

Implementierungsaufgabe:

## Gradienten-Verfahren

für die Lehrveranstaltung: Optimierung 1, UE  
gelesen von: Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Bettina Klinz  
am Institut für Diskrete Mathematik  
Technische Universität Graz

Moritz Mossböck

11820925

[moritz.mossboeck@student.tugraz.at](mailto:moritz.mossboeck@student.tugraz.at)

24. August 2023



# 1 Das Verfahren

Beim Gradientenverfahren handelt es sich wohl um die direkteste Umsetzung des generischen Abstiegsverfahren. Die Abstiegsrichtung  $\mathbf{d}$  kann hierbei als Lösung der Optimierungsaufgabe

$$\min \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle \quad \text{u.d.NB} \quad \left\{ \|\mathbf{d}\| = 1 \right.$$

Mit Hilfe der Winkelbedingung aus [Gei99, S. 27f] folgt aus der Cauchy-Schwarzschen Ungleichung, dass

$$\mathbf{d} = -\frac{1}{\|\nabla f(\mathbf{x})\|} \nabla f(\mathbf{x})$$

gilt. In der einfachsten Variante des Verfahrens verwenden wir die Armijo-Schrittweitenregel zur Bestimmung der  $t_k$ . Damit ergibt sich der folgende Algorithmus.

---

```

1  input :  $f: \mathbb{R}^n \rightarrow \mathbb{R}, \nabla f: \mathbb{R}^n \rightarrow \mathbb{R}^n, \mathbf{x}^0 \in \mathbb{R}^n, \sigma, \beta \in (0, 1), \varepsilon \geq 0$ 
2  output :  $\mathbf{x}^* \in \mathbb{R}^n$ 
3
4  steepest( $f, \nabla f, \mathbf{x}^0, \beta, \sigma, \varepsilon$ ):
5       $\mathbf{x}^k \leftarrow \mathbf{x}^0$ 
6      while  $\|\nabla f(\mathbf{x}^k)\| \geq \varepsilon$  do
7           $\mathbf{d}^k \leftarrow -\nabla f(\mathbf{x}^k)$ 
8           $t_k \leftarrow \text{armijo}(f, \nabla f, \mathbf{x}^k, \mathbf{d}^k, \beta, \sigma)$ 
9           $\mathbf{x}^k \leftarrow \mathbf{x}^k + t_k \mathbf{d}^k$ 
10     return  $\mathbf{x}^k$ 

```

---

In der Problemstellung wurde eine weitere Schrittweitenregel vorgestellt, die eine Mischung aus der Armijo-Regel und der Wolfe-Powell-Regel darstellt und auf dem Bisketionsverfahren basiert.

## Gemischte Schrittweitenregel

---

```

1  input :  $f: \mathbb{R}^n \rightarrow \mathbb{R}, \nabla f: \mathbb{R}^n \rightarrow \mathbb{R}^n, \mathbf{x}^0 \in \mathbb{R}^n, \sigma, \beta \in (0, 1), \varepsilon \geq 0, \eta > 1$ 
2  output :  $\mathbf{x}^* \in \mathbb{R}^n$ 
3
4  mixed_step( $f, \nabla f, \mathbf{x}^0, \beta, \sigma, \varepsilon, \eta$ ):
5       $i \leftarrow 1$ 
6      while  $f(\mathbf{x}^k + \eta^i \mathbf{d}^k) < f(\mathbf{x}^k) - \sigma \eta^i \|\mathbf{d}^k\|^2$  do
7           $i \leftarrow i + 1$ 
8       $t'_k \leftarrow \eta^i$ 
9       $\tau_0 \leftarrow 0$ 
10      $\tau_1 \leftarrow t'_k$ 
11     while true do
12          $\tau \leftarrow \frac{1}{2}(\tau_0 + \tau_1)$ 
13         if  $f(\mathbf{x}^k + \tau \mathbf{d}^k) \leq f(\mathbf{x}^k) - \sigma \tau \|\mathbf{d}^k\|^2$  and  $\langle \nabla f(\mathbf{x}^k + \tau \mathbf{d}^k), \mathbf{d}^k \rangle \geq -\rho \|\mathbf{d}^k\|^2$  then
14             return  $\tau$ 
15         else if  $f(\mathbf{x}^k + \tau \mathbf{d}^k) \leq f(\mathbf{x}^k) - \sigma \tau \|\mathbf{d}^k\|^2$  and  $\langle \nabla f(\mathbf{x}^k + \tau \mathbf{d}^k), \mathbf{d}^k \rangle < -\rho \|\mathbf{d}^k\|^2$  then
16              $\tau_0 \leftarrow \tau$ 
17         else
18              $\tau_1 \leftarrow \tau$ 

```

---

# 2 Resultate

Die getesteten Funktionen sind die Rosenbrock-Funktion  $r(x_1, x_2)$ ,  $f(\mathbf{x}) = -\exp(-\|\mathbf{x}\|^2)$ ,  $d(x_1, x_2) = \sin(x_1 x_2)$ ,  $h(\mathbf{x}) = \sum_{i=1}^n x_i$ , und  $g(x_1, x_2) = 2x_1^2 + x_2^2 - 2x_1 x_2 + 2x_1^3 + x_1^4$ . Im Anhang findet sich eine analytische Bestimmung der Minima. Getestet wurden nun:

1. die vergangene Zeit bis eine Abbruchbedingung erreicht wurde
2. welche Abbruchbedingung erreicht wurde

3. ob eine korrekte Minimalstelle ermittelt wurde, und der Fehler zum nächsten Minimum wenn nicht

In der Abgabe findet sich die Datei `testing.py` welche die unten aufgeführten Plots erzeugt. Ziel dabei war es das Verhalten der Algorithmen bei verschiedenen Startpunkten zu prüfen. Dazu wurde jeweils die Umgebung  $[-5, 5]^2$  gewählt mit einer Auflösung von 10000 Punkten. Dazu wurden dazu die folgenden Standard-Werte der Parameter verwendet:

Paramter				Abbruchbedingungen		
$\varrho$	$\eta$	$\beta$	$\sigma$	$\varepsilon$	$M$	$N$
$\frac{3}{4}$	$\frac{3}{2}$	$\frac{3}{4}$	$\frac{1}{2}$	$10^{-9}$	25	100

Tabelle 1: Verwendete Parameter für alle Tests

Da insgesamt 40 Plots erzeugt wurden sind diese nicht in diesem Dokument eingefügt, sie sind aber in der Abgabe enthalten. Zu finden sind sie unter `results/<version>/`.

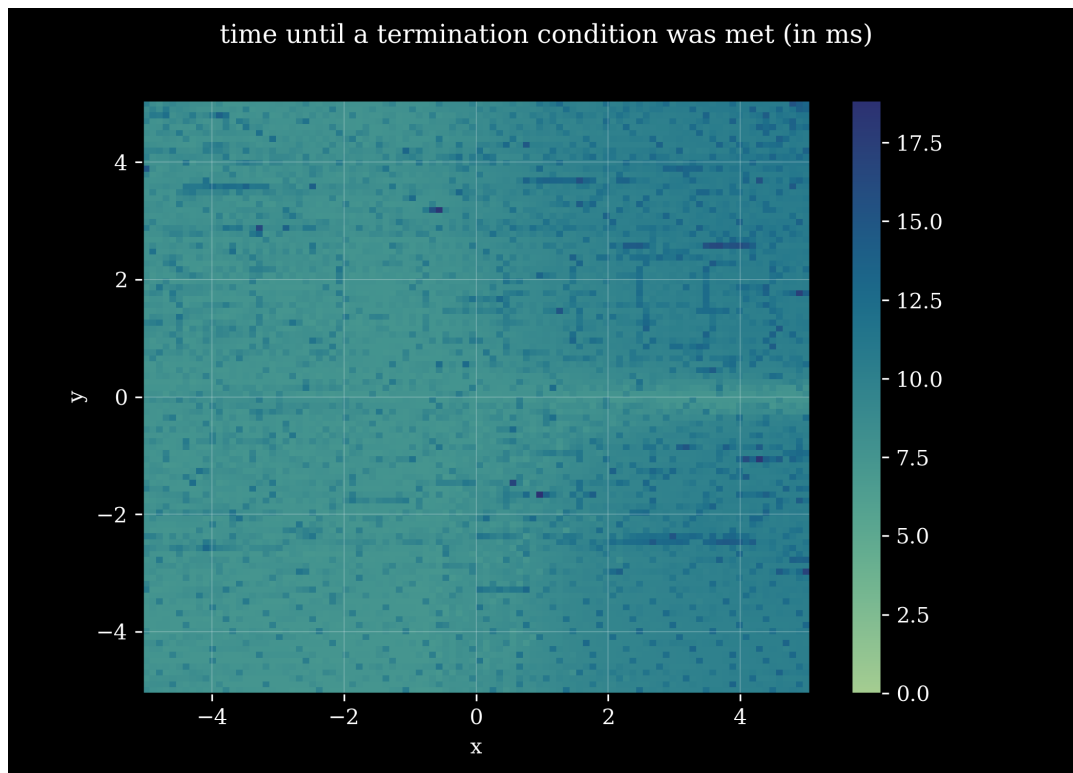


Abbildung 1: Zeit bis eine Abbruchbedingung erfüllt wurde, Armijo-Schrittweite mit Rosenbrock-Funktion

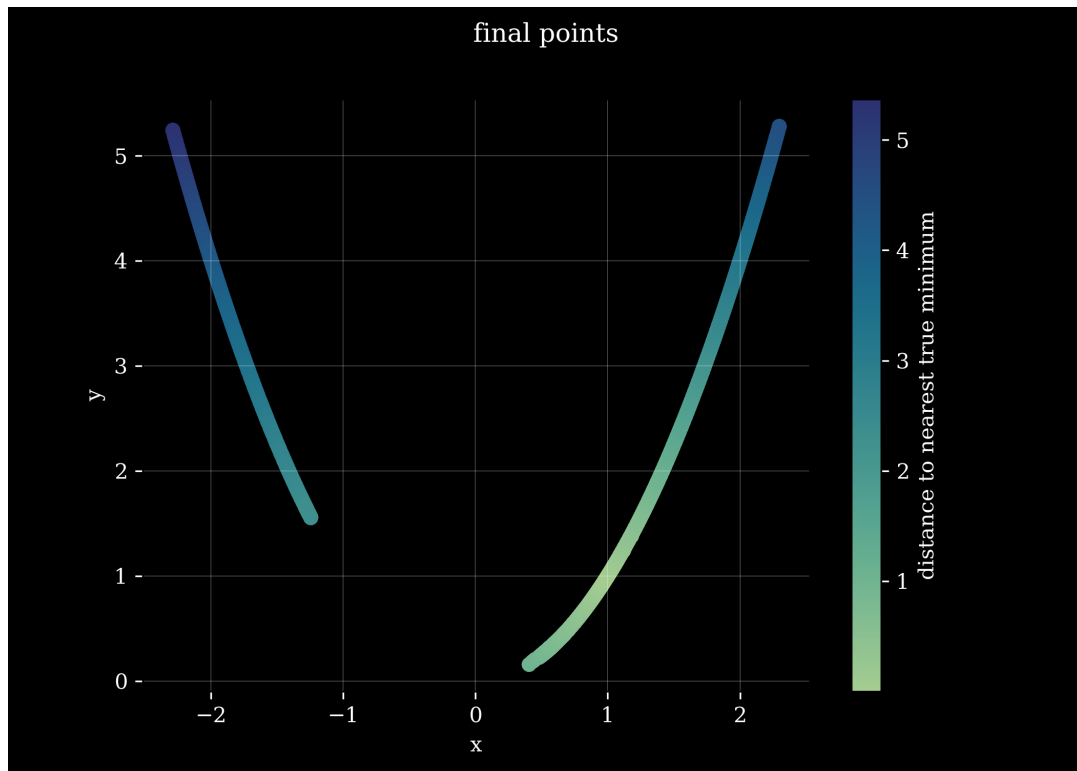


Abbildung 2: Punkte die mittels Armijo-Schrittweite als stationäre Punkte der Rosenbrock-Funktion ermittelt wurden

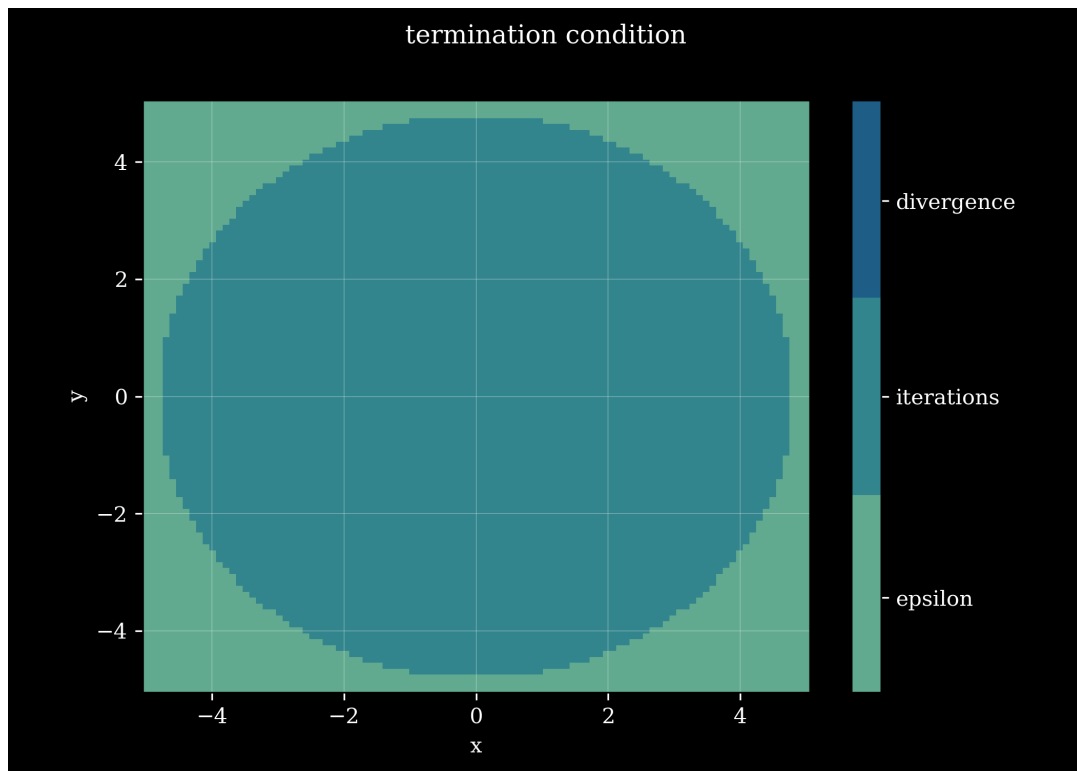


Abbildung 3: Erreichte Abbruchbedingung bei der gemischten Schrittweite und  $f$

Eine weitere Funktion die getestet wurde ist  $q(\mathbf{x}) = \frac{1}{2} \langle \mathbf{x}, \mathbf{Q}\mathbf{x} \rangle + \langle \mathbf{c}, \mathbf{x} \rangle$  wobei

$$\mathbf{Q} = \begin{bmatrix} 14 & 0 & -1 \\ 9 & 18 & 6 \\ -1 & 6 & 5 \end{bmatrix} \quad \mathbf{c} = \begin{bmatrix} 0.5 \\ 1.2 \\ 3.14 \end{bmatrix}$$

Die Armijo-Schrittweite hat für  $N = 10^4$  immer noch die maximale Anzahl der Iterationen überschritten. Allerdings wurde bereits der Punkt

$$\mathbf{x}_A^* = \begin{bmatrix} -0.85279 \\ 1.0432 \\ -2.0405 \end{bmatrix}$$

erreicht. Mit der gemischten Schrittweitenregel wurde der Punkt

$$\mathbf{x}_G^* = \begin{bmatrix} -0.85279 \\ 1.0432 \\ -2.05039 \end{bmatrix}$$

erreicht.

## 3 Anhang

### 3.1 Anmerkungen zu Absicherungen gegen Endlosschleifen

Während der Entwicklung sind einige Probleme numerischer Natur aufgetreten. Diese könnten vermutlich durch geeignete Bibliotheken gelöst werden, allerdings handelte es sich bei den auftretenden Zahlen bereits um Größen im Bereich  $10^{-12}$  und darunter. Bei der Armijo-Regel wird zusätzlich geprüft, dass  $t_k \geq 10^{-12}$  gilt. Ohne diesen Check hat das Verfahren nicht abgebrochen. Bei der gemischten Schrittweitenregel gibt es zwei mögliche Problemstellen. Einerseits bei der Ermittlung von  $t'_k$ . Hier sind teilweise sehr große Exponenten  $i$  aufgetreten, weswegen die Grenze  $i \leq 50$  eingebaut wurde. Im Anschluss wird die Breite des Suchintervalls  $[\tau_0, \tau_1]$  überprüft. Auch hier sind Fälle aufgetreten, in denen  $\tau_1 - \tau_0 < 10^{-12}$  eingetreten ist. In diesem Fall wird ebenfalls abgebrochen.

### 3.2 Verschiedenes

$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^n$	kanonischer $n$ -dimensionaler Vektorraum über $\mathbb{R}$
$\mathcal{C}(D, B)$	stetige Funktionen $f: D \rightarrow B$
$\mathcal{C}^k(D, B)$	$k$ -fach stetig-differenzierbare Funktionen $f: D \rightarrow B$
$Df$ , auch $\mathbf{J}f$	Jacobi-Matrix von $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$
$\nabla f$	Gradient einer Funktion ( $\nabla f = Df^t$ )
$\mathbb{R}^{m \times n}$	$m \times n$ Matrizen
$\mathbb{R}_s^{n \times n}$	symmetrische $n \times n$ Matrizen
$\nabla^2 f$ , auch $\mathbf{H}f$	Hesse-Matrix von $f: \mathbb{R}^n \rightarrow \mathbb{R}$
$\mathbf{A}, \mathbf{M}$	Matrix
$\mathbf{I}$	Einheitsmatrix
$\mathbf{I}_n$	Einheitsmatrix in $\mathbb{R}^{n \times n}$
$\mathbf{0}$	Nullmatrix
$\mathbf{0}_n$	Nullmatrix in $\mathbb{R}^{n \times n}$
$\mathbf{x}, \mathbf{z}$	Vektor
$\mathbf{x}^*$	stationärer Punkt
$\mathbf{0}$	Nullvektor
$\mathbf{0}_n$	Nullvektor in $\mathbb{R}^n$
$\mathbf{1}$	$[1 \ \dots \ 1] \in \mathbb{R}^n$
$\langle \cdot, \cdot \rangle$	Standard-Skalarprodukt in $\mathbb{R}^n$ mit $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^t \mathbf{y}$

Tabelle 2: Symbolindex

### 3.2.1 Analytische Untersuchung der getesteten Funktionen

mit den jeweiligen Ableitungen:

$$\begin{aligned}\nabla r(x_1, x_2) &= \begin{bmatrix} -400(x_2 - x_1^2)x_1 - 2(1 - x_1) \\ 200(x_2 - x_1^2) \end{bmatrix} \\ \nabla^2 r(x_1, x_2) &= \begin{bmatrix} 800x_1^2 - 400(x_2 - x_1^2) + 2 & -400x_1 \\ -400x_1 & 200 \end{bmatrix} \\ \nabla f(\mathbf{x}) &= 2 \exp(-\|\mathbf{x}\|^2) \mathbf{x} \\ \nabla^2 f(\mathbf{x}) &= 2 \exp(-\|\mathbf{x}\|^2) \begin{bmatrix} 1 - 2x_1^2 & -2x_1x_2 & \cdots & -2x_1x_{n-1} & -2x_1x_n \\ -2x_1x_2 & 1 - 2x_2^2 & \cdots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots & -2x_{n-1}x_n \\ -2x_1x_n & -2x_2x_n & \cdots & -2x_{n-1}x_n & 1 - 2x_n^2 \end{bmatrix} \\ \nabla g(x_1, x_2) &= 2 \begin{bmatrix} 2x_1 - x_2 + 3x_1^2 + 2x_1^3 \\ x_2 - x_1 \end{bmatrix} \quad \nabla^2 g(x_1, x_2) = 2 \begin{bmatrix} 2 + 6x_1 + 6x_1^2 & -1 \\ -1 & 1 \end{bmatrix}\end{aligned}$$

Am Gradienten von  $f$  sieht man schnell, dass nur  $\mathbf{0}$  ein stationärer Punkt ist. Die Hesse-Matrix  $\nabla^2 f(\mathbf{0})$  ist offensichtlich positiv definit, da  $\nabla^2 f(\mathbf{0}) = 2\mathbf{I}_n$  gilt. Die Rosenbrock-Funktion  $r$  hat bekanntermaßen bei  $\mathbf{x}^* = (1, 1)$  ihr Minimum. Für  $g$  bestimmen wir aus dem Gradienten  $x_1 = x_2$  für stationäre Punkte und somit

$$\begin{aligned}2x_1 - x_1 + 3x_1^2 + 2x_1^3 &\stackrel{!}{=} 0 \iff x_1(1 + 3x_1 + 2x_1^2) \stackrel{!}{=} 0 \\ \iff x_1(x_1 + 1) \left(x_1 - \frac{1}{2}\right) &= 0 \\ \nabla^2 g(0, 0) = 2 \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} &\rightsquigarrow \lambda_{1,2} = 3 \pm \sqrt{5} \implies \nabla^2 g(0, 0) > 0 \\ \nabla^2 g(-1, -1) = 2 \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} &\rightsquigarrow \lambda_{1,2} = 3 \pm \sqrt{5} \implies \nabla^2 g(0, 0) > 0 \\ \nabla^2 g\left(\frac{1}{2}, \frac{1}{2}\right) = 2 \begin{bmatrix} \frac{13}{2} & -1 \\ -1 & 1 \end{bmatrix} &\rightsquigarrow \lambda_{1,2} = \frac{15 \pm \sqrt{137}}{2} \implies \nabla^2 g\left(\frac{1}{2}, \frac{1}{2}\right) > 0\end{aligned}$$

Die Hesse-Matrix von  $g$  ist also in allen drei stationären Punkten positiv definit und hat somit drei Minima.

$h$  hat offensichtlich keine Minima:

$$\nabla h = \mathbf{1} \quad \nabla^2 h = \mathbf{0}_n$$

Für  $d$  ergibt sich

$$\nabla d(x_1, x_2) = \cos(x_1 x_2) \begin{bmatrix} x_2 \\ x_2 \end{bmatrix} \quad \nabla^2 d(x_1, x_2) = \begin{bmatrix} -x_2^2 \sin(x_1 x_2) & \cos(x_1 x_2) - x_1 x_2 \sin(x_1 x_2) \\ \cos(x_1 x_2) - x_1 x_2 \sin(x_1 x_2) & -x_1^2 \sin(x_1 x_2) \end{bmatrix}$$

Die Gleichung  $\nabla d(\mathbf{x}) = \mathbf{0}$  hat einerseits die Lösung  $\mathbf{x} = \mathbf{0}$  und für  $x_1 \neq 0$  folgt

$$x_2 = \frac{\pi(2k-1)}{2x_1} \quad k \in \mathbb{Z}$$

Offenbar ist  $\mathbf{0}$  kein Minimum, da auf der Geraden  $x_2 = 1$  bei  $x_1 = 0$  kein Minimum vorliegt.

Für  $q(\mathbf{x})$  betrachten wir

$$\nabla q(\mathbf{x}) = \mathbf{Q}\mathbf{x} + \mathbf{c} \stackrel{!}{=} \mathbf{0} \iff \mathbf{Q}\mathbf{x} = \mathbf{c}$$

Da  $\mathbf{Q}$  regulär ist, hat  $q$  nur den stationären Punkt

$$\mathbf{x}^* = \begin{bmatrix} -0.8528 \\ 1.0432 \\ -2.0504 \end{bmatrix}$$

Zusätzlich ist  $\mathbf{Q}$  positiv definit, daher folgt aus  $\nabla^2 q(\mathbf{x}) = \mathbf{Q}$ , dass  $\mathbf{x}^*$  ein Minimum ist.

## Literatur

- [Gei99] Carl Geiger. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. ger. Springer-Lehrbuch. 1999. ISBN: 9783540662204.