Geodäten im hyperbolischen Raum und Zahlentheorie

Petridis, Yiannis

Max-Planck-Institut für Mathematik, Bonn

Forschungsbereich - Mathematik
Korrespondierender Autor
Petridis, Yiannis, E-Mail: petridis@mpim-bonn.mpg.de

Zusammenfassung
Die Geometrie studiert Geodäten in verschiedenen Situationen, insbesondere auch auf hyperbolischen Flächen. Die Verteilung der Geodäten auf arithmetischen hyperbolischen Flächen führt zu Erkenntnissen über die Arithmetik von quadratischen Formen, einem wichtigen Zweig der Zahlentheorie.

Abstract
Geometry studies geodesics in various settings, in particular on hyperbolic surfaces. The distribution of geodesics on arithmetic hyperbolic surfaces gives information on the arithmetic of quadratic forms, an important branch of number theory.

Geodäten und die Sphäre.

Eine geodätische Kurve ist die Bahn eines Punktes im Raum, der sich reibungsfrei und ohne die Einwirkung äußerer Kräfte bewegt. Eine geodätische Kurve ist die Verbindungslinie kürzester Länge zwischen zwei Punkten, zumindest wenn die Punkte nahe genug beieinander liegen. Während im üblichen euklidischen Raum der Abstand zweier Punkte gegeben wird durch die Länge des Geradenstücks zwischen ihnen, ist die Situation im Allgemeinen schwieriger, aber auch interessanter. So ist z. B. (näherungsweise) die Erdoberfläche eine Sphäre. Fliegen wir von Frankfurt nach Los Angeles, führt uns der Weg über Grönland. Diese Streckenführung über einen hohen Breitengrad resultiert aus der Tatsache, dass die kürzeste Verbindungslinie (Geodäte) auf der Sphäre entlang eines Großkreises führt, das heisst entlang eines Kreises, der den gleichen Mittelpunkt hat wie die Sphäre (Erde). Ein anderes Beispiel wäre die Reise von London (auf dem Null-Meridian) nach Fiji, das im Pazifischen Ozean auf dem 180. Längengrad liegt. Die Geodäte zwischen diesen beiden Orten führt über den Nordpol (**Abb. 1**).

Ein direkter Flug nach Fiji sollte dieser Strecke folgen. Die von einem Punkt aus startenden Geodäten können in eine beliebige vorgegebene Richtung gehen (festgelegt durch den Anfangsgeschwindigkeitsvektor), aber sie treffen sich alle wieder im Antipodenpunkt. Dies zeigt die **Abbildung 1** ebenfalls.

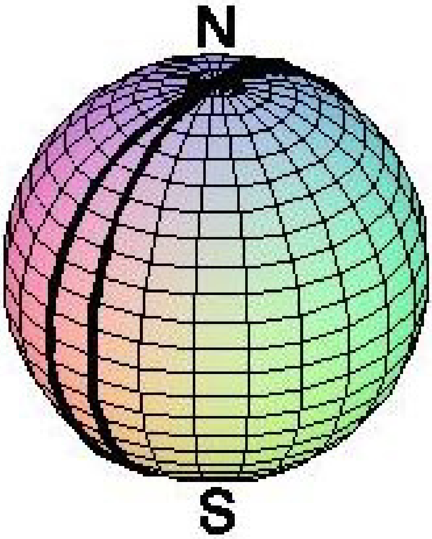

Abb. 1: *Zwei Längenkreise auf der Sphäre, vom Nordpol (N) zum Südpol (S).*
Urheber: Max-Planck-Institut für Mathematik, Y.Petridis

Der hyperbolische Raum und seine Geodäten.

Die hyperbolische Geometrie ist ein weiteres Beispiel für eine Geometrie, in der Euklids Axiome nicht gelten. Sie wurde von Bolyai und Lobatchevsky geschaffen. Hier laufen die Geodäten auseinander. Damit ist das Folgende gemeint: Der Abstand zwischen zwei Geodäten, die sich in einer Richtung annähern, aber nicht treffen, nimmt in der entgegengesetzten Richtung exponentiell zu. Der Abstand wird dabei auf den Kurven gemessen, die senkrecht auf den Geodäten stehen.

Ein einfaches Modell für die hyperbolische Geometrie ist die hyperbolische Ebene: die Punkte im Inneren des Kreises mit Radius 1, gegeben durch die Gleichung $x^2 + y^2 = 1$. Die Geodäten sind die Durchmesser und die Kreisbögen, die den Randkreis senkrecht treffen. Das fünfte euklidische Axiom wird verletzt: von P gibt es zwei Geodäten g, h parallel zum f (**Abb. 2**).

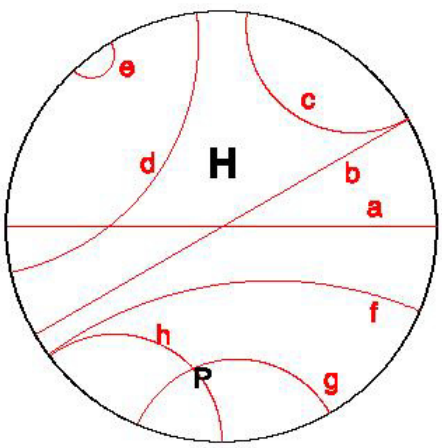

Abb. 2: *Hyperbolische Ebene (H) und ihre Geodäten.*
Urheber: Max-Planck-Institut für Mathematik, Y.Petridis

Der hyperbolische Abstand vom Ursprung zu dem Punkt (x,y) ist gleich $\log[(1+r)/(1-r)]$. Hierbei bezeichnet $r = \sqrt{(x^2 + y^2)}$ den euklidischen Abstand. Wenn sich der Punkt dem Randkreis mit Radius 1 nähert, wächst der hyperbolische Abstand über alle Maßen (und strebt gegen Unendlich). Die hyperbolische Geometrie ist für den Mathematiker genauso real wie die euklidische. Es könnte sogar so sein, dass das Universum ein hyperbolischer Raum ist, wie es eine Untersuchung von Messdaten der kosmischen Hintergrundstrahlung durch R. Aurich, S. Lustig, F. Steiner und H. Then nahelegt.

Hyperbolische Flächen und ihre Geodäten

Räume, die lokal wie eine hyperbolische Ebene aussehen, nennt man hyperbolische Flächen. Sie sind aus vielen Gründen wichtig, z. B. weil sie in Bezug zur Theorie der automorphen Formen stehen, eines der Hauptforschungsgebiete des Max-Planck-Instituts für Mathematik. Topologisch (also unter dem Gesichtspunkt der stetigen Deformierbarkeit) betrachtet, gehören sie zu den einfachsten Räumen, den Mannigfaltigkeiten. Einige der Geodäten γ auf der hyperbolischen Fläche M kommen wieder zum Ausgangspunkt zurück, sogar mit dem gleichen Geschwindigkeitsvektor. Diese heißen geschlossene Geodäten. Es stellt sich heraus, dass es bei vorgegebener Länge nur endlich viele von ihnen (wenn überhaupt welche) gibt. Die Längen $l(\gamma)$ der Geodäten bilden eine aufsteigende Zahlenfolge, die unbeschränkt ist. Man kann die geschlossenen Geodäten zählen:

$$\pi(x) = \text{Anzahl der } \gamma \text{ mit } l(\gamma) \leq x$$

Diese Funktion $\pi(x)$, die unterschiedliche geometrische Eigenschaften der Fläche erfasst, wurde in der Theorie der dynamischen Systeme mithilfe analytischer Methoden untersucht. Huber (1959) und Selberg, und in einem allgemeineren Fall Margulis (1970), bewiesen, dass die Funktion e^x/x eine gute Approximation von $\pi(x)$ ist. Und dies gilt für jede hyperbolische Fläche!

Die Selbergsche Spurformel ist ein wichtiges Instrument bei dem Studium der Funktion $\pi(x)$ und ein Objekt der Forschung vieler Gäste des MPIM und auch eines seiner Direktoren (D. Zagier). Sie ist eine Verallgemeinerung einer einfachen Tatsache, die man schon in den Grundvorlesungen über lineare Algebra lernt. Die Spur einer symmetrischen Matrix kann auf zwei verschiedene Weisen berechnet werden: Eine Methode ist, die Diagonaleinträge der Matrix zu summieren, die andere Methode besteht darin, die Summe der Eigenwerte zu bilden. Obgleich die zweite Rechenmethode im Allgemeinen die Schwierigere ist, so kann sie dennoch nützliche Informationen liefern, wenn z. B. gewisse Eigenschaften der Eigenwerte bekannt sind. Die Selbergsche Spurformel kann als Verallgemeinerung der beiden Methoden auf den Fall der Spurberechnung unendlicher Matrizen (d. h. Hilbertraumoperatoren) angesehen werden. Die Längen der geschlossenen Geodäten entsprechen den Diagonaleinträgen und die Eigenwerte sind die Eigenwerte des Laplace-Beltrami-Operators. Das asymptotische Verhalten von $\pi(x)$ wird dadurch bestimmt, dass der kleinste Eigenwert des Laplace-Beltrami-Operators 0 ist. Die Eigenwerte des Laplace-Beltrami Operators sind die Eigenfrequenzen von M: Sie bilden das Klangspektrum, das die Fläche M entwickelt, wenn man sie als Trommel benutzt.

Die Arithmetik Quadratischer Formen und Anwendungen

Wenn die hyperbolische Fläche von arithmetischer Natur ist, folgen wichtige zahlentheoretische Konsequenzen. Um diese Konsequenzen erklären zu können, muss kurz auf die Pellsche Gleichung eingegangen werden. Proclus (410-475 A.D.) bemerkte, dass die Pythagoräer einen geometrischen Ansatz, der zu einem Algorithmus führte, benutzen, um die einfachste Pellsche Gleichung

$$x^2 - 2y^2 = \pm 1$$

zu lösen. Eine Pellsche Gleichung ist eine Gleichung der Gestalt $x^2 - ky^2 = \pm 1$ mit vorgegebem k; sie zu lösen bedeutet, alle Paare (x, y) mit ganzzahligem x und y zu finden, die der Gleichung genügen. Man beginnt mit dem Paar $(x, y)=(1, 1)$, welches die ‚kleinste Lösung' von $x^2 - 2y^2 = -1$ ist. Hat man eine Lösung (x_n, y_n), so liefern die Zahlen

$$x_{n+1} = 2y_n + x_n, \quad y_{n+1} = y_n + x_n,$$

eine Lösung für die Gleichung mit entgegengesetztem Vorzeichen. Theon von Smyrna (ca. 130 A.D.) gab dieses Resultat an und er nannte die x_n Diagonalzahlen und die y_n Seitenzahlen. Die Paare, die man auf diese Art und Weise erhält, sind $(x_2,y_2) = (3,2)$, $(x_3,y_3) = (7,5)$, $(x_4,y_4) = (17,12)$, usw. Theon, ein Neoplatoniker, und Platos Schule waren an diesem Problem interessiert, da sie wussten, dass $2y^2$ keine Quadratzahl ist, falls y eine ganze Zahl ist (vgl. Proclus Kommentar zu Euklid I. 47). Dies bedeutet, dass man $\sqrt{2}$ nicht als einen Bruch mit ganzzahligem Zähler und ganzzahligem Nenner schreiben kann. Zahlen, die man nicht auf diese Art darstellen kann, heißen irrational. Daher suchte Theon nach der nächsten Möglichkeit, nähmlich dass sich $2y^2$ von dem Quadrat einer ganzen Zahl nur um ± 1 unterscheidet. In moderner algebraischer Notation funktioniert Theons Methode auf Grund von

$$x_{n+1}^2 - 2y_{n+1}^2 = (2y_n + x_n)^2 - 2(y_n + x_n)^2 = -(x_n^2 - 2y_n^2)$$

Die Paare (x_n,y_n) werden in moderner algebraischer Zahlentheorie durch die Formel

$$x_n + y_n\sqrt{2} = (1 + \sqrt{2})^n$$

angegeben.

Eine Verallgemeinerung der obigen Gleichung ist der Ausdruck

$$Q(x,y) = ax^2 + bxy + cy^2$$

mit ganzzahligen Koeffizienten a, b, c. Es handelt sich hierbei um eine quadratische Form und eine ganze Zahl N wird durch sie dargestellt, wenn es ganzzahlige x und y gibt mit $Q(x, y) = N$. Zur Vereinfachung werden im Folgenden nur solche quadratische Formen betrachtet, bei denen a, b und c relativ prim sind, d.h. die einzigen ganzen Zahlen, die a, b und c gleichzeitig teilen, sind ± 1. Die quadratische Form heisst indefinit, falls $Q(x, y)$ sowohl negative wie auch positive Zahlen darstellt. Bei diesen quadratischen Formen ist die Diskriminante $d = b^2 - 4ac > 0$, der gleiche Ausdruck ist Schülern in der Schule bekannt im Zusammenhang mit dem Lösen von $ax^2 + bx + c = 0$. Für eine positive ganze Zahl d kann es zwei verschiedene quadratische Formen $Q(x,y)$ und $Q'(x,y)$ mit Diskriminante d geben, die genau die gleichen Zahlen darstellen. Solche Formen werden identifiziert und der Mathematiker nennt sie dann äquivalent. Zum Beispiel ist $Q(x,y) = x^2 - 2y^2$ äquivalent zu $Q'(x,y) = -2x^2 + y^2$, wie man leicht durch das Vertauschen von x und y sieht. Aber quadratische Formen können auf kompliziertere Art zueinander äquivalent sein. Die obigen zwei Formen $Q(x, y)$ und $Q'(x, y)$ sind auch zu $Q''(x, y) = x^2 + 2xy - y^2$ äquivalent. Dies folgt durch einfache Rechnungen:

$$(x + y)^2 - 2y^2 = x^2 + 2xy - y^2, \quad (x - y)^2 + 2(x - y)y - y^2 = x^2 - 2y^2$$

Diese zeigen, dass $Q(x + y, y) = Q''(x, y)$ und $Q''(x-y, y) = Q(x, y)$ und folglich stellen diese beiden Formen die gleichen Zahlen dar. Man gibt die Variablentransformationen, die die äquivalenz von Q zu Q' und Q'' zeigen, mithilfe zweier Matrizen

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ und } T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

an. Diese Matrizen geben lineare Transformationen in der xy-Ebene an. Diese Transformationen sind

$$S(x, y) = (-y, x) \text{ und } T(x, y) = (x + y, y).$$

Man kann beweisen, dass es für ein vorgegebenes $d > 0$ nur endlich viele inäquivalente Formen der Diskriminante d gibt. Ihre Anzahl wird Klassenzahl genannt und mit $h(d)$ bezeichnet. Die Untersuchung quadratischer Formen, ihrer Klassenzahlen $h(d)$ und der Zahlen, die sie darstellen, ist ein aktiver Bereich mathematischer Forschung. D. Hilbert hat in der veröffentlichten Version seiner Ansprache auf dem Internationalen Kongress der Mathematiker in Paris 23 wichtige Probleme, die im 20. Jahrhundert untersucht werden sollten, aufgelistet. Nummer 11 bezieht sich auf quadratische Formen mit beliebigen algebraischen Zahlen als Koeffizienten.

Was die alten Griechen nicht wissen konnten, ist der enge Zusammenhang zwischen quadratischen Formen und Geodäten auf hyperbolischen Flächen. Dieser Zusammenhang sieht wie folgt aus: Man nimmt die Kleinste Lösung (x_d, y_d) der Gleichung

$$x^2 - dy^2 = 4$$

Ihre Grundeinheit ist definiert durch

$$\varepsilon_d = (x_d + y_d \sqrt{d}) / 2$$

Wie man aus der obigen Diskussion erkennt, ist dies für $d = 2$ die Zahl $3 + 2\sqrt{2}$. Die Länge aller geschlossenen Geodäten auf einer gewissen arithmetischen hyperbolischen Fläche (verbunden mit den Matrizen S und T) lässt sich mit $2\log \varepsilon_d$ identifizieren und jede Länge tritt $h(d)$-mal auf. Dabei werden aber nur solche d zugelassen, die keine Quadratzahl sind und bei der Division durch 4 den Rest 0 oder 1 haben. Dies ermöglichte Sarnak (1982) zu beweisen, dass für diese d

$$\sum_{\varepsilon_d \leq x} h(d) \sim x^2 / (2\ln x)$$

ist. Das Symbol ~ bedeutet, dass der Quotient von dem Ausdruck auf der linken Seite durch den Ausdruck auf der rechten Seite für $x \to \infty$ gegen 1 strebt.

Verfeinerte Verteilungen der Geodäten

Die topologische Natur der Fläche M wird durch die Anzahl ihrer Löcher angegeben. Dies wird als das Geschlecht g von M bezeichnet. Es ist leichter, den Begriff der Homologie einer Fläche vom Geschlecht 1 wie in **Abbildung 3** zu verstehen. Sie sieht wie ein Donut aus und wird Torus genannt.

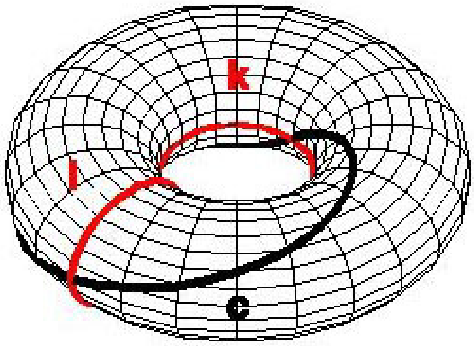

Abb. 3: *Torus mit Homologiebasis k, l und Weg c .*

Urheber: Max-Planck-Institut für Mathematik, Y.Petridis

Alle Längenkreise des Donuts lassen sich in einen festen Längenkreis, der den Donut einmal umläuft, überführen, wohingegen die Kreise eines festen Breitenkreises sich zu dem kleinen Kreis, der das innere Loch umrundet, zusammenziehen lassen. Jeder Weg γ auf dem Torus lässt sich durch den fixierten Längenkreis (l) und den kleinen inneren Kreis (k) darstellen. Jedem Weg γ kann man seine Homologie $\Phi(\gamma)$ zuordnen. Dies ist ein Paar ganzer Zahlen (n_1, n_2) das angibt, dass der Längenkreis (n_1)- mal und der innere Kreis n_2-mal durchlaufen wird. Der kompliziertere Weg aus Abbildung 3 durchläuft einmal den fixierten Längenkreis (l) und einmal den kleinen inneren Kreis (k), wenn man ihn zusammenzieht. Allgemeiner lässt sich die Homologie $H_1(M, \mathbf{Z})$ bei einer Fläche von Geschlecht g durch die Punkte $(n_1, n_2, \ldots, n_{2g})$ mit ganzzahligen Koordinaten angeben. Nun kann man eine verfeinerte Frage stellen: Wie sind die Längen der Geodäten verteilt, wenn man nur Geodäten mit Homologien in einer vorgegebenen Teilmenge von $H_1(M, \mathbf{Z})$ betrachtet? Für Geodäten, die alle dieselbe Homologie haben, stammt die Antwort von Phillips und Sarnak (1987), die die Funktion

$$\pi(x,\beta) = \text{Anzahl der } \gamma \text{ mit } l(\gamma) \leq x \text{ und } \Phi(\gamma) = \beta$$

betrachteten. Sie zeigten

$$\pi(x,\beta) \sim (g-1)^g e^x / (x^{g+1})$$

Hierbei ist bemerkenswert, dass der Ausdruck auf der rechten Seite nur noch von der topologischen Invariante g und nicht von β abhängt. In einer neueren Arbeit untersuchen Y. Petridis (MPIM) und M. S. Risager (Gast am MPI) Geodäten, dessen Homologien zu einer Teilmenge $A \subset H_1(M, Z)$ gehören. Man definiert

$$\pi(x,A) = \text{Anzahl der } \gamma \text{ mit } l(\gamma) \leq x \text{ und } \Phi(\gamma) \text{ in } A$$

Man misst die Menge A durch ihre Dichte $d(A)$, die angibt, welcher Anteil der Gitterpunkte der Homologie zu A gehören. Es stellte sich heraus, dass für viele Mengen, einschließlich zufälliger Mengen,

$$\pi(x,A) \sim d(A)\, \pi(x)$$

Der geodätische Fluss auf einer hyperbolischen Fläche ist ein Beispiel für einen chaotischen Fluss, ein Objekt von großem Interesse in der Theorie der dynamischen Systeme. Mithilfe dynamischer Systeme haben Eskin, Margulis und Mozes die Werte anderer interessanter indefiniter quadratischen Formen untersucht. Die Verbindung zwischen Zahlentheorie und den dynamischen Systemen wird in der Zukunft zweifellos noch viele neue spektakuläre Resultate liefern. (*Y. Petridis*)