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

Optimierung SoSe 2020

Dr. Alexey Agaltsov

Plan

- Quasi-Newton-Verfahren
- Beste Approximation in $\mathbb{S}^n$
- BFGS-Verfahren

- Quasi-Newton-Verfahren
- Beste Approximation in $\mathbb{S}^n$
- BFGS-Verfahren

Minimierungsproblem

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

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

- Das Gradientenverfahren benötigt nur die ersten Ableitungen
- Das Newton-Verfahren benötigt die zweiten Ableitungen

Weniger Iterationsschritte

Iterationsschritte sind aufwendiger

- Mit den Quasi-Newton-Verfahren kann man einen Kompromiss erreichen

Erinnerung: Sekantenverfahren

Löse $g(x) = 0, x \in \mathbb{R}$

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

linearisiere

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

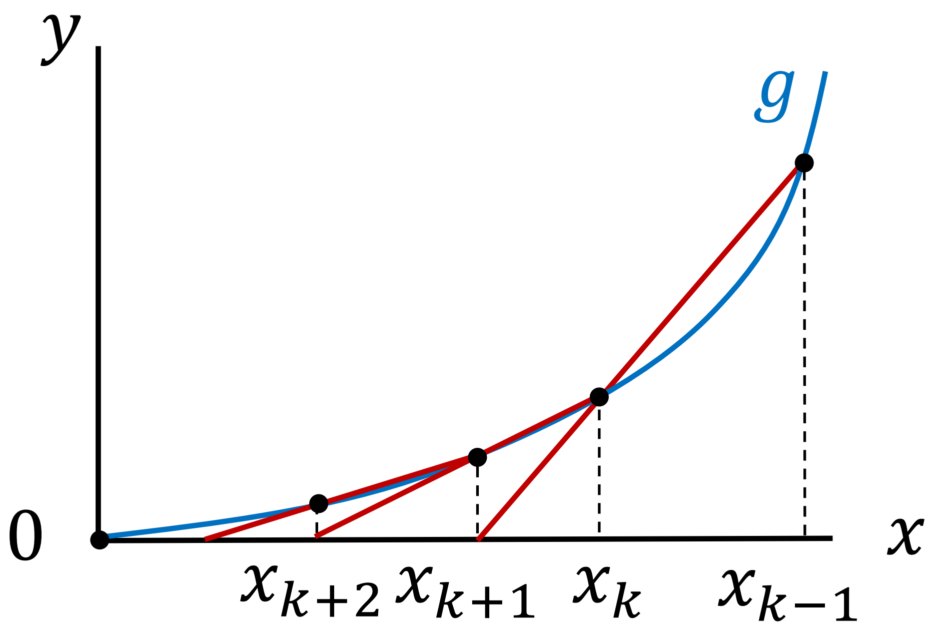
aktuelle Approximation

- Approximiere $g'(x_k)$ durch endliche Differenzen:

$$g(x_k) + \frac{g(x_k) - g(x_{k-1})}{x_k - x_{k-1}}(x - x_k) = 0$$

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

Sekantenverfahren



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

Quasi-Newton-Verfahren

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

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

Optimalitätsbedingung

$$f'(x) = 0$$

Sekantenverfahren

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

Alternative Herleitung

(Gedämpftes) Newton-Verfahren:

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

Quasi-Newton Verfahren:

$$x_{k+1} = x_k - \alpha_k B_k \nabla f(x_k)$$

wobei B_k ist eine *schnell auszurechnende* Approximation an $\nabla f(x_k)^{-1}$

Quasi-Newton-Verfahren

1. *Initialisierung:* Startwerte x_0, B_0, Toleranzwert ϵ
2. **for** $k = 0,1,2,\dots$ **do**:
3.   **if** $\|\nabla f(x_k)\|_2 < \epsilon$ **then** break
4.   bestimme eine Schrittweite α_k
5. $x_{k+1} = x_k - \alpha_k B_k \nabla f(x_k)$
6.   bestimme B_{k+1}
7. **end for**

Wie bestimmt man $B_{k+1} \approx \nabla f(x_{k+1})^{-1}$?

- Iterierten x_k, x_{k+1} und Gradienten $\nabla f(x_k), \nabla f(x_{k+1})$ lassen $\nabla^2 f(x_{k+1})$ abschätzen:

$$\underbrace{\nabla f(x_{k+1}) - \nabla f(x_k)}_{y_k} \approx \nabla^2 f(x_{k+1})\underbrace{(x_{k+1} - x_k)}_{s_k}$$

$$\nabla^2 f(x_{k+1})^{-1} y_k \approx s_k$$

- Setze:

$$B_{k+1} \in \mathbb{S}^n, \quad B_{k+1} y_k = s_k \qquad \textit{Sekantengleichung}$$

- Es bleibt noch $\frac{n(n+1)}{2} - n$ Freiheitsgrade festzulegen

Beispiel: Eindimensionaler Fall

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

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

Mehrdimensionaler Fall

$$B_{k+1} \in \mathbb{S}^n, \quad B_{k+1} y_k = s_k$$

Wie wird man die Freiheitsgrade los?

- Davidson-Fletcher-Powell (1959) – DFP

$$\left\| B_{k+1}^{-1} - B_k^{-1} \right\| \to min$$

- Broyden-Fletcher-Goldfarb-Shannon (1970er) – BFGS

$$\| B_{k+1} - B_k \| \to min$$

Eines der effizientesten Quasi-Newton-Verfahren

Plan

- Quasi-Newton-Verfahren
- Beste Approximation in $\mathbb{S}^n$
- BFGS-Verfahren

Notationen

- Gewichtungsmatrix $M \in \mathbb{R}^{n\times n}$ mit $\det M \neq 0$
- Frobenius-Skalarprodukt von $A, B \in \mathbb{S}^n$ bezüglich M:

$$\langle A, B\rangle_M := \mathrm{Spur}(MAM^T MBM^T)$$

- Frobenius-Norm von $A \in \mathbb{S}^n$ bezüglich M:

$$||A||_M^2 := \langle A, A\rangle_M$$

Lemma 4.2: Beste Approximation in $\mathbb{S}^n$

$$\text{Minimiere} \quad \|E\|_M \qquad \text{über } E \in \mathbb{S}^n$$

$$\text{u.d.N.} \quad Ey = s - r$$

$$y, s, r \in \mathbb{R}^n, \ \ s^T y \neq 0$$

$$M \in \mathbb{R}^{n \times n} \text{ ist regulär, } M^T M s = y,$$

Dann gibt es eine eindeutige optimale Lösung E_*:

$$E_* = \left(1 + \frac{r^T y}{s^T y}\right) \frac{s^T s}{s^T y} - \frac{s r^T + r s^T}{s^T y}$$

Charakterisierung von E_*

$$\text{Minimiere} \quad \|E\|_M \qquad \text{über } E \in \mathbb{S}^n$$
$$\text{u.d.N.} \quad Ey = s - r$$

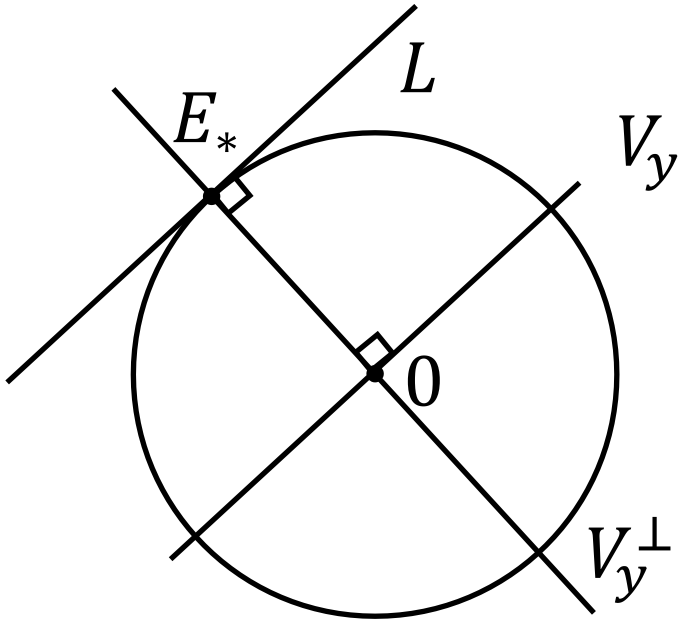
- E_* ist die Projektion von $0 \in \mathbb{S}^n$ auf:

$$L := \{E \in \mathbb{S}^n : Ey = s - r\}$$

Beispiel 2.10

Eindeutige Lösung

$$\{E_*\} = L \cap V_y^\perp$$
$$V_y = \{E \in \mathbb{S}^n : Ey = 0\}$$



Äquivalente Formulierung

$$\{E_*\} = L \cap V_y^\perp \qquad V_y^\perp = \{S \in \mathbb{S}^n \colon \langle S, E\rangle_M = 0 \ \forall E \in V_y\}$$

$$M^T M s = y$$

1. Wir zeigen zuerst, dass:

$$V_y^\perp = R_s$$

$$R_s = \{vs^T + sv^T \colon \ v \in \mathbb{R}^n\}$$

2. Danach bestimmen wir $v \in \mathbb{R}^n$ so dass:

$$E_* = vs^T + sv^T \in L = \{E \in \mathbb{S}^n \colon Ey = s - r\}$$

$$\Leftrightarrow \quad E_* y = s - r$$

Beweis

Behauptung: $R_s \subseteq V_y^\perp$

Sei $A = vs^T + sv^T \in R_s$ und $E \in V_y$:

$$\mathrm{Spur}(X_1 \dots X_n) = \mathrm{Spur}(X_2 \dots X_n X_1)$$

$$\begin{aligned}\langle E, A\rangle_M &= \mathrm{Spur}(MEM^TM(vs^T + sv^T)M^T)\\ &= 2\,\mathrm{Spur}(ME\underbrace{M^TMs}_{y}v^TM^T)\\ &= 2\,\mathrm{Spur}(M\underbrace{Ey}_{0}v^TM^T) = 0\end{aligned}$$

$$\Longrightarrow A \in V_y^\perp$$

Beweis

Behauptung: $R_s = V_y^{\perp}$

Wir haben bewiesen $R_s \subseteq V_y^{\perp}$

$s \neq 0$

$$\dim R_s = \dim\{vs^T + sv^T : v \in \mathbb{R}^n\} = n$$

$y \neq 0$

$$\dim V_y = \dim\{E \in \mathbb{S}^n : Ey = 0\} = \dim \mathbb{S}^n - n$$

$$\dim V_y^{\perp} = \dim \mathbb{S}^n - \dim V_y = n$$

$$\dim R_s = \dim V_y^{\perp}$$

$$R_s = V_y^{\perp}$$

Beweis: explizite Formel für E_*

Finde $v \in \mathbb{R}^n$: $\overbrace{(vs^T + sv^T)}^{E_*}y = s - r$

$$v = \alpha s - \beta r$$

$$(2\alpha s^T y)s - (\beta s^T y)r - (\beta r^T y)s = s - r$$

$$\alpha = \frac{1}{2s^T y}\left(1 + \frac{r^T y}{s^T y}\right), \beta = \frac{1}{s^T y}$$

$$v = \alpha s - \beta r$$
$$E_* = vs^T + sv^T$$

$$\Longrightarrow E_* = \left(1 + \frac{r^T y}{s^T y}\right)\frac{s^T s}{s^T y} - \frac{sr^T + rs^T}{s^T y}$$

$\square$

Plan

- Quasi-Newton-Verfahren
- Beste Approximation in $\mathbb{S}^n$
- BFGS-Verfahren

BFGS-Verfahren

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

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

- Angenommen, x_* ist eine optimale Lösung
- Seien x_k die Letzte Approximation an x_* und B_k eine Approximation an $\nabla^2 f(x_k)^{-1}$ und α_k eine Schrittweite

$$x_{k+1} := x_k - \alpha_k B_k \nabla f(x_k)$$

$$B_{k+1} = \operatorname{argmin}\{ \|B - B_k\|_{M_k} : B \in \mathbb{S}^n, By_k = s_k\} \quad M_k^T M_k y_k = s_k$$

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

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

Satz 4.3. BFGS-Update-Formel

$$\text{Minimiere} \quad \|B - B_k\|_{M_k} \quad \text{über } B \in \mathbb{S}^n$$

$$\text{u.d.N.} \quad By_k = s_k$$

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

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

Angenommen:

- Sei $M_k \in \mathbb{R}^{n\times n}$ regulär und so, dass $M_k^T M_k y_k = s_k$
- $y_k^T s_k > 0$
- $B_k \in \mathbb{S}^n_{\succ}$

Dann gibt es eine eindeutige Lösung B_{k+1}, sodass $B_{k+1} \in \mathbb{S}^n_{\succ}$

$$B_{k+1} = \left(1 - \rho_k s_k y_k^T\right) B_k \left(1 - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T, \qquad \rho_k = \frac{1}{y_k^T s_k}$$

Beweis: Substitution

Minimiere $\|B - B_k\|_{M_k}$ über $B \in \mathbb{S}^n$

u.d.N. $By_k = s_k$

$B := B_k + E$

Minimiere $\|E\|_{M_k}$ über $E \in \mathbb{S}^n$

u.d.N. $Ey_k = s_k - B_k y_k$

Lemma 4.2

$$E_* = \left(1 + \frac{y_k^T B_k y}{s_k^T y_k}\right)\frac{s_k^T s_k}{s_k^T y_k} - \frac{s_k (B_k y_k)^T + (B_k y_k) s_k^T}{s_k^T y_k}$$

Beweis: Rücksubstitution

$$B_{k+1} = B_k + E_*$$

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

$$\rho_k = \frac{1}{y_k^T s_k}$$

$$B_{k+1} = B_k + \rho_k s_k s_k^T + \rho_k^2 s_k y_k^T B_k y_k s_k^T - \rho_k s_k y_k^T B_k - \rho_k B_k y_k s_k^T$$

$$B_{k+1} = \left(1 - \rho_k s_k y_k^T\right) B_k \left(1 - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T$$

Beweis: $B_k \succ 0 \Rightarrow B_{k+1} \succ 0$

$$B_{k+1} = \left(1 - \rho_k s_k y_k^T\right) B_k \left(1 - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T$$

Sei $d \in \mathbb{R}^n \setminus \{0\}$

Angenommen, $s_k^T d = 0$

$$d^T B_{k+1} d = d^T B_k d > 0$$

Angenommen, $s_k^T d \neq 0$

$$d^T B_{k+1} d = \underbrace{u^T B_k u}_{\geq 0} + \underbrace{\rho_k \left(s_k^T d\right)^2}_{> 0} > 0$$

$$u = d - \rho_k y_k s_k^T d$$

$$B_{k+1} \succ 0$$

$\square$

Aufgabe 4.4

Sei $\nabla^2 f \succ 0$ und setze

$$M_k = \left(\int_0^1 \nabla^2 f(x_k + t s_k)\, dt \right)^{-1/2}$$

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

Dann ist $M_k \in \mathbb{R}^{n\times n}$ regulär und so, dass $M_k^T M_k y_k = s_k$

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

Lemma 4.5. Schrittweitenwahl

$$y_k^T s_k > 0$$
$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$
$$s_k = x_{k+1} - x_k$$

Das gilt, falls $\nabla f(x_k) \neq 0$, $B_k \succ 0$ und zumindest eine der folgenden Annahmen erfüllt ist:

- α_k wird mit der Minimierungsregel bestimmt
- α_k erfüllt die Krümmungsbedingung

Beweis

Behauptung. Werde α_k mit der Minimierungsregel gewählt, so gilt

$$\left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)^T \underbrace{(x_{k+1} - x_k)}_{\alpha_k d_k} > 0 \qquad d_k = -B_k \nabla f(x_k)$$

α_k minimiert $\varphi(\alpha) = f(x_k + \alpha d_k)$ über $\alpha > 0$

$$\frac{d}{d\alpha}\varphi(\alpha)\Big|_{\alpha=\alpha_k} = \nabla f(x_{k+1})^T d_k = 0$$

Da d_k eine Abstiegsrichtung in x_k weil $B_k \succ 0$:

$$-\nabla f(x_k)^T d_k = \nabla f(x_k)^T B_k \nabla f(x_k) > 0$$

Beweis

Behauptung. Erfülle α_k die Krümmungsbedingung, so gilt:

$$\left(\nabla f(x_{k+1}) - \nabla f(x_k)\right)^T (x_{k+1} - x_k) > 0$$

$$\nabla f(x_k + \alpha_k d_k)^T d_k \geq \mu \overbrace{\nabla f(x_k)^T d_k}^{< 0} \text{ mit } \mu \in (0,1)$$

$$> \nabla f(x_k)^T d_k$$

$$x_{k+1} - x_k = \alpha_k d_k$$

$\square$

Wahl von B_0

- $B_0 = I$

- Rechne $\nabla^2 f(x_0)$ aus und setze $B_0 = \nabla^2 f(x_0)^{-1}$

  Benötigt Implementierung von $\nabla^2 f$

- Setze $B_0 \approx \nabla^2 f(x_0)^{-1}$ durch finite Differenzen

Bemerkungen

- Typischerweise konvergiert das BFGS-Verfahren lokal superlinear [We92]

  starke Konvexität von f

  Wolfe-Bedingungen

- B_k kann viel Speicherplatz benötigen

  Es gibt die Modifikationen des Quasi-Newton-Verfahrens mit kleinerem Speicherplatzbedarf [NW99]

Zusammenfassung

- Quasi-Newton-Verfahren
- Beste Approximation in $\mathbb{S}^n$
- BFGS-Verfahren

Nächstes Video

- 4c. Newton-artige Verfahren: Ausgleichsprobleme