5. Optimierungsprobleme

- Betrachten hier nur <u>endlichdimensionale Optimierungsprobleme</u> (OP) der Form:

Finde $\underline{x}^* \in \mathbb{R}^n$:

(1) $\boxed{f(\underline{x}^*) = \min_{\underline{x} \in \Omega_{ad}} f(\underline{x})}$ bzw. $\boxed{\max_{\underline{x} \in \Omega_{ad}} (-f(\underline{x}))}$

wobei

$\Omega = \Omega_{ad} \subset \mathbb{R}^n$ – Zulässigkeitsmenge / ~bereich,

$\underline{x} = (x_1, x_2, \dots, x_n)^T \in \Omega_{ad}$ – zulässiger Punkt,

$f: \mathbb{R}^n \to \mathbb{R}$ – Zielfunktion = Kostenfunktional.

- Großdimensionale Optimierungsprobleme der Form (1) entstehen z.B. bei der FEM-Diskretisierung
  - des Optimalsteuerproblems (optimal control) bzw.
  - des Formoptimierungsproblems (optimal design)

(= unendlichdimensionale Optimierungsprobleme) aus Kapitel 1, z.B. Optimalsteuerproblem aus Kapitel 1 nach FEM-Diskretisierung (→ 39 b)

$$\min_{\underline{x} = (\underline{y}, \underline{u}) \in \mathbb{R}^{n = n_s + n_c}} \overbrace{\frac{1}{2}(M_s \underline{y}, \underline{y}) - (\underline{y}_d, \underline{y}) + \underbrace{\cancel{\frac{1}{2}\int_\Omega y_d^2(x)\,dx}}_{= \text{const}} + \frac{\alpha}{2}(M_c \underline{u}, \underline{u})}^{f(\underline{x})}$$

s.t. $K\underline{y} = M_{sc}\underline{u}$ $\qquad c_i(x) = (K\underline{y} - M_{sc}\underline{u})_i = 0, \; i = 1, \dots, n_s = m_1$

$$\left.\begin{array}{ll} \underline{y}_i \le y_i \le \overline{y}_i & \underline{y}_i - y_i \le 0 \\ & y_i - \overline{y}_i \le 0 \\ \underline{u}_i \le u_i \le \overline{u}_i & \underline{u}_i - u_i \le 0 \\ & u_i - \overline{u}_i \le 0 \end{array}\right\} \; c_i(\underline{x}) \le 0, \; i = \overline{m_1 + 1, m_2 + m_1}, \quad m_2 = 2n_s + 2n_c$$

wobei $K = \left[\int_\Omega A(x) \nabla\varphi_j \cdot \nabla\varphi_i \,dx\right]_{i,j = \overline{1, n_s}}$ – Steifigkeitsmatrix,

$M_{sc} = \left[\int_\Omega \varphi_j \cdot \varphi_i \,dx\right]_{i = \overline{1, n_s},\, j = \overline{1, n_c}}$ – Massenmatrix, $\underline{y}_d = \dots$

Beispiel aus Kapitel 1: Optimalsteuerproblem

$$\min_{y\in Y,\, u\in U} J(y,u) := \frac{1}{2}\int_\Omega (y(x)-y_d(x))^2\,dx + \frac{\alpha}{2}\int_\Omega (u(x))^2\,dx$$

$$\text{s.t.} \left.\begin{bmatrix} -\operatorname{div}(A(x)\nabla y(x)) = u(x), & x\in\Omega \\ +\ BC:\ y(x)=g(x):=0 & , x\in\Gamma \end{bmatrix}\right\}(\text{PDE constr.})$$

State constraints: $\underline{y}(x) \le y(x) \le \bar{y}(x),\ x\in\Omega$

Control constr. : $\underline{u}(x) \le u(x) \le \bar{u}(x),\ x\in\Omega$

Die Variationsformulierung der PDE ergibt schließlich das folgende <u>unendlichdim. restr. OP</u>:

$$\min_{y\in \mathring{H}^1(\Omega),\, u\in L_2(\Omega)} \frac{1}{2}\int_\Omega (y(x)-y_d(x))^2\,dx + \frac{\alpha}{2}\int_\Omega (u(x))^2\,dx$$

$$\text{s.t.} \int_\Omega A(x)\nabla y(x)\cdot\nabla v(x)\,dx = \int_\Omega u(x)v(x)\,dx \quad \forall v\in \mathring{H}^1(\Omega)$$

$$\underline{y}(x) \le y(x) \le \bar{y}(x),\ x\in\Omega$$

$$\underline{u}(x) \le u(x) \le \bar{u}(x),\ x\in\Omega$$

Die FEM-Diskretisierung

$$y\in \mathring{H}^1(\Omega) \approx y_h(x) = \sum_{i=1}^{n_s} y_i\varphi_i(x) \in Y_h,\ \dim Y_h = n_s$$

$$u\in L_2(\Omega) \approx u_h(x) = \sum_{i=1}^{n_c} u_i\varphi_i(x) \in U_h,\ \dim U_h = n_c$$

ergibt das <u>endlichdimensionale restringierte OP</u>:

$$\min_{x=(\underline{y}_h,\underline{u}_h)\in\mathbb{R}^n} \frac{1}{2}(M_s\underline{y},\underline{y}) - (\underline{y}_d,\underline{y}) + \underbrace{\cancel{\frac{1}{2}\int_\Omega (y_d(x))^2\,dx}}_{\text{const}} + \frac{\alpha}{2}(M_c\underline{u},\underline{u})$$

$$\text{s.t.}\ K_h\underline{y}_h = M_{sc,h}\underline{u}_h$$

$$\underline{y}_i \le y_i \le \bar{y}_i,\ i=\overline{1,n_s}$$

$$\underline{u}_j \le u_j \le \bar{u}_j,\ j=\overline{1,n_c}$$

K – Steifigkeitsmatrix
M_s, M_c, M_{sc} – Massenmatrizen

Zwei große Klassen von Optimierungsproblemen:

1. Frei (unrestringierte) Optimierung: $\Omega_{ad} = \mathbb{R}^n$

2. Optimierung mit Nebenbedingungen (restringierte OP):

Wir beschränken uns auf den Fall, daß sich Ω_{ad} durch m_1 Gleichungsrestriktionen und m_2 Ungleichungsrestriktionen beschreiben lässt, d.h.

$\underline{x} \mapsto x$

$$\Omega_{ad} = \{ x \in \mathbb{R}^n : \ c_i(x) = 0, \ i = \overline{1, m_1} \text{ (Gleichungsrestr.)}, \ c_i(x) \le 0, \ i = \overline{m_1+1, m_1+m_2} \text{ (Ungleichungsrestr.)} \}$$

mit gegebenen Funktionen $c_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1, 2, \ldots, m_1+m_2$:

$$(1)_{rOP} \quad \boxed{\begin{aligned} &\min_{x \in \mathbb{R}^n} f(x) \\ &\text{s.t. } c_i(x) = 0, \ i = 1, 2, \ldots, m_1 \quad \text{(equitity const.)} \\ &\phantom{\text{s.t. }} c_i(x) \le 0, \ i = m_1+1, \ldots, m_1+m_2 \quad \text{(ineq. constraints)} \end{aligned}}$$

Je nach Komplexität der beteiligten FKten $f(\cdot)$ und $c_i(\cdot)$ unterscheidet man genauer folgende OP nach aufsteigender Schwierigkeit:

Simplex Alg.

1) Lineare Optimierung: $f(\cdot)$ und $c_i(\cdot)$ affine linear.
2) Quadratische Opt.: f quadratisch, c_i linear.
3) Nichtlineare Optimierung mit linearen NB: f – nichtlinear, c_i – linear.
4) Nichtlineare Optimierung mit nichtlinearen NB: f – nichtlinear, c_i – nichtlinear,

wobei $g : \mathbb{R}^n \to \mathbb{R}$ heißt

affine

- linear, falls $g(x) = g_0 + c^T x$ mit geg. $g_0 \in \mathbb{R}$, $c \in \mathbb{R}^n$,
- quadratisch, falls $g(x) = g_0 + c^T x + \frac{1}{2} x^T H x$, $H \in \mathbb{R}^{n \times n}$.

- **Bem. 5.1:** $\Omega_{ad} = \{z_1, z_2, \ldots, z_N\}$
  1. Ω_{ad} - endlich: Kombinatorische Optimierung
  2. $\Omega_{ad} \subset \mathbb{Z}^N$: Ganzzahlige Optimierung

- **Def. 5.2:** globales und lokales Minimum (Maximum)
  1. Ein Pkt. $x^* \in \Omega_{ad}$ heißt (striktes) globales Min., falls
  $$f(x^*) \le f(x) \quad (f(x^*) < f(x)) \quad \forall x \in \Omega_{ad} : x \neq x^*.$$
  2. Ein Pkt. $x^* \in \Omega$ heißt (striktes) lokales Minimum, falls $\exists\, \varepsilon > 0$:
  $$f(x^*) \le f(x) \quad (f(x^*) < f(x)) \quad \forall x \in U_\varepsilon(x^*) : x \neq x^*,$$
  wobei $U_\varepsilon(x^*) := \{x \in \Omega_{ad} : \|x - x^*\| < \varepsilon\}$.
  ε-Umgebung von x^*

Bem.:
- $\max f(x) \iff \min(-f(x))$
- In den meisten Fällen ist man an einem globalen Min. interessiert.
- Viele Alg. garantieren aber nur die Konstr. von lok. Min. !!!

- **Bem. 5.3:** Konvexe Optimierungsprobleme
  Für OP mit konvexer Zielfunktion f und konvexem Zulässigkeitsbereich Ω_{ad} ist jedes lokale Minimum auch ein globales Minimum.
  Wh: Konvexität für f und Ω_{ad}:
  $\Omega \subset \mathbb{R}^n$ = konvex gdw. $\forall x, y \in \Omega_{ad} \wedge \forall \alpha \in [0,1]$:
  $$\alpha x + (1-\alpha) y \in \Omega_{ad}.$$
  $f : \mathbb{R}^n \to \mathbb{R}$ = konvex gdw. $\forall x, y \in \mathbb{R}^n \wedge \forall \alpha \in [0,1]$:
  $$f(\alpha x + (1-\alpha) y) \le \alpha f(x) + (1-\alpha) f(y)$$

5.1. Freie Optimierungsprobleme

- In diesem Abs. btr. wir das frei (unrestringierte) OP

$$(1)_{fOP} \quad \boxed{\text{Finde } x^* \in \Omega_{ad} = \mathbb{R}^n : f(x^*) = \min_{x \in \mathbb{R}^n} f(x)}$$

- Im Kapitel 3 haben wir schon das <u>quadratische freie OP</u>

$$(3.13) \quad f(x) := E(x) := \frac{1}{2}(\overset{\text{SPD}}{\overset{\downarrow}{A}} x, x) - (b, x) \longrightarrow \min_{x \in \mathbb{R}^n} !$$

Kennengelernt und gezeigt, daß

$$f(x^*) = \min_{x \in \mathbb{R}^n} f(x) \iff A x^* = b$$

5.1.1. Theoretische Grundlagen

- Sei $f : \mathbb{R}^n \to \mathbb{R}$ eine zumindest zweimal stetig partiell differentierbare Fkt, d.h. $f \in C^2$. Dann ist $g = \operatorname{grad} f = \nabla f : \mathbb{R}^n \to \mathbb{R}^n$ durch

$$g(x) = \nabla f(x) = f'(x)^T = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix} = \left[\frac{\partial f}{\partial x_i}(x) \right]_{i=\overline{1,n}}$$

gegeben. $g(x)$ heißt <u>Gradient</u> der Funktion f an der Stelle x.

- Der Hessian $H: \mathbb{R}^n \to \mathbb{R}^{n\times n}$ ist durch

$$H(x) = \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix} = H^T(x)$$

Satz von Schwarz

gegeben. $H(x)$ nennt man die Hesse-Matrix von f an der Stelle x. Die Hesse-Matrix ist symmetrisch

- Aus der Taylor-Darstellung

f-quadr. $= 0$

$n=1$: $f(x) = f(x^*) + f'(x^*)(x-x^*) + \frac{1}{2}(x-x^*)f''(x^*)(x-x^*) + o(\|x-x^*\|^2)$

(2) $n \geq 1$: $f(x) = f(x^*) + \underbrace{\nabla^T f(x^*)}_{\substack{= \\ 0}}(x-x^*) + \underbrace{\frac{1}{2}(x-x^*)^T H(x^*)(x-x^*)}_{\substack{> 0 \\ \geq}} + o(\|x-x^*\|^2)$

$f(x) \underset{\geq}{>} f(x^*)$

notwendig hinreichend

folgen sofort notwendige und hinreichende Bedingungen für ein lokales Minimum von f:

- Satz 5.4: (notwendige Bedingung)
  Sei $x^* \in \mathbb{R}^n$ ein lokales Minimum (Maximum) von f. Dann gilt:

  (3) $\nabla f(x^*) = 0$ (= nichtl. GS $\boxed{F(x) = 0}$ in $\mathbb{R}^n$), wobei $F(x) = \nabla f(x)$

- Satz 5.5: (hinreichend Bedingung)
  Falls $x^* \in \mathbb{R}^n$ ein kritischer Punkt von f ist, d.h. eine Lösung des nichtlinearen GS (3) ist, und $H(x^*)$ positiv (definit) ist, dann ist $x^* \in \mathbb{R}^n$ ein (striktes) lokales Minimum.

- **Bem. 5.6:**

1. Ein Pkt. $x^* \in \mathbb{R}^n$: $\nabla f(x^*) = 0$ wird stationärer oder kritischer Pkt. genannt.
2. Nach Satz 5.5 ist ein kritischer Pkt. $x^* \in \mathbb{R}^n$ mit p.d. Hesse-Matrix $H(x^*)$ ein striktes lokales Minimum.
3. Offenbar ist dann ein kritischer Pkt. $x^* \in \mathbb{R}^n$ mit n.d. Hesse-Matrix (d.h. $-H(x^*)$ ist p.d.) ein striktes lokales Maximum.
4. Ein kritischer Pkt. $x^* \in \mathbb{R}^n$ mit indefiniter Hesse-Matrix $H(x^*)$, d.h. $H(x^*)$ ist weder positiv noch negativ definit, heißt Sattelpkt.
5. Idee für numerische Lösungsverfahren:
   1) Löse nichtlineares GS

$$F(x) := \nabla f(x) = 0 \quad \text{in } \mathbb{R}^n; \tag{4}$$

   z.B. mit Newton-Verfahren (↓), um kritische Punkte x^* zu finden!

   2) Check Definitheit der Hesse-Matrix:

$$\text{①} \quad H(x^*) \begin{cases} > 0 \text{ (SPD)} & \Rightarrow \text{striktes lok. Min.} \\ \geq 0 & \Rightarrow \text{(Lokales Minimum)} \\ < 0 & \Rightarrow \text{striktes lok. Max} \\ \leq 0 & \Rightarrow \text{(Lokales Maximum)} \\ \text{indefinit} & \Rightarrow \text{Sattelpunkt} \\ = 0 & \Rightarrow \text{Check höhere Terme in der Taylorentw. (2)} \end{cases}$$

   im Punkt x^*!

6. Umgekehrt kann man Lösungen des nichtlin. GS (4) finden, indem man das OP (1) f_{OP} $\min f(x)$ bzw. $\max f(x)$ löst.

5.1.2. Abstiegsverfahren

- Ein typischer Iterationsschritt eines Abstiegsverfahren zur Lösung freier OP

$$(1)_{fOP} \quad \text{Finde } x^* \in \mathbb{R}^n : f(x^*) = \min_{x \in \mathbb{R}^n} f(x)$$

hat im Allgemeinen die Form:

(5)
1. Bestimme, ausgehend von einer Näherung $x^{(k)} \in \mathbb{R}^n$, eine Suchrichtung $s^{(k)} \in \mathbb{R}^n$ ($\rightarrow$ Abstiegsrichtung).
2. Bestimme eine Schrittweite $\alpha^{(k+1)} > 0$ mit $f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) < f(x^{(k)})$ ($\rightarrow$ Liniensuche)
3. Setze $x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)}$.

- <u>Bemerkung:</u> Das Gradienten-Verfahren und das CG-Verfahren sowie ihre präkonditionierten Versionen aus Abs. 3.3.2 sind Verfahren der Form (5).

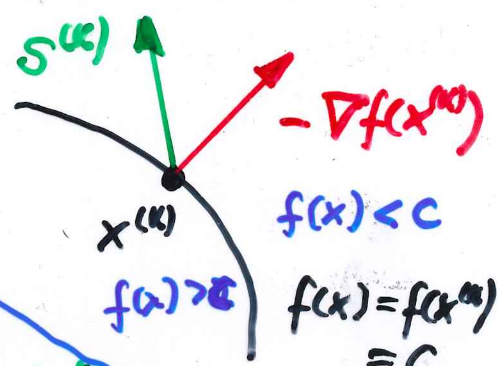
- <u>Anforderungen an die Suchrichtung ($\rightarrow$ Abstiegsricht.):</u>

Aus der Taylor-Entwicklung (vgl. auch (2))

$$f(x^{(k)} + \alpha s^{(k)}) = f(x^{(k)}) + \alpha \nabla^T f(x^{(k)}) s^{(k)} + o(\alpha)$$
$$\approx f(x^{(k)}) + \alpha \nabla^T f(x^{(k)}) s^{(k)}$$

folgt für hinreichend kleine Werte von $\alpha > 0$ sofort

$$f(x^{(k)} + \alpha s^{(k)}) < f(x^{(k)})$$



falls

$$(6) \qquad \boxed{\nabla^T f(x^{(k)}) \, s^{(k)} < 0}$$

$$-\nabla^T f(x) \cdot s^{(k)} > 0$$

Eine Suchrichtung $s^{(k)} \in \mathbb{R}^n$, die die Bd. (6) erfüllt, heißt Abstiegsricht.!

Das entsprechende Verfahren heißt <u>Abstiegsverfahren</u>!

- **Anforderungen an die Schrittweite:**
  - **Exakte Liniensuche:**

$$\alpha^{(k+1)} > 0 : f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) = \min_{\alpha > 0} f(x^{(k)} + \alpha s^{(k)})$$

Allerdings ist diese Schrittweitenwahl außer in einigen Spezialfällen (z.B. bei einer quadratischen Zielfkt. $f(x)$: siehe Abs. 3.2.2) zu aufwendig!

  - **Praktikable Liniensuchstrategie** ($\rightarrow$ Bisektion): z.B. $\alpha^{(k+1)} = 1, 2^{-1}, 2^{-2}, \ldots$

$$(7)\quad \alpha^{(k+1)} > 0 : f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) \le f(x^{(k)}) + \mu \alpha^{(k+1)} \nabla^T f(x^{(k)}) s^{(k)}$$

für ein fix vorgegebenes $\mu \in (0,1)$ ($\rightsquigarrow$ Taylor) $\wedge$

$$(8)\quad \nabla^T f(x^{(k)} + \alpha^{(k+1)} s^{(k)}) s^{(k)} \ge \sigma \nabla^T f(x^{(k)}) s^{(k)}$$

für ein fix vorgegebenes $\sigma \in (\mu, 1)$.
$\rightsquigarrow$ Keine zu kleine Schrittweite!

- **Satz 5.7:** Falls
  1. die Näherungen $x^{(k)}$ eine beschränkte Folge bilden,
  2. die Schrittweiten $\alpha^{(k)}$ die Bd. (7) und (8) erfüllen, und
  3. $\exists \varepsilon > 0 : \nabla^T f(x^{(k)}) s^{(k)} \le -\varepsilon \|\nabla f(x^{(k)})\|_2 \|s^{(k)}\|_2 < 0$,

  dann gilt: $\lim_{k \to \infty} \nabla f(x^{(k)}) = 0$.

- **Bem. 5.8:**
  1. $s^{(k)} = -\nabla f(x^{(k)})$ $\Downarrow$ GV = Verf. d. steilsten Abst.
     Vor. 3 von Satz 5.7 ist mit $\varepsilon = 1$ erfüllt!
  2. GV mit Liniensuche (7)-(8) konvergiert auf der Basis von Satz 5.7 unter schwachen Vor. global!
  3. Allerdings konvergiert das GV nur sehr langsam!

5.1.3. Das Newton-Verfahren

- **Idee:** Sei $x^{(k)} \in \mathbb{R}^n$ eine Näherung an ein lok. Min. v. f.

Aus der Taylor-Entwicklung (vgl. auch (2))

$$(9) \quad f(x) = \underbrace{f(x^{(k)}) + \nabla^T f(x^{(k)})(x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T H(x^{(k)})(x - x^{(k)})}_{=: q_k(x)} + \cancel{HOT}$$

$$HOT = o(\|x - x^{(k)}\|^2)$$

folgt, daß sich $f(x)$ in einer Umgebung $U_\varepsilon(x^{(k)})$ von $x^{(k)}$ durch die quadratische Fkt.

$$(10) \quad q_k(x) := f(x^{(k)}) + \nabla^T f(x^{(k)})(x - x^{(k)}) + \frac{1}{2}(x - x^{(k)})^T H(x^{(k)})(x - x^{(k)}) \approx f(x)$$

gut approximieren läßt!

Die nächste Iteration $x^{(k+1)}$ des Newton-Verfahrens in der Optimierung wird als Minimum der quadratischen Funktion $q_k(x)$ definiert:

$$q_k(x^{(k+1)}) = \min_{x \in \mathbb{R}^n} q_k(x). \qquad \text{SQP}$$

Unter der Annahme, daß (vgl. Satz 5.5)

$\nabla^2 q_k(x^{(k)}) = H(x^{(k)})$ <u>positiv definit</u> ist,

ist $x^{(k+1)}$ nach Satz 5.4 durch die Bedingung

$$\nabla q_k(x^{(k+1)}) = \nabla f(x^{(k)}) + H(x^{(k)})(x^{(k+1)} - x^{(k)}) = 0$$

eindeutig definiert, d.h.

$$(11) \quad x^{(k+1)} = x^{(k)} - H(x^{(k)})^{-1} \nabla f(x^{(k)})$$
$$= x^{(k)} + \alpha^{(k+1)} s^{(k)}$$

mit der Schrittweite $\alpha^{(k+1)} = 1$ und der Suchrichtung

$$(12) \quad s^{(k)} = -H(x^{(k)})^{-1} \nabla f(x^{(k)}) \quad \text{(Newton-Richtung)}$$

NEWTON-Verfahren zur Lösung von

freie OP	nichtlineare GS
$\boxed{\min_{x\in\mathbb{R}^n} f(x)}$ $\Longrightarrow$	$\nabla f(x) = 0$ in $\mathbb{R}^n$
	$\boxed{F(x) = 0 \text{ in } \mathbb{R}^n}$
$\approx$ in	$U(x^{(k)})$ $\approx$
SQP $\quad \min_{x\in\mathbb{R}^n} q_k(x) \quad x^{(k)}$	$F(x^{(k)}) + \nabla F(x^{(k)})(x - x^{(k)}) = 0$
$(11)_{\nabla f}$ $\quad x^{(k+1)} = x^{(k)} - H_f(x^{(k)})^{-1} \nabla f(x^{(k)})$	$x^{(k+1)} = x^{(k)} - J_F(x^{(k)})^{-1} F(x^{(k)})$ $\quad (11)_F$
$\uparrow$ Hesse-Matrix von f	$\uparrow$ Jacobi-Matrix von F

$$J_F(x) = \left[\frac{\partial F_i}{\partial x_j}(x)\right]_{i,j=\overline{1,n}} \overset{\text{falls } F=\nabla f}{=} \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(x)\right]_{i,j=\overline{1,n}} = H_f(x)$$

Satz 5.9: $(11)_{\nabla f}$

"bzw. 2x + H Lip"

Vor.: Sei $f: \mathbb{R}^n \to \mathbb{R}$ "dreimal" stetig partiell diffbar, $x^* \in \mathbb{R}^n$ sei ein lokales Minimum von f und $H(x^*)$ sei SPD.

Bh.: Dann ist das NEWTON-Verfahren für hinreichend gute Startwerte $x^{(0)}$, d.h. $\|x^* - x^{(0)}\|$ hinreichend klein, wohldefiniert und es konvergiert q-quadratisch gegen x^*, d.h. $\exists\, c = \text{const} > 0$:

$$(13) \quad \|x^* - x^{(k+1)}\| \le c\, \|x^* - x^{(k)}\|^2.$$

Beweis: siehe Beweis von Satz 5.10 mit $F = \nabla f$. ∎

(↓)

- **Satz 5.10:** $(11)_F \rightarrow F(x) = 0$ in $\mathbb{R}^n$

Vor.: Sei $F: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ einmal stetig partiell diff'bar

$x^* \in \mathbb{R}^n$ sei Nullstelle von F, d.h. $F(x^*) = 0$,

$\nabla F(x^*) =: J(x^*)$ sei regulär und $\beta = \| J(x^*)^{-1} \|$

Bh.: Dann ist das NEWTON-Verfahren $(11)_F$ für hinreichend gute Startwerte wohldefiniert und es konvergiert q-superlinear gegen x^*, d.h.

$$\lim_{k\to\infty} \underbrace{\frac{\| x^* - x^{(k+1)} \|}{\| x^* - x^{(k)} \|}}_{= q_k} = \lim_{k\to\infty} q_k = 0$$

$$\| x^* - x^{(k+1)} \| \le \underbrace{q_k}_{\to 0} \| x^* - x^{(k)} \|$$

Falls ∇F in x^* Lipschitz-stetig ist, d.h. $\exists L = \text{const} > 0$

$$\| \nabla F(x) - \nabla F(x^*) \| \le L \, \| x - x^* \| \quad \forall x \in U(x^*)$$

dann konvergiert das Newton-Verfahren q-quadratisch, d.h. es gilt (13).

Beweis: $x = x^{(k)}$, $F' = \nabla F = J$ - Jacobi-Matrix

$$\| x^{(k+1)} - x^* \| = \| x - F'(x)^{-1} F(x) - x^* \|$$

$$= \| F'(x)^{-1} [\underbrace{F(x^*)}_{=0} - F(x) - F'(x)(x - x^*)] \|$$

$$\le \| F'(x)^{-1} \| \; \underbrace{\| F(x^*) - F(x) - F'(x)(x - x^*) \|}_{o(\| x - x^* \|)} \quad \text{1x stetig diff'bar}$$

ZWS: $\exists \xi \in (0,1)$

$$= \| F'(x)^{-1} \| \; \underbrace{\| [F'(x + \xi(x^* - x)) - F'(x^*) + F'(x^*) - F'(x)] (x - x^*) \|}_{O(\| x - x^* \|^2)} \quad F' \text{ Lip in } x^*$$

$$\le \underbrace{\frac{2 \beta L}{1 - \beta L \| x^{(k)} - x^* \|}}_{\le C} \| x^{(k)} - x^* \|^2 \le C \, \| x^{(k)} - x^* \|^2 .$$

q.e.d.

Satz 5.10: $(11)_F \longrightarrow F(x)=0$ in $\mathbb{R}^n$

Vor.: Sei $F: \mathbb{R}^n \to \mathbb{R}^n$ einmal stetig partiell diffbar,
$x^* \in \mathbb{R}^n$ sei Nullstelle von F, d.h. $F(x^*)=0$.
$\nabla F(x^*) =: J(x^*)$ sei regulär und $\beta = \| J(x^*)^{-1} \| > 0$

Bh.: Dann ist das NEWTON-Verfahren $(11)_F$ für hinreichend gute Startwerte wohldefiniert und es konvergiert q-superlinear gegen x^*, d.h.

$$\lim_{k\to\infty} \underbrace{\frac{\|x^*-x^{(k+1)}\|}{\|x^*-x^{(k)}\|}}_{= q_k} = \lim_{k\to\infty} q_k = 0$$

$$\|x^*-x^{(k+1)}\| \le \underbrace{q_k}_{\to 0 \quad k\to\infty} \|x^*-x^{(k)}\| .$$

Falls ∇F in x^* Lipschitz-stetig ist, d.h. $\exists\, L=\text{const}>0$:

$$(*) \quad \|\nabla F(x) - \nabla F(x^*)\| \le L \|x-x^*\| \quad \forall x \in \bar{U}(x^*),$$

dann konvergiert das NEWTON-Verfahren q-quadratisch, d.h. es gilt (13).

Beweis: $x = x^{(k)}$, $F' = \nabla F = J = J_F$ = Jacobi-Matrix

$$x^{(k+1)} - x^* = \underbrace{x^{(k)} - F'(x^{(k)})^{-1} F(x^{(k)})}_{x^{(k+1)}} - x^*$$

$$= F'(x)^{-1}\left[0 - F(x) - F'(x)(x^*-x)\right]$$

$$= F'(x)^{-1}\left[\overset{\shortparallel}{F(x^*)} - (F(x) + F'(x)(x^*-x))\right]$$

$$\|x^{(k+1)}-x^*\| \le \underset{*}{\|F'(x)^{-1}\|}\ \underbrace{\|F(x^*) - (F(x)+F'(x)(x^*-x))\|}_{o(\|x-x^*\|)} \quad F\in C^1 \ !$$

$$\le \|F'(x)^{-1}\|\ \|F(x^*) - F(x) - F'(x)(x^*-x)\|$$

$\exists\, \xi \in (0,1)$

$$\le \|F'(x)^{-1}\|\ \|(F'(x+\xi(x^*-x)) - F'(x))(x^*-x)\|$$

$$\le \|F'(x)^{-1}\|\ \|(\underbrace{F'(x+\xi(x^*-x)) - F'(x^*)} + \underbrace{F'(x^*) - F'(x)})\|\ \|x^*-x\|$$

$$\le \|F'(x)^{-1}\| \left(L\|x+\xi(x^*-x)-x^*\| + L\|x^*-x\|\right) \|x^*-x\|$$

$(*)$ F' Lip

$$\le \frac{2\beta L}{1-\beta L\|x^*-x\|} \|x^*-x\|^2 \le C \|x^*-x\|^2 . \qquad \text{q.e.d.}$$

- **Bemerkung:**
  - ⊕ Im Gegensatz zum GV (q-linear) konvergiert das Newton-Verfahren sehr schnell (q-quadratisch)!
  - ⊖ Nur lokale Konvergenz, d.h. $x^{(0)} \in U(x^*)$!
    $\rightsquigarrow$ gedämpftes Newton mit Liniensuche!
  - ⊖ $H_f(x^{(k)})$ bzw. $J_F(x^{(k)})$ wird benötigt und in jedem Newton-Schritt ist lineares GS
    $$J_F(x^{(k)})\, s^{(k)} = d^{(k)}$$
    zu lösen!
    $\rightsquigarrow$ BFGS

- **Algorithmus 5.11:** $F(x)=0$ bzw. $\nabla f(x)=0$

Startnäherung: $x^{(0)} \in \mathbb{R}^n$, $B^{(0)}$, $\varepsilon = 10^{-?}$ TGS

Iteration: $k = 0, 1, \ldots, k_{stop}$

$d^{(k)} = -F(x^{(k)})$

$\|d^{(k)}\| \leq \varepsilon \|d^{(0)}\| \Longrightarrow$ STOP

$B^{(k)}\, s^{(k)} = d^{(k)}$ (Lin. GS lösen)

$\alpha^{(k+1)} =$ Schrittweitenwahl (Liniensuche)

$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)}\, s^{(k)}$

$B^{(k+1)} =$ Newton, BFGS, ... (?)

$$B^{(k)} = J_F(x^{(k)}) = H_f(x^{(k)})$$

$\alpha^{(k+1)} = 1$ – Newton, $\alpha^{(k+1)} < 1$ gedämpftes New.

$B^{(k+1)} = B^{(k)} + \ldots$ (14) BFGS $\rightarrow$ Pkt. 5.1.4.

5.1.4. Quasi-Newton-Verfahren

- **Idee:**

1. Ersetzen in $q_K(x)$ die Hesse-Matrix $H(x^{(k)})$ durch eine Approximation $B^{(k)}$.
   Dann erhält man anstelle der Newton-Richtung die Suchrichtung
   $$s^{(k)} = -B^{(k)-1}\nabla f(x^{(k)}) = -B^{(k)-1}F(x^{(k)}).$$
2. Wähle $B^{(0)}$ z.B. $B^{(0)} = H(x^{(0)})$ oder $B^{(0)} = I$.
   Für geg. $B^{(k)}$ berechnen neue Approximation $B^{(k+1)}$ der Hesse-Matrix im nächsten Iterationspkt so, daß die folgenden 3 Bedingungen erfüllt werden:

   1) Quasi-Newton-Bd.: (Sekantenbed.)
   $$B^{(k+1)}\underbrace{(x^{(k+1)}-x^{(k)})}_{=:w^{(k)}} = \underbrace{F(x^{(k+1)})-F(x^{(k)})}_{=:y^{(k)}}$$

   <u>Bem.:</u> Für $n=1$ ist das die Sekantenbed.:
   $$B^{(k+1)} = \frac{F(x^{(k+1)})-F(x^{(k)})}{x^{(k+1)}-x^{(k)}}$$
   $\rightsquigarrow$ Sekantenverfahren:
   $$x^{(k+2)} = x^{(k+1)} - \frac{x^{(k+1)}-x^{(k)}}{F(x^{(k+1)})-F(x^{(k)})}F(x^{(k+1)})$$

   2) $B^{(k+1)}$ ist SPD. Diese Bedingung sichert, daß die nächste Suchrichtung
   $$s^{(k+1)} = -(B^{(k+1)})^{-1}\nabla f(x^{(k+1)})$$
   eine Abstiegsrichtung ist. Tatsächlich
   $$\nabla f(x^{(k+1)})^T s^{(k+1)} = -\nabla f(x^{(k+1)})(B^{(k+1)})^{-1}\nabla f(x^{(k+1)}) < 0$$

   3) $B^{(k+1)} = B^{(k)} +$ Niedrigrangmatrix
   Rang-1 oder Rang-2

■ Broyden – Fletcher – Goldfarb – Shanno – Verfahrens

- Eines der wichtigsten Verfahren dieser Klasse ist das BFGS-Verfahren (= Rang-2-Verfahren),

$$(14)\quad B^{(k+1)} = B^{(k)} + \underbrace{\frac{1}{y^{(k)T} w^{(k)}}\, y^{(k)} y^{(k)T} + \frac{1}{w^{(k)T} B^{(k)} w^{(k)}}\, B^{(k)} y^{(k)} y^{(k)T} B^{(k)}}_{\text{Rang-2-Matrix}}$$

- **SATZ 5.12:**
  Falls $B^{(k)}$ SPD ist und falls $y^{(k)T} w^{(k)} > 0$, dann ist auch $B^{(k+1)}$ SPD.
  Beweis: Klar!

- **Bem. 5.13:**
  1. Erfüllt die Schrittweite $\alpha^{(k+1)}$ die Bed. (8), dann gilt $y^{(k)T} w^{(k)} > 0$.
  2. Unter geeigneten Voraussetzung läßt sich für das BFGS-Verfahren mit Liniensuche, die den Kriterien (7) und (8) genügt, zeigen, daß es global und superlinear konvergiert!

5. 2. Optimierungsprobleme (OP) mit Nebenbedingungen (NB)

- In diesem Abschnitt betrachten wir restringierte Optimierungsprobleme (rOP) mit Gleichungs- und Ungleichungsnebenbedingungen der Form

$(1)_{rOP}$

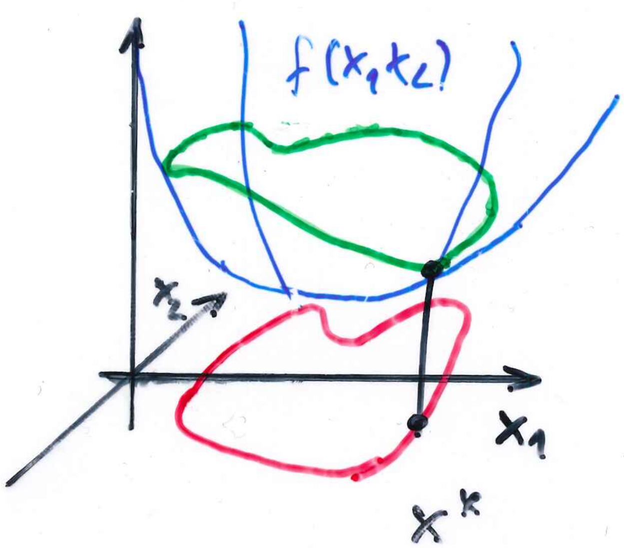
$$\boxed{\begin{array}{l} \text{Finde } x^* \in \mathbb{R}^n: \\ f(x^*) = \min\limits_{x \in \mathbb{R}^n} f(x) \\ \quad \text{s.t. } c_i(x) = 0, \; i \in I_1 = \{1, 2, \ldots, m_1\} \\ \quad \phantom{\text{s.t. }} c_i(x) \leq 0, \; i \in I_2 = \{m_1+1, \ldots, m_1+m_2\} \end{array}}$$

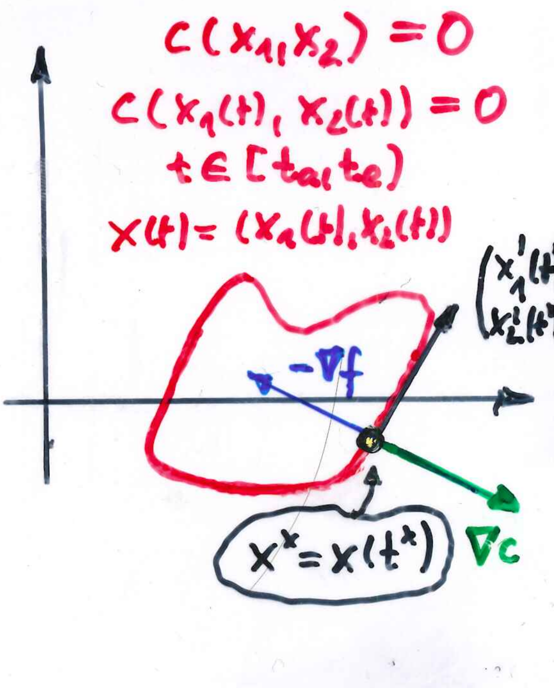
5.2.1. Lagrange's Idee (1788)

- Ges. ist ein lokales Extremum (= Min. oder Max.) der Funktion $f(x_1, x_2)$ unter der Nebenbedingung

$$c(x_1, x_2) = 0 \qquad (\text{d.h. } n = 2, \; m_1 = 1, \; m_2 = 0),$$

wobei die Fkten $f, c: \mathbb{R}^2 \to \mathbb{R}$ als stetig diffbar vorausgesetzt werden:





- Angenommen es existiert eine Parameterisierung der Lösung der NB, d.h.

$$\exists\, x = x(t) = (x_1(t), x_2(t)) : c(x_1(t), x_2(t)) = 0,\ t \in [t_a, t_e]$$

Durch Differentieren der NB nach t erhalten wir:

$$0 = \frac{d}{dt} c(x_1(t), x_2(t)) = \frac{\partial c}{\partial x_1} x_1' + \frac{\partial c}{\partial x_2} x_2' = \left(\nabla c, \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}\right)$$

Offbar gilt: $\boxed{\nabla c \perp \begin{pmatrix} x_1' \\ x_2' \end{pmatrix}}$

lok. Extr. $f(x_1, x_2)$ unter NB $c(x_1, x_2) = 0$	$\Longleftrightarrow$	freies lok. Extr. von $F(t) := f(x_1(t), x_2(t))$

Sei nun $(x_1^*, x_2^*) = (x_1(t^*), x_2(t^*))$ ein lok. Extr., d.h.

$$F'(t^*) = \frac{\partial f}{\partial x_1}(x_1^*, x_2^*)\, x_1'(t^*) + \frac{\partial f}{\partial x_2}(x_1^*, x_2^*)\, x_2'(t^*) = 0, \text{ d.h.}$$

$$\boxed{\nabla f(x_1^*, x_2^*) \perp \begin{pmatrix} x_1'(t^*) \\ x_2'(t^*) \end{pmatrix}}$$

$$\exists\, \lambda^* \in \mathbb{R} : \nabla f(x_1^*, x_2^*) = -\lambda^* \nabla c(x_1^*, x_2^*)$$

λ^* heißt Lagrange-Multiplikator.

- Resultat: Wir erhalten also für Extremalwertprobleme

$$\boxed{\begin{aligned} &f(x_1, x_2) \longrightarrow \text{extr. (min/max)} \\ &\text{s.t. } c(x_1, x_2) = 0 \end{aligned}}$$

die notwendige Extremalbedingung

$$\begin{aligned} \nabla_x L(x^*, \lambda^*) &= \\ \nabla_\lambda L(x^*, \lambda^*) &= \end{aligned} \boxed{\begin{aligned} \nabla f(x_1^*, x_2^*) + \lambda^* \nabla c(x_1^*, x_2^*) &= 0 \\ c(x_1^*, x_2^*) &= 0 \end{aligned}} = \text{nGS},$$

die wir mit Hilfe der sogenannten Lagrange-Fkt.

$$L(x, \lambda) = L(x_1, x_2, \lambda) := f(x_1, x_2) + \lambda_1 c_1(x_1, x_2)$$

auch kompakt in der Form $\boxed{\nabla L(x^*, \lambda^*) = 0}$ schreiben können.

5.2.2. Theoretische Grundlagen

- **Optimierungsprobleme mit Gleichungsnebenbed.:**
  - Btr. (1)$_{rop}$ mit $m = m_1 < n$ und $m_2 = 0$, d.h.

$m_1 = m$, $m_2 = 0$

$$(15) \quad \min_{x \in \mathbb{R}^n} f(x_1, x_2, \dots, x_n) \qquad \min_{x \in \mathbb{R}^n} f(x)$$

$$\text{s.t. } c_1(x_1, \dots, x_m, x_{m+1}, \dots, x_n) = 0 \qquad \text{s.t. } c(x) = 0$$
$$\vdots \qquad\qquad \vdots$$
$$c_m(x_1, \dots, x_m, x_{m+1}, \dots, x_n) = 0$$

Bem.: Satz über implizite Fkten: (15) ⟹ fOP

T →

- **Def. 5.15:** (Lagrange-Fkt., Lagrange-Multipl.)
  Die zu (f, c) = (Zielfkt, NB) gehörige Lagrange-Funktion ist definiert als:

$$(16) \quad L(x, \lambda) = L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) :=$$
$$:= f(x) + \underbrace{(\lambda, c(x))}_{\mathbb{R}^m} = f(x_1, \dots, x_n) + \sum_{i=1}^{m} \lambda_i c_i(x_1, \dots, x_n)$$

Parameter

Die $\lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$ heißen Lagrange-Multiplik.

- **Satz 5.16:** (Notwendige Extremalbed.)
  Vor.: 1. Seien $f: \mathbb{R}^n \to \mathbb{R}$ und $c: \mathbb{R}^n \to \mathbb{R}^m$ stetig diffbar und sei $m \le n$.
  2. x^* sei Stelle eines lokalen Extremum von f unter der NB $c(x^*) = 0$:

CQ

$$\text{rang } c'(x^*) = \text{rang}\left[\frac{\partial c}{\partial x}\right] = \text{rang}\left[\frac{\partial c_i}{\partial x_j}\right]_{\substack{i=\overline{1,m}\\ j=\overline{1,n}}} = m.$$

  Bh.: Dann $\exists! \lambda^* = (\lambda_1^*, \dots, \lambda_m^*) \in \mathbb{R}^m$:

$$(18) \quad \boxed{\nabla L(x^*, \lambda^*) = 0} \qquad \text{nGS in } \mathbb{R}^{n+m}$$

T → Bsp.: Hot Spot

- **Bem. 5.17:**

1. Ist x Stelle eines lokalen Extremums von f unter den Nebenbedingungen
$$c_1(x)=0,\ldots,c_m(x)=0,$$
so gilt unter den Voraussetzungen von Satz 4.16
$$\exists\ \lambda_1^*,\ldots,\lambda_m^*\in\mathbb{R}:$$

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

(18)

in R^{n+m}

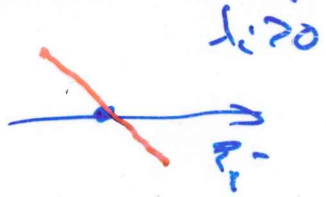
$$\boxed{\begin{aligned}\frac{\partial f}{\partial x_j}(x^*)+\sum_{i=1}^{m}\lambda_i^*\frac{\partial c_i}{\partial x_j}(x^*)&=0,\quad j=1,2,\ldots,n\\ c_i(x^*)&=0,\quad i=1,2,\ldots,m\end{aligned}}$$

Das ist ein nichtlineares Gleichungssystem von $(n+m)$ nichtlinearen Gleichungen zur Bestimmung der $(n+m)$ Unbekannten $(x^*,\lambda^*)=(x_1^*,\ldots,x_n^*,\lambda_1^*,\ldots,\lambda_m^*)$.

2. Die Gleichungen (18) sind nur notwendige Bedingungen für das Vorliegen eines lokalen Extr.!

3. Außerdem erhalten wir nur lokale Extrema, in denen die Bedingung rang $c'(x^*)=m$ erfüllt ist!

4. Die Lagrange-Multiplikatoren $(-\lambda_i)$ sind ein Maß für die Empfindlichkeit auf die Verletzung der i-ten Nebenbedingung, d.h.
$$\frac{\partial f}{\partial \xi_i}(x^*(\xi_1,\ldots,\xi_m))\Big|_{\xi=(\xi_1,\ldots,\xi_m)=0}\overset{\text{(mms)}}{=}-\lambda_i,$$
wobei $x^*(\xi_1,\ldots,\xi_m)$ ein lokales Extremum von f unter den Nebenbedingungen $c(x)=\xi$ ist, d.h.
$$\begin{aligned}c_1(x_1^*,\ldots,x_n^*)&=\xi_1,\\ &\vdots\\ c_m(x_1^*,\ldots,x_n^*)&=\xi_m.\end{aligned}$$



OP mit Gleichungs- und Ungleichungsnebenbedingungen:

- Analog zur Def. 4.15 ordnen wir auch dem OP $(1)_{rOP}$ die Lagrange-FKt. $L: \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}$ mit

$$(16) \quad L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i c_i(x) = f(x) + (\lambda, c(x))$$

zu, wobei $m = m_1 + m_2$ und $\lambda = (\lambda_1, \dots, \lambda_m)^T \in \mathbb{R}^m$ - Lag. Mult.

- Bezeichnungen:
  1) Gradient der Lagrange-FKt. bzgl. x:
  $$\nabla_x L(x,\lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla c_i(x).$$
  2) Hesse-Matrix der Lagrange-FKt. bzgl. x:
  $$\nabla_{xx} L(x,\lambda) = \nabla^2 f(x) + \sum_{i=1}^{m} \lambda_i \nabla^2 c_i(x) \quad \text{(sym.)}$$
  3) Menge der Indizes der in x aktiven NB:
  $$\mathcal{A}(x) := \{ i \in I_1 \cup I_2 = \{1,2,\dots,m\} : c_i(x) = 0 \}.$$

- Satz 5.18: (notwendige Optimalitätsbedingung)
  Seien $f: \mathbb{R}^n \to \mathbb{R}$ und $c: \mathbb{R}^n \to \mathbb{R}^m$ stetig diffbar und sei $x^* \in \mathbb{R}^n$ ein lokales Min. des restr. OP $(1)_{rOP}$.
  Falls $\{ \nabla c_i(x^*) : i \in \mathcal{A}(x^*) \}$ linear unabhängig sind (Linear Independence Constraint Qualification = LICQ),
  $(x^*, \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$: KKT-Pkt.
  dann existiert genau ein Vektor $\lambda^* \in \mathbb{R}^m$:

$$(19) \quad \boxed{\begin{aligned} \nabla_x L(x^*,\lambda^*) = \nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla c_i(x^*) &= 0 \\ c_i(x^*) &= 0 \text{ für } i \in I_1 \\ c_i(x^*) \le 0,\ \lambda_i^* \ge 0,\ \lambda_i^* c_i(x^*) &= 0,\ i \in I_2 \end{aligned}} = \text{KKT-Bd.}$$

Karush Kuhn Tucker

- Satz 5.19: (hinreichende Bedingungen)
  Falls $x^* \in \mathbb{R}^n$ und $\lambda^* \in \mathbb{R}^m$ die KKT-Bed. (19) erfüllen und $\nabla_{xx} L(x^*,\lambda^*)$ p.d. ist, dann ist x^* ein striktes lok. Min.

- Bem.: $s^T \nabla_{xx} L(x^*,\lambda^*) s > 0 \ \forall s \in \mathbb{R}^n, s \neq 0 \ \leadsto \ \forall s \in C(x^*,\lambda^*),\ s \neq 0$
  Cone of Critical Directions

5.2.3. Das SQP-Verfahren

- Beim NEWTON-Verfahren (vgl. Abs. 5.1.3) bzw. bei Quasi-NEWTON-Verfahren (vgl. Abs. 5.1.4) der freien Optimierung wurde die Suchrichtung $s^{(k)}$ im Iterationspkt. $x^{(k)}$ als Lösung des freien quadratischen Optimierungsproblems (QOP)

$f(x^{(k)}+s) =$ $$(20)_{fQOP} \quad \cancel{f(x^{(k)})} + \nabla^T f(x^{(k)})\, s + \frac{1}{2} s^T \underset{SPD}{B^{(k)}} s \longrightarrow \min_{s \in \mathbb{R}^n}$$

$\cancel{+\, o(\|s\|^2)}$

$$\Leftrightarrow \quad B^{(k)} s = -\nabla f(x^{(k)})$$

wobei $B^{(k)} := H(x^{(k)}) = \nabla_{xx} f(x^{(k)}) = \nabla^2 f(x^{(k)})$ = NEWTON

$x^{(k+1)} = x^{(k)} + \underset{\approx 1}{\alpha^{(k+1)}} s^{(k)}$ $\quad B^{(k)} := (14)$ = BFGS

- Analog werden das NEWTON-Verfahren und die Quasi-Newton-Verfahren für restringierte OP der Form $(1)_{rOP}$ formuliert:
  Zu geg. Näherungen $x^{(k)} \approx x^*$ und $\cancel{\lambda^{(k)} \approx \lambda^*}$ wird die nächste Suchrichtung $s^{(k)} \in \mathbb{R}^n$ als Lösung des folgenden quadratischen OP mit "Linearisierten" NB bestimmt:

$$(20)_{rQOP} \quad \boxed{\begin{aligned} &\cancel{f(x^{(k)})} + \nabla^T f(x^{(k)})\, s + \frac{1}{2} s^T B^{(k)} s \to \min_{s \in \mathbb{R}^n} \\ &\text{s.t. } c_i(x^{(k)}) + \nabla^T c_i(x^{(k)}) \cdot s = 0, \; i \in I_1 \\ &\qquad c_i(x^{(k)}) + \nabla^T c_i(x^{(k)})\, s \le 0, \; i \in I_2 \end{aligned}}$$

$\underline{c_i(x^{(k)}+s)} = 0$

≤ 0

Taylor

Die nächste Näherung ergibt sich dann zu:

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} s^{(k)}, \qquad \text{z.B. } \alpha^{(k+1)} = 1$$

$\lambda^{(k+1)}$ = der zu $s^{(k)}$ gehörende Lagrange-Parameter des rQOP $(20)_{rQOP}$.

Zur Wahl von $B^{(k)}$:

1. $B^{(k)} = L_{xx}(x^{(k)}, \lambda^{(k)})$ = Wilson-Verfahren
   $\uparrow$ $\hat{=}$ Newton-Verfahren
   $f_{xx} = \nabla^2 f$ (fOP)

2. $B^{(k)}$ = Quasi-Newton z.B. BFGS

Zur Bestimmung der Schrittweite $\alpha^{(k+1)}$:

$\rightarrow$ Liniensuche nach den Kriterien (7) und (8)!

- **Algorithmus 5.20:** SQP-Verfahren: Wilson

Start: Wähle Anfangsnäherung $(x^{(0)}, \lambda^{(0)}) \in \mathbb{R}^n \times \mathbb{R}^m$.

Iteration: $k = 0, 1, 2, \ldots, k_{stop}$ KKT-Syst.

Ist $(x^{(k)}, \lambda^{(k)})$ ein KKT-Pkt, d.h. Lsg. von (15): STOP!

Berechne einen KKT-Pkt. $(x^{(k+1)}, \lambda^{(k+1)}) \in \mathbb{R}^n \times \mathbb{R}^m$ des quadratischen OP = QPO:

$$\boxed{\begin{aligned} &\min_{x \in \mathbb{R}^n} \nabla f(x^{(k)})^T (x - x^{(k)}) + \tfrac{1}{2}(x - x^{(k)}) L_{xx}(x^{(k)}, \lambda^{(k)})(x - x^{(k)}) \\ &\text{s.t. } c_i(x^{(k)}) + \nabla^T c_i(x^{(k)})(x - x^{(k)}) = 0, \quad i \in I_1 \\ &\quad\;\; c_i(x^{(k)}) + \nabla^T c_i(x^{(k)})(x - x^{(k)}) \le 0, \quad i \in I_2 \end{aligned}}$$

Besitzt dieses QOP mehrere KKT-Punkte, so wählt man den KKT-Punkt $(x^{(k+1)}, \lambda^{(k+1)}) \in \mathbb{R}^n \times \mathbb{R}^m$:

$\|(x^{(k+1)}, \lambda^{(k+1)}) - (x^{(k)}, \lambda^{(k)})\| \longrightarrow \min.$

- Unter geeigneten Bedingungen kann lokale superlineare bzw. sogar quadratische Konvergenz gezeigt werden.
(siehe z.B. Satz 5.31 in Geiger / Kanzow [12]).

5.2.4. Quadratische Optimierungsprobleme

- Btr. das allgemeine restringierte quadratische Optimierungsproblem (rq OP = QP = Quadratic Programming)

(21) (QP)

$$\boxed{\begin{aligned} &\text{Ges. } x^* \in \mathbb{R}^n: \\ &q(x^*) = \min_{x \in \mathbb{R}^n} \; q(x) := \overbrace{\cancel{f} + g^T \cdot x + \tfrac{1}{2} x^T B x}^{\text{quadratisch}} \\ &\qquad \text{s.t. } c_i + a_i^T x = 0, \; i \in I_1 = \{1, \dots, m_1\} \\ &\qquad\qquad \underbrace{c_i + a_i^T x}_{\text{(affine) Linear}} \leq 0, \; i \in I_2 = \{m_1+1, \dots, \underbrace{m_1+m_2}_{=m}\} \end{aligned}}$$

mit geg. $f, c_i \in \mathbb{R}$ und $g, a_i \in \mathbb{R}^n$ sowie geg. Matrix $B \in \mathbb{R}^{n \times n}$ spd

- **Bem.:** (QP) mit reinen Gleichungsnebenbed., d.h. $I_2 = \emptyset$, führt nach Satz 5.16 / Bem. 5.17 auf die Lösung des ~~nicht~~ linearen GS

(22) = $(18)_{QP}$ $\quad \boxed{\nabla L(x, \lambda) = 0}$ $\quad$ KKT-System

mit der Lagrange-FKt. $(m = m_1)$

$$L(x, \lambda) = q(x) + \sum_{i=1}^{m} \lambda_i (c_i + \overbrace{a_i^T x}^{\sum_{k=1}^{n} a_{ik} x_k}),$$

d.h. (22) ergibt sich zu

(22)

$$\boxed{\begin{aligned} \tfrac{\partial L}{\partial x_j} = \quad & \frac{\partial q}{\partial x_j}(x) + \sum_{i=1}^{m} \lambda_i \frac{\partial}{\partial x_j}(c_i + a_i^T x) = 0, \quad j = 1, \dots, n \\ \tfrac{\partial L}{\partial \lambda_i} = \quad & c_i + a_i^T x = 0, \quad i = 1, \dots, m \end{aligned}}$$

$$\Updownarrow$$

(22)

$$\boxed{\begin{aligned} & g_j + (Bx)_j + \sum_{i=1}^{m} \lambda_i a_{ij} = 0, \quad j = 1, \dots, n \\ & c_i + a_i^T x = 0, \quad i = 1, \dots, m \end{aligned}}$$

$$\Updownarrow$$

KKT (22)

$$\begin{bmatrix} B & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -g \\ -c \end{bmatrix} \quad \text{mit } A = \begin{bmatrix} a_1^T \\ \vdots \\ a_m^T \end{bmatrix} = \underbrace{\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \\ a_{m1} & \dots & a_{mn} \end{bmatrix}}_{(m \times n)\text{-Matrix}}$$

Sei $x^{(k)}$ ein zuläßiger Pkt. des (QP)
mit reinen Gleichungsnebenbed., d.h.

NB: $A x^{(k)} = -c$ bzw. $a_i^T x = -c_i$, $i = \overline{1, m = m_1}$

Dann gilt für die Korrektur Δx in der Darstellung

$$x^{(k+1)} = x^{(k)} + \Delta x$$

wegen (22) die folgende Beziehung

$$\begin{bmatrix} B & A^T \\ A & 0 \end{bmatrix} \overset{\text{KKT-Pkt}}{\begin{bmatrix} x \\ \lambda \end{bmatrix}} = \begin{bmatrix} -g \\ -c \end{bmatrix} \qquad \text{KKT-System}$$

$$\Leftrightarrow \begin{bmatrix} B & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^{(k)} + \Delta x \\ \lambda \end{bmatrix} = \begin{bmatrix} -g \\ -c \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} B & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \lambda \end{bmatrix} = \begin{bmatrix} -g \\ -c \end{bmatrix} - \begin{bmatrix} B & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^{(k)} \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -g - B x^{(k)} \\ -c - A x^{(k)} \end{bmatrix}$$

$$= \begin{bmatrix} -\nabla q(x^{(k)}) \\ 0 \end{bmatrix}$$

$$\boxed{\begin{aligned} B \Delta x + A^T \lambda &= -\nabla q(x^{(k)}) \\ A \Delta x &= 0 \end{aligned}}$$

Algorithmus 5.21: Strategie der aktiven Mengen → (QP)

Start: Bestimme ein für (21)=(QP) zulässige Startnäherung $x^{(0)} \in \mathbb{R}^n$ sowie einen zugehörigen Lagrange-Multiplikator $\lambda^{(0)} \in \mathbb{R}^n$.

Setze $\mathcal{A}_0 := \mathcal{A}(x^{(0)}) := \{ i \in I_1 \cup I_2 : c_i + a_i^T x^{(0)} = 0 \}$.

Schritte: $k = 0, 1, 2, \ldots, k_{STOP}$ (= STOP)

(S1) Ist $(x^{(k)}, \lambda^{(k)})$ ein KKT-Pkt. von (21): STOP!

(S2) Setze $\lambda_i^{(k+1)} = 0$ für $i \notin \mathcal{A}_k$ und bestimme $(\Delta x^{(k)}, \lambda_{\mathcal{A}_k}^{(k+1)}) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{A}_k|}$:

$$\begin{bmatrix} B & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} \Delta x^{(k)} \\ \lambda_{\mathcal{A}_k}^{(k+1)} \end{bmatrix} = \begin{bmatrix} -\nabla q(x^{(k)}) \\ 0 \end{bmatrix},$$

wobei $\lambda_{\mathcal{A}_k} := [\lambda_i]_{i \in \mathcal{A}_k} \in \mathbb{R}^{|\mathcal{A}_k|}$, $A_k = [a_i^T]_{i \in \mathcal{A}_k}$.

(S3) Unterscheide die folgenden Fälle:

1) Ist $\Delta x^{(k)} = 0$ und $\lambda_i^{(k+1)} \geq 0 \;\forall i \in \mathcal{A}_k \cap I_2$: STOP!

2) Ist $\Delta x^{(k)} = 0$ und $\min\{ \lambda_i^{(k+1)} : i \in \mathcal{A}_k \cap I_2 \} < 0$, so bestimme einen Index $r \in \mathcal{A}_k \cap I_2$:

$$\lambda_r^{(k+1)} = \min\{ \lambda_i^{(k+1)} : i \in \mathcal{A}_k \cap I_2 \}.$$

Setze $x^{(k+1)} = x^{(k)}$ und $\mathcal{A}_{k+1} := \mathcal{A}_k \setminus \{r\}$: GOTO (S1)

3) Ist $\Delta x^{(k)} \neq 0$ und $x^{(k)} + \Delta x^{(k)}$ zulässig für (21), so setze $x^{(k+1)} := x^{(k)} + \Delta x^{(k)}$ und $\mathcal{A}_{k+1} = \mathcal{A}_k$: GOTO (S1)

4) Ist $\Delta x^{(k)} \neq 0$ und $x^{(k)} + \Delta x^{(k)}$ nicht zulässig für (21), so bestimme einen Index r mit

$$0 < \alpha^{(k+1)} \overset{\leq}{:=} -\frac{c_r + a_r^T x^{(k)}}{a_r^T \Delta x^{(k)}} = \min\left\{ -\frac{c_i + a_i^T x^{(k)}}{a_i^T \Delta x^{(k)}} : i \notin \mathcal{A}_k \wedge a_i^T \Delta x^{(k)} > 0 \right\}$$

Setze $x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} \Delta x^{(k)}$ und $\mathcal{A}_{k+1} := \mathcal{A}_k \cup \{r\}$: GOTO (S1)

GOTO (S1) means GOTO (S1) $\wedge$ $k \leftarrow k+1$

5.2.5. Weitere Verfahren zur Lösung von rOP mit NB in Gleichungs- und Ungleichungsform

$(1)_{rOP}$

$$\boxed{\begin{aligned} &\text{Ges. } x^* \in \mathbb{R}^n: \quad f(x^*) = \min_{x \in \mathbb{R}^n} f(x) \\ &\qquad \text{s.t. } c_i(x) = 0, \; i \in I_1 = \{1, \dots, m_1\} \\ &\qquad\quad\;\; c_i(x) \le 0, \; i \in I_2 = \{m_1+1, \dots, m_1+m_2\} \end{aligned}}$$

Penalty-Methoden:

- Idee: 1. Ersetze $(1)_{rOP}$ durch ein r.OP mit Gleichungsnebenbed.:

(23)
$$\min_{x \in \mathbb{R}^n} f(x)$$
$$\text{s.t. } \tilde{c}_i(x) := c_i(x) = 0, \; i \in I_1$$
$$\tilde{c}_i(x) := \max\{0, c_i(x)\} = 0, \; i \in I_2$$

Semi-Smooth Newton

2. Ersetze (23) durch eine Folge von fOPs mit Strafparameter $\alpha_K \nearrow \infty$

$$\min_{x \in \mathbb{R}^n} P(x, \alpha_K)$$

mit der StrafFkt. (z.B.)

$$P(x, \alpha) := f(x) + \frac{\alpha}{2} \sum_{i=1}^{m} (\tilde{c}_i(x))^2.$$

Barriere-Methoden:

$(1)_{rOP} \longmapsto (24) \quad \min_{x \in \mathbb{R}^n} B(x, \alpha_K) \qquad (\alpha_K \searrow 0)$

$$\text{s.t. } c_i(x) = 0, \; i \in I_1$$

mit $B(x, \alpha) := f(x) - \alpha \sum_{i=m_1+1}^{m_2} \ln(-c_i(x))$ oder $:= f(x) - \alpha \sum_{i=m_1+1}^{m_2} \frac{1}{c_i(x)}$

- ... siehe Lit z.B. [12] Geiger / Kanzow