Geändert am 16.12.14/
Dr. Kurt Jechlitschka

Prof. Dr. Hans G. Bartels

Optimierung multivariabler Funktionen

Inhaltsverzeichnis

1. Einführung 1
2. Geometrische Darstellung von duovariablen Funktionen 2
3. Zweidimensionale Schnitte durch multivariable Funktionen 5
4. Partielle Differentiation 6
5. Notwendige Bedingungen für das Eintreten eine lokalen Extremums 11
6. Lösung des Systems der notwendigen Bedingungen 13
7. Partielle Ableitungen höherer Ordnung 19
8. Hinreichende Bedingungen, Totale Differentiale 21
   - 8.1 Hauptabschnittsdeterminanten der Funktional- bzw. Hesse-Matrix 25
   - 8.2 Die semidefiniten Fälle 29
   - 8.3 Ermittlung der Hauptabschnitts-Determinanten durch Triangulation der Funktionalmatrix 31
   - 8.4 Triangulation bei Nullen auf der Hauptdiagonale 33

1. Einführung

Bei den allermeisten ökonomischen Problemen hängt die abhängige Variable y nicht nur von einer einzigen unabhängigen Variablen x sondern von vielen unabhängigen Variablen x_1, x_2, ..., x_n ab. Der Gewinn z.B. ist gewöhnlich eine Funktion von Fertigung und Vertrieb vieler Produkte (Mehrproduktunternehmen) und nicht nur eines Produkts (Einproduktunternehmen).

Üblicherweise indiziert man die Produkte mit dem Index j, j=1,2, ..., n.

Man muss also unbedingt in der Lage sein, auch mit multivariablen Funktionen umzugehen. Dabei kann man in vielfältiger Weise auf das Wissen zu monovariablen Funktionen zurückgreifen, denn das Meiste, was monovariabel gilt, gilt mehr oder weniger auch multivariabel.

Algebraisch schreibt man eine multivariable Funktion in allgemeiner Form wie folgt $y = f(x_1,x_2,\dots,x_n)$

oder

$y = f(\mathbf{x})$ $\mathbf{x}= (x_1,x_2, \dots ,x_n)^T$ Das fette **x** heißt verbal „Vektor x".

Das hochgestellte T heißt „transponiert", weil Vektoren sich als Spaltenvektoren verstehen; durch das Transponieren wird aus dem Spalten- ein Zeilenvektor.

Insbesondere zur konkreten algebraischen Schreibweise einer multivariablen Funktion gibt es nahezu keine Alternative.

Man kann zwar auch multivariable Funktionen in Form von **Wertetabellen** darstellen, da kommt man aber schnell vom Hundertsten ins Tausendste, d.h. man verliert den Überblick.

In Anbetracht dessen, dass man bei monovariablen Funktionen nahezu alles auch geometrisch darstellen und damit sehr gut verständlich machen kann, stellt sich die Frage, ob man multivariable Funktionen ebenfalls **geometrisch** darstellen kann?

Diese Frage ist sehr differenziert zu beantworten:

1. Duovariable Funktionen, also Funktionen des Typs $y = f(x_1,x_2)$, kann man eingeschränkt geometrisch darstellen.
2. Funktionen mit mehr als drei unabhängigen Variablen kann man allenfalls ein bisschen mit zweidimensionalen Schnitten durch das Funktionsgebirge visualisieren. Darüber hinaus ist eine geometrische Darstellung nicht möglich.

2. Geometrische Darstellung von duovariablen Funktionen

Eine duovariable Funktion hat drei Variable, die abhängige Variable y und die beiden unabhängigen Variablen x_1 und x_2. Eine zutreffende geometrische Darstellung ist damit nur im dreidimensionalen Raum möglich, was durchaus auch realisiert werden kann, etwa durch ein dreidimensionales Modell aus Draht oder als Gips-Relief.

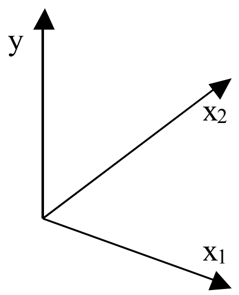
Das ist jedoch viel zu aufwendig und deswegen stellt sich die Frage, ob man solche dreidimensionalen Funktionen auch auf Papier oder auf der Tafel, also auf etwas Zwei- Dimensionalem darstellen kann. Man kann:

a. als perspektivische Zeichnung
b. in Form von Isoquanten

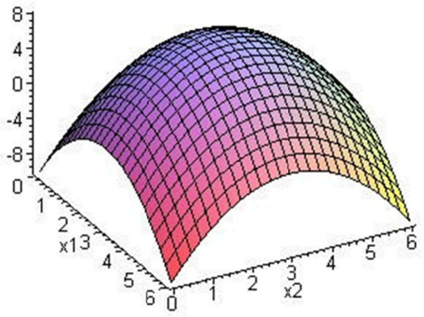
Das wollen wir uns anhand der folgenden Funktion klar machen

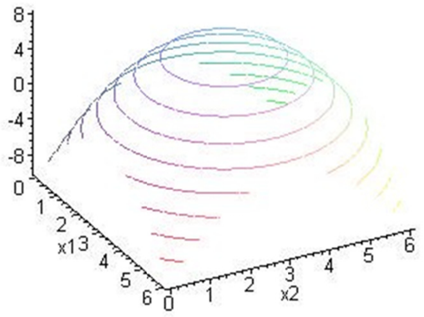
$y = -(x_1-3)^2 - (x_2-3)^2 + 8$

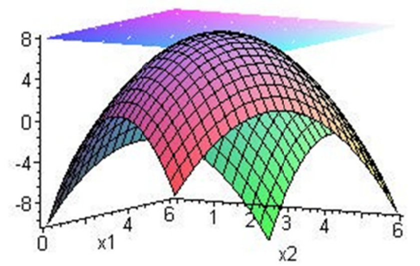
Das zugehörige Koordinatensystem hat die drei Achsen y, x_1 und x_2. Es kann auf dem zwei-dimensionalen Papier nur noch perspektivisch dargestellt werden. Üblicherweise ordnet man der senkrechten Achse die abhängige Variable y und den beiden anderen Achsen die unabhängigen Variablen x_1 und x_2 in beliebiger Reihenfolge zu:

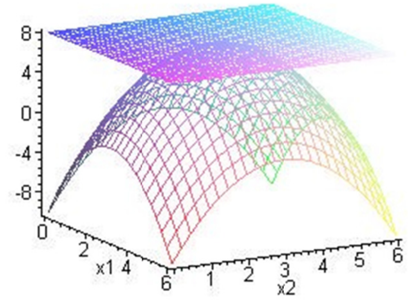


In dieses Koordinatensystem wird nun die Funktion perspektivisch eingezeichnet. Perspektivisch heißt: zweidimensional aber so, dass der Eindruck von etwas Dreidimensionalem entsteht!









Man kann den perspektivischen Eindruck durch Schnittlinien unterstützen, die alternativ zu einer der drei Achsenebenen $y=0, x_1=0$, $y=0, x_2=0$ oder $x_1=0, x_2=0$ parallel sind.

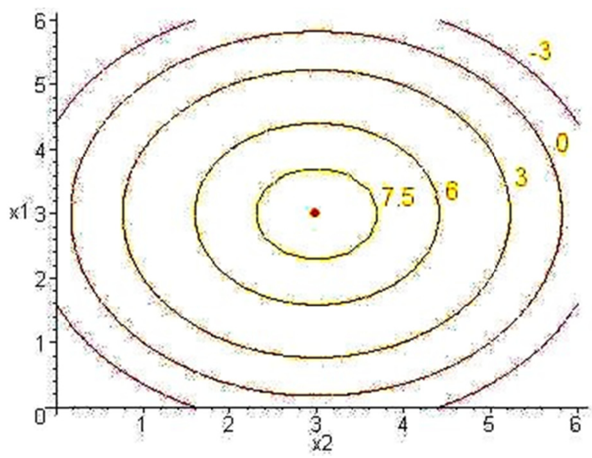
Diese perspektivischen Zeichnungen haben im Endeffekt das Ziel, im Gehirn des Betrachters die Funktion als dreidimensionales Gebilde entstehen zu lassen. Man nennt das die **dreidimensionale Anschauung.** Sie bereitet vielen Menschen Schwierigkeiten**.** Diesen Vorgang, aus einer zweidimensionalen perspektivischen Darstellung eine zutreffende dreidimensionale Vorstellung zu generieren, braucht man in vielen Lebensbereichen. In der Mathematik hilft die

dreidimensionale Vorstellung darüber hinaus, um ungefähr erahnen zu können, was sich vier- und höherdimensional abspielt.

Die zur (unteren) Ebene (x_1=0, x_2=0) parallelen Schnitte durch die dreidimensionale Funktion haben eine besondere Bedeutung: es sind Linien gleicher Werte für die abhängige Variable y, sogenannte **Isoquanten**. Solche Isoquanten gibt es in verschiedenen Lebensbereichen, z.B.

- die sog. Höhenlinien, mit denen auf den zweidimensionalen Landkarten die Täler und Höhen, also die dritte Dimension, abgebildet werden.
- die Isotimen, mit denen auf Wetterkarten die Linien gleichen Luftdrucks dargestellt und damit die Tiefdruck- und die Hochdruckgebiete kenntlich gemacht werden.

Die obige Beispielsfunktion kann in der x_1/x_2-Ebene z.B. durch folgende Isoquanten dargestellt werden.



Es stellt sich noch die Frage, ob man für duovariable Funktionen mit Excel Wertetabellen für die Isoquanten erzeugen und demnach die Isoquanten mit dem Diagramm-Assistenten geometrisch darstellen kann. Das ist dann möglich, wenn man die Gleichung nach einer der beiden unabhängigen Variablen auflösen kann. Bei der Auflösung sind die abhängige Variable y und die andere unabhängige Variable als Konstante aufzufassen.

Auch auf Grund einer solchen Isoquanten-Darstellung sollte es dem Betrachter möglich sein, im Gehirn eine ungefähre Vorstellung der zugehörigen dreidimensionalen Funktion zu erzeugen.

3. Zweidimensionale Schnitte durch multivariable Funktionen

In der Ökonomie ist die abhängige Variable zumeist eine Zielvariable, wie z.B. Gewinn, Kosten oder eine Rentabilität. Wenn eine solche Zielvariable von mehreren Variablen des Typs x_j abhängig ist, hat man zumeist keinerlei geometrische Vorstellungen über diesen Zusammenhang. Dann kann man Schnitte derart durch die Funktion legen, dass die Zielvariable nur noch von einer einzigen unabhängigen Variablen abhängig ist. So kann man sich wenigsten teilweise eine geometrische Vorstellung von der Funktion verschaffen.

Das wollen wir beispielhaft für die folgende zusammengesetzte Funktion zeigen

$$y = \begin{cases} \dfrac{x_1^2 + x_2^2 + x_1 \cdot x_2}{x_1^2 + x_2^2} & \textit{für } x_1, x_2 \neq 0 \\ 1 & \textit{für } x_1, x_2 = 0 \end{cases}$$

Insbesondere wird man sich dafür interessieren, wie die Funktion

- in der y/x_1-Ebene (x_2=0) aussieht: Es gilt durchgängig y=1
- in der y/x_2-Ebene (x_1=0) aussieht: Auch hier gilt durchgängig y=1

Man kann den Schnitt aber auch in jede beliebige Zwischenrichtung legen, und das ist das, was wir hier eigentlich vermitteln wollen.

Wir wollen den Schnitt oberhalb der Winkelhalbierenden der x_1/x_2-Ebene bilden. Das geht ganz einfach, indem wir $x_1=x_2$ setzen. Ein solches Einsetzverfahren ist jedoch für beliebige Schnitte und bei höherdimensionalen Funktionen letztendlich nicht machbar. Daher wählen wir hier ein anderes Verfahren, das für beliebige Schnitte geht und wie folgt funktioniert:

a. Vorgabe eine Punktes durch den der Schnitt gehen soll,
b. Vorgabe des Richtungsvektors, in den der Schnitt, ausgehend von dem unter a. gewählten Punkt, gelegt werden soll.

Als Ausgangspunkt können wir hier jeden Punkt auf der Winkelhalbierenden nehmen, zweckmäßigerweise wählt man hier den Punkt (0,0). Als Richtungsvektor muss man, um auf der Winkelhalbierenden zu bleiben, zwei gleiche Zahlen wählen, z.B. 1,1. Diese Richtung wird mit der (Verlängerungs-)Variablen λ multipliziert.

Dann substituiert man die einzelnen Variablen durch die jeweilige Koordinate des Anfangspunktes und die jeweilige mit der Variablen λ multiplizierte Richtung:

$$\mathrm{x}_1 = 0 + 1 \cdot \lambda$$
$$\mathrm{x}_1 = 0 + 1 \cdot \lambda$$

Eingesetzt in die Ursprungsfunktion ergibt sich

$$y = \begin{cases} \dfrac{\lambda^2 + \lambda^2 + \lambda \cdot \lambda}{\lambda^2 + \lambda^2} = 1{,}5 & \textit{für } \lambda \neq 0 \\ 1 & \textit{für } \lambda = 0 \end{cases}$$

Die Funktion hat also über der 45°-Achse der x_1/x_2-Ebene überall den Wert 1,5, nur im Nullpunkt hat sie den Wert 1. In Richtung der Achsen ist die Funktion also stetig (dort ist sie demnach

partiell differenzierbar), außerhalb der Achsen hat sie in 0,0 eine Sprungstelle und ist folglich dort nicht stetig.

Wir haben hier dieses Beispiel gewählt, um zugleich zeigen zu können, dass es bei multivariablen Funktionen den kuriosen (und wegen seiner Seltenheit nicht weiter relevanten) Fall gibt, dass eine multivariable Funktion an einzelnen Stellen partiell differenzierbar und dennoch nicht stetig sein kann.

4. Partielle Differentiation

Wie wir von den monovariablen Funktionen wissen, versteht man unter der Differentiation die Ermittlung der Steigung in jedem Punkt der Funktion. Will man dieses Konzept etwa auf eine duovariable Funktionen $y = f(x_1,x_2)$ übertragen, dann hat man insofern ein Problem, weil es für jeden Punkt der Funktion unendlich viele Richtungen gibt, für die man die Steigung bestimmen müsste.

So unlösbar dieses Problem zu sein scheint, es ist dennoch über die geometrische Anschauung lösbar. Bei den monovariablen Funktionen haben wir die Steigung dadurch ermittelt, indem wir eine tangierende Gerade (Tangente) an die Funktion gelegt haben. Das geometrische Äquivalent bei einer duovariablen Funktion ist eine **tangierende Ebene,** höherdimensional spricht man auch von einer **Hyperebene**. Das sollte man sich im Kopf vorstellen: An jedem beliebigen Punkt auf dem Graphen einer stetigen duovariablen Funktion kann man genau eine tangierende Hyperebene legen!

Algebraisch schreibt man eine solche Hyperebene allgemein wie folgt $y = a_1x_1 + a_2x_2 + a_3$

(a_1, a_2, a_3 sind Konstante)

Wir behaupten nun, dass eine solche tangentiale Hyperebene durch den Punkt auf dem Graphen und genau zwei von diesem Punkt ausgehenden tangierenden Geraden eindeutig aufgespannt wird.

Das wollen wir für unser eingangs benutztes Beispiel verifizieren: $y = -(x_1-3)^2 - (x_2-3)^2 + 8$

Als Punkt auf dem Graphen wählen wir $x_1 = 1, x_2 = 2$, dort ist $y = 3$.

Ausgehend von diesem Punkt bilden wir in der „gegriffenen" Richtung (2,1) einen Schnitt durch die Funktion. Mit der Variablen λ wird der Richtungsvektor verlängert. Dann kann man die Variablen wie folgt substituieren

$x_1 = 1+2\lambda$
$x_2 = 2+1\lambda$

Die Gleichung des Schnitts durch die Funktion lautet dann

$$y = -(1 + 2\lambda - 3)^2 - (2 + 1\lambda - 3)^2 + 8 = -(2\lambda - 2)^2 - (\lambda - 1)^2 + 8$$

Nunmehr ist die 1. Ableitung nach λ zu bilden:

$$\frac{dy}{d\lambda} = -2 \cdot (2\lambda - 2) \cdot 2 - 2 \cdot (\lambda - 1)$$

Diese Tangente hat an der Stelle $\lambda = 0$ die Steigung $8 + 2 = 10$

Wir „greifen" die weitere Richtung (-1,-1) und berechnen die Steigung des Schnitts an der Stelle (1,2):

$$x_1 = 1 - \lambda$$
$$x_2 = 2 - \lambda$$
$$y = -(1 - \lambda - 3)^2 - (2 - \lambda - 3)^2 + 8 = -(-2 - \lambda)^2 - (-1 - \lambda)^2 + 8$$
$$\frac{dy}{d\lambda} = -2(-2 - \lambda) \cdot (-1) - 2(-1 - \lambda) \cdot (-1)$$
$$\frac{dy}{d\lambda}(\lambda = 0) = -4 - 2 = -6$$

Auf den beiden Tangenten berechnen wir mit $\lambda = 1$ (dabei könnte man jedes beliebige λ vorgeben) je einen zweiten Punkt:

	Punkt	1. Richtung	Zsf Pkt u.1.Richt.	2. Richtung	Zsf Pkt u.2.Richt.
x_1	1	2	$1+2\lambda = 3$	-1	$1-1\lambda = 0$
x_2	2	1	$2+1\lambda = 3$	-1	$2-1\lambda = 1$
y	3	Steigung: 10	$3+10\lambda = 13$	Steigung: -6	$3-6\lambda = -3$

Diese drei Punkte werden nunmehr in die allgemeine Gleichung der Hyperebene eingesetzt, um die Koeffizienten a_1, a_2 und a_3 auszurechnen:

a_1	a_2	a_3	y	KdR
1	2	**1**	3	7
3	3	1	13	20
0	1	1	-3	-1

a_1	a_2	a_3	y	KdR
1	2	**1**	3	7
2	**1**	0	10	13
-1	-1	0	-6	-8

a_1	a_2	a_3	y	KdR
-3		**1**	-17	-19
2	**1**		10	13
1			4	5

a_1	a_2	a_3	y	KdR
		1	-5	-4
	1		2	3
1			4	5

Die Gleichung der tangentialen Hyperebene an den Punkten $x_1 = 1, x_2 = 2$ lautet also

$y = 4x_1 + 2x_2 - 5$

Zur Konstruktion einer tangentialen Hyperebene an einem Punkt einer duovariablen Funktion genügt also die Kenntnis des Punktes sowie die Steigung von zwei Tangenten an diesem Punkt.

Wählt man zwei andere Tangenten an dem gegebenen Punkt, dann muss sich die gleiche tangentiale Hyperebene ergeben. Das wollen wir hier zeigen, indem wir ausgehend von dem gegeben Punkt 1,2 zwei andere Richtungen wählen, und zwar diejenigen, die einerseits zu der x_1-

Achse und andererseits zu der x_2-Achse parallel sind

a. 0,1
b. 1,0

Die weiteren Berechnungen ergeben sich dann wie zuvor:

Zu a.

$x_1 = 1 + 0\lambda$
$x_2 = 2 + 1\lambda$
$y = -(x_1-3)^2 - (x_2-3)^2 + 8 = -(1-3)^2 - (2 + \lambda - 3)^2 + 8 = 4 - (\lambda - 1)^2 + 8$

$\frac{dy}{d\lambda} = -2(\lambda - 1)$ für $\lambda = 0$ ergibt sich $\frac{dy}{d\lambda} = 2$

Zu b.

$x_1 = 1 + 1\lambda$
$x_2 = 2 + 0\lambda$
$y = -(x_1-3)^2 - (x_2-3)^2 + 8 = -(1+\lambda - 3)^2 - (2-3)^2 + 8 = -(\lambda - 2)^2 - 1 + 8$
$\frac{dy}{d\lambda} = -2(\lambda - 2)$ für $\lambda = 0$ ergibt sich $\frac{dy}{d\lambda} = 4$

Bevor wir auf die Vorteile der zu den Achsen parallelen Richtungen eingehen, wollen wir zunächst zeigen, dass bei den nunmehr zugrundegelegten Richtungen die gleiche tangentiale Hyperebene entsteht.

Dazu berechnen wir zunächst den jeweils zweiten Punkt auf den Tangenten mit $\lambda = 1$.

	Punkt	1. Richtung	Zsf Pkt u.1.Richt.	2. Richtung	Zsf Pkt u.2.Richt.
x_1	1	0	$1+0\cdot\lambda = 1$	1	$1+1\cdot\lambda = 2$
x_2	2	1	$2+1\cdot\lambda = 3$	0	$2+0\cdot\lambda = 2$
y	3	Steigung: 2	$3+2\cdot\lambda = 5$	Steigung: 4	$3+4\cdot\lambda = 7$

Diese drei Punkte werden nun in die allgemeine Gleichung der Hyperebene eingesetzt, um die Koeffizienten a_1, a_2 und a_3 auszurechnen:

a_1	a_2	a_3	y	KdR
1	2	**1**	3	7
1	3	1	5	10
2	2	1	7	12

a_1	a_2	a_3	y	KdR
1	2	**1**	3	7
0	**1**	0	2	3
1	0	0	4	5

a_1	a_2	a_3	y	KdR
1	0	**1**	-1	1
0	**1**	0	2	3
1			4	5

a_1	a_2	a_3	y	KdR
0	0	**1**	-5	-4
0	**1**	0	2	3
1			4	5

Die Funktion der tangentialen Hyperebene ergibt sich also, wie behauptet, in gleicher Weise wie zuvor

$y = 4x_1 + 2x_2 - 5$

Man kann demnach ausgehend von dem gegebenen Punkt beliebige Richtungen für die anzulegenden Tangenten wählen, es ergibt sich immer die gleiche tangentiale Hyperebene. Es ist also egal, in welcher

Richtung man die Tangenten an einen Punkt anlegt, um die tangentiale Hyperebene zu konstruieren.

Diese Aussage machen wir uns zu Nutze, indem wir künftig die Tangenten nur noch parallel zu den Achsen anlegen, weil dann die Ermittlung der Steigung der Tangente besonders einfach ist.

Dieses Vorgehen bzw. dieses Verfahren wird bezeichnet als

- **partielle Differentiation.**

Das machen wir uns klar, indem wir den allgemeinen Punkt x_1, x_2 auf der Oberfläche einer duovariablen Funktion betrachten, an den wir je eine Tangente anlegen

a. einmal parallel zur x_2-Achse, also in Richtung: 0,1,
b. und einmal parallel zur x_1-Achse, also in Richtung: 1,0.

Im Weiteren betrachten wir pars pro toto nur noch den Fall a.

Die einzusetzenden Koordinaten lauten $x_1 = x_1 + 0 \cdot \lambda$

$x_2 = x_2 + 1 \cdot \lambda$

$y = -(x_1-3)^2 - (x_2 + \lambda - 3)^2 + 8$

$$\frac{dy}{d\lambda} = 2(x_2 + \lambda - 3)$$

Die Steigung an der Stelle $\lambda=0$ beträgt dann

$$\frac{dy}{d\lambda} = 2(x_2 - 3)$$

Im Weiteren betrachten wir den für x_2 eingesetzten Term

$$x_2 = x_2 + \lambda$$

Mit diesem Ausdruck wird die neue Variable λ eingeführt, nach der dann im Weiteren differenziert wird. Differenzieren heißt infinitesimal erhöhen. Wenn man λ auf der rechten Seite infinitesimal erhöht, dann erhöht man auch x_2 auf der linken Seite infinitesimal. Anstatt also nach λ zu differenzieren kann man demnach auch nach x_2 differenzieren.

Löst man die obige Gleichung nach λ auf, so ergibt sich $\lambda=0$.

Anstatt also den Term $x_2 = x_2 + \lambda$ einzusetzen, kann man λ gleich Null setzen und stattdessen nach der Variablen x_2 ableiten.

Das ist die schon zuvor erwähnte partielle Differentiation. Man leitet sukzessive nach den einzelnen Variablen ab und betrachtet dabei alle anderen Variablen als Konstante. Zum Zeichen, dass partiell abgeleitet wird, schreibt man die Differentiale nicht mehr

$$\frac{dy}{dx}$$

Sondern $\frac{\partial y}{\partial x}$ **(gesprochen: „del y nach del x")**

Die partiellen Ableitungen für die bisher behandelte duovariable Funktion

$y = -(x_1-3)^2 - (x_2-3)^2 + 8$

ergeben sich z.B. wie folgt:

$$\frac{\partial y}{\partial x_1} = -2 \cdot (x_1 - 3)$$

$$\frac{\partial y}{\partial x_2} = 2 \cdot (x_2 - 3)$$

Aus der Definition einer partiellen Differentiation, dass nur nach einer Variablen abgeleitet wird und dabei alle anderen Variablen als Konstante betrachtet werden, folgt, dass sämtliche Differentiationsregeln, die für Funktionen mit einer Veränderlichen gelten, uneingeschränkt übertragbar sind. Dabei müssen die Produkt-, Quotienten- und die Kettenregel jedoch nur angewendet werden, wenn die Variable, nach der differenziert wird, in beiden Faktoren eines Produkts, im Zähler und Nenner eines Quotienten oder in der inneren Funktion auftritt.

Insbesondere ist das Produkt aus einer Variablen, nach der partiell zu differenzieren ist, mit einer anderen Variablen kein Problem.

5. Notwendige Bedingungen für das Eintreten eine lokalen Extremums

Eines der Hauptprobleme in den Wirtschaftswissenschaften ist es, das globale Maximum, auch Optimum genannt, zu finden (bzw. das globale Minimum, etwa bei Kostenfunktionen).

Optimumverdächtig sind vor allem die lokalen Maxima (Minima).

In diesem Zusammenhang stellt sich die Frage, wie man (zunächst) bei einer duovariablen Funktion $y = f(x_1,x_2)$ das Eintreten eines lokalen Maximums erkennt. Die Antwort ergibt sich unmittelbar aus der dreidimensionalen Anschauung:

Eine duovariable Funktion hat in solchen Punkten ein lokales Maximum, wo die an diesen Punkt angelegte tangentiale Hyperebene parallel ist zur x_1,x_2-Ebene. Das gleiche gilt für lokale Minima.

Zuvor haben wir für duovariable Funktionen gezeigt, dass zwei Tangenten in beliebiger Richtung die zugehörige tangentiale Hyperebene eindeutig festlegen. Wenn also die beiden partiellen Ableitungen einer duovariablen Funktion beide den Wert Null haben, dann ist die tangentiale Hyperebene an diesem Punkt parallel zur x_1,x_2-Ebene, dann tritt dort ein lokales Maximum oder ein lokales Minimum ein.

Als dritte und letzte Möglichkeit kann an einer solchen Stelle auch ein sog. Sattelpunkt eintreten. Dieser Begriff bezieht sich tatsächlich auf diejenige Stelle auf dem Rücken eines Pferdes, wo man gemeinhin den Sattel auflegt:

- in Längsrichtung ist dieser Punkt ein lokales Minimum
- in Querrichtung ist dieser Punkt ein lokales Maximum.

Das ist sozusagen ein Sattelpunkt im engeren Sinne. Es gibt auch Fälle, bei denen in der einen Richtung ein lok. Max oder Min vorliegt, in der anderen aber ein Wendepunkt mit waagerechter Tangente, wie etwa bei der Funktion $y=x^3$. Auch solche Punkte werden (nicht ganz zutreffend) als Sattelpunkte bezeichnet.

Diese für duovariable Funktionen dargestellten Zusammenhänge gelten natürlich in gleicher Weise auch für höherdimensionale Funktionen.

Abschließend noch ein Begriff: Stellen, an denen die I. Ableitungen einer multivariablen Funktion alle gleich Null sind, werden auch als **stationäre Punkte** bezeichnet.

Zusammenfassend kann man folgendes formulieren:

Für das Eintreten eines lokalen Maximums (Minimums) ist es notwendig, dass dort alle I. partiellen Ableitungen gleich Null sind.

Oder anders herum ausgedrückt: Sind an einer Stelle einer multivariablen Funktion alle I. partiellen Ableitungen gleich Null, dann tritt dort ein

- **ein lokales Maximum, oder**
- **ein lokales Minimum, oder**
- **ein Sattelpunkt auf.**

Auf Basis der bisherigen Ableitungen wollen wir uns abschließend noch einmal mit folgendem Gedanken beschäftigen: Sind die beiden Richtungsableitungen einer duovariablen Funktion an einer Stelle gleich Null, und hat die Funktion in diesen beiden Richtungen je ein lokales Maximum, dann hat auch die duovariable Funktion dort ein Maximum.

An dieser Aussage könnte man Zweifel haben, denn die Funktion könnte ja an dieser Stelle in einer Zwischenrichtung steigend oder fallend sein. Dann läge dort weder ein lokales Maximum, noch ein lokales Minimum und auch kein Sattelpunkt vor.

So kann es aber nicht sein, weil wir gezeigt haben, dass die tangentiale Hyperebene an einem Punkt einer duovariablen Funktion stets durch zwei Tangenten beliebiger Richtung eindeutig aufgespannt wird.

Sind also die partiellen Ableitungen an einer Stelle beide gleich Null und liegt in beiden Richtungen ein lokales Maximum vor, dann kann die Funktion nicht in einer Zwischenrichtung größere Werte annehmen als in dem Punkt selbst, jedenfalls nicht mit einer Steigung, die ungleich Null ist.

6. Lösung des Systems der notwendigen Bedingungen

Sollen für eine multivariable Funktion mit n Variablen die lokalen Maxima bzw. die stationären Punkte bestimmt werden, dann sind alle n partiellen Ableitungen zu bilden und gleich Null zu setzen. Dadurch entsteht ein Gleichungssystem mit n Gleichungen und n Variablen.

Es stellt sich die Frage, ob und wie man dieses Gleichungssystem lösen kann? Dazu folgende Beispiele:

$$Z^{opt} = -x_1 + 4x_1x_2 - x_2^2 - 6x_1 - 6x_2$$

$$\frac{\partial Z}{\partial x_1} = -2x_1 + 4x_2 - 6 = 0$$

$$\frac{\partial Z}{\partial x_2} = +4x_1 - 2x_2 - 6 = 0$$

Hier ist das System der notwendigen Bedingungen ein lineares 2x2-Gleichungssystem. Lineare Gleichungssysteme lassen sich sicher mit dem sog. Gauß'schen Algorithmus lösen. Es gibt weitere Lösungsverfahren, z.B. den Algorithmus von Gauß-Seidel oder die sog. Cramer'sche Regel.

Die Lösung des Gleichungssystems lautet $x_1 = x_2 = 3$.

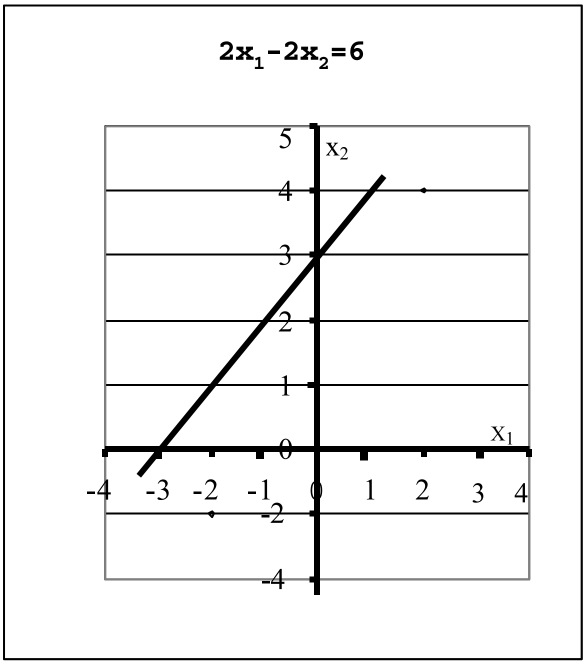
Wenn die Ursprungsfunktion ein quadratisches Polynom ist („quadratisch bedeutet hier: die höchste Potenz ist 2, maximal zwei Variable sind multiplikativ verknüpft), erhält man als notwendige Bedingungen immer ein quadratisches Gleichungssystem („quadratisch" bedeutet hier: genau so viele Variable wie Gleichungen).

$$z^{opt} = -x_1^2 + 2x_1x_2 - x_2^2 - 6x_1 + 6x_2 + 18$$

$$\frac{\partial Z}{\partial x_1} = -2x_1 + 2x_2 - 6 = 0$$

$$\frac{\partial y}{\partial x_2} = +2x_1 - 2x_2 + 6 = 0$$

Addiert man die beiden Gleichungen (=Geraden), dann erhält man 0=0 (das ist eine der beiden Komplikationen bei der Lösung linearer Gleichungssysteme). Die beiden Zeilen sind linear abhängig, in diesem Fall ist die eine Zeile das Negative der anderen. Die beiden Geraden, die durch die Gleichungen aufgespannt werden, sind einander identisch. Sie haben folgendes Aussehen



Die Ursprungsfunktion hat unendlich viele stationäre Punkte, sie liegen alle auf der Geraden x2 = x1 + 3. Man nennt das eine stationäre Kurve. Entlang dieser stationären Kurve hat die

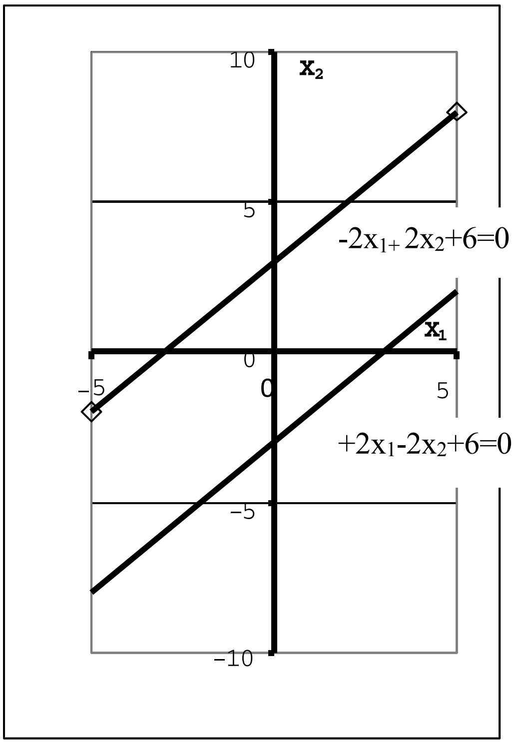
Ursprungsfunktion den Wert +27. Links und rechts der Geraden sind die Funktionswerte kleiner als +27. Die Funktion hat also entlang der Geraden ein lokales Maximum im weiteren Sinne. Man muss sich das als (waagerechten) Grat eines (unendlich langen) Bergrückens vorstellen.

$$Z^{opt} = -x_1^2 + 4x_1x_2 - x_2^2 - 6x_1 - 6x_2$$

$$\frac{\partial Z}{\partial x_1} = -2x_1 + 2x_2 - 6 = 0$$

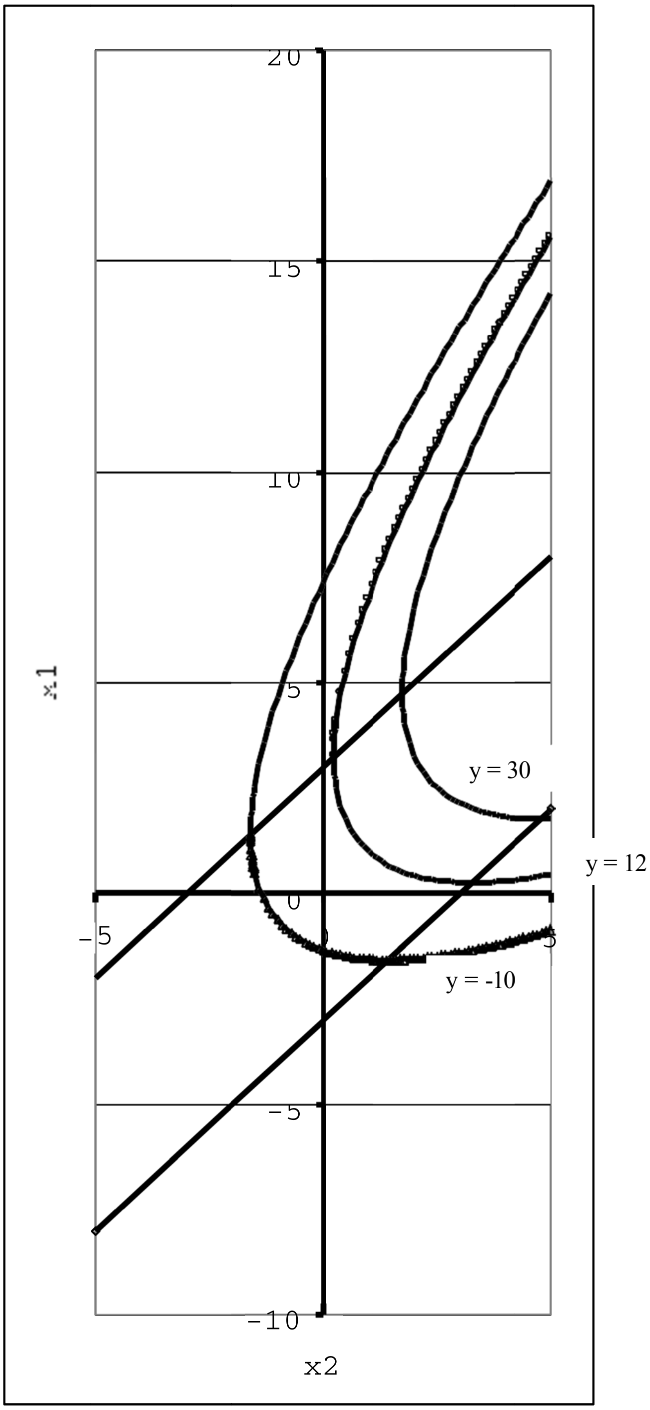
$$\frac{\partial Z}{\partial x_2} = +2x_1 - 2x_2 + 6 = 0$$

Es handelt sich ebenfalls um ein lineares Gleichungssystem, allerdings um ein widersprüchliches. Addiert man nämlich die beiden Gleichungen, so erhält man +12 = 0, was, wie gesagt, widersprüchlich ist (das ist die andere Komplikation bei der Lösung linearer Gleichungssysteme). Die beiden Geraden haben folgendes Aussehen:



Sie sind parallel zu einander, sie haben keinen Schnittpunkt. Es gibt also keinen Punkt auf der Funktion, bei der beide Ableitungen gleich Null sind. Es gibt demnach keinen stationären Punkt. Die Funktion steigt resp. fällt ständig.

Wir wollen versuchen, über diese merkwürdige Funktion eine geometrische Vorstellung zu bekommen. Dazu haben wir die Isoquanten für y = -10, y = 12 und y = 30 eingezeichnet. Es ergibt sich das folgende Bild



Es handelt sich also um einen in Richtung der 45^0-Achse steigenden Höhenrücken. Betrachtet man diesen Höhenrücken in Richtung der x_2-Achse, dann ist $x_2 = x_1 + 3$ die Kammlinie, betrachtet man den Höhenrücken in Richtung der x_1-Achse dann ist $x_2 = x_1 - 3$ die Kammlinie. Die beiden Kammlinien schneiden sich untypischer Weise nicht.

Weitere Beispiele:

$$z^{opt} = x_1^3 + x_2^3 - 3x_1 - 27x_2 + 24$$

$$\frac{\partial Z}{\partial x_1} = 3x_1^2 - 3 = 0$$

$$\frac{\partial Z}{\partial x_2} = 3x_2^2 - 27 = 0$$

Das System der notwendigen Bedingungen ist ein nichtlineares Gleichungssystem. Es gibt kein Verfahren, mit dem man nichtlineare Gleichungssysteme immer lösen kann. In diesem Fall ist das nichtlineare Gleichungssystem jedoch lösbar.

Die Lösung lautet $x_1 = \pm 1$ und $x_2 = \pm 3$. Es gibt also insgesamt 4 stationäre Punkte.

$$z^{opt} = x_1^4 - x_2^4 + 3x_1^2x_2 - x_1 + 12x_2 - 14$$

$$\frac{\partial Z}{\partial x_1} = 4x_1^3 + 6x_1x_2 - 1 = 0$$

$$\frac{\partial Z}{\partial x_2} = 3x_1^2 - 4x_2^3 + 12 = 0$$

Das System der notwendigen Bedingungen ist wiederum ein nichtlineares Gleichungssystem, diesmal ist es jedoch unlösbar, was meistens der Fall ist.

Das ist eine sehr ernste Situation, denn wenn man die stationären Punkte nicht ausrechnen kann, dann scheitert die ganze Optimierung.

Wie kann man sich aus dieser misslichen Situation befreien? Es gibt folgende Möglichkeiten:

1. Man kann versuchen, die Lösungen des nichtlinearen Gleichungssystems näherungsweise zu berechnen,
2. oder man sucht nach geeigneter Software.

Zu 1. Es gibt ein Suchverfahren, mit dem man Lösungen für ein nichtlineares Gleichungssystem finden kann, die **Methode von Newton-Raphson**. Es ist eine Erweiterung der Newton'schen Näherungsmethode zum Finden von Nullstellen für monovariable Gleichungen.

Die Methode von Newton-Raphson funktioniert wie folgt:

a. Man bringt alle Gleichungen auf die Form =0.

b. Dann ersetzt man die „0" durch eine künstliche Variable, im Weiteren mit y bezeichnet. Dadurch definiert man die Gleichungen auch dort, wo y nicht gleich Null ist (man „bläst" die Funktionen auf).

c. Im Weiteren definiert man einen „aussichtsreichen" Anfangspunkt.

d. An der Stelle dieses Anfangspunktes konstruiert man an die Funktionen tangentiale Hyperebenen, mit denen die Funktionen angenähert werden.

e. Dann sucht man den Schnittpunkt dieser tangentialen Hyperebenen in der Ebene y = 0. Das ist eine erste Näherung an einen Punkt bei dem sich die Funktionen in der

y = 0-Ebene schneiden, also eine erste Näherung an eine der Lösungen des Gleichungssystems.

f. Benutzt man den Lösungspunkt als neuen Anfangspunkt kann man sich der Lösung beliebig genau nähern, vorausgesetzt, dass das Verfahren überhaupt konvergiert, was nicht immer der Fall ist.

Dieses Verfahren ist sehr aufwendig, und vor allem garantiert es nicht, dass alle Lösungen des Gleichungssystems gefunden werden. Im Gegensatz zu linearen Gleichungssystemen ist es nämlich bei nichtlinearen Gleichungssystemen unklar, wie viel Lösungen es hat, was wir an folgenden Beispielen deutlich machen wollen

1. Zwei Kreise im zweidimensionalen Raum:
   - Sie schneiden sich nicht: keine Lösung
   - Sie tangieren sich: eine Lösung
   - Sie schneiden sich: zwei Lösungen.
2. Man kann auch unendlich viele Lösungen haben, etwa bei einer Geraden und einer Sinus-Funktion.
3. Drei Kreise im zweidimensionalen Raum: Zwei der Kreise schneiden sich in zwei Punkten. Der dritte Kreis geht durch einen dieser Punkte. Es verbleibt also nur eine Lösung. Dem entspricht algebraisch ein nichtlineares Gleichungssystem mit drei Gleichungen und zwei Variablen, das genau eine Lösung hat. Das ist eine Situation, die bei linearen Gleichungssystemen auch auftreten kann.
4. Zwei Kugeln im dreidimensionalen Raum, die sich tangieren: Dem entspricht algebraisch ein nichtlineares Gleichungssystem mit zwei Gleichungen und drei Variablen, das genau eine Lösung hat. Diese Situation kann bei linearen Gleichungssystemen nicht auftreten. Solche Lösungen könnte man übrigens mit Newton-Raphson nicht finden.

Kurz und gut: bei nichtlinearen Gleichungssystemen kann es drunter und drüber gehen!

Zu 2. Unter den zuvor beschriebenen Umständen ist es im allgemeinen besser, man lässt die Finger von der Newton-Raphson und geht auf die Suche nach geeigneter **Software**. Da gibt es so einiges. Für uns ist natürlich das interessant, was man sowieso auf seinem Rechner hat, und man hat: **In Excel gibt es unter „Extras“ den „Solver“** mit dem man ähnlich wie beim Newton- Raphson-Verfahren Lösungen für nichtlineare Gleichungssysteme suchen kann. Der Solver sucht bei Vorgabe eines Anfangspunktes eine in der Nähe des Anfangspunktes liegende Lösung. Gibt man sukzessive verschiedene Anfangspunkte vor, hat man die Chance viele und, wenn man

Glück hat, alle Lösungen zu finden.

Auf diese Software kommen wir weiter unten noch genauer zu sprechen, weil man mit ihr darüber hinaus alle (differentialrechnungsbasierten) Optimierungsprobleme lösen kann, und zwar

1. exakt, wenn die Voraussetzungen der konvexen Optimierung erfüllt sind, was meistens der Fall ist,
2. ansonsten als heuristisches Suchverfahren.

7. Partielle Ableitungen höherer Ordnung

Die Kenntnis der II. partiellen Ableitungen ist für die im Weiteren zu behandelnden hinreichenden Bedingungen unverzichtbar. Prinzipiell werden diese Ableitungen genauso gebildet wie bei monovariablen Funktionen.

Die Bildung der partiellen Ableitungen höherer Ordnung und deren Schreibweise wollen wir anhand des folgenden duovariablen Polynoms $y = f(x_1,x_2)$ kennenlernen

$$y = -4x_1^4 - x_1^2x_2^2 - 2x_2^4 + 20$$

Die I. partiellen Ableitungen lauten

$$\frac{\partial y}{\partial x_1} = -16x_1^3 - 2x_1x_2^2$$

$$\frac{\partial y}{\partial x_2} = -2x_1^2x_2 - 8x_2^3$$

Die II. partiellen Ableitungen ergeben sich dann wie folgt

$$\frac{\partial^2 y}{\partial x_1^2} = -48x_1^2 - 2x_2^2$$

$$\frac{\partial^2 y}{\partial x_2^2} = -2x_1^2 - 24x_2^2$$

Diese beiden II. partiellen Ableitungen entwickeln sich ganz normal. Das einzig Neue ist die Nomenklatur. Der Ausdruck

$$\frac{\partial^2 y}{\partial x_j^2}$$

(gesprochen: „del Quadrat y nach del x_j Quadrat“) ist kein algebraischer Ausdruck, man

kann also nicht die Wurzel ziehen, es ist einfach nur eine Bezeichnung.

Das sind jedoch noch nicht alle II. partiellen Ableitungen, man kann nämlich die I. Ableitung nach x_1 nicht nur ein zweites Mal nach x_1 sondern auch nach x_2 ableiten (sog. gemischte Ableitungen):

$$\frac{\partial^2 y}{\partial x_1 \partial x_2} = -4x_1 x_2$$

$$\frac{\partial^2 y}{\partial x_2 \partial x_1} = -4x_1 x_2$$

Bei der ersten der beiden gemischten Ableitungen haben wir zunächst nach x_1 und dann nach x_2 abgeleitet, bei der zweiten haben wir zunächst nach x_2 und dann nach x_1 abgeleitet. Das Ergebnis ist beide Male gleich. Das ist kein Zufall, das ist immer so. Bei den gemischten Ableitungen kommt es auf die Reihenfolge der Ableitungen nicht an!

Für die Bezeichnung der II. partiellen Ableitungen gibt es auch folgende Schreibweise

$$\frac{\partial^2 y}{\partial x_j^2} = f_{11} \text{ und } \frac{\partial^2 y}{\partial x_i \partial x_j} = f_{ij} = f_{ji}$$

Die III. partiellen Ableitungen mit denen wir uns im Weiteren nicht weiter befassen werden, schreibt man wie folgt:

$$\frac{\partial^3 y}{\partial x_j^3} = f_{jjj} \text{ oder } \frac{\partial^3 y}{\partial x_1 \partial x_2 \partial x_3} = f_{123} \text{ oder } \frac{\partial^3 y}{\partial x_1 \partial x_2 \partial x_2} = \frac{\partial^3 y}{\partial x_1 \partial x_2^2} = f_{122}$$

Die Ableitungen IV. und noch höherer Ordnung würde man entsprechend schreiben.

Sind die II. partiellen Ableitungen an einer Stelle negativ, so ist die Funktion dort konkav, sind sie positiv, so ist die Funktion dort konvex.

Das ist leicht gesagt. Die konkrete Berechnung, ob die II. partiellen Ableitungen an einer Stelle positiv oder negativ sind, ist schwierig. Die Details werden wir bei der Berechnung der hinreichenden Bedingungen kennenlernen.

8. Hinreichende Bedingungen, Totale Differentiale

Wenn man die stationären Punkte gefunden hat, dann stellt sich die Frage, ob an einer solchen Stelle ein lokales Minimum, ein lokales Maximum oder ein Sattelpunkt eintritt. Dieses Problem haben wir bereits für monovariable Funktionen mittels der II. Ableitungen gelöst und es stellt sich die Frage, ob wir dieses Konzept auch auf multivariable Funktionen übertragen können.

Dazu das folgende Beispiel:

$$y = -x_1^2 + 4x_1x_2 - x_2^2 - 6x_1 - 6x_2$$

Bildung des Systems der notwendigen Bedingungen:

$$\frac{\partial y}{\partial x_1} = -2x_1 + 4x_2 - 6 = 0$$
$$\frac{\partial y}{\partial x_2} = 4x_1 - 2x_2 - 6 = 0$$

Die Lösung dieses Gleichungssystems lautet: $x_1 = 3$, $x_2 = 3$

Im Weiteren wird dieser stationäre Punkt mit den II. Ableitungen daraufhin überprüft, was dort vorliegt:

$$\frac{\partial^2 y}{\partial x_1^2} = -2$$
$$\frac{\partial^2 y}{\partial x_2^2} = -2$$

Die II. Ableitungen sind beide stets (und damit auch an der Stelle 3,3) negativ, nach den monovariablen Kriterien würde man sagen: Dort tritt ein lokales Maximum ein!

Das ist aber nicht der Fall, wie wir im Weiteren zeigen werden. Dazu legen wir ausgehend vom Punkt 3,3 einen Schnitt in Richtung 1,1 durch die Funktion (also in Richtung der Winkelhalbierenden). Die zu ersetzenden Variablen ergeben sich dann wie folgt

$$x_1 = 3 + \lambda$$
$$x_2 = 3 + \lambda$$

Eingesetzt ergibt sich

$$y = -(3+\lambda)^2 + 4(3+\lambda)^2 - (3+\lambda)^2 - 12(3+\lambda) = 2(3+\lambda)^2 - 12(3+\lambda)$$

Die I. Ableitung lautet:

$$\frac{dy}{d\lambda} = 4(3+\lambda) - 12 = 0$$

Der stationäre Punkt dieser Funktion tritt ein, wie nicht anders zu erwarten, bei $\lambda = 0$, d.h. im Punkt $x_1 = x_2 = 3$.

Die II. Ableitung ergibt sich wie folgt:

$$\frac{d^2 y}{d\lambda^2} = +4$$

Nunmehr signalisiert uns die II. Ableitung, dass in Richtung der 45^0-Achse ein lokales Minimum vorliegt. Das ist eine erstaunliche Situation: In Richtung der x_1-Achse und in Richtung der x_2- Achse hat die Funktion je ein lokales Maximum, in Richtung der 45^0-Achse hat sie dagegen ein lokales Minimum!

Die Situation ist deswegen erstaunlich, weil z.B. bei der Ermittlung eines stationären Punktes zwei Tangenten in beliebiger Richtung hingelangt haben, um die tangentiale Hyperebene eindeutig zu bestimmen. Für die Analyse des stationären Punktes sind dagegen, wie dieses Beispiel zeigt, zwei Richtungen nicht hinreichend.

Wenn man also bei multivariablen Funktionen mit den II. Ableitungen analysieren will, dann braucht man eine irgendwie geartete Funktion der II. Ableitungen, die nicht nur Informationen über die II. Ableitungen in Richtungen der x_j-Achsen sondern auch Informationen über die Zwischenrichtungen enthält.

Eine solche Funktion scheint unmöglich zu sein. Es gibt sie aber doch:

Totale Differentiale!

Was totale Differentiale sind und was sie leisten, das soll im Weiteren entwickelt werden.

Totale Differentiale monovariabler Funktionen y = f(x)

Ein erster Schritt zur Entwicklung totaler Differentiale besteht darin, die Funktion in der Umgebung eines vorgegebenen Punktes x_0 durch ihre tangentiale Hyperebene anzunähern:

$$f(x) = f(x_0) + \frac{df(x_0)}{dx}(x - x_0)$$
$$f(x) - f(x_0) = \frac{df(x_0)}{dx}(x - x_0)$$

Im Weiteren werden die folgenden neuen Variablen eingeführt

$x - x_0$	= dx	diese Variable symbolisiert einen kleinen Betrag, um den die Variable x verändert wird
$f(x)-f(x_0)$	= dy	diese Variable bezeichnet die Änderung des Funktionswerts y, die sich infolge der Änderung der Variablen x um dx ergibt.

Darüber hinaus wird die Stelle x_0, an der die tangentiale Hyperebene angelegt wird, zur Variablen x verallgemeinert

$$dy = \frac{df(x)}{dx}dx = f'^{(x)} \cdot dx$$

Das ist das totale Differential I. Ordnung einer monovariablen Funktion.

Aus diesem totalen Differential I. Ordnung kann man die bereits früher entwickelten notwendigen Bedingungen herleiten:

Damit ein stationärer Punkt eintritt, darf sich der Funktionswert in einer infinitesimalen Umgebung eines Punktes x weder erhöhen noch vermindern, d.h. es muss gelten dy = 0. Dy ist dann sicher gleich Null, wenn gilt f'(x)=0. Das ist die bekannte notwendige Bedingung für das Eintreten eines stationären Punkts auf einer monovariablen Funktion: Die I. Ableitung muss gleich Null sein!

Mit dem gleichen Verfahren, mit dem man die Ursprungsfunktion mittels einer verallgemeinerten tangentialen Hyperebene annähert, kann man aus dem totalen Differential I. Ordnung durch abermaliges Annähern mittels einer verallgemeinerten tangentialen Hyperebene das totale Differential II. Ordnung bilden. Das ist der Durchbruch, denn dadurch erhält man angenähert eine Funktion der II. Ableitungen.

Wir substituieren F(dx) = dy

$$F(dx) = \frac{df(x)}{dx} = f'^{(x)} \cdot dx$$

und nähern F(dx) mit dem gleichen Verfahren an wie zuvor:

$$F(dx) = F(dx_0) + \frac{d\left(f'^{(x)}dx\right)}{dx}(dx - dx_0)$$

$$F(dx) - F(dx_0) = \frac{d\left(f'^{(x)}dx\right)}{dx}(dx - dx_0)$$

$$d^2y = f''(x) \cdot dx \cdot dx = f''(x) \cdot dx^2$$

Das ist das totale Differential II. Ordnung einer monovariablen Funktion.

Hieraus folgen die bereits bekannten hinreichenden Bedingungen für monovariable Funktionen. Bei einem lokalen Maximum (Minimum) muss die II. Ableitung, in welcher Richtung auch immer, kleiner (größer) als Null sein, es muss also gelten $d^2y < 0$ ($d^2y > 0$).

Da dx^2 immer größer/gleich Null ist, ist d^2y dann sicher kleiner (größer) Null, wenn gilt $f''(x) < 0$ ($f''(x) > 0$). Das sind die bekannten hinreichenden Bedingungen für monovariable Funktionen.

Man wird sich zu Recht fragen, warum man totale Differentiale braucht, wenn man mit ihnen genau das herleiten kann, was man sowieso schon gewusst hat. Wir haben die totalen Differentiale deswegen zunächst für monovariable Funktionen eingeführt

- einerseits, weil das auch geometrisch noch nachzuvollziehen ist,
- andererseits um zu zeigen, dass die aus den totalen Differentialen abgeleiteten Ergebnissen zu den bereits bekannten Ergebnissen für monovariable Funktionen nicht im Widerspruch stehen.

Der auf den totalen Differentialen basierende Erkenntnisfortschritt tritt aber erst für Funktionen mit zwei und mehr unabhängigen Variablen ein.

Totale Differentiale duovariabler Funktionen $y = f(x_1,x_2)$

Das totale Differential I. Ordnung entwickelt sich wie folgt

$$f(x_1, x_2) = f(x_{10}, x_{20}) + \frac{\partial f(x_{10}.x_{20})}{\partial x_1} \cdot (dx_1 - dx_{10}) + \frac{\partial f(x_{10}, x_{20})}{\partial x_2} \cdot (dx_2 - dx_{20})$$

$$F(dx_1, dx_2) = dy = \frac{\partial f(x_1, x_2)}{\partial x_1} \cdot dx_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot dx_2 = f_1 dx_1 + f_2 dx_2$$

In diesem totalen Differential I. Ordnung stecken die bereits bekannten notwendigen Bedingungen: Für das Eintreten eines lokalen Extremums ist es notwendig, dass gilt dy=0. Da dx_1 und dx_2 ganz unterschiedliche Werte annehmen können, ist dy nur dann sicher gleich Null, wenn die I. partiellen Ableitungen gleich Null sind.

Im Weiteren wird das totale Differential II. Ordnung entwickelt:

$$F(dx_{10}.dx_{20}) + \frac{\partial F}{\partial x_1}(dx_1 - dx_{10}) + \frac{\partial F}{\partial x_2}(dx_2 - dx_{20})$$

$$d^2y = \frac{\partial\left(\frac{\partial f(x_1, x_2)}{\partial x_1} \cdot dx_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot dx_2\right)}{\partial x_1} \cdot dx_1 + \frac{\partial\left(\frac{\partial f(x_1, x_2)}{\partial x_1} \cdot dx_1 + \frac{\partial f(x_1, x_2)}{\partial x_2} \cdot dx_2\right)}{\partial x_2} \cdot dx_2$$

$$d^2y = \frac{\partial^2 f}{\partial x_1^2} \cdot dx_1^2 + \frac{\partial^2 f}{\partial x_1 \partial x_2} \cdot dx_1 \cdot dx_2 + \frac{\partial^2 f}{\partial x_2 \partial x_1} \cdot dx_1 \cdot dx_2 + \frac{\partial^2 f}{dx_2^2} \cdot dx_2^2$$

$$d^2y = \frac{\partial^2 f}{\partial x_1^2} \cdot dx_1^2 + 2 \cdot \frac{\partial^2 f}{\partial x_1 \partial x_2} dx_1 dx_2 + \frac{\partial^2 f}{\partial x_2^2} dx_2^2 = f_{11} dx_1^2 + 2 \cdot f_{12} dx_1 dx_2 + f_{22} dx_2^2$$

Damit ist das totale Differential II. Ordnung für eine duovariable Funktion entwickelt. Ist es bei einem stationären Punkt negativ (positiv), dann handelt es sich um ein lokales Maximum (Minimum).

Hier können wir nun erstmals Vorteile aus dem Ansatz mit den totalen Differentialen ziehen: Das soeben abgeleitete totale Differential II. Ordnung für duovariable Funktionen zeigt, dass die hinreichenden Bedingungen für das Eintreten eines lok. Max, lok. Min oder eines Sattelpunkts bei einem stationären Punkt nicht nur von den zweiten Ableitungen nach x_1 bzw. x_2 ($f_{11}dx_1^2$ und

$f_{22}dx_2^2$) abhängig sind, sondern auch von den gemischten Ableitungen ($f_{12}dx_1dx_2$ bzw. $f_{21}dx_2dx_1$).

Damit ein lokales Maximum (Minimum) eintritt, muss gelten $d^2y < 0$ ($d^2y > 0$). Damit stellt sich die Frage, wann das soeben abgeleitete totale Differential sicher kleiner (größer) Null ist. Zwar kann man wegen der Quadrate dx_1^2 und dx_2^2 sofort sehen, dass f_{11} und f_{22} kleiner (größer) Null sein müssen, das Vorzeichen von f_{12} kann man aber nicht abschätzen, weil man nicht weiß, wie groß dx_1 und dx_2 sind. Um dieses Problem zu lösen, wird das zuvor entwickelte totale Differential II. Ordnung

$$d^2y = f_{11} dx_1^2 + 2 \cdot f_{12} dx_1 dx_2 + f_{22} dx_2^2$$

wie folgt umgeformt: Zunächst wird f_{11} ausgeklammert

$$d^2y = f_{11}\left(dx_1^2 + \frac{2 f_{12} dx_1 dx_2}{f_{11}}\right) + f_{22} dx_2^2$$

dann wird der Klammerausdruck quadratisch ergänzt

$$d^2y = f_{11}\left(dx_1 + \frac{f_{12}dx_2}{f_{11}}\right)^2 - f_{11} \cdot \frac{f_{12}^2 dx_2^2}{f_{11}^2} + f_{22}dx_2^2$$

schließlich wird noch dx_2^2 ausgeklammert

$$d^2y = f_{11}\left(dx_1 + \frac{f_{12}dx_2}{f_{11}}\right)^2 + \frac{f_{11}f_{22} - f_{12}^2}{f_{11}} \cdot dx_2^2$$

Bevor wir daran gehen, diesen Ausdruck auszuwerten und versuchen Ordnung zu schaffen, wollen wir zunächst noch das totale Differential II. Ordnung für Funktionen mit drei Variablen berechnen

$$d^2y = f_{11}dx_1^2 + f_{22}dx_2^2 + f_{33}dx_3^2 + 2f_{12}dx_1dx_3 + 2f_{13}dx_1dx_3 + 2f_{23}dx_2dx_3$$

Auch dieser Ausdruck muss zunächst so umgeformt werden, dass man das Vorzeichen sicher abschätzen kann. Eine solche Umformung führt, ohne dass wir das im einzelnen vorrechnen, zu folgendem Ergebnis

$$d^2y = f_{11}\left(dx_1 + \frac{f_{12}dx_2}{f_{11}} + \frac{f_{13}dx_3}{f_{11}}\right)^2 + \frac{f_{22}f_{11} - f_{12}^2}{f_{11}}\left(dx_2 + \frac{dx_3(f_{23}f_{11} - f_{12}f_{13})}{f_{22}f_{11} - f_{12}^2}\right)^2 + \frac{f_{11}f_{22}f_{33} + 2f_{12}f_{13}f_{23} - f_{13}^2f_{22} - f_{23}^2f_{11} - f_{12}^2f_{33}}{f_{22}f_{11} - f_{12}^2}dx_3^2$$

Nunmehr können wir die Abfolge der Bedingungen formulieren, die bei monovariablen und duovariablen Funktionen sowie bei Funktionen mit drei unabhängigen Variablen vorliegen müssen, damit bei einem stationären Punkt ein lok. Max. bzw. lok. Min. eintritt.

	$f(x)$	$f(x_1,x_2)$	$f(x_1,x_2,x_3)$
Bedingungen für ein lok. Max ($d^2y<0$)	$f_{11}<0$	$f_{11}<0$ $f_{11}f_{22}-f_{12}^2>0$	$f_{11}<0$ $f_{11}f_{22}-f_{12}^2>0$ $f_{11}f_{22}f_{33}+2f_{12}f_{13}f_{23}-f_{13}^2 f_{22}-f_{23}^2f_{11}-f_{12}^2f_{33} <0$
Bedingungen für ein lok. Min ($d^2y>0$)	$f_{11}>0$	$f_{11}>0$ $f_{11}f_{22}-f_{12}^2>0$	$f_{11}>0$ $f_{11}f_{22}-f_{12}^2>0$ $f_{11}f_{22}f_{33}+2f_{12}f_{13}f_{23}-f_{13}^2 f_{22}-f_{23}^2 f_{11}-f_{12}^2f_{33} >0$

Diese Bedingungen sehen sehr kompliziert aus und man hat den Eindruck, dass es unmöglich ist, darüber eine Ordnung zu schaffen.

So ist es aber nicht. Die formulierten Bedingungen sind die sog.

8.1 Hauptabschnittsdeterminanten der Funktional- bzw. Hesse-Matrix

Die Funktional- oder Hessematrix ist die wie folgt aufgebaute Matrix der II. Ableitungen:

	δx_1	δx_2	δx_3	...	δx_n
$\frac{\partial y}{\partial x_1}$	f_{11}	f_{12}	f_{13}		f_{1n}
$\frac{\partial y}{\partial x_2}$	f_{21}	f_{22}	f_{23}		f_{2n}
$\frac{\partial y}{\partial x_3}$	f_{31}	f_{32}	f_{33}		f_{3n}
. .					
$\frac{\partial y}{\partial x_n}$	f_{n1}	f_{n2}	f_{n3}		f_{nn}

Da es bei einer II. Ableitung egal ist, ob man zunächst nach der einen und dann nach der anderen Variablen ableitet oder umgekehrt, sind Funktionalmatrizen stets zu Hauptdiagonalen symmetrisch.

Im Weiteren wird exemplarisch gezeigt, dass die zuvor abgeleiteten hinreichenden Bedingungen den Hauptabschnittsdeterminanten der Funktionalmatrix entsprechen.

Für **monovariable Funktionen** sieht diese Matrix wie folgt aus

	δx_1
$\frac{\partial y}{\partial x_1}$	f_{11}

Die Determinante einer 1x1-Matrix ist das (einzige) Element der Matrix: f_{11}. Diese Determinante entspricht den bekannten hinreichenden Bedingungen für monovariable Funktionen:

1. II. Ableitung negativ: lok. Max
2. II. Ableitung positiv: lok. Min

Für **duovariable Funktionen** sieht die Funktionalmatrix wie folgt aus

	δx_1	δx_2
$\frac{\partial y}{\partial x_1}$	f_{11}	f_{12}
$\frac{\partial y}{\partial x_2}$	f_{21}	f_{22}

Die i'te Hauptabschnittsdeterminante (HAD) ist die Determinante der Funktionalmatrix für die ersten i Variablen.

1. i=1: Die erste HAD ist demnach f_{11}. Dem entsprechen die bekannten monovariablen Kriterien.
2. i=2: Die zweite HAD ist die Determinante der 2x2-Matrix. Die Determinante einer 2x2- Matrix besteht aus dem Produkt der Hauptdiagonalen abzüglich dem Produkt der Nebendiagonalen, es gilt also $HAD_2 = f_{11}f_{22} - f^{\ 2}$.

Diese Determinante muss immer größer Null sein, sowohl im Fall eines lok. Max. als auch

im Fall eines lok. Min.

Die Funktionalmatrix einer **Funktion mit drei Variablen** hat folgendes Aussehen:

	δx_1	δx_2	δx_3
$\frac{\partial y}{\partial x_1}$	f_{11}	f_{12}	f_{13}
$\frac{\partial y}{\partial x_2}$	f_{21}	f_{22}	f_{23}
$\frac{\partial y}{\partial x_3}$	f_{31}	f_{32}	f_{33}

Die erste HAD kennen wir bereits als monovariables Kriterium.

Die zweite HAD kennen wir von der zuvor behandelten Funktionalmatrix für duovariable Funktionen.

Die dritte HAD ist eine 3x3-Matrix. Die Determinante einer 3x3-Matrix kann man wie folgt mit der sog. **Sarrus'schen Regel** berechnen:

1. Füge der Matrix die gleiche Matrix auf der rechten Seite nochmals hinzu

f_{11}	f_{12}	f_{13}	f_{11}	f_{12}	f_{13}
f_{21}	f_{22}	f_{23}	f_{21}	f_{22}	f_{23}
f_{31}	f_{32}	f_{33}	f_{31}	f_{32}	f_{33}

2. Die Determinante berechnet sich dann wie folgt
   - Produkt der Hauptdiagonalen ($f_{11}f_{22}f_{33}$),
   - plus Produkt der darüber liegenden Diagonalen ($f_{12}f_{23}f_{31}$),
   - plus Produkt der wiederum darüber liegenden Diagonalen ($f_{13}f_{21}f_{32}$),
   - abzüglich dem Produkt der Nebendiagonalen ($f_{13}^2 f_{22}$),
   - abzüglich dem Produkt der darunter liegenden Diagonalen ($f_{23}^2 f_{11}$),
   - abzüglich dem Produkt der wiederum darunter liegenden Diagonalen ($f_{12}^2\ f_{33}$).

Zusammengefasst ergibt sich die Determinante der 3x3-Matrix wie folgt

$$f_{11}f_{22}f_{33} + 2f_{12}f_{13}f_{23} - f_{13}^2f_{22} - f_{23}^2f_{11} - f_{12}^2f_{33}$$

Das genau ist die dritte der hinreichenden Bedingungen, die bei Funktionen mit drei Variablen einzuhalten ist. Im Fall eine lok. Max muss sie negativ sein, im Fall eines lok. Min muss sie positiv sein.

Wir haben hier die hinreichenden Bedingungen bis einschließlich Funktionen mit drei Variablen abgeleitet. Sie setzen sich entsprechend fort, was man natürlich beweisen kann.

Diese Ergebnisse lassen sich in folgenden Sätzen zusammenfassen:

Sind die Hauptabschnittsdeterminanten einer Funktionalmatrix bei einem stationären Punkt im Vorzeichen alternierend, beginnend mit negativem Vorzeiche n (haben die HAD's also das Vorzeichenraster -,+,-,+, ...) so liegt ein lok. Max. vor (dann ist die Funktionalmatrix negativ definit).

Sind die Hauptabschnittsdeterminanten einer Funktionalmatrix bei einem stationären Punkt alle größer Null (haben die HAD's also das Vorzeichenraster +,+,+, ...), so liegt ein lok. Min vor (dann ist die Funktionalmatrix positiv definit).

Verstoßen die Hauptabschnittsdeterminanten einer Funktionalmatrix bei einem stationären Punkt gegen das Vorzeichenraster für lok. Max (-,+,-,+,...) und lok. Min (+,+,+,...), dann liegt ein Sattelpunkt vor (dann ist die Funktionalmatrix indefinit).

Bevor wir uns im Weiteren mit der Komplikation der sog. **semidefiniten Fälle** und darüber hinaus mit der effizienten Berechnung von Hauptabschnittsdeterminanten befassen, wollen wir uns abschließend nochmals vor Augen führen, was die I. und II. Ableitungen leisten.

Mit den I. partiellen Ableitungen kann man für einen (beliebigen) Punkt auf der Oberfläche einer Funktion sehen, wie und in welchem Ausmaß die Funktion in Richtung der Achsen steigt oder fällt. Diese Informationen sind nur für eine (infinitesimale) Umgebung des Punktes auf der Funktion richtig, darüber hinaus kommt es zu mehr oder weniger großen Abweichungen des Funktionsverlaufs von den in dem Punkt gemessenen Steigungen. Um die tangentiale Hyperebene an einen Punkt einer Funktion mit n Variablen von Typ x_j (= n-dimensionale Funktion) zu konstruieren, reicht die Kenntnis der Steigung von n an diesen Punkt angelegten Tangenten. Die Kenntnis der partiellen Ableitungen langt also hin, um die tangentiale Hyperebene zu konstruieren, denn die I. Ableitungen sind die Steigungen der Tangenten in Richtung der Achsen. Sind alle partiellen Ableitungen gleich Null, dann ist die tangentiale Hyperebene parallel zur Grundfläche. An einer solchen Stelle tritt ein stationärer Punkt ein. Mit den I. partiellen Ableitungen kann man also für einen kleinen Umkreis um einen Punkt auf der Oberfläche einer Funktion erkennen, wie sich die Funktion verhält, insbesondere kann man erkennen, ob ein stationärer Punkt vorliegt. Man kann mit den I. Ableitungen aber nicht erkennen, was bei dem stationären Punkt eintritt, ein lok. Max, ein lok. Min oder ein Sattelpunkt.

In vielen Fällen kann man mit den II. partiellen Ableitungen (=Hauptabschnittsdeterminanten der Funktionalmatrix) sehen, ob die Funktion in der Umgebung eines stationären Punktes konkav (negativ definit = lok. Max) oder konvex (positiv definit = lok. Min) oder indefinit (=Sattelpunkt) ist. Mit den II. partiellen Ableitungen kann man also einen größeren Umkreis um einen (stationären) Punkt überblicken, als mit den I. Ableitungen.

Manchmal langen die II: Ableitungen nicht hin, um zu entscheiden, was bei einem stationären Punkt eintritt, z.B. bei der Funktion $y= x_1^4 + x_2^4$. Solche Fälle könnte man mit den totalen Differentialen III. und höherer Ordnung entscheiden. Das überlassen wir den Theoretikern. Wir werden noch bessere Möglichkeiten kennenlernen, wie man lokale Maxima bzw. lokale Minima findet.

8.2 Die semidefiniten Fälle

Im Weiteren betrachten wir drei einfache duovariable Polynome

Beispiel 1	Beispiel 2	Beispiel 3
$y = -x_1^2-x_2^4$	$y = -x_1^2-x_1^2x_2^2$	$y = -x_1^2-x_2^3$
Die Funktion hat genau einen stationären Punkt bei 0,0	Die Funktion hat für $x_1=0$ eine stationäre Kurve. Der im Weiteren betrachtete Punkt 0,0 liegt auf dieser Kurve.	Die Funktion hat genau einen stationären Punkt bei 0,0
Die Funktionalmatrix lautet: $\begin{array}{\|c\|c\|}\hline -2 & 0 \\ \hline 0 & -12x_2^2 \\ \hline\end{array}$	Die Funktionalmatrix lautet: $\begin{array}{\|c\|c\|}\hline -2-2x_2^2 & -4x_1x_2 \\ \hline -4x_1x_2 & -2x_1^2 \\ \hline\end{array}$	Die Funktionalmatrix lautet: $\begin{array}{\|c\|c\|}\hline -2 & 0 \\ \hline 0 & -6x_2 \\ \hline\end{array}$

An der Stelle 0,0 hat die Funktionalmatrix in allen drei Beispielen das gleiche Aussehen:

-2	0
0	0

Die HAD's lauten: $HAD_1 = -2$ und $HAD_2 = 0$.

Da die HAD_1 das Vorzeichenraster für „negativ definit" erfüllt und die HAD_2 dem Vorzeichenraster für „negativ definit" nicht widerspricht, sollte man annehmen, dass ein lok. Max vorliegt, allenfalls vielleicht noch ein lok. Max im weiteren Sinne. Das kann der Fall sein, wie man anhand der Beispiele 1 und 2 erkennt, es kann sich aber auch um einen Sattelpunkt handeln, wie das Beispiel 3 zeigt.

Treten also im Vorzeichenraster der HAD's für lok. Max (lok. Min) Nullen auf, dann kann es sich auch um einen Sattelpunkt handeln. Diese Komplikation macht die Definition weiterer Begriffe erforderlich:

Erfüllen die HAD’s das Vorzeichenraster für negativ (positiv) definit, treten aber auch HAD’s gleich Null auf, so nennt man die Matrix negativ (positiv) semidefinit

Verstoßen die HAD’s gegen das Vorzeichenraster für negativ definit und positiv definit, treten aber auch HAD’s gleich Null auf, so ist die Matrix nach wie vor indefinit, es besteht also nicht die Notwendigkeit, den Begriff „semi-indefinit“ zu definieren.

Schließlich gibt es noch den Fall, dass alle HAD’s gleich Null sind, dafür hat sich kein Begriff etabliert (man könnte ihn vielleicht „nullfinit“ nennen). In einem solchen Fall kann alles vorliegen:

- ein lok. Max im engeren Sinne, $y = -x_1^4 - x_2^4$,
- ein lok Max im weiteren Sinne an der Stelle 0,0, $y = -x_1^2x_2^2$,
- ein lok. Min im engeren Sinne, $y = x_1^4 + x_2^4$,
- ein lok. Min im weiteren Sinne an der Stelle 0,0, $y = x_1^2x_2^2$,
- und schließlich auch ein Sattelpunkt, $y = x_1^3 + x_2^3$.

Dann gibt es noch eine Komplikation, die wir an Hand der folgenden Funktionalmatrix erläutern wollen:

	x_1	x_2	x_3
x_1	-2	0	4
x_2	0	0	0
x_3	4	0	2

Bisher haben wir bei Aufbau der Funktionalmatrizen die Variablen immer in ihrer natürlichen Reihenfolge x_1, x_2, x_3 ... enumeriert. Das ist keineswegs zwingend, man kann sie in jeder beliebigen Reihenfolge enumerieren. In diesem Fall gibt es insgesamt 6 Möglichkeiten

Reihenfolge der Enumeration			HAD 1	HAD 2	HAD 3	Analyse-Ergebnis
1	2	3	-2	0	0	negativ semidefinit
1	3	2	-2	-20	0	indefinit
2	1	3	0	0	0	„nullfinit“
2	3	1	0	0	0	„nullfinit“
3	1	2	2	-20	0	indefinit
3	2	1	2	0	0	positiv semidefinit

In den semidefiniten (und evtl. auch in den nullfiniten) Fällen, müsste man gegebenenfalls alle Anordnungen untersuchen um zu einem eindeutigen Ergebnis zu kommen.

Das ist die blanke Katastrophe, denn die Anzahl der unterschiedlichen Anordnungen beträgt n!, Die Rechenvorschrift, etwas zur Fakultät zu setzen, ist fast immer das numerische Ende.

Wir werden im Weiteren eine Möglichkeit zur Berechnung der Hauptabschnittsdeterminanten kennenlernen, mit denen man diese Probleme unterdrücken kann.

8.3 Ermittlung der Hauptabschnitts-Determinanten durch Triangulation der Funktionalmatrix

Die Determinante einer quadratischen Matrix **A** mit n Zeilen/Spalten

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{1n} \\ a_{21} & a_{22} & a_{2n} \\ \cdot & \cdot & \cdot \\ a_{nl} & a_{n2} & a_{nn} \end{pmatrix}$$

ergibt sich wie folgt

$$A = \sum_{P(\alpha,\beta,\ldots,\varpi)} (\pm)\, a_{1\alpha} \cdot a_{2\beta} \cdot \ldots \cdot a_{n\omega}$$

Diese Formel ist erklärungsbedürftig:

- P(a , $\beta,\ldots,\nu$) ist ein Ausdruck für alle Permutationen über die Komponenten des Vektors (1, 2, ... ,n).

  Die Anzahl der Permutationen beträgt n!.
- Das Vorzeichen + oder – ist wie folgt zu vergeben:

  Bestimme für die gerade betrachtete Permutation (a , β, ... ,ν) die Anzahl der Komponenten, die auf p folgen und kleiner sind als p. Dabei durchläuft p alle Komponenten der Permutation $(a, \beta, \ldots, \nu)$.

  Ist die Anzahl der Komponenten gerade, so ist das Vorzeichen des Produkts der betrachteten Permutation positiv, anderenfalls negativ.

Mit dieser Formel, die in Anbetracht der vielen Erklärungen eigentlich ein Algorithmus ist, kann man in der Tat die Determinante berechnen, was aber viel zu aufwendig ist.

Wir werden hier ein anderes, ein insbesondere für unsere Zwecke erheblich effizienteres Verfahren anwenden: die **Triangulation**.

Dieses Verfahren basiert insbesondere auf folgendem Satz

Addiert man zu einer Spalte (Zeile) ein Vielfaches einer anderen Spalte (Zeile), so ändert die Determinante ihren Wert nicht (Beckmann, Künzi: Mathematik für Ökonomen II, Heidelberg 1973, S. 51-53.

Triangulation heißt, dass man die Elemente der Matrix unterhalb der Hauptdiagonalen zu Null macht (man sagt auch, dass man insoweit die „Spalte ausräumt"). Das geschieht durch Multiplikation von Zeilen mit geeigneten Konstanten und durch Addition (bzw. Subtraktion) dieser Zeilen zu anderen Zeilen, so dass diese Nullen entstehen (dieses Vorgehen nennen wir im Weiteren auch „Gauß'sche Operation"). Durch eine Gauß'sche Operation ändert sich die Determinante wegen des obigen Satzes nicht.

Hat man die Matrix in dieser Weise trianguliert, kann man die Determinante leicht durch einmalige Anwendung der obigen Permutationsformel errechnen.

Dazu das folgende bereits triangulierte Beispiel

6	0	3	5
	4	1	0
		-3	2
			2

Wegen der Triangulation gibt es nunmehr nur noch die Permutation (1, 2, 3, 4), bei der alle Faktoren ungleich Null sind: $a_{11} \cdot a_{22} \cdot a_{33} \cdot a_{44} = 6 \cdot 4 \cdot (-3) \cdot 2 = -144$. Das Vorzeichen ist nicht zu korrigieren, weil die Indizes der Permutation aufsteigend sind.

Die Triangulation muss in folgender Reihenfolge durchgeführt werden:

- Man beginnt in der ersten Spalte und macht mit dem Element a_{11} alle darunter liegenden Elemente mit Gauß'schen Operationen zu Null
- Dann macht man das Gleiche für die zweite Spalte mit dem Element a_{22}, usw.

In der gleichen Reihenfolge würde man auch vorgehen, wenn man die einzelnen HAD's berechnet. Wenn die Matrix insgesamt trianguliert ist, kann man demnach alle HAD's der Hauptdiagonalen der triangulierten Funktionalmatrix entnehmen.

Abschließend ein Beispiel zur Berechnung der Hauptabschnittsdeterminanten mittels Triangulation:

-2	1	-2	3
1	-2	1	-1
-2	1	-4	2
3	-1	2	-7,16

-2	1	-2	3
0	**-1,5**	0	0,5
0	0	-2	-1
0	0,5	-1	-2,66

-2	1	-2	3
0	**-1,5**	0	0,5
0		**-2**	-1
0		-1	-2,5*

* gerundet

-2	1	-2	3
0	**-1,5**	0	0,5
0		**-2**	-1
0		0	**-2**

Die Hauptabschnittsdeterminanten lauten HAD_1= -2, HAD_2 = +3, HAD_3 = -6, HAD_4 = 12. Die Matrix ist demnach negativ definit.

Generell gilt: Die Funktionalmatrix ist vor Beginn der Triangulation symmetrisch. Nach der ersten Iteration ist sie dann noch ab der 2. Zeile und Spalte symmetrisch, usw. Das lässt sich auch allgemein zeigen.

Bei jeder Berechnung „per Hand" stellt sich die Frage nach der Rechenrichtigkeit. Eine Triangulation kann man mit der aus dem Gauß'schen Algorithmus bekannten **KdR**-Regel absichern (KdR = **Kontrolle der Rechenrichtigkeit**). Es fehlt allerdings die Möglichkeit, das Ergebnis, wie etwa bei einem linearen Gleichungssystem, in das Ausgangssystem einzusetzen und so definitiv die Richtigkeit des Ergebnisses zu konstatieren. Die Triangulation einer Funktionalmatrix ist also wegen des letztendlich nicht zu beseitigenden Rechenfehlerrisikos gefährlich.

Nicht zuletzt deswegen muss man froh sein, dass Excel zur Berechnung von Determinanten die Funktion MDET anbietet.

8.4 Triangulation bei Nullen auf der Hauptdiagonale

Die Triangulation ist problematisch, wenn dasjenige **Element auf der Hauptdiagonalen gleich Null** ist, mit dem die Spalte ausgeräumt werden soll. Dann kann man nämlich die darunter liegenden Elemente durch Gauß'sche Operationen nicht zu Null machen. Es gibt zwei Möglichkeiten, wie man sich helfen kann.

1. Durch eine **Zeilen- bzw. Spaltenvertauschung**. Das kann man bis auf das Vorzeichen folgenlos machen, wie folgender Satz zeigt:

Vertauscht man zwei Zeilen (Spalten) der Matrix, so ändert sich das Vorzeichen der Determinante

Man sucht also in der Zeile (Spalte), in der die Hauptdiagonale die störende Null hat, rechts (unterhalb) von der Null ein Element $\neq 0$ und vertauscht die beiden Spalten (Zeilen).

2. Durch eine **Zeilen- bzw. Spaltenaddition**. Das basiert auf dem weiter oben schon eingeführten Satz, dass man zu einer Spalte (Zeile) ein Vielfaches einer anderen Spalte (Zeile) addieren kann, ohne dass sich die Determinante ändert.

Die Vorgehensweise ist der unter 1. beschriebenen sehr ähnlich: Man sucht nach einer Zeile (Spalte), die unterhalb des Elements (rechts neben dem Element) auf der Hauptdiagonalen, das gleich Null ist, ein Element ungleich Null hat und addiert (oder subtrahiert) diese Zeile (Spalte) in die Zeile (Spalte) mit der Null in der Hauptdiagonalen.

Nach einer solchen Zeilen- bzw. Spaltenvertauschung oder nach einer Zeilen- bzw. Spaltenaddition kann man weiter triangulieren, weil nunmehr das zum Ausräumen benötigte Element auf der Hauptdiagonalen ungleich Null ist.

Wir haben es hier mit Funktionalmatrizen zu tun. Mit der Triangulation wollen wir nicht nur die Determinante der Gesamtmatrix sondern wir wollen alle HAD's von der Hauptdiagonalen

„abpflücken" können. Wenn man eine Zeilenvertauschung macht, dann ändert man die Reihenfolge der Variablen in den Zeilen. Das ist für die Ermittlung der HAD's „tödlich", weil man dann in den Zeilen eine andere Reihenfolge der Variablen hat als in den Spalten. Wenn man also eine Zeilenvertauschung macht, dann muss man auch die entsprechenden Spalten vertauschen. Wir sprechen von einer symmetrischen Zeilen/Spalten-Vertauschung. Dann hat man die gleiche Reihenfolge der Variablen sowohl in den Zeilen als auch in den Spalten. Die Folge der HAD's kann ohne Einschränkungen berechnet werden.

Entsprechendes gilt bei der Zeilen- bzw. Spaltenaddition. Hier hat man aber das zusätzliche Problem, dass man die Variablen additiv verknüpft. Dadurch wird die Folge der HAD's zwischen den Variablen, die addiert worden sind, irreparabel gestört. Die Zeilen/Spalten-Addition ist also für Funktionalmatrizen nicht geeignet, um Nullen auf der Hauptdiagonalen zu „reparieren".

Nullen auf der Hauptdiagonale von Funktionalmatrizen

Das alles brauchen wir aber nicht weiter zu vertiefen, weil wir beim Auftreten einer Null auf der Hauptdiagonalen einer Funktionalmatrix sofort erkennen können, was Sache ist.

Um das zu zeigen, unterstellen wir, dass nach der j'ten Triangulations-Iteration (also im Feld j+1, j+1) eine Null auf der Hauptdiagonalen eintritt. Im Weiteren ist zwischen zwei Fällen zu unterscheiden:

1. Rechts von der Null auf der Hauptdiagonalen befinden sich nur Nullen. Dann müssen wegen der Symmetrie der Matrix auch unterhalb der Null auf der Hauptdiagonalen nur Nullen sein

	1	**...**	**j**	**j+1**	**j+2**
1	a_{11}		a_{1j}	a_{1j+1}	a_{1j+2}
...		...			
j	0		a_{jj}	a_{jj+1}	a_{jj+2}
j+1				**0**	**0**
j+2				**0**	**b**

(den Koeffizienten $a_{j+2,j+2}$ haben wir, um das Schreiben der Indices zu vermeiden, mit b bezeichnet). Man sieht sofort, dass man sowohl mit einer Zeilen/Spalten-Vertauschung als auch mit einer Zeilen/Spalten-Addition kein Element $\neq 0$ in das Feld j+1, j+1 bringen kann. Die Funktionalmatrix ist sicher semi-definit. Man kann die betreffende Spalte und die betreffende Zeile nach hinten schieben, um die weitere Triangulation fortsetzen zu können.

2. Rechts von der Null auf der Hauptdiagonalen befindet sich mindestens ein Element $\neq 0$,

im Weiteren vereinfachend als a bezeichnet. Dieses Element kann durch eine Spaltenvertauschung unmittelbar rechts neben die Null in der Hauptdiagonalen geschoben werden. Wegen der Symmetrie der Matrix muss es das Element mit dem Wert a auch unterhalb der Null in der Hauptdiagonalen geben. Es kann durch eine symmetrische Zeilenvertauschung unmittelbar unterhalb der Null geschoben werden. Die Matrix hat dann das folgende Aussehen

	1	**...**	**j**	**j+1**	**j+2**
1	a_{11}		a_{1j}	a_{1j+1}	a_{1j+2}
...		...			
j	0		a_{jj}	a_{jj+1}	a_{jj+2}
j+1				**0**	**a**
j+2				**a**	**b**

Im Weiteren schieben wir den Wert b durch eine symmetrische Zeilen/Spalten- Vertauschung in das Feld j+1, j+1:

	1	**...**	**j**	**j+1**	**j+2**
1	a_{11}		a_{1j}	a_{1j+1}	a_{1j+2}
...		...			
j	0		a_{jj}	a_{jj+1}	a_{jj+2}
j+1				**b**	**a**
j+2				**a**	**0**

Jetzt kann mit dem Element b ausgeräumt werden. Dann erhält man

	1	**...**	**j**	**j+1**	**j+2**
1	a_{11}		a_{1j}	a_{1j+1}	a_{1j+2}
...		...			
j	0		a_{jj}	a_{jj+1}	a_{jj+2}
j+1				**b**	**a**
j+2				**0**	$\mathbf{-a^2/b}$

Durch die zuvor ausgeführten Triangulationen kann man alle HAD's bis einschließlich der HAD_j berechnen. Insofern gibt es eine Fallunterscheidung: HAD_j kann kleiner oder größer Null sein.

Des Weiteren gibt es eine Fallunterscheidung bezüglich b: Es kann kleiner oder größer Null sein.

Für diese Fallunterscheidungen kann man nun die Folge der Vorzeichen der HAD's berechnen

		HAD_j	HAD_{j+1}	HAD_{j+2}
$HAD_j>0$	b>0	+	+	-
	b<0	+	-	-
$HAD_j<0$	b>0	-	-	+
	b<0	-	+	+

Keine dieser Folgen erfüllt das Vorzeichenraster für negativ bzw. positiv definit. Die Funktionalmatrix ist indefinit.

Das führt zu folgendem Satz:

Tritt beim Triangulieren einer Funktionalmatrix auf der Hauptdiagonalen eine Null ein, dann ist zwischen zwei Fällen zu unterscheiden:

1. **Sind auch alle anderen Elemente rechts (und unterhalb) dieser Null gleich Null, dann empfiehlt es sich, die betreffende Zeile und Spalte ans Ende der Matrix zu schieben. Die letzte HAD ist dann gleich Null.**
2. **Gibt es dagegen rechts (und unterhalb) von der Null auf der Hauptdiagonalen ein Element ungleich Null, dann ist die Funktionalmatrix sicher indefinit.**

Die Null auf der Hauptdiagonalen braucht demnach nicht durch eine Zeilen/Spalten-Vertauschung oder eine Zeilen/Spalten-Addition beseitigt zu werden, um die Triangulation fortsetzen zu können.

Abschließend wollen wir noch dasjenige Beispiel triangulieren, bei dem wir weiter oben für unterschiedliche Reihenfolge der Variablen unterschiedliche Ergebnisse erhalten haben:

-2	0	4
0	0	0
4	0	2

-2	0	4
0	0	0
0	0	10

Man müsste die Triangulation mit dem Element 2,2 fortsetzen. Es ist gleich Null. Die zugehörige Zeile und auch die Spalte enthalten nur Nullen. Wir können sie daher ans Ende der Matrix verschieben

-2	4	0
0	10	0
0	0	0

Die HAD's lauten: $HAD_1 = -2$, $HAD_2 = -20$, $HAD_3 = 0$.

Die Triangulation führt unmittelbar zu dem richtigen Ergebnis: die Matrix ist indefinit.

Man kann die stationären Punkte auch mit sog. **Eigenwerten** überprüfen, was wir hier aber nicht weiter behandeln, weil dieser Weg ineffizient ist.