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

Optimierung 3, SS 2009

Übungsblatt 8

Aufgabe 8.1
Sei $g : \mathbb{R}^n \to \mathbb{R}^m$ eine stetig differenzierbare Abbildung, die mindestens eine Nullstelle besitzt.

a) Gibt es zu jedem g eine stetig differenzierbare Funktion $f : \mathbb{R}^m \to \mathbb{R}$ mit $f'(x) = g(x)$? (Die Nullstellen von g sind dann genau die stationären Punkte von f.)

b) Geben Sie eine Funktion $f : \mathbb{R}^n \to \mathbb{R}$ an, deren globale Minima genau die Nullstellen von g sind.

c) Sind auch alle stationären Punkte der Funktion f aus b) schon Nullstellen von g? Geben Sie ggf. ein Gegenbeispiel an.

d) Zeigt Aufgabe b) eine praktikable Möglichkeit, ein globales Optimierungsproblem auf die Bestimmung von Nullstellen einer Funktion zurückzuführen?

Lösung zu Aufgabe 8.1

a) Für $n \neq m$ passen die Definitionsbereiche von f' und g nicht zueinander. Sei also im Folgenden $n = m$. Wir betrachen als Gegenbeispiel die Abbildung $g(\xi_1, \xi_2) := (\xi_1\xi_2, \xi_1^2 - 1)^T$. Gäbe es eine Funktion f mit $f'(x) = g(x)$, so müsste gelten

$$\frac{\partial}{\partial \xi_1} f(\xi_1, \xi_2) = \xi_1\xi_2 \Rightarrow f(\xi_1, \xi_2) = \frac{1}{2}\xi_1^2\xi_2 + C_1(\xi_2)$$
$$\frac{\partial}{\partial \xi_2} f(\xi_1, \xi_2) = \xi_1^2 - 1 \Rightarrow f(\xi_1, \xi_2) = \xi_1^2\xi_2 - \xi_2 + C_2(\xi_1)$$

Da diese beiden Definitionen für alle $x \in \mathbb{R}^2$ übereinstimmen müssen, können wir beispielsweise $(\xi_1, 0)^T$ einsetzen. Damit muss gelten

$$C_2(\xi_1) = C_1(0) \text{ für alle } \xi_1 \in \mathbb{R}.$$

Also ist C_2 eine Konstante. Setzt man $(0, \xi_2)^T$ ein, so erhält man

$$C_1(\xi_2) = -\xi_2 + C_2.$$

Damit haben wir f bis auf eine Konstante bestimmt, es ist daher $f(x) = {}^1\!/_2\xi_1^2\xi_2 - \xi_2 + C_2 = \xi_1^2\xi_2 - \xi_2 + C_2$. Umformen ergibt ${}^1\!/_2\xi_1^2\xi_2 = \xi_1^2\xi_2$ für alle $(\xi_1, \xi_2) \in \mathbb{R}$. Das ist aber nicht möglich, daher kann es keine Funktion f mit den gewünschten Eigenschaften geben.

b) Wir definieren $f : \mathbb{R}^n \to \mathbb{R}$ durch

$$f(x) := \sum_{i=1}^{m} (g_i(x))^2 ,$$

wobei $g(x) = (g_1(x), \dots, g_m(x))^T$ ist. Dann ist $f(x) = 0$ genau dann, wenn $g_1(x) = \dots = g_m(x) = 0$, und da $f(x) \geq 0$ gilt, ist jede Nullstelle von g ein globales Minimum von f und umgekehrt.

c) Es gilt

$$f'(x) = \sum_{i=1}^{m} 2g_i(x)g_i'(x) = 2g'(x)^T g(x).$$

Also ist x insbesondere dann ein stationärer Punkt von f, wenn $g'(x) = 0$, also z.B. wenn $g : \mathbb{R} \to \mathbb{R}$ in x ein lokales Maximum hat. Das muss natürlich nicht nur für $g(x) = 0$ der Fall sein. Es gibt auch Gegenbeispiele mit $g'(x) \neq 0$.

Ein solches ist etwa $g(\xi_1, \xi_2) := (\xi_1\xi_2, \xi_1^2 - 1)^T$. Damit ergibt sich

$$g'(x)^T = \begin{pmatrix} \xi_2 & 2\xi_1 \\ \xi_1 & 0 \end{pmatrix}$$

und

$$\begin{aligned} f(x) &= \xi_1^2\xi_2^2 + (\xi_1^2 - 1)^2 \\ f'(x) &= \begin{pmatrix} 2\xi_1\xi_2^2 + 4\xi_1(\xi_1^2 - 1) \\ 2\xi_1^2\xi_2 \end{pmatrix} . \end{aligned}$$

Ein stationärer Punkt der Funktion f ist also beispielsweise $x^* = (0, 1)^T$, hier ist offenbar $g'(x^*)^T$ singulär. Es gilt jedoch $g(x^*) = (0, -1)^T$, der Punkt x^* ist also keine Nullstelle von g.

d) Die Aufgabe zeigt, dass sich die Nullstellenbestimmung auf ein Optimierungsproblem zurückführen lässt. Da Optimierungsprobleme meistens schwieriger zu lösen sind als Nullstellenprobleme, ist man natürlich eher am umgekehrten Weg interessiert. Für Funktionen f, deren globales Minimum 0 ist, könnte man tatsächlich versuchen, die Nullstellen der Funktion g zu bestimmen, um die Minima zu finden. Leider ist es in den meisten Fällen nicht so einfach, eine Darstellung von f in der Form $f(x) = \sum_{i=1}^{m}(g_i(x))^2$ (man spricht von Sum of Squares-Darstellung) anzugeben. Ist das globale Minimum von f nicht 0, müsste die Funktion zudem vorher entsprechend verschoben werden, was zumindest die Kenntnis des Minimalwertes voraussetzt. Auch das ist in vielen Problemen leider nicht gegeben. In einigen Sonderfällen ist es aber durchaus möglich, eine Sum of Squares-Zerlegung mit praktikablem Aufwand zu ermitteln.

Aufgabe 8.2

Seien die Funktionen $\varphi : \mathbb{R}_+^n \to \mathbb{R}$, $\alpha : \mathbb{R}^n \to \mathbb{R}^m, \alpha(x) := (\alpha_1(x)^T, \dots, \alpha_m(x)^T)^T$ konvex und es gelte die Slater-Bedingung. Es gebe $x^* \geq 0 \in \mathbb{R}^n$ und $y^* \geq 0 \in \mathbb{R}^m$ mit

$$\begin{array}{rrcl} & \frac{d}{dx}\Lambda_x(x^*) & \geq & 0, \\ & x^{*T}\frac{d}{dx}\Lambda_x(x^*) & = & 0, \\ & \frac{d}{dy}\Lambda_y(y^*) & \leq & 0 \\ \text{sowie} & y^{*T}\frac{d}{dy}\Lambda_y(y^*) & = & 0. \end{array}$$

Hierbei seien

$$\Lambda_x : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \Lambda(x, y^*) = \varphi(x) + {y^*}^T\alpha(x)$$

und

$$\Lambda_y : \mathbb{R}^m \to \mathbb{R},\ y \mapsto \Lambda(x^*, y) = \varphi(x^*) + y^T\alpha(x^*).$$

Beweisen Sie, dass x^* ein Minimalpunkt von φ über $F := \bigcap_{i=1}^{m} L_{\alpha_i}(0)$ ist.

Lösung zu Aufgabe 8.2
Seien $x^* \geq 0 \in \mathbb{R}^n$ und $y^* \geq 0 \in \mathbb{R}^m$ zwei Punkte mit

$$\begin{array}{rrcl} & \frac{d}{dx}\Lambda_x(x^*) & \geq & 0, \\ & x^{*T}\frac{d}{dx}\Lambda_x(x^*) & = & 0, \\ & \frac{d}{dy}\Lambda_y(y^*) & \leq & 0 \\ \text{sowie} & y^{*T}\frac{d}{dy}\Lambda_y(y^*) & = & 0. \end{array}$$

Nach Vorlesung reicht es zu zeigen, dass $(x^{*T}, y^{*T})^T$ ein Sattelpunkt der Lagrangefunktion ist, d.h., für alle $x \geq 0$ und für alle $y \geq 0$ gilt

$$\Lambda(x^*, y) \leq \Lambda(x^*, y^*) \leq \Lambda(x, y^*).$$

Wir betrachten dazu die gegebenen Funktionen

$$\Lambda_x : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \Lambda(x, y^*) = \varphi(x) + {y^*}^T\alpha(x)$$

und

$$\Lambda_y : \mathbb{R}^m \to \mathbb{R},\ y \mapsto \Lambda(x^*, y) = \varphi(x^*) + y^T\alpha(x^*).$$

Da φ und α konvexe Funktionen sind, folgt die Konvexität von Λ_x und die Linearität von Λ_y. Wir erhalten somit für $x \geq 0$

$$\Lambda(x, y^*) - \Lambda(x^*, y^*) = \Lambda_x(x) - \Lambda_x(x^*) \overset{\Lambda_x \text{ konv.}}{\geq} \frac{d}{dx}\Lambda_x(x^*)^T(x - x^*) \overset{\text{Vor.}}{=} \frac{d}{dx}\Lambda_x(x^*)^T x \overset{\text{Vor.}}{\geq} 0$$

und für $y \geq 0$

$$\Lambda(x^*, y) - \Lambda(x^*, y^*) = \Lambda_y(y) - \Lambda_y(y^*) \overset{\Lambda_y \text{ lin.}}{=} \frac{d}{dy}\Lambda_y(y^*)^T(y - y^*) \overset{\text{Vor.}}{=} \frac{d}{dy}\Lambda_y(y^*)^T y \overset{\text{Vor.}}{\leq} 0.$$