Méthodes de gradient

(P) : $\min_{x \in \Omega} f(x) = f(x_*)$ $\quad f : \Omega \to \mathbb{R}$ Différentiable

$\Omega \subset \mathbb{R}^n$ ouvert

Objectif : Construire $(x_k)_{k \in \mathbb{N}} \in \Omega^{\mathbb{N}}$ tq :

(i) $\lim_{k \to +\infty} x_k = x_*$

(ii) Le calcul de x_k à chaque étape ne doit pas être cher.

I] Méthode de descente

Déf :

$d \in \mathbb{R}^n$ est dit direction de descente pour la fonction f au point $x \in \Omega$ s'il existe $\bar{t} > 0$ tq

$$f(x + td) < f(x) \qquad \forall t \in [0, \bar{t}]$$

Algo :

$x_0 \in \Omega$.

À l'itération k :

k $\leftarrow x_k$

d_k : direction de descente

t_k : pas de descente $\left(f(x_k + t_k d_k) < f(x_k)\right)$

$x_{k+1} = x_k + t_k d_k$

$\rightarrow x_{k+1}$

Prop :

$d \in \mathbb{R}^n$, $(\nabla f(x), d) < 0 \Rightarrow d$ direction de descente de f en x

Démo :

$t > 0$

$$f(x+td) - f(x) = (\nabla f(x), td) + \|td\| \varepsilon(td) \quad \varepsilon(td) \xrightarrow[t \to 0]{} 0$$

$$= t\left[\underbrace{(\nabla f(x), d)}_{<0} + \|d\| \varepsilon(td)\right]$$

$$\underset{t \ll 1}{<} 0$$

D'où $f(x+td) < f(x)$

Exemple :

$$\boxed{d = -\nabla f(x)}$$

$$(\nabla f(x), d) = -\|\nabla f(x)\|^2 < 0$$

II - 2) Méthodes de gradient

Prop : $d \in \mathbb{R}^n$ direction de descente de f en x, $\|d\| = 1$

$$g = \frac{-\nabla f(x)}{\|\nabla f(x)\|}$$

$$(\nabla f(x), g) \leq (\nabla f(x), d) < 0$$

Démo :

$$(\nabla f(x), g) = \cdots = -\|\nabla f(x)\|$$

On $|(\nabla f(x), d)| \underset{C-S}{\leq} \|\nabla f(x)\| \times \|d\| = \|\nabla f(x)\|$

Donc $-\|\nabla f(x)\| \geq (\nabla f(x), d)$

D'où $(\nabla f(x), g) < (\nabla f(x), d) < 0$

<u>Algorithme</u>

$k \leftarrow x_k$

$d_k = -\nabla f(x_k)$

Choisir $t_k > 0$

$x_{k+1} = x_k + t_k d_k$

$|x_{k+1} - x_k| < \varepsilon$: <u>STOP</u>

$\longrightarrow x_{k+1}$

<u>Rmq</u> :

① Si le pas de descente est constant

$$t_k = t \quad \underline{\forall k}.$$

↳ On utilise <u>la méthode du gradient à pas constant</u>.

② $t_k > 0$,

$$f(x_k + t_k d_k) \leq f(x_k + t d_k) \qquad \forall t > 0$$

↳ On utilise ici <u>la méthode du gradient à pas optimal</u>.