Um método de descida simples com passo constante para $\min_{x \in \mathbb{R}^n} f(x)$

1

Objetivo: apresentar um esquema de descida simples

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

onde o passo $\alpha > 0$ é o mesmo para toda iteração; Estudar a complexidade deste algoritmo

Taylor de $1^{\underline{a}}$ ordem (aproximação linear)

$$f(x+d) = f(x) + \nabla f(x)^t d + \underbrace{o(\|d\|)}_{},$$

onde $\dfrac{o(\|d\|)}{\|d\|} \xrightarrow[\|d\| \to 0]{} 0$.

<u>Pergunta</u>: quem é $o(\|d\|)$?

Ou "qual a qualidade da aproximação linear de f em x"?

- $\varphi(t) = f(x + td)$
- $\varphi'(t) = \nabla f(x + td)^t d$
- $\varphi(0) = f(x)$, $\varphi(1) = f(x + d)$

<u>TFC</u> $\Rightarrow$ $\varphi(1) - \varphi(0) = \int_0^1 \varphi'(t)\,dt$

$= \int_0^1 [\varphi'(t) - \varphi'(0) + \varphi'(0)]\,dt$

$= \varphi'(0) + \int_0^1 [\varphi'(t) - \varphi'(0)]\,dt$

$$\Rightarrow \boxed{\begin{aligned} f(x+d) &= f(x) + \nabla f(x)^t d \\ &\quad + \underbrace{\int_0^1 \left(\nabla f(x+td) - \nabla f(x)\right)^t d \, dt}_{o(\|d\|)} \end{aligned}}$$

A fim de que o método com passo constante convirja, a função f deve ter gradiente Lipschitz.

HIPÓTESE: existe uma cte $L > 0$ tq

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x-y\|, \ \forall x, y.$$

Com esta hipótese, conseguimos simplificar a medida de qualidade da aproximação linear de f:

$$\left| f(x+d) - \left[f(x) + \nabla f(x)^t d \right] \right|$$

$$= \left| \int_0^1 (\nabla f(x+td) - \nabla f(x))^t d \, dt \right|$$

$$\leq \int_0^1 \left| (\nabla f(x+td) - \nabla f(x))^t d \right| dt$$

$$\leq \int_0^1 \|\nabla f(x+td) - \nabla f(x)\| \cdot \|d\| \, dt$$

$$\leq \int_0^1 L \|x+td-x\| \cdot \|d\| \, dt$$

$$= L \int_0^1 t \|d\|^2 dt = \frac{L}{2} \|d\|^2 .$$

Em particular, $\boxed{f(x+d) \leq f(x) + \nabla f(x)^t d + \frac{L}{2}\|d\|^2}$

Voltando ao método:

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

- $d = -\alpha \nabla f(x^k)$
- $\alpha = \dfrac{1}{L}$ (constante)
- $x = x^k$

$$\left.\begin{array}{l} d = -\alpha \nabla f(x^k) \\ \alpha = \frac{1}{L} \\ x = x^k \end{array}\right\} \begin{aligned} x + d &= x^k - \frac{1}{L}\nabla f(x^k) \\ &= x^{k+1}. \end{aligned}$$

Usando a desig. anterior,

$$f(x^{k+1}) = f\left(x^k - \frac{1}{L}\nabla f(x^k)\right)$$

$$\leq f(x^k) + \nabla f(x^k)^t \left(-\frac{1}{L}\nabla f(x^k)\right)$$

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

$$= f(x^k) - \frac{1}{L}\|\nabla f(x^k)\|^2 + \frac{L}{2L^2}\|\nabla f(x^k)\|^2$$

$$\Rightarrow \boxed{f(x^{k+1}) \leq f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|^2}$$

(f decresce!)

Isto mostra que

$$\underbrace{f(x^k) \to -\infty}_{f \text{ ilimitada}} \quad \text{ou} \quad \underbrace{\nabla f(x^k) \to 0}_{x^* \text{ solução}}$$

<u>Ponto fraco</u>: raramente conhecemos L.

↳ Uma solução: ao longo do processo, começar com $L > 0$ pequeno e ir ajustando...

Complexidade

Para f com ∇f Lipschitz, o método converge...

Pergunta: quão rápido?

ou "quantos passos são suficientes para alcançar $\|\nabla f(x^k)\| \leq \varepsilon$ para uma precisão $\varepsilon > 0$ dada"?

<u>Teorema</u>: $$\min_{0 \leq j \leq k} \|\nabla f(x^j)\| \leq \sqrt{\frac{2L(f(x^0)-f^*)}{k}},$$

onde f^* é o valor ótimo de f

(supomos que $f \in C^1$ limitada inferiormente)

Portanto, dado $\varepsilon > 0$, o número de iterações k suficientes para atingir $\|\nabla f(x^k)\| \leq \varepsilon$ é proporcional à $\frac{1}{\varepsilon^2}$.

Prova: $$f(x^{k+1}) \leq f(x^k) - \frac{1}{2L}\|\nabla f(x^k)\|^2, \quad \forall k$$

$$\Rightarrow f^* \leq f(x^{k+1}) \leq f(x^0) - \frac{(k+1)}{2L}\min_{0\leq j\leq k}\|\nabla f(x^j)\|^2.$$

$$\Rightarrow \min_{0\leq j\leq k}\|\nabla f(x^j)\| \leq \sqrt{\frac{2L(f(x^0)-f^*)}{k}}.$$

Para que $\|\nabla f(x^k)\| \leq \varepsilon$, é suficiente que

$$\sqrt{\frac{2L(f(x^0)-f^*)}{k}} \leq \varepsilon \Leftrightarrow \frac{2L(f(x^0)-f^*)}{\varepsilon^2} \leq k$$

Ou seja, a fim de $\|\nabla f(x^k)\| \leq \varepsilon$, é suficiente que o número de iterações seja pelo menos

$$[2L(f(x^0) - f^*)] \frac{1}{\varepsilon^2}.$$ ■

Alguns ~~textos~~ escrevem $O(1/\varepsilon^2)$:

"ordem de ε^{-2}". <u>Exemplos</u>: $\varepsilon = 10^{-4} \Rightarrow$

$\varepsilon^{-2} = 10^8$; $\varepsilon = 10^{-6} \Rightarrow \varepsilon^{-2} = 10^{12}$; $\varepsilon = 10^{-8} \Rightarrow \varepsilon^{-2} = 10^{16}$