Lösung 6

Hinweise

1. Vergessen Sie beim Anwenden der Kettenregel nicht, ins Differential von g den Punkt $f(x, y)$ einzusetzen.

2. Wie in Abschnitt 10.2.1 des Skripts erklärt, zeigt der Gradient ∇f in die Richtung des steilsten Anstiegs.

3. Die Hesse-Matrix im einzigen kritischen Punkt von f ist degeneriert, sagt also nichts aus. In b) setzen Sie $v = (x, y)$. Verwenden Sie dann die Methoden der Analysis I in der Variablen t (z.B. Korollar 7.37 und/oder Beispiel 8.61). Dabei sind die Fälle $y \neq 0$ und $y = 0$ zu unterscheiden. In c) betrachten Sie die Einschränkung von f auf eine Parabel durch $(0, 0)$, die so gewählt ist, dass beide Faktoren in der Definition von f nichtnegativ sind.

4. Den einzigen kritischen Punkt von f findet und klassifiziert man mit Standardmethoden. Um zu sehen, dass er kein globales Extremum ist, bemerken Sie, dass f weder nach oben noch nach unten beschränkt ist.

5. Ist $x \in U$ ein kritischer Punkt von f und $y \in U$ beliebig, so folgt aus der Konvexität und Offenheit von U, dass die Funktion $\varphi\colon t \mapsto f((1-t)x + ty)$ aus dem Hinweis auf dem Aufgabenblatt auf einer offenen Obermenge von $[0, 1]$ definiert. Ausserdem ist φ konvex und differenzierbar, da f diese Eigenschaften hat, und Sie kennen den Wert $\varphi'(0)$. Verwenden Sie all dies und die Eigenschaften von konvexen Funktionen aus Abschnitt 7.2.3 des Skripts, um die Ungleichung $f(y) \geq f(x)$ zu beweisen.

Bitte wenden!

6. Wählen Sie zu $x_0 \in U$ mit $f(x_0) \neq 0$ einen Vektor $v \in \mathbb{R}^n$ mit $\partial_v f(x_0) = f(x_0)$ und folgern Sie dann aus der Definition der Ableitung entlang v, dass für kleine $s > 0$ die Ungleichung $\|f(x_0 + sv)\|_2 > \|f(x_0)\|_2$ gilt. Was passiert im Fall $f(x_0) = 0$?

Alternativ können Sie bemerken, dass man auch das Quadrat der Norm von f betrachten kann, und zeigen, dass die Ableitung der erhaltenen Funktion $x \mapsto \|f(x)\|_2^2$ in einem Punkt $x_0 \in U$ in Richtung $v \in \mathbb{R}^n$ genau $2\langle \partial_v f(x_0), f(x_0)\rangle$ ist. Dann kann Proposition 10.29 verwendet werden um zu zeigen, dass in einem lokalen Maximum x_0 von $x \mapsto \|f(x)\|_2^2$ notwendigerweise $f(x_0) = 0$ gelten muss.

Siehe nächstes Blatt!

Musterlösungen zu ausgewählten Aufgaben

Diese Woche werden Musterlösungen zu den Aufgaben 3, 5 und 6 zur Verfügung gestellt.

3. Wir stellen zuerst fest, dass

$$f(x,y) = -2x^4 + 3x^2y - y^2.$$

a) Der Gradient von f ist gegeben durch

$$\nabla f(x,y) = \begin{pmatrix} -8x^3 + 6xy \\ 3x^2 - 2y \end{pmatrix}.$$

Damit dieser verschwindet, muss aufgrund der zweiten Zeile $y = \frac{3}{2}x^2$ sein, was eingesetzt in der ersten Zeile $x = 0$, und damit auch $y = 0$ ergibt. Der Punkt $(0,0)$ ist also der einzige kritische Punkt von f. Die Hesse-Matrix in diesem Punkt ist die Matrix

$$H(0,0) = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix},$$

welche degeneriert ist (also Determinante 0 besitzt). Somit kann Korollar 10.32 nicht angewendet werden um diesen kritischen Punkt von f zu klassifizieren.

b) Sei $v = (x,y)$. Dann ist

$$\varphi(t) := f(tv) = -2t^4x^4 + 3t^3x^2y - t^2y^2.$$

Ableiten nach t ergibt, dass φ einen kritischen Punkt in $t = 0$ besitzt. Ist $y \neq 0$, so ist $\varphi''(0) = -2y^2 < 0$, sodass $t = 0$ nach Korollar 7.37 ein striktes lokales Maximum von φ ist. Im Fall $y = 0$ muss $x \neq 0$ sein (da v ein Einheitsvektor sein soll), also besitzt auch in diesem Fall $\varphi(t) = -2t^4x^4$ ein striktes lokales Maximum in 0.

c) Wir bemerken, dass der linke Faktor in der Definition von f für $y \leq 2x^2$, und der rechte Faktor für $y \geq x^2$ nichtnegativ ist. Für $y = \frac{3}{2}x^2$ sind also beide Faktoren nichtnegativ. Die Einschränkung von f auf diese Parabel ist gegeben durch

$$t \mapsto f\big(t, \tfrac{3}{2}t^2\big) = \frac{t^4}{4},$$

was in $t = 0$ ein striktes lokales Minimum besitzt. Der Punkt $(0,0)$ ist also kein lokales Maximum von f.

Bitte wenden!

5. Sei $x \in U$ ein kritischer Punkt von f und $y \in U$ beliebig. Aufgrund der Konvexität von U ist dann $(1-t)x + ty \in U$ für alle $t \in [0,1]$. Da U aber auch offen ist, wissen wir sogar, dass dies für alle t in einem offenen Intervall $I \supset [0,1]$ gilt. Dies erlaubt die Betrachtung der Funktion

$$\varphi\colon I \to \mathbb{R},\ t \mapsto f((1-t)x + ty).$$

Als Einschränkung der konvexen Funktion f auf eine gerade Strecke in U ist es naheliegend zu vermuten, dass auch φ konvex ist. Für die formale Verifikation dieser Vermutung seien $s, t \in I$ und $\lambda \in [0,1]$. Dann ist

$$\begin{aligned}\varphi((1-\lambda)s + \lambda t) &= f\big((1 - (1-\lambda)s - \lambda t)x + ((1-\lambda)s + \lambda t)y\big)\\ &= f\big((1-\lambda)((1-s)x + sy) + \lambda((1-t)x + ty)\big)\\ &\leq (1-\lambda)f((1-s)x + sy) + \lambda f((1-t)x + ty)\\ &= (1-\lambda)\varphi(s) + \lambda\varphi(t),\end{aligned}$$

wobei wir $(1-s)x + sy, (1-t)x + ty \in U$ und die Konvexität von f verwendet haben. Des Weiteren ist φ wegen der Kettenregel differenzierbar und es gilt

$$\varphi'(t) = \mathrm{D}_{(1-t)x+ty} f(y - x)$$

für alle $t \in I$. Aus der Konvexität von φ folgt, dass diese Ableitung monoton wachsend ist (Proposition 7.40). Für jedes $t \in [0,1]$ gilt also

$$\varphi'(t) \geq \varphi'(0) = \mathrm{D}_x f(y - x) = 0,$$

wobei letztere Gleichung aus der Annahme folgt, dass x ein kritischer Punkt von f ist. Eine Anwendung des Mittelwertsatzes auf φ produziert sodann ein $\xi \in (0,1)$ mit

$$f(y) - f(x) = \varphi(1) - \varphi(0) = \varphi'(\xi) \geq 0,$$

was $f(y) \geq f(x)$ beweist. Der Punkt x ist also ein globales Minimum von f.

6. Sei $x_0 \in U$ ein lokales Maximum von $x \mapsto \|f(x)\|_2$. Ist $f(x_0) = 0$, so folgt daraus $f(x) = 0$ für alle x in einer Umgebung von x_0 und damit $\mathrm{D}_{x_0} f = 0$. Dies widerspricht aber der angenommenen Surjektivität von $\mathrm{D}_{x_0} f$.

Der Fall $f(x_0) \neq 0$ kann auf verschiedene Arten behandelt werden. Wir präsentieren die beiden im Hinweis angedeuteten Argumente.

Erstes Argument: Aufgrund der Surjektivität von $\mathrm{D}_{x_0} f$ gibt es ein $v \in \mathbb{R}^n$ mit $\partial_v f(x_0) = \mathrm{D}_{x_0} f(v) = f(x_0)$ (vgl. Proposition 10.6). Dies impliziert die Asymptotik

$$f(x_0 + sv) = f(x_0) + s\partial_v f(x_0) + o(s) = (1+s)f(x_0) + o(s)$$

Siehe nächstes Blatt!

für $s \to 0$. Sei nun $\varepsilon > 0$ so gewählt, dass für $s \in (0, \varepsilon)$ sowohl $x_0 + sv \in U$ als auch $\frac{1}{s}\|o(s)\|_2 < \frac{1}{2}\|f(x_0)\|_2$ gilt. Damit erhalten wir unter Verwendung der umgekehrten Dreiecksungleichung

$$\begin{aligned}\|f(x_0 + sv)\|_2 &\geq (1+s)\|f(x_0)\|_2 - \|o(s)\|_2 \\ &= \|f(x_0)\|_2 + s\big(\|f(x_0)\|_2 - \tfrac{1}{s}\|o(s)\|_2\big) \\ &> \|f(x_0)\|_2 + \tfrac{s}{2}\|f(x_0)\|_2 \\ &> \|f(x_0)\|_2\end{aligned}$$

für jedes $s \in (0, \varepsilon)$. Der Punkt x_0 kann also kein lokales Maximum von $x \mapsto \|f(x)\|_2$ sein.

Zweites Argument: Auch die Funktion $x \mapsto \|f(x)\|_2^2 = \langle f(x), f(x)\rangle$ hat ein lokales Maximum in x_0, und diese Funktion ist nun auf ganz U differenzierbar (als Summe der Quadrate der Komponenten von f). Nach Proposition 10.29 verschwindet das Differential dieser Funktion in x_0, und angesichts Proposition 10.6 gilt dies auch für alle Richtungsableitungen. Für $v \in \mathbb{R}^n$ folgt also unter Verwendung der Produktregel für Skalarprodukte[1] (siehe unten)

$$\begin{aligned}0 = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \langle f(x_0 + sv), f(x_0 + sv)\rangle &= \langle \partial_v f(x_0), f(x_0)\rangle + \langle f(x_0), \partial_v f(x_0)\rangle \\ &= 2\langle \partial_v f(x_0), f(x_0)\rangle.\end{aligned}$$

Allerdings erlaubt die Surjektivität von $\mathrm{D}_{x_0} f$ wie im 1. Argument $v \in \mathbb{R}^n$ so zu wählen, dass $\partial_v f(x_0) = f(x_0)$ gilt. Für diesen Vektor v ist obige Ableitung

$$2\langle f(x_0), f(x_0)\rangle = 2\|f(x_0)\|_2^2 \neq 0,$$

ein Widerspruch.

Produktregel für Skalarprodukte: Ist $I \subset \mathbb{R}$ eine nichtleere, offene Teilmenge und sind $g, h\colon I \to \mathbb{R}^m$ differenzierbare Funktionen, so ist auch deren Skalarprodukt $\langle g, h\rangle\colon I \ni s \mapsto \langle g(s), h(s)\rangle$ differenzierbar mit

$$\langle g, h\rangle'(s) = \langle g'(s), h(s)\rangle + \langle g(s), h'(s)\rangle$$

für jedes $s \in I$.

Der Beweis ist eins zu eins derselbe wie für die übliche Produktregel in Proposition 7.5. Zur Begründung der dort durchgeführten Schritte sind dabei nur die Eigenschaften eines Skalarprodukts (insbesondere dessen Stetigkeit) zu verwenden.

[1] Die Verwendung dieser Produktregel kann auch vermieden werden, indem man

$$\langle f(x_0 + sv), f(x_0 + sv)\rangle = \sum_{k=1}^{m} f_k(x_0 + sv)^2$$

schreibt, in $s = 0$ ableitet, und das Ergebnis dann als $2\langle \partial_v f(x_0), f(x_0)\rangle$ identifiziert.