9.7 Konvexität

Wir führen den in vieler Hinsicht wichtigen Begriff der Konvexität ein und beleuchten dabei auch die Rolle der zweiten Ableitung. Die ersten systematischen Untersuchungen der konvexen Funktionen stammen von dem dänischen Ingenieur und Mathematiker J. L. Jensen (1859–1925).

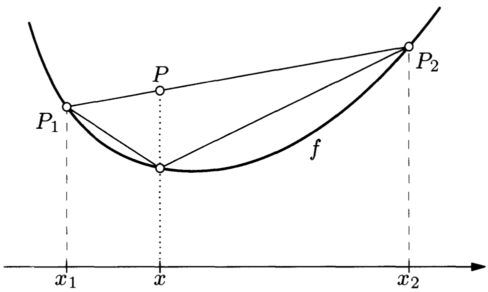
Eine reelle Funktion f heißt *konvex* auf einem Intervall I, wenn die Sekante durch je zwei Punkte P_1, P_2 des Graphen oberhalb des Graphen liegt. Da die Sekante durch P_1 und P_2 durch die lineare Funktion

$$L(x) = \frac{x_2 - x}{x_2 - x_1} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2)$$

dargestellt wird, hat man folgende analytische Formulierung:

Definition: Sei I ein Intervall. $f : I \to \mathbb{R}$ heißt *konvex auf I*, wenn für jedes Tripel $x_1, x, x_2 \in I$ mit $x_1 < x < x_2$ folgende Ungleichung gilt:

(K) $$\boxed{f(x) \le \frac{x_2 - x}{x_2 - x_1} f(x_1) + \frac{x - x_1}{x_2 - x_1} f(x_2).}$$



Die Sekante durch P_1 und P_2 liegt oberhalb des Graphen von f

Da die Punkte $x \in (x_1; x_2)$ genau die Punkte $\lambda x_1 + (1-\lambda)x_2$ mit $\lambda \in (0;1)$ sind, kann man die Konvexitätsbedingung auch so formulieren:

Für jedes Punktepaar $x_1, x_2 \in I$ mit $x_1 \ne x_2$ und jede Zahl $\lambda \in (0;1)$ gilt:

(K′) $$\boxed{f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).}$$

Gilt in (K) bzw. (K′) statt $\le$ die Relation

$<$, so heißt f *streng konvex*,

$\ge$, so heißt f *konkav*,

$>$, so heißt f *streng konkav*.

Das Beispiel $f(x) = |x|$ zeigt, daß eine konvexe Funktion nicht differenzierbar sein muß. Für differenzierbare Funktionen charakterisieren wir die Konvexität als ein Wachstum der Ableitung. Den Zusammenhang stellt der folgende Hilfssatz her, der die Konvexität durch Differenzenquotienten ausdrückt.

Hilfssatz: *f ist genau dann konvex, wenn für jedes Tripel $x_1, x, x_2 \in I$ mit $x_1 < x < x_2$ folgende Ungleichung gilt:*

$$\frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(x_2) - f(x)}{x_2 - x}. \tag{13}$$

Ist f konvex, so gilt für jedes solche Tripel genauer

$$\frac{f(x) - f(x_1)}{x - x_1} \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_2) - f(x)}{x_2 - x}; \tag{13'}$$

siehe Abbildung.

Beweis: Wir multiplizieren in (K) mit der positiven Zahl $x_2 - x_1$ und erhalten die äquivalente Ungleichung

$$\big((x_2 - x) + (x - x_1)\big)f(x) \leq (x_2 - x)f(x_1) + (x - x_1)f(x_2).$$

Diese ist weiter äquivalent zu

$$(x_2 - x)\big(f(x) - f(x_1)\big) \leq (x - x_1)\big(f(x_2) - f(x)\big).$$

Division durch die positive Zahl $(x_2 - x)(x - x_1)$ ergibt dann (13).

Wir kommen zum Beweis von (13′): Aus (K) folgt zunächst

$$f(x) - f(x_1) \leq \frac{(x - x_1)\big(f(x_2) - f(x_1)\big)}{x_2 - x_1}$$

und daraus weiter die linke Ungleichung in (13′). Analog zeigt man die rechte. □

Konvexitätskriterium: *Eine in $[a; b]$ stetige und in $(a; b)$ differenzierbare Funktion f ist genau dann konvex in $[a; b]$, wenn die Ableitung f' in $(a; b)$ monoton wächst.*

Beweis: a) Sei f konvex und seien $x_1, x_2 \in (a; b)$ Punkte mit $x_1 < x_2$. Für jeden Zwischenpunkt $x \in (x_1; x_2)$ gilt dann (13′), woraus mit $x \downarrow x_1$ einerseits und mit $x \uparrow x_2$ andererseits folgt:

$$f'(x_1) \leq \frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq f'(x_2).$$

f' wächst also monoton.

b) Sei umgekehrt f' monoton wachsend. Wir zeigen, daß f das Konvexitätskriterium (13) erfüllt. Sei dazu x_1, x, x_2 ein Tripel in $[a;b]$ mit $x_1 < x < x_2$. Nach dem Mittelwertsatz gibt es Punkte $\xi_1 \in (x_1;x)$ und $\xi_2 \in (x;x_2)$ so, daß

$$\frac{f(x) - f(x_1)}{x - x_1} = f'(\xi_1) \quad \text{und} \quad \frac{f(x_2) - f(x)}{x_2 - x} = f'(\xi_2).$$

Da $\xi_1 < \xi_2$ ist und f' monoton wächst, folgt (13). □

Folgerung 1: *Sei $f : [a;b] \to \mathbb{R}$ stetig und in $(a;b)$ 2-mal differenzierbar. Dann gilt:*

(i) *f ist genau dann konvex, wenn in $(a;b)$ $f'' \geq 0$ ist.*

(ii) *f ist streng konvex, wenn $f'' > 0$ ist.*

Beweis: (i) f' wächst genau dann monoton in $(a;b)$, wenn dort $f'' \geq 0$.
(ii) Andernfalls gibt es ein Tripel x_1, x, x_2 in $[a;b]$ mit $x_1 < x < x_2$, so daß in (13) Gleichheit gilt. Nach dem Mittelwertsatz gibt es dann weiter Punkte $\xi_1 \in (x_1;x)$ und $\xi_2 \in (x;x_2)$ so, daß

$$f'(\xi_1) = \frac{f(x) - f(x_1)}{x - x_1} = \frac{f(x_2) - f(x)}{x_2 - x} = f'(\xi_2).$$

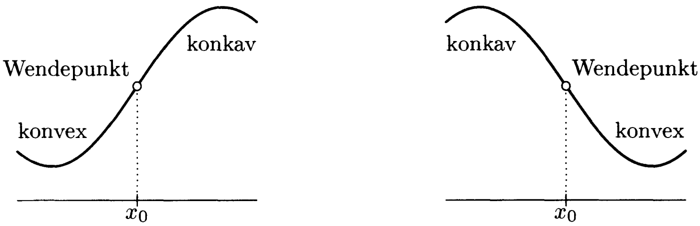
Das aber widerspricht der strengen Monotonie von f'. □

Beispiel: e^x ist streng konvex auf $\mathbb{R}$ und $\ln x$ streng konkav auf $\mathbb{R}_+$.

Wendepunkte. Sei $f : (a;b) \to \mathbb{R}$ stetig. Wir sagen, f habe in x_0 einen *Wendepunkt*, wenn es Intervalle $(\alpha; x_0)$ und $(x_0; \beta)$ gibt so, daß eine der Bedingungen (14) oder (15) erfüllt ist:

(14) *f ist in $(\alpha; x_0)$ konvex und in $(x_0; \beta)$ konkav;*

(15) *f ist in $(\alpha; x_0)$ konkav und in $(x_0; \beta)$ konvex.*



Zum Beispiel hat die Funktion $f : \mathbb{R} \to \mathbb{R}$, $f(x) := \begin{cases} \sqrt{x} & \text{für } x \geq 0, \\ -\sqrt{|x|} & \text{für } x < 0, \end{cases}$ in 0 einen Wendepunkt; f ist in 0 nicht differenzierbar.

Für eine $\mathscr{C}^2$-Funktion f sind die Bedingungen (14) und (15) äquivalent zu (14′) bzw. (15′):

(14′) $\quad f'' \ge 0$ *in* $(\alpha; x_0)$ *und* $f'' \le 0$ *in* $(x_0; \beta)$;

(15′) $\quad f'' \le 0$ *in* $(\alpha; x_0)$ *und* $f'' \ge 0$ *in* $(x_0; \beta)$.

Insbesondere ergibt sich als notwendige Bedingung $f''(x_0) = 0$.

9.8 Konvexe Funktionen und Ungleichungen

Wir beweisen in diesem Abschnitt einige fundamentale Ungleichungen. Die Grundlage bildet folgende Verallgemeinerung von (K′) aus 9.7.

Ungleichung von Jensen: *Sei $f : I \to \mathbb{R}$ konvex. Sind $\lambda_1, \ldots, \lambda_n$ positive Zahlen mit $\lambda_1 + \ldots + \lambda_n = 1$, so gilt für beliebige $x_1, \ldots, x_n \in I$:*

$$(\mathrm{K}_n) \qquad \boxed{f(\lambda_1 x_1 + \ldots + \lambda_n x_n) \le \lambda_1 f(x_1) + \ldots + \lambda_n f(x_n).}$$

Ist f streng konvex, so gilt in (K_n) Gleichheit nur, wenn $x_1 = \ldots = x_n$. Für konkaves f gilt (K_n) mit $\ge$.

Beweis durch vollständige Induktion: Für $n = 1$ ist die Aussage trivial. Für den Schluß $n \to n + 1$ setzen wir

$$\lambda_1 + \ldots + \lambda_n =: \lambda \quad \text{und} \quad \frac{\lambda_1}{\lambda} x_1 + \ldots + \frac{\lambda_n}{\lambda} x_n =: x.$$

Offensichtlich ist $x \in I$. Mittels (K_2) und (K_n) folgt nun

$$(*) \qquad \begin{aligned} f\left(\sum_{\nu=1}^{n} \lambda_\nu x_\nu + \lambda_{n+1} x_{n+1}\right) &\le \lambda f(x) + \lambda_{n+1} f(x_{n+1}) \\ &\le \lambda \sum_{\nu=1}^{n} \frac{\lambda_\nu}{\lambda} f(x_\nu) + \lambda_{n+1} f(x_{n+1}). \end{aligned}$$

Das ist gerade (K_{n+1}).

Ist f streng konvex, so ergibt $(*)$ für die Gleichheit in (K_{n+1}) aufgrund der Definition zunächst $x = x_{n+1}$ und aufgrund der Induktionsannahme weiter $x_1 = \ldots = x_n$. Hieraus folgt die Gleichheit aller $x_1, \ldots, x_{n+1}$. □

Als Anwendung beweisen wir eine weitreichende Verallgemeinerung der in 2.5 Aufgabe 4 aufgestellten Ungleichung zwischen dem arithmetischen und dem geometrischen Mittel.

Ungleichung zwischen dem gewichteten arithmetischen und dem gewichteten geometrischen Mittel: *Sind $x_1, \dots, x_n$ beliebige positive Zahlen und $\lambda_1, \dots, \lambda_n$ positive Zahlen mit $\lambda_1 + \ldots + \lambda_n = 1$, so gilt:*

$$\boxed{{x_1}^{\lambda_1} \cdots {x_n}^{\lambda_n} \le \lambda_1 x_1 + \ldots + \lambda_n x_n;}$$

insbesondere gilt

$$\sqrt[n]{x_1 \cdots x_n} \le \frac{x_1 + \ldots + x_n}{n}.$$

Das Gleichheitszeichen steht jeweils nur, wenn $x_1 = \ldots = x_n$.

Die Zahlen ${x_1}^{\lambda_1} \cdots {x_n}^{\lambda_n}$ und $\lambda_1 x_1 + \ldots + \lambda_n x_n$ heißen *mit $\lambda_1, \dots, \lambda_n$ gewichtetes geometrisches* bzw. *arithmetisches Mittel* der Zahlen $x_1, \dots, x_n$.

Beweis: Da der natürliche Logarithmus wegen $\ln''(x) < 0$ konkav ist, gilt nach der Ungleichung von Jensen

$$\ln(\lambda_1 x_1 + \ldots + \lambda_n x_n) \ge \lambda_1 \ln x_1 + \ldots + \lambda_n \ln x_n.$$

Anwendung der Exponentialfunktion ergibt die behauptete Ungleichung. Die Aussage zur Gleichheit folgt aus der strengen Konkavität des ln. □

Mit Hilfe der Ungleichung zwischen dem arithmetischen und geometrischen Mittel leiten wir weiter die sogenannte Höldersche Ungleichung her (O. Hölder, 1859–1937). Diese enthält als Spezialfall die wichtige Cauchy-Schwarzsche Ungleichung. Zur Formulierung der Ungleichungen verwenden wir die *p-Norm* eines Vektors $z = (z_1, \dots, z_n) \in \mathbb{C}^n$. Man definiert

$$\|z\|_p := \left(\sum_{\nu=1}^{n} |z_\nu|^p \right)^{1/p}, \quad p \ge 1. \tag{16}$$

Höldersche Ungleichung: *Es seien $p, q > 1$ Zahlen mit $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Dann gilt für beliebige Vektoren $z, w \in \mathbb{C}^n$:*

$$\sum_{k=1}^{n} |z_k w_k| \le \|z\|_p \cdot \|w\|_q.$$

Im Fall $p = q = 2$ ist das die sogenannte

Cauchy-Schwarzsche Ungleichung:

$$\boxed{|\langle z, w \rangle| \le \|z\| \cdot \|w\|.}$$

Hierbei bezeichnen $\langle\ ,\ \rangle$ das *Standardskalarprodukt* und $\|\ \|$ die *euklidische* Norm in $\mathbb{C}^n$:

$$\langle z, w \rangle := \sum_{k=1}^{n} z_k \overline{w_k}, \quad \|z\|_2 := \|z\| := \sqrt{\sum_{\nu=1}^{n} |z_\nu|^2}.$$

Beweis: Es genügt, den Fall $z \neq 0$ und $w \neq 0$ zu behandeln. Nach der Ungleichung zwischen dem arithmetischen und geometrischen Mittel gilt

$$\frac{|z_k w_k|}{\|z\|_p \|w\|_q} \leq \frac{1}{p}\frac{|z_k|^p}{\|z\|_p^p} + \frac{1}{q}\frac{|w_k|^q}{\|w\|_q^q}.$$

Durch Summation ergibt sich daraus bereits die behauptete Ungleichung

$$\frac{1}{\|z\|_p \|w\|_q} \sum_k |z_k w_k| \leq \frac{1}{p} + \frac{1}{q} = 1. \qquad \square$$

Bemerkung: Die Cauchy-Schwarzsche Ungleichung folgt auch sofort aus

$$\left(\sum_{k=1}^n |z_k|^2\right) \cdot \left(\sum_{k=1}^n |w_k|^2\right) - \left|\sum_{k=1}^n z_k \overline{w_k}\right|^2 = \sum_{k>l}^n |z_k w_l - z_l w_k|^2.$$

Diese Identität impliziert weiter, daß die Gleichheit $|\langle z, w\rangle| = \|z\| \cdot \|w\|$ genau dann eintritt, wenn $z_k w_l = z_l w_k$ für alle k, l gilt, d.h., wenn z und w linear abhängig sind.

Aus der Hölderschen Ungleichung leiten wir schließlich die Dreiecksungleichung der p-Norm ab.

Minkowskische Ungleichung: *Für $p \geq 1$ gilt:*

$$\boxed{\|z + w\|_p \leq \|z\|_p + \|w\|_p} \qquad z, w \in \mathbb{C}^n.$$

Beweis: Für $p = 1$ folgt die Behauptung unmittelbar aus der Dreiecksungleichung für Zahlen. Sei also im folgenden $p > 1$. Mit $s_k := |z_k + w_k|^{p-1}$, $s = (s_1, \ldots, s_n)$ und q derart, daß $1/q + 1/p = 1$ ist, ergibt die Höldersche Ungleichung ausgehend von $|z_k + w_k|^p \leq |z_k||s_k| + |w_k||s_k|$

$$(*) \qquad \|z + w\|_p^p \leq \|z\|_p \|s\|_q + \|w\|_p \|s\|_q.$$

Nun ist $s_k^q = |z_k + w_k|^p$. Damit folgt

$$\|s\|_q = \left(\sum_{k=1}^n |s_k|^q\right)^{1/q} = \left(\sum_{k=1}^n |z_k + w_k|^p\right)^{1/p \cdot p/q} = \|z + w\|_p^{p-1}.$$

In Verbindung mit $(*)$ beweist das die Behauptung. $\square$

Minkowski, Hermann (1864–1909). Wirkte zunächst in Zürich und später auf Betreiben Hilberts in Göttingen. In tiefgründigen Arbeiten entwickelt er eine *Geometrie der Zahlen.* Später wendet er sich der Mathematischen Physik zu. Sein Konzept der Raum-Zeit-Union (Minkowski-Welt) fördert entscheidend die spezielle Relativitätstheorie.