Zykeln "entlang" von 6 Kanten des Würfels

mit den Ecken $(\pm 1, \pm 1, \pm 1)$.

Dabei ist $\nabla f \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -y-z+2(x-1)_+ -2(-x-1)_+ \\ -x-z+2(y-1)_+ -2(-y-1)_+ \\ -x-y+2(z-1)_+ -2(-z-1)_+ \end{pmatrix}$

und $\nabla f \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, dito $\nabla f(v) \neq 0$ für die anderen Ecken.

<u>Modifiziertes Verfahren (Lucidi Sciandrone 1997):</u>

Wähle $\gamma > 0$, $\delta \in (0,1)$ und eine Folge $(\beta_k)_{k \in \mathbb{N}}$

mit $\beta_k > 0$, $\beta_k \searrow 0$.

In jedem Schritt des allgemeinen line-search-Verfahrens werde $\alpha = \alpha_k$ so gewählt, dass

$$(2) \begin{cases} f(x^k + \alpha s^k) \leq f(x^k) - \gamma \alpha^2 \|s^k\|_2^2 \\ \min\{f(x^k + \hat{\alpha} s^k), f(x^k - \hat{\alpha} s^k)\} \geq f(x^k) - \gamma \hat{\alpha}^2 \|s^k\|_2^2 \end{cases}$$

wobei $\hat{\alpha} := \frac{\alpha}{\delta}$.

Dabei werde zuerst $\alpha = \pm 1$ in die erste Ungleichung von (2) eingesetzt. <u>Falls</u> diese erfüllt ist, aber die zweite nicht, so erfüllt auch $\alpha := \frac{\alpha}{\delta}$ oder $\alpha := -\frac{\alpha}{\delta}$ die erste Ungleichung. Der Schritt $\alpha \to \pm \frac{\alpha}{\delta}$ wird wiederholt.

Entweder gilt dann $|\hat{\alpha}| \longrightarrow \infty$ und $f(x^k + \hat{\alpha} s^k) \longrightarrow -\infty$ (d.h. es gibt keine endliche globale Minimalstelle) oder es wird eine Schrittweite erzeugt, die beide Bed. aus (2) erfüllt.

Falls $\alpha = \pm 1$ die erste Bedingung in (2) nicht erfüllt, so liefert $\delta\alpha$ eine Schrittweite, die die zweite Bed. erfüllt. Wiederhole den Übergang $\alpha \longrightarrow \pm\delta\alpha$ so lange bis entweder die erste Bed. erfüllt ist oder bis

(3a) $\|\alpha s^k\| < \frac{1}{2}\beta_k$ gilt. (β_k ist hier fest und $\alpha \searrow 0$.)

In letzterem Fall wähle $\alpha = 0$.
(Update $x^{k+1} = x^k + \alpha s^k$ wie auf S. 1 und Wechsel $k \to k+1$.)

<u>Beh:</u> Sei Voraussetzung 1 erfüllt und f stetig differenzierbar und x^* ein Häufungspunkt der Folge $\{x^k\}$. Dann gilt $\nabla f(x^*) = 0$.

<u>Bew:</u> (schwächere Voraussetzung als bei den Abstiegsverfahren in Einf. Opti.)
$f(x^k)$ ist monoton fallend, nach unten beschränkt durch $f(x^*)$. Wegen (2) Teil 1 gilt

$$f(x^k) - f(x^{k+1}) \geq \gamma \|x^k - x^{k+1}\|_2^2 . \qquad (3)$$

(Trivial, falls $\alpha = 0$ gewählt wurde)

Aufsummieren liefert, dass $\|x^k - x^{k+1}\|$ eine Nullfolge ist (sonst ergäbe $f(x^k) \longrightarrow -\infty \le f(x^*)$ einen Widerspruch).

Widerspruchsannahme: $\|\nabla f(x^*)\|_2 =: \eta > 0$.

Sei $\varepsilon > 0$ so, dass $\|Df(x) - Df(x^*)\| \; (*) \; \le \frac{c\eta}{2}$ $\forall x$ mit $\|x - x^*\| \le \varepsilon$, wobei c aus Voraussetzung 1 sei. Sei nun γ aus (2).

Da $\beta_k \to 0$ $\exists \tilde{k}_0 : \beta_k < \min\left\{\frac{\varepsilon}{2\ell}, \frac{c\eta}{2\gamma}\right\}$ für $k \ge \tilde{k}_0$ und ℓ auch aus Vorauss. 1. Da $\|x^k - x^{k+1}\|$ eine Nullfolge ist gibt es $k_0 \ge \tilde{k}_0$: $\|x^k - x^{k+1}\|_2 \le \frac{\varepsilon}{2\ell}$ $\forall k \ge k_0$.

Ferner gilt ∞ oft: $\|x^k - x^*\| \le \frac{\varepsilon}{2}$ (Häufungspunkt x^*).

Für jedes solche $k \ge k_0$ ist $\|x^{k+j} - x^*\| \le \varepsilon$ $(0 \le j \le \ell)$.

Zu $v := \nabla f(x^*)$ erfüllt mind. eine der Suchrichtungen s^{k+j} $(0 \le j \le \ell - 1)$ nach Voraussetzung 1:

$$\frac{|v^T s^{k+j}|}{\|s^{k+j}\|_2} > c\|v\|_2 = c\eta. \quad (**)$$

Setze $\Phi(\alpha) := f(x^{k+j} + \alpha s^{k+j})$ mit obigem j und $|\alpha| \|s^{k+j}\| \le \frac{\varepsilon}{2\ell}$.

$$\Rightarrow |\Phi'(\alpha)| = |Df(x^{k+j} + \alpha s^{k+j}) s^{k+j}|$$

$$\overset{\Delta\text{-Ungl.}}{\ge} |Df(x^*) s^{k+j}| - \underbrace{\|Df(x^{k+j} + \alpha s^{k+j}) - Df(x^*)\|_2}_{\le \frac{c}{2}\eta \text{ (wegen } *)} \|s^{k+j}\|_2$$

$$\overset{(**)}{\ge} \frac{1}{2} c\eta \|s^{k+j}\|_2.$$

Insbesondere hat Φ' keine Nullstelle, das Vorzeichen ist für $|\alpha| \le \varepsilon / (2\ell\|s^{k+j}\|)$ konstant.

Mit $\Phi(\alpha) = \Phi(0) + \int_0^\alpha \Phi'(t)\,dt$ folgt

(5) $\quad f(x^{k+j} + \alpha s^{k+j}) \le f(x^{k+j}) - |\alpha| \frac{c\eta}{2} \|s^{k+j}\|_2$ falls $\|\alpha s^{k+j}\|_2 \le \frac{\varepsilon}{2\ell}$.

passendes Vorzeichen wählen $\alpha \cdot \Phi'(0) < 0$

Dabei beschreibt (5) einen gewissen Mindestabstieg.

Falls nun $\|\alpha s^{k+j}\| < \beta_{k+j}$ wäre, so wäre mit

$$
\begin{aligned}
& f(x^{k+j} + \alpha s^{k+j}) \\
\le\; & f(x^{k+j}) - \frac{c\eta}{2} \|\alpha s^{k+j}\|_2 \\
\le\; & f(x^{k+j}) - \frac{c\eta}{2} \frac{\|\alpha s^{k+j}\|_2^2}{\beta_{k+j}} \\
\le\; & f(x^{k+j}) - \gamma \|\alpha s^{k+j}\|_2^2
\end{aligned}
$$

Def. $\tilde{k}_0 \Rightarrow \beta_{k+j} \le \frac{c\eta}{2\gamma}$

die 1. Bed. von (2) erfüllt, und $|\alpha|$ würde nicht weiter reduziert. Die Schranke $\|\alpha s^{k+j}\| \overset{(3a)}{\le} \frac{1}{2} \beta_{k+j}$ wird nicht erreicht.

$\Rightarrow$ Der Fall $\alpha = 0$ kann in maximal $\ell - 1$ aufeinanderfolgenden Iterationen auftreten.

Für $\alpha \neq 0$ impliziert die 2. Bed. aus (2) (mit k statt $k+j$)

$$- \gamma \|\hat{\alpha} s^k\|_2^2 \le f(x^k \underset{(-)}{+} \hat{\alpha} s^k) - f(x^k) \underbrace{\le - \frac{1}{2} c \eta \|\hat{\alpha} s^k\|_2}_{(5) \text{ gilt auch für } \hat{\alpha},\ |\hat{\alpha}| \le \frac{\varepsilon}{2\ell\|s^k\|}}$$

d.h. $\gamma \|\hat{\alpha} s^k\| \ge \frac{1}{2} c\eta$

bzw. $\|x^{k+1} - x^k\| \overset{\alpha = \delta\hat{\alpha}}{=} \|\delta \hat{\alpha} s^k\| \ge \frac{\delta c \eta}{2\gamma} > 0$ im Widerspruch dazu dass $\|x^{k+1} - x^k\|$ eine Nullfolge ist.

#