Technische Universität München
Zentrum Mathematik
Prof. Dr. P. Gritzmann, Dipl.-Inf. Dipl.-Math. S. Borgwardt

Optimierung 3, SS 2009

Übungsblatt 12

Aufgabe 12.1
Sei $\varphi : \mathbb{R}^n \to \mathbb{R}$ stetig differenzierbar und nach unten beschränkt, $x_0 \in \mathbb{R}^n$, s eine Abstiegsrichtung sowie $\rho \in (0,1)$ und $\kappa \in (\rho, 1)$.

a) Konstruieren Sie einen möglichst schnellen Algorithmus zur Bestimmung einer Schrittweite, die die Wolfe-Powell Bedingungen für ρ und κ erfüllt. (*Hinweis: Lemma* 9.2.14)

b) Begründen Sie, dass Ihr Verfahren nach endlich vielen Schritten eine Schrittweite findet, die den Wolfe-Powell Bedingungen genügt.

c) Betrachten Sie die Funktion $\varphi(x) := \xi_1^2 + 2\xi_1\xi_2 + 5\xi_2^2$. Berechnen Sie die ersten drei Schritte des Gradientenverfahren mit Startpunkt $x_0 = (3, -1)^T$ für diese Funktion. Benutzen Sie hierbei Ihr Verfahren für die Schrittweitenbestimmung, mit Parametern $\rho = 1/2, \kappa = 3/4$.

Lösung zu Aufgabe 12.1

a) Wir konstruieren eine Schrittweite, die die Wolfe-Powell Bedingungen erfüllt durch Bisektion. Dazu bestimmen wir analog zu Lemma 9.2.14 Folgen $(\mu_i)_{\mathbb{N}}$ und $(\nu_i)_{\mathbb{N}}$ nach folgendem Verfahren erzeugt:

- Wir konstruieren zunächst Startpunkte μ_0, ν_0 (beispielsweise) wie folgt:

  Wähle $\mu_0 \in \{2^{-l} : l \in \mathbb{N}_0\}$ maximal, so dass μ_0 die Ungleichung $\varphi(x_0 + \mu_0 s) \leq \varphi(x_0) + \mu_0\rho\varphi'(x_0)^T s$ erfüllt. Setze dann $\nu_0 := 2\mu_0$, somit erfüllt ν_0 die Ungleichung $\varphi(x_0 + \nu_0 s) > \varphi(x_0) + \nu_0\rho\varphi'(x_0)^T s$. Das ist natürlich nur dann sichergestellt, wenn $\mu_0 \neq 1$ ist. Ist das nicht der Fall, dann wählen wir das minimale $\nu_0 \in \{2^l : l \in \mathbb{N}\}$, so dass ν_0 die Ungleichung $\varphi(x_0 + \nu_0 s) > \varphi(x_0) + \nu_0\rho\varphi'(x_0)^T s$ erfüllt und setzen dann $\mu_0 := 1/2\nu_0$.

  Für unsere Startpunkte $\mu_0 < \nu_0$ gilt nach dieser Konstruktion:

  $$\begin{aligned}\varphi(x_0 + \mu_0 s) &\leq \varphi(x_0) + \mu_0\rho\varphi'(x_0)^T s\\ \varphi(x_0 + \nu_0 s) &> \varphi(x_0) + \nu_0\rho\varphi'(x_0)^T s\end{aligned}$$

- Sind die ersten k Folgenglieder bereits bekannt, so betrachten wir $\lambda = \frac{\mu_k + \nu_k}{2}$. Erfüllt λ die Wolfe-Powell Bedingungen, so brechen wir den Algorithmus ab. Sonst prüfen wir, ob λ die Ungleichung $\varphi(x_0 + \lambda s) \leq \varphi(x_0) + \lambda\rho\varphi'(x_0)^T s$ erfüllt. Ist das der Fall, so setzen wir $\mu_{k+1} := \lambda$ und $\nu_{k+1} := \nu_k$, sonst $\mu_{k+1} := \mu_k$ und $\nu_{k+1} := \lambda$.

b) Mit dem oben ausgeführten Algorithmus erhalten wir (falls wir nicht abbrechen) monoton wachsende bzw. fallende Folgen (μ_i) bzw. (ν_i) mit $\lim_{i\to\infty}\mu_i = \lim_{i\to\infty}\nu_i$, für die alle Voraussetzungen von Lemma 9.2.14 gelten (die Ungleichungen gelten nach Konstruktion). Damit gibt es ein $i_0 \in \mathbb{N}$, so dass für alle $i \geq i_0$ die μ_i beide Wolfe-Powell Bedingungen erfüllen. Wir brechen den Algorithmus ab, sobald die erste solche Schrittweite gefunden wird.

c) Wir berechnen zunächst

$$\varphi'(x) = \begin{pmatrix} 2\xi_1 + 2\xi_2 \\ 2\xi_1 + 10\xi_2 \end{pmatrix}$$

und erhalten damit die Suchrichtung $s^{(1)} = -\varphi'(x^{(0)}) = (-4,4)^T$. Damit können wir μ_0 bestimmen; wir probieren einfach $\mu_0 = 2^{-l}, l = 0,1,2,\ldots$, bis die geforderte Ungleichung erfüllt ist (natürlich könnte man die Ungleichung auch nach μ_0 auflösen – das ist in der Praxis bei komplizierterem φ aber meist nicht mehr möglich). Hier gilt:

$$\begin{array}{lllr}
\mu_0 = 1: & \varphi(x^{(0)} + s^{(1)}) \leq \varphi(x^{(0)}) + \frac{1}{2}\varphi'(x^{(0)})^T s^{(1)} & \Leftrightarrow 40 \leq 8 - \frac{32}{2} & \text{false} \\
\mu_0 = \frac{1}{2}: & \varphi(x^{(0)} + \frac{1}{2}s^{(1)}) \leq \varphi(x^{(0)}) + \frac{1}{4}\varphi'(x^{(0)})^T s^{(1)} & \Leftrightarrow 8 \leq 8 - \frac{32}{4} & \text{false} \\
\mu_0 = \frac{1}{4}: & \varphi(x^{(0)} + \frac{1}{4}s^{(1)}) \leq \varphi(x^{(0)}) + \frac{1}{8}\varphi'(x^{(0)})^T s^{(1)} & \Leftrightarrow 4 \leq 8 - \frac{32}{8} & \text{true}
\end{array}$$

Also starten wir die Bisektion mit $\mu_0 = \frac{1}{4}$ und $\nu_0 = 2\mu_0 = \frac{1}{2}$. Wir prüfen zunächst, ob μ_0 bereits beide Wolfe-Powell Bedingungen erfüllt, und stellen fest, dass dies zutrifft. Somit ist der nächste Iterationspunkt

$$x^{(1)} = x^{(0)} + \frac{1}{4}s^{(1)} = (2,0)^T.$$

Die nächste Suchrichtung ist dann $s^{(2)} = -\varphi'(x^{(1)}) = (-4,-4)^T$ und wir bestimmen wieder μ_0 durch Einsetzen:

$$\begin{array}{lllr}
\mu_0 = 1: & \varphi(x^{(1)} + s^{(2)}) \leq \varphi(x^{(1)}) + \frac{1}{2}\varphi'(x^{(1)})^T s^{(2)} & \Leftrightarrow 100 \leq 4 - \frac{32}{2} & \text{false} \\
\mu_0 = \frac{1}{2}: & \varphi(x^{(1)} + \frac{1}{2}s^{(2)}) \leq \varphi(x^{(1)}) + \frac{1}{4}\varphi'(x^{(1)})^T s^{(2)} & \Leftrightarrow 20 \leq 4 - \frac{32}{4} & \text{false} \\
\mu_0 = \frac{1}{4}: & \varphi(x^{(1)} + \frac{1}{4}s^{(2)}) \leq \varphi(x^{(1)}) + \frac{1}{8}\varphi'(x^{(1)})^T s^{(2)} & \Leftrightarrow 4 \leq 4 - \frac{32}{8} & \text{false} \\
\mu_0 = \frac{1}{8}: & \varphi(x^{(1)} + \frac{1}{8}s^{(2)}) \leq \varphi(x^{(1)}) + \frac{1}{16}\varphi'(x^{(1)})^T s^{(2)} & \Leftrightarrow 2 \leq 4 - \frac{32}{16} & \text{true}
\end{array}$$

Auch diesmal stellen wir fest, dass $\mu_0 = \frac{1}{8}$ die Wolfe-Powell-Bedingungen bereits erfüllt.

Der nächste Iterationspunkt ist also

$$x^{(2)} = x^{(1)} + \frac{1}{8}s^{(2)} = \left(\frac{3}{2}, -\frac{1}{2}\right)^T,$$

und damit bekommen wir die Suchrichtung $s^{(3)} = -\varphi'(x^{(2)}) = (-2, 2)^T$.

$$\begin{array}{lllr}
\mu_0 = 1: & \varphi(x^{(2)} + s^{(3)}) \leq \varphi(x^{(2)}) + \frac{1}{2}\varphi'(x^{(2)})^T s^{(3)} & \Leftrightarrow 10 \leq 2 - \frac{8}{2} & \text{false} \\
\mu_0 = \frac{1}{2}: & \varphi(x^{(2)} + \frac{1}{2}s^{(3)}) \leq \varphi(x^{(2)}) + \frac{1}{4}\varphi'(x^{(2)})^T s^{(3)} & \Leftrightarrow 2 \leq 2 - \frac{8}{4} & \text{false} \\
\mu_0 = \frac{1}{4}: & \varphi(x^{(2)} + \frac{1}{4}s^{(3)}) \leq \varphi(x^{(2)}) + \frac{1}{8}\varphi'(x^{(2)})^T s^{(3)} & \Leftrightarrow 1 \leq 2 - \frac{8}{8} & \text{true}
\end{array}$$

Die Wolfe-Powell Schrittweite ist diesmal $\mu_0 = \frac{1}{4}$, damit bekommen wir

$$x^{(3)} = x^{(2)} + \frac{1}{4}s^{(3)} = (1, 0)^T.$$

Aufgabe 12.2

Sei $\gamma(t) := (1+2^{-t})\begin{pmatrix}\cos 2\pi t\\ \sin 2\pi t\end{pmatrix}$ und $\Gamma := \{\gamma(t) : t \in \mathbb{R}\}$ (vgl. Abbildung). Für $x \in \mathbb{R}^2$ mit $\|x\|_2 > 1$ sei $\lambda(x) := \max\{\lambda \leq 1 : \lambda x \in \Gamma\}$ und $\mu(x) := \min\{\mu > 1 : \mu x \in \Gamma\}$. Wir betrachten die Funktion $f(x) := 2^{-20}(1-\|x\|_2)^8 + d(x)$, wobei

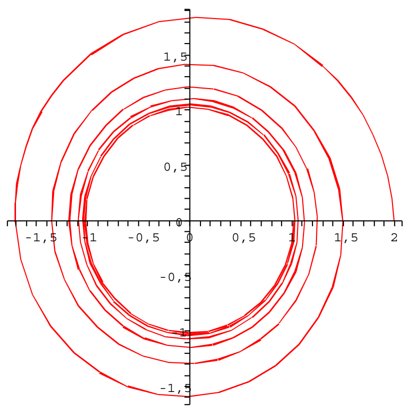
$$d(x) := \begin{cases} \|x - x\lambda(x)\|_2^4 \cdot \|x - x\mu(x)\|_2^4, & \text{für } \|x\|_2 > 1. \\ 0, & \text{sonst} \end{cases}$$

eine Abstandsfunktion bezüglich Γ ist.

a) Skizzieren Sie qualitativ den Verlauf von f entlang des Strahls $\{(\sigma,0)^T : \sigma \geq 0\}$. Beschreiben Sie mit Hilfe Ihrer Skizze informell das Aussehen von f auf ganz $\mathbb{R}^2$.

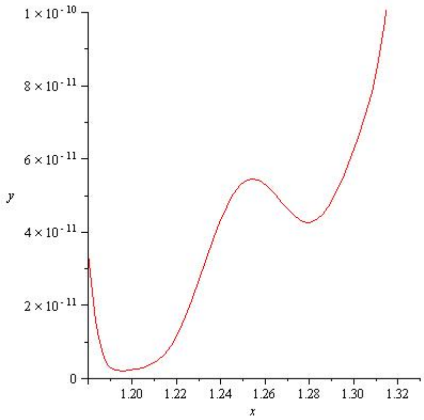
b) Wie sieht (qualitativ) der Verlauf des Gradientenverfahrens mit exakter line search für f aus, wenn man bei $x_0 = (2,0)^T$ startet? Wo liegen die Häufungspunkte, die das Verfahren erzeugt?

c) Welche Rolle spielt das Verhalten von f im Inneren des Einheitskreises für den Verlauf des Gradientenverfahrens aus b)?



Lösung zu Aufgabe 12.2

a) Für unsere qualitative Skizze nehmen wir an, dass die durch Γ definierte Spirale die x-Achse bei 1.2 und 1.3 trifft. Die folgende Zeichnung zeigt den Verlauf von f im entsprechenden Bereich. Man erkennt die zwei zu 1.2 und 1.3 gehörigen „Täler“ von f, die sich aus der Abstandsfunktion d und dem für größer werdende x steigenden additiven Teil von f ergeben. Die Abstandsfunktion dominiert den Restteil von f zwischen den Spiralarmen deutlich. Qualitativ lässt sich die Zeichnung ähnlich fortsetzen, die Höhe der Abstandsfunktion nimmt dabei aber sehr stark zu, die des Restteils nur leicht.



Anhand dieser Überlegung kann man sich gut klarmachen, wie die Funktion aussieht: Die Spirale wird mit einer leichten Neigung versehen (durch den Faktor $1 + 2^{-t}$ zieht sie sich bei steigendem t näher an den Einheitskreis heran, und das führt zu einem kleiner werdenden Restteil $2^{-20}(1 - ||x||)^8$ in f), zwischen je zwei Spiralarmen wird die Funktion zu einer Art Rinne mit sehr hohen „Grenzwänden“. Man bekommt also so etwas wie eine spiralförmige Rutsche, die sich asymptotisch dem Rand des Einheitskreises nähert. Zwar sind benachbarte Rutschenwindungen miteinander verbunden, die Höhe der Wand, welche die Rinne begrenzt, ist aber sehr hoch.

b) Der Punkt $(2, 0)^T$ liegt auf der Spirale, nach links und rechts erheben sich steile „Wände“. Im ersten Schritt laufen wir (in Richtung zum Nullpunkt) in den Talpunkt der Senke, in der wir uns befinden, und danach (senkrecht dazu) in Richtung des steilsten Abstiegs, entlang der ablaufenden Spirale. Danach laufen wir erneut senkrecht (parallel zur ersten Suchrichtung!) zur aktuellen Talsenke, und danach wieder senkrecht dazu (parallel zur zweiten Suchrichtung!) weiter entlang der Spirale. Das Gradientenverfahren mit lokal exakter Schrittweitenbestimmung folgt damit der Spiralrichtung in kleinen orthogonalen Zick-Zack-Schritten. Die Häufungspunkte liegen dann natürlich auf dem Rand des Einheitskreises.

c) Wenn man das Gradientenverfahren außerhalb des Einheitskreises startet, wird die „Rutschrinne“ nie verlassen, das Verfahren gelangt also nie ins Innere des Einheitskreises. Das Verhalten von f in diesem Bereich spielt somit keine Rolle für den Verlauf des Gradientenverfahrens. Das bedeutet auch, dass man f so manipulieren könnte, dass lokale Minima nur im Inneren des Einheitskreises angenommen werden, so dass das Gradientenverfahren zwar gegen stationäre Punkte konvergiert („Terassenpunkte“), nicht aber gegen ein Minimum (weder global noch lokal).