Grundlagen der Optimierung

Das Subgradientenverfahren

18 Das Subgradientenverfahren

Literatur: [Alt, 2004, Kapitel 4]

Wir betrachten noch ein einfaches Verfahren für die *unbeschränkte* Minimierungsaufgabe

$$\text{Minimiere} \quad f(x) \quad \text{über } x \in \mathbb{R}^n \tag{18.1}$$

mit konvexer Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$.

Idee: Verallgemeinere das Gradientenverfahren (Algorithmus 6.1) und benutze als Suchrichtung jeweils ein Element aus dem Subdifferential $\partial f(x_k)$.

Algorithmus 18.1 **(Subgradientenverfahren)**

1: Wähle $x_0 \in \mathbb{R}^n$ und setze $k := 0$
2: **while** Abbruchkriterium nicht erfüllt **do**
3: Wähle einen Subgradienten $s_k \in \partial f(x_k)$ und setze $d_k := -\frac{s_k}{\|s_k\|}$
4: Bestimme eine Schrittweite $t_k > 0$
5: Setze $x_{k+1} := x_k + t_k d_k$ und $k := k + 1$
6: **end while**

Bemerkung 18.2 (a) Die Suchrichtung d_k muss keine Abstiegsrichtung sein! Daher kann man die Schrittweite t_k nicht mit Hilfe der Armijo-Regel bestimmen. Das Subgradientenverfahren ist kein Abstiegsverfahren, d.h., $\{f(x_k)\}$ ist i.A. nicht monoton fallend.

(b) Eine Routine zur Bestimmung eines Subgradienten $s_k \in \partial f(x_k)$ nennt man auch ein **Orakel**, da man keine Kontrolle hat, welcher Subgradient zurückgeliefert wird.

Wir untersuchen jetzt Bedingungen an die Wahl der Schrittweite t_k und betrachten einen Schritt des Subgradientenverfahrens (Schritt 3–5 in Algorithmus 18.1).

Lemma 18.3

Sei x^* ein (globales) Minimum von (18.1). Ist x_k kein Minimum von (18.1), dann gibt es ein $T_k > 0$ mit der Eigenschaft:

$$\|x_{k+1} - x^*\| < \|x_k - x^*\| \quad \text{für alle } t \in (0, T_k). \tag{18.2}$$

Beweis: Für beliebiges $t_k > 0$ gilt

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \left\| x_k - t_k \frac{s_k}{\|s_k\|} - x^* \right\|^2 \\
&= \|x_k - x^*\|^2 - 2\, t_k \underbrace{\left(\frac{s_k}{\|s_k\|}\right)^\top (x_k - x^*)}_{=: b_k} + t_k^2 \underbrace{\left(\frac{s_k}{\|s_k\|}\right)^\top \frac{s_k}{\|s_k\|}}_{=1} \\
&= \|x_k - x^*\|^2 - 2\, t_k\, b_k + t_k^2. \qquad (*)
\end{aligned}
$$

Da x_k nicht optimal ist, gilt $f(x^*) < f(x_k)$. Wegen $s_k \in \partial f(x_k)$ folgt aus der Subgradientenungleichung (15.1)

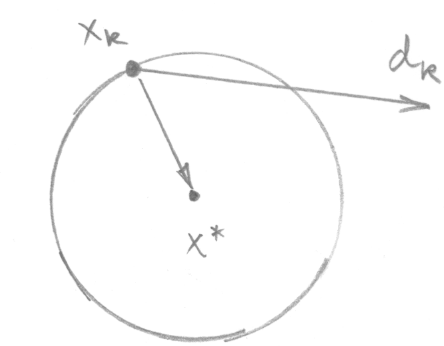
$$0 > f(x^*) - f(x_k) \ge s_k^\top (x^* - x_k) = -b_k\, \|s_k\| \quad \Rightarrow \quad b_k > 0.$$

Zurück zu $(*)$:

$$
\begin{aligned}
\|x_{k+1} - x^*\|^2 &= \|x_k - x^*\|^2 - 2\, t_k \underbrace{b_k}_{>0} + t_k^2 \\
&< \|x_k - x^*\|^2 \quad \text{für alle } 0 < t_k < T_k := 2\, b_k.
\end{aligned}
$$

□

Die Beobachtung $b_k > 0$ lässt sich wie folgt interpretieren: Der Winkel zwischen der Suchrichtung d_k und der Idealrichtung $x^* - x_k$ ist kleiner als 90°.



Da man b_k nicht kennt, liefert Lemma 18.3 nur die abstrakte Bedingung, dass die Schrittweite t_k hinreichend klein sein muss, um (18.2) sicherzustellen.

Um Konvergenz des Subgradientenverfahrens zu erreichen, verlangt man

$$\lim_{k\to\infty} t_k = 0 \quad \text{und} \quad \sum_{k=0}^{\infty} t_k = \infty, \tag{18.3}$$

also zum Beispiel $t_k = 1/(k+1)$ für $k \in \mathbb{N}_0$.

Beachte: Man kann also die Schrittweiten vor dem Start des Verfahrens festlegen! Die zweite Bedingung braucht man, da alle Iterierten in der Kugel $\overline{B_r(x_0)}$ mit Radius $r := \sum_{k=0}^{\infty} t_k$ liegen, denn:

$$\|x_0 - x_k\| = \Big\| \sum_{i=0}^{k-1} x_i - x_{i+1} \Big\| \le \sum_{i=0}^{k-1} \|x_i - x_{i+1}\| \overset{\|d_i\|=1}{=} \sum_{i=0}^{k-1} t_i \le r.$$

Da die Folge $\{f(x_k)\}$ i.A. nicht monoton fallend ist, tritt an ihre Stelle die „Folge der bisher besten Funktionswerte“:

$$\overline{f}_k := \min\{f(x_i) : 0 \le i \le k\}.$$

Satz 18.4 (Globaler Konvergenzsatz für das Subgradientenverfahren) Erfüllen die Schrittweiten t_k die Bedingung (18.3) und stoppt das Subgradientenverfahren nicht nach endlich vielen Schritten, dann gilt

$$\lim_{k\to\infty} \overline{f}_k = f^* := \inf_{x\in\mathbb{R}^n} f(x),$$

wobei f^* auch $-\infty$, d.h. (18.1) unbeschränkt sein kann.

Beweis: [Alt, 2004, Satz 4.2.5] □

Gründe für einen Abbruch nach endlich vielen Schritten können sein:

- $\partial f(x_k) = \emptyset$ (wenn f nicht auf ganz $\mathbb{R}^n$ definiert ist)
- $s_k = 0$, d.h., x_k ist ein Minimum von f.

Bemerkung 18.5 (a) Das Subgradientenverfahren kann zur Lösung der Aufgabe (17.1), also

$$\text{Minimiere} \quad f(x) \quad \text{über } x \in C \tag{18.4}$$

mit konvexer Grundmenge $C \subset \mathbb{R}^n$ verallgemeinert werden. Dazu ist Schritt 5 von Algorithmus 18.1 zu ersetzen durch

$$x_{k+1} := \operatorname{proj}_C \big(x_k + t_k\, d_k\big),$$

wobei $\operatorname{proj}_C(\cdot)$ die orthogonale Projektion auf die Menge C bezeichnet.

(b) Das Subgradientenverfahren konvergiert in der Praxis langsam, weil ein einzelner Subgradient s_k nur wenig Information über das lokale Verhalten der Funktion f in der Nähe von x_k enthält. Eine Verbesserung sind sogenannte Bundle-Verfahren, die mit approximativen Abstiegsrichtungen arbeiten, siehe [Alt, 2004, Kapitel 6-7].

Literatur

W. Alt. *Numerische Verfahren der konvexen, nichtglatten Optimierung.* Teubner, Stuttgart, 2004.