I Quelques définitions

I.1 Fonction coercive

Définition 1.1. Si f est continue sur X, on dit que est coercive sur un compact X [a] si

$$\lim_{\|x\|\to\infty} f(x) = +\infty$$

a. Les compacts de $\mathbb{R}^n$ sont les segments

Propriété 1.1. Si f coercive sur X un compact alors f admet un minimum

I.2 Formule de Taylor, gradient, hessienne

Définition 1.2. *Développement de Taylor à l'ordre 2*
Si f est de classe $\mathcal{C}^2$, de $D \subseteq \mathbb{R}^n \to \mathbb{R}$ alors on a

$$f(x+h) = f(x) + \nabla f(x) \cdot h + \frac{1}{2} h \cdot H_f(x) h$$

Définition 1.3. *Gradient*
Si f est de classe $\mathcal{C}^1$, de $D \subseteq \mathbb{R}^n \to \mathbb{R}$ alors on a le gradient de f en x, noté ∇f comme étant définie ci-dessous

$$\nabla f(\cdot) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\cdot) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\cdot) \end{pmatrix}$$

Définition 1.4. *Hessienne et hessien*
Si f est de classe $\mathcal{C}^2$, de $D \subseteq \mathbb{R}^n \to \mathbb{R}$ alors on a la matrice hessienne de f en x, noté $H_f(x)$ comme étant définie ci-dessous

$$H_f(\cdot) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\cdot) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\cdot) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\cdot) \end{pmatrix}$$

Le hessien de f en x est le déterminant de la matrice hessienne

I.3 Défini positif

Définition 1.5. Soit $A \in \mathcal{M}_n(\mathbb{R})$, A est semi-définie positive (resp. négative) si

$$\forall x \in \mathbb{R}^n \; x \cdot Ax \geqslant 0 \,(\text{resp. } \leqslant 0)$$

Définition 1.6. Soit $A \in \mathcal{M}_n(\mathbb{R})$, A est définie positive (resp. négative) si

$$\forall x \in \mathbb{R}^n \; x \cdot Ax > 0 \,(\text{resp. } < 0)$$

Comment montrer qu'une matrice est définie positive ?

Propriété 1.2. *avec des valeurs propres*
- A est semi-définie positive (resp. négative) ssi toutes ses valeurs propres sont positives (resp. négatives)
- A est définie positive (resp. négative) ssi toutes ses valeurs propres sont strict. positives (resp. négatives)

Propriété 1.3. *avec mineurs principaux*
- A est semi-définie positive ssi tout ses mineurs principaux (excepté le déterminant) sont strict. positif, et ayant un determinant nul
- A est définie positive ssi tout ses mineurs principaux sont strict. positif

I.4 Fonction convexe

Définition 1.7. $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$ avec C convexe est dites convexe si

$$\forall (x,y) \in \mathbb{C}^2 \, \forall t \in [0,1] \, f\left((1-t)x + ty\right) \leqslant (1-t)f(x) + tf(y)$$

f est dites strict. convexe si l'inégalité ci-dessus est stricte.

Propriété 1.4. *comment déterminer qu'une fonction est convexe en général*
f est convexe ssi $H_f(x)$ est semi-défini positif pour tout x
f est strict. convexe ssi $H_f(x)$ est défini positif pour tout x

Propriété 1.5. Si f est convexe et s'il existe un minimum local alors celui-ci est l'unique minimum global de f

Propriété 1.6. *Inégalité AG*
Soit $X \in \mathbb{R}^{\ltimes}$ et $(\lambda_i)_{i\in\mathbb{N}} \in \left(\mathbb{R}_+^*\right)^n$ tel que $\sum_{i=1}^n \lambda_i = 1$

$$\sum_{i=1}^n \lambda_i x_i \geqslant \prod_{i=1}^n x_i^{\lambda_i}$$

I.5 Cas quadratique

Définition 1.8. f est dites de la forme quadratique si, avec A une matrice symétrique on ait

$$f(x) = \frac{1}{2}x \cdot Ax + b \cdot x + c$$

Propriété 1.7. Si f est de la forme quadratique alors :
- $\nabla f(x) = Ax + b$
- $H_f(x) = A$

II Extremum

Définition 2.9. Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ et $x^* \in D$, on dit que f admet en x^* un *minimum global* sur D si

$$\forall x \in D \; f(x^*) \leqslant f(x)$$

Si l'inégalité est stricte alors on dit que le minimum est strict.

Définition 2.10. Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ et $x^* \in D$, on dit que f admet en x^* un *minimum local* sur D si il existe un rayon ε tel que

$$\forall x \in D \cap \mathcal{B}(x^*, \varepsilon) \; f(x^*) \leqslant f(x)$$

Si l'inégalité est stricte alors on dit que le minimum est strict.

Définition 2.11. Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ et $x^* \in D$, on dit que f admet en x^* un *maximum global* sur D si

$$\forall x \in D \; f(x^*) \geqslant f(x)$$

Si l'inégalité est stricte alors on dit que le minimum est strict.

Définition 2.12. Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ et $x^* \in D$, on dit que f admet en x^* un *maximum local* sur D si il existe un rayon ε tel que

$$\forall x \in D \cap \mathcal{B}(x^*, \varepsilon) \; f(x^*) \geqslant f(x)$$

Si l'inégalité est stricte alors on dit que le minimum est strict.

Conditions nécessaires d'optimalité

Propriété 2.8. *Condition de premier ordre*
Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^1$ sur D ouvert, si x^* est un minimum local de f alors

$$\nabla f(x^*) = 0$$

Propriété 2.9. *Condition du deuxième ordre*
Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$ sur D ouvert, si x^* est un minimum local de f alors

$$H_f(x^*) \text{ définie positive}$$

Conditions suffisantes d'optimalité

Propriété 2.10. *optimalité locale*
Soit $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ de classe $\mathcal{C}^2$ sur D ouvert,
Si x^* est un point critique ($\nabla f(x^*) = 0$) et si $H_f(x^*)$ est définie positive (resp. semi définie positive) alors x^* est un minimum local strict (resp. minimum local).

Propriété 2.11. *optimalité globale*
Si x^* est un point critique ($\nabla f(x^*) = 0$) et si $\forall z \in D \; H_f(z)$ est définie positive (resp. semi définie positive) alors x^* est un minimum local strict (resp. minimum local).

III Méthodes itératives

III.1 Méthode de Cauchy

Définition 3.13. Soit f numérique définie sur un voisinage V d'un point a, soit u un vecteur de norme 1. On appelle la dérivée de f au point a dans la direction u
$f'(a;u) = \lim_{t\to 0} \frac{f(a+tu)-f(a)}{t}$
Si f est différentiable en a alors $f'(a;u) = \nabla f(x) \cdot u$

Algorithme

Initialisation :
On choisit un point de départ $x^{(0)}$ et on construit la liste de points :

Caractère héréditaire :
$x^{(k+1)} = x^{(k)} - t^{(k)}\nabla f(x^{(k)})$ avec $t^{(k)}$ minimise $t \mapsto x^{(k)} - t\nabla f(x^{(k)}$ avec $t > 0$
On arrête la suite dès que $\nabla f(x^{(k)} = 0$

III.2 Méthode des directions conjuguées

Définition 3.14. Soit A une matrice symétrique, deux vecteurs non nul $d^{(0)}$ et $d^{(1)}$ sont conjugués par rapport à A ssi

$$d^{(0)} \cdot Ad^{(1)} = 0$$

Algorithme

Initialisation :
On choisit un point de départ $x^{(0)}$ et une direction initiale $d^{(0)}$. On construit la liste de points :

Caractère héréditaire :
On note $A = H_f(x^k)$
$x^{(k+1)} = x^{(k)} + \lambda^{(k)}d^{(k)}$
$\lambda^{(k)} = \frac{-\nabla q(x^{(k)})\cdot d^{(k)}}{d^{(k)}\cdot Ad^{(k)}}$
$d^{(k+1)}$ conjugué à $d^{(k)}$ par rapport à A.

III.3 Méthode du gradient conjugué

Algorithme

Initialisation :
On choisit un point de départ $x^{(0)}$ et une direction initiale $d^{(0)} = -\nabla f(x^{(0)})$. On construit la liste de points :

Caractère héréditaire :
On note $A = H_f(x^k)$
$\lambda^{(k)} = \frac{-\nabla f(x^{(k)})\cdot d^{(k)}}{d^{(k)}\cdot Ad^{(k)}}$
$x^{(k+1)} = x^{(k)} + \lambda^{(k)}d^{(k)}$
$\beta^{(k)} = \frac{\nabla f(x^{(k+1)})\cdot Ad^{(k)}}{d^{(k)}\cdot Ad^{(k)}}$
$d^{(k+1)} = -\nabla f(x^{(k+1)}) + \beta^{(k)}d^{(k)}$

III.4 Méthode de descente linéaire

Algorithme

Initialisation :
On choisit un point de départ $x^{(0)}$ et une direction initiale $d^{(0)} = -\nabla f(x^{(0)})$. On construit la liste de points :

Caractère héréditaire :
$\lambda^{(k)}$ minimise $t \mapsto x^{(k)} + td^{(k)}$
$x^{(k+1)} = x^{(k)} + \lambda^{(k)}d^{(k)}$
$\beta^{(k)} = \frac{\left\|\nabla f(x^{(k+1)})\right\|^2}{\left\|\nabla f(x^{(k)})\right\|^2}$
$d^{(k+1)} = -\nabla f(x^{(k+1)}) + \beta^{(k)}d^{(k)}$

III.5 Méthode de Newton

Algorithme

Initialisation :

On choisit un point de départ $x^{(0)}$ et une direction initiale $d^{(0)} = -\nabla f(x^{(0)})$. On construit la liste de points :

Caractère héréditaire :

$x^{(k+1)} = x^{(k)} - \left(H_f(x^{(k)}\right)^{-1} \nabla f(x^{(k)})$

III.6 Méthode de Newton améliorée

Algorithme

Initialisation :

On choisit un point de départ $x^{(0)}$ et une direction initiale $d^{(0)} = -\nabla f(x^{(0)})$. On construit la liste de points :

Caractère héréditaire :

$x^{(k+1)} = x^{(k)} - \lambda^{(k)} \left(H_f(x^{(k)}\right)^{-1} \nabla f(x^{(k)})$ où avec $\lambda^{(k)}$ minimise f dans la direction $-\left(H_f(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$

III.7 Méthode de quasi-Newtoniennes

Définition 3.15. Soit a et b deux vecteur, on définit le produit tensoriel de a et b par :

$$a \otimes b = ab^t \in \mathcal{M}_n(\mathbb{R})$$

Algorithme - correction de rang 1

Initialisation :

$x^{(0)}$ et $S^{(0)} = I$

Caractère héréditaire :

$x^{(k+1)} = x^{(k)} - \lambda^{(k)} S^{(k)} \left(\nabla f(x^{(k)})\right)$
$\lambda^{(k)}$ minimise f dans la direction $-S^{(k)} \nabla f(x^{(k)})$
$d^{(k)} = x^{(k+1)} - x^{(k)}$ et $y^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$
$B^{(k)} = d^{(k)} - S^{(k)} y^{(k)}$
$S^{(k+1)} = S^{(k)} + \frac{B^{(k)} \otimes B^{(k)}}{y^{(k)} \cdot B^{(k)}}$

Algorithme - correction de rang 2 - DFP

Initialisation :

$x^{(0)}$ et $S^{(0)} = I$

Caractère héréditaire :

$x^{(k+1)} = x^{(k)} + \lambda^{(k)} S^{(k)} \left(\nabla f(x^{(k)})\right)$
$\lambda^{(k)}$ minimise f dans la direction $-S^{(k)} \nabla f(x^{(k)})$
$d^{(k)} = x^{(k+1)} - x^{(k)}$ et $y^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$
$B^{(k)} = S^{(k)} y^{(k)}$
$S^{(k+1)} = S^{(k)} + \frac{d^{(k)} \otimes d^{(k)}}{y^{(k)} \cdot d^{(k)}} - \frac{B^{(k)} \otimes B^{(k)}}{y^{(k)} \cdot B^{(k)}}$

Algorithme - correction de rang 2 - BFGS

Initialisation :

$x^{(0)}$ et $S^{(0)} = I$

Caractère héréditaire :

$x^{(k+1)} = x^{(k)} + \lambda^{(k)} S^{(k)} \left(\nabla f(x^{(k)})\right)$
$\lambda^{(k)}$ minimise f dans la direction $-S^{(k)} \nabla f(x^{(k)})$
$d^{(k)} = x^{(k+1)} - x^{(k)}$ et $y^{(k)} = \nabla f(x^{(k+1)}) - \nabla f(x^{(k)})$
$B^{(k)} = S^{(k)} y^{(k)}$
$S^{(k+1)} = \left(I - \frac{sy^t}{y^t s}\right) S^{(k)} \left(I - \frac{ys^t}{y^t s}\right) + \frac{ss^t}{y^t s}$