Tehnici de Optimizare

Facultatea de Matematica si Informatica

Universitatea Bucuresti

- Department Informatica-

2021

Clasificare metode de optimizare

Informatia ce indica comportamentul unei functii $f \in \mathbb{R}^n \to \mathbb{R}$ intr-un punct $x \in \mathbb{R}^n$ se poate clasifica:

- Informatie de ordin 0: $f(x)$
- Informatie de ordin 1: $f(x), \nabla f(x)$
- Informatie de ordin 2: $f(x), \nabla f(x), \nabla^2 f(x)$
- ...

Fie algoritmul iterativ definit de $x_{k+1} = \mathcal{M}(x_k)$; in functie de ordinul informatiei utilizate in expresia lui $\mathcal{M}$:

- Metode de ordin 0: $f(x_k)$
- Metode de ordin 1: $f(x_k), \nabla f(x_k)$
- Metode de ordin 2: $f(x_k), \nabla f(x_k), \nabla^2 f(x_k)$

Metoda Gradient

Cea mai "simpla" metoda de ordinul I: Metoda Gradient

- Prima aparitie in lucrarea [1] a lui Augustin-Louis Cauchy, 1847
- Cauchy rezolva un sistem neliniar de ecuatii cu 6 necunoscute, utilizand Metoda Gradient

[1] A. Cauchy. *Methode generale pour la resolution des systemes dequations simultanees.* C. R. Acad. Sci. Paris, 25:536-538, 1847

Metoda Gradient - iteratie

Iteratia MG:

$$x^{k+1} = x^k - \alpha_k \nabla f(x^k)$$

Variatii ale MG se obtin prin alegeri diferite ale pasului, in functie de context.

Exemplu: daca $f(x) = \frac{1}{2}\|Ax - b\|^2 \geq 0, A \in R^{m \times n}$, atunci

$$x^{k+1} = x^k - \alpha A^T(Ax^k - b)$$

Per iteratie: $O(mn)$

Garantii de convergenta

- Fie f diferentiabila cu ∇f *Lipschitz continuu* (constanta Lipschitz $L > 0$)
- Alegem lungimea pasului $\alpha_k = \frac{1}{L}$

Atunci rata de convergenta globala a sirului x_k generat de metoda gradient este subliniara, data de:

$$\min_{0 \leq i \leq k} \|\nabla f(x_i)\| \leq \frac{1}{\sqrt{k}} \sqrt{2L(f(x_0) - f^*)}$$

Garantii de convergenta

$$\min_{\{1 \leq i \leq k\}} \left\| \nabla f(x^i) \right\| \leq \frac{c}{\sqrt{k}}$$

Pentru a atinge acuratete $\left\| \nabla f(x^k) \right\| \leq 10^{-3}$ atunci deducem:

$$\min_{\{1 \leq i \leq k\}} \left\| \nabla f(x^i) \right\| \leq \frac{c}{\sqrt{k}} \leq 10^{-3} \Rightarrow k \geq \frac{c^2}{10^{-6}} = o(10^6)$$

Garantii de convergenta

$$\min_{\{1\le i\le k\}} \left\|\nabla f(x^i)\right\| \le \frac{c}{\sqrt{k}}$$

Pentru a atinge acuratete $\left\|\nabla f(x^k)\right\| \le 10^{-2}$ atunci deducem:

$$\min_{\{1\le i\le k\}} \left\|\nabla f(x^i)\right\| \le \frac{c}{\sqrt{k}} \le 10^{-2} \Rightarrow k \ge \frac{c^2}{10^{-2}} = o(10^4)$$

Sau in general, pentru $\left\|\nabla f(x^k)\right\| \le \epsilon$ sunt necesare:

$$\min_{\{1\le i\le k\}} \left\|\nabla f(x^i)\right\| \le \frac{c}{\sqrt{k}} \le \epsilon \Rightarrow k \ge \frac{c^2}{\epsilon^2} \text{ iteratii}$$

Garantii de convergenta

Ratele de convergenta de tipul: $p > 0$

$$\|x^k - x^*\| \leq \frac{C}{k^p} \qquad\qquad f(x^k) - f^* \leq \frac{C}{k^p}$$

$$\|\nabla f(x^k)\| \leq \frac{C}{k^p}$$

se numesc **subliniare.** Necesarul de iteratii pentru ϵ acuratete se obtine:

$$\|\nabla f(x^k)\| \leq \frac{C}{k^p} \leq \epsilon \quad \Rightarrow k \geq \left(\frac{C}{\epsilon}\right)^{\frac{1}{p}}$$

Garantii de convergenta

Sub presupunerea ca ∇f Lipschitz continuu avem:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}\|y - x\|^2 \quad x, y \in \operatorname{dom} f.$$

Considerand $x = x_k, y = x_{k+1} = x_k - (1/L)\nabla f(x_k)$ avem:

$$f(x_{k+1}) \leq f(x_k) - \frac{1}{2L}\|\nabla f(x_k)\|^2.$$

Insumam dupa $i = 0, \dots k-1$ si rezulta

$$\frac{1}{2L}\sum_{i=0}^{k-1}\|\nabla f(x_i)\|^2 \leq f(x_0) - f(x_k) \leq f(x_0) - f^*$$

In concluzie, observam

$$k \min_{0 \leq i \leq n}\|\nabla f(x_i)\|^2 \leq \sum_{i=0}^{k-1}\|\nabla f(x_i)\|^2 \leq 2L(f(x_0) - f^*)$$

Garantii de convergenta a sirului

- Fie f diferentiabila cu ∇f *Lipschitz continuu* (constanta Lipschitz $L > 0$)
- Exista un punct de minim local x^*, astfel incat Hessiana in acest punct satisface

$$\sigma I_n \preceq \nabla^2 f(x^*) \preceq L I_n$$

- Punctul initial x_0 al iteratiei metodei gradient cu pas $\alpha_k = \frac{2}{\sigma + L}$ este suficient de aproape de punctul de minim, i.e.

$$\|x_0 - x^*\| \leq \frac{2\sigma}{L}$$

Atunci rata de convergenta locala a sirului x_k generat de metoda gradient este liniara (i.e de ordinul $\mathcal{O}\left(\log(\frac{1}{\epsilon})\right)$), data de:

$$\|x_k - x^*\| \leq \beta \left(1 - \frac{2\sigma}{L + 3\sigma}\right)^k \qquad \text{cu } \beta > 0$$

Garantii de convergenta

Ratele de convergenta de tipul: fie $\gamma \in (0,1), C > 0$

$$\left\|\nabla f(x^k)\right\| \quad sau \;\; f\left(x^k\right) - f^* \quad sau \;\; \left\|x^k - x^*\right\| \leq (1-\gamma)^k C$$

se numesc **liniare (geometrice).**

Necesarul de iteratii pentru ϵ acuratete se obtine:

$$\left\|\nabla f(x^k)\right\| \leq (1-\gamma)^k C \leq \epsilon$$

$$klog\left(\frac{1}{1-\gamma}\right) \geq \log\frac{C}{\epsilon} \Rightarrow k \geq \frac{1}{\log(1/1-\gamma)}\log\frac{C}{\epsilon}$$

Mai simplu, folosind $\log(1-\gamma) \leq -\gamma$, e sufficient:

$$k \geq \frac{1}{\gamma}\log\frac{C}{\epsilon}$$

Garantii de convergenta

Mai simplu, folosind $\log(1-\gamma) \leq -\gamma$, e sufficient:

$$k \geq \frac{1}{\gamma} \log \frac{C}{\epsilon}$$

Pentru a atinge acuratete $\epsilon = 10^{-10}$, sunt necesare

$$k \geq \frac{1}{\gamma}(10 + \log(C))$$

Parametrul γ (de obicei, nr. de conditionare) este dominant!

Garantii de convergenta – sub convexitate

- Fie f functie convexa, diferentiabila cu ∇f Lipschitz continuu (constanta Lipschitz $L > 0$). Daca alegem lungimea pasului constanta $\alpha_k = \frac{1}{L}$, atunci rata de convergenta globala a sirului x_k generat de metoda gradient este subliniara, data de:

$$f(x_k) - f^* \leq \frac{L\|x_0 - x^*\|^2}{2k}$$

- Daca in plus functia este tare convexa cu constanta $\sigma > 0$, atunci rata de convergenta globala a sirului x_k generat de metoda gradient cu pas $\alpha_k = \frac{1}{L}$ este liniara, data de:

$$\|x_k - x^*\|^2 \leq \left(\frac{L - \sigma}{L + \sigma}\right)^k \|x_0 - x^*\|^2$$

Garantii de convergenta a sirului

$$\frac{L-\sigma}{L+\sigma} = \frac{L\left(1-\frac{\sigma}{L}\right)}{L\left(1+\frac{\sigma}{L}\right)} = \frac{\left(1-\frac{\sigma}{L}\right)}{\left(1+\frac{\sigma}{L}\right)} \leq 1-\frac{\sigma}{L} = 1-\frac{1}{\mu}$$

Pentru a atinge ϵ-acuratete, ratele de convergenta precedente ale MG indica necesitatea a $\approx \mu \log\left(\frac{1}{\epsilon}\right)$ iteratii!

Rata MG este determinata de $\boldsymbol{\mu = \frac{L}{\sigma}}$ (numarul de conditionare al problemei)!

Observati:

$\mu \to \infty$, atunci $\left(1-\frac{1}{\mu}\right)^k \to 1^k$

$\mu \to 1$, atunci $\left(1-\frac{1}{\mu}\right)^k \to 0^k$

Garantii de convergenta a sirului

$$H = \begin{bmatrix} 0.5 & 0 \\ 0 & \gamma \end{bmatrix} = \begin{bmatrix} 0.5 & 0 \\ 0 & 5/3 \end{bmatrix}$$

$$\mu(H) = \frac{L}{\sigma} = \frac{10}{3} \approx 3{,}33$$

$$\left|x^k - x^*\right|^2 \leq \left(1 - \frac{3}{10}\right)^k \frac{7}{25} = \left(\frac{7}{10}\right)^k \frac{7}{25}$$

$$|x^{13} - x^*|^2 \leq \left(\frac{7}{10}\right)^{13} \frac{7}{25} \leq \left(\frac{7}{10}\right)^{13} \approx 0.0097$$

Metoda Newton

- Fie iteratia x^k aproximam functia obiectiv $f(x)$ (in jurul lui x^k) cu termenul Taylor de ordin II:

$$f(x) \approx f\left(x^k\right) + \nabla f\left(x^k\right)^T\left(x - x^k\right) + \frac{1}{2}\left(x - x^k\right)^T \nabla^2 f(x^k)(x - x^k)$$

- Partea drepta este o functie patratica (rezolvam prin cond. de ordin I)

$$\nabla^2 f\left(x^k\right)\left(x - x^k\right) + \nabla f\left(x^k\right) = 0$$

Deci:

$$x = x^k - \left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right) \qquad (:= x^{k+1})$$

Metoda Newton

$$x^{k+1} = x^k - \alpha_k \left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right)$$

Alegeri ale pasului:

- Constant $\alpha = 1$
- Ideal: $\alpha_k = \arg\min f(x^k - \alpha \left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f\left(x^k\right))$
- Backtracking: descrestere α_k pana satisfacem $(\gamma > 0)$
- $f\left(x^{k+1}\right) \leq f\left(x^k\right) - \alpha_k \gamma \nabla f\left(x^k\right)^T \left[\nabla^2 f\left(x^k\right)\right]^{-1} \nabla f(x^k)$

Metoda Newton

THEOREM 1. Let $f(x)$ be twice differentiable, let $\nabla^2 f(x)$ satisfy a Lipschitz condition with constant L, let $f(x)$ be strongly convex with constant ℓ, and let the initial approximation satisfy the condition

$$q = (L\ell^{-2}/2)\|\nabla f(x^0)\| < 1 . \qquad (5)$$

Then method (2) converges to the global minimum point x^* with the quadratic rate:

$$\|x^k - x^*\| \leq (2\ell/L)q^{2^k} . \qquad (6)$$

Metoda Newton

Presupunerile se pot relaxa la functii neconvexe:

- f dublu diferentiabila
- Exista un minim local nesingular: $\nabla^2 f(x^*) \succ \sigma I_n$
- x^0 in vecinatatea lui x^*

In acest caz, MN converge patratic la x^*!

Metoda Newton

Convergenta patratica:

$$q < 1 \ (e.g.\ q = 0,8)$$

- $k = 1$: $q^2 \ (e.g. = 0,64)$
- $k = 2$: $q^4 \ (e.g. = 0,4096)$
- $k = 3$: $q^8 \ (e.g. = 0,1678)$
- $k = 4$: $q^{16} \ (e.g. = 0,0282)$
- $k = 5$: $q^{32} \ (e.g. = 0,0008)$
- $k = 6$: $q^{64} \ (e.g. = 6,3e - 7)$