

MTMM.00.202 Optimiseerimismeetodid

Evely Kirsiaed

`evely.kirsiaed@ut.ee`

sügis 2020

4 Gradientmeetodid

- Gradientmeetodite idee
- Gradientmeetodite koonduvus

Gradientmeetodid

Olgu $\Omega \subset \mathbb{R}^n$ lahtine hulk, $f : \Omega \rightarrow \mathbb{R}$ diferentseeruv. Valime $h \in \mathbb{R}^n$, $h \neq 0$, ja moodustame funktsiooni

$$\varphi(t) = f(x + th), \quad t \in (-\delta, \delta).$$

Hulga Ω lahtisuse tõttu leidub selline $\delta > 0$, et φ on määratud. Kehtib $\varphi(0) = f(x)$. Leiame

$$\begin{aligned} \varphi'(0) &= \lim_{\substack{t \rightarrow 0 \\ t \neq 0}} \frac{\varphi(0+t) - \varphi(0)}{t} = \lim_{\substack{t \rightarrow 0 \\ t \neq 0}} \frac{f(x+th) - f(x)}{t} \\ &= \lim_{\substack{t \rightarrow 0 \\ t \neq 0}} \frac{(f'(x), th) + \alpha(x; th)}{t} = (f'(x), h) + \|h\| \underbrace{\lim_{\substack{t \rightarrow 0 \\ t \neq 0}} \frac{\alpha(x; th)}{t\|h\|}}_{=0}. \end{aligned}$$

Seega $(f'(x), h)$ näitab funktsiooni f kasvamise kiirust, kui liikuda punktist x suunas h . (Erinevate suundade võrdlemisel tuleb valida $\|h\|_2 = \text{const.}$)

Gradientmeetodid

Vaatame kõikvõimalikke suundasid $h \in \mathbb{R}^n$, mille korral $\|h\|_2 = \text{const.}$

Cauchy-Bunjakowski-Schwarzi võrratuse põhjal

$$-\|f'(x)\|_2\|h\|_2 \leq (f'(x), h) \leq \|f'(x)\|_2\|h\|_2$$

ning vähemalt üks võrdus realiseerub vaid juhul kui $f'(x)$ ja h on lineaarselt sõltuvad (näiteks $f'(x) = \lambda h$, $\lambda \in \mathbb{R}$).

Kui $\lambda > 0$ (st $f'(x)$ ja h on samasuunalised), siis realiseerub parempoolne võrdus: $(f'(x), h)$ on maksimaalne kui liikuda gradiendi suunas. Kui $\lambda < 0$ (st $f'(x)$ ja h on vastassuunalised), siis realiseerub vasakpoolne võrdus: funktsiooni f kasvamise kiirus on minimaalne (kahanemise kiirus on maksimaalne) kui liikuda gradiendi vastassuunas ehk suunas $-f'(x)$.

Gradientmeetodid

Olgu $f : \mathbb{R}^n \rightarrow \mathbb{R}$ diferentseeruv. Vaatleme ülesannet

$$\min_{x \in \mathbb{R}^n} f(x).$$

Definitsioon

Iteratsioonimeetodeid

$$x_{k+1} = x_k - \alpha_k f'(x_k), \quad \alpha_k > 0,$$

miinimumülesande lahendamisel nimetatakse gradientmeetoditeks.

Kui gradientmeetodis $\alpha_k > 0$ määratakse tingimusest

$$f(x_k - \alpha_k f'(x_k)) = \min_{\alpha > 0} f(x_k - \alpha f'(x_k))$$

(suunas $-f'(x_k)$ liigutakse seni, kuni jõutakse funktsiooni f vähima väärtuseni sellel suunal), siis räägitakse **kiireima languse meetodist**.

Kiireima languse meetod

Vaatleme kiireima languse meetodit

$$x_{k+1} = x_k - \alpha_k f'(x_k), \quad \alpha_k > 0,$$

kus $\alpha_k > 0$ määratakse tingimusest

$$f(x_k - \alpha_k f'(x_k)) = \min_{\alpha > 0} f(x_k - \alpha f'(x_k)).$$

- Kaks järjestikust liikumissuunda on ortogonaalsed, st $(x_{k+1} - x_k, x_{k+2} - x_{k+1}) = 0$ (Ül 10).
- Kui $f'(x_k) \neq 0$, siis $f(x_{k+1}) < f(x_k)$ (Ül 11).

Kiireima languse meetod

Näide 1

Vaatleme ruutfunktsionaali

$$f(x) = \frac{1}{2}(Ax, x) - (b, x),$$

kus A on sümmeetriline positiivselt määratud ruutmaatriks, minimiseerimist juhul $x \in \mathbb{R}^2$, $A = \begin{pmatrix} 2 & 0 \\ 0 & 10 \end{pmatrix}$, $b = \begin{pmatrix} 0 \\ 10 \end{pmatrix}$, st

$$f(x) = x_1^2 + 5x_2^2 - 10x_2.$$

Kiireima languse meetodi arvutuseeskiri on (vt Ül 12)

$$x^{k+1} = x^k - \frac{\|f'(x^k)\|^2}{(Af'(x^k), f'(x^k))} f'(x^k), \text{ kusjuures } f'(x^k) = Ax^k - b.$$

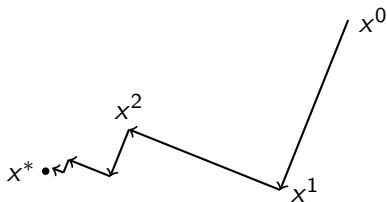
Kiireima languse meetod

Näide 1 (jätkub)

Valime alglähendi $x^0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$, siis $f'(x^0) = Ax^0 - b = \begin{pmatrix} 4 \\ 10 \end{pmatrix}$ ning

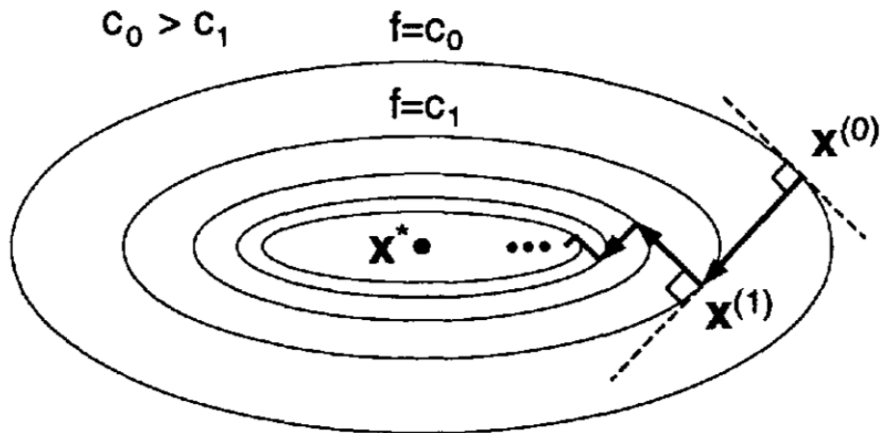
$$x^1 = x^0 - \frac{\|f'(x^0)\|^2}{(Af'(x^0), f'(x^0))} f'(x^0) = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{116}{1032} \begin{pmatrix} 4 \\ 10 \end{pmatrix} \approx \begin{pmatrix} 1.55 \\ 0.88 \end{pmatrix}$$

analoogiliselt $x^2 = x^1 - \frac{\|f'(x^1)\|^2}{(Af'(x^1), f'(x^1))} f'(x^1) \approx \begin{pmatrix} 0.55 \\ 1.28 \end{pmatrix}$ jne.



$$x^k \rightarrow x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Kiireima languse meetod



Kiireima languse meetod

Märkus. Ülesandes 12 on eeldus, et ruutfunktsionaalis

$$f(x) = \frac{1}{2}(Ax, x) - (b, x)$$

peab maatriks A olema sümmeetriline. Mittesümmeetrilise A korral

$$(Ax, x) = \left(\frac{1}{2}(A + A^T)x, x \right) = \frac{1}{2}(Ax, x) + \frac{1}{2}(x, A^T x) = \frac{1}{2}(\underbrace{(A + A^T)}_{=:B} x, x),$$

kus B on sümmeetriline, st alati saab viia kujule, kus A on sümmeetriline. Seega sümmeetrilisus ei ole kitsendav tingimus, küll aga on oluline, et A oleks positiivselt määratud.

Ligikaudne kiireima languse meetod

Kiireima languse meetodis esineval ülesandel

$$\min_{\alpha > 0} f(x_k - \alpha f'(x_k))$$

ei tarvitse lahendit olla või on seda täpselt raske leida, siis leitakse $\alpha_k > 0$ nii, et

$$f(x_k - \alpha_k f'(x_k)) \leq \inf_{\alpha > 0} f(x_k - \alpha f'(x_k)) + \delta_k, \quad \delta_k \geq 0, \quad \lim_{k \rightarrow \infty} \delta_k = 0.$$

Nimetame seda ligikaudseks kiireima languse meetodiks. Kui

$\inf_{\alpha > 0} f(x_k - \alpha f'(x_k)) > -\infty$, siis on selline valik teostatav.

Kiireima languse meetod on erijuht ligikaudsest kiireima languse meetodist ($\delta_k = 0$).

Gradientmeetodite koonduvus

Lemma (jääkliikme hinnang)

Olgu $f : \mathbb{R}^n \rightarrow \mathbb{R}$ selline, et mingi L korral

$$\|f'(x) - f'(y)\|_2 \leq L\|x - y\|_2 \quad \forall x, y \in \mathbb{R}^n$$

(tuletis rahuldab Lipschitzi tingimust). Siis Taylori arendises

$$f(x + h) = f(x) + (f'(x), h) + \alpha(x; h)$$

on jääkliige hinnatav $|\alpha(x; h)| \leq \frac{L}{2}\|h\|_2^2$.

Teoreem 1

Kui f on alt tõkestatud, f' rahuldab Lipschitzi tingimust ja ligikaudses kiireima languse meetodis $\sum_{k=1}^{\infty} \delta_k < \infty$, siis $\lim_{k \rightarrow \infty} f'(x_k) = 0$.

Gradientmeetodite koonduvus

Funktsiooni $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nimetatakse **kumeraks**, kui iga $x, y \in \mathbb{R}^n$, $x \neq y$, ja iga $\lambda \in (0, 1)$ korral

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Funktsiooni $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nimetatakse **rangelt kumeraks**, kui iga $x, y \in \mathbb{R}^n$, $x \neq y$, ja iga $\lambda \in (0, 1)$ korral

$$f(\lambda x + (1 - \lambda)y) < \lambda f(x) + (1 - \lambda)f(y).$$

Funktsiooni $f : \mathbb{R}^n \rightarrow \mathbb{R}$ nimetatakse **tugevalt kumeraks**, kui eksisteerib $\gamma > 0$ nii, et

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y) - \lambda(1 - \lambda)\gamma\|x - y\|^2$$

iga $x, y \in \mathbb{R}^n$, $x \neq y$, ja iga $\lambda \in (0, 1)$ korral.

Gradientmeetodite koonduvus

Teoreem 2

Kui f on tugevalt kumer ja f' rahuldab Lipschitzi tingimust, siis ülesandel $\min_{x \in \mathbb{R}^n} f(x)$ on olemas ühene lahend x^ ja kiireima languse meetodi korral kehtivad võrratused*

$$\begin{aligned}f(x_k) - f(x^*) &\leq (f(x_0) - f(x^*))q^k, \\ \|x_k - x^*\|^2 &\leq \frac{1}{\gamma}(f(x_0) - f(x^*))q^k,\end{aligned}$$

kus $q = 1 - \frac{2\gamma}{L}$ (γ ja L on arvud vastavalt f tugeva kumeruse võrratusest ja f' Lipschitzi tingimusest).

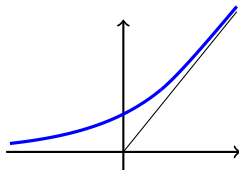
Gradientmeetodite koonduvus

Märkus 1. Ülesande $\min_{x \in \mathbb{R}^n} f(x)$ ühese lahendi olemasolu tagab f tugev kumerus, diferentseeruvust ja Lipschitzi tingimuse rahuldatust tuletise poolt selleks ei vajata.

Märkus 2. Nii $f(x_k) - f(x^*)$ kui $\|x_k - x^*\|$ kahanevad geomeetrilise progressiooni kiirusega teguriga vastavalt q ja $q^{1/2} > q$, st $\|x_k - x^*\|$ kahanemine on aeglasem.

Märkus 3. Teoreemis ei saa tugeva kumeruse eeldust asendada range kumerusega. Nt

$$f(x) = \begin{cases} e^{\frac{x}{2}}, & x \leq 1, \\ \frac{e^{1/2}}{4} \left(3x + \frac{1}{x}\right), & x \geq 1, \end{cases}$$



on rangelt kumer ja f' rahuldab Lipschitzi tingimust, aga miinimumpunkti ei ole.

Gradientmeetodite koonduvus

Märkus 4. Ülesande $\min_{x \in \mathbb{R}^n} f(x)$ lahendamisel on kasutatav Newtoni meetod

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

mis tegelikult lahendab võrrandisüsteemi $f'(x) = 0$. See meetod on erijuht gradientmeetodi üldistusest

$$x_{k+1} = x_k - A_k f'(x_k), \quad A_k \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n),$$

kusjuures gradientmeetodis $A_k = \alpha_k I$. Newtoni meetod on ruutkoonduvusega, kui $(f')'$ rahuldab Lipschitzi tingimust, samas on probleemiks alglähendi valik.

Gradientmeetodite koonduvus

Märkus 5. Maatriks $f''(x)$ on sümmeetriline, seega on tema omaväärtused reaalsed, $f''(x^*)$ omaväärtused on mittenegatiivsed, kuid välistatud ei ole omaväärtus $\lambda = 0$ (siis $f''(x^*)$ ei ole pööratav). Kui $f''(x_k)$ ei ole pööratav (esineb omaväärtus $\lambda_k < 0$), aga x_k on x^* lähedal, siis $|\lambda_k|$ on väike ning saab kasutada meetodit

$$x_{k+1} = x_k - (f''(x_k) + \mu_k I)^{-1} f'(x_k),$$

kus $\mu_k > 0$ on väike, aga selline, et $\mu_k + \lambda_k > 0$ maatriksi $f''(x_k)$ iga omaväärtuse λ_k korral. Siis $f''(x_k) + \mu_k I$ on pööratav.