Gradient Descent Algorithm

Theory :

Let F be a function defined and differentiable in the neighborhood of a point $a \in \mathbb{R}^n$ and value in $\mathbb{R}$. We note $dom(F)$ its domain.

We have the followed property :

$$F(a - \gamma \nabla F(a)) \leq F(a) \qquad \text{for } \gamma \in \mathbb{R}^{+} \text{ small enough.} \tag{1}$$

<u>Proof</u> :
With a Taylor development at the first order we have :

$$F(a - \gamma \nabla F(a)) = F(a) - \gamma \nabla F(a)^{T} \nabla F(a) + o(\|\gamma \nabla F(a)\|) \tag{2}$$

We want to show :

$$-\gamma . \nabla^2 F(a) + o(\gamma \|\nabla F(a)\|) < 0 \tag{3}$$

We have by definition :

$$\lim_{\gamma \to 0^+} \frac{o(\gamma \|\nabla F(a)\|)}{\gamma \|\nabla F(a)\|} = 0 \tag{4}$$

We can deduce that :

$$\lim_{\gamma \to 0^+} \frac{o(\gamma \|\nabla F(a)\|)}{\gamma \|\nabla^2 F(a)\|} = 0 \tag{5}$$

Which means that $\exists \ M > 0$ such that $\forall \ \gamma < M$:

$$|o(\gamma . \nabla F(a))| < |\gamma \|\nabla^2 F(a)\|| \tag{6}$$

equivalent to

$$|o(\gamma . \nabla F(a))| < \gamma \ \nabla^2 F(a) \tag{6}$$

because $\gamma > 0$ and $\|\nabla^2 F(a)\| = \nabla^2 F(a) > 0$ and so we have :

$$-\gamma \nabla^2 F(a) + |o(\gamma . \nabla F(a))| < 0 \tag{7}$$

Which shows (3) and proves (1).

Inequality (1) being true, to find a local minimum or a global minimum (if F is strictly convex for example) we can build an algorithm performing that way :

Algorithm :

Input : $a_O \in dom(F)$, γ small, ε very small, $n = 0$

While $desc > \varepsilon$ **do** :

$a_{n+1} = a_n - \gamma \nabla F(a_n)$

$desc = |F(a_{n+1}) - F(a_n)|$

$n \mathrel{+}= 1$

done

Algorithm with optimized step :

Input : $a_O \in dom(F)$, ε very small, $n = 0$

While $desc > \varepsilon$ **do** :

$\gamma_n = argmin_\gamma F(a_n - \gamma \nabla F(a_n))$

$a_{n+1} = a_n - \gamma_n \nabla F(a_n)$

$desc = |F(a_{n+1}) - F(a_n)|$

$n \mathrel{+}= 1$

done