



Kurven

VERSTEHEN – ENTWERFEN – BERECHNEN – DARSTELLEN

WENDELIN L. F. DEGEN



Scientific
Publishing

Wendelin L. F. Degen

Kurven

Verstehen – Entwerfen – Berechnen – Darstellen

Kurven

Verstehen – Entwerfen – Berechnen – Darstellen

von

Wendelin L. F. Degen

Impressum



Karlsruher Institut für Technologie (KIT)
KIT Scientific Publishing
Straße am Forum 2
D-76131 Karlsruhe

KIT Scientific Publishing is a registered trademark
of Karlsruhe Institute of Technology.

Reprint using the book cover is not allowed.

www.ksp.kit.edu



*This document – excluding the cover, pictures and graphs – is licensed
under a Creative Commons Attribution-Share Alike 4.0 International License
(CC BY-SA 4.0): <https://creativecommons.org/licenses/by-sa/4.0/deed.en>*



*The cover page is licensed under a Creative Commons
Attribution-No Derivatives 4.0 International License (CC BY-ND 4.0):
<https://creativecommons.org/licenses/by-nd/4.0/deed.en>*

Print on Demand 2018 – Gedruckt auf FSC-zertifiziertem Papier

ISBN 978-3-7315-0662-1

DOI 10.5445/KSP/1000069243

Inhaltsverzeichnis

1 Was ist eine Kurve?

1.1	Begriff einer Kurve	1
1.2	Parameterdarstellungen und reguläre Kurven	4
1.3	Parametertransformationen	6
1.4	Die Bogenlänge	7
1.5	Segmentierung – Geometrische Splines	9
1.6	Die Berührordnung	11

2 Elemente der Differentialgeometrie ebener Kurven

2.1	Begleitendes Bezugssystem – Frenet-Formeln	17
2.2	Die Krümmung bei allgemeinem Parameter	22
2.3	Evoluten, Evolventen	24
2.4	Natürliche Gleichung, Integration	27
2.5	Beispiele und Anwendungen	29

3 Geschlossene Kurven in der Ebene

3.1	Jordankurven	33
3.2	Flächeninhalt und Isoperimetrie	35
3.3	Konvexität	37

4 Kurven im euklidischen Raum

4.1	Frenet-Formeln	40
4.2	Natürliche Gleichungen	41

5 Polynomiale Kurven

5.1	Definition der polynomialen Kurven	43
5.2	Kubische Hermite-Splines	45
5.3	Kurven mit pythagoräischen Hodographen	47

6 Rationale Kurven

6.1	Allgemeine rationale Kurven	49
6.2	Homogene Koordinaten und Fernpunkte	51
6.3	Anwendung, insbesondere auf Kegelschnitte	55
6.4	G^k -Stetigkeit homogener dargestellter Kurven	58

7 Bézier-Darstellung

7.1	Parameterdarstellung und Kontrollpunkte	63
7.2	Zusammenhang mit polynomialen Kurven	66
7.3	Ableitungen	68
7.4	Glatter Anschluss polynomialer Segmente	69
7.5	Konstruktion kubischer G^2 -Splines	70
7.6	Konvexitätsstreue	75
7.7	Unterteilung des Bézier-Polygons	77
7.8	Interpolation mit polynomialen Kurven	79

8 B-Splines

8.1	B-Splines	83
8.2	Spline-Kurven	87
8.3	Der de Boor-Algorithmus	89
8.4	B-Splines mit mehrfachen Knoten	92
8.5	Rationale Spline-Kurven (NURBS)	93

9 Algebraische Kurven

9.1	Grundbegriffe	95
9.2	Tangenten	98

9.3	Singularitäten	99
9.4	Schnittpunkte mit Geraden	101
9.5	Die Resultante zweier Polynome	103
9.6	Rationale Kurven als algebraische Kurven	105
9.7	Numerische Berechnung von Punkten	109
9.8	Schnittpunkte zweier algebraischer Kurven	115

10 Weitere Spline-Konstruktionen

10.1	Der Overhauser-Spline	117
10.2	Unterteilungsalgorithmen	118
10.3	Abschneiden der Ecken (Corner cutting)	121

11 Spezielle Kurven

11.1	Die Kettenlinie (Katenoid)	123
11.2	Die Traktrix	127
11.3	Hundekurven	131
11.4	Spiralen	132
11.5	Die Cassinischen Ovalen	135
11.6	Die Zykloide und die Trochoiden	136
11.7	Die Kleeblattschlinge	142

Literatur

Geleitwort

Lebhafte Diskussionen während eines Seminars über Architekturgeometrie, das wir 2010 zusammen mit Wendelin Degen für Studierende der Architektur und Informatik durchgeführt hatten, regten Wendelin Degen an, eine Einführung speziell zur Geometrie der Kurven zu schreiben, der in anderen Lehrbüchern meist wenig Raum gegeben wird.

So entstand ein Werk, aufbauend auf in Karlsruhe und Stuttgart gehaltenen Vorlesungen Wendelin Degens, das theoretische Resultate aus der Differentialgeometrie und Algebraischen Geometrie mit anwendungsnahen Techniken vereint und Beweise enthält, die man nicht oder nur schwer wo anders findet. Neben speziellen Kurven, wie der Kettenlinie, der Hundekurve, der Kleeblattschlinge und einigen weiteren, werden polynomiale und rationale Kurven, sowie Splines besprochen, die insbesondere im rechnergestützten geometrischen Design Anwendung finden, die aber auch weit darüber hinaus verbreitet sind. Mit vielen Illustrationen ausgestattet kann das Buch als anregende Ergänzung zu anderen Lehrbüchern dienen.

Dieses Buch über Kurven ist Wendelin Degens letztes Werk geworden. Kurz vor seinem Tod 2014 übergab er es dem unten Letztgenannten zur Publikation, der es mit Hilfe der später durch Michael Degen aufgespurten TeX-Dateien

für den Druck editieren konnte und versucht hat, Fehler zu korrigieren, sowie Notationen und Illustrationen zu vereinheitlichen. Dabei waren insbesondere die Korrekturhinweise von Bert Jüttler hilfreich und wir danken ihm für seine kritische Durchsicht.

Karlsruhe, Februar 2017

Udo Beyer, Klaus Meirer, Hartmut Prautzsch

Vorwort

Das vorliegende Bändchen über Kurven ist aus verschiedenen Lehrveranstaltungen hervorgegangen, die der Autor an den Universitäten Stuttgart und Karlsruhe, nicht nur für Mathematiker, sondern auch für Studenten der Informatik, der Architektur und des Bauingenieurwesens sowie für Hörer aller Fakultäten gehalten hat. Dementsprechend sind die Ansprüche an mathematische Vorkenntnisse des Lesers möglichst gering gehalten, oder es wurde auf Beweise verzichtet, wenn tiefer gehende Methoden dazu erforderlich sind.

Andererseits wird bei weitem nicht nur die Differentialgeometrie der Kurven behandelt, sondern es werden vor allem auch die neueren Methoden des CAGD (Computer Aided Geometric Design) zum Entwerfen und Zeichnen von Kurven dargestellt und im letzten Abschnitt eine Reihe von Kurven beschrieben, welche eine besondere Bedeutung erlangt haben, sei es aus historischen Gründen oder wegen ihrer Anwendungen.

Karlsruhe, Oktober 2014

Wendelin L. F. Degen

1 Was ist eine Kurve?

1.1 Begriff einer Kurve

Kurven begegnen einem im Alltag auf Schritt und Tritt bei Straßen, Eisenbahnen, Fluglinien, Flussläufen, als Silhouetten von Gegenständen und Umrisslinien von Flächen, als Randlinien von Bereichen, in Kunstwerken und auf Konstruktionszeichnungen. Hieraus hat sich durch Abstraktion der mathematische Begriff einer Kurve entwickelt. Aber auch dort ist er – ohne zusätzliche Attribute gebraucht – noch schillernd. Muss eine Kurve in einem umgebenden Raum liegen? Welchen Einfluss hat gegebenenfalls dieser Raum auf die Kurve? Kann eine Kurve einen „flächenhaften“ Bereich (z. B. in der Ebene) ganz ausfüllen? Was ist überhaupt *eine* Kurve? Kann sie aus mehreren getrennten Teilen („Ästen“) bestehen (man denke z. B. an eine Hyperbel) und warum gehören diese dennoch zu *einer* Kurve?

Alle diese Fragen bedürfen einer sorgsam mathematischen Klärung. Diese hat auch in der Geschichte geraume Zeit in Anspruch genommen und zu vielen Begriffsbildungen beigetragen. Andererseits sind Kurven und ihre automatische Erzeugung, Berechnung und Verarbeitung in Computern – vor allem im technischen Bereich – so vielfältig und wichtig, dass man auf gesicherte Erkenntnisse, vereinfachende Voraussetzungen und bewährte Verfahren ange-

wiesen ist. Hierzu möchte dieses Bändchen beitragen. Eine erste Fassung des Begriffs *Kurve* ist folgende

Definition 1. *Eine Kurve ist eine Punktmenge im Raum,¹ welche lokal² bijektiv und stetig³ auf ein nicht-leeres Intervall der reellen Zahlen abgebildet werden kann.*

Bemerkungen:

1. Diese Definition kann ohne Änderung auf topologische Räume verallgemeinert werden, da neben den reellen Zahlen nur die Begriffe der Abbildung und der Stetigkeit benötigt werden.
2. Die in der Definition genannte Punktmenge „erbt“ vom topologischen Raum, in dem sie liegt, per Durchschnittsbildung ein System von Umgebungen, welches sie selbst zu einem topologischen Raum macht (sogenannte *Relativtopologie*). Erst dadurch ist der Begriff der stetigen Abbildung in dieser Allgemeinheit sinnvoll.
3. Für Kurven in der euklidischen Ebene ist dies alles ganz einfach: Die Umgebungen einer reellen Zahl sind die *offenen Intervalle*, die diese Zahl enthalten, und die Umgebungen eines Punktes \mathbf{p} in der euklidischen Ebene sind die offenen Mengen, die diesen Punkt enthalten, (wobei man sich so-

¹ „Raum“ wird hier im mathematischen Sinn gebraucht; es kann auch eine Ebene oder eine räumliche Fläche oder aber ein mehrdimensionaler Raum oder ein solcher mit einer anderen als der euklidischen Struktur (etwa ein affiner oder projektiver Raum) sein.

² Eine Eigenschaft wird hier und im Folgenden *lokal* genannt, wenn jeder Punkt der Kurve eine Umgebung besitzt, in der die Eigenschaft erfüllt ist.

³ Gemeint ist, dass auch die Umkehrabbildung stetig ist; solche Abbildungen nennt man *topologisch*.

gar auf das System der offenen, konzentrischen Kreisscheiben mit diesem Punkt als Mittelpunkt beschränken kann).

4. Die in der Definition genannten lokalen topologischen Abbildungen unterliegen sonst keiner weiteren Einschränkung. Sie können durch beliebige andere ersetzt werden, wenn diese nur ebenfalls topologisch sind und die ganze Kurve überdecken. (Dies ist der tiefere Grund für die noch zu besprechende Möglichkeit der *Parametertransformationen*.)
5. Natürlich besteht der Wunsch, diese Abbildungen – es können sehr viele sein! – zu einer einzigen zusammenzufassen. In den meisten Fällen gelingt das auch, aber schon beim Kreis oder bei einer Hyperbel geht es nicht! Jeder Kreis ist eine abgeschlossene und beschränkte (d. h. *kompakte* Menge, die reellen Zahlen (mit der ihnen aufgeprägten Topologie) besitzen hingegen diese Eigenschaft nicht. Nun ist es ein wichtiger Satz der Topologie, dass die Kompaktheit bei topologischen Abbildungen *invariant* ist.

Daher kann es keine globale reelle Parametrisierung der Kreise geben. (Dasselbe gilt natürlich für alle anderen geschlossenen Kurven.)

Bei Hyperbeln liegt ein anderer Hinderungsgrund vor: Sie sind *nicht zusammenhängend*, sondern bestehen aus zwei Ästen. Auch der Zusammenhang einer Punktmenge ist eine topologisch invariante Eigenschaft, die das System der reellen Zahlen besitzt; also kann es auch in diesem Fall keine globale topologische Abbildung geben.

6. Manch einer wird einwenden, dass Kreise doch sehr wohl in einheitlicher Weise vermöge

$$x(t) = \cos(t), y(t) = \sin(t), \quad t \in [0, 2\pi), \quad (1.1)$$

dargestellt werden können. Dies ist zwar richtig, aber die in (1.1) gegebene Abbildung ist *nicht topologisch*! Das Urbild einer Umgebung des Punktes $(1,0)$ ist kein offenes Intervall in \mathbb{R} !

1.2 Parameterdarstellungen und reguläre Kurven

Die Beschreibung einer Kurve wie oben in (1.1) für den Kreis durch eine (*vektorierte*) stetige Abbildung eines Intervalls $I \subset \mathbb{R}$

$$\mathbf{x}: t \rightarrow \mathbf{x}(t) \in \mathbb{R}^2, \quad t \in I, \quad (1.2)$$

(bezüglich eines vorgegebenen Koordinatensystems) heißt *Parameterdarstellung* der (ebenen) Kurve. Das Analoge gilt für räumliche Kurven, wenn die Ebene \mathbb{R}^2 durch den \mathbb{R}^n ersetzt wird. Das Intervall I kann aus *allen* reellen Zahlen ($I = \mathbb{R}$) oder aus einer nicht leeren zusammenhängenden Teilmenge von \mathbb{R} bestehen; es gibt dann nur die Möglichkeiten

1. I ist offen ($I = (a, b)$, $I = (-\infty, b)$, $I = (a, \infty)$),
2. I ist kompakt ($I = [a, b]$),
3. Mischformen von a) und b), wobei der eine Endpunkt des Intervalls dazu gehört, der andere aber nicht (z. B. $I = [a, b)$), jeweils mit $a < b$, $a, b \in \mathbb{R}$.

In (1.2) bedeutet \mathbf{x} zunächst nur die *Bezeichnung* der Abbildung; im konkreten Einzelfall muss sie explizit durch ihre Koordinatenfunktionen angegeben werden. Dabei ist es nicht erforderlich, dass die Abbildung global bijektiv ist;

die (Stetigkeit und) lokale Injektivität reicht (für theoretische Zwecke) völlig aus. In der *Praxis* ist allerdings diese weite Fassung des Kurvenbegriffs von Definition 1 untauglich, weil sie viel zu viele Möglichkeiten von kuriosen Kurven, z. B. solche, die in jedem noch so kleinen Teilstück unendlich viele Ecken haben, oder solche, deren Länge nicht berechenbar ist, zulässt. Daher machen wir fortan die

Einschränkende Voraussetzung: *Die Koordinatenfunktionen $t \mapsto x(t)$ und $t \mapsto y(t)$, $t \in I$, der Abbildung \mathbf{x} seien (zumindest stückweise) stetig differenzierbar.*

Eine solche Kurve wird (stückweise) *glatt* genannt.¹ (Man spricht in diesem Fall auch einfach von einer *tangentenstetigen* Kurve). An den eventuell vorhandenen Ausnahme-Stellen müssen die einseitigen Grenzwerte (beiderseits) existieren und gleich den einseitigen Grenzwerten der Ableitungsvektoren sein. Das Analoge gelte auch für Kurven im \mathbb{R}^n .

Auch das reicht im Allgemeinen nicht aus, sondern es wird zusätzlich noch folgendes verlangt:

Regularität: *Eine glatte Kurve heißt regulär, wenn für alle $t \in I$ gilt:*

$$\mathbf{x}'(t) \neq \mathbf{0}. \quad (1.3)$$

D. h. die Ableitung bzw. deren einseitiger Grenzwert darf also an keiner Stelle der Nullvektor sein. Nach den Sätzen über implizite Funktionen folgt hieraus die lokale Injektivität der Parameterdarstellung, so dass man sich darum nicht zu kümmern braucht.

¹ „stückweise“ bedeutet: bis auf endlich viele Stellen, an denen dies nicht gilt

1.3 Parametertransformationen

Es sei eine Kurve in einer Parameterdarstellung (1.2) gegeben. Wie aus Definition 1 einer Kurve hervorgeht, ist diese keineswegs eindeutig durch die Kurve selbst (als Punktmenge betrachtet) bestimmt: Vielmehr kann jede beliebige topologische und stetig differenzierbare Abbildung

$$\phi : \bar{I} \longrightarrow I, \quad (1.4)$$

deren Umkehrabbildung ebenfalls stetig differenzierbar ist, für eine neue Parameterdarstellung

$$\mathbf{x} : \bar{t} \rightarrow \bar{\mathbf{x}}(\bar{t}) = \mathbf{x}(\phi(\bar{t})) \in \mathbb{R}^2, \quad \bar{t} \in \bar{I}, \quad (1.5)$$

verwendet werden. Diese Darstellung ist der ersteren völlig äquivalent.

Es sei darauf hingewiesen, dass die Bedingung über die Differenzierbarkeit der Umkehrabbildung nicht ignoriert werden darf. So ist z. B. die Zuordnung $t \mapsto t^3$, $t \in \mathbb{R}$ eine topologische und stetig differenzierbare Abbildung von \mathbb{R} auf \mathbb{R} , aber die durch $s \mapsto \sqrt[3]{s}$ gegebene Umkehrabbildung ist an der Stelle $s = 0$ *nicht differenzierbar*!

Außerdem hatten wir oben die Regularität als zusätzliche Forderung an die (hier zu behandelnden) Kurven gestellt. Durch Ableiten von (1.5) folgt nun nach der Kettenregel

$$\frac{d\bar{\mathbf{x}}(\bar{t})}{d\bar{t}} = \frac{d\mathbf{x}(\phi(\bar{t}))}{dt} \frac{d\phi(\bar{t})}{d\bar{t}}, \quad (1.6)$$

also bleibt die Regularität nur dann erhalten, wenn

$$\frac{d\phi(\bar{t})}{d\bar{t}} \neq 0 \quad \text{für alle } \bar{t} \in \bar{I} \quad (1.7)$$

erfüllt ist. Diese Bedingung ist dafür auch hinreichend und garantiert zugleich die stetige Differenzierbarkeit der Umkehrabbildung.

Solche Abbildungen (1.4), die topologisch und stetig differenzierbar sind und darüber hinaus die Bedingung (1.7) erfüllen, heißen *Diffeomorphismen* (von \bar{I} auf I). Wenn also von einer Parametertransformation die Rede ist, wird immer stillschweigend vorausgesetzt, dass ein Diffeomorphismus der betreffenden Intervalle vorliegt.

1.4 Die Bogenlänge

Um die (*Bogen-*)*Länge* eines kompakten Kurvenstücks in einem euklidischen Raum \mathbb{R}^n zu messen, unterteilt man das Parameter-Intervall $[a, b]$ durch endlich viele Zwischenstellen $t_0 = a, t_1, t_2, \dots, t_n = b$, welche streng monoton wachsend gewählt werden müssen, und summiert die Länge der dadurch entstehenden Sehnen:

$$\bar{L} = \sum_{i=0}^{n-1} \|\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i)\| \quad (1.8)$$

(Hierbei bedeutet $\|\mathbf{v}\|$ die euklidische *Norm* des Vektors \mathbf{v}). Die Größe \bar{L} ist ein *Näherungswert* für die Länge des Kurvenbogens von $\mathbf{x}(a)$ nach $\mathbf{x}(b)$. Dieser hängt naturgemäß von der Wahl der Zwischenstellen ab und ist stets etwas kleiner als die wirkliche Länge L dieses Bogens, denn die Sehnen sind stets die kürzesten Verbindungen ihrer Endpunkte. Infolgedessen gilt

$$L = \sup_{\mathcal{T}} \bar{L}, \quad (1.9)$$

wobei \mathcal{T} die Menge aller Unterteilungen (d. h. die Menge aller endlichen Folgen von Zwischenstellen wie oben beschrieben) bedeutet.

In der Analysis wird gezeigt, dass bei stetig differenzierbaren Kurven dieses Supremum existiert (d. h. endlich ist) und dass sein Wert durch das Integral

$$L = \int_a^b \|\mathbf{x}'\| dt \quad (1.10)$$

gegeben ist. Bei allgemeineren Kurven – die den zusätzlichen Voraussetzungen der Glattheit und Regularität nicht zu genügen brauchen – kann es vorkommen, dass die Menge der Werte \bar{L} *nicht beschränkt* ist, also das Supremum (1.9) nicht existiert. Solche Kurven nennt man *nicht rektifizierbar*. Man kann also folgern:

Satz 2. *Glatte und reguläre Kurven mit kompakten Parameterintervallen in euklidischen Räumen sind rektifizierbar; d. h. sie haben eine wohlbestimmte endliche Länge, und diese kann durch das Integral (1.10) berechnet werden.*

Dass dem so ist, erkennt man in groben Zügen bereits an (1.7), indem man die Differenzen $\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i)$ (für jede Vektorkomponente einzeln) durch Differenzenquotienten $(\mathbf{x}(t_{i+1}) - \mathbf{x}(t_i))/\Delta t_i$ mit $\Delta t_i := (t_{i+1} - t_i)$ und diese nach dem Mittelwertsatz der Differentialrechnung durch $\mathbf{x}'(\tilde{\tau})$ ersetzt und danach wieder mit Δt_i multipliziert.¹

Nun braucht man nur noch eine unendliche Folge von Unterteilungen, deren Feinheitsgrad $\max_i \{|\Delta t_i|\}$ gegen Null konvergiert, heranzuziehen, um die Gleichheit von (1.9) mit dem Integral (1.10) einzusehen.

¹ Die Tilde über dem Zwischenwert τ soll andeuten, dass dieser bei jeder Vektorkomponente separat zu nehmen ist.

Aufgrund von Satz 2 kann man von einem beliebigen Anfangspunkt $\mathbf{x}(a)$ der Kurve aus die Bogenlänge s zu einem variablen Punkt $\mathbf{x}(t)$ durch

$$s = \int_a^t \|\mathbf{x}'\| dt \quad (1.11)$$

messen. Damit hat man eine Abbildung $t \mapsto s$ von $I = [a, b]$ auf $[0, L]$; diese ist offensichtlich streng monoton wachsend und stetig differenzierbar (mit $ds/dt = \|\mathbf{x}'\|$), also nach (1.7) ein Diffeomorphismus. Daher kann man sie als Parametertransformation benutzen:

Korollar 3. *Bei glatten und regulären Kurven in euklidischen Räumen kann man stets (gegebenenfalls mit Hilfe einer Parametertransformation) die Bogenlänge als Parameter zugrunde legen.*

Die Bedeutung dieses Ergebnisses liegt eher in der Theorie, denn die tatsächliche Berechnung der Bogenlänge ist zwar für die Gerade und den Kreis trivial, aber für die meisten anderen Kurven nicht einfach und nur in wenigen Fällen durch elementare Funktionen auszudrücken; die Integration von (1.10) führt entweder auf höhere Funktionen oder ist nur durch Näherungsverfahren zu erreichen.

1.5 Segmentierung – Geometrische Splines

Für die Praxis bedeuten die zusätzlichen Voraussetzungen der Glattheiten und Regularität keine wesentlichen Einschränkungen, da ohnehin in den meisten Fällen polynomiale oder rationale Funktionen (siehe die Kapitel 4 und 5) verwendet werden. Nur bei den letzteren hat man dafür zu sorgen, dass keine Nullstelle des Nenners in das Parameterintervall fällt.

Andererseits neigen aber polynomiale Funktionen bei größeren Intervallen und höheren Graden zu sehr großen Schwankungen und Abweichungen; um solche Effekte zu vermeiden, hat man schon seit längerem in der numerischen Mathematik bei der Interpolation und Approximation von Funktionen *stückweise polynomiale Funktionen*, sogenannte *Splines* eingesetzt. Diese sind wie folgt definiert:

Definition 4. Eine Funktion $f : [a, b] \rightarrow \mathbb{R}$ heißt Spline(funktion) vom Grad n mit der Knotenfolge $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$, wenn gilt:

1. Die Einschränkung von f auf jedes der Teilintervalle $[x_i, x_{i+1}]$, ($i = 0, 1, \dots, n-1$) ist ein Polynom vom Grad n .
2. Die Funktion f ist in $[a, b]$ (mindestens) $(n-1)$ -mal stetig differenzierbar. Wegen a) kann die n -te Ableitung nur an den Knoten Sprünge haben.

Theorie und Anwendungen der Splinefunktionen sind weit verbreitet und werden hier nicht weiter erörtert (s. z. B. [3]).

Da die oben erwähnten Effekte für polynomiale Funktionen auch bei polynomialen *Kurven* (s. Kapitel 4) auftreten, sind jene zum Design ästhetisch ansprechender Kurven ebenfalls ungeeignet. Deshalb werden auch im Bereich des CAGD (Computer Aided Geometric Design) fast ausschließlich Kurven verwendet, die sich aus (meist vielen) kompakten polynomialen *Segmenten* zusammensetzen.

Dieser Bereich der angewandten Mathematik und Informatik ist den Fragen gewidmet, wie die Koeffizienten zu bestimmen sind, damit das betreffende Segment eine gewünschte Form besitzt, wie man effiziente Algorithmen findet, um diese Segmente zu berechnen, und wie man sie zusammensetzen muss,

damit eine glatte Kurve entsteht, die die Übergangsstellen nicht mehr erkennen lässt.

Eine aus mehreren glatt aneinanderschließenden Segmenten bestehende Kurve heißt *ein geometrischer Spline* oder kurz ein G^1 -*Spline*.¹ Wenn jedoch die Segmente nur aneinanderschließen und Ecken zugelassen sind, spricht man von einem G^0 -*Spline*. Bei einem geometrischen Spline werden oft auch die Parameterintervalle ihrer polynomialen Segmente „aneinandergehängt“. Die Übergangsstellen innerhalb des Gesamt-Intervalls werden wie im Falle der Funktionen *Knoten* genannt (vgl. Kapitel 4).

In vielen Fällen reicht es nicht aus, dass *nur die Tangenten* an den Übergangsstellen übereinstimmen, sondern man wünscht sich einen noch „geschmeidigeren“ Übergang. Dieser muss nun *nicht in der Übereinstimmung der höheren Ableitungen* (sogenannte C^k -Stetigkeit) bestehen, denn diese Bedingung wäre bei segmentierten Kurven zu restriktiv und würde überdies von der Parametrisierung abhängen. Das adäquate Konzept ist das der *Berührordnung*, welches im folgenden Abschnitt behandelt wird.

1.6 Die Berührordnung

Um dies zu präzisieren und zu vereinheitlichen, setzt man zunächst voraus, dass jedes Segment noch in einem *offenen Intervall* \bar{I} , welches das kompakte Intervall $I = [a, b]$ seiner Parameterdarstellung enthält, definiert ist (bzw. dort-

¹ In manchen Fällen werden auch andere als polynomialen oder rationale Segmente zugelassen, und auch wenn es solche vom Grad n sind, muss nicht unbedingt die Berührordnung (s. Abschnitt 1.6) k gleich $n - 1$ sein.

hin fortgesetzt werden kann) und in \bar{I} beliebig oft differenzierbar ist.¹ Sodann erklärt man:

Definition 5. *Zwei beliebig oft stetig differenzierbare Kurven \mathcal{C} und \mathcal{D} mit den Parameterdarstellungen $\mathbf{x} : I \rightarrow \mathbb{R}^n, t \mapsto \mathbf{x}(t)$ und $\mathbf{y} : J \rightarrow \mathbb{R}^n, s \mapsto \mathbf{y}(s)$, berühren sich im Punkt $\mathbf{a} \in \mathbb{R}^n$ in k -ter Ordnung, wenn es eine Parametertransformation $f : J \rightarrow I, t = f(s)$ derart gibt, dass an der Stelle $t_0 \in I$ bzw. an dessen Urbild $s_0 \in J$ ($t_0 = f(s_0)$) die Kurven durch diesen Punkt \mathbf{a} gehen (d. h. $\mathbf{x}(t_0) = \mathbf{y}(f(s_0)) = \mathbf{a}$) und dort auch noch alle Ableitungen bis zur Ordnung k übereinstimmen.*

Diese Definition ist ohne Änderungen auch auf den Fall des Anschlusses von zwei Segmenten eines geometrischen Splines anwendbar, da ja vorausgesetzt wurde, dass die Segmente über ihre Endpunkte hinaus differenzierbar fortgesetzt werden können:

Definition 6. *Ein geometrischer Spline, bei welchem sich je zwei aufeinanderfolgende Segmente in ihren Anschlusspunkten in k -ter Ordnung berühren, heißt ein G^k -Spline.*

Bemerkungen:

1. Diese „ G^k -Stetigkeit“ ersetzt die aus der Analysis bekannte C^k -Stetigkeit von reellen Funktionen, bei welchen an den Knoten nur die Übereinstimmung der Ableitungen bis zur k . Ordnung gefordert ist. Da aber bei Kurven die Parameterverteilung (in gewissen Grenzen) willkürlich ist, muss man diesem Umstand Rechnung tragen. Man könnte die obige Definition somit auch kurz wie folgt fassen: „ G^k -Stetigkeit ist C^k -Stetigkeit

¹ Dies hat den Vorteil, dass man auch in den Endpunkten von Ableitungen sprechen kann, ohne auf einseitige Grenzwerte zurückgreifen zu müssen.

nach Anwendung einer geeigneten Parametertransformation auf eines der beiden Segmente“.

2. Nach Definition 5 ist klar, dass die G^k -Stetigkeit stets auch die G^m -Stetigkeit für alle Zahlen $m < k$ impliziert („ $G^k \dots$ “ bedeutet also soviel wie „mindestens $G^k \dots$ “).
3. Es sei ausdrücklich betont, dass der Begriff der Berührordnung *keinen* Bezug auf die (evtl. gar nicht vorhandene) *euklidische* Struktur des \mathbb{R}^n nimmt. Sie ist hingegen *affin, ja sogar projektiv invariant*.
4. Offensichtlich sind die zuvor eingeführten Begriffe der G^0 - und G^1 -Stetigkeit Spezialfälle der G^k -Stetigkeit für $k = 0$ bzw. für $k = 1$.
5. Es wird sich im Abschnitt 2.1 herausstellen, dass die G^2 -Stetigkeit vorliegt, wenn der Spline G^1 -stetig ist und in allen Anschlusspunkten die *Krümmungen der beiden Segmente übereinstimmen*. (Dies hat oft zu der irrtümlichen Meinung geführt, dass die G^2 -Stetigkeit eine euklidische Struktur des Raumes voraussetzt.)

Die Bedingungen der Berührordnung scheinen im konkreten Fall schwer nachprüfbar zu sein, da nichts darüber ausgesagt ist, wie man zu einer „geeigneten Parametertransformation“ gelangen soll. Daher nimmt man – wie so oft in der Mathematik – einmal an, diese sei schon gegeben und zieht daraus Folgerungen.

Durch Ableitung von $\mathbf{x}(f(s))$ nach der Kettenregel und Gleichsetzung mit den entsprechenden Ableitungen von $\mathbf{y}(s)$ an der Stelle $s = s_0$ erhält man (mit $t_0 = f(s_0)$):

$$\begin{aligned}
 \mathbf{x}(t_0) &= \mathbf{y}(s_0) \\
 \mathbf{x}'(t_0) \cdot \dot{f}(s_0) &= \dot{\mathbf{y}}(s_0) \\
 \mathbf{x}''(t_0) \cdot (\dot{f})^2 + \mathbf{x}'(t_0) \cdot \ddot{f}(s_0) &= \ddot{\mathbf{y}}(s_0) \\
 \mathbf{x}'''(t_0) \cdot (\dot{f}(s_0))^3 + 3\mathbf{x}''(t_0)\dot{f}(s_0)\ddot{f}(s_0) + \mathbf{x}'(t_0)\ddot{\ddot{f}} &= \ddot{\ddot{\mathbf{y}}}(s_0) \\
 \dots\dots\dots &= \dots
 \end{aligned} \tag{1.12}$$

Diese Bedingungen haben die allgemeine Struktur

$$\sum_{i=1}^m \frac{d^{m+1-i}\mathbf{x}(t_0)}{dt^{m+1-i}} P_{m,i} = \frac{d^m\mathbf{y}(s_0)}{ds^m}, \quad m = 1, \dots, k, \tag{1.13}$$

wobei $P_{m,i}$ (mit $i \leq m$) gewisse Polynome der Konstanten

$$f_j := \frac{d^j f(s_0)}{ds^j}, \quad j = 1, \dots, k, \tag{1.14}$$

sind und diese darin nur bis zum Index $j = i$ vorkommen. Nach (1.12), 1. Ableitung, gilt

$$P_{1,1} = f_1, \tag{1.15}$$

und die weiteren Polynome $P_{m,i}$ berechnen sich *rekursiv* – wie man aus der Kettenregel genauer erschließen kann – gemäß

$$P_{m+1,i} = P_{m,i} \cdot f_1 + \sum_{j=1}^{i-1} \frac{\partial P_{m,i-1}}{\partial f_j} \cdot f_{j+1}, \quad i = 1, \dots, m+1, \tag{1.16}$$

(wobei die für $i = 1$ die Summe und für $i = m + 1$ das erste Glied entfällt). Man hat also insbesondere

$$P_{m,0} = f_1^m \quad \text{und} \quad P_{m,m} = f_m, \quad m = 1, \dots, k. \quad (1.17)$$

Von der unbekannten Parametertransformation f benötigt man somit nur die Koeffizienten (1.14), und kann sie z. B. einfach als Taylor-Polynom

$$f(s) = t_0 + \sum_{j=1}^k \frac{1}{j!} f_j (s - s_0)^j \quad (1.18)$$

ansetzen; wichtig ist nur, dass $f_1 > 0$ ist, damit sie in einer Umgebung von s_0 (gleichsinnig) bijektiv ist.

Zusammenfassend ist damit gezeigt:

Satz 7. *Zwei Kurvensegmente \mathcal{C} und \mathcal{D} (s. Definition 5) berühren sich genau dann im gemeinsamen Punkt (1.12) in k . Ordnung bzw. haben dort einen G^k -stetigen Anschluss, wenn es Koeffizienten f_1, \dots, f_k mit $f_1 > 0$ derart gibt, dass die Bedingungen (1.13) gelten; dabei sind die Polynome durch (1.15), (1.16) rekursiv definiert.*

In den folgenden Abschnitten wird von diesem Satz noch mehrfach Gebrauch gemacht werden.

2 Elemente der Differentialgeometrie ebener Kurven

Die Differentialgeometrie von Kurven befasst sich mit (meist unendlich oft) differenzierbaren, regulären Kurven und untersucht vor allem die lokalen Eigenschaften wie z. B. Tangenten, Krümmungen, Normalen etc. (Letzteres sofern die Kurve in einem euklidischen Raum liegt). Das hauptsächliche Hilfsmittel hierzu sind *begleitende Bezugssysteme*. Zunächst werden *Kurven in der euklidischen Ebene* behandelt. Im folgenden Abschnitt werden analoge Fragen im Raum angesprochen.

2.1 Begleitendes Bezugssystem – Frenet-Formeln

Es sei I ein offenes Intervall reeller Zahlen und – wie in Kapitel (1.2) – eine differenzierbare Kurve durch ihre Parameterdarstellung gegeben:

$$\mathbf{x} : s \rightarrow \mathbf{x}(s) \in \mathbb{R}^2, \quad s \in I. \quad (2.1)$$

Wie im vorigen Kapitel gezeigt wurde, kann man stets die *Bogenlänge als Parameter* benutzen, was hier von vornherein geschehen sei. (Deshalb wurde der Parameter mit s bezeichnet.)

Ferner wurde das Intervall als offen vorausgesetzt, damit man nicht zu umständlichen Formulierungen mit einseitigen Grenzwerten gezwungen ist, denn zu jedem Parameterwert s gibt es immer noch eine Umgebung, die ganz im Intervall I enthalten ist.

Dass die Bogenlänge als Parameter benützt wird, drückt sich durch

$$\|\mathbf{x}'(s)\| = 1 \quad \text{für alle } s \in I \quad (2.2)$$

aus. Der Vektor $\mathbf{x}'(s)$ ist also ein *Einheitsvektor*, der in Richtung der *Tangente* weist. Diese hat (an einer bestimmten Stelle $s \in I$) die Darstellung

$$\mathbf{y} = \mathbf{x}(s) + t \cdot \mathbf{x}'(s), \quad (2.3)$$

wobei $t \in \mathbb{R}$ der laufende Parameter auf der Tangenten ist.

Es bietet sich nun an, mit dem betreffenden Kurvenpunkt $\mathbf{x}(s)$ ein *begleitendes Bezugssystem* zu verknüpfen, dessen Ursprung eben dieser Kurvenpunkt und dessen erster Basisvektor $\mathbf{t}(s) := \mathbf{x}'(s)$ ist. Der zweite Basisvektor soll mit $\mathbf{t}(s)$ ein Orthonormalsystem bilden und ist dadurch eindeutig bestimmt. Er wird *Normalenvektor* genannt und mit $\mathbf{n}(s)$ bezeichnet.

Man kann sich vorstellen, dass sich dieses Bezugssystem in Abhängigkeit von s mit gleichmäßiger Geschwindigkeit entlang der Kurve fortbewegt.

Die Ableitungen der Basisvektoren $\mathbf{t}(s), \mathbf{n}(s)$ lassen sich als Linearkombinationen ihrer selbst darstellen. Da es sich um ein Orthonormalsystem handelt, ist die Koeffizientenmatrix *schiefssymmetrisch*, so dass nur ein einziger Koeffizient, der mit $\kappa(s)$ bezeichnet wird, übrig bleibt. Wenn man noch – wie das allgemein üblich ist – die (selbstverständliche) Abhängigkeit vom Kurvenpa-

parameter s weglässt, schreiben sich diese sogenannte *Frenet'schen Ableitungsgleichungen* wie folgt:

$$\begin{aligned} \mathbf{x}' &= \mathbf{t} \\ \mathbf{t}' &= \kappa \mathbf{n} \\ \mathbf{n}' &= -\kappa \mathbf{t}. \end{aligned} \quad (2.4)$$

Die Größe $\kappa(s)$ heißt *Krümmung* der Kurve an der Stelle s . Wenn $\kappa(s) \neq 0$ ist, heißt die Größe $r(s) := 1/\kappa(s)$ der *Krümmungsradius* an dieser Stelle. Diese Bezeichnung ist durch folgende Eigenschaft gerechtfertigt:

Satz 8. *Unter allen Kreisen, welche die Kurve an der Stelle s berühren, hat jener mit dem Krümmungsradius (als einziger) eine Berührung 2. Ordnung.*

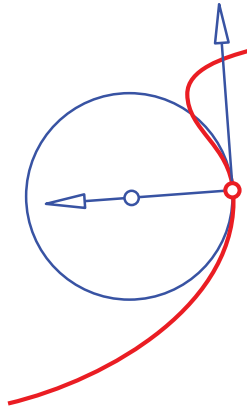


Abbildung 2.1: Begleitendes Bezugssystem und Krümmungskreis.

Beweis: Die genannten Kreise (ein sogenanntes *Kreisbündel* bildend) haben ihren Mittelpunkt auf den Normalen im Abstand r ihres Radius, also in lokalen Koordinaten ξ, η bezüglich des begleitenden Bezugssystem die Gleichung

$$\xi^2 + (\eta - r)^2 - r^2 = 0. \quad (2.5)$$

Andererseits können die lokalen Koordinaten der Kurve nach Taylor in Reihen (bzw. Polynome von s mit Restgliedern) entwickelt werden: Aufgrund der Ableitungsgleichungen ergibt sich

$$\begin{aligned}\xi &= s & -1/6 \cdot \kappa^2 s^3 & \dots[4] \\ \eta &= 1/2 \cdot \kappa s^2 & +1/6 \cdot \kappa' s^3 & \dots[4],\end{aligned}\tag{2.6}$$

wobei das Symbol $\dots[4]$ weitere Glieder mit mindestens 4. Potenz von s andeuten soll. Nun gilt folgende **Regel**: Wenn eine Kurve durch eine implizite Gleichung, die andere durch eine Taylor-Entwicklung nach ihrem Kurvenparameter s gegeben ist,¹ und man diese Entwicklungen in die linke der Gleichung einsetzt, so ist die Berührordnung gleich k , wenn alle Glieder bis zur Potenz k verschwinden (also nur noch solche mit höherer als der k -ten Potenz verbleiben).²

Das Einsetzen von (2.6) in (2.5) liefert

$$\xi^2 + \eta^2 - 2r\eta = (1 - r\kappa) \cdot s^2 \dots[3].\tag{2.7}$$

Also verschwinden genau dann die quadratischen Glieder, wenn

$$r = 1/\kappa.\tag{2.8}$$

□

Bemerkungen:

1. Der in Satz 8 genannte Kreis heißt *Krümmungskreis* der Kurve an der betrachteten Stelle, und sein Mittelpunkt wird *Krümmungsmittelpunkt* genannt.

¹ Der Kurvenparameter braucht i. Allg. nicht die Bogenlänge zu sein.

² Siehe hierzu [8].

2. Man kann zeigen, dass bei Kurvensegmenten, deren Krümmung streng monoton wächst, die Krümmungskreise jeweils ganz ineinander liegen (Der Beweis ist nicht einfach).
3. Bei oben genannter Regel muss natürlich der betreffende Kurvenpunkt $\mathbf{x}(s)$ die implizite Gleichung erfüllen (d. h. auf der 1. Kurve liegen) und diese Gleichung muss an der betreffenden Stelle nach mindestens einer Variablen auflösbar sein, damit sie wenigstens lokal eine Kurve darstellt.
4. Dass diese Regel mit der in Abschnitt 1.5 gegebenen Definition der Berührrordnung in Einklang steht, wird in [8] nachgewiesen.
5. Die lokalen Reihenentwicklungen (2.6) gelten nur dann, wenn die Kurve mindestens zweimal stetig differenzierbar ist. Bei höherer Differentiationsordnung können sie entsprechend fortgesetzt werden. Obwohl es keine Taylor-Reihen (sondern nur Taylor-Polynome mit Restgliedern) zu sein brauchen, spricht man häufig von „lokalen Reihenentwicklungen“.
6. Das Fehlen des Absolutglieds und desjenigen erster Ordnung in (2.7) bedeutet, dass alle Kreise des Büschels die Kurve im Punkt $\mathbf{x}(s)$ in 1. Ordnung berühren.
7. Wie man leicht nachrechnet, haben Kreise selbst immer die *konstante Krümmung* $1/r$. Insbesondere hat der Krümmungskreis die Krümmung κ , also den Radius $r = 1/\kappa$ (s. (2.8)).
8. Verschwindet an einer Stelle s_0 die Krümmung ($\kappa(s_0) = 0$), so sind die Vektoren $\mathbf{x}'(s_0)$ und $\mathbf{x}''(s_0)$ nach (2.4) linear abhängig, d. h. *es liegt ein Wendepunkt* vor.

9. Das Vorzeichen der Krümmung (sofern $\kappa(s_0) \neq 0$ ist) entscheidet darüber, wie sich die Kurve im Vergleich zur Tangente verhält: Bei $\kappa(s_0) > 0$ liegt sie lokal links von ihr, bei $\kappa(s_0) < 0$ rechts von ihr.
10. Bei $\kappa(s_0) \neq 0$ und $\kappa'(s_0) \neq 0$ kann man aus (2.7) entnehmen, dass die Kurve den Krümmungskreis lokal *durchsetzt*, d. h. entweder von seinem Inneren zu seinem Äußeren verläuft oder umgekehrt.
11. Kurvenstellen, an denen $\kappa'(s_0) = 0$ ist, (an denen also die Krümmung einen Extremwert annimmt) heißen *Scheitelpunkte* der Kurve. (Die Scheitelpunkte der Kegelschnitte sind auch Scheitelpunkte in diesem Sinne!)

2.2 Die Krümmung bei allgemeinem Parameter

Oftmals sind jedoch Kurven – wie schon erwähnt – *nicht* mit der Bogenlänge parametrisiert, und die Umparametrisierung ist mühevoll. Daher ist es wünschenswert, wenigstens die *Krümmung bei allgemeiner Parametrisierung* zu berechnen. Dies leistet die folgende Formel

$$\kappa = [\mathbf{x}', \mathbf{x}''] / \|\mathbf{x}'\|^3, \quad (2.9)$$

wobei das Symbol $[\dots]$ die Determinante der Matrix der darin stehenden Vektoren bedeutet.

Beweis: Man denke sich die Umparametrisierung auf die Bogenlänge gemäß (1.5) mit einer Funktion $t = f(s)$, $s \in \bar{I}$ bereits durchgeführt: $\bar{\mathbf{x}}(s) = \mathbf{x}(f(s))$. Dann gilt (unter Weglassung der Argumente)

$$\mathbf{t} = \dot{\bar{\mathbf{x}}} = \mathbf{x}' \cdot \dot{f}$$

und

$$\kappa \mathbf{n} = \ddot{\bar{\mathbf{x}}} = \mathbf{x}'' \cdot (\dot{f})^2 + \mathbf{x}' \cdot \ddot{f}$$

(links nach den Frenet-Formeln, rechts durch Ableitung nach der Kettenregel). Hieraus folgt (bei $\dot{f} > 0$, d. h. gleichsinniger Umparametrisierung) wegen $\|\bar{\mathbf{t}}\| = 1$

$$\dot{f} = 1/\|\mathbf{x}'\| \quad (2.10)$$

und

$$\kappa = [\dot{\bar{\mathbf{x}}}, \ddot{\bar{\mathbf{x}}}] = [\mathbf{x}' \dot{f}, \mathbf{x}'' (\dot{f})^2], \quad (2.11)$$

woraus nach (2.10) die Formel (2.9) folgt. \square

Die Krümmung kann auch für den G^2 -stetigen Anschluss herangezogen werden; es gilt nämlich

Satz 9. *Zwei Kurvensegmente, die einen G^1 -stetigen Anschluss (im Endpunkt des ersten Segments, der zugleich der Anfangspunkt des zweiten ist) haben dort sogar einen G^2 -Anschluss, wenn in diesem Punkt auch noch die Krümmungen übereinstimmen.*

Beweis: Die beiden Segmente können ohne Beschränkung der Allgemeinheit auf die Bogenlänge bezogen werden. Dann gelten (mit den Bezeichnungen wie in Abschnitt 1.5) wegen der G^1 -Stetigkeit die Beziehungen (1.12) mit $f_1 = 1$, denn die Parametertransformation ist lediglich die Verschiebung der Anfangs-

punkte $t - t_0 = s - s_0$; somit sind beide Vektoren gleich dem gemeinsamen Tangenten-Einheitsvektor $\mathbf{t}(t_0)$. Nach den Frenet-Formeln folgt

$$\mathbf{x}''(t_0) = \kappa(t_0) \mathbf{n}(t_0) \quad \text{und} \quad \mathbf{y}''(s_0) = \bar{\kappa}(s_0) \bar{\mathbf{n}}(s_0).$$

Nun sind die Vektoren $\mathbf{n}(t_0)$ und $\bar{\mathbf{n}}(s_0)$ ebenfalls einander gleich, da sie beide zusammen mit \mathbf{t} ein orthonormiertes Rechtssystem bilden. Also gilt die G^2 -Bedingung (1.12), 2. Ableitung, $\mathbf{x}''(t_0) = \mathbf{y}''(s_0)$ (mit $f_2 = 0$) genau dann, wenn $\kappa(t_0) = \bar{\kappa}(s_0)$ erfüllt ist. \square

2.3 Evoluten, Evolventen

Eine einparametrische Schar von Geraden in der Ebene hat im allgemeinen eine *Hüllkurve*. Dies gilt insbesondere auch für die Schar der Normalen einer Kurve.¹

Man beschreibt eine Schar von Geraden durch eine *Leitkurve*, welche die Geraden der Schar *transversal* schneidet, d. h. von jeder Geraden genau einen Punkt enthält und dort die Gerade nicht berührt. Ferner muss eine auf denselben Parameter bezogene Vektorfunktion gegeben sein, welche die Richtung der betreffenden Geraden angibt:

$$t \mapsto \mathbf{x}(t), \quad t \mapsto \mathbf{v}(t), \quad t \in I. \quad (2.12)$$

Für die Hüllkurve macht man den Ansatz

$$t \mapsto \mathbf{y}(t) = \mathbf{x}(t) + h(t) \cdot \mathbf{v}(t), \quad t \in I, \quad (2.13)$$

¹ „Schar“ wird synonym zu „Menge“ oder „Familie“ gebraucht.

und findet durch Ableiten (unter Weglassung der Argumente t)

$$\mathbf{y}' = \mathbf{x}' + h' \cdot \mathbf{v} + h \cdot \mathbf{v}'. \quad (2.14)$$

Nun war die Leitkurve transversal vorausgesetzt, d. h. die Vektoren \mathbf{x}' und \mathbf{v} sind (an jeder Stelle $t \in I$) linear unabhängig. Also existieren Funktionen $t \mapsto \alpha(t)$ und $t \mapsto \beta(t)$, mit

$$\mathbf{v}' = \alpha \cdot \mathbf{x}' + \beta \cdot \mathbf{v}. \quad (2.15)$$

Setzt man dies in (2.14) ein, so folgt

$$\mathbf{y}' = (1 + h \cdot \alpha) \cdot \mathbf{x}' + (h' + \beta) \cdot \mathbf{v}. \quad (2.16)$$

Daher berührt die von \mathbf{y} beschriebene Kurve genau dann die Geraden der Schar, wenn der Koeffizient bei \mathbf{x}' identisch verschwindet:

$$1 + h \cdot \alpha = 0. \quad (2.17)$$

Wenn also α im Intervall I keine Nullstelle hat, ist h durch (2.17) eindeutig bestimmt ($h = -1/\alpha$), und (2.13) beschreibt die Hüllkurve. Bei einer Nullstelle von α sind die Vektoren \mathbf{v}, \mathbf{v}' nach (2.15) linear abhängig, d. h. die Geraden der Schar sind dort „momentan parallel“, h geht in der Umgebung dieser Stelle gegen unendlich und die Hüllkurve hat eine Asymptote.

Soweit die allgemeine Theorie der Hüllkurve einer einparametrischen Geraden-schar in der Ebene. Nun zu der Schar der Normalen einer Kurve! Als Leitkurve wählen wir die Kurve selbst, parametrisieren sie nach ihrer Bogenlänge und bezeichnen sie wie im allgemeinen Fall mit \mathbf{x} . Nach (2.4) ist dann $\mathbf{x}' = \mathbf{t}$ und $\mathbf{v} = \mathbf{n}$. Da \mathbf{y}' ein Vielfaches von \mathbf{v} ist, ergibt sich aus (2.14) mit (2.4), dass $1 - h\kappa = 0$, und weiter $h = 1/\kappa = r$. Dies bedeutet:

Satz 10. *Die Hüllkurve der Normalen eines glatten Kurvenstücks ohne Wendepunkte ist der geometrische Ort ihrer Krümmungsmittelpunkte.*

Diese Hüllkurve heißt *Evolute* der ursprünglichen Kurve (s. Abb. 2.2). Die Darstellung der Evolute lautet also

$$\mathbf{e}(s) = \mathbf{x}(s) + r(s) \cdot \mathbf{n}(s) \quad (2.18)$$

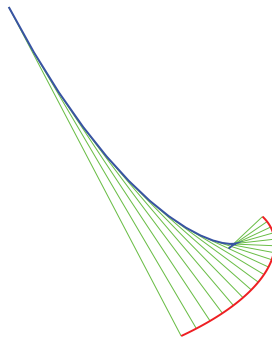


Abbildung 2.2: Die Evolute (blau) einer Kurve (rot).

Die Evolute einer Kurve hat noch eine andere interessante Eigenschaft; es gilt nämlich der

Satz 11. *(Chr. Huygens, 1673) Die Länge eines Teilbogens der Evolute ist gleich der Differenz der Krümmungsradien an den entsprechenden Stellen der Ausgangskurve, sofern diese dazwischen keinen Scheitelpunkt hat.*

Beweis: Durch Ableitung von (2.18) folgt

$$\mathbf{e}' = r' \cdot \mathbf{n},$$

da die Glieder mit \mathbf{t} nach den Frenet-Formeln und wegen (2.8) herausfallen. Also ist $\|\mathbf{e}'\| = r'$ (sofern r' positiv ist) und somit erhält man die Länge des Bogens nach (1.11) zu

$$\int_{s_1}^{s_2} \mathbf{e}(s) = r(s_2) - r(s_1). \quad \square \quad (2.19)$$

Man kann sich diesen Sachverhalt dadurch veranschaulichen, dass man sich die Evolute als eine Rolle vorstellt, wobei in deren Endpunkt ein Faden befestigt wird; wickelt man diesen bis zum Anfangspunkt der Evolute auf und gibt noch den Krümmungsradius hinzu, so gelangt man zum Anfangspunkt der ursprünglichen Kurve \mathcal{C} . Hält man nun den Faden gespannt und wickelt ihn wieder ab, so beschreibt das Ende des Fadens den gesamten Bogen der Kurve \mathcal{C} .

Die abgewickelte Kurve nennt man *Evolvente*; die Kurve \mathcal{C} ist also die Evolvente ihrer eigenen Evolute. (Die Bezeichnungen „Evolute“ und „Evolvente“ kommen von lat. *evolvere* = abwickeln und wurden ebenfalls von Huygens eingeführt.)

2.4 Natürliche Gleichung, Integration

Während man bei der Herleitung der Frenet-Formeln von einer Kurve (in Bogenlängen-Parametrisierung) ausgeht und zur Krümmung gelangt, kann man sich umgekehrt fragen, ob es zu einer gegebenen Funktion

$$s \mapsto \kappa(s), \quad s \in I, \quad (2.20)$$

eine Kurve gibt, welche diese als Krümmungsfunktion besitzt. Diese Frage ist in der Tat zu bejahen, denn man braucht dazu nur die Frenet-Formeln als

lineares System von Differentialgleichungen für die Koordinatenfunktionen $s \mapsto x(s)$, $s \in I$, und $s \mapsto y(s)$, $s \in I$, aufzufassen und die diesbezüglichen Existenzsätze anzuwenden. Die einzige Voraussetzung hierzu ist die *Stetigkeit* der Funktion (2.20).

Zur Integration der Frenet-Formeln geht man davon aus, dass $\mathbf{x}'(s)$ ein Einheitsvektor ist; daher lassen sich dessen Komponenten in der Form

$$x'(s) = \cos(\phi(s)), \quad y'(s) = \sin(\phi(s)) \quad (2.21)$$

darstellen, wobei $\phi(s)$ der (gegen die positive x -Achse und im Bogenmaß gemessene) Winkel der Tangente ist. Durch Ableitung nach s folgt daraus

$$x''(s) = -\sin(\phi(s)) \cdot \phi'(s), \quad y''(s) = \cos(\phi(s)) \cdot \phi'(s). \quad (2.22)$$

Durch Vergleich mit (2.4) erhält man

$$\frac{d\phi(s)}{ds} = \kappa(s), \quad (2.23)$$

also

$$\phi(s) = \int_0^s \kappa(\sigma) d\sigma. \quad (2.24)$$

Ist diese erste Integration einmal ausgeführt, setzt man die erhaltene Funktion $\phi(s)$ in (2.23) ein und integriert auch diese Gleichungen:

$$x(s) = \int_0^s \cos(\phi(\sigma)) d\sigma, \quad y(s) = \int_0^s \sin(\phi(\sigma)) d\sigma. \quad (2.25)$$

Auf diese Weise gewinnt man eine explizite Darstellung der Kurve, welche (2.20) als Krümmungsfunktion besitzt.

Die bei diesen Integrationen auftretenden Konstanten arrangieren sich, wie man aufgrund der Additionstheoreme der trigonometrischen Funktionen so-

fort erkennt, zu einer noch willkürlichen festen Bewegung in der Ebene; d. h. es gilt der

Satz 12. *Zu einer gegebenen stetigen Funktion (2.20) existiert bis auf eine Bewegung in der Ebene eindeutig eine Kurve, welche diese als Krümmungsfunktion besitzt, und diese Kurve ist durch (2.23) und (2.25) dargestellt.*

2.5 Beispiele und Anwendungen

Bei Modelleisenbahnen werden meistens (kreisförmig) gekrümmte Schienen unmittelbar an gerade Schienen (und umgekehrt) angeschlossen. Das geht bei wirklichen Eisenbahnen ganz und gar nicht, weil durch die Unstetigkeit der Fliehkraft (die mit dem Quadrat der Krümmung einhergeht!) ein zu starker *Ruck* entstehen würde. (Das Analoge gilt natürlich auch im Straßenbau.) Also muss man ein *Übergangsstück* einfügen, bei welchem die Krümmung *stetig* von Null auf $\kappa = 1/r$ anwächst (wobei r der vorgesehene Bogenradius ist).

Hierzu kann man die sogenannte *Klothoide* (siehe Abb. 2.3) verwenden, die dadurch definiert ist, dass ihre Krümmung eine lineare Funktion der Bogenlänge ist. Den Anfangspunkt der Bogenlängenmessung wählt man bei der Krümmung Null und die noch frei verfügbare Konstante (aus Dimensionsgründen) so, dass gilt:

$$\kappa(s) = 1/a^2 \cdot s \quad (2.26)$$

Wie man dem vorigen Abschnitt entnimmt, lässt sich die *natürliche Gleichung* (2.26) – wenn auch nicht mit elementaren Funktionen, jedoch mit Reihenentwicklungen oder geeigneten Näherungsverfahren – integrieren, und es kann daraus ein passendes Übergangsstück entnommen werden.

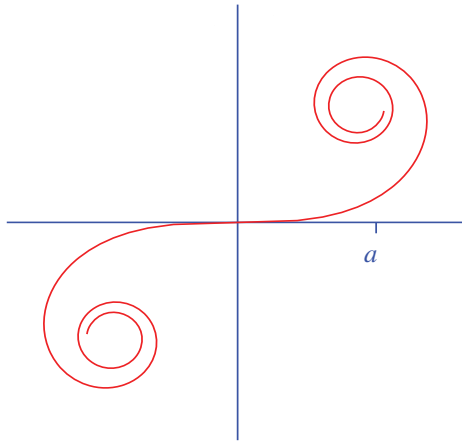


Abbildung 2.3: Die Klothoide.

Man kann in beiden Fällen auch kubische Kurven, z. B. in Bézier-Darstellung (s. Kapitel 6), verwenden. Heutzutage sind fertige Lösungen solcher Probleme meist in den einschlägigen Softwareprodukten integriert, so dass man nur noch die Daten einzugeben hat.

Für manche Zwecke ist es günstiger, eine Kurve in *Polarkoordinaten* statt in vektorieller Form darzustellen. Hierbei wird r (was nun die 1. Polarkoordinate, nicht mehr den Krümmungsradius bezeichnet) als Funktion des Polarwinkels ϕ angegeben. Offenbar ist z. B. der Kreis durch eine *konstante* Funktion ($r = \text{const}$) gegeben.

Auch die *Kegelschnitte* lassen sich in Polarkoordinaten auf einfache Weise darstellen, wenn der Ursprung in einen der Brennpunkte gelegt wird (man vgl. die

sogenannte Scheitelgleichungen der Kegelschnitte). Durch Umrechnung erhält man:

$$r = \frac{p}{1 - \varepsilon \phi}, \quad (2.27)$$

wobei p der *Parameter*, d. h. die Ordinate im *Brennpunkt*, und ε die *numerische Exzentrizität* bedeuten und $p > 0$ sein muss. (Für $0 < \varepsilon < 1$ handelt es sich um Ellipsen, bei $\varepsilon = 1$ um Parabeln und bei $\varepsilon > 1$ um Hyperbeln).

Ein Musterbeispiel für die Darstellung einer Kurve in Polarkoordinaten ist die *logarithmische Spirale*

$$r = a \cdot e^{c\phi}. \quad (2.28)$$

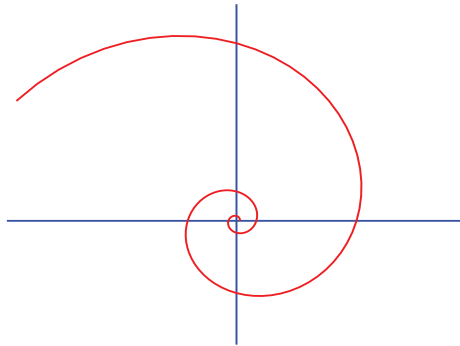


Abbildung 2.4: Die logarithmische Spirale.

Sie hat die (kennzeichnende) Eigenschaft, dass sie die Strahlen (Halbgeraden) durch den Ursprung *unter konstanten Winkeln* schneidet. Dies ist eine unmittelbare Konsequenz aus folgendem

Satz 13. *Jede Drehstreckung*

$$\begin{aligned}x &= p(\cos(\alpha)\bar{x} + \sin(\alpha)\bar{y}) \\ y &= p(-\sin(\alpha)\bar{x} + \cos(\alpha)\bar{y})\end{aligned}\tag{2.29}$$

mit $p = a \cdot e^{c\alpha}$ führt die *logarithmische Spirale* (2.28) in sich über:

Beweis: Es gilt nach (2.29) – wie man leicht nachrechnet – $r = \sqrt{x^2 + y^2} = p \cdot \bar{r}$ und $\phi = \bar{\phi} + \alpha$; somit geht (2.28) in $p \cdot \bar{r} = a \cdot e^{c(\bar{\phi} + \alpha)}$ über. Mit $p = a \cdot e^{c\alpha}$ entsteht dieselbe Gleichung wie (2.28). \square

3 Geschlossene Kurven in der Ebene

3.1 Jordankurven

Man nennt eine ebene Kurve *geschlossen*, wenn sie das Bild eines kompakten Intervalls ist, dessen Anfangspunkt mit dem Endpunkt übereinstimmt. Wenn sie darüber hinaus eine *stetige* Parametrisierung besitzt, die – bis auf die Gleichheit von Anfangspunkt und Endpunkt – *umkehrbar eindeutig* ist, wird sie eine *Jordankurve* genannt.¹ Hierüber gilt der

Satz 14. (*Jordanscher Kurvensatz*): *Eine (geschlossene) Jordankurve teilt die euklidische Ebene in drei disjunkte Punktmengen ein: Das Innere, das Äußere der Kurve und die Punkte der Kurve selbst. Dabei ist das Innere eine beschränkte Menge, das Äußere dagegen nicht; beide Mengen sind offen.*

Bemerkungen

1. Obwohl dieser Satz anschaulich ohne jeden Zweifel gültig erscheint, ist ein strenger mathematischer Beweis nicht ganz einfach.

¹ Nach C. Jordan (1838 - 1922).

2. Eine geschlossene Kurve, die bijektiv auf einen Kreis abgebildet werden kann, heißt *einfach geschlossen*; eine solche Kurve hat also keine Selbstüberschneidungen oder Selbstberührungen.
3. Ein stetiges und bijektives Bild eines kompakten Intervalls (also *keine* geschlossene Kurve) heißt *Jordanbogen*.
4. Anstelle eines kompakten Intervalls und der Zusatzbedingung, dass der Anfangspunkt gleich dem Endpunkt ist, kann man in der Definition einer Jordankurve auch eine *stetige und periodische* Abbildung $\mathbb{R} \rightarrow \mathbb{E}^2$ – etwa mit der Periode 1 – setzen, welche in einem (also auch in jedem anderen) *Fundamentalintervall* $[a, a + 1) \subset \mathbb{R}$ bijektiv ist. (Diese Definition vermeidet den Anschein, dass der Anfangs- oder Endpunkt irgendeine Sonderrolle besitzt.)
5. Geschlossene Kurven, die *keine* Jordankurven sind, können alle Arten von Selbstüberschneidungen, Selbstberührungen oder gar gemeinsamen Teilbögen besitzen. Wenn eine solche Kurve *stetig* ist, kann man jedem Punkt der Ebene, der *nicht auf der Kurve liegt* auf folgende Weise eine *Umlaufszahl* zuordnen:

$$\omega(\mathbf{p}) = \frac{1}{2\pi} \int_{t=a}^b d\phi.$$

Dabei ist $\phi(t)$ der (positiv orientierte) Winkel des „Fahrhalbstrahls“ von \mathbf{p} zum Kurvenpunkt $\mathbf{x}(t)$. Diese Umlaufszahl ist stets eine *ganze Zahl*.

6. Der Zusammenhang mit Jordankurven ist folgender: Wenn die Jordankurve im mathematisch positiven Sinn durchlaufen wird, kommen nur die Umlaufszahlen 1 und 0 vor, wobei 1 genau für die Punkte im Inneren und 0 genau für jene im Äußeren zutrifft. (Wie H. Hopf bemerkt hat, ist dies

schon der wesentliche Beweisschritt des Jordanschen Kurvensatzes, denn bei stetigen geschlossenen Kurven, die keine Jordankurven sind, kann die Umlaufszahl eine beliebige ganze Zahl sein.)

3.2 Flächeninhalt und Isoperimetrie

Die Vereinigung des Inneren einer Jordankurve mit den Punkten der Kurve selbst ist stets ein kompakter Bereich B der Ebene. Diesem kann man einen *Flächeninhalt* (auch „Jordan-Inhalt“ genannt) zuordnen. Dies geschieht dadurch, dass man eine Schar paralleler Geraden $g(x)$ in der Ebene wählt, auf jeder von ihnen die Gesamtlänge $L(g)$ der Strecken $g \cap B$ bestimmt und dieses integriert:

$$F = \int_{xu}^{xo} L(g(x)) dx.$$

(xu, xo sind die x -Werte der ersten und letzten Geraden, die mit B einen nicht-leeren Durchschnitt haben). Dass das Integral existiert, folgt aus der Stetigkeit der Jordankurve; es kommt aber darauf an zu zeigen, dass dieses Integral *unabhängig von der gewählten Richtung* der Geradenschar ist. Darüber hinaus muss gezeigt werden, dass der Wert F sich nicht ändert, wenn man auf den Bereich eine Bewegung anwendet (*Bewegungsinvarianz des Flächeninhalts*).

Diesem Inhaltsbegriff für krummlinig begrenzte Bereiche liegt die Vorstellung zugrunde, die Fläche dadurch zu approximieren, dass man den Bereich in (viele) schmale Streifen zerlegt und deren Inhalte aufsummiert. In der Grenze, wo die Streifenbreite gegen Null konvergiert, geht auch der Fehler, der von der Ersetzung der krummen oberen und unteren Enden des Streifens durch geradlinige Strecken herrührt, aufgrund der Stetigkeit gegen Null.

Eine andere Möglichkeit, den Inhalt zu berechnen, besteht darin, einen beliebigen inneren Punkt \mathbf{p} zu wählen und die Inhalte aller *dreieckigen* „Sektoren“ $\Delta(\mathbf{p}, \mathbf{x}(t), \mathbf{x}(t + dt))$ aufsummieren; dies führt auf die Inhaltsformel

$$F = \frac{1}{2} \int_{t=t_0}^{t=t_1} \|\mathbf{x}(t) - \mathbf{p}\|^2 \phi'(t) dt,$$

wobei das Integral über einen vollen Umlauf der Parameterdarstellung der Kurve zu nehmen ist und $\phi(t)$ der im positiven Sinn zu messende Winkel des „Fahrstrahls“ $\mathbf{px}(t)$ bedeutet. Diese Darstellung hat den Vorteil, dass man keine weiteren Schnittpunkte zu berechnen braucht: Wenn der Fahrstrahl vor- und zurück läuft, werden die nicht zum Bereich gehörende Zwischenstücke automatisch negativ genommen. (Allerdings setzt diese Formel voraus, dass die Funktion $t \mapsto \phi(t)$ differenzierbar ist.)

Näheres zum Inhaltsbegriff findet man in den Büchern über Analysis, etwa [25] oder – ausführlicher und mit Beweisen der Invarianzeigenschaften des Inhalts in [15].

In diesem Zusammenhang muss auch die berühmte *isoperimetrische Eigenschaft* des Kreises erwähnt werden:

Satz 15. *Unter allen ebenen Bereichen, die von einer rektifizierbaren Jordankurve der festen Länge L umschlossen werden, hat der Kreis den größten Flächeninhalt.*

Das „isoperimetrische Problem“ wurde schon im Altertum diskutiert: Nach der griechischen Sage wurde der Königin Dido (um 900 v. Chr.), der Gründerin von

Karthago, so viel Land zugesprochen, wie sie an einem Tage mit einem Ritt auf einem Esel umschließen könne.¹

Obwohl die isoperimetrische Eigenschaft des Kreises anschaulich ohne Zweifel richtig erscheint, ist ein strenger Beweis erst relativ spät dann aber von mehreren Autoren unter verschiedenen Voraussetzungen und mittels unterschiedlicher Methoden geliefert worden (siehe z. B. [2], [1]).

3.3 Konvexität

Beim Begriff der Konvexität geht man nicht von Kurven, sondern von *beliebigen Punktmengen* im \mathbb{R}^d aus:

Definition 16. Eine Punktmenge M des \mathbb{R}^d heißt *konvex*, wenn sie mit je zwei ihrer Punkte $\mathbf{a}, \mathbf{b} \in M$ auch die gesamte Verbindungsstrecke enthält.

Eine der wichtigsten Eigenschaften konvexer Mengen ist:

Satz 17. Der Durchschnitt je zweier konvexer Mengen ist konvex.

Dies folgt unmittelbar aus der Definition. (Beachte, dass die leere Menge konvex ist!)

Über die Theorie konvexer Mengen gibt es eine umfangreiche Literatur, auf welche lediglich durch das Zitat [19] hingewiesen wird. Hier interessieren konvexe Mengen *in der Ebene* nur deshalb, weil das Innere einer Jordankurve eine

¹ Daher nennt man es auch „Dido’sches Problem“.

konvexe Menge sein kann. In diesem Fall ist dann auch die Vereinigung des Inneren mit der Kurve selbst eine (kompakte) konvexe Menge. Man spricht dann von einem *ebenen konvexen Bereich*. (Eine solche Kurve heißt gelegentlich auch eine „Eilinie“ und der Bereich ein „Eibereich“ .

Umgekehrt sei darauf hingewiesen, dass eine konvexe Menge M in der Ebene nicht notwendigerweise von einer Jordankurve begrenzt ist; selbst dann nicht, wenn sie beschränkt ist. (Die topologische Struktur der Menge M kann sehr kompliziert sein!) Erst wenn man die Bedingung der Abgeschlossenheit von M hinzunimmt oder die (abgeschlossene) Menge *aller* Randpunkte (nicht nur die zu M gehörenden) betrachtet, ist die obige Aussage gültig.

Über konvexe Jordankurven, welche mindestens 3-mal stetig differenzierbar sind, gilt eine weitere globale Eigenschaft, die man den „Vierscheitelsatz“ nennt:

Satz 18. *Eine konvexe C^3 -Jordankurve besitzt mindestens vier Extremscheitel.*¹

Einen Beweis findet man z. B. in [6].

¹ „Extremscheitel“ soll hier andeuten, dass an dieser Stelle nicht nur $\kappa' = 0$ ist, sondern die Krümmung einen Extremwert besitzt.

4 Kurven im euklidischen Raum

Für Raumkurven verhält sich vieles analog wie für ebene Kurven; insbesondere gilt das über Parametertransformationen, Differenzierbarkeit und Regularität. Gesagte genauso, wobei die Vektoren nun drei Komponenten haben. Es kommt jedoch eine weitere Bedingung hinzu:

Eindeutigkeit der Schmiegeebene, begleitendes Bezugssystem. *Es wird vorausgesetzt, dass die Parameterdarstellung mindestens zweimal stetig differenzierbar ist und dass im gesamten Intervall I*

\mathbf{x}' und \mathbf{x}'' linear unabhängig voneinander sind.

Dann definieren der Kurvenpunkt und diese beiden Vektoren an jeder Stelle $t \in I$ eine Ebene, welche *die Schmiegeebene* genannt wird. In dieser Schmiegeebene wählt man diejenige Gerade, welche auf der Tangenten senkrecht steht und nennt sie die *Normale*. Diese Konfiguration wird durch die *Binormale* ergänzt, d.h. durch die Gerade, die auf der Schmiegeebene senkrecht steht und durch den Kurvenpunkt geht. Damit hat man wieder ein begleitendes Koordinatensystem mit Ursprung im Kurvenpunkt, wenn noch die Richtung des Normalenvektors so festgelegt wird, dass \mathbf{x}'' auf der positiven Seite der Tangente (in der Schmiegeebene) liegt.

4.1 Frenet-Formeln

Die Einheitsvektoren (Vektorfunktionen!) in den Richtungen der (orientierten) Tangente, Normale und Binormale werden mit \mathbf{t}, \mathbf{n} bzw. \mathbf{b} bezeichnet, wobei die Richtung von \mathbf{b} so fixiert wird, dass ein orthonormiertes Rechtssystem entsteht.

Wenn die Kurve mit der Bogenlänge parametrisiert wird, gelten die folgenden *Frenet-Formeln für Raumkurven*:

$$\begin{aligned}\mathbf{x}' &= \mathbf{t} \\ \mathbf{t}' &= \kappa \mathbf{n} \\ \mathbf{n}' &= -\kappa \mathbf{t} + \tau \mathbf{b} \\ \mathbf{b}' &= -\tau \mathbf{n}.\end{aligned}\tag{4.1}$$

Bei der genannten Orientierung ist $\kappa > 0$ und wegen der Orthogonalität der Basisvektoren ist die Koeffizientenmatrix der Ableitungen schiefsymmetrisch; es bleiben daher nur zwei Größen, κ und τ , übrig. Die erste von ihnen ist – wie im ebenen Fall – die *Krümmung* der Raumkurve; die zweite, τ , wird *Torsion* genannt. Wenn diese an einer (isolierten) Stelle verschwindet, ist die Schmiegeebene *momentan stationär*. (Sie ändert ihren Drehsinn um die Tangente.) Bei $\tau > 0$ windet sich die Kurve wie eine Rechtsschraube (wie die Weinrebe), bei $\tau < 0$ wie eine Linksschraube (wie der Hopfen). Ist in einem (Teil-)Intervall die Torsion identisch Null, so liegt ein *ebener Bogen* vor. (Die Frenet-Formeln gehen in der Tat in jene für ebene Kurven über.)

4.2 Natürliche Gleichungen

Wie auch im ebenen Fall bestimmen Krümmung und Torsion (als gegebene Funktionen der Bogenlänge)

$$\kappa = \kappa(s), \quad \tau = \tau(s), \quad s \in I,$$

eine Raumkurve bis auf eine Bewegung eindeutig; der Grund hierfür ist der gleiche: man fasst die Frenet-Formeln als lineares System von gewöhnlichen Differentialgleichungen für die (unbekannten) Komponenten-Funktionen des Kurvenvektors \mathbf{x} auf.

Die Schraublinie (Helix)

Eine der einfachsten Raumkurven ist die Schraublinie; sie ist durch *konstante Funktionen* der Krümmung und der Torsion gekennzeichnet. Die Frenet-Gleichungen lassen sich in diesem Fall leicht integrieren und führen auf die Parameterdarstellung

$$\mathbf{x}(t) = \begin{pmatrix} a \cos(t) \\ a \sin(t) \\ pt \end{pmatrix}, \quad (4.2)$$

wobei $s = \sqrt{a^2 + p^2}t$.

Die Schraublinie liegt auf einem geraden Kreiszylinder, dessen Achse mit der z -Achse des bei (4.2) zugrunde gelegten Koordinatensystems zusammenfällt; der Radius dieses Zylinders ist a . Wie der Name besagt, besitzt die Schraublinie die Eigenschaft, dass sie durch die einparametrische Gruppe von *Schraubun-*

gen (einer Untergruppe der euklidischen Bewegungsgruppe) in sich übergeführt wird:

$$\begin{aligned}\bar{x} &= \cos(\alpha)x - \sin(\alpha)y \\ \bar{y} &= \sin(\alpha)x + \cos(\alpha)y \\ \bar{z} &= z + c\end{aligned}\tag{4.3}$$

(mit $c = p\alpha$). Nach (4.2) ist nämlich

$$x(t + \alpha) = a \cos(t + \alpha) = a(\cos(t) \cos(\alpha) - \sin(t) \sin(\alpha)) = \bar{x}(t),$$

$$y(t + \alpha) = a \sin(t + \alpha) = a(\sin(t) \cos(\alpha) + \cos(t) \sin(\alpha)) = \bar{y}(t),$$

und

$$z(t + \alpha) = p(t + \alpha) = z + c = \bar{z}(t).$$

Also liegt der Bildpunkt auf derselben Schraublinie, wobei sich nur der Parameter t um α erhöht hat. Aus der letzten Gleichung (4.3) ergibt sich wegen $c = p\alpha$ noch, dass sich bei einer vollen Umdrehung die ganze Kurve um $h = 2\pi p$ in Richtung der z -Achse verschiebt. Die Größe h heißt die *Ganghöhe* der Schraublinie.

5 Polynomiale Kurven

Polynome sind nun einmal die einfachsten Funktionen, und daher ist es naheliegend, diese auch zu Parameterdarstellungen von Kurven zu verwenden. Allerdings neigen aber polynomiale Funktionen – wie schon erwähnt – bei höheren Graden oder größeren Werten ihrer Argumente zu sehr großen Funktionswerten und erheblichen Schwankungen, so dass sie beim praktischen Kurvendesign durch günstigere Methoden ersetzt wurden. Andererseits kann man aber Methoden und Ergebnisse, die für *Funktionen* entwickelt wurden, auf jede Komponente einer vektoriellen Parameterdarstellung anwenden und so direkt für *Kurven* nutzbar machen. Dies gilt insbesondere für die Interpolation und Approximation (s. Abschnitt 9.2).

5.1 Definition der polynomialen Kurven

Definition 19. Eine Kurve im \mathbb{R}^d mit einer Parameterdarstellung

$$\mathbf{x}(t) = \sum_{k=0}^n t^k \mathbf{a}_k \quad (5.1)$$

heißt *polynomiale Kurve*.

Hierbei gilt:

- \mathbf{a}_0 ist ein *Punkt* (Anfangspunkt für $t = 0$), die anderen fetten Buchstaben \mathbf{a}_k bezeichnen *Vektoren*; insbesondere ist \mathbf{a}_1 der Tangentenvektor im Anfangspunkt.
- Das Parameterintervall ist i. Allg. nicht $[0, 1]$, es kann ein beliebiges – auch ein ein- oder zweiseitig unbeschränktes (bis ∞ oder bis $-\infty$ gehendes) Intervall sein.
- Wenn $\mathbf{a}_n \neq 0$ ist, heißt n der *Grad* der Kurve; bei $n = 0$ entartet die „Kurve“ in einen einzigen Punkt.
- Ersetzt man in (5.1) den Parameter t durch ein Polynom $t = f(s)$ vom Grad $k > 1$, so entsteht eine Parameterdarstellung vom Grad $k * n$ (welcher man diesen („überhöhten Grad“ nur schwerlich ansehen kann). Der „wahre Grad“ ist also die *minimale Zahl* n , mit welcher sich die Kurve in der Form (5.1) darstellen lässt.
- Für $t \rightarrow \infty$ läuft auch die Kurve „ins Unendliche“ (sofern der Grad > 0 ist); ebenso für $t \rightarrow -\infty$.

Parametertransformationen bei polynomialen Kurven

Es war in Abschnitt 1.3 darauf hingewiesen worden, dass die Parametrisierung einer Kurve kein wesentlicher Bestandteil der Kurve ist, sondern weitgehend willkürlich gewählt werden kann. Wenn allerdings die polynomiale Form gewahrt werden und der Grad gleich bleiben soll, sind nur die linearen Transformationen $t = a + b \cdot \tilde{t}$ zulässig.

5.2 Kubische Hermite-Splines

Wie am Beginn dieses Kapitels erwähnt, kann man Methoden der Interpolation und Approximation auf polynomiale Kurven komponentenweise übertragen. Dies soll am Beispiel der Hermite-Interpolation erläutert werden.

Gegeben ist ein Intervall $[a, b] \subset \mathbb{R}$ und darin eine Knotenfolge

$$a = t_0 < t_1 < \dots < t_{m-1} < t_m = b,$$

ferner eine Folge von Punkten \mathbf{P}_i ($i = 0, 1, \dots, m$) und schließlich noch eine Folge von Tangentenvektoren \mathbf{P}'_i ($i = 0, 1, \dots, m$) im \mathbb{R}^d .

Gesucht wird ein stückweise polynomialer kubischer C^1 -Spline

$$\mathbf{X} : t \in [a, b] \mapsto \mathbf{X}(t) \in \mathbb{R}^d,$$

mit den Knoten t_i , welcher für $t = t_i$ durch den Punkt \mathbf{P}_i hindurch läuft und dort den Tangentenvektor \mathbf{P}'_i hat (Übergangsstellen nur an den gegebenen Knoten). Jedes Kurvensegment

$$\mathbf{H}_i : t \in [t_i, t_{i+1}] \mapsto \mathbf{X}(t) \in \mathbb{R}^d, \quad i = 0, \dots, m-1,$$

ist also eine reine *polynomiale Kurve*, und die oben genannten Forderungen können dadurch (und auch nur auf diese Weise!) erfüllt werden, dass

$$\mathbf{H}_i(t_i) = \mathbf{P}_i, \quad \mathbf{H}_i(t_{i+1}) = \mathbf{P}_{i+1}, \quad \mathbf{H}'_i(t_i) = \mathbf{P}'_i, \quad \mathbf{H}'_i(t_{i+1}) = \mathbf{P}'_{i+1} \quad (5.2)$$

gilt. Diese Bedingungen führen mit dem Ansatz

$$\mathbf{H}_i(t) = h_{i,1}(t)\mathbf{P}_i + h_{i,2}(t)\mathbf{P}_{i+1} + h_{i,3}(t)\mathbf{P}'_i + h_{i,4}(t)\mathbf{P}'_{i+1} \quad (5.3)$$

auf die sogenannten *Hermite-Basisfunktionen*

$$\begin{aligned}
 h_{i,1}(t) &= 2 \frac{(t-t_i)^3}{\Delta_i^3} - 3 \frac{(t-t_i)^2}{\Delta_i^2} + 1, \\
 h_{i,2}(t) &= -2 \frac{(t-t_i)^3}{\Delta_i^3} + 3 \frac{(t-t_i)^2}{\Delta_i^2}, \\
 h_{i,3}(t) &= \frac{(t-t_i)^3}{\Delta_i^3} - 2 \frac{(t-t_i)^2}{\Delta_i^2} + (t-t_i), \\
 h_{i,4}(t) &= 2 \frac{(t-t_i)^3}{\Delta_i^3} - 3 \frac{(t-t_i)^2}{\Delta_i^2},
 \end{aligned} \tag{5.4}$$

womit das Problem gelöst ist. Dasselbe gilt auch für die globale Aufgabe, da dieses auf die Bestimmung der einzelnen Segmente reduziert werden kann.

Wenn man noch die Werte der Funktionen \mathbf{H}_i jeweils außerhalb des Intervalls $(t_i, t_{i+1}]$ auf Null setzt und \mathbf{H}_0 außerhalb des Intervalls $[t_0, t_1]$, kann man die Lösung in der Form

$$\mathbf{X} = \sum_{i=0}^{m-1} \mathbf{H}_i$$

schreiben (wobei für jeden Wert von $t \in [a, b]$ immer genau nur *ein Summand* zur Auswertung beiträgt).

Dieses Verfahren ist zwar recht einfach und liefert einen C^1 -Spline, verlangt aber die Vorgabe aller Ableitungsvektoren (nicht nur deren Richtungen!) an den Knoten. Die Gestalt der Kurve hängt wesentlich von deren Längen ab, und diese Abhängigkeit ist schwer im Vorhinein abzuschätzen.

Trotz dieses schwerwiegenden Nachteils kann man die Übertragung von Methoden für Splinefunktionen auf Splinekurven noch einen Schritt weiter treiben, indem man auf die Vorgabe der Ableitungsvektoren verzichtet und statt dessen einen C^2 -Übergang an den inneren Knoten verlangt.

Durch zweimalige Ableitung von (5.3) und Vergleich von $\mathbf{H}_i''(t_{i+1})$ mit $\mathbf{H}_{i+1}''(t_{i+1})$ folgt für die Komponenten von \mathbf{P}_i' je ein lineares Gleichungssystem.¹ Deren Matrizen sind alle gleich und *tridiagonal*, d. h. sie haben nur in der Diagonale und in den beiden Schrägzeilen darüber und darunter von Null verschiedene Elemente; ferner sind sie diagonal dominant und somit stets eindeutig lösbar.

Auf die weiteren Details wird hier verzichtet, da man nur die entsprechenden Formeln der Splinefunktion umzuschreiben hat (s. z. B. [16]).

5.3 Kurven mit pythagoräischen Hodographen

Eine interessante Frage wurde von R. Farouki [13, 12, Kap. 17] angeschnitten: Gibt es polynomiale Kurven, bei welchen das quadratische Bogenelement $(x'(t))^2 + (y'(t))^2$ ein vollständiges Quadrat ist? – Dann ist die Ableitung ds/dt der in (1.11) definierten Bogenlänge ein Polynom und daher die Bogenlänge selbst ebenso. Man nennt solche Kurven *Pythagoräische-Hodograph-Kurven* oder kurz *PH-Kurven*.

Beim Grad $n = 2$ findet man leicht, dass es (außer den „falsch parametrisierten“ Geraden $\mathbf{x}(t) = \mathbf{a} + \mathbf{b} \cdot (t + \lambda t^2)$) keine PH-Kurven gibt. Bei den Kubiken wird die Suche erstmals erfolgreich: $x = 1 - t^2$, $y = \frac{1}{\sqrt{3}} t (1 - t^2)$ stellt eine PH-Kurve dar, denn es ist $(x'(t))^2 + (y'(t))^2 = 4t^2 + 1/3 (1 - 6t^2 + 9t^4) = 1/3 (1 + 3t^2)^2$. Diese Kurve wird *Tschirnhausen-Kubik* genannt. Sie ist bis auf den Faktor $\frac{1}{\sqrt{3}}$ identisch mit dem Newtonschen Knoten (s. Abb. 9.2 in 9.6). Die Tschirnhausen-Kubik ist (bis auf Bewegungen) die einzige PH-Kubik. Bei

¹ An den Enden bleiben zwei Bedingungen frei wählbar, etwa $\mathbf{H}_0''(t_0) = \mathbf{H}_{m-1}''(t_m) = 0$.

höheren Graden gibt es eine reichhaltigere Menge von PH-Kurven. Näheres findet man in [12, Kap. 17].

6 Rationale Kurven

6.1 Allgemeine rationale Kurven

Viele, auch sehr einfache Kurven wie z. B. den Kreis kann man nicht als eine polynomiale Kurve darstellen. Daher hat man die Klasse der polynomialen Kurven wie folgt erweitert:

Definition 20. Eine rationale Kurve vom Grad $n > 0$ ist eine solche, die eine Parameter-Darstellung der Art

$$\mathbf{x}(t) = \frac{1}{\sum_{k=0}^n \gamma_k t^k} \cdot \left(\sum_{k=0}^n \mathbf{a}_k t^k \right), \quad t \in I, \quad (6.1)$$

besitzt. Der Parameter durchläuft dabei ein beliebiges Intervall I ($I = \mathbb{R}$ zulässig). Damit der Grad wirklich n ist, muss $\gamma_n \neq 0$ oder $\mathbf{a}_n \neq \mathbf{0}$ sein (oder beides) und n muss die kleinste Zahl sein, mit der eine Darstellung (6.1) möglich ist.

Der (Einheits-)Kreis lässt sich nun (bis auf den Punkt $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$) in der Form

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \frac{1}{1+t^2} \cdot \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 0 \end{pmatrix} \cdot t + \begin{pmatrix} 0 \\ -1 \end{pmatrix} \cdot t^2 \right), \quad t \in \mathbb{R}, \quad (6.2)$$

darstellen, was natürlich dasselbe ist wie die wohl geläufigere Form

$$x = 2t/(1+t^2), \quad y = (1-t^2)/(1+t^2). \quad (6.3)$$

Analog lassen sich alle Ellipsen, Hyperbeln und Parabeln (deren Achsen *nicht parallel zur y-Achse* sind) als rationale Kurven darstellen.

Bemerkungen:

1. Bei allen rationalen Kurven hat man darauf zu achten, dass die Nullstellen des Nenners *nicht* in das Intervall $[a, b]$ fallen (oder dass gar keine existieren).
2. Die Darstellung (6.1) gilt in jeder Dimension d gleichermaßen; die Vektoren \mathbf{x} , und \mathbf{a}_k haben dann eben d Komponenten.
3. Im allgemeinen wird man verlangen, dass d der Vektoren $\mathbf{a}_1, \dots, \mathbf{a}_n$ *linear unabhängig* sind, damit die Kurve den \mathbb{R}^d aufspannt; andernfalls liegt die Kurve ganz in einem Unterraum des \mathbb{R}^d (z. B. bei $d = 3$ und genau zwei linear unabhängigen Vektoren: in einer *Ebene* des \mathbb{R}^3).
4. Die Darstellung (6.1) lässt sich – wie es schon beim Kreis (6.3) getan wurde – auch allgemein in ihre d Komponenten aufspalten.
5. Die rationalen Kurven vom Grad $n = 2$ (also „die quadratischen rationalen Kurven“) heißen *Kegelschnitte*; jene vom Grad $n = 3$ (die *kubischen rationalen Kurven*) werden kurz *Kubiken* genannt. Für die höheren Grade sind auch Bezeichnungen wie *Quartiken*, *Quintiken* etc. gebräuchlich.

6. Schließlich sei darauf hingewiesen, dass bei rationalen Kurven – wie auch bei polynomialen Kurven – die Parametertransformationen, welche die Rationalität und den Grad der Darstellung erhalten, nur die „gebrochen-linearen“ Transformationen sind:

$$t = \frac{a\bar{t} + b}{c\bar{t} + d}.$$

Da diese Transformationen das Doppelverhältnis invariant lassen, macht es Sinn, auf einer rationalen Kurve vom Doppelverhältnis eines Punkte-Quadrupels zu sprechen.

7. Wie man schon am Beispiel des Kreises leicht feststellen kann, ist die Verteilung des Parameters t auf die Kurve i. Allg. *sehr ungleichmäßig*. Beim Zeichnen von rationalen Kurven hat man dies zu berücksichtigen (insbesondere wenn man geradlinige Verbindungen zwischen den berechneten Punkten benutzt!).

6.2 Homogene Koordinaten und Fernpunkte

Die euklidischen (oder affinen) Räume \mathbb{R}^d , in denen man normalerweise Geometrie treibt, sind – topologisch gesehen – *keine abgeschlossenen Mengen*, daher gibt es in ihnen unendliche Folgen aus lauter verschiedenen Punkten, die keinen Häufungspunkt besitzen (z. B. die ganzen Zahlen auf der reellen Zahlgeraden \mathbb{R}^1). Um diesem Mangel abzuhelpfen, hat man diese Räume \mathbb{R}^d durch *Hinzufügung neuer Punkte* (die zuvor nicht da waren und die man sich als „unendliche ferne“ Punkte vorstellen kann) zu kompakten topologischen Räumen erweitert (Kompaktifizierung des \mathbb{R}^d). Rein topologisch gesehen, gibt es dafür viele Möglichkeiten; um jedoch möglichst viele der gewohnten geometrischen Eigenschaften (insbesondere die der linearen Unterräume) fortbe-

stehen zu lassen, fgt man dem \mathbb{R}^d eine einzige Hyperebene, *Fernhyperebene* genannt, hinzu.

Analytisch geschieht dies dadurch, dass man die affinen Koordinaten x_1, x_2, \dots, x_d als *Quotienten*

$$x_1 = \frac{\xi_1}{\xi_0}, \dots, x_d = \frac{\xi_d}{\xi_0} \quad (6.4)$$

mit einem beliebigen Nenner $\xi_0 \neq 0$ schreibt. (Man definiert also bei gegebenem ξ_0 die Gren ξ_k durch $\xi_k = x_k \cdot \xi_0$.) Diese (nur bis auf einen gemeinsamen Faktor bestimmten!) Gren ξ_k , $k = 1, \dots, d$ heien *homogene Koordinaten* des Punkts $(x_1, x_2, \dots, x_d)^T$.

Nun lassen sich alle Gleichungen und Formelausdrcke, in denen Koordinaten von Punkten involviert sind, durch Heraufmultiplizieren eventuell vorkommender Nenner „homogen machen“.

Man kann dieses Homogen-machen auch dadurch bewerkstelligen, dass man einem Vektor $\mathbf{x} = (x_1, x_2, \dots, x_d)$ eine nullte Komponente mit dem Wert 1 voranstellt, und diesen mit einem beliebigen Faktor $\rho \neq 0$ (z. B. mit einem vorkommenden Nenner) multipliziert ¹

$$\mathbf{X} = \rho \cdot (1, x_1, x_2, \dots, x_d). \quad (6.5)$$

¹ Man schreibt statt (6.5) gelegentlich auch einfach $\mathbf{X} = \rho \cdot (1, \mathbf{x})$.

So lassen sich z. B. die wohlbekannten linearen Gleichungen von Hyperebenen

$$a_0 + \sum_{k=1}^d a_k x_k = 0$$

(wobei mindestens einer der Koeffizienten a_k , $k = 1, \dots, d \neq 0$ sein muss) in der Form

$$\sum_{k=0}^d a_k \xi_k = 0 \quad (6.6)$$

schreiben. (Beachte, dass die Summe nun beim Index $k = 0$ beginnt, und *alle Glieder die homogenen Koordinaten in der ersten Potenz enthalten!*)

Bis hierhin sind noch keine neuen Punkte eingeführt worden; es ist alles nur in anderer Form geschrieben worden. Nun lässt man aber zu, dass *alle Koeffizienten* a_k , $k = 1, \dots, d$ verschwinden und nur $a_0 \neq 0$ ist. Dann geht (6.6) in die Gleichung

$$\xi_0 = 0 \quad (6.7)$$

über (wenn diese noch durch a_0 dividiert wird). Wenn man also *neue Punkte* durch (homogene) Koordinaten- $(d+1)$ -Tupel

$$(0, \xi_1, \dots, \xi_d)$$

eingführt, so erfüllen diese automatisch eine *lineare Gleichung* (6.7) und können daher als *in einer Hyperebene liegend* aufgefasst werden.

In analoger Weise kann man die analytischen Darstellungen aller anderen geometrischen Konfigurationen in homogenen Koordinaten umschreiben und mit der höchsten vorkommenden Potenz von ξ_0 heraufmultiplizieren. (Die entstehenden Ausdrücke und Gleichungen müssen dann alle *homogen* bezüglich der Variablen ξ_k sein; d. h. die Multiplikation aller dieser Variablen mit demselben

Faktor muss diese Ausdrücke bzw. Gleichungen invariant lassen.) Wenn man dann $\xi_0 = 0$ setzt, erhält man die Fernpunkte der betreffenden Figur.

Dies soll an einigen Beispielen erläutert werden:

1. Zwei parallele Geraden in der Ebene mit den Gleichungen $ax + by = c$ und $ax + by = d$ haben in homogenen Koordinaten die Gleichungen $a\xi_1 + b\xi_2 = c\xi_0$ und $a\xi_1 + b\xi_2 = d\xi_0$. Setzt man $\xi_0 = 0$ so werden beide Gleichungen identisch, und die Lösung ist

$$\xi_1 = \rho \cdot b, \quad \xi_2 = \rho \cdot (-a).$$

Da der Vektor $(a, b)^T$ ein Normalenvektor ist, ist der Vektor $(\xi_1, \xi_2)^T = (b, -a)^T$ ein Vektor in Richtung der beiden Geraden (der ja ebenfalls nur bis auf einen Faktor $\neq 0$ bestimmt ist.) Es gilt also der

Satz: *Alle Geraden einer Schar paralleler Geraden haben denselben Fernpunkt als Schnittpunkt. Dieser Fernpunkt hat die homogenen Koordinaten $(0, \xi_1, \xi_2)^T$, wobei $(\xi_1, \xi_2)^T$ ein Richtungsvektor dieser Geraden ist.*

2. Dieser Satz gilt genau so in allen Räumen \mathbb{R}^d , allein mit dem Unterschied, dass der Fernpunkt nun $d + 1$ Komponenten hat: $(0, \xi_1, \dots, \xi_d)^T$.
3. Hat die Ellipse $(x/a)^2 + (y/b)^2 - 1 = 0$ Fernpunkte (wo sie doch ganz in Endlichen liegt)? Der Übergang zu homogenen Koordinaten liefert die Gleichung $(\xi_1/a)^2 + (\xi_2/b)^2 - \xi_0^2 = 0$, und Einsetzen von $\xi_0 = 0$ (Schnitt mit der Ferngeraden) führt auf $(\xi_1/a)^2 + (\xi_2/b)^2 = 0$. Also: *Die Ellipse schneidet die Ferngerade in einem Paar konjugiert komplexer Punkte, das durch $\rho \cdot (0, a, \pm ib)^T$ gegeben ist (mit $\rho \in \mathbb{R}$, $\rho \neq 0$ beliebig).*

4. Nimmt man statt dessen die Hyperbel, so ändert sich nur das Vorzeichen beim Glied $(y/b)^2$; somit lautet die Gleichung für die Schnittpunkte mit der Ferngeraden nun $(\xi_1/a)^2 - (\xi_2/b)^2 = 0$. Diese hat die beiden *reellen* Lösungen: $\rho \cdot (0, a, \pm b)^T$, und das sind die Fernpunkte *der Asymptoten*. Es gilt also: *Die Asymptoten berühren die Hyperbel in ihren Fernpunkten.*
5. Nimmt man schließlich *die Parabel* (wie auch die andern beiden Kegelschnitte in der Normalform ihrer Gleichung) $y^2 - 2px = 0$, so findet man nach dem oben beschriebenen „Rezept“ die homogene Form: $\xi_2^2 - \xi_0 \xi_1 = 0$, und Einsetzen von $\xi_0 = 0$ liefert $\xi_2^2 = 0$. Der Fernpunkt ist also $(0, 1, 0)^T$ (oder – wie oben – ein Vielfaches davon). Da diese Lösung doppelt zählt, folgt: *Die Parabel berührt die Ferngerade, und der Berührungspunkt ist der Fernpunkt der Parabelachse.*

6.3 Anwendung auf rationale Kurven, insbesondere Kegelschnitte

Die Darstellung in homogenen Koordinaten lässt sich insbesondere auf rationale Kurven anwenden: Man erweitert alle Vektoren bei einer Darstellung einer rationalen Kurve (s. (6.1)) in *inhomogenen* Koordinaten zu Vektoren im \mathbb{R}^{d+1} (homogene Vektoren), indem man ihnen eine nullte Komponente mit Wert 1 voranstellt (vgl. (6.5)). Dann multipliziert man die Gleichung mit dem gemeinsamen Nenner

$$N(t) = \sum_{k=0}^n \gamma_k t^k$$

herauf und erhält

$$(N(t), N(t)\mathbf{x}(t)) = (N(t), \sum_{k=0}^n \mathbf{a}_k t^k)$$

also

$$\mathbf{X}(t) = \sum_{k=0}^n \mathbf{A}_k t^k \quad (6.8)$$

mit

$$\mathbf{X}(t) = (N(t), N(t)\mathbf{x}(t)) \quad \text{und} \quad \mathbf{A}_k = (\gamma_k, \mathbf{a}_k).$$

Daher gilt:

Bemerkung 21. *Jede rationale Kurve lässt sich unter Verwendung homogener Koordinaten in polynomialer Form darstellen.*

Als Beispiel seien die *Kegelschnitte* betrachtet: Wählt man in der (affinen oder auch projektiven) Ebene drei beliebige, nicht kollineare Grundpunkte $\mathbf{P}, \mathbf{Q}, \mathbf{R}$ und drei linear unabhängige quadratische Formen $F_0(t), F_1(t), F_2(t)$,¹ so beschreibt

$$\mathbf{X}(t) = F_0(t)\mathbf{P} + F_1(t)\mathbf{Q} + F_2(t)\mathbf{R} \quad (6.9)$$

stets einen Kegelschnitt, wobei $\mathbf{X}(t)$ auf der Verbindungsgeraden zweier Grundpunkte liegt, wenn t Nullstelle der dem dritten Grundpunkt zugeordneten Form ist. Wählt man zwei neue Grundpunkte, etwa \mathbf{A} und \mathbf{C} , auf dem Kegelschnitt liegend und einen dritten, \mathbf{B} , im Schnittpunkt der Tangenten, die in den ersten beiden Punkten an den Kegelschnitt gelegt wurden, so erhält man – wenn man noch den Parameter t so einrichtet, dass den Grundpunkten \mathbf{A} und \mathbf{C} die Parameterwerte 0 und ∞ zugeordnet werden, die vereinfachte Darstellung

$$\mathbf{X}(t) = \mathbf{A} + 2t \cdot \mathbf{B} + t^2 \cdot \mathbf{C}. \quad (6.10)$$

¹ Um sprachliche Überladungen zu vermeiden, wird nicht zwischen „Punkten“ und den sie darstellenden „homogenen Vektoren“ unterschieden.

Diese nennt man die (homogene oder projektive) *Normalform* der Kegelschnittsgleichung. (Sie gilt auch im Falle, dass einer oder gar zwei der Grundpunkte Fernpunkte sind.)

Sind hingegen alle drei Grundpunkte im Endlichen gelegen, so kann man zunächst die Normierung von \mathbf{A} (d. h. dessen 0. Komponente) zu eins machen: $\mathbf{A} = (1, \mathbf{a})$. Mit einer Parametertransformation $t = \rho \bar{t}$ ergibt sich $\bar{\mathbf{C}} = \rho^2 \mathbf{C}$, und man kann erreichen, dass $\mathbf{C} = \sigma(1, \mathbf{c})$ mit $\sigma = \pm 1$ wird.¹ Setzt man noch $\mathbf{B} = \omega(1, \mathbf{b})$, so folgt

$$\begin{aligned} \mathbf{X}(t) &= (1, \mathbf{a}) + 2\bar{t} \cdot (\omega, \omega\mathbf{b}) + \bar{t}^2 \cdot \sigma(1, \mathbf{c}) \\ &= (1 + 2\omega\bar{t} + \sigma\bar{t}^2, \mathbf{a} + 2\bar{t} \cdot \omega\mathbf{b} + \bar{t}^2 \cdot \sigma\mathbf{c}). \end{aligned} \quad (6.11)$$

Dies liefert nun die *rationale* Normalform

$$\mathbf{x}(t) = \frac{\mathbf{a} + 2\bar{t} \cdot \omega\mathbf{b} + \bar{t}^2 \cdot \sigma\mathbf{c}}{1 + 2\bar{t}\omega + \sigma\bar{t}^2} \quad (6.12)$$

der Kegelschnittsgleichung. Variiert man den Parameter ω , stellt sie eine *einparametrische Schar* von Kegelschnitten dar; man nennt die Schar ein *Kegelschnittbüschel*. Dies wird auch dadurch verständlich, dass ein Kegelschnitt i. Allg. durch fünf Punkte bestimmt ist; da die beiden Berührungspunkte \mathbf{A}, \mathbf{B} je doppelt zählen, sind vier Punkte „verbraucht“, und ein fünfter kann beliebig gewählt werden.

¹ Bei $\sigma = -1$ berühren die Kegelschnitte die beiden Tangenten von außen.

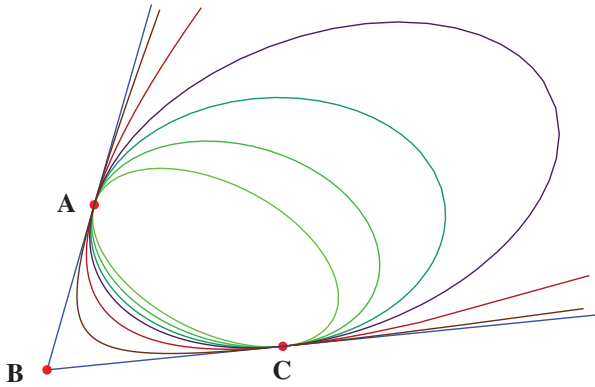


Abbildung 6.1: Ein Kegelschnittbüschel: Die Grundpunkte sind rot markiert; die gemeinsamen Tangenten sind blau. Ellipsen sind in grün, blaugrün, und violett gezeichnet, die Parabel in rot und ein Hyperbelast in braun.

6.4 G^k -Stetigkeit homogen dargestellter Kurven

Wenn zwei Kurvenssegmente in homogenen Koordinaten vorliegen, stellt sich die Frage, unter welchen Bedingungen sie G^k -stetig aneinander anschließen. Hier können nicht einfach die im Abschnitt 1.6 hergeleiteten Bedingungen (1.12) angewendet werden, da neben einer Parameter-Transformation auch noch eine mögliche *Umnormung* berücksichtigt werden muss. Die genauere Definition lautet:

Definition 22. Sind

$$t \mapsto \mathbf{X}(t), \quad t \in I,$$

und

$$s \mapsto \mathbf{Y}(s), \quad s \in J,$$

zwei (hinreichend oft stetig differenzierbare) Kurvendarstellungen in homogenen Koordinaten, so berühren sich diese an der Stelle t_0 in k -ter Ordnung, wenn es eine Parametertransformation

$$f : I \rightarrow J, \quad s = f(t)$$

und eine Umnormungsfunktion

$$\rho : I \rightarrow \mathbb{R}$$

(mit $\rho(t) \neq 0$ für alle $t \in I$) gibt, so dass die Funktionen $t \mapsto \mathbf{X}(t)$ und $t \mapsto \rho(t) \cdot \mathbf{Y}(f(t))$ selbst und alle ihre Ableitungen bis zur Ordnung k in einer gewissen Umgebung U von t_0 übereinstimmen.

Analog zu den in Abschnitt 1.6 hergeleiteten Bedingungen erhält man bei homogenen die folgenden Beziehungen, wobei alle Argumente weggelassen wurden, t durch t_0 und s durch $s_0 = f(t_0)$ zu ersetzen sind und $\phi = f'$ bedeutet:

$$\begin{aligned} \mathbf{X} &= \rho \mathbf{Y} \\ \mathbf{X}' &= \rho' \mathbf{Y} + \rho \phi \dot{\mathbf{Y}} \\ \mathbf{X}'' &= \rho'' \mathbf{Y} + (2\rho' \phi + \rho \phi') \dot{\mathbf{Y}} \\ \dots &\dots\dots\dots \end{aligned} \tag{6.13}$$

Auch hier gilt die allgemeine Formel

$$\mathbf{X}^{(k)} = \sum_{j=0}^k \Phi_j^k \mathbf{Y}^{(j)}, \tag{6.14}$$

wobei sich die Koeffizienten Φ_j^k *rekursiv* gemäß

$$\begin{aligned}\Phi_0^0 &= \rho \\ \Phi_j^{k+1} &= \frac{d}{dt}\Phi_j^k + \phi\Phi_{j-1}^k\end{aligned}\quad (6.15)$$

berechnen.¹ (Dabei sind $\Phi_{k+1}^k = 0$ und $\Phi_{-1}^k = 0$ zu setzen.)

Dies möge am Beispiel von zwei aneinander anschließenden Kegelschnittbögen verdeutlicht werden, welche an der Anschlussstelle eine G^2 -Berührung haben sollen. Diese seien in Normalform (s. (6.10) in Abschnitt 5.3) mit den Grundpunkten $\mathbf{A} = (1, a_1, a_2)$, $\mathbf{B} = \omega(1, b_1, 0)$, $\mathbf{C} = (1, 0, 0)$ bzw. \mathbf{C} , $\mathbf{D} = \chi(1, d_1, 0)$, $\mathbf{E} = (1, e_1, e_2)$ gegeben (damit sind die Berührbedingungen G^0 und G^1 bereits erfüllt).

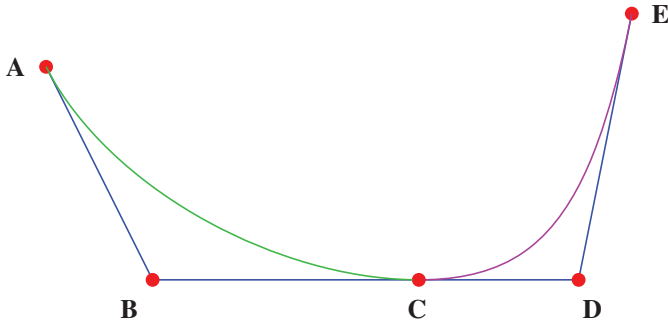


Abbildung 6.2: G^2 -Anschluss von 2 Kegelschnittbögen.

¹ Die Ableitungen sind „formal“ zu verstehen, d. h. die entsprechenden Polynome Φ_j von ρ, ϕ und deren Ableitungen sind zu differenzieren, noch *ehe* die festen Argumente t_0, s_0 eingesetzt werden.

Wegen

$$\mathbf{X}(t) = t^2 \mathbf{A} + 2t \mathbf{B} + \mathbf{C}, \quad \mathbf{Y}(s) = s^2 \mathbf{E} + 2s \mathbf{B} + \mathbf{C} \quad (6.16)$$

ist

$$\mathbf{X}(0) = \mathbf{Y}(0) = \mathbf{C}$$

und

$$\mathbf{X}'(0) = 2\mathbf{B}, \quad \dot{\mathbf{Y}}(0) = 2\mathbf{D},$$

sowie

$$\mathbf{X}''(0) = 2\mathbf{A}, \quad \ddot{\mathbf{Y}} = 2\mathbf{E},$$

woraus nach (6.13) zunächst $\rho(0) = 1$ und

$$2\mathbf{B} = \rho'(0) \mathbf{C} + 2\rho(0)\phi(0) \mathbf{D}$$

folgt. Mit den oben angegebenen Koordinaten ergibt sich

$$\phi(0) = \frac{\omega b_1}{\chi d_1}$$

und

$$\rho'(0) = 2\omega - 2\chi \phi(0).$$

Dies beinhaltet noch keine Bedingungen für die Kegelschnittbögen, sondern liefert lediglich die Werte $\phi(0)$ und $\rho'(0)$. Erst die dritte (G^2 -)Bedingung

$$2\mathbf{A} = \rho''(0) \mathbf{C} + (2\rho'(0)\phi(0) + \rho(0)\phi'(0)) 2\mathbf{D} + \rho(0)\phi(0)^2 2\mathbf{E}$$

führt neben den Werten für $\phi'(0)$ und $\rho''(0)$ auf die einzige Beziehung

$$a_2 = \phi(0)^2 e_2 = \left(\frac{\omega b_1}{\chi d_1} \right)^2 e_2 \quad (6.17)$$

zwischen den Koordinaten der Grundpunkte. Somit gilt:

Satz 23. *Die Beziehung (6.17) ist (bei der angegebenen Lage des Koordinatensystems) notwendig und hinreichend für einen G^2 -Übergang der beiden Kegelschnittbögen (6.16).*

7 Bézier-Darstellung

In vielen CAD-Systemen und Zeichenprogrammen werden heute sogenannte *Bézier-Kurven* verwendet. Diese sind nichts anderes als *polynomiale Kurven* (s. Kapitel 4), die nicht mit Potenzfunktionen, sondern mit Hilfe der Bernstein-Polynome dargestellt sind.

7.1 Parameterdarstellung und Kontrollpunkte

Eine sogenannte „Bézier-Kurve“ vom Grad n ist eine polynomiale Kurve in der Darstellung:

$$\mathbf{x}(t) = \sum_{k=0}^n \binom{n}{k} (1-t)^{n-k} t^k \mathbf{c}_k. \quad (7.1)$$

Dabei:

- bedeuten fett gedruckte Buchstaben Punkte im \mathbb{R}^d ($d = 2$: Ebene, $d = 3$: Raum, etc.),
- ist t der Parameter,

- heißen $\mathbf{c}_0, \mathbf{c}_1, \dots, \mathbf{c}_n$ die *Bézier-* oder *Kontrollpunkte* (sie bestimmen die Lage und Gestalt der Kurve und sind die Eckpunkte des *Bézier-Polygons*),
- die Polynome $\binom{n}{k}(1-t)^{n-k}t^k$ *Bernstein-Polynome* (sie werden oft mit $B_k^n(t)$ bezeichnet) und
- die Kurve $\mathbf{x}(t)$ für $t \in [0, 1]$ zusammen mit ihrem Bézier-Polygon ein *Bézier-Segment*.

Die Kurve $\mathbf{x}(t)$ (mit $t \in [0, 1]$) beginnt bei \mathbf{c}_0 und endet bei \mathbf{c}_n ; sie berührt dort die Geraden $\mathbf{c}_0\mathbf{c}_1$ bzw. $\mathbf{c}_{n-1}\mathbf{c}_n$, sofern diese Punkte voneinander verschieden sind. Der Kurvenverlauf ähnelt dem des Kontroll- oder Bézier-Polygons (das durch Verbinden der Kontrollpunkte entsteht). Die Kurve liegt ganz in der konvexen Hülle der Kontrollpunkte. Dies folgt aus dem

Satz 24. Sind $\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_n$ beliebige Punkte im \mathbb{R}^n und $\alpha_0, \alpha_1, \dots, \alpha_n$ beliebige, nicht negative reelle Zahlen, deren Summe $= 1$ beträgt, so liegt der Punkt

$$\mathbf{x} = \sum_{k=0}^n \alpha_k \mathbf{a}_k \tag{7.2}$$

in der konvexen Hülle dieser Punkte.

Bei Anwendung auf Bézier-Polygone hat man die Bézier-Punkte anstelle der Punkte \mathbf{a}_k und $\alpha_k = \binom{n}{k}(1-t)^{n-k}t^k$ (also den Wert des betrachteten Bernstein-Polynoms an der Stelle t) zu nehmen; für $t \in [0, 1]$ sind diese ≥ 0 .

Zur Berechnung eines Kurvenpunkts aus den Bézier-Punkten verwendet man den **Algorithmus von de Casteljau**:¹

```
Global:   $n, \{c_k, k = 0, \dots, n\}$     // Grad und Kontrollpunkte
Input :   $t \in [0 \dots 1]$                 // Parameter
Output:   $x(t)$                           // Kurvenpunkt
Lokal :   $\{b_k, k = 0, \dots, n\}, k, i$ 

begin
  for  $k = 0$  to  $n$  do                      // Laden der Kontrollpunkte
     $b[k] := c[k]$ 
  end for
  for  $i = 0$  to  $n$  do
    for  $k = 0$  to  $n - i$  do
       $b[k] = (1 - t) * b[k] + t * b[k + 1]$ 
    end for
  end for
  return  $x = b[0]$ 
end
```

Bemerkung: Es ist zu beachten, dass das lokale Array b bei jeder Stufe i überschrieben wird. Da der überschriebene Vektor $b[k]$ beim nächsten Schritt $k + 1$ nicht mehr benötigt wird, entsteht dadurch kein Schaden.

Übungen:

1. Man sehe sich (mit irgendeinem Programm) den Verlauf der Bernstein-Polynome an.

¹ Es handelt sich bei dieser Formulierung um keine echte Programmiersprache.

2. Man wähle einige (wenige) Punkte z. B. in der Ebene, bilde damit *Konvexkombinationen* wie in (7.2) und sehe zu, wie sich der Punkt \mathbf{x} mit den Werten der Koeffizienten verändern. (Wenn $n = 2$ ist und $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2$ nicht auf einer Geraden liegen, heißen die Koeffizienten $\alpha_0, \alpha_1, \alpha_2$ *baryzentrische Koordinaten* von \mathbf{x} .)
3. Man mache sich klar, dass ein Bézier-Segment vom Grad $n = 1$ mit der Strecke von \mathbf{c}_0 nach \mathbf{c}_1 übereinstimmt; der Parameter t ist dabei das *Teilverhältnis* des Zwischenpunkts $\mathbf{x}(t)$.

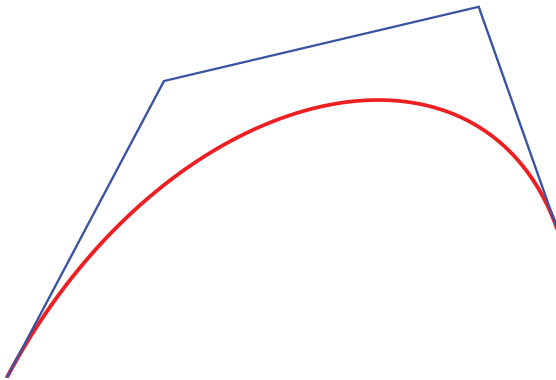


Abbildung 7.1: Eine typische Kurve vom Grad 3 mit ihrem Bézier-Polygon.

7.2 Zusammenhang mit polynomialen Kurven

Darüber gilt folgender

Satz 25. *Jeder endliche Abschnitt einer polynomialen Kurve (d. h. mit kompaktem Parameterintervall) hat eine Bézier-Darstellung. Umgekehrt lässt sich*

jede Bézier-Darstellung in eine mit Monomen (bis zum gleichen oder kleineren Grad) umwandeln.

Der **Beweis des ersten Teils** gliedert sich in zwei Schritte:

1. („Parameter-Transformation“) Der endliche Abschnitt bestehe aus jenem Teil der (vollständigen) Kurve, dessen Parameter- Intervall $[a, b]$ (mit $a < b$) ist. Dann definiert man einen *neuen Parameter* durch $\bar{t} = (t - a)/(b - a)$. Man setzt also $t = a + (b - a) \cdot \bar{t}$ in die Darstellung (5.1) ein. Durch Ausmultiplizieren aller Potenzen und Umordnen nach den Potenzen von \bar{t} entsteht wieder eine Darstellung derselben polynomialen Kurve, aber mit dem Parameter \bar{t} , welcher *nun von 0 bis 1 läuft*. Der Grad bleibt dabei erhalten.
2. (Umentwicklung nach Bernstein Polynomen) Man denkt sich die Bernstein-Polynome B_k^n „ausgeschrieben“; also z. B. für $n = 2$:

$$\begin{aligned} B_0^2 &= 1 - 2t + t^2, \\ B_1^2 &= -2t + 2t^2, \\ B_2^2 &= t^2, \end{aligned} \tag{7.3}$$

und fasst das als lineares Gleichungssystem für die „Unbekannten“ $1, t, t^2, \dots, t^n$ auf. Da dieses für jedes n Dreiecksgestalt hat und alle Koeffizienten in der Hauptdiagonalen seiner Matrix von Null verschieden sind, ist es stets eindeutig lösbar. So erhält man eine „Entwicklung der Potenzen $1, t, t^2, \dots, t^n$ nach den Bernstein-Polynomen“. Dies wird in die im ersten Schritt (Parameter-Transformation) erhaltene Darstellung (wobei der Querstrich bei t wieder weggelassen werden kann) eingesetzt. Die Umordnung nach den Bernstein-Polynomen ergibt schließlich die gewünschte Bézier-Darstellung.

Zum Beweis der Umkehrung hat man nur die Bernstein-Polynome nach Potenzen von t zu entwickeln und die Glieder mit gleichen Potenzen zusammenzufassen. \square

„Bézier-Kurven“ sind also nichts anderes als (endliche) Segmente polynomialer Kurven in einer *besonderen Darstellung*.

7.3 Ableitungen

Man rechnet leicht nach, dass sich die Ableitung einer Bézier-Kurve nach folgender Formel berechnet:

$$\mathbf{x}'(t) = n \cdot \sum_{k=0}^{n-1} \binom{n-1}{k} (1-t)^{n-1-k} t^k (\mathbf{c}_{k+1} - \mathbf{c}_k). \quad (7.4)$$

Die Ableitung hat also – abgesehen vom Faktor n – dieselbe Form wie eine Bézier-Kurve vom Grad $n-1$, jedoch stehen anstelle von Kontroll-Punkten hier deren Differenz-Vektoren.

Insbesondere gilt für $t = 0$

$$\mathbf{x}'(0) = n \cdot (\mathbf{c}_1 - \mathbf{c}_0) \quad (7.5)$$

und analog für $t = 1$

$$\mathbf{x}'(1) = n \cdot (\mathbf{c}_n - \mathbf{c}_{n-1}), \quad (7.6)$$

d. h. es gilt der

Satz 26. *Eine polynomiale Kurve berührt ihr Bézier-Polygon in dessen beiden Endpunkten.*

Analog zu (7.4) gelten auch für die höheren $(k+1)$ -ten Ableitungen entsprechende Formeln, wobei anstelle der Kontrollpunkte deren höhere Differenzen stehen und statt des Vorfaktors n das Produkt $n(n-1)\dots(n-k)$ zu setzen ist.

7.4 Glatter Anschluss polynomialer Segmente

Wie schon in Abschnitt 1.5 angesprochen, werden größere Kurvenbögen in (meist viele) kleinere Segmente unterteilt und als polynomiale Kurven genommen. Damit die gesamte Kurve *glatt* ist, müssen die einzelnen Segmente mindestens G^1 -stetig (s. Abschnitt 1.6) aneinanderschließen. Nach Satz 26 ist dies gewährleistet, wenn

$$\mathbf{c}_n = \mathbf{d}_0 \text{ und } \mathbf{c}_{n-1}, \mathbf{c}_n, \mathbf{d}_1 \text{ in dieser Reihenfolge kollinear sind.} \quad (7.7)$$

Hierbei wurden die Kontrollpunkte des folgenden Segments mit \mathbf{d}_k bezeichnet.¹

Bei höheren ästhetischen Ansprüchen reicht diese einfache G^1 -Stetigkeit nicht aus: Man verlangt meist mindestens G^2 -Stetigkeit. Daher erhebt sich die Frage, wie man die Kontrollpunkte der beiden Segmente einrichten muss, dass diese vorliegt.

Aus der Ableitungsformel (7.4) und der analogen Formel für die zweite Ableitung erkennt man, dass es nur auf die Kontrollpunkte $\mathbf{c}_{n-2}, \mathbf{c}_{n-1}, \mathbf{c}_n = \bar{\mathbf{c}}_0, \bar{\mathbf{c}}_1, \bar{\mathbf{c}}_2$ ankommt und (7.7) erfüllt sein muss. Bezeichnet man die Differenz-Vektoren der Kontrollpunkte mit $\mathbf{v}_i = \mathbf{c}_{i+1} - \mathbf{c}_i$ ($i = 0, \dots, n-1$) bzw. diejenigen des

¹ Bei m Segmenten zählt man die Kontrollpunkte einfach von 0 bis $m \cdot n$ durch, wobei diejenigen, deren Indizes ein Vielfaches von n sind, die gemeinsamen Anschlusspunkte aufeinanderfolgender Segmente sind.

zweiten Segments mit $\mathbf{w}_i = \mathbf{d}_{i+1} - \mathbf{d}_i$ ($i = 0, \dots, n-1$), so folgen aus (1.13) die Bedingungen

$$\mathbf{c}_n = \mathbf{d}_0 \quad (7.8)$$

$$\mathbf{v}_{n-1} \cdot f_1 = \mathbf{w}_0 \quad (7.9)$$

$$(\mathbf{v}_{n-1} - \mathbf{v}_{n-2}) \cdot f_1^2 + (f_2/(n-1))\mathbf{v}_{n-1} = \mathbf{w}_1 - \mathbf{w}_0, \quad (7.10)$$

wobei stillschweigend angenommen wurde, dass beide Segmente denselben Grad n haben. Es gilt also der

Satz 27. *Zwei Bézier-Segmente vom gleichen Grad haben genau dann einen G^2 -Anschluss, wenn die Gleichungen (7.8), (7.9) und (7.10) erfüllt sind.*

Die Gleichung (7.10) kann man auch kurz in der Form

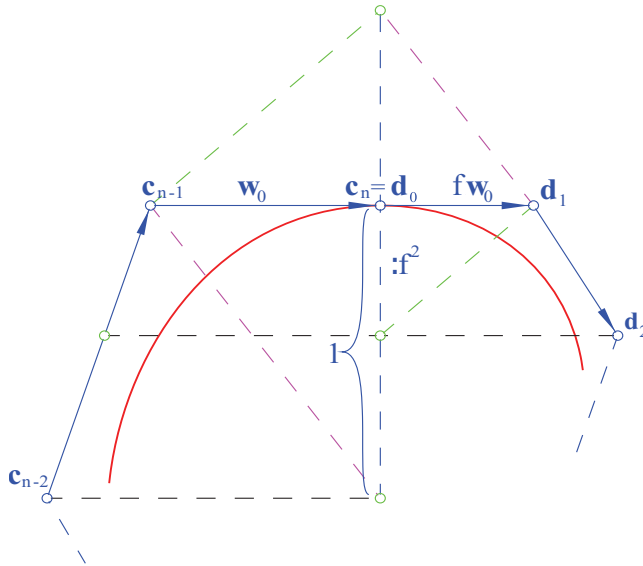
$$-\mathbf{v}_{n-2} \cdot f_1^2 \equiv \mathbf{w}_1 \pmod{\mathbf{w}_0} \quad (7.11)$$

schreiben.¹ Sie lässt sich z. B. durch die folgende Abbildung 7.2 veranschaulichen (statt f_1 steht nur f ; durch Projektion von \mathbf{c}_{n-2} , \mathbf{c}_n und \mathbf{d}_2 auf eine Hilfsgerade orthogonal zu $\mathbf{c}_{n-1}\mathbf{d}_1$ lässt sich ein Paar von Strecken im Verhältnis $1 : f^2$ erzeugen, so dass (7.11) gilt).

7.5 Konstruktion kubischer G^2 -Splines

Die in Kapitel 4 behandelten Übertragungen von Methoden für Splinefunktionen auf Splinekurven sind in gewisser Weise „ungeometrisch“, da sie nicht in-

¹ Das Zeichen \equiv bedeutet: „Die Gleichheit gilt bis auf Zusatzglieder, die im Modul stehen“ (hier also Vielfache von \mathbf{w}_0).

Abbildung 7.2: Zum G^2 -Anschluss von Bézier-Segmenten.

variant gegenüber Parametertransformationen sind und immer noch einer Knotenfolge bedürfen. Darüber hinaus ist die Abhängigkeit von dieser nicht linear und daher schwer zu überblicken. Es ist daher wünschenswert, einen G^2 -Spline zu konstruieren, welcher eine gegebene Punktfolge $\mathbf{p}_i (i = 0, 1, \dots, m)$ interpoliert (d. h. durch diese Punkte hindurch geht). Es liegt nahe, zwischen je zwei der gegebenen Punkte ein kubisches Segment einzuspannen. Dann braucht man nur noch dafür zu sorgen, dass für je zwei aufeinanderfolgende Segmente ein G^2 -Anschluss besteht. Üblicherweise nummeriert man die Bézier-Punkte aller Segmente von 0 bis $3m$ durch, wobei die Anfangs- und Endpunkte der Segmente die Nummern $3i$ haben; man hat also $\mathbf{b}_{3i} = \mathbf{p}_i (i = 0, 1, \dots, m)$.

Den Ergebnissen vom vorigen Abschnitt (insbesondere Satz (27)) entnimmt man, dass für je zwei aufeinanderfolgende Segmente *zwei Parameter* (sogenannte „Design Parameter“), frei gewählt werden können, zum einen das Verhältnis $f = |\mathbf{b}_{3i}\mathbf{b}_{3i+1}|/|\mathbf{b}_{3i-1}\mathbf{b}_{3i}|$ und zum anderen die Lage des Kontrollpunkts \mathbf{b}_{3i+2} auf der Parallelen zur gemeinsamen Tangente $\mathbf{b}_{3i-1}\mathbf{b}_{3i+1}$ im Abstand f^2 auf derselben Seite wie \mathbf{b}_{3i-2} (s. Abb. 7.2).

Etliche Wissenschaftler, die in der Pionierzeit des CAGD an dieser für die Praxis wichtigen Frage gearbeitet hatten, haben für die Wahl dieser Parameter geeignete Vorschläge erarbeitet (meist so, dass einer der Design-Parameter festgelegt wurde und der andere willkürlich blieb) und den so erhaltenen Splines mit einem speziellen Namen bezeichnet; auf diese Weise haben β -, γ -, ν -, τ -Splines etc. Eingang in die Literatur gefunden (siehe z. B. [16], [12]).

Ohne auf diese Splines näher einzugehen, werde hier eine andere Konstruktion eingeführt: Es sollen nämlich an den zu interpolierenden Punkten auch noch die *Tangentenrichtungen*, etwa repräsentiert durch die Einheitsvektoren \mathbf{v}_i ($i = 0, 1, \dots, m$) vorgegeben sein.

Die Aufgabe ähnelt der im Abschnitt 4.2 besprochenen Hermiteschen Interpolation, jedoch mit dem Unterschied, dass nur die Tangentenrichtungen (nicht aber die Ableitungsvektoren) vorgeschrieben sind (d. h. kein C^1 -Übergang!). Daher verläuft die Konstruktion des betrachteten Splines auch ganz anders als im genannten Abschnitt.

Der erste Design-Parameter f bleibt auch bei dieser Konstruktion frei wählbar und legt den Kontrollpunkt \mathbf{b}_{3i+1} fest. Damit ist auch die oben erwähnte Parallele zur gemeinsamen Tangente $\mathbf{b}_{3i}\mathbf{b}_{3i+1}$ im Abstand f^2 wohldefiniert, und auf dieser muss der nächste Kontrollpunkt \mathbf{b}_{3i+2} liegen.

Da nun aber die Tangente im Endpunkt $\mathbf{b}_{3i+3} = \mathbf{p}_{i+1}$ durch diesen Punkt und den vorgegebenen Richtungsvektor \mathbf{v}_{i+1} bestimmt ist, muss \mathbf{b}_{3i+2} auch auf dieser Geraden, also im Schnittpunkt mit jener Parallelen liegen. Die einzige Bedingung, die diese Konstruktion ermöglicht, ist, dass sich diese beiden Geraden treffen und der Schnittpunkt auf dem *rückwärtigen* Halbstrahl $\mathbf{b}_{3i+2} - \xi \cdot \mathbf{v}_{i+1}$ ($\xi > 0$) liegt.¹

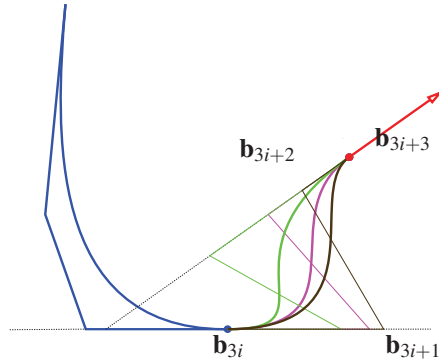


Abbildung 7.3: Konstruktion eines G^2 -Bézier-Anschluss-Segments mit gegebenen Punkten und Tangentenrichtungen.

Diese Konstruktion wird sukzessive für jeden inneren Anschlusspunkt \mathbf{b}_{3i} durchgeführt.

Die obige Abbildung 7.3 zeigt eine ziemlich ungünstige Konstellation eines einzelnen Anschlusses, welche einen Wendepunkt erzwingt, aber die G^2 -Bedingung dennoch zu erfüllen erlaubt. Es wurden drei mögliche Bézier-

¹ Der Fall, dass die Krümmung im Anschlusspunkt verschwindet, ist ebenfalls möglich. Dann liegen alle fünf Punkte $\mathbf{b}_{3i-2}, \dots, \mathbf{b}_{3i+2}$ in dieser Reihenfolge auf einer Geraden, und der erwähnte Halbstrahl muss diese in \mathbf{b}_{3i+2} treffen.

Segmente mit den λ -Werten 0.8, 1.0 und 1.1 in den Farben grün, violett und schwarz gezeichnet.

Die folgende Abbildung 7.4 zeigt einen kompletten Spline mit acht Segmenten (in verschiedenen Farben), die nach dieser Methode konstruiert wurden. Die Konstruktionsschritte sind in die Zeichnung mit aufgenommen.

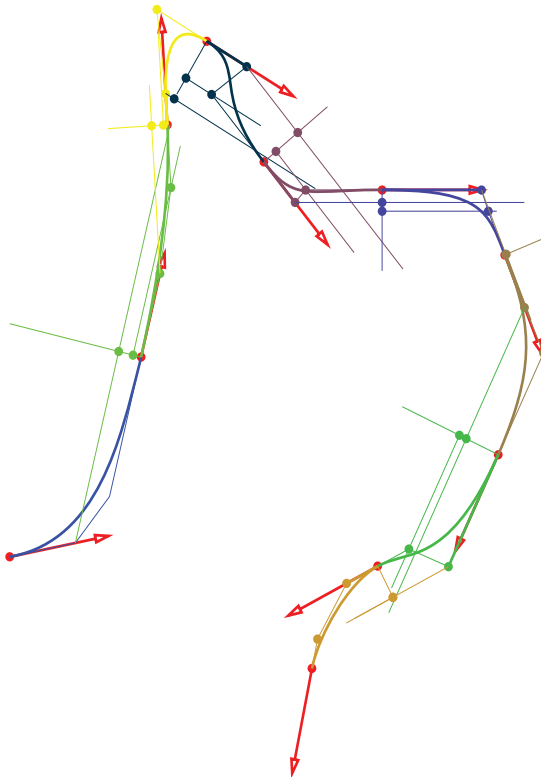


Abbildung 7.4: Konstruktion der Bézier-Punkte eines G^2 -Splines mit gegebenen Punkten und Tangentenrichtungen.

7.6 Konvexitätstreue

In vielen Fällen ist es wünschenswert, dass eine Kurve nicht mehr als unvermeidlich viele Wendepunkte besitzt; insbesondere dass ein langgestreckter Bogen, dessen Tangentenvektoren am Anfang und am Ende sich nur wenig voneinander unterscheiden, gar keine solchen besitzt, d. h. konvex ist. So entsteht die Frage, ob eine Kurve stets konvex ist, wenn ihr Bézier-Polygon diese Eigenschaft hat. Dies hängt eng mit der sogenannten *Variationsminderungseigenschaft* (engl. variation diminishing property) zusammen, welche Folgendes besagt:

Satz 28. *Eine Kurve hat mit einer beliebigen Geraden nicht mehr Schnittpunkte, als es Schnittpunkte dieser Geraden mit dem Bézier-Polygon gibt.*

(Einen Beweis findet man z. B. in [23] oder [18]).

Schließt man nun das Kontrollpolygon durch Verbinden des Endpunkts mit dem Anfangspunkt (sofern diese nicht schon von vornherein zusammenfallen), und entsteht auf diese Weise ein konvexes (und doppelpunktfreies) Polygon, so hat jede Gerade höchstens zwei Schnittpunkte mit diesem. Dasselbe gilt dann nach Satz 28 auch für die Bézier-Kurve, welche infolgedessen konvex ist.

Der Umweg über die „variation diminishing property“ erscheint – zumindest bei kleinem Grad – „von weit hergeholt“ und weckt das Bedürfnis nach einem einfacheren und direkten Beweis. Dieser soll hier für kubische Bézier-Kurven gegeben werden:

Satz 29. Eine kubische Kurve, deren Bézier-Polygon konvex ist und dessen orientierte Seitenvektoren in einen Winkelbereich von höchstens 180 Grad liegen, ist eine konvexe Kurve.¹

Beweis: Man bezeichne die Seitenvektoren des Kontrollpolygons der Reihe nach mit **a**, **b**, **c**, dann gilt nach der Ableitungsformel (7.4)

$$\mathbf{x}' = 3 (B_0^2(t)\mathbf{a} + B_1^2(t)\mathbf{b} + B_2^2(t)\mathbf{c}) \quad (7.12)$$

$$\mathbf{x}'' = 6 (B_0^1(t)(\mathbf{b} - \mathbf{a}) + B_1^1(t)(\mathbf{c} - \mathbf{b})). \quad (7.13)$$

Hieraus folgt für die Determinante dieser beiden Vektoren

$$[\mathbf{x}', \mathbf{x}''] = 18 (E \cdot [\mathbf{a}, \mathbf{b}] + F \cdot [\mathbf{b}, \mathbf{c}] + G \cdot [\mathbf{a}, \mathbf{c}]), \quad (7.14)$$

wobei

$$\begin{aligned} E &= B_0^2(t)B_0^1(t) + B_1^2(t)B_0^1(t) - B_0^2(t)B_1^1(t), \\ F &= B_1^2(t)B_1^1(t) + B_1^2(t)B_1^1(t) - B_2^2(t)B_0^1(t), \\ G &= B_2^2(t)B_0^1(t) + B_0^2(t)B_1^1(t). \end{aligned} \quad (7.15)$$

Die linke Seite von (7.14) hat nach (2.9) dasselbe Vorzeichen wie die *Krümmung* der Kurve an der Stelle t , da die Norm von $\mathbf{x}'(t)$ stets positiv ist. Es kommt also darauf an, zu zeigen, dass die rechte Seite keinen Vorzeichenwechsel aufweist. Nach Voraussetzung sind alle drei Determinanten in (7.14) positiv (man beachte die Bedingung über den Winkelbereich!). Folglich genügt es zu

¹ Der Zusatz über den Winkelbereich ist notwendig, weil sich sonst das Kontrollpolygon (und damit auch die Kurve) selbst überschneiden kann und somit nicht mehr konvex im strengen Sinn ist.

zeigen, dass die Größen E, F, G bis auf die Stellen $t = 0$ und $t = 1$, wo alle drei verschwinden, *positiv* sind.

Bei den Bernstein-Polynomen ist das generell der Fall; somit hat man $G \geq 0$. Um $E \geq 0$ zu zeigen, wird die Differenz

$$\begin{aligned} B_1^2(t)B_0^1(t) - B_0^2(t)B_1^1(t) &= \binom{2}{1}(1-t)t \cdot \binom{1}{0}(1-t) - \binom{2}{0}(1-t)^2 \cdot \binom{1}{1}t \\ &= (2-1)(1-t)^2t \geq 0 \end{aligned}$$

umgerechnet. Da der erste Summand von E ohnehin nicht negativ ist, folgt die behauptete Ungleichung.

Analog geht man bei F vor:

$$\begin{aligned} B_1^2(t)B_1^1(t) - B_2^2(t)B_0^1(t) &= \binom{2}{1}(1-t)t \cdot \binom{1}{1}t - \binom{2}{2}t^2 \cdot \binom{1}{0}(1-t) \\ &= (2-1)(1-t)t^2 \geq 0 \end{aligned}$$

In allen drei Fällen gilt $\dots = 0$ nur bei $t = 0$ und bei $t = 1$. □

7.7 Unterteilung des Bézier-Polygons

Wenn beim Design einer Kurve ein gewisses Segment zu groß geraten ist und etwa an der Stelle t_0 unterteilt werden muss, um den Anforderungen besser gerecht zu werden, entsteht die Frage, wo die Kontrollpunkte der beiden Teilbögen liegen. Diese wird durch den folgenden Satz beantwortet:

Satz 30. *Bezeichnet man die beim Algorithmus von de Casteljau entstehenden Zwischenpunkte der Stufen k von 0 bis n mit $\mathbf{c}_{k,i}$ (mit $i = 0, \dots, k$),¹ so sind die Punkte $\mathbf{c}_{0,0}, \mathbf{c}_{0,1}, \dots, \mathbf{c}_{0,n}$ die Kontrollpunkte des 1. Teilsegments und die Punkte $\mathbf{c}_{0,n}, \mathbf{c}_{1,n-1}, \dots, \mathbf{c}_{n,0}$ die Kontrollpunkte des 2. Teilsegments.*

Beweis: Die im Satz genannten Zwischenpunkte haben die explizite Darstellung

$$\mathbf{c}_{0,j} = \sum_{k=0}^j \binom{j}{k} (1-t_0)^{j-k} t_0^k \mathbf{c}_k. \quad (7.16)$$

Führt man nun für den 1. Teilbogen den Parameter $\tau := t/t_0$ ein (damit dieser für $t \in [0, t_0]$ das Intervall $[0, 1]$ durchläuft), so folgt $t = t_0 \tau$ und $1-t = 1-\tau + (1-t_0)\tau$. Setzt man dies in die Bézier-Darstellung der Kurve ein, so folgt (mit den Abkürzungen $\sigma = 1-\tau$ und $s_0 = 1-t_0$)

$$\begin{aligned} \mathbf{x}(t) &= \sum_{k=0}^n \binom{n}{k} (\sigma + s_0 \tau)^{n-k} (t_0 \tau)^k \mathbf{c}_k \\ &= \sum_{k=0}^n \sum_{i=0}^{n-k} \binom{n}{k} \binom{n-k}{i} \sigma^{n-k-i} \tau^{i+k} s_0^i t_0^k \mathbf{c}_k. \end{aligned} \quad (7.17)$$

Umordnen der Doppelsumme und Ersetzen der Indizes gemäß $j := i+k$, $i := j-k$ liefert:

$$\bar{\mathbf{x}}(\tau) = \mathbf{x}(t) = \sum_{j=0}^n \binom{n}{j} \sigma^{n-j} \tau^j \left(\sum_{k=0}^j \binom{j}{k} s_0^{j-k} t_0^k \mathbf{c}_k \right), \quad (7.18)$$

wobei die Beziehung $\binom{n}{k} \binom{n-k}{j-k} = \binom{n}{j} \binom{j}{k}$ verwendet wurde. Der in der großen Klammer stehende Ausdruck stimmt nach (7.16) mit dem Zwischenpunkt $\mathbf{c}_{0,j}$ überein, und die übrigen Terme ergeben die Bézier-Darstellung für das 1. Teilsegment. Womit der erste Teil des Satzes bewiesen ist. Der zwei-

¹ Bei $k=0$ sind es die ursprünglichen Kontrollpunkte \mathbf{c}_i (mit $i = 0, \dots, k$).

te ergibt sich analog, wenn man die Kurven in der Gegenrichtung durchlaufen lässt und die Bezeichnungen entsprechend ändert. \square

7.8 Interpolation mit polynomialen Kurven

Mit einer polynomialen Kurve vom Grad n kann man – wenn noch die Parameterwerte, an denen dies geschehen soll – $n + 1$ Punkte $\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_n$ interpolieren; i. Allg. werden die Parameterwerte äquidistant, d. h. $t_i = i/n$ ($i = 0, \dots, n$) genommen. Trivialerweise entsprechen \mathbf{a}_0 und \mathbf{a}_n den Bézier-Punkten $\mathbf{c}_0, \mathbf{c}_n$ und bei $n = 1$ sind dies schon alle.

Bei $n = 2$ kommt der „mittlere Punkt“ $\mathbf{a}_1 = \mathbf{x}(1/2) = 1/4\mathbf{c}_0 + 1/2\mathbf{c}_1 + 1/4\mathbf{c}_2$ hinzu. Eliminiert man hieraus den Kontrollpunkt \mathbf{c}_1 , so erhält man

$$(1-t)(1-2t)\mathbf{a}_0 + 4t(1-t)\mathbf{a}_1 + t(2t-1)\mathbf{a}_2 \quad (7.19)$$

als interpolierende Kurve. Bei genauerem Hinsehen erkennt man, dass dies nichts anderes ist als die wohlvertraute Interpolationsformel von Lagrange für die Stützstellen $0, 1/2, 1$, hier nur in Vektorform aufgeschrieben!

Das Analoge gilt natürlich auch für jeden Grad n ; man braucht die entsprechenden Umrechnungen von der Bézier-Form nicht einmal durchzuführen, denn das Ergebnis muss stets die vektorielle Form der Interpolationsformel von Lagrange sein (weil diese eindeutig ist):

Satz 31. *Die polynomiale Kurve*

$$\mathbf{x}(t) = \sum_{k=0}^n L_k(t) \mathbf{a}_k, \quad (7.20)$$

wobei $L_i(t)$ die Lagrangeschen Basis-Funktionen

$$L_k(t) = \prod_{\substack{i=0,1,\dots,n \\ i \neq k}} \frac{t-t_i}{t_k-t_i}$$

für die Stützstellen t_0, t_1, \dots, t_n bedeuten, interpoliert die (gegebenen) Punkte $\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_n$ an diesen Stellen.

Von „Bézier-Kurven“ ist zwar nicht mehr viel übrig geblieben, vielmehr liegt wieder ein Beispiel einer direkten Übertragung einer Methode der Numerik von Funktionen auf eine Frage bei Kurven vor, wie es schon in Abschnitt 4.2 aufgetreten war.

Wenn die Stützstellen, wie oben erwähnt, *äquidistant* sind, kann man selbstverständlich auch darauf zugeschnittene Formeln, wie z. B. diejenige von Newton-Gregory verwenden:

$$\mathbf{x}(t) = \sum_{k=0}^n \binom{s}{k} \Delta_0^k \mathbf{a}.$$

Hierbei ist $s = (t - t_0)/h$, $t_k = t_0 + kh$ und $\binom{s}{k}$ bedeuten die „verallgemeinerten Binomialkoeffizienten“

$$\binom{s}{k} = \frac{s(s-1)(s-2) \dots (s-k+1)}{1 \cdot 2 \cdot \dots k}.$$

Ferner bezeichnet das Symbol $\Delta_0^k \mathbf{a}$ bei $k = 0$ den Punkt \mathbf{a}_0 , bei $k = 1$ den ersten Differenzvektor $\mathbf{a}_1 - \mathbf{a}_0$ und bei $k > 1$ die „höheren Differenzvektoren“, die man durch sukzessive Differenzbildung aus den vorigen erhält.

Statt – wie oben geschehen – die Kontrollpunkte durch die zu interpolierenden Punkte zu ersetzen, kann man auch die Kontrollpunkte durch die Datenpunkte ausdrücken und damit die Bézier-Darstellung des entstehenden Splines erhalten.

Es seien also der Grad n und $n \cdot m + 1$ Datenpunkte

$$\mathbf{p}_0, \mathbf{p}_1, \dots, \mathbf{p}_{n \cdot m}$$

gegeben. Für je $(n + 1)$ aufeinanderfolgende Punkte $\mathbf{p}_{n \cdot k}, \dots, \mathbf{p}_{(n+1) \cdot k}$ (mit $k = 0, \dots, m - 1$) wird ein Bézier-Segment gebildet, welches diese Punkte interpoliert. Da der Endpunkt des vorigen Segments mit dem Anfangspunkt des nächsten übereinstimmt, entsteht zumindest ein G^0 -Spline.

Wie oben werden äquidistante Stützstellen $t_i = i/n$ genommen; es müssen dann die Interpolationsbedingungen

$$\mathbf{p}_{n \cdot k + j} = \sum_{i=0}^n B_i^n(t_j) \mathbf{c}_i \quad (j = 0, \dots, n) \quad (7.21)$$

erfüllt sein. Aus diesen Gleichungen lassen sich nun die Kontrollpunkte berechnen, denn die Koeffizienten-Matrix mit den Elementen $B_i^n(t_j)$ hat eine von Null verschiedene Determinante. Für $n = 3$ erhält man z. B.:

$$\begin{aligned} \mathbf{c}_0 &= \mathbf{p}_{3n} \\ \mathbf{c}_1 &= (-15 \mathbf{p}_{3n} + 54 \mathbf{p}_{3n+1} - 27 \mathbf{p}_{3n+2} + 6 \mathbf{p}_{3n+3}) / 18 \\ \mathbf{c}_2 &= (6 \mathbf{p}_{3n} - 27 \mathbf{p}_{3n+1} + 54 \mathbf{p}_{3n+2} - 15 \mathbf{p}_{3n+3}) / 18 \\ \mathbf{c}_3 &= \mathbf{p}_{3n+3} \end{aligned} \quad (7.22)$$

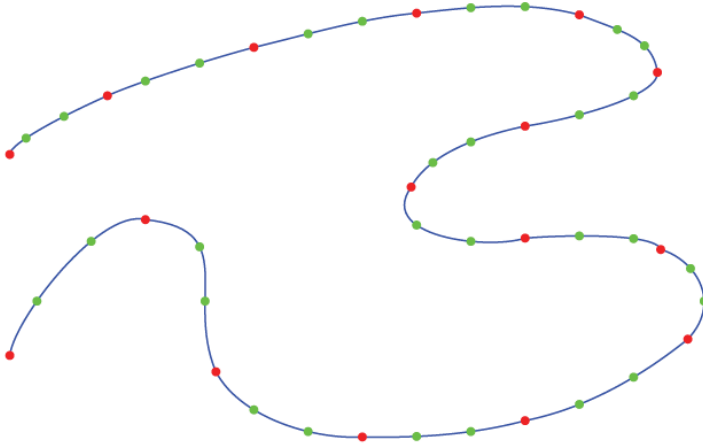


Abbildung 7.5: Ein interpolierender Spline.

Man mag befremdet sein, wenn hier ein Verfahren angegeben wird, welches nur einen stetigen, nicht einmal glatten Spline liefert. Dieses ist jedoch sehr schnell und interpolierend, folglich für einen Rohentwurf gut geeignet. Außerdem darf man nicht vergessen, dass beim praktischen Zeichnen bzw. Plotten meistens statt eines aufwändig berechneten G^k -Splines sowieso nur ein *Polygonzug* (oder gar nur ein „Treppchen“) aufs Papier kommt ...

Die Abbildung unten zeigt einen mit dieser Methode konstruierten Spline. Die Anfangs- und Endpunkte sind rot markiert, die anderen Datenpunkte grün.

8 B-Splines

Spline-Kurven sind als spezielle stückweise polynomiale (oder rationale) und G^k -stetige Kurven definiert. Im Gegensatz zum Vorgehen in 7.8 werden Spline-Kurven jedoch *nicht aus einzelnen Segmenten* zusammengesetzt, sondern die Zusammensetzung geschieht bereits auf der Ebene der Koordinaten-Funktionen, die man *B-Spline-Funktionen* (oder auch kurz B-Splines) nennt. Daher muss man sich zunächst mit B-Spline-Funktionen vertraut machen.

8.1 B-Splines

Man geht von einer monoton wachsenden Folge von *Knoten* $\{t_k \mid k \in \mathbb{Z}\}$ auf der reellen Achse aus. Vorerst denke man an eine *streng* monotone Folge, d. h. $t_k < t_{k+1}$ für alle $k \in \mathbb{Z}$; später werden auch *mehrfache Knoten* zugelassen. Diese Knoten sind die *Übergangsstellen*, an denen eine Spline-Funktion von einem Polynom zum einem andern C^k -stetig wechselt (und nur an den Knoten darf eine Spline-Funktion einen solchen Wechsel haben). Ferner wird verlangt, dass alle beteiligten Polynome vom Grad $\leq n$ sind; diesen nennt man den *Grad der Spline-Funktion*.

Die Grundidee ist nun, dass man für jeden Grad n (bei gegebener Knotenfolge) ein System von *Spline-Basisfunktionen*

$$N_k^n : \mathbb{R} \rightarrow \mathbb{R}, \quad t \mapsto N_k^n(t)$$

mit folgenden Eigenschaften konstruiert:

1. N_k^n ist eine stückweise polynomiale Funktion vom Grade n , die außerhalb des Intervalls $[t_k, t_{k+n+1}]$ identisch verschwindet und in dessen Innerem nur positive Werte hat.¹
2. N_k^n ist überall $(n-1)$ -mal stetig differenzierbar und hat an den Knoten $t_k, t_{k+1}, \dots, t_{k+n+1}$ Sprungstellen der n . Ableitung (N_k^0 ist dort unstetig).
3. Für alle $t \in \mathbb{R}$ und alle Grade $n \geq 0$ gilt:

$$\sum_{k \in \mathbb{Z}} N_k^n(t) = 1. \quad (8.1)$$

Bemerkungen:

1. Man erkennt sofort, dass die Basisfunktionen N_k^0 in $[t_k, t_{k+1}]$ den konstanten Wert 1 haben muss und sonst überall verschwindet (s. Abb. 8.1).
2. Ebenso leicht folgert man, dass die Basisfunktionen N_k^1 im ersten Teilintervall $[t_k, t_{k+1}]$ linear von Null bis Eins ansteigt und im zweiten Teilintervall $[t_{k+1}, t_{k+2}]$ wieder linear auf Null zurückgeht (s. Abb. 8.2).

¹ Aus b) folgt, dass der Funktionswert bei $n > 0$ in den Endpunkten ebenfalls gleich Null ist; bei $n = 0$ wird er für $t = t_k$ gleich Eins und für $t = t_{k+1}$ gleich Null gesetzt.

3. Bei den höheren Graden ist nicht unmittelbar zu erkennen, ob es überhaupt ein solches System von Basisfunktionen gibt und wie gegebenenfalls Teilpolynome berechnet werden können. (Beides wird in Folgenden positiv beantwortet werden.)
4. Die Existenz einmal vorweggenommen, folgt schon aus den Eigenschaften a) und b), dass die Graphen der Spline-Basisfunktion für $n \geq 2$ stets eine *Glockenform* haben (s. Abb. 8.3).
5. Funktionen, die nur in einem endlichen Intervall $[a, b]$ von Null verschiedene Werte haben, nennt man *Funktionen mit kompaktem Träger*; sie spielen auch in anderen Gebieten der Mathematik eine wichtige Rolle.
6. Die Gleichung (8.1) nennt man „Zerlegung der Eins“; sie erlaubt gewichtete Mittelbildungen, was später bei den Spline-Kurven verwendet werden wird.

Die folgenden Abbildungen zeigen die Basisfunktionen der Grade 0, 1, 2.

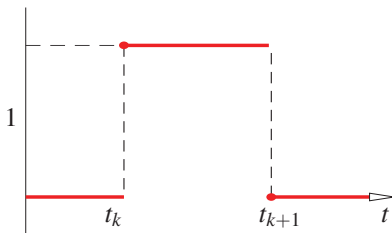


Abbildung 8.1: Die Basisfunktion N_k^0 .

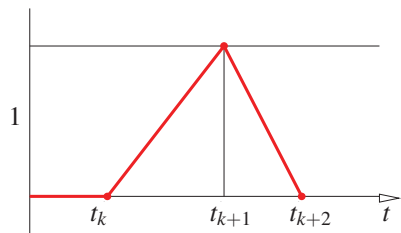
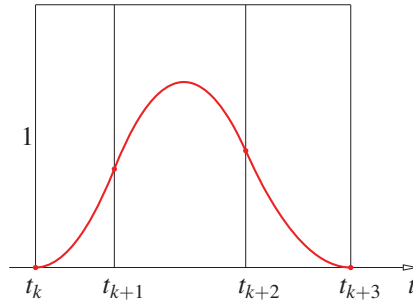


Abbildung 8.2: Die Basisfunktion N_k^1 .

Abbildung 8.3: Die Basisfunktion N_k^2 .

In der mathematischen Theorie der Splines (vgl. [3, 4]) wurde bewiesen:

Satz 32. (I.J. Schoenberg u. a., um 1946) *Es gibt für jedes $n \in \mathbb{N}$ genau ein System von Funktionen N_k^n , welches den Bedingungen a) bis c) genügt.*

Es wird hier auf einen ausführlichen Beweis verzichtet und auf die angegebene Literatur verwiesen. Es sei jedoch eine Idee genannt, die das Ergebnis plausibel erscheinen lässt: Das Polynom im ersten Teilintervall $[t_k, t_{k+1}]$, in welchem N_k^n nicht verschwindet, muss die Form $y = a(t - t_k)^n$ haben, da N_k^n einen C^{n-1} -stetigen Anschluss an das vorausgehende Nullpolynom haben soll. Im Endpunkt (Knoten t_{k+1}) hat man dann die Werte

$$\begin{aligned} y(t_{k+1}) &= a(\Delta t_k)^n, \\ y'(t_{k+1}) &= n a(\Delta t_k)^{n-1}, \\ &\vdots \\ y^{(n-1)}(t_{k+1}) &= (n-1) \dots 1 \cdot (\Delta t_k), \end{aligned}$$

(mit der Abkürzung $\Delta t_k = (t_{k+1} - t_k)$, welche wegen b) mit dem Funktionswert bzw. den Ableitungen des nächsten Teilpolynoms übereinstimmen müssen. Daher kann man dieses als Taylorpolynom mit diesen Anfangswerten ansetzen,

und es bleibt nur noch der Koeffizient bei $(t - t_{k+1})^n$ frei wählbar. So geht es weiter bis zum vorletzten Teilpolynom; jedes Mal bleibt der Koeffizient bei der höchsten Potenz unbestimmt. Beim letzten Teilpolynom jedoch müssen nicht nur die Anfangswerte mit den Endwerten des vorigen passend sein, sondern auch die Endwerte (bei t_{k+n+1}) einen C^{n-1} -stetigen Anschluss an das darauf folgende Nullpolynom haben. Dies liefert ein homogenes lineares Gleichungssystem mit n Gleichungen für die bislang $n + 1$ unbestimmten Koeffizienten, welches somit stets nichttrivial lösbar ist und genau einen Proportionalitätsfaktor offen lässt. Dieser wird dazu verwendet, die letzte Bedingung c) zu erfüllen.

Die explizite Berechnung der B-Spline-Basisfunktionen nach dieser Idee wäre allerdings recht umständlich, zumal es folgende rekursive Methode gibt:

Satz 33. *Die Spline-Basisfunktionen lassen sich rekursiv wie folgt berechnen:*

$$N_k^0(t) = \begin{cases} 1 & \text{für } t_k \leq t < t_{k+1} \\ 0 & \text{sonst} \end{cases}, \quad (8.2)$$

$$N_k^{n+1}(t) = \frac{t - t_k}{t_{k+n} - t_k} N_k^n(t) + \frac{t_{k+n+1} - t}{t_{k+n+1} - t_{k+1}} N_{k+1}^n(t). \quad (8.3)$$

8.2 Spline-Kurven

Eine *Spline-Kurve* im \mathbb{R}^d entsteht nun dadurch, dass man zunächst eine Folge von *Kontrollpunkten* $\{\mathbf{d}_k \in \mathbb{R}^d \mid k \in \mathbb{Z}\}$ wählt und mit diesen ein vom Kurvenparameter $t \in \mathbb{R}$ abhängiges *gewogenes Mittel* bildet:

$$\mathbf{x}(t) = \sum_{k=-\infty}^{+\infty} N_k^n(t) \mathbf{d}_k. \quad (8.4)$$

Da bei gegebenem Wert von t nur endlich viele Basisfunktionen von Null verschieden sind, ist diese Summe stets *endlich*, so dass keine Konvergenzprobleme auftreten. Genauer folgt aus der definierenden Eigenschaft a) der

Satz 34. Für $t \in [t_k, t_{k+1}]$ liegt der laufende Kurvenpunkt in der konvexen Hülle der Kontrollpunkte $\mathbf{d}_{k-n}, \mathbf{d}_{k-n+1}, \dots, \mathbf{d}_k$, da nur diese Punkte zur genannten Mittelbildung beitragen.

Bemerkungen:

1. Da die Spline-Basisfunktionen im Innern der Trägerintervalle positive Werte haben, geht die Spline-Kurve im Allgemeinen *nicht durch die Kontrollpunkte*.¹
2. Da nur die oben genannten Kontrollpunkte zur Summe in (8.4) beitragen, sofern t in $[t_k, t_{k+1}]$ liegt, beeinflussen nur diese die Lage von $\mathbf{x}(t)$ („lokale Abhängigkeit von den Kontrollpunkten“). Dies ist einer der wesentlichen Vorteile der Spline-Kurven, da man einen unerwünschten Verlauf an einer Stelle durch Verschieben einiger Kontrollpunkte korrigieren kann, ohne die ganze Kurve neu berechnen zu müssen.
3. Der zweite Hauptvorteil ist, dass automatisch ein G^{n-1} -Spline entsteht, sofern keine mehrfachen Knoten vorhanden sind (man braucht sich also nicht um die Glattheit an den Knoten zu kümmern).
4. Man muss eine Spline-Kurve nicht über die gesamte reelle Achse verfolgen, sondern kann sie auch nur in einem Teilintervall $[a, b]$ betrach-

¹ Dies kann z. B. am Anfangspunkt bzw. am Endpunkt vermieden werden, wenn später *mehrfache Knoten* eingeführt werden.

- ten. Allerdings ist dabei zu beachten, dass *sie nicht durch die Kontrollpunkte läuft* (s. Bem. 1).
5. Um diesem letztgenannten Nachteil zu begegnen, hat man die Theorie der B-Splines erweitert, indem man *mehrfache Knoten* eingeführt hat (siehe Abschnitt 7.4).
 6. Wenn eine Spline-Kurve nur in einem gewissen Indexbereich $k_{\min} \leq k \leq k_{\max}$ benötigt wird, kann man die restlichen Kontrollpunkte alle gleich 0 setzen. (Man braucht dann die Knoten t_k auch nur in diesem Bereich zu definieren.)
 7. Ferner gibt es für B-Splines viele hilfreiche Techniken, wie z. B. das Einfügen von Knoten oder Algorithmen zur Unterteilung etc. Wegen dieser und weiterer Details muss auf die Literatur verwiesen werden.
 8. Die Bezeichnungen der Spline-Basisfunktionen sind in der Literatur nicht ganz einheitlich. Insbesondere wird oftmals statt des Grades n die *Ordnung* $m = n + 1$ benutzt oder beide Indizes unten gesetzt.

8.3 Der de Boor-Algorithmus

Zur effektiven praktischen Berechnung einer Spline-Kurve verwendet man einen Algorithmus, der dem de Casteljau-Algorithmus entspricht und nach de Boor benannt wird. Er beruht auf der Rekursionsformel (8.2) der Spline-Basisfunktionen und der Darstellung (8.4) einer Spline-Kurve.

Ersetzt man nämlich die Spline-Basisfunktionen in der Darstellung (8.2), so entsteht

$$\mathbf{x}(t) = \sum_{k=-\infty}^{+\infty} (\alpha_k^0 N_k^{n-1} + (1 - \alpha_{k+1}^0) N_{k+1}^{n-1}) \mathbf{d}_k, \quad (8.5)$$

wobei

$$\alpha_k^j = \frac{t - t_k}{t_{k+n+1-j} - t_k} \quad (8.6)$$

gesetzt wird und diese Abkürzung zunächst nur für $j = 0$ verwendet wurde. Ordnet man diese Summe nach den Basisfunktionen N_k^{n-1} um (was darauf heraus läuft, beim 2. Glied den Index k durch $k - 1$ zu ersetzen), so folgt

$$\mathbf{x}(t) = \sum_{k=-\infty}^{+\infty} (\alpha_k^0 \mathbf{d}_k + (1 - \alpha_k^0) \mathbf{d}_{k-1}) N_k^{n-1}. \quad (8.7)$$

Daher ist es sinnvoll, auch die Kontrollpunkte *rekursiv* gemäß

$$\begin{aligned} \mathbf{d}_k^0 &= \mathbf{d}_k \\ \mathbf{d}_k^{j+1} &= \alpha_k^j \mathbf{d}_k^j + (1 - \alpha_k^j) \mathbf{d}_{k-1}^j \end{aligned} \quad (8.8)$$

zu berechnen. Damit wird (8.7) zu

$$\mathbf{x}(t) = \sum_{k=-\infty}^{+\infty} N_k^{n-1} \mathbf{d}_k^1. \quad (8.9)$$

Diese Darstellung hat dieselbe Form wie (8.4) nur mit dem Unterschied, dass die „nächste“ Serie von Kontrollpunkten \mathbf{d}_k^1 mit den Spline-Basisfunktion vom Grad $n - 1$ kombiniert wird. Offensichtlich kann man diesen Prozess aufgrund der Rekursionsformeln beider Faktoren solange wiederholen, bis der Grad auf 0 reduziert ist und für $t \in [t_k, t_{k+1}]$ nur noch der *eine* Summand $N_k^0 = 1$ übrig bleibt; d. h.

$$\mathbf{x}(t) = \mathbf{d}_k^n \quad \text{für } t \in [t_k, t_{k+1}]. \quad (8.10)$$

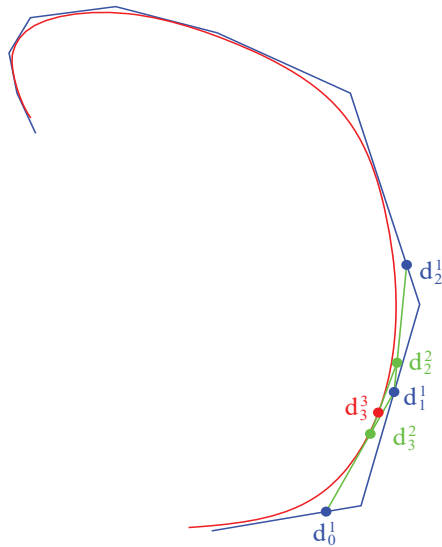


Abbildung 8.4: Eine Spline-Kurve mit Erläuterung des de Boor Algorithmus.

Daher sind nur die „höheren Kontrollpunkte“ bis zur Ordnung n nach (8.8) unter Verwendung von (8.6) zu berechnen. Programmiertechnisch kann dies analog wie beim de Casteljau Algorithmus geschehen.

Die Wahl der Knoten hat nur wenig Einfluss auf die Gestalt der B-Spline-Kurve (und dieser Einfluss ist noch wenig erforscht); klar ist jedoch, dass eine lineare Transformation $\bar{t} = at + b$ die Kurve gar nicht verändert.

8.4 B-Splines mit mehrfachen Knoten

Wie schon erwähnt, wurde die Theorie der B-Splines dahingehend erweitert, dass *mehrfache Knoten* zugelassen werden. Dabei heißt t_k ein m -facher Knoten, wenn

$$t_{k-1} < t_k = t_{k+1} = \dots = t_{k+m-1} < t_{k+m} \quad (8.11)$$

gilt. (Es ist nur $1 \leq m < n$ sinnvoll; Knoten mit $m = 1$ sind einfache Knoten.)

Bei der rekursiven Berechnung der Spline-Basisfunktionen werden Glieder, deren Nenner aufgrund mehrfacher Knoten verschwinden, einfach weggelassen.

Man kann Folgendes zeigen:

Satz 35. *Ist t_k ein m -facher Knoten, so ist eine Spline-Kurve an dieser Stelle nur noch C^{n-m-1} -stetig;*

und ferner

Satz 36. *Ist t_k ein $(n-1)$ -facher Knoten, so geht eine Spline-Kurve vom Grad n durch diesen Punkt (bzw. beginnt oder endet dort).*

Mit Hilfe dieser Ergebnisse lassen sich z. B. Kurven modellieren, die an bestimmten Stellen *Ecken* haben sollen (man wähle $m = n - 1$).

8.5 Rationale Spline-Kurven (NURBS)

Wie bei polynomialen Kurven kann man auch bei Spline-Kurven zu rationalen Kurven übergehen, wobei der Nenner nun nach den *Basis-Splines* entwickelt wird und die Kontrollpunkte mit den Entwicklungskoeffizienten als Gewichten versehen werden. Dies liefert die Darstellung

$$\mathbf{x}(t) = \frac{\sum_{k \in \mathbb{Z}} \gamma_k N_k^n(t) \mathbf{c}_k}{\sum_{k \in \mathbb{Z}} \gamma_k N_k^n(t)} \quad (8.12)$$

der *rationalen Spline-Kurven* (mit im Allgemeinen nicht gleichmäßig verteilten Knoten), die kurz „NURBS“ (= „non uniform rational B-splines“) genannt werden.¹ Auch hier können die Gewichte γ_k als zusätzliche Design-Parameter genutzt werden.

¹ Der Begriff „NURBS“ stellt eine Verballhornung dar; denn es gibt keine rationalen B-Splines und die rationalen Splines vom Grad n bilden nicht einmal einen linearen Raum.

9 Algebraische Kurven

Dieses Thema ist sehr umfangreich und füllt etliche Bände an mathematischer Literatur. Es können daher hier nur die allerersten Begriffe und Eigenschaften besprochen werden.

9.1 Grundbegriffe

Definition 37. Eine ebene algebraische Kurve vom Grad $n > 0$ ist eine solche, deren Punkte (mit Koordinaten x, y) eine polynomiale Gleichung der Art

$$\sum_{i+k \leq n} a_{i,k} x^i y^k = 0 \quad (9.1)$$

erfüllen. Damit der Grad wirklich n ist, muss mindestens einer der Koeffizienten $a_{n-k,k}$ ($k = 0, \dots, n$) von Null verschieden sein.

Wenn der Grad $= 1$ ist, ist (9.1) eine *lineare* Gleichung und stellt also eine *Gerade* dar.

Dagegen haben bekanntlich alle *Kegelschnitte* eine quadratische Gleichung; sie sind also nicht nur polynomiale, sondern auch algebraische Kurven vom Grad 2. Das Umgekehrte gilt aber nicht ohne Einschränkungen: Eine algebraische

Kurve 2. Grades kann *zerfallen*, d. h. das Polynom auf der linken Seite in (9.1) kann in ein Produkt aus zwei linearen Faktoren zerlegt werden: Dann wird die Gleichung bereits erfüllt, wenn die Koordinaten x, y nur *einen* dieser Faktoren annullieren; d. h. *die Kurve besteht aus zwei Geraden*. (Diese Figur kann nicht in einer einzigen polynomialen oder rationalen Parameterdarstellung erfasst werden!) Trotzdem zählt man die Geradenpaare und auch die sogenannte „Doppelgeraden“ (wo beide Faktoren bis auf eine Konstante gleich sind) zu den Kegelschnitten und nennt sie „entartet“.

Bei den höheren Graden gibt es selbstverständlich viele Möglichkeiten des Zerfallens. Die (Teil-)Kurven, die durch Nullsetzen eines (nicht konstanten!) Faktors entstehen, heißen „Bestandteile“ der betreffenden algebraischen Kurve. Es gilt aber stets der

Satz 38. *Die Summe der Grade aller Bestandteile einer ebenen algebraischen Kurve vom Grad n ist gleich n .*

Für $n = 3$ (Kubiken) gibt es also nur die beiden folgenden Fälle des Zerfallens:

1. Zerfall in einen nicht entarteten Kegelschnitt und eine Gerade,
2. Zerfall in drei (nicht notwendig verschiedene) Geraden.

Definition 39. *Wenn das die Kurve darstellende Polynom in (9.1) sich nicht in (mindestens lineare) Faktoren zerlegen lässt, heißt die Kurve (und auch das Polynom) irreduzibel.*

Für eine algebraische Kurve bedeuten also „irreduzibel“ und „nicht zerfallend“ dasselbe.

Bemerkungen:

1. Die Frage, ob sich ein Polynom in Faktoren zerlegen lässt, hängt entscheidend vom zugrunde gelegten Zahlkörper ab: Z. B. lässt sich das Polynom $x^2 + y^2$ im reellen Zahlkörper *nicht zerlegen*, jedoch gibt es im komplexen Zahlkörper sehr wohl die Zerlegung $x^2 + y^2 = (x + iy)(x - iy)$. Die Irreduzibilität einer algebraischen Kurve bezieht sich also stets auf den zugrunde gelegten Zahlkörper.

2. Auch die Frage „wie viele Punkte“ eine algebraische Kurve besitzt, ist ebenfalls vom zugrunde gelegten Zahlkörper abhängig: Z.B. hat die durch

$$x^2 + y^2 + 1 = 0 \tag{9.2}$$

dargestellte Kurve *überhaupt keinen* Punkt im \mathbb{R}^2 ist aber in \mathbb{C}^2 (der komplexen Ebene) ein nicht entarteter Kegelschnitt (da sich das Polynom in (9.2) auch in \mathbb{C} nicht zerlegen lässt).

3. Betrachtet man hingegen die Gleichung

$$x^2 + y^2 = 0, \tag{9.3}$$

so hat diese im Reellen genau eine Lösung, nämlich $x = 0, y = 0$. Die „Kurve“ besteht also im Reellen nur aus diesem einzigen Punkt. Wie schon in der ersten Bemerkung ausgeführt wurde, lässt sich ihr Polynom jedoch in zwei lineare Faktoren zerlegen, d. h. in der komplexen Ebene stellt (9.3) ein *Geradenpaar* dar (der oben genannte reelle Punkt ist deren Schnittpunkt).

4. Das Zerfallen einer algebraischen Kurve darf nicht mit der Möglichkeit verwechselt werden, dass eine Kurve im Reellen *nicht immer zusammenhängend* zu sein braucht. Z.B. kann es schon bei einer Kubik vorkommen,

dass ihr Bild aus einem ringförmigen Teil und (getrennt von diesem!) aus einem gestreckten Teil (mit einer geradlinigen Asymptote) besteht, obwohl die Kurve *irreduzibel* ist (siehe Abb. 9.1).

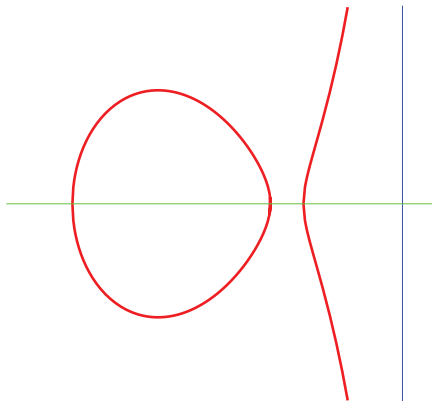


Abbildung 9.1: Eine „zweiteilige“ Kubik.

9.2 Tangenten

Bezeichnet man das Polynom, welches eine ebene algebraische Kurve \mathcal{C} darstellt mit P , was man auch durch die „Gleichung“

$$\mathcal{C}: P(x, y) = 0 \tag{9.4}$$

ausdrückt, so kann man in einem Punkt $(x_0, y_0)^T$ von \mathcal{C} (für welchen also $P(x_0, y_0) = 0$ ist!) eine Umentwicklung nach Potenzen von $(x - x_0)$ und $(y - y_0)$

vornehmen und die höheren (als die ersten) Potenzen weglassen.¹ Es bleibt dann eine *lineare Gleichung*

$$\frac{\partial P}{\partial x}(x_0, y_0)(x - x_0) + \frac{\partial P}{\partial y}(x_0, y_0)(y - y_0) = 0 \quad (9.5)$$

stehen. Diese stellt – sofern nicht beide Koeffizienten bei $(x - x_0)$ und $(y - y_0)$ verschwinden – eine *Gerade* dar, welche durch den Kurvenpunkt $(x_0, y_0)^T$ geht. Da es sich um die bestmögliche Approximation der Kurve durch eine Gerade in diesem Punkt handelt, ist es die *Tangente an \mathcal{C} im Punkt $(x_0, y_0)^T$* .

9.3 Singularitäten

Wie schon erwähnt, kann es vorkommen, dass beide Koeffizienten in (9.5) verschwinden, die Tangente also unbestimmt ist.

Definition 40. Ein Kurvenpunkt $\mathbf{p} := (x_0, y_0)^T$ heißt *singulär* (oder „eine Singularität der Kurve“), wenn

$$\frac{\partial P}{\partial x}(x_0, y_0) = 0, \quad \frac{\partial P}{\partial y}(x_0, y_0) = 0 \quad (9.6)$$

gilt. Die nicht singulären Punkte heißen regulär.

¹ Das Transpositionszeichen T soll andeuten, dass es sich um einen *Spalten*-Vektor handeln soll.

Wenn man nun bei einem singulären Punkt (quasi als „Ersatz“ für die lineare Approximation durch die Tangente) die *quadratischen Glieder* der oben genannten Umentwicklung, nämlich

$$\frac{\partial^2 P}{\partial x^2}(x_0, y_0)\xi^2 + 2 \cdot \frac{\partial^2 P}{\partial x \partial y}(x_0, y_0)\xi\eta + \frac{\partial^2 P}{\partial y^2}(x_0, y_0)\eta^2 = 0 \quad (9.7)$$

(wobei zur Abkürzung $\xi = x - x_0$ und $\eta = y - y_0$ gesetzt wurde) betrachtet, so stellt (9.7) ein *Geradenpaar* dar. Wenn die Diskriminante dieser quadratischen Form, d. h. die aus den Koeffizienten gebildete Determinante

$$\text{diskr} := \begin{vmatrix} \frac{\partial^2 P}{\partial x^2}(x_0, y_0) & \frac{\partial^2 P}{\partial x \partial y}(x_0, y_0) \\ \frac{\partial^2 P}{\partial x \partial y}(x_0, y_0) & \frac{\partial^2 P}{\partial y^2}(x_0, y_0) \end{vmatrix} \quad (9.8)$$

von Null verschieden ist, sind auch die beiden Geraden voneinander verschieden (sie können aber konjugiert komplex sein!).

Definition 41. Wenn $\text{diskr} \neq 0$ ist, heißt die Singularität ein *gewöhnlicher Doppelpunkt*; die beiden Geraden, in welche das genannte Geradenpaar zerfällt, heißen „*Tangenten im Doppelpunkt*“ (oder „*Doppeltangenten*“).

Man erkennt unschwer, dass es noch viele weitere Entartungsmöglichkeiten einer Singularität gibt; nicht nur die, dass $\text{diskr} = 0$ ist, was einer *Selbstberührung* entspricht. Im Buch von Walker [24] finden sich weitere Beispiele: Spitzen, Schnabelspitzen, Vielfach-Knoten etc. Jedem Typ einer Singularität kann man mit feineren algebraischen Methoden eine *Multiplizität* zuordnen; so haben insbesondere reguläre Punkte die Multiplizität 1, gewöhnliche Doppelpunkte die Multiplizität 2 etc. Diese Multiplizitäten sind bei der Zählung von Schnittpunkten und auch bei anderen Gelegenheiten (wie z. B. im folgenden Satz) zu berücksichtigen. Hinsichtlich der weiteren Einzelheiten muss auf die Literatur verwiesen werden. Es sei nur noch auf folgendes Resultat hingewiesen werden:

Satz 42. *Eine irreduzible algebraische Kurve vom Grad n hat höchstens $(n-1)(n-2)/2$ singuläre Punkte, wobei jeder mit seiner Multiplizität zu zählen ist.*

Hieraus folgt:

Korollar 43. *Wenn eine algebraische Kurve vom Grad n mehr als $(n-1)(n-2)/2$ singuläre Punkte besitzt, zerfällt sie.*

Aus diesem Grund kann ein nicht entarteter (=irreduzibler) *Kegelschnitt* (Grad $n = 2$) also *keinen singulären Punkt* enthalten; da $(n-1)(n-2)/2 = 0$ ist. Bei irreduziblen *Kubiken* gibt es höchstens *einen* singulären Punkt (vgl. das Beispiel der Kubik in Abbildung 9.2, S. 57).

9.4 Schnittpunkte mit Geraden

Die Schnittpunkte einer algebraischen Kurve mit einer Geraden erhält man einfachsten, wenn man für die Gerade eine lineare Parameter-Darstellung (vgl. Kapitel 4, (5.1) mit $n = 1$) wählt und diese in die Gleichung (9.1) einsetzt. Es entsteht dann *ein Polynom Q vom Grad n für die Variable t* ; die Nullstellen dieses Polynoms liefern die Parameter-Werte für die Schnittpunkte. Hieraus folgt sofort

Satz 44. *Eine Gerade schneidet eine ebene algebraische Kurve von Grad n in höchstens n Punkten, oder sie liegt ganz auf ihr; im zweiten Fall zerfällt die Kurve in diese Gerade und einen anderen Bestandteil vom Grad $n-1$ (der seinerseits auch wieder zerfallen kann).*

Ähnlich wie beim Fundamentalsatz der Algebra kann man für die verschiedenen Arten von Schnittpunkten *Vielfachheiten* (auch „Multiplizitäten“ genannt) einführen, jeden Schnittpunkt mit seiner Vielfachheit zählen und ebenso mit allen komplexen Schnittpunkten verfahren: Dann lässt sich dieser Satz dahingehend verschärfen, dass es *genau n Schnittpunkte* sind. (Dies ist ein Spezialfall des unten stehenden Satzes von Bézout mit $m = 1$.)

Bemerkungen:

1. Die Vielfachheit beträgt z. B. 1, wenn der Schnittpunkt ein regulärer Punkt der Kurve ist und die Gerade *von der Tangente im Schnittpunkt verschieden ist*.
2. Wenn die Gerade die Kurve in einem regulären Punkt berührt (also die Tangente in diesem Punkt ist), beträgt die Vielfachheit zwei. Die Vielfachheit beträgt ebenfalls zwei, wenn die Gerade durch einen gewöhnlichen Doppelpunkt geht.
3. Die oben erwähnte Verschärfung scheint nicht zu gelten, wenn man z. B. die Schnittpunkte der Parabel $y - x^2 = 0$ mit einer Geraden parallel zu ihrer Achse $x - a = 0$ betrachtet: Man findet sofort den Schnittpunkt $(a, a^2)^T$, aber wo ist der zweite? Es ist der Fernpunkt der Geraden, den sie mit der Achse und der Parabel gemeinsam hat, siehe 6.2. Beispiel 5.

Aus alledem ergibt sich die

Regel: *Bei allen Punkten auf algebraischen Kurven (einschließlich der Geraden), insbesondere bei Fragen über Schnittpunkte von solchen, sind nicht nur mögliche komplexe Punkte, sondern auch alle Fernpunkte mit in Betracht zu ziehen.*

9.5 Die Resultante zweier Polynome

Für den folgenden Abschnitt wird ein (auch sonst sehr nützliches) Ergebnis benötigt:

Satz 45. *Zwei Polynome*

$$f = a_0 + a_1 t + \dots a_n t^n \quad \text{mit } a_n \neq 0$$

und

$$g = b_0 + b_1 t + \dots b_m t^m \quad \text{mit } b_m \neq 0$$

haben genau dann einen gemeinsamen nicht konstanten Faktor, wenn die folgende Determinante verschwindet:¹

$$\begin{vmatrix} a_0 & a_1 & a_2 & \dots & a_n & & & \\ & a_0 & a_1 & a_2 & \dots & a_n & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ & & a_0 & \dots & & a_{n-1} & a_n & \\ b_0 & b_1 & b_2 & \dots & b_m & & & \\ & b_0 & b_1 & b_2 & \dots & b_m & & \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \\ & & b_0 & \dots & & b_{m-1} & b_m & \end{vmatrix} = 0. \quad (9.9)$$

¹ Dieser Satz gilt nicht nur für Polynome über \mathbb{R} , sondern allgemein für Polynomringe von kommutativen Ringen mit eindeutiger Primfaktorzerlegung.

Bemerkungen:

1. Diese Determinante nennt man *die Resultante oder die Sylvestersche Determinante* der beiden Polynome f und g .
2. Wenn zwei Polynome *eine gemeinsame Nullstelle* t_0 haben, besitzen sie den gemeinsamen Faktor $t - t_0$, also muss die Resultante verschwinden. Die Umkehrung gilt nur, wenn der Koeffizientenkörper algebraisch abgeschlossen ist, denn nur in solchen Körpern lassen sich alle Polynome in Linearfaktoren zerlegen. Dann impliziert ein gemeinsamer Faktor also stets eine gemeinsame Nullstelle.
3. Etwas allgemeiner als im Satz 45 behauptet, kann man zeigen, dass der Rang der Sylvesterschen Matrix gleich $n + m - k$ ist, wenn die beiden Polynome f und g einen gemeinsamen Faktor vom Grad k haben.

Einen vollständigen *Beweis* dieses Satzes findet man z. B. in [24]. Um einzusehen, wie man auf die obige Determinante kommt, soll wenigstens gezeigt werden, dass die Resultante verschwindet, wenn die beiden Polynome eine gemeinsame Nullstelle t_0 besitzen:

Man hat dann

$$a_0 + a_1 t_0 + \dots + a_n t_0^n = 0$$

und

$$b_0 + b_1 t_0 + \dots + b_m t_0^m = 0.$$

Man multipliziere die erste dieser beiden Gleichungen noch $(n - 1)$ -mal und die zweite noch $(m - 1)$ -mal mit t_0 . Daraus entsteht ein System von $n + m$ Gleichungen. Fasst man nun die Potenzen $t_0^0, \dots, t_0^{n+m-1}$ als *Unbekannte* auf, so hat man ein homogenes, lineares Gleichungssystem, welches nicht-trivial lösbar

sein soll. Dies trifft genau dann zu, wenn dessen Determinante verschwindet. Diese Determinante ist aber die Resultante.

9.6 Rationale Kurven als algebraische Kurven

Es soll nun gezeigt werden:

Satz 46. *Jede polynomiale oder rationale Kurve ist zugleich auch eine algebraische Kurve vom gleichen Grad $n > 0$.*

Beweis: Es genügt, die Behauptung für rationale Kurven nachzuweisen, denn die polynomialen sind auch rational mit dem Nenner 1. Es kommt darauf an, ein Polynom in x, y zu finden, welches von allen Punkten der Kurve und nur von diesen (mit endlich vielen Ausnahmen) „erfüllt“ wird, ohne dass der Kurvenparameter darin vorkommt; d. h. man muss den Parameter t *eliminieren*. Ausnahmen müssen zugelassen werden, weil z. B. die Nullstellen t_i des Nenners einer rationalen Kurve (d. h. deren *Fernpunkte*) keinem reellen Koordinatenpaar (x, y) zugeordnet werden können.

Man schreibt (6.1) für x und y getrennt, wobei $\mathbf{a}_k = (\alpha_k, \beta_k)^T$ gesetzt und mit dem Nenner heraufmultipliziert wird:

$$\begin{aligned}\sum_{k=0}^n (\alpha_k - x \cdot \gamma_k) t^k &= 0, \\ \sum_{k=0}^n (\beta_k - y \cdot \gamma_k) t^k &= 0.\end{aligned}\tag{9.10}$$

Dies sind zwei Polynome in t vom Grad n , und für jeden Wert $t = t_0$, der von einer Nullstelle des Nenners verschieden ist, ergeben sich nach (6.1) zwei Werte $x = x_0$ und $y = y_0$, welche (9.10) erfüllen. D. h. diese Polynome haben

die gemeinsame Nullstelle t_0 , wenn darin $x = x_0$ und $y = y_0$ eingesetzt wird. Nach Satz 45 verschwindet also die Sylvestersche Determinante. Wenn man diese Determinante auswertet, entsteht ein Polynom P in den Variablen x_0, y_0 vom Grad n , dessen Koeffizienten nur von $\alpha_k, \beta_k, \gamma_k$ ($k = 0, \dots, n$) abhängen und welches von allen Punkten der Kurve annulliert wird.

Ist umgekehrt für ein bestimmtes Wertepaar x, y die Gleichung $P(x, y) = 0$ erfüllt, so haben nach Satz 45 die Polynome (9.10) eine gemeinsame Nullstelle t , d. h. dieser Punkt liegt auf der rationalen Kurve, sofern es sich nicht um eine Nullstelle des Nenners handelt. \square

Dies soll an einem einfachen Beispiel illustriert werden: Der Kreis ist eine rationale Kurve mit der Darstellung

$$x = \frac{1-t^2}{1+t^2}, \quad y = \frac{2t}{1+t^2} \quad (9.11)$$

(wobei alle Punkte bis auf $(-1, 0)$ erreicht werden und dieser letzte Punkt als Grenzpunkt für $t \rightarrow \infty$ approximiert wird).

Nach (9.10) lauten die beiden Polynome f, g

$$f = (1-x) - (1+x)t^2, \quad g = y - 2t + yt^2,$$

woraus sich die Sylvestersche Determinante zu

$$D = \begin{vmatrix} 1-x & 0 & -(1+x) & 0 \\ 0 & 1-x & 0 & -(1+x) \\ y & -2 & y & 0 \\ 0 & y & -2 & y \end{vmatrix} \quad (9.12)$$

ergibt. Die Auswertung liefert

$$D = 4(x^2 + y^2 - 1),$$

und $D = 0$ stellt, wie erwartet, den (Einheits-)Kreis dar.

Es erhebt sich nun die Frage, ob auch die Umkehrung von Satz 46 gilt, d. h. ob etwa jede algebraische Kurve auch rational ist. Dies ist keineswegs der Fall! Es gilt vielmehr:

Satz 47. *Eine irreduzible algebraische Kurve vom Grad n ist rational, wenn*

$$(n-1) \cdot (n-2) = \sum_i m_i(m_i-1) \quad (9.13)$$

gilt, wobei die Summe über die singulären Punkte mit ihren Multiplizitäten m_i zu erstrecken ist.

Aus diesem Satz kann oft die Rationalität einer Kurve gefolgert werden:

1. Bei *nicht-entarteten Kegelschnitten* ist $n = 2$, die linke Seite von (9.13) also 0 und ebenso die rechte, da es keine Singularitäten gibt. Die Bedingung ist erfüllt, daher sind alle nicht-entarteten Kegelschnitte rational.
2. Sei \mathcal{C} eine irreduzible Kubik mit genau einem singulären Punkt der Multiplizität $m_1 = 2$. Auf der linken Seite von (9.13) steht 2 wegen $n = 3$; rechts ebenso, wie man sofort sieht. Daher ist \mathcal{C} rational. (Dies gilt insbesondere für die Kubik von Abbildung 9.2, welche man „Newtonscher Knoten“ nennt.)
3. Sei \mathcal{C} eine irreduzible algebraische Kurve vom Grad $n > 2$ mit der *Maximalzahl* $k = (n-1)(n-2)/2$ von singulären Punkten, alle mit der Multi-

plizität $m_i = 2$. Dann steht auf der rechten Seite von (9.13) $((n-1)(n-2)/2) \cdot 2$, was genau mit deren linker Seite übereinstimmt. Es gilt also das folgende Korollar.

Korollar 48. *Eine irreduzible algebraische Kurve vom Grad $n > 2$ mit der Maximalzahl $k = (n-1)(n-2)/2$ von singulären Punkten ist rational.*

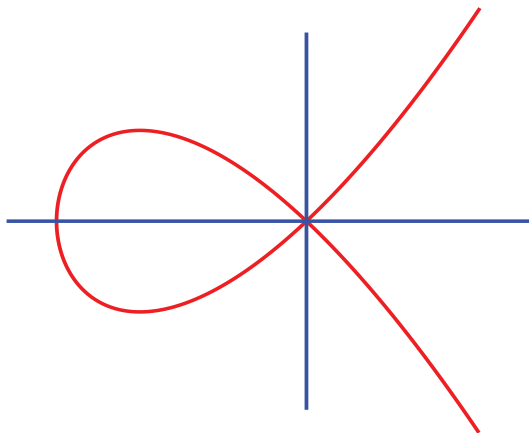


Abbildung 9.2: Der Newtonsche Knoten.

Der Satz sagt (leider!) nichts darüber aus, wie man die rationale Parameterdarstellung erhält. Bei Kubiken ist diese Aufgabe noch relativ einfach: Da eine rationale Kubik einen singulären Punkt haben muss, kann man das Geradenbüschel mit Zentrum in diesem Punkt betrachten: Von den drei Schnittpunkten jeder Büschelgeraden fallen zwei mit dem singulären Punkt zusammen, der dritte durchläuft alle übrigen Punkte der Kubik (vgl. Satz 44). Dadurch entsteht eine lokal bijektive Abbildung von \mathbb{R} (Büschelparameter) auf die Kubik, die als die gesuchte Parameterdarstellung dienen kann.

Als **Beispiel** für dieses Vorgehen werde die Kubik

$$x^2 + x^3 - y^2 = 0$$

genommen.¹ Sie hat bei $(0,0)$ einen gewöhnlichen Doppelpunkt. Daher nimmt man das Geradenbüschel

$$y = tx$$

durch diesen Punkt (mit t als Büschelparameter). Einsetzen in die Gleichung der Kubik liefert

$$x^2 + x^3 - t^2 x^2 = 0.$$

Die Doppellösung $x = 0$ gehört zur Singularität, die dritte Lösung

$$x = t^2 - 1$$

ist bereits die erste Gleichung der gewünschten Parameterdarstellung, und die zweite folgt daraus mit der obigen Gleichung für die Büschelgerade zu

$$y = t^3 - t.$$

9.7 Numerische Berechnung von Punkten

Bei algebraischen Kurven steht – im Gegensatz zu polynomialen Kurven, Splines oder, allgemeiner, rationalen Splines – *kein einfacher und allgemein gültiger Algorithmus* zur Berechnung ihrer Punkte Verfügung.

Es gibt aber eine Reihe von Fällen, in denen man die Punkte auf einfache Weise berechnen kann:

¹ Es handelt sich um den Newtonschen Knoten von Abb. 9.2.

1. Die algebraische Kurve ist *rational*. Dann besitzt sie eine Parameterdarstellung (6.1) (s. Kapitel 5). Damit können ihre Punkte unmittelbar berechnet werden.
2. In der Gleichung (9.1) kommt wenigstens eine der beiden Variablen x oder y nur in der ersten Potenz vor. Dann kann man die Gleichung nach dieser Variablen auflösen und erhält die betreffende Koordinate als eine Funktion der anderen.
3. Das Analoge gilt, wenn x oder y in kleinerer als 5. Potenz vorkommt. Im quadratischen Fall kann man auf (9.1) die Lösungsformel für quadratische Gleichungen (sogenannte „Mitternachtsformel“) anwenden; die Koeffizienten sind dabei Polynome in der anderen Variablen. Man erhält zwei „Äste“ der Kurve, je nach dem Vorzeichen der Wurzel. In den beiden restlichen Fällen der Grade 3 und 4 gibt es auch noch Auflösungsformeln, die man nach Cardano benennt (s. z. B. [20]). Diese sind allerdings recht umständlich und erfordern diverse Fallunterscheidungen, so dass deren Benutzung zwar möglich aber nicht ratsam ist.

Eine polynomiale Gleichung von höherem als viertem Grad lässt sich aber nicht mehr in „geschlossener Form“ (durch Wurzelausdrücke) auflösen (dies wird in der Algebra unter dem Stichwort „Galois-Theorie“ bewiesen). Daher ist man in diesen Fällen auf *numerische Näherungsverfahren* angewiesen.

4. Die relativ einfache Berechnung von Schnittpunkten mit Geraden durch Nullstellenbestimmung eines Polynoms kann insofern als Ersatz für die explizite Berechnung von Punkten der Kurve dienen, als man eine hinreichend dichte Schar von Geraden (etwa die Parallelen zu einer der Achsen) heranzieht und deren Schnittpunkte mit der Kurve berechnet. Die erhalte-

nen Punkte müssen dann zur gesamten Kurve in geeigneter Weise zusammengesetzt werden.

5. Betrachtet man die linke Seite von (9.1) als polynomiale Funktion

$$P : \mathbb{R}^2 \longrightarrow \mathbb{R},$$

so teilt diese die Ebene in drei Punktmengen ein, je nachdem, ob der Funktionswert $P(x, y) > 0$, < 0 oder $= 0$ ist. Die letzte Punktmenge ist definitionsgemäß die algebraische Kurve, welche die andern beiden Bereiche voneinander trennt. Der Graph der Funktion P kann als „Gebirge über der Ebene“ aufgefasst werden, wobei die algebraische Kurve die „Höhenlinie auf dem Niveau“ $z = 0$ ist.

Diese Vorstellung kann auf zweierlei Weise zur Berechnung der Kurvenpunkte ausgenutzt werden:

a) Wenn man bereits zwei (nahe beieinanderliegende) Punkte $\mathbf{p}_1 = (x_1, y_1)^T$ und $\mathbf{p}_2 = (x_2, y_2)^T$ mit $P(p_1) < 0$ bzw. $P(p_2) > 0$ kennt, so liegt zwischen diesen ein Kurvenpunkt, den man z. B. durch ein Halbierungsverfahren nun leicht berechnen kann.

b) Die beiden Koeffizienten in der Gleichung (9.5) stellen den *Normalenvektor* (orthogonal zur Tangente) dar; dies gilt nicht nur in einem Punkt \mathbf{p}_0 der Kurve, *sondern auch bezüglich jeder andern Höhenlinie!* Dieser Vektor (den man auch den *Gradientenvektor*

$$\text{grad } P(x_0, y_0) = \left(\frac{\partial P}{\partial x}(x_0, y_0), \frac{\partial P}{\partial y}(x_0, y_0) \right)^T$$

nennt) weist in Richtung des „steilsten Anstiegs“. Wenn also $P(\mathbf{p}_0) < 0$ ist, gelangt man in dieser Richtung am schnellsten auf das Nullniveau. Wenn $P(\mathbf{p}_0) > 0$ ist, steigt man in der Gegenrichtung am schnellsten ab. (Diese Richtung sollte dabei laufend im Sinne eines „Kompasses“ korrigiert werden.)

6. Schließlich sei noch auf folgendes Verfahren (das man gelegentlich als „Marching Algorithmus“ bezeichnet, s. Abbildung 9.3) hingewiesen: Man geht davon aus, dass bereits ein Anfangspunkt \mathbf{p}_0 und eine mutmaßliche Richtung der Kurventangente, etwa in Form eines Einheitsvektors \mathbf{v}_0 zur Verfügung stehe. (Diese Vorgaben können auf irgendeine andere Weise berechnet worden sein.)

Dann geht man Schritt für Schritt in Richtung der Tangente mit einer gegebenen Weite ds , welche die gewünschte Dichte der zu berechnenden Punktfolge auf der Kurve repräsentiert, voran und berechnet auf beiden Seiten des Endpunkts mit einer gewissen „Auslenkung“ in Normalenrichtung zwei weitere Punkte $\mathbf{r}_1, \mathbf{r}_2$.

Wenn nun die Kurve keine allzu große Krümmung besitzt, kann man erwarten, dass sie *zwischen diesen beiden Punkten hindurch läuft*. Man betrachtet nun die Funktion $f(t) := P((1-t) * \mathbf{r}_1 + t * \mathbf{r}_2)$, $t \in [0, 1]$, wobei P (wie beim vorigen Verfahren) die linke Seite von (9.1) ist. Diese hat beim Kurvenpunkt, dem ein t -Wert in $0 < t_0 < 1$ entspricht, eine *Nullstelle*. Dies kann mit einem allgemeinen Nullstellen-Bestimmungs-Verfahren (z. B. mit der „Regula Falsi“) (meist in wenigen Schritten) mit der gewünschten Genauigkeit berechnet werden. So erhält man also den nächsten Kurvenpunkt \mathbf{p}_1 .

Eine erste Entwurfsskizze dieses Algorithmus findet man hier im Anschluss. Er bedarf noch mancherlei Verfeinerungen; z. B. kann die Schrittweite durch die Krümmung (oder eine Näherung davon) gesteuert werden, damit das „Ausbrechen“ der Kurve aus dem Suchdreieck möglichst vermieden wird.



Global:	P, n	// Polynom auf der linken Seite
		// von (9.1) und dessen Grad
Input :	$\mathbf{p}_0, \mathbf{v}_0,$	// Start-Punkt und -Richtung
	ds, dn, \mathbf{z}	// Schrittweite, Auslenkung, Zielpunkt
Output:	\mathbf{p}, nn	// Array von Kurvenpunkten, „Füllstand“
Lokal :	$k, fin, error, \mathbf{q},$	
	$\mathbf{n}, \mathbf{r}_1, \mathbf{r}_2,$	
	$w_1, w_2, sig_1, sig_2,$	
	t_0	

begin $k := 0$ $fin := \text{false}; \text{error} := \text{false}$ **while** not fin and not $error$ **do** $\mathbf{q} := \mathbf{p}_i + ds \cdot \mathbf{v}$ $\mathbf{n} := \mathbf{v}^\perp$ $\mathbf{r}_1 := \mathbf{q} + dn \cdot \mathbf{n}$ $\mathbf{r}_2 := \mathbf{q} - dn \cdot \mathbf{n}$ $w_1 := P(\mathbf{r}_1); w_2 := P(\mathbf{r}_2)$ $sig_1 := \text{sign}(w_1); sig_2 := \text{sign}(w_2)$ **if** $sig_1 * sig_2 \neq -1$ **then** $\text{errormessage}(\text{'Marching gerät au\ss er Kontrolle'})$ $\text{error} := \text{true}; \text{exit}$ **else** $\text{regula falsi}(f, 0, 1, w_1, w_2, t_0, ok);$ **if** not ok **then** $\text{errormessage}(\text{'Regula Falsi versagt'})$ $\text{error} := \text{true}; \text{exit}$ **else** $i := i + 1$ $\mathbf{p}_i := (1 - t_0) \cdot \mathbf{r}_1 + t_0 \cdot \mathbf{r}_2$ $\mathbf{v} := (\mathbf{p}_i - \mathbf{p}_{i-1})$ $\mathbf{v} := \mathbf{v} / \|\mathbf{v}\|$ // Normierung **if** $\text{dist}(\mathbf{p}_i, \mathbf{z}) < ds$ **then** $\text{melde}(\text{'Zielpunkt erreicht'})$ $fin := \text{true}$ $nn := i$ **end if** **end if****end if**

```
end while
return p, nn
end
```

9.8 Schnittpunkte zweier algebraischer Kurven

Analog zum Satz über die Schnittpunkte einer algebraischen Kurve mit einer Geraden gibt es einen grundlegenden Satz über die Schnittpunkte *zweier* algebraischer Kurven:

Satz 49. (*Satz von Bézout*) *Zwei algebraische Kurven mit den Graden n und m haben entweder höchstens $n \cdot m$ Schnittpunkte oder sie (zerfallen beide und) haben einen gemeinsamen Bestandteil (und somit unendlich viele „Schnittpunkte“).*

Bemerkung: Auch dieser Satz lässt sich (allerdings mit noch tieferen algebraischen Methoden und feinerer Analyse der Multiplizitäten der Schnittpunkte) zu „genau $n \cdot m$ Schnittpunkte“ verschärfen.

Zur tatsächlichen *Berechnung der Schnittpunkte* ist vorweg zu bemerken, dass das Einsetzen einer Parameterdarstellung der einen Kurve in das Polynom der anderen, was bei einer Geraden zum Ziel führte, sich nur dann in analoger Weise durchführen lässt, wenn der kleinere der beiden Grade wenigstens bezüglich einer der beiden Variablen (x oder y) ≤ 4 ist. Dies liegt daran, dass man bei einem Grad > 1 die Gleichung der Kurve zuerst nach dieser Variablen auflösen muss, um zu einer Parameterdarstellung zu gelangen. Eine polynomiale Gleichung von höherem als viertem Grad lässt sich aber nicht mehr in „geschlossener Form“ (durch Wurzelausdrücke) auflösen.

Wie schon erwähnt, sind die Auflösungsformeln („Cardano’sche Formeln“) schon beim Grad 3 – und erst recht die vom Grad 4 – umständlich und werden daher kaum zu diesem Zweck benutzt.

Folglich muss man i. Allg. *numerische Verfahren* zur Schnittpunktberechnung anwenden. Z. B. kann man der einen Kurve „entlang gehen“ und dabei auf Vorzeichenwechsel des Polynoms der anderen Kurve achten. Dazwischen liegt dann ein Schnittpunkt, der ebenfalls mit der Regula Falsi leicht berechnet werden kann. Auf weitere Details muss im Rahmen dieser Zusammenstellung leider verzichtet werden.

Als ein Beispiel zum Satz von Bézout zeigt die folgende Abbildung die 6 Schnittpunkte einer Kubik (Grad 3) mit einer Ellipse (Grad 2).

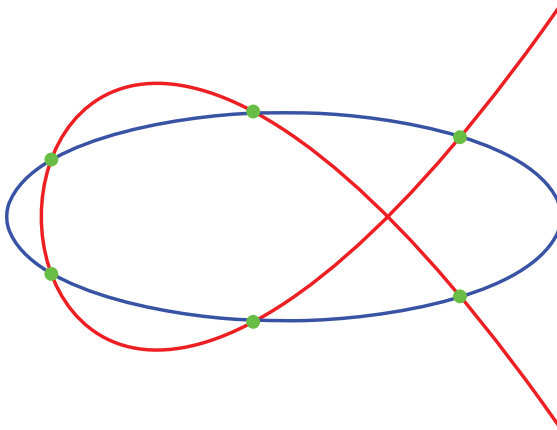


Abbildung 9.4: Schnitt einer Kubik mit einer Ellipse.

10 Weitere Spline-Konstruktionen

Von Splines war schon in den ersten vier Kapiteln die Rede. Insbesondere wurde im dritten Kapitel die Frage des glatten und die des krümmungsstetigen Anschlusses von Kurvensegmenten angesprochen (G^1 - und G^2 -Splines). Daneben besteht in der Praxis häufig die Anforderung, dass die Kurve durch eine Serie von gegebenen Punkten (in derselben Reihenfolge) hindurchgeht. Solche Splines nennt man *interpolierend*. Diese Fragen sind in den Anwendungen sehr bedeutsam; deshalb kommen wir hier nochmals auf sie zurück. Vorab sei daran erinnert, dass Splines *nicht* durch ihre Kontrollpunkte gehen, sondern lediglich in deren Nähe vorbei laufen (als ob sie von diesen angezogen würden). Daher kommen Splines bei Interpolations-Anforderungen weniger in Frage.

10.1 Der Overhauser-Spline

Eine sehr einfache Methode, zu einem interpolierenden G^1 -Spline zu gelangen, beruht auf folgender Idee: Man lege je eine Parabel durch drei aufeinanderfolgende Daten-Punkte, also etwa \mathcal{P}_0 durch die Punkte $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2$ und \mathcal{P}_1 durch $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$. Dann gehen offensichtlich zwei Parabelbögen von \mathbf{p}_1 nach \mathbf{p}_2 , und man kann einen *gleitenden Übergang* vom einen zum anderen wie folgt konstruieren: Seien

$$\mathcal{P}_0: \quad t \in [0, 1] \longrightarrow \mathbf{f}(t) \in \mathbb{R}^2$$

und

$$\mathcal{P}_1: t \in [0, 1] \longrightarrow \mathbf{g}(t) \in \mathbb{R}^2.$$

Parameterdarstellungen dieser beiden Parabelbögen, welche die Datenpunkte \mathbf{p}_1 und \mathbf{p}_2 verbinden, so leistet die Vektorfunktion

$$\mathcal{O}: \mathbf{x}(t) = (1-t)\mathbf{f}(t) + t\mathbf{g}(t), \quad t \in [0, 1] \quad (10.1)$$

das Gewünschte: Für $t = 0$ liefert (10.1) $\mathbf{x}(0) = \mathbf{f}(0) = \mathbf{p}_1$ und für $t = 1$ folgt $\mathbf{x}(1) = \mathbf{g}(1) = \mathbf{p}_2$, also interpoliert sie ebenfalls diese Punkte. Aber es gilt noch viel mehr: Durch Ableiten von (10.1) erhält man

$$\mathbf{x}'(t) = \mathbf{g}(t) - \mathbf{f}(t) + (1-t)\mathbf{f}'(t) + t\mathbf{g}'(t),$$

und weil die Differenz $\mathbf{g}(t) - \mathbf{f}(t)$ an den Intervallenden verschwindet, folgt

$$\mathbf{x}'(0) = \mathbf{f}'(0) \quad \text{und} \quad \mathbf{x}'(1) = \mathbf{g}'(1),$$

d. h. die *Gleitkurve* \mathcal{O} schließt bei \mathbf{p}_1 tangenstetig an den vorderen Bogen der Parabel \mathcal{P}_0 an, und das Analoge gilt bei \mathbf{p}_2 bezüglich des hinteren Bogens von \mathcal{P}_1 .

10.2 Unterteilungsalgorithmen

Um eine Kurve zu zeichnen (von einem Computer zeichnen zu lassen oder mittels eines graphischen Ausgabe-Geräts z. B. auf Papier zu plotten), wird in den meisten Fällen eine Punktfolge auf der Kurve gewählt (d. h. eine hinreichend dichte und monotone Folge von Parameterwerten festgelegt, und damit werden die entsprechenden Kurvenpunkte berechnet), welche dann stückweise geradlinig verbunden wird.

Eine gänzlich andere Technik wurde von N. Dyn, J. Gregory und D. Levin [10] eingeführt: Man geht aus von einer Punktfolge (\mathbf{p}_i) , die von einer glatten Kurve interpoliert werden soll. Dann „erfindet“ man ein Schema, wie man aus 3, 4 oder mehr benachbarten Punkten dieser Folge einen Zwischenpunkt konstruiert, und führt diese Konstruktion entlang der gegebenen (ursprünglichen) Folge durch. Man hat dann eine neue Folge mit etwa doppelt so vielen Punkten und kann die Konstruktion der Zwischenpunkte auf diese neue Folge anwenden. Dieses Verfahren wird solange wiederholt, bis eine visuell nicht mehr von einer glatten Kurve zu unterscheidenden Linie entsteht. Man nennt es *Unterteilungsalgorithmus* oder *-schema*.

Unterteilungsalgorithmen resultieren in einer sehr schnellen Generierung von Kurven. Das von den genannten Autoren vorgeschlagene *Vier-Punkte-Schema* benutzt (außer am Anfang und am Ende) jeweils vier aufeinanderfolgende Punkte und fügt einen Zwischenpunkt \mathbf{z} zwischen dem 2. und 3. Punkt ein; es lautet

$$\mathbf{z} = (-\mathbf{p}_{i-1} + 9\mathbf{p}_i + 9\mathbf{p}_{i+1} - \mathbf{p}_{i+2})/16. \quad (10.2)$$

Wie erwartet, liegt eine Affinkombination dieser vier Punkte vor, d. h. die Koeffizientensumme ist 1. Aber woher kommen die „Gewichte“ $-1, 9, 9, -1$? Sieht man sich eine passende Interpolationsformel, z. B. die von Lagrange oder die von Newton-Gregory an,¹ (siehe Abschnitt 6.8), so erkennt man: Es handelt sich um eine polynomiale Kubik, die diese 4 Punkte an den Stellen $0, 1/3, 2/3, 1$ interpoliert, und von dieser Kubik wird der „mittlere“ Punkt \mathbf{z} an der Stelle $t = 1/2$ genommen. Also eine ganz und gar klassische Angelegenheit der Interpolation!

¹ Beim zweiten Namen handelt es sich um James Gregory (1637-1675), *nicht* um den 3. Autor von [10] !

Beim ersten und letzten Punktepaar hat man keine vier Punkte in der angegebenen Lage zur Verfügung: Man benutzt dann entweder eine Parabel, die (z. B. am Anfang) $\mathbf{p}_0, \mathbf{p}_1, \mathbf{p}_2$ an den Stellen 0, $1/2, 1$ interpoliert und nimmt den Zwischenpunkt bei $t = 1/4$, d. h. $\mathbf{z} = (3\mathbf{p}_0 + 6\mathbf{p}_1 - \mathbf{p}_2)/8$ (siehe (7.19)) oder man verwendet eine geeignete Schätzung des Tangentenvektors bei \mathbf{p}_0 und berechnet die kubische Hermite-Interpolante an dieser Stelle; am Ende wird analog verfahren.

Die folgende Abbildung zeigt ein Beispiel des Vier-Punkte-Schemas, wobei die ursprünglichen Punkte in rot, die der 1. bis 4. Unterteilung in violett, gelb, grün und blau gezeichnet (und durch kleine Kreisscheiben hervorgehoben) sind.

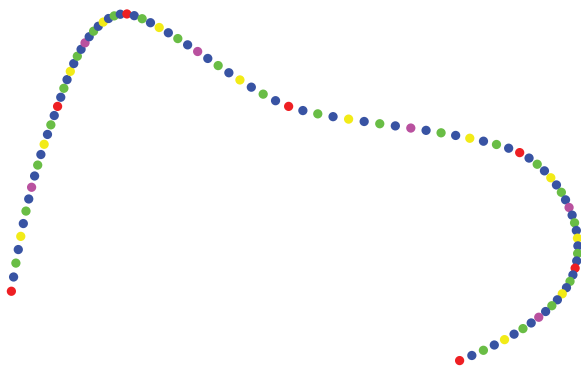


Abbildung 10.1: Eine Anwendung des Vier-Punkte-Schemas.

Für Kurven gibt es eine Reihe weiterer Unterteilungsschemata; die Idee der Unterteilung kommt aber hauptsächlich bei *Flächen* wegen der viel reichhaltigeren topologischen Möglichkeiten zum Tragen (vgl. [11], [12], [22]).

10.3 Abschneiden der Ecken (Corner cutting)

Eines der bekanntesten Eckenschnittverfahren (engl. corner cutting scheme) ist das von Chaikin [7]. Man geht wieder von einem Kontrollpolygon aus, dem sich die konstruierende Kurve annähern soll. Dann bestimmt man auf jeder Kante zwei neue Punkte bei $1/4$ und bei $3/4$ der Länge dieser Kante. Wie der Name des Verfahrens besagt, schneidet man die Ecke dadurch ab, dass man den $3/4$ -Punkt der vorigen Kante mit dem $1/4$ -Punkt der anschließenden Kante verbindet (und die Ecke weglässt). Die Zahl der Ecken verdoppelt sich dadurch annähernd. Dieses Verfahren wird solange wiederholt, bis keine Ecken mehr sichtbar sind.

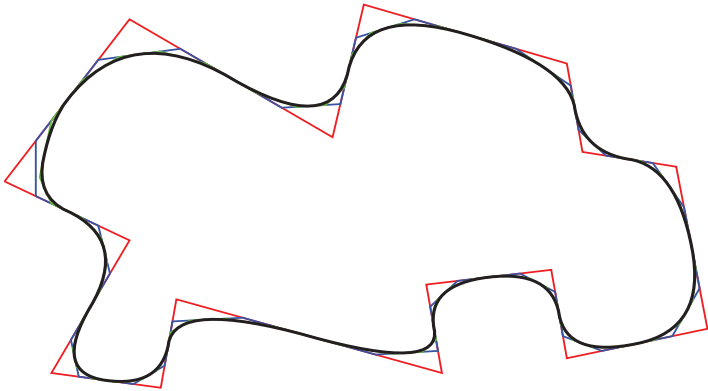


Abbildung 10.2: Eine Anwendung des Chaikin-Algorithmus.

Die mathematische Analyse zeigt, dass das Verfahren gegen einen quadratischen Spline konvergiert [17]. Übrigens liegen die Kantenmittelpunkte des ursprünglichen Kontrollpolygons auf der Grenzkurve, und die Kanten sind Tangenten an diese mit jenen als Berührungspunkte.

Allgemeines „Corner cutting“ wurde in [14, 5, 21] untersucht. Speziell Chai-kins Verfahren wurde auch auf Flächen übertragen [9] und stand Pate für die Klasse der stationären Unterteilungsalgorithmen. Einen Überblick findet man in [11] und in [12, Kap. 12/13].

11 Spezielle Kurven

Viele Kurven haben im Laufe der Geschichte der Mathematik eine herausragende Bedeutung erlangt, sei es wegen besonderer geometrischer Eigenschaften oder wegen technisch-mechanischer Anwendungen. Einige davon wie z. B. die logarithmische Spirale, die Schraublinie, die Klothoide oder der Newtonsche Knoten wurden in den vorangegangenen Kapiteln bereits besprochen.

Im Folgenden werden weitere dieser bekannten Kurven vorgestellt, ohne jedoch Vollständigkeit anzustreben oder mit dieser Auswahl eine Rangfolge zu präjudizieren. Außerdem werden die Kegelschnitte bewusst weggelassen, da diese in jedem Lehrbuch über Analytische Geometrie ausführlich behandelt werden.

11.1 Die Kettenlinie (Katenoid)

Die Kettenlinie ist schlichtweg ein geeigneter Ausschnitt der hyperbolischen Cosinus-Funktion in einem passenden Maßstab:

$$y = \cosh(x) = 1/2(e^x + e^{-x}). \quad (11.1)$$

Es handelt sich also um eine transzendente Kurve (weder rational noch algebraisch). Ihr Name ist dadurch gerechtfertigt, dass eine an ihren Enden aufgehängte Kette (annähernd) diese Form annimmt (siehe Abb. 11.1).

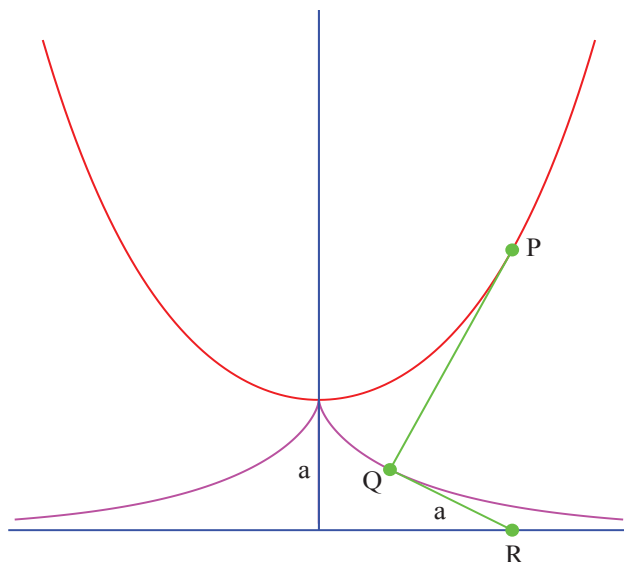


Abbildung 11.1: Die Kettenlinie mit Traktrix.

Es soll hier kurz skizziert werden, warum das so ist: Man denke sich die Kette ohne jeglichen Widerstand biegsam, so dass sie nur ihrem eigenen Gewicht unterworfen ist. Nach den Prinzipien der Mechanik nimmt sie dabei diejenige Form an, welche die *minimale potentielle Energie* realisiert. Wenn man – eine reale Kette weiter idealisierend – annimmt, dass das Material völlig gleichmäßig über die Länge der Kette verteilt ist, errechnet sich die potentielle Energie einer durch die Funktion $x \mapsto y(x)$ gegebenen Form der Kette (bis auf physikalische Konstanten wie z. B. das spezifische Gewicht etc.) durch das Integral

$$I = \int_{-a}^{+a} y \sqrt{1 + (y')^2} dx, \quad (11.2)$$

denn $y(x)$ gibt die Höhe über dem *Nullniveau* (x -Achse) an, und das Linienelement $\sqrt{1 + (y')^2} dx$ bestimmt das Gewicht eines infinitesimalen Stückchens der Kette.

Das Problem besteht nun darin, diejenige Funktion zu finden, welche das Integral I (bei festem Wert a der Intervallgrenzen) zum Minimum macht. Genau für derartige Aufgaben, ein Integral eines Ausdrucks, der von einer unbekannten Funktion y und deren Ableitung y' abhängt, zu minimieren, wurde die mathematische Disziplin der *Variationsrechnung* geschaffen (und weit über die geschilderte Aufgabe hinaus verallgemeinert).

Die Grundidee der Variationsrechnung ist die folgende: Man denke sich die Lösung bereits gegeben und biete sie in ein Feld von *Nachbarfunktionen* ein; d. h. man betrachtet eine Funktion von zwei Variablen x und ε

$$(x, \varepsilon) \mapsto y(x, \varepsilon), \quad x \in [a, b], \quad \varepsilon \in [-1, +1],$$

so dass für $\varepsilon = 0$ die gedachte Lösung angenommen wird. Dann hängt der Wert des Integrals von ε ab und nimmt bei $\varepsilon = 0$ ein Minimum an. Wenn alles stetig differenzierbar ist, folgt die Bedingung

$$\left. \frac{dI}{d\varepsilon} \right|_{\varepsilon=0} = 0 \tag{11.3}$$

(wofür man kurz $\delta I = 0$ schreibt.)¹

Ohne nun weiter auf die Details der Variationsrechnung einzugehen, sei sogleich deren Hauptergebnis (für den vorliegenden Fall) genannt: Eine notwen-

¹ Man nennt den linksseitigen Ausdruck „die (erste) Variation von I “.

dige Bedingung für (11.3) ist das Bestehen der Differentialgleichung (sogenannte „Euler-Lagrange- Gleichung“):

$$\frac{\partial F}{\partial y} - \frac{d}{dx} \left(\frac{\partial F}{\partial y'} \right) = 0, \quad (11.4)$$

wobei der „Ausdruck“ des Integranden mit F bezeichnet wurde und natürlich eine stetig differenzierbare Funktion der *drei* Veränderlichen x, y, y' sein muss.

Im vorliegenden Fall ist

$$F(x, y, y') = y \sqrt{1 + (y')^2};$$

damit errechnet sich die Euler-Lagrangesche Differentialgleichung nach einigen Umformungen zu

$$1 + (y')^2 - yy'' = 0. \quad (11.5)$$

Es müsste nun gezeigt werden, dass die hyperbolische Cosinusfunktion die einzige ist, welche (bei passenden Randbedingungen) dieser Differentialgleichung genügt. Es wird hier jedoch auf den Eindeutigkeitsnachweis verzichtet und nur gezeigt, dass \cosh (11.5) tatsächlich erfüllt.

Dies folgt einfach aus den folgenden Beziehungen für $y = \cosh(x)$:

$$y'(x) = \sinh(x), \quad y''(x) = \cosh(x), \quad 1 + \sinh(x)^2 = \cosh(x)^2. \quad (11.6)$$

Damit ist die physikalische Bedeutung der Kettenlinie erklärt.

Die Kettenlinie hat aber noch weitere interessante Eigenschaften: Erstens folgt aus den Gleichungen (11.6) unmittelbar, dass das Bogenelement $ds = \sqrt{1 + y'(x)^2} dx = \cosh(x) dx$ ist, so dass dieses wegen $\cosh(x) = \sinh(x)'$ elementar

integrierbar ist; die Länge des vom Scheitel aus gemessenen Bogens bis zum Punkt $(x, \cosh(x))$ ist demnach einfach

$$s = \sinh(x). \quad (11.7)$$

Zweitens wird sich gleich zeigen, dass diejenige Evolvente der Kettenlinie, welche durch ihren Scheitel (Tiefpunkt) geht, genau mit einer Kurve – die zuvor auf andere Weise definiert wurde und „Traktrix“ genannt wird – übereinstimmt.

Diese Kurve wird im folgenden Abschnitt besprochen. Zur Vorbereitung dieser zweiten Eigenschaft wird hier schon einmal die genannte Evolvente berechnet. Man braucht dazu nur die Länge s nach rückwärts auf der Tangente im Punkt P (s. Abb. 11.1) abzutragen; zunächst berechnet sich der Tangenten-Einheitsvektor \mathbf{v} zu

$$\mathbf{v} = 1/\cosh(x)(1, \sinh(x)),$$

und daraus folgt sofort

$$Q = P - \sinh(x)\mathbf{v} = (x - \tanh(x), 1/\cosh(x)). \quad (11.8)$$

Dies ist also eine Parameter-Darstellung der Evolvente, wobei die x -Koordinate des „laufenden“ Punkts P auf der Kettenlinie als Parameter dient.

11.2 Die Traktrix

Bereits im Altertum, genauer in der griechischen Blütezeit der Mathematik um 300 v. Chr., entstand die Frage nach einer Kurve mit der Eigenschaft, dass für jeden Punkt P der Kurve, der Abschnitt von dem im Punkt P an die Kurve gelegten Tangente bis zu deren Schnittpunkt S mit einer *Leitgeraden* eine

konstante, d. h. von P unabhängige, Länge hat. Man hatte zwar eine ungefähre Vorstellung, wie eine solche Kurve, die man „Traktrix“ nannte,¹ aussehen müsste, aber eine exakte mathematische Darstellung konnte man nicht finden. Erst Chr. Huygens glückte 1693 eine Lösung dieses Problems.

Es mag sein, dass diese Aufgabenstellung von folgender Alltagssituation herührt: Ein Herrchen geht mit seinem Hund an der Leine (fester Länge) auf einem kerzengeraden Weg spazieren. Der Hund springt auf die daneben liegende Wiese und schnüffelt, wobei die gestreckte Leine orthogonal zur Richtung des Weges ist. Herrchen will nun weiterlaufen und zieht den (unwilligen) Hund an der Leine zu sich; dieser bewegt sich (notgedrungen) stets genau in Richtung der Leine bis er (annähernd!) im Abstand der Leinenlänge hinter Herrchen herläuft. Welche Kurve beschreibt der Weg des Hundes?

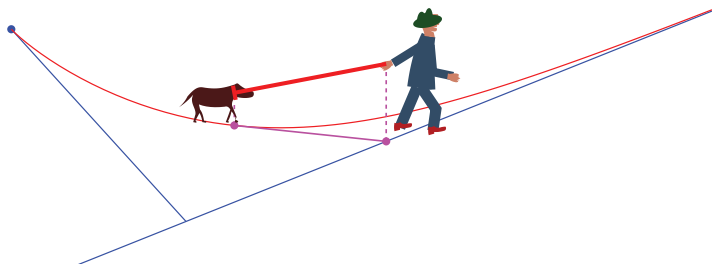


Abbildung 11.2: Illustration zur Traktrix.

Diese kleine Geschichte erklärt zugleich den Namen der fraglichen Kurve, denn er kommt von „trahere“ (lat. = ziehen).

¹ Eine Traktix ist in Abb. 11.1 in violett zu sehen.

Mit derzeitigen Methoden stellt man aus der geometrischen Beschreibung die folgende Bedingung für die Funktion $x \mapsto y(x)$ ($x \geq 0$), welche den rechten Ast der Traktrix darstellen soll, auf:¹

$$\left(\frac{y}{y'}\right)^2 + y^2 = a^2; \quad (11.9)$$

dabei ist a die konstante Länge des genannten Tangentenabschnitts. Es handelt sich also wieder um eine (gewöhnliche) Differentialgleichung, diesmal von erster Ordnung, welche man auch in der Form

$$y' = -\frac{y}{\sqrt{a^2 - y^2}} \quad (11.10)$$

schreiben kann. Durch *Trennung der Variablen* (eine gängige Methode zur Lösung von Differentialgleichungen von bestimmter Gestalt, z. B. wie im vorliegenden Fall, wo die Variable x nicht explizit auf der rechten Seite vorkommt) folgt durch Kehrwertbildung und (unbestimmte) Integration

$$x = \int \frac{\sqrt{a^2 - y^2}}{y} dy. \quad (11.11)$$

Das auf der rechten Seite stehende Integral ist elementar auswertbar und liefert

$$x = \sqrt{a^2 - y^2} - a \operatorname{arccosh}\left(\frac{a}{y}\right) \quad (0 < y \leq a), \quad (11.12)$$

wobei der Anfangswert $x = 0, y = a$ bereits berücksichtigt ist.

Damit kann das Problem als gelöst gelten, auch wenn mit (11.12) nur die Umkehrfunktion $y \mapsto x(y)$ dargestellt wird.

¹ Der linke Ast ist zur y -Achse spiegelsymmetrisch.

Es bleibt noch zu zeigen, dass die in Abschnitt 9.1 genannte Evolvente der Kettenlinie tatsächlich die Traktrix (mit $a = 1$) ist. Dazu wird Q nach x abgeleitet und der zugehörige Einheitsvektor gebildet: Nach (11.8) folgt

$$Q' = \frac{\sinh(x)}{\cosh^2(x)} (\sinh(x), -1)$$

und damit

$$\mathbf{w} = \frac{1}{\|Q'\|} Q' = \frac{1}{\cosh(x)} (\sinh(x), -1). \quad (11.13)$$

Man muss nun den Faktor λ so bestimmen, dass der Punkt

$$R = Q + \lambda \mathbf{w}$$

auf der x -Achse liegt. Da Q nach (11.8) und \mathbf{w} nach (11.13) gegen gleiche y -Koordinaten haben, folgt

$$\lambda = 1 \quad (\text{unabhängig von } x!),$$

womit die Behauptung bewiesen ist. Zugleich wurde mit (11.8) eine *Parameterdarstellung der Traktrix* gewonnen.

Das der Traktrix zugrunde liegende Problem wurde mehrfach verallgemeinert, insbesondere dahingehend, dass die Leitlinie auch eine gekrümmte Kurve sein kann oder dass der Abstand des Verfolgers auch abnehmen kann. Man spricht dann von allgemeinen „Verfolgungsproblemen“, welche vielfältige Anwendungen im Straßenbau (die Lenkung an den Vorderachsen zieht ein längeres Fahrzeug hinter sich her) bis hin zu Abwehrraketen haben.

11.3 Hundekurven

Hundekurven wurden ebenfalls schon im Altertum betrachtet, vor allem von Nikomedes, einem Mathematiker aus dem Königsgeschlecht von Bythien. Wegen ihrer Form nennt man sie auch *Konchoiden*, zu deutsch „Muschelkurven“. Einige von ihnen sind der (echten) Traktrix sehr ähnlich (und werden deshalb öfters mit dieser verwechselt); sie sind aber sehr viel einfacher und werden hier nur besprochen, weil sie in den Fragenkreis der Verfolgungsprobleme gehören aber von der eigentlichen Traktrix sehr wohl zu unterscheiden sind.

Um die Familie der Hundekurven zu beschreiben, gehe man wieder von einer festen Geraden, der *Leitlinie* l , und einem festen Punkt A im Abstand $a > 0$ von ihr aus. Dann betrachte man das Geradenbüschel mit Zentrum A und trage auf jeder Geraden g des Büschels, die nicht zur Leitlinie parallel ist, vom Schnittpunkt $S = g \cap l$ je eine Strecke der festen Länge b nach beiden Seiten hin ab; die Endpunkte X_1, X_2 beschreiben (in Abhängigkeit des Büschelpara-

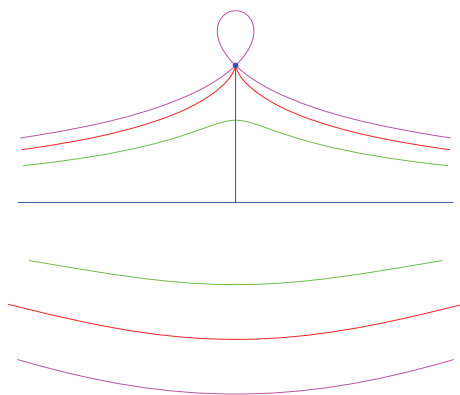


Abbildung 11.3: Hundekurven.

metrs) zwei Äste der Hundekurve (s. Abb. 11.3). Je nach dem Verhältnis b/a entstehen verschiedene Formen dieser Kurven.

11.4 Spiralen

Die logarithmische Spirale wurde schon im Abschnitt 2.5 als Beispiel einer Kurvendarstellung in Polarkoordinaten vorgestellt. Es wurde darauf hingewiesen (Satz 8), dass sie die Geraden des Büschels durch den Ursprung unter konstanten Winkeln schneidet. Dasselbe gilt dann auch für die konzentrischen Kreise um diesen Punkt.

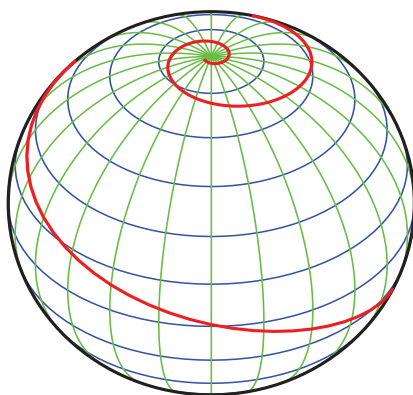


Abbildung 11.4: Eine Loxodrome.

Wenn man nun die logarithmische Spirale *stereographisch auf die Kugel projiziert*,¹ so entsteht eine räumliche Kurve auf der Kugeloberfläche, welche man

¹ Die stereographische Projektion bildet einen Punkt X der x - y -Ebene auf den zweiten Schnittpunkt \bar{X} der Geraden NX mit der Einheitskugel ab; dabei ist N der „Nordpol“ der Kugel (mit Koordinaten $(0,0,1)$).

Loxodrome nennt. Da die stereographische Projektion eine *winkeltreue Abbildung* ist, schneidet auch die Loxodrome sowohl die Meridiane als auch die Breitenkreise unter *konstanten Winkeln*. Dies war in früheren Zeiten, als es weder Funkpeilung noch GPS gab, ein wichtiges Hilfsmittel zu Navigation auf hoher See, da das Einhalten eines festen Kurswinkels relativ gut mittels Theodoliten möglich war.

Es sei jedoch darauf hingewiesen, dass ein Teilbogen der Loxodrome *nicht die kürzeste Entfernung* (auf der Kugeloberfläche zwischen ihren Endpunkten ist). Auf einer gekrümmten Fläche kommen dafür nur die *geodätischen Linien* in Frage. Diese sind durch folgende Eigenschaft gekennzeichnet: Man bilde längs einer Kurve auf der Fläche die von den Tangentenebenen eingehüllte Torse und wickle diese in die Ebene ab; wenn sich dabei das Bild der Flächenkurve als *geradlinig* erweist (und nur dann), ist die ursprüngliche Flächenkurve eine geodätische Linie.

Man erkennt unschwer, dass auf der Kugel alle *Großkreise* (und nur diese) geodätische Linien sind.

Bemerkenswert ist, dass sich die Loxodrome nicht nur um den Südpol, dem Bild des Koordinatenursprungs (= Mittelpunkt der Kugel), sondern auch um den Nordpol unendlich oft spiralig herumwindet. Das rührt daher, dass bei der stereographischen Projektion die „Fernpunkte“ der Geraden durch den Ursprung *auf einen einzigen Punkt*, dem Nordpol der Kugel, abgebildet werden. Dies entspricht der sogenannte „Ein-Punkt-Kompaktifizierung“ der euklidischen Ebene, die diese also topologisch äquivalent zu einer Kugel macht, ganz im Gegensatz zum *projektiven Abschluss* (vgl. Abschnitt 7.4), welcher eine *nicht orientierbare Fläche vom Geschlecht Null* liefert.

Es gibt noch eine *zweite Spirale*, die man „Archimedische Spirale“ nennt; sie ist dadurch gekennzeichnet, dass der Radius r der Kurve (bei Polarkoordinaten) nicht exponentiell, sondern *linear* mit dem Polarwinkel ϕ anwächst:

$$r = c\phi \quad \text{oder} \quad x = c\phi \cos(\phi), \quad y = c\phi \sin(\phi). \quad (11.14)$$

Dies hat zur Folge, dass Satz 8 für die Archimedische Spirale nicht mehr gilt; andererseits haben aber die einzelnen Windungen annähernd gleichen Abstand voneinander, da der Polarradius bei einer Umdrehung um $2\pi c$ anwächst.

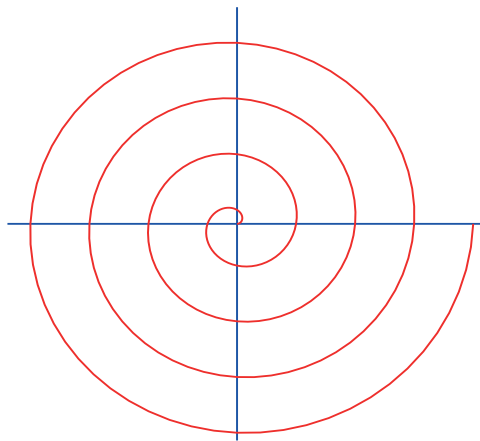


Abbildung 11.5: Die Archimedische Spirale.

11.5 Die Cassinischen Ovalen

Die Cassinischen Ovalen sind ebenfalls eine Familie von Kurven, welche eine gewisse Verwandtschaft zu den Kegelschnitten aufweisen. Sie haben – wie jene – zwei Brennpunkte im Abstand $2a$, um welche sie herum verlaufen. Im Gegensatz zu den Kegelschnitten ist jedoch nicht die Summe der Abstände eines Kurvenpunkts zu den Brennpunkten, sondern *das Produkt* eine Konstante b^2 . Bei $b > a$ haben diese Kurven eine *ovale Form*, was ihnen den besagten Namen gab; bei $0 < b < a$ bestehen sie aus *zwei Ästen*, die etwas eiförmig um je einen Brennpunkt herum verlaufen. Interessant ist der Fall $b = a$, bei dem *die Lemniskate* entsteht: Diese hat die Form einer liegenden Acht, deren Doppelpunkt in der Mitte zwischen den Brennpunkten liegt (s. Abb. 11.6).

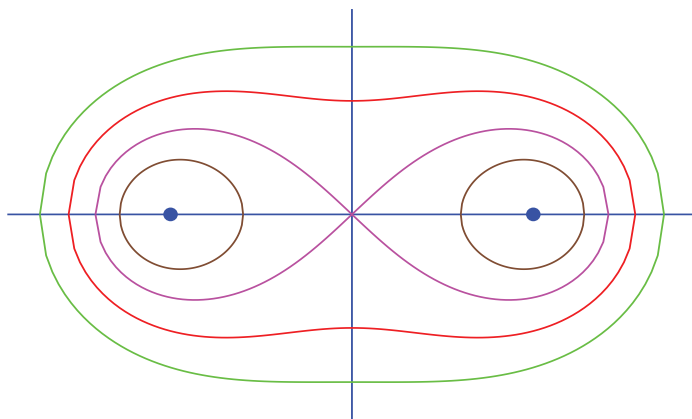


Abbildung 11.6: Einige Ovalen von Cassini.

Die Gleichung der Cassinischen Ovalen erhält man sofort aus der Definition zu

$$(x^2 - a^2)^2 + 2y^2(x^2 + a^2) + y^4 = b^4. \quad (11.15)$$

Es handelt sich also um algebraische Kurven 4. Ordnung (und dieses Beispiel zeigt einmal mehr die Reichhaltigkeit dieser Klasse von Kurven).

11.6 Die Zykloide und die Trochoiden

Eine ganz andere Klasse von Kurven kommt durch die *Kinematik* ins Spiel. Befestigt man auf der Peripherie eines Kreises einen Punkt **P** und lässt den Kreis auf einer Geraden ohne zu gleiten rollen, so beschreibt **P** eine Kurve, die man *Zykloide* (s. Abb. 11.7) nennt. Ihre Parameterdarstellung ergibt sich aus dem geschilderten Bewegungsvorgang unmittelbar zu

$$x = r\phi - \rho \sin(\phi) \quad y = -\rho \cos(\phi) \quad (11.16)$$

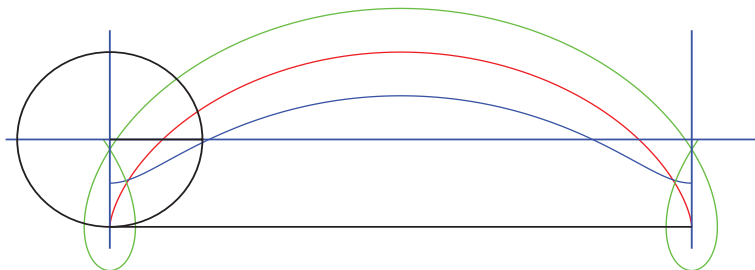


Abbildung 11.7: Gewöhnliche Zykloiden.

Die Zykloide erlangte in der Vergangenheit aus drei Gründen eine gewisse Bedeutung:

1. Das Zykloidenpendel

Stellt man eine Zykloide „auf den Kopf“ und lässt einem in der Spitze befestigten Pendel nur den Platz zwischen den beiden von dort ausgehenden

Zykloidenbögen frei, und zwar so, dass die biegsame Leine immer einen dieser Bögen berührt, so – wie man durch Nachrechnen bestätigen kann – die Frequenz *unabhängig vom Pendelausschlag* (Ch. Huygens 1673). Man nennt die gewöhnliche Zykloide aus diesem Grunde auch „Tautochrone“.¹

Man hat versucht, diese Tatsache zur Erhöhung der Ganggenauigkeit von Uhren auszunutzen. Dies hat sich aber nicht sonderlich bewährt, da einerseits die Führung der Leine (oder Kette) entlang der Zykloidenbögen zu aufwendig ist und andererseits weitere Einflüsse (wie z. B. die Wärmeausdehnung der Pendellänge) überwiegen.

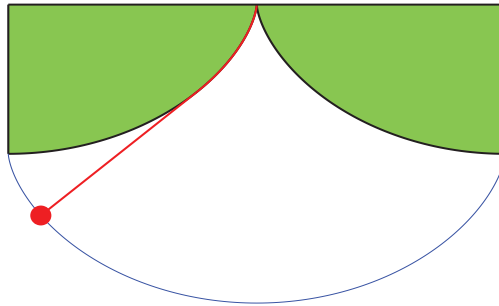


Abbildung 11.8: Das Zykloiden-Pendel.

2. Zykloidenverzahnung

Bei der Konstruktion von Zahnrädern kommt es darauf an, dass die beiden Zahnflanken, die sich momentan im Eingriff befinden, nach Möglichkeit auf einander *abrollen*, ohne zu gleiten, und dass der Eingriffspunkt einen möglichst konstanten Abstand von den beiden Mittelpunkten der Zahnräder einhält (damit das Übersetzungsverhältnis ebenfalls konstant bleibt und das Getriebe nicht rattert). Ohne die Details hier näher auszuführen,

¹ Es sei noch erwähnt, dass die Bahn des Pendelmittelpunkts (also die Evolvente der Zykloide) ebenfalls eine Zykloide ist, die zur ersten kongruent ist.

sei erwähnt, dass es dafür im wesentlichen zwei Lösungen gibt: Die eine heißt *Zykloidenverzahnung*, weil die Zahnflanken als gewisse Teilbögen einer Zykloide ausgebildet sind; die andere heißt *Evolventenverzahnung*. Näheres findet man z. B. in [26].

3. Brachystochrone

Die dritte bemerkenswerte Eigenschaft der Zykloide ist ebenfalls mechanischer Natur: Denkt man sich einen Massenpunkt auf einer abschüssigen Bahn reibungslos hinabgleiten (etwa bis zu einem Nullniveau), so entsteht die Frage, welche Form diese Bahn haben muss, damit der Massenpunkt in der *kürzesten Zeit* unten ankommt. Die Lösung ist wiederum die gewöhnliche Zykloide, wobei die Spitze im Anfangspunkt liegt. Die Zykloide ist also eine *Brachystochrone* (von griech. *brachys* = kurz, *chronos* = Zeit).

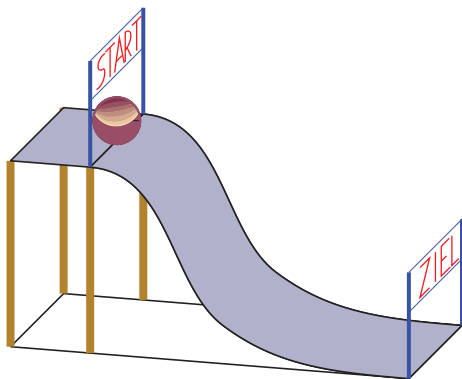


Abbildung 11.9: Illustration zur Brachystochrone.

Diese interessante Eigenschaft der Zykloide verdient, dass auf ihre mathematische Herleitung näher eingegangen wird.

Hierzu wird die nach unten gerichtete Distanz des Massenpunkts von der Ausgangslage (Plattform in Höhe h) mit η bezeichnet; diese Größe η kann als (monoton wachsende) Funktion einer der folgenden Parameter betrachtet werden:

- der Bogenlänge s entlang der Bahn,
- der Zeit t vom Beginn des Laufs an,
- der Geschwindigkeit v des Massenpunkts,
- der Koordinate x in horizontaler Richtung
der aktuellen Lage des Massenpunkts.

Welcher Parameter gerade gemeint ist, ergibt sich aus dem Zusammenhang, so dass das Argument weggelassen werden kann. Die folgende Skizze verdeutlicht dies:

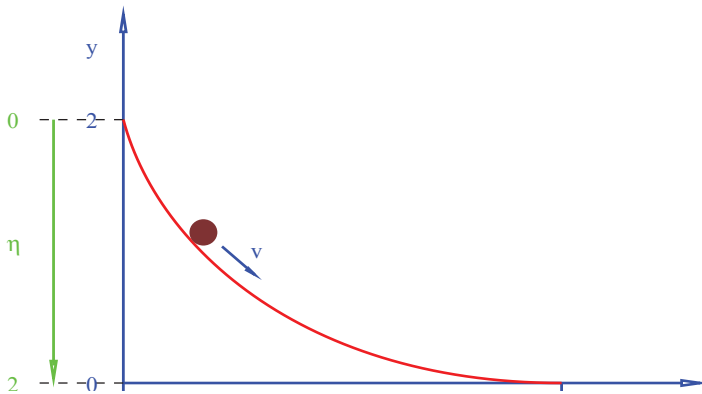


Abbildung 11.10: Zu den Bezeichnungen.

Die Grundidee zur Lösung ist es, die *Energiegleichung*

$$m g \eta = \frac{1}{2} m v^2$$

heranzuziehen. Aus dieser folgt sofort

$$v = \frac{ds}{dt} = \sqrt{2g\eta} \quad \text{oder} \quad \frac{dt}{ds} = \frac{1}{\sqrt{2g\eta}}.$$

Deren Integration liefert

$$T = t_1 - t_0 = \int_0^a \frac{ds}{\sqrt{2g\eta}} \quad \text{oder} \quad T = \int_0^a \sqrt{\frac{1 + (\eta')^2}{2g\eta}} \stackrel{!}{=} \min.$$

Auch hier liegt wiederum ein Problem der Variationsrechnung vor! Abgesehen vom konstanten Faktor $2g$, lautet die Intergranthenfunktion

$$F(\eta, \eta') = \frac{\sqrt{1 + (\eta')^2}}{\eta} \tag{1}$$

(mit $\eta' = d\eta/dx$). Die Euler-Lagrange-Gleichung (s. Abschnitt 10.1) kann nach einigen Umrechnungen auf die folgende Form

$$1 + (\eta')^2 + 2\eta\eta'' = 0$$

gebracht werden. Durch Anwendung einschlägiger Lösungsverfahren für gewöhnliche Differentialgleichungen und weiterer Rechnung erhält man schließlich unter Berücksichtigung der Anfangsbedingungen die Lösung

$$\eta = 1 - \cos(\phi), \quad x = \phi - \sin(\phi).$$

Man erkennt nun, dass es sich dabei – wie erwartet – um die Parameterdarstellung der (normalisierten) Zykloide handelt (welche an der Grundlinie gespiegelt ist)!

Es gibt noch eine weitere – sehr erstaunliche – Eigenschaft der Zykloide als Brachystochrone:

Bemerkung 50. *Die Zeit, die der Massenpunkt benötigt, um an den Endpunkt der Bahn (auf dem Nullniveau) zu gelangen, ist unabhängig vom Anfangspunkt (auch wenn dieser niedriger als die maximale Höhe h liegt).*

Hierauf beruht das oben schon erwähnte Zykloidenpendel von Huygens.

Trochoiden

Die Erzeugung der Zykloide kann man in vielerlei Hinsicht verallgemeinern: Statt den Punkt **P** auf der Peripherie eines Kreises zu fixieren, kann man ihn in dessen Innerem oder im Äußeren befestigen; dementsprechend entstehen wellenförmige Kurven oder solche mit Schleifen. Weiterhin kann man den Kreis statt auf einer Geraden auch auf einem anderem Kreis – entweder von innen oder auch von außen – rollen lassen. Die entstehenden Kurven heißen Hypozykloiden bzw. Epizykloiden. Solche Rollungen von Kreisen auf Kreisen kann man natürlich iterieren, wodurch die *höheren Radlinien* mit einer Vielzahl von Formen entstehen. Allesamt (einschließlich der Zykloiden) werden *Trochoiden* genannt (Ph. de la Hire 1706 und L. Euler 1781).

Schließlich sei erwähnt, das die Trochoiden i. Allg. *zwei Erzeugungen* besitzen; z. B. kann bei der Epizykloide der bewegliche Kreis außen auf dem festen Kreis abrollen oder ein größerer Kreis kann den festen Kreis in seinem Inneren enthalten und sich um ihn herumwinden (wie ein Hula-Hoop-Reifen). Näheres findet man in [26].

11.7 Die Kleeblattschlinge

Die Kleeblattschlinge (s. Abb. 11.11) ist eine der wenigen Kurven mit einer dreizähligen Symmetrie (Rotationsuntergruppe C_3 , die von einer Drehung um den singulären Punkt mit Winkel 120 Grad erzeugt wird). Ihre Gleichung lautet

$$(x^2 + y^2)^2 + 3x^2y - y^3 = 0. \quad (11.17)$$

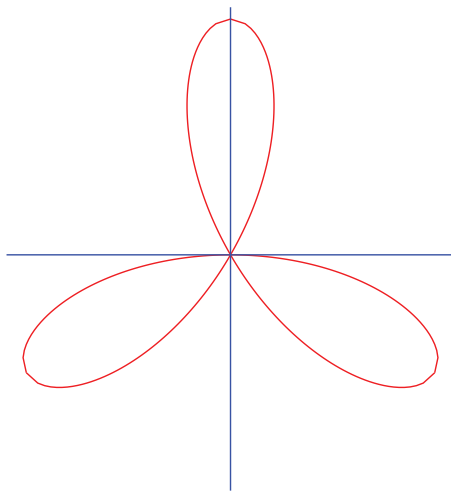


Abbildung 11.11: Die Kleeblattschlinge.

Es handelt sich also um eine algebraische Kurve 4. Ordnung. Im Ursprung des Koordinatensystems hat sie einen (gewöhnlichen) Tripelpunkt. Die drei Tangenten der Kurvenzweige, die durch diese Singularität hindurch laufen, erhält man durch weglassen der Glieder 4. Ordnung (analog zur Beschreibung der Tangenten in einem gewöhnlichen Doppelpunkt in Abschnitt 7.3) zu

$$y = 0, \quad y - \sqrt{3}x = 0, \quad y + \sqrt{3}x = 0. \quad (11.18)$$

Interessant ist hierbei auch die folgende Frage: Welche algebraischen Kurven können überhaupt diese Symmetriegruppe C_3 besitzen? - Offenbar kann man stets den Koordinatenursprung in das Drehzentrum legen, so dass keine linearen Glieder in der Gleichung vorkommen; auch das Absolutglied spielt für diese Frage keine Rolle: Wenn es (wie bei der Kleeblattschlinge) verschwindet, liegt der Ursprung auf der Kurve und diese hat dort eine Singularität. Die Gleichung einer fraglichen Kurve der in Rede stehenden Art kann man nun nach dem Grad ihrer Glieder ordnen:

$$F_2(x,y) + F_3(x,y) + \dots F_n(x,y) = 0,$$

wobei $F_k(x,y)$ *homogene Formen* vom Grad k bez. x,y sind.

Da die C_3 von der Drehung

$$\begin{aligned} x &= -1/2\bar{x} - \sqrt{3}/2\bar{y} \\ y &= \sqrt{3}/2\bar{x} - 1/2\bar{y} \end{aligned} \quad (11.19)$$

erzeugt wird und diese sowohl die Homogenität als auch den Grad invariant lässt, müssen alle Formen F_k bei (11.19) in sich selbst übergehen. Man rechnet leicht nach, dass aus dieser Bedingung

$$\begin{aligned} F_2(x,y) &= \alpha(x^2 + y^2) \quad \text{und} \\ F_3(x,y) &= 3x^2y - y^3 \quad \text{oder} \quad F_3(x,y) = 3xy^2 - x^3 \end{aligned} \quad (11.20)$$

folgt. Die erstere der Beziehungen (11.20) besagt, dass quadratische Formen mit dieser dreizähligen Symmetrie schon die volle Rotationssymmetrie besitzen; dies ist der tiefere Grund dafür, dass es keine Kegelschnitte gibt, welche C_3 -symmetrisch, nicht aber rotationssymmetrisch sind.

Um zu entscheiden, welche der beiden invarianten kubischen Formen für die Gleichung der Kleeblattschlinge in Frage kommt, kann man deren zusätzliche

Spiegelsymmetrie an der y -Achse ausnützen; dies bedeutet, dass in ihrer Gleichung keine ungeraden Potenzen von x vorkommen können, also kann es nur die erste der beiden in (11.20) (2. Zeile) genannten Formen sein.

Aus demselben Grund kann man bei den Formen 4. Grades davon ausgehen, dass sie die Gestalt

$$F_4(x, y) = \alpha x^4 + 2\beta x^2 y^2 + \gamma y^4.$$

haben. Wendet man jetzt das obige Invarianzprinzip an, so folgt

$$\alpha = \beta = \gamma$$

d. h.

$$F_4(x, y) = \alpha(x^2 + y^2)^2.$$

Der Faktor α bestimmt lediglich die Länge der „Flügel“ und kann zu 1 normiert werden. Damit ist die Gleichung (11.17) der Kleeblattschlinge erklärt und zugleich gezeigt, dass sie im wesentlichen die einzige algebraische Kurve 4. Ordnung mit den genannten Symmetrie-Eigenschaften ist.

Stichwortverzeichnis

Symbole

G^0 -Spline	11
G^1 -Spline	11
G^2 -Spline	70
G^k -Spline	12
β -Spline	71
γ -Spline	71

A

Ableitungsformeln für die	
Bézier-Darstellung	68
Ableitungsgleichungen	
von Frenet	19
algebraische Kurven	95, 98
Algorithmus von	
Chaikin	121
de Boor	89
de Casteljau	65

B

B-Spline	83–89
Bézier-	
Darstellung	63

Kurve	68
Punkte	64
Segment	63

begleitendes Bezugssystem	17
Berührordnung	11
Berührung k -ter Ordnung	12
Bernstein-Polynome	63
Bestandteile algebr. Kurven	96
Binormale	33
Bogenlänge	7, 9
de Boor-Algorithmus	89
Brachystochrone	138
Brennpunkt	31

C

Cassinische Ovale	135
de Casteljau-Algorithmus	65
Chaikins Algorithmus	121
Corner Cutting	121

D

Dido	36
Dido'sches Problem	36

Diffeomorphismus 7
 Doppelpunkt einer
 algebraischen Kurve 102
 Drehstreckung 32

E

Eibereich 36, 38
 Eigenschaft, lokale 2
 Eilinie 36
 einfach geschlossen 34
 Euler-Lagrange-Gleichung ... 126
 Evolute 24, 26
 Evolvente 24

F

Fahrstrahl 36
 Ferngerade 51
 Fernpunkt 51
 Flächeninhalt 35
 Frenet'sche
 Ableitungsgleichungen ... 19
 Frenet-Formeln
 für ebene Kurven 17, 19
 für Raumkurven 40

G

geometrischer Spline 9
 geschlossene Kurve 33
 glatt 5
 Glattheit 8
 gleitender Übergang 117

H

Hüllkurve 24
 Helix 41
 Hermite-
 Basisfunktion 46
 Spline 44
 homogene Koordinaten 52
 Hundekurven 130

I

Interpolation
 mit polynom. Kurven 79
 nach Lagrange 79
 isoperimetrische Eigenschaft .. 36
 isoperimetrisches Problem 36

J

Jordan-
 bogen 34
 kurve 33
 Jordanscher Kurvensatz 33

K

Kegelschnitt 30
 Kettenlinie 123
 Kleeblattschlinge 142
 Klothoide 29
 Knoten 11
 Knotenfolge 10
 Konchoiden 131
 konstante Krümmung 21

konvex	37
Konvexe-Hüllen-Eigenschaft ..	64
Konvexitätstreue	75
Krümmung	19, 22
Krümmungs-	
formel	23
kreis	20 f.
mittelpunkt	20
radius	19
krümmungsstetig	117
Kubik	50
Kurve (Definition)	2

L

Leitkurve	24
logarithmische Spirale	31, 132
lokale Eigenschaft	2
Loxodrome	133

M

Marching Algorithmus	112 f.
mehrfache Knoten	92
Mitternachtsformel	110
Multiplizität	100
Muschelkurven	131

N

natürliche Gleichung	29
Normalenvektor	18
numerische Exzentrizität	31
NURBS	93

P

Parameterdarstellung	
einer Kurve	4
Parametertransformation	6
PH-Kurve	47
Polarkoordinaten	30
polynomiale Kurven	43

Q

Quartik	50
Quintik	50

R

rationale	
Kurve	49
PH-Kurve	47
Spline-Kurven	93
Rationalität (Bedingungen) ...	107
reguläre Kurve	5
rektifizierbar	8
Relativtopologie	2
Resultante	103

S

Satz von Bézout	115
Scheitelpunkt	22
Schmiegeebene	39
Schraublinie	41
Segment einer Kurve	9
Singularität (einer	
algebraischen Kurve)	99

Spline 10
 Splinefunktion 10
 Splinekurve 87
 stückweise glatt 5
 Sylvestersche Determinante .. 104

T

Tangente 98
 einer algebr. Kurve 98
 im Sinne der
 Differentialgeometrie 18
 tangenzenstetig 5
 Topologie (des euklidischen
 Raumes) 3
 topologische Abbildung 2 f.
 Traktrix 124, 127
 transversal schneiden 24
 Trennung der Variablen 129

U

Umlaufszahl 34

Unterteilung 118
 Unterteilungsalgorithmus 118
 Unterteilungsschema 119

V

Variationsminderungs-
 eigenschaft 75
 Variationsrechnung 125
 Vier-Punkte-Schema 119
 Vierscheitelsatz 38

W

Wendepunkt 21

Z

Zerfallen einer algebraischen
 Kurve 96
 Zykloide 136
 Zykloidenpendel 136
 Zykloidenverzahnung 137

Literatur

- [1] Wilhelm Blaschke, Kurt Leichtweiß.
Elementare Differentialgeometrie.
Springer: Berlin, 1973.
- [2] Wilhelm Blaschke.
Kreis und Kugel.
de Gruyter: Berlin, 1956.
- [3] Klaus Böhmer.
Spline-Funktionen: Theorie und Anwendungen.
Teubner: Stuttgart, 1974.
- [4] Carl de Boor.
A practical guide to splines.
Applied mathematical sciences; 27. Springer: New York, 1978.
- [5] Carl de Boor.
Cutting corners always works.
Computer Aided Geometric Design 4. 1987. 125-131.
- [6] Heinrich Brauner.
Differentialgeometrie.
Vieweg: Braunschweig, 1981.
- [7] George Merrill Chaikin.
An algorithm for high-speed curve generation.
Computer Graphics and Image Processing, 3. 1974. 346-349.

- [8] Wendelin L. F. Degen.
Berührordnung.
Technischer Bericht am Math. Institut B der Universität Stuttgart.
- [9] Daniel Doo, Malcolm Sabin.
Behaviour of recursive division surfaces near extraordinary points.
Computer Aided Design 10. 1978. 356-360.
- [10] Nira Dyn, John A. Gregory, David Levin.
A 4-point interpolatory subdivision scheme for curve design.
Computer Aided Geometric Design 4. 1987. 257-268.
- [11] Nira Dyn.
Subdivision Schemes in CAGD.
In: Advances in Numerical Analysis II,
Wavelets, Subdivisions and Radial Functions. W. A. Light, Hrsg.,
Oxford University Press, 1992, S. 36–104.
- [12] Gerald Farin, Josef Hoschek, Myung-Soo Kim. (Hrsg.)
Handbook of Computer Aided Geometric Design.
Elsevier: Amsterdam, 2002.
- [13] Rida Farouki.
Pythagorean-Hodograph Curves.
Springer: Berlin, Heidelberg, New York, 2008.
- [14] John A. Gregory, Ruibin Qu.
Nonuniform corner cutting.
Computer Aided Geometric Design 13. 1996. 763-772.
- [15] Hugo Hadwiger.
Vorlesungen über Inhalt, Oberfläche und Isoperimetrie.
Springer: Berlin, 1957.
- [16] Josef Hoschek, Dieter Lasser.
Grundlagen der geometrischen Datenverarbeitung.
Teubner: Stuttgart, 1992.

- [17] Jeffrey M. Lane, Richard F. Riesenfeld.
*A theoretical development for the computer generation
and display of piecewise polynomial surfaces.*
IEEE Trans. on Pattern Analysis
and Machine Intelligence 2. 1980. 35-46.
- [18] Jeffrey M. Lane, Richard F. Riesenfeld.
*A Geometric Proof for the
Variation Diminishing Property of B-Spline Approximation.*
Journal of Approximation Theory 37. 1983. 35-46.
- [19] Kurt Leichtweiß.
Konvexe Mengen.
VEB Deutscher Verlag der Wissenschaften, 1980.
- [20] Heinrich Netz.
Formeln der Mathematik.
Westermann: Braunschweig, 1965.
- [21] Marco Paluszny, Hartmut Prautzsch, Martin Schäfer.
A geometric look at corner cutting.
Computer Aided Geometric Design 14. 1997. 421-447.
- [22] Jörg Peters, Ulrich Reif.
Subdivision surfaces.
Springer: Berlin, 2008.
- [23] Isaac Jacob Schoenberg.
On spline functions (with a supplement by T. N. T. Greville).
In: Inequalities, O. Shisha, Hrsg.
Academic Press: New York, 1967, 255-291, 1967, S. 255-291.
- [24] Robert J. Walker.
Algebraic curves.
Princeton Univ. Press: Princeton, N. J., 1950.

- [25] Rolf Walter.
Einführung in die Analysis.
de Gruyter: Berlin, New York, 2007.
- [26] Walter Wunderlich.
Ebene Kinematik.
Bibliogr. Inst.: Mannheim, 1970.

Kurven wohin das Auge blickt: als Konturlinien, als Umrisse von Buchstaben, in Straßenverläufen, als Wurf- und Bewegungsbahnen, im Design, in Graphiken, Kunstwerken, in der Architektur usw. Vielfach sind sie geplant und am Rechner entworfen; oft visualisieren Kurven Zusammenhänge oder ergeben sich durch physikalische Gesetzmäßigkeiten. So sind Kurven seit langem Gegenstand mathematischer Betrachtungen und bieten mit zunehmenden technischen Möglichkeiten zur Berechnung und Fertigung auch heute immer wieder Anlass zu weiteren Untersuchungen und Entwicklungen.

Dieses Buch führt in die Mathematik der Kurven ein, ohne viele Kenntnisse vorauszusetzen und verzichtet auf tiefer gehende Beweise. Es wendet sich an Studierende nicht nur der Mathematik, sondern auch der Informatik, der Architektur und technischer Fächer. Behandelt werden Grundlagen aus der Differentialgeometrie, der algebraischen Geometrie, sowie der Geometrieverarbeitung (Computer Aided Geometric Design – CAGD). Abgerundet wird das Buch durch ein Kapitel über spezielle Kurven, die historisch oder ihrer Anwendungen wegen von Bedeutung sind.

ISBN 978-3-7315-0662-1



9 783731 506621 >