



DIPLOMARBEIT

Methoden zur Knotenschätzung von B-Spline-Kurven für die Modellierung profilhafter Laserscanndaten

Ausgeführt zum Zwecke der Erlangung
des akademischen Grades eines

Diplom-Ingenieurs

unter der Leitung von

Univ.Prof. Dr.-Ing. Hans-Berndt Neuner

(Department für Geodäsie und Geoinformation, TU Wien)

und

Univ.-Ass. MSc. Claudius Schmitt

(Department für Geodäsie und Geoinformation, TU Wien)

eingereicht an der Technischen Universität Wien

Fakultät für Mathematik und Geoinformation

von

Manuel Grecher, B.Sc.

Matr.Nr.: 1025589

Breiteneich 92

3580 Horn

Wien, im August 2016

Manuel Grecher

Kurzfassung

Durch die Zunahme von flächenhaften Messtechniken ist besonders in der Ingenieurgeodäsie eine genaue Approximation von großen Punktwolken für spätere Anwendungszwecke notwendig. Hierfür bieten sich im Speziellen B-Spline Kurven bzw. B-Spline Flächen an, welche sich bei einer geeigneten Parameterwahl an lokale Gegebenheiten geometrisch bestmöglich anpassen können. Das Anpassungsvermögen von B-Splines ist dabei im Wesentlichen von zwei Parametersätzen, den Kontrollpunkten und den Knoten, abhängig. Das gleichzeitige Schätzen dieser Parameter führt zu einem hochgradig nichtlinearen Gleichungssystem. Deshalb werden typischerweise die Kontrollpunkte in einem linearen Ausgleichungsmodell geschätzt, nachdem zuvor die Knoten bestimmt wurden. Vor allem die Anzahl an Knoten als auch die Verteilung dieser im Parameterraum hat einen großen Einfluss auf das Approximationsergebnis.

Im ersten Teil dieser Arbeit werden deshalb, neben einem Grundlagenkapitel zu B-Splines und vorstellen der benutzten Daten, insgesamt drei unterschiedliche Methoden für das Auffinden geeigneter Knotenpositionen untersucht und die erzielten Ergebnisse präsentiert. Zur Evaluierung der Methoden stehen vier, aus B-Spline Kurven, extrahierte Punktmengen zur Verfügung, bei denen die Parameter (Kontrollpunkte und Knoten) bekannt sind. Die Vorteile und Nachteile der einzelnen Methoden bei der Knotenauffindung können somit durch das Vorhandensein einer Soll-Form und den bekannten Parametern aufgezeigt werden. Am Schluss dieses Evaluierungsprozesses zeigt sich eine klare Präferenz für jene Methode, welche einen Teil des natürlichen Immunsystems nachahmt und sich auf die *Clonal Selection* Theorie stützt.

Im zweiten Teil der Arbeit wird untersucht, ob mit Hilfe von unterschiedlichen Informationskriterien (AIC, BIC) die Anzahl an Knoten und somit die benötigte Anzahl an Parametern für ein zufriedenstellendes Approximationsergebnis bestimmt werden können. Es stellt sich heraus, dass mit der Kombination aus *Clonal Selection* Algorithmus und dem Bayesianischen Informationskriterium BIC die besten Ergebnisse erzielt werden. Alle gewonnen Erkenntnisse werden schlussendlich ebenfalls an realen Daten erprobt.

Abstract

As a result of the increase of areal measurement techniques, an accurate approximation of large point clouds for future applications is necessary. This is particularly true in the engineering geodesy. Suitable for this purpose are B-spline curves and B-spline surfaces, which provide with a suitable choice of the parameters a good fit to local features of the terrain. The adaptability of B-splines is essentially depending on two parameter sets, namely the control points and the knots. The simultaneous estimation of these parameters results in a highly non-linear system of equations. Therefore, in most applications, the knots are determined first. Subsequently the control points are estimated by solving a linear system of equations. The number of knots and their distribution of these in the parameter space have a significant impact on the quality of the approximation result.

At the beginning of the first part of this work, an introduction to the B-spline approximation and the data set used in the present work is given. Then three different methods for finding appropriate knot positions are analyzed and the corresponding results are presented. To evaluate these methods four point sets have been constructed from a given B-spline curve. Since the parameters, i.e. the control points and knots, of these approximations are known, the advantages and disadvantages of each method in identifying appropriate knots can be quantified. The evaluation process shows that the method which mimics a part of the natural immune system and is based on the clonal selection theory is superior among the methods considered here.

In the second part of this work different information criteria (AIC, BIC) are used in order to evaluate if it is possible to estimate the number of knots required to obtain a sufficiently accurate approximation result. It turns out that the best results are achieved with the combination of clonal selection algorithm and the bayesian information criterion BIC. Finally, all findings are validated against the results obtained from a realistic data set.

Danksagung

Im Folgenden darf ich mich bei jenen Personen bedanken, welche mich bei der Entstehung und Fertigstellung dieser Arbeit maßgeblich unterstützt haben.

Ein besonderer Dank gilt meinen beiden Betreuern, welche mich während der Entstehungszeit meiner Arbeit durch ihre fachliche und persönliche Unterstützung begleitet haben.

Martina, dir danke ich für die Unterstützung beim Verfassen der Arbeit, für die unzähligen Verbesserungsvorschläge während der Entstehungsphase und den großen persönlichen Rückhalt.

Darüber hinaus möchte ich mich noch besonders bei meiner ganzen Familie, insbesondere bei meinen Eltern, bedanken, die mir mein Studium durch ihre Unterstützung ermöglicht haben.

Inhaltsverzeichnis

Kurzfassung	i
Abstract	ii
Danksagung	iii
Inhaltsverzeichnis	iv
Abkürzungen	vi
1. Einleitung	1
1.1. Motivation	1
1.2. Aufbau der vorliegenden Arbeit	3
2. Grundlagen der B-Spline	4
2.1. B-Spline Basisfunktionen	4
2.2. B-Spline Kurven	6
2.2.1. Position und Vielfachheit der Knoten	7
2.2.2. Position der Kontrollpunkte	9
2.2.3. Grad der B-Spline Kurve	10
2.3. Ableitung von B-Spline Kurven	11
2.4. Eigenschaften von eingespannten B-Spline Kurven	12
2.5. Globale Kurvenapproximation mit Hilfe von B-Splines	12
2.5.1. Parametrisierung der Datenpunkte	13
2.5.2. Ermittlung eines Knotenvektors	14
2.5.3. Kleinste Quadrate Ausgleich	16
2.5.4. Motivation der Knotenoptimierung	19
3. Daten	21
3.1. Simulierte Datensätze	21
3.2. Reale Datensätze	23
4. Knotenschätzung	26
4.1. B-Spline Approximation mit dominanten Punkten (DOM)	26
4.1.1. Parametrisierung der Datenpunkte	28
4.1.2. Selektion von dominanten Punkten	28
4.1.3. Knotengenerierung auf Basis von dominanten Punkten	32

4.1.4.	Krümmungsberechnung anhand einer Basis B-Spline Kurve	32
4.1.5.	Evaluierung der Methode	38
4.2.	Gleichverteilung des Rauschverhältnis (GDR)	43
4.2.1.	Evaluierung der Methode	44
4.3.	Elitist Clonal Selection Algorithm (CSA)	51
4.3.1.	Evaluierung der Methode	55
5.	Modellselektion mit Hilfe von Informationskriterien	67
5.1.	Das Akaike Informationskriterium (AIC)	67
5.2.	Das Bayesianische Informationskriterium (BIC)	70
5.3.	Ermittlung der Knotenanzahl anhand von Informationskriterien	72
6.	Modellselektion und Knotenschätzung anhand realer Daten	78
7.	Zusammenfassung	87
	Literaturverzeichnis	90
	Abbildungsverzeichnis	92
	Tabellenverzeichnis	96

Abkürzungen

Ab	Anitbody (Antikörper)
Ag	Antigen
AIC	Akaike Information Criterium (Akaike Informationskriterium)
AICc	Akaike Information Criterium (Akaike Informationskriterium für kleine Stichprobengröße)
AIS	Artificial Immun System
AVG	knot averaging Technik
B-Spline	Basis-Spline
BIC	Bayesianisches Information Criterium (Bayesianisches Informationskriterium)
CSA	Methode der Knotenfindung mit Clonal Selection Algorithmus
DOM	Methode der Knotenfindung mit dominanten Punkten
GDR	Methode der Knotenfindung mit gleichverteilen des Rauschverhältnis
GMM	Gauß Markov Modell
KL	Kullback-Leibler
KTP	knot placement Technik
kumSum	kumulierte Summe
RMSE	Root Mean Square Error (Wurzel der mittleren Residuenquadratsumme)
RRR	Rot Ring Reflektor
RSS	Residual Sum of Squares (Residuenquadratsumme)
TLS	terrestrischer Laserscanner

1. Einleitung

1.1. Motivation

Mit der Entwicklung flächenhafter Messtechniken, wie beispielsweise dem terrestrischen Laserscanner TLS, gewinnen flächenhafte Auswertemethoden in der Ingenieurgeodäsie, neben den bestehenden herkömmlichen punktbasierten Methoden, stark an Bedeutung. Damit die Information und die Redundanz einer Punktwolke für spätere Prozessionschritte genutzt werden können, müssen diese zuerst in einer geeigneten Form modelliert werden. Kurven bzw. Flächen an Datenpunkte anzupassen, um somit ein digitales analytisches Modell für die Geometrie eines realen Objektes zu bekommen, ist dabei eine Schlüsseltechnologie im Bereich des Reverse Engineering. Ziel dabei ist es ein raumkontinuierliches mathematisches Modell zu erzeugen, welches das natürliche Objekt bestmöglich approximiert.

Hierfür werden zumeist Freiformkurven oder Freiformflächen herangezogen, welche sowohl einfache als auch komplexe Formgestalten modellieren können, wobei Freiformkurven die Basis für die Freiformflächen bilden. B-Splines sind dabei für gewöhnlich die am häufigsten verwendete Funktionsgruppe, da diese in ihrer geometrischen Anpassungsmöglichkeit sehr flexibel sind und positive mathematische Eigenschaften mit sich bringen. Im Allgemeinen werden bei der Approximation von Datenpunkten mittels B-Spline Kurve oder Fläche zuerst die unbekannt Parameterwerte der geordneten Datenpunkte berechnet. Auf dieser Basis wird anschließend ein Knotenvektor ermittelt und schlussendlich erhält man die fehlenden Kontrollpunkte in einem linearen Ausgleichungsschritt. Der Grad der B-Spline Kurve, welcher gemeinsam mit dem Knotenvektor mitentscheidend für die Berechnung der B-Spline Basisfunktionen ist, wird meistens empirisch auf einen festen Wert gesetzt. Dabei spielen vor allem die zu erwartenden weiteren Anwendungen, die Anforderung an die Kurvenkontinuität aber auch die Erfahrung aus den gemachten Beobachtungen eine entscheidende Rolle.

In der bisherigen Forschung hat sich gezeigt, dass neben der Verteilung der inneren Knoten, auch die Anzahl an benötigten Knoten bzw. Kontrollpunkten einen großen Einfluss auf die Genauigkeit des Approximationsergebnisses hat. Beide Einflussfaktoren sind jedoch a-priori unbekannt. Generell nimmt die Approximationsgenauigkeit mit steigender Modellkomplexität und somit mit steigender Anzahl an Knoten bzw. Kontrollpunkten zu. Ab wann jedoch ein Kurvenmodell die Daten zufriedenstellend approximiert, wird bisher meistens durch ein Qualitätskriterium festgelegt, welches oftmals die Komplexität des Modells außer Acht lässt.

Ziel dieser Arbeit ist es daher verschiedene Ansätze der Knotenpositionsbestimmung und der Modellselektion zu untersuchen und zu vergleichen. Mit Hilfe der in dieser Arbeit untersuchten Methoden kann schlussendlich jene Methode aufgezeigt werden, die sowohl die a-priori unbekannte Verteilung als auch die Anzahl an benötigten Knoten bestmöglich bestimmen kann.

1.2. Aufbau der vorliegenden Arbeit

Nachdem im Kapitel 1 die Motivation und die Struktur der vorliegenden Arbeit vorgestellt wird, wird im Kapitel 2 auf die Grundlagen von B-Spline Basisfunktionen und B-Spline Kurven eingegangen. Neben den wichtigsten Eigenschaften bezüglich der Basisfunktionen, wird vor allem darauf eingegangen, wie sehr und in welchem Ausmaße einzelne Parameter die Kurvengestalt von B-Spline Kurven beeinflussen können. Am Ende des Kapitels wird die gängige Praxis bei der Approximation von Datenpunkten durch eine B-Spline Kurve aufgezeigt.

In Kapitel 3 stehen neben simulierten Datensätzen auch reale Datensätze bereit, welche in den nachfolgenden Kapiteln benutzt werden um die einzelnen Methoden zu evaluieren. Dazu werden die einzelnen Datensätze beschrieben und die Vorgehensweise bei der Messung der realen Daten kurz erklärt.

Ein Kernthema dieser Arbeit ist die optimale Bestimmung der Knotenpositionen. Kapitel 4 stellt dazu insgesamt drei unterschiedliche Methoden für das Auffinden von Knotenpositionen vor. Die einzelnen Methoden basieren dabei auf der Veränderung des Krümmungsverhaltens, auf die Gleichverteilung des Rauschverhältnisses und auf evolutionäre Strategien. Die unterschiedlichen Vorgehensweisen und die entsprechenden Formeln jeder Methode werden dabei detailliert beschrieben. Anhand der simulierten Datensätze aus Kapitel 3 können die einzelnen Methoden bezüglich deren Grenzen zur Auffindung passender Knoten untersucht werden. Wesentliche Einflussgrößen, welche die Knotenauffindung der jeweiligen Methoden negativ oder positiv beeinflussen, werden durch Beispiele aufgezeigt.

Ein weiterer Schwerpunkt der vorliegenden Arbeit liegt in der Festlegung der Modellkomplexität. Da die Parameteranzahl und somit die zur Verfügung stehende Anzahl an Knoten das Approximationsergebnis wesentlich beeinflussen, werden im Kapitel 5 zwei Informationskriterien vorgestellt, welche das Problem der Modellselektion lösen sollen.

Im darauffolgenden Kapitel 6 werden die gewonnen Erkenntnisse anhand von realen Messdaten angewendet.

Kapitel 7 fasst die entsprechenden Resultate dieser Arbeit zusammen. Abschließend wird ein Ausblick auf weitere Fragestellungen, welche über den Umfang der vorliegenden Arbeit deutlich hinausgehen, gegeben.

2. Grundlagen der B-Spline

2.1. B-Spline Basisfunktionen

B-Spline Basisfunktionen werden in weitere Folge als Gewichte für die Bestimmung von B-Spline Kurvenpunkte benutzt. Um Basisfunktionen aufzustellen, wird zuallererst ein Knotenvektor $\mathbf{U} = \{u_0, u_1, \dots, u_m\}$ benötigt, wobei die einzelnen Elemente als Knoten u_i bezeichnet werden. Die Elemente des Knotenvektors müssen in aufsteigender Reihenfolge sortiert sein. Sind die Abstände zwischen den Knoten gleich, so werden die Knoten als äquidistant bezeichnet. Das halb-offene Intervall $[u_i, u_{i+1})$ wird als die i -te Knotenspanne bezeichnet. Ein Knoten, welcher k -mal auftritt ($u_i = u_{i+1} = \dots = u_{i+k-1}$), hat zur Folge, dass einige Knotenspannen nicht existieren. Außerdem beeinflusst ein Knoten mit der Vielfachheit k im späteren Verlauf nicht nur das Aussehen sondern auch die Kontinuität der B-Spline Kurve (siehe Kapitel 2.2.1). Der Definitionsbereich der Basisfunktionen wird durch den Anfangs- und Endknoten vorgegeben. Oft wird für $u_0 = 0$ und für $u_m = 1$ herangezogen, wodurch sich das geschlossene Definitionsintervall $[0, 1]$ ergibt (Piegl & Tiller, 1997, S.50 ff.).

Um B-Spline Basisfunktionen zu definieren, wird außerdem der Grad p der Basisfunktion benötigt. Die i -te B-Spline Basisfunktion $N_{i,p}(u)$ vom Grad p kann rekursiv über den *Cox-de Boor* Algorithmus berechnet werden (Cox, 1972; de Boor, 1972).

$$N_{i,0}(u) = \begin{cases} 1 & \text{wenn } u_i \leq u < u_{i+1} \\ 0 & \text{ansonsten} \end{cases} \quad (2.1)$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u)$$

Ist der Grad $p = 0$, so nimmt die Basisfunktion die einfache Form einer Rechtecksfunktion an, wobei die i -te Basisfunktion nur in der i -ten Knotenspanne den Wert 1 annimmt.

Die Basisfunktion $N_{i,p}(u)$ wird mit Hilfe von zwei Koeffizienten als Linearkombination der Basisfunktionen $N_{i,p-1}(u)$ und $N_{i+1,p-1}(u)$ gebildet. Beide Koeffizienten sind linear in u und liegen im Wertebereich von 0 bis 1 (siehe Abb. 2.1). Daraus folgt, dass die Basisfunktion $N_{i,p}(u)$ im halboffenen Bereich $[u_i, u_{i+p+1})$ nicht null ist oder anders ausgedrückt $N_{i,p}(u)$ ist ungleich null in $p+1$ Knotenspannen $[u_i, u_{i+1}), [u_{i+1}, u_{i+2}), \dots, [u_{i+p}, u_{i+p+1})$ (siehe Abb. 2.2 rote Umrandung) (Piegl & Tiller, 1997, S.50 ff.).

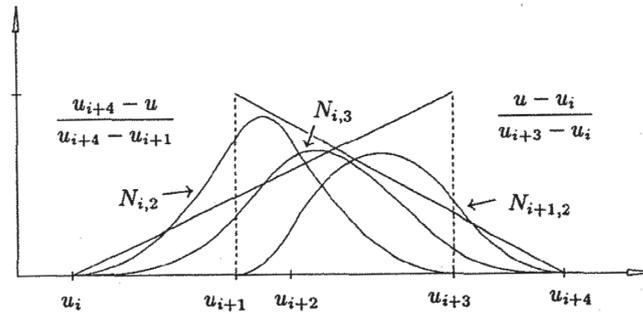


Abb. 2.1: Darstellung der rekursiven Definition von B-Spline Basisfunktionen (Piegl & Tiller 1997, S.51)

Das rekursive Berechnungsschema von B-Spline Basisfunktionen kann in folgender Abbildung betrachtet werden.

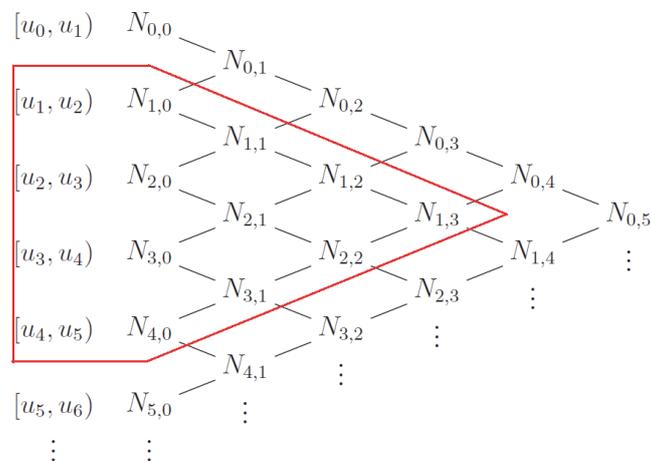


Abb. 2.2: Dreiecksschema zur Berechnung von B-Spline Basisfunktionen

Im Folgenden sind die wichtigsten Eigenschaften von B-Spline Basisfunktionen zusammengefasst (vgl. Piegl & Tiller, 1997, S.63 ff.):

1. $N_{i,p}(u)$ ist ein Polynom vom Grad p in u
2. $N_{i,p}(u)$ ist für alle $i \geq 0, p \geq 0$ und u nicht negativ
3. $N_{i,p}(u)$ ist nicht null im Intervall $[u_i, u_{i+p+1})$
4. In einer Knotenspanne $[u_i, u_{i+1})$ sind höchstens $p+1$ Basisfunktionen vom Grad p ungleich null (nämlich $N_{i-p,p}(u), N_{i-p+1,p}(u), \dots, N_{i,p}(u)$)
5. Die Summe aller $p+1$ Basisfunktionen, welche an einer Stelle u nicht null sind, ist immer 1
6. Ist die Anzahl an Knoten $m+1$, der Grad der Basisfunktionen p und die Anzahl an Kontrollpunkte $n+1$, so muss die Gleichung $m = n + p + 1$ erfüllt werden
7. An einem Knoten u_i mit Vielfachheit k weist die Basisfunktion $N_{i,p}(u_i)$ nur eine C^{p-k} Stetigkeit auf

2.2. B-Spline Kurven

Eine B-Spline Kurve wird über $n+1$ Kontrollpunkte P_0, P_1, \dots, P_n , einen Knotenvektor $U = \{u_0, u_1, \dots, u_m\}$ mit $m+1$ Knoten u_i und dem Grad p der B-Spline Kurve bestimmt. Zu beachten ist, dass der Grad p der Kurve nicht, wie etwa bei Bézier Kurven, durch die Anzahl der Kontrollpunkte vorgegeben ist, sondern frei wählbar ist. Dadurch wird eine größere Freiheit bei der Modellierung von Kurven gewährleistet. Oft wird anstatt des Grades p die Ordnung $o = p + 1$ der Kurve angegeben (Piegl & Tiller, 1997, S.81 ff.). In dieser Arbeit wird immer der Grad p und nicht die Ordnung benutzt.

B-Spline Kurvenpunkte $C(u)$ können anschließend wie folgt berechnet werden:

$$C(u) = \sum_{i=0}^n N_{i,p}(u) P_i \quad (2.2)$$

Dabei läuft der Parameter u zumeist über das Intervall $[0,1]$. Ein Kurvenpunkt $C(u)$ am Parameterwert u berechnet sich dabei durch eine Linearkombination aus $p+1$ Kontrollpunkten, welche durch positive Basisfunktionen gewichtet werden. Die restlichen Basisfunktionen $N_{i,p}(u)$, welche mit Hilfe des Knotenvektors U und dem Grad p gebildet werden (siehe Kapitel 2.1), haben an dieser Stelle den Wert null, wodurch auch die entsprechenden Kontrollpunkte keinen Einfluss auf den Kurvenpunkte haben. Zu beachten ist, dass n , m und p folgende Bedingung erfüllen müssen.

$$m = n + p + 1 \quad (2.3)$$

Wird die Anzahl an Kontrollpunkten und Knoten vorgegeben, so ist automatisch der Grad der B-Spline Kurve festgelegt. Jene Punkte auf der Kurve, welche mit den Parametern der Knoten u_i gebildet werden, bezeichnet man als Knotenpunkte. Diese wiederum unterteilen die B-Spline Kurve in Kurvensegmente $C_i(u)$ (Abb. 2.3) vom Grad p (Piegl & Tiller, 1997, S.81 ff.).

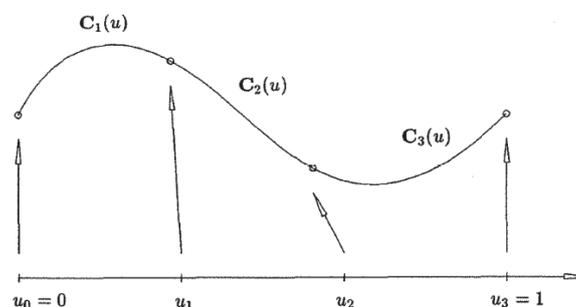


Abb. 2.3: Stückweise kubische polynomiale Kurve mit drei Segmenten (Piegl & Tiller 1997, S.48)

Die Form von B-Spline Kurven wird maßgeblich von folgenden Kontrollparametern bestimmt:

- Position und Vielfachheit der Knoten u_i
- Position der Kontrollpunkte P_i
- Grad p der Kurve

Im folgenden Abschnitt wird gezeigt, wie sich die einzelnen Kontrollparameter auf die Form der Kurve auswirken. Um die einzelnen Auswirkungen bzw. Unterschiede klar und deutlich sichtbar zu machen, wird eine 2D-Darstellung in der vorliegenden Arbeit bei B-Spline Kurven bevorzugt.

2.2.1. Position und Vielfachheit der Knoten

Anhand der Formel (2.2) ist zu sehen, dass die B-Spline Kurve erst dann durch den ersten und letzten Kontrollpunkt verläuft, sobald $N_{0,p}(u_0) = N_{n,p}(u_m) = 1$ gilt. Dies wiederum wird dadurch erreicht, dass der Anfangs- und Endknoten eine Vielfachheit von $p+1$ aufweist.

$$U = \{u_0 = u_1 = \dots = u_p, u_{p+1}, \dots, u_{m-p-1}, u_{m-p} = u_{m-p+1} = \dots = u_m\} \quad (2.4)$$

Die daraus resultierende Kurve wird eingespannte B-Spline Kurve genannt und ist im Bereich $[u_0, u_m]$ definiert. Fehlt diese Vielfachheit des Anfangs- und Endknoten so ist die daraus resultierende offene B-Spline Kurve nur im Bereich $[u_p, u_{m-p}]$ definiert und geht nicht durch den ersten und letzten Kontrollpunkt (Piegl & Tiller, 1997, S.572). In den nachfolgenden Betrachtungen werden immer eingespannte B-Spline Kurven verwendet.

Die Menge $\{u_{p+1}, \dots, u_{m-p-1}\}$ bildet die inneren Knoten. Auch diese können eine Vielfachheit $k > 1$ aufweisen. Generell ist die Kontinuität und Differenzierbarkeit der B-Spline Kurve $C(u)$ direkt von den Basisfunktionen $N_{i,p}(u)$ abhängig, da wie bereits erwähnt, $C(u)$ als Linearkombination der Basisfunktionen und Kontrollpunkten gebildet wird. Daraus folgt, dass eine B-Spline Kurve vom Grad p an inneren Knoten mit Vielfachheit k , C^{p-k} stetig ist (Piegl & Tiller, 1997, S.88 f.).

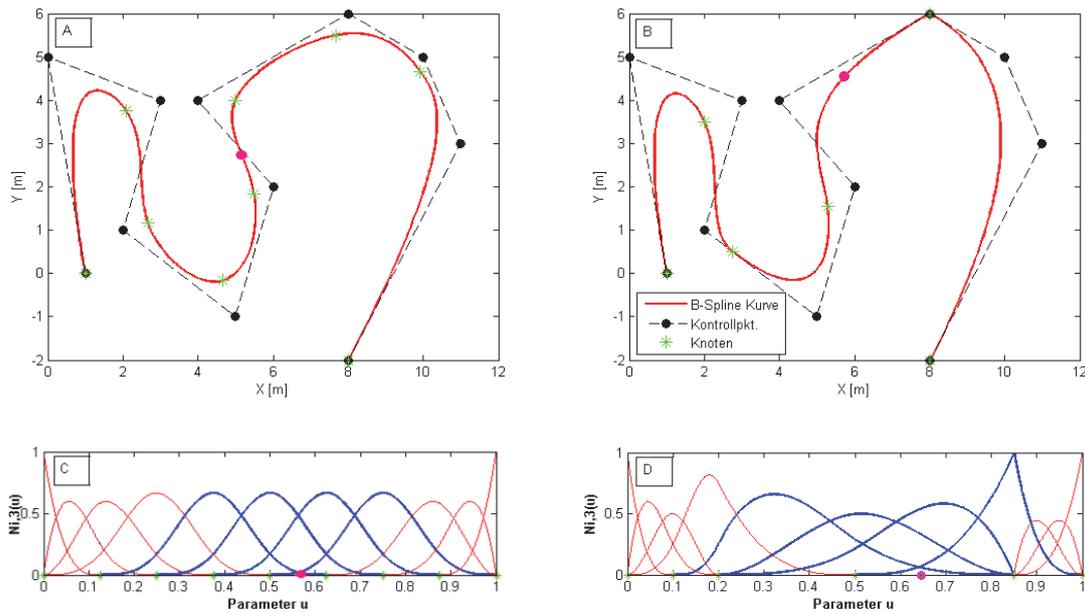


Abb. 2.4: A) B-Spline mit äquidistantem Knotenvektor U_1 C) und zugehörigen Basisfunktionen
 B) B-Spline mit nicht äquidistantem Knotenvektor U_2 D) und zugehörigen Basisfunktionen
 beide B-Spline Kurven vom Grad $p=3$ benutzen die gleichen Kontrollpunkte P

$$U_1 = \left[0, 0, 0, 0, \frac{1}{8}, \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}, \frac{3}{4}, \frac{7}{8}, 1, 1, 1, 1 \right] \quad U_2 = \left[0, 0, 0, 0, \frac{1}{10}, \frac{1}{5}, \frac{1}{5}, \frac{1}{2}, \frac{17}{20}, \frac{17}{20}, \frac{17}{20}, 1, 1, 1, 1 \right]$$

$$P = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 & 2 & 5 & 6 & 4 & 8 & 10 & 11 & 8 \\ 0 & 5 & 4 & 1 & -1 & 2 & 4 & 6 & 5 & 3 & -2 \end{bmatrix} \quad p = 3$$

In Abb. 2.4 wird eine B-Spline Kurve vom Grad 3 mit zwei unterschiedlichen Knotenvektoren U gebildet, wobei beide Male das gleiche Kontrollpolygon (=Verbindung der einzelnen Kontrollpunkte) zur Anwendung kommt. Die linke Abbildung zeigt jene Kurve mit einem äquidistanten Knotenvektor. Die dazugehörigen Basisfunktionen vom Grad 3 und die Knotenverteilung im Intervall $[0,1]$ sind im darunterliegenden Bild zu sehen. Die Position des eingezeichneten rosa Kurvenpunkts ist, bei gleichbleibendem Kontrollpolygon, von $p+1$, von null verschieden, blau eingefärbten Basisfunktionen abhängig. Wird die Position der Knoten geändert, so hat dies Auswirkungen auf die Basisfunktionen (Abb. 2.4 D). Bei der Abbildung B wurde der Knotenvektor so geändert, dass die Knotenabstände nicht mehr äquidistant sind. Außerdem wurde an der Stelle $u = \frac{1}{5}$ ein Knoten mit Vielfachheit $k = 2$ und an der Stelle $u = \frac{17}{20}$ ein Knoten mit Vielfachheit $k = 3$ eingefügt. An diesen Stellen weist die B-Spline Kurve nur noch eine C^1 bzw. C^0 Stetigkeit auf. So tangiert die Kurve im Parameter $u = \frac{1}{5}$ (entspricht zweiter innerer Knoten) das Kontrollpolygon. An der Stelle $u = \frac{17}{20}$ bekommt die Kurve einen Knick (entspricht C^0 Stetigkeit) und verläuft durch den Kontrollpunkt P_7 , da an dieser Stelle nur noch eine Basisfunktion ungleich null ist $N_{7,3}\left(\frac{17}{20}\right) = 1$ und somit der Kurvenpunkt $C\left(\frac{17}{20}\right)$ dem Kontrollpunkt entspricht.

2.2.2. Position der Kontrollpunkte

Eine wichtige Eigenschaft von B-Spline Kurven ist die lokale Modifikation der Kurve über Kontrollpunkte. Diese Eigenschaft folgt direkt daraus, dass $N_{i,p}(u)$ nur im Intervall $[u_i, u_{i+p+1})$ ungleich null ist (siehe Absatz 2.1 bzw. Abb. 2.2). Liegt der Parameter u nicht in diesem Intervall, so wirkt sich eine Änderung des Kontrollpunktes P_i nicht auf die B-Spline Kurve $C(u)$ aus. Somit hat eine Änderung des Kontrollpunktes P_i lediglich Auswirkungen auf das Kurvenintervall $[u_i, u_{i+p+1})$ (Piegl & Tiller, 1997, S.84 ff.). Dieses Verhalten kann in Abb. 2.5 betrachtet werden. Es wurde der gleiche Knotenvektor, der gleiche Grad und die gleichen Kontrollpunkte wie in Abb. 2.4 A benutzt. Lediglich der sechste Kontrollpunkt wurde verschoben. Dabei verändert sich das Aussehen der Kurve nur in vier ($= p + 1$) Kurvensegmenten (blaue Linie).

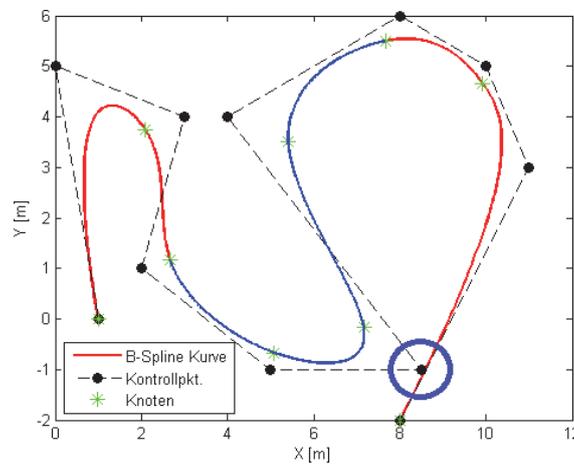


Abb. 2.5: Änderung der Position des sechsten Kontrollpunktes (vgl. Abb. 2.4 A) bewirkt eine lokale Änderung der B-Spline Kurve im Intervall $[u_6, u_{10})$

Dieses Modifikationsschema ist ein wichtiges Instrument um Kurven zu designen, da die Kurve nur lokal und nicht global verändert wird. Dies lässt sich auch mit folgenden Formeln zeigen, wobei $(P_i + v)$ den veränderten Kontrollpunkt darstellt.

$$\begin{aligned}
 M(u) &= \sum_{j=0}^{i-1} N_{j,p}(u)P_j + N_{i,p}(u)(P_i + v) + \sum_{j=i+1}^n N_{j,p}(u)P_j \\
 &= \sum_{j=0}^n N_{j,p}(u)P_j + N_{i,p}(u)v \\
 &= C(u) + N_{i,p}(u)v
 \end{aligned} \tag{2.5}$$

Die neu entstandene Kurve $M(u)$ besteht dabei aus der „alten“ Kurve $C(u)$ und einem Translationsvektor $N_{i,p}(u)v$, welcher nur im Intervall $[u_i, u_{i+p+1})$ ungleich null ist.

Es ist ersichtlich, dass durch das Hinzufügen von weiteren Kontrollpunkten, wobei auch gleichzeitig die Anzahl der Knoten erhöht werden muss, damit die Bedingung (2.3) erfüllt

wird, der betroffene Bereich eingegrenzt werden kann und somit eine feinere Modellierung der Kurve möglich ist.

Ähnlich wie bei Knoten mit Vielfachheit k , kann auch ein Kontrollpunkt öfters als einmal auftreten. Liegt dabei der Parameter u im Intervall $[u_i, u_{i+1})$, so liegt der B-Spline Kurvenpunkt in der konvexen Hülle, aufgespannt durch die Kontrollpunkte $P_i, \dots, P_{i-p+1}, P_{i-p}$. Werden p Kontrollpunkte ident ($P_i = P_{i-1} = \dots = P_{i-p+1}$), so bildet sich die konvexe Hülle zu einem Liniensegment $P_{i-p}P_i$ um und die B-Spline Kurve muss den Kontrollpunkt P_i passieren, wobei die Kurve an diesem Punkt eine C^0 Stetigkeit aufweist (Abb. 2.6 rechts) (Piegl & Tiller, 1997, S.90 ff.).

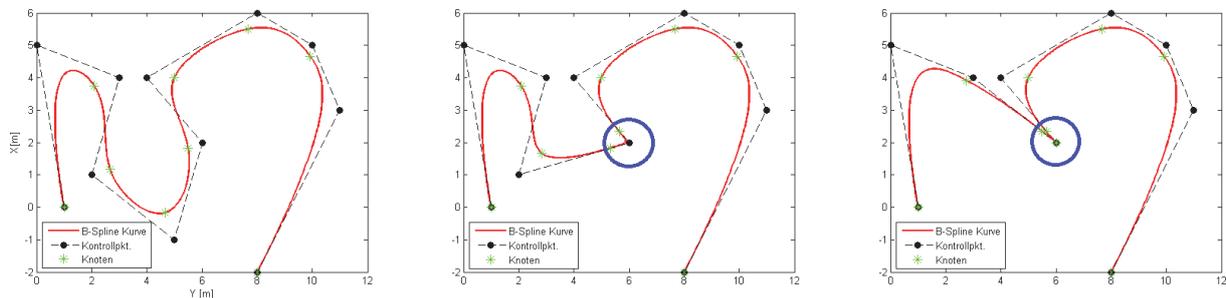


Abb. 2.6: B-Spline Kurve mit einfachem, doppeltem ($P_4 = P_5$) und dreifachem Kontrollpunkt ($P_3 = P_4 = P_5$)

2.2.3. Grad der B-Spline Kurve

Der Grad einer B-Spline Kurve ändert das Aussehen der Kurve global (Abb. 2.7). Ausschlaggebend hierfür sind wiederum die B-Spline Basisfunktionen. Wie bereits in Abschnitt 2.1 gezeigt wurde, ist eine Basisfunktion $N_{i,p}(u)$ nur im Intervall $[u_i, u_{i+p+1})$ nicht null. Wird also der Grad p der Basisfunktion und somit der B-Spline Kurve erhöht, so wird auch gleichzeitig das Intervall, in welchem die Basisfunktion ungleich null ist, vergrößert (zu sehen in Abb. 2.2). Ein Kurvenpunkt $C(u)$ wird anschließend durch $p+1$ Basisfunktionen und deshalb auch durch $p+1$ Kontrollpunkte beeinflusst. Ausgehend von der Anzahl an Kontrollpunkten $n+1$, kann der Grad einer B-Spline Kurve höchstens $n = p$ sein. In diesem Fall wird jeder Kurvenpunkt durch alle Kontrollpunkte beeinflusst und die B-Spline Kurve verwandelt sich zu einer Bézier-Kurve (Abb. 2.7 D). Es existieren keine inneren Knoten mehr wodurch die gesamte Kurve aus einem einzelnen Kurvensegment vom Grad p besteht. Die Bézier-Kurve ist deshalb ein Spezialfall der B-Spline Kurve (Piegl & Tiller, 1997, S.84 ff.).

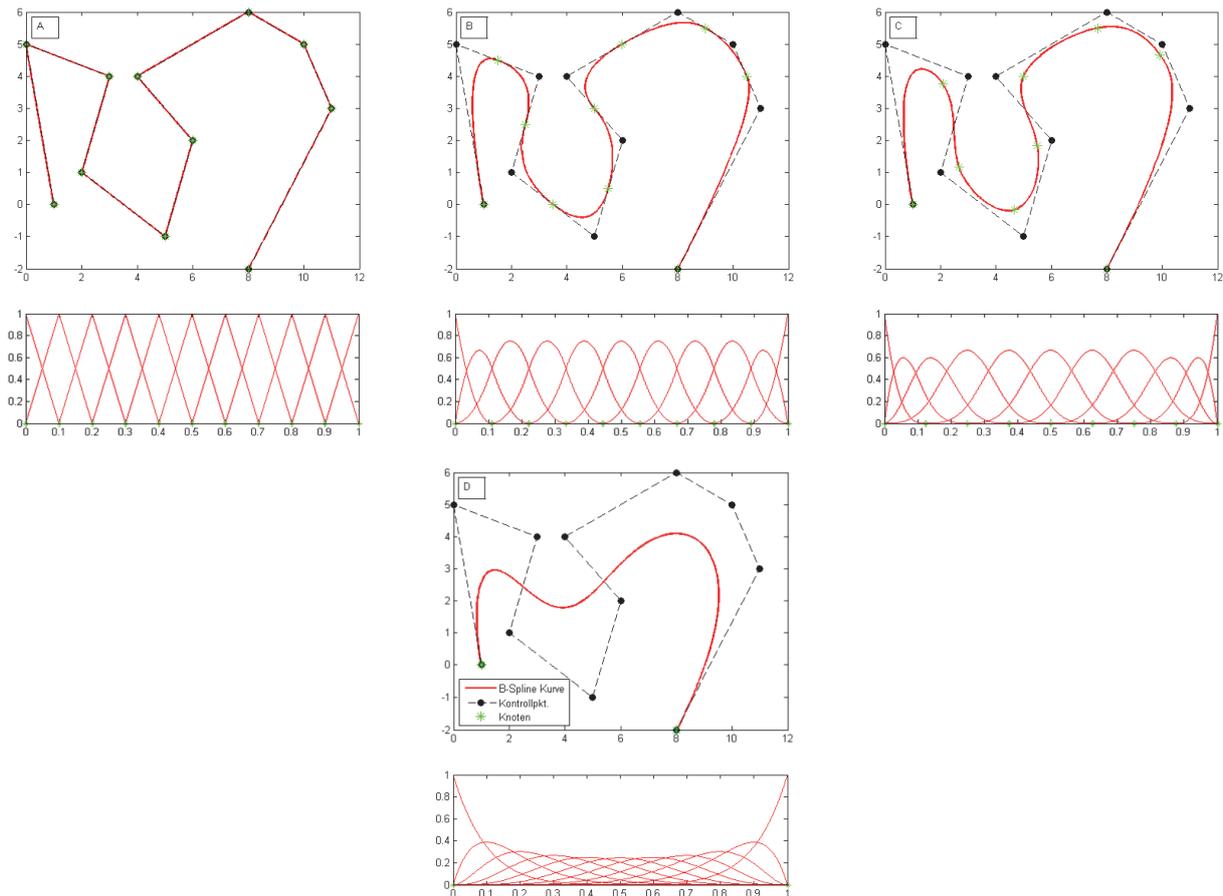


Abb. 2.7: B-Spline Kurve vom Grad A) $p=1$; B) $p=2$; C) $p=3$
 D) B-Spline Kurve vom Grad $p=10=n$ (bei $n+1 = 11$ Kontrollpunkten \rightarrow Bézier-Kurve)
 unter jeder Kurvenabbildung sind die entsprechenden Basisfunktionen zu sehen

2.3. Ableitung von B-Spline Kurven

Die Ableitung einer B-Spline Kurve kann durch das Ableiten der Basisfunktionen geschehen. Die k -te Ableitung einer Basisfunktion kann wie die Basisfunktion selbst über eine rekursive Formel berechnet werden.

$$N_{i,p}^{(k)}(u) = p \left(\frac{N_{i,p-1}^{(k-1)}(u)}{u_{i+p} - u_i} - \frac{N_{i+1,p-1}^{(k-1)}(u)}{u_{i+p+1} - u_{i+1}} \right) \quad (2.6)$$

Durch einsetzen von (2.6) in (2.2) bekommt man anschließend die k -te Ableitung $C^{(k)}(u)$ der B-Spline Kurve $C(u)$. Dabei existieren lediglich die k -te Ableitung solange $k \leq p$ gilt. Alle höheren Ableitungen sind gleich null (Piegl & Tiller, 1997, S.61).

Die Ableitung der B-Spline Kurve kann auch auf eine weitere Art geschehen. Hierfür wird zuerst die Vielfachheit der Anfangs- und Endknoten im Knotenvektor \mathbf{U} (2.4) der B-Spline Kurve $C(u)$ verringert, indem jeweils ein Knoten entfernt wird, wodurch ein neuer Knotenvektor $\mathbf{U}^|$ mit $m - 1$ Knoten entsteht.

$$\mathbf{U}^| = \{u_0 = u_1 = \dots = u_{p-1}, u_p, \dots, u_{m-p-1}, u_{m-p} = u_{m-p+1} = \dots = u_{m-1}\} \quad (2.7)$$

Die durch diesen Knotenvektor berechneten Basisfunktionen $N_{i,p-1}(u)$ stellen gemeinsam mit dem neu eingeführten Kontrollpolygon \mathbf{Q} die B-Spline Kurve vom Grad $p - 1$ dar.

$$C^1(u) = \sum_{i=0}^{n-1} N_{i,p-1}(u) \mathbf{Q}_i \quad \text{mit} \quad \mathbf{Q}_i = p \frac{\mathbf{P}_{i+1} - \mathbf{P}_i}{u_{i+p+1} - u_{i+1}} \quad (2.8)$$

Diese Art der Ableitungsberechnung ist wesentlich effizienter als jene über die rekursive Berechnung der Basisfunktionen mit Formel (2.6) (Piegl & Tiller, 1997, S.93 f.).

2.4. Eigenschaften von eingespannten B-Spline Kurven

Nachfolgend werden die wichtigsten Eigenschaften von eingespannten B-Spline Kurven zusammengefasst, wobei von einer B-Spline Kurve $C(u)$ vom Grad p mit $n+1$ Kontrollpunkten \mathbf{P}_i und einem Knotenvektor $\mathbf{U} = \{u_0, u_1, \dots, u_m\}$ der Länge $m+1$ ausgegangen wird (Piegl & Tiller, 1997, S.81 f.).

1. Eine B-Spline Kurve vom Grad p ist eine an den Knotenpunkten aus Kurvensegmenten stückweise zusammengesetzte Kurve, wobei jedes Kurvensegment ebenfalls eine Kurve vom Grad p ist.
2. Sind $p+1$ Anfangs- bzw. Endknoten gleich, so entsteht eine eingespannte B-Spline Kurve, welche den Anfangs- und Endkontrollpunkt enthält.
3. Ein Kurvenpunkt $C(u)$ befindet sich in der konvexen Hülle der Kontrollpunkte $\mathbf{P}_{i-p}, \mathbf{P}_{i-p+1}, \dots, \mathbf{P}_i$, falls sich u im Intervall $[u_i, u_{i+1})$ befindet. Deshalb verläuft die gesamte B-Spline Kurve immer in der konvexen Hülle, welche durch die Kontrollpunkte aufgespannt wird.
4. Ändert sich die Position eines Kontrollpunktes \mathbf{P}_i so ändert sich die Kurve nur lokal im Intervall $[u_i, u_{i+p+1})$.
5. Eine B-Spline Kurve $C(u)$ weist an Knoten mit Vielfachheit k eine C^{p-k} Stetigkeit auf.
6. Wenn der Grad p der B-Spline Kurve gleich der Anzahl der Kontrollpunkte minus eins ist ($n = p$) und es jeweils $p+1$ gleiche Anfangs- und Endknoten gibt, so reduziert sich die B-Spline Kurve zu einer Bézier-Kurve.
7. Wird eine B-Spline Kurve $C(u)$ einer affinen Transformation unterworfen, so kann das gleiche Resultat erzielt werden, wie wenn nur die Kontrollpunkte \mathbf{P}_i der affinen Transformation unterworfen werden.

2.5. Globale Kurvenapproximation mit Hilfe von B-Splines

Soll eine Menge an Datenpunkten $\{\mathbf{D}_0, \mathbf{D}_1, \dots, \mathbf{D}_h\}$ durch eine B-Spline Kurve vom Grad p approximiert werden, so werden typischerweise die Kontrollpunkte \mathbf{P}_i in einem linearen

Ausgleichungsmodell geschätzt (siehe Abschnitt 2.5.3). Die zugehörigen Parameterwerte der Datenpunkte $\{\bar{u}_0, \bar{u}_1, \dots, \bar{u}_h\}$, welche die Positionen auf der Kurve festlegen, werden zuvor meist durch die *uniform*, die *chord* bzw. die *centripetal* Methode bestimmt (Piegl & Tiller, 1997, 364 ff.). Die Knoten, welche für die Berechnung der Basisfunktionen notwendig sind, werden dabei in einem weiteren separaten Abschnitt, zumeist ausgehend von den Parameterwerten der Datenpunkte, ermittelt.

2.5.1. Parametrisierung der Datenpunkte

Die einfachste Methode um Parameterwerte für die Datenpunkte zu ermitteln ist die *uniform* Methode. Wie bereits der Name erahnen lässt, werden in einem Intervall $[a, b]$ $h+1$ gleichmäßig verteilte Parameter ermittelt.

$$\begin{aligned}\bar{u}_0 &= a \\ \bar{u}_i &= \frac{i}{h} \quad \text{für } 1 \leq i \leq h-1 \\ \bar{u}_h &= b\end{aligned}\tag{2.9}$$

Diese Art der Parameterermittlung ist meistens ungeeignet, da es zu unkontrollierten Kurvenverläufen kommen kann, sobald die zu den Parameterwerten zugehörigen Datenpunkte nicht gleichmäßig verteilt sind. Um dies auszuschließen wurde die *chord* und die *centripetal* Methode entwickelt. Die *chord* Methode nimmt an, dass die approximierende Kurve in der Nähe der Datenpunkte verläuft. Somit sollte die Länge eines Kurvenabschnittes zwischen zwei benachbarten Kurvenpunkten ähnlich zu der Sehnenlänge der beiden Datenpunkte sein. Daraus ergeben sich folgende Formeln zur Berechnung der Parameterwerte.

$$\begin{aligned}\bar{u}_0 &= a \\ \bar{u}_k &= a + (b - a) * \frac{1}{L} \left(\sum_{i=1}^k |D_i - D_{i-1}|^d \right) \quad \text{für } k = 1 \leq k \leq h-1 \\ \bar{u}_m &= b\end{aligned}\tag{2.10}$$

wobei $L = \sum_{i=1}^h |D_i - D_{i-1}|^d$

Für die *chord* Methode wird $d = 1$ und für die *centripetal* Methode $d = 0.5$ benutzt. Die *centripetal* Methode erzielt bessere Ergebnisse als die *chord* Methode sobald mit Datenpunkten gerechnet wird, welche ein plötzliches Wendeverhalten aufweisen (Lee, 1989). Die Parametrisierung der Datenpunkte bei der Kurvenschätzung ist dabei unabhängig von der Orientierung der Kurve im Raum.

Wie sehr die Parametrisierung die Form einer B-Spline Kurve beeinflussen kann ist anhand der Abb. 2.8 zu sehen. Die *chord* Methode führt in diesem Beispiel zu einem unbrauchbaren und unerwünschten Ausschwingen der Kurve. Dieses Verhalten ist typisch bei dieser Methode, sollte eine Sehne zwischen zwei Datenpunkten deutlich länger sein als andere. Die *centripetal* Methode versucht dieses Ungleichgewicht, welches durch die unterschiedlichen Sehnenlängen entsteht, auszugleichen. Am besten jedoch erfolgt in diesem Beispiel die Interpolation der sieben Datenpunkte mit Hilfe der *uniform* Methode. Es sei darauf hingewiesen, dass durch dieses Beispiel nicht davon ausgegangen werden kann, dass die *uniform* Methode als die Geeignetste gilt. Vielmehr soll darauf hingewiesen werden, wie sehr bereits die Parametrisierung der Datenpunkte die Form der B-Spline Kurve beeinflussen kann.

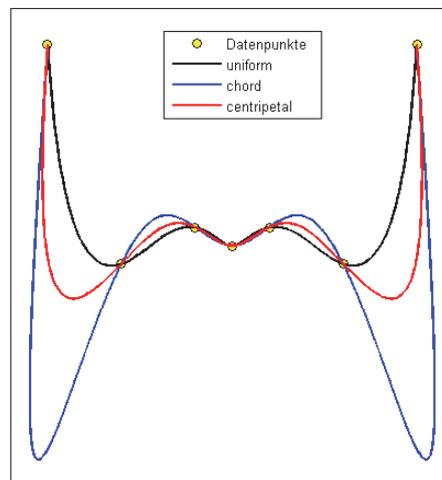


Abb. 2.8: Einfluss unterschiedlicher Parametrisierungsmethoden auf die Approximation von Datenpunkten durch eine B-Spline Kurve

2.5.2. Ermittlung eines Knotenvektors

Sobald ein Set an Parameterwerten für die $h + 1$ Datenpunkte ermittelt wurde, kann ein Knotenvektor U generiert werden. Die Anzahl der Knoten ($m + 1$) muss die Gleichung (2.3) erfüllen. Deshalb muss a priori bereits die Anzahl an gewünschten Kontrollpunkten und der Grad p der späteren B-Spline Kurve festgelegt sein. Da in der vorliegenden Untersuchung eingespannte Kurven betrachtet werden, müssen dafür insgesamt $n - p$ innere Knoten aufgefunden werden. Grundsätzlich kann ein Knotenvektor auch ohne Parameterwerte ermittelt werden. Dies sollte jedoch vermieden werden, da es dabei vorkommen kann, dass das Gleichungssystem (2.2) singulär wird, da zur Lösung mindestens ein Datenpunkt in jeder Knotenspanne benötigt wird (Piegl & Tiller, 1997, S.365). Die herkömmliche Methode zur Bestimmung eines Knotenvektors wird nachfolgend aufgezeigt.

Bei der *knot averaging* (AVG) Technik (Anzahl an Knoten $m+1 =$ Anzahl an Kontrollpunkten $n+1$) werden die inneren Knoten durch die Mittelung von p Datenparameterwerten ermittelt. Die so ermittelten Knotenparameter u geben die Verteilung der Datenparameter \bar{u} wieder.

$$\begin{aligned} u_0 &= \dots = u_p = 0 \\ u_{i+p} &= \frac{1}{p} \sum_{j=i}^{i+p-1} \bar{u}_j \quad (i = 1, \dots, n-p) \\ u_{m-p} &= \dots = u_m = 1 \end{aligned} \quad (2.11)$$

Sobald $m > n$ gilt, wird in den meisten Fällen die *knot placement* (KTP) Technik eingesetzt.

$$\begin{aligned} u_0 &= \dots = u_p = 0 \\ u_{i+p} &= (1 - \alpha)\bar{u}_{i-1} + \alpha\bar{u}_i \quad (i = 1, \dots, n-p) \\ u_{m-p} &= \dots = u_m = 1 \end{aligned} \quad (2.12)$$

$$\text{wobei } \alpha = id - j \quad \text{mit } j = \text{int}(id) \quad \text{und } d = \frac{h+1}{n-p+1}$$

Dadurch ist sichergestellt, dass in jeder zukünftigen Knotenspanne zumindest ein Datenpunkt vorhanden ist, wodurch die Normalgleichungsmatrix $N^T N$ bei der Ausgleichung positiv definit ist und eine gute Kondition aufweist (Piegl & Tiller, 1997, S.412; Schoenberg & Whitney, 1953). Die *knot placement* Technik verfolgt dabei das Ziel, in Bereichen mit vielen Datenpunkten die Knotendichte zu erhöhen, da davon ausgegangen wird, dass in diesen Bereichen viel Information vorhanden ist, welche es zu modellieren gilt. Bei äquidistant verteilten Datenpunkten würde mit dieser Technik ein äquidistanter Knotenvektor generiert werden. Zumeist liefert diese Vorgehensweise jedoch nicht die erwünschte oder geforderte Approximationsgenauigkeit. Besonders bei komplexen Kurvenformen zeigen sich die Nachteile dieser Technik.

Neben dieser konventionellen Art der Knotengenerierung gibt es unzählige Methoden um einen Knotenvektor zu generieren. Bei der Knotengenerierung besteht allgemein das Problem, dass es einerseits noch keine analytische Beschreibung für das Auffinden einer optimalen Knotenposition gibt und andererseits mehrere lokale Optima im Ausgleichungsmodell existieren (Galvez et al., 2015). Im Kapitel 4 werden drei unterschiedliche Methoden vorgestellt, welche für die Knotengenerierung nicht primär die Parametrisierung der Datenpunkte benutzen.

2.5.3. Kleinste Quadrate Ausgleich

Wurde für die $h + 1$ Datenpunkte $\{\mathbf{D}_0, \mathbf{D}_1, \dots, \mathbf{D}_h\}$ ein Set an Parameterwerten $\{\bar{u}_0, \bar{u}_1, \dots, \bar{u}_h\}$ und ein Knotenvektor \mathbf{U} ermittelt, so können die $n + 1$ Kontrollpunkte \mathbf{P}_i einer approximierenden B-Spline Kurve vom Grad p in einem linearen Ausgleichungsproblem bestimmt werden. Dabei wird zumeist gefordert, dass die Kurve den Anfangs- und Endpunkt des Kontrollpolygons enthält:

$$\mathbf{C}(\bar{u}_0) = \mathbf{P}_0 = \mathbf{D}_0 \quad \mathbf{C}(\bar{u}_h) = \mathbf{P}_n = \mathbf{D}_h. \quad (2.13)$$

Die restlichen gesuchten Kontrollpunkte $\mathbf{P}_1, \dots, \mathbf{P}_{n-1}$ sollen so bestimmt werden, dass der Abstand zwischen Daten- und korrespondierenden Kurvenpunkten ein Minimum annimmt.

$$f(\mathbf{P}_1, \dots, \mathbf{P}_{n-1}) = \sum_{k=1}^{h-1} |\mathbf{D}_k - \mathbf{C}(\bar{u}_k)|^2 \quad (2.14)$$

Dies führt, solange keine Nebenbedingungen erfüllt werden müssen, zu einem Ausgleich im Gauß-Markov-Modell (GMM). Somit gilt

$$(\mathbf{N}^T \mathbf{N}) \mathbf{P} = \mathbf{R}, \quad (2.15)$$

wobei \mathbf{N} eine $(h - 1) \times (n - 1)$ Matrix der Form

$$\mathbf{N} = \begin{bmatrix} N_{1,p}(\bar{u}_1) & \cdots & N_{n-1,p}(\bar{u}_1) \\ \vdots & \ddots & \vdots \\ N_{1,p}(\bar{u}_{h-1}) & \cdots & N_{n-1,p}(\bar{u}_{h-1}) \end{bmatrix}, \quad (2.16)$$

\mathbf{R} ein Vektor mit $(n - 1)$ Elementen ist

$$\mathbf{R} = \begin{bmatrix} \sum_{k=1}^{h-1} N_{1,p}(\bar{u}_k) \mathbf{Q}_k \\ \vdots \\ \sum_{k=1}^{h-1} N_{n-1,p}(\bar{u}_k) \mathbf{Q}_k \end{bmatrix} \quad (2.17)$$

und \mathbf{Q}_k wie folgt gebildet wird.

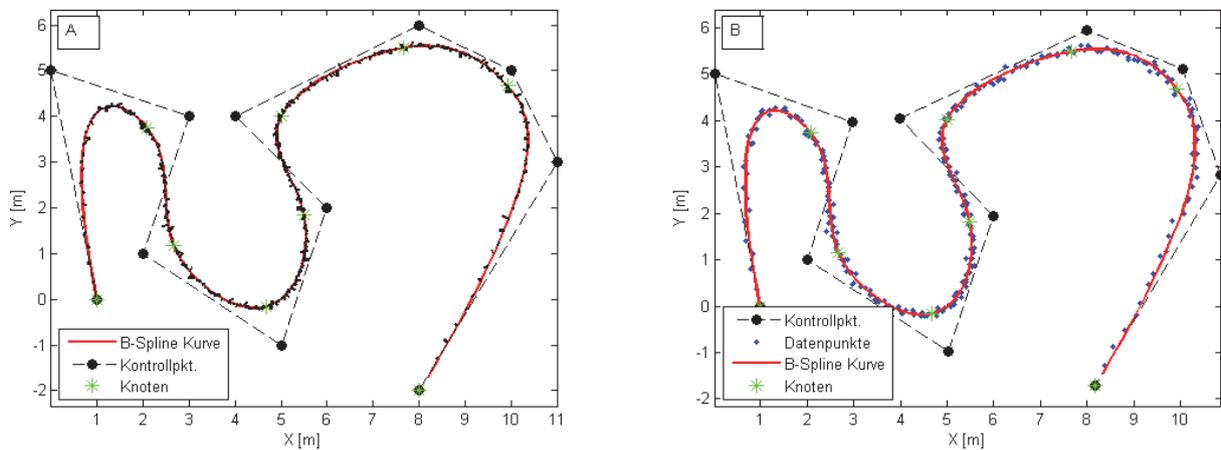
$$\mathbf{Q}_k = \mathbf{D}_k - N_{0,p}(\bar{u}_k) \mathbf{D}_0 - N_{n,p}(\bar{u}_k) \mathbf{D}_h \quad \text{für } k = 1, \dots, h - 1 \quad (2.18)$$

Die Kontrollpunkte \mathbf{P}_i können schlussendlich über

$$\mathbf{P} = (\mathbf{N}^T \mathbf{N})^{-1} \mathbf{R} \quad (2.19)$$

berechnet werden. Im hier gezeigten Formelapparat werden alle Beobachtungen als gleichwertig angesehen, wodurch die Gewichtsmatrix der Beobachtungen im GMM zur Einheitsmatrix wird. Die gemeinsame Schätzung von Kontrollpunkten und Knotenvektor führt zu einem hochgradig, nichtlinearen Gleichungssystem, welches gegenwärtig nur unzureichend bzw. nicht zufriedenstellend gelöst werden kann. Somit ist die Ermittlung von Kontrollpunkten von der Anzahl und der Verteilung der zuvor bestimmten Knoten abhängig. Dadurch kann die volle Flexibilität, welche die B-Spline Kurve mit sich bringt, nicht ausgenutzt werden (Piegl & Tiller, 1997, S.410 ff.).

Im Allgemeinen muss davon ausgegangen werden, dass die vorhandenen Datenpunkte aus einer Messung stammen und folglich mit einem Rauschen belegt sind. Es ist zu untersuchen, wie sehr verrauschte Messdaten als auch die Anzahl an Daten das Schätzergebn bei einem kleinsten Quadrate Ausgleich beeinflussen. Anhand einer B-Spline-Kurve, wobei die Anzahl an Kontrollpunkte, der Grad der Kurve und der Knotenvektor als bekannt vorausgesetzt werden, kann die Approximationsgüte des kleinsten Quadrate Ausgleichs bei unterschiedlich stark verrauschten Beobachtungsdaten ermittelt werden. Dazu wird die bekannte B-Spline-Kurve aus Abb. 2.4 A mit dem äquidistanten Knotenvektor benutzt. Ausgehend von dieser Kurve werden in unterschiedlichen äquidistanten Abständen (sowohl im Parameterraum als auch im euklidischen Raum) Punkte generiert und mit einem künstlichen normalverteilten Rauschen versehen (Abb. 2.9 A). Durch diese Vorgehensweise kann auf die Berechnung der Parameterwerte, wie in Kapitel 2.5.1, verzichtet werden und im weiteren Berechnungsschritt können diese ebenfalls als bekannt angesehen werden.



**Abb. 2.9: A) Ursprüngliche B-Spline Kurve C_2 mit 284 verrauschten Datenpunkten ($\sigma = 0.05m$)
B) aus verrauschten Datenpunkten gewonnene approximierende B-Spline Kurve C_1**

Anschließend werden die Kontrollpunkte im linearen Ausgleichsproblem aus den verrauschten Datenpunkten (mit bekannten Parameterwerten) ermittelt und die approximierende Kurve C_1 (Abb. 2.9 B) mit der ursprünglichen Kurve C_2 verglichen. Mittels der RMSE (root mean square error) Funktion wird ein geeignetes Fehlermaß berechnet. Diese Funktion berücksichtigt bei der Berechnung auch die Anzahl an benutzten Datenpunkten.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (C_1(u_i) - C_2(u_i))^2}{n}} \quad (2.20)$$

Die Anzahl an Punkten, welche für die B-Spline Approximation benutzt werden, ergibt sich durch die Vorgabe eines bestimmten Abstandsmaßes im euklidischen Raum. Die Kurvenlänge der ausgewählten B-Spline Kurve beträgt in diesem Beispiel 28.351m. Wird ein Punktabstand von 0.1m gewählt, so werden 284 Punkte für die Approximation benutzt. Diese

Punkte besitzen zwar im euklidischen Raum einen gleichmäßigen Punktabstand zueinander, nicht jedoch im zugehörigen Parameterraum. Wird hingegen ein äquidistanter Parameterabstand gefordert, indem im Intervall $[0,1]$ 284 gleichabständige Werte ermittelt werden, so sind die Punktabstände im euklidischen Raum nicht mehr gleichmäßig. Dieser Umstand hat jedoch auf die Kurvenapproximation und somit auf den berechneten RMSE Wert keinen großen Einfluss, da stets genügend Punkte relativ gleichmäßig auf der Kurve und zwischen den Knoten verteilt sind.

In Abb. 2.10 ist der berechnete RMSE-Wert gegenüber dem an die Daten angebrachten Datenrauschen ersichtlich. Das Datenrauschen, in der Abbildung als Standardabweichung dargestellt, erstreckt sich dabei im Intervall von null bis einem Meter. Die blauen Marker entsprechen dabei einem Punktabstand von 0.001m, die Roten einem Punktabstand von 0.01m und die Schwarzen einem Punktabstand von 0.1m. Die dazugehörigen Linien zeigen den linearen Trend an.

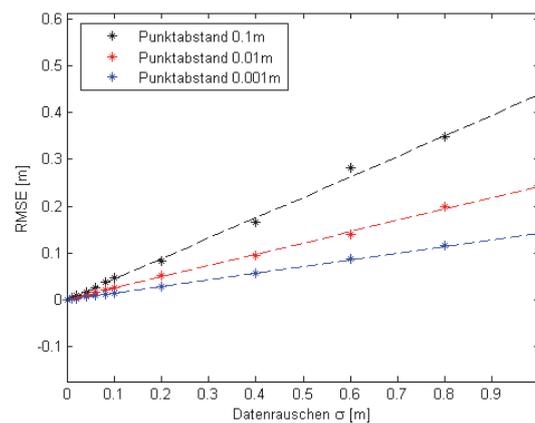


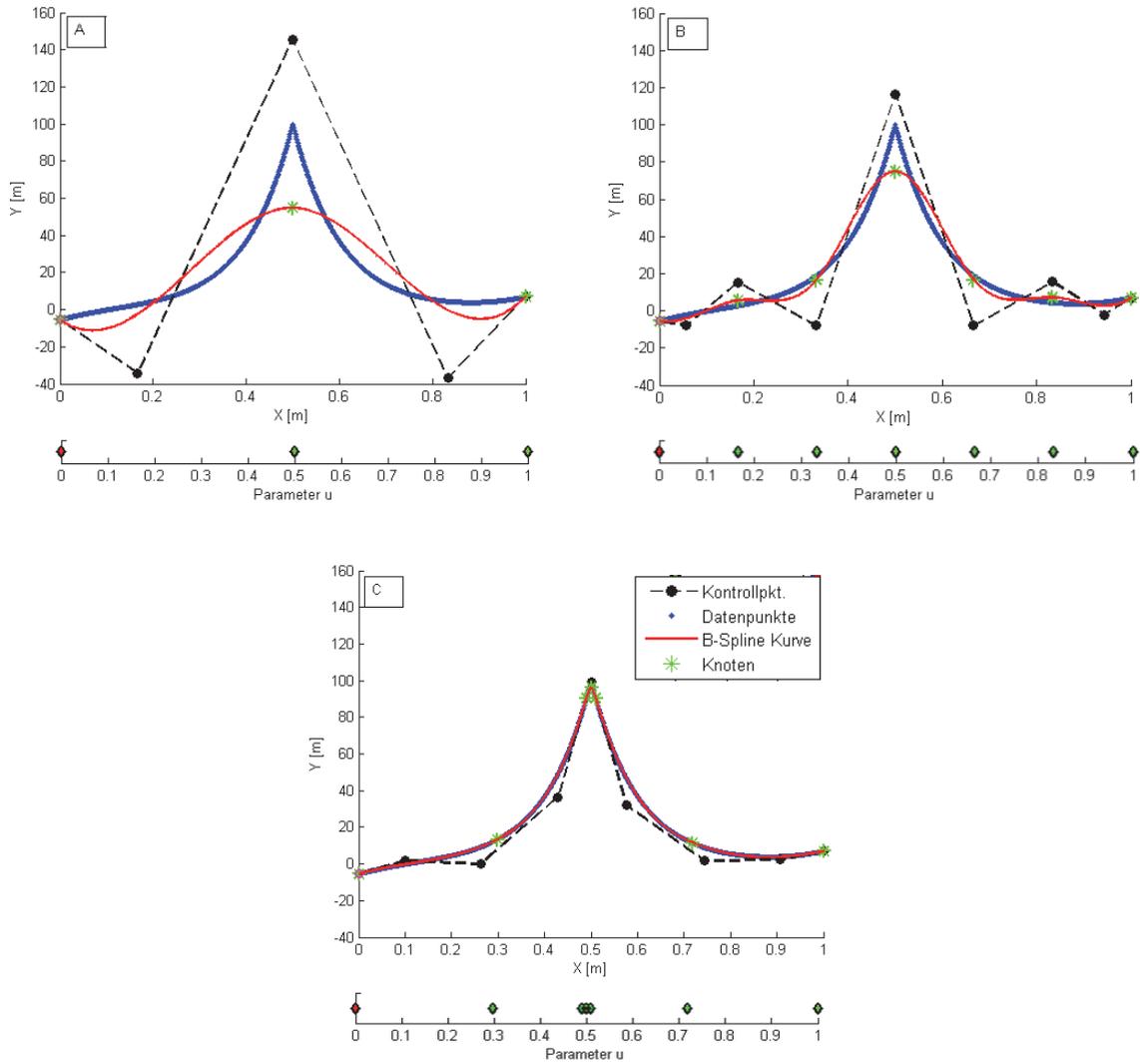
Abb. 2.10: RMSE-Wert in Abhängigkeit des Datenrauschens und des Punktabstandes

Es ist ersichtlich, dass mit zunehmend verrauschten Daten der RMSE-Wert linear ansteigt. Dies ergibt sich zwangsläufig, da die Kontrollpunkte aus stärker verrauschten Daten im Ausgleich und somit die approximierende B-Spline Kurve nicht genauer als aus weniger stark verrauschten Daten geschätzt werden können. Zudem ist erkennbar, dass eine Erhöhung der Punktdichte und somit eine Erhöhung der Datenpunkte die Approximationsgüte erhöht, da die Anzahl an Unbekannten gleich bleibt und sich somit die Redundanz vergrößert, sofern ,wie in diesem Fall, von unkorrelierten Beobachtungen ausgegangen wird.

2.5.4. Motivation der Knotenoptimierung

Neben den Kontrollpunkten, welche im linearen Ausgleichungsproblem zufriedenstellend berechnet werden können, hat vor allem die Anzahl an inneren Knoten auf das Approximationsergebnis eine entscheidende Rolle. Durch die Erhöhung der Knotenanzahl steigt, bei gleichbleibendem Grad p , über die Beziehung (2.3) die Anzahl an Kontrollpunkten. Durch das Hinzufügen zusätzlicher Knoten ist eine signifikante Verbesserung im Ergebnis erkennbar (siehe Abb. 2.11 B). Von großem Einfluss ist dabei auch die Positionierung jedes einzelnen Knotens innerhalb des Parameterraums u (siehe Abb. 2.11 C).

In Abb. 2.11 A ist eine unterbestimmte B-Spline Approximation mit nur einem inneren Knoten erkennbar. Die Knotenverteilung jeder Kurve ist in der darunterliegenden Grafik sichtbar. In Blau ist die zu approximierende Punktmenge dargestellt. Die B-Spline Kurve in rot kann bei Abb. 2.11 A die vorgegebene Punktmenge bei weitem nicht zufriedenstellen approximieren. Durch die Erhöhung der Anzahl an Knoten vergrößert sich auch die Anzahl an Kontrollpunkten und ein besseres Approximationsergebnis wird erzielt (Abb. 2.11 B). Der Grad der B-Spline Kurve ist bei allen Abbildungen gleichbleibend drei. Jeder zusätzliche Knoten erweitert dabei nicht nur das funktionale System der B-Spline Kurve mit einer zusätzlichen Basisfunktion, sondern ändert auch gleichzeitig p (=Grad der B-Spline Kurve) aufsteigende Basisfunktionen, ausgehend von dem neu hinzugefügten Knoten (siehe Formel (2.1)). Der Einfluss der Knotenpositionierung kann anhand der Abb. 2.11 B und C beobachtet werden. Ausgehend von einer gleichverteilten Knotenaufteilung in B kann mit der gleichen Anzahl an Knoten durch eine Verschiebung der einzelnen Knotenpositionen ein deutlich besseres Resultat erreicht werden (siehe Abb. 2.11 C). Es ist eindeutig ersichtlich, dass neben einer korrekten Anzahl an Knoten auch eine passende Knotenverteilung für die Approximationsgenauigkeit entscheidend ist.



**Abb. 2.11: A) Unterbestimmte B-Spline Approximation mit einem inneren Knoten
 B) unzufriedene B-Spline Approximation mit der korrekten Anzahl an inneren Knoten,
 jedoch falscher Knotenverteilung
 C) zufriedenstellende B-Spline Approximation mit korrekter Knotenanzahl und Knotenverteilung**

3. Daten

3.1. Simulierte Datensätze

Um die in Kapitel 4 vorgestellten Methoden vergleichen zu können, welche alle das Ziel verfolgen, die optimalsten Knotenpositionen zu finden, müssen zuerst geeignete Testdaten erzeugt werden. Hierfür wurden die in Abb. 3.1 dargestellten B-Spline Kurven vom Grad 3 benutzt.

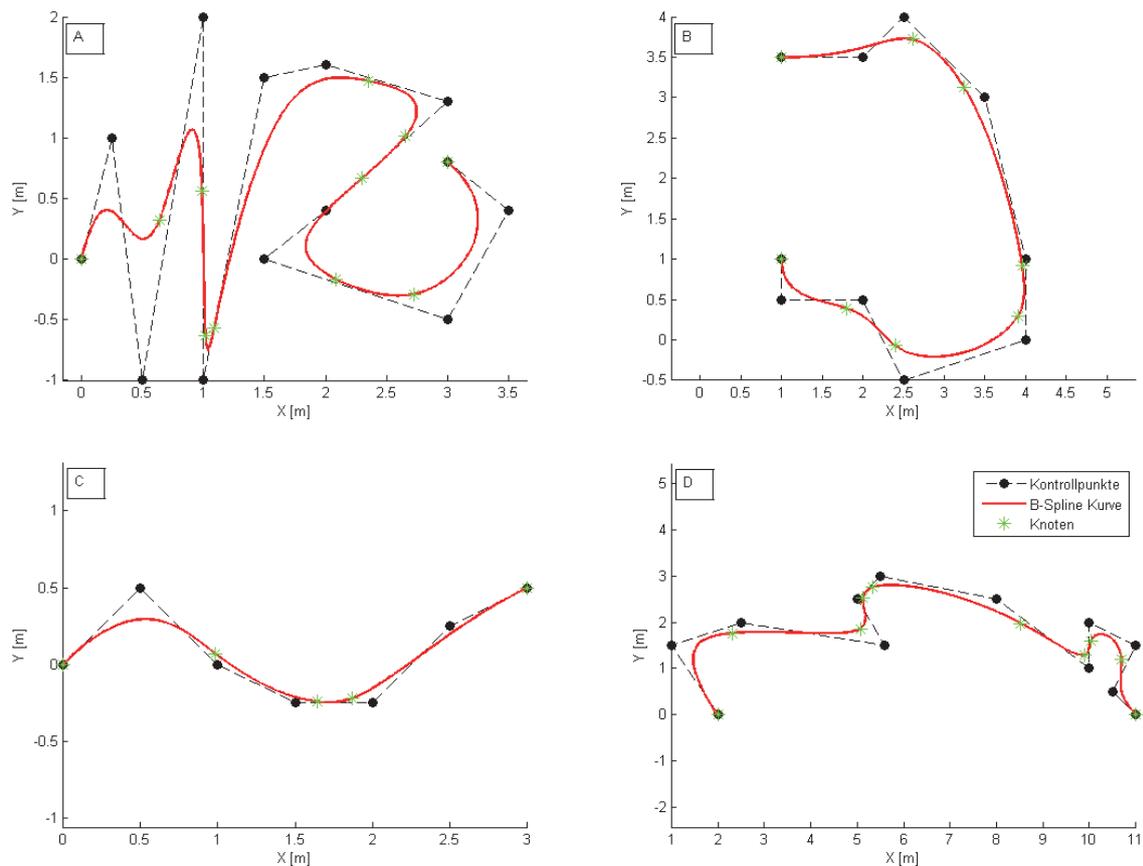


Abb. 3.1: Verwendete B-Spline Kurven für die Evaluierung der unterschiedlichen Methoden in Kapitel 4

Der Datensatz A soll herausfinden, ob die einzelnen Methoden mit komplexen Kurvenstrukturen, welche eng beisammen liegen und unterschiedlich starke Krümmungen aufweisen, zurechtkommen. Die Knotenpositionen wurden dabei so verteilt, dass diese nicht immer in Bereichen mit starken Krümmungen liegen. Eine solche Kurvenform wird unter normalen Umständen nicht zu modellieren sein. Trotzdem ist es wichtig das Verhalten der einzelnen Methoden bezüglich solcher extremen Kurven zu kennen. Der Datensatz B hingegen weist im Gegensatz zu A nur schwache Krümmungen auf. Die Krümmungsänderungen innerhalb der inneren Knoten fallen dabei gering aus. Mit dem einfachen Datensatz C soll gezeigt werden, ob eine zufriedenstellende Approximation der

Datenpunkte auch dann gelingen kann, wenn die ermittelten Knotenpositionen deutlich von den originalen Knoten abweichen. Datensatz B und C stellen dabei Kurven dar, welche durchaus in der Praxis zu modellieren sind. Beim letzten Datensatz D wechseln sich Kurvenbereiche mit einer starken Krümmung und einer sehr geringen Krümmung regelmäßig ab.

In Tab. 3.1 sind die Kontrollpunkte P und der Knotenvektor U der einzelnen Kurven ersichtlich. Da es sich um eingespannte B-Spline Kurven handelt, werden beim Knotenvektor die ersten und letzten p Knotenwerte nicht angegeben.

Tab. 3.1: Kontrollpunkte P und Knotenvektor U der einzelnen Datensätze aus Abb. 3.1

Abb. 3.1 A	P	X[m]	0	0,25	0,5	1	1	1,5	2	3	2	1,5	3	3,5	3
		Y[m]	0	1	-1	2	-1	1,5	1,6	1,3	0,4	0	-0,5	0,4	0,8
	U		0	0,1	0,16	0,18	0,2	0,5	0,65	0,7	0,85	0,9	1		
Abb. 3.1 B	P	X[m]	1	1	2	2,5	4	4	3,5	2,5	2	1			
		Y[m]	1	0,5	0,5	-0,5	0	1	3	4	3,5	3,5			
	U		0	0,1	0,2	0,45	0,55	0,75	0,85	1					
Abb. 3.1 C	P	X[m]	0	0,5	1	1,5	2	2,5	3						
		Y[m]	0	0,5	0	-0,25	-0,25	0,25	0,5						
	U		0	0,3	0,6	0,7	1								
Abb. 3.1 D	P	X[m]	2	1	2,5	5,6	5	5,5	8	10	10	11	10,5	11	
		Y[m]	0	1,5	2	1,5	2,5	3	2,5	1	2	1,5	0,5	0	
	U		0	0,1	0,25	0,35	0,4	0,6	0,7	0,75	0,9	1			
Grad p der B-Spline Kurven bei allen Abbildungen = 3															

Mit Hilfe der vorgestellten B-Spline Kurven konnten vier Datensätze erstellt werden. Dabei bestand die Forderung darin, dass der Punkt zu Punktabstand im euklidischen Raum bei allen Daten 0,015m groß sein soll. Für die einzelnen Kurven ergeben sich somit folgende Anzahlen an Punkte.

Kurve A \triangleq Datensatz 1 \rightarrow 733 Punkte

Kurve B \triangleq Datensatz 2 \rightarrow 604 Punkte

Kurve C \triangleq Datensatz 3 \rightarrow 689 Punkte

Kurve D \triangleq Datensatz 4 \rightarrow 919 Punkte.

Durch diese Vorgehensweise waren auch gleichzeitig die Parameterwerte \bar{u} für alle Punkte bekannt. Jeder einzelne Datensatz wurde anschließend mit einem normalverteilten weißen Rauschen ($\sigma = 0,005m$) verrauscht um somit reale Bedingungen für die unterschiedlichen Methoden zu simulieren.

3.2. Reale Datensätze

Neben den simulierten Daten stehen auch reale Messdaten zur Verfügung. In dieser Arbeit wird hierfür das in Abb. 3.2 dargestellte Objekt benutzt, welches in etwa eine Größe von 0.4x0.4x0.2m aufweist und einer B-Spline Fläche vom Grad 3 entspricht. Eine B-Spline Fläche besitzt im Gegensatz zu einer B-Spline Kurve zwei Parameterrichtungen u und v . Die einzelnen Parameterrichtungen am Messobjekt werden im späteren Kapitel 6 aufgezeigt. Der aus einem Aluminiumblock herausgefräste Gegenstand stellt ein Blatt einer Gurkenpflanze dar. Durch die Produktionsweise des Messobjektes konnte die Form der Fläche mit einer Genauigkeit von $<0,1\text{mm}$ hergestellt werden und ist dementsprechend gemeinsam mit den Referenzpunkten (siehe Abb. 3.2 rote Zahlen) hochgenau bekannt. Mit Hilfe eines terrestrischen Scanners (Leica Nova MS50) wurde die Oberfläche abgetastet. Die Aluminiumoberfläche ist dabei so beschaffen, dass es zu keiner gerichteten Reflexion kommen kann. Um einzelne Scans miteinander vergleichen zu können bzw. diese in ein einheitliches Koordinatensystem zu transformieren um somit die B-Spline Resultate mit der bekannten Fläche zu überprüfen, sind am Rand des Messobjektes fünf fixe Halterungen für Kugelprismen angebracht, deren Referenzpunkte ebenfalls hochgenau bekannt sind.

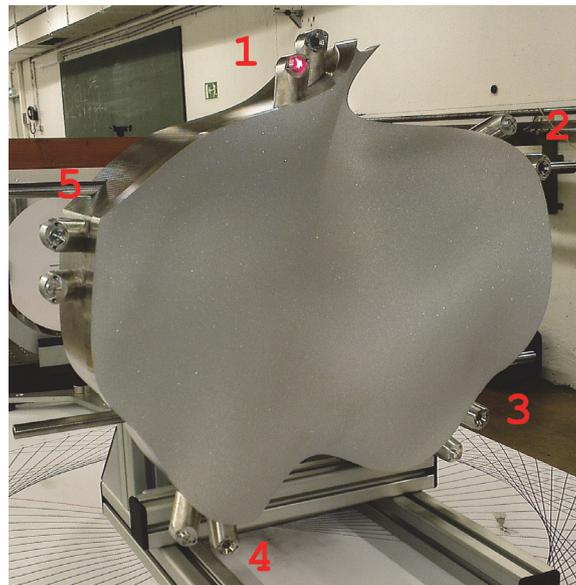


Abb. 3.2: Messobjekt mit seitlich angebrachten Halterungen für Kugelprismen

Insgesamt wurde die Fläche drei Mal gescannt, wobei eine Distanz von etwa vier Metern zum Messobjekt bestand und die Standpunkte leicht versetzt wurden um unterschiedliche Ansichten des Objekts zu bekommen. Die Auflösung des Scans wurde so gewählt, dass Punkte mit einer Maschenweite von 2mm in einem regelmäßigen Raster anfielen. Somit konnten in einem Durchgang zwischen 29804 bis 32928 Punkte am Objekt gemessen werden, wobei die Messungen im 250 Hz Modus stattfanden. Laut dem Datenblatt von Leica ergibt sich dabei für die Distanz ein Messrauschen von $\sigma_s = 0.8\text{mm}$ bei einer Entfernung von

bis zu 50m. Durch die Kombination einer Totalstation und eines Scanners in einem Gerät (MS50) konnten die einzelnen Punktwolken direkt in einem lokalen Koordinatensystem gemessen werden, indem vor dem Scan eine freie Stationierung über fünf Passpunkte erfolgte. Diese waren im Raum rund um das Messobjekt verteilt. Die Identpunkte am Messobjekt, welche mit Hilfe eines 1,5" RRR (Rot-Ring Reflektor) aufgezeigt und bei jedem Scanvorgang mitgemessen wurden, dienten dabei lediglich der Kontrolle der freien Stationierung, wobei eine maximale Abweichung von 0,4mm zwischen Identpunkten unterschiedlicher Scans zu beobachten war. Durch die Kombination aller Scans standen schlussendlich 90750 Punkte zur Verfügung, wobei Punkte mit einem geringen Intensitätswert zuvor entfernt wurden.

Im Allgemeinen sind Pflanzenteile nicht ausreichend durch einfache geometrische Primitive darstellbar. Eine Modellierung mit Freiformflächen ist dabei auf jeden Fall vielversprechender, wobei jedoch der Aufwand steigt. Da in dieser Arbeit die Knotenschätzung von B-Spline Kurven im Mittelpunkt steht und Kurven die Grundlage für B-Spline Flächen sind, werden aus den flächenhaften Scans horizontale und vertikale Scanprofile extrahiert (Abb. 3.3). Die im Profil liegenden Punkte können für die Approximation durch eine B-Spline Kurve über die zugehörige Z- oder X-Koordinaten sortiert werden. Da die Parameterwerte der Datenpunkte unbekannt sind, erfolgte eine Parametrisierung mit Hilfe der *centripetal* Methode.

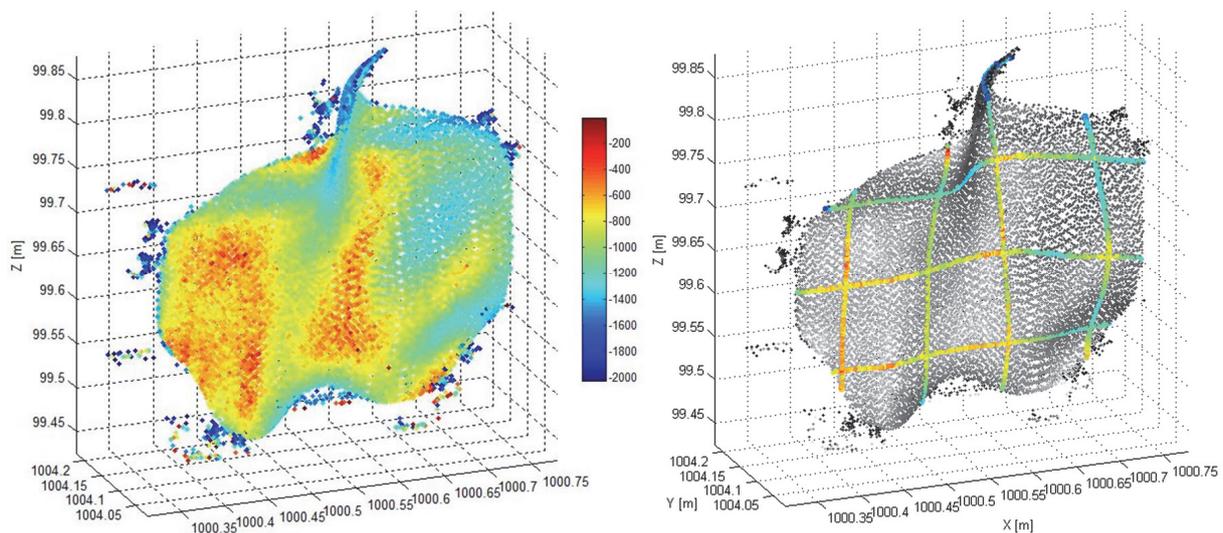


Abb. 3.3: links) Aus drei Scans zusammengefasste Punktwolke (Intensitätswerte farblich dargestellt)
rechts) Aus Punktwolke extrahierte Scanprofile

Für die einzelnen Profile (4 horizontale und 4 vertikale Profile) stehen folgende Anzahlen an Punkte bereit.

Tab. 3.2: Anzahl an extrahierte Punkte für einzelne Profile

XY-Ebene	Punkteanzahl	YZ-Ebene	Punkteanzahl
Z = 99.55 [m]	346	X = 1000.4 [m]	380
Z = 99.65 [m]	456	X = 1000.5 [m]	476
Z = 99.75 [m]	541	X = 1000.6 [m]	402
Z = 99.85 [m]	100	X = 1000.7 [m]	399

4. Knotenschätzung

4.1. B-Spline Approximation mit dominanten Punkten (DOM)

Neben der herkömmlichen Art der Knotengenerierung (siehe Kapitel 2.5.2) kann die Forminformation, welche in den Datenpunkten vorhanden ist, für das Auffinden von passenden Knotenpositionen verwendet werden.

Razdan (1999) beispielsweise benutzt ausgewählte Punkte einer vorgegebenen Kurve, welche durch eine kubische B-Spline Kurve interpoliert werden, um die gegebene Kurve bestmöglich zu approximieren. Die ausgewählten Kurvenpunkte sollen dabei die Gestaltinformation der Kurve innehaben und somit diese bestmöglich repräsentieren. Für die Auswahl geeigneter Interpolationspunkte, welche auch als Knotenpunkte fungieren, wird sowohl die Bogenlängeninformation als auch die Krümmungsinformation verwendet, wobei die benötigte Anzahl in einem vorherigen Schritt abgeschätzt wird. Dieser Algorithmus ist jedoch wegen der fix vorgegebenen Interpolationspunkte sensitiv gegenüber verrauschten Daten und somit für verrauschte Daten nicht geeignet, da die Abweichung zwischen approximierender und vorgegebener Kurve beträchtlich groß sein.

Li et al. (2005) verwendet die lokale Krümmungsinformation der Messdaten um geeignete Knotenpositionen für die B-Spline Kurvenapproximation festzulegen. Die Krümmung eines jeden Datenpunktes wird dabei so gewonnen, dass ein Kreis durch den Punkt und seine beiden Nachbarpunkte gelegt wird. Anschließend wird durch Invertieren des Radius die lokale Krümmung berechnet. Für nicht verrauschte Daten ist diese Art der Krümmungsermittlung zulässig. Wenn jedoch von verrauschten Daten ausgegangen werden muss, so liefert diese Methode unzuverlässige Krümmungsberechnungen, welche in weiterer Folge schwierig zu handhaben sind.

Ein weiterer Lösungsvorschlag von Park & Lee (2007), welcher auf den beiden zuvor genannten Ansätzen aufbaut, transformiert das Problem der B-Spline Kurvenapproximation auf das Auffinden passender dominanter Punkte. Dominante Punkte, welche speziell ausgewählte Datenpunkte sind, sollen dafür sorgen, dass in komplexen (einfachen) Kurvenregionen mehr (weniger) Kurvensegmente generiert werden. Dabei spielt vor allem die Krümmungsinformation und die Bogenlängeninformation, welche in den Datenpunkten vorhanden ist, für das Auffinden von dominanten Punkten eine entscheidende Rolle. Mit einer anschließenden geeigneten Knotenplatzierung, welche abhängig von der Verteilung der dominanten Punkte ist, soll eine zufriedenstellende Kurvenapproximation erreicht werden. Der von Park & Lee entwickelte Algorithmus lässt im Gegensatz zu den Methoden

von Razdan und Li et al. eine *error-bounded* B-Spline Kurvenapproximation zu. Dies bedeutet, dass iterativ die Anzahl an dominanten Punkten und somit auch die Anzahl an Knoten so lange erhöht wird, bis eine gewisse Toleranz eingehalten wird.

Der Algorithmus weist vier grundlegende Arbeitsschritte auf:

- 1) Parametrisierung der Datenpunkte
- 2) Selektion von dominanten Punkten
- 3) Knotengenerierung auf Basis von dominanten Punkten
- 4) Kontrollpunkte im GMM berechnen

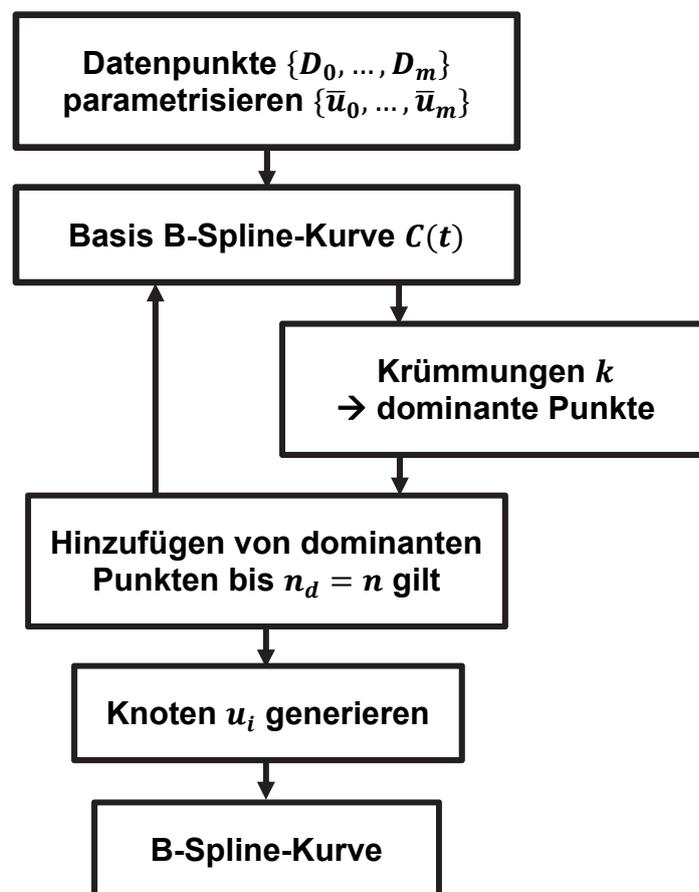


Abb. 4.1: Flussdiagramm der DOM-Methode

Im Folgenden werden die einzelnen Arbeitsschritte näher erklärt. Der vierte Schritt des Algorithmus entspricht jenem aus Kapitel 2.5.3.

4.1.1. Parametrisierung der Datenpunkte

Sollen $h + 1$ Datenpunkte $\{D_0, \dots, D_m\}$ durch eine B-Spline Kurve approximiert werden, so müssen zuerst $h + 1$ Parameterwerte $\{\bar{u}_0, \dots, \bar{u}_m\}$ im Definitionsbereich der Kurve $[a, b]$ gefunden werden. Für die Parametrisierung der Datenpunkte wurden bei Park & Lee die bekannten *chord* und *centripetal* Methoden benutzt (siehe Kapitel 2.5.1).

4.1.2. Selektion von dominanten Punkten

Die Qualität der Kurvenapproximation ist maßgeblich von der Bestimmung von $n + 1$ dominanten Punkten d_j ($j = 0, \dots, n$) abhängig. Dabei ist die Anzahl an dominanten Punkten gleich der Anzahl an Kontrollpunkten, welche im linearen Ausgleichungsschritt bestimmt wird. Der Algorithmus basiert auf dem *Adaptive Refinement* Paradigma und beinhaltet folgende wesentliche Schritte.

1. Setze $n_d = -1$, wobei n_d den höchsten Index der momentanen dominanten Punkte angibt. Es werden so lange dominante Punkte gesucht bis $n_d = n$ gilt.
2. Anhand der Krümmungsinformation der einzelnen Datenpunkte werden potentielle dominante Punkte ermittelt und in einer Liste nach Signifikanz in absteigender Reihenfolge geordnet. Solange die Liste nicht leer ist oder ein Qualitätskriterium noch nicht erfüllt wird und $n_d < n$ ist wird ein potentieller Punkt mit der größten Signifikanz ausgewählt und fortan als dominanter Punkt behandelt. Dabei erhöht sich der Index um eins ($n_d \rightarrow n_d + 1$).
3. Solange $n_d < n$ gilt oder ein Qualitätskriterium nicht erreicht wird, werden zusätzliche dominante Punkte d_j wie folgt bestimmt
 - 3.1. finde jene Region zwischen zwei dominanten Punkten $d_k = D_s$ & $d_{k+1} = D_e$ (wobei $|e-s| > 1$ gelten muss) mit der größten Abweichung
 - 3.2. wähle einen neuen dominanten Punkt d_w ($s < w < e$) und erhöhe den Index $n_d \rightarrow n_d + 1$

zu Punkt 2) – Auffinden von potentiellen dominanten Punkten

Ausgehend von einer Basis-B-Spline Kurve $C(u)$, welche die Datenmenge im ersten Iterationsschritt nur näherungsweise approximiert, wird die Krümmung k_i in jedem Datenpunkt D_i berechnet. Eine geeignete Basis-B-Spline Kurve kann mit der *Bisection*-Methode gefunden werden, welche in der Berechnungszeit schneller als eine inkrementelle Methode ist (Park, 2004). Dabei wird ausgehend von einem oberen und unteren Limit an Kontrollpunkten iterativ die Anzahl an Kontrollpunkten so angepasst, dass eine vorgegebene Toleranz δ zwischen Basis-B-Spline Kurve und Datenpunkten eingehalten wird. Das untere Limit besteht dabei zu Beginn aus $p + 1$ Kontrollpunkten wodurch eine Bézier-Kurve entsteht

und das obere Limit ist durch $h + 1$ Kontrollpunkte (= Anzahl an Datenpunkten) vorgegeben (entspricht Interpolation). Dieses Intervall wird nun bei jedem Iterationsschritt halbiert und eine approximierende B-Spline Kurve mit der entsprechenden Anzahl an Kontrollpunkten generiert (Abb. 4.2). Ist dabei die Kurve innerhalb der vorgegebenen Toleranz, so approximiert diese die Daten zufriedenstellend und das obere Limit kann durch das mittlere Intervall ersetzt werden. Andernfalls wird das untere Limit durch das mittlere Intervall ersetzt, da die Kurve die Daten nicht genau genug approximiert. Dieser binäre Suchprozess wird so lange fortgesetzt bis der Abstand zwischen den beiden Limits gleich eins ist. Die Toleranz δ muss dabei so gewählt werden, dass die wesentlichsten Merkmale der Kurve noch zufriedenstellend dargestellt werden, jedoch das Rauschen nicht mitmodelliert wird.

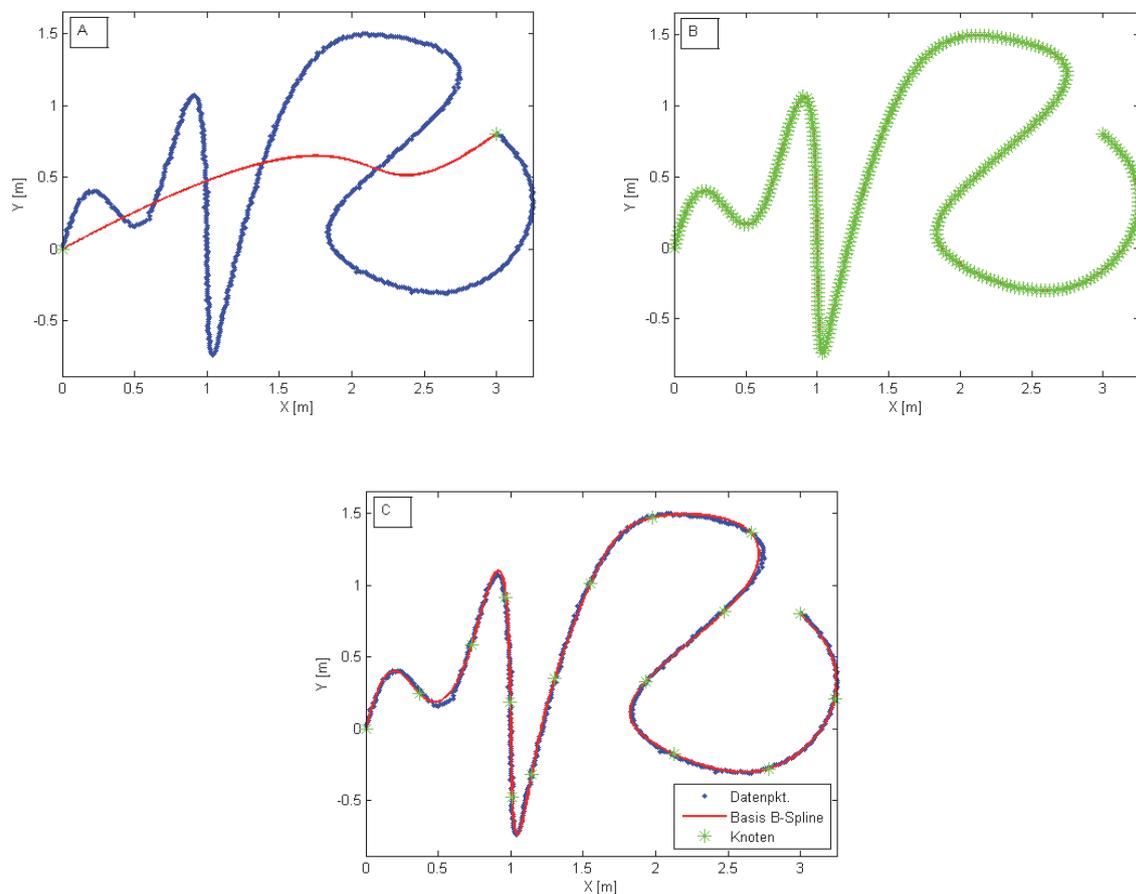


Abb. 4.2: Datenapproximation durch Bisection-Methode mit eingezeichneten Knotenpunkten (grüne Marker); alle Datenpunkte wurden mit einem Rauschen belegt ($\sigma^2 = 0,005m^2$)
 A) unteres Limit an Kontrollpunkten (\cong Bézier-Kurve) B) oberes Limit an Kontrollpunkten (Interpol.)
 C) Basis B-Spline Kurve welche Toleranz $\delta = 0,05m$ nicht überschreitet

Als einzuhaltende Toleranz bei der Approximation kann beispielsweise die maximale Abweichung zwischen den Datenpunkten \mathbf{D} und der approximierenden Kurve \mathbf{C} gewählt werden.

$$\text{maximale Abweichung} = \max_{0 \leq k \leq h} \left(\min_{0 \leq u \leq 1} |\mathbf{D}_k - \mathbf{C}(u)| \right) \leq \delta \quad (4.1)$$

$$\text{mittlere Abweichung} = \frac{1}{h} \sum_{k=0}^h \left(\min_{0 \leq u \leq 1} |\mathbf{D}_k - \mathbf{C}(u)| \right) \leq \delta \quad (4.2)$$

Dabei wird für jeden Datenpunkt der minimale Abstand zur bestehenden Kurve berechnet. Dieser Abstand entspricht genähert dem Normalabstand zwischen Datenpunkt und Kurve. Anschließend wird von allen Datenpunkten die größte bestehende Abweichung zur Kurve mit dem Toleranzwert verglichen. Ist die größte Abweichung kleiner als die vorgegebene Toleranz, so ist gesichert, dass die berechnete B-Spline Kurve keine größere Abweichung zu den Datenpunkten als die Toleranz aufweist. Die mittlere Abweichung gibt gemeinsam mit der maximalen Abweichung Auskunft darüber, wie gut die Approximation ausgefallen ist.

Die Basis-B-Spline Kurve hat also als Ziel die verrauschten Eingangsdaten so zu approximieren, dass das Rauschen in den Messdaten keinen Einfluss auf die anschließende Krümmungsberechnung hat.

Die Krümmung kann nun bei einer vorgegebenen Kurve leicht berechnet werden,

$$k_i = \|\dot{\mathbf{C}}(\bar{u}_i) \times \ddot{\mathbf{C}}(\bar{u}_i)\| / \|\dot{\mathbf{C}}(\bar{u}_i)\|^3 \quad (4.3)$$

wobei $\dot{\mathbf{C}}(\bar{u}_i)$ und $\ddot{\mathbf{C}}(\bar{u}_i)$ die erste und zweite Ableitung der Kurve (siehe Kapitel 2.3) am Parameterwert \bar{u}_i darstellen. Es besteht auch die Möglichkeit die Krümmung für jeden Datenpunkt über eine lokale Methode zu berechnen, indem beispielsweise ein Kreis durch einen Punkt und seinen beiden Nachbarpunkten gelegt wird. Anschließend kann die Krümmung durch Invertieren des Kreisradius bestimmt werden (LI et al., 2005; siehe Kapitel 4.1). Diese Methode der Krümmungsberechnung ist jedoch, wie bereits anfangs erwähnt, für verrauschte Daten nicht zufriedenstellend, da das Rauschen die lokale Krümmungsberechnung stark beeinflusst. Durch die Verwendung einer Basis-B-Spline Kurve kann dieses Problem umgangen werden.

Punkte \mathbf{D}_i in denen sich die Krümmung k_i zu den Nachbarpunkten wie folgt verhält $k_{i-1} < k_i > k_{i+1}$ werden als potentielle dominante Punkte gespeichert und in eine Liste aufgenommen. Diese Punkte entsprechen den Maximalwerten in einem Krümmungsplot (siehe Abb. 4.3 in Kapitel 4.1.5). Punkte mit einer hohen Krümmung haben dabei eine hohe Signifikanz und werden bevorzugt als dominante Punkte ausgewählt. Es werden jedoch nur Punkte berücksichtigt, welche einen Krümmungswert von $k_{min} = k_{mean}/4$ überschreiten wobei k_{mean} der gemittelte Krümmungswert über alle Datenpunkte ist. Nähere Informationen zum Krümmungsplot in Abb. 4.3 befinden sich in Kapitel 4.1.5.

Ausgehend von den bisher gefundenen dominanten Punkten wird ein Knotenvektor mit der Formel (4.7) generiert und die Kontrollpunkte der Kurve, und somit die B-Spline Kurve selbst, durch einen linearen Ausgleich (siehe Kapitel 2.5.3) bestimmt. Dabei wird auch der Krümmungswert für jeden Datenpunkt aktualisiert. Wurde die vorgegebene Anzahl an dominanten Punkten bzw. ein Qualitätskriterium noch nicht erreicht, so werden zusätzliche dominante Punkte in einem iterativen Prozess gesucht (siehe Punkt 3).

zu Punkt 3) – hinzufügen von dominanten Punkten

Werden weitere dominante Punkte hinzugefügt, so wird zuerst jener Abschnitt zwischen zwei dominanten Punkten d_s d_e gesucht, in welchem ein Datenpunkt D_i die größte Abweichung zur momentanen Kurve $C(\bar{u}_i)$ hat. Die Abweichung wird dabei über $\|C(\bar{u}_i) - D_i\|$ definiert. Sind zumindest drei Punkte im jeweiligen Abschnitt vorhanden, so kann mit Hilfe eines Formfaktors ein neuer dominanter Punkt platziert werden. Der Formfaktor $\lambda_{s,e}$, welcher die Formkomplexität eines Segmentes angibt, wird wie folgt berechnet.

$$\lambda_{s,e} = r \frac{K_{s,e}}{K_{0,h}} + (1-r) \frac{L_{s,e}}{L_{0,h}} \quad 0 \leq r \leq 1 \quad (4.4)$$

$K_{s,e}$ gibt dabei die totale Krümmung und $L_{s,e}$ die Bogenlänge des jeweiligen Abschnittes wieder (siehe Formel(4.5)). Der Parameter r steuert dabei den Einfluss der Krümmung bzw. der Bogenlänge auf die Formfaktorberechnung. Wird $r = 0$ gesetzt, so entspricht dies einer Gleichverteilung der Bogenlänge wohingegen $r = 1$ einer Gleichverteilung der Krümmung gleichkommt. Die Krümmung wird mit Hilfe der Trapezregel und Formel (4.3), die Bogenlänge näherungsweise mit Hilfe der Datenpunkte berechnet.

$$K_{s,e} = \sum_{i=s}^{e-1} (|k_i| + |k_{i+1}|)(\bar{u}_{i+1} - \bar{u}_i)/2 \quad L_{s,e} = \sum_{i=s}^{e-1} \|D_{i+1} - D_i\| \quad (4.5)$$

Die Nenner in (4.4) geben die totale Krümmung bzw. die Bogenlänge, berechnet über alle gegebenen Datenpunkte, wieder.

Unter Ausnutzung der Formel (4.4) kann nun ein neuer dominanter Punkt d_w im jeweiligen Abschnitt gefunden werden, welcher die Differenz zweier Formfaktoren der Segmente $S_{s,w}$ & $S_{w,e}$ minimiert.

$$\min_w |\lambda_{s,w} - \lambda_{w,e}| \quad s < w < e \quad (4.6)$$

Anschließend wird wiederum ein neuer Knotenvektor, ausgehend von den dominanten Punkten, mit Formel (4.7) erzeugt und die Kontrollpunkte für eine approximierende Kurve im linearen Ausgleichsproblem berechnet. Dabei ändern sich durch die Hinzunahme eines dominanten Punktes bzw. eines Kontrollpunktes höchstens p innere Knoten. Die restlichen Knoten bleiben unverändert. Dadurch wird erreicht, dass sich die approximierende B-Spline Kurve nur in einem begrenzten Bereich verändert.

4.1.3. Knotengenerierung auf Basis von dominanten Punkten

Wurden n dominante Punkte d_j ($j = 0, \dots, n$) ausgewählt, so können die Parameter der inneren Knoten u_i durch Mittelbildung von p Parameterwerten der dominanten Punkte wie folgt berechnet werden.

$$u_{p+i} = \frac{1}{p} \sum_{j=i}^{i+p-1} \bar{u}_{f(j)} \quad (i = 1, \dots, n - p) \quad (4.7)$$

wobei p ... Grad d. B – Spline, n ... Anzahl an dominanten Punkten

Dabei ist \bar{u}_k der Parameterwert des Punktes D_k und $f(j)$ eine simple Funktion, die den Index des Punktes D_k , welcher mit dem dominanten Punkt d_j korrespondiert, zurückgibt. Werden die dominanten Punkte so ausgewählt, dass sie die Kurvengestalt der Datenpunkte grob wiedergeben, so findet sich dies auch im aufgestellten Knotenvektor wieder, da dieser aus den dominanten Punkten abgeleitet wird.

Die Eigenschaften der Knotengenerierung mit Hilfe dominanter Punkte lassen sich wie folgt zusammenfassen:

- nicht singuläre Systemmatrix während Least Square Minimierung
- gute Kurvenapproximation auch wenn $|m-n|$ klein ist
- wenn $m=n$ so stimmt Knotenvektor mit jenem für Interpolation überein
- wird ein neuer dominanter Punkt hinzugefügt so ändert sich der Knotenvektor höchstens an p innere Knoten

4.1.4. Krümmungsberechnung anhand einer Basis B-Spline Kurve

Der beschriebene Algorithmus findet für eine geordnete Messdatenreihe eine B-Spline-Kurve vom Grad p , welche die Forminformation der Daten bei der Knotengenerierung berücksichtigt. Als Übergabeparameter wird neben den Datenpunkten sowohl der Grad p der B-Spline-Kurve als auch die Anzahl an dominanten Punkten n (= Anzahl an Kontrollpunkten) bzw. die Anzahl an Knoten erwartet. Soll eine vorgegebene Fehlertoleranz zwischen der approximierenden Kurve und den Datenpunkten eingehalten werden so ist diese Methode auch für eine *error bounded* B-Spline Kurvenapproximation geeignet. Dabei werden iterativ so lange neue dominante Punkte ermittelt bis die Fehlertoleranz eingehalten wird und eine zufriedenstellende Approximation gewährleistet ist.

Dominante Punkte und schließlich davon abgeleitete Knotenpunkte sollen dabei vorwiegend an Stellen mit starker Krümmung platziert werden, um in Regionen mit komplexen Kurvensegmenten eine bessere B-Spline Kurvenapproximation zu gewährleisten. Das Ergebnis und die Auswahl dominanter Punkte ist dabei im Wesentlichen von der

Krümmungsberechnung und somit von der Basis B-Spline Kurve, welche am Anfang von Punkt 2 (Kapitel 4.1.2) für die erste Krümmungsberechnung benötigt wird, abhängig. Ist diese anfängliche Approximation der Datenpunkte nicht zufriedenstellend, so werden dominante Punkte an Orte platziert, welche keinen komplexen Kurvenverlauf aufweisen bzw. es werden zu viele dominante Punkte bestimmt. In weiterer Folge wird dadurch die Knotengenerierung negativ beeinflusst.

Es stellte sich heraus, dass die Approximation der Datenpunkte durch die Basis B-Spline Kurve, stark von der Parametrisierung der Datenpunkte abhängig ist. Sind die Parameterwerte der Datenpunkte bekannt, so ist der Kurvenverlauf der Basis B-Spline glatt. Dadurch ist auch die berechnete Krümmungskurve glatt und Krümmungsmaxima werden einfach und zufriedenstellen gefunden. Werden jedoch die Parameterwerte der Datenpunkte durch die in Kapitel 2.5.1 beschriebenen Methoden ermittelt, so kann es passieren, dass die Basis B-Spline Kurve, in Bereichen wo eigentlich ein glatter Kurvenverlauf durch die Datenpunkte vorgegeben wird, ungewollt leicht ausschwingt und somit mehrere Krümmungsmaxima in kurzen aufeinanderfolgenden Datenpunkten hat.

In Abb. 4.3 wurde die Krümmung für jeden Datenpunkt der Kurve aus Abb. 4.2 C in einer Grafik dargestellt, wobei in der Abbildung A bei der Berechnung der Basis B-Spline Kurve die Parameter der Datenpunkte bekannt waren. In der Abbildung B wurden die Parameterwerte mit Hilfe der *centripetal* Methode ermittelt. Für die anfängliche Approximation der Punktdaten durch die Basis B-Spline Kurve wurde eine Toleranz von $\delta = 0.05m$ gewählt. Dies entspricht der zehnfachen Standardabweichung an Rauschen, welches an die Daten angebracht wurde. Damit kann in den meisten Fällen sichergestellt werden, dass eine ausreichend genaue Approximation der Daten erfolgt. Eine Herabsetzung dieser Grenze sollte möglichst nicht erfolgen, da besonders bei Kurven mit starken Krümmungen ansonsten unverhältnismäßig viele Knoten für die Basis B-Spline Kurve benötigt werden, was wiederum zu einem ungewollten Ausschwingen der Kurve führen kann. Wird die Toleranz zu hoch angesetzt, so kann es passieren, dass wichtige Krümmungsmaxima nicht modelliert werden.

Die Krümmungsmaxima, welche die Schranke $k_{min} = k_{mean}/4$ überschreiten, sind mit einem roten Marker gekennzeichnet (Erklärung dazu siehe Kapitel 4.1.2. Punkt 2). Wie zu sehen ist, wurden in der Grafik A deutlich weniger Maxima gefunden als in der Grafik B, da der Krümmungsverlauf der Basis B-Spline Kurve glatter ist. Sind die Parameter der Datenpunkte nicht bekannt und müssen diese erst ermittelt werden, so werden eindeutig mehr Maxima gefunden.

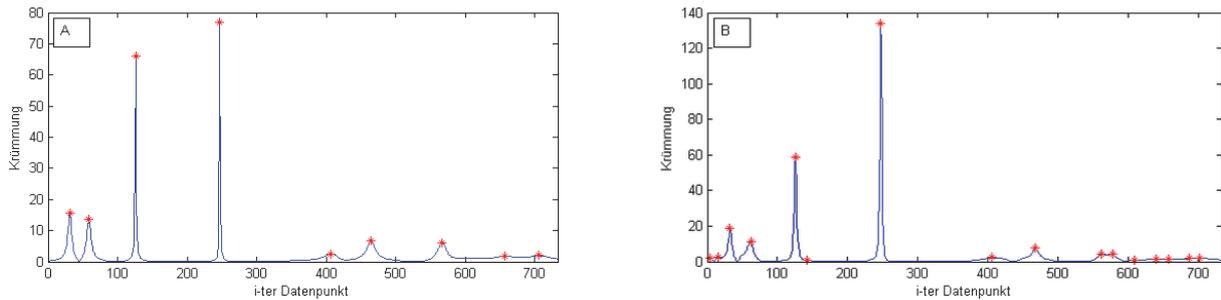


Abb. 4.3: Krümmungsplot der Basis B-Spline Kurve aus Abb. 4.2 mit eingezeichneten Krümmungsmaxima
A) Parameterwerte u der Datenpunkte bekannt
B) Parameterwerte u der Datenpunkte unbekannt

Dies liegt daran, dass für die anfängliche Approximation der Punktdaten deutlich mehr Kontrollpunkte bzw. Knoten benötigt wurden (Abb. 4.4). Dies wiederum führt dazu, dass die Basis B-Spline Kurve, in Kurvenbereichen wo eine mehr oder weniger konstante Krümmung der Daten vorliegt, zusätzlich leicht zu schwingen beginnt. Wird anschließend die Krümmung aus dieser Kurve berechnet und fällt die Schranke k_{min} zu gering aus, so werden bei Kurvensegmenten mit eigentlich konstanter Krümmung mehrere Krümmungsmaxima gefunden. Zu sehen ist dies in Abb. 4.3 B, wo ab dem Datenpunkt 550 bis zum letzten Datenpunkt, im Gegensatz zur Abb. 4.3 A, mehr dominante Punkte gesetzt werden. Auch zu Beginn wurden zwei dominante Punkte gefunden, welche jedoch an dieser Stelle unnötig sind und in weiterer Folge die Knotenberechnung negativ beeinflussen. Des Weiteren ist zu bemerken, dass die Schranke k_{min} einen wesentlichen Einfluss auf die Auswahl von dominanten Punkten hat. Liegt eine Kurve mit einzelnen starken Krümmungsmaxima vor, so fällt die Schranke sehr klein aus, wodurch auch unbedeutende Maxima als dominante Punkte ausgewählt werden. In Abb. 4.3 lag der Wert für k_{min} bei beiden Grafiken bei etwa 0.4.

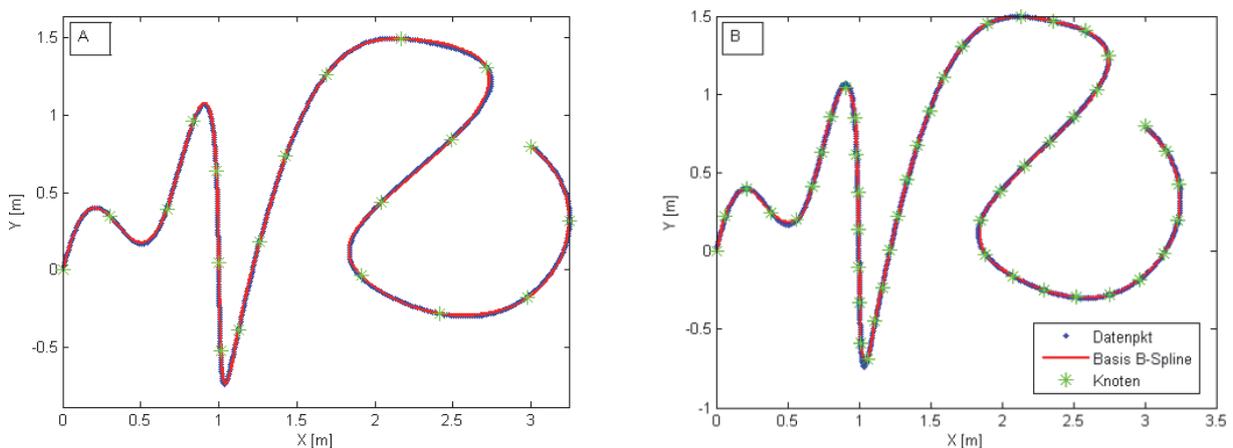


Abb. 4.4: Approximation der Datenpunkte durch Basis B-Spline Kurve (Toleranz $\delta = 0,05m$)
A) Parameterwerte u der Datenpunkte bekannt
B) Parameterwerte u der Datenpunkte unbekannt

In Abb. 4.5 sind die gefundenen Krümmungsmaxima aus Abb. 4.3, welche anschließend als dominante Punkte aufgefasst werden, in den Datenpunkten eingezeichnet. In der Abbildung A passen diese sehr gut mit dem Kurvenverlauf überein. An Stellen mit starker Krümmung wurde jeweils ein dominanter Punkt gesetzt. In der Abbildung B hingegen wurden dominante Punkte an Stellen platziert, welche keine nennenswerte Krümmung aufweisen. Beispielsweise wurden bereits zu Beginn der Kurve zwei weitere dominante Punkte an Orte mit geringer Krümmung gefunden. Die weiteren Orte wurden rot markiert.

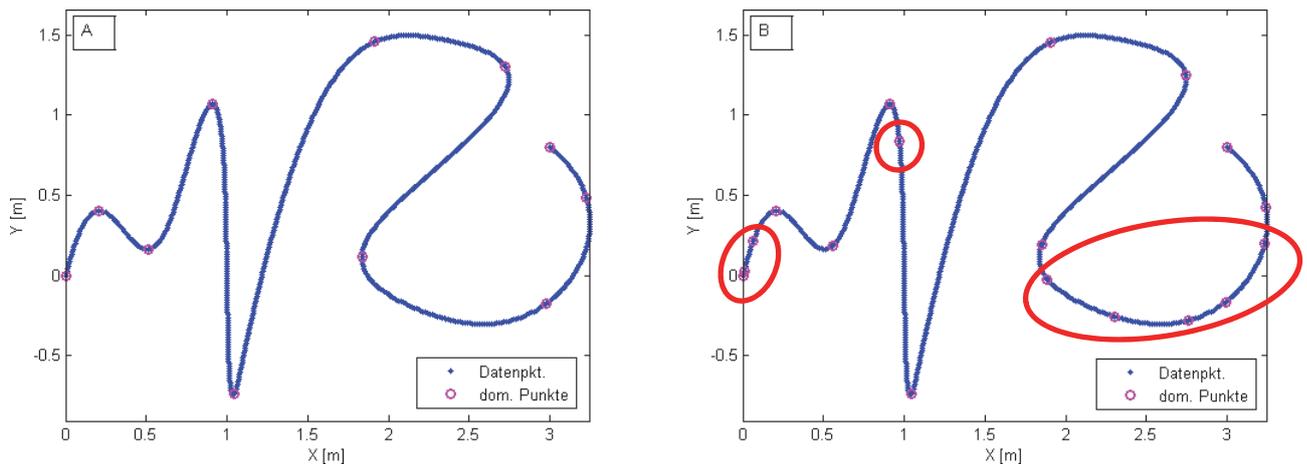


Abb. 4.5: Aus Krümmungsplot der Basis B-Spline Kurve ermittelte dominante Punkte (=Krümmungsmaxima)
A) Parameterwerte u der Datenpunkte bekannt
B) Parameterwerte u der Datenpunkte unbekannt

Werden alle gefundenen Krümmungsmaxima aus dem Krümmungsplot als dominante Punkte verwendet und wird eine vorgegebene Anzahl an dominanten Punkten bzw. ein Qualitätskriterium noch nicht erreicht, so müssen weitere dominante Punkte aufgefunden werden. Die dabei benutzte Vorgehensweise wurde bereits in Kapitel 4.1.2 beschrieben.

Dabei beeinflusst der Parameter r in Formel (4.4) die Auswahl bzw. die Platzierung weiterer dominanter Punkte. Dies wird in den nächsten zwei Grafiken (Abb. 4.6 und Abb. 4.7) aufgezeigt. Zur besseren Übersicht wird hier die besagte Formel (4.4) nochmals angeschrieben.

$$\lambda_{s,e} = r \frac{K_{s,e}}{K_{0,h}} + (1 - r) \frac{L_{s,e}}{L_{0,h}} \quad 0 \leq r \leq 1$$

In der Abb. 4.6 wurde der Parameter r gleich null gesetzt. Dadurch wird der neue dominante Punkt an jene Stelle gesetzt, welche die Bogenlänge zwischen den bestehenden dominanten Punkten halbiert. Nachdem bereits alle Krümmungsmaxima aus dem Krümmungsplot für eine Knotengenerierung benutzt wurden und die Daten noch nicht zufriedenstellend approximiert sind (Abb. 4.6 A), wird nun zwischen dem vorletzten und letzten dominanten Punkt, wo auch die größte Abweichung zu den Datenpunkten zu finden ist, ein neuer dominanter Punkt gesucht. Dazu wird für jeden Datenpunkt innerhalb des Kurvensegmentes,

aufgespannt durch die beiden dominanten Punkte, der Formfaktor λ berechnet. An jener Stelle wo sich die beiden Formfaktoren $\lambda_{s,w}$ und $\lambda_{w,e}$ schneiden wird ein neuer dominanter platziert (Abb. 4.6 C). Dieser wird bei der nächsten Knotenberechnung berücksichtigt, wodurch die Anzahl an inneren Knoten um eins steigt. Es ändern sich dabei jedoch, wie bereits in Kapitel 4.1.3 erwähnt, höchstens p innere Knoten. Die Approximation der Daten ist im nächsten Iterationsschritt an dieser Stelle deutlich besser (Abb. 4.6 B).

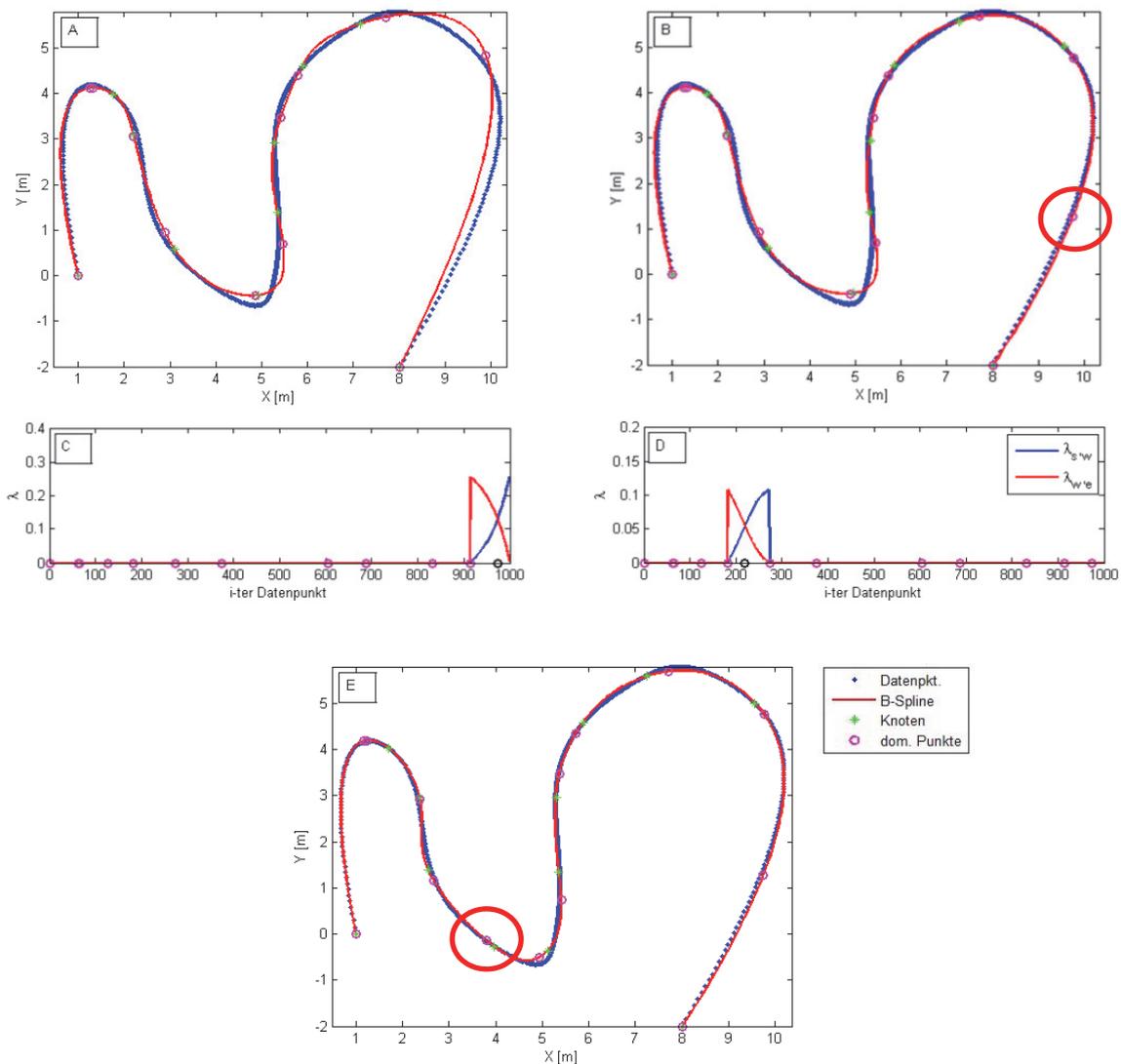


Abb. 4.6: Ermittlung von neuen dominanten Punkten mit $r = 0$ [siehe Formel (4.4)]
A) Ausgangskurve um neuen dominanten Punkt zu berechnen
C) geplotteter Formfaktor λ für i-ten Datenpunkt
und eingezeichneter neuer dominanter Punkt (schwarzer Kreis)
B) B-Spline Kurve mit neuem dominantem Punkt
D) neuerliche Ermittlung eines neuen dominanten Punktes
E) B-Spline Kurve mit neuem dominantem Punkt, welche vorgegebene Toleranz einhält

In der zweiten Grafik (Abb. 4.7) wurde der Parameter r gleich eins gesetzt. Dadurch wird in der Formel (4.4) lediglich die Krümmung der Kurve betrachtet und der neue dominante Punkt wird an jene Stelle gesetzt, welche die Krümmung im jeweiligen Abschnitt auf zwei neue

Abschnitte gleichverteilt (Abb. 4.7 C und D). Im Gegensatz zu Abb. 4.6 kommen die neu ermittelten dominanten Punkte an Orte zu liegen, an denen die Kurvenkrümmung groß ist (Abb. 4.7 B und E).

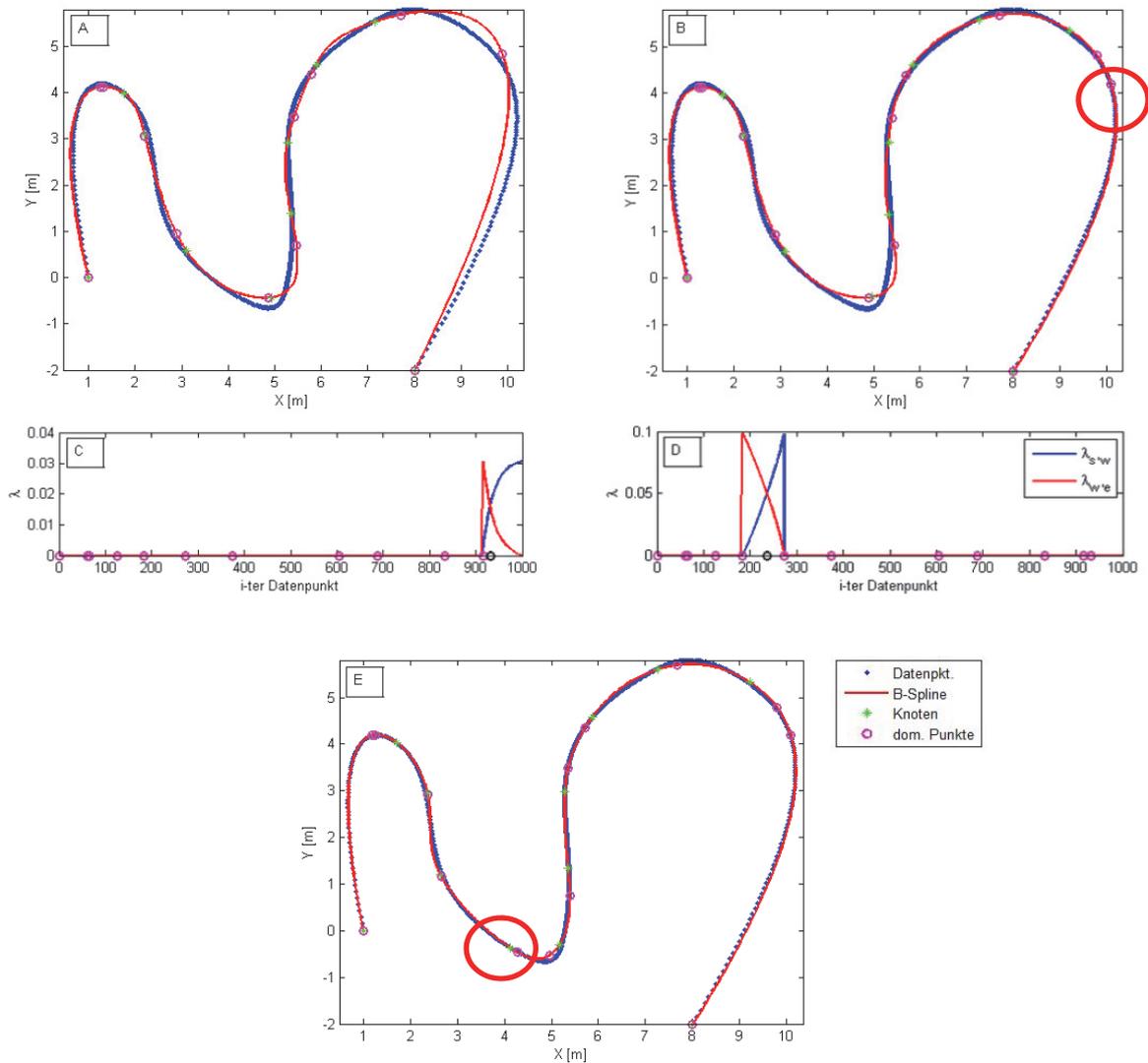


Abb. 4.7: Ermittlung von neuen dominanten Punkten mit $r = 1$ [siehe Formel (4.4)]
A) Ausgangskurve um neuen dominanten Punkt zu berechnen
C) geplotteter Formfaktor λ für i-ten Datenpunkt
und eingezeichneter neuer dominanter Punkt (schwarzer Kreis)
B) B-Spline Kurve mit neuem dominantem Punkt
D) neuerliche Ermittlung eines neuen dominanten Punktes
E) B-Spline Kurve mit neuem dominantem Punkt, welche vorgegebene Toleranz einhält

Da sich gezeigt hat, dass der Einfluss der Bogenlänge bei der Formfaktorberechnung im Allgemeinen größer als jener der Krümmung ist, wurde der Parameter r , wenn nicht anders erwähnt, in allen Berechnungen mit dominanten Punkten gleich 0,8 gesetzt um diesen Effekt auszugleichen.

4.1.5. Evaluierung der Methode

Um zu zeigen, wie gut die vorgestellte Methode unterschiedliche Datenpunkte approximieren kann, werden die in Kapitel 3.1 angegebenen B-Spline Kurven vom Grad 3 benutzt. Der Punkt zu Punktabstand und das angebrachte Rauschen bleiben bei 0,015m und 0,005m bestehen. Die Anzahl an inneren Knoten wird ebenfalls nicht verändert um zu zeigen wie gut das Auffinden passender Knotenpositionen ist. Um die Ergebnisse besser vergleichen zu können, werden die Kontrollpunkte bei den Abbildungen weggelassen. In Abb. 4.8 sind bei allen Berechnungsschritten die Parameterwerte der Datenpunkte bekannt. Die einzigen unbekannt sind deshalb die Knotenpositionen und die Kontrollpunkte.

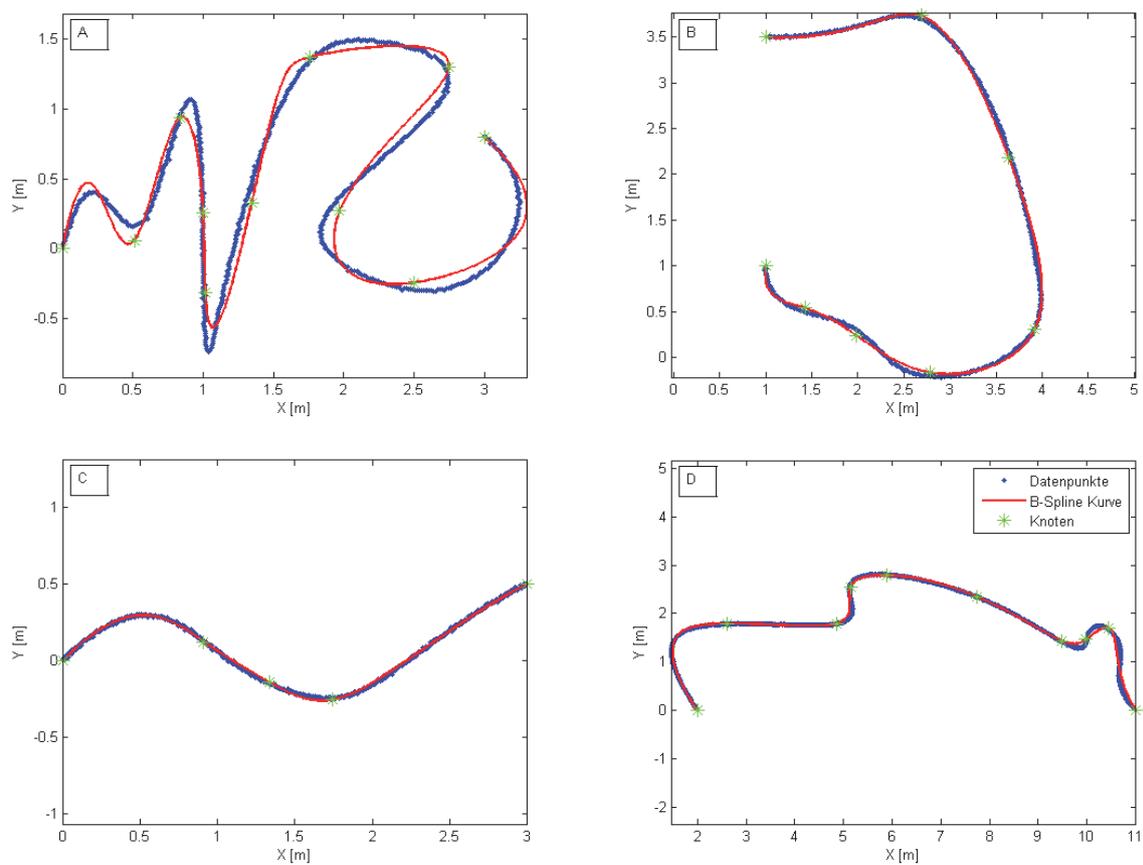


Abb. 4.8: Parameterwerte \bar{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve bekannt

A) 9 innere Knoten B) 6 innere Knoten
C) 3 innere Knoten D) 8 innere Knoten

Tab. 4.1: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierende B-Spline Kurve bei bekannten Parameterwerten \bar{u} der Datenpunkte

Abb. 4.8	A	B	C	D
Anzahl an innere Knoten	9	6	3	8
max. Abweichung [m]	0,176	0,062	0,027	0,118
mittlere Abweichung [m]	0,040	0,021	0,007	0,024

Die Approximation der Daten bei Kurve A (Abb. 4.8) ist wenig zufriedenstellend. Zwar wird die allgemeine Form wiedergegeben, die maximale und mittlere Abweichungen (Formel (4.1) & (4.2)) zwischen der approximierenden B-Spline Kurve und den Datenpunkten (siehe Tab. 4.1) fallen zu groß aus, da diese nicht das an die Daten angebrachte Rauschen ($\sigma = 0,005m$) repräsentieren. Auch mit der Hinzunahme zusätzlicher dominanter Punkte konnte der Fehler nicht wesentlich verringert werden. Dies liegt vor allem an der suboptimalen Verteilung der Knoten im Vergleich zu den originalen Knotenpositionen (siehe Abb. 4.10). Die Kurve B, jene Kurve mit geringen Krümmungen und Krümmungsänderungen, konnte wesentlich besser angepasst werden auch wenn die ermittelten Knotenpositionen nicht mit den Ursprünglichen zusammenpassen. Bei Kurve C fallen die Abweichungen ebenfalls auf Grund der Einfachheit der Form relativ gering aus. Bei der Approximation der Kurve D ist die größte Abweichung an jener Stellen mit der größten Krümmung ($X=10m$, $Y=1,5m$) zu finden.

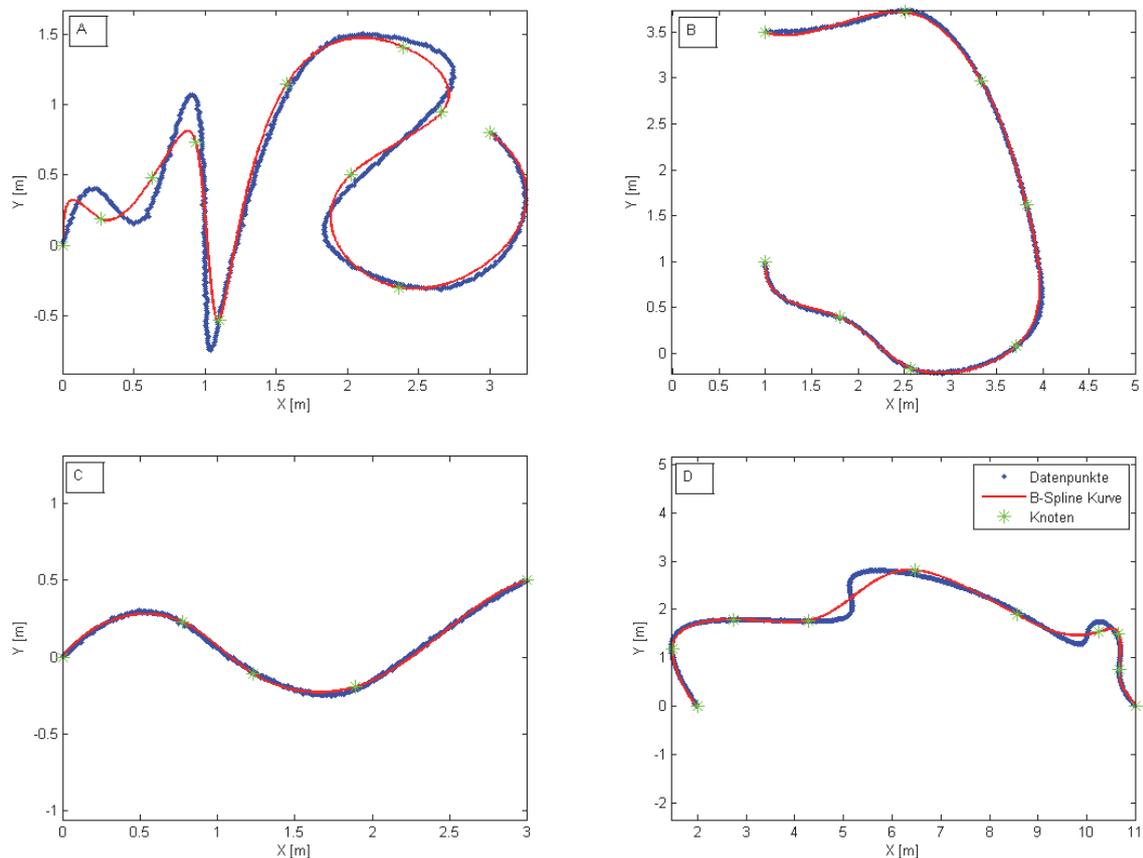


Abb. 4.9: Parameterwerte \bar{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve unbekannt

**A) 9 innere Knoten B) 6 innere Knoten
C) 3 innere Knoten D) 8 innere Knoten**

Tab. 4.2: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierende B-Spline Kurve bei unbekanntem Parameterwerten \bar{u} der Datenpunkte

Abb. 4.9	A	B	C	D
Anzahl an innere Knoten	9	6	3	8
max. Abweichung [m]	0,260	0,046	0,031	0,334
mittlere Abweichung [m]	0,044	0,013	0,010	0,064

Sind die Parameterwerte der Datenpunkte nicht bekannt, sondern müssen zuvor erst z.B. mit der *centripetal* Methode ermittelt werden, so ist die Approximationsgüte im Allgemeinen (bis auf Kurve B) schlechter als bei bekannten Parameterwerten. Deutlich verschlechtert haben sich dabei die Abweichungen bei Kurve A und D. Die Verschlechterung bei Kurve A ist auf die unzureichende Parametrisierung der Datenpunkte mit der *centripetal* Methode zurückzuführen, welche mit der ursprünglichen Parametrisierung in Kurvenbereichen mit starken Krümmungen nicht zusammenpasst. Bei Kurve D ist hingegen, neben der schlechten Parametrisierung der Punkte, vor allem die Basis B-Spline Kurve für das ungünstigere Ergebnis, im Gegensatz zu Abb. 4.8 D, verantwortlich. Durch das leichte Ausschwingen dieser in Bereichen geringer Krümmung werden dominante Punkte falsch gesetzt, welche wiederum die Knotengenerierung dauerhaft negativ beeinflussen. So kommt es schlussendlich in Bereichen starker Krümmung (siehe Abb. D x=4 bis 6m) zu keiner Knotenauffindung. Zu bemerken ist außerdem, dass obwohl am Ende der Kurve (X = 10 bis 11m) mehrere Knoten gesetzt werden, die stärkere Krümmung nicht zufriedenstellend modelliert wird. In Abb. 4.10 können die berechneten Knotenpositionen, sowohl bei bekanntem als auch bei unbekanntem Parametervektor der Daten, mit den originalen Knotenpositionen verglichen werden.

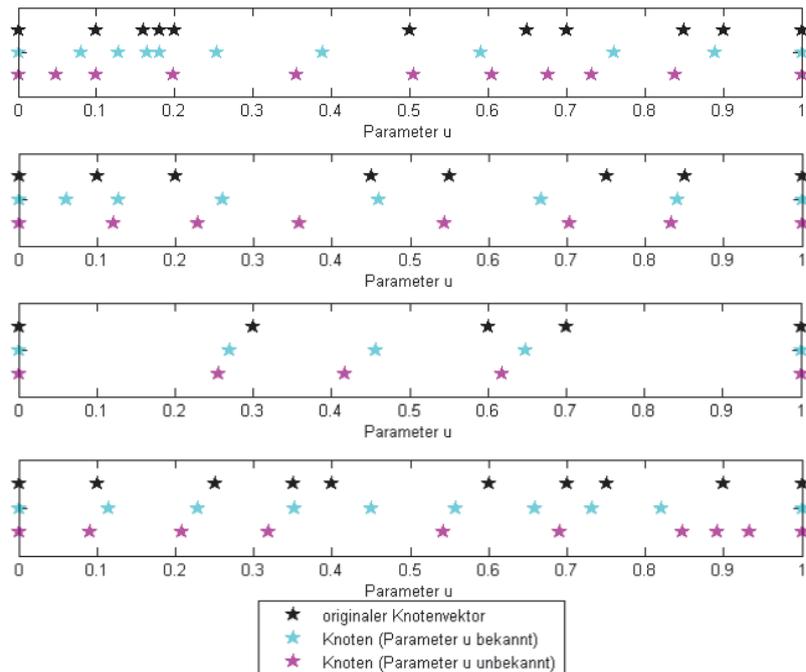


Abb. 4.10: Knotenverteilung für approximierende B-Spline Kurve A, Kurve B, Kurve C und Kurve D der Abb. 4.8 & Abb. 4.9

Da die originalen Knotenpositionen mit Hilfe von dominanten Punkten nicht aufgefunden werden können, stellt sich die Frage wie sehr sich dies auf die Approximationsgüte auswirkt. Dazu wurde für jeden einzelnen Datenpunkt die vorhandene Abweichung zur B-Spline Kurve in X- und Y-Richtung ermittelt. Die Parameterwerte der Datenpunkte \bar{u} waren bei den Berechnungen bekannt. Somit konnten die Residuen einfach über die zugehörigen Punkte ermittelt werden, wobei die Zuordnung über die Parameterwerte \bar{u} erfolgte, da diese die Punkte auf der Kurve eindeutig festlegen. Bei allen vier Datensätzen konnten systematische Effekte beobachtet werden, wenn die ursprüngliche Anzahl an inneren Knotenpunkten für die Approximation der verrauschten Datenpunkte benutzt wurden. Anhand der Abb. 4.11 ist dieser Effekt beim Datensatz 3 klar erkennbar. Sowohl in der X- als auch in der Y-Richtung treten systematische Abweichungen im Bereich von $\pm 3\text{cm}$ zur Soll-Kurve auf, wobei wegen der Kurvenform diese in X-Richtung geringer ausfallen. Dies ist der unzureichenden Positionierung der Knoten (grüne Linien in Abb. 4.11) durch die verwendete Methode zuzuordnen. Es kann darauf geschlossen werden, dass die Modellkomplexität und somit die Anzahl an inneren Knoten nicht ausreicht um die Datenmenge zufriedenstellend zu modellieren.

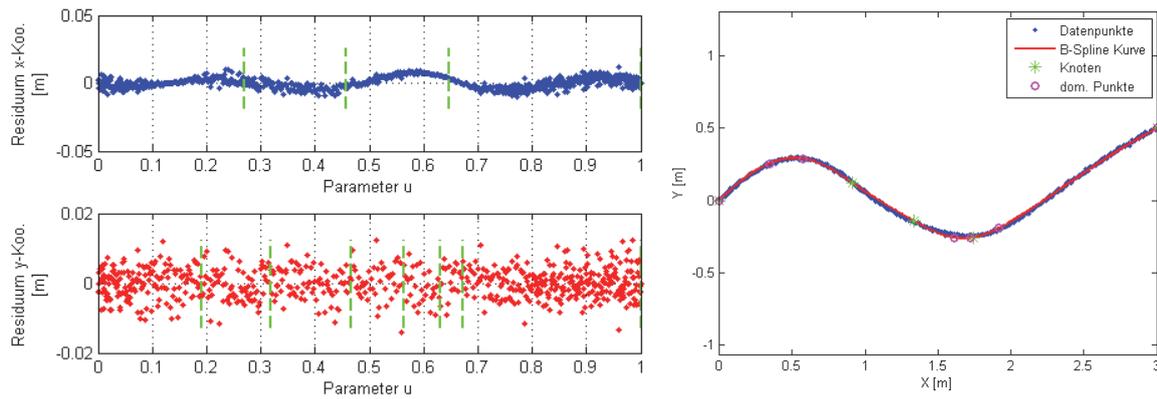


Abb. 4.11: Auftretende Residuen [m] in X- bzw. Y-Richtung mit deutlich erkennbaren systematischen Effekten bei 3 inneren Knoten

Wird die Modellkomplexität erhöht, indem mehr Knoten bei der Approximation zugelassen werden, so kann der systematische Anteil in den Residuen stetig verkleinert werden. Ab einem gewissen Modell (siehe Kapitel 5) ist schlussendlich nur noch das Rauschen beobachtbar (Abb. 4.12) und der entsprechende Datensatz kann optimal approximiert werden. Beim Datensatz 3, unter Verwendung der DOM-Methode, wird dies ab einer inneren Knotenanzahl von sechs erreicht. Die Anzahl an benötigten inneren Knoten ist jedoch doppelt so hoch wie anfänglich für die tatsächliche B-Spline Kurve benötigt wurde.

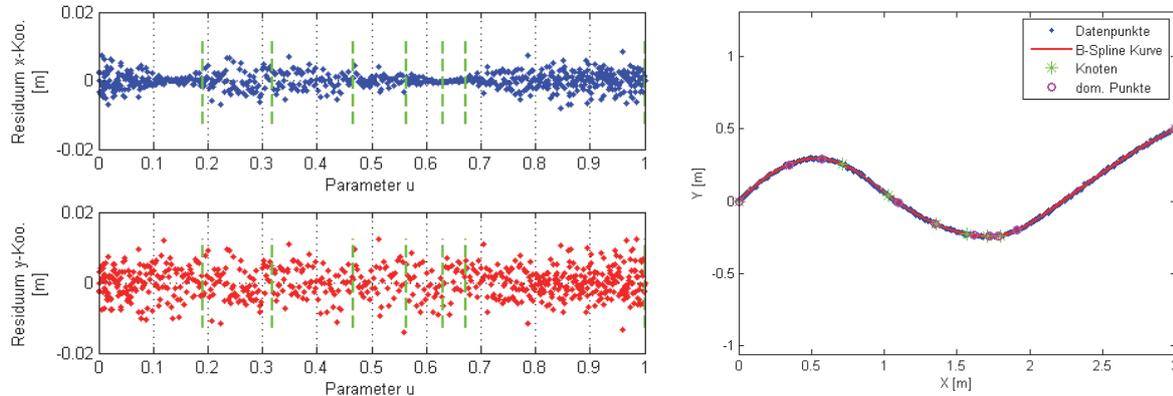


Abb. 4.12: Auftretende Residuen [m] in X- bzw. Y-Richtung mit keinem erkennbaren systematischen Effekt bei 6 inneren Knoten

4.2. Gleichverteilung des Rauschverhältnis (GDR)

Schmitt & Neuner (2014) gehen, bei der von ihnen vorgestellten Methode der Knotengenerierung und Knotenoptimierung, davon aus, dass alle Punkte, welche durch eine B-Spline Kurve approximiert werden sollen, eine gleiche Varianz aufweisen. Dadurch soll sichergestellt werden, lediglich systematische Effekte basierend auf der Geometrie und nicht der Messtechnik durch den neuen Ansatz zu optimiert. Durch diese Annahme wird außerdem die Gleichverteilung der Residuen in jeder Knotenspanne gewährleistet, wodurch im späteren Verlauf im stochastischen Modell die Einheitsmatrix angenommen werden kann.

Die allgemeine Vorgehensweise der Methode ist in Abb. 4.13 schematisch dargestellt. Es handelt sich dabei um einen iterativen Prozess, wobei innerhalb des Iterationsschrittes I (Schritt 3 – 5) ein weiterer Iterationsschritt II (Schritt 5) abläuft. Zu Beginn werden die Parameterwerte der Beobachtungen ermittelt. Hierfür stehen die bereits bekannten Methoden aus Kapitel 2.5.1 zur Verfügung. Anschließend werden die Kontrollpunkte der B-Spline Kurve bei einer Minimalkonfiguration des Knotenvektors im linearen Ausgleichsmodell ermittelt. Die dabei erzeugte eingespannte Kurve entspricht einer Bézier-Kurve, wobei diese den ersten und letzten Datenpunkt berührt und somit in diesen Punkten auch ein Kontrollpunkt liegt (siehe Abb. 4.14). Wird durch diese anfängliche Kurvenapproximation der gegebene Datensatz noch nicht zufriedenstellend approximiert, bzw. ist ein gewisses Kriterium noch nicht erfüllt, so werden weitere Knoten hinzugefügt. Dabei wird in einem Iterationsdurchlauf immer nur ein weiterer Knoten hinzugefügt und anschließend überprüft, ob das Kriterium eingehalten wird oder nicht. Als mögliches Kriterium steht beispielsweise eine maximale Anzahl an Kontrollpunkten oder ein Qualitätsmaß zur Verfügung.

Um die Position des neuen Knotens zu bestimmen, wird in jeder Knotenspanne die kumulierte Summe der quadrierten Residuen zuerst von links nach rechts $kumSum_{l,k}(\bar{u}_k)$ und anschließend von rechts nach links $kumSum_{k,r}(\bar{u}_k)$ berechnet. Stellen $u_l, u_r = u_{l+1}$ zwei aufeinanderfolgende Knoten dar, so ergibt sich die kumulierte Summe zu einem Datenpunkt mit dem Parameterwert \bar{u}_k , wobei dieser innerhalb der Knotenspanne liegt, zu

$$\begin{aligned}
 kumSum_{l,k}(\bar{u}_k) &= \sum_{\bar{u}_n=u_l}^{\bar{u}_k} (\mathbf{C}(\bar{u}_n) - \mathbf{D}_k)^2 \\
 kumSum_{k,r}(\bar{u}_k) &= \sum_{\bar{u}_n=\bar{u}_k}^{u_r} (\mathbf{C}(\bar{u}_n) - \mathbf{D}_k)^2
 \end{aligned} \tag{4.8}$$

wobei $u_l < \bar{u}_k < u_r$ gilt.

Die kumulierte Summe (4.8) kann als potentielle Energie der Residuen innerhalb einer Knotenspanne verstanden werden. Eine mögliche Knotenposition innerhalb jeder Knotenspanne ist am Parameterwert \bar{u}_k zu finden, an der die beiden Versionen der kumulierten Summen (von links und von rechts) denselben Wert aufweisen und somit folgende Bedingung erfüllt wird:

$$\min_k |kumSum_{l,k} - kumSum_{k,r}| \quad l < k < e \quad (4.9)$$

Wurde für jede Knotenspanne eine Knotenposition und der zugehörige Wert gefunden, wird anschließend ein neuer Knoten in jene Knotenspanne eingefügt, welche den größten Wert aufweist (siehe Abb. 4.14 D). Durch diese Vorgehensweise soll die höchste vorhandene potentielle Energie verringert werden. Anschließend werden wiederum die Kontrollpunkte im Ausgleich mit dem neuen Knotenvektor ermittelt und der neu gefundene Knoten in einem nichtlinearen iterativen Schritt verbessert. Hierfür wird ein beschränktes lineares GMM benutzt, in welchem die Position des neuen Knotens frei ist, die restlichen Knoten jedoch durch Bedingungen festgehalten werden.

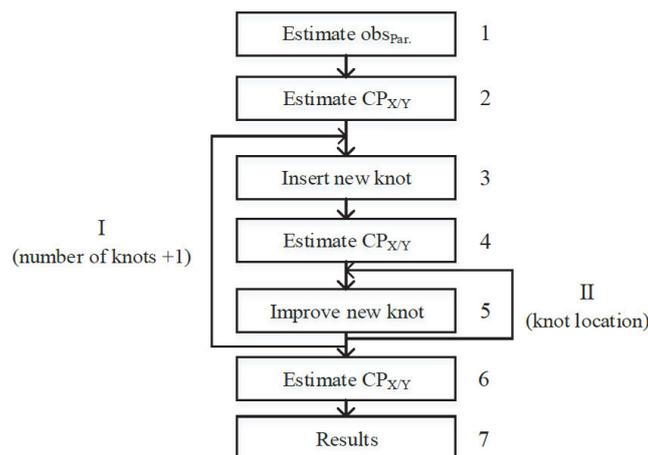
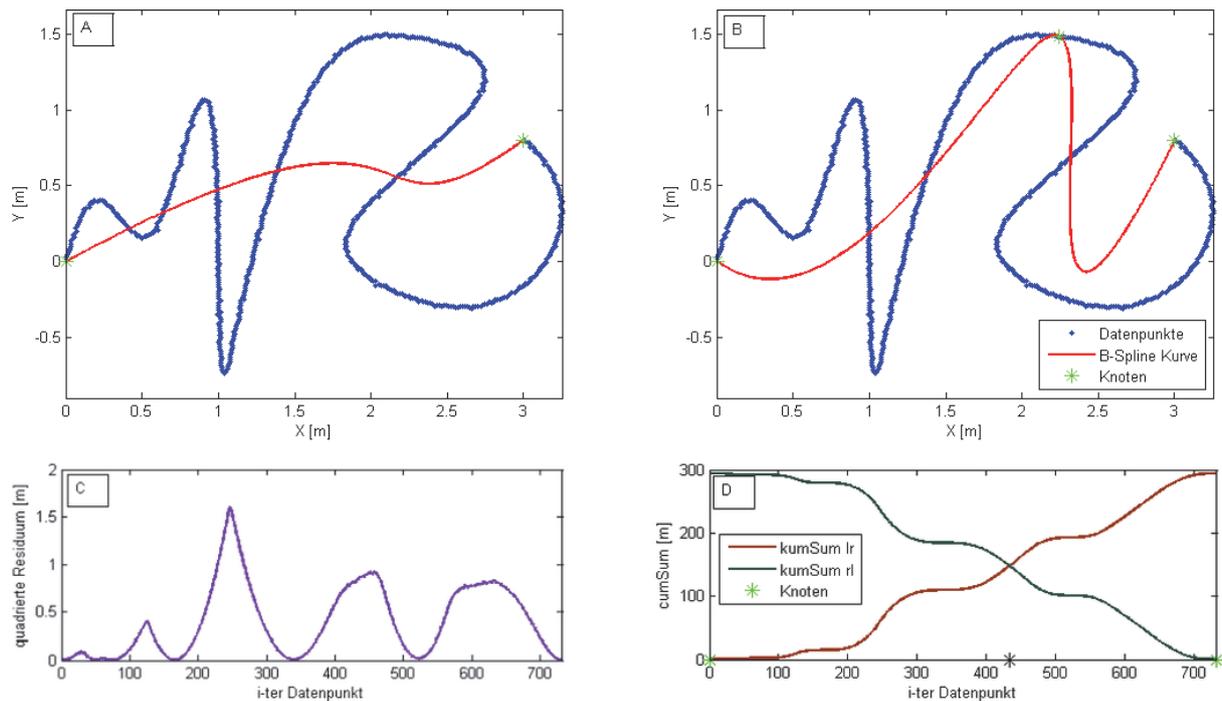


Abb. 4.13: Schematischer Ablauf der GDA-Methode

4.2.1. Evaluierung der Methode

Die Vorgehensweise des eben beschriebenen Algorithmus wird im Folgenden an einem praktischen Beispiel demonstriert, wobei jedoch der nichtlineare iterative Prozessionsschritt II (Schritt 5) in der Knotengenerierung nicht benutzt wird. Es werden die Datenpunkte der bekannten B-Spline Kurve aus Abb. 3.1 A verwendet, welche im euklidischen Raum einen regelmäßigen Punkt zu Punktabstand von 0,015m aufweisen. Außerdem wird das normalverteilte Rauschen von $\sigma = 0,005m$ an den Daten belassen (Abb. 4.14).



**Abb. 4.14: A) Datenpunkte durch Bézier-Kurve approximiert
 B) Datenpunkte mit B-Spline Kurve (1 innerer Knoten) approximiert
 C) quadrierte Abweichung zwischen Bézier-Kurve und Datenpunkten
 D) von links bzw. von rechts aufsummierte quadratische Abweichung (kumSum)
 und eingezeichneter neuer Knoten (schwarzer Stern)**

Zu Beginn der Methode wird eine Bézier-Kurve an die Datenpunkte approximiert, wobei für diese Berechnung die Parameterwerte der Datenpunkte, wegen der benutzten B-Spline Kurve, bekannt sind. Dadurch soll zuerst die Fähigkeit des Algorithmus getestet werden, die Knotenpositionen richtig aufzufinden. Anschließend wird ausgehend von der Bézier-Kurve die quadrierte Abweichung für jeden Datenpunkt berechnet und diese wiederum über die Datenpunkte, der zurzeit einzig bestehenden Knotenspanne, aufsummiert (siehe Formel (4.8)). An jenem Ort, an dem sich die von links bzw. von rechts berechneten kumulierten Summen schneiden, wird ein neuer Knoten platziert (siehe Formel (4.9)). Die neue Knotenposition wird in Abb. 4.14 mit einem schwarzen Stern markiert. Durch diese Berechnungsmethode teilt der neue Knoten die bestehende Abweichung innerhalb des Intervalls gleichmäßig auf zwei neue Intervalle auf. In Abb. 4.14 B ist jene B-Spline Kurve zu sehen, welche mit Hilfe des neu ermittelten Knotens berechnet wird. Dieser Prozess wird anschließend so lange wiederholt, bis die ursprüngliche Anzahl an inneren Knoten der Ausgangs-B-Spline Kurve erreicht wird (Abb. 4.15 A). In diesem Fall hat die B-Spline Kurve 9 innere Knoten.

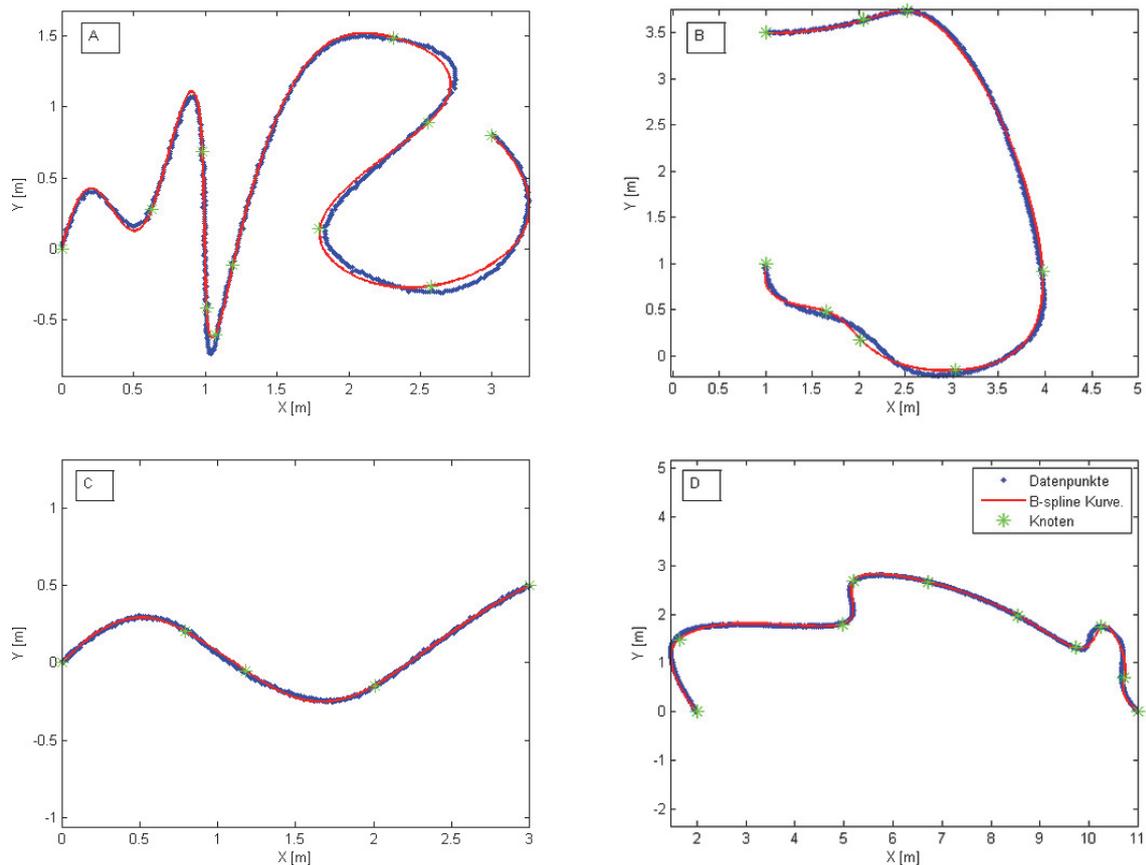


Abb. 4.15: Parameterwerte \bar{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve bekannt
A) 9 innere Knoten B) 6 innere Knoten
C) 3 innere Knoten D) 8 innere Knoten

Der Algorithmus wird wiederum an den unterschiedlichen Kurvenformen aus dem Kapitel 3.1 ausprobiert (Abb. 4.15). Um die einzelnen Kurven zu vergleichen, wird für alle Kurven der Punkt zu Punktabstand bei $0,015\text{m}$ und das Rauschen bei $\sigma = 0,005\text{m}$ belassen. In Abb. 4.15 sind die Parameterwerte der Datenpunkte bei den Berechnungen bekannt. Alle Kurven sind dabei B-Spline Kurven vom Grad 3. Für jede Kurve werden insgesamt so viele innere Knoten ermittelt, wie für die ursprüngliche B-Spline Kurve notwendig waren. Bei der Abb. 4.15 und Abb. 4.16 sind die Kontrollpunkte nicht eingezeichnet, um die Ergebnisse besser vergleichen zu können.

In der folgenden Tabelle (Tab. 4.3) ist neben der maximalen Abweichung (siehe Formel (4.1)) auch die mittlere Abweichung der einzelnen B-Spline Kurven zu den jeweiligen Datenpunkten ersichtlich. Diese wird mit Hilfe der Formel (4.2) berechnet.

Tab. 4.3: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierende B-Spline Kurve bei bekannten Parameterwerten \bar{u} der Datenpunkte

Abb. 4.15	A	B	C	D
Anzahl an innere Knoten	9	6	3	8
max. Abweichung [m]	0,112	0,090	0,024	0,102
mittlere Abweichung [m]	0,017	0,022	0,006	0,019

In Abb. 4.16 wurden hingegen die Parameterwerte der Datenpunkte mit Hilfe der *centripetal* Methode ermittelt. In der Tab. 4.4 ist wiederum die maximale und mittlere Abweichung angegeben.

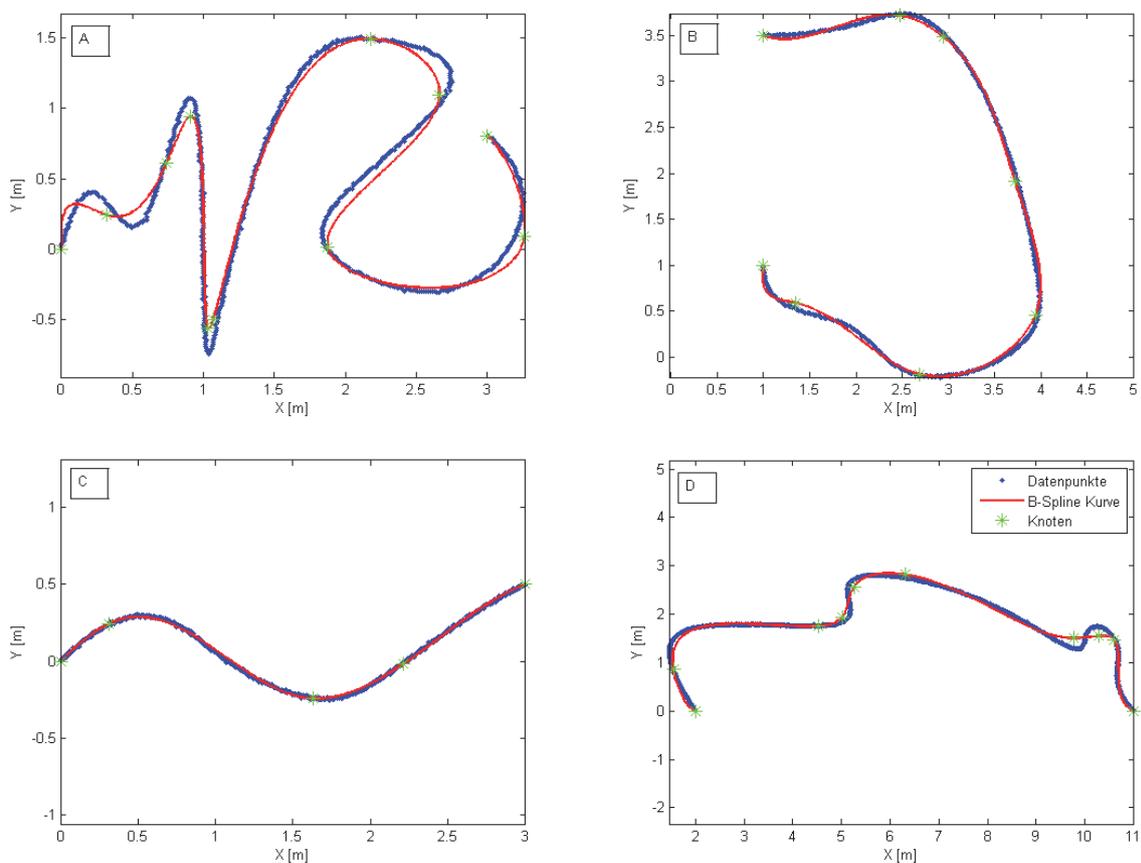


Abb. 4.16: Parameterwerte \bar{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve unbekannt

A) 9 innere Knoten B) 6 innere Knoten
C) 3 innere Knoten D) 8 innere Knoten

Tab. 4.4: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierender B-Spline Kurve bei unbekanntem Parameterwerten \bar{u} der Datenpunkte

Abb. 4.16	A	B	C	D
Anzahl an innere Knoten	9	6	3	8
max. Abweichung [m]	0,176	0,079	0,023	0,227
mittlere Abweichung [m]	0,035	0,023	0,007	0,049

Insgesamt kann wie bereits in Kapitel 4.1.5 beobachtet werden, dass sobald die Parameterwerte der Datenpunkte bekannt sind, eine bessere Knotengenerierung und Datenapproximation stattfindet als mit unbekanntem Parameterwerten. Dies liegt daran, dass vor allem die Parameterwerte, welche mit Hilfe der *centripetal* Methode ermittelt wurden, in Kurvenbereichen mit starker Krümmung beträchtlich von den ursprünglichen Parameterwerten abweichen. Kurven mit geringer und nicht stark wechselnder Krümmung (Abb. 4.15 B; C) können deshalb besser approximiert werden als jene mit einer starken und schnell springenden Krümmung (Abb. 4.15 A; D). Dies ist vor allem in Abb. 4.16 A zu sehen, wo im Bereich von $X = 1m$ eine starke Krümmung zu beobachten ist, welche unzureichend durch die B-Spline Kurve approximiert wird. Trotzdem liegt die mittlere Abweichung der B-Spline Kurve bei drei von vier Kurven bei nur etwa 0,02m, wenn die ursprüngliche Anzahl an inneren Knoten benutzt wird. Erhöht man die Anzahl an inneren Knoten, so passt sich die Kurve stetig besser den Datenpunkten an. Sind die Parameterwerte der Datenpunkte jedoch nicht bekannt, so muss zumeist mit einer schlechteren Datenapproximation gerechnet werden. So verschlechtert sich bei allen vier Kurven die mittlere Abweichung, wobei jedoch die Verschlechterung bei Kurve B und C sehr gering ausfällt.

In der Abb. 4.17 sind die für die Kurven in Abb. 4.15 und Abb. 4.16 entsprechenden Knoten entlang der Parameterachse aufgezeichnet. Die Knoten unterscheiden sich dabei vom originalen Knotenvektor mitunter stark. Zumeist kann die originale Knotenposition nicht aufgefunden werden. Lediglich bei Kurve A und bei bekanntem Parametervektor konnten die ursprünglichen Knoten näherungsweise gefunden werden.

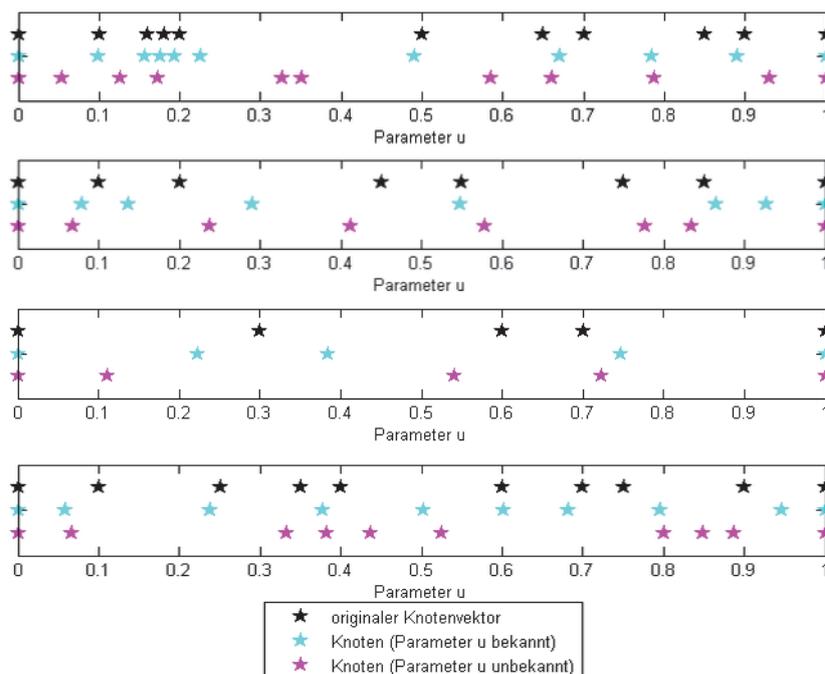


Abb. 4.17: Knotenverteilung für approximierende B-Spline Kurve A, Kurve B, Kurve C und Kurve D der Abb. 4.15 & Abb. 4.16

Im folgenden Abschnitt wird überprüft, wie sehr das Auffinden der Knotenposition vom Datenrauschen abhängig ist. Dazu werden die Daten der Kurve aus Abb. 3.1 A mit einem immer größer werdenden normalverteiltem Rauschen verrauscht und anschließend die Knotenposition, wie im Kapitel 4.2 beschrieben, ermittelt. Die Anzahl an zu bestimmenden inneren Knoten, wird mit jener gleichgesetzt, welche für die Datengenerierung notwendig war. Die Parameterwerte der Datenpunkte sind bei den Berechnungen bekannt. Der Vorgang der Knotenermittlung wird dabei für jeden Rauschlevel 100-mal wiederholt, wobei die Daten bei jedem Durchlauf neu verrauscht werden. Die Knotenpositionen pro Rauschlevel, welche in Abb. 4.18 zu sehen sind, stellen dabei den Mittelwert dar. Zu jeder Knotenposition wird aus den 100 Realisierungen auch der minimale bzw. maximale Knotenwert ermittelt, welcher bei jeder Knotenposition in Abb. 4.18 als farbliche Linie dargestellt ist. Diese gibt Auskunft darüber, wie sehr sich die Knotenposition innerhalb eines Rauschlevels bei 100-mal unterschiedlich verrauschten Daten verändern kann.

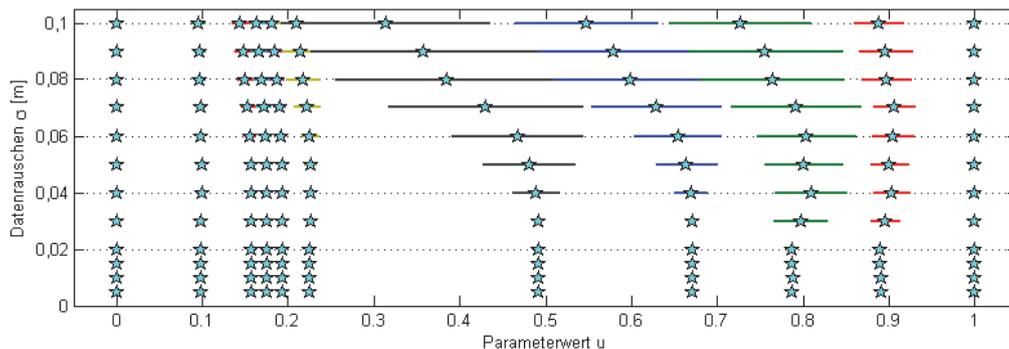


Abb. 4.18: Mittlere Knotenposition (Stern) plus zugehörige Standardabweichung (gefärbte Linie) für unterschiedliche Rauschlevels (von $\sigma = 0,005m$ bis $\sigma = 0,1m$)

Wie anhand von Abb. 4.18 zu sehen ist, ist die Knotenauffindung bei leicht verrauschten Daten (bis $\sigma = 0,02m$) robust. Sowohl die Veränderung der mittleren Knotenpositionen zwischen den ersten vier Rauschlevels (von $\sigma = 0,005m$ bis $\sigma = 0,02m$) als auch die minimale und maximale Abweichung jeder Knotenposition innerhalb eines Rauschlevels fallen sehr gering aus. Bei stärker verrauschten Daten (ab $\sigma = 0,03m$) verändern sich die mittleren Knotenpositionen zwischen unterschiedlichen Rauschlevels immer deutlicher. Besonders betroffen davon sind dabei die inneren Knoten sechs bis neun, bei denen auch gleichzeitig die Abweichungen mit zunehmendem Rauschen ansteigen. Daraus ergibt sich, dass die verschiedenen Knotenpositionen mit der Methode von Schmitt & Neuner, bei schwach unterschiedlich verrauschten Daten, wiedergefunden werden können. Die originalen Knotenpositionen können hingegen nicht aufgefunden werden.

Es stellt sich somit die Frage wie sich das nicht Auffinden der originalen Knotenpositionen auf die B-Spline Approximation auswirkt. Wird versucht die Datenpunkte mit der ursprünglichen Anzahl an inneren Knoten zu modellieren, so treten zwischen den einzelnen

Knotenspannen (grüne Linien in nachfolgenden Abbildungen) systematische Effekte bei den Residuen in X- und Y-Richtung auf (Abb. 4.19), da nicht der tatsächliche Knotenvektor aufgefunden werden konnte (siehe Abb. 4.17). Diese Effekte sind ein Hinweis darauf, dass sich das funktionale Modell den Daten noch nicht gut genug anpassen kann und somit ein Underfitting besteht. Es muss daher die Modellkomplexität zunehmen, indem die Parameteranzahl und somit die Anzahl an Knoten erhöht wird.

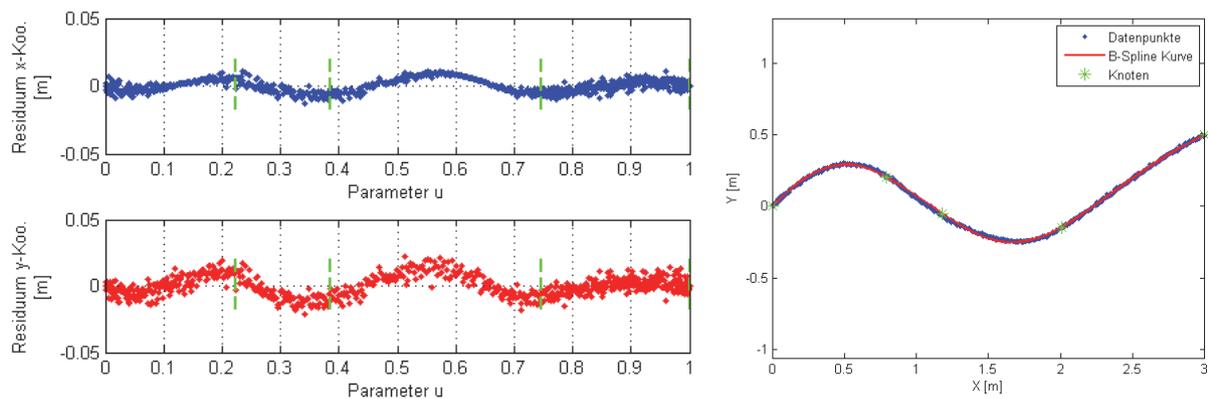


Abb. 4.19: Auftretende Residuen [m] in X- bzw. Y-Richtung mit deutlich erkennbaren systematischen Effekten bei 3 inneren Knoten

Komplexere Modelle bergen jedoch die Gefahr des Overfittings. Dabei wird nicht nur das eigentliche Signal sondern auch bereits das Rauschen der Daten mitmodelliert. Eine Möglichkeit dies zu verhindern bieten so genannte Informationskriterien welche in Kapitel 5 ausführlich behandelt werden. Durch die Auswahl eines passenden Modells mit ausreichender Komplexität ($\hat{=}$ passende Anzahl an Parametern) ist schlussendlich nur noch das Rauschen im Residuum-Plot zu sehen (Abb. 4.20). Die Anzahl an benötigten inneren Knoten ist jedoch dreimal so hoch wie für die anfängliche B-Spline Kurve. Es ist jedoch klar ersichtlich, dass die Hinzunahme von zusätzlichen Knoten die Residuen und somit auch die potentielle Energie innerhalb einer Knotenspanne deutlich reduzieren.

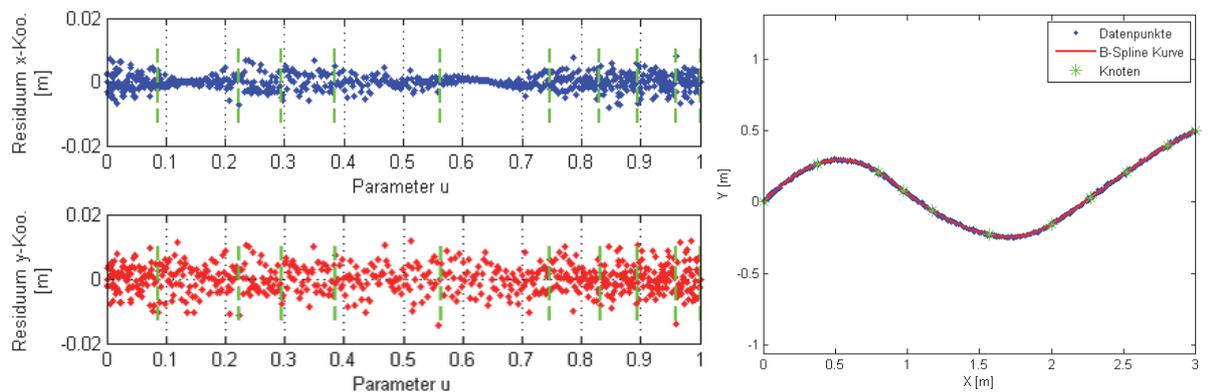


Abb. 4.20: Auftretende Residuen [m] in X- bzw. Y-Richtung mit keinem erkennbarem systematischem Effekt bei 9 inneren Knoten

4.3. Elitist Clonal Selection Algorithm (CSA)

Der *Elitist Clonal Selection Algorithm (CSA)* von Galvez et al. (2015) ist ein AIS (*Artificial Immun System*) Schema, welches ein Teil des natürlichen Immunsystems nachahmt und sich auf die *Clonal Selection* Theorie stützt.

Die *Clonal Selection* Theorie, vorgestellt und beschrieben von Castro & Zuben (2000), erklärt die grundlegende Charakteristik einer lernfähigen Immunreaktion auf einen Antigenreiz. Gelangt ein krankheitserregender Mikroorganismus (=Antigen Ag) in den Körper, so aktiviert das Immunsystem des Wirts die Produktion von Antigenrezeptoren (=Antikörper Ab), welche zur Verteidigung des Immunsystems dienen. Diese Antikörper sind auf bestimmte Antigene spezialisiert. Ein Maß für die Spezialisierung eines Ab bezüglich eines Ag ist die Affinität, die für ein bestimmtes Ab-Ag-Paar spezifisch ist. Ein Antikörper mit einer hohen Affinität ist im hohen Grad auf das vorhandene Antigen angepasst. Deshalb vermehren sich jene Antikörper mit der höchsten Affinität am stärksten. Während der Vermehrung steigt die mittlere Affinität bezüglich des Antigens an, da während der Vermehrung zufällige kleine Mutationen des Antikörpers auftreten und durch einen Selektionsmechanismus gewährleistet wird, dass nur jene Antikörper sich weiter vermehren, welche eine erhöhte Affinität aufweisen. Antikörper mit einer geringen Affinität werden vernichtet. Die Anzahl an spezialisierten Antikörpern steigt bei dieser Vorgehensweise rasch. Für eine tiefere Auseinandersetzung mit der *Clonal Selection* Theorie wird der interessierte Leser auf die beiden Fachartikel von Castro & Zuben (2000 & 2002) verwiesen.

Im Folgenden werden die einzelnen Arbeitsschritte des *Elitist Clonal Selection Algorithm*, welcher von Galvez et al. durch adaptieren des CSA entwickelt wurde und in Abb. 4.21 schematisch aufgezeigt sind, beschrieben. Dabei wird für die Vorstellung der Methode bei Galvez et al. eine Punktmenge $\{x_i, y_i\}_{i=0, \dots, h}$ benutzt, welche die Form

$$y_i = \varphi(x_i) + \sigma_i \quad i = 0, \dots, h \quad (4.10)$$

hat. $\varphi(x)$ stellt die zu Grunde liegende unbekannt Funktion dar, welche in diesem Fall der B-Spline Kurvenfunktion (siehe Formel (2.2)) entspricht. Außerdem werden zur Generierung von Datenpunkten, welche anschließend verrauscht werden, bekannte Funktionen benutzt. Durch diese Vorgehensweise ist die Parametrisierung der Punkte bereits vorab bekannt, wodurch nur noch die Kontrollpunkte und die inneren Knoten als unbekannt gelten.

Die Notation (4.10) wird im Nachfolgenden bei der Vorstellung des Algorithmus nicht benutzt. Vielmehr wird der allgemeinere Fall aufgezeigt, bei dem die Parametrisierung der Punktmenge nicht bekannt ist. Antigene werden dabei mit Ag und Antikörper mit Ab abgekürzt.

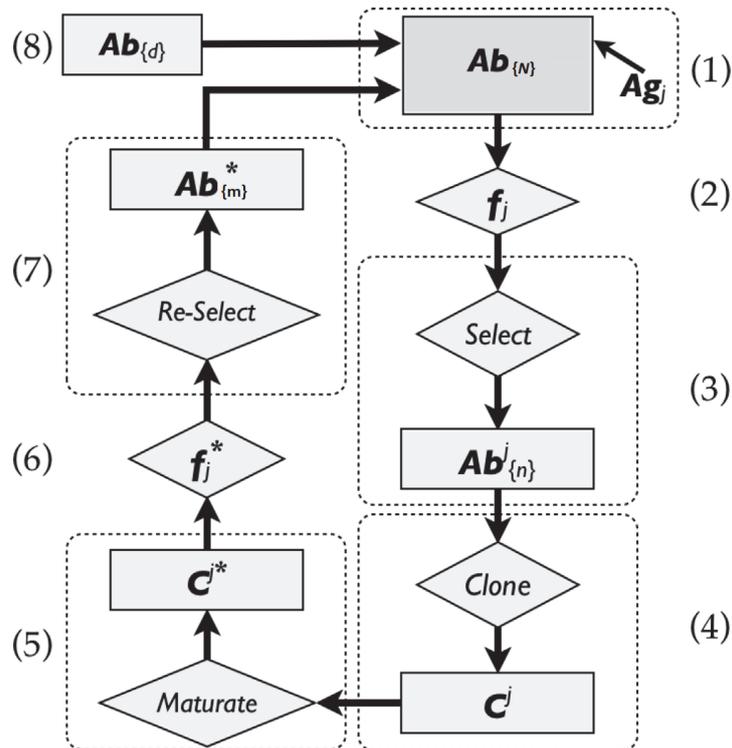


Abb. 4.21: Flussdiagramm des Elitist Clonal Selection Algorithmus für Optimierungsprobleme, Castro & Zuben (2002)

- 1) Der ursprüngliche Zweck des CSA war es Muster zu erkennen, weshalb es ein Set an mehreren unterschiedlichen Antigenen gab. Bei Optimierungsproblemen gibt es keine Antigen Population sondern lediglich ein Antigen Ag_j (z.B. eine zu optimierende Funktion $g(-)$ bzw. eine zu approximierende Datenmenge), sodass jeder einzelne Antikörper (Ab_m) ein Element des Lösungsraums und die gesamte Antikörper-Population das Memory-Set darstellt.

Im Falle der Knotenschätzung bilden $m_i = n - 2p - 1$ innere Knoten ein Antikörper, wobei n die Anzahl an Basisfunktionen und p den Grad der Basisfunktionen darstellen. Dabei werden die inneren Knoten zu Beginn gleichverteilt und zufällig im Intervall $[0,1]$ ausgewählt. Anschließend werden diese in aufsteigender Reihenfolge zu einem inneren Knotenvektor $U_i = \{u_{p+1}, u_{p+2}, \dots, u_{n-p-1}\}$ sortiert. Ein Knotenvektor stellt ein Antikörper dar, wobei am Anfang und am Ende des inneren Knotenvektors zusätzlich $p + 1$ Knoten hinzugefügt werden müssen (siehe Kapitel 2.2.1 „eingespannte Kurve“). Die Anzahl N an Knotenvektoren (=Populationsgröße) muss a priori festgelegt werden und beeinflusst die Rechenzeit, das Konvergenzverhalten als auch die Größe des Suchraums.

- 2) Im zweiten Arbeitsschritt wird der Affinitätsvektor $\mathbf{f} = [f_1, f_2, \dots, f_N]$ für alle N Knotenvektoren berechnet, wobei $f_j = Af(Ab_j, Ag)$ die Affinität des j-ten Knotenvektors Ab_j bezüglich des Antigens Ag (=die gemessenen Datenpunkte \mathbf{D}) angibt. Ziel ist es die *Least Square Error* Funktion,

$$\Theta(\mathbf{U}_j) = \sum_{i=0}^h (\mathbf{D}_i - \mathbf{C}(\bar{u}_i))^2 \quad (4.11)$$

\mathbf{D}_i ... Datenpunkt

$\mathbf{C}(\bar{u}_i)$... B Spline Kurvenpunkt

\mathbf{U}_j ... Knotenvektor

welche die Affinität für einen bestimmten Knotenvektor \mathbf{U}_j angibt, zu minimieren. Deshalb ist bei der Knotenschätzung jener Knotenvektor zu bevorzugen, welcher die Affinität und somit den Abstand der B-Spline Kurve zu den Daten minimiert. Die *Least Square Error* Funktion wird in diesem Kontext auch als *Residual Sum of Squares* (RSS) bezeichnet.

Die Anzahl an Datenpunkten wird bei dieser Funktion nicht berücksichtigt, weshalb auch die *Root Mean Square Error* (RMSE) Funktion berechnet wird.

$$RMSE(\mathbf{U}_j) = \sqrt{\frac{\sum_{i=0}^h (\mathbf{D}_i - \mathbf{C}(\bar{u}_i))^2}{h + 1}} \quad (4.12)$$

- 3) Um einen optimalen Knotenvektor zu bekommen werden nur die n besten Antikörper (Knotenvektoren) mit den geringsten Affinitätswerten ausgewählt, zusammengefasst $Ab_{\{n\}}$ und im nächsten Schritt geklont. Das Ziel besteht nämlich nicht darin, lokale Optima, sondern das globale Optimum zu finden. Die Größe von n liegt typisch im Bereich von 10-30% der Populationsgröße N.
- 4) Die Gesamtanzahl an Klonen N_c ist abhängig von der Populationsgröße N und von einem Multiplikationsfaktor a.

$$N_c = \sum_{i=1}^n \text{round}[(v_i + 1) * a * N] \quad (4.13)$$

wobei $v_1 = 5, v_2 = 3, v_3 = 1$ und $v_i = 0$ für $i \geq 4$ a ... Multiplikator

Antikörper mit geringen Affinitätswerten werden dabei öfters geklont, da die Wahrscheinlichkeit ein globales Optimum gefunden zu haben größer ist. Alle Knotenvektoren werden in eine Matrix C zusammengefasst.

- 5) Im nächsten Schritt wird in jedem Knotenvektor ein zufällig ausgewählter innerer Knoten u_i mutiert. Der ausgewählte Knoten erfährt eine Mutationsschrittweite welche wie folgt berechnet wird:

$$u_i \rightarrow u_i - \frac{\Delta}{2}(\sigma - 0.5) \quad (4.14)$$

$$\text{wobei } \Delta = \min\left\{\frac{u_i}{2}, 1 - u_i\right\}, \sigma = N(0,1)$$

Alle restlichen Knoten bleiben unverändert. Dabei kann der Zähler Δ in (4.14), unter der Voraussetzung, dass die Knoten im Intervall $[0,1]$ definiert wurden, maximal den Wert 0.33 annehmen. Folglich ist deshalb die maximale Mutationsschrittweite eines inneren Knotens mit ± 0.16 beschränkt. Die inneren Knoten müssen anschließend neuerlich in aufsteigender Reihenfolge sortiert werden. Die mutierten Knotenvektoren werden in der Matrix C^* zusammengefasst.

- 6) Anschließend wird die Affinität (siehe Punkt 2) für jeden mutierten Klon berechnet und im Affinitätsvektor f^* zusammengefasst.
- 7) Mehrere mutierte Klone, welche eine hohe Affinität aufweisen, werden daraufhin ausgewählt und zu einer neuen Generation $Ab_{\{m\}}^*$ zusammengefasst.
- 8) Um die Durchmischung der Population zu fördern werden d Knotenvektoren mit den geringsten Affinitätswerten durch zufällig generierte Knotenvektoren $Ab_{\{d\}}$ ersetzt. Die Zahl d ist mit Bedacht zu wählen, da diese die Untersuchungsmöglichkeit des Algorithmus erhöhen aber auch verringern kann.

Außerdem werden N_e Knotenvektoren einer Generation durch gute nicht mutierte Knotenvektoren der vorhergehenden Generation ersetzt, um eine schnellere Konvergenz zu erreichen.

Die Arbeitsschritte 1) – 8) werden solange iterativ abgearbeitet bis die vorgegebene Anzahl an Iterationsschritten N_{iter} erreicht wurde.

Eine a priori sinnvolle Festlegung der Parameter der Methode kann ohne experimenteller Untersuchung nicht geschehen und ist von Fall zu Fall unterschiedlich. Sobald alle Parameter $\{N, d, n, a, N_c, N_e, N_{iter}\}$ der Methode feststehen, muss lediglich der Grad p der approximierenden B-Spline Kurve und die Anzahl an inneren Knoten m_i vorgegeben werden. Eine angemessene Wahl des Grades ist dabei wichtig, da diese die Approximation global beeinflusst (siehe Kapitel 2.2.3). Wird dieser zu gering angesetzt, so kann die Flexibilität der B-Spline Kurve nicht ausgenutzt werden, wohingegen ein zu großer Grad ein Ausschwingen der Kurve begünstigt und gleichzeitig auch die Berechnungszeit ansteigt.

Die Anzahl an inneren Knoten ist hingegen ein Modellselektionsproblem (siehe Kapitel 5). Die unter Punkt 2 genannten Funktionen (4.11) und (4.12) berücksichtigen nicht die Komplexität des Modells. Das bedeutet, dass der geringste Fehler in den Funktionen dann auftritt, wenn die Anzahl der frei verfügbaren Variablen (Knoten und Kontrollpunkte) ansteigt und sich somit die Komplexität des Modells erhöht. Um diesen Sachverhalt abzufangen, können zusätzliche Informationskriterien (AIC & BIC) für die Modellselektion implementiert werden, welche die Anzahl an Unbekannten berücksichtigen. Diese Funktionen bestimmen, welches Modell am besten zu den tatsächlichen Messwerten passt. In Kapitel 5 wird auf die Modellselektion und die zugehörigen Informationskriterien näher eingegangen.

An die inneren Knoten werden im Laufe des Berechnungsprozesses keine Bedingungen gestellt, sodass auch Knoten mit Vielfachheit l auftreten können. An Knoten mit Vielfachheit l ist die B-Spline Kurve anschließend C^{p-l} stetig. Eine menschliche Intervention bei der Knotengenerierung ist nicht notwendig. Nach Ablauf aller Iterationsschritte N_{iter} wird jener Knotenvektor mit der geringsten Affinität als Lösung herangezogen.

Anschließend können die Arbeitsschritte 1) – 8) für unterschiedliche Anzahlen an inneren Knoten wiederholt werden, wobei der Grad p der approximierenden B-Spline Kurve für Vergleichszwecke gleich bleiben muss. Führt man die Berechnungen für n unterschiedliche Anzahlen an inneren Knoten durch, so bekommt man n optimale Knotenvektoren, wobei anschließend jener Knotenvektor mit dem geringsten AIC- bzw. BIC-Wert als Optimum angesehen wird (siehe Kapitel 5.3).

4.3.1. Evaluierung der Methode

Die vorgestellte Methode wird wieder an den verschiedenen Punktdaten, welche aus den bekannten B-Spline Kurven aus dem Kapitel 3.1 entnommen werden, ausprobiert. Der Punkt zu Punktabstand wird bei $0,015m$ belassen und die Daten werden auch dieses Mal mit einem Rauschen von $\sigma = 0,005m$ verrauscht. Zuerst müssen die Parameter der Methode so bestimmt werden, dass bei allen Kurvenapproximationen eine zufriedenstellende optimale Knotenauffindung gewährleistet werden kann. Diese Bestimmung muss empirisch durchgeführt werden, da im Allgemeinen bei einem Zufallsprozess (Knoten werden zufällig generiert und Mutation ist zufällig) nicht vorausgesagt werden kann wie die Parameter zu wählen sind. Ein Indiz dafür, ob die Parameter $\{N, d, n, a, N_c, N_e, N_{iter}\}$ richtig gewählt wurden, kann die berechnete Affinität (4.9) des besten Knotenvektors innerhalb einer Generation sein. In Abb. 4.22 sind für die vier Kurven pro Iterationsschritt die logarithmierten Affinitätswerte dargestellt, wobei bei jeder Approximation der Daten die Anzahl an inneren Knoten jener entspricht, welche für die ursprüngliche B-Spline Kurve benutzt wurde. Bei bekannten Parameterwerten \bar{u} der Datenpunkte (Abb. 4.22 A) ändert sich ab dem 60.

Iterationsschritt die Affinität kaum noch. Die Grafik mit den Affinitätswerten wurde für verschiedene Parameterwerte (bezüglich des CSA) erstellt und anschließend jene Parameterwerte (siehe Tab. 4.5) für alle Berechnungen übernommen, welche eine schnelle Verringerung der Affinität pro Generation ermöglichten. Dabei musste ein Kompromiss eingegangen werden, da nicht alle Datensätze die gleiche Anzahl an innere Knoten aufweisen und somit beispielsweise bei Datensatz 3 nach wenigen Iterationsschritten die Affinität gleichbleibt. Es sei darauf hingewiesen, dass durch diese Vorgehensweise nicht sichergestellt werden kann die besten Parameterwerte gefunden zu haben, da durch den Zufallsprozess im CSA die Affinitätsgrafiken bei jeder neuerlichen Berechnung leicht unterschiedlich ausfallen. Die Parameterwerte können mitunter mit jedem Optimierungsproblem stark variieren und müssen für jede Approximation neu angepasst werden. Sind die Parameterwerte \bar{u} der Datenpunkte nicht bekannt, so ist bereits nach wenigen Iterationsschritten ein minimaler Affinitätswert erreicht (Abb. 4.22 B). Der Grund dafür liegt offensichtlich in der unzureichenden Parametrisierung der Datenpunkte. Wie sich in den folgenden Absätzen noch zeigen wird hat dies negative Einflüsse auf die Approximationsgenauigkeit.

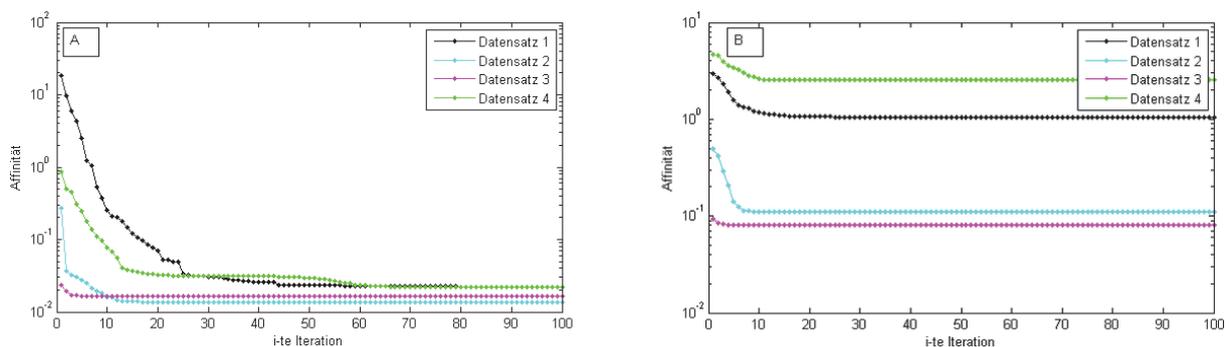


Abb. 4.22: Affinitätswert des besten Knotenvektors pro Generation (Iterationsschritt) für jene vier Datensätze aus Kapitel 3.1 bei Verwendung der Parameterwerte aus Tab. 4.5 Parameterwerte \bar{u} der Datenpunkte A) bekannt; B) unbekannt

Folgende Muster konnten bezüglich der einzelnen Parameterwerte des CSA beobachtet werden. Sie sollen ein allgemeines Verständnis darüber geben wie sich die einzelnen Parameter auf das Resultat auswirken. Eine kurze Zusammenfassung zu der Bedeutung der einzelnen Parameterwerte ist in Tab. 4.5 angegeben.

N ... Eine Erhöhung von der Populationsgröße N führt zu längeren Berechnungszeiten, da sich gleichzeitig auch die Anzahl an Klone pro Generation erhöht, sollte keine Anpassung über den Parameter α erfolgen. Im Verhältnis zur benötigten Berechnungszeit steigt die Chance ein globales Optimum bei Erhöhung der Populationsgröße zu finden nur sehr gering an. Es hat sich gezeigt, dass es besser ist die Populationsgröße klein zu belassen dafür jedoch die Anzahl an Generationen zu erhöhen. Dadurch wird die Durchmischung mit neuen Knotenvektoren gefördert und der durchsuchte Lösungsraum vergrößert. Sobald sich jedoch

die Anzahl an gesuchten inneren Knoten erhöht und somit die Modellkomplexität ansteigt, muss die Populationsgröße bzw. die Anzahl an Klone erhöht werden. Dies soll anhand eines kleinen logischen Beispiels erklärt werden.

Geht man davon aus, dass lediglich ein innerer Knoten innerhalb einer Generation (eines Iterationsschrittes) zu mutieren sei, so wird dieser Knoten 228-mal mutiert, sollten die Parameterwerte aus Tab. 4.5 für den CSA benutzt werden. Muss jedoch ein Knotenvektor mit zwei inneren Knoten mutiert werden und ändern sich die Parameterwerte des CSA nicht, so wird jeder einzelne innere Knoten im Idealfall nur 114-mal mutiert. Dadurch dass sich die einzelnen Knoten weniger oft verändern braucht es im Allgemeinen mehr Iterationsschritte um ein Optimum zu finden. Dies ist auch anhand des Affinitätsplots (Abb. 4.22) erkennbar. Der CSA braucht beim Datensatz 1, welcher gleichzeitig mit 9 inneren Knoten das komplexeste Modell ist, im Gegensatz zu den restlichen Datensätzen die meisten Iterationsschritte um einen optimalen Knotenvektor zu finden.

d ... Die Durchmischung mit neuen zufällig erzeugten Knotenvektoren in jeder Generation ist wichtig um nicht nur einen lokalen, sondern einen globalen optimalen Knotenvektor zu bekommen. Wird *d* zu groß angesetzt so besteht die Gefahr, dass die Affinitätswerte nur zögerlich mit zunehmender Iterationsanzahl konvergieren. Eine zu kleine Anzahl an neuen Knotenvektoren kann dazu beitragen, dass der CSA für längere Zeit (über mehrere Generationen hinweg) an einem lokalen optimalen Knotenvektor festhält.

n ... Die Anzahl an Knotenvektoren, welche entsprechend den Affinitätswerten in einer Generation geklont werden, sollte im Bereich von 10% bis 30% der Populationsgröße *N* liegen. Eine größere Anzahl an zu klonenden Knotenvektoren steigert nur den Berechnungsaufwand nicht jedoch den Erfolg der Methode

a ... Der Multiplikationsfaktor regelt gemeinsam mit der Populationsgröße die Anzahl an Klone.

N_c ... Die Anzahl an zu mutierenden Klone in jeder Generation sollte dann erhöht werden, wenn die Anzahl an zu optimierenden inneren Knoten ansteigt. Der Grund liegt darin, dass sich beim Mutationsschritt lediglich ein zufällig ausgewählter Knotenwert pro Klon verändert. Hat ein Knotenvektor nur zwei innere Knoten, so werden die einzelnen Knoten deutlich häufiger mutiert als etwa zehn innere Knoten eines anderen Knotenvektors, sollte *N_c* gleich bleiben.

N_e ... Dieser Parameter ist schwierig zu handhaben. Man möchte neben einer schnellen Konvergenz auch unnötige Rechenoperationen vermeiden. Der Parameter *N_e*, welcher angibt wie viele nicht mutierte Knotenvektoren einer Generation in die nächste Generation mitgenommen werden sollen, kann eine schnellere Konvergenz ermöglichen, wurden

geeignete Knotenvektoren ausgewählt. Genauso möglich ist es, dass er keinen Einfluss auf die Berechnungen hat und lediglich Rechenressourcen verschlingt.

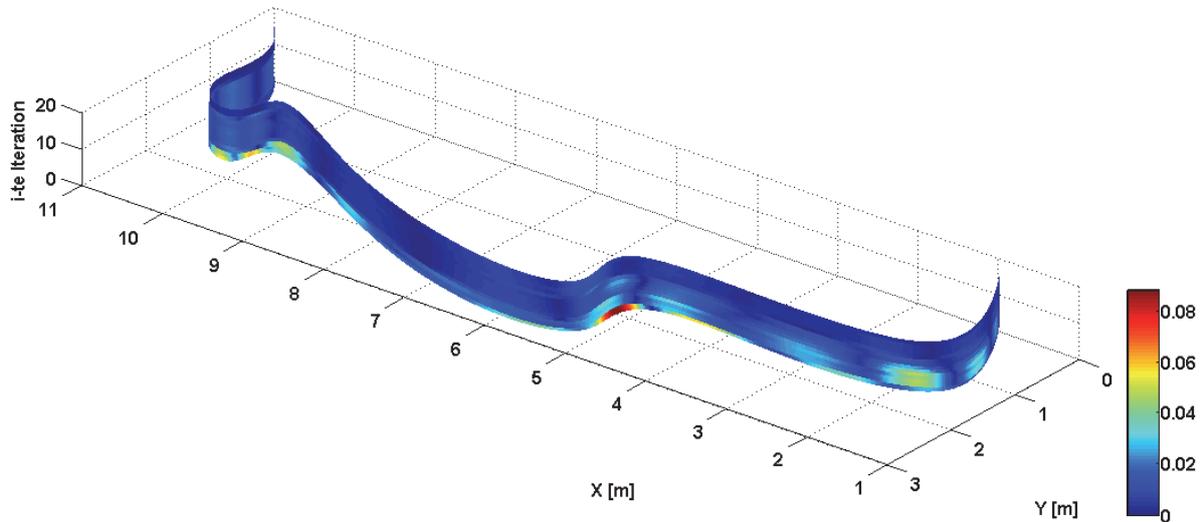
N_{iter} ... Sobald die Anzahl an zu optimierenden inneren Knoten ansteigt, muss im Allgemeinen auch länger iteriert werden, um einen globalen optimalen Knotenvektor zu finden, wenn die restlichen Parameterwerte gleichbleiben. Anhand des Affinitätsplots (Abb. 4.22) ist ersichtlich ab wann der CSA einen solchen optimalen Zustand erreicht hat. Bleibt nämlich die Affinität über mehrere Iterationsschritte (=Generationen) hinweg gleich, so kann der CSA keinen optimaleren Knoten mehr auffinden.

Zwischen den einzelnen Parametern konnten keine eindeutigen Verhältniswerte festgestellt werden. Für die nachfolgenden Grafiken werden im CSA die Parameterwerte aus Tab. 4.5 benutzt.

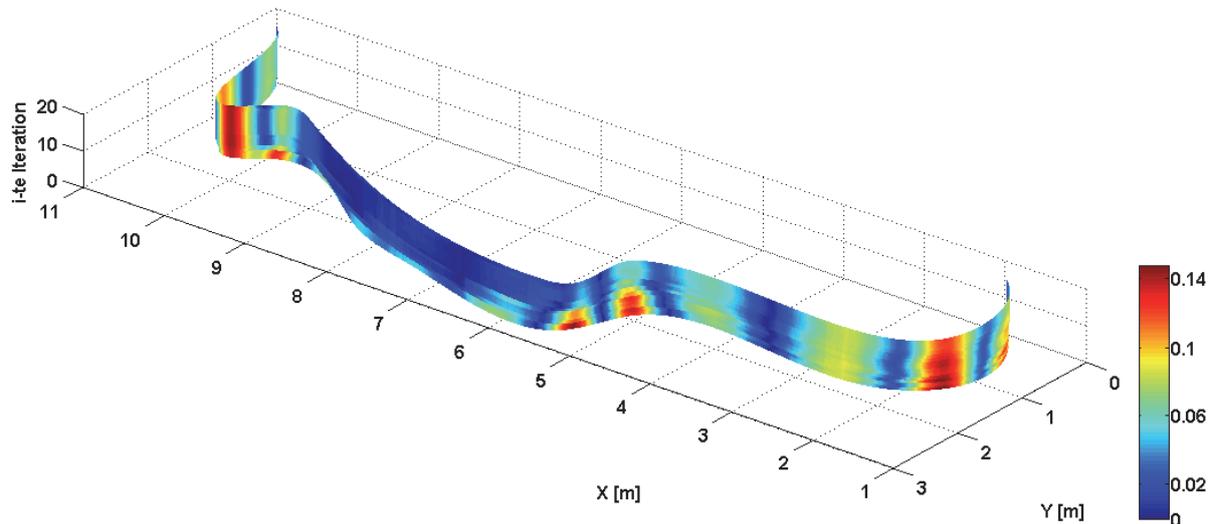
Tab. 4.5: Benutzte Parameterwerte für CSA bei simulierten Datensätzen

Parameter	Bedeutung	verwendeter Wert
N	Anzahl an Knotenvektoren pro Generation	120
d	Anzahl an Knotenvektoren welche in jeder Generation durch neue Knotenvektoren ersetzt werden	15
n	Anzahl an Knotenvektoren welche in einer Generation geklont werden	10
a	Multiplikationsfaktor in (4.13)	0.1
N_c	Anzahl an Klone pro Generation	228
N_e	Anzahl an Knotenvektoren welche in die nächste Generation mitgenommen werden	15
N_{iter}	Anzahl an Generationen (Iterationsschritte)	100

Für den Datensatz 4 (entspricht Abb. 3.1 D) zeigt sich in Abb. 4.22, dass bereits nach etwa 15 Iterationsschritten ein optimaler Knotenvektor und somit eine optimale Approximation der Daten erreicht wird, da sich der entsprechende Affinitätswert, sowohl bei bekanntem als auch bei unbekanntem Parametervektor \bar{u} der Datenpunkte, nicht mehr stark ändert. Dargestellt ist dies in Abb. 4.23, in welcher die Parameterwerte der Datenpunkte bei den Berechnungen bekannt sind. In jeder Generation wird mit Hilfe des besten Knotenvektors eine approximierende B-Spline Kurve berechnet und die Abweichung der Kurve zu den entsprechenden Datenpunkten bestimmt, wobei die Zugehörigkeit über die Parameter der Datenpunkte eindeutig ist. Die Abweichung entspricht dabei der euklidischen Distanz und ist in der Abbildung farblich kodiert dargestellt. Nach 100 Iterationen beträgt die maximale Abweichung 0.017m, was in etwa der dreifachen Standardabweichung des Rauschens entspricht (siehe Tab. 4.6).



**Abb. 4.23: Entwicklung der Kurvenapproximation bei 20 Iterationsschritten;
Parameterwerte \bar{u} der Datenpunkte bekannt;
Abweichung in [m] der Ist- zur Soll-Kurve farblich kodiert**



**Abb. 4.24: Entwicklung der Kurvenapproximation bei 20 Iterationsschritten;
Parameterwerte \bar{u} der Datenpunkte unbekannt;
Abweichung in [m] der Ist- zur Soll-Kurve farblich kodiert**

Müssen hingegen die Parameterwerte \bar{u} der Datenpunkte mit Hilfe der *centripetal* Methode ermittelt werden, so kann zwar der CSA einen (im Sinne der Affinität) optimalen Knotenvektor liefern, jedoch ist die Approximation der Daten in den meisten Fällen nicht zufriedenstellend (siehe Abb. 4.27). Für den Datensatz 4 (Abb. 4.24) vergrößert sich daher die maximale Abweichung auf 0,122m (siehe Tab. 4.8). Nach 15 Iterationen verändert sich bei Abb. 4.23 und Abb. 4.24 die Form und somit auch die Abweichung der Kurve zu den gegebenen Daten nur noch minimal, was auch anhand des Affinitätsplots erkennbar ist, da sich die Affinitätswerte bei den nächsten Iterationsschritten nicht mehr stark ändern (Abb. 4.22 B).

Mit Hilfe des CSA können für die einzelnen Datensätze die originalen Knotenpositionen der ursprünglichen B-Spline Kurven weitestgehend wieder ermittelt werden (vergleiche Abb. 4.30), wobei bei den Berechnungen die Parameterwerte \bar{u} der Datenpunkte bekannt sind. Gemeinsam mit dem besten Knotenvektor nach der 100. Iteration konnte die Datenmenge optimal approximiert werden (siehe Tab. 4.6 bzw. Abb. 4.25). Für die Anzahl an inneren Knoten wird vorerst das vorhandene a-priori Wissen benutzt.

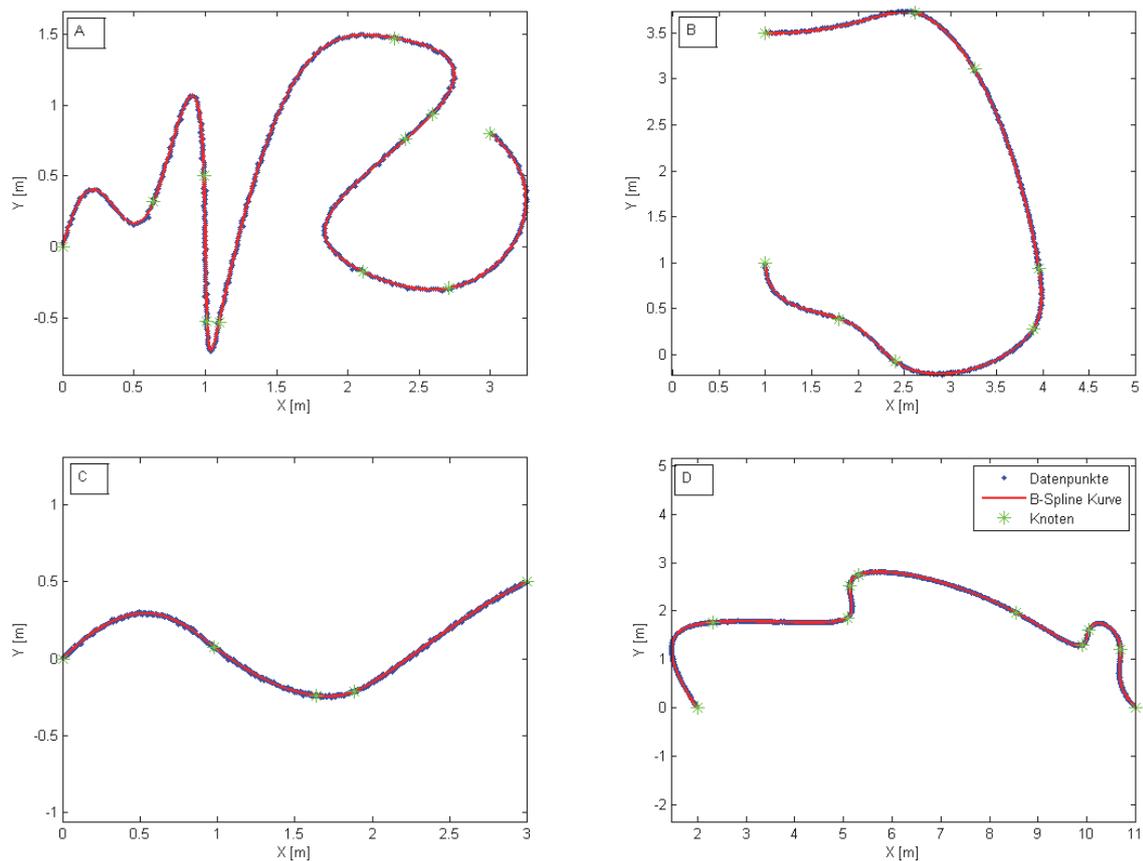


Abb. 4.25: Parameterwerte \bar{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve bekannt
A) 9 innere Knoten B) 6 innere Knoten
C) 3 innere Knoten D) 8 innere Knoten

Die maximale bzw. mittlere Abweichung für alle vier Kurven entspricht in etwa der dreifachen bzw. einfachen Standardabweichung des an die Daten angebrachten Rauschens.

Tab. 4.6: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierender B-Spline Kurve bei bekannten Parameterwerten \bar{u} der Datenpunkte

Abb. 4.25	A	B	C	D
Anzahl an innere Knoten	9	6	3	8
max. Abweichung [m]	0,016	0,014	0,016	0,017
mittlere Abweichung [m]	0,004	0,003	0,003	0,004

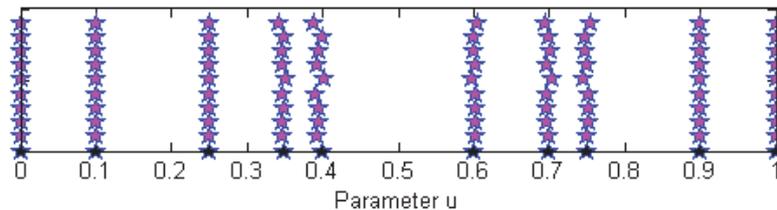
Die Datenpunkte jeder Kurve werden dabei unter Verwendung des CSA 10-mal approximiert. In Abb. 4.25 wird aus den 10 Durchführungen jener Knotenvektor für die Approximation benutzt, welcher den besten Affinitätswert aufweist. Da es sich beim CSA um keine deterministische Methode handelt, sondern die Knotengenerierungen und die Mutationen zufällig ablaufen, fallen die Ergebnisse bei jeder Realisierung leicht unterschiedlich aus. Die Resultate sind dabei eng mit den benutzten Parameterwerten $\{N, d, n, a, N_c, N_e, N_{iter}\}$ gekoppelt. Werden diese falsch angenommen so fallen die einzelnen Knotenvektoren deutlich unterschiedlich aus und eine einheitliche Lösung wird nicht erreicht. Ausgehend von den 10 Realisierungen wird in der Tab. 4.7 für jeden Datensatz der minimale, maximale und mittlere RSS- bzw. RMSE-Wert (Formeln (4.11) & (4.12)) angegeben, um zu zeigen, wie robust der CSA bei mehrfacher Wiederholung der Approximation ist. Der Algorithmus hat lediglich beim Datensatz 1 in einem von zehn Fällen einen unbrauchbaren Knotenvektor geliefert, wie anhand des maximalen RSS- und RMSE-Wertes bzw. anhand der maximalen und mittleren Abweichung (Formeln (4.1) & (4.2)) zu sehen ist. Dies ist dadurch erklärbar, dass die Parameterwerte des CSA anhand dieses Datensatzes festgelegt wurden, da dieser mit 9 inneren Knoten gleichzeitig der Komplexeste ist. In einem von zehn Durchläufen reicht die Anzahl an Iterationen nicht aus, da das Problem beseitigt werden kann, indem N_{iter} in Tab. 4.5 auf 150 erhöht wird.

Tab. 4.7: Fehlerwerte bei zehnmaliger Approximation der einzelnen Datensätze durch CSA; Parameterwert \bar{u} der Datenpunkte bekannt

		max Abweichung [m]	mittlere Abweichung [m]	RSS [m ²]	RMSE [m]
Datensatz 1	min	0,016	0,004	0,019	0,005
	max	0,092	0,012	0,665	0,030
	mean	0,032	0,006	0,106	0,009
Datensatz 2	min	0,014	0,003	0,013	0,005
	max	0,014	0,004	0,013	0,005
	mean	0,014	0,004	0,013	0,005
Datensatz 3	min	0,016	0,003	0,017	0,005
	max	0,016	0,04	0,017	0,005
	mean	0,016	0,04	0,017	0,005
Datensatz 4	min	0,017	0,004	0,022	0,005
	max	0,017	0,004	0,023	0,005
	mean	0,017	0,004	0,022	0,005

Bei den restlichen Datensätzen kommt es beim Auffinden eines Knotenvektors zu keinen Problemen. Die Lage der inneren Knoten kann durch den CSA ebenfalls wieder aufgefunden werden, wie anhand des Datensatzes 4 in Abb. 4.26 zu sehen ist. Anhand der zusätzlichen Tabelle in der Abb. 4.26 kann die Schwankung der einzelnen Knoten im Parameterraum

betrachtet werden. Die inneren Knoten variieren bei den Wiederholungen dabei nur minimal um die tatsächliche Knotenposition. Diese minimalen Variationen haben, wie aus der Tab. 4.7 hervorgeht, auf die Ergebnisse keinen nennenswerten Einfluss. Anscheinend haben Knoten ein flexibles Intervall innerhalb dessen diese das Ergebnis bzw. die Approximationsgenauigkeit nicht negativ beeinflussen. Piegł und Tiller (2000) nutzen diese Eigenschaft bei der Approximation von Punktwolken aus.



min	0.00	0.10	0.25	0.34	0.39	0.60	0.69	0.75	0.90	1.00
max	0.00	0.10	0.25	0.35	0.40	0.61	0.70	0.75	0.90	1.00

Abb. 4.26: Ermittelte Knotenpositionen des Datensatzes 4 bei 10-maliger Wiederholung der Approximation und Schwankung der einzelnen Knoten im Parameterraum (Tabelle)
Parameterwerte \bar{u} der Datenpunkte bekannt
 (schwarze Sterne \triangleq ursprüngliche Knotenpositionen)

Sind hingegen die Parameterwerte der Datenpunkte nicht bekannt so fällt die Approximation der Daten, wie bei den anderen Methoden erkennbar war, schlechter aus (Abb. 4.27). Besonders bei Kurve A und D, jene Kurven mit starken Krümmungsänderungen, fällt die maximale Abweichung groß aus (Tab. 4.8), da, wie bereits erwähnt wurde, die mit Hilfe der *centripetal* Methode berechneten Parameterwerte \bar{u} deutlich von den ursprünglichen Werten abweichen. Die größten Abweichungen sind an jenen Stellen mit den größten Krümmungen zu finden. Selbst eine völlig freie Knotenplatzierung, welche durch den CSA gegeben ist, kann letztendlich die fehlerhafte Parametrisierung der Datenpunkte nicht kompensieren. Vor allem bei Kurvenverläufen mit starker Krümmung scheint eine korrekte Parametrisierung wichtig zu sein.

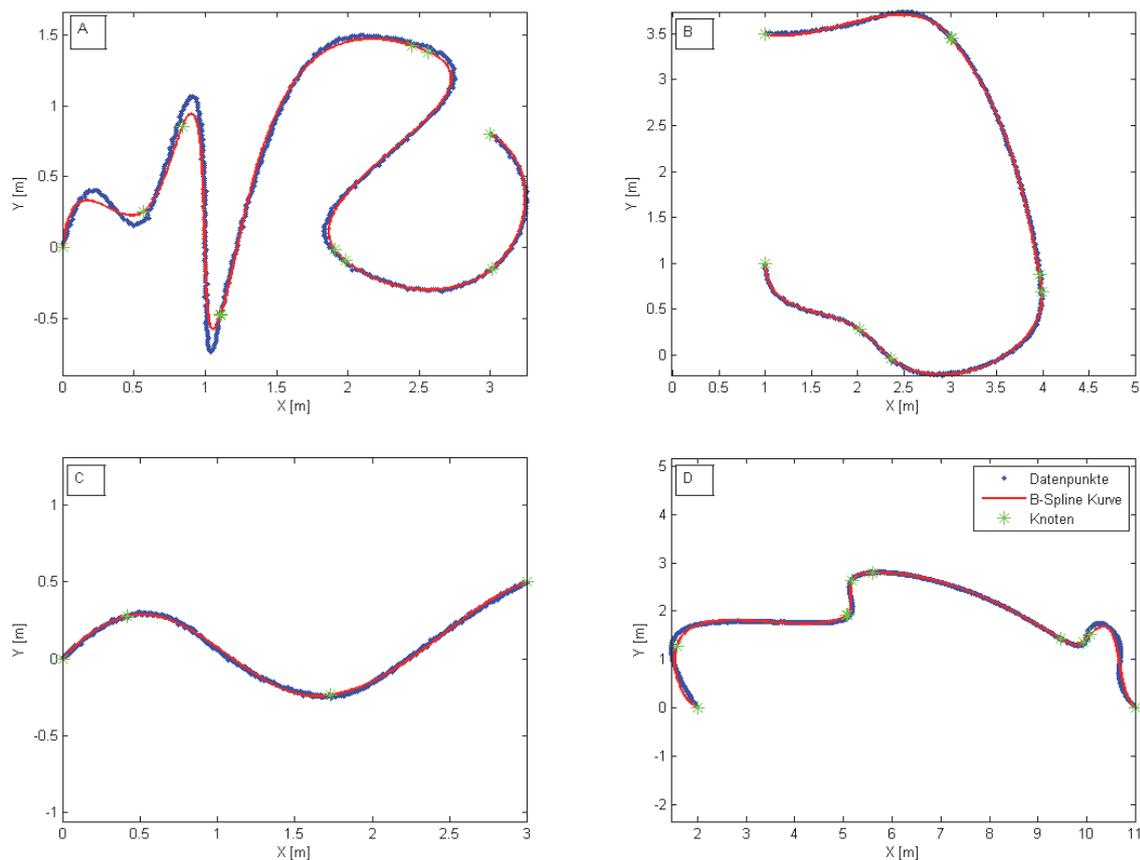


Abb. 4.27: Parameterwerte \bar{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve unbekannt

A) 9 innere Knoten B) 6 innere Knoten

C) 3 innere Knoten D) 8 innere Knoten

Tab. 4.8: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierender B-Spline Kurve bei unbekanntem Parameterwerten \bar{u} der Datenpunkte

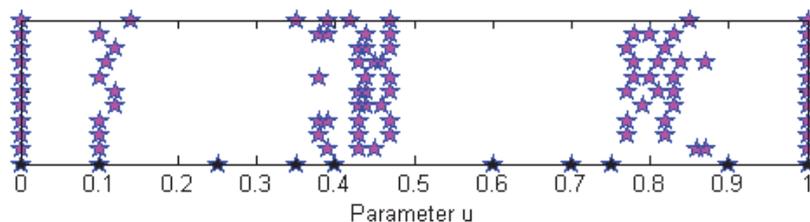
Abb. 4.27	A	B	C	D
Anzahl an innere Knoten	9	6	3	8
max. Abweichung [m]	0,142	0,031	0,026	0,122
mittlere Abweichung [m]	0,018	0,009	0,008	0,027

Im Fall von unbekanntem Parameterwerten der Datenpunkte wurde ebenfalls jeder Datensatz unter Verwendung des CSA 10-mal approximiert um zu sehen wie robust dieser bei unbekannter Parametrisierung ist. Anhand der Tab. 4.9 ist erkennbar, dass in diesem Fall der CSA nicht mehr so robust ist, wie wenn die Parameterwerte der Datenpunkte bekannt sind. Die einzelnen Fehlerwerte schwanken zwar in einem gewissen Bereich, diese sind jedoch nach oben und unten hin beschränkt wenn die Parameterwerte des CSA ausreichend gut gewählt werden. Die Knotenpositionen variieren hingegen deutlich stärker (Abb. 4.28). Es hat sich gezeigt, dass für deutlich unterschiedliche Knotenpositionen vergleichbare Fehlermaße auftreten können. Dies lässt darauf schließen, dass das Auffinden eines globalen eindeutigen Knotenvektors deutlich erschwert wird, sollten die Parameterwerte der

Datenpunkte nicht zufriedenstellend sein. Anscheinend dürften also mehrere unterschiedliche Knotenvektoren einen ähnlich guten Affinitätswert liefern. Daraufhin ist es dem CSA nicht mehr möglich einen eindeutigen Knotenvektor zu bestimmen. Dieser Effekt ist umso ausgeprägter, je stärker die Krümmung der Kurve bzw. je schlechter die Parametrisierung der Datenpunkte ausfällt.

Tab. 4.9: Fehlerwerte bei zehnmaliger Approximation der einzelnen Datensätze durch CSA; Parameterwert \bar{u} der Datenpunkte unbekannt

		max Abweichung [m]	mittlere Abweichung [m]	RSS [m ²]	RMSE [m]
Datensatz 1	min	0,142	0,017	1,038	0,036
	max	0,166	0,020	1,240	0,041
	mean	0,152	0,018	1,108	0,038
Datensatz 2	min	0,031	0,009	0,111	0,014
	max	0,032	0,009	0,112	0,014
	mean	0,032	0,009	0,112	0,014
Datensatz 3	min	0,027	0,008	0,081	0,011
	max	0,027	0,008	0,081	0,011
	mean	0,027	0,008	0,081	0,011
Datensatz 4	min	0,122	0,026	1,885	0,045
	max	0,139	0,028	2,546	0,053
	mean	0,129	0,027	2,352	0,051



min	0.00	0.10	0.12	0.35	0.39	0.42	0.47	0.79	0.82	1.00
max	0.00	0.14	0.43	0.45	0.46	0.78	0.86	0.87	0.87	1.00

Abb. 4.28: Ermittelte Knotenpositionen des Datensatzes 4 bei 10-maliger Wiederholung der Approximation und Schwankung der einzelnen Knoten im Parameterraum (Tabelle)
Parameterwerte \bar{u} der Datenpunkte unbekannt
(schwarze Sterne \triangleq ursprüngliche Knotenpositionen)

In Abb. 4.29 wird die Knotenverteilung mit dem besten Affinitätswert pro Kurve aufgezeigt. Sind die Datenparameter bekannt, so können die originalen Knotenpositionen, welche einem globalen Optimum entsprechen, aufgefunden werden. Bei unbekanntem Datenparameterwert können nur vereinzelt ursprüngliche Knotenpositionen ausgemacht werden.

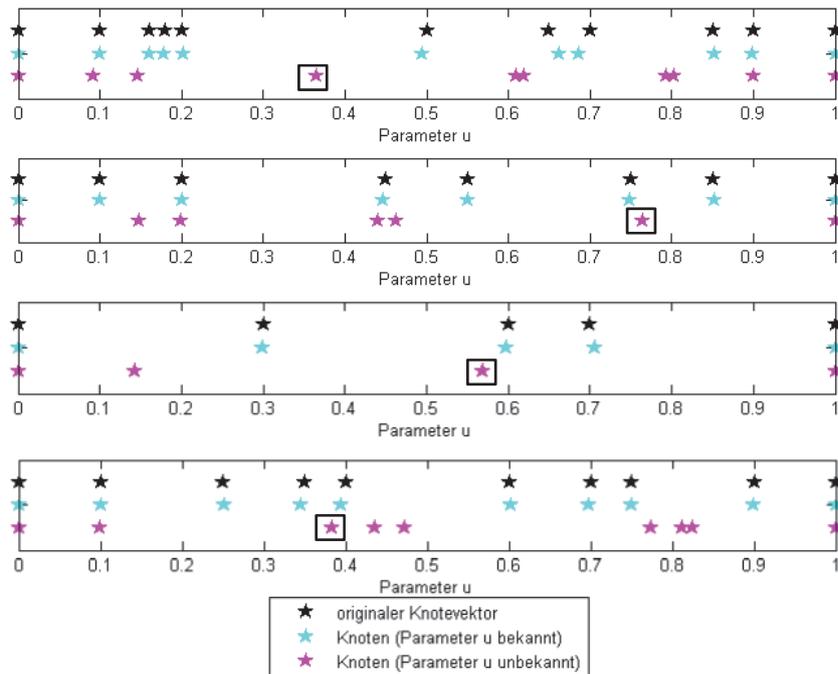


Abb. 4.29: Knotenverteilung für approximierende B-Spline Kurve A, Kurve B, Kurve C und Kurve D der Abb. 4.25 & Abb. 4.27 (Knoten mit Vielfachheit 2 sind schwarz umrahmt)

Da der CSA bei bekannten Datenparameterwerten die originalen Knotenpositionen wieder auffinden kann, ist es weiters nicht verwunderlich, dass bei der Approximation der Datenpunkte lediglich das Rauschen im Residuenplot zu sehen ist (Abb. 4.30). Bei allen vier Datensätzen konnten keine systematischen Effekte in den zugehörigen Grafiken beobachtet werden. Im Gegensatz zu den beiden anderen Methoden für die Knotenermittlung (DOM und GDR) ist für die Modellierung der Datenpunkte ein deutlich einfacheres Modell notwendig. Dies ist auf das Wiederauffinden der ursprünglichen Knotenpositionen zurückzuführen.

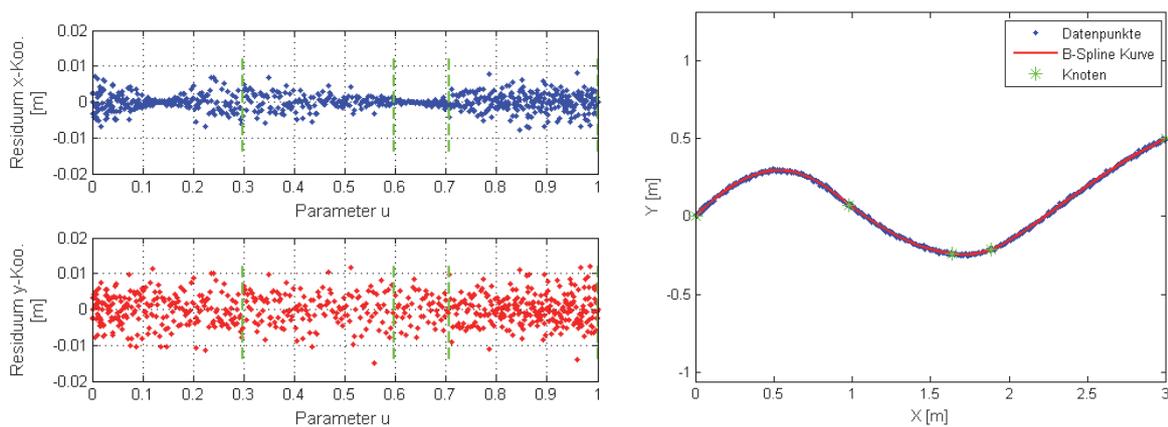


Abb. 4.30: Auftretende Residuen [m] in X- bzw. Y-Richtung mit keinem erkennbarem systematischem Effekt bei 3 inneren Knoten

Die noch unbekanntenen Kontrollpunkte können im linearen Ausgleichsproblem einfach und zufriedenstellend ermittelt werden (siehe Kapitel 2.5.3). Ein Problem besteht jedoch noch darin, dass a-priori die Komplexität und somit die Anzahl der Knoten der B-Spline

Kurve unbekannt ist. Wird etwa ein zu komplexes Modell benutzt, so besteht die Gefahr des Overfittings und umgekehrt die Gefahr des Underfittings. Mit Hilfe von Informationskriterien soll im nachfolgenden Kapitel 5 eine geeignete Modellselektion für B-Spline Kurven erfolgen.

5. Modellselektion mit Hilfe von Informationskriterien

Bei den bisherigen Methoden der Knotenschätzung war die Anzahl der zu bestimmenden Knoten und Kontrollpunkte, auf Grund der benutzten B-Spline Kurven, bekannt. Im Allgemeinen ist jedoch bei der Kurven- und Flächenschätzung die Anzahl an zu ermittelnden Parametern (Kontrollpunkte und Knoten) a-priori nicht bekannt da keine Informationen über das tatsächliche funktionale Modell vorliegen. Ziel der Modellselektion ist es aus einer Menge an möglichen Modellen ein „bestes“ Modell zu finden, welches eine optimale Modellkomplexität aufweist. Dabei kommt das Prinzip der Sparsamkeit zur Anwendung. Ein optimales Modell soll die Anzahl an benötigten Parametern geringhalten, gleichzeitig aber noch die Daten so gut wie möglich approximieren. Diese Forderung ist gleichbedeutend mit einem ausbalancieren des Varianz/Bias Trade-Off und kann im Zuge der Kurvenapproximation wie folgt beschrieben werden (Burnham & Anderson, 2002).

Wird die Anzahl an zu bestimmenden Parametern erhöht, so kann das Modell (B-Spline Kurve) diese Freiheiten nutzen und die gemessenen Daten besser approximieren, wobei der Bias abnimmt, die Komplexität des Modells hingegen zunimmt. Ein Minimum des Bias ist dann erreicht, wenn das Modell die Daten interpoliert. Dies entspricht einem „Overfitten“, da dabei auch das Rauschen, welches in den Daten steckt, mitmodelliert wird und somit gleichzeitig die Varianz ein Maximum aufweist (Harmening & Neuner, 2014).

Für diese Problematik stehen viele sogenannte Informationskriterien zur Verfügung, welche allesamt auf der Maximum-Likelihood Theorie aufbauen. Wie der Name schon verrät, werden bei der Maximum-Likelihood Schätzung jene Modellparameter gesucht, welche die Likelihood-Funktion maximieren. Die Likelihood-Funktion ist dabei die bedingte Wahrscheinlichkeit $L(\theta|\mathbf{D})$ der Beobachtungen \mathbf{D} in Abhängigkeit der unbekanntem Modellparameter θ (Koch, 1997).

Im Nachfolgendem werden die Grundlagen des Akaike- und des Bayesianischen Informationskriterium vorgestellt, jene Methoden welche im späteren Verlauf für die Bestimmung der Anzahl an inneren Knoten benutzt werden.

5.1. Das Akaike Informationskriterium (AIC)

Die Kullback-Leibler-Funktion $I(f, g)$ bildet die Grundlage für das Akaike Informationskriterium (AIC) und wird im stetigen Fall für zwei Dichteverteilungen f und g wie folgt berechnet.

$$I(f, g) = \int f(x) \log\left(\frac{f(x)}{g(x|\theta)}\right) dx \quad (5.1)$$

Die Funktion kann als Informationsverlust oder auch als Distanz zwischen den beiden Verteilungen interpretiert werden, welche anfällt, wenn versucht wird die wahre Verteilung $f(x)$ durch die Verteilung $g(x|\theta)$ zu modellieren, wobei θ jene Parameter darstellt welche für die approximierende Verteilung benutzt werden (Burnham & Anderson, 2002, S.50 ff.). Das AIC übernimmt nun die Aufgabe, jenes Modell zu finden, welches den Informationsverlust geringhält und somit die Kullback-Leibler-Funktion minimiert. Da zumeist die wahre Verteilung $f(x)$ unbekannt ist muss ein Übergang auf relative Distanzen erfolgen. Dadurch kann nicht mehr festgestellt werden wie gut ein Modell das „wahre“ Modell approximiert, sondern lediglich welches Modell aus der Menge an möglichen Modellen das „Beste“ im Sinne des Informationsverlustes ist. Akaike (1973) konnte zeigen, dass das Maximum der logarithmierten Likelihood $\log(L(\hat{\theta}|\mathbf{D}))$ für die geschätzten Parameterwerte $\hat{\theta}$ ein verzerrter Schätzer für die relative Kullback-Leibler-Distanz (KL) ist, wobei der zugehörige Bias approximativ gleich der Anzahl an zu schätzenden Parametern K für das jeweilige Modell ist. Der Schätzwert für die KL-Distanz lautet somit:

$$\hat{E}(KL) = \log(L(\hat{\theta}|\mathbf{D})) - K. \quad (5.2)$$

Wird die Gleichung (5.2) mit 2 multipliziert so ergibt sich die bekannte Form des AICs.

$$AIC = -2 \log(L(\hat{\theta}|\mathbf{D})) + 2K \quad (5.3)$$

Dabei ist zu erkennen, dass der erste Term der Gleichung (5.3) kleiner wird, sobald mehr Parameter bei der Schätzung zugelassen werden, da sich das Modell den Daten besser anpassen kann. Gleichzeitig nimmt jedoch die Unsicherheit bei der Schätzung der KL-Distanz zu. Deshalb wird die Hinzunahme von weiteren Parametern durch den zweiten Term ($2K$) bestraft. Durch diese Vorgehensweise kann das Ausbalancieren des Varianz/Bias Trade-Offs erreicht und ein „Over- bzw. Underfitting“ vermieden werden. Dabei ist zu beachten, dass ein einzelner AIC-Wert für sich nicht interpretierbar ist, da erst durch Differenzbildung über alle möglichen Kandidaten eines Modellsets ein „bestes“ Modell gefunden werden kann. Deshalb ist nicht der absolute, sondern der relative AIC-Wert, insbesondere die AIC-Differenzen Δ_i , für die Modellselektion wichtig.

$$\Delta_i = AIC_i - AIC_{min} \quad (5.4)$$

Diese Differenzen sind einfach zu interpretieren und können auch für die Sortierung der Modelle benutzt werden. Jenes Modell mit dem niedrigsten AIC-Wert ($\Delta_i \equiv \Delta_{min} \equiv 0$) wird als das „beste“ Modell angesehen. Sind hingegen alle Modelle von schlechter Qualität so findet das AIC trotzdem ein relatives „bestes“ Modell. Dieses Modell ist jedoch absolut gesehen noch immer ein sehr schlechtes. Es sollte daher eine große Anstrengung auf die Auswahl eines geeigneten Modellsets gelegt werden, um mit einer hohen Wahrscheinlichkeit ein gutes Modell im Modellset zu haben (Burnham & Anderson, 2002, S.60 ff.).

Eine Sonderstellung bei der Ermittlung der AIC-Werte nehmen Modelle ein, bei welchen anstatt des Maximum-Likelihood Schätzers die Kleinste Quadrate Schätzung, unter der Annahme normalverteilter Residuen, verwendet wird. Dann nämlich berechnet sich das AIC als einfache Funktion in Abhängigkeit der Verbesserungsquadratsumme zu

$$AIC = N \ln(\hat{\sigma}_{ML}^2) + 2K \quad (5.5)$$

$$\text{mit } \hat{\sigma}_{ML}^2 = \frac{1}{N} \sum_{i=1}^N (\hat{\varepsilon}_i)^2 \quad \text{und} \quad \hat{\varepsilon}_i = \hat{y}_i - y_i. \quad (5.6)$$

N ist dabei die Anzahl an Beobachtungen und $\hat{\sigma}_{ML}^2$ der Maximum-Likelihood Schätzer der Varianz der Gewichtseinheit. Bei einer kleinen Stichprobe sollte eine überarbeitete Version des AICs benutzt werden, welche vom AIC ausgeht und einen weiteren Bias-Korrekturterm aufweist.

$$AIC_c = AIC + \frac{2K(K+1)}{N-K-1} \quad (5.7)$$

Das AIC_c sollte laut Burnham und Anderson immer dann verwendet werden, wenn das Verhältnis $\frac{N}{K} > 40$ ist, um ein Overfitting (welches dem AIC bei kleinen Stichprobenumfängen ständig innewohnt) zu verhindern, wobei K die Parameteranzahl des komplexesten Modells im Modellset entspricht. Dabei muss bei K auch die Varianz der Gewichtseinheit berücksichtigt werden, woraus sich bei der Vorgabe von mi inneren Knoten folgender Bias-Korrekturterm ergibt:

$$2K = 2(2mi + p + 2). \quad (5.8)$$

Dieser setzt sich aus der Anzahl an inneren Knoten, der Anzahl an Kontrollpunkten und der Varianz der Gewichtseinheit zusammen. Da bei der B-Spline Kurvenapproximation relativ wenig Datenpunkte zur Verfügung stehen und das Verhältnis $\frac{N}{K} > 40$ ist, kann K kein idealer Schätzer für den Bias-Term sein. Deshalb wird anstelle des AIC_s der AIC_c bevorzugt und für die Modellauswahl benutzt (Burnham & Anderson, 2002, S.63 ff.).

Zusammenfassend kann gesagt werden, dass das AIC jenes Modell aus einem Modellset auswählt, welches den KL-Informationsverlust minimiert. Dabei wird die Existenz eines „besten“ Modells im Set angenommen, welches den günstigsten Varianz/Bias Trade-Off aufweist. Das AIC sollte laut Burnham & Anderson dann zur Anwendung kommen, wenn sich mit steigender Parameteranzahl die KL-Distanz wie in Abb. 5.1 verringert, wobei für die 12 Modelle die Annahme

$$g_1 = K_1, \dots, g_{12} = K_{12} \quad \text{und} \quad K_1 < K_2 < \dots < K_{12} \quad (5.9)$$

zutrifft (Burnham & Anderson, 2004).

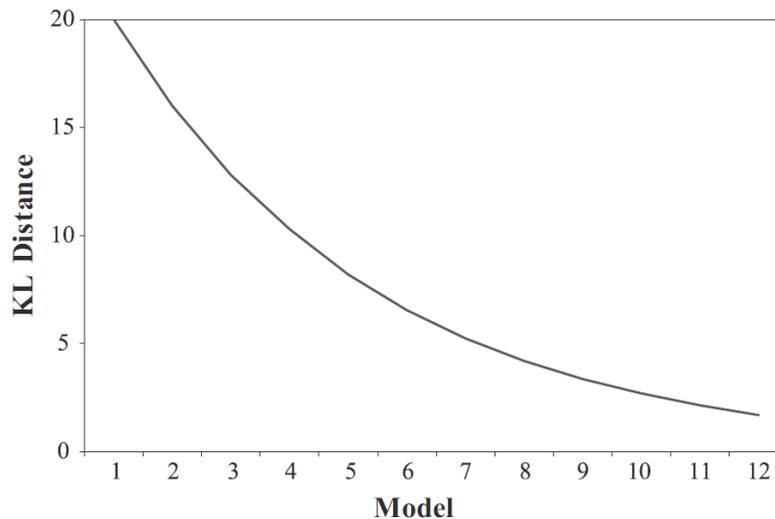


Abb. 5.1: KL-Informationsverlust bei steigender Parameteranzahl
typischer Anwendungsfall des AICs
[Burnham & Anderson, 2004]

5.2. Das Bayesianische Informationskriterium (BIC)

Beim BIC handelt es sich um ein bayesianisch fundiertes Informationskriterium, welches von Gideon Schwarz (1978) entwickelt wurde und wie folgt berechnet wird:

$$BIC = N \ln(\hat{\sigma}_{ML}^2) + \ln(N) K \quad (5.10)$$

$$\text{mit } \hat{\sigma}_{ML}^2 = \frac{1}{N} \sum_{i=1}^N (\hat{\varepsilon}_i)^2 \quad \text{und} \quad \hat{\varepsilon}_i = \hat{y}_i - y_i.$$

Neben der Benennung des Informationskriteriums als BIC sind auch die Bezeichnungen *Schwarz Kriterium SIC* und *Schwarz Bayes Kriterium SBC* geläufig. Der auffälligste Unterschied zwischen den Formeln (5.5) und (5.10) ist die unterschiedliche Gewichtung des Bestrafungsterms (zweiter Term auf der rechten Seite der Gleichung). Beim AIC ist dieser lediglich von der Anzahl der zu schätzenden Parametern K abhängig und entspricht dabei dem Bias, der bei der Schätzung der erwarteten Kullback-Leibler-Information auftritt. Die Stichprobengröße N wird lediglich bei der Likelihood $L(\hat{\theta}|\mathcal{D})$ berücksichtigt, sodass mit steigendem N , Modelle mit zunehmender Parameteranzahl begünstigt werden. Der Bestrafungsterm beim BIC ist hingegen vom Stichprobenumfang N abhängig. Deshalb fällt dieser bereits ab 8 Beobachtungen größer aus als jener beim AIC wie folgendes Beispiel zeigt.

AIC	für $K = 3$	BIC
$2 \cdot 3 = 6$	<	$\ln(8) \cdot 3 \approx 6,24$

Analog zum AIC werden beim BIC jene Modelle mit dem geringsten BIC-Wert bevorzugt. Die Differenzbildung kann wieder für die Sortierung der Modelle benutzt werden.

$$\Delta_i = BIC_i - BIC_{min} \quad (5.11)$$

Es stellt sich dabei die Frage, welches Modell mit Hilfe des BICs im Gegensatz zu dem AIC nun schlussendlich ausgewählt wird. Die Antwort lautet: Es wird das sparsamste Modell (hinsichtlich der Parameteranzahl) aus dem Modellset mit dem gleichzeitig geringsten KL-Informationsverlust ausgewählt. Dazu ist folgende Grundannahme wichtig.

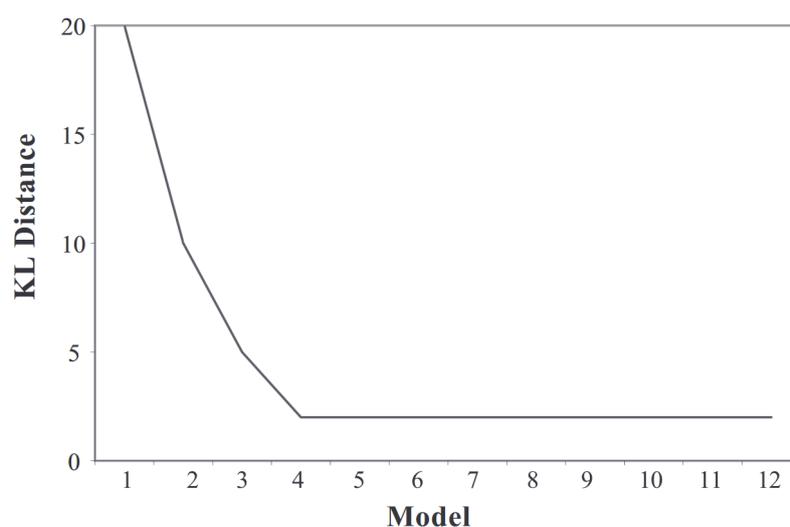
Es gibt insgesamt R Modelle g_i ($i = 1, \dots, R$), welche das Modellset bilden und sich mit Hilfe der Parameteranzahl K hierarchisch ordnen lassen, sodass

$$g_1 = K_1 < g_2 = K_2 < \dots < g_R = K_R \quad (5.12)$$

gilt. Außerdem gilt, dass die zugehörige KL-Distanz mit steigender Parameteranzahl sinkt oder gleich bleibt.

$$I(f, g_1) \geq I(f, g_2) \geq \dots \geq I(f, g_R) \quad (5.13)$$

Nach diesem Szenario würde das BIC das Modell g_t (mit $t = K_i$ und $1 \leq t \leq R$) bevorzugen, welches, wie bereits erwähnt, das sparsamste Modell mit dem zugleich niedrigsten Informationsverlust ist. In Abb. 5.2 ist dies grafisch dargestellt. Bei steigender Modellzahl und somit steigender Parameteranzahl sinkt zuerst die KL-Distanz bis zum Modell 4. Ab dem 5. Modell verringern die zusätzlichen Parameter den Informationsverlust nicht mehr, wodurch nach dem BIC das Modell 4 als das beste Modell angesehen wird. Da g_t nicht identisch mit f ist und somit die KL-Distanz $I(f, g)$ nicht null ist wird das Modell 4 auch als „quasi wahre“ Modell bezeichnet. Das BIC selektierte Modell konvergiert bei einem festen Modellset und einer vorhandenen i.i.d. Stichprobe N bei $N \rightarrow \infty$ mit Wahrscheinlichkeit 1 gegen das Modell g_t . Das durch das BIC ausgewählte „quasi wahre“ Modell birgt jedoch bei einem kleinen Stichprobenumfang die Gefahr des Underfittings (Burnham & Anderson, 2004).



**Abb. 5.2: KL-Informationsverlust bei steigender Parameteranzahl
typischer Anwendungsfall des BICs
[Burnham & Anderson, 2004]**

5.3. Ermittlung der Knotenanzahl anhand von Informationskriterien

Die Ermittlung einer passenden Knotenanzahl bei der Bestimmung einer approximierenden B-Spline Kurve ist von entscheidender Bedeutung. Über die Beziehung (2.3) wird, sobald der Grad der B-Spline Kurve und die Anzahl an Knoten a-priori festgelegt wurde, auch gleichzeitig die Anzahl an Kontrollpunkte gesteuert. Im Grundlagenkapitel 2 ist zu sehen, dass bei einer Erhöhung der Parameteranzahl die Datenpunkte im Allgemeinen besser approximiert werden. Ab einer bestimmten Parameteranzahl hingegen bewirken zusätzliche Parameter keine nennenswerte Steigerung der Approximationsgenauigkeit. Ab diesem Zeitpunkt besteht die Gefahr des Overfittings, da durch die zusätzlichen Parameter nicht mehr die eigentliche Form der Kurve sondern das Rauschen mitmodelliert wird. Geeignete Informationskriterien, welche in den letzten beiden Kapiteln 5.1 und 5.2 vorgestellt wurden, sollen ein Overfitting aber auch ein Underfitting verhindern. Anhand der 4 verrauschten Datensätze aus Kapitel 3.1 wird überprüft, ob die ursprüngliche Anzahl an inneren Knoten mit Hilfe der beiden Informationskriterien bei den unterschiedlichen Methoden zur Knotenschätzung gefunden werden können. Bei jeder Methode und jedem Datensatz steht ein Modellset, angefangen von 0 bis 25 inneren Knoten, zur Verfügung. Ausgehend von diesem Modellset wird für jede Knotenanzahl das AIC (bzw. AICc) und das BIC (siehe Formeln (5.5), (5.7) und (5.10)) berechnet und anschließend die Differenzen Δ_i ermittelt, um das „beste“ bzw. das „quasi beste“ Modell zu finden (siehe Tab. 5.2). Ein vollständiges Berechnungsexemplar ist durch die Tab. 5.1 gegeben. Es wird darauf hingewiesen, dass bei allen folgenden Ergebnissen die Parameterwerte der Datenpunkte bekannt sind. Durch diese Vorgehensweise hat eine möglicherweise falsche Parametrisierung keinen Einfluss auf die Modellselektion. Außerdem soll festgestellt werden, ob eine Modellselektion mit Hilfe der DOM- bzw. der GDR-Methode möglich ist, da diese aus Sicht des Berechnungs- und Zeitaufwands im Gegensatz zum CSA deutlich besser sind. Die Grundidee dahinter ist, dass zuerst ein Modell unter Verwendung von Informationskriterien ausgewählt wird und anschließend mit dem CSA ein optimaler Knoten aufgestellt wird. Dadurch könnte wertvolle Zeit eingespart werden.

Die Gesamtanzahl an Daten würde diese Arbeit unübersichtlich machen. Deshalb wird für Demonstrationszwecken lediglich der Datensatz 2 benutzt, wobei für die Knotenschätzung der *Elitist Clonal Selection Algorithm* zur Anwendung kommt und der Grad der B-Spline Kurve bei 3 belassen wird. Die Ergebnisse sind in der Tab. 5.1 ersichtlich, wobei im gesamten Berechnungsprozess die Parameterwerte der Datenpunkte \bar{u} bekannt sind. Bei der Erhöhung der Anzahl an inneren Knoten m_i verringert sich zunächst die maximale bzw. die mittlere Abweichung und auch die RSS- und RMSE-Werte nehmen deutlich ab. Die

Bestrafungsterme bei den einzelnen Informationskriterien können vorerst diese starke Abnahme der Varianz der Gewichtseinheit ($\hat{\sigma}_{ML}^2 = \frac{RSS}{n} = RMSE^2$) nicht ausgleichen. Ab dem Modell mit 6 inneren Knoten bleiben jedoch die Abweichungen und somit die Gewichtseinheit weitestgehend konstant. Dies ist ein Hinweis darauf, dass bei den nachkommenden, komplexeren Modellen das Rauschen der Daten bei der Approximation mitmodelliert wird. Bei den Informationskriterien steigen deshalb bei komplexeren Modellen (B-Spline Kurven mit größerer Anzahl an inneren Knoten) die Bestrafungsterme im Verhältnis gesehen zu den ersten Termen auf der rechten Seite der jeweiligen Gleichungen (vgl. (5.5), (5.7) und (5.10)) stärker an. Somit ist beim Modell mit 6 inneren Knoten bei allen Informationskriterien ein Minimum zu finden (entsprechende Zeile in Tab. 5.1 fett formatiert). Durch die Differenzbildung ist dieser Umstand sofort ersichtlich, da jenes Modell, für welches $\Delta_i \equiv \Delta_{min} \equiv 0$ gilt, als das „beste“ bzw. „quasi beste“ Modell angesehen wird. Der Unterschied zwischen dem AIC- und dem AICc-Wert ist ebenfalls bei den Differenzen erkennbar. Da die Parameteranzahl K relativ gering ist (bei $m_i = 25 \rightarrow K = 55$), fällt der zusätzliche Bestrafungsterm bei $N = 604$ Beobachtungen klein aus und erhöht den AICc-Wert nur minimal. Das BIC bestraft zusätzliche Parameter deutlich stärker als das AIC bzw. das AICc.

Tab. 5.1: Exemplarisches Berechnungsergebnis eines Modellsets (0 bis 25 innere Knoten) für den Datensatz 2

mi	Abweichung [m]		RSS [m ²]	RMSE [m]	AIC	Δ_i (AIC)	AICc	Δ_i (AICc)	BIC	Δ_i (BIC)
	maximale	mittlere								
0	0,331	0,132	40,801	0,260	-1619,70	4818,98	-1619,63	4818,12	-1594,72	4763,81
1	0,237	0,066	8,777	0,121	-2543,79	3894,89	-2543,65	3894,10	-2509,61	3848,92
2	0,147	0,033	2,003	0,058	-3432,15	3006,53	-3431,91	3005,85	-3388,77	2969,75
3	0,130	0,029	1,766	0,054	-3504,30	2934,38	-3503,93	2933,82	-3451,74	2906,79
4	0,074	0,023	1,056	0,042	-3810,76	2627,92	-3810,23	2627,52	-3749,00	2609,53
5	0,021	0,006	0,034	0,007	-5882,80	555,88	-5882,09	555,67	-5811,84	546,69
6	0,014	0,004	0,013	0,005	-6438,68	0,00	-6437,76	0,00	-6358,53	0,00
7	0,014	0,004	0,013	0,005	-6434,87	3,81	-6433,71	4,05	-6345,53	13,00
8	0,014	0,004	0,013	0,005	-6431,49	7,19	-6430,06	7,70	-6332,95	25,57
9	0,014	0,004	0,013	0,005	-6428,17	10,51	-6426,44	11,32	-6320,44	38,09
10	0,014	0,004	0,013	0,005	-6424,77	13,91	-6422,71	15,05	-6307,84	50,69
11	0,014	0,004	0,013	0,005	-6426,07	12,61	-6423,64	14,11	-6299,95	58,58
12	0,014	0,004	0,013	0,005	-6420,41	18,27	-6417,60	20,16	-6285,10	73,43
13	0,014	0,004	0,013	0,005	-6416,84	21,84	-6413,61	24,15	-6272,33	86,20
14	0,014	0,004	0,013	0,005	-6425,04	13,64	-6421,35	16,40	-6271,33	87,20
15	0,015	0,004	0,013	0,005	-6417,22	21,46	-6413,05	24,70	-6254,32	104,21
16	0,015	0,004	0,013	0,005	-6414,41	24,27	-6409,73	28,03	-6242,31	116,21
17	0,014	0,004	0,013	0,005	-6415,23	23,45	-6410,00	27,76	-6233,94	124,59
18	0,016	0,004	0,013	0,005	-6411,08	27,60	-6405,27	32,48	-6220,59	137,93
19	0,014	0,004	0,013	0,005	-6399,88	38,80	-6393,46	44,29	-6200,20	158,33
20	0,015	0,004	0,013	0,005	-6401,32	37,36	-6394,26	43,49	-6192,45	166,08
21	0,016	0,004	0,013	0,005	-6403,33	35,35	-6395,59	42,17	-6185,26	173,27
22	0,016	0,004	0,013	0,005	-6401,56	37,12	-6393,11	44,64	-6174,30	184,23
23	0,015	0,004	0,013	0,005	-6405,97	32,71	-6396,78	40,97	-6169,52	189,01
24	0,016	0,004	0,013	0,005	-6399,10	39,58	-6389,13	48,62	-6153,45	205,08
25	0,014	0,004	0,013	0,005	-6399,09	39,59	-6388,31	49,45	-6144,25	214,28

Alle drei Informationskriterien (AIC, AICc und BIC) können die exakte Anzahl an inneren Knoten der ursprünglichen B-Spline Kurve auffinden. Der RMSE-Wert bei der B-Spline Kurve mit 6 inneren Knoten stimmt dabei mit dem Rauschen ($\sigma = 0.005m$), welches an die Daten angebracht wurde, überein. Die maximale Abweichung zwischen der Kurve und den zu approximierenden Daten entspricht in etwa der dreifachen Standardabweichung des Rauschlevels. Der RMSE-Wert kann gemeinsam mit der maximalen Abweichung bei der Kurvenapproximation benutzt werden, um sicherzustellen, dass im Modellset ein „gutes“ Modell vorhanden ist und das Informationskriterium eben dieses gefunden hat. Stimmt nämlich der dreifache RMSE-Wert nicht in etwa mit der maximalen Abweichung überein ($3 \cdot RMSE \approx \text{max. Abw.}$) so ist davon auszugehen, dass zwar das Informationskriterium innerhalb des Modellsets ein Modell gefunden hat, dieses aber aus absoluter Sicht gesehen die Daten nicht zufriedenstellend approximieren kann.

Es stellt sich die Frage, welches Informationskriterium nun bevorzugt für das Auffinden eines passenden Modells benutzt werden soll. Dazu werden zuerst einmal für die einzelnen Datensätze die KL-Distanzen geplottet (Abb. 5.3). Die einzelnen Modelle können dabei über die Parameteranzahl hierarchisch geordnet werden (vgl. Bedingung (5.12)). Das erste Modell hat demnach null innere Knoten, das zweite Modell einen inneren Knoten usw..

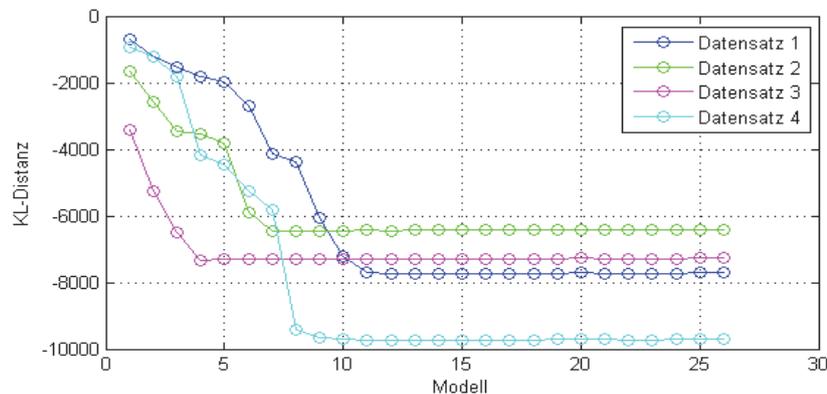


Abb. 5.3: KL-Distanz der einzelnen Datensätze aus Kapitel 3.1; Knotenermittlung mit Hilfe der CSA-Methode

Bei allen Datensätzen sinkt zu Beginn, bei zunehmender Parameteranzahl (Modellkomplexität steigt, vgl. Bedingung (5.13)), die KL-Distanz. Ab einem bestimmten Modell bleibt diese weitestgehend konstant und zusätzliche Parameter verringern den Informationsverlust nicht mehr. Die einzelnen Kurven in Abb. 5.3 ähneln sehr stark jenem Kurvenmodell aus Abb. 5.2. Gemeinsam mit der Tatsache, dass die Bedingungen (5.12) und (5.13) ebenfalls erfüllt werden, wird erwartet, dass das BIC das geeignetste Informationskriterium bei der Modellselektion und somit bei der Bestimmung der Anzahl an inneren Knoten von B-Spline Kurven ist.

Wie sich in Tab. 5.2 zeigt ist diese Erwartung nur in Kombination mit dem *Elitist Clonal Selection Algorithm* (CSA) zutreffend. Die Tabelle zeigt dabei die mit Hilfe der zwei Informationskriterien AICc und BIC ermittelte Anzahl an innere Knoten (entspricht „beste“ und „quasi beste“ Modell) für die unterschiedlichen Datensätze, wobei bei der Knotenermittlung die in Kapitel 4 gezeigten Methoden zur Anwendung kommen. Für jede Methode und jeden Datensatz wurde eine Tab. 5.1 erzeugt und schlussendlich die Resultate in der Tab. 5.2 zusammengefasst.

Tab. 5.2: Zusammenfassung der Ergebnisse der Modellselektion unter der Verwendung von Informationskriterien (AICc & BIC) bei unterschiedlichen Datensätzen und unterschiedlichen Methoden der Knotenermittlung; Parameterwerte der Datenpunkte bei allen Berechnungsschritten bekannt

Datensatz x – tatsächliche Knotenanzahl	ermittelte Anzahl an innere Knoten					
	CSA		DOM		GDR	
	BIC	AICc	BIC	AICc	BIC	AICc
Datensatz 1 - 9 innere Knoten	11	12	23	24	24	25
Datensatz 2 - 6 innere Knoten	6	6	12	14	18	20
Datensatz 3 - 3 innere Knoten	3	3	6	7	9	9
Datensatz 4 - 8 innere Knoten	10	11	19	20	17	17

CSA ... *Elitist Clonal Selection Algorithm*; DOM ... *dominante Punkte*; GDR ... *Gleichverteilung des Rauschverhältnis*

Das BIC selektiert im Gegensatz zum AIC, durch den stärkeren Bestrafungsterm, Modelle mit einer kleineren Anzahl an inneren Knoten, wobei nie weniger Knoten, als für die ursprüngliche B-Spline Kurve benötigt wurden, ausgewählt werden. Das bezüglich der Knotenanzahl beste Ergebnis ist in Tab. 5.2 grün markiert dargestellt. Bei zwei von vier Datensätzen konnte unter Anwendung des *Elitist Clonal Selection Algorithm* (CSA) und des BICs die tatsächliche Knotenanzahl wiedergefunden werden. Beim Datensatz 1 und 4 hingegen ist die ermittelte etwas größer als die tatsächliche Knotenanzahl. Die anderen Methoden zur Knotenermittlung schneiden deutlich schlechter ab. Der Grund dafür ist, dass es sich hierbei um deterministische Methoden handelt und somit bei der Knotenermittlung kein globaler optimaler Knotenvektor aufgefunden werden kann. Zumeist wird lediglich ein lokaler optimaler Knotenvektor aufgefunden (vgl. Kapitel 4.1.4 und 4.2.1) und die Datenpunkte somit nicht optimal approximiert. Bei der Erhöhung der Modellkomplexität sinkt zwar die Varianz der Gewichtseinheit und somit die KL-Distanz kontinuierlich (Abb. 5.4), jedoch braucht man im Gegensatz zum CSA komplexere Modelle um eine ähnlich minimale KL-Distanz wie in Abb. 5.3 zu erreichen. Eine Modellselektion unter Zuhilfenahme der DOM- bzw. GDR-Methode, wie dies zu Beginn des Kapitels angedacht war, ist deshalb leider nicht sinnvoll. Der Bestrafungsterm fällt bei den jeweiligen Informationskriterien meistens zu schwach aus, wodurch erst bei Modellen mit einer hohen Parameteranzahl ein Minimum erreicht wird. Durch die Überparametrisierung besteht jedoch die Gefahr, dass es zu ungewollten Kurvenverläufen kommt. Außerdem ist ja das Ziel, mit einer möglichst geringen Anzahl an Parametern, die Datenpunkte bestmöglich durch eine B-Spline Kurve zu approximieren. Da dies am besten unter Verwendung des CSA gelingt wird im Kapitel 6 nur noch diese Methode für die Knotenermittlung benutzt.

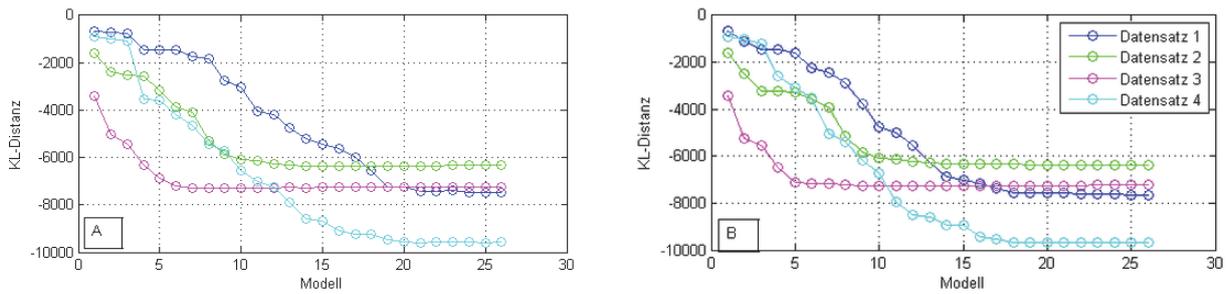


Abb. 5.4: KL-Distanz der einzelnen Datensätze aus Kapitel 3.1;
A) Knotenermittlung mit Hilfe der DOM-Methode
B) Knotenermittlung mit Hilfe der GDR-Methode

Zuvor muss noch die Tatsache berücksichtigt werden, dass bisher alle Methoden ein schlechteres Approximationsergebnis geliefert haben, wenn die Parameterwerte der Datenpunkte unbekannt waren. Wie anhand der Tab. 5.3 zu erkennen ist hat dies auch gravierende Auswirkungen auf die Selektion des „besten“ bzw. „quasi besten“ Modells. Erfolgt die Parametrisierung der Datenpunkte über die *centripetal* Methode so tritt, selbst bei der CSA-Methode, eine zum Teil extreme Überparametrisierung ein. Bei den anderen Methoden reicht zumeist nicht einmal das Modellset aus um ein geeignetes Modell zu finden, da oftmals durch das Informationskriterium das komplexeste Modell mit 25 inneren Knoten ausgewählt wird.

Tab. 5.3: Zusammenfassung der Ergebnisse der Modellselektion unter der Verwendung von Informationskriterien (AICc & BIC) bei unterschiedlichen Datensätzen und unterschiedlichen Methoden der Knotenermittlung; Parameterwerte der Datenpunkte bei allen Berechnungsschritten unbekannt

Datensatz x – tatsächliche Knotenanzahl	ermittelte Anzahl an innere Knoten					
	CSA		DOM		GDR	
	BIC	AICc	BIC	AICc	BIC	AICc
Datensatz 1 - 9 innere Knoten	23	25	25	25	25	25
Datensatz 2 - 6 innere Knoten	19	22	20	20	18	25
Datensatz 3 - 3 innere Knoten	8	17	9	25	14	17
Datensatz 4 - 8 innere Knoten	24	24	25	25	24	24

CSA ... *Elitist Clonal Selection Algorithm*; DOM ... *dominante Punkte*; GDR ... *Gleichverteilung des Rauschverhältnis*

6. Modellselektion und Knotenschätzung anhand realer Daten

Im Kapitel 5 konnte gezeigt werden, dass unter Verwendung des *Elitist Clonal Selection Algorithm* (CSA) und mit Hilfe des Bayesianischen Informationskriteriums (BIC) ein passendes funktionales Modell und anschließend ein zufriedenstellender Knotenvektor aufgefunden wird, wenn die Parameterwerte der Datenpunkte bekannt sind. Da jedoch bisher nur simulierte Datensätze zur Anwendung kamen soll in diesem Kapitel die Praxistauglichkeit bei der Auffindung einer passenden Knotenanzahl bzw. entsprechender Knotenpositionen und somit eines zufriedenstellenden Modells untersucht werden. Als Datensatz stehen, die aus einem Scan einer B-Spline Fläche vom Grad 3, extrahierte Profile in horizontaler und vertikaler Richtung bereit (siehe Kapitel 3.2). Da eine B-Spline Fläche zwei Parameterrichtungen aufweist, soll durch die Anordnung der Profile untersucht werden, ob für die entsprechenden Parameterrichtungen zuerst die korrekte Anzahl an Knoten und schließlich die Knotenpositionen der B-Spline Fläche aufgefunden werden können. Die Parameterwerte der Punkte sind jedoch im Allgemeinen nicht bekannt, sondern müssen zuvor erst ermittelt werden. Bei diesen Profilen kommt deshalb wieder die *centripetal* Methode zur Anwendung. Da auch die Anzahl an inneren Knoten und somit das funktionale Modell a-priori nicht bekannt ist muss zuerst einmal, ausgehend von einer B-Spline Kurve vom Grad 3, für unterschiedliche Knotenanzahlen (0 bis 15 innere Knoten) ein Modellset aufgestellt werden. Beim CSA kommen bei alle nachfolgenden Profile die folgenden Parametereinstellungen zur Anwendung.

$$\{N, d, n, a, N_c, N_e, N_{iter}\} = \{90, 10, 10, 0.1, 178, 10, 90\}$$

Diese wurden auf Grundlage der berechnete Affinität (4.9) des besten Knotenvektors innerhalb einer Generation bestimmt, wobei das komplexeste Modell mit 15 inneren Knoten bei den Berechnungen benutzt wurde (siehe Kapitel 4.3.1). Anhand des BICs erfolgt anschließend die Auswahl eines passenden Modells (vgl. Tab. 5.1). Die Ergebnisse der Modellselektion für die einzelnen Profile sind in Tab. 6.1 ersichtlich. Horizontale Profile bekommen dabei die Bezeichnung Hx und vertikale Profile die Bezeichnung Vx. Die Positionen dieser sind in Abb. 6.2 zu sehen

Tab. 6.1: Anhand des BICs ermittelte Anzahl an inneren Knoten für die extrahierten Profile aus Kapitel 3.2 Parametrisierung der Datenpunkte über *centripetal* Methode

Profil	XY-Ebene	innere Knoten	Profil	YZ-Ebene	innere Knoten
H1	Z = 99.55 [m]	11	V1	X = 1000.4 [m]	7
H2	Z = 99.65 [m]	14	V2	X = 1000.5 [m]	9
H3	Z = 99.75 [m]	13	V3	X = 1000.6 [m]	12
H4	Z = 99.85 [m]	2	V4	X = 1000.7 [m]	13

Nachdem nun die Anzahl an inneren Knoten für jedes einzelne Profil feststehen, kann mit Hilfe des CSA ein optimaler Knotenvektor und anschließend über den linearen Ausgleichsschritt die Kontrollpunkte für die jeweiligen B-Spline Kurven ermittelt werden. Die Form des gescannten Objekts ist hochgenau bekannt und kann deshalb als Soll-Form für die ermittelten B-Spline Kurven benutzt werden. Vorerst liegen jedoch die B-Spline Kurven der extrahierten Profile und die Soll-Fläche in unterschiedlichen Koordinatensystemen. Um die Systeme für Vergleichszwecke zusammenzuführen, können die am Messobjekt angebrachten Identpunkte benutzt werden. Bei der Soll-Fläche besteht jedoch das Problem, dass vorerst lediglich die Referenzpunkte für das 0,5" große Kugelprisma (RRR) gemeinsam mit der Fläche in einem System bekannt sind. Da die einzelnen Identpunkte beim Scanvorgang jedoch mit einem 1,5" großen Kugelprisma aufgenommen wurden, besteht nach der Transformation zwischen der Soll-Fläche und der Punktwolke eine konstante Differenz. Deshalb werden in einem eigenen Messvorgang die Halterungen der Identpunkte mit beiden Kugelprismen mittels eines Lasertrackers eingemessen, um somit die Soll-Fläche nun vollständig über eine 3D-Helmert-Transformation in das Scan-Koordinatensystem zu transformieren. Aus der Soll-Fläche werden anschließend die exakt gleichen Profile herausgeschnitten. Für Vergleichszwecke muss die Soll-Fläche zuvor diskretisiert werden, indem auf dieser selbst 102400 Punkte berechnet werden. Die Abweichungsberechnung erfolgt so, dass zuerst für einen Punkt der B-Spline Kurve die zwei nächstgelegenen Punkte des Soll-Profiles ermittelt werden. Ausgehend davon entspricht der Abstand zwischen dem B-Spline Kurvenpunkte und dem Lotfußpunkt auf die Verbindungsgeraden der beiden Punkte des Soll-Profiles der Abweichung (Abb. 6.1). Der dabei durch die lineare Interpolation entstehende Fehler ist durch die hohe Punktdichte vernachlässigbar klein.

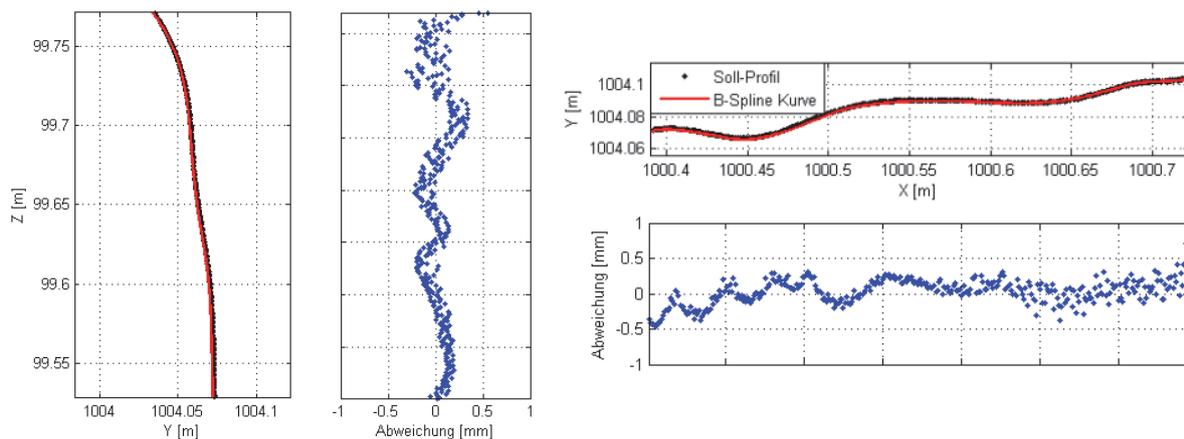


Abb. 6.1: B-Spline Kurve und Soll-Profil der extrahierten Profile V1 und H1 mit berechneter Abweichung

Die Abweichungen befinden sich bei allen Profilen ausschließlich im Bereich von $\pm 0,6 \text{ mm}$ und somit bereits unter der zu erwartenden Standardabweichung der Streckengenauigkeit

des Scans von $\sigma_s = 0.8\text{mm}$. Die etwas divergierende Abweichung, welche in den beiden Abweichungsgrafiken (Abb. 6.1) zu sehen ist und sich bei allen Profilen zeigt, rührt daher, dass bei der Profilextraktion, bedingt durch die hohe Punktdichte der Soll-Fläche, oftmals zwei nebeneinander liegende Punkte eines Soll-Profiles ausgeschnitten wurden. Dadurch kann es zu leicht unterschiedlichen Distanzberechnungen zwischen einzelnen nah beieinander liegenden Punkten kommen, wodurch die Abweichungsberechnung leicht streut. Die Grenzen für die einzelnen Soll-Profile können nicht enger gewählt werden, da ansonsten stellenweise keine Punkte für Vergleichszwecke zur Verfügung stehen.

Durch die homogene Verteilung der Messpunkte und der glatten Gestalt der Messoberfläche kann also durch die Parametrisierung der Datenpunkt mit Hilfe der *centripetal* Methode eine zufriedenstellende B-Spline Kurve aufgefunden werden. Dieses Verhalten hat sich auch schon bei den einfachen Datensätzen 2 und 3 in Kapitel 4.3.1 gezeigt. Die Anzahl an Knoten (siehe Tab. 6.1), ausgewählt durch das BIC, ist hingegen nicht zufriedenstellend. Für die Erzeugung der B-Spline Fläche werden nämlich 3 innere Knoten in Parameterichtung u und 5 innere Knoten in Parameterichtung v benutzt. Die ursprüngliche Knotenverteilung in beiden Parameterichtungen ist in Abb. 6.2 dargestellt. Es werden also, wie bereits bei den simulierten Datensätzen, deutlich komplexere Modelle benötigt und durch das BIC ausgewählt.

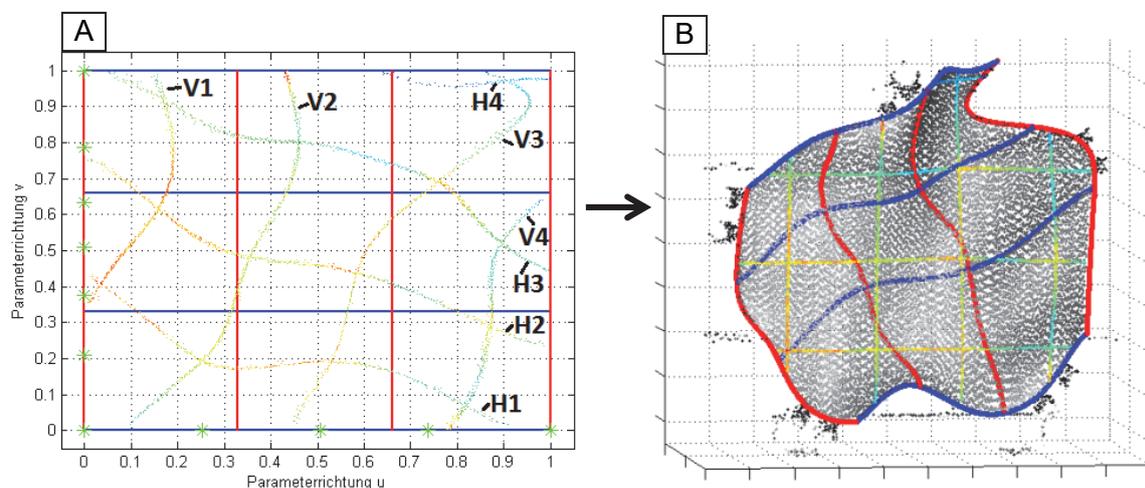


Abb. 6.2: A) Parameterraum zur Veranschaulichung der Parameterwertverteilung der unterschiedlichen Profile; grüne Marker zeigen die Knotenverteilung entlang der Parameterrichtungen
B) B-Spline Fläche mit Profilen konstantem Parameterwerts (blaue & rote Linien) und horizontalen und vertikalen extrahierten Profilen (vgl. Abb. 3.3)

Durch eine willkürliche Selektion von horizontalen und vertikalen Profilen kann schlussendlich die ursprüngliche Anzahl an inneren Knoten nicht aufgefunden werden, selbst dann nicht wenn die Parameterwerte der Datenpunkte mit Hilfe einer Basisfläche berechnet wurden und somit bekannt sind, wie anhand der Tab. 6.2 zu sehen ist. Auf die Ermittlung der

Parameterwerte mit einer Basisfläche wird im nächsten Absatz eingegangen. In Abb. 6.2 sind neben den bereits bekannten horizontalen und vertikalen Profilen (vgl. Abb. 3.3) auch jene Profile eingezeichnet, welche entstehen, wenn ein Parameterwert bei der B-Spline Fläche konstant gehalten wird. Anhand des Parameterraums in Abb. 6.2 A ist zu erkennen, dass die Profile H1, H3, H4, V1 und V4 nicht über eine gesamte Parameterrichtung reichen. Beispielsweise kommt die Parametrisierung der Datenpunkte des Profils H1 von $v = 0$ bis $v = 0,9$. Diese Profile können deshalb nicht benutzt werden um die ursprüngliche Anzahl an Knoten zu bestimmen. Ein Extremfall wird beim Profil H4 erreicht, welches gerade einmal ein Drittel des Parameterraums in Parameterrichtung u abdeckt. Dadurch lässt sich auch erklären warum in Tab. 6.1 nur zwei innere Knoten bei diesem Profil durch das BIC ausgewählt werden.

Tab. 6.2: Anhand des BICs ermittelte Anzahl an inneren Knoten für die extrahierten Profile aus Kapitel 3.2 Parametrisierung der Datenpunkte durch Basisfläche

Profil	XY-Ebene	innere Knoten	Profil	YZ-Ebene	innere Knoten
H1	Z = 99.55 [m]	---	V1	X = 1000.4 [m]	---
H2	Z = 99.65 [m]	8	V2	X = 1000.5 [m]	13
H3	Z = 99.75 [m]	---	V3	X = 1000.6 [m]	8
H4	Z = 99.85 [m]	---	V4	X = 1000.7 [m]	---

Das nicht Auffinden der ursprünglichen Knotenanzahl, sowohl bei bekannten als auch bei unbekanntem Parameterwerten der Datenpunkte, ist darauf zurückzuführen, dass die horizontalen und vertikalen Profile nicht mit Profilen konstantem Parameterwert zusammenfallen (Abb. 6.2 B). Diese Profile, welche entstehen, wenn ein Parameterwert konstant gehalten wird, sind nichts anderes als B-Spline Kurven im euklidischen Raum. Wird der Parameter u (bzw. v) festgehalten, so kann nur noch der Parameterwert v (bzw. u) variiert werden, wodurch im Allgemeinen eine gekrümmte Kurve entsteht. Bei den extrahierten Profilen handelt es sich hingegen um Profile, welche gedanklich mit Hilfe einer Ebene aus dem Scan herausgeschnitten wurden. Diese können deshalb nicht Profile mit einer konstanten Parameterrichtung repräsentieren wie in Abb. 6.2 A zu sehen ist.

Es stellt sich somit die Frage, ob die korrekten Anzahlen an Knoten und auch die Knotenpositionen dann aufgefunden werden können wenn Punkte eines Profils entlang eines konstanten Parameterwertes für die B-Spline Approximation benutzt werden (Abb. 6.2 A). Für die Profilextraktion entlang eines Parameterwertes müssen jedoch die Parameterwerte jedes einzelnen Punktes bekannt sein. Dieser Berechnungsschritt erfolgte in diesem Fall durch eine MATLAB-Funktion, welche von Corinna Harmening implementiert wurde. Dabei werden die gescannten Punkte auf eine Basisfläche mit bekannter Parameterform projiziert. Durch diese Vorgehensweise bekommt jeder Datenpunkt zwei Parameterwerte u und v zugeordnet. Als Basisfläche dient dabei ein sogenannter Coons Patch. Die genaue Vorgehensweise kann

bei Harmening und Neuner (2014) nachgelesen werden. Durch die bekannten Parameterwerte der einzelnen Punkte des Scans können nun Profile (P1 und P2) ermittelt werden bei denen ein Parameterwert konstant ist (siehe Abb. 6.3). Anschließend erfolgt eine Kurvenapproximation mit unterschiedlich komplexen B-Spline Kurven. Anhand des BICs findet eine Modellauswahl aus dem erstellten Modellset statt. Insgesamt stehen 16 Modelle, angefangen bei 0 inneren Knoten bis hin zu 15 inneren Knoten, zur Verfügung. Für das Auffinden der jeweiligen Knotenpositionen wird der CSA, mit den zu Beginn dieses Kapitels genannten Einstellungen, benutzt. Die Ergebnisse bezüglich der Modellselektion bzw. der Knotenanzahl und die Anzahl an extrahierten Punkten sind in Tab. 6.3 grün markiert.

Tab. 6.3: Durch BIC ermittelte Knotenanzahl für Profil P1 und P2

Profil		# Punkte	innere Knoten
P1	$u = 0,5 = konst$	462	7
P2	$v = 0,5 = konst$	553	5

innere Knoten	Profil P1 $u = 0,5 = konst$			Profil P2 $v = 0,5 = konst$		
	RMSE [m]	BIC	Δ_i (BIC)	RMSE [m]	BIC	Δ_i (BIC)
0	0,0017	-5843,06	732,78	0,0058	-5677,42	2158,43
1	0,0017	-5830,78	745,06	0,0032	-6322,67	1513,18
2	0,0014	-6028,61	547,23	0,0012	-7400,29	435,57
3	0,0009	-6373,67	202,17	0,0010	-7622,74	213,11
4	0,0008	-6503,25	72,59	0,0008	-7747,97	87,88
5	0,0007	-6567,01	8,83	0,0008	-7835,85	0,00
6	0,0007	-6562,98	12,86	0,0008	-7824,60	11,26
7	0,0007	-6575,84	0,00	0,0008	-7812,23	23,62
8	0,0007	-6564,61	11,23	0,0008	-7802,33	33,52
9	0,0007	-6553,33	22,51	0,0008	-7791,79	44,06
10	0,0007	-6544,23	31,61	0,0008	-7781,49	54,36
11	0,0007	-6532,00	43,84	0,0008	-7769,93	65,92
12	0,0007	-6521,67	54,17	0,0008	-7763,44	72,41
13	0,0007	-6515,33	60,51	0,0008	-7751,12	84,73
14	0,0007	-6509,60	66,24	0,0008	-7744,53	91,32
15	0,0007	-6492,17	83,67	0,0008	-7733,38	102,48

Bei beiden Profilen P1 und P2 kommt es zu einer leichten Überparametrisierung. Anstelle von 5 inneren Knoten in Parameterrichtung v ($u = 0,5 = konst$) werden durch das BIC 7 Knoten ausgewählt. In Parameterrichtung u ($v = 0,5 = konst$) sind es anstellen von ursprünglichen 3 inneren Knoten nun 5 Knoten. Diese leichte Überparametrisierung war auch bereits bei den Datensätzen 1 und 4 in Kapitel 4.3.1 beobachtbar. In diesem Fall

könnte die Parametrisierung der Datenpunkte durch die Basisfläche möglicherweise ein Grund dafür sein, dass nicht die ursprüngliche Anzahl an Knoten aufgefunden werden konnte. Bei Harmening & Neuner (2014) wird nämlich die Parametrisierung ausgehend von einem Coons Patch iterativ so lange verbessert, bis sich diese nicht mehr signifikant ändern.

Beim Profil P1 bzw. P2 zeigt sich, dass sich die RMSE-Werte der B-Spline Kurven ab dem 5. bzw. 4. inneren Knoten in der vierten Nachkommastelle nicht mehr ändern. Das hat wiederum zur Folge, dass die BIC-Werte bei diesen Modellen sehr eng beieinander liegen, wodurch auch die einzelnen BIC-Differenzen Δ_i klein ausfallen. Beim Profil P1 hat die B-Spline Kurve mit 5 inneren Knoten den zweitniedrigsten BIC-Wert und würde nach dem „quasi besten“ Modell (B-Spline Kurve mit 7 inneren Knoten) durch das BIC als das zweitbeste Modell gelten. Da die Differenz so klein ausfällt obliegt es dem Anwender eines der beiden Modelle auszuwählen. Dabei können die weiteren Fehlermaße wie z.B. die maximale bzw. mittlere Abweichung bei der Entscheidungsfindung helfen. Dies setzt voraus, dass es keine Ausreißer mehr in den Daten vorhanden sind. Bei Profil P2 zeigt sich dies nicht. In dieser Arbeit werden zuerst die vom BIC selektierten Modelle für die Approximation verwendet und im weiteren Verlauf mit den tatsächlichen Profilen verglichen.

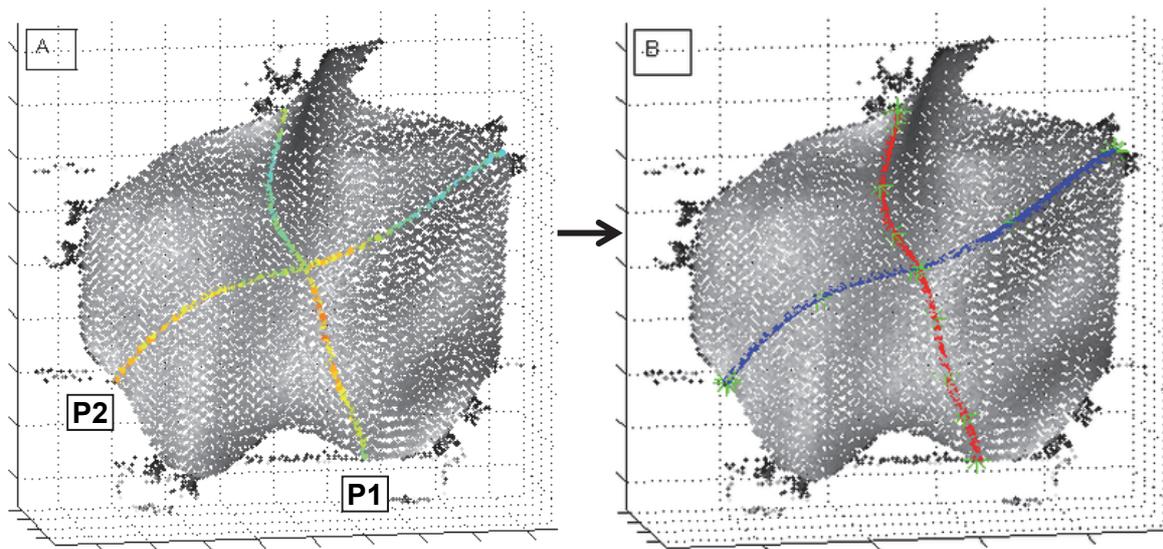


Abb. 6.3: A) Ausgeschnitten Profile mit einem konstanten Parameterwert ($u = v = 0,5$)
B) Approximation der Datenpunkte durch B-Spline Kurve mit 5 bzw. 7 inneren Knoten (vgl. Abb. 6.4)

Ein Vergleich der ermittelten Knotenpositionen zu den tatsächlichen Knotenpositionen der B-Spline Fläche pro Parameterrichtung ist in Abb. 6.4 zu sehen. Trotz der leichten Überparametrisierung der B-Spline Kurve durch zwei zusätzliche innere Knoten, können die einzelnen Knotenpositionen der B-Spline Fläche weitestgehend aufgefunden werden.

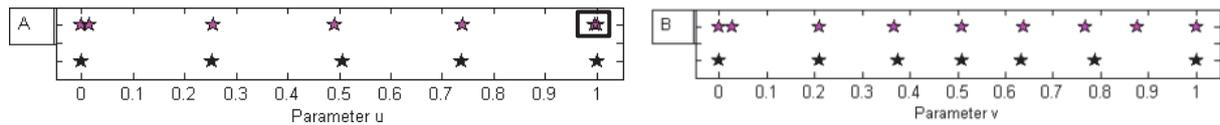


Abb. 6.4: Ursprüngliche Knotenanzahl und Knotenpositionen der B-Spline Fläche (schwarze Marker); durch BIC festgestellte Knotenanzahl und durch CSA ermittelte Knotenpositionen (farbige Marker) für A) Profil P2 und B) Profil P1 [doppelter Knoten schwarz umrahmt]

Der größte Unterschied zwischen den jeweiligen Knotenvektoren tritt in Parameterrichtung v an der Stelle $v \approx 0,8$ auf. Diese minimalen Variationen haben, wie bereits in Kapitel 4.3.1 gezeigt wurde, auf die Ergebnisse zumeist keinen nennenswerten Einfluss. Es fällt jedoch auf, dass die zusätzlichen Knoten, welche zu einer leichten Überparametrisierung führen, am Rande des jeweiligen Parameters u und v zu finden sind. So hat der erste bzw. letzte Knoten in Parameterrichtung u die Parameterwerte $u = 0,015$ bzw. $u = 0,995$. In Parameterrichtung v entsprechend die Werte $v = 0,028$ bzw. $v = 0,876$. Wird hingegen mit Hilfe des CSA die ursprüngliche Anzahl an Knoten gesucht, so fallen die äußersten inneren Knoten weg, wie anhand der Abb. 6.5 zu sehen ist. Gleichzeitig fällt auf, dass die anfänglichen Knotenpositionen der B-Spline Fläche nicht mehr genau aufgefunden werden. Auch bei mehrmaliger Approximation der Datenpunkte bleiben die jeweiligen Knotenpositionen innerhalb einer gewissen Toleranz konstant und behalten die ersichtliche Abweichungen bei.

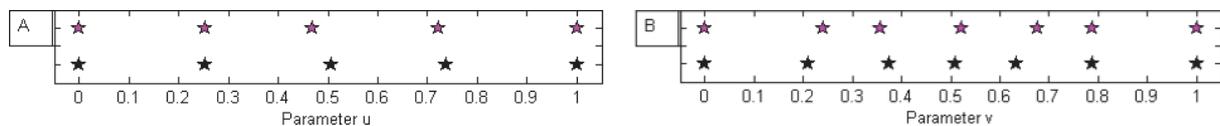


Abb. 6.5: Ursprüngliche Knotenanzahl und Knotenpositionen der B-Spline Fläche (schwarze Marker); ermittelte Knotenpositionen durch CSA bei Benutzung der ursprünglichen Anzahl an Knoten (farbige Marker) für A) Profil P2 und B) Profil P1

Es stellt sich somit die Frage, warum gerade am Rand des Parameterraums zusätzliche Knoten eingefügt werden. Normalerweise wird nur an jenen Stellen, wo eine große Abweichung zwischen der B-Spline Kurve und den zu approximierenden Datenpunkten besteht, ein neuer Knoten eingefügt, da dadurch die Affinität und somit der Abstand der B-Spline Kurve zu den Datenpunkten minimiert werden kann. Deshalb wurde zuerst für jeden Datenpunkt der Normalabstand zur B-Spline Kurve berechnet und in Abb. 6.6 dargestellt. Die benutzte Knotenverteilung bei den approximierenden B-Spline Kurven für die Grafiken A und B entspricht dabei der aus Abb. 6.5 bzw. für die Grafiken C und D der aus Abb. 6.4. Vor allem bei Abb. 6.6 A, jener Kurve mit der ursprünglichen Anzahl an Knoten, ist ein Anstieg der Abweichungen am Anfang und am Ende der Kurve deutlich bemerkbar. Bei Kurve B fällt dies nicht so sehr auf. Betrachtet man hingegen die Abweichungen bei den B-Spline Kurven mit zwei zusätzlichen Knoten (Abb. 6.6 C und D), so verringern sich die Abweichungen vor allem bei Grafik C am Anfang und am Ende der Kurve deutlich. Durch die zusätzlichen äußeren Knoten kann also effektiv die Approximationsgenauigkeit gesteigert werden. Dieser

Effekt kann bei der B-Spline Kurve in Parameterrichtung u ($v = 0,5 = konst$) auch anhand des RMSE-Wertes beobachtet werden. So weist die B-Spline Kurve mit 3 inneren Knoten einen RMSE-Wert von 1mm auf, wohingegen die B-Spline Kurve mit 5 inneren Knoten einen kleineren Wert von 0,8mm hat. Die Abweichungsgrafiken für die B-Spline Kurven in Parameterrichtung v (Abb. 6.6 B und D) unterscheiden sich nicht wesentlich voneinander. Dies ist auch anhand der RMSE-Werte aus Tab. 6.3 zu sehen, da beide Modelle einen RMSE-Wert von 0,7mm haben. Der auslösende Faktor für das Einfügen zusätzlicher Knoten am Rand des Parameterraums ist jener, dass die approximierenden B-Spline Kurven durch den ersten und letzten Datenpunkt verlaufen, da eingespannten B-Spline Kurven (siehe Kapitel 2.2.1) benutzt werden. Sind diese Datenpunkte nun stark verrauscht oder beinhalten grobe Messfehler, so ist die anfängliche Abweichung der B-Spline Kurve zu den Datenpunkten groß. In diesem Fall scheint es so, dass die Messdaten am Rand des gescannten Objekts stärker verrauscht sind als in der Mitte. Der CSA versucht diese Abweichungen durch zusätzliche Knoten am Anfang und am Ende des Parameterraums abzufangen. Diese äußeren Knoten können einen Großteil der Abweichungen beseitigen wodurch anschließend der CSA für die restlichen inneren Knoten die ursprünglichen Knotenpositionen auffinden kann (Abb. 6.4). Stehen diese äußeren Knoten nicht zur Verfügung, so müssen die inneren Knoten die am Rande auftretenden Abweichungen kompensieren, indem sie die Positionen zu den originalen Knoten verändern (Abb. 6.5).

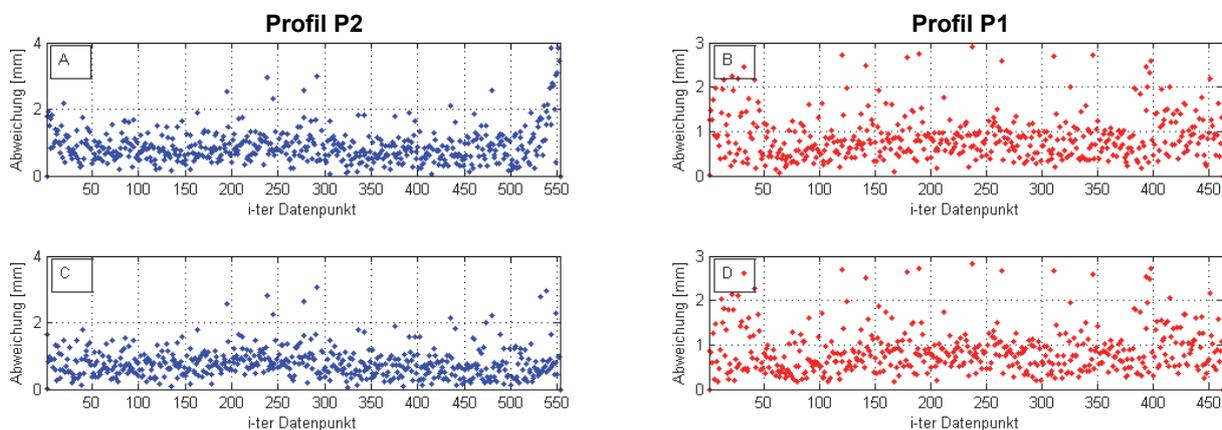
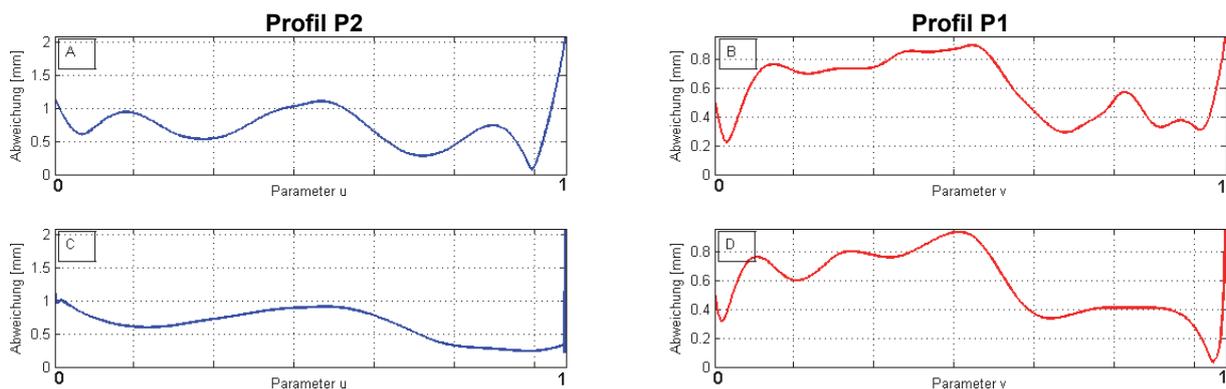


Abb. 6.6: Abweichung der Datenpunkte zur approximierenden B-Spline Kurve bei bekannter Parametrisierung der Datenpunkte in Parameterrichtung u (A und C) bzw. v (B und D)
A) 3 innere Knoten B) 5 innere Knoten (siehe Abb. 6.5)
C) 5 innere Knoten D) 7 innere Knoten (siehe Abb. 6.4)

Am Schluss dieses Kapitels werden die ermittelten B-Spline Kurven mit der Soll-Fläche verglichen. Dazu werden über die Parametrisierung der Soll-Fläche die jeweiligen Soll-Profile ausgeschnitten und mit den approximierenden B-Spline Kurven verglichen. In Abb. 6.7 sind die jeweiligen Abweichungen zur Soll-Kurve entlang des Parameters u bzw. v ersichtlich, wobei die Abweichung dem kürzesten Abstand eines Punktes auf der B-Spline Kurve zur

Soll-Kurve entspricht. Bei den Kurven A und C in Abb. 6.7 (B-Spline Kurve in Parameterrichtung $u = \text{Profil 2}$) ist zu sehen, dass durch das Hinzufügen von zwei zusätzlichen Knoten am Anfang und am Ende des Parameterintervalls eine deutliche Verbesserung eintritt und die Abweichungskurve glatter wird. Generell ist die B-Spline Kurve in Parameterrichtung u nicht weiter als 1mm von der Soll-Kurve entfernt, sollten 5 innere Knoten benutzt werden. Bei der Kurve B und D (B-Spline Kurve in Parameterrichtung $v = \text{Profil 1}$) ist eine Verbesserung bei Erhöhung der Knotenanzahl nur am Ende der Kurve bemerkbar, wo auch der Verlauf der Abweichungskurve dementsprechend glatter wird. Die Abweichung zum Soll-Profil liegt auch hier über den gesamten Verlauf deutlich unter 1mm. Die Verbesserungen am Rande der B-Spline Kurve durch die zusätzlichen inneren Knoten sind deutlich erkennbar. Eine maximale Abweichung am Rand der Profile von 2mm bei Profil P2 bzw. 0,9 mm bei Profil P1 bleibt trotz zusätzlicher Knoten bestehen (Abb. 6.7 C und D). Dies lässt sich darauf zurückführen, dass beim Scan des Messobjekts in einem 2mm großem Raster Punkte gemessen wurden und somit die Randbereiche des Objekts nicht zwingend in den Messungen vorhanden sind.



**Abb. 6.7: Abweichung der ermittelten B-Spline Kurve mit
A) 3 B) 5 C) 5 D) 7
inneren Knoten zur Soll-Kurve**

Es ist anhand der Abb. 6.7 klar ersichtlich, dass mit Hilfe des BICs und dem CSA eine zufriedenstellende Rekonstruktion der ursprünglichen Kurvengestalt möglich ist, sobald die Parameterwerte der Datenpunkte bei der B-Spline Berechnung bekannt sind oder zumindest bestmöglich mit Hilfe von Basisflächen ermittelt wurden. Auch das Auffinden der ursprünglichen Knotenpositionen ist möglich, sollten Profile entlang einer Parameterrichtung für Approximationszwecke zur Verfügung stehen.

7. Zusammenfassung

Diese Arbeit beschäftigte sich mit der Approximation von Punktdaten durch eine B-Spline Kurve um die Redundanz, welche beim Scannen von Objekten in der Ingenieurgeodäsie zumeist auftritt, für Modellierungszwecke auszunutzen. Die Approximationsgenauigkeit ist, sobald der Grad der B-Spline Kurve vorgegeben wird, maßgeblich von der Anzahl und von der Verteilung der inneren Knoten abhängig, mit Hilfe derer wiederum die Basisfunktionen erstellt werden. Die Kontrollpunkte werden typischerweise in einem linearen Ausgleichungsmodell geschätzt. Eine gleichzeitige Ermittlung von Knoten und Kontrollpunkten führt zu einem hochdimensionalen nichtlinearen Ausgleichungsproblem, welches zurzeit nicht gelöst werden kann.

Für die Ermittlung der Knotenpositionen stehen verschiedene Verfahren bereit. Eine derzeit übliche Vorgehensweise bei der Knotenbestimmung ist die Verwendung der Parameterwerte der Datenpunkte, welche zuvor in einem separaten Schritt im Wesentlichen über drei unterschiedliche Methoden berechnet werden können. Zwar besitzt durch diese einfache Methode die im Ausgleichungsproblem aufgestellte Normalgleichungsmatrix eine gute Kondition, jedoch ist eine zufriedenstellende Approximationsgenauigkeit in den meisten Fällen nur durch eine hohe Überparametrisierung und somit mit einer großen Anzahl an inneren Knoten erreichbar.

Anhand von vier verschiedenen B-Spline Kurven vom Grad 3, welche als Testdatensätze fungierten, konnte das Verhalten von insgesamt drei unterschiedlichen Verfahren zur Knotenbestimmung untersucht werden. Die erste Methode (DOM) benutzt für das Auffinden von Knoten sogenannte dominante Punkte, welche im Wesentlichen die Krümmungsinformationen der Punktdaten repräsentieren. Ziel dieser Methode ist es in Kurvenregionen mit starker Krümmung vermehrt Knoten zu platzieren, um somit eine global zufriedenstellende B-Spline Approximation zu gewährleisten. Wie sich jedoch gezeigt hat, werden vor allem bei komplexen Kurvenformen durch die fehleranfällige Krümmungsberechnung auch in einfachen Kurvenabschnitten dominante Punkte platziert, welche wiederum nachhaltig die Knotengenerierung negativ beeinflussen.

Die zweite Methode (GDR) geht bei der Ermittlung der Knotenposition von einer Minimalkonfiguration des Knotenvektors aus und erhöht sukzessive die Anzahl an inneren Knoten bis ein definiertes Kriterium erfüllt wird. Die hinzugefügten Knoten werden dabei so platziert, dass sich die kumulierte Summe der quadrierten Residuen innerhalb einer Knotenspanne mit dem größten Wert verringert. Anhand der berechneten Abweichungen konnte festgestellt werden, dass diese Methode im Gegensatz zur DOM-Methode bessere Approximationsresultate erzielt.

Bei beiden Methoden musste jedoch die Modellkomplexität und somit die Anzahl an inneren Knoten für eine zufriedenstellende B-Spline Approximation erhöht werden, da systematische Effekte in den Residuen auftraten sobald die ursprüngliche Anzahl an inneren Knoten für die Approximation der Testdatensätze benutzt wurde. Grund dafür war die unzureichende Lokalisierung der inneren Knoten bei beiden Methoden.

Die letzte untersuchte Methode (CSA) ahmt einen Teil des natürlichen Immunsystems nach und versucht dabei durch zufällige Mutationsschritte und einer geeigneten Selektion von Knoten einen globalen optimalen Knotenvektor zu finden. Das Konvergenzverhalten der Methode wird im Wesentlichen von 7 Parametern bestimmt. In der Arbeit wurden die Auswirkungen der einzelnen Parameter auf das Konvergenzverhalten als auch auf das Auffinden eines optimalen Knotenvektors behandelt. Die Ergebnisse fielen durch die zufällige Mutation und Knotengenerierung bei jeder Realisierung leicht unterschiedlich aus. Es konnte jedoch gezeigt werden, dass bei bekannter Parametrisierung der Datenpunkte die CSA-Methode die ursprünglichen Knotenpositionen auffinden kann. Sind hingegen die Parameterwerte der Datenpunkte vorab nicht bekannt, sondern müssen erst über einen separaten Berechnungsschritt ermittelt werden, so fallen bei allen drei Methoden die Ergebnisse deutlich schlechter aus. Die Auswirkung der Parametrisierung auf das Approximationsergebnis gäbe Grundlage für weiterführende Untersuchungen.

Für die Ermittlung der Anzahl an inneren Knoten und somit der Modellkomplexität wurden in dieser Arbeit das Akaike (AIC) mit dem Bayesianische (BIC) Informationskriterium verglichen. Dabei stellte sich heraus, dass das BIC im Gegensatz zum AIC durch den stärkeren Bestrafungsterm, bei Hinzunahme von zusätzlichen Knoten, Modelle mit einer niedrigeren Anzahl an inneren Knoten auswählt und dadurch am ehesten die ursprüngliche Knotenanzahl auffinden kann. Die besten Ergebnisse, bei der Approximation der Punktdaten durch eine B-Spline Kurve, konnten schlussendlich hinsichtlich der Knotenanzahl und der Knotenverteilung in Kombination mit der CSA-Methode erzielt werden.

Die zuvor gewonnen Erkenntnisse wurden anhand von realen Daten überprüft. Als Datensatz standen die aus einem Scan einer B-Spline Fläche vom Grad 3 extrahierte horizontale und vertikale Kurvenprofile bereit. Dabei stellte sich heraus, dass durch eine willkürliche Selektion von Profilen, trotz der Verwendung des BICs und der CSA-Methode, die ursprüngliche Knotenanzahl und Knotenverteilung nicht aufgefunden werden konnten. Kommen jedoch Profile in einer Parameterrichtung für die B-Spline Approximation zur Anwendung, so können zumindest die ursprünglichen Knotenpositionen aufgefunden werden. Es kommt, bedingt durch das größere Rauschen am Rande des gescannten Objekts, zu einer leichten Überparametrisierung. Diese Überparametrisierung hat jedoch in diesem Fall seine Berechtigung und wird in der vorliegenden Arbeit auch behandelt.

Die Parametrisierung der Punktdaten, welche einen signifikanten Einfluss auf alle gezeigten Approximationsergebnisse hat, sollte in weiteren Forschungsarbeiten unbedingt untersucht werden. Da die Evaluierung der vorgestellten Methoden nur anhand von Punktdaten von B-Spline Kurven stattfand, wäre es außerdem von Interesse, wie diese bei anderen Kurvenformen funktionieren

Literaturverzeichnis

- Gálvez A. (2015). Elitist clonal selection algorithm for optimal choice of free knots in B-spline data fitting. In: Applied Soft Computing 26, Elsevier-Verlag, S. 90-106.
- De Castro L.N, Von Zuben F.J. (2002). Learning and optimization using the clonal selection principle. IEEE Trans. Evolut. Comput. 6 (3), S. 239-251.
- De Castro L.N, Von Zuben F.J. (2000). The clonal selection algorithm with engineering applications. In: Proceedings of GECCO00, Workshop on Artificial Immune Systems and their Applications, Las Vegas, USA, S. 36-37.
- Park H. (2004). An error-bounded approximate method for representing planar curves in B-splines. In: Computer Aided Geometric Design, Elsevier, Vol. 21(5), S. 479-497.
- Park H, Lee J.H. (2007). B-spline curve fitting based on adaptive curve refinement using dominant points. In: Comput.-Aided Design, S. 439-451.
- Lee E. T. Y. (1989). Choosing nodes in parametric curve interpolation. In: Comput.-Aided Design, S. 363-370.
- Piegl L. A, Tiller W. (1997). The Nurbs Book, 2nd ed. Springer. 639 Seiten.
- Piegl L. A, Tiller W. (2000). Surface Approximation to Scanned Data. In: The Visual Computer 16. S. 386-395.
- Schoenberg I. J. and Whitney A. (1953). On Polya Frequency Function. III. The Positivity of Translation Determinants with an Application to the Interpolation Problem by Spline Curves. Trans. Am. Math. Soc., 74, 2, S. 246-259.
- Razdan A. (1999). Knot placement for B-spline curve approximation. Report. Arizona State University.
- Li W, Xu S. et al. (2005). Adaptive knot placement in B-Spline curve approximation. Elsevier Verlag, Computer-Aided Design , (37), S. 791-797.
- Schmitt C, Neuner H. (2015). Knot estimation on B-Spline curves. In: Österreichische Zeitschrift für Vermessung & Geoinformation 103 (2+3), S. 188-197.
- Cox M. G. (1972). The Numerical Evaluation of B-Splines. In: IMA Journal of Applied Mathematics, 10, S. 134–149.
- De Boor C. (1972). On calculating with B-splines. In: Journal of Approximation Theory, 6, S. 50-62.
- Koch K-R. (2000). Einführung in die Bayes-Statistik. Springer, Berlin-Heidelberg, 225 Seiten.

- Akaike H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. (Petrov & Csaki, Eds.) 2nd International Symposium on Information Theory, S. 267-281.
- Schwarz G.E. (1978). Estimating the dimension of a model. *Ann. Stat.* 6 (2), S. 461–464.
- Burnham K. P, Anderson D. R, (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. In: *Sociological Methods and Research* 33/2, S. 261-304.
- Burnham K. P, Anderson D. R. (2002). *Model selection and Multimodel inference: A practical information-theoretic approach*. 2nd Edition. New York: Springer Verlag.
- Harmening C, Neuner H.-B. (2014). Raumkontinuierliche Modellierung mit Freiformflächen, In: *Terrestrisches Laserscanning 2014*, Wissner 78, S. 105-122.
- Leica Geosystems, Datenblatt MS50, Abgefragt am 04.03.2016 von http://www.leica-geosystems.us/downloads123/zz/tps/nova_ms50/brochures-datasheet/Leica_Nova_MS50_DAT_en.pdf

Abbildungsverzeichnis

Abb. 2.1: Darstellung der rekursiven Definition von B-Spline Basisfunktionen (Piegl & Tiller 1997, S.51)	5
Abb. 2.2: Dreiecksschema zur Berechnung von B-Spline Basisfunktionen	5
Abb. 2.3: Stückweise kubische polynomiale Kurve mit drei Segmenten (Piegl & Tiller 1997, S.48)	6
Abb. 2.4: A) B-Spline mit äquidistantem Knotenvektor $\mathbf{U1}$ C) und zugehörigen Basisfunktionen B) B-Spline mit nicht äquidistantem Knotenvektor $\mathbf{U2}$ D) und zugehörigen Basisfunktionen beide B-Spline Kurven vom Grad $p=3$ benutzen die gleichen Kontrollpunkte \mathbf{P}	8
Abb. 2.5: Änderung der Position des sechsten Kontrollpunktes (vgl. Abb. 2.4 A) bewirkt eine lokale Änderung der B-Spline Kurve im Intervall $[u_6, u_{10})$	9
Abb. 2.6: B-Spline Kurve mit einfachem, doppeltem ($\mathbf{P4} = \mathbf{P5}$) und dreifachem Kontrollpunkt ($\mathbf{P3} = \mathbf{P4} = \mathbf{P5}$)	10
Abb. 2.7: B-Spline Kurve vom Grad A) $p=1$; B) $p=2$; C) $p=3$ D) B-Spline Kurve vom Grad $p=10=n$ (bei $n+1 = 11$ Kontrollpunkten \rightarrow Bézier-Kurve) unter jeder Kurvenabbildung sind die entsprechenden Basisfunktionen zu sehen	11
Abb. 2.8: Einfluss unterschiedlicher Parametrisierungsmethoden auf die Approximation von Datenpunkten durch eine B-Spline Kurve	14
Abb. 2.9: A) Ursprüngliche B-Spline Kurve $\mathbf{C2}$ mit 284 verrauschten Datenpunkten ($\sigma = 0.05m$) B) aus verrauschten Datenpunkten gewonnene approximierende B-Spline Kurve $\mathbf{C1}$	17
Abb. 2.10: RMSE-Wert in Abhängigkeit des Datenrauschens und des Punktabstandes	18
Abb. 2.11: A) Unterbestimmte B-Spline Approximation mit einem inneren Knoten B) unzufriedene B-Spline Approximation mit der korrekten Anzahl an inneren Knoten, jedoch falscher Knotenverteilung C) zufriedenstellende B-Spline Approximation mit korrekter Knotenanzahl und Knotenverteilung	20
Abb. 3.1: Verwendete B-Spline Kurven für die Evaluierung der unterschiedlichen Methoden in Kapitel 4	21
Abb. 3.2: Messobjekt mit seitlich angebrachten Halterungen für Kugelprismen	23
Abb. 3.3: links) Aus drei Scans zusammengefasste Punktwolke (Intensitätswerte farblich dargestellt) rechts) Aus Punktwolke extrahierte Scanprofile	24
Abb. 4.1: Flussdiagramm der DOM-Methode	27
Abb. 4.2: Datenapproximation durch Bisection-Methode mit eingezeichneten Knotenpunkten (grüne Marker); alle Datenpunkte wurden mit einem Rauschen belegt ($\sigma_2 = 0,005m_2$) A) unteres Limit an Kontrollpunkten ($\hat{=}$ Bézier-Kurve) B) oberes Limit an Kontrollpunkten (Interpol.) C) Basis B-Spline Kurve welche Toleranz $\delta = 0,05m$ nicht überschreitet	29
Abb. 4.3: Krümmungsplot der Basis B-Spline Kurve aus Abb. 4.2 mit eingezeichneten Krümmungsmaxima A) Parameterwerte u der Datenpunkte bekannt B) Parameterwerte u der Datenpunkte unbekannt	34
Abb. 4.4: Approximation der Datenpunkte durch Basis B-Spline Kurve (Toleranz $\delta = 0,05m$) A) Parameterwerte u der Datenpunkte bekannt B) Parameterwerte u der Datenpunkte unbekannt	34
Abb. 4.5: Aus Krümmungsplot der Basis B-Spline Kurve ermittelte dominante Punkte (=Krümmungsmaxima) A) Parameterwerte u der Datenpunkte bekannt B) Parameterwerte u der Datenpunkte unbekannt	35

Abb. 4.6: Ermittlung von neuen dominanten Punkten mit $r = 0$ [siehe Formel (4.4)] A) Ausgangskurve um neuen dominanten Punkt zu berechnen C) geplotteter Formfaktor λ für i -ten Datenpunkt und eingezeichneter neuer dominanter Punkt (schwarzer Kreis) B) B-Spline Kurve mit neuem dominantem Punkt D) neuerliche Ermittlung eines neuen dominanten Punktes E) B-Spline Kurve mit neuem dominantem Punkt, welche vorgegebene Toleranz einhält _____	36
Abb. 4.7: Ermittlung von neuen dominanten Punkten mit $r = 1$ [siehe Formel (4.4)] A) Ausgangskurve um neuen dominanten Punkt zu berechnen C) geplotteter Formfaktor λ für i -ten Datenpunkt und eingezeichneter neuer dominanter Punkt (schwarzer Kreis) B) B-Spline Kurve mit neuem dominantem Punkt D) neuerliche Ermittlung eines neuen dominanten Punktes E) B-Spline Kurve mit neuem dominantem Punkt, welche vorgegebene Toleranz einhält _____	37
Abb. 4.8: Parameterwerte u der Datenpunkte bei Datenapproximation durch B-Spline Kurve bekannt A) 9 innere Knoten B) 6 innere Knoten C) 3 innere Knoten D) 8 innere Knoten _____	38
Abb. 4.9: Parameterwerte u der Datenpunkte bei Datenapproximation durch B-Spline Kurve unbekannt A) 9 innere Knoten B) 6 innere Knoten C) 3 innere Knoten D) 8 innere Knoten _____	39
Abb. 4.10: Knotenverteilung für approximierende B-Spline Kurve A, Kurve B, Kurve C und Kurve D der Abb. 4.8 & Abb. 4.9 _____	41
Abb. 4.11: Auftretende Residuen $[m]$ in X- bzw. Y-Richtung mit deutlich erkennbaren systematischen Effekten bei 3 inneren Knoten _____	42
Abb. 4.12: Auftretende Residuen $[m]$ in X- bzw. Y-Richtung mit keinem erkennbaren systematischen Effekt bei 6 inneren Knoten _____	42
Abb. 4.13: Schematischer Ablauf der GDA-Methode _____	44
Abb. 4.14: A) Datenpunkte durch Bézier-Kurve approximiert B) Datenpunkte mit B-Spline Kurve (1 innerer Knoten) approximiert C) quadrierte Abweichung zwischen Bézier-Kurve und Datenpunkten D) von links bzw. von rechts aufsummierte quadratische Abweichung (kumSum) und eingezeichneter neuer Knoten (schwarzer Stern) _____	45
Abb. 4.15: Parameterwerte u der Datenpunkte bei Datenapproximation durch B-Spline Kurve bekannt A) 9 innere Knoten B) 6 innere Knoten C) 3 innere Knoten D) 8 innere Knoten _____	46
Abb. 4.16: Parameterwerte u der Datenpunkte bei Datenapproximation durch B-Spline Kurve unbekannt A) 9 innere Knoten B) 6 innere Knoten C) 3 innere Knoten D) 8 innere Knoten _____	47
Abb. 4.17: Knotenverteilung für approximierende B-Spline Kurve A, Kurve B, Kurve C und Kurve D der Abb. 4.15 & Abb. 4.16 _____	48
Abb. 4.18: Mittlere Knotenposition (Stern) plus zugehörige Standardabweichung (gefärbte Linie) für unterschiedliche Rauschlevels (von $\sigma = 0.005m$ bis $\sigma = 0.1m$) _____	49
Abb. 4.19: Auftretende Residuen $[m]$ in X- bzw. Y-Richtung mit deutlich erkennbaren systematischen Effekten bei 3 inneren Knoten _____	50
Abb. 4.20: Auftretende Residuen $[m]$ in X- bzw. Y-Richtung mit keinem erkennbarem systematischen Effekt bei 9 inneren Knoten _____	50
Abb. 4.21: Flussdiagramm des Elitist Clonal Selection Algorithmus für Optimierungsprobleme, Castro & Zuben (2002) _____	52

Abb. 4.22: Affinitätswert des besten Knotenvektors pro Generation (Iterationsschritt) für jene vier Datensätze aus Kapitel 3.1 bei Verwendung der Parameterwerte aus Tab. 4.5 Parameterwerte \mathbf{u} der Datenpunkte A) bekannt; B) unbekannt	56
Abb. 4.23: Entwicklung der Kurvenapproximation bei 20 Iterationsschritten; Parameterwerte \mathbf{u} der Datenpunkte bekannt; Abweichung in [m] der Ist- zur Soll-Kurve farblich kodiert	59
Abb. 4.24: Entwicklung der Kurvenapproximation bei 20 Iterationsschritten; Parameterwerte \mathbf{u} der Datenpunkte unbekannt; Abweichung in [m] der Ist- zur Soll-Kurve farblich kodiert	59
Abb. 4.25: Parameterwerte \mathbf{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve bekannt A) 9 innere Knoten B) 6 innere Knoten C) 3 innere Knoten D) 8 innere Knoten	60
Abb. 4.26: Ermittelte Knotenpositionen des Datensatzes 4 bei 10-maliger Wiederholung der Approximation und Schwankung der einzelnen Knoten im Parameterraum (Tabelle) Parameterwerte \mathbf{u} der Datenpunkte bekannt (schwarze Sterne \triangleq ursprüngliche Knotenpositionen)	62
Abb. 4.27: Parameterwerte \mathbf{u} der Datenpunkte bei Datenapproximation durch B-Spline Kurve unbekannt A) 9 innere Knoten B) 6 innere Knoten C) 3 innere Knoten D) 8 innere Knoten	63
Abb. 4.28: Ermittelte Knotenpositionen des Datensatzes 4 bei 10-maliger Wiederholung der Approximation und Schwankung der einzelnen Knoten im Parameterraum (Tabelle) Parameterwerte \mathbf{u} der Datenpunkte unbekannt (schwarze Sterne \triangleq ursprüngliche Knotenpositionen)	64
Abb. 4.29: Knotenverteilung für approximierende B-Spline Kurve A, Kurve B, Kurve C und Kurve D der Abb. 4.25 & Abb. 4.27 (Knoten mit Vielfachheit 2 sind schwarz umrahmt)	65
Abb. 4.30: Auftretende Residuen [m] in X- bzw. Y-Richtung mit keinem erkennbaren systematischen Effekt bei 3 inneren Knoten	65
Abb. 5.1: KL-Informationsverlust bei steigender Parameteranzahl typischer Anwendungsfall des AICs [Burnham & Anderson, 2004]	70
Abb. 5.2: KL-Informationsverlust bei steigender Parameteranzahl typischer Anwendungsfall des BICs [Burnham & Anderson, 2004]	71
Abb. 5.3: KL-Distanz der einzelnen Datensätze aus Kapitel 3.1; Knotenermittlung mit Hilfe der CSA-Methode	75
Abb. 5.4: KL-Distanz der einzelnen Datensätze aus Kapitel 3.1; A) Knotenermittlung mit Hilfe der DOM-Methode B) Knotenermittlung mit Hilfe der GDR-Methode	77
Abb. 6.1: B-Spline Kurve und Soll-Profil der extrahierten Profile V1 und H1 mit berechneter Abweichung	79
Abb. 6.2: A) Parameterraum zur Veranschaulichung der Parameterwertverteilung der unterschiedlichen Profile; grüne Marker zeigen die Knotenverteilung entlang der Parameterrichtungen B) B-Spline Fläche mit Profilen konstantem Parameterwerts (blaue & rote Linien) und horizontalen und vertikalen extrahierten Profilen (vgl. Abb. 3.3)	80
Abb. 6.3: A) Ausgeschnitten Profile mit einem konstanten Parameterwert ($\mathbf{u} = \mathbf{v} = \mathbf{0},5$) B) Approximation der Datenpunkte durch B-Spline Kurve mit 5 bzw. 7 inneren Knoten (vgl. Abb. 6.4)	83
Abb. 6.4: Ursprüngliche Knotenanzahl und Knotenpositionen der B-Spline Fläche (schwarze Marker); durch BIC festgestellte Knotenanzahl und durch CSA ermittelte Knotenpositionen (farbige Marker) für A) Profil P2 und B) Profil P1 [doppelter Knoten schwarz umrahmt]	84

Abb. 6.5: Ursprüngliche Knotenanzahl und Knotenpositionen der B-Spline Fläche (schwarze Marker); ermittelte Knotenpositionen durch CSA bei Benutzung der ursprünglichen Anzahl an Knoten (farbige Marker) für A) Profil P2 und B) Profil P1 _____ 84

Abb. 6.6: Abweichung der Datenpunkte zur approximierenden B-Spline Kurve bei bekannter Parametrisierung der Datenpunkte in Parameterrichtung u (A und C) bzw. v (B und D) A) 3 innere Knoten B) 5 innere Knoten (siehe Abb. 6.5) C) 5 innere Knoten D) 7 innere Knoten (siehe Abb. 6.4) _____ 85

Abb. 6.7: Abweichung der ermittelten B-Spline Kurve mit A) 3 B) 5 C) 5 D) 7 inneren Knoten zur Soll-Kurve _____ 86

Tabellenverzeichnis

Tab. 3.1: Kontrollpunkte P und Knotenvektor U der einzelnen Datensätze aus Abb. 3.1	22
Tab. 3.2: Anzahl an extrahierte Punkte für einzelne Profile	25
Tab. 4.1: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierende B-Spline Kurve bei bekannten Parameterwerten \mathbf{u} der Datenpunkte	38
Tab. 4.2: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierende B-Spline Kurve bei unbekannt Parameterwerten \mathbf{u} der Datenpunkte	40
Tab. 4.3: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierende B-Spline Kurve bei bekannten Parameterwerten \mathbf{u} der Datenpunkte	47
Tab. 4.4: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierender B-Spline Kurve bei unbekannt Parameterwerten \mathbf{u} der Datenpunkte	47
Tab. 4.5: Benutzte Parameterwerte für CSA bei simulierten Datensätzen	58
Tab. 4.6: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierender B-Spline Kurve bei bekannten Parameterwerten \mathbf{u} der Datenpunkte	60
Tab. 4.7: Fehlerwerte bei zehnmaliger Approximation der einzelnen Datensätze durch CSA; Parameterwert \mathbf{u} der Datenpunkte bekannt	61
Tab. 4.8: Maximale und mittlere Abweichung zwischen Datenpunkten und approximierender B-Spline Kurve bei unbekannt Parameterwerten \mathbf{u} der Datenpunkte	63
Tab. 4.9: Fehlerwerte bei zehnmaliger Approximation der einzelnen Datensätze durch CSA; Parameterwert \mathbf{u} der Datenpunkte unbekannt	64
Tab. 5.1: Exemplarisches Berechnungsergebnis eines Modellsets (0 bis 25 innere Knoten) für den Datensatz 2	74
Tab. 5.2: Zusammenfassung der Ergebnisse der Modellselektion unter der Verwendung von Informationskriterien (AICc & BIC) bei unterschiedlichen Datensätzen und unterschiedlichen Methoden der Knotenermittlung; Parameterwerte der Datenpunkte bei allen Berechnungsschritten bekannt	76
Tab. 5.3: Zusammenfassung der Ergebnisse der Modellselektion unter der Verwendung von Informationskriterien (AICc & BIC) bei unterschiedlichen Datensätzen und unterschiedlichen Methoden der Knotenermittlung; Parameterwerte der Datenpunkte bei allen Berechnungsschritten unbekannt	77
Tab. 6.1: Anhand des BICs ermittelte Anzahl an inneren Knoten für die extrahierten Profile aus Kapitel 3.2 Parametrisierung der Datenpunkte über centripetal Methode	78
Tab. 6.2: Anhand des BICs ermittelte Anzahl an inneren Knoten für die extrahierten Profile aus Kapitel 3.2 Parametrisierung der Datenpunkte durch Basisfläche	81
Tab. 6.3: Durch BIC ermittelte Knotenanzahl für Profil P1 und P2	82