4a. Newton-artige Verfahren
Das Newton-Verfahren

Optimierung SoSe 2020

Dr. Alexey Agaltsov

Plan

- Die Newton-Richtung
- Das Newton-Verfahren
- Konvergenz

Mathematisches Programm

Minimiere $f(x)$ über $x \in U$

$f \in C^2(U),\ U \subseteq \mathbb{R}^n$ offen

- Man kann dieses Problem mit dem Gradientenverfahren lösen
- Das benötigt nur die ersten Ableitungen von f und garantiert lineare Konvergenz
- Benutzt man auch die zweiten Ableitungen, so kann man ein effizienteres Verfahren erstellen

Newton-Verfahren

Das Newton-Verfahren

Das Newton-Verfahren lässt verschiedene Herleitungen zu:

- Löse die Optimalitätsbedingung $\nabla f(x) = 0$ für x_* mit dem Newton-Verfahren für nichtlineare Gleichungen
- Minimiere konsekutive quadratische Approximationen an f

Newton-Verfahren für nichtlineare Gleichungen

Löse $g(x) = 0$

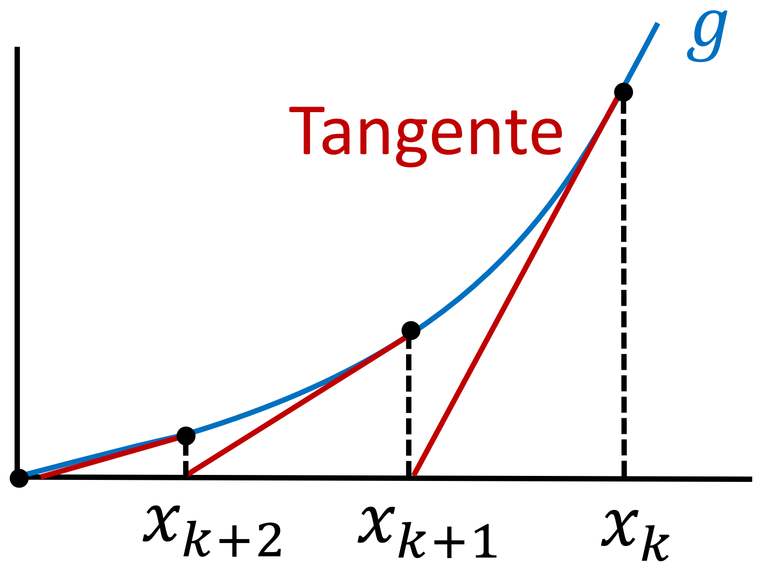
$g \in C^1(\mathbb{R})$

linearisiere

$$g(x_k) + g'(x_k)(x - x_k) = 0$$

aktuelle Näherungslösung

$$x_{k+1} := x_k - \frac{g(x_k)}{g'(x_k)}$$



Newton-Verfahren für Optimierungsprobleme

Minimiere $f(x)$ über $x \in \mathbb{R}$

$f \in C^2(\mathbb{R})$

Optimalitätsbedingung

$$f'(x) = 0$$

Newton-Verfahren

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$

Quadratische Approximation

Minimiere $f(x)$ über $x \in \mathbb{R}$

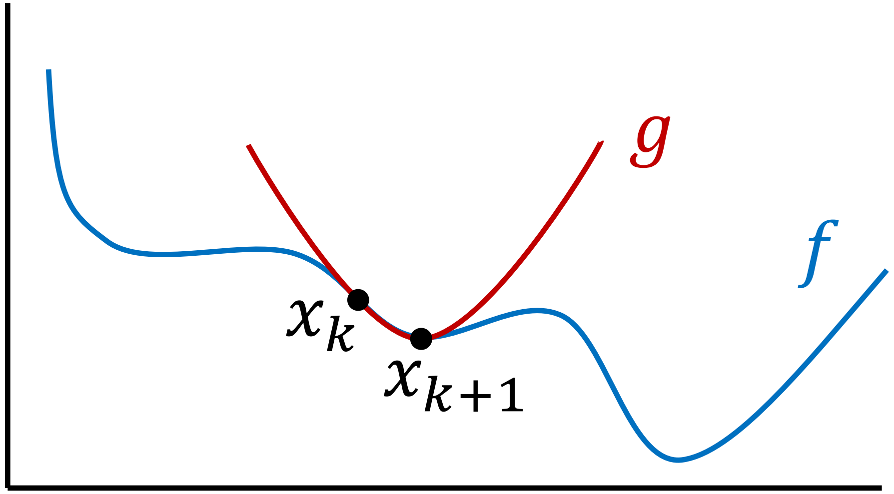
$f \in C^2(\mathbb{R})$

quadratische Approximation

$$\text{Minimiere } g(x) = f(x_k) + f'(x_k)(x - x_k) + \frac{1}{2} f''(x_k)(x - x_k)^2$$

aktuelle Approximation

$$x_{k+1} = x_k - \frac{f'(x_k)}{f''(x_k)}$$



Mehrdimensionaler Fall

Minimiere $f(x)$ über $x \in \mathbb{R}^n$

$f \in C^2(\mathbb{R}^n)$

quadratische Approximation

Minimiere $g(x)$ über $x \in \mathbb{R}^n$

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

aktuelle Approximation

Optimalitätsbedingung

$$\nabla f(x_k) + \nabla^2 f(x_k)(x_{k+1} - x_k) = 0$$

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

Die Newton-Schritt

- Als Newton-Schritt bezeichnen wir die Aufdatierungsformel:

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

- Als Newton-Richtung im Punkt x_k bezeichnen wir:

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

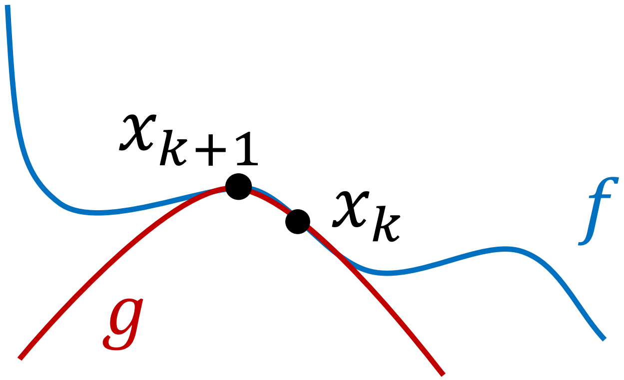
Die Newton-Richtung

$$d_k = -\nabla^2 f(x_k)^{-1} \nabla f(x_k), \quad \nabla f(x_k) \neq 0$$

- Sei $\nabla^2 f(x_k) \succ 0$, so ist die Newton-Richtung d_k eine Abstiegsrichtung:

$$-\nabla f(x_k)^T d_k = \nabla f(x_k)^T \nabla^2 f(x_k)^{-1} \nabla f(x_k) > 0$$

Vorsicht: *Ist* $\nabla^2 f(x_k) \nsucc 0$, *so kann* d *keine Abstiegsrichtung sein*



Plan

- Die Newton-Richtung
- Das Newton-Verfahren
- Quadratische Konvergenz

Koordinatentransformation

Minimiere $f(x)$ über $x \in \mathbb{R}^n$
$f \in C^2(\mathbb{R}^n)$

$\bar{x} := P^{1/2}x$ mit $P \in \mathbb{S}^n_{\succ}$
$\bar{f}(\bar{x}) := f\left(P^{-1/2}\bar{x}\right) = f(x)$

Minimiere $\bar{f}(\bar{x})$ über $\bar{x} \in \mathbb{R}^n$

- Die Newton-Richtung in $\bar{x}$ ist:

$$\bar{d} = -\nabla^2 \bar{f}(\bar{x})^{-1} \nabla \bar{f}(\bar{x}) = -P^{1/2} \nabla^2 f(x)^{-1} \nabla f(x)$$

- Die entsprechende Richtung in x ist:

$$d = P^{-1/2}\bar{d} = -\nabla^2 f(x)^{-1} \nabla f(x)$$

Die Newton-Richtung ist linear- (und sogar affin-) invariant

Das NV ist unempfindlich ggb. der Kondition der Unterniveau-Mengen

Abbruchsregel

- Sei x_k eine aktuelle Näherungslösung und $\varepsilon > 0$ ein Toleranzwert
- Gewöhnliches Abbruchskriterium $||\nabla f(x_k)||_2 < \varepsilon$
- Es wäre wünschenswert, eine affin-invariante Abbruchsregel bei dem Newton-Verfahren zu haben

Das Newton-Dekrement

- Als Newton-Dekrement am Punkt x_k bezeichnen wir:

$$\begin{aligned}\lambda(x_k) &= \|\nabla f(x_k)\|_{\nabla^2 f(x_k)^{-1}} \\ &= \left(\nabla f(x_k)^T \nabla^2 f(x_k)^{-1} \nabla f(x_k)\right)^{1/2}\end{aligned}$$

- Das Newton-Dekrement ist affin-invariant
- Affin-invariante Abbruchsregel:

$$\lambda(x_k) < \varepsilon \implies \text{Breche die Iteration ab}$$

Das Newton-Verfahren

1. *Initialisierung:* Startwert $x_0 \in \mathbb{R}^n$, Toleranzwert $\epsilon > 0$
2. **for** $k = 0,1,2,\ldots$ **do**:
3. **if** $\lambda(x_k) < \epsilon$ **then** break
4. $x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k)$
5. **end for**

Beispiel: 1D

Minimiere $f(x) = x - \log x$ über $x > 0$

$f'(x) = 1 - \frac{1}{x} \implies x_* = 1$ ist stationär

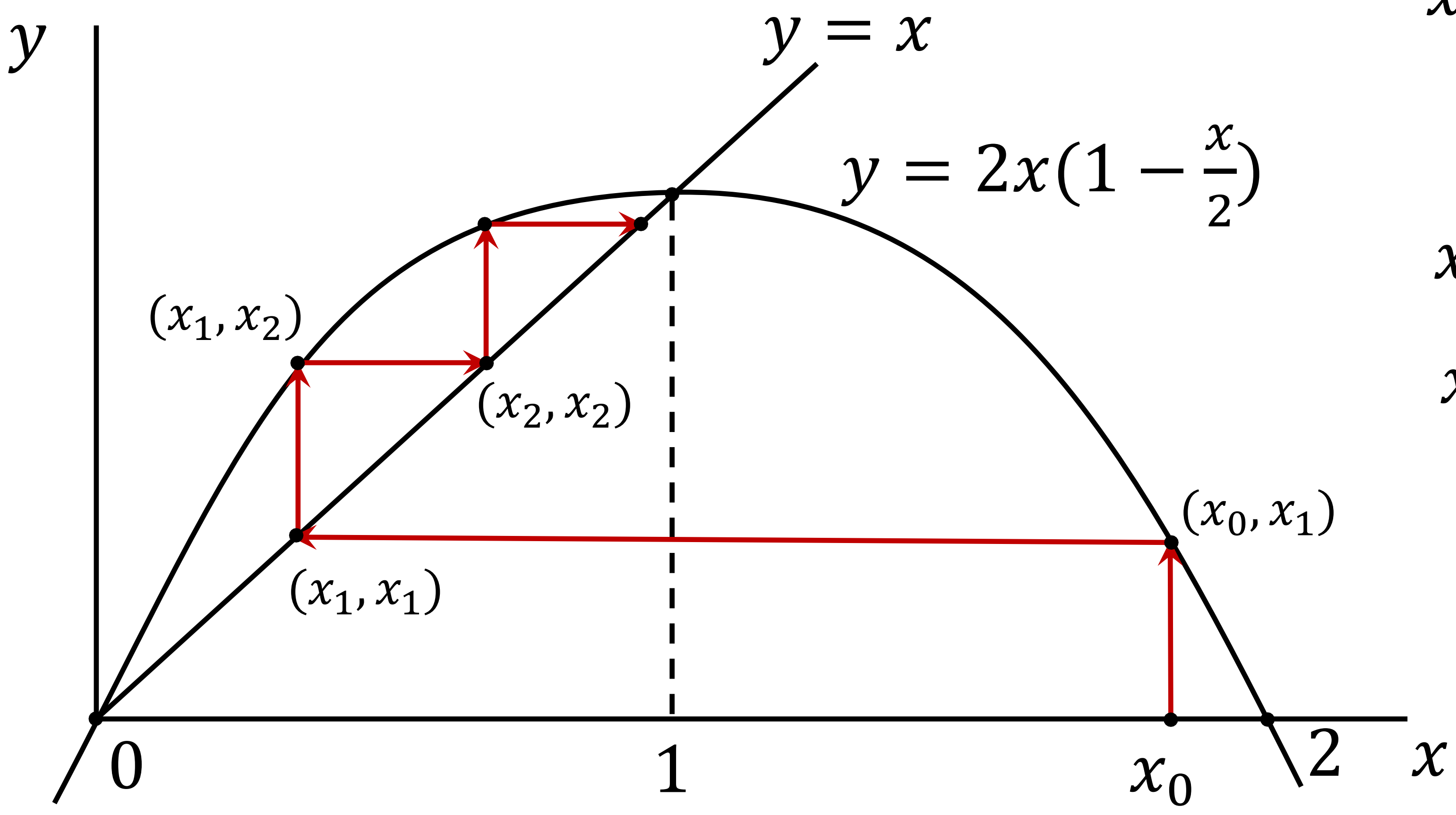
$f''(x) = \frac{1}{x^2} > 0 \implies f$ ist strikt konvex

x_* ist die optimale Lösung

$$d(x) = -\left(\frac{1}{x^2}\right)^{-1}\left(1 - \frac{1}{x}\right) = x - x^2$$

$$x_{k+1} = x_k + \left(x_k - x_k^2\right) = 2x_k - x_k^2$$

Beispiel: 1D



$$x_{k+1} = 2x_k(1 - \frac{x_k}{2})$$

$$x_0 \in (0,2) \Rightarrow x_k \to 1$$

$$x_0 = 2 \Rightarrow x_k \equiv 2$$

Beispiel: 2D

Minimiere $f(x_1, x_2) = -\log(1 - x_1 - x_2) - \log(x_1) - \log(x_2)$

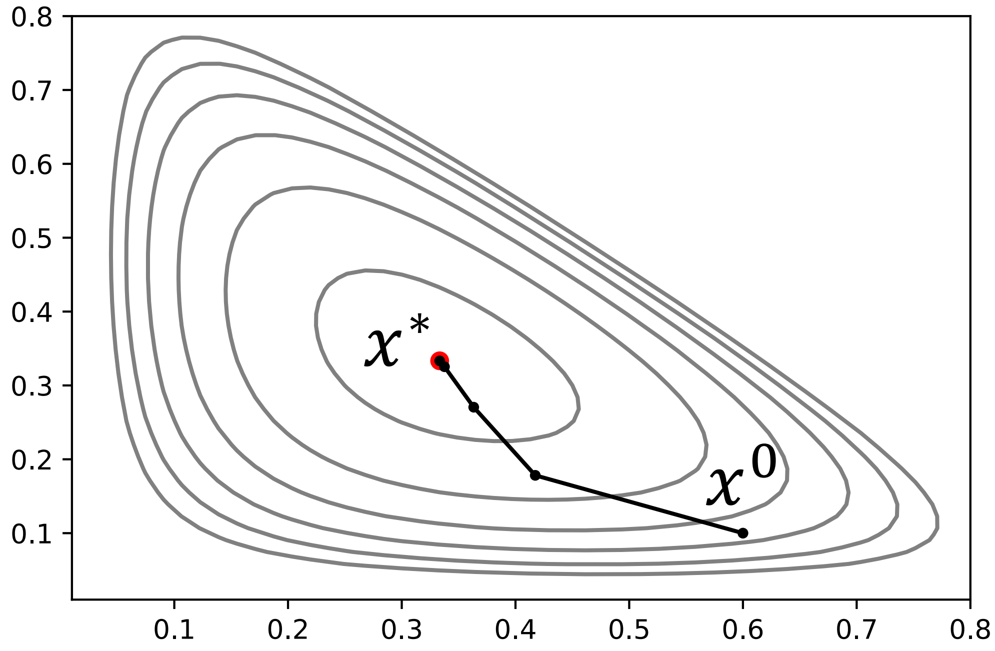
über $x_1, x_2 > 0, x_1 + x_2 < 1$

$$\nabla f(x_1, x_2) = \begin{bmatrix} \frac{1}{1-x_1-x_2} - \frac{1}{x_1} \\ \frac{1}{1-x_1-x_2} - \frac{1}{x_2} \end{bmatrix} \Longrightarrow \text{stationärer Punkt } x_1^* = x_2^* = \frac{1}{3}$$

$$\nabla^2 f(x_1, x_2) = \begin{bmatrix} \left(\frac{1}{1-x_1-x_2}\right)^2 + \left(\frac{1}{x_1}\right)^2 & \left(\frac{1}{1-x_1-x_2}\right)^2 \\ \left(\frac{1}{1-x_1-x_2}\right)^2 & \left(\frac{1}{1-x_1-x_2}\right)^2 + \left(\frac{1}{x_2}\right)^2 \end{bmatrix} \succ 0$$

$\Longrightarrow f$ ist konvex $\quad$ $\left(\frac{1}{3}, \frac{1}{3}\right)$ ist die optimale Lösung

Beispiel: 2D



$$x^0 = (0.6, 0.1), \qquad x^* = \left(\tfrac{1}{3}, \tfrac{1}{3}\right)$$

k	$\lVert x^k - x^* \rVert_2$	$f(x^k) - f_*$	c^k
0	0,35	0,72	
1	0,18	0,21	0,4
2	0,07	0,029	0,66
3	0,009	0,00044	0,53
4	0,00011	6,9E-08	0,36

$$f(x^k) - f_* = c^k \left(f(x^{k-1}) - f_*\right)^2$$

Wir werden auch theoretisch zeigen, dass die Konvergenz quadratisch ist

Plan

- Die Newton-Richtung
- Das Newton-Verfahren
- Konvergenz

Mathematisches Program

$$\text{Minimiere } f(x) \text{ über } x \in \mathbb{R}^n$$
$$f \in C^2(\mathbb{R}^n)$$

- Sei x_* eine optimale Lösung
- Sei $x_0, x_1, \ldots$ eine mit dem Newton-Verfahren erzeugte Folge
- Was bestimmt die Konvergenz und ihre Geschwindigkeit?

Was bestimmt die Konvergenz?

Hauptfaktoren, die die Konvergenz bestimmen:

- Qualität der quadratischen Approximation
- Wie nahe x_0 an x_* liegt
- Positive Definitheit der Hesse-Matrix

Qualität der quadratischen Approximation

$$f(x) \approx f(x_k) + \nabla f(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T \nabla^2 f(x_k)(x - x_k)$$

- Die Approximation ist gut falls $\nabla^2 f(x)$ sich langsam ändert
- Lipschitz-Stetigkeit von $\nabla^2 f(x)$ quantifiziert es durch $L > 0$:

$$\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2$$

Beispiel: Quadratische Zielfunktion

Minimiere $f(x) = \frac{1}{2} x^T Q x - c^T x + r \quad$ mit $Q \in \mathbb{S}^n_{\succ}$

$$\nabla f(x) = Qx - c$$
$$\nabla^2 f(x) = Q \succ 0$$

- Nach Aufgabe 2.21 ist $x_* = Q^{-1} c$ ein eindeutiges globales Minimum
- Das Newton-Verfahren findet x_* in einem Schritt $\forall x_0$:

$$\begin{aligned} x_1 &= x_0 - Q^{-1}(Q x_0 - c) \\ &= Q^{-1} c \end{aligned}$$

Wahl des Startwertes

- Die Methode des steilsten Abstiegs bzgl. $||\cdot||_P$ konvergiert schnell falls

$$P^{-1/2}(\nabla^2 f)P^{-1/2} \approx I \quad \text{auf } S = \{x: f(x) \leq f(x_0)\}$$

- Ist $x_0 \approx x_*$, so ist $P = \nabla^2 f(x_*)$ eine optimale Wahl

- Die Newton-Richtung in x ist die Richtung des steilsten Abstiegs bzgl. $||\cdot||_P$ mit $P = \nabla^2 f(x)$

schnelle Konvergenz für $x_0 \approx x_*$

Positive Definitheit der Hesse-Matrix

$$d = -\nabla^2 f(x)^{-1} \nabla f(x)$$

- $\nabla^2 f(x)$ muss regulär sein, damit ist d wohldefiniert
- $\nabla^2 f(x) \succ 0$ garantiert, dass d eine Abstiegsrichtung ist
- $\nabla^2 f(x) \succcurlyeq mI$ lässt die Konvergenzgeschwindigkeit abschätzen

Konvergengeschwindigkeit steigt mit m

Satz 4.1. Quadratische Konvergenz

Sei $f \in C^2(\mathbb{R}^n)$ und sei x_* ein lokales Minimum. Angenommen:

- $\nabla^2 f(x) \succcurlyeq mI \succ 0 \quad \forall x \in B_\delta(x_*)$
- $\|\nabla^2 f(x) - \nabla^2 f(y)\|_2 \leq L\|x - y\|_2 \quad \forall x, y \in B_\delta(x_*)$

Qualität der quadratischen Approximationen

Ist $\|x_0 - x_*\|_2 < \min\left(\delta, \frac{2m}{L}\right)$, so gilt:

$$\|x_{k+1} - x_*\|_2 \leq \frac{L}{2m}\|x_k - x_*\|^2 \quad \forall k \geq 0$$

Beweis

$$r_k := x_* - x_k$$

Ziel: Drücke r_{k+1} durch r_k aus

$$r_{k+1} = r_k + \nabla^2 f(x_k)^{-1} \nabla f(x_k)$$

$\nabla f(x_*) = 0$

$$= r_k - \nabla^2 f(x_k)^{-1} (\nabla f(x_*) - \nabla f(x_k))$$

Taylor-Formel

$$= r_k - \nabla^2 f(x_k)^{-1} \int_0^1 \nabla^2 f(x_k + r_k t)\, r_k\, dt$$

$$= \nabla^2 f(x_k)^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x_k + r_k t)] r_k\, dt$$

Beweis

Behauptung: $\|r_{k+1}\|_2 \leq \frac{L}{2m}\|r_k\|_2^2$

$$r_{k+1} = \nabla^2 f(x_k)^{-1} \int_0^1 [\nabla^2 f(x_k) - \nabla^2 f(x_k + r_k t)] r_k \, dt$$

$$\|r_{k+1}\|_2 \leq \underbrace{\|\nabla^2 f(x_k)^{-1}\|_2}_{\preccurlyeq \frac{1}{m}} \int_0^1 \underbrace{\|\nabla^2 f(x_k) - \nabla^2 f(x_k + r_k t)\|_2}_{\leq L\|r_k\|_2 t} \|r_k\|_2 \, dt$$

$$\|r_{k+1}\|_2 \leq \frac{L}{m} \int_0^1 \|r_k\|_2^2 \, t \, dt \leq \frac{L}{2m}\|r_k\|_2^2$$

Globale Konvergenz

- Das Newton-Verfahren entspricht den Schrittweiten $\alpha_k \equiv 1$
- Liegt x_0 weit von x_* so kann die Konvergenz fehlschlagen
- Schrittweitenstrategien können die Konvergenz globalisieren

Gedämpftes Newton-Verfahren

Zusammenfassung

- Die Newton-Richtung
- Das Newton-Verfahren
- Quadratische Konvergenz

Nächstes Video

- 4b. Newton-artige Verfahren: Quasi-Newton-Verfahren