Trust-Region- und Quasi-Newton-Verfahren

Laslo Hunhold

Seminar über industrielle Anwendungen
Mathematisches Institut
Universität zu Köln

bei

PD Dr. Thomas Mrziglod
Bayer AG, Leverkusen

8. Januar 2018

Einführung

„Programmierung“

- Formulierung des Problems
- Entwicklung von Algorithmen
- Analyse der Algorithmen

Problemeigenschaften

- Unbeschränkt
- Stetig
- Deterministisch
- Lokal

Anforderung an Algorithmen

- Robust
- Effizient
- Genau

Mathematische Formulierung

Problem

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{gemäß} \quad & x \in \Omega \end{aligned}$$

Parameter

- Lösungsraum Ω (überabzählbar)
- Bewertungsfunktion $f : \Omega \to \mathbb{R}$ („glatt“)
- Variable x

Mathematische Formulierung

Lösungsbegriff und Strategie

Lokaler Minimierer $x^\star$ $\exists \mathcal{N}(x^\star) \subseteq \Omega \colon \forall x \in \mathcal{N}(x^\star) : f(x^\star) \leq f(x)$

Stationärer Punkt $x^\star$ $\nabla f(x^\star) = 0$

Taylorformel $\forall x, p \in \Omega \colon x + p \in \Omega \colon$
$$f(x+p) \approx f(x) + \nabla f(x)^T p + \frac{1}{2} p^T \nabla^2 f(x + tp) p$$

$\rightarrow$ NB 1. Ordnung $x^\star$ lokaler Minimierer $\wedge f \in \mathcal{C}^1(\mathcal{N}(x^\star)) \Rightarrow$
$\nabla f(x^\star) = 0$

$\rightarrow$ NB 2. Ordnung $x^\star$ lokaler Minimierer $\wedge f \in \mathcal{C}^2(\mathcal{N}(x^\star)) \Rightarrow$
$\nabla f(x^\star) = 0 \wedge \nabla^2 f(x^\star)$ positiv semidefinit

$\rightarrow$ HB 2. Ordnung
$f \in \mathcal{C}^2(\mathcal{N}(x^\star)) \wedge \nabla f(x^\star) = 0 \wedge \nabla^2 f(x^\star)$ pos. def. $\Rightarrow$
$x^\star$ streng lokaler Minimierer

$\rightarrow$ $\boxed{\text{Finde } x^\star \text{ mit } \nabla f(x^\star) = 0}$

Algorithmen

Überblick

Parameter

- Startpunkt $x_0 \in \Omega$
- Iterierte $\{x_k\}_{k=0}^{\infty}$

Schritt

- $x_{k+1} \leftarrow \{x_k, \ldots, x_0,\ \text{Verhalten von } f \text{ in } x_k\}$
- 2 Strategien Liniensuchverfahren und Trust-Region-Verfahren

Ziel

- monotoner Algorithmus $f(x_{k+1}) < f(x_k)$
- nichtmonotoner Algorithmus $\exists m > 0\colon f(x_{k+m}) < f(x_k)$

Algorithmen

Liniensuchverfahren

Schritt $k \to k+1$

- *Wähle* Suchrichtung p_k
- *Suche* Schrittweite $\alpha_k = \mathrm{argmin}_{\alpha>0}(f(x_k + \alpha_k \cdot p_k))$
  - „Liniensuche“
  - Approximativ
  - Sample-Bewertung mit Wolfe- bzw. Goldstein-Bedingungen
  - Nicht näher betrachtet
- *Setze* $x_{k+1} = x_k + \alpha_k \cdot p_k$

Algorithmen

Trust-Region-Verfahren

Schritt $k \to k+1$

- *Formuliere* Modellfunktion
  $m_k(p) = f_k + (\nabla f_k)^T p + \frac{1}{2} p^T B_k p \approx f(x_k + p)$
  - $f_k := f(x_k)$
  - $\nabla f_k := \nabla f(x_k)$
  - B_k symmetrisch
- *Wiederhole*
  - *Suche* Suchrichtung $p_k = \operatorname{argmin}_{\|p\| \leq \Delta_k}(m_k(p))$
  - *Berechne* Vereinbarkeit $\rho_k(p_k) = \frac{f(x_k) - f(x_k + p_k)}{m(0) - m(p_k)}$

  solange $\rho_k < 0$
- *Bestimme* Trust-Region-Radius Δ_{k+1} aus ρ_k
- *Wenn* $\rho_k > \eta$
  - $x_{k+1} = x_k + p_k$

  Sonst
  - $x_{k+1} = x_k$

Algorithmen

Vergleich

Unterschied im Ansatz

- Jeweils umgekehrte *Reihenfolge*
- Liniensuchverfahren
  Wähle Suchrichtung p_k und suche Schrittweite α_k
- Trust-Region-Verfahren
  Wähle maximale Schrittweite $\Delta_k \geq \alpha_k$ und suche Suchrichtung p_k

Gemeinsames Problem

- Wahl bzw. Suche der Suchrichtung

Liniensuchverfahren

Wahl der Suchrichtung

Steilster Abstieg

$$\rightarrow \boxed{p_k^S = -\frac{\nabla f_k}{\|\nabla f_k\|}}$$

- 1. Ordnung $\rightarrow$ Einfache Berechnung, langsame Konvergenz
- $\forall \tilde{p}_k \colon |\angle(p_k^S, \tilde{p}_k)| < \frac{\pi}{2} \colon \forall \tilde{\alpha}_k$ klein: $f(x_k + \tilde{\alpha}_k \cdot \tilde{p}_k) < f(x_k)$

Newton-Richtung

- *Setze* $m_k(p) = f_k + (\nabla f_k)^T p + \frac{1}{2} p^T \nabla^2 f_k p \approx f(x_k + p)$
- *Löse* $0 = \nabla m_k(p) = \nabla f_k + (\nabla^2 f_k)^T p = \nabla f_k + \nabla^2 f_k p$

$$\rightarrow \boxed{p_k^N = -\nabla^2 f_k^{-1} \nabla f_k}$$

- 2. Ordnung $\rightarrow$ Schnelle Konvergenz, langsame Berechnung
- Finde Kompromiß $\rightarrow$ Quasi-Newton-Verfahren

Liniensuchverfahren

Quasi-Newton-Verfahren

Approximiere $B_k \approx \nabla^2 f_k \rightarrow \boxed{p_k^{QN} = -B_k^{-1} \nabla f_k}$

Bedingungen an B_k

- Sekantengleichung $\nabla^2 f_{k+1}(x_{k+1} - x_k) \approx \nabla f_{k+1} - \nabla f_k \rightarrow$
  $\boxed{B_{k+1} s_k = y_k}$ mit $s_k := x_{k+1} - x_k$, $y_k = \nabla f_{k+1} - \nabla f_k$
- Kurvaturbedingung
  B_k positiv definit $\rightarrow \boxed{s_k^T y_k > 0}$
- $\boxed{\text{Symmetrisch}}$ wie $\nabla^2 f_k$
- $\boxed{\text{rank}(B_{k+1} - B_k) \text{ klein}}$

Verfahren

- BFGS („Broyden-Fletcher–Goldfarb-Shanno“)
- SR1 („Symmetric Rank-One“)

Liniensuchverfahren

Quasi-Newton-Verfahren (BFGS)

Herleitung

- $B_{k+1} = B_k + U_k + V_k$ mit U_k, V_k symmetrisch-Rang-1
- B_{k+1} symmetrisch $\rightarrow$ Setze $U_k = \gamma u u^T,\ V_k = \mu v v^T$ mit $\gamma, \mu \in \mathbb{R},\ u, v \in \mathbb{R}^n$
- Wähle $u = B_k s_k,\ v = y_k$
- Sekantengleichung $\rightarrow \gamma = -\frac{1}{s_k^T B_k s_k},\ \mu = \frac{1}{y_k^T s_k}$

$$\rightarrow \boxed{B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{y_k^T s_k}}$$

Eigenschaften

- B_k positiv definit
- rank$(B_{k+1} - B_k) = rank(U_k + V_k) = rank(U_k) + rank(V_k) = 2$
- Global superlinear konvergent

Liniensuchverfahren

Quasi-Newton-Verfahren (SR1)

Herleitung

- $B_{k+1} = B_k + \sigma u u^T$ mit $\sigma \in \{-1, 1\}$, $u \in \mathbb{R}^n$
- Sekantengleichung $\rightarrow$
  $B_{k+1} s_k = y_k = B_k s_k + \sigma u u^T s_k = B_k s_k + \left[\sigma u^T s_k\right] u$
  $\rightarrow u = \alpha(y_k - B_k s_k)$
- $$\begin{aligned} y_k - B_k s_k &= \sigma\alpha^2(y_k^T - s_k^T B_k) s_k (y_k - B_k s_k) \\ &= \left\{\sigma\alpha^2 \left[s_k^T (y_k - B_k s_k)\right]\right\} (y_k - B_k s_k) \end{aligned}$$
  $\rightarrow \sigma = \operatorname{sgn}\left(\left[s_k^T (y_k - B_k s_k)\right]\right), \ \alpha = \pm\left|s_k^T (y_k - B_k s_k)\right|^{-\frac{1}{2}}$

$$\rightarrow \boxed{B_{k+1} = B_k + \frac{(y_k - B_k s_k)(y_k - B_k s_k)^T}{s_k^T (y_k - B_k s_k)}}$$

Eigenschaften

- B_k i.A. nicht positiv definit
- $\operatorname{rank}(B_{k+1} - B_k) = rank(\sigma u u^T) = 1$

Trust-Region-Verfahren

Suche der Suchrichtung

Erinnerung

$p_k = \text{argmin}_{\|p\| \leq \Delta_k}(m_k(p))$ mit $m_k(p) = f_k + (\nabla f_k)^T p + \frac{1}{2} p^T B_k p$

Lineares Problem $B_k = 0$

$\rightarrow m_k(p) = f_k + (\nabla f_k)^T p$

$\rightarrow \boxed{p_k^L = -\Delta_k \frac{\nabla f_k}{\|\nabla f_k\|}} = -\Delta_k p_k^S \rightarrow$ Steilster Abstieg

Trust-Region-Newton-Verfahren $B_k = \nabla^2 f_k$

- $\|p\| \leq \Delta_k \rightarrow \nabla^2 f_k$ muß nicht positiv definit sein

$\rightarrow$ Trust-Region-Quasi-Newton-Verfahren mit SR1

Trust-Region-Verfahren

Cauchy-Punkt

Motivation

- Exakter Schritt p_k nicht nötig, nur ausreichende Reduktion
- Quantifiziere Reduktion mit Cauchy-Punkt p_k^C

Vorgehen

- *Berechne* Lösung $p_k^L = -\Delta_k \frac{\nabla f_k}{\|\nabla f_k\|}$ des linearen Problems mit $B_k = 0$
- *Bestimme* $\tau_k = \text{argmin}_{\tau \in (0,1]} \left(m_k \left(\tau p_k^L \right) \right)$

$\rightarrow \boxed{p_k^C = \tau_k p_k^L}$

Nutzen

- Einfach zu berechnen
- Globale Konvergenz, falls Verfahren pro Schritt mindestens Cauchy-Reduktion erreicht

Trust-Region-Verfahren

Analyse

- Global konvergent für $\eta = 0$ und $\eta \in \left(0, \frac{1}{4}\right)$ unter jeweils schwachen Bedingungen
- Kreisförmige Trust-Regions nicht geeignet für schlecht-skalierende Funktionen

Ausblick

- Sherman-Morrison-Formel für analytischen Ausdruck $H_k := B_k^{-1}$
- Nicht-euklidische Trust-Regions

Ende

Vielen Dank für Ihre Aufmerksamkeit!