I.1 Allgemeine unrestringierte Minimierungsverfahren

Sei $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ stetig differenzierbar.

(1.1) Minimierungsalgorithmus (Prototyp)

S0) Wähle Start-Iterierte $x^0 \in \mathbb{R}^n$, $\varepsilon > 0$, $C > 0$. Setze $k = 0$.

S1) Falls $\|Df(x^k)\| < \varepsilon$ STOP (x^k numerische Lösung)
Falls $\|x^k\| > C$ STOP (Divergenz)

S2) Wähle Suchrichtung $d^k \in \mathbb{R}^n$

S3) Bestimme Schrittweite $t_k > 0$ mit

$$f(x^k + t_k d^k) < f(x^k)$$

S4) Update: $x^{k+1} = x^k + t_k d^k$, $k := k+1$. Gehe zu S1).

Beispiel: Gradienten-Verfahren $d^k = -Df(x^k)^T = -\nabla f(x^k)$

I.1 Allgemeine unrestringierte Minimierungsverfahren

(1.2) Der Minimierungsalgorithmus heißt

a) *wohldefiniert*, wenn die Auswahlregeln zur Suchrichtung und zur Schrittweite immer durchführbar sind;

b) *konsistent*, wenn das Verfahren stoppt, falls $x^k = x^*$ ein Minimum ist;

c) *(lokal) konvergent*, wenn zu jedem Minimum x^* ein $\delta > 0$ existiert, sodass zu jedem $\varepsilon > 0$ ein $k_0 > 0$ existiert mit

$$\|x^0 - x^*\| < \delta \quad \Rightarrow \quad \|x^k - x^*\| < \varepsilon \text{ für } k > k_0;$$

d) *global konvergent*, wenn für alle $x^0 \in \mathbb{R}^n$ und alle $\varepsilon > 0$ ein Minimum x^* und ein $k_0 > 0$ existiert mit

$$\|x^k - x^*\| < \varepsilon;$$

e) *effizient*, wenn die Schrittweitenwahl eine echte Reduktion von f bewirkt und ein $\theta > 0$ existiert, sodass zu jedem $x^0 \in \mathbb{R}^n$ die Iterierten x^k, Suchrichtungen d^k und Schrittweiten t_k

$$f(x^k + t_k d^k) \leq f(x^k) - \theta \left(\frac{\nabla f(x^k)^T d^k}{\|d^k\|} \right)^2$$

erfüllen.

I.1 Allgemeine unrestringierte Minimierungsverfahren

(1.3) Für einen Minimierungsalgorithmus gelte:

a) Winkelbedingung: Es existiert $\delta > 0$ mit $-\dfrac{\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\| \, \|d^k\|} \geq \delta$;

b) die Schrittweiten $t_k > 0$ sind effizient.

Dann ist jeder Häufungspunkt x^* der Folge $\{x^k\}_{k\in\mathbb{N}}$ ein stationärer Punkt (d.h. $\nabla f(x^*) = 0$).

(1.4) Für f gelte:

i) Die Levelmenge $\mathscr{L}(x^0) = \{x \in \mathbb{R}^n \colon f(x) \leq f(x^0)\}$ sei konvex

ii) f sei gleichmäßig konvex auf $\mathscr{L}(x^0)$, d.h. es existiert $c_0 > 0$ mit

$$f(x) - f(y) \geq \nabla f(y)^T (x - y) + c_0 \|x - y\|^2, \qquad x, y \in \mathscr{L}(x^0).$$

Für den Minimierungsalgorithmus gelte:

a) Schwache Winkelbedingung: $\displaystyle\sum_{k=0}^{\infty} \delta_k = \infty$ für $\delta_k = \left(\dfrac{\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\| \, \|d^k\|}\right)^2$

b) Die Schrittweiten sind effizient.

Dann konvergiert $\{x^k\}_{k\in\mathbb{N}}$ gegen das eindeutige Minimum x^*.

I.1 Allgemeine unrestringierte Minimierungsverfahren

Schrittweitenregel: Wähle $\sigma \in (0,1)$, $\rho \in (\sigma,1)$.

1) Armijo-Regel: wähle $t > 0$ mit
$f(x+td) \leq f(x) + t\sigma\nabla f(x)^T d$

2) Wolfe-Powell-Regel: wähle $t > 0$ mit
$f(x+td) \leq f(x) + t\sigma\nabla f(x)^T d$ und $\nabla f(x+td)^T d \geq \rho\nabla f(x)^T d$

3) Strenge Wolfe-Powell-Regel: wähle $t > 0$ mit
$f(x+td) \leq f(x) + t\sigma\nabla f(x)^T d$ und $|\nabla f(x+td)^T d| \leq -\rho\nabla f(x)^T d$

(1.5) Sei $\nabla f(x)^T d < 0$.

a) Die Armigo-Regel ist wohldefiniert.

b) Falls f nach unten beschränkt ist, sind die Wolfe-Powell-Regeln wohldefiniert.

(1.6) ∇f sei Lipschitz-stetig auf $\mathscr{L}(x^0)$, d.h.

$$\|\nabla f(x) - \nabla f(y)\| \leq L\,\|x-y\| \qquad x,y \in \mathscr{L}(x^0).$$

Dann sind die Wolfe-Powell-Regeln effizient.

I.1 Allgemeine unrestringierte Minimierungsverfahren

(1.7) Eine Folge $\{x^k\}$ konvergiert gegen x^* mindestens

a) *linear*, wenn $c \in (0,1)$ und $k_0 > 0$ existiert mit

$$\|x^{k+1} - x^*\| \le c\,\|x^k - x^*\| \qquad \text{für } k \ge k_0$$

b) *superlinear*, wenn zu jedem $\varepsilon > 0$ ein $k_0 > 0$ existiert mit

$$\|x^{k+1} - x^*\| \le \varepsilon\,\|x^k - x^*\| \qquad \text{für } k \ge k_0$$

c) *quadratisch*, wenn $x^k \to x^*$ konvergiert, und wenn $C > 0$ und $k_0 > 0$ existieren mit

$$\|x^{k+1} - x^*\| \le C\,\|x^k - x^*\|^2 \qquad \text{für } k \ge k_0.$$

I.2 Newton-Verfahren

(2.1) Globalisiertes Newton-Verfahren

S0) wähle $x^0 \in \mathbb{R}^n, \rho > 0, p > 2, \beta \in (0,1), \sigma \in \left(0, \frac{1}{2}\right), \varepsilon \geq 0$, setze $k = 0$

S1) falls $\|\nabla f(x^k)\| \leq \varepsilon$ STOP

S2) bestimme d^k mit $D^2 f(x^k) d^k = -\nabla f(x^k)$ falls d^k nicht berechenbar ist oder die Bedingung

$$\nabla f(x^k)^T d^k \leq -\rho \|d^k\|^p$$

verletzt ist, setze $d^k = -\nabla f(x^k)$

S3) bestimme $t_k \in \{\beta^j : j = 0, 1, 2, \dots\}$ maximal mit

$$f(x^k + t_k d^k) < f(x^k) + t_k\, \sigma \nabla f(x^k)^T d^k$$

S4) setze $x^{k+1} = x^k + t_k d^k$, $k := k + 1$, gehe zu S1).

(2.2) Jeder Häufungspunkt x^* von $\{x^k\}_{k \in \mathbb{N}}$ ist ein stationärer Punkt von f.

I.2 Newton-Verfahren

(2.3) Wenn x^* ein isolierter Häufungspunkt von $\{x^k\}$ ist, dann konvergiert bereits die gesamte Folge $\{x^k\}_{k\in\mathbb{N}}$ gegen x^*.

(2.4) Die Levelmenge $\mathscr{L}(x^0)$ sei konvex, ∇f sei Lipschitz-stetig und f sei gleichmäßig konvex auf $\mathscr{L}(x^0)$, d.h.

$$d^T D^2 f(x) d \geq 2c_0 \|d\|^2 \quad \text{für} \quad x \in \mathscr{L}(x^0), \quad d \in \mathbb{R}^n.$$

Dann ist das gedämpfte Newton-Verfahren mit $d^k = -D^2 f(x^k)^{-1} \nabla f(x^k)$ und $t_k = \frac{c_0}{L - c_0}$ konvergent.

(2.5) Sei x^* ein stationärer Punkt von f, und sei $D^2 f(x^*)$ positiv definit. Dann existiert ein $\delta > 0$, sodass gilt:

a) $D^2 f(x)$ positiv definit für $\|x - x^*\| < \delta$

b) für x^0 mit $\|x^0 - x^*\| < \delta$ konvergiert das Newton-Verfahren mit

$$d^k = -D^2 f(x^k)^{-1} \nabla f(x^k), \quad t_k = 1$$

superlinear.

c) Falls $D^2 f$ Lipschitz stetig ist, d.h.

$$\|D^2 f(x) - D^2 f(y)\| \leq \gamma \|x - y\| \; \text{für alle} \; \|x - x^*\| < \delta \; \text{und} \; \|y - x^*\| < \delta$$

ist die Konvergenz quadratisch.

I.2 Newton-Verfahren

(2.6) Satz von Newton-Kantorovich
Für $x^0 \in \mathbb{R}^n$ und $\alpha, \beta, \gamma > 0$ gelte:
a) $\|\nabla f(x^0)\| \leq \alpha$
b) $d^T D^2 f(x^0) d \geq \beta \|d\|^2 \qquad d \in \mathbb{R}^n$
c) $\|D^2 f(x) - D^2 f(y)\| \leq \gamma \, \|x - y\|$ für $\|x - x^0\| \leq r_0$, $\|y - x^0\| \leq r_0$, $r_0 = \frac{2\alpha}{\beta}$
d) $2\alpha\gamma \leq \beta^2$.
Dann ist das ungedämpfte Newton-Verfahren mit $x^{k+1} = x^k - D^2 f(x^k) \nabla f(x^k)$ wohldefiniert, und $\{x^k\}$ konvergiert gegen x^* mit $\nabla f(x^*) = 0$,

$$\|x^* - x^0\| \leq r_1 := \frac{r_0}{1 + \sqrt{1 - \frac{2\alpha\gamma}{\beta^2}}} = \frac{2\alpha}{\beta + \sqrt{\beta^2 - 2\alpha\gamma}}$$

und

$$d^T D^2 f(x^*) d \geq (\beta - \gamma\|x^* - x^0\|) \, \|d\|^2 \quad d \in \mathbb{R}^n$$

Falls sogar $2\alpha\gamma < \beta^2$ ist x^* lokales Minimum.

3. Inexakte Newton-Verfahren

(3.1) Lokales inexakte Newton-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^n$, $\varepsilon \geq 0$, setze $k = 0$

S1) falls $\|\nabla f(x^k)\| \leq \varepsilon$ STOP

S2) Wähle $\eta_k > 0$ und berechne $d^k \in \mathbb{R}^n$ mit

$$\|\nabla f(x^k) + D^2 f(x^k) d^k\| \leq \eta_k \|\nabla f(x^k)\|$$

S3) setze $x^{k+1} = x^k + d^k$, $k := k+1$, gehe zu S1)

(3.2) Sei x^* ein strenges lokales Minimum.
Dann existiert $\delta > 0$, sodass für alle $\|x^0 - x^*\| < \delta$ gilt:

a) Es existiert $\bar{\eta} \in (0,1)$, sodass für $\eta_k \leq \bar{\eta}$ die Folge $\{x^k\}$ wohldefiniert ist und linear konvergiert.

b) Für $\eta_k \longrightarrow 0$ konvergiert $\{x^k\}$ superlinear.

c) Falls $D^2 f$ lokal Lipschitz-stetig und $\eta_k = O\big(\|\nabla f(x^k)\|\big)$, so konvergiert $\{x^k\}$ quadratisch.

3. Inexakte Newton-Verfahren

(3.3) PCG-Algorithmus zur Lösung von $Hd = r$

Sei H symmetrisch positiv definit, und sei B ein symmetrisch positiv definiter Vorkonditionierer.

S0) Wähle $\eta \geq 0$, setze $\varepsilon = \eta \|r\|, d_{cg}^0 = 0,\ r^0 = r - Hd_{cg}^0, p^0 = Br^0, m = 0$

S1) falls $\|r^m\| \leq \varepsilon$ STOP

S2) Setze

$$\begin{aligned}
t_m &= \frac{(r^m)^T p^m}{(p^m)^T H p^m} \\
d_{cg}^{m+1} &= d_{cg}^m + t_m p^m \\
r^{m+1} &= r^m - t_m H p^m \\
\beta_m &= \frac{(r^{m+1})^T B r^{m+1}}{(r^m)^T B r^m} \\
p^{m+1} &= B r^{m+1} - \beta_m p^m
\end{aligned}$$

Setze $m := m + 1$ und gehe zu S1)

3. Inexakte Newton-Verfahren

(3.4) Globalisiertes inexaktes Newton-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^n, \rho > 0, p > 2, \beta \in (0,1)$, $\sigma \in \left(0, \frac{1}{2}\right), \varepsilon \geq 0$. Setze $k = 0$.

S1) Setze $H_k = D^2 f(x^k)$, wähle symmetrisch positiv definiten Vorkonditionierer B_k
falls $\|B_k \nabla f(x^k)\| < \varepsilon$ STOP

S2) Wähle $\eta_k > 0$ und bestimme d^k mit $\|H_k d^k + \nabla f(x^k)\| < \eta_k \|\nabla f(x^k)\|$;
falls d^k nicht berechenbar ist oder $\nabla f(x^k)^T d^k \leq -\rho \|d^k\|^p$ nicht gilt,
setze $d^k = -B_k \nabla f(x^k)$

S3) Bestimme $t_k \in \{\beta^j : j = 0,1,2,\dots\}$ maximal mit

$$f(x^k + t_k d^k) < f(x^k) + t_k\, \sigma \nabla f(x^k)^T d^k$$

S4) Setze $x^{k+1} = x^k + t_k d^k,\ k := k+1$, gehe zu S1)

4. Quasi-Newton-Verfahren

(4.1) Sei $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ zweimal stetig differenzierbar. Dann gilt:
Für eine Folge $\{x^k\}$ mit $H_k(x^{k+1} - x^k) = -\nabla f(x^k)$ ist äquivalent:
a) $x^k \longrightarrow x^*$ konvergiert superlinear und $\nabla f(x^*) = 0$
b) $\|\big(H_k - D^2 f(x^k)\big)(x^{k+1} - x^k)\| = o(\|x^{k+1} - x^k\|)$

(4.2) Sei H symmetrisch positiv definit und für $s, y \in \mathbb{R}^n$ gelte $s^T y > 0$.
Dann gilt für:

$$
\begin{aligned}
H_+ &= H + \frac{1}{s^T y} y y^T - \frac{1}{s^T H s}(Hs)(Hs)^T \\
B_+ &= B + \frac{1}{s^T y}(r s^T + s r^T) - \frac{r^T y}{(s^T y)^2} s s^T \qquad r = s - By,\ B = H^{-1}
\end{aligned}
$$

(BFGS = Broyden-Fletcher-Goldfarb-Shanno-Formel)

a) $H_+ s = y,\ B_+ y = s.$
b) $H_+ B_+ = I$
c) H_+, B_+ symmetrisch positiv definit.
d) $\det H_+ = \frac{y^T s}{s^T H s} \det H.$

4. Quasi-Newton-Verfahren

(4.3) Globalisiertes BFGS-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n,n}$ symmetrisch positiv definit, $\sigma \in (0, \frac{1}{2})$, $\rho \in (\sigma, 1)$, $\varepsilon \geq 0$.
Setze $k := 0$.

S1) falls $\|\nabla f(x^k)\| \leq \varepsilon$ STOP

S2) $d^k = -B_k \nabla f(x^k)$

S3) bestimme $t_k > 0$ mit $f(x^k + t_k d^k) \leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k$ und $\nabla f(x^k + t_k d^k)^T d^k \geq \rho \nabla f(x^k)^T d^k$

S4) setze $x^{k+1} = x^k + t_k d^k$, $s^k = x^{k+1} - x^k$, $y^k = \nabla f(x^{k+1}) - \nabla f(x^k)$, $r^k = s^k - B_k y^k$ und

$$B_{k+1} = B_k + \frac{1}{(s^k)^T y^k} \left(r^k (s^k)^T + s^k (r^k)^T \right) - \frac{(r^k)^T y^k}{((r^k)^T y^k)^2} s^k (s^k)^T$$

S5) setze $k := k+1$ und gehe zu S1)

(4.4) Sei $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}$ stetig differenzierbar und nach unten beschränkt. Dann gilt:
a) $(s^k)^T y^k > 0$ für alle k
b) B_k symmetrisch positiv definit für alle k
c) Das Verfahren ist wohldefiniert.

4. Quasi-Newton-Verfahren

(4.5) Lokale Konvergenz des BFGS-Verfahrens
Sei $f \in C^2(\mathbb{R}^n, \mathbb{R})$, $D^2 f$ lokal Lipschitz-stetig und sei x^* striktes Minimum. Dann existiert $\delta > 0$ und $\varepsilon > 0$, sodass für $x^0 \in \mathbb{R}^n$ und B_0 symmetrisch positiv definit mit $\|x^0 - x^*\| \le \delta$, $\|B_0 - D^2 f(x^*)\| \le \varepsilon$ das ungedämpfte BFGS-Verfahren superlinear konvergiert.

(4.6) Globale Konvergenz des BFGS-Verfahrens
Sei $f \in C^2(\mathbb{R}^n, \mathbb{R})$, die Levelmenge $\mathscr{L}(x^0)$ sei konvex und f sei auf $\mathscr{L}(x^0)$ gleichmäßig konvex. Dann konvergiert Algorithmus (4.3) für beliebige $x^0 \in \mathbb{R}^n$ und symmetrisch positiv definite B_0 gegen das Minimum x^*.

5. Nichtlineare cg-Verfahren

(5.1) Fletcher-Reeves-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^n$, $\varepsilon \geq 0$, $0 < \sigma < \rho < \frac{1}{2}$, $d^0 = -\nabla f(x^0)$. Setze $k = 0$.

S1) falls $\|\nabla f(x^k)\| \leq \varepsilon$ STOP

S2) bestimme $t_k > 0$ mit

$$f(x^k + t_k d^k) \leq f(x^k) + \sigma t_k \nabla f(x^k)^T d^k$$

$$|\nabla f(x^k + t_k d^k)^T d^k| \leq -\rho\, \nabla f(x^k)^T d^k$$

S3) setze

$$\begin{aligned} x^{k+1} &= x^k + t_k d^k \\ \beta_k &= \frac{\|\nabla f(x^{k+1})\|^2}{\|\nabla f(x^k)\|^2} \\ d^{k+1} &= -\nabla f(x^{k+1}) + \beta_k d^k \end{aligned}$$

S4) setze $k := k+1$, gehe zu S1)

Variante von Polak-Ribière: $\beta_k = \dfrac{\left(\nabla f(x^{k+1}) - \nabla f(x^k)\right)^T \nabla f(x^{k+1})}{\|\nabla f(x^k)\|^2}$

5. Nichtlineare cg-Verfahren

(5.2) Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar, und sei f nach unten beschränkt. Dann ist Algorithmus (5.1) wohldefiniert, und es gilt:

$$-\frac{1}{1-\rho} \le \frac{\nabla f(x^k)^T d^k}{\|\nabla f(x^k)\|^2} \le -2 + \frac{1}{1-\rho}$$

(5.3) Sei zusätzlich ∇f Lipschitz-stetig auf $\mathscr{L}(x^0)$.
Dann gilt: $\liminf\limits_{k\to\infty} \|\nabla f(x^k)\| = 0$.

(5.4) Sei f zweimal stetig differenzierbar, $\mathscr{L}(x^0)$ konvex und f gleichmäßig konvex auf $\mathscr{L}(x^0)$.
Dann konvergiert Algorithmus (5.1) gegen das eindeutige Minimum x^*.

6. Trust-Region-Verfahren

(6.1) $d^* \in \mathbb{R}^n$ ist genau dann eine Lösung von

$$(P) \qquad \text{minimiere } q(d) = g^T d + \frac{1}{2} d^T H d \text{ unter der Bedingung } \|d\| \leq \Delta,$$

wenn $\lambda^* \in \mathbb{R}$ existiert mit

a) $(H + 2\lambda^* I)d^* = -g$

b) $\lambda^* \geq 0,\ \|d^*\| \leq \Delta,\ \lambda^*(\|d^*\| - \Delta) = 0$

c) $H + 2\lambda^* I$ positiv semidefinit.

Zu einem Minimum d^* ist der Lagrangeparameter $\lambda^* = -\dfrac{(g + Hd^*)^T d^*}{2\Delta^2}$ eindeutig.

6. Trust-Region-Verfahren

(6.2) Trust-Region-Newton-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^n$ $\Delta_0 > 0$, $\Delta_{\min} > 0$, $0 < \rho_1 < \rho_2 < 1$, $0 < \sigma_1 < 1 < \sigma_2$, $\varepsilon \geq 0$. Setze $k = 0$.

S1) falls $\|\nabla f(x^k)\| \leq \varepsilon$ STOP

S2) Bestimme ein Minimum d^k von

$$q_k(d) = f(x^k) + \nabla f(x^k)^T d + \frac{1}{2} d^T D^2 f(x^k) d \quad \text{unter} \quad \|d\| \leq \Delta_k$$

S3) berechne $r_k = \dfrac{f(x^k) - f(x^k + d^k)}{q_k(0) - q_k(d^k)}$ und setze

$$x^{k+1} = \begin{cases} x^k + d^k & \text{falls } r_k \geq \rho_1 \\ x^k & \text{sonst} \end{cases}$$

(Sprechweise: Im Fall $r_k \geq \rho_1$ heißt der k-te Iterationsschritt *erfolgreich*.)

S4)

$$\Delta_{k+1} = \begin{cases} \sigma_1 \, \Delta_k & \text{falls } r_k < \rho_1 \\ \max\{\Delta_{\min}, \Delta_k\} & \text{falls } r_k \in [\rho_1, \rho_2] \\ \max\{\Delta_{\min}, \sigma_2 \Delta_k\} & \text{falls } r_k > \rho_2 \end{cases}$$

S5) $k := k + 1$, gehe zu S1).

6. Trust-Region-Verfahren

(6.3) Für $d^k \neq 0$ gilt $q_k(0) - q_k(d^k) \geq \frac{1}{2}\|\nabla f(x^k)\| \min\left\{\Delta_k, \frac{\|\nabla f(x^k)\|}{\|H_k\|}\right\}$.

(6.4) Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ zweimal stetig differenzierbar und sei $\{x^k\}$ die Folge aus (6.2).
Sei $\{x^k\}_K$ eine gegen x^* konvergente Teilfolge.
Wenn x^* kein stationärer Punkt ist, dann gilt $\liminf\limits_{k \in k} \Delta_k > 0$.

(6.5) Es gibt unendlich viele erfolgreiche Iterationsschritte.

(6.6) Jeder Häufungspunkt von $\{x^k\}$ ist stationärer Punkt von f.

(6.7) Sei x^* Häufungspunkt von $\{x^k\}$, und sei $D^2 f(x^*)$ positiv definit. Dann gilt:

a) Die gesamte Folge $\{x^k\}$ konvergiert gegen x^*.
b) Ex existiert $k_0 > 0$, sodass alle Schritte $k \geq k_0$ erfolgreich sind.
c) Ex existiert $\bar{\Delta} > 0$ mit $\Delta_k \geq \bar{\Delta}$ für alle $k \in \mathbb{N}$.
d) Die Konvergenzrate ist mindestens superlinear.
e) Falls $D^2 f$ lokal Lipschitz-stetig ist, ist die Konvergenz quadratisch.

7. Innere-Punkte-Verfahren

(7.1) Seien $x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$. Dann gilt:
Wenn $(x,y,z) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ das KKT-System

$$(L) \qquad \begin{array}{rcl} 0 & = & A^T y + z - c \\ 0 & = & Ax - b \\ 0 & = & x^T z, \qquad x \geq 0, \qquad z \geq 0 \qquad (\text{also } x_j z_j = 0 \text{ für } j = 1, \dots, n) \end{array}$$

löst, dann löst $x \in \mathbb{R}^n$

$$(P) \qquad \text{minimiere } c^T x \text{ unter } x \geq 0,\ Ax = b$$

und $y \in \mathbb{R}^m$ löst

$$(D) \qquad \text{maximiere } b^T y \text{ unter } A^T y \leq c.$$

(7.2) Sei $\operatorname{rang} A = m$ und $\begin{pmatrix} x \\ z \end{pmatrix} > 0$. Betrachte $F \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} A^T y + z - c \\ Ax - b \\ \operatorname{diag}(x_j z_j) \end{pmatrix}$.

Dann ist $\mathrm{DF} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ \operatorname{diag}(z) & 0 & \operatorname{diag}(x) \end{pmatrix}$ regulär.

7. Innere-Punkte-Verfahren

(7.3) Sei $\tau > 0$. Wenn $(x_\tau, y_\tau, z_\tau) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n$ das System

$$(L_\tau) \qquad \begin{array}{lll} A^T y + z = c & & \\ Ax = b & & \\ x_j z_j = \tau \qquad (j = 1, \dots, n) & \qquad x > 0, & \quad z > 0 \end{array}$$

löst, dann löst $x_\tau \in \mathbb{R}^n$ das primale Barriere-Problem

$$(P_\tau) \quad \text{minimiere } c^T x - \tau \sum_{j=1}^n \log x_j \text{ unter } Ax = b, x > 0$$

und $(y_\tau, z_\tau) \in \mathbb{R}^m \times \mathbb{R}^n$ löst das duale Barriere Problem

$$(D_\tau) \quad \text{maximiere } b^T y + \tau \sum_{i=1}^n \log z_j \text{ unter } A^T y + z = c,\ z > 0.$$

(7.4) $$\mathscr{F} = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^n \colon Ax = b, A^T y + z = c,\ x \geq 0,\ z \geq 0 \right\}$$

(primal-dual) zulässige Menge

$$\mathscr{F}^0 = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathscr{F} \colon x > 0, z > 0 \right\}$$ *(primal-dual) strikt zulässige Menge*

7. Innere-Punkte-Verfahren

(7.5) Sei $\mathscr{F}^0 \neq \emptyset, \tau > 0$. Dann gilt:
a) (P_τ) ist lösbar

b) Es existiert eine Lösung $\begin{pmatrix} x_\tau \\ y_\tau \\ z_\tau \end{pmatrix}$ von (L_τ).

Dabei ist x_z, z_τ eindeutig, und y_z eindeutig falls $\operatorname{rang} A = m$.

(7.6) Die Abbildung $\tau \longmapsto \begin{pmatrix} x_\tau \\ y_\tau \\ z_\tau \end{pmatrix}$ heißt *zentraler Pfad*.

(7.7) Es gilt: $\tau = \frac{1}{n} x_\tau^T z_\tau = \frac{1}{n}(c^T x_\tau - b^T y_\tau)$.
$\frac{1}{n} x_\tau^T z_\tau$ heißt *gewichtete Dualitätslücke*.

7. Innere-Punkte-Verfahren

(7.8) Zulässiges Innere-Punkte-Verfahren

S0) Wähle $\gamma \in (0,1), \sigma \in (0,1), \varepsilon \in (0,1)$

$$\begin{pmatrix} x^0 \\ y^0 \\ z^0 \end{pmatrix} \in \mathscr{N}(\gamma) := \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathscr{F}^0 : x_j z_j \geq \gamma \frac{1}{n} x^T z \text{ für } j = 1, \dots, n \right\}$$

Setze $k = 0$.

S1) falls $\mu_k := \frac{1}{n}(x^k)^T z^k \leq \varepsilon$ STOP

S2) bestimme $\begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta z^k \end{pmatrix}$ als Lösung von

$$\begin{pmatrix} 0 & A^T & I \\ A & 0 & 0 \\ Z^k & 0 & X^k \end{pmatrix} \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta z^k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -X^k Z^k e + \sigma \mu_k e \end{pmatrix}$$

mit $X^k = \mathrm{diag}(x^k),\ Z^k = \mathrm{diag}(z^k)$.

S3) wähle t_k als größte Schrittweite $t \in [0,1]$ mit $\begin{pmatrix} x^k \\ y^k \\ z^k \end{pmatrix} + t \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta z^k \end{pmatrix} \in \mathscr{N}(\gamma)$.

S4) setze $\begin{pmatrix} x^{k+1} \\ y^{k+1} \\ z^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \\ z^k \end{pmatrix} + t_k \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta z^k \end{pmatrix}$, setze $k := k+1$ und gehe zu S1).

7. Innere-Punkte-Verfahren

(7.9) Es gilt $(\Delta x^k)^T \Delta z^k = 0$ und $\mu_k(t) = \big(1 - t(1-\sigma)\big)\mu_k$.

(7.10) Wenn $\delta > 0, \omega > 0$ existiert, sodass für die Schrittweite $t_k \in [0,1]$

$$t_k \geq \frac{\delta}{n^\omega}$$

erfüllt ist, dann gilt für alle $\varepsilon \in (0, \min\{1, \mu_0\})$

$$\mu_k \leq \varepsilon \quad \text{für} \quad k \geq \log\left(\frac{\mu_0}{\varepsilon}\right)\frac{n^\omega}{\delta(1-\sigma)}.$$

(7.11) Es gilt $\begin{pmatrix} x^k(t) \\ y^k(t) \\ z^k(t) \end{pmatrix} = \begin{pmatrix} x^k \\ y^k \\ z^k \end{pmatrix} + t \begin{pmatrix} \Delta x^k \\ \Delta y^k \\ \Delta z^k \end{pmatrix} \in \mathcal{N}(\gamma)$ für alle $t \in [0, \bar{t}_k]$ mit

$$\bar{t}_k = 2^{3/2}\gamma\frac{\sigma}{n}\,\frac{1-\gamma}{1+\gamma}.$$

(7.12) Für $u, v \in \mathbb{R}^n$ mit $u^T v \geq 0$ gilt

$$\|UVe\| \leq 2^{-3/2}\|u+v\|^2 \qquad \big(U = \operatorname{diag}(u),\ V = \operatorname{diag}(v)\big).$$

(7.13) Es gilt $\|\Delta X^k\, \Delta Z^k e\| \leq 2^{-3/2}\left(1 + \frac{1}{\gamma}\right) n\mu_k$.

7. Innere-Punkte-Verfahren

$S^n = \{A \in \mathbb{R}^{n\times n} \colon A^T = A\}$

$S^n_+ = \{A \in S^n \colon A \geq 0\}$ symmetrisch positiv semidefinit

$S^n_{++} = \{A \in S^n_+ \colon A > 0\}$ symmetrisch positiv definit

(7.14) Seien $C, A_1, \ldots, A_m \in S^n$, $b \in \mathbb{R}^n$. Dann heißt

(P) $C : X$ unter $X \geq 0$, $A_i : X = b_i$ $(i = 1, \ldots, m)$

semidefinites Programm.

(7.15) a) $g(X) = -\lambda_{\min}(X)$ ist konvex auf S^n

b) $S^n_+ \subset S^n$ ist konvex und abgeschlossen

c) für $X \in S^n$ gilt: $X \geq 0 \iff X : Y \geq 0 \quad \forall\, Y \in S^n_+$.

(7.16) Dualitätssatz für semidefinite Programme

a) Für alle zulässigen $X \in M$ und $b \in N$ gilt: $C : X \geq b^T y$.

b) Falls ein strikt zulässiges Tripel $(\hat X, \hat y, \hat Z) \in S^n_{++} \times \mathbb{R}^n \times S^n_{++}$ existiert mit $A_i : \hat X = b_i$ $(i = 1, \ldots, m)$ und $\sum_{i=1}^m \hat y_i A_i + \hat Z = C$, dann sind (P) und (D) lösbar mit $\min(P) = \max(D)$.

7. Innere-Punkte-Verfahren

Betrachte zu $\tau > 0$ die Barriereprobleme

(P_τ) minimiere $C : X + \tau \log\det X$ unter $A_i : X = b_i \quad (i = 1,\dots,m)$

(D_τ) maximiere $b^T y + \tau \log\det Z$ unter $\sum_{i=1}^m y_i A_i + Z = C$.

(7.17) Sei (X_τ, y_τ, Z_τ) Lösung des KKT-Systems

$$\begin{aligned}
\sum_{i=1}^m y_i A_i + Z_\tau &= C \\
A_i : X_\tau - b_i &= 0 \\
X_\tau Z_\tau + Z_\tau X_\tau &= 2\tau I
\end{aligned}$$

dann ist X_τ Lösung von (P_τ) und y_τ Lösung von (D_τ).

7. Innere-Punkte-Verfahren

(7.18) Zulässiges Innere-Punkte-Verfahren für semidefinite Programme

S0) Wähle X^0, y^0, Z^0 mit $X^0, Z^0 > 0,\ \Sigma y_i A_i + Z^0 = C$
$\tau_0 > 0, \varepsilon \geq 0$ $\qquad A_i : X^0 = b_i$

S1) Falls $\|XZ\|_F \leq \varepsilon$ $\qquad$ STOP

S2) Bestimme $\Delta X^k, \Delta y^k, \Delta Z^k$ mit

$$\begin{aligned}
\sum_{i=1}^{m} \Delta y_i^k A_i + \Delta Z^k &= 0 \\
A_i : \Delta X &= 0 \qquad (i = 1, \dots, m) \\
Z^k \Delta X^k + X^k \Delta Z^k + \Delta X^k Z^k + \Delta Z^k X^k &= 2\tau_k I - X^k Z^k - Z^k X^k
\end{aligned}$$

S3) Bestimme $t_k > 0$ mit $X^k + t_k \Delta X^k > 0$ und $Z^k + t_k \Delta Z^k > 0$

S4) Setze $\begin{pmatrix} X^{k+1} \\ y^{k+1} \\ Z^{k+1} \end{pmatrix} = \begin{pmatrix} X^k \\ y^k \\ Z^k \end{pmatrix} + t_k \begin{pmatrix} \Delta X^k \\ \Delta y^k \\ \Delta Z^k \end{pmatrix}$
wähle $\tau_{k+1} \in (0, \tau_k)$; setze $k := k+1$, gehe zu S1).

7. Innere-Punkte-Verfahren

(7.19) Seien $A_1,\ldots,A_m$ linear unabhängig in S^n, $X^k, Z^k \in S^n_{++}$.
Dann gilt für die Lösung von S2) aus (7.18):

a) Die Matrix $M_k = \big(\mathrm{Spur}(A_i (Z^k)^{-1} A_j X^k)\big)_{ij} \in \mathbb{R}^{m,m}$ ist symmetrisch positiv definit.

b) $M_k \Delta y^k = c^k$ mit $c^k = \big(\mathrm{Spur}(A_i X^k - \tau A_i (Z^k)^{-1})\big)_{i=1\ldots n}$.

c) $Z^k \Delta X^k + \Delta X^k Z^k = 2\tau I - X^k Z^k - Z^k X^k + \sum_{i=1}^m \Delta y_i^k (A_i X^k + X^k A_i)$.

(7.20) Sei $Z \in S^n_{++}$ und $B \in S^n$. Dann besitzt die Lyapunov-Gleichung

$$Z\,\Delta X + \Delta X\, Z = B$$

genau eine Lösung $\Delta X \in S^n$.

8. Penalty-Verfahren

Zu $f\colon \mathbb{R}^n \to \mathbb{R}$ und $h\colon \mathbb{R}^n \to \mathbb{R}^m$ betrachte

$$(P) \quad \text{Minimiere} \quad f(x) \text{ auf } M = \{x \in \mathbb{R}^n\colon\ h(x) = 0\}.$$

(8.1) S0) Wähle $\varepsilon \geq 0$, $\alpha_0 > 0$, setze $k = 0$.

S1) Bestimme eine Lösung $x^k \in \mathbb{R}^n$ der unrestringierten Aufgabe

$$(P_k) \quad \text{Minimiere} \quad f_k(x) := f(x) + \frac{\alpha_k}{2}\|h(x)\|^2$$

S2) falls $\|h(x^k)\| \leq \varepsilon$ STOP

S3) Wähle $\alpha_{k+1} > \alpha_k$, setze $k := k+1$, gehe zu S1).

(8.2) Seien f, h stetig, α_k streng monoton steigend mit $\lim\limits_{k\to\infty} \alpha_k = \infty$, sei (P) zulässig (d.h. $M \neq \emptyset$) und (8.1) wohldefiniert (d.h. in S1) existiert eine Lösung x^k). Dann gilt:

a) $\{f_k(x^k)\}$ ist monoton wachsend.

b) $\{\|h(x^k)\|\}$ ist monoton fallend.

c) $\{f(x^k)\}$ ist monoton wachsend.

d) $\lim\limits_{k\to\infty} h(x^k) = 0$.

e) Jeder Häufungspunkt von $\{x^k\}$ löst (P).

(8.3) Die Folge $\{x^k\}$ aus (8.1) konvergiert gegen x^* und $Dh(x^*)$ habe maximalen Rang. Dann gilt:
$y^k := \alpha_k h(x^k)$ konvergiert gegen ein y^*, und (x^*, y^*) ist KKT-Punkt von (P).

8. Exakte Penalty-Verfahren

Nun betrachte zu $f\colon \mathbb{R}^n \to \mathbb{R}$, $y\colon \mathbb{R}^n \to \mathbb{R}^p$, $h\colon \mathbb{R}^n \to \mathbb{R}^m$

$$(P) \quad \text{Minimiere } f(x) \text{ auf } M = \{x \in \mathbb{R}^n\colon y(x) \le 0 \text{ und } h(x) = 0\}.$$

(8.4)

a) Eine stetige Funktion $r\colon \mathbb{R}^n \to \mathbb{R}$ heißt *Penalty-Funktion* zu M, wenn $r(x) \ge 0$ für $x \in \mathbb{R}^n$ und

$$r(x) = 0 \iff x \in M.$$

b) Sei heißt *exakt* in einem Minimum $x^* \in M$ von (P), wenn $\bar\alpha > 0$ existiert, sodass für alle $\alpha \ge \bar\alpha$ gilt: x^* ist Minimum von

$$(P_\alpha) \quad \text{Minimiere } f_{r,\alpha}(x) := f(x) + \alpha r(x).$$

(8.5) Sei x^* ein lokales Minimum von (P) mit $\nabla f(x^*) \ne 0$, und sei r exakt in x^*. Dann ist r nicht differenzierbar in x^*.

(8.6) Seien f, g stetig differenzierbar und konvex, $h(x) = Ax - b$, und sei (x^*, y^*, z^*) ein KKT-Punkt von (P). Dann ist die ℓ_1-Penalty-Funktion

$$r_1(x) = \sum_{k=0}^{p} \max\{0, g_k(x)\} + \sum_{i=0}^{m} |h_i(x)|$$

exakt in x^*.

8. Multiplier-Penalty-Verfahren

(8.7) Sei (x^*, y^*) ein KKT-Punkt zur Minimierung von $f(x)$ unter $h(x) = 0$ mit

$$d^T D_x^2 L(x^*, y^*) d > 0 \text{ für } d \neq 0 \text{ mit } Dh(x^*) d = 0\,.$$

Dann existiert $\bar{\alpha} > 0$, sodass für $\alpha \geq \bar{\alpha}$ gilt:
x^* ist striktes lokales Minimum der *erweiterten Lagrange-Funktion*

$$L_\alpha(x, y^*) = f(x) + h(x)^T y^* + \frac{\alpha}{2} \|h(x)\|^2.$$

(8.8) Multiplier-Penalty-Verfahren

S0) Wähle $x^0 \in \mathbb{R}^n$, $y^0 \in \mathbb{R}^m$, $\alpha_0 > 0$, $\varepsilon \geq 0$, $c \in (0,1)$, $\rho > 1$. Setze $k = 0$.
S1) Falls $\|\nabla_x L(x^k, y^k)\| \leq \varepsilon$, $\|h(x^k)\| \leq \varepsilon$ STOP
S2) Bestimme ein Minimum x^{k+1} von

$$(L_k) \quad \text{Minimiere} \quad L_{\alpha_k}(x, y^k), \quad x \in \mathbb{R}^n$$

und setze $y^{k+1} = y^k + \alpha_k h(x^{k+1})$.
S3) Falls $\|h(x^{k+1})\| > c\|h(x^k)\|$, setze $\alpha_{k+1} = \rho\alpha_k$ sonst setze $\alpha_{k+1} = \alpha_k$.
S4) Setze $k := k+1$ und gehe zu S1).

Erweiterte Lagrange-Funktion für Ungleichungsrestriktionen:

$$L_\alpha(x, y, z) = f(x) + y^T h(x) + \frac{\alpha}{2}\|h(x)\|^2 + \frac{1}{2\alpha}\sum_{k=1}^{p}\left(\max\{0, z_k + \alpha g_k(x)\}^2 - z_k^2\right)$$

9. Aktive-Mengen Strategien

Sei $H \in \mathbb{R}^{n,n}$ symmetrisch positiv definit, $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m,n}, b \in \mathbb{R}^m$.
Betrachte das Quadratische Programm

$$(P) \quad \text{Minimiere} \quad f(x) := \frac{1}{2} x^T H x + c^T x \text{ unter } Ax \le b.$$

(9.1) Primal-duales Aktive-Mengen-Verfahren

S0) Wähle $(x^0, y^0) \in \mathbb{R}^n \times \mathbb{R}^m, \alpha > 0$, $\varepsilon \ge 0$. Setze $k = 0$

S1) falls $\|Hx^k + c + A^T y^k\| \le \varepsilon$ und $\|\max\{0, Ax^k - b\}\| \le \varepsilon$ STOP

S2) setze

$$\begin{aligned} \mathscr{A}_k &= \{i \in \{1,\dots,m\} \colon y_i^k + \alpha(Ax^k - b)_i > 0\}, \\ \mathscr{I}_k &= \{i \in \{1,\dots,m\} \colon y_i^k + \alpha(Ax^k - b)_i \le 0\}. \end{aligned}$$

Sei (x^{k+1}, y^{k+1}) Lösung von

$$(P_k) \quad \begin{array}{ll} Hx^{k+1} + c + A^T y^{k+1} = 0 & \\ (Ax^{k+1})_i = b_i & i \in \mathscr{A}_k \\ y_i^{k+1} = 0 & i \in \mathscr{I}_k \end{array}$$

S3) setze $k := k+1$ und gehe zu S1).

9. Aktive-Mengen Strategien

(9.2) $H = (h_{ij}) \in \mathbb{R}^{n,n}$ heißt *M-Matrix*, wenn gilt:

a) $h_{ii} > 0$

b) $h_{ij} \leq 0$ für $i \neq j$

c) $d \geq 0 \implies H^{-1} d \geq 0$

Dann gilt auch: $d_{\mathscr{A}} \geq 0 \implies H^{-1}_{\mathscr{A},\mathscr{A}} d_{\mathscr{A}} \geq 0$.

(9.3) Sei H eine M-Matrix und $A = I$. Dann gilt:

a) $x^* \leq x^{k+1} \leq x^k$ für $k \geq 1$

b) $x^k \leq b$ für $k \geq 2$

c) es existiert $k_0 > 0$ mit $y^k \geq 0$ für $k \geq k_0$.

d) $\lim_{k\to\infty} x^k = x^*$

(9.4) Eine Funktion $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$ heißt *Komplementaritätsfunktion* zu (P), wenn gilt:

$$\varphi(x,y) = 0 \iff Ax \leq b,\ y \geq 0,\ (Ax-b)^T y = 0$$

(9.5) Zu $\alpha > 0$ ist $\varphi(x,y) = y - \max\{0, y + \alpha(Ax-b)\}$ Komplementaritätsfunktion zu (P).

9. Verallgemeinertes Newton-Verfahren

(9.7) Eine Abbildung $F\colon \mathbb{R}^n \to \mathbb{R}^m$ heißt *Newton-differenzierbar*, wenn eine Abbildung $G\colon \mathbb{R} \to \mathbb{R}^{m,n}$ *(Newton-Ableitung)* existiert, sodass gilt:

$$\lim_{d\to 0} \frac{1}{\|d\|} \|f(x+d) - F(x) - G(x+d)d\| = 0.$$

(9.7) Sei $F\colon \mathbb{R}^n \to \mathbb{R}^m$ Newton-differenzierbar mit Newton-Ableitung $G\colon \mathbb{R} \to \mathbb{R}^{m,n}$, und sei $h\colon \mathbb{R}^k \to \mathbb{R}^n$ stetig differenzierbar. Dann ist $F \circ h$ Newton-differenzierbar mit Newton-Ableitung $y \mapsto F(h(y))Dh(y)$.

(9.8) Sei x^* Nullstelle von F (d.h. $F(x^*) = 0$), und sei F Newton-differenzierbar mit Newton-Ableitung G.
Zu einem $\varepsilon > 0$ sei $G(x^* + d)$ invertierbar für alle d mit $\|d\| \le \varepsilon$ und es existiere $C > 0$ mit $\|G(x^* + d)^{-1}\| \le C$ für $\|d\| \le \varepsilon$.
Dann existiert $\delta \in (0, \varepsilon)$, sodass für $x^0 \in \mathbb{R}^n$ mit $\|x^0 - x^*\| \le \delta$ die Newton-Iteration

$$x^{k+1} = x^k - G(x^k)^{-1} F(x^k)$$

wohldefiniert ist und superlinear gegen x^* konvergiert.

10. SQP-Verfahren

Seien $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^p$ und $h\colon \mathbb{R}^n \to \mathbb{R}^m$ stetig differenzierbar.
Betrachte

$$(P) \qquad \text{Minimiere} \quad f(x) \text{ unter } g(x) \le 0 \text{ und } h(x) = 0.$$

Dann ist $(x^*, y^*, z^*) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ KKT-Punkt zu (P), falls

$$F\begin{pmatrix} x^* \\ y^* \\ z^* \end{pmatrix} = 0 \text{ für } F\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \nabla_x L(x,y,z) \\ h(x) \\ \varphi(x,z) \end{pmatrix} = \begin{pmatrix} \nabla f(x) + Dh(x)^T y + Dg(x)^T z \\ h(x) \\ (\min\{-g_i(x), z_i\}) \end{pmatrix}.$$

(10.1) S0) Wähle $(x^0, y^0, z^0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p$ mit $z^0 \ge 0$, $\quad H_0 \in \mathbb{R}^{n,n}$ symmetrisch, setze $k := 0$.

S1) Falls (x^k, y^k, z^k) KKT-Punkt von (P) $\qquad$ STOP

S2) Berechne KKT-Punkt $(x^{k+1}, y^{k+1}, z^{k+1})$ von

$$\begin{aligned} (Q_k) \qquad \text{Minimiere} \quad & \nabla f(x^k)(x - x^k) + \tfrac{1}{2}(x - x^k)^T H_k (x - x^k) \\ \text{unter } & g(x^k) + Dg(x^k)(x - x^k) \le 0 \\ \text{und } & h(x^k) + Dh(x^k)(x - x^k) = 0 \end{aligned}$$

S3) Wähle $H_{k+1} \in \mathbb{R}^{n,n}$ symmetrisch, setze $k := k+1$ und gehe zu S1).

10. SQP-Verfahren

(10.2) In Algorithmus (10.1) wähle $H_k = D_x^2 L(x^k, y^k, z^k)$, und falls (Q_k) mehrere KKT-Punkte besitzt, wähle den KKT-Punkt $(x^{k+1}, y^{k+1}, z^{k+1})$ mit

$$\|(x^{k+1}, y^{k+1}, z^{k+1}) - (x^k, y^k, z^k)\| \qquad (*)$$

minimal.

Sei (x^*, y^*, z^*) Lösung von (P) mit

1) strikte Komplementarität: $g_i(x^*) + z_i^* \neq 0 \qquad (i = 1, \ldots, p)$

2) $(CQ2)$: $\{\nabla h_j(x^*),\ j = 1, \ldots, m\} \cup \{\nabla g_i(x^*) \colon i \in I(x^*)\}$ linear unabhängig mit $I(x^*) := \{i \colon g_i(x^*) = 0\}$

3) hinreichende Bedingung 2. Ordnung:

$$d^T D_x^2 L(x^*, y^*, z^*) d > 0 \text{ für } d \neq 0 \text{ mit}$$
$$Dh(x^*) d = 0,\ \nabla g(x^*)^T d = 0 \quad \big(i \in I(x^*)\big)$$

Dann existiert $\varepsilon > 0$, sodass für alle (x^0, y^0, z^0) mit $\|(x^0, y^0, z^0) - (x^*, y^*, z^*)\| < \varepsilon$ gilt:

a) (Q_k) ist wohldefiniert und $\{(x^k, y^k, z^k)\}$ konvergiert gegen (x^*, y^*, z^*).

b) Die Konvergenz ist superlinear.

c) Sind $D^2 f,\ D^2 h_j,\ D^2 g_i$ lokal Lipschitz-stetig, so ist die Konvergenz quadratisch.

10. SQP-Verfahren

Definiere $\mathscr{F}(\tau) = \{x \in \mathbb{R}^n \colon |h_j(x)| \le \tau \ (j = 1,\dots,m),\ g_i(x) \le \tau \ (i = 1,\dots,p)\}$.

(10.3) S0) Wähle (x^0, y^0, z^0) und $\tau > 0$ mit $x^0 \in F(\tau), z^0 \ge 0$, $\delta > 0$, $\rho \in (0,1)$, $\sigma \in (0,1)$, H_0 symmetrisch, $\beta_j^0 = \delta$ $(j = 1,\dots,m)$, $\alpha_i^0 = \delta$ $(i = 1,\dots,p)$; $k = 0$.

S1) Wenn (x^k, y^k, z^k) KKT-Punkt von (P) STOP

S2) Bestimme KKT-Punkt (d^k, y^{k+1}, z^{k+1}) von

$$(Q_k) \qquad \text{minimiere} \quad \nabla f(x^k)^T d + \tfrac{1}{2} d^T H_k d \qquad \text{unter} \quad h(x^k) + Dh(x^k) d = 0, \quad g(x^k) + Dg(x^k) d \le 0$$

S3) Anpassen der Penalty-Parameter: setze

$$\alpha_i^{k+1} = \begin{cases} z_i^{k+1} + 2\delta, & \alpha_i^k \le z_i^{k+1} + \delta, \\ \alpha_i^k, & \text{sonst}, \end{cases} \qquad \beta_j^{k+1} = \begin{cases} |y_j^{k+1}| + 2\delta, & \beta_j^k \le |y_j^{k+1}| + \delta, \\ \beta_j^k, & \text{sonst}, \end{cases}$$

und definiere $f_{k+1}(x) = f(x) + \sum_{i=1}^p \alpha_i^{k+1} \max\{0, g_i(x)\} + \sum_{j=1}^m \beta_j^{k+1} |h_j(x)|$

S4) Wähle Schrittweite $t_k \in \{\rho^\ell \colon \ell = 0,1,2,\dots\}$ maximal mit

$$f_{k+1}(x^k + t_k d^k) \le f_{k+1}(x^k) - t_k \sigma \Big((d^k)^T H_k d^k + \delta \sum_{i=1}^p \max\{0, g_i(x^k)\} + \delta \|h(x^k)\|_\infty \Big)$$

und $x^k + t_k d^k \in \mathscr{F}(\tau)$. Setze $x^{k+1} = x^k + t_k d^k$, wähle H_{k+1}.
Setze $k := k+1$ und gehe zu S1).

10. SQP-Verfahren

(10.4) Es gelte:

a) f, g_i konvex, $h(x) = Ax - b$ linear affin.
b) Die Slaterbedingung sei erfüllt.
c) $\rho_1 \|d\|^2 \le d^T H_k d \le \rho_2 \|d\|^2$.
d) $\mathscr{F}(\tau_0)$ kompakt für ein $\tau_0 > 0$.

Dann ist (Q_k) lösbar, der Algorithmus (10.3) ist wohldefiniert, und die Iterierten $\{d^k\}, \{y^k\}, \{z^k\}$ sind beschränkt.

(10.5) Unter den Voraussetzungen von (10.4) ist jeder Häufungspunkt von $\{x^k\}$ ein KKT-Punkt (und damit Lösung) von (P).

Zu x^k, H_k, und Δ_k definiere $h_k(d) = h(x^k) + Dh(x^k)d$ und

$$(NP) \qquad \text{Minimiere } \|h_k(d^n)\| \text{ unter } \|d^n\| \le \Delta_k$$

und zu $q_k(d) = \nabla f(x^k)^T d + \frac{1}{2} d^T H_k d$

$$(TP) \qquad \text{Minimiere } q_k(d^n + d^t) \text{ unter } Dh(x^k)d^t = 0 \text{ und } \|d^t\| \le \Delta_k\,.$$

Setze $L_\alpha(x, y) = f(x) + h(x)^T y + \alpha \|h(x)\|^2$ und
$L_{k,\alpha}(d, \Delta y) = q_k(x) + h_k(d)^T \Delta y + \alpha \|h_k(d)\|^2$

10. SQP-Verfahren

(10.6) Trust-Region-SQP-Verfahren

S0) Wähle $\Delta_{\min} > 0,\ 0 < \underline{\rho} < \bar{\rho} < 1,\ 0 < \sigma_1 < 1 < \sigma_2,$
$\varepsilon > 0,\ x^0 \in \mathbb{R}^n,\ \Delta_0 \geq \Delta_{\min},\ \alpha_0 > 0,\ \beta > 1$; setze $k = 0$.

S1) Bestimme Z_k. Falls $\|Z_k^T \nabla f(x^k)\| + \|h(x^k)\| \leq \varepsilon$ STOP

S2) Wähle symmetrische Approximation H_k von $D_x^2 L(x^k, y^k)$, löse (NP) und (TP), setze $d = d^n + d^t$, berechne Δy.

S3) Solange $L_{k,\alpha_k}(0,0) - L_{k,\alpha_k}(d,\Delta y) \geq \frac{1}{2}\big(q_k^L(d^n) - q_k^L(d)\big)$
$+\frac{\alpha_k}{2}\big(\|h(x^k)\|^2 - \|h(x^k) - Dh(x^k)d\|^2\big)$. Setze $\alpha_k := \beta \cdot \alpha_k$.

S4) Berechne $\rho_k = \dfrac{L_{\alpha_k}(x^k, y^k) - L_{\alpha_k}(x^k + d, y^k + \Delta y)}{L_{k,\alpha_k}(0,0) - L_{k,\alpha_k}(d,\Delta y)}$

S5) Falls $\rho_k \geq \underline{\rho}$, setze $x^{k+1} = x^k + d,\ y^{k+1} = y^k + \Delta y$
sonst $x^{k+1} = x^k,\ y^{k+1} = y^k$

$$\alpha_{k+1} = \alpha_k \text{ und } \Delta_{k+1} = \begin{cases} \sigma_1 \Delta_k & \rho_k < \underline{\rho} \\ \max\{\Delta_m, \Delta_k\} & \rho_k \in [\underline{\rho}, \bar{\rho}] \\ \max\{\Delta_{\min}, \sigma_2 \Delta_k\} & \rho_k > \bar{\rho} \end{cases}$$

setze $k := k+1$ und gehe zu S1).

11. Nicht-glatte Optimierung

(11.1) Sei $M \subset \mathbb{R}^n$ konvex und sei $f\colon M \to \mathbb{R}$ konvex.
Damit ist f im Inneren von M lokal Lipschitz-stetig.

(11.2) Sei $M \subset \mathbb{R}^n$ offen und konvex, $f\colon M \to \mathbb{R}$ konvex, $d \in \mathbb{R}^n, x \in M$. Dann gilt:

a) $q(t) = \frac{1}{2}\big(f(x+td) - f(x)\big)$ ist monoton fallend für $t \to 0+$.

b) Die Richtungsableitung von f in x nach d existiert, und es gilt

$$Df(x;d) = \inf_{t>0} \frac{1}{t}\big(f(x+td) - f(x)\big).$$

(11.3) Sei $M \subset \mathbb{R}^n$ offen, konvex, und $f\colon M \to \mathbb{R}^n$ konvex, $x \in M$. Dann heißt

$$\partial f(x) = \{s \in \mathbb{R}^n\colon f(y) \geq f(x) + s^T(y-x) \quad \forall y \in M\}$$

das (konvexe) *Subdifferenzial* und $s \in \partial f(x)$ heißt *Subgradient*.

(11.4) Sei $M \subset \mathbb{R}$ offen, konvex, $f\colon M \to \mathbb{R}^n$ konvex, $x \in M$. Dann gilt:

a) $\partial f(x)$ ist nicht-leer, konvex und kompakt

b) $\partial f(x) = \{s \in \mathbb{R}^n\colon s^T d \leq Df(x;d) \quad \forall d \in \mathbb{R}^n\}$

c) $Df(x;d) = \max\limits_{s \in \partial f(x)} s^T d$ für alle $d \in \mathbb{R}^n$.

(11.5) Sei $M \subset \mathbb{R}^n$ offen, konvex, $f\colon M \to \mathbb{R}$ konvex, $x^* \in M$. Dann ist äquivalent:

a) f nimmt sein Minimum in M an, d.h. $f(x) \geq f(x^*)$ für alle $x \in M$.

b) $0 \in \partial f(x^*)$

c) $Df(x^*;d) \geq 0$ für alle $d \in \mathbb{R}^n$.

11. Nicht-glatte Optimierung

(11.6) a) Eine Abb. $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *erweiterte Funktion*.

b) $\mathrm{dom}(f) = \{x \in \mathbb{R}^n \colon f(x) < \infty\}$ heißt der *wesentliche Definitionsbereich*
f heißt *echt* (proper), wenn $\mathrm{dom}(f) \neq \emptyset$.

c) Eine erweiterte Funktion heißt *konvex*, wenn für $x, y \in \mathbb{R}^n$, $\lambda \in (0,1)$ gilt:
$f(1-\lambda)x + \lambda y) \leq (1-\lambda)f(x) + \lambda f(y)$.

(11.7) $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ konvex $\implies$ $\mathrm{dom}(f) \subset \mathbb{R}^n$ konvex.

(11.8) Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ eine echte erweiterte Funktion.
Dann heißt f *nach unten halbstetig* in $x \in \mathbb{R}^n$, falls gilt: $\liminf\limits_{y \to x} f(y) \geq f(x)$.

f heißt nach unten halbstetig (l.s.c. = lower semi-continuous) in $X \subset \mathbb{R}^n$, wenn f l.s.c. in allen $x \in X$.

(11.9) Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ eine echte erweiterte Funktion. Dann ist äquivalent:

a) f l.s.c. in $\mathbb{R}^n$.

b) $\mathscr{L}(c) := \{x \in \mathbb{R}^n \colon f(x) \leq c\}$ ist abgeschlossen für alle $c \in \mathbb{R}$.

c) $\mathrm{epi}(f) = \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \colon \alpha \geq f(x)\}$ ist abgeschlossen in $\mathbb{R}^n \times \mathbb{R}$.

(11.10) Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ eine echte *erweiterte* konvexe Funktion. Dann heißt
$\partial f(x) = \{s \in \mathbb{R}^n \colon f(y) \geq f(x) + s^T(y - x)\ \forall\, y \in \mathbb{R}^n\}$ das *Subdifferenzial* von f.

11. Nicht-glatte Optimierung

(11.11) Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ echt und konvex. Dann gilt:
x^* Minimum von $f \iff 0 \in \partial f(x^*)$.

(11.12) Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ echt, konvex, $x^1, x^2 \in \mathbb{R}^n$, $s^1 \in \partial f(x^1)$, $s^2 \in \partial f(x^2)$.
Dann gilt: $(x^1 - x^2)^T(s^1 - s^2) \geq 0$ *(Monotonie)*.

(11.12) Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ echt, konvex. Betrachte

$$(P) \qquad \min f(x) \text{ auf } \mathbb{R}^n.$$

Dann gilt:
a) Die Lösungsmenge von (P) ist konvex.
b) Ist f strikt konvex auf $\mathrm{dom}(f)$, dann existiert höchstens eine Lösung.
c) Ist f l.s.c. und gleichmäßig konvex auf $\mathrm{dom}(f)$, dann existiert genau eine Lösung.

11. Nicht-glatte Optimierung: Regularisierungsverfahren

(11.13) Moreau-Yosida-Regularisierung

Sei $f\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ echt, l.s.c., konvex. Wähle $\gamma > 0$. Dann heißt

$f\gamma(x) := \min\limits_{y\in\mathbb{R}^n} \big(f(y) + \dfrac{1}{2\gamma}\|x-y\|^2\big)$ die *Moreau-Yosida-Regularisierung* von f.

Es gilt: Für festes x ist $g_{\gamma,x}(y) = f(y) + \frac{1}{2\gamma}\|x-y\|^2$ gleichmäßig konvex auf $\mathrm{dom}(f)$, d.h. es exisitert eine eindeutige Lösung $p_\gamma(x)$ mit

$g_{\gamma,x}\big(p_\gamma(x)\big) = \min\limits_{y\in\mathbb{R}} g_{\gamma,x}(y)$ und $f_\gamma(x) = f\big(p_\gamma(x)\big) + \dfrac{1}{2\gamma}\|p_\gamma(x) - x\|^2$.

(11.14) Proximal-Punkt-Verfahren

S0) Wähle $x^0 \in \mathrm{dom}(f), \gamma > 0$. Setze $k := 0$.
S1) Ist x^* Minimum von f STOP
S2) Berechne Minimierung x^{k+1} von $g_k(x) = f(x) + \frac{1}{2\gamma}\|x-x^k\|^2$
S3) Setze $k := k+1$, gehe zu S1).

(11.15) a) $s^k := \frac{1}{\gamma}(x^{k-1} - x^k) \in \partial f(x^k)$

b) $\|s^k\|$ monoton fallend

c) $f(x^k) - f(x) \le \frac{2}{k\gamma}(\|x-x^0\|^2 - \|x-x^k\|^2) - \frac{k\gamma}{2}\|s^k\|^2$ für allle $x \in \mathbb{R}^n$

d) falls die Lösungsmenge $S = \{x \in \mathbb{R}^n\colon f(x) = \inf\limits_{y\in\mathbb{R}^n} f(y)\} \neq \emptyset$,

dann konvergiert $\{x^k\}$ gegen ein $x^* \in S$.

11. Nicht-glatte Optimierung: Regularisierungsverfahren

(11.16) Tikhonov-Regularisierungs-Verfahren

S0) Wähle $x^0 \in \mathrm{dom}(f)$ setze $k := 0$.

S1) Falls x^k Minimum von f STOP

S2) Wähle $\varepsilon_k > 0$ und bestimme x^{k+1} als globales Minimum von

$$f_{\varepsilon_k}(x) := f(x) + \frac{\varepsilon_k}{2}\|x\|^2$$

S3) setze $k := k+1$ und gehe zu S1).

(11.17) Sei $\{\varepsilon_k\}$ monoton fallend und $\lim \varepsilon_k = 0$. Dann gilt:

a) Jeder Häufungspunkt von $\{x^k\}$ in Minimum von f.

b) $s^k := -\varepsilon_{k-1} x^k \in \partial f(x^k)$.

c) Falls die Lösungsmenge $S = \{x \in \mathbb{R}^n : f(x) = \inf_{y \in \mathbb{R}} f(y)\} \neq \emptyset$,

dann konvergiert $\{x^k\}$ gegen ein $x^* \in S$.

11. Nicht-glatte Optimierung: Subgradienten-Verfahren

Sei $f\colon \mathbb{R}^n \to \mathbb{R}$ konvex und $M \subset \mathbb{R}^n$ konvex und abgeschlossen.
Sei $P_M\colon \mathbb{R}^n \to M$ die Orthogonalprojektion auf M.

(11.18) Subgradienten-Verfahren

S0) Wähle $x^0 \in M, m_0 := f(x^0)$, setze $k := 0$.
S1) Falls x^k Minimum von f STOP
S2) Berechne $s^k \in \partial f(x^k)$; setze $d^k := -\frac{1}{\|s^k\|} s^k$,
wähle $t_k > 0$ und $x^{k+1} = P_M(x^k + t_k d^k)$
S3) $m_{k+1} := \min\{f(x^{k+1}), m_k\}$, setze $k := k+1$ und gehe zu S1).

(11.19) Es sei $f^* = \inf\limits_{y \in M} f(y) > -\infty$, und die Lösungsmenge $S = \{x \in \mathbb{R}^n\colon f(x) = f^*\}$ sei nicht leer. Die Folge $\{t_k\}$ sei monoton fallend mit

$$\lim_{k\to\infty} t_k = 0 \text{ und } \sum_{k=0}^{\infty} t_k = \infty.$$

Dann konvergiert die Folge $\{m_k\}$ gegen f^*.