Bachelorstudiengänge Wirtschaftswissenschaften und Informatik

Mathematik II

Prof. Dr. D. Klatte

Universität Zürich

Vorlesungsmaterialien Frühjahrssemester 2009

Version 22.05.09

Inhaltsverzeichnis

Die Kapitel 1-6 wurden im Fach **Mathematik I** behandelt.
Die Kapitel 7-11 sind als Skript im Studentenladen zu beziehen.

12. Spezielle Kapitel der linearen Algebra

12.1. Verbrauchsmatrizen (Input-Output-Modell) [SH §16.9]

Wir betrachten eine Volkswirtschaft mit n Produktionszweigen, in der der Output jedes Zweiges als Produktionsfaktor (Input) im eigenen wie in den anderen Zweigen eingesetzt und verbraucht werden kann.

Beispiel: Volkswirtschaft mit $n = 3$ Produktionszweigen

En	Energieproduktion,
Ch	Chemieproduktion,
Ba	Bauwesen

und einer *Verbrauchsmatrix* A

$$\begin{pmatrix} \text{En-Verbrauch} \\ \text{Ch-Verbrauch} \\ \text{Ba-Verbrauch} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.1 \end{pmatrix}}_{=:\,A} \begin{pmatrix} \text{Input für En} \\ \text{Input für Ch} \\ \text{Input für Ba} \end{pmatrix}.$$

Z.B. betrifft die erste Zeile den Energieverbrauch: Der **Verbrauch**

zur Produktion von 1 Einheit in En beträgt 0.2 En-Einheiten
zur Produktion von 1 Einheit in Ch beträgt 0.4 En-Einheiten
zur Produktion von 1 Einheit in Ba beträgt 0.3 En-Einheiten.

Nennen wir den *Produktionsvektor (Inputvektor)*

$$\underline{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} \qquad \begin{array}{l} \textit{Energie} \\ \textit{Chemieprodukte} \\ \textit{Produkte des Bauwesens} \end{array}$$

so berechnet sich der *verfügbare Output* π_1 an Energie durch

$$\pi_1 = p_1 - (0.2p_1 + 0.4p_2 + 0.3p_3),$$

analog die verfügbaren Outputs π_2, π_3 aus Chemie und Bauwesen, insgesamt

$$\begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} - \begin{pmatrix} 0.2 & 0.4 & 0.3 \\ 0.3 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.1 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

das heisst

$$\underline{\pi} = \underline{p} - A\underline{p}.$$

In diesem Modell wird berücksichtigt, dass ein Teil des Inputs $\underline{p}$, nämlich $A\underline{p}$, während der Produktion verbraucht wird.

Mit I bezeichnen wir wieder die Einheitsmatrix (hier: vom Typ $(3,3)$).

Sei nun ein *Nachfragevektor*

$$\underline{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \qquad \begin{array}{l} \mathit{Energie} \\ \mathit{Chemieprodukte} \\ \mathit{Produkte\ des\ Bauwesens} \end{array}$$

gegeben mit $\underline{y} \geq \underline{o}$. Gesucht ist in dem Modell

$$\underline{p} - A\underline{p} = \underline{y}, \quad \text{d.h., } (I - A)\underline{p} = \underline{y}, \tag{1}$$

ein geeigneter Input $\underline{p} \geq \underline{o}$, so dass die Nachfrage $\underline{y}$ befriedigt werden kann (*"produktive Volkswirtschaft"*).

Literaturhinweis: Simon-Blume *Mathematics for Economists* §8.5

Mathematisch ergibt sich also ein lineares Gleichungssystem

$$(I - A)\underline{p} = \underline{y} \tag{2}$$

mit bekannter rechter Seite $\underline{y}$ und gesuchtem Vektor $\underline{p}$.

Weiss man, dass $I - A$ invertierbar ist, dann folgt sofort, dass

$$\underline{p} = (I - A)^{-1}\underline{y}$$

zu jedem $\underline{y}$ die eindeutige Lösung des LGS (2) ist.

Wann ist in diesem Falle bei nichtnegativem Nachfragevektor $\underline{y}$ auch der Lösungsvektor $\underline{p}$ nichtnegativ?

Antwort: Sicher ist das so, wenn die Matrix $(I - A)^{-1}$ nur nichtnegative Elemente enthält.

Mathematica liefert als Inverse von $I - A$ in unserem Beispiel die Matrix (sie enthält nur positive Elemente!)

$$(I - A)^{-1} = \frac{1}{29} \begin{pmatrix} 96 & 90 & 52 \\ 70 & 120 & 50 \\ 66 & 80 & 72 \end{pmatrix}.$$

Also ist in unserem Beispiel erfüllt, dass die Lösungsvektoren

$$\underline{p} = (I - A)^{-1}\underline{y} \qquad \text{des LGS } (I - A)\underline{p} = \underline{y}$$

nichtnegativ sind, wenn die Nachfragevektoren $\underline{y}$ nichtnegativ sind.

Gibt es da eine allgemeine Gesetzmässigkeit? Ja!

Satz: Sei A eine (n,n)-Matrix mit nichtnegativen Elementen a_{ij}, so dass alle Zeilensummen s_i die Eigenschaft

$$\boxed{s_i := \sum_{j=1}^{n} a_{ij} < 1} \qquad (i = 1, \ldots, n)$$

haben. Dann existiert $(I - A)^{-1}$ (wobei I die (n,n)-Einheitmatrix ist), und es gilt

$$(I - A)^{-1} = I + A + A^2 + \ldots + A^m + \ldots$$

Folglich hat auch $(I - A)^{-1}$ nur nichtnegative Elemente.

Wegen der Analogie zur Summenformel der geometrischen Reihe

$$1/(1 - q) = 1 + q + q^2 + \ldots + q^m + \ldots \qquad \text{für } -1 < q < 1.$$

spricht deshalb auch von der **geometrischen Reihe für Matrizen** (auch *Neumannsche Reihe* genannt).

Beweis:

Wir betrachten die Partialsummen S_m und das Produkt AS_m

$$\begin{aligned} S_m &= I + A + A^2 + \ldots + A^m \\ AS_m &= \phantom{I + {}} A + A^2 + \ldots + A^m + A^{m+1} \end{aligned}$$

und bilden die Differenz $S_m - AS_m$. Als Ergebnis erhalten wir

$$(I - A)S_m = I - A^{m+1}. \tag{3}$$

Man sagt, dass eine Folge von (n, n)-Matrizen $\{S_m = (s_{ij}^{(m)})\}$ gegen die (n, n)-Matrix $S = (s_{ij})$ *konvergiert*, geschrieben $S := \lim_{m\to\infty} S_m$, wenn für jedes Indexpaar (i, j)

$$\lim_{m\to\infty} s_{ij}^{(m)} = s_{ij}$$

gilt. Analog versteht man die Konvergenz der Folge $\{A, A^2, \ldots, A^m, \ldots\}$.

Da alle Zeilensummen kleiner als 1 sind, konvergiert die Folge der Matrizen

$$\{A, A^2, \ldots, A^m, A^{m+1}, \ldots\} \qquad \text{für } m \to \infty$$

gegen die Nullmatrix.

Machen Sie sich das in unserem Beispiel klar! (Computerexperiment oder Nachdenken!) Zusatzfrage: Was passiert, wenn nicht $s_i < 1$ gilt?

Damit gilt aber nach (3) und den Rechengesetzen für Limites

$$I = \lim_{m\to\infty}(I - A^{m+1}) = \lim_{m\to\infty}(I - A)S_m = (I - A)S,$$

wobei $S := \lim_{m\to\infty} S_m$ ist (die Existenz des Limes ist gesichert *).

Nach Definition der Inversen folgt damit, dass S die Inverse zu $(I - A)$ und folglich gleich $I + A + A^2 + \ldots + A^m + \ldots$ ist!!! q.e.d.

*Beweis unter Ausnutzung der Konvergenz der geometrischen (Zahlen-) Reihe

12.2. Lineare Algebra und Differentialrechnung im $\mathbb{R}^n$

12.2.1. Optimierung unter linearen Restriktionen

[Ro II §7.3], [SH §14.5]*

Beispiel 12.1: *Abstand eines Punktes von einer Geraden im Raum*

Gegeben sei eine Gerade G, definiert durch

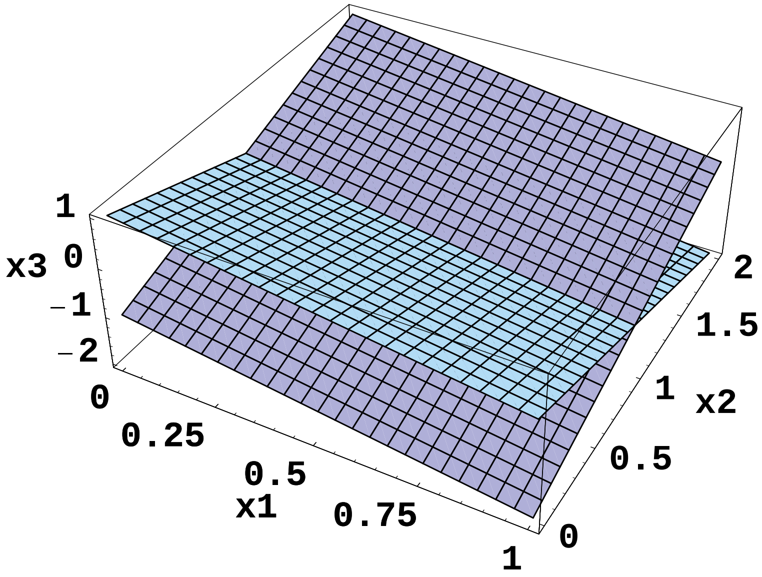
$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ -x_1 + x_2 - x_3 &= 1 \end{aligned} \quad \text{d.h.} \quad \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, \quad t \in \mathbb{R}.$$

Man bestimme denjenigen Punkt $\underline{x}^* = (x_1^*, x_2^*, x_3^*)^\mathsf{T}$ auf G, der den kleinsten euklidischen Abstand zum Koordinatenursprung hat, also

$$x_1{}^2 + x_2{}^2 + x_3{}^2 \;\rightarrow\; \text{min} \quad \text{bezüglich } (x_1, x_2, x_3)^\mathsf{T} \in G. \qquad (*)$$

Hinweis: die euklidische Norm des Vektors (x_1, x_2, x_3) ist die Wurzel aus $x_1{}^2 + x_2{}^2 + x_3{}^2$ - das liefert den gleichen Lösungspunkt.

*Ro II = Rommelfanger, Band 2, SH = Sydsaeter/Hammond



Durch *Einsetzen der Parameterform* erhält man

$$x_1{}^2 + x_2{}^2 + x_3{}^2 = (0-t)^2 + (1+0\cdot t)^2 + (0+t)^2 = 2t^2+1,$$

also ist die Aufgabe

$$2t^2+1 \;\rightarrow\; \min \qquad \text{bezüglich } t \in \mathbb{R}$$

zu lösen (die eindeutige Lösung ist offenbar $t^* = 0$), also ist

$$\underline{x}^* = (-t^*, 1, t^*)^\mathsf{T} = (0,1,0)^\mathsf{T} \text{ eindeutige Lösung von } (*).$$

Alternativ kann man die *Lagrangemethode* benutzen, man setzt

$$L(x_1, x_2, x_3, \lambda_1, \lambda_2) := {x_1}^2 + {x_2}^2 + {x_3}^2 - \lambda_1(x_1 + x_2 + x_3 - 1) - \lambda_2(-x_1 + x_2 - x_3 - 1)$$

und löst mit den partiellen Ableitungen L_{x_i}, L_{λ_j} das Lagrangesystem

$$\begin{aligned} L_{x_i}(x_1, x_2, x_3, \lambda_1, \lambda_2) &= 0 \qquad (i = 1, 2, 3) \\ -L_{\lambda_j}(x_1, x_2, x_3, \lambda_1, \lambda_2) &= 0 \qquad (j = 1, 2). \end{aligned}$$

Das Lagrangesystem ist in diesem Falle ein inhomogenes LGS, nämlich

$$\begin{array}{rrrrrcl} 2x_1 & & & -\lambda_1 & +\lambda_2 & = & 0 \\ & 2x_2 & & -\lambda_1 & -\lambda_2 & = & 0 \\ & & 2x_3 & -\lambda_1 & +\lambda_2 & = & 0 \\ x_1 & +x_2 & +x_3 & & & = & 1 \\ -x_1 & +x_2 & -x_3 & & & = & 1 \end{array}$$

Wie man leicht ausrechnet, ist die einzige Lösung dieses LGS

$$(x_1^*, x_2^*, x_3^*, \lambda_1^*, \lambda_2^*) = (0, 1, 0, 1, 1).$$

Im allgemeinen hat die Lagrangemethode den Nachteil, dass sie nur *stationäre Punkte* $\underline{x}^*$ bestimmt, aber in unserem Beispiel liefert der geometrische Hintergrund, dass $\underline{x}^*$ globaler Minimalpunkt ist.

Beispiel 12.2: *Wohlfahrtsanalyse* (vgl. Varian "Mikroökonomie" §13)

In einem Markt mit einem x-Gut und einem y-Gut habe man drei Konsumenten K_i $(i = 1, 2, 3)$ und zwei Unternehmen U_j $(j = 1, 2)$, die das x-Gut produzieren. Das y-Gut stehe "für alles andere", was konsumiert wird.

Jedem Konsumenten K_i $(i = 1, 2, 3)$ seien zugeordnet:

x_i	durch K_i konsumierte Menge des x-Gutes
y_i	durch K_i konsumierte Menge des y-Gutes
ω_i	Anfangsausstattung von K_i mit dem y-Gut
$u_i(x_i) + y_i$	Nutzenfunktion des Konsumenten K_i

Jedem Unternehmen U_j $(j = 1, 2)$ seien zugeordnet:

z_j	produzierte Menge des x-Gutes in U_j
$c_j\, z_j$	Kosten zur Produktion von z_j in U_j

Wohlfahrtsmaximierung: Gesucht ist die *Allokation*

$$(x_1^*, x_2^*, x_3^*, y_1^*, y_2^*, y_3^*, z_1^*, z_2^*) \in \mathbb{R}^8,$$

die die Nutzensumme maximiert unter der Bedingung, dass die insgesamt produzierte und konsumierte Menge des x-Gutes gleich sind. Ferner nimmt man an, dass der Verbrauch des y-Gutes in U_j gleich den Kosten $c_j\, z_j$ ist, womit die zweite Gleichung sinnvoll wird:

$$\text{Maximiere} \quad \sum_{i=1}^{3} u_i(x_i) + \sum_{i=1}^{3} y_i$$

$$\text{unter den Nebenbedingungen} \quad \begin{aligned} \sum_{i=1}^{3} x_i &= \sum_{j=1}^{2} z_j \\ \sum_{i=1}^{3} y_i &= \sum_{i=1}^{3} \omega_i - \sum_{j=1}^{2} c_j\, z_j \end{aligned}$$

bezüglich der Variablen $x_1, x_2, x_3, y_1, y_2, y_3, z_1, z_2$.

Laut Modell sind c_j, ω_i und die Funktionen $u_i(\cdot)$ gegeben.

Hier empfiehlt sich natürlich gleich die Substitution

$$\sum_{i=1}^{3} y_i = \sum_{i=1}^{3} \omega_i - \sum_{j=1}^{2} c_j \, z_j \, ,$$

so dass sich nur noch ein Problem in den Variablen x_i und z_j ergibt:

$$\text{Maximiere} \quad \sum_{i=1}^{3} u_i(x_i) + \sum_{i=1}^{3} \omega_i - \sum_{j=1}^{2} c_j \, z_j$$

$$\text{unter der Nebenbedingung} \quad \sum_{i=1}^{3} x_i = \sum_{j=1}^{2} z_j \, .$$

Beim Lösen der Optimierungsaufgabe kann man die Konstante $\sum_{i=1}^{3} \omega_i$ zunächst weglassen, das ergibt die Lagrangefunktion

$$L(x_1, x_2, x_3, z_1, z_2, \lambda) := \sum_{i=1}^{3} u_i(x_i) - \sum_{j=1}^{2} c_j \, z_j - \lambda \left(\sum_{i=1}^{3} x_i - \sum_{j=1}^{2} z_j \right) .$$

Übung: Die Funktionen $u_i(\cdot)$ seien stetig differenzierbar. Man gebe die Lagrangebedingungen an und zeige für die Lösung $(x_1^*, x_2^*, x_3^*, z_1^*, z_2^*)$ mit zugehörigem λ^*, dass $\lambda^* = u_i'(x_i^*) = c_j \;\; (\forall i, j)$ gilt.

Definition:
(Optimierungprobleme unter linearen Gleichungsrestriktionen)

Gegeben sind eine nach allen Variablen stetig partiell differenzierbare Funktion

$$(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \mapsto f(x_1, x_2, \ldots, x_n) \in \mathbb{R}$$

sowie $m \cdot n + m$ reelle Zahlen a_{ij} und b_i für $i = 1, \ldots, m$, $j = 1, \ldots, n$. Wir betrachten das Optimierungsproblem (PMIN/LIN)

Minimiere $f(x_1, x_2, \ldots, x_n)$

unter den m linearen Gleichungsrestriktionen

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \ldots & + & a_{mn}x_n & = & b_m \end{array}$$

Analog bezeichnen wir mit (PMAX/LIN) das entsprechende Maximierungsproblem.

Sei $f_{x_j}(x_1, x_2, \ldots, x_n)$ die partielle Ableitung von f bezüglich der Variablen x_j im Punkt $\underline{x} = (x_1, x_2, \ldots, x_n)^{\mathsf{T}}$.

Satz: (Notwendige Optimalitätsbedingung nach Lagrange)

Sei $\underline{x}^* = (x_1^*, x_2^*, \ldots, x_n^*)^\mathsf{T}$ ein lokaler Lösungspunkt von (PMIN/LIN) oder (PMAX/LIN). Dann existiert ein Vektor $\underline{\lambda}^* = (\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^*)^\mathsf{T}$, so dass

$$\begin{aligned} f_{x_j}(x_1^*, x_2^*, \ldots, x_n^*) - \sum_{i=1}^{m} \lambda_i^* \, a_{ij} &= 0, \;\; j = 1, 2, \ldots, n. && (*) \\ \sum_{j=1}^{n} a_{ij} \, x_j^* &= b_i, \;\; i = 1, 2, \ldots, m. && (**) \end{aligned}$$

Das entspricht den Bedingungen

$$L_{x_j}(\underline{x}^*, \underline{\lambda}^*) = 0, \;\; L_{\lambda_i}(\underline{x}^*, \underline{\lambda}^*) = 0 \quad (i = 1, \ldots, m, \;\; j = 1, \ldots, n),$$

wobei L wieder die übliche *Lagrangefunktion*

$$L(\underline{x}, \underline{\lambda}) := f(\underline{x}) - \sum\nolimits_{i=1}^{m} \lambda_i (\sum\nolimits_{j=1}^{n} a_{ij} \, x_j - b_i)$$

ist mit $\underline{x} = (x_1, x_2, \ldots, x_n)^\mathsf{T}$ und $\underline{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)^\mathsf{T}$.

Ist f eine quadratische Funktion, so ist $(*)$, $(**)$ ein LGS!

(PMIN/LIN) bzw. (PMAX/LIN) in *Vektor-Matrix-Schreibweise*:

$$\text{Minimiere/Maximiere} \quad f(\underline{x})$$

$$\text{unter der Restriktion} \quad A\underline{x} = \underline{b}\,,$$

dabei werden hier und im folgenden alle Vektoren als Spaltenvektoren aufgefasst. Die Lagrangefunktion hat dann die Form

$$L(\underline{x}, \underline{\lambda}) = f(\underline{x}) - \underline{\lambda}^{\mathsf{T}}(A\underline{x} - \underline{b}).$$

Die Lagrangebedingungen lauten dann

$$\operatorname{grad} f(\underline{x}^*) - A^{\mathsf{T}}\underline{\lambda}^* = 0, \quad A\underline{x}^* = \underline{b}.$$

Substitutionsmethode (am Beispiel (PMIN/LIN)):

Die allgemeine Lösung des LGS $A\underline{x} = \underline{b}$ lautet in Vektorschreibweise (falls die Lösungsmenge nichtleer ist):

$$\underline{x} = \underline{x}^0 + \sum_{k=1}^{r} t_k\, \underline{x}^k, \qquad t_1, \ldots, t_r \in \mathbb{R},$$

wobei $\underline{x}^0$ eine spezielle Lösung des LGS $A\underline{x} = \underline{b}$ ist und die Vektoren $\underline{x}^1, \ldots, \underline{x}^r$ eine Basis der Lösungsmenge von $A\underline{x} = \underline{o}$ bilden.

Satz über die Substitutionsmethode für (PMIN/LIN):

Das LGS $A\underline{x} = \underline{b}$ sei lösbar und habe die allgemeine Lösung

$$\underline{x} = \underline{x}^0 + \sum_{k=1}^{r} t_k \, \underline{x}^k, \qquad t_1, \ldots, t_r \in \mathbb{R}.$$

Ist $\underline{t} = (t_1, t_2, \ldots, t_r)^\mathsf{T}$ und ist X die Matrix mit den Spaltenvektoren $\underline{x}^1, \ldots, \underline{x}^r$, so schreibt sich das $\underline{x} = \underline{x}^0 + X\underline{t}$. Wir definieren

$$h(\underline{t}) := f(\underline{x}^0 + X\underline{t}), \;\; \underline{t} \in \mathbb{R}^r,$$

und betrachten eine Optimierungsaufgabe ohne Restriktionen:

$$(\widetilde{P}) \qquad \text{min} \quad h(\underline{t}) \;\; \text{bezüglich } \underline{t} \in \mathbb{R}^r.$$

Dann gilt:

$\underline{x}^* = \underline{x}^0 + X\underline{t}^*$ ist lokale (bzw. globale) Minimalstelle von (PMIN/LIN)

genau dann, wenn

$\underline{t}^*$ lokale (bzw. globale) Minimalstelle von $(\widetilde{P})$ ist.

12.2.2. Semi-definite Matrizen und Optimalitätsbedingungen 2. Ordnung bei freien Extremalaufgaben [Ro II, §6.1, §7.2]

Wir betrachten die Optimierungsaufgabe(n) ohne Nebenbedingungen

$$\text{Minimiere/Maximiere } f(x_1, \ldots, x_n),$$

wobei f stetige partielle Ableitungen 2. Ordnung $f_{x_i x_j}(x_1, \ldots, x_n)$ [damit auch stetige partielle Ableitungen 1. Ordnung $f_{x_j}(x_1, \ldots, x_n)$] bezüglich aller Variablen haben möge.

Da es keine Nebenbedingungen gibt, spricht man davon, es seien **freie Extremalaufgaben**. Wir schreiben wieder $\underline{x} = (x_1, \ldots, x_n)^\mathsf{T}$.

Für einen Punkt $\underline{x}^0 \in \mathbb{R}^n$ sagt man, f habe in $\underline{x}^0$ ein globales Minimum bzw. $\underline{x}^0$ sei ein globaler Minimalpunkt von f, falls

$$f(\underline{x}^0) \leq f(\underline{x}) \quad \text{für alle } \underline{x} \in \mathbb{R}^n. \tag{4}$$

Ersetzt man in (4) "$\leq$" durch "$\geq$", so heisst $\underline{x}^0$ globaler Maximalpunkt von f, und man sagt, f habe ein globales Maximum in $\underline{x}^0$.

Relative Extrema: Man sagt, f habe in $\underline{x}^0$ ein relatives Minimum bzw. $\underline{x}^0$ sei ein relativer Minimalpunkt von f, falls ein $\varepsilon > 0$ existiert, so dass

$$f(\underline{x}^0) \leq f(\underline{x}) \quad \text{für alle } \underline{x} \text{ mit } \|\underline{x} - \underline{x}^0\| < \varepsilon. \tag{5}$$

Ersetzt man in (5) "$\leq$" durch "$\geq$", so heisst $\underline{x}^0$ relativer Maximalpunkt von f, und man sagt, f habe ein relatives Maximum in $\underline{x}^0$.

Sammelname für relative Minima und Maxima: relative Extrema.

Ersetzt man in (5) "$\leq$" durch "$<$" bzw. "$>$", so spricht man von einem strikten relativen Extremum.

Wie im Falle $n = 2$, den wir aus der *Mathematik I* kennen, beweist man sehr einfach den folgenden Satz.

Satz: Notwendige Bedingungen erster Ordnung [Ro II Satz 7.2]

Falls f in $\underline{x}^0$ ein relatives Extremum hat, so gilt

$$\operatorname{grad} f(\underline{x}^0) := (f_{x_1}(\underline{x}^0), \ldots, f_{x_n}(\underline{x}^0))^{\mathsf{T}} = \underline{o}, \tag{6}$$

das heisst, der Gradient grad $f(\underline{x}^0)$ von f im Punkt $\underline{x}^0$ ist gleich dem Nullvektor. Begriff: Ein Punkt $\underline{x}^0$ mit (6) heisst stationärer Punkt.

Natürlich braucht man hier nur vorauszusetzen, dass f partielle Ableitungen 1. Ordnung hat, denn wenn f in $\underline{x}^0$ ein relatives Extremum hat, dann hat auch jede Funktion

$$x_j \mapsto f(x_1^0, \ldots, x_{j-1}^0, x_j, x_{j+1}^0, \ldots, x_n^0)$$

in $\underline{x}^0$ ein relatives Extremum als Funktionen in einer Variablen – dafür ist als notwendige Bedingung wohlbekannt: $f_{x_j}(\underline{x}^0) = 0$.

Zur Anwendung des Satzes vgl. auch Abschnitt 12.3.

Repetition: Hinreichende Bedingungen 2. Ordnung für n=2

Hat $f : \mathbb{R}^2 \to \mathbb{R}$ stetige zweite partielle Ableitungen nach x_1 und x_2 und ist $\underline{x}^0$ ein stationärer Punkt von f, dann gelten mit der

$$\text{Hesse-Matrix } H(\underline{x}^0) = (f_{x_i x_j}(\underline{x}^0)\,;\ 1 \leq i,j \leq 2)$$

(sie ist symmetrisch!) die folgenden Aussagen:

1. Falls $\det H(\underline{x}^0) > 0$ und $f_{x_1 x_1}(\underline{x}^0) > 0$, so hat f in $\underline{x}^0$ ein striktes relatives Minimum.
2. Falls $\det H(\underline{x}^0) > 0$ und $f_{x_1 x_1}(\underline{x}^0) < 0$, so hat f in $\underline{x}^0$ ein striktes relatives Maximum.
3. Falls $\det H(\underline{x}^0) < 0$, so besitzt f in $\underline{x}^0$ kein relatives Extremum ($\underline{x}^0$ heisst dann Sattelpunkt von f).

4 Falls $\det H(\underline{x}^0) = 0$ gilt, kann noch keine Aussage über das Vorliegen eines relativen Extremums gemacht werden.

Beim Beweis der Aussagen 1 und 2 in der Vorlesung *Mathematik I* hatten wir davon Gebrauch gemacht, dass folgende Äquivalenzen gelten

$$
\begin{aligned}
&\left(\det H(\underline{x}^0) > 0 \text{ und } f_{x_1x_1}(\underline{x}^0) > 0 \right) \\
\Leftrightarrow\ & \underline{v}^\top H(\underline{x}^0)\underline{v} > 0\ \forall \underline{v} \neq \underline{o} \\
\\
&\left(\det H(\underline{x}^0) > 0 \text{ und } f_{x_1x_1}(\underline{x}^0) < 0 \right) \\
\Leftrightarrow\ & \underline{v}^\top H(\underline{x}^0)\underline{v} < 0\ \forall \underline{v} \neq \underline{o}.
\end{aligned}
\tag{7}
$$

Verallgemeinerung

Sie wird mit Beweis im Fach Mathematik III - Analysis für Ökonomen behandelt. Einen kurzen Beweis der folgenden Aussage 1' – und damit auch 2' – findet man z.B. auch im Lehrbuch F. Jarre, J. Stoer, *Optimierung*, Springer 2004, Satz 6.0.3.

Hat $f : \mathbb{R}^n \to \mathbb{R}$ stetige zweite partielle Ableitungen nach allen n Variablen, so bezeichnen wir wieder für $\underline{x}^0 = (x_1^0, \ldots, x_n^0)^\top$ den Gradienten von f in $\underline{x}^0$ mit $\operatorname{grad} f(\underline{x}^0)$ und die Hesse-Matrix (sie ist wieder symmetrisch!) von f in $\underline{x}^0$ mit

$$H(\underline{x}^0) := (f_{x_i x_j}(\underline{x}^0)\,;\; 1 \leq i, j \leq n).$$

Die Aussagen 1, 2 oben gelten dann in der äquivalenten Form (7), Aussage 3 gilt in einer modifizierten Form:

Satz. [Ro II Satz 7.6]

Sei wieder $\underline{x}^0$ stationärer Punkt von f, d.h., grad $f(\underline{x}^0) = \underline{o}$.

1' Falls $\underline{v}^\top H(\underline{x}^0)\underline{v} > 0 \ \forall \underline{v} \neq \underline{o}$, so hat f in $\underline{x}^0$ ein striktes relatives Minimum.

2' Falls $\underline{v}^\top H(\underline{x}^0)\underline{v} < 0 \ \forall \underline{v} \neq \underline{o}$, so hat f in $\underline{x}^0$ ein striktes relatives Maximum.

3' Falls sowohl $\underline{u}^\top H(\underline{x}^0)\underline{u} > 0$ für mindestens ein $\underline{u} \neq \underline{o}$ als auch $\underline{v}^\top H(\underline{x}^0)\underline{v} < 0$ für mindestens ein $\underline{v} \neq \underline{o}$ gilt, so besitzt f in $\underline{x}^0$ kein relatives Extremum (d.h., $\underline{x}^0$ ist ein Sattelpunkt).

4' Falls nur $\underline{v}^\top H(\underline{x}^0)\underline{v} \geq 0 \ \forall \underline{v}$ bzw. $\underline{v}^\top H(\underline{x}^0)\underline{v} \leq 0 \ \forall \underline{v}$ gilt, kann noch keine Aussage über das Vorliegen eines relativen Extremums gemacht werden.

Man spricht bei den Kriterien in den Aussagen 1 und 2 (bzw. 1' und 2') von positiver bzw. negativer Definitheit,

bei Aussage 3 (bzw. 3') von Indefinitheit

der symmetrischen Matrix $H(\underline{x}^0)$ bzw. der betreffenden quadratischen Form $\underline{v} \mapsto \underline{v}^\top H(\underline{x}^0)\underline{v}$.

Gelten die Bedingungen in 4', so heisst die Matrix $H(\underline{x}^0)$ positiv bzw. negativ *semi*-definit.

Diese Begriffe werden wir nun in der Sprache der linearen Algebra für beliebige symmetrische n-reihige Matrizen formulieren.

Definition. Eine n-reihige symmetrische Matrix A heisst

- **positiv definit**, wenn $\underline{v}^\mathsf{T} A \underline{v} > 0$ für alle $\underline{v} \in \mathbb{R}^n \setminus \{\underline{o}\}$ gilt;
- **positiv semidefinit**, wenn $\underline{v}^\mathsf{T} A \underline{v} \geq 0$ für alle $\underline{v} \in \mathbb{R}^n$ gilt;
- **negativ definit**, wenn $\underline{v}^\mathsf{T} A \underline{v} < 0$ für alle $\underline{v} \in \mathbb{R}^n \setminus \{\underline{o}\}$ gilt;
- **negativ semidefinit**, wenn $\underline{v}^\mathsf{T} A \underline{v} \leq 0$ für alle $\underline{v} \in \mathbb{R}^n$ gilt;
- **indefinit**, wenn keine dieser vier Eigenschaften gilt.

Übung: Man bestimme die symmetrische Matrix A im Falle $n = 2$, falls $\underline{v}^\mathsf{T} A \underline{v} > 0$ mit $\underline{v} = (v_1, v_2)^\mathsf{T}$ z.B. die folgende Gestalt hat:

(a) $v_1^2 + v_2^2$ $\quad$ (b) $4v_1^2 - 4v_1v_2 + 2v_2^2$ $\quad$ (positiv definit)

(c) $-v_1^2 - v_2^2$ $\quad$ (d) $-v_1^2 - 2v_1v_2 - 2v_2^2$ $\quad$ (negativ definit)

(e) $(v_1 - 2v_2)^2$ $\quad$ (f) $-v_1^2 + 2v_1v_2 - v_2^2$ $\quad$ (pos. bzw. neg. semidefinit)

Diskussion der Begriffe an einem Beispiel mit ökonomischem Bezug.*

Beispiel. Eine Unternehmung XYZ produziere 3 Güter mit den Outputmengen x_1, x_2, x_3 (in Mengeneinheiten ME1, ME2, ME3) nach der Gesamtkostenfunktion

$$K(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 3x_3^2 + x_1x_2 + x_2x_3 + 100.$$

Die Marktpreise der Güter seien exogen vorgegeben mit

$$p_1 = 40 \text{ CHF/ME1}, \; p_2 = 50 \text{ CHF/ME2}, \; p_3 = 80 \text{ CHF/ME3}.$$

Bei welchem Outputvektor (x_1, x_2, x_3) operiert XYZ gewinnmaximal?

*J. Tietze, *Einf. in die angewandte Wirtschaftsmathematik*, Vieweg, 1995, S. 7-47.

Lösung: Gegeben ist

$$K(x_1, x_2, x_3) = x_1^2 + 2x_2^2 + 3x_3^2 + x_1 x_2 + x_2 x_3 + 100.$$

Die Gewinnfunktion lautet

$$G(x_1, x_2, x_3) = 40x_1 + 50x_2 + 80x_3 - K(x_1, x_2, x_3).$$

Die notwendigen Optimalitätsbedingungen 1. Ordnung lauten

$$\begin{array}{rcrrrrcl} G_{x_1} & = & -2x_1 & -x_2 & & +40 & = & 0 \\ G_{x_2} & = & -x_1 & -4x_2 & -x_3 & +50 & = & 0 \\ G_{x_3} & = & & -x_2 & -6x_3 & +80 & = & 0 \end{array}$$

Das ist ein LGS mit der eindeutigen Lösung (Nachrechnen!)

$$\underline{x}^0 := (x_1^0, x_2^0, x_3^0) = (17.5\,,\,5\,,\,12.5).$$

Handelt es sich wirklich um ein relatives Maximum? Ist es sogar ein globales Maximum?

Die Hesse-Matrix $H(\underline{x})$ der Funktion G lautet

$$A := H(\underline{x}^0) = \begin{pmatrix} -2 & -1 & 0 \\ -1 & -4 & -1 \\ 0 & -1 & -6 \end{pmatrix}.$$

Sie ist nicht vom Punkt $\underline{x}^0$ abhängig. Mit $\underline{v} := (v_1, v_2, v_3)^\mathsf{T}$ rechnet man leicht aus

$$\begin{aligned} \underline{v}^\mathsf{T} A \underline{v} &= -2v_1^2 - 4v_2^2 - 6v_3^2 - 2v_1v_2 - 2v_2v_3 \\ &= -v_1^2 - 2v_2^2 - 5v_3^2 - (v_1 + v_2)^2 - (v_2 + v_3)^2 \\ &< 0 \qquad (\text{falls } \underline{v} \neq \underline{o}). \end{aligned}$$

Damit ist die Matrix A negativ definit und folglich $\underline{x}^0$ ein (strikter) relativer Maximalpunkt.

Es handelt sich sogar um ein globales Maximum. Man wende das folgende Theorem an (vgl. Mathematik III): Ist die Hesse-Matrix $H(\underline{x})$ einer Funktion $f : \mathbb{R}^n \to \mathbb{R}$ negativ semidefinit für alle $\underline{x}$, so ist f konkav und somit jedes relative Maximum auch globales Maximum.

Ein praktisches Kriterium für positive bzw. negative Definitheit: Hauptabschnittsdeterminanten

Entsteht $A^{[k]}$ aus einer n-reihigen symmetrischen Matrix $A = (a_{ij})$ durch Streichen sowohl der letzten k Zeilen als auch der letzten k Spalten von A, dann heisst

$$\det A^{[k]} \quad \text{eine Hauptabschnittsdeterminante von } A.$$

Alle Teilmatrizen $A^{[k]}$ sind wieder symmetrisch. Insbesondere gilt also

$$\det A = \det A^{[0]} \text{ und } a_{11} = \det A^{[n-1]}.$$

Zum Beispiel sind für eine (3×3)-Matrix $A = (a_{ij})$ die Zahlen

$$a_{11}, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

die Hauptabschnittsdeterminanten von A.

In der Repetition waren die hinreichenden Optimalitätskriterien für Aufgaben mit 2 Variablen mit Hauptabschnittsdeterminanten formuliert.

Satz. Sei A eine n-reihige symmetrische Matrix. Dann ist

A positiv definit genau dann, wenn
alle Hauptabschnittsdeterminanten von A positiv sind;

A negativ definit genau dann, wenn
die Hauptabschnittsdeterminanten $\det A^{[k]}$ von A
positiv im Falle gerader Ordnung von $A^{[k]}$,
aber negativ im Falle ungerader Ordnung von $A^{[k]}$ sind.

Zum Beispiel ist eine symmetrische (3×3)-Matrix $A = (a_{ij})$ negativ definit, falls

$$a_{11} < 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \qquad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0$$

gilt. Eine symmetrische (2×2)-Matrix $A = (a_{ij})$ ist negativ definit, falls

$$a_{11} < 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0\,.$$

Anwendung der Hauptabschnittsdeterminanten auf das Beispiel

Wir betrachten die Determinante der Matrix $A = H(\underline{x}^0)$ des Beispiels:

$$\det A = \begin{vmatrix} -2 & -1 & 0 \\ -1 & -4 & -1 \\ 0 & -1 & -6 \end{vmatrix}.$$

Dann gilt $\det A^{[2]} = -2 < 0$,

$$\begin{aligned} \det A^{[1]} &= \begin{vmatrix} -2 & -1 \\ -1 & -4 \end{vmatrix} = 8 - 1 = 7 > 0, \\ \det A^{[0]} = \det A &= \begin{vmatrix} -2 & -1 & 0 \\ -1 & -4 & -1 \\ 0 & -1 & -6 \end{vmatrix} \\ &= \begin{vmatrix} 0 & 7 & 2 \\ -1 & -4 & -1 \\ 0 & -1 & -6 \end{vmatrix} = -(-1) \begin{vmatrix} 7 & 2 \\ -1 & -6 \end{vmatrix} = -40 < 0. \end{aligned}$$

Damit ist A negativ definit – wie bereits oben überprüft.

Notwendiges Optimalitätskriterium 2. Ordnung

In der Ökonomie (vgl. z.B. Varian, *Mikroökonomie*) wird auch oft das folgende notwendige Optimalitätskriterium benutzt:

Habe $f : \mathbb{R}^n \to \mathbb{R}$ stetige zweite partielle Ableitungen nach allen Variablen. Dann gelten folgende Kriterien:

1. Hat f in $\underline{x}^0$ ein relatives Minimum, so ist

$$\operatorname{grad} f(\underline{x}^0) = \underline{o} \text{ und } H(\underline{x}^0) \text{ ist positiv semidefinit.}$$

2. Hat f in $\underline{x}^0$ ein relatives Maximum, so ist

$$\operatorname{grad} f(\underline{x}^0) = \underline{o} \text{ und } H(\underline{x}^0) \text{ ist negativ semidefinit.}$$

Wieder ist also die Semidefinitheit wichtig. Zum Beweis vgl. z.B. F. Jarre, J. Stoer, *Optimierung*, Springer 2004, Satz 6.0.2.

Man mache sich die Kriterien z.B. an folgenden Funktionen klar: $f(x, y, z) = (x + y + z)^2$ und $g(x, y, z) = -(x - 2y + 3z)^2$.

12.2.3 Hinreichende Bedingungen 2. Ordnung der Optimierung unter einer Gleichungsnebenbedingung (optional)

[Ro II §7.4], [SH §14.4]

Wir betrachten die Optimierungsaufgabe

$$\text{(P)} \qquad \text{Minimiere } f(x_1, x_2) \quad \text{bezüglich} \quad g(x_1, x_2) = 0,$$

wobei f und g zweimal stetig differenzierbar auf $\mathbb{R}^2$ sind.

Lagrangefunktion der Aufgabe (P):

$$L(x_1, x_2, \lambda) := f(x_1, x_2) - \lambda g(x_1, x_2)$$

Hessematrix von L bezüglich (x_1, x_2) im Punkt (x_1^0, x_2^0, λ), λ fest:

$$H := \begin{pmatrix} f_{x_1x_1}(x_1^0, x_2^0) - \lambda g_{x_1x_1}(x_1^0, x_2^0) & f_{x_1x_2}(x_1^0, x_2^0) - \lambda g_{x_1x_2}(x_1^0, x_2^0) \\ f_{x_1x_2}(x_1^0, x_2^0) - \lambda g_{x_1x_2}(x_1^0, x_2^0) & f_{x_2x_2}(x_1^0, x_2^0) - \lambda g_{x_2x_2}(x_1^0, x_2^0) \end{pmatrix}$$

Der Nullpunkt im $\mathbb{R}^2$ sei mit $\underline{o}$ bezeichnet.

Satz: Hinreichende Bedingung 2. Ordnung für lokales Minimum

Seien in (x_1^0, x_2^0, λ) die Lagrange-Bedingungen der Aufgabe (P) erfüllt und bezeichne $\operatorname{grad} g(x_1^0, x_2^0)$ wieder den Gradienten von g in (x_1^0, x_2^0).

Falls $\underline{y}^\mathsf{T} H \underline{y} > 0$ für alle $\underline{y} = (y_1, y_2)^\mathsf{T} \neq \underline{o}$ mit

$$\operatorname{grad} g(x_1^0, x_2^0)^\mathsf{T} \underline{y} = 0 \tag{8}$$

gilt, so ist (x_1^0, x_2^0) strikter lokaler Minimalpunkt von (P), d.h., es gilt

$$f(x_1, x_2) > f(x_1^0, x_2^0)$$

für alle Punkte (x_1, x_2) "nahe" (x_1^0, x_2^0), die $g(x_1, x_2) = 0$ erfüllen.

Im Falle strikter lokaler Maxima dreht sich das Zeichen $>$ jeweils um.

Definitheit unter linearen Nebenbedingungen:

Die hinreichende Bedingung für lokale Minima ist von der Form

$$\underline{y}^\mathsf{T} H \underline{y} > 0 \quad \forall \underline{y} \neq \underline{o}: \ \underline{c}^\mathsf{T} \underline{y} = 0$$

mit einem Vektor $\underline{c}^\mathsf{T} = (c_1, c_2)$ und symmetrischer $(2,2)$-Matrix H, denn H ist Hesse-Matrix mit stetigen zweiten partiellen Ableitungen.

Die hinreichende Bedingung für lokale Maxima ist von der Form

$$\underline{y}^\mathsf{T} H \underline{y} < 0 \quad \forall \underline{y} \neq \underline{o}: \ \underline{c}^\mathsf{T} \underline{y} = 0$$

Man sagt im ersten Falle, dass die Matrix H *positiv definit* - im zweiten Falle *negativ definit* - *auf der Geraden* $\{\underline{y} \in \mathbb{R}^2 \,|\, \underline{c}^\mathsf{T} \underline{y} = 0\}$ ist.

Beispiel: Man betrachte die Aufgabe

$$\begin{array}{ll} \text{min} & f(x_1,x_2) := \frac{1}{2}(x_1^2 - x_2^2) - x_1 \\ \text{bezüglich} & g(x_1,x_2) := -x_1 + 2x_2 = 0 \end{array} \tag{9}$$

Lösung durch Substitution: Es gilt $x_1 = 2x_2$, eingesetzt in die Zielfunktion ergibt sich

$$\varphi(x_2) := f(2x_2, x_2) = \frac{1}{2}(4x_2^2 - x_2^2) - 2x_2 = \frac{3}{2}x_2^2 - 2x_2.$$

Folglich

$$\varphi'(x_2) = 3x_2 - 2 = 0 \;\Leftrightarrow\; x_2 = \frac{2}{3}, \quad \text{somit } x_1 = \frac{4}{3}.$$

Die hinreichende Bedingung 2. Ordnung für freie Extrema lautet

$$\varphi''(\tfrac{2}{3}) = 3 > 0,$$

d.h., $\underline{x}^0 = (\frac{4}{3}, \frac{2}{3})$ ist strikter lokaler Minimalpunkt.

Die Lagrange-Funktion ist im Beispiel gleich

$$L(x_1, x_2, \lambda) = \frac{1}{2}(x_1^2 - x_2^2) - x_1 - \lambda(-x_1 + 2x_2)$$

Die Lagrange-Bedingungen lauten

$$\begin{array}{rrrrcl} (i) & x_1 & & +\lambda & = & 1 \\ (ii) & & -x_2 & -2\lambda & = & 0 \\ (iii) & -x_1 & +2x_2 & & = & 0 \end{array}$$

Die Lösung des Systems der Lagrange-Bedingungen ist

$$x_1^0 = \frac{4}{3}, \quad x_2^0 = \frac{2}{3}, \quad \lambda = -\frac{1}{3}.$$

Die Bedingung (8) ist hier sehr einfach:

$$-y_1 + 2y_2 = 0, \quad \text{d.h. } y_2 = \frac{1}{2}y_1, \tag{10}$$

da grad $g(x_1, x_2) = (-1, 2)^\mathsf{T}$ für alle (x_1, x_2).

Die Hesse-Matrix von L bezüglich (x_1, x_2) in (x_1^0, x_2^0, λ) lautet hier

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \qquad \text{(Hesse-Matrix).}$$

Sei nun $\underline{y} = (y_1, y_2)^\mathsf{T} \neq \underline{o}$ ein beliebiger Vektor mit (10). Dann gilt

$$\underline{y}^\mathsf{T} H \underline{y} = (y_1, y_2) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = y_1^2 - y_2^2$$

und nach Einsetzen von (10)

$$\underline{y}^\mathsf{T} H \underline{y} = y_1^2 - y_2^2 = y_1^2 - \frac{1}{4}y_1^2 = \frac{3}{4}y_2^2 > 0 \quad \forall \underline{y} \neq \underline{o} \ \text{ mit } y_2 = \frac{1}{2}y_1.$$

Folglich ist $\underline{x}^0 = (\frac{4}{3}, \frac{2}{3})$ strikter lokaler Minimalpunkt.

Spezialfall eines Kriteriums von Debreu

In der ökon. Literatur sind von Gerard Debreu (1983 Nobelpreis für Ökonomie) bewiesene Kriterien zum Checken der positiven/negativen Definitheit einer symmetrischen Matrix H unter linearen Nebenbedingungen

$$B\underline{y} = \underline{o}$$

(B muss dabei eine Matrix passender Ordnung sein) sehr beliebt, die mit Determinanten/Unterdeterminanten einer "geränderten" Matrix

$$\begin{pmatrix} O & B \\ B^\top & H \end{pmatrix}$$

arbeiten, vgl. am besten die Originalarbeit *G. Debreu, Definite and semidefinite quadratic forms, Econometrica 20 (1952) 295-300*.

Wir betrachten hier nur den Spezialfall für die Aufgabe (P).

Sei (x_1^0, x_2^0, λ) ein Punkt, der den Lagrange-Bedingungen von (P) genügt, und sei $\operatorname{grad} g(x_1^0, x_2^0) \neq \underline{o}$.

Sei wieder H die Hesse-Matrix der Lagrange-Funktion L bezüglich (x_1, x_2) in (x_1^0, x_2^0, λ):

$$H = \begin{pmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{pmatrix}$$

(H symmetrisch!) und

$$\underline{c} = \operatorname{grad} g(x_1^0, x_2^0).$$

Satz.

Unter den gestellten Voraussetzungen gelten folgende Kriterien:

1. Die Bedingung
$$\det\begin{pmatrix} 0 & \underline{c}^{\mathsf{T}} \\ \underline{c} & H \end{pmatrix} < 0$$
ist notwendig und hinreichend dafür, dass $\underline{y}^{\mathsf{T}} H \underline{y} > 0$ für alle $\underline{y} \neq \underline{o}$ mit $\underline{c}^{\mathsf{T}} \underline{y} = 0$ gilt. Also ist unter dieser Bedingung (x_1^0, x_2^0) strikter lokaler Minimalpunkt von (P).

2. Die Bedingung
$$\det\begin{pmatrix} 0 & \underline{c}^{\mathsf{T}} \\ \underline{c} & H \end{pmatrix} > 0,$$
ist notwendig und hinreichend dafür, dass $\underline{y}^{\mathsf{T}} H \underline{y} < 0$ für alle $\underline{y} \neq \underline{o}$ mit $\underline{c}^{\mathsf{T}} \underline{y} = 0$ gilt. Also ist unter dieser Bedingung (x_1^0, x_2^0) strikter lokaler Maximalpunkt von (P).

Beweisidee (Notwendigkeit).

Sei $\underline{y} = (y_1, y_2)^{\mathsf{T}}$ ein Vektor mit

$$\underline{y} \neq \underline{o}, \quad \underline{c}^{\mathsf{T}}\underline{y} = 0 \;\text{ und }\; \underline{y}^{\mathsf{T}} H \underline{y} > 0. \tag{11}$$

Nach Voraussetzung ist $\underline{c} \neq \underline{o}$, sei zum Beispiel $c_1 \neq 0$. Da $\underline{c}^{\mathsf{T}}\underline{y} = 0$ und $(y_1, y_2)^{\mathsf{T}} \neq \underline{o}$ gilt, darf dann y_2 nicht Null sein, also folgt $y_2 \neq 0$.

Seien

$$\widehat{H} := \begin{pmatrix} 0 & c_1 & c_2 \\ c_1 & h_{11} & h_{12} \\ c_2 & h_{12} & h_{22} \end{pmatrix} \;\text{ und }\; Q := \begin{pmatrix} 0 & 1 & 0 \\ y_1 & 0 & 1 \\ y_2 & 0 & 0 \end{pmatrix}, \quad \text{d.h.,} \quad \det Q = y_2.$$

Unter Beachtung der Determinantengesetze folgt dann

$$\det(Q^{\mathsf{T}} \widehat{H} Q) = (\det Q)^2 \det \widehat{H} = y_2^2 \det \widehat{H}. \tag{12}$$

Durch Ausmultiplizieren folgt unter Beachtung von $\underline{c}^{\top}\underline{y} = 0$ (Übungsaufgabe: Ausrechnen!)

$$Q^{\top}\widehat{H}Q = \begin{pmatrix} \underline{y}^{\top}H\underline{y} & 0 & h_{11}y_1 + h_{12}y_2 \\ 0 & 0 & c_1 \\ h_{11}y_1 + h_{12}y_2 & c_1 & h_{11} \end{pmatrix},$$

also mit (12) und nach Ausrechnen von $\det(Q^{\top}\widehat{H}Q)$:

$$y_2^2 \det \widehat{H} = \det(Q^{\top}\widehat{H}Q) = -c_1^2\, \underline{y}^{\top}H\underline{y},$$

Es folgt wegen $c_1^2 > 0$ und $y_2^2 > 0$ sowie $\underline{y}^{\top}H\underline{y} > 0$ sofort

$$\det \widehat{H} < 0, \qquad \text{was zu zeigen war.}$$

Anwendung auf Beispiel (9)

Man muss $\underline{c}$ als den Gradienten $\operatorname{grad} g(x_1^0, x_2^0) = (-1, 2)^\mathsf{T}$ nehmen und die oben ausgerechnete Hesse-Matrix einsetzen:

$$\det \begin{pmatrix} 0 & c^\mathsf{T} \\ c & H \end{pmatrix} = \begin{vmatrix} 0 & -1 & 2 \\ -1 & 1 & 0 \\ 2 & 0 & -1 \end{vmatrix} = -3 < 0,$$

also ist auch nach diesem Kriterium bestätigt, dass $\underline{x}^0$ strikter lokaler Minimalpunkt ist.

12.3 (optional) Matrizenrechnung und Regressionsanalyse

Literatur: Schira *Statistische Methoden der VWL und BWL*, Pearson 2003.

Das folgende *lineare* Modell widerspiegele eine ökonomische Hypothese für die Beziehung zwischen zwei Variablen X und Y:

$$Y = \beta_0 + \beta_1 X .$$

Die *Parameter* β_0 und β_1 sind auf der Grundlage von Beobachtungen zu "bestimmen".

Beispiel (Konsumhypothese von Keynes). Der gesamtwirtschaftliche Konsum C sei eine Funktion des verfügbaren Einkommens Y^{verf}:

$$C = \beta_0 + \beta_1 Y^{verf}.$$

Dabei wird $\beta_1 \in (0, 1)$ als *marginale Konsumneigung*, $\beta_0 > 0$ als *autonomer Konsum* (einkommensunabhängiger Konsum) interpretiert.

Im allgemeinen Modell oben sind dann $Y := C$ und $X := Y^{verf}$.

Der ökonometrische Modellansatz berücksichtigt weitere Einflüsse auf Y im ökonomischen Modell $Y = \beta_0 + \beta_1 X$ und fügt eine *Störvariable U* (eine *Zufallsvariable*) hinzu,

$$Y = \beta_0 + \beta_1 X + U .$$

Geht man zu einer Beobachtung der Werte von X und Y über, so gilt für eine *Stichprobe* aus n Werten analog

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad (i = 1, \ldots, n), \tag{13}$$

dabei heissen

y_i **endogene** Variablen (beobachtbar),
x_i **exogene** Variablen (beobachtbar),
u_i **latente** Variablen (nicht beobachtbar),
β_0, β_1 **Modellparameter** oder **Koeffizienten**, ("wahre" Werte)
$\widehat{\beta}_0$, $\widehat{\beta}_1$ **Schätzparameter** oder **Schätzer** (geschätzte Werte).

Die wahren Werte der Parameter β_0 und β_1 sind unbekannt, sie können nur *geschätzt* werden - das ist die Aufgabe der *Regressionsanalyse*.

Die Wahrscheinlichkeitsverteilungen der latenten Variablen sind unbekannt (oft wird aber Normalverteilung vorausgesetzt), aber man unterstellt für Erwartungswert E, Varianz V und Kovarianz Cov, dass $E(u_i) = 0$ und $V(u_i) = \sigma^2$=const. für alle i sowie $Cov(u_i, u_k) = 0$ für alle $i, k = 1, \dots, n$, $i \neq k$.

Bei der Schätzung geht es darum, eine Schätzgerade $y = \widehat{\beta}_0 + \widehat{\beta}_1 x$ zu finden, die der unbekannten Modellgeraden $y = \beta_0 + \beta_1 x$ nahe kommt. Man vergleicht

$$
\begin{aligned}
y_i &= \beta_0 + \beta_1 x_i + u_i && \text{(Beobachtungswerte)},\\
\widehat{y}_i &= \widehat{\beta}_0 + \widehat{\beta}_1 x_i && \text{(geschätzte Werte)}.
\end{aligned}
$$

Methode der kleinsten Quadrate

Gegeben beobachtete x_i, y_i, $i = 1, \dots, n$, finde $(\widehat{\beta}_0, \widehat{\beta}_1)$, so dass

$$
\varphi(\widehat{\beta}_0, \widehat{\beta}_1) := \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i)^2 \tag{14}
$$

minimal wird.

Schreiben wir nun das System (13)

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad (i = 1, \ldots, n)$$

in Matrixschreibweise. Man setze - **wir schreiben hier Vektoren fett** -

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Dann hat (13) die völlig gleichbedeutende Darstellung

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}. \tag{15}$$

Die Zielfunktion (14) in der Methode der kleinsten Quadrate lautet dann, äquivalent umgeschrieben mit Matrizen und Vektoren,

$$\varphi(\hat{\boldsymbol{\beta}}) = (\mathbf{y} - X\hat{\boldsymbol{\beta}})^\top(\mathbf{y} - X\hat{\boldsymbol{\beta}}) = \mathbf{y}^\top\mathbf{y} - 2\hat{\boldsymbol{\beta}}^\top X^\top \mathbf{y} + \hat{\boldsymbol{\beta}}^\top X^\top X\hat{\boldsymbol{\beta}},$$

wobei $\hat{\boldsymbol{\beta}}$ der (gesuchte!) geschätzte $\boldsymbol{\beta}$ -Vektor ist

Der Gradient einer gegebenen quadratischen Funktion

$$\begin{aligned} g(p,q) &= a_1 p + a_2 q \quad + \quad b_{11}p^2 + 2b_{12}pq + b_{22}q^2 \\ &= (p,q)\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + (p,q)\begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} \end{aligned}$$

lautet offenbar (ausrechnen!)

$$\operatorname{grad} g(p,q) = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + 2\begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix}.$$

Damit gilt für den Gradienten der Kleinste-Quadrate-Funktion φ im Minimum $\widehat{\boldsymbol{\beta}^{\,*}}$ der Aufgabe (14)

$$\operatorname{grad} \varphi(\widehat{\boldsymbol{\beta}^{\,*}}) = -2X^\top \mathbf{y} + 2X^\top X \widehat{\boldsymbol{\beta}^{\,*}} = \mathbf{o}.$$

Die Matrix

$$X^\top X = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}$$

(mit $\sum = \sum_{i=1}^n$) ist genau dann invertierbar, wenn

$$\det X^\top X = n \sum x_i^2 - \left(\sum x_i\right)^2 \neq 0\,.$$

Es gilt in diesem Falle sogar $\det X^\top X > 0$, wie man leicht ausrechnet, also ist dann wegen $n > 0$ die Matrix $X^\top X$ positiv definit und somit φ (strikt) konvex, d.h., der Punkt $\widehat{\boldsymbol{\beta}}^{\,*}$ aus

$$\mathbf{o} = -2X^\top \mathbf{y} + 2X^\top X \widehat{\boldsymbol{\beta}}^{\,*} \quad \Leftrightarrow \quad \widehat{\boldsymbol{\beta}}^{\,*} = (X^\top X)^{-1} X^\top \mathbf{y} \tag{16}$$

definiert den eindeutig bestimmten Minimalpunkt der Aufgabe (14).

Die Komponenten $\widehat{\beta_0}^*$ und $\widehat{\beta_1}^*$ von $\widehat{\boldsymbol{\beta}}^*$ kann man als Lösung des LGS $X^\top X \boldsymbol{\beta} = X^\top \mathbf{y}$ mit der Cramerschen Regel leicht ausrechnen:

$$X^\top \mathbf{y} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum y_i \\ \sum x_i y_i \end{pmatrix}$$

ergibt

$$\widehat{\beta_0}^* = \frac{\begin{vmatrix} \sum y_i & \sum x_i \\ \sum x_i y_i & \sum x_i^2 \end{vmatrix}}{\det X^\top X} = \frac{(\sum y_i)(\sum x_i^2) - (\sum x_i)(\sum x_i y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

$$\widehat{\beta_1}^* = \frac{\begin{vmatrix} n & \sum y_i \\ \sum x_i & \sum x_i y_i \end{vmatrix}}{\det X^\top X} = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n \sum x_i^2 - (\sum x_i)^2}$$

Mit den in der Statistik üblichen Bezeichnungen für Mittelwerte

$$\bar{x} = \tfrac{1}{n}\sum x_i, \quad \bar{y} = \tfrac{1}{n}\sum y_i, \quad \overline{x^2} = \tfrac{1}{n}\sum x_i^2, \quad \overline{xy} = \tfrac{1}{n}\sum x_i y_i$$

folgen sofort die in Lehrbüchern oft angegebenen Formeln

$$\widehat{\beta_0}^* = \frac{\bar{y}\,\overline{x^2} - \bar{x}\,\overline{xy}}{\overline{x^2}\,\bar{x}^2}, \qquad \widehat{\beta_1}^* = \frac{\overline{xy} - \bar{x}\,\bar{y}}{\overline{x^2}\,\bar{x}^2}.$$

Multiple lineare Regressionsanalyse

Lineare ökonomische Modelle mit zwei Variablen sind häufig zu einschränkend, deshalb spezifiziert man oft Modelle mit zwei oder mehr exogenen Variablen, d.h.,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U\,.$$

wobei U wieder als Störvariable (Zufallsvariable) angesehen wird.

Beispiel (Erweiterte Konsumhypothese). Der gesamtwirtschaftliche Konsum C wird als Funktion des verfügbaren Einkommens Y^{verf} und des Vermögens W angesehen und wir spezifizieren deshalb

$$C = \beta_0 + \beta_1 Y^{verf} + \beta_2 W + U, \quad 0 < \beta_1 < 1,\ \beta_0, \beta_2 > 0.$$

Mit einem solchen Modell kann man argumentieren, dass die Aktienkurse einen Einfluss auf die private Konsumnachfrage ausüben.

Beobachtet man nun wieder Stichprobenvektoren $\mathbf{x}^j$, $j = 1, \ldots, k$, und $\mathbf{y}$ aus je n Elementen, so definiert man via

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & {x^1}_1 & \ldots & {x^k}_1 \\ 1 & {x^1}_2 & \ldots & {x^k}_2 \\ \vdots & \vdots & & \vdots \\ 1 & {x^1}_n & \ldots & {x^k}_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

das multiple lineare Regressionsmodell in Matrixschreibweise

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}. \tag{17}$$

Zu den oben getroffenen Annahmen über die latenten Variablen u_i wird zusätzlich gefordert, dass sie unabhängig und normalverteilt seien. Ferner setzen wir voraus:

$$X^\top X \text{ ist regulär,}$$

offenbar ist dabei $X^\top X$ eine $(k+1, k+1)$-Matrix.

Methode der kleinsten Quadrate (MKQ)

Gegeben $\mathbf{y}$ und X, finde $\widehat{\boldsymbol{\beta}} \in \mathbb{R}^{k+1}$, so dass

$$\Phi(\widehat{\boldsymbol{\beta}}) = (\mathbf{y} - X\widehat{\boldsymbol{\beta}})^\top(\mathbf{y} - X\widehat{\boldsymbol{\beta}})$$

minimal wird. In den Klammern stehen Vektoren der Länge n, $\Phi(\widehat{\boldsymbol{\beta}})$ ist also für jedes $\widehat{\boldsymbol{\beta}}$ eine reelle Zahl !

Ausgerechnet, erhalten wir

$$\Phi(\widehat{\boldsymbol{\beta}}) = \mathbf{y}^\top\mathbf{y} - 2\widehat{\boldsymbol{\beta}}^\top X^\top\mathbf{y} + \widehat{\boldsymbol{\beta}}^\top X^\top X\widehat{\boldsymbol{\beta}} ,$$

somit hat die notwendige Optimalitätsbedingung die Form

$$\operatorname{grad}\Phi(\widehat{\boldsymbol{\beta}}) = -2X^\top\mathbf{y} + 2X^\top X\widehat{\boldsymbol{\beta}} = \mathbf{o}.$$

Die Voraussetzungen sichern die (strikte) Konvexität von Φ (Beweis in *Mathematik III*). Damit hat der eindeutige Optimalpunkt in der MKQ die gleiche Form wie bei (14), (16) - nur der Typ von X bzw. die Länge von $\boldsymbol{\beta}$ unterscheiden sich -, nämlich

$$\widehat{\boldsymbol{\beta}^*} = (X^\top X)^{-1}X^\top\mathbf{y}.$$