

Optimisation

I. Extremums et points critiques en optimisation

Dans tout le chapitre, E désigne un espace euclidien de dimension n , identifié à $\mathbb{R}^n$ grâce à une base orthonormée. De plus, U est un ouvert de E et $f:U \rightarrow \mathbb{R}$ une fonction numérique. On cherche les points où f est maximale ou minimale. Cette recherche est l'objet de l'optimisation.

Définition :

Soit A une partie de U et $a \in A$. On dit que f admet en a un **maximum local** sur A s'il existe $r > 0$ tel que $f(x) \leq f(a)$ pour tout $x \in A \cap B(a,r)$. Le maximum est **global** sur A si $f(x) \leq f(a)$ pour tout $x \in A$. On définit de même un minimum local ou global. Un extremum est un maximum ou un minimum ; il est **strict** si l'inégalité est stricte pour $x \neq a$.

Définition :

Si f est différentiable sur U , un point $a \in U$ est un **point critique** de f lorsque $df(a) = 0$, autrement dit lorsque $\nabla f(a) = 0$. Dans une base orthonormée, cela revient à $\partial_1 f(a) = \dots = \partial_n f(a) = 0$.

1. La condition nécessaire du premier ordre

Théorème :

Soit $f:U \rightarrow \mathbb{R}$ différentiable sur l'ouvert U . Si f admet un extremum local en $a \in U$, alors a est un point critique : $\nabla f(a) = 0$.

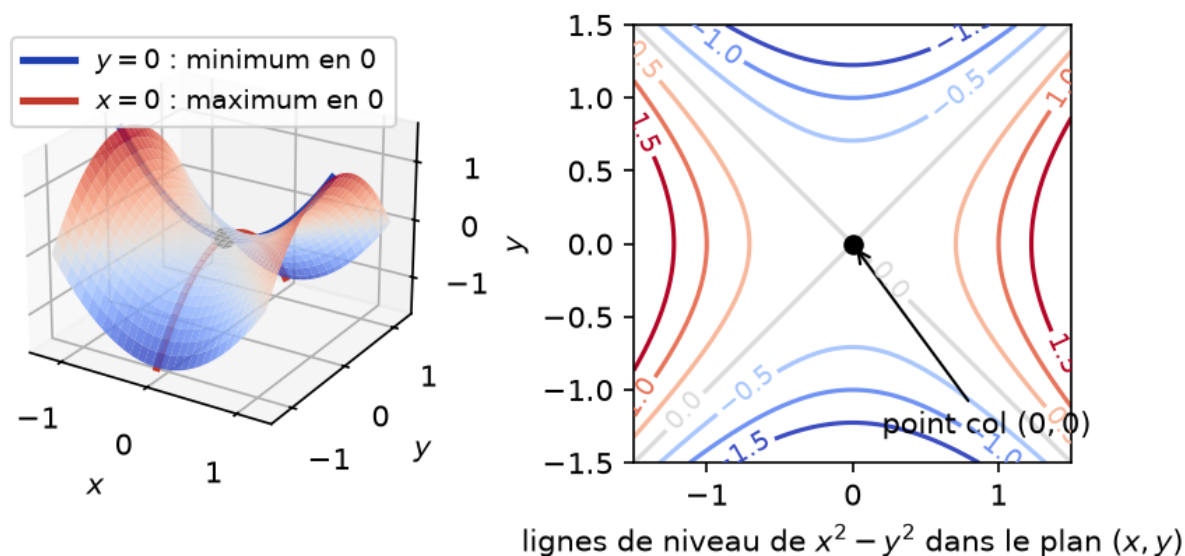
Démonstration :

Supposons que a soit un minimum local. Fixons $h \in E$. Comme U est ouvert, la fonction $\varphi(t) = f(a+th)$ est définie sur un intervalle ouvert contenant 0 . Elle est dérivable en 0 avec $\varphi'(0) = df(a)(h)$. De plus, φ admet un minimum local en 0 , qui est intérieur à son intervalle de définition. Donc $\varphi'(0) = 0$. Ainsi $df(a)(h) = 0$ pour tout h , c'est-à-dire $df(a) = 0$.

Attention :

La réciproque est fausse. Par exemple, $f(x,y) = x^2 - y^2$ vérifie $\nabla f(0,0) = 0$. Cependant, $f(x,0) = x^2 > 0$ et $f(0,y) = -y^2 < 0$ pour x, y non nuls. L'origine n'est donc ni un minimum ni un maximum : c'est un **point col**. En outre, le théorème exige un point intérieur : sur $[0,1]$, $x \rightarrow x$ atteint son minimum en 0 sans dérivée nulle.

La figure ci-dessous montre ce point col. En effet, le long de l'axe des abscisses, la surface présente un minimum ; en revanche, le long de l'axe des ordonnées, elle présente un maximum. Les lignes de niveau sont des hyperboles qui se croisent en l'origine.



Exemple :

Soit $f(x,y) = x^2 + y^2 - 2x + 4y$. Ses dérivées partielles sont $2x - 2$ et $2y + 4$. Le seul point critique est donc $(1, -2)$. Or $f(x,y) = (x-1)^2 + (y+2)^2 - 5$. Par conséquent, f admet en $(1, -2)$ un minimum global strict égal à -5 .

2. Extremum global sur un compact

Une fonction continue sur un compact non vide est bornée et atteint ses bornes. C'est le théorème des bornes atteintes, vu avec la compacité. Il garantit l'existence d'un extremum global. Ensuite, le calcul différentiel sert à le localiser.

Méthode :

Pour trouver les extremums globaux d'une fonction continue f sur un compact K :

- justifier que K est compact (fermé et borné en dimension finie) et que f est continue, donc que les extremums existent ;

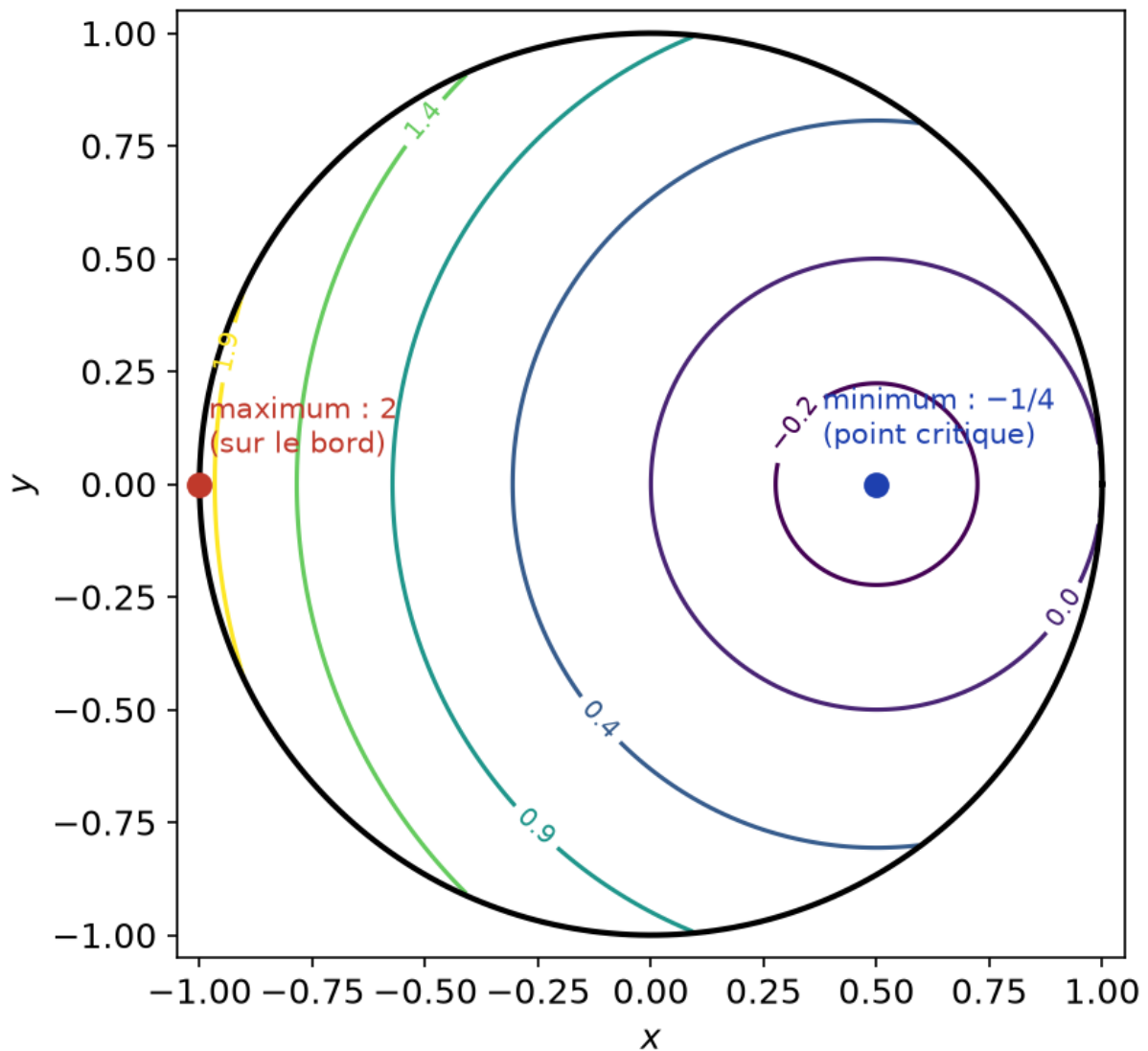
- chercher les points critiques de f dans l'intérieur de K ;
- étudier f sur la frontière de K , souvent grâce à un paramétrage ;
- comparer toutes les valeurs obtenues.

Sur un domaine non borné, on se ramène à un compact : on montre que f est plus grande qu'une valeur déjà atteinte en dehors d'une boule.

Exemple :

Cherchons les extremums de $f(x,y) = x^2 + y^2 - x$ sur le disque fermé D de centre 0 et de rayon 1 . D'abord, D est compact et f continue. Ensuite, le seul point critique est $(1/2, 0)$, intérieur à D , avec $f(1/2, 0) = -1/4$. Sur le cercle, $x^2 + y^2 = 1$ donc $f = 1 - x$, qui varie de 0 à 2 . Finalement, le minimum vaut $-1/4$ et le maximum vaut 2 , atteint en $(-1, 0)$.

Comme le montre la figure suivante, le minimum est atteint en un point critique intérieur. En revanche, le maximum est atteint sur le bord, où le gradient ne s'annule pas.



II. Extremums sur une partie et extremums liés

Souvent, la variable est astreinte à rester sur une partie X de U : un cercle, une sphère, une surface. On parle alors d'optimisation sous contrainte. Le gradient ne s'annule plus forcément ; seule sa composante tangente à X s'annule.

1. Vecteurs tangents et condition nécessaire

Définition :

Un vecteur v est **tangent** à X en a s'il existe $\varepsilon > 0$ et un arc $\gamma:]-\varepsilon, \varepsilon[\rightarrow X$, dérivable en 0 , tel que $\gamma(0) = a$ et $\gamma'(0) = v$. L'ensemble de ces vecteurs se note $T_a X$.

Théorème :

Soit $f: U \rightarrow \mathbb{R}$ différentiable et $X \subset U$. Si la restriction de f à X admet un extremum local en $a \in X$, alors $df(a)(v) = 0$ pour tout $v \in T_a X$. Autrement dit, $\nabla f(a)$ est orthogonal à $T_a X$.

Démonstration :

Soit $v \in T_a X$ et γ un arc comme dans la définition. La fonction $\varphi = f \circ \gamma$ est définie sur $]-\varepsilon, \varepsilon[$. Par la règle de la chaîne, elle est dérivable en 0 avec $\varphi'(0) = df(a)(\gamma'(0)) = df(a)(v)$. Or γ prend ses valeurs dans X , donc φ admet un extremum local en 0 , point intérieur. Ainsi $\varphi'(0) = 0$.

2. Optimisation sous une contrainte

Le chapitre précédent a établi le résultat suivant. Si $g: U \rightarrow \mathbb{R}$ est de classe C^1 , si $X = g^{-1}(\{0\})$ et si $dg(a) \neq 0$, alors $T_a X = \ker dg(a) = \nabla g(a)^\perp$. On en déduit la condition des extremums liés.

Théorème :

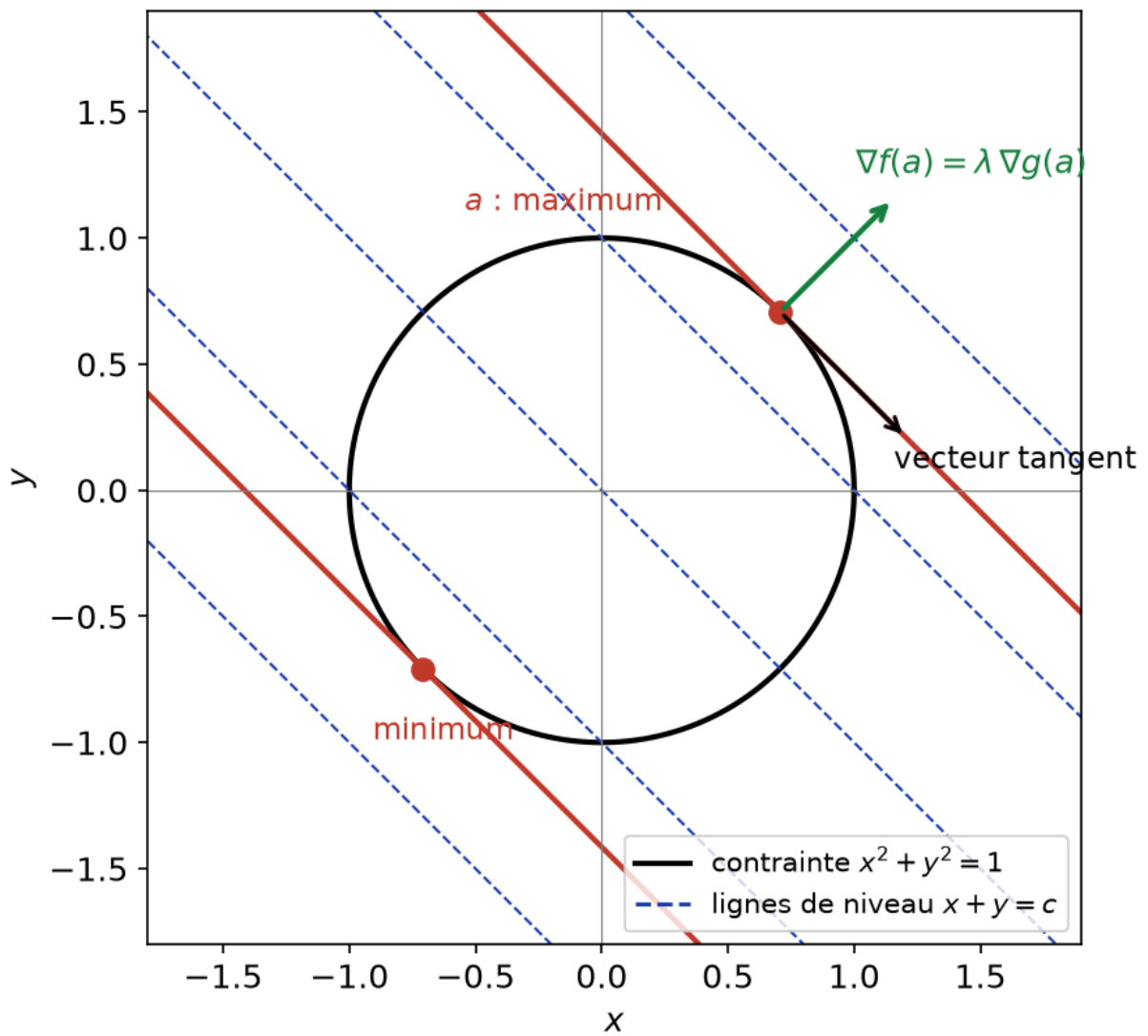
Soit $f, g: U \rightarrow \mathbb{R}$ de classe C^1 et $X = \left\{ x \in U, \begin{matrix} g(x) = 0 \end{matrix} \right\}$. Soit $a \in X$ tel que $\nabla g(a) \neq 0$. Si $f|_X$ admet un extremum local en a , alors il existe un réel λ , appelé multiplicateur de Lagrange, tel que

$$\nabla f(a) = \lambda \nabla g(a).$$

Démonstration :

D'après le théorème précédent, $\nabla f(a)$ est orthogonal à $T_a X = \nabla g(a)^\perp$. Donc $\nabla f(a) \in \left(\nabla g(a)^\perp\right)^\perp$. En dimension finie, ce biorthogonal est la droite $\mathbb{R} \nabla g(a)$, d'où le résultat.

Géométriquement, en un extremum lié, la ligne de niveau de f est tangente à la contrainte. La figure ci-dessous l'illustre avec $f(x,y) = x + y$ sur le cercle unité : les droites $x + y = c$ touchent le cercle exactement aux deux extremums.



Méthode :

Pour optimiser f sur $X = \{g = 0\}$:

- vérifier que ∇g ne s'annule pas sur X (sinon, traiter à part les points où il s'annule) ;
- résoudre le système $\nabla f(x) = \lambda \nabla g(x)$, $g(x) = 0$, d'inconnues x et λ ;
- justifier l'existence de l'extremum, souvent parce que X est compact ;
- comparer les valeurs de f aux points trouvés.

Exemple :

Cherchons les extremums de $f(x,y) = x + y$ sur le cercle $g(x,y) = x^2 + y^2 - 1 = 0$. Le gradient $\nabla g = (2x, 2y)$ ne s'annule pas sur le cercle. La condition $(1,1) = \lambda (2x, 2y)$ impose $x = y$, donc $x = y = \pm \frac{\sqrt{2}}{2}$. Le cercle est compact, donc les extremums existent. Par conséquent, le maximum vaut $\sqrt{2}$ et le minimum vaut $-\sqrt{2}$.

III. Fonctions de classe C^k et théorème de Schwarz

Pour aller au second ordre, il faut dériver deux fois. On définit donc les dérivées partielles successives.

Définition :

Les dérivées partielles d'ordre k de f sont, lorsqu'elles existent, les fonctions $\partial_{j_k} \dots \partial_{j_1} f$ obtenues par k dérivations partielles successives. On note aussi $\frac{\partial^2 f}{\partial x_i \partial x_j} = \partial_i \partial_j f$. La fonction f est de **classe** C^k sur U si toutes ses dérivées partielles d'ordre k existent et sont continues sur U . Elle est de classe C^∞ si elle est C^k pour tout k .

Propriété :

Les combinaisons linéaires, les produits et les composées de fonctions de classe C^k sont de classe C^k . De même, le quotient de deux fonctions C^k est C^k là où le dénominateur ne s'annule pas. En particulier, les fonctions polynomiales en les coordonnées sont de classe C^∞ .

Théorème :

Théorème de Schwarz. Si f est de classe C^2 sur l'ouvert U , alors pour tous i, j et tout

$x \in U :$

$$\partial_i \partial_j f(x) = \partial_j \partial_i f(x).$$

Ce théorème est admis. Ainsi, l'ordre des dérivations ne compte pas pour une fonction C^2 . Plus généralement, pour une fonction C^k , une dérivée d'ordre k ne dépend que du nombre de dérivations effectuées selon chaque variable.

Attention :

L'hypothèse C^2 est indispensable. En effet, la fonction $f(x,y) = \frac{xy(x^2-y^2)}{x^2+y^2}$, prolongée par $f(0,0) = 0$, vérifie $\partial_2 \partial_1 f(0,0) = -1$ et $\partial_1 \partial_2 f(0,0) = 1$. Elle n'est donc pas de classe C^2 .

Exemple :

Existe-t-il f de classe C^2 sur $\mathbb{R}^2$ telle que $\partial_1 f = y$ et $\partial_2 f = -x$? Si c'était le cas, on aurait $\partial_2 \partial_1 f = 1$ et $\partial_1 \partial_2 f = -1$. Cela contredit le théorème de Schwarz. Donc une telle fonction n'existe pas. Autrement dit, le champ $(y, -x)$ n'est pas un gradient.

IV. Exemples d'équations aux dérivées partielles

Une équation aux dérivées partielles (EDP) relie une fonction inconnue de plusieurs variables à ses dérivées partielles. Le programme demande de savoir résoudre quelques-unes grâce à un changement de variables bien choisi.

1. Principe du changement de variables

Méthode :

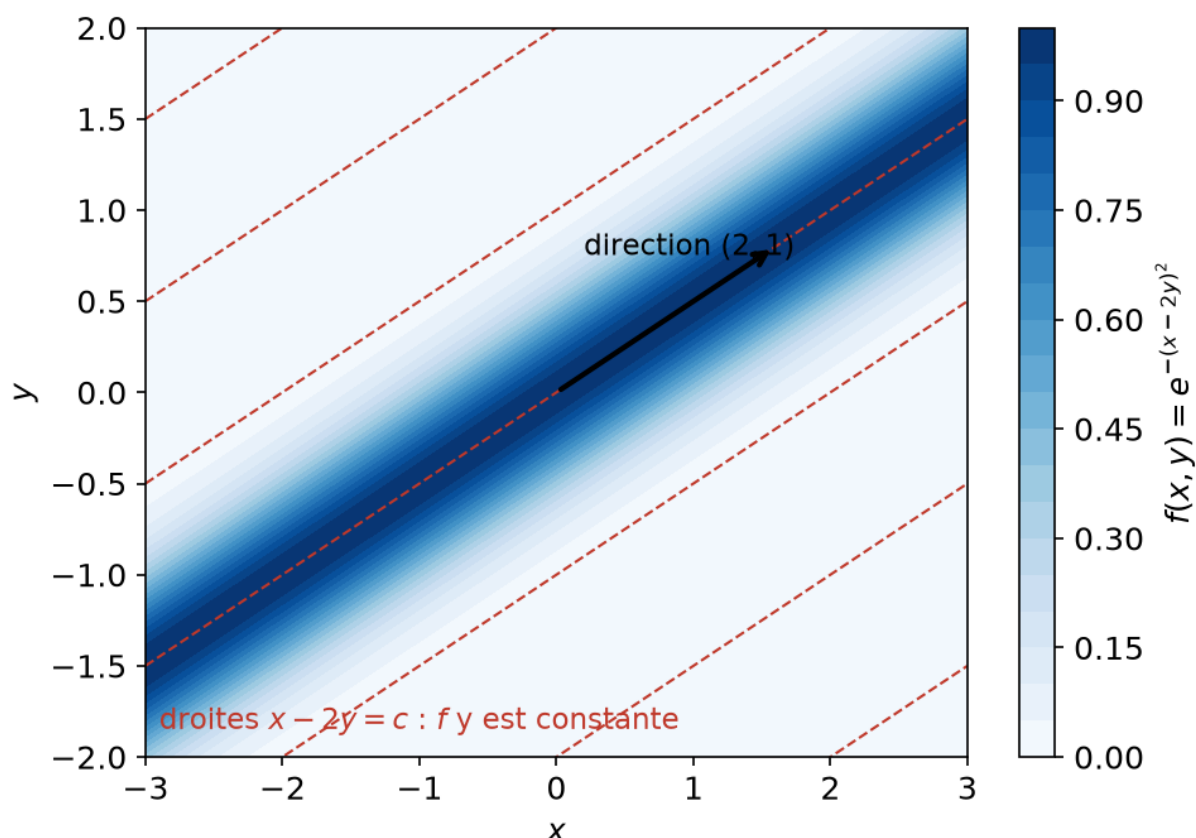
Pour résoudre une EDP sur un ouvert U :

- choisir une bijection $\Phi : V \rightarrow U$, $(u,v) \rightarrow (x,y)$, de classe C^1 ainsi que sa réciproque ;
- poser $g = f \circ \Phi$, de sorte que f est C^1 si et seulement si g l'est ;
- exprimer les dérivées partielles de g par la règle de la chaîne ;
- traduire l'EDP en une équation simple sur g , la résoudre, puis revenir à $f = g \circ \Phi^{-1}$.

Exemple :

Réolvons $2 \partial_x f + \partial_y f = 0$ pour f de classe C^1 sur $\mathbb{R}^2$. Posons $g(u,v) = f(u + 2v, v)$. L'application $(u,v) \rightarrow (u+2v, v)$ est linéaire bijective, de réciproque $(x,y) \rightarrow (x - 2y, y)$. Par la règle de la chaîne, $\partial_v g(u,v) = 2 \partial_x f + \partial_y f = 0$. Ainsi, pour u fixé, $v \rightarrow g(u,v)$ est constante sur l'intervalle $\mathbb{R}$. Donc $g(u,v) = \varphi(u)$, avec φ de classe C^1 . Finalement, les solutions sont les fonctions $f(x,y) = \varphi(x - 2y)$.

La figure ci-dessous représente une telle solution. Comme on le voit, elle est constante sur chaque droite $x - 2y = c$: l'information est « transportée » dans la direction $(2,1)$.



2. Équations classiques

D'autres équations reviennent souvent aux concours. L'équation des ondes $\partial_x^2 f = \partial_y^2 f$ se résout par $u = x + y$, $v = x - y$. Elle devient $\partial_u \partial_v g = 0$, ce qui donne

$f(x, y) = A(x + y) + B(x - y)$. De même, le laplacien $\Delta f = \partial_x^2 f + \partial_y^2 f$ s'étudie en coordonnées polaires. En outre, le théorème de Schwarz sert à chaque calcul de dérivée seconde d'une composée. Ces exemples sont traités dans les exercices.

V. Hessienne et formule de Taylor-Young à l'ordre 2

Définition :

Soit f de classe $\mathcal{C}^2$ sur U et $a \in U$. La **matrice hessienne** de f en a est

$$H_f(a) = (\partial_i \partial_j f(a))_{1 \leq i, j \leq n} \in \mathcal{M}_n(\mathbb{R}).$$

D'après le théorème de Schwarz, c'est une matrice **symétrique**.

Théorème :

Formule de Taylor-Young à l'ordre 2. Si f est de classe $\mathcal{C}^2$ sur U et $a \in U$, alors, quand $h \rightarrow 0$:

$$f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} h^\top H_f(a) h + o(|h|^2).$$

On admet cette formule ; elle s'obtient en appliquant la formule de Taylor à $t \rightarrow f(a+th)$. En dimension 2, on note souvent $r = \partial_x^2 f(a)$, $s = \partial_x \partial_y f(a)$ et $t = \partial_y^2 f(a)$ (notations de Monge). La formule s'écrit alors avec $h = (h_1, h_2)$:

$$f(a+h) = f(a) + \partial_x f(a) h_1 + \partial_y f(a) h_2 + \frac{1}{2}(r h_1^2 + 2 s h_1 h_2 + t h_2^2) + o(|h|^2).$$

Exemple :

Pour $f(x,y) = e^x \cos y$ en $(0,0)$, on trouve $\nabla f(0,0) = (1,0)$, $r = 1$, $s = 0$ et $t = -1$. Donc

$$e^x \cos y = 1 + x + \frac{x^2 - y^2}{2} + o(x^2 + y^2).$$

On retrouve le produit des développements $1 + x + \frac{x^2}{2}$ et $1 - \frac{y^2}{2}$.

VI. Conditions d'extremum au second ordre

En un point critique, le terme linéaire disparaît de la formule de Taylor-Young. Le signe de $f(a+h) - f(a)$ dépend alors de la forme quadratique $h \rightarrow h^\top H_f(a) h$. C'est ici qu'intervient le théorème spectral.

1. Condition nécessaire et condition suffisante

Théorème :

Soit f de classe C^2 sur l'ouvert U et $a \in U$.

- **Condition nécessaire.** Si f admet un minimum local en a , alors $\nabla f(a) = 0$ et $H_f(a) \in \mathcal{S}_n^+(\mathbb{R})$.
- **Condition suffisante.** Si $\nabla f(a) = 0$ et $H_f(a) \in \mathcal{S}_n^{++}(\mathbb{R})$, alors f admet un minimum local strict en a .

Pour un maximum, on remplace $H_f(a)$ par $-H_f(a)$.

Démonstration :

Condition nécessaire. On sait déjà que $\nabla f(a) = 0$. Fixons h . Pour t réel proche de 0, Taylor-Young donne $0 \leq f(a+th) - f(a) = \frac{t^2}{2} h^\top H_f(a) h + o(t^2)$. En divisant par t^2 puis en faisant tendre t vers 0, on obtient $h^\top H_f(a) h \geq 0$.

Condition suffisante. D'après le théorème spectral, $H_f(a)$ est diagonalisable en base orthonormée. Sa plus petite valeur propre α est strictement positive, et $h^\top H_f(a) h \geq \alpha |h|^2$. Écrivons le reste sous la forme $|h|^2 \varepsilon(h)$ avec $\varepsilon(h) \rightarrow 0$. Il existe $r > 0$ tel que $|\varepsilon(h)| \leq \alpha/4$ pour

$|h| < r$. Alors, pour $0 < |h| < r$,

$$f(a+h) - f(a) \geq \frac{\alpha}{2}|h|^2 - \frac{\alpha}{4}|h|^2 = \frac{\alpha}{4}|h|^2 > 0.$$

Corollaire :

Si a est un point critique et si $H_f(a)$ possède une valeur propre strictement positive et une valeur propre strictement négative, alors f n'a pas d'extremum local en a . En effet, f croît le long du premier vecteur propre et décroît le long du second. On parle de point col (ou point selle).

Attention :

Si $H_f(a)$ est positive mais non définie, on ne peut pas conclure. Par exemple, $x^2 + y^4$, $x^2 + y^3$ et $x^2 - y^4$ ont la même hessienne en $(0,0)$. Pourtant, la première y admet un minimum, les deux autres non. Il faut alors étudier directement $f(a+h) - f(a)$.

2. Le cas $n = 2$: trace et déterminant

En dimension 2, les deux valeurs propres λ_1, λ_2 de $H_f(a)$ vérifient $\lambda_1 \lambda_2 = \det H_f(a) = rt - s^2$ et $\lambda_1 + \lambda_2 = \text{tr } H_f(a) = r + t$. On en déduit une règle très pratique.

Propriété :

Soit a un point critique de f de classe C^2 sur un ouvert de $\mathbb{R}^2$.

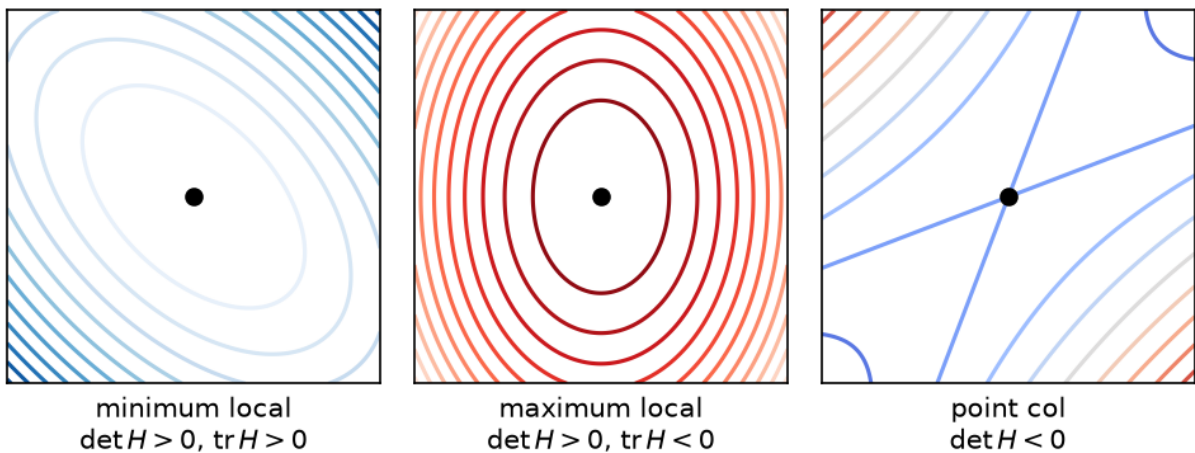
- Si $rt - s^2 > 0$ et $r + t > 0$, alors f admet un minimum local strict en a .

- Si $rt - s^2 > 0$ et $r + t < 0$, alors f admet un maximum local strict en a .
- Si $rt - s^2 < 0$, alors a est un point col.
- Si $rt - s^2 = 0$, on ne peut pas conclure sans étude supplémentaire.

Démonstration :

Si le déterminant est strictement positif, les deux valeurs propres sont non nulles et de même signe, celui de leur somme. Il suffit alors d'appliquer le théorème précédent. Si le déterminant est strictement négatif, les valeurs propres sont de signes contraires, et le corollaire s'applique.

La figure ci-dessous résume les trois situations. Autour d'un minimum ou d'un maximum, les lignes de niveau ressemblent à des ellipses. En revanche, autour d'un point col, elles ressemblent à des hyperboles.



Méthode :

Pour étudier les extremums locaux d'une fonction $\mathcal{C}^2$ sur un ouvert :

- résoudre $\nabla f = 0$ pour obtenir les points critiques ;
- calculer la hessienne en chacun d'eux ;
- en dimension 2, conclure par le signe de $rt - s^2$ et de $r + t$; sinon, chercher le signe des valeurs propres ;
- si la hessienne est dégénérée, étudier f le long de droites ou de courbes ;
- pour un extremum global, compléter par une inégalité ou par un argument de compacité.

Exemple :

Soit $f(x,y) = x^3 + y^3 - 3xy$. Le système $3x^2 - 3y = 0$, $3y^2 - 3x = 0$ donne $y = x^2$ et $x = x^4$. Les points critiques sont donc $(0,0)$ et $(1,1)$. La hessienne vaut $\begin{pmatrix} 6x & -3 \\ -3 & 6y \end{pmatrix}$. En $(0,0)$, $rt - s^2 = -9 < 0$: c'est un point col. En $(1,1)$, $rt - s^2 = 27 > 0$ et $r + t = 12 > 0$: c'est un minimum local strict, égal à -1 . Ce minimum n'est pas global, car $f(x,0) = x^3$ tend vers $-\infty$.

Ce qu'il faut retenir

- En un point intérieur, un extremum local d'une fonction différentiable est un point critique ; la réciproque est fausse.
- Sur un compact, une fonction continue atteint ses bornes : on compare les points critiques intérieurs et les valeurs au bord.
- Si $f|_X$ a un extremum en a , alors $\nabla f(a)$ est orthogonal à $T_a X$.
- Sous la contrainte $g = 0$ avec $\nabla g(a) \neq 0$, un extremum lié vérifie $\nabla f(a) = \lambda \nabla g(a)$.
- Une fonction est C^k si ses dérivées partielles d'ordre k existent et sont continues.
- Théorème de Schwarz : pour f de classe C^2 , $\partial_i \partial_j f = \partial_j \partial_i f$, donc la hessienne est

symétrique.

- Taylor-Young : $f(a+h) = f(a) + \langle \nabla f(a), h \rangle + \frac{1}{2} h^\top H_f(a) h + o(|h|^2)$.
- En un point critique, hessienne définie positive donne un minimum local strict ; un minimum local impose une hessienne positive.
- Pour $n = 2$: $rt - s^2 > 0$ donne un extremum (minimum si $r + t > 0$), $rt - s^2 < 0$ un point col, $rt - s^2 = 0$ aucune conclusion.
- Une EDP se résout par un changement de variables C^1 bijectif qui la ramène à $\partial_v g = 0$ ou $\partial_u \partial_v g = 0$.

Questions fréquentes sur optimisation

Un point critique est-il toujours un extremum local ?

Non. La condition $\nabla f(a) = 0$ est seulement nécessaire, et seulement en un point intérieur. Par exemple, $(0,0)$ est critique pour $f(x,y) = x^2 - y^2$, mais c'est un point col : f prend des valeurs positives et négatives au voisinage.

Que faire quand le déterminant de la hessienne est nul ?

La hessienne ne permet alors pas de conclure. On étudie directement le signe de $f(a+h) - f(a)$, par exemple le long de droites ou de courbes bien choisies. Ainsi, $x^2 + y^3$, $x^2 + y^4$ et $x^2 - y^4$ ont la même hessienne en $(0,0)$ mais des comportements différents.

Comment traiter un extremum sous une contrainte $g = 0$?

Si f restreinte à $X = \{g = 0\}$ admet un extremum local en a et si $\nabla g(a) \neq 0$, alors $\nabla f(a)$ est colinéaire à $\nabla g(a)$. On résout donc $\nabla f(a) = \lambda \nabla g(a)$ avec $g(a) = 0$. Ensuite, on compare les valeurs obtenues, souvent grâce à la compacité de X .

Pourquoi le théorème de Schwarz demande-t-il une fonction de classe C^2 ?

Sans cette hypothèse, les dérivées croisées peuvent différer. La fonction $xy(x^2 - y^2)/(x^2 + y^2)$, prolongée par 0, a des dérivées croisées égales à -1 et 1 en l'origine. Elle est de classe C^1 mais pas C^2 .

Pour aller plus loin en maths spé

- Les énoncés : [exercices de maths spé sur optimisation](#)
- À maîtriser avant : [Calcul différentiel dans les espaces normés](#), [Endomorphismes autoadjoints et théorème spectral](#)
- Chapitre précédent : [Calcul différentiel dans les espaces normés](#)
- Tester vos connaissances : [QCM de maths spé par chapitre](#)
- Le sommaire : [tous les chapitres de maths spé](#) et les [chapitres de maths sup](#)