2 Lokale Maxima und Minima unter Nebenbedingungen

Sit.: $U \subset \mathbb{R}^n$ offen, $x_0 \in U$

$f: U \to \mathbb{R}$

$g: U \to \mathbb{R}^m$ $\quad g(x) = \begin{pmatrix} g_1(x) \\ \vdots \\ g_m(x) \end{pmatrix}$ $\quad M = \{x \in U \mid g(x) = 0\}$

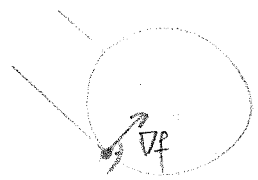
betrachte Minima von $f|_M$

x_0 heißt <u>lokales Minimum von f unter der Nebenbed. $g(x)=0$</u>, falls $g(x_0) = 0$ und falls es eine Umgebung U_0 von x_0 gibt, sodaß

$$\forall x \in U_0 \text{ mit } g(x) = 0 : \quad f(x) \geq f(x_0)$$

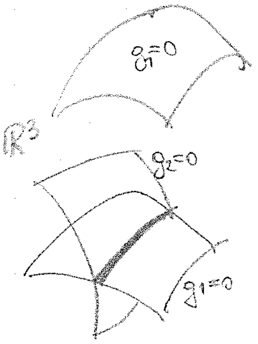
<u>Bsp</u>: $f(x_1, x_2) = x_1 + x_2$

$g(x_1, x_2) = x_1^2 + x_2^2 - 1$



typisch: $\{x \in \mathbb{R}^3 \mid g_1(x) = 0\}$ Fläche in $\mathbb{R}^3$

$\{x \in \mathbb{R}^3 \mid g_1(x) = 0,\ g_2(x) = 0\}$ Kurve in $\mathbb{R}^3$



<u>Satz</u>: (notwendige Bedingung)

Seien $f: U \to \mathbb{R}$, $g: U \to \mathbb{R}^m$ stetig diffbar

$\nabla g_1(x_0), \dots, \nabla g_m(x_0)$ seien linear unabhängig $\quad (m \leq n)$

Falls x_0 lokales Minimum von f unter NB $g(x) = 0$,

gibt es $\lambda_1, \dots, \lambda_m \in \mathbb{R}$, sodaß $\quad \nabla f(x_0) - \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0) = 0$

(λ_i heißen Lagrange-Multiplikatoren.)

Bem: zusammen mit $g(x_0) = 0$ $\quad n+m$ Gleichungen in

$n+m$ Unbekannten $x_1, \dots, x_n, \lambda_1, \dots, \lambda_m$

Bem: $\lambda_1, \dots, \lambda_m$ eindeutig bestimmt (wegen l.U. der $\nabla g_i(x_0)$)

brauchen Hilfsresultat aus linearer Algebra:

für Unterraum $V \leq \mathbb{R}^n$ bez. $V^\perp = \{u \in \mathbb{R}^n \mid \langle u,v\rangle = 0 \ \forall v \in V\}$

wissen: $\mathbb{R}^n = V \oplus V^\perp$ $\qquad$ $\langle\cdot,\cdot\rangle$ euklid. Skp.

für Matrix $G \in \mathbb{R}^{m\times n}$:

$$\operatorname{Ker} G = \{v \in \mathbb{R}^n \mid Gv = 0\} \qquad \text{Kern von } G$$

$$\operatorname{Im} G^T = \{u \in \mathbb{R}^n \mid u = G^T\mu \text{ für ein } \mu \in \mathbb{R}^m\} \qquad \text{Bild von } G^T$$

<u>HS 1:</u> $(\operatorname{Ker} G)^\perp = \operatorname{Im} G^T$

<u>Beweis:</u> Sei $u = G^T\mu \in \operatorname{Im} G^T$, $v \in \operatorname{Ker} G$

$$\langle u,v\rangle = u^Tv = \mu^T \underbrace{Gv}_{} = 0$$

daher $\operatorname{Im} G^T \subset (\operatorname{Ker} G)^\perp$

$\dim(\operatorname{Ker} G)^\perp = n - \dim \operatorname{Ker} G = \operatorname{Rang} G = \operatorname{Rang} G^T = \dim \operatorname{Im} G^T$

somit $\operatorname{Im} G^T = (\operatorname{Ker} G)^\perp$ $\square$

$v \in \mathbb{R}^n$ heißt <u>Tangentialvektor</u> an $M = \{x \in U \mid g(x) = 0\}$ in x_0, falls es einen stetig diffbaren Weg $x: I \to M$ (I offenes Intervall um ...) gibt, sodaß

$$\left.\begin{aligned} g(x(t)) &= 0 \quad \forall t \in I \\ x(0) = x_0, &\quad \dot{x}(0) = v \end{aligned}\right\} \quad (*)$$

$T_{x_0}M$ bezeichne die Menge aller Tangentialvektoren an M in x_0.

<u>HS 2:</u> $T_{x_0}M = \operatorname{Ker} G$, wobei $G = Dg(x_0) \in \mathbb{R}^{m\times n}$

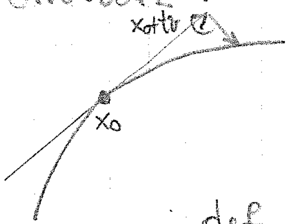
Bem: $T_{x_0}M$ ist also Vektorraum, heißt <u>Tangentialraum</u> an M in ...

<u>Beweis:</u> ($\subseteq$) Sei $x: I \to M$ stetig diffbarer Weg mit $(*)$

$$\Rightarrow \quad 0 = \frac{d}{dt}\Big|_{t=0} g(x(t)) = Dg(x_0)\,\dot{x}(0) = Gv \quad \Rightarrow v \in \operatorname{Ker} G$$

(2) zeige: zu jedem $v \in \operatorname{Ker} G$ gibt es stetig diffbaren Weg
$x : I \to \mathcal{M}$ mit $(*)$

Ansatz: $x(t) = x_0 + \underbrace{tv}_{\in \operatorname{Ker} G} + u(t)$ mit $u(t) \in (\operatorname{Ker} G)^{\perp} \underset{HS1}{=} \operatorname{Im} G^T$



$u = G^T \mu$

def. $\gamma(t,\mu) = g(x_0 + tv + G^T\mu)$

haben $\gamma(0,0) = g(x_0) = 0$

$\frac{\partial \gamma}{\partial \mu}(0,0) = \underbrace{GG^T}_{m \times m}$ invertierbar, denn:

$$GG^T\mu = 0 \Rightarrow 0 = \mu^T GG^T \mu = \|G^T\mu\|_2^2$$

$$\Rightarrow 0 = G^T\mu = \sum_{i=1}^{m} \mu_i \nabla g_i(x_0) \xRightarrow[\nabla g_i(x_0) \text{ l.u.}]{} \mu_1 = \dots = \mu_m = 0$$

nach Satz über implizite Funktionen: $\exists$ offenes Intervall I um 0
$\exists\, \mu : I \to \mathbb{R}^m$ stetig diffbar mit $\mu(0) = 0$ und

$$\gamma(t, \mu(t)) = 0 \quad \forall t \in I$$
$$\| $$
$$g(x_0 + tv + G^T\mu(t))$$

$$0 = \frac{\partial \gamma}{\partial t}(0,0) + \frac{\partial \gamma}{\partial \mu}(0,0)\,\dot\mu(0) = \overset{0}{\cancel{Gv}} + GG^T\dot\mu(0) \Rightarrow \dot\mu(0) = 0$$

$\Rightarrow x(t) := x_0 + tv + G^T\mu(t)$ erfüllt $(*)$ □

<u>Beweis des Satzes:</u> Sei x_0 lokales Minimum von f unter NB $g(x)=0$

Sei $v \in T_{x_0}\mathcal{M} \underset{HS2}{=} \operatorname{Ker} G$, sei $x : I \to \mathcal{M}$ stetig diffbaren Weg mit

$\Rightarrow t = 0$ ist lokales Minimum von $\varphi(t) := f(x(t))$

$\Rightarrow 0 = \dot\varphi(0) = f'(x_0)v = \langle \nabla f(x_0), v \rangle$, gilt $\forall v \in \operatorname{Ker} G$

$\Rightarrow \nabla f(x_0) \in (\operatorname{Ker} G)^{\perp} \underset{HS1}{=} \operatorname{Im} G^T$

$\Rightarrow \exists\, \lambda \in \mathbb{R}^m : \nabla f(x_0) = G^T\lambda = \sum_{i=1}^{m} \lambda_i \nabla g_i(x_0)$ □

Bem: geometrische Interpretation

Falls x_0 lokales Minimum (oder Maximum) von f unter NB $g(x) = 0$, so steht $\nabla f(x_0)$ orthogonal auf alle Tangentialvektoren von $M = \{x \mid g(x) = 0\}$ in x_0.

Bem: Lagrange-Funktion

$(x_0, \lambda_0) \in \mathbb{R}^{n+m}$ mit $\lambda_0 = (\lambda_1, \dots, \lambda_m)^T$ ist Sattelpunkt von

$$L(x, \lambda) = f(x) - \sum_{i=1}^{m} \lambda_i g_i(x)$$

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_i \lambda_i \frac{\partial g_i}{\partial x_j}$$

$$\frac{\partial L}{\partial \lambda_k} = g_k$$

Bem: hinreichende Bedingung

$$v^T \left(\nabla^2 f(x_0) - \sum_{i=1}^{m} \lambda_i \nabla^2 g_i(x_0) \right) v > 0 \qquad \forall \underset{\neq 0}{v} \in T_{x_0} M$$

≥ 0 auch notwendig für Minimum unter NB

(ohne Beweis)

III. Nichtlineare Optimierung mit NB

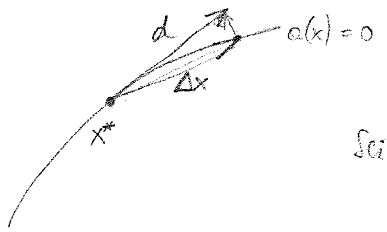
§1 Minimierung unter Gleichheits-NB

$f(x)$ minimal! NB $a(x)=0$ $\qquad f: \mathbb{R}^n \to \mathbb{R}$ $\quad C^\infty$

$\qquad a: \mathbb{R}^n \to \mathbb{R}^m$

$g(x) = \frac{\partial f}{\partial x}(x)^T \qquad A(x) = \frac{\partial a}{\partial x}(x) \quad m \times n$

$H(x) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right)_{i,j=1}^n$



$\cancel{a(x^* + \Delta x) = 0}$

Sei $x^* \in \mathbb{R}^n$ zulässig: $a(x^*) = 0$

betrachten zulässige Punkte um x^*: $\quad x = x^* + \Delta x \qquad a(x) = 0$

$$\underset{\overset{\shortparallel}{0}}{a(x)} = a(x^* + \Delta x) = \underset{\overset{\shortparallel}{0}}{a(x^*)} + A(x^*) \cdot \Delta x + O(\|\Delta x\|^2)$$

Tangentialvektoren in x^*: $\quad d \in \operatorname{Ker} A(x^*)$

benötigen als Vorbereitung Hilfsmittel, um Krümmung der Mf. $\{x \mid a(x)=0\}$ behandeln zu können

Sit. wie oben

<u>HS</u>: $A(x^*)$ habe l.u. Zeilen

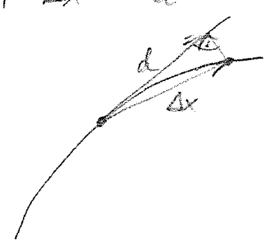
Dann ist für geeignete Null-Umgebungen $U, V \subset \mathbb{R}^n$ folgende Abb. ein Diffeomorphismus

$$\{\Delta x \in \mathbb{R}^n : a(x^* + \Delta x) = 0\} \cap U \longrightarrow \operatorname{Ker} A(x^*) \cap V$$

$$\Delta x \longmapsto P \cdot \Delta x = d$$

wobei P og. Projektion auf $\operatorname{Ker} A(x^*)$

Es ist weiters $d = \Delta x + O(\|\Delta x\|^2)$



<u>Beweis</u>: a) Sei Δx mit $a(x^* + \Delta x) = 0$, $\|\Delta x\|$ gen. klein

setze $d = P \cdot \Delta x$, wissen: d ist Lösung von

$$\begin{pmatrix} I & A(x^*)^T \\ A(x) & 0 \end{pmatrix} \begin{pmatrix} d \\ \mu \end{pmatrix} = \begin{pmatrix} \Delta x \\ 0 \end{pmatrix}$$

d.h. $\quad d = \left(I - A^T (A A^T)^{-1} A(x^*)\right) \cdot \Delta x$

$$A(x) := \left(\frac{\partial a_i(x^*)}{\partial x}\right)_{i \in E \cup I_0} = \left(\nabla a_i(x^*)\right)^T_{i \in E}$$

$$\mathbb{R}^n = \operatorname{Ker} A(x) \oplus \operatorname{Im} A(x)^T$$

Zeige: $\exists \mu \in \mathbb{R}^n$: $\underbrace{a(x^* + \varepsilon d + A^T(x^*)\mu)}_{= \varphi(\varepsilon, \mu)} = 0$

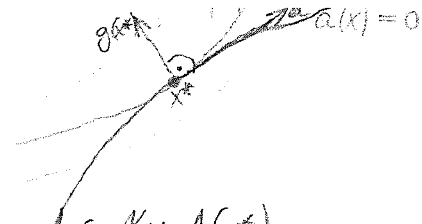
$$\varphi(0,0) = 0$$

$$\frac{\partial \varphi}{\partial \mu}(0,0) = A(x^*)\, A^T(x^*)$$ inv., falls $A(x^*)$ l.u. Zeilen hat

$$\frac{\partial \varphi}{\partial \varepsilon}(0,0) = A(x^*)\, d = 0$$

$\Rightarrow$ lokal für kleine ε $\exists_1\, \mu(\varepsilon) = \underbrace{\left(\frac{\partial \varphi}{\partial \mu}\right)^{-1} \left(\frac{\partial \varphi}{\partial \varepsilon}\right)^{(0,0)}}_{0} + O(\varepsilon^2)$

$$\mu = O(\varepsilon^2)$$



daher $\quad (g(x^*), \varepsilon d) = O(\varepsilon^2) \qquad \forall \varepsilon > 0$

$\Rightarrow \quad (g(x^*), d) = 0$ gilt für alle $d \in \operatorname{Ker} A(x^*)$

somit $\quad g(x^*) = (\operatorname{Ker} A(x^*))^\perp \underset{\text{l.A.}}{=} \operatorname{Im} A(x^*)^T$

$\Rightarrow \quad \exists_1 \lambda^* \in \mathbb{R}^m : \qquad g(x^*) = A(x^*)^T \lambda^*$

"Lagrange-Multiplikatoren" $\qquad$ (λ^* eindeutig, weil ~~Spalten von~~ $A(x^*)^T$ l.u.)

erhalten ~~Gls~~ Gls. als notwendige Bed. für Minimum:

$$\begin{cases} g(x) - A^T(x)\lambda = 0 \\ -a(x) = 0 \end{cases} \qquad F(x,\lambda) = 0$$

kann aufgefaßt werden als

$$\nabla L(x,\lambda) = 0$$

mit $\quad L(x,\lambda) = f(x) - a(x)^T \lambda \qquad$ Lagrange-Funktion

$$\left(\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} - \sum_{k=1}^m \frac{\partial a_k}{\partial x_j} \lambda_k \, , \quad \frac{\partial L}{\partial \lambda_i} = -a_i \right)$$

Newton-Verfahren: benötige Matrix

$$\begin{pmatrix} M(x,\lambda) & -A^T(x) \\ -A(x) & 0 \end{pmatrix}$$

mit $\quad M(x,\lambda) = \left(\frac{\partial^2 L}{\partial x_i \partial x_j}(x) \right)_{i,j=1}^n = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(x) - \sum_{k=1}^m \frac{\partial^2 a_k}{\partial x_i \partial x_j}(x) \lambda_k \right)_{i,j} \quad$ symm.

Hesse-Matrix der Lagrange-Funktion bzgl. x

§2 Lagrange-Newton Methode (Sequential Quadratic Programming)

verwende NV zur numer. Lösung von $\nabla L(x,\lambda) = 0$

$$\Leftrightarrow \begin{cases} g(x) - A^T(x)\lambda = 0 \\ -a(x) = 0 \end{cases}$$

Seien x_0, λ_0 geg.
Löse

$$\begin{pmatrix} M_k & -A_k^T \\ -A_k & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \Delta \lambda_k \end{pmatrix} = \begin{pmatrix} -g(x_k) + A_k^T \lambda_k \\ + a(x_k) \end{pmatrix} \qquad k = 0,1,2,\ldots$$

$\left(\text{wobei} \quad M_k = M(x_k, \lambda_k) \; , \quad A_k = A(x_k) \right)$

und setze

$$x_{k+1} = x_k + \Delta x_k \quad , \quad \lambda_{k+1} = \lambda_k + \Delta \lambda_k$$

äquiv: Lösung von $\begin{pmatrix} M_k & -A_k^T \\ -A_k & 0 \end{pmatrix} \begin{pmatrix} \Delta x_k \\ \lambda_k \end{pmatrix} = \begin{pmatrix} -g_k \\ a_k \end{pmatrix}$,

kann aufgefaßt werden als

$$(*) \begin{cases} \text{minimiere } \underline{\text{quadrat.}} \text{ Funktion} & \frac{1}{2} \Delta x^T M_k \Delta x + g_k \\ \text{unter } \underline{\text{linearen}} \text{ NB} & A_k \cdot \Delta x + a_k = 0 \end{cases}$$

lokal quadratische Konvergenz

<u>Quasi-Newton-Variante</u> (möchte nicht 2. Ableitungen in M_k berechnen)

ersetze M_k durch QN-Approximation (BFGS):

$$M_{k+1} = M + \frac{\Delta g \cdot \Delta g^T}{\Delta x^T \Delta g} - \frac{M \Delta x \cdot \Delta x^T M}{\Delta x^T M \Delta x} \Big|_k$$

mit $\Delta g_k = g(x_{k+1}) - g(x_k) - (A(x_{k+1})^T - A(x_k)^T) \cdot \lambda_{k+1}$, $\Delta x_k = x_{k+1} - x_k$

17.6.1991

§3 Straffunktionen ("Penalty functions")

(P) minimiere $f(x)$ unter NB $a(x) = 0$ Lösung: x^*

(P_ε) minimiere $\underbrace{f(x) + \frac{1}{\varepsilon}\|a(x)\|_2^2}_{=:\ \phi(x,\varepsilon)}$ $(\varepsilon > 0)$ Lösung: $x(\varepsilon)$

Abweichen von NB wird bestraft

löse Folge von Minimierungsproblemen <u>ohne</u> NB:

(i) wähle Folge $\{\varepsilon_k\} \to 0$ z. B. $\varepsilon_k = 10^{-k}$

(ii) finde (lokales) Minimum von $\phi(\cdot,\varepsilon)$: $x_k = x(\varepsilon_k)$

(iii) breche ab, falls $a(x_k)$ genügend klein

<u>Satz</u>: (Konvergenz der Straffunktionsmethode)

Sei $f^* := \inf\{f(x) : a(x) = 0\} \in \mathbb{R}$

Falls $\varepsilon_k \searrow 0$ (monoton) und x_k <u>globales</u> Minimum von $\phi(\cdot,\varepsilon_k)$, so gilt

(i) $\|a(x_k)\| \searrow 0$ (monoton)

(ii) jeder Häufungspunkt x^* von $\{x_k\}$ ist globales Minimum von f unter NB $a = 0$.

<u>Beweis</u>: (i) Sei $l > k$

haben $\phi(x_k,\varepsilon_k) \le \phi(x_l,\varepsilon_k) \le \phi(x_l,\varepsilon_l) \le \phi(x_k,\varepsilon_l)$

es folgt: $\{\phi(x_k,\varepsilon_k)\}_k$ monoton $\nearrow$

$$\left.\begin{aligned} 0 &\le \phi(x_k,\varepsilon_l) - \phi(x_k,\varepsilon_k) = \left(\frac{1}{\varepsilon_l} - \frac{1}{\varepsilon_k}\right)\|a(x_k)\|^2 \\ 0 &\ge \phi(x_l,\varepsilon_l) - \phi(x_l,\varepsilon_k) = \left(\frac{1}{\varepsilon_l} - \frac{1}{\varepsilon_k}\right)\|a(x_l)\|^2 \end{aligned}\right\} -$$

$$\Rightarrow \quad 0 \le \underbrace{\left(\frac{1}{\varepsilon_l} - \frac{1}{\varepsilon_k}\right)}_{>0}\left(\|a(x_k)\|^2 - \|a(x_l)\|^2\right) \quad \Longrightarrow \quad \{\|a(x_k)\|\} \text{ monoton}$$

$$\Phi(x_k, \varepsilon_k) \le \Phi(x_\ell, \varepsilon_k) \Rightarrow f(x_k) \le f(x_\ell) \Rightarrow \{f(x_k)\} \text{ monoton}$$

$$\Phi(x_k, \varepsilon_k) = \inf_x \Phi(x, \varepsilon_k) \le \inf_{x: a(x)=0} \Phi(x, \varepsilon_k) = \inf_{x: a(x)=0} f(x) = f^*$$

$$\| $$

$$\underbrace{f(x_k)}_{\ge f(x_0^*)} + \frac{1}{\varepsilon_k} \|a(x_k)\|^2 \quad \underset{\varepsilon_k \to 0}{\Longrightarrow} \quad \|a(x_k)\| \searrow 0$$

(ii) Sei x^* HP von $\{x_k\}$, dann: $a(x^*) = 0$

$$f(x_k) \le \Phi(x_k, \varepsilon_k) \le f^*$$

$$\Rightarrow f(x^*) \le f^* = \inf_{a(x)=0} f(x) \le f(x^*) \Rightarrow f(x^*) = f^*$$

$\Rightarrow$ Beh. (ii) □

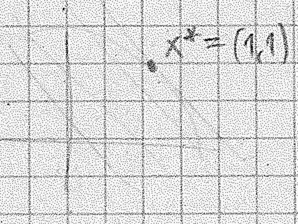
Bem: Unter den hinreichenden Bedingungen von S.2, §1 kann man zeigen

$$x(\varepsilon) = x^* + O(\varepsilon)$$

! <u>aber</u>: praktische Schwierigkeiten bei Minimierung von $\Phi(\cdot, \varepsilon)$ für kleine ε

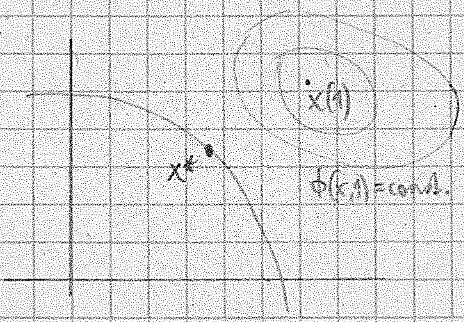
Bsp: $-x_1 - x_2 = \min!$

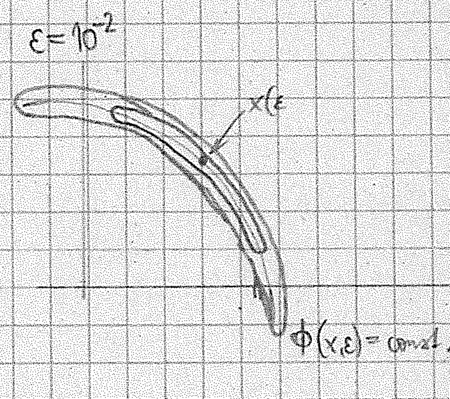
$1 - x_1^2 - x_2^2 = 0$



$$\Phi(x, \varepsilon) = -x_1 - x_2 + \frac{1}{\varepsilon}\left(1 - x_1^2 - x_2^2\right)^2$$

$\varepsilon = 1$:





Penalty-Funktionsmethode i.a. nur für grobe Näherungen geeignet

schlecht konditioniert
schwierig zu lösen

§4 Erweiterte Lagrangefunktions-Methoden („augmented Lagrangian“)

§3: minimiere $\Phi(x,\varepsilon) = f(x) + \frac{1}{\varepsilon}\cdot\frac{1}{2}\sum_{i=1}^{m} a_i(x)^2$ $\qquad \varepsilon \downarrow 0$ (bzw $\frac{1}{\varepsilon}\to\infty$)

§4: $\Phi(x,\theta,\sigma) = f(x) + \frac{1}{2}\sum_{i=1}^{m} \sigma_i\,(a_i(x)-\theta_i)^2$

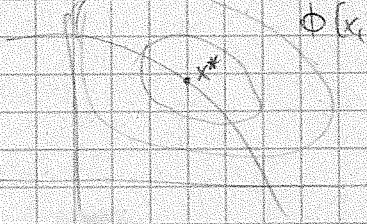
$$= f(x) + \frac{1}{2}\,(a(x)-\theta)^T S\,(a(x)-\theta) \qquad S = \begin{pmatrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_m\end{pmatrix}$$

$$\sigma = \begin{pmatrix}\sigma_1\\ \vdots\\ \sigma_m\end{pmatrix},\ \theta = \begin{pmatrix}\theta_1\\ \vdots\\ \theta_m\end{pmatrix}$$

möchte θ so wählen, daß für <u>endliches</u> σ das Minimum von $\Phi(\cdot,\theta,\sigma)$ bei der Lösung x^* des Minimierungsproblems mit NB auftritt $\rightarrow$ vermeide Schlechtkonditioniertheit

<u>Bsp</u>: $\quad -x_1 - x_2 = \min!$

$\qquad 1 - x_1^2 - x_2^2 = 0$



z.B.: $\sigma = 1$

$$\Phi(x,\theta,\sigma) = -x_1 - x_2 + \frac{1}{2}\left(1 - x_1^2 - x_2^2 - \theta\right)^2$$

Minimum bei $x^* = (1,1)$, falls $\theta = \cancel{\frac{1}{\sqrt{2}}}$

führe neue Parameter ein: $\lambda_i = \sigma_i\theta_i$ $\quad (i=1,\dots,m)$ $\qquad (\lambda = S\theta)$

$$\Phi(x,\sigma,\theta) = f(x) - \underbrace{\theta^T S}_{\lambda^T}\cdot a(x) + \frac{1}{2}\,a(x)^T S\,a(x) + \underbrace{\frac{1}{2}\,\theta^T S\,\theta}_{\text{unabh. von } x}$$

minimiere daher (bzgl. x)

$$\phi(x,\lambda,\sigma) := \underbrace{f(x) - \lambda^T a(x)} + \frac{1}{2}\,a(x)^T S\,a(x)$$

„<u>erweiterte Lagrange-funktion</u>“

bez. Minimum $x(\lambda,\sigma)$ $\qquad$ (Hestenes 1947)

(in §3: nur f „erweitert“)

<u>Satz 1:</u> Es seien die hinreichenden Bedingungen von S. 2, §1 bei x^*, λ^* erfüllt

(Erinn: $\nabla L(x^*, \lambda^*) = 0$, $d^T M(x^*, \lambda^*) d > 0 \quad \forall \underset{\neq 0}{d} \in \operatorname{Ker} A(x^*)$)

dann: $\exists\, \sigma_0 \geq 0 \quad \forall\, \sigma > \sigma_0$:

x^* ist isoliertes lokales Minimum von $\phi(\cdot, \lambda^*, \sigma)$

m. a. W. : $\underline{x^* = x(\lambda^*, \sigma)}$

<u>Beweis:</u> (setzen voraus : rang $A(x^*) = m$, Satz gilt auch ohne diese VS)

$$\phi(x, \lambda, \sigma) = L(x, \lambda) + \tfrac{1}{2}\, a(x)^T S\, a(x) \qquad S = \operatorname{diag}(\sigma_i)$$

$$\frac{\partial \phi}{\partial x}(x, \lambda, \sigma) = \underbrace{\frac{\partial L}{\partial x}(x, \lambda)}_{=0} + \underbrace{a(x)}_{=0}{}^T S A(x) \qquad A = \frac{\partial a}{\partial x}$$

für $x = x^*, \lambda = \lambda^*$

$$\frac{\partial^2 \phi}{\partial x^2}(x, \lambda, \sigma) = \underbrace{\frac{\partial^2 L}{\partial x^2}(x, \lambda)}_{M(x,\lambda)} + A(x)^T S A(x) + \underbrace{\{\ldots\}}_{\substack{=0\\ x=x^*}} =: M_\sigma(x, \lambda)$$

zeigen noch : $M_\sigma(x^*, \lambda^*)$ positiv definit

Sei $0 \neq u \in \mathbb{R}^n$, zeige $u^T M_\sigma(x^*, \lambda^*)\, u > 0$

zerlege $u = \underset{\operatorname{Ker} A(x^*)}{\underset{\cap}{v}} + \underset{(\operatorname{Ker} A(x^*))^\perp = \operatorname{Im} A(x^*)^T}{\underset{\cap}{w}} = v + A(x^*)^T \mu$

(lasse Argumente x^*, λ^* weg) :

$$u^T M_\sigma u = v^T M v + 2 v^T M w + w^T M w + w^T A^T S A w$$

$$= v^T M v + 2 v^T M A^T \mu + \mu^T A M A^T \mu + \mu^T (AA^T) S (AA^T) \mu$$

$$\geq \underset{\substack{\vee\\ 0}}{\alpha} \|v\|^2 - 2 \|v\| \cdot \|MA^T\| \cdot \|\mu\| - \|AMA^T\| \cdot \|\mu\|^2 + s \|(AA^T)\mu\|^2$$

VS S. 2, §1 $\qquad (s = \min_i \sigma_i)$

A hat vollen Rang $\Rightarrow$ $B = AA^T$ invertierbar

$\|\mu\| = \|B^{-1} B\mu\| \leq \|B^{-1}\| \cdot \|B\mu\|$

$\Rightarrow \|B\mu\| \geq \frac{1}{\|B^{-1}\|} \|\mu\|$

quadratische Form $\underset{\vee \atop 0}{\alpha} \xi^2 - 2\beta \cdot \xi\eta + \gamma\eta^2 > 0 \quad \forall (\xi,\eta) \neq 0$

falls $\alpha > 0,\ \gamma > 0,\ \alpha\gamma > \beta^2$ $\quad \left(\det\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix} > 0 \right)$

hier: $\underset{\vee \atop 0}{\alpha} \left(\sigma - \|AMA^T\| \right) > \|MA^T\|^2$

erfüllt, falls σ genügend groß

somit $\frac{\partial \phi}{\partial x}(x^*, \lambda^*, \sigma) = 0$

$\frac{\partial^2 \phi}{\partial x^2}(x^*, \lambda^*, \sigma)$ pos. def, falls σ gen. groß

$\Longrightarrow$ x^* striktes lokales Minimum von $\Phi(\cdot, \lambda^*, \sigma)$ □

also: falls λ^* bekannt, so ist durch obigen Satz 1 die Minimierung mit NB zurückgeführt auf eine Minimierung ohne NB

wie berechne λ^* ?

Satz 2: (duales Problem)

Es seien die hinreichenden Bed. von S. 2, §1 in x^*, λ^* erfüllt

Dann ist λ^* ~~isoliertes~~ lokales Maximum von

$$\psi_\sigma(\lambda) := \Phi(x(\lambda,\sigma), \lambda, \sigma),$$

(global, falls $x(\lambda,\sigma)$ globales Minimum: dann $\Phi(x(\lambda,\sigma),\lambda,\sigma) = \min_x \Phi(x,\lambda,\sigma$

<u>Beweis:</u> (lasse festes σ in Notation weg, σ gen. groß)

$x(\lambda)$ Lösung von $\frac{\partial \phi}{\partial x}(x,\lambda) = 0$

da $\frac{\partial^2 \phi}{\partial x^2}(x^*, \lambda^*)$ pos. def. (Beweis von S.1)

existiert nach impl. FT C^∞-Abb $\lambda \mapsto x(\lambda)$ in Umgebung von λ^*

$$\psi(\lambda) = \phi(x(\lambda), \lambda) \le \phi(x^*, \lambda) \underset{\underset{a(x^*)=0}{\uparrow}}{=} \phi(x^*, \lambda^*) = \psi(\lambda^*)$$

$\Rightarrow$ λ^* (lokales) Maximum $\square$

numerisch: löse duales Problem für λ^* : maximiere $\psi(\lambda)$!

benötige $\nabla\psi(\lambda)$, evtl. Hessematrix $\frac{\partial^2 \psi}{\partial \lambda^2}$

$$\frac{\partial \psi}{\partial \lambda}(\lambda) = \underbrace{\frac{\partial \phi}{\partial x}(x(\lambda), \lambda)}_{=0} \cdot \frac{\partial x}{\partial \lambda} + \frac{\partial \phi}{\partial \lambda}(x(\lambda), \lambda)$$

$$\phi(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i a_i(x) + \tfrac{1}{2} a^T(x) S a$$

$$\frac{\partial \phi}{\partial \lambda_i} = -a_i$$

somit $\nabla\psi(\lambda)^T = -a(x(\lambda))$

Weitere Rechnung ergibt

$$\frac{\partial^2 \psi}{\partial \lambda^2}(\lambda) = -\left(A M_\sigma^{-1} A^T\right)(x(\lambda), \lambda)$$

und $\left(A M_\sigma^{-1} A^T\right)^{-1} = \left(A M^{-1} A^T\right)^{-1} + S$

Für große σ ist das $\approx S$. Dies motiviert die Iteration (anstelle Newton angewandt auf $\nabla\psi(\lambda) = 0$):

$$\underline{\lambda_{k+1} = \lambda_k - S a(x(\lambda))}$$ (Hestenes, Powell 1969)

möglicher Algorithmus (Powell)

1) initialisiere λ, σ

2) finde ~~das~~ Minimum $x(\lambda,\sigma)$ von $\Phi(x,\lambda,\sigma)$

(Varianten: wie genau?

Minimalvariante: 1 Schritt eines Quasi-Newton-Verfahrens)

setze $a := a(x(\lambda,\sigma))$

3) für Komponenten $i = 1, \ldots, m$: (außer im ersten Schritt)

falls $|a_i| > \frac{1}{4} |\tilde{a}_i|$ setze $\sigma_i := 10\sigma_i$

falls $|a_i| < \frac{1}{100} |\tilde{a}_i|$ setze $\sigma_i := \frac{1}{10}\sigma_i$

4) setze $\lambda := \lambda - Sa$

$\tilde{a} := a$

gehe nach 2), (oder breche ab, falls $\|Sa\|_\infty \le TOL$)

zielt auf Konvergenzrate $\frac{1}{4}$ oder besser

evtl. Schwierigkeit in (ii): schlechtkonditioniert, falls σ groß wir

andere Möglichkeit: Bei Berechnung von $x(\lambda,\sigma)$ mit Quasi-Newt

erhalte Approximation B_σ an $M_\sigma^{-1} = \left(\frac{\partial^2 \Phi}{\partial x^2}\right)^{-1}$

verwende diese in Berechnung von λ_{k+1}:

$$\lambda_{k+1} = \lambda_k + (AB_\sigma A^T)^{-1} a(x_k)$$

(aufwendiger, aber konvergiert schneller

interessant, falls nur wenige NB)

§5 Minimierung mit Ungleichheits-Nebenbedingungen

(P) minimiere $f(x)$ unter NB $a_i(x) \geq 0 \quad i \in \mathcal{U}$

$a_i(x) = 0 \quad i \in \mathcal{G}$ $\qquad \mathcal{U} \cup \mathcal{G} = \{1, \dots, m\}$

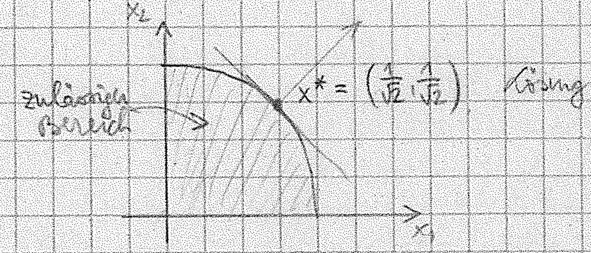
$(f: \mathbb{R}^n \to \mathbb{R}, \quad a_i: \mathbb{R}^n \to \mathbb{R} \quad C^\infty)$

Bsp: $-x_1 - x_2 = \min!$

NB 1: $1 - x_1^2 - x_2^2 \geq 0$

2: $x_1 \geq 0$

3: $x_2 \geq 0$



im Lösungspunkt x^*:

$1 - x_1^{*2} - x_2^{*2} \underline{= 0}$ bzgl. der NB1 zulässige Richtungen: alle Richtungen, die in das Innere des Einheitskreises zeigen

d mit $-\binom{1}{1}^T d \geq 0$

$x_1^* \underline{> 0}$ alle Richtungen zulässig

$x_2^* \underline{> 0}$ — " —

} "inaktive" NB, spielen lokal um x^* keine Rolle

$\nabla f(x^*) = \left(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right) = \frac{1}{2}\left(-2 \cdot \frac{1}{\sqrt{2}}, -2\frac{1}{\sqrt{2}}\right) = \underbrace{\frac{1}{2}}_{\text{positives Vielfaches}} \nabla a_1(x^*)$

---

definieren allgemein für zulässiges x:

$\mathcal{A}(x) = \{ i \in \mathcal{U} \mid a_i(x) = 0 \}$ <u>aktive</u> Ungl. NB in x

$\mathcal{I}(x) = \{ i \in \mathcal{U} \mid a_i(x) > 0 \}$ <u>inaktive</u> — " —

Notwendige Bedingungen für lokales Minimum liefert folgender grundlegender S

<u>Satz 1</u>: (Kuhn - Tucker 1951) Sei x^* lokales Minimum von (P), und gelte Regularitäts-VS: $\nabla a_i(x^*)$ für $i \in \mathcal{A}(x^*) \cup \mathcal{G}$ sind linear unabhängig

Dann gilt: $\exists_1 \lambda \in \mathbb{R}^m$ mit $\nabla f(x^*) = \sum_{i=1}^m \lambda_i \nabla a_i(x^*)$

$\lambda_i \geq 0 \quad \forall i \in \mathcal{U}$ (Ungl. NB)

$\lambda_i a_i(x^*) = 0 \quad \forall i = 1, \dots, m$ (Komplement

Beweis in mehreren Schritten:

Sei $x = x^* + \Delta x$, Δx "klein"

$$f(x) = f(x^*) + \underbrace{\nabla f(x^*)\cdot \Delta x}_{\text{möchte } < 0\ !} + O\left(\|\Delta x\|^2\right)$$

<u>Abstiegsrichtung</u>: $d \in \mathbb{R}^n$ mit $\nabla f(x^*) \cdot d < 0$

$$a_i(x) = \underbrace{a_i(x^*)}_{\substack{\| \\ 0 \\ \text{für } i \in \mathcal{A}(x^*) \cup \mathcal{G}}} + \nabla a_i(x^*) \cdot \Delta x + O\left(\|\Delta x\|^2\right) \begin{array}{ll} \geq 0 & i \in \mathcal{U} \\ = 0 & i \in \mathcal{G} \end{array}$$

<u>zulässige Richtung</u>: $d \in \mathbb{R}^n$ mit $\nabla a_i(x^*) \cdot d \geq 0 \quad \forall i \in \mathcal{A}(x^*)$

$\nabla a_i(x^*)\, d = 0 \quad \forall i \in \mathcal{G}$

(jede beliebige Richtung ist bzgl inaktiver NB zulässig)

<u>HS 2</u>: Falls x^* lokales Minimum von (P), so gibt es in x^* keine zulässige Abstiegsrichtung.

<u>Beweis</u>: indirekt: Sei $d \in \mathbb{R}^n$ mit $\nabla f(x^*)\, d < 0$ (Abstiegsrichtun

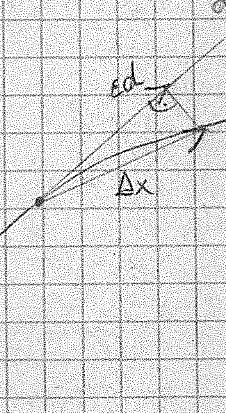
$\nabla a_i(x^*)\, d \geq 0 \quad \forall i \in \mathcal{A}(x^*)$ (zul

$= 0 \quad \forall i \in \mathcal{G}$

Sei $\mathcal{A}_0 = \{ i \in \mathcal{A}(x^*) \mid \nabla a_i(x^*)\, d = 0 \}$

$\mathcal{A}_+ = \{ \qquad > \qquad \} = \mathcal{A}(x^*) \setminus \mathcal{A}_0$

nach VS: $\nabla a_i(x^*)$, $i \in \mathcal{A}_0$, sind linear unabh.



aus HS, §1: Zu gen. kleinem $\varepsilon > 0$ gibt es $\Delta x = \varepsilon d + O(\varepsilon^2)$

sodaß für $x = x^* + \Delta x$ gilt: $a_i(x) = 0 \quad \forall i \in \mathcal{A}_0$

Sei nun $i \in \mathcal{A}_+$:

$$a_i(x) = \underbrace{a_i(x^*)}_{0} + \varepsilon \underbrace{\nabla a_i(x^*) d}_{> 0} + O(\varepsilon^2) > 0 \,, \quad \varepsilon > 0 \text{ gen. klein}$$

ebenso $a_i(x) > 0$ für inaktive NB $i \in \mathcal{I}(x^*)$

somit : $x = x^* + \Delta x = x^* + \varepsilon d + O(\varepsilon^2)$ zulässig

aber :

$$f(x) = f(x^*) + \varepsilon \underbrace{\nabla f(x^*) d}_{<0} + O(\varepsilon^2) < f(x^*) \,, \quad \varepsilon \text{ klein}$$

$\Longrightarrow$ x^* kein lokales Minimum $\square$

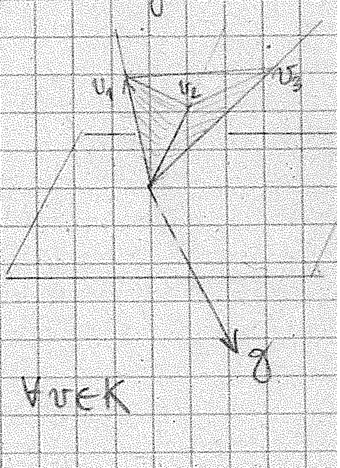
<u>HS 3</u>: (Lemma von Farkas) Seien $g, v_1, \dots, v_k \in \mathbb{R}^n$.

Dann sind äquivalent:

(1) $\nexists\, d \in \mathbb{R}^n$: $(g, d) < 0$ und
$(v_i, d) \geq 0 \quad \forall i = 1, \dots, k$

(2) $\exists\, \lambda_i \geq 0 \; (i = 1, \dots, k)$: $\quad g = \sum_{i=1}^{k} \lambda_i v_i$

<u>Beweis</u>: bezeichne $K = \left\{ \sum_{i=1}^{k} \lambda_i v_i \;\middle|\; \lambda_i \geq 0 \right\}$ "Kegel"



K konvex, abgeschlossen (Ü)

Aussage kann umformuliert werden als:

$$g \notin K \iff \exists\, d \in \mathbb{R}^n : \; (g, d) < 0, \quad (v, d) \geq 0 \;\; \forall v \in K$$

geometrische Interpretation: fasse d auf als Normalvektor der Hyperebene $\{ x \in \mathbb{R}^n \mid (x, d) = 0 \}$

$g \notin K \iff \exists$ Hyperebene, die g und K trennt (s. Bild!)

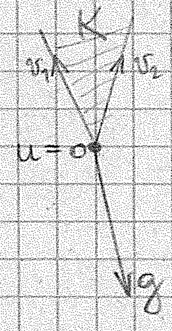
anschaulich klar

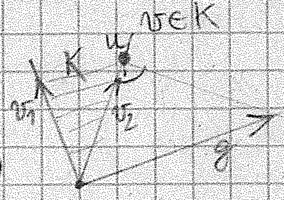
Beweis: $(\Leftarrow)$ klar

$(\Rightarrow)$ Sei $g \notin K$

Sei $u \in K$ jener Punkt in K mit dem kleinsten Abstand zu g:

$$\|u-g\| = \min_{v \in K} \|v-g\|$$ (existiert, weil K abgeschlossen (!

(i) 

(ii) 

<u>2 Fälle</u>: (i) $u=0$

dann $\forall v \in K$: $\quad \frac{d}{d\lambda}\|\lambda v - g\|^2\Big|_{\lambda=0} \geq 0$

$$\| \quad \frac{d}{d\lambda}\left[\lambda^2\|v\|^2 - 2\lambda(v,g) + \|g\|^2\right]\Big|_{\lambda=0}$$

$$\| \quad 2\lambda\|v\|^2 - 2(v,g)\Big|_{\lambda=0} = -2(v,g)$$

$\Rightarrow \quad (v,g) \leq 0 \quad \forall v \in K$

wähle $d=-g$: $\quad (g,d)<0 \quad , \quad (v,d) \geq 0 \quad \forall v \in K$

(ii) $u \neq 0$

wegen $\lambda u \in K \quad \forall \lambda \geq 0$ und $\|\lambda u - g\|^2$ minimal bei $\lambda = 1$:

$$0 = \frac{d}{d\lambda}\|\lambda u - g\|^2\Big|_{\lambda=1} \underset{\text{s. oben}}{=} 2(u,u) - 2(u,g) = 2(u,u-g)$$

$\Rightarrow \quad (u,u-g) = 0$

setze $\underline{d = u-g}$: $\quad$ dann $u = g+d$

$$0 = (u,u-g) = (g+d,d) = (g,d) + \underbrace{(d,d)}_{>0} \Rightarrow \underline{(g,d)<0}$$

Sei $v \in K$ bel.

K konvex $\Rightarrow$ $u + \theta(v-u) \in K \quad \forall \theta \in (0,1)$

$$\Rightarrow \quad \| u + \theta(v-u) - g \|^2 \geq \cancel{\| u-g \|^2}$$

$$\| u + \theta(v-u) - g \|^2 = \cancel{\| u-g \|^2} + 2\theta\,(u-g, v-u) + \theta^2 \|v-u\|^2$$

dividiere durch θ, lasse $\theta \searrow 0$:

$$(\underbrace{u-g}_{d},\ v-u) \geq 0$$

$$(u-g,\ v-u) = (d,v) - \underbrace{(d,u)}_{0} = (d,v)$$

$$\Rightarrow \quad \underline{(d,v) \geq 0 \quad \forall v \in K} \qquad \square$$

Beweis von Satz 1: x^* lokales Minimum

$$\underset{HS\,2}{\Longrightarrow} \quad \nexists\, d \in \mathbb{R}^n : \quad \nabla f(x^*) \cdot d < 0 \qquad (\nabla f(x^*)^T, d) < 0$$

$$\nabla a_i(x^*) \cdot d \geq 0 \quad \forall i \in \mathcal{A}(x^*)$$

$$\left.\begin{array}{r} \nabla a_i(x^*)\, d \geq 0 \\ -\nabla a_i(x^*)\, d \geq 0 \end{array}\right\} \forall i \in \mathcal{G}$$

$$\underset{HS\,3}{\Longrightarrow} \quad \exists\ \lambda_i \geq 0 \ (i \in \mathcal{A}(x^*))$$

$$\lambda_i^+, \lambda_i^- \geq 0 \quad (i \in \mathcal{G}) \qquad \text{mit}$$

$$\nabla f(x^*) = \sum_{i \in \mathcal{A}(x^*)} \lambda_i \nabla a_i(x^*) + \sum_{i \in \mathcal{G}} \underbrace{(\lambda_i^+ - \lambda_i^-)}_{=:\lambda_i} \nabla a_i(x^*)$$

setze $\lambda_i = 0$ für $i \in \mathcal{I}(x^*)$

damit: $\displaystyle \nabla f(x^*) = \sum_{i=1}^{m} \lambda_i \nabla a_i(x^*)$

$$\lambda_i \geq 0 \qquad i \in \mathcal{U} = \mathcal{A}(x^*) \cup \mathcal{I}(x^*)$$

$$\lambda_i = 0 \quad \text{für } i \text{ mit } a_i(x^*) > 0 \Rightarrow \lambda_i a_i(x^*) = 0 \quad \forall i$$

möchten notwendige und hinreichende Bedingungen für Minimum von (P) mittels <u>zweiter</u> Ableitungen

x^* erfülle Kuhn-Tucker-Bed: $\nabla f(x^*) = \sum_{i=1}^{m} \lambda_i^* \nabla a_i(x^*)$

$\lambda_i^* \geq 0 \quad \forall i \in \mathcal{U}, \quad \lambda_i^* a_i(x^*) = 0 \quad \forall i$

$\overset{0}{\neq}$ $d \in \mathbb{R}^n$ heißt <u>strikt zulässige Richtung</u> in x^* :

$\Longleftrightarrow$ d zulässige Richtung in x^* und $\nabla a_i(x^*) d = 0 \quad \forall i \in \mathcal{U}$ mit $\lambda_i^* >$

$\Longleftrightarrow$ d zulässige Richtung für das Minimierungsproblem, in dem alle Ungleichheits-NB in (P) mit nichtverschwindenden Lagrange-Multiplikatoren durch Gleichheits-NB ersetzt werden

<u>Satz 4:</u> (<u>notwendige</u> Bedingungen)

Sei x^* lokales Minimum von (P), und gelte Regularitäts-VS von Sa

$\Longrightarrow \quad \exists \lambda^* \in \mathbb{R}^m$, sodaß (x^*, λ^*) die Kuhn-Tucker-Bed. erfüllen. (S

und es gilt

$$\underline{d^T M(x^*, \lambda^*) d \geq 0} \quad \forall \text{ strikt zulässigen } d \in \mathbb{R}^n$$

(Hierbei ist wieder $M = \frac{\partial^2 L}{\partial x^2}$, mit $L(x,\lambda) = f(x) - \sum_{i=1}^{m} \lambda_i a_i(x)$ Lagrang

<u>Beweis:</u> Sei d strikt zulässige Richtung: $\nabla a_i(x^*) d = 0 \quad \forall i \in \mathcal{U}$ mit λ_i^*

— " — $\quad \forall i \in \mathcal{G}$

nach Regularitäts-VS sind diese $\nabla a_i(x^*)$ linear unabhängig, und $a_i(x^*$ (Komplement

nach HS, §1: $\forall \varepsilon > 0$ gen. klein $\exists \Delta x = \varepsilon d + O(\varepsilon^2)$

sodaß für $x = x^* + \Delta x$ gilt: $a_i(x) = 0 \quad \forall i \in \mathcal{U}$ mit $\lambda_i^* >$

$\forall i \in \mathcal{G}$

$\Longrightarrow \quad \lambda_i^* a_i(x) = 0 \quad \forall i = 1, \ldots, m$

damit

$$f(x) = f(x) - \sum_{i=1}^{m} \lambda_i^* a_i(x) = L(x, \lambda^*)$$

weiter wie in S.1, §1 (Taylorentwicklung von $L(x,\lambda^*)$ um x^*) $\Rightarrow$ Beh.

Satz 5: (*hinreichende* Bedingungen)

(x^*, λ^*) erfülle Kuhn-Tucker-Bedingungen, und es gelte

$$\underline{d^T M(x^*, \lambda^*) d > 0 \quad \forall \text{ strikt zulässigen Richtungen } d}$$

$\Rightarrow$ x^* ist striktes lokales Minimum von (P).

Beweis: indirekt: falls x^* *kein* striktes lokales Minimum von (P

so existiert Folge zulässiger Punkte $\underset{x^*}{\overset{}{x_k}} \to x^*$ mit ($x_k \neq x^*$)

$$f(x_k) \leq f(x^*)$$

schreibe $x_k = x^* + \varepsilon_k d_k$ mit $\|d_k\| = 1$, $\varepsilon_k \to 0$

beschränkte Folge $\{d_k\}$ hat Häufungspunkt:

$\exists$ konvergente Teilfolge $\{d_{k_\ell}\}$ von $\{d_k\}$

$$\lim_{\ell \to \infty} d_{k_\ell} =: d \qquad \text{oEdA}: \quad d_k \to d \ (\neq 0)$$

zeigen: d ist strikt zulässige Richtung

x_k zulässig $\Rightarrow 0 \leq a_i(x_k) = a_i(x^*) + \varepsilon_k \nabla a_i(x^*) d_k + O(\varepsilon_k^2)$, i

falls i aktive NB: $a_i(x^*) = 0$ dividiere durch ε_k, lasse $\varepsilon_k \to$

$$\left.\begin{array}{lll} \Rightarrow & \nabla a_i(x^*)\, d \geq 0 & \forall i \in \mathcal{A}(x^*) \\ \text{ebenso} & \nabla a_i(x^*)\, d = 0 & \forall i \in \mathcal{G} \end{array}\right\} d \text{ zulässige R}$$

$$\left.\begin{array}{l} f(x_k) \leq f(x^*) \\ \parallel \\ f(x^*) + \varepsilon_k \nabla f(x^*) d_k + O(\varepsilon_k^2) \end{array}\right\} \nabla f(x^*) d \leq 0$$

$$0 \geq \nabla f(x^*) d \underset{\substack{\uparrow \\ \text{Kuhn-Tucker}}}{=} \sum_{i=1}^{m} \underbrace{\lambda_i^* \nabla a_i(x^*) d}_{\geq 0 \ \forall i}$$

$$\Rightarrow \lambda_i^* \nabla a_i(x^*) d = 0 \quad \forall i = 1, \dots, m$$

$$\Rightarrow \nabla a_i(x^*) d = 0 \quad \forall i \in \mathcal{U} \text{ mit } \lambda_i^* > 0$$

somit: d strikt zulässige Richtung

x_k zulässig $\Rightarrow$

$$\begin{array}{l} f(x_k) \geq f(x_k) - \sum_{i\in\mathcal{U}} \underset{\overset{\vee\!\mid}{0}}{\lambda_i^*}\, \underset{\overset{\vee\!\mid}{0}}{a_i(x_k)} - \sum_{i\in\mathcal{G}} \lambda_i^* \underset{\overset{\shortparallel}{0}}{a_i(x_k)} = \\ \quad \wedge\!\mid \\ f(x^*) \end{array}$$

$$= L(x_k, \lambda^*) = \underbrace{L(x^*, \lambda^*)}_{f(x^*)} + \cancelto{0}{\frac{\partial L}{\partial x}(x^*, \lambda^*)}\, \varepsilon_k d_k +$$

$$+ \frac{1}{2} \varepsilon_k^2\, d_k^T M(x^*, \lambda^*) d_k + O(\varepsilon_k^3)$$

$$\Rightarrow \quad \cancel{\varepsilon_k}\; d_k^T M(x^*, \lambda^*) d_k + O(\varepsilon_k) \leq 0$$

$$\underset{k\to\infty}{\Rightarrow} \quad d^T M(x^*, \lambda^*) d \leq 0$$ im Wspr. zur VS $\square$

§6 Sequentielle quadratische Optimierung

(P) minimiere $f(x)$ unter NB $\quad a_i(x) = 0 \quad i \in \mathcal{G}$

$\qquad\qquad\qquad\qquad\qquad\qquad\quad a_i(x) \geq 0 \quad i \in \mathcal{U}$

Erinnerung: §2 (Lagrange-Newton bei Gleichheits-NB)

Iteration $\quad x_{k+1} = x_k + \Delta x_k$

bestimme Δx_k als Lösung des quadratischen Minimierungsproblems

minimiere $\frac{1}{2} \Delta x_k^T M_k \Delta x_k + g_k \quad$ ($M_k = M(x_k, \lambda_k)$, $g_k = \nabla f(x_k)^T$)

unter linearen NB

$$\nabla a_i(x_k) \cdot \Delta x_k + a_i(x_k) = 0 \qquad i \in \mathcal{G}$$

(wieder Modifi… Quasi-Newt… gedämpftes …

jetzt: zusätzlich Ungleichheits-NB

$$\nabla a_i(x_k) \cdot \Delta x_k + a_i(x_k) \geq 0 \qquad i \in \mathcal{U}$$

Wie löst man solche quadratische Minimierungsprobleme mit Ungleichheits-NB

("Active Set Method")

§7 Aktive Nebenbedingungs-Verfahren zur quadratischen Optimierung

(ähnlich Simplex-Verfahren der linearen Optimierung)

$$(Q) \quad \text{minimiere} \quad \frac{1}{2} x^T M x + c^T x \qquad (x \in \mathbb{R}^n, M \in \mathbb{R}^{n \times n}, c \in \mathbb{R}$$

$$\text{unter NB:} \quad a_i^T x \geq b_i \quad i \in \mathcal{U}$$

$$a_i^T x = b_i \quad i \in \mathcal{G}$$

$(a_i \in \mathbb{R}^n, b_i \in \mathbb{R})$

$\mathcal{U} \cup \mathcal{G} = \{1, \dots, m\}$

nehmen an: $d^T M d > 0 \quad \forall d$ mit $a_i^T d = 0$ für $i \in \mathcal{G}$

(nach S. 5, §5: Falls es x^* gibt, das die Kuhn-Tucker-Bed. erfüllt,
so ist es lokales Minimum
hier tatsächlich: globales Minimum)

Sei x_0 zulässiger Punkt (evtl. erhalten mit Phase I des Simplex-Verfahr

aktive NB in x_0: $\mathcal{A} = \{ i \in \mathcal{U} \mid a_i^T x_0 = b_i \}$

ersetze Ungleichung in (Q) durch Gleichheits-NB für $i \in \mathcal{A}$
und suche Lösung des quadratischen Minimierungsproblems
mit Gleichheits-NB:

$$\begin{cases} \frac{1}{2} x^T M x + c^T x = \min! \\ a_i^T x = b_i \quad \text{für } i \in \mathcal{A} \cup \mathcal{G} \end{cases}$$

bzw. äquivalent für $d = x - x_0$ und $g = Mx + c$

$$\begin{cases} \frac{1}{2} d^T M d + g^T d = \min! \\ a_i^T d = 0 \quad \text{für } i \in \mathcal{A} \cup \mathcal{G} \end{cases}$$

(erhalte d als Lösung eines linearen Gleichungssystems)

Falls $a_i^T(x_0+d) \geq b_i \quad \forall i \in \mathcal{U}\setminus\mathcal{A}$

so ist x_0+d zulässig, setze

$x_1 := x_0 + d$

~~beginne von vorn~~

sonst: setze $x_1 := x_0 + \rho d$

wobei $0 \leq \rho < 1$ maximal so, daß $x_0+\rho d$ zulässig ist

$$a_i^T(x_0+\rho d) \geq b_i \quad \forall i \in \mathcal{U}\setminus\mathcal{A}$$

erhalte $\rho = \min\limits_{\substack{i\in\mathcal{U}\setminus\mathcal{A} \\ \text{mit } a_i^T d < 0}} \dfrac{b_i - a_i^T x_0}{a_i^T d}$

Für jenen Index j, bei dem das Minimum angenommen wird

$$a_j^T x_1 = b_j$$

erweitere aktive NB $\mathcal{A}$ um j: $\quad \mathcal{A} := \mathcal{A}\cup\{j\}$

2 Möglichkeiten:

a) x_1 ist Lösung von $\begin{cases} \frac{1}{2}x^T Mx + c^T x = \min! \\ a_i^T x = b_i \qquad i\in\mathcal{A}\cup\mathcal{G} \end{cases}$

betrachte zugehörige Lagrange-Mult. λ_i, $i\in\mathcal{A}$

falls $\lambda_i \geq 0 \ \forall i\in\mathcal{A}$: x_1 erfüllt Kuhn-Tucker-Bed

$\underset{\text{VS, S. 5, §5}}{\Longrightarrow}$ x_1 Lösung von (Q) fertig

sonst: sei k mit $\lambda_k = \min\limits_i \lambda_i < 0$

entferne k aus $\mathcal{A}$, beginne von vorne

(8) x_1 keine Lösung mit NB $i \in \mathcal{A} \cup \mathcal{G}$

Beginne von vorne

erhalte zusammenfassend : <u>Aktive - NB - Verfahren</u>

1) finde zulässigen Punkt x (z.B. mit Simplex Phase I)

zugehörige aktive Indexmenge : $\mathcal{A} = \{ i \in \mathcal{U} \mid a_i^T x = b_i \}$

2) löse
$$\begin{pmatrix} M & -A(\mathcal{A})^T \\ -A(\mathcal{A}) & 0 \end{pmatrix} \begin{pmatrix} d \\ \lambda \end{pmatrix} = \begin{pmatrix} -g \\ 0 \end{pmatrix} \qquad g = Mx + c$$

mit $A(\mathcal{A}) = (a_i^T)_{i \in \mathcal{A} \cup \mathcal{G}}$

3) falls $d = 0$:

falls $\lambda_i \geq 0 \quad \forall i \in \mathcal{A}$: x ist Lösung, fertig

falls $\lambda_k = \min\limits_{i \in \mathcal{A}} \lambda_i < 0$: $\mathcal{A} := \mathcal{A} \setminus \{k\}$

gehe nach 2)

4) setze $x := x + \rho d$

mit $$\rho = \min \left(1, \min_{\substack{i \in \mathcal{U} \setminus \mathcal{A} \\ a_i^T d < 0}} \frac{b_i - a_i^T x}{a_i^T d} \right)$$

Falls $\rho < 1$: $\mathcal{A} := \mathcal{A} \cup \{j\}$ wobei j jener Index, bei dem das innere Minimum angenommen wird

gehe nach 2)

Bem: Gls. in 2) unter gegebener VS an M immer eindeutig nach d auflösb[ar]

λ eindeutig, falls $A(\mathcal{A})$ maximalen Rang hat liefert i.a. nach endlich vielen Schritten die Lösung von (Q)

denn: in jedem Schritt minimiert x $\begin{cases} \frac{1}{2}x^T M x + c^T x = \min! \\ A(\mathcal{A})x = b(\mathcal{A}) \end{cases}$

Zielfunktion wird in jedem Schritt verringert (außer falls $\rho = 0$)

$\exists$ nur endlich viele Wahlmöglichkeiten für $\mathcal{A}$ und damit für (*

daher: falls Algorithmus nicht bei einem nichtoptimalen x wegen $\rho = 0$ in jedem Schritt steckenbleibt ("Cycling", vgl. Sim
tritt praktisch nicht au
so bricht er nach endlich vielen Schritten ab.