

Aufgaben zum Tut am 14.05.2007

Thomas Pajor

15. Mai 2007

Aufgabe 1.

Beweisen Sie: Sei $f : \mathbb{R}^n \rightarrow \mathbb{R}$ eine differenzierbare Funktion, dann zeigt der negative Gradient $-\nabla f(x_0)$ stets in Richtung des steilsten Abstiegs für beliebiges $x_0 \in \mathbb{R}^n$.

Lösung.

Um diese Aussage zu beweisen, müssen wir uns einem Spezialfall der Kettenregel¹ bedienen. Diese besagt:

Satz (Spezialfall der Kettenregel). Sei $\gamma : \mathbb{R} \rightarrow \mathbb{R}^n$ differenzierbar in $t_0 \in \mathbb{R}$ und $f : \mathbb{R}^n \rightarrow \mathbb{R}$ differenzierbar in $\gamma(t_0)$. Dann ist $f \circ \gamma$ in t_0 differenzierbar und es gilt

$$\begin{aligned}(f \circ \gamma)'(t_0) &= f'(\gamma(t_0)) \cdot \gamma'(t_0) \\ &= (\nabla f)(\gamma(t_0)) \cdot \gamma'(t_0) \\ &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\gamma(t_0)) \gamma'_i(t_0)\end{aligned}$$

Dabei bezeichnet „ \cdot “ das Standardskalarprodukt auf dem \mathbb{R}^n -Vektorraum, und $\gamma'_i(t_0)$ den i -ten Eintrag des n -Vektors der Ableitung von γ (γ bildet ja in den \mathbb{R}^n ab, das heißt die Ableitung von γ ist ein n -dimensionaler Spaltenvektor).

Kehren wir zunächst zum eigentlichen Problem zurück. Wir wollen zeigen, dass $-\nabla f(x_0)$ die Richtung des *steilsten* Abstiegs der Funktion vorgibt. Das bedeutet, dass das Gradientenabstiegs-Optimierungsverfahren gierig in dem Sinne vorgeht, dass es immer den lokal bestmöglichen Optimierungsschritt durchführt. Um diese Aussage zu zeigen, wollen wir beweisen, dass die

¹Man erinnere sich noch dunkel an die HM II Vorlesung!

Richtungsableitung an der Stelle x_0 unter allen Richtungen $v \in \mathbb{R}^n$ ($|v| = 1$) für den negative Gradienten minimal wird.

Halten wir also mal einen beliebigen Punkt $x_0 \in \mathbb{R}^n$ fest, und betrachten alle möglichen Richtungen $v \in \mathbb{R}^n$ mit $|v| = 1$. Die Richtungsableitung von f in Richtung v am Punkt x_0 ist nun gerade

$$\left. \frac{d}{dt} f(x_0 + tv) \right|_{t=0}$$

Anschaulich „variieren“ wir um den Funktionswert $f(x_0)$ in v -Richtung in dem wir uns mit tv um x_0 „herumbewegen“. Für $t = 0$ ergibt das dann gerade die Richtungsableitung im Punkt x_0 . Sei nun $\gamma(t) := x_0 + tv$. Dann ist γ eine Funktion $\gamma: \mathbb{R} \rightarrow \mathbb{R}^n$ mit $t \mapsto x_0 + tv$ und f ist eine Funktion $f: \mathbb{R}^n \rightarrow \mathbb{R}$. Unsere Voraussetzungen entsprechen gerade dem obigen Spezialfall der Kettenregel, und es gilt

$$\begin{aligned} \left. \frac{d}{dt} f(x_0 + tv) \right|_{t=0} &= \left. \frac{d}{dt} (f \circ \gamma)(t) \right|_{t=0} \\ &= (f \circ \gamma)'(0) \\ &\stackrel{\text{KR}}{=} (\nabla f)(\gamma(0)) \cdot \gamma'(0) \\ &= (\nabla f)(x_0) \cdot v \end{aligned}$$

An der Stelle, wo „KR“ steht, kommt die Kettenregel zur Anwendung. Vielleicht außerdem noch eine kurze Erklärung warum $\gamma'(0) = v$ gilt. Es ist

$$\begin{aligned} \gamma'(0) &= \left. \frac{d}{dt} \gamma(t) \right|_{t=0} \\ &= \left. \frac{d}{dt} (x_0 + tv) \right|_{t=0} \\ &= v|_{t=0} \\ &= v \end{aligned}$$

Nun wissen wir, dass die Richtungsableitung in Richtung v am Punkt x_0 gerade durch

$$(\nabla f)(x_0) \cdot v$$

beschrieben wird. Wir suchen jetzt eine Richtung (also ein v), für das dieser Term minimal wird. Es gilt

$$\begin{aligned} (\nabla f)(x_0) \cdot v &= |(\nabla f)(x_0)| \cdot \underbrace{|v|}_{=1} \cdot \cos \omega \\ &= |(\nabla f)(x_0)| \cdot \cos \omega \end{aligned}$$

Dieser Term wird genau dann minimal, wenn $\cos \omega$ minimal wird. Für $\omega = \pi$ ist $\cos \omega = -1$. Dies entspricht einem Winkel von 180° zwischen $(\nabla f)(x_0)$ und der Optimierungsrichtung v . Der negative Gradient $-(\nabla f)(x_0)$ zeigt also genau in die Richtung des steilsten Abstiegs von f an der Stelle x_0 . \square

Aufgabe 2 (Nachklausur 2003).

Sei $n \in \mathbb{N}$. Auf einer Kreisscheibe mit Radius r sollen n Punkte $p_i = (x_i, y_i)$ möglichst gleichmäßig verteilt werden. Hierzu soll ein iteratives Optimierungsverfahren eingesetzt werden.

- (a) Skizzieren Sie eine Bewertungsfunktion $c(q)$, die für gute Konfigurationen q minimale Werte liefert. Erläutern Sie Ihren Entwurf.
- (b) Welches Optimierungsverfahren würden Sie wählen? Genauer: Wie erzeugen Sie aus einer Konfiguration q eine neue Konfiguration q' ?
- (c) Wie könnte eine initiale Konfiguration aussehen?
- (d) Brauchen Sie auch eine Straffunktion $l(q)$? Wenn ja: welche und wozu? Wenn nein: warum nicht?

Lösung.

Eine Konfiguration q in unserem Optimierungsproblem entspricht gerade einer Menge von Punkten (p_1, \dots, p_n) . Ein Übergang von einer Konfiguration q in eine Konfiguration q' entspricht also dem Neusetzen der Koordinaten der jeweiligen Punkte.

- (a) Natürlich gibt es hier kein „Patentrezept“ für die Lösung. Ein möglicher Lösungsansatz wäre der Folgende. Die Punkte werden vermutlich dann eine gute und gleichmäßige Verteilung annehmen, wenn sich keine zwei Punkte zu nahe kommen. Das heißt für alle Punktepaare p_i und p_j für $i \neq j$ soll deren Abstand maximiert werden. Um das als Minimierungsproblem zu formulieren, können wir folgende Funktion als Bewertungsfunktion definieren:

$$c(q) = c(p_1, \dots, p_n) := - \min_{i \neq j} |p_i - p_j|$$

Der Wert der Bewertungsfunktion ist also gerade der Abstand der zwei am nächsten beieinander liegenden Punkte, und da wir ein Minimierungsproblem formulieren wollen, muss dieser Abstand mit -1 multipliziert werden (wir wollen ja, dass die Punkte weit auseinander liegen).

- (b) Es bietet sich ein Gradientenabstiegsverfahren an. Dadurch werden die nächstliegenden Punkte sukzessive auseinandergeschoben. Es gilt also für die Berechnung einer neuen Konfiguration

$$q' = q - (\nabla c)(q) \cdot h$$

wobei wir die Schrittweite h empirisch bestimmen könnten.

- (c) Da der Gradient von c Null wird, wenn zwei Punkte identisch sind, würde das Gradientenabstiegsverfahren in solch einem Falle nicht mehr funktionieren. Dies ist also unbedingt zu

vermeiden; zum Beispiel mit der folgenden spiralförmigen Anfangskonfiguration:

$$p_i := \left(\frac{i}{n} r \cos i, \frac{i}{n} r \sin i \right)$$

Dadurch wird ebenfalls vermieden, dass Punkte kollinear liegen, denn auch dies könnte einer gleichmäßigen Verteilung abträglich sein.

Eine Zufällige Anfangskonfiguration würde, unter Voraussetzung eines gleichverteilten Zufallsgenerators, ebenfalls gut funktionieren.

- (d) Würden wir keine Straffunktion wählen, so würde kein Optimum existieren, denn die Punkte könnten beliebig weit auseinander geschoben werden. Um den Rand der Kreisscheibe zu modellieren lässt sich beispielsweise folgende Straffunktion benutzen

$$l(q) = l(p_1, \dots, p_n) := \sum_{i=1}^n k \cdot \max\{r, |p_i|\}$$

für eine Konstante $k > 1$. Lässt man diese additiv in die Bewertungsfunktion einfließen, also optimiert man nach

$$c'(q) = c(q) + l(q)$$

so würde dies verhindern, dass Punkte den Rand der Kreisscheibe verlassen.