



Analyse I et II

Licence Informatique

Guillaume Metzler

Institut de Communication (ICOM)
Université de Lyon, Université Lumière Lyon 2
Laboratoire ERIC UR 3083, Lyon, France

guillaume.metzler@univ-lyon2.fr

Résumé

L'objectif de ce cours est de présenter les outils nécessaires à l'étude d'une fonction : cela peut être l'étude de ses variations, la recherche d'extrema ou encore d'éventuelles propriétés comme la convexité de ces fonctions. On étudiera également des objets comme les suites numériques réelles.

La première partie de ce document est destinée aux étudiants de première année de Licence Informatique et concerne l'étude des suites numériques réelles ou encore des fonctions réelles à une variable réelle. Nous reviendrons sur des généralités sur les suites numériques réelles : étude de la monotonie, sa nature (récurrente, arithmétique, géométrique), la convergence ou divergence des suites et donc de limite. Enfin, nous introduisons des nouveaux concepts comme la notion de comparaison des suites. Une partie conséquente est ensuite consacrée à l'étude de fonctions réelles. On commencera par présenter quelques fonctions usuelles comme les fonctions linéaires, quadratiques, logarithme et exponentielles, circulaires et hyperboliques. L'objectif étant d'offrir un spectre large des fonctions que l'on pourra rencontrer par la suite. Nous étudierons ensuite les propriétés de continuité et surtout de dérivabilité qui sera essentiel dans l'étude des problèmes en *Science des Données*. Enfin, des bases sur l'intégration seront également fournies pour des besoins ultérieures en probabilités et statistiques.

Dans la deuxième partie, destinée principalement aux étudiants de deuxième année Licence Informatique, nous effectuerons une étude analogue mais pour les fonctions à plusieurs variables. Nous allons profiter des connaissances acquises dans la première partie de ce document mais aussi dans le cours d'*Algèbre Linéaire*¹ afin d'étendre les définitions vues en première année à ce contexte. Nous souhaitons nous diriger vers un point essentiel qui est l'étude des problèmes d'optimisation et les algorithmes de descente de gradient que l'on retrouvera ultérieurement dans des cours de statistiques ou encore d'apprentissage machine.

1. Un cours d'Algèbre Linéaires, dispensé à des étudiants de L2 MIASHS est disponible à l'[adresse suivante](#).

Remerciements Un grand merci à Julien Ah-Pine dont les ressources ont été d'une aide précieuse pour la préparation de la deuxième partie de ce cours.

Table des matières

I	Introduction	6
II	Analyse des fonctions à une variable réelle	12
1	Suites réelles	14
1.1	Généralités sur les suites	14
1.2	Convergence des suites	18
1.3	Divergence des suites	23
1.4	Limite d'une suite monotone	24
1.5	Relations de comparaison	25
1.5.1	Suite négligeable	25
1.5.2	Suite dominée	27
1.5.3	Suites équivalentes	27
2	Fonctions usuelles	31
2.1	Logarithme	31
2.2	Fonction exponentielle	34
2.3	Fonctions puissances	38
2.4	Fonctions circulaires	40
2.5	Fonctions hyperboliques	44
3	Continuité des fonctions réelles	48
3.1	Quelques caractéristiques des fonctions	48
3.2	Continuité	51
3.3	Extrema d'une fonction et fonction lipschitzienne	58
4	Dérivabilité	63
4.1	Définition de la dérivabilité	63
4.2	Dérivation et extrema	66
4.3	Etude des accroissements d'une fonction	67
4.4	Régularité des fonctions réelles	70

5 Etude des fonctions convexes	73
5.1 Caractérisations de la convexité	73
5.2 Lien avec la recherche d'extremum	79
6 Intégration	81
6.1 Généralités sur l'intégrale	81
6.2 Calculs de primitives et intégrales	81
6.3 Quelques propriétés de l'intégrale	88
7 Développements Limités	91
7.1 Formule de Taylor	91
7.2 Développements limités	91
8 Formulaire	94
8.1 Dérivées des fonctions usuelles	94
8.2 Formules de dérivations générales	95
8.3 Développements limités	96
8.4 Primitives des fonctions usuelles	97
III Analyse des fonctions à plusieurs variables	98
9 Généralités sur les fonctions à plusieurs variables	100
10 Continuité et Dérivabilité	103
10.1 Continuité	103
10.2 Fonctions différentiables et dérivables	104
10.3 Fonction différentiables : ordre 2	115
11 Recherche d'extrema et applications aux fonctions convexes	120
11.1 Retour sur la convexité	120
11.2 Conditions d'optimalité	121
11.3 Problème d'optimisation	129

12 Optimisation : descente de gradient	132
12.1 Introduction à la descente de gradient	132
12.2 Descente de gradient à pas optimal	135
12.3 Descente de Gradient à pas conjugué	140
12.4 Descente de Newton	142
12.5 Approximation de la méthode de Newton	143
13 Intégration	148
13.1 Fonctions réelles intégrables	148
13.2 Fonctions à plusieurs variables	151
13.3 Changement de variable	156

Première partie

Introduction

L'analyse des fonctions à variables réelles constituera la base de la plupart des études que l'on pourra mener dans le domaine de la *Science des Données*. Connaître les fonctions, leurs comportements et leurs propriétés sont des éléments essentiels qui permettront de résoudre des tâches complexes comme l'extraction de données ou l'estimation des valeurs optimales des paramètres d'un modèle d'apprentissage automatique. Les modèles d'apprentissage se traduisent très souvent par des fonctions mathématiques dont les propriétés sont connues et simples à étudier et qui vont permettre de relater un phénomène observé dans les données ou ce que l'on cherche à faire. On peut, par exemple, chercher à prédire la valeur du cours d'une action en fonction du temps, trouver une solution la moins coûteuse possible, déterminer le meilleur arrangement possible lorsque l'on dispose de différentes contraintes, etc.

Il faudra donc connaître des propriétés sur la fonction comme la continuité, la dérivabilité voire déterminer des intégrales. En réalité, il y a plusieurs facettes de la Sciences des Données qui sont à prendre en compte :

- pour faire de l'analyse de données de façon générale, ce qui nous amènera à étudier des problèmes d'optimisation.
- la construction de modèles prédictifs qui reposent sur des outils mathématiques et statistiques. Cela permettra d'effectuer des tâches de régression ou de classification
- Etudier les propriétés des modèles sur le plan mathématique permettra également de prouver leur bon fonctionnement, *i.e.* avoir des garanties théoriques sur leurs performances mais aussi que les algorithmes développés convergent bien la solution optimale.

Etude de la complexité des algorithmes Ces études de suites ou de fonctions sont également intéressantes pour tout bon informaticien qui souhaite développer des algorithmes efficaces pour la gestion ou le traitement sur des bases de données. Il n'aura pas échapper à l'apprenant dans ce domaine qu'il est possible d'implémenter une solution à un problème de plusieurs façon différentes. Prenons un exemple simple où l'on souhaite calculer $n!$ (lire *factorielle n*). On peut alors développer une approche dite *itérative* ou *réursive* pour calculer cette valeur

Apprentissage Faire en sorte qu'une intelligence artificielle soit capable de réaliser certaines tâches est loin d'être aisé, car cela nécessite non seulement des données mais aussi des outils analytique, algébriques mais aussi des connaissances en statistiques pour quiconque souhaite développer ou encore étudier des algorithmes sophistiqués. Si l'on retourne même à un usage plus primaire des données, la connaissance des outils

Algorithm 1: Approche récursive

Entrée: $n \in \mathbb{N}$
Sortie : $n!$
Factorielle(n)
1: **if** $n == 1$ **then**
2: return(n)
3: **else**
4: return($n \times$ Factorielle($n - 1$))
5: **end if**

Algorithm 2: Approche itérative

Entrée: $n \in \mathbb{N}$
Sortie : $n!$
1: $a = n$
2: **for** $i = 1, \dots, n - 1$ **do**
3: $a = a * n - i$
4: **end for**

algébriques, comme la manipulation de matrices est capitale pour développer des codes plus efficaces qui permettent de limiter le nombre d'opérations et donc de réduire le temps d'exécution des différents calculs. En ce sens, la mathématique ne peut être dissociée, dans le cadre de la manipulation des données, de l'outil informatique et tous les deux se révèlent être complémentaires dans ce domaine.

L'apprentissage nécessite deux grands outils fondamentaux : l'algèbre linéaire mais aussi les outils d'analyses de fonctions à plusieurs variables ce qui pourra permettre de mettre en place une intelligence capable de discriminer les spams dans une série de mails.

Pour simplifier la représentation, nous allons considérer que les mails sont décrits par deux variables, *i.e.* représentés par deux valeurs et on peut alors les représenter dans un plan, comme cela est montré en Figure 1. Pour simplifier la tâche, on va supposer que l'on souhaite apprendre une droite qui sépare de façon optimale les points rouges des points bleus, dans le sens où l'on effectue le minimum d'erreur.

Nous avons donc besoin de deux éléments ici

- Savoir définir une droite dans un plan (ou un hyperplan dans un espace de dimension quelconque)
- Définir le critère à minimiser, *i.e.* la fonction à minimiser, dont les valeurs dépendent de la droite considéré.

Le premier point est plutôt simple, vous vous souvenez que l'équation d'une droite dans un plan est de la forme

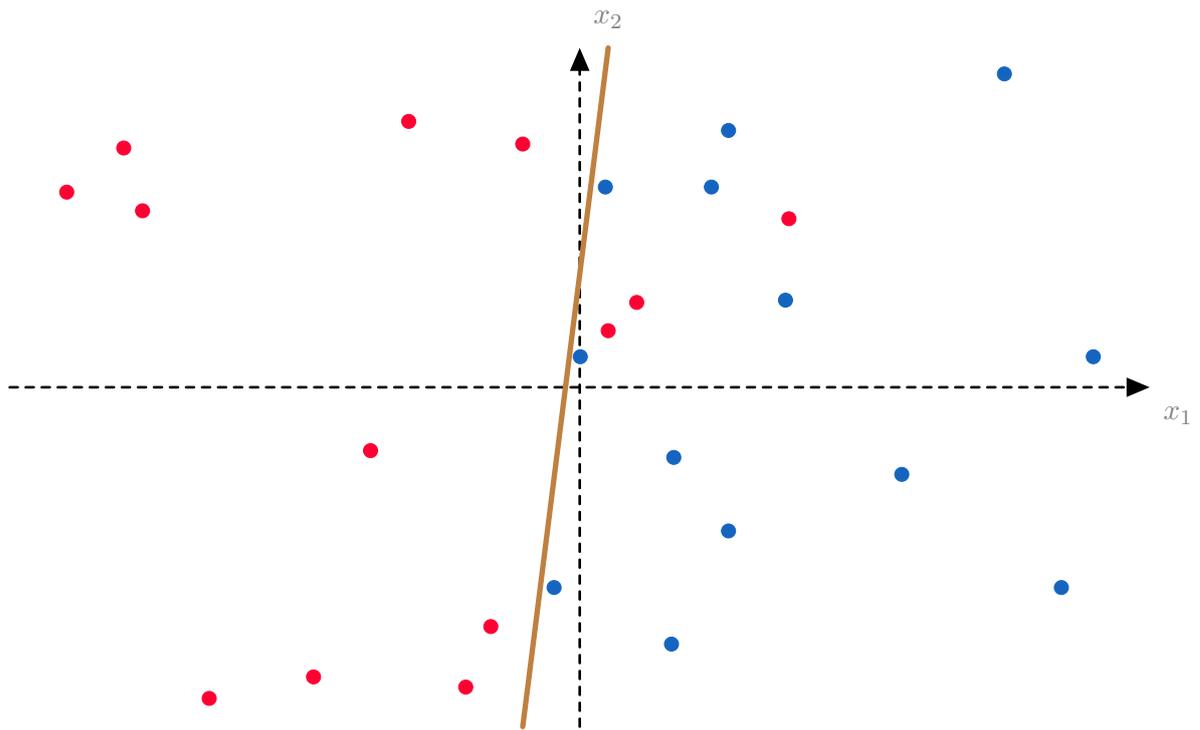


FIGURE 1 – Représentation des différents mails dans un espace à deux dimensions. Les spams sont représentés en rouge alors que les mails classiques sont représentés en bleu. La droite représentée en gris représente une potentielle droite optimale permettant de séparer les spams des mails standards.

$$y = ax + b$$

Si on reprend la représentation effectuée dans la Figure 1, notre droite se présentera sous la forme

$$ax_1 + bx_2 + c = 0,$$

ce qui est exactement la même forme que la précédente. On va noter $h_{a,b,c}(\mathbf{x}) = ax_1 + bx_2 + c$ le membre de gauche de l'équation précédente. Et il faudrait donc déterminer les bonnes valeurs des paramètres a, b et c de la droite.

Ces paramètres seront déterminés de façon à réduire le nombre d'erreurs de classification effectués par notre algorithme qui apprend une droite séparant les données en deux groupes. Pour fixer les idées, on va dire que si on se trouve à gauche de la droite $ax_1 + bx_2 + c = 0 \iff h_{a,b,c}(\mathbf{x}) = 0$, *i.e.* si $h_{a,b,c}(\mathbf{x}) < 0$ alors le mail est catégorisé comme étant un spam. Dans le cas contraire, le mail est catégorisé comme étant "désirable", *i.e.* lorsque $h_{a,b,c}(\mathbf{x}) > 0$.

A partir de cela, on peut alors proprement définir le critère que l'on souhaite minimiser, à savoir notre nombre d'erreurs. Pour cela, on se rappelle que les exemples considérés possèdent une étiquette (colorée), s'ils sont bleus, on va dire que les mail sont "positifs", *i.e.* ils sont standards, on va leur attribuer une valeur $y = 1$ Dans le cas contraire, s'ils sont rouges, ils sont dits "négatifs" donc considérés comme des spams et on va leur une valeur $y = -1$.

Ainsi on va chercher à minimiser le nombre d'erreurs qui s'écrit comme suit :

$$\frac{1}{m} \sum_{i=1}^m \mathbb{1}_{\{y_i h_{a,b,c}(\mathbf{x}_i) < 0\}}.$$

Une erreur correspond donc à l'attribution du mauvais signe à une donnée \mathbf{x} , *i.e.* au fait que la donnée se trouve du mauvais côté de la droite.

Fonctions à plusieurs variables Ce simple exemple montre que chercher à minimiser un nombre d'erreurs revient à étudier une fonction à plusieurs variables, ici a, b et c et à chercher le minimum de cette fonction.

Vous avez pu voir, ou verrez que chercher les valeurs de a, b et c pour lesquelles la fonction atteint son minimum nécessite de trouver des points critiques et d'étudier la nature de ses points critiques via la résolution de l'équation d'Euler et l'étude de

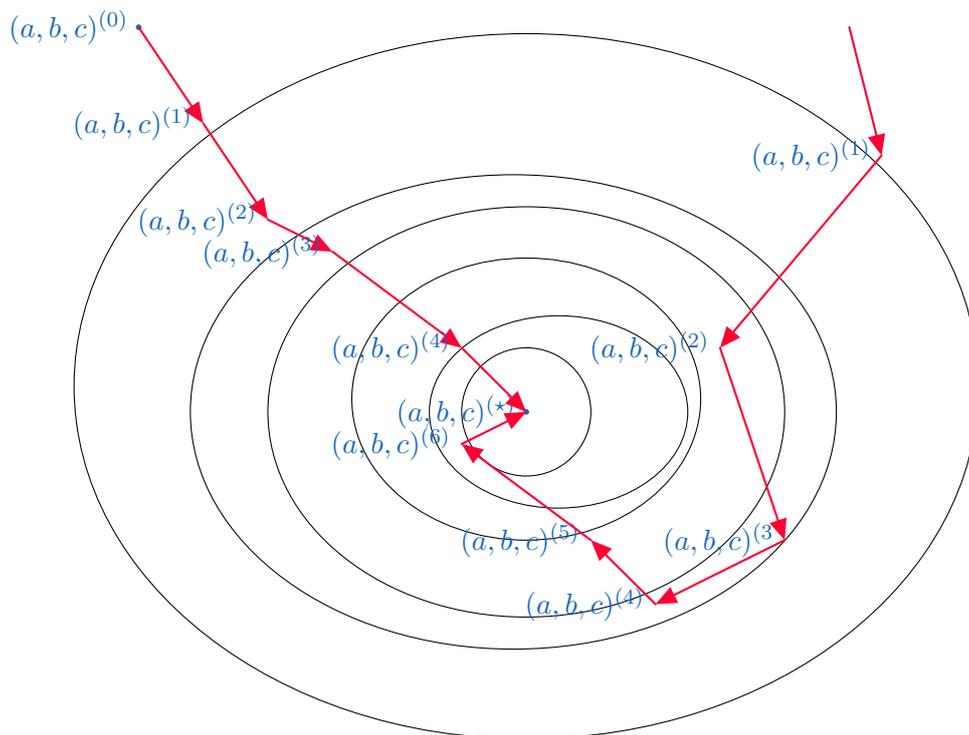


FIGURE 2 – Illustration de la descente de gradient vers la solution optimale $(a, b, c)^{(*)}$ de la fonction g que l'on souhaite minimiser en fonction des paramètres a, b et c . Les différentes ellipses représentent des lignes de niveau de la fonction g .

la dérivée seconde de la fonction, qui se présentera sous la forme d'une matrice. Plus formellement, on dit que l'on va résoudre un problème d'optimisation :

$$\min_{a,b,c} g(a, b, c) = \min_{a,b,c} \frac{1}{m} \sum_{i=1}^m \mathbb{1}_{\{y_i h_{a,b,c}(\mathbf{x}_i) < 0\}}.$$

Bien évidemment quand les points critiques sont très complexes à trouver mais que la fonction possède de bonnes propriétés, on peut passer par des algorithmes de descente de gradient, qui vont, de façon itérative, approcher la solution optimale comme cela est illustré en Figure 2.

Vous verrez que cette approche itérative s'écrit de la façon suivante

$$(a, b, c)^{(n+1)} = (a, b, c)^{(n)} - \eta \nabla g \left((a, b, c)^{(n)} \right),$$

i.e. les valeurs des paramètres a, b et c à l'itération $n + 1$ est une transformation des valeurs de ces mêmes paramètres à l'itération n qui met en jeu les données mais aussi le *gradient* de la fonction que l'on cherche à optimiser (et donc de l'équation de la

droite). Cette procédure simple, appelée descente de gradient nécessite de manipuler des *vecteurs* voir des matrices ou des inverses de matrices pour des versions plus sophistiquées.

Ce n'est bien sûr pas la seule application mais c'est une première application directe qui pourra se faire plus tard avec ces outils d'Analyse et d'Algèbre Linéaire.

Vous étudierez également, vraisemblablement l'année prochaine, comment exploiter l'algèbre linéaire pour faire de l'*Analyse de Données*.

Commençons sans plus tarder !

Deuxième partie

Analyse des fonctions à une variable réelle

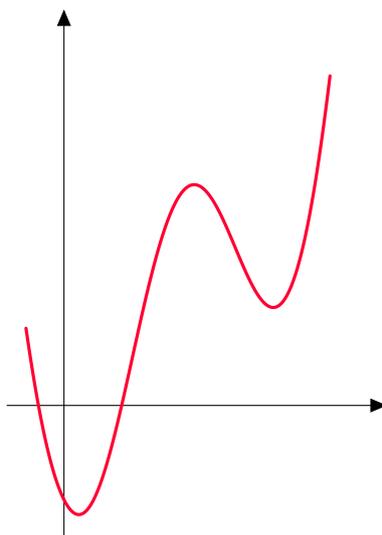
Dans cette première partie, nous nous intéresserons à l'étude des fonctions réelles d'une variable réelle ainsi que des suites numériques réelles.

On souhaite tout d'abord faire quelques rappels sur les suites numériques qui seront très souvent utilisées dans la prochaine partie de ce cours.

Les suites sont omniprésentes dans la conception des algorithmes et on les retrouve à chaque fois que l'on effectue des boucles dans son programme. En outre, combinée à l'étude des fonctions, elles permettent d'étudier les propriétés des algorithmes et notamment leur convergence ou encore leur complexité. Enfin, pour cette dernière il est également important de connaître des ordres de grandeurs, ou plus précisément des équivalents de suites ou de fonctions afin de mieux appréhender la complexité d'un algorithme : son temps de calcul ainsi que son coût en mémoire.

Ce sont autant d'éléments qui vont permettre de développer des solutions efficaces dans de multiples applications avec la prise en compte des contraintes matérielles.

Une grosse partie est ensuite consacrée à l'étude des fonctions, avec une présentation non exhaustive de fonctions usuelles en mathématiques et que le lecteur sera susceptible de rencontrer dans son futur en tant que qu'ingénieur/chercheur en Science des Données.



Nous présenterons les concepts de continuités et de dérivabilités de ces fonctions ainsi que quelques grands résultats concernant la recherche d'extrema pour des fonctions

continues voire même pour des fonctions convexes.

Nous verrons aussi comment approcher une fonction dérivable, ou plusieurs dérivables, à l'aide d'une approximation linéaire, quadratique, voire plus pour en simplifier l'étude localement autour d'un point. L'ensemble de ces outils constituent des éléments fondamentaux dans la recherche d'extrema d'une fonction sur le plan algorithmique. Ce sont ces méthodes d'approximation qui vont permettre de développer des algorithmes de recherche d'extrema de fonctions complexes et cela de façon automatique. Les propriétés des fonctions vont ensuite nous permettre de garantir la convergence de l'algorithme ou encore la nature de la solution.

Une dernière partie est consacrée à l'intégration des fonctions continues et quelques théorèmes généraux et des méthodes de calculs permettant de calculer une grande variété d'intégrale. Cela permettra de donner quelques outils qui pourront s'avérer utiles lors de la présentation des concepts comme l'**espérance** ou la **variance** d'une variable aléatoire ayant une fonction de densité continue.

1 Suites réelles

Cette section est dédiée à l'étude des suites réelles. Elle rappelle les définitions élémentaires sur les suites : variations et bornes. On abordera ensuite les notions de convergence et limite des suites. On finira enfin par quelques rappels et études des suites arithmétiques, géométriques et arithmético-géométriques.

Cette première section est primordiale car elle permet d'introduire des outils nécessaires à l'étude des fonctions. En effet, on pourra voir les suites comme des fonctions évaluées sur l'ensemble des nombres entiers \mathbb{N} .

1.1 Généralités sur les suites

Définition 1.1: Suites réelles

On appelle *suite réelle* toute famille de nombres réels indexés par \mathbb{N} . On notera $u = (u_n)_{n \in \mathbb{N}}$ une telle suite, où pour tout $n \in \mathbb{N}$, $u_n \in \mathbb{R}$ est le terme de rang n de la suite u .

Exemple 1.1. Les suites u, v, w définies pour tout entier $n \in \mathbb{N}$ par

$$u_n = n + 3, \quad v_n = 4n, \quad \text{et} \quad w_n = 2^n$$

sont des suites réelles au sens défini précédemment.

Dans l'exemple ci-dessus, les suites u et v sont des suites **arithmétiques** et la suite w est une suite **géométrique**. Nous reviendrons sur ces points là un peu plus tard dans cette section.

La définition de suite suppose que cette dernière est définie pour tout entier naturel n , mais il n'est pas rare que l'on souhaite définir une suite à partir d'un certain rang $n_0 \in \mathbb{N}$ afin que les objets soient bien définies.

Exemple 1.2. Considérons la suite u définie par

$$u_n = \sqrt{2n - 6}.$$

Pour la suite u soit bien définie, il faut que l'argument dans la racine carrée soit positif ou nul, i.e. il faut que $2n - 6 \geq 0$ soit $n \geq 3$. Ainsi la suite u est définie à partir du rang $n_0 = 3$.

On pourra également introduire des suites plus *exotiques* dont la définition pourrait varier selon la parité de n , comme cela serait le cas avec la suite u définie pour tout entier n par :

$$u_n = \begin{cases} 2n - 3 & \text{si } n \text{ est pair,} \\ n + 3 & \text{si } n \text{ est impair.} \end{cases}$$

On pourra également effectuer des opérations sur les suites, tout comme on peut effectuer des opérations sur les nombres réels.

Définition 1.2: Opérations sur les suites

Soient deux suites de nombres réels $u = (u_n)_{n \in \mathbb{N}}$ et $v = (v_n)_{n \in \mathbb{N}}$ et soit λ un nombre réel quelconque. On peut alors définir :

1. la suite $\lambda u = (\lambda u_n)_{n \in \mathbb{N}}$,
2. la suite $u + v = (u_n + v_n)_{n \in \mathbb{N}}$,
3. la suite $u \times v = (u_n \times v_n)_{n \in \mathbb{N}}$,
4. la suite $\frac{u}{v} = \left(\frac{u_n}{v_n} \right)_{n \in \mathbb{N}}$, à condition que la suite $(v_n)_{n \in \mathbb{N}}$ soit non nulle pour tout entier n

Exemple 1.3. *Considérons les suites u et v définies pour tout entier $n \in \mathbb{N}$ par*

$$u_n = n + 3 \quad \text{et} \quad v_n = 2^n$$

et λ un nombre réel quelconque. Alors les suite $u + v$ et $u \times v$ sont définies pour tout entier naturel n par

$$u_n + v_n = 2^n + n + 3 \quad \text{et} \quad u_n \times v_n = 2^n(n + 3).$$

Après avoir défini les suites, nous pouvons maintenant nous intéresser à l'évolution de leurs valeurs et plus précisément à leurs variations.

Définition 1.3: Variations

Une suite de nombres réels $(u_n)_{n \in \mathbb{N}}$ est dite :

1. croissante si pour tout $n \in \mathbb{N}$: $u_{n+1} \geq u_n$,
2. décroissante si pour tout $n \in \mathbb{N}$: $u_{n+1} \leq u_n$,
3. constante si pour tout $n \in \mathbb{N}$: $u_{n+1} = u_n$.

On pourra également reprendre cette définition pour introduire les notions de monotonie stricte en remplaçant les inégalités larges par des inégalités strictes.

A nouveau, il n'est pas rare d'étudier la monotonie d'une fonction à partir d'un certain rang n_0 , *i.e.* en ne se préoccupant pas du comportement de la suite sur ses premiers termes.

Exemple 1.4. On considère les suites u, v, w définies pour tout entier $n \in \mathbb{N}$ par

$$u_n = 2n, \quad v_n = \left(\frac{1}{2}\right)^n, \quad \text{et} \quad w_n = (-1)^n.$$

sont des suites réelles au sens défini précédemment.

Alors la suite u est strictement croissante pour tout n . En effet $u_{n+1} = 2(n+1) = 2n+2 > 2n = u_n$. De la même façon on remarque que la suite v est strictement décroissante et que la suite w n'est ni croissante, ni décroissante.

Etudier les valeurs successives des suites revêt d'un enjeu important dans l'étude de certains algorithmes afin d'étudier d'éventuelles propriétés de convergence de ces derniers. Pour étudier cette convergence, nous avons donc besoin de connaître le comportement de la suite mais aussi de savoir si cette dernière prend des valeurs bornées. En effet, la monotonie d'une suite et son caractère bornée vont nous permettre d'affirmer que cette dernière est convergente.

Nous illustrerons plus tard dans ce cours, après avoir étudié les fonctions, l'utilité des suites et de l'étude de leur convergence.

Définition 1.4: Suites minorées, majorées et bornées

Soit $u = (u_n)_{n \in \mathbb{N}}$, alors cette suite est dite :

1. *minorée* s'il existe un réel m tel que : $\forall n \in \mathbb{N}, m \leq u_n$,
2. *majorée* s'il existe un réel M tel que : $\forall n \in \mathbb{N}, M \geq u_n$,
3. *bornée* si elle est à la fois minorée et majorée.

De cette définition découle la proposition immédiate suivante :

Proposition 1.1: Suite bornée

Soit une suite réelle $u = (u_n)_{n \in \mathbb{N}}$, alors la suite u est bornée si et seulement si la suite $|u| = (|u_n|)_{n \in \mathbb{N}}$ est majorée.

Démonstration. La proposition se démontre en montrant les deux implications.

- Supposons que $|u|$ soit majorée, il existe donc un réel M tel que pour tout n , $|u_n| \leq M$. En utilisant la définition de la valeur absolue, on a pour tout entier n

$$-M \leq u_n \leq M.$$

Ainsi la suite u est donc bien minorée et majorée, elle est donc bornée.

- Supposons que la suite u soit bornée, il existe donc des réels m, M tels que pour tout n

$$m \leq u_n \leq M.$$

En posant $K = \max(|m|, |M|)$, nous avons, pour tout entier n

$$-K \leq m \leq u_n \leq M \leq K.$$

Ainsi, en utilisant la définition de la valeur absolue, nous avons $|u_n| \leq K$ pour tout n , ce qui montre bien que la suite $|u|$ est majorée. □

Nous allons maintenant définir la convergence d'une suite, les propriétés de convergence ainsi que les liens entre la monotonie, le caractère borné et la convergence d'une suite.

1.2 Convergence des suites

La notion de convergence est très importante lorsqu'il s'agira, plus tard, d'étudier des algorithmes d'optimisation.

En effet, nous serons souvent amenés à construire des algorithmes pour trouver une solution à un problème que l'on souhaite résoudre. Le problème sera résolu de façon itérative et on espère que notre algorithme convergera vers une (bonne) solution.

On va donc étudier à quelle(s) condition(s) une suite converge, ce qui permettra de donner quelques outils pour étudier, plus tard, la convergence de méthodes plus complexes.

Définition 1.5: Convergence

Soit une suite réelle $(u_n)_{n \in \mathbb{N}}$ et L un nombre réel. On dit que la suite $(u_n)_{n \in \mathbb{N}}$ tend vers L si :

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ tel que } \forall n \in \mathbb{N}, n \geq n_0 \implies |u_n - L| \leq \varepsilon.$$

On note alors $\lim_{n \rightarrow +\infty} (u_n) = L$.

Dit autrement, une suite admet pour limite $L \in \mathbb{R}$, si tous les termes de la suite, peuvent être aussi proche de L qu'on le souhaite, à partir d'un certain rang. On peut montrer que, lorsqu'elle existe, **la limite d'une suite est unique**.

Définition 1.6: Suite convergente

Soit une suite réelle $(u_n)_{n \in \mathbb{N}}$, on dit que cette suite est *convergente* s'il existe un réel L tel que :

$$\lim_{n \rightarrow +\infty} (u_n) = L.$$

Dans le cas contraire, elle est dite *divergente*.

Regardons quelques exemples pour illustrer des cas de suite convergentes et divergentes.

Exemple 1.5. On considère la suite $(u_n)_{n \in \mathbb{N}}$ définie pour tout entier naturel n par

$$u_n = 1 + \frac{1}{n+1}.$$

Notons que cette suite $(u_n)_{n \in \mathbb{N}}$ converge vers 1 en appliquant nos connaissances sur les calculs de limites. En effet $\lim_{n \rightarrow +\infty} \frac{1}{n+1} = 0$.

Si nous devons employer la définition, cela nous donnerait

$$\forall \varepsilon > 0, \exists n_0 \in \mathbb{N} \text{ tel que } \forall n \in \mathbb{N}, n \geq n_0 \implies \left| 1 + \frac{1}{n+1} - 1 \right| \leq \varepsilon.$$

Il faudrait donc montrer l'existence d'un tel n_0 qui dépendrait donc de ε .

$$\frac{1}{n+1} \leq \varepsilon \iff n \geq \frac{1}{\varepsilon} - 1.$$

Ainsi, la définition est vérifiée en prenant $n_0 = \frac{1}{\varepsilon} - 1$.

Bien évidemment, nous n'utilisons pas la définition, en pratique, pour montrer qu'une suite converge. On se contente d'appliquer ce que l'on connaît pour le calcul de limites ou, quand cela est moins visible, d'utiliser des propriétés que nous verrons plus tard.

Exemple 1.6. *Considérons les suites $(v_n)_{n \in \mathbb{N}}$ et $(w_n)_{n \in \mathbb{N}}$ suivantes*

$$v_n = (-1)^n \quad \text{et} \quad w_n = 2^n$$

Ces deux suites divergent. En effet, la suite $(v_n)_{n \in \mathbb{N}}$ alterne entre deux valeurs -1 et 1 selon la parité de n et ne peut donc pas être convergente (cela contredirait l'unicité de la limite).

La suite $(w_n)_{n \in \mathbb{N}}$ est également divergente car elle peut prendre des valeurs aussi grandes qu'on le souhaite.

Nous verrons un peu plus tard, comment définir une suite qui diverge. Mais regardons d'abord quelques propriétés des suites convergentes.

Proposition 1.2: Convergence et bornée

Toute suite convergente est bornée.

Démonstration.

□

Cette proposition, bien que très simple ne s'utilise que très rarement. Elle sert surtout à montrer qu'une suite est divergente en utilisant la **contraposée** : *une suite non bornée est divergente*, ce que l'on a pu voir directement avec la suite $(w_n)_{n \in \mathbb{N}}$ de l'Exemple 1.6.

Regardons maintenant quelques résultats sur la convergence des suites et les opérations élémentaires (somme et produit)

Proposition 1.3: Suites convergentes vers 0

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ convergeant toutes deux vers 0 et soit λ un nombre réel. Alors

1. la suite $(\lambda u_n)_{n \in \mathbb{N}}$ converge vers 0
2. la suite $(u_n + v_n)_{n \in \mathbb{N}}$ converge vers 0

Démonstration. La preuve consiste à appliquer la définition de limite dans le cas où $L = 0$ sur les deux suites étudiées

- 1.
- 2.

□

Ce n'est pas surprenant intuitivement mais regardons quelques exemples pour visualiser ces propriétés.

Exemple 1.7. On considère deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ définies pour tout n par :

$$u_n = \frac{1}{n+1} \quad \text{et} \quad v_n = \frac{2}{n+1}$$

Ces deux suites convergent bien vers 0. Soit également un nombre réel λ . Alors,

$$\lim_{n \rightarrow +\infty} \lambda u_n = \lambda \lim_{n \rightarrow +\infty} u_n = \lambda \times 0 = 0.$$

De la même façon

$$\lim_{n \rightarrow +\infty} (u_n + v_n) = \lim_{n \rightarrow +\infty} \left(\frac{1}{n+1} + \frac{2}{n+1} \right) = \lim_{n \rightarrow +\infty} \left(\frac{3}{n+1} \right) = 0.$$

Ce premier résultat permet d'avoir des propriétés plus générales suivantes sur les suites convergentes.

Proposition 1.4: Opérations sur les suites convergentes

Soient deux suites réelles convergentes $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ de limites respectives L_1 et L_2 et λ un nombre réel. Alors :

1. la suite $(\lambda u_n)_{n \in \mathbb{N}}$ converge vers λL_1 ,
2. la suite $(u_n + v_n)_{n \in \mathbb{N}}$ converge vers $L_1 + L_2$,
3. la suite $(u_n \times v_n)_{n \in \mathbb{N}}$ converge vers $L_1 L_2$
4. si $L_2 \neq 0$, la suite $\left(\frac{u_n}{v_n}\right)_{n \in \mathbb{N}}$ converge vers $\frac{L_1}{L_2}$.

A nouveau cette proposition est plutôt intuitive, elle ne devrait pas choquer le lecteur et généralise ce que l'on a vu précédemment aux suites convergentes (mais pas nécessairement vers 0).

Elle est en revanche très pratique lorsqu'il s'agit de calculer la limite de *somme*, *produit* ou encore *quotient* de deux suites.

Il se peut parfois que la limite d'une suite soit très difficile à calculer car son expression est trop complexe, il peut donc se révéler utile de trouver une suite majorante dont la limite est plus simple à déterminer. C'est d'ailleurs ce que l'on fait, dans un cadre plus général en employant le **Théorème des Gendarmes**.

On dit que la suite $(v_n)_{n \in \mathbb{N}}$ est **une suite majorante** de $(u_n)_{n \in \mathbb{N}}$, si pour tout entier n , nous avons

$$u_n \leq v_n.$$

On peut également parler de suite *majorante* à partir d'un certain rang. On peut également, de façon analogue, définir la notion de suite minorante.

Proposition 1.5: Convergence et suite majorante

Soit $(u_n)_{n \in \mathbb{N}}$ une suite réelle et supposons qu'il existe une suite $(v_n)_{n \in \mathbb{N}}$ telle que pour tout entier naturel n , $|u_n| \leq v_n$ et $\lim_{n \rightarrow +\infty} v_n = 0$, alors $\lim_{n \rightarrow +\infty} u_n = 0$.

La démonstration n'est qu'une simple écriture de la définition de limite. En revanche, noter que l'on étudie la suite des *valeurs absolues* de la suite $(u_n)_{n \in \mathbb{N}}$, i.e. on étudie une suite dont les valeurs sont *positives*. Cette hypothèse, si elle n'est pas respectée, peut mettre en défaut le résultat précédent.

En effet, considérons les suites $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ définies par

$$u_n = -2 + \frac{1}{n+1} \quad \text{et} \quad v_n = \frac{1}{n+1}.$$

On a bien $u_n \leq v_n$ pour tout entier n et $(v_n)_{n \in \mathbb{N}}$ converge vers 0, mais cela n'est pas le cas de la suite $(u_n)_{n \in \mathbb{N}}$ qui converge vers -2 .

Proposition 1.6: Théorème des Gendarmes

Soient trois suites réelles $(u_n)_{n \in \mathbb{N}}$, $(v_n)_{n \in \mathbb{N}}$ et $(w_n)_{n \in \mathbb{N}}$ telles que pour tout entier n , nous avons

$$u_n \leq v_n \leq w_n.$$

Alors si $\lim_{n \rightarrow +\infty} u_n = \lim_{n \rightarrow +\infty} w_n = L \in \mathbb{R}$, on a

$$\lim_{n \rightarrow +\infty} v_n = L$$

Le résultat de cette proposition reste inchangée si l'on a $u_n \leq v_n \leq w_n$ à partir d'un certain rang n .

Exemple 1.8. *Considérons la suite $(u_n)_{n \in \mathbb{N}}$ définie pour tout entier $n \in \mathbb{N}$ par*

$$u_n = \frac{\sin(n)}{n+1}$$

et appliquons le théorème des gendarmes pour étudier la limite de cette suite. Notons que pour tout entier n , nous avons

$$-1 \leq \sin(n) \leq 1 \iff \frac{-1}{n+1} \leq \frac{\sin(n)}{n+1} \leq \frac{1}{n+1}.$$

$$\text{Or } \lim_{n \rightarrow +\infty} \frac{1}{n+1} = \lim_{n \rightarrow +\infty} \frac{-1}{n+1} = 0.$$

On en déduit donc, à l'aide du théorème des gendarmes, que la suite $(u_n)_{n \in \mathbb{N}}$ converge vers 0.

C'est donc un résultat très intéressant à employer en pratique. On retrouvera un résultat analogue lorsque l'on reviendra sur les études de fonctions.

1.3 Divergence des suites

Toutes les suites étudiées ne sont pas forcément convergentes et nous avons pu en voir quelques exemples dans la section précédente. Dans la présente section, on cherchera simplement à définir proprement une suite divergente ainsi qu'une proposition utile en pratique

Définition 1.7: Suite divergente

Soit $(u_n)_{n \in \mathbb{N}}$ une suite de nombres réels.

- On dit que $(u_n)_{n \in \mathbb{N}}$ tend vers $+\infty$ et on note $\lim_{n \rightarrow +\infty} u_n = +\infty$, si :

$$\forall A \in \mathbb{R}, \exists n_0 \in \mathbb{N} \text{ tel que } \forall n \geq n_0 \implies u_n \geq A.$$

- On dit que $(u_n)_{n \in \mathbb{N}}$ tend vers $-\infty$ et on note $\lim_{n \rightarrow +\infty} u_n = -\infty$, si :

$$\forall A \in \mathbb{R}, \exists n_0 \in \mathbb{N} \text{ tel que } \forall n \geq n_0 \implies u_n \leq A.$$

Dans les deux cas, la suite $(u_n)_{n \in \mathbb{N}}$ est dite *divergente*.

Nous avons déjà rencontré un exemple d'une telle suite lorsque dans l'Exemple [?], notamment à travers la suite $(w_n)_{n \in \mathbb{N}}$ définie pour tout entier n par $w_n = 2^n$.

Si on reprend la définition, nous avons bien pour tout $A \in \mathbb{R}$, $2^n \geq A$ pour tout $n \geq \frac{\ln(A)}{2}$. Ainsi la définition est bien vérifiée pour tout $n \geq n_0 = \frac{\ln(A)}{2}$.

Remarque Une suite divergente ne doit pas forcément tendre vers $+\infty$, il faut simplement qu'elle ne tende pas vers une valeur fixe. Donc les suites

$$u_n = (-1)^n \quad \text{et} \quad v_n = \sin(n)$$

ne tendent pas vers $\pm\infty$ mais elles divergent pour autant.

Proposition 1.7: Divergence et Relation d'ordre

Soient $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ deux suites réelles telles que pour tout entier n , $u_n \leq v_n$ alors

1. si $\lim_{n \rightarrow +\infty} u_n = +\infty$, on a $\lim_{n \rightarrow +\infty} v_n = +\infty$,
2. si $\lim_{n \rightarrow +\infty} v_n = -\infty$, on a $\lim_{n \rightarrow +\infty} u_n = -\infty$.

Cette proposition reste valide si l'hypothèse $u_n \leq v_n$ n'est vérifiée qu'à partir d'un certain rang.

Ces premières définitions et résultats ne permettant pas forcément de trouver la limite d'une suite ou encore de prouver que cette dernière converge.

Le seul théorème qui nous permis de faire cela est celui des gendarmes, mais on ne dispose pas toujours de tels moyens pour déterminer la convergence d'une suite. En revanche, lorsque la suite possède de bonne propriété comme le fait d'être *monotone*, *i.e.* croissante ou décroissante, on peut alors montrer sa convergence.

1.4 Limite d'une suite monotone

Lorsque que l'on étudie des suites monotones, ce qui arrivent souvent lorsque l'on regarde l'évolution des solutions d'un problème au cours des itérations d'un algorithme, nous sommes capables de montrer la convergence de ces dernières. C'est ce que nous allons voir avec les deux résultats suivants.

Proposition 1.8: Existence limite suite croissante

Soit $(u_n)_{n \in \mathbb{N}}$ une suite réelle **croissante**. Alors

1. si $(u_n)_{n \in \mathbb{N}}$ est *majorée*, alors elle converge et $\lim_{n \rightarrow +\infty} u_n = \sup_n \{u_n \mid n \in \mathbb{N}\}$.
2. si $(u_n)_{n \in \mathbb{N}}$ n'est *pas majorée*, alors elle diverge et $\lim_{n \rightarrow +\infty} u_n = +\infty$.

On en déduit un résultat analogue pour les suites qui sont décroissantes

Proposition 1.9: Existence limite suite minorée

Soit $(u_n)_{n \in \mathbb{N}}$ une suite réelle **décroissante**. Alors

1. si $(u_n)_{n \in \mathbb{N}}$ est *minorée*, alors elle converge et $\lim_{n \rightarrow +\infty} u_n = \inf_n \{u_n \mid n \in \mathbb{N}\}$.
2. si $(u_n)_{n \in \mathbb{N}}$ n'est *pas minorée*, alors elle diverge et $\lim_{n \rightarrow +\infty} u_n = -\infty$.

Ainsi, pour étudier la convergence d'une suite, il suffit parfois de montrer qu'elle est croissante et majorée ou encore décroissante et minorée pour montrer qu'elle converge.

Ces deux résultats ne montrent que l'existence de la limite pour une suite, mais on ne dispose pas de moyens pour la calculer pour le moment. La plupart pas du temps, il

faudra se laisser guider par l'énoncé ou les hypothèses dont on dispose sur la suite.

On termine cette présentation sur les suites en abordant un point important qui est la notion de *relation d'ordre* et d'*équivalences* entre des suites.

1.5 Relations de comparaison

Pour calculer des limites de suite, plus précisément pour lever une indétermination, on a parfois besoin de savoir quelles sont les suites prépondérantes, négligeables ou encore équivalentes à d'autres suites, à l'aide de suites de références.

Cela est par exemple le cas lorsque l'on souhaite étudier les limites suivantes :

$$\lim_{n \rightarrow +\infty} \frac{n^2}{6n + 3}, \quad \lim_{n \rightarrow +\infty} \frac{\ln(6n^2 + 3n)}{2n - 4}, \quad \lim_{n \rightarrow +\infty} \frac{e^{-2n}}{\frac{1}{n}}.$$

Si l'intérêt est très limité dans le cadre de ce cours. Cela se révélera particulièrement intéressant de disposer de ce genre de notions, notamment lorsque l'on cherchera à évaluer la *complexité* d'un algorithme, *i.e.* la complexité d'une opération effectuée ou la complexité globale pour la résolution d'un problème.

Connaître cet ordre pour certaines opérations, permettra de sélectionner l'algorithme le plus efficace pour la résolution d'une tâche.

Ces aspects algorithmes sont particulièrement intéressants à prendre en compte dans un cadre industriel avec des contraintes de temps lors de l'apprentissage de modèles en Machine Learning.

Mais commençons par définir ces relations de comparaisons.

1.5.1 Suite négligeable

Définition 1.8: Suite négligeable

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$. On dit que $(u_n)_{n \in \mathbb{N}}$ est négligeable devant $(v_n)_{n \in \mathbb{N}}$ s'il existe une suite $(w_n)_{n \in \mathbb{N}}$ qui tend vers 0 et telle que $u_n = v_n \times w_n$ à partir d'un certain rang.

On note alors $u_n = o(v_n)$.

Regardons tout de suite une proposition plus concrète pour caractériser une suite négligeable devant une autre.

Proposition 1.10: Caractérisation négligeable

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ qui est non nulle à partir d'un certain rang. On dit que $(u_n)_{n \in \mathbb{N}}$ est négligeable devant $(v_n)_{n \in \mathbb{N}}$ si et seulement si

$$\lim_{n \rightarrow +\infty} \left(\frac{u_n}{v_n} \right) = 0.$$

Exemple 1.9. La suite $(u_n)_{n \in \mathbb{N}}$ définie pour tout entier n par $u_n = n + 2$ est négligeable devant la suite $(v_n)_{n \in \mathbb{N}}$ définie pour tout entier n par $v_n = 2^n$.

En effet, la limite de la suite $\left(\frac{u_n}{v_n} \right)_{n \in \mathbb{N}} = \frac{n+2}{2^n}$ est nulle lorsque n tend vers $+\infty$.

Ainsi, pour savoir si une suite est négligeable devant une autre, il suffit simplement d'étudier le quotient de ces deux suites. On pourra, par la suite, utiliser les comparaisons de références ci-dessous dans un cadre pratique

Comparaison des suites de référence Si on considère des nombres réels α, β et γ strictement positifs, alors :

$$\ln(n)^\alpha = o(n^\beta), \quad n^\beta = o(e^{\gamma n}) \quad \text{et} \quad e^{\gamma n} = o(n!).$$

Ces relations sont normalement déjà connues, mais on va les montrer afin de pouvoir les utiliser par la suite. Dans les trois cas, on va s'intéresser au quotient entre les deux suites

• $\ln(n)^\alpha = o(n^\beta)$: ...

• $n^\beta = o(e^{\gamma n})$: Il suffit simplement de réécrire le quotient comme suit

$$\frac{n^\beta}{e^{\gamma n}} = e^{\beta \ln(n) - \gamma n}.$$

Or la suite $u_n = n$ croît plus rapidement que la suite $v_n = \ln(n)$, donc $\lim_{n \rightarrow +\infty} (\beta \ln(n) - \gamma n) = -\infty$. On en déduit donc que $\lim_{n \rightarrow +\infty} \frac{n^\beta}{e^{\gamma n}} = 0$.

• $e^{\gamma n} = o(n!)$: Soit n_0 un entier tel que $n_0 \geq 2e^\gamma$, alors pour tout entier naturel n tel que $n \geq n_0$:

$$\left| \frac{e^{\gamma n}}{n!} \right| = \frac{e^{\gamma n}}{n!} = \frac{e^{\gamma n_0}}{n_0!} \times \frac{e^{\gamma}}{n} \times \frac{e^{\gamma}}{n-1} \times \dots \times \frac{e^{\gamma}}{n_0+1} \leq \left(\frac{1}{2} \right)^{n-n_0} \frac{e^{\gamma n_0}}{n_0!}.$$

Or $\lim_{n \rightarrow +\infty} \left(\frac{1}{2} \right)^{n-n_0} = 0$, donc $\lim_{n \rightarrow +\infty} \frac{e^{\gamma n}}{n!} = 0$.

1.5.2 Suite dominée

Définition 1.9: Suite dominée

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$. On dit que $(u_n)_{n \in \mathbb{N}}$ est dominée par $(v_n)_{n \in \mathbb{N}}$ s'il existe une suite $(w_n)_{n \in \mathbb{N}}$ bornée telle que $u_n = v_n \times w_n$ à partir d'un certain rang.

On note alors $u_n = \mathcal{O}(v_n)$.

Regardons tout de suite une proposition plus concrète pour caractériser une suite négligeable devant une autre.

Proposition 1.11: Caractérisation domination

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ qui est non nulle à partir d'un certain rang. On dit que $(u_n)_{n \in \mathbb{N}}$ est dominée par $(v_n)_{n \in \mathbb{N}}$ si et seulement si le quotient $\left(\frac{u_n}{v_n} \right)$ est borné à partir d'un certain rang.

Ainsi, pour savoir si une suite est dominée devant une autre, il suffit simplement d'étudier le quotient de ces deux suites et de montrer que ce quotient reste borné.

Exemple 1.10. La suite $(u_n)_{n \in \mathbb{N}}$ définie pour tout entier n par $u_n = 6n + 2$ est dominée par la suite $(v_n)_{n \in \mathbb{N}}$ définie pour tout entier n par $v_n = 7n + 4$.

En effet, la limite de la suite quotient $\left(\frac{u_n}{v_n} \right)_{n \in \mathbb{N}}$ est égale $\frac{6}{7}$, elle est donc bien bornée à partir d'un certain rang.

1.5.3 Suites équivalentes

Il s'agit d'un dernier point à étudier et qui sera aussi très utile au moment de l'analyse de fonctions. Cette notion d'équivalence est importante lorsque l'on cherche à faire des approximations asymptotiques des suites ou même des fonctions.

Définition 1.10: Suites équivalentes

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$. On dit que les deux suites sont équivalentes s'il existe une suite $(w_n)_{n \in \mathbb{N}}$ tendant vers 1 telle que $u_n = v_n \times w_n$ à partir d'un certain rang.

On note alors $u_n \sim v_n$.

Noter que la relation d'équivalence entre deux suites est une relation symétrique. Ainsi si la suite $(u_n)_{n \in \mathbb{N}}$ est équivalente à la suite $(v_n)_{n \in \mathbb{N}}$, on pourra également dire que la suite $(v_n)_{n \in \mathbb{N}}$ est équivalente à la suite $(u_n)_{n \in \mathbb{N}}$ (ce qui n'était pas le cas avec les notions *dominées* et *négligeables*).

Proposition 1.12: Caractérisation équivalence

Soient deux suites réelles $(u_n)_{n \in \mathbb{N}}$ et $(v_n)_{n \in \mathbb{N}}$ qui est non nulle à partir d'un certain rang. On dit que $(u_n)_{n \in \mathbb{N}}$ est équivalente à $(v_n)_{n \in \mathbb{N}}$ si et seulement si $\lim_{n \rightarrow +\infty} \frac{u_n}{v_n} = 1$.

Plus que les définitions en soit, ce sont des surtout les exemples des équivalents de références qu'il sera bon de mémoriser pour l'usage pratique.

Exemple 1.11. On considère une suite réelle $(u_n)_{n \in \mathbb{N}}$ telle que sa limite lorsque n tend vers $+\infty$ est égale à 0. On a alors les équivalents suivants :

1. $\sin(u_n) \sim u_n$,
2. $\cos(u_n) - 1 \sim \frac{u_n^2}{2}$,
3. $\text{sh}(u_n) \sim u_n$,
4. $\text{ch}(u_n) - 1 \sim \frac{u_n^2}{2}$,
5. $\ln(1 + u_n) \sim u_n$,
6. $e^{u_n} - 1 \sim u_n$,
7. $(1 + u_n)^\alpha - 1 \sim \alpha u_n$.

Pour toutes les démonstrations, on se rappelle que la suite $(u_n)_{n \in \mathbb{N}}$ est une suite qui converge vers 0.

- **Pour** $\sin(u_n) \sim u_n$:

On utilisera la définition du nombre dérivé en une valeur. Plus précisément, on utilisera le fait que f est une fonction dérivable en a , on alors

$$\lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = f'(a).$$

En appliquant cette définition à la fonction $f = \sin$ et en prenant $a = 0$ et $x = u_n$, nous avons alors

$$\lim_{n \rightarrow +\infty} \frac{\sin(u_n) - \sin(0)}{u_n - 0} = \lim_{n \rightarrow +\infty} \frac{\sin(u_n)}{u_n} = \sin'(0) = \cos(0) = 1.$$

- **Pour** $\cos(u_n) - 1 \sim -\frac{u_n^2}{2}$:

On pourrait être tenté de refaire la même procédure que précédemment vu que $\cos(0) = 1$, en écrivant

$$\lim_{n \rightarrow +\infty} \frac{\cos(u_n) - \cos(0)}{u_n - 0} = \lim_{n \rightarrow +\infty} \frac{\cos(u_n) - 1}{u_n} = \cos'(0) = -\sin(0) = 0.$$

Mais on remarque cette limite tend vers 0 et nous n'avons pas montré une quelconque équivalence entre les deux suites. Il va donc falloir travailler un peu et notamment ... utiliser quelques relations trigonométriques.

Pour cela, on se rappelle que pour tout x nous avons

$$\begin{aligned} \cos(2x) &= \cos(x)^2 - \sin(x)^2, \\ &\quad \downarrow \cos(x)^2 + \sin(x)^2 = 1 \text{ soit } \cos(x)^2 = 1 - \sin(x)^2 \\ &= 1 - 2\sin(x)^2, \\ \cos(2x) - 1 &= -2\sin(x)^2. \end{aligned}$$

En prenant $x = \frac{u_n}{2}$ et en utilisant l'équivalence démontrée précédemment, nous avons

$$\begin{aligned} \cos(u_n) - 1 &= -2\sin\left(\frac{u_n}{2}\right)^2, \\ &\quad \downarrow \sin(u_n) \sim u_n \\ &\sim -2\frac{u_n^2}{4}, \\ &\sim -\frac{u_n^2}{2}. \end{aligned}$$

- Le principe est exactement le même pour les fonctions sh et ch. En revanche, nous ne savez pas comment elles sont définies actuellement donc on s'en occupe pas (on les manipulera peu dans le cadre de ce cours).
- **Pour** $\ln(1 + u_n) \sim u_n$:

On se sert à nouveau de la définition du nombre dérivé dans ce cas pour trouver notre équivalent. On veut montrer que $\lim_{n \rightarrow +\infty} \frac{\ln(1 + u_n)}{u_n} = 1$, or

$$\lim_{n \rightarrow +\infty} \frac{\ln(1 + u_n)}{u_n} = \lim_{n \rightarrow +\infty} \frac{\ln(1 + u_n) - \ln(1)}{(1 + u_n) - 1} = \ln'(1) = 1.$$

- **Pour** $e^{u_n} - 1 \sim u_n$:
C'est exactement pareil dans ce cas

$$\lim_{n \rightarrow +\infty} \frac{e^{u_n} - 1}{u_n} = \lim_{n \rightarrow +\infty} \frac{e^{u_n} - e^0}{u_n - 0} = e^0 = 1.$$

- **Pour** $(1 + u_n)^\alpha - 1 \sim \alpha u_n$:
A nouveau, on va employer notre définition du nombre dérivé et on se rappelle que la dérivée de $(1+x)^{\text{alpha}}$ est donnée par $\alpha(1+x)^{\alpha-1}$ dont la valeur est égale à α lorsque $x = 0$.
D'où

$$\lim_{n \rightarrow +\infty} \frac{(1 + u_n)^\alpha - 1}{u_n} = \lim_{n \rightarrow +\infty} \frac{(1 + u_n)^\alpha - 1^\alpha}{(1 + u_n) - 1} = \alpha.$$

Quelques remarques importantes pour finir sur ces notions d'équivalents.

- Il est possible de faire le produits de deux équivalents, *i.e.* si $a_n \sim b_n$ et $u_n \sim v_n$ alors $a_n u_n \sim b_n v_n$
- Cela reste également valable si on souhaite faire le quotient de deux équivalents.
- On peut également prendre la puissance de deux équivalents (c'est d'ailleurs ce que nous utilisé à un moment), *i.e.* si $u_n \sim v_n$, alors pour tout $\alpha \in \mathbb{R}$, on a aussi $u_n^\alpha \sim v_n^\alpha$

En revanche de façon générale, la notion d'équivalents n'est pas compatible avec la somme ou la différence.

Exemple 1.12. *Considérons le suites $(u_n)_{n \in \mathbb{N}^*}$ et $(v_n)_{n \in \mathbb{N}^*}$ définies par*

$$u_n = \frac{1}{n^3} + \frac{1}{n} \quad \text{et} \quad v_n = \frac{1}{n^2} - \frac{1}{n}.$$

Alors $u_n \sim \frac{1}{n}$ et $v_n \sim \frac{-1}{n}$ lorsque n tend vers $+\infty$. Si on somme les deux équivalents, nous trouverions un équivalent à 0, ce qui est absurde ! En réalité, nous avons $u_n + v_n = \frac{1}{n^3} + \frac{1}{n^2} \sim \frac{1}{n^2}$ lorsque n tend vers $+\infty$.

2 Fonctions usuelles

2.1 Logarithme

On commence par présenter le logarithme népérien et on présentera rapidement la définition du logarithme en base quelconque, qui présentera un intérêt dans l'étude ultérieure de certains algorithmes.

Logarithme Népérien

Définition 2.1: Logarithme népérien

La fonction *logarithme népérien*, notée \ln , est l'unique primitive sur $]0, +\infty[$ de la fonction $x \mapsto \frac{1}{x}$ qui s'annule en 1.

Par définition, pour tout $x \in]0, +\infty[$ nous avons

$$\ln(x) = \int_1^x \frac{1}{t} dt.$$

De cette même définition, on peut déduire que la fonction $x \mapsto \ln(x)$ est continue et dérivable sur $]0, +\infty[$ et que pour tout x dans ce même intervalle, nous avons

$$\ln'(x) = \frac{1}{x}.$$

On se rappelle de la définition de fonction $x \mapsto \frac{1}{x}$, cette fonction est **strictement** positive sur $]0, +\infty[$. Ceci implique la fonction \ln est **strictement** croissante sur ce même intervalle.

Avant de poursuivre sur l'étude de la fonction en elle-même, regardons quelques unes de ses propriétés, notamment vis-à-vis du produit et du quotient. Ces propriétés sont résumées par la proposition suivante :

Proposition 2.1: Propriétés logarithme népérien

Pour tout $(x, y) \in \mathbb{R}^2$ et pour tout $n \in \mathbb{Z}$, nous avons

- i) $\ln(xy) = \ln(x) + \ln(y)$,
- ii) $\ln\left(\frac{1}{x}\right) = -\ln(x)$,
- iii) $\ln\left(\frac{x}{y}\right) = \ln(x) - \ln(y)$,
- iv) $\ln(x^n) = n \ln(x)$.

Démonstration. On va uniquement démontrer le premier point de cette proposition. En effet, les trois autres points n'en sont que des conséquences.

Pour tout $y \in]0, +\infty[$ on considère la fonction $f : x \mapsto \ln(xy)$ pour tout $x \in]0, +\infty[$. La fonction f est dérivable sur l'intervalle $]0, +\infty[$ comme composée de deux fonctions dérivables sur ce même intervalle. Ainsi, pour tout $x \in]0, +\infty[$, nous avons

$$f'(x) = \frac{y}{yx} = \frac{1}{x}.$$

Ainsi, $f'(x) = \ln'(x)$ pour tout réel $x \in]0, +\infty[$. Comme les fonctions f et \ln ont la même dérivée sur cet intervalle, elles sont donc **égales à une constante près sur ce même intervalle**. On en déduit

$$\forall x \in]0, +\infty[, f(x) = \ln(x) + c,$$

où c est une certaine constante. Cependant, pour $x = 1$, nous avons $f(1) = \ln(y) = \ln(1) + c = c$.

On en déduit que $c = \ln(y)$, donc

$$\forall x \in]0, +\infty[, f(x) = \ln(xy) = \ln(x) + \ln(y).$$

□

Nous avons vu que la fonction \ln était strictement croissante entre $]0, +\infty[$, il nous reste donc pas à voir quelles sont les limites de cette fonction en les bornes de cet intervalle. Ces limites sont données par le résultat suivant :

Proposition 2.2: Limites

La fonction \ln admet des limites en 0 et $+\infty$ qui sont :

$$\lim_{x \rightarrow 0^+} \ln(x) = -\infty \quad \text{et} \quad \lim_{x \rightarrow +\infty} \ln(x) = +\infty.$$

Démonstration. Pour montrer ces deux résultats, on va utiliser la croissance de la fonction \ln ainsi que la quatrième propriété de cette fonction énoncée dans la Proposition 2.1.

La croissance stricte de la fonction \ln implique $\ln(2) > \ln(1) = 0$. Ainsi, en utilisant la quatrième partie, on a

$$\lim_{n \rightarrow +\infty} \ln(2^n) = \lim_{n \rightarrow +\infty} n \ln(2) = +\infty.$$

La fonction \ln est une fonction croissante qui n'est donc pas majorée, on a donc

$$\lim_{x \rightarrow +\infty} \ln(x) = +\infty.$$

De plus,

$$\lim_{x \rightarrow 0^+} \ln(x) = \lim_{x \rightarrow 0^+} -\ln\left(\frac{1}{x}\right) = \lim_{t \rightarrow +\infty} -\ln(t) = -\lim_{t \rightarrow +\infty} \ln(t) = -\infty.$$

□

Ce dernier résultat nous permet d'affirmer que la fonction \ln réalise une bijection croissante et continue de $]0, +\infty[$ dans \mathbb{R} . On pourra également dresser le tableau de variation de cette fonction \ln et en donner une représentation graphique (voir Figure 3).

x	0	1	$+\infty$
$\frac{1}{x}$		+	0
\ln		$-\infty$	$+\infty$

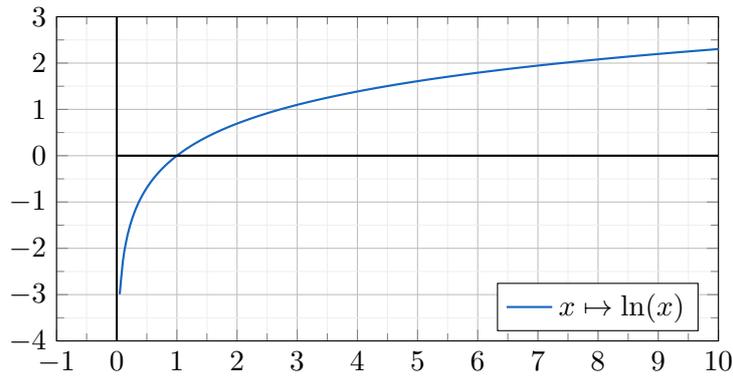


FIGURE 3 – Représentation de la fonction logarithme népérien sur \mathbb{R}^{+*}

x	0	1	$+\infty$
$\frac{1}{x}$		+	0
ln		0	$+\infty$

$-\infty$ \nearrow 0 \nearrow $+\infty$

Logarithme en base quelconque

Définition 2.2: Logarithme en base a

Soit a un nombre réel strictement positif et différent de 1. On appelle *logarithme de base a* , que l'on note \log_a , la fonction définie sur $]0, +\infty[$ par :

$$\log_a(x) = \frac{\ln(x)}{\ln(a)}.$$

Exemple 2.1. *Mettre quelques exemples d'utilisation du logarithme en base quelconques*

2.2 Fonction exponentielle

Définition 2.3: Exponentielle

La fonction exponentielle, notée \exp , est la fonction réciproque de la fonction \ln . Elle est donc bijective et strictement croissante de \mathbb{R} dans $]0, +\infty[$.

Remarque Il est également possible de définir la fonction exponentielle comme la seule solution de l'équation différentielle avec la condition suivante

$$y'(x) = y(x), \quad y(0) = 1.$$

Mais nous n'aborderons pas le thème des équations différentielles dans le cadre de ce cours. On gardera cependant en tête que la dérivée de la fonction exponentielle est elle-même, *i.e.*

$$\forall x \in \mathbb{R}, \quad \exp'(x) = \exp(x)$$

Ecrire la preuve ou pas ?

Mais qu'est-ce qu'une fonction bijective ? Commençons par donner une définition formelle avant de l'expliquer plus en détail et d'en trouver quelques caractéristiques simple.

Définition 2.4: Applications injectives, surjectives et bijectives

Soient I et J deux intervalles de \mathbb{R} et considérons une fonction $f : I \rightarrow J$. Alors la fonction f est dite

- **injective** : si pour tout $y \in J$ il existe au plus un $x \in I$ tel que $f(x) = y$, *i.e.* f est injective si tout point de l'espace d'arrivée admet **au plus** un antécédent par f .
- **surjective** : si pour tout $y \in J$ il existe au moins un $x \in I$ tel que $f(x) = y$, *i.e.* f est surjective si tout point de l'espace d'arrivée admet **au moins** un antécédent par f .
- **bijective** : si elle est à la fois injective et surjective, *i.e.* si pour tout $y \in J$ il existe un seul et unique $x \in I$ tel que $f(x) = y$, donc tout point de J admet **un seul et unique** antécédent par f .

Soient $x_1, x_2 \in I$ alors f est injective si et seulement si on a l'implication suivante

$$f(x_1) = f(x_2) \implies x_1 = x_2.$$

Pour caractériser les fonctions bijectives, on pourra se contenter de montrer que cette dernière est strictement monotone (*i.e.* strictement croissante ou strictement décroissante) de I dans J .

Revenons à notre fonction exponentielle, cette dernière est donc bijective de \mathbb{R} dans \mathbb{R}_+^* et sa fonction réciproque n'est rien d'autre que la fonction \ln . Nous avons donc les relations suivantes :

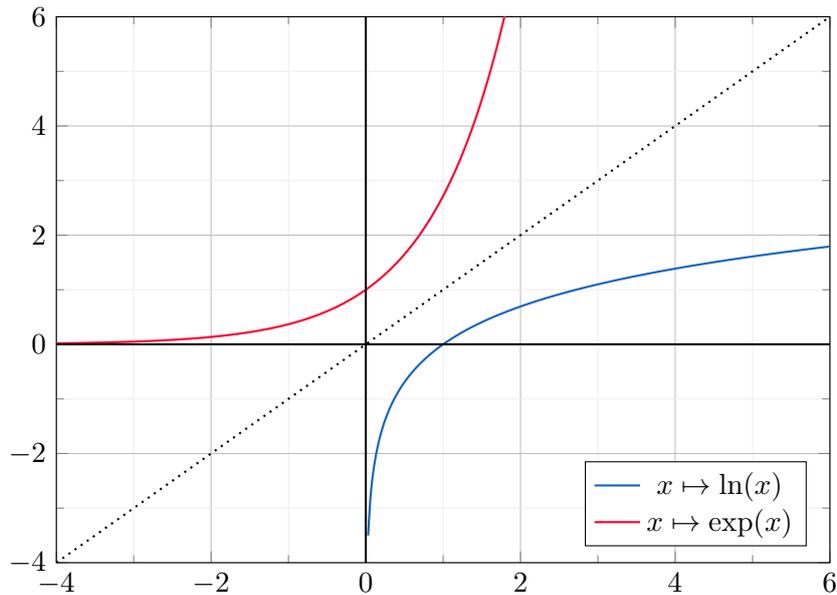


FIGURE 4 – Représentation de la fonction logarithme népérien sur $\mathbb{R}^{+\ast}$. Le graphique met également le caractère réciproque des fonctions logarithme et exponentielle via la symétrie par rapport à la droite d'équation $y = x$.

$$\forall x \in \mathbb{R}, \ln(\exp(x)) = x \quad \text{et} \quad \forall x \in]0, +\infty[, \exp(\ln(x)) = x.$$

On peut alors dresser le tableau suivant et représenter la fonction exponentielle ainsi que sa fonction réciproque en Figure 5. En tant que fonction réciproque l'une de l'autre, il est important de garder à l'esprit que ces deux fonctions sont donc symétriques par rapport à la droite d'équation $y = x$.

x	$-\infty$	0	$+\infty$
$\exp'(x)$		$+$	
\exp	0	1	$+\infty$

Les limites de la fonction exponentielle se déduisent directement des limites du logarithme népérien par symétrie avec la droite d'équation $y = x$. Tout comme nous l'avons fait pour le logarithme, nous pouvons également donner quelques propriétés de la fonction exponentielle.

Proposition 2.3: Propriétés de l'exponentielle

Pour tous réels $x, y \in \mathbb{R}$ et pour tout $n \in \mathbb{Z}$, la fonction \exp vérifie les propriétés suivantes

- i) $\exp(x + y) = \exp(x) \exp(y)$,
- ii) $\exp(-x) = \frac{1}{\exp(x)}$,
- iii) $\exp(x - y) = \frac{\exp(x)}{\exp(y)}$,
- iv) $\exp(nx) = (\exp(x))^n$.

Démonstration. Pour démontrer les différents points, on utilisera le fait que l'exponentielle se définit comme la fonction réciproque du logarithme népérien.

i) Pour tous $x, y \in \mathbb{R}$, nous avons

$$\ln(\exp(x) \exp(y)) = \ln(\exp(x)) + \ln(\exp(y)) = x + y = \ln(\exp(x + y)).$$

En utilisant la bijectivité de \ln , on en déduit la relation : $\exp(x+y) = \exp(x) \exp(y)$.

Les points ii), iii) et iv) sont des conséquences immédiates du point i). En effet

$$\exp(x) \exp(-x) = \exp(x - x) = \exp(0) = 1$$

et on déduit directement la relation ii).

Pour obtenir la relation iii), il suffit d'écrire

$$\exp(x - y) = \exp(x + (-y)) = \exp(x) \exp(-y) = \frac{\exp(x)}{\exp(y)}.$$

Enfin, pour la relation iv), on va à nouveau utiliser la fonction \ln . Ce qui nous donne, pour tout $x \in \mathbb{R}$ et pour tout $n \in \mathbb{Z}$

$$\ln(\exp(x)^n) = n \ln(\exp(x)) = nx = \ln(\exp(nx)).$$

□

2.3 Fonctions puissances

Maintenant que l'on a défini les fonctions \exp et \ln , nous sommes capables de définir en toute généralité les fonctions *puissances*.

Jusqu'à présent, les fonctions puissances $f(x) = x^n$, ont été définies pour tout entier $n \in \mathbb{N}$ voir pour des entiers $n \in \mathbb{Z}$.

Or, nous avons vu que pour tout entier n , $x^n = e^{n \ln(x)}$ qui est donc définie pour tout $x > 0$.

On pourra naturellement étendre cette définition au cas où n est un nombre réel, mais uniquement sur $]0, +\infty[$.

Définition 2.5: Fonction puissance α

Soit α un nombre réel. On définit sur $]0, +\infty[$ la fonction *puissance* α par :

$$\forall x \in]0, +\infty[, x^\alpha = e^{\alpha \ln(x)}.$$

Proposition 2.4: Propriétés de la fonction puissance

Soient α et β deux nombres réels, alors pour tout $x, y > 0$, nous avons

- i) $1^\alpha = 1$,
- ii) $x^\alpha y^\alpha = (xy)^\alpha$,
- iii) $\ln(x^\alpha) = \alpha \ln(x)$,
- iv) $x^\alpha x^\beta = x^{\alpha+\beta}$,
- v) $(x^\alpha)^\beta = x^{\alpha\beta}$.

Démonstration. kl,kl

□

On se propose maintenant d'étudier les variations de cette fonction puissance selon le signe α . Pour cela considérons, pour tout réel α , la fonction $f_\alpha(x)$ définie pour tout réel x strictement positif. On a donc

$$f_\alpha(x) = x^\alpha = e^{\alpha \ln(x)}.$$

On en déduit que la fonction f_α est dérivable pour tout $x > 0$, comme composée de deux fonctions dérivables (nous verrons ce point là ultérieurement donc nous l'admettons pour le moment). La dérivée de cette fonction est donnée par :

$$f'_\alpha(x) = \frac{\alpha}{x} e^{\alpha \ln(x)} = \frac{\alpha}{x} x^\alpha = \alpha x^{\alpha-1}.$$

Etudions le signe de cette dérivée selon le signe de α . nous avons donc trois cas à distinguer.

- **Cas où $\alpha = 0$:**

Dans ce cas, pour tout réel x , nous avons $f_0(x) = 1$ et la fonction puissance est donc la fonction constante égale à 1.

La dérivée est également constante et égale à 0 pour tout réel x .

- **Cas où $\alpha > 0$:**

Dans ce deuxième cas, la dérivée $f'_\alpha(x) > 0$ pour tout réel $x > 0$, et la fonction f_α est donc strictement croissante. De plus

$$\begin{cases} \lim_{x \rightarrow 0} \alpha \ln(x) = -\infty, \\ \lim_{x \rightarrow +\infty} \alpha \ln(x) = +\infty, \end{cases} \quad \text{donc} \quad \begin{cases} \lim_{x \rightarrow 0} f_\alpha(x) = 0, \\ \lim_{x \rightarrow +\infty} f_\alpha(x) = +\infty. \end{cases}$$

Remarquons que l'on peut même prolonger la fonction par continuité en 0 en lui attribuant la valeur 0. On en déduit alors directement le tableau des variations de cette fonction.

- **Cas où $\alpha < 0$:**

Dans ce troisième cas, la dérivée $f'_\alpha(x) < 0$ pour tout réel $x > 0$, et la fonction f_α est donc strictement décroissante. De plus

$$\begin{cases} \lim_{x \rightarrow 0} \alpha \ln(x) = +\infty, \\ \lim_{x \rightarrow +\infty} \alpha \ln(x) = -\infty, \end{cases} \quad \text{donc} \quad \begin{cases} \lim_{x \rightarrow 0} f_\alpha(x) = +\infty, \\ \lim_{x \rightarrow +\infty} f_\alpha(x) = 0. \end{cases}$$

Nous représentons les différentes fonctions pour différentes valeurs α en Figure 5. Notons également, dans le cas $\alpha > 0$, la position relative de $x \mapsto x^\alpha$ selon que $x > 1$ ou $x < 1$.

Considérons deux réels α_1 et α_2 tels que $\alpha_1 < 1 < \alpha_2$, alors, pour tout $x > 1$

$$x^{\alpha_1} \leq x \leq x^{\alpha_2}.$$

De la même façon, si l'on considère deux réels α_1 et α_2 tels que $\alpha_1 < 1 < \alpha_2$, nous avons, pour tout $x < 1$

$$x^{\alpha_2} \leq x \leq x^{\alpha_1}.$$

Nous pourrions effectuer une étude similaire dans le cas où l'exposant α est négatif.

Nous poursuivons notre étude de fonctions avec les fonctions dites *circulaires*. Ce sont des fonctions moins courantes mais que l'on peut retrouver fréquemment, notamment lorsque l'on sera amené à travailler sur des données périodiques ou sinusoïdales ou plus généralement dans le domaine de la *théorie du signal*.

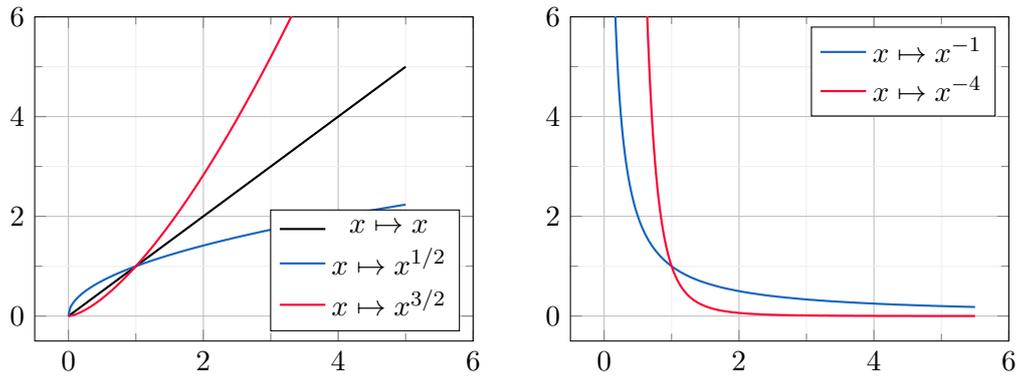


FIGURE 5 – Représentation de la fonction puissance α pour différentes valeurs de α .

2.4 Fonctions circulaires

Les fonctions circulaires sont des fonctions déjà rencontrées dans des contextes géométriques et notamment lors de mesures d'angles dans un triangle à partir des côtés *adjacent*, *opposé* ou encore de l'*hypoténuse*. On se rappelle que dans un triangle rectangle, nous pouvons obtenir ces différentes mesures d'un angle α par les relations :

$$\sin(\alpha) = \frac{\text{opposé}}{\text{hypoténuse}}, \quad \cos(\alpha) = \frac{\text{adjacent}}{\text{hypoténuse}} \quad \text{et} \quad \tan(\alpha) = \frac{\text{opposé}}{\text{adjacent}}.$$

De ces définitions vous avez également pu étudier, de ces relations découlent le théorème de Pythagore ou une version généralisée appelée le théorème d'Al-Kashi.

Vous vous souviendrez également de quelques valeurs particulières de ces fonctions en fonction de l'angle étudié, ces différentes valeurs sont rappelées en Figure 6.

Nous mettons ici de côté l'aspect géométrique pour nous focaliser sur l'aspect analytique. On rappelle que les fonctions *sinus* et *cosinus* sont définies sur \mathbb{R} et sont 2π -périodiques, ce qui signifie que pour tout réel x

$$\sin(2\pi + x) = \sin(x) \quad \text{et} \quad \cos(2\pi + x) = \cos(x).$$

La fonction cosinus est une fonction *paire* et la fonction sinus est fonction *impaire* et vérifient donc pour tout réel x

$$\sin(-x) = -\sin(x) \quad \text{et} \quad \cos(x) = \cos(-x).$$

Ces deux propriétés peuvent facilement se voir sur les graphiques de ces fonctions présentées en Figure 7. Elles sont également dérivables pour tout $x \in \mathbb{R}$ et

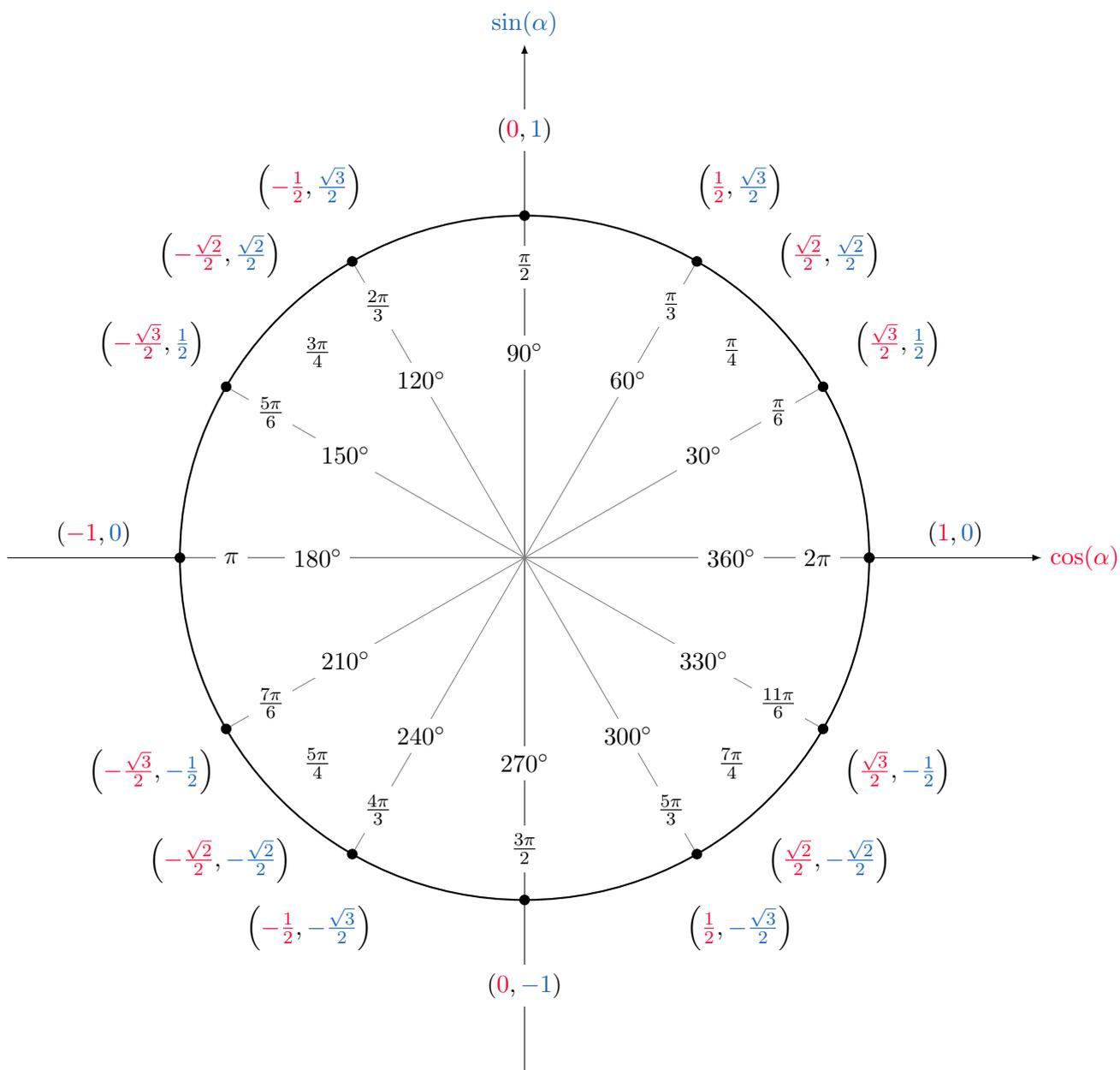


FIGURE 6 – Cercle trigonométrique. On lit abscisse le **cosinus** de l'angle α (première coordonnée) et en ordonnée le **sinus** de ce même angle α (deuxième coordonnée)

$$\sin'(x) = -\cos(x) \quad \text{et} \quad \cos'(x) = \sin(x).$$

A partir de ces relations là, on peut dresser les tableaux de variations de ces deux fonctions circulaires

Si ces formules de dérivées paraissent pour le moment sortir de nul part, elles sont en fait issues d'une autre définition de ces fonctions circulaires en utilisant l'*exponentielle complexe*.

Fonction directe Notre objectif n'est pas de rentrer dans le détail de l'exponentielle complexe mais simplement d'utiliser les liens existants entre ces deux fonctions.

Définition 2.6: Fonctions circulaire

Les fonctions *sinus* et *cosinus* peuvent se définir à l'aide de l'exponentielle complexe, *i.e.* pour tout $x \in \mathbb{R}$ nous avons

$$\cos(x) = \frac{\exp(ix) + \exp(-ix)}{2} \quad \text{et} \quad \sin(x) = \frac{\exp(ix) - \exp(-ix)}{2i},$$

où i est le nombre vérifiant $i^2 = -1$.

S'il ne s'agit pas là de la "vraie" définition de ces deux fonctions circulaires, elles seront amplement suffisantes ici et suffiront à démontrer ce que sont les dérivées des fonctions sin et cos.

En revanche de ces relations, on pourra montrer que pour tout réel x nous avons

$$\cos^2(x) + \sin^2(x) = 1. \tag{1}$$

Remarque Bien que les nombres complexes ne soient pas étudiés dans le cadre, nous attirons l'attention du lecteur sur le fait que les nombres complexes, sont également employés dans d'autres champs des mathématiques comme les probabilités, pour déterminer une *fonction caractéristique* d'une variable aléatoire qui suit une distribution connue, permettant ainsi de la reconnaître ou encore de calculer des moments de ceete loi (espérance, variance).

Il nous reste à étudier la fonction tangente tan. Elle est définie pour tout $x \in \mathbb{R} \setminus \left\{ \frac{\pi}{2} + k\pi \right\}$ par

$$\tan(x) = \frac{\sin(x)}{\cos(x)}.$$

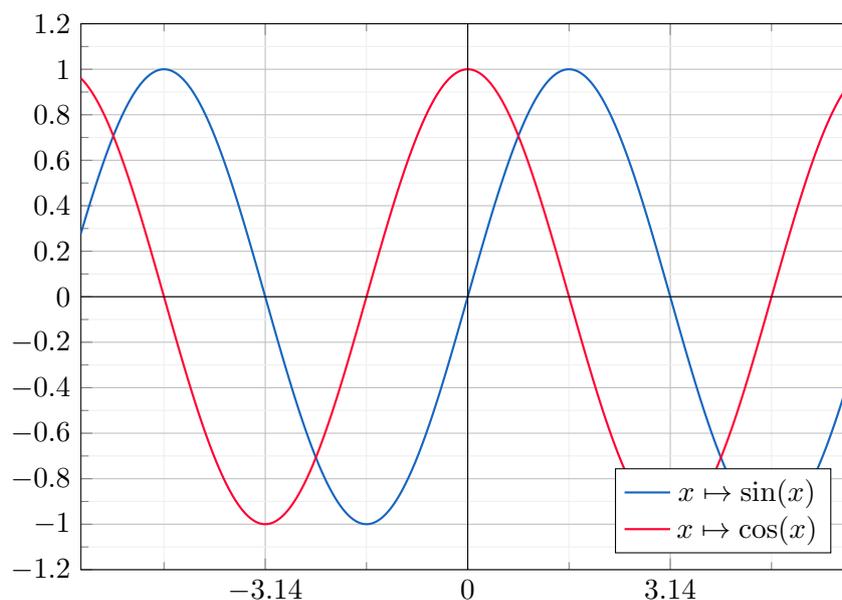


FIGURE 7 – Représentation graphique des fonctions trigonométriques sinus et cosinus dans l'intervalle $[-2\pi, 2\pi]$.

La fonction tan est donc par définition une fonction impaire et elle est également dérivable pour tout $x \in \mathbb{R} \setminus \left\{ \frac{\pi}{2} + k\pi \right\}$. Pour ces même réels x , sa dérivée est donnée par $\tan'(x) = \frac{1}{\cos^2(x)} = 1 + \tan^2(x)$. La deuxième égalité est donnée par la relation (1).

Une représentation graphique de la fonction tangente est donnée en Figure 8. Sur cette même figure, on montre qu'il est également possible de lire graphiquement la valeur de la tangente d'un angle α directement à partir du cercle trigonométrique, en considérant une droite formant un angle α avec l'axe des abscisses et la droite d'équation $x = 1$.

Fonction réciproque Précédemment, nous avons vu que les fonctions logarithme et exponentielle sont réciproques l'une de l'autre. Dans le cas présent, on ne peut pas dire que les fonctions circulaires sont bijectives de \mathbb{R} dans \mathbb{R} en raison de leur périodicité. On pourra revanche noter que les restrictions des fonctions $\sin : [-\pi/2, \pi/2] \rightarrow [-1, 1]$, $\cos : [0, \pi] \rightarrow [-1, 1]$ et $\tan : [-\pi/2, \pi/2] \rightarrow]-\infty, +\infty[$ sont des bijections. On peut donc définir des fonctions réciproques que l'on appelle respectivement arcsin, arccos et arctan. Il n'existe malheureusement pas d'expression de ces fonctions, on se contentera donc uniquement de leur existence.

Terminons cette présentation des fonctions usuelles en présentant les fonctions dites *hyperboliques*.

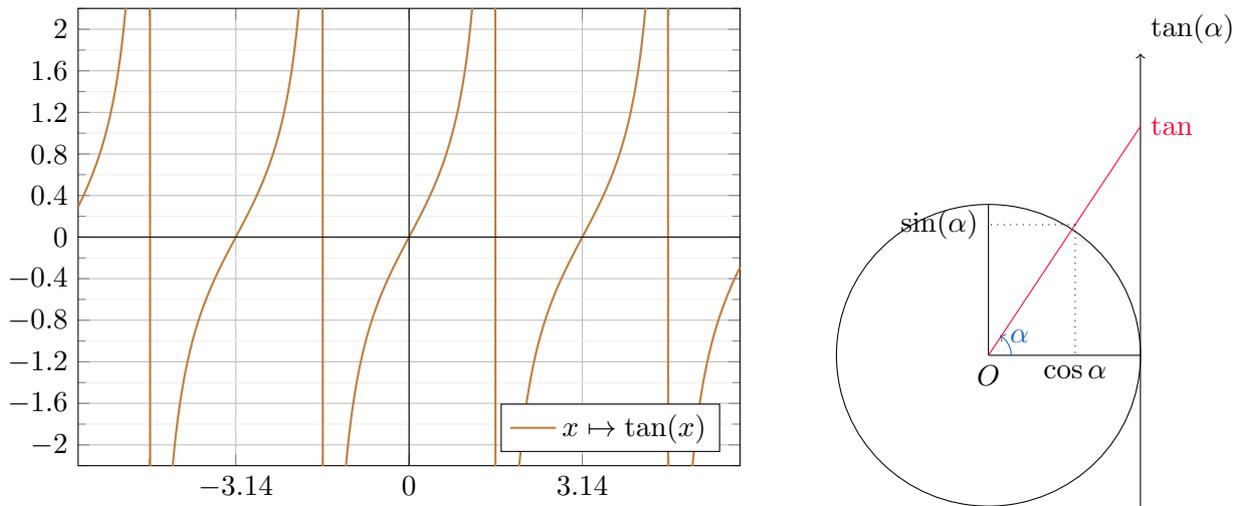


FIGURE 8 – Représentation de la fonction tangente, \tan , à gauche et illustration de l'interprétation géométrique de la tangente. La valeur de la tangente est représentée par la longueur du segment rouge sur la figure de droite.

2.5 Fonctions hyperboliques

Ces fonctions ressemblent fortement à celles étudiées précédemment, elles portent d'ailleurs des noms très proches. On les retrouve plus rarement que les précédentes mais l'une d'elle, la tangente hyperbolique est souvent utilisé dans des modèles d'apprentissage que l'on appelle *réseaux de neurones*. Commençons par les définir.

Définition 2.7: Fonctions hyperboliques

Les fonctions *sinus hyperbolique* et *cosinus hyperbolique*, notées respectivement sh et ch , sont définies pour tout réel x par les relations

$$\text{sh}(x) = \frac{\exp(x) - \exp(-x)}{2} \quad \text{et} \quad \text{ch}(x) = \frac{\exp(x) + \exp(-x)}{2}.$$

A partir de ces fonctions, nous pouvons également introduire la fonction *tangente hyperbolique*, notée th , définie pour tout réel x par la relation

$$\text{th}(x) = \frac{\text{sh}(x)}{\text{ch}(x)} = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}.$$

Les fonctions hyperboliques et plus particulièrement la tangente sont définies sur \mathbb{R} tout entier contrairement à son homologue circulaire. Une représentation graphique de ces fonctions est donnée en Figure 10.

De ces définitions, on peut directement décrire quelques liens entre ces fonctions

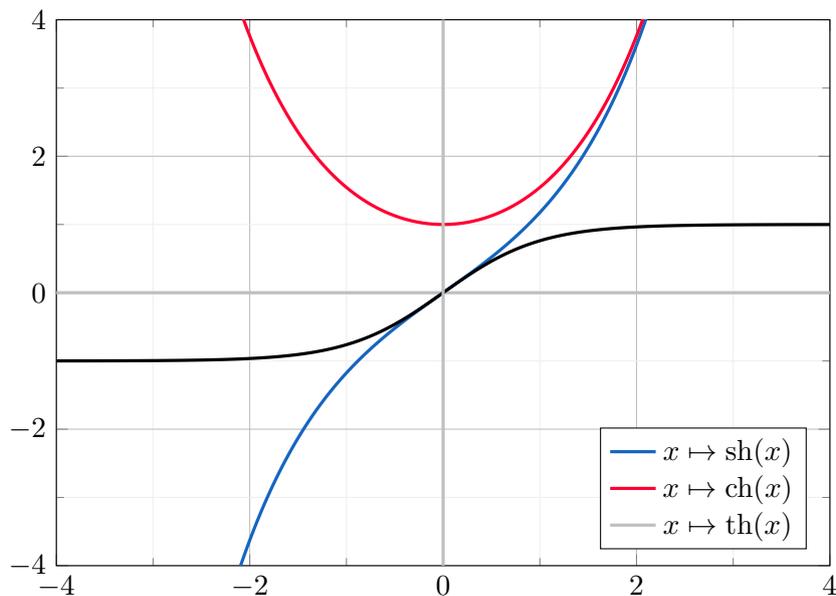


FIGURE 9 – Représentation des fonctions *sinus hyperbolique*, *cosinus hyperbolique* et *tangente hyperbolique*

hyperboliques, tout comme nous l’avons fait pour les fonctions circulaires. Par exemple, pour tout réel x nous avons les relations

$$\text{ch}(x) + \text{sh}(x) = e^x \quad \text{et} \quad \text{ch}^2(x) - \text{sh}^2(x) = 1. \quad (2)$$

Etudions maintenant ces fonctions d’un peu plus et notamment leurs propriétés.

Concernant le signe Pour tout $x \in \mathbb{R}$, la fonction ch est positive comme somme de deux fonctions positives. La fonction sh est positive pour tout $x \geq 0$ et négative sinon. En effet, $\text{sh}(x) \geq 0 \iff e^x \geq e^{-x}$ et cette dernière inégalité est vérifiée si et seulement si $x \geq 0$. Enfin, tout comme la fonction sh , la fonction th est positive pour tout $x \geq 0$ et négative dans le cas contraire.

Concernant la dérivabilité Ces différentes fonctions circulaires sont dérivables sur \mathbb{R} comme la somme ou le quotient de fonctions dérivables sur \mathbb{R} . En effet, nous avons vu plus tôt que la fonction \exp est une fonction définie et dérivable sur \mathbb{R} . Leurs dérivées sont données par la proposition suivante

Proposition 2.5: Dérivées fonctions hyperboliques

Les fonctions ch , sh et th sont dérivables sur \mathbb{R} et pour tout $x \in \mathbb{R}$, nous avons :

$$\text{sh}'(x) = \text{ch}(x), \quad \text{ch}'(x) = \text{sh}(x) \quad \text{et} \quad \text{th}'(x) = \frac{1}{\text{ch}^2(x)}.$$

Démonstration. On se contentera de montrer le premier point uniquement et les deux autres dérivées seront laissés à titre d'exercices.

Pour tout réel x , la fonction sh est dérivable. On utilisera simplement le fait que dérivée de la fonction $x \mapsto \exp(u(x)) = u'(x) \exp(u(x))$. Ce qui nous donne

$$\text{sh}'(x) = \left(\frac{\exp(x) - \exp(-x)}{2} \right)' = \frac{\exp(x) - (-) \exp(-x)}{2} = \frac{\exp(x) + \exp(-x)}{2} = \text{ch}(x).$$

La deuxième relation se montre de la même façon et la troisième nécessitera, en plus, d'utiliser la dérivée d'un quotient de deux fonctions. \square

On peut également réécrire la dérivée de la fonction th en fonction de la fonction th elle-même à l'aide de la relation (2), ce qui nous donne

$$\forall x \in \mathbb{R}, \text{th}'(x) = \frac{\text{ch}^2(x) - \text{sh}^2(x)}{\text{ch}^2(x)} = 1 - \text{th}^2(x).$$

Concernant les limites Les limites de ces fonctions s'obtiennent à partir des limites de la fonction exponentielle. Ainsi :

- $\lim_{x \rightarrow +\infty} \text{ch}(x) = +\infty$ et $\lim_{x \rightarrow -\infty} \text{ch}(x) = +\infty$.
- $\lim_{x \rightarrow -\infty} \text{sh}(x) = -\infty$ et $\lim_{x \rightarrow +\infty} \text{sh}(x) = +\infty$.
- $\lim_{x \rightarrow -\infty} \text{th}(x) = -1$ et $\lim_{x \rightarrow +\infty} \text{th}(x) = 1$.

Fonctions réciproques On finit cette présentation des fonctions en étudiant l'existence de fonctions réciproques en se basant sur la représentation graphique des fonctions circulaires données en Figure 10.

La fonction sh est une fonction strictement croissante de \mathbb{R} dans \mathbb{R} car sa dérivée est positive et ne s'annule qu'en un seul point. Elle est donc *bijective* et admet une fonction réciproque qui est également définie de \mathbb{R} dans \mathbb{R} . Elle est notée argsh pour *argument sinus hyperbolique* et est définie par

$$\operatorname{argsh}(x) = \ln \left(x + \sqrt{1 + x^2} \right).$$

La fonction ch est une fonction strictement croissante de \mathbb{R}^+ dans $[1, +\infty[$ car sa dérivée est positive et s'annule qu'en un seul point. Elle est donc *bijective* de \mathbb{R}^+ dans $[1, +\infty[$ et admet une fonction réciproque qui est également définie de $[1, +\infty[$ dans \mathbb{R}^+ . Elle est notée argch pour *argument cosinus hyperbolique* et est définie par

$$\operatorname{argsh}(x) = \ln \left(x + \sqrt{x^2 - 1} \right).$$

La fonction th est une fonction strictement croissante de \mathbb{R} dans \mathbb{R} car sa dérivée est positive et s'annule qu'en un seul point. Elle est donc *bijective* et admet une fonction réciproque qui est également définie de \mathbb{R} dans \mathbb{R} . Elle est notée argth pour *argument tangente hyperbolique* et est définie par

$$\operatorname{argth}(x) = \frac{1}{2} \ln \left(\frac{1+x}{1-x} \right).$$

En réutilisant le fait qu'une fonction et sa réciproque sont symétriques par rapport à la droite d'équation $y = x$, il est très facile d'obtenir une représentation graphique des fonctions réciproques hyperboliques. Nous pourrions également étudier les dérivées mais nous allons nous arrêter là car elles ne présentent que peu d'intérêt pour la suite et elles sont peu utilisées dans un cadre d'analyse de données.

La prochaine section s'intéresse à l'étude de la continuité et de la dérivabilité des fonctions en général.

3 Continuité des fonctions réelles

Dans cette section, nous présentons les définitions de continuité et de dérivabilité d'une fonction d'une variable réelle. Cela sera également l'occasion de revenir sur l'étude des variations des fonctions et le lien qui existe entre dérivée d'une fonction et sa monotonie.

3.1 Quelques caractéristiques des fonctions

Commençons par définir quelques propriétés simples des fonctions, comme la monotonie, la parité ou encore la périodicité.

Définition 3.1: Variations et Monotonies

Soit I un intervalle de \mathbb{R} et f une fonction définie de I dans \mathbb{R} . La fonction f est dite :

- croissante si $\forall(x, y) \in I^2, x \leq y \implies f(x) \leq f(y)$,
- décroissante si $\forall(x, y) \in I^2, x \leq y \implies f(x) \geq f(y)$,
- strictement croissante $\forall(x, y) \in I^2, x \leq y \implies f(x) < f(y)$,
- strictement décroissante $\forall(x, y) \in I^2, x \leq y \implies f(x) > f(y)$.

On pourra aussi caractériser les fonctions constantes en disant que ce sont des fonctions à la fois croissantes et décroissantes. L'étude des variations d'une fonction va permettre de mieux comprendre les relations qui existent une variable d'entrée (un *input* en informatique) et la variable de sortie (l'*output*). On pourra exemple étudier l'évolution d'une variable en fonction d'une autre variable sur le plan statistique (c'est ce que l'on fera dans un contexte de *régression linéaire*). Etudier les variations d'une fonction va également nous permettre de construire des algorithmes dont le but sera de trouver le minimum (ou maximum) d'une fonction (sous réserves de bonnes propriétés de cette dernière). Mais aussi d'étudier la convergence de certaines suites qui interviennent naturellement dans certains algorithmes.

Mais ce sont pas les seules propriétés que l'on peut étudier. On pourra également s'intéresser à d'autres propriétés comme les notions de parité ou de périodicité des fonctions : utiles pour le calcul d'intégrale ou encore pour simplifier certains calculs (utiles aussi sur le plan théorique).

Commençons par regarder la notion de parité

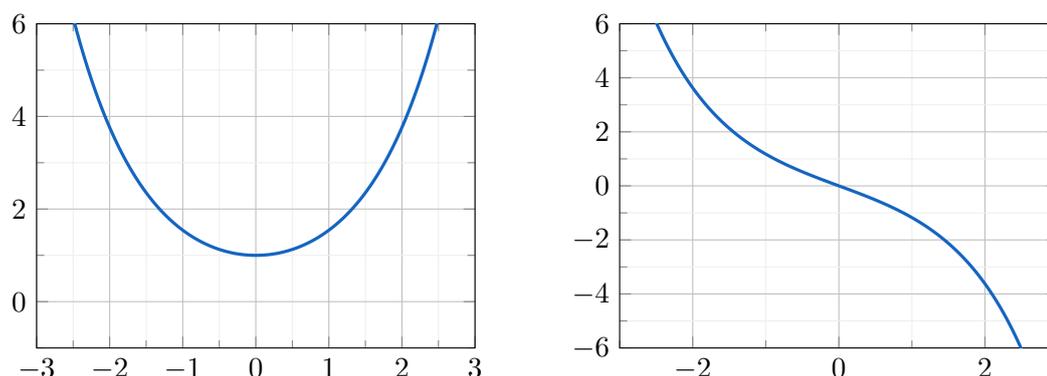


FIGURE 10 – Exemple d’une fonction paire (à gauche) et d’une fonction impaire (à droite)

Définition 3.2: Parité d’une fonction

Soit I un intervalle de \mathbb{R} et f une fonction définie de I dans \mathbb{R} . La fonction f est dite :

- **paire** si pour tout $x \in I$, $f(x) = f(-x)$,
- **impaire** si pour tout $x \in I$, $f(-x) = -f(x)$.

Sur le plan graphique (voir Figure 10) une fonction paire se reconnaît par son graphe symétrique par rapport à l’axe des ordonnées. Une fonction impaire est elle symétrique par rapport à l’origine du repère, *i.e.* si l’on effectue une rotation du graphe d’un angle de 180, ce dernier est invariant.

Exemple 3.1. Les fonctions $f : x \mapsto x^2, x^4$ ou plus généralement $f : x \mapsto x^{2n}$, $n \in \mathbb{N}$ sont des fonctions paires.

En effet, posons $f(x) = x^{2n}$, $n \in \mathbb{N}$ alors

$$f(-x) = (-x)^{2n} = ((-1)^2)^n (x^{2n}) = x^{2n} = f(x).$$

Les fonctions \cos et \cosh le sont également. Plus généralement la somme de deux fonctions paires reste une fonction paire.

Exemple 3.2. Les fonctions $f : x \mapsto x, x^3$ ou plus généralement $f : x \mapsto x^{2n+1}$, $n \in \mathbb{N}$ sont des fonctions paires.

En effet, posons $f(x) = x^{2n+1}$, $n \in \mathbb{N}$ alors

$$f(-x) = (-x)^{2n+1} = (-1) ((-1)^2)^n (x^{2n+1}) = -x^{2n+1} = -f(x).$$

Les fonctions \sin et sh le sont également. Plus généralement la somme de deux fonctions impaires reste une fonction impaire.

Définition 3.3: Périodicité d'une fonction

Soit I un intervalle de \mathbb{R} , T un nombre réel tels que pour tout $x \in I$, $x + T \in I$. Soit f une fonction définie de I dans \mathbb{R} . On dit que la fonction f est **périodique** de période T (ou encore que f est T -périodique) si pour tout $x \in I$, $f(x + T) = f(x)$.

Cette notion de périodicité est importante en *Analyse de Fourier* lorsque l'on cherche à étudier des signaux notamment, ce qui est très courant dans le domaine médical lorsque l'on effectue des électro-cardiogrammes ou des électro-encéphalogrammes mais aussi dans l'acoustique.

Nous avons déjà rencontrés des fonctions périodiques dans la section précédente lorsque nous avons introduit les fonctions circulaires, comme les fonctions \cos et \sin représentées en Figure 11.

Exemple 3.3. On peut aussi créer soit même des fonctions qui ne sont pas forcément 2π -périodiques. Si on considère la fonction $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par

$$f(x) = \begin{cases} x & \text{si } x \in [0, 1[, \\ x - 1 & \text{si } x \in [1, 2[, \\ \dots & \\ x - k & \text{si } x \in [k, k + 1[, \\ x - (k + 1) & \text{si } x \in [k + 1, k + 2[, \\ \dots & \end{cases},$$

pour n'importe quel entier naturel k . Alors la fonction f est 1-périodique. En effet, pour tout réel x , il existe un entier k tel que $x \in [k, k + 1[$, donc $x + 1 \in [k + 1, k + 2[$

$$f(x + 1) = x + 1 - (k + 1) = x - k = f(x),$$

où la dernière découle du fait que $x \in [k, k + 1[$.

On peut également changer la période d'une fonction périodique en multipliant la variable x par une certaine valeurs. En effet, soit $a > 0$, et considérons une fonction f qui est T -périodique. Alors la fonction $g : x \mapsto f(ax)$ sera $\frac{T}{a}$ -périodique. Ce sont des notions que l'on retrouve très souvent en analyse harmonique.

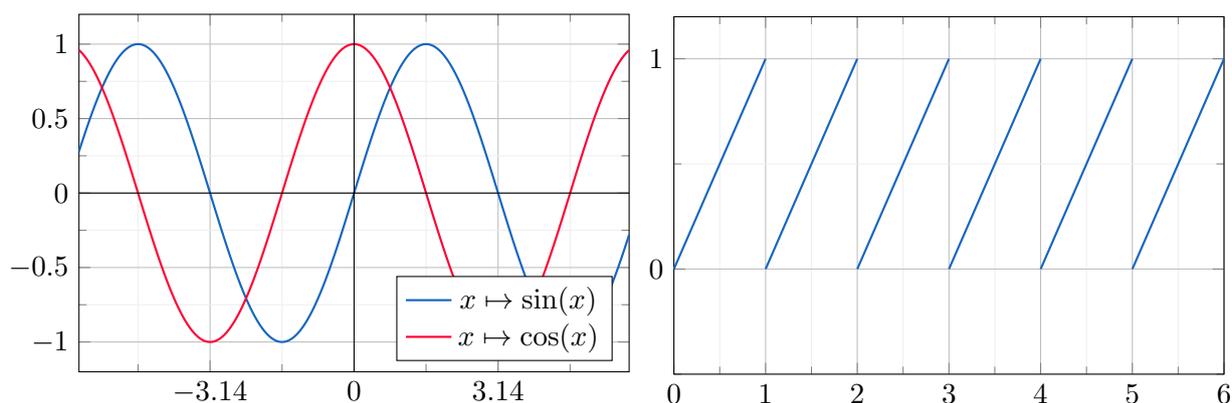


FIGURE 11 – Exemples de deux fonctions 2π -périodiques à gauche et d’une fonction 1-périodique à droite.

Exemple 3.4. *Considérons la fonction cos qui est 2π -périodique, et soit $T > 0$. Alors la fonction $f : x \mapsto \cos\left(\frac{2\pi}{T}x\right)$ est T -périodique. En effet, pour tout réel x nous avons*

$$f(x + T) = \cos\left(\frac{2\pi}{T}(x + T)\right) = \cos\left(\frac{2\pi}{T}x + 2\pi\right) = \cos\left(\frac{2\pi}{T}x\right) = f(x).$$

Regardons maintenant une autre caractéristique des fonctions que l’on appelle la continuité.

3.2 Continuité

C’est une caractéristique élémentaire que l’on a tous rencontrés dans sa vie. On a toujours présenté la continuité d’une fonction à l’aide de sa représentation graphique. Nous disons naïvement qu’une fonction est continue si nous pouvons la représenter graphiquement d’un seul trait ou encore sans lever le crayon.

Ce n’est pas une définition rigoureuse de la continuité mais on comprend l’idée que l’on peut finalement approcher n’importe quel point du graphe en suivant la courbe sans que cette dernière ne présente d’interruption, que cela soit dans un sens de parcours, ou dans l’autre.

Essayons de traduire cela avec la notion de limite et regardons ce qu’il se passe en un point de notre courbe pour une fonction f donnée.

Définition 3.4: Continuité en un point

Soit f une fonction réelle définie sur un intervalle I de \mathbb{R} et soit a un point de I .
On dit que f est continue en a si $\lim_{x \rightarrow a} f(x) = f(a)$.

Lorsque la fonction f est définie en a , si f admet une limite en a , cette limite ne peut pas être différente de $f(a)$.
Ainsi f est continue en a si et seulement si elle admet une limite a qui est égale à $f(a)$.

Exemple 3.5. Soit a un nombre réel et considérons la fonction $f : x \mapsto \frac{x}{x-a}$, la fonction f n'est pas définie en a , elle ne peut donc pas être continue en a .

Exemple 3.6. Considérons la fonction f définie pour tout réel x par

$$f(x) = \begin{cases} x^2 & \text{si } x \leq 0, \\ x + 1 & \text{si } x > 0. \end{cases}$$

La fonction f est bien définie en 0, mais n'est pourtant pas continue en 0. Elle présente un saut de discontinuité en ce point.

On représente aussi graphiquement des fonctions discontinues en Figure 12 avec des sauts de discontinuité ou encore dans le cas d'une fonction, la fonction *sinus cardinal* qui est la fonction $x \mapsto \frac{\sin(x)}{x}$, où la discontinuité, sur le plan graphique, est moins évidente.

Dans l'Exemple 3.6, la fonction est bien définie pour tout nombre réel mais elle n'est pourtant pas continue en 0.

En effet, si on considère les limites de la fonction f en 0 suivantes :

$$\lim_{x \rightarrow 0^+} f(x) = 1 \quad \text{et} \quad \lim_{x \rightarrow 0^-} f(x) = 0.$$

On va donc introduire des notions de limite à gauche et de limite à droite pour une fonction, ce qui peut servir à étudier la continuité de cette dernière en un point donné.

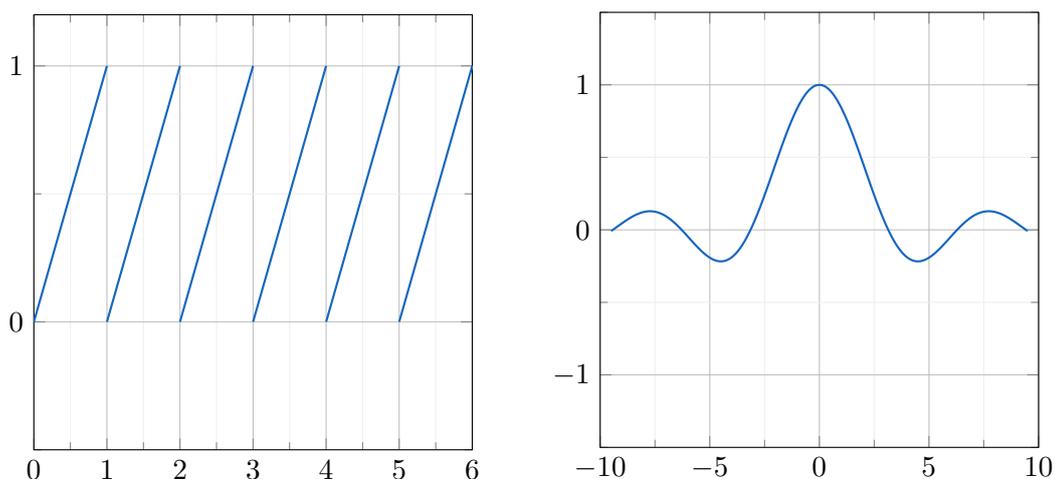


FIGURE 12 – Exemple de deux graphes de fonctions discontinues. La fonction représentée à gauche est celle de l'Exemple 3.3 qui présente des discontinuités lorsque pour toute valeur entière de x . Le graphe de droite est celui de la fonction *sinus cardinal*, définie par $x = \frac{\sin(x)}{x}$ pour tout x non nul. Cette fonction n'est pas définie en 0 donc pas continue en ce point.

Définition 3.5: Limite à gauche (à droite)

Soit f une fonction réelle définie sur un intervalle I de \mathbb{R} et soit a un point de I . On dit que :

- f est continue à gauche en a si $\lim_{x \rightarrow a^-} f(x) = f(a)$,
- f est continue à droite en a si $\lim_{x \rightarrow a^+} f(x) = f(a)$.

f sera alors continue en a si sa limite à gauche et à droite en a est bien la même. A l'aide de cette notion de limite à gauche et à droite, nous pourrions même définir des fonctions en des points en lesquels elles ne sont pas initialement continues. C'est ce que l'on appelle le **prolongement par continuité**.

Soit a un nombre réel et considérons une fonction f définie pour tout réel $x \in \mathbb{R} \setminus \{a\}$.

$$\lim_{x \rightarrow a^-} f(x) = \lim_{x \rightarrow a^+} f(x).$$

Exemple 3.7. Soit f la fonction définie pour tout réel x par

$$f(x) = \begin{cases} x^2 \sin(1/x) & \text{si } x \neq 0, \\ 0 & \text{si } x = 0. \end{cases}$$

La fonction f est continue en 0 car pour tout réel $x \in \mathbb{R}^*$, nous avons

$$|f(x)| = x^2 |\sin(1/x)| \leq x^2.$$

Comme $\lim_{x \rightarrow 0} x^2 = 0$, on a $\lim_{x \rightarrow 0} f(x) = 0 = f(0)$.

Exemple 3.8. Considérons la fonction sinus cardinal, sinc définie pour tout réel x non nul par

$$\text{sinc}(x) = \frac{\sin(x)}{x}.$$

Cette fonction n'est pas définie en 0, mais nous avons que son graphe (voire graphique de droite en Figure 12) semblait continue, i.e. que les limites à gauche et à droite 0 et même égales à 1.

Nous pourrions justifier pourquoi au moment où l'on aura introduit la notion de nombre dérivé d'une fonction.

Maintenant que nous avons caractérisé la continuité d'une fonction en un point a d'un intervalle I . Nous pouvons caractériser la continuité sur tout l'intervalle.

Définition 3.6: Continuité sur un intervalle

Soit f une fonction définie de I dans \mathbb{R} , f est dite continue sur I si elle est continue en tout point de I , i.e. si pour tout $a \in I$, on a bien $\lim_{x \rightarrow a} f(x) = f(a)$.

Une telle fonction f sera alors dite C^0 (pour dire continue) ou plus précisément, on écrira que $f \in C^0(I, \mathbb{R})$.

Bien évidemment, la somme ou encore le produit de fonctions continues reste une fonction continue!

On peut donner une autre définition de la continuité beaucoup plus abstraite mais qui nous sera utile pour caractériser d'autres formes de la continuité pour des fonctions.

Définition 3.7: Continuité

Soit f une fonction définie dans un intervalle I de \mathbb{R} à valeurs dans \mathbb{R} , la fonction f est alors dite continue en tout point $x \in I$, si

$$\forall x \in I, \forall \varepsilon > 0, \exists \eta > 0, \text{ tel que}$$

$$\forall y \in I, |x - y| \leq \eta \implies |f(x) - f(y)| \leq \varepsilon.$$

Cette définition ne sera pas utilisée en pratique mais nous l'utilisons pour des soucis de rigueur de présentation de futurs résultats.

Le résultat suivant permet d'étudier l'existence d'un zéro pour une fonction de données ou, plus généralement l'existence d'une solution de l'équation $f(x) = a$, $a \in \mathbb{R}$ pour une fonction f continue.

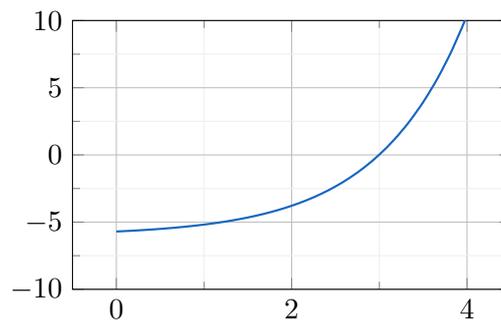
Théorème 3.1: Existence

Soit f une fonction continue de I dans \mathbb{R} , et soient a et b deux points de I tels que $a < b$.

Si $f(a)f(b) \leq 0$, alors il existe $c \in [a, b]$ tel que $f(c) = 0$.

La démonstration se fait par *dichotomie*. Elle est *constructiviste* ce qui peut faire un excellent exercice de programmation.

Démonstration. Soit $a < b$ deux nombres réels et f une fonction continue sur l'intervalle $[a, b]$ telle que $f(a)$ et $f(b)$ soient de signes opposés. Par exemple $f(a) > 0 > f(b)$.



Dichotomie : On va chercher à couper, itérativement, notre intervalle de recherche en deux. On considère deux suites $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ qui vont définir les bornes de notre

intervalle de recherche, telles que $a_0 = a$ et $b_0 = b$.

On pose $m_1 = \frac{a_0 + b_0}{2}$ et on évalue $f(m_1)$. On a donc deux possibilités :

1. $f(m_1)$ et $f(a_0)$ sont de signes contraires,
2. $f(m_1)$ et $f(b_0)$ sont de signes contraires.

Dans le premier cas, cela signifie que la valeur c recherchée se trouve dans l'intervalle $[a_0, m_1]$, sinon, deuxième cas, elle se trouve dans l'intervalle $[m_1, b_0]$.

- Si on est dans le cas 1), on posera alors $a_1 = a$ et $b_1 = m_1$
- Si on est dans le cas 2), on posera alors $a_1 = m_1$ et $b_1 = b$

Et on recommence le processus! On posera ensuite $m_2 = \frac{a_1 + b_1}{2}$ et on évalue $f(m_2)$ et on applique la même règle que précédemment.

On a donc deux possibilités :

1. $f(m_2)$ et $f(a_1)$ sont de signes contraires,
2. $f(m_2)$ et $f(b_1)$ sont de signes contraires.

Dans le premier cas, cela signifie que la valeur c recherchée se trouve dans l'intervalle $[a_1, m_2]$, sinon, deuxième cas, elle se trouve dans l'intervalle $[m_2, b_1]$...

L'algorithme de dichotomie est appliqué à chaque fois au sous-intervalle dans lequel on observe un changement de signe. On peut y voir là un algorithme récursif au cours duquel la longueur de l'intervalle étudié est divisée par 2 à chaque itération.

L'erreur absolue de cette méthode ; après n étapes est donc au plus égale à

$$\frac{b - a}{2^{n+1}}.$$

□

Enfin, si on cherche à approcher la solution avec une erreur au plus égale à $\varepsilon > 0$, on peut alors dénombrer le nombre d'itérations nécessaires pour cela :

$$\frac{b - a}{2^N} \leq \varepsilon \implies N = \left\lceil \log_2 \frac{b - a}{\varepsilon} \right\rceil,$$

où $\lceil x \rceil$ désigne le plus petit entier supérieur à x .

Remarque : sur le plan algorithmique, ce n'est pas la méthode la plus efficace pour rechercher un 0 d'une fonction, sa convergence est lente. Il existe d'autres méthodes (comme la méthode de Newton) qui permet de rechercher des zéros de façon plus efficace. En revanche cette méthode a l'avantage d'être applicable à beaucoup plus de situations. Le pseudo-code de cette recherche par dichotomie est présentée en Algorithme 3

Algorithm 3: Procédure de recherche par dichotomie

Entrée: $a < b \in \mathbb{R}$, une fonction continue f sur $[a, b]$, $\varepsilon > 0$

Sortie : un zéro de la fonction f

Factorielle(n)

```

1: while  $(b - a) \geq \varepsilon$  do
2:    $m = \frac{b + a}{2}$ 
3:   if  $f(a)f(m) \leq 0$  then
4:      $b = m$ 
5:   else
6:      $a = m$ 
7:   end if
8: end while

```

Une application directe de ce résultat permet de justifier que tout polynôme de degré impair s'annule au moins une fois.

Exemple 3.9. Soient a et b deux nombres réels. Considérons un polynôme P de degré impair, par exemple $P(x) = x^{2n+1} + ax + b$. Lorsque $x \rightarrow +\infty$, alors $P(x) > 0$ et lorsque $x \rightarrow -\infty$, alors $P(x) < 0$. Une application du résultat nous assure l'existence d'un nombre réel x_0 tel que $P(x_0) = 0$.

On peut alors utilisée une approche par dichotomie pour déterminer la valeur de x_0 .

Ce résultat permet par exemple de montrer que si deux fonctions f et g sont (i) toutes deux continues sur un intervalle I et (ii) dont la différence change de signe dans ce même intervalle. Alors il existe un réel $x_0 \in I$ tel que $f(x_0) = g(x_0)$.

Exemple 3.10. Considérons la fonction f définie pour tout réel x par $I = [0, \ln(6)]$, définie par $f(x) = e^x - 6x$. On peut montrer que cette fonction f s'annule au moins une fois sur cet intervalle.

En effet, nous avons $f(0) = 1 > 0$ et $f(\ln(6)) = 6 - 6 \ln(6) = 6(1 - \ln(6)) < 0$.

On peut également utiliser ce résultat pour montrer qu'une fonction f admet un point fixe en appliquant ce même résultat à la fonction $f(x) - x$.

Le résultat précédent peut se généraliser de la façon suivante

Théorème 3.2: Théorème des valeurs intermédiaires

Pour toute fonction f continue sur un intervalle I de \mathbb{R} à valeurs réelles, l'image $f(I)$ est un intervalle de \mathbb{R} .

Ou encore, de façon équivalente :

Pour toute application continue f de $[a, b]$ dans \mathbb{R} et tout réel u compris entre $f(a)$ et $f(b)$, il existe au moins un réel c compris entre a et b tel que $f(c) = u$.

Regardons maintenant les principaux résultats que l'on peut tirer de la continuité d'une fonction mais aussi les différentes continuités qui existent.

3.3 Extrema d'une fonction et fonction lipschitzienne

Le premier résultat porte sur le caractère borné des fonctions continues sur un intervalle I de \mathbb{R} .

Théorème 3.3: Théorème de Weierstrass

Soit f une fonction continue sur un intervalle fermé I de \mathbb{R} . Alors la fonction f est bornée et atteint ses bornes.

Il est fait mention d'un intervalle fermé I dans ce résultat, il s'agit d'intervalle I de la forme $[a, b]$ où $a, b \in \mathbb{R}$. A contrario, un intervalle I est dit ouvert s'il est de la forme $]a, b[,] - \infty, b[,]a, +\infty[$

A noter qu'un intervalle peut être ni ouvert ni fermé, comme c'est le cas des intervalles de la forme $[a, b[,]a, b]$ ou encore $] - \infty, b]$ ou encore $[a, +\infty[$.

Comme une fonction f continue est bornée et atteint ses bornes, on peut également avoir des informations sur la nature de l'ensemble image par la fonction f .

Théorème 3.4: Image d'un segment par une fonction continue

Soit f une fonction continue sur un intervalle fermé I de \mathbb{R} . Alors l'image de l'intervalle I par la fonction f est également un intervalle fermé et borné.

La définition et le résultat suivants sont plus anecdotiques mais permettent simplement de montrer que la notion de continuité évoquée en Définition 3.7 peut être spécialisée, comme le montre la définition suivante ?

Définition 3.8: Continuité Uniforme

Soit f une fonction définie sur un intervalle I dans \mathbb{R} , f est dite **uniformément continue** sur I si :

$$\forall \varepsilon > 0, \exists \eta > 0 \text{ tels que} \\ \forall (x, y) \in I^2, |x - y| \leq \eta \implies |f(x) - f(y)| \leq \varepsilon.$$

Exemple 3.11. On peut considérer la fonction $f : I = [1, +\infty[\rightarrow \mathbb{R}$, alors f est uniformément continue. En effet, pour tout $\varepsilon > 0$, nous avons, pour tout $x, y \in I$ nous avons

$$|f(x) - f(y)| = \left| \frac{1}{x} - \frac{1}{y} \right| = \frac{|x - y|}{xy} \leq |x - y|.$$

En posant $\eta = \varepsilon$, on a démontré le résultat souhaitait.

Remarquons que la continuité uniforme implique la continuité simple. Pourquoi selon vous? Notons que la réciproque est fautive, comme nous le voyons dans l'exemple ci-dessous.

Exemple 3.12. Considérons la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = x^2$. La fonction f n'est pas uniformément continue sur \mathbb{R} .

Pour montrer cela, nous devons définir ce que cela signifie de ne pas être uniformément continue en prenant la négation la Définition 3.8. Nous devons donc montrer que

$$\exists \varepsilon > 0, \forall \eta > 0 \text{ tels que} \\ \exists (x, y) \in I^2, |x - y| \leq \eta \implies |f(x) - f(y)| \geq \varepsilon.$$

Dit autrement, il suffit d'exhiber un couple (x, y) dont la différence peut être rendue aussi proche de 0 que l'on souhaite mais pour lequel la différence des images, i.e. $|f(x) - f(y)|$ ne sera jamais aussi petite qu'on le souhaite (ce "souhait" est déterminé par la valeur de ε).

Pour cela, on va considérer les suites $(x_n)_{n \in \mathbb{N}^*}$ et $(y_n)_{n \in \mathbb{N}^*}$ définies par

$$x_n = n + \frac{1}{n} \quad \text{et} \quad y_n = n.$$

Ainsi, pour tout $n > 0$, nous avons

$$|x_n - y_n| = \frac{1}{n} \quad \text{et} \quad |f(x_n) - f(y_n)| = \left| \left(n + \frac{1}{n} \right)^2 - n^2 \right| = 2 + \frac{1}{n^2}.$$

En particulier lorsque $n \rightarrow +\infty$, nous avons $\lim_{n \rightarrow +\infty} |x_n - y_n| = 0$ et $\lim_{n \rightarrow +\infty} |f(x_n) - f(y_n)| = 2$.

Nous avons donc montré que l'on peut rendre x_n et y_n infiniment proche de 0 pour autant la différence des images sera toujours plus grande que 2.

Exemple 3.13. On va reprendre notre fonction $f : x \mapsto x^2$ mais nous considérons cette sa restriction sur l'intervalle $[0, a]$ où $a > 0$. Dans ce cas, la fonction f est uniformément continue.

Dans ce cas

$$|f(x) - f(y)| = |x^2 - y^2| = |x - y|(x + y) \leq 2a|x - y|.$$

Ainsi, en choisissant $\eta = \frac{\varepsilon}{2a}$, on obtient directement $|f(x) - f(y)| < \varepsilon$.

Ce troisième exemple semble indiquer que si une fonction est continue sur un intervalle fermé, *i.e.* un segment alors elle est uniformément continue. C'est le *théorème de Heine*.

Théorème 3.5: Théorème de Heine

Toute fonction continue sur un segment est uniformément continue sur ce segment.

On va maintenant exiger une propriété plus forte de continuité qui contrôle aussi les variations de notre fonction. Il s'agit du caractère lipschitzien.

Définition 3.9: Fonction Lipschitzienne

Soit I un intervalle de \mathbb{R} , une fonction f définie sur I est dite **lipschitzienne** s'il existe un réel $k > 0$ tel que :

$$\forall (x, y) \in I^2, |f(x) - f(y)| \leq k|x - y|.$$

On dit alors que f est *k-lipschitzienne*.

Cette propriété aura une grande importance lorsque l'on abordera, l'année suivante, les algorithmes de descente de gradient. On la retrouvera également dans certaines théories de l'apprentissage machine, notamment pour étudier les algorithmes.

Exemple 3.14. *Considérons la fonction $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ définie par $f(x) = \frac{1}{1+x}$. Alors la fonction f est 1-lipschitzienne.*

En effet, pour tout $x, y \geq 0$, nous avons

$$\left| \frac{1}{1+x} - \frac{1}{1+y} \right| = \left| \frac{x-y}{(1+x)(1+y)} \right| \leq |x-y|.$$

La dernière inégalité est due au fait que pour tout $x \geq 0$, $1+x \geq 1$, donc $\frac{1}{1+x} \leq 1$.

Dans le cas où la fonction f est k -lipschitzienne avec $0 \leq k < 1$, la fonction f est dite **contractante**.

En outre, on peut montrer que toute application lipschitzienne est uniformément continue.

En effet, si on se donne une fonction f qui est k -lipschitzienne, alors pour tout x, y nous avons

$$|f(x) - f(y)| \leq k|x - y|.$$

Il suffit alors de choisir $\eta = \frac{\varepsilon}{k}$ pour retrouver la définition de continuité uniforme.

Ainsi, on pourra également montrer qu'il existe un lien étroit entre le fait d'être lipschitzien et le fait d'être dérivable car la constante de lipschitz kM n'est rien d'autre que, par définition

$$\max_{x,y} \frac{|f(x) - f(y)|}{|x - y|}.$$

et en prenant, en particulier, $y = x + h$ et en faisant tendre h vers 0, on retrouve notre définition de notre nombre dérivé, objet d'étude de notre prochaine section et comme illustré en Figure 13.

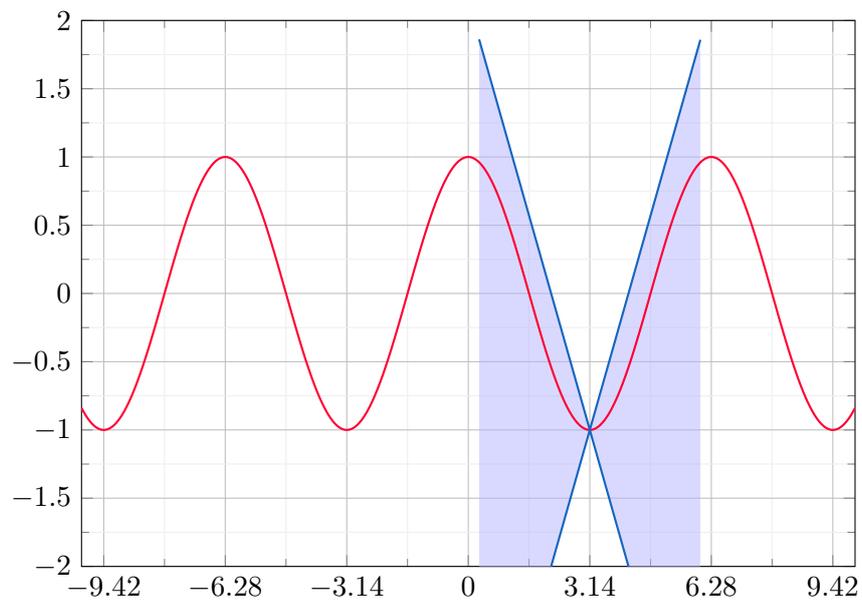


FIGURE 13 – Illustration du caractère lipschitzien d'une fonction. En déplaçant le cône le long de la courbe, ce dernier ne viendra jamais intersecter la courbe. On peut aussi voir la caractère lipschitz comme le fait que la courbe représentative de la fonction sera toujours dans la "zone bleue" quelque soit la position du cône, *i.e.* la courbe ne se trouvera jamais dans la "zone blanche".

4 Dérivabilité

La dérivabilité des fonctions va nous permettre d'étudier le comportement de cette dernière, *i.e.* sa monotonie, aussi bien globalement que localement.

4.1 Définition de la dérivabilité

Définissons tout d'abord ce qu'est un nombre dérivé.

Définition 4.1: Nombre dérivé

Soit une fonction f définie de I dans \mathbb{R} , et soit a un point de I . La fonction f est dite dérivable en a si la quantité $\frac{f(x) - f(a)}{x - a}$ admet une limite finie lorsque x tend vers a . On appelle alors cette limite *nombre dérivée de f en a* et on la note $f'(a)$.

La Figure 14 illustre cette définition. On y compare le graphe de la fonction f , la fonction définissant le nombre dérivée (D) et la tangente de f au point a , T_a . Plus x est proche de a , plus la droite D et la tangente en a se confondent. On rappelle que la tangente en a a pour équation

$$y = f(a) + f'(a)(x - a),$$

où $f'(a)$ traduit le taux d'accroissement de la tangente.

On peut définir la notion de nombre dérivée à gauche et à droite pour une fonction f selon que l'on approche du point a par valeurs supérieures ou par valeurs inférieures.

Ainsi, si la fonction f est définie à **droite de** a et si la quantité $\frac{f(x) - f(a)}{(x - a)}$ admet une limite à droite en a ; on appelle cette limite à droite le *nombre dérivée de f en a à droite* et on le note $f'(a^+)$.

On peut faire exactement la même chose à gauche ! Cette distinction est importante car on peut potentiellement trouver des limites différentes à gauche et à droite pour cette fonction, comme ce fut le cas pour la continuité. Bien évidemment, si ces limites sont différentes, la fonction n'est donc pas continuellement dérivable en ce point a .

La notion de dérivabilité est également étroitement liée à celle de la continuité

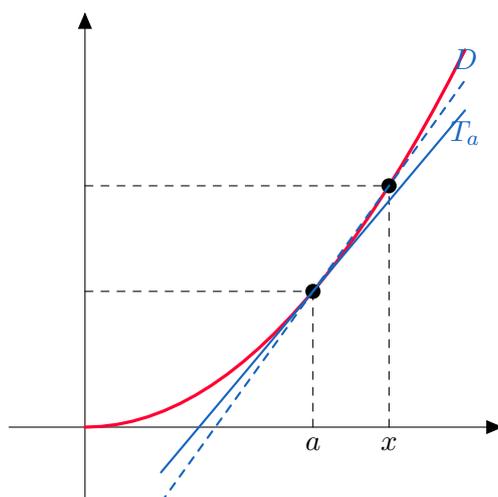


FIGURE 14 – Représentation d’une fonction f quelconque (en rouge). On compare la droite (en pointillés) reliant deux points de coordonnées $(x, f(x))$ et $(a, f(a))$ à la tangente à la courbe représentative f au point a .

Proposition 4.1: Lien avec la continuité

Soit f une fonction définie de I dans \mathbb{R} et soit a un point de I . Si la fonction f est dérivable en a , alors elle est continue en a

Démonstration. Supposons que f est dérivable en a . Alors la quantité $\frac{f(x) - f(a)}{x - a}$ admet une limite en a et elle est donc bornée au voisinage de a . Ainsi, il existe $M > 0$ et $\alpha > 0$ tels que pour tout $x \in [a - \alpha, a + \alpha] \cap I$ et $x \neq a$, nous ayons

$$\left| \frac{f(x) - f(a)}{x - a} \right| \leq M.$$

On a alors, pour tout $x \in [a - \alpha, a + \alpha] \cap I$, $|f(x) - f(a)| \leq M|x - a|$. Or $\lim_{x \rightarrow a} |x - a| = 0$, donc $\lim_{x \rightarrow a} |f(x) - f(a)| = 0$, c’est-à-dire $\lim_{x \rightarrow a} f(x) = f(a)$; la fonction f est donc continue en a . \square

On s’intéresse à quelques résultats généraux sur la dérivation de fonctions. Il s’agit essentiellement de faire quelques rappels sur la dérivation de somme, produit ou de composé de fonctions.

Proposition 4.2: Opérations sur la dérivation

Soient f et g deux fonctions définies sur un intervalle I et dérivables en $a \in I$.
Alors :

- pour tout $\lambda \in \mathbb{R}$, λf est dérivable en a et $(\lambda f)'(a) = \lambda f'(a)$,
- la fonction $f + g$ est dérivable en a et $(f + g)'(a) = f'(a) + g'(a)$,
- la fonction fg est dérivable en a et $(fg)'(a) = f'(a)g(a) + f(a)g'(a)$
- si $g(a) \neq 0$, la fonction $1/g$ est dérivable en a et $(1/g)'(a) = -g'(a)/g(a)^2$.
- si $g(a) \neq 0$, la fonction f/g est dérivable en a et $(f/g)'(a) = \frac{f'(a)g(a) - g'(a)f(a)}{g(a)^2}$.

Exemple 4.1. *Considérons la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ qui, à tout réel x , associe $(2x^2 + 1)e^x$.*

La fonction se présente comme un produit de deux fonctions dérivables u et v , où $u(x) = 2x^2 + 1$ et $v(x) = e^x$, toutes deux dérivables sur \mathbb{R} .

Ainsi, pour tout réel x , nous avons

$$f'(x) = u'(x)v(x) + u(x)v'(x) = 4xe^x + (2x^2 + 1)e^x = e^x(2x^2 + 4x + 1).$$

Le plus souvent, nous rencontrons des fonctions qui sont composées avec d'autres fonctions.

Proposition 4.3: Dérivation composition

Soit f une fonction définie de I dans \mathbb{R} , soit g une fonction définie d'un intervalle J de \mathbb{R} dans \mathbb{R} tel que $f(I) \subset J$, et $a \in I$.

Si f est dérivable en a et si g est dérivable en $f(a)$, alors $g \circ f$ est dérivable en a et

$$(g \circ f)'(a) = f'(a) \times (g' \circ f)(a).$$

Exemple 4.2. *Considérons la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$, qui à tout réel x associe $\ln(3x^2 + 2)$.*

La fonction f est bien dérivable sur \mathbb{R} car $3x^2 + 2 > 0$ pour tout réel x .

La fonction se présente sous la forme $\ln(u(x))$ où $u(x) = 3x^2 + 2$. Ainsi, la dérivée est donnée par

$$f'(x) = u'(x) \frac{1}{u(x)} = \frac{6}{3x^2 + 2}.$$

De la dérivation d'une fonction composée découle immédiatement la proposition suivante.

Proposition 4.4: Dérivation fonction réciproque

Soit f une bijection strictement monotone et continue dans I dans un intervalle J ; on note sa réciproque f^{-1} .

Si f est dérivable en a et si $f'(a) \neq 0$ alors f^{-1} est dérivable en $b = f(a)$ et :

$$(f^{-1})'(b) = \frac{1}{f' \circ f^{-1}(b)}.$$

Ces résultats peuvent bien sûr s'étendre à n'importe quel réel a d'un intervalle $I \subset \mathbb{R}$.

4.2 Dérivation et extrema

On peut également lier la notion de dérivée à celle d'extrema. Par exemple :

- on dit que f admet un minimum local en a s'il existe un voisinage de a sur lequel la restriction de f admet un minimum en a :

$$\exists \alpha > 0, \forall x \in [a - \alpha, a + \alpha] \cap I, f(x) \geq f(a).$$

- on dit que f admet un maximum local en a s'il existe un voisinage de a sur lequel la restriction de f admet un maximum en a :

$$\exists \alpha > 0, \forall x \in [a - \alpha, a + \alpha] \cap I, f(x) \leq f(a).$$

La recherche de ses minima ou maxima d'une fonction sont des points essentiels en optimisation mais constituent les principaux problèmes que l'on cherchera à résoudre en analyse de données ou encore en apprentissage machine.

Exemple 4.3. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ définie par $f(x) = x^2$. Cette fonction f admet un unique minimum (qui est global), égal à 0 et qui est atteint en $x = 0$.

Il est possible de caractériser l'existence d'un extremum en étudiant la dérivée première d'une fonction, plus précisément en cherchant une valeur pour laquelle elle s'annule :

Proposition 4.5: Caractérisation extremum

Soit une fonction f définie de I dans \mathbb{R} , et soit a un point de I tel que I contienne un segment de la forme $[a - \alpha, a + \alpha]$, avec $\alpha > 0$.

Si f est dérivable en a et admet un extremum local en a , alors $f'(a) = 0$.

Exemple 4.4. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ la fonction définie par

$$f(x) = \frac{1}{1+x^2}$$

Cette fonction admet un extremum qui est atteint en $x = 0$ et qui est égal à 1. En effet, la fonction f est dérivable sur \mathbb{R} et pour tout réel x , nous avons

$$f'(x) = \frac{-2x}{(1+x^2)^2}.$$

La fonction f' s'annule en $x = 0$, la fonction f admet donc un extremum en ce point, qui est un maximum après une étude rapide de la fonction.

Remarque La proposition précédente permet de caractériser un extremum pour une fonction, elle nous dit bien que si la fonction admet un extremum en un point, alors elle s'annule en ce point.

Mais la réciproque n'est pas forcément vraie ! Il faudra donc faire attention et il est parfois nécessaire de procéder à une étude plus fine de la fonction pour savoir si cette dernière admet un extremum.

Exemple 4.5. Considérons la fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ qui à x associe $f(x) = x^3$.

Cette fonction n'admet pas d'extremum bien que sa dérivée, $f' : x \mapsto 3x^2$ s'annule en $x = 0$.

4.3 Etude des accroissements d'une fonction

Les résultats suivants vont nous permettre de justifier les variations d'une fonction telles qu'elles ont été présentées dans le secondaire. Le premier d'entre eux, le Théorème de Rolle, illustré en Figure 15, permet d'établir l'existence d'un point pour lequel la dérivée d'une fonction continue sur un segment s'annule.

Théorème 4.1: Théorème de Rolle

Soit une fonction f définie dans un segment $[a, b]$ à valeurs dans \mathbb{R} , telle que f soit continue sur $[a, b]$ et dérivable sur $]a, b[$. Si $f(a) = f(b)$, alors il existe un réel c tel que $f'(c) = 0$.

Démonstration. On suppose que f est continue sur $[a, b]$, dérivable sur $]a, b[$ et que $f(a) = f(b)$.

Cette démonstration repose sur le fait que si f admet un extremum local en $c \in]a, b[$ et que f est dérivable en c , on peut alors conclure que $f'(c) = 0$.

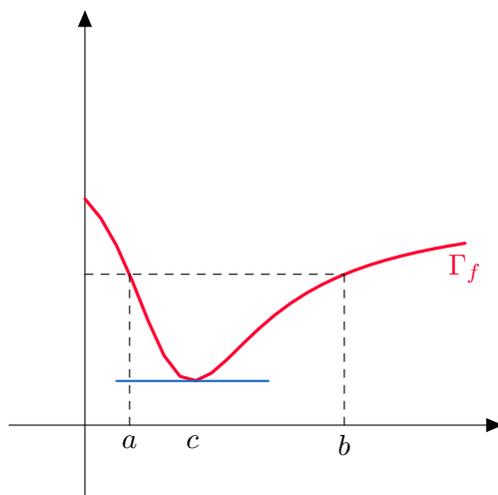


FIGURE 15 – Illustration du Théorème de Rolle

- **Premier cas** : si f est constante, tout point $c \in]a, b[$ est donc un extremum de f , on a donc $f'(c) = 0$ en tout point $c \in]a, b[$.
- **Deuxième cas** : si f n'est pas constante. Comme f est continue sur un segment, elle est donc bornée et atteint ses bornes (Weierstrass!). Il existe donc un minimum m et un maximum M sur $[a, b]$. De plus, f n'étant pas constante, l'une au moins des deux valeurs n'est pas égale à $f(a) = f(b)$, elle l'est alors forcément en une valeur $c \in]a, b[$, où la fonction f admet un extremum, donc $f'(c) = 0$.

□

Le prochain résultat permet de décrire la variation de f entre deux points à l'aide de la valeur d'une de ces dérivées entre ces deux points.

Théorème 4.2: Egalité des accroissements finis

Soit une fonction f définie dans un segment $[a, b]$ dans \mathbb{R} , telle que f soit continue sur $[a, b]$ et dérivable sur $]a, b[$. Il existe $c \in]a, b[$ tel que $f(b) - f(a) = f'(c)(b - a)$.

Ce théorème est illustré en Figure 16

Cette égalité des accroissements finis permet d'obtenir un encadrement de la variation de f entre a et b à l'aide d'un encadrement de f' sur $]a, b[$. C'est ce que l'on appelle l'inégalité des accroissements finis.

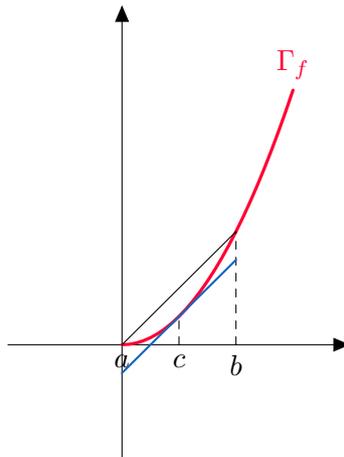


FIGURE 16 – Illustration de l'égalité des accroissements finis

Corollaire 4.1: Inégalité des accroissements finis

Soit f une fonction définie d'un segment $[a, b]$ dans \mathbb{R} , telle que f soit continue sur $[a, b]$ et dérivable sur $]a, b[$.

- si m et M sont des réels tels que pour tout $x \in]a, b[$, $m \leq f'(x) \leq M$, alors on a :

$$m(b - a) \leq f(b) - f(a) \leq M(b - a).$$

- si M est un réel tel pour tout $x \in]a, b[$, $|f'(x)| \leq M$ alors on a

$$|f(b) - f(a)| \leq M|b - a|.$$

De cette même égalité des accroissements finis, on peut obtenir une autre caractérisation des fonctions lipschitziennes !

Corollaire 4.2: Dérivabilité et Lipschitz

Soit f une fonction définie dans un intervalle I de \mathbb{R} et dérivable sur I . Si k est un tel que pour tout $x \in I$, $|f'(x)| \leq k$ alors f est k -lipschitzienne.

Enfin, la dernière application mais la plus importante est l'étude des variations d'une fonction f sur un intervalle I de \mathbb{R} . L'égalité des accroissements finis permet de montrer que

- f est **constante** si et seulement si pour tout $x \in I$, $f'(x) = 0$.
- f est **croissante** si et seulement si pour tout $x \in I$, $f'(x) \geq 0$.

iii) **décroissante** si et seulement si pour tout $x \in I$, $f'(x) \leq 0$.

On souhaite étudier les propriétés de continuité des dérivées d'une fonction.

4.4 Régularité des fonctions réelles

On souhaite maintenant caractériser les fonctions dérivables et dont la ou les dérivées successives sont également continues.

Définition 4.2: Fonction de classe C^1

Soit f une fonction définie de I dans \mathbb{R} et dérivable sur I . Si la fonction f' est continue sur I alors la fonction f est dite *continuellement dérivable* sur I , ou encore de classe C^1 sur I .

On notera alors $C^1(I, \mathbb{R})$ l'ensemble des fonctions définies de I dans \mathbb{R} et de classe C^1 sur I .

Proposition 4.6: Fonction composée

Soient des fonctions $f \in C^1(I, \mathbb{R})$ et $g \in C^1(J, \mathbb{R})$ où J est un intervalle tel que $f(I) \subset J$. Alors la fonction $f \circ g$ est de classe C^1 sur I .

Proposition 4.7: Fonction réciproque

Soit f une bijection de classe C^1 définie de I dans J . Si f' ne s'annule pas sur I , alors f^{-1} est de classe C^1 sur J .

On peut également définir, par une relation de récurrence, les dérivées n -ème d'une fonction f si ces dernières existent. Dit autrement, si f est une fonction de classe C^1 et que sa dérivée f' est dérivable, alors on a $f'' = (f')'$. Et de façon générale, si f est une fonction de classe C^n dont la dérivée n -ième $f^{(n)}$ est dérivable, alors $f^{(n+1)} = (f^{(n)})'$.

Proposition 4.8: Dérivées successives

Soit $n \in \mathbb{N}$ et soient f et g deux fonctions définies sur un intervalle I à valeurs dans \mathbb{R} que l'on suppose de classe C^n sur I , alors

- i) Pour tout réel λ , la fonction λf est de classe C^n sur I et $(\lambda f)^{(n)} = \lambda f^{(n)}$.
- ii) La fonction $f + g$ est aussi de classe C^n sur I et $(f + g)^{(n)} = f^{(n)} + g^{(n)}$.
- iii) La fonction $f \times g$ est de classe C^n sur I et $(f \times g)^{(n)} = \sum_{k=0}^n \binom{n}{k} f^{(k)} \times g^{(n-k)}$.
- iv) Si la fonction g ne s'annule pas sur I , les fonctions $\frac{1}{g}$ et $\frac{f}{g}$ sont de classe C^n sur I .

Démonstration. Les trois premiers points de la proposition se démontrent par récurrence dont les deux premiers découlent directement de la linéarité de la dérivation.

Le troisième point (appelé **formule de Leibniz**) est intéressant à démontrer par récurrence mais très calculatoire : il s'agit de démontrer que la fonction $f \times g$ est de classe C^n sur I et que $(f \times g)^{(n)} = \sum_{k=0}^n \binom{n}{k} f^{(k)} \times g^{(n-k)}$. Pour cela, on partira de la relation

$$(f \times g)^{(n)} = \sum_{k=0}^n \binom{n}{k} (f^{(k)} \times g^{(n-k)})'$$

Il s'agit ensuite de connaître les relations entre les coefficients binomiaux pour arriver au résultat souhaité.

$$\begin{aligned} (f \times g)^{(n)} &= \sum_{k=0}^n \binom{n}{k} (f^{(k)} \times g^{(n-k)})', \\ &= \sum_{k=0}^n \binom{n}{k} (f^{(k+1)} \times g^{(n-k)} + f^{(k)} \times g^{(n+1-k)})', \\ &\quad \downarrow \text{ on sépare notre somme en deux} \\ &= \sum_{k=0}^n \binom{n}{k} (f^{(k+1)} \times g^{(n-k)}) + \sum_{k=0}^n \binom{n}{k} (f^{(k)} \times g^{(n+1-k)}), \\ &\quad \downarrow \text{ on réindexe la première somme en posant } j = k + 1 \\ &= \sum_{j=1}^{n+1} \binom{n}{j-1} (f^{(j)} \times g^{(n+1-j)}) + \sum_{k=0}^n \binom{n}{k} (f^{(k)} \times g^{(n+1-k)}), \\ &\quad \downarrow \text{ on sépare les termes en } j = n + 1 \text{ et } k = 0 \end{aligned}$$

$$\begin{aligned}
&= \binom{n}{n} f^{(n+1)} g^{(0)} + \sum_{j=1}^n \binom{n}{j-1} f^{(j)} g^{(n+1-j)} \\
&\quad + \sum_{k=1}^n \binom{n}{k} f^{(k)} g^{(n+1-k)} + \binom{n}{0} f^{(0)} g^{(n+1)}, \\
&\quad \downarrow \text{on factorise} \\
&= \binom{n+1}{n+1} f^{(n+1)} g^{(0)} + \sum_{k=1}^n \left(\binom{n}{k-1} + \binom{n}{k} \right) f^{(k)} g^{(n+1-k)} \\
&\quad + \binom{n+1}{0} f^{(0)} g^{(n+1)}, \\
&\quad \downarrow \text{propriété des coefficients binomiaux} \\
&= \binom{n+1}{n+1} f^{(n+1)} g^{(0)} + \sum_{k=1}^n \binom{n+1}{k} f^{(k)} g^{(n+1-k)} \\
&\quad + \binom{n+1}{0} f^{(0)} g^{(n+1)}, \\
&\quad \downarrow \text{on rassemble} \\
&= \sum_{k=0}^n \binom{n+1}{k} \left(f^{(k)} \times g^{(n+1-k)} \right).
\end{aligned}$$

□

On peut, de façon similaire démontrer les résultats suivants qui sont identiques à ceux énoncés dans le cas des fonctions une fois dérivables.

Proposition 4.9: Fonction composée

Soit $n \in \mathbb{N}$ et soient des fonctions $f \in C^n(I, \mathbb{R})$ et $g \in C^n(J, \mathbb{R})$ où J est un intervalle tel que $f(I) \subset J$. Alors la fonction $f \circ g$ est de classe C^n sur I .

Proposition 4.10: Fonction réciproque

Soit $n \in \mathbb{N}$ et soit f une bijection de classe C^n définie de I dans J . Si f' ne s'annule pas sur I , alors f^{-1} est de classe C^n sur J .

5 Etude des fonctions convexes

Les fonctions convexes sont d'une importance extrême dans le domaine de la Data Science et notamment en Machine Learning lors de la formulation de certains problèmes.

Elles ont des propriétés intéressantes pour une majeure partie des algorithmes d'optimisation. Plus précisément, lorsque l'on cherche à minimiser une quantité comme une certaine "erreur" que l'on exprime avec une fonction convexe, on est sûr de pouvoir atteindre son minimum, donc de minimiser l'erreur.

5.1 Caractérisations de la convexité

On va, dans cette partie, essentiellement donner des propriétés et comment caractériser des fonctions convexes dans le cas où l'erreur que l'on cherche à minimiser ne dépend que d'un seul paramètre.

Définition 5.1: Fonction convexe

Soit f une fonction définie d'un intervalle I dans \mathbb{R} . Cette fonction f est dite convexe si elle vérifie la propriété suivante :

$$\forall x, y \in \mathbb{R}, \forall \lambda \in [0, 1], f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

Cette définition peut très facilement se représenter graphiquement (voire Figure 17) et signifie que *les cordes d'une fonction se retrouvent au dessus de sa courbe représentative*.

Exemple 5.1. La fonction $x \mapsto x^2$ est une fonction convexe. Pour montrer cela, d'après la définition, il faudrait montrer que pour tout $x, y \in \mathbb{R}$ et pour tout $\lambda \in [0, 1]$, nous avons

$$(\lambda x + (1 - \lambda)y)^2 \leq \lambda x^2 + (1 - \lambda)y^2.$$

Nous allons montrer que la différence est négative

$$\begin{aligned} & (\lambda x + (1 - \lambda)y)^2 - (\lambda x^2 + (1 - \lambda)y^2), \\ &= \lambda^2 x^2 + (1 - \lambda)^2 y^2 + 2\lambda(1 - \lambda)xy - \lambda x^2 - (1 - \lambda)y^2, \\ &= -\lambda(1 - \lambda)x^2 - \lambda(1 - \lambda)y^2 + 2\lambda(1 - \lambda)xy, \\ &= -\lambda(1 - \lambda)(x - y)^2, \\ &\leq 0. \end{aligned}$$

Il existe bien d'autres fonctions dont on pourra admettre la convexité :

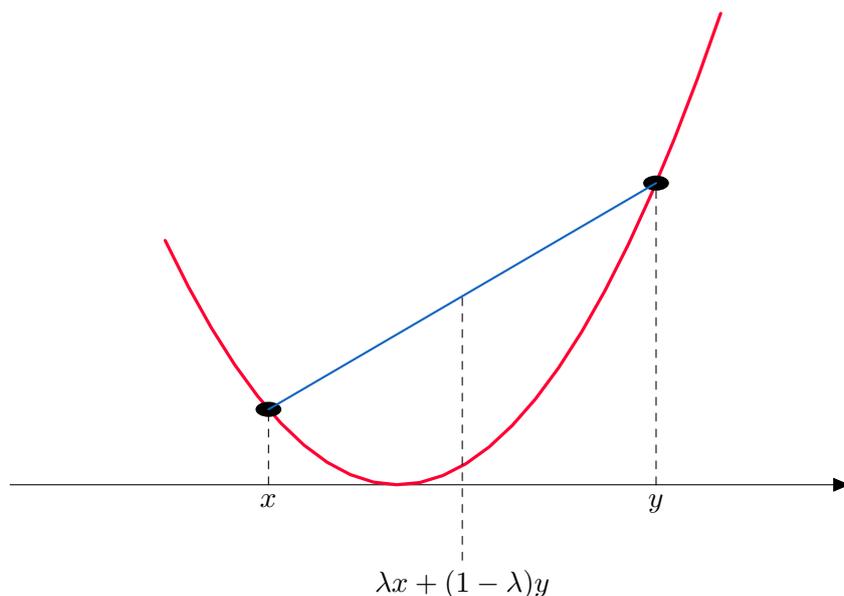


FIGURE 17 – Représentation d’une fonction convexe. On observe que sur le segment $[x, y]$, la courbe représentation de la fonction (en rouge) se trouve en dessous de la corde (en bleue)

- toute les fonctions affines, *i.e.* $x \mapsto ax + b$, $a, b \in \mathbb{R}$
- la fonction exponentielle
- la fonction $x \mapsto x^2$ ou plus généralement $x \mapsto x^{2n}$, $n \in \mathbb{N}$
- la restriction de la fonction $x \mapsto 1/x$ sur l’intervalle $x > 0$

A l’inverse, une fonction f sera dite **concave** si son opposé $-f$ est une fonction **convexe**.

Définition 5.2: Fonction concave

Soit f une fonction définie d’un intervalle I dans \mathbb{R} . Cette fonction f est dite concave si elle vérifie la propriété suivante :

$$\forall x, y \in \mathbb{R}, \forall \lambda \in [0, 1], f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y).$$

Cette définition peut également se représenter graphiquement (voire Figure 18 et signifie que *les cordes d’une fonction se retrouve en dessous de sa courbe représentative*.

Exemple 5.2. *Les fonctions suivantes sont des fonctions concaves sur leur ensemble de définition*

- les fonctions linéaires et affines, donc de la forme $ax + b$
- la fonction $x \mapsto \ln(x)$

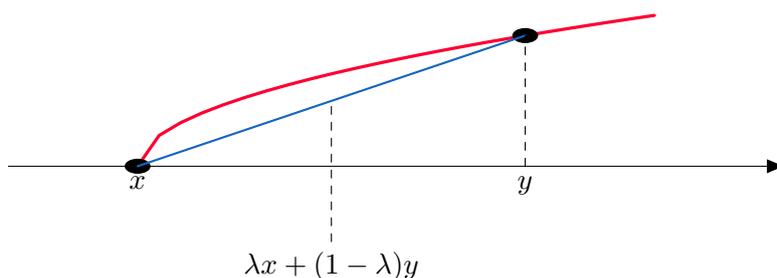


FIGURE 18 – Représentation d’une fonction concave. On observe que sur le segment $[x, y]$, la courbe représentation de la fonction (en rouge) se trouve en dessous de la corde (en bleue)

- la fonction $x \mapsto \sqrt{x}$
- la fonction $x \mapsto -x^{2n}$

Dans la suite, nous étudierons uniquement les fonctions convexes. Les définitions étant analogues pour les fonctions concaves en remplaçant la fonction par son opposé. Lors de l’étude des extrema des fonctions, nous prendrons par contre le soin de distinguer les propriétés liées aux fonctions convexes de celles des fonctions concaves.

La définition de convexité peut également être généralisée à l’utilisation de plusieurs points x_k dans notre intervalle I comme le montre la proposition suivante (que vous retrouverez en probabilités)

Proposition 5.1: Inégalité de Convexité

Soit f une fonction définie de I dans \mathbb{R} telle que f soit convexe sur I . Soit $n \in \mathbb{N}^*$, soit x_1, x_2, \dots, x_n des éléments de I et $\lambda_1, \dots, \lambda_n$ des nombres réels tels que pour tout $k \in \llbracket 1, n \rrbracket$, $\lambda_k \geq 0$ et $\sum_{k=1}^n \lambda_k = 1$, alors

$$f\left(\sum_{k=1}^n \lambda_k x_k\right) \leq \sum_{k=1}^n \lambda_k f(x_k).$$

Cette définition n’est que très peu utilisée pour montrer qu’une fonction est convexe car elle devient trop complexe à manipuler pour la plupart des fonctions, comme la fonction *exponentielle* ou encore la fonction *inverse*.

On va cependant voir qu’ils existent d’autres caractérisations de la convexité qui sont plus simples à utiliser pour montrer la convexité d’une fonction.

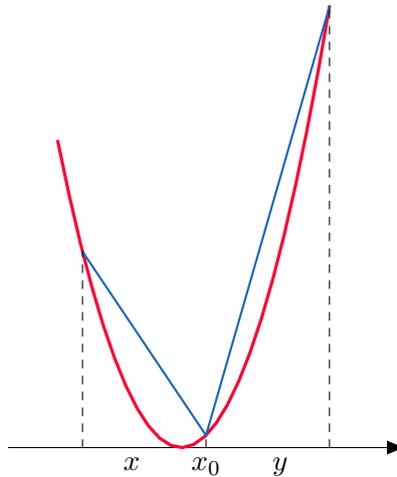


FIGURE 19 – Représentation de la fonction φ_{x_0} de la Proposition 5.2 par les pentes des droites en bleues. On note que la pente en $x < y$ est plus faible que la pente en y .

On commencera par fournir une autre caractérisation de la convexité en étudiant les pentes des sécantes dont une des extrémités est fixée (voir Figure 19).

Proposition 5.2: Caractérisation convexité, ordre 0

Soit f une fonction définie de I dans \mathbb{R} . Alors les propositions suivantes sont équivalentes

- i) La fonction f est convexe.
- ii) Pour tout $x_0 \in I$, la fonction $\varphi_{x_0} : I \rightarrow \mathbb{R}$ par $\varphi_{x_0}(x) = \frac{f(x) - f(x_0)}{x - x_0}$ est croissante.

On peut fournir une deuxième caractérisation des fonctions convexes en étudiant ses tangentes (voir Figure 20), si la fonction étudiée est dérivable.

Proposition 5.3: Caractérisation convexité, ordre 1

Soit f une fonction définie de I dans \mathbb{R} et dérivable sur I . Alors la fonction f est convexe si et seulement si sa fonction dérivée f' est croissante.

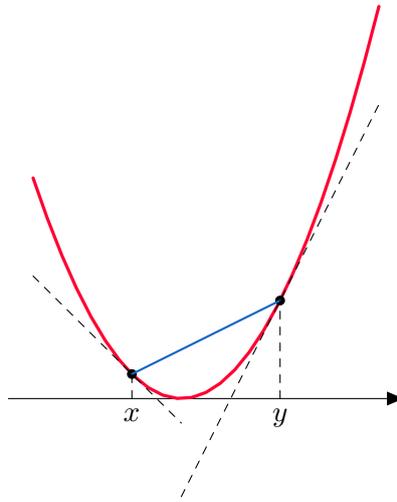


FIGURE 20 – Représentation de tangentes d’une fonction convexe (en noires pointillées). Pour une telle fonction, les tangentes se trouvent toujours sous la courbe représentative de la fonction, contrairement à la corde (en bleue) se trouve au dessus de la courbe représentative de la fonction.

Proposition 5.4: Caractérisation convexité, ordre 1 bis

Soit f une fonction définie de I dans \mathbb{R} et dérivable sur I . Si f est convexe sur I , alors la courbe représentative de f est au dessus de chacune de ses tangentes.

Démonstration. On considère x_0 un élément de I . La tangente à la courbe C_f , représentative de la fonction f , en x_0 a pour équation

$$y = f(x_0) + f'(x_0)(x - x_0).$$

Définissons sur I la fonction s pour tout $x \in I$ par

$$s(x) = f(x) - f(x_0) - f'(x_0)(x - x_0).$$

Cette fonction s représente, pour l’abscisse x donnée, la différence entre la courbe et sa tangente. On va montrer que cette quantité est toujours positive. Or, pour tout $x \in I$, $s'(x) = f'(x) - f'(x_0)$ et, puisque la fonction f est croissante (car f est convexe) on peut alors dresser le tableau de variation suivant

x	x_0		
$s'(x)$	-	0	+
s			

Ainsi $\forall x \in I$, $s(x) \geq 0$, donc la courbe de f est toujours au dessus de sa tangente en x_0 . □

Ce dernier résultat permet de donner une dernière caractérisation de la convexité en utilisant la dérivée seconde de la fonction f et qui exploite directement le lien entre convexité et croissante de la dérivée.

Corollaire 5.1: Caractérisation des fonctions convexes, ordre 2

Soit f une fonction définie d'un intervalle I dans \mathbb{R} et deux fois dérivable sur I . Alors la fonction f est convexe sur I si et seulement si sa fonction dérivée seconde f'' est positive sur I .

C'est cette dernière caractérisation que nous serons essentiellement amenés à utiliser pour montrer la convexité d'une fonction

Exemple 5.3. Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ qui à x associe $\exp(x^2)$.

Cette fonction est deux fois dérivable sur \mathbb{R} et ses dérivées sont données par

$$f'(x) = 2x \exp(x^2) \quad \text{et} \quad f''(x) = (2 + 4x^2) \exp(x^2).$$

La dérivée seconde est positive donc la fonction f est convexe sur \mathbb{R} .

Exemple 5.4. Soit $f : \mathbb{R}_+^* \rightarrow \mathbb{R}$ qui à x associe $\ln(x)$.

Cette fonction est deux fois dérivable sur \mathbb{R}_+^* et ses dérivées sont données par

$$f'(x) = \frac{1}{x} \quad \text{et} \quad f''(x) = -\frac{1}{x^2}.$$

La dérivée seconde est négative donc la fonction f est concave sur \mathbb{R}_+^* .

Exemple 5.5. On considère maintenant la fonction f définie pour tout réel x par

$$f(x) = \frac{1}{3}x^3 - 9x^2 + 4.$$

La fonction f est deux fois dérivable sur \mathbb{R} en tant que polynôme du second degré. Ses dérivées sont données par

$$f'(x) = x^2 - 18x = x(x - 18) \quad \text{et} \quad f''(x) = 2x - 18 = 2(x - 9).$$

La dérivée seconde s'annule en $x = 9$. Elle est négative pour tout $x < 9$, donc la fonction f est concave sur cet intervalle. Elle est positive pour tout $x > 9$, elle est donc convexe sur cet intervalle.

5.2 Lien avec la recherche d'extremum

Dans cette section, on ne va pas énoncer de résultats particuliers, on va simplement spécifier la recherche d'extrema aux cas des fonctions convexes et concaves.

En reprenant ce que nous avons vu dans les sections précédentes, nous pouvons énoncer les résultats suivants concernant les fonctions convexes.

Lemme 5.1: Extrema fonction convexe, C^0

Soit f une fonction convexe sur un intervalle I de \mathbb{R} . Si f admet un minimum local, alors ce minimum est un minimum global.

Exemple 5.6. La fonction $x \mapsto x^2$, est convexe et admet un minimum local en $x = 0$. C'est donc aussi le minimum global de notre fonction.

Lemme 5.2: Extrema fonction convexe, C^1

Soit f une fonction convexe sur un intervalle I de \mathbb{R} que l'on suppose dérivable. Si f admet un point critique, *i.e.* s'il existe $x_0 \in I$ tel que $f'(x_0) = 0$, alors x_0 est point en lequel la fonction f atteint son minimum.

Exemple 5.7. La fonction f , définie de \mathbb{R} dans \mathbb{R} , par $f(x) = x^4 + 4x^2 - 6$ est convexe sur \mathbb{R} . Elle est également une fois dérivable et sa dérivée est donnée par $f'(x) = 4x^3 + 8x$ qui s'annule en $x = 0$.

La fonction f admet donc un minimum en $x = 0$.

Remarquez que le fait d'être convexe n'implique pas nécessairement l'existence d'un minimum. Il ne faut pas oublier l'existence du point critique pour garantir l'existence du minimum, comme le montre l'exemple de la fonction exponentielle qui est bien convexe mais qui n'admet pas de minimum.

Il n'est donc pas forcément nécessaire de passer par le tableau de variations pour savoir si un point critique est un minimum ou non, l'étude de la convexité est suffisante.

Pour les fonctions concaves, nous avons les résultats analogues suivants.

Lemme 5.3: Extrema fonction concave, C^0

Soit f une fonction concave sur un intervalle I de \mathbb{R} . Si f admet un maximum local, alors ce maximum est un maximum global.

Lemme 5.4: Extrema fonction concave, C^1

Soit f une fonction concave sur un intervalle I de \mathbb{R} que l'on suppose dérivable. Si f admet un point critique, *i.e.* s'il existe $x_0 \in I$ tel que $f'(x_0) = 0$, alors x_0 est point en lequel la fonction f atteint son maximum.

6 Intégration

L'objectif de cette dernière section n'est pas de présenter la construction de l'intégrale à l'aide des fonctions continues par morceaux, mais simplement de donner quelques outils permettant leurs calculs : comme **l'intégration par parties** ou encore **le changement de variables**.

6.1 Généralités sur l'intégrale

Le calcul d'intégrale peut se voir comme l'opération complémentaire au calcul de dérivée.

Si la dérivée permettait de nous renseigner quant à l'évolution d'une fonction, *i.e.* sa monotonie, l'intégrale de cette dernière nous donnera une évaluation de cette fonction, à savoir, l'aire comprise entre la courbe représentative de la fonction et l'axe des abscisses.

Par exemple, en Figure 21 ; on cherchera à évaluer l'intégrale de la fonction de la fonction f , définie sur l'intervalle $[-2, 4]$ par

$$f(x) = (x + 2)\sqrt{3 + x} - x^2,$$

dont on cherchera la valeur de l'intégrale sur le segment $[-1, 3]$.

Cette intégrale n'est pas calculer de façon directe par nos ordinateurs, elle est simplement approchée en créant des subdivisions de l'intervalle sur lequel on cherche à calculer la valeur de notre intégrale. A l'aide de cette subdivision, on définit une fonction f qui sera continue par morceaux (voire même constante par morceaux) permettant la définition de *rectangles* qui vont servir à approcher la valeur de notre intégrale, comme illustré en Figure 22.

Plus la subdivision sera fine, meilleure sera l'approximation de notre intégrale. C'est ce principe qui est appliqué par nos ordinateurs pour le calcul d'une intégrale. La valeur retournée n'est qu'une approximation de la valeur recherchée à l'aide de ce principe, que l'on étudiera en TP.

Bien évidemment, les outils mathématiques vont nous permettre de calculer la valeur de la plupart des intégrales que l'on rencontrera par la suite, à l'aide du calcul de *primitives*.

6.2 Calculs de primitives et intégrales

On présente maintenant la notion de primitives et quelques outils de calculs de primitives et donc d'intégrales.

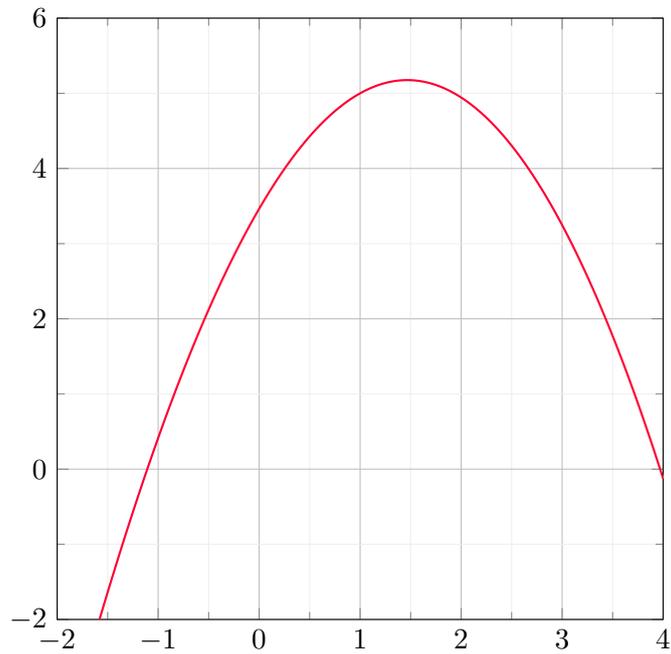


FIGURE 21 – Représentation de la fonction continue $f : x \mapsto (x + 2)\sqrt{3 + x - x^2}$ dont on cherche à évaluer l'intégrale sur le segment $[-1, 3]$

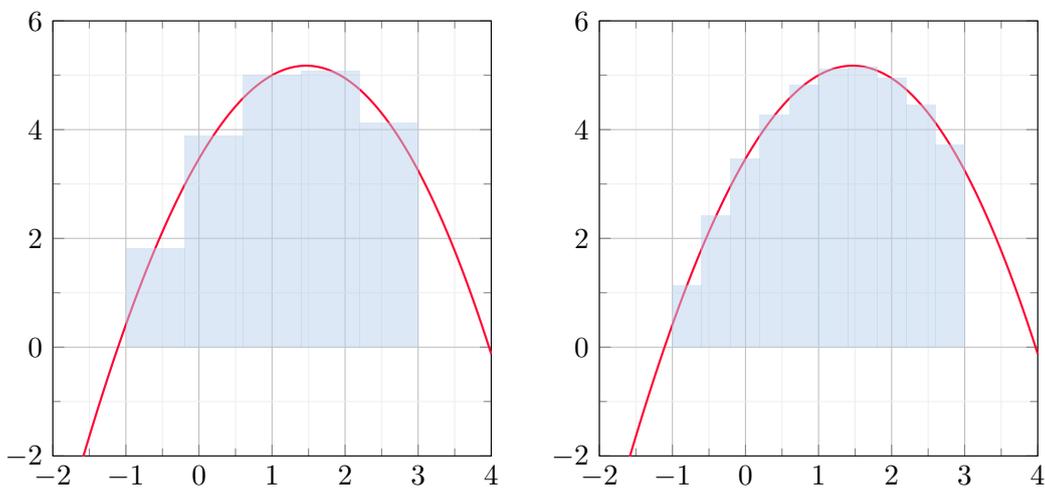


FIGURE 22 – Calcul approché de l'intégrale à l'aide de fonctions continues par morceaux, ici des fonctions dites *étagées* car constantes par morceaux.

Primitives La notion de primitive, notée F , est primordiale pour le calcul d'intégrale d'une fonction f continue sur un intervalle I de \mathbb{R} . C'est, en effet, en évaluant les valeurs de cette primitive en les bornes de l'intervalle de I que l'on pourra déterminer la valeur de l'intégrale de notre fonction f sur I .

Définition 6.1: Primitive

Soit f une fonction réelle définie et continue sur un intervalle I de \mathbb{R} . On appelle *primitive* de f sur I , toute fonction F définie et dérivable sur I telle que $F' = f$

Cette définition montre que la primitive d'une fonction f n'est pas unique! En effet, si F est une primitive de f , alors $\forall k \in \mathbb{R}$ la fonction $F + k$ est aussi une primitive de f .

Démonstration. Soit F une primitive de f , et considérons la fonction une autre primitive de la fonction f dérivable sur I .

Alors on a

$$G' = f \quad \text{et} \quad F' = f \quad F' - G' = 0.$$

Ce qui veut donc dire qu'il existe une constante $k \in \mathbb{R}$ telle que $G = F + k$. \square

Exemple 6.1. Soit f la fonction définie sur \mathbb{R} par $f(x) = 3x^2$.

Alors la fonction F définie sur \mathbb{R} par $F(x) = x^3 + k$, $k \in \mathbb{R}$ est une primitive de f

En effet, pour tout réel x , nous avons $F'(x) = 3x^2 = f(x)$.

La proposition suivante permet de définir une primitive d'une fonction f de façon unique.

Proposition 6.1: Primitive qui s'annule en un point

Soit f une fonction réelle définie et continue sur un intervalle I de \mathbb{R} , et soit a un point de I .

Alors la fonction F définie sur I par

$$F(x) = \int_a^x (t) dt$$

est l'unique primitive de f qui s'annule en a .

Démonstration. Si l'existence d'une telle primitive est difficile à démontrer, l'unicité est simple à prouver.

En effet, si f admet une autre primitive G qui s'annule également en a , alors il existe une constante $k \in \mathbb{R}$ telle que $G = F + k$. Ainsi

$$0 = G(a) = F(a) + k = k$$

Donc $F = G$. □

De ce résultat découle directement les deux corollaires suivants, dont le premier permet de voir une intégrale comme la différence des valeurs de la primitive en les bornes de l'intervalle d'intégration.

Corollaire 6.1: Un premier calcul d'intégrale

Soit f une fonction réelle définie et continue sur I , et soit F une primitive de f sur I alors :

$$\forall a, b \in I, \int_a^b f(t) dt = F(b) - F(a).$$

Soient $a, b \in I$ et soit G définie sur I par

$$G(x) = \int_a^x f(t) dt.$$

G est une primitive de f donc il existe une constante $k \in \mathbb{R}$ telle que $G = F + k$, or $0 = G(a) = F(a) + k$, donc $G(a) = -F(a)$.

On a donc

$$\int_a^b f(t) dt = G(b) = F(b) + k = F(b) - F(a).$$

En particulier, si on suppose maintenant que la fonction f est de classe C^1 , nous avons le résultat suivant

Corollaire 6.2: Lien avec la dérivée

Soit f une fonction de classe C^1 sur un intervalle I de \mathbb{R} , alors

$$\forall a, b \in \mathbb{R}, f(b) - f(a) = \int_a^b f'(t) dt.$$

Regardons maintenant quelques méthodes de calculs.

Méthodes de calculs Pour déterminer des primitives d'une fonction donnée, il sera bon de connaître toutes les formules de dérivation, mais dans "l'autre sens" pour les fonctions les plus classiques.

Dans des cas plus complexes, deux outils s'offrent à nous pour le calcul pratique d'intégrales et de primitives : *l'intégration par parties* et *le changement de variables*.

Proposition 6.2: S

ient u et v deux fonctions de classe C^1 sur un intervalle I de \mathbb{R} , alors

$$\int_a^b u(t)v'(t) dt = [u(t)v(t)]_a^b - \int_a^b u'(t)v(t) dt.$$

Démonstration. La démonstration repose sur la formule de dérivation d'un produit et ensuite l'intégration des différents termes.

Soient u et v des fonctions de classe C^1 sur I et $(uv)' = u'v + uv'$, donc d'après le corollaire précédent, nous avons

$$[u(t)v(t)]_a^b = \int_a^b (uv)'(t) dt = \int_a^b u'(t)v(t) dt + \int_a^b u(t)v'(t) dt,$$

Ainsi, en réarrangeant les termes :

$$\int_a^b u(t)v'(t) dt = [u(t)v(t)]_a^b - \int_a^b u'(t)v(t) dt.$$

□

Regardons deux exemples d'applications, dont le premier est le calcul de la primitive de la fonction $x \mapsto \ln(x)$.

Exemple 6.2. *Considérons la fonction $f : x \mapsto \ln(x)$ définie et continue sur \mathbb{R}_+^* . Utilisons la formule de l'intégration par parties pour déterminer la primitive de cette fonction qui s'annule en 1*

$$\begin{aligned} \int_1^x \ln(t) dt &= \int_1^x 1 \ln(t) dt, \\ &\downarrow \text{On considère ensuite } u'(t) = 1 \text{ et } v(t) = \ln(t) \\ &= [t \ln(t)]_1^x - \int_1^x t \times \frac{1}{t} dt, \\ &= x \ln(x) - \int_1^x 1 dt, \end{aligned}$$

$$\begin{aligned}
&= x \ln(x) - [t]_1^x dt, \\
&= x \ln(x) - x + 1.
\end{aligned}$$

Ainsi, une primitive de la fonction $x \mapsto \ln(x)$ est donnée par la fonction $x \mapsto x \ln(x) - x$.

Cette proposition peut également servir plus directement au calcul d'intégrale

Exemple 6.3. *Considérons la fonction f définie sur l'intervalle $I = [0, e-1]$ par $f(x) = \frac{x}{(x+1)^2}$ et déterminons son intégrale sur I .*

$$\begin{aligned}
\int_0^{e-1} \frac{t}{(t+1)^2} dt &= \int_0^{e-1} t \times \frac{1}{(t+1)^2} dt, \\
&\downarrow \text{ On considère ensuite } u'(t) = \frac{1}{(t+1)^2} \text{ et } v(t) = t \\
&= \left[-\frac{t}{t+1} \right]_0^{e-1} - \int_0^{e-1} -\frac{1}{t+1} dt, \\
&= -\left(\frac{e-1}{e} - 0 \right) - [\ln(t+1)]_0^{e-1}, \\
&= -1 + \frac{1}{e} + 1, \\
&= \frac{1}{e}.
\end{aligned}$$

On la retrouvera très souvent dans des calculs en probabilités, notamment pour le calcul d'espérances de variances qui nécessiteront de calculer des quantités de la forme

$$\int t f(t) dt \quad \text{ou encore} \quad \int t^2 f(t) dt,$$

où f sera une fonction que l'on appellera densité de probabilité. Par exemple, pour calculer la moyenne d'une variable aléatoire X distribuée selon une loi normale de densité

$$f(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(t-\mu)^2\right),$$

nous devons effectuer le calcul d'intégrale suivant

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} t f(t) dt = \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(t-\mu)^2\right) dt.$$

Pour évaluer cette intégrale, nous aurons besoin de la formule d'intégration parties mais aussi de la formule de changement de variable qui fait l'objet de la proposition suivante

Proposition 6.3: Changement de variable

Soit f une fonction réelle définie sur un intervalle I de \mathbb{R} , et soit φ une fonction de classe C^1 définie sur un segment $[a, b]$ et à valeurs dans I , alors

$$\int_a^b f(\varphi(t))\varphi'(t) dt = \int_{\varphi(a)}^{\varphi(b)} f(u) du.$$

Démonstration. Définissons une fonction F sur l'intervalle I par $F(x) = \int_{\varphi(a)}^x f(t) dt$.

F est alors la primitive de f qui s'annule $\varphi(a)$. F est de classe C^1 sur $[a, b]$ et $(F \circ \varphi)' = \varphi'(F' \circ \varphi)$.

Ainsi

$$\int_{\varphi(a)}^{\varphi(b)} f(u) du = F(\varphi(b)) - F(\varphi(a)) = \int_a^b (F \circ \varphi)'(t) dt = \int_a^b f(\varphi(t))\varphi'(t) dt.$$

□

Un cas intéressant de cette fonction, notamment pour la pratique, est le cas où la fonction φ est bijective, donc strictement monotone sur son intervalle de définition.

En posant, dans la proposition précédente, $\alpha = \varphi(a)$ et $\beta = \varphi(b)$, on a alors

$$\int_{\alpha}^{\beta} f(u) du = \int_{\varphi^{-1}(\alpha)}^{\varphi^{-1}(\beta)} f(\varphi(t))\varphi'(t) dt = \int_a^b f(\varphi(t))\varphi'(t) dt.$$

Cela revient à effectuer le changement de variable $u = \varphi(t)$.

Prenons un cas d'usage pour rendre cela plus clair.

Exemple 6.4. *Considérons la fonction $f : x \mapsto \cos(\sqrt{x})$ définie et continue sur l'intervalle $[1, 4]$. Déterminons la valeur de l'intégrale de cette fonction sur cet intervalle.*

Cela est de suite moins évident, la fonction $x \mapsto \sqrt{x}$ nous empêche de déterminer directement une primitive de notre fonction. Qu'à cela ne tienne, nous n'avons qu'à appliquer le changement de variable $u = \sqrt{x}$, qui est une fonction bijective de $[1, 4]$ dans $[1, 2]$

Il faut donc penser à changer les bornes de notre intégrale et à la fonction que l'on intègre.

Les bornes : *lorsque x est égal 1, u prend la valeur 1. De la même façon, lorsque $x = 4$, u prend la valeur 2.*

La fonction intégrer : notre variable initiale est la variable x , nous devons donc intégrer avec l'opérateur " dx ", mais notre nouvelle variable devient la fonction u , d'où

$$u(x) = \sqrt{x} \implies u'(x) = \frac{du}{dx} = \frac{\sqrt{x}}{dx} = (\sqrt{x})' = \frac{1}{2\sqrt{x}} = \frac{1}{2u} \iff \frac{du}{dx} = \frac{1}{2u}.$$

Ainsi l'opérateur " dx " devient $2udu$.

C'est une façon plutôt physicienne d'écrire le changement de variables, mais elle permet de limiter les erreurs liées à la formule de changement de variable.

En combinant tout cela on a donc

$$\int_1^4 \cos(\sqrt{x}) \, dx = \int_1^2 \cos(u) 2udu = \int_1^2 2u \cos(u) \, du.$$

On applique une intégration par parties pour conclure ce calcul.

$$\int_1^2 2u \cos(u) \, du = [2u \sin(u)]_1^2 \int_1^2 2 \sin(u) \, du = 2(2 \sin(2) - \sin(1) + \cos(2) - \cos(1)).$$

En guise d'entraînement, le lecteur est invité à calculer l'intégrale suivante :

$$\int_{-\infty}^{+\infty} t f(t) \, dt = \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(t - \mu)^2\right) \, dt,$$

en effectuant le changement de variable linéaire $u = t - \mu$ (qui est bien un changement de variable bijectif de classe C^1 de \mathbb{R} dans \mathbb{R}) et en utilisant le résultat suivant

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x)^2\right) \, dx = 1.$$

Ce dernier résultat pourra être démontré l'année prochaine, à l'aide d'un changement de variables (en coordonnées polaires) et à l'aide d'une *intégrale double*.

6.3 Quelques propriétés de l'intégrale

On termine cette présentation de l'intégrale par quelques propriétés ou résultats généraux.

Une des premières chose à retenir est la relation suivante

$$\int_a^b f(t) dt = - \int_b^a f(t) dt.$$

Ainsi, parcourir l'intégrale dans un sens ou dans l'autre change le signe de la valeur de cet intégrale. C'est une conséquence directe des propositions de la section précédente.

Proposition 6.4: Linéarité

Soient f et g deux fonctions réelles définies sur un intervalle I de \mathbb{R} et soient deux réels λ et μ et a, b deux points de I .

Alors

$$\int_a^b (\lambda f + \mu g)(t) dt = \lambda \int_a^b f(t) dt + \int_a^b g(t) dt.$$

La notion de *linéarité* sera pleinement étudiée l'année prochaine, mais pour résumer, concernant l'intégrale, on dit simplement que *l'intégrale d'une somme de deux fonctions est égale à la somme des intégrales de ces deux fonctions*.

Proposition 6.5: Relation de Chasles

Soit f une fonction définie et continue sur un intervalle I de \mathbb{R} et soient $a < b < c$ trois nombres réels dans I , alors :

$$\int_a^c f(t) dt = \int_a^b f(t) dt + \int_b^c f(t) dt$$

Cela peut être utile de connaître cette relation pour certains calculs d'intégrales

Exemple 6.5. *Supposons que l'on souhaite calculer la valeur de l'intégrale de la fonction $x \mapsto \sin(x)$ sur l'intervalle $[-\pi, \pi]$, alors*

$$\begin{aligned} \int_{-\pi}^{\pi} \sin(t) dt &= \int_{-\pi}^0 \sin(t) dt + \int_0^{\pi} \sin(t) dt, \\ &\downarrow \text{ en posant le changement } u = -t \text{ dans la première intégrale} \\ &= \int_{\pi}^0 -\sin(-u) du + \int_0^{\pi} \sin(t) dt, \\ &\downarrow \text{ pour tout } x, \sin(-x) = -\sin(x) \\ &= \int_{\pi}^0 \sin(u) du + \int_0^{\pi} \sin(t) dt, \end{aligned}$$

$$\begin{aligned} &= - \int_0^\pi \sin(u) \, du + \int_0^\pi \sin(t) \, dt, \\ &= 0. \end{aligned}$$

Les deux dernières propositions permettent d'établir quelques inégalités simples sur la valeur d'une intégrale qui peuvent s'avérer utiles pour démontrer d'autres résultats dans de futurs cours.

Proposition 6.6: Inégalité de la moyenne

Soit f une fonction réelle définie sur un intervalle I de \mathbb{R} et soient a et b deux points de I , alors

$$\left| \int_a^b f(t) \, dt \right| \leq \sup_{x \in I} |f(x)| |b - a|$$

Proposition 6.7: Une dernière inégalité

Soit f une fonction réelle définie sur un intervalle I de \mathbb{R} et soient a et b deux points de I , alors :

$$\left| \int_a^b f(t) \, dt \right| \leq \int_a^b |f(t)| \, dt.$$

7 Développements Limités

Cette courte section est une introduction aux développements limités d'une fonction d'une variable réelle.

Les développements limités sont des approximations des fonctions à l'aide de polynômes qui permettent, dans certaines situations, de simplifier des analyses de problèmes complexes. Cela nous permettra également d'introduire le *Développement de Taylor* d'une fonction que l'on sera amené à réutiliser lorsque l'on présentera des algorithmes d'optimisation de fonctions convexes.

7.1 Formule de Taylor

La Formule de Taylor² permettent d'effectuer une étude locale d'une fonction f au voisinage d'un point a à l'aide de ses dérivées successives en ce point.

Présentons, sans plus tarder, la *Formule de Taylor-Young*

Proposition 7.1: Formule de Taylor-Young

Soit n un entier naturel, soit f une fonction réelle définie et de classe C^n sur I , un intervalle de \mathbb{R} et $a \in I$.

Alors il existe une fonction h définie sur I et vérifiant $\lim_{x \rightarrow a} h(x) = 0$ telle que :

$$\forall x \in I, f(x) = \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k + (x-a)^n h(x),$$

où $f^{(k)}$ désigne la dérivée k -ème de la fonction f .

Cette proposition présente ce que l'on appelle une forme de développement limité à l'ordre n d'une fonction. Bien évidemment, en pratique, nous n'irons jamais jusque là et on se contentera principalement d'effectuer des développements limités à l'ordre 1 ou 2 pour les problèmes étudiés.

7.2 Développements limités

Regardons une définition plus formelle de ce qu'est un développement limité et faisons ensuite le lien avec le développement de Taylor.

2. Il serait plus exact de dire les Formules de Taylor car de nombreuses formulations existent selon la qualité selon ce que l'on souhaite tirer comme information de notre fonction. Nous présentons ici la formule dite de *Taylor-Young* mais le lecteur curieux pourra également regarder les formules de *Taylor-Lagrange* ou de *Taylor avec reste intégral*.

Considérons tout d'abord une fonction f définie sur un intervalle I de \mathbb{R} et un point a de I .

Définition 7.1: Développement limité

Soit $n \in \mathbb{N}$. On dit que la fonction f admet un développement limité à l'ordre n au voisinage de a s'il existe des nombres réels $a_0, a_1, a_2, \dots, a_n$ et une fonction ε définie sur I et vérifiant $\lim_{x \rightarrow a} \varepsilon(x) = 0$, tels que

$$\forall x \in I, f(x) = \sum_{k=0}^n a_k (x-a)^k + (x-a)^n \varepsilon(x).$$

Dans le cas particulier où $a = 0$, on obtiendra un développement limité de la fonction au voisinage de 0.

Cela est utile afin de pouvoir également fournir des développements limités de fonction au voisinage de l'infini, ce que l'on appelle aussi un *développement asymptotique*.

Développement asymptotique On considère une fonction f définie sur un intervalle de la forme $[\alpha, +\infty[$. Déterminer un développement asymptotique de la fonction f consiste à écrire f sous la forme

$$f(x) = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \dots + \frac{a_n}{x^n} + o\left(\frac{1}{x^n}\right).$$

Pour obtenir un tel développement, on définit sur l'intervalle $\left]0, \frac{1}{\alpha}\right]$, la fonction $g : t \mapsto f\left(\frac{1}{t}\right)$ définissant ainsi la fonction g au voisinage de 0.

Ainsi, si $g(t) = a_0 + a_1 t + \dots + a_n t^n + t^n \varepsilon(t)$ avec $\lim_{t \rightarrow 0} \varepsilon(t) = 0$, alors :

$$f(x) = g(1/x) = a_0 + \frac{a_1}{x} + \frac{a_2}{x^2} + \dots + \frac{a_n}{x^n} + \frac{1}{x^n} \varepsilon\left(\frac{1}{x}\right), \quad \text{où} \quad \lim_{x \rightarrow +\infty} \varepsilon\left(\frac{1}{x}\right) = 0$$

Formule de Taylor

Proposition 7.2: DL à l'ordre n

Soit n un entier naturel et soit f une fonction de classe C^n sur un intervalle I de \mathbb{R} . Alors f admet un développement limité à l'ordre n au voisinage de tout point a de I et ce développement est :

$$\forall x \in I, \quad f(x) = \sum_{k=0}^n a_k (x-a)^k + o((x-a)^n).$$

En réalité, le lecteur a déjà effectué plusieurs développement limité de fonctions au voisinage d'un point ... lorsqu'il a déterminé l'équation de la tangente à une fonction en un point !

Nous terminons cette présentation des développements limités (même s'il reste plein de propriétés et d'opérations sur les développements limités que nous pourrions montrer) en présentant quelques développements limités de références.

Lemme 7.1: Quelques DL usuels

Soit n un entier naturel. On a alors les développements limités suivants au voisinage de 0

i) $\frac{1}{1-x} = \sum_{k=0}^n x^k + o(x^n).$

ii) $e^x = \sum_{k=0}^n \frac{x^k}{k!} + o(x^n).$

iii) $\ln(1+x) = \sum_{k=1}^n (-1)^{k-1} \frac{x^k}{k} + o(x^n).$

iv) $\sin(x) = \sum_{k=0}^n (-1)^k \frac{x^{2k+1}}{(2k+1)!} + o(x^{2n+2}).$

v) $\cos(x) = \sum_{k=0}^n (-1)^k \frac{x^{2k}}{(2k)!} + o(x^{2n+1}).$

vi) $\operatorname{sh}(x) = \sum_{k=0}^n \frac{x^{2k+1}}{(2k+1)!} + o(x^{2n+2}).$

vii) $\operatorname{ch}(x) = \sum_{k=0}^n \frac{x^{2k}}{(2k)!} + o(x^{2n+1}).$

viii) Soit $\alpha \in \mathbb{R}$, alors

$$(1+x)^\alpha = \sum_{k=0}^n \frac{\prod_{j=0}^{k-1} (\alpha-j)}{k!} x^k + o(x^n).$$

L'obtention de ces derniers est une simple application de la Proposition précédente.

8 Formulaire

Cette section rassemble les quelques formules usuelles que le lecteur pourra être amené à utiliser au cours de sa formation.

8.1 Dérivées des fonctions usuelles

Fonction	Ensemble def.	Dérivée	Ensemble dér.
$a \in \mathbb{R}$	\mathbb{R}	0	\mathbb{R}
$ax + b, a, b \in \mathbb{R}$	\mathbb{R}	a	\mathbb{R}
$x^n, n \in \mathbb{N}$	\mathbb{R}	nx^{n-1}	\mathbb{R}
$1/x$	\mathbb{R}	nx^{n-1}	\mathbb{R}
$x^{-n}, n \in \mathbb{N}$	\mathbb{R}^*	$-nx^{-(n+1)}$	\mathbb{R}^*
\sqrt{x}	\mathbb{R}_+	$\frac{1}{2\sqrt{x}}$	\mathbb{R}_+^*
$x^\alpha, \alpha \in \mathbb{R}$	\mathbb{R}_+^*	$\alpha x^{\alpha-1}$	\mathbb{R}_+^*
e^x	\mathbb{R}	e^x	\mathbb{R}
$\ln(x)$	\mathbb{R}_+^*	$1/x$	\mathbb{R}_+^*
$\cos(x)$	\mathbb{R}	$-\sin(x)$	\mathbb{R}
$\sin(x)$	\mathbb{R}	$\cos(x)$	\mathbb{R}
$\tan(x)$	$\mathbb{R} \setminus \left\{ \frac{\pi}{2} + k\pi \right\}, k \in \mathbb{Z}$	$1 + \tan^2(x) = \frac{1}{\cos^2(x)}$	$\mathbb{R} \setminus \left\{ \frac{\pi}{2} + k\pi \right\}, k \in \mathbb{Z}$
$\text{ch}(x)$	\mathbb{R}	$\text{sh}(x)$	\mathbb{R}
$\text{sh}(x)$	\mathbb{R}	$\text{ch}(x)$	\mathbb{R}
$\text{th}(x)$	\mathbb{R}	$1 - \text{th}^2(x) = \frac{1}{\text{ch}^2(x)}$	\mathbb{R}

8.2 Formules de dérivations générales

Dans cette section, on supposera que u et v sont des fonctions dérivables sur \mathbb{R} . On supposera également que u^{-1} , la fonction réciproque de u est aussi dérivable sur \mathbb{R} . Seuls quelques exemples sur la dérivation de composée de fonctions sont données avec quelques fonctions usuelles, car il s'agit essentiellement de combiner la dérivée de la composition de fonction avec la dérivée de la fonction usuelle.

Fonction	Dérivée	Ensemble der.
$u(x)^2$	$2u'(x)u(x)$	\mathbb{R}
$u(x)^n, n \in \mathbb{N}$	$nu'(x)u(x)^{n-1}$	\mathbb{R}
$\sqrt{u(x)}$	$\frac{u'(x)}{2\sqrt{u(x)}}$	$x \in \mathbb{R} \mid u(x) > 0$
$u(x)^\alpha, \alpha \in \mathbb{R}$	$\alpha u'(x)u(x)^{\alpha-1}$	$x \in \mathbb{R} \mid u(x) > 0$
$\frac{1}{u(x)}$	$-\frac{u'(x)}{u(x)^2}$	$x \in \mathbb{R} \mid u(x) \neq 0$
$\frac{u(x)}{v(x)}$	$\frac{u'(x)v(x) - v'(x)u(x)}{v(x)^2}$	$x \in \mathbb{R} \mid v(x) \neq 0$
$u(x)v(x)$	$u'(x)v(x) + u(x)v'(x)$	\mathbb{R}
$u \circ v(x)$	$v'(x)u'(v(x))$	\mathbb{R}
$u^{-1}(x)$	$\frac{1}{u'(u^{-1}(x))}$	$x \in \mathbb{R} \mid u'(u^{-1}(x)) \neq 0$
$e^{u(x)}$	$u'(x)e^{u(x)}$	\mathbb{R}
$\ln(u(x))$	$\frac{u'(x)}{u(x)}$	$x \in \mathbb{R} \mid u(x) > 0$
$\cos(u(x))$	$-u'(x)\sin(u(x))$	\mathbb{R}
$\sin(u(x))$	$u'(x)\cos(u(x))$	\mathbb{R}
$\text{ch}(u(x))$	$u'(x)\text{sh}(u(x))$	\mathbb{R}
$\text{sh}(u(x))$	$u'(x)\text{ch}(u(x))$	\mathbb{R}

8.3 Développements limités

Les développements limités sont donnés à l'ordre n pour les fonctions classiques ou encore $2n + 1$ et $2n + 2$ pour les fonctions circulaires et hyperboliques.

Fonction	Développement limité
$\frac{1}{1-x}$	$1 + x + x^2 + x^3 + \dots + x^n + o(x^n)$
$\frac{1}{1+x}$	$1 - x + x^2 - x^3 + \dots + (-1)^n x^n + o(x^n)$
$(1+x)^\alpha$	$1 + \alpha x + \frac{\alpha(\alpha-1)}{2!} x^2 + \dots + \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} x^n + o(x^n)$
$\ln(1+x)$	$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots + (-1)^{n+1} \frac{x^n}{n} + o(x^n)$
$\ln(1-x)$	$-x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} + \dots - \frac{x^n}{n} + o(x^n)$
e^x	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots + \frac{x^n}{n!} + o(x^n)$
$\cos(x)$	$1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \dots + (-1)^n \frac{x^{2n}}{2n!} + o(x^{2n+1})$
$\sin(x)$	$x - \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+2})$
$\text{ch}(x)$	$1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \dots + \frac{x^{2n}}{2n!} + o(x^{2n+1})$
$\text{sh}(x)$	$x + \frac{x^3}{3!} + \frac{x^5}{5!} + \dots + \frac{x^{2n+1}}{(2n+1)!} + o(x^{2n+2})$

8.4 Primitives des fonctions usuelles

Fonction	Primitive	Ensemble def.
$a \in \mathbb{R}$	$ax + C, C \in \mathbb{R}$	\mathbb{R}
$ax + b, a, b \in \mathbb{R}$	$\frac{1}{2}ax^2 + cx + C, C \in \mathbb{R}$	\mathbb{R}
$x^n, n \in \mathbb{N}$	$\frac{1}{n+1}x^{n+1} + C, C \in \mathbb{R}$	\mathbb{R}
$1/x$	$\ln(t)$	\mathbb{R}^*
\sqrt{x}	$\frac{2}{3}x^{3/2} + C, C \in \mathbb{R}$	\mathbb{R}
$x^\alpha, \alpha \in \mathbb{R} \setminus \{-1\}$	$\frac{x^{\alpha+1}}{\alpha+1} + C, C \in \mathbb{R}$	\mathbb{R}
$a^x, a \in \mathbb{R}^+ \setminus \{1\}$	$\frac{a^x}{\ln(a)} + C, C \in \mathbb{R}, a \in \mathbb{R}_+^* \setminus \{1\}$	\mathbb{R}
e^x	$e^x + C, C \in \mathbb{R}$	\mathbb{R}
$\ln(x)$	$x \ln(x) - x + C, C \in \mathbb{R}$	\mathbb{R}^*
$\cos(x)$	$\sin(x) + C, C \in \mathbb{R}$	\mathbb{R}
$\sin(x)$	$-\cos(x) + C, C \in \mathbb{R}$	\mathbb{R}
$\tan(x)$	$-\ln(\cos(t)) + C, C \in \mathbb{R}$	$\mathbb{R} \setminus \left\{ \frac{\pi}{2} + k\pi \right\}, k \in \mathbb{Z}$
$\text{ch}(x)$	$\text{sh}(t) + C, C \in \mathbb{R}$	\mathbb{R}
$\text{sh}(x)$	$\text{ch}(t) + C, C \in \mathbb{R}$	\mathbb{R}
$\text{th}(x)$	$\ln(\text{ch}(t)) + C, C \in \mathbb{R}$	\mathbb{R}
$\frac{1}{a^2 + x^2}, a \in \mathbb{R}^*$	$\frac{1}{a} \arctan\left(\frac{x}{a}\right) + C, C \in \mathbb{R}, a \in \mathbb{R}^*$	\mathbb{R}

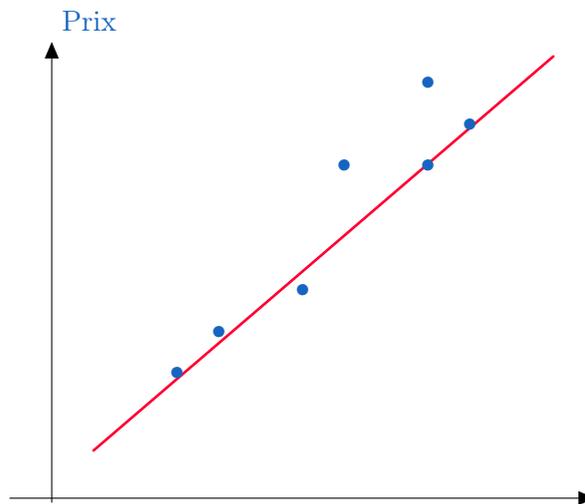
Troisième partie

Analyse des fonctions à plusieurs variables

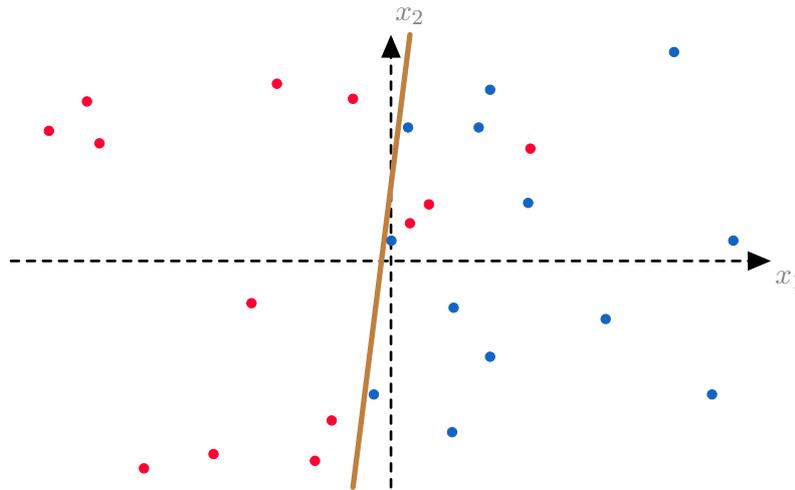
Cette deuxième partie se concentre sur l'analyse des fonctions à plusieurs variables et consiste essentiellement à étendre les concepts présentés en première partie à de telles fonctions.

La plupart des problèmes concrets que l'on va étudier mettent un jeu un grand nombres de variables et la plupart des quantités étudiées que l'on va chercher à estimer peuvent dépendre de plusieurs facteurs.

C'est par exemple le cas lorsque l'on cherche à estimer le prix d'un logement en fonction de ces différentes caractéristiques.



C'est aussi le cas lorsque l'on cherche à classer des individus dans certaines catégories en fonction de leurs multiples caractéristiques, ce que l'on peut couramment faire en marketing.



Pour étudier et résoudre ces problèmes là, nous aurons besoin des outils d'algèbre linéaires : comme la manipulation de vecteurs et de matrices ainsi que quelques notions de géométries.

Nous commencerons donc par présenter ce que sont les fonctions à plusieurs variables, étudier leur continuité et définir ce que signifie être *dérivable* pour de telles fonctions. Ces points là seront d'une extrême importance pour la résolution de problèmes d'optimisations à l'aide d'algorithme de descentes de gradient que nous présenterons dans un troisième temps :

$$\lim_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}).$$

L'étude plus approfondie des propriétés de la fonction f , comme la convexité ou encore l'étude de sa dérivée seconde vont nous permettre de déterminer les solutions du précédent problème. Cela permettra également d'établir des garanties de convergence des algorithmes que nous développerons dans le cadre de ce cours.

Enfin, il est aussi intéressant de voir comment la théorie de l'intégration peut s'étendre à de telles fonctions, et comment on peut intégrer sur différents domaines qui ne sont pas forcément linéaires.

9 Généralités sur les fonctions à plusieurs variables

Dans cette première partie, on commence par définir ce qu'est une fonction de plusieurs variables.

Définition 9.1: Fonctions à valeurs vectorielles

Une fonction de plusieurs variables à valeurs vectorielles f est une application d'un domaine $D \subset \mathbb{R}^n$ dans \mathbb{R}^p , où D est supposé non vide.

Exemple 9.1. La fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^3$ définie par

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

est une fonction à valeurs vectorielles, i.e., l'image du couple (x, y) par f est un vecteur.

Dans cet exemple, il est d'usage de désigner les composantes du vecteur image comme les composantes des images de la fonction f . Dans le cas où l'espace image est un sous ensemble de \mathbb{R} , la fonction est dite à **valeurs réelles**.

Définition 9.2: Fonctions à valeurs réelles

Une fonction f de plusieurs variables est dite à *valeurs réelles* si l'espace d'arrivée de cette fonction est un sous-ensemble de \mathbb{R} .

C'est ce type de fonction que nous rencontrerons le plus dans le cadre de ce cours.

Exemple 9.2. La fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = x^2 - y^2,$$

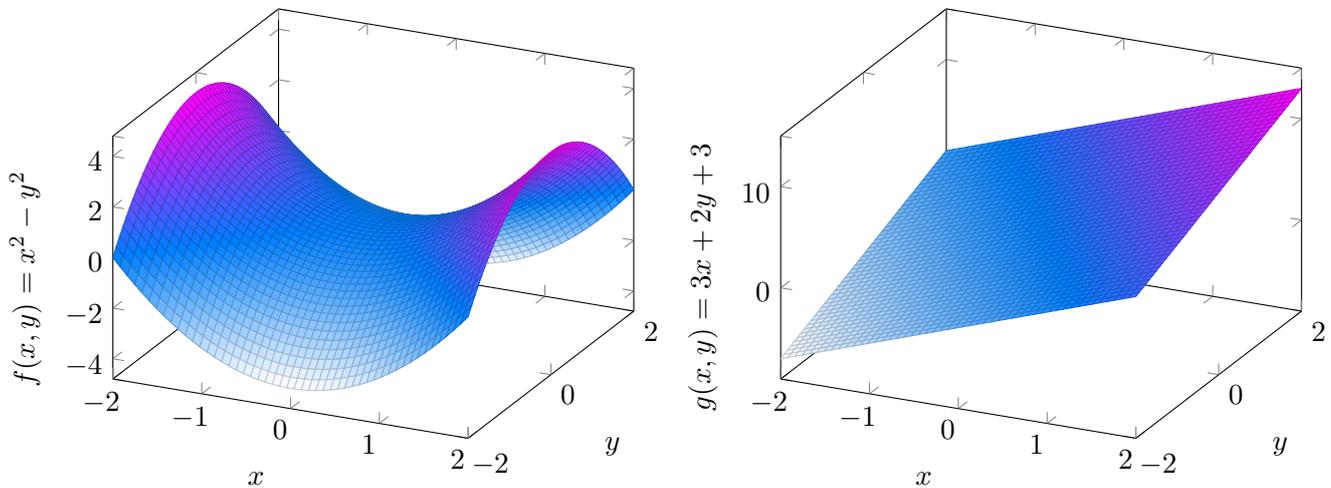
est une fonction à valeurs réelles.

Un autre exemple de fonction à valeurs réelles que l'on peut rencontrer sont les affines de la forme

$$f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle + c = a_1x_1 + \dots + a_nx_n + c, \text{ où } \mathbf{a} = (a_1, \dots, a_n) \text{ et } \mathbf{x} = (x_1, \dots, x_n).$$

On peut représenter graphiquement les fonctions définies par plusieurs variables mais nous sommes bien évidemment limité à une vue en trois dimensions. Ainsi nous ne pouvons représenter graphiquement que des fonctions réelles qui dépendent de deux variables.

Exemple 9.3. Les graphes des fonctions $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $f(x, y) = x^2 - y^2$ et $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par $g(x, y) = 3x + 2y + 3$, où g est une fonction affine :



Plus généralement, on définit un graphe comme suit :

Définition 9.3: Graphe d'une fonction

Soit f une fonction définie sur un sous-ensemble D de \mathbb{R}^n à valeurs réelles. On appelle *graphe* de la fonction f , l'ensemble :

$$\{(\mathbf{x}, f(\mathbf{x})) \in \mathbb{R}^{n+1} : \mathbf{x} = (x_1, \dots, x_n) \in D\}.$$

Le graphe d'une fonction affine, comme nous avons pu le voir dans l'exemple précédent, est un hyperplan de l'espace d'arrivée.

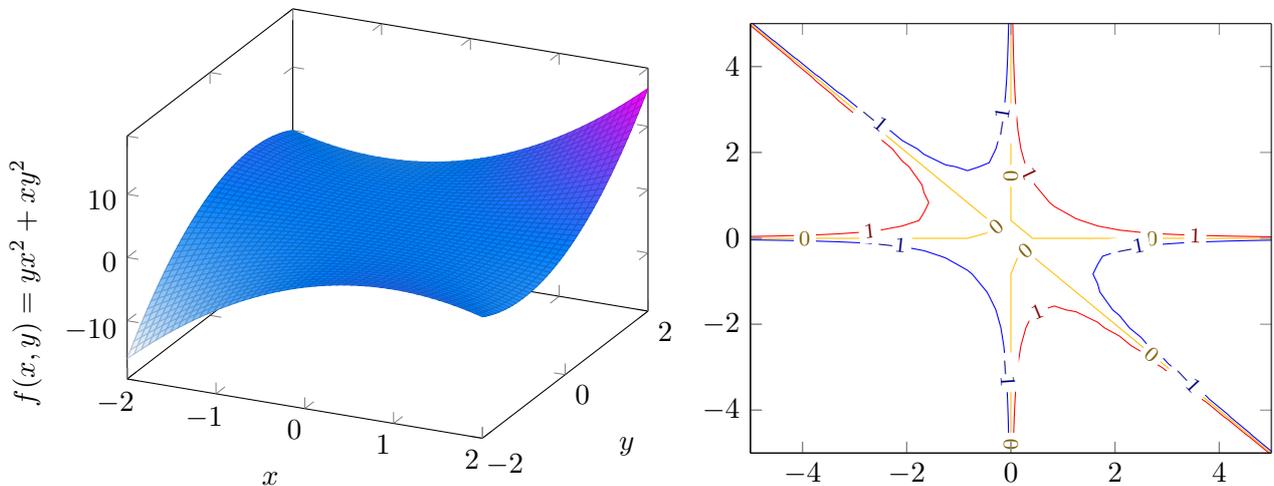
Lorsque l'on étudie le graphe d'une fonction à plusieurs variables, nous sommes parfois amenés à étudier ce que l'on appelle les *lignes de niveaux* de la fonction, comme on peut le retrouver en cartographie. C'est notamment employé dans certains outils dans le domaine de l'optimisation de fonctions.

Définition 9.4: Ligne de niveau

Considérons une fonction $f : \mathbb{R}^n \rightarrow \mathbb{R}$. On appelle ligne de niveau de la fonction f pour un réel $c \in \text{Im}(f)$, l'ensemble des points

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) = c\}.$$

Exemple 9.4. Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $f(x, y) = yx^2 + xy^2$. Le graphe ainsi que la ligne de niveau $f(x, y) = 1$ sont représentés ci-dessous.



A votre avis, à quoi ressemblent les lignes de niveau des fonctions affines ? Prenez une fonction affine $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ et montrez que ses lignes de niveau sont des hyperplans de \mathbb{R}^2 .

Opérations sur les fonctions à plusieurs variables Pour finir, cette section, on présente quelques propriétés des opérations effectuées sur les fonctions à plusieurs variables. De façon générale, les opérations définies pour les fonctions définies de \mathbb{R} dans \mathbb{R} restent valables pour les fonctions à plusieurs variables.

Considérons maintenant deux fonctions f et g définies sur des domaines D et D' et λ un nombre réel.

Alors

- La fonction $\lambda(f + g)$ est définie sur $D \cap D'$ par

$$\lambda(f + g)(x) = \lambda f(x) + \lambda g(x).$$

- La fonction fg est définie sur $D \cap D'$ par :

$$fg(x) = f(x)g(x).$$

- La fonction f/g est définie sur $D \cap D' \setminus \{\mathbf{x} \in D' \mid g(\mathbf{x}) = 0\}$ par

$$\frac{f}{g}(x) = \frac{f(x)}{g(x)}.$$

10 Continuité et Dérivabilité

On va maintenant revenir sur les notions fondamentales de continuité et de dérivabilité en étudiant leurs définitions dans le cas des fonctions à plusieurs variables.

10.1 Continuité

On commence par définir la continuité.

Définition 10.1: Continuité

Considérons une suite $(\mathbf{x}_n)_{n \in \mathbb{N}}$ à valeurs dans \mathbb{R}^n dont la limite existe est égale à $\mathbf{x}_0 \in \mathbb{R}^n$. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$.

Alors f est dite continue en \mathbf{x}_0 si et seulement si

$$\lim_{n \rightarrow \infty} f(\mathbf{x}_n) = f\left(\lim_{n \rightarrow \infty} \mathbf{x}_n\right) = f(\mathbf{x}_0).$$

Ou, de façon équivalente, si et seulement si $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$.

Nous avons déjà rencontré plusieurs exemples de fonctions continues

- les fonctions linéaires et affines de la forme $f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle + c$, qui sont continues sur \mathbb{R} .
- les fonctions polynomiales de la forme $f(\mathbf{x}) = (\langle \mathbf{a}, \mathbf{x} \rangle + c)^d$.
- les fonctions exponentielles de la forme $f(\mathbf{x}) = e^{\langle \mathbf{a}, \mathbf{x} \rangle + c}$ le sont aussi sur \mathbb{R}
- les projections sur un sous-espace (ou sur une composante) sont des fonctions continues car linéaires et continues en 0.

On peut également énoncé un résultat sur la continuité concernant la composition de deux deux fonctions continues.

Proposition 10.1: Composition et continuité

Soit f est une fonction définie sur un sous-ensemble D de \mathbb{R}^n à valeurs dans \mathbb{R} et g une fonction définie sur un intervalle I de \mathbb{R} .

Si f est une fonction continue en tout point $\mathbf{x} \in D$ et si g est continue sur I , alors $g \circ f$ est continue sur D .

Cette proposition permettra de montrer que des fonctions complexes sont continues et les écrivant comme la composition de fonctions continues plus *simples*.

Exemple 10.1. *Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par*

$$f(x, y) = \frac{1}{x^2 + y^2 + 1}$$

Alors la fonction f est continue sur \mathbb{R}^2 comme la composition des fonctions $(x, y) \mapsto x^2 + y^2 + 1$, continue sur \mathbb{R}^2 car polynomiale et $z \mapsto \frac{1}{z}$, continue sur \mathbb{R}_+^* .

Avec ce même résultat, on peut même montrer que la fonction norme $\mathbf{x} \mapsto \|\mathbf{x}\|$ est continue.

10.2 Fonctions différentiables et dérivables

Le nombre dérivée d'une fonction réelle à valeurs réelles f en un point x_0 a été définie comme décrivant l'évolution de la fonction f au voisinage de ce point. Plus précisément, nous l'avons défini comme la variation de f au voisinage de x_0 par la quantité

$$f'(x_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Dans le cas unidimensionnel, nous n'avons que deux façons de faire tendre h vers 0, par valeurs inférieures ou valeurs supérieures. Dans le cas multi-dimensionnel il ne faudra plus considérer des h réels mais vectoriels car il existe une infinité de directions qui pointent vers un point \mathbf{x}_0 , ce qui nous ramène à notre étude des espaces vectoriels et la décomposition des vecteurs dans une base.

La direction \mathbf{h} considérée étant à présent un vecteur de \mathbb{R}^n , on peut donc le décomposer dans une base, la base canonique $(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n)$ par exemple et écrire

$$\mathbf{h} = h_1 \mathbf{e}_1 + h_2 \mathbf{e}_2 + \dots + h_n \mathbf{e}_n.$$

Ainsi, en étudiant les variations rapport aux différents vecteurs de la base, on pourra étudier les variations de la fonctions par rapport à n'importe quel vecteur \mathbf{h} et donc dans n'importe quelle direction.

Ce dernier point laisse donc à penser que la dérivée, que l'on l'appellera **gradient**, d'une fonction à plusieurs variables sera donc un **vecteur**.

De cette remarque, on va donc s'intéresser aux **dérivées partielles** de la fonction par à ces différentes composantes.

Définition 10.2: Fonctions partielles

Soit f une fonction définie sur un domaine de D de \mathbb{R}^n et $\mathbf{x} \in D$.
Pour tout $i \in \llbracket 1, n \rrbracket$, on appelle $f_{\mathbf{x}}^i$ les fonctions partielles définies par

$$f_{\mathbf{x}}^i(u) = f(x_1, x_2, \dots, x_{i-1}, u, x_{i+1}, \dots, x_n),$$

où u est tel que $\{(x_1, x_2, \dots, x_{i-1}, u, x_{i+1}, \dots, x_n)\} \subset D$.

Une fonction partielle peut finalement être vue comme une fonction d'une seule variable, *i.e.*, une fonction pour laquelle *toutes les variables sont figées sauf une*.

Dans le cas où ces différentes fonctions partielles sont dérivables, on parle alors de dérivées partielles.

Définition 10.3: Dérivées partielles

Soit f une fonction définie d'un ensemble D de \mathbb{R} dans \mathbb{R} . Si la fonction f admet des dérivées partielles en tout point $\mathbf{x} \in D$, alors on note $\frac{\partial f}{\partial x_i}$ l'application

$$\mathbf{x} \mapsto \frac{\partial f}{\partial x_i}(\mathbf{x}) = \frac{\partial f(x_1, x_2, \dots, x_i, \dots, x_n)(\mathbf{x})}{\partial x_i}.$$

On peut facilement faire une analogie avec la définition de nombre dérivée pour les fonctions d'une variable réelle

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_i) - f(\mathbf{x})}{h}.$$

On regardera seulement la variation dans une direction, ou selon une composante, du vecteur \mathbf{x} .

Exemple 10.2. Soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ une fonction affine définie par $f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle + c$, où $\mathbf{a} \in \mathbb{R}^n$ et $c \in \mathbb{R}$.

Alors les dérivées partielles de f existent et sont définies, pour tout $i \in \llbracket 1, n \rrbracket$ par

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) = \mathbf{a}$$

Exemple 10.3. Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $f(x, y) = e^{-x^2+y^2} - 3xy + \frac{1}{x^2 + y^2 + 1}$.

La fonction f admet des dérivées partielles par rapport à x et à y qui sont données par

$$\frac{\partial f}{\partial x}(x, y) = -2xe^{-x^2+y^2} - 3y - \frac{2x}{(x^2 + y^2 + 1)^2} \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = -2ye^{-x^2+y^2} - 3x - \frac{2y}{x^2 + y^2 + 1}.$$

On peut également définir le concept de fonctions de classe C^1 pour les fonctions à plusieurs variables, il faut et il suffit que les différentes dérivées partielles soient continues.

Cependant, la généralisation à des fonctions de classe C^n n'est pas aussi évidente. nous limiterons au cas des fonctions de classe C^2 que nous étudierons ultérieurement.

Maintenant que nous avons introduit les dérivées partielles, on peut introduire la notion de **gradient** qui représente la valeur de la dérivée d'une fonction f selon toutes ses composantes.

Définition 10.4: S

Soit f une fonction définie sur un domaine D de \mathbb{R}^n à valeurs dans \mathbb{R} et soit \mathbf{a} un point de D . Si f admet des dérivées partielles d'ordre 1 par rapport aux différentes variables, alors, on appelle **gradient de f en \mathbf{a}** et on note $\nabla f(\mathbf{a})$, le vecteur de \mathbb{R}^n définie par

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

De plus, si f est une fonction de classe C^1 , alors le gradient est une application continue de D dans \mathbb{R}^n .

Exemple 10.4. *Considérons la fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ définie par*

$$f(x, y, z) = 3x^2yz - 2x^3y^2z^4 + e^x.$$

La fonction f admet des dérivées partielles continues en tout point de D et son gradient est donnée par

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 6xyz - 6(xy z^2)^2 + e^x \\ 3x^2z - 4x^3y z^4 \\ 3x^2y - 8x^3y^2 z^3 \end{pmatrix}.$$

Lorsque la fonction est à valeurs vectorielles la notion de gradient change car l'image n'est plus un simple nombre réel mais un vecteur. Il sera donc important de regarder

comment se comporte les différentes composantes de l'image par la fonction selon les différentes variables.

Exemple 10.5. *Considérons la fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ définie par*

$$f(x, y, z) = (3xy, 6xy, -2yz).$$

On voit que chaque composante évolue différemment selon les variables x, y et z . On va donc regarder chaque composante de l'image une par une et leurs différentes dérivées partielles.

On sera donc amenés à calculer un nombre de dérivées partielles qui dépend à la fois du nombre de variables, mais aussi de la dimension de l'espace d'arrivée. On pourra alors obtenir, dans le cas d'une fonction à valeurs vectorielles, une **matrice**, que l'on appelle la **matrice Jacobienne**.

Définition 10.5: Matrice Jacobienne

Soit $\mathbf{a} \in \mathbb{R}^n$ et soit f une fonction définie sur un ensemble de \mathbb{R}^n dans \mathbb{R}^p telle que

$$f(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_p(\mathbf{x})),$$

où les fonctions f_j sont des fonctions à valeurs réelles qui admettent des dérivées partielles d'ordre 1.

On appelle **matrice jacobienne** de f en \mathbf{a} , notée $Jac_f(\mathbf{a}) \in \mathbb{R}^{p \times n}$ définie par

$$Jac_f(\mathbf{a}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_1}{\partial x_n}(\mathbf{a}) \\ \vdots & & \vdots \\ \frac{\partial f_p}{\partial x_1}(\mathbf{a}) & \dots & \frac{\partial f_p}{\partial x_n}(\mathbf{a}) \end{pmatrix}$$

La matrice Jacobienne est alors une généralisation gradient pour les fonctions à valeurs réelles.

Exemple 10.6. *Soit $\mathbf{A} \in \mathbb{R}^{p \times n}$, soit $f : \mathbb{R}^n \rightarrow \mathbb{R}^p$ définie par $f(\mathbf{x}) = \mathbf{A}\mathbf{x}$.*

Alors le Jacobien de f est donné par la matrice \mathbf{A} elle-même.

De façon générale, la dérivée d'une application linéaire est elle-même.

Exemple 10.7. *Considérons la fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ définie par*

$$f(x, y, z) = (3xy, 6xy, -2yz).$$

alors la matrice Jacobienne de f en un point (x, y, z) de \mathbb{R}^3 est donnée par

$$Jac_f(\mathbf{x}) = \begin{pmatrix} 3y & 3x & 0 \\ 6y & 6x & 0 \\ 0 & -2z & -2y \end{pmatrix}$$

Regardons maintenant quelques opérations sur les dérivées partielles.

Opération sur les dérivées partielles Les propriétés des opérations sur les dérivées partielles sont très semblables à celles pour les fonctions réelles à valeurs réelles.

Considérons une fonction $f : D \rightarrow \mathbb{R}$ et $g : D' \rightarrow \mathbb{R}$ et soit $\lambda \in \mathbb{R}$. Si les fonctions f et g admettent des dérivées partielles d'ordre 1 en un point $\mathbf{a} \in D \cap D'$, alors nous avons

- la fonction $\lambda(f + g)$ possède des dérivées partielles d'ordre 1 en \mathbf{a} et

$$\forall i \in \llbracket 1, n \rrbracket, \frac{\partial(\lambda(f + g))}{\partial x_i}(\mathbf{a}) = \lambda \frac{\partial f}{\partial x_i}(\mathbf{a}) + \lambda \frac{\partial g}{\partial x_i}(\mathbf{a}).$$

- La fonction fg possèdent des dérivées partielles d'ordre 1 en \mathbf{a} et

$$\forall i \in \llbracket 1, n \rrbracket, \frac{\partial(fg)}{\partial x_i}(\mathbf{a}) = g(\mathbf{a}) \frac{\partial f}{\partial x_i}(\mathbf{a}) + f(\mathbf{a}) \frac{\partial g}{\partial x_i}(\mathbf{a}).$$

- La fonction f/g possède également des dérivées partielles d'ordre en \mathbf{a} si $g(\mathbf{a}) \neq 0$ et

$$\frac{\partial}{\partial x_i} \left(\frac{f}{g} \right) (\mathbf{a}) = \frac{1}{g(\mathbf{a})^2} \left(g(\mathbf{a}) \frac{\partial f}{\partial x_i}(\mathbf{a}) - f(\mathbf{a}) \frac{\partial g}{\partial x_i}(\mathbf{a}) \right).$$

Les règles de dérivations sont identiques au cas d'une seule variable lorsqu'il s'agit de calculer les dérivées partielles.

Il nous reste à étudier comment cela se comporte avec la composition.

Proposition 10.2: Dérivation de composée de fonctions

Soit DS un sous ensemble de \mathbb{R}^n et soient $f : D \rightarrow \mathbb{R}$ et $\phi : \mathbb{R} \rightarrow \mathbb{R}$ et \mathbf{a} un point de D .

Si f et ϕ sont de classe C^1 , alors la fonction composée $\phi \circ f : D \rightarrow \mathbb{R}$ est aussi de classe C^1 sur D et on a

$$\forall i \in \llbracket 1, n \rrbracket \frac{\partial}{\partial x_i}(\phi \circ f)(\mathbf{a}) = (\phi' \circ f)(\mathbf{a}) \frac{\partial f}{\partial x_i}(\mathbf{a}) = \phi'(f(\mathbf{a})) \frac{\partial f}{\partial x_i}(\mathbf{a}).$$

Le gradient de la composée s'écrit alors

$$\nabla(\phi \circ f) = (\phi' \circ f) \nabla f.$$

On peut également l'étendre à des fonctions à valeurs vectorielles.

Proposition 10.3: Dérivation de fonctions composée

Soit D un sous ensemble de \mathbb{R}^n et soient $f : D \rightarrow \mathbb{R}$ une fonction de classe C^1 sur D et $g : \mathbb{R} \rightarrow \mathbb{R}^n$, une fonction réelle à valeurs vectorielles, $g(x) = (g_1(x), \dots, g_n(x))$ qui admet des dérivées partielles, et \mathbf{a} un point de D .

La fonction composée $f \circ g$ (à valeurs réelles) est alors dérivable sur \mathbb{R} et pour tout réel t on a

$$\begin{aligned} \frac{d(f \circ g)}{dt}(t) &= \sum_{i=1}^n \frac{\partial f}{\partial x_i}(g(t)) g'_i(t), \\ &= \nabla f(g(t))^T \text{Jac}_g(t), \\ &= \langle \nabla f(g(t)), \text{Jac}_g(t) \rangle, \end{aligned}$$

où $\text{Jac}_g(t) \in \mathbb{R}^{n \times 1}$ désigne la matrice jacobienne de g en t .

Jusqu'à présent, nous avons étudié les dérivées de d'une fonction différentiable f par rapport au vecteur de la base de \mathbb{R}^n . C'est-à-dire que nous avons étudié l'évolution de la fonction f par rapport aux vecteurs de bases, *i.e.* dans n directions différentes.

A partir de cela, nous sommes maintenant capables d'étudier la dérivée de cette même fonction f dans n'importe quelle direction $\mathbf{d} \in \mathbb{R}^n$ qui va s'écrire

$$\mathbf{d} = \sum_{j=1}^n d_j \mathbf{e}_j.$$

On parle alors de **dérivée directionnelle**.

Définition 10.6: Dérivée directionnelle

Soit f une fonction définie sur un domaine $D \subset \mathbb{R}^n$, \mathbf{x} un point de D et \mathbf{d} un vecteur de norme unitaire^a. Considérons également la fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ par $\phi(t) = f(\mathbf{x} + t\mathbf{d})$.

Si ϕ est dérivable en 0, alors f est dérivable en \mathbf{x} dans la direction \mathbf{d} et sa dérivée est noté $D_{\mathbf{d}}f(\mathbf{x})$

$$\begin{aligned} D_{\mathbf{d}}f(\mathbf{x}) &= \phi'(0), \\ &= \lim_{t \rightarrow 0} \frac{\phi(t) - \phi(0)}{t}, \\ &= \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t}. \end{aligned}$$

a. Pour rappel, cela signifie que $\|\mathbf{d}\| = \sum_{j=1}^n d_j^2 = 1$.

Dans le cas où \mathbf{d} n'est rien d'autre qu'un vecteur de base, on retrouve la définition de dérivée partielle.

Proposition 10.4: Dérivée directionnelle

Soit D un sous ensemble de \mathbb{R}^n et soit $f : D \rightarrow \mathbb{R}$ une fonction de classe C^1 sur D , \mathbf{x} un point de D et \mathbf{d} un vecteur unitaire de \mathbb{R}^n .

Alors la fonction f possède une dérivée dans la direction \mathbf{d} égale à

$$D_{\mathbf{d}}f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle.$$

Démonstration. Repartons de la définition et considérons $\phi(t) = f(\mathbf{x} + t\mathbf{d}) = f(g(t))$ où $g : \mathbb{R} \rightarrow D$ définie par $g(t) = \mathbf{x} + t\mathbf{d}$. On réécrit ainsi la fonction ϕ comme une composée de deux fonctions. En appliquant la Proposition ??, nous avons

$$\phi'(t) = (f \circ g)'(t) = \langle \nabla f(g(t)), J(t) \rangle.$$

En $t = 0$, nous avons :

$$\phi'(0) = \langle \nabla f(g(0)), J(0) \rangle = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle.$$

□

Regardons un petit exemple avec une fonction affine.

Exemple 10.8. Soit $\mathbf{a} \in \mathbb{R}^n$ et c un nombre réel. Considérons f la fonction affine définie de \mathbb{R}^n dans \mathbb{R} par

$$f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle + c = a_1x_1 + \dots + a_nx_n + c.$$

Considérons maintenant une direction $\mathbf{d} = (d_1, d_2, \dots, d_n)$ de \mathbb{R}^n . Alors, pour tout réel t

$$\begin{aligned}\phi(h) &= f(\mathbf{x} + t\mathbf{d}), \\ &= \langle \mathbf{a}, \mathbf{x} + t\mathbf{d} \rangle + c, \\ &= a_1(x_1 + td_1) + a_2(x_2 + td_2) + \dots + a_n(x_n + td_n).\end{aligned}$$

Donc la dérivée directionnelle est donnée par

$$a_1d_1 + \dots + a_nd_n = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle.$$

Ces notions de dérivées, de gradient, nous servent principalement, comme en dimension 1 à déterminer une approximation de notre fonction au voisinage d'un point. Plus précisément, pour le moment, une approximation d'ordre 1 soit une approximation affine.

Cela suppose de deux choses. Si on note T l'approximation affine de la fonction f en un point \mathbf{a} , alors T doit vérifier

- Il faut que la fonction et son approximation coïncide en le point \mathbf{a} où la fonction est approchée

$$f(\mathbf{a}) = T(\mathbf{a}).$$

- Il faut aussi T devienne aussi proche de f lorsque \mathbf{x} tend \mathbf{a} et cela plus rapidement que la vitesse de convergence de \mathbf{x} vers \mathbf{a} , ce que l'on peut traduire par

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - T(\mathbf{x})}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Pour obtenir une expression de cette approximation, nous avons besoin d'introduire une dernière définition, qui est celle de **différentiabilité**, même si le terme est déjà apparu auparavant. Cette définition va faire apparaître une fonction linéaire qui sera importante dans la définition de notre approximation T .

Définition 10.7: Différentielle des fonctions à plusieurs variables

Soit D un sous-ensemble de \mathbb{R}^n , f une application de D à valeurs réelles et \mathbf{a} un point de D . On dit que f est différentiable en \mathbf{a} s'il existe une fonction linéaire L telle que

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - (f(\mathbf{a}) + L(\mathbf{a} - \mathbf{a}))}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

Comme la fonction L est linéaire, elle peut s'écrire

$$L(\mathbf{x}) = \langle \mathbf{u}, \mathbf{x} \rangle = u_1x_1 + u_2x_2 + \dots + u_nx_n.$$

On doit donc avoir

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \frac{f(\mathbf{x}) - (f(\mathbf{a}) + \langle \mathbf{u}, \mathbf{x} - \mathbf{a} \rangle)}{\|\mathbf{x} - \mathbf{a}\|} = 0.$$

On peut même montrer que le vecteur \mathbf{u} dont il fait mention ici n'est rien d'autre que le vecteur des dérivées partielles de la fonction f évaluées en \mathbf{a} , *i.e.*,

$$\forall i \in \llbracket 1, n \rrbracket, \quad u_i = \frac{\partial f}{\partial x_i}(\mathbf{a}).$$

Ainsi, notre fonction T est définie par

$$T(\mathbf{x}) = f(\mathbf{a}) + \sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{a})(x_i - a_i).$$

C'est cette dernière quantité

$$\sum_{i=1}^n \frac{\partial f}{\partial x_i}(\mathbf{a})(h) = \frac{\partial f}{\partial x_1}(\mathbf{a})h_1 + \dots + \frac{\partial f}{\partial x_n}(\mathbf{a})h_n,$$

que l'on appelle différentielle de f en \mathbf{a} évaluée en \mathbf{h} .

On remarque de suite que cette différentielle est aussi égale à

$$\langle \nabla f(\mathbf{a}), \mathbf{h} \rangle == D_{\mathbf{h}}f(\mathbf{a}).$$

Prenons quelques exemples pour mieux comprendre les objets que l'on manipule.

Exemple 10.9. Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $f(x_1, x_2) = x_1^2 - x_2^2$.

Et considérons le vecteur de \mathbb{R}^2 $\mathbf{d} = (1/2, \sqrt{3}/2)$, qui est bien unitaire.

Alors le gradient de la fonction f est donnée par

$$\nabla f(\mathbf{x})^T = \begin{pmatrix} 2x_1 & -2x_2 \end{pmatrix}.$$

Ainsi, la dérivée directionnelle de f dans la direction \mathbf{d} , qui est aussi la différentielle de f en \mathbf{x} évaluée en \mathbf{d} , est donnée par

$$D_{\mathbf{d}}f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = \begin{pmatrix} 2x_1 & -2x_2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ \frac{\sqrt{3}}{2} \end{pmatrix} = x_1 - \sqrt{3}x_2.$$

Ainsi, on peut évaluer l'accroissement de la fonction f en n'importe quel point \mathbf{x} dans la direction \mathbf{d} donnée.

Exemple 10.10. Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $f(x, y) = \cos\left(\frac{x_1^2}{2}\right) + x_2^2$.

Et considérons le vecteur de \mathbb{R}^2 $\mathbf{d} = (4/5, 3/5)$, qui est bien unitaire.

Alors le gradient de la fonction f est donnée par

$$\nabla f(\mathbf{x})^T = \begin{pmatrix} -x_1 \sin\left(\frac{x_1^2}{2}\right) & 2x_2 \end{pmatrix}.$$

Ainsi, la dérivée directionnelle de f dans la direction \mathbf{d} , qui est aussi la différentielle de f en \mathbf{x} évaluée en \mathbf{d} , est donnée par

$$D_{\mathbf{d}}f(\mathbf{x}) = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = \begin{pmatrix} -x_1 \sin\left(\frac{x_1^2}{2}\right) & 2x_2 \end{pmatrix} \begin{pmatrix} \frac{4}{5} \\ \frac{3}{5} \end{pmatrix} = -\frac{4x_1}{5} \sin\left(\frac{x_1^2}{2}\right) + \frac{6}{5}x_2.$$

Par exemple, dans la direction donnée, cet accroissement ne cesse de changer de signe selon la valeur de x_1 .

Avec ces notions, on peut alors définir l'approximation linéaire d'une fonction f au voisinage d'un point.

Définition 10.8: Approximation affine

Soit D un sous-ensemble de \mathbb{R}^n et f une fonction de D dans \mathbb{R} .

Alors f possède une approximation affine en $\mathbf{a} \in D$, s'il existe une forme linéaire L telle que pour tout vecteur \mathbf{h} vérifiant $\mathbf{a} + \mathbf{h} \in D$ on ait

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + L(\mathbf{h}) + o(\|\mathbf{h}\|),$$

où $o(\|\mathbf{h}\|)$ représente une fonction qui tend vers 0 lorsque \mathbf{h} tend vers $\mathbf{0}$ plus vite que $\|\mathbf{h}\|$.

De plus, si f est de classe C^1 sur D , alors on

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{h} \rangle + o(\|\mathbf{h}\|).$$

Si on pose $\mathbf{h} = \mathbf{x} - \mathbf{a}$ dans la définition précédente, nous sommes alors en mesure d'approcher la valeur de la fonction f en n'importe quel point \mathbf{x} à partir de sa valeur en \mathbf{a} de sa dérivée en ce point.

Ainsi, comme dans le cas uni-dimensionnel, l'équation

$$f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle$$

définit l'**hyperplan tangent** à f en \mathbf{a} .

Exemple 10.11. Considérons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par $f(x_1, x_2) = x_1^2 + x_2^2$.

Et considérons le vecteur de \mathbb{R}^2 $\mathbf{a} = (1/2, 1/2)$.

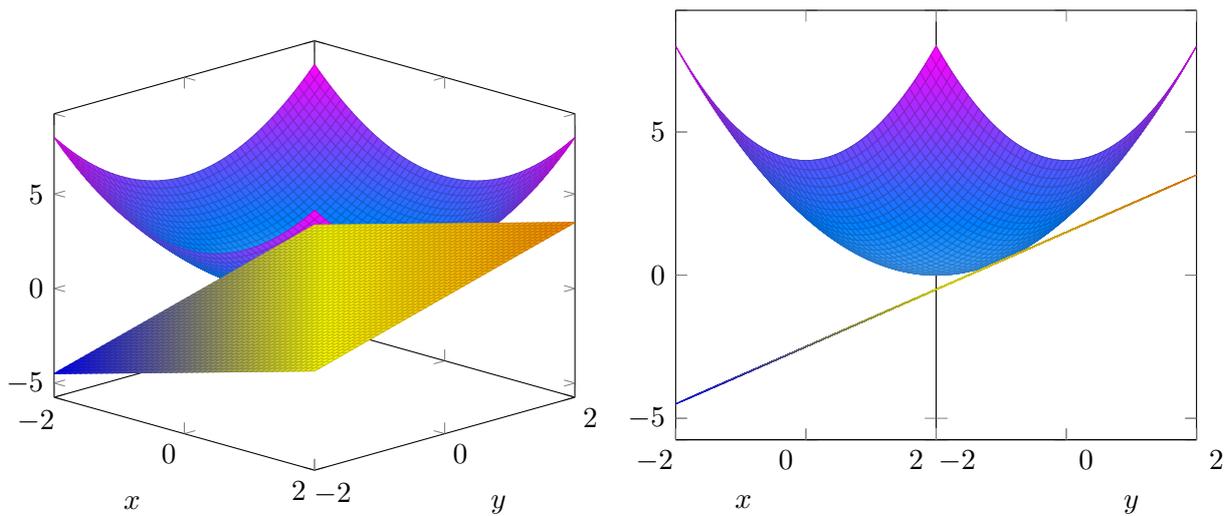
Alors le gradient de la fonction f est donnée par

$$\nabla f(\mathbf{x})^T = \begin{pmatrix} 2x_1 & 2x_2 \end{pmatrix}.$$

Et le plan tangent de f au point \mathbf{a} est donné par l'équation

$$T(\mathbf{x}) = f(\mathbf{a}) + \langle \nabla f(\mathbf{a}), \mathbf{x} - \mathbf{a} \rangle = -\frac{1}{2} + x + y.$$

Ce que l'on peut représenter graphiquement comme suit



Dans la suite, nous allons nous concentrer sur les fonctions qui sont cette fois-ci deux fois différentiables.

10.3 Fonction différentiables : ordre 2

Le fait, pour une fonction d'être deux fois différentiable va nous permettre de fournir une caractérisation des fonctions convexes. Nous pourrons aussi déterminer la nature d'un extremum d'une fonction, comme nous l'avons fait dans le cas des fonctions réelles à valeurs réelles.

Commençons par regarder ce qu'est une fonction deux fois différentiables en définissant les dérivées partielles d'ordre 2.

Définition 10.9: Dérivées partielles d'ordre 2

Soit D un sous-ensemble de \mathbb{R}^n et $\mathbf{x} \in \mathbb{R}^n$. Considérons une fonction f de D dans \mathbb{R} dont les dérivées partielles d'ordre 1 admettent des dérivées partielles, *i.e.*, telle que pour $i \in \llbracket 1, n \rrbracket$, $\frac{\partial f}{\partial x_i}(\mathbf{x})$ existent et soient différentiables.

Cette dérivée est notée $\frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x})$.

Si les dérivées premières partielles admettent des dérivées partielles d'ordre 1, alors les fonctions

$$\forall i, j \in \llbracket 1, n \rrbracket, \frac{\partial^2 f}{\partial x_j \partial x_i}$$

sont appelées **dérivées partielles d'ordre 2**

Exemple 10.12. Reprenons la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ définie par

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

Cette fonction admet des dérivées partielles d'ordre 2. Les dérivées partielles d'ordre 1 sont

$$\frac{\partial f}{\partial x}(x, y) = y^2 - 2yx \quad \text{et} \quad \frac{\partial f}{\partial y}(x, y) = 2xy - x^2.$$

Ainsi, les dérivées partielles d'ordre 2 sont définies par, lorsque l'on dérive deux fois par rapport à la même variable

$$\frac{\partial^2 f}{\partial x \partial x}(x, y) = \frac{\partial^2 f}{\partial x^2}(x, y) = -2y \quad \text{et} \quad \frac{\partial^2 f}{\partial y \partial y}(x, y) = \frac{\partial^2 f}{\partial y^2}(x, y) = 2x.$$

Et les dérivées dites croisées sont données par

$$\frac{\partial f}{\partial y} \left(\frac{\partial f}{\partial x} \right) (x, y) = 2y - 2x \quad \text{et} \quad \frac{\partial f}{\partial x} \left(\frac{\partial f}{\partial y} \right) (x, y) = 2y - 2x.$$

Dans cet exemple, on remarque que les dérivées partielles croisées sont égales. Ce n'est pas propre à cet exemple et c'est une conséquence d'un résultat plus général que l'on appelle le **Théorème de Schwarz**.

Théorème 10.1: Théorème de Schwarz

Soit D un sous-ensemble de \mathbb{R}^n et f une application de classe C^2 de D dans \mathbb{R}^n .
Alors

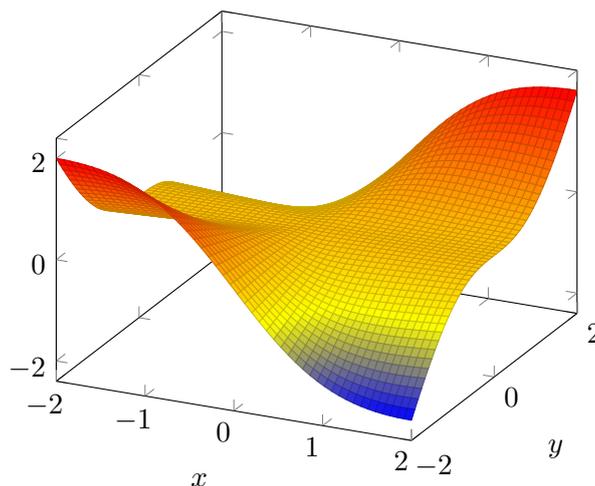
$$\forall i, j \in \llbracket 1, n \rrbracket \frac{\partial f}{\partial y} \left(\frac{\partial f}{\partial x} \right) = \frac{\partial f}{\partial x} \left(\frac{\partial f}{\partial y} \right).$$

Ce théorème nous dit que dans le cas où la fonction étudiée est de classe C^2 , donc les dérivées partielles d'ordre 2 existent et sont continues, alors l'ordre de dérivation n'a pas d'importance.

Ce résultat peut également servir à montrer qu'une fonction f n'est pas de classe C^2 .

Exemple 10.13. Soit f la fonction définie de \mathbb{R}^2 dans \mathbb{R} par

$$f(x, y) = \frac{xy^3}{x^2 + y^2} \text{ si } (x, y) \neq (0, 0) \text{ et } f(0, 0) = 0$$



La fonction f est bien de classe C^1 . Les dérivées partielles sont nulles en $(0, 0)$ et en dehors, nous avons

$$\frac{\partial f}{\partial x}(x, y) = y^3 \frac{y^2 - x^2}{(x^2 + y^2)^2} \text{ et } \frac{\partial f}{\partial y}(x, y) = xy^2 \frac{3x^2 + y^2}{(x^2 + y^2)^2}.$$

Pour tout y non nul et x non nul respectivement, nous avons

$$\frac{\partial f}{\partial x}(0, y) = -y \text{ et } \frac{\partial f}{\partial y}(x, 0) = x.$$

Ainsi

$$\lim_{y \rightarrow 0} \frac{\frac{\partial f}{\partial x}(0, y) - \frac{\partial f}{\partial x}(0, 0)}{y - 0} = -1$$

et

$$\lim_{x \rightarrow 0} \frac{\frac{\partial f}{\partial y}(x, 0) - \frac{\partial f}{\partial y}(0, 0)}{x - 0} = 1$$

Les dérivées partielles secondes en $(0, 0)$ ne sont pas égales. La fonction n'est donc pas de classe C^2 .

Tout comme nous avons défini le Jacobien d'une fonction f , quand cette dernière était de classe C^1 , nous définissons ce que l'on appelle une **matrice hessienne**, pour les fonctions de classe C^2 .

Définition 10.10: Matrice Hessienne

Soit f une fonction définie sur un sous-ensemble D de \mathbb{R}^n à valeurs dans \mathbb{R} et de classe C^2 . On appelle **matrice hessienne** ou parfois **hessien** de f en un point $\mathbf{x} \in \mathbb{R}^n$, la matrice notée

$$\nabla^2 f(\mathbf{x}) = \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \right)_{i,j=1,\dots,n}.$$

On peut également rencontrer la notation $Hess$ pour désigner le hessien de la fonction f .

Plus généralement, si une fonction f est de classe C^k , on peut à nouveau approcher cette fonction par un développement de Taylor d'ordre au plus k .

Définition 10.11: Formules de Taylor-Young

Soit f une fonction définie sur un sous-ensemble D de \mathbb{R}^n à valeurs réelles de classe C^k et \mathbf{x} un point de \mathbb{R}^n ,

Ainsi, une approximation d'ordre 2 d'une fonction f de classe C^2 au voisinage de $\mathbf{x} \in \mathbb{R}^2$ est donnée, $\forall \mathbf{h} = (h_1, h_2) \in \mathbb{R}^2$, par

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + h_1 \frac{\partial f}{\partial x_1}(\mathbf{x}) + h_2 \frac{\partial f}{\partial x_2}(\mathbf{x})$$

$$\begin{aligned}
& + \frac{1}{2} \left(h_1^2 \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) + 2h_1 h_2 \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) \right) + h_2^2 \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) + o(\|\mathbf{h}\|^2), \\
& = f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{h} \rangle + \frac{1}{2} \langle \nabla^2 f(\mathbf{x}) \mathbf{h}, \mathbf{h} \rangle + o(\|\mathbf{h}\|^2).
\end{aligned}$$

On servira de cette approximation lorsque l'on cherchera à étudier le fonctionnement d'algorithmes d'optimisation qui utilisent la dérivée seconde de la fonction.

Exemple 10.14. Soit la fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie par

$$f(x, y) = ye^x + \cos(xy).$$

Cette fonction f est bien de classe C^2 sur \mathbb{R}^2 et admet donc un développement limité d'ordre 2 en $\mathbf{0}$ donnée par

$$f(\mathbf{x}) = f(\mathbf{0}) + x \frac{\partial f}{\partial x}(\mathbf{0}) + y \frac{\partial f}{\partial y}(\mathbf{0}) + \frac{1}{2} \left(x^2 \frac{\partial^2 f}{\partial x^2}(\mathbf{0}) + 2xyx \frac{\partial^2 f}{\partial x \partial y}(\mathbf{0}) + y^2 x \frac{\partial^2 f}{\partial y^2}(\mathbf{0}) \right),$$

où la gradient est donné par

$$\nabla f(\mathbf{x}) = \begin{pmatrix} ye^x - y \sin(xy) & e^x - x \sin(xy) \end{pmatrix} \quad \text{soit} \quad \nabla f(\mathbf{0}) = \begin{pmatrix} 0 & 1 \end{pmatrix}.$$

La matrice hessienne est donnée par

$$\nabla^2 f(\mathbf{x}) = \begin{pmatrix} ye^x - y^2 \cos(xy) & e^x - \sin(xy) - xy \cos(xy) \\ e^x - \sin(xy) - xy \cos(xy) & -x^2 \cos(xy) \end{pmatrix} \quad \text{soit} \quad \nabla^2 f(\mathbf{0}) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Donc une approximation de f d'ordre 2 au point $\mathbf{0}$ est donnée par

$$f(\mathbf{x}) = 1 + y + 2xy.$$

Nous avons maintenant tous les outils nécessaires à la caractérisations des extrema d'une fonction, ainsi qu'à l'étude des fonctions convexes.

11 Recherche d'extrema et applications aux fonctions convexes

On souhaite maintenant regarder comment les outils de calculs différentielles, *i.e.*, l'étude de la dérivée des fonctions peut nous permettre de localiser et caractériser les extrema d'une fonction.

Nous reviendrons également sur les notions de convexité et la caractérisation de la convexité.

11.1 Retour sur la convexité

Les définitions que nous avons rencontré pour la convexité dans les cas des fonctions d'une seule variable se généralise pour les fonctions à plusieurs variables.

Définition 11.1: Convexité

Soit f une fonction définie sur un domaine D de \mathbb{R}^d à valeurs dans \mathbb{R} . On dit que la fonction f est **convexe** sur D si pour tout $\mathbf{u}, \mathbf{v} \in \mathbb{R}^d$ et pour tout $\alpha \in [0, 1]$ nous avons

$$f(\alpha\mathbf{u} + (1 - \alpha)\mathbf{v}) \leq \alpha f(\mathbf{u}) + (1 - \alpha)f(\mathbf{v}).$$

Dans cette définition, si on remplace le signe \leq par \geq , on définit une fonction **concave**. De plus, si les inégalités larges sont remplacées par des inégalités strictes, la fonction sera dite **strictement convexe** ou **strictement concave**. On rappelle également que si une fonction f est convexe alors $-f$ est concave et réciproquement.

On peut également caractériser la convexité d'une fonction en se basant sur le gradient de cette dernière ou encore en utilisant la matrice hessienne.

Définition 11.2: Convexité, caractérisation ordre 1

Soit f une fonction de classe C^1 sur un ensemble $D \in \mathbb{R}^d$, alors

- f est **convexe** si tout hyperplan tangent à f est en **dessous** de son graphe, *i.e.*,

$$\forall \mathbf{u}, \mathbf{v} \in D, \langle \mathbf{u}, \mathbf{v} \rangle \geq f(\mathbf{v}) + \langle \mathbf{u} - \mathbf{v}, \nabla f(\mathbf{v}) \rangle.$$

- f est **concave** si tout hyperplan tangent à f est au **dessus** de son graphe, *i.e.*,

$$\forall \mathbf{u}, \mathbf{v} \in D, \langle \mathbf{u}, \mathbf{v} \rangle \leq f(\mathbf{v}) + \langle \mathbf{u} - \mathbf{v}, \nabla f(\mathbf{v}) \rangle.$$

Enfin, la caractérisation d'ordre 2.

Définition 11.3: Convexité, caractérisation ordre 2

Soit f une fonction de classe C^2 sur un ensemble $D \in \mathbb{R}^d$, alors

- f est **convexe** si sa matrice hessienne est *semie-définie positive, i.e.*,

$$\forall \mathbf{u}, \mathbf{v} \in D, \mathbf{u}^T \nabla^2 f(\mathbf{v}) \mathbf{u} \geq 0.$$

- f est **concave** si sa matrice hessienne est *semie-définie négative, i.e.*,

$$\forall \mathbf{u}, \mathbf{v} \in D, \mathbf{u}^T \nabla^2 f(\mathbf{v}) \mathbf{u} \leq 0.$$

Exemple 11.1. Soit f la fonction définie de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(x, y) = (42y)^2 + 5x^2 + x + 3y + 4xy.$$

On va étudier la convexité de cette dernière.

Commençons par voir que la fonction f est une forme quadratique que l'on peut donc réécrire sous forme matricielle.

$$\begin{aligned} f(x, y) &= (42y)^2 + 5x^2 + x + 3y + 4xy, \\ &= 16 - 16y + 4y^2 5x^2 + x + 3y + 4xy, \\ &= 5x^2 + 4y^2 + 4xy + x - 13y + 16, \\ &= \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{b}, \mathbf{u} \rangle + 16, \end{aligned}$$

$$\text{où } \mathbf{u} = \begin{pmatrix} x & y \end{pmatrix}^T, \mathbf{A} = \begin{pmatrix} 5 & 2 \\ 2 & 4 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 1 & -13 \end{pmatrix}^T$$

Pour cette forme quadratique, la matrice hessienne est donnée par $2\mathbf{A}$, or cette matrice \mathbf{A} a une trace et un déterminant positif, ses valeurs propres sont donc positives et la matrice \mathbf{A} est donc positive. Ainsi f est convexe.

11.2 Conditions d'optimalité

Après ce bref rappel sur les fonctions convexes, nous lancer dans la recherche d'extrema d'une fonction f définie de $D \subset \mathbb{R}^d$ à valeurs dans \mathbb{R} .

Let us first define the minimum of a function f .

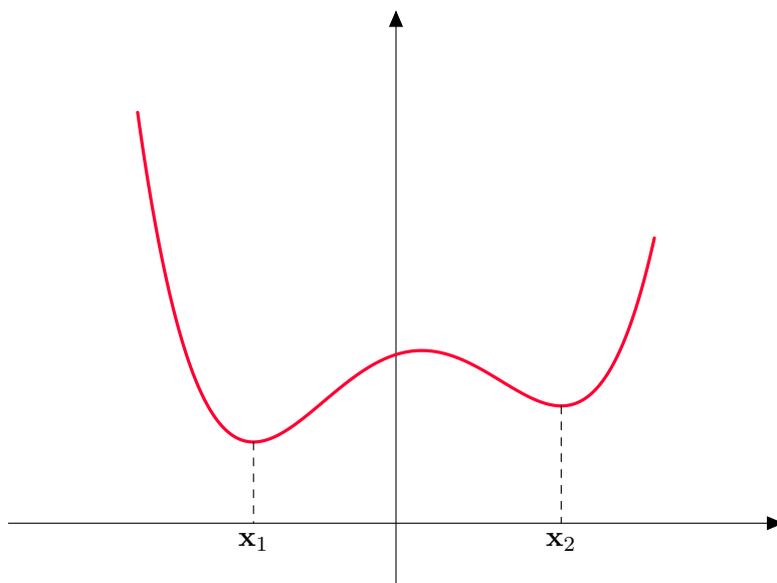


FIGURE 23 – Illustration du concept de minimum local et global d’une fonction. Le point x_1 est le minimum global de la fonction alors que x_2 est un simple minimum local.

Définition 11.4: Minimus local et global

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction continue. On dit que $\mathbf{u} \in \mathbb{R}^d$ est **minimum local** de f si pour tout voisinage $V \subset \mathbb{R}^d$ de \mathbf{u} (c’est une ensemble qui contient de le point \mathbf{u}) on a :

$$f(\mathbf{u}) \leq f(\mathbf{v}), \quad \forall \mathbf{v} \in V,$$

Le point \mathbf{u} est un **minimum global** est une minimum global de la fonction f si et seulement si :

$$f(\mathbf{u}) \leq f(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbb{R}^d.$$

Noter bien la différence entre local et global, dans le premier cas, on supposera que la fonction a atteint son minimum dans un espace donné. Dans le second cas, elle doit pas atteindre une valeur plus faible que celle atteinte en \mathbf{u} .

En d’autres termes, le minimum d’une fonction f est juste la valeur (\mathbf{u}) où \mathbf{u} est le point en lequel la fonction f atteint son minimum (local ou global) comme illustré par la Figure 23.

Lorsque l’on peut représenter le graphe de la fonction, il est très simple de localiser

un minimum (ou un maximum). Mais cela devient plus difficile en dimension supérieure.

La proposition suivante permet de donner une première caractérisation d'un minimum (local) d'une fonction f . C'est ce que l'on appelle l'**inéquation d'Euler**

Proposition 11.1: Inéquation d'Euler

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction continue \mathcal{U} un ensemble convexe non vide de \mathbb{R}^d .
Considérons également $\mathbf{u} \in \mathcal{U}$ un minimum de relatif de f par rapport à \mathcal{U} . Si f est différentiable en \mathbf{u} , alors

$$\nabla f(\mathbf{u})(\mathbf{v} - \mathbf{u}) \geq 0, \forall \mathbf{v} \in \mathcal{U}$$

Cela signifie simplement que si on évalue le gradient de f en \mathbf{u} alors la fonction sera croissante dans n'importe quelle direction $\mathbf{v} - \mathbf{u}$

Démonstration. Supposons que \mathbf{u} est un minimiseur relatif par rapport à notre ensemble \mathcal{U} . Alors pour tout $\alpha > 0$ et tout vecteur \mathbf{v} tels que $\alpha(\mathbf{u} - \mathbf{v}) \in \mathcal{U}$, nous avons

$$f(\mathbf{u}) \leq f(\mathbf{u} + \alpha(\mathbf{u} - \mathbf{v})).$$

Ainsi

$$\frac{f(\mathbf{u}) - f(\mathbf{u} + \alpha(\mathbf{u} - \mathbf{v}))}{\alpha} \geq 0.$$

En prenant la limite lorsque α tend vers 0, on trouve que

$$D_{\mathbf{d}}f(\mathbf{u}) \geq 0.$$

□

Mais ce résultat est rarement utilisé en pratique Si on revient à la Figure 23, on remarque que les extrema sont localisés en les points où le gradient de la fonction f s'annule. Cette condition d'optimalité est connue sous le d'**équation d'Euler**.

Proposition 11.2: Equation d'Euler

Soit $f : \mathbb{R}^d \rightarrow \mathbb{R}$ une fonction continue et différentiable en $\mathbf{u} \in \mathbb{R}^d$. Si \mathbf{u} est un extremum alors :

$$\nabla f(\mathbf{u}) = 0.$$

Démonstration. On fait la preuve dans le cas où \mathbf{u} est un minimum, mais elle est analogue dans le cas où \mathbf{u} est un maximum.

En utilisant la définition de minimum dans un voisinage de \mathbf{u} , i.e., $\forall \mathbf{v} \in \mathbb{R}^d, \exists t > 0$ tels que $\mathbf{u} + t\mathbf{v} \in V$, nous avons

$$\begin{aligned} f(\mathbf{u}) &\leq f(\mathbf{u} + t\mathbf{v}) = f(\mathbf{u}) + \nabla f(\mathbf{u})^T(t\mathbf{v}) + t\mathbf{v}^T \varepsilon(t\mathbf{v}), \quad t \ll 1 \\ \Leftrightarrow 0 &\leq \nabla f(\mathbf{u})^T(t\mathbf{v}) + t\mathbf{v}^T \varepsilon(t\mathbf{v}), \end{aligned}$$

où ε est une fonction vectorielle dont les composantes tendent vers 0 lorsque t tend vers 0.

En divisant par $t > 0$ et en prenant la limite lorsque $t \rightarrow 0$, on obtient :

$$0 \leq \nabla f(\mathbf{u})\mathbf{v}$$

En remplaçant maintenant \mathbf{v} par $-\mathbf{v}$ on a, de façon similaire :

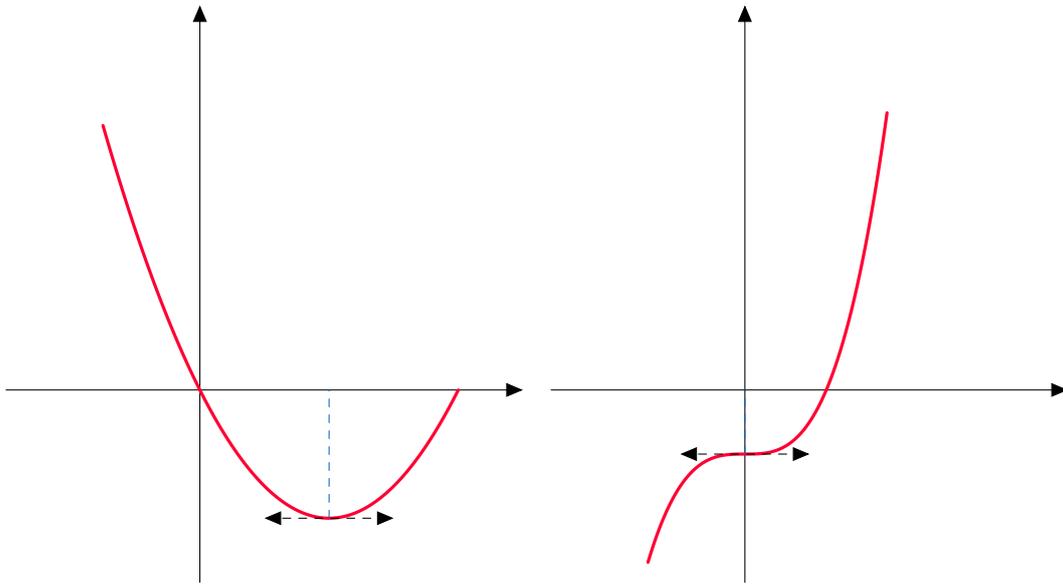
$$0 \leq -\nabla f(\mathbf{u})^T \mathbf{v}.$$

Ainsi pour tout $\mathbf{v} \in \mathbb{R}^d$ on a $\nabla f(\mathbf{u})^T \mathbf{v} = 0$ et donc $\nabla f(\mathbf{u}) = 0$. □

Les solutions de l'équation d'Euler définissent ce que l'on appelle des *points critiques* ou des *valeurs critiques*.

Faites bien attention ! Cette proposition ne donne qu'une condition nécessaire pour trouver un extremum d'une fonction, mais un peu point critique n'est pas forcément un extremum, la condition n'est pas suffisante comme nous avons pu le voir dans le cas unidimensionnel.

Exemple 11.2. Let us consider the functions f and g respectively defined by $\frac{1}{2}(x-2)^2 - 2$ and $\frac{1}{2}x^3 - 1$ and represented below.



Les points \mathbf{x} des deux graphes sont solutions de l'équation d'Euler $\nabla f(\mathbf{x}) = 0$ et $\nabla g(\mathbf{x}) = 0$ respectivement. Cependant, s'il s'agit bien d'un point en lequel la fonction f atteint son minimum, on ne peut pas dire que le point \mathbf{x} soit un point en lequel la fonction g atteint un extremum.

Cet exemple montre qu'il est important de définir des critères qui permettent de caractériser la nature d'un point critique : minimum, maximum ou ... ni l'un ni l'autre.

Etant donnée une solution \mathbf{u} de $\nabla f(\mathbf{u}) = 0$, on peut dire que :

- \mathbf{u} est un **minimum local** si $\nabla^2 f(\mathbf{u}) = Hess_f(\mathbf{u}) \geq 0$, i.e. si la matrice hessienne de la fonction f évaluée en \mathbf{u} est semie-définie positive. Cela signifie qu'en ce point, la fonction f est localement convexe!
Ce point est un minimum global si f est **convexe** sur son ensemble de définition ou si pour tout $\mathbf{v} \neq \mathbf{u}$ on a $f(\mathbf{u}) \leq f(\mathbf{v})$.
- \mathbf{u} est un **maximum local** si $\nabla^2 f(\mathbf{u}) = Hess_f(\mathbf{u}) \leq 0$, si la matrice hessienne de la fonction f évaluée en \mathbf{u} est semie-définie négative. Cela signifie qu'en ce point, la fonction f est localement concave!
Ce point est un maximum global si f est **concave** sur son ensemble de définition ou si pour tout $\mathbf{v} \neq \mathbf{u}$ on a $f(\mathbf{u}) \geq f(\mathbf{v})$.
- Dans les autres cas, on ne peut rien dire ou il faut procéder à une étude plus poussée.

Exemple 11.3. Soit f la fonction définie de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(x, y) = (42y)^2 + 5x^2 + x + 3y + 4xy.$$

On va étudier la convexité de la fonction f et chercher ses extremum.

Commençons par voir que la fonction f est une forme quadratique que l'on peut donc réécrire sous forme matricielle.

$$\begin{aligned} f(x, y) &= (42y)^2 + 5x^2 + x + 3y + 4xy, \\ &= 16 - 16y + 4y^2 + 5x^2 + x + 3y + 4xy, \\ &= 5x^2 + 4y^2 + 4xy + x - 13y + 16, \\ &= \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{b}, \mathbf{u} \rangle + 16, \end{aligned}$$

$$\text{où } \mathbf{u} = \begin{pmatrix} x & y \end{pmatrix}^T, \mathbf{A} = \begin{pmatrix} 5 & 2 \\ 2 & 4 \end{pmatrix} \text{ et } \mathbf{b} = \begin{pmatrix} 1 & -13 \end{pmatrix}^T$$

Pour cette forme quadratique, la matrice hessienne est donnée par $2\mathbf{A}$, or cette matrice \mathbf{A} a une trace et un déterminant positif, ses valeurs propres sont donc positives et la matrice \mathbf{A} est donc positive. Ainsi f est convexe.

Pour rechercher ses extrema, on commence par résoudre l'équation d'Euler

$$\nabla f(\mathbf{u}) = 2\mathbf{A}\mathbf{u} + \mathbf{b} = \mathbf{0}.$$

Cela nous amènes à résoudre le système

$$\begin{cases} 10x + 4y + 1 = 0, \\ 4x + 8y - 13 = 0. \end{cases}$$

La solution de ce système est donnée par le vecteur

$$\mathbf{u} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -\frac{30}{32} \\ \frac{67}{32} \end{pmatrix}.$$

La fonction f étant convexe, il s'agit du minimum global de la fonction f .

Etudions un deuxième exemple qui n'implique pas une forme quadratique.

Exemple 11.4. Soit f la fonction définie de \mathbb{R}^2 dans \mathbb{R} définie par

$$f(\mathbf{u}) = f(x, y) = 2x^2 + 4(y^2)^2 + 4x + 6y2xy + 2y^3.$$

On va reprendre la même étude que précédemment.

Nous avons un terme cubique donc on ne peut pas exprimer cette fonction à l'aide d'une matrice comme nous l'avons fait dans l'exemple précédent.

La matrice hessienne de notre fonction est donnée par

$$\text{Hess}_f(\mathbf{u}) = \begin{pmatrix} 4 & -2 \\ -2 & 12y + 8 \end{pmatrix}$$

Remarquons que la trace et le déterminant de cette matrice sont respectivement égales à $12y+12$ et $48y+28$. Ces deux quantités sont positives si et seulement si $y \geq -\frac{7}{12}$.

Ainsi la fonction est convexe sur $\mathbb{R} \times \left[-\frac{7}{12}, +\infty\right[$.

Elle est concave si $12y + 12$ est négatif et $48y + 28$ est positif, donc si y vérifie $y \leq -1$ et $y \geq -\frac{7}{12}$. Ces deux conditions sont incompatibles donc la fonction n'est ni convexe, ni

concave en dehors de l'ensemble $\mathbb{R} \times \left[-\frac{7}{12}, +\infty\right[$.

Pour trouver les extrema de cette fonction, on doit résoudre le système

$$\nabla f(\mathbf{u}) = 0 \iff \begin{cases} 4x + 4 - 2y & = & 0, \\ 6y^2 + 8y - 2x - 10 & = & 0. \end{cases} \iff \begin{cases} 2x & = & y - 2, \\ 6y^2 + 7y - 8 & = & 0. \end{cases}$$

Ce système admet deux solutions \mathbf{u}_1 et \mathbf{u}_2 qui sont

$$\mathbf{u}_1 = \begin{pmatrix} \frac{-31}{24} - \frac{\sqrt{241}}{24} \\ -\frac{7}{12} - \frac{\sqrt{241}}{12} \end{pmatrix} \quad \text{et} \quad \mathbf{u}_2 = \begin{pmatrix} \frac{-31}{24} + \frac{\sqrt{241}}{24} \\ -\frac{7}{12} + \frac{\sqrt{241}}{12} \end{pmatrix}$$

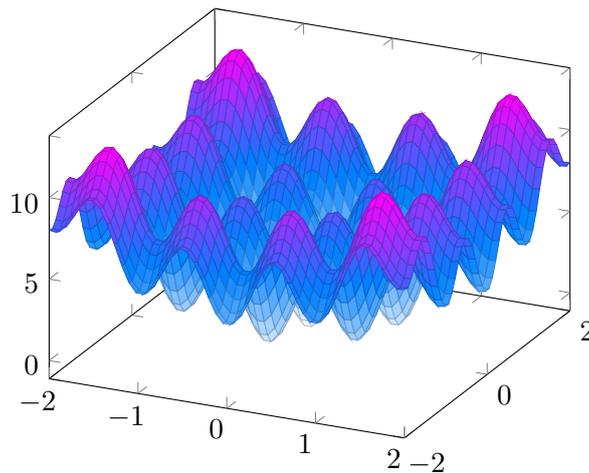
Le premier point \mathbf{u}_1 voit sa deuxième composante être plus petite que -1 , il ne s'agit donc ni d'un minimum, ni d'un maximum. Quant au deuxième point, sa deuxième composante est strictement positive, il s'agit donc d'un **minimum local** de cette fonction. Ce n'est pas un minimum global, car la fonction f tend vers $-\infty$ lorsque y tend vers $-\infty$.

Reargons un dernier exemple

Exemple 11.5. *Etudions $f : [-2, 2]^2 \rightarrow \mathbb{R}$ définie par*

$$f(\mathbf{x}) = 4 + (x_1^2 - 2 \cos(2\pi x_1)) + (x_2^2 - 2 \cos(2\pi x_2)).$$

Cette fonction est intéressante car elle admet plusieurs minima locaux mais un seul minimum global. Elle est souvent utilisée pour tester l'efficacité d'un algorithme d'optimisation.



On peut voir que cette fonction n'est ni convexe, ni concave globalement. On peut regarder le gradient de cette fonction

$$\nabla f(\mathbf{x}) = \begin{pmatrix} 2x_1 + 4\pi \sin(2\pi x_1) & 2x_2 + 4\pi \sin(2\pi x_2) \end{pmatrix}.$$

Les solutions de l'équation d'Euler sont données par la résolution du système

$$\begin{aligned} x_1 &= -2\pi \sin(2\pi x_1), \\ x_2 &= -2\pi \sin(2\pi x_2). \end{aligned}$$

Les deux équations sont indépendantes mais sont globalement difficiles à résoudre ... sauf si on garde l'esprit que $\sin(x) \in [-1, 1]$. Ainsi, on peut voir que le vecteur nul est un point critique, qui est aussi le minimum global de la fonction.

On va maintenant regarder quelques exemples concrets de problèmes que l'on va chercher à résoudre en Science des Données.

11.3 Problème d'optimisation

On va considérer que l'on se trouve dans un espace E de dimension det soit $f : E \rightarrow \mathbb{R}$. Un problème d'optimisation, ici de minimisation se présente comme

$$\min_{\mathbf{x} \in E} f(\mathbf{x}).$$

La fonction f s'appelle souvent **la fonction de coût** ou **risque**. Cela peut représenter le coût associé au stockage de produits dans un entrepôt (représenté par le paramètre \mathbf{x}) ou le risque lié à une prise de décision.

Regardons quelques problèmes d'optimisations classiques que vous étudierez plus en détail l'année prochaine. Ils sont là uniquement pour motiver la prochaine section.

La régression linéaire Etant donnée un vecteur de valeurs $\mathbf{y} \in \mathbb{R}^m$ et une matrice de données $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_m)$, $\mathbf{x}_i \in \mathcal{R}^d$ où $d < m$, On souhaite trouver un vecteur $\boldsymbol{\theta} \in \mathbb{R}^d$ qui permet d'expliquer au mieux \mathbf{y} à l'aide de \mathbf{X} en utilisant le modèle suivant :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \quad \text{où } \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}),$$

où $\boldsymbol{\varepsilon}$ représente les erreurs du modèle.

Pour trouver le meilleur vecteur $\boldsymbol{\theta}$ on cherche à résoudre le problème suivant :

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^d} \sum_{i=1}^m \varepsilon_i = \min_{\boldsymbol{\theta} \in \mathbb{R}^d} \|\mathbf{y} - \mathbf{X}\boldsymbol{\theta}\|_2^2.$$

On peut observer que ce problème d'optimisation est convexe. En effet, son gradient est donné par

$$\nabla_{\boldsymbol{\theta}} \boldsymbol{\varepsilon} = -2\mathbf{X}^T(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}),$$

et la solution de l'équation d'Euler est $\boldsymbol{\theta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$.

La dérivée seconde, égale à

$$\nabla_{\boldsymbol{\theta}}^2 = 2\mathbf{X}^T \mathbf{X}$$

est une matrice de $\mathbb{R}^{d \times d}$ symétrique et semi-définie positive. Notre problème est donc convexe.

Montrons que la matrice hessienne est bien semi-définie positive. Soit λ une valeur propre de $\mathbf{X}^T \mathbf{X}$ et \mathbf{u}_λ le vecteur propre associé. En utilisant la définition de vecteur propre, nous avons

$$\begin{array}{rcl}
 \mathbf{X}^T \mathbf{X} \mathbf{u}_\lambda & = & \lambda \mathbf{u}_\lambda, \\
 \downarrow \text{ en pré-multipliant par } \mathbf{u}_\lambda^T & & \\
 \mathbf{u}_\lambda^T \mathbf{X}^T \mathbf{X} \mathbf{u}_\lambda & = & \lambda \mathbf{u}_\lambda^T \mathbf{u}_\lambda, \\
 \downarrow \text{ droite : définition de norme} & & \\
 \downarrow \text{ gauche : propriété de la transposition} & & \\
 (\mathbf{X} \mathbf{u}_\lambda)^T \mathbf{X} \mathbf{u}_\lambda & = & \lambda \|\mathbf{u}_\lambda\|_2^2, \\
 \downarrow \text{ gauche : définition de norme} & & \\
 \|\mathbf{X} \mathbf{u}_\lambda\|_2^2 & = & \lambda \|\mathbf{u}_\lambda\|_2^2.
 \end{array}$$

Comme un vecteur propre est nécessairement non-nul, on a

$$\lambda = \frac{\|\mathbf{X} \mathbf{u}_\lambda\|_2^2}{\|\mathbf{u}_\lambda\|_2^2} \geq 0.$$

Dans ce cas, nous avons obtenu une solution analytique à notre problème, mais ce n'est pas toujours possible.

Régression logistique La régression logistique est aussi problème de régression. En revanche le but ne sera plus de prédire une valeur réelle, mais plutôt l'appartenance à un groupe pour un individu donné, ce que l'on appellera *une classe*.

Pour prédire l'appartenance à une classe, on utilisera le modèle suivant

$$g(\mathbf{x}, \boldsymbol{\theta}) = \log \left(\frac{\mathbb{P}(y = 1 | \mathbf{x})}{1 - \mathbb{P}(y = 1 | \mathbf{x})} \right) = \theta_0 + \theta_1 x_1 + \dots + \theta_d x_d.$$

Les paramètres du modèle sont estimés par une méthode que l'on appelle, le *maximum de vraisemblance*, une quantité très utilisée en statistiques.

Plus précisément, on va se concentrer l'optimisation de la *negative log-vraisemblance* :

$$\ell(\mathbf{X}, \boldsymbol{\theta}) = - \sum_{i=1}^m y_i \log(p_i) + (1 - y_i) \log(1 - p_i), \quad \text{où } p_i = \frac{1}{1 + \exp(- \sum_{j=1}^d \theta_j X_{ij})},$$

où p_i désigne une probabilité X_{ij} désigne l'élément en i -ème ligne et j -ème colonne de la matrice \mathbf{X} .

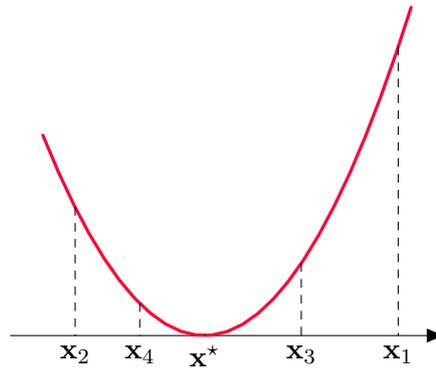
Malheureusement, il n'existe pas de solutions analytiques à ce problème. Pour trouver une solution, nous aurons besoin d'utiliser des algorithmes d'optimisation.

12 Optimisation : descente de gradient

Etant donnée une fonction **function** $f : \mathbb{R}^d \rightarrow \mathbb{R}$, nous cherchons à résoudre :

$$\mathbf{x}^* = \arg \min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}).$$

Notre objectif est de présenter quelques algorithmes qui permettent de construire une suite $(\mathbf{x}_n)_{n \in \mathbb{N}}$ qui converge cette solution \mathbf{x}^* , la valeur \mathbf{x} pour laquelle f atteint son minimum.



Pour cela, on nous utiliserons une technique appelée la descente de gradient qui va permettre d'orienter la direction de l'espace de recherche des solutions, dans la construction de notre suite.

12.1 Introduction à la descente de gradient

Etant donnée une fonction f et un ensemble \mathcal{U} non vide et sachant que le problème suivant admet une solution

$$\mathbf{u}^* = \arg \min_{\mathbf{v} \in \mathcal{U}} f(\mathbf{v}).$$

l'idée est de construire une suite $(\mathbf{u}_k)_{k \in \mathbb{N}}$ qui converge vers \mathbf{u}^* en utilisant le principe suivant :

- choisir une valeur \mathbf{u}_0 : c'est le point de départ de notre algorithme d'optimisation.
- $\mathbf{u}_k \rightarrow \mathbf{u}_{k+1}$: on choisit une direction \mathbf{d}_k et on minimise la fonction objective f le long de cette direction.
- on résout $\arg \min_{\rho > 0} f(\mathbf{u}_k - \rho \mathbf{d}_k) = \rho_k$: on cherche à quel point on doit se déplacer dans la direction donnée pour minimiser la fonction f . Cette constante ρ est souvent appelé **pas d'apprentissage**.
- $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \mathbf{d}_k$: on met à jour la valeur de notre suite.

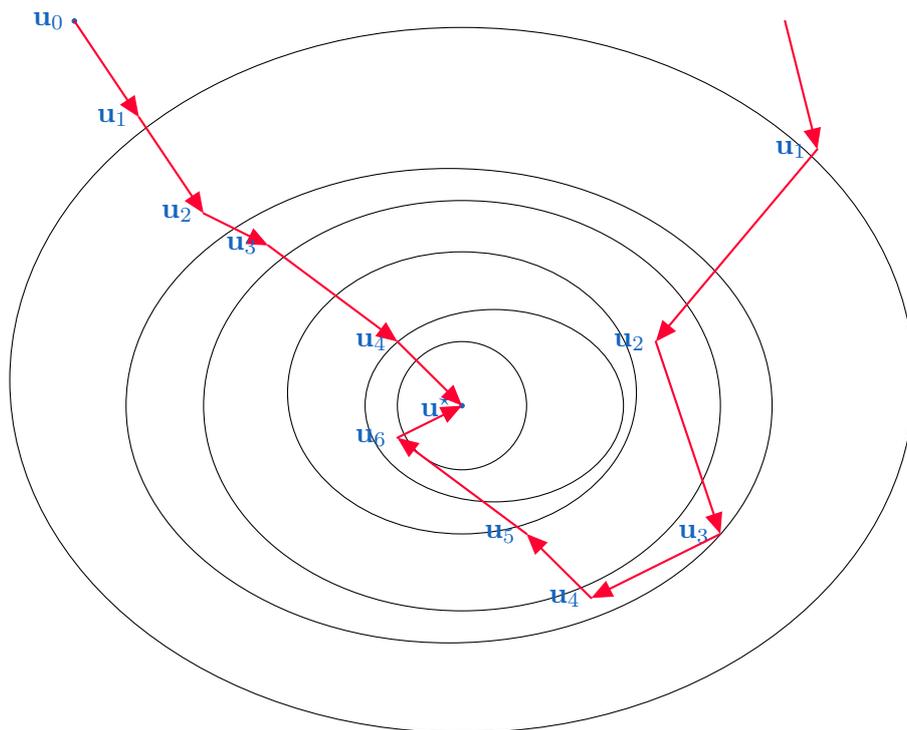


FIGURE 24 – Illustration de la convergence vers \mathbf{u}^* de la fonction f via plusieurs chemins. Les différentes ellipses représentent les lignes de niveau de la fonction f .

Les trois dernières étapes du processus sont itérées jusqu'à ce que l'on soit suffisamment proche de la solution \mathbf{u}^* .

Plusieurs questions se posent

- i) comment choisir la direction de descente \mathbf{d}_k ?
- ii) jusqu'à quel point devant nous avancer dans la direction indiquée, *i.e.*, quelle valeur donnée à ρ_k .

Cette question est tout à fait légitime dans la mesure où, comme le montre la Figure 24, plusieurs chemins peuvent mener à la solution optimale. Cependant il y en a certainement un qui est préférable à un autre. Comment pouvons-nous utiliser la connaissance que nous avons de la fonction à optimiser, et de ses dérivées pour construire un tel chemin.

Pour répondre à la première question *i*), nous nous intéressons à la quantité suivante

$$f(\mathbf{u}_k - \rho \mathbf{d}_k).$$

Plus précisément, nous allons nous intéresser à l'approximation d'ordre 1 de cette

fonction. Ainsi, lorsque ρ est proche de 0, nous avons le développement limité d'ordre 1 suivant

$$f(\mathbf{u}_k - \rho \mathbf{d}_k) = f(\mathbf{u}_k) - \rho \langle \nabla f(\mathbf{u}_k), \mathbf{d}_k \rangle + \rho \varepsilon(\rho),$$

où ε est une fonction qui tend 0 lorsque ρ tend vers 0.

Pour minimiser $f(\mathbf{u}_k - \rho \mathbf{d}_k)$ nous devons choisir la direction \mathbf{d}_k qui maximise le produit scalaire $\langle \nabla f(\mathbf{u}_k), \mathbf{d}_k \rangle$. Une application de l'inégalité de **Cauchy-Schwartz** nous donne

$$\langle \nabla f(\mathbf{u}_k), \mathbf{d}_k \rangle \leq \|\nabla f(\mathbf{u}_k)\| \|\mathbf{d}_k\|.$$

mais surtout, que l'égalité est atteinte lorsque les deux vecteurs sont colinéaires! Ainsi, il suffit de choisir \mathbf{d}_k **colinéaires au gradient de la fonction** $\nabla f(\mathbf{u}_k)$ donc, par exemple, prendre

$$\mathbf{d}_k = \nabla f(\mathbf{u}_k).$$

Nous pourrions alors écrire notre procédure comme suit

- choisir \mathbf{u}_0 pour initialiser notre algorithme,
- poser $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \nabla f(\mathbf{u}_k)$ pour $\rho_k > 0$ déterminé de façon à minimiser $f(\mathbf{u}_k - \rho_k \nabla f(\mathbf{u}_k))$

Il faut aussi regarder à quel moment notre algorithme s'arrête, car il n'est certainement pas possible, numériquement, d'atteindre exactement la solution optimale. Pour cela, on utilisera le fait que la solution \mathbf{u}^* est une valeur critique qui vérifie

$$\nabla f(\mathbf{u}^*) = \mathbf{0}.$$

Donc, lorsque la norme du gradient f est proche de 0 on peut supposer que l'on est proche de la solution optimale.

En pratique, on se fixe une constante strictement positive ε . On stoppe notre algorithme lorsque l'on a

$$\|\nabla f(\mathbf{u}_k)\| \leq \varepsilon.$$

Le dernier point à étudier est le choix de la valeur de ρ_k pour les différentes itérations de l'algorithme.

Un choix plutôt simple consisterait à prendre $\rho_k = \rho > 0$ pour n'importe quelle valeur de k .

Ce qui donne notre premier algorithme de descente de gradient

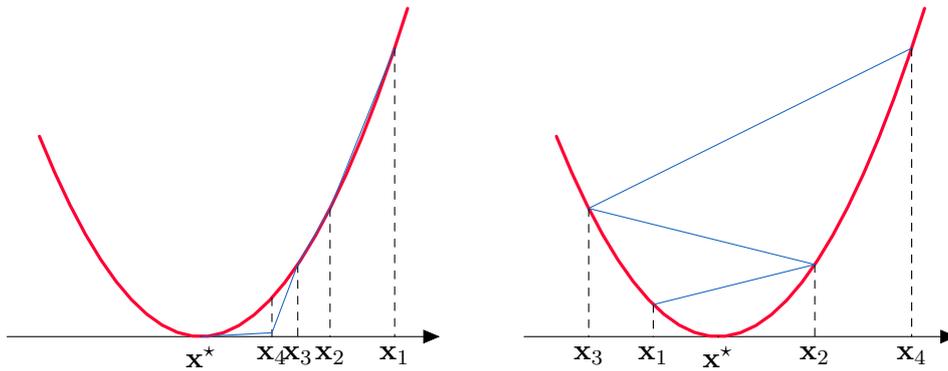


FIGURE 25 – Illustration de l'importance du choix du pas d'apprentissage ρ . Une valeur trop faible entraîne convergence très lente de l'algorithme (à gauche). Une valeur trop grande peut entraîner une divergence de l'algorithme (à droite).

Définition 12.1: Descente de gradient à pas constant

Soit f défini sur un sous-ensemble D de \mathbb{R}^n à valeurs dans \mathbb{R} et soit ρ, ε des nombres réels positifs.

Alors la **descente de gradient à pas constant** est décrit par

- choisir \mathbf{u}_0 pour initialiser notre algorithme,
- tant que $\|\nabla f(\mathbf{u}_k)\| \geq \varepsilon$
 1. calculer $\nabla f(\mathbf{u}_k)$
 2. poser $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho \nabla f(\mathbf{u}_k)$

Cependant, choisir la descente de gradient à pas constant n'est pas le meilleur choix. Cela peut conduire à une convergence extrêmement lente voire à une divergence de notre algorithme comme illustré en Figure 25.

On peut par contre choisir une suite de valeurs $(\rho_k)_{k \in \mathbb{N}}$ qui soit décroissante selon k , par exemple $\rho_k = \frac{1}{k}$. Mais il existe des choix plus performants.

12.2 Descente de gradient à pas optimal

Un des choix les plus intéressants est de choisir ρ_k , $k \in \mathbb{N}$ qui minimise la fonction selon la direction de descente $\mathbf{d}_k = \nabla f(\mathbf{u}_k)$. Cela conduit à un autre algorithme de descente de gradient que l'on appelle *la descente de gradient à pas optimal*.

Définition 12.2: Descente de gradient à pas optimal

Soit f défini sur un sous-ensemble D de \mathbb{R}^n à valeurs dans \mathbb{R} et soit $\varepsilon > 0$.

Alors la **descente de gradient à pas optimal** est définie par

- choisir \mathbf{u}_0 pour initialiser l'algorithme, $k = 0$
- tant que $\|\nabla f(\mathbf{u}_k)\| \geq \varepsilon$
 1. calculer $\nabla f(\mathbf{u}_k)$
 2. résoudre $\arg \min_{\rho > 0} f(\mathbf{u}_k - \rho \nabla f(\mathbf{u}_k))$
 3. poser $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho \nabla f(\mathbf{u}_k)$

On voit qu'à chaque itération de l'algorithme, nous devons résoudre un autre problème de minimisation mais selon une variable réelle uniquement. Cela peut rendre l'algorithme plus lent, mais il nécessite généralement peu d'itérations pour converger.

Néanmoins, on peut montrer que pour les fonctions dites *fortement convexes*, cet algorithme converge.

Nous avons déjà vu la notion de convexité, mais pas celle de convexité forte.

Définition 12.3: Fonction fortement convexe

Soit f une fonction continue et différentiable sur un sous-ensemble D de \mathbb{R}^d à valeurs dans \mathbb{R} . On dit que f est **fortement convexe**, **α -fortement convexe** ou encore **α -elliptique** s'il existe une constante $\alpha > 0$ telle que

$$\langle \nabla f(\mathbf{v}) - \nabla f(\mathbf{u}), \mathbf{v} - \mathbf{u} \rangle \geq \alpha \|\mathbf{v} - \mathbf{u}\|^2, \forall \mathbf{u}, \mathbf{v} \in \mathbb{R}^d$$

A l'aide de la dérivée seconde, on dit qu'une fonction f est *α -elliptique* si nous avons

$$\nabla^2 f \succeq \alpha \mathbf{I},$$

i.e. si pour tout $\mathbf{u} \in \mathbb{R}^d$, $\mathbf{u}^T \nabla^2 f \mathbf{u} \geq \alpha \|\mathbf{u}\|^2$.

Proposition 12.1: Convergence de l'algorithme à pas optimal

Si f est une fonction de classe C^1 et α -fortement convexe, $\alpha > 0$, alors la descente de gradient à pas optimal converge.

Démonstration. La démonstration de cette proposition sera effectuée en exercice. \square

Regardons cet algorithme et plus précisément les liens entre les directions de descente des algorithmes.

Que pouvons-nous dire, pour tout entier $k \in \mathbb{N}$ du produit scalaire $\langle \nabla f(\mathbf{u}_{k+1}), \nabla f(\mathbf{u}_k) \rangle$?

On va utiliser le fait que pour tout entier $k \in \mathbb{N}$, $\rho_k = \arg \min_{\rho > 0} f(\mathbf{u}_k - \rho \mathbf{d}_k)$?

Si ρ_k minimise $f(\mathbf{u}_k - \rho_k \mathbf{d}_k)$, nous avons

$$\begin{aligned} & \frac{\partial}{\partial \rho} f(\mathbf{u}_k - \rho \nabla f(\mathbf{u}_k)) \Big|_{\rho=\rho_k} = 0, \\ & \downarrow \text{ en utilisant la définition de dérivation de fonction composée} \\ \Leftrightarrow & \langle \nabla f(\mathbf{u}_k - \rho_k \nabla f(\mathbf{u}_k)), \nabla f(\mathbf{u}_k) \rangle = 0, \\ & \downarrow \text{ on se rappelle que } \mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \nabla f(\mathbf{u}_k) \\ \Leftrightarrow & \langle \nabla f(\mathbf{u}_{k+1}), \nabla f(\mathbf{u}_k) \rangle = 0. \end{aligned}$$

Ainsi, deux directions de descente successives sont orthogonales entre elles. Si cela ne semble pas d'un grand intérêt à première vue, elle présente un grand intérêt pour certains problèmes en apprentissage machine où nous serons amenés à travailler avec des *fonctions quadratiques*.

Regardons cela sur un exemple.

Exemple 12.1. Soit \mathbf{A} une matrice symétrique semi-définie positive, et soit $\mathbf{b} \in \mathbb{R}^d$. Considérons la forme quadratique f définie pour tout $\mathbf{u} \in \mathbb{R}^d$ par :

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

Regardons ce qu'il se passe si on cherche le pas optimal, dans la descente de gradient, lorsque l'on travaille avec des fonctions quadratiques.

Pour tout $k \in \mathbb{N}$ le gradient de la fonction f est donné par $\nabla f(\mathbf{u}_k) = \mathbf{A}\mathbf{u}_k - \mathbf{b}$.

Ainsi,

$$\begin{aligned} \nabla f(\mathbf{u}_{k+1}) &= \mathbf{A}\mathbf{u}_{k+1} - \mathbf{b}, \\ & \downarrow \text{ on utilise la définition de } \mathbf{u}_{k+1} \\ &= \mathbf{A}(\mathbf{u}_k - \rho_k(\mathbf{A}\mathbf{u}_k - \mathbf{b})) - \mathbf{b}, \\ & \downarrow \text{ on développe} \\ &= \mathbf{A}\mathbf{u}_k - \mathbf{b} - \rho_k \mathbf{A}(\mathbf{A}\mathbf{u}_k - \mathbf{b}). \end{aligned}$$

Utilisons le fait que pour tout entier k , nous avons $\langle \nabla f(\mathbf{u}_{k+1}), \nabla f(\mathbf{u}_k) \rangle = 0$.

$$\begin{aligned}
\langle \mathbf{A}\mathbf{u}_k - \mathbf{b}, \mathbf{A}\mathbf{u}_k - \mathbf{b} - \rho_k \mathbf{A}(\mathbf{A}\mathbf{u}_k - \mathbf{b}) \rangle &= 0, \\
\iff \langle \mathbf{A}\mathbf{u}_k - \mathbf{b}, \mathbf{A}\mathbf{u}_k - \mathbf{b} \rangle &= \langle \mathbf{A}\mathbf{u}_k - \mathbf{b}, \rho_k \mathbf{A}(\mathbf{A}\mathbf{u}_k - \mathbf{b}) \rangle, \\
\iff \rho_k &= \frac{\langle \mathbf{A}\mathbf{u}_k - \mathbf{b}, \mathbf{A}\mathbf{u}_k - \mathbf{b} \rangle}{\langle \mathbf{A}\mathbf{u}_k - \mathbf{b}, \mathbf{A}(\mathbf{A}\mathbf{u}_k - \mathbf{b}) \rangle}
\end{aligned}$$

Finalemnt, pour les formes quadratiques, le pas d'apprentissage optimal à chaque itération $k \in \mathbb{N}$ est égale à

$$\rho_k = \frac{\|\mathbf{A}\mathbf{u}_k - \mathbf{b}\|^2}{\|\mathbf{A}\mathbf{u}_k - \mathbf{b}\|_{\mathbf{A}}^2},$$

où $\|\mathbf{u}\|_{\mathbf{A}}^2 = \mathbf{u}^T \mathbf{A} \mathbf{u}$.

Pour les formes quadratiques, on peut ainsi aisément déterminer le pas optimal, sans avoir à résoudre un autre problème d'optimisation. Regardons un exemple d'application.

Exemple 12.2. Soit f la fonction définie de \mathbb{R}^2 dans \mathbb{R}^2 par

$$f(\mathbf{u}) = f(x, y) = 4x^2 - 4xy + 2y^2 = \frac{1}{2} \mathbf{u}^T \mathbf{A} \mathbf{u} - \mathbf{b}^T \mathbf{u} = \frac{1}{2} \mathbf{u}^T \begin{pmatrix} 8 & -4 \\ -4 & 4 \end{pmatrix} \mathbf{u} - \begin{pmatrix} 0 \\ 0 \end{pmatrix} \mathbf{u}.$$

On va effectuer quelques itérations de l'algorithme de descente de gradient à pas optimal à cette fonction qui est convexe sur \mathbb{R}^2 (on le vérifie facilement) en partant du

point $\mathbf{u}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

Il s'agit d'une forme quadratique, donc le pas optimal ρ est donné à chaque itération par

$$\rho_k = \frac{\|\mathbf{A}\mathbf{u}_k - \mathbf{b}\|}{\|\mathbf{A}\mathbf{u}_k - \mathbf{b}\|_{\mathbf{A}}}.$$

On rappelle aussi que la mise à jour est définie par

$$\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \nabla f(\mathbf{u}_k),$$

où $\nabla f(\mathbf{u}_k) = \mathbf{A}\mathbf{u}_k - \mathbf{b}$

Ainsi, pour les premières itérations nous avons

- **Première itération :**

$$\rho_0 = \frac{\|\mathbf{A}\mathbf{u}_0\|}{\|\mathbf{A}\mathbf{u}_0\|_{\mathbf{A}}} = \frac{16}{128} = \frac{1}{8} \quad \text{et} \quad \nabla f(\mathbf{u}_0) = \mathbf{A}\mathbf{u}_0 = \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

Ce nous donne

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{8} \begin{pmatrix} 4 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}.$$

- **Deuxième itération :**

$$\rho_1 = \frac{\|\mathbf{A}\mathbf{u}_1\|}{\|\mathbf{A}\mathbf{u}_1\|_{\mathbf{A}}} = \frac{4}{16} = \frac{1}{4} \quad \text{et} \quad \nabla f(\mathbf{u}_1) = \mathbf{A}\mathbf{u}_1 = \begin{pmatrix} 0 \\ 2 \end{pmatrix}.$$

Ce nous donne

$$\mathbf{u}_2 = \begin{pmatrix} 0.5 \\ 1 \end{pmatrix} - \frac{1}{4} \begin{pmatrix} 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}.$$

- **Troisième itération :**

$$\rho_2 = \frac{\|\mathbf{A}\mathbf{u}_2\|}{\|\mathbf{A}\mathbf{u}_2\|_{\mathbf{A}}} = \frac{4}{32} = \frac{1}{8} \quad \text{et} \quad \nabla f(\mathbf{u}_2) = \mathbf{A}\mathbf{u}_2 = \begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

Ce nous donne

$$\mathbf{u}_3 = \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} - \frac{1}{8} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}.$$

Il existe également d'autres méthodes de recherche dans une direction donnée, *i.e.*, d'autres façons de déterminer la valeur de ρ avec, par exemple

- la **règle d'Armijo**

- le **critère de Wolfe**

Mais nous ne détaillerons pas ces approches ici.

12.3 Descente de Gradient à pas conjugué

On va étudier une autre méthode descente de gradient dite à *pas conjugués* qui va choisir différemment la directions de descente. Pour cela, dans cette section, on va considérer une matrice \mathbf{A} symétrique et définie positive et la forme quadratique définie pour tout $\mathbf{u} \in \mathbb{R}^d$ par

$$f(\mathbf{u}) = \frac{1}{2} \langle \mathbf{A}\mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{b}, \mathbf{u} \rangle.$$

Définition 12.4: S

Soit \mathbf{A} une matrice symétrique définie positive de $\mathcal{M}_d(\mathbb{R})$ et \mathbf{u}, \mathbf{v} deux vecteurs de \mathbb{R}^d .

Les vecteurs \mathbf{u} et \mathbf{v} sont dits **conjugués** si $\langle \mathbf{A}\mathbf{u}, \mathbf{v} \rangle = 0$.

Regardons maintenant le fonction de l'algorithme.

- Soit $\mathbf{u}_0 \in \mathbb{R}^d$. Si $\nabla f(\mathbf{u}_0) = \mathbf{0}$, on arrête l'algorithme car nous avons atteint le minimum de la fonction.
Sinon, on définit la direction de descente $\mathbf{d}_0 = -\nabla f(\mathbf{u}_0)$ et on cherche à minimiser la fonction f dans cette direction. Cela revient à chercher la valeur ρ qui minimise

$$f(\mathbf{u}_0 - \rho \mathbf{d}_0),$$

ce que l'on obtient, d'après ce que l'on a vu dans la section précédente, en choisissant

$$\rho_0 = \frac{\langle \nabla f(\mathbf{u}_0), \mathbf{d}_0 \rangle}{\langle \mathbf{A}\mathbf{d}_0, \mathbf{d}_0 \rangle}.$$

- On pose ensuite $\mathbf{u}_1 = \mathbf{u}_0 - \rho_0 \mathbf{d}_0$. Si $\nabla f(\mathbf{u}_1) = \mathbf{0}$, on arrête l'algorithme car nous avons atteint le minimum de la fonction.
Sinon, on pose $\mathbf{d}_1 = -\nabla f(\mathbf{u}_1) + \alpha_0 \mathbf{d}_0$, où α_0 sera choisi tel que

$$\langle \mathbf{A}\mathbf{d}_1, \mathbf{d}_0 \rangle = 0 \iff \langle \mathbf{A}(-\nabla f(\mathbf{u}_1)), \mathbf{d}_0 \rangle + \langle \mathbf{A}\alpha_0 \mathbf{d}_0, \mathbf{d}_0 \rangle = 0.$$

Donc

$$\alpha_0 = -\frac{\langle \mathbf{A}\nabla f(\mathbf{u}_1), \mathbf{d}_0 \rangle}{\langle \mathbf{A}\mathbf{d}_0, \mathbf{d}_0 \rangle}.$$

On a donc défini notre direction de descente et notre pas d'apprentissage sera à nouveau donné par

$$\rho_1 = \frac{\langle \nabla f(\mathbf{u}_1), \mathbf{d}_1 \rangle}{\langle \mathbf{A}\mathbf{d}_1, \mathbf{d}_1 \rangle}.$$

On posera ensuite $\mathbf{u}_2 = \mathbf{u}_1 - \rho_1 \mathbf{d}_1$.

- Plus généralement, supposons que $\mathbf{u}_0, \mathbf{d}_0, \mathbf{u}_1, \mathbf{d}_1, \dots, \mathbf{u}_{k-1}, \mathbf{d}_{k-1}$ sont connues. Si $\nabla f(\mathbf{u}_k) = \mathbf{0}$, l'algorithme s'arrête. Sinon, on pose $\mathbf{d}_k = \nabla f(\mathbf{u}_k) + \rho_{k-1} \mathbf{d}_{k-1}$, où α_{k-1} sera choisi tel que

$$\langle \mathbf{A}\mathbf{d}_k, \mathbf{d}_{k-1} \rangle = 0 \iff \langle \mathbf{A}\nabla f(\mathbf{u}_k), \mathbf{d}_{k-1} \rangle + \langle \mathbf{A}\alpha_{k-1} \mathbf{d}_{k-1}, \mathbf{d}_{k-1} \rangle = 0.$$

Donc

$$\alpha_{k-1} = -\frac{\langle \mathbf{A}\nabla f(\mathbf{u}_k), \mathbf{d}_{k-1} \rangle}{\langle \mathbf{A}\mathbf{d}_{k-1}, \mathbf{d}_{k-1} \rangle}.$$

Et notre pas d'apprentissage sera à nouveau donné par

$$\rho_k = \frac{\langle \nabla f(\mathbf{u}_k), \mathbf{d}_k \rangle}{\langle \mathbf{A}\mathbf{d}_k, \mathbf{d}_k \rangle}.$$

Plus de calculs sont nécessaires pour mettre en oeuvre l'algorithme, mais ces dernières restent simples.

On peut également montrer le résultats suivant concernant les directions de descente \mathbf{d}_k de l'algorithme à pas conjugués.

Proposition 12.2: Propriété de la descente à pas conjugués

Supposons que pour tout entier $1 \leq k \leq n$ nous ayons $\nabla f(\mathbf{u}_k) \neq \mathbf{0}$. Alors nous avons les relations suivantes :

$$\forall 0 \leq l \leq k-1 \langle \nabla f(\mathbf{u}_k), \nabla f(\mathbf{u}_l) \rangle = 0 \text{ et } \langle \mathbf{A}\mathbf{d}_k, \mathbf{d}_l \rangle.$$

La démonstration se fait par récurrence et le développement précédent. Nous pourrions la regarder en TD.

Cet algorithme est intéressant car il permet de converger plus rapidement vers la solution optimale par rapport à la méthode à pas optimale.

On a d'ailleurs le résultat de convergence suivant

Proposition 12.3: Convergence de la descente à pas conjugués

Soit \mathbf{A} une matrice symétrique définie et positive. Alors la méthode du gradient à pas conjugués converge en au plus n itérations.

On va maintenant regarder un algorithme de descente de gradient qui utilise aussi des informations sur la dérivée seconde de la fonction.

12.4 Descente de Newton

Dans cette dernière partie, on va regarder comment utiliser les propriétés des fonctions qui sont deux fois différentiables ou, encore mieux, de classe C^2 pour établir d'autres algorithmes.

Mais avant cela, quelques mots sur la méthode de Newton. Elle est utilisée pour trouver un zéro d'une fonction h donnée en utilisant un développement du premier ordre, *i.e.* en approchant localement, en un point x_0 , la fonction par une droite affine

$$h(x) \simeq h(x_0) + h'(x_0)(x - x_0).$$

Partant de cette approximation, pour trouver une solution à $h(x) = 0$, nous devons résoudre l'équation

$$h(x_0) + h'(x_0)(x - x_0) = 0$$

On obtient alors un point $x_1 = x_0 - \frac{h(x_0)}{h'(x_0)}$ et ce point va se rapprocher de la solution $h(x) = 0$.

On peut alors partir à nouveau de ce point afin de trouver une solution à l'équation, qui utilise l'approximation affine de h au point x_1

$$h(x_1) + h'(x_1)(x - x_1) = 0$$

pour déterminer x_2 .

On va ainsi construire itérativement une suite $(x_k)_{k \in \mathbb{N}}$ définie par

$$x_{k+1} = x_k - \frac{h(x_k)}{h'(x_k)}$$

qui convergera vers la solution de l'équation $h(x) = 0$. On illustre graphiquement le processus en Figure 26

Mais comment utiliser ce principe là pour notre descente de gradient ?

Ce que l'on cherche dans le cas d'une descente de gradient, c'est non pas de trouver la valeur \mathbf{u} en laquelle la fonction f étudiée s'annule, mais en quelle valeur \mathbf{u} le gradient $\nabla f(\mathbf{u})$ s'annule !

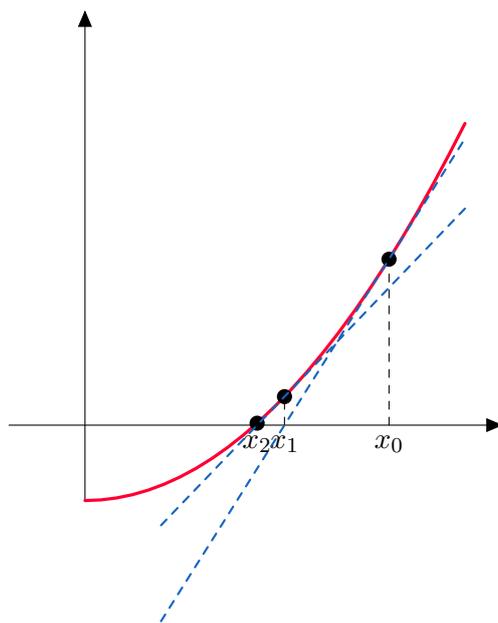


FIGURE 26 – Illustration de la méthode de Newton dans le cas unidimensionnel

On va donc appliquer la méthode de Newton au gradient ∇f de la fonction f .

Ainsi notre mise à jour s'écrit

$$\mathbf{u}_{k+1} = \mathbf{u}_k - (\nabla^2 f(\mathbf{u}_k))^{-1} \nabla f(\mathbf{u}_k).$$

Cet algorithme est bien plus efficace et convergera plus rapidement car il repose sur une approximation plus fine de la fonction.

On pourra même montrer que, pour les formes quadratiques, cet algorithme convergence en une seule itération !

En revanche, elle demande à ce que la fonction soit de classe C^2 et que la matrice hessienne soit inversible. Le calcul de l'inverse d'une matrice est également une opération très coûteuse $\mathcal{O}(d^3)$, rendant la méthode peu efficace dans certains cas.

Pour éviter ce problème d'inversion de matrices, on peut avoir recours à des méthodes d'approximation de la méthode de Newton dites des méthodes de **Quasi-Newton**.

12.5 Approximation de la méthode de Newton

L'inconvénient principale de la méthode de Newton réside dans l'inversion de la hessienne $(\nabla_f^2(\mathbf{x}))^{-1}$. Pour éviter ce problème là, on considère le système suivant

$$\begin{cases} \mathbf{u}_{k+1} &= \mathbf{u}_k - \mathbf{M}_k \nabla f(\mathbf{u}_k), \\ \mathbf{M}_{k+1} &= \mathbf{M}_k + \mathbf{C}_k. \end{cases}$$

L'idée est donc d'approcher l'inverse de la matrice hessienne $(\nabla_f^2(\mathbf{x}_k))^{-1}$ par une matrice \mathbf{M}_k qui sera également définie positive et à laquelle on va ajouter, itérativement, une matrice de correction \mathbf{C}_k afin de corriger l'erreur à la vraie matrice hessienne.

En utilisant un développement de Taylor à l'ordre 1 du gradient $\nabla f(\mathbf{u}_k)$, nous avons

$$\begin{aligned} \nabla f(\mathbf{u}_k) &= \nabla f(\mathbf{u}_{k+1} + (\mathbf{u}_k - \mathbf{u}_{k+1})), \\ &\sim \nabla f(\mathbf{u}_{k+1}) + \nabla^2 f(\mathbf{u}_{k+1})(\mathbf{u}_k - \mathbf{u}_{k+1}). \end{aligned}$$

Nous avons alors l'approximation suivante

$$(\nabla^2 f(\mathbf{u}_{k+1}))^{-1} (\nabla f(\mathbf{u}_{k+1}) - \nabla f(\mathbf{u}_k)) \sim \mathbf{u}_{k+1} - \mathbf{u}_k.$$

Ainsi, en posant $\mathbf{M}_{k+1} = (\nabla^2 f(\mathbf{u}_{k+1}))^{-1}$, $\boldsymbol{\gamma}_k = \nabla f(\mathbf{u}_{k+1}) - \nabla f(\mathbf{u}_k)$ et $\boldsymbol{\delta}_k = \mathbf{u}_{k+1} - \mathbf{u}_k$, on obtient ce que l'on appelle la **Condition de Quasi-Newton** :

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

Regardons comment définir une telle matrice de correction \mathbf{C}_k . On en présentera deux méthodes, une première consistant à définir une correction de rang 1 et une seconde avec une correction de rang 2.

En pratique, c'est la correction de rang 2 qui est la plus souvent utilisée dans les solveurs.

On décrira simplement les méthodes, mais elles seront, si le temps le permet étudiées plus en détail, en TD.

Correction de rang 1 .

La matrice de correction \mathbf{C}_k est supposée de rang égal à 1. On peut donc écrire \mathbf{C}_k comme étant égale à $\mathbf{v}_k \mathbf{v}_k^T$ où $\mathbf{v}_k \in \mathbb{R}^d$.

La règle de mise à jour de la matrice \mathbf{M}_k peut être ainsi être écrite

$$\mathbf{M}_{k+1} = \mathbf{M}_k + \mathbf{v}_k \mathbf{v}_k^T,$$

et la condition de quasi-Newton nous donne

$$\begin{aligned}(\mathbf{M}_k + a_k \mathbf{v}_k \mathbf{v}_k^T) \boldsymbol{\gamma}_k &= \boldsymbol{\delta}_k, \\ \mathbf{M}_k \boldsymbol{\gamma}_k + a_k \mathbf{v}_k \mathbf{v}_k^T \boldsymbol{\gamma}_k &= \boldsymbol{\delta}_k.\end{aligned}$$

En effectuant le produit scalaire avec le vecteur $\boldsymbol{\gamma}_k$ dans l'égalité précédente, on obtient

$$(\boldsymbol{\gamma}_k^T \mathbf{v}_k)^2 = (\boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k)^T \boldsymbol{\gamma}_k.$$

Puis, en utilisant l'égalité, qui permet de faire apparaître le terme $(\boldsymbol{\gamma}_k^T \mathbf{v}_k)^2$

$$\mathbf{v}_k \mathbf{v}_k^T = \frac{\mathbf{v}_k (\mathbf{v}_k^T \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T \mathbf{v}_k) \mathbf{v}_k^T}{(\mathbf{v}_k^T \boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T \mathbf{v}_k)} = \frac{\mathbf{v}_k \mathbf{v}_k^T \boldsymbol{\gamma}_k (\mathbf{v}_k \mathbf{v}_k^T \boldsymbol{\gamma}_k)^T}{(\mathbf{v}_k \boldsymbol{\gamma}_k)^2},$$

où la dernière égalité utilise le fait que $(\mathbf{ab})^T = \mathbf{b}^T \mathbf{a}^T$ et que $\mathbf{a}^T \mathbf{b} = \mathbf{b}^T \mathbf{a}$.

En repartant du fait que $\mathbf{v}_k \mathbf{v}_k^T \boldsymbol{\gamma}_k = \boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k$ et que $(\boldsymbol{\gamma}_k^T \mathbf{v}_k)^2 = (\boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k)^T \boldsymbol{\gamma}_k$ dans l'égalité précédente, on obtient la formule dite de *Broyden*.

Ce qui nous conduit à un premier algorithme d'approximation dit de *Broyden*.

Définition 12.5: Broyden Algorithm

Soit f une fonction définie sur un ensemble D de \mathbb{R}^n et $\varepsilon > 0$.

Alors l'algorithme de Broyden

1. choisir \mathbf{u}_0 pour initialiser l'algorithme et $\mathbf{M}_0 = \mathbf{I}$.
2. tant que $\|\nabla f(\mathbf{u}_k)\| \geq \varepsilon$
 - (a) set $\rho_k = \arg \min_{\rho \in \mathbb{R}} f(\mathbf{u}_k - \rho \mathbf{M}_k \nabla f(\mathbf{u}_k))$,
 - (b) set $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \mathbf{M}_k \nabla f(\mathbf{u}_k)$,
 - (c) set $\mathbf{M}_{k+1} = \mathbf{M}_k + \frac{(\boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k)(\boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k)^T}{(\boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k)^T \boldsymbol{\gamma}_k}$,

L'inconvénient de cette méthode est qu'elle ne garantit pas le caractère défini positif de la matrice \mathbf{M}_k au cours des itérations et qu'elle devient instable dès lors que $(\boldsymbol{\delta}_k - \mathbf{M}_k \boldsymbol{\gamma}_k)^T \boldsymbol{\gamma}_k$ est proche de 0.

On présente une autre méthode d'approximation qui elle n'utilise pas le fait d'approcher l'inverse de la matrice hessienne, mais directement la matrice hessienne et à l'aide d'une approximation de rang 2.

Correction de rang 2 L'idée reste semblable, mais au lieu d'utiliser une matrice de correction \mathbf{C}_k de rang 1, on va l'écrire comme la somme deux matrices de rang 1 chacun, mais dans des bases différentes, on va donc effectuer une correction de rang 2.

$$\mathbf{C}_k = \alpha \mathbf{v}_k \mathbf{v}_k^T + \beta \mathbf{w}_k \mathbf{w}_k^T.$$

Dans ce qui suit, on va également chercher à approcher, non plus l'inverse de la hessienne, mais d'abord, la matrice hessienne elle-même, que l'on va noter H_k par soucis de clarté. On reviendra ensuite à l'inverse de la matrice Hessienne. Ce qui nous donne.

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \mathbf{C}_k = \mathbf{H}_k + \mathbf{v}_k \mathbf{v}_k^T + \mathbf{w}_k \mathbf{w}_k^T.$$

En appliquant à nouveau la Condition de Quasi-Newton, qui s'écrit cette fois

$$\mathbf{H}_{k+1} \boldsymbol{\delta}_k = \boldsymbol{\gamma}_k,$$

on obtient la formule dite de *Davidon-Fletcher-Powell*

$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\boldsymbol{\gamma}_k \boldsymbol{\gamma}_k^T}{\boldsymbol{\gamma}_k^T \boldsymbol{\delta}_k} - \frac{\mathbf{H}_k \boldsymbol{\delta}_k \boldsymbol{\delta}_k^T \mathbf{H}_k^T}{\boldsymbol{\delta}_k^T \mathbf{H}_k \boldsymbol{\delta}_k}.$$

Mais ce qui nous intéresse, c'est une façon d'avoir l'inverse de la matrice directement, pour cela, il faudrait appliquer la formule de *Sherman-Morrison* qui dit que pour toute matrice inverible \mathbf{A} et pour tout vecteur \mathbf{x}, \mathbf{y} tels que $1 + \mathbf{x}^T \mathbf{A} \mathbf{y} \neq 0$, nous avons

$$(\mathbf{A} + \mathbf{y} \mathbf{x}^T)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{y} \mathbf{x}^T \mathbf{A}^{-1}}{1 + \mathbf{x}^T \mathbf{A}^{-1} \mathbf{y}}.$$

Cela conduit à une règle de mise à jour qui donne lieu à l'algorithme de *Broyden-Fletcher-Goldfarb-Shanno* (BFGS).

Définition 12.6: BFGS Algorithm

Soit f une fonction définie sur un ensemble D de \mathbb{R}^n et $\varepsilon > 0$.

Alors l'algorithme de BFGS est donné par

1. choisir \mathbf{u}_0 pour initialiser l'algorithme, $\mathbf{M}_0 = \mathbf{I}$ and $k = 0$
2. tant que $\|\nabla f(\mathbf{u}_k)\| \geq \varepsilon$
 - (a) poser $\rho_k = \arg \min_{\rho \in \mathbb{R}} f(\mathbf{u}_k - \rho \mathbf{M}_k \nabla f(\mathbf{u}_k))$,
 - (b) poser $\mathbf{u}_{k+1} = \mathbf{u}_k - \rho_k \mathbf{M}_k \nabla f(\mathbf{u}_k)$,
 - (c) poser $\mathbf{M}_{k+1} = \mathbf{M}_k + \left[1 + \frac{\langle \mathbf{M}_k \boldsymbol{\gamma}_k, \boldsymbol{\gamma}_k \rangle}{\langle \boldsymbol{\delta}_k, \boldsymbol{\gamma}_k \rangle} \right] \frac{\langle \boldsymbol{\delta}_k, \boldsymbol{\delta}_k \rangle}{\langle \boldsymbol{\delta}_k, \boldsymbol{\gamma}_k \rangle} - \frac{\boldsymbol{\delta}_k \boldsymbol{\gamma}_k^T \mathbf{M}_k + \mathbf{M}_k \boldsymbol{\gamma}_k \boldsymbol{\delta}_k^T}{\langle \boldsymbol{\delta}_k, \boldsymbol{\gamma}_k \rangle}$.

Nous pourrions également définir une expression plus simple de BFGS, en cherchant directement à approcher l'inverse de la matrice Hésienne à l'aide formule de Davidon-Fletcher-Powell.

La réalité est en fait beaucoup plus complexe car la plupart des problèmes que l'on cherche à résoudre font l'objet de **contraintes**. On peut, par exemple, chercher à maximiser la rentabilité d'une stratégie marketing, sous contraintes de coûts.

Cela vous conduira plus tard à étudier des problèmes de la forme

$$\begin{aligned} & \min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}), \\ \text{t.q. } & g(\mathbf{x}) \leq \mathbf{0}, \\ & h(\mathbf{x}) = \mathbf{0}, \end{aligned}$$

où f est une fonction à valeurs réelles, g et h sont des fonctions à valeurs vectorielles que l'on appelle des contraintes.

13 Intégration

Dans cette dernière section, nous présentons le calcul d'intégrale pour des fonctions dépendantes de plusieurs variables. Nous verrons, entre autre, à quelles conditions nous pouvons intervertir l'ordre dans lequel nous effectuons les différentes intégrales afin que le résultat obtenu soit juste.

Mais avant cela, nous revenons sur l'intégrale en une dimension pour introduire la notion de *fonction intégrable* que l'on retrouvera pour les fonctions à plusieurs variables.

13.1 Fonctions réelles intégrables

Dans la précédente partie, nous avons effectué plusieurs calculs d'intégrales sans jamais nous soucier de la bonne définition de ces dernières. Nous avons simplement exigé que les fonctions que l'on cherche à intégrer sont bien continues et, la plupart du temps, nous cherchions à intégrer les fonctions sur des segments de la forme $[a, b]$ où $a < b \in \mathbb{R}$.

La réalité est en fait plus complexe que cela et toutes les fonctions, même continues ne sont pas intégrables. **Mais qu'est-ce que cela signifie être intégrable ?**

Définition 13.1: Fonction intégrable

Une fonction f définie et continue (une hypothèse moins forte est suffisante en pratique) sur un intervalle I de \mathbb{R} est intégrable si l'on a

$$\int_I |f(t)| dt < +\infty$$

Dit autrement, **une fonction est intégrable si l'intégrale de sa valeur absolue est finie.**

Dans la pratique, toutes les fonctions que l'on rencontrera sont intégrables et, l'an dernier, nous nous posions peu la question car toutes les fonctions étudiées étaient continues et définies sur des segments $[a, b]$, ainsi, par Théorème de Weierstrass, cette fonction est alors bornée ce qui permettait d'avoir

$$\int_a^b |f(t)| dt \leq \sup_{x \in [a, b]} |f(x)| |b - a| < +\infty.$$

Mais que se passe-t-il si on souhaite maintenant intégrer sur d'autres ensembles comme des intervalles ouverts ?³ C'est par exemple, ce que l'on cherche à faire lorsque l'on

3. C'est ce que l'on appellera des **intégrales impropres** car les bornes de l'intervalle de d'intégrations font intervenir une valeur qui n'appartient pas à l'ensemble de définition de l'intégrande

souhaite évaluer l'intégrale suivante, permettant de calculer l'espérance d'une variable aléatoire X dont les valeurs sont distribuées selon une loi gaussienne :

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} tf(t) dt = \int_{-\infty}^{+\infty} t \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(t-\mu)^2\right) dt.$$

Ici on intègre sur \mathbb{R} tout entier ! Mais est-ce bien valable ? Il se trouve que oui dans le cas présent (mais on ne cherchera pas à expliquer pourquoi dans le cas de ce cours). Mais ce n'est pas toujours le cas, et cela peut dépendre aussi de l'intervalle sur lequel la fonction est intégrée.

Exemple 13.1. Soit $\alpha > 1$, alors la fonction $f : x \mapsto \frac{1}{x^\alpha}$ est intégrable sur $[1, +\infty[$.

En effet, la fonction f est continue et positive sur cet intervalle. Il suffit alors de vérifier que cette dernière converge. Calculons son intégrale pour cela.

$$\int_1^{+\infty} t^{-\alpha} dt = \left[\frac{1}{1-\alpha} t^{1-\alpha} \right]_{t=1}^{t \rightarrow +\infty} = \frac{0-1}{1-\alpha} = \frac{1}{\alpha-1}.$$

L'avant dernière inégalité est due au fait que $1-\alpha < 0$ donc $t^{1-\alpha}$ tend vers 0 lorsque $t \rightarrow +\infty$.

Cela montre donc que lorsque $\alpha < 1$, nous avons $1-\alpha > 0$ et cette fois-ci $t^{1-\alpha}$ tend vers $+\infty$ lorsque $t \rightarrow +\infty$ et l'intégrale ne converge donc plus.

Exemple 13.2. Soit $\alpha < 1$, alors la fonction $f : x \mapsto \frac{1}{x^\alpha}$ est intégrable sur $]0, 1]$.

En effet, la fonction f est continue et positive sur cet intervalle. Il suffit alors de vérifier que cette dernière converge. Calculons son intégrale pour cela.

$$\int_0^1 t^{-\alpha} dt = \left[\frac{1}{1-\alpha} t^{1-\alpha} \right]_{t \rightarrow 0}^{t=1} = \frac{1-0}{1-\alpha} = \frac{1}{\alpha-1}.$$

L'avant dernière inégalité est due au fait que $1-\alpha > 0$ donc $t^{1-\alpha}$ tend vers 0 lorsque $t \rightarrow 0$.

Cela montre également que lorsque $\alpha > 1$, nous avons $1-\alpha < 0$ et cette fois-ci $t^{1-\alpha}$ tend vers $+\infty$ lorsque $t \rightarrow 0$ et l'intégrale ne converge donc plus.

La fonction $x \mapsto \frac{1}{x^\alpha}$ est un exemple parmi d'autres, nous pourrions dresser des constats analogues avec la fonction $x \mapsto e^x$ qui sera intégrable sur \mathbb{R}_- mais sur \mathbb{R}_+ par exemple.

Il y a même une dernière sorte d'intégrale, on les appelle *intégrale semi-convergente*, ce sont des fonctions qui ne sont pas intégrables au sens de la définition, mais dont l'intégrale est bien convergente, *i.e.* admet une valeur finie.

Exemple 13.3. *Considérons la fonction f définie pour tout réel $x > 0$ par*

$$f(x) = \frac{\sin(x)}{x}$$

et étudions l'intégrale de cette dernière sur \mathbb{R}_+^ . On va montrer que cette fonction n'est pas intégrable, *i.e.*,*

$$\int_0^{+\infty} |f(t)| dt = +\infty.$$

mais qu'elle est bien semi-convergente donc que,

$$\int_0^{+\infty} f(t) dt < +\infty.$$

- **Non intégrable** : on considère la suite $(u_n)_{n \in \mathbb{N}}$ définie pour tout $n \in \mathbb{N}$ par

$$\int_{n\pi}^{(n+1)\pi} \left| \frac{\sin(t)}{t} \right| dt \geq \frac{1}{(n+1)\pi} \int_{n\pi}^{(n+1)\pi} |\sin(t)| dt = \frac{1}{(n+1)\pi},$$

$$\text{car } \int_{n\pi}^{(n+1)\pi} |\sin(t)| dt = \int_0^\pi |\sin(u)| du = 1.$$

$$\text{On a donc } u_n \geq \frac{1}{(n+1)\pi}.$$

$$\text{Or } \int_0^{+\infty} |f(t)| dt = \sum_{n=0}^{\infty} u_n = \frac{1}{\pi} \sum_{n=0}^{\infty} \frac{1}{n+1} = +\infty.$$

On a donc montré que la fonction n'est pas intégrable.

- **Semi-convergence** : pour cela on va simplement utiliser l'intégration par partie.

$$\int_0^{+\infty} \frac{\sin(t)}{t} dt = \left[\frac{1 - \cos(t)}{t} \right]_{t=0}^{t=+\infty} + \int_0^{+\infty} \frac{1 - \cos(t)}{t^2} dt.$$

Nous avons $\left[\frac{1 - \cos(t)}{t} \right]_{t=0}^{t=+\infty} = 0$ (il suffit de d'étudier la limite en $+\infty$ et, en 0, de se souvenir d'un équivalent de $1 - \cos(t)$ au voisinage de 0.) et lorsque $t \rightarrow +\infty$, $0 < \frac{1 - \cos(t)}{t} < \frac{1}{t}$ et on en déduit que la limite tend vers 0.

Pour le second terme, il suffit de remarquer que $\frac{1 - \cos(t)}{t^2}$ tend vers 1 au

voisinage de 0 (développement limité) et que pour tout t nous avons aussi $\frac{1 - \cos(t)}{t^2} \leq \frac{1}{t^2}$ dont l'intégrale sur $[1, +\infty[$ est finie (elle vaut 1).

Donc notre intégrale est bien semi-convergente mais pas convergente.

Les intégrales impropres seront très nombreuses en probabilité, mais heureusement elles seront toutes convergentes (à quelques exceptions près)⁴. Cette notion de convergence va se retrouver dans un théorème d'intégration des fonctions à plusieurs variables : le Théorème de Fubini-Tonelli.

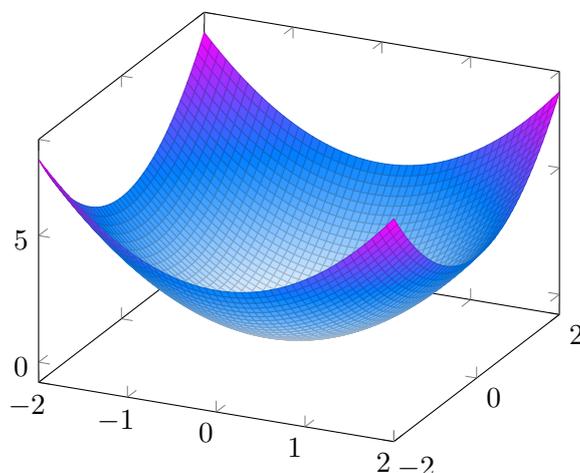
13.2 Fonctions à plusieurs variables

L'intégration des fonctions à plusieurs variables est analogue à l'intégration des fonctions à une seule variable, le calcul des primitives se fait exactement de la même façon en considérant que seule la variable d'intérêt est réellement une variable alors que toutes les autres seront fixes.

Imaginons que l'on souhaite intégrer la fonction f définie sur \mathbb{R}^2 par

$$\forall x, y \in \mathbb{R}, f(x, y) = x^2 + x^2$$

que l'on peut représenter graphiquement comme suit :



4. En réalité les intégrales foisonnent dans la théorie du signal que l'on retrouvera Analyse de Fourier, dans l'analyse de sons, ou encore l'étude des ondes comme les ECG ou les EEG. Les intégrales qui entrent vont être de différentes formes et peuvent aussi faire intervenir des paramètres ce qui posera des questions de définition, de dérivabilité de telles intégrales. D'autres questions plus exotiques impliquent l'interversion entre un calcul de limite et une intégrale, ou encore une somme infinie et une intégrale lorsque l'on s'intéresse à une suite de fonctions.

Comment déterminer la valeur d'une telle intégrale si on cherche à l'évaluer sur un rectangle de la forme $[a, b] \times [c, d]$ ou encore sur \mathbb{R}^2 ? Est-ce que cette dernière existe? Quel sens pouvons nous donner à cette intégrale?

On commence par étudier l'intégrale sur un rectangle avant de s'intéresser à des intégrales plus générales. Les exemples sont données pour des fonctions à deux variables, mais se généralisent pour des fonctions à plusieurs variables.

Intégrales sur un rectangle

On commence par le cas simple où les fonctions sont définies sur des rectangles avec le premier résultat.

Théorème 13.1: Théorème de Fubini-Tonelli sur un rectangle

Soit a, b, c et d trois nombres réels et soit f une fonction continue sur $[a, b] \times [c, d]$ à valeurs dans \mathbb{R} .

Alors

$$\int_a^b \left(\int_c^d f(x, y) dy \right) dx = \int_c^d \left(\int_a^b f(x, y) dx \right) dy.$$

Ce théorème nous dit simplement que si on intègre une fonction sur un rectangle dont les bords sont compris, l'ordre d'intégration importe peu.

Exemple 13.4. On considère la fonction $f : R \rightarrow \mathbb{R}$ définie par $f(x, y) = ye^{xy}$ où $R = [1, 2] \times [0, 2]$.

Déterminons la valeur de $I \int_R f(x, y) dx dy = \int_0^2 \int_1^2 f(x, y) dx dy$.

On peut appliquer le théorème de Fubini et dire que

$$I = \int_0^2 \left(\int_1^2 ye^{yx} dx \right) dy = \int_0^2 [e^{xy}]_{x=1}^{x=2} dy = \int_0^2 e^{2y} - e^y dy.$$

En intégrant maintenant par rapport à la variable y , nous obtenons

$$I = \left[\frac{1}{2}e^{2y} - e^y \right]_0^2 = \frac{1}{2}e^4 - e^2 + \frac{1}{2}.$$

Intégrales sur des ensembles plus généraux Nous ne sommes pas toujours amenés à intégrer sur des rectangles comme nous avons pu le voir en début de cette section dans

le cas de la dimension 1 mais aussi sur des surfaces ou volumes plus généraux comme \mathbb{R}^2 par exemple. Parfois les bornes des intervalles d'intégration sont elles mêmes définies par des équations.

Proposition 13.1: Intégrale sur des domaines généraux

Soient $a < b \in \mathbb{R}$ et $[a, b]$ un segment de \mathbb{R} . Soient u et v deux fonctions continue sur $[a, b]$ telles que pour tout $x \in [a, b]$ nous avons $u(x) \leq v(x)$.

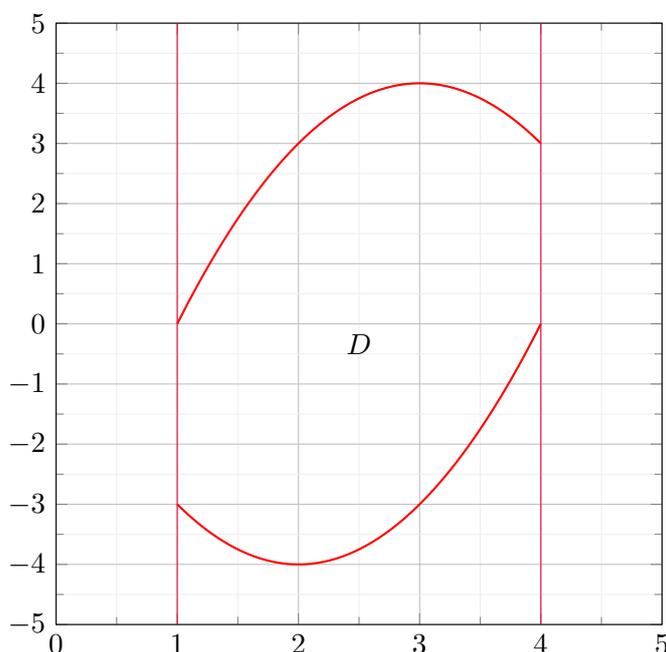
Considérons D un domaine de \mathbb{R}^2 défini par $D = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b, u(x) \leq y \leq v(x)\}$.

Si f est une fonction continue de D dans \mathbb{R} , alors f est intégrable sur D et on a

$$\int_D f(x, y) = dx dy = \int_a^b \left(\int_{u(x)}^{v(x)} f(x, y) dy \right) dx.$$

Exemple 13.5. Soit D le domaine de \mathbb{R}^2 compris entre les droites d'équations $x = 1$, $x = 4$ et les courbes d'équations $y = (x - 2)^2 - 4$ et $y = -(x - 3)^2 + 4$. On cherche à évaluer l'intégrale sur D de la fonction f définie sur D par $f(x, y) = 3x - 2y + 1$.

On commence par représenter le domaine D d'intégration de notre fonction.



En posant $u(x) = (x - 2)^2 - 4$ et $v(x) = -(x - 3)^2 + 4$, il suffit d'appliquer la

proposition précédente pour calculer notre intégrale.

$$\begin{aligned}
 I &= \int_D f(x, y) \, dx \, dy, \\
 &= \int_1^4 \left(\int_{u(x)}^{v(x)} 3x - 2y + 1 \, dy \right) dx, = \int_1^4 [3xy - y^2 + y]_{u(x)}^{v(x)} dx, \\
 &= \int_1^4 2x^3 2x^2 + 55x 30 \, dx, \\
 &= \left[\frac{1}{2} x^4 \frac{2}{3} x^3 + \frac{55}{2} x^2 30x \right]_{x=1}^{x=4} dx, \\
 &= 153
 \end{aligned}$$

Le théorème de Fubini précédemment énoncé peut également se généraliser comme suit

Théorème 13.2: Théorème de Fubini-Tonelli sur \mathbb{R}^2

Soit X et Y deux intervalles de \mathbb{R} . Soit f une fonction définie et continue sur $X \times Y$. Si la fonction $|f|$ est intégrable, *i.e.*, si on a

$$\int \int_{X \times Y} |f(x, y)| \, dx \, dy < +\infty,$$

alors

$$\int_X \left(\int_Y f(x, y) \, dy \right) dx = \int_Y \left(\int_X f(x, y) \, dx \right) dy.$$

Ce théorème peut très facilement se généraliser à plus que deux intégrales, mais pour un ensemble de $n > 0$ intégrales sur des intervalles ouverts. On peut l'utiliser par exemple, pour évaluer l'intégrale suivante qui intervient dans la distribution gaussienne.

$$\int_{-\infty}^{+\infty} e^{-x^2} \, dx.$$

Exemple 13.6. Soit f la fonction $x \mapsto e^{-x^2}$ définie et continue sur \mathbb{R} . Cette fonction est intégrable mais il est impossible de trouver une primitive de cette fonction, *i.e.*, cette dernière n'est pas exprimable à l'aide des fonctions usuelles.

En revanche on peut estimer l'intégrale $\int_{-\infty}^{+\infty} e^{-x^2} \, dx$ en passant par une intégrale double.

En effet, si on note $I = \int_{-\infty}^{+\infty} e^{-x^2} dx$ et que l'on considère $(x, y) \mapsto e^{-x^2-y^2}$, également continue sur \mathbb{R}^2 et intégrable, alors

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-x^2-y^2} dx dy &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-x^2-y^2} dx \right) dy, \\ &\downarrow \text{l'exponentielle d'une somme est égal au produit des exponentielles} \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-x^2} e^{-y^2} dx \right) dy, \\ &\downarrow e^{-y^2} \text{ est indépendant de } x \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} e^{-x^2} dx \right) e^{-y^2} dy, \\ &= \left(\int_{\mathbb{R}} e^{-x^2} dx \right) \left(\int_{\mathbb{R}} e^{-y^2} dy \right), \\ &\downarrow \text{chaque intégrale est égale à } I \\ &= I^2. \end{aligned}$$

En évaluant ainsi I^2 on sera en mesure de trouver la valeur de I ! On peut montrer, on détaillera les calculs dans un exemple ultérieur, que $\int_{\mathbb{R}^2} e^{-x^2-y^2} dx dy = \pi$ donc $I = \sqrt{\pi}$.

On peut également utiliser ce résultat pour montrer qu'une fonction n'est pas intégrable et montrant que l'on ne peut pas échanger l'ordre d'intégration sans prendre un minimum de précautions !

Exemple 13.7. Considérons la fonction f définie de $\mathbb{R}_+ \times [0, 1]$ par

$$f(x, y) = 2e^{-2xy} - e^{-xy}.$$

La fonction f est continue sur $\mathbb{R}_+ \times [0, 1]$ mais on va voir qu'elle n'est pas intégrable sur cette région. En effet, si on intègre d'abord par rapport à x .

$$\int_0^1 \left(\int_{\mathbb{R}_+} 2e^{-2xy} - e^{-xy} dx \right) dy = \int_0^1 \left[\frac{-1}{y} (e^{-2xy} - e^{-xy}) \right]_{x=0}^{x \rightarrow +\infty} dy = \int_0^1 0 dy = 0.$$

Mais si on intègre cette fois-ci par rapport à y

$$\int_{\mathbb{R}_+} \left(\int_0^1 2e^{-2xy} - e^{-xy} dy \right) dx = \int_{\mathbb{R}_+} \left[\frac{-1}{x} (e^{-2xy} - e^{-xy}) \right]_{y=0}^{y=1} dx = \int_{\mathbb{R}_+} \frac{e^{-x} - e^{-2x}}{x} dx > 0.$$

Cette dernière intégrale est positive car la fonction à intégrer est positive sur \mathbb{R}_+ .

On en déduit donc que $|f|$ n'est pas convergente, donc la fonction f n'est pas intégrable.

Intégrales multiples Les résultats précédents restent valables lorsque les fonctions étudiées sont définies sur des sous ensemble des \mathbb{R}^n où $n > 2$ et les résultats sont notamment employés pour évaluer le volume d'une boule en dimension n !

Un excellent exercice consiste à évaluer le volume V_n de cette boule unité (ou de la boule de rayon r) pour tout entier $n > 0$, c'est très technique et cela demande de faire des raisonnements par récurrence. Mais on a déjà les résultats connus suivants pour les premières valeurs de n

$$V_1(r) = 2r, \quad V_2(r) = 2\pi r, \quad V_3(r) = \frac{4}{3}\pi r^3.$$

Mais il faut, pour cela, regarder quel est le lien entre $V_{n+1}(r)$ et $V_n(r)$ à l'aide du théorème de Fubini ! .

Il ne nous reste plus qu'à regarder comment opérer un changement de variable !

13.3 Changement de variable

Dans cette dernière partie, on regarde comment effectuer un changement de variable dans le cas d'intégrales multiples. Comme dans le cas unidimensionnel, il va falloir changer les opérateurs dx et dy qui entrent en jeu. Avant d'énoncer le résultat principal, on rappelle la définition de Jacobien qui interviendra dans la formule de changement de variable.

Définition 13.2: Jacobien

Soit ϕ une application de \mathbb{R}^2 dans \mathbb{R}^2 définie par

$$\phi(u, v) = (x(u, v), y(u, v)).$$

On supposera que les fonctions x et y admettent des dérivées partielles sur leur ensemble de définition. Ainsi, le *Jacobien* de ϕ au point (u, v) , que l'on note $Jac_\phi(u, v)$ est le déterminant de la matrice des dérivées partielles de ϕ , *i.e.*,

$$Jac_\phi(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u}(u, v) & \frac{\partial x}{\partial v}(u, v) \\ \frac{\partial y}{\partial u}(u, v) & \frac{\partial y}{\partial v}(u, v) \end{vmatrix}$$

Pour que ce Jacobien existe, il faut que la fonction ϕ soit différentiable et qu'elle réalise une bijection, *i.e.* qu'il s'agisse d'un difféomorphisme⁵.

Théorème 13.3: Changement de variables

Soient U et V deux ouverts^a de \mathbb{R}^2 et soit $\phi : U \rightarrow V$ un difféomorphisme de classe C^1 de U sur V .

Alors pour toute fonction $f : V \rightarrow \mathbb{R}^p$ continue et intégrable, on a :

$$\int_V f(x, y) dx dy = \int_U f(\phi(u, v)) |Jac_\phi(u, v)| du dv.$$

^a. ce sont des ensembles qui ne contiennent pas leur frontière, comme une boule sans ses bords, ou un intervalle ouvert des deux côtés, dans le cas unidimensionnel.

Remarquons que le changement de variable fait intervenir un Jacobien où l'on exprime **les anciennes variables en fonction des nouvelles variables**. Regardons un premier exemple comme application

Exemple 13.8. *Considérons la fonction $f : D \rightarrow \mathbb{R}^2$ définie par $f(x, y) = \frac{1}{1+x+y}$,*

où $D = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1, 0 \leq x+y \leq 1\}$.

On cherche à évaluer l'intégrale de la fonction f sur le domaine D . Pour cela, on effectuera le changement de variable suivant :

$$\begin{aligned} u &= x, \\ v &= x + y. \end{aligned}$$

5. il s'agit d'une application bijective différentiable, dont la réciproque est elle-même différentiable.

Remarquons que ce changement de variable définit la fonction ϕ^{-1} et non la fonction ϕ . Pour obtenir la fonction ϕ , il suffit d'exprimer les anciennes variables x et y en fonction des nouvelles u et v . On obtient donc

$$\begin{aligned}x &= u, \\y &= v - u.\end{aligned}$$

Notre Jacobien est alors égal à

$$\text{Jac}_\phi(u, v) = \begin{vmatrix} \frac{\partial x}{\partial u}(u, v) & \frac{\partial x}{\partial v}(u, v) \\ \frac{\partial y}{\partial u}(u, v) & \frac{\partial y}{\partial v}(u, v) \end{vmatrix} = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = |1| = 1.$$

De plus, lorsque x parcourt l'intervalle $[0, 1]$, la variable u . Comme on a $v = x + y$, alors v parcourt l'intervalle $[0, 1]$ également.

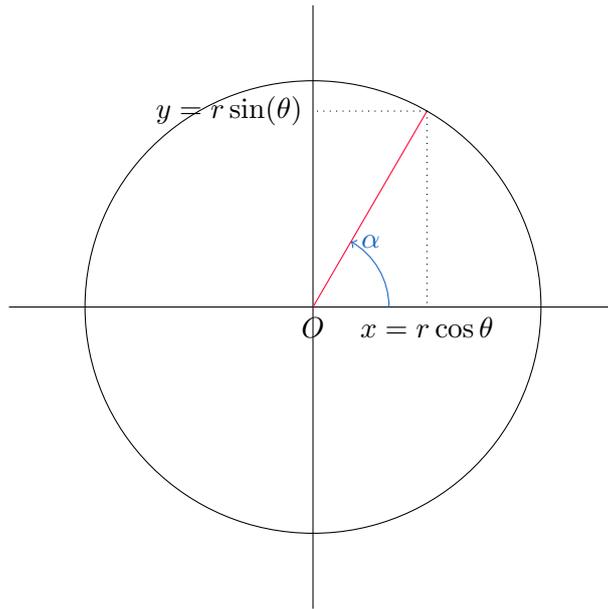
$$\int_D \frac{1}{1+x+y} dx dy = \int_0^1 \int_0^1 \frac{1}{1+v} du dv = \left(\int_0^1 \frac{1}{1+v} dv \right) \left(\int_0^1 du \right) = \ln(2).$$

Nous aurions également pu calculer cette intégrale sans changement de variable.

Ce résultat, plutôt pratique, va également nous permettre de déterminer l'intégrale calculée à dans l'Exemple 13.6.

Exemple 13.9. Reprenons le calcul de l'intégrale $\int_{\mathbb{R}^2} e^{-x^2-y^2} dx dy$. On ici effectuer un changement de variable dit en coordonnées polaires, i.e., on va poser

$$\begin{aligned}x &= r \cos(\theta), \\y &= r \sin(\theta).\end{aligned}$$



Ce changement de variable est bien un C^1 difféomorphisme et le jacobien est

$$Jac_{\phi}(u, v) = \begin{vmatrix} \frac{\partial x}{\partial r}(r, \theta) & \frac{\partial x}{\partial \theta}(r, \theta) \\ \frac{\partial y}{\partial r}(r, \theta) & \frac{\partial y}{\partial \theta}(r, \theta) \end{vmatrix} = \begin{vmatrix} \cos(\theta) & -r \sin(\theta) \\ \sin(\theta) & r \cos(\theta) \end{vmatrix} = |-r(\cos^2(\theta) + \sin^2(\theta))| = r.$$

De plus, lorsque x et y parcourent \mathbb{R}^2 tout entier, il faut, si on reprend le chemin que l'on a vu au dessus, que r parcourent toutes les valeurs de 0 à $+\infty$ et que θ parcourent l'intervalle $[-\pi, \pi]$ (ou $[0, 2\pi]$). Ce qui nous donne

$$\int_{\mathbb{R}^2} e^{-x^2-y^2} dx dy = \int_0^{2\pi} \int_0^{+\infty} e^{-r^2} r dr d\theta.$$

Par Fubini (la nouvelle fonction est intégrable), on peut écrire

$$\left(\int_0^{+\infty} e^{-r^2} r dr \right) \left(\int_0^{2\pi} d\theta \right) = 2\pi \left[-\frac{1}{2} e^{-r^2} \right]_{r=0}^{r \rightarrow +\infty} = \pi.$$