BFGS Quasi-Newton

(Broyden, Fletcher, Goldfarb, Shanno)

GD: $x_{k+1} = x_k - \eta I \nabla f(x_k)$

N: $x_{k+1} = x_k - \eta \nabla^2 f(x_k)^{-1} \nabla f(x_k)$

- Qualquer aproximação quadrática define uma direção

$$f(x_k + d) \approx m_k(d) = f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T B_k d$$

(B_k: definida positiva)

$$d_k = -B_k^{-1} \nabla f(x_k)$$

Line search

QN: $x_{k+1} = x_k - \eta_k B_k^{-1} \nabla f(x_k)$

↓

Atualizado iterativamente

* Propriedades desejadas

① $B_k^{-1} \nabla f(x_k)$ deve ser fácil de computar

② B_k deve aproximar a Hessiana

③ $m_k(d)$ de capturar informação sobre a curvatura de f na direção d.

BFGS:

$$f(x_k + d) \approx m_k(d) \quad , \quad f(x_{k+1} + d) \approx m_{k+1}(d)$$

Requerimento:

m_{k+1} deve ter o mesmo gradiente de f nas últimas duas iterações.

(i) $\nabla m_{k+1}\Big|_{d=0} = \nabla f(x_{k+1})$

$$\left[x_{k+1} = x_k + \eta_k d_k \right]$$
$$x_k = x_{k+1} - \eta_k d_k \Big]$$

(ii) $\nabla m_{k+1}\Big|_{d=-\eta d_k} = \nabla f(x_k)$

$$m_{k+1}(d) = f(x_{k+1}) + \nabla f(x_{k+1})^T d + \frac{1}{2} d^T B_k d$$

$$\nabla m_{k+1}(d) = \nabla f(x_{k+1}) + B_{k+1} d$$

(i) $\nabla m_{k+1}(0) = \nabla f(x_{k+1})$ - satisfeito automaticamente

(ii) $\nabla m_{k+1}(-\eta_k d_k) = \nabla f(x_{k+1}) - \eta_k B_{k+1} d_k = \nabla f(x_k)$

$$\eta_k B_k d_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

$$x_{k+1} = x_k + \eta_k d_k \leftrightarrow \eta_k d_k = x_{k+1} - x_k$$

$$B_{k+1} \underbrace{(x_{k+1} - x_k)}_{s_k} = \underbrace{\nabla f(x_{k+1}) - \nabla f(x_k)}_{y_k}$$

$$B_{k+1} s_k = y_k \quad (2)$$

* Definindo a aproximação da Hessiana

$B_{k+1} = \arg\min \|B - B_k\|$ (B_{k+1} é próxima de B_k)

$B = B^T$ (simétrica)

$B_{k+1} s_k = y_k$

Each choice of norm gives a different B_{k+1}

BFGS

$H_k = B_k^{-1}$

① H_{k+1} também é próxima de H_k

② H_{k+1} é simétrica

③ $B_{k+1} s_k = y_k \rightarrow s_k = (B_{k+1})^{-1} y_k$

$\rightarrow s_k = H_k y_k$

$H_{k+1} = \arg\min \|H - H_k\|$

s.a. $H = -H^T$

$H y_k = s_k$

Norma de frobenius

$\|A\|_F^2 = \sum_{i,j} \|A_{ij}\|^2$

Norma de frobenius ponderada

$\|A\|_W^2 = \|W^{1/2} A W^{1/2}\|_F^2$

BFGS usa

$$W = \int_0^1 \nabla^2 f(x_k + t\,\eta_k d_k)\, dt$$

↳ definida positiva

Então:

$$H_{k+1} = \left(I - \rho_k s_k y_k^T\right) H_k \left(I - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T$$

$$\rho_k = \frac{1}{y_k^T s_k}$$

Algoritmo BFGS

Dado x_0

$k = 0$

Repetir enquanto $\|\nabla f(x_k)\| > \varepsilon$

$$d_k = -H_k \nabla f(x_k)$$

$$x_{k+1} = x_k + \eta_k d_k = x_k - \eta_k H_k \nabla f(x_k)$$

$$s_k = x_{k+1} - x_k$$

$$y_k = \nabla f(x_{k+1}) - \nabla f(x_k)$$

$$H_{k+1} = \left(I - \rho_k s_k y_k^T\right) H_k \left(I - \rho_k y_k s_k^T\right) + \rho_k s_k s_k^T$$

$$\rho_k = \frac{1}{y_k^T s_k}$$

Busca em Linha

- Decremando suficientemente

$$f(x_k + \eta_k d_k) \leq f(x_k) + \alpha_1 \eta_k \nabla f(x_k)^T d_k$$

- condição de declive

$$\nabla f(x_k + \eta_k d_k)^T d_k \geq \alpha_2 \cdot \nabla f(x_k)^T d_k$$

$0 < \alpha_1 < 1$

$0 < \alpha_2 < 1$