Gradient ja Hessiaan

Konstantin Tretjakov

1 Gradient

Definitsioon Olgu $f : \mathbb{R}^n \to \mathbb{R}^m$. Funktsiooni f nimetatakse diferentseeruvaks punktis $\mathbf{x}_0$, kui leidub lineaarteisendus $\mathbf{A}(\mathbf{x}_0)$ selline, et

$$f(\mathbf{x}_0 + \mathbf{\Delta x}) = f(\mathbf{x}_0) + \mathbf{A}(\mathbf{x}_0)\mathbf{\Delta x} + o(\mathbf{\Delta x})$$

Funktsiooni f nimetatakse diferentseeruvaks hulgal $Q \subset \mathbb{R}^n$, kui ta on diferentseeruv igas hulga Q punktis. Kui f on diferentseeruv hulgal $\mathbb{R}^n$, siis ütleme lihtsalt, et f on diferentseeruv. Maatriksit $\mathbf{A}(\mathbf{x}_0)$ nimetatakse funktsiooni f *tuletiseks või Jakobi maatriksiks* (punktis $\mathbf{x}_0$) ning tähistatakse $\frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{x}}$ või $f'(\mathbf{x}_0)$.

Ülesanne: Mida tähendab eeltoodud definitsioonis $o(\mathbf{\Delta x})$ ja miks pole see $o(\|\mathbf{\Delta x}\|)$?

Teoreem 1.1

$$(\mathbf{A}(\mathbf{x}_0))_{ij} = \frac{\partial f_i(\mathbf{x}_0)}{\partial x_j}$$

Ülesanne*: Tõesta seda.

Vihje: alguses vaatle juhtu $f : \mathbb{R}^n \to \mathbb{R}$ ning kasuta täisdiferentsiaali valemit $\mathrm{d}f(\mathbf{x}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i} \mathrm{d}x_i$

Teoreem 1.2

$$(f(g(\mathbf{x}))' = f'(g(\mathbf{x}))g'(\mathbf{x})$$

Ülesanne*: Tõesta seda.

Vihje: alguses vaatle juhtu $f : \mathbb{R}^n \to \mathbb{R}$ ning kasuta valemit $\frac{\partial f \circ g}{\partial x_j}(\mathbf{x}) = \sum_{i=1}^n \frac{\partial f}{\partial x_i}(g(\mathbf{x})) \cdot \frac{\partial g_i}{\partial x_j}(\mathbf{x})$.

Olgu $f : \mathbb{R}^n \to \mathbb{R}$ diferentseeruv funktsioon. Tema tuletis (punktis $\mathbf{x}_0$) on siis $1 \times n$ maatriks

$$\mathbf{A}(\mathbf{x}_0) = \begin{pmatrix} \frac{\partial f(\mathbf{x}_0)}{\partial x_1} & \frac{\partial f(\mathbf{x}_0)}{\partial x_2} & \dots & \frac{\partial f(\mathbf{x}_0)}{\partial x_n} \end{pmatrix}$$

Vektorit $\mathbf{A}(\mathbf{x}_0)^T$ nimetatakse funktsiooni f *gradiendiks* ning tähistatakse $\mathrm{grad}_{\mathbf{x}} f(\mathbf{x}_0)$ või $\nabla_{\mathbf{x}} f(\mathbf{x}_0)$ või lihtsalt $\nabla f(\mathbf{x}_0)$.

Teoreem 1.3

$$\begin{aligned} \nabla(f(\mathbf{x})g(\mathbf{x})) &= \nabla f(\mathbf{x})g(\mathbf{x}) + f(\mathbf{x})\nabla g(\mathbf{x}) \\ \nabla(f(\mathbf{x})/g(\mathbf{x})) &= (\nabla f(\mathbf{x})g(\mathbf{x}) - f(\mathbf{x})\nabla g(\mathbf{x}))/g^2(\mathbf{x}) \\ \nabla(f(g(\mathbf{x}))) &= \nabla f(g(\mathbf{x}))\nabla g(\mathbf{x}) = f'(g(\mathbf{x}))\nabla g(\mathbf{x}) \end{aligned}$$

Ülesanne: Veendu valemite õigsuses.

Vihje: kasuta teoreeme (1.1) ja (1.2).

Ülesanne*: Veendu, et gradiendi saab vaadelda kui vektorit, mis näitab funktsiooni pinnal suurima tõusu poole.

Näited

$$\frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}^T$$
$$\frac{\partial \mathbf{x}^T\mathbf{A}\mathbf{x}}{\partial \mathbf{x}} = \mathbf{x}^T(\mathbf{A}+\mathbf{A}^T)$$

Ülesanne: Veendu valemite õigsuses.

2 Hessiaan

Definitsioon Olgu $f : \mathbb{R}^n \to \mathbb{R}$ diferentseeruv ja $\nabla f : \mathbb{R}^n \to \mathbb{R}^n$ diferentseeruv (s.t. f on kaks korda diferentseeruv). Funktsiooni ∇f tuletis punktis $\mathbf{x}_0$ on $n \times n$ maatriks $\mathbf{H}(\mathbf{x}_0)$, mida nimetatakse funktsiooni f *teiseks tuletiseks* ehk *Hessiaaniks* (punktis $\mathbf{x}_0$) ja tähistatakse $\frac{\partial^2 f(\mathbf{x}_0)}{\partial^2 \mathbf{x}}$ või $\nabla^2 f(\mathbf{x}_0)$.

Teoreem 2.1

$$(\mathbf{H}(\mathbf{x}_0))_{ij} = \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_i \partial x_j}$$

Ülesanne: Tõesta seda.

Vihje: väide järeldub teoreemist (1.1).

Teoreem 2.2 *Kui funktsiooni f osatuletised $\frac{\partial^2 f}{\partial x_i \partial x_j}$, $\frac{\partial^2 f}{\partial x_j \partial x_i}$ on pidevad punktis $\mathbf{x}_0$, siis nad on võrdsed selles punktis.*

Seega piisavalt sileda funktsiooni Hessiaan on sümmeetriline maatriks.

Teoreem 2.3 *Sümmeetrilise maatriksi kõik omaväärtused on reaalarvud.*

Teoreem 2.4 *Sümmeetrilise maatriksi omavektoritest saab koostada ortonormeeritud baasi.*

Ülesanne*: Leia mõnest õpikust nende teoreemide tõestusi ja saa nendest aru (vt. näiteks M. Kilp. *Algebra I*).

Lõpuks üks tähtsamaid tulemusi: kui funktsioon on kaks korda diferentseeruv, me saame arendada teda Taylor'i ritta:

Teoreem 2.5

$$f(\mathbf{x}_0 + \mathbf{\Delta x}) = f(\mathbf{x}_0) + \frac{\partial f(\mathbf{x}_0)}{\partial \mathbf{x}}\mathbf{\Delta x} + \frac{1}{2}\mathbf{\Delta x}^T\left(\frac{\partial^2 f(\mathbf{x}_0)}{\partial^2 \mathbf{x}}\right)\mathbf{\Delta x} + o(\|\mathbf{\Delta x}\|^2)$$

Ülesanne: Tõesta seda.

Vihje: vaja on rakendada diferentseeruvuse definitsiooni kaks korda.

Viimane teoreem väidab, et kui meil on mingi piisavalt sile funktsioon, ning me vaatleme teda ainult mingi punkti $\mathbf{x}$ väikeses ümbruses, me saame asendada teda ruutfunktsiooniga. Seega on kasuks kujutada ette kuidas näeb välja funktsioon kujul $f(\mathbf{x}) = c + \mathbf{a}^T\mathbf{x} + \mathbf{x}^T\mathbf{H}\mathbf{x}$, kus $\mathbf{H}$ on sümmeetriline maatriks.

Ülesanne: Vaatleme alguses funktsiooni $f(\mathbf{x}) = \mathbf{x}^T\mathbf{H}\mathbf{x}$, kus $\mathbf{x} \in \mathbb{R}^2$ ning $\mathbf{H}$ on sümmeetriline maatriks. Olgu maatriksi $\mathbf{H}$ ortogonaalsed ühikupikkusega omavektorid $\mathbf{v}_1$ ja $\mathbf{v}_2$ ning vastavad omaväärtused λ_1 ja λ_2. Joonista tasandile vektorid $\mathbf{v}_1$ ja $\mathbf{v}_2$. Uuri kuidas funktsioon käitub nende vektorite poolt määratud sirgete peal, s.t. mis on funktsiooni väärtused punktides $t\mathbf{v}_1$, $t\mathbf{v}_2$ ($t \in \mathbb{R}$). Kas näed, et f on sisuliselt kas kumer paraboloid, või "sadulpunktiga" pind? Mis on arvude λ_i roll?

Ülesanne: Analüüsi kuidas peab välja nägema funktsioon $f(\mathbf{x}) = c + \mathbf{a}^T\mathbf{x} + \mathbf{x}^T\mathbf{H}\mathbf{x}$.

3 Lokaalse miinimumi mõiste

Tähistagu f funktsiooni $\mathbb{R}^n \to \mathbb{R}$.

Definitsioon Punkt $\mathbf{x}^*$ on funktsiooni f *lokaalne miinimum* (või lihtsalt miinimum), kui leidub selle punkti ümbrus $U(\mathbf{x}^*)$ niisugune, et $f(\mathbf{x}) \geq f(\mathbf{x}^*) \quad \forall \mathbf{x} \in U(\mathbf{x}^*)$.

Teoreem 3.1 (Fermat) *Olgu $\mathbf{x}^*$ funktsiooni f miinimumi punkt ning f on diferentseeruv selles punktis. Siis $\nabla f(\mathbf{x}^*) = \mathbf{0}$.*

Ülesanne: Tõesta seda.

Üldiselt vastupidine väide ei kehti, s. t. gradiendi võrdus nulliga ei ole miinimumi piisav tingimus. Kehtivad aga järgmised väited:

Teoreem 3.2 *Kui f on kumer funktsioon, s.t.*

$$\forall\ \mathbf{x}, \mathbf{y} \quad f(\lambda \mathbf{x} + (1-\lambda)\mathbf{y}) \leq \lambda f(\mathbf{x}) + (1-\lambda) f(\mathbf{y})$$

siis tingimus $\nabla f(\mathbf{x}^*) = \mathbf{0}$ on samaväärne sellega, et $\mathbf{x}^*$ on funktsiooni f globaalne miinimum.

Teoreem 3.3 *Olgu $\mathbf{x}^*$ funktsiooni f miinimumi punkt, ning f on kaks korda diferentseeruv punktis $\mathbf{x}^*$. Siis $\nabla^2 f(\mathbf{x}^*) \geq 0$ (s.t. funktsiooni Hessiaan on mittenegatiivselt määratud selles punktis).*

Teoreem 3.4 *Kui f on kaks korda diferentseeruv punktis $\mathbf{x}^*$, $\nabla f(\mathbf{x}^*) = \mathbf{0}$ ning $\nabla^2 f(\mathbf{x}^*) > 0$, siis $\mathbf{x}^*$ on funktsiooni f miinimum.*

Ülesanne: Sõnasta eelmiste kahe väidete erijuhud funktsiooni $f : \mathbb{R} \to \mathbb{R}$ jaoks.