Fonctions de plusieurs variables

Dérivées Partielles

Edouard Marchais[a]

[a] *EPITA, 14-16 Rue Voltaire, 94270 Le Kremlin-Bicêtre, France.*

E-mail: edouard.marchais@epita.fr

Abstract: La notion de dérivée partielle est présentée avec une emphase sur les applications en deux et trois dimensions. Le point de vue géométrique est d'abord abordé à travers les notions de surfaces et de lignes de niveau. Ensuite les concepts de limites et de continuité pour des fonctions de plusieurs variables sont discutés, en évitant une présentation trop formelle, puis l'idée de dérivée partielle (de premier et second ordre) est présentée. On applique par la suite cet outil à l'étude des plans tangents à des surfaces et au calcul d'approximations linéaires. Les notions de dérivées directionnelles, de gradient et de dérivée composée (*chain rule*) sont également présentées en détails avec de nombreux exemples. Enfin, le problème de minimisation et maximisation d'une fonction de plusieurs variables est abordé avec, encore une fois, de nombreux exemples d'application. Des exercices accompagnant chaque thématique sont fournis en annexe.

Table des Matières

1 Introduction **2**

2 Surfaces et lignes de niveau **3**

3 Limites et continuité **7**
3.1 Limite d'une fonction à deux variables 7
3.2 Points intérieurs et points de bord 12
3.3 Continuité des fonctions de deux variables 15
3.4 Fonctions de trois variables ou plus 16

4 La dérivée partielle **18**
4.1 Définitions et exemples 18
4.2 La dérivée seconde 21

5 Plans tangents et approximations linéaires **24**
5.1 Différentielles 27
5.2 Approximation linéaire 28
5.3 Méthode de Newton pour deux équations 31

6 Dérivées directionnelles et gradients **33**
6.1 Calcul de la dérivée directionnelle 34
6.2 Le vecteur gradient 36
6.3 Dérivés le long de chemins courbes 38
6.4 Gradients sans coordonnées 40

7 La dérivée composée ou « chain rule » **41**
7.1 Les dérivées de $f(g(x, y))$ 42
7.2 Les dérivées de $f(x(t), y(t))$ 42
7.3 Dérivées secondes 44
7.4 Les dérivées de $f(x(t, u), y(t, u))$ 44
7.5 Un paradoxe 45
7.6 Variables non indépendantes 46

8 Maxima, minima et point-cols **48**
8.1 Point stationnaire et plan tangent 48
8.2 Le problème de Steiner 50
8.3 Extremum sur la frontière 52
8.4 Maximum, minimum et point-col 53
8.5 La série de Taylor 56

A Contraintes et multiplicateurs de Lagrange **59**
A.1 Maximum et minimum avec deux contraintes 61
A.2 Contraintes d'inégalité 62

B Normes **65**

C Série de Taylor pour une fonction à n variables **70**

D Exercices **72**
D.1 Surfaces et lignes de niveau 72
D.2 Limite et continuité 73
D.3 La dérivée partielle 74
D.4 Plans tangents et approximations linéaires 76
D.5 Dérivées directionnelles et gradients 77
D.6 La dérivée composée ou « chain rule » 78
D.7 Maxima, minima et point-cols 79

1 Introduction

Position du problème. De manière peu surprenante, la notion de dérivée partielle est centrale dans l'étude de fonctions de plusieurs variables. Pour se convaincre de cela il suffit que considérer le rôle fondamental, tant du point de vue historique que mathématique, que joue la notion de dérivée pour une fonction d'une seule variable, avec laquelle nous sommes déjà familier. Considérons par exemple une fonction f dépendante d'un seule variable réelle x. Une question naturelle consiste à savoir comment varie f lorsque sa variable x varie également? Nous connaissons la réponse à cette question. Maintenant nous souhaitons considérer des fonctions du type $f(x, y)$ ou $F(x, y, z)$ avec deux ou trois variables, ou plus, variant indépendamment. De manière analogue au cas à une variable, on souhaite à présent relier les petites (infinitésimales) variations des paramètres x et y (et z), que l'on note Δx et Δy (et Δz), à une petite variation de f (ou F), notée Δf (ou ΔF).

Démarche de solution. Ce problème se résout de manière totalement analogue à la construction de la dérivée pour une fonction à une seule variable, c'est-à-dire par passage à la limite. En effet, nous savons que df/dx est la limite du taux d'accroissement $\Delta f/\Delta x$ lorsque Δx devient infiniment petit. Dans le cadre multidimensionnel, ce taux d'accroissement est toujours associé à la pente d'une tangente mais il s'agit plutôt d'une tangente à une *surface*. Notre but est donc de donner un sens à la notion dérivée afin d'étudier les variations de fonctions telles que

$$f(x, y) = x^2 - y^2 \qquad \text{ou} \qquad f(x, y) = \sqrt{x^2 + y^2} \qquad \text{ou} \qquad f(x, y, z) = 2x + 3y + 4z$$

On peut aisément se convaincre que ces fonctions possèdent des graphiques ainsi que des dérivées, elles doivent donc admettre des tangentes associées à ces dérivées que nous cherchons à évaluer.

Résumé. Notre sujet est donc le calcul différentiel pour les fonctions de plusieurs variables, c'est-à-dire l'étude des variations de fonctions à plusieurs paramètres en fonction des (petites) variations de ceux-ci. Notre présentation peut être synthétiser en six point essentiels présenté dans le tableau ci-dessous. Dans celui-ci on a mis en correspondance les opérations habituelles concernant le calcul différentiel pour des fonctions d'une variable avec leurs contreparties pour une fonction à deux variables. Il peut servir de guide pour se repérer dans l'avancement de notre présentation.

Courbe $y = f(x)$	vs.	Surface $z = f(x, y)$
$\frac{df}{dx}$	devient deux dérivées partielles	$\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$
$\frac{d^2 f}{dx^2}$	devient quatre dérivées secondes	$\frac{\partial^2 f}{\partial x^2}, \frac{\partial^2 f}{\partial x \partial y}, \frac{\partial^2 f}{\partial y \partial x}, \frac{\partial^2 f}{\partial y^2}$
$\Delta f \approx \frac{df}{dx}\Delta x$	devient l'approximation linéaire	$\Delta f \approx \frac{df}{dx}\Delta x + \frac{df}{dy}\Delta y$
La tangente	devient le plan tangent	$z - z_0 = \frac{\partial f}{\partial x}(x - x_0) + \frac{\partial f}{\partial y}(y - y_0)$
$\frac{dy}{dt} = \frac{\partial y}{\partial x}\frac{dx}{dt}$	devient la « chain rule »	$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}$
$\frac{df}{dx} = 0$	devient deux équations maximum-minimum	$\frac{\partial f}{\partial x} = 0$ et $\frac{\partial f}{\partial y} = 0$

2 Surfaces et lignes de niveau

Généralités. Le graphique de $y = f(x)$ est une courbe dans le plan xy. Il y a deux variables, x est indépendant et libre, y dépend de x. Au-dessus de l'axe des x se trouve, par exemple, le point (x, y) sur la courbe. La courbe peut être affichée sur une page imprimée en deux dimensions. Le graphe de $z = f(x, y)$ est une *surface dans l'espace* xyz. Il y a *trois variables*, x et y sont indépendantes, z dépend de celles-ci. Au-dessus du point (x, y) dans le *plan de base* (*base plane*) se trouve le point (x, y, z) sur la surface (**Figure 1 (gauche)**). Puisque la page imprimée reste bidimensionnelle, on peut ombrer ou colorer ou projeter la surface. Si cela est effectué avec soin, on doit être alors en mesure de convertir ces images bidimensionnelles en une compréhension tridimensionnelle. Afin de mener cette conversion correctement notre cerveau est en fait déjà en train de manipuler et évaluer, implicitement, les **dérivées partielles** des fonctions associées à ces surfaces. Cette section utilise des exemples et des figures pour illustrer des **surfaces** et leurs **lignes de niveau**. La section suivante sera consacré aux notions de limite et continuité en plusieurs dimensions.

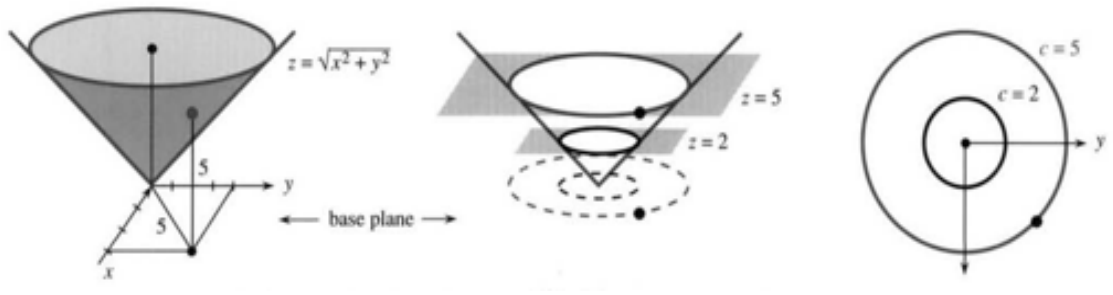


Figure 1: La surface pour $z = f(x, y) = \sqrt{x^2 + y^2}$ est un cône. Les lignes de niveau sont des cercles.

Exemples 1. On s'intéresse à la surface et aux les lignes de niveau associées à l'équation

$$z = f(x, y) = \sqrt{x^2 + y^2}$$

- La surface est un **cône** et la raison est que la fonction $\sqrt{x^2 + y^2}$ correspond à la distance dans le plan de base du point $(0, 0)$ au point (x, y). Lorsque nous nous éloignons d'une distance 5 de l'origine, dans le plan de base, nous montons d'un hauteur 5 également jusqu'à la surface. Le cône monte avec une pente de 1. La distance jusqu'à (x, y) est égale à la distance jusqu'à z (c'est un cône faisant un angle de 45°).
- Les lignes de niveau sont des **cercles**. À la hauteur 5, le cône contient un cercle de points, qui sont tous au même « niveau » en surface. Le plan $z = 5$ rencontre la surface $z = \sqrt{x^2 + y^2}$ en ces points (**Figure 1 (milieu et droite)**). Le cercle en dessous (dans le plan de base) est la ligne de niveau.

Définition. Une **ligne (ou courbe) de niveau (level curve)** de $z = f(x, y)$ contient tous les points (x, y) qui partagent la même valeur

$$c = f(x, y)$$

Au-dessus de ces points, la surface est à la hauteur $z = c$. Les lignes de niveau sont parfois appelées **lignes de contour**. Il existe différentes lignes de niveau pour différents c. Pour voir la courbe

pour $c = 2$, coupez la surface avec le plan horizontal $z = 2$. Le plan rencontre la surface au-dessus des points où $f(x, y) = 2$. *La ligne de niveau dans le plan de base a pour équation* $f(x, y) = 2$. Au-dessus se trouvent tous les points au « niveau 2 » ou « niveau c » en surface. Chaque courbe $f(x, y) = c$ est labellisée par sa constante c. Ceci produit un **carte isoligne (contour map)**. Pour le cône, les lignes de niveau sont données par $\sqrt{x^2 + y^2} = c$, et la carte isoligne est constituée de cercles de rayon c.

Exemples 2.

(1) Les lignes de niveau associée à l'équation

$$z = f(x, y) = x^2 + y^2$$

sont des **cercles fixes**. Mais la surface n'est pas un cône (elle est courbe). Le cercle de rayon 3 est la ligne de niveau $x^2 + y^2 = 9$. Sur la surface ci-dessus, la hauteur est de 9.

(2) La surface associée à la fonction linéaire

$$f(x, y) = 2x + y$$

est un **plan**. Ses lignes de niveau sont des lignes droites. La surface $z = 2x + y$ rencontre le plan $z = c$ dans la ligne $2x + y = c$. Cette ligne est au-dessus du plan de base lorsque c est positif, et en dessous lorsque c est négatif. Les lignes de contour sont *dans* le plan de base. Sur la **Figure 2 (milieu)** ces lignes parallèles sont labellisées en fonction de leur hauteur dans la surface.

(3) Si les lignes de niveau sont toutes droites alors celles-ci **ne sont pas** automatiquement parallèles. En effet la surface $z = y/x$ a des lignes de niveau $y/x = c$. Ces lignes $y = cx$ tournent autour de l'origine, alors que la surface grimpe comme un toboggan en spirale.

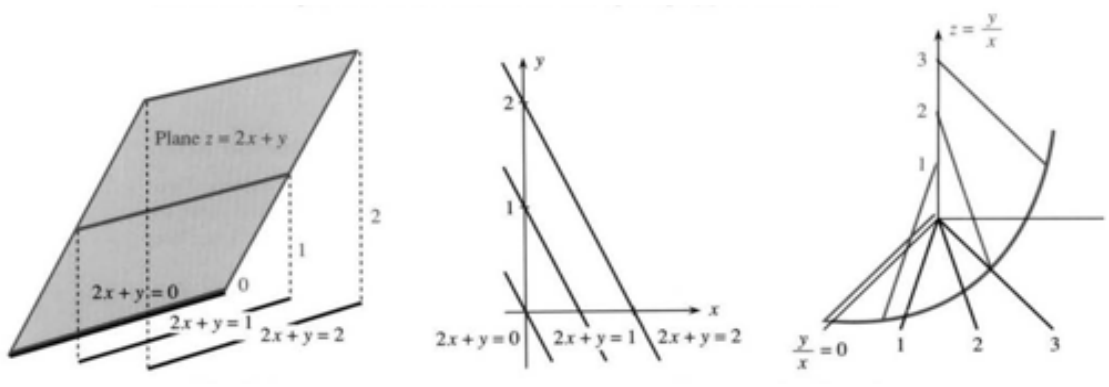


Figure 2: Un plan a des lignes de niveau parallèles. La glissière en spirale $z = y/x$ a des lignes $y/x = c$.

Exemples 3.

(1) L'un des exemples les plus courants de carte isoligne est une *carte topographique* comme celle illustrée à la **Figure 3**. Elle donne l'altitude de la région et elle permet d'avoir une vue d'ensemble du terrain : où sont les montagnes, où sont les zones plates. Ces cartes topographiques sont fréquemment colorées en vert aux altitudes inférieures et en marron, rouge ou blanc aux altitudes les plus élevées.

(2) À partir d'une carte isoligne, comment trouver le point le plus élevé ?
Les lignes de niveau forment des *boucles* autour du point maximum. À mesure que c augmente, les boucles se resserrent. De même, les courbes se resserrent jusqu'au point le plus bas lorsque c diminue.

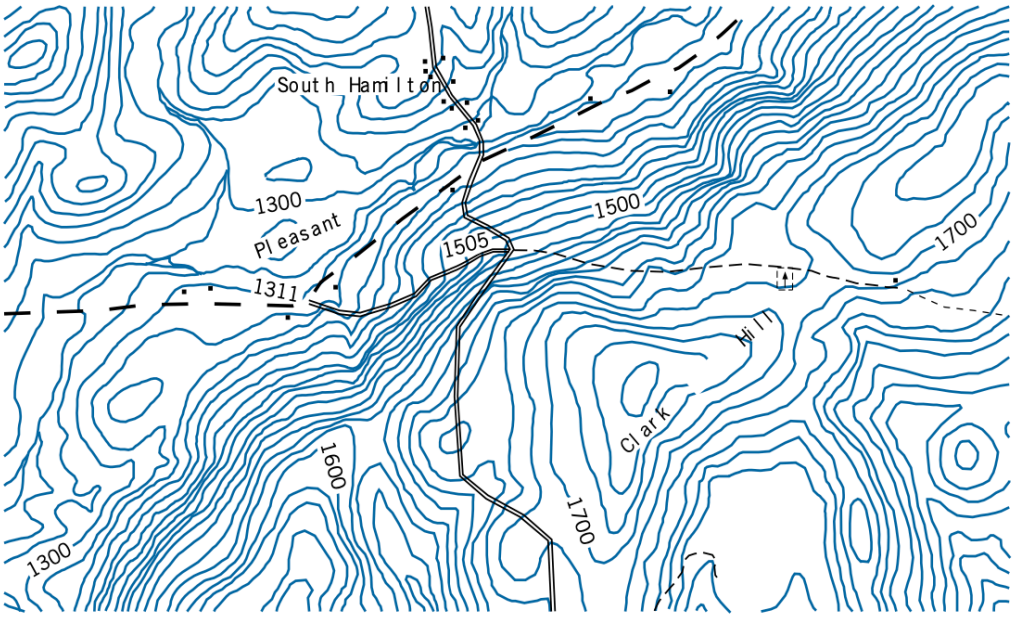


Figure 3: Plan topologique de la région située autour de la ville de South Hamilton, dans l'état de New York (US).

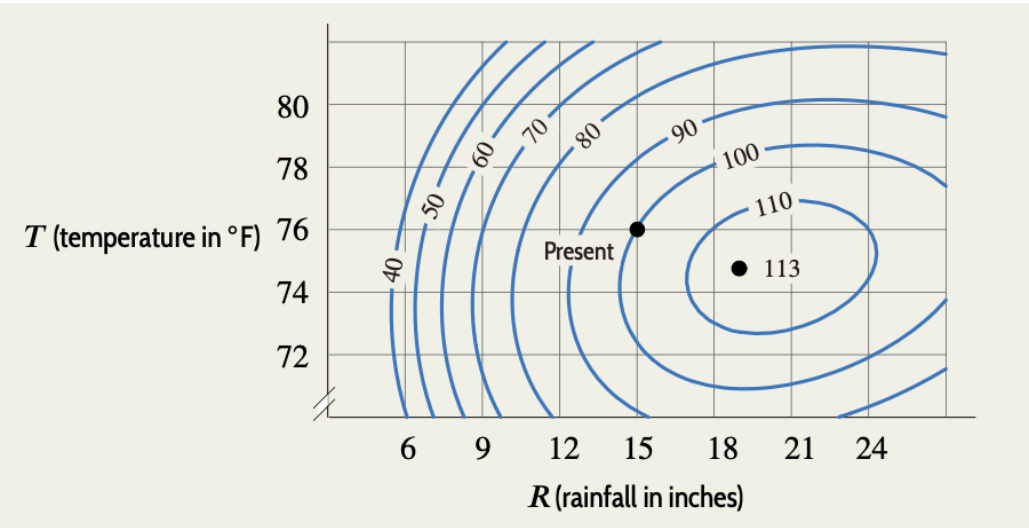


Figure 4: Production de maïs, C, en fonction des précipitations (*rainfall*) en pouces (*inches*) et de la température (en degrés Fahrenheit).

Exemples 4.

(1) Les cartes isolignes peuvent également donner des informations sur une fonction de deux variables sans référence à une surface. Considérons l'effet des conditions météorologiques sur la production de maïs aux États-Unis. La **Figure 4** donne la production de maïs $C = f(R, T)$ en fonction des précipitations totales, R, en pouces (*inches*), et de la température moyenne, T, en degrés Fahrenheit, pendant la saison de croissance. Au temps présent, $R = 15$ pouces et $T = 76°F$. La production est mesurée en pourcentage de la production actuelle; ainsi, le contour passant par $R = 15, T = 76$, a la valeur 100, c'est-à-dire $C = f(15, 76) = 100$.

(2) Quelle direction dans le plan de base est en montée sur la surface ?
De manière générale, les directions des pentes les plus fortes (descendantes ou ascendantes) sont orthogonales aux lignes de niveau. Nous reviendrons sur cette question importante par la suite.

Exemples 5. En économie, la fonction x^2y est qualifiée de *fonction d'utilité* (*utility function*) et sa ligne de niveau associée $x^2y = c$ est appelée *courbe d'indifférence* (*indifference curve*). La fonction d'utilité x^2y varie en fonction du nombre x d'heures d'éveil et y d'heures d'endormissement. Par exemple deux heures d'éveil et quinze minutes de sommeil correspondent à la valeur $f = (2^2)(1/4) = 1$. Ce qui est équivalent, du point de vue mathématique, à une heure de chaque, c'est-à-dire $f = (1^2)(1) = 1$. Ces deux points se trouvent sur la même ligne de niveau dans la **Figure 5 (gauche)**. La courbe d'indifférence est « convexe ». Afin de faire un choix raisonnable *on préférera donc faire la moyenne entre deux points.* Dans ce cas on note que la ligne entre deux points est situé au-dessus de la courbe et donc le « point moyen » se trouve sur une ligne de niveau supérieure. La **Figure 5 (milieu)** montre un cas extrême. Les lignes de niveau sont des droites $4x + y = c$. Quatre « quarter » peuvent être librement échangés contre un dollar. La valeur est $f = 4x+y$ dollars. **Figure 5 (droite)** montre l'autre extrême. Les chaussures gauches ou droites supplémentaires sont inutiles. La valeur (ou l'utilité) est donc le *plus petit* nombre entre x et y. Cela permet de *compter* les paires de chaussures.

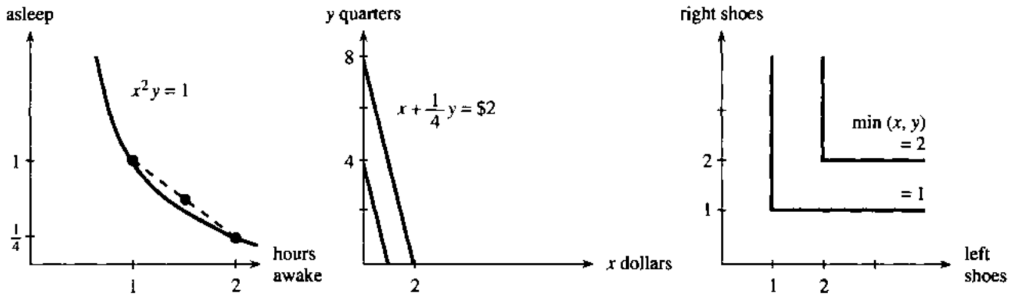


Figure 5: Fonctions d'utilité x^y, $4x + y$, $\min(x, y)$. Substitution convexe, droite, compléments.

3 Limites et continuité

Introduction. Nous venons de voir des exemples concrets de fonctions à plusieurs variables avec leurs représentations graphiques. À présent nous allons aborder la notion de limite pour une fonction de plus d'une variable, et ce que cela signifie pour une telle fonction d'être continue en un point de son domaine. Dans ce cadre nous rencontrerons des situations qui ne se produisent tout simplement pas avec les fonctions d'une seule variable.

3.1 Limite d'une fonction à deux variables

Rappel. On rappelle d'abord la définition d'une limite d'une fonction à une variable. On considère une fonction $f(x)$ définie pour tout $x \neq a$ dans un intervalle ouvert contenant a. Soit L un nombre réel. Dans ce cas on a

$$\lim_{x \to a} f(x) = L$$

si pour chaque $\varepsilon > 0$, il existe un $\delta > 0$, tel que si $0 < |x - a| < \delta$ pour tout x dans le domaine de f, entraînant que

$$|f(x) - L| < \varepsilon$$

Avant de pouvoir adapter cette définition pour définir une limite d'une fonction de deux variables, nous devons d'abord voir comment étendre l'idée d'un intervalle ouvert dans le cas d'une variable à un intervalle ouvert dans le cas de deux variables.

Distance et métrique. Cette notion d'intervalle fait en effet intervenir, implicitement, la notion de *distance* entre les nombres réels x et a, elle-même définie grâce à la fonction *valeur absolue*. Cependant la notion de distance ou *métrique*, est un concept plus général. En effet, on appelle **distance** ou **métrique** sur un ensemble $\mathcal{E}$ (correspondant le plus souvent à $\mathbb{R}$) toute fonction $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ associant à la paire d'élément $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{E}$ le nombre réel $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y})$ et possédant les propriétés suivantes.

(1) Si $\boldsymbol{x} \neq \boldsymbol{y}$ alors $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) > 0$ et $\mathcal{D}(\boldsymbol{x}, \boldsymbol{x}) = 0$

(2) $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) = \mathcal{D}(\boldsymbol{y}, \boldsymbol{x})$

(3) $\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) \leq \mathcal{D}(\boldsymbol{x}, \boldsymbol{z}) + \mathcal{D}(\boldsymbol{z}, \boldsymbol{y})$ pour tout $\boldsymbol{z} \in \mathcal{E}$

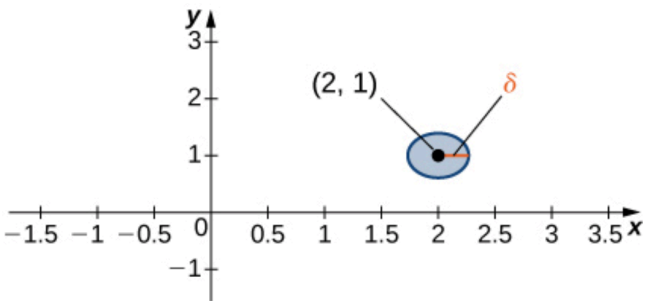


Figure 6: Exemple d'un disque centré autour du point $\boldsymbol{x} = (2, 1)$.

Notation et généralisation. Naturellement, un ensemble muni d'une métrique est qualifié d'*espace métrique*, c'est-à-dire un espace sur lequel la notion de distance a un sens. Dans notre cas, cet ensemble correspondra à $\mathbb{R}^n$ avec n un entier fini ($n = 2, 3$ le plus souvent) et les éléments $\boldsymbol{x}$ et $\boldsymbol{y}$ correspondent alors à des *points* dans cet espace que l'on notera plutôt en gras de sorte que

$$\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \qquad \text{et} \qquad \boldsymbol{y} = (y_1, y_2, \ldots, y_n)$$

avec $x_i, y_i \in \mathbb{R}$ pour $i = 1, 2, \ldots, n$. Évidemment, il existe une infinité de fonctions pouvant remplir les conditions définissant une métrique. Il s'agit donc ici d'en choisir une qui soit la plus naturelle ou la plus commode, en fonction du type de problème. Étant donné que $\mathbb{R}^n$ est un *espace vectoriel de dimension fini* il est naturel d'utiliser la *norme euclidienne*, ou norme ℓ^2, de sorte que

$$\mathcal{D}(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

D'autre normes sont également disponibles et nous renvoyons à l'**Appendice B** pour une discussion plus poussée sur ce sujet. Nous pouvons à présent désormais généraliser la notion d'intervalle ouvert au plan $\mathbb{R}^2$ grâce à la notion de *disque (ouvert)*.

Notion de disque. Un **disque (ouvert)** centré au point $\boldsymbol{a} = (a, b) \in \mathbb{R}^2$ et de rayon δ correspond à l'ensemble des points $\boldsymbol{x} = (x, y)$ défini par

$$\left\{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x} - \boldsymbol{a}\| < \delta\right\}$$

comme indiqué sur la **Figure 6**. L'idée d'un disque est nécessaire à la définition de la limite d'une fonction à deux variables. En effet, si le rayon δ est petit, alors tous les points $\boldsymbol{x} = (x, y)$ du disque sont proches du point $\boldsymbol{a} = (a, b)$. La situation est en fait analogue au cas à une dimension où x est proche de a dans la définition d'une limite d'une fonction à une variable réelle. En une dimension, on exprime cette restriction comme $a - \delta < x < a + \delta$ ou encore, grâce à la fonction valeur absolue, par $|x - a| < \delta$. En deux dimensions, on utilisera dorénavant la notion de disque.

Limite en deux dimensions. Soit f une fonction des deux variables (réelles) x et y. La limite de $f(x, y)$ lorsque (x, y) approche (a, b) est L et s'écrit

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

si pour chaque $\varepsilon > 0$ il existe un nombre $\delta > 0$ assez petit tel que pour tous les points (x, y) dans un disque de rayon δ autour de (a, b), sauf peut-être pour (a, b) lui-même, la différence entre $f(x, y)$ et L soit plus petite que la quantité ε (**Figure 7**). En utilisant des symboles puis en notant $\boldsymbol{x} = (x, y)$ et $\boldsymbol{a} = (a, b)$, on peut écrire que: Pour tout $\varepsilon > 0$, il existe un nombre $\delta > 0$ tel que

$$|f(x, y) - L| < \varepsilon$$

pour tout $\boldsymbol{x} \in \mathbb{R}^2$ vérifiant

$$0 < \|\boldsymbol{x} - \boldsymbol{a}\| < \delta$$

Propriétés. Similairement au cas à une seule variable, il peut s'avérer compliqué ou fastidieux de montrer qu'une fonction à deux variables possède une limite en un point donné à partir de la définition. Pour remédier à cela, un certain nombre de résultats intermédiaires, très utiles dans la pratique, sont donnés ci-dessous, afin d'éviter d'avoir recours à la définition fondamentale dans les cas les plus courants. On considère ici des fonctions $f(x, y)$ et $g(x, y)$ définies pour tout $(x, y) \neq (a, b)$

dans un voisinage autour de (a, b), et on assume que ce voisinage soit complètement inclus dans le domaine de f. Supposons également que L et M sont des nombres réels tels que

$$\lim_{(x,y)\to(a,b)} f(x,y) = L \qquad \textit{et} \qquad \lim_{(x,y)\to(a,b)} g(x,y) = M$$

et soit c une constante. Dans ce cas, chacune des « lois » suivantes est vraie.

(1) Constante : $\displaystyle\lim_{(x,y)\to(a,b)} c = c$

(2) Identité : $\displaystyle\lim_{(x,y)\to(a,b)} x = a$ et $\displaystyle\lim_{(x,y)\to(a,b)} y = b$

(3) Somme : $\displaystyle\lim_{(x,y)\to(a,b)} f(x,y) + g(x,y) = L + M$

(4) Différence : $\displaystyle\lim_{(x,y)\to(a,b)} f(x,y) - g(x,y) = L - M$

(5) Multiplication : $\displaystyle\lim_{(x,y)\to(a,b)} c\, f(x,y) = c\, L$

(6) Produits : $\displaystyle\lim_{(x,y)\to(a,b)} f(x,y)\, g(x,y) = L\, M$

(7) Quotient : $\displaystyle\lim_{(x,y)\to(a,b)} \frac{f(x,y)}{g(x,y)} = \frac{L}{M}$ pour $M \neq 0$

(8) Puissance : $\displaystyle\lim_{(x,y)\to(a,b)} (f(x,y))^n = L^n$

pour tout entier positif n.

(9) Racine : $\displaystyle\lim_{(x,y)\to(a,b)} \sqrt[n]{f(x,y)} = \sqrt[n]{L}$

pour tout L si n est impair et positif, et pour $L \geq 0$ si n est pair et positif.

Exemples 6.

(1) On cherche à déterminer la limite suivante

$$\lim_{(x,y)\to(2,-1)} (x^2 - 2xy + 3y^2 - 4x + 3y - 6)$$

Pour ce cas on utilise d'abord les lois **(3)** et **(4)** pour séparer les termes

$$\begin{aligned}\lim_{(x,y)\to(2,-1)} (x^2 - 2xy + 3y^2 - 4x + 3y - 6) = &\lim_{(x,y)\to(2,-1)} x^2 - \lim_{(x,y)\to(2,-1)} 2xy + \lim_{(x,y)\to(2,-1)} 3y^2 \\ &- \lim_{(x,y)\to(2,-1)} 4x + \lim_{(x,y)\to(2,-1)} 3y - \lim_{(x,y)\to(2,-1)} 6\end{aligned}$$

Ensuite, on utilise la loi **(5)** de la deuxième à la cinquième limite

$$\begin{aligned}= &\left(\lim_{(x,y)\to(2,-1)} x^2\right) - 2\left(\lim_{(x,y)\to(2,-1)} xy\right) + 3\left(\lim_{(x,y)\to(2,-1)} y^2\right) \\ &-4\left(\lim_{(x,y)\to(2,-1)} x\right) + 3\left(\lim_{(x,y)\to(2,-1)} y\right) - \left(\lim_{(x,y)\to(2,-1)} 6\right)\end{aligned}$$

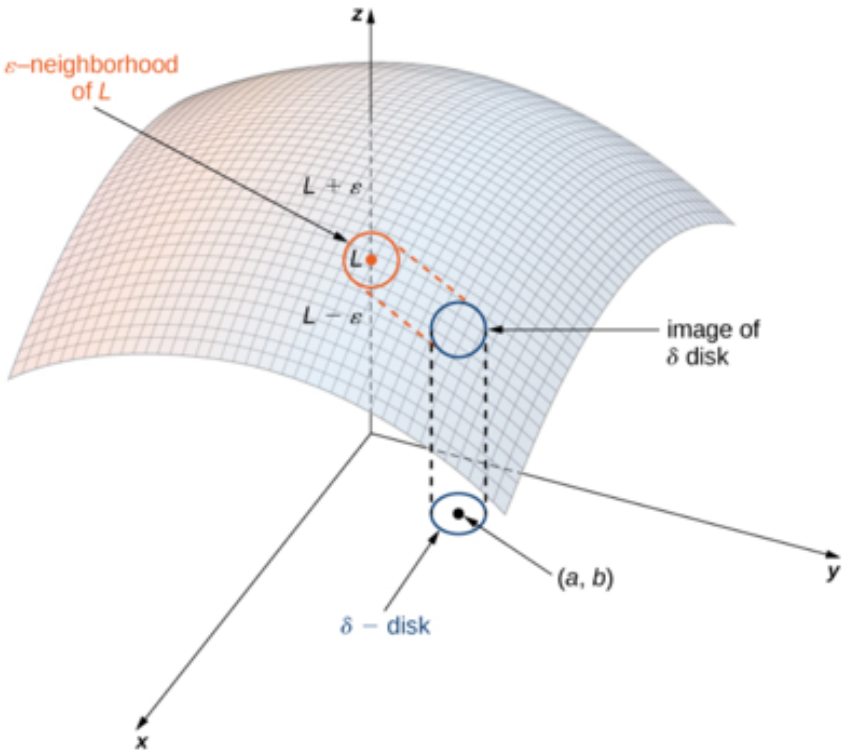


Figure 7: La limite d'une fonction impliquant deux variables requiert que $f(x, y)$ soit dans un voisinage (neighborhood) ε de L pour tout point $\boldsymbol{x} = (x, y)$ situé à une distance inférieure à δ du point $\boldsymbol{a} = (a, b)$. Plus la valeur de ε est petite, plus la valeur de δ est petite.

L'utilisation de **(8)** sur les première et troisième limites et de **(6)** sur la seconde limite donne

$$= \left(\lim_{(x,y)\to(2,-1)} x\right)^2 - 2\left(\lim_{(x,y)\to(2,-1)} x\right)\left(\lim_{(x,y)\to(2,-1)} y\right) + 3\left(\lim_{(x,y)\to(2,-1)} y\right)^2$$
$$-4\left(\lim_{(x,y)\to(2,-1)} x\right) + 3\left(\lim_{(x,y)\to(2,-1)} y\right) - \left(\lim_{(x,y)\to(2,-1)} 6\right)$$

Enfin, on utilise **(2)** sur les six premières limites et **(1)** sur la dernière limite

$$\lim_{(x,y)\to(2,-1)} (x^2 - 2xy + 3y^2 - 4x + 3y - 6) = (2)^2 - 2(2)(-1) + 3(-1)^2 - 4(2) + 3(-1) - 6 = -6$$

(2) On cherche à déterminer la limite suivante

$$\lim_{(x,y)\to(2,-1)} \frac{2x + 3y}{4x - 3y}$$

Ici, avant d'appliquer **(7)**, nous devons vérifier que la limite du dénominateur est non nulle. En utilisant **(4)**, **(5)** et **(2)** on obtient

$$\lim_{(x,y)\to(2,-1)} (4x - 3y) = \left(\lim_{(x,y)\to(2,-1)} 4x\right) + \left(\lim_{(x,y)\to(2,-1)} -3y\right)$$
$$= 4\left(\lim_{(x,y)\to(2,-1)} x\right) - 3\left(\lim_{(x,y)\to(2,-1)} y\right) = 4(2) - 3(-1) = 11$$

Puisque la limite du dénominateur est différente de zéro, la loi **(7)** s'applique. Nous calculons maintenant la limite du numérateur en utilisant **(4)**, **(5)** et **(2)**. Cela donne

$$\begin{aligned}\lim_{(x,y)\to(2,-1)}(2x+3y) &= \left(\lim_{(x,y)\to(2,-1)} 2x\right) + \left(\lim_{(x,y)\to(2,-1)} 3y\right) \\ &= 2\left(\lim_{(x,y)\to(2,-1)} x\right) + 3\left(\lim_{(x,y)\to(2,-1)} y\right) = 2(2)+3(-1) = 1\end{aligned}$$

Par conséquent, selon la loi **(7)**, nous obtenons finalement

$$\lim_{(x,y)\to(2,-1)} \frac{2x+3y}{4x-3y} = \frac{\lim_{(x,y)\to(2,-1)}(2x+3y)}{\lim_{(x,y)\to(2,-1)}(4x-3y)} = \frac{1}{11}$$

Puisque nous prenons la limite d'une fonction à deux variables, le point (a, b) appartenant en effet à $\mathbb{R}^2$, il est donc possible d'approcher ce point à partir d'un nombre infini de directions. Ainsi, lors du calcul d'une limite, la réponse varie parfois en fonction du chemin emprunté vers (a, b). Si tel est le cas, *la limite n'existe pas*. En d'autres termes, *la limite doit être unique*, quel que soit le chemin emprunté.

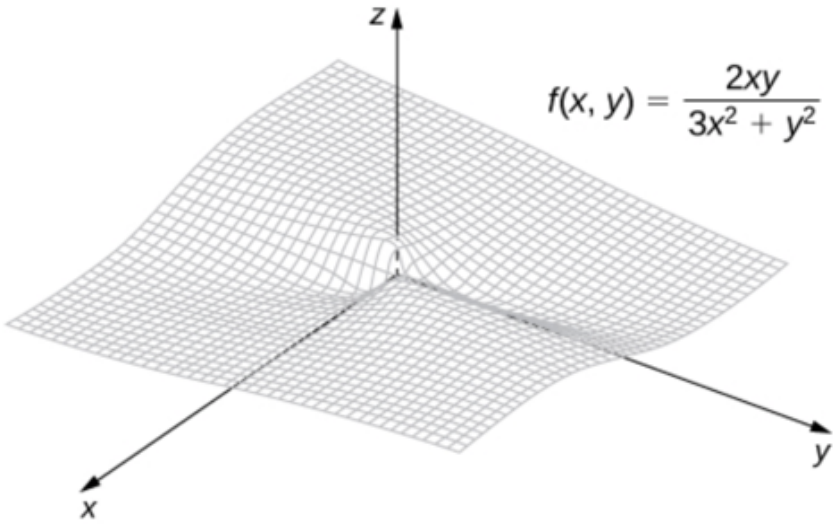


Figure 8: Graphique de la fonction $f(x, y) = (2xy)/(3x^2 + y^2)$. Le long de la ligne $y = 0$, la fonction est égale à zéro; le long de la ligne $y = x$, la fonction est égale à $1/2$.

Exemples 7.

(1) On cherche à déterminer la limite suivante

$$\lim_{(x,y)\to(0,0)} \frac{2xy}{3x^2+y^2}$$

Le domaine de la fonction $f(x, y) = 2xy/(3x^2 + y^2)$ se compose de tous les points du plan xy sauf pour le point $(0, 0)$(**Figure 8**). Pour montrer que la limite n'existe pas lorsque (x, y) s'approche de $(0, 0)$, nous notons qu'il est impossible de satisfaire la définition d'une limite d'une fonction à deux variables du fait que la fonction prend différentes valeurs le long de différentes lignes passant par le point $(0, 0)$. Tout d'abord, considérons la ligne $y = 0$ dans le

plan xy. Substituer $y = 0$ en $f(x, y)$ donne

$$f(x, 0) = \frac{2x \times 0}{3x^2 + 0^2} = 0$$

pour toute valeur de x. Par conséquent, la valeur de f reste constante pour tout point sur l'axe des x, et lorsque y s'approche de zéro, la fonction reste fixe à zéro. Ensuite, considérons la ligne $y = x$. Substituer $y = x$ dans $f(x, y)$ donne

$$f(x, x) = \frac{2x \times x}{3x^2 + x^2} = \frac{2x^2}{4x^2} = \frac{1}{2}$$

Ceci est vrai pour tout point sur la ligne $y = x$. Si nous laissons x approcher de zéro tout en restant sur cette ligne, la valeur de la fonction reste fixée à $1/2$ quelle que soit la taille de x. On choisit une valeur pour ε inférieure à $1/2$, disons $1/4$. Ensuite, quelle que soit la taille d'un disque que nous dessinons autour de $(0, 0)$, les valeurs de $f(x, y)$ pour les points à l'intérieur de ce disque incluront à la fois 0 et 2. Par conséquent, la définition d'une limite en un point n'est jamais satisfaite et la limite n'existe pas.

(2) On veut déterminer la limite suivante

$$\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2 + 3y^4}$$

De la même manière que précédemment, nous pouvons approcher l'origine le long de toute ligne droite passant par l'origine. Si nous essayons l'axe x (c'est-à-dire $y = 0$), alors la fonction reste fixe à zéro. La même chose est vraie pour l'axe y. Supposons à présent que nous approchons de l'origine le long d'une droite de pente k. L'équation de cette ligne est $y = kx$. Alors la limite devient

$$\begin{aligned}\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2 + 3y^4} &= \lim_{(x,y)\to(0,0)} \frac{4x \times (kx)^2}{x^2 + 3 \times (kx)^4} = \lim_{(x,y)\to(0,0)} \frac{4k^2x^3}{x^2 + 3k^4x^4} \\ &= \lim_{(x,y)\to(0,0)} \frac{4k^2x}{1 + 3k^4x^2} = \frac{\lim_{(x,y)\to(0,0)}(4k^2x)}{\lim_{(x,y)\to(0,0)}(1 + 3k^4x^2)} = 0\end{aligned}$$

quelle que soit la valeur de k. Il semblerait que la limite soit égale à zéro. Et si on choisissait plutôt une courbe passant par l'origine? Par exemple, on peut considérer la parabole donnée par l'équation $x = y^2$. La substitution de y^2 à la place de x dans $f(x, y)$ donne

$$\lim_{(x,y)\to(0,0)} \frac{4xy^2}{x^2 + 3y^4} = \lim_{(x,y)\to(0,0)} \frac{4y^2 \times y^2}{(y^2)^2 + 3y^4} = \lim_{(x,y)\to(0,0)} \frac{4y^4}{y^4 + 3y^4} = \lim_{(x,y)\to(0,0)} = 1$$

Par la même logique que précédemment, il est impossible de trouver un disque autour de l'origine qui satisfasse la définition de la limite pour toute valeur de $\varepsilon < 1$. Par conséquent, la limite n'existe pas.

3.2 Points intérieurs et points de bord

Topologie élémentaire dans le plan. Pour étudier la continuité et la différentiabilité d'une fonction de deux ou plusieurs variables, nous devons d'abord nous familiariser avec du vocabulaire issu de la *topologie*. Dans les définitions qui suivent on pourra se reporter à la **Figure 9** afin de mieux les comprendre. Soit $\mathcal{S}$ un sous-ensemble de $\mathbb{R}^2$, dans ce cas :

(1) Un point P est appelé un **point intérieur** de $\mathcal{S}$ s'il existe un disque centré autour de P contenu complètement dans $\mathcal{S}$.

(2) Un point P est appelé un **point de bord** ou **point-frontière** (*boundary point*) de $\mathcal{S}$ si chaque disque centré autour de P contient des points à la fois à l'intérieur et à l'extérieur de $\mathcal{S}$.

(3) $\mathcal{S}$ est appelé un **ensemble ouvert** si chaque point de $\mathcal{S}$ est un point intérieur.

(4) $\mathcal{S}$ est appelé un **ensemble fermé** s'il contient tous ses points de bord.

(5) Un ensemble ouvert $\mathcal{S}$ est un **ensemble connexe** s'il ne peut pas être représenté comme l'union de deux ou plusieurs sous-ensembles ouverts disjoints et non vides.

(6) $\mathcal{S}$ est appelé une **region** si il est ouvert, connexe et non vide.

Un exemple simple d'ensemble ouvert est un disque. Si nous incluons la frontière du disque, alors il devient un ensemble fermé. Un ensemble qui contient certains de ses points de bord, mais pas tous, n'est ni ouvert ni fermé. Par exemple, si nous incluons la moitié de la limite d'un disque mais pas l'autre moitié, alors l'ensemble n'est ni ouvert ni fermé.

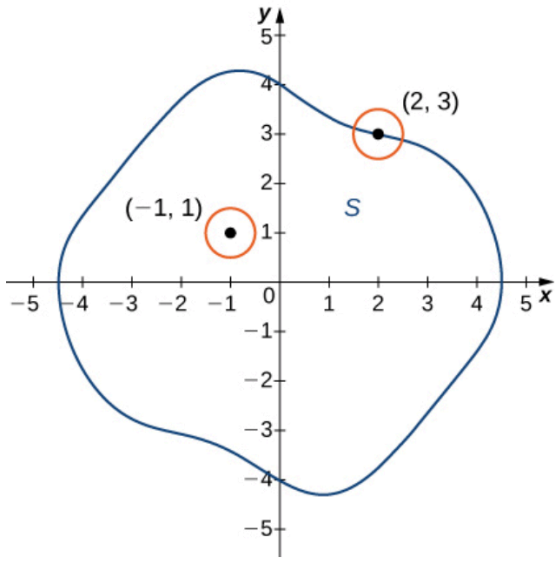


Figure 9: Dans l'ensemble $\mathcal{S}$ montré, $(-1, 1)$ est un *point intérieur* et $(2, 3)$ est un *point frontière*.

Notion de limite. La définition d'une limite d'une fonction à deux variables nécessite que le disque soit contenu dans le domaine de la fonction. Cependant, si nous souhaitons trouver la limite d'une fonction à un point de bord du domaine, le disque n'est pas automatiquement contenu dans le domaine. Par définition, certains des points du disque sont à l'intérieur du domaine et d'autres à l'extérieur. Par conséquent, nous avons besoin de modifier notre définition afin de ne considérer que les points qui sont à la fois à l'intérieur du disque et qui font parti du domaine de définition de la fonction. Cela conduit à la définition de la limite d'une fonction en un point de bord.

Définition améliorée. Soit f une fonction de deux variables (réelles) x et y, et supposons que (a, b) est à la limite du domaine de f, c'est-à-dire que c'est un point du bord. Alors, la limite de $f(x, y)$ lorsque (x, y) s'approche de (a, b) est L et s'écrit

$$\lim_{(x,y)\to(a,b)} f(x, y) = L$$

si pour chaque $\varepsilon > 0$, il existe un nombre $\delta > 0$ tel que pour tout point (x, y) à l'intérieur du domaine de f et appartenant au disque de rayon δ de (a, b), la différence entre $f(x, y)$ et L soit

plus petite que la quantité ε (**Figure 7**). En utilisant des symboles puis en notant $\boldsymbol{x} = (x, y)$ et $\boldsymbol{x}_0 = (a, b)$, on peut écrire que: Pour tout $\varepsilon > 0$, il existe un nombre $\delta > 0$ tel que

$$|f(x, y) - L| < \varepsilon$$

pour tout $\boldsymbol{x} \in \mathbb{R}^2$ vérifiant

$$0 < \|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$$

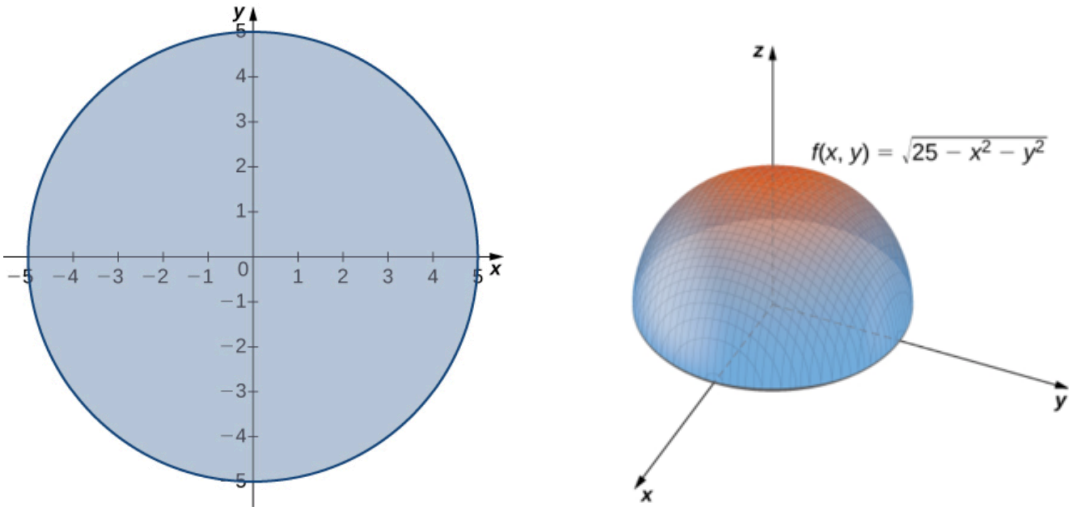


Figure 10: Domaine (gauche) et graphe (droite) de la fonction $f(x, y) = \sqrt{25 - x^2 - y^2}$.

Exemples 8. On veut montrer que la limite suivante

$$\lim_{(x,y)\to(4,3)} \sqrt{25 - x^2 - y^2} = 0$$

où le point $(4, 3)$ est un point de bord. Le domaine de la fonction $f(x, y) = \sqrt{25 - x^2 - y^2}$, parfois noté $\mathsf{Dom}(f)$, est donné par l'ensemble

$$\mathsf{Dom}(f) = \{(x, y) \in \mathbb{R}^2, x^2 + y^2 < 25\}$$

Cela correspond géométriquement à un cercle de rayon 5 centré à l'origine. Son domaine intérieur est indiqué sur la **Figure 10** (à gauche). Nous pouvons utiliser les propriétés de la limite, qui s'appliquent aux limites du bord des domaines ainsi qu'aux points intérieurs, voir la **Figure 10 (droite)**. On écrit alors

$$\begin{aligned}\lim_{(x,y)\to(4,3)} \sqrt{25 - x^2 - y^2} &= \sqrt{\lim_{(x,y)\to(4,3)} (25 - x^2 - y^2)} \\ &= \sqrt{\lim_{(x,y)\to(4,3)} 25 - \lim_{(x,y)\to(4,3)} x^2 - \lim_{(x,y)\to(4,3)} y^2} = \sqrt{25 - 4^2 - 3^2} = 0\end{aligned}$$

3.3 Continuité des fonctions de deux variables

Définition. Nous savons que la continuité d'une fonction d'une variable dépend de la limite d'une fonction d'une variable. En particulier, pour que $f(x)$ soit continue au point $x = x_0$, il faut que $f(x_0)$ et la limite de $f(x)$ quand $x \to x_0$ existent, et enfin il faut que ces valeurs coïncident. Ces trois conditions sont également nécessaires pour la continuité d'une fonction de deux variables (ou plus). On peut donc directement énoncé qu'une fonction $f(\boldsymbol{x}) = f(x, y)$ est continue en un point $\boldsymbol{x}_0 = (x_0, y_0)$ de son domaine si les conditions suivantes sont remplies :

(1) $f(\boldsymbol{x}_0)$ existe

(2) $\lim\limits_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x})$ existe

(3) $\lim\limits_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = f(\boldsymbol{x}_0)$

Exemples 9. Montrons que la fonction $f(x, y) = f(\boldsymbol{x})$ ci-dessous est continue au point $\boldsymbol{x}_0 = (5, -3)$.

$$f(\boldsymbol{x}) = \frac{3x + 2y}{x + y + 1}$$

Selon la définition, les conditions **(1)**, **(2)** et **(3)** sont à vérifier avec $x_0 = 5$ et $y_0 = -3$.

(1) Cette condition est remplie car le domaine de la fonction f est constitué de ces paires ordonnées pour lesquelles le dénominateur est différent de zéro (c'est-à-dire $x+y+1 \neq 0$). Le point $(5, -3)$ satisfait cette condition. En outre,

$$f(\boldsymbol{x}_0) = f(5, -3) = \frac{3(5) + 2(-3)}{5 + (-3) + 1} = \frac{15 - 6}{2 + 1} = 3$$

(2) Cette condition est remplie puisque

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} f(\boldsymbol{x}) = \lim_{(x,y)\to(5,-3)} \frac{3x + 2y}{x + y + 1} = \frac{\lim_{(x,y)\to(5,-3)}(3x + 2y)}{\lim_{(x,y)\to(5,-3)}(x + y + 1)} = \frac{15 - 6}{5 - 3 + 1} = 3$$

(3) Cette condition est remplie étant donné que nous venons de montrer que les deux côtés de cette équation valent trois.

La continuité d'une fonction de plusieurs variables peut également être définie en termes de delta et epsilon. Par exemple, en deux dimensions, une fonction $f(x, y)$ est continue en un point $\boldsymbol{x}_0 = (x_0, y_0)$ de son domaine si pour tout $\varepsilon > 0$ il existe une quantité $\delta > 0$ tel que, chaque fois que $\|\boldsymbol{x} - \boldsymbol{x}_0\| < \delta$ alors l'inégalité $|f(x, y) - f(x_0, y_0)| < \varepsilon$ est vérifiée.

Propriétés. La définition précédente a des conséquences immédiates (et voulues) par rapport aux opérations communes que l'on peut avoir a effectuées sur des fonctions de deux variables. En effet, si $f(x, y) = f(\boldsymbol{x})$ et $g(\boldsymbol{x}) = g(x, y)$ sont des fonctions continues en $\boldsymbol{x}_0 = (x_0, y_0)$ et $u(x), v(y)$ sont deux fonctions également continues en x_0 et y_0 respectivement, alors :

(1) La fonction $h(x, y) = f(x, y) + g(x, y)$ est également continue en (x_0, y_0).

(2) La fonction $w(x, y) = u(x) \cdot v(x)$ est également continue en (x_0, y_0).

(3) Si $f(z)$ est une fonction (d'une variable) continue en $z_0 = g(x_0, y_0) = g(\boldsymbol{x}_0)$ alors la fonction

$$w(x, y) = w(\boldsymbol{x}) = (f \circ g)(\boldsymbol{x}) = f(g(\boldsymbol{x}))$$

est continue en $\boldsymbol{x}_0 = (x_0, y_0)$, voir la **Figure 11**.

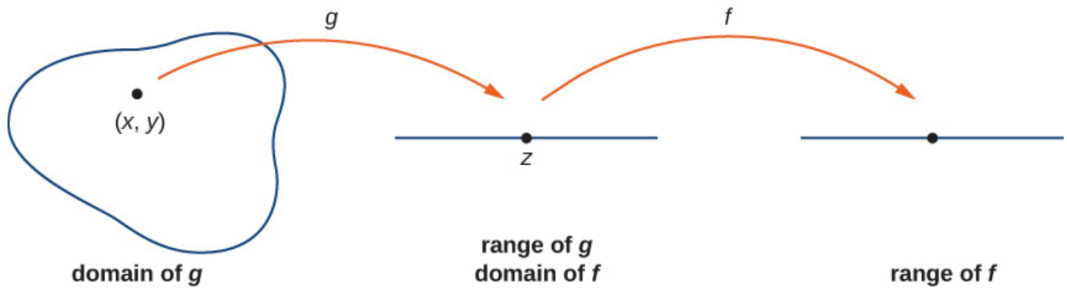


Figure 11: La composition de deux fonctions continues est continue.

Exemples 10. On veut montrer que les fonctions suivantes sont continues partout.

$$f(\boldsymbol{x}) = f(x,y) = 4x^3y^2 \qquad \text{et} \qquad g(\boldsymbol{x}) = g(x,y) = \cos(4x^3y^2)$$

Pour ce faire, nous remarquons d'abord que les polynômes $g(x) = 4x^3$ et $h(y) = y^2$ sont continus pour toute valeur réelle, et donc par **(2)**, la fonction $f(\boldsymbol{x}) = 4x^3y^2$ est continue en tout point (x,y) dans le plan xy. Puisque $f(\boldsymbol{x}) = 4x^3y^2$ est continue en tout point (x,y) dans le plan xy et $g(x) = \cos x$ est continu pour toute valeur réelle x, le résultat **(3)** implique que la fonction $g(\boldsymbol{x}) = \cos(4x^3y^2)$ est continue en tout point $\boldsymbol{x} = (x,y)$ dans le plan xy.

3.4 Fonctions de trois variables ou plus

Motivation. Il est fréquent, lors d'applications concrètes, de devoir évaluer la limite d'une fonction de trois variables ou plus. Par exemple, une fonction $F(\boldsymbol{x}) = F(x,y,z)$ peut correspondre à la température au point de coordonnées $\boldsymbol{x} = (x,y,z)$ dans l'espace physique à trois dimensions. Ou encore une fonction $G(\boldsymbol{x}) = G(x,y,z,t)$ peut indiquer la pression atmosphérique en un point $\boldsymbol{x} = (x,y,z,t)$ de l'espace et au temps t. Comment prendre la limite en un point dans $\mathbb{R}^3$? Que signifie être continu en un point en quatre dimensions? Afin de répondre à ces questions nous devons d'abord étendre les concepts que nous avons établies en deux dimensions en commençant par la notion de disque, c'est-à-dire de domaine centré un point. À partir de là, les idées de limite et de continuité sont facilement généralisable à une fonction de trois variables ou plus, notamment grâçe à la notation $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ expliquée auparavant.

Notion de boule. Une **boule** de rayon δ en trois dimensions est constituée de tous les points de $\mathbb{R}^3$ situés à une distance inférieure à δ du point $\boldsymbol{a} = (x_0, y_0, z_0)$ de $\mathbb{R}^3$, soit

$$\left\{\boldsymbol{x} = (x,y,z) \in \mathbb{R}^3 \mid \|\boldsymbol{x} - \boldsymbol{a}\| < \delta\right\}$$

La généralisation de ce concept à des dimensions supérieures est immédiate puisqu'il suffit, pour une boule centrée au point $\boldsymbol{a} = (a_0, \ldots, a_n) \in \mathbb{R}^n$, de considérer l'ensemble des points $\boldsymbol{x} = (x_0, \ldots, x_n) \in \mathbb{R}^n$ tels que $\|\boldsymbol{x} - \boldsymbol{a}\| < \delta$ en utilisant la norma euclidienne dans $\mathbb{R}^n$, c'est-à-dire

$$\|\boldsymbol{x} - \boldsymbol{a}\| = \sqrt{\sum_{i=1}^{n} (x_i - a_i)^2}$$

Pour montrer qu'il existe une limite d'une fonction à trois variables en un point $\boldsymbol{x}_0 = (x_0, y_0, z_0)$, il suffit de montrer que pour tout point d'une boule centrée sur $\boldsymbol{x}_0$, la valeur de la fonction à ce point

est arbitrairement proche d'une valeur fixe (la valeur limite). Il est important de noter que toutes les propriétés sur les limites pour les fonctions de deux variables sont également valables pour les fonctions de trois variables et plus.

Exemples 11. Supposons que nous voulions déterminer la limite suivante.

$$\lim_{(x,y,z)\to(4,1,-3)} \frac{x^2y-3z}{2x+5y-z}$$

Avant de pouvoir appliquer la loi quotient **(7)**, nous devons vérifier que la limite du dénominateur est différente de zéro. En utilisant les lois différence **(4)**, identité **(2)** et constante **(1)**,

$$\begin{aligned}\lim_{(x,y,z)\to(4,1,-3)} 2x+5y-z &= 2\left(\lim_{(x,y,z)\to(4,1,-3)} x\right)+5\left(\lim_{(x,y,z)\to(4,1,-3)} y\right)-\left(\lim_{(x,y,z)\to(4,1,-3)} z\right)\\ &= 2(4)+5(1)-(-3) = 16\end{aligned}$$

Puisque ce n'est pas nul, on trouve ensuite la limite du numérateur. En utilisant les lois produit **(6)**, différence **(4)**, multiplication **(5)** et identité **(1)**, on trouve

$$\begin{aligned}\lim_{(x,y,z)\to(4,1,-3)} x^2y-3z &= \left(\lim_{(x,y,z)\to(4,1,-3)} x\right)^2\left(\lim_{(x,y,z)\to(4,1,-3)} y\right)-3\left(\lim_{(x,y,z)\to(4,1,-3)} z\right)\\ &= (4^2)(1)-3(-3) = 16+9 = 25\end{aligned}$$

Enfin, en appliquant la loi quotient **(7)** on a

$$\lim_{(x,y,z)\to(4,1,-3)} \frac{x^2y-3z}{2x+5y-z} = \frac{\lim_{(x,y,z)\to(4,1,-3)}(x^2y-3z)}{\lim_{(x,y,z)\to(4,1,-3)}(x^2y-3z)} = \frac{25}{16}$$

4 La dérivée partielle

Introduction. L'idée centrale du calcul différentiel est la dérivée. Pour une fonction d'une seule variable, un changement en x produit un changement en f et le rapport $\Delta f/\Delta x$ s'approche de la dérivée, ou de la pente, ou du taux d'accroissement. La question naturelle qui se pose est que faire si f dépend à la fois de x et de y? L'idée nouvelle est de faire varier x et y *séparément*. Premièrement, considérons uniquement des variations de x. Si la fonction est $f = x + xy$, alors Δf est $\Delta x + y\,\Delta x$ et le taux d'accroissement $\Delta f/\Delta x$ est alors $1 + y$. On en conclut que la dérivée par rapport à x de $x + xy$ est $1 + y$. Pour toutes les fonctions, la méthode est la même. La variable y reste fixe et x varie, puis on prend la limite dans le taux d'accroissement $\Delta f/\Delta x$.

4.1 Définitions et exemples

Définition. La **dérivée partielle** de la fonction $f(x, y)$ par rapport à x est donnée par

$$\frac{\partial f}{\partial x}(x, y) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}$$

Sur la gauche se trouve un nouveau symbole $\frac{\partial f}{\partial x}$ qui signale que seul x est autorisé à varier. La forme différente « ∂ » de la lettre « d » est un rappel que x n'est pas la seule variable mais elle se prononce de la même manière. L'autre variable y est présente mais elle reste fixe et joue le rôle d'un paramètre extérieur dans la procédure de limite.

Exemples 12. Considérons la fonction f suivante et sa dérivée partielle par rapport à x.

$$f(x, y) = x^2y^2 + xy + y \qquad \text{et} \qquad \frac{\partial f}{\partial x}(x, y) = 2xy^2 + y + 0$$

Il ne faut surtout pas traiter y comme zéro! Cette variable doit être considérée comme une constante. Sa dérivée par rapport à x est zéro. Si $f(x) = \sin 6x$ alors $\partial f/\partial x = 6\cos 6x$. Si $f(x, y) = \sin xy$ alors $\partial f/\partial x = y \cos xy$.

Le terme $\partial f/\partial x$ se prononce « *df* sur *dx* » et c'est une fonction de x et de y. Lorsque l'on a besoin d'être plus précis on dira « *la dérivée partielle de f par rapport à x* ». Le symbole f' n'est plus disponible, car il ne donne aucune indication particulière sur x. On le remplacera par f_x et se prononce « *f x* » ou « *f* indice *x* », ce qui est plus court que $\partial f/\partial x$ et signifie la même chose. On peut aussi vouloir indiquer le point (x_0, y_0) où la dérivée est calculée. Dans ce cas on écrira

$$\frac{\partial f}{\partial x}(x_0, y_0) \quad \text{ou} \quad f_x(x_0, y_0) \quad \text{ou} \quad \left.\frac{\partial f}{\partial x}\right|_{(x_0, y_0)} \quad \text{ou juste} \quad \left(\frac{\partial f}{\partial x}\right)_0$$

Exemples 13. On considère la fonction $f(x, y) = \sin 2x \cos y$ dans ce cas

$$f_x = 2\cos 2x \cos y$$

car $\cos y$ est constant pour $\partial/\partial x$. Le point particulier (x_0, y_0) à considérer est ici $(0, 0)$. La hauteur de la surface est $f(0, 0) = 0$. La pente dans la direction des x est $f_x = 2$. En un point différent, par exemple au point $(x_0, y_0) = (\pi, \pi)$ on aurait $f_x(\pi, \pi) = -2$.

Maintenant, *on garde x constant et on fait varier y*. Le rapport $\Delta f/\Delta y$ approche $\partial f/\partial y$: nous avons alors la formule équivalente à la définition précédente pour la variable y, c'est-à-dire

$$f_y(x, y) = \lim_{\Delta y \to 0} \frac{\Delta f}{\Delta y} = \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}$$

C'est la pente dans la direction des y. Veuillez noter qu'une surface peut monter dans la direction des x et descendre dans la direction des y. Le plan $f(x,y) = 3x - 4y$ a $f_x = 3$ (haut) et $f_y = -4$ (bas). Par la suite nous allons voir comment déterminer la pente dans la direction des 45° correspondant à la droite $x = y$ (dans le plan xy).

Exemples 14. On considère les fonctions suivantes

$$f(x,y) = \sqrt{x^2+y^2} \quad , \quad \frac{\partial f}{\partial x} = \frac{x}{\sqrt{x^2+y^2}} = \frac{x}{f} \quad , \quad \frac{\partial f}{\partial y} = \frac{y}{\sqrt{x^2+y^2}} = \frac{y}{f}$$

La dérivée par rapport à x de $\sqrt{x^2+y^2}$ est en réalité un calcul à une variable, car y est constant. L'exposant passe de $1/2$ à $-1/2$, et il y a $2x$ provenant de la dérivée composée (*chain rule*). On remarque que cette « fonction distance » possède la dérivée remarquable $\frac{\partial f}{\partial x} = \frac{x}{f}$ que l'on peut voir comme une *équation différentielle* définissant f par rapport à x. Le graphique associé à f est un cône, voir la **Figure 12 (gauche)**. Au-dessus du point $(0,2)$, le point sur la surface du cône est à la hauteur $\sqrt{0^2+2^2} = 2$. Les dérivées partielles en ce point sont $f_x = 0/2$ et $f_y = 2/2$. Donc, en ce point, la surface monte dans la direction des y mais reste constante dans la direction des x. On peut comprendre cela de la manière suivante. Pour un pas de Δx la hauteur passera de 0^2+2^2 à $(\Delta x)^2+2^2$ avec un changement d'ordre $(\Delta x)^2$ (ordre quadratique et non linéaire) et donc la dérivée par rapport à x de f en $(0,2)$ est nulle.

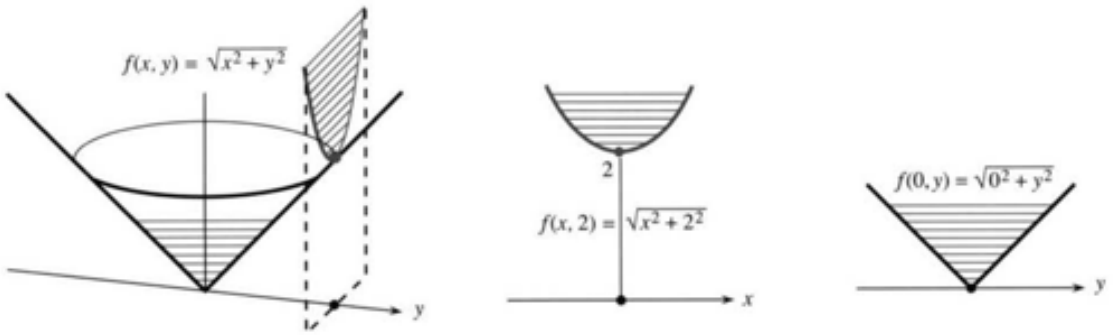


Figure 12: Fonctions partielles $\sqrt{x^2+y^2}$ et $\sqrt{0^2+2^2}$ de la fonction de distance $f = \sqrt{x^2+y^2}$.

Analyse graphique et fonction partielles. La **Figure 12** est assez importante, elle montre comment $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ sont les dérivées ordinaires de $f(x,y_0)$ et $f(x_0,y)$ dans les direction parallèles aux axes $y = y_0$ et $x = x_0$ respectivement. En effet on peut vérifier que

$$\frac{d}{dx} f(x,y_0) = \frac{\partial f}{\partial x}\Big|_{y=y_0} \qquad \text{et} \qquad \frac{d}{dy} f(x_0,y) = \frac{\partial f}{\partial y}\Big|_{x=x_0}$$

Dans ce sens il est naturel de les appeler « *fonctions partielles* ». La première possède une ordonnée fixée à $y = y_0$ tandis que x varie. La seconde fonction possède son abcisse fixée à $x = x_0$ tandis que y varie. Leurs graphiques correspondent à des *coupes transversales sur la surface*, c'est-à-dire découpées par les plans verticaux $y = y_0$ et $x = x_0$, voir la **Figure 12 (centre et droite)** pour

$$x_0 = 0 \qquad \text{et} \qquad y_0 = 2$$

N'oubliez pas que la ligne de niveau est découpée par le plan horizontal $z = c$. Les limites de $\frac{\partial f}{\partial x}$ et $\frac{\partial f}{\partial y}$ sont calculées comme d'habitude. Avec les fonctions partielles, nous revenons à une seule

variable. *La dérivée partielle est la dérivée ordinaire d'une fonction partielle* (avec y constante ou x constant). Pour le cône, $\frac{\partial f}{\partial y}$ existe en tous points sauf $(0,0)$. La **Figure 12 (droite)** montre comment la section transversale au milieu du cône produit la fonction valeur absolue c'est-à-dire

$$f(0,y) = |y|$$

Il a des dérivées unilatérales mais pas de dérivée bilatérale. De même $\frac{\partial f}{\partial x}$ n'existera pas à la pointe du cône. À travers ce raisonnement on commence à développer l'idée d'une *fonction continue* $f(x,y)$ en fonction de nos besoins (la définition fait l'objet de l'**Exercice 18**). Il faut bien garder à l'esprit que chaque dérivée partielle implique une direction, mais les limites et la continuité en un point impliquent toutes les directions. On notera au final que la fonction distance $\sqrt{x^2+y^2}$ est continue en $(0,0)$ mais elle n'est pas différentiable en ce point.

Exemples 15. On considère la fonction $f(x,y) = y^2 - x^2$ dont les dérivées partielles sont

$$\frac{\partial f}{\partial x} = -2x \qquad \text{et} \qquad \frac{\partial f}{\partial y} = 2y$$

Si on se déplace dans la direction des x à partir du point $(1,3)$ alors $f = x^2 - y^2$ correspond à la fonction partielle $9 - x^2$, avec $y = 3$, c'est-à-dire une parabole orientée vers le bas. Si maintenant on se déplace dans la direction des y, en fixant $x = 1$, la fonction partielle est $y^2 - 1$ correspond à une parabole orientée vers le haut. La surface ainsi décrite, correspondant à la fonction f, est illustrée sur **Figure 13** et est appelée une *paraboloïde hyperbolique*, car les lignes de niveau $x^2 - y^2 = c$ sont des hyperboles. Il est courant d'appeler une telle configuration un *col*, et le point spécial à l'origine est appelé *point-col* ou *point-selle* (*saddle point*). Il est situé aux extrema des paraboles décrites précédemment à partir du point $(0,0)$. En effet, l'origine $(0,0)$ joue un rôle particulier pour la fonction $f = x^2 - y^2$ car ses deux dérivées partielles sont nulles en ce point. De plus *le bas de la parabole y en $(0,0)$ est en haut de la parabole x*. La surface est donc localement plate en ce point dans toutes les directions. C'est à la fois le sommet d'une colline et le bas d'une chaîne de montagnes. Il est important de noter qu'un point-col n'est *ni un maximum ni un minimum*, bien que les deux dérivées soient nulles.

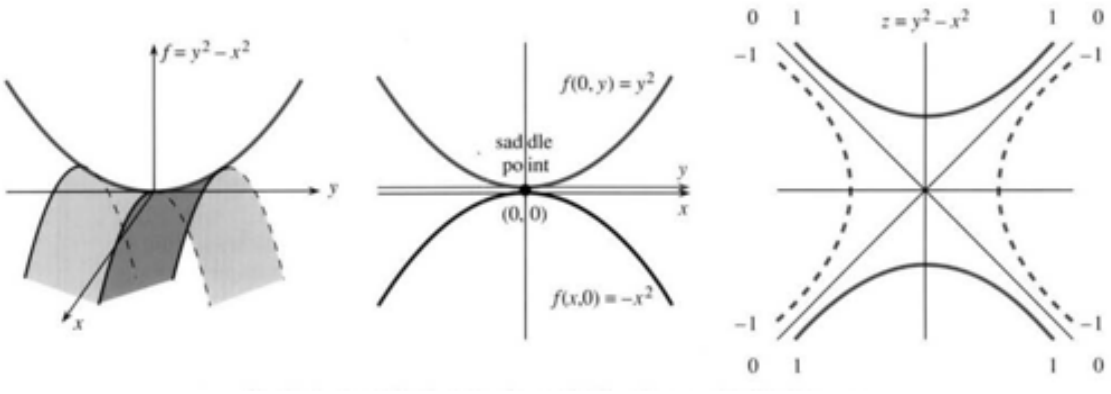


Figure 13: Une fonction de selle, ses fonctions partielles et ses lignes de niveau.

Attention, il ne faut pas croire que $f(x,y)$ doit contenir y^2 et x^2 pour avoir un point-col, la fonction $2xy$ le fait tout aussi bien. Les lignes de niveau $2xy = c$ sont toujours des hyperboles. Les fonctions partielles $2xy_0$ et $2x_0y$ donnent maintenant des lignes droites (ce qui est remarquable). Le long de la ligne des 45°, d'équation $x = y$, la fonction devient $2x^2$ et augmente pour des x croissants. Au

contraire, le long de la ligne $-45°$, d'équation $x = -y$, la fonction devient $-2x^2$ et elle diminue donc pour des x croissants. Le graphe de $2xy$ est donnée sur la **Figure 13** tournée de $45°$.

Exemples 16. On considère les deux fonctions suivantes

$$f(x, y, z) = x^2 + y^2 + z^2 \qquad \text{et} \qquad P(T, V) = \frac{nRT}{V}$$

La fonction f montre plus de variables cependant la fonction P démontre que les variables ne sont pas toujours nommées x et y, mais également que la fonction n'est pas toujours nommée f! La pression, la température et le volume sont P, T et V. Les lettres changent mais les règles du calcul différentiel reste bien sûr inchangées. Par exemple on peut écrire

$$\frac{\partial P}{\partial T} = \frac{nR}{V} \qquad \text{et} \qquad \frac{\partial P}{\partial V} = -\frac{nRT}{V^2}$$

où on notera la dérivée de $1/V$. Il n'y a pas de $\frac{\partial P}{\partial R}$ car r est une constante de chimie (la constante des gaz parfaits) et ce n'est pas une variable par rapport à laquelle on peut différencier.

Conclusion. À ce stade les exemples sont innombrables. En physique on caractérise le mouvement d'un corps par un jeu de six variables, ses trois coordonnées dans l'espace x, y, z ainsi les trois composantes p_x, p_y, p_y de sa quantité de mouvement, ou impulsion (*momentum*), $\vec{p} = m\vec{v}$ ou $\boldsymbol{p} = m\,\boldsymbol{v}$. Autre exemple, en économie, si on souhaite décrire les variations de prix d'une centaine de produits, on a besoin d'introduire un nombre de variables égal afin de calculer le profit global sur la vente de ces produits. De plus, rien n'interdit que ce profit soit une fonction (très) compliquée de ces variables. Typiquement, les dérivées partielles donnent accès aux *profit marginal*, c'est-à-dire associé à un produit spécifique. Une simple feuille de calcul permet de relever le changement sur une fonction puis, sur une feuille de papier millimétré, on peut ensuite calculer la dérivée.

4.2 La dérivée seconde

Notation. Le passage aux dérivées secondes ne présente pas de sérieuses difficultés techniques si ce n'est que les deux dérivées premières f_x et f_y conduisent à quatre dérivées secondes

$$f_{xx} = \frac{\partial^2 f}{\partial x^2} \quad , \quad f_{yy} = \frac{\partial^2 f}{\partial y^2} \quad , \quad f_{xy} = \frac{\partial^2 f}{\partial x \partial y} \quad , \quad f_{yx} = \frac{\partial^2 f}{\partial y \partial x}$$

En suivant la notation on remarque que f_{xx} est la dérivée par rapport à x de la dérivée par rapport à x. Heureusement f_{xy} est égal à f_{yx}, comme nous allons le voir d'abord par l'exemple.

Exemples 17.

(1) Si $f = x/y$ alors on a $f_x = 1/y$ puis $f_{xx} = 0$ et $f_{xy} = -1/y^2$. La fonction x/y est linéaire en x (ce qui explique $f_{xx} = 0$). Ensuite on calcule que $f_y = -x/y^2$ et donc $f_{xy} = -1/y^2$. *Les dérivées mixtes f_{xy} et f_{yx} sont égales.* Dans la direction des y, la dérivée seconde est $f_{yy} = 2x/y^3$. Le calcul à une variable est suffisant pour toutes ces dérivées, car une seule variable est en mouvement.

(2) Si $f = 4x^2 + 3xy + y^2$ alors $f_x = 8x + 3y$ et $f_y = 3x + 2y$. On peut directement observer que les dérivées mixtes f_{xy} et f_{yx} sont donc égales à 3. La dérivée seconde dans la direction des x est $\frac{\partial^2 f}{\partial x^2} = f_{xx} = 8$, de même $\frac{\partial^2 f}{\partial y^2} = f_{yy} = 2$.

(3) On veut calculer les quatre dérivées partielles du second ordre de $f(x, y) = xy^2 + 3x^2e^y$. De $f_x = y^2 + 6xe^y$ on obtient

$$f_{xx} = \frac{\partial}{\partial x}\left(y^2 + 6xe^y\right) = 6e^y \qquad \text{et} \qquad f_{xy} = \frac{\partial}{\partial y}\left(y^2 + 6xe^y\right) = 2y + 6xe^y$$

Puis de $f_y = 2xy + 3x^2e^y$ on obtient (on observe que $f_{xy} = f_{yx}$)

$$f_{yx} = \frac{\partial}{\partial x}\left(2xy + 3x^2e^y\right) = 2y + 6xe^y \qquad \text{et} \qquad f_{yy} = \frac{\partial}{\partial y}\left(2xy + 3x^2e^y\right) = 2x + 3x^2e^y$$

Théorème de Schwarz. La règle générale est la suivante, si $f(x, y)$ possède des dérivées secondes *continues* alors

$$f_{xy} = f_{yx} \qquad \text{ou encore} \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

L'**Exercice 17** esquisse une preuve basée sur le théorème de la valeur moyenne. Il est a noté que ce résultat est souvent appelé le *Théorème de Schwarz* pour des fonctions différentiables. Incidemment, pour les dérivées troisièmes, on peut vérifier que presque pour tous les exemples on a $f_{xxy} = f_{xyx} = f_{yxx}$ est différent de $f_{yyx} = f_{yxy} = f_{xyy}$.

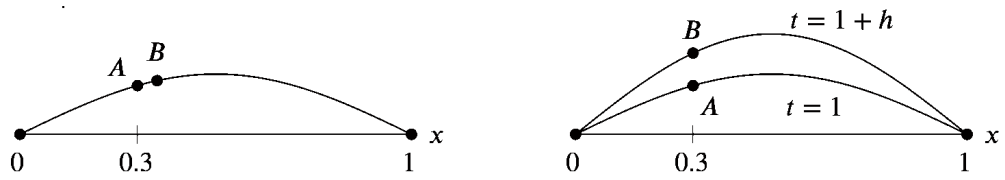


Figure 14: Gauche: Pour $f_{xx}(0.3, 1) < 0$ la concavité de la corde à $t = 1$. La pente en B est moins grande que celle en A. Droite: Pour $f_{xt}(0.3, 1) > 0$. La pente en un point de la corde à deux instants différents. La pente en B est plus grande que celle en A.

Exemples 18. Prenons l'exemple d'une corde de guitare d'un mètre de long. À un point d'abscisse x le long de cette corde et à l'instant t, celle-ci montre un déplacement vertical décrit par la fonction

$$f(x, t) = 0.003 \sin(\pi x) \sin(2765\, t)$$

On veut ici calculer les quatre dérivées partielles secondes de f au point $(x, t) = (0.3, 1)$ et interpréter le sens de leur signe en termes pratiques. On trouve d'abord que $f_x = 0.003\,\pi \cos(\pi x) \sin(2765\, t)$ donc

$$\begin{aligned} f_{xx} &= -0.003\,\pi^2 \sin(\pi x) \sin(2765\, t) && \text{soit} \quad f_{xx}(0.3, 1) \simeq -0.01 \\ f_{xt} &= (0.003)\,(2765)\,\pi \cos(\pi x) \cos(2765\, t) && \text{soit} \quad f_{xt}(0.3, 1) \simeq 14 \end{aligned}$$

La dérivée première $f_x(x, t)$ donne en fait la pente de la corde en tout point à tout instant. La dérivée seconde $f_x(x, t)$ indique le sens de la *concavité* de la corde. Le fait que $f_{xx}(0.3, 1) < 0$ implique que la concavité de la corde est orientée vers le bas au point $x = 0.3$ quand $t = 1$, voir la **Figure 14 (gauche)**. D'un autre coté, $f_{xt}(x, t)$ is le taux d'accroissement de la pente de la corde par rapport au temps. Donc $f_{xt}(0.3, 1) > 0$ indique qu'au temps $t = 1$ la pente au point $x = 0.3$ augmente, voir la **Figure 14 (droite)**. Maintenant on trouve que

$$f_t = (0.003)\,(2765)\,\pi \sin(\pi x) \cos(2765\, t)$$

On en tire donc que

$$\begin{aligned} f_{tx} &= (0.003)\,(2765)\,\pi \cos(\pi x) \cos(2765\, t) && \text{soit} \quad f_{tx}(0.3, 1) \simeq 14 \\ f_{tt} &= -(0.003)\,(2765)^2 \sin(\pi x) \sin(2765\, t) && \text{soit} \quad f_{tt}(0.3, 1) \simeq -7200 \end{aligned}$$

La dérivée première $f_t(x,t)$ donne la vitesse de la corde en tout point et à tout instant. Donc les dérivées secondes f_{tx} et f_{tt} indiquent toutes les deux le taux d'accroissement de la vitesse. Si $f_{tx}(0.3,1) > 0$ indique qu'au temps $t = 1$ la vitesse des points juste à la droite de $x = 0.3$ est plus grand celle en $x = 0.3$, voir la **Figure 15 (gauche)**. Si $f_{tt}(0.3,1) < 0$ alors la vitesse du point $x = 0.3$ diminue at temps $t = 1$. On a alors l'accélération

$$f_{tt}(0.3,1) = -7200\,\mathsf{m\,s}^{-2}$$

en ce point, voir la **Figure 15 (droite)**.

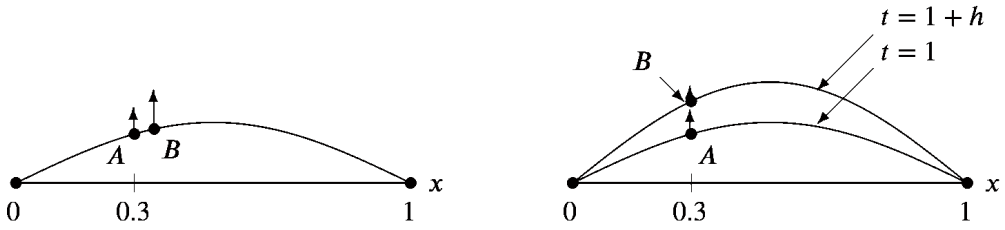


Figure 15: Gauche: Pour $f_{tx}(0.3,1) < 0$ on a la vitesse de différents points au temps $t = 1$. La pente en B est moins grande que celle en A. Droite: Pour $f_{tt}(0.3,1) < 0$ on a une accélération négative. La figure montre ;a vitesse en un point de la corde à deux instants différents. La pente en B est plus grande que celle en A.

5 Plans tangents et approximations linéaires

Introduction. Sur une courte distance, une courbe d'équation $y = f(x)$ variant peu est presque droite. La courbe peut changer de direction, mais la droite tangente à la courbe en (x_0, y_0) d'équation $y - y_0 = f'(x_0)(x - x_0)$ garde la même pente. On dit alors que la tangente forme l'*approximation linéaire* de $y = f(x)$ au point (x_0, y_0). On écrit cela sous la forme

$$y \approx y_0 + f'(x_0)(x - x_0)$$

Que se passe-t-il avec deux variables? La fonction est $z = f(x, y)$, et son graphe, correspond à une *surface*. Si nous nous plaçons à un point sur cette surface et considérons que notre vue ne porte pas loin. Nous ne savons alors pas si la surface peut se courbe au-delà de notre vision, en prenant la forme d'une sphère ou d'un col par exemple. Pour notre vision limitée, la surface semble plate. Nous en concluons donc que nous sommes situés sur un plan (pas nécessairement horizontal) et notre tâche ici est de déterminer l'équation de ce plan que l'on qualifiera de **plan tangent** (*tangent plane*).

Convention et définition du plan tangent. On suivra la plupart du temps la convention habituelle en dénotant les coordonnées du **point de base** (*base point*) par x_0 et y_0. Sa hauteur sur la surface est donc $z_0 = f(x_0, y_0)$. D'autres lettres sont possibles : le point peut être (a, b) de hauteur w. L'indice $_0$ indique la valeur prise par les variables x, y, z ou les dérivées $\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}$ en ce point. Contrairement au cas à une variable, où la tangente en un point donné possède une pente $\frac{df}{dx}$, une fonction $f(x, y)$ de deux variables possède deux dérivées, notées $\left(\frac{\partial f}{\partial x}\right)_0$ et $\left(\frac{\partial f}{\partial y}\right)_0$ en un point (x_0, y_0) de la surface d'équation $z = f(x, y)$. Ces dérivées correspondent aux pentes du **plan tangent** à cette surface et l'équation (importante) de ce plan « linéarise » la surface d'équation $z = f(x, y)$ au voisinage du point (x_0, y_0, z_0). En résumé, le plan tangent à (x_0, y_0, z_0) possède les mêmes pentes que la surface $z = f(x, y)$ au voisinage de ce point et l'équation (linéaire) du plan tangent est donnée par

$$z - z_0 = \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0) \tag{5.1}$$

Le **vecteur normal**, ou « normale » à ce plan est $\boldsymbol{N} = \left(\frac{\partial f}{\partial x}\right)_0 \boldsymbol{i} + \left(\frac{\partial f}{\partial y}\right)_0 \boldsymbol{j} - \boldsymbol{k}$

Exemples 19. On veut trouvez le plan tangent à la surface $z = 14 - x^2 - y^2$ en $(x_0, y_0, z_0) = (1, 2, 9)$. Les dérivées sont $(\partial f/\partial x) = -2x$ et $(\partial f/\partial y) = -2y$. Lorsque $x = 1$ et $y = 2$, elles deviennent $(\partial f/\partial x)_0 = -2$ et $(\partial f/\partial y)_0 = -4$. L'équation du plan tangent est donc

$$z - 9 = -2(x - 1) - 4(y - 2) \qquad \text{ou} \qquad z + 2x + 4y = 19$$

À partir de l'équation du plan on remarque que la fonction $z = f(x, y) = 19 - 2x - 4y$ possède des dérivés -2 et -4, tout comme la surface, le plan est donc tangent. Le vecteur normal $\boldsymbol{N}$ a des composantes -2, -4, -1. L'*équation paramétrique* de la droite normale à la surface est alors

$$(x, y, z) = (1, 2, 9) + t(-2, -4, -1) \qquad \text{soit} \qquad \begin{cases} x = 1 - 2t \\ y = 2 - 4t \\ z = 9 - t \end{cases}$$

Sous cette forme, on en conclut que la droite normale passe par le point $(1, 2, 9)$ et sort le long de $\boldsymbol{N}$, perpendiculairement au plan et à la surface.

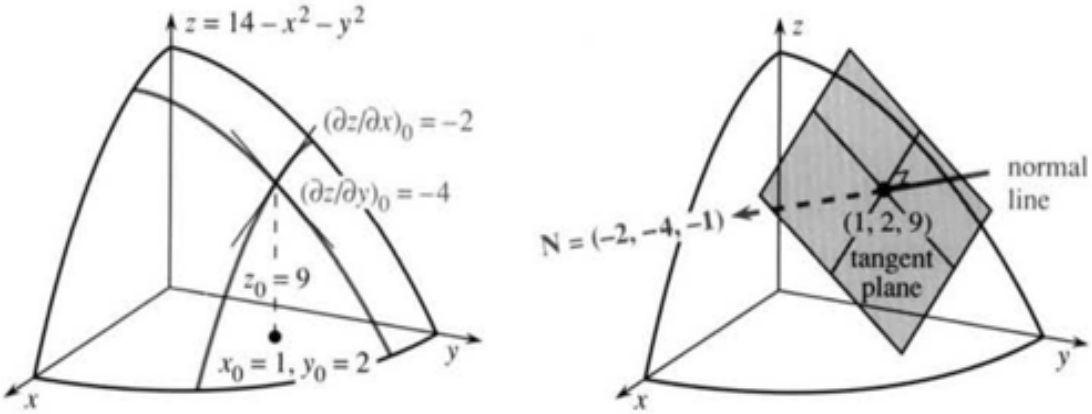


Figure 16: Le plan tangent contient les lignes tangentes x et y, perpendiculaires à $\boldsymbol{N}$.

La **Figure 16** montre plus de détails sur le plan tangent. Les lignes pointillées sont les lignes tangentes en x et en y, elles appartiennent au plan tangent. Évidemment, toutes les lignes tangentes se trouvent dans le plan tangent. Ces lignes particulières sont également tangentes aux *fonctions partielles*, où y est fixé à $y_0 = 2$ ou x est fixé à $x_0 = 1$. Le plan est « en équilibre » sur la surface et touche celle-ci au *point tangent*. Cela est d'autant plus vrai que, sur la surface, *chaque courbe passant par le point tangent est tangente au plan*. Géométriquement, la courbe monte jusqu'au point et « touche » le plan [1]. La tangente $\boldsymbol{T}$ à la courbe et la normale $\boldsymbol{N}$ à la surface sont perpendiculaires, c'est-à-dire que l'on a la relation vectorielle $\boldsymbol{T} \cdot \boldsymbol{N} = 0$ en tout point d'une surface.

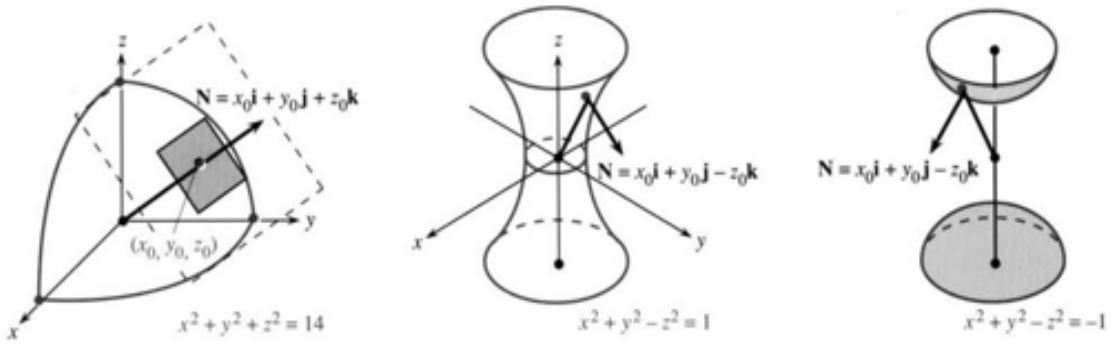


Figure 17: Plan tangent et normale $\boldsymbol{N}$ pour une sphère. Hyperboloïdes à 1 et 2 nappes.

Exemples 20. On veut trouver le plan tangent à la sphère $z^2 = 14 - x^2 - y^2$ en $(1, 2, 3)$. Au lieu d'utiliser l'équation donnée, on utilise $z = \sqrt{14 - x^2 - y^2}$. Au point $(x_0, y_0) = (1, 2)$ la hauteur est $z_0 = 3$. La surface est une sphère de rayon $\sqrt{14}$. La racine carrée possède les dérivées (compliquées)

[1] Une désignation plus savante celui-ci est *plan osculateur*. Dans une configuration de point-col, par exemple, ce plan est « touché » des deux côtés par les courbes appartenant à la surface.

suivantes

$$\frac{\partial z}{\partial x} = \frac{\partial}{\partial x}\sqrt{14 - x^2 - y^2} = \frac{\frac{1}{2}(-2x)}{\sqrt{14 - x^2 - y^2}} \qquad \text{et} \qquad \frac{\partial z}{\partial y} = \frac{\frac{1}{2}(-2y)}{\sqrt{14 - x^2 - y^2}} \tag{5.2}$$

Au point $(1, 2)$, les pentes (de la surface ou des tangentes) sont donc $-1/3$ dans la direction des x et $-2/3$ dans la direction des y. L'équation (linéaire) du plan tangent est $z - 3 = -1/3(x - 1) - 2/3(y - 2)$. Il est tentant de réécrire l'équation, en multipliant par 3 et en déplaçant tous les termes vers la gauche on obtient une forme commode à interpréter

$$1(x - 1) + 2(y - 2) + 3(z - 3) = 0 \tag{5.3}$$

correspondant à un *plan tangent à la sphère*. Si les mathématiques sont « *science des motifs* », l'équation (5.3) est un candidat de choix pour ce type d'approche. Les nombres $1, 2, 3$ apparaissent deux fois. Le point de $(x_0, y_0, z_0) = (1, 2, 3)$ appartient à ce plan (et à la sphère) et le vecteur normal est $1\,\boldsymbol{i} + 2\,\boldsymbol{j} + 3\,\boldsymbol{k}$. L'équation tangente est $1\,x + 2\,y + 3\,z = 14$. Rien de tout cela ne peut être un accident, mais la racine carrée de $14 - x^2 - y^2$ a donné à un motif simple un aspect compliqué. *Cette racine carrée n'est pas nécessaire.* On peut calculer directement la dérivée dz/dx en effectuant une *différenciation implicite*. Pour cela il suffit de différencier simplement chaque terme tel qu'il est dans l'équation originale de la surface sphérique en trois dimension, à savoir

$$x^2 + y^2 + z^2 = 14 \qquad \text{mène à} \qquad 2x + 2z\,\frac{\partial z}{\partial x} = 0 \qquad \text{et} \qquad 2y + 2z\,\frac{\partial z}{\partial y} = 0$$

En divisant par 2, les dérivées sur une sphère deviennent $\frac{\partial z}{\partial x} = -\frac{x}{z}$ et $\frac{\partial z}{\partial y} = -\frac{y}{z}$. Ce sont les mêmes que dans (5.2). À partir de cela on peut écrire l'équation du plan tangent sous une forme extrêmement symétrique, c'est-à-dire

$$z - z_0 = -\frac{x_0}{z_0}(x - x_0) - \frac{y_0}{z_0}(y - y_0) \qquad \text{ou} \qquad x_0(x - x_0) + y_0(y - y_0) + z_0(z - z_0) = 0$$

On peut lire directement dans la seconde forme que $\boldsymbol{N} = x_0\boldsymbol{i} + y_0\boldsymbol{j} + z_0\boldsymbol{k}$. En fait ce calcul montre quelque chose que nous savions déjà, c'est-à-dire que *le vecteur normal à la sphère pointe vers l'extérieur le long du rayon.*

Deuxième formulation. La sphère suggère une question importante pour d'autres surfaces. Supposons que l'équation soit de la forme $F(x, y, z) = c$ au lieu de $z = f(x, y)$. Les dérivées partielles et la tangente peuvent-elles être trouvé directement à partir de f? La réponse est *oui*. Il n'est pas nécessaire de résoudre l'équation $F(x, y, z) = c$ d'abord pour z. Les dérivées de F, calculées en (x_0, u_0, z_0), donnent une deuxième formule pour le plan tangent et le vecteur normal. Dans ce cas, on dira que le plan tangent à la surface $F(x, y, z) = c$ a pour équation linéaire

$$\left(\frac{\partial F}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial F}{\partial y}\right)_0 (y - y_0) + \left(\frac{\partial F}{\partial z}\right)_0 (z - z_0) = 0 \tag{5.4}$$

Le vecteur normal est $\boldsymbol{N} = \left(\frac{\partial F}{\partial x}\right)_0 \boldsymbol{i} + \left(\frac{\partial F}{\partial y}\right)_0 \boldsymbol{j} + \left(\frac{\partial F}{\partial z}\right)_0 \boldsymbol{k}$. On remarque que la formule (5.4) inclut le cas d'origine $z = f(x, y)$. Il suffit simplement de poser $F = f(x, y) - z$. Ses dérivées partielles sont alors $\partial f/\partial x$ et $\partial f/\partial y$ et -1 (le -1 provient de la dérivée de $-z$). L'équation (5.4) est alors la même que notre équation tangente d'origine (5.1).

Exemples 21. On souhaite trouver le plan tangent d'une *hyperboloïde* dont l'équation est donnée par $F = x^2 + y^2 - z^2 = c$. Les dérivées partielles sont $F_x = 2x, F_y = 2y, F_z = -2z$. L'équation (5.4) est donc $2x_0(x - x_0) + 2y_0(y - y_0) - 2z_0(z - z_0) = 0$ soit

$$x_0(x - x_0) + y_0(y - y_0) - z_0(z - z_0) = 0$$

Le vecteur normal est donc de la forme $\boldsymbol{N} = x_0\,\boldsymbol{i} + y_0\,\boldsymbol{j} - z_0\,\boldsymbol{k}$. Pour $c > 0$ cette hyperboloïde possède une « *nappe* » (voir la **Figure 17**), pour $c = 0$ c'est un cône et enfin pour $c < 0$, elle se divise en deux nappes.

5.1 Différentielles

Définition. Revenons à l'équation linéaire (5.1) pour le plan tangent. Celle-ci est de la forme

$$z - z_0 = \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0)$$

et c'est l'une des formules les plus importantes que nous verrons. On imagine à présent que l'on se déplace le long du plan tangent au lieu de la surface courbe. Les mouvements dans le plan sont dx et dy et dz, tandis que Δx et Δy et Δz sont des mouvements dans la surface. Les variations « d » sont régis par l'équation tangente tandis que les variations « Δ » sont régis par l'équation $z = f(x, y)$. Par convention les « d » sont qualifiés de *différentielles* le long de la tangente. Si on se rappelle le cas à une variable avec une courbe du plan décrite par $y = f(x)$, alors on a les relations suivantes

$$dy = \frac{dy}{dx}dx \quad \text{(ligne droite)} \qquad \text{et} \qquad dy \approx \frac{dy}{dx}\Delta x \quad \text{(sur la courbe)}$$

qui lient « d » et « Δ ». Maintenant, dans le cas à plusieurs variables, y et x sont toutes les deux des variables *indépendantes*. La variable *dépendante* est z. L'idée reste la même, les distances $x - x_0$ et $y - y_0$ et $z - z_0$ (sur le plan tangent) sont données par dx et dy et dz. L'équation du plan est donc

$$dz = \frac{\partial z}{\partial x}\,dx + \frac{\partial z}{\partial y}\,dy \qquad \text{ou} \qquad df = f_x dx + f_y dy \tag{5.5}$$

Remarques. Que ce soit pour dz ou df, il s'agit de *différentielles totales*. On peut écrire dz, df ou encore dw, mais on n'utilise jamais les notations de dérivées partielles ∂z et ∂f pour ce type de différentielles. Les différences dx et dy suggèrent de petits mouvements en x et y; alors que dz est le mouvement résultant dans z. Sur le plan tangent, l'équation (5.5) est exactement vraie. Il est intéressant de noter qu'une « transformée de centrage » permet de placer le point (x_0, y_0, z_0) au centre des coordonnées. Ensuite, une « transformation de zoom », ou *changement d'échelle*, permet « d'étirer » la surface de manière à ce qu'elle se confonde avec son plan tangent.

Exemples 22.

(1) L'aire d'un triangle est donner par $A = \frac{1}{2}ab\sin\theta$ où la base est de longueur b, le côté incliné est de longueur a et l'angle entre eux est θ. Vous pouvez préférer $A = \frac{1}{2}bh$, où h est la hauteur perpendiculaire $a\sin\theta$. Trouvez alors la différentielle totale dA. Dans tous les cas, nous avons besoin des dérivées partielles. Si $A = \frac{1}{2}ab\sin\theta$, alors

$$\frac{\partial A}{\partial a} = \frac{1}{2}\,b\,\sin\theta \qquad \frac{\partial A}{\partial b} = \frac{1}{2}\,a\,\sin\theta \qquad \frac{\partial A}{\partial \theta} = \frac{1}{2}\,ab\,\cos\theta$$

Cela conduit conduit immédiatement à la *différentielle totale* dA (en regardant A comme une fonction des trois variables a, b, θ)

$$dA = \frac{\partial A}{\partial a}\,da + \frac{\partial A}{\partial b}\,db + \frac{\partial A}{\partial \theta}\,d\theta = \frac{1}{2}\,b\,\sin\theta\,da + \frac{1}{2}\,a\,\sin\theta\,db + \frac{1}{2}\,ab\,\cos\theta\,d\theta$$

(2) Le volume du cylindre est $V = \pi r^2 h$. V est-il plus sensible à un changement de $r = 1.0$ à $r = 1.1$ ou de $h = 1.0$ à $h = 1.1$? Pour répondre à cela on évalue d'abord les dérivées partielles $\partial V/\partial R = 2\pi rh$ et $\partial V/\partial h = \pi r^2$. Elles mesurent la *sensibilité au changement*. Physiquement,

elles correspondent à l'aire latérale et l'aire de base du cylindre. La différentielle du volume dV correspond au volume infinitésimal d'une couche latérale sur le côté (d'épaisseur dr) et de celui d'une couche extrémale située « sur le dessus » du cylindre et d'épaisseur dh. On écrit alors

$$dV = \text{couche latérale} + \text{couche extrémale} = 2\pi rh\,dr + \pi r^2\,dh$$

Comme $r = h = 1$, la différentiel devient $dV = 2\pi\,dr + \pi\,dh$. Avec $dr = dh = 0.1$, le volume de la couche latérale est $0.2 \times \pi$ tandis que le volume de la couche extrémale n'est que $0.1 \times \pi$. Donc V est sensible à la variation dr. Pour un « cylindre court » assimilable à une pièce de monnaie (h est *petit* par rapport à r), la couche extrémale a un plus grand volume. V est plus sensible à dh. Dans notre cas, $V = \pi r^2 h$ passe à $\pi(1)^3$ à $\pi(1.1)^3$. On peut alors *comparez* ΔV à dV, cela donne

$$\Delta V = \pi \times (1.1)^3 - \pi(1)^3 = 0.331 \times \pi \qquad \text{et} \qquad dV = 2\pi \times (0.1) + \pi \times (0.1) = 0.300 \times \pi$$

La différence est $\Delta V - dV = 0.031\,\pi$. Les couches latérale et extrémale ont donc manquées un petit volume dans **Figure 18 (à gauche)**, juste au-dessus de la couches latérale et autour de la couche extrémale. L'erreur est d'ordre $(dr)^2 + (dh)^2$. Pour la fonction $V = \pi r^2 h$, la différentielle $dV = 2\pi rh\,dr + \pi r^2\,dh$ est une *approximation linéaire* du vrai changement ΔV. Nous allons expliquer cela correctement maintenant.

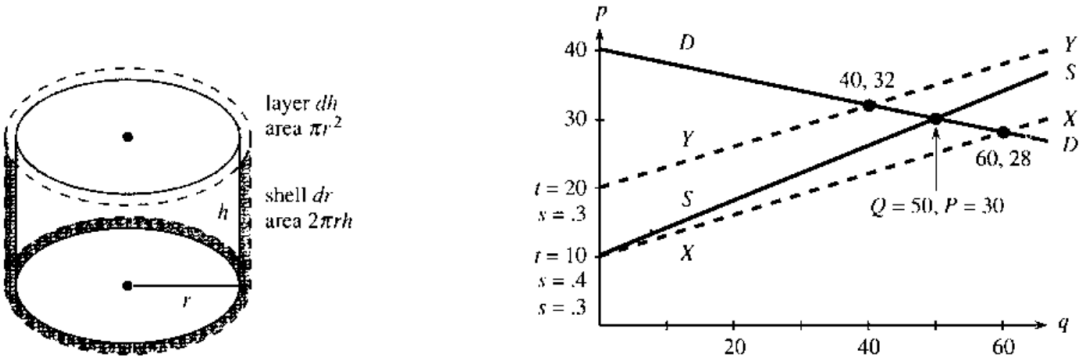


Figure 18: *Gauche* : Les couches latérale (*shell*) plus extrémale (*layer*) donne $dV = 0.300\pi$. Inclure l'anneau supérieur donne $\Delta V = 0,311\pi$. *Droite* : Quantité Q et prix$P$ mov avec les lignes.

5.2 Approximation linéaire

Introduction. De fait, *les tangentes conduisent immédiatement à des approximations linéaires.* Cela est vrai pour les plans tangents comme pour des lignes tangentes. Pour que cette approximation reste correcte il faut cependant bien veiller à ce qu'on reste dans une situation où le plan reste proche de la surface, de la même manière qu'une ligne (la tangente) reste proche de la courbe qu'elle approxime. L'avantage est que des fonctions linéaires sont plus simples (à manipuler et comprendre) que des fonctions arbitraires $f(x)$ ou $f(x, y)$ ou $F(x, y, z)$. Tout ce dont nous avons besoin pour construire ces approximations ce sont les dérivées premières *en un point* et celles-ci sont garanties d'être correctes *près de ce point.*

Formule d'approximation. Cette idée clé du calcul infinitésimal est déjà présente dans les différentielles. Sur le plan tangent on a df qui est exactement égal à $f_x\,dx + f_y\,dy$. Cependant sur

la surface courbe on a une approximation linéaire de $f(x, y)$ valable au voisinage du point tangent (x_0, y_0) de la forme

$$f(x, y) \approx f(x_0, y_0) + \left(\frac{\partial f}{\partial x}\right)_0 (x - x_0) + \left(\frac{\partial f}{\partial y}\right)_0 (y - y_0) \tag{5.6}$$

Ce qui revient à écrire $\Delta f \approx f_x \Delta x + f_y \Delta y$. Le côté droit de (5.6) est une fonction linéaire $f_L(x, y)$. En (x_0, y_0), les fonctions f et f_L ont les mêmes pentes. Alors $f(x, y)$ s'éloigne de f_L, avec une erreur du « second ordre » telle que

$$|f(x, y) - f_L(x, y)| \leq M \left[(x - x_0)^2 + (y - y_0)^2\right] \tag{5.7}$$

Cela suppose que f_{xx}, f_{xy} et f_{yy} sont continues et bornées par M le long de la ligne de (x_0, y_0) à (x, y). L'**Exemple 6.2** de la **Section 7** montre que $|f_{tt}| \ll 2M$ le long de cette ligne. Prenons l'exemple du volume d'un cylindre où les paramètres r et h passent de 1.0 à 1.1. Les dérivées secondes de $V = \pi r^2 h$ sont $V_{rr} = 2\pi h$ et $V_{rh} = 2\pi r$ et $V_{hh} = 0$. Elles sont inférieures à $M = 2.2\,\pi$. Alors (5.7) donne la borne d'erreur $2.2\pi(0.1^2 + 0.1^2) = (0.044)\pi$, un peu supérieure à l'erreur réelle $(0.031)\pi$. La remarque principale est que *l'erreur de l'approximation linéaire vient des termes quadratiques*, ce sont les premiers termes ignorés par f_L.

Exemples 23. On souhaite trouver une approximation linéaire de la fonction de distance $r = \sqrt{x^2 + y^2}$. En premier lieu on calcule les dérivées partielles $r_x = x/r$ et $r_y = y/r$. Alors $\Delta r \approx (x/r)\,\Delta x + (y/r)\,\Delta y$. Pour (x, y, r) près de $(1, 2, \sqrt{5})$ on a donc

$$\sqrt{x^2 + y^2} \approx \sqrt{1^2 + 2^2} + (x - 1)/\sqrt{5} + 2(y - 2)/\sqrt{5}$$

Si y est fixé à 2, il s'agit d'une approximation à une variable de $\sqrt{x^2 + 2^2}$. Si x est fixé à 1, c'est une approximation linéaire en y. En faisant varier les deux variables on pourrait penser que dr impliquerait dx et dy dans une racine carrée, ce n'est cependant pas le cas. La distance implique x et y dans une racine carrée, mais le « changement de distance » est *linéaire* en Δx et Δy, à une première approximation près. On note également un *point stationnaire* en $x = 0, y = 0$ où $r_x = r_y = 0$. Tout mouvement à partir de ce point $(0, 0)$ donne $\Delta r = \sqrt{(\Delta x)^2 + (\Delta y)^2}$ et ici on note que la racine carrée est de retour. La raison est que les dérivées partielles x/r et y/r ne sont *pas continues* en $(0, 0)$. Le cône d'équation $r = \sqrt{x^2 + y^2}$ possède en effet un sommet pointu sans plan tangent, en d'autres termes *l'approximation linéaire ne marche pas* au point $(0, 0)$.

L'exemple suivant montre comment approximer Δz à partir de Δx et Δy, lorsque l'équation est sous la forme $F(x, y, z) = c$. Dans ce cas on utilise les dérivées implicites (5.4) au lieu des dérivées explicites (5.1). L'idée est toujours la même : regardez l'équation tangente comme un moyen de trouver Δz, au lieu d'une équation pour z. Voici un exemple avec une autre notation.

Exemples 24. À partir de $F(x, y, z) = -x^2 - y^2 + z^2 = 0$ trouver une approximation linéaire de Δz. En utilisant les dérivées de F on obtient $-2x\,\Delta x - 2y\,\Delta y + 2z\,\Delta z \approx 0$. Ensuite, on résout pour Δz, ce qui donne $\Delta z \approx (x/z)\,\Delta x + (y/z)\,\Delta y$, le même que dans l'**Exemple 4.2**.

L'exemple suivant, issu de l'économie, correspond au problème (classique) d'offre et de la demande.

Exemples 25. Comment le *prix d'équilibre* change-t-il lorsque la *courbe d'offre* change? Le prix d'équilibre est à l'intersection des courbes d'offre et de demande (c'est-à-dire lorsque *offre* = *demande*), voir la **Figure 18 (droite)**. Lorsque le prix p augmente, la demande q baisse (la pente

est de -0.2) soit

$$\textit{ligne de demande } DD: \quad p = -0.2\,q + 40$$

L'offre (notée q) augmente avec le prix. La pente s est positive (ici $s = 0.4$):

$$\textit{ligne d'approvisionnement } SS: \quad p = s\,q + t = 0.4\,q + 10$$

Ces lignes se rencontrent au **prix d'équilibre** $P = 30\$$. La quantité $Q = 50$ est disponible à un prix P (sur SS) et demandée à ce même prix P (sur DD), cette quantité est donc vendue. Comment les dérivées partielles interviennent dans ce problème? La réalité est que ces lignes DD et SS ne sont pas fixes pour toujours. Les changements technologiques, les changements de concurrence ainsi que la valeur de l'argent change. Par conséquent, les lignes bougent et donc le point de croisement (Q, P) se déplace également. Il faut bien comprendre que des dérivées se cachent dans toutes ces phrases. La remarque principale est que *le prix d'équilibre P est une fonction de s et t*. Réduire s grâce à une meilleure technologie abaisse la ligne d'approvisionnement à

$$p = 0.3\,q + 10$$

La ligne de demande n'a pas changée. Le client est toujours aussi impatient ou avare. Mais le prix P et la quantité Q sont différents. Le nouvel équilibre est à $Q = 60$ et $P = 28\$$, où la nouvelle ligne XX croise DD. Si la technologie est chère, le fournisseur augmentera t en réduisant s. La ligne YY est

$$p = 0.3\,q + 20$$

Cela donne un équilibre plus élevé $P = 32\$$ à une quantité inférieure $Q = 40$, c'est-à-dire que la demande était trop faible pour la technologie. On désire à présent déterminer $\partial P/\partial s$ et $\partial P/\partial T$. La difficulté est que P n'est pas donné en fonction de $\mathcal{S}$ et t. On prend donc des dérivées implicites des équations offre = demande, soit

$$\begin{array}{rlll} \textit{offre} = \textit{demande}: & P = -0.2\,Q + 40 & = s\,Q + t & \\ \hookrightarrow \text{dérivée par rapport à } s: & P_s = -0.2\,Q_s & = s\,Q_s + Q & (\text{on note que } t_s = 0) \\ \hookrightarrow \text{dérivée par rapport à } t: & P_t = -0/2\,Q_t & = s\,Q_t + 1 & (\text{on note que } t_t = 1) \end{array} \tag{5.8}$$

Maintenant on remplace par $s = 0.4$, $t = 10$, $P = 30$, $Q = 50$. C'est le point de départ autour duquel on trouve une approximation linéaire. Les deux dernières équations donnent $P_s = 50/3$ et $P_t = 1/3$ (voir **Exercice 21**). L'approximation linéaire est donc

$$P = 30 + 50(s - 0.4)/3 + (t - 10)/3 \tag{5.9}$$

Il s'agit d'un exemple concret où les équations sont linéaires et leurs dérivées sont faciles à obtient. Néanmoins il demeure une difficulté puisqu'il n'y a pas de formule explicite pour P et donc la fonction $P(s, t)$ n'est pas connue. Au lieu d'un point sur une surface, nous suivons l'intersection de deux lignes. *La solution change lorsque l'équation change* et *La dérivée de la solution vient de la dérivée de l'équation.*

Résumé. En conclusion on notera que le fondement de cette section est l'équation (5.1) pour le plan tangent. Tout repose sur ce que cette relation contient c'est-à-dire la *différentielle totale*, l'*approximation linéaire* et la sensibilité aux petits variations ou changements. Les sections ultérieures abordent la dérivation composée, ou *chain rule*, les « dérivées directionnelles » ainsi que les « gradients ». On gardera à l'esprit que l'idée centrale du calcul différentiel est $\Delta f \approx f_x \Delta x + f_y \Delta y$.

5.3 Méthode de Newton pour deux équations

Contexte. L'approximation linéaire peut également être utilisée *pour résoudre des équations.* Pour savoir où une fonction est nulle, regardez d'abord où son approximation s'annule. Pour savoir où un graphe intersecte le plan xy, regardez où son plan tangent intersecte le plan xy. On se rappelle la *méthode de Newton* pour résoudre $f(x) = 0$. L'estimation de départ est x_n et autour de ce point, $f(x)$ est proche de $f(x_n) + (x - x_n)f'(x_n)$. Ensuite on a $f(x) = 0$ à la prochaine estimation donc $x_{n+1} = x_n - f(x_n)/f'(x_n)$. C'est là que la droite tangente coupe l'axe des x.

Approximations. Avec deux variables, l'idée est la même, mais deux inconnues x et y nécessitent *deux équations.* Nous résolvons $g(x, y) = 0$ et $h(x, y) = 0$. Les deux fonctions ont des approximations linéaires qui partent du point courant (x_n, y_n), où les dérivées sont calculées, on a donc

$$g(x,y) \approx g(x_n,y_n) + \left(\frac{\partial g}{\partial x}\right)(x - x_n) + \left(\frac{\partial g}{\partial y}\right)(y - y_n)$$
$$h(x,y) \approx h(x_n,y_n) + \left(\frac{\partial h}{\partial x}\right)(x - x_n) + \left(\frac{\partial h}{\partial y}\right)(y - y_n)$$

L'idée naturelle est d'*égaliser ces approximations à zéro.* Cela donne des équations linéaires pour $x - x_n$ et $y - y_n$. Ce sont les pas Δx et Δy qui nous amènent à la prochaine estimation dans le cadre de la méthode de Newton. En résumé on peut dire que cette méthode, permettant de résoudre $g(x, y) = 0$ et $h(x, y) = 0$ simultanément, est associée à des équations linéaires avec des pas Δx et Δy qui vont de (x_n, y_n) à (x_{n+1}, y_{n+1}). On a les relation linéaires

$$\left(\frac{\partial g}{\partial x}\right)\Delta x + \left(\frac{\partial g}{\partial y}\right)\Delta y = -g(x_n,y_n) \quad \text{et} \quad \left(\frac{\partial h}{\partial x}\right)\Delta x + \left(\frac{\partial h}{\partial y}\right)\Delta y = -h(x_n,y_n) \tag{5.10}$$

Exemples 26. Les équations $g = x^3 - y = 0$ et $h = y^3 - x = 0$ ont 3 solutions $(1, 1)$, $(0, 0)$, $(-1, -1)$. On commence en différents points (x_0, y_0) et à partir de ceux-ci, la prochaine estimation donne $x_1 = x_0 + \Delta x$, $y_1 = y_0 + \Delta y$. Il est intéressant de repérer quelle solution la méthode de Newton choisira, lorsque celle-ci converge bien entendu. On considère trois cas.

(1) Supposons $(x_0, y_0) = (2, 1)$. À ce stade, $g = 2^3 - 1 = 7$ et $h = 1^3 - 2 = -1$. Les dérivées sont $g_x = 3x^2 = 12$, $g_y = -1$, $h_x = -1$, $h_y = 3y^2 = 3$. Les pas Δx et Δy sont solutions de (5.10), soit

$$\begin{cases} 12\Delta x - \Delta y = -7 \\ -\Delta x + 3\Delta y = +1 \end{cases} \Rightarrow \begin{cases} \Delta x = -4/7 \\ \Delta y = -1/7 \end{cases} \Rightarrow \begin{cases} x_1 = x_0 + \Delta x = 10/7 \\ y_1 = y_0 + \Delta y = 8/7 \end{cases}$$

Ce nouveau point $(10/7, 8/7)$ est plus proche de la solution en $(1, 1)$. Le point suivant est $(1.1, 1.05)$ et la convergence est claire. La convergence est par la suite rapide.

(2) Si on commence en $(x_0, y_0) = (1/2, 0)$ alors on trouve $g = 1/8$ et $h = -1/2$, soit

$$\begin{cases} (3/4)\,\Delta x - \Delta y = -1/8 \\ -\Delta x + 0\Delta y = +1/2 \end{cases} \Rightarrow \begin{cases} \Delta x = -1/2 \\ \Delta y = +1/4 \end{cases} \Rightarrow \begin{cases} x_1 = x_0 + \Delta x = 0 \\ y_1 = y_0 + \Delta y = -1/4 \end{cases}$$

La méthode de Newton nous fait passer de $(1/2, 0)$ sur l'axe x à $(0, -1/4)$ sur l'axe y. L'étape suivante passe à $(1/32, 0)$, de retour sur l'axe x. On dit que nous sommes situé dans le *bassin d'attraction* de $(0, 0)$.

(3) On commence maintenant plus loin sur l'axe en $(1, 0)$, où $g = 1$ et $h = -1$, alors

$$\begin{cases} 3\Delta x - \Delta y = -1 \\ -\Delta x + 0\Delta y = +1 \end{cases} \Rightarrow \begin{cases} \Delta x = -1 \\ \Delta y = -2 \end{cases} \Rightarrow \begin{cases} x_1 = x_0 + \Delta x = 0 \\ y_1 = y_0 + \Delta y = -2 \end{cases}$$

La méthode de Newton nous fait passer de $(1,0)$ à $(0,-2)$ à $(16,0)$ et ne converge donc plus, elle est en échec. Ce danger est toujours présent, quand on part loin d'une solution.

Discussion. Il est important de noter que même un petit ordinateur (avec un processeur modeste) permet de découvrir des comportements étonnants. On peut partir de centaines de points de départ (x_0, y_0) différents et suivre la méthode de Newton. Chaque solution possède un *bassin d'attraction*, contenant tous les (x_0, y_0) conduisant à cette solution. Il existe aussi un bassin menant à l'infini.

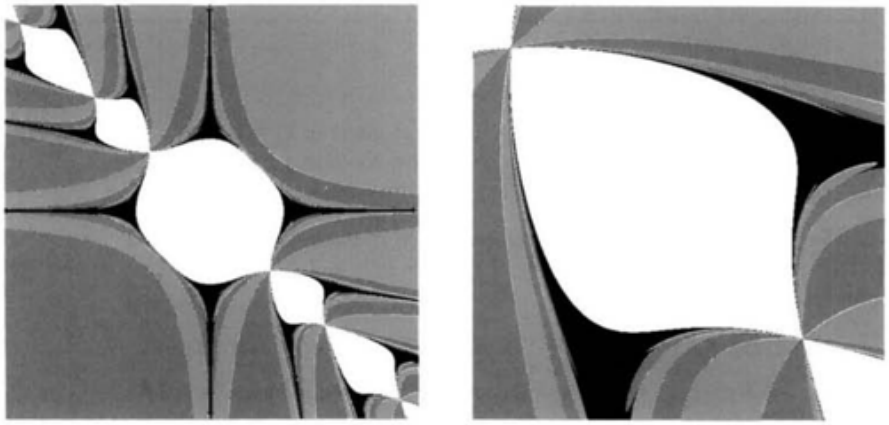


Figure 19: Les bassins d'attraction vers $(1,1)$, $(0,0)$, $(-1,-1)$ et l'infini.

Sur la **Figure 19** les bassins sont complètement mélangés, une figure en couleur ferait apparaître des **fractales**. Le comportement le plus extrême se situe à la frontière entre les bassins, lorsque la méthode de Newton ne peut pas décider de la voie à suivre. C'est dans ce type de situation que l'on utilise le terme de *comportement chaotique* pour caractériser le fait qu'un petit changement initial peut entrainer plusieurs comportements ultérieurs très différents et parfois extrêmes. Le chaos est un mouvement irrégulier qui suit une règle définie. Dans notre cas, c'est la méthode de Newton qui détermine une *itération* de chaque point (x_n, y_n) au suivant. Dans les problèmes scientifiques bien posés, cette méthode converge normalement vers la solution que l'on recherche (en commencant assez près). Mais l'ordinateur permet d'étudier des itérations à partir de points plus lointains. Cette démarche correspond à une partie des mathématiques appelée les *mathématiques expérimentales* où chaque expérience effectuée est susceptible d'être originale. On peut également retrouver un comportement chaotique en essayant de résoudre $x^2 + 1 = 0$. Dans ce cas précis on pourrait penser que la méthode de Newton est un échec cependant, c'est la meilleure méthode pour résoudre des équations non linéaires. L'erreur est mise au carré lorsque l'algorithme converge, car les approximations linéaires ont des erreurs d'ordre $(\Delta x)^2 + (\Delta y)^2$. Chaque étape double le nombre de chiffres corrects, près de la solution. Cette exemple montre pourquoi il est important d'être relativement proche de la solution.

6 Dérivées directionnelles et gradients

Introduction. Lorsque x change, nous savons à présent comment $f(x,y)$ change puisque la dérivée partielle $\partial f/\partial x$ traite y comme une constante. De même lors de l'évaluation de $\partial f/\partial y$, la variable x est considérée constante et donne la pente dans la direction des y. Mais la direction des x (est-ouest) et celle des y(nord-sud) ne sont pas les seules directions possibles. On peut par exemple suivre une ligne faisant 45° par rapport à l'horizontale correspondant pour laquelle $\Delta x = \Delta y$. Cette idée est renforcée par le fait que, avant de dessiner des axes, aucune direction n'est privilégiée et, de plus, le graphe d'une surface possède des pentes dans *toutes* les directions. Sur cette surface, on recherche donc par le calcul le taux d'accroissement (ou la pente) dans une direction donnée auquel on va associer une *dérivée directionnelle*. Dans le cas de la ligne à 45° on pourrait être tenter de diviser Δf par Δx, mais cela reviendrait à construire la dérivée dans la direction des x.

Formulation. Tentons d'énoncer le problème clairement. On se donne une fonction $f(x,y)$ autour d'un point $P = (x_0, y_0)$. On se donne également une direction spécifique sous la forme d'un vecteur unitaire $\boldsymbol{u}$. Il doit alors exister une définition naturelle de $D_{\boldsymbol{u}} f$, la dérivée dans la direction $\boldsymbol{u}$. Pour calculer cette pente en P, nous avons besoin d'une formule, de préférence basée sur $\partial f/\partial x$ et $\partial f/\partial y$, que nous connaissons déjà. Notez que la direction des 45° correspond au vecteur (unitaire) $\boldsymbol{u} = \boldsymbol{i}/\sqrt{2} + \boldsymbol{j}/\sqrt{2}$. La racine carrée de 2 va donc entrer dans la dérivée. Cela montre d'emblée que diviser Δf par Δx est faux. Nous devons diviser par la longueur du pas Δs dans la direction à 45° et pas par Δx.

Exemples 27. On reste sur la surface $z = xy$. Quand (x,y) se déplace d'une distance Δs dans la direction 45° de $(1,1)$, que vaut $\Delta z/\Delta s$? Le pas est la quantité Δs fois le vecteur unitaire $\boldsymbol{u}$. À partir de $x = y = 1$, l'étape se termine à $x = y = 1 + \Delta s/\sqrt{2}$. (Les composants de $\boldsymbol{u}\,\Delta s$ sont $\Delta s/\sqrt{2}$.) Alors $z = xy$ peut s'écrire

$$z = (1 + \Delta s/\sqrt{2})^2 = 1 + \sqrt{2}\,\Delta s + \frac{1}{2}(\Delta s)^2 \quad \text{ce qui signifie} \quad \Delta z = \sqrt{2}\,\Delta s + \frac{1}{2}(\Delta s)^2$$

Le rapport $\Delta z/\Delta s$ approche $\sqrt{2}$ comme $\Delta s \to 0$. C'est la pente dans la direction des 45°.

Définition. La dérivée de f dans la direction unitaire $\boldsymbol{u}$ en P est notée $D_{\boldsymbol{u}} f(P)$ et définie par

$$D_{\boldsymbol{u}} f(P) = \lim_{\Delta s \to 0} \frac{\Delta f}{\Delta s} = \lim_{\Delta s \to 0} \frac{f(P + \boldsymbol{u}\,\Delta s) - f(P)}{\Delta s}$$

Le pas à partir de $P = (x_0, y_0)$ a une longueur Δs et cela amène au point $(x_0 + u_1 \Delta s, y_0 + u_2 \Delta s)$. Ainsi nous calculons le changement Δf et divisons par Δs. Il est également possible de formuler la dérivée directionnelle en fonction de f_x et f_y et, dans ce sens, la formule (6.2) ci-dessous fait gagner du temps. Dans ce sens, on peut facilement retrouver les dérivées par rapport à x et y à partir de la définition ci-dessus. En effet, la direction des x est $\boldsymbol{u} = (1,0)$ donc $\boldsymbol{u}\,\Delta s$ devient $(\Delta s, 0)$ et on récupère $\partial f/\partial x$, soit

$$\frac{\Delta f}{\Delta s} = \frac{F(x_0 + \Delta s, y_0) - F(x_0, y_0)}{\Delta s} \qquad \text{approche} \qquad D_{(1,0)} f = \frac{\partial f}{\partial x}$$

De même $D_{\boldsymbol{u}} f = \partial f/\partial y$, quand $\boldsymbol{u} = (0,1)$, c'est-à-dire dans la direction des y. Qu'est-ce que $D_{\boldsymbol{u}} f$ quand $\boldsymbol{u} = (0,-1)$? C'est la direction vers les y négatifs, donc $D_{\boldsymbol{u}} f = -\partial f/\partial y$.

6.1 Calcul de la dérivée directionnelle

Simplification. Géométriquement, $D_{\boldsymbol{u}}f$ est la pente de la surface $z = f(x, y)$ dans la direction $\boldsymbol{u}$. Comme mentionnée précédemment cette pente peut être évaluée à partir des dérivées $\partial f/\partial x$ et $\partial f/\partial y$, dans les deux directions spéciales des x et des y. Dans ce sens la *dérivée directionnelle* $D_{\boldsymbol{u}}f$ de la fonction $f(x, y)$ dans la direction unitaire $\boldsymbol{u} = (u_1, u_2) = u_1\,\boldsymbol{i} + u_2\,\boldsymbol{j}$, avec $\|\boldsymbol{u}\| = 1$, est

$$D_{\boldsymbol{u}}f = \frac{\partial f}{\partial x}\,u_1 + \frac{\partial f}{\partial y}\,u_2 \tag{6.1}$$

Cette expression se généralise facilement à un nombre arbitraire de variables. En effet, en posant $\boldsymbol{u} = (u_1, u_2, \ldots, u_n) \in \mathbb{R}^n$ et pour $f \equiv f(x_1, x_2, \ldots, x_n)$ avec n un entier fini, on a

$$D_{\boldsymbol{u}}f = \frac{\partial f}{\partial x_1}\,u_1 + \cdots + \frac{\partial f}{\partial x_n}\,u_n = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\,u_i \tag{6.2}$$

On notera que si l'on souhaite calculer la dérivée directionnelle dans une direction donnée par un **vecteur non unitaire**, noté $\boldsymbol{v}$ par exemple, il faut d'abord construire un vecteur $\boldsymbol{u}$ unitaire parallèle à $\boldsymbol{v}$. Pour ce faire on calcule simplement que

$$\boldsymbol{u} = \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|}$$

Ce processus de normalisation reste inchangé quelque soit le nombre de variables.

Commentaire. Cette formule se retrouve facilement, pour deux ou plusieurs variables, à partir de l'approximation linéaire de Δf. Pour deux variables, par exemple, on peut exprimer les pas Δx et Δy en fonction du pas Δs. Cela donne

$$\Delta f \approx \frac{\partial f}{\partial x}\,\Delta x + \frac{\partial f}{\partial y}\,\Delta y = \frac{\partial f}{\partial x}\,u_1\,\Delta s + \frac{\partial f}{\partial y}\,u_2\,\Delta s$$

En divisant par Δs et en laissant Δs approcher de zéro, on retrouve la formule (6.2) qui est la limite de $\Delta f/\Delta s$, car l'approximation devient exacte dans cette limite. Un argument plus prudent garantit cette limite à condition que f_x et f_y soient continues au point de base (x_0, y_0). L'idée principale ici est que *les pentes dans toutes les directions sont connues à partir des pentes dans deux directions.*

Exemples 28.

(1) On considère, encore une fois, la surface $f = xy$ et $P = (1, 1)$ et $\boldsymbol{u} = (1/\sqrt{2}, 1/\sqrt{2})$. Afin de trouver $D_{\boldsymbol{u}}f(P)$ on calcule d'abord les dérivées $f_x = y$ et $f_y = x$. Celles-ci sont donc égales à 1 en P. La dérivée dans la direction des 45° est donc bien

$$D_{\boldsymbol{u}}f(P) = f_x u_1 + f_y u_2 = 1 \times (1/\sqrt{2}) + 1 \times (1/\sqrt{2}) = \sqrt{2}$$

(2) La fonction linéaire $f = 3x + y + 1$ possède une pente (constante) donnée par $D_{\boldsymbol{u}}f = 3u_1 + u_2$. Comme la direction des x correspond au vecteur $u = (1, 0)$ alors dans ce cas $D_{\boldsymbol{u}}f = 3 = \partial f/\partial x$. De la même manière, dans la direction des y on trouve $D_{\boldsymbol{u}}f = 1$. Deux autres directions jouent également un rôle spécial, celles qui sont *parallèles* et *perpendiculaires* aux *lignes de niveau* de $f(x, y)$. En effet

- Dans une direction parallèle aux lignes de niveau (*level direction*) on a $D_{\boldsymbol{u}}f = 0$ car f est constant.

- Dans une direction perpendiculaire aux lignes de niveau le module de $D_{\boldsymbol{u}}f$ est aussi grand que possible (avec $u_1^2 + u_2^2 = 1$), cela indique la pente la plus raide (*steepest direction*).

On souhaite trouver ces directions pour la fonction $f = 3x + y + 1$. La direction parallèle $\boldsymbol{u}$ à la ligne de niveau correspond à $D_{\boldsymbol{u}}f = 3u_1 + u_2 = 0$ donc le vecteur $\boldsymbol{u}$ doit être proportionnel à $(1, -3)$. A posteriori on remarque que le remplacement de x par 1 et y par -3 ne produit aucun changement dans $f = 3x + y + 1$. Afin de trouver la direction avec la pente la plus raide, on recherche une direction orthogonale aux lignes de niveau et celles-ci sont parallèles au vecteur $(1, -3)$. À partir de là, on remarque que le produit scalaire de $(3, 1)$ et $(1, -3)$ est nul, la direction de la pente est donc proportionnelle au vecteur $\boldsymbol{u} = (3, 1)$. Au passage, on aurait également pu choisir $(-3, -1)$. Ensuite, pour faire de $(3, 1)$ un vecteur unitaire, on le divise par sa longueur (norme) $\sqrt{10}$. On notera aussi que les composantes de $\boldsymbol{u}$ correspondent aux dérivées partielles $f_x = 3$ et $f_y = 1$. Les deux vecteurs candidats $(3, 1)$ et $(-3, 1)$ correspondent en fait à la montée et à la descente associées à la pente la plus raide. Ces deux vecteurs sont donc, naturellement, parallèles et de sens opposé. Pour les distinguer on effectue les calculs suivants.

- *Montée la plus raide* : $D_{\boldsymbol{u}}f = 3 \times (3/\sqrt{10}) + 1 \times (1/\sqrt{10}) = 10/\sqrt{10} = \sqrt{10} > 0$
- *Descente la plus forte* : Inverser vers $\boldsymbol{u} = (-3/\sqrt{10}, -1/\sqrt{10})$ et $D_{\boldsymbol{u}}f = -\sqrt{10} < 0$

Par exemple les lignes de niveau autour d'une montagne suivent $D_{\boldsymbol{u}}f = 0$. Les ruisseaux sont perpendiculaires. Sur un plan comme $f = 3x + y + 1$, ces directions restent les mêmes en tous points (**Figure 20**). Sur une montagne, la direction la plus raide change à mesure que les pentes changent car la dérivée directionnelle n'est pas constante.

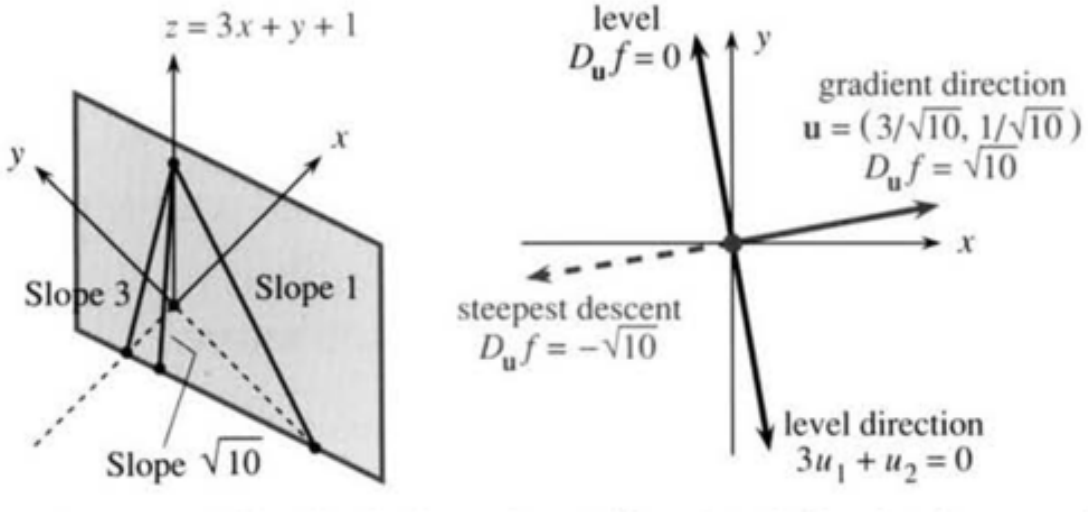


Figure 20: La direction la plus raide est le long du dégradé. La direction du niveau est perpendiculaire.

6.2 Le vecteur gradient

Introduction. On considère à nouveau l'expression de la dérivée directionnelle $D_{\boldsymbol{u}} f = f_x u_1 + f_y u_2$. Après inspection, il n'est pas difficile de voir que celle-ci peut s'écrire sous la forme d'un *produit scalaire* de deux vecteurs. Un vecteur des vecteurs est $\boldsymbol{u} = (u_1, u_2)$, définissant la direction, tandis que l'autre vecteur est (f_x, f_y), provenant de la fonction. Ce dernier vecteur est appelé le *gradient* de la fonction f.

Définition. Le **gradient** d'une fonction f, noté $\boldsymbol{grad}\, f \equiv \boldsymbol{\nabla} f$, est le vecteur dont les composantes sont les dérivées partielles de f par rapport à chacune de ses variables. Par exemple, en deux et trois dimensions, les gradients des fonctions $f(x, y)$ et $F(x, y, z)$ sont de la forme

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x}\,\boldsymbol{i} + \frac{\partial f}{\partial y}\,\boldsymbol{j} \qquad \text{et} \qquad \boldsymbol{\nabla} F = \frac{\partial F}{\partial x}\,\boldsymbol{i} + \frac{\partial F}{\partial y}\,\boldsymbol{j} + \frac{\partial F}{\partial x}\,\boldsymbol{k}$$

La généralisation pour un nombre de variables arbitraire est immédiate. En effet, le système de vecteurs $\{\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n\}$ formant une base de $\mathbb{R}^n$ et $f \equiv f(x_1, x_2, \ldots, x_n)$ avec n un entier fini, on a

$$\boldsymbol{\nabla} f = \frac{\partial f}{\partial x_1}\,\boldsymbol{u}_1 + \cdots + \frac{\partial f}{\partial x_n}\,\boldsymbol{u}_n = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\,\boldsymbol{u}_i$$

Le symbole *nabla* $\boldsymbol{\nabla}$ est une notation concise qui est lu comme « grad » (le terme « del » est également employé en anglais).

Commentaire. Pour la fonction linéaire $f = 3x + y + 1$, le gradient est le vecteur constant $\boldsymbol{\nabla} f = (3, 1) = 3\,\boldsymbol{i} + \boldsymbol{j}$ et cela caractérise le déplacement nécessaire afin de « monter » lorsque l'on se situe sur le plan. Pour la fonction non linéaire $f = x^2 + xy$, le gradient est le vecteur non constant $\boldsymbol{\nabla} f = (2x + y, x)$. Notez que $\boldsymbol{\nabla} f$ partage les deux dérivées de $\boldsymbol{N} = (f_x, f_y, -1)$. Mais le gradient n'est pas le vecteur normal, $\boldsymbol{N}$ est en effet en trois dimensions, pointant loin de la surface $z = f(x, y)$. Le vecteur gradient ne possède que deux composantes, car f ne dépend que de deux variables, et donc il appartient au plan xy. Le gradient $\boldsymbol{\nabla} f$ indique simplement la *direction* dans laquelle la pente augmente en un point donné, et sa longueur $\|\boldsymbol{\nabla} f\|$ donne la magnitude de la *pente*, c'est-à-dire, plus $\|\boldsymbol{\nabla} f\|$ est grand plus la pente est élevée en ce point. On peut à présent définir la dérivée directionnelle directement à partir de la notion de gradient.

Dérivée directionnelle et gradient. La *dérivée directionnelle* d'une fonction f dans la direction donnée par le vecteur unitaire $\boldsymbol{u}$ est

$$D_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u} \tag{6.3}$$

Une *ligne de niveau* est donc, par (6.3), perpendiculaire au vecteur gradient $\boldsymbol{\nabla} f$, puisque

$$D_{\boldsymbol{v}} f = 0 = \boldsymbol{\nabla} f \cdot \boldsymbol{v}$$

où $\boldsymbol{v}$ indique un vecteur unitaire *parallèle* à la ligne de niveau. Ici, tout déplacement dans la direction $\boldsymbol{v}$ se fait à « hauteur » constante. De même, par (6.3), la *pente* $D_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u}$ est la plus grande possible lorsque $\boldsymbol{u}$ est parallèle à $\boldsymbol{\nabla} f$. Comme $\boldsymbol{u}$ est unitaire, cette pente maximale est donc le module du gradient de f, c'est-à-dire

$$\max_{\boldsymbol{u}} D_{\boldsymbol{u}} f = \|\boldsymbol{\nabla} f\|$$

et le vecteur unitaire $\boldsymbol{u}$ est donc de la forme

$$\boldsymbol{u} = \frac{\boldsymbol{\nabla} f}{\|\boldsymbol{\nabla} f\|} \qquad \text{de sorte que} \qquad \boldsymbol{\nabla} f \cdot \boldsymbol{u} = \frac{\|\boldsymbol{\nabla} f\|^2}{\|\boldsymbol{\nabla} f\|} = \|\boldsymbol{\nabla} f\|$$

Dans l'exemple $f = 3x + y + 1$ avait $\boldsymbol{\nabla} f = (3, 1)$. Sa pente la plus raide est dans la direction $\boldsymbol{u} = (3, 1)/\sqrt{10}$. La pente maximale était de $\sqrt{10}$. En effet,

$$\|\boldsymbol{\nabla} f\| = \sqrt{f_x^2 + f_y^2} = \sqrt{9 + 1}$$

Le point important à garder à l'esprit est que *le maximum de* $\boldsymbol{\nabla} f \cdot \boldsymbol{u}$ *est la longueur* $\|\boldsymbol{\nabla} f\|$. Dans les exemples non linéaires, le gradient, les directions et la pente varieront. Mais si on prend un point particulier, par exemple celui sur la **Figure 21**, alors près de ce point, comme pour n'importe quel autre point, l'image linéaire prend le dessus. Sur le graphique de f les vecteurs spéciaux sont la direction du niveau $\boldsymbol{L} = (f_x, -f_x, 0)$, la direction ascendante $\boldsymbol{U} = (f_x, f_y, f_x^2 + f_y^2)$ et la normale $\boldsymbol{N} = (f_x, f_y, -1)$.

Exemples 29. Le gradient de $f(x, y) = (14 - x^2 - y^2)/3$ est $\boldsymbol{\nabla} f = (-2x/3, -2y/3)$. Sur la surface, le vecteur normal est $\boldsymbol{N} = (-2x/3, -4y/3, -1)$. Au point $(1, 2, 3)$, ce vecteur est $\boldsymbol{N} = (-2/3, -4/3, -1)$. Au point $(1, 2)$, en bas de la base, le gradient est $(-2/3, -4/3)$. La longueur de $\boldsymbol{\nabla} f$ est la pente $\sqrt{20}/3$.

La plupart du temps un randonneur ne monte pas tout droit. Une « pente » de $\sqrt{20}/3$ est assez raide (presque 150%). Pour estimer la pente dans d'autres directions, on peut mesurer la distance le long du chemin entre deux lignes de niveau. Si $\Delta f = 1$ sur une distance $\Delta s = 3$ la pente est d'environ 1/3. Ce calcul n'est exact que dans la limite, lorsque $\Delta f/\Delta s$ devient $D_{\boldsymbol{u}} f$.

Exemples 30. On veut déterminer le gradient de la fonction

$$f(x, y, z) = xy + yz + xz$$

On peut généraliser le calcul de $f(x, y)$ à $f(x, y, z)$ et le gradient est maintenant le vecteur tridimensionnel (f_x, f_y, f_z). On trouve alors

$$\boldsymbol{\nabla} f = (y + z, x + z, x + y)$$

Pour dessiner le graphique de $w = f(x, y, z)$, il faudrait une image à quatre dimensions, avec des axes dans les directions $xyzw$. On note également les dimensions. Le graphe est une « surface » de dimensions 3, dans un espace de dimension 4. Le gradient est en dessous dans la base tridimensionnelle. Les lignes de niveaux de f proviennent de $xy + yz + zx = c$ et elle sont bidimensionnels. Le vecteur gradient est perpendiculaire à cet ensemble de lignes de niveaux (qui sont en dimensions 3). Au passage, il ne faut le gradient avec le vecteur $\boldsymbol{N}$. En quatre dimension, le vecteur normal est $\boldsymbol{N} = (f_x, f_y, f_z, -1)$ et toujours perpendiculaire à la surface dans un espace dimensionnel 4.

Exemples 31. On veut trouver $\boldsymbol{\nabla} z$ quand $z(x, y)$ est donné implicitement par la fonction $F(x, y, z) = x^2 + y^2 - z^2 = 0$. Dans ce cas, on trouve $z = \pm\sqrt{x^2 + y^2}$ et les dérivées sont $\pm x/\sqrt{x^2 + y^2}$ et $\pm y/\sqrt{x^2 + y^2}$, qui vont servir à déterminer $\boldsymbol{\nabla} z$. Mais on peut procéder autrement, en effet, pour trouver ce gradient plus rapidement, il peut différencier $F(x, y, z)$ tel quel, puis divisez par f_z, soit

$$F_x\, dx + F_y\, dy + F_z\, dz = 0 \qquad \text{ou} \qquad dz = (-F_x\, dx - F_y\, dy)/F_z$$

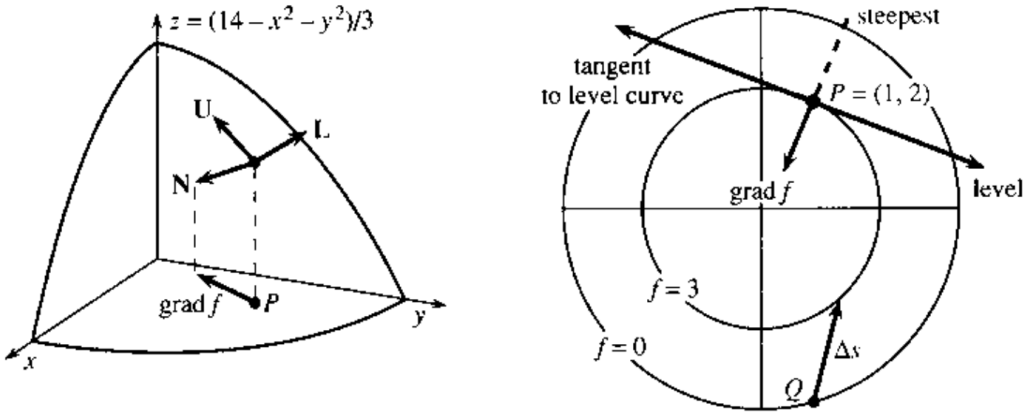


Figure 21: $\boldsymbol{N}$ perpendiculaire à la surface et $\boldsymbol{\nabla} f$ perpendiculaire à la ligne de niveau (dans la base).

Le gradient est donc $\boldsymbol{\nabla} z = (-F_x/F_z, -F_y/F_z)$. Ces dérivées sont évaluées en (x_0, y_0). Le calcul n'a pas besoin de la fonction explicite $z = f(x, y)$. On a

$$F = x^2 + y^2 - z^2 \quad \Rightarrow \quad F_x = 2x, F_y = 2y, F_z = -2z \quad \Rightarrow \quad \boldsymbol{\nabla} z = (x/z, y/z)$$

La surface correspond à celle d'un cône d'axe Oz. On peut se déplacer vers le somment du cône en suivant la direction $(x/z, y/z)$. Cette direction du gradient est radiale et orienté vers l'extérieur. La pente du cône est la longueur du vecteur gradient. On trouve

$$\boldsymbol{\nabla} z = \sqrt{(x/z)^2 + (y/z)^2} = 1 \quad \text{parce que} \quad z^2 = x^2 + y^2 \quad \text{sur le cône}$$

6.3 Dérivés le long de chemins courbes

Cinématique. Sur un chemin droit (c'est-à-dire dans une direction fixe), la dérivée de f est $D_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u}$ où $\boldsymbol{u}$ est parallèle au chemin. Quelle est la dérivée sur un chemin courbe? Dans un tel cas la direction du chemin $\boldsymbol{u}$ est indiquée par le *vecteur tangent* $\boldsymbol{T}$ à celui-ci. Dans le calcul on remplace donc $\boldsymbol{u}$ par $\boldsymbol{T}$, où $\boldsymbol{T}$ donne la « direction » de la courbe. Un chemin correspond formellement à une courbe dont l'ensemble des points est décrit par le vecteur position

$$\boldsymbol{R}(t) = x(t)\,\boldsymbol{i} + y(t)\,\boldsymbol{j}$$

Dans cette modélisation, on imagine un point (x, y) appartenant à une courbe dont les coordonnées x et y sont vues comme des fonctions du temps $t \geq 0$ qui joue le rôle d'un paramètre extérieur. Au fur et à mesure que t croît l'ensemble des points du chemin sont parcourus et décrit par le vecteur position $\boldsymbol{R}$.

Formules. La vitesse de déplacement le long de ce chemin est donc $\boldsymbol{v} = (dx/dt)\,\boldsymbol{i} + (dy/dt)\,\boldsymbol{j}$ et le vecteur tangent est alors défini par $\boldsymbol{T} = \boldsymbol{v}/\|\boldsymbol{v}\|$. Remarquez le choix, de se déplacer à n'importe quelle vitesse (avec $\boldsymbol{v}$) ou d'aller à vitesse unitaire (avec $\boldsymbol{T}$). Le même choix existe pour la dérivée

de $f(x,y)$ le long de cette courbe. On a

$$\textit{taux d'accroissement} = \frac{df}{dt} = \boldsymbol{\nabla} f \cdot \boldsymbol{v} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} \tag{6.4}$$

$$\textit{pente} = \frac{df}{ds} = \boldsymbol{\nabla} f \cdot \boldsymbol{T} = \frac{\partial f}{\partial x}\frac{dx}{ds} + \frac{\partial f}{\partial y}\frac{dy}{ds} \tag{6.5}$$

La première dérivée concerne le *temps*. Si nous avançons plus vite, df/dt augmente. La seconde dérivée concerne la *distance*. Si nous nous déplaçons d'une distance ds, à n'importe quelle vitesse, la fonction varie de df et la pente dans cette direction est donc df/ds.

Commentaires. Pour un mouvement uniforme sur une ligne droite on a $\boldsymbol{R} = \boldsymbol{R}_0 + \boldsymbol{v}t$ où la vitesse $\boldsymbol{v}$ est constante. La direction $\boldsymbol{T} = \boldsymbol{u} = \boldsymbol{v}/\|\boldsymbol{v}\|$ est également constante. La dérivée directionnelle est donc $\boldsymbol{\nabla} f \cdot \boldsymbol{u}$, mais le taux d'accroissement est $\boldsymbol{\nabla} f \cdot \boldsymbol{v}$. Les équations (6.4) et (6.5) correspondent à des *dérivées composées* (*chain rule*) et sont des généralisation de la formule bien connue $df/dt = (df/dx)(dx/dt)$ à plus de variables. Nous reviendrons sur leur preuves dans la section suivante. Pour le moment, nous nous concentrons, à travers des exemples, sur le sens de df/ds qui correspond à la dérivée de f dans la direction $\boldsymbol{u} = \boldsymbol{T}$ le long de la courbe.

Exemples 32. Trouvez df/dt et df/ds pour $f = r = \sqrt{x^2+y^2}$ et où la courbe est donnée par $x = t^2$, $y = t$, voir la **Figure 22 (gauche)**. La vitesse le long de la courbe est $\boldsymbol{v} = 2t\,\boldsymbol{i} + \boldsymbol{j}$. Au point typique $t = 1$ c'est $\boldsymbol{v} = 2\boldsymbol{i} + \boldsymbol{j}$. Le vecteur tangent unitaire est $\boldsymbol{T} = \boldsymbol{v}/\sqrt{5}$ et le gradient est un vecteur unitaire $\boldsymbol{i}/\sqrt{2} + \boldsymbol{j}/\sqrt{2}$ pointant vers l'extérieur, et $f(x,y)$ est la distance r à partir du centre. Le produit scalaire avec $\boldsymbol{v}$ est $df/dt = 3/\sqrt{2}$. Le produit scalaire avec $\boldsymbol{T}$ est $df/ds = 3/\sqrt{10}$. Lorsque on ralentit à la vitesse 1 (avec $\boldsymbol{T}$), les changements dans $f(x,y)$ ralentissent également.

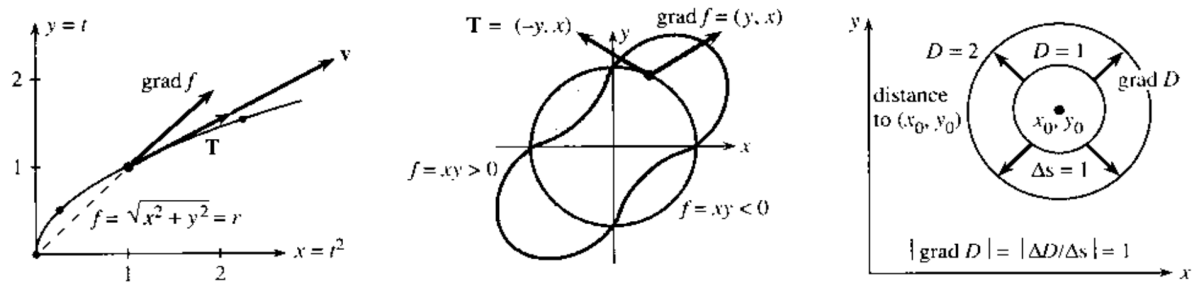


Figure 22: La distance $f = r$ change le long de la courbe. La pente des montagnes russes est $\boldsymbol{\nabla} f \cdot \boldsymbol{T}$. Pour la distance D de (x_0, y_0) on a $\mathbf{grad}\, D =$ vecteur unitaire.

Exemples 33. Trouvez df/ds pour $f = xy$ le long du chemin circulaire $x = \cos t$, $y = \sin t$. On adopte d'abord une approche directe. Sur le cercle, $xy = (\cos t)(\sin t)$. La dérivée de f provient de la règle du produit, soit $df/dt = \cos^2 t - \sin^2 t$. Normalement, c'est différent de df/ds, car le temps t n'a pas besoin d'être égal à la longueur de l'arc s. Il faut diviser par un facteur de vitesse ds/dt, mais ici la vitesse est de 1. (Un cercle de longueur $s = 2\pi$ est complété à $t = 2\pi$). Ainsi la pente df/ds le long des « montagnes russes » dans **Figure 22 (centre)** est $df/ds = \cos^2 t - \sin^2 t = \cos 2t$. Une seconde approche utilise les vecteurs $\boldsymbol{\nabla} f$ et $\boldsymbol{T}$. Le gradient de $f = xy$ est $\boldsymbol{\nabla} f = (x, y) = (\sin t, \cos t)$. Le vecteur tangent (unitaire) au chemin est $\boldsymbol{T} = (-\sin t, \cos t)$. Leur produit scalaire est le même

que précédemment, soit

$$\textit{pente le long du chemin} = \boldsymbol{\nabla} f \cdot \boldsymbol{T} = -\sin^2 t + \cos^2 t = \cos 2t$$

6.4 Gradients sans coordonnées

Caractérisation. On termine à présent par des considérations plus abstraites. Quelle est la définition sans coordonnées du gradient? Jusqu'à présent, $\boldsymbol{\nabla} f = (f_x, f_y)$ dépendait totalement du choix des axes x et y. Mais la pente d'une surface est indépendante des axes. Ceux-ci sont ajoutés plus tard, pour nous aider à calculer. De plus, la pente df/ds n'implique que f et la direction, rien d'autre. Le gradient est en fait ce qu'on appelle un *tenseur* au sens où il ne dépend pas du système de coordonnées. Le gradient a différentes formules dans différents systèmes (cartésien xy, polaire $r\theta$, etc ...), mais la direction et la longueur de $\boldsymbol{\nabla} f$ sont déterminés par df/ds, sans aucune référence à un axe particulier. En effet, d'après ce que nous avons discuter précédemment, nous savons que

- La *direction* de $\boldsymbol{\nabla} f$ est celle dans laquelle df/ds est le plus grand.
- La *longueur* $\|\boldsymbol{\nabla} f\|$ est la magnitude de la plus grande pente.

La relation clé est que (*changement de* f) $\approx$ (*gradient de* f) $\cdot$ (*changement de position*). C'est une autre façon d'écrire $\Delta f \approx f_x\,\Delta x + f_y\,\Delta y$ qui est la forme « multivariée », car nous avons utilisé deux variables, de l'approximation linéaire de base $\Delta y \approx (dy/dx)\,\Delta x$.

Exemples 34. On définit la fonction suivante

$$D(x, y) = \text{distance du point } (x, y) \text{ au point } (x_0, y_0)$$

(a) Sans dérivées, prouver que $\|\boldsymbol{\nabla} D\| = 1$.
$\hookrightarrow$ Le graphe de $D(x, y)$ est un cône de pente 1 avec un somment en (x_0, y_0).

(b) Dans quelle direction la distance $D(x, y)$ augmente-t-elle le plus rapidement?
$\hookrightarrow$ En s'éloignant le plus directement de (x_0, y_0). C'est donc la direction de $\boldsymbol{\nabla} D$.

(c) À quelle vitesse d augmente-t-il dans cette direction la plus raide?
$\hookrightarrow$ Un pas de longueur Δs augmente D de Δs. Par conséquent $\|\boldsymbol{\nabla} D\| = \Delta s/\Delta s = 1$.

En conclusion on voit que $\boldsymbol{\nabla} D$ est un *vecteur unitaire*. On revient sur la fonction D dans l'**Exercice 26**. Les dérivées de D sont $(x - x_0)/D$ et $(y - y_0)/D$. La somme de leurs carrés est 1, car $(x - x_0)^2 + (y - y_0)^2$ vaut D^2.

7 La dérivée composée ou « chain rule »

Introduction. Jusqu'à présent nous nous sommes concentrer sur la dérivation de fonction à plusieurs variables où chaque variable apparaissaient explicitement dans la fonction. Cependant, il est très commun dans les applications de rencontrer une situation où un paramètre extérieur est introduit, tel que le temps par exemple, dont va dépendre implicitement chaque variable d'une fonction donnée. Dans ce sens il est fondamental d'avoir à notre disposition une règle (ou formule) de dérivation composée (*chain rule*) afin de pouvoir calculer les variations d'un tel type de fonction. On peut également se convaincre de l'utilité d'un telle règle en considérant le cas d'une fonction $F(g(x))$, que l'on souhaite dériver par rapport à x, ce type de situation apparaît tout le temps dans les applications. On peut également se poser la même question pour $F(g(x,y))$ ou $F(x(t),y(t))$ ou pire encore... Il est donc impératif de savoir différentier de telles fonctions si l'on veut pouvoir progresser dans la résolution de problèmes pratiques.

Cas typiques. Fondamentalement, la dérivation composée s'applique à une *fonction d'une fonction*. Auparavant, pour une variable, on avait $F(g(x))$ cependant, avec deux variables, il y a maintenant plus de possibilités, par exemple

(1) $f(z)$ avec $z = g(x,y)$. Que valent $\partial f/\partial x$ et $\partial f/\partial y$?

(2) $f(x,y)$ avec $x = x(t)$, $y = y(t)$. Que vaut df/dt ?

(3) $f(x,y)$ avec $x = x(t,u)$, $y = y(t,u)$. Que valent $\partial f/\partial t$ et $\partial f/\partial u$?

Toutes les dérivées sont supposées continues. Plus exactement, les « dérivées d'entrée » comme $\partial g/\partial x$ et dx/dt et $\partial x/\partial u$ sont continues. Alors que les dérivées de sortie comme $\partial f/\partial x$ et df/dt et $\partial f/\partial u$ seront continues à partir de la formule de dérivation composée. On évitera par exemple des points comme $r = 0$, en coordonnées polaires, car on diviserai par zéro dans $\partial r/\partial x = x/r$.

Premier exemple. On commence par un problème typique avec $f(x,y) = xy$ où on remplace x par $r\cos\theta$ et y par $r\sin\theta$. La fonction devient alors $f = (r\cos\theta)\times(r\sin\theta)$ et on veut obtenir ses dérivées par rapport à r et θ. En premier lieu, pour ne pas se tromper, il est préférable de ne pas réutiliser la lettre f lorsque les variables ont changées. On introduit alors une nouvelle fonction F telle que

$$f(x,y) = xy = f(r\cos\theta, r\sin\theta) = (r\cos\theta)\times(r\sin\theta) = F(r,\theta)$$

Pourquoi ne pas l'appeler $f(r,\theta)$? Parce qu'à proprement parler c'est r fois θ. Si nous suivons les règles, alors $f(x,y)$ est xy et $f(r,\theta)$ devrait être $r\theta$. La nouvelle fonction F fait ce qu'il faut, elle multiplie $(r\cos\theta)(r\sin\theta)$. Malgré l'évidence de ce point, il est à noter que généralement, f reste très souvent inchangée bien que ses variables changent, c'est un abus de notation très commun et important à mentionner. Nous avons déjà commis cet abus de notation. Dans les exemples **(2)** et **(3)** on aurait du écrire $\partial F/\partial t$. Au lieu de cela, on a mis $\partial f/\partial t$. Le but de la dérivation composée ou *chain rule* est de trouver des dérivées dans les nouvelles variables t et u (ou r et θ). Dans notre exemple, on veut la dérivée de $F(r,\theta)$ par rapport à r. La chain rule donne

$$\frac{\partial F}{\partial r} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial r} = y\,\cos\theta + x\,\sin\theta = 2r\,\sin\theta\,\cos\theta = r\,\sin 2\theta$$

On a remplacé $r\sin\theta$ et $r\cos\theta$ par y et x. On peut vérifier immédiatement la réponse puisque $F(r,\theta) = r^2\cos\theta\sin\theta$ conduit à $\partial F/\partial r = 2r\cos\theta\sin\theta$. La dérivée est donc correcte. La seule chose incorrecte, mais nous le faisons quand même, est d'écrire f au lieu de F. Par le même raisonnement on trouve que

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = r^2(\cos^2\theta - \sin^2\theta) = r^2\cos 2\theta$$

7.1 Les dérivées de $f(g(x,y))$

Formules. Ici, g dépend de x et y, et f dépend de g. Supposons que x varie de dx, tandis que y reste constant. Alors g varie de $dg = (\partial g/\partial x)dx$. Quand g change, f change aussi, soit $df = (df/dg)dg$. Maintenant, on remplace dg pour former la « chaîne », c'est-à-dire que $df = (df/dg)(\partial g/\partial x)dx$. On a alors les formules de dérivation composée, ou « chain rule », pour $f(g(x,y))$ suivantes

$$\frac{\partial f}{\partial x} = \frac{df}{dg}\frac{\partial g}{\partial x} \qquad \text{et} \qquad \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y}$$

Exemples 35. Toute fonction $F(x + c\,y)$ satisfait l'équation $\partial f/\partial y = c\,\partial f/\partial x$ où c est une constante. La fonction intérieure est $g = x + cy$ et la fonction extérieure F peut être n'importe quoi, g^2 ou $\sin g$ ou encore e^g. La fonction composée est alors $(x + cy)^2$ ou $\sin(x + cy)$ ou e^{x+cy}. Dans chaque cas séparé, on peut vérifier que $\partial f/\partial y = c\partial f/\partial x$. La chain rule reproduit l'équation d'onde dans chaque cas, à partir de $\partial g/\partial x = 1$ et $\partial g/\partial y = c$. On a

$$\frac{\partial f}{\partial x} = \frac{df}{dg}\frac{\partial g}{\partial x} = 1\,\frac{\partial f}{\partial g} \quad \text{et} \quad \frac{\partial f}{\partial y} = \frac{df}{dg}\frac{\partial g}{\partial y} = c\,\frac{\partial f}{\partial g} \quad \text{donc} \quad \frac{\partial f}{\partial y} = c\,\frac{\partial f}{\partial x}$$

Il s'agit d'un exemple important puisque $\partial f/\partial y = c\,\partial f/\partial x$ est notre premier exemple d'une *équation aux dérivées partielles*. L'inconnu $f(x,y)$ possède deux variables et ses deux dérivées partielles entrent dans l'équation.

Équations différentielles. Jusqu'à présent, nous avons travaillé avec des dérivées à une seule variable du type dy/dt et des *équations différentielles ordinaires*. La variable indépendante était le temps ou une coordonnées de l'espace. Pour les équations aux dérivées partielles, les variables sont le temps et l'espace (éventuellement plusieurs dimensions dans l'espace) simultanément. Les grandes équations de la physique mathématique telles que l'*équation de la chaleur*, l'*équation des ondes* ou l'*équation de Laplace*, sont des équations aux dérivées partielles. Pour finir, regardons comment la chain rule s'applique à la fonction $f = \sin xy$. Sa dérivée par rapport à x est $y\cos xy$. Si on était patient on pourrais pu poser $f = \sin g$ et $g = xy$ puis utiliser $f_x = f_x\,g_x$. Cela permet de vérifier a posteriori la dérivée (supposée connue) de $\sin xy$. Nous passons maintenant à la chain rule dans le cas où la fonction externe dépend de deux fonctions internes, et chacune d'elles dépend de t, on veut alors déterminer df/dt.

7.2 Les dérivées de $f(x(t), y(t))$

Cas à deux variables. Avant la formule, voici l'idée. Supposons que t varie de Δt. Cela affecte x et y; ils varient de Δx et Δy. Il y a un effet domino sur f qui varie de Δf. Si on fait le cheminement inverse, on a alors

$$\Delta f \approx \frac{\partial f}{\partial x}\,\Delta x + \frac{\partial f}{\partial y}\,\Delta y \qquad \text{et} \qquad \Delta x \approx \frac{dx}{dt}\,\Delta t \qquad \text{et} \qquad \Delta y \approx \frac{dy}{dt}\,\Delta t$$

Maintenant, si on remplace Δx et Δy par leur expressions en fonction de Δt puis en formant $\Delta f/\Delta t$ alors la limite $\Delta t \to 0$ produit la formule de dérivation composée, ou chain rule, pour $f(x(t), y(t))$ suivante

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\,\frac{dx}{dt} + \frac{\partial f}{\partial y}\,\frac{dy}{dt}$$

Cette formule est proche de la règle pour une variable $dz/dx = (dz/dy)(dy/dx)$. Il suffit de prendre dy égal à zéro (ou plutôt prendre Δy égal à zéro dans $(\Delta z/\Delta y)(\Delta y/\Delta x)$ avant de prendre la limite).

On remarque que Δt affecte Δf de deux manières, via x et via y. *La chain rule possède donc deux termes.* Mentionnons à nouveau que le vrai nom pour $f(x(t), y(t))$ devrait être $F(t)$ et non pas $f(t)$.

Cas à trois variables et plus. Pour $f(x, y, z)$, la règle comporte trois termes et on a une expression de la forme

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial z}\frac{dz}{dt} \qquad \text{ou} \qquad f_t = f_x\, x_t + f_y\, y_t + f_z\, z_t$$

Encore une fois, la généralisation de cette formule à un nombre arbitraire de variable est immédiate. On considère une fonction $f(x_1(t), \ldots, x_n(t))$ où n est un entier fini et les fonctions $x_i(t)$ sont des fonctions (réelles) continues et différentiables sur intervalle donné. La *différentielle totale* de f est

$$df = \frac{\partial f}{\partial x_1}\, dx_1 + \cdots + \frac{\partial f}{\partial x_n}\, dx_n = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\, dx_i$$

La *dérivée totale* $f'(t) = df/dt$ est alors simplement

$$\frac{df}{dt} = \frac{\partial f}{\partial x_1}\frac{dx_1}{dt} + \cdots + \frac{\partial f}{\partial x_n}\frac{dx_n}{dt} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\frac{dx_i}{dt}$$

Exemples 36. À quelle vitesse la température varie-t-elle lorsque l'on se rend de Paris à Marseille? Supposons que la température en Île-de-France soit de 18°C et qu'à Marseille il fasse 29°C. Le déplacement (à vol d'oiseau) entre les villes est de 661 km et l'augmentation de la température, notée $f(x, y)$, est de 11°C, soit environ 0.02 degrés par km. En imaginant que l'on puisse rouler en ligne droite vers le sud à 70 km/h, le taux d'augmentation est de $0.02 \times 70 = 1.4$ degrés par heure environ. Notez comment (degrés/kilomètre) fois (kilomètre/heure) est égal à (degrés/heure). C'est la chain rule ordinaire

$$\frac{df}{dx}\frac{dx}{dt} = \frac{df}{dt}$$

où il n'y a pas de variable y puisque l'on se déplace vers le sud. Si la route va vers le sud-est, on suppose que la température au cours du voyage est alors $f = 18 + 0.05 \times x + 0.01 \times y$. Maintenant $x(t)$ correspond à un déplacement vers sud et $y(t)$ vers l'est. Qu'elle est la valeur de df/dt si la vitesse est toujours de 70 km/h? On écrit

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 0.02\,\frac{70}{\sqrt{2}} + 0.01\,\frac{70}{\sqrt{2}} \approx 1.5\,\text{degrés/heure}$$

En réalité, il existe un autre terme car la température dépend aussi directement de t, à cause de la nuit et du jour. La température comprend, par exemple, un facteur $\cos(2\pi t/24)$ qui possède une période de 24 heures, et il apporte un terme supplémentaire dans la chain rule. Pour $f(x, y, t)$ la chain rule est alors

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} + \frac{\partial f}{\partial t} \tag{7.1}$$

C'est la *dérivée totale* df/dt, à partir de toutes les « causes », c'est-à-dire des changements de x, y, t qui affectent f. La dérivée partielle $\partial f/\partial t$ n'est qu'une partie de la dérivé totale df/dt. (Notez que $dt/dt = 1$). Si la nuit et le jour ajoutent $12\cos(2\pi t/24)$ à f, le terme supplémentaire est

$$\frac{\partial f}{\partial t} = -\pi\,\sin\left(2\pi\,\frac{t}{24}\right)$$

À la tombée de la nuit, c'est $-\pi$ degrés par heure. Il faut conduire plus vite que 70 km/h pour se réchauffer plus rapidement.

7.3 Dérivées secondes

Que vaut d^2f/dt^2? Pour cela nous avons besoin de la dérivée de (7.1), ce qui n'est pas évident. On commence donc par l'exemple d'un mouvement en ligne droite pour lequel on veut déterminer l'accélération. Supposons que $x = x_0 + t\cos\theta$ et $y = y_0 + t\sin\theta$. Nous nous déplaçons avec un angle θ fixe et avec une vitesse $\|\boldsymbol{v}\| = 1$. Les dérivées sont $x_t = \cos\theta$ et $y_t = \sin\theta$, on a donc bien $\|\boldsymbol{v}\| = \sqrt{x_t^2 + y_y^2} = \sqrt{\cos^2\theta + \sin^2\theta} = 1$. Par la chain rule on obtient alors

$$f_t = f_x\, x_t + f_y\, y_t = f_x \cos\theta + f_y \sin\theta \tag{7.2}$$

Pour obtenir la dérivée seconde f_{tt} on applique encore cette règle mais cette fois à f_t. On obtient

$$\begin{aligned} f_{tt} &= (f_t)_x \cos\theta + (f_t)_y \sin\theta \\ &= (f_{xx}\cos\theta + f_{yx}\sin\theta)\cos\theta + (f_{xy}\cos\theta + f_{yy}\sin\theta)\sin\theta \\ &= f_{xx}\cos^2\theta + 2f_{xy}\cos\theta\sin\theta + f_{yy}\sin^2\theta \end{aligned} \tag{7.3}$$

Pour obtenir f_{rr} en coordonnées polaires il suffit de remplacer t par r. En effet, lorsque nous nous déplaçons dans la direction r, θ est fixe. L'équation (7.3) donne f_{rr} à partir de f_{xx}, f_{xy}, f_{yy}. Les dérivées secondes sur des chemins courbes viennent avec les termes supplémentaire dus à la courbure et sont abordées en exercice.

Exemples 37. Les dérivées f_{xx}, f_{xy}, f_{yy} sont toutes continues et bornées par M, trouvez alors une borne sur f_{tt}, qui est la dérivée seconde sur un chemin arbitraire. L'équation (**??**) donne $|f_{tt}| \leq M\cos^2\theta + M\sin 2\theta + M\sin^2\theta \leq 2M$. Cette borne supérieure $2M$ était nécessaire dans l'équation (5.7), pour évaluer l'erreur de l'approximation linéaire.

7.4 Les dérivées de $f(x(t,u), y(t,u))$

Formules. Supposons qu'il y ait deux fonctions internes x et y, chacune dépendant de t et u. Lorsque t varie, x et y varient également tous les deux et on a $dx = x_t dt$ et $dy = y_t dt$. Alors dx et dy forcent un changement dans f tel que $df = f_x dx + f_y dy$. La chain rule pour $\partial f/\partial t$ dans ce cas n'est donc pas une surprise. La formules de dérivation composée pour $f(x(t,u), y(t,u))$ est

$$\boxed{\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}}$$

Cette règle a $\partial/\partial t$ au lieu de d/dt, à cause de la variable supplémentaire u dont dépendent x et y. Le symbole « ∂ » nous rappellent que u est constant. De même, si t est constant pendant que u varie on a

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}$$

Finalement on peut construire la *différentielle totale* de f, soit

$$df = \frac{\partial f}{\partial t}\,dt + \frac{\partial f}{\partial u}\,du = \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial t}\right)dt + \left(\frac{\partial f}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial u}\right)du$$

Exemples 38. En coordonnées polaires, trouvez f_θ et $f_{\theta\theta}$ avec $f(x,y) = f(r\cos\theta, r\sin\theta)$. La chain rule donne

$$\frac{\partial f}{\partial \theta} = \frac{\partial f}{\partial x}\frac{\partial x}{\partial \theta} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial \theta} = \frac{\partial f}{\partial x}(-r\sin\theta) + \frac{\partial f}{\partial y}(r\cos\theta) \tag{7.4}$$

Le seconde dérivée par rapport à θ est plus difficile, car (7.4) possède quatre termes qui dépendent de θ. On applique la chain rule au premier terme $\partial f/\partial x$. C'est une fonction de x et y, et x et y sont des fonctions de θ, on a donc

$$\frac{\partial}{\partial \theta}\left(\frac{\partial f}{\partial x}\right) = \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\frac{\partial x}{\partial \theta} + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial x}\right)\frac{\partial y}{\partial \theta} = f_{xx}(-r\sin\theta) + f_{xy}(r\cos\theta)$$

La dérivée par rapport à θ de $\partial f/\partial y$ est similaire. On applique ensuire la règle du produit à (7.4), ce qui donne

$$\begin{aligned} f_{\theta\theta} &= [f_{xx}(-r\sin\theta) + f_{xy}(r\cos\theta)]\,(-r\sin\theta) + f_x(-r\cos\theta) \\ &\quad + [f_{yx}(-r\sin\theta) + f_{yy}(r\cos\theta)]\,(r\cos\theta) + f_y(-r\sin\theta) \end{aligned} \tag{7.5}$$

Cette formule n'est pas attractive. En mathématiques, une formule désordonnée est souvent un signe que l'on pose la mauvaise question. En fait, la combinaison $f_{xx} + f_{yy}$ est beaucoup plus spéciale que les dérivées f_{xx} et f_{yy} prisent séparément. On pourrait espérer la même chose de $f_{rr} + f_{\theta\theta}$, mais dimensionnellement c'est impossible, en effet r est une longueur et θ est un angle (sans dimension). Les dimensions de f_{xx} et f_{yy} correspondent aux dimensions de f_{rr} et f_r/r et $f_{\theta\theta}/r^2$. Dans ce sens, on pourrait espérer que

$$f_{xx} + f_{yy} = f_{rr} + \frac{1}{r}f_r + \frac{1}{r^2}f_{\theta\theta}$$

Cette équation est en fait correcte. En effectuant (7.2) + (7.3) + (7.5) avec t changé en r on obtient l'*équation de Laplace* $f_{xx} + f_{yy} = 0$ exprimée en coordonnées polaires, soit $f_{rr} + f_r/r + f_{\theta\theta}/r^2 = 0$.

7.5 Un paradoxe

Le problème. Avant de continuer il est intéressant de revenir sur un de nos premiers calculs. Cela remonte à $\partial r/\partial x$ avec $r = \sqrt{x^2+y^2}$, qui était pratiquement le premier exemple de dérivées partielles que nous avons vu. On a

$$\frac{\partial r}{\partial x} = \frac{\partial}{\partial x}\sqrt{x^2+y^2} = \frac{x}{\sqrt{x^2+y^2}} = \frac{x}{r} \tag{7.6}$$

Le problème est le suivant. Nous savons que $x = r\cos\theta$ donc $\partial r/\partial x = x/r = \cos\theta$. Par contre $r = x/\cos\theta$ donc $\partial r/\partial x$ vaut aussi $1/\cos\theta$. Comment $\partial r/\partial x$ peut-il conduire à $\cos\theta$ dans un sens et à $1/\cos\theta$ dans l'autre sens? La réponse n'est pas évidente et demande un peu de réflexion.

La solution. En fait la question importante à se poser lorsque l'on calcule des dérivées partielles est *quelle variable est maintenue constante ?* Dans l'équation (7.6), y est constant. Mais quand on pose $r = x/\cos\theta$ afin de calculer $\partial r/\partial x = 1/\cos\theta$ alors θ était constant. Dans les deux cas, nous changeons x et examinons l'effet sur r. Le mouvement se fait sur une ligne horizontale (où y est constant) ou sur une ligne radiale (où θ est constant) et la **Figure 23** montre la différence. Cet exemple montre que $\partial r/\partial x$ est différent de $1/(\partial x/\partial r)$. En effet, la formule élégante

$$\frac{\partial r}{\partial x}\frac{\partial x}{\partial r} = 1$$

n'est généralement pas vraie. Peut-on dire ce qui la remplace? Puisque r dépend de x et de y, on doit inclure $(\partial r/\partial y)(\partial y/\partial r)$. Étant donné que les variables x, y et les variables r, θ interviennent, il

est naturel d'utiliser des matrices de taille 2×2 afin d'obtenir une formule compacte. Celle-ci est de la forme

$$\begin{bmatrix} \partial r/\partial x & \partial r/\partial y \\ \partial \theta/\partial x & \partial \theta/\partial y \end{bmatrix} \begin{bmatrix} \partial x/\partial r & \partial x/\partial \theta \\ \partial y/\partial r & \partial y/\partial \theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Nous reviendrons sur cette représentation matricielle dans le cadre des changements de (plusieurs) variables.

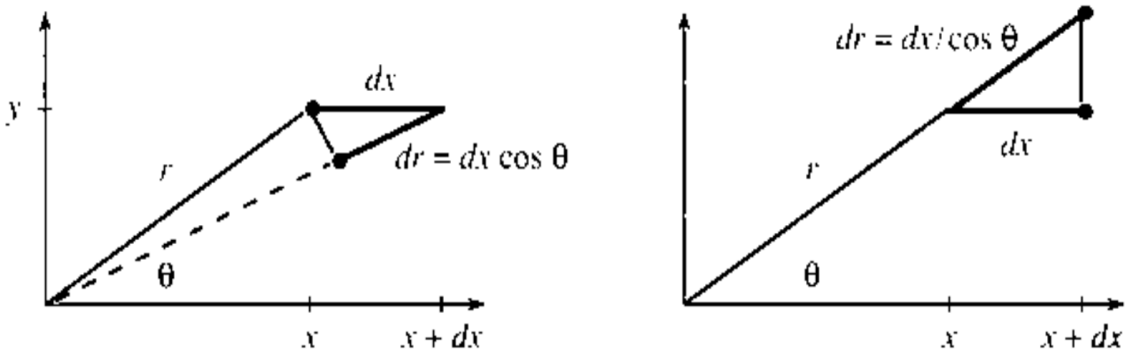


Figure 23: $dr = dx \cos\theta$ quand y est constant, $dr = dx/\cos\theta$ quand θ est constant.

7.6 Variables non indépendantes

La section précédente indique un problème sérieux. En calculant les dérivées partielles de $F(x, y, z)$, nous avons supposé que les variables x, y, z étaient indépendantes. Jusqu'à présent, x pouvait bouger tandis que y et z étaient considérées comme fixes. En physique, chimie ou en économie, cela peut ne pas être le cas. De fait, si il existe une relation entre x, y, z, alors x ne peut pas varier indépendamment.

Exemples 39.

(1) La loi des gaz parfaits $PV = nRT$ relie la pression au volume et à la température. P, V, T ne sont pas indépendants. Quelle est la signification de $\partial V/\partial P$? Est-ce égal à $1/(\partial P/\partial V)$? En l'état, ces questions n'ont pas de réponse, jusqu'à ce que nous disions ce qui est maintenu constant. Dans le paradoxe, $\partial r/\partial x$ avait une signification pour y fixe et une autre signification pour θ fixe. Pour indiquer ce qui est maintenu constant, on utilise alors un indice supplémentaire (ne désignant pas une dérivée) indiquant la variable maintenue constante lors du calcul de la dérivée. On a alors

$$\left(\frac{\partial r}{\partial x}\right)_y = \cos\theta \qquad \text{et} \qquad \left(\frac{\partial r}{\partial x}\right)_\theta = \frac{1}{\cos\theta}$$

Donc, dans le cas de la loi des gaz parfait, $(\partial f/\partial P)_V$ est à volume constant et $(\partial f/\partial P)_T$ est à température constante. Dans ce sens, l'habituel $\partial f/\partial P$ signifie que V et T sont tous les deux constants. Mais alors la loi des gaz parfaits ne nous permettra pas de changer P puis la pression ne dépend que de V et T selon elle.

(2) Soit $f = 3x + 2y + z$, calculez alors $\partial f/\partial x$ sur le plan $z = 4x + y$. On considère les trois approches suivantes.

- On prend x et y comme étant indépendants et on remplace z par $4x+y$, on a donc

$$f = 3x + 2y + (4x + y) \qquad \text{donc} \qquad \left(\frac{\partial f}{\partial x}\right)_y = 7$$

- On garde x et y indépendants puis on traite z par la chain rule, soit

$$\left(\frac{\partial f}{\partial x}\right)_y = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial z}\frac{\partial z}{\partial x} = 3 + 1 \times 4 = 7$$

- (Différent) On prend x et z indépendants, alors $y = z - 4x$ et on écrit

$$\left(\frac{\partial f}{\partial x}\right)_z = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial y}{\partial x} = 3 + 2 \times (-4) = -5$$

Sans indice, $\partial f/\partial x$ signifie prendre la dérivée par rapport à x de la manière habituelle. La réponse dans ce cas est $\partial f/\partial x = 3$, quand y et z ne bougent pas. Mais sur le plan $z = 4x+y$, l'un d'eux doit bouger. Le résultat 3 n'est donc qu'une partie de la réponse totale, qui est $(\partial f/\partial x)_y = 7$ ou $(\partial f/\partial x)_z = -5$. L'interprétation géométrique est la suivante. Nous sommes sur le plan $z = 4x + y$ et la dérivée $(\partial f/\partial x)_y$ change x mais pas y. Pour rester dans le plan, dz vaut $4\,dx$. La variation de $f = 3x + 2y + z$ est donc $df = 3\,dx + 0 + dz = 7\,dx$.

(3) Sur la « ligne d'univers » $x^2 + y^2 + z^2 = t^2$, trouvez $(\partial f/\partial y)_{x,z}$ pour $f = xyzt$. Les indices x, z signifient que x et z sont fixes. La chain rule ignore $\partial f/\partial y$ et $\partial f/\partial z$, on écrit donc

$$\left(\frac{\partial f}{\partial y}\right)_{x,z} = \frac{\partial f}{\partial y} + \frac{\partial f}{\partial t}\frac{\partial t}{\partial y} = xzt + xyz \times \frac{y}{t}$$

(4) À partir de la loi $PV = T$, calculer le produit $(\frac{\partial P}{\partial V})_T(\frac{\partial V}{\partial T})_P(\frac{\partial T}{\partial P})_V$. Il est tentant d'éliminer ∂V, ∂T et ∂P de ce produit pour obtenir 1. Cependant, la bonne réponse est -1, en effet

$$\left(\frac{\partial P}{\partial V}\right)_T = -\frac{T}{V^2} \qquad \left(\frac{\partial V}{\partial T}\right)_P = \frac{1}{P} \qquad \left(\frac{\partial T}{\partial P}\right)_V = V$$

Donc le produit est

$$\left(\frac{\partial P}{\partial V}\right)_T \left(\frac{\partial V}{\partial T}\right)_P \left(\frac{\partial T}{\partial P}\right)_V = -\frac{T}{PV} = -1$$

car $PV = T$. La chain rule peut donc être délicate à appliquer (voir **Exercice 31**).

(5) Voici un cas de différenciation implicite expliquée par la chain rule.

$$\text{Si} \quad F(x, y) = 0 \quad \text{alors} \quad F_x + F_y y_x = 0 \quad \text{donc} \quad \frac{dy}{dx} = -\frac{F_x}{F_y}$$

8 Maxima, minima et point-cols

Position du problème. L'une des équation fondamentale du calcul différentiel est également la plus simple

$$\frac{df}{dx} = 0$$

La pente est nulle et la droite tangente est horizontale. Très probablement, nous sommes en haut ou en bas d'un graphique, un maximum ou un minimum. C'est le point que recherche l'ingénieur, le gestionnaire, le scientifique ou l'investisseur, celui-ci correspondant à un maximum de contrainte ou de production, de vitesse ou de profit. Avec plus de variables dans $f(x, y)$ et $f(x, y, z)$, le problème (généralement) devient plus réaliste. Cependant la question demeure, *comment localiser le maximum et le minimum ?*

Cas possibles. La réponse utilise bien évidemment les dérivées partielles. Lorsque le graphique est toujours au même niveau, celles-ci sont nulles. Obtenir les équations $f_x = 0$ et $f_y = 0$ est un problème de mathématiques pur, les résoudre et comprendre leurs solutions, en revanche est la partie plus difficile. En particulier nous ferons attention aux « point-cols » mais également au minimum en un point limite, cela prend généralement plus de temps à étudier. On rappelle ici les étapes de la recherche d'extremum pour une fonction $f(x)$ d'une seule variable avec $a \leq x \leq b$.

- On recherche en premier les **points stationnaires**, appelés aussi *points critiques*, dont l'abscisse est solution de $df/dx = 0$.
- Ensuite on considère les **points approximatifs** où la dérivée n'existe pas, ainsi que les **points extrêmes** sur le domaine de recherche (ici a et b).
- Enfin le signe de la *dérivée seconde* permet de d'identifier si il s'agit d'un maximum ou d'un minimum.

En deux dimensions, un point stationnaire nécessite $\partial f/\partial x = 0$ et $\partial f/\partial y = 0$ et la droite tangente devient un plan tangent. Les extrémités a et b sont remplacées par une **(courbe) frontière**. En pratique, il est plus rare de trouver des extrema sur la frontière mais ceux-ci demande sensiblement plus de travail afin d'être localisé et caractérisé. Enfin, il y a trois dérivées secondes f_{xx}, f_{xy}, et f_{yy}. Elles indiquent comment le graphe s'éloigne du plan tangent, vers le haut au minimum, vers le bas au maximum, dans les deux sens en un **point-col**, appelé aussi **point-selle**. Ceci sera déterminé en comparant $(f_{xx})(f_{yy})$ avec $(f_{xy})^2$.

8.1 Point stationnaire et plan tangent

Introduction. Supposons que $f(x, y)$ possède un minimum au point (x_0, y_0). Cela peut être un **minimum absolu** ou seulement un **minimum local**. Dans les deux cas on a $f(x_0, y_0) \leq f(x, y)$ près du point. Pour un minimum absolu, cette inégalité est valable partout où f est défini. Pour un minimum local, l'inégalité peut échouer loin de (x_0, y_0). Dis simplement, à vos pieds se situe un minimum absolu tandis qu'au sommet de votre tête se trouve un minimum local. Nous supposons pour l'instant que (x_0, y_0) est un *point intérieur* du domaine de f. En un *point limite*, on ne peut avoir une tangente horizontale et des dérivées nulles. La remarque principale est que à un minimum ou un maximum (absolu ou local), une dérivée différente de zéro est impossible, sinon le plan tangent s'inclinerait. Dans une certaine direction, au point tangent, f diminuerait dans un sens et augmenterai dans le sens opposé. On note au passage que le *point minimum* est (x_0, y_0), et la *valeur minimum* est $f(x_0, y_0)$.

Critère pour un extremum. Si des dérivées existent en un minimum ou un maximum intérieur alors celles-ci elles sont nulles. Pour une fonction $f(x, y)$ on a alors

$$\frac{\partial f}{\partial x} = 0 \quad \text{et} \quad \frac{\partial f}{\partial y} = 0 \qquad \text{ou} \qquad \boldsymbol{\nabla} f = \mathbf{0}$$

Pour une fonction $f(x, y, z)$ de trois variables, ajoutez la troisième équation $\partial f/\partial z = 0$. Il est à noter que la condition $\boldsymbol{\nabla} f = \mathbf{0}$ est valide pour un nombre arbitraire, mais fini, de variables. Le raisonnement remonte en fait au cas à une variable. C'est parce que nous regardons le long des lignes $x = x_0$ et $y = y_0$. Le minimum de $f(x, y)$ est au point de rencontre des lignes. C'est donc aussi le minimum *le long de chacune de ces lignes prisent séparément.*

Interprétation graphique. En nous déplaçant dans la direction des x le long de $y = y_0$, on trouve $\partial f/\partial x = 0$. Puis en se déplaçant dans la direction des y, $\partial f/\partial y = 0$ au même point. *La pente dans toutes les directions est nulle*, en d'autres termes $\boldsymbol{\nabla} f = \mathbf{0}$. Graphiquement, le minimum (x_0, y_0) est le point situé « en bas » de la surface $z = f(x, y)$. Deux coupes transversales sur la **Figure 24** contiennent ce point. La phrase « si des dérivés existent » exclut le sommet d'un cône, qui est un *point approximatif.* La valeur absolue $f = |x|$ a un minimum sans $df/dx = 0$, tout comme la distance $f = r$ où le point approximatif est $(0, 0)$.

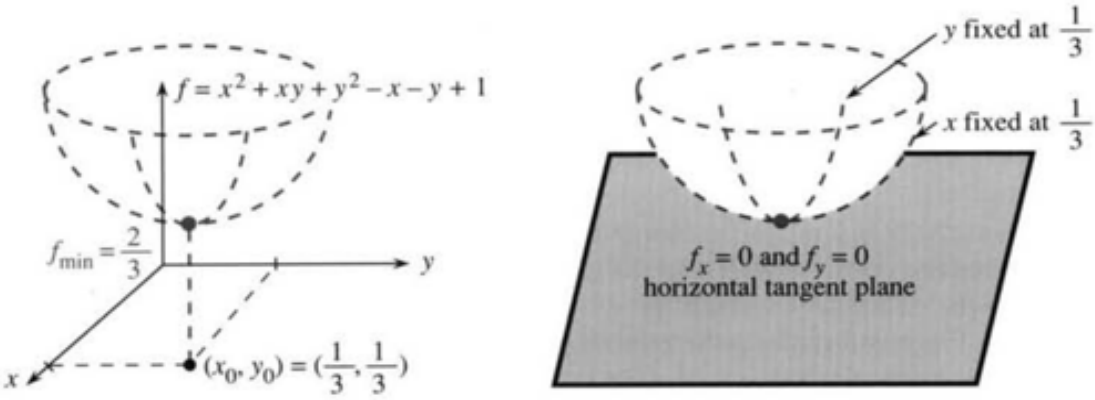


Figure 24: $\partial f/\partial x = 0$ et $\partial f/\partial y = 0$ au minimum. Quadratic f a des dérivées linéaires.

Exemples 40. On souhaite minimisez le quadratique $f(x, y) = x^2 + xy + y^2 - x - y + 1$. Pour localiser le minimum (ou maximum), on calcule $f_x = 0$ et $f_y = 0$, soit

$$f_x = 2x + y - 1 = 0 \qquad \text{et} \qquad f_y = x + 2y - 1 = 0$$

Il s'agit d'un système deux équations pour deux inconnues x et y. Puisque f est quadratique, les équations sont linéaires. Leur solution est $x_0 = 1/3, y_0 = 1/3$ ce qui donne la position du point stationnaire. C'est en fait un minimum et nous verrons dans la suite comment montrer cela. La constante 1 dans f affecte la valeur minimale $f = 2/3$, mais pas la position du minimum. Le graphique augmente simplement de 1. Les termes linéaires $-x - y$ affectent f_x et f_y. Ils éloignent le minimum de $(0, 0)$. La partie quadratique $x^2 + xy + y^2$ rend la surface courbe vers le haut. Sans cette partie incurvée, le plan aurait son minimum ou son maximum aux points extrêmes.

8.2 Le problème de Steiner

Position du problème. Le problème posé est simple, on veut trouvez *le point le plus proche* de trois points donnés dans un plan. Cet exemple est important. En effet on peut vouloir localiser un aéroport proche de trois villes ou encore choisir une maison proche de trois emplois. Le problème est de se rapprocher le plus possible des coins d'un triangle. Le meilleur point dépend de la signification du mot « proche ». La distance au premier coin (x_1, y_1) est $D_1 = \sqrt{(x-x_1)^2 + (y-y_1)^2}$. Les distances aux autres coins (x_2, y_2) et (x_3, y_3) sont D_2 et D_3. Selon que l'on considère $(distance)$ ou $(distance)^2$ ou $(distance)^p$, notre problème reviendra à

$$\textit{minimiser} \;\; d_1 + d_2 + d_3 \qquad \textit{ou} \qquad d_1^2 + d_2^2 + d_3^2 \qquad \textit{ou} \qquad d_1^p + d_2^p + d_3^p$$

Le deuxième problème est le plus simple avec D_1^2 et D_2^2 et D_3^2 car nous avons la forme quadratique

$$f(x,y) = (x-x_1)^2 + (y-y_1)^2 + (x-x_2)^2 + (y-y_2)^2 + (x-x_3)^2 + (y-y_3)^2$$

avec les dérivées partielles suivantes

$$\frac{\partial f}{\partial x} = 2(x - x_1 + x - x_2 + x - x_3) = 0 \qquad \text{et} \qquad \frac{\partial f}{\partial y} = 2(y - y_1 + y - y_2 + y - y_3) = 0$$

Résoudre $\partial f/\partial x = 0$ donne $x = \frac{1}{3}(x_1+x_2+x_3)$ et $\partial f/\partial y = 0$ donne $y = \frac{1}{3}(y_1+y_2+y_3)$. Le meilleur point est alors le *barycentre* ou *centre de gravité* du triangle (**Figure 25 (gauche)**). C'est le point le plus proche des coins avec le choix de minimiser $(distance)^2$. Notez comment la quadrature rend les dérivées linéaires. Ici il s'agit en fait d'un exemple élémentaire de la *méthode des moindres carrés* (*least square method*) qui forme une part importante des mathématiques appliquées.

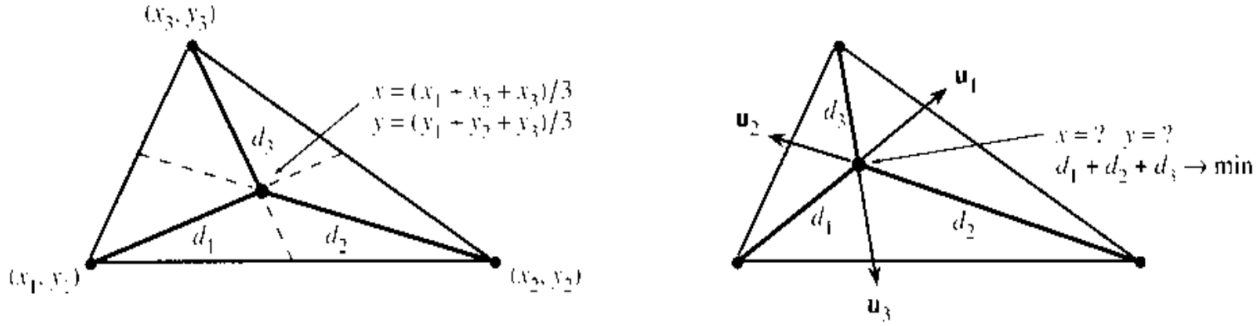


Figure 25: Le centre de gravité minimise $D_1^2 + d_2^2 + d_3^2$. Le point de Steiner minimise $d_1 + d_2 + d_3$.

Généralisation. Le « vrai » problème de Steiner consiste en fait à minimiser $f(x,y) = d_1 + d_2 + d_3$. Cela est du à la raison pratique que lorsque l'on construit des routes depuis les coins, le coût est proportionnel à la longueur de ces routes, et non du carré de ces longueurs . Les équations $f_x = 0$ et $f_y = 0$ semblent compliquées à cause des racines carrées. Mais la solution, c'est-à-dire le point le plus proche des trois sommets, est remarquable (**Figure 25 (droite)**). Dans le cas présent il plus simple de considérer les dérivées de $d_1^2 = (x-x_1)^2 + (y-y_1)^2$ afin de déterminer celles de d_1. En effet on a $\partial_x d_1^2 = 2d_1\, \partial_x d_1 = 2(x - x_1)$ et similairement pour $\partial_y d_1^2$. On peut donc écrire

$$\frac{\partial d_1}{\partial x} = \frac{x - x_1}{d_1} \qquad \text{et} \qquad \frac{\partial d_1}{\partial y} = \frac{y - y_1}{d_1} \qquad \text{soit} \qquad \boldsymbol{\nabla} d_1 = \left(\frac{x - x_1}{d_1}, \frac{y - y_1}{d_1}\right)$$

Ce gradient est donc un *vecteur unitaire*, en effet on a

$$\frac{(x-x_1)^2}{d_1^2}+\frac{(y-y_1)^2}{d_1^2}=\frac{D_1^2}{d_1^2}=1$$

Cela était déjà mentionné dans l'**Exemple 6.1**, la distance augmente avec une pente de 1 lorsque l'on s'éloigne du centre. Le gradient de D_1 est un vecteur unitaire noté $\boldsymbol{u}_1$, partant du point (x_1, y_1) sur (**Figure 25 (droite)**). Si on procède de la même manière avec les gradients de D_2 et D_3, notés $\boldsymbol{u}_2$ et $\boldsymbol{u}_3$, ceux-ci pointent directement à l'opposé des autres coins du triangle. Comme on a $f(x, y) = d_1 + d_2 + d_3$, on peut donc ajouter les gradients et on retrouve la contrainte initiale stipulant qu'au minimum on doit avoir

$$\boldsymbol{\nabla} f = \boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = \mathbf{0}$$

La somme des trois vecteurs (gradients) unitaires est égale à zéro. S'éloigner d'un coin nous rapproche d'un autre. Le point le plus proche des trois coins est l'endroit où ces mouvements s'annulent. C'est la signification de $\boldsymbol{\nabla} f = \mathbf{0}$ au minimum.

Solutions possibles. Il est inhabituel que la somme de trois vecteurs unitaires s'annule, cela ne peut se produire que d'une seule manière, les trois directions doivent former des angles de 120°. Notre solution possède cette propriété et ceci est illustré sur la **Figure 26 (droite)**. Les vecteurs unitaires s'annulent donc au *point de Steiner* (ou *point de Fermat*), les routes menant aux coins font alors 120° angles. Ce point optimal résout donc le problème, à l'exception d'une autre possibilité.

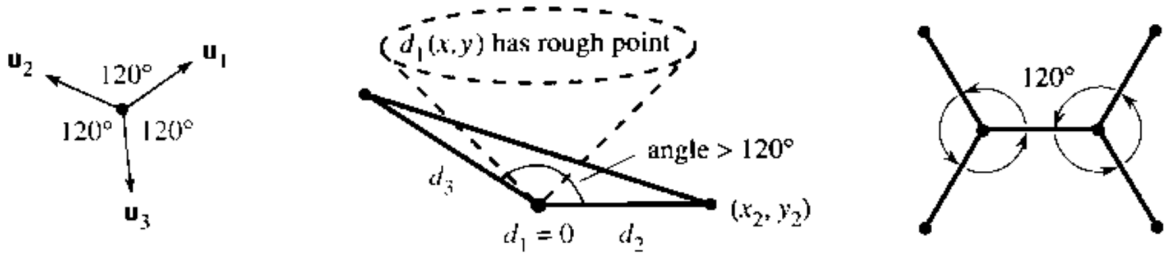


Figure 26: gradients $\boldsymbol{u}_1 + \boldsymbol{u}_2 + \boldsymbol{u}_3 = \mathbf{0}$ pour 120° angles. Corner gagne au grand angle. *Quatre coins.* Dans ce cas, deux points de branchement valent mieux 120°.

L'autre possibilité est un minimum à un *point approximatif.* En effet, le graphique de la fonction de distance $d_1(x, y)$ est un cône possédant un sommet au centre (x_1, y_1). Les trois coins du triangle sont des points approximatifs pour $d_1 + d_2 + d_3$, donc tous sont des minimiseurs possibles. Cette solution devient viable lorsque l'angle en un coin du triangle dépasse 120°. Dans cette configuration il n'y a en effet plus de point de Steiner. La solution correspond alors à un point approximatif, l'un des coins du triangle et c'est le coin avec le grand angle qui la solution du problème. Sur la **Figure 26 (milieu)**, cela signifie $d_1 = 0$ et dans ce cas la somme devient $d_2 + d_3 = f$ provient des deux arêtes les plus courtes. En résumé on a donc deux types de solution pour le « vrai » problème de Steiner, une située à l'intérieur du triangle si les angles de celui-ci sont inférieurs à 120° (voir la **Figure 26 (droite)**), et une autre située au sommet avec un angle supérieur à 120°.

Remarques. Le problème de Steiner pour *quatre points* est surprenant car minimiser $d_1 + d_2 + d_3 + d_4$ ne mène pas à la meilleure solution. En fait il faut relier les quatre points par des routes, en minimisant toujours leur longueur totale, mais cette fois si en autorisant les routes à *bifurquer*,

c'est-à-dire qu'il peut y avoir des points de branchement. Une solution typique est montrée dans la **Figure 26 (droite)**. On note que les angles aux points de branchement sont de 120° et il y a au plus deux points de branchement (deux de moins que le nombre de coins). Dans le cas générale où la quantité à minimiser est $d_1^p + d_2^p + d_3^p$, la dérivée par rapport à x donne

$$\frac{\partial df}{\partial x} = p\, d_1^{p-2}(x - x_1) + p\, d_2^{p-2}(x - x_2) + p\, d_3^{p-2}(x - x_3)$$

Les équations clés sont toujours $\partial f/\partial x = 0$ et $\partial f/\partial y = 0$ mais il nécessaire d'avoir recourt à des méthodes numériques pour les résoudre. Dans un tel cas l'algorithme typiquement utilisé est la *méthode de Newton* et c'est un excellent exemple d'application. La situation est typique des problèmes réels, on connaît les dérivées $\partial f/\partial x$ et $\partial f/\partial y$ mais pas le point où elles sont nulles. On peut alors calculer la position de point le plus proche des trois sommets et observer qu'elle change à mesure que p change.

8.3 Extremum sur la frontière

Contraintes. Le problème de Steiner n'avait pas de frontières. Les routes peuvent aller n'importe où. Mais la plupart des applications ont des restrictions sur x et y, comme $x \geq 0$ ou $y \leq 0$ ou $x^2 + y^2 \geq 1$. Le minimum avec ces restrictions est probablement supérieur au minimum absolu. Il existe trois possibilités suivantes

(1) *point stationnaire* **(2)** *point approximatif* **(3)** *point extrême*

Dans le cas d'un *point extrême* on doit maximiser ou minimiser $f(x, y)$ le long de la frontière.

Exemples 41. On veut minimiser $f(x, y) = x^2 + xy + y^2 - x - y + 1$ sur les différents domaines suivants.

(a) *Dans le demi-plan droit $x \geq 0$*
$\hookrightarrow$ La valeur minimum obtenue dans l'**Example 7.4** était $f = 2/3$ pour $x = x_0 = 1/3, y = y_0 = 1/3$. *Ce point est toujours autorisé.* Il satisfait la restriction $x \geq 0$. Le minimum ne s'est donc pas déplacé.

(b) *Dans demi-plan inférieur $y \leq 0$*
$\hookrightarrow$ Désormais, le minimum absolu situé en $(1/3, 1/3)$ n'est plus autorisé et il n'y a pas de points approximatifs. On recherche donc un minimum sur la frontière $y = 0$, voir **Figure 27**. On procède par étape en posant d'abord $y = 0$ de manière à ce que f ne dépende que de x puis on détermine x tel que

$$f(x, 0) = x^2 + 0 - x - 0 + 1 \qquad \text{soit} \qquad f_x = 2x - 1 = 0$$

Le minimum est est donc en $x = 1/2$ et $y = 0$ et la valeur de f en ce minimum est $3/4$, soit en hausse de $2/3$ par rapport au cas précédent.

(c) *Sur ou à l'extérieur du cercle $x^2 + y^2 = 1$*
$\hookrightarrow$ La solution $(1/3, 1/3)$ de $f_x = 0, f_y = 0$ est ici situé à l'intérieur du cercle. Une autre possibilité est donc que le minimum soit un point extrême, c'est-à-dire qu'il soit situé *sur le cercle*, voir la **Figure 27**. Pour suivre cette frontière, on définit $y = \sqrt{1 - x^2}$ et on remplace dans f. Cependant la fonction f se complique beaucoup et df/dx également. Il existe néanmoins une approche afin d'éviter les racines carrées. On défini $x = \cos t$ et $y = \sin t$ de sorte que $f = x^2 + xy + xy + 1$ et sa dérivée deviennent des fonctions de l'angle t de la forme

$$\begin{aligned} f(t) &= 1 + \cos t \sin t - \cos t - \sin t + 1 \\ \frac{df}{dt} &= \cos^2 t - \sin^2 t + \sin t - \cos t = (\cos t - \sin t)(\cos t + \sin t - 1) \end{aligned}$$

Maintenant la condition $df/dt = 0$ permet de localiser un minimum ou un maximum le long de la frontière en fonction de t. Le premier facteur $(\cos t - \sin t)$ est nul lorsque $x = y$. Le deuxième facteur est nul lorsque $\cos t + \sin t = 1$, ou $x + y = 1$. Ces points *sur le cercle* sont donc les candidats et dans l'**Exercise 37 (4)** ceux-ci sont triés. On peut également trouver un extremum sur une frontière en utilisant la méthode des *multiplicateurs de Lagrange* et celle-ci est discutée en détails dans l'**Appendice A**. Il est à noté que la minimisation/maximisation sur une frontière est un problème majeur, qui peut rapidement devenir assez difficile, pour lequel la méthodes des multiplicateurs de Lagrange conduit souvent à une solution satisfaisante.

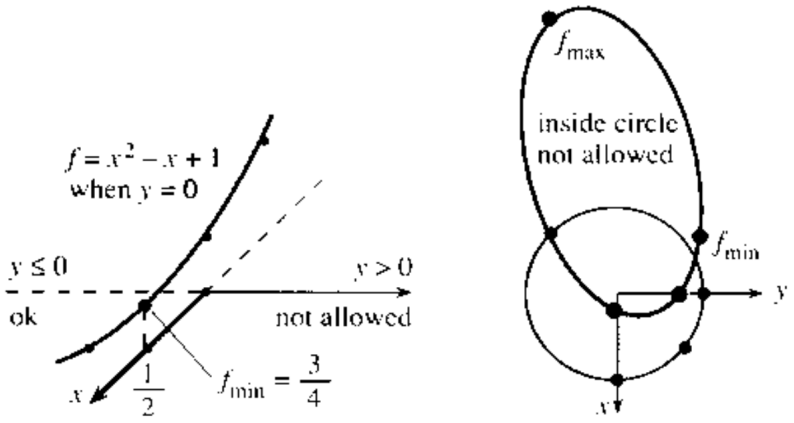


Figure 27: Les limites $y = 0$ et $x^2 + y^2 = 1$ contiennent le minimum.

8.4 Maximum, minimum et point-col

Nature d'un extremum. Comment distinguer le maximum d'un minimum? Une approche possible, et fastidieuse, consiste à évaluer la fonction f au voisinage d'un point stationnaire et de comparer entre elles les valeurs obtenues afin de dresser une *topographie* locale et approximative de f. L'analyse mathématique propose une autre approche, basée sur l'analyse et l'interprétation des *dérivées secondes*. Dans le cas d'une fonction à une variable le test est simple. Si $f_{xx} > 0$ on a un minimum, si $f_{xx} = 0$ on a un point d'inflexion et si $f_{xx} < 0$ on a un maximum. Il s'agit cependant d'un *test local*, la différentiation étant une opération locale, ne donnant pas d'information sur la forme globale de la courbe et ne permettant pas de conclure sur la nature local ou global d'un extremum. Mais ce test permet néanmoins de décider si la pente est croissante (bas du graphique) ou décroissante (haut du graphique) et on va voir à présent un test similaire pour une fonction $f(x, y)$.

Cas en deux dimensions. Ce nouveau test implique les trois dérivées secondes que l'on va évaluer au point pour lequel $f_x = 0$ et $f_y = 0$, c'est-à-dire là où le plan tangent est horizontal. La question fondamentale est alors de savoir si le graphe de f est *au-dessus* ou *en dessous* de ce plan. Par exemple, si on a $f_{xx} > 0$ et $f_{yy} > 0$ alors le point testé est un minimum pour les directions x et y, mais il faut gardé à l'esprit qu'il existe d'*autres directions*, c'est là toute la complexité engendrée par une fonction de plusieurs variables. Avant de présenter les règles générales qui permettent de déterminer la nature d'un extremum pour une fonction de deux variables, on discute à travers des exemples (simples) comment ce problème peut se révéler plus subtil que prévu.

Exemples 42.

(1) On considère la fonction suivante

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

dont les dérivées partielles, facilement calculables, sont

$$f_{xx} = 2 \quad , \quad f_{xy} = 10 \quad , \quad f_{yy} = 2$$

Quelle est alors la nature du point stationnaire (si celui-ci existe)? On remarque tout d'abord que toutes les dérivées secondes sont positives puis, en résolvant $f_x = f_y = 0$, on trouve que l'origine $(0,0)$ est un point stationnaire. La fonction est la somme de $x^2 + y^2$, qui augmente lorsque l'on s'éloigne de l'origine, et de $10\,xy$ qui forme une *selle* ou un *col*. Les dérivées secondes doivent nous permettre de décider si c'est $x^2 + y^2$ ou $10\,xy$ qui domine. Le long de l'axe x, c'est-à-dire en fixant $y = 0$ de sorte que $f = x^2$, on remarque que l'origine est minimum, mais uniquement dans la direction des x. On obtient le même résultat en posant $y = 0$. Cependant $(0,0)$ *n'est pas un minimum* pour la fonction f complète, à cause du terme $10\,xy$, il faut donc faire attention à ne pas conclure hâtivement. Si on essaie par exemple $x = 1, y = -1$. Alors $f = 1 - 10 + 1 = -8$ et donc le graphique de f est situé sous le plan xy en ce point. En fait, le point stationnaire à $x = y = 0$ est un *point-col*, voir la **Figure 28 (droite)**.

(2) On considère maintenant la fonction suivante

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

dont les dérivées partielles sont

$$f_{xx} = 2 \quad , \quad f_{xy} = 1 \quad , \quad f_{yy} = 2$$

Quelle est alors la nature du point stationnaire (si celui-ci existe)? Les dérivées secondes sont à nouveau toutes positives cependant le graphique se courbe cette fois vers le haut dans toutes les directions, voir la **Figure 28 (gauche)**. La grande différence par rapport à l'exemple ci-dessus est que la dérivée seconde f_{xy}, c'est-à-dire le coefficient devant xy, est réduit de 10 à 1. C'est donc la *magnitude* de f_{xy} et non pas son signe qui fait la différence. Sin on ajoute les termes $-x - y + 1$ à f on retombe sur l'exemple précédent, dès lors on voit que l'ajout de ces termes ne fait que déplacer le point stationnaire de l'origine vers $(1/3, 1/3)$. En effet, ces termes n'étant que linéaires, les dérivées secondes restent inchangées, et le test ci-dessous pour des *fonctions quadratiques* donne le même résultat.

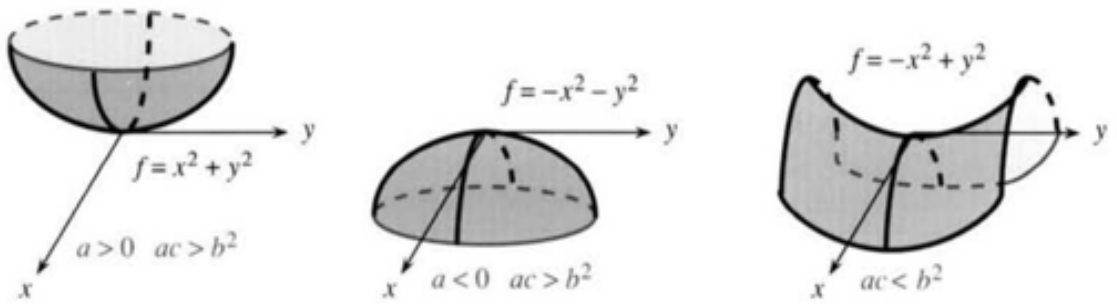


Figure 28: Configurations de minimum, maximum et point-col basées sur les signes a et $ac - b^2$.

Critère pour une fonction quadratique. On considère la *fonction quadratique*

$$f(x,y) = \mathsf{r}\,x^2 + 2\mathsf{s}\,xy + \mathsf{t}\,y^2$$

où les coefficients r,s,t sont des nombres réels. Dans ce cas, f possède un *point stationnaire* à l'origine $(0,0)$ et sa nature est déterminé par le test suivant.

- Si $\mathsf{s}^2 - \mathsf{rt} < 0$ alors l'origine est un **extrémum** local.
  - Si $\mathsf{r} > 0$ alors l'origine est un **minimum** local
  - Si $\mathsf{r} < 0$ alors l'origine est un **maximum** local
- Si $\mathsf{s}^2 - \mathsf{rt} > 0$ alors l'origine n'est pas un extrémum local, c'est un **point col**
- Si $\mathsf{s}^2 - \mathsf{rt} = 0$ alors **on ne peut pas conclure** directement.
  Dans ce cas on peut utiliser d'autres moyens (par exemple faire un graphique, etc...)

Il n'est pas difficile de démontrer ce critère. Pour cela il suffit d'écrire $f(x,y)$ sous la forme d'une somme de deux carrés, appelée *forme canonique*, afin d'éliminer le terme *mixte* proportionnel à xy. De la sorte, on dit que l'on *complète le carré*, on obtient

$$\mathsf{r}\,x^2 + 2\mathsf{s}\,xy + \mathsf{t}\,y^2 = \mathsf{r}\left(x + \frac{\mathsf{s}}{\mathsf{r}}\,y\right)^2 + \frac{\mathsf{rt} - \mathsf{s}^2}{\mathsf{r}}\,y^2$$

Sous cette *forme canonique* il est facile de se convaincre que les signe a et $ac - b^2$ sont les paramètres essentiels afin de juger de la nature du point stationnaire à l'origine. On note que puisque le test implique le *carré* de b, son signe n'a aucune importance. Précédemment, dans l'**Exemple 8.4 (1)**, nous avions $b = 5$ et un point-col tandis que dans l'**Exemple 8.4 (1)** on avait $b = 1/2$ et un minimum. Naturellement si on inverse le dernier exemple pour obtenir $-x^2 - xy - y^2$ alors l'origine devient un maximum et il en va de même pour la fonction $-x^2 + xy - y^2$. Vient maintenant la dernière étape, avec le passage de $\mathsf{r}\,x^2 + 2\mathsf{s}\,xy + \mathsf{t}\,y^2$ à une fonction générale et arbitraire $f(x,y)$. Pour toutes les fonctions, quadratiques ou non, ce sont les *termes du second ordre*, c'est-à-dire dont la dérivée seconde est non-nulle, que nous testons.

Exemples 43. On veut déterminer la nature du point stationnaire à l'origine de la fonction

$$f(x,y) = e^x - x - \cos y$$

Tout d'abord on peut vérifier rapidement que les équations

$$f_x = e^x - 1 = 0 \qquad \text{et} \qquad f_y = \sin y = 0$$

ont pour solution $(0,0)$. Ensuite, les dérivées secondes évaluées à l'origine sont $f_{xx} = e^x = 1$ et $f_{yy} = \cos y = 1$ et $f_{xy} = 0$. Il est suffisant de n'utiliser que les dérivées *au point stationnaire*. Comme les dérivées premières sont nulles, on remarque que ce sont donc les termes du second ordre viennent en premier dans la série pour $e^x - x - \cos y$, en effet

$$\left(1 + x + \frac{1}{2}x^2 + \dots\right) - x - \left(1 + y + \frac{1}{2}y^2 + \dots\right) = \frac{1}{2}x^2 + \frac{1}{2}y^2 + \text{termes d'ordre supérieur}$$

Selon ce calcul, il y a un *minimum* à l'origine puisque la *partie quadratique* $\frac{1}{2}x^2 + \frac{1}{2}y^2$ va « vers le haut ». Les termes x^3 et y^4 sont trop petits pour changer cela, puisque l'on reste au voisinage de l'origine afin que la série reste valide. Finalement, ces termes deviendront grands, mais près d'un point stationnaire (où $x = 0$, $y = 0$), c'est le terme quadratique qui prévaut. Nous n'avons pas besoin de toute la série, car à partir de $f_{xx} = f_{yy} = 1$ et $f_{xy} = 0$, nous savions que cela commencerait par $\frac{1}{2}x^2 + \frac{1}{2}y^2$ et cela est suffisant pour notre analyse.

Critère pour une fonction arbitraire. On considère une fonction arbitraire $f(x, y)$ et on admet que le point (x_0, y_0) est un *point stationnaire* de f. On définit également les coefficients suivants

$$\mathsf{r} = \left(\frac{\partial^2 f}{\partial^2 x}\right)_0 \qquad \mathsf{s} = \left(\frac{\partial^2 f}{\partial x \partial y}\right)_0 \qquad \mathsf{t} = \left(\frac{\partial^2 f}{\partial^2 y}\right)_0$$

où l'indice 0 indique que les dérivées secondes sont évaluées au point stationnaire. Dans ce cas la nature de ce point stationnaire est donner par le même critère que dans le cas quadratique, c'est-à-dire

- Si $\mathsf{s}^2 - \mathsf{rt} < 0$ alors le point (x_0, y_0) est un **extrémum** local.
  - Si $\mathsf{r} > 0$ alors le point (x_0, y_0) est un **minimum** local
  - Si $\mathsf{r} < 0$ alors le point (x_0, y_0) est un **maximum** local
- Si $\mathsf{s}^2 - \mathsf{rt} > 0$ alors le point (x_0, y_0) n'est pas un extrémum local, c'est un **point col**
- Si $\mathsf{s}^2 - \mathsf{rt} = 0$ alors **on ne peut pas conclure** directement.

Exemples 44. On veut déterminer la nature du point stationnaire à l'origine de la fonction

$$f(x, y) = e^{xy}$$

dont les dérivées premières sont $f_x = ye^{xy}$ et $f_y = xe^{xy}$. Les dérivées secondes à l'origine donnent $a = f_{xx} = 0$, $b = f_{xy} = 1$ et $c = f_{yy} = 0$. Le test $b^2 > ac$ conduit donc à une configuration en point-col. Si on examine la série autour de l'origine on obtient

$$e^{xy} = 1 + xy + \frac{1}{2}x^2y^2 + \dots$$

où il n'y a pas de terme linéaire car $f_x = f_y = 0$. De cela on confirme que l'origine est bien un *point stationnaire*. Il n'y a pas de terme proportionnel à x^2 ou y^2 mais seulement àxy), on en conclut que le point stationnaire est un *point-col*. Si on applique le test en $(2, -2)$ on trouve $f_{xx}f_{yy} > (f_{xy})^2$ alors on peut dire que le graphique est *concave vers le haut* à ce stade, mais ce n'est pas un minimum puisque les dérivées premières ne sont pas nulles en ce point.

Commentaire. Dans ce dernier exemple, la série commence par un terme constant qui n'a pas d'impact sur la forme de la fonction autour de n'importe quel point sur la surface $z = f(x, y)$, on l'ignore donc. Viennent ensuite les termes linéaires qui, au contraire, jouent un rôle extrêmement important. Ces termes sont mis en valeur par les dérivées premières et ils permettent de déterminer un plan tangent. Ce n'est qu'aux points stationnaires, lorsque la partie linéaire s'annule et que le plan tangent est horizontal, que les dérivées secondes prennent le relais. Autour de n'importe quel point de base, ces termes constants, linéaires et quadratiques constituent en fait le début d'une *série de Taylor*.

8.5 La série de Taylor

Introduction. Nous allons maintenant considérer les séries tronquées de la section précédente dans leur entièreté. Ce type de série de puissance, appelée de manière générique *série de Taylor*, fait intervenir l'ensemble des dérivées partielles à tous les ordres d'une fonction $f(x, y)$ évaluées à un *point de base* noté (x_0, y_0). Dans le cas d'une fonction d'une seule variable $f(x)$, ce type de série incluent des puissances entières x^n lorsque le point de base est $x_0 = 0$. Dans le cas d'une fonction $f(x, y)$, chacun des termes de la série est proportionnel à $x^n y^m$ lorsque le point de base est $(x_0, y_0) = (0, 0)$. Ensuite, dans le cas à une variable, on sait que le coefficient devant chaque x^n

correspond à la n-ième dérivée de $f(x)$, notée $d^n f/dx^n$, évaluée au point $x_0 = 0$ et multipliée par $1/n!$. De manière analogue, pour le cas à deux variables, le coefficient devant chaque terme $x^n y^m$ correspond à la dérivée mixte $(\partial/\partial x)^n(\partial/\partial y)^m f(x,y)$ évaluée au point de base (x_0, y_0) (ce que l'on indique avec un indice $_0$) et multiplié par $1/(n!m!)$.

Le développement. La *série de Taylor* d'une fontion $f(x,y)$ au *point de base* $(x_0, y_0) = (0,0)$ est une *double somme*, notée de manière concise $\sum\sum a_{nm} x^n y^m$, dont le coefficient du terme principal a_{nm} est la *dérivée mixte* de f d'ordre $n+m$ évaluée au point de base. Au voisinage de $(0,0)$ on peut donc écrire

$$\begin{aligned} f(x,y) &= f(0,0) + x\left(\frac{\partial f}{\partial x}\right)_0 + y\left(\frac{\partial f}{\partial y}\right)_0 + \frac{x^2}{2}\left(\frac{\partial^2 f}{\partial x^2}\right)_0 + xy\left(\frac{\partial^2 f}{\partial x \partial y}\right)_0 + \frac{y^2}{2}\left(\frac{\partial^2 f}{\partial y^2}\right)_0 \\ &\quad + \sum_{n+m>2} \frac{x^n y^m}{n!m!}\left(\frac{\partial^{n+m} f}{\partial x^n \partial y^m}\right)_0 \end{aligned}$$

Le sommation ici se fait pour $n+m > 2$, c'est-à-dire que si $m+n = 3$, on a (par exemple) les termes avec $(n=0, m=3)$, $(n=1, m=2)$, $(n=2, m=1)$ et $(n=3, m=0)$. Si le point de base passe en (x_0, y_0), les puissances deviennent $(x-x_0)^n(y-y_0)^m$, et toutes les dérivées sont calculées à ce nouveau point de base.

Commentaire. Dans le développement de Taylor ci-dessus, les trois premiers termes correspondent l'**approximation linéaire** de $f(x,y)$. Ils donnent le plan tangent à la surface définie par $z = f(x,y)$ au point de base. On note que, selon la notation, le terme x^2 correspond à $n = 2$ et $m = 0$ donc $n!m! = 2$ et le terme xy correspond à $n = m = 1$ soit $n!M! = 1$. Dans la continuité de ce que nous avons discuté précédemment, on remarque c'est la *partie quadratique* $\frac{1}{2}(ax^2 + 2bxy + cy^2)$ qui contrôle lorsque la partie linéaire est nulle. Ceci justifie a posteriori le rôle proéminent jouer par les dérivées secondes pour déterminer la position de la surface $z = f(x,y)$ par rapport au plan tangent (horizontal) au point stationnaire.

Exemples 45. On peut facilement vérifier que toutes les dérivées de $f(x,y) = e^{x+y}$ sont égales à un à l'origine. La série Taylor est donc

$$e^{x+y} = 1 + x + y + \frac{x^2}{2} + xy + \frac{y^2}{2} + \cdots = \sum\sum \frac{x^n y^m}{n!m!}$$

Dans ce cas précis on a $ac = b^2$, le cas spécial qui a été omis jusqu'à présent. C'est la version en deux dimensions d'un *point d'inflexion* que l'on a pour $f(x) = x^3$ en $x = 0$. En fait, les dérivées secondes ne permettent pas de décider de la concavité de la fonction. Lorsque $f_{xx} f_{yy} = (f_{xy})^2$, il faut en fait examiner les dérivées d'ordre supérieures.

La généralisation de la série de Taylor à un **nombre fini de variables** se formule plus simplement utilisant une **notation matricielle**. Cette formulation est présentée brièvement dans l'**annexe C**.

Conclusion. Une dernière question en guise de conclusion, *comment calculer un minimum numériquement ?* Un bon moyen est de résoudre $f_x = 0$ et $f_y = 0$ où les fonctions $g = f_x$ et $h = f_y$ sont celles qui interviennent dans la méthode de Newton présentée dans la **Section 5.3**. Au point courant (x_n, y_n), les dérivées de g et h donnent des équations linéaires pour les pas Δx et Δy. Ensuite, le point suivant $x_{n+1} = x_n + \Delta x, y_{n+1} = y_n + \Delta y$ provient de l'ajout de ces pas au point courant. En entrée on a donc le point (x_n, y_n), en sortie le nouveau point (x_{n+1}, y_{n+1}) et le système d'équations linéaires est de la forme

$$\begin{cases} (g_x)\,\Delta x + (g_y)\,\Delta y = -g(x_n, y_n) \\ (h_x)\,\Delta x + (h_y)\,\Delta y = -h(x_n, y_n) \end{cases} \quad \text{ou encore} \quad \begin{cases} (f_{xx})\,\Delta x + (f_{xy})\,\Delta y = -f_x(x_n, y_n) \\ (f_{xy})\,\Delta x + (f_{yy})\,\Delta y = -f_y(x_n, y_n) \end{cases} \tag{8.1}$$

Donc, si les dérivées secondes de f sont disponibles, on peut utiliser la méthode de Newton. Lorsque le problème devient lourd et que les dérivées du deuxième ordre et au-delà sont d'une forme complexe, difficile à implémenter numériquement, il existe une méthode alternative appelée *méthode de la descente de gradient (steepest descent method)*. De manière triviale, le principe de cette approche consiste à « descendre » la surface définit par l'équation $z = f(x, y)$, comme un rocher dégringolant le versant d'une montagne, afin de trouver un minimum. Dans cette démarche on utiliser le fait que la direction avec la pente la plus raide en tout point est donnée par le *gradient*, avec un signe moins pour descendre au lieu de monter. On se déplace alors dans la direction $\Delta x = -s\,\partial f/\partial x$ et $\Delta y = -s\,\partial f/\partial y$. À ce stade la question est de savoir jusqu'où faut-il descendre? Comme pour un rocher, ce n'est parce que l'on démarre avec la pente la plus raide que l'on va automatiquement atteindre le minimum. La taille du pas doit être surveillée, pour terminer l'étape lorsque la fonction f recommence à augmenter (voir **Exercice 40**). À la fin de chaque étape, il est donc nécessaire de calculer les dérivées premières et de recommencer dans une nouvelle direction avec la pente la plus raide.

A Contraintes et multiplicateurs de Lagrange

Introduction. Cette section fait face à un problème pratique. Nous minimisons souvent une fonction $f(x, y)$ tandis qu'une autre fonction $g(x, y)$ est fixe. Il existe une **contrainte** sur x et y, donnée par $g(x, y) = k$. Cela limite le matériel disponible ou les fonds disponibles ou l'énergie disponible. Avec cette contrainte, le problème est de faire le mieux possible ($f_{\min}$ ou $f_{\min}$). Au minimum absolu de $f(x, y)$, l'exigence $g(x, y) = k$ est probablement violée. Dans ce cas, le point minimum n'est pas autorisé. Nous ne pouvons pas utiliser $f_x = 0$ et $f_y = 0$, en effet ces équations ne tiennent pas compte de g.

Première étape. Trouvez des équations pour le minimum contraint ou le maximum contraint. Elles impliqueront f_x et f_y ainsi que g_x et g_y, qui donnent des informations locales sur f et g. Pour voir les équations, regardez deux exemples.

Exemples 46. On cherche à minimiser

$$f = x^2 + y^2$$

sous la contrainte $g = 2x + y = k$. On procède à quelques essais, la contrainte autorise $x = 0$, $y = k$, où $f = k^2$. De plus, $(k, 0)$ satisfait la contrainte, et $f = \frac{1}{4}k^2$ est plus petit. Aussi $x = y = \frac{1}{3}k$ donne $f = \frac{2}{9}k^2$ (meilleur jusqu'à présent). Une idée de solution se profile alors. On regarde les courbes de niveau de $f(x, y)$ dans la **Figure 29**. Ce sont des cercles $x^2 + y^2 = c$. Lorsque c est petit, les cercles ne touchent pas la ligne $2x + y = k$. Il n'y a aucun point qui satisfait la contrainte, lorsque c est trop petit. On augmente maintenant c. Finalement, les cercles croissants $x^2 + y^2 = c$ toucheront simplement la ligne $x + 2y = k$. Le point où ils touchent est le gagnant. Il donne la plus petite valeur de c qui peut être obtenue sur la ligne. Le point de contact est $(x_{\min}, x_{\min})$, et la valeur de c est $f_{\min}$.

Notion de multiplicateur. Quelle équation décrit le point de contact? Lorsque le cercle touche la ligne, ils sont tangents. Ils ont la même pente. *Les perpendiculaires au cercle et à la ligne vont dans le même sens.* C'est une remarque essentielle que l'on observe sur la **Figure 29** (gauche). La direction perpendiculaire à $f = c$ est donnée par $\boldsymbol{\nabla} f = (f_x, f_y)$. La direction perpendiculaire à $g = k$ est donnée par $\boldsymbol{\nabla} g = (g_x, g_y)$. Ces deux vecteurs sont parallèles il existe donc un facteur de proportionnalité λ (*lambda*) inconnu entre eux. On en conclut donc que au minimum de $f(x, y)$ soumis à $g(x, y) = k$, le gradient de f est parallèle au gradient de g, avec un nombre inconnu λ comme **multiplicateur**:

$$\boldsymbol{\nabla} f = \lambda \boldsymbol{\nabla} g \qquad \text{donc} \qquad \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x} \quad \text{et} \quad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}$$

Deuxième étape. Il y a maintenant trois inconnues x, y, λ. Il y a aussi trois équations:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \lambda \frac{\partial g}{\partial x} \quad \text{est} \quad & 2x &= 2\lambda \\ \frac{\partial f}{\partial y} &= \lambda \frac{\partial g}{\partial y} \quad \text{est} \quad & 2y &= 2\lambda \\ g(x, y) &= k \quad \text{est} \quad & 2x + y &= k\lambda \end{aligned}$$

Dans la troisième équation, remplacez 2λ par $2x$ et $\frac{1}{2}\lambda$ par y. Alors $2x+y$ est égal à $\frac{2}{5}\lambda$ qui égal à k. Connaissant $\lambda = \frac{2}{5}k$, revenons aux deux premières équations pour x, y et $f_{\min}$:

$$x = \lambda = \frac{2}{5}k, \qquad y = \lambda = \frac{1}{2}\lambda = \frac{1}{5}k, \qquad f_{\min} = \left(\frac{2}{5}k\right)^2 + \left(\frac{1}{5}k\right)^2 = \frac{5}{25}k^2 = \frac{1}{5}k^2$$

Le point gagnant $(x_{\min}, y_{\min})$ est $(\frac{2}{5}k, \frac{1}{5}k)$. Il minimise la « distance au carré », $f = x^2 + y^2 = \frac{1}{5}k^2$, de l'origine à la ligne.

Intepréttation. Quelle est la signification du multiplicateur de Lagrange λ? Une réponse possible est que *la dérivée de $\frac{1}{5}k^2$ est $\frac{2}{5}k$, ce qui équivaut à λ. Le multiplicateur λ est la dérivée de f_{min}, par rapport à k.* Déplacez la ligne par Δk, et $f_{\min}$, change d'environ $\lambda\Delta k$. Ainsi le « multiplicateur de Lagrange » mesure la **sensibilité** par rapport à k.

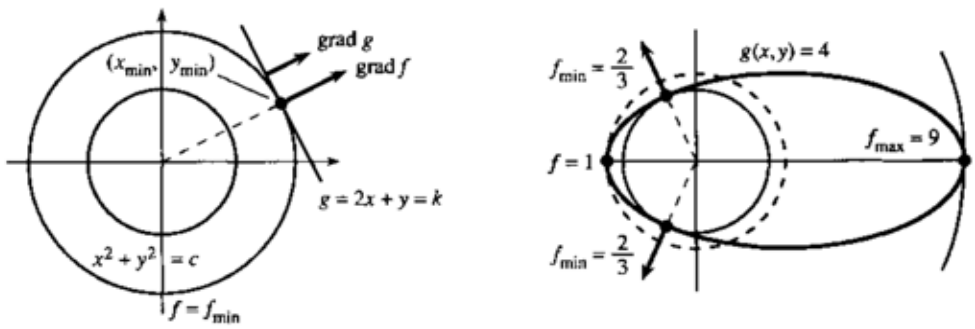


Figure 29: Cercles $f = c$ tangents à la droite $g = k$ et ellipse $g = 4$: gradients parallèles.

Exemples 47. On veut maximiser et minimiser la fonciton

$$f = x^2 + y^2$$

sur l'ellipse $g = (x-1)^2 + 4y^2 = 4$. Les cercles $x^2 + y^2 = c$ augmentent jusqu'à toucher l'ellipse. Le point de contact est $(x_{\min}, y_{\min})$ et la plus petite valeur de c est $f_{\min}$. Au fur et à mesure que les cercles grandissent, ils coupent l'ellipse. Enfin, il y a un point $(x_{\max}, y_{\max})$ où le dernier cercle se touche. Cette plus grande valeur de c est $f_{\max}$. Le minimum et le maximum sont décrits par la même règle: le cercle est tangent à l'ellipse (**Figure 29** droite). Les perpendiculaires vont dans le même sens. Par conséquent (f_x, f_y) est un multiple de (g_x, g_y), et le multiplicateur inconnu est λ:

$$\begin{aligned} f_x &= \lambda g_x \quad : & 2x &= \lambda 2(x-1) \\ f_y &= \lambda g_y \quad : & 2y &= \lambda 8y \\ g &= k \quad : & (x-1)^2 + 4y^2 &= 4 \end{aligned} \tag{A.1}$$

La deuxième équation permet deux possibilités: $y = 0$ ou $\lambda = \frac{1}{4}$. Alors $y = 0$, la dernière équation donne $(x-1)^2 = 4$. Donc $x = 3$ ou $x = -1$. Alors la première équation donne $\lambda = \frac{3}{2}$ ou $\lambda = \frac{1}{2}$. Les valeurs de f sont $x^2 + y^2 = 3^2 + 0^2 = 9$ et $x^2 + y^2 = (-1)^2 + 0^2 = 1$. Suivez maintenant $\lambda = \frac{1}{4}$. La première équation donne $x = -\lambda = \frac{1}{3}$. Alors la dernière équation nécessite $y^2 = \lambda = \frac{5}{9}$. Puisque $x^2 = \frac{1}{9}$, on trouve $x^2 + y^2 = \frac{6}{9} = \frac{2}{3}$. Il s'agit de $f_{\min}$. Conclusion: les équations (A.1) ont quatre solutions, auxquelles le cercle et l'ellipse sont tangents. Les quatre points sont $(3, 0)$, $(-1, 0)$, $(-\frac{1}{3}, \frac{\sqrt{5}}{3})$ et $(-\frac{1}{3}, -\frac{\sqrt{5}}{3})$. Les quatre valeurs off sont $9, 1, \frac{2}{3}, \frac{2}{3}$.

En résumé, les trois équations sont $f_x = \lambda g_x$ et $f_y = \lambda g_y$ et $g = k$. Les inconnues sont x, y et λ. Il n'y a pas de système absolu pour résoudre les équations (sauf si elles sont linéaires; utilisez alors l'élimination ou la *règle de Cramer*). Souvent, les deux premières équations donnent x et y en termes de λ, et la substitution par $g = k$ donne une équation pour λ.

Fonction de Lagrange. Au minimum, la courbe de niveau $f(x, y) = c$ est tangente à la courbe de contrainte $g(x, y) = k$. Si cette courbe de contrainte est donnée paramétriquement par $x(t)$ et $y(t)$, alors pour minimiser $f(x(t), y(t))$ on utilise la règle de la chaîne:

$$\frac{df}{dt} = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dt} = 0 \quad \text{ou} \quad (\mathbf{grad}\, f) \cdot (\text{tangente à la courbe}) = 0$$

Ce calcul prouve que $\boldsymbol{\nabla} f$ est perpendiculaire à la courbe. Ainsi $\boldsymbol{\nabla} f$ *est parallèle à* $\boldsymbol{\nabla} g$. Cela signifie que $(f_x, f_y) = \lambda(g_x, g_y)$. Nous avons perdu $f_x = 0$ et $f_y = 0$. Mais une nouvelle fonction $\mathcal{L}$, appelée **fonction de Lagrange,** a trois dérivées nulles et est donnée par

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot (g(x, y) - k)$$

Ses trois dérivées sont $\mathcal{L}_x = \mathcal{L}_y = \mathcal{L}_\lambda = 0$ à la solution:

$$\frac{\partial \mathcal{L}}{\partial x} = \frac{\partial f}{\partial x} - \lambda\frac{\partial g}{\partial x} = 0 \quad , \quad \frac{\partial \mathcal{L}}{\partial y} = \frac{\partial f}{\partial y} - \lambda\frac{\partial g}{\partial y} = 0 \quad , \quad \frac{\partial \mathcal{L}}{\partial \lambda} = -g + k = 0$$

Notez que $\frac{\partial \mathcal{L}}{\partial \lambda}$ produit automatiquement $g - k$. La contrainte est « intégrée » à L. *Lagrange a inclus un terme* $\lambda \cdot (g - k)$, qui est destiné à être nul, mais ses dérivées sont absolument nécessaires dans les équations! À la solution, $g = k$ et $L = f$ et $\frac{\partial \mathcal{L}}{\partial k} = \lambda$. Ce qui est important sont les relations $f_x = \lambda g_x$ et $f_y = \lambda g_y$, provenant de $\mathcal{L}_x = \mathcal{L}_y = 0$. C'est-à-dire: La contrainte $g = k$ force $dg = g_x dx + g_y dy = 0$. *Cela restreint les mouvements* dx *et* dy. Ils doivent suivre la courbe. Les équations disent que $df = f_x dx + f_y dy$ est égal à λdg. Ainsi df est nul *dans la direction voulue* ce qui est le point clé.

A.1 Maximum et minimum avec deux contraintes

Introduction. Tout le sujet de la min(max)isation s'appelle l'**optimisation**. Ses applications aux décisions d'affaires constituent la *recherche opérationnelle*. Le cas particulier des fonctions linéaires est importante, ce domaine des mathématiques est appelé la ***programmation linéaire***. Une discussion sur ces vastes sujets dépasse largement le cadre de notre discussion, mais nous pouvons faire un pas de plus afin de permettre une deuxième contrainte. La fonction à minimiser ou maximiser est maintenant $f(x, y, z)$. Les contraintes sont $g(x, y, z) = k_1$ et $h(x, y, z) = k_2$. Les multiplicateurs sont λ_1 et λ_2. Nous avons besoin d'au moins trois variables x, y, z car deux contraintes détermineraient complètement x et y.

Méthode. Pour minimiser $f(x, y, z)$ sous réserve de $g(x, y, z) = k_1$ et $h(x, y, z) = k_2$, il faut résoudre cinq équations avec pour solutions $x, y, z, \lambda_1, \lambda_2$. Pour cela il faut cobiner $g = k_1$ et $h = k_2$ avec

$$\frac{\partial f}{\partial x} = \lambda_1\frac{\partial g}{\partial x} + \lambda_2\frac{\partial h}{\partial x} = 0 \quad , \quad \frac{\partial f}{\partial y} = \lambda_1\frac{\partial g}{\partial y} + \lambda_2\frac{\partial h}{\partial y} = 0 \quad , \quad \frac{\partial f}{\partial z} = \lambda_1\frac{\partial g}{\partial z} + \lambda_2\frac{\partial h}{\partial z} = 0 \tag{A.2}$$

La **Figure 30** (gauche) montre la géométrie derrière ces équations. Par commodité, f est $x^2+y^2+z^2$, donc nous minimisons la distance (au carré). Les contraintes $g = x+y+z = 9$ et $h = x+2y+3z = 20$ sont linéaires, leurs graphes sont des plans. Les contraintes conservent (x, y, z) sur les deux plans, et donc sur la ligne où elles se rencontrent. on trouve la *distance au carré de* $(0, 0, 0)$ *à une ligne.*

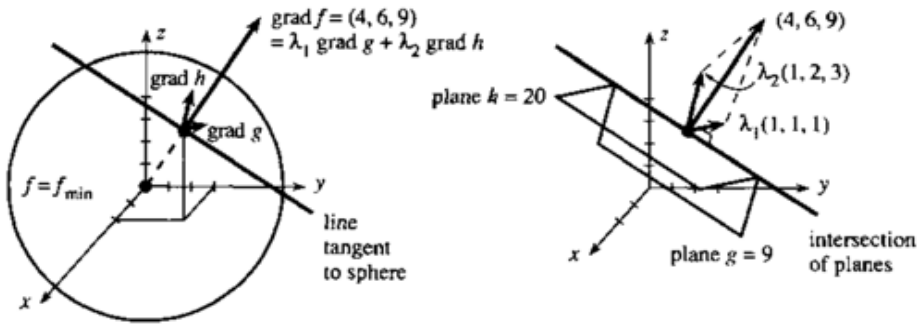


Figure 30: Le vecteur perpendiculaire $\boldsymbol{\nabla} f$ est une combinaison $\lambda_1 \boldsymbol{\nabla} g + \lambda_2 \boldsymbol{\nabla} h$.

Quelle équation résolvons-nous? Les surfaces de niveau $x^2 + y^2 + z^2 = c$ sont des sphères. Ils grandissent à mesure que c augmente. La première sphère à toucher la ligne y est tangente. Ce point de contact donne la solution (le plus petit c). **Les trois vecteurs $\boldsymbol{\nabla} f$, $\boldsymbol{\nabla} g$, $\boldsymbol{\nabla} h$ sont perpendiculaires à la ligne**:

$$\begin{aligned} \textit{ligne tangente à la sphère} \quad &\Rightarrow \quad \boldsymbol{\nabla} f \text{ perpendiculaire à la ligne} \\ \textit{ligne dans les deux plans} \quad &\Rightarrow \quad \boldsymbol{\nabla} g \text{ et } \boldsymbol{\nabla} h \text{ perpendiculaire à la ligne} \end{aligned}$$

Ainsi $\boldsymbol{\nabla} f$, $\boldsymbol{\nabla} g$, $\boldsymbol{\nabla} h$ sont dans le même plan, perpendiculaire à la ligne. Avec trois vecteurs dans un plan, $\boldsymbol{\nabla} f$ est une combinaison de $\boldsymbol{\nabla} g$ et $\boldsymbol{\nabla} h$:

$$(f_x, f_y, f_z) = \lambda_1 (g_x, g_y, g_z) + \lambda_2 (h_x, h_y, h_z) \tag{A.3}$$

C'est l'équation clé (A.3). Il s'applique aussi bien aux surfaces courbes qu'aux plans.

Exemples 48. On veut minimiser la fonction

$$x^2 + y^2 + z^2$$

lorsque $x + y + z = 9$ et $x + 2y + 3z = 20$. Dans la **Figure 30** (droite), les normales à ces plans sont $\boldsymbol{\nabla} g = (1, 1, 1)$ et $\boldsymbol{\nabla} h = (1, 2, 3)$. Le gradient of $f = x^2 + y^2 + z^2$ est $(2x, 2y, 2z)$. Les équations (A.2) et (A.3) sont

$$2x = \lambda_1 + \lambda_2 \quad , \quad 2y = \lambda_1 + \lambda_2 \quad , \quad 2z = \lambda_1 + 3\lambda_2.$$

On remplace x, y, z dans les deux autres équations $g = x + y + z = 9$ et $h = 20$:

$$\frac{\lambda_1 + \lambda_2}{2} + \frac{\lambda_1 + 2\lambda_2}{2} + \frac{\lambda_1 + 3\lambda_2}{2} = 9 \quad \text{et} \quad \frac{\lambda_1 + \lambda_2}{2} + 2\frac{\lambda_1 + 2\lambda_2}{2} + 3\frac{\lambda_1 + 3\lambda_2}{2} = 20$$

Après avoir multiplié par 2, ceux-ci se simplifient à $3\lambda_1 + 6\lambda_2 = 18$ et $6\lambda_1 + 14\lambda_2 = 40$. Les solutions sont $\lambda_1 = 2$ et $\lambda_2 = 2$. Les équations précédentes donnent maintenant $(x, y, z) = (2, 3, 4)$. La fonction de Lagrange à deux contraintes est $\mathcal{L}(x, y, z, \lambda_1, \lambda_2) = f - \lambda_1(g - k_1) - \lambda_2(h - k_2)$. Ses cinq dérivées sont nulles, ce sont nos cinq équations. Lagrange a augmenté le nombre d'inconnues de 3 à 5, en ajoutant λ_1 et λ_2. Le meilleur point $(2, 3, 4)$ donne $f_{\min} = 29$. Les λ donnent $\frac{\partial \lambda}{\partial k}$, la sensibilité aux changements de 9 et 20.

A.2 Contraintes d'inégalité

Introduction. En pratique, les applications impliquent des **inégalités** aussi bien que des équations. Les contraintes peuvent être $g \leq k$ et $h \geq 0$. Le premier signifie: il n'est pas nécessaire

d'utiliser toute la ressource k, mais vous ne pouvez pas en utiliser plus. Le second signifie: h mesure une quantité qui ne peut pas être négative. **Au point minimum, les multiplicateurs doivent satisfaire les mêmes inégalités**: $\lambda_1 \leq 0$ et $\lambda_2 \geq 0$. Il y a des inégalités sur les λ quand il y a des inégalités dans les contraintes.

Raisonnement. À cause de la relation $g \leq k$, le minimum peut être *sur ou à l'intérieur* de la courbe de contrainte. À l'intérieur de la courbe, où $g < k$, nous sommes libres de nous déplacer dans toutes les directions. La contrainte n'est pas vraiment contraignante. Cela ramène $f_x = 0$ et $f_y = 0$ et $\lambda = 0$, il s'agit d'un minimum ordinaire. Sur la courbe, où $g = k$ contraint le minimum à descendre, nous avons $\lambda < 0$. Nous ne savons pas à l'avance à quoi s'attendre. Pour 100 contraintes $g_i \leq k_i$, il y a donc 100 valeurs de λ possible. Certains λ sont nuls (lorsque $g_i < k_i$) et certains sont différents de zéro (lorsque $g_i = k_i$). Ce sont ces 2^{100} possibilités qui rendent l'optimisation intéressante. En **programmation linéaire** à deux variables, les contraintes sont $x \geq 0$, $y \geq 0$.

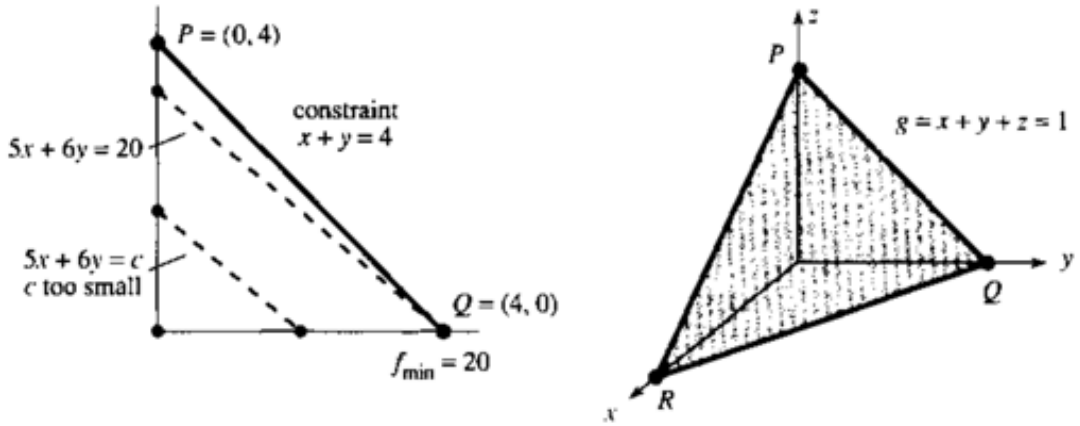


Figure 31: Programmation linéaire: f et g sont linéaires, les inégalités limitent les valeurs de x, y.

Exemples 49. On veut minimiser la fonction

$$f = 5x + 6y$$

avec les contraintes $g = x + y = 4$ et $h = x \geq 0$ et $H = y \geq 0$. La contrainte $g = 4$ est une équation, h et H donnent des inégalités. Chacun a son propre multiplicateur de Lagrange, et les inégalités nécessitent $\lambda_2 \geq 0$ et $\lambda_3 \geq 0$. Les dérivées de f, g, h, H ne posent aucun problème à calculer:

$$\begin{aligned} \frac{\partial f}{\partial x} &= \lambda_1 \frac{\partial g}{\partial x} + \lambda_2 \frac{\partial h}{\partial x} + \lambda_3 \frac{\partial H}{\partial x} \quad \text{conduit à} \quad 5 = \lambda_1 + \lambda_2 \\ \frac{\partial f}{\partial y} &= \lambda_1 \frac{\partial g}{\partial y} + \lambda_2 \frac{\partial h}{\partial y} + \lambda_3 \frac{\partial H}{\partial y} \quad \text{conduit à} \quad 5 = \lambda_1 + \lambda_3 \end{aligned}$$

Ces équations font λ_3 plus grand que λ_2. Par conséquent $\lambda_3 > 0$, ce qui signifie que la contrainte sur H doit être une équation (l'inégalité pour le multiplicateur signifie l'égalité pour la contrainte). En d'autres termes $H = y = 0$. Alors $x + y = 4$ conduit à $x = 4$. La solution est à $(x_{\min}, y_{\min}) = (4, 0)$, où $f_{\min} = 20$. A ce minimum, $h = x = 4$ est au-dessus de zéro. Le multiplicateur de la contrainte $h \geq 0$ doit être $\lambda_2 = 0$. Alors la première équation donne $\lambda_1 = 5$. Comme toujours, le multiplicateur mesure la « sensibilité ». Lorsque $g = 4$ est augmenté de Δk, le coût $f_{\min} = 20$ est augmenté de $5\Delta k$. En économie, $\lambda_1 = 5$ est appelé un **prix fictif**, *c'est le coût de l'augmentation de la contrainte.*

Derrière cet exemple se cache un problème de géométrie intéressant. La courbe de contrainte $x + y = -4$ est une ligne. Les inégalités $x \geq 0$ et $y \geq 0$ laissent un morceau de cette droite, de P à Q sur la **Figure 31**. Les courbes de niveau$f = 5x + 6y = c$ se déplacent à mesure que c augmente, jusqu'à ce qu'elles touchent la ligne. **Le premier point de contact est $Q = (4, 0)$, qui est la solution. C'est toujours un point final**, ou un coin du triangle PQR. Il donne le plus petit coût $f_{\min}$, qui est $c = 20$.

B Normes

Introduction. La position relative d'un point à un autre point peut être représenté par un vecteur $\boldsymbol{v}$. Ce vecteur $\boldsymbol{v}$ (différent du vecteur nul $\mathbf{0}$) peut ensuite être associé à un nombre positif noté $||\boldsymbol{v}||$ appelé sa **norme**. Ce nombre correspond à la « longueur » du vecteur, c'est-à-dire la distance entre les deux points mentionnés. Il existe de nombreuses manières différentes de mesurer la longueur (c'est-à-dire de nombreuses normes différentes). Cependant toute norme de vecteurs (ou fonctions ou matrices ...) doit partager ces deux propriétés de la valeur absolue $|c|$ d'un nombre:

- Multiplier $\boldsymbol{v}$ par λ (**rescaling**[2]) : $||c \cdot \boldsymbol{v}|| = |\lambda| \cdot \|\boldsymbol{v}\|$
- Additionner $\boldsymbol{v}$ à $\boldsymbol{w}$ (**inégalité triangulaire**) : $\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$

Nous commençons par trois normes spéciales, de loin les plus importantes. Elles sont les normes $\boldsymbol{\ell^2}$, $\boldsymbol{\ell^1}$ et $\boldsymbol{\ell^\infty}$ du vecteur $\boldsymbol{v} = (v_1, \ldots, v_n)$. Le vecteur $\boldsymbol{v}$ est dans $\mathbb{R}^n$ (v_i réels) ou dans $\mathbb{C}^n$ (v_i complexes):

- $\boldsymbol{\ell^2}$ **norm** = **norme Euclidienne** : $\|\boldsymbol{v}\|_2 = \sqrt{|v_1|^2 + \cdots + |v_n|^2}$
- $\boldsymbol{\ell^1}$ **norm** = **1-norme** : $\|\boldsymbol{v}\|_2 = |v_1| + \cdots + |v_n|$
- $\boldsymbol{\ell^\infty}$ **norm** = **norme max** : $\|\boldsymbol{v}\|_\infty$ = **maximum parmi** $|v_1|, \ldots, |v_n|$

Le vecteur $\boldsymbol{v} = (1, 1, \ldots, 1)$ a la norme $\|\boldsymbol{v}\|_2 = \sqrt{n}$ et $\|\boldsymbol{v}\|_1 = n$ et $\|\boldsymbol{v}\|_\infty = 1$. Ces trois normes sont les cas particuliers $p = 2$ et $p = 1$ et $p = \infty$ de la norme ℓ^p $\|\boldsymbol{v}\|_p = (|v_1|^p + \cdots + |v_n|^p)^{1/p}$. La **Figure 32** montre des vecteurs avec une **norme unité**: $p = 1/2$ est illégale.

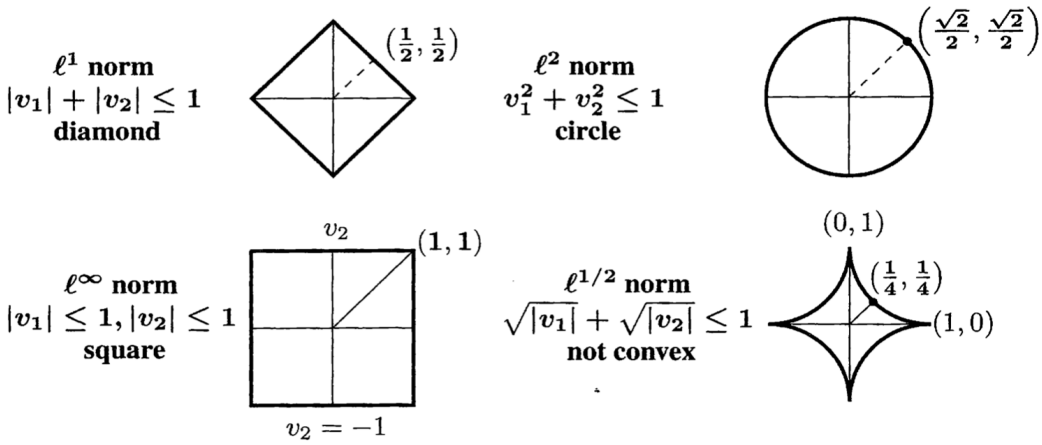


Figure 32: Les normes vectorielles $\|\boldsymbol{v}\|_1$, $\|\boldsymbol{v}\|_2$, $\|\boldsymbol{v}\|_\infty$, et un échec ($p = 0$ ne marche pas également).

L'échec pour $p = 1/2$ provient de l'inégalité triangulaire : $(1, 0)$ et $(0, 1)$ sont de norme 1, mais leur somme $(1, 1)$ a la norme $2^{1/p} = 4$. Seules les valeurs $1 < p < \infty$ produisent une norme acceptable $\|\boldsymbol{v}\|_\infty$.

Le minimum de $\|\boldsymbol{v}\|_p$ sur la ligne $a_1 v_1 + a_2 v_2 = 1$. Quel point sur une ligne diagonale comme $3v_1 + 4v_2 = 1$ est le plus proche de $(0, 0)$? La réponse (et la signification de « plus proche ») dépendra de la norme. C'est une autre façon de voir les différences importantes entre $\boldsymbol{\ell^1}$ et $\boldsymbol{\ell^2}$ et

[2] mise à l'échelle

$\boldsymbol{\ell^\infty}$. Nous verrons un premier exemple d'une fonctionnalité très spéciale: **La minimisation dans $\boldsymbol{\ell^1}$ produit des solutions éparses**. Pour voir le point le plus proche de $(0,0)$, développez le cercle $\boldsymbol{\ell^1}$ et $\boldsymbol{\ell^2}$ et $\boldsymbol{\ell^2}$ carré jusqu'à ce qu'ils touchent la ligne diagonale. Pour chaque p, ce point de contact $\boldsymbol{v}^*$ résoudra notre *problème d'optimisation*:

Minimiser $\|\boldsymbol{v}\|_p$, parmi les vecteurs (v_1, v_2) sur la ligne $3v_1 + 4v_2 = 1$

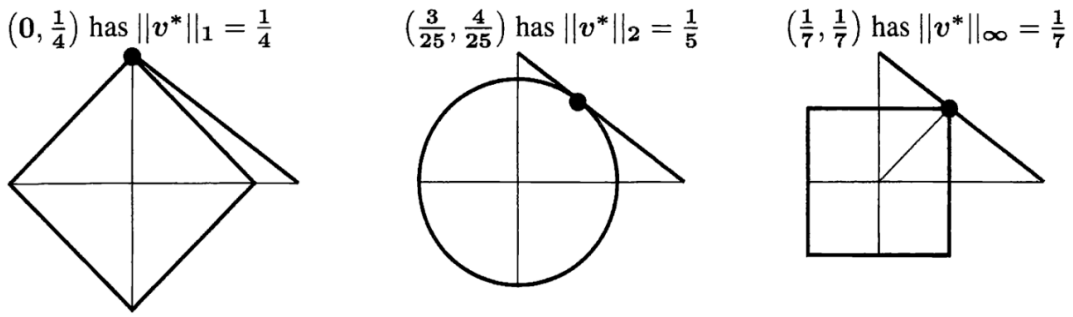


Figure 33: Les solutions $\boldsymbol{v}^*$ minimisation de $\boldsymbol{\ell^1}$ et $\boldsymbol{\ell^2}$ et $\boldsymbol{\ell^\infty}$ minimizations. Le premier cas est *épars*.

Interprétation. La première figure montre une propriété très importante de la solution de minimisation du problème $\boldsymbol{\ell^1}$: **Cette solution $\boldsymbol{v}^*$ n'a aucune composante**. Le vecteur $\boldsymbol{v}^*$ est « *épars*[3] ». C'est parce que **un losange touche une ligne en un point pointu**. La ligne (ou hyperplan en dimensions supérieures) contient les vecteurs qui résolvent les m contraintes $\boldsymbol{av} = \boldsymbol{b}$. La surface du diamant contient des vecteurs avec la même norme $\boldsymbol{\ell^1}$. Le diamant s'agrandit pour rencontrer la ligne en un coin du diamant! Le point essentiel est que **les solutions à ces problèmes sont rares**. Ils ont peu de composantes différentes de zéro et ces composantes ont une signification. En revanche, la solution des moindres carrés (utilisant $\boldsymbol{\ell^2}$) a de nombreux composantes petites et non intéressantes. En mettant au carré, ces composants deviennent très petits et n'affectent guère la distance $\boldsymbol{\ell^2}$. Une dernière observation: la **norme $\boldsymbol{\ell^0}$** d'un vecteur $\boldsymbol{v}$ compte le nombre de composantes différents de zéro. Mais ce n'est pas une vraie norme. Les points avec $\|\boldsymbol{v}\|_0 = 1$ se trouvent sur l'axe x ou l'axe y, une seul composante diffère de zéro. La **Figure 33** pour $p = 1/2$ devient encore plus extrême, juste une croix ou un « squelette » le long des deux axes. Bien entendu, ce squelette n'est pas du tout convexe. La « norme zéro » viole l'exigence fondamentale que $||2\boldsymbol{v}|| = 2||\boldsymbol{v}||$. En fait $||2\boldsymbol{v}||_0 = \|\boldsymbol{v}\|_0 =$ nombre de composantes non nulles dans $\boldsymbol{v}$. L'observation intéressante est que nous pouvons trouver la solution la plus éparse à $\boldsymbol{av} = \boldsymbol{b}$ *en utilisant la norme* $\boldsymbol{\ell^1}$. Nous avons « convexifié » ce squelette $\boldsymbol{\ell^0}$ le long des deux axes. Nous avons rempli le squelette et le résultat est le diamant $\boldsymbol{\ell^1}$.

Produits scalaires et angles. La norme $\boldsymbol{\ell^2}$ a une place spéciale. Quand on écrit $||\boldsymbol{v}||$ sans indice, cela indique une norme. Cela connecte la géométrie ordinaire aux produits scalaires[4] $(\boldsymbol{v}, \boldsymbol{w}) = \boldsymbol{v}^\intercal \boldsymbol{w}$ et aux angles θ entre vecteurs:

(1) Produit scalaire = longueur au carré : $\boldsymbol{v} \cdot \boldsymbol{v} = \boldsymbol{v}^\intercal \boldsymbol{v} = \|\boldsymbol{v}\|^2$

(2) Angle θ entre vecteurs $\boldsymbol{v}$ et $\boldsymbol{w}$: $\boldsymbol{v}^\intercal \boldsymbol{v} = \|\boldsymbol{v}\| \, \|\boldsymbol{w}\| \cos\theta$

[3] *sparse* en anglais, c'est-à-dire que beaucoup de ses composantes sont nulles ou encore il contient beaucoup de 0.
[4] Ici le vecteur $\boldsymbol{v}^\intercal$ est le transposé de $\boldsymbol{v}$.

Alors $\boldsymbol{v}$ est orthogonal à $\boldsymbol{w}$ lorsque $\theta = 90°$ et $\cos\theta = 0$ et $\boldsymbol{v}^\intercal\boldsymbol{w} = 0$. Ces deux connexions conduisent à deux inégalités parmi les plus importantes des mathématiques :

- **Cauchy-Schwarz** : $|\boldsymbol{v}^\intercal\boldsymbol{w}| \leq \|\boldsymbol{v}\| \cdot \|\boldsymbol{w}\|$
- **Inégalité triangulaire** : $\|\boldsymbol{v}+\boldsymbol{w}\| \leq \|\boldsymbol{v}\| + ||\boldsymbol{w}\|$

Nous pouvons relier l'égalité de Cauchy-Schwarz à l'équation du produit scalaire avec le cosinus: $|\cos\theta| \leq 1$ signifie que $|\boldsymbol{v}^\intercal\boldsymbol{w}| \leq \|\boldsymbol{v}\|\,\|\boldsymbol{w}\|$. Et cela conduit à son tour à l'inégalité triangulaire, reliant les côtés $\boldsymbol{v}$, $\boldsymbol{w}$, et $\boldsymbol{v}+\boldsymbol{w}$ d'un triangle ordinaire en dimensions n:

(a) Égalité : $\|\boldsymbol{v}+\boldsymbol{w}\|^2 = (\boldsymbol{v}+\boldsymbol{w}, \boldsymbol{v}+\boldsymbol{w}) = \boldsymbol{v}^\intercal\boldsymbol{v} + \boldsymbol{v}^\intercal\boldsymbol{w} + \boldsymbol{w}^\intercal\boldsymbol{v} + \boldsymbol{w}^\intercal\boldsymbol{w}$

(b) Inégalité : $\|\boldsymbol{v}+\boldsymbol{w}\|^2 \leq \|\boldsymbol{v}\|^2 + 2\,\|\boldsymbol{v}\|\,\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2 = (\|\boldsymbol{v}\| + ||\boldsymbol{w}\|)^2$

Cela confirme notre intuition: toute longueur de côté dans un triangle est inférieure à la somme des deux autres longueurs de côté : $\|\boldsymbol{v}+\boldsymbol{w}\| \leq \|\boldsymbol{v}\| + ||\boldsymbol{w}\|$. Égalité dans le cas de la norme $\boldsymbol{\ell^2}$ n'est possible que lorsque le triangle est totalement plat et que tous les angles ont $|\cos\theta| = 1$.

Produits scalaires et $\mathcal{S}$-normes. Une dernière question sur les normes vectorielles. Est-ce que $\boldsymbol{\ell^2}$ est la seule norme liée aux produits scalaires et aux angles? Il n'y a pas de produits scalaires pour $\boldsymbol{\ell^1}$ et $\boldsymbol{\ell^\infty}$. Mais nous pouvons trouver d'autres produits internes qui correspondent à d'autres normes. Par exemple, nous considérons toute matrice définie positive symétrique $\mathcal{S}$ alors

- $\|\boldsymbol{v}\|_{\boldsymbol{S}}^2$ donne une norme pour $\boldsymbol{v}$ in $\mathbb{R}$ (appelée $\mathcal{S}$-norme)
- $(\boldsymbol{v}, \boldsymbol{w})_{\boldsymbol{S}} = \boldsymbol{v}^\intercal\boldsymbol{S}\boldsymbol{w}$ donne un $\mathcal{S}$-produit scalaire pour $\boldsymbol{v}$, $\boldsymbol{w}$ dans $\mathbb{R}^n$

Le produit scalaire $(\boldsymbol{v}, \boldsymbol{w})_{\boldsymbol{S}}$ est d'accord avec $\|\boldsymbol{v}\|_{\boldsymbol{S}}^2$. Nous avons des angles de **(2)**. Nous avons des inégalités de **(b)**. La preuve est dans **(b)** lorsque chaque norme inclut la matrice $\mathcal{S}$. Nous savons que toute matrice définie positive $\mathcal{S}$ peut être factorisée en $\boldsymbol{a}^\intercal\boldsymbol{a}$. Alors la $\mathcal{S}$-norme et le $\mathcal{S}$-produit scalire pour $\boldsymbol{v}$ et $\boldsymbol{w}$ sont exactement la norme $\boldsymbol{\ell^2}$ et le produit scalaire standard pour $\boldsymbol{a}\boldsymbol{v}$ et $\boldsymbol{a}\boldsymbol{w}$. Par conséquent, nous pouvons écrire

$$(\boldsymbol{v}, \boldsymbol{w})_{\boldsymbol{S}} = \boldsymbol{v}^\intercal\boldsymbol{S}\boldsymbol{w} = (\boldsymbol{S}\boldsymbol{v})^\intercal(\boldsymbol{a}\boldsymbol{w}) \quad \text{car} \quad \boldsymbol{S} = \boldsymbol{a}^\intercal\boldsymbol{a}$$

Ce n'est pas une idée impressionnante mais c'est pratique. Les matrices $\mathcal{S}$ et $\boldsymbol{a}$ « pondèrent » les vecteurs et leurs longueurs. Les moindres carrés pondérés ne sont donc que les moindres carrés ordinaires dans cette norme pondérée. Einstein avait besoin d'une nouvelle définition de la longueur et de la distance dans un espace-temps de dimension 4. Lorentz a proposé celle-ci, qu'Einstein a accepté ($c =$ vitesse de la lumière):

$$\boldsymbol{v} = (x, y, z, t) \qquad \text{avec} \qquad \|\boldsymbol{v}\| = x^2 + y^2 + z^2 - c^2t^2$$

S'agit-il d'une norme pour $\mathbb{R}^4$?

Normes et produits scalaires de fonctions. Une fonction $f(x)$ est un « vecteur dans l'espace de fonction » . Cette idée simple donne à l'algèbre linéaire une importance qui va au-delà de l'espace à n dimensions $\mathbb{R}^n$. Toute l'intuition associée à la linéarité se transpose des dimensions finies aux dimensions infinies. La condition fondamentale pour un espace vectoriel est d'autoriser les *combinaisons linéaires* $c\,\boldsymbol{v}+d\,\boldsymbol{w}$ de vecteurs $\boldsymbol{v}$ et $\boldsymbol{w}$. Cette idée s'étend directement aux combinaisons linéaires $c\,f + d\,g$ de fonctions f et g. C'est précisément avec les normes que de nouvelles questions se posent en dimensions infinies. Pensez aux vecteurs particuliers $\boldsymbol{v}_n = (1, 1/2, \dots (1/2)^n, 0, 0, \dots)$ dans la norme habituelle $\boldsymbol{\ell^2}$. Ces vecteurs se rapprochent car $\|\boldsymbol{v}_n - \boldsymbol{v}_M\| \to 0$ lorsque $n \to \infty$ et $N \to \infty$. Pour qu'un espace vectoriel soit « complet », chaque suite convergente $\boldsymbol{v}_n$ doit avoir une limite $\boldsymbol{v}_\infty$ dans l'espace: $\|\boldsymbol{v}_n - \boldsymbol{v}_\infty\| \to 0$.

(i) L'espace des vecteurs infinis $\boldsymbol{v} = (v_1, \dots, v_N, 0, 0, \dots)$ se terminant par tous les zéros n'est ***pas complet*** .

(ii) L'espace des vecteurs avec $\|\boldsymbol{v}\|^2 = |v_1|^2 + |v_2|^2 + \cdots < +\infty$ est ***complet***. Un vecteur comme $\boldsymbol{v}_\infty = (1, 1/2, 1/4, 1/8, \dots)$ est inclus dans cet espace mais pas le vecteur $\mathbf{1}$. Cela ne se termine pas par des zéros.

Deux noms célèbres sont associés à des espaces vectoriels de dimension infinie complets:

- Un ***espace de Banach*** est un espace vectoriel complet avec une norme $\|\boldsymbol{v}\|$ satisfaisant les règles **(i)** et **(ii)**.
- Un ***espace de Hilbert*** est un espace de Banach également muni d'un produit scalaire tel que $(\boldsymbol{v}, \boldsymbol{v}) = \|\boldsymbol{v}\|^2$.

Ces espaces sont de dimension infinie lorsque les vecteurs ont une infinité de composantes:

- $\boldsymbol{\ell^1}$ est un espace de Banach de norme $\|\boldsymbol{v}\| = |v_1| + |v_2| + \dots$
- $\boldsymbol{\ell^2}$ est un espace de Hilbert car il est muni du produit scalaire $(\boldsymbol{v}, \boldsymbol{w}) = v_1 w_1 + v_2 w_2 + \dots$
- $\boldsymbol{\ell^\infty}$ est un espace de Banach de norme $\|\boldsymbol{v}\|_\infty$ = supremum (borne supérieure) des nombres $|v_1|, |v_2|, \dots$

Notre intérêt particulier est dans les espaces de fonctions. Les vecteurs peuvent être, par exemple, des fonctions $f(x)$ définies pour $0 < x < 1$.

- $\boldsymbol{L^1}[0,1]$ est un espace de Banach avec $||\boldsymbol{f}|| = \int_0^1 |f(x)|\, dx$
- $\boldsymbol{L^2}[0,1]$ est un espace de Hilbert avec $(f, g) = \int_0^1 f(x)g(x)dx$ et $\|\boldsymbol{f}\| = \int_0^1 |f(x)|^2 dx$.
- $\boldsymbol{L^\infty}[0,1]$ est un espace de Banach avec $\|\boldsymbol{f}\|_\infty$ = supremum de $|f(x)|$.

Notez le parallèle entre les sommes des composantes dans $\boldsymbol{\ell^1}$ et les intégrales de fonctions dans $\boldsymbol{L^1}$. De même pour les sommes de carrés dans $\boldsymbol{\ell^2}$ et les intégrales de $|f(x)|^2$ dans $\boldsymbol{L^2}$. *Ajouter ou intégrer.*

Régularité des fonctions. Il existe de nombreux types d'espaces fonctionnels. Cette partie des mathématiques est l'« **analyse fonctionnelle** » . Souvent, un espace de fonctions regroupe toutes les fonctions avec un niveau de régularité (c'est-à-dire de continuité et de différentiabilité) spécifique. Un exemple remarquable est l'espace $\boldsymbol{C}[0,1]$ contenant toutes les **fonctions continues**, en effet:

$$f \text{ appartient à } \boldsymbol{C}[0,1] \text{ et } \|\boldsymbol{f}\|_{\boldsymbol{C}} = \max |f(x)| \text{ si } f(x) \text{ est continue pour tout } 0 < x < 1$$

La norme max dans l'espace de fonction $\boldsymbol{C}$ est comme la norme $\boldsymbol{\ell^\infty}$ pour les vecteurs. Nous pouvons augmenter le niveau de régularité à $\boldsymbol{C^1}[0,1]$ ou $\boldsymbol{C^2}[0,1]$. Ensuite, la première dérivée ou la deuxième dérivée doivent également être continues. Ce sont des espaces de Banach mais pas des espaces de Hilbert. Leurs normes ne proviennent pas des produits scalaires, comparez les deux formules suivantes:

$$\|\boldsymbol{f}\|_{\boldsymbol{C^1}} = \|\boldsymbol{f}\|_{\boldsymbol{C}} + \left\| \frac{d\boldsymbol{f}}{dx} \right\|_{\boldsymbol{C}} \qquad \|\boldsymbol{f}\|_{\boldsymbol{C^2}} = \|\boldsymbol{f}\|_{\boldsymbol{C}} + \left\| \frac{d^2\boldsymbol{f}}{dx^2} \right\|_{\boldsymbol{C}}$$

Si on veut un espace de Hilbert $\boldsymbol{H^1}$ alors nous construisons sur l'espace habituel $\boldsymbol{L^2}$ (qui est $\boldsymbol{H^0}$):

$$\|\boldsymbol{f}\|_{\boldsymbol{H^1}} = \|\boldsymbol{f}\|^2 + \left\| \frac{d^2\boldsymbol{f}}{dx^2} \right\|^2 \qquad \text{et} \qquad (\boldsymbol{f}, \boldsymbol{g})_{\boldsymbol{H^1}} = \int_0^1 f(x)g(x)\, dx + \int_0^1 \frac{df}{dx}\frac{dg}{dx}\, dx$$

Nous terminons cette excursion sauvage dans l'espace fonctionnel avec trois exemples.

(1) Le vecteur infini $\boldsymbol{v} = (1, 1/2, 1/3, 1/4, \dots)$ est dans $\boldsymbol{\ell^2}$ et $\boldsymbol{\ell^\infty}$. Mais ce n'est pas dans $\boldsymbol{\ell^1}$. La somme de ses composants est infinie.

(2) La fonction « marche » (ou *de Heaviside*) appartient à $\boldsymbol{L}^1$ et $\boldsymbol{L}^2$ et $\boldsymbol{L}^\infty$, mais pas à $\boldsymbol{C}$. La fonction fait un « saut ».

(3) La fonction $\max(0, x)$ appartient à $\boldsymbol{C}$ et $\boldsymbol{H}^1$ mais pas à $\boldsymbol{C}^1$. La pente de cette fonction fait un « saut ».

C Série de Taylor pour une fonction à n variables

Notation. Pour le développement de Taylor d'une fonction de plus de deux variables il est commode d'utiliser la **notation matricielle**. Par exemple la **partie linéaire** de la série de Taylor peut s'écrire (on omet pour l'instant l'indice $(\,)_0$ sur les dérivées pour alléger la notation) comme un *produit scalaire*:

$$x\left(\frac{\partial f}{\partial x}\right)+y\left(\frac{\partial f}{\partial y}\right)=\begin{pmatrix} x & y \end{pmatrix}\cdot\begin{pmatrix} f_x \\ f_y \end{pmatrix}=\boldsymbol{x}\cdot\boldsymbol{\nabla} f$$

Ce qui se généralise de manière évidente puisque pour n variables on a toujours

$$\boldsymbol{x}\cdot\boldsymbol{\nabla} f=x_1\left(\frac{\partial f}{\partial x_1}\right)+x_2\left(\frac{\partial f}{\partial x_2}\right)+\cdots+x_n\left(\frac{\partial f}{\partial x_n}\right)$$

avec le vecteur transposé $\boldsymbol{x}^\intercal=(x_1,x_2,\ldots,x_n)$. La **partie quadratique** s'écrit de la manière suivante

$$\frac{x^2}{2}\left(\frac{\partial^2 f}{\partial x^2}\right)+xy\left(\frac{\partial^2 f}{\partial x\partial y}\right)+\frac{y^2}{2}\left(\frac{\partial^2 f}{\partial y^2}\right)=\frac{x^2}{2}\left(\frac{\partial^2 f}{\partial x^2}\right)+\frac{xy}{2}\left(\frac{\partial^2 f}{\partial x\partial y}\right)+\frac{yx}{2}\left(\frac{\partial^2 f}{\partial y\partial x}\right)+\frac{y^2}{2}\left(\frac{\partial^2 f}{\partial y^2}\right)$$

$$=\frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\cdot\begin{pmatrix} xf_{xx}+yf_{xy} \\ xf_{yx}+yf_{yy} \end{pmatrix}=\frac{1}{2}\begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\frac{1}{2}\boldsymbol{x}^\intercal\boldsymbol{H}\boldsymbol{x}$$

où l'on a utilisé les propriétés du produit scalaire et du *produit d'un vecteur avec une matrice*. Dans cette réécriture on a utilisé le vecteur $\boldsymbol{x}$ et son vecteur transposé $\boldsymbol{x}^\intercal$ correspondant respectivement à

$$\boldsymbol{x}=\begin{pmatrix} x \\ y \end{pmatrix}\qquad\text{et}\qquad\boldsymbol{x}^\intercal=\begin{pmatrix} x & y \end{pmatrix}$$

Matrice Hessienne. On a également introduit la **matrice Hessienne** des dérivées partielles du second ordre de la forme

$$\boldsymbol{H}=\begin{pmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{pmatrix}$$

De la même manière que précédemment cette notation demeure inchangée lorsque l'on pass à n variables, en effet

$$\boldsymbol{x}^\intercal\boldsymbol{H}\boldsymbol{x}=\begin{pmatrix} x_1 & x_2 & \ldots & x_n \end{pmatrix}\begin{pmatrix} f_{x_1x_1} & f_{x_1x_2} & \cdots & f_{x_1x_n} \\ f_{x_2x_1} & f_{x_2x_2} & \cdots & f_{x_2x_n} \\ \vdots & \vdots & \ddots & \vdots \\ f_{x_nx_1} & f_{x_nx_2} & \cdots & f_{x_nx_n} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

La série de Taylor l'ordre quadratique pour n variables est donc de la forme générale

$$f(\boldsymbol{x})\simeq f(\boldsymbol{x}_0)+\boldsymbol{x}\cdot\boldsymbol{\nabla} f+\frac{1}{2}\boldsymbol{x}^\intercal\boldsymbol{H}\boldsymbol{x}+\ldots$$

où $\boldsymbol{x}_0$ est le *point de base* (dans $\mathbb{R}^n$) autour duquel on approxime la fonction f. On remarque que (en deux dimensions) le déterminant de la matrice Hessienne est

$$\mathbf{det}\,\boldsymbol{H}=\begin{vmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{vmatrix}=f_{xx}f_{yy}-f_{xy}f_{yx}=f_{xx}f_{yy}-f_{xy}^2$$

ce qui revient à écrire $\mathbf{det}\,\boldsymbol{H}=\mathsf{rt}-\mathsf{s}^2$ dans la notation de la **Section 8.4**. On peut donc en conclure que si $\mathbf{det}\,\boldsymbol{H}>0$ on a minimum ou un maximum selon le signe de r. Et si $\mathbf{det}\,\boldsymbol{H}<0$ on a un point-col. *Cette analyse reste valable pour n variables.* Le **polynôme caractéristique** de $\boldsymbol{H}$ est de la forme suivante

$$\mathbf{det}\,\boldsymbol{H}-\lambda\boldsymbol{I}=(f_{xx}-\lambda)(f_{yy}-\lambda)-f_{xy}^2=(\mathsf{r}-\lambda)(\mathsf{t}-\lambda)-\mathsf{s}^2$$

Analyse spectrale. Les **valeurs propres** de la matrice Hessienne sont des racines de ce polynôme caractéristique, c'est-à-dire solutions de l'équation $\mathbf{det}\,\boldsymbol{H} - \lambda \boldsymbol{I} = 0$. Résoudre cette équation conduit aux deux valeurs propres

$$\lambda_{\pm} = \frac{1}{2}\left(f_{xx} + f_{yy} \pm \sqrt{(f_{xx} - f_{yy})^2 + 4f_{xy}^2}\right)$$

La condition $\mathbf{det}\,\boldsymbol{H} \geq 0$ implique que les valeurs propres $\lambda_{\pm}$ sont réelles et qu'il existe une base dans laquelle la matrice $\boldsymbol{H}$ est diagonale et de la forme

$$\boldsymbol{H} = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \qquad \text{avec} \qquad \mathbf{det}\,\boldsymbol{H} = \mathbf{Tr}\,\boldsymbol{H} = \lambda_+\lambda_-$$

Dans ce sens, analyser si l'on a un minimum, un maximum ou un point-col pour une fonction donnée en un point peut se traduire en terme d'**analyse du signe des valeurs propres de la matrice Hessienne**. Encore une fois cette démarche reste inchangée en dimension n, puisque l'on a jusqu'à n valeurs propres réelles distinctes λ_i $(1 \leq i \leq n)$ avec une matrice Hessienne diagonale de la forme

$$\boldsymbol{H} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix} \qquad \text{avec} \qquad \mathbf{det}\,\boldsymbol{H} = \mathbf{Tr}\,\boldsymbol{H} = \lambda_1\lambda_2\cdots\lambda_n$$

D Exercices

D.1 Surfaces et lignes de niveau

Exercice 1.

(1) Dessinez la surface $z = f(x, y)$ pour ces quatre fonctions:

$$f_1 = \sqrt{4 - x^2 - y^2} \qquad f_2 = 2 - \sqrt{x^2 + y^2} \qquad f_3 = 2 - \frac{1}{2}(x^2 + y^2) \qquad f_4 = 1 + e^{-x^2 - y^2}$$

(2) Les lignes de niveau des quatre fonctions sont _____. Ils entourent le maximum à _____. Dessinez les quatre courbes $f(x, y) = 1$ et classez-les en augmentant le rayon.

(3) Définissez $y = 0$ et calculez la dérivée par rapoprt à x de chaque fonction à $x = 2$. Quelle montagne est la plus plate et laquelle est la plus raide à ce point ?

(4) Définissez $y = 1$ et calculez la dérivée par rapoprt à x de chaque fonction à $x = 1$.

Exercice 2. Sans calculatrice, indiquer la fonction correspondant à chaque figure ci-dessous.

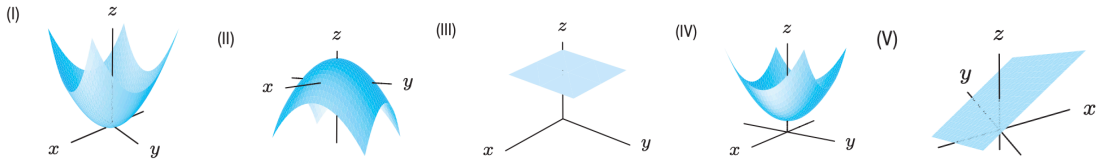
(a) $z = 2 + x^2 + y^2$

(b) $z = 2(x^2 + y^2)$

(c) $z = 2$

(d) $z = 2 - x^2 - y^2$

(e) $z = 2 + 2x - y$



Exercice 3.

(1) Supposons que les lignes de niveau soient des droites parallèles. Dans ce cas la surface doit-elle être un plan ?

(2) Construire une fonction pour laquelle $f = 0$ est un cercle et $f = 1$ ne l'est pas.

(3) Dessinez la carte isoligne pour $f = xy$ avec des lignes de niveau $f = -2, -1, 0, 1, 2$. Décrivez la surface.

Exercice 4.

(1) Le graphe de $w = F(x, y, z)$ est une surface _____-dimensionnelle dans l'espace $xyzw$. Ses lignes de niveaux $F(x, y, z) = c$ sont des surfaces _____-dimensionnelles dans l'espace xyz. Pour $w = x - 2y + z$, ces lignes de niveaux sont _____. Pour $w = x^2 + y^2 + z^2$ ces lignes de niveaux sont _____.

(2) Les lignes de niveaux de $f = x^2 + y^2 + qz^2$ ressemblent à des ballons de rugby lorsque q est _____, comme des ballons de basket lorsque q est _____, et comme les frisbees quand q est _____.

(3) **(a)** Les lignes de niveau de $f(x, y) = \sin(x - y)$ sont _____.

(b) Les lignes de niveau de $g(x, y) = \sin(x^2 - y^2)$ sont _____.

(c) Les lignes de niveau de $h(x, y) = \sin(x - y^2)$ sont _____.

Exercice 5. On considère les fonctions

$$f_1(x_1, x_2) = x_1^2 - x_2^2 \qquad \text{et} \qquad f_2(x_1, x_2) = 2x_1x_2$$

Faire un sketch des lignes de niveaux associées à $f_1(x_1, x_2) = 12$ et $f_2(x_1, x_2) = 16$ sur le même diagramme. Indiquer sur celui-cui les valeurs de $\boldsymbol{x} = [x_1, x_2]^\intercal$ pour lesquelles $\boldsymbol{f}(\boldsymbol{x}) = [f_1(x_1, x_2), f_2(x_1, x_2)]^\intercal = [12, 16]^\intercal$ où l'exposant $^\intercal$ indique un vecteur colonne.

D.2 Limite et continuité

Exercice 6. Dans les cas suivants, évaluez les limites aux valeurs indiquées de x et y. Si la limite n'existe pas, indiquez-le et expliquez pourquoi.

(1) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{4x^2 + 10y^2 + 4}{4x^2 - 10y^2 + 6}$

(2) $\displaystyle\lim_{(x,y)\to(0,1)} \frac{y^2 \sin x}{x}$

(3) $\displaystyle\lim_{(x,y)\to(\pi/4,1)} \frac{y \tan x}{y + 1}$

(4) $\displaystyle\lim_{(x,y)\to(2,5)} \left(\frac{1}{x} - \frac{5}{y}\right)$

(5) $\displaystyle\lim_{(x,y)\to(4,4)} e^{-x^2-y^2}$

(6) $\displaystyle\lim_{(x,y)\to(1,2)} \left(x^2y^3 - x^3y^2 + 3x + 2y\right)$

(7) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy + 1}{x^2 + y^2 + 1}$

(8) $\displaystyle\lim_{(x,y)\to(0,0)} \ln(x^2 + y^2)$

Exercice 7. Dans les cas suivants, utilisez des techniques algébriques pour évaluer la limite.

(1) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^4 - 4y^4}{x^2 + 2y^2}$ **(2)** $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2 - xy}{\sqrt{x} - \sqrt{y}}$ **(3)** $\displaystyle\lim_{(x,y,z)\to(0,0,0)} \frac{x^2 - y^2 - z^2}{x^2 + y^2 - z^2}$

Exercice 8. Pour les cas suivants, évaluez la limite de la fonction en déterminant la valeur que la fonction approche le long des chemins indiqués. Si la limite n'existe pas, expliquez pourquoi.

(1) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy + y^3}{x^2 + y^2}$

(a) Le long de l'axe x $(y = 0)$ **(b)** Le long de l'axe y $(x = 0)$ **(c)** Le long du chemin $y = 2x$

(2) Évaluez $\displaystyle\lim_{(x,y)\to(0,0)} \frac{xy + y^3}{x^2 + y^2}$ en utilisant les résultats de la question précédente.

(3) $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4 + y^2}$

(a) Le long de l'axe x $(y = 0)$ **(b)** Le long de l'axe y $(x = 0)$ **(c)** Le long du chemin $y = x^2$

(4) Évaluez $\displaystyle\lim_{(x,y)\to(0,0)} \frac{x^2y}{x^4 + y^2}$ en utilisant les résultats de la question précédente.

Exercice 9. Discutez de la continuité des fonctions suivantes. Trouvez la plus grande région du plan xy dans laquelle les fonctions suivantes sont continues.

(1) $f(x,y) = \ln(x+y)$

(2) $f(x,y) = \dfrac{1}{xy}$

Exercice 10.

(1) Déterminer la région du plan xy dans laquelle la fonction $g(x,y) = \arctan(\frac{xy^2}{x+y})$ est continue.

(2) En quel point dans l'espace la fonction $g(x,y,z) = x^2 + y^2 - 2z^2$ est-elle continue?

Exercice 11. On considère la fonction $\langle\cdot|\cdot\rangle_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$, définie par

$$\langle \boldsymbol{x}|\boldsymbol{y}\rangle_2 = 2x_1y_1 + 3x_2y_1 + 3x_1y_2 + 5x_2y_2$$

où $\boldsymbol{x} = (x_1, x_2)^\intercal$ et $\boldsymbol{y} = (y_1, y_2)^\intercal$ sont des vecteurs colonnes. Après avoir déterminé la matrice $\boldsymbol{Q}$ telle que $\langle \boldsymbol{x}|\boldsymbol{y}\rangle_2 = \boldsymbol{x}^\intercal \boldsymbol{Q}^2 \boldsymbol{y}$, montrer les résultats suivants ($\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \mathbb{R}^2$ et $\lambda \in \mathbb{R}$) :

(1) $\langle \boldsymbol{x}|\boldsymbol{x}\rangle_2 \geq 0$

(2) $\langle \boldsymbol{x}|\boldsymbol{y}\rangle_2 = \langle \boldsymbol{y}|\boldsymbol{x}\rangle_2$

(3) $\langle \boldsymbol{x}+\boldsymbol{y}|\boldsymbol{z}\rangle_2 = \langle \boldsymbol{x}|\boldsymbol{z}\rangle_2 + \langle \boldsymbol{y}|\boldsymbol{z}\rangle_2$

(4) $\langle \boldsymbol{x}|\lambda\boldsymbol{y}\rangle_2 = \lambda\langle \boldsymbol{x}|\boldsymbol{y}\rangle_2$

Exercice 12. Montrer que pour tout vecteurs $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ on a

$$|\,\|\boldsymbol{x}\| - \|\boldsymbol{y}\|\,| \leq \|\boldsymbol{x} - \boldsymbol{y}\|$$

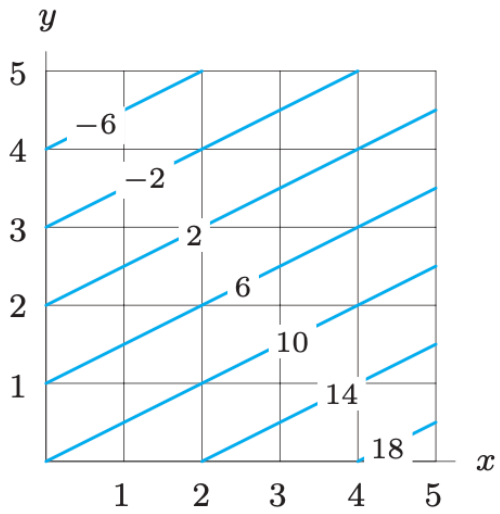
Indice : écrire $\boldsymbol{x} = (\boldsymbol{x} - \boldsymbol{y}) + \boldsymbol{y}$, et utiliser l'inégalité triangulaire. Même chose pour $\boldsymbol{y}$.

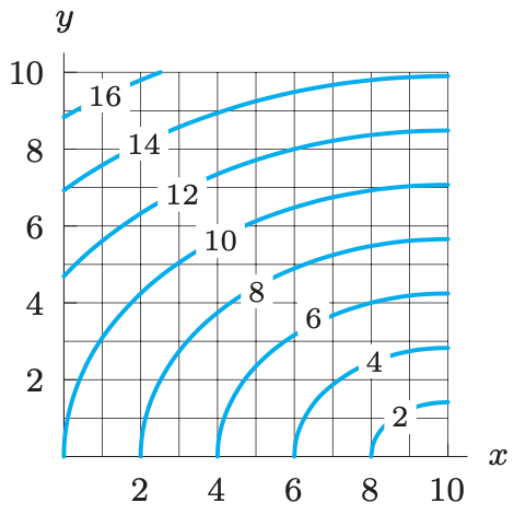
D.3 La dérivée partielle

Exercice 13.

(1) La figure ci-dessous à gauche est un diagramme isoligne pour une fonction $z = f(x,y)$. f_x est positif ou négative? Même question pour f_y. Estimer $f(2,1)$, $f_x(2,1)$ et $f_y(2,1)$.

(2) Donner la valeur approximative de $f_x(3,5)$ en utilisant le diagramme isoligne de $f(x,y)$ ci-dessous à droite.





Exercice 14. Trouvez $\partial f/\partial x$ et $\partial f/\partial y$ pour les fonctions suivantes.

(1) $3x - y + x^2y^2$

(2) $\sin(3x - y) + y$

(3) $x^3y^2 - x^2 - e^y$

(4) xe^{x+4}

(5) $(x + y)/(xy)$

(6) $1/\sqrt{x^2 + y^2}$

(7) $(x^2 + y^2)^{-1}$

(8) $\ln(x + 2y)$

(9) $\ln \sqrt{x^2 + y^2}$

(10) y^x

(11) $\tan^{-1}(y/x)$

(12) $\ln(xy)$

Exercice 15. Calculez $f_{xx}, f_{xy} = f_{yx}$ et f_{yy} pour les fonctions suivantes.

(1) $x^2 + 3xy + 2y^2$

(2) $(x + 3y)^2$

(3) $(x + iy)^3$

(4) e^{ax+by}

(5) $1/\sqrt{x^2 + y^2}$

(6) $(x + y)^n$

(7) $\cos ax \sin by$

(8) $1/(x + iy)$

Exercice 16.

(1) Montrer que $t^{-1/2}e^{-x^2/4t}$ satisfait l'**équation de la chaleur**$f_t = f_{xx}$. Ce $f(x, t)$ est la température à la position x et au temps t due à une source ponctuelle de chaleur à $x = 0$, $t = 0$.

(2) L'équation de la chaleur dans le plan xy est $f_t = f_{xx} + f_{yy}$. Montrer que $f(x, y, t) = e^{-2t} \sin x \sin y$ est une solution. Pour quelle valeur de l'exposant α la fonction $f = e^{\alpha t} \sin 2x \sin 3y$ est une solution de l'équation de la chaleur ?

(3) Trouver les solutions $f(x, y, t) = e^{\beta t} \sin mx \cos ny$ de l'équation de la chaleur $f_t = f_{xx} + f_{yy}$. Montrez que $t^{-1}e^{-x^2/4t}e^{-y^2/4t}$ est aussi une solution.

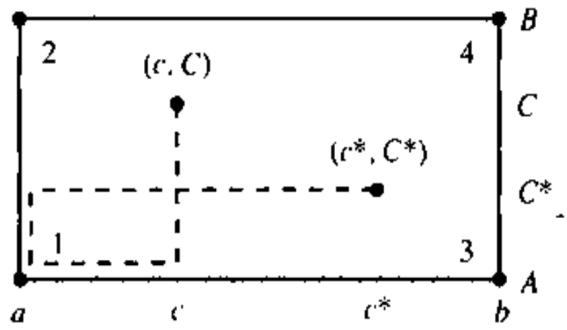
(4) L'*équation d'onde* est $f_{tt} = f_{xx}$. Vérifiez que $f(x, t) = \sin(x + t)$ et $f(x, t) = \sin(x - t)$ sont des solutions. Dessinez les deux graphiques à $t = \pi/4$. Quelle « onde » s'est déplacée vers la gauche et laquelle s'est déplacée vers la droite ?

Exercice 17. La preuve de $f_{xy} = f_{yx}$ se base sur l'étude de $f(x, y)$ dans un petit rectangle. La différence de haut en bas est $g(x) = f(x, B) - f(x, A)$. La différence aux coins de $1, 2, 3, 4$ est :

$$\begin{aligned} Q &= [f_4 - f_3] - [f_2 - f_1] \\ &= g(b) - g(a) \quad \text{(définition de } g) \\ &= (b - a)g_x(c) \quad \text{(Théorème de la Valeur Moyenne)} \\ &= (b - a)[f_x(c, B) - f_x(c, A)] \text{ (calculer } g_x) \\ &= (b - a)(B - A)f_{xy}(c, C) \quad \text{(TVM encore)} \end{aligned}$$

(1) La différence droite-gauche est $h(y) = f(b, y) - f(a, y)$. Le même Q est $h(B) - h(A)$. Modifiez les étapes pour atteindre $Q = (B - A)(b - a)f_y(c^*, C^*)$.

(2) Les deux formes de Q rendent f_{xy} en (c, C) égal à f_{yx} en (c^*, C^*). Réduisez le rectangle vers (a, A). Quelle hypothèse donne $f_{xy} = f_{yx}$ à ce point typique ?



Exercice 18.

(1) Définition de la continuité : $f(x, y)$ est continue en (a, b) si $f(a, b)$ est définie et $f(x, y)$ approche la limite _____ quand (x, y) approche (a, b). Construire une fonction qui n'est pas continue en $(1, 2)$.

(2) Montrer que $x^2y/(x^4 + y^2) \to 0$ le long de toute ligne droite $y = mx$ vers l'origine. Mais lorsque l'on se déplace vers le bas sur la parabole $y = x^2$, le rapport est égal à _____.

(3) Peut-on définir $f(0, 0)$ tel que $f(x, y)$ soit continue en $(0, 0)$?

(a) $f = |x| + |y - 1|$ **(b)** $f = (1 + x)^y$ **(c)** $f = x^{1+y}$

D.4 Plans tangents et approximations linéaires

Exercice 19. Trouvez le plan tangent et le vecteur normal en P.

(1) $z = \sqrt{x^2 + y^2}$, $P(0, 1, 1)$

(2) $x + y + z = 17$, $P(3, 4, 10)$

(3) $z = x/y$, $P(6, 3, 2)$

(4) $z = e^{x+2y}$, $P(0, 0, 1)$

(5) $x^2 + y^2 + z^2 = 6$, $P(1, 2, 1)$

(6) $x^2 + y^2 + 2z^2 = 7$, $P(1, 2, 1)$

(7) $z = x^y$, $P(1, 1, 1)$

(8) $V = \pi r^2 h$, $P(2, 2, 8\pi)$

Exercice 20.

(1) La normale $\boldsymbol{N}$ à la surface $F(x, y, z) = 0$ a des composantes F_x, F_y, F_z. La *ligne normale* a $x = x_0 + F_x t$, $y = y_0 + F_y t$, $z =$ _____. Pour la surface $xyz - 24 = 0$, trouvez le plan tangent et la ligne normale en $(4, 2, 3)$.

(2) Pour la sphère $x^2 + y^2 + z^2 = 9$ trouver l'équation du plan tangent passant par $(2, 1, 2)$. Trouvez également l'équation de la droite normale et montrez qu'elle passe par $(0, 0, 0)$.

(3) Pour $w = xy$ près de (x_0, y_0), l'approximation linéaire est $dw =$ _____. Cela ressemble à la règle _____ pour les dérivées. La différence entre $\Delta w = xy - x_0y_0$ et cette approximation est _____.

Exercice 21.

(1) Si le fournisseur réduit $\mathcal{S}$, **Figure 18** montre que P diminue et Q _____.

(a) Trouvez $P_s = 50/3$ et $P_t = 1/3$ dans l'équation d'économie (5.9) en résolvant les équations au-dessus pour Q_s et Q_t.

(b) Quelle est l'approximation linéaire de Q autour de $s = 0.4$, $t = 10$, $P = 30$, $Q = 50$?

(2) Résouder les équations $P = -0,2\,Q + 40$ et $P = s\,Q + t$ pour P et Q. Ensuite, trouver $\partial P/\partial s$ et $\partial P/\partial t$ explicitement. Pour les même s, t, P, Q vérifiez que 50/3 et 1/3.

(3) Si l'équation « offre = demande » (5.8) change en $P = s\,Q + t = Q + 50$, Trouvez P_s et P_t à $s = 1$, $t = 10$.

D.5 Dérivées directionnelles et gradients

Exercice 22. Utiliser la figure ci-dessous afin d'estimer les dérivées directionnelles suivantes,

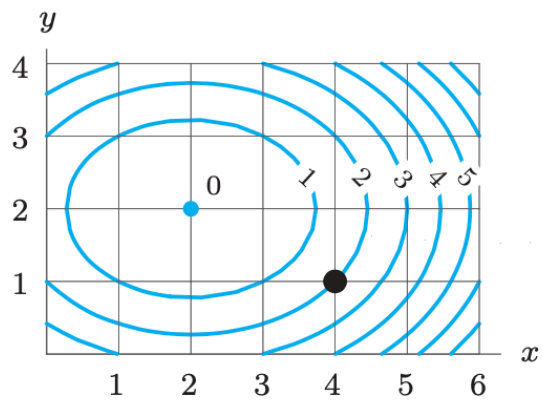
(1) $f_{\boldsymbol{i}}(4,1)$

(2) $f_{\boldsymbol{j}}(4,1)$

(3) $f_{\boldsymbol{u}}(4,1)$ avec $\boldsymbol{u} = (\boldsymbol{i} - \boldsymbol{j})/\sqrt{2}$

(4) $f_{\boldsymbol{u}}(4,1)$ avec $\boldsymbol{u} = (-\boldsymbol{i} + \boldsymbol{j})/\sqrt{2}$

(5) $f_{\boldsymbol{u}}(4,1)$ avec $\boldsymbol{u} = (-2\boldsymbol{i} + \boldsymbol{j})/\sqrt{2}$



Exercice 23. Calculez $\boldsymbol{\nabla} f$, puis $D_{\boldsymbol{u}} f = \boldsymbol{\nabla} f \cdot \boldsymbol{u}$, puis $D_{\boldsymbol{u}} f$ en P.

(1) $f(x,y) = x^2 - y^2$ avec $\boldsymbol{u} = (\sqrt{3}/2, 1/2)$ et $P(1,0)$

(2) $f(x,y) = 3x + 4y + 7$ avec $\boldsymbol{u} = (3/5, 4/5)$ et $P(0, \pi/2)$

(3) $f(x,y) = e^x \cos y$ avec $\boldsymbol{u} = (0,1)$ et $P(0, \pi/2)$

(4) $f(x,y) = y^{10}$ avec $\boldsymbol{u} = (0,-1)$ et $P(1,-1)$

(5) $f(x,y) =$ distance de (x,y) à $(0,3)$, avec $\boldsymbol{u} = (1,0)$ et $P(1,1)$

Exercice 24. Trouvez $\boldsymbol{\nabla} f = (f_x, f_y, f_z)$ pour les fonctions suivantes, issues de la physique (plus précisément de l'électromagnétisme).

(1) $1/\sqrt{x^2 + y^2 + z^2}$ (source ponctuelle à l'origine)

(2) $\ln(x^2 + y^2)$ (ligne source le long de l'axe z)

(3) $1/\sqrt{(x-1)^2 + y^2 + z^2} - 1/\sqrt{(x+1)^2 + y^2 + z^2}$ (dipôle)

Exercice 25. Trouvez la direction $\boldsymbol{u}$ dans laquelle f augmente le plus rapidement à $P = (1,2)$, à quelle vitesse ?

(1) $f(x,y) = ax + by$

(2) $f(x,y) =$ plus petit entre $2x$ et y

(3) $f(x,y) = e^{x-y}$

(4) $f(x,y) = \sqrt{5 - x^2 - y^2}$ (attention...)

Exercice 26. La distance D de (x,y) à $(1,2)$ est $D^2 = (x-1)^2 + (y-2)^2$. Montrer que $\partial D/\partial x = (x-1)/D$ et $\partial D/\partial y = (y-2)/D$ et $\|\boldsymbol{\nabla} D\| = 1$. Le graphe de $D(x,y)$ est _____ avec ses vertex sont en _____.

Exercice 27. On considère la fonction de plusieurs variables

$$f(\boldsymbol{x}) = (\boldsymbol{a}^\intercal \boldsymbol{x})\,(\boldsymbol{b}^\intercal \boldsymbol{x})$$

où $\boldsymbol{a}$, $\boldsymbol{b}$ et $\boldsymbol{x}$ sont des vecteurs de $\mathbb{R}^n$. Déterminer $\boldsymbol{\nabla} f(\boldsymbol{x})$ et la matrice Hessienne $\boldsymbol{H}$ associée à f.

D.6 La dérivée composée ou « chain rule »

Exercice 28. Trouvez df/dt à partir de la chain rule dans les cas suivants.

(1) $f = x^2 + y^2$ avec $x = t$ et $y = t^2$

(2) $f = \sqrt{x^2 + y^2}$ avec $x = t$ et $y = t^2$

(3) $f = xy$ avec $x = 1 - \sqrt{t}$ et $y = 1 + \sqrt{t}$

(4) $f = x/y$ avec $x = e^t$ et $y = 2e^t$

(5) $f = \ln(x + y)$ avec $x = e^t$ et $y = e^t$

(6) $f = x^4$ avec $x = t$ et $y = t$

Exercice 29.

(1) Sur la ligne $x = u_1 t$, $y = u_2 t$, $z = u_3 t$, quelle combinaison de f_x, f_y, f_z donne df/dt? C'est la dérivée directionnelle en 3D.

(2) Pour $f(x,y,t) = x + y + t$ trouvez $\partial f/\partial t$ et $\partial f/\partial t$ quand $x = 2t$ et $y = 3t$. Expliquez la différence.

(3) Supposons que $x_t = t$ et $y_t = 2t$, non constants comme dans (7.2) et (7.3). Pour $f(x,y)$, Trouvez f_t et f_{tt}. La réponse implique $f_x, f_y, f_{xx}, f_{xy}, f_{yy}$.

(4) Déduire $\partial f/\partial r = (\partial f/\partial x)\cos\theta + (\partial f/\partial y)\sin\theta$ à partir de la chain rule. Pourquoi prenons-nous $\partial x/\partial r$ comme $\cos\theta$ et non $1/\cos\theta$?

Exercice 30.

(1) Pour $f = x^2 + y^2 + z^2$ calculer $\partial f/\partial x$ (pas d'indice, x, y, z tous indépendants).

(2) Lorsqu'il y a une autre relation $z = x^2 + y^2$, utilisez-la pour remplacer z et calculer $(\partial f/\partial x)$.

(3) Calculer $(\partial f/\partial x)_y$ en utilisant la chain rule $(\partial f/\partial x) + (\partial f/\partial z)(\partial z/\partial x)$.

(4) Pourquoi cette chain rule ne contient-elle pas $(\partial f/\partial y)(\partial y/\partial x)$?

Exercice 31.

(1) La loi des gaz parfait en physique (thermodynamique) est $PV = nRT$ ou une relation plus générale $F(P,V,T) = 0$. Montrer que la multiplication des trois dérivées dans **Exemple 7.2** donne toujours -1. Trouvez d'abord $(\partial P/\partial V)_T$ à partir de $\partial P/\partial V + (\partial F/\partial P)(\partial P/\partial V)_T = 0$.

(2) Si nous changeons le problème en passant à quatre variables liées par $F(x,y,z,t) = 0$, quel est alors le produit correspondant de quatre dérivées?

Exercice 32. On considère les fonctions $f : \mathbb{R}^2 \to \mathbb{R}$ et $\boldsymbol{g} : \mathbb{R} \to \mathbb{R}^2$ telles que

$$f(\boldsymbol{x}) = \frac{x_1^2}{6} + \frac{x_2^2}{4} \qquad \text{et} \qquad \boldsymbol{g}(t) = [3t+5, 2t-6]^\intercal$$

Soit $F : \mathbb{R} \to \mathbb{R}$ donné par $F(t) = f(\boldsymbol{g}(t))$. Calculer $\frac{dF}{dt}(t)$ en utilisant la « chain rule ».

Exercice 33. On considère les fonctions

$$f(\boldsymbol{x}) = \frac{x_1 x_2}{2} \qquad \text{et} \qquad \boldsymbol{g}(s,t) = [4s+3t, 2s+t]^\intercal$$

Calculer $\frac{\partial}{\partial s} f(\boldsymbol{g}(s,t))$ et $\frac{\partial}{\partial t} f(\boldsymbol{g}(s,t))$ en utilisant la « chain rule ».

Exercice 34. On considère les fonctions

$$\boldsymbol{x}(t) = [e^t + t^3, t^2, t+1]^\intercal \quad (t \in \mathbb{R}) \qquad \text{et} \qquad f(\boldsymbol{x}) = x_1^3 x_2 x_3^2 + x_1 x_2 + x_3$$

avec $\boldsymbol{x} = [x_1, x_2, x_3]^\intercal \in \mathbb{R}^3$. Trouver $\frac{d}{dt} f(\boldsymbol{x}(t))$ en fonction de t.

D.7 Maxima, minima et point-cols

Exercice 35. Trouvez tous les points stationnaires ($f_x = f_y = 0$) dans les cas suivants. Déterminer si c'est un minimum, un maximum ou un point-col. Le test avec a, b, c s'applique avec $a = f_{xx}, b = f_{xy}, c = f_{yy}$.

(1) $x^2 + 2xy + 3y^2$

(2) $xy - x + y$

(3) $x^2 + 4xy + 3y^2 - 6x - 12y$

(4) $x^2 - y^2 + 4y$

(5) $x^2y^2 - x$

(6) $xe^y - e^x$

(7) $-x^2 + 2xy - 3y^2$

(8) $(x+y)^2 + (x+2y-6)^2$

(9) $x^2 + y^2 + z^2 - 4z$

(10) $(x+y)(x+2y-6)$

(11) $(x-y)^2$

(12) $(1+x^2)/(1+y^2)$

(13) $(x+y)^2 - (x+2y)^2$

(14) $\sin x - \cos y$

(15) $x^3 + y^3 - 3x^2 + 3y^2$

(16) $8xy - x^4 - y^4$

Exercice 36.

(1) Un rectangle a des côtés sur les axes x et y et un coin sur la ligne $x + 3y = 12$. Trouvez sa superficie maximale.

(2) Une boîte a un coin à $(0,0,0)$ et toutes les arêtes sont parallèles aux axes. Si le coin opposé (x,y,z) est sur le plan $3x + 2y + z = 1$, quelle position donne le volume maximum? Montrez d'abord que le problème maximise $xy - 3x^2y - 2xy^2$.

(3) (Ajustement en ligne droite) Trouvez x et y pour minimiser l'erreur

$$E = (x+y)^2 + (x+2y-5)^2 + (x+3y-4)^2$$

Montrez que cela donne un minimum et pas un point-col.

(4) (Moindres carrés) Quels nombres x, y se rapprochent le plus de la solution des trois équations $x - y = 1$, $2x + y = -1$, $x + 2y = 1$? Mettre au carré et ajouter les erreurs, $(x - y - 1)^2 +$ _____ $+$ _____. Puis minimisez.

Exercice 37.

(1) Minimisez $f = x^2 + y^2 - x - y$ restreint par

(a) $x \leq 0$ **(b)** $y \geq 1$ **(c)** $x \leq 0$ et $y \geq 1$

(2) Minimisez $f = x^2 + y^2 + 2x + 4y$ dans les régions

(a) $x \leq 0$ **(b)** $y \geq 1$ **(c)** $x \leq 0$ et $y \geq 1$

(3) Maximise et minimise $f = x + \sqrt{3}y$ sur le cercle $x = \cos t$, $y = \sin t$.

(4) L'**Exemple 8.1** a suivi $f = x^2 + xy + y^2 - x - y + 1$ autour du cercle $x^2 + y^2 = 1$. Les quatre points stationnaires ont $x = y$. Calculez f à ces points et localisez le minimum.

Exercice 38. Trouvez, dans les cas suivants, $f_x, f_y, f_{xx}, f_{xy}, f_{yy}$ au point de base. Écrivez l'approximation quadratique de $f(x, y)$, ce qui signifie la série de Taylor à travers les termes du second ordre.

(1) $f = e^{x+y}$ en $(0, 0)$

(2) $f = e^{x+y}$ à $(1, 1)$

(3) $f = \sin x \cos y$ en $(0, 0)$

(4) $f = x^2 + y^2$ à $(1, -1)$

Exercice 39.

(1) La série de Taylor autour de (x, y) s'écrit également avec les incréments h et k: $F(x+h, y+k) = f(x, y) + h_____ + k_____ + \frac{1}{2}h^2_____ - hk_____ + \ldots$. Remplissez ces quatre blancs.

(2) Trouvez les lignes le long desquelles $f(x, y)$ est constant (ces fonctions ont $f_{xx}f_{yy} = f_{xy}^2$ ou $ac = b^2$):

(a) $f = x^2 - 4xy + 4y^2$ **(b)** $f = e^x e^y$

(3) Pour $F(x, y, z)$, les trois premiers termes après $F(0, 0, 0)$ dans la série de Taylor sont _____. Les six termes suivants sont _____.

(4) **(a)** Pour l'erreur $f - f_L$ en approximation linéaire, la série de Taylor à $(0, 0)$ commence par les termes quadratiques _____.

(b) Le graphe de f remonte de son plan tangent (et $f > f_L$) si _____. Alors f est concave vers le haut.

(c) Pour que $(0, 0)$ soit un minimum, nous avons également besoin de _____.

Exercice 40.

(1) Le gradient de $x^2 + 2y^2$ au point $(1, 1)$ est $(2, 4)$. La descente la plus forte se fait le long de la ligne $x = 1 - 2s$, $y = 1 - 4s$ (signe moins pour descendre). Minimisez $x^2 + 2y^2$ par rapport à la taille du pas $\mathcal{S}$. Cela localise le point suivant _____, où la descente la plus forte recommence.

(2) La méthode de Newton minimise $x^2 + 2y^2$ en une seule étape. À partir de $(x_0, y_0) = (1, 1)$, Trouvez Δx et Δy à partir de l'équation (8.1).

(3) Si $f_{xx} + f_{yy} = 0$, montrez que $f(x, y)$ ne peut pas avoir de maximum ou de minimum intérieur (uniquement des point-cols).

(4) La valeur de x théorèmes et y exercices est donné par la fonction $f = x^2 y$ (on suppose). Le maximum auquel un étudiant ou un auteur peut faire face est de $4x + y = 12$. Remplacez $y = 12 - 4x$ et maximisez f. Montrez que la droite $4x + y = 12$ est tangente à la ligne de niveau $x^2 y = f_{\max}$.

Exercice 41. Trouver, en utilisant l'approche présentée dans l'**Appendice C**, les maxima et minima de la fonction

$$f(\boldsymbol{x}) = f(x_1, x_2) = \frac{x_1^3}{3} - 4x_1 + \frac{x_2^3}{3} - 16x_2$$

Exercice 42. Ecrire le développements de Taylor, en utilisant l'approche présentée dans l'**Appendice C**, des fonctions suivantes au point $\boldsymbol{x}_0$. Négliger les termes d'ordre trois supérieur.

(1) $f(\boldsymbol{x}) = x_1^4 + 2x_1^2x_2^2 + x_2^4$, $\boldsymbol{x}_0 = [1, 1]^\intercal$

(2) $f(\boldsymbol{x}) = e^{x_1 - x_2} + e^{x_1 + x_2} + x_1 + x_2 + 1$, $\boldsymbol{x}_0 = [1, 0]^\intercal$

Exercice 43. On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$ définie par

$$f(\boldsymbol{x}) = \boldsymbol{x}^\intercal \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} \boldsymbol{x} + \boldsymbol{x}^\intercal \begin{bmatrix} 3 \\ 5 \end{bmatrix} + 6$$

(1) Trouver le gradient et la matrice Hessienne de f au point $[1, 1]^\intercal$.

(2) Trouver la dérivée directionnelle de f en $[1, 1]^\intercal$ dans la direction (unitaire) de la pente positive maximale.

(3) Trouver un point satisfaisant les conditions d'un extremum. S'agit-il d'un maximum ou d'un minimum?

Exercice 44. On considère la fonction

$$f(\boldsymbol{x}) = f(x_1, x_2) = x_1^2x_2 + x_2^3x_1$$

(a) Dans quelle direction la fonction f décroit-elle le plus rapidement au point $\boldsymbol{x}^{(0)} = [2, 1]^\intercal$?

(b) Quel est le taux d'accroissement de f au point $\boldsymbol{x}^{(0)}$ dans la direction de la pente maximum de f ?

(c) Trouver le taux d'accroissement de f au point $\boldsymbol{x}^{(0)}$ dans la direction $\boldsymbol{d} = [3, 4]^\intercal$.

Exercice 45. On considère la fonction $f : \mathbb{R}^2 \to \mathbb{R}$:

$$f(\boldsymbol{x}) = \boldsymbol{x}^\intercal \begin{bmatrix} 2 & 2 \\ 2 & 1 \end{bmatrix} \boldsymbol{x} + \boldsymbol{x}^\intercal \begin{bmatrix} 3 \\ 4 \end{bmatrix} + 7$$

(a) Trouver la dérivée directionnelle de f en $[0, 1]^\intercal$ dans la direction $[1, 0]^\intercal$.

(b) Trouver tous les points qui satisfont la condition pour un extremum de f. La fonction f possède-t-elle des minima? Si c'est le cas, trouver les tous; sinon, expliquer pourquoi il n'y a pas de minima.