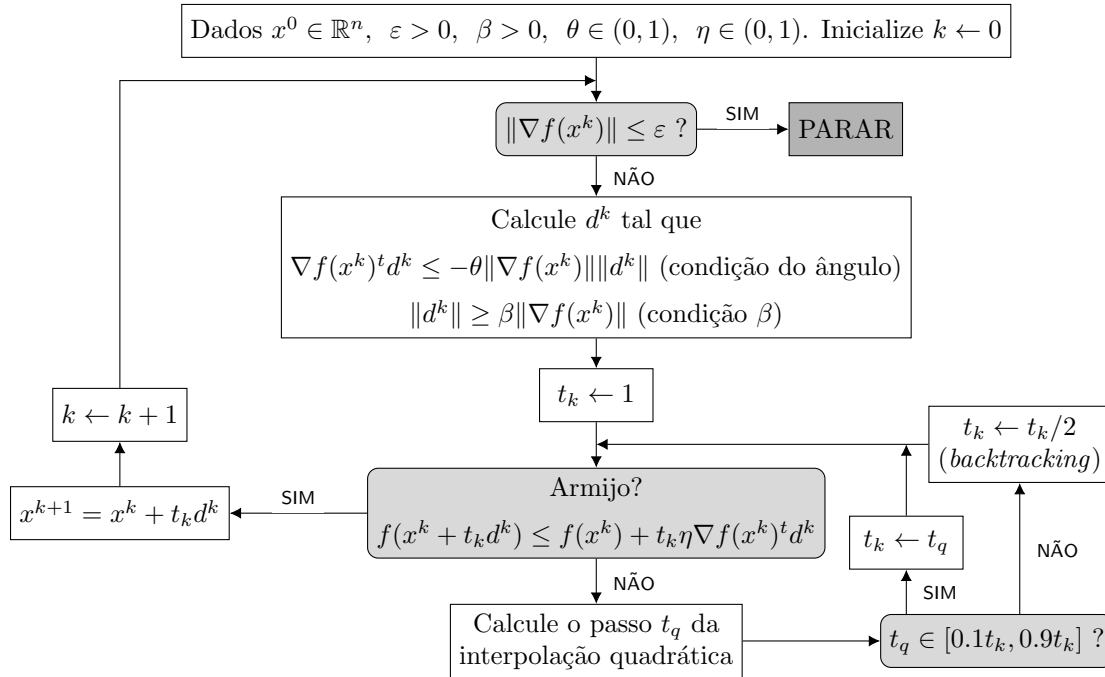


11.1 Método geral de descida com convergência global



Algoritmo 11.1: Esquema geral de descida.

No teorema a seguir, analisamos a convergência global do Algoritmo 11.1. Para tanto, negligenciamos o critério de parada $\|\nabla f(x^k)\| \leq \varepsilon$, permitindo que uma sequência infinita seja gerada.

Teorema 3. *Suponha que f seja continuamente diferenciável. Seja x^* um ponto de acumulação da sequência $\{x^k\}$ gerada pelo esquema geral de descida (Algoritmo 11.1), desconsiderando o critério de parada. Então*

$$\nabla f(x^*) = 0.$$

Demonstração. Seja $\{x^k\}_{k \in N_1}$, $N_1 \subseteq \mathbb{N}$, uma subsequência convergindo à x^* . Primeiramente, podemos supor sem perda de generalidade que $\nabla f(x^k) \neq 0$ para todo $k \in N_1$. Assim, a condição β garante que $d^k \neq 0$ para todo $k \in N_1$. A prova será dividida em dois casos:

Caso 1: $\|t_k d^k\| \geq \varepsilon_1$ para algum $\varepsilon_1 > 0$ e infinitos $k \in N_1$, digamos, $k \in N_2 \subseteq N_1$. Das condições de Armijo e do ângulo, temos, para todo $k \in N_2$,

$$\begin{aligned} f(x^{k+1}) = f(x^k + t_k d^k) &\leq f(x^k) + t_k \eta \nabla f(x^k)^t d^k \\ &\leq f(x^k) - t_k \eta \theta \|\nabla f(x^k)\| \|d^k\| \leq f(x^k) - \eta \theta \varepsilon_1 \|\nabla f(x^k)\|, \end{aligned}$$

o que implica

$$\|\nabla f(x^k)\| \leq \frac{f(x^k) - f(x^{k+1})}{\eta \theta \varepsilon_1} \quad \forall k \in N_2.$$

Como $\nabla f(x^k)^t d^k \leq 0$ e $t_k > 0$, a condição de Armijo obriga a sequência $\{f(x^k)\}$ ser não crescente. Portanto, se $p(k) \geq k + 1$ é o índice em N_2 logo após k , temos

$$0 \leq f(x^k) - f(x^{k+1}) \leq \dots \leq f(x^k) - f(x^{p(k)}),$$

donde segue que

$$\|\nabla f(x^k)\| \leq \frac{f(x^k) - f(x^{p(k)})}{\eta \theta \varepsilon_1} \quad \forall k \in N_2.$$

Como $\lim_{k \in N_2} f(x^k) = \lim_{k \in N_2} f(x^{p(k)}) = f(x^*)$, concluímos da desigualdade acima que $\lim_{k \in N_2} \nabla f(x^k) = 0$, e assim $\nabla f(x^*) = 0$.

Caso 2: $\lim_{k \in N_1} \|t_k d^k\| = 0$. Consideramos dois subcasos:

Caso 2.1: $t_k = 1$ é aceito infinitas vezes, digamos, para $k \in N_3 \subseteq N_1$. Neste caso, só pode ser $\lim_{k \in N_3} \|d^k\| = 0$. Assim, da condição β segue que

$$\|\nabla f(x^*)\| = \lim_{k \in N_3} \|\nabla f(x^k)\| \leq \lim_{k \in N_3} \frac{\|d^k\|}{\beta} = 0,$$

o que implica $\nabla f(x^*) = 0$.

Caso 2.2: $t_k = 1$ é rejeitado para todo $k \geq k_0$, $k \in N_1$. Para cada um desses índices k , seja \tilde{t}_k o passo rejeitado imediatamente antes do passo aceito $t_k < 1$. Independentemente se t_k foi calculado por interpolação quadrática ou *backtracking*, t_k satisfaz as salvaguardas em relação à tentativa anterior \tilde{t}_k , isto é,

$$t_k \in [0.1\tilde{t}_k, 0.9\tilde{t}_k].$$

Em particular, $t_k \geq 0.1\tilde{t}_k$, e logo, para todo $k \geq k_0$, $k \in N_1$, temos $\|\tilde{t}_k d^k\| \leq 10\|t_k d^k\|$, o que implica

$$\lim_{k \in N_1} \|\tilde{t}_k d^k\| = 0. \quad (11.1)$$

Pela definição de \tilde{t}_k , a condição de Armijo falha nestes passos, isto é,

$$f(x^k + \tilde{t}_k d^k) > f(x^k) + \tilde{t}_k \eta \nabla f(x^k)^t d^k, \quad \forall k \geq k_0, k \in N_1. \quad (11.2)$$

Como estamos supondo $d^k \neq 0$ para todo $k \in N_1$, e como os passos gerados pelo método são sempre positivos, a sequência $\frac{\tilde{t}_k d^k}{\|\tilde{t}_k d^k\|}$ está bem definida e é limitada, visto que todos seus termos têm norma igual a 1. Considere assim um ponto de acumulação seu, digamos

$$v^* = \lim_{k \in N_4} \frac{\tilde{t}_k d^k}{\|\tilde{t}_k d^k\|},$$

onde $N_4 \subseteq N_1$. Da condição do ângulo temos

$$\nabla f(x^k)^t \frac{\tilde{t}_k d^k}{\|\tilde{t}_k d^k\|} \leq -\theta \|\nabla f(x^k)\| \left\| \frac{\tilde{t}_k d^k}{\|\tilde{t}_k d^k\|} \right\| = -\theta \|\nabla f(x^k)\|$$

para todo $k \in N_4$. Passando ao limite sobre N_4 , obtemos da continuidade do gradiente de f que

$$\nabla f(x^*)^t v^* \leq -\theta \|\nabla f(x^*)\|. \quad (11.3)$$

Fixado $k \in N_4$, defina a função em uma variável

$$\varphi(\delta) = f(x^k + \delta \tilde{t}_k d^k).$$

Aplicando o Teorema do Valor Médio relativo ao intervalo $[0, 1]$, garantimos a existência de um $\delta_k \in (0, 1)$ tal que

$$\varphi'(\delta_k) = \frac{\varphi(1) - \varphi(0)}{1 - 0},$$

o que implica

$$\nabla f(x^k + \delta_k \tilde{t}_k d^k)^t (\tilde{t}_k d^k) = f(x^k + \tilde{t}_k d^k) - f(x^k).$$

Dividindo esta expressão por $\|\tilde{t}_k d^k\|$ e usando (11.2), obtemos

$$\eta \nabla f(x^k)^t \frac{\tilde{t}_k d^k}{\|\tilde{t}_k d^k\|} \leq \nabla f(x^k + \delta_k \tilde{t}_k d^k)^t \frac{\tilde{t}_k d^k}{\|\tilde{t}_k d^k\|}.$$

Isso vale para todo $k \in N_4$. Como $\delta_k \in (0, 1)$ para todo $k \in N_3$ e $\lim_{k \in N_3} \|\tilde{t}_k d^k\| = 0$ por (11.1), passamos ao limite sobre N_3 na expressão acima para obter $\eta \nabla f(x^*)^t v^* \leq \nabla f(x^*)^t v^*$. Como $\eta < 1$, deve ser $\nabla f(x^*)^t v^* \geq 0$. Com isso, a expressão (11.3) implica

$$-\theta \|\nabla f(x^*)\| \geq 0,$$

donde segue que $\nabla f(x^*) = 0$ visto que $\theta > 0$. Isso conclui a demonstração. \square

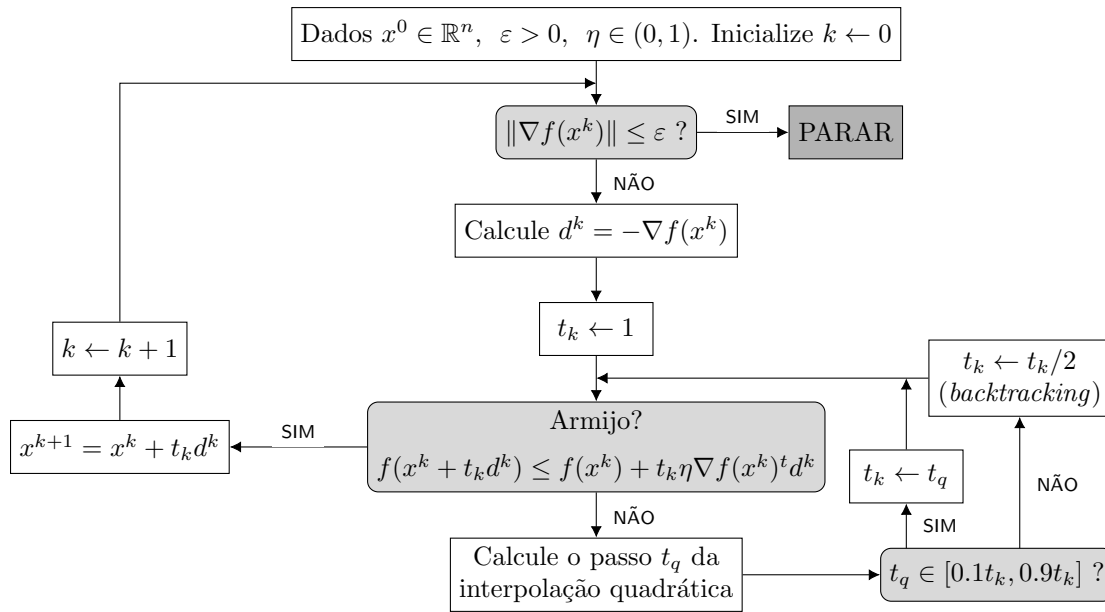
11.2 Método do gradiente

Uma primeira especialização do método geral de descida é relativa à escolha $d^k = -\nabla f(x^k)$. O método resultante é chamado *método do gradiente*. A escolha $d^k = -\nabla f(x^k)$ naturalmente satisfaz as condições do ângulo e β , conforme o resultado a seguir, cuja demonstração é trivial e ficará a cargo do leitor.

Lema 1. *Suponha que f seja diferenciável em x e seja $d = -\nabla f(x)$. Então*

1. $\nabla f(x)^t d \leq -\theta \|\nabla f(x)\| \|d\|$ para todo $\theta \in (0, 1)$;
2. $\|d\| \geq \beta \|\nabla f(x)\|$ para todo $\beta \in (0, 1]$.

O lema acima permite negligenciar a condição do ângulo e escolher $\beta = 1$ no esquema geral de descida (Algoritmo 11.1). O método resultante é resumido no Algoritmo 11.2.



Algoritmo 11.2: Método do gradiente.

Como o método do gradiente é um caso particular do esquema geral de descida, goza da mesma convergência global descrita pelo Teorema 3.

Teorema 4. *Suponha que f seja continuamente diferenciável. Seja x^* um ponto de acumulação da sequência $\{x^k\}$ gerada pelo método do gradiente (Algoritmo 11.2), desconsiderando o critério de parada. Então*

$$\nabla f(x^*) = 0.$$