Optimierung

Vorlesung 4

Newton- und Quasi-Newton-Verfahren (Teil II)

Newton-Verfahren

Taylor-Approximation 1. Ordnung von ∇f

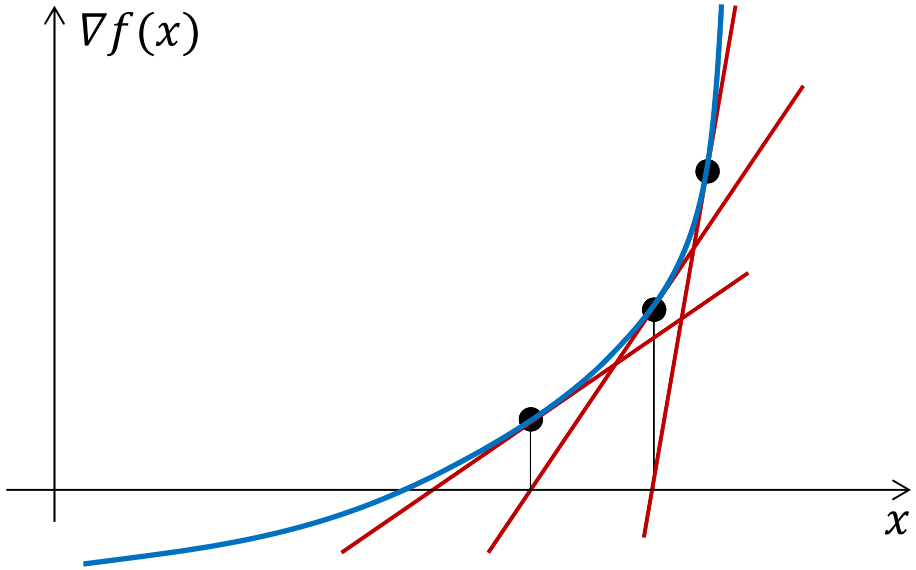
$$\nabla f(\underline{x_k + d_k}) \approx \nabla f(\underline{x_k}) + H(x_k) \cdot d_k$$

Newton-Verfahren

$$\underline{0 \stackrel{!}{=} \nabla f(x_k + d_k)} \approx \underline{\nabla f(x_k) + H(x_k) \cdot d_k}$$

Newton-R.

- Setze $\nabla f(x_k) + H(x_k) d_k = 0$ und berechne $\underline{\boldsymbol{x_{k+1} = x_k + d_k}}$
- Löse lineares Gleichungssystem für d_k: $\boldsymbol{d_k = \underline{-\underline{H(x_k)}^{-1} \cdot \nabla f(x_k)}}$



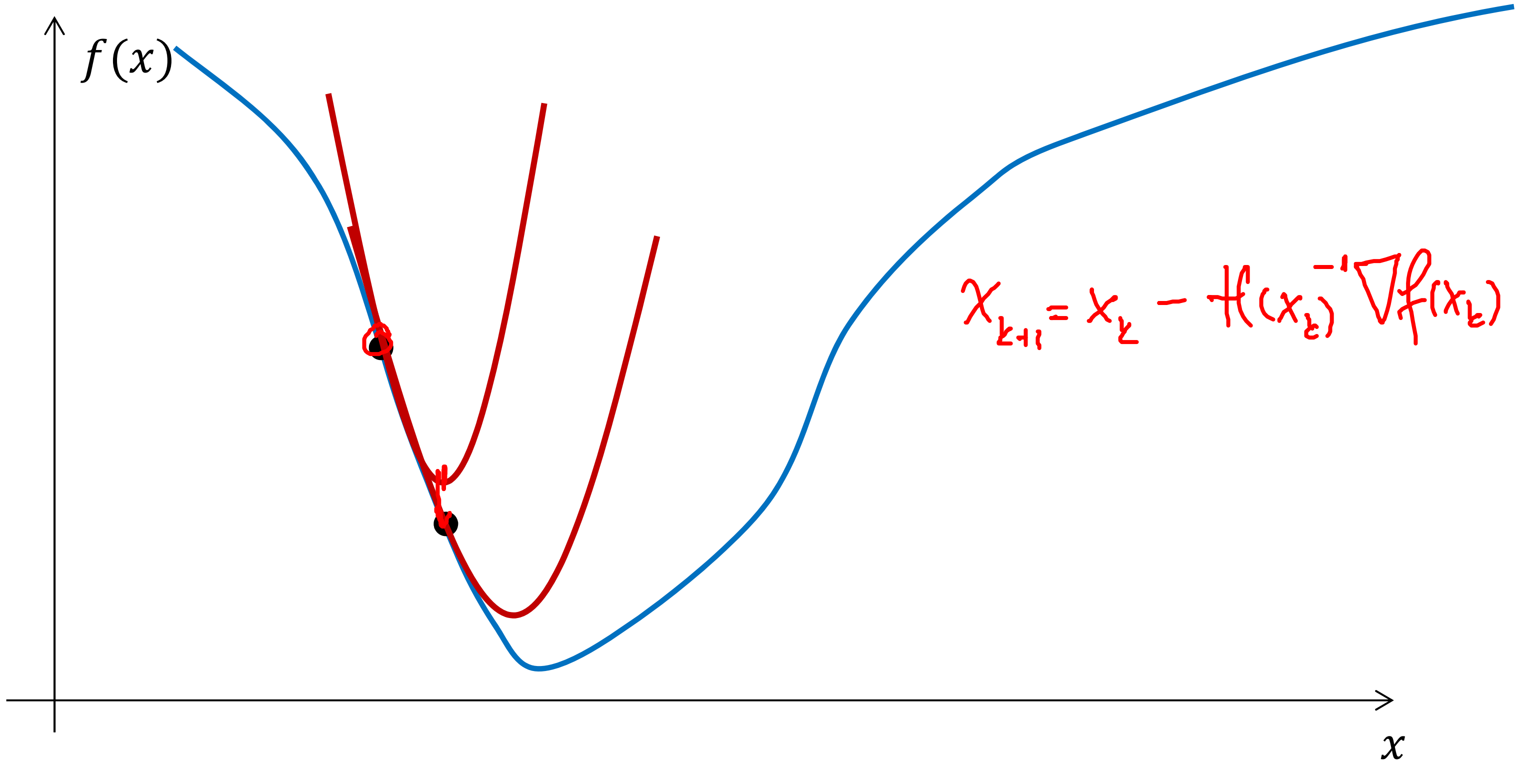
Newton-Verfahren (alternative Sicht)

Taylor-Approximation 2. Ordnung von f

$$f(x_k + d_k) \approx f(x_k) + d_k^\top \cdot \nabla f(x_k) + \frac{1}{2} d_k^\top \cdot H(x_k) \cdot d_k$$

Newton-Verfahren

- Neuer Wert x_{k+1}: minimiere 2. Ordnung Taylor-Approximation



Fortschritt beim Newton-Verfahren

Abstiegsrichtung

- Eine Richtung d ist eine Abstiegsrichtung für einen Wert x_k, falls

$$d^\top \nabla f(x_k) < 0$$

  - Beispiel: Neg. Gradient $d = -\nabla f(x_k)$

- Taylor-Approximation: $f(x_k + \varepsilon \cdot d) \approx f(x_k) + \varepsilon \cdot d^\top \nabla f(x_k)$

Newton-Richtung $\boldsymbol{d_k = -H(x_k)^{-1} \nabla f(x_k)}$

- Falls $H(x_k)$ positiv definit ist, ist d_k eine Abstiegsrichtung

Fortschritt beim Newton-Verfahren

Konvexe Funktionen $\boldsymbol{f(x)}$:

- Newton-Richtung ist eine Abstiegsrichtung
- Newton-Verfahren konvergiert zu lokalem Minimum

Konvexe, quadratische Funktionen $\boldsymbol{f(x)}$:

- Newton-Schritt geht zu Minimum der quadratischen Approx. von $f(x)$
- Verfahren erreicht lokales Minimum in einem Schritt!

Allgemein:

- Konvergiert sobald man *nahe genug* bei einem lokalen Minimum ist
- Kann mit ***line search*** (wie beim Gradientenverfahren) kombiniert werden:

Newton-Richtung: $d_k = -H(x_k)^{-1} \cdot \nabla f(x_k)$

Bestimme $x_{k+1} = x_k + \tau_k \cdot d_k$, so dass $f(x_k + \tau_k d_k)$ minimal ist

Konvergenz

- Wie schnell kommt man zum lokalen Minimum?

Konvergenz von Folgen $s_1, s_2, \ldots,$ $\lim_{k\to\infty} s_k = \bar{s}$

Lineare Konvergenz:

$$\limsup_{k\to\infty} \frac{|s_{k+1} - \bar{s}|}{|s_k - \bar{s}|} = C$$

$s_i = 1/2^i$

Superlineare Konvergenz:

$$\limsup_{k\to\infty} \frac{|s_{k+1} - \bar{s}|}{|s_k - \bar{s}|} = 0$$

$s_i = 1/i!$

Quadratische Konvergenz:

$$\boxed{\limsup_{k\to\infty} \frac{|s_{k+1} - \bar{s}|}{|s_k - \bar{s}|^2} = C}$$

$s_i = \frac{1}{2^{2^i}}$

Konvergenz der wichtigsten Verfahren

Gradientenverfahren (steilster Abstieg)

- lineare Konvergenz

Newtonverfahren

- quadratische Konvergenz
- minimiert quadratische Funktionen in 1 Schritt

Konjugierte Gradienten

- minimiert n-dimensionale quadratische Funktionen in n Schritten
- quadratische Konvergenz alle n Schritte (superlinear)

Quasi-Newtonverfahren

- superlineare Konvergenz
- einfacher als Newton-Verfahren

Quadratische Konvergenz des Newton-Verfahren

Ziel: Die Sequenz $x_1, x_2, \ldots$ konvergiert quadratisch gegen x^* ← stat. Punkt

$\nabla f(x^*) = 0$

Matrix-Norm $\|M\|$ einer Matrix M

$\|x\| = \sqrt{\sum x_i^2}$

$$\|M\| := \max_{x:\|x\|=1} \|Mx\|$$

- $\forall M, M', x$: $\|Mx\| \leq \|M\| \cdot \|x\|$, $\|M \cdot M'\| \leq \|M\| \cdot \|M'\|$

$|f''(x) - f''(y)| \leq L \cdot |x-y|$

Annahmen: Für x, y nahe genug bei x^* gilt $\|\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{y})\| \leq \boldsymbol{L} \cdot \|\boldsymbol{x} - \boldsymbol{y}\|$, sowie $\|\boldsymbol{H}(\boldsymbol{x})^{-1}\| \leq \boldsymbol{C}$ für Konstanten $L, C \geq 0$

Lemma:

$$\nabla f(z) - \nabla f(x) = \int_0^1 \left[H\big(x + t(z-x)\big)\right](z-x)\,dt$$

Quadratische Konvergenz des Newton-Verfahren

Lemma:

$\phi(1) \quad - \quad \phi(0)$

$$\nabla f(z) - \nabla f(x) = \int_0^1 \left[H\big(x + t(z-x)\big)\right](z-x)\,dt$$

Beweis:

- Definiere $\phi(t) := \nabla f\big(x + t(z-x)\big)$ $\quad \phi(0) = \nabla f(x)\,,\ \phi(1) = \nabla f(z)$
- Ableitung $\phi'(t) = \left[H\big(x + t(z-x)\big)\right](z-x)$

$$\nabla f(z) - \nabla f(x) = \phi(1) - \phi(0) = \int_0^1 \phi'(t)\,dt = \int_0^1 \left[H(x + t(z-x))\right](z-x)\,dt$$

Quadratische Konvergenz des Newton-Verfahren

Ziel: Die Sequenz $x_1, x_2, \ldots$ konvergiert quadratisch gegen x^* $\quad \nabla f(x^*) = 0$

$x_i : x \qquad x_i - x^* \qquad x_{i+1} - x^*$

$x_{i+1} : x'$

Lemma:

$$\nabla f(z) - \nabla f(x) = \int_0^1 \left[H\big(x + t(z-x)\big)\right](z-x)\,dt$$

■

$x' = x - H(x)^{-1} \nabla f(x)$

$$x' - x^* = x - H(x)^{-1} \nabla f(x) - x^*$$

$$= x - x^* + H(x)^{-1}\big(\nabla f(x^*) - \nabla f(x)\big)$$

$$= x - x^* + H(x)^{-1} \int_0^1 [H(x + t(x^* - x))](x^* - x)\,dt$$

$$= H(x)^{-1} \int_0^1 \big[H(x + t(x^* - x)) - H(x)\big](x^* - x)\,dt$$

$$x - x^* = H(x)^{-1} H(x)(x - x^*)$$

$$= H(x)^{-1} \cdot \int_0^1 H(x)(x - x^*)\,dt$$

$$\int_0^1 c\,dt = c$$

Quadratische Konvergenz des Newton-Verfahren

Matrix-Norm $\|M\|$ einer Matrix M

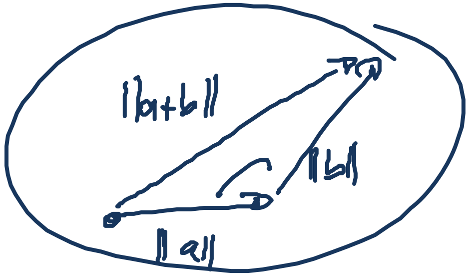
- $\forall M, M', x$: $\|Mx\| \le \|M\| \cdot \|x\|$, $\quad \|M \cdot M'\| \le \|M\| \cdot \|M'\|$

Annahmen: Für x, y nahe genug bei x^* gilt $\|\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{y})\| \le \boldsymbol{L} \cdot \|\boldsymbol{x} - \boldsymbol{y}\|$,
sowie $\left\|\boldsymbol{H}(\boldsymbol{x})^{-1}\right\| \le \boldsymbol{C}$ für Konstanten $L, C \ge 0$

$$x' - x^* = H(x)^{-1} \cdot \int_0^1 \left[H\big(x + t(x^* - x)\big) - H(x)\right](x^* - x)dt$$

$$\|x' - x^*\| \le \|H(x)^{-1}\| \int_0^1 \left\|H\big(x + t(x^* - x)\big) - H(x)\right\| \cdot \|x^* - x\|dt$$

$$\|x' - x^*\| \le \|H(x)^{-1}\| \cdot \left\| \int_0^1 \ldots dt \right\|$$



$$\left\| \int_0^1 [H(x+t(x^*-x)) - H(x)](x^*-x)\, dt \right\| \le \int_0^1 \| (H(x \ldots) - H(x))(x^*-x) \| dt$$

$$\|a+b\| \le \|a\| + \|b\|$$

Quadratische Konvergenz des Newton-Verfahren

Matrix-Norm $\|M\|$ einer Matrix M

- $\forall M, M', x$: $\|Mx\| \leq \|M\| \cdot \|x\|$, $\|M \cdot M'\| \leq \|M\| \cdot \|M'\|$

Annahmen: Für x, y nahe genug bei x^* gilt $\|\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{H}(\boldsymbol{y})\| \leq \boldsymbol{L} \cdot \|\boldsymbol{x} - \boldsymbol{y}\|$,
sowie $\|\boldsymbol{H}(\boldsymbol{x})^{-1}\| \leq \boldsymbol{C}$ für Konstanten $L, C \geq 0$

$$\|x' - x^*\| \leq \|H(x)^{-1}\| \int_0^1 \|H(x + t(x^* - x)) - H(x)\| \cdot \|x^* - x\| dt$$

$$\leq \|x^* - x\| \|H(x)^{-1}\| \int_0^1 \underbrace{\|H(x + t(x^* - x)) - H(x)\|}_{\leq L \cdot \|t(x^* - x)\|} dt$$

$$\leq \|x^* - x\| \|H(x)^{-1}\| \int_0^1 L \cdot t \cdot \|x^* - x\| dt = \|x^* - x\|^2 \cdot \underbrace{\|H(x)^{-1}\|}_{\leq C} \cdot L \cdot \int_0^1 t \, dt$$

$$\|x' - x^*\| \leq \frac{L \cdot C}{2} \cdot \|x - x^*\|^2 \qquad \lim_{i \to \infty} \frac{\|x^{i+1} - x^*\|}{\|x^i - x^*\|^2} \leq \frac{LC}{2}$$

Gradientenverfahren vs. Newton-Verfahren

Gradientenverfahren

- Lineare Konvergenz
  - Benötigt i.A. deutlich mehr Schritte (vor allem, wenn die Hesse'sche Matrix schlecht konditioniert ist)
- 1. Ableitung nötig
- 1 Iterationsschritt:
  - Evaluation Gradient + line search
- "Billige" Schritte, "schlechte" Konv.
- Bessere Konvergenz:

  Konjugierte Gradientenverfahren

Newton-Verfahren

- Quadratische Konvergenz
  - konvergiert in wenigen Schritten (bei quadr. Problem in 1 Schritt)
- 1. & 2. Ableitung nötig
- 1 Iterationsschritt: $-H^{-1}\nabla f$
  - Evaluation Gradient + Hesse'sche Matrix
  - n-dim. lineares Gl.-system
- "Teure" Schritte, "gute" Konvergenz
- Einfachere Iterationen:

  Quasi-Newton-Verfahren

Quasi-Newton-Verfahren

- Newton-Verfahren + Line Search:

$$x_{k+1} = x_k + \tau_k d_k, \qquad \text{wobei } d_k = -H(x_k)^{-1} \nabla f(x_k)$$

- **Idee:** Ersetze $H(x_k)^{-1}$ durch eine Matrix D_k, welche einfacher berechnet werden kann

$D_k \approx H(x_k)^{-1}$

- Taylor-Approximation: $\boxed{\nabla f(x_{k+1}) - \nabla f(x_k) \approx H(x_{k+1})(x_{k+1} - x_k)}$

$\nabla f(x+\delta) \approx \nabla f(x) + H(x) \cdot \delta \qquad f'(x+\delta) \approx f'(x) + \delta \cdot f''(x)$

  - Idee: D_{k+1}^{-1} sollte dies auch erfüllen...

- Quasi-Newton Bedingung:

$$\boxed{\boldsymbol{D_{k+1} \cdot q_k = p_k}}, \quad \text{wobei} \quad \boldsymbol{p_k = x_{k+1} - x_k}, \ \boldsymbol{q_k = \nabla f(x_{k+1}) - \nabla f(x_k)}$$

$$\underbrace{\nabla f(x_{k+1}) - \nabla f(x_k)}_{q_k} = D_{k+1}^{-1} \underbrace{(x_{k+1} - x_k)}_{p_k} \qquad q_k = D_{k+1}^{-1} p_k$$

Quasi-Newton-Verfahren:

$\approx H(x_k)^{-1}$

$$\boldsymbol{x_{k+1} = x_k + \tau_k \cdot d_k,} \quad \textbf{wobei} \quad \boldsymbol{d_k = -D_k \nabla f(x_k)}$$

– τ_k wird typischerweise durch line search ermittelt

Quasi-Newton Bedingung:

$D_1, D_2, \ldots$

$$D_{k+1} \cdot q_k = p_k, \quad \text{wobei} \quad p_k = x_{k+1} - x_k, \; q_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

Berechnung von D_{k+1}:

- Alte Matrix + Korrekturterm: $D_{k+1} = D_k + C_k$

$$(D_k + C_k) q_k = p_k \quad \Longrightarrow \quad C_k q_k = p_k - D_k q_k$$

- Zum Beispiel ($\phi \in [0,1]$ ist ein Parameter)

$$C(\phi) = \frac{pp^\top}{p^\top q} - \frac{Dqq^\top D}{q^\top Dq} + \phi\eta vv^\top, \qquad \text{wobei } v = \frac{p}{p^\top q} - \frac{Dq}{\eta}, \eta = q^\top Dq$$

Quasi-Newton-Verfahren:

$H(x_k)^{-1}$

$$x_{k+1} = x_k - \tau_k \cdot D_k \nabla f(x_k)$$

$$D_{k+1} = D_k + C_k, \qquad C_k q_k = p_k - D_k q_k$$

$D_{k+1} q_k = p_k$

- Parameter $\phi = 0$ (Davidson-Fletcher-Powell, erstes Quasi-Newton-Verf.)

$$C(0) = \frac{pp^\top}{p^\top q} - \frac{Dqq^\top D}{q^\top Dq}$$

- Parameter $\phi = 1$ (Broyden-Fletcher-Goldfarb-Shanno)

$$C(1) = \frac{pp^\top}{p^\top q}\left[1 + \frac{q^\top Dq}{p^\top q}\right] - \frac{Dqp^\top + pq^\top D}{p^\top q}$$

- Quasi-Newton-Verfahren konvergieren oft superlinear
- Für quadratische Funktionen in n Schritten
  - entspricht dann dem konjugierten Gradientenverfahren