

## Steilster Abstieg

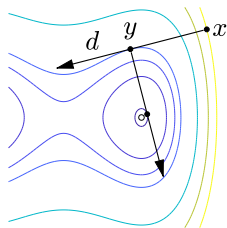
Die Methode des steilsten Abstiegs dient zur Minimierung multivariater Funktionen  $f(x_1, \dots, x_n)$ . Zur Durchführung eines Iterationsschritts  $x \rightarrow y$  wird zunächst der negative Gradient

$$d = -\text{grad } f(x)$$

als lokal beste Abstiegsrichtung berechnet. Dann bestimmt man  $y$  als eine Minimalstelle von  $f$  in Richtung von  $d$ :

$$f(y) = \min_{t \geq 0} f(x + td).$$

Wie in der Abbildung illustriert, ist die Suchrichtung orthogonal zu der Niveaumenge durch  $x$  und berührt eine Niveaumenge zu einem kleineren Funktionswert in  $y$ .



Die Konvergenz der durch die Methode des steilsten Abstiegs erzeugten Folge  $x_0, x_1, \dots$  kann unter sehr allgemeinen Voraussetzungen gezeigt werden. Hinreichend ist, dass  $f$  nach unten beschränkt ist und  $\text{grad } f$  in einer Umgebung  $U$  der Menge  $\{x : f(x) \leq f(x_0)\}$  Lipschitz-stetig ist, d.h.

$$\|\text{grad } f(x) - \text{grad } f(\tilde{x})\| \leq L\|x - \tilde{x}\|, \quad x, \tilde{x} \in U.$$

Dann gilt

$$\sum_{\ell=0}^{\infty} \|\text{grad } f(x_\ell)\|^2 < \infty;$$

insbesondere ist  $\|\text{grad } f(x_\ell)\|$  eine Nullfolge. Dies impliziert, dass jeder Häufungspunkt  $x_*$  der Folge  $x_0, x_1, \dots$  ein kritischer Punkt von  $f$  ist, d.h.  $\text{grad } f(x_*) = (0, \dots, 0)^t$ . Dass es sich um ein lokales Minimum handelt ist statistisch gesehen fast sicher, kann jedoch nicht zwingend gefolgert werden.

In dem Algorithmus braucht die eindimensionale Minimierung nur näherungsweise durchgeführt werden. Die Suchrichtung  $d$  muss nicht als der negative Gradient gewählt, und eine globale Minimalstelle  $y$  nicht

bestimmt werden. Entscheidend für die Konvergenz ist lediglich, dass in jedem Iterationsschritt eine Reduktion des Funktionswertes proportional zu  $\|\text{grad } f(x)\|^2$  erreicht wird.

---

## Beispiel

---

Steilster Abstieg für eine quadratische Funktion

$$f(x) = \frac{1}{2}x^tAx - b^tx$$

mit einer symmetrischen positiv definiten Matrix  $A$

---

Iterationsschritt  $x \rightarrow y = x + td$  mit

$$d = -\text{grad } f(x) = b - Ax$$

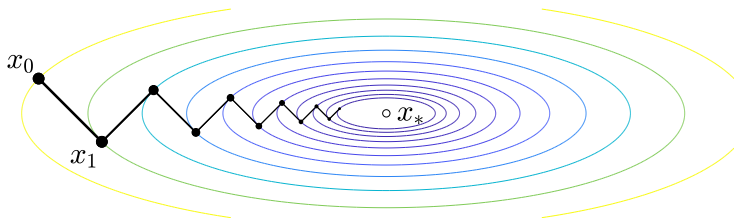
und  $t$  der Minimalstelle der univariaten Funktion

$$\begin{aligned} f(x + td) &= \frac{1}{2}(x + td)^tA(x + td) - b^t(x + td) \\ &= \frac{1}{2}d^tAd t^2 + (x^tAd - b^td) t + \frac{1}{2}(x^tAx - 2b^tx) \end{aligned}$$

Nullsetzen der Ableitung nach  $t \rightsquigarrow$  Formel für den  
Halbgeradenparameter  $t$

$$0 = d^t A d t - (b - Ax)^t d = d^t A d t - d^t d \quad \Longrightarrow \quad t = \frac{d^t d}{d^t A d},$$

d.h. man erhält einen expliziten Ausdruck für  $y$   
unerwünschte Oszillationen bei Eigenwerten stark unterschiedlicher  
Größenordnung von  $A$



konkretes Beispiel

$$A = \begin{pmatrix} 1 & 0 \\ 0 & 100 \end{pmatrix}, \quad b = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$x = (c, 1)^t \rightsquigarrow$$

$$d = -(Ax - b) = - \begin{pmatrix} c \\ 100 \end{pmatrix}, \quad Ad = - \begin{pmatrix} c \\ 10^4 \end{pmatrix}$$

und

$$d^t d = c^2 + 10^4, \quad d^t Ad = c^2 + 10^6, \quad t = \frac{d^t d}{d^t Ad} = \frac{c^2 + 10^4}{c^2 + 10^6}$$

sowie

$$y = x + td = \begin{pmatrix} c \\ 1 \end{pmatrix} - \frac{c^2 + 10^4}{c^2 + 10^6} \begin{pmatrix} c \\ 100 \end{pmatrix} = \frac{99c^2}{c^2 + 10^6} \begin{pmatrix} 10^4/c \\ -1 \end{pmatrix}$$

$c = 100 \rightsquigarrow$  Verbesserung um weniger als 1%:

$$y = \frac{99}{101} \begin{pmatrix} x_1 \\ -x_2 \end{pmatrix}$$