ANSD, Dakar
ISE 1

# Optimisation

## Séquence 4 : Méthodes de gradient

Dr. Oumar Diop
oumardiop32@yahoo.fr

16 mai 2025

# Généralités

Pour la méthode du gradient, on choisit comme direction de descente, la direction de la plus grande pente de $f$ au point $x^{(k)}$ . Intuitivement, cela permet d'espérer une réduction significative d'énergie pour $f(x^{(k+1)})$ . La direction $d^{(k)}$ est donc donnée par :

$$d^{(k)} = -\nabla f(x^{(k)}).$$

En effet, en utilisant le développement de Taylor, on obtient :

$$f\left(x^{(k)} - \rho\nabla f(x^{(k)})\right) = f(x^{(k)}) + \langle \nabla f(x^{(k)}), -\rho\nabla f(x^{(k)})\rangle + o(\rho).$$

Ce qui permet d'obtenir :

$$f\left(x^{(k)} - \rho\nabla f(x^{(k)})\right) - f(x^{(k)}) = -\rho\|\nabla f(x^{(k)})\|^2 + o(\rho).$$

Si $\rho > 0$, le membre de droite de l'égalité devient négative.
Les méthodes de gradient diffèrent alors selon l'expression du facteur de descente $\rho_k$.
On distingue plusieurs méthode de type gradient. Une méthode de gradient est définie par la formule suivante :

$$x^{(k+1)} = x^{(k)} - \rho_k \nabla f(x^{(k)}). \tag{1}$$

# Algorithme du gradient à pas fixe



L'algorithme de la méthode du gradient à pas fixe est donné par les trois étapes suivants.

- Etape 1 : Initialisation
  - ☐ Choisir $k = 0$ ;
  - ☐ choisir $x^{(0)} \in \mathbb{R}^n$, par exemple $x^{(0)} = 0$ ;
  - ☐ choisir le nombre maximum d'itération (par exemple $k_{max} = 10^7$) ;
  - ☐ fixer la tolérance $\epsilon = 10^{-7}$ par exemple.
- Etape 2 : Boucle Tant que ("test d'arrêt" est faux) et $(k < k_{max})$ faire :
  - ☐ $x^{(k+1)} = x^{(k)} - \rho\nabla f(x^{(k)}$ ;
  - ☐ Incrémenter le compteur : $k = k + 1$.
- Etape 3 : Convergence
  - ☐ Si ("test darrêt" est vrai ) alors la suite est convergente c'est à dire $x^{(k)}$ est une approximation du minimum recherché ;
  - ☐ Sinon la méthode n'a pas convergé.

Le choix du facteur de descente $\rho$ est déterminant opour assurer la convergence. On donne le résultat suivant.

# Algorithme du gradient à pas fixe



## Théorème 1

*Soit* $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ *une application* $\mathcal{C}^1$ *vérifiant :*

$$\exists \alpha > 0, \langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n.$$

*et*

$$\exists M > 0, \|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|, \quad \forall x, y \in \mathbb{R}^n.$$

*DE plus, si*

$$0 < \rho < \frac{2\alpha}{M^2},$$

*alors la méthode du gradient à pas fixe donnée par la formule*

$$x^{(k+1)} = x^{(k)} - \rho \nabla f(x^{(k)}),$$

*converge vers l'unique minimum* $\bar{x}$ *de* $f$*.*

# Méthode de gradient à pas optimal



Dans cette méthode, on choisit $\rho_k \in \mathbb{R}$ vérifiant :

$$f\left(x^{(k)} - \rho_k \nabla f(x^{(k)})\right) = \min_{\rho \in \mathbb{R}} f\left(x^{(k)} - \rho \nabla f(x^{(k)})\right), \qquad (2)$$

et

$$x^{(k+1)} = x^{(k)} - \rho_k \nabla f(x^{(k)}. \qquad (3)$$

## Théorème 2

*Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction elliptique. c'est à dire :*

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \alpha \|x - y\|^2, \ \ \forall\, x, y \in \mathbb{R}^n.$$

*Alors la méthode de gradient donnée par (2) et (3) est bien définie et elle converge.*

Pour la preuve, revoir le séquence précédent.

# Plan

# Application aux fonctions quadratiques



On considère la fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par :

$$f(x) = \frac{1}{2} < Ax, x > - < b, x > +c, \tag{4}$$

où $A$ est une matrice symétrique définie positive, $b \in \mathbb{R}^n$ et $c \in \mathbb{R}$.
La méthode de gradient à pas optimal consiste à calculer $\rho_k \in \mathbb{R}$ qui minimise la fonction

$$g(\rho) = f\left(x^{(k)} - \rho \nabla f(x^{(k)})\right)$$

Dans ce cas $\rho_k$ satisfait l'équation d'Euler en $g$, c'est dire :

$$g'(\rho_k) = 0.$$

On peut vérifier que :

$$g'(\rho_k) = \langle \nabla f\left(x^{(k)} - \rho_k \nabla f(x^{(k)})\right), \nabla f(x^{(k)}\rangle.$$

On rappelle que pour ce cas particulier $\nabla f(x^{(k)} = Ax^{(k)} - b$, on a alors :

$$\begin{aligned} g'(\rho) &= -\langle A\left(x^{(k)} - \rho_k \nabla f(x^{(k)})\right) - b, A(x^{(k)} - b\rangle \\ &= -\|Ax^{(k)} - b\|^2 + \rho\langle A(A(x^{(k)} - b), A(x^{(k)} - b\rangle. \end{aligned}$$

On obtient

$$\rho_k = \frac{\|Ax^{(k)} - b\|^2}{\langle A(Ax^{(k)} - b), Ax^{(k)} - b\rangle}. \tag{5}$$

Exercice : Vérifier que $\rho_k > 0$.

# Plan

# Fonctions non quadratique



Dans le cas d'une fonction $f$ non quadratique, $\nabla f$ est en général une application non linéaire alors il peut être difficile de trouver $\rho_k$ solution de problème

$$f\left(x^{(k)} - \rho_k \nabla f(x^{(k)})\right) = \min_{\rho \in \mathbb{R}} f\left(x^{(k)} - \rho \nabla f(x^{(k)})\right), \tag{6}$$

Il faudra alors utiliser à chaque pas $k$ une méthode numérique pour approcher $\rho_k$. Ceci va évidement alourdir les calculs. On pourra alors se satisfaire d'une estimation de $\rho_k$. On donne le théorème suivant :

## Théorème 3

*Soit $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ une fonction elliptique de constante d'ellipticité $\alpha$ et vérifiant $\nabla f$ une application lipschitzienne de constante de lipschitz $M$*
*Supposons de plus que la suite $\{\rho_k\}_{k\in\mathbb{N}} \subset \mathbb{R}$ satisfait à la proposition suivant : il existe $a_1, a_2$ vérifiant : $0 < a_1 \leq a_2 < \frac{2\alpha}{M^2}$ tels que : $a_1 \leq \rho_k \leq a_2, \quad \forall\, k \in \mathbb{N}$. Alors la méthode générale de gradient*

$$x^{(k+1)} = x^{(k)} - \rho_k \nabla f(x^{(k)}.$$

*converge et la convergence est au moins géométrique, c'est à dire : $\exists \beta \in [0, 1]$ tels que :*

$$\|x^{(k)} - x^*\| \leq \beta^k \|x^0 - x^*\|, \quad \forall\, k \in \mathbb{N}.$$

*où $x^*$ est l'unique minimum de $f$ sur $\mathbb{R}^n$.*

Exercice 2

Soit $C \in \mathcal{M}_{p,n}(\mathbb{R}), b\mathbb{R}^p$. On définit la fonction $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ par :

$$f(x) = \frac{1}{2}\|Cx - b\|^2.$$

1. Montrer que $f$ est une fonction quadratique et déterminer la matrice $A$, le vecteur $b$ et le réel $c$ associés à $f$.
2. Vériifer que pour toute matrice carrée $A$, la matrice $AA^t$ est symétrique
3. Donner l'expression du gradient $\nabla f(x)$ et de la matrice hessienne $\nabla^2 f(x)$ en tout point $x \in \mathbb{R}^n$. (on pourra se référer à un résultat du cours plutot que de faire les calculs).
4. Montrer que $f$ est convexe.
5. Montrer que $f$ est $\alpha-$ elliptique si et seulement si $ker(A) = \{0\}$.
6. Soit $x \in \mathbb{R}^n$ tel que $\nabla f(x) \neq 0$. Rappeler la définition d'une direction de descente $d \in \mathbb{R}^n$ au point $x$ et montrer que $d \in \mathbb{R}^n$ est une direction de descente si et seulement si

$$\langle Ax - b, Ad \rangle < 0.$$

(En particulier $Ad \neq 0$.

# TP : Comparaison de méthodes



L'objectif de ce tp est de comparer différentes méthodes d'optimisation sur la fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par :

$$f(x, y) = (x - 1)2 + 10(x^2 - y)^2$$

1. Calculer $\nabla f(x, y)$ puis $\nabla^2 f(x, y)$ ;
2. Cette fonction est-elle convexe ? En quels points est-elle minimale ?
3. Visualiser la fonction en 3D (fonctions plot3d ou plot3d1), puis en 2D par niveaux de gris (fonction grayplot). Visualiser également ses lignes de niveau (fonction contour2d).
4. Qu'observez-vous ? Pourquoi $f$ n'est-elle pas simple à minimiser ?
5. Implémenter la méthode de relaxation et l'appliquer à la fonction $f$ en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement.
6. Implémenter la méthode du gradient à pas optimal et l'appliquer à la fonction $f$ en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement.
7. Implémenter la méthode de Newton pour trouver les zéros de $\nabla f$ en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement.
8. Implémenter la méthode du gradient conjugué et l'appliquer à la fonction $f$ en partant du point $(-1, 1)$. Afficher la trajectoire des points calculés successivement.