Optimisation TD 7

La méthode de descente du gradient

Si on considère un programme d'optimisation **convexe** noté :

$$(P) = \begin{cases} \min f(x) \\ x \in R^n \end{cases}$$

On suppose que la fonction f est continûment différentiable.

Si nous ne disposons pas de méthode analytique pour optimiser une fonction objectif (parce que le nombre de paramètres est élevé par exemple, ou parce que le calcul serait trop coûteux), nous nous rabattons sur une méthode numérique itérative. L'une d'elles est la méthode de **descente du gradient** qui sert à optimiser des problèmes non-linéaires **convexes** non contraints.

L'idée de base consiste à chercher une suite $\{x_k\}_{x \in N} \in \Re$ dont le successeur de x_k doit satisfaire la condition : $f(x_{k+1}) < f(x_k)$. Afin d'établir cette suite, on exploitera la dérivée directionnelle définie précédemment. Le terme descente vient du fait que cette méthode recherche le minimum suivant une direction opposée à celle du gradient de la fonction objectif.

Algorithme du gradient

1. Initialiser avec x_0 (au hasard)
2. Répéter : $x_{k+1} = x_k - \alpha \nabla f(x_k)$
3. jusqu'à convergence

\- minimisation

\+ maximisation

* au hasard : difficile de suggérer des valeurs "intelligentes".

* ∇ : Le gradient, Indique la direction et l'importance de la pente au voisinage de x_k .

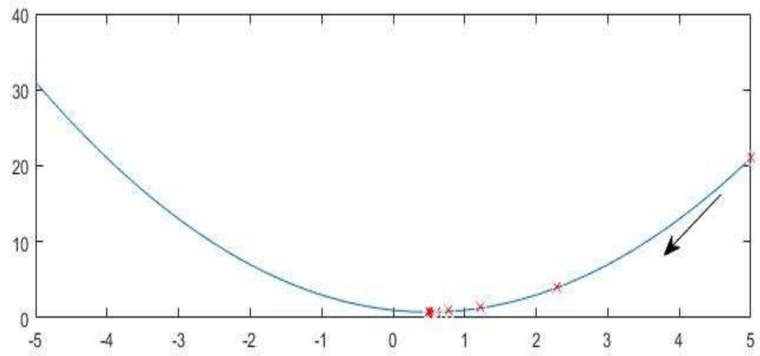
* α : est un paramètre qui permet de moduler la correction (α trop faible, lenteur dans la convergence; α trop grand, oscillations dans la convergence;

* Nbre d'itération fixé, ou $\|x_{k+1} - x_k\|$,ou $\|\nabla f(x_k)\|$ très petit

Exemple 1: fonction à une seule variable $f(x) = x^2 - x + 1$ $avec\ x_0 = 5$ $\alpha = 0.3$

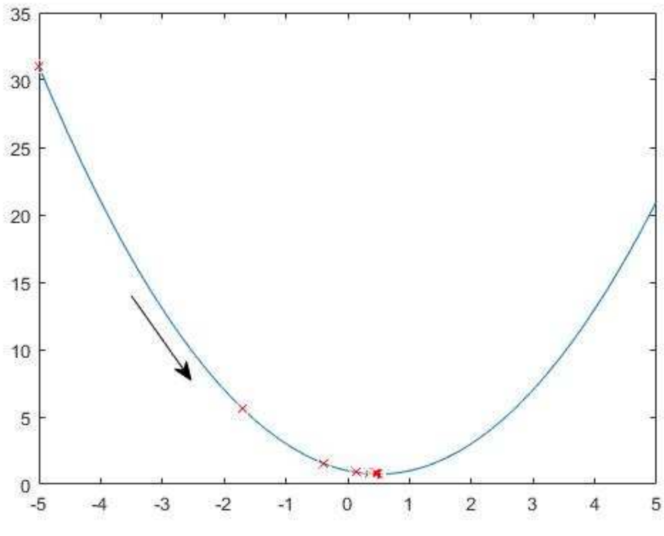
x	df_x	f(x)
5.0000	9.0000	21.0000
2.3000	3.6000	3.9900
1.2200	1.4400	1.2684
0.7880	0.5760	0.8329
0.6152	0.2304	0.7633
0.5461	0.0922	0.7521
0.5184	0.0369	0.7503
0.5074	0.0147	0.7501

0.5029	0.0059	0.7500
0.5012	0.0024	0.7500
0.5005	0.0009	0.7500



on aurait pu partir de l'autre côté

X	df_x	f(x)
-5.0000	-11.0000	31.0000
-1.7000	-4.4000	5.5900
-0.3800	-1.7600	1.5244
0.1480	-0.7040	0.8739
0.3592	-0.2816	0.7698
0.4437	-0.1126	0.7532
0.4775	-0.0451	0.7505
0.4910	-0.0180	0.7501
0.4964	-0.0072	0.7500
0.4986	-0.0029	0.7500
0.4994	-0.0012	0.7500
0.4998	-0.0005	0.7500



Exemple 2: fonction à deux variables : $f(x,y)=\frac{1}{2}x^2+\frac{1}{7}y^2$ *avec* $x_0=(7.5,\ 2.2)$

avec un α optimal

x	y	df_x	df_y	f(x,y)	alpha_optimal
7.5000	2.2000	7.5000	15.4000	45.0650	0.1709
6.2179	-0.4326	6.2179	-3.0282	19.9861	0.4651
3.3262	0.9757	3.3262	6.8298	8.8637	0.1709
2.7576	-0.1919	2.7576	-1.3430	3.9310	0.4651
1.4752	0.4327	1.4752	3.0290	1.7434	0.1709
1.2230	-0.0851	1.2230	-0.5956	0.7732	0.4651
0.6542	0.1919	0.6542	1.3433	0.3429	0.1709
0.5424	-0.0377	0.5424	-0.2641	0.1521	0.4651
0.2901	0.0851	0.2901	0.5958	0.0674	0.1709