106. Betrachten Sie das lineare Regressionsproblem: Gegeben seien n Punkte (x_1, y_1), …, (x_n, y_n) im $\mathbb{R}^n$. Gesucht ist eine bestmögliche Annäherung durch eine Gerade $y = d + kx$ in bezug auf die Summe der quadratischen Abstände. Dies führt auf das Problem

$$\min_{(d,k)\in\mathbb{R}^2} \sum_{i=1}^{n} (d + kx_i - y_i)^2.$$

Wir nehmen an, dass Indizes i und j existieren sodass $x_i \neq x_j$.

(a) Schreiben Sie das Problem in ein unrestringiertes Optimierungproblem folgender Form um:

$$\min_{z\in\mathbb{R}^2} (Az - b)^t (Az - b)$$

wobei A eine $n \times 2$ Matrix über $\mathbb{R}$ ist und $b \in \mathbb{R}^n$.

(b) Zeigen Sie, dass einen eindeutigen stationären Punkt gibt und bestimmen Sie diesen (als Funktion der Punkte (x_1, y_1), …, (x_n, y_n)).

107. Zeigen Sie, dass sich die Intervalllängen $b_j - a_j$ der Suchintervalle $[a_j, b_j]$ im in der Vorlesung vorgestellten Algorithmus zur Umsetzung der Wolfe-Powell Schrittweitenstrategie um mindestens den Faktor $\max\{1 - \tau_1, 1 - \tau_2\}$ reduzieren.

108. Gesucht ist ein (Gegen)beispiel, das zeigt, dass die Schrittweitenstrategie von Armijo i.a. keine effiziente Schrittweite liefert und zwar auch dann nicht, wenn die selben Voraussetzungen getroffen werden, die für die Wolfe-Powell Schrittweitenstrategie die Effizienz garantieren.

Hilfestellung:

(a) Wählen Sie die Suchrichtungen $d^{(k)}$ sodass $||d^{(k)}||$ (zu) rasch gegen 0 geht, z.B. $d^{(k)} := -2^{-k}\nabla f(x^{(k)})$.

(b) Zeigen Sie, dass es sich um Abstiegsrichtungen handelt und dass

$$\lim_{k\to\infty} \frac{\nabla f(x^{(k)})^t d^{(k)}}{||d^{(k)}||} = 0 \qquad \Longrightarrow \qquad \lim_{k\to\infty} \nabla f(x^{(k)}) = 0.$$

Ist die Winkelbedingung für alle k erfüllt?

(c) Untersuchen Sie das Verhalten der Armijo-Regel in der Basisversion mit

$$t_A := \max\{\beta^\ell \,|\, \ell \in \{0, 1, 2, \ldots\} \text{ sodass } f(x + td) \leq f(x) + \sigma t \nabla f(x)^t d\}$$

für die folgende Situation: Funktion $f(x) = \frac{x^2}{4}$, Startpunkt $x^{(0)} \neq 1$, Parameter $\sigma \leq \frac{3}{4}$. Welche Schrittweite t_A liefert der Armijo-Algorithmus in dieser Situation?

(d) Zeigen Sie, dass t_A nicht effizient ist, jedoch alle Voraussetzungen, die an im Satz über die Effizienz der Wolfe-Powell Schrittweitenstrategie verlangt werden, erfüllt sind.

(e) Konvergiert der Algorithmus gegen eine Minimimalstelle?

109. Sei $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$, $\sigma \in (0, 1)$, $\beta \in (0, 1)$, $c > 0$ und $x^{(0)} \in \mathbb{R}^n$ fest vorgegeben. Die skalierte Armijo-Regel wählt zu $x \in \mathcal{L}(x^{(0)})$ und $d \in \mathbb{R}^n$ mit $\nabla f(x)^t d < 0$ einen Skalierungsparameter $s > 0$ und bestimmt die Schrittweite t als $\max\{\, s\beta^l \,|\, l = 0, 1, 2, \ldots\}$ sodass $f(x + td) \leq f(x) + \sigma t \nabla f(x)^t d$ gilt. Zeigen Sie

(a) Die skalierte Armijo-Regel ist wohldefiniert.

(b) Ist f nach unten beschränkt, der Gradient von f auf $\mathcal{L}(x^{(0)})$ Lipschitz-stetig und wird s so gewählt, dass

$$s \geq -c\frac{\nabla f(x)^t d}{||d||^2}$$

gilt, so ergibt sich eine effiziente Schrittweitenregel.

110. Sei $f \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R})$ und $x^{(0)} \in \mathbb{R}^n$. Die Minimierungsregel (exakte Schrittweitenstrategie) bestimmt zu $x \in \mathcal{L}(x^{(0)})$ und $d \in \mathbb{R}^n$ mit $\nabla f(x)^t d < 0$ eine Schrittweite $t_{\text{min}} > 0$ mit

$$f(x + t_{\text{min}}) = \min\{f(x + td) \,|\, t \geq 0\}$$

Zeigen Sie dass wenn $\mathcal{L}(x^{(0)})$ kompakt ist und der Gradient ∇f Lipschitz-stetig auf $\mathcal{L}(x^{(0)})$ ist, diese Schrittweitenstrategie wohldefiniert und effizient ist.

111. Angenommen die aktuelle Testschrittweite $\bar{t}$ erfülle nicht die Armijobedingung

$$f(x + td) \leq f(x) + \sigma t \nabla f(x)^t d$$

Eine mögliche Strategie ist die Funktion $\varphi(t) = f(x+td)$ durch ein quadratisches Polynom $P(t)$ zu approximieren von dem man $P(0) = \varphi(0)$, $P'(0) = \varphi'(0)$ und $P(\bar{t}) = \varphi(\bar{t})$ fordert.

(a) Bestimmen Sie die Koeffizienten von P.

(b) Zeigen Sie, dass unter den genannten Voraussetzungen P ein globales Minimum besitzt und zeigen Sie, dass die Stelle $\hat{t}$ für die dieses angenommen wird, sich wie folgt ergibt

$$\hat{t} = -\frac{\bar{t}^2 \varphi'(0)}{2(\varphi(\bar{t}) - \varphi(0) - \bar{t}\varphi'(0)}$$

$\hat{t}$ kann dann als neuer Testkandidat verwendet werden.

(c) Zeigen Sie, dass

$$\hat{t} < \frac{\bar{t}}{2(1-\sigma)}$$

gilt.

112. Bestimmen Sie die Q-Konvergenzordnung folgender Folgen $\{x_k\}$ (Anmerkung: Um die Lesbarkeit zu erleichtern wird hier und in Folgebeispielen wo dies angebracht ist und $n = 1$ gilt x_k und nicht $x^{(k)}$ verwendet.)

(a) $x_k = \dfrac{1}{k}$

(b) $x_k = k^{-k}$

(c) $x_k = 1 + \left(\dfrac{1}{2}\right)^{2^k}$

(d) $x_k = \left(\dfrac{1}{k}\right)^k$

(e) $x_0 := 4$ und $x_{k+1} = \dfrac{1}{2}\left(x_k + \dfrac{4}{x_k}\right)$ für $k \geq 0$.

113. Bestimmen Sie die R-Konvergenzordnung Ordnung folgender Folge

$$x_k = \begin{cases} 1 + \frac{1}{2^k} & \text{für k gerade ,} \\ 1 & \text{für k ungerade.} \end{cases}$$

114. Beweisen Sie die folgenden Aussagen betreff Q-Konvergenz

(a) Der Q-Faktor hängt von der verwendeten Norm ab, die Q-Konvergenzordnung jedoch nicht.

(b) Es existiert stets ein Wert p_0 mit

$$Q_p(\{x^{(k)}\}) = \begin{cases} 0 & \text{für } p \in [1, p_0), \\ \infty & \text{für } p \in [p_0, \infty). \end{cases}$$

115. Betrachten Sie die Folge $\{x_k\}$ mit

$$x_k = \begin{cases} (\frac{1}{4})^{2^k} & \text{für k gerade ,} \\ \frac{x_{k-1}}{k} & \text{für k ungerade.} \end{cases}$$

(a) Liegt superlineare Q-Konvergenz vor?

(b) Liegt quadratische Q-Konvergenz vor?

(c) Liegt quadratische R-Konvergenz vor?

116. Zeigen Sie dass die Folge $\{x_k\}$ mit $x_k = \frac{1}{k}$ nicht Q-linear konvergent ist (man spricht hier von sublinearer Konvergenz).

117. Gegeben sei eine Folge $\{x^{(k)}\}$ die gegen x^* Q-superlinear konvergiert und sei ferner $x^{(k)} \neq x^*$ für alle $k \geq k_0$ für ein $k_o \in \mathbb{N}$. In der Vorlesung wurde gezeigt, dass dann gilt

$$\lim_{k\to\infty} \frac{||x^{(k+1)} - x^{(k)}||}{||x^{(k)} - x^*||} = 1.$$

Zeigen Sie an Hand der Folge $\{x^{(k)}\} = \{x_k\}$ mit $x_{2k-1} = \frac{1}{k!}$ für $k = 1, 2, \ldots$ und $x_{2k} = 2x_{2k-1}$ für $k = 1, 2, \ldots$, dass die Umkehrung nicht gilt (hier wird wiederum wegen $n = 1$ und der Lesbarkeit x_k statt $x^{(k)}$ verwendet).

118. Gegeben sei die Funktion $f : \mathbb{R}^2 \to \mathbb{R}$ mit

$$f(x_1, x_2) = x_1^2 + 2x_2^2.$$

(a) Bestimmen Sie die Richtung des steilsten Abstiegs.

(b) Angenommen wir starten in $x^{(0)} = (2, 1)^t$. Welcher Punkt $x^{(1)}$ ergibt sich im Rahmen des steilsten Abstiegsverfahrens mit exakter Schrittweitensuche im nächsten Schritt?

(c) Geben Sie eine Formel für den Punkt $x^{(k)}$ in Iteration k an.

119. Wenden Sie das steilste Abstiegsverfahren auf die quadratische Funktion

$$f(x) := 1/2x^tQx + c^tx + \gamma$$

mit

$$Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 25 \end{pmatrix} \qquad c = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

mit exakter Schrittweitensuche an. Starten Sie um Ursprung $x^{(0)} = (0, 0, 0)^t$.

(a) Geben Sie $x^{(k)}$, $f(x^{(k)})$, $\nabla f(x^{(k)})$, $||\nabla f(x^{(k)})||$ für die ersten 3 Iterationen an.

(b) Wieviele Iterationen benötigt man um die (euklidische) Norm des Gradienten unter 10^-8 zu drücken? (Für diese Aufgabe ist nichts abzugeben - jedoch kann ein Codestück für eine Implementationsaufgabe ebenfalls nützlich sein.)

(c) Was ändert sich, wenn der Eintrag q_{22} auf γ und der Eintrag q_{33} auf γ^2 gesetzt wird? Betrachten Sie unterschiedliche Werte von γ $(1, 10, 100, 1000, 10000, 10^8)$. Wie erklären Sie sich das unterschiedliche Verhalten?