Kapitel V
Allerleirauh

Als der Tanz zuende war, ließ sich der König die Suppe bringen und aß sie, und sie schmeckte ihm so gut, daß er meinte, niemals eine bessere Suppe gegessen zu haben.

Brüder Grimm

Hier tragen wir einiges nach, was doch auch jeder gebildete Mathematiker weiß und oft, wenn er die Anfangsgründe der Analysis erklärt, als Leitstern vor Augen hat. Da ließe sich noch mancherlei anfügen.

§ 1. Eine nicht meßbare Menge

Wir betrachten $\mathbb{R}$ als Maßraum. Zu den meßbaren Mengen soll jedenfalls das Einheitsintervall $I = [0, 1)$ gehören, und sein Maß sei $\mu(I) = 1$. Auch soll das betrachtete Maß μ translationsinvariant und σ-additiv sein. Wir zeigen, daß es eine Menge $M \subset I$ gibt, die für kein solches Maß meßbar ist. Freilich kann man diese Menge nicht wirklich vorzeigen, konstruieren: Man wählt mit dem Auswahlaxiom, Sie werden schon sehen.

Die additive Gruppe $\mathbb{R}/\mathbb{Z}$ hat die Untergruppe $\mathbb{Q}/\mathbb{Z}$, und sie zerfällt in Restklassen nach dieser Untergruppe, also in Restklassen für die Äquivalenzrelation auf $\mathbb{R}/\mathbb{Z}$

$$x \sim y \;:\Longleftrightarrow\; x - y \in \mathbb{Q}.$$

Aus jeder Klasse wählen wir genau ein Element in I aus. Die Klasse trifft ja I, weil $x \sim x - q$ für alle $q \in \mathbb{Q}$. Die so ausgewählte Menge M ist eine Teilmenge von $I \subset \mathbb{R}$.

Nun hat jedes $r \in \mathbb{R}$ einen wohlbestimmten Repräsentanten $r' \in I$ modulo $\mathbb{Z}$, nämlich $r' = r - [r]$. Insbesondere wird $\mathbb{Q}/\mathbb{Z}$ durch $\mathbb{Q} \cap I$ repräsentiert, also modulo $\mathbb{Z}$ erhält man eine disjunkte Zerlegung

$$\mathbb{R}/\mathbb{Z} = \bigcup_q M + q, \quad q \in \mathbb{Q} \cap I.$$

Zwar ist $M + q$ noch nicht in I, aber das korrigieren wir, wir setzen

$$M + q = A_q \sqcup B_q, \quad A_q = (M + q) \cap I,$$

und wir haben die disjunkte Zerlegung

$$[0,1) = \bigcup_q A_q \sqcup (B_q - 1), \quad q \in \mathbb{Q} \cap I. \tag{1.1}$$

Nun beachte, daß $M + q$ aus M durch Translation mit q hervorgeht. Wäre M meßbar, so mit gleichem Maß auch $M + q$ also auch $A_q \sqcup B_q$ und schließlich wäre

$$\mu(M) = \mu(A_q) + \mu(B_q - 1). \tag{1.2}$$

Ist nun $\mu(M) = 0$, so wäre $\mu(I) = 0$, und ist $\mu(M) > 0$, so wäre $\mu(I) = \infty$ nach (1.1), und daher kann M überhaupt nicht meßbar sein.

Die Konstruktion ist leichter zu durchschauen, wenn man den eigentlichen Ursprung des Gedankens aufsucht: Das Intervall I ist als Parametrisierung der Kreislinie, der multiplikativen Gruppe S^1, durch $t \mapsto \exp(2\pi it)$ anzusehen. Diese Gruppe hat die abzählbare Untergruppe

$$Q \;=\; \{\exp(2\pi it) \mid t \in \mathbb{Q}\}$$

der rationalen Drehungen. Die böse Menge ist ein Repräsentantensystem M der Zerlegung von S^1 in Restklassen modulo Q. Für ein unter Drehungen invariantes Maß μ auf S^1 mit $0 < \mu(S^1) < \infty$ kann M nicht meßbar sein. Es hätten ja alle Mengen qM für $q \in Q$ gleiches Maß weil sie durch Drehungen aus M hervorgehen. Die Mengen qM sind für verschiedene $q \in Q$ disjunkt, und ihre Vereinigung ist ganz S^1. Wäre nun $\mu(M) = 0$, so folgt $\mu(S^1) = 0$, weil Q abzählbar ist. Wäre $\mu(M) > 0$, so folgt $\mu(S^1) = \infty$, weil Q unendlich ist.

Von diesem Beispiel kommt man zum zuerst erklärten, wenn man überall den Kreis S^1 und seine Punkte durch das Parameterintervall I und die entsprechenden Punkte ersetzt.

So ist der Erweiterung von Maßen eine prinzipielle Grenze gesetzt, wenn man es mit dem Auswählen so hält, wie es in der Mathematik gebräuchlich ist. Freilich, sagen die Logiker, führt es auch nicht zu Widersprüchen, wenn man das Auswahlaxiom aufgibt und dafür postuliert, daß jede Teilmenge von $\mathbb{R}^n$ meßbar ist ...

§ 2. Der Rangsatz

Die Elementargeometrie der differenzierbaren Abbildungen beruht zunächst vor allem auf dem Satz über die Umkehrabbildung. Er sagt, welche Transformationen lokale Koordinatentransformationen sind: diejenigen nämlich, deren Differential als lineare Abbildung umkehrbar ist.

Die wesentliche Invariante einer linearen Abbildung ist ihr Rang. Ist $A : V \to W$ linear vom Rang r, so gibt es lineare Isomorphismen (Basisisomorphismen) $B : V \xrightarrow{\cong} \mathbb{R}^m$ und $C : W \xrightarrow{\cong} \mathbb{R}^n$, sodaß die transformierte Abbildung

$$\begin{array}{ccc} V & \xrightarrow{A} & W \\ {\scriptstyle B}\downarrow{\scriptstyle\cong} & & {\scriptstyle\cong}\downarrow{\scriptstyle C} \\ \mathbb{R}^m & \longrightarrow & \mathbb{R}^n \end{array}$$

die Gestalt

$$CAB^{-1} : \mathbb{R}^m \to \mathbb{R}^n, \quad (x_1, \dots, x_m) \mapsto (x_1, \dots, x_r, 0, \dots, 0)$$

hat. Man wählt eine Basis $v_1, \dots, v_m$ von V, so daß $v_{r+1}, \dots, v_m$ den Kern von A aufspannen, und eine Basis $w_1, \dots, w_n$ von W mit $w_j = Av_j$ für $j \leq r$.

Auch für eine lokale Beschreibung differenzierbarer Abbildungen bis auf Koordinatentransformationen ist der Rang die erste Invariante, aber hier ist auch der Rang eine Funktion, und ist sie nicht konstant, so kann man die gegebene Funktion auch nicht in ein so einfaches Musterexemplar, das ja konstanten Rang hat, transformieren.

(2.1) Rangsatz. *Sei U offen in $\mathbb{R}^m$ und V offen in $\mathbb{R}^n$, und sei $f : U \to V$ eine C^ℓ-Abbildung, $\ell \geq 1$, von lokal um $p \in U$ konstantem Rang r. Dann gibt es C^ℓ-Karten $h : U_1 \to \mathbb{R}^m$ und $k : V_1 \to \mathbb{R}^n$ von Umgebungen U_1 von p in U und V_1 von $q = f(p)$ in V, mit $h(p) = 0$, $k(q) = 0$, und*

$$k \circ f \circ h^{-1} : (x_1, \dots, x_m) \mapsto (x_1, \dots, x_r, 0, \dots, 0)$$

lokal um 0 in $\mathbb{R}^m$.

Lokal um die betrachteten Punkte p und q sieht es so aus:

$$\begin{array}{ccc} U & \xrightarrow{f} & V \\ {\scriptstyle h}\downarrow & & \downarrow{\scriptstyle k} \\ \mathbb{R}^m & \to & \mathbb{R}^n, \quad (x_1, \dots, x_m) \mapsto (x_1, \dots, x_r, 0, \dots, 0). \end{array}$$

Der Satz über die Umkehrabbildung ist der Spezialfall $m = n = r$.

Beweis: Wir dürfen gleich annehmen: $p = 0 \in \mathbb{R}^m$, $q = 0 \in \mathbb{R}^n$. Wir finden eine reguläre $(r \times r)$-Untermatrix von $Df(0)$, und nach Vertauschen der Koordinaten von $\mathbb{R}^m$ und $\mathbb{R}^n$ ist

$$(\partial f_i/\partial x_j), \quad 1 \leq i, j \leq r$$

am Ursprung regulär. Die lokal um den Ursprung definierte Transformation

$$h : (x_1, \dots, x_m) \mapsto \big(f_1(x), \dots, f_r(x), x_{r+1}, \dots, x_m\big)$$

hat die Jacobimatrix

$$\begin{array}{cl} \left[\begin{array}{c|ccc} \multicolumn{4}{c}{\partial f_i/\partial x_j} \\ \hline & 1 & & \\ 0 & & \ddots & \\ & & & 1 \end{array}\right] & \begin{array}{l} \left.\vphantom{\frac{\partial}{\partial}}\right\} r \\ \left.\vphantom{\begin{array}{c}1\\ \ddots\\ 1\end{array}}\right\} m - r \end{array} \end{array}$$

weiße Stellen sind Null.

Ihre Determinante ist $\det(\partial f_i/\partial x_j)_{i,j \leq r} \neq 0$, also h ist eine zulässige Koordinatentransformation am Ursprung, und das Diagramm

$$\begin{tikzcd} (x_1, \dots, x_m) \arrow[rr, mapsto, "f"] \arrow[dr, mapsto, "h"'] & & \big(f_1(x), \dots, f_n(x)\big) \\ & \begin{array}{c} \big(f_1(x), \dots, f_r(x), x_{r+1}, \dots, x_m\big) \\ = (z_1, \dots, z_r, z_{r+1}, \dots, z_m) \end{array} \arrow[ur, mapsto, "f \circ h^{-1} =: g"'] & \end{tikzcd}$$

zeigt:

$$g := f \circ h^{-1} : (z_1, \dots, z_m) \mapsto \big(z_1, \dots, z_r, g_{r+1}(z), \dots, g_n(z)\big).$$

Soweit führt die Transformation im Urbildraum. Bisher haben wir erst $\mathrm{rg}_p f \geq r$ benutzt. Die Jacobimatrix von g hat die Gestalt

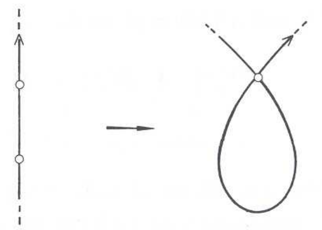
$$\left[\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ \hline & ? & & A(z) \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\ \ddots\\ 1\end{array}}\right\} r \\ \left.\vphantom{A}\right\} m-r \end{array} \qquad A(z) = (\partial g_i/\partial z_j)_{i,j>r}\ .$$

Die Abbildung $g = f \circ h^{-1}$ hat den gleichen Rang r wie f, und daher muß die Teilmatrix $A(z)$ lokal um den Ursprung verschwinden, also $\partial g_i/\partial z_j = 0$ für $i, j > r$. Betrachten wir die Funktionen auf einem Würfel $\{|z_j| < \varepsilon\}$ um den Ursprung, so hängen die Funktionen g_i, $i > r$ damit von den letzten Komponenten z_j, $j > r$ nicht ab, wir können schreiben $g_i = g_i(z_1, \ldots, z_r)$. Dann aber haben wir lokal um den Ursprung von $\mathbb{R}^n$ die invertierbare Transformation

$$\begin{aligned} &k\colon (z_1, \ldots, z_n) \mapsto \\ &\quad \big(z_1, \ldots, z_r, z_{r+1} - g_{r+1}(z_1, \ldots, z_r), \ldots, z_n - g_n(z_1, \ldots, z_r)\big), \end{aligned}$$

deren inverse ebenso mit $+$ statt $-$ aussieht, und $k \circ g = k \circ f \circ h^{-1}$ hat die verlangte Gestalt. □

Der Rang $\mathrm{rg}_x f$ kann in einer Umgebung von p nicht kleiner als $\mathrm{rg}_p f$ sein. Ist $\mathrm{rg}_p f = r$, so hat die Jacobimatrix $Df(p)$ eine $(r \times r)$-Untermatrix mit nicht verschwindender Determinante, und die bleibt in einer Umgebung von p ungleich Null. Wohl aber kann $\mathrm{rg}_x f > \mathrm{rg}_p f$ für x nahe p sein, wie die Abbildung $f(x) = x^2$ für $p = 0$ zeigt. Der Rangsatz beschreibt also lokal eine Abbildung bis auf Koordinatentransformation, falls der Rang der Funktion lokal nicht steigt. Zwei Fälle gibt es, wo man dessen sicher sein kann, nämlich wenn $\mathrm{rg}_p f = m$ und wenn $\mathrm{rg}_p f = n$ ist. Eine Abbildung wie im Satz vom Rang m in jedem Punkt heißt eine **Immersion** (**immersiv**), eine Abbildung vom Rang n heißt **Submersion** (submersiv). Eine Immersion ist lokal injektiv, wie der Rangsatz zeigt. Im Großen muß sie nicht injektiv sein:



Eine Submersion ist nach dem Rangsatz offen: Bilder von offenen Mengen sind offen.

§ 3. Das Morse-Lemma

Manches grundlegende Theorem heißt Lemma, wie manche würdige Person noch mit ihrem Kindernamen gerufen wird; so auch dieses. Wir haben gelernt, daß eine C^2-Funktion lokal um einen singulären Punkt mit nicht ausgearteter Hesseform ebenso aussieht, wie diese Hesseform. Das Lemma von Morse sagt, daß in der Tat die Funktion sich lokal durch eine differenzierbare Koordinatentransformation in ihre Hesseform überführen läßt.

(3.1) Theorem *(M. Morse). Sei U offen in $\mathbb{R}^n$ und $f : U \to \mathbb{R}$ sei eine C^∞-Funktion. Sei $p \in U$ ein kritischer Punkt von f mit nicht ausgearteter Hessematrix $2H$. Dann gibt es eine Umgebung V von p in U und einen C^∞-Diffeomorphismus $\varphi : V \to V'$ mit $\varphi(p) = 0 \in \mathbb{R}^n$ und $D\varphi(p) = \mathrm{id}$, sodaß*

$$f \circ \varphi^{-1}(x) \;=\; f(p) + {}^t xHx.$$

Zunächst wollen wir die Transformationen der symmetrischen Matrizen selbst, also der quadratischen Formen, betrachten.

(3.2) Lemma. *Sei H eine reguläre symmetrische reelle $(n \times n)$-Matrix, dann gibt es eine Umgebung U von H im Raum S aller symmetrischen reellen $(n \times n)$-Matrizen und eine C^∞-Abbildung*

$P : U \to \mathrm{End}(\mathbb{R}^n)$, *mit* $P(H) = \mathrm{id}$ *und*

$$^t\!P(A) \cdot A \cdot P(A) \;=\; H$$

für alle $A \in U$.

Mit anderen Worten: Wenn A sich wenig von H unterscheidet, läßt sich A in H transformieren, und die Transformation hängt C^∞ von A ab.

Beweis: Jedenfalls gibt es ja eine Transformation T, sodaß ^tTHT eine Diagonalmatrix ist. Es genügt also, das Lemma für Diagonalmatrizen H zu zeigen. In diesem Fall schreiben wir $A = H + X$, wobei X eine Umgebung von 0 im Raum S der symmetrischen Matrizen durchlaufen soll, und wir suchen $P(A) = 1 + Y$ mit einer oberen Dreiecksmatrix Y. Hier ist $1 = \mathrm{id}$ die Einheitsmatrix. Erreichen wollen wir:

$$F(X,Y) := \;{}^t(1+Y)\cdot(H+X)\cdot(1+Y) - H = 0.$$

Dies ruft nach dem Satz über das Auflösen von Gleichungen: Sei also S der Vektorraum der symmetrischen $(n \times n)$-Matrizen, V der Vektorraum der oberen Dreiecksmatrizen (gleicher Dimension $\frac{1}{2}n(n{+}1)$), und

$$F : S \times V \to S, \quad (X,Y) \mapsto F(X,Y),$$

wie oben definiert. Wir wollen die Gleichung $F(X,Y) = 0$ lokal um $(0,0)$ durch $Y = Y(X)$ lösen. Für den Satz über das Auflösen von Gleichungen müssen wir also $D_Y F(0,0) \in \mathrm{Hom}_{\mathbb{R}}(V,S)$ berechnen. Nun, für $X = 0$ ist

$$F(0,Y) = {}^tYH + HY + {}^tYHY,$$

also

$$D_Y F(0,0) : Y \mapsto {}^tYH + HY.$$

Wir müssen zeigen, daß diese lineare Abbildung den Kern 0 hat, also ${}^tYH + HY = 0 \Longrightarrow Y = 0$. Weil aber H regulär diagonal, Y eine obere und damit ${}^t Y$ eine untere Dreiecksmatrix ist, stimmt das.

□

Weil hier nur quadratische Gleichungen gelöst werden, kann man die Lösung auch durch eine Folge quadratischer Ergänzungen hinschreiben.

Beweis (3.1). Wir können $p = 0$, $f(0) = 0$ annehmen und

$$f(x) = \sum_{i,j=1}^{n} a_{ij}(x)\, x_i x_j, \quad a_{ij}(0) = h_{ij},$$

mit C^∞-Funktionen a_{ij} und $(h_{ij}) = H$ schreiben. Das zeigt zum Beispiel die Integraldarstellung des Restglieds zweiter Ordnung der Taylorentwicklung. Setzen wir $A(x) = \big(a_{ij}(x)\big)$ und wählen P wie im Lemma, und $Q(x) := P\big(A(x)\big)$, so steht da

$$f(x) = {}^t x \cdot A(x) \cdot x, \quad {}^t Q(x) \cdot A(x) \cdot Q(x) = H,$$

also lokal um Null

$$f(x) = {}^t(Q(x)^{-1} \cdot x) \cdot H \cdot (Q(x)^{-1} \cdot x).$$

Bleibt also, $\varphi(x) = Q(x)^{-1} \cdot x$ zu setzen, was wegen

$$Q(0) = PA(0) = P(H) = \mathrm{id}$$

in der Tat bei Null die Ableitung id hat. Siehe unsere Definition der Ableitung. □

Auch hier braucht man nicht, daß f eine C^∞-Funktion ist: Die Differenzierbarkeitsordnung von φ hängt an der von $x \mapsto A(x)$, und diese kann höchstens zwei geringer als die von f sein.

Betrachten wir noch einmal Funktionen einer Variablen, etwa lokal um den Ursprung. Verschwindet der $(k-1)$-Jet von f am Ursprung, so ist, wie gesagt:

$$f(x) = \varphi(x) \cdot x^k, \quad \varphi(x) = \frac{1}{(k-1)!} \int_0^1 (1-t)^{k-1} f^{[k]}(tx)\, dt,$$

wie man sieht, wenn man die Integraldarstellung des Restglieds der Taylorentwicklung (Bd. 1, IV, 2.5) auf die Funktion $x \mapsto f(x \cdot h)$ für $x = 1$ anwendet, und dann wieder x statt h schreibt. Uns interessiert nur, daß jedenfalls φ eine C^∞-Funktion ist, falls dasselbe für f gilt, und $a := \varphi(0)$ ist der k-te Taylorkoeffizient von f. Ist $a \neq 0$, so können wir schreiben:

$$f(x) = a \cdot \big(\psi(x)\big)^k, \quad \psi(x) := x \cdot \sqrt[k]{\varphi(x)/a}.$$

Weil $\varphi(0)/a = 1$, ist ψ lokal um Null eine C^∞-Funktion, und es ist $\psi'(0) = 1$, also haben wir gezeigt:

(3.3) Bemerkung. *Ist die reelle C^∞-Funktion f lokal um $p \in \mathbb{R}$ definiert und beginnt ihre Taylorentwicklung mit dem Term $a \cdot x^k$, $a \neq 0$, $k > 0$, so gibt es lokal um p eine invertierbare Transformation φ mit $\varphi(p) = 0$, $\varphi'(p) = 1$, sodaß*

$$f(x) = a \cdot \varphi(x)^k. \qquad \square$$

Auch in höherer Dimension sind fast alle Funktionen lokal durch ein endliches Taylorpolynom an der betreffenden Stelle bis auf Transformation bestimmt, aber das ist nicht so leicht zu zeigen, ja nicht einmal leicht zu sagen, was das heißen soll. Immerhin: Das Morselemma ist ein erster und der wichtigste Schritt.

§ 4. Der Satz von Sard

Ist U offen in $\mathbb{R}^n$ und $f : U \to \mathbb{R}^p$ stetig differenzierbar, so hat die Gleichung

$$f(x) = q$$

eine p-kodimensionale Untermannigfaltigkeit von U als Lösungsmenge, falls q ein regulärer Wert von f ist. Wie groß aber ist die Aussicht, daß man bei zufälliger Wahl von q einen regulären Wert von f trifft? Sehr groß, das sagt eben der

(4.1) Satz von Sard. *Sei U offen in $\mathbb{R}^n$, sei $f : U \to \mathbb{R}^p$ eine C^∞-Abbildung, und sei $D \subset U$ die Menge der kritischen Punkte der Abbildung f, dann hat $f(D) \subset \mathbb{R}^p$ das Maß Null.*

Beweis: Induktion nach n; für $n = 0$ ist $\mathbb{R}^n$ ein Punkt, $f(U)$ höchstens ein Punkt, der Satz also richtig. Da er jedenfalls auch für $p = 0$ gilt, nehmen wir jetzt $p > 0$ an.

Für den Induktionsschritt sei $D_i \subset U$ die Menge der Punkte $u \in U$, wo alle partiellen Ableitungen von f der Ordnung $\leq i$ verschwinden. Die D_i bilden eine absteigende Folge abgeschlossener Teilmengen

$$D \supset D_1 \supset D_2 \supset \cdots$$

von U, und wir zeigen:

(i) $f(D \smallsetminus D_1)$ ist dünn, d.h. ist eine Nullmenge.
(ii) $f(D_i \smallsetminus D_{i+1})$ ist dünn.
(iii) $f(D_k)$ ist dünn für genügend große k.

Alle diese Aussagen müssen wir nur lokal zeigen, also jeder Punkt $u \in D \smallsetminus D_1$ hat eine Umgebung V, so daß $f(V \cap (D \smallsetminus D_1))$ dünn ist, und so auch in den anderen Fällen. Abzählbar viele solche Umgebungen überdecken ja dann die betroffene Menge.

Beweis (i): Man kann $p \geq 2$ annehmen, denn für $p = 1$ ist $D = D_1$. Sei $u \in D \smallsetminus D_1$, dann verschwindet eine partielle Ableitung von f nicht am Punkt u, und wir dürfen annehmen $\partial f_1 / \partial x_1(u) \neq 0$. Dann ist die Abbildung

$$h : U \to \mathbb{R}^n, \quad x \mapsto \big(f_1(x), x_2, \ldots, x_n\big)$$

bei u lokal invertierbar, ihre Einschränkung auf eine Umgebung V von u ist ein Diffeomorphismus $h : V \to V'$, und die transformierte Abbildung $g := f \circ h^{-1}$ hat lokal um $h(u)$ die Gestalt

$$g : (z_1, \ldots, z_n) \mapsto \big(z_1, g_2(z), \ldots, g_p(z)\big).$$

Wir müssen die Behauptung für g zeigen. Diese Abbildung überführt die Hyperebene $\{z \mid z_1 = t\}$ jeweils auf ihrem Definitionsgebiet in die Hyperebene $\{y \mid y_1 = t\}$. Sei

$$g^t\colon \bigl(\{t\} \times \mathbb{R}^{n-1}\bigr) \cap V' \to \{t\} \times \mathbb{R}^{p-1}$$

die Einschränkung von g. Dann ist ein Punkt aus $\bigl(\{t\} \times \mathbb{R}^{n-1}\bigr) \cap V'$ genau dann kritisch für g, wenn er kritisch für g^t ist, weil g die Jacobimatrix

$$Dg = \left[\begin{array}{c|ccc} 1 & 0 & \cdots & 0 \\ \hline ? & & Dg^t & \\ & & & \end{array}\right]$$

hat. Nun hat aber nach Induktionsvoraussetzung die Menge der kritischen Werte von g^t das Maß Null in $\{t\} \times \mathbb{R}^{p-1}$, also hat die Menge der kritischen Werte von g dünnen Durchschnitt mit jeder Hyperebene $\{y \mid y_1 = t\}$, hat also selbst nach Fubini das Maß Null, und das zeigt (i).

Beweis (ii): Hier verfahren wir ähnlich wie im Beweis von (i). Für jeden Punkt $u \in D_k \smallsetminus D_{k+1}$ gibt es eine $(k+1)$-te Ableitung, die im Punkt u nicht verschwindet, wir dürfen annehmen:

$$\partial^{k+1} f_1/\partial x_1 \partial x_{\nu_1} \ldots \partial x_{\nu_k}(u) \neq 0.$$

Sei $w : U \to \mathbb{R}$ die Funktion

$$w = \partial^k f_1/\partial x_{\nu_1} \ldots \partial x_{\nu_k}.$$

Dann ist also $w(u) = 0$ und $\partial w/\partial x_1(u) \neq 0$, und wie eben definiert die Abbildung

$$h : x \mapsto \bigl(w(x), x_2, \ldots, x_n\bigr)$$

eine Karte $h : V \xrightarrow{\cong} V'$ um u, und

$$h(D_k \cap V) \;\subset\; \{0\} \times \mathbb{R}^{n-1} \;\subset\; \mathbb{R}^n.$$

Betrachten wir also wieder die transformierte Abbildung $g := f \circ h^{-1} : V' \to \mathbb{R}^p$ und ihre Einschränkung $g^0 \colon \big(\{0\} \times \mathbb{R}^{n-1}\big) \cap V' \to \mathbb{R}^p$, so hat die Menge der kritischen Werte von g^0 nach Induktionsvoraussetzung das Maß Null. Aber jeder Punkt aus $h(D_k \cap V)$ ist kritisch für g^0, weil alle partiellen Ableitungen von g, also auch von g^0, der Ordnung $\leq k$, insbesondere erster Ordnung, verschwinden. Also ist $f(D_k \cap V) = g \circ h(D_k \cap V)$ dünn.

Beweis (iii): Sei $W \subset U$ ein Würfel der Kantenlänge a, und sei

$$k > \frac{n}{p} - 1.$$

Dann zeigen wir, daß $f(W \cap D_k)$ dünn ist. Die Taylorformel liefert die Abschätzung

$$f(u+h) = f(u) + R(u,h), \quad |R(u,h)| \leq c \cdot |h|^{k+1}$$

für $u \in D_k \cap W$ und $u+h \in W$, wobei die Konstante c bei gegebenem f und W jetzt fest gewählt sei. Hier benutzen wir, daß f eine C^{k+1}-Funktion ist, vergleiche (I, 3.6).

Nun zerlege W in r^n Würfel der Kantenlänge a/r. Ist W_1 ein Würfel dieser Zerlegung, der einen Punkt $u \in D_k$ enthält, so schreibt sich jeder Punkt aus W_1 als $u+h$ mit $|h| \leq \sqrt{n}\, a/r$, und nach der obigen Restgliedabschätzung liegt $f(W_1)$ in einem Würfel der Kantenlänge

$$2 \cdot c \cdot \frac{(\sqrt{n}\, a)^{k+1}}{r^{k+1}} = \frac{b}{r^{k+1}},$$

mit einer Konstante b, die nur von W und f, nicht aber von der Zerlegung abhängt. Alle diese Würfel zusammen haben eine Volumensumme $s \leq r^n \cdot b^p / r^{p(k+1)}$, und für $p(k+1) > n$ konvergiert dieser Ausdruck mit wachsendem r gegen Null. Die Volumensumme kann also durch Wahl einer genügend feinen Zerlegung beliebig klein gemacht werden. $\square$

Schauen wir den Beweis noch einmal an, so finden wir, daß wir zuletzt die Taylorentwicklung einer C^{k+1}-Funktion f benutzt haben, also voraussetzen müssen, daß f eine C^ℓ-Funktion mit $\ell > n/p$ ist. Aber in der Induktion kommen dann ja auch die Dimensionspaare $(n-1, p-1), \ldots$ und schlimmstens $(n-p+1, 1)$ vor, sodaß wir also $\ell > \max\{0, n-p+1\}$ benutzt haben. Tatsächlich genügt

$$\ell > \max\{0, n-p\},$$

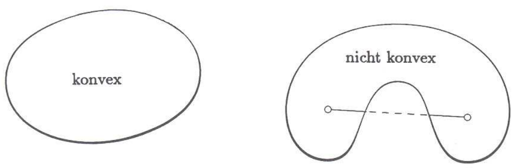
aber das braucht man auch. H. Whitney hat eine C^1-Funktion auf der Ebene konstruiert, die auf einer topologisch eingebetteten Strecke das Differential Null hat aber dort nicht konstant ist, sodaß also die Menge der kritischen Werte ein Intervall in $\mathbb{R}$ enthält (Duke Math. J. 1 (1935), 514-517).

§ 5. Konvexe Funktionen

Eine Teilmenge $K \subset \mathbb{R}^n$ heißt **konvex**, wenn sie mit je zwei Punkten p, q auch deren Verbindungsstrecke

$$\{\lambda p + \mu q \mid \lambda, \mu \geq 0, \quad \lambda + \mu = 1\}$$

enthält. Man kann auch $(1-t)p + tq = p + t(q-p)$, $0 \leq t \leq 1$, als Parametrisierung der Verbindungsstrecke wählen.

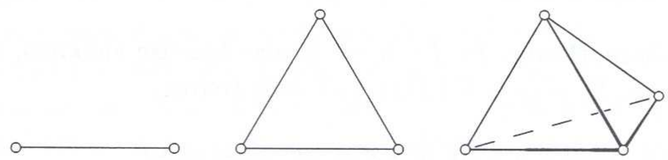


Beliebige Durchschnitte konvexer Mengen sind offenbar konvex, und daher liegt jede Teilmenge $A \subset \mathbb{R}^n$ in einer kleinsten konvexen Teilmenge, nämlich dem Durchschnitt aller konvexen Teilmengen, in denen sie liegt. Diese Menge bezeichnet man als die **konvexe Hülle** der Menge A.

(5.1) Beispiel. *Die konvexe Hülle von k Punkten $p_1, \ldots, p_k$ ist die Menge der Punkte*

$$\lambda_1 p_1 + \cdots + \lambda_k p_k,$$

mit $\lambda_1 + \cdots + \lambda_k = 1$ und $\lambda_j \geq 0$ für $j = 1, \ldots, k$.



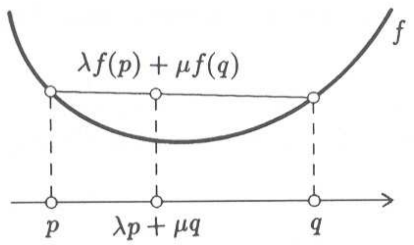
Beweis: Die beschriebene Menge ist offenbar konvex und enthält alle p_j. Umgekehrt schließt man durch Induktion: Ist $\lambda_k \neq 1$ und $\mu := 1 - \lambda_k = \lambda_1 + \cdots + \lambda_{k-1}$, so setze $\lambda'_j := \lambda_j/\mu$, dann liegt

$$\lambda_1 p_1 + \cdots + \lambda_k p_k = \mu(\lambda'_1 p_1 + \cdots + \lambda'_{k-1} p_{k-1}) + \lambda_k p_k$$

auf der Verbindungsstrecke von p_k und einem Punkt der konvexen Hülle von $\{p_1, \ldots, p_{k-1}\}$, also in der konvexen Hülle von $p_1, \ldots, p_k$. $\square$

Sei nun $K \subset \mathbb{R}^n$ eine konvexe Menge. Eine Funktion $f : K \to \mathbb{R}$ heißt **konvex**, wenn die Menge der Punkte über f, also die Menge $\{(x, y) \mid y \geq f(x)\}$ konvex ist, und das bedeutet offenbar, wenn für alle $p, q \in K$ gilt: Für $\lambda, \mu \geq 0$ und $\lambda + \mu = 1$ ist

$$f(\lambda p + \mu q) \ \leq \ \lambda f(p) + \mu f(q). \tag{5.2}$$



Die Definition einer konvexen Funktion ist nur sinnvoll, wenn das Definitionsgebiet der Funktion konvex ist, und das wollen wir jetzt immer voraussetzen.

(5.3) Bemerkung. *Ist* $f : K \to \mathbb{R}$ *eine konvexe Funktion, so ist die Menge* $K_c = \{x \in K \mid f(x) \leq c\}$ *auch konvex.*

Beweis: Sind $p, q \in K_c$ und λ, μ wie oben, so ist

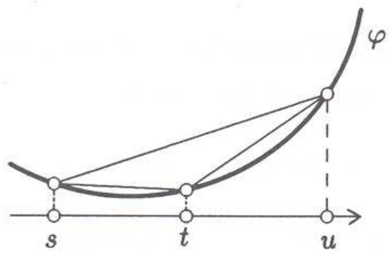
$$f(\lambda p + \mu q) \leq \lambda f(p) + \mu f(q) \leq \lambda c + \mu c = c,$$

also $\lambda p + \mu q \in K_c$. □

Die Definition zeigt unmittelbar, daß eine Funktion genau dann konvex ist, wenn ihre Einschränkung auf jede Strecke konvex ist. Man kann sich daher in vielen Situationen auf das Eindimensionale zurückziehen, und konvexe Funktionen einer Variablen sind die wichtigsten. Für eine Funktion $\varphi : (a, b) \to \mathbb{R}$ ist die Konvexitätsbedingung (5.2) äquivalent zu der Bedingung

$$\text{(5.4)} \qquad \frac{\varphi(t) - \varphi(s)}{t - s} \leq \frac{\varphi(u) - \varphi(t)}{u - t} \quad \text{für} \quad a < s < t < u < b.$$

Die Bedingung der Konvexität sagt ja, daß zwischen beiden Termen die Steigung $\big(\varphi(u) - \varphi(s)\big)/(u - s)$ der Strecke liegt, die $\big(s, \varphi(s)\big)$ mit $\big(u, \varphi(u)\big)$ verbindet.



(5.5) Satz. *Eine konvexe Funktion auf einer offenen Teilmenge von $\mathbb{R}^n$ ist stetig.*

Beweis: Sei $f : K \to \mathbb{R}$ die Funktion und $p \in K$. Wir können nach Abziehen einer Konstante und einer Translation in K annehmen: $p = 0$ und $f(p) = 0$. Weil K offen ist, liegt noch ein Würfel W der Kantenlänge $2s > 0$ um 0 in K, und auf den (endlich vielen) Ecken dieses Würfels sei $f \leq c$ für ein $c > 0$. Dann liegen diese Ecken in K_c, also ihre konvexe Hülle W auch, also insbesondere

$$f(x) \;\leq\; c \quad \text{für} \quad |x| = s.$$

Für ein festes solches x betrachte die konvexe Funktion einer Variablen

$$\varphi(t) := f(tx), \quad -1 \leq t \leq 1.$$

Es ist $\varphi(0) = 0$ und $\varphi(1) \leq c$, $\varphi(-1) \leq c$. Daraus folgt

$$\varphi(t) \leq ct, \quad \varphi(-t) \geq -ct \quad \text{für} \quad 0 \leq t \leq 1,$$

aus (5.4) mit $0, t, 1$ bzw. $-t, 0, 1$ für s, t, u. Weil aber $\varphi(-t)$ dieselben Voraussetzungen wie $\varphi(t)$ erfüllt, gilt $-ct \leq \varphi(t) \leq ct$, also $|\varphi(t)| \leq ct$, und das heißt:

$$|f(tx)| \leq ct \quad \text{für } |x| = s \text{ und } 0 \leq t \leq 1\,.$$

Für beliebiges x mit $0 < |x| \leq s$ ist daher

$$|f(x)| = \big|f\big((|x|/s) \cdot sx/|x|\big)\big| \leq c|x|/s,$$

und $f(0) = 0$, was die Behauptung zeigt. □

Es ist wesentlich, daß K offen ist, siehe die Funktion $\varphi\colon [0,1] \to \mathbb{R}$ mit Wert 0 am Rand und -1 im Inneren.

Aus (5.4) folgt mit dem Mittelwertsatz sofort:

(5.6) Satz. *Eine differenzierbare Funktion $f : (a,b) \to \mathbb{R}$ ist genau dann konvex, wenn ihre Ableitung monoton wächst. Existiert f'' überall, so ist f genau dann konvex, wenn stets $f'' \geq 0$ ist.* □

Im Höherdimensionalen entnimmt man daraus:

(5.7) Satz. *Eine C^2-Funktion $f : K \to \mathbb{R}^n$ ist genau dann konvex, wenn die Matrix $(\partial^2 f/\partial x_i \partial x_j)$ überall positiv semidefinit ist.*

Beweis: Die Funktion $t \mapsto f(x + th)$ hat nach (I, 3.3) bei $t = 0$ die zweite Ableitung

$$\sum_{i,j=1}^{n} \partial^2 f/\partial x_i \partial x_j(x)\, h_i h_j,$$

und die Bedingung des Satzes ist, daß dieses nie negativ ist, was nach (5.6) dazu äquivalent ist, daß die Einschränkung von f auf jede Gerade konvex ist. □

Weil der Durchschnitt einer Familie konvexer Mengen konvex ist, ist das Supremum einer Familie konvexer Funktionen konvex. So haben wir alles in allem einen großen Vorrat konvexer Funktionen. Die wichtigste Aussage über sie ist folgende:

(5.8) Jensens Ungleichung. *Sei X ein Maßraum mit Maß μ, sodaß $\mu(X) = 1$. Sei $f : X \to (a,b)$ integrabel und $\varphi : (a,b) \to \mathbb{R}$ konvex. Dann gilt*

$$\varphi\Big(\int_X f\, d\mu\Big) \;\leq\; \int_X \varphi \circ f\, d\mu.$$

Es kann $a = -\infty$ oder $b = \infty$ sein, und die rechte Seite kann ∞ sein.

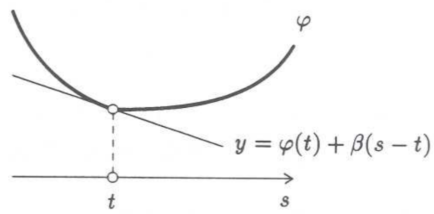
Beweis: Sei $t = \int_X f\, d\mu$, dann ist $a < t < b$. Sei β das Supremum über s der Differenzenquotienten $\big(\varphi(t)-\varphi(s)\big)/(t-s)$ auf der linken Seite von (5.4), dann steht dort:

$$\frac{\varphi(t)-\varphi(s)}{t-s} \le \beta \le \frac{\varphi(t)-\varphi(u)}{t-u}, \quad s < t < u, \quad \text{d.h.}$$

$$\varphi(t)-\varphi(s) \le (t-s)\beta, \quad \varphi(t)-\varphi(u) \le (t-u)\beta,$$

weil $(t-u) < 0$. Demnach gilt für alle $s \in (a,b)$:

$$\varphi(s)-\varphi(t)-\beta(s-t) \ge 0.$$



Dies sagt nur, daß φ oberhalb der Geraden durch $\varphi(t)$ mit Steigung β verläuft. Eine solche Stützhyperebene findet man auch im Höherdimensionalen durch jeden Punkt einer konvexen Funktion. Nun setze $s = f(x)$, dann haben wir

$$\varphi f(x) - \varphi(t) - \beta\,\big(f(x)-t\big) \ge 0,$$

und weil die so als nicht negativ erkannte Funktion μ-meßbar ist, können wir integrieren, und wegen $\mu(X) = 1$ ergibt sich

$$\int \varphi \circ f\, d\mu \;\ge\; \varphi(t) + \beta\Big(\int f\, d\mu - t\Big),$$

was nach Bestimmung von t die Behauptung ist. □

Der Satz gilt mit gleichem Beweis auch im Höherdimensionalen. Das Integral $\int \varphi \circ f\, d\mu$ ist so zu deuten, daß jedenfalls $(\varphi \circ f)_-$ immer endliches Integral hat, wenn f integrabel ist.

Man kann die Idee im Satz und Beweis so fassen: Die Menge

$$K := \{(s, y) \mid \varphi(s) \leq y\}$$

ist konvex, und die Abbildung $X \to (a, b) \times \mathbb{R}$, $x \mapsto \big(f(x), \varphi f(x)\big)$ führt in diese Menge, also liegt auch ihr Mittelwert $(\int f, \int \varphi f)$ in K, und das heißt $\varphi \int f \leq \int \varphi f$.

Jedoch weiß man die Ungleichung erst durch ihre Anwendungen und Spezialisierungen zu würdigen.

Die Funktion $\varphi(s) = e^s$ ist konvex weil $\varphi'' > 0$, also

$$\exp \int_X f \, d\mu \;\leq\; \int_X e^f \, d\mu. \tag{5.9}$$

Wählt man $X = \{p_1, \ldots, p_n\}$ mit $\mu(p_j) = \lambda_j$ und $\lambda_1 + \cdots + \lambda_n = 1$ und schreibt $f(p_j) = x_j$, so besagt die Jensensche Ungleichung:

$$\varphi(\lambda_1 x_1 + \cdots + \lambda_n x_n) \;\leq\; \lambda_1 \varphi(x_1) + \cdots + \lambda_n \varphi(x_n). \tag{5.10}$$

In unserem Fall also:

$$\exp(\lambda_1 x_1 + \cdots + \lambda_n x_n) \;\leq\; \lambda_1 \exp(x_1) + \cdots + \lambda_n \exp(x_n),$$

oder wenn wir $y_j = \exp(x_j)$ setzen:

$$\begin{aligned} &y_1^{\lambda_1} \cdot \ldots \cdot y_n^{\lambda_n} \leq \lambda_1 y_1 + \cdots + \lambda_n y_n, \\ &\text{für } 0 \leq \lambda_j \text{ und } \lambda_1 + \cdots + \lambda_n = 1. \end{aligned} \tag{5.11}$$

Das ist die Ungleichung zwischen dem geometrischen und arithmetischen Mittel. Man nennt darum allgemein in der Ungleichung

$$\exp \int_X \log(g) \, d\mu \;\leq\; \int_X g \, d\mu \tag{5.12}$$

die linke Seite das **geometrische** und die rechte das **arithmetische Mittel**.

Zwei positive reelle Zahlen p und q bilden ein **Paar konjugierter Exponenten**, wenn

$$\frac{1}{p} + \frac{1}{q} = 1, \quad \text{d.h. } p + q = pq\,.$$

Es ist dann $1 < p, q < \infty$, und weil p gegen ∞ geht für q gegen 1, nennt man auch $(1, \infty)$ und $(\infty, 1)$ ein konjugiertes Paar. Ein wichtiger Spezialfall ist $p = q = 2$, und die vertraute Dreiecksungleichung und Schwarzsche Ungleichung in diesem Fall verallgemeinern sich wie folgt:

(5.12) Satz. *Sei X ein Maßraum und seien p, q konjugierte Exponenten mit $1 < p, q < \infty$. Seien $f, g : X \to [0, \infty]$ meßbare Funktionen. Dann gilt:*

(i) Hölders Ungleichung.

$$\int\limits_X f \cdot g\, dx \;\le\; \Big(\int\limits_X f^p dx\Big)^{1/p} \cdot \Big(\int\limits_X g^q dx\Big)^{1/q}.$$

(ii) Minkowskis Ungleichung.

$$\Big(\int\limits_X (f+g)^p dx\Big)^{1/p} \;\le\; \Big(\int\limits_X f^p dx\Big)^{1/p} \;+\; \Big(\int\limits_X g^p dx\Big)^{1/p}.$$

Wir schreiben $\langle f, g\rangle := \int_X f \cdot g\, dx$ und $\|f\|_p = \big(\int_X f^p dx\big)^{1/p}$, dann heißt das für nicht negative Funktionen f und g:

(i) $$\langle f, g\rangle \;\le\; \|f\|_p \cdot \|g\|_q,$$

(ii) $$\|f+g\|_p \;\le\; \|f\|_p + \|g\|_p.$$

Beweis (i): Ist $\|f\|_p = 0$ so ist $f = 0$ fast überall, also gilt die Ungleichung, ebenso falls $\|g\|_q = 0$. Andernfalls kann man f durch $f/\|f\|_p$ und g durch $g/\|g\|_q$ ersetzen, und es genügt, die Behauptung für $\|f\|_p = \|g\|_q = 1$ zu zeigen. Nun ist

$$f(x) \cdot g(x) \le p^{-1} f(x)^p + q^{-1} g(x)^q$$

für alle $x \in X$ nach (5.11), und daraus ergibt sich (i) durch Integration, es ist ja jetzt $\int f^p = \int g^q = 1$.

Beweis (ii): Sei $h = (f+g)^{p-1}$, dann gilt:

$$(*)\quad \left\|f+g\right\|_p^p = \langle f+g, h\rangle = \langle f,h\rangle + \langle g,h\rangle \leq \|f\|_p \cdot \|h\|_q + \|g\|_p \cdot \|h\|_q$$

Nun ist $h^q = (f+g)^p$, also

$$\begin{aligned}\|h\|_q &= \Big(\int h^q\Big)^{1/q} = \Big(\int (f+g)^p\Big)^{1/q} = \Big(\int (f+g)^p\Big)^{(1/p)\cdot(p/q)} \\ &= \Big(\int (f+g)^p\Big)^{(1/p)(p-1)} = \left\|f+g\right\|_p^{p-1}.\end{aligned}$$

Setzt man dies in $(*)$ ein und dividiert durch $\|f+g\|_p^{p-1}$, so folgt die Behauptung jedenfalls, wenn $\|f+g\|_p$ nicht 0 oder ∞ ist.

Ist $\|f+g\|_p = 0$, so ist die Behauptung trivial, ebenso wie wenn $\|f\|_p + \|g\|_p = \infty$ ist. Aber weil die Funktion $t \mapsto t^p$ für $t \geq 0$ konvex ist, gilt

$$2^{-p}(f+g)^p \leq \tfrac{1}{2}(f^p + g^p),$$

also wenn $\|f\|_p + \|g\|_p$ endlich ist, so ist auch $\|f+g\|_p$ endlich, und damit gilt die Behauptung immer. $\square$

Wie die Argumente schon zeigen, die Zuordnung

$$f \mapsto \|f\|_p := \big\||f|\big\|_p$$

definiert eine Norm, die p-Norm. Man bildet dazu die L^p-Räume, wie wir den Raum $L^1(\mu)$ gebildet haben. Sie spielen eine wichtige Rolle in der Analysis.