

Diskrete Geometrie und Visualisierung

Teilnehmer:

Stefanie Cattes	Herder-Oberschule
Benjamin Drost	Heinrich-Hertz-Oberschule
Franz Löchner	Heinrich-Hertz-Oberschule
Ertan Özel	Herder-Oberschule
Julius Seiberl	Georg-Forster-Oberschule
Katja Wolff	Heinrich-Hertz-Oberschule

Gruppenleiter:

Konstantin Poelke, Carsten Lange, Konrad Polthier	Freie Universität Berlin und DFG-Forschungszentrum MATHEON „Mathematik für Schlüsseltechnologien“
---	---

Dieser Kurs soll eine Einführung in Visualisierungstechniken geben. Zunächst werden parametrisierte Kurven betrachtet, mit deren Hilfe viele wichtige Konzepte leicht eingeführt werden können. Im zweiten Abschnitt werden Bezierkurven betrachtet, deren Form sich durch ein Kontrollpolygon leicht bearbeiten, berechnen und parametrisieren lässt. Abschnitt drei behandelt die Parametrisierung von Flächen. Im folgenden vierten Abschnitt wird ein erstaunliches Phänomen beschrieben: Ein Zylinder wird punktweise durch eine triangulierte Fläche approximiert, doch abhängig von der Parameterwahl konvergiert der Flächeninhalt gegen den des Zylinders oder nicht. Zuletzt wird in Abschnitt fünf eine Methode zur Bildverzerrung skizziert, indem wir versuchen zu klären, wie sich effizient Gitter deformieren lassen und welche Abbildungen sich dafür eignen.

Inhaltsverzeichnis

1	Parametrisierte Kurven	91
1.1	Definition	91
1.2	Der (Einheits-)Kreis	91
1.3	Das Unendlichzeichen	92
1.4	Die Sanduhr	93
1.5	Geschwindigkeiten, die erste Ableitung und Tangentialvektoren	93
1.6	Bogenlänge	94
1.7	Die Betragsfunktion	95
1.8	Zykloide	96
1.9	Epizykloide	97
1.10	Hypozykloide	98
2	Bezierkurven	99
2.1	Der de Casteljau-Algorithmus	99
2.2	Bernstein-Polynome	100
3	Parametrisierte Flächen	102
3.1	Kugeloberfläche	102
4	Von parametrisierten Flächen zur Schwarz'schen Laterne	104
4.1	Die Schwarz'sche Laterne	105
4.2	Mantelfläche der Schwarz'schen Laterne	105
5	Bilderverzerrung	107
5.1	Normen	107
5.2	„Energie“-Funktionen	108
5.3	Die Abbildung	108
5.4	Gitternetz	109

1 Parametrisierte Kurven

1.1 Definition

Parametrisierte Kurven laufen in Abhängigkeit eines Parameters t , welcher nicht im Koordinatensystem eingezeichnet ist. Sie entsprechen folgender Definition:

$$f : I \subset \mathbb{R} \rightarrow \mathbb{R}^n, t \mapsto \begin{pmatrix} f_1(t) \\ f_2(t) \\ \vdots \\ f_n(t) \end{pmatrix}, f_i : I \rightarrow \mathbb{R}$$

Ein Beispiel für eine Parametrisierung einer Kurve ist die Parametrisierung von $g(x) = x^2$

Diese Kurve liegt im \mathbb{R}^2 , sodass die beiden Funktionen f_1 und f_2 die x- bzw. y-Koordinate unserer Kurve darstellen.

Die x-Koordinate wird dabei einfach mit $f_1(t) = t$, die y-Koordinate mit $f_2(t) = t^2$ beschrieben.

Die Parameterdarstellung lautet also für $t \in \mathbb{R}$:

$$t \rightarrow \begin{pmatrix} t \\ t^2 \end{pmatrix}.$$

1.2 Der (Einheits-)Kreis

Da der Einheitskreis die Kreisfunktionen Sinus und Cosinus beschreibt, ist es nur logisch, dass Sinus und Cosinus in der entsprechenden parametrisierten Darstellung enthalten sind. Der Cosinus ist die Projektion auf die x-Achse und bestimmt deshalb die x-Koordinate. Er entspricht $f_1(t)$. Der Sinus ist die Projektion auf die y-Achse und bestimmt deshalb die y-Koordinate. Er entspricht $f_2(t)$.

Die Parameterdarstellung eines Kreises ist also folgende:

$$t \rightarrow \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}, t \in [0; 2\pi].$$

Will man noch einen Radius ungleich Eins berücksichtigen, so tritt dieser einfach

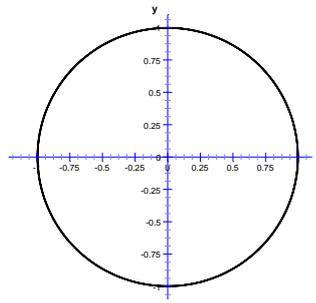


Abbildung 1: Der Einheitskreis

als Vorfaktor auf:

$$t \rightarrow r \cdot \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}, t \in [0; 2\pi].$$

1.3 Das Unendlichzeichen

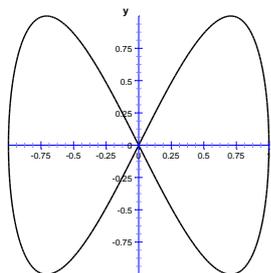


Abbildung 2: Das „Unendlichzeichen“

Hierbei hatten wir zwei Ansätze. Der erste war, das Zeichen einfach durch Aneinanderreihung zweier Kreise zu erhalten. Dazu stellten wir als erstes fest, dass wir, um den Kreis um Eins nach rechts zu verschieben, einfach nur eine Eins zur x-Koordinate addieren mussten, da ja jeder Punkt auf der selben Höhe bleibt und sich nur parallel zur x-Achse verschiebt. Für die Verschiebung nach links muss mit der selben Begründung eine Eins von der x-Koordinate subtrahiert werden. Also:

$$t \rightarrow \begin{pmatrix} \cos(t) + 1 \\ \sin(t) \end{pmatrix}, \text{ für } t \in [-\pi; \pi]$$

$$t \rightarrow \begin{pmatrix} \cos(t) - 1 \\ \sin(t) \end{pmatrix}, \text{ für } t \in [\pi; 3\pi]$$

Die zweite Version gründet sich darauf, dass zum Erhalten der Kurve die y-Koordinate doppelt so schnell laufen muss, wie die x-Koordinate. Während also der Cosinus nur einmal alle Werte zwischen $-\pi$ und π durchläuft muss der Sinus die Werte zwischen $-\pi$ und 3π durchlaufen. Also:

$$t \rightarrow \begin{pmatrix} \cos(t) \\ \sin(2t) \end{pmatrix}$$

1.4 Die Sanduhr

Um nun von der vorangegangenen Figur auf eine Sanduhr zu kommen, muss man die Figur an der Identität spiegeln, was mit der Parameterdarstellung besonders einfach geht. Man muss einfach nur die Komponenten von x- und y-Eintrag vertauschen. Ausgehend von der zweiten Version also:

$$t \rightarrow \begin{pmatrix} \sin(2t) \\ \cos(t) \end{pmatrix}$$

1.5 Geschwindigkeiten, die erste Ableitung und Tangentialvektoren

Da unsere Kurve von einem im Koordinatensystem nicht vorhandenen Parameter t abhängt, können wir diesen auch als die Zeit ansehen, die aufgewendet wird um die Kurve entlang zu gehen. Wenn $t=0$ ist, so läuft man also gerade los, während zum Beispiel der Punkt bei $t=1$ den Ort beschreibt, bei dem man sich zur Zeit 1 befindet. Man kann bestimmte Kurven auch mit unterschiedlichen Geschwindigkeiten entlanglaufen. In Abschnitt 1.3 hat man sehr schön gesehen, wie eine Verdopplung des pro Zeiteinheit zurückgelegten Weges zu einer Stauchung der Kurve führte. Würde man allerdings f_1 und f_2 um den selben Faktor beschleunigen oder verlangsamen, so würde man z.B. den Kreis in der bisherigen Zeit einfach nur mehrmals oder nur teilweise durchlaufen.

Ganz physikalisch gesprochen kann man diese Geschwindigkeit auch als Vektor auffassen. Dieser sogenannte Tangentialvektor würde dann an den entsprechenden Kurvenpunkt angesetzt. Betrachtet man die Kurve des Unendlichzeichens so sieht man, dass am Punkt $(0|0)$ gleich zwei Tangentialvektoren anliegen, da wir diesen Punkt zwei Mal mit unserer Kurve treffen, was nicht an der Periodizität von Sinus und Cosinus liegt, und auch nicht an der Länge des Intervalls.

Die Geschwindigkeit ist außerdem der Ort (in unserem Fall die Parameterdarstellung der Kurve) abgeleitet nach der Zeit (bei uns der Parameter).

Am Beispiel unseres Unendlichzeichens sähe diese Ableitung wie folgt aus:

$$\frac{d}{dt}\gamma(t) = \begin{pmatrix} \frac{d}{dt}\cos(t) \\ \frac{d}{dt}\sin(2t) \end{pmatrix} = \begin{pmatrix} -\sin(t) \\ 2\cos(2t) \end{pmatrix}.$$

Man kann übrigens auch hier äquivalent zu einer klassischen Funktion einen Differenzenquotienten aufstellen. Während es bei Funktionen dabei nämlich darum geht, wie sich der Funktionswert verändert, wenn man den x-Wert minimal variiert, geht es bei Kurven darum, wie sich der Ort ändert, wenn man den Parameter minimal variiert.

Also:

$$\lim_{h \rightarrow 0} \left(\frac{\gamma(t+h) - \gamma(t)}{h} \right)$$

Verallgemeinert lautet der Ausdruck so:

$$\gamma : I \rightarrow \mathbb{R}^2$$

$$t \rightarrow \begin{pmatrix} f_1(t) \\ f_2(t) \end{pmatrix}, \text{ der Ort zum Zeitpunkt } t$$

$$\frac{d}{dt}\gamma : I \rightarrow \mathbb{R}^2$$

$$t \rightarrow \begin{pmatrix} \frac{d}{dt}f_1(t) \\ \frac{d}{dt}f_2(t) \end{pmatrix}, \text{ die Geschwindigkeit zum Zeitpunkt } t$$

Diese Darstellung gibt uns, da es der Vektor der Geschwindigkeit zum Zeitpunkt t ist, sowohl die Richtung (Richtung des Vektors), als auch die Größe (Länge des Vektors) der Geschwindigkeit an. Wollen wir allerdings nur die Geschwindigkeit, also die Länge des Vektors, haben, so können wir diese über den Satz des Pythagoras bestimmen.

$$\left| \frac{d}{dt}\gamma(t) \right| = \sqrt{\left(\frac{d}{dt}f_1(t)\right)^2 + \left(\frac{d}{dt}f_2(t)\right)^2}$$

Natürlich ist dabei auch eine Geschwindigkeit von Null möglich. Zu einem solchen Zeitpunkt bleibt der Punkt sozusagen 'stehen'.

Definition 1.1 Eine Kurve γ heißt regulär, falls $\frac{d}{dt}\gamma \neq 0$ für alle Zeitpunkte t gilt.

1.6 Bogenlänge

Ebenfalls äquivalent zur klassischen Funktion kann man die Bogenlänge einer Kurve bestimmen. Diese ist bestimmt durch das Integral über die Geschwindigkeiten im

Intervall $[a; b]$.

Also

$$\int_b^a \left(\left| \frac{d}{dt} \gamma(t) \right| \right) = \int_b^a \left(\sqrt{\left(\frac{d}{dt} f_1(t) \right)^2 + \left(\frac{d}{dt} f_2(t) \right)^2} \right).$$

Wie man leicht erkennt, führt diese Schreibweise für Parameterdarstellungen einfacher Funktionen (siehe Beispiel x^2) zu der üblichen Bogenlänge, da ja $f_1(t) = x$ und demnach $\frac{d}{dt} f_1(t) = 1$ ist.

1.7 Die Betragsfunktion

Die Betragsfunktion ist bekanntermaßen an der Stelle $x = 0$ nicht differenzierbar, da die Funktion erst mit der konstanten Steigung von -1 fällt, und nach der Null sprunghaft zu einer Steigung von $+1$ wechselt.

In der Parameterdarstellung kann man dieses Problem umgehen. Dort muss die entsprechende Kurve nämlich nur aus zwei gleichen Funktionen f_1 und f_2 bestehen, wobei f_2 den Betrag von f_1 beschreibt. Dadurch können die Funktionen so gewählt werden, dass kein sprunghafter Steigungswechsel stattfindet, sondern man sozusagen vor der Ecke langsamer wird, um danach wieder zu beschleunigen. Eine mögliche Darstellung dafür wäre diese:

$$\gamma(t) = \begin{pmatrix} t^3 \\ |t^3| \end{pmatrix} \quad \text{mit} \quad \frac{d}{dt} \gamma(t) = \begin{pmatrix} 3t^2 \\ \frac{t^3}{|t^3|} \cdot 3t^2 \end{pmatrix}.$$

Zur näheren Erläuterung:

Annäherung von links:

$$|t^3| = -t^3 \text{ für } t < 0 \Rightarrow \begin{pmatrix} 3t^2 \\ \frac{t^3}{|t^3|} \cdot 3t^2 \end{pmatrix} = \begin{pmatrix} 3t^2 \\ \frac{t^3}{-t^3} \cdot 3t^2 \end{pmatrix} = \begin{pmatrix} 3t^2 \\ -3t^2 \end{pmatrix} \xrightarrow{t \rightarrow 0} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Annäherung von rechts:

$$|t^3| = t^3 \text{ für } t > 0 \Rightarrow \begin{pmatrix} 3t^2 \\ \frac{t^3}{|t^3|} \cdot 3t^2 \end{pmatrix} = \begin{pmatrix} 3t^2 \\ \frac{t^3}{t^3} \cdot 3t^2 \end{pmatrix} = \begin{pmatrix} 3t^2 \\ 3t^2 \end{pmatrix} \xrightarrow{t \rightarrow 0} \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Die klassische Betragsfunktion hingegen sieht so aus:

$$f(t) = \begin{pmatrix} t \\ |t| \end{pmatrix} \quad \text{mit} \quad \frac{d}{dt} f(t) = \begin{pmatrix} 1 \\ \frac{t}{|t|} \end{pmatrix}$$

Sie liefert folgende Ergebnisse:

Annäherung von links:

$$|t| = -t \text{ für } t < 0 \Rightarrow \begin{pmatrix} 1 \\ \frac{t}{|t|} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{t}{-t} \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

Annäherung von rechts:

$$|t| = t \text{ für } t < 0 \Rightarrow \begin{pmatrix} 1 \\ \frac{t}{|t|} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{t}{t} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Hier sind die Ergebnisse in beiden Fällen konstant.

Die Potenzen in der ersten Darstellung sorgen dafür, dass die Ableitung gegen Null geht.

Wie man also sieht, muss man sich bei der Arbeit mit Kurven von dem üblichen Gedanken abwenden, differenzierbare Funktionen seien „rund“. Auch eine eckige Kurve wie die des Betrages ist differenzierbar.

1.8 Zykloide

Zykloide sind sogenannte Abrollfiguren. Sie beschreiben die Kurve, die ein Punkt beschreibt, der auf einem Kreis mit Radius r liegt und eine bestimmte Bahn „entlangrollt“.

Dazu versuchten wir zuerst einmal, die Bahn des Punktes auf dem Kreis zu beschreiben, ohne den Kreis dabei auch noch fortzubewegen. Als erstes wurde zum Erreichen der benötigten Rechtsdrehung die folgende Version aufgestellt:

$$r \cdot \begin{pmatrix} \cos(t) \\ -\sin(t) \end{pmatrix}.$$

Sie begründet sich darin, dass der Cosinus zwar seiner alten Bahn folgt, der Sinus jedoch in negative Richtung wandert. Allerdings würde der Kreispunkt dann nicht „unten“, sondern „rechts“ am Kreis beginnen. Um „unten“ zu beginnen, müsste man das ganze um $-\frac{\pi}{2}$ verschieben, was dank der Periodizität zu folgender korrekten Darstellung führt:

$$r \cdot \begin{pmatrix} -\sin(t) \\ -\cos(t) \end{pmatrix}.$$

Um nun noch die Rechtsdrehung einzubinden, müssen wir die Bahn des Mittelpunktes betrachten. Diese wird einfach durch $t \rightarrow \begin{pmatrix} t \\ 1 \end{pmatrix}$ beschrieben und ist eine Parallele zu unserer Geraden.

Durch Addition können wir diese Kräfte nun überlagern und so die Bewegung des Punktes bestimmen:

$$\begin{pmatrix} t \\ 1 \end{pmatrix} + r \cdot \begin{pmatrix} -\sin(t) \\ -\cos(t) \end{pmatrix}.$$

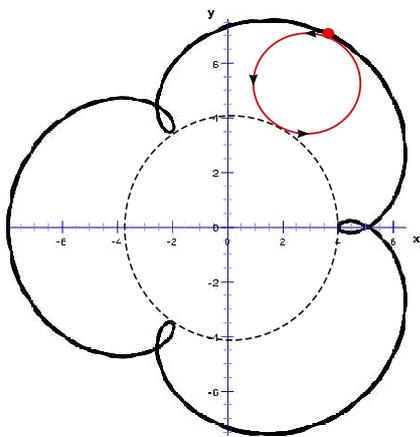
Allerdings stellt sich hierbei die Frage, warum man nicht einfach $5t$ schreiben könnte. Dies ließ sich jedoch einfach damit begründen, dass sich dann der Mittelpunkt im Verhältnis zum Rad viel zu schnell fortbewegen würde. Nichtsdestotrotz fiel auf, dass unser Vorfaktor Eins nicht unbedingt richtig sein müsste, da ja das Rad eine andere Strecke zurücklegt als der Kreismittelpunkt. Während letzterer nämlich zum Beispiel bei $t = 2\pi$ die Strecke 2π zurücklegt, ist der Kreispunkt gerade eine Strecke von $2\pi \cdot r$ weit gekommen, da er ja gerade einmal um den Kreis gewandert ist. Dementsprechend muss die Geschwindigkeit des Mittelpunktes angepasst werden. Dazu bilden wir das Verhältnis aus Umfang des Kreises und Bogenmaß, $\frac{2\pi \cdot r}{2\pi}$. Der benötigte Faktor ist also r .

Die exakte Form lautet also

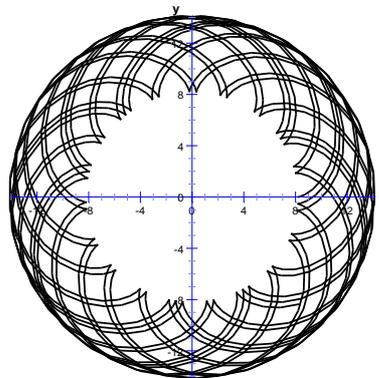
$$\begin{pmatrix} rt \\ r \end{pmatrix} + r \cdot \begin{pmatrix} -\sin(t) \\ -\cos(t) \end{pmatrix} = r \cdot \begin{pmatrix} t - \sin(t) \\ 1 - \cos(t) \end{pmatrix}.$$

1.9 Epizykloide

Eine weitere mögliche Abrollfigur beschreibt die Bahn eines Punktes auf einem Kreis mit Radius r der auf einem größeren Kreis mit Radius R entlang rollt, wobei ($R \geq r$).



(a) Epizykloide(schematisch)



(b) Epizykloide mit irrationaler Phase

Abbildung 3: Epizykloide

Als erstes stellten wir hierzu wieder die Formel für die Bahn des Punktes auf dem kleinen Kreis dar, sollte sich der Kreis nicht fortbewegen. Da wir sozusagen „links“ an unserem Kreis anfangen, müssen wir die „normale“ Form anpassen. Die passende Formel lautet also wie folgt:

$$r \cdot \begin{pmatrix} -\cos(t) \\ -\sin(t) \end{pmatrix}.$$

Die Bahn des Mittelpunktes wird hierbei jedoch nicht einfach als Gerade beschrieben, sondern ist die Bahn eines Kreises mit Radius $(R + r)$. Sie lautet also

$$(R + r) \cdot \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$

Auch hierbei muss eine Anpassung der Geschwindigkeit stattfinden. Und zwar durch das Verhältnis aus Umfang des kleinen Kreises und Mittelpunktsbahn:

$$\frac{2\pi \cdot (R + r)}{2\pi \cdot r} = \frac{R + r}{r} = \frac{R}{r} + 1.$$

Diese Variation muss nun als verändernder Faktor für die Geschwindigkeit der Kreisbahn auftreten. Nach Überlagerung der Kräfte lautet die komplette Formel folgendermaßen:

$$t \rightarrow \begin{pmatrix} (R + r) \cos(t) - r \cdot \cos\left(\left(\frac{R}{r} + 1\right) \cdot t\right) \\ (R + r) \sin(t) - r \cdot \sin\left(\left(\frac{R}{r} + 1\right) \cdot t\right) \end{pmatrix}$$

1.10 Hypozykloide

Eine weitere sehr ähnliche Abrollfigur beschreibt die Bahn eines Punktes auf einem Kreis mit Radius r , der dieses Mal *in* einem größeren Kreis mit Radius R entlang rollt ($R \geq r$).

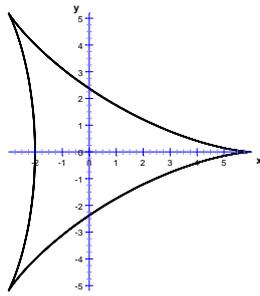
Die Bahn des Punktes auf dem kleinen Kreis lässt sich in diesem Fall durch eine Abwandlung der Formel aus 1.1 da wir zwar rechts starten aber im Uhrzeigersinn drehen.

Damit kommt man auf folgende Formel:

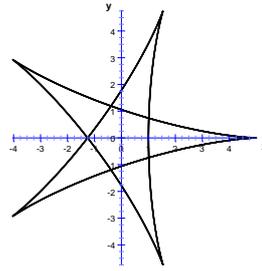
$$r \cdot \begin{pmatrix} \cos(t) \\ -\sin(t) \end{pmatrix}.$$

Die Bahn des Mittelpunktes folgt hierbei analog zu den Epizykloiden einer Kreisbahn, diesmal im inneren des Kreises mit dem Radius $(R - r)$. Die Gleichung lautet also

$$(R - r) \cdot \begin{pmatrix} \cos(t) \\ \sin(t) \end{pmatrix}.$$



(a) $R=6, r=2$



(b) $R=5, r=2$

Abbildung 4: Hypozykloide mit unterschiedlichen Radien-Verhältnissen

Wieder muss eine Anpassung der Geschwindigkeit stattfinden. Der Faktor hierbei lässt sich wie bei den Epizykloiden durch das Verhältnis aus Umfang des kleinen Kreises und Mittelpunktsbahn berechnen:

$$\frac{2\pi \cdot (R - r)}{2\pi \cdot (r)} = \frac{R - r}{r} = \frac{R}{r} - 1.$$

Die komplette Formel lautet nun:

$$t \rightarrow \begin{pmatrix} (R - r) \cdot \cos(t) - r \cdot \cos\left(\left(\frac{R}{r} - 1\right) \cdot t\right) \\ (R - r) \cdot \sin(t) - r \cdot \sin\left(\left(\frac{R}{r} - 1\right) \cdot t\right) \end{pmatrix}.$$

2 Bezierkurven

Da man für die Bildbearbeitung die vorangegangenen, komplizierten Kurven nur schlecht benutzen kann, benötigt man einfachere Kurven, die sich kontrolliert verändern lassen. Bezierkurven haben sich dabei als besonders nützlich erwiesen. Durch bestimmte Stütz- bzw. Fixpunkte können sie auf zwei verschiedenen Arten erzeugt werden, die im Folgenden erläutert werden.

2.1 Der de Casteljau-Algorithmus

Das erste Verfahren arbeitet mit dem Algorithmus von de Casteljau und ist konstruktiv. Bei diesem Algorithmus sind $n + 1$ Punkte P_0, \dots, P_n (die wir die Punkte der 0. Generation nennen) und ein $z \in [0; 1]$ gegeben. z gibt dabei ein Verhältnis an, in dem die Strecken $\overline{P_0P_1}, \overline{P_1P_2}, \dots, \overline{P_{n-1}P_n}$ jeweils später geteilt werden sollen. Bei

dieser Teilung erhält man nun $n - 1$ neue Punkte (der 1. Generation). Auf diese Punkte wendet man das Verfahren erneut an und erzeugt so Punkte der 2., 3., ..., n -ten Generation, wobei zuletzt nur noch ein einziger Punkt übrig bleibt. Das ganze lässt sich mit folgendem Pseudocode beschreiben:

```

1  For  $i = 0, \dots, n$ 
2     $P_i^0 := P_i$ 
3  For  $j = 1, \dots, n$ 
4    For  $i = 0, \dots, (n - j)$ 
5       $P_i^j := (1 - z) \cdot P_i^{j-1} + z \cdot P_{i+1}^{j-1}$ 
6  Return  $P_0^n$ 

```

In Zeile 1 und 2 werden die Ursprungspunkte formgerecht umbenannt. In der Zeile 3 zählt man mit dem Laufindex j die jeweiligen Generationen. Da in jeder Generation ein Punkt weniger vorhanden ist, gibt es dementsprechend genau n Generationen. j läuft daher von 1 bis n . In Zeile 4 läuft der Index i von 0 bis $n - j$, was die jeweiligen Anzahlen an Punkten in der Generation j sind. Die wichtigste Zeile ist die fünfte: Jeder neue Punkt P_i^j entsteht auf der Strecke zwischen zwei Punkten P_i^{j-1} und P_{i+1}^{j-1} der vorangegangenen Generation. Man erzeugt ihn, indem man die Differenz $P_i^{j-1} - P_{i+1}^{j-1}$ (also die Strecke zwischen beiden Punkten), multipliziert mit z , zu dem Punkt P_i^{j-1} addiert:

$$P_i^j = P_i^{j-1} + z \cdot (P_i^{j-1} - P_{i+1}^{j-1}) = (1 - z) \cdot P_i^{j-1} + z \cdot P_{i+1}^{j-1}.$$

In der 6. Zeile wird dann schließlich der errechnete letzte Punkt ausgegeben. Für jedes z , das man zu den gegebenen Punkten wählt, entsteht ein anderer Punkt auf der Kurve. So lässt sich die Kurve letztendlich konstruieren.

2.2 Bernstein-Polynome

Mit dem zweiten Verfahren lässt sich die Bezierkurve berechnen.

Zunächst betrachten wir das Problem für vier Punkte P_0, P_1, P_2, P_3 für $z \in [0; 1]$. Wir errechnen die Punkte der ersten Generation wie unter 2.1 angegeben.

$$\begin{aligned} P_0^1 &= (1 - z) \cdot P_0 + z \cdot P_1 \\ P_1^1 &= (1 - z) \cdot P_1 + z \cdot P_2 \\ P_2^1 &= (1 - z) \cdot P_2 + z \cdot P_3 \end{aligned}$$

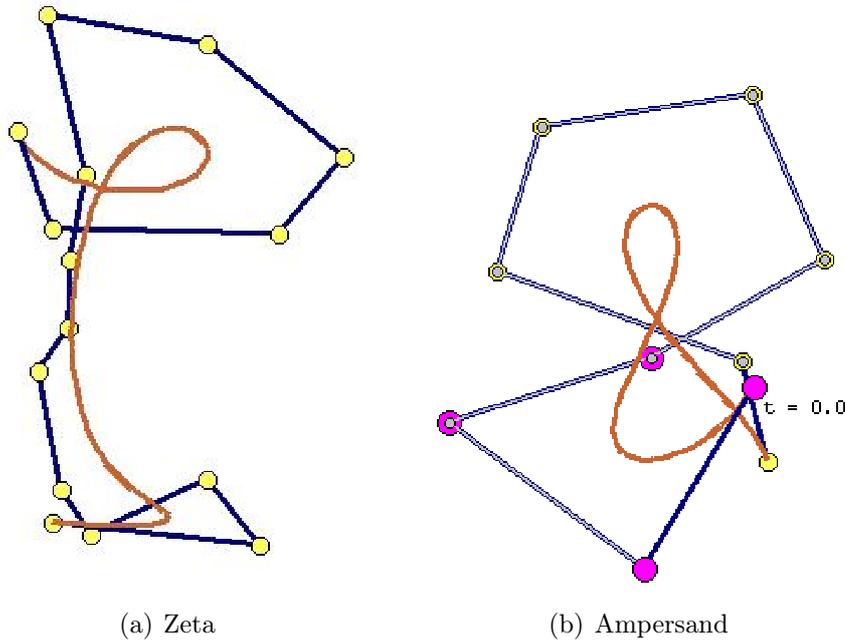


Abbildung 5: Design mit Bezierkurven

Aus denen lassen sich wiederum die Punkte der zweiten Generation errechnen:

$$\begin{aligned}
 P_0^2 &= (1 - z) \cdot P_0^1 + z \cdot P_1^1 \\
 &= (1 - z)((1 - z) \cdot P_0 + z \cdot P_1) + z((1 - z) \cdot P_0 + z \cdot P_1) \\
 &= (1 - z)^2 \cdot P_0 + 2(1 - z)z \cdot P_1 + z^2 \cdot P_2 \\
 \text{Analog } P_1^2 &= (1 - z)^2 \cdot P_1 + 2(1 - z)z \cdot P_2 + z^2 \cdot P_3
 \end{aligned}$$

Nun wird der letzte Punkt aus diesen errechnet:

$$\begin{aligned}
 P_0^3 &= (1 - z) \cdot P_0^2 + z \cdot P_1^2 \\
 &= (1 - z)((1 - z)^2 \cdot P_0 + 2(1 - z)z \cdot P_1 + z^2 \cdot P_2) + z((1 - z)^2 \cdot P_1 + 2(1 - z)z \cdot P_2 + z^2 \cdot P_3) \\
 &= (1 - z)^3 \cdot P_0 + 3(1 - z)^2 z \cdot P_1 + 3(1 - z)z^2 \cdot P_2 + z^3 \cdot P_3 \\
 &= \sum_0^3 \binom{3}{i} (1 - z)^{3-i} z^i \cdot P_i
 \end{aligned}$$

Dies lässt sich nun auf eine beliebige Anzahl von Punkten verallgemeinern:

$$\sum_{i=0}^n \binom{n}{i} (1 - z)^{n-i} z^i \cdot P_i$$

Dabei nennt man $\binom{n}{i}(1-z)^{n-i}z^i$ das i -te Bernsteinpolynom vom Grad n .

3 Parametrisierte Flächen

Neben den Kurven kann man auch Flächen parametrisieren und im Raum darstellen. Hierfür ist es erforderlich, einen zweiten Parameter einzuführen.

Der Definitionsbereich ist diesmal z.B. ein (kartesisches) Produkt zweier Intervalle, also eine Fläche.

3.1 Kugeloberfläche

Wir versuchen nun, eine Kugeloberfläche zu parametrisieren. Unser erster Versuch führte zu

$$(t, u) \mapsto \begin{pmatrix} \cos(t) \\ \sin(t) \\ \sin(u) \end{pmatrix}$$

mit $t \in [0, 2\pi]$ und $u \in [-\pi/2, \pi/2]$. Das Resultat ist jedoch eine Zylinderoberfläche.

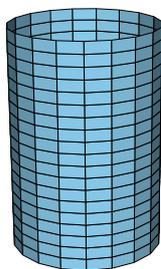


Abbildung 6: Parametrisierter Zylinder

Dies liegt daran, dass der Radius des Kreises, welcher auf der x-y-Ebene liegt, bei uns fest ist, bei der Kugel jedoch natürlich abhängig vom 2. Parameter variiert.

$$(t, u) \rightarrow \begin{pmatrix} \cos(t) \cdot \cos(u) \\ \sin(t) \cdot \cos(u) \\ \sin(u) \end{pmatrix} \text{ mit } t \in [0; 2\pi] \text{ und } u \in [\pi/2, \pi/2]$$

Mithilfe einer Dreh-Matrix sind wir auf eine 2. Möglichkeit gekommen, die Kugeloberfläche darzustellen.

Dabei drehen wir einen in der x-z-Ebene liegenden Halbkreis um die z-Achse.

Die Änderung der kanonischen Einheitsvektoren ist im Folgenden beschrieben:

$$e_1 \mapsto \tilde{e}_1 := \begin{pmatrix} \cos(v) \\ \sin(v) \\ 0 \end{pmatrix}, e_2 \mapsto \tilde{e}_2 := \begin{pmatrix} -\sin(v) \\ \cos(v) \\ 0 \end{pmatrix}, e_3 \mapsto \tilde{e}_3 := \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Die Drehmatrix ist dann: $\mu(\tilde{e}_1, \tilde{e}_2, \tilde{e}_3)$

$$\text{Halbkreis: } G_H = \begin{pmatrix} \cos(u) \\ 0 \\ \sin(u) \end{pmatrix}$$

Nun muss der Graph, welcher den Halbkreis auf der x-z-Ebene darstellt, rotieren bzw. mit der Drehmatrix multipliziert werden.

$$\begin{aligned} \mu \cdot G_H &= \begin{pmatrix} \cos(v) & -\sin(v) & 0 \\ \sin(v) & \cos(v) & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \cos(u) \\ 0 \\ \sin(u) \end{pmatrix} \\ &= \begin{pmatrix} \cos(v) \cdot \cos(u) - \sin(v) \cdot 0 + 0 \cdot \sin(u) \\ \sin(v) \cdot \cos(u) + \cos(v) \cdot 0 + 0 \cdot \sin(u) \\ 0 \cdot \cos(u) + 0 \cdot 0 + 1 \cdot \sin(u) \end{pmatrix} = \begin{pmatrix} \cos(u) \cdot \cos(v) \\ \cos(u) \cdot \sin(v) \\ \sin(u) \end{pmatrix}. \end{aligned}$$

$$\text{Die Kugeloberfläche ist also } \begin{pmatrix} \cos(u) \cdot \cos(v) \\ \cos(u) \cdot \sin(v) \\ \sin(u) \end{pmatrix}.$$

Ableiten kann man die Koordinatenfunktionen der Fläche, indem man partiell ableitet, d.h. ein Parameter wird als Konstante betrachtet und nach dem anderen Parameter wird abgeleitet:

Definition 3.1 Für eine Funktion $g : I \times J \subset \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$

$(u, v) \mapsto g(u, v)$ bezeichnet man mit

$$\frac{\partial}{\partial u} g(u, v) := \lim_{h \rightarrow 0} \frac{g(u+h, v) - g(u, v)}{h} \text{ bzw.}$$

$$\frac{\partial}{\partial v} g(u, v) := \lim_{h \rightarrow 0} \frac{g(u, v+h) - g(u, v)}{h}$$

die partielle Ableitung in u- bzw. v-Richtung, sofern der Limes existiert.

Probleme können auftreten, wenn die Fläche einen „spitzen Gipfel“ hat. Die Ableitung an einem bestimmten Punkt spannt unter gewissen Voraussetzungen eine Ebene

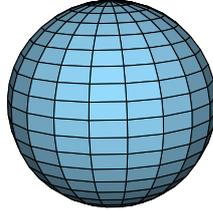


Abbildung 7: Kugeloberfläche

auf, welche durch die Bilder der Einheitsvektoren zu jedem Punkt (u, v) beschrieben werden kann. Dazu wurde das sogenannte „Totale Differential“ eingeführt. Es hat folgende Form:

$$Df := \begin{pmatrix} \frac{\partial}{\partial u} f_1 & \frac{\partial}{\partial v} f_1 \\ \frac{\partial}{\partial u} f_2 & \frac{\partial}{\partial v} f_2 \\ \frac{\partial}{\partial u} f_3 & \frac{\partial}{\partial v} f_3 \end{pmatrix}.$$

Die bekannten partiellen Ableitungen ergeben sich durch Multiplikation mit den Einheitsvektoren $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$:

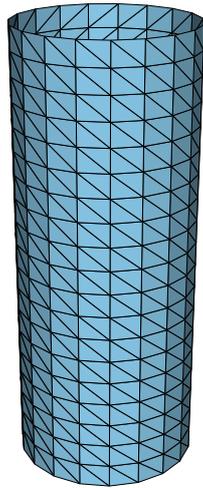
$$\frac{\partial}{\partial u} f|_{(u,v)} = Df|_{(u,v)} \cdot e_1$$

und

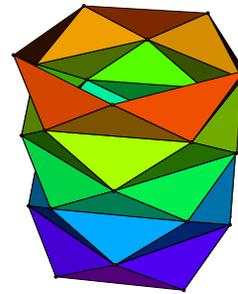
$$\frac{\partial}{\partial v} f|_{(u,v)} = Df|_{(u,v)} \cdot e_2.$$

4 Von parametrisierten Flächen zur Schwarz'schen Laterne

Um eine parametrisierte Fläche zu diskretisieren (damit sie etwa mit dem Computer visualisiert werden kann), wird die Fläche in der Regel trianguliert, d.h. durch Dreiecke angenähert. Im Folgenden wird die Triangulierung an einem Kreiszyylinder verdeutlicht. Der Zylinder wird im ersten Schritt in k gleich hohe Scheiben mit Höhe H/k zerlegt, wobei H die Gesamthöhe angibt. Die Kreisfläche jeder Scheibe wird



(a) diskretisierter Zylinder



(b) Schwarz' Laterne

Abbildung 8: Diskretisierung von Flächen durch Triangulierung

dann mit einem regelmäßigen n -Eck angenähert. Schließlich werden die Eckpunkte der n -Ecke mit Dreiecken verbunden.

4.1 Die Schwarz'sche Laterne

Die Schwarz'sche Laterne, benannt nach dem Mathematiker Schwarz, geht von einer weiteren Methode zur Annäherung eines Zylinders aus und zeigt dabei ein verblüffendes Resultat. Der Unterschied zur ersten Vorgehensweise besteht in der Anordnung der n -Ecke untereinander: Die regelmäßigen Vielecke stehen diesmal verdreht zueinander.

4.2 Mantelfläche der Schwarz'schen Laterne

Die Besonderheit der Figur entsteht bei Annäherung der diskreten Dreiecke zu einem runden Kreiszyylinder, welche auf zwei Wege erreicht werden kann. Zum einen kann die Anzahl k der Scheiben gegen unendlich divergieren, andererseits kann auch die Eckenzahl jedes n -Ecks divergieren.

Gegeben:

- k Anzahl der Scheiben
- n Anzahl der Ecken jedes Vielecks

- H Höhe des Zylinders

In jeder der k Scheiben befinden sich $2n$ kongruente Dreiecke; der Flächeninhalt jedes Dreiecks lässt sich mit $\frac{1}{2} \cdot$ Grundseite $g \cdot$ Höhe h berechnen. Die Grundseite entspricht der Länge einer Seite eines regelmäßigen n -Ecks, ausgedrückt als:

$$g = 2r \sin(\pi/n).$$

Der Satz des Pythagoras liefert nun die Höhe h des Dreiecks. Während das eine Vieleck mit einer Ecke die Kreislinie auf dem Zylinder berührt, beschreibt die darüber liegende Kante eine Sehne des darüberliegenden Bogens. Der Abstand b dieser Sehne zur Kreislinie ergibt sich zu:

$$b = r(1 - \cos(\pi/n)).$$

Somit gilt für die Höhe des Dreiecks:

$$h = \sqrt{b^2 + (H/k)^2} = \sqrt{r^2(1 - \cos(\pi/n))^2 + (H/k)^2}.$$

Man berechnet die gesamte Mantelfläche zu:

$$A_{Zylinder} = 2nk \cdot r \sin(\pi/n) \cdot \sqrt{r^2(1 - \cos(\pi/n))^2 + (H/k)^2}.$$

Für $k \rightarrow \infty$ läuft die Mantelfläche gegen ∞ , da der Term im Wurzelargument gegen einen festen Wert konvergiert, während der Faktor vor der Wurzel gegen ∞ divergiert.

Im Gegensatz dazu konvergiert die Mantelfläche $A_{Zylinder}$ gegen den Flächeninhalt des angenäherten Zylinders. Mit der Regel von de L'Hospital gilt:

$$\begin{aligned} \lim_{n \rightarrow \infty} n \cdot \sin(\pi/n) &= \pi \\ \lim_{n \rightarrow \infty} r^2(1 - \cos(\pi/n))^2 + (H/k)^2 &= (H/k)^2 \end{aligned}$$

Sogar für den Fall, dass beide Variablen gegen ∞ laufen, wobei k langsamer gegen ∞ strebt als n^2 , in O - Notation geschrieben als $k \in o(n^2)$.

Beweis:

Man definiert sich die Abhängigkeit von k zu n mit $k(n)$ und schätzt das Wachstum von dieser Funktion ab. Da

$$\begin{aligned} &2rnk(n) \sin(\pi/n) \sqrt{r^2(1 - \cos(\pi/n))^2 + \left(\frac{H}{k(n)}\right)^2} \\ &= 2rn \sin(\pi/n) \sqrt{r^2k(n)^2(1 - \cos(\pi/n))^2 + H^2} \end{aligned}$$

gilt, folgt:

$$\lim_{n \rightarrow \infty} k(n)(1 - \cos(\pi/n)) = \lim_{n \rightarrow \infty} k(n) \left(- \sum_{i=1}^{\infty} \frac{(-1)^i}{(2i)!} \left(\frac{\pi}{n}\right)^{2i} \right) \leq \lim_{n \rightarrow \infty} k(n) \frac{\pi^2}{2n^2}$$

Der letzte Grenzwert ist genau dann Null, wenn $k(n) \in o(n^2)$. Damit gilt

$$\lim_{n \rightarrow \infty} \sqrt{r^2 k(n)^2 (1 - \cos(\pi/n))^2 + H^2} = H$$

und damit konvergiert

$$\lim_{n \rightarrow \infty} 2rnk(n) \sin(\pi/n) \sqrt{r^2 (1 - \cos(H/k))^2 + (H/k(n))^2} = 2\pi r H.$$

5 Bilderverzerrung

Zu einer der Anwendungen im Bereich der diskreten Geometrie gehört die Bildverarbeitung. Hier im speziellen betrachten wir die Verzerrung von dergleichen. Bilder bestehen je nach Auflösung aus einer bestimmten Anzahl an Bildpunkten (Pixel). Wir wollen mit unserer Methode einige dieser Punkte zu Kontrollpunkten bestimmen. Diese können zum Beispiel per Maus verschoben werden. Die anderen Bildpunkte rundherum sollen daraufhin "glatt" mit verzerrt werden. Die anderen unbewegten Kontrollpunkte jedoch sollen so beibehalten werden, wie sie sind.

5.1 Normen

Für diese Methode müssen wir uns vorerst klar darüber werden, wie wir Längen messen oder Abstände definieren. Als Beispiel sei hier der Abstand über den Pythagoras erwähnt

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}.$$

Wir erwarten von den Längenmessungsarten folgende Eigenschaften:

- (A) $Abstand(P; Q) = 0 \Rightarrow P = Q$
- (B) $Abstand(P; Q) \leq Abstand(P; R) + Abstand(R; Q)$
- (C) Skalierbarkeit: $Abstand(sP; sQ) = |s| * Abstand(P; Q)$,

außerdem sollten die Längen stets nicht negativ sein.

Das gleiche können wir auch für den Vektor zwischen den Punkten formulieren.

Eine Längenmessung, die diese Eigenschaften erfüllt, nennen wir Norm. Neben dem oben benannten euklidischen Abstand sei an dieser Stelle noch eine weitere Norm vorgestellt:

Die Manhattan-Norm

Bezogen auf die rechtwinklige Siedlungsstruktur des Ortsteils von New York ist der Abstand zweier Punkte nicht unbedingt die direkte Streckenlänge zwischen ihnen (wie beim euklidischen), sondern die Summe aus x-Abstand und y-Abstand:

$$\left\| \begin{pmatrix} P_x \\ P_y \end{pmatrix} - \begin{pmatrix} Q_x \\ Q_y \end{pmatrix} \right\|_M = |P_x - Q_x| + |P_y - Q_y|.$$

5.2 „Energie“-Funktionen

Für unsere "Glattheit" brauchen wir nun eine bestmögliche Approximation (Annäherung). Hierzu verwenden wir sogenannte Energiefunktionen $E(x) = \|x - z\|$, deren Begriff aus der Physik vielleicht bekannt ist. Hiervon versuchen wir das Minimum zu erreichen. z ist hierbei das Objekt, z.B. ein Punkt, zu dem der kürzeste Normabstand gesucht wird. Das merken wir uns am besten erst einmal.

5.3 Die Abbildung

Während wir die Kontrollpunkte durch eine affine Abbildung (also Spiegelung, Rotation und Translation) verschieben, müssen wir auch für jeden anderen Punkt eine Abbildung finden, die dann diesen Punkt "glatt" mitverzerrt. Wir nennen diese einfach mal α_x für jeden der Punkte. Diese sollte auch affin sein, denn diese sind ganz nett durch Matrizen und Vektoren beschreibbar:

$$\begin{pmatrix} U_x \\ U_y \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} * \begin{pmatrix} V_x \\ V_y \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix},$$

wobei U der neue Bildpunkt und V der vergangene, die abcd-Matrix für Rotation und Spiegelung und das additive t für die Verschiebung verantwortlich ist.

Um nun eine schön "glatte" Abbildung zu finden, suchen wir das Minimum folgender Energiefunktion:

$$E(\alpha_x) = \sum_{i=1}^n G * \|\alpha_x(P_i) - Q_i\|,$$

mit α_x auf P_i (die Ausgangskontrollpunkte) angewendet, Q_i als die wirklich abgebildeten Kontrollpunkte und G als Gewicht für die Beibehaltung unveränderter Kontrollpunkte. Für G gibt es folgende Möglichkeit:

$$G = \frac{1}{\|P_i - x\|^c} \text{ mit } c \geq 1$$

Sollte x nun ein P_i sein, wird G unendlich groß. Zur Minimierung der Energiefunktion muss also $\alpha_{P_i}(P_i) - Q_i = 0$ sein. Das heißt α_{P_i} bildet P_i wieder auf Q_i ab, was wir erreichen wollten.

Wir können nun ein Minimum der Energiefunktion ermitteln. Dies ist auf Dauer beziehungsweise bei komplizierten Normen kein Zuckerschlecken, aber es funktioniert. Wir können allerdings etwas Beschleunigung in unser System bringen.

5.4 Gitternetz

Wir nutzen die Möglichkeit eines Gitternetzes, um weniger Abbildungen aus Energiefunktionen herleiten zu müssen. Im Grunde suchen wir jetzt nur noch das Minimum der Abbildungsenergiefunktionen für die Gitternetzpunkte. Alle Punkte auf den Gitternetzlinien werden durch die zwei jeweiligen Eckpunkte der Strecke interpoliert. Die restlichen Punkte können ebenso ihre Abbildung durch die Interpolation mit entweder den vertikalen oder horizontalen Gitternetzlinienpunkten bekommen.

Die Interpolation läuft dann wie folgt:

$$\begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix} \text{ und } \begin{pmatrix} a_2 & b_2 \\ c_2 & d_2 \end{pmatrix}$$

für die zwei Punkte mit bereits berechneter Abbildung, dann folgt für den Punkt, der im Verhältnis $t \in [0; 1]$ dazwischen liegt:

$$\begin{pmatrix} t * a_1 + (1 - t) * a_2 & t * b_1 + (1 - t) * b_2 \\ t * c_1 + (1 - t) * c_2 & t * d_1 + (1 - t) * d_2 \end{pmatrix}.$$

Diese Interpolation ist natürlich schneller, allerdings auch nicht so genau, aber wer braucht bei Approximation schon Genauigkeit? :). Umso feiner das Gitter ist, umso genauer wird die "glatte" Verzerrung, umso höher ist allerdings auch die benötigte Rechenleistung.

Als Anwendung der diskreten Geometrie ist dies weniger ein Visualisierungsversuch, sondern ein schönes Beispiel zur Bearbeitung beziehungsweise zur Manipulation von Bildern und macht Lust auf mehr...

6 Quellen

S. Schaefer, T. McPhail, J. Warren, *Image deformation Using Moving Least Squares*, ACM SIGGRAPH 2006, 533–540.

Geometrie der Brüche

Teilnehmer:

Ssohrab Borhanian	Heinrich-Hertz-Oberschule
Sebastian Biegel	Georg-Forster-Oberschule
Georg Hieronimus	Herder-Oberschule
Julius Mielke	Herder-Oberschule
Thu Hien Nguyen	Heinrich-Hertz-Oberschule
Joscha Podlesny	Herder-Oberschule

Gruppenleiter:

Heino Hellwig	Humboldt-Universität zu Berlin, Mitglied im DFG-Forschungszentrum MATHEON „Mathematik für Schlüsseltechnologien“
---------------	--

Unsere Gruppe untersuchte die rationalen Zahlen \mathbb{Q} auf geometrische Weise. Es mag überraschen sich eine Woche lang mit Brüchen zu beschäftigen, von denen man meint, alles in der Sekundarstufe I bereits gelernt zu haben. Wie der Bericht jedoch zeigt, ermöglicht eine geometrische Deutung der Brüche ein tieferes Verständnis ihrer Eigenschaften und einen einfachen, anschaulichen Zugang zur elementaren Zahlentheorie. Anhand der Originalarbeiten von Pick und Ford studierten wir das Wechselspiel zwischen geometrischer und arithmetischer Interpretation der Brüche. Die Ergebnisse ermöglichten uns u.a. lineare diophantische Gleichungen zu lösen und die Berechnung der Kettenbrüche geometrisch zu verstehen.

Inhaltsverzeichnis

1	Gitter und der Satz von Pick	19
2	Farey-Folgen	22
3	Lineare diophantische Gleichungen	27
4	Kettenbrüche	28
5	Ford-Kreise	32

1 Gitter und der Satz von Pick

Der Satz von Pick und die Theorie der Gitter sind einfache aber nützliche Hilfsmittel um Brüche und deren Eigenschaften geeignet zu visualisieren und geometrisch darzustellen.

Definition:

Ein **Gitter** wird durch die äquidistante Parallelverschiebung in Richtung zweier Geraden, die sich in einem Punkt schneiden, um zwei feste, nicht notwendigerweise gleiche, diskrete Werte erzeugt. Diese parallelen Geraden bezeichnet man als Hauptgitterstrahlen, deren Schnittpunkte wir Gitterpunkte nennen. Gitterpolygone sind Polygone, deren Ecken auf Gitterpunkten liegen. Wir definieren als **Flächengrundeinheit** A_G des Gitters die Fläche des Parallelogramms, welches durch den Ursprung und seine Verbindung mit den nächsten Gitterpunkten auf den erzeugenden Geraden aufgespannt wird.

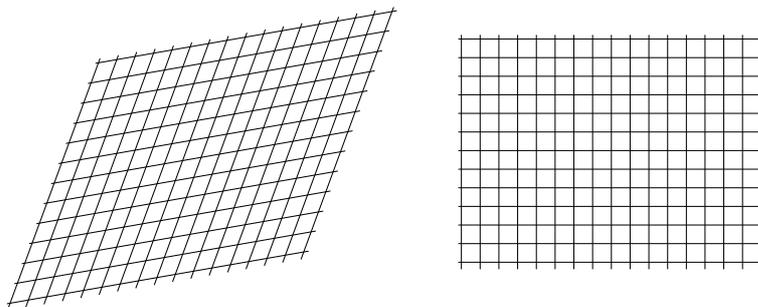


Abbildung 1: Beispiele zweier Gitter.

Satz:

Die Fläche eines (nicht notwendigerweise konvexen) Polygons $P \subseteq \mathbb{R}^2$ mit Gitterpunkten als Ecken ist durch

$$A(P) = i + \frac{u}{2} - 1$$

gegeben, wobei i die Anzahl der inneren Gitterpunkte und u die Anzahl der Randpunkte ist.

Beweis:

Für ein Gitterpolygon P definieren wir die Gitterpunktzahl des Polygons als:

$$G(P) = i + \frac{u}{2} - 1$$

Wir zerschneiden das Polygon durch das Verbinden zweier Randpunkte, sodass zwei neue Polygone entstehen und definieren δ als Anzahl der Gitterpunkte, welche zwischen den beiden Randpunkten auf der Verbindungsgeraden liegen. Wir zeigen nun,

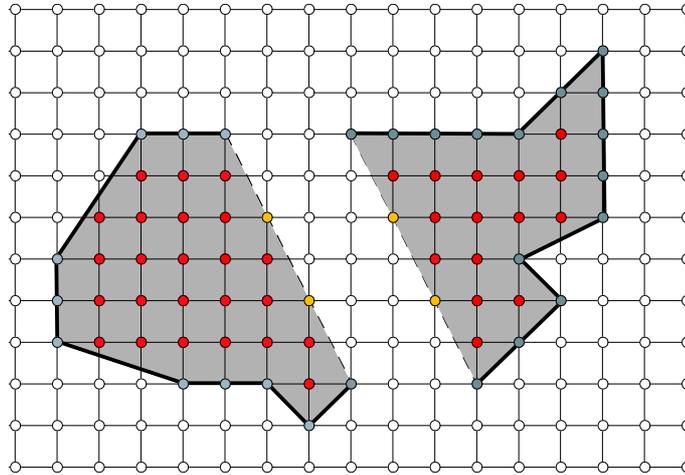


Abbildung 2: Die Zerlegung eines Gitterpolygons P in zwei Teilpolygone P_1 und P_2 .

dass sich die Gitterpunktzahl additiv verhält, d.h. das gilt:

$$G(P) = G(P_1) + G(P_2)$$

Die Gitterpunktzahl der beiden Teilpolygone ist:

$$G(P_1) = i_1 + \frac{u_1}{2} - 1$$

$$G(P_2) = i_2 + \frac{u_2}{2} - 1$$

Weiter berechnet sich:

$$i = i_1 + i_2 + \delta$$

$$u = u_1 + u_2 - 2 \cdot \delta - 2$$

Also gilt:

$$\begin{aligned}
 G(P) &= i + \frac{u}{2} - 1 \\
 &= i_1 + i_2 + \delta + \frac{u_1 + u_2}{2} - \delta - 2 \\
 &= i_1 + \frac{u_1}{2} - 1 + i_2 + \frac{u_2}{2} - 1 \\
 &= G(P_1) + G(P_2)
 \end{aligned}$$

Den Flächeninhalt eines Hauptgitterpolygons P_H , dessen Seiten parallel zu den Hauptgitterstrahlen sind, können wir nun über ein ganzzahliges Vielfaches der Flächeneinheit ausdrücken. Für die Gitterpunktzahl G_G der Flächeneinheit mit $i=0$ und $u=4$ folgt:

$$G_G = i + \frac{u}{2} - 1 = 1 = A_G$$

Demnach folgt für den Flächeninhalt eines Hauptgitterpolygons P_H :

$$A(P_H) = G(P_H) = n \cdot A_G = n ; n \in \mathbb{N}$$

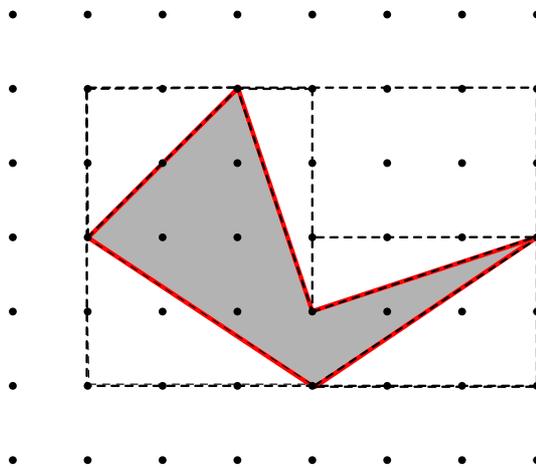


Abbildung 3: Die Fläche eines beliebigen Gitterpolygons wurde mittels Dreiecke so ergänzt, dass die Flächenberechnung auf die Flächenberechnung von Hauptgitterpolygonen hinausläuft.

Nun ist die Fläche eines beliebigen Gitterpolygons so zu ergänzen, dass die Figur nur von Hauptgitterstrahlen begrenzt wird, indem man über allen Seiten die nicht

schon Hauptgitterstrahlen sind, Dreiecke konstruiert, deren andere Seiten Hauptgitterstrahlen sind. Diese Seiten, die nicht Hauptgitterstrahlen sind, können somit als Diagonalen von Parallelogrammen, deren Flächeninhalt dem Doppelten des Flächeninhalts des ergänzenden Dreiecks entspricht, aufgefasst werden. Daraus folgt, dass die Gitterpunktzahl der durch Dreiecke erweiterten Figur gleich dessen Flächeninhalt und somit auch für das ursprüngliche Gitterpolygon die Gitterpunktzahl gleich dem Flächeninhalt ist.

Der Satz von Pick wird im folgenden dazu verwendet, die Haupteigenschaft einer fundamentalen Folge von Brüchen, der Farey-Folge, zu beweisen.

2 Farey-Folgen

Der Geologe J. Farey schrieb im Jahr 1816 an das *Philosophical Magazine* einen Brief, in dem er von einer erstaunlichen Eigenschaft der Brüche berichtete [1]. Er war im Laufe seiner Arbeiten auf ein mathematisches Phänomen gestoßen, dass er sich nicht erklären konnte:

If all the possible vulgar fractions of different values, whose greatest denominator (when in their lowest terms) does not exceed any given number, be arranged in the order of their values, or quotients; then if both the numerator and the denominator of any fraction therein, be added to the numerator and the denominator, respectively, of the fraction next but one to it (on either side), the sums will give the fraction next to it; although, perhaps, not in its lowest terms. For example, if 5 be the greatest denominator given; then are all the possible fractions, when arranged, $1/5, 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4$, and $4/5$; taking $1/3$, as the given fraction, we have $(1+1)/(5+3) = 2/8 = 1/4$ the next smaller fraction than $1/3$; or $(1+1)/(3+2) = 2/5$, the next larger fraction to $1/3$. Again, if 99 be the largest denominator, then, in a part of the arranged Table, we should have $15/52, 28/97, 13/45, 24/83, 11/38$, etc.; and if the third of these fractions be given, we have $(15+13)/(52+45) = 28/97$ the second: or $(13+11)/(45+38) = 24/83$ the fourth of them: and so in all the other cases. I am not acquainted, whether this curious property of vulgar fractions has been before pointed out; or whether it may admit of any easy or general demonstration; which are points on which I should be glad to learn the sentiments of some of your mathematical readers; and am Sir, Your obedient humble servant, J. Farey.

Er erhoffte sich, dass ein Leser des Magazins seine Vermutungen beweisen könnte. Cauchy fand in der Tat bald einen Beweis und benannte die Folgen nach Farey. Im Folgenden werden wir nun auf die Definition und die Haupteigenschaften der Farey-Folgen eingehen.

Definition:

Eine **Farey-Folge** der Ordnung n ist die aufsteigende Folge der unkürzbaren Brüche zwischen 0 und 1 mit Nenner $\leq n$.

Definition:

Der **Mediant** zweier Brüche $\frac{a}{b}, \frac{c}{d}$ mit positiven Nennern ist der Bruch $\frac{a+c}{b+d}$.

Satz:

Ist $\frac{a}{b} < \frac{c}{d}$, so liegt der Mediant genau zwischen diesen Brüchen, d.h. es gilt:

$$\frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}.$$

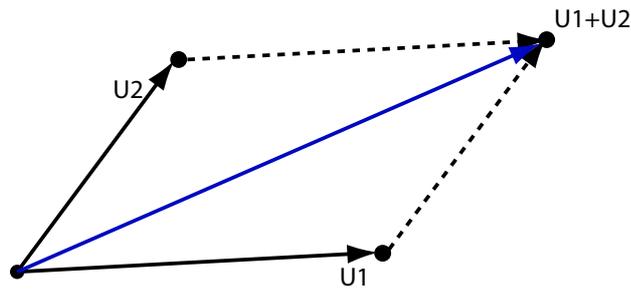
Beweis:

$$\begin{aligned} & \frac{a}{b} < \frac{c}{d} \\ \Leftrightarrow & a \cdot d < b \cdot c \\ 1. \Leftrightarrow & a \cdot d^2 < b \cdot c \cdot d \\ \Leftrightarrow & a \cdot d^2 + a \cdot b \cdot d < b \cdot c \cdot d + a \cdot b \cdot d \\ \Leftrightarrow & a \cdot d \cdot (b+d) < b \cdot d \cdot (a+c) \\ 2. \Leftrightarrow & a \cdot b \cdot d < b^2 \cdot c \\ \Leftrightarrow & a \cdot b \cdot d + b \cdot c \cdot d < b \cdot c \cdot (b+d) \\ \Leftrightarrow & b \cdot d \cdot (a+c) < b \cdot c \cdot (b+d) \end{aligned}$$

Aus 1. und 2. folgt:

$$\begin{aligned} \Leftrightarrow & a \cdot d \cdot (b+d) < b \cdot d \cdot (a+c) < b \cdot c \cdot (b+d) \\ \Leftrightarrow & \frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d} \end{aligned}$$

Die Medianteneigenschaft hat eine einfache geometrische Deutung: wir ordnen den Brüchen $\frac{a}{b}$ bzw. $\frac{c}{d}$ die Punkte $U_1 = (b; a)$ bzw. $U_2 = B(d; c)$ des Fundamentalgitters zu. Der Mediant $\frac{a+c}{b+d}$ entspricht dann der Diagonalen des Additionsparallelogramms $U_1 + U_2$.



Satz (Haupteigenschaften der Farey-Folgen):

I. Für zwei aufeinanderfolgende Brüche $\frac{a}{b}, \frac{c}{d} \in \mathfrak{F}_n$ gilt:

$$|bc - ad| = 1. \quad (1)$$

II. Für drei aufeinanderfolgende Brüche $\frac{a}{b}, \frac{x}{y}, \frac{c}{d} \in \mathfrak{F}_n$ gilt:

$$\frac{x}{y} = \frac{a+c}{b+d}. \quad (2)$$

Bemerkung:

Hauptsatz I. und Hauptsatz II. sind äquivalent.

Beweis:

(I \Rightarrow II) Seien $\frac{a}{b}, \frac{x}{y}, \frac{c}{d} \in \mathfrak{F}_n$ drei aufeinanderfolgende Brüche beliebig gewählt. Nach 1. gilt:

$$\begin{aligned} & bx - ay = 1 \wedge cy - xd = 1 \\ \Rightarrow & x(b+d) = y(a+c) \\ \Leftrightarrow & \frac{x}{y} = \frac{a+c}{b+d} \end{aligned}$$

(II \Rightarrow I) beweisen wir durch Induktion.

Der 2. Satz gilt für alle Farey-Folgen der Ordnung $n \in \mathbb{N}$.

Induktionsanfang: $n=1$ $|1 \cdot 1 - 1 \cdot 0| = 1 \rightarrow 1$. gilt für F_1

Induktionsschritt:

Induktionsannahme: 1. gilt für \mathfrak{F}_n

Induktionsbehauptung: 1. gilt auch für \mathfrak{F}_n . Beweis : Seien $\frac{a}{b}; \frac{x}{y}; \frac{c}{d}$ drei beliebige benachbarte Brüche der Farey-Folge \mathfrak{F}_n .

Nach 2. gilt: $\frac{x}{y} = \frac{a+c}{b+d}$.

Also sind $\frac{a}{b}$ und $\frac{a+c}{b+d}$ zwei benachbarte Brüche.

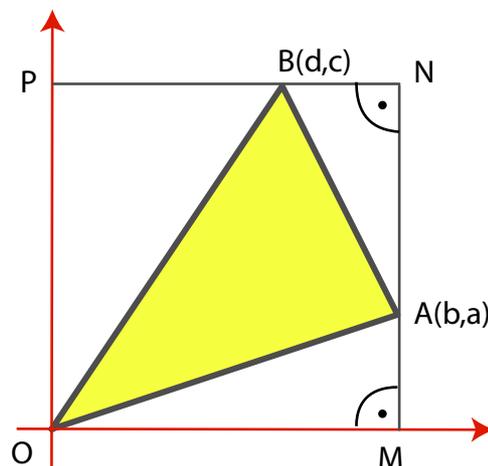
$$|b \cdot (a+c) - a \cdot (b+d)| = |ba + bc - ab - ad| = |bc - ad|$$

Nach der Induktionsannahme gilt:

$$|bc - ad| = 1 \rightarrow |b \cdot (a+c) - a \cdot (b+d)| = 1$$

\rightarrow 1. gilt auch für \mathfrak{F}_n

Für den Beweis der Haupteigenschaften reicht es also, zu beweisen, dass eine Eigenschaft gilt, da sie äquivalent sind. Wir beweisen die erste Eigenschaft geometrisch.



Seien $\frac{a}{b}$ und $\frac{c}{d}$ sind zwei benachbarte Brüche der Farey-Folge \mathfrak{F}_n . Wir ordnen den Brüchen $\frac{a}{b}$ bzw. $\frac{c}{d}$ die Punkte A(b;a) bzw. B(d;c) des Fundamentalgitters zu und bekommen ein Dreieck $\triangle AOB$. Im Inneren des Dreiecks $\triangle AOB$ können keine weitere Gitterpunkte liegen, da $\frac{a}{b}$ und $\frac{c}{d}$ benachbarte Brüche sind.

Auf \overline{OB} und \overline{OA} können auch keine weitere Gitterpunkte liegen, da die Brüche der Farey-Folge unkürzbar sind. Also können wir den Flächeninhalt des Dreiecks $\triangle AOB$ mit dem Satz von Pick wie folgt berechnen:

$$A(\triangle AOB) = 0 + \frac{3}{2} - 1 = \frac{1}{2}$$

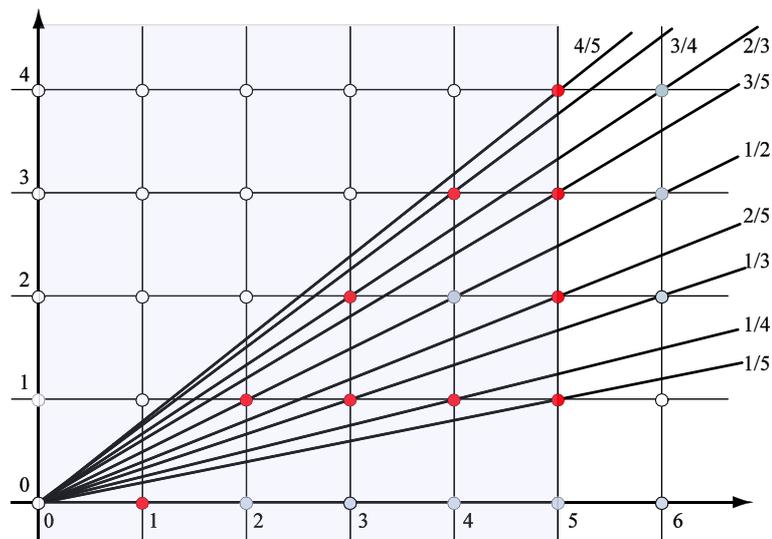


Abbildung 4: Die den Brüchen der Farey-Folge F_5 zugeordneten Ursprungsgeraden. Beachte, dass der Ordnung der Farey-Folge nach Größe der Ordnung der Ursprungsgeraden nach aufsteigender Steigung entspricht.

Wir können außerdem diesen Flächeninhalt auch analytisch berechnen:

$$\begin{aligned}
 A(\triangle AOB) &= A(\square MNP0) - A(\triangle BNA) - A(\triangle AMO) - A(\triangle PBO) \\
 &= \left| b \cdot c - \frac{1}{2} \cdot (c - a) \cdot (b - d) - \frac{1}{2} \cdot ab - \frac{1}{2} \cdot cd \right| \\
 &= \frac{1}{2} \cdot |bc - ad|
 \end{aligned}$$

Also folgt:

$$\begin{aligned}
 A(\triangle AOB) &= \frac{1}{2} \cdot |bc - ad| = \frac{1}{2} \\
 \Rightarrow |bc - ad| &= 1 \quad (q.e.d)
 \end{aligned}$$

Die Haupteigenschaften der Farey-Folgen sind wichtige Hilfsmittel um die Größe von Brüchen miteinander zu vergleichen ohne sie auszurechnen zu müssen.

Beispiel

Welcher Bruch ist größer:

$$A = \frac{100^{100} + 1}{100^{90} + 1} \quad \text{oder} \quad B = \frac{100^{99} + 1}{100^{89} + 1} ?$$

3 Lineare diophantische Gleichungen

Eine lineare Gleichung $15x + 17y = 1$ beschreibt eine Gerade in der Euklidischen Ebene. Gibt es Punkte mit ganzzahligen Koordinaten, die auf dieser Geraden liegen? Zur praktischen Berechnung haben wir das Eulersche Reduktionsverfahren eingeführt. Hier stellen wir nur kurz den zahlentheoretischen Haupteigenschaften zusammen.

Satz: Die lineare diophantische Gleichung der Form

$$ax + by = 1 \quad (3)$$

mit $a, b \in \mathbb{N}$ und $ggT(a; b) = 1$ ist immer ganzzahlig lösbar.

Beweis: o.B.d.A. gilt: $0 < a < b \Rightarrow \frac{a}{b} < 1$ (1)

Aus $ggT(a; b) = 1$ folgt, dass $\frac{a}{b}$ unkürzbar ist. (2)

Aus (1) und (2) schlussfolgern wir, dass der Bruch $\frac{a}{b}$ in der Farey-Folge der Ordnung b (also \mathfrak{F}_b) liegen muss.

Daraus folgt, dass es einen benachbarten Bruch $\frac{h}{k} \in F_b$ existiert mit: $\frac{h}{k} < \frac{a}{b}$.

Nach der Eigenschaft 1. der Farey Folge gilt also: $ak - bh = 1$

Also bekommen wir folgende Lösungen für (3): $x = k$ und $y = -h$.

Folgerung: Die allgemeine lineare diophantische Gleichung der Form $ax + by = c$ mit $a, b \in \mathbb{N}$ und $ggT(a; b) = 1$ ist ganzzahlig lösbar.

Begründung: Sei $x_0, y_0 \in \mathbb{Z}$ die Lösung für die Gleichung (*), also $ax_0 + by_0 = 1$. Dann ist $x = x_0 \cdot c$ und $y = y_0 \cdot c$ eine mögliche Lösung für die Gleichung $ax + by = c$.

Für den folgenden Satz benötigen wir das fundamentale

Lemma von Euklid:

Falls $ggT(a; b) = 1$ und a teilt das Produkt $b \cdot c$, so muss a die Zahl c teilen.

Bezeichnung: x teilt y schreiben wir als $x|y$.

Beweis:

Da $ggT(a; b) = 1$ hat $ax + by = 1$ nach obigem Satz eine ganzzahlige Lösung x, y . Dann gilt $acx + bcy = c$ und da $b \cdot c$ ein Vielfaches von a ist, können wir das Produkt in der Form $b \cdot c = a \cdot d$ schreiben. Ersetzen wir dies in $acx + bcy = c$ so erhalten wir

$acx + ady = c$, woraus ersichtlich ist, dass $a|c$.

Mithilfe dieses Lemmas von Euklid können wir alle Lösungen von $ax + by = 1$ bestimmen:

Satz:

Sei $\text{ggT}(a; b) = 1$, dann haben alle Lösungen der $ax + by = 1$ die Form:

$$x = x_0 - bt$$

und

$$y = y_0 + at,$$

wobei x_0, y_0 eine beliebige Lösung von (3) und $t \in Z$ ist.

Beweis:

Einfaches Einsetzen zeigt, dass $x = x_0 - bt$ und $y = y_0 + at$ Lösungen von $ax + by = 1$ sind. Das auch alle Lösungen diese Form haben zeigt man wie folgt:

sei x, y eine weitere ganzzahlige Lösungen für $ax + by = 1$. Dann folgt:

$$\begin{aligned} a \cdot x_0 + b \cdot y_0 &= a \cdot x + b \cdot y = 1 \\ \Rightarrow a \cdot (x - x_0) &= -b(y - y_0) \end{aligned}$$

Aber $\text{ggT}(a, b) = 1$ impliziert nach Euklids Lemma, dass $a|(y - y_0)$, also $a \cdot t = y - y_0$. Damit folgt aber sofort: $-b \cdot t = x - x_0$.

4 Kettenbrüche

Ein großes Problem der Mathematik war lange Zeit die rationale Approximation von irrationalen Zahlen und Brüchen mit großen Nennern. Mit Hilfe von Kettenbrüchen ist dies möglich. Schon im alten Griechenland entdeckte Euklid eine Methode, die Grundlage für die Erstellung der Kettenbrüche ist:

Euklidischer Algorithmus

Berechnung des größten gemeinsamen Teilers der Zahlen $x, y \in N$ mit $y \geq x$:

$$\begin{aligned} y &= a_1 \cdot x + r_1 \\ x &= a_2 \cdot r_1 + r_2 \\ r_1 &= a_3 \cdot r_2 + r_3 \\ &\dots \\ r_{n+1} &= a_n \cdot r_n + 0 \end{aligned}$$

Dann ist der $ggT(x, y) = r_n$. Über diesen Algorithmus kommen wir auch zur Kettenbruchdarstellung:

Sei $x \in \mathbb{R}_{>0}$ und a_0 der ganzzahlige Anteil von x . Dann ist

$$x = a_0 + r_0, \quad (4)$$

wobei $0 \leq r_0 < 1$ ist. Ist $r_0 \neq 0$ so können wir

$$\frac{1}{r_0} = a_1 + r_1 \quad (5)$$

berechnen, wobei wieder $0 \leq r_1 < 1$ ist und so fort. Wir erhalten durch Zusammensetzung der einzelnen Ausdrücke dann die regelmässige Kettenbruchentwicklung von x :

$$x = [a_0; a_1, a_2, a_3, \dots] := a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}} \quad (6)$$

Die Glieder a_i heißen die **Teil- oder Partialnenner** der Kettenbruchentwicklung von x . Berechnet man die Kettenbruchentwicklung bis zum i -ten Partialbruch, so heißen die entstehenden gekürzten Brüche P_i/Q_i die **(Haupt-)Konvergenten** von x :

$$\begin{aligned} \frac{P_0}{Q_0} &= \frac{a_0}{1}, \\ \frac{P_1}{Q_1} &= a_0 + \frac{1}{a_1} = [a_0; a_1], \\ \frac{P_i}{Q_i} &= [a_0; a_1, a_2, a_3, \dots, a_i]. \end{aligned}$$

Es gilt:

Satz:

1. Für $k \geq 1$ gilt:

$$\begin{cases} P_{k+1} = a_{k+1} \cdot P_k + P_{k-1} \\ Q_{k+1} = a_{k+1} \cdot Q_k + Q_{k-1}. \end{cases}$$

2. (Determinantensatz der Kettenbrüche)

Für $k \geq 0$ gilt:

$$P_k \cdot Q_{k+1} - P_{k+1} \cdot Q_k = (-1)^k.$$

Beweis: siehe Ford [2].

Beispiel:

Der Goldener Schnitt $\tau = \frac{\sqrt{5}+1}{2}$ hat die Kettenbruchentwicklung $\tau = [1; 1, 1, 1, 1, 1, 1, 1, \dots]$

$$\tau = [1; 1, 1, 1, 1, 1, 1, \dots] = \frac{1}{1 + \frac{1}{1 + \frac{1}{1 + \dots}}} \quad (7)$$

mit den Konvergenten F_{k+1}/F_k , wobei F_k die k -te Fibonacci-Zahl ist.

Mit Hilfe der Kettenbruchdarstellung kann jede reelle Zahl dargestellt und approximiert werden.

Beispiel:

Bei der Erstellung des Kalenders ergibt sich die Frage, wie stark dieser von dem astronomischen Jahr abweicht und wie man die Anzahl der Schalttage berechnet. Pro Jahr ergibt sich eine Abweichung $F = \frac{48443}{200000}$ Tagen. Als Kettenbruch dargestellt ergibt sich:

$$F = [0; 4, 7, 1, 3, \dots]$$

Daraus folgen die Konvergenten:

$$F \approx 0; \frac{1}{4}; \frac{7}{29}; \frac{8}{33}; \dots$$

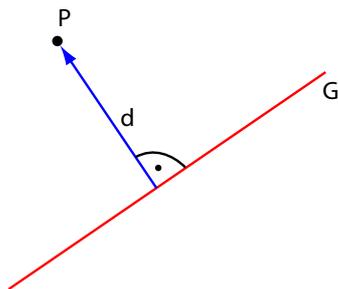
Der Julianische Kalender(46 v.Chr.) führt alle 4 Jahre ein Schaltjahr ein, entsprechend der Approximation $\frac{1}{4}$. Der Persische Kalender(1079 n.Chr.) arbeitet mit 8 Schalttagen alle 33 Jahre, benutzt also die genauere Approximation $\frac{8}{33}$.

Geometrische Deutung des Kettenbruchalgorithmus (nach Felix Klein, 1907)

Sei G die Ursprungsgerade $y=\alpha \cdot x$. Die Menge der nächsten Punkte von G ist die Menge aller Punkte $(p, q) \in \mathbb{Z}^2, q \geq 1$, sodass für alle $m, n \in \mathbb{Z}$ mit $0 < n \leq q$ gilt:

$$\text{dist}\{(q, p); G\} \leq \text{dist}\{(n, m); G\} ,$$

wobei $\text{dist}\{A;B\}$ den euklidischen Abstand zwischen den Punkten A und B bezeichne.



Satz:

Gegeben sei eine Ursprungsgerade G und zwei Punkte U_1, U_2 auf verschiedenen Seiten von G .

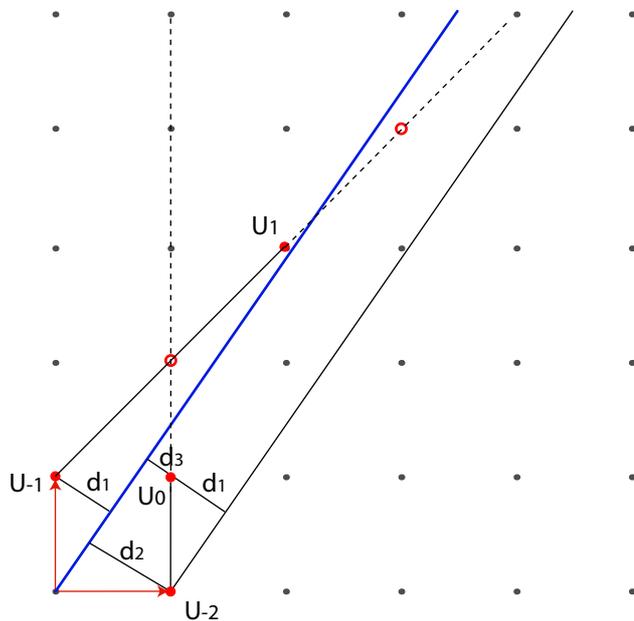
Es gibt dann eine eindeutige ganze Zahl, sodass entweder $\vec{U}_1 + a\vec{U}_2$ auf G liegt oder $\vec{U}_1 + a\vec{U}_2$ und $\vec{U}_1 + (a + 1)\vec{U}_2$ auf verschiedenen Seiten von G liegen.

Der Punkt $\vec{U}_1 + a\vec{U}_2$ ist näher an G als U_1 und falls $\vec{U}_1 + a\vec{U}_2$ nicht auf G liegt, so ist $\vec{U}_1 + a\vec{U}_2$ auf derselben Seite von G wie U_1 .

Ist \vec{U}_1 näher an G als \vec{U}_2 , dann ist $a = 0$, während für den Fall das \vec{U}_2 näher oder gleich nah an G ist wie \vec{U}_1 gilt $a \leq 1$.

Beweis:

Betrachtet wird die Ursprungsgerade G , die durch das Gitter der beiden erzeugenden Vektoren \vec{U}_1 und \vec{U}_2 verläuft.



Falls $d_1 < d_2$, so liegt $\vec{U}_1 + \vec{U}_2$ auf der U_1 gegenüberliegenden Seite von $G \Rightarrow a = 0$.

Falls $d_1 = d_2$, so liegt $\vec{U}_1 + a\vec{U}_2$ auf G .

Falls $d_1 > d_2$, so liegt $\vec{U}_1 + a\vec{U}_2$ auf derselben Seite wie U_1 und hat einen Abstand von $d_1 - d_2$ etc. $\vec{U}_1 + a\vec{U}_2$ den Abstand $d_1 - ad_2$.

Allgemein gilt also: $\vec{U}_n = a_n\vec{U}_{n-1} + \vec{U}_{n-2}$ mit $\vec{U}_{-2} = (1,0)$ und $\vec{U}_{-1} = (0,1)$.

5 Ford-Kreise

Die Ford-Kreise können als "Landkarte" der rationalen Zahlen aufgefasst werden. Jedem irreduziblen Bruch p/q wird ein Ford-Kreis K wie folgt zugeordnet.

Definition:

Sei p/q ein irreduzibler Bruch. Der **Ford-Kreis** $K(\frac{p}{q})$ über $\frac{p}{q}$ ist der Kreis mit Mittelpunkt $M(\frac{p}{q}, \frac{1}{2q^2})$ und Radius $r = \frac{1}{2q^2}$.

Satz:

Sind $\frac{p}{q}$ und $\frac{P}{Q}$ zwei verschiedene Brüche, dann berühren sich die zugehörigen Ford-Kreise tangential oder sie sind disjunkt.

Beweis:

Betrachten wir die allgemeine Situation:

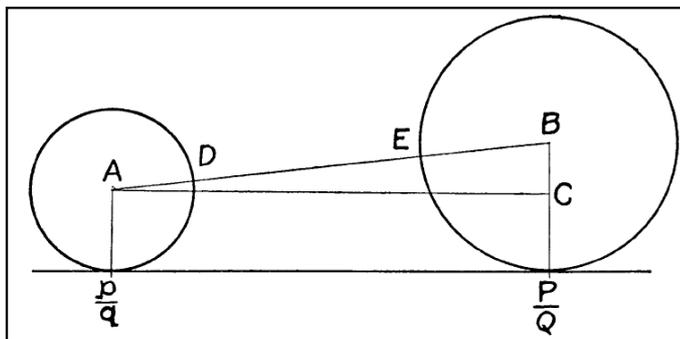


Abbildung 5: Die Abstandsberechnung zwischen zwei Ford-Kreisen (aus [2]).

$$\begin{aligned}
|\overline{AB}^2| &= |\overline{BC}^2| + |\overline{AC}^2| \\
|\overline{AB}^2| &= \left(\frac{P}{Q} - \frac{p}{q}\right)^2 + \left(\frac{1}{2Q^2} - \frac{1}{2q^2}\right)^2 \\
|\overline{AB}^2| &= \left(\frac{Pq - Qp}{Q^2p^2}\right)^2 + \frac{1}{4Q^4} - \frac{1}{2Q^2q^2} + \frac{1}{4q^4} \\
|\overline{AB}^2| &= \left(\frac{Pq - Qp}{Q^2p^2}\right)^2 + \frac{1}{4Q^4} - \frac{1}{2Q^2q^2} + \frac{1}{4q^4} + \frac{1}{2Q^2q^2} - \frac{1}{2Q^2q^2} \\
|\overline{AB}^2| &= \frac{(pQ - qP)^2 - 1}{Q^2q^2} + \underbrace{\left(\frac{1}{2Q^2} + \frac{1}{2q^2}\right)^2}_{(|\overline{AD}| + |\overline{EB}|)^2}
\end{aligned}$$

I. Fall : $|pQ - qP| = 1 \Rightarrow K\left(\frac{p}{q}\right)$ und $K\left(\frac{P}{Q}\right)$ berühren sich tangential

II. Fall : $|pQ - qP| > 1 \Rightarrow$ Kreise sind disjunkt.

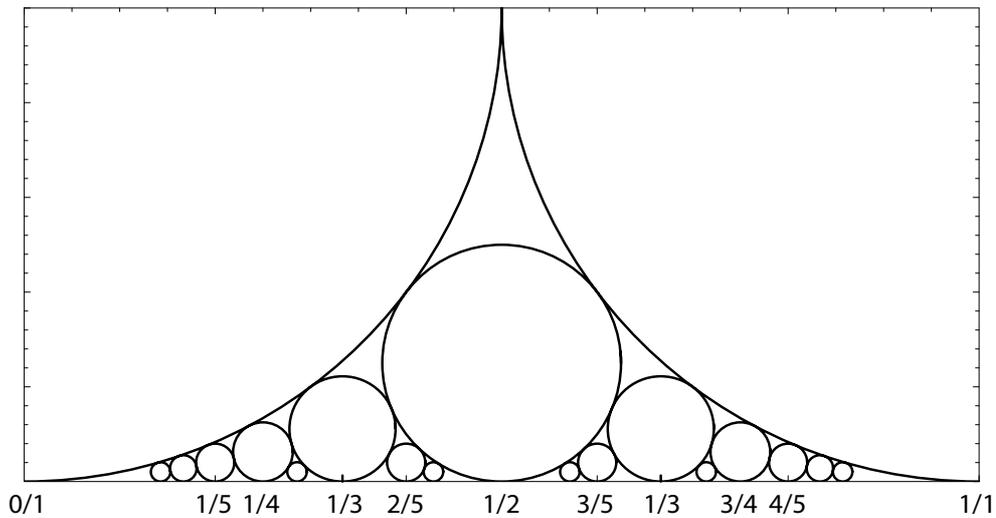


Abbildung 6: Die Ford-Kreise über der Farey-Folge \mathfrak{F}_5 .

Jeder Ford-Kreis wird von unendlich vielen Kreisen tangential berührt. Weiter zeigt sich, dass die erste Haupteigenschaft der Farey-Folge nicht unbedingt nur für aufeinander folgende Brüche gilt.

Am Beispiel

$$\mathfrak{F}_5 = \frac{0}{1}; \frac{1}{5}; \frac{1}{4}; \frac{1}{3}; \frac{2}{5}; \frac{1}{2}; \frac{3}{5}; \frac{2}{3}; \frac{3}{4}; \frac{4}{5}; \frac{1}{1}$$

erkennt man, dass $\frac{1}{2}$ und $\frac{2}{3}$ in \mathfrak{F}_5 nicht aufeinander folgen, aber trotzdem gilt

$$2 \cdot 2 - 1 \cdot 3 = 1.$$

Die Bedeutung der Ford-Kreise zeigt sich auch in der hierarchischen Anordnung der Kreise. Angenommen man lässt eine Gerade $y = c$ sich durch stetiges verkleinern von c der x-Achse parallel nähern, so würde diese Gerade die Ford-Kreise über den Brüchen der Farey-Folgen nacheinander in ihrer hierarchischen Reihenfolge $\{\mathfrak{F}_1, \mathfrak{F}_2, \dots\}$ schneiden. Weiter lassen sich auch die Lösungen der diophantischen Gleichung

$$a \cdot x + b \cdot y = \pm 1; a, b \in \mathbb{Z}$$

mit $\text{ggT}(a, b) = 1$ mithilfe der Ford-Kreise veranschaulichen. Diese werden durch die den Ford-Kreis $K(a)$ tangential berührenden Ford-Kreise repräsentiert.

Literatur

- [1] Farey, J.: On a curious property of vulgar fractions, Philos. Mag. Journal, London **47**, 385-386, 1816.
- [2] Ford, L.: Fractions. Am. Math. Monthly **45**, 586-601, 1938.
- [3] Rademacher, H.: Lectures on Elementary Number Theory. Blaisdell, New York u.a., 1964.

Konvexe Mengen und konvexe Funktionen

Teilnehmer:

Moritz Butz	Herder-Gymnasium
Franziska Ihlefeldt	Herder-Gymnasium
Johannes Jendersie	Georg-Forster-Gymnasium
Marie Lambert	Georg-Forster-Gymnasium
Eike Müller	Herder-Gymnasium
Gregor Pasemann	Georg-Forster-Gymnasium
Konstantin Rohde	Heinrich-Hertz-Gymnasium

Gruppenleiter:

Konrad Gröger	Humboldt-Universität zu Berlin
---------------	--------------------------------

Im Rahmen des Projektes haben wir uns mit konvexen Mengen und konvexen Funktionen im Raum \mathbb{R}^N beschäftigt. Innerhalb dieses Raumes haben wir Begriffe und Sachverhalte erörtert, die mit Konvexität im Zusammenhang stehen. Beispiele hierfür sind die konvexe Hülle einer beliebigen Teilmenge von \mathbb{R}^N , konvexe Kombinationen von Punkten aus \mathbb{R}^N , Polytope und Polyeder. Für konvexe Mengen haben wir einen Trennungssatz bewiesen und mit dessen Hilfe gezeigt, dass jede abgeschlossene konvexe Teilmenge von \mathbb{R}^N der Durchschnitt von abgeschlossenen Halbräumen ist.

Nach der Betrachtung der konvexen Mengen haben wir uns konvexen Funktionen zugewandt. Mit Hilfe von Epigraphen konnten wir Fragestellungen für konvexe Funktionen auf die konvexen Mengen zurückführen. Außerdem haben wir das Subdifferential von konvexen Funktionen definiert und an konkreten Beispielen veranschaulicht. Das Subdifferential, eine mehrdeutige Abbildung, lässt sich als ein Ersatz für die Ableitung von nicht überall differenzierbaren konvexen Funktionen ansehen. Abschließend haben wir komplexe Optimierungsprobleme, mit denen man im Zusammenhang mit konvexen Mengen und Funktionen konfrontiert werden kann, angedeutet und deren Lösbarkeit nachgewiesen.

1 Konvexe Mengen

Der N -dimensionale Raum \mathbb{R}^N ist definiert als das N -fache kartesische Produkt der reellen Zahlen mit sich selbst. Ein Punkt x dieses Raumes kann durch N Koordinaten x_1 bis x_N dargestellt werden:

$$x := (x_1, x_2, \dots, x_N) \in \mathbb{R}^N.$$

Folgende Regeln sind für das Rechnen mit Punkten aus \mathbb{R}^N erklärt:

Sind $x = (x_1, x_2, \dots, x_N)$ und $y = (y_1, y_2, \dots, y_N)$ zwei Punkte aus \mathbb{R}^N und ist t eine reelle Zahl, so setzt man

$$x + y := (x_1 + y_1, x_2 + y_2, \dots, x_N + y_N)$$

und

$$tx := (tx_1, tx_2, \dots, tx_N).$$

Definition 1 Eine Teilmenge $E \subset \mathbb{R}^N$ ¹⁾ heißt **konvex**, wenn die Verbindungsstrecke von je zwei Punkten $x, y \in E$ zu E gehört.

Eine Teilmenge $E \subset \mathbb{R}^1$ ist genau dann konvex, wenn E ein Intervall ist, denn die Verbindungsstrecke von zwei Punkten aus dem Intervall liegt auch wieder in dem Intervall. Beispiele für konvexe und nicht konvexe Teilmengen von \mathbb{R}^2 zeigt die Abbildung 1.

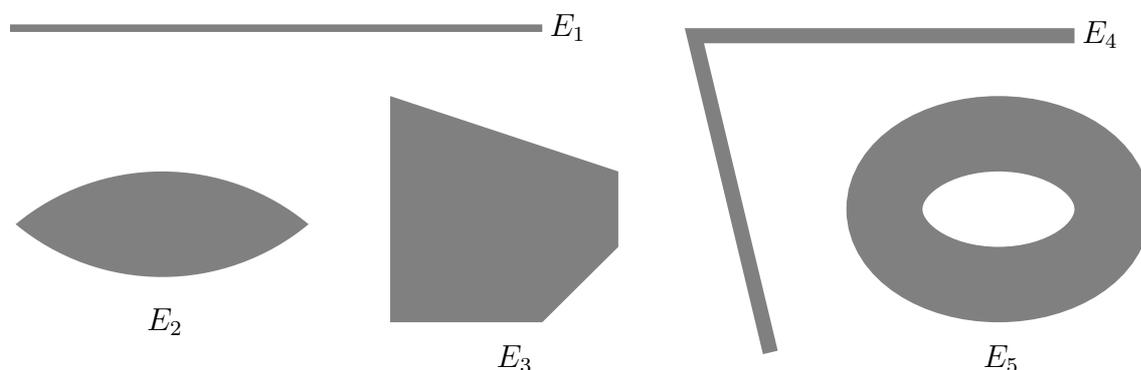


Abbildung 1: E_1, E_2, E_3 sind konvex, E_4, E_5 sind nicht konvex

Die leere Menge und alle einelementigen Mengen sind konvex, denn es existieren keine zwei Punkte in diesen Mengen, somit müssen diese Mengen keine Bedingung erfüllen, um konvex zu sein.

Satz 1 Der Durchschnitt beliebig vieler konvexer Teilmengen aus \mathbb{R}^N ist konvex.

Beweis:

Liegen x und y in allen beteiligten konvexen Teilmengen, so liegt die Verbindungsstrecke auch in allen beteiligten konvexen Teilmengen und somit im Durchschnitt. \square

¹⁾ $E \subset \mathbb{R}^N$ bedeutet, dass E eine Teilmenge von \mathbb{R}^N ist. E kann dabei auch gleich \mathbb{R}^N sein.

Satz 2 Zu jeder Menge $E \subset \mathbb{R}^N$ existiert eine kleinste konvexe Menge, die E enthält.

Beweis:

Nach Satz 1 ist der Durchschnitt aller konvexen Mengen, die E enthalten, wieder konvex. Dieser Durchschnitt ist gleich der kleinsten konvexen Menge, die E enthält. Gäbe es noch eine kleinere konvexe Menge, die E enthält, so wird mit ihr auch die Schnittmenge gebildet und somit kann sie nicht kleiner sein als der Durchschnitt. \square

Definition 2 Die kleinste konvexe Menge, die E enthält, wird **konvexe Hülle** von E genannt und mit $\text{conv}(E)$ bezeichnet. Eine Teilmenge von \mathbb{R}^N wird **Polytop** genannt, wenn sie die konvexe Hülle einer endliche Teilmenge von \mathbb{R}^N ist.

Beispielsweise sind konvexe n -Ecke in \mathbb{R}^2 Polytope, ebenso Quader, Tetraeder und Oktaeder in \mathbb{R}^3 . Dagegen sind Kreisscheiben und Kreiskegel zwar konvex, aber keine Polytope.

Definition 3 Man nennt einen Punkt $x \in \mathbb{R}^N$ **konvexe Kombination** der Punkte v_1, \dots, v_n aus \mathbb{R}^N , wenn nicht negative Zahlen t_1, \dots, t_n mit der Eigenschaft $\sum_{i=1}^n t_i = 1$ existieren, sodass $x = \sum_{i=1}^n t_i v_i$ gilt.

Satz 3 Ist E eine konvexe Teilmenge von \mathbb{R}^N , so gehören zu E alle konvexen Kombinationen von je endlich vielen Punkten v_1, \dots, v_n aus E .

Beweis:

Der Beweis wird mit Hilfe der vollständigen Induktion geführt. Für $n = 1$ ist die Behauptung trivial. Nehmen wir nun an, dass die Behauptung für alle konvexen Kombinationen von $n - 1$ Punkten gilt. Es sei

$$x = \sum_{i=1}^n t_i v_i, \quad t_i > 0, \quad \sum_{i=1}^n t_i = 1, \quad v_1, \dots, v_n \in E.$$

Dann gilt

$$x = (1 - t_n) \sum_{i=1}^{n-1} \frac{t_i v_i}{1 - t_n} + t_n v_n.$$

Es ist $\sum_{i=1}^{n-1} \frac{t_i}{1 - t_n} = 1$ und $\frac{t_i}{1 - t_n} > 0$. Nach der Induktionsannahme gehört $y := \sum_{i=1}^{n-1} \frac{t_i}{1 - t_n} v_i$ zu E . Da auch v_n zu E gehört und $x = (1 - t_n)y + t_n v_n$ Punkt der Verbindungsstrecke von y und v_n ist, gilt $x \in E$. Somit ist die Induktion abgeschlossen. \square

Satz 4 Ist $E \subset \mathbb{R}^N$ beliebig, so ist die Menge K aller konvexen Kombinationen von je endlich vielen Punkten von E konvex.

Beweis:

Es sei $x, y \in K$. Es gelte etwa

$$x = \sum_{i=1}^n r_i v_i, \quad r_i \geq 0, \quad \sum_{i=1}^n r_i = 1, \quad y = \sum_{i=1}^n s_i v_i, \quad s_i \geq 0, \quad \sum_{i=1}^n s_i = 1;$$

dabei seien v_1, \dots, v_n Punkte von E . Wir können bei beiden Punkten x und y deshalb mit den gleichen v_i arbeiten, weil man auch "überflüssige" Punkte mit dem Koeffizienten 0 mitnehmen kann. Dann ist für $t \in [0, 1]$

$$(1-t)x + ty = \sum_{i=1}^n ((1-t)r_i + ts_i)v_i = \sum_{i=1}^n t_i v_i \quad \text{für } t_i := (1-t)r_i + ts_i.$$

Offensichtlich ist $t_i \geq 0$ und $\sum_{i=1}^n t_i = (1-t)\sum_{i=1}^n r_i + t\sum_{i=1}^n s_i = 1-t+t = 1$. Folglich gehören die Punkte $(1-t)x + ty$ für $t \in [0, 1]$ zu K , d. h., K ist konvex. \square

Folgerung 1 Die konvexe Hülle $\text{conv}(E)$ einer beliebigen Menge $E \subset \mathbb{R}^N$ ist gleich der Menge alle konvexen Kombinationen von Punkten aus E .

Beweis:

Nach Satz 3 muss $\text{conv}(E)$ als konvexe Menge die Menge K der konvexen Kombinationen von Punkten aus E enthalten. Nach Satz 4 ist K konvex und daher $\text{conv}(E)$ als kleinste konvexe Menge, die E enthält, in K enthalten.

Für $x = (x_1, x_2, \dots, x_N)$, $y = (y_1, y_2, \dots, y_N) \in \mathbb{R}^N$ definiert man das **Skalarprodukt** $x \cdot y$ und den **Betrag** $|x|$ durch die Vorschriften

$$x \cdot y := \sum_{i=1}^N x_i y_i, \quad |x| := \sqrt{x \cdot x} = \left(\sum_{i=1}^N x_i^2 \right)^{\frac{1}{2}}. \quad (1)$$

Der Betrag $|x|$ ist der Abstand des Punktes x vom Nullpunkt bzw. die Länge des Vektors x . Für $N = 1, 2, 3$ ist das mit dem Satz des Pythagoras zu begründen, für größere N ist das als *Definition* des Abstands anzusehen.

Das Skalarprodukt $x \cdot y$ ist symmetrisch in x und y , d. h., es gilt $y \cdot x = x \cdot y$. Der Wert $x \cdot y$ hängt sowohl von x als auch von y linear ab, d. h., es ist

$$(x + \xi) \cdot y = x \cdot y + \xi \cdot y, \quad x \cdot (y + \eta) = x \cdot y + x \cdot \eta, \quad (tx) \cdot y = x \cdot (ty) = t(x \cdot y) \quad (2)$$

für beliebige $x, y, \xi, \eta \in \mathbb{R}^N$ und $t \in \mathbb{R}$. Den Nullpunkt von \mathbb{R}^N bezeichnen wir wie die Zahl Null mit 0. Welche Bedeutung das Symbol 0 jeweils hat, wird aus dem Zusammenhang hervorgehen. Gilt für $x, y \in \mathbb{R}^N$ die Beziehung $x \cdot y = 0$, so ist

$$|y - x|^2 = (y - x) \cdot (y - x) = y \cdot y - y \cdot x - x \cdot y + x \cdot x = |y|^2 + |x|^2,$$

d. h., in dem Dreieck, das von x , $(y - x)$ und y gebildet wird, gilt ein Analogon des Satzes von Pythagoras. Deshalb werden Vektoren $x, y \in \mathbb{R}^N$, für die $x \cdot y = 0$ ist, zueinander senkrecht oder zueinander orthogonal genannt.

Hilfssatz Für beliebige $x, y \in \mathbb{R}^N$ gilt

$$x \cdot y \leq |x||y|. \quad (3)$$

Beweis:

Für $y = 0$ ist die Behauptung (3) trivial. Es sei nun $y \neq 0$ und $t := \frac{x \cdot y}{|y|^2}$. Dann ist

$$0 \leq |x - ty|^2 = (x - ty) \cdot (x - ty) = x \cdot x - 2tx \cdot y + t^2 y \cdot y = |x|^2 - 2 \frac{(x \cdot y)^2}{|y|^2} + \frac{(x \cdot y)^2}{|y|^2}.$$

Folglich ist $0 \leq |x|^2|y|^2 - (x \cdot y)^2$, d. h., es gilt die Ungleichung (3). \square

Die Beziehung (3) nennt man **Cauchy-Schwarzsche Ungleichung**.

Für beliebige $x, y \in \mathbb{R}^N$ ergibt sich unter Benutzung der Cauchy-Schwarzschen Ungleichung $|x + y|^2 = (x + y) \cdot (x + y) = |x|^2 + 2x \cdot y + |y|^2 \leq (|x| + |y|)^2$. Folglich ist

$$|x + y| \leq |x| + |y| \quad \text{für beliebige } x, y \in \mathbb{R}^N. \quad (4)$$

Die Ungleichung (4) bezeichnet man als **Dreiecksungleichung**.

Die Menge

$$B_r(\xi) = \{a \in \mathbb{R}^N; |a - \xi| < r\} \quad (5)$$

wird **Kugel** in \mathbb{R}^N genannt mit dem Mittelpunkt ξ und dem Radius r . Existiert zu einem Punkt $\xi \in E$ ein $r > 0$, sodass $B_r(\xi) \subset E$ gilt, so wird ξ **innerer Punkt** von E genannt.

Es sei $x_n = (x_{n1}, x_{n2}, \dots, x_{nN})$ für $n \in \mathbb{N}$ eine Folge von Punkten im \mathbb{R}^N . Die Folge (x_n) heißt **konvergent** gegen $\xi = (\xi_1, \dots, \xi_N)$, falls $\lim_{n \rightarrow \infty} |x_n - \xi| = 0$ gilt. Dies ist der Fall, wenn die Folge koordinatenweise konvergiert. Es muss somit für alle $i = 1, \dots, N$ gelten $\lim_{n \rightarrow \infty} |x_{ni} - \xi_i| = 0$.

Die Folge (x_n) heißt **Cauchyfolge**, wenn $\lim_{m, n \rightarrow \infty} |x_m - x_n| = 0$ ist. Jede Cauchyfolge in \mathbb{R}^N besitzt einen Grenzwert. Das folgt daraus, dass eine Folge (x_n) im \mathbb{R}^N genau dann Cauchyfolge ist, wenn die Koordinatenfolgen (x_{ni}) ($i = 1, \dots, N$) Cauchyfolgen reeller Zahlen sind.

Definition 4 Eine Menge $E \subset \mathbb{R}^N$ heißt **abgeschlossen**, wenn für jede konvergente Folge (x_n) aus E deren Grenzwert $\lim_{n \rightarrow \infty} x_n$ auch in E liegt. Eine Menge $E \subset \mathbb{R}^N$ heißt **offen**, wenn jeder Punkt $\xi \in E$ innerer Punkt von E ist.

Der Durchschnitt beliebig vieler abgeschlossener Mengen in \mathbb{R}^N ist abgeschlossen. Zu jeder Menge $E \subset \mathbb{R}^N$ gibt es eine kleinste abgeschlossene Teilmenge und eine kleinste abgeschlossene, konvexe Teilmenge von \mathbb{R}^N , die E enthält. Die erste nennt man die **abgeschlossene Hülle** von E , und man bezeichnet sie meist mit \overline{E} . Die zweite nennt man die **abgeschlossene konvexe Hülle** von E , und man bezeichnet sie meist mit $\overline{\text{conv}}(E)$. Zur Begründung

kann man wie bei der konvexen Hülle argumentieren (vgl. die Aussagen und Beweise von Satz 1 und Satz 2).

Man kann beweisen, dass die abgeschlossene Hülle einer konvexen Menge $E \subset \mathbb{R}^N$ ebenfalls konvex ist. Polytope sind stets abgeschlossen. Das kann man mit Hilfe von Folgerung 1 beweisen.

Wir kommen zu weiteren Beispielen für konvexe Mengen. Ist $K \subset \mathbb{R}^N$ eine konvexe, nicht leere Menge und $r > 0$, so sind auch folgende Mengen konvex:

$$\begin{aligned} U_r(K) &:= \{ \xi \in \mathbb{R}^N; |\xi - x| < r \text{ für ein } x \in K \} \\ \overline{U}_r(K) &:= \{ \xi \in \mathbb{R}^N; |\xi - x| \leq r \text{ für ein } x \in K \} \end{aligned}$$

Begründung:

Es sei $\xi \in U_r(K)$ und $\eta \in U_r(K)$. Wir wählen x und y in K so, dass $|\xi - x| < r$ und $|\eta - y| < r$ gilt. Solche Punkte x und y müssen nach der Definition von $U_r(K)$ existieren. Für $t \in [0, 1]$ ist dann $z := (1 - t)x + ty \in K$ und für $\zeta := (1 - t)\xi + t\eta$ gilt

$$|\zeta - z| = |(1 - t)(\xi - x) + t(\eta - y)| \leq (1 - t)|\xi - x| + t|\eta - y| < (1 - t)r + tr = r. \quad (6)$$

Folglich gehört ζ zu $U_r(K)$. Mit zwei Punkten ξ, η liegt also auch deren Verbindungsstrecke in $U_r(K)$, d. h., $U_r(K)$ ist konvex. Für $\overline{U}_r(K)$ argumentiert man analog. \square

Diese Mengen werden **Umgebungen** von K genannt.

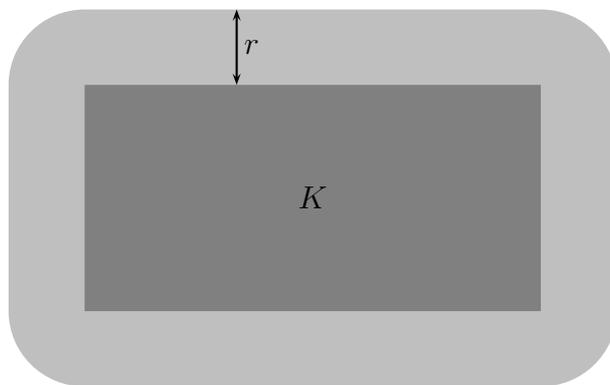


Abbildung 2: Umgebung einer konvexen Menge K

Weil eine aus einem Punkt bestehende Menge K konvex ist, zeigt die Feststellung insbesondere, dass offene und abgeschlossene Kugeln konvex sind.

Es gibt noch weitere spezielle konvexe Mengen. Seien $a \in \mathbb{R}^N \setminus \{0\}$ und $z \in \mathbb{R}^N$ beliebig aber fest. Die Menge

$$H_{a,z} := \{ x \in \mathbb{R}^N; a \cdot (x - z) = 0 \}$$

besteht aus den Punkten x , für die der Vektor $(x - z)$ senkrecht zu a steht. Im \mathbb{R}^2 ist dies somit eine Gerade und im \mathbb{R}^3 eine Ebene. Im \mathbb{R}^N wird eine solche Menge **Hyperebene** genannt.

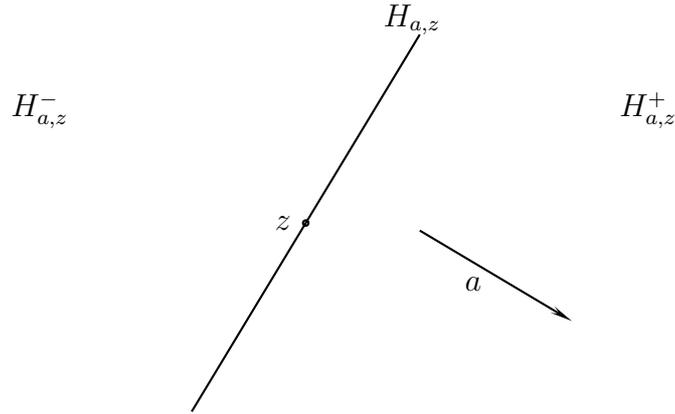


Abbildung 3: Hyperebene und Halbräume für $N = 2$

Die Hyperebene $H_{a,z}$ teilt den \mathbb{R}^N in die Mengen

$$\begin{aligned} H_{a,z}^+ &:= \{x \in \mathbb{R}^N; a \cdot (x - z) > 0\} \\ H_{a,z}^- &:= \{x \in \mathbb{R}^N; a \cdot (x - z) < 0\} \end{aligned}$$

auf, die **Halbräume** genannt werden.

Für beliebige $x, y \in \mathbb{R}^N$ und $t \in [0, 1]$ gilt:

$$a \cdot (tx + (1-t)y - z) = a \cdot (t(x-z) + (1-t)(y-z)) = ta \cdot (x-z) + (1-t)a \cdot (y-z)$$

Daher sind $H_{a,z}$, $H_{a,z}^+$ und $H_{a,z}^-$ konvexe Mengen. Die Hyperebene $H_{a,z}$ ist abgeschlossen und sowohl $H_{a,z}^+$ als auch $H_{a,z}^-$ sind offen. Die abgeschlossene Hülle von $H_{a,z}^+$ und $H_{a,z}^-$ erhält man, indem man die Hyperebene $H_{a,z}$ hinzunimmt.

Satz 5 *Es sei $K \subset \mathbb{R}^N$ konvex, abgeschlossen und nicht leer. Dann gibt es zu jedem Punkt $\xi \in \mathbb{R}^N$ genau einen Punkt $x \in K$, der ξ am nächsten liegt, für den also die Beziehung*

$$|\xi - x| < |\xi - y| \quad \text{für alle } y \in K \setminus \{x\}$$

gilt. Dieser Punkt x wird auch durch die Beziehung

$$x \in K, \quad (\xi - x) \cdot (y - x) \leq 0 \quad \text{für alle } y \in K \tag{7}$$

charakterisiert.

Beweis:

1. Für beliebige Punkte $x, y \in \mathbb{R}^N$ gilt

$$\begin{aligned} |x+y|^2 + |x-y|^2 &= (x+y) \cdot (x+y) + (x-y) \cdot (x-y) \\ &= x \cdot x + 2x \cdot y + y \cdot y + x \cdot x - 2x \cdot y + y \cdot y \\ &= 2|x|^2 + 2|y|^2. \end{aligned} \tag{8}$$

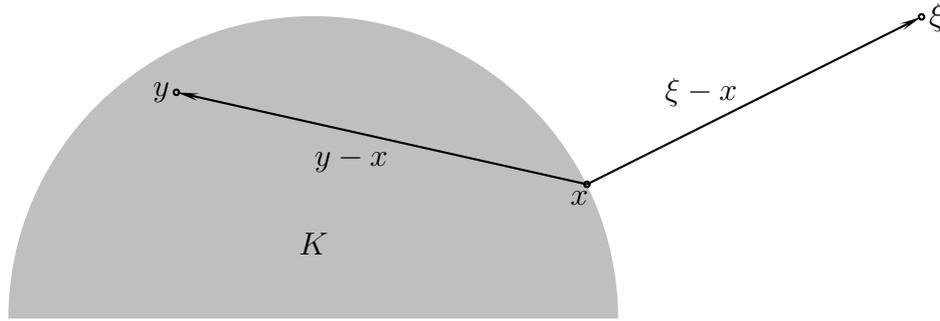


Abbildung 4: Konvexe Projektion

Wir setzen $d := \inf_{x \in K} |\xi - x|$ und wählen eine Folge (x_n) in K so, dass $\lim_{n \rightarrow \infty} |\xi - x_n| = d$ ist. Unter Verwendung der Formel (8) für $(\xi - x_n)$ und $(\xi - x_m)$ anstatt von x und y ergibt sich

$$\begin{aligned}
 |x_m - x_n|^2 &= |(\xi - x_n) - (\xi - x_m)|^2 \\
 &= 2|\xi - x_n|^2 + 2|\xi - x_m|^2 - |2\xi - (x_n + x_m)|^2 \\
 &= 2|\xi - x_n|^2 + 2|\xi - x_m|^2 - 4 \left| \xi - \frac{x_n + x_m}{2} \right|^2 \\
 &\leq 2|\xi - x_n|^2 + 2|\xi - x_m|^2 - 4d^2.
 \end{aligned}$$

Daher gilt

$$\lim_{m, n \rightarrow \infty} |x_m - x_n|^2 \leq 2d^2 + 2d^2 - 4d^2 = 0.$$

Somit gilt auch $\lim_{m, n \rightarrow \infty} |x_m - x_n| = 0$, d. h., die Folge (x_n) ist eine Cauchyfolge in \mathbb{R}^N . Der Grenzwert dieser Folge sei x . Dann gilt

$$|\xi - x| = \lim_{n \rightarrow \infty} |\xi - x_n| = d.$$

2. Nun muss noch gezeigt werden, dass für den Punkt x die Beziehung (7) gilt. Ist $y \in K$, so gehören auch die Punkte $z_t := (1-t)x + ty$, $0 \leq t \leq 1$, zu K . Wir setzen

$$\begin{aligned}
 \varphi(t) &:= |\xi - z_t|^2 \\
 &= |\xi - (1-t)x - ty|^2 \\
 &= (\xi - (1-t)x - ty) \cdot (\xi - (1-t)x - ty) \\
 &= (\xi - x + tx - ty) \cdot (\xi - x + tx - ty) \\
 &= |\xi - x|^2 - 2t(\xi - x) \cdot (y - x) + t^2|y - x|^2.
 \end{aligned}$$

Dann ist

$$\begin{aligned}
 \varphi'(t) &= -2(\xi - x) \cdot (y - x) + 2t|y - x|^2, \\
 \varphi'(0) &= -2(\xi - x) \cdot (y - x).
 \end{aligned}$$

Wäre $(\xi - x) \cdot (y - x) > 0$, so wäre $\varphi'(0) < 0$, und $\varphi(t) = |\xi - z_t|^2$ wäre für kleine $t > 0$ echt kleiner als $\varphi(0) = |\xi - x|^2 = d^2$. Das kann nach Definition von d nicht sein. Der Widerspruch beweist die Beziehung (7).

3. Es gelte die Beziehung (7). Für beliebiges $y \in K$ ist dann

$$|\xi - y|^2 = |\xi - x + x - y|^2 = |\xi - x|^2 + 2(\xi - x) \cdot (x - y) + |x - y|^2 \geq |\xi - x|^2 + |x - y|^2.$$

Folglich ist $|\xi - x| < |\xi - y|$ für alle $y \in K \setminus \{x\}$. Der Punkt x liegt also echt näher an ξ als alle anderen Punkte $y \in K$. \square

Wir definieren die Abbildung $P_K : \mathbb{R}^N \longrightarrow K$ durch die Vorschrift

$$P_K(\xi) := x, \quad \text{falls } x \in K, \quad |\xi - x| < |\xi - y| \text{ für alle } y \in K \setminus \{x\}.$$

Die Abbildung P_K ordnet also jedem Punkt $\xi \in \mathbb{R}^N$ den Punkt x der konvexen Menge K zu, der am nächsten zu ξ liegt. Man nennt P_K die **konvexe Projektion** von \mathbb{R}^N auf K . Die konvexe Projektion verallgemeinert die senkrechte Projektion auf eine Gerade bzw. Ebene.

Satz 6 *Es sei $K \subset \mathbb{R}^N$ konvex, abgeschlossen und nicht leer. Ferner sei ξ ein Punkt von \mathbb{R}^N , der nicht zu K gehört. Dann gibt es eine Hyperebene $H_{a,z}$ in \mathbb{R}^N , so dass $K \subset H_{a,z}^-$ und $\xi \in H_{a,z}^+$ gilt.*

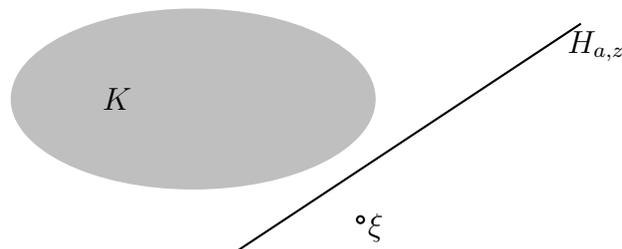


Abbildung 5: Trennung von konvexer Menge und Punkt

Beweis:

Es sei $x = P_K(\xi)$. Es ist $x \neq \xi$, weil ξ nicht zu K gehört. Wir setzen $a := \xi - x$ und $z := \frac{\xi + x}{2}$. Nach Wahl von a und z ist

$$a \cdot (\xi - z) = (\xi - x) \cdot \left(\xi - \frac{\xi + x}{2} \right) = \frac{1}{2} |\xi - x|^2 > 0.$$

Folglich gilt $\xi \in H_{a,z}^+$. Für $y \in K$ ergibt sich mit Hilfe von (7)

$$\begin{aligned} a \cdot (y - z) &= (\xi - x) \cdot \left(y - \frac{\xi + x}{2} \right) = (\xi - x) \cdot \left(y - x - \frac{\xi - x}{2} \right) \\ &= (\xi - x) \cdot (y - x) - \frac{1}{2} |\xi - x|^2 \leq -\frac{1}{2} |\xi - x|^2 < 0. \end{aligned}$$

Damit ist gezeigt, dass K in $H_{a,z}^-$ liegt. \square

Man sagt, die Hyperebene $H_{a,z}$ **trenne** den Punkt ξ und die konvexe Menge K .

Man kann den Satz verallgemeinern und statt ξ eine konvexe, abgeschlossene und beschränkte Menge L betrachten. Dann lässt sich eine Hyperebene finden, die K und L trennt. Der Beweis wird hier jedoch nicht angeführt.

Folgerung 2 *Jede konvexe abgeschlossene Teilmenge K von \mathbb{R}^N ist der Durchschnitt von Halbräumen, die K enthalten.*

Beweis:

Ist $K = \mathbb{R}^N$, so kann man diese Menge als Durchschnitt von null Halbräumen auffassen. Ist $K \subsetneq \mathbb{R}^N$, so sei D der Durchschnitt aller K enthaltenden abgeschlossenen Halbräume. Dann gilt $K \subset D$. Die Menge D kann aber nicht mehr Punkte enthalten als K . Ist nämlich $\xi \notin K$, so gibt es nach Satz 6 einen abgeschlossenen Halbraum, der K , aber nicht ξ enthält. \square

Als Anwendung stellen wir eine Variante des Mittelwertsatzes der Differentialrechnung für Funktionen auf \mathbb{R} mit Werten in \mathbb{R}^N vor. Es sei

$$f(t) = (f_1(t), \dots, f_N(t)) \quad \text{für } t \in \mathbb{R}.$$

Man nennt f stetig differenzierbar, wenn f_1, \dots, f_N stetig differenzierbar sind, und setzt $f'(t) := (f'_1(t), \dots, f'_N(t))$ für $t \in \mathbb{R}$.

Satz 7 *Es sei $f : \mathbb{R} \rightarrow \mathbb{R}^N$ stetig differenzierbar und $t_0, t_1 \in \mathbb{R}$, $t_0 < t_1$. Dann gilt*

$$\frac{f(t_1) - f(t_0)}{t_1 - t_0} \in \overline{\text{conv}}(E) \quad \text{für } E := \{f'(t); t_0 < t < t_1\}.$$

Beweis:

Wegen der Stetigkeit von f' ist E in \mathbb{R}^N beschränkt. Es sei $\overline{H_{a,z}^-}$ ein abgeschlossener Halbraum in \mathbb{R}^N , der E enthält. Dann gilt

$$a \cdot (f'(t) - z) = \sum_{i=1}^N a_i (f'_i(t) - z_i) \leq 0 \quad \text{für alle } t \in]t_0, t_1[.$$

Integriert man diese Ungleichung von t_0 bis t_1 und dividiert durch $t_1 - t_0$, so findet man

$$a \cdot \left(\frac{f(t_1) - f(t_0)}{t_1 - t_0} - z \right) = \sum_{i=1}^N a_i \left(\frac{f_i(t_1) - f_i(t_0)}{t_1 - t_0} - z_i \right) \leq 0.$$

Diese Ungleichung zeigt, dass $\frac{f(t_1) - f(t_0)}{t_1 - t_0}$ ebenfalls zu $\overline{H_{a,z}^-}$ gehört. Aufgrund der Folgerung 2 ist $\overline{\text{conv}}(E)$ der Durchschnitt aller E enthaltenden abgeschlossenen Halbräume. Folglich gilt $\frac{f(t_1) - f(t_0)}{t_1 - t_0} \in \overline{\text{conv}}(E)$. \square

Man kann durch Beispiele für $N = 2$ zeigen, dass $\frac{f(t_1) - f(t_0)}{t_1 - t_0}$ im Allgemeinen nicht gleich einem der Werte $f'(t)$, $t_0 < t < t_1$, ist. Man muss deshalb mit weniger zufrieden sein. Ist $|f'(t)| \leq r$ für $t_0 < t < t_1$, so liegt E und damit auch $\overline{\text{conv}}(E)$ in der abgeschlossenen Kugel $\overline{B_r(0)}$. Es ist also $\left| \frac{f(t_1) - f(t_0)}{t_1 - t_0} \right| \leq r$, d. h., der Betrag des Differenzenquotienten lässt sich durch die Beträge der Ableitungen in den Zwischenpunkten abschätzen.

Zum Abschluss des Abschnitts über konvexe Mengen werden wir den Begriff Polyeder klären. Eine Teilmenge von \mathbb{R}^N nennt man **polyedrisch**, wenn sie der Durchschnitt von endlich vielen abgeschlossenen Halbräumen ist. Unter einem **Polyeder** in \mathbb{R}^N versteht man eine beschränkte polyedrische Teilmenge von \mathbb{R}^N .

2 Konvexe Funktionen

Nach der Betrachtung der konvexen Mengen folgt die Beschreibung von konvexen Funktionen im N -dimensionalen Raum.

In diesem Abschnitt bezeichnet D eine nicht leere konvexe Menge in \mathbb{R}^N .

Definition 5 Für eine Funktion $f : D \rightarrow \mathbb{R}$ nennt man die Menge

$$\text{epi}(f) := \{(x, s) \in \mathbb{R}^N \times \mathbb{R}; x \in D, f(x) \leq s\}$$

den **Epigraphen** von f . Man nennt die Funktion f **konvex**, wenn ihr Epigraph eine konvexe Teilmenge von $\mathbb{R}^N \times \mathbb{R} = \mathbb{R}^{N+1}$ ist.

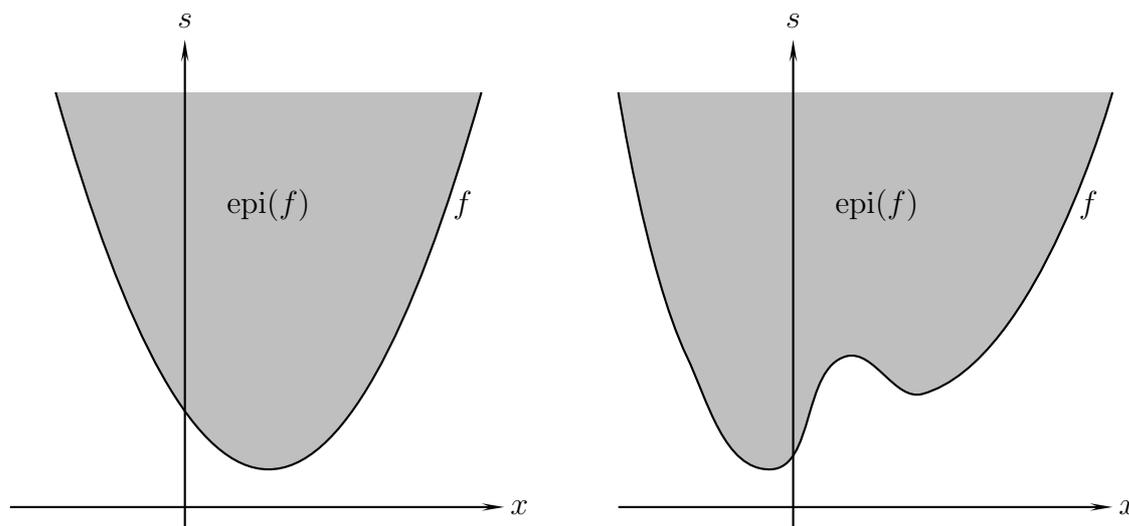


Abbildung 6: Epigraphen einer konvexen und einer nicht konvexen Funktion

Bemerkung 1 Eine Funktion $f : D \rightarrow \mathbb{R}$ ist genau dann konvex, wenn sie die *Jensensche Ungleichung* erfüllt:

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \text{ für alle } x, y \in D, t \in [0, 1]. \quad (9)$$

Begründung:

1. Es sei $f : D \rightarrow \mathbb{R}$ konvex. Für $x, y \in D$ gehören die Punkte $(x, f(x))$ und $(y, f(y))$ zur Menge $\text{epi}(f)$. Da $\text{epi}(f)$ konvex ist, gehört auch die Verbindungsstrecke dieser Punkte zu $\text{epi}(f)$. Die Punkte der Verbindungsstrecke sind

$$(1-t)(x, f(x)) + t(y, f(y)) = ((1-t)x + ty, (1-t)f(x) + tf(y)), \quad 0 \leq t \leq 1.$$

Die Zugehörigkeit dieser Punkte zu $\text{epi}(f)$ besagt, dass die Beziehung (9) gilt.

2. Die Funktion $f : D \rightarrow \mathbb{R}$ erfülle die Bedingung (9). Ferner seien (x, r) und (y, s) beliebige Punkte von $\text{epi}(f)$. Für $t \in [0, 1]$ gilt dann auf Grund von (9)

$$f((1-t)x + ty) \leq (1-t)f(x) + tf(y) \leq (1-t)r + ts.$$

Daher gehören die Punkte

$$((1-t)x + ty, (1-t)r + ts) = (1-t)(x, r) + t(y, s), \quad 0 \leq t \leq 1,$$

zu $\text{epi}(f)$. Da dies die Punkte der Verbindungsstrecke von (x, r) und (y, s) sind, ist $\text{epi}(f)$ konvex, also f eine konvexe Funktion.

Beispiel 1. Es sei $D = \mathbb{R}^N$ und $g(x) := a \cdot x + b$ für $a \in \mathbb{R}^N$, $b \in \mathbb{R}$. Solche Funktionen heißen **affin**. Für $t \in [0, 1]$ und $x, y \in D$ gilt

$$g((1-t)x + ty) = a \cdot ((1-t)x + ty) + b = (1-t)(a \cdot x + b) + t(a \cdot y + b) = (1-t)g(x) + tg(y).$$

Folglich genügt g der Bedingung (9), d. h., g ist konvex. Die Graphen von affinen Funktionen bilden Hyperebenen in \mathbb{R}^{N+1} . Dies bedeutet, dass die Epigraphen abgeschlossene Halbräume sind.

Beispiel 2. Ist $D = \mathbb{R}^N$ und $f(x) := |x|$, so ist f konvex, denn es ist

$$f((1-t)x + ty) = |(1-t)x + ty| \leq (1-t)|x| + t|y| = (1-t)f(x) + tf(y) \quad \text{für } x, y \in \mathbb{R}^N, \quad t \in [0, 1].$$

Definition 6 Ein Element $a \in \mathbb{R}^N$ heißt **Subgradient** einer konvexen Funktion $f : D \rightarrow \mathbb{R}$ im Punkt $x^* \in D$, wenn die durch

$$g(x) := a \cdot (x - x^*) + f(x^*) \quad \text{für } x \in \mathbb{R}^N \tag{10}$$

definierte affine Funktion g die Eigenschaft $g(x) \leq f(x)$ für alle $x \in D$ besitzt. Ist a Subgradient von f in x^* , so nennt man diese affine Funktion g **Stützfunktion** von f in x^* (siehe Abbildung 7 unten). Die Menge aller Subgradienten von f im Punkt x^* bezeichnet man mit $\partial f(x^*)$, und man nennt diese Menge das **Subdifferential** von f in x^* . Dabei ist es möglich, dass $\partial f(x^*) = \emptyset$ ist.

Bemerkung 2 Für eine konvexe Funktion f ist $a \in \partial f(x^*)$ genau dann, wenn die Beziehung

$$a \cdot (x - x^*) \leq f(x) - f(x^*) \quad \text{für alle } x \in D \tag{11}$$

gilt. Die Formel (11) ist für $g(x) := a \cdot (x - x^*) + f(x^*)$ gleichwertig zur Ungleichung $g(x) \leq f(x)$ für alle $x \in D$.

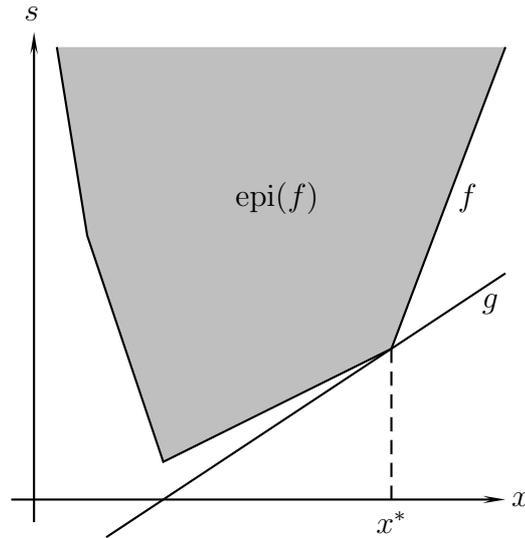


Abbildung 7: Stützfunktion g einer konvexen Funktion f im Punkt x^*

Satz 8 *Es sei $N = 1$, d. h., D sei ein Intervall in \mathbb{R} . Ferner sei $f : D \rightarrow \mathbb{R}$ eine konvexe Funktion, die in dem inneren Punkt x^* von D differenzierbar ist. Dann ist $f'(x^*)$ der einzige Subgradient von f im Punkt x^* .*

Beweis:

1. Nach Bemerkung 2 ist $a \in \partial f(x^*)$, wenn die Beziehung

$$a(x - x^*) \leq f(x) - f(x^*) \quad \text{für alle } x \in D$$

gilt. Da x^* innerer Punkt von D ist, gehört $x = x^* + h$ für h mit hinreichend kleinen Beträgen zu D . Folglich ist $ah \leq f(x^* + h) - f(x^*)$ für hinreichend kleine $|h|$. Für $h > 0$ ergibt sich nach Division durch h und Ausführung des Grenzübergangs $h \rightarrow 0$ die Ungleichung $a \leq f'(x^*)$. Für $h < 0$ ergibt sich analog $a \geq f'(x^*)$. Folglich kann nur $f'(x^*)$ Subgradient von f in x^* sein.

2. Wir zeigen, dass $f'(x^*)$ tatsächlich Subgradient ist. Es sei $x^* + h \in D$. Aufgrund der Konvexität von f folgt aus der Beziehung $x^* + th = (1 - t)x^* + t(x^* + h)$ für $t \in [0, 1]$ die Ungleichung

$$f(x^* + th) \leq (1 - t)f(x^*) + tf(x^* + h).$$

Für $t \in]0, 1]$ ist daher

$$\frac{1}{t}(f(x^* + th) - f(x^*)) \leq f(x^* + h) - f(x^*).$$

Das ist gleichwertig mit

$$\frac{1}{th}(f(x^* + th) - f(x^*))h \leq f(x^* + h) - f(x^*).$$

Für $t \rightarrow 0$ und damit auch $th \rightarrow 0$ ergibt sich

$$f'(x^*)h \leq f(x^* + h) - f(x^*). \quad (12)$$

Nach Bemerkung 2 zeigt diese Ungleichung, dass $f'(x^*) \in \partial f(x^*)$ ist. \square

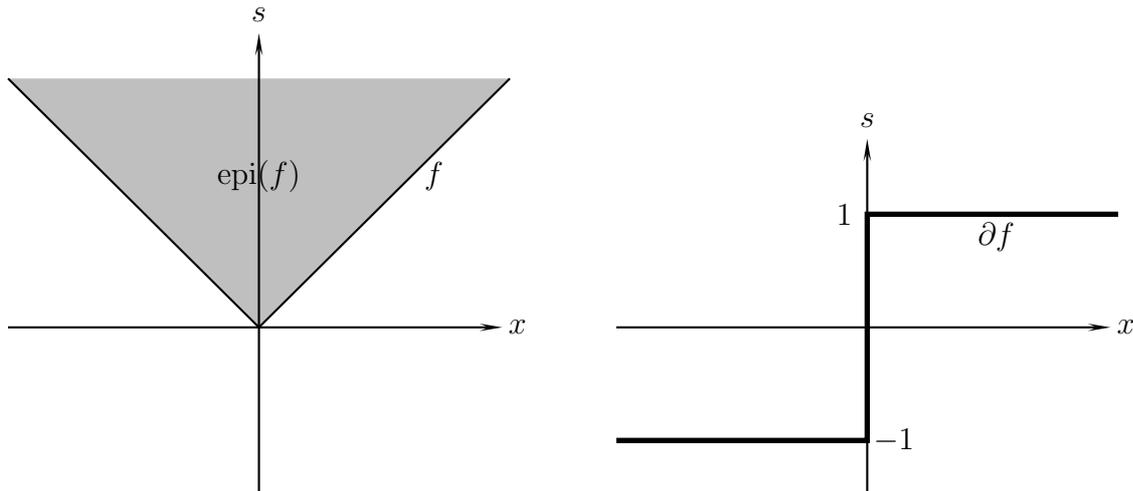


Abbildung 8: Epigraph und Subdifferential der Betragsfunktion

Satz 8 macht deutlich, dass man bei konvexen Funktionen das Subdifferential als eine Verallgemeinerung der Ableitung ansehen kann. Als weiteres Beispiel für $N = 1$ ist in Abbildung 8 Epigraph und Subdifferential der durch $f(x) := |x|$ für $x \in \mathbb{R}$ definierten Betragsfunktion f dargestellt. Es ist $\partial f(0) = [-1, 1]$, weil genau diejenigen Geraden durch den Nullpunkt Stützfunktionen für f liefern, deren Anstieg im Intervall $[-1, 1]$ liegt.

Wir kehren nun wieder zum Fall beliebiger Dimensionen N zurück.

Für eine konvexe Funktion $f : D \rightarrow \mathbb{R}$ interpretiert man ∂f als eine „mehrdeutige Abbildung“ von D in \mathbb{R}^N , d. h., als eine Vorschrift A , die jedem Punkt $x \in D$ eine (möglicherweise leere) Teilmenge $A(x)$ von \mathbb{R}^N zuordnet. Mehrdeutige Abbildungen sind gewöhnungsbedürftig. Das Arbeiten mit mehrdeutigen Abbildungen gewährt mehr Freiheit, als man sie beim Arbeiten allein mit eindeutigen Abbildungen hat.

Satz 9 *Es sei $f : D \rightarrow \mathbb{R}$ konvex. Dann und nur dann, wenn $0 \in \partial f(x^*)$ ist, nimmt die Funktion f im Punkt x^* ihr Minimum an.*

Beweis:

Nach Bemerkung 2 ist $0 \in \partial f(x^*)$ genau dann, wenn $0 \cdot (x - x^*) \leq f(x) - f(x^*)$ ist für alle $x \in D$. Das ist offensichtlich genau dann der Fall, wenn $f(x^*) = \min_{x \in D} f(x)$ ist. \square

Beispiel 3. Es sei $f(x) := p \cdot x + q$ für $x \in D$ und Parameter $p \in \mathbb{R}^N$, $q \in \mathbb{R}$. Für $x^* \in D$ ist

$$\begin{aligned} a \in \partial f(x^*) &\iff a \cdot (x - x^*) \leq f(x) - f(x^*) = p \cdot (x - x^*) \text{ für alle } x \in D \\ &\iff (a - p) \cdot (x - x^*) \leq 0 \text{ für alle } x \in D. \end{aligned}$$

Insbesondere ist $0 \in \partial f(x^*)$ und damit x^* Minimalstelle von f auf D genau dann, wenn

$$-p \cdot (x - x^*) \leq 0 \text{ für alle } x \in D \tag{13}$$

gilt.

Ist D zusätzlich beschränkt und abgeschlossen, so existiert (mindestens) ein $x^* \in D$ derart, dass $f(x^*) = d := \inf_{x \in D} f(x)$ ist. Das zeigt man wie folgt: Man wählt eine Folge (x_n)

aus D mit der Eigenschaft $\lim_{n \rightarrow \infty} f(x_n) = d$. Dann besitzt die Folge (x_n) einen gegen einen Punkt $x^* \in D$ konvergente Teilfolge (x_{n_j}) (Satz von Bolzano-Weierstraß), und es gilt $f(x^*) = \lim_{j \rightarrow \infty} f(x_{n_j}) = d$.

Wir zeigen nun, dass die Minimierung affiner Funktionen auf konvexen Mengen, wie sie in Beispiel 3 beschrieben wird, auch für Optimierungsaufgaben der Ökonomie von Interesse ist.

Es seien E_1, \dots, E_m Erzeuger einer Ware (beispielsweise Molkereien, die Milch erzeugen) und G_1, \dots, G_n Geschäfte einer Ladenkette, in denen die Ware zum Verkauf angeboten wird. Die Kapazitäten der Erzeuger seien c_1, \dots, c_m Wareneinheiten pro Zeiteinheit (beispielsweise Liter/Tag). Der Bedarf von Geschäft G_j betrage b_j Wareneinheiten pro Zeiteinheit. Die Kosten für den Transport einer Wareneinheit vom Erzeuger E_i zum Geschäft G_j seien p_{ij} . Werden x_{ij} Wareneinheiten von E_i nach G_j transportiert, so sind die Gesamtkosten also $\sum_{i=1}^m \sum_{j=1}^n p_{ij} x_{ij}$. Es liegt im Interesse der Ladenkette, diesen Wert zu minimieren.

Dabei sind folgende Nebenbedingungen zu erfüllen:

1. $x_{ij} \geq 0$ für $i = 1, \dots, m, j = 1, \dots, n$: Es gibt keine negativen Warenmengen.
2. $\sum_{i=1}^m x_{ij} = b_j, j = 1, \dots, n$: Der Bedarf der Geschäfte soll gedeckt werden.
3. $\sum_{j=1}^n x_{ij} \leq c_i, i = 1, \dots, m$: Es kann nicht mehr abtransportiert als erzeugt werden.

Man fasst $x = (x_{ij})$ als Element von \mathbb{R}^{mn} auf. Durch die Nebenbedingungen 1 - 3 wird in \mathbb{R}^{mn} ein Polyeder D definiert. Es handelt sich nämlich um den Durchschnitt endlich vieler abgeschlossener Halbräume. (Jede lineare Gleichung ist zu zwei Ungleichungen äquivalent, und jede lineare Ungleichung definiert einen abgeschlossenen Halbraum.) Für $x \in D$ und für $i = 1, \dots, m, j = 1, \dots, n$ ist $0 \leq x_{ij} \leq b_i$. Daher ist D in \mathbb{R}^{mn} beschränkt. Es kann sein, dass D leer ist. Das ist der Fall, wenn die Gesamtkapazität nicht ausreicht, um den Gesamtbedarf zu befriedigen. Wir nehmen an, dass $D \neq \emptyset$ ist. Die Aufgabe besteht darin, die durch die Vorschrift $f(x) := p \cdot x := \sum_{i=1}^m \sum_{j=1}^n p_{ij} x_{ij}$ für $x \in \mathbb{R}^{mn}$ definierte lineare Funktion

f auf dem Polyeder D zu minimieren. Wie wir im Zusammenhang mit Beispiel 3 gesehen haben, ist dieses Problem lösbar. Auf die praktisch wichtige Frage nach einem Algorithmus zur Bestimmung einer Lösung sind wir aus Zeitgründen nicht eingegangen.

Schlusswort

Die Theorie der konvexen Mengen und der konvexen Funktionen ist umfangreicher, als unser Text zeigen kann. Obwohl die räumliche Vorstellungskraft für den N -dimensionalen Raum nicht ausreicht, ist es möglich Rechenregeln und Definitionen vom dreidimensionalen Raum auf höherdimensionale Räume zu übertragen. Dadurch sind wir in die Lage versetzt worden, hoch komplexe Optimierungsaufgaben zu erfassen und als lösbar nachzuweisen.

Mehr zu dem Thema findet man in Literatur über *Konvexe Analysis* ¹⁾.

¹⁾ z. B.: "Convex Analysis" von R. Tyrrell Rockafellar, Princeton University Press, 1970

Gelöste und ungelöste Probleme in der Zahlentheorie

Teilnehmer:

Kristin Emmrich	Heinrich-Hertz-Oberschule
Loryn Fechner	Herder-Oberschule
Felix Neumann	Georg-Forster-Oberschule
Alexandra Stepanova	Herder-Oberschule
Roman Stolzenburg	Herder-Oberschule
Walter Unger	Georg-Forster-Oberschule

Gruppenleiter:

Jürg Kramer	Humboldt-Universität zu Berlin, Mitglied im DFG-Forschungszentrum MATHEON
Anna v. Pippich	Humboldt-Universität zu Berlin

Als *pythagoreisches Zahlentripel* (x, y, z) bezeichnet man drei positive natürliche Zahlen x, y, z , die der Gleichung $x^2 + y^2 = z^2$ genügen. Geometrisch können x, y, z als Seiten eines rechtwinkligen Dreiecks aufgefasst werden. In unserem Sommerschul-Kurs werden wir zunächst der Frage nachgehen, ob es endlich oder unendlich viele pythagoreische Zahlentripel gibt; dazu werden wir für pythagoreische Zahlentripel eine allgemeine Formel herleiten.

Das berühmte *Fermat-Problem* stellt in Verallgemeinerung der Suche nach pythagoreischen Zahlentripeln die Frage, ob es positive natürliche Zahlen x, y, z gibt, die der Gleichung

$$x^n + y^n = z^n$$

für einen natürlichen Exponenten $n > 2$ genügen. Die *Vermutung von Fermat* aus dem Jahr 1637, dass es keine solchen natürlichen Zahlen gibt, wurde erst 1995 von Andrew Wiles vollständig bewiesen. In unserem Kurs werden wir die Vermutung von Fermat für den Exponenten $n = 4$ beweisen.

Es zeigt sich, dass sich hinter dem Satz von Wiles eine noch viel tiefer liegende Vermutung der Zahlentheorie verbirgt: die sogenannte *abc-Vermutung*. Diese Vermutung ist bis heute unbewiesen und stellt ein sehr aktives Forschungsfeld dar. Zunächst werden wir die *abc-Vermutung* für Polynome in einer Variablen (*Satz von Mason*) formulieren und beweisen. Als Folgerung werden wir das Analogon der Fermat-Vermutung für Polynome beweisen. Motiviert durch die Analogie zwischen den ganzen Zahlen und Polynomen in einer Variablen können wir nun die *abc-Vermutung* für ganze Zahlen präzise formulieren und die Grenzen ihrer Gültigkeit ausloten. Schließlich zeigen wir, dass die *abc-Vermutung* den Satz von Wiles für große Exponenten n impliziert.

1 Geschichte

Der Satz von Pythagoras: Dieser Satz wurde bereits im Altertum bei der Landvermessung und beim Bau angewendet. Pythagoras hat den Satz wahrscheinlich auf einer Reise nach Ägypten gefunden. Man nimmt an, dass der Satz schon lange vor Pythagoras bekannt war. Die Gleichung $4961^2 + 6480^2 = 8161^2$ wurde bereits ca. 1500 v. Chr. in babylonischen Keilschrifttexten erwähnt.

Weitere Rechenvorschriften zur Erzeugung pythagoreischer Tripel finden sich im Buch des griechischen Mathematikers Diophant (ca. 250 v. Chr., in Alexandria). Deswegen werden Probleme, bei denen es um ganzzahlige Lösungen von Gleichungen geht, auch „diophantische Probleme“ genannt.

Pierre de Fermat (17. Jh.), Hobby-Mathematiker, schrieb seine mathematischen Entdeckungen auf Buchrändern und in Briefen an bedeutende Gelehrten seiner Zeit, unter anderem die Vermutung, dass es keine positiven natürliche Zahlen x, y, z gibt, die der Gleichung $x^n + y^n = z^n$ für einen natürlichen Exponenten $n > 2$ genügen. Über dieses Problem schrieb er auf einem Buchrand: „Für diese Behauptung habe ich einen wahrhaft wunderbaren Beweis gefunden, aber dieser Rand ist zu schmal, um ihn zu fassen.“

Kummer (19. Jh.) bewies die Fermat-Vermutung für alle Primzahlexponenten kleiner als 100 (mit Ausnahme zweier Fälle). Bis etwa 1990 gelang es, die Gültigkeit der Fermat-Vermutung für alle Exponenten n kleiner als 4 Millionen zu beweisen.

Am 23. Juni 1993 wurde ein Beweis für die Fermat-Vermutung von Wiles veröffentlicht, jedoch wurde bereits 1994 die Unvollständigkeit dieses Beweises gezeigt. Es dauerte über ein Jahr, bis schließlich Andrew Wiles mit einem korrigierten Beweis an die Öffentlichkeit trat. Am 27. Juni 1997 wurde der Beweis als richtig anerkannt und so bekam Andrew Wiles den Wolfskehl-Preis (der von Wolfskehl für die Erbringung des Beweises ausgesetzt worden war).

2 Pythagoreische Tripel

Unser Ziel ist es, alle primitiven pythagoreischen Tripel (x, y, z) , die die Gleichung $x^2 + y^2 = z^2$ ganzzahlig erfüllen, zu berechnen. Dazu benötigen wir Grundlagen über die Rationalität von grafischen Objekten.

Definition 2.1. *Ein rationaler Punkt ist ein Punkt, der aus rationalen Koordinaten besteht.*

Definition 2.2. *Eine rationale Gerade ist eine Gerade der Form $aX + bY + c = 0$, bei der die Koeffizienten rational sind.*

Lemma 2.3. *Eine Gerade, die zwei rationale Punkte verbindet, ist eine rationale Gerade.*

Beweis. Seien $P_1(x_1, y_1)$ und $P_2(x_2, y_2)$ rationale Punkte. Die Steigung m der Geraden $Y = mX + t$ durch diese Punkte berechnet sich zu

$$m = \frac{y_2 - y_1}{x_2 - x_1}.$$

Da diese Operationen den Ring der rationalen Zahlen nicht verlassen, ist auch m rational. Die Verschiebung t in Y-Richtung ist

$$t = Y - mX = Y - \frac{y_2 - y_1}{x_2 - x_1} \cdot X.$$

Einsetzen von P_1 ergibt

$$t = y_1 - \frac{y_2 - y_1}{x_2 - x_1} \cdot x_1.$$

Somit ist auch t rational und es handelt sich also um eine rationale Gleichung. \square

Lemma 2.4. *Der Schnittpunkt zweier rationaler Geraden ist ein rationaler Punkt.*

Beweis. Ohne der Beschränkung der Allgemeinheit seien die beiden rationalen Geraden von der Form

$$\begin{aligned} f(X) = Y &= -\frac{a_1}{b_1}X - \frac{c_1}{b_1}, \\ g(X) = Y &= -\frac{a_2}{b_2}X - \frac{c_2}{b_2}. \end{aligned}$$

Durch Gleichsetzen erhält man

$$\begin{aligned} \frac{a_1}{b_1}X + \frac{c_1}{b_1} &= \frac{a_2}{b_2}X + \frac{c_2}{b_2} \iff \\ X &= \left(\frac{c_2}{b_2} - \frac{c_1}{b_1}\right) \cdot \left(\frac{a_1}{b_1} - \frac{a_2}{b_2}\right)^{-1}. \end{aligned}$$

Daraus folgt, dass die X -Koordinate des Schnittpunkts rational ist. Durch Einsetzen in f erhält man

$$Y = -\frac{a_1}{b_1} \cdot \left(\frac{c_2}{b_2} - \frac{c_1}{b_1}\right) \cdot \left(\frac{a_1}{b_1} - \frac{a_2}{b_2}\right)^{-1} - \frac{c_1}{b_1}.$$

Also ist auch die Y -Koordinate rational. Der Schnittpunkt ist folglich rational, denn seine beiden Koordinaten sind rational. \square

Definition 2.5. Ein rationaler Kegelschnitt ist gegeben durch eine Gleichung der Form

$$aX^2 + bXY + cY^2 + dX + eY + f = 0,$$

wobei die Koeffizienten rational sind.

Lemma 2.6. Die Schnittpunkte eines rationalen Kegelschnitts mit einer rationalen Geraden sind nicht immer rational. Aber, wenn einer von beiden rational ist, dann ist es auch der andere.

Beweis. Gegeben seien die Gleichungen

$$aX^2 + bXY + cY^2 + dX + eY + f = 0 \tag{2.1}$$

und

$$Y = mX + t \tag{2.2}$$

mit rationalen Koeffizienten. Durch Einsetzen von (2.2) in (2.1) erhält man

$$aX^2 + bX(mX + t) + c(mX + t)^2 + dX + e(mX + t) + f = 0.$$

Durch Umformen erhält man eine Gleichung der Form

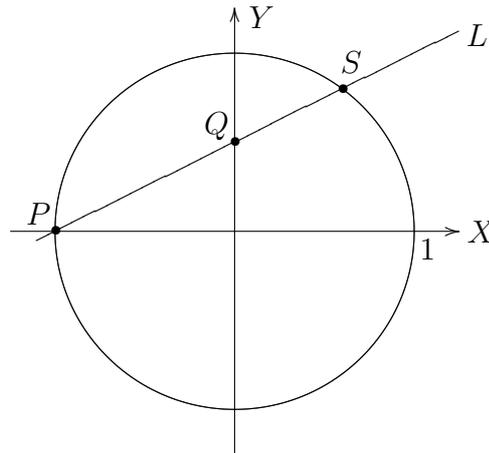
$$AX^2 + BX + C = 0$$

mit $A, B, C \in \mathbb{Q}$. Die Lösungen dieser Gleichung sind gegeben durch

$$x_{1/2} = -\frac{B}{2A} \pm \sqrt{\frac{B^2}{4A^2} - \frac{C}{A}}.$$

Da die auftretende Quadratwurzel nicht zwangsläufig eine rationale Zahl ist, sind auch $x_{1/2}$ nicht immer rational. Angenommen x_1 sei rational. Dann ist folglich die zweite Lösung x_2 auch rational. Und wenn die X -Koordinate rational ist, dann ist es nach Einsetzen in (2.2) auch die Y -Koordinate. Das heißt, wenn der eine Schnittpunkt rational ist, dann ist es auch der zweite. \square

Gegeben sei der Einheitskreis $X^2 + Y^2 = 1$. Wir nehmen einen rationalen Punkt auf dem Kreis, zum Beispiel $P(-1, 0)$, und einen rationalen Punkt $Q(0, t)$ auf der Y -Achse. Nach 2.3 ist die Gerade L , die diese beiden Punkte verbindet, eine rationale Gerade. Diese Gerade schneidet unseren Kreis in zwei Punkten, nämlich in P und in $S(x, y)$. Da der Kreis ein spezieller rationaler Kegelschnitt ist, folgt nach 2.6, dass S ebenfalls rational ist, weil Q ja rational ist.



Die Gleichung der rationalen Geraden L ist $L(X) = t(1 + X)$. Durch Umformen und Gleichsetzen von der Kreisgleichung und $L(X)$ erhält man

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

Wie man sieht, ist die erste Lösung der Gleichung $X = -1$, weil der Punkt P auf dem Kreis und auf L liegt. Um nun auf die zweite Lösung zu kommen, teilen wir durch $(X + 1)$ und erhalten nach Umformen

$$1 - X = t^2(1 + X) \iff X(1 + t^2) = 1 - t^2 \iff X = \frac{1 - t^2}{1 + t^2}.$$

Durch Einsetzen in $L(X)$ erhält man schließlich

$$Y = \frac{2t}{1 + t^2}.$$

Diese Formeln für X und Y bieten uns letztendlich die Möglichkeit, die primitiven pythagoreischen Tripel zu charakterisieren.

Definition 2.7. Ein primitives pythagoreisches Tripel (x, y, z) besteht aus positiven natürlichen Zahlen x, y, z , die teilerfremd sind und der Gleichung

$$x^2 + y^2 = z^2$$

genügen.

Bemerkung. Wenn (x, y, z) ein primitives pythagoreisches Tripel ist, so sind x, y, z sogar paarweise teilerfremd. Dies sieht man wie folgt ein: Angenommen, y und z hätten einen gemeinsamen Teiler. Dann hätte auch x wegen $x^2 = z^2 - y^2$ diesen Teiler. Dies widerspricht aber der Teilerfremdheit von x, y, z .

Theorem 2.8. *Es sei (x, y, z) ein primitives pythagoreisches Tripel. Dann existieren positive natürliche Zahlen m, n verschiedener Parität mit $n > m$ und $\text{ggT}(m, n) = 1$ derart, dass*

$$x = n^2 - m^2, \quad y = 2mn, \quad z = m^2 + n^2$$

gilt.

Beweis. Es sei (x, y, z) ein beliebiges primitives pythagoreisches Tripel. Damit setzen wir

$$x' = \frac{x}{z}, \quad y' = \frac{y}{z}.$$

Der Punkt $S(x', y')$ definiert dann einen Punkt mit rationalen Koordinaten auf dem Einheitskreis $X^2 + Y^2 = 1$, denn es besteht die Äquivalenz

$$x'^2 + y'^2 = 1 \iff x^2 + y^2 = z^2.$$

Da x und y teilerfremd sind, können x und y nicht beide gerade sein. Die Zahlen x und y können aus folgendem Grund auch nicht beide ungerade sein: Angenommen x und y seien beide ungerade, d.h. $x \equiv y \equiv 1 \pmod{2}$. Dann muss z gerade sein, d.h. $z \equiv 0 \pmod{2}$, und wir erhalten den Widerspruch

$$2 \equiv x^2 + y^2 = z^2 \equiv 0 \pmod{4}.$$

Die Zahlen x und y müssen also unterschiedliche Parität haben. Wir können im folgenden ohne Beschränkung der Allgemeinheit annehmen, dass x ungerade und y gerade ist. Da nun $S(x', y')$ ein rationaler Punkt auf dem Einheitskreis ist, gibt es eine rationale Zahl t , so dass

$$x' = \frac{1 - t^2}{1 + t^2}, \quad y' = \frac{2t}{1 + t^2} \tag{2.3}$$

gilt. Da t eine rationale Zahl ist, kann diese als Quotient zweier teilerfremder natürlicher Zahlen m und n dargestellt werden, d.h.

$$t = \frac{m}{n}. \tag{2.4}$$

Wenn wir jetzt (2.4) in (2.3) einsetzen, erhalten wir

$$\frac{x}{z} = x' = \frac{1 - m^2/n^2}{1 + m^2/n^2} = \frac{n^2 - m^2}{n^2 + m^2}, \tag{2.5}$$

$$\frac{y}{z} = y' = \frac{2 \cdot m/n}{1 + m^2/n^2} = \frac{2mn}{n^2 + m^2}. \tag{2.6}$$

Da $(n^2 - m^2)$ und $(n^2 + m^2)$ aus (2.5) sowie $2mn$ und $(n^2 + m^2)$ aus (2.6) nicht zwangsläufig teilerfremd sein müssen, gibt es ein $\lambda \in \mathbb{N}$, so dass

$$\lambda x = n^2 - m^2, \quad \lambda y = 2mn, \quad \lambda z = n^2 + m^2$$

gilt. Da λ die Zahlen $(n^2 + m^2)$ und $(n^2 - m^2)$ teilt, muss es auch die Summe $2n^2$ und die Differenz $2m^2$ teilen. Da m und n aber teilerfremd sind, folgt daraus, dass λ die Zahl 2 teilen muss. Die natürliche Zahl λ kann somit nur die Werte 1 oder 2 annehmen. Im folgenden zeigen wir indirekt, dass $\lambda = 1$ sein muss. Dazu nehmen wir an, dass $\lambda = 2$ gilt. Da x nach obiger Annahme ungerade ist, folgt somit mit $\lambda x = n^2 - m^2$ die Kongruenz $n^2 - m^2 = \lambda x \equiv 2 \pmod{4}$. Da jedoch n^2 und m^2 entweder kongruent zu 0 oder 1 mod 4 sind, ist das nicht möglich. Also gilt $\lambda = 1$.

Somit kommen wir auf die Bildungsvorschrift

$$x = n^2 - m^2, \quad y = 2mn, \quad z = n^2 + m^2, \tag{2.7}$$

womit der Satz vollständig bewiesen ist. \square

3 Die Vermutung von Fermat

Fermat hat das Problem der Suche nach pythagoreischen Tripeln dahingehend verallgemeinert, dass er die Vermutung aufgestellt hat, dass keine von Null verschiedenen ganzzahligen Lösungen x, y, z für die Gleichung $x^n + y^n = z^n$ für ein $n \in \mathbb{N}$, $n \geq 3$, existieren.

Einen Beweis seiner Vermutung hat Fermat uns allerdings nicht hinterlassen, sondern nur in einer Randnotiz geschrieben, er habe einen wahrhaft wunderbaren Beweis gefunden, leider reiche der Platz hier aber nicht aus, um ihn aufzuschreiben.

Theorem 3.1. *Die Vermutung von Fermat ist für $n = 4$ korrekt.*

Beweis. Im Gegensatz zur Behauptung nehmen wir an, dass teilerfremde $x, y, z \in \mathbb{N}_{>0}$ mit der Eigenschaft $x^4 + y^4 = z^4$ existieren. Wir haben diese Annahme zu einem Widerspruch zu führen.

Indem wir $w = z^2$ schreiben, erhalten wir die Gleichung

$$(x^2)^2 + (y^2)^2 = w^2$$

und stellen fest, dass (x^2, y^2, w) ein primitives pythagoreisches Tripel ist. Gemäß Satz 2.8 existieren positive natürliche Zahlen p, q verschiedener Parität mit $p > q$ und $\text{ggT}(p, q) = 1$ derart, dass

$$x^2 = 2pq, \quad y^2 = p^2 - q^2, \quad w = p^2 + q^2$$

gilt. Umstellen der zweiten Gleichung ergibt

$$q^2 + y^2 = p^2.$$

Da p und q teilerfremd sind, ist (q, y, p) ein primitives pythagoreisches Tripel, wobei p ungerade ist. Eine erneute Anwendung von Satz 2.8 liefert positive natürliche Zahlen a, b verschiedener Parität mit $a > b$ und $\text{ggT}(a, b) = 1$ derart, dass

$$q = 2ab, \quad y = a^2 - b^2, \quad p = a^2 + b^2$$

gilt. Einsetzen von $q = 2ab$ und $p = a^2 + b^2$ in die Gleichung $x^2 = 2pq$ liefert

$$x^2 = 2pq = 2(a^2 + b^2) \cdot 2ab = 4ab(a^2 + b^2) \iff \left(\frac{x}{2}\right)^2 = ab(a^2 + b^2).$$

Damit ist das Produkt $ab(a^2 + b^2)$ also ein Quadrat. Da die Zahlen $a, b, (a^2 + b^2)$ paarweise teilerfremd sind, müssen a, b und $(a^2 + b^2)$ selbst Quadratzahlen sein, d.h.

$$a = x_1^2, \quad b = y_1^2, \quad a^2 + b^2 = w_1^2$$

mit positiven natürlichen Zahlen x_1, y_1, w_1 . Ausgehend von unserer Annahme haben wir damit positive natürliche Zahlen x_1, y_1, w_1 gefunden, die der Gleichung

$$x_1^4 + y_1^4 = w_1^2$$

genügen. Überdies stellen wir fest:

$$\begin{aligned} x_1 < x, & \quad \text{denn} \quad x_1^2 = a < 2ab = q < 2pq = x^2, \\ y_1 < y, & \quad \text{denn} \quad y_1^2 = b < a + b < ((a + b)(a - b))^2 = y^2, \\ w_1 < w, & \quad \text{denn} \quad w_1^2 = a^2 + b^2 = p < p^2 + q^2 = w < w^2. \end{aligned}$$

Damit greift nun die *Methode des unendlichen Abstiegs*: Ausgehend von den positiven natürlichen Zahlen x, y, w mit $x^4 + y^4 = w^2$ haben wir kleinere positive natürliche Zahlen x_1, y_1, w_1 mit $x_1^4 + y_1^4 = w_1^2$ konstruiert. Diesen Prozess kann man offensichtlich fortsetzen und eine unendliche Folge von Tripeln (x_j, y_j, w_j) von immer kleiner werdenden positiven natürlichen Zahlen konstruieren, die der Gleichung $x_j^4 + y_j^4 = w_j^2$ genügen. Dies ist ersichtlich ein Widerspruch, da es nur endlich viele positive natürliche Zahlen gibt, die kleiner als x sind. \square

4 Der Satz von Mason

In diesem Kapitel beweisen wir den Satz von Mason, der auch als *abc*-Vermutung für komplexe Polynome in einer Variablen bezeichnet wird. Dazu stellen wir zunächst einige Begriffe bereit und beweisen zwei Lemmas.

Es sei $f \in \mathbb{C}[X]$ ein Polynom mit komplexen Koeffizienten, d.h.

$$f(X) = c_n X^n + c_{n-1} X^{n-1} + \dots + c_1 X + c_0,$$

wobei $c_1, \dots, c_n \in \mathbb{C}$ und $c_n \neq 0$ gilt. Der Grad von f ist hierbei durch $\deg(f) = n$ gegeben. Der *Fundamentalsatz der Algebra* besagt, dass wir $f \in \mathbb{C}[X]$ schreiben können als

$$f(X) = c \cdot (X - \alpha_1)^{a_1} \cdot \dots \cdot (X - \alpha_m)^{a_m}, \quad (4.1)$$

wobei $c \in \mathbb{C}$ mit $c \neq 0$ ist, $\alpha_1, \dots, \alpha_m \in \mathbb{C}$ die verschiedenen Nullstellen von f sind und $a_1, \dots, a_m \in \mathbb{N}_{>0}$ die jeweiligen Vielfachheiten der Nullstellen bezeichnen. Damit erhalten wir für den Grad von f die Gleichheit

$$\deg(f) = n = a_1 + \dots + a_m = \sum_{j=1}^m a_j.$$

Definition 4.1. Für $f \in \mathbb{C}[X]$ mit der Zerlegung (4.1) definieren wir

$$n_0(f) := \text{Anzahl der verschiedenen Nullstellen von } f = m.$$

Bemerkung. Es gilt $n_0(f) \leq \deg(f)$. Die Differenz der beiden Zahlen $n_0(f)$ und $\deg(f)$ kann allerdings sehr groß sein. Zum Beispiel ist für $f(X) := (X - 1)^{10101}$ der Grad $\deg(f) = 10101$ und $n_0(f) = 1$.

Lemma 4.2. Sei $f \in \mathbb{C}[X]$ und $f \neq 0$. Dann gilt

$$n_0(f) = \deg(f) - \deg(\text{ggT}(f, f')). \quad (4.2)$$

Beweis. Sei $f \in \mathbb{C}[X]$ und $f \neq 0$. Wir schreiben f wie in (4.1) als

$$f(X) = c \cdot (X - \alpha_1)^{a_1} \cdot \dots \cdot (X - \alpha_m)^{a_m}, \quad (4.3)$$

wobei $c \in \mathbb{C}$ mit $c \neq 0$ ist und $\alpha_1, \dots, \alpha_m \in \mathbb{C}$ die verschiedenen Nullstellen mit den Vielfachheiten $a_1, \dots, a_m \in \mathbb{N}_{>0}$ bezeichnen. Ableiten von f unter Verwendung der Produktregel liefert

$$\begin{aligned} f'(X) &= c \cdot a_1 \cdot (X - \alpha_1)^{a_1-1} \cdot (X - \alpha_2)^{a_2} \cdot \dots \cdot (X - \alpha_m)^{a_m} \\ &\quad + c \cdot (X - \alpha_1)^{a_1} \cdot \left[(X - \alpha_2)^{a_2} \cdot \dots \cdot (X - \alpha_m)^{a_m} \right]'. \end{aligned} \quad (4.4)$$

Schreiben wir $(X - \alpha_1)^{a_1} = (X - \alpha_1)^{a_1-1} \cdot (X - \alpha_1)$, so zeigt die Gleichheit (4.1), dass das Polynom $(X - \alpha_1)^{a_1-1}$ das Polynom f teilt. Aus der Gleichheit (4.4) folgt, dass das Polynom $(X - \alpha_1)^{a_1-1}$ auch das Polynom f' teilt. Daraus folgt

$$(X - \alpha_1)^{a_1-1} \mid \text{ggT}(f, f').$$

Analog erhalten wir für jeden Wert $j \in \{2, \dots, m\}$, dass

$$(X - \alpha_j)^{a_j-1} \mid \text{ggT}(f, f')$$

gilt. Insgesamt erhalten wir damit die Aussage

$$(X - \alpha_1)^{a_1-1} \cdot \dots \cdot (X - \alpha_m)^{a_m-1} \mid \text{ggT}(f, f').$$

Aus Gradgründen folgt nun

$$\text{ggT}(f, f') = (X - \alpha_1)^{a_1-1} \cdot \dots \cdot (X - \alpha_m)^{a_m-1}.$$

Daraus ergibt sich

$$\deg(\text{ggT}(f, f')) = \sum_{j=1}^m (a_j - 1) = \sum_{j=1}^m a_j - m = \deg(f) - n_0(f), \quad (4.5)$$

woraus die Behauptung unmittelbar folgt. \square

Lemma 4.3. *Seien $f, g \in \mathbb{C}[X]$. Dann ist*

$$n_0(f \cdot g) \leq n_0(f) + n_0(g),$$

wobei die Gleichheit genau dann gilt, wenn f und g teilerfremd sind.

Beweis. Diese Behauptung ergibt sich unmittelbar aus dem Fundamentalsatz der Algebra. \square

Theorem 4.4 (Satz von Mason). *Für alle nicht-konstanten, teilerfremden Polynome $a, b, c \in \mathbb{C}[X]$ mit $a + b = c$ gilt die Ungleichung*

$$\max(\deg(a), \deg(b), \deg(c)) \leq n_0(a \cdot b \cdot c) - 1.$$

Beweis. Aus der Gleichheit

$$a + b = c \quad (4.6)$$

folgt durch Ableiten die Gleichheit

$$a' + b' = c'. \quad (4.7)$$

Multiplizieren von (4.6) mit a' und von (4.7) mit a liefert die Gleichheiten

$$a'a = a'c - a'b \quad \text{und} \quad aa' = ac' - ab'.$$

Nach Gleichsetzen und Umformen erhalten wir

$$ab' - a'b = ac' - a'c. \quad (4.8)$$

Nun teilen $\text{ggT}(a, a')$ und $\text{ggT}(b, b')$ offensichtlich die linke Seite von (4.8). Da jetzt $\text{ggT}(c, c')$ entsprechend die rechte Seite der Gleichung (4.8) teilt, muss $\text{ggT}(c, c')$ auch die linke Seite von (4.8), d.h. $ab' - a'b$, teilen. Insgesamt haben wir also die Teilbarkeiten

$$\begin{aligned} \text{ggT}(a, a') &| (ab' - a'b), \\ \text{ggT}(b, b') &| (ab' - a'b), \\ \text{ggT}(c, c') &| (ab' - a'b). \end{aligned}$$

Aufgrund der paarweisen Teilerfremdheit der Polynome a, b, c ergibt sich somit die Teilbarkeitsbeziehung

$$\text{ggT}(a, a') \cdot \text{ggT}(b, b') \cdot \text{ggT}(c, c') | (ab' - a'b). \quad (4.9)$$

Aus folgendem Grund ist die Differenz $ab' - a'b$ ungleich dem Nullpolynom: Wäre nämlich $ab' - a'b = 0$, so könnte man die Gleichung schreiben als $a'b = ab'$. Daraus würde $a|a'b$ folgen. Da aber a und b nach Voraussetzung teilerfremd sind, ergäbe sich daraus $a|a'$. Da dies jedoch nur für $a' = 0$ möglich ist, müsste a ein konstantes Polynom sein. Dies widerspricht aber unserer Voraussetzung.

Da nun $ab' - a'b$ nicht das Nullpolynom ist und aufgrund von (4.9) von dem Produktpolynom $\text{ggT}(a, a') \cdot \text{ggT}(b, b') \cdot \text{ggT}(c, c')$ geteilt wird, besteht die Ungleichung

$$\deg(\text{ggT}(a, a') \cdot \text{ggT}(b, b') \cdot \text{ggT}(c, c')) \leq \deg(ab' - a'b). \quad (4.10)$$

Unter Berücksichtigung der Eigenschaft $\deg(f \cdot g) = \deg(f) + \deg(g)$ ($f, g \in \mathbb{C}[X]$) erhalten wir aus (4.10) die Ungleichung

$$\deg(\text{ggT}(a, a')) + \deg(\text{ggT}(b, b')) + \deg(\text{ggT}(c, c')) \leq \deg(ab' - a'b). \quad (4.11)$$

Aufgrund der weiteren Gradeigenschaft

$$\deg(f + g) \leq \max(\deg(f), \deg(g)) \quad (f, g \in \mathbb{C}[X])$$

erhalten wir für die rechte Seite von (4.11) die Abschätzung

$$\deg(ab' - a'b) \leq \max(\deg(ab'), \deg(a'b)).$$

Die Grade $\deg(ab')$ und $\deg(a'b)$ berechnen wir unter Berücksichtigung der bekannten Formel $\deg(f') = \deg(f) - 1$ ($f \in \mathbb{C}[X]$, $\deg(f) > 0$) zu

$$\begin{aligned}\deg(ab') &= \deg(a) + \deg(b') = \deg(a) + \deg(b) - 1 \\ &= \deg(a) - 1 + \deg(b) = \deg(a') + \deg(b) = \deg(a'b),\end{aligned}$$

d.h.

$$\deg(ab' - a'b) \leq \max(\deg(ab'), \deg(a'b)) = \deg(a) + \deg(b) - 1. \quad (4.12)$$

Die Ungleichungen (4.11) und (4.12) zusammengenommen ergeben

$$\deg(\text{ggT}(a, a')) + \deg(\text{ggT}(b, b')) + \deg(\text{ggT}(c, c')) \leq \deg(a) + \deg(b) - 1.$$

Addition von $\deg(c)$ zur vorhergehenden Ungleichung liefert nach Umstellen

$$\begin{aligned}\deg(c) &\leq -1 + \deg(a) - \deg(\text{ggT}(a, a')) \\ &\quad + \deg(b) - \deg(\text{ggT}(b, b')) \\ &\quad + \deg(c) - \deg(\text{ggT}(c, c')).\end{aligned}$$

Nach dreimaliger Anwendung von Lemma 4.2 ergibt sich damit die Ungleichung

$$\deg(c) \leq n_0(a) + n_0(b) + n_0(c) - 1.$$

Da nun a, b, c paarweise teilerfremd sind, ergibt sich mit Hilfe von Lemma 4.3

$$n_0(a) + n_0(b) + n_0(c) = n_0(a \cdot b \cdot c),$$

und damit die Ungleichung

$$\deg(c) \leq n_0(a \cdot b \cdot c) - 1.$$

Aus Symmetriegründen beweist man in analoger Weise

$$\deg(a) \leq n_0(a \cdot b \cdot c) - 1 \quad \text{und} \quad \deg(b) \leq n_0(a \cdot b \cdot c) - 1,$$

woraus insgesamt die Behauptung des Satzes von Mason folgt. □

5 Der Satz von Fermat für Polynome

Mit Hilfe des Satzes von Mason (Theorem 4.4) können wir das Analogon der Fermat-Vermutung für Polynome beweisen.

Theorem 5.1 (Fermat für Polynome). *Für $n \geq 3$ gibt es keine nicht-konstanten, teilerfremden Polynome $f, g, h \in \mathbb{C}[X]$ mit*

$$f^n + g^n = h^n.$$

Beweis. Wir nehmen an, dass $f, g, h \in \mathbb{C}[X]$ nicht-konstante, teilerfremde Polynome mit $f^n + g^n = h^n$ sind. Da dann auch die Polynome f^n, g^n, h^n nicht-konstant und teilerfremd sind, gilt nach dem Satz von Mason (Theorem 4.4) die Ungleichung

$$\max(\deg(f^n), \deg(g^n), \deg(h^n)) \leq n_0(f^n \cdot g^n \cdot h^n) - 1. \quad (5.1)$$

Wegen

$$n_0(f^n \cdot g^n \cdot h^n) = n_0(f \cdot g \cdot h) \leq \deg(f \cdot g \cdot h) = \deg(f) + \deg(g) + \deg(h)$$

und unter Berücksichtigung der Eigenschaft $\deg(q^n) = n \cdot \deg(q)$ ($q \in \mathbb{C}[X]$) erhalten wir aus (5.1) die Ungleichung

$$n \cdot \max(\deg(f), \deg(g), \deg(h)) \leq \deg(f) + \deg(g) + \deg(h) - 1. \quad (5.2)$$

Da weiter $\deg(q) \leq \max(\deg(f), \deg(g), \deg(h))$ für $q = f, g, h$ gilt, liefert (5.2) die Abschätzungen

$$n \cdot \deg(f) \leq \deg(f) + \deg(g) + \deg(h) - 1 \quad (5.3)$$

$$n \cdot \deg(g) \leq \deg(f) + \deg(g) + \deg(h) - 1 \quad (5.4)$$

$$n \cdot \deg(h) \leq \deg(f) + \deg(g) + \deg(h) - 1. \quad (5.5)$$

Addition von (5.3), (5.4) und (5.5) ergibt nun die Ungleichung

$$\begin{aligned} n \cdot (\deg(f) + \deg(g) + \deg(h)) &\leq 3 \cdot (\deg(f) + \deg(g) + \deg(h)) - 3 \\ \iff (n - 3) \cdot (\deg(f) + \deg(g) + \deg(h)) &\leq -3. \end{aligned}$$

Dies ist für $n \geq 3$ offensichtlich ein Widerspruch, da dann die linke Seite der letzten Ungleichung nicht negativ ist. Damit ist der Satz von Fermat für Polynome bewiesen. \square

6 Die Analogie zwischen den ganzen Zahlen und Polynomen

In diesem Abschnitt vergleichen wir Eigenschaften der ganzen Zahlen mit Eigenschaften von Polynomen mit komplexen Koeffizienten. Insbesondere werden wir das Analogon des Grades eines Polynoms für ganze Zahlen entdecken und schließlich die Funktion $n_0(f)$ (siehe Definition 4.1) auch für ganze Zahlen definieren.

Eigenschaft	\mathbb{Z}	$\mathbb{C}[X]$
Elemente	ganze Zahlen a	Polynome $f = f(X)$ mit komplexen Koeffizienten
Teilbarkeit	$a \in \mathbb{Z}$ $b \in \mathbb{Z}$ $a b \iff \exists c \in \mathbb{Z} : b = a \cdot c$	$f = a_n X^n + a_{n-1} X^{n-1} + \dots + a_0$ $g = b_m X^m + b_{m-1} X^{m-1} + \dots + b_0$ $f g \iff \exists h \in \mathbb{C}[X] : g = f \cdot h$
Division mit Rest	$a, b \in \mathbb{Z}, b \neq 0$ $\implies \exists q, r \in \mathbb{Z},$ $0 \leq r < b :$ $a = q \cdot b + r$	$f, g \in \mathbb{C}[X], g \neq 0$ $\implies \exists q, r \in \mathbb{C}[X],$ $r = 0$ oder $\deg(r) < \deg(g):$ $f = q \cdot g + r$
Analogie z.B.:	$a \in \mathbb{Z}, a \neq 0$ $\ln(a)$ $\ln(a \cdot b) = \ln(a) + \ln(b)$	$f \in \mathbb{C}[X], f \neq 0$ $\deg(f)$ $\deg(f \cdot g) = \deg(f) + \deg(g)$
Division mit Rest (revisited)	$a, b \in \mathbb{Z}, b \neq 0$ $\implies \exists q, r \in \mathbb{Z},$ $r = 0$ oder $0 \leq \ln(r) < \ln(b):$ $a = q \cdot b + r$	$f, g \in \mathbb{C}[X], g \neq 0$ $\implies \exists q, r \in \mathbb{C}[X],$ $r = 0$ oder $\deg(r) < \deg(g):$ $f = q \cdot g + r$

ggT(.,.)	$a, b \in \mathbb{Z}$, nicht beide Null: ggT(a, b) definiert; Berechnung mit Hilfe des Euklidischen Algorithmus	$f, g \in \mathbb{C}[X]$, nicht beide Null: ggT(f, g) definiert; Berechnung mit Hilfe des Euklidischen Algorithmus
Einheiten	$\{-1, +1\}$	konst. Polynome $\neq 0$ kurz: $\mathbb{C}^\times := \mathbb{C} \setminus \{0\} \subseteq \mathbb{C}[X]$
Primzahlen, bzw. Prim- polynome	2, 3, 5, 7, 11, 13, ..., 229, ...	$X - \alpha, \alpha \in \mathbb{C}$
Fundamen- talsatz	der Arithmetik $a \in \mathbb{Z}, \ln(a) > 0 (\Leftrightarrow a > 1) :$ $a = \pm p_1^{a_1} \cdot \dots \cdot p_m^{a_m},$ wobei p_1, \dots, p_m paarweise verschiedene Primzahlen, $a_1, \dots, a_m \in \mathbb{N}_{>0}$	der Algebra $f \in \mathbb{C}[X], \deg(f) > 0:$ $f(X) = c \cdot (X - \alpha_1)^{a_1} \cdot \dots \cdot (X - \alpha_m)^{a_m},$ wobei $c \in \mathbb{C}^\times, \alpha_1, \dots, \alpha_m \in \mathbb{C}$ paar- weise verschieden, $a_1, \dots, a_m \in \mathbb{N}_{>0}$
$n_0(\cdot)$	$a \in \mathbb{Z}, \ln(a) > 0 (\Leftrightarrow a > 1):$ $n_0(a) = \sum_{p a} \ln(p)$	$f \in \mathbb{C}[X], \deg(f) > 0:$ $n_0(f) = \#$ versch. Nst. von f $= \#$ versch. Primpolynome p von f $= \sum_{p f} 1 = \sum_{p f} \deg(p)$

1. *Bemerkung (Division mit Rest/Division mit Rest (revisited)):*

Es ist leicht zu erkennen, dass eine Analogie zwischen dem Betrag $|\cdot|$ einer ganzen Zahl und dem Grad $\deg(\cdot)$ eines komplexen Polynoms besteht. Diese ist jedoch noch nicht quantifiziert. Deshalb ersetzen wir den Betrag $|a|$ der ganzen Zahl a durch den natürlichen Logarithmus ihres Betrages $\ln(|a|)$. Eine Begründung hierfür liegt darin, dass sich der Logarithmus auf einem Produkt analog zum Grad additiv verhält: es gilt $\ln(|a \cdot b|) = \ln(|a|) + \ln(|b|)$ und $\deg(f \cdot g) = \deg(f) + \deg(g)$. Weiter gilt zum Beispiel $\ln(|a|) = -\infty$ für $a = 0$ und $\deg(f) := -\infty$ für $f = 0$. Damit können wir die Division mit Rest für ganze Zahlen und komplexe Polynome symmetrisch formulieren (siehe Division mit Rest (revisited)).

2. *Bemerkung (Größter gemeinsamer Teiler ggT):*

Zur Veranschaulichung geben wir ein Beispiel, wie man mit Hilfe des Euklidischen Algorithmus den größten gemeinsamen Teiler $\text{ggT}(a, b)$ zweier ganzer Zahlen a und b berechnen kann: Sei $a = 29$ und $b = 17$. Dann gilt:

$$29 = 1 \cdot 17 + 12;$$

$$17 = 1 \cdot 12 + 5;$$

$$12 = 2 \cdot 5 + 2;$$

$$5 = 2 \cdot 2 + 1;$$

$$2 = 2 \cdot 1 + 0.$$

Der $\text{ggT}(a, b)$ ist der letzte Rest ungleich Null. Es gilt also $\text{ggT}(29, 17) = 1$. In analoger Weise berechnet man unter Verwendung der Polynomdivision mit Hilfe des Euklidischen Algorithmus den $\text{ggT}(f, g)$ zweier Polynome $f, g \in \mathbb{C}[X]$.

3. *Bemerkung (Einheiten):*

Einheiten sind alle Elemente einer Menge, die bezüglich der gegebenen Verknüpfung ein inverses Element innerhalb der Menge besitzen. Die Einheiten von \mathbb{Z} bzgl. der Multiplikation sind $+1$ und -1 . Die Einheiten von $\mathbb{C}[X]$ bzgl. der Multiplikation sind die konstanten Polynome $c \in \mathbb{C}^\times$.

4. *Bemerkung (Primzahlen/Primpolynome):*

Primzahlen bzw. Primpolynome sind jene Elemente einer Menge, die nur durch sich selbst, ihre Gegenzahl und die Einheiten teilbar sind. Die Primpolynome sind somit genau die unzerlegbaren bzw. irreduziblen Polynome. Es gibt unendlich viele Primzahlen und Primpolynome.

5. *Bemerkung (Definition von $n_0(\cdot)$):*

Für ein Polynom f ist $n_0(f)$ definiert als die Anzahl der verschiedenen Nullstellen von f . Im Raum $\mathbb{C}[X]$ der Polynome mit komplexen Koeffizienten bedeutet dies, dass $n_0(f)$ gleich der Anzahl der verschiedenen Primpolynome von f ist. Daher wirkt zunächst eine Übertragung von n_0 für $a \in \mathbb{Z}$ mit $n_0(a) :=$ Anzahl der verschiedenen Primfaktoren von a als logisch. Im Raum $\mathbb{R}[X]$ kann jedoch die Anzahl der verschiedenen Nullstellen von der Anzahl der Primpolynome abweichen. Beispielsweise hat $f(X) = X^2 + 3$ zwei komplexe Nullstellen, lässt sich jedoch in $\mathbb{R}[X]$ nicht weiter faktorisieren. Glücklicherweise können wir $n_0(f)$ auch als $n_0(f) = \sum_{p|f} \deg(p)$ darstellen, wobei p die Primpolynome von f durchläuft. Analog definieren wir nun unter Berücksichtigung der 1. *Bemerkung* die Funktion $n_0(a)$ wie folgt.

Definition 6.1. Für $a \in \mathbb{Z}, \ln(|a|) > 0$ definieren wir

$$n_0(a) := \sum_{\substack{p|a \\ p \text{ Primzahl}}} \ln(p).$$

7 Die abc -Vermutung

Mithilfe der Analogie aus Kapitel 6 formulieren wir das Analogon des Satzes von Mason (Theorem 4.4) für ganze Zahlen.

Vermutung 7.1. Für alle teilerfremden, ganzen Zahlen $a, b, c \in \mathbb{Z} \setminus \{0\}$ mit $a + b = c$ gilt die Ungleichung

$$\max(\ln(|a|), \ln(|b|), \ln(|c|)) \leq n_0(a \cdot b \cdot c) - 1.$$

Definition 7.2. Für $a \in \mathbb{Z}, \ln(|a|) > 0$ definieren wir

$$N_0(a) := \prod_{\substack{p|a \\ p \text{ Primzahl}}} p.$$

Wir formen die Vermutung 7.1 wie folgt äquivalent um.

$$\begin{aligned} \max(\ln(|a|), \ln(|b|), \ln(|c|)) &\leq n_0(a \cdot b \cdot c) - 1 \\ \Leftrightarrow \ln(\max(|a|, |b|, |c|)) &\leq \sum_{p|abc} \ln(p) - 1 \\ \Leftrightarrow \ln(\max(|a|, |b|, |c|)) &\leq \ln\left(\prod_{p|abc} p\right) - 1 \\ \Leftrightarrow \max(|a|, |b|, |c|) &\leq K \cdot \prod_{p|abc} p = K \cdot N_0(a \cdot b \cdot c), \end{aligned} \quad (7.1)$$

wobei $K := \exp(1)^{-1}$ ist und wir die Äquivalenzen unter Berücksichtigung der Definition 6.1, des Logarithmusgesetzes $\ln(a) + \ln(b) = \ln(a \cdot b)$ ($a, b \in \mathbb{N}_{>0}$) und anschließendem Exponentieren erhalten haben.

Wir überprüfen die Vermutung 7.1 mit folgendem Beispiel: Für $n \in \mathbb{N}_{>0}$ seien

$$a := 3^{2^n}, \quad b := -1, \quad c := 3^{2^n} - 1.$$

Zunächst versuchen wir, das Produkt der Primteiler $N_0(a \cdot b \cdot c)$ von $a \cdot b \cdot c$ in (7.1) abzuschätzen. Die Zahl a hat als einzigen Primfaktor die 3, die Zahl b hat keinen und die Zahl c hat mindestens den Primfaktor 2, da c eine gerade Zahl ist. Wir zeigen nun mit Hilfe von vollständiger Induktion, dass sogar

$$2^n | (3^{2^n}) - 1$$

für alle $n \in \mathbb{N}$ gilt. Der Induktionsanfang $n = 0$ ist klar: es gilt $2^0 = 1 | 2 = 3 - 1 = 3^{2^0} - 1$. Wir nehmen nun an, dass $2^n | (3^{2^n} - 1)$ für ein festes $n \in \mathbb{N}$ gilt. Wir zeigen $2^{n+1} | (3^{2^{n+1}} - 1)$ wie folgt:

$$3^{2^{n+1}} - 1 = 3^{2^n \cdot 2} - 1 = (3^{2^n})^2 - 1^2 = (3^{2^n} - 1) \cdot (3^{2^n} + 1).$$

Der Faktor $3^{2^n} - 1$ ist nach Induktionsvoraussetzung durch 2^n teilbar. Der Faktor $3^{2^n} + 1$ ist durch 2 teilbar, da $3^{2^n} + 1$ eine gerade Zahl ist. Damit ist das Produkt der beiden Faktoren $(3^{2^n} - 1) \cdot (3^{2^n} + 1) = c$ durch 2^{n+1} teilbar, womit die Behauptung bewiesen ist. Der Faktor $c/2^n$ besitzt möglicherweise weitere Primfaktoren, welche in ihrer Vielfachheit eingehen, jedoch nicht den Primfaktor 3. Damit erhalten wir die folgende Abschätzung:

$$\begin{aligned} N_0(a \cdot b \cdot c) &\leq 3 \cdot 1 \cdot 2 \cdot \frac{c}{2^n} = \frac{6}{2^n} (3^{2^n} - 1) \\ &\leq \frac{6}{2^n} \cdot 3^{2^n}. \end{aligned} \tag{7.2}$$

Da $a + b = c$ mit a, b, c teilerfremd und trivialerweise $|a| = 3^{2^n} \leq \max(|a|, |b|, |c|)$ gilt, erhalten wir aus der Vermutung 7.1 die Ungleichung

$$\begin{aligned} 3^{2^n} &\leq K \cdot N_0(a \cdot b \cdot c) \\ &\leq K \cdot \frac{6}{2^n} \cdot 3^{2^n}, \end{aligned}$$

wobei die zweite Ungleichung aufgrund von (7.2) besteht. Division durch 3^{2^n} liefert die Abschätzung

$$1 \leq K \cdot \frac{6}{2^n}.$$

Für $n \rightarrow \infty$ gilt jedoch $6/2^n \rightarrow 0$. Demnach müsste $1 \leq 0$ gelten. Das ist ein Widerspruch! Die Vermutung 7.1 ist also in dieser Form falsch.

Die richtige Formulierung der bis heute noch unbewiesenen *abc*-Vermutung lautet:

Vermutung 7.3 (Die *abc*-Vermutung). *Zu $\epsilon > 0$ existiert eine Konstante $K(\epsilon)$, so dass für alle teilerfremden, ganzen Zahlen $a, b, c \in \mathbb{Z} \setminus \{0\}$ mit $a + b = c$ die Ungleichung*

$$\max(|a|, |b|, |c|) \leq K(\epsilon) \cdot N_0(a \cdot b \cdot c)^{1+\epsilon}$$

gilt.

8 Die *abc*-Vermutung impliziert die Fermat-Vermutung

Wäre die *abc*-Vermutung bewiesen, so erhielte man einen sehr kurzen Beweis für die Richtigkeit der Vermutung von Fermat für große Exponenten $n \in \mathbb{N}$.

Proposition 8.1. *Aus der *abc*-Vermutung 7.3 folgt, dass die Vermutung von Fermat für große Exponenten $n \in \mathbb{N}$ korrekt ist.*

Beweis. Seien $x, y, z \in \mathbb{N}_{>0}$ teilerfremd mit

$$x^n + y^n = z^n.$$

Wir setzen $a := x^n, b := y^n$ und $c := z^n$. Dann gilt

$$N_0(x^n \cdot y^n \cdot z^n) = N_0(x \cdot y \cdot z) \leq x \cdot y \cdot z.$$

Diese Abschätzung liefert zusammen mit *abc*-Vermutung 7.3 die Ungleichungen

$$x^n \leq K(\epsilon) \cdot (x \cdot y \cdot z)^{1+\epsilon}, \quad (8.1)$$

$$y^n \leq K(\epsilon) \cdot (x \cdot y \cdot z)^{1+\epsilon}, \quad (8.2)$$

$$z^n \leq K(\epsilon) \cdot (x \cdot y \cdot z)^{1+\epsilon}. \quad (8.3)$$

Multiplikation von (8.1), (8.2) und (8.3) ergibt die Ungleichung

$$(xyz)^n \leq K(\epsilon)^3 \cdot (x \cdot y \cdot z)^{3+3\epsilon}. \quad (8.4)$$

Wir setzen $\epsilon' := 3\epsilon$ und dividieren (8.4) durch $(x \cdot y \cdot z)^{3+\epsilon'}$. Man erhält

$$(x \cdot y \cdot z)^{n-(3+\epsilon')} \leq K(\epsilon'/3)^3.$$

Logarithmieren dieser Abschätzung liefert

$$(n - (3 + \epsilon')) \cdot \ln(x \cdot y \cdot z) \leq 3 \cdot \ln(K(\epsilon'/3)).$$

Da $a, b, c \in \mathbb{N}_{>0}$ und $a + b = c$ ist, gilt $x \cdot y \cdot z \geq 2$ und damit $1/\ln(x \cdot y \cdot z) \leq 1/\ln(2)$.
Damit folgt

$$\begin{aligned} n - (3 + \epsilon') &\leq \frac{3}{\ln(2)} \cdot \ln(K(\epsilon'/3)) \\ \implies n &\leq \frac{3}{\ln(2)} \cdot \ln(K(\epsilon'/3)) + (3 + \epsilon'). \end{aligned}$$

Wir erhalten also eine Schranke für n , so dass sich die Fermatsche Vermutung für alle genügend großen natürlichen Zahlen n als korrekt herausstellt. \square

☺ Schön, dass ihr uns bis hierher gefolgt seid. Leider ist es für uns an der Zeit, Schluss zu machen, aber wir hoffen, es hat euch gefallen und es war klar verständlich. Wir hatten jedenfalls eine Menge Spaß hier und finden es schön, dass noch andere Interesse an der Mathematik zeigen. Sämtliche Rechtschreibfehler sind copyrighted. Missbrauch hat gerichtliche Konsequenzen. Sollte noch etwas unklar sein, schließen wir mit Einsteins Worten: „Man sollte nie aufhören, Fragen zu stellen!“ ☺

Hyperbolische Geometrie

Teilnehmer:

Jakob Hoffmann	Heinrich-Hertz-Oberschule
Luise Jungnickel	Georg-Forster-Oberschule
Tim Klose	Heinrich-Hertz-Oberschule
Victor Kojoukhov	Herder-Oberschule
Sascha Kühl	Georg-Forster-Oberschule
Rami Nagi	Herder-Oberschule
Felix Rehn	Heinrich-Hertz-Oberschule

Gruppenleiter:

Thomas Neukirchner	Humboldt-Universität zu Berlin
--------------------	--------------------------------

Ziel des Kurses war, die hyperbolische Geometrie der Ebene aus Sicht ihrer Isometrien zu erkunden. Als Ausgangspunkt haben wir zunächst die abstands- und orientierungserhaltenden Abbildungen der Euklidischen Ebene untersucht, und gezeigt, dass sie sich als Verknüpfung zweier Spiegelungen an Geraden darstellen lassen. Je nachdem, ob die Geraden parallel sind oder sich schneiden, entsteht als Abbildung eine Verschiebung oder Drehung.

Für das Poincaré-Scheiben-Modell der hyperbolischen Geometrie haben wir Geraden und Spiegelungen an ihnen definiert und daraus die Geometrie von hyperbolischen Kreisen und Dreiecken abgeleitet. Das geometrische Werkzeug dafür bestand im Wesentlichen aus der Inversion am Kreis. Abschließend wurde mit den erarbeiteten Mitteln bewiesen, dass der Flächeninhalt eines sphärischen bzw. hyperbolischen Dreiecks sich aus dessen Winkelsumme bestimmt. Diese Tatsache zeigt sich eindrucksvoll in einigen Bildern von M.C.Escher.

Neben der theoretischen Kursarbeit haben sich die Teilnehmer intensiv mit der dynamischen Geometrie Software *Cinderella* beschäftigt, mit der auch alle Illustrationen des vorliegenden Textes erstellt wurden.

Inhaltsverzeichnis

1	Euklidische Geometrie der Ebene	71
1.1	Parallelen-Axiom	71
1.2	Isometrien	71
2	Sphärische Geometrie	72
2.1	Winkelsumme eines sphärischen Dreiecks	72
3	Inversion am Kreis	74
3.1	Geometrische Konstruktion	75
3.2	Kreisverwandtschaft und Winkeltreue	76
4	Poincaré-Scheiben-Modell der hyperbolischen Geometrie	78
4.1	Ω -Geraden und Parallelen	78
4.2	Spiegelung und Abstandsbegriff	79
5	Steiner's Kreiskette	80
6	Winkelsumme im hyperbolischen Dreieck	82
7	Konstruktionen mit Zirkel und Lineal	85
7.1	Ω -Gerade durch zwei Punkte A und B	85
7.2	Ω -Kreis durch A mit Mittelpunkt M	86
7.3	M.C.Escher	86
8	Schlussbetrachtung	87

1 Euklidische Geometrie der Ebene

1.1 Parallelen-Axiom

Die Euklidische Geometrie befasst sich mit den Eigenschaften von Punkten, Geraden, Kreisen und Winkeln. Ihr axiomatischer Aufbau hat sich über zwei Jahrtausende bewährt, führte aber auch zu einer Frage, auf die lange Zeit keine Antwort gefunden werden konnte: Ist das Parallelenaxiom eine logische Folgerung der anderen Euklidischen Axiome?

Parallelenaxiom der Euklidischen Geometrie: Gegeben seien eine Gerade l und ein Punkt $P \notin l$. Dann existiert genau eine Gerade m durch den Punkt P , die l nicht schneidet. Eine solche Gerade wird als Parallele bezeichnet: $l \parallel m$.

Heutzutage stehen die analytischen Methoden der Geometrie im Vordergrund, da sie sich wesentlich leichter handhaben lassen als eine aus Axiomen synthetisch aufgebaute Geometrie. Eine wesentliche Rolle spielen dabei Isometrien.

1.2 Isometrien

Eine Isometrie ist eine Abbildung der Ebene auf sich selbst, bei der Längen und somit auch Winkel unverändert bleiben. Beispiele für Isometrien sind:

- **Spiegelung:** Die Spiegelung an einer Geraden l bezeichnen wir mit S_l . Die Spiegelung kehrt die Orientierung um.
- **Translation:** Sie ist die Verknüpfung $S_l \circ S_m$ zweier Spiegelungen an parallelen Geraden $l \parallel m$ und entspricht der Verschiebung längs einer Geraden senkrecht zu l, m . Die Translation erhält die Orientierung.
- **Rotation:** Sie ist die Verknüpfung $S_l \circ S_m$ zweier Spiegelungen an nichtparallelen Geraden $l \not\parallel m$ und entspricht der Drehung um deren Schnittpunkt $P = l \cap m$.

Allgemein gilt:

Satz 1.1. *Jede Isometrie der Ebene ist die Verknüpfung von 1, 2 oder 3 Spiegelungen an Geraden. Insbesondere ist eine orientierungserhaltende Isometrie entweder eine Translation oder eine Rotation.*

All das führt uns zu der Frage, wie eine Geometrie beschaffen sein müsste, bei der es keine Parallelen bzw. viele Parallelen gibt?

2 Sphärische Geometrie

Als Modell der sphärischen Geometrie betrachten wir die Oberfläche einer Kugel, auch Sphäre genannt, und definieren:

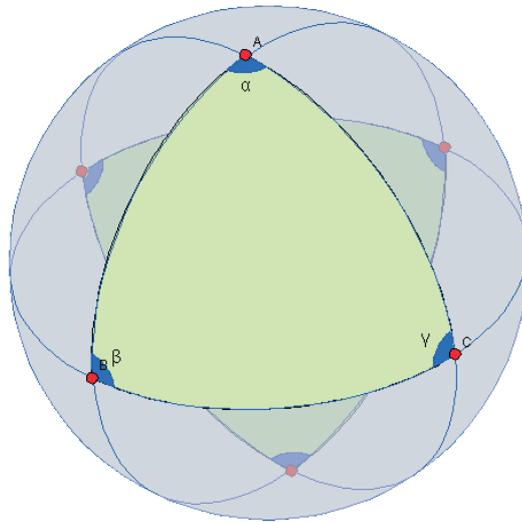
- **S-Geraden** sind Großkreise, d.h. Schnitte der Sphäre mit Ebenen durch den Mittelpunkt.
- **S-Punkte** sind Paare antipodaler, also einander gegenüberliegender Punkte auf einer S-Geraden.
- **S-Winkel** entsprechen den Euklidischen Schnittwinkel von Großkreises, d.h. dem Schnittwinkel der entsprechenden Ebenen.

Anhand dieses Modells wird deutlich, dass sich zwei S-Geraden stets schneiden, es also keine Parallelen geben kann (ein Breitenkreis hat zwar konstanten Abstand zum Äquator, ist aber keine S-Gerade und damit auch kein Kandidat für eine Parallele). Daraus folgt, dass die Verknüpfung zweier Spiegelungen $S_l \circ S_m$ immer eine Rotation ist um den Schnittpunkt $l \cap m$ der beiden S-Geraden l, m . In der sphärischen Geometrie gibt es keine Translationen.

2.1 Winkelsumme eines sphärischen Dreiecks

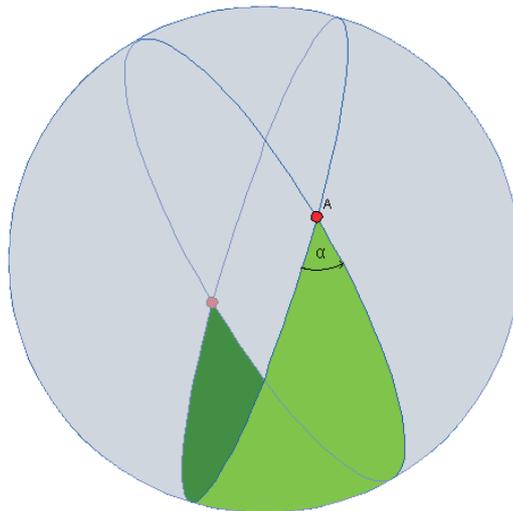
Die Innenwinkelsumme eines Dreiecks in der Ebene beträgt 2π . Das sphärische Dreieck ist hingegen gewölbt und hat daher augenscheinlich eine größere Innenwinkelsumme. Insbesondere ist die Innenwinkelsumme nicht konstant und es stellt sich folgende Frage:

Können wir die Innenwinkelsumme eines sphärischen Dreiecks aus dem Verhältnis seines Flächeninhalts gegenüber dem Flächeninhalt der gesamten Sphäre bestimmen?



A_{Dreieck} und $A_{\text{Sphäre}}$ sind bekannt. $\Rightarrow \alpha + \beta + \gamma = ?$

Eine Sphäre lässt sich mit Hilfe von mehreren S-Geraden durch einen Punkt wie eine Mandarine in Stückchen zerteilen. Diese Stückchen kann man auch als Zweiecke bezeichnen, da sie jeweils zwei gleiche Ecken besitzen, die auf der Sphäre sich genau gegenüber liegen.



Die Fläche eines solchen sphärischen Zweiecks verhält sich zu der Gesamtfläche der Sphäre wie einer seiner beiden Winkel zum Vollwinkel.

$$A_{\text{Zweieck}} = A_{\text{Sphäre}} \cdot \frac{\alpha}{2\pi} \quad (1)$$

Zurück zum sphärischen Dreieck. Das sphärische Dreieck kann als eine Fläche angesehen werden, wo sich drei sphärische Zweiecke überlappen. Somit können wir die Gesamtfläche der Figur, die von diesen drei Zweiecken gebildet wird, bestimmen, indem wir die Flächen der einzelnen sphärischen Zweiecke summieren und zweimal die Fläche des sphärischen Dreiecks abziehen. Schneiden wir diese Figur gedanklich aus der Sphäre aus, so bleibt von der Sphäre eine Figur übrig, die zu unserer Figur kongruent ist. Somit können wir die halbe Fläche der Sphäre mit der Fläche unserer Figur gleichsetzen:

$$\begin{aligned} \frac{A_{\text{Sphäre}}}{2} &\stackrel{(1)}{=} A_{\text{Sphäre}} \cdot \frac{\alpha}{2\pi} + A_{\text{Sphäre}} \cdot \frac{\beta}{2\pi} + A_{\text{Sphäre}} \cdot \frac{\gamma}{2\pi} - 2 \cdot A_{\text{Dreieck}} \\ &= A_{\text{Sphäre}} \cdot \frac{\alpha + \beta + \gamma}{2\pi} - 2 \cdot A_{\text{Dreieck}} \\ \Leftrightarrow \frac{1}{2} &= \frac{\alpha + \beta + \gamma}{2\pi} - 2 \cdot \frac{A_{\text{Dreieck}}}{A_{\text{Sphäre}}} \\ \Leftrightarrow \pi &= \alpha + \beta + \gamma - 4\pi \cdot \frac{A_{\text{Dreieck}}}{A_{\text{Sphäre}}} \end{aligned}$$

und wir erhalten:

Satz 2.1. *Der Flächeninhalt eines sphärischen Dreiecks im Verhältnis zur Gesamtfläche der Sphäre bestimmt sich aus der Innenwinkelsumme des Dreiecks durch folgende Gleichung:*

$$4\pi \cdot \frac{A_{\text{Dreieck}}}{A_{\text{Sphäre}}} = \alpha + \beta + \gamma - \pi$$

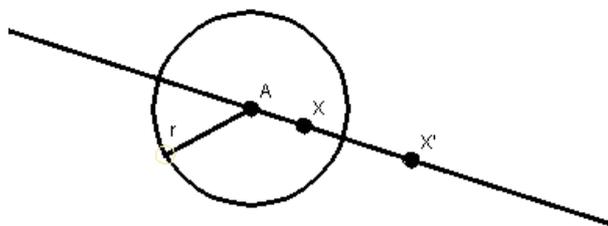
3 Inversion am Kreis

Die Inversion am Kreis ist eines der elementaren Hilfsmittel, um später das Poincaré-Modell der hyperbolischen Ebene einführen zu können. Da diese Operation jedoch nicht so geläufig ist und zudem auch Euklidisch gesehen jede Menge interessanter Eigenschaften hat, widmen wir ihr einen eigenen Abschnitt.

Definition 3.1. Sei Σ ein Kreis um P mit Radius r . Für $X \neq P$ ist der an Σ invertierte Punkt X' charakterisiert durch:

$$|PX| \cdot |PX'| = r^2$$

$$\overrightarrow{PX'} = \lambda \cdot \overrightarrow{PX} \quad \text{für ein } \lambda > 0$$



Die Abbildung

$$F_{\Sigma} : \mathbb{R}^2 \setminus \{P\} \rightarrow \mathbb{R}^2 \setminus \{P\}$$

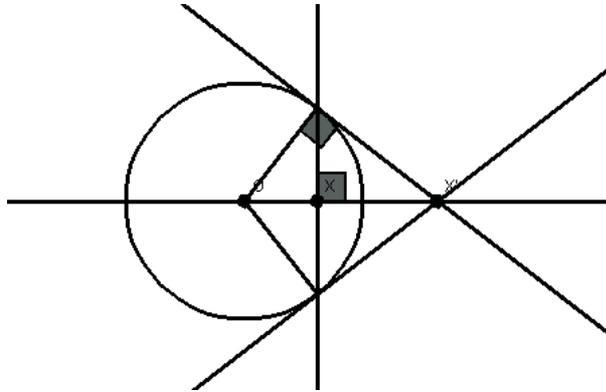
$$X \mapsto X'$$

heißt Inversion am Kreis Σ . Formal setzen wir:

$$F(P) = \infty, \quad F(\infty) = P$$

3.1 Geometrische Konstruktion

Es ist möglich, die Inversion am Kreis zu konstruieren, falls X im Innern des Kreises liegt. Dazu zeichnet man eine Senkrechte zum Strahl \overrightarrow{OX} Durch X . An den Schnittpunkten P der entstandenen Sekante mit der Peripherie legt man Tangenten an den Kreis. Der Schnittpunkt dieser Tangenten ist der gesuchte Bildpunkt X' .



Will man einen Punkt außerhalb des Kreises invertieren, so muss man diese Schritte in umgekehrter Reihenfolge ausführen.

Beweis. Man beachte, dass die Dreiecke OXP und OPX' ähnlich sind. Aus $|OP| = r$ folgt dann:

$$\frac{|OX|}{r} = \frac{r}{|OX'|} \Leftrightarrow |OX| \cdot |OX'| = r^2$$

□

3.2 Kreisverwandtschaft und Winkeltreue

Die Inversion am Kreis besitzt viele nützliche Eigenschaften. Einige davon werden wir in diesem Teilabschnitt vorstellen.

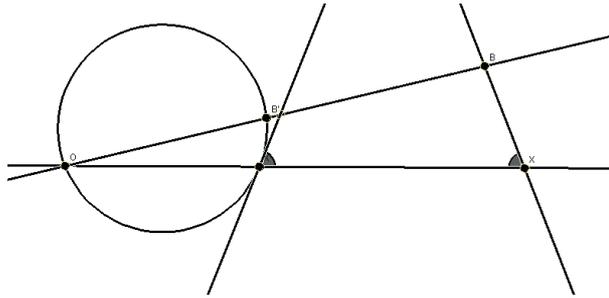
Satz 3.1. *Sei F eine Inversion am Kreis mit Zentrum O . Dann gilt:*

- (i) *F bildet Kreise, die nicht durch O gehen, wieder auf solche ab.*
- (ii) *F bildet Kreise durch O auf Geraden ab (und umgekehrt).*

Dies zeigt man, indem man dem zu invertierenden Kreis ein rechtwinkliges Dreieck einbeschreibt, dessen Hypotenuse also ein Durchmesser des Kreises ist, und zeigt, dass nun die invertierten Dreiecke ähnlich sind. Auf den genauen Beweis verzichten wir.

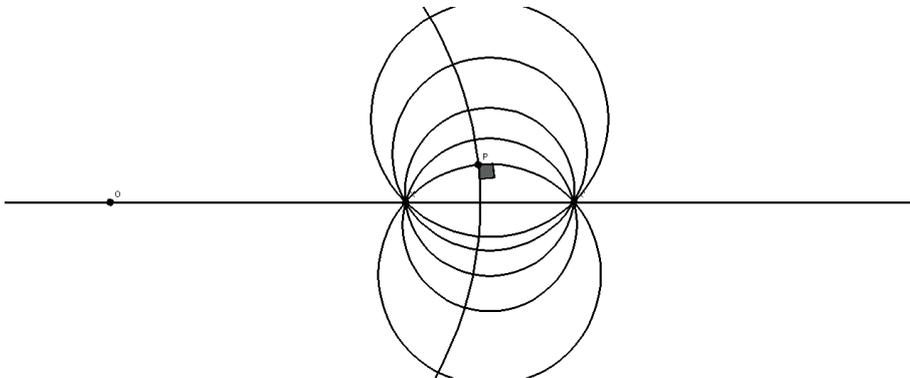
Satz 3.2. *Die Inversion am Kreis erhält Schnittwinkel von Kurven, d.h. die Inversion am Kreis ist winkeltreu.*

Zum Beweis genügt es zu zeigen, dass der Schnittwinkel eines zentralen Strahls \overline{OX} mit einer Geraden \overline{XB} gleich ist dem Schnittwinkel des selben Strahls mit dem Bild der Geraden \overline{XB} unter Inversion, also dem Kreis durch die Punkte $OX'B'$.



Auch hier verzichten wir auf einen ausführlichen Beweis.

Satz 3.3. *Sei $X' = F_\Sigma(X)$, $X \neq X'$. Dann gilt: Jeder Kreis durch X, X' schneidet Σ orthogonal. Und umgekehrt geht jeder Kreis durch X , der Σ orthogonal schneidet auch durch X' .*



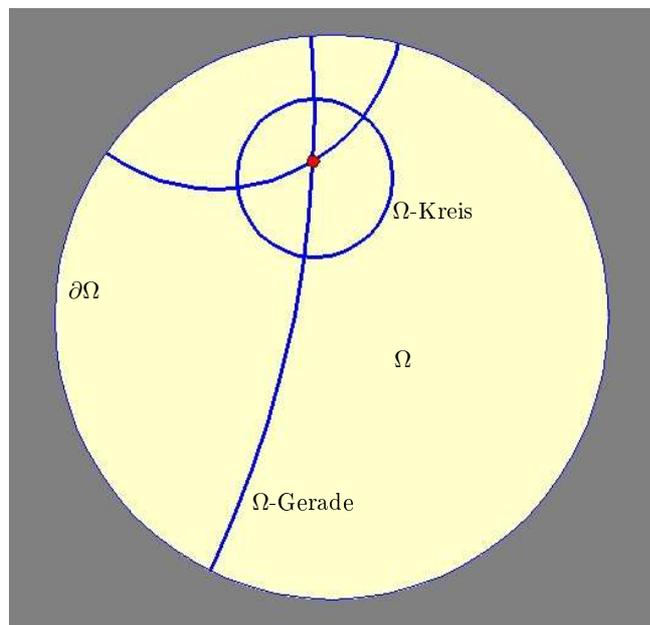
Beweis. Sei k ein Kreis durch X, X' und bezeichne mit $P \in \Sigma \cap k$ einen Schnittpunkt mit dem Inversionskreis. Es gilt:

$$F_\Sigma(X) = X', \quad F_\Sigma(X') = X, \quad F_\Sigma(P) = P$$

Aus Satz 3.1 folgt nun $F_\Sigma(k) = k$ und nach Satz 3.2 vertauschen sich die komplementären Schnittwinkel von k und Σ unter F_Σ und sind somit rechte Winkel. \square

4 Poincaré-Scheiben-Modell der hyperbolischen Geometrie

Da wir nun wissen, was Inversion am Kreis ist, können wir uns an die hyperbolische Geometrie heranwagen. Als Modell betrachten wir eine Kreisscheibe Ω und beschreiben nun, was wir in diesen Modell unter Ω -Punkten, Ω -Geraden, Ω -Winkel etc. verstehen wollen:



Ω	Kreisscheibe
$\partial\Omega$	Kreislinie, die Ω berandet (gehört nicht mit zu Ω)
Ω -Punkte	Punkte in der Kreisscheibe
Ω -Geraden	Kreissegmente, die $\partial\Omega$ senkrecht schneiden
Ω -Winkel	Euklidischer Winkel
Ω -Kreise	Durchmesser \perp Tangenten

4.1 Ω -Geraden und Parallelen

Ω -Geraden teilen grundlegende Eigenschaften mit Geraden der Euklidischen Geometrie, wie hier anhand zweier Beispiele gezeigt werden soll.

(i) **Vor.:** Seien A und B Ω -Punkte.

Beh.: Es existiert genau eine Ω -Gerade g , für die gilt $A, B \in g$.

Bew.: Sei g eine Ω -Gerade durch A und bezeichne mit k den ihr zugrunde liegenden Euklidischen Kreis. Da $k \perp \partial\Omega$, muss nach Satz 3.3 der Kreis k auch durch den Punkt A' gehen, der durch Inversion von A an $\partial\Omega$ entsteht. Liegt nun auch B auf g , so geht der Kreis k durch die drei Punkte A, B, A' und ist damit eindeutig bestimmt.

(ii) **Vor.:** Seien g und h Ω -Geraden.

Beh.: Zwei Ω -Geraden $g \neq h$ schneiden sich höchstens in einem Punkt.

Bew.: Beweis durch Widerspruch: Würden sie sich in mindestens zwei Punkten P, Q schneiden, wären g, h zwei verschiedene Ω -Geraden durch diese Punkte. Dies widerspricht der Eigenschaft (i).

Das Parallelenaxiom der Euklidischen Geometrie ist hingegen nicht erfüllt. Anstelle dessen gilt:

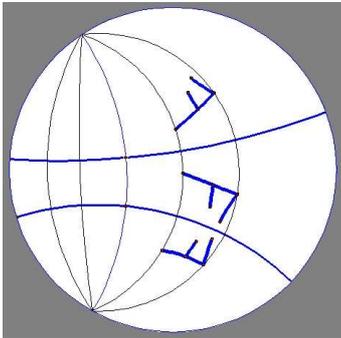
Parallelenaxiom der hyperbolischen Geometrie: Sei g eine Ω -Gerade und P ein Ω -Punkt, der nicht auf g liegt. Dann gibt es unendlich viele Ω -Geraden durch P , die g nicht schneiden. Solche Geraden l werden als Parallele bezeichnet.

4.2 Spiegelung und Abstandsbegriff

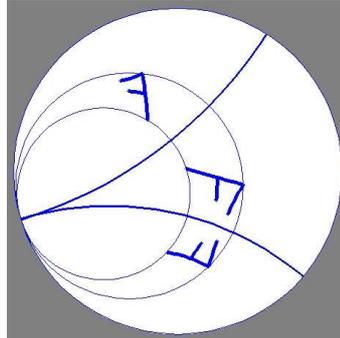
Da wir den Abstand in der hyperbolischen Geometrie nicht einfach mit konventionellen Werkzeugen (Lineal, Maßband, etc. ..) ausmessen können und die Berechnung nicht trivial ist, brauchen wir andere Mittel, um uns unter Abständen in der hyperbolischen Geometrie etwas vorstellen zu können.

Definition 4.1. Die Spiegelung S_g an einer Ω -Geraden g sei die Inversion F_k an dem g zugrunde liegenden Kreis k . Diese bildet Ω auf sich selbst ab.

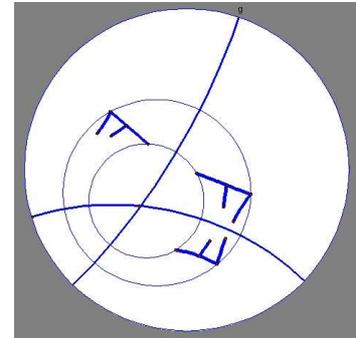
Wir fordern also von unseren Spiegelungen an Ω -Geraden, dass sie abstandserhaltend sind, und können auf diese Weise Strecken mit einander vergleichen. Insbesondere erhalten wir 3 verschiedene Typen von Isometrien der Form $S_m \circ S_l$:



Translation:
 l, m ultraparallel



Horozyklus-Rotation:
 l, m asymptotisch



Rotation:
 l, m schneiden sich

5 Steiner's Kreiskette

Steiner's Kreiskette (engl. *porism*) ist eine schöne Anwendung und Veranschaulichung der hyperbolischen Geometrie. Es handelt sich hier um einen Sachverhalt, der aus Euklidischer Sicht sehr erstaunlich ist, im Poincaré-Modell der hyperbolischen Geometrie jedoch fast trivial.

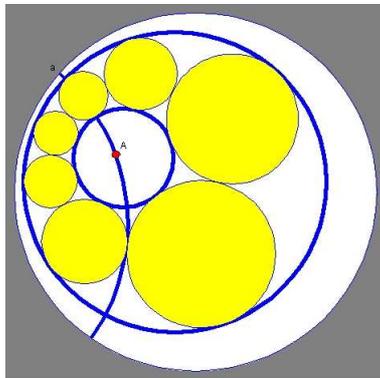


Abb.1

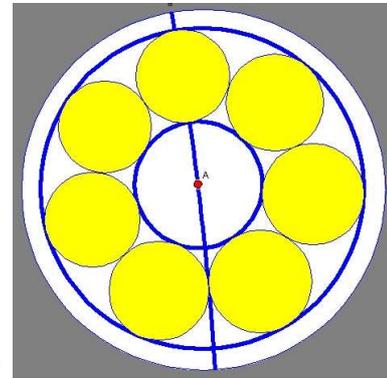


Abb.2

Wir betrachten einen äußeren Kreis k_1 und darin enthalten einen inneren Kreis k_2 . Wir können nun das Größenverhältnis dieser beiden Kreise so wählen, dass ihr Zwischenraum von einer Kreiskette ausgefüllt wird. Es gilt nun:

Satz 5.1. *Sind die Kreise k_1, k_2 so gewählt, dass es **eine** geschlossene Kreiskette gibt, so schließt sich **jede** Kreiskette zwischen k_1 und k_2 .*

Beweis. Für konzentrische Kreise k_1 und k_2 (Abb.2) ist die Aussage von Steiner offensichtlich. Seien nun die Kreise k_1, k_2 in beliebiger Lage. Bei

entsprechender Betrachtungsweise dieser Kreise im Poincaré-Modell der hyperbolischen Geometrie können diese immer noch als konzentrisch betrachtet werden, d.h. alle Kreise der Kette sind - hyperbolisch gesehen - gleich groß. Um dies nun auch Euklidisch zu verstehen, müssen wir durch eine geeignete (hyperbolische) Isometrie die Figur so überführen, dass die Kreise auch Euklidisch gleich groß sind. Dies ist genau dann der Fall, wenn der (hyperbolische) Mittelpunkt der Kreise k_1 und k_2 mit dem Mittelpunkt der Kreisscheibe Ω übereinstimmt. Wir zeigen:

Für zwei disjunkte Kreise k_1 und k_2 existiert ein Inversionskreis Σ derart, dass $F_\Sigma(k_1)$ und $F_\Sigma(k_2)$ konzentrische Kreise sind.

Da F_Σ Kreise stets auf Kreise abbildet und dabei Schnittwinkel (und die dazugehörigen Schnittpunkte) erhalten werden, kann also jede Kreiskonfiguration wie in Abb.1 auf eine entsprechende symmetrische wie in Abb.2 abgebildet werden. Der Beweis kann auf zwei Arten erfolgen: analytisch oder geometrisch, wobei bei beiden Methoden die Konstruktion des Inversionskreises das Ziel ist.

Analytisch betrachten wir einen Strahl durch die Mittelpunkte der Kreise k_1, k_2 , bezüglich dessen wir den Schnittpunkten mit k_1 und k_2 die Parameter a, b und c, d , dem gesuchten Mittelpunkt von Σ die Variable m und dem Radius von Σ die Variable r zuordnen. Für die an Σ gespiegelten Schnittpunkte gilt dann:

$$a' - m = \frac{r^2}{a - m}; \quad b' - m = \frac{r^2}{b - m}; \quad c' - m = \frac{r^2}{a - m}; \quad d' - m = \frac{r^2}{b - m}$$

Die Forderung, dass die invertierten Kreise konzentrisch sind, ist äquivalent zu:

$$\frac{a' + b'}{2} = \frac{c' + d'}{2}$$

Setzen wir die ersten vier Gleichungen in die letzte ein, so erhalten wir

$$\frac{r^2}{a - m} + \frac{r^2}{b - m} = \frac{r^2}{c - m} + \frac{r^2}{d - m}$$

und Auflösen nach m ergibt:

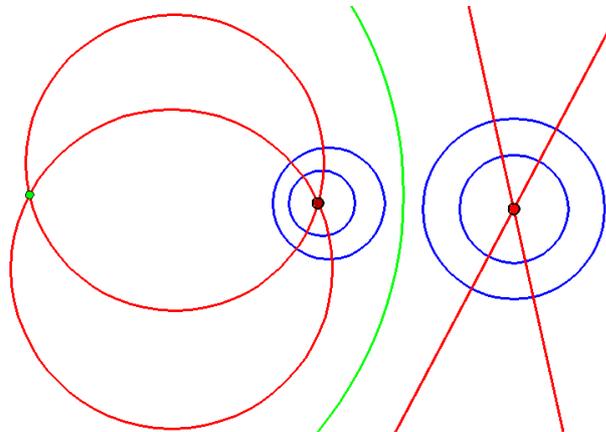
$$m = \frac{ac - bd \pm \sqrt{(a - c)(a - d)(b - c)(b - d)}}{a + b - c - d}$$

Für den Radius r können wir einen beliebigen Wert wählen. Damit haben wir konstruktiv die Existenz eines (bzw. im Normalfall zweier) solcher möglicher Mittelpunkte für Inversionskreise gezeigt.

Der geometrische Beweis beruht auf der Tatsache, dass die Inversion am Kreis eine Involution ist, d.h.

$$F_{\Sigma} = F_{\Sigma}^{-1}$$

Daher betrachten wir nun die Konstruktion rückwärts und gehen von zwei konzentrischen Kreisen k_1 und k_2 aus, durch deren gemeinsamen Mittelpunkt wir noch zwei Geraden g_1 und g_2 legen. Dieses Konstrukt invertieren wir nun an einem Kreis Σ , der allerdings nicht zu k_1 und k_2 konzentrisch sein soll (sonst bleiben diese Kreise konzentrisch) noch durch deren gemeinsamen Mittelpunkt gehen darf, da wir sonst Geraden erhalten.



Dabei gehen die konzentrischen Kreise k_1, k_2 über in disjunkte Kreise $k'_1 = F_{\sigma}(k_1)$ und $k'_2 = F_{\sigma}(k_2)$, während aus den Geraden g_1, g_2 zwei sich schneidende Kreise $g'_1 = F_{\sigma}(g_1)$ und $g'_2 = F_{\sigma}(g_2)$ werden (Erhalt von Schnittpunkten), die auch weiterhin k'_1 und k'_2 im rechten Winkel schneiden (Erhalt von Schnittwinkeln).

In umgekehrter Reihenfolge bleibt also die Aufgabe zu lösen, zu zwei gegebenen Kreisen ein Paar sich schneidender, orthogonaler Kreise zu finden. Dies kann man durch Tangentenkonstruktionen erreichen. \square

6 Winkelsumme im hyperbolischen Dreieck

Ziel dieses Abschnitts ist es das Gegenstück zum Satz 2.1 über sphärische Dreiecke in der hyperbolischen Geometrie zu beweisen:

Satz 6.1. *Der Flächeninhalt eines hyperbolischen Dreiecks mit Innenwinkeln α, β, γ ist proportional zu $\pi - \alpha - \beta - \gamma$. Insbesondere gilt $\alpha + \beta + \gamma < \pi$.*

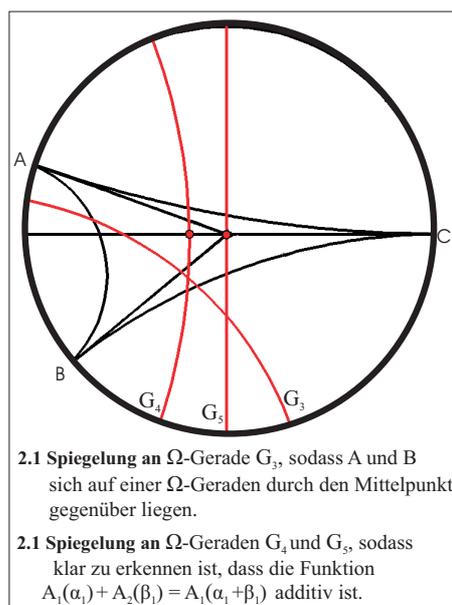
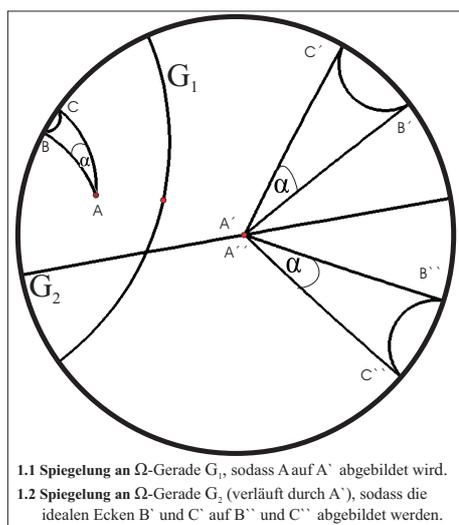
Der Beweis gliedert sich in mehrere Schritte:

1. Schritt: Alle hyperbolischen Dreiecke mit 2 idealen Ecken und gleichem Winkel sind kongruent.

Dabei bezeichnet man eine Ecke als *ideal*, wenn sie auf dem Rand $\partial\Omega$ liegt. Wir betrachten also zwei solche Dreiecke ABC und $A''B''C''$, die bei A den gleichen Winkel α einschließen. Dann findet man zwei Spiegelungen, die die Dreiecke ineinander überführt.

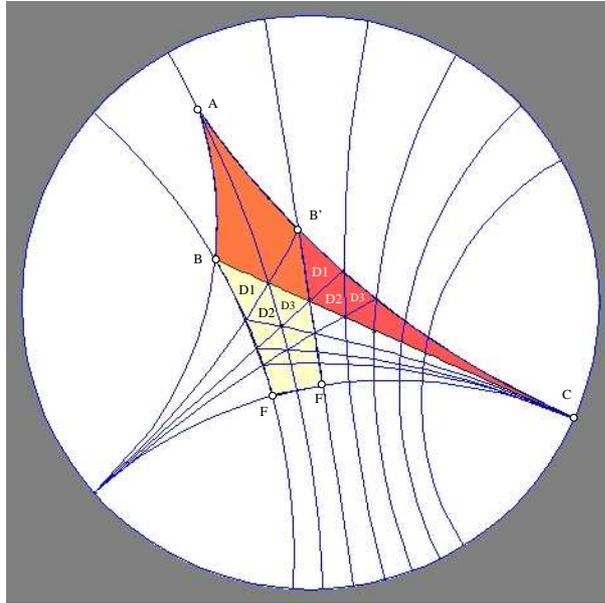
2. Schritt: Bezeichne $A(\alpha_1)$ den Flächeninhalt eines Dreiecks mit 2 idealen Ecken und Außenwinkel α_1 . Nach Schritt 1 sind diese alle kongruent. Dann gilt:

$$A(\alpha_1) + A(\beta_1) = A(\alpha_1 + \beta_1) \quad (2)$$



3. Schritt: Ein Dreieck mit einer idealen Ecke hat einen endlichen Flächeninhalt.

In der Skizze hat das Dreieck ABC eine ideale Ecke bei C . Dieses lässt sich nun so in Dreiecke D_k zerlegen, die nach Spiegelungen ganz im 5-Eck $B'ABFF'$ Platz finden, dessen Flächeninhalt endlich ist. Dabei entsteht B' durch Spiegelung an der Winkelhalbierenden bei A und so fort.



4. Schritt: Alle Dreiecke mit 3 idealen Ecken haben den selben endlichen Flächeninhalt A_{ideal} .

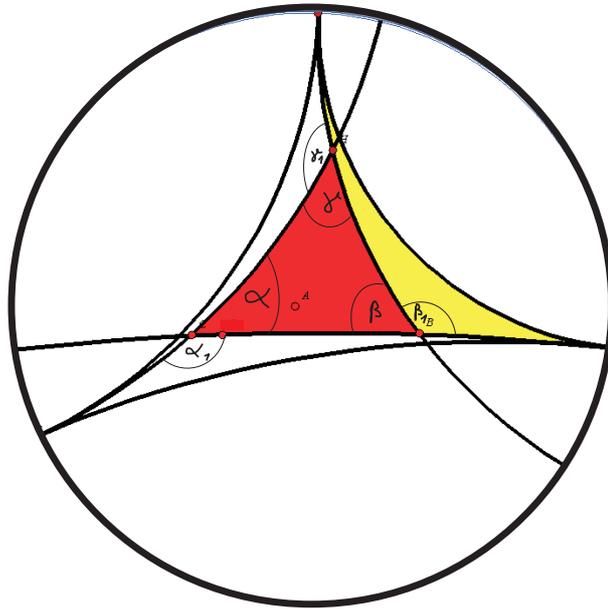
Ein solches ideales Dreieck lässt sich in 3 Dreiecke mit jeweils 2 idealen Ecken zerlegen. Jedes davon zerlegt sich wiederum in 2 Dreiecke mit nur einer idealen Ecke, von denen wir im Schritt 3 gezeigt haben, dass sie endlichen Flächeninhalt haben. Dass der Flächeninhalt immer derselbe ist für ein ideales Dreieck, lässt sich nun leicht mit (2) zeigen.

5. Schritt: Es gilt $A(0) = 0$, sowie $A(\pi) = A_{\text{ideal}}$. Nach Schritt 2 ist $A(\varphi)$ additiv und - wenn man A als stetige Funktion voraussetzt - damit linear: $A(\varphi) = m \cdot \varphi$. Für die Steigung m gilt also $m = \frac{A_{\text{ideal}}}{\pi}$, d.h.

$$A(\varphi) = A_{\text{ideal}} \cdot \frac{\varphi}{\pi} \quad (3)$$

Daraus folgt:

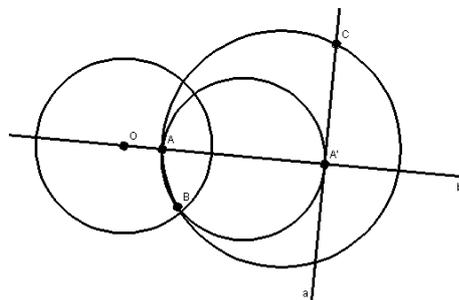
$$\begin{aligned} A_{\Delta} &= A_{\text{ideal}} - A(\alpha) - A(\beta) - A(\gamma) \\ &\stackrel{(3)}{=} A_{\text{ideal}} - A_{\text{ideal}} \cdot \frac{\alpha}{\pi} - A_{\text{ideal}} \cdot \frac{\beta}{\pi} - A_{\text{ideal}} \cdot \frac{\gamma}{\pi} \\ &= \frac{A_{\text{ideal}}}{\pi} (\pi - \alpha - \beta - \gamma) \end{aligned}$$



7 Konstruktionen mit Zirkel und Lineal

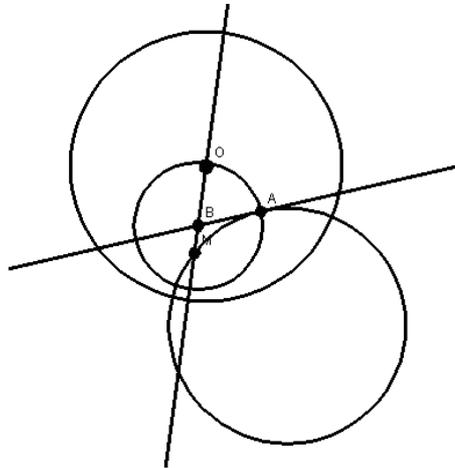
In Abschnitt 4 haben wir die hyperbolischen Entsprechungen für Kreis und Gerade in der Poincaré-Kreisscheibe gefunden. In diesem Abschnitt soll es nun um deren Konstruktion mit Euklidischen Mitteln gehen.

7.1 Ω -Gerade durch zwei Punkte A und B



Wir invertieren einen der Punkte an $\partial\Omega$, z.B. $A \rightarrow A'$ (siehe Abschnitt 3.1). Der Kreis durch A, B, A' liefert dann die gewünschte Ω -Gerade durch A, B .

7.2 Ω -Kreis durch A mit Mittelpunkt M

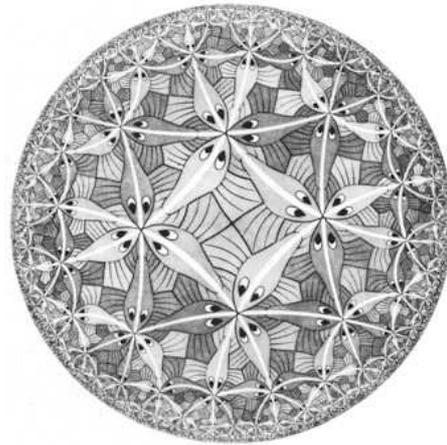


Wir wissen, dass ein Ω -Kreis k einem Euklidischen Kreis entspricht. Sei M dessen hyperbolischer und B dessen Euklidischer Mittelpunkt. Dann liegt B auf der Geraden \overline{OM} . Ferner schneidet die Ω -Gerade durch A und M als Durchmesser den Kreis k orthogonal. Daher konstruieren wir eine Tangente an die Ω -Gerade durch A, M (vgl. Abschnitt 7.1) in A und setzen den gesuchten Mittelpunkt auf den Schnittpunkt dieser mit der Geraden \overline{OM} .

Aus Satz 3.3 wissen wir, dass die Mittelpunkte von Kreisen durch einen Punkt A und senkrecht zu $\partial\Omega$ alle auf der Mittelsenkrechten zu $\overline{AA'}$ liegen, außerhalb von Ω . Diese Gerade nennt man *Polare a* zum Punkt A . Umgekehrt entspricht einer Ω -Geraden m ein Punkt außerhalb von Ω , nämlich der Mittelpunkt M des zugrundeliegenden Euklidischen Kreises. M wird auch als *Pol* zur Ω -Geraden m bezeichnet.

7.3 M.C.Escher

Allein mit Zirkel, Lineal und viel Geduld ausgestattet schuf der Künstler M.C.Escher um 1960 mittels Dreiecksparkettierung in der hyperbolischen Ebene phantastische Grafiken in seiner Reihe "Circle Limits". Vom Grundprinzip her werden spezielle, am Mittelpunkt der Poincaré-Kreisscheibe regelmäßig angeordnete Dreiecke immer wieder an ihren Seiten (d.h. Ω -Geraden) gespiegelt. Dabei muss für die Winkel $\frac{\pi}{p}, \frac{\pi}{q}, \frac{\pi}{r}$ im Ausgangsdreieck gelten, dass p, q, r natürliche Zahlen sind, sowie $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} < 1$ wegen Satz 6.1.



"There is no end of possibilities. It is wonderfully satisfying to make these pictures by hand, with pencil and paper, compass and straightedge. I encourage you to test this theorem for yourself!"

Chaim Goodman-Strauss 2001

8 Schlussbetrachtung

Der geneigte Leser wird, nachdem er bis an diese Stelle gekommen ist, wohl ermüdet sein von der hyperbolischen Welt. Das wollen wir ihm nicht übelnehmen, sondern vielmehr nach weiteren Anwendungsmöglichkeiten suchen. Tatsächlich lassen sich auch kosmologische Phänomene mittels hyperbolischer Vorstellung zwar nicht unbedingt einfacher lösen, so aber doch geometrischer verstehen. Wir als Seminarteilnehmer haben zumindest gelernt, dass hyperbolische Ameisen beim Betrachten von Eschers Bildern nicht unbedingt zu beneiden sind und die anscheinend Euklidische Welt vielleicht hyperbolisch besser verständlich wäre.