10 Das Gradientenverfahren

Ziel: Untersuchung von Vorschriften für die Richtungswahl. Einfachstes Verfahren zur Lösung von (P) ist das Gradientenverfahren: $p^k := -\nabla f(x^k)$.
Offensichtlich gilt dann: $\nabla f(x^k)^T p^k = -\|\nabla f(x^k)\|^2 < 0$.

Das Gradientenverfahren mit semi-effizienter Schrittweitenregel lautet:

Algorithmus 10.1

0) Wähle eine semi-effiziente Schrittweitenregel und ein $x^0 \in \mathbb{R}^n$. Setze $k := 0$.
1) Falls $\nabla f(x^k) = 0$: STOP! x^k ist kritische Lösung von Problem (P)
2) Setze $p^k := -\nabla f(x^k)$.
3) Bestimme eine Schrittweite $t_k > 0$ gemäß Schrittweitenregel und $x^{k+1} := x^k + t_k \cdot p^k$.
4) Setze $k := k+1$ und gehe zu 1).

Im Fall von einer beliebigen semi-effizienten Schrittweitenregel sind mit $H_k := I$ und $m := M := 1$ die Voraussetzungen von Satz 8.17 erfüllt, so dass aus diesem Satz diese Konvergenzaussage folgt:

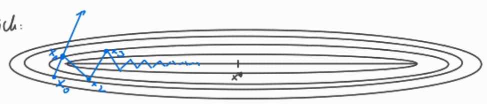
Satz 10.2

Es gelten (V1)-(V3). Wenn Algorithmus 10.1 nicht nach endlich vielen Schritten ab, so erzeugt er eine unendliche Folge $(x^k)_k$ mit

i) Jeder Häufungspunkt von $(x^k)_k$ ist kritische Lösung von (P)
ii) Besitzt (P) genau eine kritische Lösung x^* in N_0, so ist $\lim_{k\to\infty} x^k = x^*$.
iii) Ist auch (V4) erfüllt und x^* die dann existierende Lösung von (P), so folgt $\lim_{k\to\infty} x^k = x^*$ und mit der Semieffizienzkonstante $\vartheta > 0$ aus der Schrittweitenregel und einem $c > 0$ gilt: $0 \le f(x^{k+1}) - f(x^*) \le (1 - 2\beta\vartheta) \cdot [f(x^k) - f(x^*)]$, $k \in \mathbb{N}_0$
sowie: $\|x^k - x^*\| \le c \cdot (\sqrt{1 - 2\beta\vartheta})^k$, $k \in \mathbb{N}_0$

Das Gradientenverfahren konvergiert für jede Schrittweitenregel aus Kapitel 8 mit einer Teilfolge gegen eine kritische Lösung (wenn (V1)-(V3) erfüllt). Ist auch noch (V4) erfüllt so konvergiert die gesamte Iteriertenfolge und zwar (mindestens) R-linear sowie die Folge der Funktionswerte Q-linear.
Langsame Konvergenz kann nicht mehr ausgeschlossen werden, wenn $2\beta\vartheta$ sehr klein ist. Die Konstante ϑ hängt dabei von der jeweiligen Schrittweitenregel ab. Für die exakten Schrittweiten (Minimum-, Curry-Schrittweite) gilt nach Satz 8.3 $\vartheta := \vartheta_M = \vartheta_C = \frac{1}{2\gamma}$. Für die Schrittweiten kann langsame Konvergenz eintreten, falls $\frac{\beta}{\gamma}$ klein ist.

Untersuche das Gradientenverfahren mit exakter Schrittweite für den Fall einer quadratischen Funktion $f(x) := \frac{1}{2}x^T Q x + c^T x + \kappa$, $x \in \mathbb{R}^n$ mit Q positiv definit und symmetrisch.
Dann hat (P) wegen $\nabla f(x^*) = 0 \Leftrightarrow Qx^* + c = 0$ die eindeutige Lösung $x^* := -Q^{-1} \cdot c$. Aus Beispiel 8.7 wissen wir, dass $\text{cond}(Q) = \frac{\lambda_{max}(Q)}{\lambda_{min}(Q)} = \frac{\gamma}{\beta}$. Mit der Beziehung $\vartheta = \vartheta_M = \vartheta_C = \frac{1}{2\gamma}$ kann man schreiben: $1 - 2\beta\vartheta = 1 - 2\beta\frac{1}{2\gamma} = 1 - \frac{\beta}{\gamma} = 1 - \frac{1}{\text{cond}(Q)} = \frac{\text{cond}(Q)-1}{\text{cond}(Q)}$
Durch Einsetzen in die beiden Ungleichungen von Satz 10.2 iii) erhält man: $0 \le f(x^{k+1}) - f(x^*) \le \left(\frac{\text{cond}(Q)-1}{\text{cond}(Q)}\right)(f(x^k) - f(x^*))$, $k \in \mathbb{N}_0$ und $\|x^k - x^*\| \le c \cdot \left(\sqrt{\frac{\text{cond}(Q)-1}{\text{cond}(Q)}}\right)^k$, $k \in \mathbb{N}_0$.
Die Abschätzung beschreiben zwar nur ein worst-case-Verhalten, dies entspricht jedoch durchaus dem realen Verhalten: die Folge der Funktionswerte $(f(x^k))_k$ und der Iterierten $(x^k)_k$ konvergiert umso langsamer, je größer die Konditionszahl von Q ist.
Diese Konvergenzaussagen gelten nicht nur für quadratische Funktionen, sondern qualitativ auch für lokale Minimalpunkte x^* beliebiger $C^2(\mathbb{R}^n)$-Funktionen f, welche die hinreichende Optimalitätsbedingung 2. Ordnung aus Satz 7.15 erfüllen. In diesem kann f lokal um x^* durch ein quadratisches Taylorpolynom $g^*(x) = f(x^*) + \nabla f(x^*) \cdot (x - x^*) + (x - x^*)^T \nabla^2 f(x) \cdot (x - x^*)$ mit positiv definiter, symmetrischer Matrix $\nabla^2 f(x)$ angenähert werden. Langsame Konvergenz für das Gradientenverfahren ist dann für f zu erwarten, wenn die Kondition der Hesse-Matrix $\nabla^2 f(x^*)$ groß ist.
Anschaulich:



Zick-Zack-Kurs der Iterierten.

Beispiel 10.3

Wenden Gradientenverfahren mit exakter Schrittweite an auf die quadratische Funktion $f(x) = \frac{1}{2}x^T Q x$ mit $Q := \begin{pmatrix} 2 & 0 \\ 0 & 2 \cdot 10^3 \end{pmatrix}$ und $x \in \mathbb{R}^2$.
Problem (P) hat die Lösung $x^* = (0,0)$ mit Minimalwert $f(x^*) = 0$. Die Konditionszahl von Q ist $\text{cond}(Q) = 10^3$. Für die obige Konstante in den Ableitungen erhält man $\frac{\text{cond}(Q)-1}{\text{cond}(Q)} = 0{,}999$ was auf eine möglicherweise langsame Konvergenz hinweist. Die Abstiegsrichtung in x ist $p := -\nabla f(x) = -Qx$, woraus sich gemäß Beispiel 9.5 die Minimumschrittweite ergibt: $t_M = -\frac{\nabla f(x)^T p}{p^T Q p} = \frac{-\nabla f(x)^T \nabla f(x)}{-\nabla f(x)^T Q(-\nabla f(x))} = \frac{(x^T Q)(Qx)}{(x^T Q) Q (Qx)} = \frac{x^T Q^2 x}{x^T Q^3 x}$.
Es ist $Q^2 = \begin{pmatrix} 4 & 0 \\ 0 & 4 \cdot 10^6 \end{pmatrix}$ und $Q^3 = \begin{pmatrix} 8 & 0 \\ 0 & 8 \cdot 10^9 \end{pmatrix}$, also $t_M = \frac{x_1^2 + 10^6 \cdot x_2^2}{2(x_1^2 \cdot 10^9 x_2^2)}$. Setze $t_k := t_M(x^k, p^k)$ und $x^k := (x_1^k, x_2^k)$ gilt dann: $x^{k+1} = x^k - t_k \cdot Q \cdot x^k = (I - t_k Q)x^k = \begin{pmatrix} 1 - z_k & 0 \\ 0 & 1 - 10^3 z_k \end{pmatrix} \cdot x^k$
mit $z_k := 2t_k := \frac{x_1^2 + 10^6 \cdot x_2^2}{(x_1^2 \cdot 10^9 x_2^2)}$.
Wählt man als Startpunkt $x^0 = (1, 10^{-3})^T$, so ist $z_0 = 2t_0 = \frac{2}{1 + 10^3} = \frac{2}{1001} \approx 2 \cdot 10^{-3}$. Damit dann $(x_1^1, x_2^1) \approx (x_1^0, -x_2^0)$. Damit ist $z_1 \approx z_0$ somit $(x_1^2, x_2^2) \approx (x_1^0, x_2^0)$. Der Fortschritt dürfte somit gering ausfallen.

konkret:

k	x_1^k	x_2^k	$f(x_1^k, x_2^k)$	z_k
0	1 000 000	0,001000	1,001000	0,001998
1	0,998002	-0,000998	0,997004	0,001998
2	0,996008	0,000996	0,993024	0,001998
		⋮		

Startet man mit $x^0 = (0,1)^T$ und $f(x^0) = 1000$, so ist $t_0 = 10^{-3}$ und somit $x^1 = (0,0)^T = x^*$. Dito für $x^0 = (1,0)^T$ und $f(x^0) = (1)$, so ist $t_0 = 1$ und ebenso $x^1 = (0,0)^T = x^*$.
Die Lösung wird mit einer Iteration des Verfahrens erreicht. Dies passiert beim Gradientenverfahren mit der Minimumschrittweite und einer quadratischen Funktion f mit positiv definiter Matrix Q immer dann, wenn x^0 so gewält wurde, dass $x^0 = v - Q^{-1}$, wobei v Eigenvektor zu einem Eigenwert $\lambda > 0$ von Q ist, d.h. $Qv = \lambda v$. Denn bei einer solchen Wahl von x^0 ist: $\nabla f(x^*) = Qx^0 + c = Q(v - Q^{-1}c) + c = Qv = \lambda v$ und $t_0 = \frac{\nabla f(x^0)^T \nabla f(x^0)}{\nabla f(x^0)^T Q \nabla f(x^0)} = \frac{\lambda^2 v^T v}{\lambda^2 v^T Q v} = \frac{1}{\lambda}$. Es folgt $x^1 := x^0 + t_0(-\nabla f(x^0)) = x^0 - \frac{1}{\lambda}(\lambda \cdot v) = x^0 - v = (v - Q^{-1}c) - v = -Q^{-1} \cdot c = x^*$.

In der Praxis sind Gradientenverfahren am Anfang recht schnell, werden in der Nähe der Lösung langsam. Bis 1960 gab es aber keine guten Alternativen dazu.

11 Verfahren der konjugierten Richtungen / konjugierte Gradienten (CG-Verfahren)

CG-Verfahren ("conjugate gradients") konvergieren in der Praxis oftmals schneller als Gradientenverfahren, aber langsamer als die globalisierten Newton-Verfahren bzw. Quasi-Newton-Verfahren (Kap. 12/13)

→ wir skippen das Kapitel aus Zeitgründen