Optimalizace

Metody hledání volných lokálních extrémů – pokračování

Tomáš Kroupa, Mirko Navara

11. 4. 2023

FEL ČVUT

Newtonova metoda minimalizace funkce

Hledáme možné lokální minimum dvakrát diferencovatelné funkce f jako řešení rovnice $\mathbf{g}(\mathbf{x}) = \mathbf{0}$, kde $\mathbf{g}(\mathbf{x}) := \nabla f(\mathbf{x}) = f'(\mathbf{x})^{\mathsf{T}}$.

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\,\mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1}\,f'(\mathbf{x}_k)^{\mathsf{T}}\,.$$

- Používá affinní aproximaci zobrazení $\mathbf{g} = f'$, a tedy kvadratickou aproximaci funkce f.
- Zatímco při řešení obecné soustavy $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ může být Jacobiho matice $\mathbf{g}'(\mathbf{x}_k)$ zobrazení $\mathbf{g}$ libovolná, zde je *symetrická*, neboť je to Hessova matice $f''(\mathbf{x}_k)$ funkce f.
- Přesto může být výpočet této matice a její inverze pracný.
- Je-li $f'(\mathbf{x}_k) = \mathbf{0}$, pak jsme našli stacionární bod (řešení?).

Kdy je Newtonův směr sestupný (mimo stacionární body)

Pro $f'(\mathbf{x}_k) \neq \mathbf{0}$ je Newtonův směr $-f''(\mathbf{x}_k)^{-1}\, f'(\mathbf{x}_k)^\mathsf{T}$ sestupný, pokud je matice $f''(\mathbf{x}_k)$ pozitivně definitní.

Obecněji:

Tvrzení

Je-li $\mathbf{x} \in \mathbb{R}$, $\mathbf{x} \neq \mathbf{0}$, a $\mathbf{A} \in \mathbb{R}^{n \times n}$ symetrická pozitivně definitní matice, pak úhel $\angle(\mathbf{x}, \mathbf{A}\,\mathbf{x})$ je ostrý.

Důkaz

Výraz

$$\mathbf{x}^\mathsf{T}\,\mathbf{A}\,\mathbf{x}$$

je současně skalárním součinem vektorů $\mathbf{x}$, $\mathbf{A}\,\mathbf{x}$ i pozitivně definitní kvadratickou formou, takže je kladný.

Totéž platí i pro inverzní matici $\mathbf{A}^{-1}$.

Nelineární metoda nejmenších čtverců pro řešení soustavy rovnic

Nechť $\mathbf{g}\colon \mathbb{R}^n \to \mathbb{R}^m$ je diferencovatelné zobrazení. Hledáme přibližné řešení přeurčené soustavy $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ ve smyslu nejmenších čtverců.

Úloha: Minimalizujte funkci

$$f(\mathbf{x}) := \|\mathbf{g}(\mathbf{x})\|^2 = \sum_{i=1}^{m} g_i(\mathbf{x})^2, \qquad \mathbf{x} \in \mathbb{R}^n .$$

Gaussova-Newtonova metoda

Speciální případ: přeurčená nehomogenní soustava lineárních rovnic

$$\mathbf{g}(\mathbf{x}) = \mathbf{A}\mathbf{x} - \mathbf{b}$$

vede na optimální řešení

$$\mathbf{x} = \underbrace{(\mathbf{A}^\top \mathbf{A})^{-1} \mathbf{A}^\top}_{\mathbf{A}^+} \mathbf{b}\,.$$

Zobrazení $\mathbf{g}$ aproximujeme v okolí bodu $\mathbf{x}_k$ afinním zobrazením $\mathbf{T}_1$,

$$\mathbf{g}(\mathbf{y}) \approx \mathbf{T}_1(\mathbf{y}) = \underbrace{\mathbf{g}(\mathbf{x}_k)}_{-\mathbf{b}} + \underbrace{\mathbf{g}'(\mathbf{x}_k)}_{\mathbf{A}}\,\underbrace{(\mathbf{y} - \mathbf{x}_k)}_{\mathbf{x}}$$

a místo hledání minima $\|\mathbf{g}(\mathbf{y})\|^2$ volíme

$$\begin{aligned}\mathbf{x}_{k+1} - \mathbf{x}_k &= \operatorname*{argmin}_{\mathbf{y}} \|\mathbf{T}_1(\mathbf{y})\|^2 = \mathbf{A}^+\,\mathbf{b} = \\ &= -\mathbf{g}'(\mathbf{x}_k)^+\,\mathbf{g}(\mathbf{x}_k) = -(\mathbf{g}'(\mathbf{x}_k)^\top\,\mathbf{g}'(\mathbf{x}_k))^{-1}\,\mathbf{g}'(\mathbf{x}_k)^\top\,\mathbf{g}(\mathbf{x}_k)\,.\end{aligned}$$

Gaussova-Newtonova metoda

Iterace Gaussovy-Newtonovy metody (přibližného řešení soustavy m rovnic o n neznámých)

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^+\,\mathbf{g}(\mathbf{x}_k) = \mathbf{x}_k - \left(\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\,\mathbf{g}'(\mathbf{x}_k)\right)^{-1}\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\,\mathbf{g}(\mathbf{x}_k)\,.$$

Jacobiho matice $\mathbf{g}'(\mathbf{x}_k)$ musí mít LN sloupce, pak je $\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\,\mathbf{g}'(\mathbf{x}_k)$ pozitivně definitní.
Speciální případ:

Iterace Newtonovy metody (řešení soustavy n rovnic o n neznámých)

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{g}'(\mathbf{x}_k)^{-1}\,\mathbf{g}(\mathbf{x}_k)\,.$$

Nelineární metoda nejmenších čtverců jako optimalizační úloha

Minimalizujeme $f(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2 = \mathbf{g}(\mathbf{x})^\mathsf{T}\, \mathbf{g}(\mathbf{x})$, kde

$$f'(\mathbf{x}) = 2\, \mathbf{g}(\mathbf{x})^\mathsf{T}\, \mathbf{g}'(\mathbf{x})\,,$$
$$f''(\mathbf{x}) = 2\, \mathbf{g}'(\mathbf{x})^\mathsf{T}\, \mathbf{g}'(\mathbf{x}) + 2 \sum_i g_i(\mathbf{x})\, g_i''(\mathbf{x})\,.$$

Iterace Gaussovy-Newtonovy metody

$$\begin{aligned}\mathbf{x}_{k+1} &= \mathbf{x}_k - \left(\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}'(\mathbf{x}_k)\right)^{-1} \mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}(\mathbf{x}_k)\\ &= \mathbf{x}_k - \left(2\, \mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}'(\mathbf{x}_k)\right)^{-1} f'(\mathbf{x}_k)^\mathsf{T}\,.\end{aligned}$$

Iterace Newtonovy metody (optimalizační)

$$\mathbf{x}_{k+1} = \mathbf{x}_k - f''(\mathbf{x}_k)^{-1}\, f'(\mathbf{x}_k)^\mathsf{T}\,.$$

Nelineární metoda nejmenších čtverců jako optimalizační úloha

Iterace Gaussovy-Newtonovy metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left(2\,\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\,\mathbf{g}'(\mathbf{x}_k)\right)^{-1} f'(\mathbf{x}_k)^\mathsf{T}\,.$$

Má-li $\mathbf{g}'(\mathbf{x}_k)$ LN sloupce, je $\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\,\mathbf{g}'(\mathbf{x}_k)$ pozitivně definitní.

Iterace gradientní metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k\, f'(\mathbf{x}_k)^\mathsf{T}\,.$$

Levenbergova-Marquardtova metoda

Iterace Gaussovy-Newtonovy metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left(\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}'(\mathbf{x}_k)\right)^{-1} \mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}(\mathbf{x}_k)\,.$$

Iterace Levenbergovy-Marquardtovy metody

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left(\mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}'(\mathbf{x}_k) + \mu_k\, \mathbf{I}\right)^{-1} \mathbf{g}'(\mathbf{x}_k)^\mathsf{T}\, \mathbf{g}(\mathbf{x}_k)\,.$$

Regularizační parametr $\mu_k > 0$ umožňuje plynule kombinovat mezi

- Gaussovou-Newtonovou metodou (pro μ_k malé),
- gradientní metodou (pro μ_k velké).

Zvolíme konstantu $q > 1$, např. 2, začneme velkou hodnotou μ_0,

- když se účelová funkce sníží, krok přijmeme a $\mu_{k+1} := \mu_k / q$,
- když se účelová funkce zvýší, krok odmítneme a $\mu_{k+1} := \mu_k\, q$.

Jak snadno najít sestupný směr diferencovatelné funkce

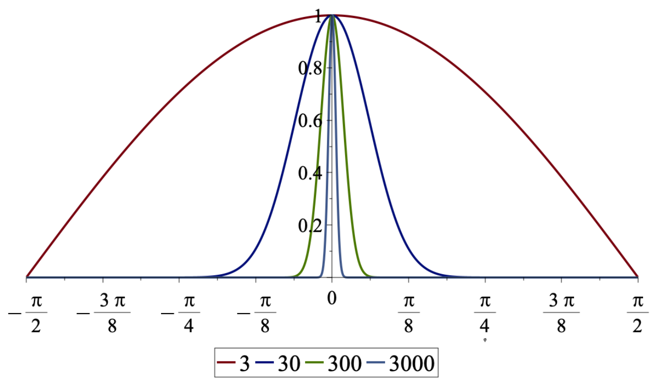
Pro náhodně zvolený vektor $\mathbf{v}_k \in \mathbb{R}^n$ vypočteme směrovou derivaci

- je-li nulová, možná už jsme ve stacionárním bodě;
- je-li záporná, $\mathbf{v}_k$ je sestupný směr;
- je-li kladná, $-\mathbf{v}_k$ je sestupný směr.

Mohli bychom tedy směr volit náhodně, ale snažili jsme se postupovat lépe a rychleji.

Prokletí dimenzionality

Ve velké dimenzi jsou skoro všechny směry skoro kolmé:



Tvar (neznormované!) hustoty rozdělení zeměpisných šířek z rovnoměrného rozdělení na sféře v *n* dimenzích

Prokletí dimenzionality

Alternativa: Soustřeďme se na menší počet dimenzí, v tom umíme vybrat směr a nedá to moc práce.

Opakujme pro další dimenze
$\implies$ *BCD = block coordinate descent*.

Někdy se dá úloha rozložit na více menších, které se neovlivňují, nebo se aspoň ovlivňují málo.

Vhodné dimenze umí vybrat *PCA = principal component analysis*.

Reference

- T. Werner. *Optimalizace* (kapitola 8). Elektronická skripta. FEL ČVUT, 17. 10. 2022. https://cw.fel.cvut.cz/wiki/_media/courses/b0b33opt/opt.pdf
- J. Hamhalter a J. Tišer. *Diferenciální počet funkcí více proměnných*. FEL ČVUT, 2005. https://math.fel.cvut.cz/en/people/tiser/dipocet.pdf
- G. Goh. *Why Momentum Really Works.* https://distill.pub/2017/momentum/
- *Matlab Activities for Multivariable Calculus.* https://mse.redwoods.edu/darnold/math50c/matlab/
- M. Navara. *Matematika spojitého světa*. Elektronická skripta. FEL ČVUT, 2011. https://cmp.felk.cvut.cz/~navara/MSS/MSS_print.pdf