 Taylor Expansion of Tangent Lengths

We address in this section conditions on the Taylor expansion of the tangent length functions to achieve $O(\alpha^6)$ convergence. The discussion is structured as follows:

- The first step is a lemma that express $q(\mathbf{p}(t))$ in a 6-th degree Bernstein basis.

- Then the main result of this subsection is stated in theorem 90. Here three different pairs of Taylor expansions of $L_0(\alpha)$ and $L_1(\alpha)$ are given that give an $O(\alpha^6)$ convergence of the ellipse approximation.

- Then we discuss the error of these and conclude that the one referenced in (7.4) and (7.5) has the smallest error functions.

- In circle approximation, symmetry in the approximation is often required, thus demanding that $L_0(\alpha) \equiv L_1(\alpha)$. In corollary 91 we look at this problem and deduct the tangent length function in (7.6).

To simplify the theorem to follow, we give the following lemma expressing $q(\mathbf{p}(t))$ in a 6-th degree Bernstein basis.

Lemma 89 *Let the Hermite interpolant $\mathbf{p}(t)$ to the ellipse segment be described as in (7.3) and let the algebraic description of the ellipse be $q(\mathbf{x}) = 0$ as describe in (7.2) then*

$$q(\mathbf{p}(t)) = 15a(\alpha)(1-t)^4 t^2 + 20b(\alpha)(1-t)^3 t^3 + 15c(\alpha)(1-t)^2 t^4, \quad (7.7)$$

where

$$a(\alpha) = \frac{3L_0(\alpha)^2 + 2L_1(\alpha)\sin\alpha - 2(1 - \cos\alpha)}{5}$$

$$b(\alpha) = \frac{-9L_1(\alpha)L_0(\alpha)\cos\alpha + 9(L_1(\alpha) + L_0(\alpha))\sin\alpha - 10(1 - \cos\alpha)}{10}$$

$$c(\alpha) = \frac{3L_1(\alpha)^2 + 2L_0(\alpha)\sin\alpha - 2(1 - \cos\alpha)}{5}.$$

$$(7.8)$$

Proof. Can be verified by expanding the expressions. ∎

The theorem following shows that three different pairs of Taylor expansion exists for $L_0(\alpha)$ and $L_1(\alpha)$ that give an $O(\alpha^6)$ convergence of the error function.

Theorem 90 *Let the cubic Hermite interpolant $\mathbf{p}(t)$ and the ellipse segment $\mathbf{c}(\alpha)$ be defined as in lemma 89. If the tangent length functions $L_0(\alpha)$ and $L_1(\alpha)$ satisfy either*

$$L_0(\alpha) = \frac{\alpha}{3} + \frac{\left(\frac{1}{24} + c_1\right)\alpha^3}{6} + \frac{c_2\alpha^4}{24} + O(\alpha^5)$$

$$L_1(\alpha) = \frac{\alpha}{3} + \frac{\left(\frac{1}{24} - c_1\right)\alpha^3}{6} - \frac{c_2\alpha^4}{24} + O(\alpha^5),$$

or

$$L_0(\alpha) = \frac{\alpha}{3} + \frac{1}{2\sqrt{3}}\alpha^2 + \frac{\left(-\frac{1}{3}+c_1\right)\alpha^3}{6} + \frac{\left(-\sqrt{3}c_1+c_2\right)\alpha^4}{24} + O(\alpha^5)$$

$$L_1(\alpha) = \frac{\alpha}{3} - \frac{1}{2\sqrt{3}}\alpha^2 + \frac{\left(-\frac{1}{3}-c_1\right)\alpha^3}{6} + \frac{\left(-\sqrt{3}c_1-c_2\right)\alpha^4}{24} + O(\alpha^5),$$

or

$$L_0(\alpha) = \frac{\alpha}{3} - \frac{1}{2\sqrt{3}}\alpha^2 + \frac{\left(-\frac{1}{3}+c_1\right)\alpha^3}{6} + \frac{\left(\sqrt{3}c_1+c_2\right)\alpha^4}{24} + O(\alpha^5)$$

$$L_1(\alpha) = \frac{\alpha}{3} + \frac{1}{2\sqrt{3}}\alpha^2 + \frac{\left(-\frac{1}{3}-c_1\right)\alpha^3}{6} + \frac{\left(\sqrt{3}c_1-c_2\right)\alpha^4}{24} + O(\alpha^5),$$

then the approximation $\mathbf{p}(t)$ *to the ellipse segment is* $O(\alpha^6)$.

The corresponding Taylor expansions of $a(\alpha)$, $b(\alpha)$ *and* $c(\alpha)$ *defined in (7.8) are respectively*

$$a(\alpha) = \left(\frac{7}{57600} + \frac{1}{80}c_1 + \frac{1}{60}c_1^2 + \frac{1}{300}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(\frac{1}{320}c_2 + \frac{1}{120}c_2c_1 + \frac{1}{1800}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right)$$

$$b(\alpha) = \left(-\frac{23}{38400} + \frac{1}{40}c_1^2 + \frac{1}{200}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(\frac{1}{80}c_2c_1 + \frac{1}{1200}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right)$$

$$c(\alpha) = \left(\frac{7}{57600} - \frac{1}{80}c_1 + \frac{1}{60}c_1^2 + \frac{1}{300}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(-\frac{1}{320}c_2 + \frac{1}{120}c_2c_1 + \frac{1}{1800}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right),$$

and

$$a(\alpha) = \left(\frac{11}{1800} - \frac{1}{40}c_1 + \frac{\sqrt{3}}{120}c_2 + \frac{1}{60}c_1^2 + \frac{1}{300}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(\frac{\sqrt{3}}{180}c_1 - \frac{\sqrt{3}}{120}c_1^2 + \frac{1}{120}c_2c_1 - \frac{\sqrt{3}}{1800} + \frac{\sqrt{3}}{600}L_0^{(5)} + \frac{1}{1800}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right)$$

$$b(\alpha) = \left(-\frac{49}{1200} + \frac{\sqrt{3}}{80}c_2 + \frac{1}{40}c_1^2 + \frac{1}{200}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(-\frac{7\sqrt{3}}{240}c_1 + \frac{1}{80}c_2c_1 + \frac{\sqrt{3}}{800}\left(L_0^{(5)} - L_1^{(5)}\right) + \frac{1}{1200}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right)$$

$$c(\alpha) = \left(\frac{11}{1800} + \frac{1}{40}c_1 + \frac{\sqrt{3}}{120}c_2 + \frac{1}{60}c_1^2 + \frac{1}{300}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(\frac{\sqrt{3}}{180}c_1 + \frac{\sqrt{3}}{120}c_1^2 + \frac{1}{120}c_2c_1 + \frac{\sqrt{3}}{1800} - \frac{\sqrt{3}}{600}L_1^{(5)} + \frac{1}{1800}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right),$$

and

$$a(\alpha) = \left(\tfrac{11}{1800} - \tfrac{1}{40}c_1 - \tfrac{\sqrt{3}}{120}c_2 + \tfrac{1}{60}c_1^2 + \tfrac{1}{300}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(-\tfrac{\sqrt{3}}{180}c_1 + \tfrac{\sqrt{3}}{120}c_1^2 + \tfrac{1}{120}c_2 c_1 + \tfrac{\sqrt{3}}{1800} - \tfrac{\sqrt{3}}{600}L_0^{(5)} + \tfrac{1}{1800}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right)$$

$$b(\alpha) = \left(-\tfrac{49}{1200} - \tfrac{\sqrt{3}}{80}c_2 + \tfrac{1}{40}c_1^2 + \tfrac{1}{200}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(\tfrac{7\sqrt{3}}{240}c_1 + \tfrac{1}{80}c_2 c_1 + \tfrac{\sqrt{3}}{800}\left(L_1^{(5)} - L_0^5\right) + \tfrac{1}{1200}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right)$$

$$c(\alpha) = \left(\tfrac{11}{1800} + \tfrac{1}{40}c_1 - \tfrac{\sqrt{3}}{120}c_2 + \tfrac{1}{60}c_1^2 + \tfrac{1}{300}\left(L_0^{(5)} + L_1^{(5)}\right)\right)\alpha^6$$
$$+ \left(-\tfrac{\sqrt{3}}{180}c_1 - \tfrac{\sqrt{3}}{120}c_1^2 + \tfrac{1}{120}c_1 c_2 - \tfrac{\sqrt{3}}{1800} + \tfrac{\sqrt{3}}{600}L_1^{(5)} + \tfrac{1}{1800}\left(L_0^{(6)} + L_1^{(6)}\right)\right)\alpha^7 + O\left(\alpha^8\right).$$

Proof. Since we already know than an $O(\alpha^6)$ approximation exists, we just have to find conditions for $L_0(\alpha)$ and $L_1(\alpha)$ that ensure that the terms of degree 5 or lower in the Taylor expansion of $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$ vanish. The proof has the following steps:

- First we make the Taylor expansion with seven terms of $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$. Here we notice that the only unknown expressions are derivatives of $L_0(\alpha)$ and $L_1(\alpha)$ up to order 5 for $\alpha = 0$.

- Then we successively look at conditions that make the terms of degree five or lower in the Taylor expansion $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$ vanish.

Taylor expanding respectively $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$ to 7 terms we get:

$$a(\alpha) = \left(\tfrac{3}{5}L_0(0)^2\right) + \left(\tfrac{6}{5}L_0(0)L_0'(0) + \tfrac{2}{5}L_1(0)\right)\alpha$$
$$+ \left(\tfrac{3}{5}L_0(0)L_0''(0) + \tfrac{3}{5}L_0'(0)^2 - \tfrac{1}{5} + \tfrac{2}{5}L_1'(0)\right)\alpha^2$$
$$+ \left(\tfrac{1}{5}L_0(0)L_0'''(0) + \tfrac{3}{5}L_0'(0)L_0''(0) - \tfrac{1}{15}L_1(0) + \tfrac{1}{5}L_1''(0)\right)\alpha^3$$
$$+ \left(\begin{array}{l} -\tfrac{1}{15}L_1'(0) + \tfrac{1}{15}L_1'''(0) + \tfrac{1}{60} + \tfrac{3}{20}L_0''(0)^2 \\ \qquad + \tfrac{1}{5}L_0'(0)L_0'''(0) + \tfrac{1}{20}L_0(0)L_0^{(4)}(0)\end{array}\right)\alpha^4$$
$$+ \left(\begin{array}{l} \tfrac{1}{10}L_0''(0)L_0'''(0) + \tfrac{1}{20}L_0^{(4)}(0)L_0'(0) + \tfrac{1}{100}L_0(0)L_0^{(5)}(0) \\ \qquad + \tfrac{1}{60}L_1^{(4)}(0) + \tfrac{1}{300}L_1(0) - \tfrac{1}{30}L_1''(0)\end{array}\right)\alpha^5$$
$$+ O(\alpha^6),$$

and

$$
\begin{aligned}
b(\alpha) \;=\; & \left(-\tfrac{9}{10}L_1(0)\,L_0(0)\right) \\
& + \left(-\tfrac{9}{10}L_1(0)\,L_0'(0) - \tfrac{9}{10}L_1'(0)\,L_0(0) + \tfrac{9}{10}L_1(0) + \tfrac{9}{10}L_0(0)\right)\alpha \\
& + \begin{aligned}[t] &\left(-\tfrac{9}{20}L_1(0)\,L_0''(0) - \tfrac{9}{20}L_1''(0)\,L_0(0) - \tfrac{9}{10}L_1'(0)\,L_0'(0)\right.\\ &\left.+\tfrac{9}{20}L_1(0)\,L_0(0) - \tfrac{1}{2} + \tfrac{9}{10}L_0'(0) + \tfrac{9}{10}L_1'(0)\right)\alpha^2 \end{aligned} \\
& + \begin{aligned}[t] &\left(\tfrac{9}{20}L_1(0)\,L_0'(0) + \tfrac{9}{20}L_1'(0)\,L_0(0) - \tfrac{3}{20}L_1(0)\,L_0'''(0)\right.\\ &- \tfrac{9}{20}L_1'(0)\,L_0''(0) - \tfrac{3}{20}L_1'''(0)\,L_0(0) - \tfrac{9}{20}L_1''(0)\,L_0'(0)\\ &\left.- \tfrac{3}{20}L_1(0) - \tfrac{3}{20}L_0(0) + \tfrac{9}{20}L_0''(0) + \tfrac{9}{20}L_1''(0)\right)\alpha^3 \end{aligned} \\
& + \begin{aligned}[t] &\left(-\tfrac{3}{20}L_0'(0) - \tfrac{3}{20}L_1'(0) + \tfrac{3}{20}L_0'''(0) + \tfrac{3}{20}L_1'''(0) + \tfrac{1}{24}\right.\\ &+\tfrac{9}{40}L_1(0)\,L_0''(0) + \tfrac{9}{40}L_1''(0)\,L_0(0) + \tfrac{9}{20}L_1'(0)\,L_0'(0)\\ &- \tfrac{3}{80}L_1(0)\,L_0(0) - \tfrac{9}{40}L_1''(0)\,L_0''(0) - \tfrac{3}{20}L_1'(0)\,L_0'''(0)\\ &\left.- \tfrac{3}{80}L_1(0)\,L_0^{(4)}(0) - \tfrac{3}{80}L_1^{(4)}(0)\,L_0(0) - \tfrac{3}{20}L_1'''(0)\,L_0'(0)\right)\alpha^4 \end{aligned} \\
& + \begin{aligned}[t] &\left(-\tfrac{3}{40}L_1''(0)\,L_0'''(0) - \tfrac{3}{80}L_1^{(4)}(0)\,L_0'(0) - \tfrac{3}{400}L_1(0)\,L_0^{(5)}(0)\right.\\ &- \tfrac{3}{80}L_1'(0)\,L_0^{(4)}(0) - \tfrac{3}{400}L_1^{(5)}(0)\,L_0(0) - \tfrac{3}{40}L_1'''(0)\,L_0''(0)\\ &- \tfrac{3}{80}L_1(0)\,L_0'(0) - \tfrac{3}{80}L_1'(0)\,L_0(0) + \tfrac{3}{40}L_1(0)\,L_0'''(0)\\ &+\tfrac{9}{40}L_1'(0)\,L_0''(0) + \tfrac{3}{40}L_1'''(0)\,L_0(0) + \tfrac{9}{40}L_1''(0)\,L_0'(0)\\ &+\tfrac{3}{80}L_1^{(4)}(0) + \tfrac{3}{80}L_0^{(4)}(0) + \tfrac{3}{400}L_1(0)\\ &\left.+\tfrac{3}{400}L_0(0) - \tfrac{3}{40}L_0''(0) - \tfrac{3}{40}L_1''(0)\right)\alpha^5 \end{aligned} \\
& +O(\alpha^6),
\end{aligned}
$$

and

$$
\begin{aligned}
c(\alpha) \;=\; & \left(\tfrac{3}{5}L_1(0)^2\right) + \left(\tfrac{6}{5}L_1(0)\,L_1'(0) + \tfrac{2}{5}L_0(0)\right)\alpha \\
& + \left(\tfrac{3}{5}L_1(0)\,L_1''(0) + \tfrac{3}{5}L_1'(0)^2 - \tfrac{1}{5} + \tfrac{2}{5}L_0'(0)\right)\alpha^2 \\
& + \left(\tfrac{1}{5}L_1(0)\,L_1'''(0) + \tfrac{3}{5}L_1'(0)\,L_1''(0) - \tfrac{1}{15}L_0(0) + \tfrac{1}{5}L_0''(0)\right)\alpha^3 \\
& + \begin{aligned}[t] &\left(-\tfrac{1}{15}L_0'(0) + \tfrac{1}{15}L_0'''(0) + \tfrac{1}{60}\right.\\ &\left.+\tfrac{3}{20}L_1''(0)^2 + \tfrac{1}{5}L_1'(0)\,L_1'''(0) + \tfrac{1}{20}L_1(0)\,L_1^{(4)}(0)\right)\alpha^4 \end{aligned} \\
& + \begin{aligned}[t] &\left(\tfrac{1}{10}L_1''(0)\,L_1'''(0) + \tfrac{1}{20}L_1^{(4)}(0)\,L_1'(0)\right.\\ &\left.+\tfrac{1}{100}L_1(0)\,L_1^{(5)}(0) + \tfrac{1}{60}L_0^{(4)}(0) + \tfrac{1}{300}L_0(0) - \tfrac{1}{30}L_0''(0)\right)\alpha^5 \end{aligned} \\
& +O(\alpha^6).
\end{aligned}
$$

Now we successively look at terms in the expansions of higher and higher degree to establish conditions that make the terms of degree five or lower

vanish. First we look at the coefficients of the terms of degree zero

$$\left(\frac{3}{5}L_0\left(0\right)^2\right) = 0$$

$$\left(-\frac{9}{10}L_1\left(0\right)L_0\left(0\right)\right) = 0$$

$$\left(\frac{3}{5}L_1\left(0\right)^2\right) = 0.$$

We see that only assigning $L_0\left(0\right) = 0$ and $L_1\left(0\right) = 0$ set the coefficients of the zero degree terms to 0. Remember these values for use later in the proof. The first degree terms are all zero by these assignments. The second degree terms are reduced to

$$\left(\frac{3}{5}L_0'\left(0\right)^2 - \frac{1}{5} + \frac{2}{5}L_1'\left(0\right)\right) = 0$$

$$\left(-\frac{9}{10}L_1'\left(0\right)L_0'\left(0\right) - \frac{1}{2} + \frac{9}{10}L_0'\left(0\right) + \frac{9}{10}L_1'\left(0\right)\right) = 0$$

$$\left(\frac{3}{5}L_1'\left(0\right)^2 - \frac{1}{5} + \frac{2}{5}L_0'\left(0\right)\right) = 0.$$

These equations only vanish when $L_0'(0) = L_1'(0) = \frac{1}{3}$. Now the third degree terms are reduced to the condition

$$L_0''\left(0\right) + L_1''\left(0\right) = 0.$$

Further all the fourth degree terms are reduced to the condition

$$L_0'''\left(0\right) + L_1'''\left(0\right) = \frac{1}{12} - \frac{9}{4}\left(L_0''\left(0\right)\right)^2,$$

and the fifth degree terms reduce to

$$6L_0''\left(0\right)L_0'''\left(0\right) + L_0^{(4)}\left(0\right) + L_1^{(4)}\left(0\right) + 2L_0''\left(0\right) = 0$$
$$3L_0''\left(0\right)\left(L_0'''\left(0\right) - L_1'''\left(0\right)\right) + L_1^{(4)}\left(0\right) + L_0^{(4)}\left(0\right) = 0$$
$$-6L_0''\left(0\right)L_1'''\left(0\right) + L_1^{(4)}\left(0\right) + L_0^{(4)}\left(0\right) - 2L_0''\left(0\right) = 0.$$

These three equations are linearly dependent, and we can substract the last from the first to get:

$$L_0''\left(0\right)\left(L_0'''\left(0\right) + L_1'''\left(0\right) + \frac{2}{3}\right) = 0.$$

144

Combining with the condition from the fourth degree terms, we get

$$L_0'' (0) \left(\frac{1}{12} - \frac{9}{4} (L_0'' (0))^2 + \frac{2}{3} \right) = 0,$$

resulting in

$$L_0'' (0) = 0 \text{ or } L_0'' (0) = \pm \frac{1}{\sqrt{3}}.$$

Above in the proof we have already found that

$$
\begin{aligned}
L_0(0) &= L_1(0) = 0 \\
L_0'(0) &= L_1'(0) = \frac{1}{3}.
\end{aligned}
$$

Analyzing the above three alternatives for $L_0'' (0)$, and applying these on the forth and fifth degree terms, we get remembering that $L_0''(0) + L_1''(0) = 0$:

1. First alternative $L_0'' (0) = L_1'' (0) = 0$, giving the following requirements

$$
\begin{aligned}
L_0''' (0) + L_1''' (0) &= \frac{1}{12} \\
L_0^{(4)} (0) + L_1^{(4)} (0) &= 0.
\end{aligned}
$$

   These are satisfied by

$$
\begin{aligned}
L_0(\alpha) &= \frac{\alpha}{3} + \frac{\left(\frac{1}{24} + c_1 \right) \alpha^3}{6} + \frac{c_2 \alpha^4}{24} \\
L_1(\alpha) &= \frac{\alpha}{3} + \frac{\left(\frac{1}{24} - c_1 \right) \alpha^3}{6} - \frac{c_2 \alpha^4}{24}.
\end{aligned}
$$

2. Second alternative $L_0'' (0) = \frac{1}{\sqrt{3}}$ and $L_1'' (0) = -\frac{1}{\sqrt{3}}$, giving the following requirements

$$
\begin{aligned}
L_0''' (0) + L_1''' (0) &= -\frac{2}{3} \\
\sqrt{3} \left(L_0''' (0) - L_1''' (0) \right) + L_1^{(4)} (0) + L_0^{(4)} (0) &= 0.
\end{aligned}
$$

   These are satisfied by

$$
\begin{aligned}
L_0(\alpha) &= \frac{\alpha}{3} + \frac{1}{2\sqrt{3}} \alpha^2 + \frac{\left(-\frac{1}{3} + c_1 \right) \alpha^3}{6} + \frac{\left(-\sqrt{3}c_1 + c_2 \right) \alpha^4}{24} \\
L_1(\alpha) &= \frac{\alpha}{3} - \frac{1}{2\sqrt{3}} \alpha^2 + \frac{\left(-\frac{1}{3} - c_1 \right) \alpha^3}{6} + \frac{\left(-\sqrt{3}c_1 - c_2 \right) \alpha^4}{24}.
\end{aligned}
$$

3. Third alternative $L_0''(0) = -\frac{1}{\sqrt{3}}$ and $L_1''(0) = \frac{1}{\sqrt{3}}$, giving the following requirements

$$L_0'''(0) + L_1'''(0) = -\frac{2}{3}$$
$$-\sqrt{3}\left(L_0'''(0) - L_1'''(0)\right) + L_1^{(4)}(0) + L_0^{(4)}(0) = 0.$$

These are satisfied by

$$L_0(\alpha) = \frac{\alpha}{3} - \frac{1}{2\sqrt{3}}\alpha^2 + \frac{\left(-\frac{1}{3} + c_1\right)\alpha^3}{6} + \frac{\left(\sqrt{3}c_1 + c_2\right)\alpha^4}{24}$$

$$L_1(\alpha) = \frac{\alpha}{3} + \frac{1}{2\sqrt{3}}\alpha^2 + \frac{\left(-\frac{1}{3} - c_1\right)\alpha^3}{6} + \frac{\left(\sqrt{3}c_1 - c_2\right)\alpha^4}{24}.$$

Higher order terms in $L_0(\alpha)$ and $L_1(\alpha)$ do not influence the annihilation of the low order terms, thus we can add a $O(\alpha^5)$ term to the expansion of $L_0(\alpha)$ and $L_1(\alpha)$.

The Taylor expansions of $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$ are now straight forward by using the proper pairs of Taylor expansion of $L_0(\alpha)$ and $L_1(\alpha)$. ∎

Looking at the sixth order terms of the alternative Taylor expansions of $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$, we see that the constants in the first expansion are much smaller than in the other two, thus we prefer to concentrate the examples in the next section on the first alternative. Tests have shown a much smaller error function when using the first alternative compared to the others. For circle approximation the first alternative is the one giving the best approximation. In addition symmetry of tangent lengths for circle approximation is impossible with the second and third alternative.

In circle approximation it is natural to require symmetry of tangent length and thus requiring $L_0(\alpha) \equiv L_1(\alpha)$. The following corollary presents the consequence of this simplification.

Corollary 91 *Let the approximation problem be defined as in theorem 90 and in addition impose that*

$$L_0(\alpha) \equiv L_1(\alpha) \equiv L(\alpha).$$

Then to achieve $O(h^6)$ convergence $L(\alpha)$ has to satisfy

$$L(\alpha) = \frac{\alpha}{3} + \frac{\alpha^3}{144} + O(\alpha^5). \tag{7.9}$$

In addition $a(\alpha) \equiv c(\alpha)$. The Taylor expansion of $a(\alpha)$ and $b(\alpha)$ satisfy

$$a(\alpha) = \left(\frac{7}{57600} + \frac{1}{150}L^{(5)}(0)\right)\alpha^6 + O(\alpha^7) \tag{7.10}$$
$$b(\alpha) = \left(-\frac{103}{38400} + \frac{29}{2400}L^{(5)}(0)\right)\alpha^6 + O(\alpha^7). \tag{7.11}$$

Circle approximation	Example	Inside error	Outside error
Simplest	94	-6.6×10^{-5}	3.4×10^{-6}
One sided outside	95	0	1.9×10^{-5}
Equioccilating	96	-1.4×10^{-5}	1.4×10^{-5}
Flat	98	0	6.2×10^{-5}
Zero integral	100	-1.9×10^{-5},	1.2×10^{-5}
Minimum square sum	101	0	2.4×10^{-5}

Table 7.1: The table shows a summary of different cubic Hermit circle approximations and how the error behaves. The approximation with least variation of error is the "one sided outside" that has an error variation of 1.9×10^{-5}, the "equioccilating" has a variation in error 2.8×10^{-5}. The circle segment approximateded is a segment with opening angle 1 of a unit circle.

Proof. To get the required symmetry $c_1 = c_2 = 0$. This gives the Taylor expansion of the tangent length function in (7.9). The Taylor expansion of $a(\alpha)$ and $b(\alpha)$ is now straight forward. ∎

7.2 Some Cubic Hermite Circle Interpolants

The background for the work on circle approximation methods described in this section was accuracy requirements to the spline library SISL developed at SINTEF in Oslo in 1988-1989. The basic idea behind these methods is to model the shape of the error function of the cubic Hermit interpolants, and is described in [Dokken:90]. In [Moerken:91-2] Quadratic Interpolation of circles is addressed, while in [Lyche:94] these ideas are extended to odd degrees n and an error function $e(t) = O(t^{2n})$. The goal in this section is to combine a required geometric shape in the error function and a high convergence rate for cubic Hermite interpolants.

In corollary 91 we addressed conditions to be imposed on the Taylor expansion of the tangent length at the start and end of a cubic Hermit circle interpolant to achieve $O(\alpha^6)$ convergence. This is the basis for a family of $O(\alpha^6)$ circle interpolants. We illustrate the flexibility of this family by the following examples:

- The near equioscillating Hermite circle approximation with two internal zeroes.

- The one sided circle approximation with a double zero in the middle.

- The circle approximation with simplest Taylor expansion of the tangent length function.

- An $O(\alpha^6)$ circle approximation with no internal zeroes.

- A circle approximation with zero integral of $q(\mathbf{p}(t))$.

- A circle approximation with a minimum of the square of the Bezier coefficients of $q(\mathbf{p}(t))$. Also in this there are no internal zeroes.

For a brief overview of the methods table 7.1 summarizes the error variation of the methods. As usual with approximation methods, which method is the best depends upon the actual use of the approximation.

The Hermite interpolant we use, is described in (7.3). Since we have $L_1(\alpha) \equiv L_2(\alpha) \equiv L(\alpha)$ as in corollary 91, we here get a somewhat simpler description of the Hermite interpolant than on page 138.

$$\mathbf{p}(t) = (x(t), y(t)) = \sum_{i=0}^{3} \mathbf{p}_i \binom{3}{i}(1-t)^{3-i}t^i, \qquad (7.12)$$

where

$$\mathbf{p}_0 = \mathbf{c}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$\mathbf{p}_1 = \mathbf{c}(0) + L(\alpha)\mathbf{c}'(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix} + L(\alpha)\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$\mathbf{p}_2 = \mathbf{c}(\alpha) - L(\alpha)\mathbf{c}'(\alpha) = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix} - L(\alpha)\begin{pmatrix} -\sin\alpha \\ \cos\alpha \end{pmatrix}$$

$$\mathbf{p}_3 = \mathbf{c}(\alpha) = \begin{pmatrix} \cos\alpha \\ \sin\alpha \end{pmatrix}.$$

In this case the expression $q(\mathbf{p}(t))$ reduces to

$$q(\mathbf{p}(t)) = 15a(\alpha)(1-t)^4t^2 + 20b(\alpha)(1-t)^3t^3 + 15a(\alpha)(1-t)^2t^4,$$

where

$$a(\alpha) = \tfrac{1}{5}(3L(\alpha)^2 + 2L(\alpha)\sin\alpha - 2(1-\cos\alpha)) \qquad (7.13)$$
$$b(\alpha) = \tfrac{1}{10}(-9L(\alpha)^2\cos\alpha + 18L(\alpha)\sin\alpha - 10(1-\cos\alpha)). \qquad (7.14)$$

These values for $a(\alpha)$ and $b(\alpha)$ can also be found in [Dokken:90]. For the approximation to be $O(\alpha^6)$, corollary 91 stated that

$$L(\alpha) = \frac{\alpha}{3} + \frac{\alpha^3}{144} + O(\alpha^5).$$

148

One requirement that has to be satisfied for the interpolant to be $O(\alpha^6)$, was that the gradient of the algebraic curve must be nonvanishing close to the curve being approximated. For a circle the gradient is only vanishing at the circle center. Thus, imposing that $0 < \alpha < \pi$, and only allow positive values for $L(\alpha)$, keeps the approximation away from the origin.

The algebraic description of the unit circle, that is used in this section, is $q(x, y) = 0$ with

$$q(x, y) = x^2 + y^2 - 1. \tag{7.15}$$

Before giving examples, we give a theorem that connects the behavior of $q(\mathbf{p}(t))$ to the approximation error. The theorem also shows that the maximum value of $\frac{|q(\mathbf{p}(t))|}{2}$ is a good error estimate for the circle approximations to be used in the examples. When we say error function in the example in this section, and in Section 7.4, we use this error estimate.

Theorem 92 *Let the Hermite interpolant $\mathbf{p}(t)$ to the circle segment be described as in (7.12) and let the algebraic description of the unit circle $q(\mathbf{x}) = 0$ be described as in (7.15). Provided that the error*

$$\rho(t) = \sqrt{x(t)^2 + y(t)^2} - 1$$

in the circle approximation satisfy

$$\rho_{\max} = \max_{0 \leq t \leq 1} |\rho(t)| < 1, \tag{7.16}$$

then

$$|\rho(t)| \leq \frac{|q(\mathbf{p}(t))|}{2(1 - \rho_{\max})}. \tag{7.17}$$

Proof. The Hermite interpolant $\mathbf{p}(t)$ can be decomposed such that $\mathbf{p}(t) = \mathbf{p}_0(t) + \rho(t)\mathbf{n}_0(t)$, where $q(\mathbf{p}_0(t)) = 0$ is an exact representation of the circle segment and $\mathbf{n}_0(t)$ is the unit normal of the circle segment at $\mathbf{p}_0(t)$. Using this decomposition and Taylor expanding to two terms with respect to $\rho(t)$ we get

$$
\begin{aligned}
q(\mathbf{p}(t)) &= q(\mathbf{p}_0(t) + \rho(t)\mathbf{n}_0(t)) \\
&= q(\mathbf{p}_0(t)) + \nabla q(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t)) \cdot \mathbf{n}_0(t))\rho(t) \\
&= \nabla q(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t)) \cdot \mathbf{n}_0(t))\rho(t).
\end{aligned}
$$

Here $\theta(t)\rho(t) \geq 0$, $|\theta(t)| \leq |\rho(t)|$. We have that $\nabla q(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t))$ is parallel to $\mathbf{n}_0(t)$, resulting in

$$|q(\mathbf{p}(t))| = \|\nabla q(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t))\|_2 |\rho(t)|.$$
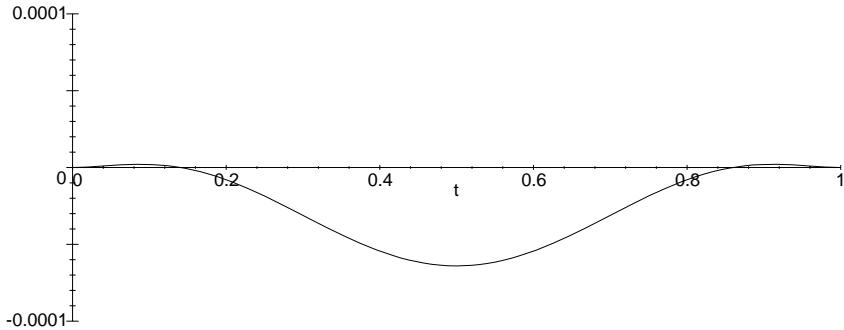
Figure 7.1: Error of the simplest circle approximation with convergence rate $O(\alpha^6)$. The circle segment approximated has opening angle $\alpha = 1$ and unit radius.

Now assuming that $\nabla q(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t)) \neq \mathbf{0}$, we get

$$|\rho(t)| = \frac{|q(\mathbf{p}(t))|}{\|\nabla q(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t))\|_2}. \tag{7.18}$$

Going back to the definition of $q(\mathbf{x}) = 0$ in (7.15), we get

$$\nabla q\left(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t)\right) = 2\mathbf{n}_0(t)(1 + \theta(t)).$$

Thus

$$\|\nabla\left(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t)\right)\|_2 = 2\left|1 + \theta(t)\right|.$$

Since $|\theta(t)| \leq |\rho(t)| \leq \rho_{max} < 1$, we get

$$\|\nabla\left(\mathbf{p}_0(t) + \theta(t)\mathbf{n}_0(t)\right)\|_2 > 2(1 - \rho_{max}).$$

Inserting this into (7.18) we get (7.17). ∎

Corollary 93 *All circle approximations methods defined as in (7.12) satisfying conditions (7.9) and (7.16) are $O(\alpha^6)$.*

Proof. Corollary 91 establish $O(\alpha^6)$, while theorem 92 connects the convergence rate to the error. ∎

The first example is the circle approximation with the simplest Taylor expansion of the tangent length function $L(\alpha)$.
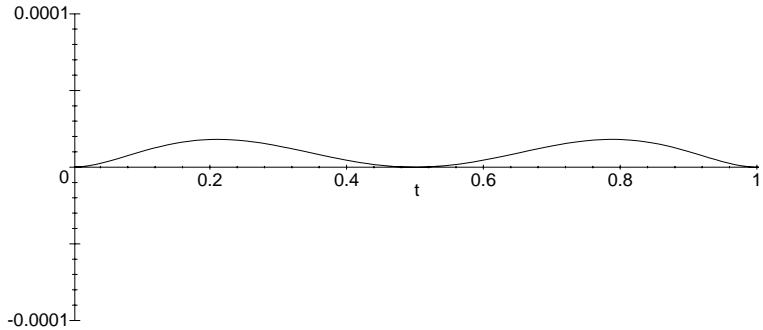
Figure 7.2: The error for the outside one sided approximation for opening angle $\alpha = 1$ and unit radius.

Example 94 (Simplest Circle Approximation.) *We chose*

$$L(\alpha) = \frac{1}{3}\alpha + \frac{1}{144}\alpha^3.$$

By corollary 91 this is a circle approximation with an $O(\alpha^6)$ convergence rate. The values of $a(\alpha)$ and $b(\alpha)$ and their corresponding Taylor expansions to eleven terms are:

$$
\begin{aligned}
a(\alpha) &= \tfrac{1}{15}\alpha^2 + \tfrac{1}{360}\alpha^4 + \tfrac{1}{34560}\alpha^6 + \tfrac{2}{15}\left(\sin\alpha\right)\alpha + \tfrac{1}{360}\left(\sin\alpha\right)\alpha^3 - \tfrac{2}{5} + \tfrac{2}{5}\cos\alpha \\
&= \tfrac{7}{57600}\alpha^6 + \tfrac{1}{151200}\alpha^8 + O(\alpha^{10}) \\
b(\alpha) &= -\tfrac{1}{10}\left(\cos\alpha\right)\alpha^2 - \tfrac{1}{240}\left(\cos\alpha\right)\alpha^4 - \tfrac{1}{23040}\left(\cos\alpha\right)\alpha^6 \\
&\qquad\qquad + \tfrac{3}{5}\left(\sin\alpha\right)\alpha + \tfrac{1}{80}\left(\sin\alpha\right)\alpha^3 - 1 + \cos\alpha \\
&= -\tfrac{23}{38400}\alpha^6 - \tfrac{1}{322560}\alpha^8 + O\left(\alpha^{10}\right).
\end{aligned}
$$

Note that the coefficients are $O(\alpha^6)$. The error function for $\alpha = 1$ is plotted in figure 7.1. Note that $L(\alpha) > 0$ for $\alpha > 0$.

Now two different circle approximation methods from [Dokken:90] are addressed.

Example 95 (One Sided Outside Circle Approximation.) *In the article [Dokken:90] an one sided outside interpolant was established by choosing $L(\alpha) = \frac{4}{3}\tan\frac{1}{4}\alpha$. The values and Taylor expansion of respectively $L(\alpha)$, $a(\alpha)$*
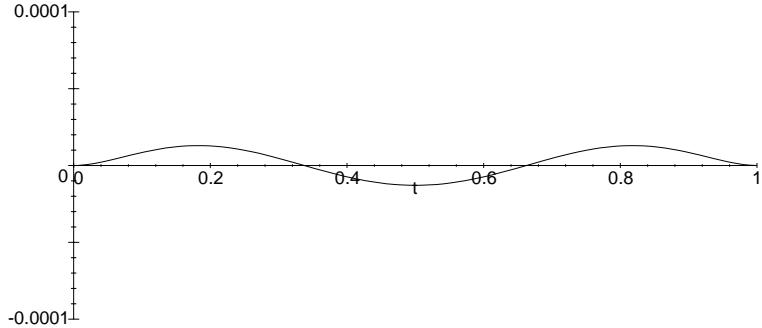
Figure 7.3: The the error of the near equioscillating approximant for opening angle $\alpha = 1$ and unit radius.

and $b(\alpha)$ are:

$$
\begin{aligned}
L(\alpha) &= \tfrac{4}{3}\tan\tfrac{1}{4}\alpha \\
&= \tfrac{1}{3}\alpha + \tfrac{1}{144}\alpha^3 + \tfrac{1}{5760}\alpha^5 + O\left(\alpha^7\right) \\
a(\alpha) &= \tfrac{16}{15}\tan^2\tfrac{1}{4}\alpha + \tfrac{8}{15}\tan\tfrac{1}{4}\alpha\sin\alpha - \tfrac{2}{5} + \tfrac{2}{5}\cos\alpha \\
&= \tfrac{1}{3840}\alpha^6 + O\left(\alpha^{10}\right)
\end{aligned}
$$

$$
\begin{aligned}
b(\alpha) &= -\tfrac{3}{2}a(\alpha) \\
&= -\tfrac{1}{2560}\alpha^6 + O\left(\alpha^{10}\right).
\end{aligned}
$$

The error of the one sided circle interpolant is shown in figure 7.2 for $\alpha = 1$. Note that $L(\alpha) > 0$ for $0 < \alpha < 2\pi$. Note that all the terms of degree 7, 8 and 9 in the Taylor expansion of $a(\alpha)$ and $b(\alpha)$ vanish.

Example 96 (Equioscillating Circle Approximation.) *An interpolant with three equioscillations of $q(\mathbf{p}(t))$ was established in [Dokken:90] by choosing*

$$
\begin{aligned}
L(\alpha) &= \frac{(9-2K)\sin\alpha - \sqrt{((9-2K)\sin\alpha)^2 - 6(2K+3\cos\alpha)(5-2K)(1-\cos\alpha)}}{3(2K+3\cos\alpha)} \\
K &= \tfrac{1}{2} - \sqrt[3]{3 - 2\sqrt{2}} - \sqrt[3]{3 + 2\sqrt{2}}.
\end{aligned}
$$

Taylor expanding in Maple, we get

$$
\begin{aligned}
L(\alpha) &= \tfrac{1}{3}\alpha + \tfrac{1}{144}\alpha^3 + 1.39140882187655 \times 10^{-4}\alpha^5 + O\left(\alpha^7\right) \\
a(\alpha) &= 2.32840483527903 \times 10^{-4}\alpha^6 + O\left(\alpha^8\right) \\
b(\alpha) &= -4.3198927470815 \times 10^{-4}\alpha^6 + O\left(\alpha^8\right).
\end{aligned}
$$

152

The error of the near equioscillating circle interpolant is shown in figure 7.3 for opening angle $\alpha = 1$ and radius 1. Also in this case $L(\alpha) > 0$ for $0 < \alpha < 2\pi$.

The next circle approximation we look at is one that satisfies

$$\left.\frac{dq(\mathbf{p}(t))}{dt}\right|_{t=\frac{1}{2}} = \left.\frac{d^2q(\mathbf{p}(t))}{dt^2}\right|_{t=\frac{1}{2}} = \left.\frac{d^3q(\mathbf{p}(t))}{dt^3}\right|_{t=\frac{1}{2}} = 0.$$

We see that within the interval $[0, 1]$ this approximation method has only four zeroes when counting multiplicity although the convergence is $O(\alpha^6)$. Thus, we disconnect the convergence rate from the number of zeroes in the error function in the approximation interval. Before the example we give a lemma proving that the function $L(\alpha)$ resulting from these requirements, gives a circle approximation with $O(\alpha^6)$ convergence.

Lemma 97 *Let $0 \leq \alpha < \arccos\frac{1}{3}$, and the tangent length of the approximation be determined by*

$$L(\alpha) = \tfrac{1}{3}\alpha + \tfrac{1}{144}\alpha^3 + \tfrac{1}{144}\left(\frac{48\alpha+\alpha^3-144(\cos\alpha)\alpha-3(\cos\alpha)\alpha^3}{(3(\cos\alpha)-1)}\right.$$

$$\left. + \frac{480\sin\alpha-96\sqrt{(-36\cos\alpha+2\cos^2\alpha+34)}}{(3(\cos\alpha)-1)}\right)$$

$$= \tfrac{1}{3}\alpha + \tfrac{1}{144}\alpha^3 + \tfrac{31}{92160}\alpha^5 + \tfrac{1069}{123863040}\alpha^7 + O\left(\alpha^9\right),$$

then we have that

$$\left.\frac{dq(\mathbf{p}(t))}{dt}\right|_{t=\frac{1}{2}} = \left.\frac{d^2q(\mathbf{p}(t))}{dt^2}\right|_{t=\frac{1}{2}} = \left.\frac{d^3q(\mathbf{p}(t))}{dt^3}\right|_{t=\frac{1}{2}} = 0,$$

and the choice of $L(\alpha)$ gives an $O(\alpha^6)$ approximation to the circle.

Proof. Since $q(\mathbf{p}(t))$ is symmetric around $t = \frac{1}{2}$ the first and third derivative have to be zero for $t = \frac{1}{2}$. Now by choosing

$$L(\alpha) = \frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{C}{5760}\alpha^5, \tag{7.19}$$

then

$$\left.\frac{d^2q(\mathbf{p}(t))}{dt^2}\right|_{t=\frac{1}{2}} = 9 - 9\cos\alpha + \frac{27}{4}\left(\frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{1}{5760}C\alpha^5\right)^2\cos\alpha$$

$$-\frac{9}{4}\left(\frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{1}{5760}C\alpha^5\right)^2$$

$$+15\left(\frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{1}{5760}C\alpha^5\right)\sin\alpha.$$

153

The requirement that

$$\frac{d^2 q(\mathbf{p}(t))}{dt^2}\bigg|_{t=\frac{1}{2}} = 0$$

gives a second degree equation in the unknown C. The solutions of this are

$$C_1 = 40\frac{48\alpha + \alpha^3 - 144(\cos\alpha)\alpha - 3(\cos\alpha)\alpha^3 + 480\sin\alpha - 96\sqrt{(-36\cos\alpha + 2\cos^2\alpha + 34)}}{(3(\cos\alpha)-1)\alpha^5}$$

$$C_2 = 40\frac{48\alpha + \alpha^3 - 144(\cos\alpha)\alpha - 3(\cos\alpha)\alpha^3 + 480\sin\alpha + 96\sqrt{(-36\cos\alpha + 2\cos^2\alpha + 34)}}{(3(\cos\alpha)-1)\alpha^5}.$$

Inserting C_1 into (7.19) and performing Taylor expansion, we get

$$L_1(\alpha) = \frac{1}{3}\alpha + \frac{1}{144}\alpha^3$$
$$+ \frac{1}{144}\frac{48\alpha + \alpha^3 - 144(\cos\alpha)\alpha - 3(\cos\alpha)\alpha^3 + 480\sin\alpha - 96\sqrt{(-36\cos\alpha + 2\cos^2\alpha + 34)}}{(3(\cos\alpha)-1)}$$
$$= \frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{31}{92160}\alpha^5 + \frac{1069}{123863040}\alpha^7 + O\left(\alpha^8\right).$$

Inserting $C_2(\alpha)$ into (7.19) and performing Taylor expansion, we get

$$L_2(\alpha) = \frac{1}{3}\alpha + \frac{1}{144}\alpha^3$$
$$+ \frac{1}{144}\frac{48\alpha + \alpha^3 - 144(\cos\alpha)\alpha - 3(\cos\alpha)\alpha^3 + 480\sin\alpha + 96\sqrt{(-36\cos\alpha + 2\cos^2\alpha + 34)}}{(3(\cos\alpha)-1)}$$
$$= 3\alpha + \frac{31}{16}\alpha^3 + \frac{13081}{10240}\alpha^5 + \frac{11602939}{13762560}\alpha^7 + O\left(\alpha^8\right).$$

We see that $L_2(\alpha)$ can be discarded since the factor before α is not $\frac{1}{3}$. ∎

Example 98 (Flat Circle Approximation.) *In figure 7.4 we show for opening angle $\alpha = 1$ and unit radius the approximation resulting from choosing the tangent length function $L(\alpha)$ as given in lemma 97. The Taylor expansions of $a(\alpha)$ and $b(\alpha)$ for this choice of $L(\alpha)$ is*

$$a(\alpha) = \frac{1}{2560}\alpha^6 - \frac{1}{163840}\alpha^8 + O\left(\alpha^{10}\right)$$
$$b(\alpha) = -\frac{1}{5120}\alpha^6 + \frac{1}{327680}\alpha^8 + O\left(\alpha^{10}\right).$$

Since $a(\alpha) \neq 0$, the interpolation is only of multiplicity two at the start and the end. The plot in figure 7.4 shows that there are no internal zeroes in $[0, 1]$. Thus, we have an $O(\alpha^6)$ approximation with only four interpolation conditions in the interval $[0, 1]$.
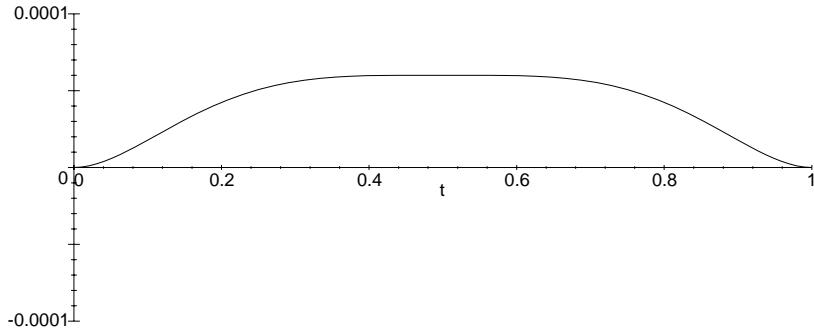
Figure 7.4: Plot of the error function for the circle approximation with first, second and third derivative at $t = \frac{1}{2}$ equal to zero. The example is a segment of a unit circle with opening angle $\alpha = 1$.

Note that the error is only a factor 3 greater than the error in the near equioscillating circle approximation in figure 7.3. Also in this case $L(\alpha) > 0$ for $0 < \alpha < 2\pi$.

By choosing a factor for the α^5 term greater than

$$\frac{1}{144} \frac{48\alpha + \alpha^3 - 144(\cos\alpha)\alpha - 3(\cos\alpha)\alpha^3 + 480\sin\alpha - 96\sqrt{(-36\cos\alpha + 2\cos^2\alpha + 34)}}{(3(\cos\alpha) - 1)}$$

we can produce more sharp approximations. However, the sharper the approximation the larger the errors are.

In the next example we force two extra interpolation points into the Hermite approximation by requiring that the integral of $q(\mathbf{p}(t))$ is zero. Also in this case we get an $O(\alpha^6)$ approximation. First we give a lemma establishing the tangent length function giving a zero integral.

Lemma 99 *The cubic Bezier function defined by (7.12) with tangent length given by*

$$L(\alpha) = \frac{-\frac{13}{5}\sin\alpha + \frac{1}{5}\sqrt{\left(169\sin^2\alpha + 216 - 378\cos\alpha + 162\cos^2\alpha\right)}}{2\left(\frac{6}{5} - \frac{9}{10}\cos\alpha\right)}$$

satisfy

$$L(\alpha) = \frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{41}{322560}\alpha^5 + O\left(\alpha^7\right),$$

155

have

$$\int_0^1 q(\mathbf{p}(t))dt = 0,$$

and is an $O(\alpha^6)$ circle approximation.

Proof. The first step is to find the integral of $q(\mathbf{p}(t))$. Remember that the integral from 0 to 1 of a Bernstein basis function of degree n is $\frac{1}{n+1}$. Thus,

$$
\begin{aligned}
\int_0^1 q(\mathbf{p}(t))dt &= \frac{1}{7}(2a + b) \\
&= \frac{1}{7}\left(\left(\frac{6}{5} - \frac{9}{10}\cos\alpha\right)L^2(\alpha) + \frac{13}{5}L(\alpha)\sin\alpha - \frac{9}{5} + \frac{9}{5}\cos\alpha\right).
\end{aligned}
$$

The values of $L(\alpha)$ making the integral zero are

$$
\begin{aligned}
L_1(\alpha) &= \frac{-\frac{13}{5}\sin\alpha + \frac{1}{5}\sqrt{\left(169\sin^2\alpha + 216 - 378\cos\alpha + 162\cos^2\alpha\right)}}{2\left(\frac{6}{5} - \frac{9}{10}\cos\alpha\right)} \\
L_2(\alpha) &= \frac{-\frac{13}{5}\sin\alpha - \frac{1}{5}\sqrt{\left(169\sin^2\alpha + 216 - 378\cos\alpha + 162\cos^2\alpha\right)}}{2\left(\frac{6}{5} - \frac{9}{10}\cos\alpha\right)}.
\end{aligned}
$$

The Taylor expansion of these are respectively

$$
\begin{aligned}
L_1(\alpha) &= \frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + \frac{41}{322560}\alpha^5 + O\left(\alpha^7\right) \\
L_2(\alpha) &= -9\alpha + \frac{231}{16}\alpha^3 - \frac{817953}{35840}\alpha^5 + O\left(\alpha^7\right).
\end{aligned}
$$

The first solution satisfy the requirement for an $O(\alpha^6)$ approximation to the circle segment. ∎

Example 100 (Zero Integral Circle Approximation.) *Figure 7.5 shows a plot of the error for angle $\alpha = 1$ of a unit circle using the approximation with zero integral, given in lemma 99. Also in this case $L(\alpha) > 0$ for $0 < \alpha < 2\pi$.*

In theorem 87 it was shown that by minimizing the square sum of the coefficient of $q(\mathbf{p}(t))$, the Hermite interpolant would be $O(\alpha^6)$. In the following example we look at this strategy for finding a circle approximation.
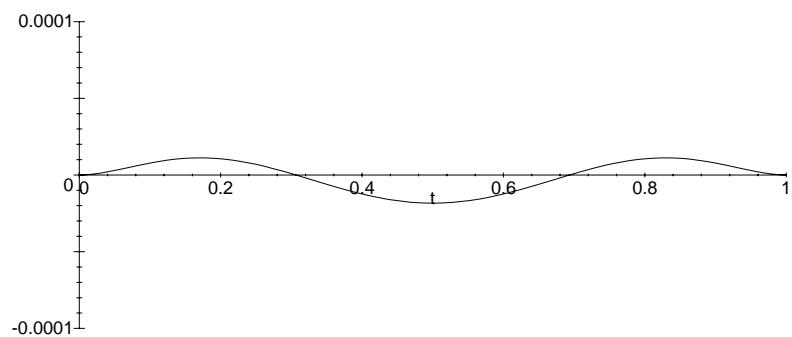
Figure 7.5: The error of the circle approximation with zero integral of $q(\mathbf{p}(t))$ for opening angle $\alpha = 1$ and unit radius.
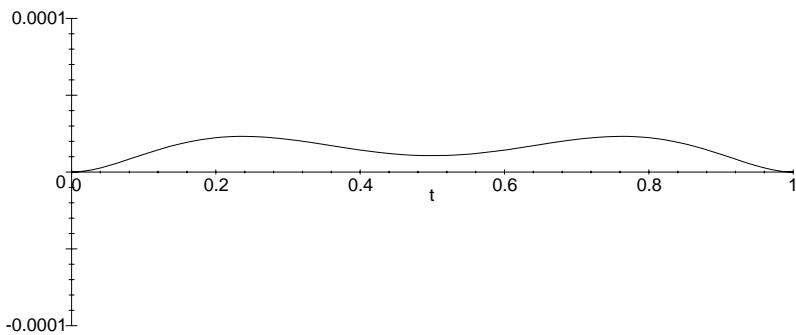
Figure 7.6: An example of the error function for the circle segment approximation with minimum square sum of the Bezier coefficients of $q(\mathbf{p}(t))$ for opening angle $\alpha = 1$ and radius 1.

Example 101 (Minimum Square Sum Circle Approximation.) *We want for a fixed opening angle α of the circular segment, to find the tangent length minimizing the expression*

$$p(\alpha) = 2(a(\alpha))^2 + (b(\alpha))^2.$$

To do this, we express the tangent length function

$$L(\alpha) = \frac{1}{3}\alpha + \frac{1}{144}\alpha^3 + c,$$

and

$$a(\alpha) = \tfrac{1}{5}(3L(\alpha)^2 + 2L(\alpha)\sin\alpha - 2(1 - \cos\alpha))$$
$$b(\alpha) = \tfrac{1}{10}(-9L(\alpha)^2\cos\alpha + 18L(\alpha)\sin\alpha - 10(1 - \cos\alpha)).$$

This minimum has to satisfy

$$\frac{dp(\alpha)}{dc} = 0,$$

which is a third degree equation in the variable c. This equation has one real solution and two imaginary solutions. Using Maple integrated in Scientific Word the exact representation of this solution can be found. However, the expression is so complex that it is not practical to include it in the example. However, using Taylor expansion, also in Maple, we get the following expression

$$\begin{aligned}L(\alpha) &= \tfrac{1}{3}\alpha + \tfrac{1}{144}\alpha^3 + 1.9275939542483660130718954245 \times 10^{-4}\alpha^5 \\ &\quad + 1.5291920868633401072847439645 \times 10^{-5}\alpha^7 \\ &\quad - 3.4158537015349191694598296 \times 10^{-7}\alpha^9 + O(\alpha^{11}).\end{aligned}$$

Note that the condition for $O(\alpha^6)$ approximation of the circular segment is satisfied. Now let $\alpha = 1$, we then get the following approximative values of

$$a(1) = 0.00028258357301723175949436545227$$
$$b(1) = -0.00035584482584010198092192238703.$$

This is in accordance with what should be expected, because the signs of $a(1)$ and $b(1)$ are opposite, and they are of similar absolute value. In the figure 7.6 we have plotted the errors for this circle approximation with opening angle $\alpha = 1$ and radius 1. Theorem 87 on page 133 also shows that this minimization of the square sums is $O(\alpha^6)$.

Ellipse approximation	Example	Inside error	Outside error
Simplest zero integral	102	-7.7×10^{-5}	8.8×10^{-5}
Simplest min. square sum	103	-6.7×10^{-5}	3.4×10^{-6}
Proj. Equi.circle	96	-2.8×10^{-5}	2.8×10^{-5}
Modified min. square sum	104	0	2.6×10^{-5}
Modified zero integral	105	-2.0×10^{-5},	1.4×10^{-5}

Table 7.2: The table shows a summary of different cubic Hermit ellipse approximations when approximating the standard ellipse segment of this section. The error variation is shown as inside and outside error. The approximation with least variation of error is the modified minimum square sum approximation, this is better than making a circle of radius 2, approximating with the equioscilating circle approximation and then projecting on to the plane of the ellipse.

7.3 Examples of Ellipse Approximations

We now give some examples on how to control the Taylor expansion of the tangent length functions to model the error in ellipse approximations. These examples illustrate the possibilities of these approximations, however they only address a limit number of the family of approximation methods that can be based on theorem 90. Note that in the examples numeric calculations have been performed in Maple integrated in Scientific Workplace, thus all digits displayed might not be significant. Rerunning the example with a different number of significant digits gives slightly different answers.

To compensate for the variation in the gradient of the algebraic representation of the ellipse, we introduce weights. Remember the ellipse in (7.2)

$$q(x,y) = \left(\frac{x}{b}\right)^2 + (b\,y)^2 - 1 = 0, \ b \neq 0,$$

which has the gradient

$$\nabla q(x,y) = \left(\frac{2x}{b^2}, 2b^2 y\right),$$

and define weights by

$$w_i = \frac{1}{\left\|\nabla q(\mathbf{c}(\tau + \frac{\alpha}{6}i))\right\|_2} = \frac{1}{2\sqrt{\frac{1}{b^2}\cos^2\left(\tau + \frac{\alpha}{6}i\right) + b^2 \sin^2\left(\tau + \frac{\alpha}{6}i\right)}}, \ i = 0, \ldots, 6$$

with $\mathbf{c}(\theta)$ given in (7.1), the parametric description of the ellipse to be approximated.

w_0	w_1	w_2	w_3	w_4	w_5	w_6
1	.841310	.619477	.474166	.385308	.329369	.293343

Table 7.3: The weights defined for $b = 2$, $\tau = 0$ and $\alpha = 1$.

Now let $b = 2$, $\tau = 0$ and $\alpha = 1$, then we get the weight values shown in table 7.3.

Example 102 (Simpelst Zero Integral Ellipse Approximation.) *The sum of the Bernstein coefficients of $q(\mathbf{p}(t))$ was in example 100 shown to be equal to the integral of $q(\mathbf{p}(t))$. An approximation to the circle was made by finding a zero of this integral. A weighted analog of this would be to find the zeroes of a weighted sum of coefficients. This can be done by posing the conditions*

$$w_2 a(\alpha) + w_3 b(\alpha) + w_4 c(\alpha) = 0. \tag{7.20}$$

Now search for a solution of type 1 in theorem 90. Here we had

$$a(\alpha) = \left(\tfrac{7}{57600} + \tfrac{1}{80}c_1 + \tfrac{1}{60}c_1^2\right)\alpha^6 + \left(\tfrac{1}{320}c_2 + \tfrac{1}{120}c_2 c_1\right)\alpha^7 + O\left(\alpha^8\right)$$

$$b(\alpha) = \left(-\tfrac{23}{38400} + \tfrac{1}{40}c_1^2\right)\alpha^6 + \left(\tfrac{1}{80}c_2 c_1\right)\alpha^7 + O\left(\alpha^8\right)$$

$$c(\alpha) = \left(\tfrac{7}{57600} - \tfrac{1}{80}c_1 + \tfrac{1}{60}c_1^2\right)\alpha^6 + \left(-\tfrac{1}{320}c_2 + \tfrac{1}{120}c_2 c_1\right)\alpha^7 + O\left(\alpha^8\right).$$

We observe that by setting $c_2 = 0$ we cancel the degree 7 term. Thus, what remains is a second degree equation in the variable c_1. (7.20) has two solution $c_1 = -0.1421624$ and $c_1 = 3.981784 \times 10^{-2}$. The last of these alternatives give the plot in figure 7.7 for the approximative error estimate defined by

$$\rho(t) = \frac{q(\mathbf{p}(t))}{\|\nabla q(\mathbf{p}(t))\|_2} = \frac{15a(\alpha)(1-t)^4 t^2 + 20b(\alpha)(1-t)^3 t^3 + 15c(\alpha)(1-t)^2 t^4}{\sqrt{\left(\frac{2x(t)}{b^2}\right)^2 + (2b^2 y(t))^2}}.$$

The tangent lengths defined by $c_1 = 3.981784 \times 10^{-2}$ and $c_2 = 0$ are

$$L_0(1) = 0.3469140844$$
$$L_1(1) = 0.3336414711.$$

Points from the approximation is plotted if figure 7.8 together with the true ellipse segment.

To evaluate how good this approximation is, we can make an approximation by viewing the ellipse as a projection of a circle with radius two. The
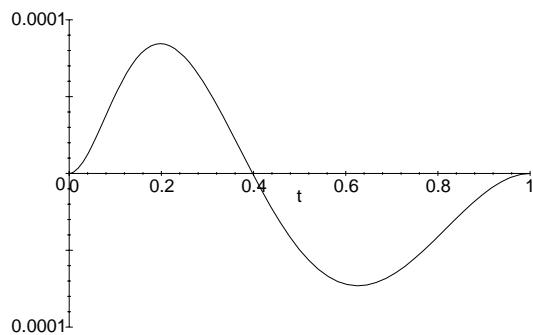
Figure 7.7: The plot of an approximative error function $\frac{q(p(t))}{\|\nabla q(p(t))\|_2}$ for the simplest zero integral approximation of a cubic Bezier curve to an ellipse $\left(\frac{x}{2}\right)^2 + (2y)^2 - 1 = 0$ for the segment $\left(2\cos\varphi, \frac{1}{2}\sin\varphi\right)$, $0 \leq \varphi \leq 1$.
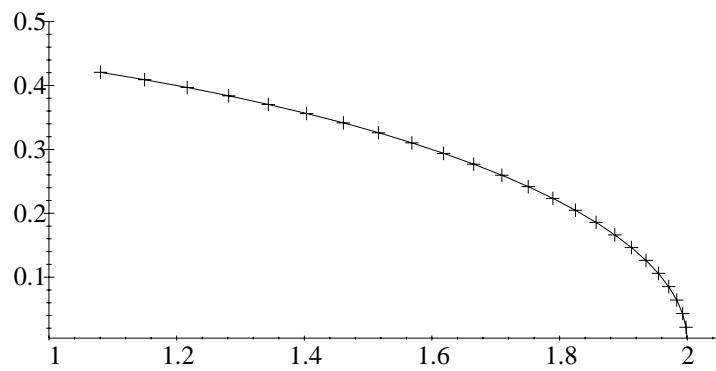
Figure 7.8: The plot of the exact ellipse segment with 25 points sampled from the approximation.

error estimate for best approximation given in figure 7.3 for a unit circle and radius 1 is ± 0.000014. The circle, that can be used to make the ellipse in this example by projection, has radius 2. Thus, we get an error of ± 0.000028 for the projected circle approximation, while the error for the ellipse approximation in this example is ± 0.00008. However, the approximation is better than the error estimates for the simplest circle approximation given in figure 7.1 that after scaling to radius 2 is between $+0.00001$ to -0.00012.

Example 103 (Simplest Minimum Square Sum Ellipse Approximation.)
Let as in the previous example the solution alternative chosen be the first, set $c_2 = 0$ giving

$$a(\alpha) = \left(\tfrac{7}{57600} + \tfrac{1}{80}c_1 + \tfrac{1}{60}c_1^2\right)\alpha^6$$

$$b(\alpha) = \left(-\tfrac{23}{38400} + \tfrac{1}{40}c_1^2\right)\alpha^6$$

$$c(\alpha) = \left(\tfrac{7}{57600} - \tfrac{1}{80}c_1 + \tfrac{1}{60}c_1^2\right)\alpha^6,$$

and find the minimum of the sum of the square of the weighted coefficients. Thus, we want to solve

$$\frac{d\left(w_2\left(a(\alpha)\right)^2 + w_3\left(b(\alpha)\right)^2 + w_4\left(c(\alpha)\right)^2\right)}{dc_1} = 0.$$

This equation have one real solution that here is approximated by

$$c_1 = -2.427947742 \times 10^{-3}.$$

The tangent lengths calculated based on this value for c_1 are

$$L_0(1) = 0.3398731198$$
$$L_1(1) = 0.3406824357.$$

The plot of the estimated error function is shown in figure 7.9. Note that this error is between -0.000067 and 0.0000034 which is significantly smaller than in the previous example.

Let as in the previous examples the solution alternative chosen be the first, and use the value of $c_1 = -2.427947742 \times 10^{-3}$ used in the previous example. Set as before $c_2 = 0$. Now add $O(\alpha^5)$ terms to $L_0(\alpha)$ and $L_1(\alpha)$

$$L_0(\alpha) = \frac{\alpha}{3} + \frac{\left(\tfrac{1}{24} + c_1\right)\alpha^3}{6} + \frac{\left(1.669690586 \times 10^{-2} + c_3\right)\alpha^5}{120}$$

$$L_1(\alpha) = \frac{\alpha}{3} + \frac{\left(\tfrac{1}{24} - c_1\right)\alpha^3}{6} + \frac{\left(1.669690586 \times 10^{-2} - c_3\right)\alpha^5}{120}.$$
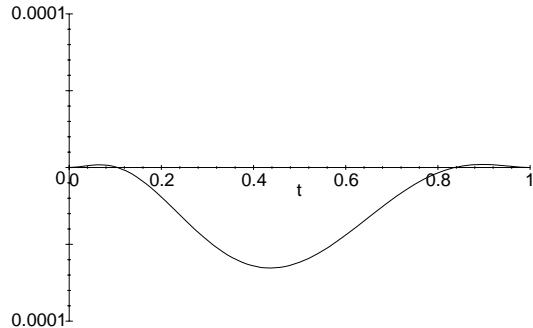
Figure 7.9: The estimated error for the simplest minimum square sum ellipse approximation. The ellipse segment approximated is the same as in the previous figure.

If we set $c_1 = 0$ and $c_3 = 0$ in this equation, we end up with the equioscillating circle approximation from example 96. Now, Taylor expanding the coefficient of $q(\mathbf{p}(t))$ we get

$$
\begin{aligned}
a(\alpha) &= \left(1.469342676 \times 10^{-4} - 3.333333333 \times 10^{-3} c_3\right) \alpha^6 \\
&\quad + \left(-1.313497581 \times 10^{-6} + 5.555555556 \times 10^{-4} c_3\right) \alpha^8 + O\left(\alpha^{10}\right) \\
b(\alpha) &= \left(-5.153264428 \times 10^{-4} - .005 c_3\right) \alpha^6 \\
&\quad + \left(-3.992837576 \times 10^{-6} + 4.904852295 \times 10^{-5} c_3\right) \alpha^8 + O\left(\alpha^{10}\right) \\
c(\alpha) &= \left(2.076304753 \times 10^{-4} - 3.333333333 \times 10^{-3} c_3\right) \alpha^6 \\
&\quad + \left(6.49202194 \times 10^{-6} - 7.349085829 \times 10^{-5} c_3\right) \alpha^8 + O\left(\alpha^{10}\right).
\end{aligned}
$$

Example 104 (Modified Least Square Sum Ellipse Approximation.)
A first alternative for finding the value of c_3 is the value giving the minimum of the square sum of the weighted coefficients

$$
\frac{d\left((w_2 a(\alpha))^2 + (w_3 b(\alpha))^2 + (w_4 c(\alpha))^2\right)}{dc_3} = 0.
$$

This is a linear equation with solution approximated to be

$$
c_3 = -3.085218937 \times 10^{-2}.
$$

The resulting estimated error is shown if figure 7.10.
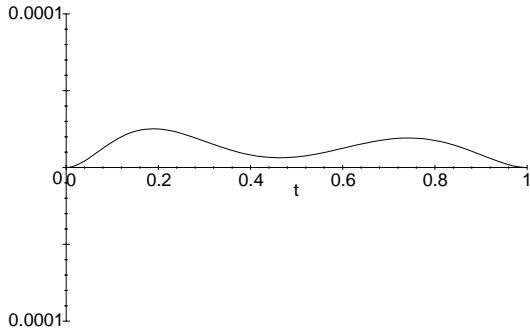
Figure 7.10: The estimated error function for calculating the constant c_3 in the ellipse approximation by the minimum square sum of the weighted coefficients.

Example 105 (Modified zero integral ellipse approximation.) *Another alternative for finding the value of c_3 is as the value giving the integral of the weighted coefficients the value 0*

$$w_2 a(\alpha) + w_3 b(\alpha) + w_4 c(\alpha) = 0.$$

This is a linear equation with one solution

$$c_3 = -1.366501622 \times 10^{-2}.$$

The resulting estimated error function is shown if figure 7.11.

This finale approximation is smaller than the one we would achieve by making an approximation of a circle of radius 2 and projecting this onto an ellipse, as discussed in the end of example 102. This error was ranging between ± 0.00003, while the error in the final example is ranging between -0.00002 and 0.000015.

7.4 Examples of Degree 4 Circle Approximation

In [Schaback:92] also the approximation with higher degree polynomials than cubic was addressed. Here we address fourth degree circle approximations methods with $O(\alpha^8)$ convergence. The methods addressed are:

- The simplest Taylor expansion of the free variables.

164
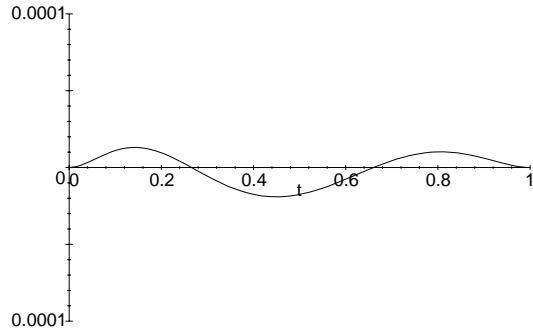
Figure 7.11: The estimated error function for calculating the constant c_3 in the ellipse approximation by a zero integral of $q(\mathbf{p}(t))$ with weighted coefficients.

Circle approximation	Example	Inside error	Outside error
Simplest	110	-1.1×10^{-7}	2.6×10^{-8}
Minimum square sum	111	0	6.6×10^{-8}
Equioccilating	112	-1.8×10^{-8}	1.8×10^{-8}

Table 7.4: The table shows a summary of different fourth degree circle approximations and how the error behaves. The circle segment approximated has opening angle 1 and is a part of a unit circle.

- Minimizing the square sum of the coefficients of $q(\mathbf{p}(t))$ when expressed in a eight degree Bernstein basis.

- Approximated equioscillation of $q(\mathbf{p}(t))$.

Table 7.4 gives a short overview of the methods and their error behavior when approximating a segment with opening angle 1 of a unit circle.

Description of Circle Segment. The circle segment is described by

$$c(\theta) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \ \theta \in [-\frac{\alpha}{2}, \frac{\alpha}{2}] \qquad (7.21)$$

with the algebraic description $q(x,y) = 0$, where

$$q(x,y) = x^2 + y^2 - 1. \qquad (7.22)$$

165

Description of Interpolant. We interpolate the position and tangent of the circle segment for $\theta = -\frac{\alpha}{2}$ and $\theta = \frac{\alpha}{2}$, and use the following interpolant $\mathbf{p}(t)$ using a fourth degree Bernstein basis

$$\mathbf{p}(t) = (x(t), y(t)) = \sum_{i=0}^{4} \mathbf{p}_i \binom{4}{i} (1-t)^{4-i} t^i, \qquad (7.23)$$

where

$$\mathbf{p}_0 = \begin{pmatrix} \cos\frac{\alpha}{2} \\ -\sin\frac{\alpha}{2} \end{pmatrix}$$

$$\mathbf{p}_1 = \begin{pmatrix} \cos\frac{\alpha}{2} \\ -\sin\frac{\alpha}{2} \end{pmatrix} + L(\alpha) \begin{pmatrix} \sin\frac{\alpha}{2} \\ \cos\frac{\alpha}{2} \end{pmatrix}$$

$$\mathbf{p}_2 = \begin{pmatrix} 1 + d(\alpha) \\ 0 \end{pmatrix}$$

$$\mathbf{p}_3 = \begin{pmatrix} \cos\frac{\alpha}{2} \\ \sin\frac{\alpha}{2} \end{pmatrix} - L(\alpha) \begin{pmatrix} -\sin\frac{\alpha}{2} \\ \cos\frac{\alpha}{2} \end{pmatrix}$$

$$\mathbf{p}_4 = \begin{pmatrix} \cos\frac{\alpha}{2} \\ \sin\frac{\alpha}{2} \end{pmatrix}.$$

The only unknown quantities in this description are $L(\alpha)$ and $d(\alpha)$. The problem is how to control these tangents to get a best possible approximation.

Lemma 106 *Let the fourth degree interpolant $\mathbf{p}(t)$ to the circle segment be described as in (7.23) and let the algebraic description of the circle $q(\mathbf{x}) = 0$ be described as in (7.22) then*

$$\begin{aligned} q(\mathbf{p}(t)) = {} & 28a(\alpha)(1-t)^6 t^2 + 56b(\alpha)(1-t)^5 t^3 \\ & + 70c(\alpha)(1-t)^4 t^4 + 56b(\alpha)(1-t)^3 t^5 + 28a(\alpha)(1-t)^2 t^6, \end{aligned}$$

where

$$a(\alpha) = \frac{1}{7}\left(3\cos\frac{\alpha}{2} + 3d(\alpha)\cos\frac{\alpha}{2} - 3 + 4(L(\alpha))^2\right)$$

$$b(\alpha) = \frac{1}{7}\begin{pmatrix} 2(\cos\frac{\alpha}{2} + L(\alpha)\sin\frac{\alpha}{2})\cos\frac{\alpha}{2} - 8 \\ +6\left(\cos\frac{\alpha}{2} + L(\alpha)\sin\frac{\alpha}{2}\right)(1 + d(\alpha)) \end{pmatrix}$$

$$c(\alpha) = \frac{1}{35}\begin{pmatrix} 34\cos^2\frac{\alpha}{2} - 52 + 64\left(\sin\frac{\alpha}{2}\right)L(\alpha)\cos\frac{\alpha}{2} \\ +16(L(\alpha))^2 - 32(L(\alpha))^2(\cos\frac{\alpha}{2})^2 + 18(1 + d(\alpha))^2 \end{pmatrix}$$

Proof. Can be verified by expansion of the expressions. ∎

Theorem 107 *Let the circle segment and the fourth degree interpolant* $\mathbf{p}(t)$ *be defined as in lemma 106. If* $d(\alpha)$ *and* $L(\alpha)$ *have either the Taylor expansion*

$$d(\alpha) = \tfrac{1}{24}\alpha^2 + \tfrac{1}{24}\left(\tfrac{35}{48} - \tfrac{\sqrt{2}}{2}\right)\alpha^4 + O(\alpha^7)$$

(7.24)

$$L(\alpha) = \tfrac{1}{4}\alpha + \tfrac{1}{6}\left(\tfrac{3\sqrt{2}}{16} - \tfrac{1}{4}\right)\alpha^3 + \tfrac{1}{120}\left(\tfrac{5\sqrt{2}}{32} - \tfrac{167}{768}\right)\alpha^5 + O(\alpha^6),$$

or

$$d(\alpha) = \tfrac{1}{24}\alpha^2 + \tfrac{1}{24}\left(\tfrac{35}{48} + \tfrac{\sqrt{2}}{2}\right)\alpha^4 + O(\alpha^7)$$

(7.25)

$$L(\alpha) = \tfrac{1}{4}\alpha - \tfrac{1}{6}\left(\tfrac{1}{4} + \tfrac{3\sqrt{2}}{16}\right)\alpha^3 - \tfrac{1}{120}\left(\tfrac{167}{768} + \tfrac{5\sqrt{2}}{32}\right)\alpha^5 + O(\alpha^6),$$

and $a(\alpha)$, $b(\alpha)$ *and* $c(\alpha)$ *are all* $O(\alpha^8)$.

Proof. First let

$$d(\alpha) = \tfrac{1}{24}\alpha^2 + \tfrac{1}{24}\left(\tfrac{35}{48} - \tfrac{\sqrt{2}}{2}\right)\alpha^4 + \tfrac{1}{5040}d_7\alpha^7 + \tfrac{1}{40320}d_8\alpha^8 + O(\alpha^9)$$

$$L(\alpha) = \tfrac{1}{4}\alpha + \tfrac{1}{6}\left(\tfrac{3\sqrt{2}}{16} - \tfrac{1}{4}\right)\alpha^3 + \tfrac{1}{120}\left(\tfrac{5\sqrt{2}}{32} - \tfrac{167}{768}\right)\alpha^5$$
$$- \tfrac{1}{3360}d_7\alpha^6 + \tfrac{1}{5040}L_7\alpha^7 + \tfrac{1}{40320}L_8\alpha^8 + O(\alpha^9).$$

Then the Taylor expansions to 10 terms of respectively $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$ are

$$a(\alpha) = \left(\tfrac{1}{17640}L_7 + \tfrac{46139}{216760320} - \tfrac{43\sqrt{2}}{286720} + \tfrac{1}{94080}d_8\right)\alpha^8 + O\left(\alpha^9\right)$$

$$b(\alpha) = \left(-\tfrac{125221}{108380160} + \tfrac{263\sqrt{2}}{322560} + \tfrac{1}{8820}L_7 + \tfrac{1}{47040}d_8\right)\alpha^8 + O(\alpha^9)$$ (7.26)

$$c(\alpha) = \left(\tfrac{1}{7350}L_7 + \tfrac{160379}{90316800} - \tfrac{809\sqrt{2}}{645120} + \tfrac{1}{39200}d_8\right)\alpha^8 + O(\alpha^9).$$

The 9th degree term contains the d_7 constant. Then let

$$d(\alpha) = \tfrac{1}{24}\alpha^2 + \tfrac{1}{24}\left(\tfrac{35}{48} + \tfrac{\sqrt{2}}{2}\right)\alpha^4 + \tfrac{1}{5040}d_7\alpha^7 + \tfrac{1}{40320}d_8\alpha^8 + O(\alpha^9)$$

$$L(\alpha) = \tfrac{1}{4}\alpha - \tfrac{1}{6}\left(\tfrac{1}{4} + \tfrac{3\sqrt{2}}{16}\right)\alpha^3 - \tfrac{1}{120}\left(\tfrac{167}{768} + \tfrac{5\sqrt{2}}{32}\right)\alpha^5$$
$$- \tfrac{1}{3360}d_7\alpha^6 + \tfrac{1}{5040}L_7\alpha^7 + \tfrac{1}{40320}L_8\alpha^8 + O(\alpha^9),$$

Then the Taylor expansion to 10 terms of respectively $a(\alpha)$, $b(\alpha)$ and $c(\alpha)$

$$a(\alpha) = \left(\tfrac{1}{17640}L_7 + \tfrac{46139}{216760320} + \tfrac{43\sqrt{2}}{286720} + \tfrac{1}{94080}d_8\right)\alpha^8 + O(\alpha^9)$$

$$b(\alpha) = \left(-\tfrac{125221}{108380160} - \tfrac{263\sqrt{2}}{322560} + \tfrac{1}{8820}L_7 + \tfrac{1}{47040}d_8\right)\alpha^8 + O(\alpha^9)$$

$$c(\alpha) = \left(\tfrac{1}{7350}L_7 + \tfrac{160379}{90316800} + \tfrac{809\sqrt{2}}{645120} + \tfrac{1}{39200}d_8\right)\alpha^8 + O(\alpha^9).$$

The 9th degree term contains the d_7 constant. ■

Corollary 108 *All circle approximations methods of the type given in equation (7.23) satisfying conditions (7.24) or (7.25) are $O(\alpha^8)$.*

Proof. Theorem 107 establish $O(\alpha^8)$, while theorem 92 connects the convergence rate to the error. ■

Corollary 109 *Choosing*

$$d(\alpha) = \tfrac{1}{24}\alpha^2 + \tfrac{1}{24}\left(\tfrac{35}{48} - \tfrac{1}{2}\sqrt{2}\right)\alpha^4$$

$$L(\alpha) = \tfrac{1}{4}\alpha + \tfrac{1}{6}\left(\tfrac{3}{16}\sqrt{2} - \tfrac{1}{4}\right)\alpha^3 + \tfrac{1}{120}\left(\tfrac{5}{32}\sqrt{2} - \tfrac{167}{768}\right)\alpha^5$$

establish a simple circle approximation method with convergence rate $O(\alpha^8)$.

The Bernstein coefficient of the approximation in corollary 109 and the Taylor expansion with 11 terms are

$$a(\alpha) = \left(\tfrac{46139}{216760320} - \tfrac{43}{286720}\sqrt{2}\right)\alpha^8 + O\left(\alpha^{10}\right)$$
$$\approx 7.646487806 \times 10^{-7}\alpha^8 + O\left(\alpha^{10}\right)$$

$$b(\alpha) = \left(-\tfrac{125221}{108380160} + \tfrac{263}{322560}\sqrt{2}\right)\alpha^8 + O\left(\alpha^{10}\right)$$
$$\approx -2.304627712 \times 10^{-6}\alpha^8 + O\left(\alpha^{10}\right)$$

$$c(\alpha) = \left(\tfrac{160379}{90316800} - \tfrac{809}{645120}\sqrt{2}\right)\alpha^8 + O\left(\alpha^{10}\right)$$
$$\approx 2.271691708 \times 10^{-6}\alpha^8 + O\left(\alpha^{10}\right).$$

Example 110 *The error of this simple degree eight circle approximation method, is shown in figure 7.12.*

We now look at the circle approximation resulting from minimizing the square sum of the coefficients of $q(\mathbf{p}(t))$. Since we already have got an $O(\alpha^8)$, based on (7.23), see theorem 107, then theorem 80 states that minimizing the square sum of the coefficient of $q(\mathbf{p}(t))$ gives an $O(\alpha^8)$ approximation.

Example 111 (Minimum Square Sum Degree Four Circle approx.) *Let $\alpha = 1$. Since α is fixed, we now denote the coefficients of $q(\mathbf{p}(t))$ respectively a, b and c. Using ten digits accuracy, we get the following description*
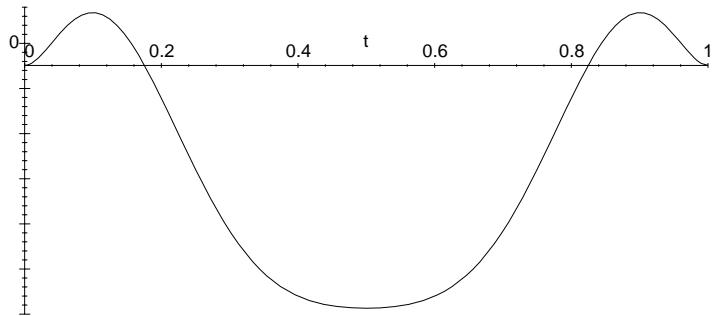
Figure 7.12: Approximated error estimate for the simplest approximation with 4th degree Bezier curve to a circle segment of opening angle $\alpha = 1$ and radius 1. The error varies between -1.1×10^{-7} and 2.6×10^{-8}. Compared to the error the simples 3rd degree approximation which was between 6.5×10^{-5} and 3.4×10^{-6}.

of the coefficients:

$$
a = -.05246\,46163 + .37610\,68123d + \tfrac{4}{7}L^2
$$

$$
b = -.92281\,39564 + .12021\,01407L
$$

$$
+ \tfrac{6}{7}\left(.87758\,25619 + .47942\,55386L\right)(1+d)
$$

$$
c = -.73756\,74517 + .76934\,49006L - .24699\,53399L^2 + \tfrac{18}{35}\left(1+d\right)^2 . \tag{7.27}
$$

The square sum of the coefficients is

$$
f(d, L) = 2a^2 + 2b^2 + c^2 .
$$

This expression attains a local minimum for

$$
d = 4.25835\,2237 \times 10^{-2}
$$

$$
L = 0.25256\,09794 .
$$

We obtain the following values for a, b and c

$$
a = 1.07846 \times 10^{-6}
$$

$$
b = -.00000\,20001
$$

$$
c = .00000\,24376 .
$$

As before the expression $\frac{q(\mathbf{p}(t))}{2}$ is a good estimate of the error. In figure 7.13 we plot the error estimate of this circle approximation. The maximum
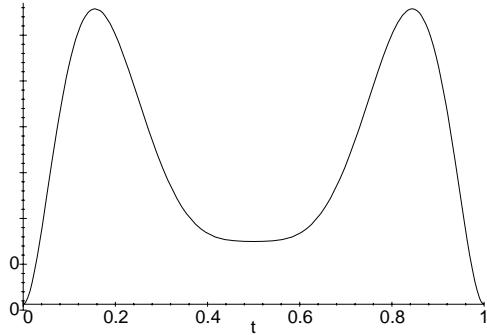
Figure 7.13: The error in the circle approximation error resulting from finding an approximate minimum of the square sum of the coefficients of $q(\mathbf{p}(t))$ for opening angle $\alpha = 1.0$. The maximal error is less than 6.6×10^{-8}. Which is approximately half of the error the simplest circle approximation.

value of the error is approximately half of the error for the simplest $O(\alpha^8)$ approximation method.

As in Section 7.2 we can produce a wide variety of circle approximations with the 4th degree Bezier curve. E.g. an equioscillating circle approximation is assumed to exist. The error estimate $\frac{q(\mathbf{p}(t))}{2}$ of this has the shape of the 8th degree Bezier function equioscillating function defined by the vertices

$$(0, 0, 1, -u, v, -u, 1, 0, 0),$$

where u and v can be shown to be by lengthy calculations to be

$$
\begin{aligned}
v &= (28 + 20u + 9u^2 - 3\sqrt{3}(2 + u)^2)/10 \\
u &= 2(\sqrt{3} - 6)/9 + 2\sqrt{2}(K_2^{3/4} + K_3^{1/2})/(9K_1^{1/6}K_2^{1/4}) \\
K_3 &= 54\sqrt{2}(1 + \sqrt{3})K_1^{1/2} + (5 + 32\sqrt{3} + (16 + 14\sqrt{3} - K_1^{1/3})K_1^{1/3})K_2^{1/2} \\
K_2 &= -5 - 32\sqrt{3} + (8 + 7\sqrt{3} + K_1^{1/3})K_1^{1/3} \\
K_1 &= -172 + 3\sqrt{3}(-55 + 54\sqrt{2(1 + \sqrt{3})}).
\end{aligned}
$$

A numeric approximation to u and v with 50 digits in Maple gives

$$
\begin{aligned}
u &= 3.3975842919587644617862134245043548479555002743432 \\
v &= 4.8459627840494672713677587801149176249864674640035.
\end{aligned}
$$

170
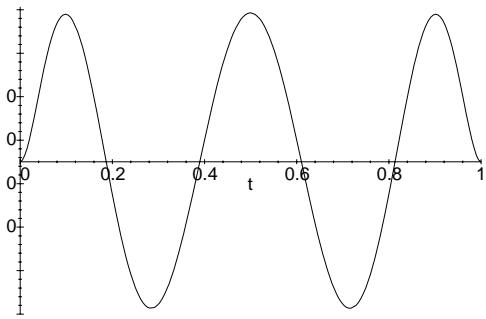
Figure 7.14: The error in the degree four circle approximation with equioscillation of $q(\mathbf{p}(t))$ for opening angle $\alpha = 1.0$. The maximal error is less than 1.8×10^{-8}. This is approximately $\frac{1}{6}$ of the error of the simplest degree four circle approximation, and $\frac{1}{3}$ of the circle approximation with the minimum square sum of the coefficients of $q(\mathbf{p}(t))$.

Example 112 (Fourth Degree Equioscillating Circle Approximation.)
We now make an approximation to the equioscillating error function by using a, b and c as defined in (7.27) with ten digit accuracy. We then solve the equation

$$a\,u = -b$$
$$a\,v = c$$

to enforce the equioscillation. The approximative solution we find is

$$L = 0.25255\,0802$$
$$d = 4.25900\,3677 \times 10^{-2},$$

which gives the following values

$$a = 5.9101 \times 10^{-7}$$
$$b = -.00000\,20073$$
$$c = 0.00000\,28641.$$

Then in figure 7.14 we show the plot of error estimate $\frac{q(\mathbf{p}(t))}{2}$. Note that the error is less than $\frac{1}{3}$ of the error for the approximation with minimum square sum of the coefficients.

171

Bibliography

[APS:87]   Dokken, T., Reference Manual, APS B-spline Library
                    version 3.3., ISBN-7267-930-2, Senter for Industriforskn-
                    ing, Oslo, Norway, (1987).

[Bajaj:93]   Bajaj, C., The Emergence of Algebraic Curves and Sur-
                    faces in Geometric Design, in Directions in *Geometric
                    Computing*, Information Geometers, (1993),1-28.

[Oslo-algorithm]   Cohen, E., T. Lyche and R. Riesenfeld, Discrete B-
                    Splines and Subdivision Techniques in Computer-Aided
                    Geometric Design and Computer Graphics, Computer
                    Graphics and Image Processing, 14(1980), 87-111.

[Chionh:92]   Chionh, E.W. and R. N. Goldmann, Using multivariate
                    resultants to find the implicit equation of a rational sur-
                    face, The Visual Computer 8(1992), 171-180.

[Cox:72]   Cox, M. G., The numerical evaluation of B-splines, J.
                    Inst. Math, Appl., 10(1972), 134-149.

[Cox:92,Cox:97]   Cox, D., Little, J. and O'Shea, D., Ideals, Varieties and
                    Algorithms, Springer-Verlag, New York, (1992) & (1997).

[Cox:98]   Cox, D., Little, J. and O'Shea, D., Using Algebraic
                    Geometry, Springer-Verlag, New York, (1998).

[Cox:99]   Cox D., Goldman R. and Zhang M., On the Validity
                    of Implicitization by Moving Quadrics for Rational Sur-
                    faces with No Base Points, J. Symbolic Computation 11
                    (1999).

[de Boor:73]   de Boor, C., The Quasi-Interpolant as a Tool in El-
                    ementary Polynomial Spline Theory, in *Approximation
                    Theory*, G.G. Lorenz (ed), Academic Press, New York,
                    (1973), 269-276.

[de Boor:78]   de Boor, C., *A practical Guide to Splines*, Springer-Verlag, (1978).

[de Boor:87-1]   de Boor, C., B-form basics, in *Geometric Modeling: Algorithms and New Trends*, G. Farin (ed), SIAM, (1987), 131-148.

[de Boor:87-2]   de Boor, C., K. Höllig and M. Sabin. High accuracy geometric Hermite interpolation, CAGD 4(1987), 269-278.

[Degen:92]   Degen, W. L. F., Best Approximation of parametric curves by splines, in *Mathematical Methods in Computer Aided Geometric Design II*, T. Lyche and L.L. Schumaker(eds) Academic Press, Boston, (1992), 171–184.

[do Carmo:92]   do Carmo, M. P., *Riemannian Geometry*, Birkhäuser, Boston·Basel·Berlin, (1992).

[Dokken:85]   Dokken, T., Finding Intersections of B-spline represented geometries using recursive subdivision techniques, CAGD 2(1985), 189-195.

[Dokken:89]   Dokken, T., V. Skytt and A-M. Ytrehus, Recursive Subdivision and Iteration in Intersections and Related Problems, in *Computer Aided Geometric Design*, T. Lyche and L.L. Schumaker(eds), Academic Press, (1989), 207-214.

[Dokken:90]   Dokken, T., M. Dæhlen, T. Lyche and K. Mørken, Good approximation of circles by curvature-continuous Bezier curves, CAGD, 7(1990), 33-41.

[Dokken:97]   Dokken, T., Aspects of Algorithms for Manifold Intersection, in *Numerical Methods and Software Tools in Industrial Mathematics*, M. Dæhlen, and A. Tveito (eds), to appear.

[Farin:92]   Farin G., *Curves and Surfaces for Computer Aided Geometric Design*, Academic Press Inc, San Diego (1992).

[Farin:86]   Farin, G., Triangular Bernstein-Bezier patches, CAGD 3(1986), 83-128.

[Fuggeli:93]   Fugelli, P., Knot insertion for B-splines - Evaluation of new and established algorithms, Master Thesis (in Norwegian), University of Oslo, Norway, (1993).

[Farouki:87]   Farouki, R. T. and V. T. Rajan, On the numerical condition of polynomials in Bernstein from, CAGD 4(1987), 191-216.

[Farouki:88]   Farouki, R. T. and V. T. Rajan, Algorithms for polynomials in Bernstein form, CAGD 5(1988), 1-26.

[Floater:95-1]   Floater, M. S., High-order approximation of conic sections by quadratic splines, CAGD 12(1995), 617-637.

[Floater:95-2]   Floater, M. S., Rational Cubic Implicitization, in *Mathematical Methods for Curves and Surfaces*, M. Dæhlen, T. Lyche and L. L. Schumaker (eds.), Vanderbilt University Press (1995), 151-159.

[Floater:96]   Floater, M. S., An analysis of cubic approximation schemes for conic sections, Advances in Computational Mathematics 5(1996), 361-379.

[Floater:97]   Floater, M. S., An $O(h^{2n})$ Hermite approximation for conic sections, CAGD 14(97), 135-151.

[Golub & Loan:83] Golub, G. H. and C. F. van Loan, Matrix Computations, North Oxford Academic, (1983).

[GPM-16:80]   Hasund, K., T. Dokken, S. Ulfsby, GPM Report No 16, System Specifications, Sculptured Surfaces, Sentralinstitutt for industriell forskning, Oslo, Norway, (1980).

[GPM-26:83]   Dokken T. et. al, Reference Manual, B-spline Library, Version 1.0, Sentralinstitutt for industriell forskning, Oslo, Norway, (1983).

[Hoschek:93]   Hoschek J. and D. Lasser, *Fundamentals of computer aided geometric design*, A K Peters, (1993).

[Hohmeyer:92]   Hohmeyer, M. E., Robust and Efficient Surface Intersection for Solid Modelling, Report No. UCB/CSD 92/681, Computer Science Division, University of California, (1992).

174

[Hoffmann:90]   Hoffmann, C.M., "Algebraic and Numerical Techniques for Offsets and Blends," in *Computations of Curves and Surfaces,* S. Micchelli, M. Gasea and W. Dahmen (eds), Kluwer Academic, Boston, (1990), 499-528.

[Hoffmann:89]   Hoffmann, C. M., *Geometric & Solid Modelling, An Introduction*, Morgan Kaufmann Publishers, Inc, San Mateo, California, (1989).

[Hoffmann:93]   Hoffmann, C. M., Implicit Curves and Surfaces in CAGD. IEEE Computer Graphics & Applications, 13(1993), 79-88.

[Hollig:88]   Höllig, K., Algorithms for rational spline curves. ARO-Report 88-1, (1988) 287-300.

[Höllig:95]   Höllig, K. and J. Koch, Geometric Hermite Interpolation, CAGD 12(1995), 567-580.

[Höllig:96]   Höllig, K. and J. Koch, Geometric Hermite Interpolation with maximal order and smoothness, CAGD 13 (1996), 681-695.

[Kriezes:90-1]   Kriezis, G. A., Algorithms for Rational Spline Surface Intersection, Ph. D. Thesis, Massachusetts Institute of Technology, Boston, (1990).

[Kriezes:90-2]   Kriezis, G. A. and N. M. Patrakalakis. Rational polynomial surface intersection, Technical Report 90-9, MIT Ocean Engineering Design Laboratory, Boston, June (1990).

[Kriezes:91]   Kriezis, G. A. and N. M. Patrakalakis. Rational polynomial surface intersections, in Proceedings of the 17 ASME Design Automation Conference, September (1991).

[Krishnan:96]   Krishnan, S. and D. Manocha, An Efficient Surface Intersection Algorithm based on the Lower Dimensional Formulation, ACM Trans. on Graphics 16(1997), 74–106.

[Lyche:94]   Lyche, T. and K. Mørken, A Metric for Parametric Approximation, in Curves and Surfaces in Geometric Design, P-J. Laurent, A. Le Méhauté and L.L. Schumaker

(eds), A.K. Peter, Wellesley, Massachusetts, (1994), 311-318.

[Mountaudouin:89] Mountaudouin, Y. D., Cross product of cones of revolution. CAD, 21(6)(1989).

[Moerken:91-1] Mørken, K., Some identities for products and degree raising of splines, Constr. Approx 7 (1991), 195-208.

[Moerken:91-2] Mørken, K., Best Approximation of Circle Segments by Quadric Bézier curves, in Curves and Surfaces, P-J. Laurent, A. Le Méhauté and L.L. Schumaker (eds), Academic Press, Boston, (1991), 331-336.

[Numres:88] Press, Flannery, Teukolsky and Vetterling, *Numerical Recipes in C*, Cambridge University Press, Cambridge, (1988), 60-71.

[Sederberg:83] Sederberg, T. W., Implicit and Parametric Curves and surfaces for Computer Aided Geometric Design, Ph.D. Thesis, Purdue University, (1983).

[Sederberg:84-1] Sederberg, T. W., D. C. Anderson and R. N. Goldman, Implicit Representation of parametric curves and surfaces, Computer Vision, Graphics and Image Processing 28(1984), 72-84.

[Sederberg:84-2] Sederberg, T. W., Planar Piecewise Algebraic Curves. CAGD (1984), 241-255.

[Sederberg:88] Sederberg, T. W. and R. J.. Meyers, Loop detection in surface patch intersections. CAGD 5(1988), 161-171.

[Sederberg:89-1] Sederberg, T. W., H. N. Christiansen and S. Katz. An improved test for closed loops in surface intersection, CAD 21(1989), 505-508.

[Sederberg:89-2] Sederberg, T. W., An algorithm for algebraic curve intersection, CAD 21(1989), 547-554.

[Sederberg:89-3] Sederberg, T. W., J. Zhao and A.K. Zundel, Approximative parametrization of algebraic curves, in *Theory and Practice of Geometric Modeling* W. Strasser and H.-P. Seidel (eds), Springer, Berlin, (1989), 33-54.

[Sederberg:89-4]   Sederberg, T. W., S.R. White and A.K. Zundel, Fat arcs: A bounding region with cubic convergence, CAGD 6(1989), 205-218.

[Sederberg:90]   Sederberg, T. W., Techniques for cubic algebraic Surfaces, IEEE CG & A July 1990, 14-25.

[Sederberg:91]   Sederberg, T. W., and Tomoyuki Nishita, Geometric Hermite approximation of surface patch intersection curves, Computer Aided Geometric Design 8(1991), 97-114.

[Sederberg:95]   Sederberg, T. W., and F. Chen, Implicitization using moving curves and surfaces, Computer Graphics Annual Conference Series (1995), 301–308.

[Sederberg:96]   Sederberg, T.W. and A.K. Zundel, Pyramides that bound surface patches. CVGIP: Graphics Models and Image Processing, (1996), 75-81.

[Sederberg:97]   Sederberg, T. W., Goldman R.N. and Du H., Implicitization Rational Curves by the Method of Moving Algebraic Curves, J. Symbolic Computation, 23(1997), 153-175.

[Sederberg:99]   Sederberg, T.W., Zheng J., Klimaszewski K. and Dokken T, Approximate Implicitization Using Monoid Curves and Surfaces, Graphical Models and Image Processing 61, 177-198 (1999).

[Seidel:92]   Seidel, H.P., Presenting Piecewise Polynomials as Linear Combinations of Multivariate Splines, in *Mathematical Methods in Computer Aided Geometric Design II*, T. Lyche and L. L. Schumaker(eds.) Academic Press, Boston, (1992), 559-566.

[SISL:94]   SISL - The SINTEF Spline Library, version 4.0, SINTEF, Oslo, Norway, December 1994.

[Stewart:73]   Stewart, G. W., Introduction to Matrix Computations, Academic Press, New York & London, (1973).

[Sinha:85]   Sinha, P., E. Klassen and K.K. Wang, Exploiting topological and geometric properties for selective subdivision. In *ACM Symposium on Computational Geometry*, ACM Press, (1985), 39-45.

177

[Spivak:79]   Spivak, M., *A comprehensive Introduction to DIFFER-ENTIAL GEOMETRY*, volume 1, second edition, Publish or Perish Inc, (1979).

[IVF:94]   Wandebäck, F. and H. Johansson, Kvalitetssäkring av CAD-geometri. IVF-RAPPORT 94028, Institutet för Verkstadsteknisk forskning, Göteborg,Sweden, (1994).

[Schaback:92]   Schaback, R., Rational Geometric Curve Interpolation, in *Mathematical Methods in Computer Aided Geometric Design II*, T. Lyche and L. L. Schumaker (eds), Academic Press, Boston, (1992), 517-535.

[Weiler:86]   Weiler, K. J., Topological Structures for Geometric Modelling, Ph.D. Thesis, Renesselaer Polytechnic Institute, Report No. TR-86032, Troy, New York, (1986).

[Zundel:94]   Zundel, A.K, Surface-Surface Intersection: Loop Destruction Using Bézier Clipping and Pyramidal Bounds. Thesis for Doctor of Philosophy, Brigham Young University, (1994).

[Zundel:96]   Zundel, A.K and T. W. Sederberg, Surface Intersection Loop Destruction, Technical report from Brigham Young University, Department of Computer Science (1996).

Appendix A

PosProd Basis Functions

In this appendix we look into the most used basis functions in CAD/CAM-systems and show that these are Pos Prod Basis functions. In Appendix A.1 we address tensor product B-splines. Then in respectively appendix A.2 and A.3 we address Bernstein Bases defined over a simplex and the tensor product Bernstein Bases.

A.1 Tensor Product B-splines

The B-spline basis functions are piecewise polynomials of degree $n = k - 1$ defined over a knot vector $\{t_1, \ldots, t_{N+k}\}$ where $t_{i+k} > t_i$ and $t_{i+1} \geq t_i$ and satisfy

$$\sum_{i=1}^{N} B_{i,k}(t) = 1$$

$$B_{i,k}(t) > 0, \ t_i < t < t_{i+k} \ i = 1, \ldots, N \qquad \text{(A.1)}$$

$$B_{i,k}(t) \geq 0, \ t_k \leq t \leq t_{N+1}.$$

From the total knot sequence $\{t_1, \ldots, t_{N+k}\}$ we take the smaller sequence $\{t_k, \ldots, t_{N+1}\}$, if $t_i = \ldots = t_{i+r-1}$ we say that r is the multiplicity of the knot value t_i. A knot vector where all knots have multiplicity equal to the polynomial order k is called a *Bernstein knot vector*.

Condition 1. in definition 24 on page 46 is satisfied by (A.1). By theorem 3.1 in [Moerken:91-1] condition 2. in definition 24 is satisfied. Thus, the B-spline basis is a Pos Prod Basis. In addition the theorem in [Moerken:91-1] can be used as the basis for a PosProd Algorithm for B-splines.

Basis functions built by tensor products of B-spline basis functions

$$B_{\mathbf{i},\mathbf{n}}(\mathbf{s}) = \prod_{l=1}^{g} B_{i_p,(n_p+1)}(s_p)$$

also satisfy the conditions 1. and 2. in definition 24, because we can apply the results of the univariate case to each variable separately.

Thus, all manifolds defined by NURBS with positive weights are R-positive manifolds. The most used of these are:

- 2D NURBS curves with positive weights.

- 3D NURBS curves with positive weights.

- 3D NURBS surfaces with positive weights.

A.2   Bernstein Bases over Simplices

Bernstein basis functions built over a simplex S

$$S = \{\beta_1\mathbf{q}_1 + \ldots + \beta_{g+1}\mathbf{q}_{g+1} \mid \beta_1 + \ldots + \beta_{g+1} = 1, \ 0 \leq \beta_i\}$$

with corners $\mathbf{q}_1,\ldots,\mathbf{q}_{g+1}$, in $\mathbb{R}^g$ can be defined by barycentric coordinates $(\beta_1(\mathbf{s}),\ldots,\beta_{g+1}(\mathbf{s}))$ that satisfy

$$\begin{pmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_{i-1} & \mathbf{q}_i & \mathbf{q}_{i+1} & \cdots & \mathbf{q}_{g+1} \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{g+1} \end{pmatrix} = \begin{pmatrix} \mathbf{s} \\ 1 \end{pmatrix}.$$

Here $\beta_1(\mathbf{s}) + \ldots + \beta_{g+1}(\mathbf{s}) = 1$ is just the last equation. For references see [Sederberg:84-2], [Farin:86] and [de Boor:87-1]. When $\mathbf{s} \in S$, $\mathbf{s}$ has to be a convex combination of the corners of the simplex. By Cramers rule we get

$$\beta_i(\mathbf{s}) = \frac{\begin{vmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_{i-1} & \mathbf{s} & \mathbf{q}_{i+1} & \cdots & \mathbf{q}_{g+1} \\ 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \end{vmatrix}}{\begin{vmatrix} \mathbf{q}_1 & \cdots & \mathbf{q}_{g+1} \\ 1 & \cdots & 1 \end{vmatrix}}, \ i = 1,\ldots,g+1.$$

Bernstein basis functions built over a simplex S are defined by

$$B_{\mathbf{i},n}(\mathbf{s}) = n! \prod_{l=1}^{g+1} \frac{(\beta_l(\mathbf{s}))^{i_l}}{i_l!} = \frac{n!}{i_1!\ldots i_{g+1}!} (\beta_1(\mathbf{s}))^{i_1} \ldots (\beta_{g+1}(\mathbf{s}))^{i_{g+1}}. \quad \text{(A.2)}$$

Here we have used n instead of $\mathbf{n}$ since we have the same maximal polynomial degree in all variables. The index set $\mathcal{I}$ is defined by

$$\mathcal{I} = \{(i_1, \ldots, i_{g+1}) \mid i_1 + \ldots + i_{g+1} = n,\ i_l \in \mathbb{N}_0,\ l = 1, \ldots, g+1\}.$$

The number of multi indices in $\mathcal{I}$ are $N = \binom{n+g}{g}$, which can be proved by recursion on $n + g$.

Lemma 113 *The Bernstein basis functions built over a simplex defined as in equation (A.2) satisfy*

$$\sum_{\mathbf{i}\in\mathcal{I}} B_{\mathbf{i},n}(\mathbf{s}) = 1 \tag{A.3}$$

$$B_{\mathbf{i},n}(\mathbf{s}) \geq 0,\ \mathbf{i}\in\mathcal{I},\ \mathbf{s}\in\Omega.$$

Proof. See [Farin:86] or [de Boor:87-1]. ∎

Lemma 114 *Let the functions $p(\mathbf{s})$ and $q(\mathbf{s})$ in g-variables be represented by Bernstein basis functions over a simplex.*

$$p(\mathbf{s}) = \sum_{\mathbf{i}\in\mathcal{I}_p} p_{\mathbf{i}} B_{\mathbf{i},n_p}(\mathbf{s})$$

$$q(\mathbf{s}) = \sum_{\mathbf{j}\in\mathcal{I}_q} q_{\mathbf{j}} B_{j,n_q}(\mathbf{s}).$$

Then the product of $p(\mathbf{s})q(\mathbf{s})$ can be expressed as

$$p(\mathbf{s})q(\mathbf{s}) = \sum_{\mathbf{i}\in\mathcal{I}_p}\sum_{\mathbf{j}\in\mathcal{I}_q} p_{\mathbf{i}} q_{\mathbf{j}} B_{\mathbf{i},n_p}(\mathbf{s}) B_{j,n_q}(\mathbf{s})$$

$$= \sum_{\mathbf{k}\in\mathcal{I}_{pq}} g_{\mathbf{k}} B_{\mathbf{k},n_p+n_q}(\mathbf{s}),$$

where $\mathbf{k} = (k_1 \ldots, k_{g+1})$, $\mathbf{j} = (j_1, \ldots, j_{g+1})$ and $\mathbf{i} = (i_1, \ldots, i_{g+1})$ with

$$g_{\mathbf{k}} = \frac{n_p!\, n_q!}{(n_p + n_q)!} \sum_{\substack{i_l=\max(0,k_l-n_q) \\ l=1,\ldots,g+1}}^{\min(k_l,n_p)} \prod_{l=1}^{g+1} \binom{k_l}{i_l} p_{\mathbf{i}}\, q_{\mathbf{k}-\mathbf{i}}. \tag{A.4}$$

If the coefficients satisfy

$$p_{\mathbf{i}} \geq 0,\ \mathbf{i}\in\mathcal{I}_p$$

$$q_{\mathbf{j}} \geq 0,\ \mathbf{j}\in\mathcal{I}_q,$$

then

$$g_{\mathbf{k}} \geq 0,\ \mathbf{k}\in\mathcal{I}_{pq}.$$

181

Proof. See [Farin:86] or [de Boor:87-1]. ∎

Equation (A.4) is a PosProd Algorithm since, as stated in theorem 29, $g_{\mathbf{k}}$ are calculated as a convex combination of all $p_{\mathbf{i}} \, q_{\mathbf{j}}$.

Thus manifolds described by simplical rational Bernstein Bases with positive weights are R-positive manifolds. The most used of these are:

- 2D Bezier curves with positive weights. (This is a special instance of 2D NURBS curves)

- 3D Bezier curves with positive weights. (This is a special instance of 3D NURBS curves)

- 3D rational triangular Bezier surfaces with positive weights.

3D rational tetrahedral Bezier volumes with positive weights are not included. This as the definition of R-positive manifolds requires that the dimension of the manifold is lower than the dimension of the space in which the manifold lies.

A.3 Tensor Product Bernstein Basis

The tensor product Bernstein basis is a special instance of the B-spline bases. Knot insertion can bring a B–spline represented function to a Bernstein knot vector and thus to the representation as a set of Bernstein basis represented functions. The overhead of the extra knots in a Bernstein knot vector is not so severe when products of functions are calculated as it is in the representation of the original functions. Since the algorithms for Bernstein basis represented objects are simpler than with B-spline basis represented objects, we go into some detail here, although the properties are already shown in Appendix A.1 on tensor product B-splines.

Example 115 *Let $p_i(s)$, $i = 1, \ldots, m$ be univariate B-spline represented functions with N coefficients of order k with k-tupple knots at the ends and single internal knots. The piecewise polynomial functions consist of $N - k + 1$ polynomial segments. The same function represented with a Bernstein knot vector requires $(N - k + 1)k$ coefficients. The overhead compared to the B-spline representation is*

$$(N - k + 1)k - N.$$

When we make products of m such functions the overhead will not grow while the total storage for the Bernstein knot vector representation will be $(N - k + 1)((k - 1)m + 1)$. Now making the quotient between the overhead

and what is needed for storage of the product of m functions with a Bernstein knot vector we get

$$\frac{(N-k+1)k-N}{(N-k+1)((k-1)m+1)} = \frac{k-\frac{N}{N-k+1}}{(k-1)m+1} \leq \frac{k-1}{(k-1)m+1} < \frac{1}{m}.$$

Thus, as the number of functions being multiplied increase, the relative effect of the overhead is small compared to the total number of coefficients. When the functions are at least C^1 relative simple techniques can be used to store the functions with a reduced Bernstein knot vector with $k-2$ multiple internal knots, thus bringing the overhead further down.

An algorithm for calculating the coefficients of the product of two bi-variate Bernstein basis represented functions is given in [Farouki:88]. This algorithm is generalized in the following lemma.

Lemma 116 *Let the functions $p(\mathbf{s})$ and $q(\mathbf{s})$ in g-variables defined over the unit cube in $\mathbb{R}^g$ be represented in a tensor product Bernstein basis*

$$p(\mathbf{s}) = \sum_{\substack{i_l=0 \\ l=1,\ldots,g}}^{n_l} p_{i_1,\ldots,i_g} \prod_{r=1}^{g} B_{i_r}^{n_r}(s_r)$$

$$q(\mathbf{s}) = \sum_{\substack{i_l=0 \\ l=1,\ldots,g}}^{m_l} q_{i_1,\ldots,i_g} \prod_{r=1}^{g} B_{i_r}^{m_r}(s_r)).$$

Here

$$B_i^m(u) = \binom{m}{i} u^i (1-u)^{(m-i)},$$

and

$$\sum_{\substack{i_l=0 \\ l=1,\ldots,g}}^{n_l} = \sum_{i_1=0}^{n_1} \cdots \sum_{i_g=0}^{n_g}.$$

Then the product of $p(\mathbf{s})$ and $q(\mathbf{s})$ can be expressed as

$$p(\mathbf{s})q(\mathbf{s}) = \sum_{\substack{i_l=0 \\ l=1,\ldots,g}}^{n_l} \sum_{\substack{j_k=0 \\ k=1,\ldots,g}}^{m_k} p_{i_1,\ldots,i_g} q_{j_1,\ldots,j_g} \prod_{r=1}^{g} \left(B_{i_r}^{n_r}(s_r) B_{j_r}^{m_r}(s_r) \right)$$

$$= \sum_{\substack{i_l=0 \\ l=1,\ldots,g}}^{n_l+m_l} g_{i_1,\ldots,i_g} \prod_{r=1}^{g} B_{i_r}^{n_r+m_r}(s_r),$$

where

$$g_{i_1 \dots, i_g} = \sum_{\substack{p_l=\max(0, i_l - m_l) \\ l=1,\dots,g}}^{\min(i_l, n_l)} c_{p_1, \dots, p_g} d_{i_1 - p_1, \dots, i_g - p_g} \frac{\prod_{r=1}^{g} \left[\binom{n_r}{p_r} \binom{m_r}{i_r - p_r} \right]}{\prod_{r=1}^{g} \binom{n_r + m_r}{i_r}}. \qquad (\text{A.5})$$

Proof. See [de Boor:87-1]. ∎

The product algorithm in (A.5) is a PosProd Product Algorithms since, as stated in remark 29, the coefficients $g(s)$ are convex combinations of the coefficients of $p(s)$ and $q(s)$.

Appendix B

Reformulation of Geometric Interrogations to Manifold Intersections

In the main part of the thesis we have concentrated the discussions on problems related to the intersection of smooth bounded manifolds in $\mathbb{R}^l$. We address in this appendix a number of geometric interrogations problems that can be reformulated to intersection of manifolds. Many of the reformulations addressed are implemented in the SISL library [SISL:94]. This appendix is based on the work in [Dokken:85].

The intersection algorithms in SISL - SINTEF Spline Library are based on an intersection kernel that address the intersection of two manifolds of possibly different dimensions. The manifolds being intersected in SISL are:

- 0-manifolds (points) in $\mathbb{R}^2$ and $\mathbb{R}^3$.

- NURBS represented 1-manifolds (curves) in $\mathbb{R}^2$ and $\mathbb{R}^3$.

- NURBS represented 2-manifolds (surfaces) in $\mathbb{R}^3$.

A special feature in SISL is the possibility to find common zeroes of two NURBS represented functions defined on the same 1-dimensional or 2-dimensional compact domain. The functions are expanded to 1-manifolds in $\mathbb{R}^2$ or 2-manifolds in $\mathbb{R}^3$ when properties of manifold intersections have to be used. This feature has enabled the implementation of a wide range of intersection related functions in SISL.

We address in this appendix such functionality integrated into SISL, but also look at possibilities emerging with an intersection kernel handling higher dimensional manifolds than 2-manifolds and higher dimensional real spaces than $\mathbb{R}^3$. The discussion is structured as follows:

- In Appendix B.1 we address:

  - The intersection of algebraic curves and parametric curves in $\mathbb{R}^2$.
  - The intersection of algebraic surfaces and parametric curves in $\mathbb{R}^3$.
  - The intersection of algebraic surfaces and parametric surfaces in $\mathbb{R}^3$.

- Generating silhouette curves of different kinds is addressed in Appendix B.2.

- Projection of a manifold onto another manifold is addressed in Appendix B.3.

- Finding extremal points on a curve or surface with respect to a direction is in SISL converted to searching for extremal values of a function defined over the parameter domain of respectively a curve or a surface. The same is the case for finding the points on a curve or a surface being closest to a given point, or the points on two curves being closest. Aspects of searching for extremal values of such functions are addressed in Appendix B.4.

B.1 Intersection of Parametric Represented Manifolds and Hypersurfaces

In Section 2.5.5 we showed that the dimensionality of an intersection problem can be reduced by combining parametric described manifolds and hypersurfaces. Many of the interrogation problems in CAD-systems can be formulated in such a way. In SISL the following intersections have been implemented by using such combinations:

- Intersection of a NURBS curve and a straight line in $\mathbb{R}^2$.

- Intersection of a NURBS curve and a circle in $\mathbb{R}^2$.

- The intersection of a general second degree algebraic curve and a NURBS curve.

- The intersection of a NURBS surface and a plane in $\mathbb{R}^3$.

- The intersection of a NURBS surface and a sphere in $\mathbb{R}^3$.

- The intersection of a NURBS curve and a torus in $\mathbb{R}^{3.}$

- The intersection of a general second degree algebraic surface and a NURBS surface in $\mathbb{R}^3$.

- The intersection of a straight line and a NURBS surface.

- The intersection of a circle and a NURBS surface.

- The intersection of a torus and a NURBS surface.

The simplest extensions to higher dimensional problems is the intersection of 1- and 2-manifolds with hypersurfaces. This results in finding zeroes of functions in 1 or 2 variables and can for NURBS represented geometries be addressed by SISL. However, the intersection of manifolds of dimension $g > 2$ and hypersurface requires an intersection kernel handling manifolds of dimension $g > 2$.

The main challenge of combining algebraic and parametric representations is to ensure that the functions resulting from the combination have minimal rounding errors. In Section 4.1 we addressed the problem of combining R-positive manifolds with hypersurfaces, and showed in Section 4.2 that the use of PosProd Algorithms resulted in a controlled growth of the relative rounding errors. An alternative to PosProd Algorithms is to use methods that reproduce polynomial behavior. One such method is B-spline interpolation. However, as the algebraic degree is growing, the condition number of the coefficient matrix of the B-spline interpolation increases. Although the condition number reflects a worst case, practical experience has shown that the use of PosProd Algorithms is to be preferred over interpolation methods.

B.2   Finding Silhouette Curves in $\mathbb{R}^3$

In SISL three different silhouette calculations have been implemented by reformulation approach:

- Parallel silhouettes.

- Perspective silhouettes.

- Circular silhouettes.

To give a complete silhouette calculation we have to consider both:

- The zeroes of the function representing the reformulated silhouette problem.

- Edges and break lines.

In all of the silhouette calculations the normal vector of the surface play a central role. Thus, we address the normal vector in Section B.2.1.

B.2.1 The normal vector

When a surface $\mathbf{p}(s_1, s_2)$ of degree (n_1, n_2) is nonrational the normal is as we know

$$\mathbf{n}(s_1, s_2) = \frac{\partial \mathbf{p}(s_1, s_2)}{\partial s_1} \times \frac{\partial \mathbf{p}(s_1, s_2)}{\partial s_2}.$$

Resulting in a polynomial degree for the normal of $(2n_1 - 1, 2n_2 - 1)$.

In case $\mathbf{p}(s_1, s_2)$ is rational, the degrees of the numerator of $\mathbf{n}(s_1, s_2)$ are $(4n_1 - 1, 4n_2 - 1)$, and of the denominator $(4n_1, 4n_2)$. However, the denominator $h(s, t)$ is a factor in both the numerator and denominator of $\mathbf{n}(s_1, s_2)$. Thus, cancelling this we get degrees $(3n_1 - 1, 3n_2 - 1)$ in the numerator and $(3n_1, 3n_2)$ of the denominator.

To control the propagation of the rounding error, PosProd Algorithms can be used for making the normal vector.

B.2.2 Parallel Projection Silhouette in $\mathbb{R}^3$

The silhouettes of the parallel projection in a direction $\mathbf{e}$ are points where

$$\mathbf{n}(s_1, s_2) \cdot \mathbf{e} = 0.$$

The polynomial order is related to the polynomial order of $\mathbf{n}(s_1, s_2)$ which is given above.

B.2.3 Perspective Silhouette in $\mathbb{R}^3$

The silhouette curves resulting from a perspective projection with respect to a point $\mathbf{p}_0$ have to satisfy

$$\mathbf{n}(s_1, s_2) \cdot (\mathbf{p}(s_1, s_2) - \mathbf{p}_0) = 0.$$

The polynomial degrees of the numerator is $(4n_1 - 1, 4n_2 - 1)$ and of the denominator $(4n_1, 4n_2)$. If the surface is nonrational the polynomial degree is $(3n_1 - 1, 3n_2 - 1)$.

B.2.4 Circular Silhouette in $\mathbb{R}^3$

Let $\mathbf{p}(s_1, s_2)$ be a point on a surface and $\mathbf{n}(s_1, s_2)$ be the normal in $\mathbf{p}(s_1, s_2)$. Let $\mathbf{q}$ be a point on an axis and $\mathbf{b}$ be the direction vector of the axis. The circular silhouette curves are defined by

$$\mathbf{n}(s_1, s_2) \times (\mathbf{p}(s_1, s_2) - \mathbf{q}) \cdot \mathbf{b} = 0.$$

The polynomial degrees of the numerator is $(4n_1 - 1, 4n_2 - 1)$ and of the denominator $(4n_1, 4n_2)$. If the surface is nonrational, the polynomial degree is $(3n_1 - 1, 3n_2 - 1)$.

B.3 Projection of a Manifold onto another Manifold

These functions are not implemented in SISL but can be of use in geometric modeling systems. Two problems are addressed:

- Normal projection of a manifold onto another manifold.

- Parallel projection of a manifold onto another manifold.

B.3.1 Normal Projection

The normal projection of a g_1-manifold $\mathbf{p}(\mathbf{s})$ onto a g_2-manifold $\mathbf{q}(\mathbf{t})$ can be formulated as

$$(\mathbf{q}(\mathbf{t}) - \mathbf{p}(\mathbf{s})) \cdot \frac{\partial \mathbf{q}(\mathbf{t})}{\partial t_i} = 0, \ i = 1, \ldots, g_2,$$

resulting in g_2 functions in $g_1 + g_2$ variables.

Remark 26 *For $g_1 = g_2 = 1$ we have the normal projection of a curve onto another curve.*

Remark 27 *For $g_1 = 1$ and $g_2 = 2$ then we have the normal projection of a parametric curve onto a parametric surface.*

Remark 28 *For $g_1 = 2$ and $g_2 = 1$ we have an the intersection of the normal plane of a parametric curve with a parametric surface.*

B.3.2 Parallel Projection

The parallel projection of a parametric g_1-manifold $\mathbf{p}(\mathbf{s})$ onto a parametric g_2-manifold $\mathbf{q}(\mathbf{t})$ in $\mathbb{R}^l$ satisfies

$$(\mathbf{q}(\mathbf{t}) - \mathbf{p}(\mathbf{s})) \cdot \mathbf{n}_j = 0, \ j = 1, \ldots, l - 1.$$

Here $\mathbf{n}_j, \ j = 1, \ldots, l-1$ are linearly independent vectors normal to projection direction $\mathbf{e}$. The result is $l - 1$ functions in $g_1 + g_2$ variables.

Remark 29 *If $\mathbf{p}(\mathbf{s})$ and $\mathbf{q}(\mathbf{t})$ both are in $\mathbb{R}^3$, $g_1 = 1$ and $g_2 = 2$, then we have the parallel projection of a curve onto a surface. Interchanging g_1 and g_2 does not change the geometric interpretation.*

B.4 Reformulation to extremal value problems.

Many problems can be reformulated to finding the extremal value of a function defined over the parameter domain of a manifold or of the tensor product of the parameter domains of two manifolds.

Let $\Omega \in \mathbb{R}^g$, be a bounded set over which we define a continuous function $f : \Omega \to \mathbb{R}$ that is piecewise smooth, and where the smooth pieces are limited by a set Ω_b that can be decomposed into $(g - 1)$-manifolds contained in Ω. Ω_b is thus also containing the boundary of Ω. To find the maximal values of the function the following approaches can be used:

- Use a tradition analyses of extremal values based on (partial) derivatives and analysis of the behavior of the function along the manifolds describing the piecewise smooth structure Ω_b.

- Apply a direct search for the extremal values by combining iteration and subdivision techniques.

In SISL the last approach is employed for efficiency reasons. This to make a generic tool for finding extremal values of NURBS based functions in one or two variables. However, this approach is also too slow when repeated use of the extremal value calculation is necessary. In these cases an iteration using as start point the result of the previous search for an extremal value, can be used. In these cases it is often too expensive to evaluate the function f for each search, and the iteration is performed on the original geometries. However, the iterations can fail and fallback strategies have to be implemented.

In SISL the following problems are addressed using the search for extremal values:

- Find the point closest to a given point $\mathbf{q}$ on a NURBS represented curve $\mathbf{p}(s)$. The function searched for extremal values is

$$f(s) = (\mathbf{p}(s) - \mathbf{q}) \cdot (\mathbf{p}(s) - \mathbf{q}).$$

  This is the same as the function for the intersection of a circle or sphere of radius 0 and a curve.

- Find the point closest to a given point $\mathbf{q}$ on a NURBS represented surface $\mathbf{p}(s)$. The function searched for extremal values is

$$f(s, t) = (\mathbf{p}(s, t) - \mathbf{q}) \cdot (\mathbf{p}(s, t) - \mathbf{q}).$$

  This is the same as the function for the intersection of a circle or sphere of radius 0 and a surface.

- Find the closest points between two NURBS represented curves $\mathbf{p}(s)$ and $\mathbf{q}(t)$. The function searched for extremal values is

$$f(s, t) = (\mathbf{p}(s) - \mathbf{q}(\mathbf{t})) \cdot (\mathbf{p}(s) - \mathbf{q}(\mathbf{t})).$$

- Find the extremal point(s) on a NURBS represented curve $\mathbf{p}(s)$ in the direction $\mathbf{d}$. The function searched for extremal values is

$$f(s) = \mathbf{p}(s) \cdot \mathbf{d}.$$

- Find the extremal point(s) on a NURBS represented surface $\mathbf{p}(s, t)$ in the direction $\mathbf{d}$. The function searched for extremal values is

$$f(s, t) = \mathbf{p}(s, t) \cdot \mathbf{d}.$$