Kapitel 1: Was ist Topologie? 1
Kapitel 2: Topologische Räume, Stetigkeit und Konvergenz 9
Kapitel 3: Zusammenhängende und kompakte Mengen 22
Kapitel 4: Auf dem Weg zur algebraischen Topologie 33
§1: Das Königsberger Brückenproblem 33
§2: Kartenfärbungsprobleme 35
§3: Was ist eine Karte? 42
§4: Graphen 48
§5: Sechs Farben reichen 49
§6: Fünf Farben reichen 53
§7: Wie wurde der Vierfarbensatz bewiesen? 56
§8: Ist der Vierfarbensatz damit bewiesen? 60
§9: Literaturhinweise zum Vierfarbenproblem 64
Kapitel 5: Simpliziale Komplexe 66
Kapitel 6: Die Homologie simplizialer Komplexe 71
Kapitel 7: Anwendungen der Homologietheorie 99
Kapitel 8: Der Fixpunktsatz von Kakutani 114
Kapitel 9: Wirtschaftliche Gleichgewichte 135
§1: Wirtschaftssysteme 136
§2: Abstrakte Ökonomien 143
§3: Beweis von Satz 1 145
§4: Der Satz von Arrow und Debreu 153
Kapitel 10: Variationen 167
§1: Nashs zweiter Beweis für die Existenz von Gleichgewichten .. 167
§2: Andere Beweise des Fixpunktsatzes von Brouwer 173

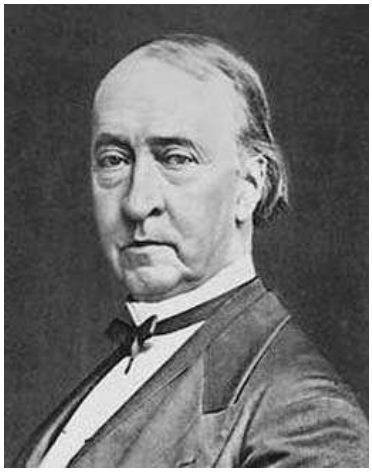
Kapitel 1
Was ist Topologie?

Die Topologie ist, verglichen etwa mit der Geometrie und der Zahlentheorie, ein sehr junges Teilgebiet der Mathematik. Der erste mathematische Satz, den man nach heutigem Verständnis der Topologie zurechnen würde, war EULERs Lösung des Königsberger Brückenproblems, mit der wir uns zu Beginn von Kapitel 4 genauer beschäftigen werden. Er veröffentlichte sie 1736 unter dem Titel *Solutio problematis ad geometriam situs pertinentis* (Lösung eines Problems bezüglich der Geometrie der Lage). Später wurden ähnliche Probleme oft unter dem Oberbegriff *Analysis situs* zusammengefaßt; unter diesem Titel erschien auch ab 1895 eine Reihe von Arbeiten von HENRI POINCARÉ, in denen er als Erster das Gebiet systematisch behandelte, das wir heute als algebraische Topologie bezeichnen.

LEONHARD EULER (1707–1783) wurde in Basel geboren, ging dort auch zur Schule und, im Alter von 14 Jahren, zur Universität. Zwei Jahre später legte er die Magisterprüfung in Philosophie ab und begann mit dem Studium der Theologie. Daneben hatte er sich auch seit Beginn seines Studium unter Anleitung von JOHANN BERNOULLI mit Mathematik beschäftigt. 1726 beendete er sein Studium in Basel und bekam eine Stelle an der Petersburger Akademie der Wissenschaften, die er 1727 antrat. Auf Einladung FRIEDRICHS DES GROSSEN wechselte er 1741 an die preußische Akademie der Wissenschaften. Nachdem sich das Verhältnis zu diesem dramatisch verschlechtert hatte, kehrte er 1766 nach St. Petersburg zurück. Im gleichen Jahr erblindete er vollständig. Trotzdem schrieb er rund die Hälfte seiner zahlreichen Arbeiten (Seine gesammelten Abhandlungen umfassen 73 Bände) danach. Sie enthalten bedeutende Beiträge zu zahlreichen Teilgebieten der Mathematik, Physik, Astronomie und Kartographie.

JULES HENRI POINCARÉ (1854–1912) wurde im lothringischen Nancy geboren, besuchte dort das heute nach ihm benannte Gymnasium und studierte von 1873 bis 1875 an der École Polytechnique, danach an der École des Mines in Paris. Sodann arbeitete er gleichzeitig als Mineningenieur und an einer Dissertation über Differentialgleichungen bei CHARLES HERMITE (1822–1901); letztere wurde 1879 von der Universität Paris angenommen. Seine erste mathematische Stelle war an der Universität Caen; danach folgten Professuren in Paris, die er bis zu seinem frühen Tod innehatte. Er leistete wesentliche Beiträge zur Theorie der Differentialgleichungen, begründete die algebraische Topologie, publizierte über Physik, blieb Zeit seines Lebens dem Bergbau verbunden und veröffentlichte auch einige philosophische Abhandlungen. Das Grab der Familie POINCARÉ auf dem Friedhof von Montparnasse ist heute noch erhalten.

Der Name *Topologie* von τοπος, der Ort, und λογος, die Lehre, geht zurück auf JOHANN BENEDICT LISTING, der 1847 eine Arbeit mit dem Titel *Vorstudien zur Topologie* veröffentlichte, das Wort *Topologie* aber bereits ab etwa 1836 immer wieder in seinen Briefen gebraucht hatte. In den allgemeinen Gebrauch kam das Wort erst ab etwa 1940, nicht zuletzt auf Grund der Bemühungen des BOURBAKI-Kreises, der in seinem (noch immer unvollendeten und wohl auch unvollendet bleibenden) Werk *Les fondations de l'Analyse* die Analysis in größtmöglicher Allgemeinheit neu aufbaute auf Grundlage der Algebra und der Topologie.



JOHANN BENEDICT LISTING (1808–1882) wurde in Frankfurt geboren und ging auch dort aufs Gymnasium; 1830 begann er mit dem Studium der Mathematik und der Architektur an der Universität Göttingen, wo er allerdings auch Vorlesungen aus zahlreichen anderen Gebieten belegte. Sein wichtigster mathematischer Lehrer war GAUSS, der zwar selbst nie etwas über Topologie veröffentlichte, aber doch eine ganze Reihe von Ideen zu diesem Gebiet beisteuerte. Als 1837 Victoria Königin von England und ihr Onkel Herzog von Hannover wurde, protestierte unter anderem der Göttinger Physiker WILHELM EDUARD WEBER (1804–1891), der gemeinsam mit GAUSS den Telegraphen erfunden hatte, gegen die Aufhebung der liberalen Hannoveraner Verfassung und verlor deshalb seine Stelle. Auf Fürsprache von GAUSS wurde LISTING sein Nachfolger und publizierte seither auch auf dem Gebiet der Physik.

NICOLAS BOURBAKI ist das Pseudonym einer Gruppe junger französischer Mathematiker, die ab 1935 an einen neuen Neuaufbau der Mathematik arbeiteten. Obwohl dabei anfangs vor allem die Lehre im Vordergrund stand, war ihr wesentlicher Grundsatz, alles so allgemein wie möglich zu formulieren und zu beweisen und zusätzliche Annahmen nur einzuführen, wo sie unbedingt erforderlich sind. Um Alterserscheinungen zu vermeiden, müssen alle Mitarbeiter im Alter von vierzig Jahren ausscheiden, so daß die aktiven Mathematiker stets jünger sind. Die Arbeit am großen Grundlagenwerk scheint derzeit eingestellt zu sein; inzwischen organisiert BOURBAKI im wesentlichen nur noch ein viermal jährlich in Paris stattfindendes Seminar, in dem prominente Mathematiker über ein Thema ihrer Wahl, aber nicht ihre eigenen Arbeiten, vortragen.

Die Topologie untersucht, informell ausgedrückt, jene Eigenschaften von Kurven, Flächen, Körpern und auch komplizierteren Mengen, die unter stetigen Deformationen erhalten bleiben.

Wenn wir uns der Einfachheit halber zunächst auf Teilmengen der Räume $\mathbb{R}^n$ beschränken, können wir das folgendermaßen präzisieren:

Definition: Zwei Mengen $X \subseteq \mathbb{R}^n$ und $Y \subseteq \mathbb{R}^m$ heißen *homöomorph,* wenn es stetige Abbildungen $f\colon X \to Y$ und $g\colon Y \to X$ gibt derart, daß $f \circ g$ die Identität auf Y ist und $g \circ f$ die Identität auf X.

Insbesondere sind f und g also bijektiv; wichtig ist aber vor allem, daß *beide* Abbildungen stetig sind: Die Abbildung

$$f\colon \begin{cases} [0,\, 2\pi) \to \left\{(x,y) \in \mathbb{R}^2 \;\middle|\; x^2 + y^2 = 1\right\} \\ \qquad t \mapsto (\cos t, \sin t) \end{cases}$$

etwa vom halboffenen Intervall $[0,\, 2\pi)$ auf die Kreislinie ist stetig und bijektiv, aber ihre Umkehrabbildung ist im Punkt $(1,0)$ nicht stetig, denn die Punkte $\big(\cos(2\pi - \frac{1}{n}), \sin(2\pi - \frac{1}{n})\big)$ haben zwar für jedes $n \in \mathbb{N}$ Urbild $2\pi - \frac{1}{n}$, aber ihr Grenzwert $(1,0)$ hat die Null als Urbild.

Als erstes Beispiel zweier homöomorpher Mengen betrachten wir den n-dimensionalen Würfel

$$W = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; 0 \le x_i \le 1 \;\text{ für alle } i\right\}$$

und für beliebige positive reelle Zahlen $a_1, \ldots, a_n$ den Quader

$$Q = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; 0 \le x_i \le a_i \;\text{ für alle } i\right\}.$$

Offensichtlich sind die Abbildungen

$$f\colon \begin{cases} W \to Q \\ (x_1, \dots, x_n) \mapsto (a_1 x_1, \dots, a_n x_n) \end{cases} \quad \text{und}$$

$$g\colon \begin{cases} Q \to W \\ (x_1, \dots, x_n) \mapsto \left(\dfrac{x_1}{a_1}, \dots, \dfrac{x_n}{a_n}\right) \end{cases}$$

stetig und zueinander invers; der Würfel und der Quader sind also homöomorph. Damit ist schon einmal klar, daß Längen und Längenverhältnisse keine topologischen Eigenschaften sind.

Allgemeiner können wir beliebige affine Abbildungen

$$\varphi\colon \begin{cases} \mathbb{R}^n \to \mathbb{R}^n \\ x \mapsto Ax + b \end{cases}$$

mit einer invertierbaren $n \times n$-Matrix A und einem Vektor $b \in \mathbb{R}^n$ betrachten. Offensichtlich ist dann jede Teilmenge $Z \subseteq \mathbb{R}^n$ homöomorph zu ihrem Bild $\varphi(Z)$, denn sowohl φ als auch seine Umkehrabbildung, die $y \in \mathbb{R}^n$ abbildet auf $A^{-1}(y - b)$, sind stetig. Somit ist auch die Lage im $\mathbb{R}^n$ topologisch irrelevant, genauso wie Winkel, denn wir können beispielsweise durch eine affine Abbildung ein Rechteck in ein Parallelogramm scheren.

Auch die Kugel

$$K = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n \;\middle|\; x_1^2 + \dots + x_n^2 \le 1\right\}$$

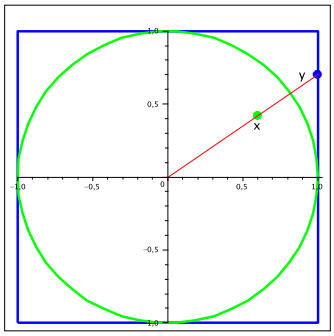
ist homöomorph zu W (und damit auch zu Q); hier lassen sich die Abbildungen $K \to W$ und $W \to K$ am besten geometrisch beschreiben: Zunächst einmal ist W homöomorph zum Würfel

$$W' = \left\{(x_1, \dots, x_n) \in \mathbb{R}^n \;\middle|\; -1 \le x_i \le 1 \ \text{ für alle } i\right\}$$

über die affine Transformation $\mathbb{R}^n \to \mathbb{R}^n$, die jede Koordinate x_i auf $2x_i - 1$ abbildet, und diesen Würfel bilden wir dann folgendermaßen ab auf die Kugel:

Der Nullpunkt des $\mathbb{R}^n$ wird auf sich selbst abgebildet. Für jeden anderen Punkt $x = (x_1, \dots, x_n) \in W'$ betrachten wir den Strahl vom Nullpunkt

durch diesen Punkt; sein Schnittpunkt mit dem Rand des Würfels sei y, und d sei sein Abstand zum Nullpunkt. Dann wird x abgebildet auf den Punkt dx; wir strecken also den Strahl vom Nullpunkt durch x in so einem Verhältnis, daß der Schnittpunkt mit der Kugel auf den Schnittpunkt mit dem Würfel abgebildet wird. Bei der Umkehrabbildung wird der Strahl entsprechend um den Faktor d gestaucht.

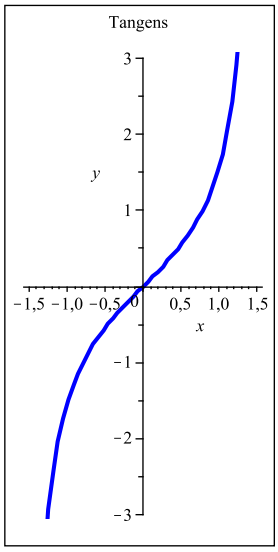


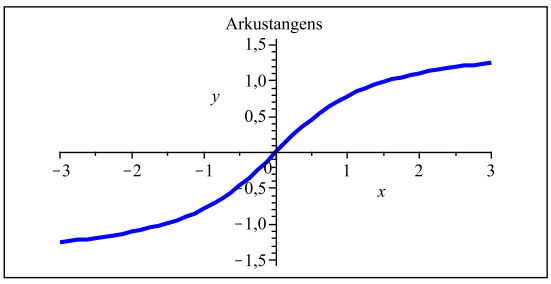
Damit ist auch klar, daß Kanten, Ecken und ähnliches keine Begriffe sind, die unter Homöomorphismen erhalten bleiben.

Auch Beschränktheit ist keine topologische Eigenschaft, denn die Tangensfunktion bildet das Intervall $(-\frac{\pi}{2}, \frac{\pi}{2})$ stetig und bijektiv ab auf ganz $\mathbb{R}$ mit dem ebenfalls stetigen Arkustangens als Umkehrfunktion. Entsprechend können wir den offenen Würfel

$$W'' = \left\{(x_1, \ldots, x_n) \;\middle|\; -\tfrac{\pi}{2} < x_i < \tfrac{\pi}{2} \quad \text{für alle } i\right\}$$

homöomorph abbilden auf den ganzen $\mathbb{R}^n$, indem wir die i-te Koordinate x_i ersetzen durch $\tan x_i$.



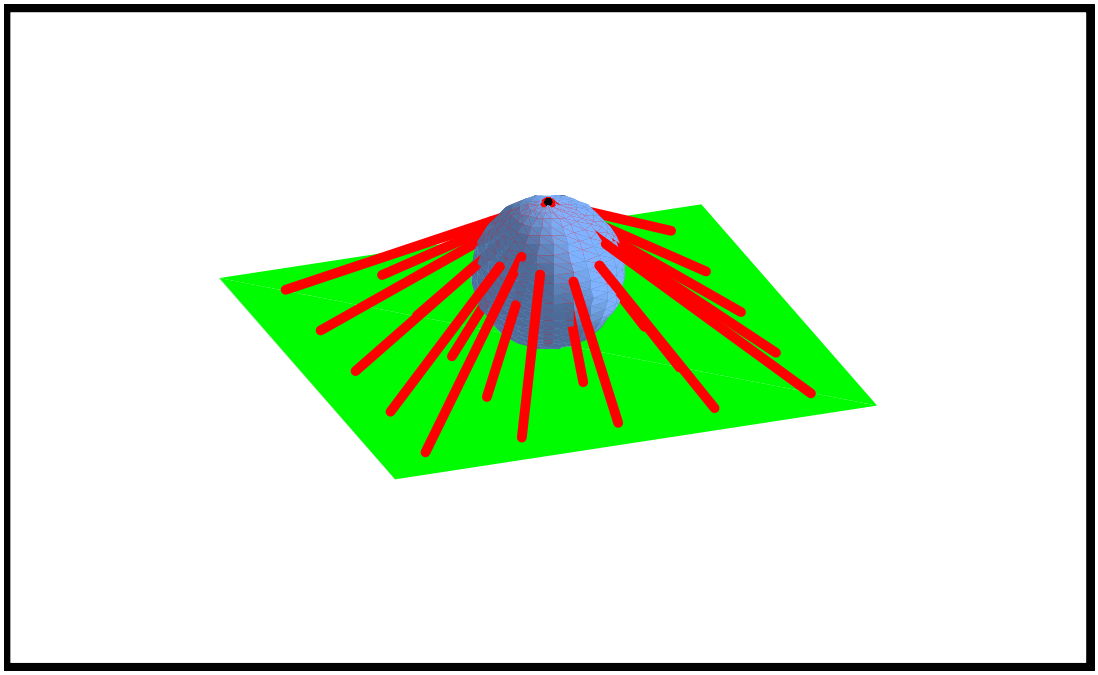


Der offene Würfel W'' ist allerdings weder homöomorph zu W noch zu W', denn diese beiden Würfel sind kompakt und werden damit auch von jeder stetigen Abbildung wieder auf kompakte Mengen abgebildet. Daher kann es weder eine surjektive stetige Abbildung von W nach W'' noch eine von W' nach W'' geben. Aus dem gleichen Grund sind beiden Würfel nicht homöomorph zu $\mathbb{R}^n$.

Die Kugeloberfläche

$$\mathbb{S}^2 = \left\{(x, y, z) \in \mathbb{R}^3 \;\middle|\; x^2 + y^2 + z^2 = 1\right\}$$

kann nicht homöomorph zu $\mathbb{R}^2$ sein, denn auch $\mathbb{S}^2$ ist eine kompakte Menge. Entfernt man aber nur einen Punkt von $\mathbb{S}^2$, etwa den Nordpol $N = (0, 0, 1)$, so ist $\mathbb{S}^2 \setminus \{N\}$ homöomorph zu $\mathbb{R}^2$: Wir identifizieren $\mathbb{R}^2$ mit der Ebene $z = -1$ in $\mathbb{R}^3$, also der Menge aller Punkte der Form $(x, y, -1) \in \mathbb{R}^3$, und bilden den Punkt $P \in \mathbb{S}^2 \setminus \{N\}$ ab auf den Schnittpunkt der Geraden NP mit der Ebenen E. In Gegenrichtung wird der Punkt $Q \in E$ abgebildet auf den Schnittpunkt der Geraden NQ mit $\mathbb{S}^2$. Offensichtlich ist auch diese *stereographische Projektion* $\mathbb{S}^2 \setminus \{N\} \to E$ ein Homöomorphismus.

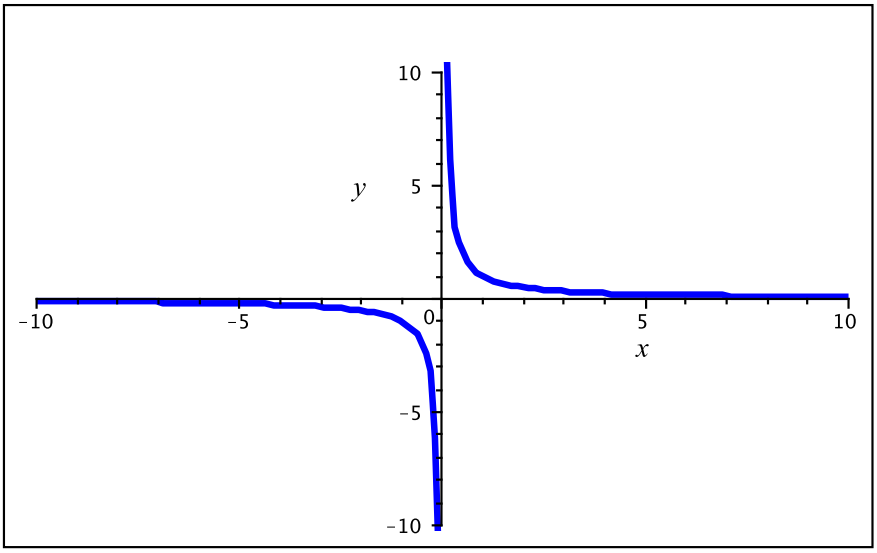


Die Hyperbel $H = \left\{(x, y) \in \mathbb{R}^2 \;\middle|\; xy = 1\right\}$ ist homöomorph zu $\mathbb{R} \setminus \{0\}$

via der beiden Abbildungen

$$f\colon \begin{cases} H \to \mathbb{R} \setminus \{0\} \\ (x,y) \mapsto x \end{cases} \quad \text{und} \quad g\colon \begin{cases} \mathbb{R} \setminus \{0\} \to H \\ x \mapsto \left(x, \dfrac{1}{x}\right) \end{cases} ;$$

eine abgeschlossene Teilmenge von $\mathbb{R}^2$ kann also homöomorph sein zu einer offenen Teilmenge von $\mathbb{R}$.



Angesichts dieser Beispiele könnte man sich fragen, ob ein Begriff, der so viele verschiedene Mengen miteinander identifiziert, überhaupt nützlich sein kann. Tatsächlich ist genau diese Vielfalt ein großer Vorteil der Topologie: Auf einem Würfel etwa können wir gut rechnen und damit viele Probleme recht einfach entscheiden. Andererseits gibt es aber kaum Probleme der „wirklichen“ Welt, die sich sinnvoll mit einem einfachen Würfel modellieren lassen. Es gibt aber gerade in der Ökonomie durchaus Ansätze, ein Wirtschaftssystem durch eine Teilmenge eines $\mathbb{R}^n$ zu modellieren, die zwar sehr viel komplizierter ist als ein Würfel, aber trotzdem homöomorph zu einem Würfel in einem geeigneten $\mathbb{R}^m$. Über den Homöomorphismus lassen sich dann Fragen wie die nach der Existenz eines Gleichgewichts zurückführen auf Fragen über einen Würfel, bei denen es erheblich größere Chancen gibt, eine Antwort zu finden.

Ein wesentlicher Teil der Vorlesung wird sich deshalb damit beschäftigen, kompliziertere Mengen und Abbildungen via Homöomorphie

zurückzuführen auf Polyeder und stückweise lineare Abbildungen zwischen ihnen. Insbesondere lassen sich so viele Probleme reduzieren auf Fragen aus der Linearen Algebra, und für deren Beantwortung steht uns ein weites Instrumentarium zur Verfügung.

Kapitel 2
Topologische Räume, Stetigkeit und Konvergenz

Wie wir im letzten Kapitel gesehen haben, kann eine offene Teilmenge eines $\mathbb{R}^n$ homöomorph sein zu einer abgeschlossenen Teilmenge eines $\mathbb{R}^m$; zwei Einbettungen einer Menge in einen $\mathbb{R}^n$ können also dort recht verschiedene Eigenschaften haben. Außerdem macht der umgebende $\mathbb{R}^n$ die Situation oft eher unübersichtlich: Wenn wir uns auf der Erdoberfläche bewegen, könnten wir zwar mit den kartesischen Koordinaten eines umgebenden $\mathbb{R}^3$ rechnen, aber natürlich bevorzugen wir zweidimensionale Koordinatensysteme wie Längen- und Breitengrade oder UTM-Koordinaten. Obwohl das Rechnen mit reellen Zahlen im weiteren Verlauf der Vorlesung eine große Rolle spielen wird, insbesondere auch bei den Anwendungen auf Gleichgewichte, definieren wir daher unseren Grundbegriff, den topologischen Raum, abstrakt und ohne Bezug zu einem $\mathbb{R}^n$.

Topologische Eigenschaften sind solche, die unter Homöomorphismen invariant bleiben; um diese definieren zu können, brauchen wir stetige Abbildungen. Stetige Abbildungen werden in der Analysis meist mit Hilfe von ε und δ definiert; da Abstände keine topologischen Eigenschaften sind, ist das für uns kein sinnvoller Zugang. Wir wissen aber aus der Analysis, daß eine Abbildung auch genau dann stetig ist, wenn das Urbild jeder offenen Menge wieder offen ist; es reicht also zur Definition von Stetigkeit, wenn wir offene Mengen haben.

Eine Teilmenge des $U \subseteq \mathbb{R}^n$ heißt bekanntlich genau dann offen, wenn es zu jedem Punkt $x \in U$ ein $\varepsilon > 0$ gibt, so daß jedes $y \in \mathbb{R}^n$ mit Abstand kleiner ε von x ebenfalls in U liegt. Damit ist klar, daß die Vereinigung beliebig vieler offener Mengen wieder offen ist, genauso wie

zumindest der Durchschnitt endlich vieler offener Mengen. (Der Durchschnitt unendlich vieler offener Mengen muß nicht mehr offen sein; beispielsweise ist der Durchschnitt der offenen Intervalle $\left(-\frac{1}{n},\ 1+\frac{1}{n}\right)$ das abgeschlossene Intervall $[0,\ 1]$.) Außerdem ist der gesamte $\mathbb{R}^n$ offen und natürlich auch die leere Menge. Diese Eigenschaften wollen wir zur Grundlage unserer Theorie machen:

Definition: *a)* Ein *topologischer Raum* X ist eine Menge X zusammen mit einer Menge $\mathcal{T}$ von Teilmengen von X, genannt die *Topologie* von X, für die gilt:

1.) $X \in \mathcal{T}$ und $\emptyset \in \mathcal{T}$.
2.) Ist I eine beliebige Indexmenge und ist $U_i \in \mathcal{T}$ für alle $i \in I$, so liegt auch die Vereinigung $\bigcup\limits_{i \in I} U_i$ der U_i in $\mathcal{T}$.
3.) Sind für eine natürliche Zahl r die Mengen $U_1, \ldots, U_r$ Elemente von $\mathcal{T}$, so auch ihr Durchschnitt $\bigcap\limits_{i=1}^{r} U_i$.

b) Die Elemente von $\mathcal{T}$ heißen *offene Mengen.*
c) Eine Teilmenge $A \subseteq X$ heißt *abgeschlossen,* wenn $X \setminus A$ offen ist.
d) Eine Abbildung $f\colon X \to Y$ zwischen zwei topologischen Räumen X und Y heißt *stetig*, wenn für jede offene Teilmenge $U \subset Y$ auch die Urbildmenge $f^{-1}(U)$ offen in X ist.
e) Eine stetige Abbildung $f\colon X \to Y$ heißt *Homöomorphismus,* wenn es eine stetige Abbildung $g\colon Y \to X$ gibt, so daß $f \circ g$ die Identität auf Y ist und $g \circ f$ die Identität auf X. (f und g sind dann natürlich bijektiv.)
f) Zwei topologische Räume heißen *homöomorph,* wenn es einen Homöomorphismus $f\colon X \to Y$ gibt.

Nach der Diskussion zu Beginn dieses Kapitels ist klar, daß der $\mathbb{R}^n$ ein topologischer Raum ist, wenn wir für $\mathcal{T}$ die Menge aller (im Sinne der Analysis) offenen Teilmengen nehmen.

Wir können jede beliebige Menge zum topologischen Raum machen, indem wir als $\mathcal{T}$ einfach die kleinstmögliche Menge nehmen, nämlich $\mathcal{T} = \{\emptyset, X\}$. Diese Topologie bezeichnen wir als die *triviale* Topologie. Umgekehrt könnten wir für $\mathcal{T}$ auch die größtmögliche Teilmenge der Potenzmenge $\mathfrak{P}(X)$ nehmen, also $\mathfrak{P}(X)$ selbst; dann ist *jede* Teilmengen von X offen. In diesem Fall reden wir von der *diskreten* Topologie.

Schon diese Beispiele zeigen, daß wir eigentlich nicht die Menge X, sondern das Paar $(X, \mathcal{T})$ als topologischen Raum bezeichnen sollten, denn zumindest für den $\mathbb{R}^n$ kennen wir ja bereits jetzt drei verschiedene mögliche Topologien $\mathcal{T}$. Meist wird allerdings aus dem Zusammenhang klar sein, was $\mathcal{T}$ ist; im Falle des $\mathbb{R}^n$ beispielsweise werden wir zumindest in dieser Vorlesung immer davon ausgehen, daß er – wenn nicht explizit etwas anderes gesagt wird – seine „übliche“ Topologie trägt. Entsprechend werden wir, sofern keine Mißverständnisse zu befürchten sind, auch sonst meist kurz die Menge X als topologischen Raum bezeichnen.

Zum besseren Verständnis des Stetigkeitsbegriffs für allgemeine topologische Räume wollen wir aber doch kurz sehen, was passiert, wenn wir auf dem $\mathbb{R}^n$ verschiedene Topologien betrachten: X sei der $\mathbb{R}^n$ mit der diskreten Topologie, Y sei der $\mathbb{R}^n$ mit der üblichen Topologie und Z schließlich der $\mathbb{R}^n$ mit der trivialen Topologie. Als Abbildungen zwischen diesen Räumen betrachten wir nur die identische Abbildung. Dann sind die Abbildung $X \to Y$, $X \to Z$ und $Y \to Z$ stetig, denn das Urbild einer Menge ist ja gerade die Menge selbst, und wenn eine Teilmenge des $\mathbb{R}^n$ offen ist bezüglich der trivialen Topologie, ist sie erst recht offen bezüglich der üblichen Topologie, und bezüglich der diskreten Topologie ist ohnehin *jede* Menge offen. Von den Abbildungen $Z \to Y$, $Z \to X$ und $Y \to X$ ist aber keine stetig, denn im Zielraum gibt es jeweils offene Mengen, die im Urbildraum nicht offen sind.

Wenn wir zeigen wollen, daß eine Abbildung $f\colon X \to Y$ stetig ist, müssen wir nach obiger Definition nachweisen, daß das Urbild *jeder* offenen Menge aus Y offen in X ist. Im Falle $Y = \mathbb{R}^n$ etwa gibt es (bezüglich der üblichen Topologie) sehr viele offene Mengen, über deren allgemeine Struktur wir nur wenig sagen können. Für alle praktischen Zwecke ist für Abbildungen $\mathbb{R}^m \to \mathbb{R}^n$ die aus der Analysis gewohnte ε-δ-Definition viel einfacher nachzuprüfen als die Offenheit der Urbilder offener Mengen.

Um auch für beliebige topologische Räume die Stetigkeit einer Abbildung lokal formulieren zu können, führen wir den Begriff der Umgebung ein:

Definition: X sei ein topologischer Raum. Eine *offene Umgebung* U eines Punktes $x \in X$ ist eine offene Menge, die x enthält. Eine *Umgebung* V von x ist eine Teilmenge $V \subseteq X$, die eine offene Umgebung von x enthält.

Lemma: Das System aller Umgebungen eines Punktes x erfüllt die folgenden vier HAUSDORFFschen Umgebungsaxiome:
a) Ist U eine Umgebung von x, so liegt x in U. Der gesamte Raum X ist Umgebung aller Punkte $x \in X$.
b) Eine Teilmenge $V \subseteq X$, die eine Umgebung U von x enthält, ist selbst eine Umgebung von x.
c) Der Durchschnitt endlich vieler Umgebungen von X ist selbst eine Umgebung von x.
d) Jede Umgebung V von x enthält eine Umgebung U von x, die gleichzeitig Umgebung eines jeden Punkts $y \in U$ ist.

Beweis: a) und *b)* folgen sofort aus der Definition. Sind $V_1, \ldots, V_r$ Umgebungen von x, so gibt es nach Definition offene Umgebungen $U_1 \subseteq V_1, \ldots, U_r \subseteq V_r$ von x. Der Durchschnitt $\bigcap_{i=1}^{r} U_i$ ist offen und enthält x, ist also eine offene Umgebung von x. Da dieser Durchschnitt im Durchschnitt der V_i liegt, ist damit auch *c)* bewiesen. Für *d)* schließlich müssen wir für U einfach eine in V enthaltene offene Umgebung von x nehmen; diese ist nach Definition Umgebung aller ihrer Punkte. ■

FELIX HAUSDORFF (1868–1942) wurde in Breslau als Sohn eines jüdischen Kaufmanns geboren. Zwei Jahre später übersiedelte die Familie nach Leipzig, wo er auch Mathematik studierte und mit Arbeiten über Anwendungen auf die Astronomie sowohl 1891 promovierte als auch 1895 habilitierte. Erst ab 1904 beschäftigte er sich mit Mengenlehre und der Topologie, wofür er heute vor allem bekannt ist. 1910–1913 lehrte er in Bonn, wohin er nach einer Professur in Greifswald 1921 auch wieder zurückkehrte. 1935 mußte er auf Grund der nationalsozialistischen Gesetze seine Stelle aufgeben, beschäftigte sich aber weiterhin mit Topologie. Bevor er 1942 interniert und nach Polen deportiert werden sollte, begingen er, seine Frau und deren Schwester Selbstmord. 2006

wurde in Bonn das nach ihm benannte Hausdorff-Zentrum für Mathematik gegründet, das erste und bis 2018 einzige Exzellenzcluster im Bereich der Mathematik.

Umgekehrt lassen sich topologische Räume auch charakterisieren durch diese Umgebungsaxiome: Ist für jedes Element $x \in X$ eine Menge $\mathfrak{U}_x$ von Teilmengen gegeben, die wir als Umgebungen von x bezeichnen, so definieren wir eine Menge $U \subseteq X$ als offen, wenn sie mit jedem Punkt $x \in U$ auch mindestens eine der Mengen aus $\mathfrak{U}_x$ enthält. Falls die HAUSDORFFschen Umgebungsaxiome gelten, erfüllen die so definierten offenen Mengen offenbar alle drei Eigenschaften aus der Definition eines topologischen Raums.

Zur lokalen Charakterisierung der Stetigkeit definieren wir:

Definition: Eine Abbildung $f\colon X \to Y$ zwischen zwei topologischen Räumen heißt *stetig im Punkt* $x \in X$, falls es zu jeder Umgebung V von $f(x)$ eine Umgebung U von x gibt mit $f(U) \subseteq V$.

Man beachte, daß dies genau der ε, δ Definition aus der Analysis entspricht.

Lemma: Eine Abbildung $f\colon X \to Y$ zwischen zwei topologischen Räumen ist genau dann stetig, wenn sie in jedem Punkt $x \in X$ stetig ist.

Beweis: Sei zunächst f stetig, $x \in X$ und V eine Umgebung von $f(x)$. Dann enthält V eine offene Umgebung V' von $f(x)$, und wegen der Stetigkeit von f ist deren Urbild $U = f^{-1}(V')$ eine offene Umgebung von x. Da $f(U) \subseteq V' \subseteq V$ ist, folgt die Stetigkeit von f in x.

Sei umgekehrt f stetig in jedem Punkt $x \in X$ und $V \subset Y$ offen. Da V Umgebung jedes seiner Punkte ist, hat jeder Punkt $x \in U = f^{-1}(V)$ eine Umgebung, deren Bild in V liegt, und damit erst recht eine offene Umgebung $U_x \subseteq U$. Als Vereinigung aller dieser offener Mengen U_x ist auch U offen, womit die Stetigkeit von f bewiesen ist. ∎

Um interessantere Beispiele topologischer Räume zu bekommen, wollen wir als nächstes beliebige Teilmengen von $\mathbb{R}^n$ oder allgemeiner noch von *irgendeinem* topologischen Raum zu topologischen Räumen machen:

Definition: X sei ein topologischer Raum und $Z \subseteq X$ sei eine Teilmenge von X. Eine Menge $U \subseteq Z$ heißt *offen bezüglich der Spurtopologie von Z bezüglich X*, wenn es eine offene Teilmenge $V \subseteq X$ gibt, so daß $U = V \cap Z$ ist.

Bevor wir mit dieser Definition arbeiten, sollten wir zunächst überprüfen, daß Z durch diese Definition zu einem topologischen Raum wird, daß die Spurtopologie also wirklich eine Topologie ist.

Der gesamte Raum $Z = X \cap Z$ und die leere Menge $\emptyset = \emptyset \cap Z$ sind offen in der Spurtopologie. Sind $U_i, i \in I$ irgendwelche dort offene Mengen, so gibt es in X offene Mengen V_i mit $U_i = V_i \cap X$ und

$$\bigcup_{i \in I} U_i = \bigcup_{i \in I} (V_i \cap Z) = \left(\bigcup_{i \in I} V_i \right) \cap Z$$

ist offen. Entsprechend ist für endlich viele offene Mengen $U_i = V_i \cap Z$

$$\bigcap_{i=1}^{r} U_i = \bigcap_{i=1}^{r} (V_i \cap Z) = \left(\bigcap_{i=1}^{r} V_i \right) \cap Z$$

eine offene Menge.

Man beachte, daß offene Mengen $U \subseteq Z$ bezüglich der Spurtopologie als Teilmengen von X im allgemeinen *nicht* offen sind: So ist beispielsweise Z selbst stets offen bezüglich der Spurtopologie, auch wenn Z keine offene Teilmenge von X ist.

Betrachten wir etwa als Beispiel noch einmal die Abbildung

$$f\colon \begin{cases} [0,\, 2\pi) \to \left\{ (x, y) \in \mathbb{R}^2 \;\middle|\; x^2 + y^2 = 1 \right\} \\ \qquad t \mapsto (\cos t, \sin t) \end{cases}$$

aus dem letzten Kapitel, jetzt aber bezüglich der Spurtopologien.

Im halboffenen Intervall $X = [0,\, 2\pi)$ sind dann insbesondere auch alle Teilintervalle $[0,\, a)$ mit $a \leq 2\pi$ offen, denn wir können sie ja schreiben als $[0,\, a) = (-1,\, a) \cap X$. Für $a < 2\pi$ ist ihr Urbild unter der Umkehrabbildung von f, also ihr Bild unter f, gleich der Menge aller Punkte (x, y) aus $\mathbb{R}^2$, für die es ein $t \in [0,\, a)$ gibt, so daß $x = \cos t$ und $y = \sin t$ ist. Insbesondere liegt also der Punkt $(1, 0) = (\cos 0, \sin 0)$ dort.

Wäre die Menge offen, so gäbe es eine offene Menge $U \subseteq \mathbb{R}^2$, so daß sie der Durchschnitt der Kreislinie mit U wäre. Insbesondere müßte U eine ε-Umgebung von $(1,0)$ enthalten, und deren Schnitt mit der Kreislinie enthielte auch die Punkte $\big(\cos(2\pi - \delta), \sin(2\pi - \delta)\big) = \big(\cos\delta, -\sin\delta\big)$ mit $0 \leq \delta < \arcsin\varepsilon$. Zumindest für $\delta < 2\pi - a$ liegen diese aber nicht in der Menge. Somit ist die Umkehrabbildung nicht stetig und f kein Homöomorphismus.

Die Stetigkeit von f dagegen läßt sich leicht mit dem lokalen Kriterium beweisen: Ist $t_0 \in (0,\ 2\pi)$ und ist U irgendeine Umgebung des Punkts $f(t_0) = (\cos t_0, \sin t_0)$, so enthält U insbesondere eine Teilmenge bestehend aus allen Punkten der Form $(\cos t, \sin t)$ mit $t_1 < t < t_2$; außerdem ist $0 < t_1 < t_0 < t_2 < 2\pi$. Somit wird das Intervall $(t_1,\ t_2)$ von f nach U abgebildet.

Bleibt noch der Punkt $t_0 = 0$ zu betrachten. Jede Umgebung U seines Bilds $(1,0)$ enthält eine Teilmenge bestehend aus allen Punkten $(\cos t, \sin t)$ mit $0 \leq t < t_1$ oder $t_2 < t < 2\pi$ für geeignete reelle Zahlen $0 < t_1 < t_2 < 2\pi$. Die Menge $[0,\ t_1) \cup (t_2,\ 2\pi)$ wird von f auf diese Teilmenge und somit nach U abgebildet und ist, als Vereinigung zweier in der Spurtopologie offener Teilmengen von X selbst offen in X. Damit ist die Stetigkeit von f bewiesen.

Aus der Analysis sind wir es gewohnt, die Stetigkeit einer Funktion auch oft durch die Vertauschbarkeit von f mit Grenzwerten von Folgen nachzuprüfen. Da die Konvergenz einer der wesentlichen Grundbegriffe der Analysis ist, der unter anderem sowohl bei der Definition des Differentialquotienten als auch bei der des bestimmten Integrals eine wichtige Rolle spielt, sollten wir auch für allgemeine topologische Räume erklären, was konvergente Folgen sind und schauen, ob wir Stetigkeit auch damit charakterisieren können. Wir ersetzen dazu einfach in der aus der Analysis bekannten Definition die ε-Umgebungen durch beliebige offene Umgebungen:

Definition: Eine Folge $(x_n)_{n\in\mathbb{N}}$ von Punkten x_n eines topologischen Raums X konvergiert gegen den Punkt $x \in X$, wenn es zu jeder offenen Umgebung U von x ein $N \in \mathbb{N}$ gibt, so daß $x_n \in U$ für alle $n \geq N$.

Für Folgen reeller Zahlen ist das äquivalent zur klassischen Definition, bei exotischen Topologien kann sie aber auf den ersten Blick unerwartete Konsequenzen haben: Versehen wir etwa eine Menge X mit der trivialen Topologie, betrachten also nur die leere Menge und X selbst als offen, so ist X für jeden seiner Punkte $x \in X$ die einzige offene Umgebung, und somit konvergiert jede Folge von Elementen aus X gegen jedes Element $x \in X$.

Etwas komplizierter ist das folgende Beispiel: Wir betrachten die Menge $X = (\mathbb{R} \setminus \{0\}) \cup \{a, b\}$ mit zwei abstrakten Punkten a und b. Eine Teilmenge $U \subseteq \mathbb{R} \setminus \{0\}$ soll genau dann offen sein, wenn sie auch als Teilmenge von $\mathbb{R}$ offen ist. Eine Teilmenge $U \subseteq X$, die a und/oder b enthält, können wir schreiben als $U = V \cup W$ mit $V \subseteq \mathbb{R} \setminus \{0\}$ und $W \subseteq \{a, b\}$; sie soll genau dann offen sein, wenn $V \cup \{0\} \subseteq \mathbb{R}$ offen ist. Man überlegt sich leicht, daß auf dieser sogenannten *Knüppelgeraden* die Folge der Zahlen $1/n$ sowohl gegen a als auch gegen b konvergiert.

Wenn wir solche Phänomene vermeiden wollen, müssen wir von unserem topologischen Raum X mehr fordern als in der obigen Definition. Die notwendige Zusatzbedingung geht zurück auf FELIX HAUSDORFF:

Definition: Ein topologischer Raum X heißt HAUSDORFFsch oder ein HAUSDORFF-Raum, wenn es zu je zwei Punkten $x, y \in X$ offene Umgebungen U von x und V von y gibt, so daß $U \cap V = \emptyset$ ist.

Offensichtlich ist jeder $\mathbb{R}^n$ sowie jede Teilmenge eines $\mathbb{R}^n$ HAUSDORFFsch, denn dort können wir vom Abstand d der beiden Punkte x und y reden. Nehmen wir als U und V die ε-Umgebungen mit $\varepsilon = \frac{d}{3}$, so haben wir zwei disjunkte Umgebungen.

Allgemeiner gilt auch:

Lemma: Jede Teilmenge eines HAUSDORFF-Raums ist (bezüglich der Spurtopologie) selbst HAUSDORFFsch.

Beweis: X sei ein HAUSDORFF-Raum und $Z \subset X$. Zu $x \neq y$ aus Z gibt es in X offene Umgebungen U von x und V von y mit $U \cap V = \emptyset$. Offensichtlich sind dann $U \cap Z$ und $V \cap Z$ zwei bezüglich der Spurtopologie offene disjunkte Umgebungen in Z.

■

Uns interessiert hier besonders die folgende Aussage:

Lemma: In einem HAUSDORFF-Raum hat jede Folge $(x_n)_{n\in\mathbb{N}}$ höchstens einen Grenzwert.

Beweis: Angenommen, x und y wären zwei verschiedene Grenzwerte der Folge. Nach Definition eines HAUSDORFF-Raums gäbe es dazu disjunkte Umgebungen U von x und V von y, und nach Definition der Konvergenz müßte es Indizes $N, M \in \mathbb{N}$ geben, so daß $x_n \in U$ für alle $n \geq N$ und $x_n \in V$ für alle $n \geq M$. Das ist aber offensichtlich unmöglich, denn $U \cap V = \emptyset$.

■

In der Analysis können wir die Stetigkeit von Funktionen mit Hilfe konvergenter Folgen *definieren.* Nachdem wir gesehen haben, daß Konvergenz in beliebigen topologischen Räumen etwas ganz anderes bedeuten kann als das, was wir aus der Analysis gewohnt sind, wollen wir der entsprechenden Eigenschaft für beliebige topologische Räume vorsichtshalber einen anderen Namen geben:

Definition: Eine Abbildung $f\colon X \to Y$ zwischen zwei topologischen Räumen heißt *folgenstetig,* wenn für jede konvergente Folge $(x_n)_{n\in\mathbb{N}}$ von Punkten aus X mit Grenzwert $x \in X$ auch die Folge der Bildpunkte $f(x_n)$ gegen $f(x)$ konvergiert.

Lemma: Jede stetige Funktion $f\colon X \to Y$ zwischen zwei topologischen Räumen ist folgenstetig.

Beweis: $(x_n)_{n\in\mathbb{N}}$ konvergiere in X gegen den Punkt x, und U sei eine offene Umgebung von $f(x)$. Wegen der Stetigkeit von f ist dann $f^{-1}(U)$ eine offene Umgebung von x, und wegen der Konvergenz der Folge $(x_n)_{n\in\mathbb{N}}$ gibt es ein $N \in \mathbb{N}$, so daß $x_n \in f^{-1}(U)$ für alle $n \geq N$. Damit liegt $f(x_n)$ für alle $n \geq N$ in U, was die Konvergenz der Folge der $f(x_n)$ gegen $f(x)$ beweist.

■

Für reelle Funktionen, egal ob in einer oder in mehreren Veränderlichen, gilt bekanntlich auch die Umkehrung dieses Lemmas. Beim Beweis spielen dort aber die speziellen ε-Umgebungen mit $\varepsilon = \frac{1}{n}$ eine wesentliche

Rolle, und für die haben wir in der allgemeinen Situation kein Analogon. So verwundert es kaum, daß wir auch hier noch zusätzliche Forderungen an den topologischen Raum stellen müssen. (Man kann explizite Gegenbeispiele konstruieren von Funktionen, die folgenstetig, aber nicht stetig sind; da dies aber mit unseren derzeitigen Kenntnissen recht langwierig wäre, sei hier darauf verzichtet.)

Definition: *a)* Eine Menge $\mathfrak{U}$ von Umgebungen eines Punktes x heißt *Umgebungsbasis* von x, wenn jede Umgebung von x eine Umgebung aus $\mathfrak{U}$ enthält.
b) Ein topologischer Raum X erfüllt das erste Abzählbarkeitsaxiom und heißt *1-abzählbar,* wenn jeder Punkt $x \in X$ eine abzählbare Umgebungsbasis hat.

Beispiele 1-abzählbarer Räume sind die Räume $\mathbb{R}^n$ oder allgemeiner metrische Räume, wo die ε-Umgebungen mit $\varepsilon = \frac{1}{n}$ eine solche Basis bilden. Typische Gegenbeispiele sind Funktionenräume wie die Menge *aller* Abbildungen $\mathbb{R} \to \mathbb{R}$. In dieser Vorlesung werden wir uns mit solchen Räumen nicht beschäftigen.

Angenommen, die offenen Umgebungen $V_1, V_2, \ldots$ bilden eine abzählbare Umgebungsbasis des Punktes $x \in X$. Dann gilt dasselbe für die Durchschnitte

$$W_1 = V_1, \quad W_2 = V_1 \cap V_2, \quad W_3 = V_1 \cap V_2 \cap V_3, \quad \ldots,$$

und offensichtlich ist $W_1 \supseteq W_2 \supseteq W_3 \supseteq \cdots$. Somit hat in einem 1-abzählbaren topologischen Raum jeder Punkt auch eine abzählbare Umgebungsbasis, deren Elemente immer kleiner werden. Eine solche Umgebungsbasis kann in Beweisen oft die ε-Umgebungen mit $\varepsilon = \frac{1}{n}$ ersetzen, beispielsweise im folgenden

Lemma: Ist X ein 1-abzählbarer und Y ein beliebiger topologischer Raum, so ist jede folgenstetige Abbildung $f\colon X \to Y$ stetig.

Beweis: x sei ein beliebiger Punkt aus X und $U \subseteq Y$ eine Umgebung von $f(x)$. Wir müssen zeigen, daß es eine Umgebung V von x gibt, so daß $f(V)$ in U liegt.

Angenommen, es gibt keine solche Umgebung V. Auch für eine abzählbare Umgebungsbasis $\{V_1, V_2, \dots\}$ von x aus immer kleiner werdenden offenen Umgebungen liegt dann das Bild keiner dieser Mengen in U. Wir können daher Elemente $x_n \in V_n$ finden derart, daß $f(x_n)$ nicht in U liegt. Da die V_n eine Umgebungsbasis bilden, konvergiert die Folge der x_n gegen x; wegen der Folgenstetigkeit konvergiert also die Folge der $f(x_n)$ gegen $f(x)$. Das ist aber unmöglich, denn kein einziges der Folgenglieder liegt in der offenen Umgebung U von $f(x)$. ■

Das letzte Argument dieses Beweises hätten wir auch anders formulieren können: Die Punkte $f(x_n)$ liegen allesamt in der abgeschlossenen Menge $Y \setminus U$, und deshalb muß auch ihr Grenzwert dort liegen nach dem folgenden

Lemma: $(x_n)_{n\in\mathbb{N}}$ sei eine konvergente Folge von Elementen der abgeschlossenen Teilmenge Z des topologischen Raums X. Konvergiert die Folge gegen ein $x \in X$, so muß x in Z liegen.

Beweis: Läge x im Komplement $U = X \setminus Z$ von Z, so müßte es wegen der Konvergenz der Folge ein $N \in \mathbb{N}$ geben, so daß auch alle x_n mit $n \geq N$ in dieser offenen Umgebung von x lägen. Tatsächlich liegt dort aber kein einziges dieser Elemente. ■

Für 1-abzählbare topologische Räume läßt sich dieses Lemma auch umkehren zu einer Charakterisierung abgeschlossener Mengen:

Definition: Ein Punkt $x \in X$ heißt *Randpunkt* der Teilmenge $Z \subseteq X$ des topologischen Raums X, wenn jede Umgebung von x sowohl Punkte aus Z als auch Punkte aus $X \setminus Z$ enthält.

Lemma: *a)* Eine Teilmenge $Z \subseteq X$ ist genau dann abgeschlossen, wenn sie jeden ihrer Randpunkte enthält.
b) In einem 1-abzählbaren topologischen Raum ist eine Teilmenge $Z \subseteq X$ genau dann abgeschlossen, wenn für jede gegen ein $z \in X$ konvergente Folge von Elementen $z_n \in Z$ der Grenzwert z in Z liegt.

Beweis: a) Ist $Z \subseteq X$ abgeschlossen und $x \in X$ ein Randpunkt von Z, der nicht in Z liegt, so liegt x im offenen Komplement von Z, hat also

dort eine offene Umgebung U. In dieser kann aber kein Punkt aus Z liegen. Liegt umgekehrt jeder Randpunkt von Z in Z, so kann kein Punkt aus dem Komplement von Z Randpunkt sein. Somit hat jeder solche Punkt eine offene Umgebung, die keinen Punkt aus Z enthält. Damit ist das Komplement von Z als Vereinigung dieser offenen Umgebungen offen, Z selbst also abgeschlossen.

b) Ist Z abgeschlossen, so konvergiert nach dem vorigen Lemma jede in X konvergente Folge von Elementen aus Z gegen ein Element aus Z. Umgekehrt habe Z diese Eigenschaft, und x sei ein Randpunkt von Z. Wir betrachten eine abzählbare Umgebungsbasis von x bestehend aus ineinander liegenden offenen Mengen $U_1 \supset U_2 \supset \cdots$. In jedem dieser U_n liegen Punkte aus Z. Wir wählen für jede Umgebung U_n einen solchen Punkt $z_n \in U_n \cap Z$. Da die U_n eine Umgebungsbasis von x bilden, konvergiert die Folge der z_n gegen x, also liegt x in Z. Somit enthält Z alle seine Randpunkte, ist also nach *a)* abgeschlossen.

■

Wenn es ein erstes Abzählbarkeitsaxiom gibt, sollte es auch ein zweites geben. Wir werden es zwar nur selten brauchen, aber der Vollständigkeit halber sei es trotzdem erwähnt:

Definition: *a)* Eine Menge $\mathfrak{B}$ von offenen Teilmengen eines topologischen Raums X heißt *Basis der Topologie von* X, wenn sich jede offene Menge als Vereinigung von Mengen aus $\mathfrak{B}$ darstellen läßt.
b) Ein topologischer Raum X erfüllt das zweite Abzählbarkeitsaxiom und heißt *2-abzählbar,* wenn seine Topologie eine abzählbare Basis hat.

Beispielsweise ist der $\mathbb{R}^n$ 2-abzählbar, denn die offenen Mengen

$$\left\{ y \in \mathbb{R}^n \;\middle|\; \|y - x\| < \frac{1}{m} \right\} \quad \text{mit} \quad x \in \mathbb{Q}^n, \;\; m \in \mathbb{N}$$

bilden für jede Norm $\|.\|$ des $\mathbb{R}^n$ eine abzählbare Basis seiner Topologie.

Offensichtlich ist das System der offenen Mengen eines topologischen Raums durch eine Basis seiner Topologie eindeutig bestimmt: Eine Menge ist genau dann offen, wenn sie sich als Vereinigung von Basismengen schreiben läßt. Dies können wir dazu verwenden, um auf einfache Weise das Produkt zweier topologischer Räume zu einem topologischen Raum zu machen:

Definition: X und Y seien topologische Räume. Die *Produkttopologie* auf $X \times Y$ ist jene Topologie, die die Menge aller Produkte der Form $U \times V$ einer offenen Teilmengen $U \subseteq X$ mit einer offenen Teilmenge $V \subseteq Y$ als Basis hat.

Lemma: Ein topologischer Raum X ist genau dann HAUSDORFFsch, wenn die Diagonale

$$\Delta \underset{\text{def}}{=} \left\{(x, y) \in X \times X \;\middle|\; x = y\right\}$$

bezüglich der Produkttopologie abgeschlossen in $X \times X$ ist.

Beweis: Δ ist genau dann abgeschlossen, wenn das Komplement

$$W = X \times X \setminus \Delta = \left\{(x, y) \in X \times X \;\middle|\; x \neq y\right\}$$

offen ist. In diesem Fall hat jeder Punkt (x, y) mit $x \neq y$ eine ganz in U liegende offene Umgebung, und diese wiederum enthält eine Basisumgebung der Form $U \times V$, wobei U eine offene Umgebung von x und U_2 eine von y in X ist. Da $U \times V$ in W liegt, ist $U \cap V = \emptyset$; der Raum X ist also HAUSDORFFsch.

Ist umgekehrt X HAUSDORFFsch, so gibt es zu je zwei verschiedenen Punkten $x, y \in X$ offene Umgebungen U von x und V von y mit $U \cap V = \emptyset$, d.h. $U \times V \subseteq W$. Somit ist W offen und Δ abgeschlossen. ∎

Kapitel 3
Zusammenhängende und kompakte Mengen

Den Begriff des Zusammenhangs können wir für allgemeine topologische Räume genauso definieren wie für Teilmengen eines $\mathbb{R}^n$:

Definition: *a)* Ein topologischer Raum X heißt zusammenhängend, wenn es keine zwei nichtleere, von X verschiedenen offene Teilmengen $U, V \subset X$ gibt mit $U \cap V = \emptyset$ und $U \cup V = X$.
b) Ein topologischer Raum X heißt *wegzusammenhängend,* wenn es zu je zwei Punkten $x, y \in X$ eine stetige Abbildung $f\colon I \to X$ aus dem Einheitsintervall $I = [0,\ 1] \subset \mathbb{R}$ gibt, so daß $f(0) = x$ und $f(1) = y$ ist.
c) Eine Teilmenge $Y \subseteq X$ heißt zusammenhängend *bzw.* wegzusammenhängend, wenn sie als topologischer Raum bezüglich der Spurtopologie diese Eigenschaft hat.

Ein topologischer Raum ist also genau dann zusammenhängend, wenn er sich nicht als Vereinigung zweier disjunkter echter offener Teilmengen U und V darstellen läßt: Ist $X = U \cup V$, so muß eine der beiden beiden Mengen gleich X sein und die andere leer. In einem wegzusammenhängenden Raum gibt zu je zwei Punkten x, y einen *Weg* von x nach y, nämlich das Bild der stetigen Funktion f.

Für zusammenhängende Mengen läßt sich *c)* auch äquivalent so formulieren, daß $Y \subseteq X$ genau dann zusammenhängend ist, wenn gilt: Sind U, V zwei offene Teilmengen von X mit leerem Durchschnitt $U \cap V \cap Y$ und liegt Y in $U \cup V$, so muß Y bereits in einer der beiden offenen Mengen liegen.

Lemma: Jeder wegzusammenhängende topologische Raum X ist zusammenhängend.

Beweis: U und V seien zwei offene Mengen und $X = U \cup V$. Falls $X = \emptyset$, folgt $X = U = V = \emptyset$, und wir sind fertig. Andernfalls gibt es mindestens einen Punkt $x \in X$. Dieser muß entweder in U oder in V liegen; indem wir gegebenenfalls die Bezeichnungen vertauschen, können wir annehmen, daß er in U liegt. Wir müssen zeigen, daß dann auch alle anderen Punkte $y \in X$ in U liegen.

Sei also $y \neq x$ ein beliebiger Punkt aus X. Wegen des Wegzusammenhangs gibt es eine stetige Abbildung $f\colon I \to X$ mit $f(0) = x$ und $f(1) = y$. Wir wollen uns überlegen, daß das gesamte Bild von f in U liegt. Dazu betrachten wir

$$s \underset{\text{def}}{=} \sup\big\{u \in [0,1] \;\big|\; f(t) \in U \text{ für alle } t \in [0,u)\big\}\,.$$

Falls $s < 1$ wäre, könnte $f(s)$ nicht in U liegen, denn wegen der Stetigkeit von f ist $f^{-1}(U)$ eine offene Menge, enthält also zu jedem ihrer Punkte auch eine offene Umgebung. Der Punkt $f(s)$ könnte aber auch nicht in V liegen, denn auch $f^{-1}(V)$ ist offen, und für alle $t \in [0,s)$ liegt $f(t)$ in U, also t in $f^{-1}(U)$ und damit nicht in der dazu disjunkten Menge $f^{-1}(V)$. Das ist ein Widerspruch, denn $X = U \cup V$, so daß $f(s)$ in einer der beiden Mengen liegen muß.

Somit ist $s = 1$ und $y = f(1) \in U$, denn läge $f(1)$ in V, gäbe es wegen der Stetigkeit von f ein $u < 1$, so daß $f(t)$ für alle $t > u$ in V und damit nicht in U läge. Dies widerspricht der Definition von s.

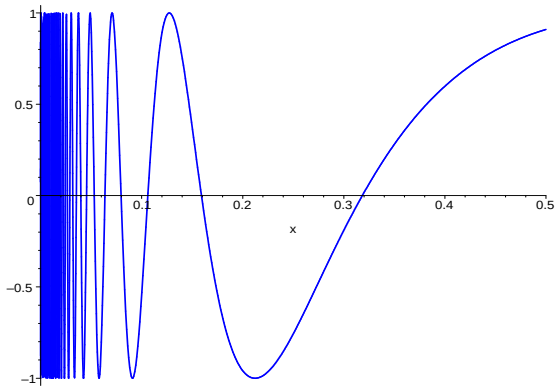
Da y ein beliebiger Punkt aus X war, liegt somit ganz X in U, d.h. $V = \emptyset$, und $X = U$ ist zusammenhängend.

■

Die Umkehrung dieses Lemmas gilt nicht; ein populäres Gegenbeispiel ist der Raum

$$X = \big\{(x, \sin\tfrac{1}{x}) \in \mathbb{R}^2 \;\big|\; x > 0\big\} \cup \big\{(0,y) \;\big|\; |y| \leq 1\big\} \subset \mathbb{R}^2$$

bestehend aus einer Sinuslinie mit für $x \to 0$ immer kleiner werdendem Abstand zwischen aufeinanderfolgenden Bergen und Tälern sowie dem Intervall $[-1,\ 1]$ auf der y-Achse, dem sich diese Sinuslinie immer mehr annähert.



Diese Menge ist nicht wegzusammenhängend; beispielsweise gibt es keinen Weg vom Punkt $\left(\frac{1}{\pi}, 0\right)$ auf der x-Achse zum Punkt $(0, 1)$ auf der y-Achse, obwohl beide Punkte in X liegen: Angenommen, es gäbe es eine stetige Funktion $f\colon [0,\ 1] \to X$ mit $f(0) = (\frac{1}{\pi}, 0)$ und $f(1) = (0, 1)$. Wir betrachten beispielsweise für $\varepsilon = \frac{1}{2}$ die Umgebung $(-\varepsilon,\ \varepsilon) \times (1 - \varepsilon,\ 1 + \varepsilon)$ von $(0, 1)$ in $\mathbb{R}^2$. Ihr Durchschnitt U mit X ist eine offene Umgebung von $(0, 1)$ in X, also müßte $f^{-1}(U)$ eine offene Umgebung der Eins in $[0,\ 1]$ sein. Insbesondere enthielte $f^{-1}(U)$ ein halboffenes Intervall $(u, 1]$ mit $u < 1$, für das $f(u)$ nicht auf der y-Achse liegt. Das ist aber nicht möglich, da $\sin \frac{1}{x}$ in jedem Intervall $(0,\ \varepsilon)$ alle Werte zwischen -1 und 1 annimmt. Somit ist X nicht wegzusammenhängend.

Die beiden Teilmengen

$$X_1 = \left\{(x, \sin \tfrac{1}{x}) \in \mathbb{R}^2 \;\middle|\; x > 0\right\} \quad \text{und} \quad X_2 = \left\{(0, y) \;\middle|\; |y| \leq 1\right\}$$

sind natürlich wegzusammenhängend und damit erst recht zusammenhängend. Wenn es zwei offene Mengen U, V mit leerem Durchschnitt gäbe, so daß $X \subseteq U \cup V$ weder in U noch in V läge, müßte daher X_1 in U und X_2 in V liegen oder umgekehrt. Das ist aber nicht möglich, denn in jeder offenen Menge, die X_2 enthält, gibt es auch Punkte aus X_1. Daher ist X zusammenhängend.

Lemma: Ist X ein zusammenhängender topologischer Raum und ist $Z \subseteq X$ sowohl offen als auch abgeschlossen, so ist $Z = \emptyset$ oder $Z = X$.

Beweis: Ist Z sowohl offen als auch abgeschlossen, so gilt das gleiche auch für sein Komplement $X \setminus Z$, und X ist die Vereinigung dieser beiden disjunkten offenen Mengen. Da X zusammenhängend ist, muß somit eine der beiden leer sein und die andere gleich X. ∎

Lemma: Für eine Teilmenge $X \subset \mathbb{R}$ sind die folgenden vier Aussagen äquivalent:
a) X ist ein Intervall.
b) X ist konvex.
c) X ist wegzusammenhängend.
d) X ist zusammenhängend.

Beweis: Die Folgerungen *a)* $\Rightarrow$ *b)* $\Rightarrow$ *c)* $\Rightarrow$ *d)* sind offensichtlich *bzw.* wurden (im letzten Fall) gerade allgemein gezeigt; wirklich beweisen müssen wir daher nur, daß jede zusammenhängende Teilmenge $X \subseteq \mathbb{R}$ ein Intervall ist. Da wir die leere Menge *per definitionem* als Intervall betrachten, können wir dazu annehmen, daß X nicht leer ist.

Falls X nach oben beschränkt ist, hat X ein Supremum $b \in \mathbb{R}$; wir wollen uns als erstes überlegen, daß X dann für jedes $z \in X$ das Intervall $[z, b)$ enthält: Gäbe es nämlich ein $c \in [z, b)$, das nicht in X läge, so wäre X die Vereinigung der beiden disjunkten nichtleeren offenen Mengen $U = \{x \in X \mid x < c\}$ und $V = \{x \in X \mid x > c\}$, also nicht zusammenhängend. Entsprechend folgt, daß X, sofern es unbeschränkt ist, zu jedem $z \in X$ auch alle reellen Zahlen $x \geq z$ enthalten muß, denn läge $x > z$ nicht in X, könnten wir wie eben argumentieren.

Dasselbe Argument zeigt auch, daß X, falls es nach unten beschränkt ist, für jedes $z \in X$ das Intervall $(a, z]$ enthalten muß, wobei a das Infimum von X bezeichnet; falls X nicht nach unten beschränkt ist, muß es entsprechend zu jedem $z \in X$ auch alle reellen Zahlen $x \leq z$ enthalten.

Ist also X eine beschränkte zusammenhängende Teilmenge von $\mathbb{R}$ mit Infimum a und Supremum b, so enthält X das offene Intervall (a, b). Da X keine Punkte $z < a$ oder $z > b$ enthalten kann, können dazu höchstens noch einer oder beide der Punkte a, b kommen; X ist also eines der vier Intervalle (a, b), $(a, b]$, $[a, b)$ oder $(a, b]$.

Falls X nur nach unten beschränkt ist mit Infimum a, enthält X auf jeden Fall alle reellen Zahlen $x > a$, zusätzlich eventuell nach den Punkt a. Somit ist $X = (a, \infty)$ oder $X = [a, \infty)$. Entsprechendes gilt für eine nur nach oben beschränkte Menge.

Bleibt noch der Fall, daß X weder nach oben noch nach unten beschränkt ist; dann ist $X = \mathbb{R} = (-\infty, \infty)$.

■

Die für uns wichtigste Anwendung zusammenhängender Mengen ist die folgende Verallgemeinerung des Zwischenwertsatzes:

Satz: Ist $f\colon X \to Y$ eine stetige Abbildung und ist X zusammenhängend, so ist auch $f(X)$ zusammenhängend.

Beweis: U und V seien zwei disjunkte offene Teilmengen von Y, und $f(X)$ liege in der Vereinigung $U \cup V$. Wegen der Stetigkeit von f sind die Urbilder $f^{-1}(U)$ und $f^{-1}(V)$ offene Teilmengen von X, die natürlich ebenfalls disjunkt sind, deren Vereinigung aber gleich X ist. Somit muß X gleich einer dieser beiden Mengen $f^{-1}(U)$ oder $f^{-1}(V)$ sein, also $f(X)$ ganz in U oder in V liegen.

■

Speziell für $Y = \mathbb{R}$ folgt

Korollar: Ist $f\colon X \to \mathbb{R}$ eine stetige Abbildung eines zusammenhängenden topologischen Raums X nach $\mathbb{R}$, so ist $f(X)$ ein Intervall.

■

Bei der Kompaktheit ist die Situation etwas komplizierter: In manchen Lehrbüchern der Analysis werden kompakte Mengen definiert als abgeschlossene und beschränkte Teilmengen eines $\mathbb{R}^n$. Damit können wir nichts anfangen, denn in allgemeinen topologischen Räumen können wir nicht von Beschränktheit reden. Nach dem Satz von HEINE-BOREL ist aber eine Teilmenge eines $\mathbb{R}^n$ genau dann abgeschlossen und beschränkt, wenn jede offene Überdeckung eine endliche Teilüberdeckung hat. Dies könnten wir für beliebige topologische Räume zur Definition der Kompaktheit verwenden. Da wir im Zusammenhang mit Kompaktheit aber vor allem Aussagen über Konvergenz beweisen wollen, empfiehlt es sich, noch zusätzlich die (für Teilmengen eines $\mathbb{R}^n$ selbstverständliche)

HAUSDORFF-Eigenschaft zu fordern. Es gibt inzwischen allerdings auch einige (wenige) Lehrbücher der Topologie, die darauf verzichten; im Zweifelsfall sollte man sich daher immer die Definitin im jeweiligen Lehrbuch anschauen. In der überwältigenden Mehrheit der Literatur ist die Definition die folgende:

Definition: X sei ein topologischer Raum
a) Ein System $\mathfrak{U} = \{U_i \mid i \in I\}$ von offenen Teilmengen $U_i \subseteq X$ mit i aus einer beliebigen Indexmenge I heißt *offene Überdeckung* von X, wenn $X = \bigcup_{i \in I} U_i$ die Vereinigungsmenge aller U_i ist.
b) Ist $J \subseteq I$ eine Teilmenge von I und ist X bereits die Vereinigung aller U_i mit $i \in J$, bezeichnen wir $\mathfrak{V} = \{U_i \mid i \in J\}$ als eine *Teilüberdeckung* von $\mathfrak{U}$. Ist speziell J eine endliche Menge, so sprechen wir von einer *endlichen Teilüberdeckung.*
c) X heißt *quasikompakt,* wenn jede offene Überdeckung $\mathfrak{U}$ von X eine endliche Teilüberdeckung hat.
d) Ein quasikompakter topologischer Raum X heißt *kompakt,* wenn er HAUSDORFFsch ist.
e) Eine Teilmenge $Y \subseteq X$ eines topologischen Raums heißt quasikompakt *bzw.* kompakt, wenn sie bezüglich der Spurtopologie diese Eigenschaft hat.

Letzteres können wir wieder wie beim Zusammenhang auch unabhängig vom Begriff der Spurtopologie formulieren: $Y \subseteq X$ ist quasikompakt genau dann, wenn es für jedes System von (in X) offenen Mengen U_i mit $Y \subseteq \cup_{i \in I} U_i$ eine endliche Teilmenge $J \subseteq I$ gibt, so daß Y bereits in der Vereinigung der endlich vielen U_i mit $i \in J$ liegt. Wir reden auch in dieser Situation von einer offenen Überdeckung von Y und einer endlichen Teilüberdeckung.

Man beachte, daß es auch Lehrbücher gibt, die bei der Definition der Kompaktheit auf die HAUSDORFF-Bedingung verzichten, also nur das fordern, was hier als Quasikompaktheit bezeichnet wird. Für Teilmengen eines $\mathbb{R}^n$ (oder allgemeiner eines HAUSDORFF-Raums) gibt es keinen Unterschied zwischen Quasikompaktheit und Kompaktheit; insbesondere haben wir im $\mathbb{R}^n$ genau die aus der Analysis bekannten kompakten Mengen.

Lemma: Jede abgeschlossene Teilmenge $Z \subset X$ eines kompakten topologischen Raums X ist kompakt.

Beweis: Da X HAUSDORFFsch ist, gilt das gleiche für Z. Weiter sei $\mathfrak{U} = \{U_i \mid i \in I\}$ eine offene Überdeckung von Z. Die U_i sind offen bezüglich der Spurtopologie; es gibt also offene Mengen $V_i \subseteq X$, so daß $U_i = V_i \cap Z$ ist. Wegen der Abgeschlossenheit von Z ist auch $V = X \setminus Z$ offen; nehmen wir V noch mit dazu, erhalten wir eine offene Überdeckung $\mathfrak{V} = \{V_i \mid i \in I\} \cup \{V\}$ von X, denn jeder Punkt aus $X \setminus Z$ liegt in $V = X \setminus Z$, und jeder Punkt aus Z liegt in mindestens einer der Mengen U_i, also erst recht im zugehörigen V_i.

Da X kompakt ist, hat $\mathfrak{V}$ eine endliche Teilüberdeckung (von X). Wenn diese Teilüberdeckung ohne die Menge V auskommt, bilden die mit Z geschnittenen Überdeckungsmengen gleichzeitig eine endliche Teilüberdeckung von $\mathfrak{U}$ für die Menge Z. Andernfalls betrachten wir die Teilüberdeckung *ohne* die Menge V. Das ist dann eine endliche Teilmenge von $\mathfrak{U}$, und es ist auch eine offene Überdeckung von Z, denn jeder Punkt von $Z \subseteq X$ muß in einer der offenen Mengen aus der Teilüberdeckung liegen, und er liegt sicher nicht in $V = X \setminus Z$. Somit hat $\mathfrak{U}$ eine endliche Teilüberdeckung von Z.

■

Umgekehrt gilt

Lemma: Jede kompakte Teilmenge Z eines HAUSDORFF-Raums X ist abgeschlossen.

Zum *Beweis* zeigen wir, daß $X \setminus Z$ offen ist. Dazu sei x ein beliebiger Punkt aus diesem Komplement und z ein beliebiger Punkt aus Z. Wegen der HAUSDORFF-Eigenschaft gibt es disjunkte offene Umgebungen U_z von z und V_z von x.. Insbsondere ist auch $U_z \cap (V_z \cap Z) = \emptyset$. Die sämtlichen Mengen $V_z \cap Z$ bilden eine offene Überdeckung von Z in seiner Spurtopologie, haben also eine Teilüberdeckung durch endlich viele Mengen $V_{z_1} \cap Z, \ldots, V_{z_m} \cap Z$. Dann ist der Durchschnitt der Umgebungen $U_{z_1}, \ldots, U_{z_m}$ disjunkt zu jedem V_{z_i}, also auch zu Z, und ist somit eine Umgebung von x, die ganz in $X \setminus Z$ liegt. Also ist $X \setminus Z$ offen und Z abgeschlossen.

■

Lemma: $f\colon X \to Y$ sei eine stetige Abbildung in den HAUSDORFF-Raum Y, und $Z \subseteq X$ sei kompakt. Dann ist auch $f(Z) \subseteq Y$ kompakt.

Beweis: $\mathfrak{U} = \{U_i \mid i \in I\}$ sei ein System von in Y offenen Teilmengen, die eine Überdeckung von $f(Z)$ bilden, d.h.

$$f(Z) \subseteq \bigcup_{i \in I} U_i \,.$$

Da f eine stetige Abbildung ist, sind dann auch die Urbilder $f^{-1}(U_i)$ offen in X, und natürlich bilden sie eine Überdeckung von Z. Diese Überdeckung hat wegen der Kompaktheit von Z eine endliche Teilüberdeckung $\{f^{-1}(U_{i_1}), \ldots, f^{-1}(U_{i_r})\}$. Damit überdecken $U_{i_1}, \ldots, U_{i_r}$ die Menge $f(Z)$; diese ist also quasikompakt. Als Teilmenge eines HAUSDORFF-Raums ist sie auch HAUSDORFFsch, also kompakt. ∎

Lemma: $f\colon X \to \mathbb{R}$ sei eine stetige Abbildung, und $Z \subseteq X$ sei kompakt. Dann nimmt f auf Z sowohl sein Maximum als auch sein Minimum an; es gibt also Elemente z_m und z_M aus Z, so daß für alle $z \in Z$ gilt: $f(z_m) \leq f(z) \leq f(z_M)$.

Beweis: Nach dem vorigen Lemma ist $f(Z)$ eine kompakte Teilmenge von $\mathbb{R}$. Diese muß beschränkt sein, denn auch eine Überdeckung durch offene Einheitskugeln muß eine endliche Teilüberdeckung haben. Somit existieren sowohl das Infimum m als auch das Supremum M von $f(Z)$. Wir müssen zeigen, daß beide in $f(Z)$ liegen.

Nach den beiden vorigen Lemmata ist $f(Z)$ abgeschlossen. Läge einer dieser beiden Punkte nicht in $f(Z)$, läge er daher im offenen Komplement $\mathbb{R} \setminus f(Z)$ und hätte damit eine ε-Umgebung, die ganz in $\mathbb{R} \setminus f(Z)$ läge. Im Falle des Supremums würde dies bedeuten, daß beispielsweise auch $M - \frac{\varepsilon}{2}$ eine obere Schranke von $f(Z)$ wäre, im Widerspruch zur Definition des Supremums als *kleinster* oberer Schranke; im Falle des Infimums wäre entsprechend $m + \frac{\varepsilon}{2}$ eine untere Schranke.

Daher müssen m und M in $f(Z)$ liegen, es gibt also Elemente z_m und z_M in Z, so daß $f(z_m) = m$ und $f(z_M) = M$ ist. ∎

Als Anwendung können wir für kompakte Teilmengen von $\mathbb{R}^n$ die Existenz der sogenannten LEBESGUE-Zahl beweisen.

Definition: Der Durchmesser einer Teilmenge $M \subseteq \mathbb{R}^n$ ist

$$d(M) = \sup_{x,y \in M} d(x,y)\,,$$

wobei $d(x,y)$ den EUKLIDischen Abstand zweier Punkte $x, y \in \mathbb{R}^n$ bezeichnet. Der Abstand eines Punkts $x \in \mathbb{R}^n$ von einer Teilmenge $M \subseteq \mathbb{R}^n$ ist

$$d(x,M) = \inf_{y \in M} d(x,y)\,.$$

Lemma: Ist $Z \subset \mathbb{R}^n$ kompakt und ist $\mathfrak{U}$ eine offene Überdeckung von Z, so gibt es eine reelle Zahl $\delta > 0$, die LEBESGUE-Zahl, derart, daß jede Teilmenge $M \subseteq Z$ mit $d(M) < \delta$ ganz in einer der Mengen $U \in \mathfrak{U}$ enthalten ist.

Beweis: Wegen der Kompaktheit von Z hat $\mathfrak{U}$ eine endliche Teilüberdeckung $\{U_1, \ldots, U_r\}$. Wir betrachten die Komplemente $A_i = Z \setminus U_i$ dieser Überdeckungsmengen und die Funktion

$$f\colon \begin{cases} Z \to \mathbb{R} \\ x \mapsto \dfrac{1}{r} \displaystyle\sum_{i=1}^{r} d(x, A_i) \end{cases}.$$

Wegen der Stetigkeit von d ist auch f stetig, nimmt also auf Z sein Minimum δ an, etwa im Punkt z. Da die U_i eine Überdeckung bilden, gibt es mindestens eine Menge U_j, die z enthält.

Wir wollen uns zunächst überlegen, daß δ positiv ist: Wäre $d(z, A_j) = 0$, so gäbe es für jedes $n \in \mathbb{N}$ ein $a_n \in A_j$ mit $d(z, a_n) < \frac{1}{n}$, und die Folge der a_n würde gegen z konvergieren. Da A_j kompakt ist, müßte der Grenzwert z in A_j liegen, was wegen $z \in U_j$ nicht möglich ist. Somit ist $d(z, A_j) > 0$, und da alle $d(z, A_i) \geq 0$ sind, ist auch $\delta = \frac{1}{r} \sum_{i=1}^{r} d(z, A_i) > 0$.

Ist $M \subseteq Z$ eine Teilmenge mit $d(M) < \delta$, und ist $x \in M$ ein beliebiges Element, so liegt M ganz in der Kugel um x mit Radius $d(M)$, also erst recht in der mit Radius δ. Nun ist δ das Minimum von f, also muß $f(x) \geq \delta$ sein. Der durchschnittliche Abstand von x zu den A_i ist also mindestens gleich δ; daher muß es ein A_j geben mit $d(x, A_j) \geq \delta$.

Damit liegt die Kugel um x mit Radius $d(M)$ außerhalb von A_j, also liegt diese Kugel und damit erst recht die Menge M ganz in U_j.

■

HENRI LÉON LEBESGUE (1875–1941) studierte ab 1894 an der Ecole Normale Supérieure. Nach seinem ersten Abschluß 1897 studierte er noch zwei Jahre auf eigene Faust weiter, hauptsächlich in der Bibliothek, bis er 1899 eine Stelle als Gymnasiallehrer in Nancy erhielt. Dort schrieb er 1901 seine Arbeit über die Definition des LEBESGUE-Integrals. 1902 promovierte er mit einer weiteren Arbeit in Paris und bekam daraufhin eine Stelle an der Universität von Rennes. 1906 wechselte er nach Poitiers, 1910 an die Sorbonne in Paris. Ab 1921 war er Professor am Collège de France. Seine Arbeiten befassen sich mit Integrationstheorie, Potentialtheorie, Variationsrechnung, Topologie und anderen Gebieten.

Wie aus der Analysis gewohnt, gilt auch im allgemeinen Fall

Lemma: X sei ein topologischer Raum, Z eine kompakte Teilmenge von X und $(z_n)_{n\in\mathbb{N}}$ eine Folge von Elementen aus Z. Dann gibt es (mindestens) eine gegen einen Punkt $z \in Z$ konvergente Teilfolge $(z_{n_\nu})_{\nu\in\mathbb{N}}$.

Beweis: Falls es nur endlich viele verschiedene Werte z_n gibt, können wir eine konstante und damit erst recht konvergente Teilfolge finden. Wir können daher annehmen, daß es unendlich viele verschiedene z_n gibt. Gäbe es keine gegen einen Punkt $z \in Z$ konvergente Teilfolge, hätte jeder Punkt $z \in Z$ eine offene Umgebung U_z, die höchstens endlich viele z_n enthielte. Das System dieser Umgebungen wäre eine Überdeckung der Menge Z, die wegen deren Kompaktheit eine endliche Teilüberdekkung hätte. Z wäre somit enthalten in einer endlichen Vereinigung von Mengen, von denen jede höchstens eine endliche Anzahl der unendlich vielen Elemente $z_n \in Z$ enthielte, ein Widerspruch.

■

Lemma: Ist X ein HAUSDORFF-Raum und Z eine kompakte Teilmenge von X, so gibt es zu jedem Punkt $x \notin Z$ offene Mengen $U, V \subseteq X$, so daß $Z \subset U$, $x \in V$ und $U \cap V = \emptyset$ ist.

Beweis: Da X ein HAUSDORFF-Raum ist, gibt es zu jedem $z \in Z$ offene Umgebungen U_z von z und V_z von x, so daß $U_z \cap V_z = \emptyset$

ist. Die sämtlichen U_z überdecken natürlich Z; da dies eine kompakte Menge ist, reichen dazu bereits endlich viele dieser Umgebungen. Es gibt daher endlich viele Punkte $z_1, \ldots, z_r$, so daß Z in der Vereinigung der Mengen U_{z_i} liegt. Da U_{z_i} leeren Durchschnitt mit V_{z_i} hat, ist diese Umgebung erst recht disjunkt zum Durchschnitt $V = V_{z_1} \cap \cdots \cap V_{z_r}$ aller V_{z_i}, und damit ist auch $U \cap V = \emptyset$. Da Z in U liegt und x in V, folgt die Behauptung.

■

Korollar: Ist X ein HAUSDORFF-Raum und Z eine kompakte Teilmenge von X, so ist Z der Durchschnitt aller in X offener Mengen, die Z enthalten.

Beweis: Wir müssen wir uns nur überlegen, daß es zu jedem Punkt $x \notin Z$ eine offene Menge $U \supseteq Z$ gibt, die x nicht enthält, und das ist nach dem gerade bewiesenen Lemma klar.

■

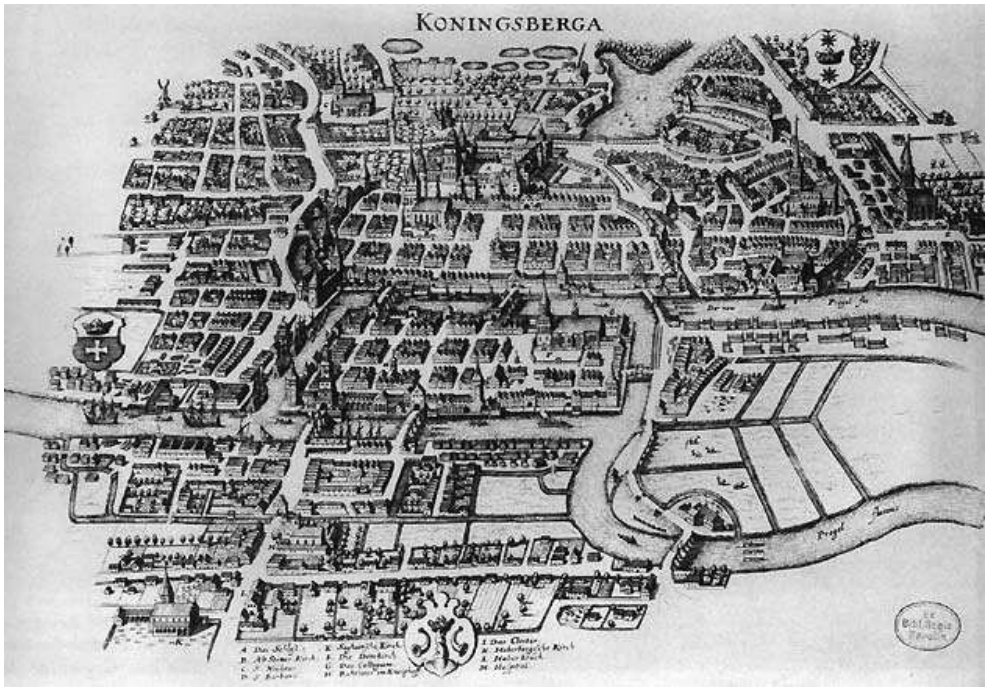
Kapitel 4
Auf dem Weg zur algebraischen Topologie

Die Aussagen der Topologie, die wir bislang kennengelernt haben, sind eher qualitativer Natur; die uralte Maxima der Pythagoräer *Alles ist Zahl* scheint hier nicht zu gelten. Andererseits zeigt die Erfahrung aus mehreren Jahrtausenden Mathematik, daß die erfolgreiche Lösung komplexer mathematischer Probleme nur selten ganz ohne Rechnung möglich ist. Schon lange bevor die Topologie als Teilgebiet der Mathematik etabliert war, wurden daher topologische Fragestellungen in Zahlen übersetzt und zumindest teilweise so gelöst. In diesem Kapitel soll dies anhand einiger einfach zu formulierender klassischer Probleme illustriert werden.

§1: Das Königsberger Brückenproblem

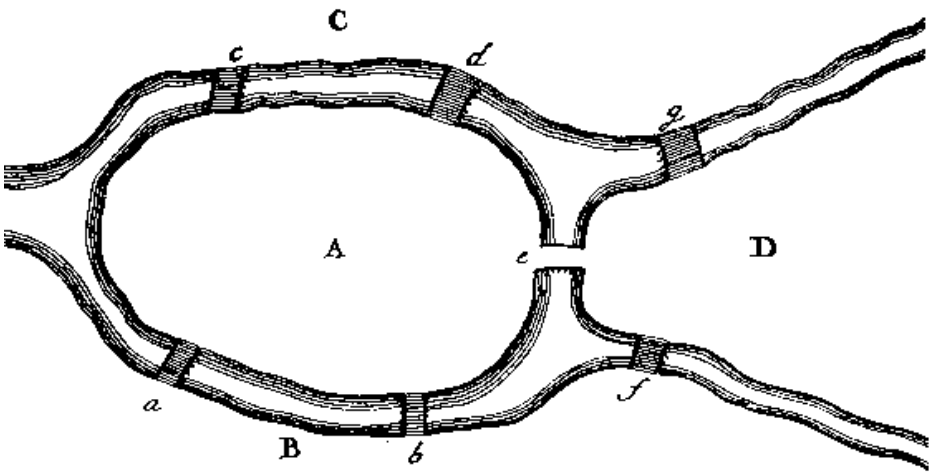
Wie bereits zu Beginn der Vorlesung erwähnt, war EULERs Lösung des Königsberger Brückenproblems die erste Arbeit, die nach heutigem Verständnis der Topologie zuzuordnen ist. Die Abbildung auf der nächsten Seite zeigt Königsberg nach einem 1652 von MERIANs Söhnen veröffentlichten Kupferstich.

Damals führten sieben Brücken über den Pregel, der sich innerhalb der Stadt zudem noch teilte und so eine Insel abtrennte, den Kneiphof. Eine der damals in Königsberg (sicherlich nicht mit höchster Priorität) diskutierten Fragen war, ob man seinen Sonntagsspaziergang so legen könne, daß man jede der sieben Brücken genau einmal überquere. Da es $7! = 5040$ mögliche Reihenfolgen gibt, wäre es unrealistisch gewesen, diese allesamt systematisch durchzuprobieren: Kaum jemand erlebt 5040 Sonntage, und auch das theoretische Durchprobieren war, da es damals noch keine Navigationscomputer gab, kaum praktikabel. Durch Zufall oder Intuition hatte noch niemand einen solchen Weg gefunden.



In seiner Arbeit *Solutio problematis ad geometriam situs pertinentis,* die 1736 in den Annalen der St. Petersburger Akademie erschien, erklärte EULER warum: Es kann keinen solchen Sonntagsspaziergang geben.

Als echter Mathematiker vereinfachte er zunächst die Geographie von Königsberg radikal auf das, was für die Lösung des Brückenproblems wirklich wesentlich ist. Die unten reproduzierte Zeichnung aus seiner Arbeit ist ästhetisch sicherlich keine Konkurrenz zu MERIAN, für die mathematische Behandlung des Problems aber deutlich nützlicher.



Er bennent dort auch die vier durch Arme des Pregel getrennten Teile Königsbergs durch die Großbuchstaben A bis D, die sieben Brücken durch die Kleinbuchstaben a bis g. Sodann zählt er, wie viele Brücken in jedes der vier Gebiete führen. Der oben erwähnte Kneiphof beispielsweise ist das Gebiet A, und dorthin führen die fünf Brücken a, b, c, d und e.

Nehmen wir an, ein Königsberger beginnt seinen Sonntagsspaziergang auf dem Kneiphof. Dann muß jeder Weg, der genau einmal über jede dieser fünf Brücken führt, außerhalb des Kneiphofs enden. Umgekehrt muß auch jeder entsprechende Weg, der außerhalb des Kneiphofs beginnt, in diesem enden. Somit kann es keinen Sonntagsspaziergang mit gleichem Anfangs- und Endpunkt geben, der genau einmal über jede der sieben Brücken führt.

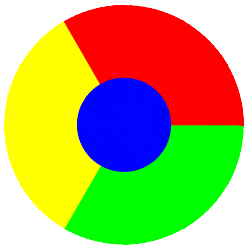
Nun betrachtete aber wohl nicht jeder Königsberger seinen Sonntagsspaziergang als Rundgang; mancher Bürger hatte vielleicht eher das Haus eines Verwandten oder eine Gaststätte als Ziel. Aber auch ein solcher Weg ist unmöglich, denn für jedes der vier Gebiete A, B, C, D gibt es eine ungerade Anzahl von Brücken, die dieses Gebiet mit dem Rest von Königsberg verbinden. Jedes der vier Gebiete müßte daher entweder Anfangs- oder Endpunkt des Spaziergangs sein, was natürlich nicht möglich ist.

§2: Kartenfärbungsprobleme

Mehr als hundert Jahre später, wahrscheinlich um 1852, kolorierte FRANCIS GUTHRIE eine Karte der Grafschaften von England und gewann dabei den Eindruck, daß er jede Karte mit nur vier Farben so einfärben könne, daß keine zwei benachbarte Gebiete die gleiche Farbe haben. Mit weniger als vier Farben kann man offensichtlich nicht auskommen: Betrachtet man etwa auf einer Europakarte die Staaten Deutschland, Frankreich, Belgien und Luxemburg, so hat jeder der vier mit jedem anderen eine gemeinsame Grenze, so daß jeder dieser vier Staaten seine eigene Farbe braucht. Schematisch ist diese Situation in der folgenden

Karte dargestellt.

Karte 1



Natürlich war GUTHRIE ein Mathematiker: Ein Kartograph, der Farben benutzt, um möglichst viel Information darzustellen, käme nie auf die Idee, die Vielfalt der zur Verfügung stehenden Farben künstlich einzuschränken.

FRANCIS GUTHRIE (1831–1899) studierte zunächst Mathematik (B.A. 1850), dann Jura (LL.B. 1852) am University College London; 1856 wurde er *fellow*. Er arbeitete zunächst als Rechtsanwalt in London, bis er 1861 eine Mathematikprofessur am Graaf-Reinet College in Südafrika annahm. Später wechselte er nach Kapstadt. Er interessierte sich nicht nur für Mathematik, sondern gab auch öffentliche Vorlesungen über Botanik und erforschte die südafrikanische Pflanzenwelt. Mehrere Pflanzenarten sind nach ihm benannt. Auf seinen Wanderungen durch die Berge erforschte er auch das Gelände für den Bau einer geplanten Eisenbahn.

FRANCIS GUTHRIE hatte sein Mathematikstudium bereits 1850 abgeschlossen; 1852 studierte er Jura. Deshalb wandte er sich an seinen jüngeren Bruder FREDERICK GUTHRIE, der damals, ebenfalls am University College, bei AUGUSTUS DE MORGAN Mathematik studierte. DE MORGAN war sofort begeistert, als ihm FREDERICK GUTHRIE am 23. Oktober 1852 die Vermutung seines Bruders präsentierte; er schrieb noch am selben Tag einen Brief an den wohl bekanntesten Mathematiker der damaligen Zeit auf den britischen Inseln, Sir WILLIAM ROWAN HAMILTON vom Trinity College in Dublin, dessen Entdeckung der Quaternionen ihn auch bei einem großen nichtmathematischen Publikum berühmt gemacht hatte. HAMILTON hatte allerdings kein großes Interesse und schrieb zurück: *I am not likely to attempt your quaternion of colour very soon.*

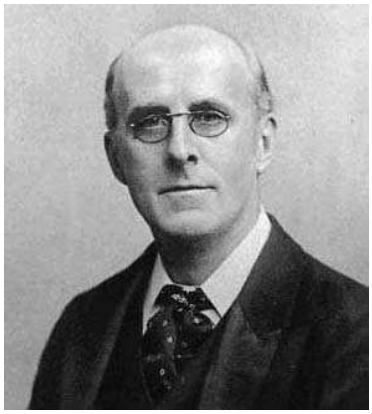
AUGUSTUS DE MORGAN (1806–1871) begann 1823 im Alter von 16 Jahren sein Mathematikstudium in Cambridge, das er nur mit einem Bachelor-Grad abschloß, da er die für einen Master notwendige Theologieprüfung ablehnte. Danach begann er in London ein Jurastudium, bewarb sich aber 1827 auch für eine Professur am neu gegründeten University College, wo er 1828 der erste Mathematikprofessor wurde. Zweimal, 1831 und 1866, legte er diese Professur aus Protest gegen die Diskriminierung von Kollegen nieder; 1836 wurde er neu berufen. Seine Arbeiten beschäftigen sich hauptsächlich mit Analysis und Logik.

SIR WILLIAM ROWAN HAMILTON (1805–1865) ging nie zur Schule. Trotzdem sprach er schon mit fünf Jahren griechisch, lateinisch und hebräisch; mit 14 beherrschte er über ein Dutzend Sprachen. Als 13-Jähriger las er das (französischsprachige) Algebra-Buch von CLAIRAUT. Mit 18 kam er ans Trinity College in Dublin, wo er bereits als Student Arbeiten über Optik publizierte. 1827, noch vor Abschluß seines Studiums, ernannte ihn das Trinity College zum Professor der Astronomie. Sein Interesse an Astronomie war allerdings gering, und er verbrachte den Großteil seiner Zeit mit Mathematik. Seine bekannteste Entdeckung sind die Quaternionen.

Briefe von DE MORGAN zeigen, daß dieser sich auch in den folgenden Jahren noch häufig mit dem Problem beschäftigte und es auch bekannt machte, danach gab es jedoch keine weiteren Aktivitäten mehr bis ARTHUR CAYLEY 1879 einen Übersichtsartikel in den *Proceedings of the Royal Geographical Society* veröffentlichte.

Ein Jahr später veröffentlichte ein ehemaliger Student CAYLEYs, der Mathematiker und Jurist Sir ALFRED BRAY KEMPE, im *American Journal of Mathematics* einen Beweis der Vermutung. Diese Zeitschrift war gerade erst 1878 von dem englischen Mathematiker JAMES JOSEPH SYLVESTER an der selbst erst zwei Jahre vorher in Baltimore gegründeten *Johns Hopkins University* begründet worden. Diese Universität sollte sich zwar in den folgenden Jahrzehnten zu einem der führenden Zentren der amerikanischen Mathematik entwickeln, und das *American Journal* gehört heute zu den international angesehensten mathematischen Fachzeitschriften, damals aber war es naturgemäß insbesondere in Europa noch kaum be-

kannt. Da CAYLEY und SYLVESTER nicht nur mathematisch zusammenarbeiteten, sondern auch gute Freunde waren, sollte KEMPE seine Arbeit wohl dort veröffentlichen, um die Zeitschrift bekannter zu machen. Tatsächlich führte dies wohl eher dazu, daß die Arbeit kaum gelesen wurde – auch wenn allgemein anerkannt wurde, daß KEMPE das Problem gelöst hatte.



SIR ALFRED BRAY KEMPE (1849-1922) studierte Mathematik am Trinity College in Cambridge; nach seinem Abschluß 1872 begann er jedoch eine juristische Ausbildung und arbeitete auch Zeit seines Lebens als Jurist. Insbesondere war er einer der führenden Experten für Kirchenrecht und hatte in dieser Eigenschaft viele Funktionen. Mathematik und Musik waren seine beiden großen Hobbies; seine mathematischen Arbeiten befassen sich vor allem mit Kinematik und der Philosophie der Mathematik, einige auch mit Algebra und eine mit dem Vierfarbenproblem.

Erst 1890 fand der englische Mathematiker JOHN PERCY HEAWOOD einen Fehler im Beweis. Dabei handelte es sich nicht nur um eine Lücke, sondern HEAWOOD konnte ein Beispiel einer Karte konstruieren, für die KEMPEs Konstruktion definitiv nicht funktionierte – auch wenn zumindest diese Karte trotzdem mit vier Farben gefärbt werden konnte. Damit war die Vierfarbenvermutung wieder offen. Immerhin reichten KEMPEs Argumente aus, um zu beweisen, daß man stets mit fünf Farben auskommt. Außerdem verallgemeinerte HEAWOOD das Problem auf Karten, die nicht mehr eben (oder auf dem Globus, d.h. einer Kugel) sind, sondern beispielsweise auf einem Torus, einer dicken Acht oder einer Brezel, *usw.* (Mathematiker reden von Flächen vom Geschlecht eins, zwei, drei, *usw.;* die Kugel hat Geschlecht null.) Er zeigte, daß man für eine Fläche vom Geschlecht $g \geq 1$ immer mit

$$\left[\frac{7+\sqrt{1+48g}}{2}\right]$$

auskommt und behauptete, daß man diese auch wirklich braucht – was erst 1968 von RINGEL und YOUNG vollständig bewiesen wurde. In dieser Formel bezeichnet $[x]$ die größte ganze Zahl kleiner oder gleich x, für

den Torus braucht man also $\frac{7+7}{2} = 7$ Farben, für die Acht

$$\left[\frac{7+\sqrt{97}}{2}\right] = \left[\frac{7+9{,}8488578\ldots}{2}\right] = 8$$

und für die Brezel

$$\left[\frac{7+\sqrt{145}}{2}\right] = \left[\frac{7+12{,}01459\ldots}{2}\right] = 9\,,$$

und so weiter. Für $g = 0$ erhält man den erwarteten Wert $\left[\frac{4+4}{2}\right] = 4$, aber wie wir in §5 sehen werden, funktioniert HEAWOODs Beweis nur für $g \geq 1$.

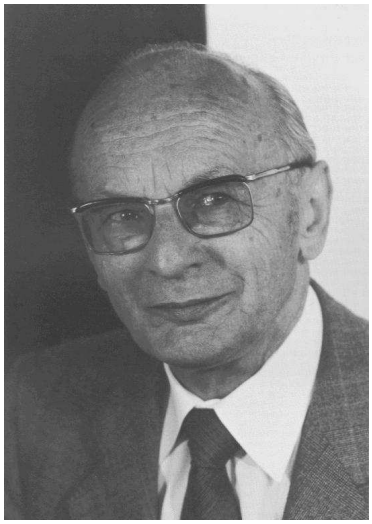
PERCY JOHN HEAWOOD O.B.E. (1861–1955) studierte am Exeter College in Oxford. 1887 wurde er Dozent an der heutigen Universität Durham, an der er bis zum Alter von 78 Jahren arbeitete, seit 1911 auf einem Lehrstuhl für Mathematik, von 1926-1928 auch als *Vice Chancellor.* Seine mathematische Forschung drehte sich ausschließlich um das das Vierfarbenproblem, bis zu seiner letzten Arbeit, der er 1949 als 88-Jähriger veröffentlichte. Daneben schrieb er auch mehrere Arbeiten für Schüler und interessierte sich sehr für den Mathematikunterricht. Seinen O.B.E. erhielt er für seine Aktivitäten zur Rettung des Schlosses von Durham.

Was das eigentliche Vierfarbenproblem betrifft, gab es lange Zeit nur kleine Fortschritte. Die grundlegende Idee der KEMPEschen Arbeit wurde von verschiedenen Mathematikern verfeinert und führte zu Beweisen für immer größere Klassen von Graphen, aber von einer allgemeinen Lösung war man immer noch weit entfernt.

Einen wesentlich neuen Aspekt brachte erst Anfang kurz nach 1960 der deutsche Mathematiker HEINRICH HEESCH ins Spiel: Er fand eine Möglichkeit, wie man Computer an der Beweisarbeit beteiligen konnte.

Erste Experimente zeigten, daß der Ansatz grundsätzlich funktionieren könnte, daß allerdings der Rechenaufwand für eine vollständige Lösung enorm sein müßte: Die meisten betrachteten Fälle führten auf den Computern der damaligen Zeit zu stundenlangen Rechenzeiten. Da Computer damals nicht nur sehr langsam, sondern vor allem auch extrem teuer

waren, hatte auch eine technische Hochschule wie Hannover nur ein zentrales Rechenzentrum mit einem einzigen Computer für die gesamte Universität. Langwierige Rechnungen waren somit sehr teuer und nur über Forschungsgelder der DfG (Deutsche Forschungsgemeinschaft) finanzierbar. Aus diesem Grund, und auch weil einige DfG-Gutachter Zweifel am endgültigen Erfolg des Projekts hatten und die Mittel daher eher zögerlich flossen, kamen die Forschungen von HEESCH und seinen Mitarbeitern nur sehr langsam voran.

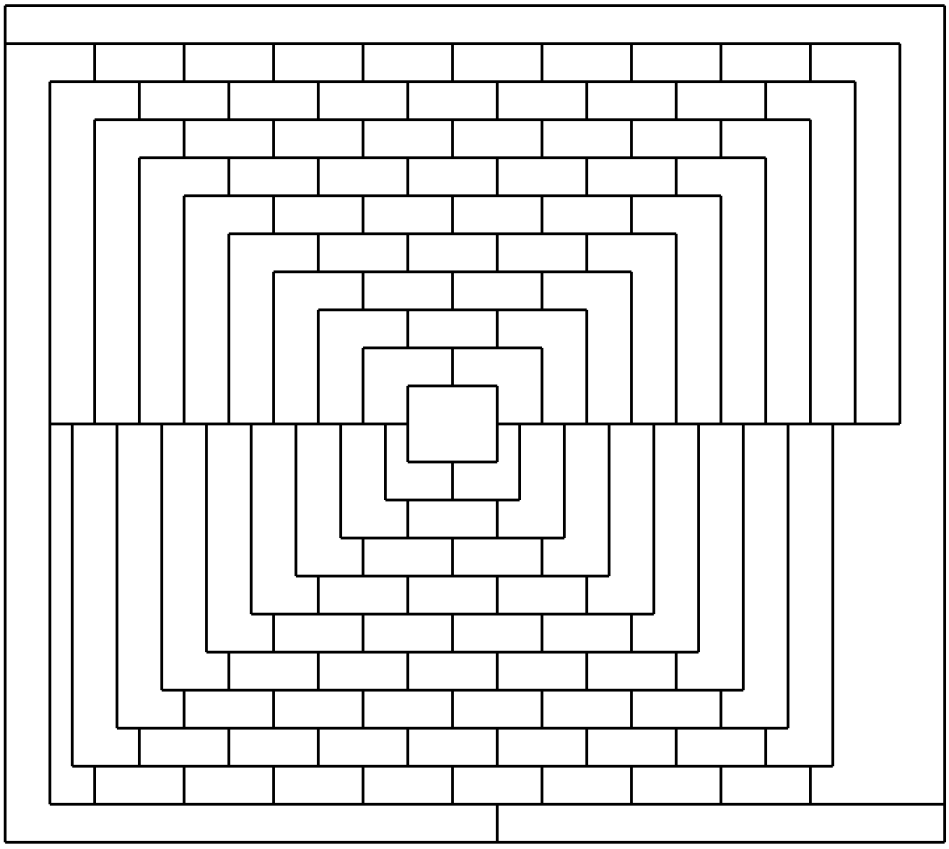


HEINRICH HEESCH (1906–1995) promovierte 1929 in Zürich mit einer Arbeit über Kristallographie; danach wurde er Assistent von HERMANN WEYL in Göttingen. Nachdem dieser 1933 Deutschland verlassen hatte, wurde er mit seiner Vertretung beauftragt. Da er sich weigerte, dem nationalsozialistischen Dozentenbund beizutreten, konnte er sich aber nicht habilitieren und mußte 1935 die Universität verlassen. Bis nach Kriegsende arbeitete er als Mathematik- und Musiklehrer. 1955 ging er an die TH Hannover, habilitierte 1958 und wurde 1966 zum Professor ernannt. Seine Arbeiten beschäftigen sich nicht nur mit dem Vierfarbenproblem, sondern auch mit anderen geometrischen Fragen wie etwa regulären Parkettierungen der Ebene.

In mindestens einem Lehrbuch der damaligen Zeit wurden allerdings auch Zweifel geäußert, ob die Vierfarbenvermutung überhaupt richtig sei, und in der Tat veröffentlichte MARTIN GARDNER am ersten April 1975 in seiner Mathematikkolumne im *Scientific American* das unten abgedruckte Beispiel einer Karte mit 110 Ländern, für die man mindestens fünf Farben braucht.

Der Artikel rief ein großes Echo hervor: Mehr als Tausend Leserbriefe gingen beim *Scientific American* ein, darunter auch einige hundert, die eine Vierfärbung der angegebenen Karte enthielten. GARDNERs Kolumne in der fraglichen Ausgabe hatte also mehr mit dem Datum zu tun als mit der Korrektheit der Vierfarbenvermutung.

Ein Jahr später veröffentlichten WOLFGANG HAKEN und KENNETH APPEL einen computergestützten Beweis der Vierfarbenvermutung, zurückgehend auf sowohl eine frühere Zusammenarbeit von HAKEN mit HEESCH, als auch auf zahlreiche darüber hinausgehende mathematische



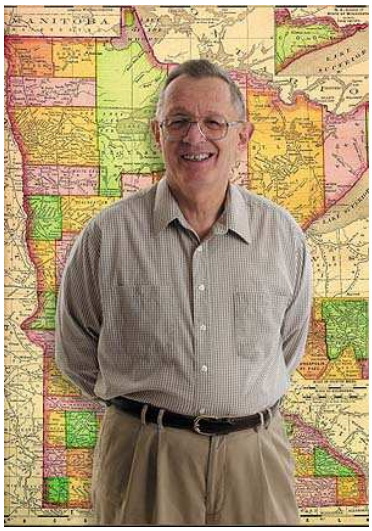
MARTIN GARDNER (1914–2010) wurde in Tulsa, Oklahoma geboren und studierte an der Universität Chicago, wo er 1936 mit einen Bachelor in Philosophie abschloß. Danach wurde er Reporter bei der Tulsa Tribune, war im Krieg bei der Navy und arbeitete anschließend als freier Schriftsteller. Von 1956 bis 1986 hatte er eine regelmäßige Kolumne über unterhaltsame Mathematik im *Scientific American.* Er schrieb über hundert Bücher zu Themen aus der Mathematik, Philosophie, Naturwissenschaften, Literatur, außerdem auch zwei Romane und mehrere Kurzgeschichten. Von 2004 bis zu seinem Tod lebte er wieder in Oklahoma.

wie auch programmiertechnische Optimierungen. Die University of Illinois at Urbana-Champaign, an der HAKEN und APPEL lehrten, kündigte dies unter anderem dadurch an, daß sie ihre Frankiermaschinen anstelle

einer Absenderangabe die Worte

FOUR COLORS
SUFFICE

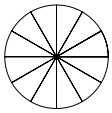
auf alle ausgehenden Briefe stempeln ließ.



KENNETH APPEL (1932–2013, Bild) war einer der ersten amerikanischen Mathematiker, die Computer in der mathematischen Forschung auch jenseits von klassischen Anwendungen wie der Numerik einsetzten. Er arbeitete u.a. über rekursive Funktionen, Kombinatorik und Gruppentheorie.
WOLFGANG HAKEN (1928–2022) wurde in Berlin geboren und studierte in Kiel, wo er einen Vortrag von HEESCH über den Vierfarbensatz hörte. 1965 emigrierte er nach USA, wo HEESCH mehrfach bei ihm zu Gast war. Seine sonstigen Arbeiten beschäftigen sich vor allem mit dreidimensionalen Mannigfaltigkeiten; dafür verlieh ihm die Universität Frankfurt 1993 die Ehrendoktorwürde.

§3: Was ist eine Karte?

Bevor wir uns daran machen können, irgendetwas über Kartenfärbungen zu beweisen, müssen wir zunächst die Begriffe klären. Wörter wie „Länder“ oder „Territorien“ und „benachbart“ scheinen zwar umgangssprachlich klar genug zu sein, aber wenn wir sie auf Grenzfälle anwenden, wird schnell klar, daß noch Klärungsbedarf besteht. Dabei wird sich herausstellen, daß einige anschaulich ziemlich klare Sachverhalte mathematisch erstaunlich komplex sein können.



Betrachten wir zunächst n Gebiete die in einem Kreis wie Kuchenstücke nebeneinander liegen. Alle diese Gebiete treffen sich im Mittelpunkt des Kreises; sollen wir sie deshalb als benachbart bezeichnen?

Wenn ja, muß jedes der n Gebiete seine eigene Farbe bekommen; dann kann also offensichtlich kein Vierfarbensatz gelten und, da wir n beliebig groß wählen können, reicht auch keine sonstige endliche Anzahl

von Farben für *jede* Karte. Wenn wir das Problem nicht von vornherein unlösbar machen wollen, dürfen wir also Länder, die sich nur in einem Punkt (oder in zwei oder allgemeiner in endlich vielen Punkten) berühren, nicht als benachbart betrachten: Zwei benachbarte Länder müssen mindestens eine gemeinsame Grenzlinie haben. Dann können wir im obigen Beispiel nicht benachbarte Sektoren gleich färben. Allgemein reichen in dieser Situation offenbar für gerade n zwei Farben, für ungerade $n > 1$ brauchen wir drei.

In der realen Welt gibt es Gebiete, die sich nur in einem Punkt berühren: Im Zentrum der Vereinigten Staaten von Amerika etwa sind die Grenzen zwischen zwei Bundesstaaten oft weitgehend durch Längen- und Breitengrade statt durch geographische Gegebenheiten definiert; in einem Punkt namens *Four Corners* kommen dabei (im Uhrzeigersinn) die Staaten Utah, Colorado, New Mexico und Arizona zusammen, wobei in unserem Sinne weder Utah und New Mexico noch Colorado und Arizona benachbart sind.

Eine wirkliche *Definition* von benachbart haben wir damit allerdings noch nicht: Dazu müssen wir erst festlegen, was eine „Grenzlinie“ sein soll. In gewisser Weise sind wir damit sogar vom Regen in die Traufe gekommen, denn wie der inzwischen verstorbene Karlsruher Mathematikprofessor HARRO HEUSER (1927–2011) in seinem Lehrbuch *Analysis 2* (§179) schreibt:

> Die im Alltag auftretenden Bögen sind so harmlos und übersichtlich gebaut, daß die mathematische Welt einen Schock erlitt, als ihr Peano im Jahre 1890 einen Bogen Γ im $\mathbb{R}^2$ vorführte, der ein ganzes Quadrat Q ausfüllte, für den also $\Gamma = Q$ war. Solche *flächenfüllenden Bögen* werden Peanobögen (oder Peanokurven) genannt. Die Anschauung gerät in die allergrößte Verlegenheit, wenn sie versucht, in den Bau derart pathologischer Bögen einzudringen. Die bloße Existenz der Peanobögen zeigt, daß der Begriff des Bogens, in dem ja nicht mehr als die Forderung der Stetigkeit steckt, viel zu allgemein ist, um in der Mathematik und ihren Anwendungen von Nutzen sein zu können. Es ist eine Tatsache von erheblicher Bedeutung, daß bereits Jordanbögen keine Peanopathologien mehr aufweisen. Immerhin können auch sie noch so unübersichtlich sein, daß die Anschauung sich weigert, ihrem Verlauf zu folgen. Einen hochexotischen Jordanbogen hat H. v. Koch im Jahre 1906 konstruiert, und eine kleine Modifikation dieser Konstruktion liefert eine Jordankurve von anschauungslähmender Ausgefallenheit. Ungebärdige Kurven dieser Art haben das ihrige dazu beigesteuert, das Vertrauen in die „anschauliche Evidenz“ zu untergraben und letztere als Beweisquelle völlig auszutrocknen.

Wir müssen also, wie so oft in der Mathematik, vorsichtig sein mit unseren Begriffen und auch mit der Art und Weise, wie wir mit ihnen umgehen. Der Einwand, daß wir es bei Karten mit „anständigen“ Kurven zu tun haben, löst unsere Probleme leider nicht: Erstens müßten wir dazu definieren, was eine „anständige“ Kurve sein soll, und zweitens verhalten sich zumindest die Ländergrenzen in der Natur eher wie die von HEUSER erwähnte KOCH-Kurve als wie die uns als typische Kurven bekannten Kreise, Parabeln oder Graphen differenzierbarer Funktionen.

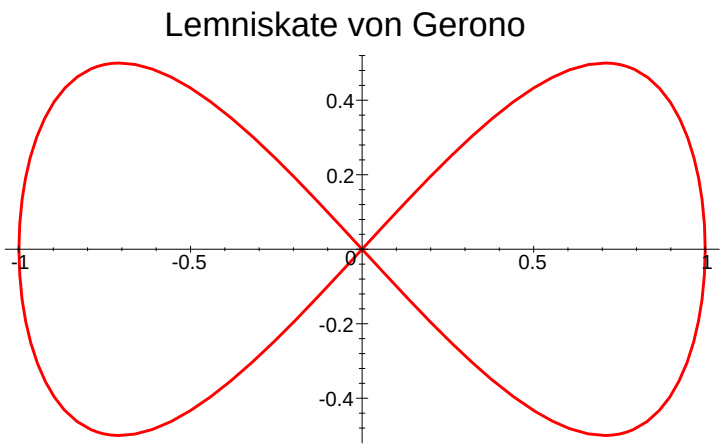
Letzteres exemplifizierte beispielsweise BENOIT B. MANDELBROT in seinem erstmalig 1977 erschienenen Buch *The fractal geometry of nature* mit der Frage: *How long is the coast of Britain?* Die Antworten, die in verschiedenen Nachschlagewerken zu finden sind, unterscheiden sich zum Teil beträchtlich, denn da die Küste in *jeder* Vergrößerung ungefähr dieselben Verästelungen hat, hängt das Ergebnis der Messung ganz wesentlich von der Länge des verwendeten Maßstabs ab. So ist auch nicht

verwunderlich, daß die gemeinsame Grenze zwischen Spanien und Portugal laut portugiesischer Enzyklopädie etwa 20% länger ist als laut spanischer: Die Maßstäbe waren jeweils verschieden, und, wie MANDELBROT schreibt, erscheint es durchaus verständlich, daß Portugal als kleineres Land mit einen feineren Maßstab mißt als Spanien.

Wir brauchen daher allgemeinere Kurvenbögen als die altbekannten Funktionsgraphen. Wir könnten beispielsweise die bei der Definition des Wegzusammenghangs benutzten Wege nehmen, jedoch sind diese nun wieder zu allgemein: Betrachten wir etwa die Abbildung

$$\varphi\colon \begin{cases} [-\pi, \pi] \to \mathbb{R}^2 \\ \qquad t \mapsto \big(\cos t, \cos t \cdot \sin t\big) \end{cases},$$

so erhalten wir die sogenannte Lemniskate von GERONO:



Diese Kurve schneidet sich selbst im Nullpunkt, und sie begrenzt zwei verschiedene Innengebiete. Ländergrenzen die so aussehen, wollen wir nicht zulassen; deshalb müssen wir die obige Definition noch etwas einschränken:

Definition: Ein Weg $\varphi\colon [0,\ 1] \to \mathbb{R}^2$ heißt JORDAN-Bogen, wenn er sich nicht selbst überschneidet, d.h. wenn $\varphi(s) = \varphi(t)$ nur gelten kann,

wenn $s = t$ oder (im Falle einer geschlossenen Kurve) $s, t \in \{0, 1\}$ ist. Eine JORDAN-Kurve ist eine geschlossene Kurve mit dieser Eigenschaft.

Die Lemniskate ist somit keine JORDAN-Kurve, denn hier ist sowohl $\varphi(-\frac{\pi}{2})$ als auch $\varphi(\frac{\pi}{2})$ gleich dem Nullpunkt. (Die Umparametrisierung auf das Intervall [0, 1] wird hoffentlich niemandem Schwierigkeiten bereiten.)

Bei einer JORDAN-Kurve sollte es nicht, wie bei der Lemniskate, möglich sein, daß das Innengebiet in zwei Teile zerfällt, und in der Tat besagt der JORDANsche Kurvensatz, daß eine solche Kurve die Ebene in genau zwei Gebiete zerlegt, das Äußere und das Innere. So klar dieser Satz anschaulich auch scheint, sein Beweis erfordert einen erstaunlich hohen Aufwand und wird daher meist, wenn überhaupt, nur in Spezialvorlesungen aus dem Bereich der geometrischen Topologie behandelt.

MARIE ENNEMOND CAMILLE JORDAN (1838–1922) war Sohn eines Ingenieurs. Er studierte zwar Mathematik an der École polytechnique in Paris, arbeite dann aber zunächst auch als Ingenieur. Daneben beschäftigte er sich aber weiterhin viel mit Mathematik und wurde 1876 Mathematikprofessor an der École polytechnique und von 1883–1885 auch am Collège de France. Seine mathematischen Arbeiten beschäftigen sich mit praktisch allen Teilgebieten der damaligen Mathematik, von nichtlinearen Gleichungen über die Analysis zur Geometrie und der damals *analysis situs* genannten Topologie.

Wir vereinbaren, daß zwei Gebiete benachbart sein sollen, wenn sie mindestens einen JORDAN-Bogen als gemeinsame Grenze haben. Aber was genau soll ein Gebiet sein?

Nicht jeder Staat hat ein zusammenhängendes Territorium: Beispielsweise ist Alaska vom Kerngebiet der Vereinigten Staaten aus auf dem Landweg nur über Kanada erreichbar; genauso ist das Gebiet um Kaliningrad, dem ehemaligen Königsberg, vom restlichen Rußland aus nur via Lettland oder Weißrußland und Litauen oder Polen erreichbar. Frankreich hat gar seine über die ganze Welt verstreuten *Départements d'Outremer,* und falls wir das Meer als eigenständiges Gebiet betrach-

ten, sorgen Inseln dafür, daß auch Deutschland, Italien und viele andere Staaten keine zusammenhängenden Landgebiete sind.

Falls wir hier verlangen wollten, daß zusammengehörende, aber räumlich auseinander liegende Teilgebiete jeweils dieselbe Farbe haben, kommen wir sicher nicht immer mit vier Farben aus: Es ist einfach, eine Geographie zu entwerfen, in der jeder Staat jeden anderen als Nachbar hat, indem man ihm einfach in jedem anderen Staat eine mit dem Kerngebiet unverbundene Enklave gibt.

Das Vierfarbenproblem kann daher höchstens dann lösbar sein, wenn wir alle Territorien als zusammenhängend annehmen. (HEAWOOD bewies auch Abschätzungen für die Anzahl von Farben, die man für Karten mit einer gewissen Anzahl unzusammenhängender Staaten braucht, aber die entstehenden Zahlen sind natürlich deutlich größer als vier.)

Für die exakte Definition einer Karte machen wir es wie die Kartographen, die im Gegensatz zu den Politikern von den Grenzen ausgehen und die Staaten als das betrachten, was zwischen den Grenzen liegt:

Wir *definieren* daher eine Karte als eine endliche Menge von Kurvenbögen, die sich höchstens in ihren Endpunkten schneiden dürfen.

Die Länder sind diejenigen Gebiete, die von den Kurvenbögen begrenzt werden. Um dies mathematisch exakt zu definieren, betrachten wir die Ebene *ohne* die Kurvenbögen. Anschaulich zerfällt diese Menge M in mehrere Gebiete, die Länder, wobei es zwischen je zwei Punkten, die zum selben Land gehören, einen Weg geben sollte, der die beiden verbindet, d.h. einen Kurvenbogen, der ganz im Land liegt und den einen der beiden Punkte als Anfang, den anderen als Ende hat.

Die können wir als Definition eines Landes nehmen: Wir sagen, zwei Punkte haben die gleiche Nationalität, wenn sie sich durch einen Kurvenbogen in M verbinden lassen, d.h. also durch einen Kurvenbogen, der keine Grenze überschneidet. Ein Land ist dann einfach die Menge von Punkten einer Nationalität.

(Um nur kurz anzudeuten, daß das nicht alles ganz so einfach ist, sei nur darauf hingewiesen, daß selbst die anschaulich klare Feststellung,

daß es auf einer solchen Karte nur endlich viele Länder gibt, nur mit ziemlichem Aufwand allgemein zu beweisen ist.)

§4: Graphen

Da sich unsere anschauliche Vorstellung von Karten, wie wir gerade gesehen haben, nur schwer in mathematischen Formalismus übersetzen läßt, ist es zur Lösung des Vierfarbenproblems geschickter, wenn wir das Problem in ein zwar etwas weniger anschauliches, dafür aber formal besser zugängliches Problem übersetzen. Auch diese Übersetzung ist anschaulich ziemlich klar und für konkrete Karten einfach durchzuführen:

Zunächst wählen wir in jedem der auf der Karte dargestellten Territorien *irgendeinen* Punkt (z.B. eine zentral gelegene Stadt oder die Hauptstadt oder den Schwerpunkt) und verbinden diesen Punkt mittels je eines Kurvenbogens mit den entsprechenden Punkten der Nachbarterritorien. Die Kurvenbögen wählen wir so, daß sie sich gegenseitig nicht schneiden – außer natürlich in den zu verbindenden Punkten, wo im allgemeinen mehrere Bögen zusammenlaufen. Anschaulich ist ziemlich klar, daß dies möglich ist; ein *Beweis,* daß dies *immer* funktioniert, erfordert leider wie der des JORDANschen Kurvensatzes einen beträchtlichen Aufwand.

Das entstehende Gebilde aus Punkten und Kurvenbögen bezeichnet man in der Mathematik als einen *Graphen:* Formal ist er gegeben durch eine Menge von Ecken, den ausgewählten Punkten, und einer Menge von Kanten, den Kurvenbögen, die gewisse Ecken miteinander verbinden.

Jede Kartenfärbung führt sofort zu einer Färbung des Graphen: Jede Ecke bekommt die Farbe des zugehörigen Gebiets. Die Kanten bekommen keine Farbe. Offensichtlich haben auf der Karte genau dann keine zwei benachbarten Gebiete dieselbe Farbe, wenn keine zwei durch eine Kante verbundenen Ecken des Graphen dieselbe Farbe haben.

Haben wir umgekehrt jeder Ecke des Graphen eine Farbe gegeben derart, daß keine zwei durch eine Kante verbundenen Ecken dieselbe Farbe haben, so führt dies sofort zu einer Kartenfärbung, in der keine zwei benachbarten Gebiete dieselbe Farbe haben.

Das Kartenfärbungsproblem wird damit äquivalent zum folgenden graphentheoretischen Problem: Man färbe die Ecken des Graphen so,

daß keine zwei durch eine Kante verbundenen Ecken dieselbe Farbe haben. In dieser Form wurde das Problem seit KEMPEs Zeiten stets bearbeitet, und in dieser Form wurde es auch gelöst.

§ 5: Sechs Farben reichen

Die Graphen, die wir im vorigen Paragraphen definiert haben, liegen natürlich in der Ebene; wir haben daher nicht nur Ecken und Kanten, sondern auch Flächen, die durch die Kanten berandet werden.

Solche Graphen erhält man auch, wenn man von zusammenhängenden Polyedern „ohne Löcher“ ausgeht, diese ins Innere einer Kugel setzt und die Kanten auf die Kugeloberfläche projeziert. Das ist dann zwar ein Graph auf der Kugeloberfläche, aber durch stereographische Projektion von einem Punkt, der auf keiner Kante liegt, erhalten wir daraus problemlos einen Graphen in der Ebene.

Für solche Graphen gilt der EULERsche Polyedersatz:

$$\#\,\text{Ecken} - \#\,\text{Kanten} + \#\,\text{Flächen} = 2\,.$$

(Zu den Flächen zählt dabei auch eine unendliche, nämlich die Außenfläche jenseits aller Kanten.)

Der *Beweis* erfolgt durch vollständige Induktion nach der Anzahl der Kanten: Im Falle einer einzigen Kanten haben wir auch nur eine Fläche, aber zwei Ecken; der Satz ist also richtig.

Hat der Graph mehr als eine Kante, so entfernen wir eine von diesen, achten aber darauf, daß der Kantenzug als ganzer auch nach dem Entfernen der Kante noch zusammenhängend bleibt.

Falls die entfernte Kante zwei Flächen voneinander trennte, fallen diese anschließend zusammen, d.h. außer der Anzahl der Kanten wird auch die Anzahl der Flächen um eins vermindert. Ecken verschwinden keine, da die verschwundene Kante bei beiden Gebieten Teil der Grenze war, also Teil eines geschlossenen Kantenzug. Somit bleibt die obige Summe in der Tat unverändert.

Falls die Kante keine Flächen voneinander trennte, war sie nicht Teil eines geschlossenen Kantenzugs, also verschwindet keine Fläche. Allerdings muß die Kante dann eine Ecke haben, von der keine weitere

Kante ausgeht, und diese Ecke verschwindet zusammen mit der Kante. Die andere bleibt, da wir stets von zusammenhängenden Kantenzügen ausgehen. Also haben wir eine Ecke und eine Kante weniger und wieder dieselbe alternierende Summe.

Diese alternierende Summe ist nach Induktionsannahme in beiden Fällen gleich zwei, womit der EULERsche Polyedersatz bewiesen wäre.

Diesen Satz wollen wir nun anwenden um uns Informationen über den zu färbenden Graphen zu verschaffen.

Die Anzahl e der Ecken ist hier natürlich die Anzahl der einzufärbenden Territorien, und falls das i-te dieser Territorien k_i Nachbarn hat, gibt es

$$k = \frac{1}{2} \sum_{i=1}^{e} k_i$$

Kanten. Über die Anzahl f der Flächen können wir allerdings zumindest auf Anhieb nichts sagen.

Um trotzdem den EULERschen Polyedersatz anwenden zu können, greifen wir zu einem Trick, der in der Mathematik paradoxerweise erstaunlich oft zum Erfolg führt: Wir machen unser Problem schwerer.

Hier geschieht dies dadurch, daß wir zusätzliche Kanten in den Graph einfügen, mit anderen Worten, wir geben einigen Territorien Nachbarn, die sie in Wirklichkeit gar nicht haben. Die vorhandenen Kanten beranden Polygone, über deren Eckenzahl wir nichts wissen. Durch Einzeichnen von Diagonalen können wir diese so unterteilen, daß wir bei gleichbleibender Gesamtzahl der Ecken nur noch Dreiecke haben. Dann hat jede Fläche genau drei Kanten, und umgekehrt gehört weiterhin jede Kante zu genau zwei Flächen. Daher ist $3f = 2k$ oder $f = \frac{2}{3}k$, und damit können wir f aus dem EULERschen Polyedersatz eliminieren:

$$e - k + f = e - \frac{k}{3} = 2 \quad \text{oder} \quad k = 3e - 6\,.$$

Im ursprünglichen Graph, bevor wir die zusätzlichen Kanten eingefügt haben, gab es eventuell weniger Kanten; für diesen gilt also nur

$$k \leq 3e - 6\,.$$

Sei nun n die Minimalzahl von Kanten, die von einer Ecke ausgehen. Da jede Kante genau zwei Ecken miteinander verbindet, ist dann

$$k \geq \frac{n}{2} e \,,$$

also ist

$$\frac{n}{2} e \leq k \leq 3e - 6 < 3e \quad \text{oder} \quad \frac{n}{2} < 3 \,.$$

Somit ist n höchstens gleich fünf.

Damit wissen wir, daß es in jedem ebenen Graphen mindestens eine Ecke gibt, von der höchstens fünf Kanten ausgehen, und das reicht, um per Induktion zu beweisen, daß sechs Farben genügen:

Satz: Jede Karte kann mit sechs Farben so eingefärbt werden, daß keine zwei benachbarten Gebiete die gleiche Farbe haben.

Zum *Beweis* können wir natürlich das äquivalente Problem für Graphen betrachten; wir lösen es durch Induktion nach der Anzahl e der Ecken.

Falls es höchstens sechs Ecken gibt, können wir jeder davon ihre eigene Farbe geben und haben das Problem gelöst.

Andernfalls nehmen wir an, wir hätten das Problem bereits gelöst für Graphen mit $e-1$ Ecken und betrachten nun einen Graphen mit e Ecken.

Wie wir oben gesehen haben, gibt es unter diesen Ecken mindestens eine, von der höchstens fünf Kanten ausgehen. Wenn wir diese Ecke und ihre Kanten aus dem Graph entfernen, bleibt ein Teilgraph mit nur $e - 1$ Ecken übrig, den wir nach Induktionsannahme mit sechs Farben so färben können, daß keine zwei Endpunkte einer Kante dieselbe Farbe haben.

Nachdem wir dies getan haben, setzen wir die herausgenommene Ecke samt ihrer Kanten wieder ein. Da höchstens fünf Kanten von ihr ausgehen, können die anderen Endpunkte dieser Kanten höchstens fünf verschiedene Farben haben, wir haben also noch mindestens eine Farbe frei, um die Ecke so zu färben, daß sie eine andere Farbe hat als jeder ihrer Nachbarn.

■

Bemerkung: Vor dem Hintergrund der Tatsache, daß tatsächlich sogar vier Farben ausreichen, erscheint dieser Satz sehr schwach. Die Beweismethodik reicht aber aus, um im Falle von Karten auf einer Fläche vom Geschlecht $g \geq 1$ die optimale Schranke zu finden: Wir können grundsätzlich genauso vorgehen, allerdings gilt der EULERsche Polyedersatz in der Form, die wir kennen, natürlich nur für Graphen auf der Kugel *bzw.* in der Ebene. Im Kapitel über die Homologie simplizialer Komplexe werden wird sehen, daß der EULERsche Polyedersatz auch für Karten auf allgemeineren Flächen gilt, allerdings muß man dann die rechte Seite ersetzen durch eine nur von der Topologie der Fläche abhängige andere Zahl. Genauer kann man zeigen, daß es zu jeder Fläche ohne Rand im $\mathbb{R}^3$ eine Zahl $g \in \mathbb{N}_0$ gibt, das sogenannte *Geschlecht* der Fläche, so daß für alle Graphen auf der Fläche gilt:

$$\#\,\text{Ecken} - \#\,\text{Kanten} + \#\,\text{Flächen} = 2 - 2g\,.$$

Damit bekommen wir, wenn wir wie oben vorgehen, die Abschätzung

$$k \leq 3e + 6g - 6$$

für die Anzahl k der Kanten und die Anzahl e der Ecken. Bezeichnen wir wieder mit n die kleinste Anzahl von Kanten, die von einer Ecke des Graphen ausgeht, so ist also hier

$$\frac{ne}{2} \leq 3e + 6g - 6 \quad \text{oder} \quad n \leq 6 + \frac{12g - 12}{e}\,.$$

Oben, im Falle $g = 0$, war der ganz rechts stehende Summand negativ; wir haben ihn einfach weggelassen und dabei aus dem Zeichen $\leq$ ein striktes Kleinerzeichen gemacht.

Für $g \geq 1$ ist $12g - 12 \geq 0$, hier können wir also nicht so vorgehen. Dafür wird jetzt eine andere Art der Abschätzung möglich: Da n die kleinste Anzahl von Kanten ist, die von einer Ecke ausgeht, gibt es mindestens eine Ecke, von der n Kanten ausgehen. Diese Kanten haben zusammen $n + 1$ Endpunkte, also ist $e \geq n + 1$ und $\frac{1}{e} \leq \frac{1}{n+1}$. Wenn der Zähler $12g - 12 \geq 0$ ist **und nur dann**, können wir also in obiger Formelzeile weiter abschätzen und erhalten

$$n \leq 6 + \frac{12g - 12}{e} \leq 6 + \frac{12g - 12}{n + 1}\,.$$

Multiplikation mit dem Nenner $n+1$ macht daraus die Ungleichung

$$n(n+1) \leq 6(n+1) + 12g - 12 \quad \text{oder} \quad n^2 - 5n - 12g + 6 \leq 0\,.$$

Mittels quadratischer Ergänzung können wir dies umschreiben zu

$$\left(n - \frac{5}{2}\right)^2 - \frac{25}{4} - 12g + 6 = \left(n - \frac{5}{2}\right)^2 - \frac{1}{4} - 12g \leq 0\,,$$

d.h.

$$\left(n - \frac{5}{2}\right)^2 \leq \frac{1}{4} + 12g = \frac{1+48g}{4}\,.$$

Ziehen wir auf beiden Seiten die positive Quadratwurzel, so erhalten wir für $n > 2$ die Ungleichung

$$n - \frac{5}{2} \leq \frac{\sqrt{1+48g}}{2} \quad \text{oder} \quad n \leq \frac{5+\sqrt{1+48g}}{2}\,.$$

Da n eine natürliche Zahl ist, können wir dies verschärfen zu

$$n \leq \left[\frac{5+\sqrt{1+48g}}{2}\right]\,.$$

Der Ausdruck rechts ist für alle $g \in \mathbb{N}$ größer oder gleich sechs, insbesondere also größer als zwei. Daher gibt es in jedem Graphen auf einer Fläche vom Geschlecht $g \geq 1$ mindestens eine Ecke mit höchstens dieser Anzahl von Nachbarn. Wie oben läßt sich daraus schnell folgern, daß eine um eins höhere Anzahl von Farben stets ausreicht, um eine Karte so einzufärben, daß keine zwei benachbarten Länder die gleiche Farbe haben, d.h. auf einer Fläche vom Geschlecht $g \geq 1$ genügen

$$1 + \left[\frac{5+\sqrt{1+48g}}{2}\right] = \left[\frac{7+\sqrt{1+48g}}{2}\right]$$

Farben.

Diese Ungleichung bewies HEAWOOD bereits 1890; seit 1968 weiß man, daß sie die bestmögliche ist, d.h. es gibt Karten, für die man wirklich so viele Farben braucht. Interessanterweise war dies hier bei weitem der schwierigste Teil, wohingegen es im klassischen Fall, für $g = 0$, trivial ist, daß man *mindestens* vier Farben braucht: Hier bestand die große Schwierigkeiten darin, zu zeigen, daß man wirklich mit vier Farben *auskommt.*

§6: Fünf Farben reichen

Wir wollen uns zunächst überlegen, warum *fünf* Farben ausreichen:

Satz: Jede Karte kann mit fünf Farben so eingefärbt werden, daß keine zwei benachbarten Gebiete die gleiche Farbe haben.

Zum *Beweis* betrachten wie wieder das äquivalente Problem für Graphen und lösen es durch Induktion nach der Anzahl e der Ecken.

Falls es höchstens fünf Ecken gibt, können wir jeder davon ihre eigene Farbe geben und haben das Problem gelöst.

Andernfalls nehmen wir an, wir hätten das Problem bereits gelöst für Graphen mit $e-1$ Ecken und betrachten nun einen Graphen mit e Ecken.

Wie wir wissen, gibt es unter diesen Ecken mindestens eine, von der höchstens fünf Kanten ausgehen. Wenn wir diese Ecke E und ihre Kanten aus dem Graph entfernen, bleibt ein Teilgraph mit nur $e-1$ Ecken übrig, den wir nach Induktionsannahme mit fünf Farben so färben können, daß keine zwei Endpunkte einer Kante dieselbe Farbe haben.

Nachdem wir dies getan haben, setzen wir die herausgenommene Ecke samt ihrer Kanten wieder ein. Falls ihre Nachbarpunkte höchstens vier verschiedene Farben haben, haben wir noch mindestens eine Farbe übrig für E und das Problem ist gelöst.

Bleibt der Fall, daß es genau fünf Nachbarecken gibt, von denen jede eine andere Farbe hat. In diesem Fall müssen wir im restlichen Graphen Farben ändern; die Methode dazu sind die sogenannten KEMPE-Ketten, mit denen KEMPE 1880 glaubte den Vierfarbensatz bewiesen zu haben.

Seien E_1, E_2, E_3, E_4 und E_5 die fünf Nachbarecken; da es auf die Namen der Farben nicht ankommt, können wir ohne Beschränkung der Allgemeinheit davon ausgehen, daß ihre Farben, in dieser Reihenfolge, rot, gelb, grün, blau und schwarz seien.

Betrachten wir nun die Menge M aller Ecken, die von E_1 aus durch einen Kantenzug erreichbar sind, in dem alle Ecken rot oder grün gefärbt sind. Diese Menge kann, muß aber nicht, die grüne Nachbarecke E_3 von E enthalten.

Falls E_3 nicht in M liegt, vertauschen wir für die Ecken aus M die Farben, d.h., die Ecken, die bislang rot waren, werden grün, und die grünen werden rot. Auch mit dieser Färbung hat keine Ecke eine Nachbarecke gleicher Farbe, denn für $P \in M$ ist auch jede rote oder grüne Nachbarecke von P in M, und da *jede* Ecke in M ihre Farbe wechselt, kann es dort keine benachbarte Ecken gleicher Farbe geben. Mit den Nachbarecken von P, die *nicht* in M liegen, gibt es erst recht keine Probleme, denn diese sind gelb, blau oder schwarz, während P rot oder grün ist.

Durch das Umfärben wechselte insbesondere E_1 seine Farbe von rot zu grün, während E_3, da es nicht in M liegt, grün blieb. Somit hat E nun zwei grüne, dafür aber keinen roten Nachbarn mehr. Wir können E also rot färben und sind fertig.

Falls E_3 in M liegt, können wir nicht so vorgehen: Zwar können wir auch dann die Farben aller Ecken aus M ändern, aber das hat dann nur zur Folge, daß E_1 grün und E_3 rot wird, es gibt also weiterhin fünf verschiedenfarbige Nachbarn.

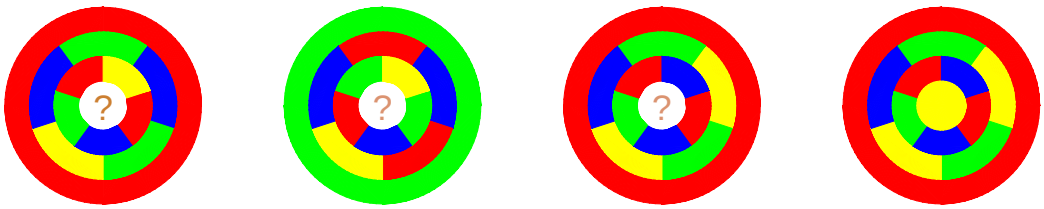
In diesem Fall betrachten wir die Menge N aller Ecken, die von E_2 aus durch einen gelb-blauen Kantenzug erreichbar sind. Wegen $E_3 \in M$ gibt es einen rot-grünen Kantenzug, der E_1 mit E_3 verbindet; zusammen mit dem Kantenzug von E_3 über E_2 nach E_1 bildet dieser eine geschlossene Kurve.

Innerhalb dieser geschlossenen Kurve muß N liegen, denn zwei Kantenzüge eines Graphen können sich höchstens ein einer Ecke schneiden, und keine Ecke kann gleichzeitig rot oder grün und blau oder gelb sein. Damit folgt insbesondere, daß E_4 *nicht* in N liegen kann, denn E_4 liegt nicht innerhalb dieser geschlossenen Kurve.

Wenn wir also in N alle bislang blauen Ecken gelb färben und alle bislang gelben blau, wird E_2 blau und E_4 bleibt blau; wir können daher E gelb färben und haben das Problem auch in diesem letzten Fall gelöst. ■

KEMPE hatte durch ein etwas komplizierteres Argument mit seinen Ketten versucht, auch den Vierfarbensatz zu beweisen; leider konnte aber

HEAWOOD durch ein explizites Gegenbeispiel zeigen, daß dieses Argument nicht immer richtig ist. Trotzdem sind KEMPE-Ketten auch für die Vierfärbung von Karten oft nützlich. Als ein Beispiel können wir auf dem die Reihe der unten abgedruckten Karten betrachten. Links haben wir eine Karte, bei der wir nur noch das Zentrum färben müssen. Da es Nachbarn mit vier verschiedenen Farben hat, ist das ohne Umfärben eines Teils der äußeren Gebiete nicht möglich.



Die roten und grünen Gebiete der linken Karte bilden eine einzige KEMPE-Kette; jedes Vertauschen von Rot und Grün in nur einem Gebiet führt also dazu, daß überall Rot und Grün vertauscht werden, und das führt zur zweiten Karte der Reihe, mit der wir genauso wenig anfangen können wie mit der ersten.

Betrachten wir allerdings die gelben und blauen Gebiete, so haben wir zwei disjunkte KEMPE-Ketten, von denen wir auf der einen die Farben vertauschen können. Falls wir etwa die nehmen, die nur aus den beiden Gebieten rechts des Zentrums besteht, kommen wir durch Umfärben zur dritten Karte der Reihe, in der das Zentrum nur noch rote, grüne und blaue Nachbarn hat. Somit können wir es gelb färben und erhalten rechts in der Reihe die gesuchte Vierfärbung der Karte.

§7: Wie wurde der Vierfarbensatz bewiesen?

Allgemein führt diese Strategie, wie gesagt, nicht zum Erfolg; für den Beweis des Vierfarbensatzes war also eine neue Idee notwendig.

Die Idee, die schließlich zu einem Beweis führte, bestand darin, zwar weiterhin rekursiv zu arbeiten, im Induktionsschritt aber nicht nur einen Punkt einschließlich seiner Verbindungskanten herauszunehmen, sondern eine größere Teilkonfiguration. Man arbeitet dabei stets mit dem

Graphen, in den so lange Kanten eingefügt wurden, bis alle Flächen von nur drei Kanten begrenzt sind.

Um die Beweislogik etwas übersichtlicher zu machen, formulieren wir die Beweisstrategie etwas um:

Angenommen, es gäbe eine Karte, die nicht mit vier Farben so eingefärbt werden kann, daß benachbarte Gebiete verschiedene Farben haben. Dann gibt es, wie wir oben gesehen haben, auch entsprechende Karten, bei denen in den daraus konstruierten ebenen Graphen alle Flächen Dreiecke sind, und natürlich gibt es unter diesen Karten auch minimale, das heißt solche mit der geringstmöglichen Anzahl von Gebieten. Wir betrachten im folgenden ein solches minimales Gegenbeispiel.

Der Beweis des Vierfarbensatzes ruht auf zwei großen Säulen:

1. Reduzibilität: Finde eine große Anzahl von Konfigurationen mit der Eigenschaft, daß kein minimales Gegenbeispiel eine solche Konfiguration enthalten kann. Solche Konfigurationen heißen *reduzibel;* Unterbegriffe wie D-Reduzibilität und C-Reduzibilität beziehen sich auf Beweisstrategien, mit denen man gewisse Klassen von Konfigurationen als reduzibel erkennen kann.

Besonders wichtig für den endgültigen Beweis war HEESCHs D-Reduzibilität, die ein Computer durch stures Nachrechnen erkennen kann.

Beim Beweis des Sechsfarbensatzes wären solche Konfigurationen etwa die Punkte mit höchstens fünf Nachbarn. Wenn wir einen solchen Punkt aus dem Graphen herausnehmen, können wir den Rest färben wie wir wollen, es bleibt immer mindestens eine zulässige Farbe für den herausgenommenen Punkt übrig.

Ähnlich ist es bei D-reduziblen Konfigurationen im Beweis des Vierfarbensatzes: Sie haben die Eigenschaft, daß man den Restgraphen, der nach ihrer Herausnahme entsteht, (zulässig) färben kann, wie man will; die Färbung läßt sich auf jeden Fall zu einer Färbung des gesamten Graphen fortsetzen.

Der Beweis, daß eine vorgegebene Konfiguration D-reduzibel ist, erfolgt durch stures Nachrechnen und ist daher sehr gut an Computer delegierbar. Auch dort gibt es allerdings Probleme:

Falls die betrachtete Konfiguration im Graphen n Nachbarpunkte hat, muß man alle Möglichkeiten für deren Färbung einzeln betrachten. Falls die Nachbarpunkte einen Ring aus n Punkten bilden, gibt es dafür $28 \cdot 3^{n-3}$ Möglichkeiten. Diese sind zwar nicht alle wesentlich verschieden, denn Konfigurationen, die aus bereits behandelten nur durch Permutation der Farben entstehen, müssen nicht noch einmal gesondert untersucht werden. Dies reduziert die Arbeit aber lediglich um knapp einen Faktor 24 und ändert nichts daran, daß ein Randpunkt mehr zu einer Verdreifachung der Möglichkeiten führt. Damit ist klar, daß Konfigurationen mit zu vielen Randpunkten vermieden werden sollten: Zu der Zeit, als HEESCH mit der Untersuchung D-reduzibler Konfigurationen begann und sein Schüler DÜRRE Konfigurationen per Computer auf D-Reduzibilität testete, brauchte der Computer beispielsweise für eine einzige (ziemlich schwierige) Konfiguration mit 14 Randpunkten 26 Stunden Rechenzeit. Da HEESCH ursprünglich schätzte, daß zum Beweis des Vierfarbensatzes etwa 10 000 Konfigurationen untersucht werden müßten, war klar, daß auch ein Beweis per Computer alles andere als einfach sein würde.

2. Unvermeidbarkeit: Hier geht es darum, eine Liste von Konfigurationen zu finden derart, daß jedes minimale Gegenbeispiel zum Vierfarbensatz mindestens eine dieser Konfigurationen als Teilgraph enthalten muß und zu zeigen, daß jede dieser Konfigurationen reduzibel ist. Falls dies gelingt, ist der Vierfarbensatz bewiesen:

Angenommen, er wäre falsch. Dann gäbe es ein minimales Gegenbeispiel im obigen Sinne. Diese müßte eine unvermeidbare Konfiguration enthalten, könnte aber keine reduzible Konfiguration enthalten. Falls jede unvermeidbare Konfiguration reduzibel ist, ist das unmöglich; es gibt daher kein minimales Gegenbeispiel und somit überhaupt keines.

Beim Beweis des Fünf- und des Sechsfarbensatzes lieferte uns der EULERsche Polyedersatz die unvermeidbaren Konfigurationen: Es waren einfach Punkte mit höchstens fünf Nachbarn. Im Falle des Sechsfarbensatzes haben diese Punkte trivialerweise eine analoge Eigenschaft zur D-Reduzibilität: Egal, wie die Nachbarpunkte eingefärbt sind, bleibt immer mindestens eine Farbe übrig für den Punkt selbst. Beim Fünffarbensatz gilt dies für einen Punkt mit fünf Nachbarn nicht mehr; hier

muß zusätzlich via KEMPE-Ketten ein Teil des Restgraphen umgefärbt werden. Dies ist ein Spezialfall der C-Reduzibilität, auf die hier jedoch nicht genauer eingegangen werden soll.

Die Suche nach unvermeidbaren Konfigurationen geht auch bei HEESCHs Strategie aus vom EULERschen Polyedersatz: Jede Ecke des Graphen erhält eine „Ladung“, die anfänglich gleich sechs minus der Anzahl der Nachbarecken ist. Wie die Rechnung, die zum Beweis des Sechsfarbensatzes führte, zeigt, ist dann die Summe aller Ladungen gleich zwölf, also insbesondere positiv.

Für Fünf- und Sechsfarbensatz reicht es, als unvermeidbare Konfigurationen die Punkte mit positiver Ladung zu nehmen; für den Vierfarbensatz muß die Ladung zunächst weiter auf dem Graphen verteilt werden ohne an der Summe etwas zu ändern. Dazu dienen sogenannte Entladungsprozeduren mit Regeln der Art:

Falls eine Ecke E mit fünf Nachbarn unter diesen Nachbarn welche mit mindestens sieben Nachbarn hat, erhält jeder dieser Nachbarn von E die Ladung $\frac{1}{2}$.

Danach ist beispielsweise eine Ecke mit fünf Nachbarn nicht mehr positiv geladen, wenn zwei oder mehr ihrer Nachbarn ihrerseits mindestens sieben Nachbarn haben; dagegen werden Ecken mit $k \geq 7$ Nachbarn positiv geladen, falls unter diesen Nachbarn mehr als $2k - 12$ ihrerseits nur fünf Nachbarn haben. (Da die Anzahl dieser Nachbarn natürlich nicht größer als k sein kann, ist dies nur für $7 \leq k \leq 11$ möglich.)

Hat man durch eine Entladungsprozedur die Ladungen neu verteilt, ist wegen der unverändert gebliebenen Ladungssumme klar, daß es stets Ecken mit positiver Ladung geben muß. Der nächste Schritt besteht nun darin, eine Liste von Konfigurationen aufzustellen derart, daß jeder Graph, der eine der positiv geladenen Ecken enthält, auch eine Konfiguration aus der Liste enthalten muß. Man konstruiert sich die Konfigurationen also um die positiv geladenen Ecken herum. Das Ergebnis ist eine Liste unvermeidbarer Konfigurationen, die natürlich stark von der gewählten Entladungsprozedur abhängt.

Das große Problem und die Kunst besteht nun darin, die Entladungsprozedur so zu wählen, daß die Liste unvermeidbarer Konfigurationen

nur reduzible Konfigurationen enthält. Dies haben APPEL und HAKEN schließlich geschafft mit einer Liste von 1818 unvermeidbaren Konfigurationen (von der sich später herausstellte, daß bereits eine Teilliste von nur 1476 Konfigurationen unvermeidbar ist.) Diese Liste wurde trotz vieler Computerexperimente per Hand aufgestellt, und auch ihre Unvermeidbarkeit wurde ohne Computerhilfe bewiesen. Die Reduzibilität der Listenelemente allerdings konnte nur per Computer nachgewiesen werden. Dessen Programmierung war im wesentlichen das Werk des damaligen Studenten JOHN KOCH (*1948), der damit 1976 promovierte und sich dazu eine ganzen Reihe von Optimierungen der Beweisstrategie einfallen lassen mußte.

§8: Ist der Vierfarbensatz damit bewiesen?

Nachdem im vorigen Abschnitt der Beweis von APPEL und HAKEN skizziert wurde, mag diese Frage seltsam erscheinen; angesichts der Tatsache, daß hier erstmalig ein wesentlicher Teil des Arguments nur auf Computerrechnungen beruht, sollten wir uns aber doch fragen, was wir uns unter einem mathematischen Beweis vorstellen wollen.

Das Konzept eines mathematischen Beweises kam aus der griechischen Geometrie. Diese ging bekanntlich aus von als unmittelbar einsichtig angesehenen Axiomen und Postulaten und leitete daraus ihre Sätze ab. Mit dem Aufkommen einer leistungsfähigen formalen Logik gegen Ende des 19. Jahrhunderts konnte man diesen Begriff der Ableitung präzisieren und es stellte sich heraus, daß tatsächlich außer den EUKLIDischen Axiomen noch weitere notwendig sind (DAVID HILBERT stellte 1899 ein entsprechendes System auf), aber für die Herleitung und das Verständnis mathematischer Sätze hatte dies keine allzu große Bedeutung: Das Wissen der Mathematik vermehrt sich kumulativ, so daß Mathematiker meist eher auf den Resultaten anderer Mathematiker aufbauen als daß sie direkt mit den Axiomen einer Theorie zu tun haben. Zwar sind sie davon überzeugt, daß sich ihre Sätze letztlich in kleinen Schritten aus den Axiomen ableiten lassen, aber sie schreiben ihre Resultate aus gutem Grund nie so auf:

Mathematik ist zwar eine formale Wissenschaft in dem Sinne, daß sie formale Methoden benutzt; jedoch sind diese Methoden nur ein Hilfs-

mittel, um interessante Probleme zu lösen. Es ist sehr einfach, aus einem Axiomensystem durch Anwendung von Ableitungsregeln neue korrekte Formeln abzuleiten, aber selbst wenn diese wider Erwarten interessant sein sollten, ist eine solche Herleitung nicht das, was sich ein Mathematiker unter einem Beweis vorstellt: Er möchte über das rein Formale hinaus einen *Grund* sehen, warum ein Satz richtig ist. Ein guter Beweis sollte daher so strukturiert sein, daß die dahinter stehenden Ideen klar erkennbar werden.

Natürlich muß auch klar werden, wie sich diese Ideen in einen formalen Beweis übersetzen lassen, aber da eine solche Übersetzung zur mathematischen Routine gehört, wird dies, je nach intendiertem Leserkreis, nur mehr oder weniger kurz angedeutet. Von Studienanfängern zu Beginn des ersten Semesters erwartet man noch, daß sie zumindest einige einfache Formeln sehr detailliert beweisen, danach gehört es zum Prozeß des mathematischen Erwachsenwerdens, daß mit zunehmender Erfahrung immer komplexere Schritte „klar“ sein sollten und weggelassen werden können. Wer mathematische Arbeiten in anerkannten Fachzeitschriften publizieren oder auch nur lesen möchte, muß diesen Prozeß beendet haben: Solche Arbeiten sind fast immer sehr knapp geschrieben. Das liegt nicht nur daran, daß die Herausgeber mit dem Platz geizen müssen: In den meisten Teilgebieten der Mathematik ist es schon seit langem schlichtweg unmöglich, einen interessanten Satz zu beweisen, ohne ganz wesentlich auf der Arbeit anderer Mathematiker aufzubauen. Schon SIR ISSAC NEWTON (1642 - 1727) schrieb am 5. Februar 1675 etwas boshaft an den (ziemlich kleinwüchsigen) ROBERT HOOKE (1635-1702), der sich beklagte, daß NEWTON seine Ideen über das Licht in den *Principia Mathematica* benutzt habe:

If I have seen further, it is by standing on the shoulders of giants.

Daß dies bei weitem nicht nur in der Mathematik und Physik so ist, sondern allgemein bei jedem Streben nach Erkenntnis, ließ schon vor knapp zwei Tausend Jahren MARCUS ANNAEUS LUCANUS (39-65) in seinem Epos *Pharsalia* über den römischen Bürgerkrieg den DIDACUS STELLA sagen: *Pigmaei gigantum humeris impositi plusquam ipsi gigantes vident.* (Pygmäen, die auf den Schultern von Riesen stehen, sehen mehr als die Riesen selbst.)

Wie ordnet sich nun der Beweis von APPEL und HAKEN in dieses Selbstverständnis der Mathematik ein?

Auch der schärfste Kritiker von Computerbeweisen kann nicht leugnen, daß dieser Beweis sehr wesentliche mathematische Ideen enthält, nicht nur von APPEL und HAKEN selbst, sondern auch von vielen ihrer Vorgänger wie KEMPE, HEAWOOD, BIRKHOFF, HEESCH und vielen anderen. Der Beweis wäre nie zustande gekommen, wenn sich die Autoren darauf beschränkt hätten, einfach einen Computer drauf los rechnen zu lassen.

Trotzdem unterscheidet sich der Beweis in beiden seiner Säulen vom Idealbild eines mathematischen Beweises:

Am deutlichsten wird dies bei den Reduzibilitätsbeweisen per Computer: Normalerweise ist es die Aufgabe eines mathematischen Beweises, auf dem entsprechenden Gebiet hinreichend kompetente weitere Mathematiker von der Richtigkeit des bewiesenen Satzes zu überzeugen: Ein entsprechender Leser sollte, nachdem er den Beweis verstanden hat, idealerweise in der Lage sein, an jeder beliebigen Stelle des Beweises Details einzufügen, um den Beweis oder Teile davon (beispielsweise für die Ausbildung von Mathematikstudenten oder einen Übersichtsartikel für Nichtspezialisten) auch Zuhörern oder Lesern eines niedrigeren Kompetenzniveaus klar zu machen. Dies ist bei einem Computerbeweis kaum möglich: Schon für eine einzige Konfiguration mit 14 Nachbarpunkten kann kein menschlicher Leser mehr überprüfen, daß sich wirklich alle Färbungen des äußeren Rings ins Innere fortsetzen lassen.

Man könnte nun einwenden, daß es genüge, wenn ein hinreichend kompetenter Leser wenigstens das Programm verstehen könne, das die Überprüfung vornimmt, aber leider ist selbst das im Allgemeinen (und insbesondere speziell bei diesem Beweis) zuviel verlangt: Aus Effizienzgründen mußte KOCH seine Programme in Assembler schreiben, und selbst wenn es heute noch Mathematiker gibt, die die Assemblersprache der IBM 360 und 370 Serie beherrschen, sind bis zum letzten durchoptimierte Assemblerprogramme so schwer zu lesen, daß auch sie kaum eine Chance haben.

In der Tat versuchte 1996 eine Gruppe von Mathematikern (N. ROBERTSON, D. P. SANDERS, P. D. SEYMOUR und R. THOMAS), den Computerbeweis nachzuvollziehen und entschloß sich ziemlich schnell, lieber ganz von vorn anzufangen und mit neuen Entladungsprozeduren und neuen Programmen zur Reduzibilität einen neuen Beweis zu konstruieren – was Ihnen auch gelang. Ein Mathematiker, der ihren Beweis nachvollziehen möchte, sieht sich wahrscheinlich vor dasselbe Problem gestellt und wird es ebenfalls einfacher finden, von vorn anzufangen statt den Beweis zu lesen. Was soll man von solchen Beweisen halten, die (außer hoffentlich den Autoren) niemand nachvollziehen kann?

Hinzu kommt, daß auch Computer alles andere als unfehlbar sind: Softwarefehler gab es auch schon vor Microsoft, und Hardwarefehler sind zwar inzwischen recht selten, kommen aber immer noch vor.

Auch die zweite Säule des Beweises, die computerunabhängige Konstruktion einer Liste unvermeidbarer Konfigurationen, entspricht nicht dem Idealbild eines mathematischen Beweises: Wer kann sicher sein, daß ein Autor bei der Betrachtung von 1818 Fällen, die (einschließlich der Zeichnungen, die in der Originalveröffentlichung zunächst nur auf Microfiche beilagen) rund 400 Druckseiten einnehmen, wirklich keinen Fehler gemacht hat?

Natürlich haben die Autoren (zum Teil mit Hilfe ihrer Familien) alle Details noch einmal überprüft und auch eine Routine zur Fehlerbehebung entwickelt, die auch bei den nach der Veröffentlichung gefundenen Fehlern bisher noch nie versagt hat, aber warum das immer funktionieren *muß,* versteht kein Mathematiker wirklich. Den meisten Mathematikern wäre daher nicht nur des Computers wegen ein „klassischer“ Beweis des Vierfarbensatzes lieber.

Angesichts der Berühmtheit des Vierfarbenproblems gab es in den letzten hundert Jahren immer wieder solche „Beweise“, meist von Autoren, die mangels Verständnis und/oder Erfahrung sehr viel elementarere Fehler machten als seinerzeit KEMPE.

§9: Literaturhinweise zum Vierfarbenproblem

Eine im großen und ganzen recht elementare Einführung in den Problemkreis geben die beiden Bücher

RUDOLF und GERDA FRITSCH: Der Vierfarbensatz. Geschichte, Topologische Grundlagen und Beweisidee, *Bibliographisches Institut Mannheim,* 1994 und

T.L. SAATY, P.C. KAINEN: The Four-Color Probem: Assaults and Conquest, *McGraw Hill, New York,* 1977; Nachdruck bei *Dover,* 1986

Den vollständigen Beweis zusammen mit einer auch für Nichtexperten gut lesbaren Einleitung findet man in

K. APPEL und W. HAKEN: Every planar map is four-colorable, Band **98** der Reihe *Contemporary Mathematics, American Mathematical Society, Providence RI,* 1989

Eine allgemeinverständliche Einführung zum neuen Beweis (mit Referenzen) bietet

N. ROBERTSON, D.P. SANDERS, P.D. SEYMOUR, R. THOMAS: The Four-Color Theorem,
`http://www.math.gatech.edu/~thomas/FC/fourcolor.html`

Auch die Biographie (stellenweise eher Hagiographie)

HANS-JÜRGEN BIGALKE: Heinrich Heesch. Kristallgeometrie, Parkettierungen, Vierfarbenforschung, *Band 3 der Reihe Vita Mathematica, Birkhäuser, Basel, Boston, Stuttgart,* 1988

enthält eine Darstellung der Grundideen des Beweises, die sich speziell an Nichtmathematiker wendet. In erster Linie mit philosophischen und wissenschaftstheoretischen Konsequenzen beschäftigen sich

DONALD MACKENZIE: Slaying the Kraken: The Sociohistory of a Mathematical Proof, *Social Studies in Science* **29** (1999), S. 7–60 und

THOMAS TYMOCZKO: The Four-Color Problem and Its Philosophical Significance, *in:* THOMAS TYMOCZKO [HRSG]: New Directions in the Philosophy of Mathematics, *Birkäuser, Boston, Basel, Stuttgart,* 1994

Beide stellen auch wesentliche Ideen des Beweises für ihre Leserschaft aus u.a. Soziologen und Philosophen vor, MACKENZIE mit erheb-

lich mehr Details als TYMOCZKO. Ansonsten geht es bei MACKENZIE eher darum, wie die wissenschaftliche Gemeinschaft auf den Beweis reagierte, bei TYMOCZKO dagegen eher um grundsätzliche Überlegungen aus philosophischer Sicht.

Kapitel 5
Simpliziale Komplexe

In diesem Kapitel geht es darum, topologische Räume aus einfachen Bausteinen zusammenzusetzen; im nächsten Kapitel werden wir dann den so konstruierten Räumen Gruppen und Zahlen zuordnen, von denen wir sehen werden, daß sie nur vom Homöomorphietyp des Raums abhängen.

Unsere Bausteine sind (offene) Simplizes:

Definition: *a)* $k+1$ Punkte $P_0, \ldots, P_k$ aus $\mathbb{R}^n$ sind *in allgemeiner Lage,* wenn die Vektoren $\overrightarrow{P_0P_1}, \ldots, \overrightarrow{P_0P_k}$ linear unabhängig sind.
b) Sind $P_0, \ldots, P_k \in \mathbb{R}^n$ Punkte in allgemeiner Lage, so heißt

$$s = \left\{ \sum_{i=0}^{k} \lambda_i P_i \;\middle|\; \lambda_i > 0, \;\; \sum_{i=0}^{k} \lambda_i = 1 \right\}$$

das von $P_0, \ldots, P_k$ aufgespannte (offene) k-*Simplex.* Die Punkte P_i bezeichnen wir als die *Ecken* von s.
c) Die *Randsimplizes* von s sind jene Simplizes, die von Teilmengen von $\{P_0, \ldots, P_k\}$ aufgespannt werden.
d) Das von $P_0, \ldots, P_k$ aufgespannte *abgeschlossene* k-*Simplex* ist

$$\bar{s} = \left\{ \sum_{i=0}^{k} \lambda_i P_i \;\middle|\; \lambda_i \geq 0, \;\; \sum_{i=0}^{k} \lambda_i = 1 \right\}.$$

Offensichtlich sind zwei Punkte genau dann in allgemeiner Lage, wenn sie verschieden sind, drei Punkte sind genau dann in allgemeiner Lage, wenn sie nicht auf einer Geraden liegen, vier Punkte genau dann, wenn

sie nicht in einer Ebenen liegen, und so weiter. Ein 1-Simplex ist eine Strecke, ein 2-Simplex ein Dreieck und ein 3-Simplex ein Tetraeder. Das von $k+1$ Punkten $P_0, \ldots, P_k$ aufgespannte abgeschlossene Simplex ist die disjunkte Vereinigung des von diesen Punkten aufgespannten Simplex zusammen mit dessen sämtlichen Randsimplizes.

Definition: *a)* Ein (endlicher) geometrischer simplizialer Komplex K ist eine endliche Menge von (offenen) Simplizes aus einem festen $\mathbb{R}^n$ mit folgenden Eigenschaften:

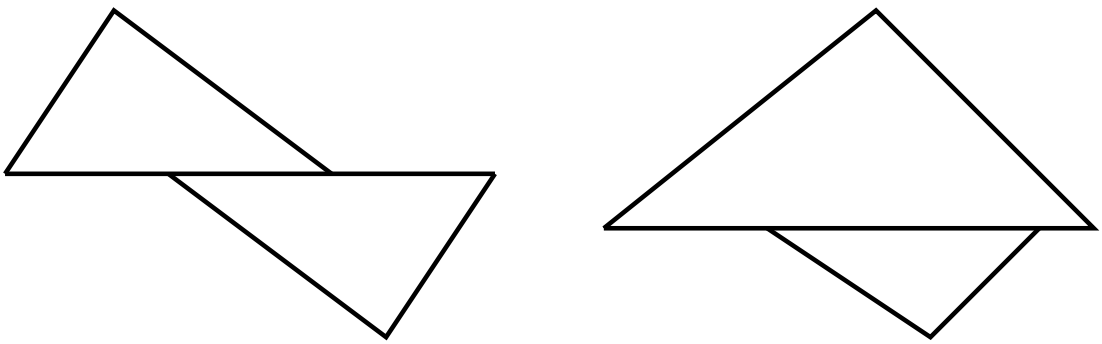
1. Für jedes Simplex $s \in K$ liegen auch dessen sämtliche Randsimplizes in K.
2. Zwei verschiedene Simplizes aus K haben leeren Durchschnitt.

b) Die *geometrische Realisierung* von K ist die Menge

$$|K| = \bigcup_{s \in K} s \subset \mathbb{R}^n ,$$

versehen mit ihrer Spurtopologie als Teilmenge von $\mathbb{R}^n$.

Die zweite Bedingung an einen Komplex besagt anschaulich interpretiert, daß wir Simplizes nur entlang gemeinsamer Randsimplizes aneinander setzen dürfen; Komplexe wie die unten eingezeichneten sind also nicht erlaubt.



Die Topologie beschränkt sich keinesfalls nur auf endliche simpliziale Komplexe; da wir aber keine unendlichen Komplexe brauchen, soll hier die Endlichkeit stets vorausgesetzt werden. Da ein simplizialer Komplex mit jedem seiner Simplizes auch dessen sämtliche Randsimplizes

enthält, enthält er auch dessen Abschluß; $|K|$ ist daher auch die Vereinigung aller $\bar{s}$ mit $s \in K$. Da abgeschlossene Simplizes kompakt sind, folgt aus der vorausgesetzten Endlichkeit von K daher insbesondere, daß $|K|$ stets eine kompakte Menge ist.

Definition: Ein kompakter topologischer Raum X heißt *triangulierbar,* wenn es einen simplizialen Komplex K und einen Homöomorphismus $f\colon |K| \to X$ gibt.

Somit ist beispielsweise die n-dimensionale *Vollkugel*

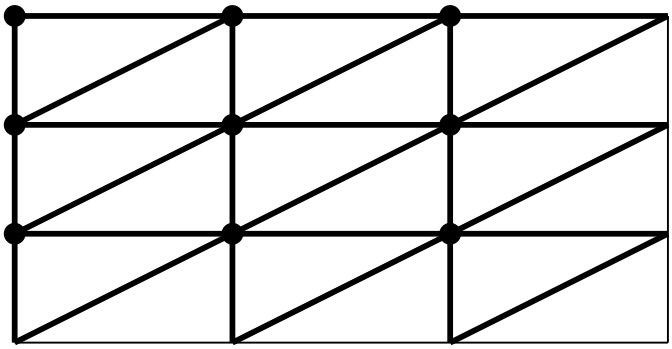
$$\mathbb{B}^n = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} x_i^2 \leq 1\right\}$$

triangulierbar, denn sie ist homöomorph zu einem abgeschlossenen n-Simplex, und dieses wiederum gehört zu jenem Komplex K, der aus dem zugehörigen offenen n-Simplex und dessen sämtlichen Randsimplizes besteht. Auch der Rand von $\mathbb{B}^n$, die $(n-1)$-*Sphäre*

$$\mathbb{S}^{n-1} = \left\{(x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; \sum_{i=1}^{n} x_i^2 = 1\right\}$$

ist triangulierbar; hierzu müssen wir einfach das einzige n-Simplex aus dem Komplex K herausnehmen.

Auch der Torus ist triangulierbar: Wir konstruieren ihn aus einem Rechteck, bei dem gegenüberliegende Seiten identifiziert werden. Zur Zerlegung in Dreiecke reicht es nun aber nicht, das Rechteck einfach in zwei Dreiecke zu zerlegen: Durch die Identifikation gegenüberliegender Seiten werden schließlich alle vier Eckpunkte des Rechtecks zu einem einzigen Punkt des Torus, die beiden „Dreiecke“ haben also auf dem Torus nur noch eine einzige Ecke, die zudem noch beiden gemeinsam ist. Es reicht auch nicht, wenn wir jede Kante des Rechtecks in der Mitte unterteilen: Ist etwa AB die obere Kante und M deren Mittelpunkt, so haben nach der Identifikation von A mit B die beiden Kanten AM und MB auf dem Torus beide die gleichen Endpunkte, wir hätten also für zwei Ecken gleich zwei Kanten, die diese verbinden. Um dies zu vermeiden, müssen wir jede Kante mindestens in drei Teilstücke zerlegen, das Rechteck also in neun Teilrechtecke. Diese können wir dann durch



eine Diagonale in jeweils zwei Dreiecke zerlegen und erhalten so einen simplizialen Komplex aus 18 Dreiecken.

Beim Zählen der Kanten müssen wir auf Identifizierungen achten: Von den vier senkrechten Strichen, die jeweils dreifach unterteilt werden, sind der erste und der letzte auf dem Torus identisch, als gibt es dort nur $3 \times 3 = 9$ Kanten, die senkrechten strichen im Rechteck entsprechen. genauso gibt es auch nur neun Kanten zu waagrechten Strichen und zusätzlich noch die neun Diagonalen. Insgesamt haben wir somit 27 Kanten, die wir zum Beispiel durch die im Bild etwas dicker eingezeichneten Kanten im Rechteck repräsentieren können.

Bleiben noch die Ecken. Da Oberkante und Unterkante sowie linke und rechte Kante des Rechtecks identifiziert werden, dürfen wir jeweils nur eine davon zählen und kommen so auf neun Ecken; auch hier sind Repräsentanten fett markiert. Für diese Triangulierung ist somit

$$\#\,\text{Ecken} - \#\,\text{Kanten} + \#\,\text{Flächen} = 9 - 27 + 18 = 0\,;$$

hier gilt also der EULERsche Polyedersatz nicht. Wir werden allerdings im nächsten Kapitel sehen, daß für jede Triangulierung des Torus die obige alternierende Summe verschwindet, so daß wir auch hier eine Gesetzmäßigkeit haben.

Da je zwei n-Simplizes homöomorph sind, ist es für die Topologie eines simplizialen Komplexes irrelevant, wo seine Ecken im $\mathbb{R}^n$ liegen; wir müssen nur sicherstellen, daß sich keine zwei Simplizes schneiden. Dies kann im Einzelfall durchaus schwierig sein; deshalb wollen wir simpliziale Komplexe auch abstrakt ohne Bezug zu einem $\mathbb{R}^n$ definieren. Auch hier beschränken wir uns wieder auf endliche Komplexe:

Definition: Ein *abstrakter simplizialer Komplex* ist ein Paar $\mathfrak{K} = (E, S)$ aus einer endlichen Menge E, deren Elemente als *Ecken* bezeichnet werden, sowie einer Menge $S \subseteq \mathfrak{P}(E)$ von nichtleeren Teilmengen von E, die mit einer Menge $\sigma \in S$ auch jede nichtleere Teilmenge von σ enthält. Eine $(q+1)$-elementige Menge $\sigma \in S$ wird als *abstraktes q-Simplex* bezeichnet.

Natürlich können wir jedem geometrischen simplizialen Komplex K einen abstrakten simplizialen Komplex zuordnen: Wir nehmen die Menge E aller Ecken von K und ordnen jedem Simplex von K als abstraktes Simplex die Menge seiner Ecken zu.

Weniger offensichtlich ist, daß wir auch jedem abstrakten simplizialen Komplex $\mathfrak{K}$ einen geometrischen zuordnen können: Dazu sei n die Anzahl der Ecken von $\mathfrak{K}$. Wir wählen im $\mathbb{R}^n$ irgendwelche n Punkte mit linear unabhängigen Ortsvektoren und ordnen jeder Ecke einen dieser Punkte zu. Wegen der linearen Unabhängigkeit der Ortsvektoren ist jede Teilmenge dieser Punktmenge in allgemeiner Lage; falls die Teilmenge zur Eckenmenge eines Simplex aus $\mathfrak{K}$ gehört, spannen die entsprechenden Punkte daher ein Simplex der richtigen Dimension auf, das wir in den Komplex K aufnehmen. Ebenfalls wegen der linearen Unabhängigkeit können zwei verschiedene solche (offene) Simplizes auch keine Punkte gemeinsamen haben; zwei abgeschlossene Simplizes schneiden sich also höchstens in einem gemeinsamen Randsimplex. Auf diese Weise können wir jedes abstrakte Simplex aus $\mathfrak{K}$ als geometrisches Simplex in $\mathbb{R}^n$ realisieren und erhalten einen geometrischen simplizialen Komplex K. (Im allgemeinen wird es natürlich geometrische Realisierungen auch schon in Räumen mit erheblich niedrigerer Dimension geben; in der Tat kann man mit etwas mehr Aufwand zeigen, daß ein abstrakter simplizialer Komplex, der kein q-Simplex mit $q > n$ enthält, durch einen geometrischen simplizialen Komplex in $\mathbb{R}^{2n+1}$ realisiert werden kann. Da uns topologische Räume in erster Linie *unabhängig* von einer Einbettung in einen umgebenden Raum interessierten, lohnt sich der Aufwand für diesen umfangreicheren Beweis jedoch nicht.)

Kapitel 6
Die Homologie simplizialer Komplexe

In diesem Abschnitt wollen wir einem simplizialen Komplex abelsche Gruppen und damit auch Zahlen zuordnen, die uns Informationen über die Geometrie des Komplexes geben. Wie wir am Ende sehen werden, hängen diese Daten nicht vom simplizialen Komplex ab, sondern nur von dessen topologischer Realisierung; sie geben also echte geometrische Information.

Beginnen wir mit einigen Aussagen über abelsche Gruppen. Eine abelsche Gruppe ist bekanntlich eine Gruppe, deren Gruppenoperation kommutativ ist; wir beschreiben diese Operation durch das Zeichen „+“ und nennen das neutrale Element Null. Auf einer abelschen Gruppe A gibt es eine natürliche Operation des Rings $\mathbb{Z}$ der ganzen Zahlen: Für $\lambda \in \mathbb{Z}$ und $a \in A$ setzen wir

$$\lambda a = \begin{cases} a + \cdots + a \text{ (}\lambda \textit{ Faktoren)} & \text{für } \lambda > 0 \\ 0 & \text{für } \lambda = 0 \\ -\big((-\lambda)a\big) & \text{für } \lambda < 0 \end{cases} .$$

In völliger Analogie zu den Vektorraumaxiomen gilt:

$$\lambda(a + b) = \lambda a + \lambda b$$

$$(\lambda + \mu)a = \lambda a + \mu a$$

$$1a = a \quad \text{und} \quad 0a = 0 .$$

Wir nennen eine Teilmenge $\{a_1, \ldots, a_r\}$ einer abelschen Gruppe *linear unabhängig,* wenn gilt: Eine Beziehung der Form

$$\lambda_1 a_1 + \cdots + \lambda_r a_r = 0$$

kann nur gelten, wenn alle λ_i verschwinden. Die maximale Mächtigkeit einer linear unabhängigen Teilmenge von A heißt *Rang von* A und

wird mit $\mathrm{rg}(A)$ abgekürzt. Beispielsweise ist der Rang einer endlichen abelschen Gruppe stets Null, denn für jedes $a \in A$ gibt es ein $\lambda \in \mathbb{Z}\backslash\{0\}$, so daß $\lambda a = 0$ ist.

Beispiele für Gruppen mit positivem Rang sind die *freien abelschen Gruppen:* Eine abelsche Gruppe A heißt *frei,* wenn sie isomorph ist zu einer der Gruppen $\mathbb{Z}^r$ für ein $r \geq 0$, das auch unendlich sein kann. Offensichtlich hat eine solche Gruppe den Rang r mit den Urbildern der Einheitsvektoren $(1, 0, 0, \dots)$, $(0, 1, 0, \dots)$ als Elementen einer maximalen linear unabhängigen Teilmenge. Allgemeiner bezeichnen wir zu beliebiger vorgegebener Menge M die Gruppe

$$\mathbb{Z}^M = \bigoplus_{m \in M} \mathbb{Z}m$$

als *freie abelsche Gruppe über* M. Die Elemente dieser Gruppe lassen sich also als endliche Linearkombinationen der Elemente von M schreiben. Die freie abelsche Gruppe über der leeren Menge soll dabei einfach die triviale Gruppe $\{0\}$ sein.

Wie bei Vektorräumen gilt auch hier:

Lemma: Alle maximalen linear unabhängigen Teilmengen einer abelschen Gruppe A haben die gleiche Mächtigkeit.

Beweis: $\{a_1, a_2, \dots\}$ und $\{b_1, b_2, \dots\}$ seien zwei maximale linear unabhängige Teilmengen von A. Dann enthält A die Untergruppe

$$U = \mathbb{Z}a_1 \oplus \mathbb{Z}a_2 \oplus \cdots \leq A,$$

und natürlich läßt sich U einbetten in den $\mathbb{R}$-Vektorraum

$$V = \mathbb{R}a_1 \oplus \mathbb{R}a_2 \oplus \cdots .$$

Da die a_i eine maximale linear unabhängige Menge bilden, muß es zu jedem b_j ein $\lambda_j \neq 0$ geben, so daß $\lambda_j b_j$ in U liegt; insbesondere bilden die $\lambda_j b_j$ also eine linear unabhängige Teilmenge von V, die wegen ihrer Maximalität in A sogar eine Basis von V sein muß. Da alle Basen eines Vektorraums die gleiche Mächtigkeit haben, ist der Satz bewiesen. ∎

Korollar: Für eine Untergruppe U einer abelschen Gruppe A gilt:

$$\mathrm{rg}(A/U) = \mathrm{rg}(A) - \mathrm{rg}(U).$$

Beweis: Ergänze eine maximale linear unabhängige Teilmenge von U zu einer maximalen linear unabhängigen Teilmenge von A; die Restklassen der neu hinzukommenden Elemente bilden eine maximale linear unabhängige Teilmenge von A/U.

■

Zum Schluß sei noch der Vollständigkeit halber der *Struktursatz für endlich erzeugbare abelsche Gruppen* angegeben:

Satz: Jede endlich erzeugbare abelsche Gruppe vom Rang r ist isomorph zu einer Gruppe der Form

$$\mathbb{Z}^r \oplus \bigoplus_{i=1}^{n} \mathbb{Z}/p_i^{e_i}$$

mit eindeutig bestimmten Primzahlpotenzen $p_i^{e_i}$.

Dieser Satz ist ein Korollar des *Elementarteilersatzes;* für seinen Beweis sei auf Vorlesungen über (kommutative) Algebra verwiesen. Für diese Vorlesung wird er nicht benötigt.

In der Topologie treten abelsche Gruppen meist in *Komplexen* auf:

Definition: Ein *Kettenkomplex* ist ein System $\mathcal{C}$ von abelschen Gruppen C_q, $q \geq 0$, zusammen mit Homomorphismen $\partial_q \colon C_q \to C_{q-1}$ derart, daß $\partial_{q-1} \circ \partial_q = 0$ ist für alle q. Dabei setzen wir formal $C_{-1} = 0$ und $\partial_0 = 0$. Die Elemente von C_q heißen *q-Ketten,* die Abbildungen ∂_q heißen *Randabbildungen.* Eine q-Kette c heißt *q-Zyklus,* wenn $\partial_q(c) = 0$ ist; sie heißt *q-Rand,* wenn sie im Bild von C_{q+1} liegt. Insbesondere ist jeder q-Rand ein q-Zyklus, denn $\partial_q \circ \partial_{q+1} = 0$. Die Gruppe aller q-Zykeln von $\mathcal{C}$ wird mit $Z_q(\mathcal{C})$ bezeichnet, die der q-Ränder mit $B_q(\mathcal{C})$. Die Faktorgruppe $H_q(\mathcal{C}) = Z_q(\mathcal{C})/B_q(\mathcal{C})$ heißt q-te *Homologiegruppe* von $\mathcal{C}$.

Da ∂_q nur auf C_q definiert ist, werden wir anstelle von $\partial_q(c)$ im Allgemeinen einfach $\partial(c)$ schreiben, denn q ist durch c eindeutig bestimmt.

Wir wollen nun jedem abstrakten simplizialen Komplex $\mathfrak{K}$ einen Kettenkomplex $\mathcal{C}(\mathfrak{K})$ zuordnen. Dazu müssen wir zunächst die Simplizes von $\mathfrak{K}$ orientieren:

Definition: Eine Orientierung eines abstrakten Simplex $\{v_0, \ldots, v_q\}$ ist eine Festlegung der Reihenfolge der v_i bis auf gerade Permutationen. Ein orientiertes Simplex $[v_0, \ldots, v_q]$ ist ein Simplex $\{v_0, \ldots, v_q\}$ zusammen mit der Orientierung, zu der die angegebene Reihenfolge der v_i gehört.

Offensichtlich hat also jedes q-Simplex mit $q \neq 0$ genau zwei verschiedene Orientierungen. Für den Rest dieses Paragraphen wollen wir annehmen, daß wir stets eine Orientierung der sämtlichen Simplizes von $\mathfrak{K}$ vorgegeben haben.

Definition: $C_q(\mathfrak{K})$, die Gruppe der q-Ketten von $\mathfrak{K}$, ist die freie abelsche Gruppe über der Menge der q-Simplizes von $\mathfrak{K}$.

Ist σ ein orientiertes Simplex aus $\mathfrak{K}$ und τ dasselbe Simplex mit der entgegengesetzten Orientierung, so identifizieren wir τ mit der Kette $-\sigma$. Es ist klar, daß $C_q(\mathfrak{K})$ nicht von der vorgegebenen Orientierung der Simplizes abhängt: Eine andere Orientierung bedeutet nur, daß wir eine andere Basen derselben Kettengruppe betrachten.

Um Randabbildungen definieren zu können, müssen wir zunächst wissen, was der Rand eines orientierten q-Simplex ist. Wir definieren

$$\partial_q\big([v_0, \ldots, v_q]\big) \underset{\text{def}}{=} \sum_{i=0}^{q} (-1)^i [v_0, \ldots, \widehat{v_i}, \ldots, v_q],$$

wobei das Dach über v_i bedeuten soll, daß diese Ecke ausgelassen wird:

$$[v_0, \ldots, \widehat{v_i}, \ldots, v_q] \underset{\text{def}}{=} [v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_q].$$

Die $(q-1)$-Kette $\partial_q[v_0, \ldots, v_q]$ hängt nur von der Orientierung des Simplex ab, nicht von der genauen Reihenfolge der v_i: Vertauschen wir zwei Ecken v_r und v_s, so ändert $[v_0, \ldots, \widehat{v_i}, \ldots, v_q]$ für $i \neq r, s$ seine Orientierung, wird also mit (-1) multipliziert. Für $i = r$ oder s werden die beiden Randsimplizes zu den Indizes r und s miteinander vertauscht

und mit $(-1)^{r-s-1}$ multipliziert, denn wir brauchen $r-s-1$ Transpositionen um v_r bzw. v_s wieder auf seine „richtige“ Position zu bringen. Außerdem wird aber auch der Vorfaktor $(-1)^i$ noch mit $(-1)^{r-s}$ multipliziert, insgesamt gesehen wird also jedes Randsimplex mit (-1) multipliziert und damit der gesamte Rand. Da sich jede gerade Permutation als Produkt einer geraden Anzahl von Transpositionen schreiben läßt, folgt die Behauptung.

Nachdem wir den Rand eines einzelnen orientierten Simplex definiert haben, können wir durch lineare Fortsetzung auch den Rand einer beliebigen q-Kette definieren: Wie ober vereinbart nehmen wir dazu an, daß alle Simplizes im Komplex $\mathfrak{K}$ irgendwie orientiert sind und betrachten die Abbildung

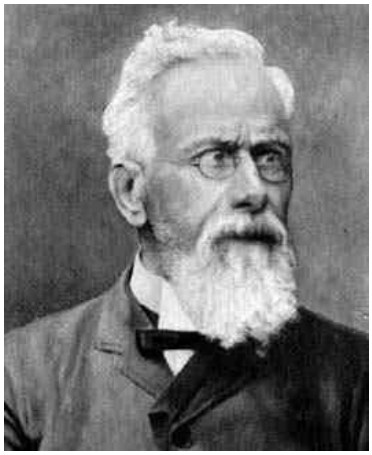
$$\partial_q\colon \begin{cases} C_q(\mathfrak{K}) & \to\ C_{q-1}(\mathfrak{K}) \\ \sum \lambda_i \sigma_i & \mapsto\ \sum \lambda_i \partial_q(\sigma_i) \end{cases} .$$

Lemma: $\partial_{q-1} \circ \partial_q = 0$.

Beweis: Es genügt natürlich, dies für jedes q-Simplex einzeln nachzurechnen:

$$\begin{aligned}
&\partial_{q-1} \circ \partial_q\big([v_0, \ldots, v_q]\big) \\
&= \partial_{q-1}\Big(\sum (-1)^i [v_0, \ldots, \widehat{v_i}, \ldots, v_q]\Big) \\
&= \sum (-1)^i \partial_{q-1}\big([v_0, \ldots, \widehat{v_i}, \ldots, v_q]\big) \\
&= \sum_i \sum_{j<i} (-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_q] \\
&\quad + \sum_i \sum_{j>i} (-1)^{i+j-1} [v_0, \ldots, \widehat{v_i}, \ldots, \widehat{v_j}, \ldots, v_q] \\
&= \sum_i \sum_{j<i} (-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_q] \\
&\quad - \sum_j \sum_{i<j} (-1)^{i+j} [v_0, \ldots, \widehat{v_j}, \ldots, \widehat{v_i}, \ldots, v_q] \\
&= 0 .
\end{aligned}$$

Damit haben wir also jedem abstrakten simplizialen Komplex $\mathfrak{K}$ einen Kettenkomplex $C(\mathfrak{K})$ zugeordnet; seine Zykeln, Ränder und Homologie bezeichnen wir kurz auch als Zykeln, Ränder und Homologie von $\mathfrak{K}$. Insbesondere schreiben wir auch $H_q(\mathfrak{K})$ für $H_q\big(C(\mathfrak{K})\big)$. Der Rang β_q von $H_q(\mathfrak{K})$ wird als q-te BETTI-Zahl des simplizialen Komplexes $\mathfrak{K}$ bezeichnet.



ENRICO BETTI (1823–1892) wurde in Pistoia in der Toskana geboren, wo er auch aufwuchs und zur Schule ging. Danach studierte er Mathematik und Physik an der Universität Pisa und bekam dann dort 1846 eine Stelle als Assistent. Da er mit den damit verbundenen Einkünften nicht leben konnte, kehrte er 1849 nach Pistoia zurück, um dort und später in Florenz als Mathematiklehrer zu arbeiten. 1857 wurde er Professor für höhere Algebra, 1859 Professor für höhere Geometrie und Analysis, 1864 Professor für theoretische Physik und 1870 Professor für Himmelsmechanik, jeweils an der Universität Pisa, deren Rektor er auch zeitweise war. Was wir (seit POINCARÉ) als BETTI-Zahlen bezeichnen kommt erstmalig 1871 in seiner Arbeit *Sopra gli spazi di un numero qualunque di dimensioni* vor. Berühmt waren auch seine Arbeiten zur exakten Begründung der GALOIS-Theorie, und natürlich arbeitete er, wie die Liste seiner Professuren zeigt, auch viel über angewandte Mathematik und Physik.

Beispiel 1: Die Kreislinie: $\mathfrak{K}$ sei ein abstrakter simplizialer Komplex, dessen topologische Realisierung die Kreislinie $\mathbb{S}^1$ ist. Ein solcher Komplex ist geometrisch gesehen natürlich nichts anderes als der Kantenzug eines n-Ecks. Wir orientieren die 1-Simplizes so, daß jede Ecke sowohl Anfangspunkt als auch Endpunkt einer Kante ist; wenn wir die Orientierung einer Kante willkürlich vorgeben, ist dadurch die Orientierung aller weiteren Kanten eindeutig bestimmt.

Da es in $\mathfrak{K}$ keine q-Simplizes mit $q \geq 2$ gibt, ist $H_q(\mathfrak{K}) = 0$ für $q \geq 2$, wir müssen also nur $H_0(\mathfrak{K})$ und $H_1(\mathfrak{K})$ berechnen.

Offensichtlich ist jedes 0-Simplex ein Zyklus, denn ∂_0 ist ja nach Definition gleich der Nullabbildung, und die Differenz zweier 0-Simplizes ist ein Rand, nämlich der Rand eines zusammenhängenden Kantenzugs, der die beiden Punkte miteinander verbindet. Also ist für zwei Nullsimplizes σ_i und σ_j die Kette $\sigma_i - \sigma_j$ ein Rand, und $B_0(\mathfrak{K})$ ist gerade der

Kern der Abbildung

$$Z_0(\mathfrak{K}) = C_0(\mathfrak{K}) \to \mathbb{Z}; \quad \sum \lambda_i \sigma_i \mapsto \sum \lambda_i,$$

d.h. $H_0(\mathfrak{K}) \cong \mathbb{Z}$.

Eine 1-Kette kann nur dann ein Zyklus sein, wenn sie mit jeder Kante auch deren Vorgänger- und Nachfolgerkante mit genau der gleichen Vielfachheit enthält; ist also c die Summe aller 1-Simplizes, so ist $Z_1(\mathfrak{K}) = \mathbb{Z}c$. Da es keine 2-Ketten gibt, gibt es auch keine 1-Ränder, d.h. $H_1(\mathfrak{K}) = Z_1(\mathfrak{K}) \cong \mathbb{Z}$.

Beispiel 2: Der Torus: Sei $\mathfrak{K}$ eine seiner Triangulierungen. Da der Torus als $\mathbb{R}^2/\mathbb{Z}^2$ geschrieben werden kann, läßt sich $\mathfrak{K}$ durch periodische Fortsetzung zu einer Triangulierung $\mathfrak{K}'$ der Ebene fortsetzen. Genauer: Das Urbild eines q-Simplex aus $|\mathfrak{K}|$ unter der Quotientenabbildung $\mathbb{R}^2 \to \mathbb{R}^2/\mathbb{Z}^2$ zerfällt in unendlich viele q-Simplizes, und die Menge aller so erhaltener Simplizes bildet eine Triangulierung von $\mathbb{R}^2$. Den zugehörigen unendlichen abstrakten simplizialen Komplex bezeichnen wir mit $\mathfrak{K}'$ und wir wählen die Orientierungen seiner Dreiecke so, daß sie alle im Uhrzeigersinn durchlaufen werden. Dies definiert dann auch eine Orientierung der Dreiecke aus $\mathfrak{K}$, denn die verschiedenen Urbilder eines solchen Dreiecks gehen durch Parallelverschiebung auseinander hervor, sind also in $\mathfrak{K}'$ konsistent orientiert. Die Kanten seien irgendwie orientiert.

Wie üblich *(?)* ist auch hier wieder $H_0(\mathfrak{K}) = \mathbb{Z}$, gehen wir also gleich zu H_1. Ein 1-Zyklus z von $\mathfrak{K}$ ist eine endliche Summe von geschlossenen zusammenhängenden Kantenzügen auf $\mathfrak{K}$; betrachten wir einen davon. O.B.d.A. sei die Identifikation des Torus mit $\mathbb{R}^2/\mathbb{Z}^2$ so gewählt, daß das Bild von $(0,0)$ eine Ecke von z ist. Dann läßt sich z liften zu einem zusammenhängenden Kantenzug c in $\mathfrak{K}'$, der in $(0,0)$ beginnt und in einem Punkt $(n,m) \in \mathbb{Z}^2$ endet. Seien c_1 und c_2 zwei fest gewählte zusammenhängende Kantenzüge aus $\mathfrak{K}'$ von $(0,0)$ nach $(0,1)$ bzw. $(1,0)$, in denen keine Kante doppelt vorkommt und keine Ecke zu mehr als zwei Kanten gehört. Dann ist

$$d = c - mc_1 - nc_2$$

ein 1-Zyklus von $\mathfrak{K}'$. Dies ist ein geschlossener endlicher Kantenzug in der Ebene, also die Vereinigung von endlich vielen geschlossenen Kantenzügen, die jeweils irgendwelche n-Ecke bilden, und die wiederum sind die Ränder der Summe der in ihnen liegenden Dreiecke, Dies zeigt, daß d sogar ein 1-Rand ist, und damit ist auch das Bild von d in $\mathfrak{K}$ ein 1-Rand. Sind also z_1 und z_2 die Bilder von c_1 und c_2 in $\mathfrak{K}$, so sind z_1 und z_2 1-Zykeln mit der Eigenschaft, daß es für jeden beliebigen 1-Zyklus z von $\mathfrak{K}$ ganze Zahlen n, m gibt, so daß $z - mz_1 - nz_2$ ein Rand ist, $H_1(\mathfrak{K}) \cong \mathbb{Z}^2$ wird also von den Äquivalenzklassen von z_1 und z_2 modulo den Rändern erzeugt. Geometrisch gesehen sind z_1 und z_2 gerade die beiden Kreise, deren Produkt der Torus ist.

Ein 2-Zyklus, schließlich, muß auch hier wieder mit jedem Dreieck auch dessen sämtliche Nachbarn enthalten, und da $\mathfrak{K}$ ein endlicher Komplex ist, ist das auch möglich: Der Zyklus muß einfach ein ganzzahliges Vielfaches der Summe aller Dreiecke sein. Da es keine 2-Ränder gibt, ist also $H_2(\mathfrak{K}) = Z_2(\mathfrak{K}) \cong \mathbb{Z}$.

In beiden hier betrachteten Beispielen hingen die Homologiegruppen nicht vom simplizialen Komplex $\mathfrak{K}$ ab, sondern nur vom topologischen Raum $|\mathfrak{K}|$. Im Rest dieses Paragraphen geht es hauptsächlich um den Beweis, daß dies allgemein so ist. Zuerst möchte ich aber eine schwache Version davon beweisen, eine Art Verallgemeinerung des EULERschen Polyedersatzes:

Satz: $\mathfrak{K}$ sei ein endlicher simplizialer Komplex der Dimension n; die Anzahl seiner q-Simplizes sei α_q. Dann ist die alternierende Summe der α_q gleich der alternierenden Summe der BETTI-Zahlen von $\mathfrak{K}$:

$$\sum_{q=0}^{n}(-1)^q\alpha_q = \sum_{q=0}^{n}(-1)^q\beta_q$$

Beweis: Da $Z_q(\mathfrak{K})$ der Kern von ∂_q ist und $B_{q-1}(\mathfrak{K})$ das Bild, ist

$$C_q(\mathfrak{K})/Z_q(\mathfrak{K}) \cong B_{q-1}(\mathfrak{K}),$$

also

$$\alpha_q = \operatorname{rg}\big(C_q(\mathfrak{K})\big) = \operatorname{rg}\big(Z_q(\mathfrak{K})\big) + \operatorname{rg}\big(B_{q-1}(\mathfrak{K})\big).$$

Genauso folgt wegen $H_q(\mathfrak{K}) = Z_q(\mathfrak{K})/B_q(\mathfrak{K})$:

$$\beta_q = \mathrm{rg}\big(H_q(\mathfrak{K})\big) = \mathrm{rg}\big(Z_q(\mathfrak{K})\big) - \mathrm{rg}\big(B_q(\mathfrak{K})\big).$$

Mithin ist

$$\beta_q = \alpha_q - \mathrm{rg}\big(B_{q-1}(\mathfrak{K})\big) - \mathrm{rg}\big(B_q(\mathfrak{K})\big),$$

wobei formal $B_{-1}(\mathfrak{K}) = 0$ zu setzen ist. Bildet man die alternierende Summe, so heben sich wegen $B_n(\mathfrak{K}) = 0$ die Ränge der Gruppen der Ränder gerade weg, womit die Behauptung gezeigt wäre. ■

Als nächstes wollen wir sehen, daß die Homologiegruppen eines simplizialen Komplexes $\mathfrak{K}$ nur vom topologischen Raum $|\mathfrak{K}|$ abhängen. Die Beweisstrategie wird folgende sein: Wir betrachten zunächst spezielle Unterteilungen von endlichen simplizialen Komplexen und zeigen, daß eine solche Unterteilung eines Komplexes dieselbe Homologie hat wie der Komplex selbst. Dann folgern wird, daß dies auch noch für allgemeinere Unterteilungen gilt und zeigen, daß zwei endliche simpliziale Komplexe eine gemeinsame verallgemeinerte Unterteilung und damit die gleiche Homologie haben. Dies beweist den Satz für kompakte Räume $|\mathfrak{K}|$.

1. Schritt: Unterteilungen

Beginnen wir mit Unterteilungen. Der Bequemlichkeit halber arbeiten wir im Rest dieses Paragraphen mit *geometrischen* simplizialen Komplexen, alle Simplizes sind also geometrische Simplizes in einem festen $\mathbb{R}^N$. Da jeder abstrakte simpliziale Komplex $\mathfrak{K}$ eine geometrische Realisierung K hat, bedeutet dies keine Einschränkung der Allgemeinheit.

Umgekehrt bestimmt ein geometrischer simplizialer Komplex K auf eindeutige Weise einen abstrakten simplizialen Komplex $\mathfrak{K}$. Da wir im weiteren Verlauf dieses Paragraphen hauptsächlich mit geometrischen simplizialen Komplexen arbeiten werden, bezeichnen wir die Ketten, Zykeln und Rändern von $\mathfrak{K}$ auch als Ketten, Zykeln und Ränder von K und setzen $H_q(K) = H_q(\mathfrak{K})$. Da sich der Leser hoffentlich schon bisher alle Ketten als Aneinanderreihungen von geometrischen Simplizes vorgestellt hat, sollte dies zu keiner Verwirrung führen.

Definition: Ein (geometrischer) simplizialer Komplex L heißt Unterteilung des Komplexes K, wenn $|L| = |K|$ ist und wenn es für jedes Simplex t aus L ein Simplex s aus K gibt, so daß $t \subset s$ ist.

Man beachte, daß die beiden Simplizes s und t nicht die gleiche Dimension haben müssen: Wenn wir das verlangen wollten, hätte L genau die gleichen Ecken wie K, und es wäre $L = K$.

Es gibt eine kanonische Möglichkeit zur Unterteilung eines geometrischen simplizialen Komplexes, die *baryzentrische Unterteilung*. Erinnern wir uns daran, daß der *Schwerpunkt* eines Simplex s mit Ecken $P_0, \ldots, P_q$ der Punkt

$$b(s) = \frac{1}{q+1} \sum_{i=0}^{q} P_i$$

ist und definieren wir

Definition: Die *erste baryzentrische Unterteilung* eines geometrischen simplizialen Komplexes K ist jener Komplex K', dessen Ecken die Schwerpunkte der sämtlichen Simplizes von K sind und dessen Simplizes aufgespannt werden von Punkten $b(s_0), \ldots, b(s_q)$, wobei s_0 ein beliebiges Simplex von K ist und s_{i+1} für jedes i ein echtes Randsimplex (nicht notwendigerweise in Kodimension eins) von s_i ist.

Da jedes Nullsimplex sein eigener Schwerpunkt ist, kommen insbesondere auch alle Ecken von K unter den Ecken von K' vor. Die q-Simplizes $\langle b(s_0), \ldots, b(s_q) \rangle$ mit $q > 0$ sind echte Teilmengen von s_0, da sie höchstens *eine* Ecke von s_0 enthalten können. $b(s_0), \ldots, b(s_q)$ sind in allgemeiner Lage, denn sie sind konvexe Linearkombinationen der Ecken von s_0, und jedes $b(s_i)$ hängt von mindestens einer Ecke mehr ab als die folgenden $b(s_j)$. Eine nichttriviale lineare Beziehung zwischen den $b(s_i)$ wäre daher automatisch auch eine nichttriviale lineare Beziehung zwischen den Ecken von s_0, und die kann es nicht geben. Also ist K' ein Komplex und somit ein Unterkomplex von K.

Völlig analog können wir für jedes $n \in \mathbb{N}$ die *n-te baryzentrische Unterteilung* $K^{(n)}$ von K rekursiv definieren durch $K^{(1)} = K'$ und

$K^{(n+1)} = K^{(n)\prime}$. Der wesentliche Punkt des Beweises, daß die Homologie eines Komplexes nur von seiner topologischen Realisierung abhängt, besteht darin, zu zeigen, daß die sämtlichen $K^{(n)}$ gleiche Homologie haben.

Wir zeigen dies zunächst in einem ganz einfachen, aber wichtigen Fall: für die baryzentrische Unterteilung eines einzelnen Simplex.

2. Schritt: Die Homologie von Kegeln

Der Komplex K bestehe aus einem Simplex s und dessen sämtlichen Randsimplizes, und K' sei seine erste baryzentrische Unterteilung. K' hat einen Unterkomplex L, der aus jenen Simplizes von K' besteht, die auf dem Rand von s liegen. Jedes Simplex t von K', das nicht zu L gehört, hat den Schwerpunkt $b(s)$ von s als Ecke, denn das ist die einzige Ecke von K', die nicht in L liegt. Zu einem solchen Simplex t gibt es daher ein Simplex $u = \langle P_1, \ldots, P_q \rangle$ aus L, so daß sich t in der Form

$$t = b(s) \cdot u \underset{\text{def}}{=} \langle b(s), P_1, \ldots, P_q \rangle$$

schreiben läßt. Diese Situation wollen wir formalisieren und verallgemeinern:

Definition: K sei ein geometrischer simplizialer Komplex in $\mathbb{R}^N$ und P sei ein Punkt aus $\mathbb{R}^N$, der entweder eine Ecke von K ist oder aber in keinem von einem Simplex aus K aufgespannten affinen Unterraum von $\mathbb{R}^N$ liegt. Der *Kegel* über K mit Spitze P ist jener geometrische simpliziale Komplex $L = P \cdot K$ mit $L_0 = K_0 \cup \{P\}$, dessen q-Simplizes für $q > 0$ einerseits die sämtlichen q-Simplizes von K sind, andererseits alle Simplizes der Form $P \cdot s$, wobei s ein $(q-1)$-Simplex von K ist, das P nicht als Ecke hat.

Im obigen Beispiel ist also K' der Kegel über L mit Spitze $b(s)$.

Im folgenden schreiben wir für eine beliebige Kette $c = \sum \lambda_i s_i$

$$P \cdot c \underset{\text{def}}{=} \sum \lambda_i \; P \cdot s_i \; ,$$

wobei $P \cdot s_i = 0$ sein soll, wenn P eine Ecke von s_i ist.

Wir beweisen nun allgemein, daß die Homologie eines Kegels trivial ist. Da die nullte Homologiegruppe eines simplizialen Komplexes immer zumindest die Äquivalenzklasse eines einzelnen Punktes enthält, kann „Trivialität“ hier natürlich nicht bedeuten, daß alle Homologiegruppen verschwinden, sondern

Definition: Wir sagen, ein Komplex K sei *azyklisch* oder habe *triviale Homologie,* wenn $H_q(K) = 0$ ist für $q > 0$ und $H_0(K) \cong \mathbb{Z}$.

Lemma: Ein Kegel $L = P \cdot K$ hat triviale Homologie.

Beweis: Da jede Ecke von L durch eine Kante mit P verbunden werden kann, ist natürlich $H_0(L) \cong \mathbb{Z}$. Für $q > 0$ hat jeder q-Zyklus z von L eine eindeutige Darstellung $z = P \cdot c + d$, wobei $c \in C_{q-1}(K)$ und $d \in C_q(K)$. Daher ist

$$0 = \partial_q z = c - P \cdot \partial_{q-1} c + \partial_q d.$$

Dabei ist $c + \partial d$ eine Kette aus K, keines ihrer Simplizes hat also den Punkt P als Ecke, wohingegen in der Kette $P \cdot \partial_{q-1} c$ jedes Simplex P als Ecke hat. Daher haben die beiden Ketten kein Simplex gemeinsam, müssen also beide verschwinden:

$$c + \partial_q d = 0 \quad \text{und} \quad P \cdot \partial_{q-1} c = 0.$$

Nach der ersten dieser Gleichungen ist

$$z = P \cdot c + d = d - P \cdot \partial_q d = \partial_{q+1}(P \cdot d)$$

ein Rand, wie behauptet.

■

Korollar: Der Komplex, der aus einem Simplex s und dessen sämtlichen Seitensimplizes besteht, hat triviale Homologie.

Das ist klar, denn für jede Ecke P von s läßt sich s als Kegel $P \cdot s'$ schreiben, wobei s' das P gegenüberliegende Simplex ist, oder auch einfach als $P \cdot \partial s$.

■

3. Schritt: Simpliziale Abbildungen und Kettenabbildungen

Zum Vergleich der Homologie zweier simplizialer Komplexe müssen wir Abbildungen zwischen diesen Komplexen betrachten und diese auf die zugehörigen Kettenkomplexe und Homologiegruppen fortsetzen. Der natürliche Abbildungsbegriff für Abbildungen zwischen Komplexen ist folgender:

Definition: Eine *simpliziale Abbildung* zwischen zwei simplizialen Komplexen K und L ist eine Abbildung $\varphi\colon K_0 \to L_0$ von der Menge K_0 der Ecken von K in die Menge L_0 der Ecken von L derart, daß die Ecken eines jeden Simplex s von K auf (nicht notwendigerweise verschiedene) Ecken eines Simplex von L abgebildet werden. Das kleinste Simplex aus L mit dieser Eigenschaft bezeichnen wir mit $\varphi(s)$.

Eine simpliziale Abbildung $\varphi\colon K_0 \to L_0$ definiert eine Abbildung $\hat{\varphi}\colon |K| \to |L|$ durch lineare Fortsetzung in das Innere der Simplizes: Jeder Punkt $P \in |K|$ ist innerer Punkt eines Simplex von K und läßt sich daher als konvexe Linearkombination $\sum \lambda_i P_i$ von dessen Ecken $P_0, \ldots, P_r$ schreiben; wir setzen

$$\hat{\varphi}(P) = \sum_{i=0}^{r} \lambda_i \varphi(P_i).$$

φ induziert auch eine Abbildung zwischen den Kettenkomplexen $\mathcal{C}(K)$ und $\mathcal{C}(L)$, aber dazu müssen wir zunächst definieren, was eine solche Abbildung sein soll:

Definition: Eine *Kettenabbildung* $\varphi\colon \mathcal{C} \to \mathcal{D}$ zwischen zwei Kettenkomplexen $\mathcal{C} = \{C_q, \partial_q\}$ und $\mathcal{D} = \{D_q, \partial'_q\}$ ist ein System von Abbildungen $\varphi_q\colon C_q \to D_q$ derart, daß $\varphi_{q-1} \circ \partial_q = \partial'_q \circ \varphi_q$ ist.

Versuchen wir also, der simplizialen Abbildung $\varphi\colon K_0 \to L_0$ eine Kettenabbildung $\tilde{\varphi}\colon \mathcal{C}(K) \to \mathcal{C}(L)$ zuordnen. Dabei stoßen wir auf das Problem, daß eine simpliziale Abbildung ein Simplex von K auf ein Simplex von L mit niedrigerer Dimension abbilden kann. Wie sich zeigen wird, verlieren wir nichts, wenn wir solche Simplizes einfach ignorieren; wir definieren also $\tilde{\varphi}(s)$ für ein q-Simplex s mit Ecken $P_0, \ldots, P_q$ aus K

als Null, falls die Punkte $\varphi(P_0), \ldots, \varphi(P_q)$ nicht paarweise verschieden sind, und sonst als das von ihnen aufgespannte q-Simplex von L.Durch lineare Fortsetzung auf die Kettengruppe erhalten wir daraus einen Homomorphismus

$$\tilde{\varphi}_q \colon C_q(K) \to C_q(L); \quad \sum \lambda_i s_i \mapsto \sum \lambda_i \tilde{\varphi}_q(s_i) \ .$$

Beim Nachweis, daß dies in der Tat eine Kettenabbildung ist, können wir uns auf ein einzelnes Simplex s aus K zu beschränken; dessen Ecken seien $P_0, \ldots, P_q$. Falls alle $\varphi(P_i)$ verschieden sind, gibt es keinerlei Probleme. Wenn zwei Punkte $\varphi(P_i)$ und $\varphi(P_j)$ übereinstimmen, ist $\tilde{\varphi}_q(s) = 0$, also auch $\partial'_q\big(\tilde{\varphi}_q(s)\big) = 0$; wir müssen zeigen, daß auch $\tilde{\varphi}_{q-1}\big(\partial_q(s)\big)$ verschwindet. Für alle Randsimplizes, in denen sowohl P_i als auch P_j vorkommen, ist das klar, $\tilde{\varphi}_{q-1}\big(\partial_q(s)\big)$ enthält also nur die beiden Randsimplizes, in denen $\varphi(P_i)$ bzw. $\varphi(P_j)$ gestrichen wird. Diese beiden Simplizes sind gleich, können aber verschieden orientiert sein: Ist j der größere der beiden Indizes, so tritt das Simplex mit gestrichenem $\varphi(P_i)$ mit Vorfaktor $(-1)^i$ auf, das andere mit $(-1)^j$. Letzteres hat aber eventuell eine andere Orientierung, denn um die Eckennummerierungen der beiden Simplizes ineinander überzuführen, müssen wir die zyklische Permutation $(i+1 \ldots j)$ anwenden. Als Zyklus der Länge $j-i$ hat sie das Vorzeichen $(-1)^{j-i-1}$; wenn wir das zweite Simplex mit der Eckenreihenfolge des ersten hinschreiben, ist also der effektive Vorfaktor $(-1)^{2j-i-1} = -(-1)^i$. Somit sind die Vorfaktoren entgegengesetzt gleich und die Summe verschwindet.

Als nächstes brauchen wir Abbildungen zwischen den Homologiegruppen; die gibt es für jede Kettenabbildung nach dem folgenden Lemma, dessen Beweis auch die Definition enthält:

Lemma: Eine Kettenabbildung $\varphi \colon \mathcal{C} \to \mathcal{D}$ induziert Homomorphismen $\varphi_{*q} \colon H_q(\mathcal{C}) \to H_q(\mathcal{D})$ für alle $q \geq 0$. Ist $\psi \colon \mathcal{D} \to \mathcal{E}$ eine weitere Kettenabbildung, so ist auch $\psi \circ \varphi \colon \mathcal{C} \to \mathcal{E}$, das System der Abbildungen $\psi_q \circ \varphi_q$, eine Kettenabbildung, und $(\psi \circ \varphi)_{*q} = \psi_{*q} \circ \varphi_{*q}$.

Beweis: Jedes $h \in H_q(\mathcal{C})$ hat einen Repräsentanten $z \in Z_q(\mathcal{C})$; wir wollen $\varphi_{*q}(h)$ als die Homologieklasse von $\varphi_q(z)$ definieren. Dies ist

wohldefiniert, denn sind ∂_q die Randabbildungen von $\mathcal{C}$ und ∂'_q die von $\mathcal{D}$, so ist zunächst

$$\partial'_q\big(\varphi_q(z)\big) = \partial'_q \circ \varphi_q(z) = \varphi_{q-1} \circ \partial_q(z) = \varphi_{q-1}(0) = 0\,,$$

so daß $\varphi_q(z)$ in $Z_q(\mathcal{D})$ liegt, und wenn wir anstelle von z einen anderen Repräsentanten z' nehmen, so unterscheiden sich z und z' durch einen Rand, d.h. $z' = z + \partial_{q+1}(c)$ für eine Kette $c \in C_{q+1}$ und

$$\varphi_q(z') = \varphi_q(z) + \varphi_q\big(\partial_{q+1}(c)\big) = \varphi_q(z) + \partial'_{q+1}\big(\varphi_{q+1}(c)\big),$$

$\varphi_q(z)$ und $\varphi_q(z')$ unterscheiden sich also nur um einen Rand und definieren daher die gleiche Restklasse in $H_q(\mathcal{D})$. Die restlichen Behauptungen sind trivial. ■

Da somit jede Kettenabbildung Homomorphismen zwischen den Homologiegruppen induziert, induziert auch jede simpliziale Abbildung φ Homomorphismen $\tilde{\varphi}_{*q}$, die wir der Einfachheit halber als φ_{*q} schreiben. Natürlich gilt auch hier wieder für zusammengesetzte Abbildungen $\psi \circ \varphi$, daß $(\psi \circ \varphi)_{*q} = \psi_{*q} \circ \varphi_{*q}$ ist.

4. Schritt: Benachbarte Abbildungen und Kettenhomotopien

Hier geht es um hinreichende Bedingungen dafür, daß zwei Abbildungen die gleichen Homomorphismen auf der Homologie induzieren.

Definition: a) Zwei simpliziale Abbildungen $\varphi, \psi\colon K_0 \to L_0$ zwischen den simplizialen Komplexen K und L heißen *benachbart,* wenn es zu jedem q-Simplex s von K ein Simplex t von L gibt, so daß $\varphi(s)$ und $\psi(s)$ beide Randsimplizes von t sind.

b) φ, ψ heißen *ähnlich,* wenn es zu jedem q-Simplex s von K einen Unterkomplex $L(s)$ von L gibt, so daß gilt:

1. $L(s)$ hat triviale Homologie.
2. Für jedes Simplex s von K liegen sowohl $\varphi(s)$ als auch $\psi(s)$ in $L(s)$.
3. Für jedes Randsimplex t eines Simplex s aus K ist $L(t)$ ein Unterkomplex von $L(s)$.

c) Zwei Kettenabbildungen $\varphi, \psi: \mathcal{C} \to \mathcal{D}$ zwischen den Kettenkomplexen $\mathcal{C} = \{C_q, \partial_q\}$ und $\mathcal{D} = \{D_q, \partial'_q\}$ heißen *homotop,* wenn es für jedes $q \geq 0$ einen Homomorphismus $\gamma_q: C_q \to D_{q+1}$ gibt, so daß gilt:

$$\partial'_{q+1} \circ \gamma_q + \gamma_{q-1} \circ \partial_q = \varphi_q - \psi_q.$$

Damit dies auch für $q = 0$ sinnvoll ist, setzen wir formal $\gamma_{-1} \circ \partial_0 = 0$.

Natürlich sind benachbarte Abbildungen ähnlich, wir müssen nur $L(s)$ gleich dem kleinsten Simplex setzen, das $\varphi(s)$ und $\psi(s)$ als Randsimplizes hat: Wie wir aus dem zweiten Schritt wissen, hat der Komplex aus einem Simplex und dessen sämtlichen Randsimplizes triviale Homologie. Wir wollen nun sehen, daß ähnliche simpliziale Abbildungen homotope Abbildungen auf den zugehörigen Kettenkomplexen induzieren und daß homotope Abbildungen zwischen Kettenkomplexen die gleichen Homomorphismen zwischen den Homologiegruppen induzieren.

Lemma: Zwei ähnliche Abbildungen φ, ψ zwischen den simplizialen Komplexen K und L induzieren homotope Kettenabbildungen $\tilde{\varphi}, \tilde{\psi}: \mathcal{C}(K) \to \mathcal{C}(L)$.

Beweis: Wir müssen für jedes q-Simplex s von K eine $(q + 1)$-Kette $\gamma_q(s)$ aus L angeben, so daß

$$\partial'_{q+1} \circ \gamma_q + \gamma_{q-1} \circ \partial_q = \varphi_q - \psi_q$$

ist. Die entsprechenden Abbildungen γ_q werden induktiv definiert; wegen ihrer Linearität genügt es, $\gamma_q(s)$ für q-Simplizes s anzugeben. $\gamma_q(s)$ wird dabei immer sogar eine Kette aus dem Unterkomplex $L(s)$ sein.

Für $q = 0$, eine Ecke s von K also, sind $\varphi(s)$ und $\psi(s)$ Ecken des azyklischen Unterkomplexes $L(s)$ von L. Da dessen nullte Homologiegruppe gleich $\mathbb{Z}$ ist, muß die 0-Kette $\varphi(s) - \psi(s)$ Rand einer 1-Kette aus $L(s)$ sein. Eine solche 1-Kette nehmen wir als $\gamma_0(s)$. Nach Konstruktion ist

$$\partial'_1 \circ \gamma_0(s) = \varphi_0(s) - \psi_0(s),$$

wie gewünscht.

Seien nun $\gamma_0, \ldots, \gamma_{q-1}$ bereits so definiert, daß die geforderte Gleichung für alle p-Simplizes mit $p < q$ gilt, und sei s ein q-Simplex aus K. Für

jedes $(q-1)$-Simplex t aus dem Rand von s ist $\gamma_{q-1}(t)$ definiert und liegt in $L(t) \subset L(s)$. Daher liegt auch $\gamma_{q-1}\big(\partial_q(s)\big)$ in $L(s)$, genauso wie $\varphi(s)$ und $\psi(s)$. Folglich ist

$$c = \varphi_q(s) - \psi_q(s) - \gamma_{q-1}\big(\partial_q(s)\big)$$

eine Kette in $L(s)$, sogar ein Zyklus, denn

$$\begin{aligned}\partial'_q(c) &= \partial'_q \circ \varphi_q(s) - \partial'_q \circ \psi_q(s) - \partial'_q \circ \gamma_{q-1}\big(\partial_q(s)\big)\\ &= \varphi_{q-1} \circ \partial_q(s) - \psi_{q-1} \circ \partial_q(s) - \partial'_q \circ \gamma_{q-1}\big(\partial_q(s)\big)\\ &= \big(\varphi_{q-1} - \psi_{q-1} - \partial'_q \circ \gamma_{q-1}\big)\big(\partial_q(s)\big),\end{aligned}$$

und das ist nach Induktionsvoraussetzung gleich

$$\gamma_{q-2} \circ \partial_{q-1} \circ \partial_q(s) = 0\ .$$

Da der q-Zyklus c im azyklischen Komplex $L(s)$ liegt, ist er Rand einer $(q+1)$-Kette aus $L(s)$. Wir setzen $\gamma_q(s)$ gleich einer solchen Kette; dann ist die Bedingung an γ_q nach Konstruktion erfüllt.

■

Das zweite Lemma ist noch einfacher:

Lemma: Für homotope Kettenabbildungen $\varphi, \psi\colon \mathcal{C} \to \mathcal{D}$ stimmen φ_{*q} und ψ_{*q} für alle q überein.

Beweis: Wir müssen zeigen, daß $\varphi_q(z) - \psi_q(z)$ für jeden Zyklus z aus $Z_q(\mathcal{C})$ in der Rändergruppe $B_q(\mathcal{D})$ liegt. Das ist aber klar, denn

$$\varphi_q(z) - \psi_q(z) = \partial'_{q+1}\big(\gamma_q(z)\big) + \gamma_{q-1}\big(\partial_q(z)\big) = \partial'_{q+1}\big(\gamma_q(z)\big),$$

denn $\partial_q(z)$ verschwindet für einen Zyklus z.

■

Insbesondere induzieren also ähnliche Abbildungen die gleichen Homomorphismen auf der Homologie. Um im nächsten Schritt zu zeigen, daß die Homologie eines Komplexes gleich der seiner baryzentrischen Unterteilungen ist, werden wir dies auf eine spezielle Abbildung anwenden:

5. Schritt: Die SPERNER-Abbildung

K sei ein simplizialer Komplex, K' seine erste baryzentrische Unterteilung. Als SPERNER-Abbildung bezeichnen wir „die“ simpliziale Abbildung $\omega\colon (K')_0 \to K_0$, die jeder Ecke $b(s)$ von K' irgendeine der Ecken von s zuordnet, sowie auch die dadurch auf den Kettengruppen induzierte Abbildung $\omega\colon \mathcal{C}(K') \to \mathcal{C}(K)$.

EMANUEL SPERNER (1905–1980) wurde im damaligen Oberschlesien geboren und ging auch dort in Neisse (heute Nysa) zur Schule. Nach seinem Abitur 1925 studierte er zunächst in Freiburg, danach in Hamburg, wo er sowohl promovierte als auch habilitierte. Er war nacheinander Professor an den Universitäten Königsberg, Straßburg, Bonn und Hamburg; dort war er von 1963 bis 1965 Rektor. Bekannt ist er vor allem durch das eng mit der SPERNER-Abbildung verknüpfte Lemma von SPERNER *(siehe Kapitel 10) sowie einem heute als* SPERNER-*Theorie* bezeichneten Teilgebiet der diskreten Mathematik, in dem es um spezielle kommbinatorische Fragen geht.

Wir wollen in diesem Schritt einsehen, daß die auf der Homologie induzierten Abbildungen

$$\omega_{*q}\colon H_q(K') \to H_q(K)$$

Isomorphismen sind. Zu diesem Zweck konstruieren wir eine Kettenabbildung $\omega^-\colon \mathcal{C}(K) \to \mathcal{C}(K')$ derart, daß ω_{*q} und ω^-_{*q} für alle q zueinander invers sind. Die linearen Abbildungen $\omega^-_q\colon C_q(K) \to C_q(K')$ werden induktiv konstruiert:

$\omega^-_0\colon C_0(K) \to C_0(K')$ ist diejenige Abbildung, die jede Ecke auf sich selbst abbildet. Für $q > 0$ soll ω^-_q jedes q-Simplex s von K abbilden auf die Summe aller q-Simplizes von K', die in s liegen. Formal erklären wir das folgendermaßen: Für eine q-Kette $c = \sum \lambda_i s_i$ aus einem simplizialen Komplex und eine Ecke P, die in keinem der s_i vorkommt, definieren wir ein $(q+1)$-Kette

$$P \cdot c = \sum \lambda_i \; P \cdot s_i,$$

wobei $P \cdot s$ für ein orientiertes q-Simplex s mit Ecken $P_0, \ldots P_q$ bekanntlich das orientierte $(q+1)$-Simplex mit Ecken $P, P_0, \ldots, P_q$ be-

zeichnet, also den Kegel über s mit Spitze P. Mit dieser Bezeichnung soll dann gelten

$$\omega_q^-(s) \underset{\text{def}}{=} b(s) \cdot \omega_{q-1}^-\big(\partial_q(s)\big).$$

Man überzeugt sich leicht durch Induktion, daß dies eine Kettenabbildung definiert, denn bezeichnet ∂ die Randabbildungen von K und ∂' die von K', so ist

$$\begin{aligned}
\partial_q'\big(\omega_q^-(s)\big) &= \partial_q'\Big(b(s) \cdot \omega_{q-1}^-\big(\partial_q(s)\big)\Big) \\
&= \omega_{q-1}^-\big(\partial_q(s)\big) - b(s) \cdot \partial_{q-1}'\big(\omega_{q-1}^-(\partial_q(s))\big) \\
&= \omega_{q-1}^-\big(\partial_q(s)\big) - b(s) \cdot \omega_{q-2}^-\big(\partial_{q-1} \circ \partial_q(s)\big) \\
&= \omega_{q-1}^-\big(\partial_q(s)\big).
\end{aligned}$$

Ebenfalls durch Induktion folgt, daß

$$\omega_q \circ \omega_q^- = \mathrm{id}_{C_q(K)}$$

ist, denn für $q = 0$ ist dies klar, und für jedes q-Simplex s mit $q > 0$ ist

$$\begin{aligned}
\omega_q \circ \omega_q^-(s) &= \omega_q\big(b(s) \cdot \omega_{q-1}^-(\partial_q s)\big) = \omega\big(b(s)\big) \cdot \big(\omega_{q-1} \circ \omega_{q-1}^-\big)(\partial_q s) \\
&= \omega\big(b(s)\big) \cdot \partial_q(s) = s,
\end{aligned}$$

weil $\omega\big(b(s)\big)$ nach Konstruktion eine Ecke von s ist und $\partial_q(s)$ insbesondere auch das dieser Ecke gegenüberliegende Randsimplex enthält.

Die umgekehrte Gleichung

$$\omega_q^- \circ \omega_q = \mathrm{id}_{C_q(K')}$$

ist offensichtlich falsch, denn natürlich kann ω_q nicht injektiv sein. Für unsere Zwecke genügt es aber, wenn wir anstelle der Gleichheit nur die Ähnlichkeit der beiden Abbildungen zeigen, und die ist fast klar: Ein Simplex s' aus K' hat die Form

$$s' = \big\langle b(s), b(s_1), \ldots, b(s_q) \big\rangle\,,$$

wobei die s_i Seitensimplizes von s sind; wir definieren $K'(s')$ als die Menge aller Simplizes aus K', die in s liegen. Wie wir im zweiten Schritt gesehen haben, hat dieser Komplex triviale Homologie, außerdem liegen

sowohl $\omega_q \circ \omega_q^-(s')$ als auch $\mathrm{id}_{C_q(K)}(s')$ in $K'(s')$, also sind die beiden Abbildungen ähnlich. Es folgt, daß sowohl

$$\omega_{*q}^- \circ \omega_{*q} \colon H_q(K') \to H_q(K')$$

als auch

$$\omega_{*q} \circ \omega_{*q}^- \colon H_q(K) \to H_q(K)$$

identische Abbildungen sind, also ist

$$\omega_{*q} \colon H_q(K) \to H_q(K')$$

ein Isomorphismus.Durch Iteration von SPERNER-Abbildungen folgt sofort

Satz: Für jeden simplizialen Komplex K und für jede baryzentrische Unterteilung $K^{(n)}$ von K ist $H_q(K) \cong H_q\big(K^{(n)}\big)$.

■

6. Schritt: Verfeinerungen endlicher simplizialer Komplexe

Hier kommen wir nun zu der angekündigten Verallgemeinerung des Unterteilungsbegriffs. Dazu brauchen wir zunächst zwei neue Begriffe aus der Theorie der simplizialen Komplexe:

Definition: K sei ein simplizialer Komplex und P sei eine Ecke von K. Der *Stern* von P ist die Vereinigung $\mathrm{st}_K(P)$ der Inneren aller jener Simplizes aus K, die P als Ecke haben. (Dazu gehört auch das 0-Simplex P.) Der *Link* von P ist die Vereinigung $\mathrm{lk}_K(P)$ aller Simplizes aus $\overline{\mathrm{st}_K(P)}$, die P nicht als Ecke enthalten.

Offensichtlich ist der abgeschlossene Stern $\overline{\mathrm{st}_K(P)}$ von P die geometrische Realisierung eines Unterkomplexes $\mathrm{St}_K(P)$, und der Link von P ist geometrische Realisierung eines Unterkomplexes $\mathrm{Lk}_K(P)$ von $\mathrm{St}_K(P)$. Weiter ist $\mathrm{St}_K(P) = P \cdot \mathrm{Lk}_K(P)$ ein Kegel; insbesondere hat $\mathrm{St}_K(P)$ also nach dem Lemma aus dem zweiten Schritt triviale Homologie.

Wenn der zugrundeliegende Komplex klar ist, werden wir in Zukunft den Index K weglassen.

Der Stern von P ist nach Konstruktion eine offene Menge, und die Sterne der sämtlichen Ecken von K bilden eine Überdeckung von $|K|$. Verfeinerungen von Komplexen sollen nun so definiert werden, daß sie Verfeinerungen dieser Überdeckung entsprechen. Einen trivialen, aber nützlichen Zusammenhang zwischen den Eigenschaften der Überdekkungsmengen und der simplizialen Struktur des Komplexes liefert das folgende

Lemma: Die Ecken $P_0, \ldots, P_q$ eines simplizialen Komplexes K gehören genau dann zu einem gemeinsamen Simplex von K, wenn ihre Sterne nichtleeren Durchschnitt haben.

Beweis: Falls $P_0, \ldots, P_q$ zu einem gemeinsamen Simplex gehören, liegt dessen Inneres natürlich im Stern eines jeden P_i, also auch im Durchschnitt dieser Sterne. Enthält umgekehrt $\cap_{i=0}^{q}\mathrm{st}(P_i)$ einen Punkt Q, so muß dieser für jedes i innerer Punkt eines Simplex mit Ecke P_i sein. Da jeder Punkt aber nur im Innern eines einzigen Simplex aus K liegen kann, gehört Q zu einem Simplex, das alle P_i als Ecken hat. ∎

Damit können wir nun endlich die Verfeinerungen eines Komplexes definieren:

Definition: Ein (geometrischer) simplizialer Komplex L heißt *Verfeinerung* eines Komplexes K, in Zeichen $L < K$, wenn $|L| = |K|$ ist und wenn es zu jeder Ecke P von L eine Ecke Q von K gibt, so daß $\mathrm{st}_L(P)$ in $\mathrm{st}_K(Q)$ liegt.

Beispielsweise ist jede baryzentrische Unterteilung eine Verfeinerung; hier erhält man Q aus P durch eine SPERNER-Abbildung.

In diesem Abschnitt soll gezeigt werden, daß je zwei (endliche) simpliziale Komplexe, die die gleiche Teilmenge von $\mathbb{R}^N$ triangulieren, eine gemeinsame Verfeinerung haben, und daß Verfeinerungen die gleichen Homologiegruppen haben wie der ursprüngliche Komplex.

Definition: Eine Teilmenge $X \subset \mathbb{R}^N$ heißt *Polyeder,* wenn sie Vereinigung der Simplizes eines endlichen geometrischen simplizialen Komplexes ist.

Wir wollen uns überlegen, daß die Homologie eines Polyeders nur von diesem selbst abhängt, nicht aber von einer Triangulierung. Dazu brauchen wir zunächst einige „metrische“ Vorbereitungen.

Definition: Die *Maschenweite* $\mu(K)$ eines endlichen simplizialen Komplexes K im $\mathbb{R}^N$ ist das Supremum der Abstände zwischen zwei Punkten eines Simplex von K bezüglich der Metrik des $\mathbb{R}^N$.

Lemma: $\mu(K)$ ist das Maximum der Abstände zwischen zwei Ecken eines Simplex aus K.

Beweis: a sei der größte Abstand zwischen zwei Ecken. Für einen beliebigen Punkt Q aus dem Simplex sei P_i die am weitesten entfernte Ecke; dann enthält die Kugel mit Radius $d(Q, P_i)$ alle Ecken, also das ganze Simplex, und somit ist

$$d(Q, R) \leq d(Q, P_i)$$

für alle Simplexpunkte R. Ist P_j diejenige Ecke, die am weitesten von P_i entfernt ist, so gilt aus dem gleichen Grund

$$d(Q, P_i) \leq d(P_j, P_i),$$

also

$$d(Q, R) \leq d(Q, P_i) \leq d(P_j, P_i) \leq a,$$

wie behauptet.

■

Bei der Konstruktion einer gemeinsamen Verfeinerung werden natürlich die baryzentrischen Unterteilungen eine große Rolle spielen; insbesondere müssen wir wissen, daß ihre Maschenweite beliebig klein werden kann. Dies folgt aus

Lemma: K sei ein geometrischer simplizialer Komplex in $\mathbb{R}^N$, dessen Simplizes alle höchstens die Dimension r haben. Dann ist

$$\mu(K') \leq \frac{r}{r+1}\mu(K).$$

Beweis: Die Maschenweite von K' ist der größte Abstand zwischen zwei Ecken eines Simplex aus K'; seien also zwei solche Ecken gegeben. Nach Konstruktion von K' gibt es dazu ein Simplex

$$s = \langle P_0, \ldots, P_q \rangle$$

von K und ein Randsimplex t von s, o.B.d.A. das von P_0 bis P_p aufgespannte, so daß die gegebenen Ecken die Schwerpunkte $b(s)$ und $b(t)$ von s und t sind. Der Vektor zwischen den beiden Ecken ist dann

$$\begin{aligned} b(t) - b(s) &= \frac{1}{p+1} \sum_{i=0}^{p} P_i - \frac{1}{q+1} \sum_{i=0}^{q} P_i \\ &= \left(\frac{1}{p+1} - \frac{1}{q+1} \right) \sum_{i=0}^{p} P_i - \frac{1}{q+1} \sum_{i=p+1}^{q} P_i \\ &= \frac{q-p}{(p+1)(q+1)} \sum_{i=0}^{p} P_i - \frac{1}{q+1} \sum_{i=p+1}^{q} P_i \\ &= \frac{q-p}{q+1} \left(\frac{1}{p+1} \sum_{i=0}^{p} P_i - \frac{1}{q-p} \sum_{i=p+1}^{q} P_i \right). \end{aligned}$$

In der Klammer steht der Verbindungsvektor zwischen einem der Punkte des Simplex $\langle P_0, \ldots, P_p \rangle$ und einem der Punkte des Simplex $\langle P_{p+1}, \ldots, P_q \rangle$, von zwei Punkten aus s also, und deren Abstand kann nicht größer sein als $\mu(K)$. Also ist

$$d\big(b(t), b(s)\big) \leq \frac{q-p}{q+1} \mu(K) \leq \frac{q}{q+1} \mu(K) \leq \frac{r}{r+1} \mu(K),$$

wie behauptet.

■

Damit können wir zeigen

Lemma: Für zwei endliche simpliziale Komplexe K und L, für die $|K| = |L|$ ist, gibt es stets eine natürliche Zahl n, so daß die n-te baryzentrische Unterteilung $K^{(n)}$ von K eine Verfeinerung von L ist. Insbesondere haben K und L eine gemeinsame Verfeinerung M.

Beweis: Die Sterne der Ecken P von L bilden eine offene Überdeckung von $|L| = |K|$. Da $|L|$ wegen der Endlichkeit von L eine kompakte Teilmenge eines $\mathbb{R}^N$ ist, gibt es zu dieser Überdeckung eine LEBESGUE-Zahl δ derart, daß jede offene Teilmenge von $\mathbb{R}^N$ mit einem Durchmesser kleiner δ in einer Überdeckungsmenge liegt. Aus dem gerade gezeigten Lemma folgt, daß wir eine Zahl n finden können, so daß $\mu(K^{(n)}) < \delta/2$ ist. Da der Durchmesser eines Sterns höchstens gleich der doppelten Maschenweite sein kann, ist $K^{(n)}$ eine Verfeinerung von L und erst recht natürlich von K.

■

Satz: Für zwei endliche geometrische simpliziale Komplexe K und L mit $|K| = |L|$ sind $H_q(K)$ und $H_q(L)$ isomorph für alle q.

Beweis: Da L und K eine gemeinsame Verfeinerung haben, können wir uns auf den Fall $L < K$ beschränken. Dann läßt sich leicht eine simpliziale Abbildung $L \to K$ angeben: Man ordnet einfach jeder Ecke $P \in L_0$ eine Ecke $Q \in K_0$ zu derart, daß $\mathrm{st}_L(P) \subset \mathrm{st}_K(Q)$ ist. Diese Abbildung definiert Homomorphismen

$$\tau(L, K)_q \colon H_q(L) \to H_q(K)$$

für alle q, und diese hängen nur von L, K und q ab, nicht von der simplizialen Abbildung $L \to K$, denn je zwei solche Abbildungen sind ähnlich, da $\mathrm{st}(Q)$ und $\mathrm{st}(Q')$ nur dann einen nichtleeren Durchschnitt haben können, wenn QQ' eine Kante von K ist. Sei nun n so gewählt, daß $K^{(n)}$ eine Verfeinerung von L ist. Dann ist $K^{(n)} < L < K$, und wegen der Eindeutigkeit von τ ist einerseits $\tau(K^{(n)}, K)_q = \tau(L, K)_q \circ \tau(K^{(n)}, L)_q$, andererseits aber wird $\tau(K^{(n)}, K)_q$ von einer Iteration von SPERNER-Abbildungen induziert, ist also ein Isomorphismus. Damit ist auch $\tau(L, K)_q$ ein Isomorphismus, und der Satz ist bewiesen.

■

7. Schritt: Simpliziale Approximation von Abbildungen

Wir wissen nun, daß die Homologie eines Polyeders nicht von dessen Triangulierung abhängt, sondern nur vom Polyeder selbst. Wir wissen aber noch nicht, daß zueinander homöomorphe Polyeder wie etwa der Würfel und das Tetraeder beide die gleiche Homologie haben, und

wir können daher insbesondere noch nicht von „der“ Homologie der Vollkugel oder eines anderen krummlinig begrenzten Körpers sprechen. Diese letzte Lücke schließt der siebte Schritt, in dem wir stetigen Abbildungen zwischen Polyedern simpliziale Abbildungen zwischen deren Triangulierungen und damit Homomorphismen zwischen den Homologiegruppen zuordnen und zeigen, daß einem Homöomorphismus Isomorphismen der Homologiegruppen zugeordnet werden.

Definition: K und L seien simpliziale Komplexe. Eine simpliziale Abbildung $\varphi\colon K_0 \to L_0$ heißt *simpliziale Approximation* der stetigen Abbildung $f\colon |K| \to |L|$, wenn für jeden Punkt $x \in |K|$ die Bildpunkte $f(x)$ und $|\varphi|(x)$ in einem gemeinsamen Simplex von L liegen.

Lemma: Jede stetige Abbildung $f\colon |K| \to |L|$ zwischen zwei kompakten Polyedern hat eine simpliziale Approximation $\varphi\colon K^{(n)} \to L$ für hinreichend großes n.

Beweis: $\mathcal{U} = \{\mathrm{st}_L(P) \mid P \in L_0\}$ ist eine offene Überdeckung von $|L|$. Die Urbilder der Überdeckungsmengen unter f bilden eine offene Überdeckung der kompakten Teilmenge $|K|$ von $\mathbb{R}^N$; deren LEBESGUE-Zahl sei δ. Wie im vorigen Schritt gibt es dann ein n, so daß $\mu(K^{(n)}) < \delta/2$ ist, und wir können eine Abbildung $\varphi\colon K_0^{(n)} \to L_0$ definieren, indem wir $P \in K_0^{(n)}$ abbilden auf irgendein $Q \in L_0$ mit

$$\mathrm{st}_{K^{(n)}}(P) \subset f^{-1}\big(\mathrm{st}_L(Q)\big).$$

Dies ist eine simpliziale Abbildung, denn sind $P_0, \ldots, P_q$ Ecken eines Simplex aus $K^{(n)}$, so haben die Sterne $\mathrm{st}_{K^{(n)}}(P)$ nichtleeren Durchschnitt, und dieser Durchschnitt liegt im Durchschnitt der Urbilder $f^{-1}\big(\mathrm{st}_L(\varphi(P_i))\big)$, so daß auch dieser Durchschnitt und damit der Durchschnitt der $\mathrm{st}_L\big(\varphi(P_i)\big)$ nicht leer ist. Die $\varphi(P_i)$ sind also Ecken eines Simplex aus L.

Schließlich ist φ auch eine simpliziale Approximation von f, denn liegt $x \in |K|$ im von $P_0, \ldots, P_q$ aufgespannten Simplex, so liegt x im Durchschnitt der Sterne der P_i in $K^{(n)}$, also liegt $f(x)$ nach Konstruktion von φ im Durchschnitt der Sterne der $\varphi(P_i)$ in L, also im Simplex,

das von $\varphi(P_0), \ldots \varphi(P_q)$ aufgespannt wird, und genau dort liegt nach Konstruktion auch $|\varphi|(x)$.

■

Satz: Zu jeder stetigen Abbildung $f\colon X \to Y$ zwischen zwei Polyedern gibt es eindeutig bestimmte Homomorphismen

$$f_{*q}\colon H_q(X) \to H_q(Y)$$

für alle q. Ist $g\colon Y \to Z$ eine weitere stetige Abbildung nach einem Polyeder Z, so ist $(g \circ f)_{*q} = g_{*q} \circ f_{*q}$. Ist f ein Homöomorphismus, so sind alle f_{*q} Isomorphismen.

Beweis: K, L seien simpliziale Komplexe, die X und Y triangulieren, und $\varphi\colon K_0^{(n)} \to L_0$ sei eine simpliziale Approximation von f. Wir setzen $f_{*q} = \varphi_{*q}$ und müssen zeigen, daß diese Abbildungen nicht von K, L, n und φ abhängen.

Die Unabhängigkeit von φ bei gegebenem K, L und n ist klar: Zwar ist die Definition von φ im obigen Beweis alles andere als eindeutig, aber zwei verschiedene Wahlen von φ sind benachbart und liefern daher die gleichen Homomorphismen auf der Homologie. Auch die Unabhängigkeit von n folgt dann sofort, da man mit SPERNER-Abbildungen leicht von einem n zu einem größeren kommen kann.

Ersetzt man K und L durch andere Triangulierungen, so haben diese nach den Ergebnissen des letzten Schritts gemeinsame Verfeinerungen mit K beziehungsweise L, es genügt also, eine Verfeinerung $\tilde{K}$ von K und eine Verfeinerung $\tilde{L}$ von L zu betrachten und zu zeigen, daß eine simpliziale Approximation $\psi : \tilde{K}^{(m)} \to \tilde{L}$ die gleichen Homomorphismen definiert. Dies ist aber klar nach dem letzten Schritt.

Die Formel $(g \circ f)_{*q} = g_{*q} \circ f_{*q}$ ist ebenfalls (ziemlich) klar, und da die Identitäten auf X und Y natürlich identische Abbildungen auf den Homologiegruppen induzieren, folgt auch, daß Homöomorphismen zwischen topologischen Räumen Isomorphismen zwischen den Homologiegruppen induzieren.

■

Damit können wir definieren

Definition: Falls es zu dem topologischen Raum X einen abstrakten simplizialen Komplex $\mathfrak{K}$ gibt, so daß X homöomorph zu $|\mathfrak{K}|$ ist, definieren wir die q-te Homologiegruppe von X als

$$H_q(X) \underset{\text{def}}{=} H_q(\mathfrak{K}).$$

Ist Y ein weiterer topologischer Raum, der homöomorph zur topologischen Realisierung eines simplizialen Komplexes $\mathfrak{L}$ sei, und ist $f: X \to Y$ eine stetige Abbildung, so definieren wir Homomorphismen

$$f_{*q}: H_q(X) \to H_q(Y)$$

wie folgt: $|K|$ und $|L|$ seien geometrische Realisierungen von $\mathfrak{K}$ und $\mathfrak{L}$; dann gibt es Homöomorphismen $\rho: |K| \to X$ und $\sigma: Y \to |L|$, die zusammen mit f eine stetige Abbildung

$$g = \sigma \circ f \circ \rho : |K| \to |L|$$

ergeben. Diese Abbildung hat nach dem gerade bewiesenen Satz eine simpliziale Approximation $\varphi: K^{(n)} \to L$, und wir setzen $f_{*q} = \varphi_{*q}$.

Man kann sich leicht (wenn auch umständlich) davon überzeugen, daß f_{*q} nur von f abhängt.

Damit haben wir jedem kompakten triangulierbaren Raume X Homologiegruppen $H_q(X)$ zugeordnet, die bis auf Isomorphie eindeutig bestimmt sind, und das soll uns genügen. Mit (erheblich) mehr Aufwand kann man auch für beliebige topologische Räume Homologiegruppen definieren, wobei man zeigen kann, daß sie für kompakte triangulierbare Räume mit den hier definierten übereinstimmen.

Natürlich kann man die Homologiegruppen eines kompakten triangulierbaren topologischen Raums X auch dadurch ausrechnen, daß man einen geometrischen simplizialen Komplex K betrachtet mit $|K|$ homöomorph zu X und dann die Homologie von K berechnet. Das ist meist anschaulicher und einfacher als das Arbeiten mit abstrakten simplizialen Komplexen.

Zu Beginn dieses Kapitels haben wir uns überlegt, daß für jeden simplizialen Komplex die alternierende Summe der Simplexanzahlen der verschiedenen Dimensionen gleich der alternierenden Summe der BETTI-

Zahlen ist. Nun können wir allgemeiner sagen, daß dies für jede Triangulieerung eines festen kompakten topologischen Raums gilt. Beispielsweise folgt

Korollar: *a)* Für jede Triangulierung der 2-Sphäre mit e Ecken, k Kanten und f Dreiecken ist $e - f + k = 2$.
b) Für jede Triangulierung des Torus mit e Ecken, k Kanten und f Dreiecken ist $e - f + k = 0$.

■

Kapitel 7
Anwendungen der Homologietheorie

Als erste und elementarste Anwendung möchte ich zeigen, daß Dimensionen zumindest bei „gutartigen“ topologischen Räumen nur von der Homöomorphieklasse des Raums abhängen. Beginnen wir mit den Sphären. Die n-Sphäre S^n läßt sich triangulieren durch einen Komplex $\mathfrak{K}$, der aus den sämtlichen Randsimplizes eines $(n+1)$-Simplex s besteht. Ist $\mathfrak{L}$ der Komplex, der zusätzlich auch noch das Simplex s enthält, so hat $\mathfrak{L}$ nach dem Korollar aus Schritt 2 im vorigen Paragraphen triviale Homologie. Für $q < n$ ist aber $H_q(\mathfrak{K}) = H_q(\mathfrak{L})$, denn $\mathfrak{K}$ und $\mathfrak{L}$ haben für $q < n$ genau die gleichen q-Ketten, und auch die Randoperatoren sind gleich. Für $q = n$ ist in $\mathfrak{L}$ wegen dessen trivialer Homologie $Z_n(\mathfrak{L}) = B_n(\mathfrak{L})$, d.h. $Z_n(\mathfrak{L})$ wird erzeugt vom Rand ∂s von s. Da $Z_n(\mathfrak{L}) = Z_n(\mathfrak{K})$ ist, $\mathfrak{K}$ aber keine n-Ränder hat, ist also $H_n(\mathfrak{K}) \cong \mathbb{Z}$ und wir erhalten

Lemma: Die n-Sphäre $\mathbb{S}^n$ hat für $n \geq 1$ die Homologiegruppen

$$H_q(\mathbb{S}^n) = \begin{cases} \mathbb{Z} & \text{für } q = 0 \text{ und } q = n \\ 0 & \text{sonst.} \end{cases}$$

■

Korollar: Die Sphären $\mathbb{S}^n$ und $\mathbb{S}^m$ sind nur dann homöomorph, wenn $n = m$ ist.

Beweis: Falls zwei topologische Räume homöomorph sind, haben sie isomorphe Homologiegruppen; dies ist aber für Sphären positiver Dimension nach dem Lemma nur für $n = m$ möglich. Die 0-Sphäre kann zu keiner Sphäre positiver Dimension homömorph sein, da sie endlich (und auch nicht zusammenhängend) ist.

■

Korollar: $\mathbb{R}^n$ und $\mathbb{R}^m$ sind für $n \neq m$ nicht homöomorph.

Beweis: Wären $\mathbb{R}^n$ und $\mathbb{R}^m$ homöomorph, so wären es auch ihre ALEXANDROFF-Kompaktifizierungen (*s.* 3. Übungsblatt, Aufgabe 2). Die ALEXANDROFF-Kompaktifizierung von $\mathbb{R}^n$ mit $n > 0$ ist aber die n-Sphäre $\mathbb{S}^n$: Identifiziert man irgendeinen Punkt N von $\mathbb{S}^n$ (den „Nordpol") mit dem Punkt ∞, so wird $\mathbb{S}^n \setminus \{N\}$ über die stereographische Projektion homöomorph zu $\mathbb{R}^n$. Damit folgt die Behauptung für positives n aus dem gerade bewiesenen Lemma, und $\mathbb{R}^0$ als einpunktige Menge kann natürlich nicht homöomorph zu einem anderen $\mathbb{R}^n$ sein. ∎

Dieses Korollar läßt sich dahingehend verallgemeinern, daß man für größere Klassen topologischer Räume einen unter Homöomorphie invarianten Dimensionsbegriff erklären kann; für uns sind jedoch andere Fragen wichtiger, so daß ich nicht näher darauf eingehen möchte.

Eine zweite wichtige Anwendung der Homologietheorie sind Fixpunktsätze. Ein möglicher Zugang dazu ist der HOPFsche Spursatz, der den EULERschen Polyedersatz verallgemeinert. Um ihn zu formulieren, müssen wir zunächst für Homomorphismen von abelschen Gruppen eine Spur definieren, ähnlich zur Spur einer linearen Abbildung zwischen Vektorräumen. Da wir uns auf endliche Komplexe beschränken, reicht es, diese Spur für endlich erzeugbare abelsche Gruppen zu definieren. Sei also A eine endlich erzeugbare abelsche Gruppe, und sei $f\colon A \to A$ ein Homomorphismus. Weiter sei $\{v_1, \ldots, v_r\}$ eine maximale linear unabhängige Teilmengen von A. Dann sind die $f(v_i)$ nicht notwendigerweise darstellbar als Linearkombinationen der v_i, denn $f(v_i)$ könnte endliche Ordnung haben, also schon für sich alleine linear abhängig sein. Auf jeden Fall ist aber ein ganzzahliges Vielfaches $n_i f(v_i)$ eine Linearkombination der v_i, denn sonst wäre $\{v_1, \ldots, v_r, f(v_i)\}$ eine linear unabhängige Menge. Daher können wir schreiben

$$n_i f(v_i) = \sum_{j=1}^{r} a_{ij} v_j.$$

Falls $f(v_i)$ endliche Ordnung hat, ist dabei natürlich $n_i f(v_i) = 0$, also verschwindet $a_{ij} = 0$ für alle j. Offensichtlich sind die Quotienten

$c_{ij} = a_{ij}/n_i$ unabhängig von der Wahl von n_i, die Matrix (c_{ij}) hängt also nur ab von f und wir bezeichnen ihre Spur als Spur von f:

$$\mathrm{Spur}(f) \underset{\mathrm{def}}{=} \sum_{i=1}^{r} c_{ii} = \sum_{i=1}^{r} a_{ii}/n_i \ .$$

Dies ist gerade die Spur jener Abbildung $\tilde{f}$, die aus f entsteht durch Fortsetzung auf den $\mathbb{R}$-Vektorraum mit Basis $\{v_1, \ldots, v_r\}$, sie entspricht also dem, was wir aus der linearen Algebra gewohnt sind. Aus diesem Grund ist sie auch unabhängig von der Wahl der v_i, denn die Spur einer linearen Abbildung zwischen Vektorräumen ist die Summe der Eigenwerte dieser Abbildung und damit unabhängig von der Basis.

Definition: $f\colon X \to X$ sei eine stetige Selbstabbildung des kompakten triangulierbaren Raums X, und f_{*q} seien die auf der Homologie von X induzierten linearen Abbildungen. Die LEFSCHETZ-Zahl von f ist

$$\Lambda(f) = \sum_{q \geq 0} (-1)^q \mathrm{Spur}(f_{*q}).$$

SOLOMON LEFSCHETZ (1884–1972) wurde in Moskau als Sohn jüdischer Eltern mit türkischer Staatsangehörigkeit geboren. Die Familie zog allerdings schon bald nach seiner Geburt nach Frankreich, so daß er dort aufwuchs und ausgebildet wurde mit Französisch als seiner Hauptsprache. Er studierte Ingenieurwissenschaften an der Ecole Centrale in Paris, emigrierte danach aber wegen der besseren Karrierechancen nach USA. Dort arbeitete er zunächst als Ingenieur, verlor aber 1907 bei einer Explosion beide Hände. Danach promovierte er 1911 an der Clark University in Worcester, MA, über ein Thema aus der algebraischen Geometrie und arbeitete dort als *fellow.* 1924 wurde er zunächst Gastprofessor, dann Professor an der Unversität Princeton. Außer auf dem Gebiet der Topologie arbeitete er, vor allem während des Krieges auch an Fragen der angewandten Mathematik und war einer der Begründer der qualitativen Theorie der Differentialgleichungen, die er nach dem Krieg auch mit topologischen Methoden weiterentwickelte.

In völliger Analogie zum Beweis des EULERschen Polyedersatzes folgt nun der

Spursatz von HOPF**:** $\varphi\colon \mathcal{C}(K) \to \mathcal{C}(K)$ sei eine simpliziale Selbstabbildung des endlichen simplizalen Komplexes K. Dann ist

$$\sum_{q\geq 0}(-1)^q \mathrm{Spur}(\varphi_q) = \Lambda(|\varphi|).$$

Beweis: Zunächst überlegt man sich leicht, daß jeder Homomorphismus abelscher Gruppen $h\colon A \to A$, der die Untergruppe B auf sich selbst abbildet, ein Homomorphismus $\bar{h}\colon A/B \to A/B$ induziert und daß gilt

$$\mathrm{Spur}(h) = \mathrm{Spur}(h_{|B}) + \mathrm{Spur}(\bar{h}).$$

Nun geht alles genauso weiter wie beim EULERschen Polyedersatz: Da $Z_q(K)$ der Kern von ∂_q ist und $B_{q-1}(K)$ das Bild, ist

$$C_q(K)/Z_q(K) \cong B_{q-1}(K),$$

also

$$\mathrm{Spur}(\varphi_q) = \mathrm{Spur}(\varphi_{q|Z_q(K)}) + \mathrm{Spur}(\varphi_{q-1|B_{q-1}(K)}),$$

und auf Grund der Definition $H_q(K) = Z_q(K)/B_q(K)$ ist

$$\begin{aligned}\mathrm{Spur}(\varphi_{q|Z_q(K)}) &= \mathrm{Spur}(\ \varphi_{*q}\ ) + \mathrm{Spur}(\varphi_{q|B_q(K)})\\ &= \mathrm{Spur}(|\varphi|_{*q}) + \mathrm{Spur}(\varphi_{q|B_q(K)}).\end{aligned}$$

Mithin ist

$$\mathrm{Spur}(|\varphi|_{*q}) = \mathrm{Spur}(\varphi_q) - \mathrm{Spur}(\varphi_{q-1|B_{q-1}(K)}) - \mathrm{Spur}(\varphi_{q|B_q(K)}),$$

wobei formal $B_{-1}(K) = 0$ zu setzen ist. Bildet man die alternierende Summe, so heben sich wegen $B_n(K) = 0$ die Spuren auf den Gruppen der Ränder gerade weg, womit die Behauptung gezeigt wäre. ■

HEINZ HOPF (1894–1971) wurde im damaligen Breslau geborten und begann dort auch 1913 sein Mathematikstudium. Bei Kriegsausbruch 1914 meldete er sich freiwillig und war den ganzen Krieg über als Leutnant an der Westfront. Danach setzte er sein Studium in Breslau fort, wechselte aber schon nach einem Jahr nach Heidelberg und 1920 nach Berlin, wo er 1925 promovierte. Danach ging er nach Göttingen, wo er 1926 habilitiert wurde. 1931 erhielt er einen Lehrstuhl an der ETH Zürich, den er bis zu seiner Emeritierung innehatte. Er verallgemeinerte unter anderem Sätze von LEFSCHETZ und trug wesentlich zum Ausbau der algebraischen Topologie bei.

Der folgende Satz erlaubt es in vielen Fällen, die Existenz eines Fixpunkts zu zeigen:

Satz: Falls die stetige Selbstabbildung $f\colon X \to X$ des kompakten triangulierbaren Raums X keinen Fixpunkt hat, ist $\Lambda(f) = 0$.

Beweis: K sei ein geometrischer simplizialer Komplex, der X trianguliert, und g sei die durch den Homöomorphismus $X \to |K|$ induzierte stetige Selbstabbildung $|K| \to |K|$. Dann hat auch g keinen Fixpunkt, für jeden Punkt $P \in |K|$ ist also der Abstand $d\big(P, g(P)\big)$ positiv. Da $|K|$ kompakt ist, gibt es daher eine Zahl $\varepsilon > 0$, so daß $d\big(P, g(P)\big) \geq \varepsilon$ für alle $P \in |K|$. Indem wir zu einer baryzentrischen Unterteilung von K übergehen, können wir o.B.d.A. annehmen, daß die Maschenweite von K kleiner als $\varepsilon/3$ ist.

Wie wir aus dem letzten Paragraphen wissen, hat g für hinreichend großes r eine simpliziale Approximation $\varphi\colon K^{(r)} \to K$, und damit gibt es auch Gruppenhomomorphismen

$$\varphi_q\colon C_q(K^{(r)}) \to C_q(K).$$

Im fünften Schritt der Konstruktion in Kapitel 6 haben wir zur SPERNER-Abbildung $\omega\colon K' \to K$ eine Kettenabbildung $C_q(K) \to C_q(K')$ definiert, die im wesentlichen jedem Simplex die Summe seiner Unterteilungssimplizes zuordnet. Wenn wir mit r solchen Abbildungen schachteln, erhalten wir aus φ eine Kettenabbildung $\psi\colon \mathcal{C}(K) \to \mathcal{C}(K)$, die auf der Homologie diesselbe Abbildung induziert wie φ. Wir wollen einsehen, daß Spur(ψ_q) für alle q verschwindet; daraus folgt dann natürlich sofort die Behauptung des Satzes.

Sei also t ein q-Simplex aus K und P ein Punkt aus t. Da die Maschenweite von K kleiner als $\varepsilon/3$ ist und φ die Abbildung g simplizial approximiert, ist $d\big(|\varphi|(P), g(P)\big) < \varepsilon/3$. Andererseits ist nach Definition von ε der Abstand $d\big(P, g(P)\big) > \varepsilon$, also $d\big(P, |\varphi|(P)\big) > 2\varepsilon/3$. Für einen beliebigen Punkt Q aus t, dessen Abstand von P ja höchstens $\varepsilon/3$ sein kann, ist also immer noch $d\big(Q, |\varphi|(P)\big) > \varepsilon/3$, daher ist $|\varphi|(t) \cap t = \emptyset$. Damit ist auch für jedes Simplex s aus $K^{(r)}$, das in t liegt, $|\varphi|(s) \cap t = \emptyset$, also ist $\varphi_q(s) \neq t$ und t tritt nicht auf in der Kette $\psi_q(t)$, die ja gerade die Summe der $\varphi_q(s)$ ist. Also kommt kein Simplex t aus K in $\psi_q(t)$

vor, der entsprechende Koeffizient in der Diagonale der Matrix von ψ_q ist also stets Null, und damit verschwindet auch die Spur von ψ_q, wie behauptet.

■

Als unmittelbare Anwendung erhalten wir den

Fixpunktsatz von BROUWER**:** Jede stetige Abbildung $f\colon \mathbb{B}^n \to \mathbb{B}^n$ der Kugel auf sich selbst hat mindestens einen Fixpunkt.

Beweis: Da $H_q(\mathbb{B}^n) = 0$ für $q > 0$, ist $\Lambda(f)$ gleich der Spur von f_{*0} auf $H_0(\mathbb{B}^n) \cong \mathbb{Z}$. Ist aber $\varphi\colon K^{(r)} \to K$ eine simpliziale Approximation von f, so wird $H_0(K)$ erzeugt von der Homologieklasse $[P]$ irgendeiner Ecke P aus K, und $\varphi_{*0}([P]) = [\varphi_0(P)] = [P]$, da je zwei Ecken durch einen Kantenzug in K verbunden werden könnens und somit in der gleichen Klasse liegen. Somit ist φ_{*0} die Identität und hat die Spur eins. Also ist $\Lambda(f) = 1$ von Null verschieden, und der Satz folgt aus dem gerade bewiesenen.

■

LUITZEN EGBERTUS JAN BROUWER (1881–1966) wurde nahe Rotterdam geboren. Er beendete seine Schulausbildung schon im Alter von 14 Jahren. Da er aber weder Latein noch Griechisch gelernt hatte, konnte er nicht an die Universität und mußte zunächst zwei Jahre lang diese Sprachen lernen. 1897 bestand er dann die Aufnahmeprüfung der Universität Amsterdam. Er interessierte sich schon früh für die Philosophie der Mathematik und promovierte auch 1907 über ein entsprechendes Thema, merkte aber bald, daß er mit solchen Themen kein großes Echo fand. Deshalb arbeitete er von 1909 bis 1913 intensiv auf seinem zweiten Interessengebiet, der Topologie, und bewies dabei weit mehr als nur seinen Fixpunktsatz. Als er 1912 vor allem wegen seiner Beiträge zur Topologie einen Lehrstuhl in Amsterdam bekam, hielt er seine Antrittsvorlesung allerdings über Intuitionismus und Formalismus. Er hielt auch nie Vorlesungen über Topologie, da die Beweismethoden seiner Sätze nicht seinen konstruktivistischen Standards genügten und er sie daher nicht mehr als wirklich bewiesen betrachtete.

Auch für die Sphäre ist $H_0(\mathbb{S}^n) \cong \mathbb{Z}$, und für jede Abbildung $f\colon \mathbb{S}^n \to \mathbb{S}^n$ ist f_{*0} die Identität, hat also Spur eins. Die einzige weitere nichttriviale

Homologiegruppe ist $H_n(\mathbb{S}^n) \cong \mathbb{Z}$; sie wird erzeugt von der Klasse c der Summe aller n-Simplizes eines Komplexes, der $\mathbb{S}^n$ trianguliert.

Definition: Der Grad $\deg(f)$ einer stetigen Abbildung $f\colon \mathbb{S}^n \to \mathbb{S}^n$ ist jene ganze Zahl, für die $f_{*n}(c) = \deg(f) \cdot c$ ist.

Dann folgt natürlich sofort

Satz: Die LEFSCHETZ-Zahl einer stetigen Abbildung $f\colon \mathbb{S}^n \to \mathbb{S}^n$ ist $1+(-1)^n \deg(f)$. Falls f keinen Fixpunkt hat, ist also $\deg(f) = (-1)^{n+1}$. ∎

Ein einfaches Beispiel einer fixpunktfreien Abbildung $f\colon \mathbb{S}^n \to \mathbb{S}^n$ ist die *Antipoden-Abbildung,* die jedem Punkt sein Bild unter der Spiegelung am Mittelpunkt der Kugel zuordnet. Diese Abbildung hat offensichtlich keinen Fixpunkt auf $\mathbb{S}^n$, ihr Grad ist also $(-1)^{n+1}$.

Satz: Eine stetige Abbildung $f\colon \mathbb{S}^n \to \mathbb{S}^n$ mit $\deg(f) \neq 0$ ist surjektiv.

Beweis: Angenommen, f sei nicht surjektiv. Da $f(\mathbb{S}^n)$ auf jeden Fall kompakt ist, gibt es dann eine abgeschlossene Teilmenge $A \subset \mathbb{S}^n$ mit positivem Durchmesser δ, so daß A nicht im Bild liegt. Der Komplex K sei eine Triangulierung von $\mathbb{S}^n$ mit Maschenweite kleiner $\delta/3$, und c sei die Summe aller n-Simplizes aus K. Dann folgt genau wie beim Beweis, daß $\Lambda(f) = 0$ für fixpunktfreie Abbildungen, daß es in A ein n-Simplex gibt, das nicht in der Kette $f_{*n}(c)$ auftritt. Da $f_{*n}(c)$ aber jedes n-Simplex mit gleicher Vielfachheit enthält, muß $f_{*n}(c) = 0$ sein, also $\deg(f) = 0$, wie behauptet. ∎

Als Beispiel können wir etwa ein Polynom n-ten Grades

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \quad a_n \neq 0$$

mit komplexen Koeffizienten a_i betrachten. Die dadurch definierte Abbildung $f\colon \mathbb{C} \to \mathbb{C}$ läßt sich fortsetzen zu einer stetigen Abbildung $\hat{f}\colon \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ der RIEMANNschen Zahlenkugel $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$; diese wiederum ist homöomorph zu $\mathbb{S}^2$. Falls wir nun wüßten, daß $\deg(\hat{f}) \neq 0$ ist, würde folgen, daß $\hat{f}$ surjektiv ist, insbesondere würde also der Wert Null angenommen, und das bedeutet wegen $\hat{f}(\infty) = \infty$, daß f eine

Nullstelle hat – wir hätten also den Fundamentalsatz der Algebra bewiesen.

Für das spezielle Polynom $f(z) = a_n z^n$ können wir den Grad leicht ausrechnen: Da die Gleichung $z^n = w/a_n$ für $w \neq 0$ genau n Lösungen hat, ist in diesem Fall $\deg(f) = n$. Für ein beliebiges Polynom n-ten Grades können wir nur dann so argumentieren, wenn wir den Fundamentalsatz der Algebra voraussetzen; für seinen Beweis müssen wir uns also etwas anderes einfallen lassen.

Die neue Idee ist folgende: Wir können ein beliebiges Polynom

$$f = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$$

vom Grad n deformieren in das spezielle Polynom $a_n z^n$ durch die Abbildung

$$F\colon \begin{cases} [0,1] \times \mathbb{C} \to \mathbb{C} \\ \qquad (t,z) \mapsto t a_n z^n + (1-t) f(z), \end{cases}$$

für die $F(0,z) = f(z)$ und $F(1,z) = a_n z^n$ ist für alle z. Diese Abbildung kann fortgesetzt werden zu einer stetigen Abbildung

$$F\colon [0,1] \times \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}\,,$$

wobei $F(t,\infty) = \infty$ für alle t. Um deren Stetigkeit in ∞ zu zeigen, können wir einfach die Kugel am Äquator spiegeln und dadurch ∞ mit Null vertauschen. Der Äquator entspricht einem Kreis in $\mathbb{C}$, also die Menge aller Punkte $z \in \mathbb{C}$ mit $z\overline{z} = c$ für eine geeignete reelle Zahl c. Wenn wir die stereographische Projektion wie in Kapitel 1 definieren, hat der Kreis den Radius 2, d.h. $c = 4$. Hier ist es aber günstiger, eine Kugel vom Radius $\frac{1}{2}$ auf die Ebene zu legen; dann entspricht der Äquator dem Einheitskreis und die Spiegelung am Äquator kann dann auf $\mathbb{C}$ realisiert werden durch die Funktion $z \mapsto 1/\overline{z}$. Für die Funktion $G(t,w)$, deren Stetigkeit im Nullpunkt für alle t nachzurechnen ist, gilt dann

$$\begin{aligned} \overline{G(t,w)} &= \frac{1}{t a_n \overline{w}^{-n} + (1-t) f(1/\overline{w})} \\ &= \frac{\overline{w}^n}{a_n + (1-t)(a_{n-1}\overline{w} + \cdots + a_0 \overline{w}^n)}\,. \end{aligned}$$

Da sowohl die komplexe Konjugation als auch die Grundrechenarten außer der Division duch Null stetig sind, ist auch diese Funktion wegen $a_n \neq 0$ stetig in der Umgebung eines jeden Punktes $(t, 0)$, und damit ist auch $G(t, w)$ dort stetig.

Wenn wir zeigen können, daß sich der Grad bei einer solchen Deformation nicht ändert, ist der Fundamentalsatz der Algebra bewiesen. Diese Invarianz des Grades wollen wir gleich etwas allgemeiner beweisen. Der Homotopiebegriff, den wir dazu einführen, wird uns auch noch für andere Anwendungen nützlich sein.

Definition: Zwei stetige Abbildungen $f, g\colon X \to Y$ zwischen zwei topologischen Räumen X und Y heißen *homotop,* in Zeichen $f \simeq g$, wenn es eine stetige Abbilung $F\colon [0,1] \times X \to Y$ gibt, so daß gilt: $F(0, x) = f(x)$ und $F(1, x) = g(x)$ für alle $x \in X$.

Satz: X und Y seien kompakte triangulierbare topologische Räume, und $f, g\colon X \to Y$ seien zueinander homotope stetige Abbildungen. Dann ist $f_{*q} = g_{*q}$ für alle q.

Beweis: Wir können o.B.d.A. annehmen, daß X und Y Polyeder sind; sie seien gegeben durch die geometrischen simplizialen Komplexe K und L. Weiter sei δ die LEBESGUE-Zahl zur Überdeckung von Y durch die Sterne der Ecken von L. Da die Abbildung $F\colon [0,1] \times X \to Y$ stetig ist auf einer kompakten Menge, ist sie gleichmäßig stetig, es gibt also ein $\varepsilon > 0$, so daß gilt

$$|t_2 - t_1| < \varepsilon \Rightarrow |F(t_2, x) - F(t_1, x)| < \delta/3 \quad \text{für alle } x \in X.$$

Wir wählen eine Unterteilung

$$0 = t_0 < t_1 < \cdots < t_r = 1$$

des Einheitsintervalls derart, daß $|t_i - t_{i-1}| < \varepsilon$ für $i = 1, \ldots, n$ und setzen $f_i(x) = F(t_i, x)$ für alle i. Insbesondere ist $f_0 = f$ und $f_r = g$, und deshalb reicht es, wenn wir zeigen, daß $(f_{i-1})_{*q} = (f_i)_{*q}$ für alle i und q.

Setzen wir der Einfachheit halber $h = f_{i-1}$ und $k = f_i$. Nach Konstruktion ist $|k(x) - h(x)| < \delta/3$ für alle x.

Die Behauptung, daß $h_{*q} = k_{*q}$ für alle q, folgt natürlich sofort, wenn wir wissen, daß h und k eine gemeinsame simpliziale Approximation haben, und genau das soll nun gezeigt werden:

$Q_1, \ldots, Q_\ell$ seien die Ecken von L. Dann haben wir zu jedem Q_i den Stern $\mathrm{st}(Q_i)$, und darin die offene Teilmenge

$$V_i = \{Q \in Y \mid d\big(Q, |L| \setminus \mathrm{st}(Q_i)\big) > \delta/3\},$$

die aus allen jenen Punkten des Sterns besteht, die einen hinreichend großen Abstand vom Rand des Stern haben. Die V_i sind immer noch groß genug, um ganz Y zu überdecken: Ist nämlich Q ein beliebiger Punkt von Y, so hat die abgeschlossene $\delta/3$-Umgebung von Q einen Durchmesser kleiner δ, liegt also nach Wahl von δ ganz in einem der Sterne $\mathrm{st}(Q_j)$, und ihr Mittelpunkt Q hat somit mindestens den Abstand $\delta/3$ von jedem Punkt aus Y, der nicht in $\mathrm{st}(Q_j)$ liegt.

Da die V_i somit eine Überdeckung von Y bilden, bilden ihre Urbilder $h^{-1}(V_i)$ eine offene Überdeckung von X; deren LEBESGUE-Zahl sei ε. Dazu gibt es, wie wir im letzten Paragraphen gesehen haben, eine baryzentrische Unterteilung $K^{(n)}$ von K, deren Maschenweite kleiner als ε ist. Wir approximieren h durch eine simpliziale Abbildung $\varphi\colon K^{(n)} \to L$, die jeder Ecke P aus $K^{(n)}$ irgendeine Ecke Q_i von L zuordnet, für die $\mathrm{st}(P)$ in $h^{-1}(V_i)$ liegt. Diese Abbildung ist auch eine simpliziale Approximation von k, denn da $h(P)$ und $\varphi(P)$ beide in V_i liegen und $k(P)$ höchstens den Abstand $\delta/3$ von $h(P)$ hat, liegt auch $k(P)$ im Stern von $Q_i = \varphi(P)$. Damit ist der Satz bewiesen. ∎

Damit zurück zum Fundamentalsatz der Algebra: Ein Polynom

$$f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0, \quad a_n \neq 0$$

vom Grad n, aufgefaßt als Funktion auf der RIEMANNschen Zahlenkugel, ist – wie wir schon vor der Definition des Homotopiebegriffs gesehen haben – homotop zur Abbildung $z \mapsto a_n z^n$ vom Grad n, hat also auch topologisch den Grad $n \neq 0$ und ist somit surjektiv. Insbesondere folgt der

Fundamentalsatz der Algebra: Jedes Polynom positiven Grades mit komplexen Koeffizienten hat mindestens eine komplexe Nullstelle. ∎

Doch nun zu den eigentlichen Anwendungen der Homotopie! Als erstes erlaubt sie es uns, eine größere „Deformationsinvarianz“ der Homologie zu zeigen, als wir sie bislang kennen:

Definition: Zwei topologische Räume X und Y heißen *homotop,* in Zeichen $X \simeq Y$, falls es stetige Abbildungen $f\colon X \to Y$ und $g\colon Y \to X$ gibt, so daß $g \circ f$ homotop zur Identität von X ist und $f \circ g$ homotop zur Identität von Y. Ein topologischer Raum X heißt *zusammenziehbar,* wenn er homotop zum einpunktigen topologischen Raum ist.

Falls wir anstelle der Homotopie zur Identität Gleichheit verlangt hätten, wäre dies gerade die Definition der Homöomorphie, Homöomorphie ist also ein Spezialfall der Homotopie. Homotopie ist allerdings sehr viel allgemeiner. Beispielsweise gilt

Lemma: Die n-dimensionale Vollkugel $\mathbb{B}^n$ ist zusammenziehbar.

Beweis: f sei die Identität von

$$\mathbb{B}^n = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 \leq 1\},$$

und g sei die konstante Abbildung, die jeden Punkt auf den Nullpunkt abbildet. Dann ist $f \circ g$ die Identität auf der einpunktigen Menge, und $g \circ f = g$ ist homotop zur Identität auf $\mathbb{B}^n$ via $F(t, x) = tx$. ∎

Genauso überlegt man sich leicht, daß auch jeder $\mathbb{R}^n$ zusammenziehbar ist, nicht aber beispielsweise eine Sphäre, denn natürlich gilt

Satz: Homotope topologische Räume haben isomorphe Homologiegruppen; insbesondere haben zusammenziehbare topologische Räume triviale Homologie.

Beweis: Sind X und Y homotop, gibt es stetige Abbildungen $f\colon X \to Y$ und $g\colon Y \to X$, so daß $g \circ f$ homotop zur Identität auf X ist und $f \circ g$ zu der auf Y. Da zueinander homotope Abbildungen die gleichen

Homomorphismen auf der Homologie induzieren, ist daher $g_{*q} \circ f_{*q}$ die Identität auf $H_q(X)$ und $f_{*q} \circ g_{*q}$ ist die auf $H_q(Y)$. Somit sind f_{*q} und g_{*q} Isomorphismen.

■

Dieser Satz läßt sich häufig anwenden, um die Berechnung von Homologiegruppen zu vereinfachen. Betrachten wir etwa den Zylinder und das MOEBIUSband. Beide sind definiert als Quotienten des Quadrats, das wir hier am besten in der Form

$$Q = [0, 1] \times [-\tfrac{1}{2}, \tfrac{1}{2}]$$

darstellen; für den Zylinder wird der Punkte $(0, y)$ identifiziert mit $(1, y)$, für das Moebiusband mit $(1, -y)$. Beide topologische Räume enthalten die Kreislinie als die durch $y = 0$ definierte Teilmenge, und beide sind homotop zur Kreislinie, wie man sich leicht überzeugt anhand der Abbildung

$$F\big(t, (x, y)\big) = (x, ty).$$

Also haben beide Räume dieselbe Homologie wie $\mathbb{S}^1$, die nullte und erste Homologiegruppe sind also isomorph zu $\mathbb{Z}$, und alle anderen sind Null.

Eine weitere Anwendung des gerade bewiesenen Satzes ist eine Verallgemeinerung des BROUWERschen Fixpunktsatzes:

Satz: Für einen zusammenziehbaren triangulierbaren (kompakten) topologischen Raum X hat jede stetige Abbildung $f\colon X \to X$ mindestens einen Fixpunkt.

Beweis: Da X zusammenziehbar ist, ist $H_0(X) \cong \mathbb{Z}$, und alle höheren Homologiegruppen verschwinden. Also können wir wörtlich genauso argumentieren wie beim BROUWERschen Fixpunktsatz.

■

Zum Abschluß betrachten wir noch eine weitere Anwendung der Homotopie auf Abbildungen von Sphären, die auch wieder vor allem für die Algebra interessant ist. Ausgangspunkt dazu ist der

Satz: Für gerade Dimensionen n gibt es keine stetige Abbildung $f\colon \mathbb{S}^n \to \mathbb{S}^n$ derart, daß die Ortsvektoren von x und $f(x)$ für jeden Punkt $x \in \mathbb{S}^n$ aufeinander senkrecht stehen.

Beweis: Angenommen, es gäbe eine solche Abbildung f. Wir betrachten $\mathbb{S}^n$ als Einheitskugel in $\mathbb{R}^{n+1}$. Dann hat der Vektor

$$F(t,x) = (1-t)f(x) + tx \in \mathbb{R}^{n+1}$$

die Länge $(1-t)^2 + t^2 > 0$,

$$G(t,x) = \frac{F(t,x)}{|F(t,x)|}$$

liegt also in $\mathbb{S}^n$ und vermittelt eine Homotopie zwischen f und der Identität, denn

$$G(0,x) = \frac{f(x)}{|f(x)|} = f(x) \quad \text{und} \quad G(1,x) = \frac{x}{|x|} = x\,.$$

Da die Identität den Grad eins hat, ist auch

$$\deg(f) = 1 \neq (-1)^{n+1},$$

f hat also mindestens einen Fixpunkt, was wegen der Orthogonalität der Ortsvektoren von x und $f(x)$ absurd ist.

■

Im Spezialfall $n = 2$ formuliert man dieses Resultat auch gelegentlich als

Satz vom Igel: Einen Igel kann man nicht kämmen.

In der Tat: Könnte man ihn kämmen, so wären die Stacheln in jedem Punkt parallel zur Schale, stünden also senkrecht auf dem Ortsvektor, und wir hätten eine der Abbildungen, deren Nichtexistenz wir gerade gezeigt haben.

■

Für ungerades n gibt es stetige Abbildungen f, bei denen die Ortsvektoren von x und $f(x)$ immer aufeinander senkrecht stehen, für $n = 1$ beispielsweise die Drehung um 90°. Für den Einheitskreis in $\mathbb{R}^2$ ist das die Abbildung

$$f\colon \begin{cases} \mathbb{S}^1 \to \mathbb{S}^1 \\ (x,y) \mapsto (y,-x) \end{cases},$$

und entsprechend hat acu für jedes ungerade n die Abbildung

$$f\colon \begin{cases} \mathbb{S}^n \to \mathbb{S}^n \\ (x_0, x_1, \ldots, x_{2j}, x_{2j+1}, \ldots, x_{n-1}, x_n) \\ \qquad \mapsto (x_1, -x_0, \ldots, x_{2j+1}, -x_{2j}, \ldots, x_n, -x_{n-1}) \end{cases}$$

diese Eigenschaft. Für die Algebra sehr interessant ist die Frage, ob es vielleicht sogar n Funktionen $f_1, \ldots, f_n$ auf $\mathbb{S}^n$ gibt, so daß alle $f_i(x)$ senkrecht stehen auf x und auf allen $f_j(x)$ für $j \neq i$. In diesem Fall nennen wir $\mathbb{S}^n$ *parallelisierbar.* Geometrisch betrachtet bilden die $f_i(x)$ in diesem Fall in jedem Punkt $x \in \mathbb{S}^n$ eine Basis des Tangentialraums (den ich hier nicht definieren möchte).

Man kann leicht zeigen, daß $\mathbb{S}^1, \mathbb{S}^3$ und $\mathbb{S}^7$ parallelisierbar sind, denn für $n = 1, 3$ und 7 lassen sich auf $\mathbb{R}^{n+1}$ eine „Multiplikation“ und eine „Division“ erklären: Für $\mathbb{R}^2$ ist das die komplexe Multiplikation, für $n = 4$ die (nicht mehr kommutative) Multiplikation der HAMILTONschen Quaternionen, und für $n = 8$ die (nicht einmal mehr assoziative) Multiplikation der CAYLEYschen Oktaven. Allgemein wollen wir unter einer Multiplikation eine bilineare Abbildung $\cdot\colon \mathbb{R}^{n+1} \times \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ verstehen mit der Eigenschaft, daß man durch jedes Element ungleich Null sowohl von links als auch von rechts dividieren kann.

Gibt es eine solche Multiplikation, und ist $\{e_0, \ldots, e_n\}$ eine Basis von $\mathbb{R}^{n+1}$, wobei o.B.d.A. $e_0 = 1$ sei, so sind für jeden Punkt $x \in \mathbb{S}^n \subset \mathbb{R}^{n+1}$ die Vektoren $e_i x$ linear unabhängig voneinander, denn ist

$$\sum_{i=0}^{n} \lambda_i \, e_i \cdot x = 0$$

für irgendwelche $\lambda_i \in \mathbb{R}$, so können wir durch x dividieren und erhalten $\sum \lambda_i e_i = 0$, was natürlich nur möglich ist, wenn alle λ_i verschwinden. Also bilden die $e_i x$ für jedes x eine Basis, die wir orthonormalisieren können zu einer Orthonormalbasis $\{f_0(x), \ldots, f_n(x)\}$, für die immer noch $f_0(x) = x$ ist und $f_i(x)$ stetig abhängt von x. Damit folgt

Satz: $\mathbb{S}^1, \mathbb{S}^3$ und $\mathbb{S}^7$ sind parallelisierbar. ∎

Umgekehrt folgt auch aus der Nichtparallelisierbarkeit von $\mathbb{S}^n$ für gerades $n > 0$ *(Was passiert für $n = 0$?)*

Satz: Für ungerades $n > 1$ kann man auf $\mathbb{R}^n$ keine Multiplikation erklären, zu der es auch eine Division gibt. Insbesondere läßt sich auf $\mathbb{R}^3$ keine solche multiplikative Verknüpfung erklären.

Denn wenn es eine solche Verknüpfung gäbe, könnten wir wie oben argumentieren und die Parallelisierbarkeit von $\mathbb{S}^{n-1}$ zeigen, obwohl es tatsächlich nicht einmal eine einzige Funktion $f(x)$ gibt, so daß $f(x)$ stets senkrecht auf x steht. ■

Für gerades n kennen wir die Gegenbeispiele $n = 2, 4$ und 8, und damit muß die Algebra passen. Auf der *topologischen* Seite konnte aber HOPF 1940 zeigen, daß $\mathbb{S}^n$ nur dann parallelisierbar sein kann, wenn $n + 1$ eine Zweierpotenz ist, und 1958 bewiesen KERVAIR und MILNOR unabhängig voneinander, daß sogar $n = 1, 3$ oder 7 sein muß. Damit ist mit topologischen Methoden gezeigt, daß es außer den drei bekannten keine weiteren Fortsetzungen der reellen Multiplikation auf einen $\mathbb{R}^n$ geben kann.

Kapitel 8
Der Fixpunktsatz von Kakutani

Im vorigen Kapitel haben wir gesehen, daß topologische Fixpunktsätze eine ganze Reihe von Anwendungen auf Probleme der Algebra, Geometrie und Analysis haben. Was uns in dieser Vorlesung besonders interessiert sind allerdings Anwendungen auf Gleichgewichte in einer Volkswirtschaft oder bei Spielen, und dazu reichen die bislang bewiesenen Sätze nicht aus. Das Hauptproblem besteht darin, daß beispielsweise ein Teilnehmer an einer Volkswirtschaft praktisch immer mehr als eine mögliche Strategie verfolgen kann; welche Strategien in Frage kommen, hängt beispielsweise ab vom Preisgefüge des betrachteten Systems.

Zur Modellierung einer solchen Situation reichen topologische Räume und stetige Abbildungen nicht aus: Die möglichen Strategien und auch die Preisniveaus lassen sich zwar in vielen Fällen als Punkte einer Teilmenge eines $\mathbb{R}^n$ betrachten und damit als Punkte eines topologischen Raums. Wenn wir zu gegebenen Preisen die möglichen Strategien betrachten wollen, haben wir aber keine Abbildung zwischen zwei topologischen Räumen, sondern eine Abbildung, die jedem Punkt des einen Raumes eine *Teilmenge* des anderen zuordnet; ihr Zielraum ist also die Potenzmenge des zweiten Raums, und diese hat keine kanonische Struktur als topologischer Raum. Außerdem werden in den meisten Fällen nur sehr spezielle Teilmengen als Werte der Abbildung auftreten.

Deshalb wollen wir solche Vorschriften als eine Art verallgemeinerte Abbildungen zwischen den Räumen selbst betrachten. Früher sprach man von „mehrdeutigen Abbildungen“, heute von Korrespondenzen:

Definition: *a)* Eine *Korrespondenz* $f\colon X \to\to Y$ zwischen zwei Mengen X und Y ist eine Abbildung, die jedem Element $x \in X$ eine

Teilmenge $f(x) \subseteq Y$ zuordnet.
b) Der *Graph* von f ist die Menge

$$\Gamma_f \underset{\text{def}}{=} \big\{(x,y) \in X \times Y \;\big|\; y \in f(x)\big\}\,.$$

c) Das *obere* oder *starke Urbild* einer Teilmenge $Z \subseteq Y$ ist die Menge

$$f^+(Z) \underset{\text{def}}{=} \big\{x \in X \;\big|\; f(x) \subseteq Z\big\}\,.$$

d) Das *untere* oder *schwache Urbild* einer Teilmenge $Z \subseteq Y$ ist die Menge

$$f^-(Z) \underset{\text{def}}{=} \big\{x \in X \;\big|\; f(x) \cap Z \neq \emptyset\big\}\,.$$

e) Ist $Z = \{z\}$ eine einelementige Menge, schreiben wir kurz $f^+(z)$ und $f^-(z)$ an Stelle von $f^+\big(\{z\}\big)$ und $f^-\big(\{z\}\big)$.

Offensichtlich ist das obere Urbild stets im unteren enthalten; falls alle Mengen $f(x)$ einelementig sind, stimmen beide überein und sind gleich der „üblichen“ Urbildmenge $f^{-1}\big(\{x\}\big)$.

Wir interessieren uns für Korrespondenzen zwischen topologischen Räumen und wollen auch für diese Stetigkeitseigenschaften definieren. Eine Abbildung zwischen zwei topologischen Räumen ist stetig, wenn das Urbild einer jeden offenen Menge wieder offen ist; da wir für Korrespondenzen zwei verschiedene Urbilder einer Menge definiert haben, ist die Situation hier offensichtlich komplexer:

Definition: *a)* Eine Korrespondenz $f\colon X \to\to Y$ zwischen zwei topologischen Räumen X und Y heißt *halbstetig nach oben,* wenn das obere Urbild $f^+(U)$ jeder offenen Teilmenge $U \subseteq Y$ offen in X ist.
b) f heißt *halbstetig nach unten,* wenn das untere Urbild $f^-(U)$ jeder offenen Teilmenge $U \subseteq Y$ offen in X ist.
c) f heißt *stetig,* wenn f sowohl halbstetig nach oben als auch halbstetig nach unten ist.

Als erstes, noch ziemlich uninteressantes Beispiel betrachten wir die Korrespondenz

$$f\colon \begin{cases} \mathbb{R} \to\to \mathbb{R} \\ \quad x \mapsto \begin{cases} [-1,\ 1] & \text{falls } x \neq 0 \\ [-2,\ 2] & \text{falls } x = 0 \end{cases} \end{cases} .$$

Sie ist offensichtlich halbstetig nach oben, denn für jede offene Teilmenge $U \subseteq \mathbb{R}$, die das Intervall $[-1,\ 1]$ nicht enthält, ist $f^+(U) = \emptyset$; falls U zwar $[-1,\ 1]$ enthält, nicht aber $[-2,\ 2]$, so ist $f^+(U) = \mathbb{R} \setminus \{0\}$, und enthält U das Intervall $[-2,\ 2]$, so ist $f^+(U) = \mathbb{R}$. In allen drei Fällen ist das obere Urbild offen, womit die Halbstetigkeit nach oben bewiesen wäre.

f ist allerdings nicht halbstetig nach unten, denn für das offene Intervall $(1,\ 2)$ ist $f^-(U)$ die einelementige Menge $\{0\}$ und somit nicht offen.

Umgekehrt ist die Korrespondenz

$$g\colon \begin{cases} \mathbb{R} \to\to \mathbb{R} \\ \quad x \mapsto \begin{cases} [-2,\ 2] & \text{falls } x \neq 0 \\ [-1,\ 1] & \text{falls } x = 0 \end{cases} \end{cases}$$

halbstetig nach unten, denn für jede offene Teilmenge $U \subseteq \mathbb{R}$, die nichtleeren Durchschnitt mit $[-1,\ 1]$ hat, ist $g^-(U) = \mathbb{R}$, und wenn U nur nichtleeren Durchschnitt mit $[-2,\ 2]$ hat, so ist $g^-(U) = \mathbb{R} \setminus \{0\}$, Im Falle $U \cap [-2,\ 2] = \emptyset$ schließlich ist $g^-(U) = \emptyset$. Da aber $g^+\big((-2,\ 2)\big)$ nur die Null enthält, ist g nicht halbstetig nach oben.

Für eine Korrespondenz $f\colon X \to\to Y$, für die alle Mengen $f(x)$ einelementig sind, stimmen obere und untere Urbilder überein und somit ist f genau dann halbstetig nach oben, wenn es halbstetig nach unten ist. Bezeichnet $h\colon X \to Y$ jene Funktion, für die $f(x) = \big\{h(x)\big\}$ ist, so ist beides offensichtlich äquivalent zur Stetigkeit von h. Die Halbstetigkeit von Korrespondenzen hat also nichts zu tun mit der Halbstetigkeit reeller Funktionen einer Veränderlichen. In der neueren englischsprachigen Literatur verwendet man daher im Falle der Korrespondenzen oft nicht mehr das auch in der Analysis gebräuchliche Wort *semicontinous,* sondern ersetzt die lateinische Vorsilbe *semi* durch die gleichbedeutende griechische Vorsilbe und sagt *hemicontinous.* Im Deutschen hat sich noch keine Konvention zur Unterscheidung der klassischen Halbstetigkeit nach oben oder unten einer reellen Funktion und der hier betrachteten Halbstetigkeit von Korrespondenzen herausgebildet.

Im Funktionenfall läßt sich die Stetigkeit, zumindest für Funktionen auf 1-abzählbaren topologischen Räumen, auch über konvergente Folgen

charakterisieren. Im Falle von Korrespondenzen müssen wir zusätzlich noch eine Bedingung an den Zielraum stellen; da dieser bei den uns interessierenden Anwendungen stets kompakt ist, wollen wir uns nur mit dieser Situation beschäftigen:

Lemma: X und Y seien 1-abzählbare topologische Räume, Y sei kompakt, und $f\colon X \to\to Y$ sei eine Korrespondenz derart, daß $f(x)$ für jedes $x \in X$ abgeschlossen in Y ist.
a) f ist genau dann halbstetig nach oben, wenn gilt: Sind $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ konvergente Folgen aus X *bzw.* Y mit Grenzwerten x und y derart, daß $y_n \in f(x_n)$ für alle $n \in \mathbb{N}$, so ist auch $y \in f(x)$.
b) f ist genau dann halbstetig nach unten, wenn gilt: Ist $(x_n)_{n\in\mathbb{N}}$ eine in X konvergente Folge mit Grenzwert x, so gibt es für jedes $y \in f(x)$ eine gegen y konvergente Folge $(y_n)_{n\in\mathbb{N}}$ in Y derart, daß $y_n \in f(x_n)$ für alle $n \in \mathbb{N}$.

Beweis: a) Sei zunächst f halbstetig nach oben und seien $(x_n)_{n\in\mathbb{N}}$ und $(y_n)_{n\in\mathbb{N}}$ konvergente Folgen aus X *bzw.* Y mit Grenzwerten x und y derart, daß $y_n \in f(x_n)$ für alle $n \in \mathbb{N}$. Für jede offene Menge $U \subseteq Y$, die $f(x)$ enthält, ist $f^+(U)$ eine offene Umgebung von x. Da die Folge der x_n gegen x konvergiert, gibt es daher einen Index $N \in \mathbb{N}$, so daß $x_n \in f^+(U)$ für alle $n \geq N$. Insbesondere liegt also $f(x_n)$ für alle $n \geq N$ in U. Erst recht liegen damit die Punkte y_n mit $n \geq N$ in U.

$f(x)$ ist nach Voraussetzung eine abgeschlossene Menge im kompakten Raum Y und damit selbst kompakt. Läge y nicht in $f(x)$, gäbe es daher, wie wir am Ende von Kapitel 3 gesehen haben, offene Mengen U und V, so daß die kompakte Menge $f(x)$ in U läge, der Punkt y in V und $U \cap V = \emptyset$. Wie wir gerade gesehen haben, gäbe es zu U einen Index N, so daß $y_n \in U$ für alle $n \geq N$. Andererseits gäbe es wegen der Konvergenz der y_n gegen y einen Index M, so daß y_n für alle $y \geq M$ in V läge. Beides gleichzeitig ist offensichtlich unmöglich; somit muß y in $f(x)$ liegen.

Umgekehrt habe f die gerade bewiesene Eigenschaft; wir müssen zeigen, daß f halbstetig nach oben ist. Dazu sei $U \subseteq Y$ eine offene Menge. Die Offenheit von $f^+(U)$ ist äquivalent zur Abgeschlossenheit des Komplements $Z = X \setminus f^+(U)$, und diese ist, wie wir gegen Ende von

Kapitel 2 gesehen haben, für 1-abzählbare topologische Räume äquivalent dazu, daß jede konvergente Folge $(z_n)_{n\in\mathbb{N}}$ von Elementen $z_n \in Z$ gegen ein Element von Z konvergiert.

Sei also $(z_n)_{n\in\mathbb{Z}}$ eine Folge von Punkten aus Z mit Grenzwert $z \in X$; wir müssen uns überlegen, daß z in Z liegt.

Falls z nicht in Z liegt, liegt es in $f^+(U)$; das Bild $f(z)$ liegt dann also ganz in U. Da die z_n nicht in $f^+(U)$ liegen, gibt es für jedes z_n ein $y_n \in f(z_n)$, das nicht in U liegt. Da Y kompakt ist, hat die Folge der y_n eine konvergente Teilfolge, die gegen $y \in Y$ konvergiere. Nach Voraussetzung ist dann $y \in f(z) \subseteq U$, und da U eine offene Umgebung von y ist, liegen ab einem gewissen Index auch alle Elemente der Teilfolge in U. Dies widerspricht aber der Konstruktion der y_n. Somit ist Z abgeschlossen und $f^+(U)$ offen.

b) Nun sei f halbstetig nach unten, $(x_n)_{n\in\mathbb{N}}$ eine in X gegen x konvergente Folge und $y \in f(x)$. Da Y 1-abzählbar ist, hat y eine abzählbare Umgebungsbasis aus offenen Mengen U_i mit

$$U_1 \supseteq U_2 \supseteq U_3 \cdots .$$

Wegen der Halbstetigkeit von f nach unten sind dann die Mengen $f^-(U_i)$ offene Umgebungen von x. Da die Folge der x_n gegen x konvergiert, gibt es für jedes $i \in \mathbb{N}$ einen Index N_i, so daß $x_n \in f^-(U_i)$ für alle $n \geq N_i$, d.h. $f(x_n) \cap U_i \neq \emptyset$. Wir wählen für jedes n mit $N_i \leq n < N_{i+1}$ ein $y_n \in f(x_n) \cap U_i$; für $n < N_1$ nehmen wir irgendein $y_n \in f(x_n)$. Dann konvergiert die Folge der y_n gegen y und $y_n \in f(x_n)$ für alle n.

Umgekehrt erfülle f diese Folgenbedingung; wir wollen zeigen, daß f dann halbstetig nach unten ist. Sei also $U \subseteq Y$ offen. Das untere Urbild $f^-(U)$ ist genau dann offen, wenn sein Komplement

$$Z = X \setminus f^-(U) = \left\{ x \in X \;\middle|\; f(x) \cap U = \emptyset \right\}$$

abgeschlossen ist, und das wiederum ist äquivalent dazu, daß jede in X konvergente Folge von Elementen aus Z gegen ein Element aus Z konvergiert.

Sei also $(x_n)_{n\in\mathbb{N}}$ eine Folge von Elementen aus Z mit Grenzwert $x \in X$. Falls x nicht in Z liegt, liegt x in $f^-(U)$, es gibt also ein Element

$y \in f(x) \cap U$. Nach Voraussetzung gibt es dazu eine gegen y konvergente Folge $(y_n)_{n\in\mathbb{N}}$ mit Grenzwert y derart, daß $y_n \in f(x_n)$ für alle n. Wegen der Konvergenz dieser Folge gegen y, gibt es einen Index N, so daß $y_n \in f^-(U)$ für alle $n \geq N$; für solche n ist also $f^-(U) \cap f(x_n) \neq \emptyset$, im Widerspruch zur Annahme $x_n \notin f^-(U)$. ■

Die Halbstetigkeit nach oben läßt sich noch anders charakterisieren:

Lemma: $f\colon X \to\to Y$ sei eine Korrespondenz zwischen zwei 1-abzählbaren topologischen Räumen, Y sei kompakt und alle Bilder $f(x)$ abgeschlossen. f ist genau dann halbstetig nach oben, wenn der Graph

$$\Gamma_f = \big\{(x,y) \in X \times Y \;\big|\; y \in f(x)\big\}$$

abgeschlossen ist.

Beweis: Sei zunächst f halbstetig nach oben und (x_n, y_n) eine konvergente Folge von Punkten aus dem Graph. Ist $(x,y) \in X \times Y$ ihr Grenzwert, so konvergiert nach Definition der Produkttopologie die Folge der x_n gegen x und die der y_n gegen y. Außerdem ist $y_n \in f(x_n)$ für alle n, da die Paare im Graph liegen. Somit ist nach dem vorigen Lemma auch $y \in f(x)$, der Grenzwert (x,y) liegt also in Γ_f.

Umgekehrt sei Γ_f abgeschlossen, die Folge $(x_n)_{n\in\mathbb{N}}$ konvergiere in X gegen x und die Folge der $(y_n)_{n\in\mathbb{N}}$ konvergiere in Y gegen y; außerdem sei $y_n \in f(x_n)$ für alle n. Dann liegen die Paare (x_n, y_n) in Γ_f, also auch ihr Grenzwert (x,y). Damit ist $y \in f(x)$, was zu beweisen war. ■

Nach diesen Vorbereitungen können wir uns an die Verallgemeinerung des BROUWERschen Fixpunktsatzes für Korrespondenzen machen; sie wurde 1941 von KAKUTANI veröffentlicht:

Fixpunktsatz von KAKUTANI: K sei der simpliziale Komplex bestehend aus einem r-Simplex s und dessen sämtlichen Randsimplizes, und $f\colon S \to\to S$ sei eine nach oben halbstetige Korrespondenz auf $S = |K|$. Dann gibt es einen Punkt $x_0 \in S$ mit $x_0 \in f(x_0)$.

Beweis: $K^{(n)}$ sei die n-te baryzentrische Unterteilung von K. Wir wählen für jede Ecke x von $K^{(n)}$ irgendeinen Punkt $y \in f(x)$ und

bezeichnen diesen als $f_n(x)$. Dies definiert eine stetige Abbildung $f_n\colon S \to S$, wenn wir jedem Punkt $x = \sum_{i=0}^{q} \lambda_i x_i$ eines offenen q-Simplex mit Ecken $x_0, \ldots, x_q$ den Punkt $f_n(x) = \sum_{i=0}^{q} \lambda_i f(x_i)$ zuordnen. Nach dem BROUWERschen Fixpunktsatz hat f_n einen Fixpunkt x_n.

Da S kompakt ist, hat die Folge der x_n eine konvergente Teilfolge $(x_{n_\nu})_{\nu\in\mathbb{N}}$; wir wollen uns überlegen, daß deren Grenzwert x_0 in seinem Bild $f(x_0)$ liegt, also der gesuchte Fixpunkt ist.

Für jedes $n \in \mathbb{N}$ sei Δ_n der Abschluß eines r-Simplex aus $K^{(n)}$, für das $x_n \in \Delta_n$ liegt; die Ecken von Δ_n seien $x_0^{(n)}, \ldots, x_r^{(n)}$. Da die Folge der Durchmesser der Simplizes Δ_n mit den Maschenweiten der $K^{(n)}$ gegen Null konvergiert, konvergieren die Teilfolgen $(x_i^{(n_\nu)})_{\nu\in\mathbb{N}}$ für alle i gegen x_0.

Das Bild $y_i^{(n)} = f_n(x_i^{(n)})$ liegt nach Konstruktion von f_n in der Menge $f(x_i^{(n)})$; ist $x_n = \sum_{i=0}^{r} \lambda_i^{(n)} x_i^{(n)}$, so ist $f(x_n) = \sum_{i=0}^{r} \lambda_i^{(n)} y_i^{(n)}$.

Wegen der Kompaktheit von S sowie des Einheitsintervalls hat die Teilfolge $(n_\nu)_\nu \in \mathbb{N}$ selbst eine Teilfolge, die wir zur Vermeidung dreifacher Indizes etwas schlampig mit $(n'_\nu)_{\nu\in\mathbb{N}}$ bezeichnen derart, sowohl die Teilfolgen $y_i^{(n'_\nu)}$ als auch die Teilfolgen $\lambda_i^{(n'_\nu)}$ konvergieren; die jeweiligen Grenzwerte seinen $y_i^{(0)}$ und $\lambda_i^{(0)}$. Dann ist $x_0 = \sum_{i=0}^{r} \lambda_i^{(0)} y_i^{(0)}$. Da die Folgen der $x_i^{(n'_\nu)}$ allesamt gegen x_0 konvergieren und $y_i^{(n'_\nu)}$ in $f(x_i^{(n'_\nu)})$ liegt, zeigt die Halbstetigkeit von f nach oben, daß alle $y_i^{(0)}$ in $f(x_0)$ liegen, also ist

$$x_0 = \sum_{i=0}^{r} \lambda_i^{(0)} y_i^{(0)} \in f(x_0)\,,$$

wie behauptet.

■

Daraus folgt leicht die folgende Verallgemeinerung:

Korollar: $f\colon S \to\to S$ sei eine nach oben halbstetige Korrespondenz auf der kompakten konvexen Teilmenge $S \subset \mathbb{R}^m$. Dann gibt es einen Punkt $x_0 \in S$, für den x_0 in $f(x_0)$ liegt.

Beweis: Als kompakte Teilmenge eines $\mathbb{R}^m$ ist S beschränkt; es gibt daher ein Simplex S', das S enthält. Wir bilden S' ab auf S wie folgt: z sei ein fester Punkt von S. Ein Punkt $x \in S'$ wird auf sich selbst abgebildet, falls er bereits in S liegt; andernfalls wird er abgebildet auf den Schnittpunkt $\rho(x)$ der Strecke von x nach z mit dem Rand von S. Die so definierte Abbildung $\rho\colon S' \to S$ ist also die Identität auf S und bildet ganz S' stetig ab auf S. (Eine solche Abbildung heißt *Retraktion* von S' auf S.) Die Korrespondenz $S' \to S'$, die jedem Punkt $x \in S'$ die Menge $f\big(\rho(x)\big)$ ist offensichtlich halbstetig nach oben; nach dem gerade bewiesenen Satz gibt es also einen Punkt $x_0 \in S'$ mit $x_0 \in f\big(\rho(x_0)\big)$. Da letztere Menge in S liegt, liegt auch x_0 in S, also ist $\rho(x_0) = x_0$ und $x_0 \in f(x_0)$.

■

SHIZUO KAKUTANI (1911–2004) wurde in Osaka geboren und studierte Mathematik an der Tohuku Universität in Sendai, wobei er sich vor allem mit analytischen Funktionen beschäftigte. Seine erste Stelle hatte er als *teaching assistant* an der Universität von Osaka. Auf Einladung HERMANN WEYLs kam er 1940 ans *Institute for Advanced Studies* in Princeton, mußte seinen Aufenthalt aber 1941 wegen des Krieges abbrechen. 1948 kehrte er zurück und arbeitete danach an der University of Illinois und dann bis zu seiner Emeritierung in Yale. Seine Arbeiten beschäftigen sich unter anderem mit Funktionentheorie, analytischen Funktionen, topologischen Gruppen, Funktionalanalysis, MARKOV-Prozessen, Maßtheorie und natürlich seinem Fixpunktsatz.

Satz: $X \subset \mathbb{R}^m$ und $Y \subset \mathbb{R}^n$ seien kompakte und konvexe Mengen, und $U, V \subseteq X \times Y \subset \mathbb{R}^{m+n}$ seien abgeschlossene Teilmengen von $X \times Y$ derart, daß die Mengen

$$U_{x_0} = \big\{y \in Y \;\big|\; (x_0, y) \in U\big\} \quad \text{und} \quad V_{y_0} = \big\{x \in X \;\big|\; (x, y_0) \in V\big\}$$

für alle $x_0 \in X$ *bzw.* $y_0 \in Y$ nicht leer, abgeschlossen und konvex sind. Dann ist $U \cap V \neq \emptyset$.

Beweis: $S = X \times Y$ ist kompakt und konvex. Die Korrespondenz

$$f\colon \begin{cases} S \to\to S \\ (x, y) \mapsto V_y \times U_x \end{cases}$$

ist halbstetig nach oben, denn ihr Graph

$$\Gamma_f = \left\{(x_0, y_0, x, y) \in X \times Y \times X \times Y \;\middle|\; (x, y) \in V_{y_0} \times U_{x_0}\right\}$$

ist homöomorph zur nach Voraussetzung abgeschlossene Menge $U \times V$: Die Bedingung $x \in V_{y_0}$ ist äquivalent zu $(x, y_0) \in U$ und $y \in U_{x_0}$ zu $(x_0, y) \in V$, also liegt (x_0, y_0, x, y) genau dann in Γ_f, wenn (x, y_0, x_0, y) in $U \times V$ liegt, und die Vertauschung von Koordinaten ist natürlich ein Homöomorphismus. Nach dem obigen Korollar gibt es daher einen Punkt $(x_0, y_0) \in X \times Y$, der in seiner Bildmenge $V_{y_0} \times U_{x_0}$ liegt. Damit ist $x_0 \in V_{y_0}$ und $y_0 \in U_{x_0}$, also $(x_0, y_0) \in U \cap V$.

■

Minimax-Theorem: $X \subset \mathbb{R}^m$ und $Y \subset \mathbb{R}^n$ seien kompakt und konvex, und $f\colon X \times Y \to \mathbb{R}$ sei eine stetige Funktion. Weiter seien für alle $x_0 \in X, y_0 \in Y$ und $\alpha, \beta \in \mathbb{R}$ die Mengen

$$\left\{y \in Y \;\middle|\; f(x_0, y) \le \alpha\right\} \quad \text{und} \quad \left\{x \in X \;\middle|\; f(x, y_0) \ge \beta\right\}$$

konvex. Dann ist

$$\max_{x \in X} \min_{y \in Y} f(x, y) = \min_{y \in Y} \max_{x \in X} f(x, y)\,.$$

Zum *Beweis* betrachten wir die beiden Mengen

$$U = \left\{(x_0, y_0) \in X \times Y \;\middle|\; f(x_0, y_0) = \min_{y \in Y} f(x_0, y)\right\}$$

und

$$V = \left\{(x_0, y_0) \in X \times Y \;\middle|\; f(x_0, y_0) = \max_{x \in X} f(x, y_0)\right\}.$$

Die Voraussetzungen des vorigen Satzes sind erfüllt, denn da ein Funktionswert genau dann kleiner oder gleich dem Minimum ist, wenn er mit diesem übereinstimmt, ist für jedes $x_0 \in X$

$$\begin{aligned} U_{x_0} &= \left\{y_0 \in Y \;\middle|\; (x_0, y_0) \in U\right\} \\ &= \left\{y_0 \in Y \;\middle|\; f(x_0, y_0) = \min_{y \in Y} f(x_0, y)\right\} \\ &= \left\{y_0 \in Y \;\middle|\; f(x_0, y_0) \le \min_{y \in Y} f(x_0, y)\right\} \end{aligned}$$

konvex und analog dazu auch V_{y_0} für jedes $y_0 \in Y$. Somit haben U und V nichtleeren Durchschnitt.

Für einen Punkt $(x_0, y_0) \in U \cap V$ ist dann

$$f(x_0, y_0) = \min_{y \in Y} f(x_0, y) = \max_{x \in X} f(x, y_0)\,,$$

also ist

$$\min_{y \in Y} \max_{x \in X} f(x, y) \leq \max_{x \in X} f(x, y_0) = \min_{y \in Y} f(x_0, y) \leq \max_{x \in X} \min_{y \in Y} f(x, y)\,,$$

womit eine Ungleichung zwischen beiden Seiten bewiesen wäre. Die andere ist klar und gilt sogar für eine beliebige Abbildung $f\colon X \times Y \to \mathbb{R}$ mit zwei beliebigen Mengen X und Y: Für alle $x \in X, y \in Y$ ist $f(x, y) \leq \max_{x \in X} f(x, y)$, also auch $\min_{y \in Y} f(x, y) \leq \min_{y \in Y} \max_{x \in X} f(x, y)$ für alle $x \in X$. Damit ist auch

$$\max_{x \in X} \min_{y \in Y} f(x, y) \leq \min_{y \in Y} \max_{x \in X} f(x, y)\,.$$

Somit sind beide Seiten gleich, wie behauptet.

■

JOHN VON NEUMANN bewies diesen Satz 1928 für bilineare Abbildungen auf Simplizes; die obige Verallgemeinerung wurde 1941 von KAKUTANI zusammen mit dessen Fixpunktsatz bewiesen.

JOHN VON NEUMANN (1903-1957) wurde als JÁNOS LAJOS NEUMANN in Budapest geboren und ging auch dort zur Schule. Nach seinem Abschluß 1921 wollte er Mathematik studieren, aber sein Vater war dagegen, da man damit nicht genug Geld verdienen könne. Als Kompromiss zur dem Vater vorschwebenden Karriere als Geschäftsmann einigte man sich schließlich auf ein Studium der Chemie, für das er nach Berlin ging. Vorher schrieb er sich allerdings noch an der Universität Budapest noch für das Fach Mathematik ein, obwohl er dort natürlich nie eine Vorlesung besuchen konnte. Trotzdem bestand er alle Prüfungen mit ausgezeichneten Ergebnissen und wurde 1926 promoviert mit einer Dissertation aus dem Gebiet der Mengenlehre. In seinem Chemiestudium wechselte er 1923 an die ETH Zürich, wo er ebenfalls 1926 ein Diplom in Chemieingenieurwesen erhielt. Neben seinem Chemiestudium hatte er dort auch Kontakte mit Mathematikern wie HERMANN WEYL (1885–1955) und GEORGE PÓLYA (1887–1985). Dank eines Stipendiums konnte er 1926/27 in Göttingen bei DAVID HILBERT (1862–1943) studieren; 1927 habilitierte er in Berlin mit Arbeiten über axiomatische Mengenlehre und über Funktionalanalysis. Daneben interessierten ihn auch Fragen der Quantentheorie, was zu einer Einladung an die Universität Princeton führte. 1931 wurde er dort Professor, und als 1933 in Princeton das *Institute for Advanced*

Study gegründet wurde, war er (zusammen mit ALBERT EINSTEIN (1879–1955) einer der sechs ersten Mathematikprofessoren dort. Er behielt diese Stelle bis zu seinem Tod. Neben seinen bisherigen mathematischen Interessen leistete er auch wichtige Beiträge zur mathematischen Begründung der Quantenmechanik, entwickelte die Spieltheorie und war auch einer der Gründerväter der Informatik.

JOHN VON NEUMANN kam auf diese Fragestellung im Zusammenhang mit der im wesentlichen von ihm entwickelten Spieltheorie. Dabei ging es, trotz des Titels *Eine mathematische Theorie der Gesellschaftsspiele* nicht um optimale Strategien für Schach oder Mensch ärgere dich nicht, sondern die „Spiele" sollten auch nützlich sein, um die Handlungen der Teilnehmer eines Wirtschaftssystems zu beschreiben und rationale Strategien für sie zu finden.

Ein Spiel wird gespielt von n Personen; dem i-ten Spieler stehen dabei die Strategien aus einer gewissen Menge S_i zur Verfügung. In Abhängigkeit von den Strategien *aller* Spieler erhält er danach seinen (nicht notwendigerweise positiven) Auszahlungsbetrag $a^{(i)}(s_1, \ldots, s_n)$. Die Funktionen $a^{(i)}$ sind im Voraus bekannt, die Strategien der anderen Spieler im Allgemeinen nicht.

Als einfaches Beispiel können wir das Spiel *Stein-Schere-Papier* betrachten: Zwei Spieler bilden hinter ihrem Rücken mit einer Hand das Symbol für Stein oder Schere oder Papier; auf ein Kommando zeigen sie gleichzeitig ihre Hände. Dabei gewinnt der Stein über die Schere *(Der Stein schleift die Schere),* die Schere über das Papier *(Die Schere schneidet das Papier)* und das Papier über den Stein *(Das Papier wickelt den Stein ein).* Haben beide Spieler das gleiche Symbol gewählt, geht das Spiel unentschieden aus.

Beschreibt man Gewinn durch die Auszahlung +1, Verlust durch -1 und Unentschieden durch die Null, wird die Auszahlungsfunktion $a^{(1)}$ für den ersten Spieler durch folgende Tabelle gegeben, in der oben die Strategie des ersten Spielers steht und unten die des zweiten:

	Stein	*Schere*	*Papier*
Stein	0	-1	1
Schere	1	0	-1
Papier	-1	1	0

Für den zweiten Spieler haben wir jeweils den betragsgleichen Auszahlungswert mit dem anderen Vorzeichen.

Kehren wir zurück zu einem allgemeinen Spiel mit n Spielern, n Strategiemengen S_i und n Auszahlungsfunktionen

$$a^{(i)}\colon S_1 \times \cdots \times S_n \to \mathbb{R}\,.$$

Welche Strategie sollte der i-te Spieler wählen, um möglichst viel zu gewinnen oder wenigstens möglichst wenig zu verlieren?

Um seine Verluste zu begrenzen, sollte er auf überoptimistische Annahmen bezüglich der Strategien der übrigen Spieler verzichten. Um eine *garantierte* Untergrenze für seine Auszahlung zu bestimmen, muß er davon ausgehen, daß die Mitspieler die (für ihn) schlimmstmöglichen Strategien wählen, d.h. sie wählen Strategien s_j^* derart, daß

$$\begin{aligned} &a^{(i)}(s_1^*, \ldots, s_{i-1}^*, s_i, s_{i+1}^*, \ldots, s_n^*) \\ = \min\ &a^{(i)}(s_1, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_n) \end{aligned}$$

ist, wobei in der zweiten Zeile alle s_j mit $j \neq i$ jeweils ihren gesamten Wertebereich S_j durchlaufen. Dieser Mindestauszahlungsbetrag $m(s_i)$ wird für verschiedene Strategien $s_i \in S_i$ im allgemeinen verschieden hoch sein; er sollte eine Strategie s_i^* wählen, die in möglichst groß macht. Dann kann er *sicher* sein, daß seine Auszahlung mindestens $m(s_i^*)$ beträgt, je nach Strategien der Mitspieler eventuell auch mehr.

Die obige Formel ist trotz ihrer Einfachheit sehr lang. Da wir in Zukunft häufig Ausdrücke ähnlicher Form brauchen werden, empfiehlt es sich daher, eine kompaktere Notation einzuführen: Für ein kartesisches Produkt $S = \prod_{j=1}^n S_j$ bezeichnen wir mit S_{-i} das entsprechende Produkt ohne seinen i-ten Faktor, und für ein n-Tupel $s \in S$ sei $s_{-i} \in S_{-i}$ die Projektion von s nach S_{-i}, also jenes $(n-1)$-Tupel, das aus s durch Streichen der i-ten Komponente entsteht. Mit $s_{-i}|u$ bezeichnen wir jenes n-Tupel aus S, in dem die i-te Komponente von s durch u ersetzt wird. Mit dieser Schreibweise wird die obige doppelzeilige Formel einfach zu

$$a^{(i)}(s_{-i}^*|s_i) = \min_{s_{-i} \in S_{-i}} a^{(i)}(s_{-i}|s_i)\,.$$

Beim Spiel *Stein-Schere-Papier* bringen diese Überlegungen leider nichts: Zu *jeder* eigenen Strategie gibt es eine Strategie des Gegners, die zur Auszahlung -1 führt; es ist also völlig egal, welche man wählt.

Nun wird freilich jedermann einwenden, daß man bei einem Spiel wie *Stein-Schere-Papier* keine feste Strategie wählen *darf,* sondern die Strategie häufig und für den Gegner unvorhersehbar wechseln sollte. Das wußte auch VON NEUMANN, und er beschränkte sich daher nicht auf die bislang betrachteten sogenannten *reinen* Strategien, sondern ließ auch *gemischte* Strategien zu. Eine gemischte Strategie eines Spielers, dem m reine Strategien zur Verfügung stehen, ist ein Vektor $(p_1, \ldots, p_m)$ aus nichtnegativen reellen Zahlen mit Summe eins, der vorschreibt, daß zufällig eine der m Strategien gewählt wird, wobei p_i die Wahrscheinlichkeit für die i-te ist. Die Menge der gemischten Strategien für einen Spieler mit m reinen Strategien ist somit ein abgeschlossenes $(m-1)$-Simplex. Als Auszahlungswert eines n-Tupels aus gemischten Strategie definiert man den Erwartungswert der Auszahlungen bezüglich der durch die Strategien definierte Wahrscheinlichkeitsverteilung auf dem Produkt der Mengen der reinen Strategien.

Beim Spiel *Stein-Schere-Papier* überlegt man sich leicht, daß die optimale gemischte Strategie für jeden Spiele darin besteht, jede der drei reinen Strategien mit der gleichen Wahrscheinlichkeit $1/3$ zu wählen; dann und nur dann ist sein Erwartungswert nicht negativ; siehe Übungsblatt. Spielen beide mit dieser Strategie, ist der Erwartungswert für beide gleich Null, was ja auch der Sinn dieses Spiel ist. Beide Spieler handeln zum eigenen Schaden, wenn sie von dieser Strategie abweichen.

Mit seinem Minimax-Theorem konnte JOHN VON NEUMANN zeigen, daß es für eine größere Klasse von Zwei-Personen-Spielen ein Paar von (im allgemeinen gemischten) Strategien gibt derart, daß sich keiner der beiden Spieler verbessern kann, wenn er von seiner Strategie abweicht, während der Gegner bei der seinigen bleibt:

Definition: *a)* Ein Zwei-Personen-Nullsummenspiel ist ein Spiel mit zwei Spielern, denen jeweils eine endliche Menge von reinen Strategien zur Verfügung stehen, und bei dem die Auszahlungsfunktion des einen Spielers das Negative der Auszahlungsfunktion des anderen ist.

b) Ein *Gleichgewicht* eines n-Personenspiels ist ein n-Tupel von Strategien derart, daß keiner der n Spieler seine Auszahlung verbessern kann, wenn er seine Strategie ändert, während die übrigen Spieler bei ihren Strategien bleiben. Bezeichnet S_i die Strategiemenge des i-ten Spielers, ist $s^* \in S_1 \times \cdots \times S_n$ also genau dann ein Gleichgewicht, wenn gilt:

$$a^{(i)}(s^*) = \max_{s \in S_i} a^{(i)}(s^*_{-i}|s_i) \quad \text{für } i = 1, \ldots, n\,.$$

Satz: In einen Zwei-Personen-Nullsummenspiel gibt es stets mindestens ein Gleichgewicht aus (im Allgemeinen) gemischten Strategien.

Beweis: Der erste Spieler betrachtet nach obigen Überlegungen zu jeder seinen Strategien $s \in S_1$ jene Strategie $t^* \in S_2$ seines Gegners, für die

$$a^{(1)}(s, t^*) = m(s) \underset{\text{def}}{=} \min_{t \in S_2} a^{(1)}(s, t)$$

ist und wählt dann jene Strategie s^*, für die $m(s^*)$ maximal wird. Wählt der Gegner auch dazu die aus Sicht des ersten Spielers schlimmstmögliche Strategie t^*, ist also

$$a^{(1)}(s^*, t^*) = \max_{s \in S_1} \min_{t \in S_2} a^{(1)}(s, t)\,.$$

Entsprechend wählt der zweite Spieler eine Strategie $\widetilde{t}$, so daß auch für die aus seiner Sicht schlimmstmögliche Strategie $\widetilde{s}$ des ersten Spielers gilt

$$a^{(1)}(\widetilde{s}, \widetilde{t}) = \max_{t \in S_2} \min_{s \in S_1} a^{(2)}(s, t)\,.$$

Da wir gemischte Strategien betrachten, sind S_1 und S_2 abgeschlossene Simplizes, also kompakt und konvex; auch ihr Produkt $S_1 \times S_2$ ist somit kompakt und konvex. Bezeichnet $a^{(1)}_{ij}$ die Auszahlung für den ersten Spieler unter der Voraussetzung, daß dieser die i-te seiner m_1 reinen Strategien und der zweite die j-te seiner m_2 reinen Strategien spielt, so ist die Auszahlung $a^{(1)}(p, q)$ für die beiden gemischten Strategien $p \in S_1$ und $q \in S_2$ gegeben durch die offensichtlich stetige Funktion

$$a^{(1)}\colon \begin{cases} S_1 \times S_2 \to \mathbb{R} \\ \quad (p, q) \mapsto \displaystyle\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} p_i q_j a^{(1)}_{ij} \end{cases}$$

Da sie in jedem ihrer beiden Argumente linear ist, sieht man auch sofort, daß

$$\left\{p \in S_1 \;\middle|\; a^{(1)}(p,q) \leq \alpha\right\}$$

für alle $q \in S_2$ und alle $\alpha \in \mathbb{R}$ konvex ist; entsprechend auch

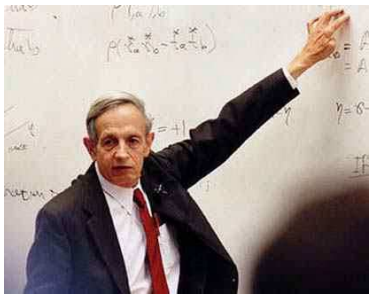
$$\left\{q \in S_1 \;\middle|\; a^{(1)}(p,q) \geq \beta\right\}$$

für alle $p \in S_1$ und $\beta \in \mathbb{R}$. Somit sind alle Voraussetzungen des Minimax-Theorems erfüllt und wir haben

$$\begin{aligned} a^{(1)}(\widetilde{s},\widetilde{t}) &= -a^{(2)}(\widetilde{s},\widetilde{t}) = -\max_{t\in S_2}\min_{s\in S_1} a^{(2)}(s,t) \\ &= -\max_{t\in S_2}\min_{s\in S_1}\left(-a^{(1)}(s,t)\right) \\ &= \min_{t\in S_2}\max_{s\in S_1} a^{(1)}(s,t) = \max_{s\in S_1}\min_{t\in S_2} a^{(1)}(s,t) \\ &= a^{(1)}(s^*,t^*)\,. \end{aligned}$$

Mithin führen die Paare (s^*,t^*) und $(\widetilde{s},\widetilde{t})$ zu denselben Auszahlungen; egal ob sie gleich sind oder nicht sind sie daher Gleichgewichte. ■

Nun sind freilich nicht alle Spiele Nullsummenspiele, und gerade im Wirtschaftsleben gibt es meist erheblich mehr als nur zwei Spieler. Wie JOHN NASH 1946 im Rahmen seiner Dissertation zeigte, kann die Existenz von Gleichgewichten auch noch unter sehr viel allgemeineren Voraussetzungen sichergestellt werden.



JOHN FORBES NASH (1928–2015) wurde in Bluefield, West Virginia, geboren und ging bereits 1941 auf das dortige College, wo er vor allem Vorlesungen über Mathematik und Naturwissenschaften belegte. Dank eines Stipendiums konnte er 1945 ans Carnegie Institute of Technology in Pittsburgh, PA wechseln, wo er eigentlich Elektrotechnik studieren wollte, dann aber doch 1948 sowohl seinen Bachelor als auch seinen Master in Mathematik machte. Danach vermittelte ihm LEFSCHETZ ein Stipendium der Universität Princeton, wo er sich mit Topologie, algebraischer Geometrie, Spieltheorie und Logik beschäftigte. Als Dissertation legte er 1950 jene Arbeit über die Existenz von Gleichgewichten vor, für die er 1994 den Alfred-Nobel-Gedächtnispreis für Wirtschaftswissenschaften erhalten sollte. Da er sich nicht sicher war, ob diese anerkannt würde, schrieb er gleichzeitig noch ebenfalls sehr fundamentale Arbeit in der er bewies, daß jede

kompakte topologische Mannigfaltigkeit homöomorph (und sogar diffeomorph) zu einer Zusammenhangskomponente einer reell-algebraischen Varietät ist. Auch in der Folgezeit arbeitete er, ab 1952 am MIT, auf beiden Gebieten weiter; dazu kommen unter anderem auch Arbeiten über partielle Differentialgleichungen, für die er beinahe die Fields-Medaille bekommen hätte, sowie Einbettungssätze aus der Differentialgeometrie. 1959 wurden erstmalig Anzeichen einer beginnenden Schizophrenie erkennbar, wegen derer er ab 1961 mehrfach in psychiatrischer Behandlung war und zwischen 1966 und 1996 keine Arbeiten schreiben konnte. Nach dem Wirtschaftsnobelpreis 1994 bekam er 2015 den zweiten nicht von ALFRED NOBEL gestifteten, aber nach ähnlichen Regeln verliehenen Preis, den 2002 von der Norwegischen Akademie der Wissenschaften gestifteten Abelpreis. Bei der Rückkehr von Preisverleihung starb er er bei einem Verkehrsunfall auf dem New Jersey turnpike.

NASH betrachtete für ein n-Personenspiel mit Strategiemenge S_i und Auszahlungsfunktionen $a^{(i)}\colon S_1 \times \cdots \times S_n \to \mathbb{R}$ für den i-ten Spieler zu jedem n-Tupel s von Strategien die Mengen

$$B_i(s) = \left\{ t^* \in S_i \;\middle|\; a^{(i)}(s_{-i}|t^*) = \max_{t \in S_i} a^{(i)}(s_{-i}|t) \right\}$$

der für den i-ten Spieler optimalen Strategien unter der Voraussetzung, daß jeder andere Spieler $j \neq i$ die Strategie s_j wählt. Offensichtlich ist s genau dann ein Gleichgewicht, wenn s für jedes i in $B_i(s)$ liegt. Definieren wir eine Korrespondenz

$$B\colon \begin{cases} S_1 \times \cdots \times S_n \to\to S_1 \times \cdots \times S_n \\ \qquad\qquad s \mapsto B_1(s) \times \cdots \times B_n(s) \end{cases},$$

so ist dies äquivalent dazu, daß $s \in B(s)$ ein Fixpunkt von B ist. Nach dem Fixpunktsatz von KAKUTANI gibt es solche Punkte s zumindest dann, wenn gilt: Das Produkt der S_i ist kompakt und konvex, B ist halbstetig nach oben und alle $B(s)$ sind abgeschlossen und konvex.

Betrachten wir zunächst die klassische Situation, daß jedem Spieler eine endliche Anzahl reiner Strategien zur Verfügung stehen; für den i-ten Spieler seien dies m_i Stück. Dann ist die Menge S_i seiner gemischten Strategien ein $(m_i - 1)$-Simplex, also insbesondere kompakt und konvex; dasselbe gilt dann auch für das Produkt aller dieser S_i.

Die Auszahlungsfunktionen $S_1 \times \cdots \times S_n \to \mathbb{R}$ haben in diesem Fall

die Erwartungswerte

$$a^{(i)}(p^{(1)},\ldots,p^{(n)}) = \sum_{j_1=1}^{m_1}\cdots\sum_{j_n=1}^{m_n}\prod_{i=1}^{n} p_{j_i}^{(i)} a_{j_1\ldots j_n}^{(i)} ,$$

wobei $p^{(i)} \in S_i$ den Wahrscheinlichkeitsvektor für den i-ten Spieler bezeichnet und $a_{i_1\ldots i_n}^{(i)}$ dessen Gewinn, falls jeder Spieler j die i_j-te seiner reinen Strategien wählt. Diese Funktion ist offensichtlich linear in jedem ihrer n Argumente, insbesondere also stetig, und jede der Mengen $B_i(s)$ ist abgeschlossen und konvex, also auch das Produkt $B(s)$ aller $B_i(s)$.

Zur Anwendbarkeit des Fixpunktsatzes von KAKUTANI fehlt somit nur noch, daß die Korrespondenz B halbstetig nach oben ist. Um dies nachzuweisen, verwenden wir die Charakterisierung durch Folgen: Wir müssen zeigen, daß für jede konvergente Folge $(s^{(k)})_{k\in\mathbb{N}}$ von n-Tupeln aus $S_1 \times \cdots \times S_n$ und jede weitere solche Folge $(t^{(k)})_{k\in\mathbb{N}}$ mit $t^{(k)} \in B(s^{(k)})$ für alle k auch der Grenzwert t der zweiten Folge in $B(s)$ liegt, wobei s den Grenzwert der ersten Folge bezeichnet. Nach Definition der Mengen $B_i(u)$ ist für jedes $k \in \mathbb{N}$ und jedes $i = 1,\ldots,n$ der Auszahlungswert an den i-ten Spieler optimal, falls alle anderen Spieler die $s^{(k)}$ entsprechende Strategie wählen, er aber die i-te Komponente von $t^{(k)}$. Wegen der Stetigkeit der Auszahlungsfunktionen gilt dies auch für den Grenzwert, also ist $t \in B(s)$, wie behauptet.

Damit sind alle Voraussetzungen des Fixpunktsatzes von KAKUTANI erfüllt; es gibt daher ein n-Tupel s von Strategien, das in $B(s)$ liegt und damit ein Gleichgewicht ist. Dies zeigt

Satz: Jedes n-Personen-Spiel, bei dem jedem Spieler eine endliche Anzahl reiner Strategien zur Verfügung stehen, hat zumindest in gemischten Strategien ein Gleichgewicht.

■

Heute bezeichnet man solche Gleichgewichte als NASH-Gleichgewichte.

Der Ansatz von NASH zeigt die Existenz von Gleichgewichten freilich auch in sehr viel allgemeineren Situationen. Um die Mengen $B_i(s)$ zu

definieren, brauchen wir nicht einmal eine Auszahlungsfunktionen, sondern es genügt, daß jeder Spieler irgendeine Präferenzstruktur auf dem Produkt aller Strategiemengen hat. Das ist beispielsweise bei manchen Anwendungen in Politik und Wirtschaft von Vorteil, wo die Handelnden zwar oft wissen, welche von zwei Situationen sie der anderen vorziehen, wo es aber oft schwer ist, den Unterschied zwischen den beiden zu quantifizieren. Natürlich müssen auch in solchen Fällen die Voraussetzungen des Satzes von KAKUTANI nachgewiesen werden, bevor man auf die Existenz eines Gleichgewichts schließen kann.

Bevor wir nun übermütig werden und glauben, wir hätten alle Probleme der Spieltheorie gelöst, wollen wir uns einige konkrete Beispiele von NASH-Gleichgewichten anschauen.

Eines der klassischsten, das in praktisch jedem Lehrbuch der Spieltheorie zu finden ist, hat als „Spieler“ zwei Personen, die verdächtigt werden, gemeinsam ein Verbrechen begangen zu haben, das man ihnen allerdings nicht nachweisen kann. In manchen Rechtssystemen ist für solche Fälle eine Kronzeugenregelung vorgesehen: Falls einer der beiden gesteht und bereit ist, im Prozeß gegen den anderen auszusagen, geht er straffrei aus (oder bekommt zumindest eine deutlich reduzierte Strafe).

Nehmen wir an, daß konkret jedem der beiden eine Strafe von zehn Jahren Gefängnis droht. Falls nur einer der beiden gesteht, geht er auf Grund der Kronzeugenregelung straffrei aus, während der andere zehn Jahre bekommt. Falls beide gestehen, braucht man keine Kronzeugen mehr, allerdings gibt es für geständige Straftäter typischerweise etwas Strafnachlaß, so daß jeder mit etwa acht Jahren rechnen kann. Gesteht schließlich keiner, so kann ihnen zwar die Tat nicht nachgewiesen werden, dafür aber einige kleinere Straftaten wie illegaler Waffenbesitz, so daß jeder ein Jahr bekommt. Wie sollten sich die beiden im (natürlich getrennten) Verhör verhalten?

Keiner weiß sicher, ob sich der andere für Zugeben oder Leugnen entschieden hat *bzw.* entscheiden wird; er muß also beide Möglichkeiten in Betracht ziehen. Wenn der andere gesteht, bekommt er acht Jahre, falls auch er gesteht; ansonsten sind es zehn Jahre. Falls der andere leugnet, geht er straffrei aus, wenn er gesteht, und bekommt ein Jahr, wenn er

leugnet. In beiden Fällen fährt er besser, wenn er gesteht; also werden wohl beide gestehen und damit für jeweils acht Jahre ins Gefängnis wandern.

Je nachdem, um was für eine Straftat es sich handelte, wird dies aus Sicht der Allgemeinheit wahrscheinlich eine gute Lösung sein, nicht aber aus Sicht der beiden Betroffenen: Hätten beide geleugnet, wären beide besser gefahren. Auch solche Situationen untersucht die Spieltheorie:

Definition: Ein n-Tupel $(s_1, \ldots, s_n)$ in einem n-Personen-Spiel heißt PARETO-*optimal* oder (besser) PARETO-*effizient,* wenn für jedes Tupel $(t_1, \ldots, t_n)$, bei dem (mindestens) einer der Spieler ein besseres Ergebnis erzielt, (mindestens) ein Spieler ein schlechteres Ergebnis erzielt.

Im obigen Beispiel, dem sogenannten *Dilemma des Gefangenen* oder *Prisoner's Dilemma* ist das NASH-Gleichgewicht (Zugeben, Zugeben) offensichtlich nicht PARETO-effizient: Mit der Strategie (Leugnen, Leugnen) würden beide wesentlich besser fahren. Letztere Strategiepaarung ist offensichtlich PARETO-effizient, allerdings gilt dies auch für (Leugnen, Zugeben) und für (Zugeben, Leugnen). PARETO-Effizienz ist somit kein Kriterium, mit dem man bestmögliche Strategien finden kann; sie liefert eher eine ziemlich schwache Minimalforderung an jede „gute" Lösung. Insbesondere ist bei einem Spiel, bei dem eine gegebene Menge an Ressourcen an die Spieler verteilt werden soll, jede Strategie PARETO-effizient, denn was ein Spieler bei einer anderen Strategie mehr bekommt, muß anderen Spielern weggenommen werden.

Trotzdem muß ein NASH-Gleichgewicht nicht einmal diese Mindestvoraussetzung erfüllen, und da zumindest im obigen Beispiel beide durch rationale Überlegungen zu ihrer Strategie kamen, läßt sich das auch nicht ändern. Wie die Spieltheorie zeigt, sind PARETO-effiziente Lösungen nur dann sinnvoll für rational handelnde Spieler, wenn die gleichen Spieler ihr Spiel mehrfach spielen, wobei die Anzahl der Wiederholungen nicht feststehen darf: Ansonsten wäre man schließlich bei der letzten Wiederholung wieder bei der alten Situation, könnte also davon ausgehen, daß jeder den eigenen Vorteil maximiert. Damit gäbe es aber auch bei der vorletzten Wiederholung keinen Anreiz, sich stattdessen am Gemeinwohl zu orientieren, und so weiter.

Situationen wie beim Dilemma des Gefangenen treten auch in der Politik auf. Als Beispiel können wir das Wettrüsten zur Zeit des Kalten Kriegs zwischen der USA und der UdSSR betrachten. Beide Regierungen mußten sich zwischen Auf- und Abrüstung entscheiden.

Hätten beide abgerüstet, hätte das ihre Wirtschaften entlastet und damit gestärkt; der Wohlstand und die Zufriedenheit der Bürger währen gestiegen, ohne daß sich im Verhältnis der beiden Staaten etwas geändert hätte. Die „Auszahlung“ wäre daher sicherlich für beide positiv gewesen wäre; sagen wir zehn auf irgendeiner Skala.

Wenn nur einer abrüstet, der andere aber aufrüstet, so stärkt dies sicherlich die Wirtschaft des abrüstenden Staats; dieser wird nun aber zumindest langfristig dominiert vom anderen, was mit einem großen Prestigeverlust einhergeht. Entsprechend nimmt das Prestige des anderen zu, und dieser kann künftig möglicherweise Forderungen stellen, die ihm auch wirtschaftlich nützen. Da in der Welt der Politik Prestige oft erheblich wichtiger ist als reale Indikatoren, wäre der Gewinn des aufrüstenden Staates auf unserer hypothetischen Skala größer als zehn, sagen wir fünfzehn, während der abrüstende eine Auszahlung von -15 zu erwarten hätte.

Wenn schließlich beide aufrüsten, ändert sich nichts an ihrem Verhältnis, aber die Aufrüstung belastet die Wirtschaft; somit erhalten beide eine Auszahlung von minus fünf auf unserer Skala.

Wieder ist beidseitige Aufrüstung das einzige NASH-Gleichgewicht, und wieder ist es die einzige Strategiepaarung, die nicht PARETO-effizient ist.

Als letztes wollen wir ein Beispiel betrachten, in dem das NASH-Gleichgewicht nicht eindeutig ist und somit nicht einmal als Anleitung zum Handeln dienen kann. Der Mathematiker und Philosoph BETRAND RUSSELL verglich das folgende Spiel *Chicken,* dessen Namen man hier mit *Feigling* übersetzen muß, mit dem Handeln der am atomaren Wettrüsten beteiligten Politiker:

Zwei Spieler fahren auf der Mittellinie einer geraden Straße mit hoher Geschwindigkeit aufeinander zu. Wer zuerst ausweicht hat verloren.

Auch wenn es sich hier um einen typischen Fall handelt, bei dem die Präferenzen der einzelnen Spieler zwar klar sind, sich aber nur schwer quantifizieren lassen, können wir der Einfachheit halber von zwei Auszahlungsfunktionen ausgehen: Weicht einer aus, und der andere fährt geradeaus weiter, hat der Ausweichende verloren und erhält somit die Auszahlung -1; sein Gegner gewinnt und erhält $+1$. Weicht keiner aus, rasen sie aufeinander, was wahrscheinlich keiner überlebt und was wir mit einer Auszahlung von -100 für beide modellieren. Weichen schließlich beide aus (was wohl gleichzeitig geschehen muß, denn sobald einer sieht, das der andere ausweicht, hat er dazu keinen Grund mehr), geht das Spiel unentschieden aus; der Auszahlungsbetrag für beide ist Null.

Hier gibt es zwei NASH-Gleichgewichte, nämlich die beiden, bei denen einer ausweicht und der andere nicht. Bei beiden Alternativen stellt sich jeder Spieler schlechter, wenn er seine Strategie ändert, während der andere die seinige beibehält. Eine Handlungsempfehlung läßt sich daraus nicht ableiten.

Kapitel 9
Wirtschaftliche Gleichgewichte

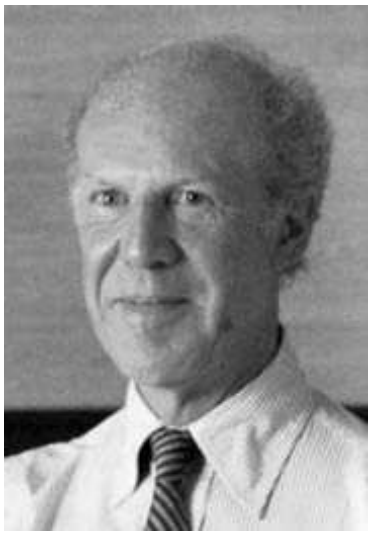
In diesem Kapitel soll es darum gehen, unter möglichst schwachen Bedingungen die Existenz von Gleichgewichten in einem Wirtschaftssystem zu zeigen. Da deren Existenz bereits gegen Ende des 19. Jahrhunderts von LEON WALRAS postuliert wurde, spricht man in der Ökonomie von WALRASschen Gleichgewichten; einen mathematisch sauberen formalen Beweis für ihre Existenz gaben aber erst KENNETH J. ARROW und GERARD DEBREU in ihrer 1954 erschienenen Arbeit

> KENNETH J. ARROW, GERARD DEBREU: Existence of an equilibrium for a competitive economy, *Econometrica* **22** (1954), 265–290

Dieses Kapitel folgt weitgehend der Darstellung in dieser Arbeit.

KENNETH JOSEPH ARROW (1921–2017) wurde in New York geboren und studiere am dortigen City College bis 1940 zum Bachelor in Sozialwissenschaften mit Schwerpunkt Mathematik. Seinen Master in Mathematik bekam er 1941 von der Columbia University in New York, danach wechselte er für seine Promotion an die Fakultät für Wirtschaftswissenschaften. 1942–1946 war er Offizier beim Wetterdienst der amerikanischen Armee. Danach kehrte er zur Fortsetzung seines Studiums zurück nach Columbia, arbeitete aber gleichzeitig auch als *research associate* für die Cowles Foundation an der Universität von Chicago. 1949, nach Abschluß seines Studiums, arbeitete er bis 1968 als Professor für Wirtschaft, Statistik und später auch Operations Research an der Stanford University, danach wechselte er nach Harvard, kehrte aber 1979 nach Stanford zurück, wo er 1991 emeritierte. Bekannt wurde er zunächst durch das in seiner Dissertation veröffentlichte ARROW-Paradoxon, wonach es kein Wahlverfahren

geben kann, das auch nur den einfachsten Anforderungen an Gerechtigkeit genügt. Für seine Arbeiten über Gleichgewichte bekam er 1972 den Wirtschaftsnobelpreis. Schon vorher wechselte sein Interesse zu Fragen der Rolle von Information in der Ökonomie und des Einflusses von Lernen auf das Wirtschaftswachstum.



GERARD DEBREU (1921–2004) wurde in Calais geboren und ging auch dort zur Schule. 1939, als er sein Abitur machte, begann der zweite Weltkrieg, und statt in Paris begann er sein Studium in Ambert, Puy-de-Dôme und ab 1940 in Grenoble, beides im unbesetzten Frankreich. 1941 bestand er die Aufnahmeprüfung der École Normale Supérieure in Paris und studierte dort bis 1944, als er sich freiwillig zur französischen Armee meldete und sowohl in Algerien als auch in Deutschland eingesetzt wurde. 1945 kehrte er nach Paris zurück und beendete sein Studium mit der Agrégation in Mathematik. Danach wechselten seine Interessen zu wirtschaftswissenschaftlichen Fragestellungen, insbesondere zu Gleichgewichtsproblemen. Dank eines Stipendiums konnte er sich unter anderem in Salzburg, Harvard, Berkeley, Chicago, Columbia, Uppsala und Oslo damit beschäftigen. 1949 wurde er *research associate* der Cowles Foundation, 1950 bekam er eine Stelle an der Universität Chicago, 1955 in Yale, 1960 in Stanford und schließlich 1962 in Berkeley, wo er auch noch nach seiner Emeritierung blieb. 1983 bekam er den Wirtschaftsnobelpreis. Außer der Arbeit mit ARROW schrieb er noch zahlreiche weitere Arbeiten über allgemeinere Gleichgewichtssätze, Eigenschaften von Gleichgewichten, und so weiter.

§1: Wirtschaftssysteme

Als erstes müssen wir ein mathematisches Modell für eine Volkswirtschaft definieren. Da die Ökonomie eine Realwissenschaft ist, wird dies nicht ohne Vereinfachungen und Abstraktionen möglich sein; wir gehen aus von den folgenden Annahmen:

Im betrachteten Wirtschaftssystem gibt es ℓ Güter, die wir einfach durch die natürlichen Zahlen $h = 1, \ldots, \ell$ bezeichnen. Zu diesen Gütern gehören auch Dienstleistungen, insbesondere also auch Arbeit. Alle Güter sind quantifizierbar durch reelle Zahlen. Besitz, Produktion und Verbrauch eines Teilnehmers am Wirtschaftssystem lassen sich somit beschreiben durch Vektoren aus $\mathbb{R}^\ell$, deren h-te Komponente jeweils die Menge des h-ten Guts angibt. Zum Vergleich solcher Vektoren verwenden wir folgende Konvention:

Definition: *a)* Für zwei Vektoren $x, y \in \mathbb{R}^\ell$ ist $x \geq y$ genau dann, wenn $x_h \geq y_h$ für alle $h = 1, \ldots, \ell$.
b) $x > y$ genau dann, wenn $x_h > y_h$ für alle $h = 1, \ldots, \ell$.
c) $\Omega \underset{\text{def}}{=} \{x \in \mathbb{R}^\ell \mid x \geq 0\}$.
d) Für $A \subseteq \mathbb{R}^\ell$ ist $-A = \{-x \mid x \in A\}$.
e) Für ν Teilmengen $A_\iota \subset \mathbb{R}^\ell$, $\iota = 1, \ldots, \nu$ ist

$$\sum_{\iota=1}^{\nu} A_\iota \underset{\text{def}}{=} \left\{ x \in \mathbb{R}^\ell \;\middle|\; x = \sum_{\iota=1}^{\nu} x_\iota \text{ für irgendwelche } x_\iota \in A_\iota \right\}.$$

a) Die Produzenten

Die am einfachsten zu beschreibenden Akteure einer Volkswirtschaft sind die *Produzenten.* Ihre Anzahl bezeichnen wir mit m, sie selbst mit den natürlichen Zahlen $j = 1, \ldots, m$. Jedem Produzenten j steht eine Menge $Y_j \subset \mathbb{R}^\ell$ von Produktionsplänen y_j zur Verfügung. Die h-te Komponente eines solchen Vektors y_j ist positiv, wenn nach diesem Produktionsplan die entsprechende Menge des Guts h produziert wird; sie ist negativ, wenn eine entsprechende Menge von h bei der Produktion eingesetzt wurde. Zu solchen Gütern h gehören nicht nur die verwendeten Rohstoffe und extern produzierte Komponenten, sondern insbesondere auch die eingesetzte Arbeit. Die Summenmenge

$$Y \underset{\text{def}}{=} \sum_{j=1}^{m} Y_j$$

ist die Menge aller in der betrachteten Volkswirtschaft realisierbarer Produktionspläne; an sie sowie die einzelnen Mengen Y_j stellen wir die folgenden Forderungen:

I.a Y_j ist eine abgeschlossene konvexe Teilmenge von $\mathbb{R}^\ell$, die den Nullvektor 0 enthält.

I.b $Y \cap \Omega = \{0\}$

I.c $Y \cap (-Y) = \{0\}$

Bedingung I.a besagt insbesondere, daß die Produktion um einen konstanten Faktor zurückgefahren werden kann, denn da der Nullvektor in jedem Y_j liegt, ist für $\lambda \in (0, 1)$ und $y_j \in Y_j$ auch

$$\lambda y_j = \lambda y_j + (1 - \lambda)0 \in Y_j .$$

Die Bedingung $0 \in Y_j$ ermöglicht es auch, daß sich ein Produzent ganz aus der Produktion zurückzieht.

Mit der Konvexitätsforderung machen wir natürlich, wie in den meisten mathematischen Modellen üblich, die Annahme, daß die reellen Zahlen etwas Reales seien. Ein Produzent kann einen Produktionsplan haben, der die Produktion von 314 PKWs pro Arbeitstag vorsieht, wofür er von einem Zulieferer 1256 Räder bezieht, er könnte aber auch 315 produzieren und dafür 1260 Räder verbrauchen. Nimmt man die Konvexitätsforderung wörtlich, müßte er dann auch 100π PKW produzieren können bei einem Einsatz von 400π Rädern. Trotzdem erweist sich eine solche Idealisierung, falls es nicht um sehr kleine Mengen geht, üblicherweise als nützlich, da die sonstigen Fehler, die in Annahmen und Modellen stecken, meist deutlich größere Auswirkungen haben.

Bedingung I.b sagt, daß eine Volkswirtschaft nichts produzieren kann, ohne dabei Ressourcen wie Rohstoffe und/oder Arbeit zu verbrauchen. Bedingung I.c schließlich besagt, daß ein Produktionsvorgang nicht rückgängig gemacht werden kann: Bezüglich der eingesetzten materiellen Güter mag das zwar eventuell möglich sein; ein perfektes Recycling strebt dies jedenfalls an. Die postulierte Asymmetrie kommt aber von der Arbeit: Sowohl für die Montage wie auch für die Demontage muß Arbeit aufgewendet werden. Ein Produktionsschritt kann zwar eventuell rückgängig gemacht werden, aber dabei wird keine alte Arbeit frei zum Einsatz für neue Produktionsschritte.

Aus Sicht eines Produzenten sind die verschiedenen Produktionspläne, die ihm zur Verfügung stehen, nicht alle gleich erstrebenswerts:; Er will (und muß) durch seine Aktivität etwas verdienen. Wir nehmen an, daß er seinen Gewinn maximieren möchte. Dieser hängt natürlich davon ab, zu welchem Preis er seine Produkte verkaufen kann, und welchen Preis er für die eingesetzten Ressourcen bezahlen muß.

Nach der reinen marktwirtschaftlichen Lehre sollte der Markt die Preise der einzelnen Güter festlegen. Wir nehmen an, daß die Marktentwicklung zu einem Preisvektor $p^* \in \mathbb{R}^\ell$ für die Preise der verschiedenen Güter führte. Der Gewinn oder Verlust der j-ten Produzenten ist dann bei Wahl des Produktionsplans y_j das Skalarprodukt $\langle p^*, y_j \rangle$ des Preisvek-

tors mit dem Produktionsplan. Gewinnmaximierung (oder Verlustminimierung) bedeutet dann, daß der Produzent dieses Skalarprodukt maximieren möchte. Seine Strategie ist demnach die folgende:

(1) *Wähle einen Produktionsplan* $y_j^* \in Y_j$ *so, daß das Skalarprodukt* $\langle p^*, y_j^* \rangle$ *das Maximum unter allen* $\langle p^*, y_j \rangle$ *mit* $y_j \in Y_j$ *annimmt.*

Diese Strategie wird in unserem Modell von allen m Produzenten verfolgt.

b) Die Konsumenten

Neben den m Produzenten gibt es auch n Konsumenten; wir bezeichnen sie durch die natürlichen Zahlen $i = 1, \ldots, n$. Jeder von ihnen hat gewisse Grundbedürfnisse, die er durch den Verbrauch von Gütern befriedigen muß. Wir gehen davon aus, daß ihm dazu eine gewisse Menge $X_i \subset \mathbb{R}^\ell$ von möglichen Verbrauchsplänen zur Verfügung steht. Für jedes $x_i \in X_i$ stehen die positiven Komponenten für konsumierte Güter und in Anspruch genommene Arbeit, die negativen für zu leistende Arbeit. Wir nehmen an:

II. Die Mengen X_i sind abgeschlossen, konvex und nach unten beschränkt. Es gibt also für jeden Verbraucher i einen (nicht notwendigerweise in X_i liegenden) Vektor $\xi_i \in \mathbb{R}^\ell$, so daß $x_i \geq \xi_i$ für alle $x_i \in X_i$.

Die Beschränktheit nach unten sorgt nicht nur dafür, daß gewisse Grundbedürfnisse befriedigt werden, sie begrenzt auch jede mögliche Arbeitsleistung. Die Abgeschlossenheit von X_i ist eine Art Stetigkeitsbedingung, und die Konvexität zeigt insbesondere die Substituierbarkeit von Gütern, die auf dem Markt gegen andere getauscht werden können.

Ein Produzent bestimmt seinen Produktionsvektor so, daß er einen maximalen Gewinn erzielt. Beim Konsumenten ist die Situation komplizierter: Wenn es ihm nur um Kostenoptimierung ginge, müßte er sich auf die Befriedigung seiner Grundbedürfnisse beschränken und müßte arbeiten soviel er kann. Für die wenigsten Konsumenten dürfte so etwas erstrebenswert sein. Wir gehen daher davon aus, daß jeder Verbraucher

eine *Nützlichkeitsfunktion* $u_i\colon X_i \to \mathbb{R}$ hat, die jedem möglichen Verbrauchsvektor eine reelle Zahl zuordnet, die umso größer ist, je mehr ihm der betreffende Verbrauchsvektor zusagt. Wir nehmen an, daß gilt

III.a u_i ist eine stetige Funktion

III.b Für jedes $x_i \in X_i$ gibt es ein $x_i' \in X_i$ mit $u_i(x_i') > u_i(x_i)$

III.c Ist $u_i(x_i') > u_i(x_i)$, so gilt auch für jedes $\lambda \in (0,\ 1)$ die Ungleichung $u_i\big((\lambda x_i' + (1-\lambda)x_i\big) > u_i(x_i)$.

Bedingung III.a besagt insbesondere, daß die Menge aller Verbrauchsvektoren, deren Nützlichkeit aus Sicht eines gegebenen Verbrauchers größer oder gleich einem vorgegebenen Wert ist, abgeschlossen ist. Durch reine Präferenzen ausgedrückt heißt dies: Ist $(x_i^{(k)})_{k\in\mathbb{N}}$ eine konvergente Folge von Verbrauchsvektoren $x_i^{(k)} \in X_i$ derart, daß Verbraucher i jedes $x_i^{(k)}$ mindestens genauso sehr schätzt wie eine Alternative $\tilde{x}_i$, so schätzt er auch den Grenzwert mindestens genauso sehr wie $\tilde{x}_i$.

Bedingung III.b formalisiert eine wohlbekannte Tatsache: Egal, wie gut es uns geht: Wir können uns immer einen Zustand vorstellen, der uns noch lieber wäre, weil der entsprechende Vektor entweder mehr von einem unserer Lieblingsgüter enthält oder ein Gut, das wir noch nicht haben, oder was auch immer.

Bedingung III.c ist insbesondere eine Konvexitätsbedingung: Sie impliziert, daß für jede reelle Zahl a die Menge

$$M = \big\{x_i \in X_i \;\big|\; u_i(x_i) \geq a\big\}$$

aller Aktionen, deren Nützlichkeit als größer oder gleich a eingeschätzt wird, konvex ist: Sind nämlich x_i und x_i' zwei Vektoren aus M, so können wir o.B.d.A. annehmen, daß $u_i(x_i') \geq u_i(x_i)$ ist. Falls sogar $u_i(x_i') > u_i(x_i)$ ist, sagt uns III.c direkt, daß auch $u_i\big(\lambda x_i + (1-\lambda)x_i'\big)$ für alle $\lambda \in [0,\ 1)$ größer ist als $u_i(x_i)$. Falls $u_i(x_i) = u_i(x_i')$ ist, müssen wir uns anders überlegen, daß es kein $x_i^* = \lambda x_i + (1-\lambda)x_i'$ gibt mit $u_i(x_i^*) < u_i(x_i)$. Gäbe es so ein x_i^*, so gäbe es wegen der Stetigkeit von u_i einen Vektor x_i^+ mit $u_i(x_i^*) < u_i(x_i^+) < u_i(x_i) = u_i(x_i')$ auf der Verbindungsstrecke von x_i^* zu x_i. Auf der Strecke von x_i nach x_i' kommen wir somit zunächst zu x_i^*, dann zu x_i^+ und schließlich zu x_i'. Daher

läßt sich x_i^* als konvexe Linearkombination von x_i und x_i^+ ausdrücken. Da $u_i(x_i) > u_i(x_i^+)$ ist, muß daher nach III.c auch $u_i(x_i^*) > u_i(x_i)$ sein, im Widerspruch zur Annahme. Somit ist $u_i(x_i^*) \geq u_i(x_i) \geq a$ für alle $x_i^* = \lambda x_i + (1-\lambda)x_i'$ zwischen x_i und x_i'.

Leider kann sich unser Verbrauch von Konsumgütern nicht nur an unseren Präferenzen orientieren, sondern muß auch unsere Mittel in Betracht ziehen. Um dies zu modellieren, betrachten ARROW und DEBREU zwei Arten von Einkünften: Zum einen verfügt Verbraucher i zu Beginn über eine gewisse Menge von Gütern; die h-te Komponente des Vektors $\zeta_i \in \mathbb{R}^\ell$ gebe deren Menge für das Gut h an. Außerdem können Verbraucher Anteile an Produzenten besitzen und damit einen Anspruch auf einen entsprechenden Anteil an deren Gewinnen. α_{ij} gebe an, welchen Anteil Verbraucher i am Produzenten j besitzt. Wir nehmen an, daß gilt

IV.a Es gibt ein $x_i \in X_i$ mit $x_i < \zeta_i$.

IV.b $\alpha_{ij} \geq 0$ für alle i, j und $\sum_{i=1}^n \alpha_{ij} = 1$.

Bedingung IV.a besagt, daß es einen möglichen Verbrauchsvektor gibt, dessen sämtliche Komponenten echt kleiner sind als die entsprechenden Komponenten von ζ_i. Wir nehmen also an, daß jeder Verbraucher von *jedem* Gut mehr besitzt als er gemäß zumindest einem seiner möglichen Verbrauchsvektoren braucht, und daß er damit *jedes* Gut des betrachteten Wirtschaftssystems gegebenenfalls zum Verkauf anbieten könnte. Diese Voraussetzung ist sicherlich unrealistisch, sie vereinfacht aber den Beweis der Existenz eines Gleichgewichts beträchtlich. ARROW und DEBREU betrachten deshalb zunächst den Fall eines Wirtschaftssystem, in dem diese Bedingung erfüllt ist, und diskutieren danach die notwendigen Modifikationen, um IV.a durch realistischere Bedingungen zu ersetzen.

Bedingung IV.b ist weitgehend unproblematisch: Die Summenbedingung besagt einfach, daß jeder Anteilseigner eines Produzenten als Verbraucher auftritt, und $\alpha_{ij} \geq 0$ verbietet den Verkauf von Anteilen, die man gar nicht besitzt.

Ein Verbraucher kann nur Güter erwerben, die er sich zu den jeweils geltenden Preisen auch leisten kann. Sein Einkommen besteht aus dem Wert der Güter (einschließlich Arbeit), die er auf dem Markt anbieten kann, sowie Gewinnbeteiligungen bei den Produzenten, von denen

er Anteile besitzt. Er wählt seinen Verbrauchsvektor x_i^* so, daß die Nützlichkeitsfunktion maximiert wird:

(2) $u_i(x_i^*) = \max\left\{u_i(x_i) \;\middle|\; \langle p^*, x_i\rangle \leq \langle p^*, \zeta_i\rangle + \sum_{j=1}^{m} \alpha_{ij} \left\langle p^*, y_j\right\rangle\right\}$

Bevor wir definieren können, was ein Gleichgewicht ist, müssen wir uns noch mit dem Preisvektor p^* beschäftigen. Wir gehen davon aus, daß er einfach die relativen Werte der einzelnen Güter angibt, wobei jedes Gut einen nichtnegativen Wert haben soll. Diese Eigenschaft des Preisvektors können wir formalisieren durch die Bedingung

(3) $p^* \in P \underset{\text{def}}{=} \left\{p \in \mathbb{R}^\ell \mid p \geq 0 \text{ und } \sum_{h=1}^{\ell} p_h = 1\right\}$.

Als letzte Forderung an ein Gleichgewicht, daß seinen Namen verdient, müssen wir noch in irgendeiner Weise formalisieren, daß sich Angebot und Nachfrage für jedes Gut die Waage halten – oder zumindest für fast jedes. Eine Ausnahme bilden etwaige frei erhältliche Güter wie beispielsweise die Nutzung natürlicher Ressourcen oder von Funkfrequenzen, für die je nach Wirtschaftssystem innerhalb gewisser Grenzen möglicherweise ein Preis von Null angesetzt werden kann. Bei solchen Gütern steht zu erwarten, daß sie in genügender Menge vorhanden sind und auch im Gleichgewicht nicht vollständig genutzt werden.

Um dies zu formalisieren, betrachten wir die Summe x aller Verbrauchsvektoren als Vektor der Gesamtnachfrage, die Summe y aller Produktionsvektoren als Vektor des Gesamtangebots und schließlich die Summe ζ aller ζ_i als Vektor der sich im Besitz der Verbraucher befindlichen Güter:

$$x = \sum_{i=1}^{n} x_i, \qquad y = \sum_{j=1}^{m} y_j. \qquad \zeta = \sum_{i=1}^{n} \zeta_i\,.$$

Aus diesen drei Vektoren können wir dann den Nachfrageüberhang

$$z \underset{\text{def}}{=} x - y - \zeta \in \mathbb{R}^\ell$$

berechnen, von dem wir für Gleichgewichtspunkte fordern:

(4) $z^* \leq 0$ und $\langle p^*, z^*\rangle = 0$.

Im Gleichgewicht soll also die Nachfrage nach jedem der ℓ Güter befriedigt werden. Die Bedingung $\langle p^*, z^* \rangle = 0$ impliziert außerdem, daß alle Güter, bei denen das Angebot über der Nachfrage liegt, Wert Null haben: Da für jedes Gut h der Preis $p_h \geq 0$ ist, während der Nachfrageüberhang $z_h \leq 0$ ist, stehen im Skalarprodukt $\langle p^*, z^* \rangle$ nur nichtpositive Summanden, eine Summe Null ist daher nur möglich, wenn jeder einzelne davon verschwindet. Somit ist $p_h z_h = 0$ für jedes Gut h, im Falle $z_h \neq 0$ muß also $p_h = 0$ sein. Ein Überangebot kann es daher nur bei freien Gütern geben.

Definition: Ein $(m+n+1)$-Tupel $(x_1^*, \ldots, x_n^*, y_1^*, \ldots, y_m^*, p^*)$ von Vektoren aus $\mathbb{R}^\ell$ heißt *Gleichgewicht*, wenn es die Bedingungen (1) – (4) erfüllt.

Satz 1: Unter den Bedingungen I bis IV gibt es ein solches Gleichgewicht.

Der Beweis dieses Satzes sollte ähnlich funktionieren wie der für die Existenz von NASH-Gleichgewichten bei Spielen, allerdings unterscheidet sich die Situation hier in einem Punkt ganz wesentlich von der bei Spielen: Dort steht jedem Spieler eine feste Menge von Strategien zur Verfügung, hier hängt aber zumindest für die Verbraucher die Menge der spielbaren Strategien vom Preisvektor und den Strategien der Produzenten ab. Auch die Auszahlungsfunktion hängt nicht nur ab von den Strategien der Verbraucher und der Produzenten, sondern zusätzlich auch noch vom Preisvektor. Letzteres Problem werden wir dadurch lösen, daß wir einen zusätzlichen Spieler einführen, der die Preise festsetzt. Das Problem mit den variablen Strategiemengen kann nicht so einfach gelöst werden: Hierzu muß der Begriff des Spiels verallgemeinert werden.

§2: Abstrakte Ökonomien

In seiner Arbeit

> GERARD DEBREU: A social equilibrium existence theorem, *Proc. N.A.S.* **38** (1952), 886–893

führte DEBREU 1952 den Begriff einer *abstrakten Ökonomie* ein als Verallgemeinerung des Spielbegriffs und bewies die Existenz von Gleich-

gewichten für abstrakte Ökonomien. Er verwendete dazu nicht den Fixpunktsatz von KAKUTANI, sondern zwei Verschärfungen, die EILENBERG und MONTGOMERY 1946 *bzw.* BEGLE 1950 publiziert haben. Da in der Arbeit von ARROW und DEBREU nur ein schwächeres Resultat benutzt wird, zu dessen Beweis der Fixpunktsatz von KAKUTANI ausreicht, werde ich mich hier auf dieses schwächere Resultat beschränken.

Eine *abstrakte Ökonomie* hat ν *Agenten* $\iota = 1, \ldots, \nu$. Jedem steht eine Menge $\mathcal{A}_\iota$ eventuell möglicher Aktionen zur Verfügung, von denen er jedoch in Abhängigkeit von den Aktionen der anderen Agenten nur die aus einer Teilmenge wirklich ausführen kann: Er kann beispielsweise von keinem Gut mehr konsumieren, als in der gesamten Ökonomie produziert wird.

Wie in der Spieltheorie betrachten wir das Produkt $\mathcal{A} \underset{\text{def}}{=} \mathcal{A}_1 \times \cdots \times \mathcal{A}_\nu$ aller $\mathcal{A}_\iota$, und auch die Bezeichnungen $\mathcal{A}_{-\iota}$, $a_{-\iota}$ *usw.* seien wie dort definiert. Für jeden Agenten ι haben wir dann eine Korrespondenz $A_\iota\colon \mathcal{A}_{-\iota} \to\to \mathcal{A}_\iota$, und für jede Wahl $a_{-\iota} \in \mathcal{A}_{-\iota}$ kann ι nur die Aktionen aus $A_\iota(a_{-\iota}) \subseteq \mathcal{A}_\iota$ ausführen. Seine Auszahlungsfunktion ist eine Funktion $f_\iota\colon \mathcal{A} \to \mathbb{R}$, wobei natürlich nur die Funktionswerte an jenen Punkten $(a_1, \ldots, a_\nu)$ relevant sind, für die a_ι in $A_\iota(a_{-\iota})$ liegt. Abgesehen von dieser Einschränkung sind Gleichgewichte genauso definiert wie das NASH-Gleichgewicht, d.h. keiner der Agenten kann seine Auszahlung verbessern, wenn alle anderen bei ihrer Wahl bleiben und nur er eine andere vor diesem Hintergrund mögliche Aktion wählt:

Definition: Ein Gleichgewichtspunkt einer abstrakten Ökonomie ist ein ν-Tupel $(a_1^*, \ldots, a_\nu^*) \in \mathcal{A} = \mathcal{A}_1 \times \cdots \times \mathcal{A}_\nu$ derart, daß für alle ι gilt:

1. $a_\iota^* \in A_\iota(a_{-\iota}^*)$ und
2. $f_\iota(a_1^*, \ldots, a_\nu^*) = \max\left\{ f_\iota(a_{-\iota}^* | a_\iota) \;\middle|\; a_\iota \in A_\iota(a_{-\iota}^*) \right\}$.

Satz von Debreu: Alle Mengen $\mathcal{A}_\iota$ seien kompakte konvexe Polyeder, und alle Korrespondenzen A_ι, alle Auszahlungsfunktionen f_ι sowie auch alle Funktionen

$$\varphi_\iota\colon \begin{cases} \mathcal{A}_{-\iota} \to \mathbb{R} \\ a_{-\iota} \mapsto \max\left\{ f_\iota(a_{-\iota} | a_\iota) \;\middle|\; a_\iota \in A_\iota(a_{-\iota}) \right\} \end{cases};$$

seien stetig. Außerdem seien für alle ι und alle $a_{-\iota} \in \mathcal{A}_{-\iota}$ die Mengen

$$M_{a_{-\iota}} = \left\{ a_\iota \in A_\iota(a_{-\iota}) \;\middle|\; f_\iota(a_{-\iota}|a_\iota) = \varphi_\iota(a_{-\iota}) \right\}$$

konvex. Dann hat die oben definierte abstrakte Ökonomie ein Gleichgewicht.

Man beachte, daß die Korrespondenzen A_ι als *stetig* vorausgesetzt sind, sie müssen also halbstetig sowohl nach oben als auch nach unten sein.

Beweis: $(a_1, \ldots, a_\nu) \in \mathcal{A}$ ist genau dann ein Gleichgewicht, wenn a_ι für jedes ι in $M_{a_{-\iota}}$ liegt, wenn $(a_1, \ldots, a_\nu)$ also ein Fixpunkt der Korrespondenz

$$\Phi\colon \begin{cases} \mathcal{A}_1 \times \cdots \times \mathcal{A}_\nu \to\to \mathcal{A}_1 \times \cdots \times \mathcal{A}_\nu \\ \quad (a_1, \ldots, a_\nu) \quad \mapsto M_{a_{-1}} \times \cdots \times M_{a_{-\nu}} \end{cases}$$

ist. Mit den Mengen $\mathcal{A}_\iota$ ist auch deren Produkt kompakt und konvexs. Um den Fixpunktsatz von KAKUTANI anwenden zu können, müssen wir daher nur zeigen, daß Φ halbstetig nach oben ist.

Dazu beweisen wir zunächst eine etwas allgemeinere Aussage:

Satz: *X und Y seien Teilmengen endlichdimensionaler reeller Vektorräume, $f\colon X \times Y \to \mathbb{R}$ sei eine stetige Abbildung, und $\varphi\colon X \to\to Y$ sei eine stetige Korrespondenz. Dann ist die Korrespondenz*

$$\mu\colon \begin{cases} X \to\to Y \\ \quad x \mapsto \left\{ y \in \varphi(x) \;\middle|\; f(x,y) = \max\limits_{z \in \varphi(x)} f(x,z) \right\} \end{cases}$$

halbstetig nach oben.

Zum Beweis verwenden wir die Charakterisierungen der jeweiligen Stetigkeitsbegriffe durch Folgen: Sei also $(x_n)_{n\in\mathbb{N}}$ eine konvergente Folge in X mit Grenzwert x und $(y_n)_{n\in\mathbb{N}}$ ein konvergente Folge in Y mit Grenzwert y derart, daß $y_n \in \mu(x_n)$ für alle n. Da φ halbstetig nach oben ist, liegt dann auch y in $\varphi(x)$. Für jeden anderen Punkt $z \in \varphi(x)$ gibt es wegen der Halbstetigkeit von φ nach unten eine Folge $(z_n)_{n\in\mathbb{N}}$ mit Grenzwert z, so daß $z_n \in \varphi(x_n)$ für alle n. Nach Definition von μ ist $f(x_n, y_n) \geq f(x_n, z_n)$ für alle n; wegen der Stetigkeit von f ist somit

auch $f(x,y) \geq f(x,z)$. Da z beliebig war, muß daher y in $\mu(x)$ liegen, was die Behauptung beweist.

■

Um im Beweis des Satzes von DEBREU fortzufahren, wenden wir dies an auf $X = \mathcal{A}_{-\iota}$ und $Y = \mathcal{A}_\iota$. Für die Abbildung $f\colon X \times Y \to \mathbb{R}$ nehmen wir die Auszahlungsfunktion des Agenten ι, und $\varphi(a_{-\iota})$ sei $A_\iota(a_{-\iota}) \subseteq \mathcal{A}_\iota$. Dann ist

$$\mu(a_{-\iota}) = \left\{a_\iota \in A_\iota(a_{-\iota}) \;\middle|\; f_\iota(a_{-\iota}|a_\iota) = \max_{a \in A_\iota(a_{-\iota})} f(a_{-\iota}|a)\right\} = M_{a_{-\iota}},$$

die Korrespondenz von $\mathcal{A}_{-\iota}$ nach $i\mathcal{A}_\iota$, die jedem $a_{-\iota}$ die Menge $M_{a_{-\iota}}$ zuordnet, ist also halbstetig nach oben. Natürlich ist die Projektion $\mathcal{A} \to \mathcal{A}_{-\iota}$ eine stetige Abbildung, und damit ist auch die Korrespondenz

$$\mu_\iota\colon \begin{cases} \mathcal{A}_1 \times \cdots \times \mathcal{A}_\nu \to\to \mathcal{A}_\iota \\ \quad (a_1, \ldots, a_\nu) \quad \mapsto M_{a_{-\iota}} \end{cases}$$

halbstetig nach oben. Die Korrespondenz Φ ist gerade das Produkt der μ_i, also ist auch sie halbstetig nach oben und hat somit nach KAKUTANI einen Fixpunkt. Damit ist der Satz von DEBREU bewiesen.

■

§3: Beweis von Satz 1

Sowohl die möglichen Strategiemengen der Verbraucher als auch die Gewinne der Produzenten hängen ab vom Preisgefüge, das nach der klassischen volkswirtschaftlichen Lehre vom Markt bestimmt wird. ARROW und DEBREU führen daher zusätzlich zu den n Verbrauchern und den m Produzenten einen weiteren Agenten ein, den *Marktteilnehmer* oder kurz *Markt.* Wir haben somit eine abstrakte Ökonomie mit $m+n+1$ Teilnehmern.

Die Strategiemenge des neuen Agenten ist die Menge

$$P = \left\{p \in \mathbb{R}^\ell \;\middle|\; p \geq 0 \text{ und } \sum_{h=1}^{\ell} p_h = 1\right\}$$

aller möglicher Preisvektoren. Die Strategiemenge des j-ten Produzenten ist Y_j. In beiden Fällen gehen wir davon aus, daß den Agenten unabhängig von den Entscheidungen der übrigen Teilnehmer alle Strategien aus der jeweiligen Menge zur Verfügung stehen, d.h. für diese Agenten ι ist $A_\iota(a_{-\iota}) = \mathcal{A}_\iota$ für alle $a \in \mathcal{A}$.

Die Strategiemenge des i-ten Verbrauchers ist natürlich X_i, tatsächlich muß er aber einen Verbrauchsplan wählen, den er auch finanzieren kann, d.h. er muß sich beschränken auf jede Vektoren $x_i \in X_i$, für die das Skalarprodukt $\langle p, x_i \rangle$ mit dem Preisvektor p höchstens gleich dem Wert seiner Güter und Beteiligungen ist:

$$\langle p, x_i \rangle \leq \langle p, \zeta_i \rangle + \sum_{j=1}^{m} \alpha_{ij} \left\langle p, y_j \right\rangle .$$

Für einen beliebigen Preisvektor p kann es durchaus sein, daß kein einziges $x_i \in X_i$ diese Bedingung erfüllt: Unter den y_j gibt es schließlich auch negative, und je nach Preisvektor könnte der zweite Summand in obiger Ungleichung negativ sein.

In einem Gleichgewicht kann dies allerdings nicht vorkommen, denn ist p^* der Preisvektor eines Gleichgewichtspunkts und y_j^* der Produktionsplan des j-ten Produzenten in diesem Gleichgewicht, so ist $\left\langle p^*, y_j^* \right\rangle \geq \langle p^*, y \rangle$ für alle $y \in Y_j$. Da dort insbesondere auch der Nullvektor liegt, ist also $\left\langle p^*, y_j^* \right\rangle \geq \langle p^*, 0 \rangle = 0$. Beim Nachweis der Existenz von Gleichgewichten können wir daher dem i-ten Verbraucher alle Verbrauchspläne erlauben, für die

$$\langle p, x_i \rangle \leq \langle p, \zeta_i \rangle + \max\left\{0, \sum_{j=1}^{m} \alpha_{ij} \left\langle p, y_j \right\rangle\right\}$$

ist, und diese Teilmenge von X_i definieren wir als $A_i(x_{-i})$. Sie kann nicht leer sein, denn nach unseren Voraussetzungen gibt es ein $x_i' \in X_i$ mit $\zeta_i \geq x_i'$, so daß zumindest x_i' in $A_i(x_{-i})$ liegt. Der Index „$-i$“ bezieht sich hier natürlich auf *alle* übrigen Agenten, also auch die Produzenten und den Marktteilnehmer.

Die Auszahlungsfunktion eines Verbrauchers ist seine Nützlichkeitsfunktion, die eines Produzenten ist das Skalarprodukt seines Produktionsplans mit dem Preisvektor. Bleibt noch die Auszahlungsfunktion des Marktteilnehmers zu definieren.

Nach unseren Forderungen soll im Gleichgewicht der Nachfrageüberhang $z^* = x^* - y^* - \zeta \leq 0$ sein und den Wert $\langle p^*, z^* \rangle = 0$ haben. Wenn wir die Auszahlungsfunktion für den Marktteilnehmer als $\langle p, z \rangle$

definieren, ist dies in jedem Gleichgewichtspunkt der betrachteten abstrakten Ökonomie der Fall: Wäre nämlich im Gleichgewicht eine der Komponenten $z_h^* > 0$, so könnte der Marktteilnehmer seine Auszahlung vergrößern, indem er die entsprechende Preiskomponente p_h zu Lasten einer anderen Komponente erhöht, im Widerspruch zur Definition eines Gleichgewichts einer abstrakten Ökonomie. Genauso könnte er im Falle $z_h^* < 0$ seine Auszahlung vergrößern, indem er p_h erniedrigt – es sei denn, p_h ist bereits null. Der Marktteilnehmer legt die Preise also nach dem Gesetz von Angebot und Nachfrage fest und erreicht damit, daß im Gleichgewicht $p_h = 0$ ist für alle h mit $z_h < 0$, und $z_h = 0$ für alle Güter h mit $p_h > 0$.

Wenn wir zeigen können, daß die so definierte abstrakte Ökonomie ein Gleichgewicht hat, folgt die Behauptung von Satz 1.

Leider erfüllt diese abstrakte Ökonomie aber nicht die Voraussetzungen des Satzes von DEBREU: Zumindest die Strategiemengen X_i der Verbraucher sind definitiv nicht kompakt, denn wir haben ja vorausgesetzt, daß es zu jedem Verbrauchsplan eine Alternative mit höherer Nützlichkeit gibt. Auf einem Kompaktum müßte die Nützlichkeitsfunktion aber ein (absolutes) Maximum annehmen.

Die Beweisstrategie von ARROW und DEBREU besteht darin, daß sie die abstrakte Ökonomie so modifizieren, daß alle Strategiemengen kompakt sind, daß die ursprüngliche und die modifizierte Ökonomie aber trotzdem dieselben Gleichgewichte haben.

Da die Strategiemenge P des Marktteilnehmers bereits kompakt ist, müssen nur die Strategiemengen der Verbraucher und der Produzenten modifiziert werden. Die Idee ist in beiden Fällen die gleiche: Wir erlauben jedem Teilnehmer nur solche Strategien, für die es zumindest grundsätzlich für jeden anderen Teilnehmer eine Strategie gibt derart, daß der Nachfrageüberhang der gesamten Ökonomie die Bedingung $z \leq 0$ erfüllt. Wir beschränken daher die Wahlmöglichkeiten des i-ten Verbrauchers auf die Menge

$$\widehat{X}_i = \left\{ x_i \in X_i \;\middle|\; \forall i' \neq i \, \exists x_{i'} \in X_{i'} \wedge \forall j \, \exists y_j \in Y_j \wedge \exists p \in P : z \leq 0 \right\}$$

und die des j-ten Produzenten auf

$$\widehat{Y}_j = \left\{ y_j \in Y_j \;\middle|\; \forall i\, \exists x_i \in X_i \wedge \forall j' \neq j\, \exists y_{j'} \in Y_{j'} \wedge \exists p \in P : z \leq 0 \right\} .$$

Wenn die betrachtete Ökonomie ein Gleichgewicht hat, müssen offensichtlich die Strategien aller Teilnehmer in den so definierten Teilmengen liegen. Da sie abgeschlossen sind, reicht es zum Nachweis der Kompaktheit, daß wir ihre Beschränktheit zeigen.

Angenommen, mindestens eine der Mengen $\widehat{Y}_j$ wäre unbeschränkt. Durch Umnummerieren können wir annehmen, daß dies für die Menge $\widehat{Y}_1$ der Fall sei. Dann gibt es eine Folge von Strategien $y_1^{(k)} \in \widehat{Y}_1$ und für alle i sowie alle $j > 1$ Strategien $x_i^{(k)} \in X_i$ und $y_j^{(k)} \in Y_j$, derart, daß gilt:

$$\lim_{k\to\infty} \left\| y_1^{(k)} \right\| = \infty \quad \text{und} \quad \sum_{j=1}^{m} y_j^{(k)} \geq \sum_{i=1}^{n} x_i^{(k)} - \zeta .$$

Als Norm verwenden wir hier und im Folgenden die Maximumsnorm.

Bezeichnet ξ die Summe der Untergrenzen ξ_i für die Strategiemengen X_i der einzelnen Verbraucher, so ist nach Voraussetzung II

$$\sum_{i=1}^{n} x_i^{(k)} \geq \xi \quad \text{und damit} \quad \sum_{j=1}^{m} y_j^{(k)} \geq \xi - \zeta .$$

Wir betrachten für jeden Index k das Maximum $\mu^{(k)}$ der Normen der $y_j^{(k)}$. Da insbesondere $\mu^{(k)} \geq \left\| y_1^{(k)} \right\|$ ist, gehen die $\mu^{(k)}$ gegen unendlich; für hinreichend große Werte von k sind sie also größer als eins. Nach Annahme I.a ist dann für jeden Produzenten j

$$\frac{y_j^{(k)}}{\mu^{(k)}} = \frac{1}{\mu^{(k)}} y_j^{(k)} + \left(1 - \frac{1}{\mu^{(k)}}\right) 0 \in Y_j .$$

Die Summe aller dieser Vektoren für ein festes solches k erfüllt daher die Ungleichung

$$\sum_{j=1}^{m} \frac{y_j^{(k)}}{\mu^{(k)}} \geq \frac{\xi - \zeta}{\mu^{(k)}} .$$

Nach Definition von $\mu^{(k)}$ hat jeder Summand höchstens die Norm eins.

Da in einem Kompaktum jede Folge eine konvergente Teilfolge hat, gibt es für jedes j eine Folge $(k_q)_{q\in\mathbb{N}}$ von Indizes, so daß die Folge der $y_j^{(k_q)}/\mu^{(k_q)}$ gegen einen Grenzwert $y_j^{(0)}$ konvergiert. Wegen der vorausgesetzten Abgeschlossenheit von Y_j liegt dort auch dieser Grenzwert, und die Summe über alle diese Grenzwerte muß wegen obiger Ungleichung und der bestimmten Divergenz der Folge der $\mu^{(k)}$ ein Vektor mit lauter nichtnegativen Komponenten sein. Nach Voraussetzung I.b muß die Summe daher gleich dem Nullvektor sein.

Für jeden einzelnen Produzenten j' ist somit

$$\sum_{j\neq j'} y_j^{(0)} = -y_{j'}^{(0)} .$$

Da $0 \in Y_{j'}$, liegt die Summe links in Y. Außerdem liegt auch $y_{j'}$ in Y, denn auch für jedes $j \neq j'$ ist $0 \in Y_j$. Somit liegen sowohl $y_{j'}^{(0)}$ als auch $-y_{j'}^{(0)}$ in Y, was nach Voraussetzung I.c nur möglich ist für $y_{j'}^{(0)} = 0$. Da j' beliebig war, gilt dies für alle Produzenten, für jedes j konvergiert also die Folge der $y_j^{(k_q)}/\mu^{(k_q)}$ gegen den Nullvektor. Somit gibt es höchstens endlich viele Indizes q, für die die Norm von $y_j^{(k_q)}$ gleich $\mu^{(k_q)}$ sein kann.

Da es nur endlich viele Produzenten gibt, kann dann auch nur für endlich viele Werte von q die Norm *irgendeines* Vektors $y_j^{(k_q)}$ gleich $\mu^{(k_q)}$ sein. Dies widerspricht aber der Definition von $\mu^{(k)}$ als dem Maximum der Normen der Vektoren $y_j^{(k)}$.

Damit führt die Annahme, es gäbe unbeschränkte Mengen $\widehat{Y}_j$ zu einem Widerspruch; zumindest die $\widehat{Y}_j$ sind also allesamt beschränkt und somit kompakt.

Daraus folgt nun leicht auch die Beschränktheit (und Kompaktheit) der $\widehat{X}_i$: Da $z = x - y - \zeta \leq 0$ ist, haben wir für jede Strategie x_i eines Verbrauchers i die Ungleichung

$$\xi_i \leq x_i \leq \sum_{j=1}^{m} y_j - \sum_{i'\neq i} x_{i'} + \zeta$$

für geeignete Strategien $x_{i'} \in X_{i'}$ und $y_j \in Y_j$. Dabei ist natürlich jedes $x_{i'} \geq \xi_{i'}$; außerdem liegen alle y_j in der jeweiligen Teilmenge $\widehat{Y}_j$, denn es gibt ja Strategien für alle anderen Teilnehmer, die $z \leq 0$ machen. Also ist

$$\xi_i \leq x_i \leq \sum_{j=1}^{m} y_j - \sum_{i' \neq i} \xi_{i'} + \zeta \,.$$

Wegen der Beschränktheit der Mengen $\widehat{Y}_j$ ist die rechte Seite beschränkt, also sind auch alle Mengen $\widehat{X}_i$ beschränkt und somit kompakt.

Da es nur endlich viele Mengen $\widehat{X}_i$ und $\widehat{Y}_j$ gibt, können wir eine reelle Zahl c finden, so daß alle diese Mengen im Innern des Würfels

$$W = \left\{ x \in \mathbb{R}^\ell \;\middle|\; \forall h : \|x_h\| \leq c \right\}$$

liegen.

Mit dieser Zahl c definieren wir eine neue abstrakte Ökonomie, von der wir zeigen wollen, daß sie alle Voraussetzungen des Satzes von DEBREU erfüllt: Für jeden Verbraucher i und jeden Produzenten j definieren wir

$$\widetilde{X}_i = X_i \cap W \quad \text{und} \quad \widetilde{Y}_j = Y_j \cap W \,.$$

Da $\widehat{X}_i \subset \widetilde{X}_i$ und $\widehat{Y}_j \subset \widetilde{Y}_j$ für alle i, j, liegen für jedes Gleichgewicht der ursprünglichen abstrakten Ökonomie die Strategien der Verbraucher in den Teilmengen $\widetilde{X}_i$ und die der Produzenten in $\widetilde{Y}_j$. Für die Verbraucher modifizieren wir entsprechend auch die Menge der möglichen Strategien zu $\widetilde{A}_i(x_{-i}) = A_i(x_{-i}) \cap \widetilde{X}_i$.

Umgekehrt ist auch jedes Gleichgewicht der modifizierten Ökonomie eines der ursprünglichen, denn in einem Gleichgewichtspunkt der modifizierten Ökonomie kann sich kein Agent verbessern, indem er allein eine Strategie wählt, die nicht in $\widetilde{X}_i$ *bzw.* $\widetilde{Y}_j$ liegt: Dies könnte er nur, wenn er auch die Strategien der anderen Agenten verändern könnte.

Zum Beweis von Satz 1 müssen wir daher nur noch zeigen, daß diese modifizierte Ökonomie in der Tat, wie behauptet, alle Voraussetzungen des Satzes von DEBREU erfüllt.

Nach I.a und II sind die Mengen X_i und Y_j abgeschlossen und konvex; der Würfel W ist kompakt und konvex; daher sind die Mengen $\widetilde{X}_i$ und $\widetilde{Y}_j$ kompakt und konvex, genau wie es die Strategiemenge P des Marktteilnehmers ohnehin schon war. Die Stetigkeit der Auszahlungsfunktionen für Verbraucher ist in III.a vorausgesetzt; für die Produzenten und den Marktteilnehmers folgen die entsprechenden Aussagen aus der Linearität dieser Funktionen. Daraus folgt für diese Teilnehmer auch die Konvexität der im Satz von DEBREU definierten Mengen $M_{a_{-\iota}}$; für Verbraucher garantiert III.c diese Konvexität.

Die Korrespondenzen zur Definition der im Kontext durchführbaren Strategien sind für Produzenten und den Marktteilnehmer konstant, also trivialerweise stetig. Da es sich jeweils um die gesamte Strategiemenge handelt, gibt es auch keine Probleme mit der Konvexität.

Im Falle der Verbraucher ist die Konvexität ebenfalls unproblematisch, da die zulässigen Teilmengen $\widetilde{A}_i(x_{-i})$ durch lineare Bedingungen definiert sind. Außerdem ist $\widetilde{A}_i(x_{-i})$ nicht leer, denn nach IV.a gibt es für jeden Verbraucher i eine Strategie $x'_i \in X_i$ mit $x'_i \leq \zeta_i$ und Setzen wir $y'_j = 0$ und nach I.a ist $0 \in Y_j$ für jeden Produzenten j. Mit diesen Strategien ist

$$\sum_{i=1}^{n} x'_i - \sum_{j=1}^{m} y'_j - \zeta \leq 0\,,$$

x'_i liegt also für jeden Verbraucher i in $\widehat{X}_i$ und damit erst recht im Würfel W. Da wir bereits zu Beginn dieses Abschnitts gesehen haben, daß x'_i in $A_i(x_{-i})$ liegt, ist somit $x'_i \in \widetilde{A}_i(x_{-i})$.

Zu zeigen bleibt die Stetigkeit der Korrespondenzen $\widetilde{A}_i$.

Die Halbstetigkeit nach oben ist äquivalent zur Abgeschlossenheit des Graphen, und die wiederum ist klar, da der Graph eine durch nichtstrikte Ungleichungen zwischen stetigen Funktionen definierte Teilmenge eines reellen Vektorraums ist.

Zu zeigen bleibt die Halbstetigkeit nach unten, und hier kommt die (problematische) Voraussetzung IV.a ins Spiel, wonach jeder Verbraucher einen möglichen Verbrauchsplan $x_i < \zeta_i$ hat, d.h. er hat, wenn er diesen

Plan wählt von *jedem* Wirtschaftsgut mehr als er braucht und damit etwas, was er verkaufen kann. Da ARROW und DEBREU diese Voraussetzung im weiteren Verlauf ihrer Arbeit noch deutlich abschwächen, beweisen sie die Halbstetigkeit nach unten tatsächlich unter einer technischen Voraussetzung, die sofort aus IV.a folgt, aber weniger stark ist: Wenn es ein $x_i \in X_i$ gibt mit $x_i < \zeta_i$, ist für diesen Verbrauchsplan insbesondere $\langle p, x_i \rangle < \langle p, \zeta_i \rangle$ für jeden Preisvektor $p \in P$; die Halbstetigkeit nach unten unter der Voraussetzung IV.a folgt also aus

Lemma: Falls gilt $\min\limits_{x_i \in X_i} \langle p, x_i \rangle < \langle p, \zeta_i \rangle$, ist $\widetilde{A}_i$ halbstetig nach unten.

Zum *Beweis* verwenden wir das Folgenkriterium. Sei also $\left(\tilde{x}_i^{(k)}\right)_{k \in \mathbb{N}}$ eine konvergente Folge aus $\widetilde{X}_i$ mit Grenzwert $\tilde{x}_i$, $\left(p^{(k)}\right)_{k \in \mathbb{N}}$ eine in P mit Grenzwert p. Weiter sei $\left(r_i^{(k)}\right)_{k \in \mathbb{N}}$ eine Folge reeller Zahlen, die gegen

$$r_i = \langle p, \zeta_i \rangle + \max\left\{0, \sum_{j=1}^{m} \alpha_{ij} \langle p, y_j \rangle\right\}$$

konvergiere, und $x_i \in \widetilde{A}(\tilde{x}_i)$ sei ein Element mit $\langle p, x_i \rangle \leq r_i$. Wir müssen zeigen, daß es eine gegen x_i konvergierende Folge $\left(x_i^{(k)}\right)_{k \in \mathbb{N}}$ gibt mit $x_i^{(k)} \in \widetilde{A}\left(\tilde{x}_i^{(k)}\right)$ und $\left\langle p, x_i^{(k)} \right\rangle \leq r_i^{(k)}$ für alle k.

Für $\langle p, x_i \rangle < r_i$ ist auch $\left\langle p^{(k)}, x_i \right\rangle < r_i^{(k)}$ für alle hinreichend großen Werte von k; setzen wir $x_i^{(k)} = x_i$ für diese k und wählen für kleinere Indizes irgendwelche zulässigen Vektoren, so haben wir eine entsprechende Folge gefunden.

Schwieriger wird es, wenn $\langle p, x_i \rangle = r_i$ ist. Nach Voraussetzung gibt es $x_i' \in \widetilde{X}_i$ mit $\langle p, x_i' \rangle < \langle p, \zeta_i \rangle$.

Wählen wir ein solches x_i', ist jedenfalls $\langle p, x_i' \rangle < r_i$, so daß für hinreichend große Werte von k auch $\left\langle p^{(k)}, x_i' \right\rangle < r_i^{(k)}$ ist. Wegen der Konvexität von $\widetilde{X}_i$ liegt für $\lambda \in [0,\ 1]$ auch $x_i(\lambda) = \lambda x_i + (1 - \lambda) x_i'$ in $\widetilde{X}_i$,

und

$$\begin{aligned}\left\langle p^{(k)}, x_i(\lambda)\right\rangle &= \lambda \left\langle p^{(k)}, x_i\right\rangle + (1-\lambda)\left\langle p^{(k)}, x_i'\right\rangle \\ &= \lambda\left(\left\langle p^{(k)}, x_i\right\rangle - \left\langle p^{(k)}, x_i'\right\rangle\right) + \left\langle p^{(k)}, x_i'\right\rangle \leq r_i^{(k)}\end{aligned}$$

genau dann, wenn gilt

$$\lambda \leq \frac{r_i^{(k)} - \left\langle p^{(k)}, x_i'\right\rangle}{\left\langle p^{(k)}, x_i\right\rangle - \left\langle p^{(k)}, x_i'\right\rangle}\,.$$

Der Nenner rechts ist positiv, denn $\left\langle p^{(k)}, x_i'\right\rangle < r_i = \left\langle p^{(k)}, x_i\right\rangle$, so daß die rechte Seite auf jeden Fall größer oder gleich Null ist. Sie könnte allerdings größer als eins sein; wir setzen daher

$$\lambda_k = \min\left\{1, \frac{r_i^{(k)} - \left\langle p^{(k)}, x_i'\right\rangle}{\left\langle p^{(k)}, x_i\right\rangle - \left\langle p^{(k)}, x_i'\right\rangle}\right\}\,.$$

Für hinreichend große k ist λ_k nach Wahl von x_i' echt größer Null; somit liegt $x_i(\lambda_k)$ dann in $\widetilde{A}_i(\tilde{x}_i^{(k)})$. Außerdem konvergiert die Folge der $r_i^{(k)}$ gegen $r_i = \lim \left\langle p^{(k)}, x_i\right\rangle$; somit muß die Folge der λ_k gegen eins konvergieren und damit die der $x_i(\lambda_k)$ gegen x_i. Damit ist auch in diesem Fall eine Folge mit den verlangten Eigenschaften gefunden, und das Lemma ist bewiesen. ■

Damit ist Satz 1 vollständig bewiesen. ■

§4: Der Satz von Arrow und Debreu

Wie bereits erwähnt, ist die Bedingung IV.a unter den Voraussetzungen für Satz 1 völlig unrealistisch: Wir nehmen hier schließlich an, daß jeder Verbraucher von jedem Gut etwas besitzt, was er verkaufen kann. Um einen nützlichen Satz zu bekommen, müssen wir diese Bedingung deutlich abschwächen. Damit jeder Verbraucher aktiv am Marktgeschehen teilnehmen kann, muß er allerdings zumindest ein nützliches Gut anbieten können; und dies formalisieren ARROW und DEBREU folgendermaßen:

Definition: *a)* Für $h = 1, \ldots, \ell$ sei $\delta^h \in \mathbb{R}^\ell$ der h-te Koordinateneinheitsvektor, d.h. seine h-te Komponente ist gleich eins, und alle sonstigen

Komponenten verschwinden.
b) $\mathcal{D}$ sei die Menge aller Güter h, für die gilt: Für jeden Verbraucher i und jeden Verbrauchsplan $x_i \in X_i$ gibt es ein $\lambda > 0$, so daß $x_i + \lambda\delta^h \in X_i$ und $u_i(x_i + \lambda\delta^h) > u_i(x_i)$.
c) $\mathcal{P}$ sei die Menge aller Güter h, für die gilt: Für jede mögliche gesamtwirtschaftliche Produktion $y \in Y$ ist $y_h \leq 0$, und es gibt mindestens eine Alternative $y' \in Y$, so daß für alle Güter $h' \neq h$ gilt: $y'_{h'} \geq y_{h'}$, wobei für mindestens ein $h'' \in \mathcal{D}$ sogar $y'_{h''} > y_{h''}$ ist.

Die Menge $\mathcal{D}$ besteht somit aus allen jenen Gütern, von denen jeder Verbraucher stets gerne noch etwas mehr hätte. Die Güter $h \in \mathcal{P}$ können wegen der Bedingung $y_h \leq 0$ nicht produziert werden, sondern sind Resourcen für die Produktion; typischerweise geht es hier um Arbeitsleistungen. Die zweite Bedingung in der Definition von $\mathcal{P}$ besagt, daß damit etwas nützliches produziert werden kann: Durch Einsatz von h läßt sich eine alternative gesamtwirtschaftliche Produktion realisieren, bei der von jedem anderen Gut mindestens genauso viel produziert (beziehungsweise nicht mehr verbraucht) wird, als im ursprünglichen Plan, wobei von mindestens einem der Güter, die jeder gerne hätte, sogar mehr produziert wird.

Bedingung IV.a besagte bekanntlich, daß jeder Verbraucher i einen Verbrauchsplan $x_i \in X_i$ hat, so daß $x_i < \zeta_i$ ist, wobei der Vektor $\zeta_i \in \mathbb{R}^\ell$ seinen Besitz beschreibt und das Kleinerzeichen so zu interpretieren ist, daß es in *jeder* Komponente gilt. Diese Bedingung wird nun abgeschwächt zu

IV'.a Es gibt ein $x_i \in X_i$, so daß $x_i \leq \zeta_i$ ist, wobei für mindestens ein Gut $h \in \mathcal{P}$ gilt: $x_{ih} < \zeta_{ih}$.

Wir verlangen also nur, daß es mindestens eine nützliche Resource (meist eine Form von Arbeit) gibt, die er auf dem Markt anbieten kann. Bedingung IV.b bleibt unverändert, und mit IV' bezeichnen wir die beiden Bedingungen IV'.a und IV.b.

Zum Nachweis der Existenz eines Gleichgewichts reichen die Bedingungen I–III und IV' noch nicht aus; der Beweis benutzt noch drei weitere Voraussetzungen. Die erste davon besagt, daß das Wirtschaftssystem zumindest grundsätzlich in der Lage ist, Überschüsse von allen

Gütern zu produzieren, die beiden anderen stellen sicher, daß es die gerade definierten „nützlichen“ Güter und Dienstleistungen auch tatsächlich gibt:

V. Es gibt ein $x \in X$ und ein $y \in Y$ derart, daß $x < y + \zeta$.

VI. $\mathcal{D} \neq \emptyset$

VII. $\mathcal{P} \neq \emptyset$

Mit diesen zusätzlichen Bedingungen läßt sich nun zeigen

Satz: Unter den Bedingungen I–III, IV’ und V–VII gibt es ein Gleichgewicht.

Der *Beweis* ist ähnlich zu dem von Satz 1: Wir konstruieren Modifikationen des gegebenen Wirtschaftssystems, auf die wir den Satz von DEBREU anwenden können, und deren Gleichgewichte auch Gleichgewichte der gegebenen Ökonomie sind.

Bei Satz 1 hatten wir angenommen, daß jeder Verbraucher von jedem Gut mehr hat als er braucht und somit etwas davon verkaufen kann; jetzt fordern wir dies nur noch für mindestens ein Gut, von dem wir allerdings verlangen, daß es in der Menge $\mathcal{P}$ liegt. Es ist klar, daß es Schwierigkeiten gibt, wenn der Marktteilnehmer den Preis dieses Gutes auf Null setzt. Deshalb erlauben wir dem Marktteilnehmer im ersten Teil des Beweises nur Strategien, bei denen die Preise der Güter aus $\mathcal{P}$ nicht unter eine vorgegebene Schranke ε fallen darf. Diese Schranke darf natürlich nicht zu groß sein, denn die Summe aller Komponenten des Preisvektors ist ja eins.

Wir bezeichnen die Elementanzahl der Menge $\mathcal{P}$ mit π und betrachten für jedes $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$ die Teilmenge

$$P^{\varepsilon} \underset{\text{def}}{=} \left\{ p \in P \;\middle|\; \forall h \in \mathcal{P} : p_h \geq \varepsilon \right\} .$$

Solange der Marktteilnehmer nur Preisvektoren aus P^{ε} wählen darf, ist sichergestellt, daß jeder Verbraucher etwas mit positivem Wert anbieten kann, denn nach IV’.a hat Verbraucher i einen Plan $x_i \in X_i$, für den $x_i \leq \zeta_i$ ist und $x_{ih'} < \zeta_{ih'}$ für mindestens ein $h' \in \mathcal{P}$. Daher ist

$$\langle p, \zeta_i - x_i \rangle = \sum_{h=1}^{\ell} p_h(\zeta_{ih} - x_{ih}) \geq p_{h'}(\zeta_{ih'} - x_{ih'}) > 0 \,;$$

es gibt also mindestens ein $x_i \in X_i$, dessen Kosten $\langle p, x_i \rangle$ kleiner sind als der Wert $\langle p, \zeta_i \rangle$ des Vermögens von i. Wir bezeichnen die abstrakte Ökonomie, bei der sie Strategiemenge des Marktteilnehmers auf P_ε beschränkt ist, mit E^ε; die, bei der er seine volle Strategiemenge P einsetzen kann, mit E.

Im ersten Teil des Beweises zeigen ARROW und DEBREU, daß E^ε für jedes $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$ ein Gleichgewicht hat; im zweiten zeigen sie, daß es mindestens ein solches ε gibt, für das ein Gleichgewicht von E^ε auch Gleichgewicht von E ist.

Für den ersten Teil zeigen sie wieder, daß jedes Gleichgewicht von E^ε in einem geeigenten Hyperwürfel liegen muß und daß ein Gleichgewicht der auf diesen Hyperwürfel beschränkten Ökonomie E^ε auch eines von E^ε sein muß.

Angenommen, E^ε hat ein Gleichgewicht $(x_1^*, \ldots, x_n^*, y_1^*, \ldots, y_m^*, p^*)$. Nach Definition der abstrakten Ökonomie zu einem Wirtschaftssystem dürfen Konsumenten nur Verbrauchspläne wählen, die sie sich auch leisten können, d.h. $x_i^* \in A_i(x_{-i}^*)$, oder konkret

$$\langle p^*, x_i^* \rangle \leq \langle p^*, \zeta_i \rangle + \sum_{j=1}^{m} \alpha_{ij} \langle p^*, y_j \rangle \ .$$

Summieren wir diese Ungleichungen über alle Verbraucher $i = 1, \ldots, n$, so erhalten wir wegen der Linearität des Skalarprodukt in jedem der beiden Argumente die Ungleichung

$$\begin{aligned} \left\langle p^*, \sum_{i=1}^{n} x_i^* \right\rangle &\leq \left\langle p^*, \sum_{i=1}^{n} \zeta_i \right\rangle + \sum_{j=1}^{m} \sum_{i=1}^{n} \alpha_{ij} \langle p^*, y_j \rangle \\ &= \left\langle p^*, \sum_{i=1}^{n} \zeta_i \right\rangle + \sum_{j=1}^{m} \langle p^*, y_j \rangle \ . \end{aligned}$$

Somit ist

$$\left\langle p^*, \sum_{i=1}^{n} x_i^* - \sum_{i=1}^{n} \zeta_i - \sum_{j=1}^{m} y_j^* \right\rangle \leq 0 \, ,$$

wobei das zweite Argument im Skalarprodukt gerade der Nachfrageüberhang $z^* = x^* - \zeta - y^*$ ist, d.h. $\langle p^*, z^* \rangle \leq 0$.

Nach Definition eines Gleichgewichts ist

$$\langle p^*, z^* \rangle = \max\left\{ \langle p, z^* \rangle \;\middle|\; p \in P^\varepsilon \right\}.$$

Da dieses Maximum nicht positiv sein kann, ist also $\langle p, z^* \rangle \leq 0$ für alle $p \in P^\varepsilon$. Ausgeschrieben ist das die Ungleichung

$$\sum_{h=1}^{\ell} p_h(x_h^* - y_h^* - \zeta_h) \leq 0\,.$$

Wir betrachten nun ein festes Gut h' und bringen alle Summanden mit Indizes $h \neq h'$ auf die rechte Seite. Dann folgt

$$p_{h'}(x_{h'}^* - y_{h'}^* - \zeta_{h'}) = p_{h'} z_{h'} \leq \sum_{h \neq h'} p_h(\zeta_h + y_h^* - x_h^*)\,.$$

Für $h \in \mathcal{P}$ ist $y_h^* \leq 0$ und damit $x *_h -y_h^* \geq x_h^* \geq \xi_h$.

Wir betrachten nun zum ausgewählten Gut h' den Preisvektor $p \in P^\varepsilon$, der h' den höchstmöglichen Preis zuordnet. Allen anderen Gütern wird dann der geringstmögliche Preis zugeordnet, d.h.

$$p_h = \begin{cases} \varepsilon & \text{falls } h \in \mathcal{P} \text{ und } h \neq h' \\ 0 & \text{falls } h \notin \mathcal{P} \text{ und } h \neq h' \end{cases}.$$

Dann ist

$$p_{h'} = 1 - \sum_{h \neq h'} p_h = \begin{cases} 1 - \pi\varepsilon & \text{falls } h' \notin \mathcal{P} \\ 1 - (\pi - 1)\varepsilon & \text{falls } h' \in \mathcal{P} \end{cases}.$$

Dieser Vektor liegt in der Tat in P^ε, denn $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$, d.h. $\pi\varepsilon \leq \frac{1}{2}$. Für $h' \notin \mathcal{P}$ ist $1 - \pi\varepsilon \geq \frac{1}{2}$ daher positiv, und für $h' \in \mathcal{P}$ ist

$$1 - (\pi - 1)\varepsilon = (1 - \pi\varepsilon) \geq \frac{1}{2} + \varepsilon \geq \varepsilon\,.$$

Nach den beiden Ungleichungen aus dem vorigen Abschnitt gilt daher für diesen speziellen Preisvektor im Falle $h' \in \mathcal{P}$ die Ungleichung

$$\begin{aligned} p_{h'} z_{h'}^* &\leq \sum_{h \neq h'} p_h(\zeta_h - x_h^* + y_h^*) \leq \sum_{h \in \mathcal{P} \setminus \{h'\}} p_h(\zeta_h - x_h^* + y_h^*) \\ &\leq \sum h \in \mathcal{P} \setminus \{h'\} \varepsilon(\zeta_h - \xi_h)\,. \end{aligned}$$

Im Falle $h' \notin \mathcal{P}$ folgt entsprechend

$$p_{h'} z^*_{h'} = (1 - \pi\varepsilon) z^*_{h'} \leq \sum_{h' \in \mathcal{P}} \varepsilon(\zeta_h - \xi_h) \,.$$

Wegen der Positivität der Nenner ist

$$\frac{\varepsilon}{1 - (\pi - 1)\varepsilon} < \frac{\varepsilon}{1 - \pi\varepsilon} \,,$$

und wegen II und IV' ist $\zeta_h - \xi_h \geq 0$ für jedes Gut h.

Nach den gerade hergeleiteten Ungleichungen für $p_{h'} z^*_{h'}$ ist deshalb für $h' \in \mathcal{P}$

$$z^*_{h'} \leq \frac{\varepsilon}{1 - (\pi - 1)\varepsilon} \sum_{h \in \mathcal{P} \setminus \{h'\}} (\zeta_h - \xi_h) \leq \frac{\varepsilon}{1 - \pi\varepsilon} \sum_{h \in \mathcal{P} \setminus \{h'\}} (\zeta_h - \xi_h) \,,$$

und für $h' \notin \mathcal{P}$ gilt ebenfalls

$$z^*_{h'} \leq \frac{\varepsilon}{1 - \pi\varepsilon} \sum_{h \in \mathcal{P} \setminus \{h'\}} (\zeta_h - \xi_h) \,;$$

die Ungleichung gilt also für alle Güter h'.

Da ε in $\left(0, \frac{1}{2\pi}\right)$ liegt, ist $2\pi\varepsilon \leq 1$ und damit $\pi\varepsilon \leq 1 - \pi\varepsilon$, d.h.

$$\frac{\varepsilon}{1 - \pi\varepsilon} \leq \frac{1}{\pi} \,.$$

Somit ist

$$z^*_{h'} \leq \frac{1}{\pi} \sum_{h \in \mathcal{P}} (\zeta_h - \xi_h) \quad \text{für alle } h' = 1, \ldots, \ell \,.$$

$1 \in \mathbb{R}^\ell$ sei der Vektor, dessen sämtliche Komponenten gleich eins sind, und

$$\zeta' = \zeta + \frac{1}{\pi} \sum_{h \in \mathcal{P}} (\zeta_h - \xi_h) 1 \,.$$

Dann ist

$$x^* - y^* - \zeta = z^* \leq \frac{1}{\pi} \sum_{h \in \mathcal{P}} (\zeta_h - \xi_h) \,,$$

also

$$x^* - y^* \leq \zeta + \frac{1}{\pi} \sum_{h \in \mathcal{P}} 1 = \zeta' .$$

Jeder Gleichgewichtspunkt von E^ε liegt also in jener Teilmenge von $\left(\mathbb{R}^\ell\right)^{n+m+1}$, in der gilt:

$$x^* - y^* \leq \zeta' \quad \text{und} \quad x_i^* \in X_i, \quad y_i^* \in Y_i, \quad p^* \in P^\varepsilon .$$

Daher sind im Gleichgewicht für jeden Agenten nur Strategien möglich, für die es Strategien der übrigen Agenten gibt derart, daß diese Ungleichung erfüllt ist. Für ein Gleichgewicht muß daher jede Komponente x_i^* in der Menge

$$\widehat{X}_i' = \left\{ x_i \in X_i \;\middle|\; \forall i' \neq i \exists x_{i'} \in X_{i'} \wedge \forall j \exists y_j \in Y_j : x - y \leq \zeta' \right\}$$

liegen, und y_j^* in der Menge

$$\widehat{Y}_j' = \left\{ y_j \in Y_j \;\middle|\; \forall i \exists x_i \in X_i \wedge \forall j' \neq j \exists y_{j'} \in Y_{j'} : x - y \leq \zeta' \right\} .$$

Dies können wir vergleichen mit der Situation im Beweis von Satz 1: Dort hatten wir gesehen, daß für ein Gleichgewicht x_i^* in der Menge

$$\widehat{X}_i = \left\{ x_i \in X_i \;\middle|\; \forall i' \neq i \exists x_{i'} \in X_{i'} \wedge \forall j \exists y_j \in Y_j : x - y \leq \zeta \right\}$$

liegen, und y_j^* in der Menge

$$\widehat{Y}_j = \left\{ y_j \in Y_j \;\middle|\; \forall i \exists x_i \in X_i \wedge \forall j' \neq j \exists y_{j'} \in Y_{j'} : x - y \leq \zeta \right\} .$$

Daraus hatten wir dort gefolgert, daß alle Mengen $\widehat{X}_i$ und $\widehat{Y}_j$ kompakt sind und es somit einen Hyperwürfel in $\mathbb{R}^\ell$ gibt, der alle diese Mengen enthält. In den Beweis ging die Forderung IV.a nicht mehr ein, und an den Besitzvektor ζ hatten wir keinerlei Forderungen gestellt, die nicht auch von ζ' erfüllt werden. Wir können den dortigen Beweis also wörtlich wiederholen mit Strichen an den $\widehat{X}$, $\widehat{Y}$ und an ζ und kommen auf die Existenz eines Würfels

$$W' = \left\{ x \in \mathbb{R}^\ell \;\middle|\; \|x_h\| \leq c' \right\}$$

für ein $c' \in \mathbb{R}$ derart, daß alle Mengen $\widehat{X}_i$ und $\widehat{Y}_j$ im Innern von W' liegen.

Wir definieren eine neue abstrakte Ökonomie $\widetilde{E}^\varepsilon$ indem wir in E^ε indem alle X_i ersetzen durch $\widetilde{X}_i' \underset{\text{def}}{=} X_i \cap W'$ und alle Y_j durch $\widetilde{Y}_j' = Y_j \cap W'$. Außerdem sei $\widetilde{A}_i'$ die Einschränkung der Korrespondenz A_i auf $\prod_{i' \neq i} \widetilde{X}_{i'}' \times \prod_{j=1}^m \widetilde{Y}_j' \times P^\varepsilon$, deren Werte dann notwendigerweise in $\widetilde{X}_i'$ liegen. Wir müssen zeigen, daß die Voraussetzungen des Satzes von DEBREU für die Existenz eines Gleichgewichts erfüllt sind.

Abgesehen von der Stetigkeit der Korrespondenz $\widetilde{A}_i'$ folgt alles Wort für Wort wie im Beweis von Satz 1. Dort hatten die Stetigkeit der Korrepondenz $\widetilde{A}_i$ gezeigt über die Aussage:

Unter der Bedingung $\langle p, \xi_i \rangle > \min_{x_i \in \widetilde{X}_i} \langle p, x_i \rangle$ folgt die Stetigkeit von $\widetilde{A}_i$ im Punkt $(x_1, \ldots, x_n, y_1, \ldots, y_m, p)$.

Wir wissen aus IV'.a, daß es für jedes $p \in P^\varepsilon$ ein $x_i \in X_i$ gibt mit $\langle p, \xi_i \rangle > \langle p, x_i \rangle \geq \min_{x_i \in \widetilde{X}_i'} \langle p, x_i \rangle$. Damit folgt auch die Stetigkeit von $\widetilde{A}_i$.

Nach dem Satz von DEBREU hat $\widetilde{E}^\varepsilon$ daher für jedes $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$ ein Gleichgewicht $(x_1^*, \ldots, x_n^*, y_1^*, \ldots, y_m^*, p^*)$. Wir müssen noch zeigen, daß dies auch ein Gleichgewicht von E^ε ist.

Nach Voraussetzung I.a liegt der Nullvektor in jeder der Mengen Y_j und damit natürlich auch in $\widetilde{Y}_j'$. Da die Gleichgewichtsstrategie den Wert der Produktion maximiert unter der Voraussetzung, daß die Strategien aller anderen Agenten gleich bleiben, ist also $\left\langle p^*, y_j^* \right\rangle \geq \langle p^*, 0 \rangle = 0$. Damit ist auch der Wert $\sum_{j=1}^m \alpha_{ij} \left\langle p^*, y_j^* \right\rangle$ der Gewinnbeteilungen jedes Vebrauchers i nicht negativ.

Oben haben wir gesehen, daß $\langle p^*, z^* \rangle \leq 0$ ist und daß daraus die Ungleichung $x^* - y^* \leq \zeta'$ folgt. Außerdem sind alle x_i^*, y_j^* innere Punkte des Würfels W'. Nach Definition eines Gleichgewichts ist $u_i(x_i^*)$ das Maximum über alle $u_i(x_i)$ mit $x_i \in \widetilde{A}_i'(x_{-i}^*)$. Wir müssen zeigen, daß es sogar sogar das Maximum über alle $u_i(x_i)$ mit $x_i \in A_i(x_{-i}^*)$ ist.

Angenommen, es gibt ein $x_i' \in A_i(x_{-i}^*)$ mit $u_i(x_i') > u_i(x_i^*)$. Nach III.c ist dann auch $u_i\big(\lambda x_i' + (1-\lambda) x_i^*)\big) > u_i(x_i^*)$ für jedes $\lambda \in (0, 1)$. Da

x_i^* ein innerer Punkt des Würfels W' ist, liegt $\lambda x_i' + (1-\lambda)x_i^*$ für hinreichend kleine Werte von λ im Würfel W'. Wegen der Konvexität von $A_i(x_{-i}^*)$ liegt dieser Vektor daher für hinreichend kleine λ in $\widetilde{A}_i'(x_{-i}^*)$, im Widerspruch zur Gleeichgewichtseigenschaft für $\widetilde{E}^\varepsilon$. Also kann es kein solches x_i' geben.

Entsprechend folgt auch, das $\langle p^*, y_j^* \rangle$ maximal ist auch für alle $\langle p^*, y_j \rangle$ mit $y_j \in Y_j$, und die Strategiemenge des Marktteilnehmers ist ohnehin dieselbe für E^ε und $\widetilde{E}^\varepsilon$. Somit ist jedes Gleichgewicht von $\widetilde{E}^\varepsilon$ auch eines von E^ε.

Damit ist gezeigt, daß E^ε für jedes $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$ ein Gleichgewicht hat.

Im zweiten Schritt muß noch gezeigt werden, daß eines dieser Gleichgewichte auch ein Gleichgewicht der Ausgangsökonomie E ist. Hier unterscheiden wir zwei Fälle:

1. Fall: Es gibt ein $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$, für das $p_h^* > \varepsilon$ ist für alle $h \in \mathcal{P}$.

Wir wollen zeigen, daß in diesem Fall jedes Gleichgewicht von E^ε auch eines von E ist.

Dazu sei p irgendein Preisvektor aus dem vollen Preissimplex P. Wir müssen zeigen, daß $\langle p, z^* \rangle \leq \langle p^*, z^* \rangle$ ist.

Angenommen, $\langle p, z^* \rangle > \langle p^*, z \rangle$ für ein $p \in P$. Da P^ε konvex ist und p^* in diesem ersten Fall als innerer Punkt von P^ε vorausgesetzt wird, liegt dann für hinreichend kleine Werte $\lambda > 0$ auch $p' = \lambda p + (1-\lambda)p^*$ in P^ε, und $\langle p', z^* \rangle > \langle p^*, z^* \rangle$. Im Gleichgewicht ist aber $\langle p, z^* \rangle$ für $p = p^*$ maximal über alle $p \in P^\varepsilon$, ein Widerspruch. Somit ist das Gleichgewicht auch eines für E, denn für die Produzenten und Konsumenten ändert sich ja nichts, wenn die Strategiemenge des Marktteilnehmers vergrößert wird.

2. Fall: Die Bedingung aus dem ersten Fall ist nicht erfüllt.

Dann gibt es also für jedes $\varepsilon \in \left(0, \frac{1}{2\pi}\right)$ und jedes Gleichgewicht $(x_1^*, \ldots, x_n^*, y_1^*, \ldots, y_m^*, p^*)$ von E^ε mindestens ein $h \in \mathcal{P}$ mit $p_h = \varepsilon$.

Wir wollen uns überlegen, daß dies nicht möglich ist.

Dazu betrachten wir eine Nullfolge aus $\left(0, \frac{1}{2\pi}\right)$ und für jedes Folgenglied ε einen Gleichgewichtspunkt von E^{ε}. Dessen Komponente für den i-ten Verbraucher liegt in $X_i' = X_i \cap W'$, und die für den j-ten Produzenten liegt in $Y_j' = Y_j \cap W'$. Sowohl W' als auch P sind kompakt, und in einem Kompaktum hat jede Folge eine konvergente Teilfolge. Gehen wir nacheinander alle Verbraucher, Produzenten und schließlich den Marktteilnehmer durch, erhalten wir also schließlich eine Teilfolge $(\varepsilon_k)_{k\in\mathbb{N}}$ mit der Eigenschaft, daß für die Gleichgewichte $(x_1^{(k)}, \ldots, x_n^{(k)}, y_1^{(k)}, \ldots, y_m^{(k)}, p^{(k)})$ alle Folgen $(x_i^{(k)})_{k\in\mathbb{N}}$, $(y_j^{(k)})_{k\in\mathbb{N}}$ und $(p^{(k)})_{k\in\mathbb{N}}$ konvergieren. Sei

$$x_i^{(0)} = \lim_{k\to\infty} x_i^{(k)}, \quad y_j^{(0)} = \lim_{k\to\infty} y_j^{(k)} \quad \text{und} \quad p^{(0)} = \lim_{k\to\infty} p^{(k)}.$$

Wegen der Kompaktheit von “ und P liegen diese Grenzwerte in W' *bzw.* P, und da die X_i und Y_j als abgeschlosssen vorausgesetzt waren, liegen die $x_i^{(0)}$ in X_i und die $y_j^{(0)}$ in Y_j.

Hier im zweiten Fall gibt es für jeden Indes k mindestens ein Gut $h \in \mathcal{P}$, für das $p_h^{(k)} = \varepsilon_k$ ist. Da es nur endlich viele $h \in \mathcal{P}$ gibt, aber unendlich viele Indizes k, gibt es mindestens ein $h \in \mathcal{P}$, für das $p_h^{(k)} = \varepsilon_k$ ist für unendlich viele Indizes k. Da die ε_k eine Nullfolge bilden und die Folge $\left(p_h^{k)}\right)_{k\in\mathbb{N}}$ konvergiert und eine Nullfolge als Teilfolge hat, ist somit auch $p_h^{(0)} = 0$.

Um die Anzahl der Buchstaben und Modifikationen nicht noch weiter zu vergrößern, nummerieren wir die Güter so um, daß für das Gut $h = 1$ der Grenzpreis $p_1^{(0)}$ verschwindet.

Nach Definition eines Gleichgewichts verschwindet $\left\langle p^{(k)}, z^{(k)}\right\rangle$ für alle k, also ist auch im Limes $\left\langle p^{(0)}, z^{(0)}\right\rangle = 0$.

Außerdem ist im Gleichgewicht $\left\langle p^{(k)}, y^{(k)}\right\rangle \geq \left\langle p^{k)}, y_j\right\rangle$ für alle $y_j \in Y_j$, also auch $\left\langle p^{(0)}, y^{(0)}\right\rangle \geq \left\langle p^{0)}, y_j\right\rangle$ für alle $y_j \in Y_j$,

Damit erfüllt der Grenzwert zumindest einige der Bedingungen an ein Gleichgewicht. Betrachten wir die Situation für die Konsumenten.

Nach Voraussetzung III.b gibt es für jeden von ihnen einen Verbrauchsplan $x_i \in X_i$ mit $u_i(x_i) > u_i(x_i^{(0)})$. Wegen der Konvergenz der Folge der

$x_i^{(k)}$ gegen $x_i^{(0)}$ und der Stetigkeit der Funktion u_i ist dann für hinreichen große Indizes k auch $u_i(x_i) > u_i(x_i^{(k)})$. Daher kann x_i nicht in $A_i(x_{-i}^{(k)})$ liegen, d.h. $\langle p^{(k)}, x_i \rangle > \left\langle p^{k)}, x_i^{(k)} \right\rangle$.

Lassen wir k gegen unendlich gehen, folgt: Falls $u_i(x_i) > u_i(x_i^{(0)})$ für ein $x_i \in X_i$, muß $\langle p^{(0)}, x_i \rangle \geq \left\langle p^{(0)}, x_i^{(0)} \right\rangle$ sein.

Zur Erinnerung: $\mathcal{P}$ enthält alle Güter h, für die gilt: $y_h \leq 0$ für alle möglichen Gesamtproduktionen $y \in Y$ des Wirtschaftssystems, und es gibt einen Produktionsplan $y' \in Y$, so daß $y'_{h'} \geq y_{h'}$ für alle $h' \neq h$, und es gibt ein $h'' \in \mathcal{D}$, so daß $y'_{h''} > y_{h''}$.

Da $1 \in \mathcal{P}$, gibt es also eine gesamtwirtschaftliche Produktion $y' \in Y$, für die $y'_h \geq y_h^{(0)}$ für alle $h \neq 1$ und $y'_{h'} > y_{h'}^{(0)}$ für mindestens ein $h' \in \mathcal{D}$.

Wie wir gerade gesehen haben, ist $\left\langle p^{(0)}, y_j^{(0)} \right\rangle \geq \langle p^{(0)}, y_j \rangle$ für alle $y_j \in Y_j$. Aufsummieren über alle Produzenten j zeigt, daß dann für $y^{(0)} = \sum_{j=1}^m y_j^{(0)}$ gilt $\langle p^{(0)}, y^{(0)} \rangle \geq \langle p^{(0)}, y' \rangle$. Somit ist

$$\begin{aligned} 0 \geq \langle p^{(0)}, y' - y^{(0)} \rangle &= \sum_{h=1}^{\ell} p_k^{(0)}(y'_h - y_h^{(0)}) = \sum_{h=2}^{\ell} p_k^{(0)}(y'_h - y_h^{(0)}) \\ &\geq p_{h'}^{(0)}(y'_{h'} - y^{(0)}) \,. \end{aligned}$$

Da $y'_{h'} - y^{(0)}$ positiv ist, muß somit $p_{h'}^{(0)}$ verschwinden. Es gibt somit mindestens ein Gut h' aus $\mathcal{D}$ mit $p_{h'} = 0$. In unserem Wirtschaftssystem gibt es also mindestens ein Gut, von dem jeder gerne mehr hätte, und das für alle umsonst zur Verfügung steht. Das ist zu schön, um wahr zu sein, und sollte uns auf einen Widerspruch führen.

Nach Definition von $\mathcal{D}$ gibt es für jeden Verbraucher i ein $\lambda > 0$, so daß $u_i(x_i^{(0)} + \lambda\delta^{(h')}) > u_i(x_i^{(0)})$ ist, wobei $\delta^{(h')}$ der Koordinateneinheitsvektor ist, der an der Stelle h' eine Eins hat und sonst lauter Nullen.

Nun sei $x_i \in X_i$ ein beliebiger Verbrauchsplan, und für $\mu \in (0, 1)$ sei $x_i(\mu) = \mu x_i + (1 - \mu)x_i^{(0)}$. Da $\lim\limits_{\mu \to 0} x_i(\mu) = x_i^{(0)}$ ist, muß dann auch

$u_i\big(x_i(\mu) + \lambda\delta^{(h')}\big) > u_i(x_i^{(0)})$ sein für hinreichend kleine positive Werte von μ.

Wie wir oben gesehen haben, folgt daraus, daß

$$\left\langle p^{(0)}, x_i(\mu) + \lambda\delta^{(h')}\right\rangle \geq \left\langle p^{(0)}, x_i^{(0)}\right\rangle$$

ist für alle hinreichend kleinen Werte von μ.

Wir können die linke Seite dieser Ungleichung noch genauer ausrechnen: Da $p_{h'}^{(0)}$ verschwindet, ist $\left\langle p^{(0)}, \delta^{(h')}\right\rangle = 0$, d.h.

$$\begin{aligned}\left\langle p^{(0)}, x_i(\mu) + \lambda\delta^{(h')}\right\rangle &= \left\langle p^{(0)}, x_i(\mu)\right\rangle = \left\langle p^{(0)}, \mu x_i + (1-\mu) + x_i^{(0)}\right\rangle \\ &= \mu\left\langle p^{(0)}, x_i\right\rangle + (1-\mu)\left\langle p^{(0)}, x_i^{(0)}\right\rangle = \left\langle p^{(0)}, x_i(\mu) + \lambda\delta^{(h')}\right\rangle .\end{aligned}$$

Dies muß also größer oder gleich $\left\langle p^{(0)}, x_i^{(0)}\right\rangle$ sein, und das geht für positive μ nur, wenn

$$\left\langle p^{(0)}, x_i\right\rangle \geq \left\langle p^{(0)}, x_i^{(0)}\right\rangle$$

ist. Somit minimiert $x_i^{(0)}$ das Skalarprodukt $\left\langle p^{(0)}, x_i\right\rangle$ über alle $x_i \in X_i$.

Dies gilt für jeden Verbraucher i; falls wir über alle Verbraucher summieren erhalten wir daher die Ungleichung

$$\left\langle p^{(0)}, x\right\rangle \geq \left\langle p^{(0)}, x^{(0)}\right\rangle \quad \text{für alle } x \in X .$$

Nach Forderung V gibt es $x \in X$ und $y \in Y$, für die $x < y + \zeta$ ist, d.h. es ist möglich, Verbrauchs- und Produktionspläne zu finden derart, daß die Nachfrage nach jedem Gut echt kleiner ist als Produktion und vorhandener Besitz. Für eine solche Wahl ist dann

$$\left\langle p^{(0)}, y + \zeta\right\rangle > \left\langle p^{(0)}, x\right\rangle \geq \left\langle p^{(0)}, x^{(0)}\right\rangle$$

und damit

$$\left\langle p^{(0)}, y\right\rangle > \left\langle p^{(0)}, x^{(0)} - \zeta\right\rangle .$$

Außerdem ist $\left\langle p^{(0)}, x^{(0)} - y^{(0)} - \zeta\right\rangle = 0$, also

$$\left\langle p^{(0)}, x^{(0)} - \zeta\right\rangle = \left\langle p^{(0)}, y^{(0)}\right\rangle .$$

Insgesamt erhalten wir so die Ungleichung

$$\left\langle p^{(0)}, y \right\rangle > \left\langle p^{(0)}, y^{(0)} \right\rangle .$$

Diese kann nur gelten, wenn es mindestens einen Produzenten j gibt, für den $\left\langle p^{(0)}, y_j \right\rangle > \left\langle p^{(0)}, y_j^{(0)} \right\rangle$ ist. Wegen der Konvergenz der Folge der $p^{(k)}$ gegen $p^{(0)}$ und der $y_j^{(k)}$ gegen $y_j^{(0)}$ ist dann auch $\left\langle p^{(k)}, y_j \right\rangle > \left\langle p^{(k)}, y_j^{(k)} \right\rangle$ für hinreichend große Indizes k, aber das ist unmöglich, da auch Produzent j im Gleichgewicht eine Strategie wählt, die $\left\langle p^{(k)}, y_j \right\rangle$ maximiert.

Dieser Widerspruch zeigt, daß der zweite Fall nicht eintreten kann; wir sind also immer im ersten Fall, und für den haben wir die Existenz eines Gleichgewichts für E gezeigt. Damit ist der Satz von ARROW und DEBREU bewiesen. ■

Nachdem wir soviel Zeit und Mühe in den Beweis dieses Satzes gesteckt haben, stellt sich natürlich die Frage, ob es dieser Mühe Wert war und ob uns dieser Satz wirklich etwas nützt, um unser reales Wirtschaftssystem zu verstehen. Im Idealfall sollte die Wirtschaft sich einem Gleichgewicht annähern um dann, da es in der Realität natürlich immer Schwankungen gibt, in dessen Nähe zu oszillieren.

Für einen Mathematiker gibt es vor allem zwei kritische Punkte, die dem im Wege stehen: Zum einen sagt der Satz nichts aus über die Eindeutigkeit des Gleichgewichts, und in der Tat wird es oft sogar eine überabzählbare kompakte und konvexe Menge von Gleichgewichten geben. Selbst wenn die Akteure ein Gleichgewicht anstreben, ist es daher eher unwahrscheinlich, daß sie alle auf dasselbe Gleichgewicht zuarbeiten.

Wichtiger noch ist, daß ein Gleichgewicht üblicherweise über einen dynamischen Prozess erreicht oder zumindest annähernd erreicht wird. Im gerade zu Ende geführstene Beweis spielt Dynamik allerdings überhaupt keine Rolle; hier geht es nur um die reine Existenz. Das Wirtschaftsleben ist allerdings ein dynamischer Prozess, und in dynamischen Prozessen gibt es durchaus unterschiedliche Arten von Gleichgewichten.

Betrachten wir etwa das Beispiel einer großen und einer kleinen Kugel. Liegt die kleine Kugel oben auf der großen, so ist das physikalisch gesehen eine Gleichgewichtslage, denn es gibt keine Kräfte, die die kleine Kugel in irgendeine Richtung lenken würden. Schon die allerkleinste Störung würde die kleine Kugel allerdings abstürzen lassen, so daß dieses Gleichgewicht völlig instabil ist.

Liegt dagegen die kleine Kugel unten im Inneren der großen Kugel, so führen Störungen zwar dazu, daß das Gleichgewicht verlassen wird, aber in der Theorie asymptotisch, in der Praxis in endlicher Zeit, wird sie in die Gleichgewichtslage zurückkehren.

Bei den Gleichgewichten, deren Existenz wir gerade bewiesen haben, verzichteten wir völlig auf dynamische Gesichtspunkte, so daß auch keinerlei Aussagen über die Stabilität von Gleichgewichten möglich sind. Trotzdem hält anscheinend die große Mehrheit der Ökonomen die Existenz dieser Gleichgewichte für relavant für das Verständis unseres Wirtschaftssystems.

Kapitel 10
Variationen

§ 1: Nashs zweiter Beweis für die Existenz von Gleichgewichten

NASHs Beweis für die Existenz von Gleichgewichten eines Spiels erschien 1950 in der Proceedings of the National Academy of Sciences und war nicht einmal eine Seite lang. Ein Jahr später veröffentlichte er in

> JOHN NASH: Non-Cooperative Games, *Annals of Mathematics* **54**, 286-295

einen zweiten Beweis, der zwar etwas länger war, den er aber für ein *considerable improvement* hielt und der an Stelle des Fixpunktsatzes von KAKUTANI direkt den BROUWERschen Fixpunktsatz benutzte. Er geht wieder aus von einem endlichen Spiel, d.h. die Anzahl n der Spieler ist endlich, und jeder von ihnen hat eine endliche Anzahl von reinen Strategien. Die reinen Strategien des i-ten Spielers bezeichnet er mit $\pi_{i\alpha}$; die Menge seiner gemischten Strategien ist somit das Simplex S_i mit den $\pi_{i\alpha}$ als Ecken. Die Indexmenge, über die α läuft, kann je nach Spieler verschieden sein; wenn von einer Summe über Indizes α die Rede ist, soll das stets bedeuten, daß wir über alle Strategien des jeweiligen Spielers summieren. Die Auszahlungsfunktion für den i-ten Spieler sei $a^{(i)}$; sie ist eine in jedem der n Argumente lineare Funktion von $S_1 \times \cdots \times S_n$ nach $\mathbb{R}$.

Für ein n-tupel $s = (s_1, \ldots, s_n)$ sagen wir, s_i und auch s *benutzen* die reine Strategie $\pi_{i\alpha}$, falls $\pi_{i\alpha}$ in s_i mit positiver Wahrscheinlichkeit auftritt.

Ein n-tupel $s^* = (s_1^*, \ldots, s_n^*)$ heißt bekanntlich ein *Gleichgewicht,* wenn

$$a^{(i)}(s_1^*, \ldots, s_n^*) = \max_{s_i \in S_i} a^{(i)}(s_{-i}^* | s_i)$$

für alle Spieler i ist, d.h. keiner kann sich verbessern, indem er einseitig seine Strategie ändert.

Die Auszahlung für eine gemischte Strategie $s_i = \sum_\alpha p_{i\alpha}\pi_{i\alpha}$ ist der Erwartungswert, d.h.

$$a^{(i)}(s_{-1}^* | s_i) = \sum_\alpha p_{i\alpha} a^{(i)}(s_{-i}^* | \pi_{i\alpha})\,.$$

Falls die reine Strategie $\pi_{i\alpha^*}$ unter allen reinen Strategien die höchste Auszahlung liefert unter der Nebenbedingung, daß jeder andere Spieler j seine Strategie s_j^* spielt, so ist $a^{(i)}(s_{-i}^* | \pi_{i\alpha^*}) \geq a^{(i)}(s_{-i}^* | \pi_{i\alpha})$ für alle reinen Strategien $\pi_{i\alpha}$ des i-ten Spielers. Da die Summe der Wahrscheinlichkeiten $p_{i\alpha}$ gleich eins ist, folgt also

$$\begin{aligned} a^{(i)}(s_{-i}^* | \pi_{i\alpha^*}) &= \sum_\alpha p_{i\alpha} a^{(i)}(s_{-i}^* | \pi_{i\alpha^*}) \\ &\geq \sum_\alpha p_{i\alpha} a^{(i)}(s_{-i}^* | \pi_{i\alpha}) = a^{(i)}(s_{-i}^* | s_i)\,. \end{aligned}$$

Für eine Strategie s_i, die $a^{(i)}(s_{-i}^* | s_i)$ maximiert, muß umgekehrt auch $a^{(i)}(s_{-i}^* | \pi_{i\alpha^*}) \leq a^{(i)}(s_{-i}^* | s_i)$ sein, so daß

$$\max_{s_i \in S_i} a^{(i)}(s_{-i}^* | s_i) = a^{(i)}(s_{-i}^* | \pi_{i\alpha^*}) = \max_\alpha a^{(i)}(s_{-i}^* | \pi_{i\alpha})$$

ist. Das Maximum wird also auch für eine reine Strategie angenommen. (Daraus folgt natürlich nicht, daß auch $s_{-i}^* | \pi_{i\alpha^*}$ ein NASH-Gleichgewicht sein muß: Für die anderen Spieler ist die Maximalitätsbedingung möglicherweise dafür nicht mehr erfüllt.)

Mit den neuen Funktionen

$$a_\alpha^{(i)} \colon \begin{cases} S_1 \times \cdots \times S_n \to \mathbb{R} \\ \quad (s_1, \ldots, s_n) \mapsto s^{(i)}(s_{-i} | \pi_{i\alpha}) \end{cases}$$

gilt dann: $s^* = (s_1^*, \ldots, s_n^*)$ ist genau dann ein NASH-Gleichgewicht, wenn

$$a^{(i)}(s^*) = \max_\alpha a_\alpha^{(i)}(s) \quad \text{für alle } \alpha\,.$$

Da die Bedingung aus Gleichungen mit stetigen Funktionen besteht, folgt daraus insbesondere, daß die Menge aller NASH-Gleichgewichte in $S_1 \times \cdots \times S_n$ abgeschlossen ist.

Ist $s_i^* = \sum_\alpha p_{i\alpha}\pi_{i\alpha}$, so ist $a^{(i)}(s^*) = \sum_\alpha p_{i\alpha}a_\alpha^{(i)}(s^*)$. In einem NASH-Gleichgewicht muß daher $p_{i\alpha}$ verwschwinden, falls $a_\alpha^{(i)}(s^*)$ kleiner ist als $\max_\beta a_\beta^{(i)}(s^*)$. Ein NASH-Gleichgewicht benutzt also nur die reinen Strategien, die für den jeweiligen Spieler optimal sind unter den vorgegebenen Strategien der übrigen Spieler. Ein Gleichgewicht kann daher auch so charakterisiert werden, daß $a_\alpha^{(i)}(s^*) = \max_\beta a_\beta^{(i)}(s^*)$ ist für alle reinen Strategien α, die in s^* benutzt werden.

Zum Nachweis der Existenz von Gleichgewichten betrachtet NASH nun die stetigen Funktionen

$$\varphi_{i\alpha}\colon \begin{cases} S_1 \times \cdots \times S_n \to \mathbb{R} \\ \quad (s_1, \ldots, s_n) \mapsto \max\{0, a_\alpha^{(i)}(s) - a^{(i)}(s)\} \end{cases}$$

sowie

$$T\colon \begin{cases} S_1 \times \cdots \times S_n \to S_1 \times \cdots \times S_n \\ s = (s_1, \ldots, s_n) \mapsto s' = (s_1', \ldots, s_n') \end{cases}$$

mit

$$s_i' = \frac{s_i + \sum_\alpha \varphi_{i\alpha}\pi_{i\alpha}}{1 + \sum_\alpha \varphi_{i\alpha}(s)}\,.$$

s_i' liegt in der Tat in S_i: Zunächst kann kein Koeffizienten einer reiner Strategie negativ werden, da die Funktionen $\varphi_{i\alpha}$ nur Werte größer oder gleich Null annehmen, und der Nenner sorgt dafür, daß die Summe aller Koeffizienten gleich eins ist.

Jedes der Simplizes S_i ist homöomorph zu einer Würfel der entsprechenden Dimension, also ist auch das Produkt aller S_i homöomorph zu einem Würfel und damit zu einer Vollkugel. Damit garantiert uns der BROUWERsche Fixpunktsatz die Existenz eines Fixpunkts von T.

NASH zeigt nun, daß jeder solche Fixpunkt ein Gleichgewicht ist:

Unter den reinen Strategien $\pi_{i\alpha}$, die der i-te Spieler für s_i benutzt, gibt es für jedes n-tupel s von Strategien der n Spieler stets welche,

für die $a_\alpha^{(i)}(s) \leq a^{(i)}(s)$ ist, denn wäre $a_\alpha^{(i)}(s) > a^{(i)}(s)$ für alle $\pi_{i\alpha}$, so wäre auch $a^{(i)}(s) > a^{(i)}(s)$, da die Komponente s_i von s ja eine konvexe Linearkombination $\sum_\alpha p_{i\alpha}\pi_{i\alpha}$ ist. Für reine Strategien mit $a_\alpha^{(i)}(s) \leq a^{(i)}(s)$ ist $\varphi_{i\alpha}(s) = 0$. Der Koeffizient von $\pi_{i\alpha}$ in s_i' für so eine Strategie ist damit gleich dem entsprechenden Koeffizienten in s_i dividiert durch den Nenner von s_i'. Für einen Fixpunkt s muß daher wegen $T(s) = s$ der Nenner gleich eins sein, d.h. alle $\varphi_{i\alpha}(s)$ müssen verschwinden. Somit ist $a_\alpha^{(i)}(s) \leq a^{(i)}(s)$ für alle α. Dies zeigt, daß jeder Fixpunkt von T ein NASH-Gleichgewicht ist.

Ist umgekehrt s^* ein NASH-Gleichgewicht, so ist $a_\alpha^{(i)}(s) \leq a^{(i)}(s)$ für alle α, also $\varphi_{i\alpha}(s) = 0$ für alle i und α und damit $T(s) = s$.

Somit ist jeder der nach BROUWER existierenden Fixpunkte von T ein NASH-Gleichgewicht, womit die Existenz solcher Gleichgewichte bewiesen ist.

■

NASH benutzt im wesentlichen das gleiche Argument, um in derselben Arbeit auch die Existenz symmetrischer Gleichgewichte zu beweisen. Um solche zu definieren, bezeichnen wir die Menge aller reinen Strategien des i-ten Spielers mit R_i, und R sei die (disjunkte) Vereinigung der R_i.

Definition: Eine *Symmetrie* oder ein *Automorphismus* eines Spiels ist eine bijektive Abbildung $\varphi\colon R \to R$ mit folgenden Eigenschaften:

1) Es gibt eine Permutation $\psi\colon \{1,\ldots,n\} \to \{1,\ldots,n\}$ der Menge aller Spieler derart, daß φ die Menge R_i auf $R_{\psi(i)}$ abbildet.
2. Für ein n-tupel $(\pi_{1\alpha_1},\ldots,\pi_{n\alpha_n})$ von reinen Strategien enthält die Menge der $\varphi(\pi_{i\alpha_i})$ wegen 1) für jeden der n Spieler genau eine Strategie. Ordnen wir diese Strategien nach Spielern, erhalten wie somit eine Abbildung

$$\chi\colon R_1 \times \cdots \times R_n \to R_1 \times \cdots \times R_n\,.$$

Dabei soll für jeden Spieler i und jedes n-tupel $\xi \in R_1 \times \cdots \times R_n$ gelten: $a^{\left(\psi(i)\right)}\left(\chi(\xi)\right) = a^{(i)}(\xi)$.

χ läßt sich linear fortsetzen auf die gemischten Strategien; wir erhalten somit eine ebenfalls mit χ bezeichnete Abbildung

$$\chi\colon \begin{cases} S_1 \times \cdots \times S_n \to S_1 \times \cdots \times S_n \\ (s_1, \ldots, s_n) \mapsto (s'_1, \ldots, s'_n) \end{cases},$$

wobei für $s_i = \sum_\alpha p_{i\alpha}\pi_{i\alpha}$ gilt: $s_{\psi(i)} = \sum_\alpha p_{i\alpha}\varphi(\pi_{i\alpha})$. Entsprechend läßt sich auch φ fortsetzen zu einer Abbildung

$$\varphi\colon S_1 \cup \cdots \cup S_n \to S_1 \cup \cdots \cup S_n\,,$$

die $\sum_\alpha p_{i\alpha}\pi_{i\alpha}$ abbildet auf $\sum_\alpha p_{i\alpha}\varphi(\pi_{i\alpha})$.

Definition: Ein n-tpel $s = (s_1, \ldots, s_n) \in S_1 \times \cdots \times S_n$ heißt *symmetrisch,* falls $\chi(s) = s$ für jede Symmetrie φ des Spiels.

Man beachte, daß nicht jedes Spiel eine von der Identität verschiedene Symmetrie hat: Zunächst können wegen Bedingung 1) Strategien eines Spielers i nur abgebildet werden auf Strategien eines Spielers j, der genauso viele reine Strategien hat wie i, und auch dann muß noch beachtet werden, daß sich bei Anwendung der Symmetrie nichts an den Auszahlungen ändert.

Satz: Jedes (endliche) Spiel hat mindestens ein symmetriesches Gleichgewicht.

Beweis: Für jeden Spieler i sei s_{i0} die gemischte Strategie, die jede reine Strategie $\pi_{i\alpha}$ mit gleicher Wahrscheinlichkeit einsetzt. Mit obigen Bezeichnungen ist also

$$s_{i0} = \sum_\alpha \frac{1}{\#R_i}\pi_{i\alpha}\,.$$

Für jede Symmetrie φ des Spiels ist dann $\varphi(s_{i0}) = s_{\psi(i)0}$, und für $s_0 = (s_{10}, \ldots, s_{n0})$ ist $\chi(s_0) = s_0$. Dies zeigt, daß es mindestens ein symmetrisches n-tupel gibt.

Sind s und t zwei symmetrische n-tupel und ist $\lambda \in [0,\ 1]$, so folgt aus den Gleichungen $\varphi(s_i) = s_{\psi(i)}$ und $\varphi(t_i) = t_{\psi(i)}$, daß

$$\varphi\big(\lambda s_i + (1-\lambda)t_i\big) = \lambda\varphi(s_i) + (1-\lambda)\varphi(t_i) = \lambda s_{\psi(i)} + (1-\lambda)t_{\psi(i)}$$

ist und damit $\chi\big(\lambda s + (1-\lambda)t\big) = \lambda s + (1-\lambda)t$.

Somit ist die Menge M aller symmetrischer n-tupel konvex, und mit der oben definierten Abbildung T ist $\chi\big(T(s)\big) = T\big(\chi(s)\big)$ für alle s aus $S_1 \times \cdots \times S_n$. Für symmetrische s ist $\chi(s) = s$, also $\chi\big(T(s)\big) = T(s)$, so daß auch $T(s)$ symmetrisch ist. T läßt sich also einschränken auf eine Abbildung $M \to M$, die natürlich weiterhin stetig ist. Da die Symmetriebedingung eine Gleichheit stetiger Funktionen ist, definiert sie eine abgeschlossene Teilmenge von $S_1 \times \cdots \times S_n$, die somit kompakt und – wie wir gerade gesehen haben – konvex ist. Insbesondere ist sie auch zusammenziehbar und hat daher nach dem BROUWERschen Fixpunktsatz mindestens einen Fixpunkt. Jeder solche Fixpunkt ist ein symmetrisches Gleichgewicht.

■

Noch ein weiteres Problem schneidet NASH in seiner zweiten Arbeit an: Wie am Ende von Kapitel 9 bemerkt, wissen die Spieler bei mehreren Gleichgewichten nicht, auf welches Gleichgewicht sie hinarbeiten sollten. Bei manchen Spielen ist das allerdings kein Problem:

Definition: *a)* Ein endliches Spiel heißt *lösbar,* wenn die Menge S^* seiner NASH-Gleichgewichte die folgende *Austauschbedingung* erfüllt: Ist $s^* \in S^*$ und gibt es zu $s_i \in S_i$ ein $t \in S = S_1 \times \cdots \times S_n$, für das $(t_{-i}|s_i)$ in S^* liegt, so liegt auch $(s^*_{-i}|s_i)$ in S^*. Die Menge S^* wird dann als die Lösung des Spiels bezeichnet.
b) Eine Strategie $s_i \in S_i$ heißt *Gleichgewichtsstrategie,* wenn es ein $t \in S$ gibt, für das $(t_{-i}|s_i)$ ein NASH-Gleichgewicht ist. *c)* Ein lösbares Spiel heißt *strikt lösbar,* wenn seine Lösung S^* noch folgende Zusatzbedingung erfüllt: Ist $s^* \in S^*$ und $s_i \in S_i$ so, daß $a^{(i)}(s^*_{-i}|s_i) = a^{(i)}(s^*)$ ist, so liegt auch $(s^*_{-i}|s_i)$ in S^*

Weniger formal ausgedrückt: Kommt s_i in einem lösbaren Spiel in *irgendeinem* NASH-Gleichgewicht vor, so kann man in *jedem* NASH-Gleichgewicht die Strategie des i-ten Spielers durch s_i ersetzen, ohne daß die Gleichgewichtseigenschaft verloren geht.

NASH zeigt nun, daß bei einem lösbaren Spiel genau dann ein Gleichgewicht entsteht, wenn jeder Spieler eine Gleichgewichtsstrategie wählt. Da nicht jedes Spiel lösbar ist, zeigt er dies etwas allgemeiner:

Definition: Eine maximale Teilmenge T von S^*, die die Austauschbedingung erfüllt, heißt eine Teillösung. Die i-te Faktormenge von T ist die Menge aller $s_i \in S_i$, für die es ein $t \in S$ gibt, so daß $(t_{-i}|s_i$ in T liegt.

Für ein lösbares Spiel ist damit die Lösung S^* die einzige Teillösung, und deren Faktormengen bestehen genau aus den Gleichgewichtslösungen.

Satz: Jede Teillösung ist das Produkt ihrer Faktormengen.

Beweis: Für $(s_1, \ldots, s_n) \in T$ kann natürlich jedes s_i zu einem Tupel aus T ergänzt werden, d.h. $s_i \in T_i$.

Umgekehrt sei $(s_1, \ldots, s_n) \in S$ und $s_i \in T_i$ für alle i. Da s_1 in T_1 liegt, gibt es ein $t^{(0)} \in T$, so das

$$t^{(1)} = (t^{(0)}_{-1}|s_1) = (s_1, t^{(0)}_2, \ldots t^{(0)}_n) \in T\,.$$

Da s_2 in T_2 liegt, folgt dann auch

$$t^{(2)} = (t^{(1)}_{-2}|s_2) = (s_1, s_2, t^{(0)}_3, \ldots t^{(0)}_n) \in T\,,$$
$$t^{(3)} = (t^{(2)}_{-3}|s_3) = (s_1, s_2, s_3, t^{(0)}_3, \ldots t^{(0)}_n) \in T\,,$$

und so weiter, bis $t^{(n)} = (t^{(n-1)}_{-n}|s_n) = (s_1, s_2, \ldots, s_n) \in T$. ∎

§2: Andere Beweise des Fixpunktsatzes von Brouwer

In Kapitel 8 hatten wir den Fixpunktsatz von BROUWER als Korollar des Fixpunktsatzes von LEFSCHETZ erhalten. Dieser Zugang hat den Vorteil, daß wir den Fixpunktsatz für beliebige zusammenziehbare triangulierbare topologische Räume erhalten, setzt aber den ganzen Apparat der Homologietheorie voraus. Viele andere Beweise funktionieren nur für kompakte konvexe Teilmengen eines $\mathbb{R}^n$, was aber für die meisten Anwendungen genügt.

BROUWER veröffentlichte seinen Satz 1911, im Jahr, in dem LEFSCHETZ über ein Thema aus der algebraischen Geometrie promovierte. Sein Fixpunktsatz erschien erst 1923, als sich BROUWER schon fast ausschließlich mit den Grundlagen der Mathematik beschäftigte. Der

BROUWERsche Beweis konnte diesen Satz daher unmöglich verwenden. Tatsächlich arbeitete BROUWER mit dem Approximationssatz von STONE-WEIERSTRASS, über den er sich auf differenzierbare Funktionen beschränken konnte.

Inzwischen gibt es eine ganze Reihe verschiedener Beweise. Spezialfälle waren auch schon vor BROUWER bekannt, und auch BROUWER selbst veröffentlichte 1909 zunächst nur einen Beweis für Teilmengen von $\mathbb{R}^3$.

Völlig elementar ist der Satz für Teilmengen von $\mathbb{R}$:

Satz: Ist X eine kompakte konvexe Teilmenge von $\mathbb{R}$, so hat jede stetige Abbildung $f\colon X \to X$ mindestens einen Fixpunkt.

Beweis: Jede konvexe Teilmenge von $\mathbb{R}$ ist ein (nicht notwendigerweise endliches) Intervall; falls es kompakt ist, muß es endlich und abgeschlossen sein, also von der Form $X = [a, b]$. Zu einer stetigen Abbildung $f\colon [a, b] \to [a, b]$ betrachten wir die Abbildung

$$g\colon \begin{cases} [a, b] \to \mathbb{R} \\ \quad x \mapsto f(x) - x \end{cases}$$

Da das Bild von f in $[a, b]$ liegt, ist insbesondere $f(a) \geq a$ und $f(b) \leq b$, also $g(a) \geq 0$ und $g(b) \leq 0$. Da g offensichtlich stetig ist, gibt es daher nach dem Zwischenwertsatz mindestens ein $x \in [a, b]$ mit $g(x) = 0$, d.h. $f(x) = x$.

■

Ein relativ elementarer Beweis des BROUWERschen Fixpunktsatzes, der ohne Homologie auskommt, beruht auf einem Satz von SPERNER, der eng verwandt ist mit dem, was wir in Kapitel 6 mit SPERNER-Abbildungen gemacht haben.

Definition: K sei ein geometrischer simplizialer Komplex, dessen topologischer Raum $|K|$ ein n-Simplex s mit Ecken $v_0, \ldots, v_n$ sei. Eine *Markierung* von K ist eine Abbildung $\lambda\colon K_0 \to \{0, \ldots, n\}$ mit der Eigenschaft, daß $v_{\lambda(v)}$ für jede Ecke $v \in K_0$ eine Ecke des kleinsten Randsimplex von s ist, das v enthält. Anders ausgedrückt: Stellt man v als konvexe Linearkombination der Ecken von s dar, darf der Koeffizient von $v_{\lambda(v)}$ nicht verschwinden. Ein Simplex aus K heißt *vollständig*

markiert, wenn λ auf der Menge seiner Ecken alle Werte von 0 bis n annimmt.

Lemma von Sperner: Die Anzahl der vollständig markierten Simplizes von K ist ungerade. Insbesondere gibt es also solche Simplizes.

Der folgende einfache *Beweis* durch Induktion nach n wurde 1968 von KUHN veröffentlicht: Für $n = 0$ bestehen sowohl s als auch K nur aus der einen Ecke v_0, so daß die Behauptung trivial ist.

Für $n > 0$ nehmen wir an, der Satz sei für $n - 1$ bewiesen. Wir nennen ein Simplex aus K *fast vollständig markiert,* wenn λ auf seinen Ecken alle Werte von 0 bis $n - 1$ annimmt, und betrachten folgende Mengen:

C ist die Menge aller vollständig markierter n-Simplizes

A ist die Menge aller fast vollständig markierter n-Simplizes

B ist die Menge aller fast vollständig markierter $(n-1)$-Simplizes, die auf dem Rand von s liegen

E ist die Menge aller fast vollständig markierter $(n - 1)$-Simplizes

Sei $D = A \cup B \cup C$. Wir definieren einen (abstrakten) Graphen mit D als Eckenmenge und E als Kantenmenge. Dabei soll $d \in A \cup C$ genau dann eine Ecke von $e \in E$ sein, denn e ein Randsimplex von d ist, und $d \in B$ ist genau dann Ecke von e, wenn $d = e$ ist. Die Kanten e sind fast vollständig markierte $(n - 1)$-Simplizes. Falls e auf dem Rand von s liegt, ist es Randsimplex genau eines n-Simplex aus K, das entweder vollständig oder fast vollständig markiert ist; dieses ist die eine Ecke von e. Die andere Ecke ist e selbst, aufgefaßt als Element von D. Falls e nicht auf dem Rand liegt, ist es gemeinsamer Rand zweier vollständig oder fast vollständig markierter n-Simplizes, die somit seine beiden Ecken sind. Damit hat jede Kante genau zwei Ecken; wir haben also wirklich einen Graphen.

Der Grad eines Ecke eines Graphen ist bekanntlich die Anzahl der Kanten, die von dieser Ecke ausgehen. Für ein Simplex $d \in A$ haben zwei Ecken von d die gleiche Markierung, so daß es genau zwei Randsimplizes gibt, die in E liegen: Man lasse jeweils eine der beiden gleich markierten Ecken weg. Der Grad einer solchen Ecke ist also zwei.

Für $d \in C$ gibt es nur ein Randsimplex in E: Weggelassen wird die mit n markierte Ecke, der Grad ist also eins. Dies gilt auch für $d \in B$, denn dann ist ja d selbst das Simplex aus E.

Da jede Kante zwei Ecken hat, ist die Summe der Grade aller Ecken gleich der doppelten Kantenzahl, d.h.

$$2\#A + \#B + \#C = 2\#E\,.$$

Somit ist $\#B + \#C$ gerade.

Nun wenden wir die Induktionsvoraussetzung an: B ist die Menge aller fast vollständig markierter $(m-1)$-Simplizes auf dem Rand. Eine Ecke v eines solchen Simplex kann durch die Markierung nur auf eine Ecke v_i abgebildet werden, die in der baryzentrischen Darstellung von v einen positiven Koeffizienten hat. Da das Simplex fast vollständig markiert ist, gibt es in diesem Simplex daher Punkte, für die keiner der Koeffizienten von $v_0, \ldots, v_{n-1}$ verschwindet. Da das Simplex auf dem Rand liegt, muß dann der Koeffizient von v_n verschwindet, d.h. das Simplex liegt im Randsimplex s' mit Ecken $v_0, \ldots, v_{n-1}$. Die Simplizes aus B liegen somit alle in dem Unterkomplex K' von K, der nur aus den in s' liegenden Simplizes aus K besteht. Dieser ist eine Triangulierung des abgeschlossenen Simplex mit Ecken $v_0, \ldots, v_{n-1}$, und als Simplex aus K' ist ein Simplex genau dann vollständig markiert, wenn es als Simplex aus K fast vollständig markiert ist. B besteht also genau aus den vollständig markierten Simplizes von K', und deren Anzahl ist nach Induktionsannahme ungerade. Da $\#B + \#C$ gerade ist, muß dann auch C eine ungerade Anzahl von Elementen enthalten, und genau das war die Behauptung des Satzes.

■

Daraus folgt nun recht einfach der BROUWERschen Fixpunktsatz für Simplizes:

Satz: Jede stetige Abbildung eines abgeschlossenen n-Simplex X auf sich selbst hat einen Fixpunkt.

Beweis: X sei das Simplex mit Ecken $v_0, \ldots, v_n$, und die Funktionen $f_i \colon X \to [0,\ 1]$ seien so definiert, daß für $x = \lambda_0 v_0 + \cdots + \lambda_n v_n$

$$f(x) = f_0(x)v_0 + \cdots + f_n(x)v_n$$

ist. Zu jedem $\varepsilon > 0$ können wir eine Triangulierung K von X mit Maschenweite kleiner ε finden. Wir definieren eine Markierung dadurch, daß wir jeder Ecke $v = \lambda_0 v_0 + \cdots + \lambda_n v_n$ von K ein $i \in \{0, \ldots, n\}$ zuordnen, für das λ_i nicht verschwindet und $f_i(v) \leq \lambda_i$ ist. Ein solcher Index i existiert, denn andernfalls wäre $f_i(v) > \lambda_i$ für alle i mit $\lambda_i \neq 0$, d.h.

$$1 = \sum_{i=0}^{n} f_i(v) \geq \sum_{i \text{ mit } \lambda_i \neq 0} f_i(v) > \sum_{i \text{ mit } \lambda_i \neq 0} \lambda_i = \sum_{i=0}^{n} \lambda_i = 1\,,$$

ein Widerspruch. Somit definiert λ eine Markierung von K, und nach dem Lemma von SPERNER gibt es mindestens ein vollständig markiertes n-Simplex $s^{(\varepsilon)}$ in K. Wir nummerieren dessen Ecken so, daß der Index der Ecke gleich ihrer Markierung ist. Für jede Ecke

$$w_i^{(\varepsilon)} = \lambda_0 v_0 + \cdots + \lambda_n v_n$$

von $s^{(\varepsilon)}$ ist $f_i\big(w_i^{(\varepsilon)}\big) \leq \lambda_i$. Da X kompakt ist, können wir für eine Nullfolge $(\varepsilon_k)_{k \in \mathbb{N}}$ eine Teilfolge finden derart, daß die entsprechende Teilfolge der Folge der $w_0^{(\varepsilon_k)}$ konvergiert. Dazu können wir wieder eine Teilfolge finden, so daß auch die entsprechende Teilfolge der $w_1^{(\varepsilon_k)}$ konvergiert, und so weiter. Dies füht auf eine Teilfolge $(\varepsilon_{k_\nu})_{\nu \in \mathbb{N}}$ mit der Eigenschaft, daß alle Folgen $w_i^{(\varepsilon_{k_\nu})}$ konvergieren. Da die Ecken $w_i^{(\varepsilon_{k_\nu})}$ des Simplex $s^{(\varepsilon_{k_\nu})}$ mit Durchmesser höchstens ε_{k_ν} sind und die ε_{k_ν} eine Nullfolge bilden, konvergieren alle diese $n + 1 >$ Folgen gegen einen gemeinsamen Grenzwert $x = \lambda_0 v_0 + \cdots + \lambda_n v_n \in X$. Wegen der Stetigkeit von f ist auch $f_i(x) \leq \lambda_i$ für alle i. Da aber sowohl die Summe der $f_i(x)$ als auch die der λ_i gleich eins ist, muß $f_i(x) = \lambda_i$ sein, d.h. $f(x) = x$.

■

Korollar: Ist der topologische Raum X homöomorph zu einem abgeschlossenen Simplex, so hat jede stetige Abbildung $f\colon X \to X$ mindestens einen Fixpunkt.

Beweis: S sei ein Simplex und $g\colon X \to S$ sei ein Homöomorphismus. Dann ist $g \circ f \circ g^{-1}$ eine stetige Abbildung von S nach S, hat also nach dem gerade bewiesenen Satz einen Fixpunkt z. Für $x = g^{-1}(z)$ ist dann $g\big(f(x)\big) = z$, also $f(x) = g^{-1}(z) = x$.

■

Korollar: Ist $X \subset \mathbb{R}^n$ kompakt und konvex, so hat jede stetige Abbildung $f\colon X \to X$ mindestens einen Fixpunkt.

Beweis: Als kompakte Menge ist X beschränkt, liegt also in einem hinreichend großen abgeschlossenen Simplex $S \subset \mathbb{R}^n$. Für $z \in S$ sei

$$d(z, X) = \min_{x \in X} d(z, x)\,,$$

wobei $d(z, x)$ den EUKLIDischen Abstand der Punkte z und x bezeichnet. Wegen der Kompaktheit von X existiert dieses Minimum, es gibt also mindestens einen solchen Punkt $x \in X$. Gäbe es zwei verschiedene solche Punkte, so läge wegen der Konvexität von X auch deren Verbindungsstrecke in X, und wegen der Dreiecksungleichung hätte der Mittelpunkt dieser Strecke einen kleineren Abstand von z als die beiden Endpunkte. Dies widerspricht der Minimalität.

Somit gibt es zu jedem $z \in S$ genau einen Punkt $x \in X$ mit minimalem Abstand zu z. Wir bezeichnen diesen Punkt mit $h(z)$ und definieren so eine Abbildung $h\colon S \to X$. Wegen der Stetigkeit des EUKLIDischen Abstands ist auch h stetig, und für $x \in X$ ist natürlich $h(x) = x$.

Da X in S liegt, kann die Hintereinanderausführung $f \circ h\colon S \to X$ auch als Abbildung von S nach S betrachtet werden, hat also einen Fixpunkt x. Für diesen ist $x = (f \circ h)(x) = f\big(h(x)\big) \in X$, also $h(x) = x$ und damit $f(x) = x$.

■

Auch diese Aussage läßt sich noch etwas verallgemeinern:

Definition: *a)* X sei ein topologischer Raum und $Z \subset X$. Eine stetige Abbildung $r\colon X \to Z$ heißt *Retraktion,* wenn $r(z) = z$ ist für alle $z \in Z$, und Z heißt ein *Retrakt* von X.
b) Allgemeiner seien X, Z zwei topologische Räume. Dann heißt Z ein r-Bild von X, wenn es stetige Abbildungen $r\colon X \to Z$ und $g\colon Z \to X$ gibt, so daß $r \circ g$ die Identität auf Z ist.

Satz: Der topologische Raum Z sei das r-Bild einer kompakten konvexen Teilmenge von $\mathbb{R}^n$. Dann hat jede stetige Abbildung $f\colon Z \to Z$ mindestens einen Fixpunkt.

Beweis: Mit den Bezeichnungen aus der Definition ist $g \circ f \circ r$ eine stetige Abbildung von X nach X, hat also einen Fixpunkt x. Für $z = r(x)$ ist dann $(g \circ f)(z) = x$, also $(r \circ g \circ f)(z) = r(x) = z$. Nach Definition des r-Bilds ist aber $r \circ g$ die Identität auf z, d.h. $f(z) = z$. ∎

Lemma: $\mathbb{S}^{n-1}$ ist kein r-Bild und damit insbesondere auch kein Retrakt von $\mathbb{B}^n$.

Beweis: Andernfalls gäbe es stetige Abbildungen $r\colon \mathbb{B}^n \to \mathbb{S}^{n-1}$ und $g\colon \mathbb{S}^{n-1} \to \mathbb{B}^n$, so daß $r \circ g$ die Identität auf $\mathbb{S}^{n-1}$ wäre. Wir identifizieren $\mathbb{B}^n$ mit der Kugel mit Radius eins um den Nullpunkt des $\mathbb{R}^n$ und können damit für einen Punkt $x \in \mathbb{B}^n$ vom Punkt $-x \in \mathbb{B}^n$ reden. Für $x \in \mathbb{B}^n$ liegt auch $f(x) \underset{\text{def}}{=} g\big(-r(x)\big)$ in $\mathbb{B}^n$, so daß wir eine stetige Abbildung $f\colon \mathbb{B}^n \to \mathbb{B}^n$ haben. Diese hat mindestens einen Fixpunkt x; es gibt also einen Punkt x, für den $g\big(-r(x)\big) = x$ ist und damit $(r \circ g)\big(-r(x)\big) = r(x)$. Da $r \circ g$ die Identität auf $\mathbb{S}^{n-1}$ ist, ist dies andererseits auch gleich $-r(x)$, d.h. $r(x) = -r(x) = 0$. Da der Nullpunkt nicht in $\mathbb{S}^{n-1}$ liegt, ist das ein Widerspruch. ∎

Umgekehrt läßt sich aus diesem Lemma auch der Fixpunktsatz von BROUWER herleiten. Um das zu sehen, betrachten wir die Situation gleich etwas allgemeiner und erinnern uns an die folgende Definition aus der Analysis:

Definition: Der Rand ∂X einer Teilmenge $X \subset \mathbb{R}^n$ ist die Menge aller Punkte $x \in \mathbb{R}^n$ mit der Eigenschaft, daß jede Umgebung von x sowohl Punkte aus X als auch solche aus $\mathbb{R}^n \setminus X$ enthält.

Lemma: $X \subset \mathbb{R}^n$ sei kompakt und konvex, und $f\colon X \to X$ sei eine stetige Abbildung. Falls f keinen Fixpunkt hat, gibt es eine Retraktion $r\colon X \to \partial X$.

Beweis: Für jeden Punkt $x \in X$ ist $f(x) \neq x$. Es gibt daher genau eine Gerade durch x und $f(x)$. Ihr Durchschnitt mit X ist als Durchschnitt zweier konvexer Mengen selbst konvex und als abgeschlossene Teilmenge des kompakten Raums X auch kompakt. Somit ist dieser

Durchschnitt eine Strecke, deren Endpunkte offensichtlich auf dem Rand von X liegen. Auf dieser Strecke wandern wir von $f(x)$ nach x und weiter, bis wir zu einem der beiden Endpunkte kommen. Diesen Endpunkt definieren wir als $r(x)$.

Falls $x \in \partial X$, ist $r(x) = x$, denn dann können wir nicht über x hinauswandern, ohne X zu verlassen. Außerdem folgt aus der Stetigkeit von f die von r. Somit ist r eine Retraktion.

■

Auch dieser Satz liefert als Korollar den Fixpunktsatz für Teilmengen von $\mathbb{R}$, denn hier ist $X = [a, b]$ und $\partial X = \{a, b\}$. Falls es eine stetige Abbildung $X \to X$ ohne Fixpunkt gäbe, gäbe es also auch eine Retraktion $r\colon [a, b] \to \{a, b\}$ mit $r(a) = a$ und $r(b) = b$. Insbesondere wäre das Bild unter r der zusammenhängenden Menge $[a, b]$ die unzusammenhängende Menge $\{a, b\}$, was natürlich nicht möglich ist.

■

Wenn wir zusätzlich den Apparat der Homologietheorie einsetzen, bekommen wir auch einen schnellen Beweis des Fixpunktsatzes für die Vollkugel:

Satz: Jede stetige Abbildung $f\colon \mathbb{B}^n \to \mathbb{B}^n$ hat mindestens einen Fixpunkt.

Beweis: Den Fall $n = 1$ haben wir bereits erledigt, und der Fall $n = 0$ ist ohnehin trivial. Sei daher $n \geq 2$. Falls es eine Abbildung f ohne Fixpunkt gibt, ist $\mathbb{S}^{n-1}$ ein Retrakt von $\mathbb{B}^n$, es gibt also eine stetige Abbildung $r\colon \mathbb{B}^n \to \mathbb{S}^{n-1}$, so daß für die Einbettung $g\colon \mathbb{S}^{n-1} \to \mathbb{B}^n$ gilt: $r \circ g = \mathrm{id}_{\mathbb{S}^{n-1}}$. Damit ist auch $r_{*(n-1)} \circ g_{*(n-1)}$ die Identität auf $H_{n-1}(\mathbb{S}^{n-1}) \cong \mathbb{Z}$. Nun ist aber $H_{n-1}(\mathbb{B}^n) = 0$, so daß $g_{*(n-1)}\colon H_{n-1}(\mathbb{S}^{n-1}) \to H_{n-1}(\mathbb{B}^n)$ und damit auch $r_{*(n-1)} \circ g_{*(n-1)}$ alles auf Null abbilden muß, ein Widerspruch.

■

Damit folgt auch der Fixpunktsatz für kompakte konvexe Teilmengen von $\mathbb{R}^n$ aus dem folgenden

Lemma: Zu jeder kompakten konvexen Teilmenge $X \subset \mathbb{R}^n$ gibt es ein $m \leq n$, so daß X homöomorph zu $\mathbb{B}^m$ ist.

Beweis: Wieder ist der Fall $n = 0$ trivial. Für $n > 0$ können wir einen inneren Punkt $x_0 \in X$ wählen, und wegen der Konvexität von X liegt für jeden Punkt $x \in X$ die Verbindungsstrecke von x_0 nach x in X. In der Menge aller Vektoren mit Anfangspunkt x_0 und Endpunkt $x \in X$ gibt es eine maximale linear unabhängige Teilmenge; diese bestehe aus den Vektoren $v_1, \ldots, v_m$. Natürlich kann m höchstens gleich n sein.

W sei der von $v_1, \ldots, v_m$ aufgespannte Vektorraum, und $\mathbb{B}^m$ sei realisiert als die Menge aller Vektoren mit EUKLIDischer Norm höchstens eins in W. Wir wollen uns überlegen, daß X homöomorph zu $\mathbb{B}^m$ ist.

Dazu definieren wir zunächst eine Abbildung $f\colon X \to \mathbb{B}^m$ wie folgt: Der Punkt x_0 wird auf den Nullvektor abgebildet. Für einen Punkt $x \neq x_0$ betrachten wir wie oben im Beweis, daß wir mit einer fixpunktfreien stetigen Abbildung eine Retraktion auf den Rand konstruieren können, die Gerade durch x_0 und x und deren Durchschnitt mit X. Wegen der Kompaktheit von X ist dieser kompakt, und wegen der Konvexität sowohl der Geraden als auch von X ist er auch konvex, also eine Strecke. Deren Endpunkte liegen auf dem Rand von X. Der Abstand von x_0 zu dem Endpunkt x_1, zu dem wir kommen, wenn wir in Richtung x laufen, sei a; der Abstand von x_0 und x sei d. Wir definieren $f(x)$ als d/a mal dem Vektor von x_0 nach x_1. Wegen der Stetigkeit der Abstandsfunktion ist f stetig. Außerdem liegt die gesamte Strecke von x_0 nach x_1 in X und wird bijektiv auf die Menge aller Vektoren λv mit $v = f(x_1)$ und $\lambda \in [0, 1]$ abgebildet. Diese Vektoren liegen allesamt in $\mathbb{B}^m$, da jeder Vektor, der x_0 mit einem Punkt $x \neq x_0$ aus X verbindet, als Linearkombination der Vektoren v_i geschrieben werden kann.

Die Injektivität von f ist klar; zum Beweis der Surjektivität beachten wir, daß jeder Vektor $v \neq 0$ aus W eine nichttriviale Linearkombination der Vektoren v_i ist. Da x_0 als innerer Punkt gewählt war, gibt es zu jedem i auch ein Vielfaches $\lambda \cdot (-v_i)$ mit hinreichend kleinem λ, so daß $\lambda \cdot (-v_i)$ Verbindungsvektor von x_0 mit einem Punkt $x \in X$ ist. Ersetzen wir in der Darstellung von v alle v_i, die einen negativen Koeffizienten haben, durch ein solches Vielfaches, haben wir eine Darstellung von v als Linearkombination mit nichtnegativen Koeffizienten. Dividieren wir in dieser Darstellung alles durch die Summe aller Koeffizienten, erhalten wir ein skalares Vielfaches von v als konvexe Linearkombination der

modifizierten v_i. Dieses Vielfache ist dann der Vektor von x_0 zur entsprechenden konvexen Linearkombination der Endpunkte der v_i, also wegen der Konvexität von X ein Vektor von x_0 zu einem Punkt aus x. Somit liegt der durch v definierte Radius von $\mathbb{B}^m$ im Bild von f.

Bleibt noch zu zeigen, daß auch die Umkehrabbildung von f stetig ist, aber das ist klar, denn auch diese Umkehrabbildung ist auf jedem Radius einfach die Multiplikation mit einer Konstanten, die stetig von der Richtung des Strahls abhängt. Damit ist der Satz bewiesen. ■

Dies zeigt noch einmal, daß jede stetige Abbildung einer kompakten konvexen Teilmenge von $\mathbb{R}^n$ auf sich selbst einen Fixpunkt hat: Wir können entweder wie beim obigen Korollar argumentieren, daß der Übergang zu einem homöomorphen Raum nichts an der Existenz von Fixpunkten ändert, oder wir können aus dem gerade bewiesenen Lemma folgern, daß die Homologiegruppen von X und ∂X isomorph sind zu den entsprechenden Homologiegruppen von $\mathbb{B}^m$ und $\mathbb{S}^{m-1}$ und dann wie im Beweis des letzten Satzes argumentieren.

Dieses Lemma hätten wir natürlich auch oben verwenden können beim Beweis, daß der BROUWERsche Fixpunktsatz für kompakte konvexe Teilmengen von $\mathbb{R}^n$ aus dem für Simplizes folgt, denn wenn ein topologischer Raum homöomorph zu $\mathbb{B}^m$ ist, ist er natürlich auch homöomorph zu einem abgeschlossenen m-Simplex. Der dortige Beweis ist allerdings deutlich kürzer, wenn auch vielleicht etwas trickreicher und weniger anschaulich.

1929 veröffentlichten KNASTER, KURATOWSKI und MAZURKIEWICZ einen weiterer Beweis des BROUWERschen Fixpunktsatzes in

> B. KNASTER, K. KURATOWSKI, S. MUZURKIEWICZ: Ein Beweis des Fixpunktsatzes für n-dimensionale Simplexe, *Fundamenta Mathematica* **14** (1929), 132–137

Ausgangspunkt ist das heute nach ihnen benannte Lemma, das sie zunächst für Simplizes beweisen – was, wie wir gleich sehen werden, für den Beweis des BROUWERschen Fixpunktsatzes ausreicht. Die Verallgemeinerung kann aber oft nützlich sein beim Nachweis der Existenz von ökonomischen Gleichgewichten.

Satz: X sei ein abgeschlossenes n-Simplex in $\mathbb{R}^n$ mit Ecken $v_0, \ldots, v_n$, und $A_0, \ldots, A_m$ seien abgeschlossene Teilmengen von S mit der Eigenschaft, daß für jede Teilmenge $I \subseteq \{0, \ldots, n\}$ gilt: Das Simplex mit Ecken v_i, $i \in I$, liegt in $\bigcup_{i \in I} A_i$. Dann ist $\bigcap_{i=0}^n A_i$ kompakt und nicht leer.

Bemerkung: Die K-K-M-Bedingung, wonach das Simplex mit Ecken v_i, $i \in I$ in der Vereinigung der A_i mit $i \in I$ liegen muß, besagt insbesondere, daß die Vereinigung *aller* A_i das gesamte Simplex X sein muß. Außerdem muß die Ecke v_i in A_i liegen, die Kante mit Ecken v_i und v_j in $A_i \cup A_j$, und so weiter.

Beweis: Als Durchschnitt abgeschlossener Mengen ist $\bigcap_{i=0}^n A_i$ abgeschlossen im kompakten Simplex X, also selbst kompakt.

Um zu sehen, daß er nicht leer ist, wählen wir zu jedem $\varepsilon > 0$ eine Triangulierung K von S mit Maschenweite höchstens ε. Für diesen simplizialen Komplex definieren wir eine Markierung wie folgt:

Die Ecke $v \in K_0$ von K liege im offenen Simplex mit Ecken $v_{i_0}, \ldots, v_{i_r}$. Da dieses nach der K-K-M-Bedingung in der Vereinigung der A_{i_ν} liegt, liegt v in mindestens einer dieser Mengen. Wähle als Markierung von v einen Index i_ν, für den dies der Fall ist.

Nach dem Lemma von SPERNER gibt es in K mindestens ein vollständig markiertes Simplex, d.h. ein n-Simplex mit der Eigenschaft, daß jede der Zahlen von 0 bis n als Markierung einer Ecke auftritt. $s^{(\varepsilon)}$ sei ein solches Simplex. Wir nummerieren seine Ecken $v_0^{(\varepsilon)}, \ldots, v_n^{(\varepsilon)}$ so, daß $v_i^{(\varepsilon)}$ die Markierng i hat. Nach Konstruktion liegt dann $v_i^{(\varepsilon)}$ in A_i.

Nun betrachten wir eine Nullfolge $(\varepsilon_k)_{k \in \mathbb{N}}$ und dazu jeweils die Simplizes $s^{(\varepsilon_k)}$ sowie die Folgen $(v_i^{(\varepsilon_k)})_{m \in \mathbb{N}}$ der Ecken. Da alle Ecken im kompakten Simplex X liegen, hat zunächst die Folge der $v_0^{(\varepsilon_k)}$ eine konvergente Teilfolge. Betrachten wir die Teilfolge von $(v_1^{(\varepsilon_k)})$ mit den entsprechenden Indizes, hat auch diese wieder eine konvergente Teilfolge, und so weiter. Wir erhalten so schließlich eine Teilfolge $(\varepsilon_{k_\nu})_{\nu \in \mathbb{N}}$ mit der Eigenschaft, daß alle $n+1$ Folgen $(v_i^{(\varepsilon_{k_\nu})})_{\nu \in \mathbb{N}}$ konvergieren. Da die Durchmesser der Simplizes $s^{(\varepsilon_{k_\nu})}$ gegen Null gehen, müssen alle

$n+1$ Folgen gegen denselben Punkt $z \in S$ konvergieren. Für jedes i liegen alle $v_i^{(\varepsilon_{k_\nu})}$ in der abgeschlossenen Menge A_i, also liegt auch der Limes z in dieser Menge. Somit liegt z im Durchschnitt aller A_i, der somit nicht leer sein kann.

■

Dies wird verallgemeinert zum

Lemma von Knaster-Kuratowski-Mazurkiewicz: Das Polyeder X sei die konvexe Hülle der $n+1$ Punkte $w_0, \ldots, w_n \in \mathbb{R}^k$, und $A_0, \ldots, A_n$ seien abgeschlossene Teilmengen von X derart, daß für jede Teilmenge $I \subseteq \{0, \ldots, n\}$ gilt: Die konvexe Hülle der Punkte w_i mit $i \in I$ liegt in $\bigcup_{i \in I} A_i$. Dann ist $\bigcap_{i=0}^n A_i$ kompakt und nicht leer.

Beweis: Wieder folgt die Kompaktheit des Durchschnitts aus der von X sowie der Abgeschlossenheit der Mengen A_i. Zu zeigen bleibt, daß der Durchschnitt nicht leer ist.

Dazu sei S ein abgeschlossenes n-Simplex mit Ecken $v_0, \ldots, v_n \in \mathbb{R}^n$. Wir betrachten die Abbildung

$$\varphi\colon \begin{cases} S \to X \\ \displaystyle\sum_{i=0}^n \lambda_i v_i \mapsto \sum_{i=0}^n \lambda_i w_i \end{cases}.$$

Sie ist offensichtlich stetig und auch surjektiv; injektiv ist sie genau dann, wenn die Punkte w_i in allgemeiner Lage sind und wir somit die Situation des gerade bewiesenen Satzes haben.

Für jedes $i \in \{0, \ldots, n\}$ sein $B_i = \varphi^{-1}(A_i)$; wegen der Stetigkeit von φ ist das eine abgeschlossene Teilmenge von S. Für eine Teilmenge $I \subseteq \{0, \ldots, n\}$ wird das Simplex S_I mit den Ecken v_i, $i \in I$, abgebildet auf die konvexe Hülle der Punkte w_i mit $i \in I$, und die liegt nach Voraussetzung in der Vereinigung der A_i mit $i \in I$. Somit liegt S_I im Durchschnitt der B_i mit $i \in I$. Damit können wir den obigen Satz anwenden und wissen, daß es mindestens einen Punkt z im Durchschnitt aller B_i gibt. Sein Bild $\varphi(z)$ liegt im Durchschnitt aller A_i, der somit nicht leer sein kann.

■

Daraus folgt nun der BROUWERsche Fixpunktsatz für Simplizes:

Satz: X sei ein abgeschlossenes n-dimensionales Simplex und $f\colon X \to X$ eine stetige Abbildung. Dann hat f mindestens einen Fixpunkt.

Beweis: Die Ecken von X seien $v_0, \ldots, v_n$, und die $n + 1$ Funktionen $f_i\colon X \to [0,\ 1]$ seien so definiert, daß für $x = \lambda_0 v_0 + \cdots + \lambda_n v_n$ gilt

$$f(x) = f_0(x)v_0 + \cdots + f_n(x)v_n\,.$$

Aus der Stetigkeit von f folgt natürlich sofort die der f_i.

Wir setzen nun $A_i = \{x \in X \mid f_i(x) \leq \lambda_i\}$ für $i = 0, \ldots, n$ und müssen zeigen, daß für jede Teilmenge $I \subseteq \{0, \ldots, m\}$ das Simplex mit Ecken v_i, $i \in I$, in der Vereinigung der A_i mit $i \in I$ liegt.

Angenommen, ein Punkt $x = \lambda_0 v_0 + \cdots + \lambda_n v_n$ aus diesem Simplex läge nicht in dieser Vereinigung. Dann wäre also $f_i(x) > \lambda_i$ für alle $i \in N$ und damit auch $\sum_{i\in I} f_i(x) > \sum_{i\in I} \lambda_i$. Da alle λ_i mit $i \notin I$ verschwinden, ist letztere Summe gleich der Summe über *alle* λ_i, also eins. Da kein $f_i(x)$ negativ werden kann, wäre also

$$1 = \sum_{i=0}^{n} f_i(x) \geq \sum_{i\in I} f_i(x) > \sum_{i\in I} \lambda_i = 1\,,$$

ein offensichtlicher Widerspruch. Somit ist die K-K-M-Bedingung erfüllt, und der obige Satz zeigt, daß der Durchschnitt aller A_i nicht leer ist. Es gibt also mindestens einen Punkt $z = \lambda_0 v_0 + \cdots + \lambda_n v_n$, für den $f_i(z) \leq \lambda_i$ für alle i gilt. Da sowohl die Summe aller $f_i(z)$ als auch die Summe aller λ_i gleich eins ist, geht das nur, wenn $f_i(z) = \lambda_i$ für alle i, d.h. $f(z) = z$.

■

Um zu sehen, wie das Lemma von KNASTER, KURATOWSKI und MAZURKIEWICZ direkt auf Gleichgewichte angewandt werden kann, betrachten wir einen einfachen Fall mit sehr stark abgeschwächten Gleichgewichtsforderungen.

Wir gehen wieder aus von einem Wirtschaftssystem mit einer gewissen Anzahl von Gütern $h = 1, \ldots, \ell$. Der Markt sollte Preise für diese Güter finden derart, daß es keinen Nachfrageüberhang gibt. Preise betrachten

wir wieder als relativ, d.h. als Elemente eines Simplex

$$P = \left\{ p \in \mathbb{R}^\ell \;\middle|\; p_h \geq 0 \;\forall h \text{ und } \sum_{h=1}^{\ell} p_h = 1 \right\} .$$

Betrachten wir zunächst den einfachsten Fall einer reinen Tauschwirtschaft, für die ein Preisvektor $p \in P$ vorgegeben sei.

Der i-te Verbraucher startet mit einem Vektor $\zeta_i \in \mathbb{R}^\ell$, der angibt, wieviel er von jedem Gut besitzt, und er möchte erreichen, daß er mit einem Vektor $x_i \in \mathbb{R}^\ell$, seinem Nachfragevektor, endet. Falls er keine Schulden machen will, kann dabei der Wert des neuen Vektors x_i nicht größer sein als der von ζ_i, d.h. $\langle p, x_i \rangle \leq \langle p, \zeta_i \rangle$. Summiert man über alle Verbraucher und vergleicht die Ausgangslage $\zeta = \sum_i \zeta_i \in \mathbb{R}^\ell$ mit der Endsituation $x = \sum_i x_i \in \mathbb{R}^\ell$, ist also $\langle p, x \rangle \leq \langle p, \zeta \rangle$. Die Nachfrage kann nur befriedigt werden, wenn der Vektor ζ komponentenweise größer oder gleich dem Vektor x ist, wenn also $x - \zeta$ keine echt positiven Komponenten hat.

Bei den Wirtschaftssystem, die wir in Kapitel neun betrachtet haben, gibt es zusätzlich noch Produzenten, die den Verbrauchern gehören, wobei der Verbraucher i einen Antal α_{ij} am Produzenten j hält, dessen Produktion durch einen Vektor $y_j \in \mathbb{R}^\ell$ beschrieben wird. Jetzt lautet die Bedingung, daß für jeden Verbraucher i

$$\langle p, x_i \rangle \leq \langle p, \zeta_i \rangle + \sum_{j=1}^{m} \alpha_{ij} \langle p, y_j \rangle$$

sein soll. Summiert man über alle i und setzt $y = \sum_{j=1}^{m} y_j$, folgt

$$\langle p, x \rangle \leq \langle p, \zeta \rangle + \langle p, y \rangle \ ,$$

denn $\sum_{i=1}^{n} \alpha_{ij} = 1$ für alle j, wenn die Produzenten ganz den Verbrauchern gehören. Die Nachfrage kann befriedigt werden, wenn komponentenweise $x \leq \zeta + y$ ist.

In beiden Fällen haben wir also zu einem gegebenen Preisvektor p einen Vektor $f(p)$, nämlich $x - \zeta$ *bzw.* $x - \zeta - y$, und wir suchen als schwächstmögliche Bedingung an ein Gleichgewicht Vektoren p, x, ζ und eventuell y derart, daß $f(p)$ keine positiven Komponenten hat und

$\langle p, f(p)\rangle \leq 0$ ist. (Letzere Bedingung bezeichnet man als die schwache WALRAS-Bedingung; bei der starken wird gefordert, daß $\langle p, f(p)\rangle$ verschwindet.)

Ein sehr schwaches Resultat über die Existenz von Gleichgewichten, bei denen weder die Präferenzen der Verbraucher noch Gewinnoptimierung der Produzenten eingeht und bei dem außerdem noch die zumindest *a priori* nicht ersichtliche Stetigkeit von f vorausgesetzt wird, liefert der folgende

Satz: $f\colon P \to \mathbb{R}^\ell$ sei stetig und $\langle p, f(p)\rangle \leq 0$ für alle $p \in P$. Dann ist $\{p \in P \mid f(p) \leq 0\}$ kompakt und nicht leer.

Beweis: Wegen der $\leq$-Bedingung an eine stetige Funktion ist die Menge abgeschlossen, als Teilmenge von P also kompakt.

Für $h = 1, \ldots, \ell$ bezeichne $f_h(p)$ die h-te Komponente von $f(p)$ und wir setzen

$$A_h = \{p \in P \mid f_h(p) \leq 0\}\,.$$

Der Vektor $p \in P$ habe genau für die Güter $h \in I \subseteq \{1, \ldots, \ell\}$ einen positiven Wert. Um das Lemma von KNASTER, KURATOWSKI und MAZURKIEWICZ anzuwenden, müssen wir zeigen, daß p in der Vereinigung der A_h mit $h \in I$ liegt. Wäre dies nicht der Fall, so wäre $f_h(p) > 0$ für alle $h \in I$, also auch

$$\langle p, f(p)\rangle = \sum_{h=1}^{\ell} p_h f_h(p) = \sum_{h \in I} p_h f_h(p) > 0\,,$$

im Widerspruch zur Voraussetzung. Damit liegt das offene Simplex, dessen Ecken die Preisvektoren sind, die für einen Index $h \in I$ die Komponente eins haben und sind nur Nullen, in der Vereinigung der A_h über alle $h \in I$, und wegen deren Abgeschlossenheit liegt dort auch der Abschluß dieses Simplex. Das Lemma sagt uns daher, daß der Durchschnitt aller A_h nicht leer ist, es gibt also Preisvektoren p, für die $f(p)$ keine positiven Komponenten hat.

■