

# MTMM.00.202 Optimiseerimismeetodid

Evely Kirsiaed

`evely.kirsiaed@ut.ee`

sügis 2020

## 5 Kaasgradientide meetod

- Ruutfunktsionaali minimiseerimine kaasgradientide meetodiga
- Üldisema funktsiooni minimiseerimine kaasgradientide meetodiga

# Ruutfunktsionaali minimiseerimine

Olgu antud ruutfunktsionaal

$$f(x) = \frac{1}{2}(Ax, x) - (b, x), \quad x \in \mathbb{R}^n,$$

kus  $A$  on sümmeetriline positiivselt määratud  $n \times n$  maatriks ja  $b \in \mathbb{R}^n$ .

Kui  $A$  ei ole sümmeetriline, võib ta asendada maatriksiga  $\frac{1}{2}(A + A^T)$ , mis on sümmeetriline (ruutfunktsionaal sellest ei muutu).

$$\begin{aligned} A \text{ on positiivselt määratud} &\Leftrightarrow (Ax, x) > 0 \quad \forall x \in \mathbb{R}^n \setminus \{0\} \\ &\Leftrightarrow \exists \gamma > 0 : (Ax, x) \geq \gamma \|x\|^2 \quad \forall x \in \mathbb{R} \\ &\Rightarrow \exists A^{-1} \text{ (sest } Ax = 0 \Rightarrow x = 0) \end{aligned}$$

Seega  $\exists! x^* \in \mathbb{R}^n : Ax^* = b$  (st  $x^* = A^{-1}b$ ) ning iga  $x \neq x^*$  korral

$$f(x) = \dots = \frac{1}{2}(A(x - x^*), x - x^*) + f(x^*) > f(x^*)$$

Ruutfunktsionaal saavutab miinimumi parajasti ühes punktis  $x^* = A^{-1}b$ .

# Ruutfunktsionaali minimiseerimine

Ruutfunktsionaali  $f$  tuletis avaldub kujul  $f'(x) = Ax - b$ , sest

$$\begin{aligned} f(x+h) - f(x) &= \frac{1}{2}(A(x+h), x+h) - (b, x+h) - \frac{1}{2}(Ax, x) + (b, x) = \\ &= (Ax - b, h) + \frac{1}{2}(Ah, h) \end{aligned}$$

ja  $\alpha(x; h) = \frac{1}{2}(Ah, h) \leq \frac{1}{2}\|A\|\|h\|^2 = o(\|h\|)$  (siin  $(Ah, h)$  ja  $\alpha(x; h)$  on mittenegatiivsed).

Märgime veel, et  $(Ax, y)$ ,  $x, y \in \mathbb{R}^n$ , on skalaarkorrutis ruumis  $\mathbb{R}^n$ , selle tagab  $A$  sümmeetrilisus ja positiivne määratus. Kasutatakse tähistust  $(x, y)_A = (Ax, y)$  ja räägitakse  $A$ -skalaarkorrutisest.

- $(x, x)_A \geq 0$ ;  $(x, x)_A = 0 \Leftrightarrow x = 0$ ;
- $(x, y)_A = (y, x)_A$ ;
- $(x + z, y)_A = (x, y)_A + (z, y)_A$ ;  $(\alpha x, y)_A = \alpha(x, y)_A$ .

# Kaasgradientide meetod ruutfunktsionaali korral

Olgu  $x_0 \in \mathbb{R}^n$ ,  $g_0 = Ax_0 - b$  (tuletis, gradient punktis  $x_0$ ),  $z_0 = -g_0$  (gradiendi vastassuund). Kui  $g_0 = 0$ , siis  $x_0 = x^*$  ja miinimumülesanne on lahendatud.

Olgu  $x_k, g_k, z_k$  leitud,  $x_k$  on punkt, kus asutakse,  $g_k = Ax_k - b$  on gradient, samuti jääkliige võrrandi  $Ax = b$  lahendamisel,  $z_k$  on liikumissuund punktist  $x_k$  väljumisel. Eeldame, et  $z_k \neq 0$ . Leiame funktsiooni  $\varphi(\alpha) = f(x_k + \alpha z_k)$  miinimumkoha. Arvutame

$$\begin{aligned}\varphi(\alpha) &= f(x_k + \alpha z_k) = \frac{1}{2}(A(x_k + \alpha z_k), x_k + \alpha z_k) - (b, x_k + \alpha z_k) = \\ &= \frac{1}{2}(Ax_k, x_k) + \alpha(Ax_k, z_k) + \frac{1}{2}\alpha^2(Az_k, z_k) - (b, x_k) - \alpha(b, z_k),\end{aligned}$$

$$\begin{aligned}\varphi'(\alpha) &= (Ax_k - b, z_k) + \alpha(Az_k, z_k) = (g_k, z_k) + \alpha(Az_k, z_k) = 0 \Leftrightarrow \\ &\Leftrightarrow \alpha = \frac{-(g_k, z_k)}{(Az_k, z_k)}\end{aligned}$$

$$\varphi''(\alpha) = (Az_k, z_k) > 0$$

# Kaasgradientide meetod ruutfunktsionaali korral

Seega  $\alpha_k = \frac{-(g_k, z_k)}{(Az_k, z_k)}$  on miinimumkoht. Määrame

$$x_{k+1} = x_k + \alpha_k z_k = x_k - \frac{(g_k, z_k)}{(Az_k, z_k)} z_k,$$

$$g_{k+1} = f'(x_{k+1}) = Ax_{k+1} - b \quad (\text{gradient, jääkliige}),$$

$$z_{k+1} = -g_{k+1} + \gamma_k z_k,$$

kus  $\gamma_k = \frac{(g_{k+1}, Az_k)}{(Az_k, z_k)}$ ; siin  $z_{k+1}$  on punktist  $x_{k+1}$  väljumisel liikumissuund, mis on gradiendi vastassuuna  $-g_{k+1}$  korrigeeritud suund.

## Näide 1

*Vaatleme ruutfunktsionaali*

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

*minimiseerimist juhul  $x \in \mathbb{R}^2$ ,  $A = \begin{pmatrix} 2 & 0 \\ 0 & 10 \end{pmatrix}$ ,  $b = \begin{pmatrix} 0 \\ 10 \end{pmatrix}$ , st*

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

*Valime alglähendi  $x^0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$ , siis  $g_0 = Ax^0 - b = \begin{pmatrix} 4 \\ 10 \end{pmatrix}$  ning*

$$z_0 = -g_0 = \begin{pmatrix} -4 \\ -10 \end{pmatrix}.$$

## Näide 1 (jätkub)

Kiireima languse meetodis  $x_{k+1} = x_k - \frac{\|g_k\|^2}{(Ag_k, g_k)} g_k$  kaasgradientide meetodis  $x_{k+1} = x_k - \frac{(g_k, z_k)}{(Az_k, z_k)} z_k$ . Kuna  $z_0 = -g_0$ , siis esimene samm on mõlemas meetodis sama

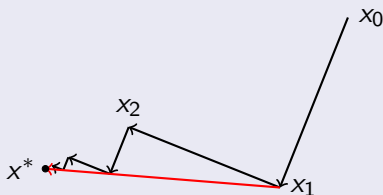
$$x_1 = x_0 - \frac{\|g_0\|^2}{(Ag_0, g_0)} g_0 = \begin{pmatrix} 2 \\ 2 \end{pmatrix} - \frac{116}{1032} \begin{pmatrix} 4 \\ 10 \end{pmatrix} = \begin{pmatrix} \frac{200}{129} \\ \frac{113}{129} \end{pmatrix}.$$

Kaasgradientide meetodis tuleb suunda  $-g_1$  korrigeerida valides

$$z_1 = -g_1 + \frac{(g_1, Az_0)}{(Az_0, z_0)} z_0 = - \begin{pmatrix} \frac{400}{129} \\ \frac{-160}{129} \end{pmatrix} + \frac{12800/129}{1032} \begin{pmatrix} -4 \\ -10 \end{pmatrix} = \begin{pmatrix} -\frac{58000}{16641} \\ \frac{4640}{16641} \end{pmatrix}$$

## Näide 1 (jätkub)

Kiireima languse meetodis  $x_2 = x_1 - \frac{\|g_1\|^2}{(Ag_1, g_1)} g_1 \approx \begin{pmatrix} 0.55 \\ 1.28 \end{pmatrix}$ , kaasgradientide meetodis  $x_2 = x_1 - \frac{(g_1, z_1)}{(Az_1, z_1)} z_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$



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

$$x_2 = x^*$$

# Kaasgradientide meetod ruutfunktsionaali korral

## Teoreem 1

*Kui ruutfunktsionaali korral lähendid  $x_k$  on leitud kaasgradientide meetodil, siis leidub  $k \leq n$  nii, et  $x_k = x^*$ .*

Tõestus.

$$\begin{aligned}(Az_{k+1}, z_k) &= (z_{k+1}, Az_k) = \\ &= (-g_{k+1}, Az_k) + \gamma_k(z_k, Az_k) = \\ &= -(g_{k+1}, Az_k) + \frac{(g_{k+1}, Az_k)}{(Az_k, z_k)}(z_k, Az_k) = 0;\end{aligned}$$

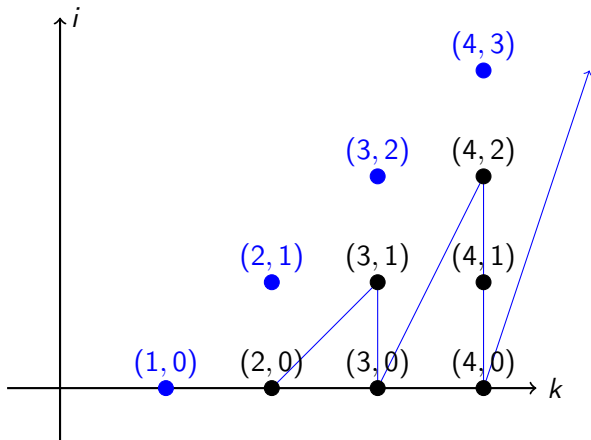
$$\begin{aligned}(g_{k+1}, z_k) &= (Ax_{k+1} - b, z_k) = \\ &= (A(x_k + \alpha_k z_k) - b, z_k) = \\ &= (Ax_k - b, z_k) + \alpha_k(Az_k, z_k) = \\ &= (g_k, z_k) + \frac{-(g_k, z_k)}{(Az_k, z_k)}(Az_k, z_k) = 0;\end{aligned}$$

$$\begin{aligned}\alpha_k &= \frac{-(g_k, z_k)}{(Az_k, z_k)} = \frac{-(g_k, -g_k + \gamma_{k-1}z_{k-1})}{(Az_k, z_k)} = \\ &= \frac{\|g_k\|^2 - \gamma_{k-1}(g_k, z_{k-1})}{(Az_k, z_k)} = \frac{\|g_k\|^2}{(Az_k, z_k)};\end{aligned}$$

$$\begin{aligned}(g_{k+1}, g_k) &= (Ax_{k+1} - b, g_k) = \\ &= (A(x_k + \alpha_k z_k) - b, g_k) = \\ &= (Ax_k - b + \alpha_k Az_k, g_k) = \\ &= (g_k + \alpha_k Az_k, g_k) = \\ &= \|g_k\|^2 + \alpha_k (Az_k, g_k) = \\ &= \|g_k\|^2 + \alpha_k (Az_k, -z_k + \gamma_{k-1}z_{k-1}) = \\ &= \|g_k\|^2 - \alpha_k (Az_k, z_k) + \alpha_k \gamma_{k-1} (Az_k, z_{k-1}) = 0;\end{aligned}$$

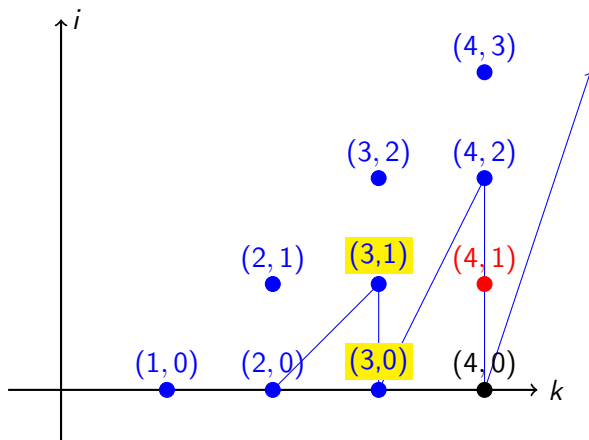
# Kaasgradientide meetod ruutfunktsionaali korral

Näitame, et  $(g_k, g_i) = 0$ ,  $k \neq i$ , ja  $(Az_k, z_i) = 0$ ,  $k \neq i$ . Võime eeldada, et  $i < k$ . Kasutame matemaatilist induktsiooni.



$$\begin{aligned}(g_{k+1}, g_i) &= (Ax_{k+1} - b, g_i) = (A(x_k + \alpha_k z_k) - b, g_i) = \\&= (Ax_k - b + \alpha_k Az_k, g_i) = (g_k + \alpha_k Az_k, g_i) = \\&= \underbrace{(g_k, g_i)}_{=0} + \alpha_k (Az_k, g_i) = \alpha_k (Az_k, -z_i + \underbrace{\gamma_{i-1} z_{i-1}}_{i=0 \text{ korral } 0}) = \\&= -\alpha_k \underbrace{(Az_k, z_i)}_{=0} + \alpha_k \gamma_{i-1} \underbrace{(Az_k, z_{i-1})}_{=0} = 0\end{aligned}$$

# Kaasgradientide meetod ruutfunktsionaali korral

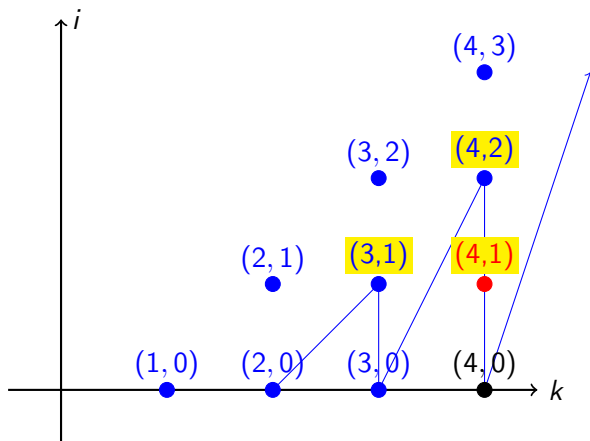


$$\begin{aligned}(Az_{k+1}, z_i) &= (z_{k+1}, Az_i) = (-g_{k+1} + \gamma_k z_k, Az_i) = \\&= (-g_{k+1}, Az_i) + \gamma_k \underbrace{(z_k, Az_i)}_{=0} = -(g_{k+1}, Az_i) \stackrel{(*)}{=} \\&= -(g_{k+1}, \frac{1}{\alpha_i}(g_{i+1} - g_i)) = \\&= -\frac{1}{\alpha_i} \underbrace{(g_{k+1}, g_{i+1})}_{=0} + \frac{1}{\alpha_i} \underbrace{(g_{k+1}, g_i)}_{=0} = 0\end{aligned}$$

$$\begin{aligned} (*) \quad g_{i+1} &= Ax_{i+1} - b = A(x_i + \alpha_i z_i) - b = Ax_i - b + \alpha_i Az_i = g_i + \alpha_i Az_i, \\ Az_i &= \frac{1}{\alpha_i}(g_{i+1} - g_i)\end{aligned}$$

Kui aga  $\alpha_i = 0$ , siis  $\alpha_i = \frac{\|g_i\|^2}{(Az_i, z_i)}$  ja  $z_i \neq 0$  alusel  $g_i = 0$ , st  $x_i = x^*$ . Kui  $z_i = 0$ , siis muidugi  $(Az_{k+1}, z_i) = 0$ .

# Kaasgradientide meetod ruutfunktsionaali korral



Vektorid  $g_k$  moodustavad paarikaupa ortogonaalse süsteemi. Nullist erinevad paarikaupa ortogonaalsed vektorid on lineaarselt sõltumatud (Ül 15). Samas ei ole võimalik, et  $g_k \neq 0$ ,  $k = 0, \dots, n$ , sest siis oleks ruumis  $\mathbb{R}^n$   $n + 1$  lineaarselt sõltumatut elementi. Seega on olemas  $g_k = 0$ ,  $k \leq n$ , mis tähendab, et  $Ax_k - b = 0$  ehk  $x_k = x^*$ . □

Vaatleme ülesannet

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

kus  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  on diferentseeruv.

Valime  $x_0 \in \mathbb{R}^n$ , leiame  $g_0 = f'(x_0)$ ,  $z_0 = -g_0$ . Olgu leitud  $x_k$ ,  $g_k = f'(x_k) \neq 0$  ja  $z_k \neq 0$ . Määrame  $\alpha_k > 0$  nii, et  $f(x_k + \alpha_k z_k) = \min_{\alpha > 0} f(x_k + \alpha z_k)$ . Seejärel leiame  $x_{k+1} = x_k + \alpha_k z_k$ ,  $g_{k+1} = f'(x_{k+1})$  ja  $z_{k+1} = -g_{k+1} + \gamma_k z_k$ , kus

$$\gamma_k = \frac{(g_{k+1}, g_{k+1} - g_k)}{\|g_k\|^2}.$$

Veendume, et ruutfunktsionaalide korral kirjeldatud meetodid ühtivad.

$$\begin{aligned}\gamma_k &= \frac{(g_{k+1}, g_{k+1} - g_k)}{\|g_k\|^2} = \frac{(g_{k+1}, Ax_{k+1} - b - (Ax_k - b))}{\|g_k\|^2} = \\ &= \frac{(g_{k+1}, A(x_k + \alpha_k z_k - x_k))}{\|g_k\|^2} = \frac{\alpha_k (g_{k+1}, Az_k)}{\|g_k\|^2} \\ &= \frac{(g_{k+1}, Az_k)}{(Az_k, z_k)},\end{aligned}$$

$$\text{sest } \alpha_k = \frac{\|g_k\|^2}{(Az_k, z_k)}.$$

Üldisema funktsiooni korral kasutatakse veel ka meetodit, kus

$$\gamma_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}.$$

Ka see ühtib ruutfunktsionaali korral algse meetodiga, sest  $(g_{k+1}, g_k) = 0$ .

# Näide

Funktsiooni  $f(x_1, x_2) = 1 - e^{-(10x_1^2 + x_2^2)}$  globaalne miinimum on punktis  $(0, 0)$ . Valime algähendiks  $x_0 = (-0.3, 0.8)$  ning leiame  $x_1, \dots, x_5$  kasutades kiireima languse meetodit ( $\gamma_k = 0$ ) ning kaasgradientide meetodit juhul  $\gamma_k = \frac{(g_{k+1}, g_{k+1} - g_k)}{\|g_k\|^2}$  ja  $\gamma_k = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$ .

