Numer. Math. 15, 122—136 (1970)

Definitheitsbedingungen für relative Extrema bei Optimierungs- und Approximationsaufgaben

W. WETTERLING

Herrn Prof. Dr. Dr. h.c. L. Collatz zum 60. Geburtstag gewidmet

Eingegangen am 6. Oktober 1969

Summary. In this paper, sufficient and necessary maximum conditions are established for a class of mathematical programming problems with an infinite set of restrictions, which is described by a finite number of inequalities. The criteria may be applied to nonlinear approximation problems and to the numerical solution of boundary value problems.

§ 1. Einleitung

In diesem Beitrag werden hinreichende und notwendige Bedingungen für relative Extrema bei einem Typ zweistufiger Optimierungsaufgaben (§3) hergeleitet; das sind Aufgaben mit unendlich vielen Ungleichungen als Restriktionen, wobei die Menge der Restriktionen durch ein endliches System von Ungleichungen beschrieben wird. Die Betrachtung dieses Aufgabentyps ist durch die Anwendungsmöglichkeiten nahegelegt: Nichtlineare Probleme der Tschebyscheff-Approximation, insbesondere solche in mehreren reellen Variablen, werden hierdurch erfaßt (§8), ferner Optimierungsprobleme, die bei der numerischen Behandlung von Randwertaufgaben nach dem Prinzip der lokal optimalen Schranken [6] auftreten. In diesen Fällen entspricht die Restriktionenmenge dem Bereich, in dem die Approximations- bzw. Randwertaufgabe formuliert ist. So ist diese Arbeit als Weiterführung des von und mit Collatz in [1] und [2], §15 und weiterhin in [6—8] verfolgten Programms der Anwendung von Approximations- und Optimierungsmethoden bei Randwertaufgaben anzusehen.

Bei den hier betrachteten Aufgaben werden über die Zielfunktion und die Restriktionsfunktionen keine generellen Konvexitätsvoraussetzungen gemacht; es ergeben sich schwache Definitheitsbedingungen für die Matrix der zweiten Ableitungen einer erweiterten Lagrange-Funktion.

Hinreichend für ein Extremum ist die Definitheit, notwendig die Semidefinitheit der quadratischen Form zu dieser Matrix im Tangentialraum an der betrachteten Stelle. Während dieser im Fall von endlich vielen Restriktionen (§2) durch ein homogenes lineares Gleichungssystem beschrieben wird, in dem die ersten Ableitungen der Restriktionsfunktionen auftreten, hat man im Fall der zweistufigen Aufgaben (§3 ff.) auch lineare Relationen, die die zweiten Ableitungen enthalten; diese sind als linearisierte Enveloppenbedingungen zu deuten.

Die wichtige Frage nach numerischen Methoden zur Behandlung solcher Aufgaben wird hier nicht erörtert. Offenbar erfordern die verschiedenen speziellen

Probleme, die sich der hier betrachteten allgemeinen Problemklasse unterordnen, verschiedenartige Algorithmen. Einige Hinweise findet man in [6] und [8]. Die in diesem Beitrag angegebenen Bedingungen bieten die Möglichkeit, das Ergebnis eines solchen Algorithmus, das im allgemeinen nur ein stationärer Wert ist, auf die Extremaleigenschaft zu prüfen.

§ 2. Finite Maximumaufgaben

Zunächst wird eine Aufgabe mit endlich vielen Restriktionen betrachtet; die hierüber bekannten Ergebnisse werden später benötigt:

Gesucht sind relative Maxima einer Funktion $F(x)$ auf einer Menge $M \subset R^n$, die durch

$$M = \{x;\ f_j(x) \leqq 0\ \ (j = 1, \ldots, m)\}$$

beschrieben ist. Dabei sind $F(x)$, $f_1(x), \ldots, f_m(x)$ reellwertig definiert und zweimal stetig differenzierbar für alle $x \in R^n$. Mit $F'(x)$ wird der Spaltenvektor grad $F(x)$ mit den Komponenten $\partial F/\partial x_i$ $(i = 1, \ldots, n)$ bezeichnet, mit $F''(x)$ die Matrix der zweiten partiellen Ableitungen von F. Entsprechend sind $f_j'(x)$ und $f_j''(x)$ zu verstehen.

Sei nun ein $\tilde{x} \in M$ gegeben; man möchte prüfen, ob F an der Stelle $\tilde{x}$ ein relatives Maximum bezüglich M hat. Hinreichend dafür sind folgende Bedingungen:

(a) *Lokale Kuhn-Tucker-Bedingung. Für $j \in J \subset \{1, \ldots, m\}$, aber nicht notwendig nur für diese Indizes, ist $f_j(\tilde{x}) = 0$, und es gibt reelle Zahlen $u_j > 0$ $(j \in J)$ mit*

$$F'(\tilde{x}) - \sum_{j \in J} u_j f_j'(\tilde{x}) = 0. \tag{1}$$

(b) *Sei $H \subset R^n$ der lineare Teilraum der Vektoren $\xi \in R^n$ mit $\xi^T f_j'(\tilde{x}) = 0$ $(j \in J)$* (ξ^T ist der zum Spaltenvektor ξ transponierte Zeilenvektor); *die quadratische Form*

$$Q(\xi) = \xi^T \left[F''(\tilde{x}) - \sum_{j \in J} u_j f_j''(\tilde{x})\right] \xi$$

ist negativ definit auf H.

Sind die Bedingungen (a) und (b) erfüllt, so gibt es eine Umgebung von $\tilde{x}$, in deren Durchschnitt mit M ($\tilde{x}$ selbst ausgenommen) $F(x) < F(\tilde{x})$ ist.

Die entsprechenden notwendigen Bedingungen lauten: Sei $\tilde{x} \in M$ und J nun die Menge aller Indizes j mit $f_j(\tilde{x}) = 0$. Folgende *constraint qualifications* an der Stelle $\tilde{x}$ seien erfüllt:

Ist $\xi \in R^n$ ein Vektor mit $\xi^T f_j'(\tilde{x}) \leqq 0$ $(j \in J)$, so gibt es ein $t_0 > 0$ und eine für $0 \leqq t \leqq t_0$ stetig differenzierbare vektorwertige Funktion $x(t)$ mit $x(0) = \tilde{x}$, $dx(0)/dt = \xi$ und $x(t) \in M$ $(0 \leqq t \leqq t_0)$. Ist $\xi^T f_j'(\tilde{x}) = 0$ $(j \in J)$, so gibt es ein solches $x(t)$, das überdies zweimal stetig differenzierbar ist und für das $f_j(x(t)) = 0$ $(j \in J,\ 0 \leqq t \leqq t_0)$ ist.

Ist dann $F(x) \leqq F(\tilde{x})$ im Durchschnitt einer Umgebung von $\tilde{x}$ mit M, so gilt:

(a*) *Es gibt $u_j \geqq 0$ $(j \in J)$ mit* (1).

(b*) *$Q(\xi)$ ist negativ semidefinit auf H.*

Man beachte jedoch, daß J und damit H und Q jetzt anders definiert sind als im Fall der hinreichenden Bedingungen.

Diese Aussagen sind in den Ergebnissen von McCormick [4] enthalten, der neben Ungleichungen auch Gleichungen als Restriktionen zuläßt.

Dort wird außerdem gezeigt, daß die *constraint qualifications* sicher dann erfüllt sind, wenn die Vektoren $f_j'(\tilde{x})$ $(j \in J)$ linear unabhängig sind. Für das von McCormick mit indirektem Beweis hergeleitete, hinreichende Kriterium ist in [8] unter der Voraussetzung der Stetigkeit der dritten Ableitungen der auftretenden Funktionen ein direkter Beweis angegeben, bei dem eine Kugelumgebung von $\tilde{x}$ mit der genannten Eigenschaft konstruiert wird.

§ 3. Zweistufige Maximumaufgaben, hinreichende Bedingungen

Eine ähnliche Bedingung soll nun für folgenden Aufgabentyp, dessen Struktur die Bezeichnung *zweistufig* nahelegt, formuliert werden: *Gesucht sind relative Maxima von* $F(x)$ *auf*

$$M = \{x;\ f(x, y) \leqq 0\ \ (y \in Y)\} \subset R^n;$$

dabei ist

$$Y = \{y;\ g_\nu(y) \leqq 0\ \ (\nu = 1, \ldots, N)\} \subset R^m.$$

Die Funktionen F, f, g_ν seien reellwertig definiert und zweimal stetig differenzierbar im R^n, $R^n \times R^m$ bzw. R^m. Wie oben werden mit $F'(x)$ und $f'(x, y)$ die Vektoren der ersten Ableitungen nach den x_i bezeichnet, ferner mit $f_y'(x, y)$ und $g_\nu'(y)$ die Vektoren der ersten Ableitungen nach den y_k, weiterhin mit F'', f_{xx}'', f_{xy}'', f_{yy}'' und g_ν'' die entsprechenden Matrizen der zweiten Ableitungen.

Wieder wird ein Punkt $\tilde{x} \in M$ betrachtet. Als hinreichend dafür, daß dort ein relatives Maximum von F bezüglich M vorliegt, erweisen sich die folgenden Bedingungen:

(A) *Lokale Kuhn-Tucker-Bedingung. Für* $y_1, y_2, \ldots, y_p \in Y$, *aber nicht notwendig nur für diese Punkte, sei* $f(\tilde{x}, y_j) = 0$, *und es gebe reelle Zahlen* $u_j > 0$ $(j = 1, \ldots, p)$ *mit*

$$F' - \sum_{j=1}^{p} u_j f_j' = 0 \tag{2}$$

(das Argument $\tilde{x}$ wird hier und im folgenden häufig weggelassen, das Argument y_j ebenfalls und durch einen Index j angedeutet).

Als Funktion von y auf der Menge Y hat $f(\tilde{x}, y)$ bei $y = y_j$ Maxima mit dem Funktionswert $f(\tilde{x}, y_j) = 0$. Diese Maxima sollen so beschaffen sein, daß ähnliche Definitheitsbedingungen wie in §2 gelten.

(B) *Für jedes* j $(1 \leqq j \leqq p)$ *gelte: Für* $\nu \in M_j \subset \{1, \ldots, N\}$ *und nur für diese Indizes ist* $g_\nu(y_j) = 0$, *die Vektoren* $g_{\nu j}' = g_\nu'(y_j)$ $(\nu \in M_j)$ *sind linear unabhängig, und es gibt reelle Zahlen* $v_{\nu j} \geqq 0$ $(\nu \in M_j)$ *mit*

$$f_j' - \sum_{\nu \in M_j} v_{\nu j}\, g_{\nu j}' = 0.$$

Auf dem linearen Teilraum

$$H_j = \{\eta;\ \eta^T g_{\nu j}' = 0\ \ (\nu \in M_j)\} \subset R^m$$

ist die quadratische Form

$$Q_j(\eta) = \eta^T \left(f_j'' - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}'' \right) \eta$$

negativ definit.

Bevor die eigentliche Definitheitsbedingung (C) formuliert werden kann, sind einige Zwischenüberlegungen nötig. Ist m_j die Anzahl der Indizes in M_j, so hat H_j die Dimension $m - m_j$.

Seien nun $\eta_j^{(1)}, \ldots, \eta_j^{(m-m_j)}$ Vektoren, die eine Basis von H_j bilden. Hiermit wird das lineare Gleichungssystem

$$\eta_j^{(k)T} \left[f_j' \xi + \left(f_j'' - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}'' \right) \eta \right] = 0 \qquad (k = 1, \ldots, m - m_j) \tag{3}$$

aufgestellt. Es zeigt sich, daß es bei gegebenem $\xi \in R^n$ genau ein $\eta \in H_j$ als Lösung dieses Systems gibt. Der Ansatz

$$\eta = \sum_{i=1}^{m-m_j} c_i \eta_j^{(i)}$$

führt nämlich auf ein lineares Gleichungssystem für die Unbekannten c_i mit der Koeffizientenmatrix

$$\left(\eta_j^{(k)T} \left[f_j'' - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}'' \right] \eta_j^{(i)} \right)_{k,\, i = 1, \ldots, m - m_j},$$

die wegen der Definitheit von $(f_j'' - \sum v_{\nu j} g_{\nu j}'')$ nichtsingulär ist. So sind also einem $\xi \in R^n$ als Lösungen von (3) für $j = 1, \ldots, p$ Vektoren $\eta_1 \in H_1, \ldots, \eta_p \in H_p$ eindeutig zugeordnet, und zwar linear:

$$\eta_j = P_j \xi \qquad (j = 1, \ldots, p)$$

mit von ξ unabhängigen $m \times n$-Matrizen P_j.

Die Definitheitsbedingung lautet dann

(C) *Die quadratische Form*

$$Q(\xi) = \xi^T \left[F'' - \sum_{j=1}^{p} u_j \left(f_j'' + 2 P_j^T f_j' + P_j^T \left(f_j'' - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}'' \right) P_j \right) \right] \xi$$

sei auf

$$H = \{ \xi ;\, \xi^T f_j' = 0 \;\; (j = 1, \ldots, p) \}$$

negativ definit.

Diese Bedingung kann auch so formuliert werden: Es sei

$$q(\xi, \eta_j) = \xi^T (F'' - \sum u_j f_j'') \xi - 2 \sum u_j \eta_j^T f_j' \xi - \sum u_j \eta_j^T (f_j'' - \sum v_{\nu j} g_{\nu j}'') \eta_j < 0 \tag{4}$$

für $\xi \in H$, $\eta_j \in H_j$, sofern ξ und die η_j durch (3) verknüpft sind und solange nicht $\xi = 0$ und daher $\eta_1 = \cdots = \eta_p = 0$ ist. Wegen (3) ist hier

$$\eta_j^T f_j' \xi + \eta_j^T (f_j'' - \sum v_{\nu j} g_{\nu j}'') \eta_j = 0$$

und daher

$$q(\xi, \eta) = \xi^T (F'' - \sum u_j f_j'') \xi + \sum u_j \eta_j^T (f_j'' - \sum v_{\nu j} g_{\nu j}'') \eta_j .$$

Die negative Definitheit von $\xi^T (F'' - \sum u_j f_j'') \xi$ (vgl. §2) ist also zusammen mit (B) hinreichend für (C), ist aber ersichtlich einschränkender als (C).

§ 4. Beweis der Maximaleigenschaft

Satz. *Sind für ein $\tilde{x}\in M$ die Bedingungen* (A), (B), (C) *erfüllt, so gibt es eine Kugel um $\tilde{x}$, in deren Durchschnitt mit M ($\tilde{x}$ selbst ausgenommen) $F(x)<F(\tilde{x})$ ist.*

Beweis (indirekt). Aus der folgenden Annahme ist ein Widerspruch herzuleiten: Es gibt eine Folge $x_1, x_2, \ldots$ von Punkten $x_k\in M$ mit $x_k\neq\tilde{x}$, $\lim\limits_{k\to\infty} x_k=\tilde{x}$ und

$$F(x_k)\geqq F(\tilde{x}). \tag{5}$$

Sei $x_k-\tilde{x}=\delta_k\xi_k$ mit $\delta_k>0$, $\|\xi_k\|=1$ mit irgendeiner Vektornorm und daher $\lim\delta_k=0$. Die Folge der ξ_k enthält eine konvergente Teilfolge. Ohne Beschränkung der Allgemeinheit sei also $\lim\xi_k=\xi$ mit $\|\xi\|=1$. Es ist

$$\frac{1}{\delta_k}[F(\tilde{x}+\delta_k\xi_k)-F(\tilde{x})]\geqq 0$$

und daher auch der Grenzwert $\xi^T F'(\tilde{x})\geqq 0$. Wegen $x_k\in M$ und $f(\tilde{x}, y_j)=0$ ist ferner

$$\frac{1}{\delta_k}[f(\tilde{x}+\delta_k\xi_k, y_j)-f(\tilde{x}, y_j)]\leqq 0$$

und daher $\xi^T f_j'\leqq 0$. Mit (2) wird

$$0\leqq\xi^T F'=\sum_{j=1}^{p} u_j\xi^T f_j'\leqq 0.$$

Wegen $u_j>0$ ist $\xi^T F'=\xi^T f_1'=\cdots=\xi^T f_p'=0$, also $\xi\in H$. Mit diesem ξ sei $\eta_j=P_j\xi$ $(j=1,\ldots,p)$.

Für jeden Index j $(1\leqq j\leqq p)$ gilt nun folgende Überlegung: Es ist $g_\nu(y_j)=0$ und $\eta_j^T g_\nu'(y_j)=0$ $(\nu\in M_j)$. In einem Intervall $0\leqq\delta\leqq\Delta$ mit hinreichend kleinem positiven Δ gibt es einen Vektor $\psi_j=\psi_j(\delta)\in R^m$ mit $\psi_j^T\eta=0$ $(\eta\in H_j)$ und

$$g_\nu(y_j+\delta\eta_j+\psi_j(\delta))=0\quad(\nu\in M_j), \tag{6}$$

der als Funktion von δ zweimal stetig differenzierbar und mit $\psi_j(0)=0$ eindeutig bestimmt ist. Um dies einzusehen, mache man den Ansatz

$$\psi_j(\delta)=\sum_{\nu\in M_j} b_\nu(\delta)\, g_{\nu j}'$$

und wende bekannte Sätze über implizite Funktionen an (s. etwa [3], Kap. X). Die hier auftretende Jacobi-Matrix $(g_{\nu j}'^T g_{\mu j}')_{\nu,\mu\in M_j}$ ist wegen der linearen Unabhängigkeit der Vektoren $g_{\nu j}'$ nichtsingulär. Aus (6) folgt weiterhin

$$g_{\nu j}'^T\left(\eta_j+\frac{d\psi_j(0)}{d\delta}\right)=0\quad(\nu\in M_j);$$

wegen $g_{\nu j}'^T\eta_j=0$ und der linearen Unabhängigkeit der $g_{\nu j}'$ ist $d\psi_j(0)/d\delta=0$ und daher $\psi_j(\delta)=\delta^2\chi_j(\delta)$ mit gleichmäßig beschränktem $\chi_j(\delta)$ für $0\leqq\delta\leqq\Delta$.

Sei $\Delta>0$ weiterhin so klein gewählt, daß $g_\nu(y_j+\delta\eta_j+\delta^2\chi_j(\delta))<0$ ist für $\nu\notin M_j$ und $0\leqq\delta\leqq\Delta$, also $y_j+\delta\eta_j+\delta^2\chi_j(\delta)\in Y$.

Die Zahlen δ_k der obigen Folge liegen für hinreichend großes k im Intervall $[0, \Delta]$. Für diese k wird mit $\chi_{jk} = \chi_j(\delta_k)$

$$\begin{aligned} F(\tilde{x} + \delta_k \xi_k) - \sum_{j=1}^{p} u_j \Big\{ & \underbrace{f(\tilde{x} + \delta_k \xi_k, y_j + \delta_k \eta_j + \delta_k^2 \chi_{jk})}_{\leqq 0} \\ & - \sum_{\nu \in M_j} v_{\nu j} \underbrace{g_\nu(y_j + \delta_k \eta_j + \delta_k^2 \chi_{jk})}_{=0} \Big\} \\ = F(\tilde{x}) - \sum_{j=1}^{p} u_j \Big\{ & \underbrace{f(\tilde{x}, y_j)}_{=0} - \sum_{\nu \in M_j} \underbrace{v_{\nu j}\, g_\nu(y_j)}_{=0} \Big\} \\ + \delta_k [\xi_k^T & \underbrace{(F' - \sum u_j f_j')}_{=0} - \sum u_j (\eta_j + \delta_k \chi_{jk})^T \underbrace{(f_j' - \sum v_{\nu j} g_{\nu j}')}_{=0}] \\ + \tfrac{1}{2} \delta_k^2 & q(\xi_k, \eta_j) + o(\delta_k^2). \end{aligned}$$

Für $k \to \infty$ hat man hiermit einen Widerspruch zwischen (4) und (5).

Bemerkung. Aus dem Beweis ist zu erkennen, daß man auch $\eta_j = \alpha_j P_j \xi$ mit beliebigen reellen α_j hätte wählen können, und zwar bereits in der Formulierung der Bedingung (C). Man erkennt aber, daß $\alpha_j = 1$ als Maximalstelle von $\alpha_j(2 - \alpha_j)$ ausgezeichnet ist und für $\alpha_j \neq 1$ zwar auch hinreichende, aber einschränkendere Bedingungen als für $\alpha_j = 1$ herauskommen. Auch aus den folgenden notwendigen Bedingungen wird das deutlich.

§ 5. Notwendige Maximalbedingungen

Bei $\tilde{x}$ habe $F(x)$ bezüglich $M = \{x; f(x, y) \leqq 0 \ (y \in Y)\}$ ein schwaches relatives Maximum, d.h. im Durchschnitt einer Umgebung von $\tilde{x}$ mit M sei $F(x) \leqq F(\tilde{x})$. Sei dabei $f(\tilde{x}, y) = 0$ genau für $y = y_1, \ldots, y_p \in Y$ (es sind also nur endlich viele Nullstellen von $f(\tilde{x}, y)$ in Y zugelassen). Außer dieser Einschränkung werden noch einige weitere Bedingungen (*constraint qualifications*) gefordert, zunächst:

(I) *Ist $\xi \in R^n$ ein Vektor mit $\xi^T f'(\tilde{x}, y_j) \leqq 0$ $(j = 1, \ldots, p)$, so gibt es ein $t_0 > 0$ und eine für $0 \leqq t \leqq t_0$ stetig differenzierbare vektorwertige Funktion $x(t)$ mit $x(0) = \tilde{x}$, $dx(0)/dt = \xi$ und $x(t) \in M$ $(0 \leqq t \leqq t_0)$.*

Für ein beliebiges solches ξ ist dann

$$\xi^T F'(\tilde{x}) = \frac{d}{dt} F\big(x(t)\big)_{(t=0)} \leqq 0,$$

es gibt also kein $\xi \in R^n$ mit $\xi^T f'(\tilde{x}, y_j) \leqq 0$ $(j = 1, \ldots, p)$ und $\xi^T F'(\tilde{x}) > 0$.

Nach dem Alternativsatz für Systeme von linearen Ungleichungen (Lemma von Farkas, [2], §5) gibt es $u_j \geqq 0$ $(j = 1, \ldots, p)$ mit

$$F'(\tilde{x}) - \sum_{j=1}^{p} u_j f'(\tilde{x}, y_j) = 0. \tag{7}$$

Weiterhin sei folgende Bedingung erfüllt:

(II) *Ist $\xi \in R^n$ ein Vektor mit $\xi^T f'(\tilde{x}, y_j) = 0$ $(j = 1, \ldots, p)$, so gibt es für $0 \leqq t \leqq t_1$ $(t_1 > 0)$ ein zweimal stetig differenzierbares $x(t) \in M$ mit $x(0) = \tilde{x}$, $dx(0)/dt = \xi$ und zugeordnete, ebenfalls zweimal stetig differenzierbare $z_j(t) \in Y$ mit $z_j(0) = y_j$ und*

$$f\big(x(t), z_j(t)\big) = 0 \quad (0 \leqq t \leqq t_1;\ j = 1, \ldots, p). \tag{8}$$

Seien schließlich die Maximalstellen y_j von $f(\tilde{x}, y)$ wie folgt beschaffen:

(III) *Für $j = 1, \ldots, p$ gilt: Es ist $g_\nu(y_j) = 0$ genau für $\nu \in M_j \subset \{1, \ldots, N\}$, die Vektoren $g'_{\nu j}$ $(\nu \in M_j)$ sind linear unabhängig, es ist*

$$f'_j - \sum_{\nu \in M_j} v_{\nu j}\, g'_{\nu j} = 0 \tag{9}$$

mit $v_{\nu j} > 0$ $(\nu \in M_j)$, und die Matrix $f''_j - \sum_{\nu \in M_j} v_{\nu j}\, g''_{\nu j}$ ist negativ definit auf $H_j = \{\eta;\ \eta^T g'_{\nu j} = 0\ (\nu \in M_j)\}$.

Sei nun $\xi \in H = \{\xi;\ \xi^T f'_j = 0\ (j = 1, \ldots, p)\}$ und hierzu $x(t)$ und $z_1(t), \ldots, z_p(t)$ gemäß (II) gegeben. Setzt man $dz_j(0)/dt = \eta_j$, so wird nach (8) $\xi^T f'_j + \eta_j^T f'_j = 0$, also $\eta_j^T f'_j = 0$ $(j = 1, \ldots, p)$. Wegen $z_j(t) \in Y$ ist $\eta_j^T g'_{\nu j} \leqq 0$ und wegen $v_{\nu j} > 0$ und (9) sogar

$$\eta_j^T g'_{\nu j} = 0 \quad (\nu \in M_j;\ j = 1, \ldots, p). \tag{10}$$

Für festes $t \in [0, t_1]$ sind die Punkte $y = z_j(t)$ Maximalstellen von $f\big(x(t), y\big)$ bezüglich Y. Es zeigt sich, daß wenigstens in einem Teilintervall von $[0, t_1]$ die notwendigen Bedingungen von McCormick (§2) gelten. Wegen der Stetigkeit der g_ν ist nämlich für $0 \leqq t \leqq t_2$ mit $0 < t_2 \leqq t_1$ $g_\nu\big(z_j(t)\big) < 0$ $(\nu \notin M_j)$. Sei $M_j(t) \subset M_j$ die Menge der Indizes ν, für die $g_\nu\big(z_j(t)\big) = 0$ ist $(0 \leqq t \leqq t_2)$. Auch die Vektoren $g'_\nu\big(z_j(t)\big)$ hängen stetig von t ab, und darum gibt es ein t_3 mit $0 < t_3 \leqq t_2$, so daß die $g'_\nu\big(z_j(t)\big)$ $\big(\nu \in M_j(t)\big)$ linear unabhängig sind, sofern $t \in [0, t_3]$ ist. Nach §2 gibt es Zahlen $w_{\nu j}(t) \geqq 0$ $\big(\nu \in M_j(t)\big)$ mit

$$f'\big(x(t), z_j(t)\big) - \sum_{\nu \in M_j(t)} w_{\nu j}(t)\, g'_\nu\big(z_j(t)\big) = 0. \tag{11}$$

Betrachtet man (9) und (11), so erkennt man: Für $t = 0$ liegt der Vektor $f'\big(x(t), z_j(t)\big)$ im relativen Inneren des von den Vektoren $g'_\nu\big(z_j(t)\big)$ aufgespannten Kegels. Alle diese Vektoren hängen stetig von t ab. Darum gibt es ein t_4 mit $0 < t_4 \leqq t_3$, so daß für $t \in [0, t_4]$ die Gl. (11) nur dann gelten kann, wenn $M_j(t) = M_j(0) = M_j$ und $w_{\nu j}(t) > 0$ ist. Es ist also auch $g_\nu\big(z_j(t)\big) = 0$ in $[0, t_4]$ für alle $\nu \in M_j$. Ebenso wie die Vektoren f' und g'_ν im Gleichungssystem (11) hängt auch dessen Lösung $w_{\nu j}(t)$ mit dem Anfangswert $w_{\nu j}(0) = v_{\nu j}$ im Intervall $[0, t_4]$ stetig differenzierbar von t ab. Aus (11) folgt dann

$$f'_j{}'\,\xi + \Big(f''_j - \sum_{\nu \in M_j} v_{\nu j}\, g''_{\nu j}\Big)\eta_j - \sum_{\nu \in M_j} \frac{dw_{\nu j}(0)}{dt}\, g'_{\nu j} = 0.$$

Ist $\{\eta_j^{(1)}, \ldots, \eta_j^{(m-m_j)}\}$ eine Basis von $H_j = \{\eta;\ \eta^T g'_{\nu j} = 0\ (\nu \in M_j)\}$, so wird

$$\eta_j^{(k)}\Big[f'_j{}'\,\xi + \Big(f''_j - \sum_{\nu \in M_j} v_{\nu j}\, g''_{\nu j}\Big)\eta_j\Big] = 0 \quad (k = 1, \ldots, m - m_j). \tag{12}$$

Als zugeordnete $z_j(t)$ in (II) kommen also nur solche in Frage, für die mit $\eta_j = dz_j(0)/dt$ das lineare Gleichungssystem (12) erfüllt ist. Weil $x(t) \in M$ für $0 \leqq t \leqq t_4$, F bei $\tilde{x} = x(0)$ maximal und

$$\frac{d}{dt} F\big(x(t)\big)_{(t=0)} = \xi^T F' = \sum_{j=1}^{p} u_j \xi^T f_j' = 0$$

ist, wird

$$\frac{d^2}{dt^2} F\big(x(t)\big)_{(t=0)} = \alpha^T F' + \xi^T F'' \xi \leqq 0,$$

wobei $\alpha = d^2x(0)/dt^2$ ist. Weiterhin ist mit $\beta_j = d^2 z_j(0)/dt^2$

$$\frac{d^2}{dt^2} f\big(x(t), z_j(t)\big)_{(t=0)} = \alpha^T f_j' + \beta_j^T f_j^{\cdot} + \xi^T f_j'' \xi + 2\xi^T f_j^{\cdot\prime} \eta_j + \eta_j^T f_j^{\cdot\cdot} \eta_j = 0 \quad (j = 1, \ldots, p)$$

und

$$\frac{d^2}{dt^2} g_\nu\big(z_j(t)\big)_{(t=0)} = \beta_j^T g_{\nu j}^{\cdot} + \eta_j^T g_{\nu j}^{\cdot\cdot} \eta_j = 0 \quad (\nu \in M_j;\ j = 1, \ldots, p).$$

Mit (7) und (9) wird

$$\begin{aligned} \frac{d^2}{dt^2} F\big(x(t)\big)_{(t=0)} = \xi^T \Big(F'' - \sum_{j=1}^{p} u_j f_j''\Big) \xi - 2 \sum_{j=1}^{p} u_j \xi^T f_j^{\cdot\prime} \eta_j \\ - \sum_{j=1}^{p} u_j \eta_j^T \Big(f_j^{\cdot\cdot} - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}^{\cdot\cdot}\Big) \eta_j \leqq 0 \end{aligned} \tag{13}$$

und zwar dies für alle $\xi \in H$ und die ihnen durch (12) eindeutig zugeordneten $\eta_j \in H_j$.

Satz. *Sei $F(x) \leqq F(\tilde{x})$ im Durchschnitt einer Umgebung von $\tilde{x}$ mit M, und sei $f(\tilde{x}, y) = 0$ genau für $y = y_1, \ldots, y_p \in Y$. Ist dann die Bedingung* (I) *erfüllt, so gilt die lokale Kuhn-Tucker-Bedingung* (7). *Sind außerdem die Bedingungen* (II) *und* (III) *erfüllt, so gilt die Semidefinitheitsbedingung* (13).

§ 6. Die constraint qualifications

Die Bedingungen (I) und (II) sind nicht so einschränkend, wie es zunächst scheint. Sie besagen, daß man an einer Maximalstelle nur dann die angegebenen notwendigen Bedingungen folgern kann, wenn diese Stelle nicht ein Randpunkt von M von dem Typ der aus der Theorie der linearen Optimierung bekannten entarteten Ecken ist. Hier gilt ein ähnlicher Satz wie im Fall der finiten Maximumaufgaben (§2).

Satz. *Sei die Menge Y beschränkt, seien ferner $f(x, y)$, $g_1(y), \ldots, g_N(y)$ dreimal stetig differenzierbar und $y_1, \ldots, y_p$ alle Punkte in Y mit $f(\tilde{x}, y) = 0$. Ist dann die Bedingung* (III) *erfüllt und sind die Vektoren $f'(\tilde{x}, y_j)$ $(j = 1, \ldots, p)$ linear unabhängig, so sind auch* (I) *und* (II) *erfüllt.*

Beweis. Sei $\xi \in R^n$ ein Vektor mit $\xi^T f'(\tilde{x}, y_j) \leqq 0$ $(j = 1, \ldots, p)$, und zwar sei

$$\xi^T f'(\tilde{x}, y_j) \begin{cases} = 0 & (j = 1, \ldots, r) \\ < 0 & (j = r+1, \ldots, p). \end{cases}$$

Zum Nachweis von (II) ist hier $r=p$ anzunehmen, während (I) für geeignetes r $(0\leqq r\leqq p)$ folgt, wenn man ohne Beschränkung der Allgemeinheit annimmt, daß die Numerierung der y_j passend gewählt ist.

Es werden vektorwertige Funktionen $x(t), z_1(t), \ldots, z_r(t)$ konstruiert, die in einem Intervall $[0, t_1]$ mit $t_1>0$ zweimal stetig differenzierbar sind und für die

$$\left.\begin{aligned} &x(0)=\tilde{x}, \qquad dx(0)/dt=\xi,\\ &z_j(0)=y_j \qquad (j=1,\ldots,r),\\ &f\big(x(t), z_j(t)\big)=0\\ &g_\nu\big(z_j(t)\big)=0 \quad (\nu\in M_j) \end{aligned}\right\} \quad (j=1,\ldots,r;\ 0\leqq t\leqq t_1) \qquad \begin{matrix}(14)\\(15)\end{matrix}$$

gilt. Bei $y=z_j(t)$ soll eine Maximalstelle von $f\big(x(t), y\big)$ bezüglich Y liegen. Dort werden daher noch die lokalen Kuhn-Tucker-Bedingungen gefordert:

$$f_y'\big(x(t), z_j(t)\big)-\sum_{\nu\in M_j} w_{\nu j}(t)\, g_\nu'\big(z_j(t)\big)=0 \quad (j=1,\ldots,r;\ 0\leqq t\leqq t_1) \tag{16}$$

mit $w_{\nu j}(0)=v_{\nu j}$. Wenn wie oben m_j die Anzahl der Indizes in M_j ist, hat man mit (14), (15) und (16)

$$r+\sum_{j=1}^{p} m_j+mr$$

Gleichungen. Für $x(t)$ wird der Ansatz

$$x(t)=\tilde{x}+t\xi+\sum_{k=1}^{r} c_k(t)\, f'(\tilde{x}, y_k) \tag{17}$$

gemacht.

Man hat dann r reellwertige Funktionen $c_k(t)$, $\sum m_j$ reellwertige Funktionen $w_{\nu j}(t)$ und mr Komponenten der vektorwertigen Funktionen $z_j(t)$ zu den Anfangswerten

$$c_k(0)=0, \qquad w_{\nu j}(0)=v_{\nu j}, \qquad z_j(0)=y_j$$

aus den Gln. (14), (15) und (16) zu bestimmen. Für die Anfangswerte sind diese Gleichungen erfüllt. Die schon in §4 benutzten Sätze über implizite Funktionen sind anwendbar, denn die Jacobi-Matrix an der Stelle $t=0$ ist nichtsingulär. Sie ist von der Form

$$\begin{matrix}\\ r\\ \sum m_j\\ mr\end{matrix}\begin{matrix}\begin{pmatrix} A & 0 & B\\ 0 & 0 & C\\ D & -C^T & G\end{pmatrix}\\ \underbrace{\qquad}_{r}\ \underbrace{\qquad\quad}_{\sum m_j}\ \underbrace{\qquad}_{mr}\end{matrix}\begin{matrix}\}\ r\\ \}\sum m_j\\ \}\ mr\end{matrix} \tag{18}$$

Dabei ist $A=(f_j'^T f_k')_{j,k=1,\ldots,r}$ wegen der linearen Unabhängigkeit der f_j' nichtsingulär. Weiterhin ist

$$B=\begin{pmatrix} f_1'^T & 0 & \ldots.\\ 0 & f_2'^T & \ldots.\\ .\ .\ .\ . & \ldots.\end{pmatrix},$$

C ist entsprechend aus den Vektoren $g_{\nu j}^{\cdot T}$ aufgebaut, D aus Vektoren $f_j^{\cdot\prime}$ f_k', und G ist eine Blockmatrix mit r $m\times m$-Matrizen $f_j^{\cdot\cdot} - \sum v_{\nu j}\, g_{\nu j}^{\cdot\cdot}$ längs der Hauptdiagonale, sonst mit Nullen ausgefüllt. Die Matrix (18) ist nichtsingulär, wenn es die Matrix

$$\begin{pmatrix} A & 0 & 0 \\ 0 & 0 & C \\ D & -C^T & G \end{pmatrix}$$

ist, denn diese entsteht wegen (9) aus (18), wenn man jeweils das $v_{\nu j}$-fache der Zeilen von $(0\ 0\ C)$ von den entsprechenden Zeilen $(A\ 0\ B)$ subtrahiert. Es bleibt zu zeigen, daß

$$\begin{pmatrix} 0 & C \\ -C^T & G \end{pmatrix}$$

nichtsingulär ist. Es genügt, dies für eine Matrix

$$\begin{pmatrix} 0 & \dots & 0 & & g_{1j}^{\cdot T} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \dots & 0 & & g_{mj}^{\cdot T} \\ g_{1j}^{\cdot} & \dots & g_{mj}^{\cdot} & f_j^{\cdot\cdot} - \sum v_{\nu j}\, g_{\nu j}^{\cdot\cdot} & \end{pmatrix}$$

unter der Annahme $M_j = \{1, \dots, m_j\}$ zu zeigen. Wäre sie singulär, so gäbe es Zahlen $a_1, \dots, a_{m_j}$ und einen Vektor $\gamma \in R^m$, nicht sämtlich $=0$, mit

$$\begin{gathered} \sum a_\nu\, g_{\nu j}^{\cdot} + (f_j^{\cdot\cdot} - \sum v_{\nu j}\, g_{\nu j}^{\cdot\cdot})\,\gamma = 0, \qquad (19) \\ g_{\nu j}^{\cdot T}\,\gamma = 0 \quad (\nu \in M_j). \end{gathered}$$

Die $g_{\nu j}^{\cdot}$ sind in (III) als linear unabhängig vorausgesetzt, darum ist $\gamma \neq 0$. Multipliziert man (19) von links mit γ^T, so wird $\gamma^T (f_j^{\cdot\cdot} - \sum v_{\nu j}\, g_{\nu j}^{\cdot\cdot})\,\gamma = 0$ mit $\gamma \neq 0$ im Widerspruch zur Definitheitsvoraussetzung in (III). Die Sätze über implizite Funktionen besagen dann, daß es ein $t_1 > 0$ und im Intervall $[0, t_1]$ zweimal stetig differenzierbare $c_k(t)$, $z_j(t)$, $w_{\nu j}(t)$ gibt, die mit $c_k(0) = 0$, $z_j(0) = y_j$, $w_{\nu j}(0) = v_{\nu j}$ eindeutig bestimmt sind. Da (16) erste Ableitungen enthält, mußte im Satz die Stetigkeit der dritten Ableitungen von f und g_ν vorausgesetzt werden.

Es ist noch $dx(0)/dt = \xi$ zu zeigen. Aus (14) folgt

$$f_j^{\cdot T}\left(\xi + \sum_{k=1}^{r} \frac{dc_k(0)}{dt}\, f_k'\right) + f_j^{\cdot T}\, \frac{dz_j(0)}{dt} = 0 \quad (j = 1, \dots, r).$$

Hier ist $f_j^{\cdot T}\xi = 0$ und auch $f_j^{\cdot T}\, dz_j(0)/dt = 0$ wegen (9) und $g_{\nu j}^{\cdot T}\, dz_j(0)/dt = 0$, was aus (15) folgt. Da die Matrix $(f_j^{\cdot T} f_k')_{j,k=1,\dots,r}$ nichtsingulär ist, wird $dc_k(0)/dt = 0$ $(k = 1, \dots, r)$ und damit $dx(0)/dt = \xi$.

Für $\nu \notin M_j$ ist $g_\nu(y_j) < 0$ und darum $g_\nu\big(z_j(t)\big) < 0$ für $0 \leqq t \leqq t_2$ mit $0 < t_2 \leqq t_1$, also $z_j(t) \in Y$ für $t \in [0, t_2]$, $j = 1, \dots, r$.

Schließlich ist zu zeigen, daß $x(t) \in M$, also $f\big(x(t), y\big) \leqq 0$ $(y \in Y)$ ist in einem Intervall $[0, t^*]$ mit $t^* > 0$. Hierzu werden Kugelumgebungen von $y_1, \dots, y_r$ und von $y_{r+1}, \dots, y_p$ konstruiert, und es ist zu unterscheiden, ob y in einer solchen Umgebung liegt oder nicht.

Sei zunächst $1 \leqq j \leqq r$. Für hinreichend kleine $t > 0$ ist $y = z_j(t)$ Maximalstelle von $f(x(t), y)$ bezüglich Y zum Maximalwert $f(x(t), z_j(t)) = 0$. Nach (16) und wegen $w_{\nu j}(0) = v_{\nu j} > 0$ $(\nu \in M_j)$ ist nämlich für kleine positive t die lokale Kuhn-Tucker-Bedingung (a) von §2 erfüllt. Die für $t = 0$ durch (III) vorausgesetzte Definitheitsbedingung (b) ist ebenfalls aus Stetigkeitsgründen für kleine positive t erfüllt. Damit gibt es für $t \in [0, t_3]$ mit $0 < t_3 \leqq t_2$ eine Umgebung von $z_j(t)$, in deren Durchschnitt mit Y ($z_j(t)$ selbst ausgenommen)

$$f(x(t), y) < f(x(t), z_j(t)) = 0 \tag{20}$$

ist. Man kann hierfür die in [8] konstruierten Kugelumgebungen wählen, und zwar so, daß deren Radius stetig von t abhängt (hier wird noch einmal die Stetigkeit der dritten Ableitungen von f und g_ν benutzt). Sei also $\varrho_j(t)$ der Radius der Kugelumgebung um $z_j(t)$, wobei $\varrho_j(t) > 0$ und stetig ist in einem Intervall $[0, t_4]$ mit $t_4 > 0$, und zwar dies für alle j mit $1 \leqq j \leqq r$. Nun wird t_5 mit $0 < t_5 \leqq t_4$ so klein gewählt, daß

$$\max_{[0, t_5]} \|y_j - z_j(t)\| \leqq \tfrac{1}{3} \min_{[0, t_5]} \varrho_j(t)$$

ist für $j = 1, \ldots, r$. Dabei ist $\|\cdot\|$ die euklidische Vektornorm im R^m. Die Kugel K_j mit dem Radius $\tau_j = \frac{1}{2} \min_{[0, t_5]} \varrho_j(t)$ um y_j enthält dann y_j und $z_j(t)$ $(0 \leqq t \leqq t_5)$ in ihrem Inneren. Andererseits ist K_j enthalten in den Kugeln um $z_j(t)$ mit den Radien $\varrho_j(t)$ für alle $t \in [0, t_5]$, denn aus $\|y - y_j\| \leqq \tau_j$ folgt

$$\|y - z_j(t)\| \leqq \tau_j + \|y_j - z_j(t)\| \leqq \tfrac{5}{6} \varrho_j(t).$$

Für $y \in Y \cap (K_1 \cup \cdots \cup K_r)$ ist damit nach (20) $f(x(t), y) \leqq 0$ $(0 \leqq t \leqq t_5)$.

Nun sei $r+1 \leqq j \leqq p$. Es ist $f(\tilde{x}, y_j) = 0$ und $f(\tilde{x}, y) < 0$ für $0 < \|y - y_j\| \leqq s$, $y \in Y$ mit hinreichend kleinem $s > 0$. Wegen $\xi^T f'(\tilde{x}, y_j) < 0$ kann $s > 0$ so klein gewählt werden, daß $\xi^T f'(\tilde{x}, y) < 0$ für $\|y - y_j\| \leqq s$ wird. Im Durchschnitt von Y mit der Kugel K_j um y_j mit dem Radius s ist dann $f(x(t), y) < 0$ (abgesehen von $f(\tilde{x}, y_j) = 0$) für $t \in [0, t_6]$ mit $t_6 > 0$.

Sei schließlich $y \in L = Y - \bigcup_{j=1}^{p} K_j$; dort ist $f(\tilde{x}, y) \leqq \sigma < 0$, denn $y_1, \ldots, y_p$ sind die einzigen Nullstellen von $f(\tilde{x}, y)$ in Y und Y ist beschränkt und abgeschlossen. Darum gibt es ein $t_7 > 0$, so daß $f(x(t), y) \leqq 0$ ist für $y \in L$ und $0 \leqq t \leqq t_7$. Insgesamt ist, wenn man $t^* = \min(t_5, t_6, t_7)$ setzt,

$$f(x(t), y) \leqq 0 \quad (y \in Y,\ 0 \leqq t \leqq t^*).$$

§ 7. Beispiel

Im Beispiel ist $n = m = 2$ und $N = 1$ bzw. $= 2$.

$$F(x) = x_2 + c\, x_1^2,$$

$$f(x, y) = (x_1 - y_1)^2 + (x_2 - y_2)^2 - 4 \quad (\leqq 0),$$

$$g_1(y) = y_1^2 + y_2^2 - 1 \quad (\leqq 0).$$

Hiermit wird $M=\{x;\ x_1^2+x_2^2\leqq 1\}$. Als Maximalstelle kommt $\tilde{x}=\begin{pmatrix}0\\1\end{pmatrix}$ in Betracht. Dann wird $f(\tilde{x}, y_1)=0$ bei $y_1=\begin{pmatrix}0\\-1\end{pmatrix}$. Weiterhin ist

$$F'(\tilde{x})=\begin{pmatrix}0\\1\end{pmatrix},\quad f'(\tilde{x}, y_1)=\begin{pmatrix}0\\4\end{pmatrix},\quad u_1=\tfrac{1}{4},$$

$$f'(\tilde{x}, y_1)=\begin{pmatrix}0\\-4\end{pmatrix},\quad g_1'(y_1)=\begin{pmatrix}0\\-2\end{pmatrix},\quad v_{11}=2.$$

Es ist

$$q(\xi, \eta_1)=2c\,\xi_1^2-\tfrac{1}{2}(\xi_1^2+\xi_2^2)+(\xi_1\eta_1+\xi_2\eta_2)+\tfrac{1}{2}(\eta_1^2+\eta_2^2).$$

$\xi\in H$ besagt $\xi_2=0$, $\eta\in H_1$ besagt $\eta_2=0$. (3) besagt schließlich $\xi_1+\eta_1=0$. Damit wird $q=(2c-1)\,\xi_1^2$. Dies ist negativ definit für $c<\frac{1}{2}$, semidefinit für $c\leqq\frac{1}{2}$. Tatsächlich hat F im Falle $c\leqq\frac{1}{2}$ bei $\tilde{x}$ ein Maximum bezüglich M, für $c>\frac{1}{2}$ jedoch nicht.

Nun werde $g_2(y)=y_1\leqq 0$ als Y beschreibende Ungleichung hinzugenommen. Dann ist

$$M=\{x;\ -\sqrt{3-x_2^2-2|x_2|}\leqq x_1\leqq\sqrt{1-x_2^2}\}.$$

Es wird $g_2'=\begin{pmatrix}1\\0\end{pmatrix}$ und damit $v_{12}=0$. $\eta\in H_1$ besagt nun $\eta_1=\eta_2=0$, wegen $H_1=\{0\}$ hat man keine Bedingungen (3), und es wird $q=(2c-\frac{1}{2})\,\xi_1^2$. Negative Definitheit hat man für $c<\frac{1}{4}$, Semidefinitheit für $c\leqq\frac{1}{4}$ und ein Maximum bei $\tilde{x}$ für $c\leqq\frac{1}{4}$. Obwohl die Bedingung (III) wegen $v_{12}=0$ nicht gilt, ist die notwendige Maximalbedingung genau für die Parameterwerte c erfüllt, für die bei $\tilde{x}$ ein Maximum liegt.

§ 8. Approximationsaufgaben

Die folgende Aufgabe der Tschebyscheff-Approximation kann als Problem des in §3 behandelten Typs geschrieben werden: *Unter den Nebenbedingungen*

$$|\varphi(y)-\Phi(y, z)|\leqq\varepsilon\quad(y\in\hat{Y}\subset R^m)\tag{21}$$

soll ε möglichst klein gemacht werden durch passende Wahl von $z\in R^{n-1}$. Dabei sei $\hat{Y}$ durch endlich viele Ungleichungen $g_\nu(z)\leqq 0$ $(\nu=1, \ldots, N)$ beschrieben. Alle auftretenden Funktionen seien zweimal stetig differenzierbar im ganzen Raum.

Wenn man (21) als

$$\left.\begin{aligned}\varphi(y)-\Phi(y, z)-\varepsilon\leqq 0\\ -\varphi(y)+\Phi(y, z)-\varepsilon\leqq 0\end{aligned}\right\}\ (y\in\hat{Y})$$

schreibt, $x=(z_1, \ldots, z_{n-1}, \varepsilon)^T$ setzt und als Y zwei Exemplare von $\hat{Y}$ wählt, hat man eine Aufgabe wie in §3. Hier sollen die hinreichenden Bedingungen des §3 für ein relatives Minimum $\tilde{\varepsilon}$ von ε an einer Stelle $z=\tilde{z}$ formuliert werden. Die Bezeichnungen entsprechen den bisher verwendeten, jedoch ist $\Phi'=(\partial\Phi/\partial z_1, \ldots, \partial\Phi/\partial z_{n-1})^T$.

(A′) *Für $y_1, \ldots, y_r\in\hat{Y}$ sei $\varphi(y_j)-\Phi(y_j, \tilde{z})=\tilde{\varepsilon}$ und für $y_{r+1}, \ldots, y_p\in\hat{Y}$ sei*

$$\varphi(y_j)-\Phi(y_j, \tilde{z})=-\tilde{\varepsilon}.$$

Im übrigen sei

$$|\varphi(y) - \Phi(y, \tilde{z})| \leqq \varepsilon \qquad (y \in \hat{Y}).$$

Es gebe reelle Zahlen $u_j > 0$ $(j = 1, \ldots, r)$, $u_j < 0$ $(j = r+1, \ldots, p)$ *mit*

$$\sum_{j=1}^{p} u_j \Phi_j' = 0, \qquad \sum_{j=1}^{p} |u_j| = 1. \tag{22}$$

(B′) *Für jedes* j $(1 \leqq j \leqq p)$ *gelte: Für* $\nu \in M_j \subset \{1, \ldots, N\}$ *und nur für diese Indizes ist* $g_\nu(y_j) = 0$. *Die Vektoren* $g_{\nu j}^{\cdot}$ *sind linear unabhängig, und es gibt reelle Zahlen* $v_{\nu j}$ $(\nu \in M_j)$, *die* $\geqq 0$ *im Falle* $1 \leqq j \leqq r$ *und* $\leqq 0$ *im Falle* $r+1 \leqq j \leqq p$ *sind, mit*

$$\varphi_j^{\cdot} - \Phi_j^{\cdot} - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}^{\cdot} = 0.$$

Auf dem linearen Teilraum

$$H_j = \{\eta; \eta^T g_{\nu j}^{\cdot} = 0 \ (\nu \in M_j)\} \subset R^m$$

ist die quadratische Form

$$Q_j(\eta) = \eta^T \Big[\varphi_j^{\cdot\cdot} - \Phi_j^{\cdot\cdot} - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}^{\cdot\cdot}\Big] \eta$$

negativ definit im Falle $1 \leqq j \leqq r$, *positiv definit im Falle* $r+1 \leqq j \leqq p$.

(C′) *Sei* $\{\eta_j^{(k)};\ k = 1, \ldots, m - m_j\}$ *eine Basis von* H_j $(j = 1, \ldots, p)$. *Durch die linearen Gleichungssysteme*

$$\eta_j^{(k)\,T} \Big[-\Phi_j^{\cdot\prime} \zeta + \Big(\varphi_j^{\cdot\cdot} - \Phi_j^{\cdot\cdot} - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}^{\cdot\cdot}\Big) \eta_j\Big] = 0$$

$$(k = 1, \ldots, m - m_j;\ j = 1, \ldots, p)$$

sind jedem $\zeta = (\zeta_1, \ldots, \zeta_{n-1}) \in R^{n-1}$ *eindeutig Vektoren* $\eta_1 \in H_1, \ldots, \eta_p \in H_p$ *zugeordnet* (N.B. das ist keine Voraussetzung, sondern Folgerung aus (B′)). *Für alle* $\zeta \in R^{n-1}$ *mit*

$$\zeta^T \Phi_1' = \cdots = \zeta^T \Phi_p' = 0 \tag{23}$$

und die jeweils zugeordneten $\eta_1, \ldots, \eta_p$ *sei*

$$q(\zeta, \eta_j) = \sum_{j=1}^{p} u_j \Big\{\zeta^T \Phi_j'' \zeta + \eta_j^T \Big(\varphi_j^{\cdot\cdot} - \Phi_j^{\cdot\cdot} - \sum_{\nu \in M_j} v_{\nu j} g_{\nu j}^{\cdot\cdot}\Big) \eta_j\Big\} < 0,$$

solange nicht $\zeta = 0$ *und daher* $\eta_1 = \cdots = \eta_p = 0$ *ist.*

Sind die Bedingungen (A′), (B′), (C′) erfüllt, so liegt bei $z = \tilde{z}$ ein relatives Minimum der Maximalabweichung

$$\varrho(z) = \operatorname*{Sup}_{y \in \hat{Y}} |\varphi(y) - \Phi(y, z)|.$$

Bemerkungen. Der lineare Teilraum H von §3 wird eigentlich durch

$$\zeta^T \Phi_1' = \cdots = \zeta^T \Phi_r' = -\zeta^T \Phi_{r+1}' = \cdots = -\zeta^T \Phi_p'$$

beschrieben. Mit (22) ergibt sich jedoch (23). Auf die Formulierung der notwendigen Bedingungen von §5 für den Spezialfall der Approximationsaufgabe soll hier verzichtet werden. Jedoch sei bemerkt, daß die lokale Kuhn-Tucker-Bedingung (7) in diesem Fall mit dem von Meinardus in [5] angegebenen Satz 84

(§ 8.1) zusammenhängt. Die Aussage, aus der (7) mit dem Lemma von Farkas folgt, besagt dasselbe wie jener Satz; allerdings sind hier nur endlich viele Extremalpunkte zugelassen und es wird eine *constraint qualification* gefordert. Daß man ohne diese hier nicht auskommt, liegt daran, daß Y nicht als beschränkt vorausgesetzt ist. Im Beispiel mit

$$\varphi(y) = \frac{1}{y_1+1}, \qquad \Phi(y,z) = z_1, \qquad g_1(y) = -y_1,$$

wo also $\hat{Y} = [0, \infty)$ und $\tilde{z}_1 = \frac{1}{2}$ mit $\tilde{\varepsilon} = \frac{1}{2}$ die eindeutig bestimmte Extremalstelle ist, gilt jener Satz von Meinardus nicht. Andererseits zeigt der Beweis jenes Satzes, daß bei beschränktem Y im Falle der Approximationsaufgabe die *constraint qualification* (I) erfüllt ist.

Beispiel. $\varphi(y) = y_1$ ist für $0 \leqq y_1 \leqq 1$ durch $\Phi(y,z) = z_1 + y_1(z_2+y_1)^2$ zu approximieren. Man findet zwei lokale Minima:

$$\tilde{z} = \begin{pmatrix} \sqrt{1/27} \\ 0 \end{pmatrix} \qquad \text{mit} \quad \tilde{\varepsilon} = \sqrt{1/27} \approx 0{,}19,$$

$$\tilde{z} = \begin{pmatrix} (10 - 7\sqrt{7})/27 \\ -2 \end{pmatrix} \quad \text{mit} \quad \tilde{\varepsilon} = \frac{7\sqrt{7}-10}{27} \approx 0{,}34.$$

In beiden Fällen sind die obigen hinreichenden Bedingungen erfüllt. Mit $\tilde{z} = (\frac{1}{2}, -1)^T$ hat man einen weiteren Punkt, an dem die lokale Kuhn-Tucker-Bedingung erfüllt ist, die notwendige Semidefinitheitsbedingung jedoch nicht.

§ 9. Verallgemeinerungen

Die in §3 angegebene Aufgabe wurde der Übersichtlichkeit wegen möglichst einfach formuliert. Einige Verallgemeinerungen liegen nahe.

a) Man kann $Y = Y_1 \cup \cdots \cup Y_s$ wählen, wobei die Y_i Teilmengen von Räumen verschiedener Dimension sind. Bei der numerischen Behandlung von Randwertaufgaben monotoner Art nach dem Prinzip der lokal optimalen Schranken [6] kann etwa Y_1 ein Bereich sein, in dem die Differentialgleichung gefordert ist, Y_2 sein Rand. Insbesondere können auch isolierte Punkte zu Y gehören oder — anders gesagt — können einzelne Ungleichungen $f_i(x) \leqq 0$ hinzutreten. Wie in diesen Fällen die Maximalbedingungen zu formulieren und zu beweisen sind, ist evident.

b) Überdies dürfen in der Beschreibung von M und Y neben Ungleichungen auch endlich viele Gleichungen vorkommen. Die zugehörigen Multiplikatoren u_i bzw. $v_{\nu j}$ sind dann keinen Vorzeichenbeschränkungen zu unterwerfen.

c) Es ist natürlich nicht wesentlich, daß die auftretenden Funktionen im ganzen Raum definiert sind.

d) Eine naheliegende, aber wohl weder für die Anwendungen noch beweistechnisch interessante Verallgemeinerung der behandelten zweistufigen Aufgaben wären mehrstufige Aufgaben.

e) Im Hinblick auf Anwendungen u.a. in der *control theory* könnte eine Verallgemeinerung der hier gefundenen Ergebnisse auf abstrakte Räume (z.B. Banachräume statt R^n und R^m) wünschenswert sein.

Literatur

1. Collatz, L.: Approximation in partial differential equations, S. 413—422 in R. E. Langer (ed.): On numerical approximation. Madison: The University of Wisconsin Press 1959.
2. — Wetterling, W.: Optimierungsaufgaben. Berlin-Göttingen-Heidelberg-New York: Springer 1966.
3. Dieudonné, J.: Foundations of modern analysis. New York-London: Academic Press 1960.
4. McCormick, G. P.: Second order conditions for constrained minima. SIAM J. Appl. Math. **15**, 641—652 (1967).
5. Meinardus, G.: Approximation von Funktionen und ihre numerische Behandlung. Berlin-Göttingen-Heidelberg-New York: Springer 1964.
6. Wetterling, W.: Lokal optimale Schranken bei Randwertaufgaben. Computing **3**, 125—130 (1968).
7. — Lösungsschranken bei elliptischen Differentialgleichungen. International Series of Numerical Mathematics **9**, 393—401 (1968).
8. — Über Minimalbedingungen und Newton-Iteration bei nichtlinearen Optimierungsaufgaben. Erscheint in: International Series of Numerical Mathematics. Basel: Birkhäuser.

Prof. Dr. W. Wetterling
Technische Hogeschool Twente
Onderafdeling der Toegepaste Wiskunde
Postbus 217
Enschede/Niederlande