Technische Universität Darmstadt
Fachbereich Mathematik
Prof. Dr. S. Ulbrich

Wintersemester 2009/10

Blatt 1

Lösungsvorschlag zur Vorlesung
Nichtglatte Optimierung und Anwendungen

G1. Versagen der exakten Schrittweitensuche bei nichtglatten Problemen

a): Es gilt

$$\nabla f_1(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}.$$

Die Richtung des steilsten Abstiegs in x ist $s = -\nabla f_1(x)$.

$$\frac{d}{dt} f_1(x+ts) = -\nabla f_1(x+ts)^T s = \begin{pmatrix} 2(x_1 - 2tx_1) \\ 4(x_2 - 4tx_2) \end{pmatrix}^T \begin{pmatrix} -2x_1 \\ -4x_2 \end{pmatrix} = -4x_1^2 - 16x_2^2 + t(8x_1^2 + 64x_2^2).$$

Die optimale Schrittweite ist daher

$$\sigma = \frac{x_1^2 + 4x_2^2}{2x_1^2 + 16x_2^2}. \tag{1}$$

Starten wir das Verfahren des steilsten Abstiegs in $x^0 = (2,1)^T$, so gilt

$$s^0 = \begin{pmatrix} -4 \\ -4 \end{pmatrix}, \quad \sigma_0 = \frac{8}{24} = \frac{1}{3}, \quad x^1 = \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \frac{1}{3}\begin{pmatrix} -4 \\ -4 \end{pmatrix} = \frac{1}{3}\begin{pmatrix} 2 \\ -1 \end{pmatrix},$$
$$s^1 = \frac{1}{3}\begin{pmatrix} -4 \\ 4 \end{pmatrix}, \quad \sigma_1 = \frac{8}{24} = \frac{1}{3}, x^2 = \frac{1}{3}\begin{pmatrix} 2 \\ -1 \end{pmatrix} + \frac{1}{9}\begin{pmatrix} -4 \\ 4 \end{pmatrix} = \frac{1}{9}\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{1}{9}x^0.$$

Wir zeigen durch vollständige Induktion, dass

$$x^{2k} = 9^{-k}x^0, \quad x^{2k+1} = 9^{-k}x^1.$$

Für x^0, x^1, x^2 ist die Formel nachgewiesen.

Induktionsschritt $k \to k+1$:

Wir haben nach Induktionsannahme $x^{2k+1} = 9^{-k}x^1$ und damit (∇f ist linear in x)

$$\nabla f(x^{2k+1}) = 9^{-k}\nabla f(x^1)$$

und wegen (1)

$$\sigma_{2k+1} = \frac{9^{-2k}}{9^{-2k}}\sigma_1 = \sigma_1 = \frac{1}{3}.$$

Dies ergibt

$$x^{2k+2} = x^{2k+1} - \sigma_{2k+1}\nabla f(x^{2k+1}) = 9^{-k}(x^1 - \sigma_1 \nabla f(x^1)) = 9^{-k}x^2 = 9^{-(k+1)}x^0.$$

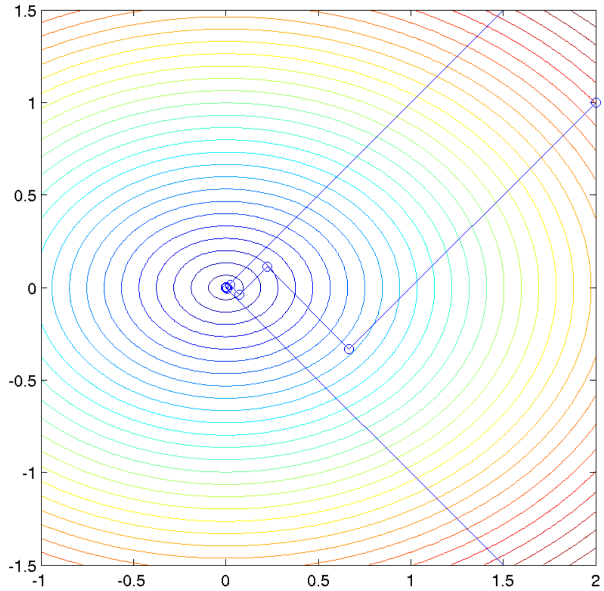

Abbildung 1: Niveaulinien von f_1 und die Iterierten

Völlig analog gilt nun

$$x^{2k+3} = x^{2k+2} - \sigma_{2k+2}\nabla f(x^{2k+2}) = 9^{-(k+1)}(x^0 - \sigma_0 \nabla f(x^0)) = 9^{-(k+1)}x^1.$$

Insbesondere also $x^k \to 0$. Weiter gilt

$$\nabla f_1(x^{2k}) = 9^{-k}\begin{pmatrix}4\\4\end{pmatrix} \to 0, \quad \nabla f_1(x^{2k+1}) = \frac{1}{3}9^{-k}\begin{pmatrix}4\\-4\end{pmatrix} \to 0$$

Abb. 1 zeigt die Niveaulinien von f_1 und die Iterierten:

b): Sei nun $f_2(x) = \sqrt{f_1(x)}$. Für $x \neq 0$ ist $f_1(x) \neq 0$ und daher f_2 stetig differenzierbar bei x mit

$$\nabla f_2(x) = \frac{1}{2f_2(x)}\nabla f_1(x),$$

d.h. die Richtung des steilsten Abstiegs von f_2 stimmt in allen $x \neq 0$ mit denen von f_1 (bis auf die Länge) überein. Minimimierung von f_2 auf $\{x - \sigma\nabla f_2(x) \ : \ \sigma \geq 0\}$ liefert denselben Minimalpunkt wie die Minimimierung von f_1 auf $\{x - \sigma\nabla f_1(x) \ : \ \sigma \geq 0\}$, da $\sqrt{\cdot}$ streng monoton wachsend of $[0, \infty$ ist. Das Verfahren des steilsten Abstiegs, gestartet in $x^0 = (2,1)^T$, erzeugt daher für f_2 dieselbe Punktefolge wie für f_1.

Da $(0,0)$ das Minimum von f_2 ist, konvergiert also das Verfahren.

c) Man prüft leicht, dass f_3 stetig und konvex ist.

Das Verfahren verhält sich genau wie für f_2: Vollständige Induktion: Natürlich stimmt x^0 überein. Induktionsannahme: x^k stimmt überein. In x^k gilt $f_3(x^k) = f_2(x^k)$ und $\nabla f_3(x^k) = \nabla f_2(x^k)$. Auf dem Strahl $S_k = \{x^k - \sigma\nabla f_3(x^k) \ : \ \sigma \geq 0\}$ stimmen f_3 und f_2 auf der Strecke von x^k bis x^{k+1} und ein Stück darüber hinaus überein. Die konvexe Funktion f_3 hat also wie f_2 auf S_k in x^{k+1} ein globales Minimum und dieses ist wie bei f_2 eindeutig. Also stimmt auch die nächste Iterierte x^{k+1} überein.

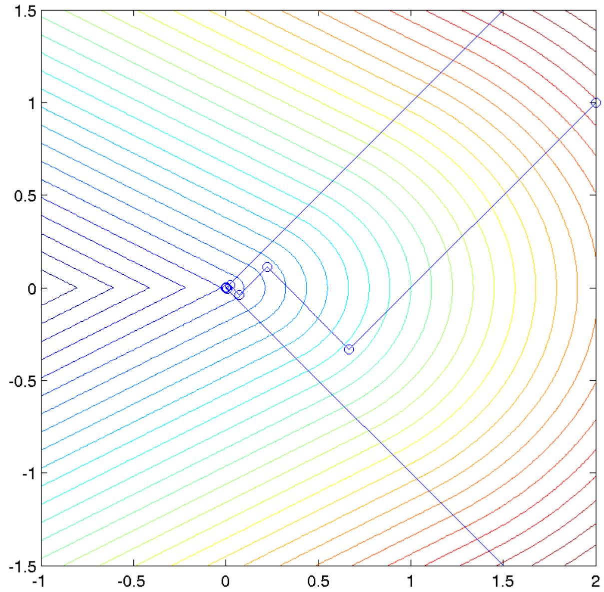

Abbildung 2: Niveaulinien von f_3 und die Iterierten

Abb. 2 zeigt die Niveaulinien von f_3 und die Iterierten:

d): Das Beispiel zeigt, dass sich das Verfahren nur im glatten Teil aufhält und gegen den falschen Punkt konvergiert. Eine Erweiterung des Verfahrens auf nichtglatte Funktionen würde aber das Verhalten im glatten Teil nicht ändern.

G2.

zu a): Es gilt $g \in \partial f(0)$ genau dann, wenn

$$f(y) \geq f(0) + g\,(y - 0) \quad \forall\, y \in \mathbb{R}.$$

Nach Definition von f ergibt dies

$$y \geq g\,y \quad \forall\, y \geq 0, \quad y^2 \geq g\,y \quad \forall\, y < 0$$

und dies ist äquivalent zu

$$g \leq 1 \quad \text{und} \quad g \geq 0,$$

also $g \in [0, 1]$.

zu b): Es gilt $g \in \partial f(0)$ genau dann, wenn

$$f(y) \geq f(0) + g^T(y - 0) \quad \forall\, y \in \mathbb{R}^n.$$

Wegen $f(x) = \|x\| = \sqrt{x^T x}$ ergibt sich

$$\|y\| \geq g^T y \quad \forall\, y \in \mathbb{R}^n. \tag{2}$$

Nach der Cauchy-Schwarzschen Ungleichung ist $g^T y \leq \|g\|\|y\|$ für alle $y \in \mathbb{R}^n$ und diese Ungleichung ist scharf für $y = g$. Daher ist (2) nur erfüllt für g mit $\|g\| \leq 1$, also $\partial f(0) = \{g \in \mathbb{R}^n \ : \ \|g\| \leq 1\}$.

G3. Epigraph und Subgradienten

zu a):

Seien X und f konvex.

Für $(x_i^T, \alpha_i)^T \in \text{epi}(f)$, $i = 0, 1$, $\lambda \in [0, 1]$ sowie $x_\lambda = (1 - \lambda)x_0 + \lambda x_1$ und $\alpha_\lambda = (1 - \lambda)\alpha_0 + \lambda\alpha_1$ ergibt sich

$$x_\lambda \in X \quad (\text{da } X \text{ konvex ist}), \alpha_\lambda \geq (1 - \lambda)f(x_0) + \lambda f(x_1) \geq f(x_\lambda).$$

Damit gilt $(x_\lambda^T, \alpha_\lambda)^T \in \text{epi}(f)$.

Sei nun epi(f) konvex.

Für $x_i \in X$, $i = 0, 1$, $\lambda \in [0, 1]$ und $\alpha_i = f(x_i)$ gilt dann mit x_λ, α_λ wie oben:

$$\begin{pmatrix} x_\lambda \\ \alpha_\lambda \end{pmatrix} \in \text{epi}(f),$$

also

$$x_\lambda \in X, \quad f(x_\lambda) \leq \alpha_\lambda = (1 - \lambda)f(x_0) + \lambda f(x_1).$$

Somit sind X und f konvex.

zu b):

Sei $g \in \partial f(x)$. Dann gilt für alle $(y^T, \alpha)^T \in \text{epi}(f)$

$$(g^T, -1)\left(\begin{pmatrix} y \\ \alpha \end{pmatrix} - \begin{pmatrix} x \\ f(x) \end{pmatrix}\right) = g^T(y - x) - \alpha + f(x) \leq g^T(y - x) - f(y) + f(x) \leq 0.$$

Es gelte umgekehrt $(*)$. Dann haben wir $(y^T, f(y))^T \in \text{epi}(f)$ für alle $y \in X$ und somit

$$0 \geq (g^T, -1)\left(\begin{pmatrix} y \\ f(y) \end{pmatrix} - \begin{pmatrix} x \\ f(x) \end{pmatrix}\right) = g^T(y - x) - f(y) + f(x),$$

d.h. $g \in \partial f(x)$.

zu c):

Aus $(*)$ und der Tatsache, dass die letzte Koordinate in der Normale von H negativ ist, folgt, dass der Epigraph von f auf oder oberhalb der Ebene H liegt (die $n + 1$-te Koordinatenachse sei nach oben angetragen). Des weiteren liegt der Punkt $(x^T, f(x))^T$ auf H. Es ist daher gerechtfertigt, zu sagen, dass H den Epigraphen in $(x^T, f(x))^T$ von unten stützt.