Mathématiques 1

TSI

4 heures Calculatrice autorisée

2025

CONCOURS CENTRALE•SUPÉLEC

Introduction à l'optimisation convexe

L'optimisation est le domaine des Mathématiques dont le but est d'établir des méthodes pour la recherche de minimum d'une fonction à valeur réelle. Le problème traite de la recherche de minimum pour les fonctions de classe $\mathscr{C}^1$ de $\mathbb{R}^2$ vers $\mathbb{R}$, et plus spécifiquement sous diverses hypothèses de convexité des applications.

Le $\mathbb{R}$-espace vectoriel $\mathbb{R}^n$ est muni de son produit scalaire usuel $\langle . \mid . \rangle$ et de sa norme associée $\| \, . \, \|$. Ainsi, pour $\vec{u} = (x_1, x_2, \dots, x_n)$ et $\vec{v} = (x'_1, x'_2, \dots, x'_n)$:

$$\langle \vec{u} \mid \vec{v} \rangle = x_1 x'_1 + x_2 x'_2 + \cdots + x_n x'_n \quad \text{et} \quad \|\vec{u}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

Le vecteur nul est noté $\vec{0}$.

Pour une application $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ on note l'image par f d'un vecteur $\vec{u} = (x, y)$, soit $f(x, y)$ soit $f(\vec{u})$. Lorsque f est de classe $\mathscr{C}^1$, le vecteur gradient de f en $\vec{u}$ est noté $\nabla f(x, y)$ ou $\nabla f(\vec{u})$; c'est le vecteur :

$$\nabla f(\vec{u}) = \nabla f(x, y) = \left(\frac{\partial f}{\partial x}(x,y) \, , \, \frac{\partial f}{\partial y}(x, y) \right) \in \mathbb{R}^2$$

On dit que f atteint en $\vec{u}_* \in \mathbb{R}^2$ un minimum $f(\vec{u}_*)$ si pour tout $\vec{u} \in \mathbb{R}^2$, $f(\vec{u}) \geqslant f(\vec{u}_*)$.
On dit alors que f admet un minimum.

Le problème comporte 4 parties. Les parties 3 et 4 sont indépendantes entre elles.

Partie A – Généralités

On établit quelques résultats préliminaires ainsi que quelques résultats généraux sur l'optimisation que l'on applique pour interpréter la loi de Descartes de réfraction de la lumière.

I – Préliminaires : propriétés de la norme euclidienne

Dans cette partie n désigne un entier naturel non nul.

Q1. Soit $\lambda \in \mathbb{R}$, $\vec{u} \in \mathbb{R}^n$ et $\vec{v} \in \mathbb{R}^n$; montrer que $\|\lambda . \vec{u}\| = |\lambda| \times \|\vec{u}\|$ et que $\|\vec{u} + \vec{v}\|^2 = \|\vec{u}\|^2 + 2 \langle \vec{u} \mid \vec{v} \rangle + \|\vec{v}\|^2$.

Q2. Rappeler l'inégalité de Cauchy-Schwarz.

Q3. En déduire l'inégalité triangulaire : $\forall \vec{u} \in \mathbb{R}^n$, $\forall \vec{v} \in \mathbb{R}^n$,

$$\|\vec{u} + \vec{v}\| \leqslant \|\vec{u}\| + \|\vec{v}\| \, . \tag{1}$$

Q4. Montrer que : $\forall \vec{u} \in \mathbb{R}^n$, $\forall \vec{v} \in \mathbb{R}^n$,

$$\big| \, \|\vec{u}\| - \|\vec{v}\| \, \big| \leqslant \|\vec{u} + \vec{v}\| \, .$$

(On pourra appliquer deux fois l'inégalité triangulaire (1) avec $(\vec{u} + \vec{v})$ et $-\vec{v}$ puis avec $(\vec{u} + \vec{v})$ et $-\vec{u}$.)

II – Tout minimum est un point critique

Q5. Soit $g : \mathbb{R} \longrightarrow \mathbb{R}$ dérivable qui atteint son minimum en x_*, c'est-à-dire tel que pour tout $x \in \mathbb{R}$, $g(x) \geqslant g(x_*)$. Soient x_1, x_2 deux réels tels que $x_1 < x_* < x_2$. Déterminer les signes des deux taux d'accroissement :

$$\frac{g(x_1) - g(x_*)}{x_1 - x_*} \qquad \text{et} \qquad \frac{g(x_2) - g(x_*)}{x_2 - x_*}$$

En déduire que $g'(x_*) = 0$.

Q6. Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction de classe $\mathscr{C}^1$ atteignant un minimum en $\vec{u}_* = (x_*, y_*)$; montrer que $\nabla f(\vec{u}_*) = \vec{0}$. (On pourra considérer les deux fonctions d'une seule variable réelle $f^{y_*} : x \longmapsto f(x, y_*)$ et $f_{x_*} : y \longmapsto f(x_*, y)$.)

Q7. En considérant la fonction $f : (x, y) \longmapsto (x - y)^3$ montrer qu'un point critique n'est pas nécessairement un point en lequel f atteint un minimum.

III – Fonctions coercives

Une fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ est dite **coercive** si :

- f est continue, et
- $\displaystyle\lim_{\|(x,y)\| \to +\infty} f(x,y) = +\infty$, c'est à dire si : $\forall m \in \mathbb{R}$, $\exists a \geqslant 0$, $\forall (x, y) \in \mathbb{R}^2$, $\|(x, y)\| \geqslant a \implies f(x, y) \geqslant m$

Le but de cette partie est de montrer le résultat :

- Toute fonction coercive admet un minimum.

Soit $\mathscr{A} = \big\{(x, y) \in \mathbb{R}^2 \mid f(x, y) \leqslant f(0, 0)\big\} = f^{-1}\big(]-\infty; f(0, 0)]\big)$; montrons d'abord que $\mathscr{A}$ est un fermé borné de $\mathbb{R}^2$.

Q8. Montrer que $\mathscr{A}$ est un ensemble non vide et borné.

Q9. Soit $\vec{u} \notin \mathscr{A}$; notons $\varepsilon = f(\vec{u}) - f(\vec{0}) > 0$. Justifier l'existence de $r > 0$ tel que :

$$\forall \vec{v} \in \mathbb{R}^2,\ \|\vec{v} - \vec{u}\| \leqslant r \implies |f(\vec{v}) - f(\vec{u})| \leqslant \frac{\varepsilon}{2}$$

En déduire que :

$$\forall \vec{v} \in \mathbb{R}^2,\ \|\vec{v} - \vec{u}\| \leqslant r \implies f(\vec{v}) - f(\vec{u}) \geqslant -\frac{\varepsilon}{2}$$

Q10. Pour cette valeur de r, soit $B_F(\vec{u}, r) = \big\{\vec{v} \in \mathbb{R}^2 \mid \|\vec{v} - \vec{u}\| \leqslant r\big\}$ la boule fermée centrée en $\vec{u}$ et de rayon r. Déduire de **Q9** que pour tout $\vec{v} \in B_F(\vec{u}, r)$, $f(\vec{v}) - f(\vec{0}) \geqslant \frac{\varepsilon}{2}$.
En déduire que $B_F(\vec{u}, r)$ est incluse dans $\mathbb{R}^2 \setminus \mathscr{A} = \big\{\vec{v} \in \mathbb{R}^2 \mid \vec{v} \notin \mathscr{A}\big\}$.

Q11. En déduire que $\mathbb{R}^2 \setminus \mathscr{A}$ est un ouvert de $\mathbb{R}^2$ puis que $\mathscr{A}$ est un fermé de $\mathbb{R}^2$.

Montrons maintenant que f admet un minimum.

Q12. Justifier que f restreinte à $\mathscr{A}$ y admet un minimum, c'est à dire que $\exists \vec{u}_* \in \mathscr{A}$ tel que $\forall \vec{u} \in \mathscr{A}$, $f(\vec{u}) \geqslant f(\vec{u}_*)$.

Q13. En déduire que la fonction f admet un minimum sur $\mathbb{R}^2$.

IV – Application : loi de réfraction de la lumière de Descartes

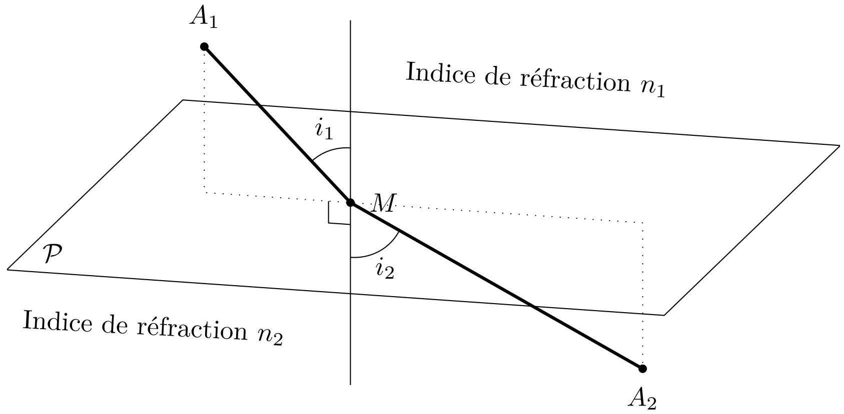

Figure 1 – Loi de réfraction de la lumière

Deux milieux d'indice de réfraction de la lumière homogènes n_1, n_2 sont séparés par un plan $\mathcal{P}$ orienté. Deux points A_1, A_2 sont situés de part et d'autre du plan $\mathcal{P}$. Selon la loi de réfraction de la lumière de Descartes (voir Figure 1), un rayon lumineux reliant les deux points A_1, A_2 se déplace de manière rectiligne dans chacun des deux milieux, et en passant par le point M du plan vérifiant :

- M est situé dans le plan perpendiculaire à $\mathcal{P}$ passant par A_1 et A_2,
- $n_1 \sin(i_1) = n_2 \sin(i_2)$ où i_1, i_2 désignent les angles entre $\overrightarrow{MA_1}$, $\overrightarrow{MA_2}$ avec une normale au plan $\mathcal{P}$.

Nous nous proposons ici de vérifier, en appliquant les résultats établis précédemment, le principe de Fermat selon laquelle ce trajet lumineux de A_1 à A_2 est celui de plus courte durée.

On rappelle que l'indice de réfraction n_i d'un milieu i ($i \in \{1,2\}$) est défini par :

$$n_i = \frac{c}{v_i}$$

où c et v_i désignent respectivement la vitesse de propagation de la lumière dans le vide et dans le milieu i.

Puisque dans un milieu homogène le trajet le plus rapide suit une ligne droite, on supposera que dans chacun des deux milieux le trajet est rectiligne. Soit M un point quelconque du plan $\mathcal{P}$; la durée du trajet lumineux $A_1 - M - A_2$ est donnée par :

$$\frac{A_1M}{v_1} + \frac{A_2M}{v_2} = \frac{n_1A_1M + n_2A_2M}{c}$$

Il s'agit donc de déterminer, s'il existe, le point M du plan où le «chemin optique» $n_1A_1M + n_2A_2M$ atteint un minimum.

Soient M_1 et M_2 les projetés orthogonaux de A_1 et A_2 sur le plan $\mathcal{P}$.

Q14. On suppose dans cette question que $M_1 = M_2$; vérifier que le chemin optique atteint un minimum au point M satisfaisant la loi de Descartes.

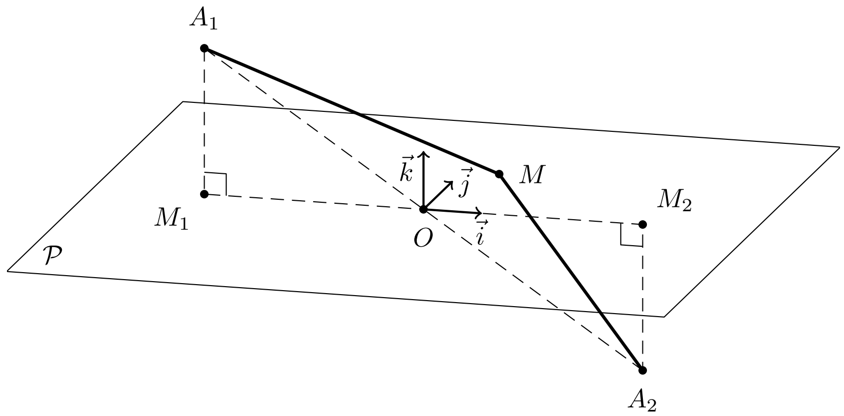

Figure 2

On supposera désormais $M_1 \neq M_2$. On construit un repère orthonormé $(O, \vec{i}, \vec{j}, \vec{k})$ de la façon suivante :

- l'origine O est l'intersection du plan $\mathcal{P}$ et de la droite (A_1A_2) ;
- le vecteur $\vec{i}$ a même direction que la droite (M_1M_2) ;
- le vecteur $\vec{j}$ est choisi de façon à ce que $(O, \vec{i}, \vec{j})$ soit un repère orthonormé direct de $\mathcal{P}$;
- le vecteur $\vec{k}$ est obtenu par le produit vectoriel $\vec{k} = \vec{i} \wedge \vec{j}$.

Ainsi dans ce repère (cf. Figure 2) les points ont pour coordonnées :

$$A_1 \begin{pmatrix} x_1 \\ 0 \\ z_1 \end{pmatrix} \quad ; \quad A_2 \begin{pmatrix} x_2 \\ 0 \\ z_2 \end{pmatrix} \quad ; \quad M_1 \begin{pmatrix} x_1 \\ 0 \\ 0 \end{pmatrix} \quad ; \quad M_2 \begin{pmatrix} x_2 \\ 0 \\ 0 \end{pmatrix} \quad ; \quad M \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}.$$

On définit alors :

$$\begin{array}{rcl} f: \quad \mathbb{R}^2 & \longrightarrow & \mathbb{R} \\ (x,y) & \longmapsto & f(x,y) = n_1 A_1M + n_2 A_2M = n_1\sqrt{(x-x_1)^2 + y^2 + z_1^2} + n_2\sqrt{(x-x_2)^2 + y^2 + z_2^2} \end{array}$$

qui donne le chemin optique en fonction de x, y et dont il s'agit de déterminer le point où un minimum est atteint.

Q15. Justifier que f est de classe $\mathscr{C}^1$ et calculer ses dérivées partielles.

Q16. Appliquer la question **Q4** pour montrer que $f(x,y) \geqslant n_1 \left| \|\overrightarrow{OM}\| - \|\overrightarrow{OA_1}\| \right| + n_2 \left| \|\overrightarrow{OM}\| - \|\overrightarrow{OA_2}\| \right|$.
En déduire que f est coercive.

Q17. Montrer que si (x,y) est un minimum de f, alors le point M est nécessairement sur le segment $[M_1M_2]$.

Q18. En déduire que si f atteint un minimum en (x,y) alors le point M satisfait :

$$n_1 \frac{M_1M}{A_1M} = n_2 \frac{M_2M}{A_2M}.$$

On appliquera pour cela l'équation $\frac{\partial f}{\partial x}(x,y) = 0$.

Q19. En déduire qu'au chemin optique minimal, le point M satisfait la loi de Descartes de réfraction de la lumière.

Partie B – Convexités de fonctions

Dans cette partie on introduit trois notions de convexité plus ou moins fortes pour les applications de $\mathbb{R}^2$ dans $\mathbb{R}$ et on montre l'intérêt que revêtent ces notions pour la recherche de minimum.

I – Fonction convexes ; strictement convexes

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction de classe $\mathscr{C}^1$.

- f est dite **convexe** si pour tous vecteurs $\vec{u}$, $\vec{v}$ de $\mathbb{R}^2$

$$f(\vec{u}) \geqslant f(\vec{v}) + \langle \nabla f(\vec{v}) \mid \vec{u} - \vec{v} \rangle$$

- f est dite **strictement convexe** si pour tous vecteurs $\vec{u}$, $\vec{v}$ de $\mathbb{R}^2$

$$\vec{u} \neq \vec{v} \implies f(\vec{u}) > f(\vec{v}) + \langle \nabla f(\vec{v}) \mid \vec{u} - \vec{v} \rangle$$

Q20. Montrer que si f est strictement convexe alors f est convexe. Montrer qu'une fonction convexe n'est pas forcément strictement convexe (on pourra considérer une fonction f constante).

Puisque f est $\mathscr{C}^1$, la surface d'équation $z = f(x,y)$ admet en tout point $(x_0, y_0, f(x_0, y_0))$ un plan tangent.

Q21. Donner l'équation du plan tangent au point $(x_0, y_0, f(x_0, y_0))$ à la surface d'équation $z = f(x,y)$.
En déduire une interprétation géométrique de f convexe (respectivement de f strictement convexe).

Q22. Montrer que si f est convexe et $\vec{u}_*$ est un point critique ($i.e$ $\nabla f(\vec{u}_*) = \vec{0}$), alors f atteint un minimum en $\vec{u}_*$.
Montrer que si de plus f est strictement convexe, alors $\vec{u}_*$ est l'unique point où f atteint un minimum.

Autrement dit, une fonction convexe atteint nécessairement un minimum en un point critique alors que ce n'est forcément le cas pour une fonction quelconque. Pour une fonction strictement convexe, ce point où le minimum est atteint est unique, s'il existe. Soit

$$f : \begin{cases} \mathbb{R}^2 & \longrightarrow \mathbb{R} \\ (x,y) & \longmapsto \mathrm{e}^x + \mathrm{e}^y \end{cases}$$

Q23. Soit $a \in \mathbb{R}$; dresser le tableau de variation de l'application

$$g_a : \begin{cases} \mathbb{R} & \longrightarrow \mathbb{R} \\ x & \longmapsto \mathrm{e}^x - \mathrm{e}^a(1 + x - a) \end{cases}$$

Q24. Montrer que f est strictement convexe et n'admet aucun minimum.

II – Fonctions fortement convexes

Pour une fonction, être convexe ou strictement convexe n'assure pas de l'existence d'un minimum. Pour cette raison on introduit une notion de convexité plus forte.

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction de classe $\mathscr{C}^1$ et soit α un réel strictement positif.

- f est dite **α-fortement convexe** si pour tous vecteurs $\vec{u}$, $\vec{v}$ de $\mathbb{R}^2$

$$f(\vec{u}) \geqslant f(\vec{v}) + \langle \nabla f(\vec{v}) \mid \vec{u} - \vec{v} \rangle + \alpha \|\vec{u} - \vec{v}\|^2 \ ;$$

- f est dite **fortement convexe** si il existe $\alpha > 0$ tel que f soit α-fortement convexe.

Q25. Montrer que si f est fortement convexe alors f est strictement convexe.

Q26. Exemple : Montrer que la fonction $f : (x,y) \longmapsto x^2 + y^2$ est fortement convexe ; pour quelles valeurs de α est-elle α-fortement convexe ?

On se propose de montrer que les fonctions fortement convexes sont coercives.

Q27. Montrer que si f est α-fortement convexe, alors pour tout $\vec{u} \in \mathbb{R}^2$, $\vec{v} \in \mathbb{R}^2$

$$f(\vec{u}) \geqslant f(\vec{v}) - \|\nabla f(\vec{v})\| \times (\|\vec{u}\| + \|\vec{v}\|) + \alpha \left(\|\vec{u}\| - \|\vec{v}\|\right)^2 .$$

(On appliquera pour cela des résultats de la partie **I.1**.)

Q28. En déduire que si f est α-fortement convexe, alors f est coercive, puis que toute fonction fortement convexe admet un minimum atteint en un unique point $\vec{u}_*$, et caractérisé par l'équation $\nabla f(\vec{u}_*) = \vec{0}$.

Partie C – Cas particulier des fonctions quadratiques

Dans cette partie on étudie la cas particulier des fonctions quadratiques, ou polynomiales de degré 2.

On note $\mathscr{M}_2(\mathbb{R})$ l'ensemble des matrices carrées ayant deux lignes et deux colonnes et à coefficients réels.

On identifiera les vecteurs de $\mathbb{R}^2$ avec la matrice colonne de leurs coordonnées dans la base canonique ; par exemple le vecteur $\vec{u} = (x,y) \in \mathbb{R}^2$ sera représenté $\begin{pmatrix} x \\ y \end{pmatrix}$. Avec cette convention, pour une matrice $M \in \mathscr{M}_2(\mathbb{R})$ et un vecteur $\vec{u} \in \mathbb{R}^2$, la notation $M\vec{u}$ est définie par le produit matriciel $M \times \begin{pmatrix} x \\ y \end{pmatrix}$ et l'application $\vec{u} \longmapsto M\vec{u}$ est l'endomorphisme de $\mathbb{R}^2$ dont la matrice dans la base canonique est M.

Une fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ est dite **quadratique** s'il existe une matrice $M \in \mathscr{M}_2(\mathbb{R})$ symétrique, un vecteur $\vec{n} \in \mathbb{R}^2$ et un réel k tel que

$$\forall \vec{u} \in \mathbb{R}^2, \quad f(\vec{u}) = \frac{1}{2} \langle M\vec{u} \mid \vec{u} \rangle - \langle \vec{n} \mid \vec{u} \rangle + k.$$

En notant $\vec{u} = \begin{pmatrix} x \\ y \end{pmatrix}$, si $M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$, $\vec{n} = \begin{pmatrix} d \\ e \end{pmatrix}$, alors : $f(x,y) = \dfrac{1}{2}\left(ax^2 + 2bxy + cy^2\right) - dx - ey + k.$

Q29. Soit f polynomiale de degré 2, $f(x,y) = \alpha x^2 + \beta xy + \gamma y^2 + \delta x + \varepsilon y + k$ avec $(\alpha, \beta, \gamma, \delta, \varepsilon, k) \in \mathbb{R}^6$.
Donner la matrice M et le vecteur $\vec{n}$ tels que $f(\vec{u}) = \frac{1}{2} \langle M\vec{u} \mid \vec{u}\rangle - \langle \vec{n} \mid \vec{u}\rangle + k$.

Dans la suite de cette partie, f désignera la fonction quadratique

$$f : \vec{u} \longmapsto f(\vec{u}) = \frac{1}{2} \langle M\vec{u} \mid \vec{u}\rangle - \langle \vec{n} \mid \vec{u}\rangle + k.$$

avec $M \in \mathscr{M}_2(\mathbb{R})$ une matrice symétrique et $\vec{n} \in \mathbb{R}^2$.

Q30. Établir que f est de classe $\mathscr{C}^1$ et que pour tout $\vec{u} \in \mathbb{R}^2$, $\nabla f(\vec{u}) = M\vec{u} - \vec{n}$.

On souhaite maintenant établir des conditions pour que f soit convexe, strictement convexe, ou fortement convexe.

Q31. Montrer que pour tout $\vec{u} \in \mathbb{R}^2$ et $\vec{v} \in \mathbb{R}^2$

$$\langle M\vec{u} \mid \vec{v}\rangle = \langle M\vec{v} \mid \vec{u}\rangle .$$

Q32. Montrer que

$$\forall \vec{v} \in \mathbb{R}^2, \forall \vec{x} \in \mathbb{R}^2, \ f(\vec{v} + \vec{x}) - f(\vec{v}) = \langle \nabla f(\vec{v}) \mid \vec{x}\rangle + \frac{1}{2} \langle M\vec{x} \mid \vec{x}\rangle .$$

Une matrice $M \in \mathscr{M}_2(\mathbb{R})$ symétrique est dite :

- **positive** si $\forall \vec{x} \in \mathbb{R}^2$, $\langle M\vec{x} \mid \vec{x}\rangle \geqslant 0$;
- **strictement positive** si $\forall \vec{x} \in \mathbb{R}^2 \setminus \{\vec{0}\}$, $\langle M\vec{x} \mid \vec{x}\rangle > 0$.

Q33. Montrer que f est :
– convexe si et seulement si M est positive ;
– strictement convexe si et seulement si M est strictement positive ;
– α-fortement convexe si et seulement si

$$\forall \vec{x} \in \mathbb{R}^2 \setminus \{\vec{0}\}, \ \langle M\vec{x} \mid \vec{x}\rangle \geqslant 2\alpha \|\vec{x}\|^2.$$

On caractérise maintenant la convexité de f à l'aide du signe des valeurs propres de la matrice M.

Q34. Justifier l'existence d'une matrice diagonale $D \in \mathscr{M}_2(\mathbb{R})$ et d'une matrice orthogonale $P \in O_2(\mathbb{R})$ tel que $M = PDP^{-1}$.

Q35. Déduire de **Q33** et **Q34** que f est convexe si et seulement si les valeurs propres de M sont toutes positives.

Q36. Montrer de même que f est strictement convexe si et seulement si toutes les valeurs propres de M sont strictement positives si et seulement si f est fortement convexe.

Q37. En déduire une condition nécessaire et suffisante portant sur la trace $\text{tr}(M)$ et le déterminant $\det(M)$ de M pour que f soit convexe, respectivement strictement convexe, fortement convexe.

Q38. Proposer une méthode n'utilisant que $\det(M)$, $\text{tr}(M)$ et les solution du système $M \begin{pmatrix} x \\ y \end{pmatrix} = \vec{n}$ pour déterminer tous les (éventuels) points en lesquels une fonction quadratique f atteint son minimum.

Partie D – Recherche approchée de minimum par une méthode de descente de gradient

Comme on l'a vu, une fonction $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ fortement convexe admet un minimum atteint en un unique point $\vec{u}_*$ et caractérisé par l'équation $\nabla f(\vec{u}_*) = \vec{0}$.

Mais résoudre cette dernière équation n'est pas toujours facile ni même possible. Aussi ont été inventées des méthodes de résolution approchées : elles consistent à construire une suite $(\vec{u}_n)_{n\in\mathbb{N}}$ de points de $\mathbb{R}^2$ qui converge vers le point $\vec{u}_*$ où f atteint son minimum, c'est-à-dire telle que $\|\vec{u}_n - \vec{u}_*\| \underset{n\to+\infty}{\longrightarrow} 0$. La méthode est d'autant meilleure que la convergence est rapide.

Parmi celles-ci, la méthode de descente de gradient (à pas fixe) construit la suite $(\vec{u}_n)_{n\in\mathbb{N}}$ définie par :

$$(\vec{u}_n)_{n\in\mathbb{N}} \ : \ \begin{cases} \vec{u}_0 \text{ est un point quelconque de } \mathbb{R}^2 \\ \forall n \in \mathbb{N},\ \vec{u}_{n+1} = \vec{u}_n - \rho.\nabla f(\vec{u}_n) \end{cases} \tag{$*$}$$

où $\rho > 0$ est un réel strictement positif.

Dans cette méthode, à chaque point $\vec{u}_n$, le point suivant est obtenu en se déplaçant d'un pas fixé ρ dans la direction locale de plus grande descente $-\nabla f(\vec{u}_n)$.

Q39. Soit $n \in \mathbb{N}$. Notons $\vec{u}_n = (x_n, y_n)$ et $\vec{u} = (x, y) \in \mathbb{R}^2$; soit $z(\vec{u})$ le réel tel que le point de coordonnées $(x, y, z(\vec{u}))$ appartienne au plan tangent à la surface représentative de f au point de coordonnées $(x_n, y_n, f(x_n, y_n))$.

Exprimer $z(\vec{u})$ en fonction de $\vec{u}_n$, $f(\vec{u}_n)$ et $\nabla f(\vec{u}_n)$.

Soit $r > 0$ fixé et $\vec{w} \in \mathbb{R}^2$ tel que $\|\vec{w}\| = 1$; montrer que si $\nabla f(\vec{u}_n) \neq \vec{0}$ alors $z(\vec{u}_n + r.\vec{w})$ atteint son minimum pour $\vec{w} = -\frac{\nabla f(\vec{u}_n)}{\|\nabla f(\vec{u}_n)\|}$. C'est ce que veut dire que la direction locale de plus grande descente est $-\nabla f(\vec{u}_n)$.

Afin que la suite $(\vec{u}_n)_{n\in\mathbb{N}}$ définie en $(*)$ converge vers le point où f atteint son minimum, le pas ρ de descente doit être choisi correctement et la fonction f vérifiera une condition supplémentaire.

On établit dans cette partie le résultat suivant :

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ une fonction α-fortement convexe. Si ∇f est M-lipschitzienne, c'est-à-dire si :

$$\exists M > 0,\ \forall \vec{u} \in \mathbb{R}^2, \forall \vec{v} \in \mathbb{R}^2,\ \|\nabla f(\vec{u}) - \nabla f(\vec{v})\| \leqslant M \|\vec{u} - \vec{v}\|$$

alors en choisissant ρ tel que $0 < \rho < \dfrac{4\alpha}{M^2}$, la suite $(\vec{u}_n)_{n\in\mathbb{N}}$ définie par $\vec{u}_0 \in \mathbb{R}^2$ et :

$$\forall n \in \mathbb{N},\ \vec{u}_{n+1} = \vec{u}_n - \rho.\nabla f(\vec{u}_n)$$

converge vers l'unique point $\vec{u}_*$ où f atteint un minimum, et la convergence est géométrique, c'est-à-dire qu'il existe $k \in [0,1[$ tel que pour tout $n \in \mathbb{N}$:

$$\|\vec{u}_n - \vec{u}_*\| \leqslant k^n \|\vec{u}_0 - \vec{u}_*\| \underset{n\to+\infty}{\longrightarrow} 0$$

On se place sous ces hypothèses : on considère dans cette partie une fonction f α-fortement convexe atteignant un minimum en $\vec{u}_*$, telle que ∇f est M-lipschitzienne, ainsi qu'une suite $(\vec{u}_n)_{n\in\mathbb{N}}$ définie par la relation de récurrence $\vec{u}_{n+1} = \vec{u}_n - \rho.\nabla f(\vec{u}_n)$ avec $\rho > 0$.

Q40. Établir que pour tout $n \in \mathbb{N}$, $\vec{u}_{n+1} - \vec{u}_* = (\vec{u}_n - \vec{u}_*) - \rho.(\nabla f(\vec{u}_n) - \nabla f(\vec{u}_*))$.

Q41. Montrer que pour toute fonction f, α-fortement convexe :

$$\forall \vec{u} \in \mathbb{R}^2,\ \forall \vec{v} \in \mathbb{R}^2,\ \langle \nabla f(\vec{u}) - \nabla f(\vec{v}) \mid \vec{u} - \vec{v} \rangle \geqslant 2\alpha \|\vec{u} - \vec{v}\|^2$$

Q42. Montrer que pour tout $n \in \mathbb{N}$, $\|\vec{u}_{n+1} - \vec{u}_*\|^2 \leqslant \left(1 - 4\alpha\rho + M^2\rho^2\right) \times \|\vec{u}_n - \vec{u}_*\|^2$.

Q43. En notant $t(\rho) = 1 - 4\alpha\rho + M^2\rho^2$, montrer que $t(\rho)$ atteint un minimum en $\dfrac{2\alpha}{M^2}$. En déduire que $\alpha \leqslant \dfrac{M}{2}$.

Q44. Montrer que si $0 < \rho < \dfrac{4\alpha}{M^2}$, alors $0 \leqslant t(\rho) < 1$, puis conclure.

◇ Fin ◇