SME0212/SME5720 – Otimização Não-Linear

Direções de descida e busca de Armijo

Elias S. Helou Neto

Diferenciabilidade

Dizemos que uma função $f : \mathbb{R}^n \to \mathbb{R}$ é diferenciável em um ponto $\boldsymbol{x} \in \mathbb{R}^n$ quando

$$f(\boldsymbol{x} + \boldsymbol{h}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{h} + \|\boldsymbol{h}\| E(\boldsymbol{x}, \boldsymbol{h})$$

onde

$$\lim_{\boldsymbol{h} \to \boldsymbol{0}} E(\boldsymbol{x}, \boldsymbol{h}) = 0.$$

Direção de descida

Dizemos que $\boldsymbol{d} \in \mathbb{R}^n$ é uma direção de descida para $f : \mathbb{R}^n \to \mathbb{R}$ a partir de $\boldsymbol{x} \in \mathbb{R}^n$ quando existe $\Lambda > 0$ tal que para todo $\lambda \in (0, \Lambda)$ temos que

$$f(\boldsymbol{x} + \lambda \boldsymbol{d}) < f(\boldsymbol{x}).$$

Direção de máxima descida

Se $f : \mathbb{R}^n \to \mathbb{R}$ é diferenciável em $\boldsymbol{x} \in \mathbb{R}^n$ e $\nabla f(\boldsymbol{x}) \neq \mathbf{0}$ então $-\nabla f(\boldsymbol{x})$ é uma direção de descida para f a partir de $\boldsymbol{x}$.

Direção de máxima descida

Se $f : \mathbb{R}^n \to \mathbb{R}$ é diferenciável em $\boldsymbol{x} \in \mathbb{R}^n$ e $\nabla f(\boldsymbol{x}) \neq \boldsymbol{0}$ então, para qualquer $\sigma \in (0, 1)$, existe $\Lambda_\sigma > 0$ tal que para $\lambda \in (0, \Lambda_\sigma)$ temos:

$$f\big(\boldsymbol{x} - \lambda \nabla f(\boldsymbol{x})\big) < f(\boldsymbol{x}) - \lambda \sigma \|\nabla f(\boldsymbol{x})\|^2$$

Algoritmo de máxima descida

Passo fixo

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\lambda > 0$, $\epsilon > 0$
Result: $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\nabla f(\boldsymbol{x})\| \leq \epsilon$
$k \leftarrow 0$;
while $\|\nabla f(\boldsymbol{x}_k)\| > \epsilon$ **do**
    $x_{k+1} \leftarrow x_k - \lambda \nabla f(\boldsymbol{x}_k)$;
    $k \leftarrow k + 1$;
end
$\boldsymbol{x} \leftarrow \boldsymbol{x}_k$;

Algoritmo 1: Máxima descida com passo fixo

Algoritmo de máxima descida

Busca linear de Armijo

O mesmo passo fixo ao longo de todas as iterações atrasa o algoritmo.

O mesmo passo fixo ao longo de todas as iterações atrasa o algoritmo.

Dado $\sigma \in (0, 1)$, sabemos que, em uma iteração k, existe $\Lambda_k > 0$ tal que para $\lambda_k \in (0, \Lambda_k)$ temos

$$f(\boldsymbol{x}_k - \lambda_k \nabla f(\boldsymbol{x}_k)) < f(\boldsymbol{x}_k) - \lambda_k \sigma \|\nabla f(\boldsymbol{x}_k)\|^2.$$

Algoritmo de máxima descida

Busca linear de Armijo

O mesmo passo fixo ao longo de todas as iterações atrasa o algoritmo.

Dado $\sigma \in (0,1)$, sabemos que, em uma iteração k, existe $\Lambda_k > 0$ tal que para $\lambda_k \in (0, \Lambda_k)$ temos

$$f(\boldsymbol{x}_k - \lambda_k \nabla f(\boldsymbol{x}_k)) < f(\boldsymbol{x}_k) - \lambda_k \sigma \|\nabla f(\boldsymbol{x}_k)\|^2.$$

A ideia de Armijo é escolher o maior valor da forma

$$\{\lambda, \lambda\beta, \lambda\beta^2, \lambda\beta^3, \dots\}$$

que satisfaça tal condição de descida. Aqui, $\beta \in (0,1)$.

Algoritmo de máxima descida

Busca linear de Armijo

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\lambda > 0$, $\beta \in (0,1)$, $\sigma \in (0,1)$, $\epsilon > 0$
Result: $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\nabla f(\boldsymbol{x})\| \leq \epsilon$

```
k ← 0;
while ‖∇f(x_k)‖ > ε do
    λ_k ← λ;
    while f(x_k − λ_k ∇f(x_k)) ≥ f(x_k) − λ_k σ ‖∇f(x_k)‖² do
        λ_k ← λ_k β;
    end
    x_{k+1} ← x_k − λ_k ∇f(x_k);
    k ← k + 1;
end
x ← x_k;
```

Algoritmo 2: Máxima descida com busca de Armijo

Algoritmo de descida

Busca linear de Armijo

Data: $\boldsymbol{x}_0 \in \mathbb{R}^n$, $\lambda > 0$, $\beta \in (0,1)$, $\sigma \in (0,1)$, $\epsilon > 0$, $\gamma \in (0,1)$, $\kappa \in (0,1)$
Result: $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\nabla f(\boldsymbol{x})\| \leq \epsilon$
$k \leftarrow 0$;
while $\|\nabla f(\boldsymbol{x}_k)\| > \epsilon$ **do**
    $\lambda_k \leftarrow \lambda$;
    Seja $\boldsymbol{d}_k$ tal que

$$\boldsymbol{d}_k^T \nabla f(\boldsymbol{x}_k) < -\gamma \|\boldsymbol{d}_k\| \|\nabla f(\boldsymbol{x}_k)\|$$

    e

$$\|\boldsymbol{d}_k\| > \kappa \|\nabla f(\boldsymbol{x}_k)\|$$

    while $f(\boldsymbol{x}_k + \lambda_k \boldsymbol{d}_k) \geq f(\boldsymbol{x}_k) + \lambda_k \sigma \boldsymbol{d}_k^T \nabla f(\boldsymbol{x}_k)$ **do**
        $\lambda_k \leftarrow \lambda_k \beta$;
    end
    $x_{k+1} \leftarrow x_k + \lambda_k \boldsymbol{d}_k$;
    $k \leftarrow k + 1$;
end
$\boldsymbol{x} \leftarrow \boldsymbol{x}_k$;

Algoritmo 3: Descida com busca de Armijo