Analyse II

Fiche de TD
Licence 2 Informatique

Guillaume Metzler
Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France
guillaume.metzler@univ-lyon2.fr

Résumé

Les exercices proposés dans cette fiche constituent une bonne base d'entraînement pour mettre en application les différentes notions vues en cours. Les exercices sont essentiellement triés par thème mais il n'est pas impossible qu'il faille avoir recours à des notions vues ultérieurement afin de pouvoir le traiter.

L'ordre des exercices ne présage pas de leur difficulté, ces derniers sont essentiellement rangés par thématique et la difficulté des exercices peut fortement varier. En revanche, tous les éléments nécessaires à la résolution des exercices figurent dans le cours. Il faudra simplement mener une réflexion plus ou moins profonde.

En cas de problème dans la résolution de ces exercices, vous pouvez toujours me solliciter par mail.

Tous les exercices ne pourront pas être traités en TD, il est donc important que vous vous entraîniez chez vous pour maîtriser ces notions et que vous refassiez les exercices traités en cours.

Table des matières

1 Continuité et Dérivabilité de fonctions à plusieurs variables

Exercice 1.1 (Définition et Dérivation).
Pour chacune des fonctions suivantes déterminer le domaine de définition ainsi que leur dérivée première, i.e. le gradient de ces fonctions où $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$

1. $f(\mathbf{x}) = x_1^2\sqrt{x_2}$
2. $f(\mathbf{x}) = \dfrac{x_1x_2}{x_1^2 + x_2^2}$
3. $f(\mathbf{x}) = \exp(x_1^2 + x_2^2)$
4. $f(\mathbf{x}) = \ln\left(x_1 + \sqrt{x_1^2 + x_2^2}\right)$

Exercice 1.2 (Dérivées premières et secondes).
Calculer les dérivées premières et secondes des fonctions f *définies sur* $\mathbb{R}^2$ *suivantes*

1. $f(\mathbf{x}) = 4x_1^2 + \exp(x_1x_2)$
2. $f(\mathbf{x}) = 7x_1x_2 + \cos(x_1) + x_1^2 + 4x_2^2$
3. $f(\mathbf{x}) = 4(x_1 - x_2)^2 + 5(x_1^2 - x_2)^2$
4. $f(\mathbf{x}) = \exp(x_1^2 + x_2^2)$

Exercice 1.3 (Matrice Hessienne).

Déterminer les matrices hessiennes associées aux applications suivantes

1. $f(x, y) = 4x^2 + 6y^2 + 3xy + 2\,(\cos(x) + \sin(x))$,
2. $f(x, y) = \ln(x + y) + x^2 + 2y + 4$,
3. $f(x, y, z) = \dfrac{6x}{1 + y} + \exp(xy) + Z$
4. *Pour tout vecteur* $\mathbf{x}, \mathbf{b} \in \mathbb{R}^n$, $f(\mathbf{x}) = \ln\left(\sum_{i=1}^n e^{\mathbf{x}_i + \mathbf{b}_i}\right)$

Exercice 1.4 (Un contre exemple à Scwharz).
On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$f(\mathbf{x}) = \begin{cases} \dfrac{x_1x_2(x_1^2 - x_2^2)}{x_1^2 + x_2^2} & \text{si } (x_1, x_2) \neq (0, 0), \\ 0 \text{ sinon.} \end{cases}$$

1. *Calculer les dérivées partielles :* $\dfrac{\partial f}{\partial x_1}(0, x_2)$ *et* $\dfrac{\partial f}{\partial x_2}(x_1, 0)$.
2. *Calculer* $\dfrac{\partial f}{\partial x_1}\left(\dfrac{\partial f}{\partial x_2}\right)(0, 0)$ *et* $\dfrac{\partial f}{\partial x_2}\left(\dfrac{\partial f}{\partial x_1}\right)(0, 0)$.
3. *Est-ce que la fonction* f *est de classe* C^2 *sur* $\mathbb{R}^2$ *?*

Exercice 1.5 (Un calcul de gradient).
Soit f *la fonction définie pour tout vecteur* $\mathbf{x} \in \mathbb{R}^n$ *par*

$$f(\mathbf{x}) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j,$$

où $a_{ij} = a_{ji}$ *pour tout* $1 \leq i, j \leq n$.

1. *Pour tout* $i \in [\![1, n]\!]$, *calculer les dérivées partielles* $\dfrac{\partial f}{\partial x_i}(\mathbf{x})$.

2. *En déduire une écriture matricielle du gradient de l'application f qui fait intervenir la matrice* $\mathbf{A} = (a_{ij})_{i=1}^{n}$.

Exercice 1.6 (Approximation)**.**
Précisez si les fonctions suivantes sont de classe C^1 sur leur domaine de définition. On calculera ensuite la matrice hessienne associée avant d'indiquer si les fonctions sont de classe C^2 sur leur domaine de définition. On donnera finalement une approximation à l'ordre 2 de ces fonctions au voisinage du point a considéré.

1. $f(\mathbf{x}) = x_1^2\sqrt{x_2}$ *et* $\mathbf{a} = (1, 1)$.
2. $f(\mathbf{x}) = \exp(x_1^2 + x_2^2 + x_3^2)$ *et* $\mathbf{a} = (0, 0, 0)$.

2 Convexité et optimalité

Exercice 2.1 (Etude de la convexité I).
Expliquer pourquoi les fonctions suivantes sont convexes (avec ou sans calculs) :

1. $f(\mathbf{x}) = \sum_{i=1}^{n} x_i^2$.
2. $f(\mathbf{x}) = 3x_1^2 + (x_2 - 3)^2 + 4x_1 + 6x_2 + 5$.
3. $f(\mathbf{x}) = x_1^4 + 6x_2^2 + 2x_2^2 + 9x_1^2 + 3$.
4. $f(\mathbf{x}) = 6x_1^2 + 5x_2^2 + 6x_1x_2$.
5. $f(\mathbf{x}) = \exp(x_1x_2)$ *sur l'intervalle* $]1, +\infty[\times] - \infty, -1[$.

Exercice 2.2 (Etude de la convexité II).

Les fonctions suivantes sont-elles convexes ? Pour les deux premières fonctions, on cherchera également les optima locaux et/ou globaux. Faire de même pour la dernière fonction.

1. $f(\mathbf{x}) = (1 - x_1)^2 + 100(x_2 - x_1^2)^2$.
2. $f(\mathbf{x}) = (x_1 + 2x_2 - 7)^2 + (2x_1 + x_2 - 5)^2$.
3. $f(\mathbf{x}) = 2x_1^2 - 1.05x_1^2 + x_1x_2 + x_2^2$.
4. $f(\mathbf{x}) = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$.
5. $f(\mathbf{x}) = 10 + (x_1^2 - \cos(2\pi x_1)) + (x_2^2 - \cos(2\pi x_2))$.

Exercice 2.3 (Une application convexe).
Soit $\mathbf{z}_1, \dots, \mathbf{z}_n, n$ *vecteurs de* $\mathbb{R}^d$. *Notons* $\mathbf{G}$ *la matrice des produits scalaires des vecteurs* $\mathbf{z}_i, i = 1, \dots, n$, *i.e.*

$$\mathbf{G} = \begin{pmatrix} \langle \mathbf{z}_1, \mathbf{z}_1 \rangle & \langle \mathbf{z}_1, \mathbf{z}_2 \rangle & \dots & \langle \mathbf{z}_1, \mathbf{z}_n \rangle \\ \langle \mathbf{z}_2, \mathbf{z}_1 \rangle & \langle \mathbf{z}_2, \mathbf{z}_2 \rangle & \dots & \langle \mathbf{z}_2, \mathbf{z}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \mathbf{z}_n, \mathbf{z}_1 \rangle & \langle \mathbf{z}_n, \mathbf{z}_2 \rangle & \dots & \langle \mathbf{z}_n, \mathbf{z}_n \rangle \end{pmatrix}$$

Cette matrice est aussi appelée matrice de Gram. On considère la fonction $f\mathbb{R}^n \to \mathbb{R}^n$ *définie par*

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{G} \mathbf{x}.$$

Montrer que la fonction f *est convexe.*

Exercice 2.4 (Condition de convexité).
On considère la fonction f *définie de* $\mathbb{R}^2$ *dans* $\mathbb{R}$ *par*

$$f(\mathbf{x}) = q \ln(x_1) + \sin(px_2).$$

Déterminer les valeurs de p *et* q *pour lesquelles* f *est convexe.*

Exercice 2.5 (Une solution au problème de régression).

On considère une matrice $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ *où* $\mathbf{1}$ *et* $\mathbf{x}_i \in \mathbb{R}^d$ *et* $\mathbf{y} = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ *une matrice dite "de données" et respectivement un vecteur contenant des étiquettes.*

Notre objectif est de trouver une droite de régression, c'est-à-dire, une fonction f *d'un paramètre* $\boldsymbol{\beta} \in \mathbb{R}^d$ *de la forme* $f_{\boldsymbol{\beta}}(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{x}$ *telle que pour tout* $\mathbf{x}_i$, *on ait* $f_{\boldsymbol{\beta}}(\mathbf{x}_i) \simeq y_i$, *i.e., on cherche une droite*

qui approxime le mieux notre nuage de points $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$.
Cela se fait en minimisant la fonction suivante

$$\ell(\boldsymbol{\beta}) = \|\mathbf{y} - f_{\boldsymbol{\beta}}(\mathbf{X})\|_2^2 = \|\mathbf{y} - \mathbf{X}^T\boldsymbol{\beta}\|_2^2 \tag{1}$$

Cas de la dimension 1. *On commence par traiter le cas où les données sont décrites dans un plan en deux dimensions défini par les coordonnées* x_i *et* y_i.

1. *Rappeler l'équation d'une droite dans un plan.*
2. *En vous aidant de la description précédente, montrer, qu'en dimension 1, le problème que l'on cherche à résoudre est de la forme*

$$\ell(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(y_i(-\beta_1 + \beta_2 x_i)\right)^2 .$$

Cas général. *On suppose maintenant que l'on est en dimension quelconque mais finie.*

1. *Etudier la convexité de la fonction* ℓ *en* $\boldsymbol{\beta}$ *définie par l'équation* (1).
2. *Déterminer la solution du problème d'optimisation*

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \ell(\boldsymbol{\beta}).$$

On pourra commencer pour tout vecteur $\mathbf{a}, \mathbf{x} \in \mathbb{R}^n$ *nous avons*

$$\frac{\partial \mathbf{x}^T\mathbf{a}}{\partial \mathbf{x}} = \frac{\partial \mathbf{a}^T\mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}.$$

Ainsi que le résultat d'un exercice qui précède.

Exercice 2.6 (Inégalité Arithmético-géométrique)**.**

A l'aide de la concavité de la fonction ln*, montrer que pour tout vecteur* $\mathbf{x}$ *de* $\mathbb{R}^n$*, nous avons*

$$\sqrt[n]{x_1 x_2 x_3 \ldots x_n} \leq \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}.$$

En déduire que pour tout vecteur $\mathbf{x}$ *de* $\mathbb{R}^n$*, nous avons*

$$\frac{n}{\dfrac{1}{x_1} + \ldots + \dfrac{1}{x_n}} \leq \sqrt[n]{x_1 x_2 x_3 \ldots x_n}.$$

Exercice 2.7 (Inégalité de Young)**.**

L'objectif de cet exercice est de prouver l'inégalité de Young (voir plus bas). On considère ici $\mathbf{x} \in \mathbb{R}^d$. *Soient* $0 < p, q < 1$ *tels que*

$$\frac{1}{p} + \frac{1}{q} = 1.$$

1. *Monter que pour tout* $x, y \in \mathbb{R}$ *on a*

$$\ln(xy) = \frac{\ln(x^p)}{p} + \frac{\ln(x^q)}{q}.$$

2. *En utilisant le fait que la fonction* exp *est convexe, montrer que l'on a, pour tout* $x, y \in \mathbb{R}$ *:*

$$|xy| \leq \frac{|x|^p}{p} + \frac{|x|^q}{q}.$$

Exercice 2.8 (Etude d'une fonction)**.**

Soit f *la fonction définie de* $[4,8] \times [-1,1]$ *dans* $\mathbb{R}$ *par*

$$f(\mathbf{x}) = 2\cos\left(\frac{x_1}{2}\right) + x_2^2.$$

1. *Déterminer le gradient de la fonction* f*.*
2. *En quelle(s) valeur(s) de* f *le gradient s'annule ? Ces points sont les points critiques de* f*, notées* $\mathbf{x}^\star$
3. *Déterminer la matrice hessienne et dire si la fonction* f *est convexe.*
4. *En déduire la nature des points* $\mathbf{x}^\star$ *et une approximation à l'ordre* 2 *de* f *autour de ces points.*

Exercice 2.9 (Une autre étude composée)**.**

Soit f *la fonction définie pour tout vecteur* $\mathbf{x} = (x, y)$ *de* $\mathbb{R}^2$ *par :*

$$f(x, y) = 2x^2 + 2y^2 + 2xy - x - y.$$

1. *On commence par étudier l'existence des points critiques*
   (a) *Calculer les dérivées partielles de* f*.*
   (b) *En déduire que le seul point critique de la fonction* f *est le point de coordonnées* $\left(\frac{1}{6}, \frac{1}{6}\right)$.
   (c) *Calculer l'image du point critique par la fonction* f*.*
2. *On cherche maintenant à étudier la nature de l'extremum.*
   (a) *Développer l'expression*

$$2\left(x + \frac{y}{2} - \frac{1}{4}\right)^2 + \frac{3}{2}\left(y - \frac{1}{6}\right)^2$$

   (b) *En déduire que* f *admet un minimul global sur* $\mathbb{R}^2$*.*
3. *On considère maintenant la fonction* g *définie de* $\mathbb{R}^2$ *dans* $\mathbb{R}$ *par*

$$g(x, y) = 2e^{2x} + 2e^{2y} + 2e^{x+y} - e^x - e^y.$$

   (a) *Utiliser la question 2 pour montrer que pour tout* $(x, y) \in \mathbb{R}^2$, $g(x, y) \geq \frac{-1}{6}$.
   (b) *En déduire que* g *possède un minimum global sur* $\mathbb{R}^2$ *et préciser en quel point ce minimum est atteint.*

3 Optimisation

Exercice 3.1 (La fonction de Matyas).
On considère la fonction f définie de $[-10, 10]^2$ dans $\mathbb{R}$ par

$$f(\mathbf{x}) = 0.26(x_1^2 + x_2^2) - 0.48x_1x_2$$

1. *La fonction f est-elle convexe ?*
2. *Déterminer les valeurs de $\mathbf{x}$ pour lesquelles le gradient de f s'annule.*
3. *Quel est le minimum global de cette fonction ?*
4. *On pose $\mathbf{x}^{(0)} = (x_1, x_2)^{(0)} = (1, 1)$ comme étant le point initial de notre descente de gradient à pas optimal. Calculer la valeur de $\mathbf{x}^{(1)}$.*

Exercice 3.2 (La fonction de Rosenbrock).
On considère la fonction f définie de $\mathbb{R}^2$ dans $\mathbb{R}$ par

$$f(\mathbf{x}) = (1 - x_1)^2 + 10(x_2 - x_1^2)^2$$

1. *La fonction f est-elle convexe ?*
2. *Déterminer les valeurs de $\mathbf{x}$ pour lesquelles le gradient de f s'annule.*
3. *Quel est le minimum global de cette fonction ?*
4. *On pose $\mathbf{x}^{(0)} = (x_1, x_2)^{(0)} = (2, 2)$ comme étant le point initial de notre descente de gradient à pas constant avec $\rho = 0.5$. Calculer les valeur de $\mathbf{x}^{(1)}$ et $\mathbf{x}^{(2)}$.*

Exercice 3.3 (La fonction de Rastrigin).
On considère la fonction f définie de $[-\pi, \pi]^2$ dans $\mathbb{R}$ par

$$f(\mathbf{x}) = 20 + (x_1^2 - 10\cos(2\pi x_1))^2 + (x_2 - 10\cos(2\pi x_2))^2.$$

1. *La fonction f est-elle convexe ?*
2. *Déterminer les valeurs de $\mathbf{x}$ pour lesquelles le gradient de f s'annule.*
3. *Quel est le minimum global de cette fonction ?*
4. *On pose $\mathbf{x}^{(0)} = (x_1, x_2)^{(0)} = (1, 1)$ comme étant le point initial de notre algorithme de descente de Newton. Calculer les valeur de $\mathbf{x}^{(1)}$ et $\mathbf{x}^{(2)}$.*

Exercice 3.4 (Une dernière fonction).
On considère la fonction f définie de $\mathbb{R}^2$ dans $\mathbb{R}$ par

$$f(\mathbf{x}) = 2x_1^2 - 1.05x_1^4 + \frac{x_1^6}{6} + x_1x_2 + x_2^2.$$

1. *Calculer la matrice hessienne.*
2. *Quelles sont les quantités que l'on peut calculer afin de montrer que la matrice hessienne est bien SDP ? Déterminer leurs valeurs.*
3. *On admettra que la fonction f est positive ou nulle et qu'elle n'est pas convexe. Montrer que l'on $\nabla f(\mathbf{0}) = 0$.*
4. *Quel est le minimum global de cette fonction ?*
5. *On pose $\mathbf{x}^{(0)} = (x_1, x_2)^{(0)} = (1, 1)$ comme étant le point initial de notre algorithme de descente deà pas constat avec $\rho = 0.5$. Calculer les valeur de $\mathbf{x}^{(1)}$ et $\mathbf{x}^{(2)}$.*
6. *L'algorithme va t'il converger vers le minimum global ? Pourquoi ?*

Exercice 3.5 (Descente de gradient à pas optimal : etude de la convergence).

Soit $\gamma > 0$ *et considérons la fonction* $f_\gamma : \mathbb{R}^2 \to \mathbb{R}$ *définie par*

$$f_\gamma(\mathbf{x}) = \frac{1}{2}(x_1^2 + \gamma x_2^2).$$

On souhaite appliquer la descente de gradient à pas optimal pour optimiser cette fonction et étudier la convergence de l'algorithme.

1. *Déterminer le minimum de la fonction* f_γ *et pour quelle valeur de* $\mathbf{x}$ *ce minimum est atteint.*
2. *Rappeler ce qu'est la descente de gradient à pas optimal.*

On initialise notre descente de gradient au point $\mathbf{x}^{(0)} = (x_1^{(0)}, x_2^{(0)}) = (\gamma, 1)$.

3. *Calculer les valeurs de* $\mathbf{x}^{(1)}$ *et* $\mathbf{x}^{(2)}$ *à l'aide de l'algorithme.*
4. *Montrer que pour tout* $k \in \mathbb{N}$, *la valeur de* $\mathbf{x}^{(k)}$ *est donnée par*

$$\mathbf{x}^{(k)} = \left(\left(\frac{\gamma - 1}{\gamma + 1}\right)^k \gamma, \left(\frac{\gamma - 1}{\gamma + 1}\right)(-1)^k \right).$$

5. *Montrer que*

$$f(\mathbf{x}^{(k)}) = \left(\frac{\gamma - 1}{\gamma + 1}\right)^{2k} f(\mathbf{x}^{(0)}).$$

6. *Est-ce que l'algorithme converge ?*

Exercice 3.6 (Une application de la méthode de Newton : la régression logistique).

On considère une matrice $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ *où* $\mathbf{x}_i \in \mathbb{R}^{d+1}$ *et* $(y_1, y_2, \ldots, y_n) \in \{0, 1\}^n$ *une matrice dite "de données" et respectivement un vecteur contenant des étiquettes.*

La régression logistique est méthode de régression dite généralisée principalement utilisée en Machine Learning pour effectuer des tâches de classification (binaire ou multi-classes). On va alors chercher à estimer la probabilité d'appartenir à une classe dite de référence (dans notre cas, la classe 1 *par exemple).*

Sur le plan de la modélisation, on cherche à exprimer le log du rapport de deux probabilités comme une fonction linéaire, i.e.

$$f(\mathbf{x}) = \ln\left(\frac{\mathbb{P}(y = 1 \mid \mathbf{x})}{\mathbb{P}(y = 0 \mid \mathbf{x})}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_d x_d. \quad (2)$$

1. *A partir de la définition de* f *donnée en Equation* (2) *et en utilisant le fait que* $\mathbb{P}(y = 0 \mid \mathbf{x}) = 1 - \mathbb{P}(y = 1 \mid \mathbf{x})$, *donner une expression de* $\mathbb{P}(y = 1 \mid \mathbf{x})$ *en fonction de* $\mathbf{x}$ *et de* $\boldsymbol{\beta}$. *On appellera* g *une telle fonction dans la suite.*
2. *Quelles sont les valeurs que peut prendre la fonction* g *?*
3. *Montrer que pour tout vecteur* $\boldsymbol{\beta} \in \mathbb{R}^{d+1}$, *nous avons*

$$\nabla_{\boldsymbol{\beta}} g(\boldsymbol{\beta}) = g(\boldsymbol{\beta})(1 - g(\boldsymbol{\beta})).$$

4. *Etudier la convexité de la fonction* g.
5. *Que peut-on dire de la convexité de la fonction* $\ln(g)$ *?*

Les questions suivantes sont optionnelles car un peu plus complexes à effectuer même si vous disposez de l'ensemble des outils nécessaires pour y répondre.

Une méthode classique pour estimer la valeur du paramètre $\boldsymbol{\beta}$ de la régression logistique et de passer par la maximisation de la vraisemblance des données[1]

La vraisemblance d'une donnée $(\mathbf{x}_i, y_i)$ est donnée par

$$\mathbb{P}(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}) = g(\boldsymbol{\beta}, \mathbf{x}_i)^{y_i}(1 - g(\boldsymbol{\beta}, \mathbf{x}_i))^{1-y_i}.$$

On peut voir que si $y_i = 1$, alors la vraisemblance de cette donnée est égale à $g(\boldsymbol{\beta}, \mathbf{x}_i)$, qui est donc la probabilité d'appartenir à la classe 1. Par opposition $1 - g(\boldsymbol{\beta}, \mathbf{x}_i)$ correspond à la probabilité de d'appartenir à la classe 0.[2]

La vraisemblance de notre jeu de données L est alors définie comme le produit de la vraisemblance de chacun de nos exemples, i.e.

$$L(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y}) = \prod_{i=1}^{n} \mathbb{P}(y_i \mid \mathbf{x}_i, \boldsymbol{\beta}).$$

6. *Donner une expression de la vraisemblance L. Dans quel intervalle la fonction L prend-elle ses valeurs ?*
7. *On note $\ell = \ln(L)$ la log-vraisemblance de nos données*[3]*. Déterminer une expression de $-\ell$.*
8. *Etudier la convexité de $-\ell$.*
9. *Contrairement à la régression linéaire classique, il n'existe pas de méthode pour trouver une expression du vecteur $\boldsymbol{\beta}$, on doit obligatoirement passer par un problème d'optimisation. Et la méthode la plus utilisée la descente de Newton. Ecrire le processus itératif de la méthode de Newton pour résoudre le problème de minimisation*

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{d+1}} -\ell(\boldsymbol{\beta}, \mathbf{X}, \mathbf{y}).$$

Exercice 3.7 (Etude d'une fonction et optimisation)**.**

Soit γ un nombre réel. L'objectif de cet exercice est d'étudier la fonction $f_\gamma : \mathbb{R}^2 \to \mathbb{R}$ définie par :

$$f_\gamma(x, y) = \frac{1}{2}(x^2 + \gamma y^2 + 2xy) + 2x + 2y.$$

Etude de la fonction f_γ. *Cette première partie est consacrée à l'étude de la fonction f_γ.*

1. *Etudier la convexité de la fonction f_γ.*
2. *Donner les solutions de l'équation d'Euler, i.e., les solutions du système linéaire $\nabla f_\gamma(x, y) = (0, 0)$ pour toutes les valeurs de γ.*
3. *Donner la nature des extrema de la fonction f_γ en fonction de la valeur de γ.*
4. *Montrer que la fonction f_γ peut s'écrire sous la forme*

$$f_\gamma(\mathbf{u}) = \frac{1}{2}\mathbf{u}^T \mathbf{A} \mathbf{u} - \mathbf{b}^T \mathbf{u},$$

où $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & \gamma \end{pmatrix}$, $\mathbf{b} \in \mathbb{R}^2$ est un vecteur à déterminer et $\mathbf{u} = (x, y)^T$.

1. Cette notion de vraisemblance sera présentée en statistique en troisième année.
2. Cela doit vous faire penser à une variable aléatoire qui sui une loi de probabilité que vous connaissez très bien : la loi de Bernoulli avec deux issues possibles : *succès* ou *échec*.
3. Cette dernière est souvent utilisée car plus simple à manipuler

Taux de convergence de la descente de gradient à pas optimal. *A partir de maintenant, on suppose que $\gamma > 1$ de telle sorte que la fonction f soit bien strictement convexe. L'objectif est d'étudier la vitesse de convergence de la descente de gradient à pas optimal. Cette vitesse de convergence dépend de ce que l'on appelle* ***le conditionnement de la matrice*** $\mathbf{A}$*, notée $Cond(\mathbf{A})$, et il est défini par*

$$Cond(\mathbf{A}) = \frac{\lambda_{\max}(\mathbf{A})}{\lambda_{\min}(\mathbf{A})},$$

où $\lambda_{\max}(\mathbf{A})$ et $\lambda_{\min}(\mathbf{A})$ sont respectivement la plus grande et la plus petite valeur propre de la matrice $\mathbf{A}$.

1. *Déterminer les valeurs propres de la matrice $\mathbf{A}$.*
2. *Donner une expression du conditionnement de la matrice $\mathbf{A}$ en fonction de γ et donner un équivalent asymptotique de ce conditionnement, i.e. pour de grandes valeurs de γ.*
   Indice : on utilisera le fait que pour de grandes valeurs de γ on a $(\gamma-1)^2 + 4 \simeq (\gamma-1)^2$.
3. *Notons $\mathbf{u}^\star$ le point pour lequel la fonction f_γ atteint son minimum et $\mathbf{u}_0$ le point initial de notre algorithme de descente de gradient.*
   On définit le taux de convergence de notre algorithme par la nombre $\eta = 1 - Cond(\mathbf{A})^{-1}$ et on

$$\|\mathbf{u}_{k+1} - \mathbf{u}^\star\|_{\mathbf{A}} \leq \eta^k \|\mathbf{u}_0 - \mathbf{u}^\star\|_{\mathbf{A}}. \tag{3}$$

   A l'aide de cette définition dire pour quelles valeurs de γ la convergence de la l'algorithme est la plus rapide.[4]
4. *On cherche maintenant à démontrer l'inégalité donnée en Equation (3). On note ρ_k le pas optimal de notre algorithme à l'itération k.*
   (a) *Montrer que*

$$\|\mathbf{u}_{k+1} - \mathbf{u}^\star\|_{\mathbf{A}}^2 \leq \|(\mathbf{I} - \rho_k\mathbf{A})(\mathbf{u}_k - \mathbf{u}^\star)\|_{\mathbf{A}}^2.$$

   Indice : on se rappelle que si $\mathbf{u}^\star$ est un minimum de f_γ, alors $\mathbf{A}\mathbf{u}^\star = \mathbf{b}$ où $\mathbf{A}$ et $\mathbf{b}$ ont été définis dans la précédente partie.

   (b) *Maintenant, on admet que pour tout $k \in \mathbb{N}$, nous avons :*

$$\|\mathbf{u}_{k+1} - \mathbf{u}^\star\|_{\mathbf{A}}^2 \leq \|(\mathbf{I} - \rho_k\mathbf{A})\|_2^2 \|(\mathbf{u}_k - \mathbf{u}^\star)\|_{\mathbf{A}}^2.$$

   Montrer que η^2 est une borne supérieure de $\|\mathbf{I} - \rho_k\mathbf{A}\|_2^2$, i.e.,

$$\|\mathbf{I} - \rho_k\mathbf{A}\|_2^2 \leq \eta^2 = \left(1 - \frac{\lambda_{\min}(\mathbf{A})}{\lambda_{\max}(\mathbf{A})}\right)^2.$$

   (c) *Conclure quant à la convergence.*

---

4. Le conditionnement d'une matrice joue également un rôle important dans la stabilité des solutions numériques données par notre ordinateur lorsque la précision numérique est limitée. Pour des problèmes dits *mal conditionnés*, une faible perturbation des données peut engendrer une modification radicale de la solution, *i.e.*, une multiplication de l'erreur.

4 Intégrations des fonctions à plusieurs variables

Exercice 4.1 (Quelques calculs I)**.**

On considère le domaine de définition $\mathscr{D}$ *défini par*

$$\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ x \geq 0, y \geq 0,\ x+y \leq 1\right\}.$$

Représenter le domaine $\mathscr{D}$ *dans le plan.*

Calculer $\displaystyle\int\int_{\mathscr{D}} f(x,y)\, dxdy$ *dans les cas suivants*

1. $f(x,y) = x^2 + y^2$,
2. $f(x,y) = xy(x+y)$.

Exercice 4.2 (Quelques calculs II)**.**

Calculer les intégrales doubles $\displaystyle\int\int_{\mathscr{D}} f(x,y)\, dxdy$ *dans les cas suivants :*

1. $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ y \geq 0, x-y+1 \geq 0,\ x+2y-4 \leq 0\right\}$ *et* $f(x,y) = x$,
2. $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ 0 \leq x \leq 1, x^2 \leq y \leq x\right\}$ *et* $f(x,y) = x+y$,
3. $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ 1 \leq x \leq 2, 0 \leq xy \leq \dfrac{\pi}{2}\right\}$ *et* $f(x,y) = \cos(xy)$,
4. $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ x \geq 0, y \geq 0,\ xy+x+y \leq 1\right\}$ *et* $f(x,y) = xy$ *(plus difficile),*
5. $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ 1 < x < 3, y > 2,\ x+y < 5\right\}$ *et* $f(x,y) = \dfrac{1}{(x+y)^3}$.

On pourra commencer par représenter les domaines $\mathscr{D}$ *si nécessaires.*

Exercice 4.3 (Quelques calculs III)**.**

On considère le domaine $\mathscr{D}$

$$\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ -1 \leq x \leq 1 \text{ et } x^2 \leq y \leq 4-x^3\right\}.$$

Calculer l'aire de ce domaine.

Exercice 4.4. *Changement de variable I : coordonnées polaires*

Calculer l'intégrale double

$$\int\int_{\mathscr{D}} \frac{1}{x^2+y^2+1}\, dxdy,$$

où $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ 0 \leq x \leq 1 \text{ et } 0 \leq y \leq 1,\ 0 < x^2+y^2 \leq 1\right\}$.

Exercice 4.5. *Changement de variable II : coordonnées polaires*

Soit $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ x^2-2x+y^2 \leq 0\right\}$.

1. *Montrer que* $\mathscr{D}$ *est un disque.*
2. *Calculer* $\displaystyle\int\int_{\mathscr{D}} \sqrt{x^2+y^2}\, dxdy$.

Exercice 4.6. *Autre changement de variables*

Soit $\mathscr{D} = \left\{(x,y) \in \mathbb{R}^2;\ x < y < 2x,\ x < y^2 < 2x\right\}$. *Calculer l'intégrale*

$$\int\int_{\mathscr{D}} \frac{y}{x}\, dxdy,$$

en utilisant le changement de variables $u = x/y$ *et* $v = y^2/x$.

Exercice 4.7 (Un dernier changement de variable)**.**

L'objectif de cet exercice est de calculer l'intégrale suivante

$$\int_{-\infty}^{+\infty} e^{-x^2} = \sqrt{2\pi}.$$

La question étant de savoir d'où vient le nombre π *du résultat ainsi que la racine carrée, sachant que l'on ne connaît pas de primitive de la fonction* $x \mapsto e^{-x^2}$.
On note $I = \displaystyle\int_{-\infty}^{+\infty} e^{-x^2}$ *et on va chercher la valeur de* I *en passant par ... une intégrale double !*

1. *Montrer que* $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-x^2-y^2} = I^2$.
2. *A l'aide d'un passage coordonnées polaires (encore une fois) calculer la valeur de* $\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-x^2-y^2}$.
3. *En déduire la valeur de* I.

5 Preuves et vitesse de convergence

Les exercices de cette section sont essentiellement techniques et son destinées aux étudiants qui souhaitent aller plus en pratiquant des raisonnements. Ils nécessitent de bien connaître le cours et ses définitions.

Exercice 5.1 (Etude d'une fonction α-elliptique)**.**

On commence par introduire ce qu'est une fonction coercive.

Définition 5.1: Fonction α-elliptique

Soit V un espace vectoriel réel de dimension n. Une fonction $f : V \to \mathbb{R}$ est dite α-elliptique si f est continuement différentiable et si pour tout vecteur $\mathbf{u}, \mathbf{v} \in V$ il existe $\alpha > 0$ tel que :

$$\langle \nabla f(\mathbf{v}) - \nabla f(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle \geq \alpha \|\mathbf{v} - \mathbf{u}\|_2^2.$$

On introduit maintenant ce qu'est une fonction coercive[5]*.*

Définition 5.2: Coercivité

Soit U un sous-ensemble non borné de $\mathbb{R}^n$. Une fonction f est dite coercive si

$$\lim_{\mathbf{u}_k \in U, \|\mathbf{u}_k\| \to \infty} f(\mathbf{u}_k) = +\infty.$$

Notre objectif, à l'aide de ces définitions est de démontrer la proposition suivante concernant les fonctions α-elliptique

Proposition 5.1: Coercivité des fonctions α-elliptiques

Une fonction α-elliptique est strictement convexe et coercive.

1. *Montrer, à l'aide de la Définition 5.1, l'inégalité suivante pour tout $\mathbf{u}, \mathbf{v} \in V$*

$$f(\mathbf{v}) \geq f(\mathbf{u}) + \langle \nabla f(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle + \frac{\alpha}{2} \|\mathbf{v} - \mathbf{u}\|_2^2.$$

   Pour cela, on va considérer la fonction $\varphi : \mathbb{R} \to \mathbb{R}$ définie par

$$\varphi(t) = f(\mathbf{u} + t(\mathbf{v} - \mathbf{u}))$$

   et on cherchera à minorer $\varphi(1) - \varphi(0)$.

2. *A l'aide de la Définition 5.2, montrer qu'une fonction α-elliptique est coercive et convexe.*

La notion de coercivité permet de garantir l'existence d'un minimum et le caractère α-elliptique va permettre de garantir l'unicité de ce minimum et donc l'unicité de la solution du problème d'optimisation.

Exercice 5.2 (Retour sur la descente de gradient à pas optimal)**.**

5. La notion de coercivité d'une fonction permet de montrer qu'une fonction admet nécessairement un minimum, ce qui n'est pas une garantie avec la seule propriété de convexité.

L'objectif de cet exercice est de montrer le résultat suivant

Théorème 5.1: Convergence

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction continuement différentiable, i.e., de classe C^1 et α-elliptique. Alors, la descente de gradient à pas optimal converge.

Pour montrer ce résultat, on va considérer une suite $(\mathbf{u}_k)_{k\in\mathbb{N}}$ et un point $\mathbf{u}^\star$ où la fonction f atteint son minimum. On souhaite montrer que, pour cet algorithme, on a bien $\lim\limits_{k\to+\infty} \mathbf{u}_k = \mathbf{u}^\star$.

1. *Montrer que $f(\mathbf{u}_k) - f(\mathbf{u}_{k+1}) \geq \dfrac{\alpha}{2}\|\mathbf{u}_k - \mathbf{u}_{k+1}\|_2^2$.*
2. *Expliquer pourquoi $\|\mathbf{u}_k - \mathbf{u}_{k+1}\|_2^2$ converge vers 0 lorsque k tend vers $+\infty$.*
3. *Supposons que $\|\mathbf{u}_k - \mathbf{u}_{k+1}\|_2^2 \to 0$ implique $\|\nabla f(\mathbf{u}_k) - \nabla f(\mathbf{u}_{k+1})\|_2^2 \to 0$. Montrer que $\|\nabla f(\mathbf{u}_k)\| \to 0$.*
4. *Conclure en montrant que*

$$\lim_{k\to+\infty} \|\mathbf{u}_k - \mathbf{u}^\star\|_2 = 0.$$

Exercice 5.3 (Descente de gradient à pas variable : convergence)**.**

L'objectif de cet exercice est de montrer le résultat suivant

Théorème 5.2: Convergence

Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction α-elliptique telle que pour tout $\mathbf{u}, \mathbf{v}$ de $\mathbb{R}^n$ on ait :

$$\|\nabla f(\mathbf{u}) - \nabla f(\mathbf{v})\|_2^2 \leq M\|\mathbf{u} - \mathbf{v}\|_2^2,$$

et

$$\langle \nabla f(\mathbf{v}) - \nabla f(\mathbf{u}), \mathbf{v} - \mathbf{u}\rangle \geq \alpha\|\mathbf{v} - \mathbf{u}\|_2^2,$$

où $\alpha, M > 0$. Alors la descente de gradient à pas variable ρ_k converge pour

$$0 < a \leq \rho_k \leq b < \frac{2\alpha}{M},$$

où a et b sont des nombres quelconques qui sont là pour restreindre l'espace des valeurs possibles de ρ.

On note $\mathbf{u}^\star$ le point pour lequel la fonction f atteint son minimum.

1. *Trouver la valeur de γ pour laquelle on a $\|\mathbf{u}_{k+1} - \mathbf{u}^\star\|_2^2 \leq \gamma\|\mathbf{u}_k - \mathbf{u}^\star\|_2^2$.*
2. *Donner une condition sur γ telle que $\lim\limits_{k\to+\infty} \|\mathbf{u}_k - \mathbf{u}^\star\| = 0$ et conclure.*

Exercice 5.4 (Inégalité de Kantorovich)**.**

L'objectif de cet exercice est de donner un taux de convergence de la méthode de descente de gradient à pas optimal pour une forme quadratique $f : \mathbb{R}^n$ dans $\mathbb{R}$ de la forme :

$$f(\mathbf{u}) = \frac{1}{2}\langle \mathbf{A}\mathbf{u}, \mathbf{u}\rangle - \langle \mathbf{b}, \mathbf{u}\rangle = \frac{1}{2}\mathbf{u}^T\mathbf{A}\mathbf{u} - \mathbf{b}^T\mathbf{u},$$

où $\mathbf{A}$ est une matrice symétrique définie positive de taille n et $\mathbf{b} \in \mathbb{R}^n$. Ce taux de convergence sera donné en fonction du conditionnement de la matrice $\mathbf{A}$, noté $Cond(\mathbf{A})$.

Pour cela, on considère une suite $(\mathbf{u}_k)_{k\in\mathbb{N}}$ obtenue à l'aide de notre algorithme.

1. *Montrer que pour tout $\mathbf{u} \in \mathbb{R}^n$, nous avons*

$$\frac{\|\mathbf{u}\|_2^4}{\|\mathbf{u}\|_{\mathbf{A}}^2 \|\mathbf{u}\|_{\mathbf{A}^{-1}}^2} \geq \frac{4\lambda_1\lambda_n}{(\lambda_1+\lambda_n)^2},$$

où λ_1 et λ_n désignent respectivement la plus petite et al plus grande valeur propre de la matrice $\mathbf{A}$.

Indication : on pourra utiliser le fait que pour tout $t > 0$, on a $\|\mathbf{u}\|_{\mathbf{A}}^2 \|\mathbf{u}\|_{\mathbf{A}^{-1}}^2 = \|\mathbf{u}\|_{\frac{1}{t}\mathbf{A}}^2 \|\mathbf{u}\|_{t\mathbf{A}^{-1}}^2$. On utilisera également le fait que $(a+b)^2 \geq 4ab$.

2. *En utilisant le résultat précédent, montrer que pour tout vecteur $\mathbf{u} \in \mathbb{R}^n$:*

$$\|\mathbf{u}_{k+1} - \mathbf{u}\|_{\mathbf{A}}^2 \leq \left(\frac{Cond(\mathbf{A})-1}{Cond(\mathbf{A})+1}\right)^2 \|\mathbf{u}_k - \mathbf{u}\|_{\mathbf{A}}^2,$$

où $Cond(\mathbf{A}) = \dfrac{\lambda_n}{\lambda_1}$.

Exercice 5.5 (L'algorithme de Davidon-Fletcher-Powell)**.**

On considère une fonction $f : \mathbb{R}^n \to \mathbb{R}$ que l'on supposera fortement convexe et de classe C^1. L'algorithme de Davidon-Fltecher-Powell se présente comme suit

1. ***Initialisation :*** *on pose $\mathbf{S}_0 = \mathbf{I}$ et $\mathbf{u}_0 \in \mathbb{R}^n$*
2. *Tant que $\|\nabla f(\mathbf{u}_{k+1})\|_2^2 \geq \varepsilon$ où $\varepsilon > 0$, on effectue les opérations suivantes :*
   (a) $\rho_k = \arg\min_{\rho>0} f(\mathbf{u}_k - \rho\mathbf{S}_k\nabla f(\mathbf{u}_k))$.
   (b) $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k\mathbf{S}_k\nabla f(\mathbf{u}_k)$.
   (c) $\mathbf{S}_{k+1} = \mathbf{S}_k + \dfrac{\boldsymbol{\delta}_k\boldsymbol{\delta}_k^T}{\boldsymbol{\delta}_k^T\boldsymbol{\gamma}_k} - \dfrac{\mathbf{S}_k\boldsymbol{\gamma}_k\boldsymbol{\gamma}_k^T\mathbf{S}_k}{\boldsymbol{\gamma}_k^T\mathbf{S}_k\boldsymbol{\gamma}_k}$.

Les vecteurs $\boldsymbol{\gamma}_k$ et $\boldsymbol{\delta}_k$ sont définies de la façon suivante :

$$\boldsymbol{\gamma}_k = \nabla f(\mathbf{u}_{k+1}) - \nabla f(\mathbf{u}_k) \quad \text{et} \quad \boldsymbol{\delta}_k = \mathbf{u}_{k+1} - \mathbf{u}_k.$$

1. *Montrer que pour tout entier $k > 0$ nous avons*

$$\mathbf{S}_{k+1}\boldsymbol{\gamma}_k = \boldsymbol{\delta}_k.$$

Cette condition est appelée Condition de Quasi-Newton[6]*.*

2. *Soit $\mathbf{u}$ un vecteur de $\mathbb{R}^n$, montrer que l'on a les deux égalités suivantes*

$$\mathbf{u}^T\mathbf{S}_k\mathbf{u} = \frac{(\mathbf{u}^T\mathbf{S}_k\mathbf{u})(\boldsymbol{\gamma}_k^T\mathbf{S}_k\mathbf{u}) - (\boldsymbol{\gamma}_k^T\mathbf{S}_k\mathbf{u})^2}{\boldsymbol{\gamma}_k^T\mathbf{S}_k\boldsymbol{\gamma}_k} + \frac{(\mathbf{u}^T\boldsymbol{\delta}_k)^2}{\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k},$$

$$\boldsymbol{\gamma}_k^T\boldsymbol{\delta}_k = \rho_k\nabla f(\mathbf{u}_k)^T\mathbf{S}_k\nabla f(\mathbf{u}_k)).$$

3. *Montrer que les matrices $\mathbf{S}_k$ sont toutes symétriques et définies positives.*

6. Cet algorithme est une méthode approchée de la méthode de Newton notamment utilisée quand la hessienne de notre problème est une matrice *singulière*, pour ainsi proche d'une matrice non inversible ou encore dont le déterminant est très proche de 0.