TECHNISCHE HOCHSCHULE MITTELHESSEN

Unrestringierte Optimierungsprobleme

Gradienten- und Trust-Region-Verfahren im Vergleich

Büsra Karaoglan

Fachbereich Mathematik, Naturwissenschaften und Datenverarbeitung
Studiengang Wirtschaftsmathematik

2. März 2017

Inhaltsverzeichnis

1 Einleitung

2 Das Gradientenverfahren

3 Das Trust-Region-Verfahren

4 Numerische Resultate

Inhaltsverzeichnis

1 Einleitung

2 Das Gradientenverfahren

3 Das Trust-Region-Verfahren

4 Numerische Resultate

Einleitung

Hier ist eine Zielfunktion $f : \mathbb{R}^n \to \mathbb{R}$ gegeben. Das Problem besteht darin, einen Punkt zu finden, in dem diese Zielfunktion minimal ist. Das heißt es wird ein Problem der Form

$$\text{Minimiere } f(x), \quad x \in \mathbb{R}^n \tag{1}$$

betrachtet. Es wird zwischen

- einer *globalen Lösung* von (1), das heißt einem Punkt $\bar{x} \in \mathbb{R}^n$ mit $f(\bar{x}) \leq f(x)$ für alle $x \in \mathbb{R}^n$,
- und einer *lokalen Lösung* von (1), das heißt einem Punkt $\bar{x} \in \mathbb{R}^n$, zu dem es eine Umgebung U gibt, so dass $f(\bar{x}) \leq f(x)$ wenigstens für alle $x \in U$ ist,

unterschieden.

Die Idee

Lineare Approximation:
Linearisierung von f in x ergibt die Funktion

$$\hat{f}(y) = f(x) + \nabla f(x)^T(y-x) \Rightarrow d = \hat{y} - x = -\frac{\nabla f(x)}{\|\nabla f(x)\|}$$

Quadratische Approximation:
f wird lokal durch

$$\hat{f}(y) = f(x) + \nabla f(x)^T(y-x) + \frac{1}{2}(y-x)^T\nabla^2 f(x)(y-x)$$

approximiert $\Rightarrow d = \hat{y} - x = -\nabla^2 f(x)^{-1}\nabla f(x)$

Inhaltsverzeichnis

1 Einleitung

2 Das Gradientenverfahren

3 Das Trust-Region-Verfahren

4 Numerische Resultate

Das Gradientenverfahren

Definition

0. Wähle $\beta \in (0,1)$, $\gamma \in (0,1)$ und einen Startpunkt $x^0 \in \mathbb{R}^n$.
   Für k = 0, 1, 2, ...:
1. Falls $\nabla f(x^k) = 0$, STOP.
2. Setze $s^k = -\nabla f(x^k)$.
3. Bestimme die Schrittweite $\sigma_k > 0$ mithilfe der Armijo-Regel (2).
4. Setze $x^{k+1} = x^k + \sigma_k s^k$.

Armijo-Schrittweitenregel:
Seien $\beta \in (0,1)$ und $\gamma \in (0,1)$ fest gewählte Parameter. Bestimme die größte Zahl $\sigma_k \in 1, \beta, \beta^2, \dots$, für die gilt:

$$f(x^k + \sigma_k s^k) - f(x^k) \leq \sigma_k \gamma \nabla f(x^k)^T s^k. \tag{2}$$

THM
TECHNISCHE HOCHSCHULE MITTELHESSEN

Satz

Sei $f : \mathbb{R}^n \to \mathbb{R}$ streng konvex und quadratisch. Weiter seien die Folgen (x^k) und (σ_k) durch das Gradientenverfahren mit Minimierungsregel erzeugt. Dann gilt:

$$f(x^{k+1}) - f(\bar{x}) \leq \left(\frac{\lambda_{max}(C) - \lambda_{min}(C)}{\lambda_{max}(C) + \lambda_{min}(C)}\right)^2 \left(f(x^k) - f(\bar{x})\right), \tag{3}$$

$$\|x^k - \bar{x}\| \leq \sqrt{\frac{\lambda_{max}(C)}{\lambda_{min}(C)}} \left(\frac{\lambda_{max}(C) - \lambda_{min}(C)}{\lambda_{max}(C) + \lambda_{min}(C)}\right)^k \|x^0 - \bar{x}\|, \tag{4}$$

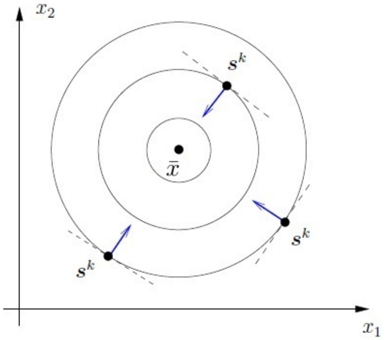
wobei $\bar{x} = -C^{-1}c$ das globale Minimum von f bezeichnet und $\lambda_{max}(C)$ & $\lambda_{min}(C)$ der maximale & minimale Eigenwert von C sind.

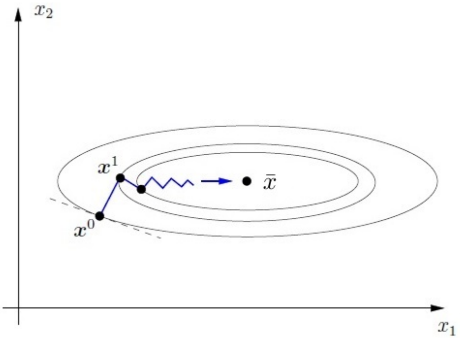
Konvergenzgeschw. mittels Kondition

Mit der *spektralen Konditionszahl* $\kappa := \kappa(\nabla^2 f) = \kappa(C) = \frac{\lambda_{max}}{\lambda_{min}}$ lassen sich die Abschätzungen wie folgt formulieren:

$$f(x^{k+1}) - f(\bar{x}) \leq \left(\frac{\kappa - 1}{\kappa + 1}\right)^2 (f(x^k) - f(\bar{x}))$$

$$\|x^k - \bar{x}\| \leq \sqrt{\kappa}\left(\frac{\kappa - 1}{\kappa + 1}\right)^k \|x^0 - \bar{x}\|$$





Inhaltsverzeichnis

1 Einleitung

2 Das Gradientenverfahren

3 Das Trust-Region-Verfahren

4 Numerische Resultate

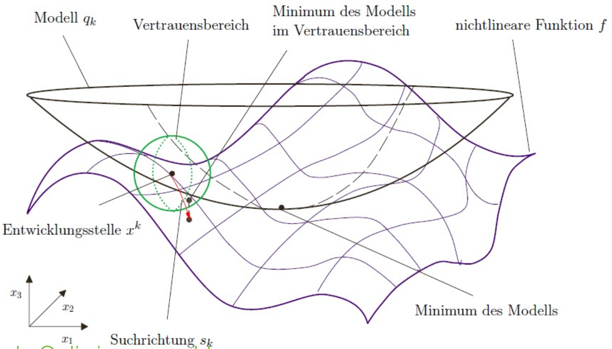
Das Trust-Region-Verfahren

Durch Taylor-Entwicklung von $f(x^k + s)$ um $s = 0$ entsteht ein *quadratisches Modell*

$$q_k(s) = f_k + {g^k}^T s + \frac{1}{2} s^T H_k s$$

mit $f_k = f(x^k),\ g^k = \nabla f(x^k)$ und $H_k = \nabla^2 f(x^k)$.
Die Schrittberechnung erfolgt durch Lösen des **Trust-Region Teilproblems:**

$$\min_{s \in \mathbb{R}^n} \{q_k(s);\, \|s\| \leq \Delta_k\}. \tag{5}$$



Anpassung des Vertrauensbereichs I

Wie sollte Δ_k gewählt werden?

Die Bewertung der Qualität des berechneten Schritts erfolgt durch Vergleich der **Abnahme der Modellfunktion** q_k **(predicted reduction)**

$$pred_k(s^k) = q_k(0) - q_k(s^k) = f_k - q_k(s^k)$$

und der **tatsächlichen Abnahme (actual reduction) der Zielfunktion**

$$ared_k(s^k) = f_k - f(x^k + s^k).$$

$$\Rightarrow \quad \rho_k(s^k) = \frac{ared_k(s^k)}{pred_k(s^k)} \tag{6}$$

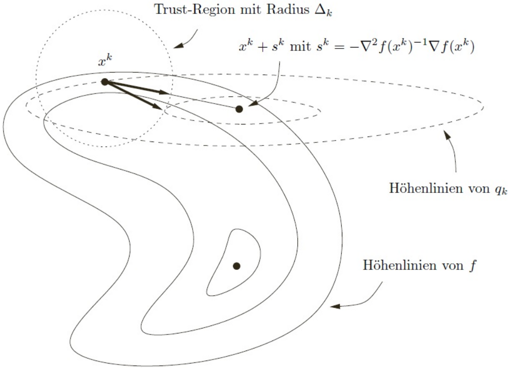
Cauchy-Abstiegsbedingung

Es gibt von k unabhängige Konstanten $\alpha \in (0,1]$ und $\beta \geq 1$ mit

$$\|s^k\| \leq \beta \Delta_k, \quad pred_k(s^k) \geq \alpha \cdot pred_k({s_c}^k), \tag{7}$$

wobei der *Cauchy-Schritt* ${s_c}^k$ die eindeutige Lösung des folgenden eindimensionalen Minimierungsproblems ist:

$$min\ q_k(s) \quad \text{u.d.N} \quad s = \tau \cdot s_l^k, \quad \tau \geq 0, \quad \|s\| \leq \Delta_k. \tag{8}$$



Trust-Region-Algorithmus

Definition

Wähle Parameter
$\alpha \in (0,1], \beta \geq 1, 0 < \eta_1 < \eta_2 < 1, 0 < \gamma_0 < \gamma_1 < 1 < \gamma_2$ und $\Delta_{min} \geq 0$.
Wähle einen Startpunkt $x^0 \in \mathbb{R}^n$ und einen Trust-Region-Radius $\Delta_0 > 0$ mit $\Delta_0 \geq \Delta_{min}$. Für $k = 0, 1, 2, \dots$:

1. Falls $g^k = 0$, dann STOP mit Resultat x^k.
2. Wähle eine symmetrische Matrix $H_k \in \mathbb{R}^{n \times n}$.
3. Berechne einen Schritt s^k, der die Cauchy-Abstiegsbedingung (7) erfüllt.
4. Berechne $\rho_k(s^k)$ gemäß (6).
5. Falls $\rho_k(s^k) > \eta_1$, dann akzeptiere den Schritt s^k, das heißt setze $x^{k+1} = x^k + s^k$. Andernfalls verwerfe den Schritt, das heißt setze $x^{k+1} = x^k$.
6. Berechne Δ_{k+1} gemäß Algorithmus 4.

Update des Trust-Region-Radius

Definition

Seien η_1, η_2 und $\gamma_0, \gamma_1, \gamma_2$ wie in Algorithmus 3 gewählt.

1. Falls $\rho_k(s^k) \leq \eta_1$, so wähle $\Delta_{k+1} \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k]$.
2. Falls $\rho_k(s^k) \in (\eta_1, \eta_2]$, so wähle $\Delta_{k+1} \in [max\left\{\Delta_{min}, \gamma_1 \Delta_k\right\}, max\left\{\Delta_{min}, \Delta_k\right\}]$.
3. Falls $\rho_k(s^k) > \eta_2$, so wähle $\Delta_{k+1} \in [max\left\{\Delta_{min}, \Delta_k\right\}, max\left\{\Delta_{min}, \gamma_2 \Delta_k\right\}]$.

Inhaltsverzeichnis

1 Einleitung

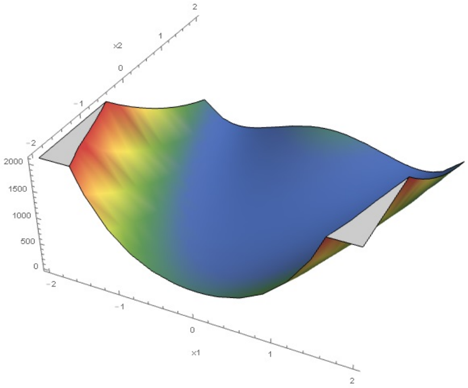
2 Das Gradientenverfahren

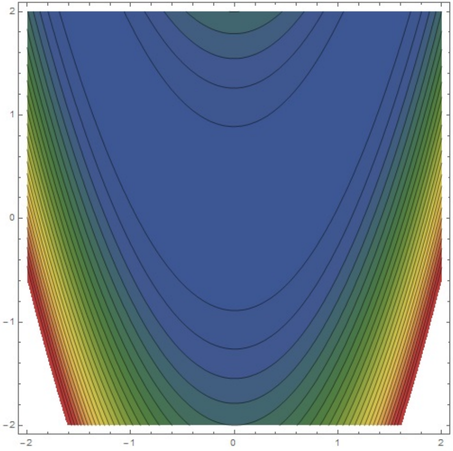
3 Das Trust-Region-Verfahren

4 Numerische Resultate

Numerische Resultate I

$$f : \mathbb{R}^2 \to \mathbb{R}, \qquad f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$





Numerische Resultate II

Die Parameter für das Gradientenverfahren:

$x^0 = (-1,9;2)^T$, $\varepsilon = 10^{-3}$, $\beta = 0,5$ und $\gamma = 10^{-4}$

Die Parameter für das Trust-Region-Verfahren:

$x^0 = (-1,9;2)^T$, $\Delta_0 = 2$, $\varepsilon = 10^{-3}$, $\beta = 2$, $\alpha = 0,5$, $\eta_1 = 0,1$, $\eta_2 = 0,9$, $\gamma_1 = 0,5$ und $\gamma_2 = 2$

$$\nabla f(x) = \begin{pmatrix} 400x_1^3 - 400x_1x_2 + 2x_1 - 2 \\ -200x_1^2 + 200x_2 \end{pmatrix} \;\&\; \nabla^2 f(x) = \begin{pmatrix} 1200x_1^2 - 400x_2 + 2 & -400x_1 \\ -400x_1 & 200 \end{pmatrix}$$

$$\nabla f(x) \stackrel{!}{=} 0 \Rightarrow \bar{x} = (1,1)^T \;\&\; \nabla^2 f(\bar{x}) = \begin{pmatrix} 802 & -400 \\ -400 & 200 \end{pmatrix} \Rightarrow \lambda_1 \approx 1001,6 \;\&\; \lambda_2 \approx 0,4$$

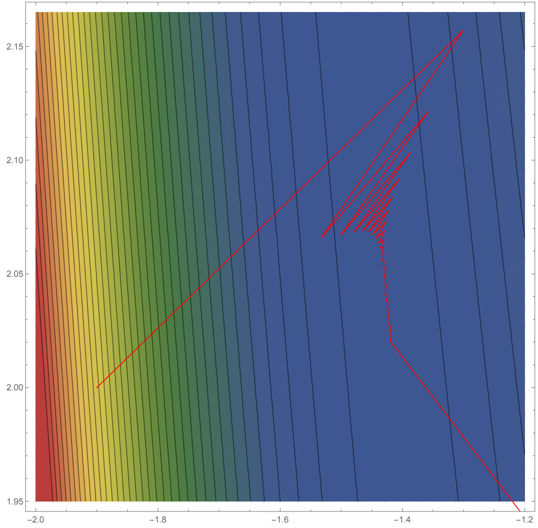
Für die Konditionszahl der Rosenbrock-Funktion gilt

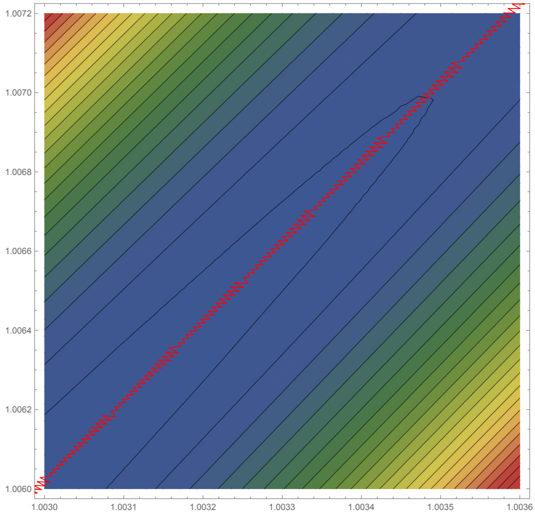
$$\kappa(\nabla^2 f(\bar{x})) = \frac{\lambda_{max}(\nabla^2 f)}{\lambda_{min}(\nabla^2 f)} = 2508,01.$$

Numerische Resultate III

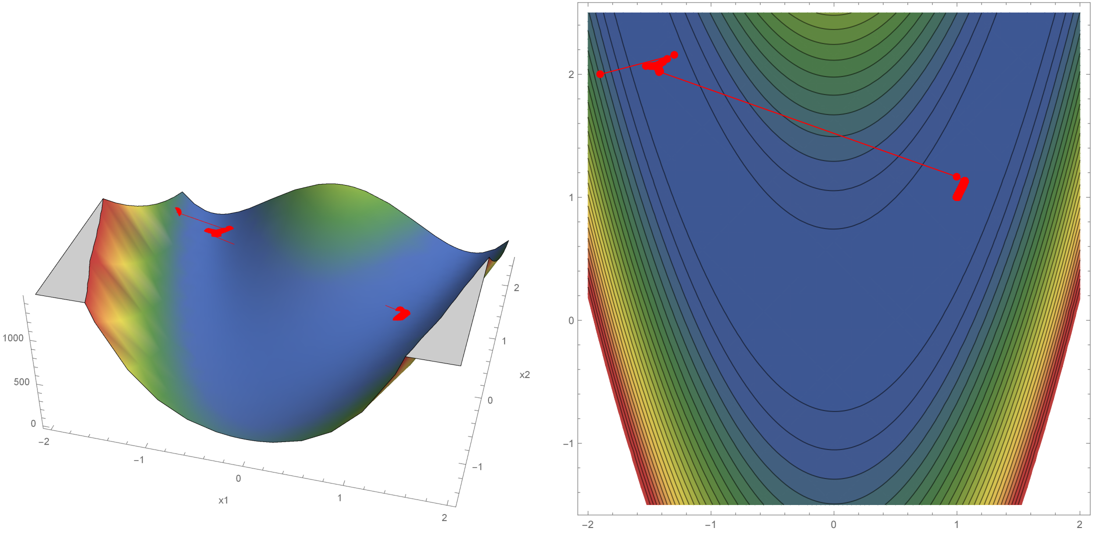
k	x^k	$f(x^k)$
0	$(-1,9;\ 2)$	$267,62$
1	$(-1,29971;\ 2,15723)$	$27,1899$
2	$(-1,53281;\ 2,06582)$	$14,4632$
3	$(-1,35800;\ 2,12123)$	$13,2361$
4	$(-1,50037;\ 2,06712)$	$9,63696$
5	$(-1,38765;\ 2,10305)$	$8,85048$
6	$(-1,47919;\ 2,06839)$	$7,57699$
$\vdots$	$\vdots$	$\vdots$
42	$(-1,41995;\ 2,02053)$	$5,85799$
43	$(0,997336;\ 1,16747)$	$2,98557$
44	$(1,06466;\ 1,13372)$	$0,00418556$
45	$(1,06362;\ 1,13231)$	$0,00415279$
46	$(1,06392;\ 1,13211)$	$0,00408911$
$\vdots$	$\vdots$	$\vdots$
4030	$(1,00078;\ 1,00157)$	$6,13034 \cdot 10^{-7}$
4031	$(1,00078;\ 1,00157)$	$6,12058 \cdot 10^{-7}$

Numerische Resultate IV





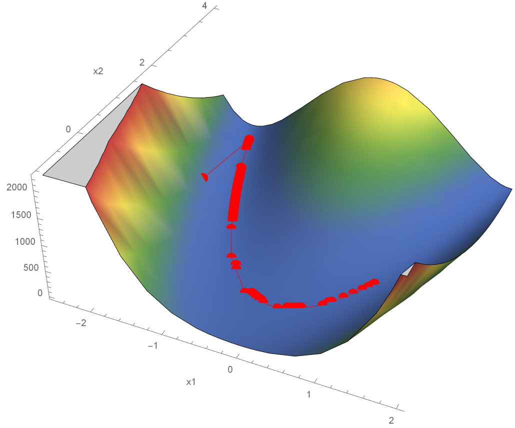
Numerische Resultate V

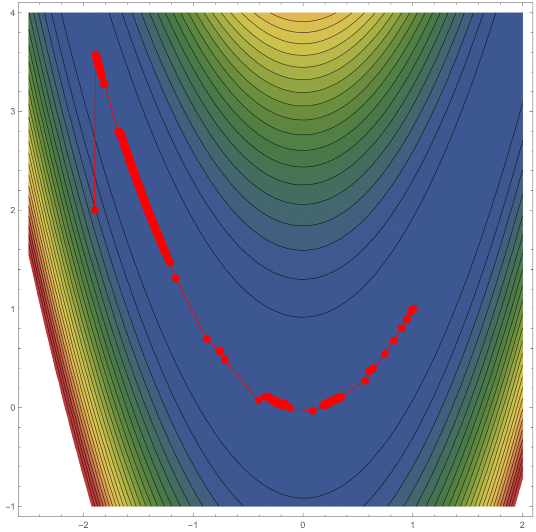


Numerische Resultate VI

k	x^k	ρ_k	s^k
0	$(-1,9;\ 2)$	1.0001	N
1	$(-1,89102;\ 3,57588)$	$1,00108$	N
2	$(-1,88898;\ 3,57589)$	$0,999999$	C
3	$(-1,88891;\ 3,56772)$	$-712,817$	N
4	$(-1,88891;\ 3,56772)$	$-712,817$	N
5	$(-1,88891;\ 3,56772)$	$1,0011$	N
6	$(-1,88683;\ 3,56773)$	$0,999996$	C
7	$(-1,88669;\ 3,55914)$	$-634,505$	N
8	$(-1,88669;\ 3,55914)$	$-634,505$	N
9	$(-1,88669;\ 3,55914)$	$1,00113$	N
10	$(-1,88456;\ 3,55918)$	$0,99999$	C
11	$(-1,88436;\ 3,55011)$	$-560,602$	N
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1188	$(0,983646;\ 0,966281)$	$1,17659$	N
1189	$(0,996671;\ 0,993183)$	$1,10757$	N
1190	$(0,999891;\ 0,999771)$	$1,02508$	N
1191	$(1;\ 1)$	$1,00109$	N

Numerische Resultate VII





Danke für Ihre Aufmerksamkeit.