Méthodes d'optimisation de gradient

1. Le principe des méthodes à direction de descente

Soit le problème d'optimisation P sans contraintes suivant :

$$\min_{X \in R^n} f(X)$$

Avec $f(X)$ est une fonction continument différentiable de n variables. On doit chercher le point X^* qui réalise le minimum de $f(X)$, càd X^* sera la solution de notre problème P.

Prenons l'exemple du bonhomme qui voulait partir d'un point X_0 vers une destination. Il devait faire deux choses essentielles :

a. Prendre **une direction $\boldsymbol{d_k}$** à l'étape k.
b. Déterminer le meilleur point d'arrêt dans cette direction c.à.d. il doit choisir aussi **le pas $\boldsymbol{a_k}$** avec lequel il va avancer vers X_{k+1} dans la direction d_k.

Ces deux opérations a. et b. vont se répéter en générant une suite des point $\{X_k\}_{k\geq 1}$ approchant une solution X^* par la récurrence.

$$X_{k+1} = X_k + a_k . d_k$$

Le choix de a_k et d_k dépend des données du problème. En optimisation, on cherche généralement à minimiser $f(X)$. On doit donc faire décroitre $f(X)$.

a_k et d_k sont alors choisis tel que :

$$f(X_{k+1}) \leq f(X_k)$$

$f(X_{k+1})$ va prendre des valeurs descendante vers le minimum, donc d_k doit être **une direction de descente** pour minimiser $f(X)$, d'où l'idée d'utiliser des méthodes à direction de descente connues pour minimiser des fonctions. Parmi ces méthodes, nous avons la méthode du Newton que nous avons vu en TP et la méthode du gradient que nous allons voir dans ce cours.

2. La méthode du gradient

Soit le problème d'optimisation P sans contraintes suivant :

$$P : \min_{X \in R^n} f(X)$$

A partir de X_k $(k \geq 1)$, on voudrait trouver un voisinage de X_k :

$$X_{k+1} = X_k + a_k.d_k \text{ tel que } f(X_{k+1}) \leq f(X_k)$$

Etape a : Le choix de la direction d_k

La méthode du gradient est une méthode à direction de descente où :

$$d_k = -\nabla f(X_k)$$

Etape b : Le choix du pas a_k

Dans cette étape, on distingue 3 méthodes :

1. **Choisir un a fixe** c.à.d : $a_1 = a_2 = a_3 = \cdots = a_k = a_{k+1}$.
2. **Choisir un a_k variable**, c.à.d. on doit choisir une valeur pour a_k différente d'une itération à une autre (a_k est donné par une expression en fonction de k).
3. **Optimiser a_k à chaque itération k** : Dans la méthode de descente, on calcul X_{k+1} tel que $f(X_{k+1}) \leq f(X_k)$ et cela en agissant sur a_k, c.à.d. on doit trouver le meilleur pas a_k tel que $f(X_{k+1})$ est le plus petit.

   Pour cela, on pose :

$$f_k(a_k) = f(X_k)$$
$$f_k(a_k) = f(X_{k-1} - a_k.\nabla f(X_{k-1}))$$

   Pour trouver a tel que $f_k(a)$ soit minimum, il suffit de résoudre le problème suivant :

$$f_k'(a_k) = 0$$

   →

$$f'\left(X_{k-1} - \boldsymbol{a_k}.\nabla f(X_{k-1})\right) = 0$$

   Donc, nous avons passé d'un problème d'optimisation dans R^n à un problème dans R dont on cherche à trouver a_k, solution de l'équation ci-dessus.

$$a_k = Argmin\, f_k(a_k)$$

   Cette expression donne la valeur du pas a_k à l'itération k pour laquelle f atteint son minimum.

3. Algorithmes de méthode du gradient

a. Méthode du gradient à pas fixe

Méthode a.1	Méthode a.2

• $k = 0$
• Choisir un point initial X_0
• Choisir un pas fixe a
• Tant que $\|\nabla f(X_k)\| > \varepsilon$
    ○ $k = k + 1$
    ○ $X_k = X_{k-1} - a.\nabla f(X_{k-1})$
• Fin tant que
	• Choisir un point initial X_0
	• $k = 1$
	• Choisir un pas fixe a
	• $X_k = X_{k-1} - a.\nabla f(X_{k-1})$
	• Tant que $\|X_k - X_{k-1}\| > \varepsilon$
	    ○ $k = k + 1$
	    ○ $X_k = X_{k-1} - a.\nabla f(X_{k-1})$
	• Fin tant que

Remarque : $\|\nabla f(X_k)\| > \varepsilon$ est une condition d'arrêt (la même chose pour $\|X_k - X_{k-1}\| < \varepsilon$)

- Si $\|\nabla f(X_k)\| < \varepsilon$ → STOP , $X^* = X_k$
- Sinon : On passe à l'itération suivante :
  - $k = k + 1$
  - $X_k = X_{k-1} - a.\nabla f(X_{k-1})$

La différence entre la méthode a.1 et la méthode a.2 réside seulement dans la condition d'arrêt (la même chose pour les méthodes b et c).

b. Méthode du gradient à pas variable

Méthode b.1	Méthode b.2

• $k = 0$
• Choisir un point initial X_0
• Tant que $\|\nabla f(X_k)\| > \varepsilon$
    ○ $k = k + 1$
    ○ $a_k = \frac{1}{3k}$ (par exemple)
    ○ $X_k = X_{k-1} - a_k..\nabla f(X_{k-1})$
• Fin tant que
	• Choisir un point initial X_0
	• $k = 1$
	• $a_k = \frac{1}{k}$ (un autre exemple)
	• $X_k = X_{k-1} - a_k.\nabla f(X_{k-1})$
	• Tant que $\|X_k - X_{k-1}\| > \varepsilon$
	    ○ $k = k + 1$
	    ○ $a_k = \frac{1}{3k}$ (par exemple)
	    ○ $X_k = X_{k-1} - a_k.\nabla f(X_{k-1})$
	• Fin tant que

c. Méthode du gradient à pas optimal

Méthode c.1	Méthode c.2

• $k = 0$
• Choisir un point initial X_0
 	
• **Tant que** $\boldsymbol{\|\nabla f(X_k)\| > \varepsilon}$
  ○ $k = k + 1$
  ○ $\boldsymbol{a_k} = \underset{a_k>0}{\text{argmin}}\, f(X_{k-1} - \boldsymbol{a_k}.\nabla f(X_{k-1}))$
  ○ $X_k = X_{k-1} - a_k.\nabla f(X_{k-1})$
• Fin tant que
	• Choisir un point initial X_0
	• $k = 1$
	• $\boldsymbol{a_k} = argmin\, f(X_{k-1} - \boldsymbol{a_k}.\nabla f(X_{k-1}))$
	• $X_{k+1} = X_k - a_k.\nabla f(X_k)$
	• **Tant que** $\boldsymbol{\|X_k - X_{k-1}\| > \varepsilon}$
	  ○ $k = k + 1$
	  ○ $\boldsymbol{a_k} = \underset{a_k>0}{\text{argmin}}\, f\left(X_{k-1} - \boldsymbol{a_k}.\nabla f(X_{k-1})\right)$
	  ○ $X_k = X_{k-1} - a_k.\nabla f(X_{k-1})$
	• Fin tant que

Exercice 1 : Soit le problème d'optimisation suivant :

$$\min f(x,y) = x^2 + \frac{1}{2}y^2 - 3(x+y)$$

Avec : $X_0 \begin{pmatrix} -2 \\ -7 \end{pmatrix}$ et $\varepsilon = 0.1$

Utiliser la méthode **du gradient à pas optimal** pour résoudre le problème.

On calcule d'abord :

$$\nabla f(x,y) = \begin{pmatrix} 2x-3 \\ y-3 \end{pmatrix}$$

Il est clair que la solution de ce problème analytiquement est : $X^*(\frac{3}{2}, 3)$

$k = 0$

$\nabla f(X_0) = \begin{pmatrix} -7 \\ -10 \end{pmatrix} \rightarrow \|\boldsymbol{\nabla f(X_0)}\| = \mathbf{12.2}$

Est-ce que $\|\boldsymbol{\nabla f(X_0)}\| > \boldsymbol{\varepsilon}$ **?** **Oui → on fait une première itération**

1ère itération

$\mathrm{k} = k+1 = 1$

$$a_k = argmin\, f(X_{k-1} - \boldsymbol{a_k}.\nabla f(X_{k-1})) \rightarrow$$

$$a_1 = argmin\, f[X_0 - \boldsymbol{a_1}.\nabla f(X_0)]$$

Donc, on doit résoudre le problème d'optimisation suivant :

$\min f(a_1) = \min\limits_{a_1>0} f[X_0 - \boldsymbol{a_1}.\nabla f(X_0)] \rightarrow$

$$\min f(a_1) = \min_{a_1>0} f\left[\begin{pmatrix} -2 \\ -7 \end{pmatrix} - \boldsymbol{a_1}.\begin{pmatrix} -7 \\ -10 \end{pmatrix}\right] = \min_{a_1>0} f\begin{pmatrix} -2+7a_1 \\ -7+10a_1 \end{pmatrix}$$

$$f\begin{pmatrix} -2+7a_1 \\ -7+10a_1 \end{pmatrix} = (-2+7a_1)^2 + \frac{1}{2}(-7+10a_1)^2 - 3(-2+7a_1-7+10a_1)$$

$$f\begin{pmatrix} -2+7a_1 \\ -7+10a_1 \end{pmatrix} = 99a_1^2 - 149a_1 + \frac{111}{2}$$

$$\min_{a_1>0} f(a_1) = 99a_1^2 - 149a_1 + \frac{111}{2}$$

Pour trouver résoudre ce problème, il faut résoudre l'équation suivante :

$f'(a_1) = 0 \rightarrow$

$f'(a_1) = 198a_1 - 149 = 0 \rightarrow$

➔ $a_1 = 0.7525$

➔ $X_1 = X_0 - a_1.\nabla f(X_0) = \begin{pmatrix} -2 \\ -7 \end{pmatrix} - 0.7526\ .\begin{pmatrix} -7 \\ -10 \end{pmatrix} \rightarrow X_1 = \begin{pmatrix} 3.2677 \\ 0.5253 \end{pmatrix}$

$\nabla f(X_1) = \begin{pmatrix} 3.5354 \\ -2.4747 \end{pmatrix}$

→ $\|\boldsymbol{\nabla f(X_1)}\| = \boldsymbol{4.3154}$

Est-ce que $\|\boldsymbol{\nabla f(X_1)}\| > \boldsymbol{\varepsilon}$ **?** **Oui → on passe à la deuxième itération**

2ème itération

$k = k + 1 = 2$

$$a_k = argmin\, f(X_{k-1} - \boldsymbol{a_k}.\nabla f(X_{k-1})) \rightarrow$$

$$a_2 = argmin\, f[X_1 - \boldsymbol{a_2}.\nabla f(X_1)]$$

Donc, on doit résoudre le problème d'optimisation suivant :

$$\min f(a_2) = \min_{a_2>0} f[X_1 - \boldsymbol{a_2}.\nabla f(X_1)] \rightarrow$$

$$\min f(a_1) = \min_{a_2>0} f\left[\begin{pmatrix} 3.267 \\ 0.525 \end{pmatrix} - \boldsymbol{a_2}.\begin{pmatrix} 3.535 \\ -2.473 \end{pmatrix}\right] = \min_{a_2>0} f\begin{pmatrix} 3.267 - 3.535.a_2 \\ 0.525 - 2.473.a_2 \end{pmatrix}$$

$$f\begin{pmatrix} 3.26 - 3.535.a_2 \\ 0.525 - 2.473.a_2 \end{pmatrix}$$

$$= (3.267 - 3.535.a_2)^2 + \frac{1}{2}(0.525 - 2.473.a_2)^2 - 3(3.267 - 3.535.a_2$$

$$+ 0.525 - 2.473.a_2)$$

$$f\begin{pmatrix} 3.26 - 3.535.a_2 \\ 0.525 - 2.473.a_2 \end{pmatrix} = (\dots).a_2^2 + (\dots).a_2 + (\dots)$$

$$\min_{a_2>0} f(a_2) = (\dots).a_2^2 + (\dots).a_2 + (\dots)$$

➔ $a_2 = 0.5984$

➔ $X_2 = X_1 - a_2.\nabla f(X_1) = \begin{pmatrix} 3.2677 \\ 0.5253 \end{pmatrix} - 0.5984 \,.\begin{pmatrix} 3.5354 \\ -2.4747 \end{pmatrix} \rightarrow X_2 = \begin{pmatrix} 1.1521 \\ 2.0061 \end{pmatrix}$

$\nabla f(X_2) = \begin{pmatrix} -0.6956 \\ -0.9939 \end{pmatrix}$

→ $\|\boldsymbol{\nabla f(X_2)}\| = \boldsymbol{1.2131}$

Est-ce que $\|\boldsymbol{\nabla f(X_3)}\| > \boldsymbol{\varepsilon}$ **? Oui → on passe à la troisième itération**

$k = k + 1 = 3$

……..

……..

……..

k	a_k	X_k	$\nabla f(X_k)$	$\|\nabla f(X_k)\|$	$\|\nabla f(X_k)\| > \varepsilon$ **???**
0	/	$\begin{pmatrix}-2\\-7\end{pmatrix}$	$\begin{pmatrix}-7\\-10\end{pmatrix}$	12.2065	Oui
1	0.7525	$\begin{pmatrix}3.2677\\0.5253\end{pmatrix}$	$\begin{pmatrix}3.5354\\-2.4747\end{pmatrix}$	4.3154	Oui
2	0.5984	$\begin{pmatrix}1.1521\\2.0061\end{pmatrix}$	$\begin{pmatrix}-0.6956\\-0.9939\end{pmatrix}$	1.2131	Oui
3	0.7526	$\begin{pmatrix}1.6757\\2.7541\end{pmatrix}$	$\begin{pmatrix}0.3515\\-0.2459\end{pmatrix}$	0.4289	Oui
4	0.5982	$\begin{pmatrix}1.4654\\2.9012\end{pmatrix}$	$\begin{pmatrix}-0.069\\-0.0988\end{pmatrix}$	0.1205	Oui
5	0.7535	$\begin{pmatrix}\mathbf{1.5175}\\\mathbf{2.9756}\end{pmatrix}$	$\begin{pmatrix}0.035\\-2.0243\end{pmatrix}$	0.0426	Non

La solution du problème est $X^* = X_5 = \begin{pmatrix}1.5175\\2.9756\end{pmatrix}$