Mathematik für Wirtschaftswissenschaftler

Vorlesungsprogramm für den 19. 04. 2007

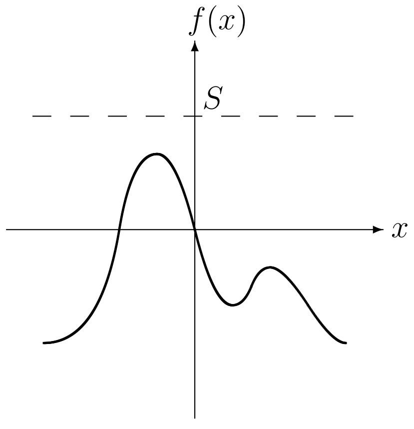
(K. Steffen, Heinrich-Heine-Universität Düsseldorf, SS 2007)

4.2 Grundeigenschaften von Funktionen

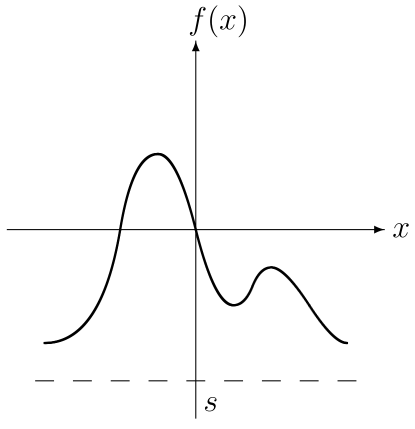
Wir besprechen in diesem Abschnitt Eigenschaften von Funktionen, die ganz besonders wichtig für die Anwendungen in der Ökonomie sind: Beschränktheit bzw. Unbeschränktheit, Positivität, monotones Wachstum, progressives bzw. degressives Wachstum und Konvexität. Diese Grundeigenschaften werden uns in der gesamten Analysis der Funktionen von einer Veränderlichen immer wieder begegnen. Für die einfachsten elementaren Funktionen (potenzfunktionen und Exponentialfunktionen) klären wir in diesem Abschnitt auch, ob sie die diskutierten Grundeigenschaften besitzen oder nicht. Ein wesentliches Ziel der Differentialrechnung ist, die Grundeigenschaften auch bei komplizierter aufgebauten elementaren Funktionen f überprüfen zu können. Dies ist oft in einfacher Weise möglich, indem man ihre Ableitung f' berechnet (siehe 4.3) und die Hauptsätze der Differentialrechnung anwendet (siehe 4.4, 4.5).

DEFINITION und DISKUSSION (***(Un–)Beschränktheit von Funktionen***):

1) Eine Funktion $f:\mathbb{R}\supset D\to\mathbb{R}$ heißt **von oben beschränkt** auf D, wenn es eine Zahl $S\in\mathbb{R}$ gibt, die von keinem Funktionswert auf D übertroffen wird, wenn also $f(x)\le S$ ist für alle $x\in D$; eine solche Zahl S heißt dann eine **obere Schranke** für f auf D. Gibt es keine obere Schranke, d.h. f nimmt beliebig große Werte auf D an, so nennt man f *von oben unbeschränkt* auf D. Analog heißt $s\in\mathbb{R}$ eine **untere Schranke** für die Funktion f auf D, wenn $f(x)\ge s$ ist für alle $x\in D$, und f heißt **von unten beschränkt** auf D, wenn eine untere Schranke existiert, bzw. *von unten unbeschränkt*, wenn es keine untere Schranke gibt, wenn also f negative Werte von beliebig großem Betrag auf D annimmt. Ist f von beiden Seiten beschränkt, also von oben und von unten beschränkt, so heißt f eine **beschränkte Funktion** auf D. Äquivalent ist, dass der Betrag von f von oben beschränkt ist (von unten sowieso durch Null), dass also $|f(x)|$ für alle $x\in\mathbb{R}$ unter einer festen endlichen Schranke bleibt. Ist dagegen f unbeschränkt auf D, so ist f von oben oder von unten oder von beiden Seiten unbeschränkt, und in jedem Fall nimmt $|f(x)|$ beliebig große Werte an, wenn x durch D läuft (was genau bedeutet, dass es zu jeder noch so groß vorgegebenen reellen Zahl r ein $x\in D$ mit $|f(x)|>r$ gibt). Die (Un–)Beschränktheit einer Funktion hängt natürlich wesentlich von dem gewählten Definitionsbereich ab!



obere Schranke S für f



untere Schranke s für f

2) Da reale Größen nie beliebig groß werden, sind auch die Funktionen, mit denen ökonomische Variablen modelliert werden, meistens auf ökonomisch sinnvolloen Definitionsbereichen beschränkt. Es gibt aber Ausnahmen wie z.B. die Stückkostenfunktion $k(x) = K(x)/x$ für die Produktion von x Mengeneinheiten eines Produkts, wenn die Kostenfunktion $K(x)$ einen positiven Fixkostenanteil K_{fix} für das Betreiben der Produktionsanlage enthält (unabhängig von der produzierten Menge). In dieser Situation ist $k(x) > K_{\text{fix}}$ für alle $x > 0$, und daher wird $K(x) > K_{\text{fix}}/x$ beliebig groß, wenn der Produktionsoutput $x > 0$ klein genug wird, und die Stückkostenfunktion k ist auf keinem Intervall $]0, x_0]$ mit $x_0 > 0$ von oben beschränkt. Zwar wird es in der Realität keinen beliebig kleinen Output geben, aber die Überlegung zeigt, dass es sinnvoll ist, eine Stückkostenfunktion auf dem Intervall $]0, x_{\text{max}}]$ zwischen 0 und dem maximal produzierbaren Output durch eine mathematische Funktion zu modellieren, die unbeschränkt ist auf jedem Intervall $]0, x_0]$ (und genauer sogar den Limes $+\infty$ hat bei $x \downarrow 0$, was eben bedeutet, dass $k(x)$ jede gegebene Zahl übertrifft, wenn nur $x > 0$ klein genug ist).

3) Dagegen ist die Stückkostenfunktion — wie die meisten “ökonomischen Funktionen” — von unten beschränkt, weil ihre Funktionswerte ja Null nicht unterschreiten. Allgemein nennen wir f eine **nichtnegative Funktion** (auf D), wenn Null eine untere Schranke für f ist, wenn also alle Funktionswerte $f(x)$ (zu $x \in D$) nichtnegativ sind, und f heißt **positive Funktion**, wenn sogar alle Funktionswerte positiv sind, wenn also $f(x) > 0$ gilt für alle x aus dem Definitionsbereich D. Geometrisch bedeutet letzteres, dass der Graph der Funktion echt oberhalb der horizontalen Koordinatenachse verläuft, während er bei einer nichtnegativen Funktion die Achse auch in gewissen Punkten von D berühren kann, die den Nullstellen von f in D entsprechen. (Aber eine nichtnegative Funktion braucht keine Nullstellen zu haben; jede positive Funktion ist erst recht nichtnegativ!) Analog erklärt man *negative Funktion* bzw. *nichtpositive Funktion*; da aber ökonomische Variablen, wie gesagt, meist nichtnegativ sind, ist dies weniger wichtig. Natürlich hängt die (Nicht–) Positivität einer Funktion wesentlich von dem zu Grunde gelegten Definitionsbereich ab.

4) Wenn eine Funktion f von oben beschränkt ist auf D, so gibt es stets eine kleinste obere Schranke $S^* \in \mathbb{R}$, nämlich die obere Grenze der Wertemenge von f, genannt das **Supremum** von f auf D und notiert $\sup_{x\in D} f(x)$ oder kurz $\sup_D f$. Jede kleinere Zahl als das Supremum wird durch Funktionswerte von f auf D überschritten (weil dieser kleinere Wert ja keine obere Schranke für f mehr sein kann) und das Supremum selbst wird von keinem Funktionswert übertroffen (weil es ja obere Schranke ist). Ist f dagegen von oben unbeschränkt, so schreibt man symbolisch $\sup_D f := \infty$; damit ist ausgedrückt, dass es keine endliche obere Schranke für f auf D gibt, so dass ∞ formal die kleinste obere Schranke ist. Analog bezeichnet $\inf_{x\in D} f(x) = \inf_D f$ die größte untere Schranke S_* der Funktion f auf D, genannt das **Infimum** von f auf D; das Infimum ist endlich, genau wenn f von unten beschränkt ist, andernfalls ist es $-\infty$.

5) Das Supremum einer Funktion muss nicht selbst Funktionswert sein, auch wenn die Funktion von oben beschränkt ist, das Supremum also endlich ausfällt. Vielmehr ist das dann und nur dann der Fall, wenn f ein **(absolutes) Maximum** in seinem Definitionsbereich D annimmt, wenn es also eine Stelle $x^* \in D$ gibt, deren Funktionswert $f(x^*)$ größer ist (oder jedenfalls nicht kleiner) als alle Funktionswerte $f(x)$ an anderen Stellen $x \in D$. In diesem Fall schreibt man auch $\max_{x\in D} f(x)$ oder kurz $\max_D f$ für diesen maximalen Funktionswert $f(x^*)$, der dann natürlich gleich dem Supremum $\sup_{x\in D} f(x)$ ist.

Analog ist das Infimum von f auf D genau dann ein Funktionswert, wenn f ein **Minimum**, also einen kleinsten Funktionswert, auf D annimmt, und in diesem Falle ist das Infimum gleich dem minimalen Funktionswert $\min_D f$.

Eine von oben unbeschränkte Funktion hat natürlich nie ein Maximum. Und ist die Funktion von oben beschränkt, so hängt die Existenz eines Maximums wesentlich von dem ins Auge gefassten Definitionsbereich ab. Die Funktion $f(x) = x$ auf $D =]0,1[$ kann das klar machen: die Wertemenge ist ebenfalls $]0,1[$, die kleinste obere Schranke ist $\sup_{]0,1[} f = 1$ und das Infimum ist $\inf_{]0,1[} f = 0$, aber die Funktion nimmt kein Maximum auf $]0,1[$ an, weil ihre Werte jede Zahl kleiner als 1 übertreffen, die Zahl 1 selbst aber nicht erreichen oder übertreffen, und ebensowenig hat die Funktion ein Minimum auf dem offenen Intervall $]0,1[$. Auf dem kompakten Intervall $[0,1]$ dagegen hat $f(x) = x$ ebenfalls das Supremum 1 und das Infimum 0, aber diese Werte werden von der Funktion nun auch angenommen und sind das Maximum bzw. das Minimum von f auf $[0,1]$.

6) *Operationen, die Beschränktheit erhalten*, sind Bildung von Summen, Differenzen, Linearkombinationen und Produkten beschränkter Funktionen mit gemeinsamem Definitionsbereich sowie die Verkettung $h \circ f$, wenn die äußere Funktion h auf der Wertemenge der innere Funktion f beschränkt ist. Das ist leicht zu sehen, und etwas genauer kann man sich entsprechende Aussagen für einseitige Beschränktheit überlegen; z.B. sind Linearkombinationen $s_1 f_1 + \ldots + s_n f_n$ von Funktionen, die von oben beschränkt sind auf D, wieder von oben beschränkt auf D, wenn die Koeffizienten s_i alle nichtnegativ sind. Problematisch ist dagegen die Bildung der reziproken Funktion $1/g$ zu einer Funktion ohne Nullstelle auf D. Hier kommt es nicht darauf an, dass g beschränkt ist, sondern dass g **von Null weg beschränkt** ist, d.h. dass alle Werte $g(x)$ einen gewissen positiven Mindestabstand δ zu 0 haben, also $|g(x)| \geq \delta$ für alle $x \in D$. Dann und nur dann ist $1/g$ beschränkt, nämlich $|1/g(x)| \leq \frac{1}{\delta}$ für alle $x \in D$. Und ein Quotient f/g ist beschränkt auf D, wenn f beschränkt ist und g von Null weg beschränkt auf D. (Allerdings kann f/g auch ohne diese Bedingungen an f und g beschränkt sein, nämlich wenn $|f(x)|$ klein ist von mindestens derselben Größenordnung wie $|g(x)|$ an Stellen x, wo $|g(x)|$ klein ist, und wenn $|g(x)|$ groß ist von mindestens derselben Größenordnung wie $|f(x)|$ an Stellen x, wo $|f(x)|$ groß ist. Ein Beispiel ist $f(x) = \ln(1+x)$, $g(x) = x$ auf $\mathbb{R}_{>0}$ mit $0 < f(x)/g(x) < 1$ für $x > 0$ wegen der fundamentalen Ungleichung $\ln(1+x) < x$ für $-1 < x \neq 0$; s. 1.4.) ■

BEISPIELE (***zur (Un–)Beschränktheit von Funktionen***):

(1) Allgemeine *Potenzfunktionen* x^s sind auf ihrem Definitionsbereich $\mathbb{R}_{>0}$ natürlich nie von oben beschränkt, mit Ausnahme der konstanten Funktion $\equiv 1$, die sich im Fall des Exponenten $s = 0$ ergibt. Im Fall $s > 0$ wird nämlich x^s beliebig groß bei $x \to \infty$; man braucht ja nur $x > S^{1/s}$ zu wählen, um zu erreichen, dass x^s größer ist als die beliebig groß gegebene Zahl S. (Das gilt übrigens auch für "Wurzelfunktionen", also $0 < s < 1$, wofür x^s zwar langsamer groß wird als x, aber immer noch beliebig groß bei $x \to \infty$.) Im Fall $s < 0$ dagegen wird x^s beliebig groß bei $x \searrow 0$; man braucht ja nur $0 < x < S^{1/s}$ zu wählen, um $x^s > S$ zu erreichen. (Für den Größenvergleich von Potenzen mit verschiedenen Basen siehe 1.4). Dagegen sind alle diese Potenzfunktionen auf $\mathbb{R}_{>0}$ von unten beschränkt, nämlich sogar positiv (im Fall $s > 0$ auf $\mathbb{R}_{\geq 0}$ aber nur nichtnegativ, da $0^s = 0$).

Will man eine beschränkte Potenzfunktion haben, so muss man das Definitionsintervall verkleinern. Da $x^s \leq b^s$ ist für $s > 0$ und $0 < x < b$, sind die Potenzfunktionen mit positivem Exponenten s beschränkt auf jedem kompakten Intervall $[0,b]$, während die mit negativem Exponenten auf jedem Intervall $[a,\infty[$ mit positivem linken Randpunkt a beschränkt sind wegen $x^s \leq a^s$ für $a \leq x$ und $s < 0$.

Für ganze Exponenten $s = n \neq 0$ sind die Potenzfunktionen x^n auf $\mathbb{R}$ nichtnegativ bzw. im Fall $n < 0$ auf $\mathbb{R}_{\neq 0}$ positiv, nur wenn n gerade ist. Bei ungeradem n dagegen ist $x^n = -(-x)^n$ auf diesen Definitionsbereichen auch von unten unbeschränkt.

(2) Allgemeine *Exponentialfunktionen* $c^x = \exp(x \ln c)$ zu Basen $c > 0$ sind positive Funktionen auf ganz $\mathbb{R}$, nehmen aber (außer für $c = 1$) beliebig große Werte an auf $\mathbb{R}$, da $c^x > S$ ist für $x \ln c > S > 0$. Wegen $c^x \le c^b$ für $x \le b$ und $c > 1$ ist die Exponentialfunktion zu einer Basis $c > 1$ auf jedem von rechts beschränkten Intervall $]-\infty, b]$ in $\mathbb{R}$ beschränkt, und die Funktion $(1/c)^x = c^{-x}$ entsprechend auf jedem von links beschränkten Intervall. (Für den Größenvergleich von Potenzen mit verschiedenen Exponenten siehe 1.4.)

(3) Die *Logarithmusfunktionen* $\log_c x$ zu Basen $0 < c \neq 1$ sind auf $\mathbb{R}_{>0}$ beidseitig unbeschränkt; denn im Fall $c > 1$ etwa ist $\log_c x \ge S$, wenn $x \ge c^S$ gemacht wird, und für $0 < x < c^{-S}$ hat man $\log_c x \le -S$. Auf kompakten Intervallen $[a, b]$, deren linker Randpunkt positiv ist, haben die Logarithmusfunktionen $\log_c x$ Werte zwischen $\log_c a$ und $\log_c b$ und sind daher beschränkt (siehe 1.4 für den Größenvergleich von Logarithmen).

(4) *Polynomfunktionen* $p(x) = c_m x^n + \ldots + c_1 x + c_0$ vom Grad $m \ge 1$ können auf ganz $\mathbb{R}$ nicht beschränkt sein, weil sie sich ja für große Werte von $|x|$ wie der führende Term $c_m x^m$ verhalten, also beliebig große Werte annehmen (wenn $c_m > 0$ oder m ungerade) oder beliebig betragsgroße negative Werte (wenn $c_m < 0$ oder m ungerade). Für x aus einem kompakten Intervall $[-b, b]$ hat man dagegen eine endliche Schranke $|p(x)| \le |c_m| b^m + \ldots + |c_1| b + |c_0|$, also ist jede Polynomfunktion beschränkt auf solchen Intervallen und überhaupt auf jedem beidseitig beschränkten Intervall (liegt in $[-b, b]$, wenn b groß genug gewählt ist). Da bei hinreichend großer Wahl von b alle Werte $p(x)$ für $|x| \ge b$ beliebig groß sind, wenn m gerade und $c_m > 0$, ist in diesem Fall (und nur dann) die Polynomfunktion immerhin noch von unten beschränkt auf ganz $\mathbb{R}$, im Fall m gerade und $c_m < 0$ ist sie analog von oben beschränkt (und nur dann).

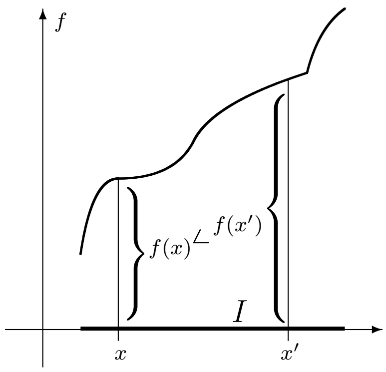
(5) Bei *rationalen Funktionen* $r(x) = \frac{p(x)}{q(x)} = \frac{c_m x^m + \ldots + c_1 x + c_0}{d_n x^n + \ldots + d_1 x + d_0}$ verläuft die Diskussion für große $|x|$ wie bei Polynomen durch Inspektion des führenden Terms $(c_m/d_n) x^{m-n}$. Im Fall $m > n$ ist die Funktion auf ihrem maximalen Definitionsbereich D (im Unendlichen) unbeschränkt. Hat die Funktion Polstellen, so ist sie auf D natürlich auch unbeschränkt, da sie bei jeder Polstelle Werte von beliebig großem Betrag annimmt. Nur wenn alle Polstellen gerade Ordnung haben und $m-n$ gerade oder nichtpositiv ist, kann die Funktion einseitig beschränkt sein (und ist es auch). Hat sie keine Pole und ist $m \le n$, so ist die rationale Funktion beschränkt auf $\mathbb{R}$; denn außerhalb eines genügend großen Intervalls $[-b, b]$ hat sie Werte mit beliebig kleinem Abstand zum Limes c_n/d_n im Unendlichen ($c_n := 0$, wenn $m < n$), auf dem kompakten Intervall $[-b, b]$ haben wir aber nach (4) eine endliche obere Schranke S für die Werte $|p(x)|$, und es gibt auch eine positive untere Schranke $s > 0$ für die Werte $|q(x)|$, so dass $|r(x)| \le \frac{S}{s}$ für $x \in [-b, b]$ folgt. Dabei haben wir die aus dem Satz von Maximum (s.u.) folgende Tatsache benutzt, dass $|q|$ auf jedem kompakten Intervall ohne Nullstelle ein positives Minimum annimmt. Beispiele für auf $\mathbb{R}$ beschränkte rationale Funktionen haben wir schon in 4.1 gesehen: $\frac{p(x)}{1+x^2}$ mit $\text{Grad}(p) \le 2$ oder $\frac{p(x)}{(1+x^2)^l}$ mit $\text{Grad}(p) \le 2l$.

(6) Die *logistische Funktion* $f(x) = \dfrac{a + b\mathrm{e}^{cx}}{1 + \mathrm{e}^{cx}}$ hat Werte echt zwischen ihren Sättigungsgrenzen $a < b$, ist also beschränkt auf $\mathbb{R}$. Das sieht man aus $a = \frac{a+a\mathrm{e}^{cx}}{1+\mathrm{e}^{cx}} < \frac{a+b\mathrm{e}^{cx}}{1+\mathrm{e}^{cx}} < \frac{b+b\mathrm{e}^{cx}}{1+\mathrm{e}^{cx}} = b$ (da $\mathrm{e}^{cx} > 0$). Weil ihre Werte die Sättigungsgrenzen nicht erreichen, ihnen aber beliebig nahe kommen (wenn $c \neq 0$), ist $a = \inf_{\mathbb{R}} f$ das Infimum und $b = \sup_{\mathbb{R}} f$ das Supremum der logistischen Funktion auf $\mathbb{R}$, aber sie nimmt auf $\mathbb{R}$ weder Maximum noch Minimum an. (Siehe die Abbildung des Graphen in 4.1.) ■

Die letzte Feststellung, dass auf $\mathbb{R}$ kein Maximum oder Minimum angenommen wird, gilt offenbar für jede streng monotone Funktion f auf $\mathbb{R}$, auch wenn sie, wie die logistische Funktion, beschränkt ist; denn zu jedem Wert $f(x)$ bekommt man größere und kleinere Werte $f(x')$, $f(x'')$, wenn man x', x'' auf verschiedenen Seiten von x wählt. Auf einem kompakten Intervall $[a, b]$ ist eine streng monotone Funktion dagegen immer beschränkt; denn ihre Werte liegen zwischen den endlichen Schranken $f(a)$ und $f(b)$. Diese Bemerkungen bringen uns zur nächsten – und für die Wirtschaftsmathematik ganz fundamentalen – Grundeigenschaft von Funktionen:

DEFINITION und DISKUSSION (***Monotonie von Funktionen***):

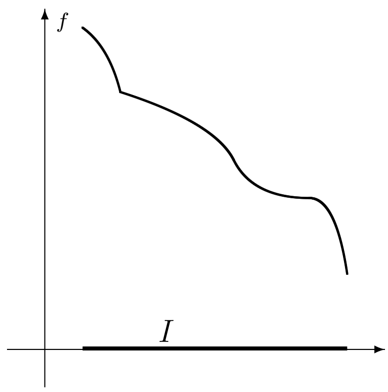
1) Eine auf einem Intervall I (mit Länge $\neq 0$) in $\mathbb{R}$ definierte reelle Funktion f heißt **(streng monoton) wachsende Funktion** (oder *isotone Funktion*) auf I, wenn $f(x)$ zunimmt bei Vergrößerung des Arguments x im Intervall I, wenn also gilt $f(x) < f(x')$ für alle $x < x'$ in I. Verlangt man dagegen nur $f(x) \leq f(x')$ für $x < x'$ in I, lässt also zu, dass f an der Stelle $x' > x$ und damit an allen Stellen im Intervall $[x, x']$ denselben Wert hat wie an der Stelle x, so nennt man f **nichtfallende Funktion** oder **schwach (monoton) wachsende Funktion**. Wenn wir von "Wachstum" ohne Präzisierung sprechen, so meinen wir stets strenges Wachstum. (Warnung: Das halten nicht alle Autoren so!)



eine wachsende Funktion f

Geometrisch bedeutet Wachstum von f, dass der Funktionsgraph ansteigt, wenn man ihn von links nach rechts durchläuft, während er bei schwachem Wachstum auch horizontale Strecken positiver Länge enthalten kann ("kann", nicht "muss"; denn eine streng wachsende Funktion ist im formalen Sinne der Definition erst recht schwach wachsend). Man stellt sich eine wachsende Funktion so vor, wie es hier abgebildet ist (obwohl sie z.B. auch Sprungstellen haben könnte).

2) Völlig analog definiert man **(streng monoton) fallende Funktionen** f (die auch *antitone Funktionen* genannt werden) auf I durch die Bedingung $f(x) > f(x')$ für alle $x < x'$ in I, und entsprechend **nichtwachsende Funktionen** bzw. **schwach (monoton) fallende Funktionen** auf I durch die schwächere Bedingung $f(x) \geq f(x')$ für $x < x'$ in I. Für den Graphen bedeutet dies, dass er beim Durchlaufen von links nach rechts abfällt, wobei er im Falle schwachen Fallens auch horizontale Strecken enthalten kann.



eine fallende Funktion f

3) Unter einer **streng monotonen Funktion** verstehen wir eine auf dem gesamten Definitionsintervall I wachsende oder auf ganz I fallende; eine *monotone Funktion* auf I ist entsprechend nichtfallend oder nichtwachsend auf ganz I. Die mathematischen Funktionen, die wir zur Modellierung ökonomischer Vorgänge verwenden, sind oft streng monoton, häufig aber auch nur auf gewissen Teilintervallen ihres Definitionsbereichs I wachsend, auf anderen Teilintervallen fallend und dann natürlich nicht auf ganz I monoton. Die (Nicht–) Monotonie einer Funktion hängt also von dem betrachteten Definitionsintervall ab.

4) Aus den Grundgesetzen für die Ordnung der reellen Zahlen und aus den Rechenregeln für Ungleichungen in 1.4 ergeben sich unmittelbar folgende Aussagen über *Operationen, die Monotonie erhalten*:

Mit f ist auch jedes Vielfache rf mit positivem Faktor $r > 0$ eine (streng) monotone Funktion auf I, und zwar *gleichsinnig monoton* mit f, d.h. f und rf sind beide nichtfallend oder beide nichtwachsend. Dagegen ist sf für $s < 0$ *gegensinnig monoton* zu f, d.h. wenn f wachsend ist, so fällt sf, und wenn f fallend ist, so wächst sf. Die Summe $f{+}g$ von zwei gleichsinnig monotonen Funktionen f, g auf I ist in demselben Sinne monoton wie f und g, und zwar streng, wenn mindestens einer der Summanden streng monoton ist. Dagegen kann man keine generelle Monotonieaussage über die die Differenz $f{-}g$ von zwei gleichsinnig monotonen Funktionen auf I machen bzw., was auf dasselbe hinausläuft, über die Summe von zwei gegensinnig monotonen Funktionen. Diese Differenz ist im Allgemeinen nicht monoton auf I und wenn sie monoton ist, so kann die Richtung der Monotonie wachsend oder fallend sein, jenachdem ob sich die Monotonie von f oder die von g stärker auswirkt. Das kann man an einfachen Beispielen wie $f(x) = x^2$, $g(x) = 2x{-}1$ erkennen; beide Funktionen sind streng wachsend auf $[0, \infty[$, aber die Differenz $f(x) - g(x) = (x{-}1)^2$ ist streng fallend auf $[0, 1]$, streng wachsend auf $[1, \infty[$ und auf $[0, \infty[$ nicht monoton.

Ein Produkt fg von positiven gleichsinnig monotonen Funktionen f, g auf I ist in demselben Sinne monoton auf I, und zwar streng, wenn mindestens einer der Faktoren streng monoton ist. Ohne die Positivitätsbedingung an die Funktionen gilt dies im Allgemeinen nicht, wie man z.B. an $f(x) = g(x) := x$ mit $f(x)g(x) = x^2$ auf $I = \mathbb{R}$ sieht. Ist g eine positive (streng) monotone Funktion, so ist die reziproke Funktion $1/g$ (streng) monoton im entgegengesetzten Sinn. Ein Quotient f/g ist folglich monoton, wenn f und g gegensinnig monotone positive Funktionen sind. Bei gleichsinniger Monotonie von f, g kann man dagegen über das Monotonieverhalten von f/g keine generelle Aussage machen.

Schließlich ist die Verkettung $h \circ f$ von zwei monotonen Funktionen $f\colon I \to J \subset \mathbb{R}$ und $h\colon J \to \mathbb{R}$ wieder monoton, und zwar im gleichen Sinne wie die innere Funktion f, wenn die äußere Funktion h wachsend ist, aber gegensinnig monoton zu f, wenn h fällt; dabei ist $h \circ f$ streng monoton, wenn f und h beide streng monoton sind. Jede streng monotone Funktion $f\colon I \to \mathbb{R}$ ist injektiv, also umkehrbar (verschiedene Argumente haben verschiedene Werte). Ist die Wertemenge $J := f(I)$ wieder ein Intervall, so ist dann auch die Umkehrfunktion $f^{-1}\colon J \to \mathbb{R}$ streng monoton im gleichen Sinne wie f; denn aus $y = f(x) < y' = f(x')$ folgt $f^{-1}(y) = x < x' = f^{-1}(y')$ bei strengem Wachstum von f (sonst gälte $x \geq x'$ und damit $f(x) \geq f(x')$) bzw. $f^{-1}(y) = x > x' = f^{-1}(y')$ bei strengem Fallen von f (sonst gälte $x \leq x'$ und damit $f(x) \geq f(x')$). Geometrisch ist auch klar, dass die Spiegelung an der Diagonalen, durch die ja der Graph von f^{-1} aus dem Graphen von f entsteht, das Aufsteigen bzw. Absteigen der Graphenkurve erhält.

5) Die Relevanz des Monotoniekonzepts für die Beschreibung ökonomischer Vorgänge liegt auf der Hand: In vielen Situationen sind ökonomische Variablen derart gekoppelt, dass mit der einen auch die andere wächst, bzw. dass die zweite fällt, wenn die erst wächst. Zum Beispiel wachsen die Kosten $K(x)$ der Herstellung eines Produkts mit dem Output x, d.h. mit der hergestellten Menge des Produkts, während normalerweise die Stückkosten $k(x) = K(x)/x$, also die Kosten pro hergestellte Einheit des Produkts, mit anwachsenden Output x abnehmen. Also wird man Kostenfunktionen $K(x)$ mathematisch meistens durch wachsende Funktionen modellieren, für welche die zugehörige Stückkostenfunktion $k(x)$ fallend ist.

Monotonie haben wir auch bei der wechselseitigen Abhängigkeit von Preis und Angebot / Nachfrage: Mit steigendem Preis p wird die Nachfrage nach einem Produkt kleiner, ist also eine fallende Funktion des Preises, den ein Monopolist festsetzt. In einer Situation, in der das produzierte Angebot eines Produkts vom erzielbaren Preis abhängt, wird dagegen der Produktionsoutput mit steigendem Marktpreis größer, ist also durch eine wachsende Funktion des Marktpreises zu modellieren. Natürlich sind aber in der Ökonomie nicht alle wechselseitigen Abhängigkeiten von Variablen monoton. Beispielsweise wird der Gewinn eines Monopolisten mit dem von ihm festgesetzten Preis wachsen, so lange der Preis eine gewisse Schwelle nicht überschreitet, aber bei höheren Preisen wird der Gewinn eine fallende Funktion des Preises sein, weil Preissteigerungen durch einen Rückgang der Nachfrage überkompensiert werden. ■

Die wirksamste Methode, um die Monotonie einer durch einen Rechenterm gegebenen Funktion festzustellen, bzw. die maximalen Teilintervalle ihres Definitionsbereichs zu bestimmen, auf denen sie wachsend oder fallend ist, gibt uns später in diesem Kapitel die Differentialrechnung. In den folgenden Beispielen diskutieren wir die Monotonie von wichtigen Grundfunktionen und einigen weiteren elementaren Funktionen, bei denen die Analyse ohne Differentialrechnung allein mit den Rechenregeln für Ungleichungen möglich ist.

BEISPIELE (***zur Monotonie von Funktionen***):

(1) Für *Potenzfunktionen* haben wir die Monotoniediskussion schon in 1.4 beim Größenvergleich von Potenzen mit gleichen Exponenten geführt. Ein Blick auf die in 4.1 abgebildeten Graphen der Potenzfunktionen veranschaulicht ihr Monotonieverhalten. Auf $\mathbb{R}_{\geq 0}$ ist x^s wachsend für alle Exponenten $s > 0$, auf $\mathbb{R}_{>0}$ ist x^s fallend für alle $s < 0$. Für $s = 0$ ist Funktion $x^s \equiv 1$ konstant, also schwach wachsend und schwach fallend, aber nicht streng monoton.

Für ungerade Exponenten $n > 0$ ist die Potenzfunktion x^n sogar auf ganz $\mathbb{R}$ wachsend; das ergibt sich aus $(-x)^n = -x^n$ für $x \in \mathbb{R}$ und $x^n < 0 < y^n$ für $x < 0 < y$. Für gerade Exponenten $n > 0$ ist x^n dagegen nicht monoton auf irgendeinem Intervall mit 0 im Inneren, da die Funktion rechts von 0 streng wächst und links von 0 streng fällt.

(2) Aus dem Vergleich von Potenzen mit gleicher Basis in 1.4 entnimmt man das Monotonieverhalten von *Exponentialfunktionen* a^x. Für Basen $a > 1$ sind sie wachsend, für Basen $0 < a < 1$ sind sie fallend auf $\mathbb{R}$ (und für $a = 1$ konstant, also nicht streng monoton). Ein Blick auf die in 4.1 abgebildeten Graphen veranschaulicht das.

Die *Logarithmusfunktion* $\log_a$ ist die Umkehrfunktion zur Expponentialfunktion mit Basis $0 < a \neq 1$, also im gleichen Sinne streng monoton auf $\mathbb{R}_{>0}$ wie die Exponentialfunktion a^x auf $\mathbb{R}$. Das heißt konkret: Für Basen $a > 1$ ist $\log_a$ streng wachsend auf $\mathbb{R}$ (und für $0 < a < 1$ ist $\log_a = -\log_{1/a}$ streng fallend auf $\mathbb{R}_{>0}$).

(3) Die Monotonie–Diskussion ist einfach für *Polynome* vom Grad ≤ 2. Konstante Funktionen sind schwach monoton (wachsend und fallend), aber nicht streng monoton. Lineare Funktionen $ax+b$ sind streng wachsend auf $\mathbb{R}$ im Fall $a > 0$ und streng fallend auf $\mathbb{R}$ für $a < 0$. Bei quadratischen Funktionen $q(x) = ax^2+bx+c$ zeigt die Umformung mit quadratischer Ergänzung $q(x) = a(x+\frac{b}{2a})^2 - \frac{b^2-4ac}{4a}$ im Fall $a > 0$ strenges Wachstum auf dem Intervall $[\frac{-b}{2a}, \infty[$ rechts von der Minimumstelle und strenges Fallen auf dem Intervall $]-\infty, \frac{-b}{2a}]$ links davon; im Fall $a < 0$ ist die Funktion dagegen links von ihrer Maximumstelle $\frac{-b}{2a}$ streng wachsend und rechts davon streng fallend. Dieses Verhalten ist bei der nach oben bzw. unten offenen Parabel $\text{Graph}(q)$ auch geometrisch klar.

Bei Polynomen vom Grad ≥ 3 ist die Monotonie–Analyse ohne Differentialrechnung nur in günstigen Fällen möglich. Zum Beispiel ist eine Polynomfunktion $p(x) = c_m x^m + \ldots + c_1 x + c_0$ vom Grad $m \geq 1$ offenbar streng wachsend auf $\mathbb{R}_{\geq 0}$, wenn alle Koeffizienten c_k nichtnegativ sind und der führende c_m positiv. Verschwinden dabei alle Koeffizienten c_k zu Potenzen x^k mit geradem Exponenten $k > 0$, so ist die Funktion sogar auf ganz $\mathbb{R}$ streng wachsend. Das gilt z.B. für die Funktion $x^5 + x$, die wir in 4.1 als Beispiel einer umkehrbaren elementaren Funktion angegeben hatten, deren Umkehrfunktion nicht elementar ist. (Es gibt keine "elementare Lösungsformel" für $x^5 + x = y$.) Sind m reelle Nullstellen $x_1 \leq \ldots \leq x_m$ bekannt (jede mit ihrer Vielfachheit aufgeführt), so zeigt die Linearfaktorzerlegung $p(x) = c_m(x - x_1) \cdot \ldots \cdot (x - x_m)$, dass p auf den Intervallen $[x_m, \infty[$ bzw. $]-\infty, x_1]$ rechts bzw. links von den Nullstellen streng monoton ist in demselben Sinn wie der führende Term $c_m x^m$ rechts bzw. links von 0. Diese Monotonieaussage kann mit Differentialrechnung auch für ein allgemeines Polynom p auf Intervallen $[a, \infty[$ bzw. $]-\infty, b]$ mit genügend großer unterer Grenze a bzw. mit genügend kleiner negativer oberer Grenze b bewiesen werden.

(4) Bei *rationalen Funktionen* $r(x) = \frac{c_m x^m + \ldots + c_1 x + c_0}{d_n x^n + \ldots + d_1 x + d_0}$ ist die Monotonie–Diskussion ohne Differentialrechnung nur in sehr einfachen Fällen möglich. Für eine gebrochen–lineare Funktion $g(x) = \frac{ax+b}{cx+d}$ mit $ad - bc \neq 0$ und $c \neq 0$ zum Beispiel zeigt Polynomdivision $g(x) = \frac{a}{c} - \frac{ad-bc}{c^2} \frac{1}{x + d/c}$, dass die Funktion auf den Intervallen $]-\infty, \frac{-d}{c}[$ und $]\frac{-d}{c}, \infty[$ links bzw. rechts von ihrer Polstelle im gleichen Sinne streng monoton ist, und zwar streng wachsend, wenn $ad - bc < 0$, und streng fallend, wenn $ad - bc > 0$. Polynomdivision kann auch in anderen einfachen Fällen Aufschluss geben. So ist das Monotonieverhalten von $r(x) = \frac{x^2-1}{x}$ auf $\mathbb{R}_{>0}$ nicht unmittelbar klar, weil Zähler- und Nennerpolynom streng wachsen (und weil der Zähler überdies einen Vorzeichenwechsel bei $x = 1$ hat). Die dividierte Darstellung $r(x) = x - \frac{1}{x}$ zeigt aber, dass die Funktion streng wächst auf $\mathbb{R}_{>0}$. Allerdings funktioniert diese Argumentation bei $\widetilde{r}(x) = \frac{x^2+x}{x} = x + \frac{1}{x}$ schon nicht mehr, weil hier eine Summe von gegensinnig monotonen Funktionen vorliegt. Wegen $\widetilde{r}(x) \to \infty$ bei $x \searrow 0$ und bei $x \to \infty$ ist $\widetilde{r}$ sicher nicht monoton auf $\mathbb{R}_{>0}$. Die Darstellung $\widetilde{r}(x') - \widetilde{r}(x) = (x' - x)(1 - \frac{1}{xx'})$ zeigt, dass $\widetilde{r}$ tatsächlich streng wächst auf $[1, \infty[$ und streng fällt auf $]0, 1]$.

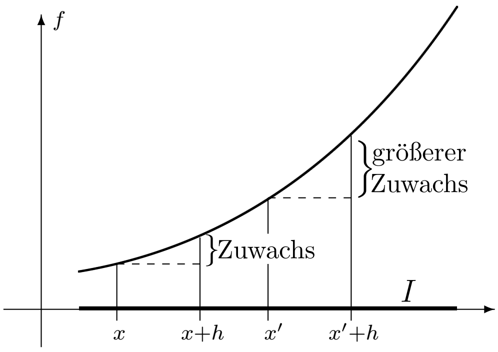
(5) Ein ökonomisches Beispiel einer gebrochen–linearen Funktion ist die Stückkostenfunktion $k(x) = K(x)/x = \kappa + K_{\text{fix}}/x$ zu einer linearen Kostenfunktion $K(x) = K_{\text{fix}} + \kappa x$ mit positiven Fixkosten K_{fix}. Auf $\mathbb{R}_{>0}$ ist $k(x)$ streng fallend (und auf dem ökonomisch bedeutungslosen Intervall $\mathbb{R}_{<0}$ ebenfalls). Übrigens erhält man dieses Ergebnis auch bei variablen Kosten κx^s, wenn diese schwächer anwachsen als x, d.h. wenn $0 < s < 1$ ist; auch dann ist die Stückkostenfunktion $k(x) = \kappa x^{s-1} + K_{\text{fix}}/x$ streng fallend auf $\mathbb{R}_{>0}$. Bei überproportional wachsenden variablen Kosten κx^s mit $s > 1$ ist dagegen diese Stückkostenfunktion Summe eines wachsenden und eines fallenden Terms und im Ganzen nicht monoton auf $\mathbb{R}_{>0}$. (Mit Differentialrechnung sieht man, dass dann $k(x)$ die Minimumstelle $x_* = (\frac{K}{(s-1)\kappa})^{1/s}$ auf $\mathbb{R}_{>0}$ hat und links davon streng fällt, rechts davon streng wächst.)

(6) Ein Beispiel für die Verkettung monotoner Funktionen ist das *Potenzieren einer monotonen Funktion.* Ist $f(x)$ streng monoton und nichtnegativ auf I, so sind alle Potenzen $f(x)^s$ mit Exponenten $s > 0$ streng monoton im gleichen Sinne wie f, während die Potenzen $f(x)^{-s}$ mit negativen Exponenten (wenn f positiv ist auf I) gegensinnig streng monoton zu f sind. Ein anderes Beispiel ist die *logistische Funktion* $\frac{a+b\mathrm{e}^{cx}}{1+\mathrm{e}^{cx}}$, die als Verkettung $h \circ g$ der im Falle $a < b$ auf $\mathbb{R}_{>-1}$ streng wachsenden gebrochen–linearen Funktion $h(y) = \frac{a+by}{1+y} = b - \frac{b-a}{1+y}$ mit der streng wachsenden Exponentialfunktion $g(x) = (\mathrm{e}^c)^x$ aufgefasst werden kann, also streng wachsend auf ganz $\mathbb{R}$ ist. ■

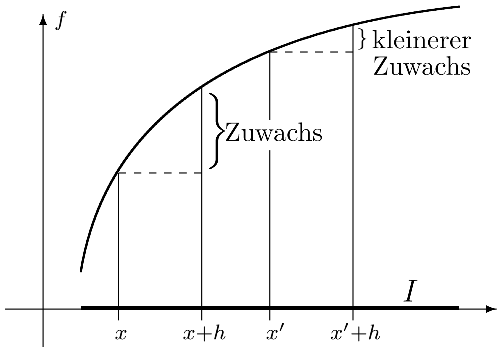
Im Falle der Monotonie einer ökonomischen Funktion steht oft nicht ihr Wachsen oder Fallen selbst im Zentrum des Interesses, sondern der *Wachstumstrend*, also die Frage ob sich das Wachstum bzw. das Abfallen verstärkt oder abschwächt. Das geht manchmal so weit, dass eine Trendumkehr schon gefeiert wird, als wäre es eine Umkehr des Monotonieverhaltens, beispielsweise wenn der Finanzminister das progressive Wachstum der Staatsverschuldung stoppen konnte (die Verschuldung aber immer noch zunimmt, nur eben langsamer). Den Wachstumstrend können wir mathematisch beschreiben, indem wir den **Zuwachs** $f(x+h)-f(x)$ betrachten, den ein Funktionswert erfährt, wenn wir das Argument x um eine feste positive *Schrittweite* h nach rechts verschieben. Dieser Zuwachs ist natürlich positiv, wenn die Funktion monoton wächst, und der Wachstumstrend wird nun einfach durch das Anwachsen oder Abnehmen der Zuwächse mit wachsendem x beschrieben. Dies führt zu einer für die Wirtschaftsmathematik sehr wichtigen Begriffsbildung:

DEFINITIONEN und DISKUSSION (***progressive und degressive Monotonie, Konvexität***):

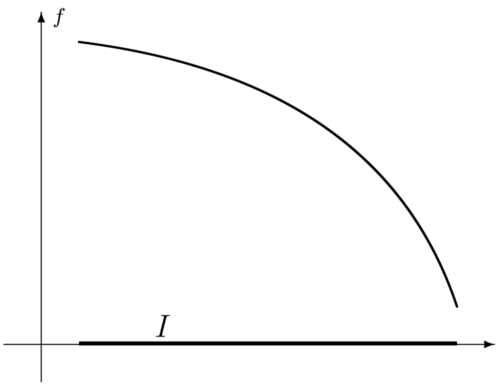
1) Wir nennen eine reelle Funktion f auf einem Intervall I **progressiv wachsend**, wenn sie wachsend ist, und wenn die Zuwächse $f(x+h)-f(x)$ für jede Schrittweite $h>0$ (kleiner als die Länge von I) ebenfalls eine wachsende Funktion von x sind (auf dem von rechts um h verkürzten Intervall $I_h := \{x \in I : x+h \in I\}$). Ist dagegen f wachsend auf I mit bei Vergrößerung von x (in I_h) abnehmenden Zuwächsen, so heißt die Funktion f **degressiv wachsend**. Entsprechend heißt eine fallende Funktion f auf I **progressiv fallend**, wenn ihre negativen "Zuwächse" $f(x+h)-f(x)$ für alle $h>0$ (und kleiner als die Länge von I) dem Betrag nach zunehmen mit wachsendem x, und f heißt **degressiv fallend**, wenn die Zuwächse bei Vergrößerung von x (in I_h) dem Betrag nach kleiner werden.



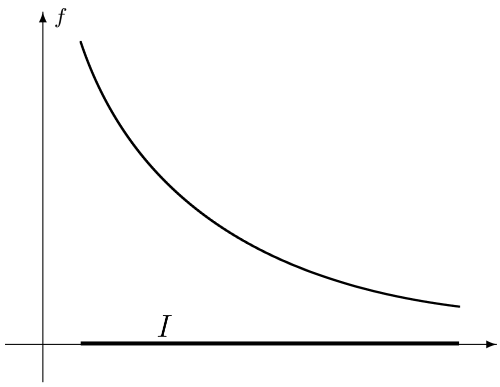
progressiv wachsende Funktion



degressiv wachsende Funktion



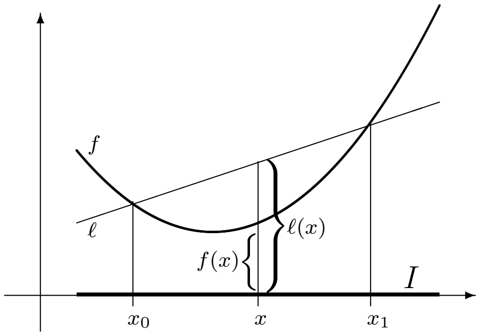
progressiv fallende Funktion



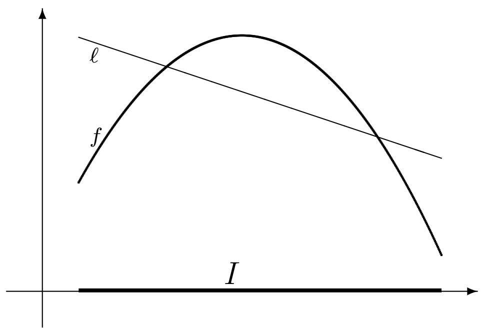
degressiv fallende Funktion

2) Natürlich muss eine wachsende Funktion nicht unbedingt auf ihrem ganzen Definitionsintervall progressiv oder degressiv wachsen, sondern das Wachstum kann z.B. auf einem Teilintervall progressiv, auf einem anderen degressiv sein. Auch kann es Teilintervalle mit *linearem Wachstum* geben, d.h. die Zuwächse $f(x+h)-f(x)$ sind unabhängig von x, wenn $x, x+h$ in einem solchen Teilintervall variieren; das Wachstum der Funktion ist dort dann weder progressiv noch degressiv, sondern konstant. Für den Graphen bedeutet es, dass er über solchen Teilintervallen auf einer Geraden verläuft. Ein Beispiel ist ein typischer *Einkommensteuertarif* $S(x)$, der für ein zu versteuerndes Einkommen von x (Euro) die darauf zu entrichtende Einkommensteuer $S(x)$ (Euro) angibt. Die Funktion $S(x)$ ist konstant Null auf einem Intervall $[0, x_0]$, wobei $x_0 > 0$ das Mindesteinkommen ist, bei dem die Steuerpflicht beginnt; daran schließt sich ein Intervall $[x_0, x_1]$ linearen Wachstums an; dann kommt ein Intervall $[x_1, x_2]$ progressiven Wachstums, genannt die *Progressionszone*; auf $[x_2, \infty[$ schließlich wächst die Funktion $S(x)$ dann wieder linear mit konstanter Steigung, die dem Spitzensteuersatz entspricht.

3) Eng verbunden mit dem Begriff der Progressivität oder Degressivität der Monotonie ist das Konzept der Konvexität einer Funktion $f : I \to \mathbb{R}$ auf einem Intervall I in $\mathbb{R}$, das auf dem Vergleich von f mit linearen Funktionen fußt. Sind $x_0 < x_1$ zwei Punkte in I, so gibt es genau eine lineare Funktion $\ell(x)$, die an den Stellen x_0 und x_1 dieselben Werte hat wie f. Wenn nun diese lineare Interpolation $\ell(x)$ der Funktionswerte zwischen beliebigen Stellen x_0, $x_1 \in I$ den wahren Funktionswert $f(x)$ immer überschätzt, wenn also $f(x) < \ell(x)$ gilt für alle x mit $x_0 < x < x_1$, so heißt f **streng konvexe Funktion** auf I. Geometrisch bedeutet dies einfach, dass *der Graph von f unterhalb seiner Sehnen liegt* (von den Sehnenendpunkten, die auf dem Graphen liegen, abgesehen). Außerhalb einer Sehne *liegt der Graph dann über der Sekante*, die durch Verlängerung der Sehne definiert wird, und da Tangenten an den Graphen Grenzlagen solcher Sekanten sind, *verläuft der Graph auch oberhalb jeder Tangente* (abgesehen vom Berührpunkt). Die Abbildung veranschaulicht den Sachverhalt. Es wird oft gesagt, dass der Graph "nach links gekrümmt" ist, wenn man ihn von links nach rechts durchläuft. Fordert man oben nur $f(x) \leq \ell(x)$ für $x_0 < x < x_1$, so heißt f **(schwach) konvexe Funktion** auf I. Geometrisch bedeutet dies, dass die Sehnen oberhalb des Graphen von f liegen oder auch auf dem Graphen.



streng konvexe Funktion



streng konkave Funktion

Analog heißt f eine **streng konkave Funktion** auf dem Intervall I, wenn für alle $x_0 < x < x_1$ in I gilt $f(x) > \ell(x)$ für die lineare Interpolationsfunktion ℓ mit denselben Werten wie f an den Stellen x_0 und x_1. Und wenn hier nur $f(x) \geq \ell(x)$ gefordert wird, so nennen wir f eine **(schwach) konkave Funktion** auf I. Geometrisch bedeutet dies, dass die Sehnen zum Graphen der Funktion unterhalb des Funktionsgraphen liegen, bei schwacher Konkavität evtl. auch auf dem Graphen, bzw. dass der Graph "nach rechts gekrümmt" ist, wenn man ihn von links nach rechts durchläuft. Lineare Funktionen (und nur diese) sind schwach konvex und zugleich schwach konkav.

Äquivalent mit (strenger) Konkavität von f ist, dass die Funktion $-f$, deren Graph aus dem von f durch Spiegelung an der horizontalen Achse entsteht, (streng) konvex ist. Daher kann die analytische Behandlung der konkaven Funktionen auf die der konvexen Funktionen zurückgeführt werden, weshalb man in der Theorie konkave Funktionen auch nur selten betrachtet.

4) Schreibt man oben den Punkt x zwischen x_0 und x_1 in der Form $(1-s)x_0 + sx_1$ mit einem Parameter $0 < s < 1$, so ergibt sich als Funktionswert der linearen Funktion mit denselben Werten wie f an den Stellen x_0 und x_1 gerade $\ell(x) = (1-s)\ell(x_0) + s\ell(x_1) = (1-s)f(x_0)+sf(x_1)$, also die entsprechende Zwischenstelle zu den beiden Funktionswerten von f in x_0 und x_1. Die Konvexität der Funktion f auf I ist daher äquivalent mit folgender **Konvexitätsungleichung**:

$$f\left((1-s)x_0 + sx_1\right) \leq (1-s)f(x_0) + sf(x_1) \qquad \text{für alle} \quad x_0, x_1 \in I \quad \text{und} \quad 0 \leq s \leq 1\,.$$

Strenge Konvexität bedeutet, dass hier sogar die strenge Ungleichung “$<$” gilt außer in den Fällen $x_0 = x_1$ oder $s = 0$ oder $s = 1$, in denen beide Seiten offensichtlich gleich sind. Bei der analytischen Behandlung von Funktionen wird ihre Konvexität oft in Form der Konvexitätsungleichung ausgenutzt. Man kann sie sich so am besten merken:

- *Der Wert der konvexen Funktion f in einem gewichteten arithmetischen Mittel zweier Stellen im Definitionsintervall ist nicht größer als das entsprechende Mittel der Funktionswerte an diesen beiden Stellen.*

Und man kann sich leicht überlegen, dass dieselbe Aussage auch für mehr als zwei Stellen richtig ist, d.h. für konvexe Funktionen f auf I gilt die sog. **Jensensche Ungleichung**:

$$f(s_1x_1 + \ldots + s_nx_n) \leq s_1f(x_1) + \ldots + s_nf(x_n)$$

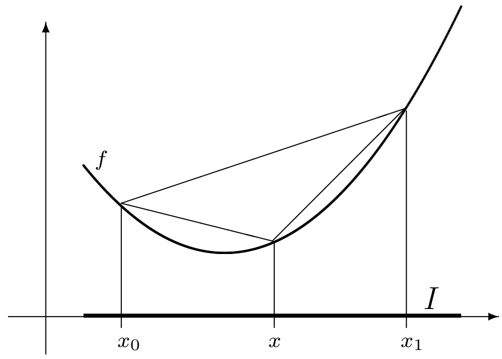
für alle $x_1, \ldots, x_n \in I$ und $s_1, \ldots, s_n \in [0,1]$ mit $s_1 + \ldots + s_n = 1$; Gleichheit tritt im Falle einer streng konvexen Funktion f und bei positiven Gewichten s_i nur dann ein, wenn alle Punkte x_i gleich sind. Für konkave Funktionen gelten natürlich dann dieselben Ungleichungen in umgekehrter Richtung.

5) Wählt man in der Konvexitätsungleichung den Parameter s außerhalb des Intervalls $[0,1]$, so dass $x = (1-s)x_0 + s_1$ ein Punkt außerhalb des Intervalls zwischen x_0 und x_1 ist, so gilt die Ungleichung in der umgekehrten Richtung, sofern x noch im Definitionsintervall I der konvexen Funktion f liegt. Dazu braucht man nur, wenn etwa $s > 1$ ist, die Rollen von x und x_1 zu vertauschen. Man hat $x_1 = \frac{1}{s}x + \frac{s-1}{s}x_0$, die Konvexitätsungleichung liefert also $f(x_1) \leq \frac{1}{s}f(x) + \frac{s-1}{s}f(x_0)$ oder $f(x) \geq (1-s)f(x_0) + sf(x_1) = \ell(x)$, wo ℓ die lineare Funktion mit denselben Werten wie f in x_0 und x_1 ist. Geometrisch bedeutet dies, wie schon in 3) oben erwähnt, dass der Graph einer konvexen Funktion außerhalb jeder Sehne oberhalb der durch Verlängerung entstehenden Sekante verläuft, sogar echt oberhalb bei einer streng konvexen Funktion.

6) Eine andere oft nützliche Beschreibung der Konvexität bezieht sich auf die **Sehnensteigungen** $[f(x_1)-f(x_0)]/(x_1-x_0)$ zu zwei Punkten $x_0 < x_1$ im Definitionsintervall der Funktion f. Die Konvexitätsungleichung $f(x) \leq (1-s)f(x_0)+sf(x_1)$ für $x = (1-s)x_0+sx_1$ mit $0 < s < 1$ kann man mit Hilfe von $s = (x - x_0)/(x_1 - x_0)$ in der äquivalenten Form

$$\frac{f(x) - f(x_0)}{x - x_0} \leq \frac{f(x_1) - f(x_0)}{x_1 - x_0} \leq \frac{f(x_1) - f(x)}{x_1 - x} \qquad \text{für} \quad x_0 < x < x_1 \;\; \text{in} \;\; I$$

formulieren. Strenge Konvexität ist gleichbedeutend damit, dass hier strikte Ungleichung "$<$" gilt. Geometrisch bedeuten diese Ungleichungen, dass die Sehnensteigungen zunehmen, wenn man einen der Sehnenendpunkte auf dem Graphen nach rechts verlagert (nämlich $(x, f(x))$ nach $(x_1, f(x_1))$ bzw. $(x_0, f(x_0))$ nach $(x, f(x))$). Natürlich gilt das dann auch, wenn man beide Sehnenendpunkte auf dem Graphen nach rechts verschiebt; denn man kann ja erst den rechten und anschließend den linken Endpunkt verlagern.



Sehnensteigungen bei konvexem f

- *Bei einer streng konvexen Funktion nehmen die Sehnensteigungen zu, wenn man einen oder beide Endpunkte einer Sehne auf dem Graphen nach rechts verlagert. Diese Eigenschaft charakterisiert strenge Konvexität.*

7) Betrachten wir speziell Sehnen, deren Endpunkte $x, x+h$ und $x', x'+h$ auf der horizontalen Achse Abschnitte derselben Länge $h > 0$ markieren, so sagt die Ungleichung für die Sehnensteigungen $\frac{1}{h}[f(x+h) - f(x)] < \frac{1}{h}[f(x'+h) - f(x')]$, wenn $x < x'$ ist, d.h. es besteht die Ungleichung $f(x+h) - f(x) < f(x'+h) - f(x')$ zwischen den Zuwächsen von f an den Stellen x und x' , wenn f streng konvex ist. Somit sind wachsende streng konvexe Funktionen progressiv wachsend und fallende streng konvexe Funktionen degressiv fallend. Es gilt auch das Umgekehrte: Ist f progressiv wachsend auf dem Intervall I und $x_0 < x_1$ in I , so ist für alle $n \in \mathbb{N}_{\geq 2}$ und $k \in \mathbb{N}_{<n}$ der mittlere Zuwachs $\frac{1}{k}[f((1-\frac{k}{n})x_0 + \frac{k}{n}x_1) - f(x_0)]$ von f auf k aufeinander folgenden Intervallen der Länge $\frac{1}{n}(x_1 - x_0)$ beginnend mit x_0 kleiner als der mittlere Zuwachs $\frac{1}{n-k}[f(x_1) - f((1-\frac{k}{n})x_0 + \frac{k}{n}x_1)]$ von f auf den $n-k$ folgenden Intervallen dieser Länge bis x_1 . Das bedeutet $f((1-\frac{k}{n})x_0 + \frac{k}{n}x_1) < \frac{n-k}{n}f(x_0) + \frac{k}{n}f(x_1)$ und ist gerade die Konvexitätsungleichung für den Zwischenpunkt $(1-s)x_0 + sx_1$ mit rationalem Parameter $s = \frac{k}{n}$. Da f wachsend ist, folgt auch für beliebige $s \in]0,1[$ dann die Konvexitätsungleichung $f((1-s)x_0 + sx_1) \leq (1-s)f(x_0) + sf(x_1)$, indem man s durch größere rationale Zahlen $\frac{k}{n}$ approximiert. Somit ist f eine konvexe Funktion auf I, und die Konvexität muss sogar streng sein, weil sonst f auf einem Teilintervall von I linear wäre, also auf diesem Intervall konstante Zuwächse hätte und nicht progressiv wachsend wäre. Damit haben wir die erste der folgenden Aussagen bewiesen, die den Zusammenhang zwischen Konvexität und Progressivität bzw. Degressivität von Wachstum herstellen. (Die anderen Aussagen folgen aus der ersten durch Übergang zu $-f(x)$ bzw. $-f(-x)$.)

- *Eine wachsende Funktion ist genau dann streng konvex, wenn sie progressiv wächst;*
- *eine fallende Funktion ist genau dann streng konvex, wenn sie degressiv fällt;*
- *Eine wachsende Funktion ist genau dann streng konkav, wenn sie degressiv wächst;*
- *eine fallende Funktion ist genau dann streng konkav, wenn sie progressiv fällt.*

Das erklärt, warum die Graphen in den Abbildungen zu 1), die zu progressiv oder degressiv monotonen Funktionen gehören, jeweils wie bei konvexen bzw. konkaven Funktionen "nach links gekrümmt" bzw. "nach rechts gekrümmt" waren: Für monoton wachsende Funktionen ist Konvexität bzw. Konkavität einfach dasselbe wie progressives bzw. degressives Wachstumsverhalten. Der Konvexitätsbegriff ist aber allgemeiner, weil er sich auch auf nichtmonotone Funktionen anwenden lässt, bei denen z.B. auf ein Intervall degressiven Abnehmens ein Intervall progressiven Wachsens folgt. Die mathematische Theorie befasst sich deshalb vor allem mit Konvexität.

8) Es gibt einige *Operationen, die Konvexität erhalten.* Aus der Charakterisierung der konvexen Funktionen durch die Konvexitätsungleichung ist klar, dass mit f und g auch Vielfache rf mit Faktor $r \in \mathbb{R}_{>0}$ und die Summe wieder konvex auf I sind, und zwar streng, wenn f sogar streng konvex ist. Entsprechendes gilt für konkave Funktionen. Über die Differenz zweier konvexer Funktionen bzw. (äquivalent) die Summe einer konvexen und einer konkaven Funktion kann man keine generell gültige Konvexitätsaussage machen.

Ebensowenig lässt sich über die Konvexität eines Produkts fg von zwei konvexen Funktionen allgemein etwas sagen, selbst wenn die Faktoren positive Funktionen sind. Das zeigen z.B. die auf $\mathbb{R}_{>0}$ konvexen Funktionen (siehe die Beispiele unten) $f(x) := \frac{1}{x}$ und $g(x) := x^s$ mit $s > 1$, wofür das Produkt x^{s-1} streng konvex ist für $s > 2$, linear und damit nicht streng konvex oder konkav für $s = 1$ und streng konkav für $1 < s < 2$. Konkavität bleibt bei Produktbildung auch nicht erhalten. Das zeigen z.B. die auf $\mathbb{R}_{\geq 0}$ streng konkaven Wurzelfunktionen x^s mit $0 < s < 1$ (siehe die Beispiele unten), deren Quadrat x^{2s} für $0 < s < \frac{1}{2}$ streng konkav ist, für $\frac{1}{2} < s < 1$ aber streng konvex. Allerdings gilt Folgendes:

- *Das Produkt fg von zwei konvexen, gleichsinnig monotonen und nichtnegativen Funktionen f und g auf einem Intervall I ist eine konvexe, im gleichen Sinn monotone, nichtnegative Funktion auf I und sogar streng monoton und streng konvex, wenn f und g streng monoton sind.*
- *Die reziproke Funktion $\frac{1}{g}$ zu einer positiven, konkaven Funktion g auf I ist konvex auf I und sogar streng konvex, wenn g auf keinem Teilintervall konstant ist.*
- *die Verkettung $h \circ f$ einer nichtfallenden konvexen Funktion h auf einem Intervall J und einer konvexen Funktion f auf einem Intervall I mit $f(I) \subset J$ ist eine konvexe Funktion auf I und streng konvex, wenn f streng konvex und h streng monoton ist.*
- *Die Umkehrfunktion f^{-1} zu einer progressiv wachsenden stetigen Funktion f auf einem Intervall ist degressiv wachsend, die zu einer degressiv wachsenden ist progressiv wachsend; die Umkehrfunktion zu einer progressiv (bzw. degressiv) fallenden stetigen Funktion auf einem Intervall ist wieder progressiv (bzw. degressiv) fallend.*

Zum Beweis der ersten Aussage rechnen wir für $0<s<1$ die Konvexitätsungleichung nach:

$$\begin{aligned} & f((1-s)x_0+sx_1)g((1-s)x_0+sx_1)-(1-s)f(x_0)g(x_0)-sf(x_1)g(x_1) \\ \leq\ & [(1-s)f(x_0)+sf(x_1)][(1-s)g(x_0)+sg(x_1)]-(1-s)f(x_0)g(x_0)-sf(x_1)g(x_1) \\ =\ & -s(1-s)[f(x_1)-f(x_0)][g(x_1)-g(x_0)] \ \leq\ 0\,, \end{aligned}$$

wobei wir bei der letzten Ungleichung die gleichsinnige Monotonie von f und g, also die gleichen Vorzeichen der Klammerausdrücke $[\ldots]$, ausgenutzt haben und in der ersten Ungleichung die Konvexität und Nichtnegativität. Sind f und g streng wachsend, so gilt die letzte Ungleichung sogar streng im Falle $x_0 < x_1$. Zur zweiten Aussage bemerken wir

$$\frac{1}{g((1-s)x_0+sx_1)} \ \leq\ \frac{1}{(1-s)g(x_0)+sg(x_1)} \ <\ (1-s)\frac{1}{x_0}+s\frac{1}{x_1}\,,$$

wobei die erste Ungleichung aus der Konkavitätsungleichung für g folgt und die zweite die Ungleichung zwischen dem gewichteten harmonischen und arithmetischen Mittel der Zahlen $1/g(x_0)$, $1/g(x_1)$ ist (siehe 1.5). Gleichheit in beiden Ungleichungen kann nur eintreten, wenn $g(x_0) = g(x_1)$ ist und in der Konkavitätsungleichung Gleichheit eintritt, was Konstanz von g auf $[x_0, x_1]$ bedeutet. Die dritte Aussage sieht man einfach so:

$$h\big(f((1-s)x_0+sx_1)\big) \leq h\big((1-s)f(x_0)+sf(x_1)\big) \leq (1-s)h(f(x_0))+sh(f(x_1))\,,$$

wobei in der ersten Ungleichung die Konvexität von f und monotones Wachstum von h ausgenutzt wurde und in der zweiten die Konvexität von h. Ist f streng konvex und h streng wachsend, so gilt die erste Ungleichung sogar strikt im Falle $x_0 \neq x_1$. Zum Beweis der letzten Aussage schließlich wendet man auf beide Seiten der Konvexitätsungleichung $f((1-t)x_0 + tx_1) < (1-t)y_0 + ty_1$ für $y_0 = f(x_0)$ und $y_1 = f(x_1)$ die Umkehrfunktion an und beachtet $x_0 = f^{-1}(y_0)$ und $x_1 = f^{-1}(y_1)$ (entsprechend bei der Konkavitätsungleichung; wenn f fallend ist, so auch f^{-1}, daher kehrt Anwendung von f^{-1} die Ungleichung dann um). Simple Beispiele zeigen, dass man die Konvexität der jeweils gebildeten Funktion nicht mehr schließen kann, wenn man auf die Monotonie–Voraussetzungen und Positivitätsvoraussetzungen an die Funktionen verzichtet.

Einfach zu sehen ist schließlich, dass die Verkettung $f(ax+b)$ einer beliebigen (streng) konvexen Funktion f mit einer nichtkonstanten linearen Funktion $ax+b$ wieder (streng) konvex ist. (Entsprechend für Konkavität.) ■

Der Nachweis der Progressivität oder Degressivität des Wachstums oder des Fallens einer vorgelegten elementaren Funktion f erfolgt über die Monotonieuntersuchung der Zuwachsfunktionen $z_h(x) := f(x+h) - f(x)$ für jede Schrittweite $h > 0$ (kleiner als die Länge des Grundintervalls, auf dem man f betrachtet). Dafür kann man alle Methoden einsetzen, die man für Monotonieuntersuchungen kennt. Oder man erbringt den Nachweis der Konvexität der vorgelegten Funktion durch Nachrechnen der Konvexitätsungleichung, was aber mühsam sein kann oder gar nicht direkt gelingt (sofern die Funktion überhaupt konvex ist). Eine meistens bessere und einfachere Methode zum Konvexitätsnachweis, die auf der Untersuchung des Monotonieverhaltens der Ableitung f' beruht, werden wir später in der Differentialrechnung kennen lernen. Wir beschränken uns daher auf wenige

BEISPIELE (***von konvexen Funktionen***):

(1) *Potenzfunktionen* x^n *mit natürlichem Exponenten* $n \in \mathbb{N}_{\geq 2}$ sind streng konvex auf $\mathbb{R}_{\geq 0}$. Das sieht man aus Bemerkung 8) oben durch Bildung der Potenzen der streng wachsenden linearen, also (schwach) konvexen Funktion x auf $\mathbb{R}_{\geq 0}$. Wegen $(-x)^n = (-1)^n x^n$ folgt daraus strenge Konkavität auf $\mathbb{R}_{\leq 0}$, wenn n ungerade ist. In diesem Fall ist x^n auf ganz $\mathbb{R}$ natürlich weder konvex noch konkav. Für gerade $n = 2m \geq 2$ ist aber x^n sogar auf ganz $\mathbb{R}$ streng konvex. Dazu rechnet man mit Hilfe der "*binomischen Ungleichung*" $2x_0x_1 < x_0^2 + x_1^2$ für $x_0 \neq x_1$ (die Differenz beider Seiten der Ungleichung ist ja $(x_1 - x_0)^2 > 0$) zunächst für die Funktion $f(x) = x^2$ die Konvexitätsungleichung nach für $0 < s < 1$:

$$\begin{aligned} ((1-s)x_0 + sx_1)^2 &= (1-s)^2x_0^2 + (1-s)s2x_0x_1 + s^2x_1^2 \\ &< (1-s)^2x_0^2 + (1-s)s(x_0^2 + x_1^2) + s^2x_1^2 = (1-s)x_0^2 + sx_1^2\,. \end{aligned}$$

(Man kann auch einfach die Ungleichung zwischen dem gewichteten arithmetischen und quadratischen Mittel von zwei Zahlen x_0, x_1 quadrieren; siehe 1.5.) Sodann fasst man x^{2m} auf als Verkettung der nichtfallenden konvexen Funktion $h(y) = y^m$ auf $\mathbb{R}_{\geq 0}$ mit der Funktion $f(x) = x^2$ auf und erhält die Behauptung wieder aus Bemerkung 8) oben. Aus der dort angegebenen Aussage über die Konvexität der reziproken Funktion ergibt sich für negative Exponenten $-n \in \mathbb{Z}_{<0}$ dann die strenge Konvexität von $x^{-n} = \frac{1}{x^n}$ auf $\mathbb{R}_{>0}$. Wegen $(-x)^n = (-1)^n x^{-n}$ folgt auch hier strenge Konkavität auf $\mathbb{R}_{<0}$ für ungerade n und strenge Konvexität auf $\mathbb{R}_{<0}$ für gerade n. Ein Blick auf die in 4.1 abgebildeten Graphen der Potenzfunktionen macht deren Konvexitätsverhalten klar.

2) Für die *allgemeinen Potenzfunktionen* x^s mit reellem Exponenten s, die für mathematische Modelle ökonomischer Vorgänge sehr wichtig sind, gilt Folgendes:

> x^s *ist auf* $\mathbb{R}_{\geq 0}$ *progressiv wachsend für* $s>1$, *und zwar um so stärker, je größer* s *ist;*
>
> x^s *ist auf* $\mathbb{R}_{\geq 0}$ *degressiv wachsend für* $0 < s < 1$ (Wurzelfunktionen), *und zwar um so langsamer, je kleiner* s *ist;*
>
> x^s *ist degressiv fallend auf* $\mathbb{R}_{>0}$ *für* $s < 0$ (reziproke Potenz- und Wurzelfunktionen), *und zwar um so schneller, je größer* $|s|$ *ist.*

Hier haben wir für $s > 0$ die Stelle $x = 0$ durch die Festsetzung $0^s := 0$, was der Grenzwert von x^s bei $x \downarrow 0$ ist, mit in den Definitionsbereich einbezogen. Die Progressivität des Wachstums für $s > 1$, also die strenge Konvexität der Potenzfunktion, erhält man durch Potenzieren der Ungleichung zwischen dem mit t und $1 - t$ gewichteten arithmetischen Mittel und dem entsprechenden gewichteten Potenzmittelwert zum Exponenten s (siehe 1.5), $(1 - t)x_0 + tx_1 < [(1 - t)x_0^s + tx_1^s]^{1/s}$ für $0 \leq x_0 < x_1$ und $0 < t < 1$. Für $0 \neq s < 1$ kann man dieselbe Ungleichung heranziehen, die dann in der entgegengesetzten Richtung gilt, um Degressivität des Wachstums für $0 < s < 1$ bzw. des Fallens für $s < 0$ zu sehen (beachte, dass sich die Ungleichung beim Potenzieren mit Exponent $s < 0$ umkehrt). Oder man bemerkt, dass die Wurzelfunktionen $x^{1/s}$ die Umkehrfunktionen zu den Potenzfunktionen x^s mit $s > 1$ sind (wegen $(x^s)^{1/s} = x$ für $x \geq 0$) und verwendet die in Bemerkung 8) oben gemachte Aussage über das degressive Wachstum der Umkehrfunktion zu einer progressiv wachsenden Funktion, sowie über die Konvexität der Reziproken von positiven konkaven Funktionen. Für die Exponenten $s = 0$ und $s = 1$ ist die Potenzfunktion konstant bzw. linear auf $\mathbb{R}_{>0}$ (sogar auf $\mathbb{R}_{\geq 0}$, wenn man $0^0 := 1$ setzt) und folglich schwach konvex und konkav auf $\mathbb{R}_{\geq 0}$, aber auf keinem Teilintervall strikt konvex oder konkav. Die Abbildungen der Graphen von allgemeinen Potenzfunktionen in 4.1 machen deren Konvexitätsverhalten für die verschiedenen Exponenten $s \in \mathbb{R}$ anschaulich klar.

(3) Allgemeine *Exponentialfunktionen* a^x sind (sehr stark) progressiv wachsend auf $\mathbb{R}$ für Basen $a>1$ und (sehr schnell gegen Null) degressiv fallend für $0<a<1$, also in beiden Fällen streng konvex auf $\mathbb{R}$. Die Konvexitätsungleichung ist hier äquivalent mit der Ungleichung zwischen dem gewichteten geometrischen und arithmetischen Mittel (s. 1.5):

$$a^{(1-s)x_0+sx_1} = (a^{x_0})^{1-s}(a^{x_1})^s \leq (1-s)a^{x_0} + sa^{x_1} \qquad (0 < s < 1)$$

mit Gleichheit nur im Falle $a^{x_0} = a^{x_1}$, also $x_0 = x_1$.

Die *Logarithmusfunktion* $\log_a$ ist im Fall $a > 1$ als Umkehrfunktion der progressiv wachsenden Funktion a^x degressiv wachsend auf $\mathbb{R}_{>0}$, also streng konkav, im Fall $0 < a < 1$ aber als Umkehrfunktion zu einer degressiv fallenden Funktion (oder wegen $\log_a = -\log_{1/a}$) wieder degressiv fallend, also streng konvex. Die Graphen der Exponentialfunktionen und Logarithmusfunktionen sind in 4.1 zu finden.

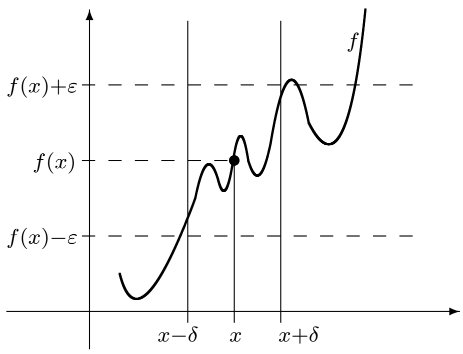
(4) Die "Investitionsfunktion" $f(x) = \frac{\mathrm{e}^{-x}}{1-\mathrm{e}^{-x}} = \frac{1}{1-\mathrm{e}^{-x}} - 1$ ist streng konvex auf $\mathbb{R}_{>0}$, weil $1 - \mathrm{e}^{-x}$ streng konkav und positiv ist auf $\mathbb{R}_{>0}$, die reziproke Funktion also streng konvex. Für die auf $\mathbb{R}$ streng wachsende logistische Funktion $g(x) = \frac{\mathrm{e}^x}{1+\mathrm{e}^x}$ gibt es kein direktes Argument dieser Art. Aber die Zuwächse $g(x+h) - g(x) = \frac{\mathrm{e}^{x+h}}{1+\mathrm{e}^{x+h}} - \frac{\mathrm{e}^x}{1+\mathrm{e}^x} = \frac{\mathrm{e}^h-1}{1+\mathrm{e}^h+\mathrm{e}^{-x}+\mathrm{e}^{x+h}}$ zu Schrittweiten $h > 0$ sind streng fallend bzgl. $x \geq 0$; denn $\mathrm{e}^{x+h} + \mathrm{e}^{-x} = (\mathrm{e}^h-1)\mathrm{e}^x + \mathrm{e}^x + \frac{1}{\mathrm{e}^x}$ ist streng wachsend in $x \geq 0$ (da $y+\frac{1}{y}$ streng wächst auf $\mathbb{R}_{\geq 1}$; siehe Beispiel (4) zu monotonen Funktionen). Somit ist g auf $\mathbb{R}_{\geq 0}$ degressiv wachsend, also konkav, und wegen $g(-x) = 1 - g(x)$ auf $\mathbb{R}_{\leq 0}$ progressiv wachsend, also konvex. Dasselbe folgt dann leicht auch für die allgemeinere logistische Funktion $\frac{a+b\mathrm{e}^{cx}}{1+\mathrm{e}^{cx}}$ mit $a < b$ und $c > 0$. ■

Als weitere grundlegende Begriffsbildung bei reellen Funktionen besprechen wir noch das Konzept der Stetigkeit. Dies ist eine mathematisch recht anspruchsvolle Begriffsbildung, die eng mit Grenzwerten zusammenhängt. Wir wollen aber Grenzwerte nicht einführen und behandeln, weil sie für die Mathematik in der Wirtschaftswissenschaft von untergeordneter Bedeutung sind. Daher begnügen wir uns hier damit, die Stetigkeit anschaulich zu erklären, ohne unnötige mathematische Strenge.

Intuitiv gesehen ist eine Funktion stetig, bei der sich die Werte $f(x)$ wenig ändern, wenn man das Argument x wenig ändert. Anders gesagt: Kleine Änderungen der unabhängigen Variablen x dürfen nicht große Änderungen der abhängigen Variablen $f(x)$ hervorrufen. Dies ist eine Eigenschaft, die man im Grunde fast immer implizit voraussetzt, wenn man Naturvorgänge — im weitesten Sinne, also unter Einschluß ökonomischer Vorgänge — durch Funktionen mathematisch modelliert. Man kennt ja in der Realität die Werte von Variablen nie genau, sondern nur bis auf (kleinere) Mess- oder Rundungsfehler bzw. in der Ökonomie oft nur bis auf (größere) Schätzfehler. Dann ist es aber völlig sinnlos, eine davon abhängige Variable $f(x)$ zu betrachten, deren Wert sich erheblich verändert, wenn man z.B. x nur durch einen anderen Mess- oder Schätzwert ersetzt oder durch einen auf- oder abgerundeten Wert. Daher ist für die mathematische Modellierung realer Vorgänge die Stetigkeit der Modellfunktionen oft eine selbstverständliche (meist gar nicht ausgesprochene) Grundanforderung. Die Schwierigkeit des Stetigkeitsbegriffs besteht darin, dass er rein qualitativ ist. (Was genau sind “kleine” bzw. “große” Änderungen der Variablen?)

DEFINITIONEN und DISKUSSION(**Stetigkeit und Unstetigkeitsstellen**):

1) Eine Funktion $f : \mathbb{R} \supset D \to \mathbb{R}$ heißt **stetig an der Stelle x** $\in D$, wenn die Abweichung $|f(x') - f(x)|$ der Funktionswerte an hinreichend nahe bei x gelegenen Stellen $x' \in D$ vom Wert $f(x)$ beliebig klein ist. Genau heißt dies, dass man eine Fehlerschranke $\varepsilon > 0$ beliebig klein vorgeben und dazu einen Abstand $\delta > 0$ bestimmen kann, derart dass $|f(x') - f(x)| < \varepsilon$ ausfällt für alle Punkte $x' \in D$ mit Abstand $|x' - x| < \delta$ zu x. (Oft wird gesagt, die Funktion sei stetig an der Stelle $x \in D$, wenn der Funktionswert $f(x)$ gleich dem Grenzwert von $f(x')$ beim Grenzübergang mit $x' \in D$ gegen x ist. Diese Definition der Stetigkeit erfordert jedoch den Grenzwertbegriff und verschleiert die hinter dem Stetigkeitskonzept stehende anschauliche Vorstellung nur.) Man nennt f eine (schlechthin) **stetige Funktion**, bzw. genauer eine stetige Funktion auf D, wenn sie an jeder Stelle ihres Definitionsbereichs D stetig ist.



Stetigkeit an der Stelle x

Stetigkeit hängt nicht nur von derAbbildungsvorschrift ab, sondern auch von der Festlegung des Definitionsbereichs; bei Verkleinerung des Definitionsbereichs durch Entfernung der Unstetigkeitsstellen kann eine unstetige Funktion stetig werden. Dabei heißt natürlich $x \in D$ eine **Unstetigkeitsstelle** der Funktion f auf D, wenn sie an dieser Stelle eben nicht stetig ist. Über die Arten der möglichen Unstetigkeitsstellen sagen wir mehr in 3) und 4) unten.

2) In 1) wird nichts darüber gesagt, wie klein δ im Verhältnis zu ε gewählt werden muss — unter Umständen sehr viel kleiner. Bei vielen Funktionen, insbesondere bei den durch einen "Rechenterm" gegebenen elementaren Funktionen (siehe 4.1), kann man jedoch eine sog. **Dehnungsschranke** $L > 0$ bestimmen (auch *Lipschitz-Konstante* genannt), d.h. es gilt $|f(x') - f(x)| \leq L \cdot |x' - x|$ für alle $x, x' \in D$ (oder wenigstens für alle x' aus D in einem Intervall $]x-\delta, x+\delta[\subset D$ mit $\delta > 0$). Mit Worten: Der Abstand der Funktionswerte ist höchstens um den Faktor L "gedehnt" im Vergleich zum Abstand der Argumente; im Fall $0 < L < 1$ ist der Abstand der Funktionswerte sogar um den Faktor L "gestaucht". Man nennt dann die Funktion **dehnungsbeschränkt** mit Dehnungsschranke L. In dieser Situation kann man zu gegebener Fehlerschranke ε wie in 1) einfach $\delta := \frac{1}{L}\varepsilon$ wählen, dann ist $|f(x') - f(x)| < \varepsilon$ für $|x' - x| < \delta$ erfüllt. Also gilt:

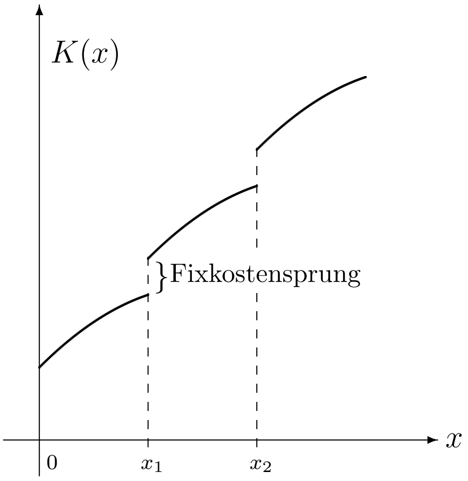
- *Jede dehnungsbeschränkte Funktion ist stetig.*

Für $f(x) = x$ gilt zum Beispiel $|f(x') - f(x) = |x' - x|$, also ist $L = 1$ eine Dehnungsschranke auf $\mathbb{R}$. Für die elementaren Grundfunktionen e^x, $\ln x$ kann man sich Dehnungsschranken (auf kompakten Teilintervallen ihres maximalen Definitionsintervalls) mit Hilfe der fundamentalen Ungleichungen in 1.4 überlegen. Für Exponenten $n \in \mathbb{N}_{\geq 2}$ ist die Potenzfunktion x^n zwar nicht dehnungsbeschränkt auf $\mathbb{R}$, aber aus der geometrischen Summenformel $x'^n - x^n = (x' - x)(x^{n-1} + x^{n-2}x' + \ldots + x'^{n-1})$ liest man ab, dass die Funktion auf einem kompakten Intervall $[-b, b]$ mit $b > 0$ die Dehnungsschranke $L = nb^{n-1}$ besitzt (weil der zweite Klammerfaktor für $|x| \leq b$ und $|x'| \leq b$ höchstens diesen Betrag haben kann). Daraus folgt weiter, dass Polynomfunktionen auf kompakten Intervallen dehnungsbeschränkt sind und rationale Funktionen auf kompakten Intervallen ohne Polstelle ebenfalls. Leicht zu sehen ist auch, dass die Verkettung von dehnungsbeschränkten Funktionen wieder dehnungsbeschränkt ist. Das Endresultat derartiger Überlegungen ist:

- *Jede elementare Funktion ist stetig und auf kompakten Teilintervallen ihres Definitionsbereichs sogar dehnungsbeschränkt.*

Das bedeutet, dass alle in mathematischen Modellen ökonomischer Vorgänge auftretenden Funktionen stetig sind, soweit sie nicht offensichtliche Unstetigkeitsstellen haben.

3) Was Stetigkeit bedeutet, kann man auch gut verstehen, wenn man Unstetigkeitsstellen studiert. Eine **hebbare Unstetigkeitsstelle** x_0 von f liegt vor, wenn man sie einfach durch geeignete andere Festsetzung des Funktionswertes $f(x_0)$ "beheben", also zu einer Stetigkeitsstelle machen kann. Bei den Funktionen der Wirtschaftsmathematik gibt es nur zwei einfache Arten von "echten", d.h. nicht hebbaren, Unstetigkeitsstellen. (In der Mathematik gibt es noch kompliziertere). Bei einer sog. **Sprungstelle** x_0 von f hat die Funktion zwei verschiedene einseitige Grenzwerte c und d, wenn man sich der Stelle x von links (von unten) bzw. von rechts (von oben) nähert. Bei Überschreiten der Stelle x_0 mit der unabhängigen Variablen x machen die Funktionswerte einen "Sprung" der Größe $d - c$. Genau heißt dies, dass für jede gegebene Fehlerschranke $\varepsilon > 0$ ein $\delta > 0$ existiert, derart dass die Funktion auf $]x_0 - \delta, x_0[$ definiert ist und dort Werte hat, die sich um weniger als ε von der Zahl c unterscheiden, und dass sie auf $]x_0, x_0 + \delta[$ definiert ist mit Werten im Abstand kleiner als ε von der Zahl d. Die einseitigen Grenzwerte c bzw. d werden oft symbolisch $f(x_0-)$ bzw. $f(x_0+)$ notiert, um anzudeuten, dass es sich um Grenzwerte von $f(x)$ mit x beliebig wenig kleiner bzw. größer als x_0 handelt. In einer Sprungstelle x_0 ist die Funktion f offenbar unstetig, und diese Unstetigkeit ist nicht hebbar, kann also nicht durch passende Abänderung des Funktionswertes $f(x_0)$ beseitigt werden. (Der Funktionswert in der Sprungstelle ist meist irrelevant; er braucht gar nicht definiert sein.)

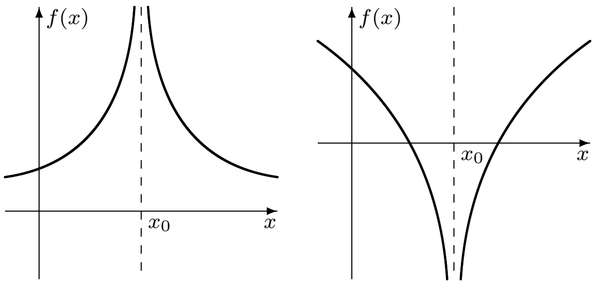


Sprungstellen der Kostenfunktion

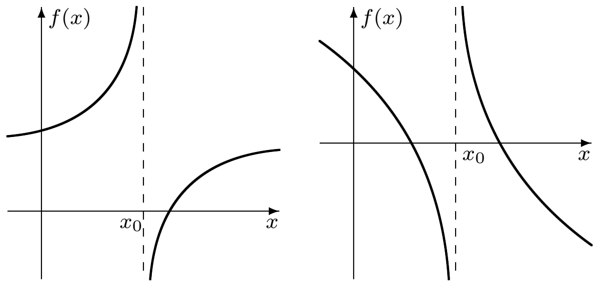
Eine realistische ökonomische Situation, die durch Funktionen mit Sprungstellen modelliert wird, liegt vor bei einer Kostenfunktion, die eine *Produktion mit quantitativer Anpassung* beschreibt (s. 4.1). Hier wird jeweils neues Produktionsaggregat zugeschaltet, wenn der Produktionsoutput bestimmte Schwellen x_i überschreitet (bei denen die laufenden Aggregate maximal ausgelastet sind). Da mit der Inbetriebnahme des neuen Aggregats positive Fixkosten entstehen, auch wenn dort noch gar nichts produziert wird, macht die Kostenfunktion $K(x)$ dann an der Stelle x_i einen Sprung nach oben in Höhe der Fixkosten des neu zugeschalteten Aggregats.

Man sieht hier, wie Sprungstellen durch *Aneinandersetzen von stetigen Funktionen* entstehen können. Wenn man eine stetige Funktion f_1 auf einem Intervall $[x_0, x_1]$ hat und eine weitere stetige Funktion f_2 auf dem rechts angrenzenden Intervall $[x_1, x_2]$, so ist die zusammengesetzte Funktion f auf dem Vereinigungsintervall $[x_0, x_2]$ definiert durch $f(x) := f_1(x)$ für $x_0 \le x < x_1$ und $f(x) := f_2(x)$ für $x_1 < x \le x_2$, wobei der Wert $f(x_1)$ z.B. als $f_1(x_1)$ oder als $f_2(x_1)$ oder als Mittelwert $\frac{1}{2}f_1(x_1) + \frac{1}{2}f_2(x_1)$ festgelegt wird. Wenn die beiden Funktionen an der Stelle x_1 "zusammenpassen", d.h. wenn $f_1(x_1) = f_2(x_1)$ ist, so ist auch die so zusammengesetzte Funktion stetig; wenn aber $f_1(x_1) \neq f_2(x_2)$ ist, so hat die zusammengesetzte Funktion an der Stelle x_1 eine Sprungstelle. Auf diese Weise ist auch die obige Kostenfunktion aus (mehreren) stetigen Funktionen zusammengesetzt, und man nennt allgemein Funktionen, die durch Aneinandersetzen stetiger Funktionen entstehen und höchstens Sprungstellen als Unstetigkeitsstellen haben, **stückweise stetig**. Die in 4.1 eingeführeten stückweise elementaren Funktionen sind also stückweise stetig.

4) Eine andere Art von Unstetigkeitsstelle liegt vor bei einer sog. **Unendlichkeitsstelle** x_0 einer Funktion f. Dies bedeutet, dass die Grenzwerte von beiden Seiten unendlich sind, und zwar unterscheidet man die **Unendlichkeitsstellen ohne Vorzeichenwechsel** von den **Unendlichkeitsstellen mit Vorzeichenwechsel**; bei ersteren sind beide einseitigen Grenzwerte $+\infty$ oder beide gleich $-\infty$, bei letzteren sind die beiden Grenzwerte entgegengesetzt unendlich. Dass $f(x)$ z.B. bei Annäherung von x an x_0 von rechts (von oben) den Grenzwert ∞ hat, bedeutet dabei genau, dass jede gegebene Zahl S von allen Funktionswerten $f(x)$ übertroffen wird, die an Stellen x in einem hinreichend kurzen Intervall $]x_0, x_0+\delta[$ mit linkem Randpunkt x_0 angenommen werden. Symbolisch schreibt man dafür $\lim_{x\downarrow x_0} f(x) = \infty$, und entsprechend erklärt man den Grenzwert $\lim_{x\downarrow x_0} f(x) = -\infty$ sowie die Grenzwerte $\lim_{x\uparrow x_0} f(x) = \infty$ bzw. $\lim_{x\uparrow x_0} f(x) = -\infty$ bei Annäherung an x_0 von links (von unten). Typische Unendlichkeitsstellen mit Vorzeichenwechsel sind Pole ungerader Ordnung bei einer rationalen Funktion wie $x_0 = 0$ bei $f(x) = 1/x$; Pole gerader Ordnung sind dagegen Unendlichkeitsstellen ohne Vorzeichenwechsel wie $x_0 = 0$ bei $f(x) = 1/x^2$. Bei diesen Funktionen ist zwar der Wert an der Polstelle nicht definiert, aber es ist klar, dass bei einer Unendlichkeitsstelle x_0 von f keine Festsetzung des Wertes $f(x_0)$ möglich ist, welche die Funktion an der Stelle x_0 stetig machen würde, dass also eine derartige Unstetigkeit nicht hebar ist. Man verlangt daher bei Unendlichkeitsstellen nicht, dass dort die Funktion überhaupt definiert ist. Die folgenden Abbildungen veranschaulichen die verschiedenen Möglichkeiten einer Unendlichkeitsstelle:



Unendlichkeitsstelle ohne Vorzeichenwechsel



Unendlichkeitsstelle mit Vorzeichenwechsel

In der Ökonomie treten Unendlichkeitsstellen selten auf, weil es unrealistisch ist, dass Variablen unendliche Grenzwerte haben. Allerdings kommt es vor, dass eine ökonomisch sinnvolle Funktion eine *einseitige Unendlichkeitsstelle* besitzt. Dies trifft z.B. auf eine Stückkostenfunktion $k(x) = K(x)/x$ an der Stelle $x > 0$ zu, wenn die zugehörige Kostenfunktion $K(x)$ für $x > 0$ von unten durch eine positive Fixkostenkonstante K_{fix} beschränkt ist. Es gilt dann $k(x) \geq K_{\text{fix}}/x$ und $k(x)$ wird daher beliebig groß, wenn $x > 0$ klein genug ist, d.h. es gilt $\lim_{x \downarrow 0} k(x) = \infty$. Da die Stückkostenfunktion nur auf $\mathbb{R}_{>0}$ definiert ist, liegt hier bei $x = 0$ eine sog. **rechtsseitige Unendlichkeitsstelle** der Funktion $k(x)$ vor.

5) Es gibt noch andere Arten von Unstetigkeitsstellen als die oben erklärten Sprungstellen und Unendlichkeitsstellen, jedoch sind diese für die Ökonomie nicht relevant. Z.B. ist denkbar, dass eine Funktion $f(x)$ an einer Stelle x_0 von links einen endlichen, von rechts aber einen unendlichen Grenzwert besitzt; dann hat man eine sog. *Sprung–Unendlichkeitsstelle*. Oder einer der beiden einseitigen Grenzwerte existiert überhaupt nicht, weil die Funktion bei entsprechender Annäherung an x_0 ein oszillierendes Verhalten zeigt und beliebig nahe bei x_0 schon auf einer Seite Werte annimmt, die weit auseinander liegen; dann liegt eine *Oszillationsstelle* vor, und diese kann kombiniert sein mit einem endlichen oder unendlichen Grenzwert oder ebenfalls mit oszillierendem Verhalten bei Annäherung an x_0 von der anderen Seite. Damit sind dann in der Tat alle möglichen Typen von Unstetigkeiten erfasst, wie in der Mathematik bewiesen wird. ■

Wir geben nun zwei wichtige Sätze über stetige Funktionen an, die der Grund sind, warum wir Stetigkeit überhaupt hier diskutiert haben. Der erste Satz garantiert uns die Existenz der Lösungen von Gleichungen der Form $f(x) = c$ mit einer stetigen Funktion f, der zweite die Existenz von Maximum- und Minimumstellen einer stetigen Funktion f unter gewissen Voraussetzungen. Die ökonomische Relevanz dieser Sätze ist klar: In vielen Situationen möchte man eine unabhängige Variable x so einstellen, dass eine davon abhängige Variable $f(x)$ einen bestimmten Wert c oder einen optimalen Wert (Maximum oder Miniumum) annimmt. Damit man an die Berechnung der Lösung(en) einer solchen Aufgabe herangehen kann, muss man aber erst einmal sicherstellen, dass eine Lösung überhaupt existiert — und das garantieren die beiden Sätze, wenn sie anwendbar sind.

SATZ (Zwischenwertsatz): *Ist f eine stetige reelle Funktion auf dem Intervall $I \subset \mathbb{R}$ und ist c eine Zahl zwischen zwei Werten $f(a)$ und $f(b)$ von f an Stellen $a < b$ in I, so hat f auch den Funktionswert c an mindestens einer Stelle zwischen a und b.*

SATZ (vom Maximum und Minimum; Extremstellensatz): *Ist f eine stetige reelle Funktion auf einem kompakten Intervall $[a, b]$ in $\mathbb{R}$, so hat f mindestens eine (absolute) Maximumstelle und mindestens eine (absolute) Minimumstelle in diesem Intervall.*

Wir erinnern daran, dass eine **(absolute) Maximumstelle** von f auf $[a,b]$ eine Stelle $x^* \in [a,b]$ mit dem größten Funktionswert ist, also mit $f(x^*) \geq f(x)$ für alle $x \in [a,b]$. Äquivalent ist, dass der Wert $f(x^*)$ gleich der kleinsten oberen Schranke für f auf $[a,b]$ ist, also gleich dem Supremum $\sup_{[a,b]} f$. Entsprechendes gilt für eine **(absolute) Minimumstelle** und das Infimum $\inf_{[a,b]} f$. **Extremstellen** sind Maximum- oder Minimumstellen.

Der *Beweis* für beide Sätze ist nicht lang und soll deshalb kurz angedeutet werden: Man konstruiert mit dem sog. Intervallhalbierungsverfahren eine Folge von Intervallen $[a_n, b_n]$ in I, indem man mit $[a_0, b_0] = [a,b]$ beginnt und als jeweils nächstes Intervall $[a_{n+1}, b_{n+1}]$ eine der beiden Hälften $[a_n, \frac{1}{2}a_n + \frac{1}{2}b_n]$ und $[\frac{1}{2}a_n + \frac{1}{2}b_n, b_n]$ von $[a_n, b_n]$ so auswählt, dass c ein Wert (schwach) zwischen $f(a_{n+1})$ und $f(b_{n+1})$ ist im Fall des Zwischenwertsatzes bzw. dass f auf $[a_{n+1}, b_{n+1}]$ keine kleinere obere Schranke hat als die obere Grenze $\sup_{[a,b]} f$ zu f auf $[a,b]$. Dann bilden die Intervalle $[a_n, b_n]$ eine Intervallschachtelung, und nach einer fundamentalen Eigenschaft des Bereichs der reellen Zahlen existiert eine Zahl x^*, die in all diesen Intervallen liegt. Ist $\varepsilon > 0$ gegeben, so gilt wegen der Stetigkeit von f an der Stelle x^* dann $|f(x) - f(x^*)| < \varepsilon$ für alle $x \in [a_n, b_n]$ mit hinreichend großem n, weil ja die Intervalllängen wegen fortgesetzter Halbierung beliebig klein werden. Da c zwischen $f(a_n)$ und $f(b_n)$ liegt, folgt $|f(x^*) - c| < 2\varepsilon$ im Falle des Zwischenwertsatzes bzw. $f(x^*) \geq \sup_{[a,b]} f - \varepsilon$, im Falle des Satzes vom Maximum. Nun war aber $\varepsilon > 0$ beliebig klein, und daher gilt $f(x^*) = c$ im ersten Fall und $f(x^*) \geq \sup_{[a,b]} f \geq f(x^*)$ im zweiten; ersteres beweist den Zwischenwertsatz und letzteres den Satz vom Maximum.

DISKUSSION: 1) Es wird oft gesagt, dass eine reelle Funktion auf einem Intervall stetig sei, wenn man ihren Graphen "ohne abzusetzen zeichnen" kann. Der Zwischenwertsatz besagt dann in anschaulicher Interpretation, dass man bei diesem Zeichnen, wenn man von einer horizontalen Geraden in der Zeichenebene zu einer anderen horizontalen Geraden fortschreitet, alle zwischen den beiden Geraden befindlichen Parallelen treffen muss. Das ist anschaulich evident, aber kein Beweis. Außerdem gibt es stetige Funktionen, deren Graph so stark oszillierend auf und ab verläuft, dass er unendliche Länge hat — dann kann man sich sehr darüber streiten, ob ein solcher Graph "ohne abzusetzen zu zeichnen" ist (nicht in endlicher Zeit, jedenfalls).

Die Voraussetzung der Stetigkeit der Funktion f ist für die Gültigkeit der Aussage des Zwischenwertsatzes wesentlich. Wenn z.B. f eine Sprungstelle zwischen a und b hat, so können die Werte $f(x)$ die Zahl c zwischen $f(a)$ und $f(b)$ ja einfach überspringen. Die Abbildung der unstetigen Kostenfunktion weiter oben demonstriert, dass eine solche Funktion eben nicht jeden Zwischenwert zwischen zwei Funktionswerten annehmen muss. (Es gibt aber auch unstetige Funktionen, für welche die Aussage des Zwischenwertsatzes gültig ist; diese haben dann aber keine Sprungstellen oder Unendlichkeitsstellen, sondern komplizierte Oszillations-Unstetigkeitsstellen.) Wichtig ist auch, dass die stetige Funktion auf einem *Intervall* betrachtet wird. Sonst könnte die (womöglich einzige) Stelle x^* zwischen a und b im Definitionsbereich fehlen, an der die Funktion den Zwischenwert c annehmen würde.

2) *Anwendung des Zwischenwertsatzes auf die Lösung von Gleichungen*: Der Zwischenwertsatz garantiert, dass eine Gleichung $f(x) = c$ (mindestens) eine Lösung im Intervall I besitzt, wenn die stetige Funktion f auf I sowohl Werte $> c$ als auch Werte $< c$ annimmt. Mit diesem Argument haben wir z.B. in Kapitel 2 schon die Lösbarkeit von Effektivzinsgleichungen begründet. Da die Gleichung auch $f(x) - c = 0$ geschrieben werden kann, gelten folgende äquivalente Formulierungen des Zwischenwertsatzes:

- **Nullstellensatz**: *Nimmt eine stetige reelle Funktion auf einem Intervall positive und negative Werte an, so hat sie mindestens eine Nullstelle in diesem Intervall.*
- **Satz vom einheitlichen Vorzeichen**: *Hat eine stetige reelle Funktion auf einem Intervall keine Nullstelle, so ist sie dort überall positiv oder überall negativ.*

Das lässt sich auf jede elementare Funktion f anwenden, insbesondere also auf Polynomfunktionen und auf rationale Funktionen (auf Intervallen ohne Polstellen); denn elementare Funktionen sind ja stetig. Zum Beispiel folgt, dass *jedes Polynom von ungeradem Grad mindestens eine reelle Nullstelle in* $\mathbb{R}$ *hat*, weil solche Funktionen positive und negative Werte annehmen (sogar von beliebig großem Betrag). Im allgemeinen gibt es mehrere verschiedene Nullstellen bzw. Lösungen einer Gleichung $f(x) = c$. Wenn aber f streng monoton ist (wachsend oder fallend), so gibt es offensichtlich nur eine, weil ja dann f an verschiedenen Stellen auch verschiedene Werte hat.

3) Der Beweis oben gibt auch ein Verfahren, wie man eine Lösung zu $f(x) = c$ bzw. eine Nullstelle (wenn $c = 0$) näherungsweise berechnen kann, nämlich das **Intervallhalbierungsverfahren**: Man startet mit $[a_0, b_0]$ so, dass z.B. $f(a_0) < c < f(b_0)$ ist, dann halbiert man das Intervall und wählt eine Hälfte $[a_1, b_1]$ so aus, dass immer noch $f(a_1) < c < f(b_1)$ ist, und so fährt man fort. Das Verfahren endet entweder, wenn f in einem Intervallmittelpunkt den Wert c genau annimmt, oder es produziert Intervalle $[a_n, b_n]$ mit schnell gegen Null gehender Länge, in denen eine Lösung x^* liegt, wobei diese durch a_n und b_n mit einem Fehler kleiner als $2^{-n}(b_0 - a_0)$ approximiert wird. Diese Methode ist schon ganz brauchbar, wird aber in der Praxis oft modifiziert. Zum Beispiel bietet sich ein *Intervallzehntelungsverfahren* an, weil man dabei mit jedem Schritt eine Dezimale mehr von der gesuchten Lösung erhält. Auch ist es sinnvoll, die Intervalle $[a_n, b_n]$ nicht durch den Mittelpunkt zu teilen, sondern im Verhältnis $(c - f(a_n)) : (f(b_n) - c)$, weil die Lösung näher bei a_n liegen wird, wenn $c - f(a_n)$ schon klein ist; diese Vorgehensweise heißt “*regula falsi*”.

4) Eine weitere äquivalente Formulierung des Zwischenwertsatzes ist folgende: *Eine stetige Funktion bildet Intervalle auf Intervalle ab*; denn ein Intervall ist ja definitionsgemäß eine Teilmenge von $\mathbb{R}$, die mit je zwei Punkten auch alle dazwischen liegenden Werte enthält. Insbesondere *ist der Wertebereich selbst ein Intervall* J, wenn die stetige Funktion auf einem Intervall I definiert ist. Nehmen wir noch zusätzlich an, dass f streng monoton wachsend ist, so gibt es dann zu jedem $y \in J$ genau ein Element $x \in I$ mit $f(x) = y$, und indem wir y diese eindeutige Stelle x zuordnen, erhalten wir die auf J definierte *Umkehrfunktion* $g = f^{-1}$ zu f mit I als Wertemenge, also $g(y) := x$, wenn $f(x) = y$. Der Graph von g entsteht aus dem von f durch Spiegelung an der Diagonalen $\{(x, y) : x = y\}$ in der Zeichenebene; das haben wir schon festgestellt. Wenn man Stetigkeit als Möglichkeit interpretiert, den Graphen “ohne abzusetzen zu zeichnen”, so ist daher anschaulich klar, dass die Umkehrfunktion wieder stetig ist. Das läßt sich auch beweisen: Ist $y = f(x) \in J$ und $\varepsilon > 0$, so betrachten wir Punkte $x_1 < x < x_2$ in $I \cap]x - \varepsilon, x + \varepsilon[$ (wenn x Randpunkt von I ist, so kann man analog argumentieren); dann liegt y im Intervall $]f(x_1), f(x_2)[$, und wenn wir $\delta > 0$ kleiner als den Abstand von y zu den Randpunkten dieses Intervalls wählen, so folgt $g(y') \in]x - \varepsilon, x + \varepsilon[$, also $|g(y') - g(y)| < \varepsilon$, für alle $y' \in J$ mit $|y' - y| < \delta$. Ähnlich kann man argumentieren, wenn f streng fallend ist. Außerdem haben wir schon gesehen, dass die Umkehrfunktion streng monoton ist im gleichen Sinne wie f. Daher gilt:

- **Satz von der stetigen Umkehrfunktion**: *Eine stetige streng wachsende (bzw. fallende) Funktion f auf einem Intervall I hat ein Intervall J als Wertemenge und besitzt eine stetige streng wachsende (bzw. fallende) Umkehrfunktion f^{-1} auf J.*

5) Die Voraussetzungen des Satzes vom Maximum und Minimum sind allesamt notwendig. Hat f eine einzige Unstetigkeitsstelle, so kann es passieren, dass f kein Maximum oder Minimum annimmt, weil dafür nur diese Stelle in Frage kommt, dort aber der Funktionswert kein Extremum ist. Beispiel: $f(x) := x$ für $0 < x < \frac{1}{2}$, $f(x) := 1 + x$ für $-\frac{1}{2} < x < 0$ und $f(0) := \frac{1}{2}$; diese Funktion nimmt auf $[-\frac{1}{2}, \frac{1}{2}]$ Werte zwischen 0 und 1 an, wobei ihre Werte 0 und 1 beliebig nahe kommen, das Maximum müsste also 1 sein und das Minimum 0, aber diese beiden Werte werden nicht angenommen, also hat die Funktion weder Maximum noch Minimum. (Eine Graphenskizze macht das klarer.) Wichtig ist auch, dass das Grundintervall I kompakt ist, d.h. beschränkt ist und seine beiden Endpunkte a, b enthält. Andernfalls kann es passieren, dass man das Supremum oder Infimum der Funktion beliebig gut approximiert, wenn man sich aus I einem Randpunkt oder $\pm\infty$ annähert, dass man es aber eben nie erreicht. Beispiele: $f(x) := x$ auf $I :=]0, 1[$ (das Supremum wird bei $x = 1$, das Infimum bei $x = 0$ erreicht, aber diese Punkte gehören nicht zu I) oder $f(x) := x/(1 + x)$ auf $I := [0, \infty[$ (das Supremum 1 wird bei $x \to \infty$ erreicht, aber auf I nicht angenommen). Außerdem braucht eine stetige Funktion auf einem nichtkompakten Intervall nicht von oben beschränkt zu sein, wie $f(x) := x$ auf $[0, \infty[$ und $f(x) = 1/x$ auf $]0, 1]$ demonstrieren; dann kann es natürlich auch kein Maximum geben (analog kein Minimum, wenn f auf I von unten unbeschränkt ist). Ein (endliches) Maximum ist offenbar eine obere Schranke und ein Minimum eine untere Schranke für eine Funktion. Daher folgt aus dem Extremstellensatz das Korollar:

- **Beschränktheitssatz**: *Auf kompakten Intervallen sind stetige Funktionen beschränkt.*

6) Oft hat man jedoch die Aufgabe, eine stetige Funktion auf einem nichtkompakten Intervall I, etwa auf $\mathbb{R} =]-\infty, \infty[$ oder auf $\mathbb{R}_{>0} =]0, \infty[$, zu minimieren. Der Satz vom Minimum läßt sich dann nicht direkt anwenden. Aber in solchen Fällen kann folgendes Prinzip helfen:

- *Wachstumsbedingungen an f können Nichtkompaktheit des Intervalls kompensieren.*

Wenn wir z.B. wissen, dass f nahe den (endlichen oder unendlichen) Randpunkten sehr große Werte hat, so brauchen wir nach Minimumstellen nur in einem kompakten Teilintervall $[a, b]$ zu suchen, außerhalb dessen die Funktionswerte schon zu groß sind, um als Minimum in Frage kommen zu können. Beispielsweise hat $f(x) = x^4 - 3x^3 + 2x$ eine Minimumstelle auf $\mathbb{R}$, weil $f(0) = 0$ ist, die Funktionswerte außerhalb eines genügend großen kompakten Intervalls $[a, b]$ aber offenbar größer als Null sind (das Intervall $[-4, 4]$ reicht hier schon; wenn man a klein genug negativ wählt und b großgenug positiv, so sind die Funktionswerte außerhalb von $[a, b]$ sogar beliebig groß), so dass das Minimum der stetigen Funktion auf $[a, b]$ auch ihr Minimum auf ganz $\mathbb{R}$ ist.

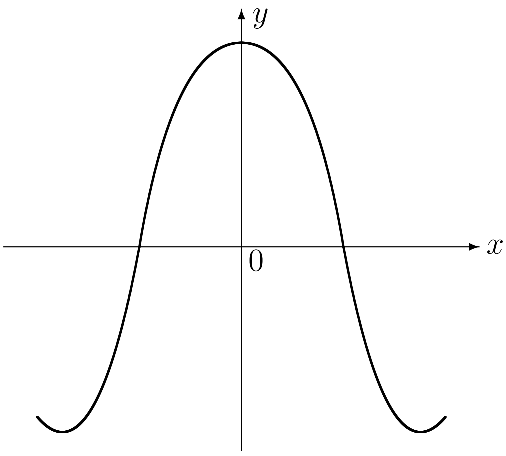
7) Anders als beim Zwischenwertsatz gibt die Intervallhalbierungsmethode kein konstruktives Verfahren zur näherungsweisen Bestimmung von Extremstellen. Der Grund ist, dass man mit der Berechnung von endlich vielen Funktionswerten niemals entscheiden kann, in welcher Hälfte eines Intervalls eine gegebene Funktion letztlich größere Werte annimmt. Gute Methoden zur Extremstellenberechnung gibt es nur, wenn man Funktionen mit "besseren" Eigenschaften als bloße Stetigkeit betrachtet, z.B. dehnungsbeschränkte oder differenzierbare Funktionen. Darauf kommen wir in der Differentialrechnung zurück. ■

Wir betonen nochmals, dass alle elementaren Funktionen stetig sind. Auf solche, durch einen Rechenterm definierte konkrete Funktionen lassen sich also alle obigen Sätze und Folgerungen anwenden!

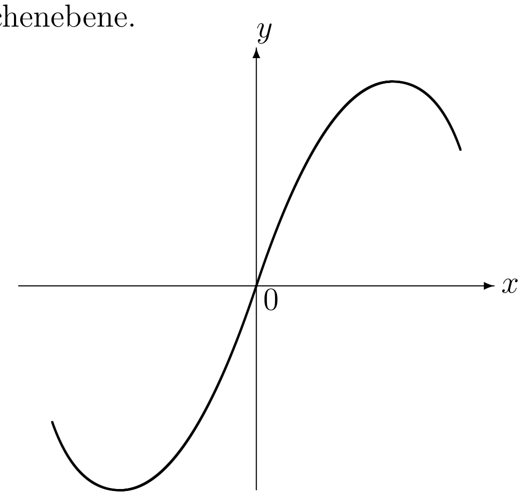
Der Vollständigkeit halber gehen wir noch kurz auf zwei weitere Grundeigenschaften von Funktionen einer reellen Variablen an, die mit Symmetrien ihres Graphen zusammenhängen und in der Mathematik große Bedeutung haben, für mathematische Modelle in der Ökonomie aber weniger wichtig sind.

DEFINITION und DISKUSSION (***Symmetrieeigenschaften von Funktionen***):

1) Ist D ein zu 0 symmetrischer Bereich in $\mathbb{R}$, d.h. D enthält mit jeder Zahl x auch die Gegenzahl $-x$, so heißt $f : D \to \mathbb{R}$ eine **gerade Funktion** auf D, wenn $f(-x) = f(x)$ ist an jeder Stelle $x \in D$, und f heißt **ungerade Funktion**, wenn $f(-x) = -f(x)$ ist für jedes $x \in D$. Eine Funktion ist gerade, genau wenn ihr Graph zur vertikalen Achse symmetrisch ist, und sie ist ungerade, genau wenn ihr Graph punktsymmetrisch zum Nullpunkt ist. Für ungerade Funktionen gilt insbesondere $f(0) = 0$, wenn $0 \in D$, d.h. der Graph geht durch den Nullpunkt $(0, 0)$ der Zeichenebene.



gerade Funktion



ungerade Funktion

2) Jede Funktion der Form $f(x) = g(x^2)$ oder $f(x) = g(|x|)$, d.h. der Wert von $f(x)$ hängt nur vom Quadrat x^2 bzw. vom Betrag $|x|$ ab (was auf dasselbe hinausläuft, weil $x^2 = |x|^2$ ist und $|x| = \sqrt{x^2}$), jede solche Funktion ist offenbar gerade. Das gilt also insbesondere für die Betragsfunktion selbst und für Potenzfunktionen $x^{2m} = (x^2)^m$ auf $\mathbb{R}$ bzw. $\mathbb{R}_{\neq 0}$ mit geradem ganzen Exponenten. Gerade Funktionen sind auch allgemeinere Polynomfunktionen $p(x)$, in denen nur Potenzen von x mit geraden Exponenten vorkommen und überhaupt alle Funktionen, die durch Rechenterme dargestellt werden können, in denen nur Potenzen x^2, $x^4 = (x^2)^2$, $x^6 = (x^2)^3, \ldots$ mit geraden Exponenten auftauchen. Umgekehrt jede gerade Funktion in der Form $f(x) = g(|x|) = g(\sqrt{x^2})$ darstellbar.

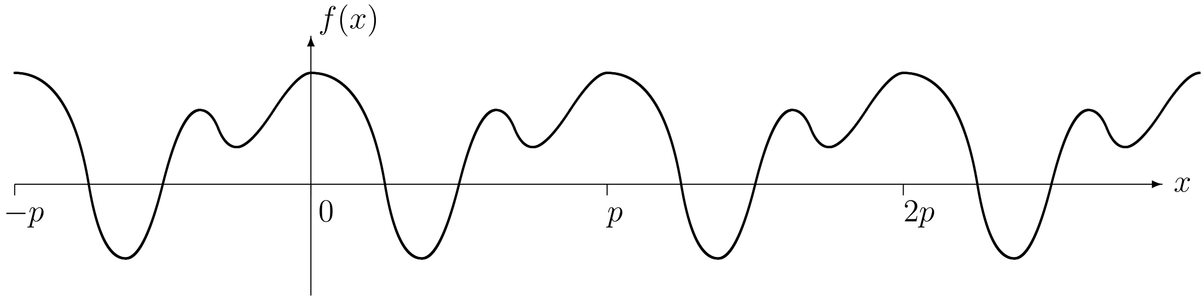
Auf der anderen Seite ist jede Funktion der Form $f(x) = xg(x)$ bzw. $f(x) = (\operatorname{sign} x)g(x)$ mit geradem g eine ungerade Funktion, und jede ungerade Funktion lässt sich so schreiben (dazu setze $g(0) := 0$ und $g(x) := \frac{1}{x}f(x)$ für $x \neq 0$ bzw. $g(x) := (\operatorname{sign} x)f(x)$; dann ist h eine gerade Funktion.) Insbesondere sind Potenzfunktionen x^n auf $\mathbb{R}$ bzw. $\mathbb{R}_{\neq 0}$ mit ungeradem ganzen Exponenten ungerade Funktionen, und allgemeiner gilt das für alle Polynomfunktionen $p(x)$, in denen nur solche Potenzen von x auftreten.

Summen von Funktionen gleicher Parität (d.h. beide gerade oder beide ungerade) haben wieder diese Parität. Produkte und Quotienten von Funktionen gleicher Parität sind gerade, bei Funktionen ungleicher Parität sind das Produkt und der Quotient ungerade. Eine Verkettung ist gerade, wenn die innere Funktion gerade ist oder die innere Funktion ungerade und die äußere gerade; die Verkettung von zwei ungeraden Funktionen ist dagegen ungerade, ebenso die Umkehrfunktion zu einer ungeraden umkehrbaren Funktion.

3) Das Konzept der *Parität* von Funktionen, d.h. der Eigenschaft, gerade oder ungerade zu sein, ist für die Wirtschaftswissenschaft von geringer Bedeutung, weil es nur Sinn macht, wenn die Funktion auf einem zum Nullpunkt symmetrischen Intervall betrachtet wird, während ökonomische Variablen meistens positive Werte haben (also auf der Zahlengeraden rechts von Null liegen) und negative Werte dieser Variablen (die links von Null liegen) nicht sinnvoll sind. Die Begriffsbildung wird aber immer wieder nützlich sein, um den Verlauf von Funktionen über ganz $\mathbb{R}$ zu verstehen, auch von solchen, die in mathematischen Modellen für ökonomische Vorgänge auftauchen, — vorausgesetzt natürlich, die Funktionen haben überhaupt eine Parität.

Die meisten Funktionen sind nämlich weder gerade noch ungerade; denn im Allgemeinen ist ein Funktionsgraph weder zur vertikalen Achse noch zum Ursprung symmetrisch. Man kann aber jede Funktion f auf einem zu 0 symmetrischen Bereich D additiv zerlegen in einen *geraden Anteil* f_{g} und einen *ungeraden Anteil* f_{u}, so dass $f(x) = f_{\mathrm{g}}(x) + f_{\mathrm{u}}(x)$ ist für alle $x \in D$. Dazu setzt man $f_{\mathrm{g}}(x) := \frac{1}{2}(f(x)+f(-x))$ und $f_{\mathrm{u}}(x) := \frac{1}{2}(f(x)-f(-x))$, und dies ist auch die einzige Möglichkeit. Zum Beispiel ist der hyperbolische Kosinus $\cosh x = \frac{1}{2}(\mathrm{e}^x + \mathrm{e}^{-x})$ der gerade Anteil und der hyperbolische Sinus $\sinh x = \frac{1}{2}(\mathrm{e}^x - \mathrm{e}^{-x})$ der ungerade Anteil der Exponentialfunktion e^x.

4) Eine Symmetrie des Graphen liegt auch vor, wenn er durch eine horizontale Verschiebung um eine Länge $p > 0$ mit sich selbst zur Deckung gebracht werden kann. Für die zugehörige Funktion f auf $\mathbb{R}$ bedeutet dies $f(x+p) = f(x)$ für alle $x \in \mathbb{R}$, und man nennt f dann eine **periodische Funktion** mit der Periode p.



eine periodische Funktion mit der Periode p

Bekannte periodische Funktionen sind die *Kreisfunktionen* Sinus und Kosinus mit dem Kreisumfang 2π als Periode, und in der Mathematik wird gezeigt, dass man jede periodische Funktion in gewisser Weise durch diese beiden periodischen Grundfunktionen ausdrücken kann.

In der Wirtschaftswissenschaft spielen periodische Funktionen keine große Rolle. Zwar gibt es den sprichwörtlichen “Schweinezyklus” — das Angebot an Schweinefleisch sinkt, der Marktpreis steigt, die Produktion wird erweitert, das Angebot wächst, der Marktpreis sinkt, die Produktion wird reduziert, das Angebot sinkt, . . . — und man kann an Konjunkturzyklen denken oder an jahreszeitlich bedingte Schwankungen gewisser Variablen in der Volkswirtschaft, aber insgesamt gesehen sind streng periodische Vorgänge in der Ökonomie wenig realistisch, und deshalb tauchen auch periodische Funktionen in mathematischen Modellen ökonomischer Vorgänge eher selten auf. Daher brauchen wir dieses Konzept hier nicht weiter zu verfolgen. ■

Wir schließen diesen Abschnitt mit einer Diskussion *ökonomischer Funktionen*, worunter wir Funktionen verstehen, die in diversen ökonomischen Situationen benutzt werden, um die gegenseitige Abhängigkeit der ökonomischen Größen mathematisch zu modellieren. Dabei ergibt sich aus der ökonomischen Situation im Allgemeinen eine *Vorinformation* über Definitionsbereich, Wertemenge und gewisse Eigenschaften der Funktionen. Die Aufgabe der Wirtschaftsmathematik ist es, einen geeigneten *mathematischen Ansatz* für eine konkrete Funktionsvorschrift zu finden, der diese Vorinformation berücksichtigt. Meistens geschieht dies durch eine Formel, die neben der unabhängigen Variablen noch einen oder mehrere Parameter (Koeffizienten) enthält, die man durch *Parameteranpassung* dann so einrichtet, dass die Funktion möglichst gut mit empirischen Daten übereinstimmt und somit als mathematisches Modell für die betrachtete ökonomische Situation tauglich erscheint. Was "möglichst gute" Übereinstimmung mit den Daten bedeutet, ist im konkreten Fall nicht immer leicht zu entscheiden. Für die angemessene Wahl der Parameter sind daher eigene mathematische Methoden in der Statisik entwickelt worden. Natürlich ist die Aufstellung eines mathematischen Modells nicht Selbstzweck, sondern man kann damit dann auch mögliche, erwartete oder erwünschte Situationen durchrechnen, über die noch keine Erfahrungen vorliegen, und dazu anhand des Modells Prognosen stellen, die einigermaßen gut begründet scheinen. (Sicherheit gibt es nie; denn im Gegensatz zur Physik richtet sich die ökonomische Wirklichkeit selten nach den mathematischen Modellen.)

Im folgenden beschreiben wir einige wichtige ökonomische Funktionen und ihre qualitativen Eigenschaften.

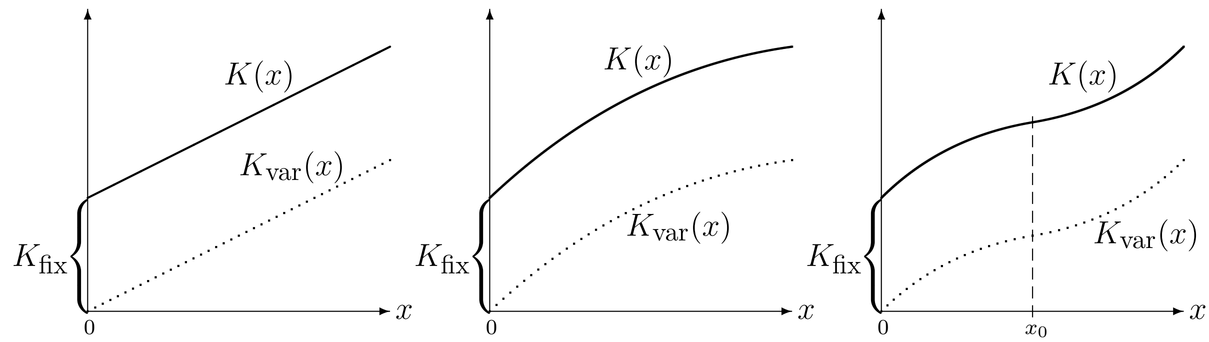
BEISPIELE (***ökonomische Funktionen***):

(1) Kostenfunktionen $K(x)$ beschreiben die Kosten der Produktion von x Mengeneinheiten eines hergestellten Produkts. Die unabhängige Variable x wird hier als **Output** bezeichnet. Sie ist nichtnegativ und meist durch eine Kapazitätsgrenze $x_{\max}$ von oben beschränkt; der Definitionsbereich der Kostenfunktion ist also $\mathbb{R}_{\geq 0}$ oder $[0, x_{\max}]$. Die nichtnegative abhängige Variable $K(x)$ gibt die **Kosten** der Produktion von x Mengeneinheiten des Produkts an.

Die (Gesamt–)Kosten $K(x)$ setzen sich additiv zusammen aus den **Fixkosten** $K_{\text{fix}}(x)$ und den **variablen Kosten** $K_{\text{var}}(x)$. Die Fixkosten sind die Bereitstellungskosten für die Produktionsstätte, die Maschinen etc. Sie sind meistens konstant, können aber bei quantitativer Anpassung der Produktion durch Zuschalten neuer Produktionsaggregate bei gewissen Schwellenwerten des Outputs auch sprunghaft ansteigen, so dass die Fixkostenfunktion eine Treppenfunktion ist. Die variablen Kosten $K_{\text{var}}(x)$ geben die unmittelbar durch die Produktion von x Einheiten entstehenden (Energie-, Rohstoff-, Verschleiß-, Arbeits- ...)Kosten an; sie sind immer wachsend in x.

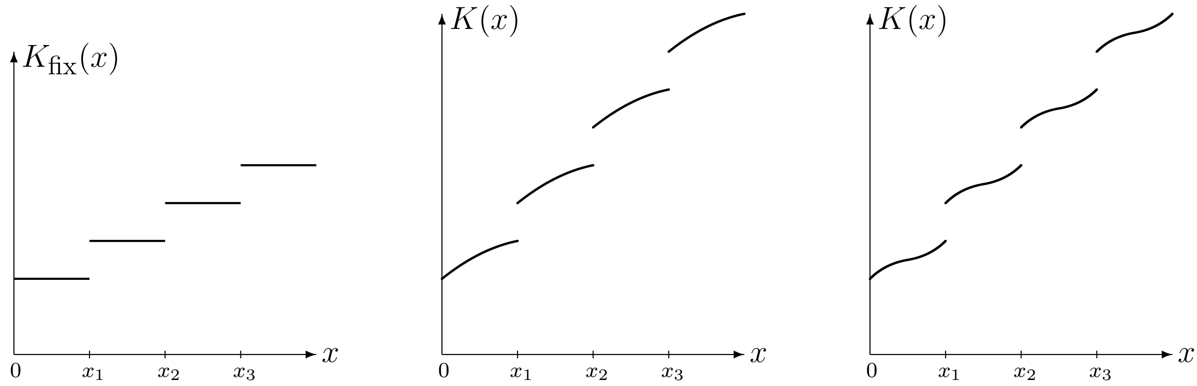
Im einfachsten Fall sind die variablen Kosten proportional zum Output, also $K_{\text{var}}(x) = \kappa \cdot x$ mit einem Proportionalitäzsfaktor $\kappa > 0$, genannt **Stück–variable Kosten**, weil er die variablen Kosten pro produzierte Einheit ("pro Stück") angibt. Bei konstanten Fixkosten hat man dann eine lineare Kostenfunktion $K(x) = K_{\text{fix}} + \kappa \cdot x$. Häufig ist aber das Wachstum der variablen Kosten nicht konstant, sondern schwächt sich mit zunehmendem Output x ab, weil die Produktionsanlagen bei besserer Auslastung kostengünstiger arbeiten. In diesem Fall sind die variablen Kosten (und bei konstanten Fixkosten auch die Gesamtkosten) also degressiv wachsend und positiv. In der Wirtschaftsmathematik wird ein solcher Funktionsverlauf **neoklassisch** genannt. Typische Modellfunktionen für dieses Verhalten sind $K(x) = \sqrt{a+bx}$ oder $K(x) = (a+bx)^s$ mit Parametern $a, b \geq 0$ und $0 < s < 1$.

In wieder anderen Fällen nehmen aber die variablen Kosten bei sehr großem Output x progressiv zu, weil die Produktionsaggregate unter Volllast ungünstiger arbeiten. Die variablen Kosten (und die Gesamtkosten) zeigen dann bis zu einem gewissen Schwellenwert x_0 des Outputs einen linear oder degressiv wachsenden Verlauf, nach dem "Wendepunkt" aber einen progressiv wachsenden. (In der Mathematik heißt ein Schwellenwert, bei dem konkaves Funktionsverhalten in konvexes umschlägt oder umgekehrt ein *Wendepunkt*.) Ein derartiger Verlauf einer positiven Funktion auf $\mathbb{R}_{>0}$ wird in der Wirtschaftsmathematik **ertragsgesetzlich** genannt (wobei diese Terminologie aus einem anderen ökonomischen Zusammenhang stammt). Typische Ansatzfunktionen mit diesem Verhalten sind kubische Polynome $ax^3 - bx^2 + cx + d$ (mit $a > 0$, $b > 0$, $d \geq 0$ und $b^2 < 3ac$; der Wendepunkt ist $x_0 = \frac{b}{3a}$) oder Varianten des hyperbolischen Sinus $a\mathrm{e}^{cx} - b\mathrm{e}^{cx} + d = \sqrt{ab}\sinh(c(x - x_0)) + d$ (mit $0 < a < b$, $c > 0$ und $d \geq b - a$; der Wendepunkt ist $x_0 = \frac{1}{2c}\ln\frac{b}{a}$).



lineare Kosten *neoklassische Kosten* *ertragsgesetzliche Kosten*

Bei quantitativer Produktionsanpassung erhält man für jedes einzelne Produktionsaggregat einen Kostenverlauf wie in diesen Abbildungen, und die Gesamtkosten zeigen abschnittsweise dasselbe Verhalten, wobei an den Stellen x_1, x_2, $x_3, \ldots$ zwischen den Abschnitten jeweils ein Fixkostensprung eintritt. Für gleiche lineare variable Kosten bei den Produktionsaggregaten haben wir die Situation schon in 4.1 abgebildet. Die folgenden Abbildungen zeigen den Verlauf bei neoklassischen oder ertragsgesetzlichen Kosten jedes einzelnen Produktionsaggregats.



Kostenfunktionen bei quantitativer Produktionsanpassung

Um die Kosten für einen konkreten Produktionsvorgang zu ermitteln, muss man für jeden *Produktionsfaktor* r (Rohstoff, Energie, Verschleiß, Arbeit, ...) die für die Produktion von x Einheiten benötigte Faktormenge $r(x)$ (in Faktoreinheiten) und den bei Einkauf von r Faktoreinheiten zu zahlenden Preis $p(r)$ pro Einheit kennen. (Oft ist $p(r) \equiv p$ konstant; es kann aber auch sein, dass $p(r)$ mit wachsendem r wegen Mengenrabatt fällt oder, wie z.B. beim Faktor Arbeit wegen Schichtzulagen, mit wachsendem r steigt.) Außerdem ist ein

Fixkostenanteil $K_{r,\text{fix}}$ für die Bereitstellung des Produktionsfaktors r zu berücksichtigen und ein Fixgrundkostenanteil $K_{0,\text{fix}}$ für die Unterhaltung der Produktionsstätte etc. Sind $r_1, \ldots, r_n$ die insgesamt benötigten Produktionsfaktoren und $p_1, \ldots, , p_n$ die zugehörigen Preisfunktionen, so lautet die Gesamtkostenfunktion dann

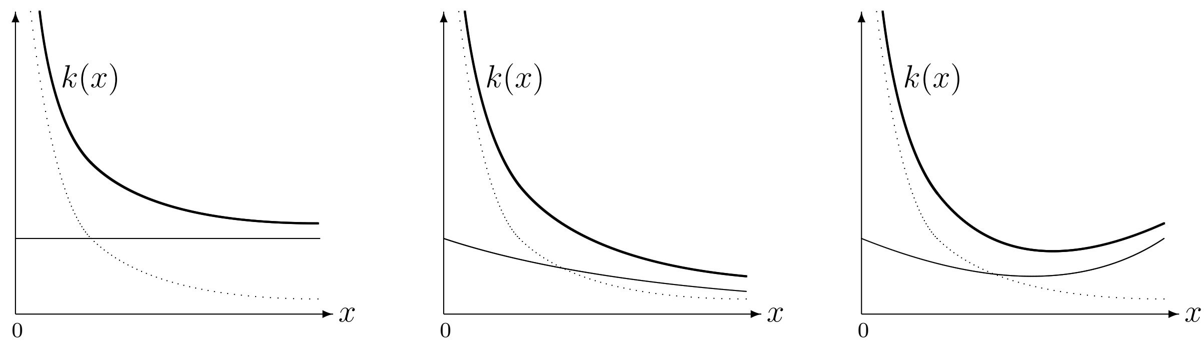
$$K(x) \quad = \quad \underbrace{K_{0,\text{fix}} + \sum_{i=1}^{n} K_{r_i,\text{fix}}}_{K_{\text{fix}}} \quad + \quad \underbrace{\sum_{i=1}^{n} p_i(r_i(x))r_i(x)}_{K_{\text{var}}} \; .$$

In der Praxis ist die Erfassung aller Kosten nicht immer einfach und nicht allein vom Produktionsvorgang abhängig. (Sind z.B. Abschreibungen für Produktionsstätten im eigenen Besitz Produktionskosten?)

(2) Die **Stückkosten** $k(x)$, auch **durchschnittliche Kosten** genannt, geben die Kosten pro produzierte Einheit bei einem Output von x Einheiten an:

$$k(x) \;=\; \frac{K(x)}{x} \;=\; k_{\text{fix}}(x) \;+\; k_{\text{var}}(x) \quad \text{mit} \quad k_{\text{fix}}(x) = \frac{K_{\text{fix}}}{x},\; k_{\text{var}}(x) = \frac{K_{\text{var}}(x)}{x}\,.$$

Dabei geben die **Stück–variablen Kosten** $k_{\text{var}}(x) = \frac{1}{x}K_{\text{var}}(x)$ die variablen Kosten pro produzierte Einheit beim Output x an und die **Stück–Fixkosten** $k_{\text{fix}}(x) = \frac{1}{x}K_{\text{fix}}$ den auf jede produzierte Einheit entfallenden Fixkostenanteil. Die Stückfixkosten sind für kleine Werte von $x > 0$ naturgemäß sehr groß, also wird man sie durch eine Funktion mit Grenzwert ∞ bei $x \searrow 0$ modellieren ($x = 0$ ist rechtsseitige Unendlichkeitsstelle von $k_{\text{fix}}(x)$ und von $k(x)$). Im allgemeinen nehmen die Stückkosten mit wachsendem Output x degressiv ab. Jedenfalls ist das so, wenn die Kosten $K(x)$ degressiv wachsen oder wenn sie linear wachsen (wenn also $k_{\text{var}} \equiv \kappa$ konstant ist); denn $k_{\text{fix}}(x) = \frac{1}{x}K_{\text{fix}}$ nimmt ja ab mit wachsendem x (wobei wir natürlich $K_{\text{fix}} > 0$ annehmen). Ist aber die Produktion bei großem Output teurer (überlastete Maschinen, Schichtzuschläge, ...), so kann es sein, dass die Stückkostenfunktion bei großen Werten von x wachsend ist. Eine Produktion mit einem großen Output, bei der die Stückkosten höher sind als bei kleinerer Produktionsmenge kann sich dennoch rentieren; das hängt von dem erzielbaren Erlös am Markt ab. Typische Ansatzfunktionen für Stückkostenfunktionen ergeben sich durch Division der in (1) angegebenen Funktionen $K(x)$ mit x. Typische Stückkostenverläufe zeigt die folgende Abbildung:

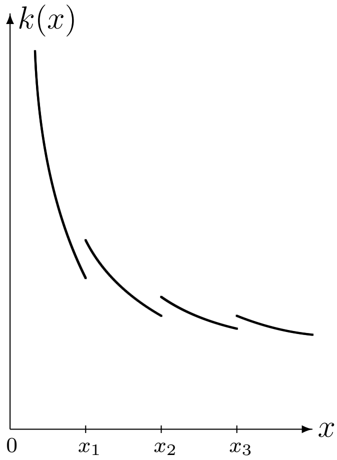


$K(x)$ *linear* $K(x)$ *neoklassisch* $K(x)$ *ertragsgesetzlich*

Stückkosten $k(x)$ —, *Stück–variable Kosten* $k_{var}(x)$ — *und Stück–Fixkosten* $k_{fix}(x)$ ······

(Der in der letzten Abbildung beschriebene Stückkostenverlauf, bei dem die Stückkosten mit großem Output wieder wachsen, stellt sich nicht bei jeder ertragsgesetzlichen Kostenfunktion $K(x)$ ein, sondern nur dann, wenn $K(x)$ bei $x \to \infty$ schneller als linear wächst.)

Bei quantitativer Produktionsanpassung ergeben sich Stückkostenfunktionen mit Sprungstellen wie hier. Bei Zuschaltung eines neuen Produktionsaggregats nehmen die Stückfixkosten und damit auch die Stückkosten wegen der Fixkostenerhöhung sprunghaft zu, wobei die Sprünge aber mit wachsender Zahl der betriebenen Aggregate kleiner werden, weil sich die Fixkosten des zugeschalteten Aggregats dann auf mehr Output–Einheiten verteilen. Weil das auch für die variablen Kosten des zugeschalteten Aggregats gilt, verläuft der Graph der Stückkostenfunktion zwischen zwei Spungstellen umso flacher, je mehr Aggregate in Betrieb sind.



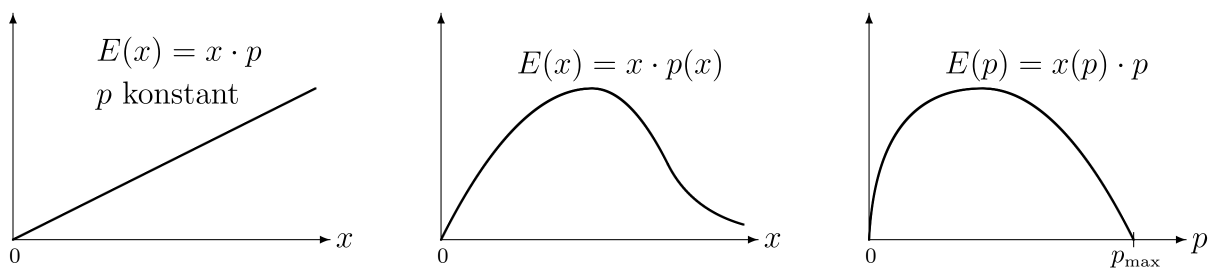
(3) Die **Umsatz–** oder **Erlösfunktion** bzw. die **Gewinnfunktion** geben den Umsatz (Erlös) $E(x)$ bzw. den Gewinn $G(x)$ (jeweils in Geldeinheiten, bezogen auf jeweils eine Rechnungsperiode) in Abhängigkeit von der abgesetzten Menge x (Mengeneinheiten) eines produzierten Gutes an. Der Umsatz / Erlös ist per Definition die für die am Markt abgesetzte Menge erzielte Einnahme, bei einem Marktpreis p pro Mengeneinheit also

$$E(x,p) \;=\; x \cdot p\,.$$

Dies ist eine Funktion von zwei Veränderlichen. Um eine davon zu eliminieren, muss man entweder den Marktpreis $p(x)$ in Abhängigkeit von der angebotenen und abgesetzten Menge x kennen — am einfachsten ist das bei konstantem Marktpreis, aber bei einem Monopol wird z.B. den Marktpreis durch die angebotene Menge beinflusst, so dass er eine fallende Funktion des Output x ist — oder die in Abhängigkeit vom festgesetzten Preis p am Markt nachgefragte Menge $x(p)$, die mit wachsendem p natürlich kleiner wird. (Dabei ist unterstellt, dass der Hersteller die nachgefragte Menge $x(p)$ auch produzieren kann und exakt diese Menge auf den Markt bringt. Für jeden Preis q ist dann $y = x(q)$ genau der Output, bei dem sich der Marktpreis $q = p(y)$ ergibt, also ist $x(p)$ die Umkehrfunktion zu $p(x)$; denn wäre $p(y)$ kleiner bzw. größer als q , so hätte der Unternehmer die bei einem kleineren bzw. größeren Preis als q entstehende Nachfrage erfüllt und bzgl. der Nachfrage bei Preis q einen zu großen bzw. zu kleinen Output produziert.) Als Funktion der Variablen des Output x bzw. des Preises p lautet nun die Erlösfunktion

$$E(x) = x \cdot p(x) \qquad \text{bzw.} \qquad E(p) = x(p) \cdot p\,.$$

Bei konstantem Marktpreis p ist $E(x) = x \cdot p$ homogen linear wachsend. Ist der Marktpreis $p(x)$ eine fallende Funktion von x, so ist $E(x) = x \cdot p(x)$ typischerweise anfangs auch linear wachsend, dann aber degressiv (höheres Angebot lässt den Preis fallen) und unter Umständen ab einem gewissen Output sogar degressiv gegen Null fallend (totaler Preisverfall bei zu großem Angebot kann den Erlös ruinieren). Entsprechendes gilt für den Erlös $E(p) = x(p) \cdot p$ als Funktion des Preises; für kleine Werte von p ist er nahe Null (wenn der Absatz bei gegen Null fallendem Preis langsamer wächst als der Preis abnimmt), wächst dann mit p, fällt aber bei größeren Werten von p wieder und erreicht Wert Null bei einem Preis $p_{\max}$ (weil man zu höherem Preis nichts mehr verkaufen kann).



einige typische Erlösfunktionen

Der **Gewinn** (bzw. *Verlust*, wenn negativ) ist definiert als Erlös minus Kosten, also lautet die Gewinnfunktion

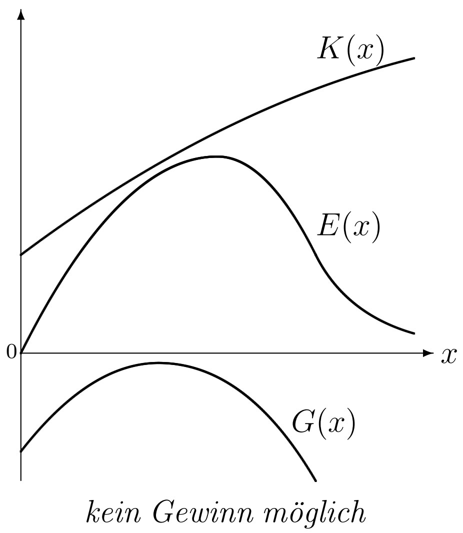
$$G(x) \;=\; E(x) - K(x) \;=\; x \cdot p(x) - K(x)\,,$$

wenn man den (absetzbaren) Output x als unabhängige Variable betrachtet, deren Wert der Hersteller direkt beeinflussen kann, beziehungsweise

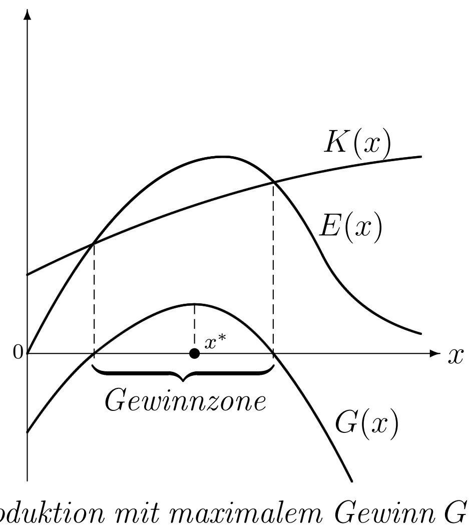
$$G(p) \;= E(p) - K(x(p)) \;=\; x(p) \cdot p - K(x(p))\,,$$

wenn man den Preis p als die Variable ansieht, deren Wert der Produzent festlegt.

Die *typische ökonomische Optimierungsaufgabe* ist nun, erstens festzustellen, ob der Gewinn überhaupt für gewisse Werte der unabhängigen Variablen positiv ist (andernfalls würde man in jedem Fall Verluste machen oder jedenfalls keinen Gewinn erzielen), und im Fall einer positiven Antwort dann festzustellen, für welchen Wert der unabhängigen Variablen der Gewinn maximal ist. Graphisch kann man die **Gewinnzone**, d.h. die Menge der Werte der unabhängigen Variablen, für die sich ein positiver Gewinn ergibt, ermitteln, indem man die Graphen von $E(x)$ und $K(x)$ aufträgt (bzw. von $E(p)$ und $K(x(p))$) und den Bereich auf der horizontalen Achse bestimmt, über dem der Graph von E oberhalb des Graphen von K verläuft (über dem also der Gewinn $E - K$ positiv ist). Meistens ergibt sich ein Intervall als Gewinnzone (wenn sie nicht leer ist); dessen Randpunkte heißen dann die *Gewinnschwellen.* Die Bestimmung des Wertes (oder der Werte) x^* (bzw. p^*) zwischen den Gewinnschwellen, für die der Gewinn maximal ausfällt, kann mit Methoden der Differentialrechnung erfolgen, wenn es nicht elementar möglich ist (wie z.B. bei einer quadratischen Gewinnfunktion).



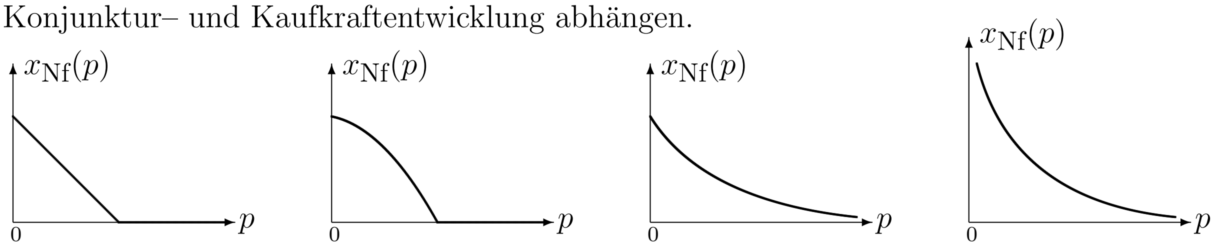
kein Gewinn möglich



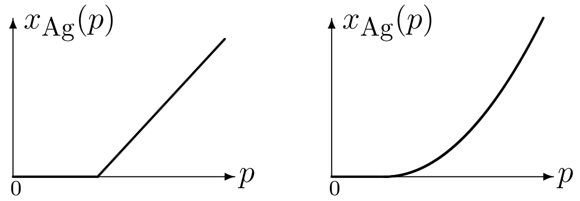
Produktion mit maximalem Gewinn $G(x^*)$

Der durchschnittliche Gewinn pro produzierter Einheit (bzw. Verlust, wenn negativ) heißt **Stückgewinnfunktion** $g(x) = \dfrac{1}{x}G(x) = e(x) - k(x)$, wo $k(x)$ die Stückkosten sind und $e(x) = \dfrac{1}{x}E(x)$ die sog. **Stückerlösfunktion**. Bei linearer Erlösfunktion $E(x) = x \cdot p$ und konstanten Stück–variablen Kosten $\kappa < p$ ist der Stückgewinn $g(x) = p - \kappa - \dfrac{1}{x}K_{\text{fix}}$ positiv, sobald mit Gewinn produziert wird ($x > \frac{1}{p-\kappa}K_{\text{fix}}$) und wächst degressiv gegen den Grenzwert $p - \kappa$ mit zunehmendem Output $x \to \infty$. Für kompliziertere Erlös– und Kostenfunktionen ergeben sich kompliziertere Stückgewinnfunktionen, die typischerweise ihr Maximum an einer von x^* verschiedenen Stelle $x^{**} < x^*$ in der Gewinnzone erreichen.

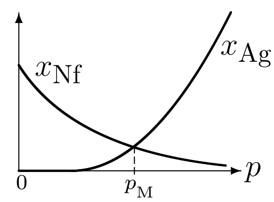
(4) Die **Nachfragefunktion** $x_{\mathrm{Nf}}(p)$ des Preises (*Preis–Nachfrage–Funktion*) gibt die am Markt absetzbare Menge eines Produkts (in Mengeneinheiten) in Abhängigkeit vom Verkaufspreis p (Geldeinheiten pro Mengeneinheit) an. Sie wird bestimmt durch das Kaufverhalten der Nachfrager und ist natürlich monoton fallend. (Eine Preiserhöhung bewirkt eine Abschwächung der Nachfrage — außer vielleicht bei sehr teuren Luxusartikeln.) Dies ist die Funktion $x(p)$, die der Unternehmer zur Ermittelung seiner Gewinnfunktion in (3) verwendet, wenn er den Preis bestimmen und den Produktionsoutput jeweils der Nachfrage anpassen kann. Die **Angebotsfunktion** (*Marktpreis–Angebots–Funktion*) beschreibt dagegen das Verhalten der Anbieter und gibt die produzierte Menge eines Gutes in Abhängigkeit von dem am Markt erzielbaren Preis p an. Sie ist typischerweise wachsend, oft aber konstant Null unterhalb einer gewissen Preisschwelle, bei der sich die Produktion noch nicht lohnt. Wenn weder auf Seiten der Anbieter noch der Nachfrager ein Monopol oder Oligopol besteht, wenn also eine Situation der *vollkommenen Konkurrenz* vorliegt, so wird sich der Marktpreis p_{M} so einstellen, dass $x_{\mathrm{Ag}}(p) = x_{\mathrm{Nf}}(p)$ ist (Angebot = Nachfrage). Graphisch kann man diesen Preis als Rechtswert des Schnittpunktes der Graphen von Nachfrage– und Angebotsfunktion bestimmen. Die Funktionen $x_{\mathrm{Nf}}(p)$ und $x_{\mathrm{Ag}}(p)$ lassen sich (zum Leidwesen der Unternehmer) nur empirisch und mit großer Unschärfe bestimmen, weil sie — wie alles im Wirtschaftsleben — von weiteren Parametern wie Konjunktur– und Kaufkraftentwicklung abhängen.



einige typische Nachfragefunktionen



typische Angebotsfunktionen

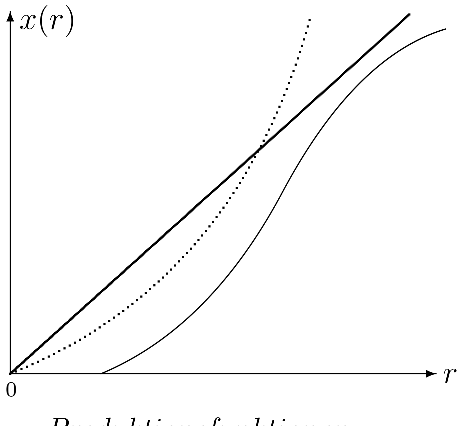


Marktpreis bei vollkommener Konkurrenz

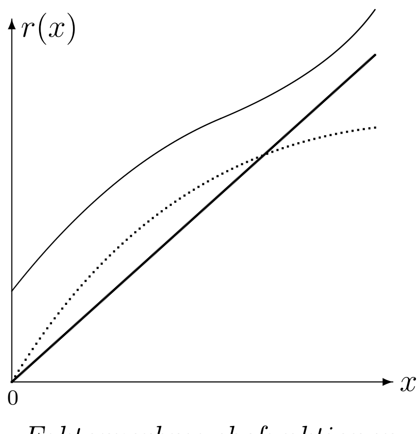
Die erste Nachfragefunktion und die erste Angebotsfunktion kann man durch stückweise lineare Funktionen modellieren, die zweite jeweils durch ein quadratisches Polynom in dem Bereich, wo die Funktion positiv ist. Für die dritte und vierte Nachfragefunktion kann man eine rationale Funktion als Ansatz nehmen (mit einem Pol in 0 im letzten Fall). Für den dritten Typ kann man aber z.B. auch eine logistische Funktion nehmen, die ihre Sättigungsgrenze 0 bei $x \to \infty$ von oben erreicht.

Oft wird auch die **Preisfunktion** der Nachfrage $p_{\mathrm{Nf}}(x)$ (*Nachfrage–Preis–Funktion*) bzw. die **Marktpreisfunktion** $p_{\mathrm{Ag}}(x)$ des Angebots (*Angebots–Marktpreis–Funktion*) angegeben, wobei der Preis als Funktion der nachgefragten bzw. angebotenen Menge x eines Produkts aufgefasst wird. Mathematisch ist das möglich, wenn jedem Wert von $x > 0$ genau ein Preis entspricht; $p_{\mathrm{Nf}}(x)$ bzw. $p_{\mathrm{Ag}}(x)$ sind dann die Umkehrfunktionen zu $x_{\mathrm{Nf}}(p)$ bzw. $x_{\mathrm{Ag}}(p)$ auf dem Intervall, wo diese Funktionen positiv sind. Es erscheint auf den ersten Blick nicht sinnvoll, die direkt kaum beeinflussbare Nachfrage x als unabhängige Variable aufzufassen, aber der Marktpreis p ist ja auch durch den Markt bestimmt und nicht willkürlich veränderbar. Insofern ist bei der Beschreibung des Zusammenhangs zwischen Preis und Nachfrage / Angebot in einer Situation vollkommener Konkurrenz jede Auszeichnung einer der beiden Variablen als unabhängige Variable willkürlich. Anders ist es in einer Monopol–Situation: Der Monopolist kann den Preis p beliebig festsetzen, die Nachfrage $x_{\mathrm{Nf}}(p)$ ist dann eine Funktion davon. Er kann auch das Angebot x alleine bestimmen und die Herausbildung des Preises $p_{\mathrm{Nf}}(x)$ dem Markt überlassen. Wenn er seine Produktion stets exakt nach der Nachfrage richtet, so ist $x_{\mathrm{Ag}}(p) = x_{\mathrm{Nf}}(p)$ und der von ihm festgesetzte Preis p stimmt mit dem am Markt gebildeten Preis $p_{\mathrm{Nf}}(x_{\mathrm{Ag}}(p))$ überein.

(5) Eine **Produktionsfunktion** $x(r)$ gibt den Produktions–Output (in Mengeneinheiten) an als Funktion des *Faktor–Input* r (in Mengeneinheiten). Natürlich hängt der Output stets von mehreren Produktionsfaktoren ab (Rohstoffe, Energie, Arbeit, ...). Hier wird angenommen, dass man einen einzigen Produktionsfaktor variiert und die anderen als konstant ansieht (sog. *c.p.–Bedingung*), man spricht auch von *partieller Faktorvariation*. Das ist eine durchaus fragwürdige Annahme, weil eine Erhöhung des Outputs x im Allgemeinen die Erhöhung des Einsatzes von *mehreren* Faktoren erfordert. Aber nur durch diese Annahme lässt sich überhaupt der Zusammenhang zwischen Output und Faktoreinsatz durch eine reelle Funktion von einer einzigen Variablen darstellen. (Die anderen Faktoren erscheinen als konstante Parameter in dieser Funktion.) Produktionsfunktionen können sehr unterschiedlich aussehen. Im Allgemeinen sind sie wachsend, aber bei gewissen Faktoren ist auch denkbar, dass der Produktionsertrag bei zu großem Faktoreinsatz abnimmt. (Man denke an den Faktor Arbeit und "zu viele Köche verderben den Brei".) Im einfachsten Fall ist der Output $x(r)$ proportional zur eingesetzten Menge r des Produktionsfaktors. In anderen Fällen ist $x(r)$ progressiv wachsend (wenn die Produktion bei höherem Faktoreinsatz effektiver verläuft) oder wechselt mit zunehmendem r von progressivem zu degressivem Wachstum (wenn die Produktion sehr großer Mengen weniger effektiv ist). Dabei ist $x(0) = 0$, wenn der Faktor wirklich für die Produktion erforderlich ist, weil dann bei Faktoreinsatz Null auch der Output Null ist. Es kann sein, dass $x(r) = 0$ ist auf einem ganzen Intervall $0 \le r \le r_{\min}$, nämlich wenn ein positiver Mindestfaktoreinsatz $r_{\min}$ erforderlich ist, um überhaupt die Produktion zu starten.



Produktionsfunktionen



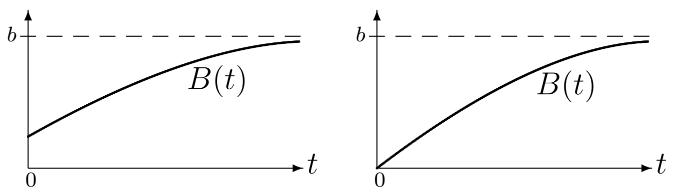
Faktorverbrauchsfunktionen

Eine **Faktorverbrauchsfunktion** $r(x)$ gibt umgekehrt an, wieviele Einheiten eines bestimmtenProduktionsfaktors gebraucht werden, wenn x Output–Einheiten produziert werden sollen. Auch solche Funktionen sind typischerweise wachsend und starten mit einem Wert $r_{\min} = r(0) \ge 0$, der für die Aufnahme der Produktion erforderlich ist. Die Faktorverbrauchsfunktionen sind genuine Funktionen von einer Veränderlichen (sofern nicht Produktionsfaktoren wechselseitig substituiert werden können, so dass verringerter Einsatz des einen durch erhöhten Einsatz eines anderen Faktors kompensiert werden kann).

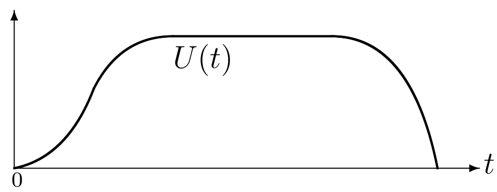
(6) Eine **Konsumfunktion** $C(Y)$ gibt die Ausgaben für Konsumgüter (in Geldeinheiten pro Jahr) eines Haushalts oder einer Population in Abhängigkeit vom Einkommen Y (Geldeinheiten pro Jahr) an. Die zugehörige **Sparfunktion** ist dann $S(Y) = Y - C(Y)$. Eine Konsumfunktion $C(Y)$ ist typischerweise wachsend und die Sparfunktion $S(Y)$ ebenfalls. Letzteres bedeutet, dass die (positiven) Steigungen der Konsumfunktionen kleiner als 1 sind, d.h. Zuwächse bei Y bewirken kleinere Zuwächse bei $C(Y)$. In mikroökonomischem Zusammenhang (einzelne Haushalte) heißen Konsumfunktionen auch *Engel–Funktionen* (nach dem Statistiker Engel). Es gibt viele verschiedene gebräuchliche Engel–Funktionen. Der einfachste Ansatz ist Proportionalität $C(Y) = cY$, $S(Y) = (1-c)Y$ mit einem Proportionalitätsfaktor $0 < c < 1$.

(7) Eine **Investitionsfunktion** $I(p)$ gibt die Investitionsausgaben einer Volkswirtschaft (in Geldeinheiten pro Jahr), einer Branche oder eines Unternehmens in Abhängigkeit vom marktüblichen Sollzinsfuß $p\%$ an. Dann ist $I(p)$ natürlich nichtnegativ und typischerweise fallend. Ein Beispiel ist etwa die Investition in den Wohnungsbau eines Landes in Abhängigkeit vom Hypothekenzinsfuß.

(8) Ein **Produktlebenszyklus** $U(t)$ beschreibt in Abhängigkeit von der Zeit t den Absatz eines Produkts (in Mengeneinheiten). Typischerweise ist der Funktionswert am Anfang Null (etwa zum Zeitpunkt $t = 0$), danach ist $U(t)$ zunächst progressiv wachsend (Einführungsphase), dann wird das Wachstum degressiv (Reifephase) und schließlich ist die Funktion ist für längere Zeit mehr oder weniger konstant Sättigungsphase); am Ende fällt $U(t)$ dann wieder, und zwar oft progressiv, bis wieder der Wert Null in endlicher Zeit erreicht ist (Degenerationsphase; es ist auch denkbar, dass Produkte langsam auslaufen, also mit degressivem Abfallen am Ende des Lebenszyklus). Der Gewinn ist meist schon in der Sättigungsphase rückläufig wegen Preisverfall. Ein elementarer Funktionsansatz für Produktlebenszyklen ist möglich, erscheint aber nicht sinnvoll. Einfacher ist es, die Einführungs– und Reifephase mit einer elementaren Funktion zu modellieren, die nach endlicher Zeit ein gewisses Niveau mit Steigung Null erreicht, dann die Funktion konstant auf diesem Niveau fortzusetzen für die angenommene Dauer der Sättigungsphase und schließlich die Degenerationsphase wieder durch eine elementare Funktion zu modellieren, die von diesem Niveau mit anfänglicher Steigung Null fällt und später den Wert Null erreicht. Die gesamte Funktion $U(t)$ ist dann eine stückweise elementare Funktion (die stetig ist, also ohne Sprungstellen, und deren Steigung sich auch stetig verändert, d.h. es gibt keine “Knickstellen” im Graphen).



logistische Funktionen



Produktlebenszyklus

(9) **Logistische Funktionen** $B(t)$ beschreiben in Abhängigkeit von der Zeit t wachsende Bestände $B(t)$ (in Mengeneinheiten), die mit zunehmender Zeit eine Sättigungsgrenze b erreichen, z.B. den PKW–Bestand in einer Volkswirtschaft. Als Ansatz wird oft $B(t) = \frac{b}{1+ae^{-ct}} = \frac{be^{ct}}{e^{ct}+a}$ mit Parametern a, b, $c > 0$ gewählt oder, um $B(0) = 0$ zu erreichen, $B(t) = \frac{b-be^{-ct}}{1+ae^{-ct}} = \frac{be^{ct}-b}{e^{ct}+a}$. Diese Funktionen erreichen die Sättigungsgrenze b zwar nur im Limes $t \to \infty$, aber sie nähern sich ihr mit exponentieller Geschwindigkeit an, so dass $B(t)$ schon für mäßig große Zeiten praktisch konstant gleich b ist. ■