Institut für Analysis und Numerik
Universität Magdeburg
Thomas Richter

12.06.2020

Lösung zu: Übung Nr. 9 zur Algorithmische Mathematik I
Wintersemester 2019/2020

Abgabe bis zum 19. Juni 2020

Das Gradientenverfahren

1. Aus welchen 2 Schritten besteht jedes Abstiegsverfahren?

Zunächst Wahl der Abstiegsrichtung d, dann Wahl der Schrittweite s.

2. Was zeichnet eine *Abstiegsrichtung* aus? Was ist die Richtung des steilsten Abstiegs?

In dieser Richtung d ist es möglich, den Funktionswert zu reduzieren, d.h. es existiert ein $\epsilon > 0$ so dass $f(x + sd) < f(x)$ für $s \in [0, \epsilon]$. Die Richtung des steilsten Abstiegs, ist die Richtung, in der der infitesimale Abstieg am schnellsten ist, also

$$\lim_{\epsilon \to 0} \frac{f(x + \epsilon d) - f(x)}{\epsilon} \to \min$$

3. Wir betrachten die Funktion $f(x, y) = x^2 - 3yx^2$ und hier den Punkt $(1, 1)$. Ist die Richtung $(1, 1)$ eine Abstiegsrichtung? In welche Richtung ist im Punkt $(1, 1)$ der steilste Abstieg?

Es ist

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

also in $(x, y) = (1, 1)$ gilt

$$\nabla f(1, 1) = \begin{pmatrix} -4 \\ -3 \end{pmatrix}.$$

es ist also

$$\langle (-4, -3), (1, 1) \rangle = -7 < 0,$$

d.h. es liegt eine Abstiegsrichtung vor. Die Richtung des steilsten Abstiegs wäre gerade $-\nabla f = (4, 3)$.

Schrittweitenwahl

1. Warum beginnt man bei der Line-Search Methode mit $s = 1$ und reduziert die Schrittweite dann? Genauer gesagt: warum glaubt man, so irgendwann eine Reduktion zu erreichen?

Abstiegsrichtung bedeutet erstmal nur, dass es ein unter umständen sehr kleies Intervall $(0, \epsilon]$ gibt, so dass für $0 < s \leqslant \epsilon$ dann $f(x + sd) < f(d)$ gilt. Es ist also sinnvoll, die Schrittweite immer stärker zu reduzieren.

2. Was ist der Vorteil der Armijo-Regel gegenüber des einfachen Line-Search?

Line-Search garantiert einen Abstieg, also $f(x+sd) < f(x)$. Die Armijo-Regel liefert darüber hinaus noch eine Quantifizierung der Art $f(x + sd) < f(x) - s\gamma\|d\|^2$. Dies kann z.B. im Konvergenzbeweis des Gradientenverfahrens zum Nachweis der Konvergenz verwendet werden.

Konvergenz des Gradientenverfahrens

1. Die Aussage von Satz 8.18 ist recht kompliziert. Können wir hieraus in jedem Fall Konvergenz gegen ein Minimum oder Abbruch durch Finden eines stationären Punktes garantieren?

Falls das Verfahren nicht abbricht, so wird in jedem Fall eine monoton fallende Folge $f(x_k)$ garantiert.

Ist die Funktion $f(x)$ nicht nach unten beschränkt, so wissen wir auch nicht mehr über die Folge.

Ist die Funktion beschränkt, so ist auch die monotone Folge $f(x_k)$ beschränkt hat also mindestens einen Häufungspunkt. Falls es nur einen Häufungspunkt gibt, so folgt also Konvergenz gegen einen stationären Punkt. Dies kann z.B. auch ein Sattelpunkt sein, z.B. bei Anwendung auf $f(x) = x^3/10$ mit $x_0 = 1$.

2. Versuchen Sie ein Szenario zu beschreiben, in dem *das Gradientenverfahren nicht mit einem stationären Punkt abbricht, die Funktion nach unten beschränkt ist, das Gradientenverfahren aber nicht gegen einen stationären Punkt konvergiert, sondern diesen nur als Häufungspunkt hat.*

Aufgabe 9.1:

Wir betrachten das Gradientenverfahren mit fester Schrittweite $s \in \mathbb{R}$ zur Minimierung von

$$f : \mathbb{R}^2 \to \mathbb{R}, \quad f(x) = \|x\|_2^2 = x_1^2 + x_2^2.$$

a) Wie lautet $\nabla f(x)$

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix} = 2x$$

b) Es sei $x^0 = (\frac{1}{2}, 0)^T$. Man berechne den negativen Gradienten $-\nabla f(x^0)$ von f in x^0. Wo liegt die nächste Iteration x^1 bei Wahl der Schrittweite $s = 1$?

Hier ist

$$-\nabla f(x_0) = -2x_0,$$

Somit gilt

$$x_1 = x_0 - 2x_0 = -x_0$$

c) Bestimmen Sie bei Wahl von $s = 1$ die Iterierten x^1, x^2, x^3.

Die Iterierten alternieren stets, also

$$x^1 = -x^0, \quad x^2 = x^0, \quad \dots$$

Programmieraufgabe 9.2:

Wir untersuchen das Gradientenverfahren zur Suche des Minimums der Rosenbrock-Funktion

$$f(x, y) = 100(y - x^2)^2 + (1 - x)^2$$

a) Man implementiere das Gradientenverfahren mit vorgegebener Schrittweite, d.h., man implementiere eine Funktion `def gradientenverfahren(x,step,tol,maxiter)`, die ausgehend vom Startvektor `x` das Gradientenverfahren durchführt und jeweils die feste Schrittweite `step` nutzt, d.h.

$$x_{k+1} = x_k - \texttt{step}\nabla f(x_k).$$

Die Iteration soll abbrechen, wenn die vorgegebene Toleranz `tol` erreicht wird, d.h.

$$\|\nabla f(x_k)\| < \texttt{tol},$$

oder wenn die maximale Anzahl an Schritten erreicht wird. Man versuche ausgehend vom Startvektor `x=np.array([1,-1])` ein Residuum von `tol=1.e-6` zu erreichen. Man experimentiere mit der geeigneten Wahl der Schrittweite `step` und wähle hinreichend viele Schritte `maxiter`.
Man dokumentiere das gefundene Minimum, die Anzahl der notwendigen Schritte und das hier erreichte Residuum.

b) Man wiederhole Aufgabe **a**) aber mit Line-Search zur Schrittweitenwahl. Hierfür wähle man $\beta = \frac{1}{2}$.

c) Man wiederhole Aufgabe **a**) aber mit der Armijo-Regel zur Schrittweitenwahl. Hierfür teste man z.B. $\beta = \frac{1}{2}$ und $\gamma = 0.01$, versuche aber noch bessere Parameter zu finden.

d) Man stelle den Konvergenzverlauf für alle drei Varianten (jeweils mit Parametern so dass Konvergenz erreicht wird) graphisch dar. Hierfür wähle man eine doppelt-Logarithmische Darstellung.