9 Schrittweitenregel

Generalvoraussetzungen (V1) - (V3). Betrachte Punkt $x \in N_0$ und Abstiegsrichtung p in x.

Gesucht: Schrittweite.

Insbesondere ist $\nabla f(x) \neq 0$ und x kein kritischer Punkt.

9.1 Exakte Schrittweite

9.1.1 Existenz und Effizienz

Naheliegende Schrittweitenregel ist die Wahl eines globalen Minimalpunkts des 1-d-Optimierungsproblem $\inf_{t \in [0,\infty)} f(x+tp)$ als Schrittweite. Das liefert den größten Fortschritt beim Reduzieren von f bei x in Richtung p.

Def 9.1 Jedes $t_M := t_M(x,p)$ mit $f(x+t_M p) = \inf_{t\in[0,\infty)} f(x+tp)$ heißt **Minimumschrittweite**.

Diese Schrittweite kann de facto nur für sehr einfache Funktionen genau bestimmt werden. In der Praxis wähle daher den kleinsten, positiven kritischen Punkt von $\varphi(t) := f(x+tp)$ als Schrittweite.

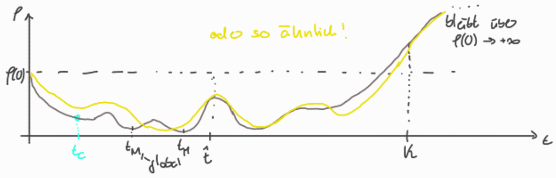
Def 9.2 Die **Curry-Schrittweite** $t_C := t_C(x,p)$ ist definiert durch $t_C(x,p) := \inf\{ t \in [0,\infty) \mid \frac{d}{dt} f(x+tp) \mid_{t=\bar{t}} = 0 \}$.

Minimum- und Curry-Schrittweite werden auch als **exakte Schrittweiten** bezeichnet.

Satz 9.3 Es gelten (V1) - (V3). Für Paare $x \in N_0$, $p \in \mathbb{R}^n$ mit $\nabla f(x)^T p < 0$ existieren eine Minimum- und die Curry-Schrittweite und diese sind positive Zahlen. Ferner sind sie effiziente Schrittweiten mit $\vartheta_M^f := \vartheta_C^f := \frac{1}{2\gamma}$.

Beweis:

Existenz: setze $\varphi(t) := f(x+tp)$. Nach Lemma 8.8 ist $\psi(t) = \varphi(0) - \varphi(t)$ auf $(0,\hat{t})$ für ein $\hat{t} := \hat{t}(x,p) > 0$, also $\varphi(0) > \varphi(t)$ $\forall t \in (0,\hat{t})$. Ferner sagt das Lemma, dass es ein $K > \hat{t}$ gibt, sodass $\varphi(0) < \varphi(t)$ $\forall t \geq K$



TODO

Somit existiert ein $t_M \in (0,K)$ mit $\varphi(t_M) = \min_{t\in[0,K]} \varphi(t) = \min_{t\in[0,\infty)} \varphi(t)$. Die Menge aller kritischen Punkte von φ in $[0,K]$, d.h. $K_\varphi := \{ t \in [0,K] \mid \varphi'(t) = 0 \}$ enthält t_M und ist somit nicht leer. Ferner ist sie kompakt (Urbilder von kompakten Mengen unter stetigen Abbildungen sind kompakt), sodass ein kleinster kritischer Punkt $t_C := \min_{t\in K_\varphi} t$ in $[0,K]$ existiert.

Aus $\varphi'(0) = \nabla f(x)^T p < 0$ folgt $t_C > 0$. Da $\varphi(t_C) < \varphi(0)$, also ist $f(x+t_C \cdot p) < f(x)$, ist $(x+t_C\cdot p) \in N_0$.

Effizienz: Weisen für jedes t_M und t_C die Effizienzungleichung nach: $f(x) - f(x+tp) \geq \vartheta \left(\frac{\nabla f(x)^T p}{\|p\|}\right)^2$

Es gilt:

$$0 = \varphi'(t) = \nabla f(x+tp)^T \cdot p = \nabla f(x)^T p + \nabla f(x+tp)^T \cdot p - \nabla f(x)^T \cdot p \overset{C.S.}{\leq} \nabla f(x)^T \cdot p + \|\nabla f(x+tp) - \nabla f(x)\| \cdot \|p\| \overset{(V3)}{\leq} \nabla f(x)^T \cdot p + \gamma \cdot t_C \cdot \|p\|^2$$

$$\Rightarrow -\nabla f(x)^T \cdot p \leq t_C \cdot \|p\| \cdot \gamma \quad \Rightarrow t_C \geq -\frac{1}{\gamma} \cdot \frac{\nabla f(x)^T \cdot p}{\|p\|^2}.$$

Nach Lemma 8.11 ist $\hat{t} \geq -\frac{2}{\gamma} \cdot \frac{\nabla f(x)^T \cdot p}{\|p\|^2} = \tilde{t} > 0$, also $t_C \geq \frac{\tilde{t}}{2} > 0$.

Es ist $\varphi'(0) = \nabla f(x)^T \cdot p < 0$ und t_C ist die kleinste positive Nullstelle von φ', also ist $\varphi'(t) < 0$ für alle $t \in [0, t_C)$, d.h. φ ist streng monoton fallend auf $[0, t_C)$. Somit ist $\varphi(\frac{\tilde{t}}{2}) \geq \varphi(t_C)$. Also durch Auswerten von φ:

$$f(x+\tfrac{\tilde{t}}{2} \cdot p) \geq f(x + t_C \cdot p) \geq \min_{t\in[0,\infty)} f(x+tp) = f(x+t_M \cdot p)$$

Daraus folgt: $f(x) - f(x+t_M \cdot p) \geq f(x) - f(x - t_C \cdot p) \geq f(x) - f(x+\frac{\tilde{t}}{2} \cdot p)$

$$= \ldots = \psi(\tfrac{\tilde{t}}{2} \cdot p) \geq -\tfrac{\tilde{t}}{2} \nabla f(x)^T \cdot p - (\tfrac{\tilde{t}}{2})^2 \tfrac{\gamma}{2} \|p\|^2 = \frac{1}{\gamma} \cdot \left(\frac{\nabla f(x)^T \cdot p}{\|p\|^2}\right)^2 - \frac{1}{2\gamma}\left(\frac{\nabla f(x)^T \cdot p}{\|p\|^2}\right)^2 = \underbrace{\frac{1}{2\gamma}}_{\vartheta_M,\, \vartheta_C} \left(\frac{\nabla f(x)^T \cdot p}{\|p\|^2}\right)^2$$

□

Ist f konvex, so ist auch $\varphi(t) := f(x+t\cdot p)$ konvex.

Bem 9.4 Sei $f: \mathbb{R}^n \to \mathbb{R}^n$ konvex. Gelten (V1) - (V3), so ist $t_C := t_C(x,p)$ auch eine Minimumschrittweite, die kleinste (kürzeste) unter allen. Ist f außerdem strikt konvex, so existiert genau eine Minimum-Schrittweite $t_M := t_M(x,p)$ und es ist $t_M = t_C$.

Eine gleichmäßig konvexe, quadratische Funktion erfüllt (V1)-(V3) (siehe Bsp. 8.7), also ist die Minimumschrittweite eindeutig. Man kann sie explizit ausrechnen.

Beispiel 9.5

Für $f(x) := \frac{1}{2}x^T Q \cdot x^T + c^T x + \alpha$, $x \in \mathbb{R}^n$ mit $Q \in \mathbb{R}^{n \times n}$ symmetrisch gilt: $Df(x) = Q \cdot x^T + c$.

Somit: $0 = \frac{d}{dt} f(x+tp) = \nabla f(x+tp)^T \cdot p$

$$= (Q(x+tp)+c)^T \cdot p$$
$$= (Qx+c)^T + tp^T \cdot Qp$$
$$= Df(x)^T \cdot p + tp^T \cdot Q \cdot p$$

Ist Q positiv definit, so ist $p^T Qp > 0$ (da $p \neq 0$ als Abstiegsrichtung).

Dann $t = -\frac{Df(x)^T \cdot p}{p^T Q p} = -\frac{(Qx+c)^T \cdot p}{p^T \cdot Q \cdot p} =: t_M := t_c$

Bemerkung 9.6

Für $f(x) := \frac{1}{2}x^T \cdot Q \cdot x + c^T x + \alpha$ $\forall x \in \mathbb{R}^n$ mit Q positiv definit und symmetrisch, ist $t_c := t_M$ berechenbar. Ist p ein zu $\lambda_{max}(Q) = \gamma$ gehörender Eigenvektor von Q, so folgt $p^T \cdot Q_p = p^T(\lambda_{max}(Q)) \cdot p = \gamma \cdot \|p\|^2$.

Sei zu diesem p ein $x \in N_0$ gewählt mit $Df(x)^T \cdot p < 0$ (z.B. $x := x^0$ und $x^0 := Q^{-1}(-p-c)$ ist ein solches x, da dann $Df(x) = -p$ gilt).

Für x und p gilt dann: $f(x) - f(x+t_c \cdot p) = (\frac{1}{2}x^T Qx + c^T x + \alpha) - (\frac{1}{2}(x+t_c \cdot p)^T \cdot Q \cdot (x+t_c p)) + c^T(x+t_c \cdot p) + \alpha$

$$= -t_c(Qx+c)^T p - \frac{1}{2}t_c^2 p^T Q p$$
$$= -t_c Df(x)^T \cdot p - t_c^2 \cdot \frac{\gamma}{2} \cdot \|p\|^2 = \frac{1}{\gamma}\left(\frac{Df(x)^T \cdot p}{\|p\|}\right)^2 - \frac{1}{2\gamma}\left(\frac{Df(x)^T \cdot p}{\|p\|}\right)^2$$
$$= \frac{1}{2\gamma}\left(\frac{Df(x)^T \cdot p}{\|p\|}\right)^2$$

Dieser Spezialfall zeigt, dass die Konstante $\vartheta_M^t = \vartheta_c^t = \frac{1}{2\gamma}$ in Satz 9.3 angenommen werden kann, also best-möglich ist.

9.1.2 Die Methode vom Goldenen Schnitt

Numerisch ist die (näherungsweise) Berechnung einer Minimumschrittweite nur für konvexe oder andere "einfache" Funktionen möglich. Eine solche Klasse "einfacher" Funktionen sind die unimodalen Funktionen.

Definition 9.7

Eine Funktion $\varphi : [a,b] \to \mathbb{R}$ heißt unimodal (auf $[a,b]$), falls genau ein $t^* \in [a,b]$ existiert, mit $\varphi(t^*) = \min_{t \in [a,b]} \varphi(t)$ und φ auf $[a, t^*]$ streng monoton fallend und auf $[t^*, b]$ streng monoton wachsend ist.

Bemerke: Eine unimodale Funktion kann Sattelpunkte besitzen und muss somit nicht unbedingt eine konvexe Funktion sein. Es gilt folgende Umkehrung:

Satz 9.8

Jede strikt konvexe Funktion $\varphi \in C[a,b]$ ist eine unimodale Funktion auf $[a,b]$

Ein ableitungsfreies Verfahren zur Berechnung des Minimalpunkts $t^* \in [a,b]$ einer auf $[a,b]$ unimodalen Funktion.

Ist $[a_k, b_k]$ ein Teilintervall von $[a,b]$ und $t^* \in [a_k, b_k]$, dann gilt für Punkte s_k, t_k mit $a_k \leq s_k \leq t_k \leq b_k$ gemäß der Definition von unimodal:

$$\varphi(s_k) > \varphi(t_k) \Rightarrow \varphi(u) > \varphi(s_k) \quad \forall u \in [a_k, s_k)$$
$$\varphi(s_k) \leq \varphi(t_k) \Rightarrow \varphi(u) > \varphi(t_k) \quad \forall u \in (t_k, b_k]$$

Im Fall $f(s_k) > f(t_k)$ liegt der Minimalpunkt f^* in $[s_k, b_k]$, im anderen Fall in $[a_k, t_k]$.

Daher neues Suchintervall: $[a_{k+1}, b_{k+1}] := \begin{cases} [s_k, b_k], & \text{falls } f(s_k) > f(t_k) \\ [a_k, t_k], & \text{falls } f(s_k) \leq f(t_k) \end{cases}$

Erstrebenswerte Unterteilen:

1) $b_k - s_k = t_k - a_k$, d.h. die Länge des Intervalls $[a_{k+1}, b_{k+1}]$ ist unabhängig vom Ausgang der Abfrage "$f(s_k) > f(t_k)$".
2) von k zu $k+1$ ist nur eine neue Funktionsauswertung nötig.

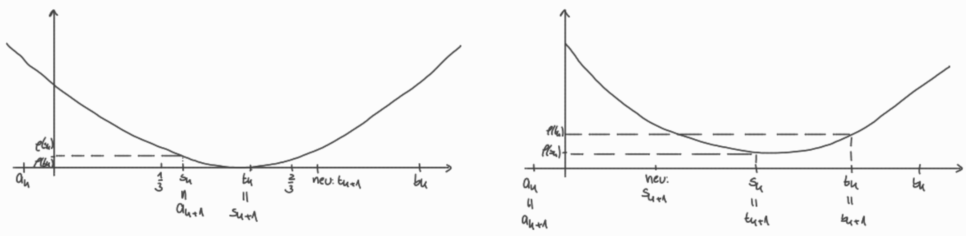
Diese Forderungen erfüllt die Aufteilung von $[a_k, b_k]$ mittels des Goldenen Schnitts.

Allgemein: teile $[a,b]$ in $[a,c]$; $[c,b]$, sodass: $\frac{\text{Länge des ganzen Intervalls}}{\text{Länge des längeren Intervalls}} = \frac{\text{Länge des längeren Intervalls}}{\text{Länge des kürzeren Intervalls}}$

Man kann damit c ausrechnen. Ist $[a,c]$ das längere Teilintervall: $c = a + F \cdot (b-a)$, $F := \frac{\sqrt{5}-1}{2} \approx 0{,}61803$.

Ist $[a,c]$ das kürzere Teilintervall: $c = a + (1-F)(b-a)$, $1 - F \approx 0{,}38197$.

Wähle für jedes k dann: $s_k := a_k + (1-F)(b_k - a_k)$, $t_k := a_k + F \cdot (b_k - a_k)$.



(Punktewahl bei der Methode des Goldenen Schnitts)

Algorithmus 9.9

0) Wähle $\varepsilon \in (0, b-a)$, setze $F := (\sqrt{5}-1)/2$, $a_0 := a$, $b_0 := b$, $s_0 := a + (1-F)(b-a)$, $t_0 := a + F(b-a)$.
   Berechne $f(s_0)$ und $f(t_0)$. Setze $k := 0$.
1) Falls $b_k - a_k \leq \varepsilon$, STOPP $\to$ es gilt: $t^* \in [a_k, b_k]$
2) (i) ist $f(s_k) > f(t_k)$, setze $a_{k+1} := s_k$, $b_{k+1} := b_k$, $s_{k+1} := t_k$, $t_{k+1} := s_{k+1} + F(b_k - s_k)$.
   Berechne $f(t_{k+1})$.
   ii) sonst (d.h. $f(s_k) \leq f(t_k)$), setze $a_{k+1} := a_k$, $s_{k+1} := t_k$, $s_{k+1} := a_k + (1-F)(t_k - a_k)$, $t_{k+1} := s_k$
   Berechne $f(s_{k+1})$.
3) Setze $k := k+1$ und gehe zu Schritt 1.

9.1.3 Anmerkungen

Das Verfahren vom Goldenen Schnitt ist in der Klasse der sogenannten ableitungsfreien Verfahren. Solche interpolieren in jedem Schritt durch Auswertungen in bestimmten Punkten und damit eine Näherung an das Minimum zu bestimmen. Es werden keine Ableitungen benötigt und manche konvergieren sogar superlinear. Allerdings sind zum Teil recht einschränkenden Vorraussetzungen nötig.

Zur Lösung des 1-d-Optimierungsproblems $\min\limits_{t \in [0,\infty)} \varphi(t) := f(x + t p)$ mit konvexer, glatter Funktion φ kann man eine Nullstelle von φ' mit numerischen Verfahren bestimmen, welche nur Funktionswerte von φ' verwenden.

Das Newton-Verfahren benötigt φ'', d.h. die Hesse-Matrix von f, was numerisch oftmals zu teuer ist. Im nicht-konvexen Fall muss noch sicher gestellt werden, dass die Nullstelle ein globaler Minimierer ist. Zum Berechnen der Curry-Schrittweite beginnt man mit einem hinreichend kleinen Intervall nahe Null und nutzt eines der oben genannten Verfahren.

In der Praxis können exakte Schrittweiten nur näherungsweise bestimmt werden, weshalb man andere Regeln verwendet.

9.2. Die Armijo-Schrittweite

Sehr populär, da sehr leicht zu berechnen.

Def 9.10

Sei $\eta \in (0,1)$ und $\xi \in (0, \frac{1}{2})$. Sei $q := q(x,p) \in \mathbb{N}_0$ die kleinste Zahl aus $\mathbb{N}_0$, so dass $f(x) - f(x + \eta^q p) \geq -\xi \cdot \eta^q \cdot \nabla f(x)^T \cdot p$ erfüllt ist. Dann liefert $t_A := t_A(x,p) := \eta^q$ die Amijo-Schrittweite

Da $\eta \in (0,1)$, ist die Amijo-Schrittweite die größte Zahl der Menge $\{1, \eta, \eta^2, \eta^3, \dots\}$, welche die Ungleichung aus Definition 9.10 erfüllt.

Da $1 > \eta > \eta^2 > \eta^3 > \dots$ muss man, mit 1 beginnend, nur überprüfen, für welches η^q diese Ungleichung zum ersten mal erfüllt ist. Die Berechnung der Amijo-Schrittweite ist also trivial.

Beispiel 9.11

Sei f die quadratische Funktion $f(x) = \frac{1}{2} x^T \cdot Q \cdot x + c^T \cdot x + \alpha$, $\forall x \in \mathbb{R}^n$ mit symmetrischer, positiv definiter Matrix Q und $x \in \mathbb{N}_0$, $p \in \mathbb{R}^n$ Abstiegsrichtung in x.

Dann ist: $f(x) - f(x + t p) = -f(Qx + c)^T p - \frac{1}{2} t^2 \cdot p^T Q p \overset{?}{\geq} -\xi \cdot t \cdot \nabla f(x)^T p$

genau dann, wenn $t = -\alpha(1-\xi) \frac{\nabla f(x)^T \cdot p}{p^T Q} = 2(1-\xi) t_c$, wobei die Curry-Schrittweite aus Beispiel 9.5 verwendet wurde. Die Amijo-Schrittweite t_A ist also die größte Zahl aus der Menge $\{1, \eta, \eta^2, \eta^3, \dots\}$, welche der Ungleichung $\eta^q = t_A \leq 2(1-\xi) t_c$ genügt. Da $\xi \in (0, \frac{1}{2})$, ist sie kleiner als $2 t_c$.

Der leichten Berechenbarkeit der Amijo-Schrittweite geht entgegen, dass sie nur semi-effizient ist.

Satz 9.12

Es gelten (V1)-(V3) und $\eta \in (0,1)$, $\xi \in (0, \frac{1}{2})$ sind gegeben. Für alle Paare $x \in \mathbb{N}_0$ und $p \in \mathbb{R}^n$ Abstiegsrichtung existiert genau eine Amijo-Schrittweite. Diese ist eine semi-effiziente Schrittweite mit der Konstanten $\vartheta_1 := \min\left(\xi, \frac{2\eta\xi(1-\xi)}{\gamma}\right)$.

Beweis: Effizienz und Eindeutigkeit:

Aus der Definition der Richtungsableitung erhält man: $-\lim\limits_{t \to 0^+} \frac{f(x+tp) - f(x)}{t} = -\nabla f(x)^T \cdot p > -\xi \nabla f(x)^T \cdot p$, falls $\xi \in (0,1)$. Somit gilt für hinreichend kleine $t > 0$, dass $f(x) - f(x + tp) > -\xi \cdot t \cdot \nabla f(x)^T p$.

Da $\lim\limits_{q \to \infty} \eta^q = 0$, kann die Ungleichung für hinreichend großes q erfüllt werden ($t = \eta^q$) und dieses q ist eindeutig.

Bleibt zu zeigen: Amijo-Schrittweite ist semi-effizient.

Ist $q = 0$, also $\eta^q = \eta^0 = 1 = t_A$, so folgt aus der Amijo-Ungleichung: $f(x) - f(x + t_A p) \geq -\xi \nabla f(x)^T \cdot p$.

Ist $q > 0$, so gilt die Amijo-Ungleichung für $\eta^q = t_A$, aber noch nicht für die direkten Vorgänger $\eta^{q-1} = \eta^{-1} \cdot t_A$, d.h. $f(x) - f(x + \eta^{-1} \cdot t_A \cdot p) < -\xi \cdot \eta^{-1} t_A \cdot \nabla f(x)^T p$ (*)

Für $\hat{t}$ aus Lemma 8.8 betrachte zwei Fälle:

1) Fall mit $\eta^{-1} \cdot t_A \leq \hat{t}$. In diesem Fall folgt aus obiger Ungleichung (*) und Lemma 8.11 ($\Psi(t) \geq \Phi(t)$):

$$-\xi \cdot \eta^{-1} t_A \cdot \nabla f(x)^T p > f(x) - f(x + \eta^{-1} t_A \cdot p) = \Psi(\eta^{-1} t_A) \geq \Phi(\eta^{-1} t_A) = -\eta^{-1} t_A \cdot \nabla f(x)^T \cdot p - (\eta^{-1} \cdot t_A)^2 \cdot \frac{\gamma}{2} \cdot \|p\|^2$$

   Division durch $\eta^{-1} t_A$ und auflösen nach t_A liefert: $t_A \geq -\frac{2\eta(1-\xi)}{\gamma} \cdot \frac{\nabla f(x)^T \cdot p}{\|p\|^2} > 0$.

   Diese Abschätzung setzen wir in die Amijo-Ungleichung ein: $f(x) - f(x + t_A p) \geq \frac{2\eta\xi(1-\xi)}{\gamma} \left(\frac{\nabla f(x)^T p}{\|p\|}\right)^2$

2) Im Fall $\eta^{-1} \cdot t_A \geq \hat{t}$ liefert Lemma 8.11 und das dort definierte $\tilde{t} := -\frac{2}{\gamma} \cdot \frac{\nabla f(x)^T \cdot p}{\|p\|^2} > 0$: $t_A \geq \eta \hat{t} \geq \eta \tilde{t} = \frac{-2\eta}{\gamma} \cdot \frac{\nabla f(x)^T p}{\|p\|} > 0$.

   Auch dies setzt man in die Amijo-Ungleichung ein: $f(x) - f(x + t_A p) \geq \frac{2\eta\xi}{\gamma} \left(\frac{\nabla f(x)^T p}{\|p\|}\right)^2$.

   Damit ist gezeigt: $f(x) - f(x + tp) \geq \vartheta_1 \cdot \min\left(-\nabla f(x)^T p, \left(\frac{\nabla f(x)^T p}{\|p\|}\right)^2\right)$, wobei $\vartheta_1 := \min\left(\frac{2\eta\xi}{\gamma}, \frac{2\eta(1-\xi)}{\gamma}, \xi\right) = \min\left(\xi, \frac{2\eta(1-\xi)}{\gamma}\right)$.

$\square$

04. Juni 2024

9.3 Wolfe-Powell-Schrittweiten

9.3.1. Definition und Effizienz

Wolfe-Powell-Schrittweiten erfüllen neben einer Ungleichung vom Amijo-Typ eine weitere.

Def. 9.13 Gegeben sei $\tau \in (0, \frac{1}{2})$ und $\sigma \in (\tau, 1)$. Dann heißt jedes Element von $T_{WP}(x,p) := \{t \in \mathbb{R}_+ \mid -\tau \cdot t \nabla f(x)^T \cdot p \leq f(x) - f(x+t\cdot p),\ -\nabla f(x+t\cdot p)^T p \leq -\sigma \nabla f(x)^T p\}$ eine **Wolfe-Powell-Schrittweite**.

Beispiel 9.14 Sei $f(x) := \frac{1}{2} x^T Q x + c^T \cdot x + \alpha$ für alle $x \in \mathbb{R}^n$, Q positiv definit. Sei p eine Abstiegsrichtung in x. Sei t_c die Curry-Schrittweite, dann gilt:

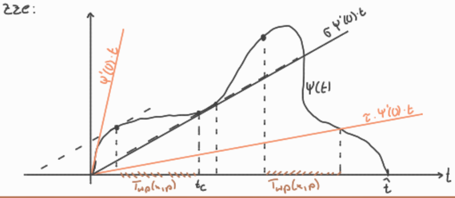
$$T_{WP}(x,p) = [(1-\sigma) t_c, 2(1-\tau) t_c]$$

Aus $\tau \in (0, \frac{1}{2})$ und $\sigma \in (\tau, 1)$ folgt: $1-\sigma < 1 < 2 - 2\tau$, also ist $T_{WP}(x,p) \neq 0$, da insbesondere $t_c \in T_{WP}(x,p)$.

Verwendet man die bekannte Funktion $\varphi(t) = f(x) - f(x+tp)$ mit $\varphi(t) = -\nabla f(x)^T p$ und $\varphi'(0) = -\nabla f(x)^T p > 0$, dann ist:

$$T_{WP}(x,p) = \{t \in \mathbb{R}_+ \mid \underline{\tau \cdot \varphi'(0) \cdot t} \leq \varphi(t),\ \varphi'(t) \leq \sigma \cdot \varphi'(0)\}$$

Skizze:



Satz 9.15 Es gelten (V1)-(V3), $t \in p$ Abstiegsrichtung in x, $\tau \in (0, \frac{1}{2})$, $\sigma \in [\tau, 1\}$ so enthält $T_{WP}(x,p)$ mindestens ein nichtleeres, abgeschlossenes Intervall. Jede Wolfe-Powell-Schrittweite $t_{WP} := t_{WP}(x,p)$ ist effizient mit $\vartheta_{WP} := \frac{1}{\gamma} \min(\frac{1-\sigma^2}{2}, 2\tau(1-\tau))$.

Beweis: Seien x, p, τ, σ wie angegeben. Setze $d(t) := \varphi(t) - \tau \cdot \varphi'(0) \cdot t$. Dann ist $d'(t) = \varphi'(t) - \tau \cdot \varphi'(0)$ und $d(0) \leq \varphi(0) - \tau \cdot \varphi(0) \cdot 0 = \ldots = f(x) - f(x + 0p) = 0$.

Ferner ist $d'(0) = \varphi'(0) - \tau \cdot \varphi'(0) = (1-\tau) \cdot \varphi'(0) > 0$ sowie mit t_c Curry-Schrittweite (lokales Maximum bzw. Sattelpunkt von φ, also $\varphi'(t_c) - 0$):

$$d'(t_c) = \varphi'(t_c) - \tau \varphi'(0) = -\tau \cdot \varphi'(0) < 0.$$

Aufgrund des Vorzeichenwechsels der Ableitung von d in $[0, t_c]$ besitzt d dort (in $(0, t_c)$) eine Stelle t_m mit $d'(t_m) = 0$, was dann ein lokales Maximum ist.

Es folgt (aus $\tau < \sigma$): $0 = d'(t_m) = \varphi'(t_m) - \tau \cdot \varphi'(0) \Rightarrow \varphi'(t_m) = \tau \cdot \varphi'(0) < \sigma \varphi'(0)$

Da nun also $d(t_m) > 0$ (da $d(0) = 0$, $d'(0) > 0$ und bei t_m ein lokales Maximum) und $\varphi'(t_m) < \sigma \cdot \varphi'(0)$ für ein $t_m \in (0, t_c)$, gibt es ein nichtleeres, abgeschlossenes Intervall $I \subseteq (0, t_c)$, sodass dort gilt:

$$d(t) > 0 \Leftrightarrow \varphi(t) - \tau \cdot \varphi(0) \cdot t > 0$$
$$\Leftrightarrow f(x) - f(x+tp) > -\tau \cdot t \nabla f(x)^T p$$
$$\varphi'(t) < \sigma \cdot \varphi'(0) \Leftrightarrow -\nabla f(x+tp)^T p < -\sigma \nabla f(x)^T p,\ \forall t \in I, \text{ d.h. } I \subseteq T_{WP}(x,p).$$

Effizienz

1. Fall: $t_{WP} \leq \frac{-2(1-\tau)}{\gamma} \cdot \frac{\nabla f(x)^T p}{\|p\|^2} =: t_r$. Nach Definition von t_{WP} ist $x + tp \in N_0$ (aus 1. WP-Ungleichung) sowie (2. Ungleichung):

$$-\nabla f(x + t_{WP} \cdot p)^T p \leq -\sigma \nabla f(x)^T p = -\nabla f(x)^T p + (1-\sigma) \nabla f(x)^T p$$
$$\Rightarrow -(1-\sigma) \nabla f(x)^T p \leq (\nabla f(x + t_{WP} \cdot p) - \nabla f(x))^T \cdot p \overset{(V3)}{\leq} \gamma \cdot t_{WP} \cdot \|p\|^2$$
$$\Rightarrow t_\ell := \frac{1-\sigma}{\gamma} \cdot \frac{\nabla f(x)^T p}{\|p\|^2} \leq t_{WP} \Rightarrow 0 < t_\ell \leq t_{WP} \leq t_r \leq \tilde{t} \leq \hat{t}.$$

Die Parabel Ψ aus Lemma 8.11 ist nach unten geöffnet. Auf einem Intervall: hier $[t_\ell, t_r]$ nimmt sie ihr Minimum an den Rändern an, d.h. in t_ℓ oder t_r. Aus Lemma 8.11 (ii) folgt $\varphi(t) \geq \Psi(t)$, $\forall t \in [0, \tilde{t}] \ni t_{WP} \Rightarrow f(x) - f(x + t_{WP} \cdot p) \geq -t_{WP} \nabla f(x)^T \cdot p - t_{WP}^2 \frac{\gamma}{2} \|p\|^2 \geq \min\limits_{t \in [t_\ell, t_r] \ni t_{WP}} (-t \nabla f(x)^T p - t^2 \frac{\gamma}{2} \|p\|^2)$

Ränder t_ℓ, t_r einsetzen: $= \frac{1}{\gamma} \min(\frac{1-\sigma^2}{2}, 2\tau(1-\tau)) \cdot (\frac{\nabla f(x)^T p}{\|p\|})^2$.

2. Fall: $t_{WP} \geq -\frac{2(1-\sigma)}{\gamma} \cdot \frac{\nabla f(x)^T p}{\|p\|^2}$. Aus der ersten WP-Ungleichung folgt: $f(x) - f(x + t_{WP} \cdot p) \geq -\tau\, t_{WP} \cdot \nabla f(x)^T p$

$$\geq \frac{2\tau(1-\sigma)}{\gamma} \cdot \left(\frac{\nabla f(x)^T p}{\|p\|}\right)^2.$$

Also der gleiche Term wie zuvor. Somit ist die Konstante ϑ_{WP} bestimmt. □

9.2.3 Numerische Berechnung

WP lässt sich numerisch bestimmen, da keine bestimmte Zahl gesucht ist, sondern nur eine beliebige aus einem Intervall. Der folgende Algorithmus berechnet eine Wolfe-Powell-Schrittweite in endlich vielen Schritten.

Idee: bestimme Intervall, dessen linker Randpunkt die Amijo-Ungleichung (erste Ungleichung in T_{WP}) erfüllt und dessen rechter Randpunkt sie nicht erfüllt. Erfüllt der linke Randpunkt auch die zweite T_{WP}-Ungleichung, dann fertig.

Anderenfalls: Intervall-Bisektion unter Beibehaltung der Eigenschaften bis linker Randpunkt beide T_{WP}-Ungleichungen erfüllt.

Algorithmus 3.16

0) Eingabe: $x \in \mathbb{N}_0$ und $p \in \mathbb{R}^n$ mit $\nabla f(x)^T p < 0$, $\tau \in (0, \frac{1}{2})$, $\sigma \in (\tau, 1)$, $k := 0$.

1) i) Ist für $t := 1$ die Amijo-Ungleichung $f(x) - f(x + tp) \geq -\tau t \cdot \nabla f(x)^T p$ erfüllt, bestimme die kleinste Zahl $b_k \in \{2, 2^2, 2^3, \dots\}$, sodass die Amijo-Ungleichung für $t = b_k$ verletzt ist, d.h. $f(x) - f(x + tp) < -\tau \cdot t \cdot \nabla f(x)^T p$. Setze $a_k := \frac{1}{2} b_k$

ii) Andernfalls, d.h. für $t := 1$ ist die Amijo-Ungleichung verletzt, bestimme die größte Zahl $a_k \in \{\frac{1}{2}, \frac{1}{2^2}, \frac{1}{2^3}, \dots\}$, sodass für $t := a_k$ die Amijo-Ungleichung gilt. setze $b_k := 2a_k$.

2) Ist für $t := a_k$ die Ungleichung $-\nabla f(x + tp)^T p \leq -\sigma \nabla f(x)^T p$ erfüllt, setze $t_{WP} := a_k$. STOP!

3) Berechne $t_k = \frac{1}{2}(a_k + b_k)$. Prüfe, ob t_k die Amijo-Bedingung erfüllt ist.

Ja – Dann setze $a_{k+1} := t_k$, $b_{k+1} := b_k$

Nein – Dann setze $a_{k+1} := a_k$, $b_{k+1} := t_k$

4) Setze $k := k+1$, gehe zu Schritt 2).

Satz 9.17

Es gelten (V1)-(V3). Dann bricht Algorithmus 3.16 nach endlich vielen Iterationen mit einer Wolfe-Powell-Schrittweite $t_{WP} \in T_{WP}(x, p)$ ab.

Beweis: Nach Lemma 8.8 ist $f(x) < f(x + tp)$ für $t \geq K$ für hinreichend große K. Also ist $f(x) - f(x + tp) < 0 < -\tau t \nabla f(x)^T p$ für solche t. Somit kann ein b_k, wie in 1) i) angegeben, gefunden werden. Im Fall von 1) ii) kann gemäß Satz 9.12 für $\eta := \frac{1}{2}$ und $\delta := \tau$ nach endlich vielen Schritten die Amijo-Schrittweite bestimmt werden, welche a_0 entspricht. Sowohl in Fall i) als auch in ii) hat man für $k = 0$ die Situation $a_k < b_k$, $t := a_k$ erfüllt die Amijo-Bedingung $f(x) - f(x + tp) \geq -\tau t \nabla f(x + tp)^T p$ und $t := b_k$ erfüllt sie nicht. Ist auch $-\nabla f(x + tp)^T p \leq -\sigma \nabla f(x)^T p$ für $t := a_k$ erfüllt, so ist $a_k \in T_{WP}(x, p)$ und das Verfahren bricht in Schritt 2) erfolgreich ab.

Anderenfalls hat man am Anfang von Schritt 3) die Situation:

"$a_k < b_k$, $t := a_k$ erfüllt Amijo-Ungleichung, $t_k := b_k$ nicht". Nun wird die Länge $[a_k, b_k]$ in Schritt 3) halbiert, wobei entweder a_k vergrößert oder b_k verkleinert wird und a_k, b_k obige Eigenschaften bewahren.

Würde der Algorithmus nicht abbrechen, würde auch Schritt 3) unendlich oft durchlaufen, so konvergieren die Folgen $(a_k)_k$ und $(b_k)_k$, da $a_0 \leq a_k \leq a_{k+1} \leq b_{k+1} \leq b_k \leq b_0$ und die Länge der Intervalle $[a_k, b_k]$ gegen 0 gehen, konvergieren beide gegen die gleiche Zahl t^*.

Idee: Im Grenzfall wäre die Amijo-Ungleichung für t^* sowohl erfüllt (linke Intervallgrenze), als auch nicht erfüllt (rechte Intervallgrenze) 5. Juli 2024

Setze $d(t) := f(x) - f(x + tp) + \tau t \nabla f(x)^T p$. Dann ist $d(t^*) = 0$. Ferner gilt $d'(t^*) \leq 0$, denn: falls $d'(t^*) > 0$, d.h. $\lim_{t \to t^*} \frac{d(t) - d(t^*)}{t - t^*} > 0$. Da f stetig differenzierbar ist, und damit d stetig, würde für hinreichend große k auch für den Differenzenquotienten gelten: $\frac{d(b_k) - d(t^*)}{b_k - t^*} > 0$. Da $b_k > t^*$ ist der Nenner positiv, $d(t^*) = 0$, also bleibt $d(b_k) > 0$ für alle hinreichend großen k.

D.h. $f(x) - f(x + tp) > -\tau t \nabla f(x)^T p$ (nach Definition von d) für t nahe an t^*, also wäre da die Amijo-Bedingung erfüllt, aber alle b_k verletzen sie. Widerspruch. Also ist $d'(t^*) \leq 0$, d.h. $-\nabla f(x + t^* p)^T p \leq -\tau \nabla f(x)^T p$.

Aus $\tau \leq \sigma$ (da $\sigma \in (\tau, 1)$) folgt dann $-\nabla f(x + t^* p)^T p \leq -\tau \cdot \nabla f(x)^T p < -\sigma \nabla f(x)^T p$. Da die Ableitung stetig ist, gilt diese Abschätzung auch für hinreichend große k, bei denen a_k hinreichend nahe an t^* ist: $-\nabla f(x + a_k p)^T p < -\sigma \nabla f(x)^T p$. Somit ist die 2. TP-Ungleichung für hinreichend großes k erfüllt mit $t := a_k$, was ein Abbruch des Algorithmus bedeutet und somit ein Widerspruch zur Annahme, dass er nicht abbricht. Also war die Annahme falsch.

der Algorithmus bricht ab. □