

Fascicule de cours de l'U.E. UL2CI007
« Des données aux modèles physicochimiques »

Responsables de l'U.E. : Patrick Gredin et Julien Toulouse

Licence de Chimie
Sorbonne Université

23 août 2025

Table des matières

Introduction générale	7
1 Premier exemple : cinétique chimique	7
2 Deuxième exemple : spectroscopie UV-Visible	11
I Fonctions et données	15
I.1 Fonctions	17
1 Fonctions d'une ou de plusieurs variables	17
1.1 Fonction d'une variable réelle	17
1.2 Fonction de plusieurs variables réelles	18
2 Systèmes de coordonnées	22
2.1 Coordonnées en dimension 2 : coordonnées polaires	22
2.2 Coordonnées en dimension 3 : coordonnées sphériques	23
3 Champs de vecteurs	24
I.2 Données	27
1 Représentation et visualisation d'une série de données	27
2 Interpolation d'une série de données	29
2.1 Interpolation linéaire	29
2.2 Interpolation polynomiale	30
2.3 Introduction à l'interpolation par splines	31
II Outils élémentaires d'analyse des fonctions et des données	33
II.1 Outils d'analyse différentielle	35
1 Dérivées	35
1.1 Dérivées d'une fonction d'une variable réelle	35
1.2 Dérivées d'une fonction de plusieurs variables réelles	38
2 Développements limités et différentielles	41
2.1 Développements limités	41
2.2 Différentielles	43
3 Opérateurs différentiels	45
4 Dérivées numériques	52
II.2 Outils d'analyse intégrale	57
1 Intégrale d'une fonction d'une variable réelle	57
1.1 Définition	57
1.2 Propriétés essentielles	59
1.3 Primitives et théorème fondamental de l'analyse	60

1.4	Autres théorèmes	62
2	Intégrale d'une fonction de plusieurs variables réelles	63
2.1	Concept général et notation	63
2.2	Intégrale double	64
2.3	Intégrale triple	66
2.4	Intégrale en coordonnées polaires et en coordonnées sphériques	68
3	Intégrale numérique	74
3.1	Méthodes des rectangles	75
3.2	Méthodes des trapèzes	76
3.3	Méthode de Simpson (optionnel)	77
4	Transformation de Fourier	78
4.1	Définition	78
4.2	Exemple simple	80
4.3	Applications (optionnel)	81
II.3	Outils d'analyse statistique	85
1	Variable aléatoire continue	85
1.1	Définition	85
1.2	Fonction de répartition	87
1.3	Espérance, variance et écart-type	89
1.4	Variable aléatoire centrée et réduite	91
1.5	Couple de variables aléatoires (optionnel)	93
1.6	Deux lois de probabilité continues importantes	95
2	Théorème central limite et intervalles de confiance	98
2.1	Théorème central limite	98
2.2	Intervalles de confiance	100
II.4	Outils d'analyse matricielle	105
1	Matrices et systèmes d'équations linéaires	105
1.1	Aspect historique	105
1.2	Représentation matricielle d'un système d'équations linéaires	106
1.3	Quelques définitions générales sur les matrices	107
1.4	Quelques définitions pour le cas des matrices carrées	110
1.5	Déterminant d'une matrice carrée	112
1.6	Calcul de l'inverse d'une matrice carrée	114
2	Valeurs propres et vecteurs propres d'une matrice	115
2.1	Définition et calculs des valeurs propres et vecteurs propres	115
2.2	Diagonalisation d'une matrice	119
3	Applications linéaires	122
3.1	Représentation matricielle des applications linéaires	122
3.2	Opérations de symétrie d'une molécule	124
III	Des données aux modèles	127
III.1	Modèles de régression	129
1	Idée générale des méthodes de régression	129
2	Régression linéaire	131
2.1	Description de la méthode de régression linéaire	131

2.2	Détermination des paramètres a_0 et b_0	131
2.3	Coefficient de détermination R^2	134
2.4	Application à la cinétique chimique	135
3	Régression polynomiale	137
3.1	Description de la méthode de régression polynomiale de degré 2	137
3.2	Détermination des paramètres a_0 , b_0 et c_0	137
III.2	Modèles physicochimiques	141
1	Modèles différentiels d'ordre 1	141
1.1	Équation différentielle linéaire d'ordre 1 sans second membre	141
1.2	Modèle cinétique d'une réaction chimique unidirectionnelle $A \longrightarrow B$	143
1.3	Équation différentielle linéaire d'ordre 1 avec second membre	143
1.4	Modèle cinétique d'une réaction chimique bidirectionnelle $A \rightleftharpoons B$	144
2	Modèles différentiels d'ordre 2	145
2.1	Équation différentielle linéaire d'ordre 2 à coefficients constants sans second membre	145
2.2	Modèle de l'oscillateur harmonique	146
2.3	Équation différentielle linéaire d'ordre 2 à coefficients constants avec second membre	147
2.4	Modèle de l'oscillateur harmonique forcé	148
3	Modèles différentiels aux dérivées partielles	149
3.1	Équation de la diffusion	150
3.2	Équation de Schrödinger	152

Ce document sert de support au cours de l'U.E. UL2CI007. Tout le contenu de ce document ne sera pas abordé dans les séances de Cours-TD. Les parties en gris seront traitées en priorité. Il est néanmoins fortement conseillé de lire également les parties non abordées en Cours-TD.

Introduction générale

Dans les sciences naturelles, et en particulier en chimie, on récolte souvent des **données** à l'aide de mesures expérimentales ou de calculs théoriques. Ces données peuvent être **représentées** par des **fonctions** mathématiques que l'on doit **visualiser** de manière adéquate. Pour extraire des informations quantitatives pertinentes de ces données et fonctions, on peut alors utiliser divers **outils mathématiques élémentaires d'analyse de données et de fonctions** :

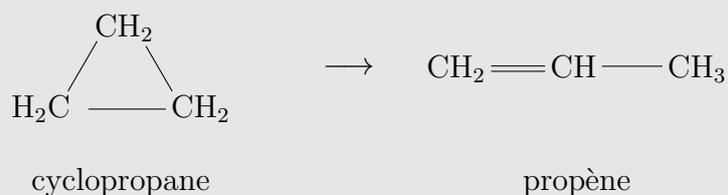
- **outils d'analyse différentielle** ;
- **outils d'analyse intégrale** ;
- **outils d'analyse statistique** ;
- **outils d'analyse matricielle**.

Ces outils mathématiques permettent aussi de proposer des **modèles** simples reproduisant au mieux les données, par exemple les **modèles de régression**. Dans le cas de la physicochimie, le graal est alors d'inférer des **modèles physicochimiques** qui permettent d'expliquer les données avec des lois générales. Ces lois prennent le plus souvent la forme d'**équations différentielles** et parfois d'**équations matricielles**.

Cette U.E. va aborder ces différentes notions. Pour illustrer l'importance de ces notions en physicochimie, nous allons dans cette introduction aborder deux exemples.

1 Premier exemple : cinétique chimique

Considérons la réaction chimique de transformation du cyclopropane en propène :



À température fixée, on mesure la concentration $c(t)$ (en mmol/L) en cyclopropane en fonction du temps t (en s). On répète l'expérience 10 fois ce qui permet de calculer des barres d'erreur sur la concentration. On obtient la **série de données** qui est visualisée

dans la Figure 1 (nous aborderons la **représentation et la visualisation des série de données** dans le Chapitre I.1).

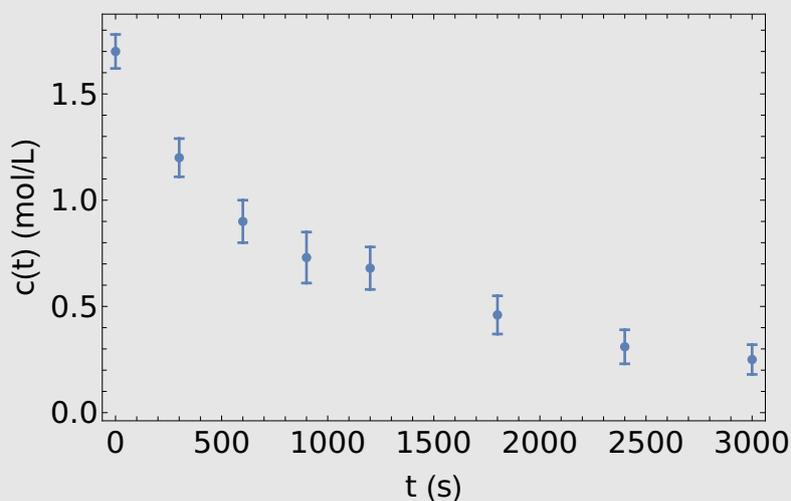
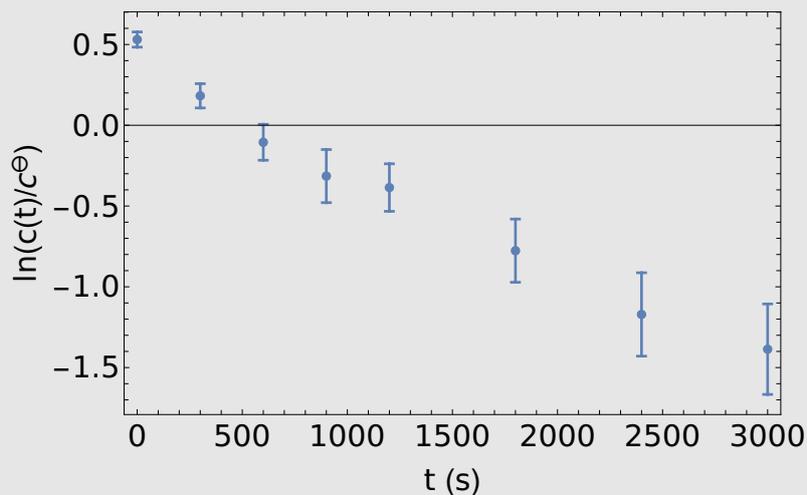


FIGURE 1 – Concentration $c(t)$ en cyclopropane en fonction du temps t .

Comment sont calculées ces **barres d'erreur** sur la concentration ? Et quelle est leur signification précise ? Si on répète l'expérience une 11ème fois, est-ce que les concentrations mesurées devront être obligatoirement dans ces barres d'erreur ? Pour répondre à ces questions, nous aurons besoin d'**outils d'analyse statistique** (Chapitre II.3).

La courbe de $c(t)$ en fonction de t n'est clairement pas une droite. Il pourrait s'agir d'une exponentielle décroissante. Pour vérifier cette hypothèse, on trace $\ln(c(t)/c^\ominus)$ en fonction de t , où $c^\ominus = 1 \text{ mmol/L}$ est une concentration de référence (diviser par c^\ominus est nécessaire pour éviter de prendre le logarithme d'une quantité ayant des unités). On obtient la courbe de la Figure 2.

FIGURE 2 – $\ln(c(t)/c^{\ominus})$ en fonction du temps t .

Comment calcule-t-on les barres d'erreur sur $\ln(c(t)/c^{\ominus})$ à partir de celles sur $c(t)$? Pour cela, nous pourrions utiliser des **outils d'analyse différentielle** (Chapitre II.1).

La courbe de $\ln(c(t)/c^{\ominus})$ en fonction de t semble pouvoir être modélisée par une droite :

$$\ln\left(\frac{c(t)}{c^{\ominus}}\right) = at + b,$$

avec a et b deux paramètres. Pour déterminer a et b , on effectue la méthode de **régression linéaire** qui donne :

$$a = -0.0006557 \text{ s}^{-1} \text{ et } b = 0.4035.$$

La droite ainsi obtenue est représentée sur la Figure 3.

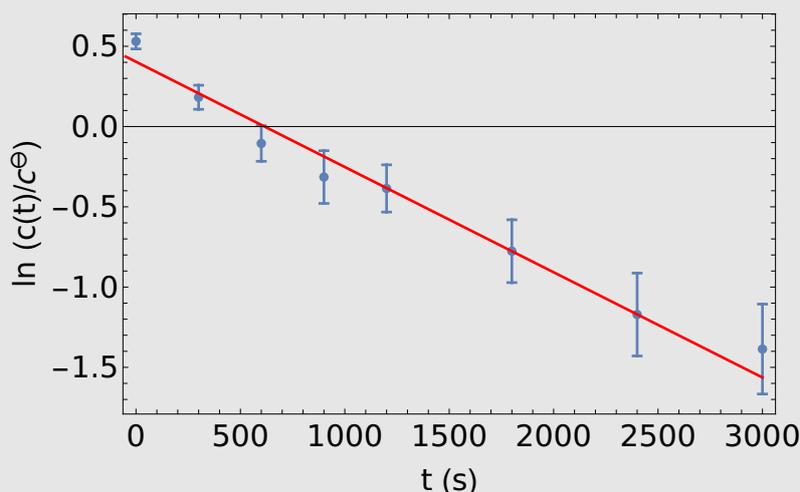


FIGURE 3 – Droite de régression linéaire $\ln(c(t)/c^{\ominus}) = -0.0006557 t + 0.4035$.

Quel est le critère utilisé pour déterminer a et b ? Quelles sont les équations qui permettent de calculer a et b ? Comment juger de la qualité de la modélisation des données par la droite de régression linéaire? Si on n'est pas satisfait par la droite, peut-on modéliser les données par une fonction plus compliquée? Nous verrons tout cela dans le Chapitre III.1 sur les **modèles de régression**, dont la mise en œuvre requière des **outils d'analyse différentielle, statistique et matricielle** (Chapitres II.1, II.3, II.4).

Si on trouve le modèle de la droite satisfaisant, cela veut dire que la concentration $c(t)$ décroît exponentiellement en fonction de t :

$$\begin{aligned} \ln\left(\frac{c(t)}{c^{\ominus}}\right) = at + b &\Leftrightarrow \frac{c(t)}{c^{\ominus}} = e^{at+b} \\ &\Leftrightarrow c(t) = c^{\ominus} e^b e^{at} \\ &\Leftrightarrow c(t) = c_0 e^{-kt}, \end{aligned}$$

où l'on a introduit les constantes positives $c_0 = c^{\ominus} e^b$ et $k = -a$. La constante c_0 est la concentration en cyclopropane à $t = 0$, c'est-à-dire $c(0) = c_0$, et la constante k est parfois appelée « constante de décroissance ».

Existe-t-il un **modèle physicochimique** conduisant à l'expression $c(t) = c_0 e^{-kt}$? Pour le deviner, on peut dériver $c(t)$ par rapport à t :

$$\frac{dc}{dt}(t) = c_0(-k)e^{-kt}.$$

On voit donc que la concentration satisfait l'**équation différentielle d'ordre 1** suivante :

$$\frac{dc}{dt}(t) = -kc(t). \quad (1)$$

Il s'agit d'une loi physicochimique assez générale : pour une réaction chimique de type $A \rightarrow B + C + \dots$, ayant un seul réactif A avec un coefficient stœchiométrique égal à 1, et se déroulant sans étapes intermédiaires à température fixée, la concentration $c(t)$ du réactif A suit l'équation 1. La constante k s'appelle la constante de vitesse de la réaction. On parle de « cinétique d'ordre 1 ». Il s'agit d'un exemple d'un **modèle physicochimique** (Chapitre III.2) s'exprimant à l'aide d'une équation différentielle. Nous verrons d'autres exemples de tels modèles.

2 Deuxième exemple : spectroscopie UV-Visible

On considère à présent l'étude de la molécule d'éthylène (C_2H_4) par spectroscopie UV-Visible. Le principe consiste à envoyer de la lumière dans la gamme UV-Visible sur des molécules d'éthylène et à mesurer l'absorbance A (une quantité sans unité, reliée à la quantité de lumière absorbée) en fonction de la longueur d'onde λ de la lumière incidente. Les résultats de ces mesures sont représentées dans la Figure 4.

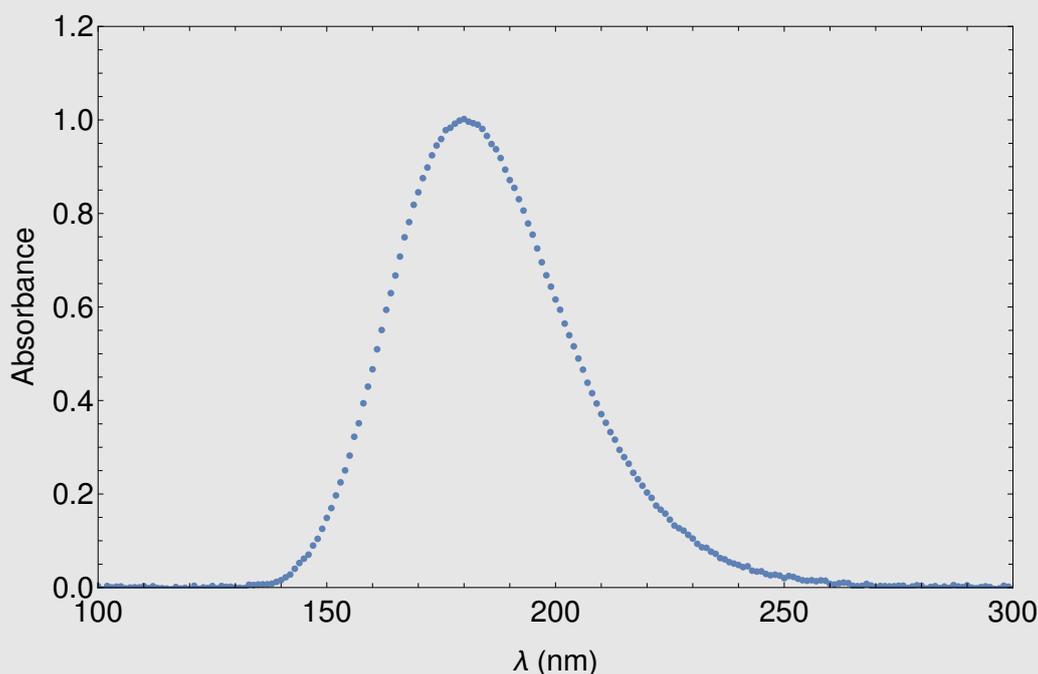


FIGURE 4 – Absorbance en fonction de la longueur d'onde λ .

Une première chose que l'on veut souvent faire est de déterminer la longueur d'onde λ_{\max} correspondant au maximum de l'absorbance. Si on disposait réellement d'une fonction $A : \lambda \mapsto A(\lambda)$ où λ pourrait prendre une continuité de valeurs (et si A était une fonction dérivable), alors la position du maximum de cette fonction pourrait être déterminé en résolvant l'équation $A'(\lambda) = 0$ où A' est la dérivée de A . Le problème est

que nous disposons pas d'une telle fonction, mais uniquement d'une **série de données**. Si on veut calculer λ_{\max} de manière automatique et fiable, il faut alors utiliser des **méthodes numériques**, et en particulier des **dérivées numériques** qui font parties des **outils d'analyse différentielle** (Chapitre II.1).

Une deuxième chose que l'on veut souvent déterminer est l'aire sous le pic d'absorbance. Cette quantité peut être utile pour comparer la contribution de différents pics et remonter à des rapports de concentration. Encore une fois, si on disposait d'une fonction $A : \lambda \mapsto A(\lambda)$ alors l'aire sous le pic d'absorbance serait donné par l'intégrale $\int_{140}^{250} A(\lambda) d\lambda$, en supposant que le pic commence à 140 nm et finit à 250 nm. Mais nous ne disposons que d'une **série de données**. Il faut alors utiliser des **intégrales numériques** qui font parties des **outils d'analyse intégrale** (Chapitre II.2).

Le **modèle physicochimique** le plus simple permettant d'expliquer l'absorption de lumière de l'éthylène autour de $\lambda_{\max} \approx 200$ nm est celui des orbitales moléculaires. Le diagramme de construction des orbitales moléculaires π et π^* est représenté dans la Figure 5. Initialement, deux électrons occupent l'orbitale liante π (orbitale HO ou HOMO pour *highest occupied molecular orbital*). Un photon UV est alors absorbé et engendre la transition d'un électron vers l'orbitale antiliante π^* (orbitale BV ou LUMO pour *lowest unoccupied molecular orbital*).

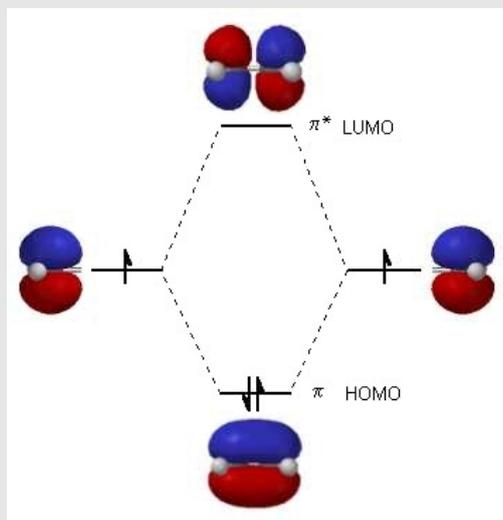


FIGURE 5 – Diagramme de construction des orbitales moléculaires π et π^* de l'éthylène à partir des orbitales atomiques p_z des atomes de carbone.

Dans la Figure 5, les formes des orbitales moléculaires π et π^* de l'éthylène, ainsi que celles des orbitales atomiques p_z des atomes de carbone de départ, ont été représentées. Mais chaque orbitale correspond à une fonction ψ des trois coordonnées d'espace $\psi : (x, y, z) \mapsto \psi(x, y, z)$. Comment **visualiser** une fonction de trois variables? Que représente exactement les images des orbitales sur la Figure 5? Nous aborderons la **visualisation des fonctions** dans le Chapitre I.2.

Ces orbitales moléculaires π et π^* peuvent être calculées à partir du modèle de Hückel. Dans ce modèle, chaque orbitale moléculaire de type π est écrite comme **combinaison linéaire** des orbitales p_z des deux atomes de carbone :

$$\psi_\pi = c_1\psi_{p_{z,1}} + c_2\psi_{p_{z,2}},$$

et les coefficients sont déterminées à partir de l'**équation matricielle aux valeurs propres**

$$\begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = E \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}.$$

Dans cette équation, α et β sont des nombres fixés représentant, respectivement, l'énergie d'une orbitale p_z isolée et l'énergie d'interaction entre les deux orbitales $p_{z,1}$ et $p_{z,2}$. Par ailleurs, E est l'énergie de l'orbitale π considérée. Il s'agit d'un exemple d'un **modèle physicochimique** (Chapitre III.2) s'exprimant à l'aide d'une équation matricielle.

Le modèle de Hückel est lui-même une approximation à une modèle plus général : l'**équation de Schrödinger**. Il s'agit d'une équation permettant en particulier de déterminer les orbitales ψ dans un système atomique ou moléculaire et qui prend la forme suivante :

$$-\frac{\hbar^2}{2m_e}\Delta\psi(x, y, z) + V(x, y, z)\psi(x, y, z) = E\psi(x, y, z),$$

où m_e est la masse de l'électron, $\hbar = h/(2\pi)$ est la constante de Planck réduite, $V(x, y, z)$ est l'énergie potentielle de l'électron, et E est l'énergie de l'orbitale ψ . Dans cette équation Δ est l'opérateur laplacien défini par la somme des trois dérivées secondes partielles par rapport à x , y et z :

$$\Delta\psi(x, y, z) = \frac{\partial^2\psi}{\partial x^2}(x, y, z) + \frac{\partial^2\psi}{\partial y^2}(x, y, z) + \frac{\partial^2\psi}{\partial z^2}(x, y, z).$$

Il s'agit d'un **opérateur différentiel**, qui fait partie des **outils d'analyse différentielle** (Chapitre II.1). L'équation de Schrödinger est un **modèle physicochimique** (Chapitre III.2) s'exprimant à l'aide d'une **équation différentielle aux dérivées partielles**.

À travers ces deux exemples (cinétique chimique et spectroscopie UV-Visible), nous avons donc vu que les outils mathématiques se trouvent au cœur de l'étude de la physicochimie.

Première partie

Fonctions et données

Chapitre I.1

Fonctions

En sciences, on utilise très souvent le concept de fonction, et en particulier le concept de fonctions de plusieurs variables. Par exemple, on peut considérer la température T dans une pièce comme fonction des trois variables d'espace x, y, z . Dans ce chapitre, nous traitons des fonctions d'une ou de plusieurs variables, et en particulier de leur visualisation.

1 Fonctions d'une ou de plusieurs variables

1.1 Fonction d'une variable réelle

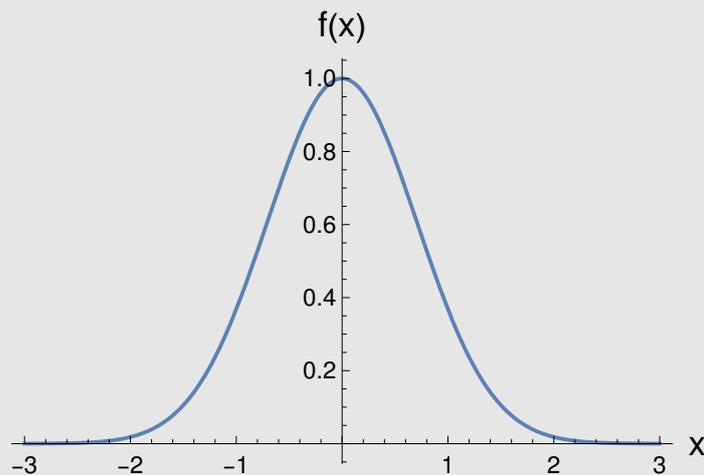
Définition (Fonction d'une variable réelle). Une fonction f d'un ensemble $D \subseteq \mathbb{R}$ dans \mathbb{R} associe à tout nombre réel $x \in D$ un nombre réel noté $f(x)$. On note :

$$\begin{aligned} f : D &\rightarrow \mathbb{R} \\ x &\mapsto f(x). \end{aligned}$$

L'ensemble de départ D est appelé domaine de définition de f .

Par abus de langage, on dit souvent que « $f(x)$ est une fonction ». Mais, en toute rigueur, $f(x)$ n'est pas une fonction mais la valeur de la fonction f au point x . Si on veut parler d'une fonction (un objet défini pour plusieurs valeurs de x), il est plus correct d'utiliser les notations f ou $x \mapsto f(x)$. Par exemple, il est plus correct de dire que « la fonction racine carrée est $x \mapsto \sqrt{x}$ » que de dire « la fonction racine carrée est \sqrt{x} ».

Bien sûr, on visualise une fonction d'une variable réelle en traçant la courbe $y = f(x)$ dans le plan (x, y) . Par exemple, le graphe de la fonction $f : x \mapsto e^{-x^2}$ est représenté sur la Figure I.1.1.

FIGURE I.1.1 – Graphe de la fonction $f : x \mapsto e^{-x^2}$.

1.2 Fonction de plusieurs variables réelles

Définition (Fonction de plusieurs variables réelles). Une fonction f d'un domaine $D \subseteq \mathbb{R}^n$ (où $n \geq 1$ est un entier) dans \mathbb{R} associe à n nombres réels $(x_1, x_2, \dots, x_n) \in D$ un nombre réel noté $f(x_1, x_2, \dots, x_n)$. On note :

$$f : \quad D \quad \rightarrow \quad \mathbb{R}$$

$$(x_1, x_2, \dots, x_n) \mapsto f(x_1, x_2, \dots, x_n).$$

On s'intéressera au cas de deux variables ($n = 2$) et au cas de trois variables ($n = 3$).

Exemple. (Enthalpie libre en thermodynamique). Pour une substance (gaz, liquide ou solide) à une pression P et à la température T , une quantité centrale est l'enthalpie libre G qui est une fonction des deux variables P et T (on suppose que le nombre de moles ne varie pas, sinon il faut l'ajouter comme troisième variable) :

$$G : \mathbb{R}_+^2 \rightarrow \mathbb{R}$$

$$(P, T) \mapsto G(P, T).$$

Ici, \mathbb{R}_+ désigne l'ensemble des réels positifs. En effet, la pression P et la température T sont toujours positives. Par exemple, pour un gaz parfait, on a (pour $P > 0$) :

$$G(P, T) = G^\ominus(T) + nRT \ln \left(\frac{P}{P^\ominus} \right),$$

où R est la constante des gaz parfaits, n est le nombre de moles (supposé fixé ici), $P^\ominus = 1$ bar est la pression standard de référence, et $G^\ominus(T)$ est l'enthalpie libre standard de référence.

Exemple. (*Orbitales en chimie quantique*). Une orbitale ψ à valeurs réelles dans un système atomique ou moléculaire est une fonction des trois coordonnées cartésiennes de l'espace :

$$\begin{aligned}\psi : \quad \mathbb{R}^3 &\rightarrow \mathbb{R} \\ (x, y, z) &\mapsto \psi(x, y, z).\end{aligned}$$

Par exemple, les orbitales $1s$ et $2p_z$ de l'atome hydrogène ont pour expressions :

$$\psi_{1s}(x, y, z) = \frac{1}{\sqrt{\pi}a_0^{3/2}} e^{-\sqrt{x^2+y^2+z^2}/a_0},$$

$$\psi_{2p_z}(x, y, z) = \frac{1}{4\sqrt{2\pi}a_0^{5/2}} z e^{-\sqrt{x^2+y^2+z^2}/(2a_0)},$$

où $a_0 \approx 0,529 \text{ \AA}$ est le rayon de Bohr.

Visualisation d'une fonction de deux variables réelles

Pour visualiser une fonction de deux variables réelles $f : (x, y) \mapsto f(x, y)$, il existe deux possibilités :

- on peut tracer la surface $z = f(x, y)$ dans l'espace à 3 dimensions (x, y, z) ;
- on peut tracer les lignes de niveau $f(x, y) = c$ dans le plan à 2 dimensions (x, y) , où c est la valeur déterminant la ligne de niveau.

Ces deux possibilités sont illustrées pour la fonction $f : (x, y) \mapsto e^{-(x^2+y^2)}$ dans la Figure I.1.2.

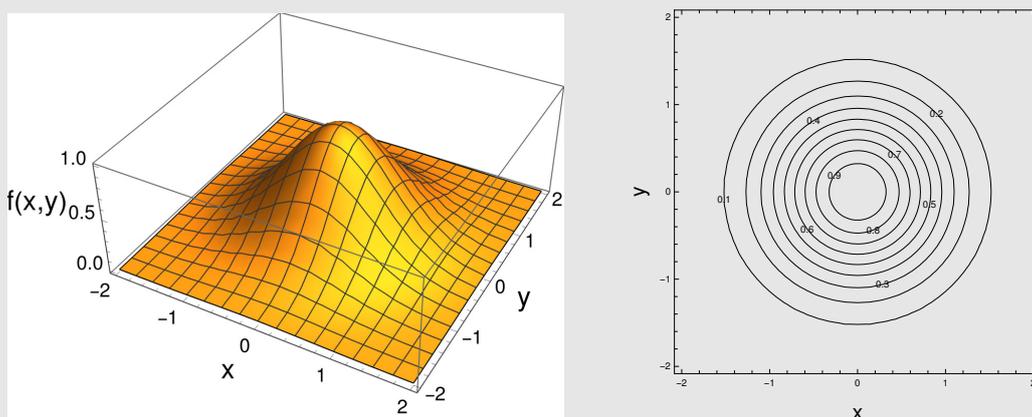


FIGURE I.1.2 – Deux visualisations possibles de la fonction $f : (x, y) \mapsto e^{-(x^2+y^2)}$. À gauche : surface $z = f(x, y)$ dans l'espace à 3 dimensions (x, y, z) . À droite : lignes de niveau $f(x, y) = c$ dans l'espace à 2 dimensions (x, y) (les valeurs de c sont indiquées sur le graphe)

En particulier, la visualisation par lignes de niveau est utilisée pour indiquer le relief sur une carte géographique. Dans ce cas, la fonction représentée est l'altitude h en fonction des coordonnées x et y sur le sol. Une ligne de niveau sur une carte correspond donc à une courbe d'altitude constante.

Visualisation d'une fonction de trois variables réelles

Pour visualiser une fonction de trois variables réelles $f : (x, y, z) \mapsto f(x, y, z)$, on peut tracer des isosurfaces $f(x, y, z) = c$ dans l'espace à 3 dimensions (x, y, z) , où c est la valeur déterminant l'isosurface.

Ceci est illustré pour les orbitales $1s$ et $2p_z$ de l'atome d'hydrogène dans les Figures I.1.3 et I.1.4. Dans le cas de l'orbitale $2p_z$, on a utilisé la couleur bleue pour tracer une isosurface correspondant à une valeur positive de la fonction ψ_{2p_z} et la couleur rouge pour tracer une isosurface correspondant à une valeur négative de la fonction ψ_{2p_z} .

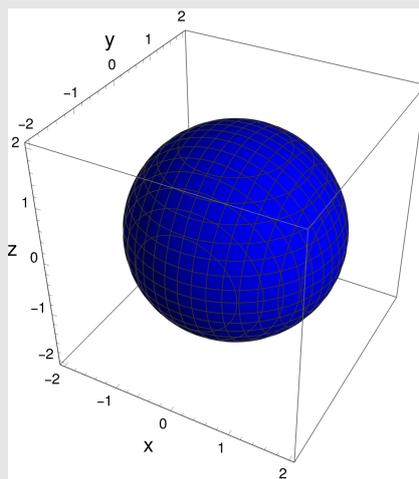


FIGURE I.1.3 – Visualisation de l'orbitale $1s$ de l'atome d'hydrogène ψ_{1s} avec des unités telles que $a_0 = 1$. La surface tracée correspond à l'isosurface $\psi_{1s}(x, y, z) = 0.1$.

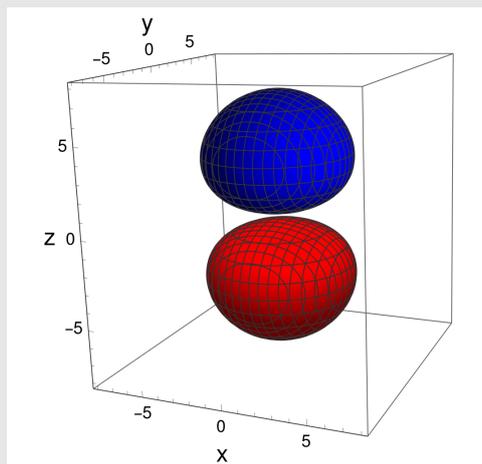


FIGURE I.1.4 – Visualisation de l'orbitale $2p_z$ de l'atome d'hydrogène ψ_{2p_z} avec des unités telles que $a_0 = 1$. La surface bleue correspond à l'isosurface $\psi_{2p_z}(x, y, z) = 0.02$ et la surface rouge correspond à l'isosurface $\psi_{2p_z}(x, y, z) = -0.02$.

Il faut garder en tête qu'une isosurface $f(x, y, z) = c$ d'une fonction f de trois variables ne permet de visualiser qu'une partie de la fonction f . Pour visualiser de façon plus complète la fonction f , il faudrait tracer beaucoup d'isosurfaces, c'est-à-dire choisir différentes valeurs de c .

2 Systèmes de coordonnées

Dans le cas de l'exemple précédent d'une orbitale $\psi : (x, y, z) \mapsto \psi(x, y, z)$, nous avons utilisé pour les trois variables d'espace les trois coordonnées cartésiennes x , y et z . Mais il est parfois plus simple d'utiliser des coordonnées alternatives aux coordonnées cartésiennes. Nous allons en voir deux : les coordonnées polaires en deux dimensions et les coordonnées sphériques en trois dimensions.

2.1 Coordonnées en dimension 2 : coordonnées polaires

On considère le plan \mathbb{R}^2 avec un repère cartésien (O, \vec{i}, \vec{j}) . Un point M de coordonnées cartésiennes (x, y) peut aussi être repéré par sa distance r à l'origine O et l'angle θ entre le vecteur unitaire \vec{i} et le vecteur \overrightarrow{OM} :

$$r = \|\overrightarrow{OM}\| \quad \text{et} \quad \theta = \widehat{(\vec{i}, \overrightarrow{OM})}.$$

Ceci est illustré sur la Figure I.1.5. Pour faire parcourir au point M tout le plan, r doit varier entre 0 et $+\infty$, et θ entre 0 et 2π .

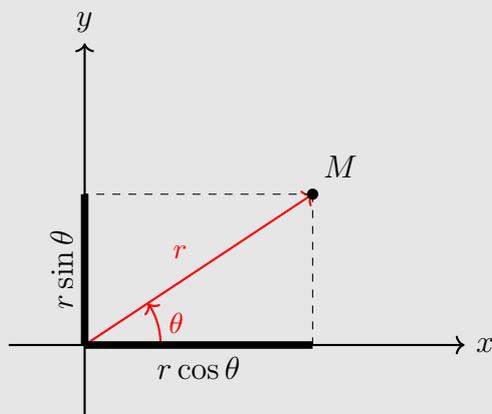


FIGURE I.1.5 – Système de coordonnées polaires.

Comme indiqué sur la Figure, les expressions de x et y en fonction de r et θ sont :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta. \end{cases}$$

Par ailleurs, d'après le théorème de Pythagore, on a : $r = \sqrt{x^2 + y^2}$.

Si on a une fonction $f : (x, y) \mapsto f(x, y)$ en coordonnées cartésiennes, on peut écrire

la fonction correspondante $\tilde{f} : (r, \theta) \mapsto \tilde{f}(r, \theta)$ en coordonnées polaires par :

$$\tilde{f}(r, \theta) = f(r \cos \theta, r \sin \theta).$$

2.2 Coordonnées en dimension 3 : coordonnées sphériques

On considère maintenant l'espace à trois dimensions \mathbb{R}^3 avec un repère cartésien $(O, \vec{i}, \vec{j}, \vec{k})$. Un point M de coordonnées cartésiennes (x, y, z) peut aussi être repéré par sa distance r à l'origine O , l'angle θ entre le vecteur unitaire \vec{k} et le vecteur \overrightarrow{OM} , l'angle ϕ entre le vecteur unitaire \vec{i} et le vecteur \overrightarrow{OP} où P est le projeté orthogonal de M sur le plan (xOy) :

$$r = \|\overrightarrow{OM}\| \text{ et } \theta = \widehat{(\vec{k}, \overrightarrow{OM})} \text{ et } \phi = \widehat{(\vec{i}, \overrightarrow{OP})}.$$

Ceci est illustré sur la Figure I.1.6. Pour faire parcourir au point M tout l'espace, r doit varier entre 0 et $+\infty$, θ entre 0 et π , et ϕ entre 0 et 2π .

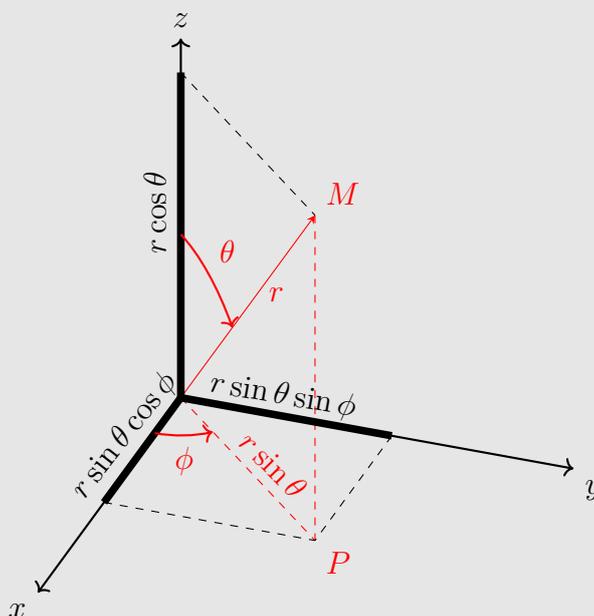


FIGURE I.1.6 – Système de coordonnées sphériques.

Comme indiqué sur la Figure, les expressions de x , y et z en fonction de r , θ et ϕ

sont :

$$\begin{cases} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta. \end{cases}$$

Par ailleurs, on a encore : $r = \sqrt{x^2 + y^2 + z^2}$.

Si on a une fonction $f : (x, y, z) \mapsto f(x, y, z)$ en coordonnées cartésiennes, on peut écrire la fonction correspondante $\tilde{f} : (r, \theta, \phi) \mapsto \tilde{f}(r, \theta, \phi)$ en coordonnées sphériques par :

$$\tilde{f}(r, \theta, \phi) = f(r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta).$$

Exemple. (*Orbitales 1s et 2p_z*). En coordonnées sphériques, les orbitales 1s et 2p_z de l'atome hydrogène ont pour expressions :

$$\tilde{\psi}_{1s}(r, \theta, \phi) = \frac{1}{\sqrt{\pi}a_0^{3/2}} e^{-r/a_0},$$

$$\tilde{\psi}_{2p_z}(r, \theta, \phi) = \frac{1}{4\sqrt{2\pi}a_0^{5/2}} r \cos \theta e^{-r/(2a_0)}.$$

Par rapport aux coordonnées cartésiennes, ces expressions sont plus simples. On voit que l'orbitale 1s ne dépend que de r et pas de θ et ϕ . L'orbitale 2p_z ne dépend que de r et θ et pas de ϕ .

3 Champs de vecteurs

On s'intéressera uniquement au cas à deux dimensions et à trois dimensions.

Définition (Champ de vecteurs à deux dimensions). Soit le plan \mathbb{R}^2 muni du repère cartésien (O, \vec{i}, \vec{j}) . Un champ de vecteurs (ou fonction vectorielle) à deux dimensions est une fonction \vec{f} de \mathbb{R}^2 dans \mathbb{R}^2 qui à un point de coordonnées $(x, y) \in \mathbb{R}^2$ associe un vecteur à deux composantes noté $\vec{f}(x, y) \in \mathbb{R}^2$. On note :

$$\begin{aligned} \vec{f} : \quad \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto \vec{f}(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} \\ &= f_1(x, y)\vec{i} + f_2(x, y)\vec{j} \end{aligned}$$

Définition (Champ de vecteurs à trois dimensions). Soit l'espace \mathbb{R}^3 muni du repère cartésien $(O, \vec{i}, \vec{j}, \vec{k})$. Un champ de vecteurs (ou fonction vectorielle) à trois dimensions est une fonction \vec{f} de \mathbb{R}^3 dans \mathbb{R}^3 qui à un point de coordonnées $(x, y, z) \in \mathbb{R}^3$ associe un vecteur à trois composantes noté $\vec{f}(x, y, z) \in \mathbb{R}^3$. On note :

$$\begin{aligned} \vec{f}: \quad \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ (x, y, z) &\mapsto \vec{f}(x, y, z) = \begin{pmatrix} f_1(x, y, z) \\ f_2(x, y, z) \\ f_3(x, y, z) \end{pmatrix} \\ &= f_1(x, y, z)\vec{i} + f_2(x, y, z)\vec{j} + f_3(x, y, z)\vec{k} \end{aligned}$$

Pour visualiser un champ de vecteurs à deux dimensions, on trace à plusieurs points (x, y) du plan le vecteur à deux composantes correspondant $\vec{f}(x, y)$. De même, pour visualiser un champ de vecteurs à trois dimensions, on trace à plusieurs points (x, y, z) de l'espace le vecteur à trois composantes correspondant $\vec{f}(x, y, z)$.

Exemple. (*Champ électrique d'une charge ponctuelle*). Le champ électrique

$$\vec{E}: (x, y, z) \mapsto \vec{E}(x, y, z)$$

est un champ de vecteurs à trois dimensions. Par exemple, une charge électrique élémentaire positive (comme un proton), supposée ponctuelle et située au centre du repère, génère un champ électrique de la forme :

$$\vec{E}(x, y, z) = \frac{e}{4\pi\epsilon_0} \begin{pmatrix} x/r^3 \\ y/r^3 \\ z/r^3 \end{pmatrix} = \frac{e}{4\pi\epsilon_0 r^3} (x\vec{i} + y\vec{j} + z\vec{k}),$$

où e est la charge élémentaire, ϵ_0 est la permittivité du vide, et $r = \sqrt{x^2 + y^2 + z^2}$. Ce champ électrique est représenté dans la Figure I.1.7.

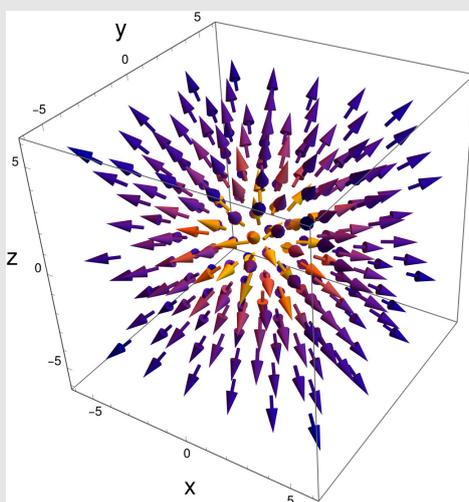


FIGURE I.1.7 – Visualisation du champ électrique $\vec{E}(x, y, z)$ généré par une charge électrique élémentaire positive ponctuelle (en unités atomiques).

Chapitre I.2

Données

En sciences, on obtient souvent des données numériques provenant de mesures expérimentales ou de calculs théoriques. Pour être étudiées et visualisées, ces données doivent être représentées sous forme d'un objet mathématique, par exemple une fonction. Une des premières opérations que l'on souhaite souvent faire sur des données est l'interpolation de ces données, ce qui permet d'estimer la valeur d'une variable entre deux points des données.

1 Représentation et visualisation d'une série de données

Considérons une série de données de la forme de n couples de points (x, y) :

x	x_1	x_2	x_3	\cdots	\cdots	x_{n-1}	x_n
y	y_1	y_2	y_3	\cdots	\cdots	y_{n-1}	y_n

Par exemple, pour l'exemple donné dans l'introduction sur la cinétique chimique, x correspond au temps t et y à la concentration c en cyclopropane :

t (s)	0	300	600	900	1200	1800	2400	3000
c (mmol/L)	1,70	1,20	0,90	0,73	0,68	0,46	0,31	0,25

Le plus souvent, on représentera une telle série de données par une fonction d'une variable réelle

$$f : D \rightarrow \mathbb{R}$$
$$x \mapsto f(x),$$

avec un domaine de définition discret $D = \{x_1, x_2, \dots, x_n\}$ et associant à chaque valeur x_i la valeur y_i :

$$f(x_1) = y_1, f(x_2) = y_2, \dots, f(x_n) = y_n.$$

Parfois, il est aussi pratique de représenter une série de données par des vecteurs ou des matrices. Dans le cas des vecteurs, on range les données dans deux vecteurs \vec{x} et \vec{y} à n composantes :

$$\vec{x} = (x_1, x_2, \dots, x_n) \text{ et } \vec{y} = (y_1, y_2, \dots, y_n).$$

Dans le cas des matrices, on range les données soit dans une matrice M à n lignes et à 2 colonnes

$$M = \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \\ \vdots & \vdots \\ x_n & y_n \end{pmatrix},$$

soit dans une matrice P à 2 lignes et à n colonnes

$$P = \begin{pmatrix} x_1 & x_2 & \dots & x_n \\ y_1 & y_2 & \dots & y_n \end{pmatrix}.$$

Remarquons que la matrice P est la transposée de la matrice M , on note $P = M^T$.

Bien sûr, pour visualiser une série de données, on dispose les points (x_i, y_i) sur le plan x, y . Ceci est illustré pour l'exemple de la cinétique chimique dans la Figure I.2.1.

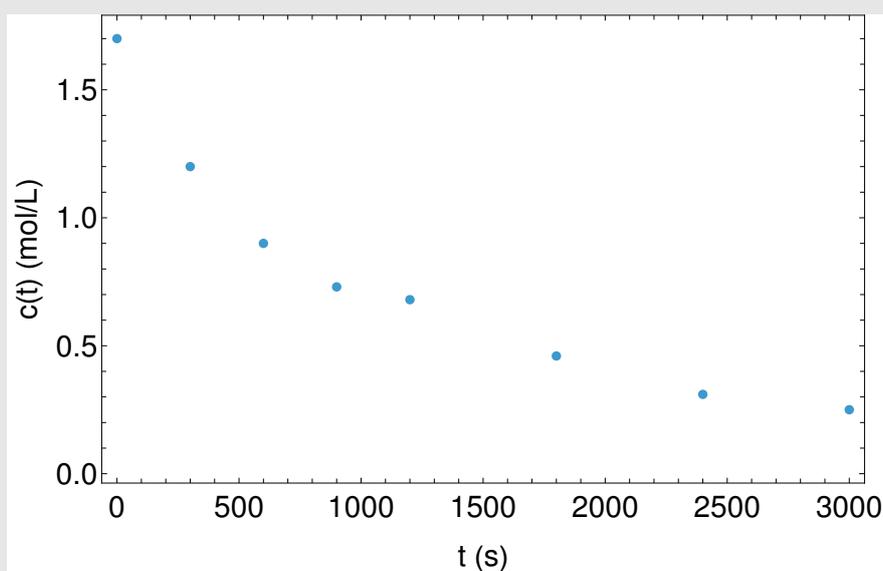


FIGURE I.2.1 – Concentration $c(t)$ en cyclopropane en fonction du temps t .

2 Interpolation d'une série de données

Supposons que l'on dispose d'une série de données $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. On a parfois besoin de considérer une valeur de x ne faisant pas partie des données mais se situant entre deux valeurs consécutives x_i et x_{i+1} des données, et on voudrait alors estimer la valeur de y correspondante. Ou, inversement, on a besoin de considérer une valeur de y ne faisant pas partie des données mais se situant entre deux valeurs consécutives y_i et y_{i+1} des données, et on voudrait alors estimer la valeur de x correspondante. Pour cela, on peut utiliser une interpolation.

Il existe différents types d'interpolation. Nous allons en voir deux assez simples : l'interpolation linéaire et l'interpolation polynomiale.

2.1 Interpolation linéaire

L'interpolation linéaire consiste à faire passer une droite entre chaque paire de points consécutifs (x_i, y_i) et (x_{i+1}, y_{i+1}) . La pente de la droite reliant les points (x_i, y_i) et (x_{i+1}, y_{i+1}) est $(y_{i+1} - y_i)/(x_{i+1} - x_i)$. L'équation de la droite entre les points (x_i, y_i) et (x_{i+1}, y_{i+1}) est donc de la forme :

$$y(x) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}x + b \text{ pour } x \in [x_i, x_{i+1}],$$

où b est une constante à déterminer. Puisque la droite doit passer par le point (x_i, y_i) , on doit avoir $y(x_i) = y_i$ et donc :

$$y_i = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}x_i + b,$$

ce qui donne

$$b = y_i - \frac{y_{i+1} - y_i}{x_{i+1} - x_i}x_i.$$

L'équation de la droite entre les points (x_i, y_i) et (x_{i+1}, y_{i+1}) est donc :

$$y(x) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}(x - x_i) + y_i \text{ pour } x \in [x_i, x_{i+1}].$$

On peut vérifier que cette droite passe bien par le point (x_{i+1}, y_{i+1}) , c'est-à-dire $y(x_{i+1}) = y_{i+1}$. L'équation de cette droite $y(x)$ donne un moyen d'estimer y pour n'importe quel $x \in]x_i, x_{i+1}[$.

On peut aussi exprimer x en fonction de y :

$$x(y) = \frac{x_{i+1} - x_i}{y_{i+1} - y_i}(y - y_i) + x_i \text{ pour } y \in [y_i, y_{i+1}].$$

Cette relation $x(y)$ donne un moyen d'estimer x pour n'importe quel $y \in]y_i, y_{i+1}[$.

L'interpolation linéaire est représentée pour l'exemple de la cinétique chimique dans la Figure I.2.2.

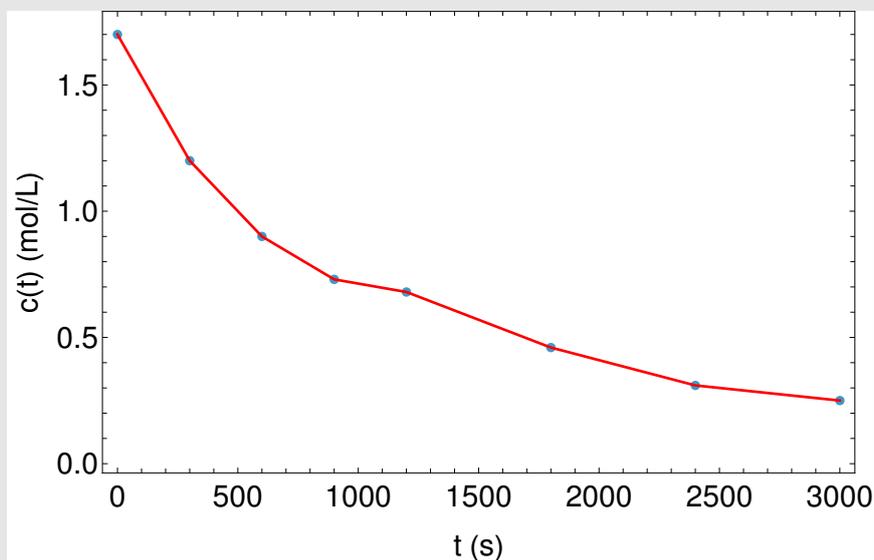


FIGURE I.2.2 – Concentration $c(t)$ en cyclopropane en fonction du temps t avec interpolation linéaire entre les points.

2.2 Interpolation polynomiale

L'interpolation polynomiale consiste à trouver un polynôme P qui passe par tous les points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, c'est-à-dire

$$P(x_1) = y_1, \quad P(x_2) = y_2, \quad \dots, \quad P(x_n) = y_n. \quad (\text{I.2.1})$$

On peut montrer qu'il existe un unique polynôme P de degré inférieur ou égal à $n - 1$ qui satisfait cela. Ce polynôme P permet d'estimer la valeur de y pour n'importe quel $x \in [x_1, x_n]$:

$$y(x) = P(x).$$

On peut exprimer ce polynôme P par

$$P(x) = y_1 \ell_1(x) + y_2 \ell_2(x) + \dots + y_n \ell_n(x),$$

où ℓ_i sont les polynômes de Lagrange définis par

$$\ell_i(x) = \frac{p_i(x)}{p_i(x_i)},$$

où

$$p_i(x) = (x - x_1)(x - x_2) \cdots (x - x_{i-1})(x - x_{i+1}) \cdots (x - x_n).$$

On peut vérifier que $l_i(x_i) = 1$ et $l_i(x_j) = 0$ pour $j \neq i$, ce qui permet de satisfaire l'équation I.2.1. Puisque chaque p_i contient un produit de $n - 1$ termes, le polynôme P est en général de degré $n - 1$, ou inférieur à $n - 1$ si le terme en x^{n-1} a un coefficient nul.

L'interpolation polynomiale est représentée pour l'exemple de la cinétique chimique dans la Figure I.2.3. Puisqu'il y a $n = 8$ points, le degré du polynôme est égal à $n - 1 = 7$. L'interpolation semble raisonnable quand t n'est pas trop grand. Aux grandes valeurs de t , l'interpolation polynomiale n'est pas raisonnable : elle conduit à une courbe oscillante qui représente certainement très mal la vraie fonction $c : t \mapsto c(t)$. Ces oscillations apparaissent souvent pour un polynôme P de degré élevé. Le problème est connu sous le nom de phénomène de Runge.

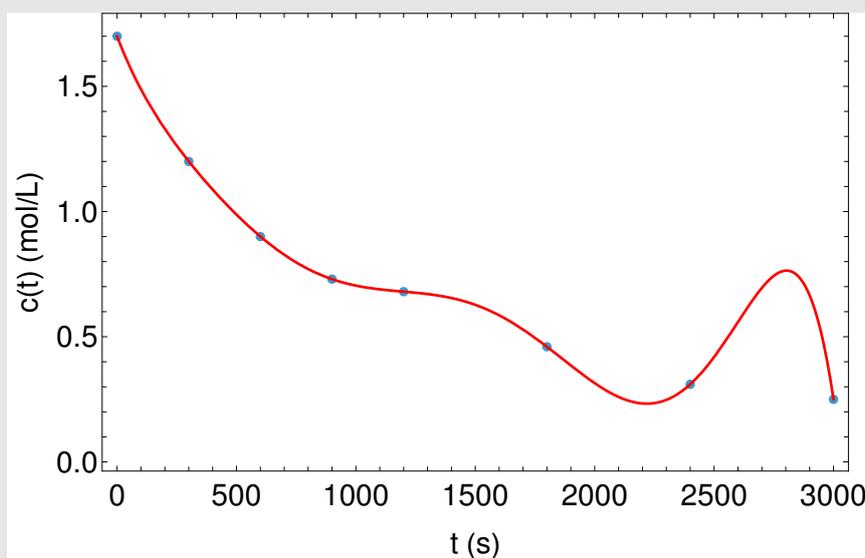


FIGURE I.2.3 – Concentration $c(t)$ en cyclopropane en fonction du temps t avec interpolation polynomiale (le degré du polynôme est égal à 7).

2.3 Introduction à l'interpolation par splines

Pour éviter le problème des oscillations fallacieuses, la solution est d'utiliser un polynôme de petit degré mais en se laissant la liberté que ce polynôme soit différent sur chaque intervalle $[x_i, x_{i+1}]$. On parle de polynôme par morceaux ou « splines ». On estime alors la valeur de y pour $x \in [x_i, x_{i+1}]$ par le polynôme s_i :

$$y(x) = s_i(x) \text{ pour } x \in [x_i, x_{i+1}].$$

On impose que le polynôme s_i passe par les points (x_i, y_i) et (x_{i+1}, y_{i+1})

$$s_i(x_i) = y_i \text{ et } s_i(x_{i+1}) = y_{i+1}.$$

De plus, on impose aussi que deux polynômes sur des intervalles consécutifs, s_{i-1} et s_i , aient la même dérivée première et dérivée seconde au point où ils se rencontrent, c'est-à-dire en x_i :

$$s'_{i-1}(x_i) = s'_i(x_i) \text{ et } s''_{i-1}(x_i) = s''_i(x_i),$$

ce qui permet d'avoir une courbe lisse aux points de raccordement des deux polynômes s_{i-1} et s_i . En pratique, on utilise le plus souvent des polynômes de degré 3. On parle alors d'interpolation par splines cubiques.

L'interpolation par splines cubiques est représentée pour l'exemple de la cinétique chimique dans la Figure I.2.4. On voit que l'on obtient une courbe interpolée lisse, qui est raisonnable sur tout le domaine d'interpolation.

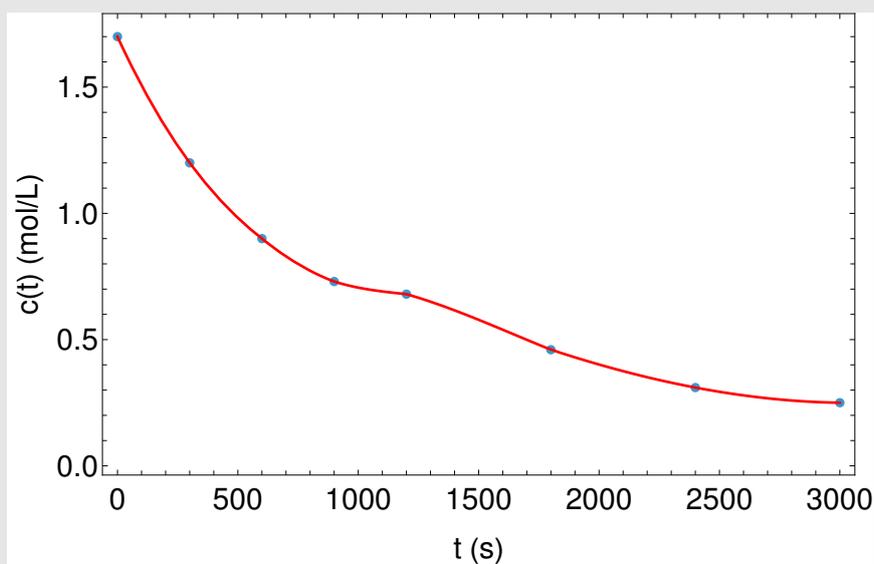


FIGURE I.2.4 – Concentration $c(t)$ en cyclopropane en fonction du temps t avec interpolation par splines cubiques.

Deuxième partie

Outils élémentaires d'analyse des fonctions et des données

Chapitre II.1

Outils d'analyse différentielle

Le calcul différentiel est un outil mathématique incontournable dans les sciences. Essentiellement, il permet d'étudier le comportement local d'une fonction. Dans ce chapitre, après avoir rappelé le concept de dérivée d'une fonction d'une variable et celui de dérivées partielles d'une fonction de plusieurs variables, nous introduirons le concept de différentielle d'une fonction qui est un outil central pour la thermodynamique et qui très utile pour le calcul de propagation d'erreur. Nous introduirons aussi les principaux opérateurs différentiels qui sont utilisés pour écrire les équations fondamentales que l'on rencontre en physicochimie. Enfin, nous verrons comment calculer numériquement une dérivée à partir de données.

1 Dérivées

1.1 Dérivées d'une fonction d'une variable réelle

Avant de définir la dérivée, rappelons le concept de continuité.

Définition (Continuité). On dit qu'une fonction $f : D \rightarrow \mathbb{R}$ est continue au point $x_0 \in D$ si

$$\lim_{x \rightarrow x_0} f(x) = f(x_0).$$

Ceci implique que les limites à gauche et à droite sont égales :

$$\lim_{x \rightarrow x_0^-} f(x) = \lim_{x \rightarrow x_0^+} f(x) = f(x_0).$$

Lorsqu'une fonction n'est pas continue en x_0 , on dit qu'elle est discontinue en x_0 . Si f est continue en tout point x_0 de D , alors on dit que f est continue sur le domaine D , ou plus simplement que f est continue. Pour faire simple, une fonction continue est telle que son graphe peut être tracé sans lever le crayon de la feuille.

Définition (Dérivabilité). On dit qu'une fonction $f : D \rightarrow \mathbb{R}$ est dérivable en $x_0 \in D$ si le taux d'accroissement de f admet une limite finie quand $x \rightarrow x_0$

$$\lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0} = f'(x_0).$$

Cette limite est notée $f'(x_0)$ et s'appelle la dérivée de f en x_0 .

Si cette limite est infinie ou si cette limite n'existe pas (par exemple, les limites à gauche et à droite sont différentes), alors f n'est pas dérivable en x_0 .

Une formule alternative souvent utilisée est :

$$\lim_{h \rightarrow 0} \frac{f(x_0 + h) - f(x_0)}{h} = f'(x_0).$$

Si f est dérivable en tout point x_0 de D , alors on dit que f est dérivable sur le domaine D , ou plus simplement que f est dérivable.

En pratique, pour calculer la dérivée d'une fonction, on utilise les règles de dérivation rappelées en annexe.

Théorème (Tangente). Soit $f : D \rightarrow \mathbb{R}$ une fonction dérivable. L'équation de la tangente à la courbe représentative de f en x_0 est : $y = f(x_0) + f'(x_0)(x - x_0)$.

Il s'agit simplement du développement limité de f en $x = x_0$ à l'ordre 1. La dérivée $f'(x_0)$ est donc le coefficient directeur (pente) de la droite tangente à la courbe de f à $x = x_0$.

La fonction $f' : x \mapsto f'(x)$ est appelée dérivée de f . On peut continuer à dériver, et on définit ainsi :

- la dérivée seconde, notée f'' , comme la dérivée de la dérivée f' , c'est-à-dire $f'' = (f')'$;
- la dérivée troisième, notée f''' , comme la dérivée de la dérivée seconde f'' , c'est-à-dire $f''' = (f'')'$;
- et, de manière générale, la dérivée d'ordre n (où n est un entier positif ou nul), notée $f^{(n)}$, comme la dérivée de la dérivée d'ordre $n - 1$, c'est-à-dire $f^{(n)} = (f^{(n-1)})'$ pour $n \geq 1$. Par convention, la dérivée d'ordre 0 est juste la fonction $f^{(0)} = f$. Ainsi, la dérivée d'ordre 1 est $f^{(1)} = f'$, la dérivée d'ordre 2 est $f^{(2)} = f''$, etc.

La notation f', f'', f''', \dots ou $f^{(1)}, f^{(2)}, f^{(3)}, \dots$ pour les dérivées est appelée notation de Lagrange. C'est la notation la plus commune en mathématiques. En sciences, on utilise aussi souvent la notation de Leibniz

$$\frac{df}{dx}, \frac{d^2f}{dx^2}, \frac{d^3f}{dx^3}, \dots$$

qui a le mérite de rappeler la définition de la dérivée avec le taux d'accroissement. Pour les fonctions du temps, $f : t \mapsto f(t)$, on utilise aussi parfois la notation de Newton :

$$\dot{f}, \ddot{f}, \overset{\cdot\cdot}{f}, \dots$$

Définition (Classe de régularité). On dit qu'une fonction $f : D \rightarrow \mathbb{R}$ est de classe C^n (où n est un entier positif ou nul) si f est dérivable sur D jusqu'à l'ordre n et la dérivée d'ordre n , $f^{(n)}$, est continue sur D . On dit que f est de classe C^∞ si f est dérivable sur D une infinité de fois.

En particulier :

- Une fonction f de classe C^0 est une fonction continue.
- Une fonction f de classe C^1 est une fonction dérivable et sa dérivée f' est continue.
- Une fonction f de classe C^2 est une fonction deux fois dérivable et sa dérivée seconde f'' est continue.
- Une fonction f de classe C^∞ est une fonction infiniment dérivable. On dit aussi que c'est une fonction lisse ou régulière.

Il est particulièrement agréable de travailler avec des fonctions de classe C^∞ puisque l'on peut les dériver sans problème autant de fois que l'on veut. Heureusement, en sciences, nous avons souvent à faire à de telles fonctions. En particulier, les fonctions usuelles $x \mapsto x^n$ (n entier positif ou négatif), $x \mapsto e^x$, $x \mapsto \ln x$, $x \mapsto \cos x$, $x \mapsto \sin x$ et leurs sommes, produits et compositions sont toutes de classe C^∞ sur leurs domaines de définition respectifs.

Exemple. (*Cinétique chimique*). Lors de la réaction chimique de transformation du cyclopropane en propène, la concentration du cyclopropane en fonction du temps est donnée par la fonction $c : t \mapsto c(t) = c_0 e^{-kt}$. Cette fonction est de classe C^∞ . En effet, les dérivées sont :

$$c'(t) = c_0(-k)e^{-kt}, \quad c''(t) = c_0(-k)^2 e^{-kt}, \quad c'''(t) = c_0(-k)^3 e^{-kt}, \quad \text{etc.}$$

La dérivée à l'ordre n existe donc pour tout entier n et vaut :

$$c^{(n)}(t) = c_0(-k)^n e^{-kt}.$$

La concentration varie donc de façon lisse avec le temps.

On a souvent besoin de trouver les extremums d'une fonction, c'est-à-dire les maximums et les minimums. Pour cela, on cherche d'abord les points stationnaires de la fonction.

Définition (Point stationnaire). Soit une fonction $f : D \rightarrow \mathbb{R}$ dérivable. On dit que le point $x_0 \in D$ est un point stationnaire de f si la dérivée est nulle en ce point : $f'(x_0) = 0$.

En un point stationnaire, la tangente à la courbe représentative de f est donc horizontale.

Théorème (Classification des points stationnaires). Soit une fonction $f : D \rightarrow \mathbb{R}$ de classe C^3 et $x_0 \in D$ un point stationnaire (on a donc $f'(x_0) = 0$). On considère trois possibilités :

- Si $f''(x_0) > 0$, alors f a un minimum en x_0 .
- Si $f''(x_0) < 0$, alors f a un maximum en x_0 .
- Si $f''(x_0) = 0$ et $f'''(x_0) \neq 0$, alors f a un point d'inflexion en x_0 .

Exemple. (Spectroscopie UV-Visible). Le pic d'absorbance UV-Visible de la molécule d'éthylène peut être modélisé par la fonction $A : \lambda \mapsto e^{-a(\lambda-180)^2}$ où $a > 0$ est une constante et λ est la longueur d'onde en nm. Pour calculer le maximum de ce pic, on cherche d'abord un point stationnaire de A :

$$A'(\lambda) = 0 \Leftrightarrow -a(\lambda - 180)e^{-a(\lambda-180)^2} = 0 \Leftrightarrow \lambda = 180.$$

Donc $\lambda_0 = 180$ nm est un point stationnaire. Pour vérifier qu'il s'agit bien d'un maximum, on calcule la dérivée seconde :

$$A''(\lambda) = -ae^{-a(\lambda-180)^2} + a^2(\lambda - 180)^2 e^{-a(\lambda-180)^2}.$$

En $\lambda = \lambda_0$, on a : $A''(\lambda_0) = -a < 0$. Il s'agit donc bien d'un maximum.

1.2 Dérivées d'une fonction de plusieurs variables réelles

Pour faire simple, nous allons donner les définitions pour le cas d'une fonction de deux variables réelles $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ définie sur tout \mathbb{R}^2 , mais tout peut se généraliser sans difficulté à une fonction définie sur un domaine $D \subset \mathbb{R}^2$ ou à une fonction dépendant de plus de variables réelles.

Définition (Continuité). On dit qu'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est continue au point $(x_0, y_0) \in \mathbb{R}^2$ si

$$\lim_{(x,y) \rightarrow (x_0,y_0)} f(x, y) = f(x_0, y_0).$$

Définition (Dérivées partielles). Soit une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. La dérivée partielle de f par rapport à x au point (x_0, y_0) est, si elle existe,

$$\frac{\partial f}{\partial x}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0 + h, y_0) - f(x_0, y_0)}{h},$$

et la dérivée partielle de f par rapport à y au point (x_0, y_0) est, si elle existe,

$$\frac{\partial f}{\partial y}(x_0, y_0) = \lim_{h \rightarrow 0} \frac{f(x_0, y_0 + h) - f(x_0, y_0)}{h}.$$

Si ces deux dérivées partielles existent (et sont donc finies) alors on dit que f est dérivable au point (x_0, y_0) .

Les dérivées partielles définissent de nouvelles fonctions de \mathbb{R}^2 dans \mathbb{R} :

$$\begin{aligned} \frac{\partial f}{\partial x} : \mathbb{R}^2 &\rightarrow \mathbb{R} & \text{et} & & \frac{\partial f}{\partial y} : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto \frac{\partial f}{\partial x}(x, y) & & & (x, y) &\mapsto \frac{\partial f}{\partial y}(x, y). \end{aligned}$$

La dérivée partielle $\partial f / \partial x$ est donc la dérivée de f par rapport à la variable x en considérant l'autre variable y comme une constante. De manière symétrique, la dérivée partielle $\partial f / \partial y$ est la dérivée de f par rapport à la variable y en considérant l'autre variable x comme une constante.

D'un point de vue géométrique, la dérivée partielle $\frac{\partial f}{\partial x}(x_0, y_0)$ est le coefficient directeur de la tangente à la surface représentative de f au point (x_0, y_0) dans la direction x . De même, la dérivée partielle $\frac{\partial f}{\partial y}(x_0, y_0)$ est le coefficient directeur de la tangente à la surface représentative de f au point (x_0, y_0) dans la direction y .

Exemple. (*Enthalpie libre d'un gaz parfait*). Nous avons vu que l'enthalpie libre G d'un gaz parfait est une fonction de la pression P et de la température T :

$$\begin{aligned} G : \mathbb{R}_+^2 &\rightarrow \mathbb{R} \\ (P, T) &\mapsto G(P, T) = G^\ominus(T) + nRT \ln \left(\frac{P}{P^\ominus} \right), \end{aligned}$$

où R est la constante des gaz parfaits, n est le nombre de moles, $P^\ominus = 1$ bar est la pression standard de référence, et $G^\ominus(T)$ est l'enthalpie libre standard de référence. La dérivée partielle de G par rapport P est :

$$\frac{\partial G}{\partial P}(P, T) = \frac{nRT}{P}.$$

D'après la loi des gaz parfait, $PV = nRT$, il s'agit du volume du système : $\partial G / \partial P = V$.

On peut continuer à dériver les dérivées partielles par rapport à x ou à y . On définit ainsi :

- la dérivée partielle seconde où l'on dérive deux fois par rapport à x :

$$\frac{\partial^2 f}{\partial x^2} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial x} \right);$$

- la dérivée partielle seconde où l'on dérive deux fois par rapport à y :

$$\frac{\partial^2 f}{\partial y^2} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial y} \right);$$

- les dérivées partielles secondes croisées où l'on dérive d'abord par rapport à y puis par rapport à x , et inversement :

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial}{\partial x} \left(\frac{\partial f}{\partial y} \right) \quad \text{et} \quad \frac{\partial^2 f}{\partial y \partial x} = \frac{\partial}{\partial y} \left(\frac{\partial f}{\partial x} \right).$$

De la même manière, on peut définir 8 dérivées partielles troisièmes, et ainsi de suite.

Définition (Classe de régularité). Une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ est de classe C^n si toutes les dérivées partielles jusqu'à l'ordre n existent et sont continues.

En général, on a donc 4 dérivées partielles secondes. Heureusement, pour des fonctions suffisamment régulières, on a le théorème suivant qui apporte une simplification :

Théorème (Théorème de Schwarz). Soit une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Si la fonction f est de classe C^2 alors les dérivées partielles secondes croisées sont égales :

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}.$$

Autrement dit, la dérivée partielle seconde croisée ne dépend pas de l'ordre dans lequel on dérive, ce qui est très pratique.

Exemple. (*Enthalpie libre d'un gaz parfait*). Reprenons la fonction $G : (P, T) \mapsto G(P, T) = G^\ominus(T) + nRT \ln \left(\frac{P}{P^\ominus} \right)$. Si la fonction $G^\ominus : T \mapsto G^\ominus(T)$ est de classe C^2 , alors la fonction G est aussi de classe C^2 sur son domaine de définition. On peut vérifier que les dérivées partielles secondes croisées sont :

$$\frac{\partial^2 G}{\partial T \partial P} = \frac{nR}{P} \quad \text{et} \quad \frac{\partial^2 G}{\partial P \partial T} = \frac{nR}{P}.$$

On trouve bien l'égalité des dérivées partielles secondes croisées.

Donnons la généralisation de la définition d'un point stationnaire.

Définition (Point stationnaire). Soit une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ dérivable. On dit que le point (x_0, y_0) est un point stationnaire de f si les deux dérivées partielles premières sont nulles en ce point :

$$\frac{\partial f}{\partial x}(x_0, y_0) = 0 \quad \text{et} \quad \frac{\partial f}{\partial y}(x_0, y_0) = 0.$$

Pour une fonction de plusieurs variables, un point stationnaire peut correspondre à un minimum, à un maximum, ou à un point selle (c'est-à-dire un maximum dans une direction et un minimum dans une autre).

2 Développements limités et différentielles

2.1 Développements limités

Théorème (Formule de Taylor). Soit une fonction $f : D \rightarrow \mathbb{R}$ dérivable n fois en $x_0 \in D$. La fonction f admet un développement de Taylor d'ordre n au voisinage de x_0 :

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + o((x - x_0)^n),$$

où $o((x - x_0)^n)$ est un reste tendant vers 0 plus vite que $(x - x_0)^n$ quand $x \rightarrow x_0$. Le développement de Taylor peut aussi s'écrire avec $h = x - x_0$:

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{f''(x_0)}{2!}h^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}h^n + o(h^n),$$

où $o(h^n)$ est un reste tendant vers 0 plus vite que h^n quand $h \rightarrow 0$.

Dire que le reste $o((x - x_0)^n)$ tend vers 0 plus vite que $(x - x_0)^n$ quand $x \rightarrow x_0$ signifie que le reste s'écrit $o((x - x_0)^n) = (x - x_0)^n \varepsilon(x)$ où ε est une fonction tendant vers 0 quand $x \rightarrow x_0$: $\lim_{x \rightarrow x_0} \varepsilon(x) = 0$.

En choisissant différentes valeurs de n , la formule de Taylor donne une hiérarchie d'approximations polynomiales à la fonction f au voisinage de x_0 , appelées développements limités (DL) :

- DL d'ordre 0 : $f(x) = f(x_0) + o(1)$,
- DL d'ordre 1 : $f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$,

— DL d'ordre 2 : $f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2}(x - x_0)^2 + o((x - x_0)^2)$,

— etc.

Le DL d'ordre 0 correspond à approcher f par la constante $f(x_0)$. Le DL d'ordre 1 correspond à approcher f par la droite $f(x_0) + f'(x_0)(x - x_0)$, c'est-à-dire la tangente de f en x_0 . Le DL d'ordre 2 correspond à approcher f par la parabole $f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2$, etc. Si on est suffisamment proche de x_0 , la précision du DL croît avec l'ordre n . Un DL approche donc une fonction f pouvant être très compliquée par un simple polynôme qui est une bonne approximation à f au voisinage d'un point. Cela peut simplifier considérablement l'étude de f au voisinage de ce point.

Exemple. (*Cinétique chimique*). Pour l'exemple de la réaction du cyclopropane en propène, nous avons vu que la concentration du cyclopropane en fonction du temps est donnée par la fonction $c : t \mapsto c(t) = c_0 e^{-kt}$ et que ses dérivées sont $c^{(n)}(t) = c_0 (-k)^n e^{-kt}$ pour tout entier n . Les développements limités d'ordre 0, 1 et 2 de c au voisinage de $t = 0$ sont donc :

— DL d'ordre 0 ($n = 0$) : $c(t) = c_0 + o(1)$,

— DL d'ordre 1 ($n = 1$) : $c(t) = c_0 - c_0 kt + o(t)$,

— DL d'ordre 2 ($n = 2$) : $c(t) = c_0 - c_0 kt + \frac{c_0 k^2}{2} t^2 + o(t^2)$.

Ces DL sont tracés sur la Figure II.1.1. On voit que, au voisinage de $t = 0$, la précision s'améliore avec l'ordre du DL.

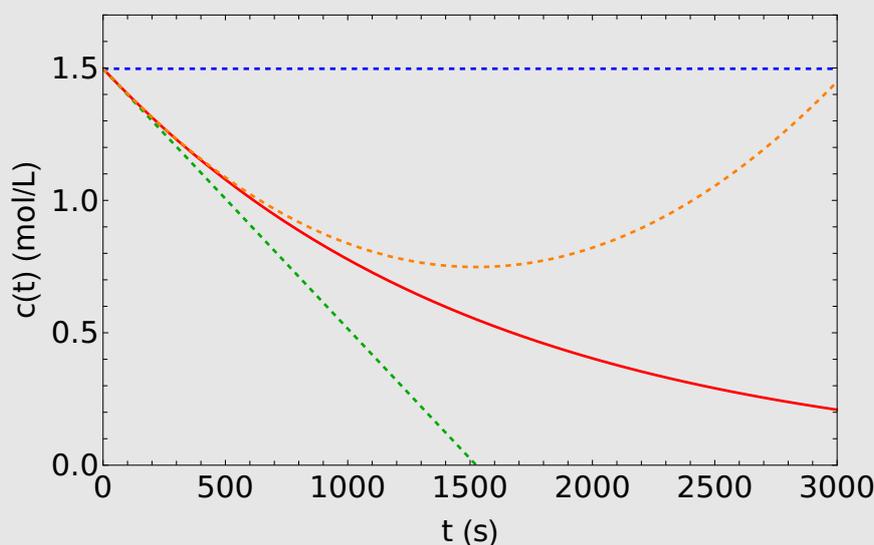


FIGURE II.1.1 – Concentration en fonction du temps $c(t) = c_0 e^{-kt}$ (en rouge) et développements limités d'ordre 0 (en bleu), 1 (en vert) et 2 (en orange) au voisinage de $t = 0$.

Développement limité d'une fonctions de deux variables (optionnel)

On peut généraliser la formule de Taylor aux fonctions de plusieurs variables. Par exemple, pour une fonction de deux variables réelles $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ de classe C^2 , le développement de Taylor au voisinage du point (x_0, y_0) à l'ordre 2 est :

$$f(x_0 + h_1, y_0 + h_2) = f(x_0, y_0) + \frac{\partial f}{\partial x}(x_0, y_0)h_1 + \frac{\partial f}{\partial y}(x_0, y_0)h_2 + \frac{1}{2} \left(\frac{\partial^2 f}{\partial x^2}(x_0, y_0)h_1^2 + \frac{\partial^2 f}{\partial y^2}(x_0, y_0)h_2^2 + 2 \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)h_1 h_2 \right) + o(h_1^2 + h_2^2).$$

2.2 Différentielles

Définition (Différentielle d'une fonction d'une variable réelle). Soit une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ dérivable. La différentielle de f au point x_0 est une fonction de \mathbb{R} dans \mathbb{R} définie par :

$$\begin{aligned} df_{x_0} : \mathbb{R} &\rightarrow \mathbb{R} \\ h &\mapsto df_{x_0}(h) = f'(x)h \end{aligned}$$

Dans cette définition, x_0 est un paramètre fixé, alors que h est la variable.

La différentielle $df_{x_0}(h)$ représente la variation de f au premier ordre induite par une variation de x d'une quantité h en partant de x_0 . Autrement dit, il s'agit du terme du premier ordre dans le développement limité de f par rapport à x au voisinage de x_0 . Puisque l'on pense habituellement à h comme une petite variation de x , on note très souvent $h = dx$, et on écrit en sciences la différentielle en notation compacte sans forcément préciser le point x_0 auquel on la calcule :

$$df = f'(x)dx.$$

Pour une fonction d'une seule variable réelle, le concept de différentielle n'apporte pas grand chose de plus que le concept de dérivée. Le concept de différentielle est plus intéressant pour une fonction de plusieurs variables. Nous donnerons le cas de deux variables réelles, mais tout se généralise sans difficulté à plus de variables réelles.

Définition (Différentielle d'une fonction de deux variables réelles). Soit une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ de classe C^1 . La différentielle de f au point (x_0, y_0) est une fonction de \mathbb{R}^2 dans \mathbb{R} définie par :

$$\begin{aligned} df_{(x_0, y_0)} : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (h_1, h_2) &\mapsto df_{(x_0, y_0)}(h_1, h_2) = \frac{\partial f}{\partial x}(x_0, y_0)h_1 + \frac{\partial f}{\partial y}(x_0, y_0)h_2. \end{aligned}$$

Dans cette définition, x_0 et y_0 sont des paramètres fixés, alors que h_1 et h_2 sont les variables.

La différentielle $df_{(x_0, y_0)}(h_1, h_2)$ représente la variation de f au premier ordre induite par une variation de x d'une quantité h_1 et par une variation de y d'une quantité h_2 , en partant du point (x_0, y_0) . Autrement dit, il s'agit du terme du premier ordre dans le développement limité de f par rapport à x et y au voisinage de (x_0, y_0) . Puisque l'on pense habituellement à h_1 et h_2 comme des petites variations de x et y , on note très souvent $h_1 = dx$ et $h_2 = dy$, et on écrit en sciences la différentielle en notation compacte sans forcément préciser le point (x_0, y_0) auquel on la calcule :

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

La différentielle permet de réunir dans un seul objet mathématique toutes les dérivées partielles premières d'une fonction. De plus, les mêmes règles de dérivation s'appliquent pour les différentielles. Par exemple, si f et g sont deux fonctions de deux variables réelles de classe C^1 , on a (partout où cela est bien défini) :

- Somme de 2 fonctions : $d(f + g) = df + dg$
- Produit de 2 fonctions : $d(f \times g) = df \times g + f \times dg$
- Fonction à une puissance (r est une constante réelle) : $d(f^r) = r f^{r-1} df$

En thermodynamique, les différentielles de fonctions à plusieurs variables sont énormément utilisées. Voici un exemple.

Exemple. (*Différentielle de l'enthalpie libre*). L'enthalpie libre G d'une substance étant une fonction de la variable pression P et de la variable température T , la différentielle de G s'écrit :

$$dG = \frac{\partial G}{\partial P} dP + \frac{\partial G}{\partial T} dT.$$

On peut interpréter dG comme la petite variation de G induite par une petite variation de P d'une quantité dP et d'une petite variation de T d'une quantité dT . La quantité $\partial G / \partial P = V$ correspond au volume V et la quantité $\partial G / \partial T = -S$ correspond à l'opposé de l'entropie S . On peut donc écrire :

$$dG = V dP - S dT.$$

D'après cette relation, si on augmente d'une petite quantité la pression à température constante, alors l'enthalpie libre augmente d'une petite quantité proportionnelle au volume. Par ailleurs, si on augmente d'une petite quantité la température à pression constante, alors l'enthalpie libre diminue d'une petite quantité proportionnelle à l'entropie.

Les différentielles sont aussi utiles pour les calculs de propagation d'erreur. Dans l'exemple ci-dessous, si on mesure la pression P avec une erreur ΔP et la température T avec une erreur ΔT , et que l'on calcule l'enthalpie libre $G(P, T)$, alors les erreurs sur P et T se propagent sur G et on peut estimer cette erreur en utilisant l'expression

de la différentielle de G :

$$\Delta G = \frac{\partial G}{\partial P} \Delta P + \frac{\partial G}{\partial T} \Delta T.$$

Cette estimation de l'erreur commise sur G sera bonne si ΔP et ΔT sont suffisamment petits.

Exemple. (*Propagation d'erreur sur des concentrations*). Pour l'exemple de la réaction chimique de transformation du cyclopropane en propène, on peut propager les erreurs Δc sur la concentration c mesurée au cours du temps vers la quantité tracée $\ln(c/c^\ominus)$ (où c^\ominus est une concentration de référence constante) en utilisant la différentielle de la fonction $f : c \mapsto \ln(c/c^\ominus)$:

$$df = f'(c)dc = \frac{1}{c}dc.$$

Les erreurs sur $\ln(c/c^\ominus)$ sont donc calculées par :

$$\Delta \ln(c/c^\ominus) = \frac{1}{c} \Delta c.$$

3 Opérateurs différentiels

Dans cette section, nous considérerons l'espace à trois dimensions \mathbb{R}^3 muni du repère cartésien $(O, \vec{i}, \vec{j}, \vec{k})$. Mais on pourrait généraliser à d'autres dimensions.

Définition (Gradient). Soit une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ dérivable. Le gradient de f , noté $\overrightarrow{\text{grad}}f$, est une fonction de \mathbb{R}^3 dans \mathbb{R}^3 définie par :

$$\begin{aligned} \overrightarrow{\text{grad}}f : \quad \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ (x, y, z) \mapsto \overrightarrow{\text{grad}}f(x, y, z) &= \begin{pmatrix} \frac{\partial f}{\partial x}(x, y, z) \\ \frac{\partial f}{\partial y}(x, y, z) \\ \frac{\partial f}{\partial z}(x, y, z) \end{pmatrix} \\ &= \frac{\partial f}{\partial x}(x, y, z)\vec{i} + \frac{\partial f}{\partial y}(x, y, z)\vec{j} + \frac{\partial f}{\partial z}(x, y, z)\vec{k}. \end{aligned}$$

Le gradient $\overrightarrow{\text{grad}}f$ est donc un champ de vecteurs (ou fonction vectorielle). Comme pour la différentielle, le gradient permet de réunir dans un seul objet mathématique toutes les dérivées partielles premières d'une fonction. La direction de $\overrightarrow{\text{grad}}f(x, y, z)$ donne, en partant du point (x, y, z) , la direction suivant laquelle la fonction f aug-

mente le plus vite.

Exemple. (*Force en mécanique classique*). En mécanique classique, une particule à la position (x, y, z) ayant une énergie potentielle $E_p(x, y, z)$ subit une force vectorielle $\vec{F}(x, y, z)$ donnée par l'opposé du gradient de E_p :

$$\vec{F}(x, y, z) = -\overrightarrow{\text{grad}}E_p(x, y, z).$$

Par exemple, pour l'énergie potentielle de pesanteur, on a (en prenant l'axe z pointant vers le haut) : $E_p(x, y, z) = mgz$, où m est la masse de la particule et g est l'accélération de la pesanteur. La force de pesanteur est alors $\vec{F}(x, y, z) = -\overrightarrow{\text{grad}}E_p(x, y, z) = -mg\vec{k}$. Il s'agit d'un champ de vecteurs constant dans l'espace, la particule subit donc la même force quelle que soit sa position dans l'espace.

Exemple. (*Champ électrique d'une charge ponctuelle*). En électrostatique, le champ électrique $\vec{E}(x, y, z)$ créé par un potentiel électrique $\phi(x, y, z)$ est donné par l'opposé du gradient de ϕ :

$$\vec{E}(x, y, z) = -\overrightarrow{\text{grad}}\phi(x, y, z).$$

Par exemple, pour une charge électrique élémentaire positive (comme un proton), supposée ponctuelle et située au centre du repère, le potentiel électrique est :

$$\phi(x, y, z) = \frac{e}{4\pi\epsilon_0 r},$$

où e est la charge élémentaire, ϵ_0 est la permittivité du vide, et $r = \sqrt{x^2 + y^2 + z^2}$. Le champ électrique correspondant est alors :

$$\vec{E}(x, y, z) = -\overrightarrow{\text{grad}}\phi(x, y, z) = \frac{e}{4\pi\epsilon_0} \left(\frac{x}{r^3}\vec{i} + \frac{y}{r^3}\vec{j} + \frac{z}{r^3}\vec{k} \right) = \frac{e}{4\pi\epsilon_0 r^3}\vec{r},$$

où $\vec{r} = x\vec{i} + y\vec{j} + z\vec{k}$.

Notons que le gradient s'applique à une fonction scalaire, c'est-à-dire à valeurs dans \mathbb{R} . Le gradient ne s'applique pas à un champ de vecteurs. Pour un champ de vecteurs, par exemple à trois dimensions,

$$\begin{aligned} \vec{A} : \quad \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ (x, y, z) &\mapsto \vec{A}(x, y, z) = \begin{pmatrix} A_1(x, y, z) \\ A_2(x, y, z) \\ A_3(x, y, z) \end{pmatrix}, \end{aligned}$$

on peut définir la divergence et le rotationnel.

Définition (Divergence). Soit un champ de vecteur $\vec{A} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ dérivable. La divergence de \vec{A} , noté $\text{div}\vec{A}$, est une fonction de \mathbb{R}^3 dans \mathbb{R} définie par :

$$\begin{aligned} \text{div}\vec{A} : \quad \mathbb{R}^3 &\rightarrow \mathbb{R} \\ (x, y, z) &\mapsto \text{div}\vec{A}(x, y, z) = \frac{\partial A_1}{\partial x}(x, y, z) + \frac{\partial A_2}{\partial y}(x, y, z) + \frac{\partial A_3}{\partial z}(x, y, z). \end{aligned}$$

La divergence $\text{div}\vec{A}$ est donc une fonction scalaire (et non pas un champ de vecteurs).

Le rotationnel pourra être considéré comme optionnel pour ce cours.

Définition (Rotationnel). Soit un champ de vecteur $\vec{A} : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ dérivable. Le rotationnel de \vec{A} , noté $\vec{\text{rot}}\vec{A}$, est une fonction de \mathbb{R}^3 dans \mathbb{R}^3 définie par :

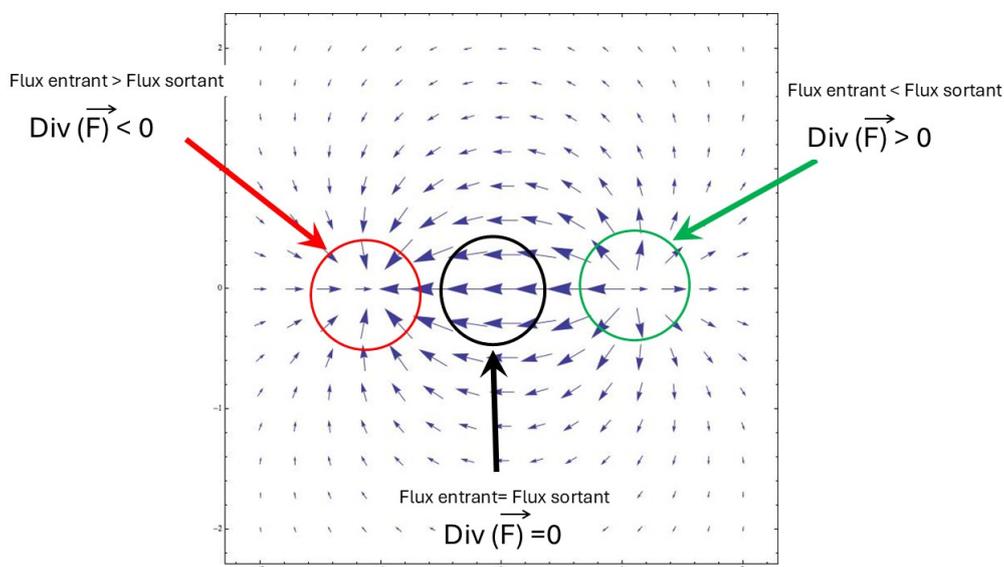
$$\begin{aligned} \vec{\text{rot}}\vec{A} : \quad \mathbb{R}^3 &\rightarrow \mathbb{R}^3 \\ (x, y, z) &\mapsto \vec{\text{rot}}\vec{A}(x, y, z) = \begin{pmatrix} \frac{\partial A_3}{\partial y}(x, y, z) - \frac{\partial A_2}{\partial z}(x, y, z) \\ \frac{\partial A_1}{\partial z}(x, y, z) - \frac{\partial A_3}{\partial x}(x, y, z) \\ \frac{\partial A_2}{\partial x}(x, y, z) - \frac{\partial A_1}{\partial y}(x, y, z) \end{pmatrix}. \end{aligned}$$

Le rotationnel $\vec{\text{rot}}\vec{A}$ est un champ de vecteurs.

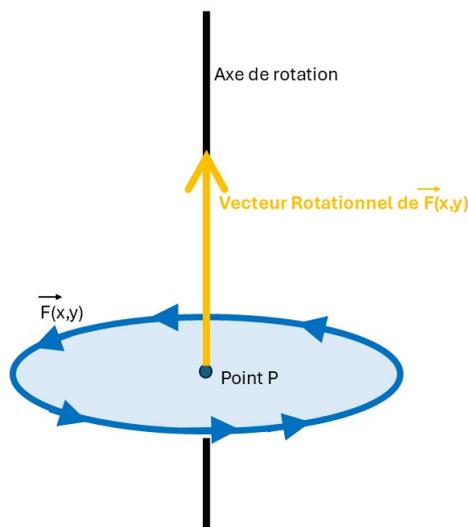
Divergence et rotationnel renseignent tous deux sur les variations du champ de vecteur \vec{A} .

Signification de la divergence et du rotationnel :

- La divergence est une mesure de la différence entre le flux vectoriel entrant et celui sortant dans une petite surface (2D) ou un petit volume (3D) autour d'un point donné.



- Le rotationnel est un vecteur perpendiculaire à un flux de vecteurs en rotation et dont la norme est une mesure de la vitesse de rotation autour d'un point donné.



Exemple. (*Divergence du champ électrique et du champ magnétique*). La divergence intervient par exemple en électromagnétisme, le champ de la physique qui étudie les interactions électriques et magnétiques, et qui est utilisé dans de nombreux domaines de la physicochimie. Un champ électrique $\vec{E}(x, y, z)$ satisfait l'équation de Maxwell-Gauss

$$\operatorname{div} \vec{E}(x, y, z) = \frac{\rho(x, y, z)}{\epsilon_0},$$

où $\rho(x, y, z)$ est la densité de charge et ϵ_0 est la permittivité du vide. De manière similaire, un champ magnétique $\vec{B}(x, y, z)$ satisfait l'équation de Maxwell-Thomson

$$\operatorname{div} \vec{B}(x, y, z) = 0.$$

Exemple. (*Rotationnel du champ électrique et du champ magnétique*). Le rotationnel intervient aussi en électromagnétisme. Un champ électrique $\vec{E}(x, y, z)$ indépendant du temps satisfait l'équation de Maxwell-Faraday

$$\vec{\operatorname{rot}} \vec{E}(x, y, z) = \vec{0}.$$

De manière similaire, un champ magnétique $\vec{B}(x, y, z)$ indépendant du temps satisfait l'équation de Maxwell-Ampère

$$\vec{\operatorname{rot}} \vec{B}(x, y, z) = \mu_0 \vec{j}(x, y, z),$$

où $\vec{j}(x, y, z)$ est la densité de courant (électrique) et μ_0 est la perméabilité du vide. Ces quatre équations (Maxwell-Gauss, Maxwell-Thomson, Maxwell-Faraday et Maxwell-

(Ampère), collectivement appelées équations de Maxwell, permettent de calculer $\vec{E}(x, y, z)$ et $\vec{B}(x, y, z)$ à partir de la densité de charge et de la densité de courant, respectivement.

Le gradient $\overrightarrow{\text{grad}}$ est un opérateur différentiel, c'est-à-dire une fonction qui prend une fonction f et qui renvoie son gradient $\overrightarrow{\text{grad}}f$:

$$\overrightarrow{\text{grad}} : f \mapsto \overrightarrow{\text{grad}}(f).$$

De même, la divergence div et le rotationnel $\overrightarrow{\text{rot}}$ sont des opérateurs différentiels, mais qui agissent sur un champ de vecteurs \vec{A} :

$$\text{div} : \vec{A} \mapsto \text{div}(\vec{A}),$$

$$\overrightarrow{\text{rot}} : \vec{A} \mapsto \overrightarrow{\text{rot}}(\vec{A}).$$

Il existe une notation alternative très pratique et très utilisée pour le gradient : la notation utilisant le symbole $\vec{\nabla}$ qui s'appelle « nabla ». Dans cette notation, on écrit :

$$\overrightarrow{\text{grad}}f = \vec{\nabla}f,$$

Le symbole $\vec{\nabla}$ est un opérateur différentiel et on peut le penser comme un « vecteur » ayant pour composantes les trois opérateurs de dérivées partielles :

$$\vec{\nabla} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} = \frac{\partial}{\partial x} \vec{i} + \frac{\partial}{\partial y} \vec{j} + \frac{\partial}{\partial z} \vec{k}.$$

Cette écriture doit être interprétée de la façon suivante : lorsque ce symbole $\vec{\nabla}$ agit sur une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$, il faut prendre les dérivées partielles de f :

$$\vec{\nabla}f = \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{pmatrix} = \frac{\partial f}{\partial x} \vec{i} + \frac{\partial f}{\partial y} \vec{j} + \frac{\partial f}{\partial z} \vec{k}.$$

Avec le même symbole $\vec{\nabla}$, la divergence s'écrit :

$$\text{div}\vec{A} = \vec{\nabla} \cdot \vec{A},$$

où \cdot désigne le produit scalaire. D'après cette écriture, on peut voir la divergence de \vec{A} comme le produit scalaire entre $\vec{\nabla}$ et \vec{A} :

$$\vec{\nabla} \cdot \vec{A} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \cdot \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \frac{\partial A_1}{\partial x} + \frac{\partial A_2}{\partial y} + \frac{\partial A_3}{\partial z}.$$

De même, le rotationnel s'écrit :

$$\overrightarrow{\text{rot}} \vec{A} = \vec{\nabla} \wedge \vec{A},$$

où \wedge désigne le produit vectoriel. D'après cette écriture, on peut voir le rotationnel de \vec{A} comme le produit vectoriel de $\vec{\nabla}$ et \vec{A} :

$$\vec{\nabla} \wedge \vec{A} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \wedge \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix} = \begin{pmatrix} \frac{\partial A_3}{\partial y} - \frac{\partial A_2}{\partial z} \\ \frac{\partial A_1}{\partial z} - \frac{\partial A_3}{\partial x} \\ \frac{\partial A_2}{\partial x} - \frac{\partial A_1}{\partial y} \end{pmatrix}.$$

Ces écritures avec $\vec{\nabla}$ permettent de retrouver facilement les expressions de la divergence et du rotationnel.

Enfin, il existe un autre opérateur différentiel très important : le laplacien.

Définition (Laplacien). Soit une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ dérivable deux fois. Le laplacien de f , noté Δf , est une fonction de \mathbb{R}^3 dans \mathbb{R} définie par :

$$\Delta f : \mathbb{R}^3 \rightarrow \mathbb{R}$$

$$(x, y, z) \mapsto \Delta f(x, y, z) = \frac{\partial^2 f}{\partial x^2}(x, y, z) + \frac{\partial^2 f}{\partial y^2}(x, y, z) + \frac{\partial^2 f}{\partial z^2}(x, y, z).$$

Le laplacien Δf est donc une fonction scalaire correspondant à la somme des trois dérivées secondes de f . Le symbole Δ est un opérateur différentiel d'ordre 2 qui peut s'écrire :

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

On peut aussi écrire le laplacien avec le symbole $\vec{\nabla}$:

$$\Delta f = \vec{\nabla}^2 f.$$

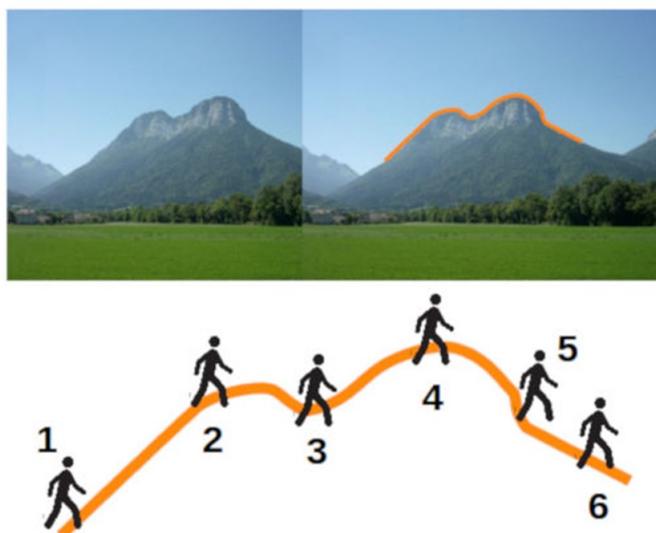
Ici, $\vec{\nabla}^2$ doit être interprété comme le produit scalaire de $\vec{\nabla}$ et $\vec{\nabla}$:

$$\vec{\nabla}^2 = \vec{\nabla} \cdot \vec{\nabla} = \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial}{\partial x} \\ \frac{\partial}{\partial y} \\ \frac{\partial}{\partial z} \end{pmatrix} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2},$$

où le “produit” des opérateurs dérivées premières doit être compris comme leur composition, c'est-à-dire leur application successive, conduisant aux opérateurs dérivées secondes.

Signification du laplacien :

Le laplacien est une mesure de la variation de pente. Une pente de valeur constante autour d'un point aboutira à une valeur de laplacien nulle.



En 1 et 6, le laplacien est de valeur nulle car la pente est régulière (pas de courbure et donc de variation de la valeur de la pente).

En 2 et 4, autour d'un sommet la pente varie et le laplacien est non nul (négatif dans ce cas).

En 3 et 5, aux alentours d'une vallée, la pente se modifie également et le laplacien est non nul (positif dans ce cas).

Exemple. (*Équation de Poisson*). En électrostatique, nous avons vu que le champ électrique $\vec{E}(x, y, z)$ est donné par l'opposé du gradient du potentiel électrique $\phi(x, y, z)$:

$$\vec{E}(x, y, z) = -\overrightarrow{\text{grad}} \phi(x, y, z).$$

Par ailleurs, d'après l'équation de Maxwell-Gauss, la divergence du champ électrique $\vec{E}(x, y, z)$ est liée à la densité de charge :

$$\text{div} \vec{E}(x, y, z) = \rho(x, y, z) / \epsilon_0.$$

En combinant ces deux équations, on arrive à :

$$\text{div} \left(-\overrightarrow{\text{grad}} \phi(x, y, z) \right) = \frac{\rho(x, y, z)}{\epsilon_0},$$

ce qui donne, d'après les définitions de div et $\overrightarrow{\text{grad}}$,

$$-\Delta \phi(x, y, z) = \frac{\rho(x, y, z)}{\epsilon_0}.$$

Cette équation permet de calculer le potentiel électrique directement à partir de la densité de charge et s'appelle l'équation de Poisson. Remarquons que le résultat $\text{div}(\overrightarrow{\text{grad}} f) = \Delta f$ est facile à voir avec les notations $\vec{\nabla}$:

$$\text{div}(\overrightarrow{\text{grad}} f) = \vec{\nabla} \cdot (\vec{\nabla} f) = (\vec{\nabla} \cdot \vec{\nabla}) f = \vec{\nabla}^2 f = \Delta f.$$

Exemple. (*Équation de Schrödinger*). Les orbitales ψ d'un système atomique ou moléculaire peuvent être déterminées par l'équation de Schrödinger :

$$-\frac{\hbar^2}{2m_e} \Delta \psi(x, y, z) + V(x, y, z) \psi(x, y, z) = E \psi(x, y, z),$$

où m_e est la masse de l'électron, $\hbar = h/(2\pi)$ est la constante de Planck réduite, $V(x, y, z)$ est l'énergie potentielle de l'électron, et E est l'énergie de l'orbitale ψ . Le premier terme de cette équation implique l'opérateur différentiel laplacien Δ et est lié à l'énergie cinétique de l'électron.

4 Dérivées numériques

Il arrive très souvent en sciences de ne pas pouvoir calculer exactement la dérivée d'une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$. Cela peut être le cas si f a une expression très compliquée ou si elle est le résultat de mesures expérimentales. On ne connaît alors f qu'en

certains points. Dans ce cas là, on peut calculer la dérivée de f de façon approchée numériquement.

Pour cela, on se base sur des développements limités de f . Pour faire simple, on supposera que la fonction f est de classe C^∞ et on peut donc faire un développement de Taylor à n'importe quel ordre. Le DL de f au voisinage de x_0 à l'ordre 1 est :

$$f(x_0 + h) = f(x_0) + f'(x_0)h + o(h).$$

Cela conduit à l'expression suivante de la dérivée de f en x_0 :

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0)}{h} + o(h^0),$$

Dans la limite $h \rightarrow 0$, on retrouve la définition de la dérivée exacte. Pour un h non nul, on obtient une approximation de la dérivée de f en x_0 :

$$f'(x_0) \approx \frac{f(x_0 + h) - f(x_0)}{h}.$$

On dit que c'est l'approximation de la dérivée par différence finie d'ordre 1 par rapport à h , ce qui veut dire que l'erreur que l'on fait sur $f'(x_0)$ est proportionnelle à h . En pratique, pour que l'approximation soit bonne il faut que h soit suffisamment petit par rapport aux variations de f .

On peut avoir une meilleure approximation de la dérivée en considérant les DL à l'ordre 2 suivants :

$$f(x_0 + h) = f(x_0) + f'(x_0)h + \frac{f''(x_0)}{2}h^2 + o(h^2),$$

et

$$f(x_0 - h) = f(x_0) - f'(x_0)h + \frac{f''(x_0)}{2}h^2 + o(h^2).$$

En faisant la différence entre ces deux DL, on obtient :

$$f(x_0 + h) - f(x_0 - h) = 2f'(x_0)h + o(h^2),$$

et donc

$$f'(x_0) = \frac{f(x_0 + h) - f(x_0 - h)}{2h} + o(h).$$

Pour un h non nul, on obtient l'approximation suivante de la dérivée de f en x_0 :

$$f'(x_0) \approx \frac{f(x_0 + h) - f(x_0 - h)}{2h}.$$

Il s'agit de l'approximation de la dérivée par différence finie d'ordre 2 par rapport à h , c'est-à-dire que l'erreur sur $f'(x_0)$ est proportionnelle à h^2 .

On peut aussi estimer numériquement la dérivée seconde. En faisant la somme des deux DL précédents à l'ordre 2, on obtient :

$$f(x_0 + h) + f(x_0 - h) = 2f(x_0) + f''(x_0)h^2 + o(h^2),$$

ce qui conduit à

$$f''(x_0) = \frac{f(x_0 + h) + f(x_0 - h) - 2f(x_0)}{h^2} + o(h^0).$$

Pour un h non nul, on arrive donc à l'approximation suivante de la dérivée seconde de f en x_0 :

$$f''(x_0) \approx \frac{f(x_0 + h) + f(x_0 - h) - 2f(x_0)}{h^2},$$

qui est l'approximation de la dérivée seconde par différence finie d'ordre 1 par rapport à h .

Exemple. (*Spectroscopie UV-Visible*). La mesure du spectre UV-Visible de la molécule d'éthylène donne une série de valeurs de l'absorbance A pour des longueurs d'onde de 100 à 300 nm par incréments de $h = 1$ nm (voir Figure 4 dans l'introduction générale). Pour déterminer la position du maximum, on peut par exemple calculer numériquement la dérivée $A'(\lambda)$ avec l'approximation par différence finie d'ordre 1

$$A'(\lambda) \approx \frac{A(\lambda + h) - A(\lambda)}{h},$$

ou l'approximation par différence finie d'ordre 2

$$A'(\lambda) \approx \frac{A(\lambda + h) - A(\lambda - h)}{2h}.$$

Ces dérivées numériques sont reportées dans la Figure II.1.2. On voit que l'approximation par différence finie d'ordre 1 donne une dérivée beaucoup plus bruitée que l'approximation par différence finie d'ordre 2. En effet, la formule de la dérivée par différence finie d'ordre 1 utilise la fonction en deux points adjacents (λ et $\lambda + h$), ce qui a tendance à amplifier les erreurs dans les données. Au contraire, la formule de la dérivée par différence finie d'ordre 2 utilise la fonction en deux points plus éloignés ($\lambda - h$ et $\lambda + h$), ce qui a tendance à partiellement annuler les erreurs dans les données. Dans les deux cas, on voit que la dérivée s'annule autour de $\lambda \approx 180$ nm, ce qui correspond au maximum de A .

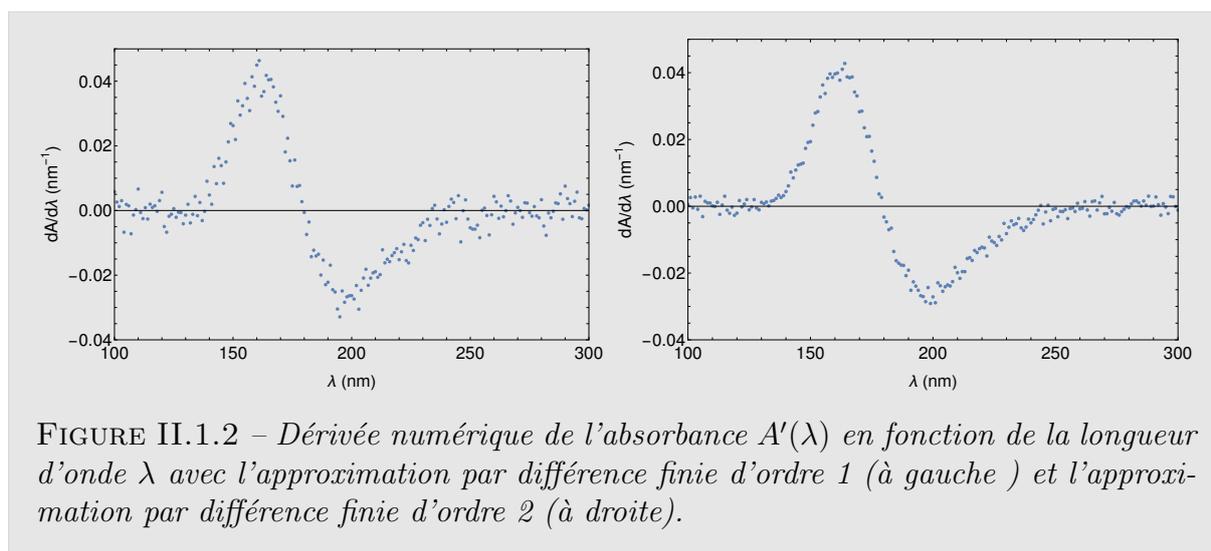


FIGURE II.1.2 – Dérivée numérique de l'absorbance $A'(\lambda)$ en fonction de la longueur d'onde λ avec l'approximation par différence finie d'ordre 1 (à gauche) et l'approximation par différence finie d'ordre 2 (à droite).

Annexe : Règles de dérivation

Nous rappelons ici les règles usuelles de dérivation. Comme il est très courant de le faire, nous utiliserons parfois l'abus de notation « $[f(x)]'$ » ou « $(f(x))'$ » pour désigner la dérivée $f'(x)$ en x .

On a 4 règles générales très importantes à retenir. Soit f et g deux fonctions dérivables, partout où cela est bien défini, les fonctions suivantes sont dérivables et on a :

1. Somme de 2 fonctions : $[f(x) + g(x)]' = f'(x) + g'(x)$
2. Produit de 2 fonctions : $[f(x) g(x)]' = f'(x) g(x) + f(x) g'(x)$
3. Fonction à une puissance (r est une constante réelle) : $[f(x)^r]' = r f'(x) f(x)^{r-1}$
4. Composition de 2 fonctions : $[f(g(x))]' = g'(x) f'(g(x))$

En appliquant la règle 3 pour $r = -1$, on peut déduire : $[1/f(x)]' = -f'(x)/f(x)^2$.

En appliquant les règles 2 et 3, on peut déduire : $[f(x)/g(x)]' = [f'(x) g(x) - f(x) g'(x)]/g(x)^2$.

En appliquant la règle 3 pour $r = 1/2$, on peut déduire la dérivée de la fonction racine carrée : $(\sqrt{x})' = (x^{1/2})' = (1/2)x' x^{1/2-1} = (1/2)x^{-1/2} = 1/(2\sqrt{x})$.

Rappelons aussi les dérivées de quelques fonctions usuelles :

- $(e^x)' = e^x$
- $(\ln x)' = 1/x$
- $(\cos x)' = -\sin x$
- $(\sin x)' = \cos x$

En utilisant la dérivée $(f/g)'$, on peut déduire la dérivée de la fonction tangente : $(\tan x)' = (\sin x / \cos x)' = (\cos x \cos x + \sin x \sin x) / (\cos x)^2 = 1 / (\cos x)^2$.

En utilisant la règle 4, on peut déduire les dérivées des fonctions composées suivantes :

- $(e^{g(x)})' = g'(x) e^{g(x)}$
- $(\ln g(x))' = g'(x) / g(x)$
- $(\cos g(x))' = -g'(x) \sin g(x)$
- $(\sin g(x))' = g'(x) \cos g(x)$

Chapitre II.2

Outils d'analyse intégrale

Le calcul intégral est un outil mathématique tout aussi important que le calcul différentiel. Alors que le calcul différentiel renseigne sur le comportement local d'une fonction, le calcul intégral donne une information globale sur une fonction. Dans ce chapitre, après avoir rappelé le concept d'intégrale d'une fonction d'une variable, nous introduirons les intégrales de fonctions de plusieurs variables, et en particulier les intégrales doubles et triples. On rencontre ce type d'intégrales par exemple en chimie quantique. Nous verrons aussi comment calculer numériquement une intégrale à partir de données. Enfin, nous aborderons la transformation de Fourier qui est un outil fondamental du calcul intégral très largement utilisé en sciences et en particulier en physicochimie.

1 Intégrale d'une fonction d'une variable réelle

1.1 Définition

La notion d'intégrale est historiquement liée à la notion du calcul d'aire de figures géométriques dont la frontière est circonscrite par une fonction d'une variable réelle. Ce sont les grecs qui en sont les précurseurs mais ce sont Mercator (1620-1687), Leibniz (1646-1716) et Newton (1643-1727) qui vont mettre en évidence que l'intégrale d'une fonction positive sur un intervalle $[a, b]$ évalue « l'aire sous la courbe » et formaliser le calcul intégral. Le concept d'intégrale a été raffiné depuis, permettant ainsi de les calculer pour des fonctions de moins en moins régulières. Dans les années 1850, Riemann eu l'idée de définir l'intégrale comme la limite d'une somme. Le concept d'intégrale n'a cessé d'être amélioré depuis permettant ainsi de les calculer pour des fonctions de moins en moins régulières (intégrale de Lebesgue et intégrale de Kurzweil-Henstock par exemple).

Pour faire simple, on donne ici la définition de Riemann pour le cas d'une fonction continue, mais il existe des généralisations à des fonctions moins régulières.

Définition (Intégrale d'une fonction d'une variable réelle continue). Soit une fonction $f : D \rightarrow \mathbb{R}$ continue sur un intervalle $[a, b] \subseteq D$. On divise l'intervalle $[a, b]$ en n sous-intervalles réguliers $a = x_0 < x_1 < \dots < x_{n-1} < x_n = b$ avec $x_i = a + i\Delta x$ et $\Delta x = (b-a)/n$. L'intégrale de f sur $[a, b]$ est la limite quand $n \rightarrow \infty$ de la somme de Riemann

$$\int_a^b f(x)dx = \lim_{n \rightarrow \infty} \sum_{i=0}^{n-1} f(x_i)\Delta x.$$

Comme illustré dans la Figure II.2.1, la somme de Riemann représente la somme d'aires de rectangles qui converge vers l'aire sous la courbe de f entre a et b dans la limite $n \rightarrow \infty$. Il s'agit de l'aire algébrique, c'est-à-dire que l'aire est positive dans les zones où $f(x) > 0$ et négative dans les zones où $f(x) < 0$.

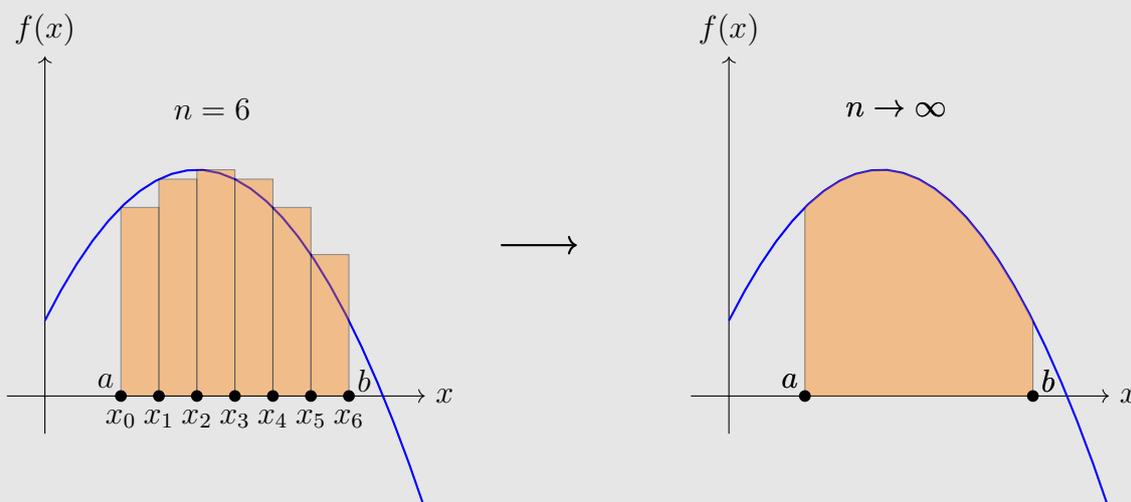


FIGURE II.2.1 – Définition de l'intégrale d'une fonction f entre a et b par limite de la somme de Riemann.

L'intégrale peut aussi être notée de façon plus compacte, sans spécifier la variable x :

$$\int_a^b f = \int_a^b f(x)dx.$$

On définit de plus l'intégrale de f sur l'intervalle inversé $[b, a]$ comme l'opposé de l'intégrale de f sur l'intervalle $[a, b]$:

$$\int_b^a f(x)dx = - \int_a^b f(x)dx.$$

Si on a une fonction f continue uniquement sur un intervalle du type $[a, b]$, c'est-à-

dire sans inclure la valeur supérieure b , alors on définit l'intégrale (dite impropre ou généralisée) de f sur $]a, b[$ comme la limite de l'intégrale de f sur $[a, y]$ quand $y \rightarrow b^-$, si cette limite existe et est finie,

$$\int_a^b f(x)dx = \lim_{y \rightarrow b^-} \int_a^y f(x)dx.$$

De même, pour une fonction f continue uniquement sur un intervalle du type $]a, b]$, c'est-à-dire sans inclure la valeur inférieure a , on définit l'intégrale de f sur $]a, b]$ comme la limite de l'intégrale de f sur $[y, b]$ quand $y \rightarrow a^+$, si cette limite existe et est finie,

$$\int_a^b f(x)dx = \lim_{y \rightarrow a^+} \int_y^b f(x)dx.$$

Enfin, pour une fonction f continue uniquement sur un intervalle du type $]a, b[$, on définit l'intégrale de f sur $]a, b[$ comme la double limite suivante, si cette limite existe et est finie,

$$\int_a^b f(x)dx = \lim_{y_1 \rightarrow a^+} \lim_{y_2 \rightarrow b^-} \int_{y_1}^{y_2} f(x)dx.$$

En particulier, ceci permet de définir l'intégrale avec une borne infinie : $a = -\infty$ et/ou $b = +\infty$.

À chaque fois que l'on peut définir l'intégrale d'une fonction f sur un intervalle, on dira que cette fonction est intégrable sur cet intervalle.

1.2 Propriétés essentielles

L'intégrale de f vérifie plusieurs propriétés essentielles.

Théorème (Additivité ou Relation de Chasles). Soit $f : D \rightarrow \mathbb{R}$ une fonction intégrable sur l'intervalle $]a, b[$. Pour tout $c \in]a, b[$, on a :

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Théorème (Linéarité). Soit $f, g : D \rightarrow \mathbb{R}$ deux fonctions intégrables sur l'intervalle $]a, b[$ et $\lambda, \mu \in \mathbb{R}$. On a :

$$\int_a^b (\lambda f(x) + \mu g(x))dx = \lambda \int_a^b f(x)dx + \mu \int_a^b g(x)dx.$$

Attention, ce dernier théorème ne doit pas laisser penser que ce qui est possible pour

l'addition de fonction l'est pour un produit de fonctions. En effet :

$$\int_a^b f(x) g(x) dx \neq \left(\int_a^b f(x) dx \right) \times \left(\int_a^b g(x) dx \right).$$

Théorème (Croissance). Soit $a \leq b$ et $f, g : D \rightarrow \mathbb{R}$ deux fonctions intégrables sur l'intervalle $]a, b[$. Si, pour tout $x \in]a, b[$, $f(x) \leq g(x)$, alors :

$$\int_a^b f(x) dx \leq \int_a^b g(x) dx.$$

En particulier, si pour tout $x \in]a, b[$, $f(x) \geq 0$, alors

$$\int_a^b f(x) dx \geq 0.$$

1.3 Primitives et théorème fondamental de l'analyse

Rappelons qu'une primitive d'une fonction f est une fonction F dérivable telle que la dérivée de F soit égale à f , c'est-à-dire $F' = f$. Si on connaît une primitive de f alors on peut facilement calculer l'intégrale de f par le théorème suivant.

Théorème (Théorème fondamental de l'analyse). Soit $f : D \rightarrow \mathbb{R}$ une fonction intégrable sur l'intervalle $]a, b[$. Si F est une primitive de f , alors on a :

$$\int_a^b f(x) dx = \left[F(x) \right]_a^b = F(b) - F(a).$$

Si $b = +\infty$, alors il faut interpréter $F(b)$ comme $\lim_{x \rightarrow +\infty} F(x)$. De même, si $a = -\infty$, alors il faut interpréter $F(a)$ comme $\lim_{x \rightarrow -\infty} F(x)$.

Toute fonction f continue admet une primitive. Or, si F est une primitive de f alors la fonction $G : x \mapsto F(x) + K$ où K est n'importe quel nombre réel (constante) est aussi une primitive de f . Donc, toute fonction f continue admet en fait une infinité de primitives qui ne diffèrent que par une constante additive.

Lorsque f est continue, alors la fonction F définie par :

$$F : t \mapsto F(t) = \int_a^t f(x) dx,$$

est la primitive de f qui s'annule en $t = a$, c'est-à-dire $F(a) = 0$. On désigne souvent

toute primitive de f par $F(x) = \int f(x)dx$, sous forcément changer le nom de la variable pour F .

Quelques primitives de fonctions usuelles :

$$\int \sin(x)dx = -\cos(x),$$

$$\int \cos(x)dx = \sin(x),$$

$$\int x^n dx = \frac{x^{n+1}}{n+1} \text{ pour } n \neq -1,$$

$$\int \frac{1}{x} dx = \ln|x|,$$

$$\int e^x dx = e^x.$$

Exemple. (*Pharmacocinétique*). L'exposition du corps à un médicament est mesurée par l'intégrale de la concentration $c(t)$ de ce médicament en fonction du temps, à partir du temps d'introduction du médicament $t = 0$ jusqu'à $t \rightarrow +\infty$

$$\int_0^{+\infty} c(t)dt.$$

Par exemple, si ce médicament suit une cinétique d'élimination d'ordre 1, la concentration est $c(t) = c_0 e^{-kt}$ où c_0 est la concentration initiale et k est la constante de vitesse. Cette fonction est intégrable sur $[0, +\infty[$ et l'intégrale vaut

$$\int_0^{+\infty} c(t)dt = \int_0^{+\infty} c_0 e^{-kt} dt = \left[c_0 \frac{e^{-kt}}{-k} \right]_0^{+\infty} = \frac{c_0}{k}.$$

Exemple. (*Intégration d'un pic en spectroscopie*). En spectroscopie, l'importance d'un pic sur un spectre est souvent quantifié en calculant l'aire sous ce pic. Par exemple, le spectre UV-Visible de la molécule d'éthylène présente un pic entre les longueurs d'onde $\lambda = 140$ et $\lambda = 250$ nm que l'on peut modéliser par la fonction $A : \lambda \mapsto e^{-a(\lambda-180)^2}$ où $a > 0$ est une constante. L'aire sous le pic est l'intégrale :

$$\int_{140}^{250} A(\lambda) d\lambda = \int_{140}^{250} e^{-a(\lambda-180)^2} d\lambda.$$

Mais on ne connaît pas de **primitive** de la fonction gaussienne $\lambda \mapsto e^{-a(\lambda-180)^2}$. On verra plus loin que l'on peut néanmoins calculer l'**intégrale** d'une fonction gaussienne. Il ne faut donc pas confondre la notion de primitive et celle d'intégrale.

1.4 Autres théorèmes

Les deux théorèmes suivants sont très utiles pour calculer des intégrales en pratique.

Théorème (Intégration par parties). Soit $u, v : D \rightarrow \mathbb{R}$ deux fonctions de classe C^1 intégrables sur l'intervalle $]a, b[$. On a :

$$\int_a^b u(x)v'(x)dx = \left[u(x)v(x) \right]_a^b - \int_a^b u'(x)v(x)dx.$$

Théorème (Intégration par changement de variable). Soit $f : D \rightarrow \mathbb{R}$ une fonction continue et intégrable sur l'intervalle $]a, b[$, et $\varphi : I \rightarrow]a, b[$ une fonction de classe C^1 sur un intervalle I . On a :

$$\int_a^b f(x)dx = \int_\alpha^\beta f(\varphi(t))\varphi'(t)dt,$$

avec α et β tels que $a = \varphi(\alpha)$ et $b = \varphi(\beta)$.

On dit que l'on a effectué le changement de variable $x = \varphi(t)$ et que $dx = \varphi'(t)dt$.

Exemple. (*Une intégrale utile en chimie quantique*). En chimie quantique, on a souvent besoin de l'intégrale :

$$I_n = \int_0^{+\infty} x^n e^{-\alpha x} dx,$$

où $\alpha > 0$ et n est un entier positif ou nul. Pour $n = 0$, on a :

$$I_0 = \int_0^{+\infty} e^{-\alpha x} dx = \left[\frac{e^{-\alpha x}}{-\alpha} \right]_0^{+\infty} = \frac{1}{\alpha}.$$

Pour $n \geq 1$, on peut exprimer I_n en fonction de I_{n-1} à l'aide d'une intégration par parties. En effet, en posant $u(x) = x^n$ et $v'(x) = e^{-\alpha x}$, on obtient :

$$\begin{aligned} I_n &= \left[x^n \frac{e^{-\alpha x}}{-\alpha} \right]_0^{+\infty} - \int_0^{+\infty} n x^{n-1} \frac{e^{-\alpha x}}{-\alpha} dx \\ &= \frac{n}{\alpha} \int_0^{+\infty} x^{n-1} e^{-\alpha x} dx = \frac{n}{\alpha} I_{n-1}. \end{aligned}$$

On a donc :

$$I_n = \frac{n}{\alpha} I_{n-1} = \frac{n}{\alpha} \frac{n-1}{\alpha} I_{n-2} = \frac{n}{\alpha} \frac{n-1}{\alpha} \dots \frac{1}{\alpha} I_0 = \frac{n!}{\alpha^{n+1}}.$$

2 Intégrale d'une fonction de plusieurs variables réelles

2.1 Concept général et notation

On peut généraliser le concept d'intégrale à une fonction de plusieurs variables réelles.

Soit une fonction f de \mathbb{R}^n dans \mathbb{R} qui aux variables $(x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ associe le nombre $f(x_1, x_2, \dots, x_n) \in \mathbb{R}$. Si la fonction f est intégrable sur un domaine $D \subseteq \mathbb{R}^n$ alors on peut définir de façon similaire à la section précédente (par limite d'une somme de Riemann) l'intégrale de f sur D qui est notée

$$\int \cdots \int_D f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n.$$

Une telle intégrale avec plusieurs variables s'appelle "intégrale multiple", par opposition à l'intégrale avec une seule variable appelée "intégrale simple". L'intégrale multiple a les mêmes propriétés essentielles que l'intégrale simple : additivité (si on décompose le domaine D en deux sous-domaines disjoints, c'est-à-dire $D = D_1 \cup D_2$ avec $D_1 \cap D_2 = \emptyset$), linéarité, croissance.

En particulier, on sera intéressé par les intégrales de fonctions de l'espace à deux dimensions \mathbb{R}^2 et à trois dimensions \mathbb{R}^3 .

2.2 Intégrale double

Pour une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ qui à (x, y) associe $f(x, y)$, si la fonction est intégrable sur un domaine $D \subseteq \mathbb{R}^2$ alors on peut définir l'intégrale de f sur D qui est notée

$$\iint_D f(x, y) dx dy \quad \text{ou} \quad \iint_D f(x, y) dS.$$

On dit que c'est une "intégrale double". La deuxième notation avec dS renvoie au fait que $dx dy$ peut être interprété comme l'aire d'une petite surface, qui est souvent appelée "élément de surface". L'intégrale double peut donc être vue comme la somme sur x et y des produits de $f(x, y)$ par l'élément de surface dS . Ceci est illustré dans la Figure II.2.2.

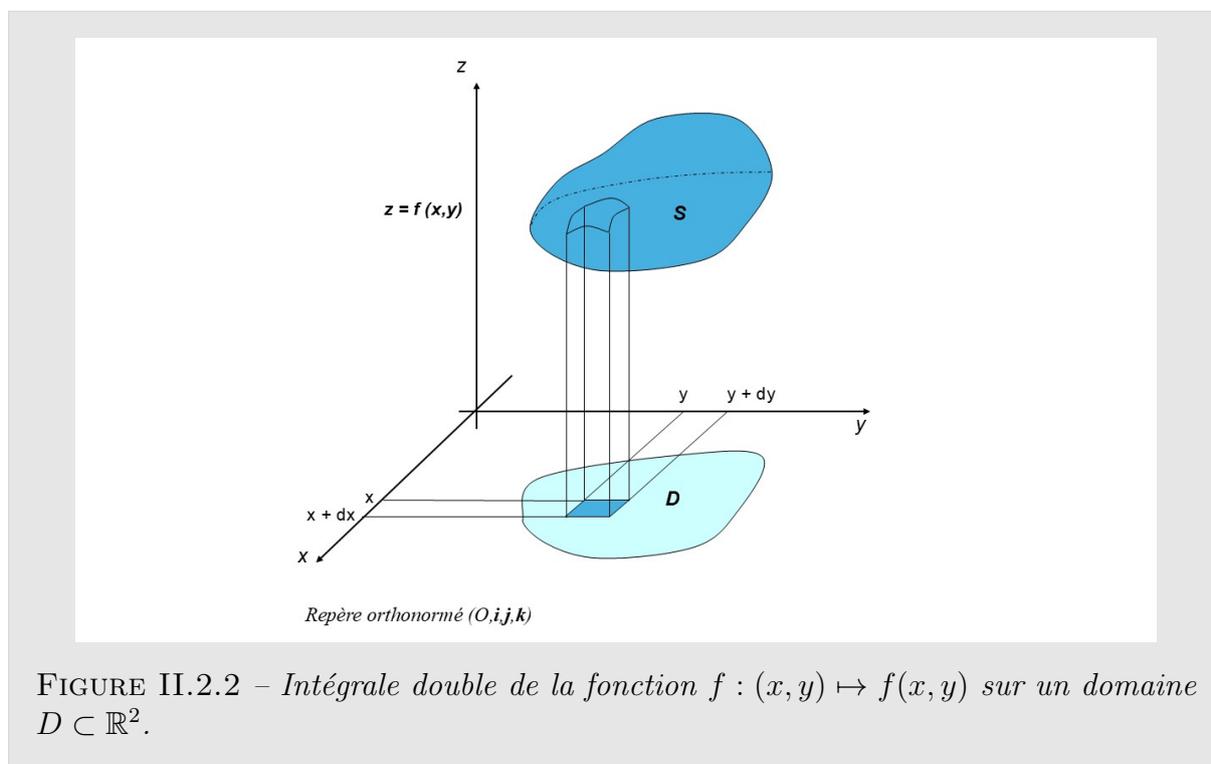


FIGURE II.2.2 – Intégrale double de la fonction $f : (x, y) \mapsto f(x, y)$ sur un domaine $D \subset \mathbb{R}^2$.

Dans le cas particulier où la fonction que l'on intègre vaut 1 partout sur le domaine D , c'est-à-dire $f(x, y) = 1$ pour tout $x, y \in D$, alors l'intégrale de f sur D donne l'aire S de la surface D :

$$\iint_D 1 dx dy = \iint_D 1 dS = S.$$

Si on intègre sur un domaine de forme rectangulaire (les variables varient indépendamment), c'est-à-dire

$$D = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b \text{ et } c \leq y \leq d\},$$

alors on peut noter l'intégrale de façon plus explicite :

$$\iint_D f(x, y) dx dy = \int_{x=a}^b \int_{y=c}^d f(x, y) dx dy.$$

D'après le théorème de Fubini, on peut considérer cette intégrale double comme deux intégrales simples consécutives, l'ordre d'intégration n'ayant pas d'importance.

Théorème (Théorème de Fubini pour une intégrale double). Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction intégrable sur le domaine $D = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b \text{ et } c \leq y \leq d\}$. Si $|f|$ est aussi intégrable sur D , alors

$$\int_{x=a}^b \int_{y=c}^d f(x, y) dx dy = \int_{x=a}^b \left[\int_{y=c}^d f(x, y) dy \right] dx = \int_{y=c}^d \left[\int_{x=a}^b f(x, y) dx \right] dy.$$

De plus, dans le cas particulier où la fonction f se décompose en produit d'une fonction qui ne dépend que de x et d'une fonction qui ne dépend que de y , c'est-à-dire $f(x, y) = g(x)h(y)$, l'intégrale double se décompose en produit d'une intégrale sur x et d'une intégrale sur y .

$$\int_{x=a}^b \int_{y=c}^d f(x, y) dx dy = \int_{x=a}^b \int_{y=c}^d g(x)h(y) dx dy = \left[\int_{x=a}^b g(x) dx \right] \times \left[\int_{y=c}^d h(y) dy \right].$$

Intégrales doubles sur des domaines non rectangles (optionnel)

Théorème (Théorème de Fubini pour une intégrale double sur un domaine non rectangle). Soit $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ une fonction intégrable sur le domaine $D = \{(x, y) \in \mathbb{R}^2 \mid a \leq x \leq b \text{ et } \phi(x) \leq y \leq \psi(x)\}$ avec ϕ et ψ deux fonctions continues sur $[a, b]$ telles de $\forall x \in [a, b], \phi(x) \leq \psi(x)$. Si $|f|$ est aussi intégrable sur D , alors

$$\int_{x=a}^b \int_{y=\phi(x)}^{\psi(x)} f(x, y) dx dy = \int_{x=a}^b \left[\int_{y=\phi(x)}^{\psi(x)} f(x, y) dy \right] dx.$$

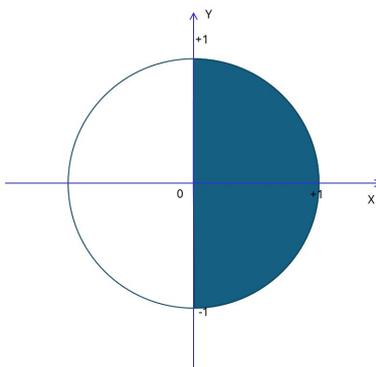
Bien que les variables ne varient pas indépendamment, on peut considérer cette intégrale double comme deux intégrales simples consécutives. On peut échanger les rôles de x et y en redéfinissant le domaine d'intégration tel que $D = \{(x, y) \in \mathbb{R}^2 \mid c \leq y \leq d \text{ et } \mu(y) \leq x \leq \nu(y)\}$ avec μ et ν deux fonctions continues sur $[c, d]$ telles de $\forall y \in [c, d], \mu(y) \leq \nu(y)$. Dans ce cas :

$$\int_{y=c}^d \int_{x=\mu(y)}^{\nu(y)} f(x, y) dx dy = \int_{y=c}^d \left[\int_{x=\mu(y)}^{\nu(y)} f(x, y) dx \right] dy.$$

Exemple. (Une intégrale sur un demi-disque). On considère l'intégrale suivante

$$I = \iint_D xy^2 dx dy,$$

sur le demi-disque $D = \{y \in \mathbb{R}, x \in \mathbb{R}^+, x^2 + y^2 \leq 1\}$.



On peut ré-exprimer le domaine comme

$$D = \left\{ (x, y) \in \mathbb{R}^2 \mid -1 \leq y \leq 1 \text{ et } 0 \leq x \leq \sqrt{1 - y^2} \right\},$$

ce qui conduit à

$$\begin{aligned} I &= \int_{y=-1}^{+1} \left[\int_{x=0}^{\sqrt{1-y^2}} xy^2 dx \right] dy \\ &= \frac{1}{2} \int_{y=-1}^{+1} \left([x^2 y^2]_{x=0}^{\sqrt{1-y^2}} \right) dy \\ &= \frac{1}{2} \int_{y=-1}^{+1} (y^2 - y^4) dy \\ &= \frac{2}{15}. \end{aligned}$$

2.3 Intégrale triple

De même, pour une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ qui à (x, y, z) associe $f(x, y, z)$, si la fonction est intégrable sur un domaine $D \subseteq \mathbb{R}^3$ alors on peut définir l'intégrale de f sur D qui est notée

$$\iiint_D f(x, y, z) dx dy dz \quad \text{ou} \quad \iiint_D f(x, y, z) dV.$$

On dit que c'est une "intégrale triple". La deuxième notation avec dV renvoie au fait que $dx dy dz$ peut être interprété comme un petit volume, qui est souvent appelé "élément de volume".

Dans le cas particulier où la fonction que l'on intègre vaut 1 partout sur le domaine D , c'est-à-dire $f(x, y, z) = 1$ pour tout $x, y, z \in D$, alors l'intégrale de f sur D donne le volume V du domaine D :

$$\iiint_D 1 dx dy dz = \iiint_D 1 dV = V.$$

Si on intègre sur un domaine de forme parallélépipède rectangle (les variables varient indépendamment), c'est-à-dire

$$D = \{(x, y, z) \in \mathbb{R}^3 \mid a \leq x \leq b \text{ et } c \leq y \leq d \text{ et } p \leq z \leq q\},$$

alors on peut noter l'intégrale de façon plus explicite :

$$\iiint_D f(x, y, z) dx dy dz = \int_{x=a}^b \int_{y=c}^d \int_{z=p}^q f(x, y, z) dx dy dz.$$

Ici encore, on peut considérer cette intégrale triple comme trois intégrales simples consécutives. De plus, dans le cas particulier où la fonction f se décompose en produit d'une fonction qui ne dépend que de x , d'une fonction qui ne dépend que de y et d'une fonction qui ne dépend que de z , c'est-à-dire $f(x, y, z) = g(x)h(y)i(z)$, l'intégrale triple se décompose en produit d'une intégrale sur x , d'une intégrale sur y et d'une intégrale sur z (à la condition aussi que les variables x , y et z varient de manière indépendante, ce qui est le cas ici puisque $a \leq x \leq b$ et $c \leq y \leq d$ et $p \leq z \leq q$).

$$\begin{aligned} \int_{x=a}^b \int_{y=c}^d \int_{z=p}^q f(x, y, z) dx dy dz &= \int_{x=a}^b \int_{y=c}^d \int_{z=p}^q g(x)h(y)i(z) dx dy dz \\ &= \left[\int_{x=a}^b g(x) dx \right] \times \left[\int_{y=c}^d h(y) dy \right] \times \left[\int_{z=p}^q i(z) dz \right]. \end{aligned}$$

De la même manière que nous avons étendu le théorème de Fubini à des domaines non rectangles, il est possible de l'étendre aussi à des volumes bien plus complexes que des parallélépipèdes rectangles.

Exemple. (*Intégrales d'orbitales en chimie quantique*). Une orbitale atomique ou moléculaire (à valeurs réelles) est une fonction $\psi : \mathbb{R}^3 \rightarrow \mathbb{R}$ qui est normalisée, ce qui veut dire que l'intégrale de ψ^2 sur tout l'espace est égale à 1 :

$$\iiint_{\mathbb{R}^3} \psi(x, y, z)^2 dx dy dz = \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} \psi(x, y, z)^2 dx dy dz = 1.$$

Par ailleurs, une quantité clé pour l'étude de la liaison chimique est l'intégrale du produit de deux orbitales atomiques ψ_1 et ψ_2 , appelée "intégrale de recouvrement" :

$$\int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} \psi_1(x, y, z) \psi_2(x, y, z) dx dy dz.$$

Exemple. (*Potentiel électrique créé par une densité de charge*). En électrostatique, une densité de charge $\rho(x, y, z)$ présente sur un domaine $D \subseteq \mathbb{R}^3$ crée un potentiel électrique $\phi(x, y, z)$ donné par

$$\phi(x, y, z) = \iiint_D \frac{\rho(x', y', z')}{4\pi\epsilon_0 \sqrt{(x' - x)^2 + (y' - y)^2 + (z' - z)^2}} dx' dy' dz',$$

où ϵ_0 est la permittivité du vide.

2.4 Intégrale en coordonnées polaires et en coordonnées sphériques

De manière générale, pour calculer une intégrale multiple, on essaie de se ramener à un produit d'intégrales simples. Mais ceci n'est pas possible que si la fonction f à intégrer s'écrit comme un produit de fonctions qui ne dépendent chacune que d'une seule variable et que le domaine D peut se décomposer en sous-domaines indépendants pour chaque variable. Souvent ce n'est pas le cas en termes des coordonnées cartésiennes, mais ça peut être le cas en termes des coordonnées polaires ou des coordonnées sphériques, si le problème présente une symétrie polaire ou sphérique. Dans ce cas, après passage dans le système de coordonnées approprié, l'intégrale multiple se factorise en produit d'intégrales simples et peut devenir calculable.

Intégrale double en coordonnées polaires

Rappelons que les coordonnées cartésiennes x et y s'expriment en fonction des coordonnées polaires r et θ par :

$$\begin{cases} x = r \cos \theta \\ y = r \sin \theta \end{cases},$$

et $r = \sqrt{x^2 + y^2}$.

Attention, si en coordonnées cartésiennes l'élément de surface est $dS = dx dy$, en coordonnées polaires l'élément de surface dS ne peut pas être égal à $dr d\theta$. Un simple raisonnement dimensionnel va nous le prouver. En effet, dS a la dimension d'une surface L^2 ce qui est cohérent avec le fait que dx et dy aient la dimension d'une longueur L . En revanche, $dr d\theta$ n'est pas cohérent avec une dimension de surface en L^2 puisque si r a une dimension d'une longueur L , $d\theta$ est un angle en radians qui est une unité sans dimension. On peut déterminer graphiquement l'expression correcte de dS en coordonnées polaires en regardant quelle est la surface balayée par une petite variation de r et de θ . Comme illustré sur la Figure II.2.3, en partant du point M , une variation de r d'une petite quantité dr crée un segment de longueur dr , et une variation de θ d'une petite quantité $d\theta$ crée un arc de cercle dont la longueur dépend de r et vaut $r d\theta$ (pour un $d\theta$ infinitésimal). L'élément de surface est donc $dS = r dr d\theta$, qui a bien la dimension d'une surface.

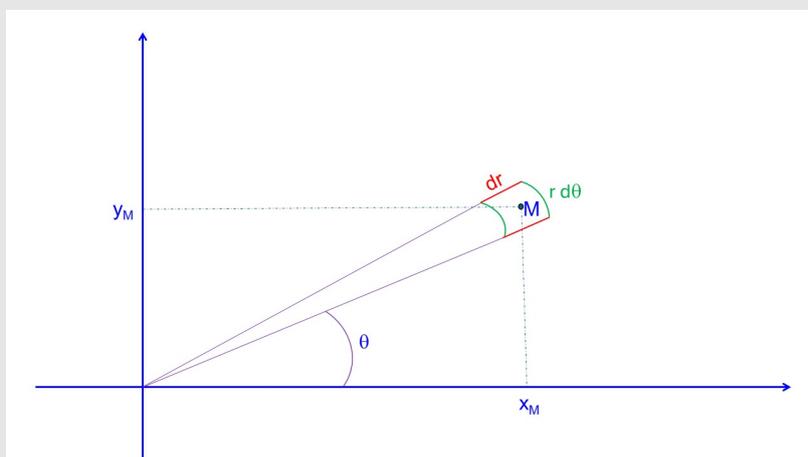


FIGURE II.2.3 – Détermination de l'expression de l'élément de surface dS en coordonnées polaires.

L'intégrale d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ des deux coordonnées cartésiennes x et y sur un domaine $D \subseteq \mathbb{R}^2$ peut donc être ré-exprimée comme une intégrale sur les coordonnées polaires de $\tilde{f}(r, \theta) = f(r \cos \theta, r \sin \theta)$

$$\iint_D f(x, y) dx dy = \iint_{\tilde{D}} \tilde{f}(r, \theta) r dr d\theta,$$

où \tilde{D} est le domaine ré-exprimé en termes des coordonnées polaires. En particulier, si le domaine D est l'ensemble du plan cartésien \mathbb{R}^2 , alors le domaine correspondant en coordonnées polaires est $\tilde{D} = \{(r, \theta) \mid 0 \leq r < +\infty \text{ et } 0 \leq \theta < 2\pi\}$ et on peut donc écrire :

$$\int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} f(x, y) dx dy = \int_{r=0}^{+\infty} \int_{\theta=0}^{2\pi} \tilde{f}(r, \theta) r dr d\theta.$$

Exemple. (*Aire d'un disque*). Si on choisit le domaine D_R suivant décrivant le disque de rayon R de centre $(0, 0)$

$$D_R = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq R^2\},$$

alors l'intégrale de 1 sur D_R est l'aire A_R de ce disque :

$$A_R = \iint_{D_R} 1 dx dy.$$

Cependant, à cause du fait que le domaine D_R ne peut pas se décomposer en deux sous-domaines indépendants pour x et y , on ne peut pas écrire cette intégrale comme le produit d'une intégrale sur x et d'une intégrale sur y . Mais en passant en coordonnées polaires le domaine devient plus simple,

$$\tilde{D}_R = \{(r, \theta) \mid 0 \leq r \leq R \text{ et } 0 \leq \theta < 2\pi\},$$

puisque les variables r et θ varient de manière indépendante. On peut donc écrire

$$\begin{aligned} A_R &= \int_{r=0}^R \int_{\theta=0}^{2\pi} r dr d\theta = \left[\int_{r=0}^R r dr \right] \times \left[\int_{\theta=0}^{2\pi} d\theta \right] \\ &= \left[\frac{r^2}{2} \right]_0^R \times \left[\theta \right]_0^{2\pi} = \pi R^2. \end{aligned}$$

On retrouve bien l'aire d'un disque de rayon R .

Exemple. (*Intégrale gaussienne*). Une intégrale célèbre est l'intégrale gaussienne suivante :

$$I = \int_0^{+\infty} e^{-ax^2} dx,$$

avec $a > 0$. On ne connaît pas une primitive de la fonction gaussienne $x \mapsto e^{-ax^2}$ mais on peut quand même calculer cette intégrale de la façon suivante. On commence par prendre le carré de cette intégrale :

$$\begin{aligned} I^2 &= \left[\int_0^{+\infty} e^{-ax^2} dx \right]^2 = \left[\int_0^{+\infty} e^{-ax^2} dx \right] \times \left[\int_0^{+\infty} e^{-ax^2} dx \right] \\ &= \left[\int_0^{+\infty} e^{-ax^2} dx \right] \times \left[\int_0^{+\infty} e^{-ay^2} dy \right], \end{aligned}$$

où l'on a utilisé la variable y au lieu de x dans la deuxième intégrale, ce qui est autorisé puisque le nom de la variable d'intégration dans une intégrale n'est pas important (on dit que c'est une variable muette). D'après le théorème de Fubini, on peut écrire cette intégrale comme une intégrale double :

$$I^2 = \int_{x=0}^{+\infty} \int_{y=0}^{+\infty} e^{-ax^2} e^{-ay^2} dx dy = \int_{x=0}^{+\infty} \int_{y=0}^{+\infty} e^{-a(x^2+y^2)} dx dy.$$

On peut alors calculer cette intégrale en passant en coordonnées polaires avec un domaine défini par $0 \leq r < +\infty$ et $0 \leq \theta \leq \pi/2$:

$$\begin{aligned} I^2 &= \int_{r=0}^{+\infty} \int_{\theta=0}^{\pi/2} e^{-ar^2} r dr d\theta \\ &= \left[\int_0^{+\infty} e^{-ar^2} r dr \right] \times \left[\int_0^{\pi/2} d\theta \right] \\ &= \left[\frac{e^{-ar^2}}{-2a} \right]_0^{+\infty} \times [\theta]_0^{\pi/2} = \frac{\pi}{4a}. \end{aligned}$$

L'intégrale gaussienne est donc :

$$I = \int_0^{+\infty} e^{-ax^2} dx = \frac{1}{2} \sqrt{\frac{\pi}{a}}.$$

On a aussi souvent besoin de l'intégrale de e^{-ax^2} de $-\infty$ à $+\infty$. Puisque la fonction $x \mapsto e^{-ax^2}$ est une fonction paire, son intégrale de $-\infty$ à $+\infty$ vaut le double de l'intégrale de 0 à $+\infty$:

$$\int_{-\infty}^{+\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}.$$

Intégrale triple en coordonnées sphériques

Rappelons que les coordonnées cartésiennes x , y et z s'expriment en fonction des coordonnées sphériques r , θ et ϕ par :

$$\begin{cases} x = r \sin \theta \cos \phi \\ y = r \sin \theta \sin \phi \\ z = r \cos \theta \end{cases},$$

et $r = \sqrt{x^2 + y^2 + z^2}$.

On peut déterminer graphiquement l'expression de l'élément de volume dV en coordonnées sphériques. Comme illustré sur la Figure II.2.4, en partant du point M , une variation de r d'une petite quantité dr crée un segment de longueur dr . Une variation de θ d'une petite quantité $d\theta$ crée un arc de cercle de longueur $r d\theta$ (pour un $d\theta$ infinitésimal). Une variation de ϕ d'une petite quantité $d\phi$ crée un arc de cercle dont la longueur dépend de la distance de la projection du point M sur le plan (xy) par rapport à l'origine qui vaut $r \sin \theta$. La longueur de l'arc de cercle a donc pour valeur $r \sin \theta d\phi$ (pour un $d\phi$ infinitésimal). L'élément de volume dV est donc $dV = r^2 \sin \theta dr d\theta d\phi$, qui a bien la dimension d'un volume.

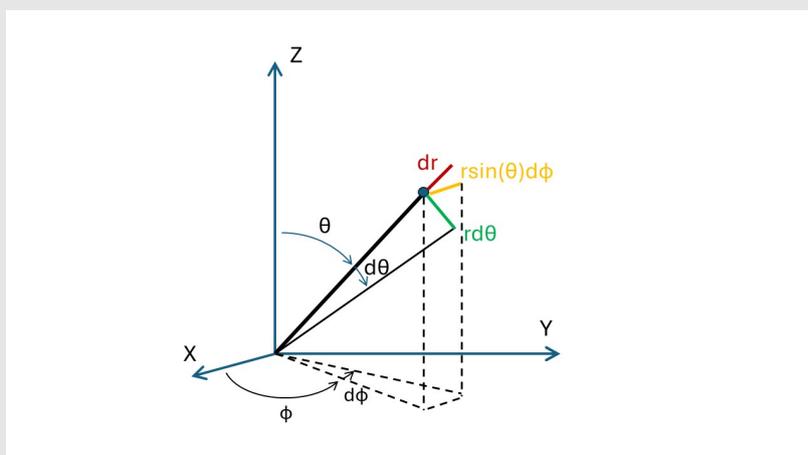


FIGURE II.2.4 – Détermination de l'expression de l'élément de volume dV en coordonnées sphériques.

L'intégrale d'une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ des trois coordonnées cartésiennes x , y et z sur un domaine $D \subseteq \mathbb{R}^3$ peut donc être ré-exprimée comme une intégrale sur les coordonnées sphériques de $\tilde{f}(r, \theta, \phi) = f(r \sin \theta \cos \phi, r \sin \theta \sin \phi, r \cos \theta)$

$$\iiint_D f(x, y, z) dx dy dz = \iiint_{\tilde{D}} \tilde{f}(r, \theta, \phi) r^2 \sin \theta dr d\theta d\phi,$$

où \tilde{D} est le domaine ré-exprimé en termes des coordonnées sphériques. En particulier, si le domaine D est l'ensemble de l'espace cartésien à trois dimensions \mathbb{R}^3 , alors

le domaine correspondant en coordonnées sphériques est $\tilde{D} = \{(r, \theta, \phi) \mid 0 \leq r < +\infty \text{ et } 0 \leq \theta \leq \pi \text{ et } 0 \leq \phi < 2\pi\}$ et on peut donc écrire :

$$\int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} f(x, y, z) dx dy dz = \int_{r=0}^{+\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} \tilde{f}(r, \theta, \phi) r^2 \sin \theta dr d\theta d\phi.$$

Exemple. (*Volume d'une boule*). Si on choisit le domaine B_R suivant décrivant la boule de rayon R de centre $(0, 0, 0)$

$$B_R = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 \leq R^2\},$$

alors l'intégrale de 1 sur B_R est le volume V_R de cette boule :

$$V_R = \iiint_{B_R} 1 dx dy dz.$$

Cependant, à cause du fait que le domaine B_R ne peut pas se décomposer en trois sous-domaines indépendants pour x , y et z , on ne peut pas écrire cette intégrale comme le produit d'une intégrale sur x , d'une intégrale sur y et d'une intégrale sur z . Mais en passant en coordonnées sphérique le domaine devient plus simple,

$$\tilde{B}_R = \{(r, \theta, \phi) \mid 0 \leq r \leq R \text{ et } 0 \leq \theta \leq \pi \text{ et } 0 \leq \phi < 2\pi\},$$

puisque les variables r , θ et ϕ varient de manière indépendante. On peut donc écrire

$$\begin{aligned} V_R &= \int_{r=0}^R \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} r^2 \sin \theta dr d\theta d\phi = \left[\int_0^R r^2 dr \right] \times \left[\int_0^{\pi} \sin \theta d\theta \right] \times \left[\int_0^{2\pi} d\phi \right] \\ &= \left[\frac{r^3}{3} \right]_0^R \times \left[-\cos \theta \right]_0^{\pi} \times \left[\phi \right]_0^{2\pi} = \frac{4\pi R^3}{3}. \end{aligned}$$

On retrouve bien le volume d'une boule de rayon R .

Exemple. (Normalisation de l'orbitale 1s). L'orbitale 1s de l'atome hydrogène a pour expression :

$$\psi_{1s}(x, y, z) = \frac{1}{\sqrt{\pi}a_0^{3/2}} e^{-\sqrt{x^2+y^2+z^2}/a_0},$$

où a_0 est le rayon de Bohr. On peut vérifier que cette orbitale est normalisée. Pour cela, on calcule l'intégrale de $\psi_{1s}(x, y, z)^2$ sur tout l'espace en utilisant les coordonnées sphériques

$$\begin{aligned} & \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} \psi_{1s}(x, y, z)^2 dx dy dz \\ &= \frac{1}{\pi a_0^3} \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} e^{-2\sqrt{x^2+y^2+z^2}/a_0} dx dy dz \\ &= \frac{1}{\pi a_0^3} \int_{r=0}^{+\infty} \int_{\theta=0}^{\pi} \int_{\phi=0}^{2\pi} e^{-2r/a_0} r^2 \sin \theta dr d\theta d\phi \\ &= \frac{1}{\pi a_0^3} \left[\int_0^{+\infty} e^{-2r/a_0} r^2 dr \right] \times \left[\int_0^{\pi} \sin \theta d\theta \right] \times \left[\int_0^{2\pi} d\phi \right]. \end{aligned}$$

L'intégrale sur r se calcule en utilisant $\int_0^{+\infty} x^n e^{-\alpha x} dx = n!/\alpha^{n+1}$ pour $n = 2$ et $\alpha = 2/a_0$. On a donc :

$$\int_0^{+\infty} e^{-2r/a_0} r^2 dr = \frac{a_0^3}{4}.$$

Par ailleurs, les intégrales sur θ et ϕ donne

$$\int_0^{\pi} \sin \theta d\theta = [-\cos \theta]_0^{\pi} = 2 \quad \text{et} \quad \int_0^{2\pi} d\phi = [\phi]_0^{2\pi} = 2\pi.$$

L'orbitale 1s est donc bien normalisée

$$\int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} \int_{z=-\infty}^{+\infty} \psi_{1s}(x, y, z)^2 dx dy dz = 1.$$

3 Intégrale numérique

Il n'est pas toujours possible de calculer analytiquement une intégrale par les méthodes classiques. Parfois même, la fonction à intégrer n'est connue que par un échantillon de valeurs issues de mesures. On doit alors se contenter d'une approximation numérique de la valeur de l'intégrale.

Considérons une fonction $f : [a, b] \rightarrow \mathbb{R}$ continue. On veut calculer une approximation numérique de l'intégrale de f sur $[a, b]$. On commence par découper l'intervalle d'in-

tégration en n sous-intervalles contigus : $[x_0, x_1], [x_1, x_2], \dots, [x_{n-1}, x_n]$ avec $x_0 = a$ et $x_n = b$. Par la relation de Chasles, on peut écrire :

$$\int_a^b f(x)dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx.$$

Sur chaque intervalle $[x_i, x_{i+1}]$, on remplace alors f par une approximation ϕ_i qui est une fonction simple que l'on sait intégrer. En sommant ces intégrales, on obtient alors une approximation numérique de l'intégrale de f :

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} \phi_i(x)dx.$$

Suivant l'approximation ϕ_i choisie, on a différentes méthodes d'intégration numérique. Pour faire simple, nous allons les décrire dans le cas d'un découpage de l'intervalle $[a, b]$ en n intervalles de même longueur $\Delta x = (b - a)/n$. Les différents points sont alors donnés par $x_i = x_0 + i\Delta x$.

3.1 Méthodes des rectangles

La méthode la plus simple consiste à approcher la fonction f sur l'intervalle $[x_i, x_{i+1}]$ par une fonction constante. Par exemple, la valeur de cette constante peut être la valeur de f en x_i ,

$$\phi_i(x) = f(x_i) \text{ pour } x \in [x_i, x_{i+1}],$$

et l'intégrale de f sur $[a, b]$ est donc approchée par la formule suivante :

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} f(x_i)\Delta x.$$

C'est la méthode des rectangles à gauche illustrée sur la Figure II.2.5.

Nous avons choisi arbitrairement de construire des rectangles en prenant la valeur de la fonction en l'extrémité gauche de chaque intervalle. On aurait pu choisir également la valeur de la fonction en l'extrémité droite de chaque intervalle, c'est-à-dire $\phi_i(x) = f(x_{i+1})$ pour $x \in [x_i, x_{i+1}]$, ce qui correspond à la méthode des rectangles à droite (Figure II.2.5). Si cela est possible, le meilleur choix est en fait de prendre la fonction au point milieu de l'intervalle, c'est-à-dire $\phi_i(x) = f((x_i + x_{i+1})/2)$ pour $x \in [x_i, x_{i+1}]$, ce qui correspond à la méthode des rectangles avec point milieu (Figure II.2.5).

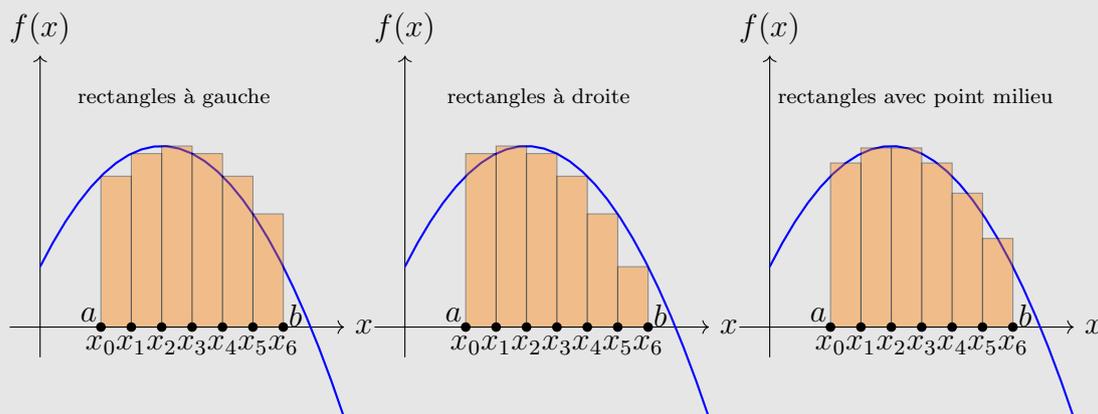


FIGURE II.2.5 – Calcul numérique de l'intégrale de f sur $[a, b]$ avec la méthode des rectangles (à gauche, à droite, et avec point milieu) dans le cas de $n = 6$ sous-intervalles.

Ces protocoles de calcul correspondent à la définition de l'intégrale au sens de Riemann. Plus le découpage de l'intervalle $[a, b]$ en sous-intervalles $[x_i, x_{i+1}]$ est fin, meilleure est l'approximation numérique de l'intégrale de f . Dans la limite d'un découpage infiniment fin, c'est-à-dire $n \rightarrow \infty$, on converge vers la valeur exacte de l'intégrale de f sur $[a, b]$.

Dans un souci d'efficacité, on souhaite obtenir une convergence vers la valeur exacte de l'intégrale de f la plus rapide possible, c'est-à-dire obtenir une valeur acceptable de l'estimation de l'intégrale pour une valeur de n la plus petite possible. Si la fonction f est de classe C^1 , on peut montrer que l'erreur de l'approximation de l'intégrale avec la méthode des rectangles décroît en $1/n$ quand n augmente. Il s'agit d'une convergence assez lente, que l'on peut accélérer avec la méthode des trapèzes ou la méthode de Simpson.

3.2 Méthodes des trapèzes

Pour avoir une convergence plus rapide de l'approximation numérique de l'intégrale de f , il faut une meilleure estimation ϕ_i de f sur chaque intervalle $[x_i, x_{i+1}]$. La méthode des trapèzes consiste à prendre pour ϕ_i l'interpolation linéaire de f entre x_i et x_{i+1} , c'est-à-dire ϕ_i est la droite passant par les points $(x_i, f(x_i))$ et $(x_{i+1}, f(x_{i+1}))$. On forme ainsi un trapèze d'aire $\Delta x(f(x_i) + f(x_{i+1}))/2$. En sommant les aires des trapèzes de chaque intervalle, on obtient l'approximation suivante pour l'intégrale de f sur $[a, b]$:

$$\int_a^b f(x)dx \approx \sum_{i=0}^{n-1} \frac{f(x_i) + f(x_{i+1})}{2} \Delta x.$$

Cette méthode des trapèzes est illustrée dans la Figure II.2.6.

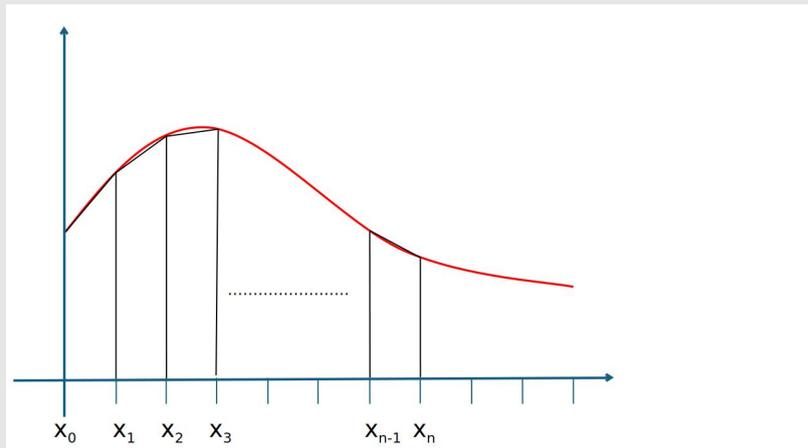


FIGURE II.2.6 – Calcul numérique de l'intégrale de f sur $[a, b]$ avec la méthode des trapèzes.

Comme pour la méthode des rectangles, l'approximation de l'intégrale avec la méthode des trapèzes converge vers la valeur exacte de l'intégrale dans la limite $n \rightarrow \infty$. Cependant, si la fonction f est de classe C^2 , l'erreur décroît en $1/n^2$ avec la méthode des trapèzes, ce qui est bien plus rapide que dans le cas de la méthode des rectangles.

3.3 Méthode de Simpson (optionnel)

Pour améliorer encore l'estimation numérique de l'intégrale de f , on peut s'orienter vers une interpolation polynomiale de degré 2 de f sur chaque intervalle $[x_i, x_{i+1}]$, c'est-à-dire que ϕ_i est une parabole de la forme $\phi_i(x) = p_i x^2 + q_i x + r_i$. Pour déterminer pour chaque intervalle $[x_i, x_{i+1}]$, les valeurs de p_i , q_i et r_i , il suffit de choisir trois valeurs dans cet intervalle prises par la fonction ϕ_i , par exemple les valeurs prises par f en x_i , x_{i+1} et au point milieu $(x_i + x_{i+1})/2$. On obtient alors l'approximation suivante pour l'intégrale de f sur $[a, b]$:

$$\int_a^b f(x) dx \approx \frac{1}{6} \sum_{i=0}^{n-1} \left(f(x_i) + 4f\left(\frac{x_i + x_{i+1}}{2}\right) + f(x_{i+1}) \right) \Delta x.$$

C'est la méthode de Simpson qui est illustrée dans la Figure II.2.7.

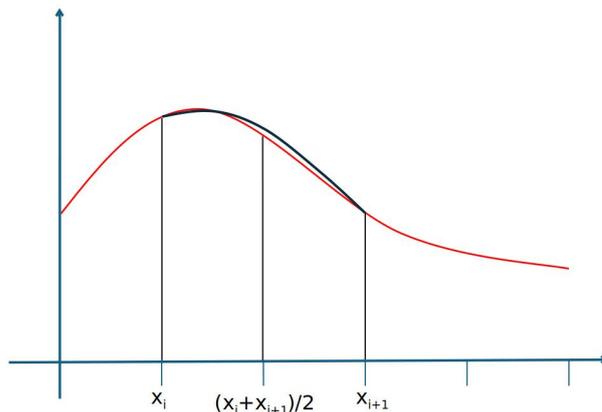


FIGURE II.2.7 – Calcul numérique de l'intégrale de f sur $[a, b]$ avec la méthode de Simpson.

Pour une fonction f de classe C^4 , l'erreur décroît en $1/n^4$ avec la méthode de Simpson, ce qui est une convergence beaucoup plus rapide que celles obtenues avec les méthodes précédentes.

Les méthodes d'intégration numérique que nous avons vu ne sont pas exhaustives mais sont les plus simples à mettre en oeuvre. Des variantes de ces méthodes ou d'autres méthodes d'intégration numérique permettent d'obtenir de meilleurs résultats, comme par exemple la méthode de Romberg ou les méthodes de quadrature de Gauss.

4 Transformation de Fourier

4.1 Définition

Supposons que nous disposions d'un signal temporel représenté par une fonction $f : t \mapsto f(t)$. Cela peut par exemple correspondre au signal mesuré par un spectromètre RMN. Ce signal temporel f est généralement une superposition de différents signaux sinusoïdaux de fréquences différentes (voir Figure II.2.8). La transformation de Fourier (TF) est l'outil mathématique qui permet d'analyser le contenu en fréquences du signal f , c'est-à-dire de calculer le spectre de fréquence correspondant à f . La transformation de Fourier est une extension aux fonctions non périodiques du développement en série de Fourier des fonctions périodiques.

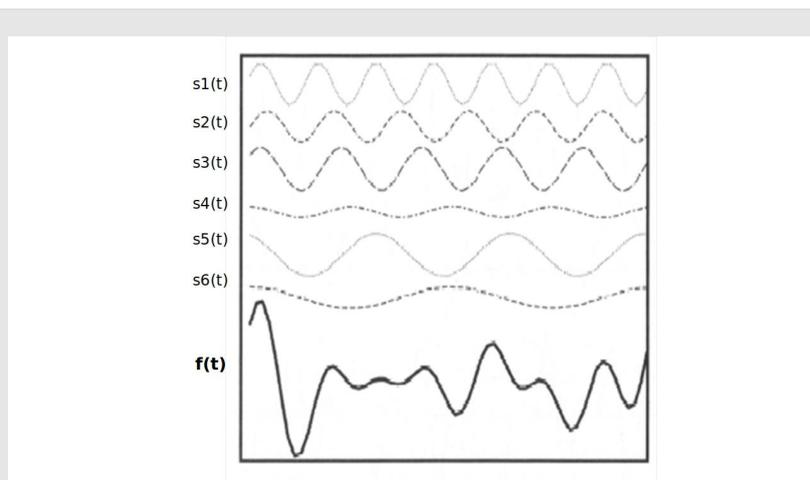


FIGURE II.2.8 – Signal temporel $f : t \mapsto f(t)$ qui est la superposition de différents signaux sinusoïdaux simples $s_1(t), \dots, s_6(t)$ d'amplitudes, de phases et de fréquences différentes.

Nous nous limiterons à la transformation de Fourier pour des fonctions d'une variable réelle et à valeurs réelles ou complexes.

Définition (Transformation de Fourier). Soit $f : t \in \mathbb{R} \mapsto f(t) \in \mathbb{C}$ une fonction intégrable sur $] -\infty, +\infty[$. La transformée de Fourier de f est la fonction $\hat{f} : \omega \in \mathbb{R} \mapsto \hat{f}(\omega) \in \mathbb{C}$ définie par :

$$\hat{f}(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt.$$

L'application qui à f associe la nouvelle fonction \hat{f} est appelée la transformation de Fourier.

D'un point de vue physique, si t représente le temps, alors ω représente la fréquence ou plus précisément la pulsation. On dit alors que f est définie dans le domaine temporel et que \hat{f} est définie dans le domaine fréquentiel. Comme la fonction $t \mapsto e^{-i\omega t}$ est à valeurs complexes, \hat{f} est en général à valeurs complexes aussi et peut se scinder en une partie réelle et une partie imaginaire. On peut étudier le spectre d'amplitude $|\hat{f}(\omega)|$ (module de \hat{f}) et le spectre de phase $\arg(\hat{f}(\omega))$. La quantité $|\hat{f}(\omega)|^2$ s'appelle la densité spectrale d'énergie.

À partir de \hat{f} , on peut revenir à la fonction d'origine f par la transformation de Fourier inverse :

$$f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{+i\omega t} d\omega,$$

et on dit que f est la transformée de Fourier inverse de \hat{f} .

On peut aussi utiliser la transformation de Fourier pour analyser un signal spatial. Dans ce cas, la variable t est remplacée par la position x et la variable ω est remplacée par l'impulsion notée k . Ceci est utilisé par exemple dans la technique d'analyse de la structure des cristaux par diffraction des rayons X.

Nous nous limiterons à la transformation de Fourier pour des fonctions d'une variable réelle et à valeurs réelles ou complexes.

4.2 Exemple simple

Nous allons calculer la transformée de Fourier de la fonction “porte” $f_a : t \mapsto f_a(t)$ représentant une impulsion rectangulaire de hauteur maximale 1, centrée à $t = 0$ et de durée $a > 0$:

$$f_a(t) = \begin{cases} 1 & \text{pour } -a/2 \leq t \leq a/2 \\ 0 & \text{sinon.} \end{cases}$$

La fonction f_a est continue par morceaux et intégrable sur $]-\infty, +\infty[$. Calculons sa transformée de Fourier. Pour $\omega \neq 0$, on a :

$$\begin{aligned} \hat{f}_a(\omega) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f_a(t) e^{-i\omega t} dt = \frac{1}{\sqrt{2\pi}} \int_{-a/2}^{+a/2} e^{-i\omega t} dt \\ &= \frac{1}{\sqrt{2\pi}} \left[\frac{e^{-i\omega t}}{-i\omega} \right]_{t=-a/2}^{t=+a/2} = \frac{1}{\sqrt{2\pi}} \frac{e^{+i\omega a/2} - e^{-i\omega a/2}}{i\omega} \\ &= \frac{1}{\sqrt{2\pi}} \frac{2 \sin\left(\frac{a}{2}\omega\right)}{\omega} = \frac{a}{\sqrt{2\pi}} \operatorname{sinc}\left(\frac{a}{2}\omega\right), \end{aligned}$$

où l'on a introduit la fonction sinus cardinal $x \mapsto \operatorname{sinc}(x) = (\sin x)/x$. Par ailleurs, pour $\omega = 0$, on a :

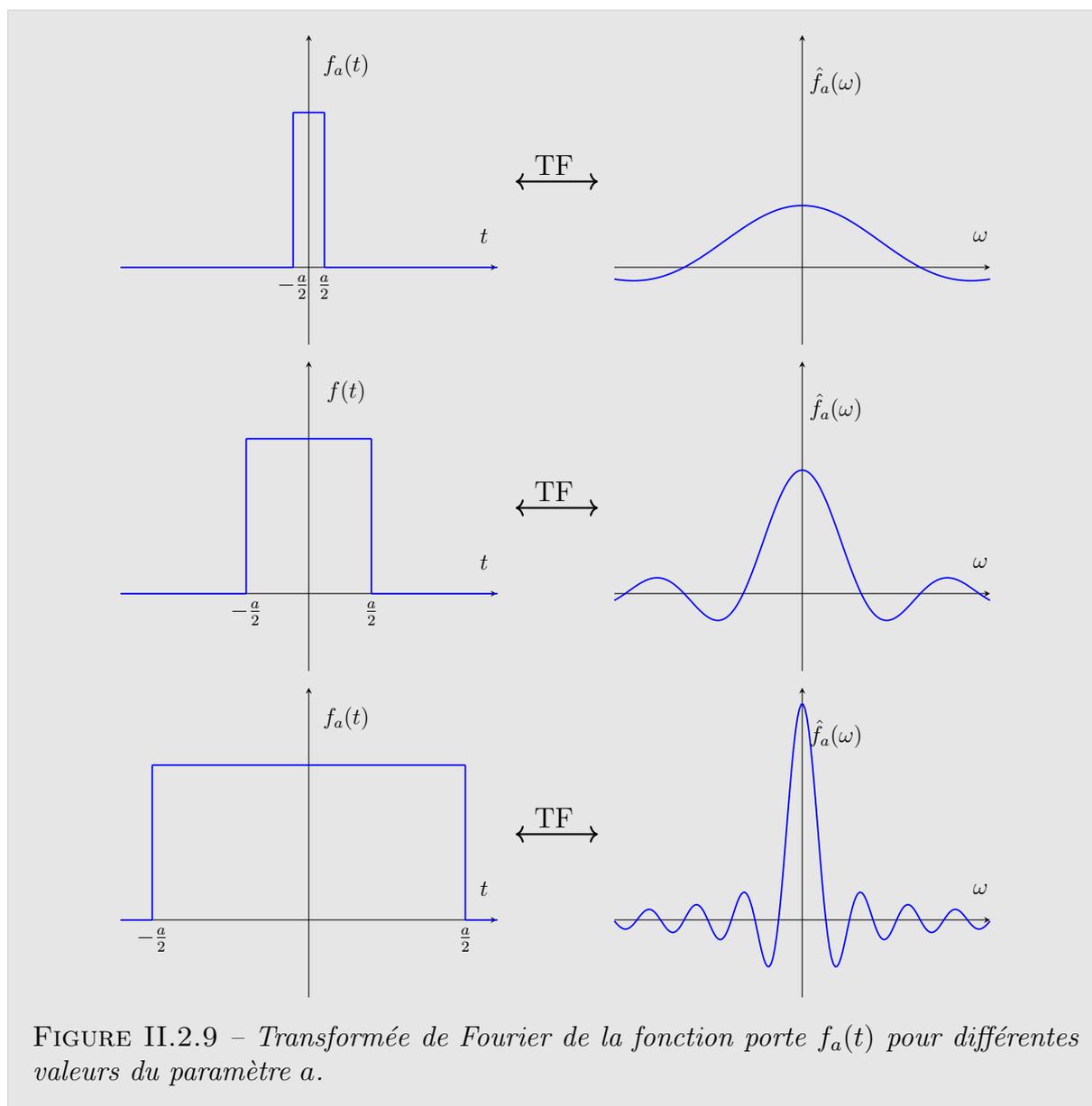
$$\hat{f}_a(0) = \frac{1}{\sqrt{2\pi}} \int_{-a/2}^{+a/2} dt = \frac{a}{\sqrt{2\pi}}.$$

Puisque $\lim_{x \rightarrow 0} \operatorname{sinc}(x) = 1$, on en déduit

$$\lim_{\omega \rightarrow 0} \hat{f}_a(\omega) = \lim_{\omega \rightarrow 0} \frac{a}{\sqrt{2\pi}} \operatorname{sinc}\left(\frac{a}{2}\omega\right) = \frac{a}{\sqrt{2\pi}} = \hat{f}_a(0),$$

donc \hat{f}_a est continue en 0 (et sur \mathbb{R} entier). Par ailleurs, \hat{f}_a est à valeurs réelles et $\lim_{\omega \rightarrow \pm\infty} \hat{f}_a(\omega) = 0$.

La fonction f_a et sa transformée de Fourier \hat{f}_a sont représentées dans la Figure II.2.9 pour différentes valeurs du paramètre a . Cet exemple permet d'illustrer le comportement général de la transformation de Fourier. Quand f_a est étroite (a petit), \hat{f}_a est étendue. Et, inversement, quand f_a est étendue (a grand), \hat{f}_a est étroite.



4.3 Applications (optionnel)

Spectroscopie infrarouge à transformée de Fourier

La spectroscopie infrarouge à transformée de Fourier (IRTF, ou FTIR en anglais) est une technique de mesure pour l'acquisition de spectres infrarouges. Au lieu d'enregistrer la quantité d'énergie absorbée lorsque la fréquence de lumière infrarouge varie (monochromateur), la lumière infrarouge passe au travers d'un interféromètre (de Michelson). Après avoir traversé l'échantillon, le signal mesuré est un interférogramme. Après que le signal ait subi une transformée de Fourier, on obtient un spectre identique à celui obtenu par une spectroscopie infrarouge conventionnelle (dispersive), mais beaucoup plus rapidement.

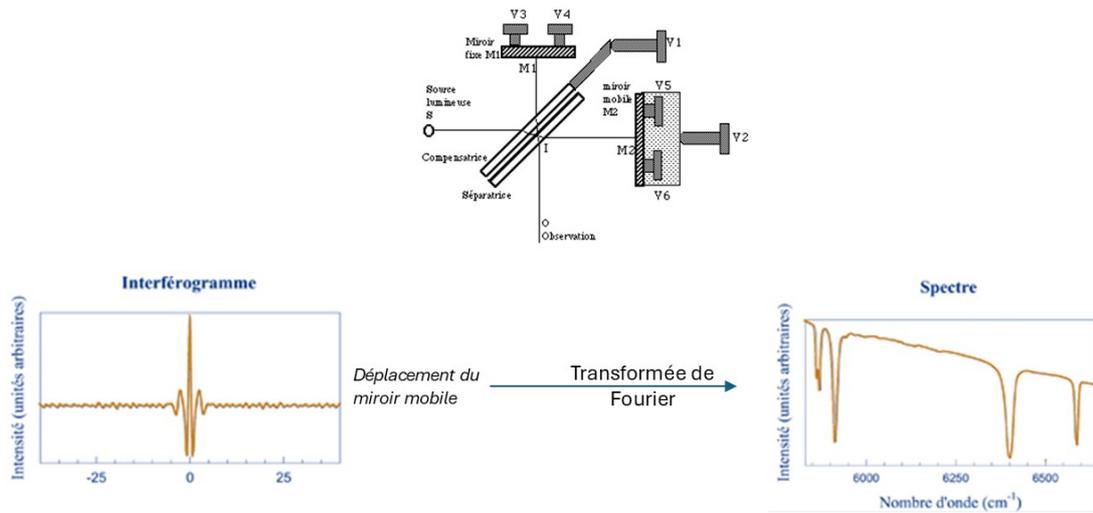


FIGURE II.2.10 – Principe d'un spectromètre infrarouge à transformée de Fourier.

Compression MP3

L'oreille humaine n'est sensible (pour les meilleures) qu'aux fréquences sonores de 20 Hz à 20 kHz. Or, les fréquences d'un signal audio analogique ou numérique de très bonne qualité s'étendent au delà de cette gamme. La compression MP3 du signal sonore va supprimer les fréquences inaudibles. Pour cela, Le convertisseur MP3 va calculer le spectre de fréquence du signal sonore en utilisant une transformation de Fourier, puis effectuer une convolution avec un filtre qui va couper le spectre au niveau des fréquences inaudibles (voir Figure II.2.11).

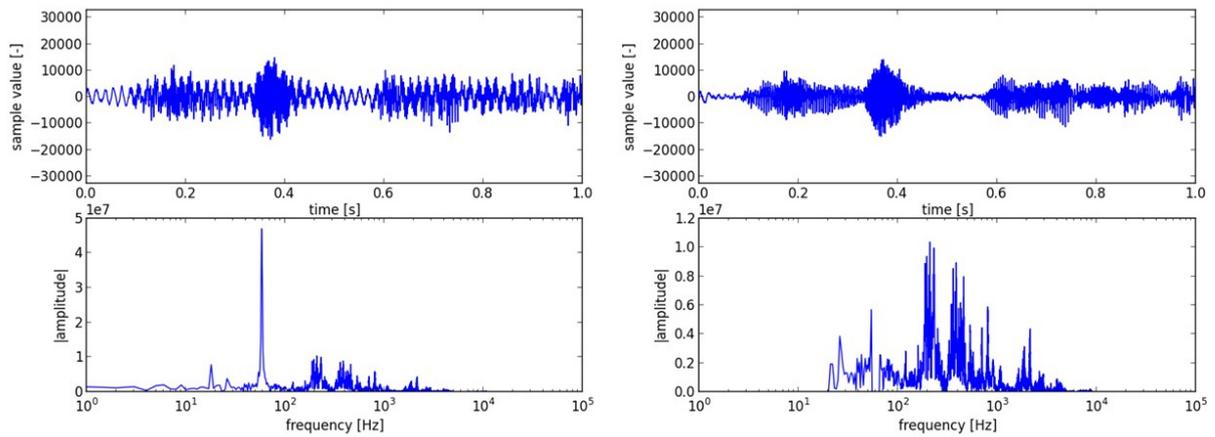


FIGURE II.2.11 – Principe de la compression MP3 : les basses et hautes fréquences du signal sonore sont supprimées.

Pour aller plus loin

Changement de variables général

Dans le cas de deux variables, on peut définir un changement de variables général en exprimant les anciennes variables x et y en fonctions des nouvelles variables u et v par

$$\begin{cases} x = \varphi_1(u, v) \\ y = \varphi_2(u, v), \end{cases},$$

où $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ est une fonction vectorielle de classe C^1 . On définit alors la matrice jacobienne (du nom du mathématicien Charles Jacobi) de φ comme la matrice des dérivées partielles premières des composantes de φ :

$$J_\varphi = \begin{pmatrix} \frac{\partial \varphi_1}{\partial u} & \frac{\partial \varphi_1}{\partial v} \\ \frac{\partial \varphi_2}{\partial u} & \frac{\partial \varphi_2}{\partial v} \end{pmatrix}.$$

L'intégrale d'une fonction $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ sur un domaine $D \subseteq \mathbb{R}^2$ peut alors être écrite avec les nouvelles variables :

$$\iint_D f(x, y) dx dy = \iint_{\tilde{D}} f(\varphi_1(u, v), \varphi_2(u, v)) |\det(J_\varphi)| du dv,$$

où \tilde{D} est le domaine d'intégration exprimé en termes des nouvelles variables u et v , et $\det(J_\varphi)$ est le déterminant de la matrice J_φ , appelé déterminant jacobien. L'élément de surface exprimé avec les nouvelles variables u et v est donc $dS = |\det(J_\varphi)| du dv$.

Par exemple, pour les coordonnées polaires, le déterminant jacobien est :

$$\det(J_\varphi) = \begin{vmatrix} \frac{\partial r \cos \theta}{\partial r} & \frac{\partial r \cos \theta}{\partial \theta} \\ \frac{\partial r \sin \theta}{\partial r} & \frac{\partial r \sin \theta}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{vmatrix} = r((\cos \theta)^2 + (\sin \theta)^2) = r,$$

et on retrouve donc $dS = |\det(J_\varphi)| dr d\theta = r dr d\theta$.

De façon similaire, dans le cas de trois variables, on peut définir un changement de variables général en exprimant les anciennes variables x , y et z en fonctions des nouvelles variables u , v et w par

$$\begin{cases} x = \varphi_1(u, v, w) \\ y = \varphi_2(u, v, w) \\ z = \varphi_3(u, v, w) \end{cases},$$

où $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ est une fonction vectorielle de classe C^1 . La matrice jacobienne de φ est :

$$J_\varphi = \begin{pmatrix} \frac{\partial \varphi_1}{\partial u} & \frac{\partial \varphi_1}{\partial v} & \frac{\partial \varphi_1}{\partial w} \\ \frac{\partial \varphi_2}{\partial u} & \frac{\partial \varphi_2}{\partial v} & \frac{\partial \varphi_2}{\partial w} \\ \frac{\partial \varphi_3}{\partial u} & \frac{\partial \varphi_3}{\partial v} & \frac{\partial \varphi_3}{\partial w} \end{pmatrix}.$$

L'intégrale d'une fonction $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ sur un domaine $D \subseteq \mathbb{R}^3$ peut alors être écrite avec les nouvelles variables :

$$\iiint_D f(x, y, z) dx dy dz = \iiint_{\tilde{D}} f(\varphi_1(u, v, w), \varphi_2(u, v, w), \varphi_3(u, v, w)) |\det(J_\varphi)| du dv dw,$$

où \tilde{D} est le domaine d'intégration exprimé en termes des nouvelles variables u , v et w . L'élément de volume exprimé avec les nouvelles variables u , v et w est donc $dV = |\det(J_\varphi)| du dv dw$.

Par exemple, pour les coordonnées sphériques, le déterminant jacobien est :

$$\begin{aligned} \det(J_\varphi) &= \begin{vmatrix} \frac{\partial r \sin \theta \cos \phi}{\partial r} & \frac{\partial r \sin \theta \cos \phi}{\partial \theta} & \frac{\partial r \sin \theta \cos \phi}{\partial \phi} \\ \frac{\partial r \sin \theta \sin \phi}{\partial r} & \frac{\partial r \sin \theta \sin \phi}{\partial \theta} & \frac{\partial r \sin \theta \sin \phi}{\partial \phi} \\ \frac{\partial r \cos \theta}{\partial r} & \frac{\partial r \cos \theta}{\partial \theta} & \frac{\partial r \cos \theta}{\partial \phi} \end{vmatrix} \\ &= \begin{vmatrix} \sin \theta \cos \phi & r \cos \theta \cos \phi & -r \sin \theta \sin \phi \\ \sin \theta \sin \phi & r \cos \theta \sin \phi & r \sin \theta \cos \phi \\ \cos \theta & -r \sin \theta & 0 \end{vmatrix} \\ &= \cos \theta (r^2 \cos \theta \sin \theta (\cos \phi)^2 + r^2 \cos \theta \sin \theta (\sin \phi)^2) \\ &\quad + r \sin \theta (r (\sin \theta)^2 (\cos \phi)^2 + r (\sin \theta)^2 (\sin \phi)^2) = r^2 \sin \theta \\ &= r^2 \sin \theta ((\cos \theta)^2 + (\sin \theta)^2) \\ &= r^2 \sin \theta, \end{aligned}$$

et on retrouve donc $dV = |\det(J_\varphi)| dr d\theta d\phi = r^2 \sin \theta dr d\theta d\phi$.

Chapitre II.3

Outils d'analyse statistique

La théorie des probabilités et les méthodes d'analyse statistique constituent des outils indispensables dans toutes les sciences. En chimie, ces outils permettent par exemple d'estimer les incertitudes statistiques lors d'une mesure expérimentale ou de concevoir des méthodes de calcul basées sur un échantillonnage statistique. En outre, ils sont au cœur de la description de la matière par la mécanique quantique et la thermodynamique statistique. Dans ce chapitre, après avoir rappelé le concept de variable aléatoire continue, nous nous intéressons à l'estimation de l'espérance et de la variance d'une variable aléatoire continue sur un échantillon.

1 Variable aléatoire continue

1.1 Définition

Définition (Variable aléatoire continue). On définit une variable aléatoire continue X comme une variable pouvant prendre une continuité de valeurs réelles $x \in \mathbb{R}$ suivant une loi de probabilité continue P déterminée par une fonction de densité de probabilité $f : \mathbb{R} \rightarrow \mathbb{R}$ de façon que :

1. la fonction f est positive ou nulle : pour tout $x \in \mathbb{R}$, $f(x) \geq 0$;
2. la fonction f est intégrable sur \mathbb{R} et son intégrale sur \mathbb{R} est égale à 1 :

$$\int_{-\infty}^{\infty} f(x)dx = 1;$$

3. la probabilité que X prenne une valeur dans un intervalle ou une réunion d'intervalles $I \subset \mathbb{R}$ est donnée par l'intégrale de f sur I :

$$P(X \in I) = \int_I f(x)dx.$$

La loi de probabilité continue P prend des valeurs réelles entre 0 et 1. Formellement,

il s'agit d'une fonction qui va de l'ensemble de tous les intervalles ou réunions d'intervalles de \mathbb{R} vers l'intervalle $[0, 1]$.

Pour un intervalle $I =]a, b] \subset \mathbb{R}$, par exemple, on note habituellement la probabilité :

$$P(X \in I) = P(a < X \leq b) = \int_a^b f(x)dx.$$

Cela correspond à l'aire sous la courbe de f entre $x = a$ et $x = b$.

Remarque : Pour une variable aléatoire continue, la probabilité que X prenne exactement une valeur x est nulle : $P(X = x) = 0$ pour tout $x \in \mathbb{R}$. Pour avoir une probabilité non nulle, il faut obligatoirement considérer la probabilité que la valeur prise par X se trouve dans un intervalle non limité à un point. De fait, pour une variable aléatoire continue, le fait d'inclure ou non les bornes de l'intervalle n'a aucun impact sur la probabilité :

$$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b).$$

Exemple. (*Loi normale centrée réduite*). Une loi de probabilité continue très importante est la loi normale centrée réduite, notée $\mathcal{N}(0, 1)$, qui a une densité de probabilité $f : \mathbb{R} \rightarrow \mathbb{R}$ donnée par :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

Il s'agit d'une fonction de Gauss ou fonction gaussienne. On peut montrer que f s'intègre bien à 1 :

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1.$$

Cette dernière intégrale a été calculée en utilisant la célèbre intégrale de Gauss $\int_{-\infty}^{\infty} e^{-y^2} dy = \sqrt{\pi}$ et en effectuant le changement de variable $y = x/\sqrt{2}$. La courbe représentative de f est tracée sur la figure II.3.1. Par exemple, la probabilité $P(1 < X \leq 2)$ correspond à l'aire sous la courbe entre $x = 1$ et $x = 2$.

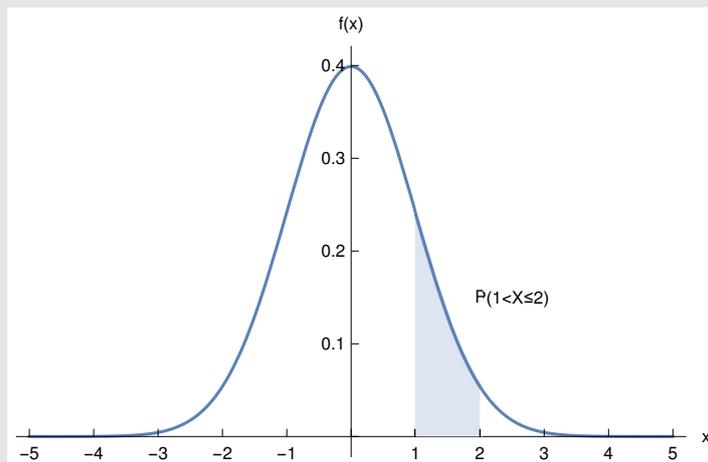


FIGURE II.3.1 – Densité de probabilité f de la loi normale centrée réduite $\mathcal{N}(0, 1)$. La probabilité $P(1 < X \leq 2)$ correspond à l'aire sous la courbe entre $x = 1$ et $x = 2$.

Exemple. (Probabilité de présence d'une particule en mécanique quantique). En mécanique quantique, la position X d'une particule (comme un électron), dans un espace à une dimension pour simplifier, est une variable aléatoire continue. La densité de probabilité associée est $x \mapsto f(x) = |\psi(x)|^2$ où $\psi : \mathbb{R} \rightarrow \mathbb{C}$ est la fonction d'onde de l'état dans lequel se trouve la particule. La probabilité de trouver la particule dans l'intervalle $[a, b]$ est donc

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

1.2 Fonction de répartition

Définition (Fonction de répartition). La fonction de répartition d'une variable aléatoire X est la fonction $F : \mathbb{R} \rightarrow \mathbb{R}$ définie par :

$$F(t) = P(X \leq t).$$

Pour une variable aléatoire continue, cela correspond à intégrer la densité de probabilité de $-\infty$ jusqu'à t :

$$F(t) = \int_{-\infty}^t f(x) dx.$$

Si f est continue, F est donc une primitive de f , ou inversement f est la dérivée de F , c'est-à-dire $F' = f$.

La fonction de répartition représente les probabilités cumulées de $-\infty$ à t . Elle caractérise entièrement la loi de probabilité.

Théorème (Propriétés de la fonction de répartition). La fonction de répartition F d'une variable aléatoire satisfait les propriétés :

- $0 \leq F(t) \leq 1$ pour tout $t \in \mathbb{R}$.
- F est croissante.
- $F(-\infty) = \lim_{t \rightarrow -\infty} F(t) = 0$ et $F(+\infty) = \lim_{t \rightarrow +\infty} F(t) = 1$.

L'utilité de la fonction de répartition est qu'elle permet d'exprimer la probabilité que X prenne une valeur dans l'intervalle $]a, b]$ sous la forme :

$$P(a < X \leq b) = F(b) - F(a).$$

En particulier, on a :

$$P(X > a) = P(a < X < +\infty) = F(+\infty) - F(a) = 1 - F(a).$$

Exemple. (*Fonction de répartition de loi normale centrée réduite*). La fonction de répartition de loi normale centrée réduite $\mathcal{N}(0, 1)$ peut être exprimée sous la forme :

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{t}{\sqrt{2}} \right) \right),$$

où $\operatorname{erf} : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction spéciale appelée « fonction d'erreur » et définie par :

$$\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx.$$

La fonction de répartition F de la loi normale centrée réduite est représentée sur la figure II.3.2.

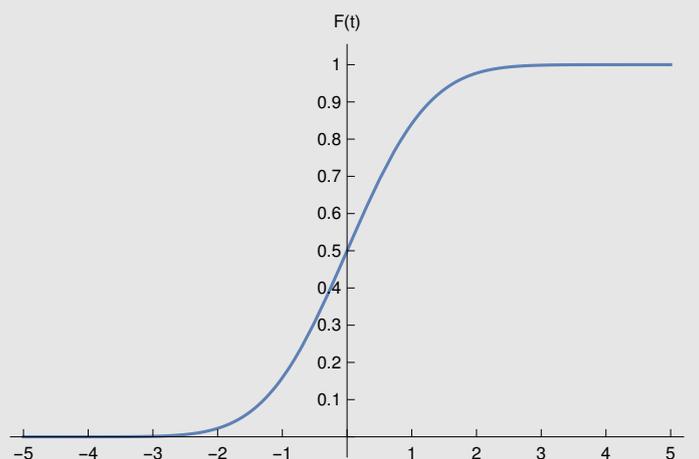


FIGURE II.3.2 – Fonction de répartition F de loi normale centrée réduite $\mathcal{N}(0, 1)$.

Remarque : Pour une loi de probabilité continue, la médiane m est définie comme la valeur telle que $\overline{F}(m) = 1/2$, c'est-à-dire $P(X \leq m) = 1/2$. En utilisant $F(+\infty) = 1 = P(X \leq m) + P(X \geq m)$, on a donc aussi $P(X \geq m) = 1/2$. La médiane est donc le point séparant la densité de probabilité en deux parties de poids égaux.

1.3 Espérance, variance et écart-type

Définition (Espérance d'une variable aléatoire continue). Soit une variable aléatoire continue X de densité de probabilité f . On définit l'espérance (ou la moyenne) de X par (si l'intégrale existe) :

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx.$$

En pratique, on parle souvent de l'espérance de la loi de probabilité sans faire explicitement référence à une variable aléatoire.

Nous verrons plus tard que l'espérance $E(X)$ correspond bien à l'interprétation intuitive suivante : si on répète une même expérience aléatoire n fois alors la moyenne des valeurs obtenues pour la variable aléatoire X tend vers $E(X)$ quand $n \rightarrow \infty$.

On peut appliquer une fonction $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ à une variable aléatoire X . On obtient une nouvelle variable aléatoire, notée $\varphi(X)$, qui suit une loi de probabilité pouvant être déduite de la loi de probabilité de X . En particulier, l'espérance de $\varphi(X)$ peut être calculée facilement. Pour une variable aléatoire continue, on a (si l'intégrale existe) :

$$E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f(x) dx.$$

En choisissant $\varphi(X) = cX$ et $\varphi(X) = X + c$ où c est une constante réelle, on peut ainsi

obtenir les espérances des variables aléatoires cX et $X + c$.

Théorème (Espérances de cX et de $X + c$). Soit une variable aléatoire X . Les espérances de cX et de $X + c$ où c est une constante réelle sont :

$$E(cX) = cE(X) \quad \text{et} \quad E(X + c) = E(X) + c.$$

En choisissant $\varphi(X) = X^2$, on peut aussi définir l'espérance du carré de X .

Pour une variable aléatoire continue, on a (si l'intégrale existe) :

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx.$$

Nous en avons besoin pour définir la variance et l'écart-type de X .

Définition (Variance et écart-type d'une variable aléatoire). Soit une variable aléatoire X . La variance de X , si elle existe, est

$$V(X) = E((X - E(X))^2) = E(X^2) - E(X)^2,$$

et l'écart-type de X est

$$\sigma(X) = \sqrt{V(X)}.$$

En pratique, on parle souvent de la variance et de l'écart-type de la loi de probabilité sans faire explicitement référence à une variable aléatoire.

La variance ou l'écart-type nous renseigne sur l'écart moyen des valeurs prises par X par rapport à la moyenne $E(X)$.

Avec les propriétés de l'espérance, on peut facilement déterminer les variances de cX et de $X + c$ où c est une constante réelle.

Théorème (Variances et écarts-types de cX et de $X + c$). Soit une variable aléatoire X . Les variances de cX et de $X + c$ où c est une constante réelle sont :

$$V(cX) = c^2 V(X) \quad \text{et} \quad V(X + c) = V(X).$$

Les écarts-types correspondants sont donc :

$$\sigma(cX) = |c| \sigma(X) \quad \text{et} \quad \sigma(X + c) = \sigma(X).$$

Exemple. (*Espérance, variance et écart-type de loi normale centrée réduite*). Pour une variable aléatoire continue X suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$, on trouve que l'espérance est nulle :

$$E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-x^2/2} dx = 0,$$

puisqu'on intègre une fonction impaire sur un intervalle symétrique autour de $x = 0$. Par ailleurs, on peut calculer $E(X^2)$ par intégration par parties en posant $u(x) = x$ et $v'(x) = x e^{-x^2/2}$, et donc $u'(x) = 1$ et $v(x) = -e^{-x^2/2}$:

$$\begin{aligned} E(X^2) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \left(x e^{-x^2/2} \right) dx \\ &= \frac{1}{\sqrt{2\pi}} \left[-x e^{-x^2/2} \right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} dx = 1. \end{aligned}$$

Ceci montre que la variance et l'écart-type sont 1 :

$$V(X) = 1 \quad \text{et} \quad \sigma(X) = 1.$$

Exemple. (*Principe d'incertitude de Heisenberg*). En mécanique quantique, la position X et la quantité de mouvement (ou impulsion) P d'une particule (dans un espace à une dimension) sont deux variables aléatoires continues. Le célèbre principe d'incertitude de Heisenberg dit que l'on ne peut pas déterminer simultanément X et P avec une précision arbitraire. Il s'agit d'une inégalité qui s'exprime avec les écarts-types de X et P :

$$\sigma(X) \sigma(P) \geq \frac{\hbar}{2},$$

où $\hbar = h/(2\pi)$ est la constante de Planck réduite.

1.4 Variable aléatoire centrée et réduite

Définition (Variable aléatoire centrée et réduite). Soit une variable aléatoire X . On dit que :

- X est centrée si son espérance est nulle : $E(X) = 0$.
- X est réduite si sa variance est égale à 1 : $V(X) = 1$.

Nous avons introduit plus haut la loi normale centrée réduite $\mathcal{N}(0, 1)$ de densité de probabilité $f : \mathbb{R} \rightarrow \mathbb{R}$ donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

qui a une espérance de 0 et une variance 1. C'est la signification du « (0,1) » dans $\mathcal{N}(0,1)$ et cela explique la qualification « centrée réduite » dans le nom de cette loi.

Il s'agit en fait d'un cas particulier d'une loi de probabilité continue plus générale : la loi normale $\mathcal{N}(\mu, \sigma^2)$ (où $\mu \in \mathbb{R}$ et $\sigma \in]0, +\infty[$ sont deux paramètres) de densité de probabilité $f_{\mu, \sigma} : \mathbb{R} \rightarrow \mathbb{R}$ donnée par

$$f_{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2},$$

qui a une espérance de μ et une variance σ^2 .

Ces deux lois sont reliées par un simple changement de variables. Si X est une variable aléatoire suivant la loi normale $\mathcal{N}(\mu, \sigma^2)$, alors la variable aléatoire Z définie par

$$Z = \frac{X - \mu}{\sigma}$$

suit la loi normale centrée réduite $\mathcal{N}(0,1)$. En effet, la probabilité d'avoir X dans l'intervalle $]x_1, x_2] \subset \mathbb{R}$,

$$P(x_1 < X \leq x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2} dx,$$

peut se réexprimer comme

$$P\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-(x-\mu)^2/2\sigma^2} dx,$$

ou, en effectuant le changement de variables $z = (x - \mu)/\sigma$ dans l'intégrale,

$$P\left(\frac{x_1 - \mu}{\sigma} < \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}} \int_{\frac{x_1 - \mu}{\sigma}}^{\frac{x_2 - \mu}{\sigma}} e^{-z^2/2} dz.$$

Finalement, en posant $z_1 = (x_1 - \mu)/\sigma$ et $z_2 = (x_2 - \mu)/\sigma$, on voit que l'on a :

$$P(z_1 < Z \leq z_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz,$$

ce qui montre bien que $Z = (X - \mu)/\sigma$ suit la loi normale centrée réduite $\mathcal{N}(0,1)$.

Pour une variable aléatoire suivant n'importe quelle loi de probabilité, on peut toujours se ramener à une variable aléatoire suivant la version centrée et réduite de cette loi.

Définition (Variable aléatoire centrée et réduite associée). Soit une variable aléatoire X d'espérance $\mu = E(X)$ et d'écart-type $\sigma = \sigma(X)$. La variable aléatoire centrée et réduite associée est :

$$Z = \frac{X - \mu}{\sigma}.$$

On a donc : $E(Z) = 0$ et $V(Z) = \sigma(Z) = 1$.

1.5 Couple de variables aléatoires (optionnel)

Définition (Couple de variables aléatoires continues). On définit un couple de variables aléatoires continues (X, Y) comme une variable pouvant prendre une continuité de valeurs réelles $(x, y) \in \mathbb{R}^2$ suivant une loi de probabilité continue P déterminée par une fonction de densité de probabilité $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$ satisfaisant les propriétés habituelles :

1. la fonction $f_{X,Y}$ est positive ou nulle : pour tout $(x, y) \in \mathbb{R}^2$, $f_{X,Y}(x, y) \geq 0$;
2. la fonction $f_{X,Y}$ est intégrable sur \mathbb{R}^2 et son intégrale sur \mathbb{R}^2 est égale à 1 :

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1;$$

3. la probabilité que (X, Y) prenne une valeur dans un domaine $D \subset \mathbb{R}^2$ est donnée par l'intégrale de $f_{X,Y}$ sur D :

$$P((X, Y) \in D) = \iint_D f_{X,Y}(x, y) dx dy.$$

Par exemple, pour un domaine donné par le produit de deux intervalles $D =]x_1, x_2] \times]y_1, y_2]$, on a la probabilité :

$$P((X, Y) \in D) = P(x_1 < X \leq x_2 \text{ et } y_1 < Y \leq y_2) = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f_{X,Y}(x, y) dx dy.$$

La fonction de répartition associée $F_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$ a alors pour expression :

$$F_{X,Y}(t, s) = P(X \leq t \text{ et } Y \leq s) = \int_{-\infty}^t \int_{-\infty}^s f_{X,Y}(x, y) dx dy.$$

Définition (Lois marginales). Soit (X, Y) un couple de variables aléatoires continues de densité de probabilité $f_{X,Y} : \mathbb{R}^2 \rightarrow \mathbb{R}$. La première variable X est une variable aléatoire continue de densité de probabilité $f_X : \mathbb{R} \rightarrow \mathbb{R}$ donnée par

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy.$$

De même, la seconde variable Y est une variable aléatoire continue de densité de pro-

babilité $f_Y : \mathbb{R} \rightarrow \mathbb{R}$ donnée par

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx.$$

Les lois de probabilité de X et Y sont appelées lois marginales.

En général, la densité de probabilité $f_{X,Y}$ du couple de variables ne peut être exprimée uniquement avec les densités de probabilité f_X et f_Y de chaque variable, sauf si X et Y sont des variables dites indépendantes.

Définition (Variables aléatoires continues indépendantes). Soit un couple de variables aléatoires continues (X, Y) . Les variables aléatoires X et Y sont dites indépendantes si la densité de probabilité du couple est donnée par le produit des densités de probabilité de chaque variable :

$$\text{pour tout } (x, y) \in \mathbb{R}^2, f_{X,Y}(x, y) = f_X(x) f_Y(y).$$

Si X et Y sont indépendantes, la probabilité d'avoir $x_1 < X \leq x_2$ et $y_1 < Y \leq y_2$ est simplement le produit des probabilités :

$$P(x_1 < X \leq x_2 \text{ et } y_1 < Y \leq y_2) = P(x_1 < X \leq x_2) P(y_1 < Y \leq y_2).$$

Comme pour le cas d'une seule variable aléatoire, on peut appliquer une fonction $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}$ à un couple de variables aléatoires (X, Y) . On obtient une nouvelle variable aléatoire, notée $\varphi(X, Y)$, qui suit une autre loi de probabilité pouvant être déduite de celle de (X, Y) . Pour le cas de variables aléatoires continues, l'espérance de $\varphi(X, Y)$ est (si l'intégrale existe) :

$$E(\varphi(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi(x, y) f_{X,Y}(x, y) dx dy.$$

Ceci permet en particulier de définir l'espérance de la somme $X + Y$ et du produit XY de variables aléatoires.

Théorème (Espérance et variance de la somme de deux variables aléatoires). Pour deux variables aléatoires X et Y , on a :

$$E(X + Y) = E(X) + E(Y) \text{ et } V(X + Y) = V(X) + V(Y) + 2 \text{Cov}(X, Y),$$

où $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ est la covariance de X et Y .

La covariance de X et Y renseigne sur les corrélations existant entre les deux variables aléatoires. Pour des variables aléatoires indépendantes, elle est nulle et la variance de $X + Y$ est simplement la somme des variances de X et de Y .

Théorème (Variance de la somme de deux variables aléatoires indépendantes). Pour deux variables aléatoires X et Y indépendantes, on a $\text{Cov}(X, Y) = 0$ et donc :

$$V(X + Y) = V(X) + V(Y).$$

Tous ces résultats sur un couple de variables aléatoires se généralisent directement à une suite de n variables aléatoires (X_1, X_2, \dots, X_n) .

Exemple. (*Probabilité de présence de deux particules en mécanique quantique*). Dans l'espace à une dimension \mathbb{R} , considérons deux particules quantiques différentes de positions X_1 et X_2 (par exemple, un électron et un proton). Les positions des deux particules constituent un couple de variables aléatoires continues (X_1, X_2) . La densité de probabilité associée est $(x, y) \mapsto f_{X_1, X_2}(x, y) = |\psi(x, y)|^2$ où $\psi : \mathbb{R}^2 \rightarrow \mathbb{C}$ est la fonction d'onde de l'état dans lequel se trouvent les deux particules. La probabilité de trouver la première particule dans l'intervalle $[a, b]$ et la deuxième particule dans l'intervalle $[c, d]$ est

$$P(a \leq X_1 \leq b \text{ et } c \leq X_2 \leq d) = \int_a^b \int_c^d f_{X_1, X_2}(x, y) dx dy.$$

La probabilité de trouver la première particule dans l'intervalle $[a, b]$ quel que soit la position de la deuxième particule est

$$P(a \leq X_1 \leq b) = \int_a^b f_{X_1}(x) dx,$$

où $f_{X_1} : x \mapsto f_{X_1}(x) = \int_{-\infty}^{\infty} f_{X_1, X_2}(x, y) dy$ où est la densité de probabilité marginale de la variable X_1 . Dans le cas simple où les particules n'interagissent pas entre elles alors les variables aléatoires X_1 et X_2 sont indépendantes et on a :

$$P(a \leq X_1 \leq b \text{ et } c \leq X_2 \leq d) = P(a \leq X_1 \leq b) P(c \leq X_2 \leq d).$$

Dans le cas plus réaliste où les particules interagissent entre elles, les variables aléatoires X_1 et X_2 ne sont pas indépendantes et on a en général :

$$P(a \leq X_1 \leq b \text{ et } c \leq X_2 \leq d) \neq P(a \leq X_1 \leq b) P(c \leq X_2 \leq d).$$

Dans ce cas, on dit qu'il y a corrélation entre les particules.

1.6 Deux lois de probabilité continues importantes

Loi normale centrée réduite (ou loi normale standard)

Nous avons déjà discuté de la loi normale centrée réduite (ou loi normale standard) $\mathcal{N}(0, 1)$ de densité de probabilité $f : \mathbb{R} \rightarrow \mathbb{R}$ donnée par

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

qui a une espérance de 0 et une variance 1. La loi normale a une importante capitale à cause du théorème central limite que nous verrons plus loin et qui permet en particulier de définir des intervalles de confiance. La fonction de répartition est :

$$F(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^t e^{-x^2/2} dx = \frac{1}{2} \left(1 + \operatorname{erf} \left(\frac{t}{\sqrt{2}} \right) \right),$$

où $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-x^2} dx$ est la fonction spéciale d'erreur. Rappelons que la fonction de répartition permet de calculer les probabilités (pour a et b réels) :

$$P(X \leq a) = F(a),$$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a),$$

$$P(a < X \leq b) = F(b) - F(a).$$

Avec un ordinateur, on peut calculer facilement la valeur de $F(t)$ pour n'importe quel $t \in \mathbb{R}$. Sans ordinateur, on utilise une table de la loi normale où l'on trouve pour plusieurs valeurs de t (disons t_1, t_2, t_3 , etc...) les valeurs de $F(t)$ correspondantes ($F(t_1), F(t_2), F(t_3)$, etc...). Si on doit calculer $F(a)$, on cherche donc dans la table la valeur de t la plus proche de a , disons $t_i \approx a$, et on approche alors $F(a) \approx F(t_i)$.

La table contient uniquement des valeurs de t positives. En effet, grâce au fait que la densité de probabilité de la loi normale est symétrique par rapport à la droite $x = 0$, c'est-à-dire $f(-x) = f(x)$ pour tout $x \in \mathbb{R}$, on peut vérifier que la fonction de répartition de la loi normale a la symétrie suivante :

$$F(-t) = 1 - F(t), \quad \text{pour tout } t \in \mathbb{R}.$$

Si on veut calculer $F(-a)$ pour a positif, on écrira donc $F(-a) = 1 - F(a)$ et on cherchera dans la table la valeur de $F(a)$.

Dans certains problèmes, on se fixe une probabilité $p = F(a)$ et on veut déterminer la valeur de a correspondante. Formellement, a est donné par la fonction inverse de F , c'est-à-dire $a = F^{-1}(p)$. On utilise alors la table en sens inverse, c'est-à-dire que l'on cherche la valeur de $F(t)$ la plus proche de p , disons $F(t_i) \approx p$, et on approche alors a par la valeur t_i correspondante : $a \approx t_i$.

Enfin, pour traiter le cas d'une variable aléatoire X suivant une loi normale non-centrée et non-réduite $\mathcal{N}(\mu, \sigma^2)$ d'espérance μ et d'écart-type σ , on fait le changement de variable $Z = (X - \mu)/\sigma$ où Z suit la loi normale centrée réduite $\mathcal{N}(0, 1)$ (voir section 1.4). Par exemple, pour calculer la probabilité $P(a < X \leq b)$ on peut alors écrire

$$P(a < X \leq b) = P\left(\frac{a - \mu}{\sigma} < Z \leq \frac{b - \mu}{\sigma}\right) = F\left(\frac{b - \mu}{\sigma}\right) - F\left(\frac{a - \mu}{\sigma}\right),$$

où F est toujours la fonction de répartition de la loi normale centrée réduite.

Loi du χ^2 (khi-2)

Considérons k (entier strictement positif) variables aléatoires X_1, X_2, \dots, X_k indépendantes suivant toutes la loi normale centrée réduite $\mathcal{N}(0, 1)$. Par définition, la variable aléatoire Q définie par la somme de leurs carrés,

$$Q = X_1^2 + X_2^2 + \dots + X_k^2,$$

suit la loi du χ^2 à k degrés de liberté, notée χ_k^2 . La loi χ_k^2 est utilisée dans le test d'hypothèse statistique dit « test du χ^2 » que nous verrons plus loin.

Il est possible de déterminer que la densité de probabilité $f_k : \mathbb{R} \rightarrow \mathbb{R}$ de la loi χ_k^2 est donnée par

$$f_k(x) = \begin{cases} 0 & \text{si } x < 0 \\ \frac{1}{2^{k/2}\Gamma(k/2)} x^{k/2-1} e^{-x/2} & \text{si } x \geq 0, \end{cases}$$

où $\Gamma :]0, +\infty[\rightarrow \mathbb{R}$ est une fonction spéciale appelée « fonction gamma » et définie par :

$$\Gamma(z) = \int_0^{+\infty} y^{z-1} e^{-y} dy.$$

La fonction gamma est la généralisation de la fonction factorielle à n'importe quel nombre réel positif. Dans le cas particulier d'un entier n strictement positif, on a $\Gamma(n) = (n-1)!$. La loi χ_k^2 a une espérance k et une variance $2k$. La densité de probabilité f_k est tracée dans la figure II.3.3 pour $k = 1$, $k = 4$ et $k = 10$. On voit en effet que, plus k augmente, plus f_k se déplace vers les grandes valeurs de x et s'étale. On peut montrer (avec le théorème central limite) que, dans la limite $k \rightarrow \infty$, la loi χ_k^2 tend asymptotiquement vers la loi normale d'espérance k et de variance $2k$.

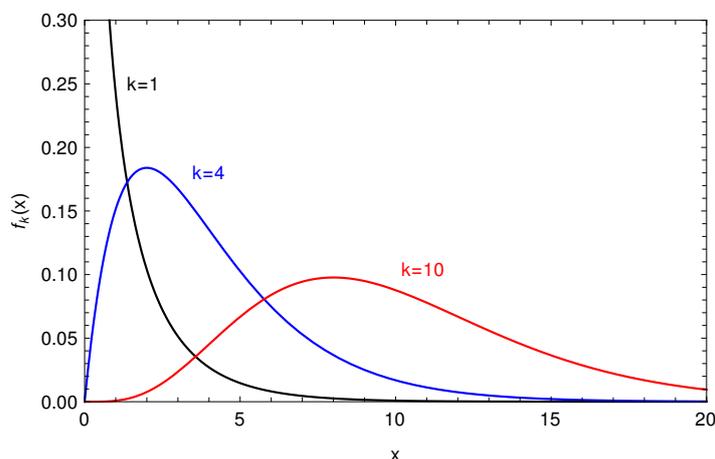


FIGURE II.3.3 – Densité de probabilité f_k de la loi χ_k^2 pour $k = 1$, $k = 4$ et $k = 10$.

Comme pour la loi normale, il existe des tables donnant la fonction de répartition de loi χ_k^2 .

2 Théorème central limite et intervalles de confiance

2.1 Théorème central limite

Considérons la situation expérimentale commune suivante. On veut mesurer une quantité physique X (par exemple, le volume à l'équivalence lors d'un dosage acido-basique). A cause des erreurs de mesure, si on répète l'expérience n fois, chaque nouvelle mesure de X conduit à des valeurs (légèrement) différentes x_1, x_2, \dots, x_n . On peut considérer X comme une variable aléatoire continue pouvant prendre différentes valeurs suivant une loi de probabilité inconnue. La valeur physique que l'on cherche à mesurer peut alors être définie comme l'espérance $E(X)$ de X . La valeur de l'espérance $E(X)$ peut être estimée par la valeur moyenne des n valeurs mesurées x_1, x_2, \dots, x_n

$$m_n = \frac{x_1 + x_2 + \dots + x_n}{n},$$

pour n suffisamment grand. Le théorème central limite permet de justifier ce résultat et donne une estimation de l'erreur commise par rapport à $E(X)$.

Commençons par introduire le concept d'échantillon aléatoire.

Définition (Échantillon aléatoire). Soit une variable aléatoire X . Un échantillon aléatoire de X de taille n est une suite de n variables aléatoires (X_1, X_2, \dots, X_n) indépendantes deux à deux et suivant chacune la même loi de probabilité que X .

Suivant l'exemple plus haut, ceci correspond à répéter l'expérience n fois, la variable aléatoire X_i correspondant à la i ème mesure. Attention à ne pas confondre les variables aléatoires (X_1, X_2, \dots, X_n) décrivant un échantillon générique (où les valeurs prises par les variables aléatoires ne sont pas encore déterminées) et la réalisation d'un échantillon particulier où les valeurs des variables aléatoires sont fixées à des valeurs particulières x_1, x_2, \dots, x_n .

On peut maintenant énoncer le théorème central limite.

Théorème (Théorème central limite). Soit une variable aléatoire X suivant une loi de probabilité quelconque d'espérance $\mu = E(X)$ et de variance $\sigma^2 = V(X)$ finies. On considère un échantillon aléatoire (X_1, X_2, \dots, X_n) de la variable aléatoire X de taille n . La moyenne des variables aléatoires de l'échantillon

$$M_n = \frac{X_1 + X_2 + \dots + X_n}{n},$$

est une variable aléatoire qui suit, asymptotiquement dans la limite $n \rightarrow \infty$, la loi normale $\mathcal{N}(\mu, \sigma^2/n)$ d'espérance μ et de variance σ^2/n .

Il est important de comprendre que M_n est une variable aléatoire car elle prend différentes valeurs sur différents échantillons. Dans l'exemple plus haut, m_n correspond à la valeur prise par la variable aléatoire M_n sur un échantillon particulier. Le théorème central limite nous dit donc que, pour n suffisamment grand, la variable aléatoire M_n suit une loi normale de même espérance que X , c'est-à-dire $E(M_n) = E(X) = \mu$, et dont la variance décroît en $1/n$, c'est-à-dire $V(M_n) = \sigma^2/n$, ou de manière équivalente dont l'écart-type décroît en $1/\sqrt{n}$, c'est-à-dire $\sigma(M_n) = \sigma/\sqrt{n}$. Pour n suffisamment grand, la densité de probabilité de M_n est donc celle de la loi normale d'espérance μ et d'écart-type σ/\sqrt{n}

$$f_{\mu, \sigma/\sqrt{n}}(x) = \frac{1}{(\sigma/\sqrt{n})\sqrt{2\pi}} e^{-(x-\mu)^2/2(\sigma/\sqrt{n})^2}.$$

Il s'agit une densité de probabilité centrée sur μ et dont l'écart-type diminue quand n augmente. Cette diminution de l'écart-type est illustrée sur la figure II.3.4.

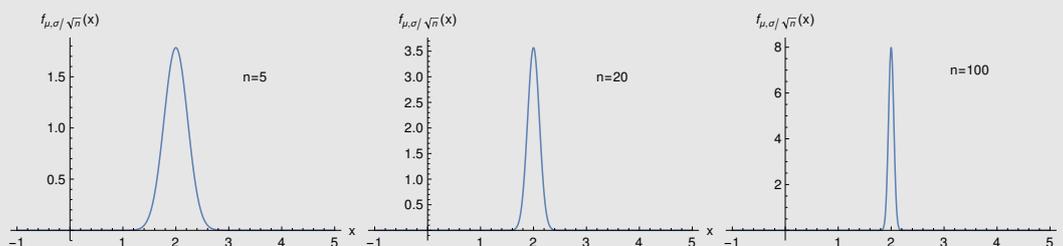


FIGURE II.3.4 – Densité de probabilité de la loi normale $f_{\mu, \sigma/\sqrt{n}}$ pour $\mu = 2$, $\sigma = 0.5$, et $n = 5$, $n = 20$, $n = 100$.

On voit donc que, quand n augmente, la variable aléatoire M_n ne peut prendre que des valeurs de plus en plus resserrées autour de la valeur $\mu = E(X)$ que l'on cherche à déterminer. Dans la limite $n \rightarrow \infty$, M_n ne peut prendre que la valeur μ . On dit que M_n est un estimateur statistique de la valeur μ avec des fluctuations décroissant en $1/\sqrt{n}$. Concrètement, cela veut donc dire que la moyenne m_n calculée sur un échantillon particulier (c'est-à-dire la valeur prise par la variable aléatoire M_n sur cet échantillon particulier) est nécessairement une bonne approximation à μ lorsque la taille de l'échantillon n est suffisamment grande. Par ailleurs, nous allons à présent voir que le fait que M_n suive asymptotiquement une loi normale permet d'estimer, pour une taille d'échantillon n fixée suffisamment grande, l'erreur commise par rapport à la valeur exacte μ , sous la forme d'un intervalle de confiance.

Remarque : Nous avons donné le théorème central limite pour la variable aléatoire $\overline{M}_n = (X_1 + X_2 + \dots + X_n)/n$ correspondant à la moyenne sur l'échantillon. On peut aussi exprimer le théorème central limite pour la variable aléatoire $\Sigma_n = X_1 + X_2 + \dots + X_n$ correspondant à la somme des variables aléatoires de l'échantillon. Dans ce cas, Σ_n suit, asymptotiquement dans la limite $n \rightarrow \infty$, la loi normale $\mathcal{N}(n\mu, n\sigma^2)$ d'espérance $n\mu$ et de variance $n\sigma^2$.

2.2 Intervalles de confiance

Définition (Intervalle de confiance). Pour la variable aléatoire M_n représentant la moyenne sur un échantillon de taille n (définie dans le théorème 2.1) d'espérance μ et d'écart-type σ/\sqrt{n} , on dit que

$$\left[M_n - \varepsilon \frac{\sigma}{\sqrt{n}}, M_n + \varepsilon \frac{\sigma}{\sqrt{n}} \right]$$

est un intervalle de confiance au seuil p (ou, de manière équivalente, au risque $1 - p$) si le nombre réel positif ε est tel que

$$P\left(|M_n - \mu| < \varepsilon \frac{\sigma}{\sqrt{n}}\right) = p.$$

Cela signifie qu'il y a une probabilité p que l'espérance μ (que l'on cherche à déterminer) se trouve dans l'intervalle $[M_n - \varepsilon\sigma/\sqrt{n}, M_n + \varepsilon\sigma/\sqrt{n}]$. Inversement, il y a un risque de probabilité $1 - p$ que μ se trouve en dehors de cet intervalle. C'est une façon de donner l'erreur statistique sur l'estimation de l'espérance μ . Notez que l'intervalle de confiance est un intervalle aléatoire puisque M_n est une variable aléatoire prenant différentes valeurs sur différentes réalisations de l'échantillon.

En pratique, on peut soit se donner une valeur ε et calculer la valeur de p correspondante, soit se donner une valeur de p et calculer la valeur ε correspondante. Pour faire ce calcul, il est pratique d'introduire la variable Z_n centrée et réduite

$$Z_n = \frac{M_n - \mu}{\sigma/\sqrt{n}},$$

qui, d'après le théorème central limite, pour n suffisamment grand, suit la loi normale centrée réduite. On peut alors écrire la probabilité $P(|M_n - \mu| < \varepsilon\sigma/\sqrt{n})$ comme

$$\begin{aligned} P\left(|M_n - \mu| < \varepsilon \frac{\sigma}{\sqrt{n}}\right) &= P(|Z_n| < \varepsilon) \\ &= P(-\varepsilon < Z_n < \varepsilon) \\ &= F(\varepsilon) - F(-\varepsilon), \end{aligned}$$

où F est la fonction de répartition de la loi normale centrée réduite. Finalement, en utilisant $F(-\varepsilon) = 1 - F(\varepsilon)$ (voir la section 1.6), on arrive à une forme pratique de l'équation reliant ε et p

$$2F(\varepsilon) - 1 = p,$$

ou encore

$$F(\varepsilon) = \frac{p + 1}{2}.$$

On rappelle que les valeurs de F peuvent être obtenues soit avec un ordinateur soit en consultant une table de la loi normale.

Voici trois exemples importants où l'on fixe la valeur de ε :

- Si on choisit $\varepsilon = 1$, on trouve $p = 2F(1) - 1 \approx 0.68$. L'intervalle à « un sigma » $[M_n - \sigma/\sqrt{n}, M_n + \sigma/\sqrt{n}]$ est donc un intervalle de confiance à environ 68 %.
- Si on choisit $\varepsilon = 2$, on trouve $p = 2F(2) - 1 \approx 0.95$. L'intervalle à « deux sigmas » $[M_n - 2\sigma/\sqrt{n}, M_n + 2\sigma/\sqrt{n}]$ est donc un intervalle de confiance à environ 95 %.
- Si on choisit $\varepsilon = 3$, on trouve $p = 2F(3) - 1 \approx 0.997$. L'intervalle à « trois sigmas » $[M_n - 3\sigma/\sqrt{n}, M_n + 3\sigma/\sqrt{n}]$ est donc un intervalle de confiance à environ 99.7 %.

Inversement, voici des exemples où l'on fixe la valeur de p :

- Si on choisit $p = 0.90$, on trouve $\varepsilon \approx 1.64$. L'intervalle $[M_n - 1.64\sigma/\sqrt{n}, M_n + 1.64\sigma/\sqrt{n}]$ est donc un intervalle de confiance à 90 %.
- Si on choisit $p = 0.95$, on trouve $\varepsilon \approx 1.96$. L'intervalle $[M_n - 1.96\sigma/\sqrt{n}, M_n + 1.96\sigma/\sqrt{n}]$ est donc un intervalle de confiance à 95 %.
- Si on choisit $p = 0.99$, on trouve $\varepsilon \approx 2.58$. L'intervalle $[M_n - 2.58\sigma/\sqrt{n}, M_n + 2.58\sigma/\sqrt{n}]$ est donc un intervalle de confiance à 99 %.

Tout ceci peut être appliqué si on connaît σ . Le problème est que habituellement on ne connaît pas σ puisque c'est l'écart-type de la variable aléatoire initiale X dont la loi de probabilité est inconnue. La solution est d'estimer σ par un estimateur statistique de l'écart-type sur l'échantillon, par exemple :¹

$$S_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - M_n)^2}.$$

L'estimateur statistique S_n est une variable aléatoire. La variable aléatoire précédemment introduite, $Z_n = (M_n - \mu)/(\sigma/\sqrt{n})$, est alors remplacée par

$$\tilde{Z}_n = \frac{M_n - \mu}{S_n/\sqrt{n}},$$

qui suit encore la loi normale centrée réduite pour n suffisamment grand. L'intervalle de confiance au seuil p devient alors

$$\left[M_n - \varepsilon \frac{S_n}{\sqrt{n}}, M_n + \varepsilon \frac{S_n}{\sqrt{n}} \right],$$

avec toujours $2F(\varepsilon) - 1 = p$ où F est la fonction de répartition de la loi normale centrée réduite.

1. La présence du facteur $n - 1$ au lieu de n au dénominateur peut a priori surprendre : il s'agit de l'estimateur de l'écart-type dit non-biaisé. Le facteur $n - 1$ au lieu de n permet de corriger un biais provenant du fait que M_n est lui-même un estimateur de l'espérance μ . Pour n suffisamment grand, on peut bien sûr approcher $n - 1$ par n .

Sur un échantillon particulier, M_n prend une valeur

$$m_n = \frac{1}{n} \sum_{i=1}^n x_i$$

et l'estimateur S_n de l'écart-type prend une valeur

$$s_n = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - m_n)^2}.$$

Cela donnera donc en pratique un intervalle de confiance :

$$\left[m_n - \varepsilon \frac{s_n}{\sqrt{n}}, m_n + \varepsilon \frac{s_n}{\sqrt{n}} \right].$$

Sur un graphique, on représente cet intervalle de confiance au seuil p par une barre d'erreur autour de la valeur m_n . Cette barre d'erreur doit être interprétée comme une estimation de l'incertitude statistique sur l'estimation de la valeur μ que l'on cherche à déterminer. On peut considérer que la valeur recherchée μ à une probabilité p de se trouver dans cette barre d'erreur. Souvent les barres d'erreur représentées sur un graphique correspondent aux intervalles de confiance à « un sigma » c'est-à-dire à $p \approx 0.68$ ou à « deux sigmas » c'est-à-dire à $p \approx 0.95$.

Exemple. (*Dosage acido-basique*). On a répété un dosage acido-basique $n = 10$ fois et on a mesuré à chaque fois un volume v à l'équivalence légèrement différent :

v_i (mL)	2.10	1.85	2.00	2.05	1.95	1.90	2.10	2.05	2.00	1.95
------------	------	------	------	------	------	------	------	------	------	------

La valeur moyenne du volume à l'équivalence sur cet échantillon est

$$\bar{v} = \frac{1}{10} \sum_{i=1}^{10} v_i = 1.995 \text{ mL.}$$

On peut estimer l'écart-type sur le volume à l'équivalence par

$$\sigma \approx s = \sqrt{\frac{1}{9} \sum_{i=1}^{10} (v_i - \bar{v})^2} \approx 0.083 \text{ mL.}$$

On a donc un écart-type sur la valeur moyenne du volume à l'équivalence de $s/\sqrt{10} \approx 0.026$ mL. En utilisant la loi normale $\mathcal{N}(0, 1)$ pour calculer les intervalles de confiance, on trouve par exemple :

— l'intervalle de confiance à « un sigma » :

$$[1.995 - 0.026, 1.995 + 0.026] \approx [1.97, 2.02] \text{ mL;}$$

— l'intervalle de confiance au seuil de 90 % :

$$[1.995 - 1.64 \times 0.026, 1.995 + 1.64 \times 0.026] \approx [1.95, 2.04] \text{ mL.}$$

Chapitre II.4

Outils d'analyse matricielle

Le calcul matriciel est un outil mathématique incontournable. Il permet de résoudre des systèmes d'équations linéaires, qui apparaissent très souvent en sciences, comme par exemple dans le modèle de Hückel en chimie quantique. Les matrices permettent aussi de représenter les opérations de symétrie d'une molécule et sont donc à la base de la théorie des groupes de symétrie. Les matrices sont aussi très utilisées dans de nombreuses méthodes d'analyse de données, comme par exemple l'analyse en composantes principales.

1 Matrices et systèmes d'équations linéaires

1.1 Aspect historique

La genèse de la théorie des matrices est un peu confuse et passe par de nombreux pays : Grande-Bretagne, France, Allemagne, Chine. Les chinois au II^e siècle avant J.C. furent les premiers à manipuler des tableaux de nombres et à appliquer un algorithme maintenant connu sous le nom de procédé d'élimination de Gauss-Jordan. La notion de matrice est intimement liée à la notion de résolution de systèmes d'équations linéaires qui est un problème très ancien. La mise en tableau des coefficients qui sont devant chaque inconnue a permis dans un premier temps de définir la notion de déterminant. Le calcul du déterminant d'un système d'équations linéaires permet alors de justifier l'existence de solutions. Cet outil est apparu avec Cardan pour les déterminants 2×2 à la fin du XVI^e siècle puis Leibniz et Takakazu au Japon ont étudié les déterminants d'ordre supérieur dès la fin du XVII^e siècle, soit 150 ans avant la définition formelle de la notion de matrice par Sylvester et Cayley. Sylvester publie un article en 1855 où il décrit les matrices comme une notation commode pour représenter les systèmes linéaires et il explique l'intérêt de cette notation matricielle pour étudier les fonctions linéaires. Trois ans plus tard, Cayley élabore une véritable théorie des matrices dans un nouvel article intitulé "A memoir on the theory of matrices".

1.2 Représentation matricielle d'un système d'équations linéaires

Considérons un système de n équations linéaires à n inconnues x_1, x_2, \dots, x_n :

$$\begin{cases} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n = b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n = b_2 \\ \vdots \\ \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,n}x_n = b_n \end{cases},$$

où les coefficients $a_{i,j}$ et b_i sont des nombres (réels ou complexes) supposés connus. On peut réécrire ce système d'équations linéaires de façon matricielle. Pour cela, on introduit la matrice \mathbf{A} associée à ce système d'équations linéaires :

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix},$$

il s'agit du tableau à n lignes et à n colonnes contenant les coefficients $a_{i,j}$. On introduit également le vecteur colonne \mathbf{X} des inconnues qui est la matrice à une colonne contenant les inconnues x_1, x_2, \dots, x_n :

$$\mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix},$$

et le vecteur colonne \mathbf{B} du second membre du système d'équations linéaires qui est la matrice à une colonne contenant les nombres b_1, b_2, \dots, b_n :

$$\mathbf{B} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

On définit le produit entre la matrice \mathbf{A} et le vecteur \mathbf{X} , noté $\mathbf{A}\mathbf{X}$ ou $\mathbf{A} \cdot \mathbf{X}$, comme

$$\mathbf{A}\mathbf{X} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ a_{2,1}x_1 + a_{2,2}x_2 + \cdots + a_{2,n}x_n \\ \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \cdots + a_{n,n}x_n \end{pmatrix},$$

c'est-à-dire $\mathbf{A X}$ est un vecteur colonne contenant sur la $i^{\text{ème}}$ ligne le membre de gauche de la $i^{\text{ème}}$ équation du système. Le système d'équations linéaires peut alors être écrit de façon matricielle comme

$$\mathbf{A X} = \mathbf{B}.$$

Cette écriture a l'avantage d'être très compacte. On peut alors utiliser le calcul matriciel pour résoudre le système d'équations.

Exemple. (*Système de réactions chimiques*). Considérons le système de trois réactions chimiques suivant :

$$\begin{aligned} (1) \quad & R_1 + R_2 = R_3 \\ (2) \quad & 2R_2 = R_1 + R_3 \\ (3) \quad & R_1 = 2R_3, \end{aligned}$$

où R_1 , R_2 et R_3 sont des réactants. Supposons que l'on connaisse, à un instant t , la variation du nombre de moles de chaque réactant par rapport à l'instant initial, Δn_{R_1} , Δn_{R_2} et Δn_{R_3} , et on veut déterminer les avancements correspondants ξ_1 , ξ_2 et ξ_3 des trois réactions. En considérant la variation de chaque réactant dans les trois réactions, on arrive au système de 3 équations linéaires à 3 inconnues suivant :

$$\begin{cases} -\xi_1 + \xi_2 - \xi_3 = \Delta n_{R_1} \\ -\xi_1 - 2\xi_2 = \Delta n_{R_2} \\ \xi_1 + \xi_2 + 2\xi_3 = \Delta n_{R_3} \end{cases}.$$

Ce système d'équations linéaires s'écrit de façon matricielle comme $\mathbf{A X} = \mathbf{B}$ avec

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & -1 \\ -1 & -2 & 0 \\ 1 & 1 & 2 \end{pmatrix} \quad \text{et} \quad \mathbf{X} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \quad \text{et} \quad \mathbf{B} = \begin{pmatrix} \Delta n_{R_1} \\ \Delta n_{R_2} \\ \Delta n_{R_3} \end{pmatrix}.$$

1.3 Quelques définitions générales sur les matrices

Commençons par donner la définition générale d'une matrice.

Définition (Matrice à m lignes et n colonnes). Une matrice \mathbf{A} à m lignes et n colonnes (ou matrice de taille $m \times n$) est un tableau de nombres (réels ou complexes) $a_{i,j}$, noté :

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{pmatrix}.$$

On utilise aussi la notation compacte $\mathbf{A} = (a_{i,j})$, où $a_{i,j}$ est l'élément de la matrice \mathbf{A} sur la ligne i et la colonne j .

Deux matrices $\mathbf{A} = (a_{i,j})$ et $\mathbf{B} = (b_{i,j})$ de taille $m \times n$ sont égales, $\mathbf{A} = \mathbf{B}$, si tous leurs éléments sont égaux, c'est-à-dire $a_{i,j} = b_{i,j}$ pour $1 \leq i \leq m$ et $1 \leq j \leq n$. On s'intéressera particulièrement aux *matrices carrées*, c'est-à-dire avec le même nombre de lignes et de colonnes ($m = n$). Une matrice à n lignes et une seule colonne s'appelle un *vecteur colonne* à n composantes.

On définit la somme de deux matrices de même taille en faisant la somme élément par élément.

Définition (Somme de deux matrices). Soit deux matrices $\mathbf{A} = (a_{i,j})$ et $\mathbf{B} = (b_{i,j})$ de même taille $m \times n$. La somme de \mathbf{A} et \mathbf{B} est la matrice $\mathbf{C} = (c_{i,j})$ de taille $m \times n$ dont les éléments sont donnés par, pour $1 \leq i \leq m$ et $1 \leq j \leq n$:

$$c_{i,j} = a_{i,j} + b_{i,j}.$$

Pour pouvoir faire la somme matricielle de \mathbf{A} et \mathbf{B} , il faut que \mathbf{A} et \mathbf{B} soient de même taille. La somme matricielle est commutative, c'est-à-dire $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.

Définition (Produit d'un scalaire par une matrice). Soit un nombre (ou scalaire) λ réel ou complexe et une matrice $\mathbf{A} = (a_{i,j})$ de taille $m \times n$. Le produit du scalaire λ par la matrice \mathbf{A} est la matrice $\lambda\mathbf{A} = (\lambda a_{i,j})$ de taille $m \times n$ obtenue en multipliant chaque élément de \mathbf{A} par λ .

Le produit de deux matrices est plus compliqué et s'effectue en multipliant les lignes de la matrice de gauche par les colonnes de la matrice de droite.

Définition (Produit de deux matrices). Soit une matrice $\mathbf{A} = (a_{i,j})$ de taille $m \times p$ et une matrice $\mathbf{B} = (b_{i,j})$ de taille $p \times n$. Le produit de \mathbf{A} par \mathbf{B} , noté $\mathbf{A}\mathbf{B}$ ou $\mathbf{A} \cdot \mathbf{B}$, est la matrice $\mathbf{C} = \mathbf{A}\mathbf{B} = (c_{i,j})$ de taille $m \times n$ dont les éléments sont donnés par, pour $1 \leq i \leq m$ et $1 \leq j \leq n$:

$$c_{i,j} = \sum_{k=1}^p a_{i,k} b_{k,j}.$$

Pour pouvoir faire le produit matriciel de \mathbf{A} par \mathbf{B} , il faut que le nombre de colonnes de \mathbf{A} soit égale au nombre de lignes de \mathbf{B} . Attention, le produit matriciel n'est généralement pas commutatif, c'est-à-dire, même dans le cas où on l'on peut faire le produit matriciel dans les deux sens, on peut avoir $\mathbf{AB} \neq \mathbf{BA}$.

Par exemple, le produit de deux matrices de taille 2×2 est

$$\begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} \begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix} = \begin{pmatrix} a_{1,1}b_{1,1} + a_{1,2}b_{2,1} & a_{1,1}b_{1,2} + a_{1,2}b_{2,2} \\ a_{2,1}b_{1,1} + a_{2,2}b_{2,1} & a_{2,1}b_{1,2} + a_{2,2}b_{2,2} \end{pmatrix}.$$

On peut se convaincre aisément sur cet exemple que le produit matriciel n'est en général pas commutatif :

$$\begin{pmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \end{pmatrix} \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix} = \begin{pmatrix} b_{1,1}a_{1,1} + b_{1,2}a_{2,1} & b_{1,1}a_{1,2} + b_{1,2}a_{2,2} \\ b_{2,1}a_{1,1} + b_{2,2}a_{2,1} & b_{2,1}a_{1,2} + b_{2,2}a_{2,2} \end{pmatrix}.$$

Il est utile de définir la matrice nulle.

Définition (Matrice nulle). La matrice nulle de taille $m \times n$, notée $\mathbf{0}$, est la matrice dont tous les éléments sont nuls.

La matrice nulle $\mathbf{0}$ joue le même rôle pour les matrices que le nombre 0 joue pour les nombres. Si \mathbf{A} est une matrice et $\mathbf{0}$ est la matrice nulle de même taille, on a $\mathbf{A} + \mathbf{0} = \mathbf{A}$. Si λ est un scalaire, on a $\lambda \mathbf{0} = \mathbf{0}$, et si on multiplie n'importe quelle matrice \mathbf{A} par le nombre 0, on obtient la matrice nulle de même taille, $0 \mathbf{A} = \mathbf{0}$. Enfin, pour n'importe quelle matrice \mathbf{A} , on a $\mathbf{A} \mathbf{0} = \mathbf{0}$ et $\mathbf{0} \mathbf{A} = \mathbf{0}$ où $\mathbf{0}$ est à chaque fois une matrice nulle de taille compatible avec le produit matriciel.

Finissons cette section par les définitions de la matrice transposée et de la matrice adjointe d'une matrice.

Définition (Matrice transposée). Soit une matrice $\mathbf{A} = (a_{i,j})$ de taille $m \times n$. La matrice transposée de \mathbf{A} , notée \mathbf{A}^T , est la matrice de taille $n \times m$ pour laquelle on a intervertit les lignes et les colonnes de \mathbf{A} :

$$\mathbf{A}^T = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{n,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{n,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,m} & a_{2,m} & \cdots & a_{n,m} \end{pmatrix}.$$

En notation compacte, on a : $\mathbf{A}^T = (a_{j,i})$.

Définition (Matrice adjointe). Soit une matrice $\mathbf{A} = (a_{i,j})$ de taille $m \times n$. La matrice adjointe (ou matrice transconjuguée) de \mathbf{A} , notée \mathbf{A}^\dagger , est la matrice de taille $n \times m$ pour laquelle on a pris le complexe conjugué de tous les éléments de \mathbf{A} et on a intervertit les lignes et les colonnes de \mathbf{A} :

$$\mathbf{A}^\dagger = \begin{pmatrix} \overline{a_{1,1}} & \overline{a_{2,1}} & \cdots & \overline{a_{n,1}} \\ \overline{a_{1,2}} & \overline{a_{2,2}} & \cdots & \overline{a_{n,2}} \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1,m}} & \overline{a_{2,m}} & \cdots & \overline{a_{n,m}} \end{pmatrix}.$$

En notation compacte, on a : $\mathbf{A}^\dagger = (\overline{a_{j,i}})$.

Pour une matrice \mathbf{A} dont tous les éléments sont des nombres réels, la matrice adjointe de \mathbf{A} est identique à la matrice transposée de \mathbf{A} .

1.4 Quelques définitions pour le cas des matrices carrées

Dans cette partie, on s'intéresse au cas importants des matrices carrées, c'est-à-dire ayant le même nombre de lignes et de colonnes.

Définition (Matrice identité). La matrice identité, notée \mathbf{I} , de taille $n \times n$ est la matrice carrée dont les éléments diagonaux sont tous égaux à 1 et les éléments non-diagonaux sont tous égaux à 0 :

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}.$$

La matrice identité \mathbf{I} joue le même rôle pour les matrices carrées que le nombre 1 joue pour les nombres. Si \mathbf{A} est une matrice carrée et \mathbf{I} est la matrice identité de même taille, on a : $\mathbf{I} \mathbf{A} = \mathbf{A} \mathbf{I} = \mathbf{A}$.

Définition (Matrice inverse). Soit une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$. On dit que \mathbf{A} est inversible s'il existe une matrice \mathbf{A}^{-1} telle que :

$$\mathbf{A} \mathbf{A}^{-1} = \mathbf{A}^{-1} \mathbf{A} = \mathbf{I}.$$

\mathbf{A}^{-1} s'appelle la matrice inverse de \mathbf{A} .

Attention, toutes les matrices ne sont pas inversibles. Une matrice non-inversible est dite singulière.

On peut facilement vérifier que, pour deux matrices carrées de même taille inversibles \mathbf{A} et \mathbf{B} , on a $(\mathbf{A}\mathbf{B})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.

Revenons à notre système d'équations linéaires écrit sous forme matricielle : $\mathbf{A}\mathbf{X} = \mathbf{B}$. Si la matrice \mathbf{A} est inversible, alors en multipliant l'équation matricielle à gauche par \mathbf{A}^{-1} , on obtient directement la solution (unique) du système : $\mathbf{X} = \mathbf{A}^{-1}\mathbf{B}$. La solution \mathbf{X} s'obtient donc en effectuant le produit matriciel de \mathbf{A}^{-1} par \mathbf{B} . La difficulté vient du calcul de l'inverse \mathbf{A}^{-1} . Nous verrons plus loin comment calculer \mathbf{A}^{-1} .

Définition (Trace d'une matrice). Soit une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$. On appelle trace de \mathbf{A} la somme de ses éléments diagonaux :

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{i,i}.$$

Il est facile de voir que la trace de la matrice transposée de \mathbf{A} est égale à la trace de \mathbf{A} : $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T)$. Par ailleurs, on peut montrer que, pour deux matrices carrées \mathbf{A} et \mathbf{B} de même taille $n \times n$, on a toujours $\text{tr}(\mathbf{A}\mathbf{B}) = \text{tr}(\mathbf{B}\mathbf{A})$ même quand $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$.

Finissons cette section par quelques définitions de certains types de matrices carrées que l'on rencontre fréquemment.

Définition (Matrice symétrique, hermitienne, orthogonale ou unitaire). Soit une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$.

- On dit que \mathbf{A} est une matrice symétrique si \mathbf{A} est égale à sa matrice transposée, c'est-à-dire $\mathbf{A} = \mathbf{A}^T$.
- On dit que \mathbf{A} est une matrice hermitienne si \mathbf{A} est égale à sa matrice adjointe, c'est-à-dire $\mathbf{A} = \mathbf{A}^\dagger$.
- On dit que \mathbf{A} est une matrice orthogonale si \mathbf{A} est inversible et son inverse est égale à sa matrice transposée, c'est-à-dire $\mathbf{A}^{-1} = \mathbf{A}^T$.
- On dit que \mathbf{A} est une matrice unitaire si \mathbf{A} est inversible et son inverse est égale à sa matrice adjointe, c'est-à-dire $\mathbf{A}^{-1} = \mathbf{A}^\dagger$.

Remarques :

- Pour une matrice dont tous les éléments sont des nombres réels, la propriété d'être symétrique est identique à être hermitienne, et la propriété d'être orthogonale est identique à être unitaire.
- Les éléments diagonaux d'une matrice hermitienne sont toujours des nombres réels (partie imaginaire nulle).
- Les vecteurs colonnes qui composent une matrice orthogonale ou unitaire sont orthonormés.

1.5 Déterminant d'une matrice carrée

Considérons une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$. Le déterminant de \mathbf{A} est un nombre que l'on calcule à partir des éléments de \mathbf{A} et que l'on note :

$$\det(\mathbf{A}) = \begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{vmatrix}.$$

Il s'appelle ainsi car il permet de déterminer en vérifiant qu'il est non-nul qu'un système d'équations linéaires admet une solution unique. Plutôt que donner une définition trop abstraite du déterminant, nous allons donner les formules explicites permettant de calculer le déterminant d'une matrice de taille 2×2 ou 3×3 , et comment calculer de manière récursive le déterminant de n'importe quelle matrice de taille $n \times n$.

Pour une matrice de taille 2×2 , le déterminant est :

$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1} a_{2,2} - a_{2,1} a_{1,2}.$$

Pour une matrice de taille 3×3 , le déterminant est :

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = a_{1,1} a_{2,2} a_{3,3} + a_{3,1} a_{1,2} a_{2,3} + a_{2,1} a_{3,2} a_{1,3} \\ - a_{3,1} a_{2,2} a_{1,3} - a_{2,1} a_{1,2} a_{3,3} - a_{1,1} a_{3,2} a_{2,3}.$$

Développement de Laplace d'un déterminant (optionnel)

Pour une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$ quelconque, on peut calculer le déterminant de façon récursive en utilisant le développement de Laplace par rapport à n'importe quelle ligne i

$$\det(\mathbf{A}) = \sum_{j=1}^n a_{i,j} C_{i,j},$$

ou par rapport à n'importe quelle colonne j

$$\det(\mathbf{A}) = \sum_{i=1}^n a_{i,j} C_{i,j},$$

où $C_{i,j}$ s'appelle le cofacteur de l'élément $a_{i,j}$ et est défini par

$$C_{i,j} = (-1)^{i+j} M_{i,j},$$

et $M_{i,j}$ s'appelle le mineur pour l'élément (i, j) de la matrice \mathbf{A} et est défini comme le déterminant de la sous-matrice carrée de taille $(n-1) \times (n-1)$ obtenue en supprimant

la ligne i et la colonne j de la matrice \mathbf{A}

$$M_{i,j} = \begin{vmatrix} a_{1,1} & \cdots & a_{1,j-1} & a_{1,j+1} & \cdots & a_{1,n} \\ a_{2,1} & \cdots & a_{2,j-1} & a_{2,j+1} & \cdots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{i-1,1} & \cdots & a_{i-1,j-1} & a_{i-1,j+1} & \cdots & a_{i-1,n} \\ a_{i+1,1} & \cdots & a_{i+1,j-1} & a_{i+1,j+1} & \cdots & a_{i+1,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & \cdots & a_{n,j-1} & a_{n,j+1} & \cdots & a_{n,n} \end{vmatrix}.$$

Le développement de Laplace d'un déterminant de taille $n \times n$ s'écrit donc avec des déterminants plus petits de taille $(n-1) \times (n-1)$. On peut si besoin réappliquer des développements de Laplace pour écrire ces déterminants $(n-1) \times (n-1)$ avec des déterminants plus petits de taille $(n-2) \times (n-2)$, et ainsi de suite, jusqu'à ce qu'on arrive à des déterminants de taille 2×2 ou 3×3 , que l'on sait calculer explicitement.

Par exemple, le développement de Laplace d'un déterminant de taille 4×4 par rapport à la première ligne est :

$$\begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{vmatrix} = a_{1,1} \begin{vmatrix} a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,2} & a_{4,3} & a_{4,4} \end{vmatrix} - a_{1,2} \begin{vmatrix} a_{2,1} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,3} & a_{4,4} \end{vmatrix} \\ + a_{1,3} \begin{vmatrix} a_{2,1} & a_{2,2} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,4} \end{vmatrix} - a_{1,4} \begin{vmatrix} a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \\ a_{4,1} & a_{4,2} & a_{4,3} \end{vmatrix}.$$

Exemple. (*Système de réactions chimiques*). Reprenons l'exemple du système de réactions chimiques associé à la matrice

$$\mathbf{A} = \begin{pmatrix} -1 & 1 & -1 \\ -1 & -2 & 0 \\ 1 & 1 & 2 \end{pmatrix}.$$

Son déterminant est

$$\det(\mathbf{A}) = \begin{vmatrix} -1 & 1 & -1 \\ -1 & -2 & 0 \\ 1 & 1 & 2 \end{vmatrix} = 5.$$

Le déterminant a les propriétés suivantes : $\det(\mathbf{A}) = \det(\mathbf{A}^T)$ et $\det(\mathbf{A}\mathbf{B}) = \det(\mathbf{B}\mathbf{A}) = \det(\mathbf{A})\det(\mathbf{B})$ où \mathbf{A} et \mathbf{B} sont deux matrices carrées de même taille.

1.6 Calcul de l'inverse d'une matrice carrée

Théorème (Calcul de l'inverse d'une matrice carrée). Soit une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$. \mathbf{A} est inversible si son déterminant est non-nul, $\det(\mathbf{A}) \neq 0$, et sa matrice inverse peut être calculée par :

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} (\text{Com}\mathbf{A})^T,$$

où $\text{Com}\mathbf{A}$ est la comatrice de \mathbf{A} , c'est-à-dire la matrice contenant les cofacteurs $C_{i,j} = (-1)^{i+j} M_{i,j}$ avec $M_{i,j}$ le mineur pour l'élément (i, j) de \mathbf{A} .

Par exemple, pour une matrice de taille 2×2 ,

$$\mathbf{A} = \begin{pmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{pmatrix}.$$

si $\det(\mathbf{A}) = a_{1,1}a_{2,2} - a_{2,1}a_{1,2} \neq 0$, alors \mathbf{A} est inversible. Sa comatrice est

$$\text{Com}\mathbf{A} = \begin{pmatrix} a_{2,2} & -a_{2,1} \\ -a_{1,2} & a_{1,1} \end{pmatrix},$$

et l'inverse de \mathbf{A} est donc

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{pmatrix} a_{2,2} & -a_{1,2} \\ -a_{2,1} & a_{1,1} \end{pmatrix}.$$

Exemple. (*Système de réactions chimiques*). Revenons à l'exemple du système de réactions chimiques. Le déterminant de la matrice est non-nul, $\det(\mathbf{A}) = 5$, donc \mathbf{A} est inversible et le système d'équations linéaires admet une solution unique. La comatrice de \mathbf{A} est

$$\text{Com}\mathbf{A} = \begin{pmatrix} -4 & 2 & 1 \\ -3 & -1 & 2 \\ -2 & 1 & 3 \end{pmatrix},$$

et l'inverse de \mathbf{A} est donc

$$\mathbf{A}^{-1} = \frac{1}{5} \begin{pmatrix} -4 & -3 & -2 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{pmatrix}.$$

Cela permet d'obtenir la solution du système d'équations linéaires :

$$\begin{aligned} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} = \mathbf{X} = \mathbf{A}^{-1} \mathbf{B} &= \frac{1}{5} \begin{pmatrix} -4 & -3 & -2 \\ 2 & -1 & 1 \\ 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} \Delta n_{R_1} \\ \Delta n_{R_2} \\ \Delta n_{R_3} \end{pmatrix} \\ &= \frac{1}{5} \begin{pmatrix} -4\Delta n_{R_1} - 3\Delta n_{R_2} - 2\Delta n_{R_3} \\ 2\Delta n_{R_1} - \Delta n_{R_2} - \Delta n_{R_3} \\ \Delta n_{R_1} + 2\Delta n_{R_2} + 3\Delta n_{R_3} \end{pmatrix}. \end{aligned}$$

2 Valeurs propres et vecteurs propres d'une matrice

Dans cette partie, nous abordons le concept très important de valeurs propres et vecteurs propres d'une matrice carrée.

2.1 Définition et calculs des valeurs propres et vecteurs propres

Définition (Valeur propre et vecteur propre d'une matrice). Soit une matrice carrée \mathbf{A} de taille $n \times n$. On dit qu'un vecteur colonne \mathbf{V} non nul ($\mathbf{V} \neq \mathbf{0}$) à n composantes est un vecteur propre de la matrice \mathbf{A} associée à la valeur propre λ (nombre réel ou complexe) si

$$\mathbf{A} \mathbf{V} = \lambda \mathbf{V}.$$

L'ensemble de toutes les valeurs propres de \mathbf{A} s'appelle le spectre de \mathbf{A} .

En d'autres termes, si on fait le produit matriciel d'une matrice carrée \mathbf{A} par un vecteur propre \mathbf{V} de \mathbf{A} , on obtient le vecteur colonne $\lambda \mathbf{V}$ qui est simplement proportionnel à \mathbf{V} , le coefficient de proportionnalité étant la valeur propre associée λ . Un

vecteur propre de \mathbf{A} est donc un vecteur colonne très particulier pour la matrice \mathbf{A} . En général, si on fait le produit matriciel de \mathbf{A} par un vecteur colonne quelconque \mathbf{X} , on obtient un autre vecteur colonne $\mathbf{X}' = \mathbf{A} \mathbf{X}$ qui n'a pas de raison d'être proportionnel à \mathbf{X} .

Notons que si \mathbf{V} est un vecteur propre de la matrice \mathbf{A} avec la valeur propre λ , alors $c \mathbf{V}$ (où c est n'importe quel nombre) est aussi vecteur propre de la matrice \mathbf{A} avec la valeur propre λ . Un vecteur propre n'est donc défini qu'à une constante multiplicative près.

L'équation $\mathbf{A} \mathbf{V} = \lambda \mathbf{V}$ s'appelle une équation aux valeurs propres. Habituellement, on connaît la matrice \mathbf{A} et on cherche les vecteurs propres et les valeurs propres. C'est un type d'équation que l'on rencontre souvent en sciences, et en particulier en chimie quantique.

Exemple. (*Modèle de Hückel de l'éthylène*). Dans le modèle de Hückel de l'éthylène ($\text{H}_2\text{C}=\text{CH}_2$), chaque orbitale moléculaire de type π est écrite comme combinaison linéaire des orbitales p_z des deux atomes de carbone :

$$\psi_\pi = c_1 \psi_{p_{z,1}} + c_2 \psi_{p_{z,2}},$$

et les coefficients sont déterminés à partir de l'équation aux valeurs propres

$$\mathbf{H} \mathbf{C} = E \mathbf{C}.$$

On arrive à cette équation aux valeurs propres en utilisant l'équation de Schrödinger et en effectuant quelques approximations. Dans cette équation, \mathbf{H} est la matrice dite hamiltonienne

$$\mathbf{H} = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix},$$

où α et β sont des nombres fixés négatifs représentant, respectivement, l'énergie d'une orbitale p_z isolée et l'énergie d'interaction entre les deux orbitales $p_{z,1}$ et $p_{z,2}$. Chaque vecteur propre \mathbf{C} contient les coefficients d'une orbitale moléculaire π

$$\mathbf{C} = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix},$$

et la valeur propre associée E correspondant à l'énergie de cette orbitale.

L'équation aux valeurs propres $\mathbf{A} \mathbf{V} = \lambda \mathbf{V}$ peut être réécrite sous la forme

$$(\mathbf{A} - \lambda \mathbf{I}) \mathbf{V} = \mathbf{0},$$

où \mathbf{I} est la matrice identité de même taille que la matrice \mathbf{A} . Si la matrice $\mathbf{A} - \lambda \mathbf{I}$ est inversible, alors la solution unique de ce système matriciel est le vecteur nul : $\mathbf{V} = (\mathbf{A} - \lambda \mathbf{I})^{-1} \mathbf{0} = \mathbf{0}$. Mais, par définition, le vecteur nul n'est jamais un vecteur

propre. Donc la seule possibilité pour avoir un vecteur non nul solution de cette équation et de trouver un nombre λ tel que la matrice $\mathbf{A} - \lambda \mathbf{I}$ ne soit pas inversible, c'est-à-dire $\det(\mathbf{A} - \lambda \mathbf{I}) = 0$. Ceci donne une équation en λ permettant de calculer toutes les valeurs propres de la matrice \mathbf{A} .

Théorème (Calculs des valeurs propres d'une matrice). Soit une matrice carrée $\mathbf{A} = (a_{i,j})$ de taille $n \times n$. Les valeurs propres λ peuvent être calculées en résolvant l'équation :

$$\det(\mathbf{A} - \lambda \mathbf{I}) = \begin{vmatrix} a_{1,1} - \lambda & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} - \lambda & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} - \lambda \end{vmatrix} = 0.$$

La quantité $\det(\mathbf{A} - \lambda \mathbf{I})$ est un polynôme en λ de degré n qui s'appelle le polynôme caractéristique de la matrice \mathbf{A} . Les valeurs propres sont les racines de ce polynôme. Une matrice \mathbf{A} de taille $n \times n$ à n valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_n$, qui peuvent être réelles ou complexes, et pas forcément toutes différentes.

Une fois que l'on a trouvé les valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_n$, on peut alors calculer les vecteurs propres. Par exemple, pour déterminer un vecteur propre \mathbf{V}_i pour la valeur propre λ_i , on résout le système d'équations linéaires donné par :

$$(\mathbf{A} - \lambda_i \mathbf{I}) \mathbf{V}_i = \mathbf{0}.$$

Puisque si λ_i est une valeur propre, la matrice de ce système $(\mathbf{A} - \lambda_i \mathbf{I})$ n'est pas inversible, il n'y a de solution unique mais une infinité de solutions. Ceci n'est pas étonnant car nous avons déjà vu qu'un vecteur propre n'est défini qu'à une constante multiplicative près.

Exemple. (Modèle de Hückel de l'éthylène). Calculons les valeurs propres E de la matrice \mathbf{H} pour le modèle de Hückel de l'éthylène. Le polynôme caractéristique de la matrice \mathbf{H} est

$$\det(\mathbf{H} - E \mathbf{I}) = \begin{vmatrix} \alpha - E & \beta \\ \beta & \alpha - E \end{vmatrix} = (\alpha - E)^2 - \beta^2.$$

Les valeurs propres E sont les solutions de l'équation

$$(\alpha - E)^2 - \beta^2 = 0,$$

ce qui donne deux valeurs propres

$$E_1 = \alpha + \beta \text{ et } E_2 = \alpha - \beta.$$

On détermine un vecteur propre \mathbf{C}_1 pour la valeur propre E_1 en résolvant le système

$$\begin{aligned} (\mathbf{H} - E_1 \mathbf{I}) \mathbf{C}_1 = \mathbf{0} &\iff \begin{pmatrix} \alpha - E_1 & \beta \\ \beta & \alpha - E_1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\iff \begin{cases} (\alpha - E_1)c_1 + \beta c_2 = 0 \\ \beta c_1 + (\alpha - E_1)c_2 = 0 \end{cases}, \end{aligned}$$

ce qui devient, en insérant la valeur $E_1 = \alpha + \beta$,

$$\begin{cases} -\beta c_1 + \beta c_2 = 0 \\ \beta c_1 - \beta c_2 = 0 \end{cases}.$$

Il s'agit de deux fois la même équation, à savoir $c_1 = c_2$. Il y a donc une infinité de solutions, par exemple on peut choisir c_1 étant n'importe quel nombre et c_2 doit alors être égal à c_1 . Habituellement, on impose en plus que le vecteur propre soit de norme égale à 1, c'est-à-dire $c_1^2 + c_2^2 = 1$, ce qui conduit par exemple aux coefficients $c_1 = 1/\sqrt{2}$ et $c_2 = 1/\sqrt{2}$.

De façon similaire, on détermine un vecteur propre \mathbf{C}_2 pour la valeur propre E_2 en résolvant le système

$$\begin{aligned} (\mathbf{H} - E_2 \mathbf{I}) \mathbf{C}_2 = \mathbf{0} &\iff \begin{pmatrix} \alpha - E_2 & \beta \\ \beta & \alpha - E_2 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ &\iff \begin{cases} (\alpha - E_2)c_1 + \beta c_2 = 0 \\ \beta c_1 + (\alpha - E_2)c_2 = 0 \end{cases}, \end{aligned}$$

ce qui devient, en insérant la valeur $E_2 = \alpha - \beta$,

$$\begin{cases} \beta c_1 + \beta c_2 = 0 \\ \beta c_1 + \beta c_2 = 0 \end{cases}.$$

On obtient deux fois la même équation, à savoir $c_1 = -c_2$. Il y a donc une infinité de solutions, par exemple on peut choisir c_1 étant n'importe quel nombre et c_2 doit alors être égal à $-c_1$. En imposant que le vecteur propre soit de norme égale à 1, c'est-à-dire $c_1^2 + c_2^2 = 1$, on peut choisir par exemple les coefficients $c_1 = 1/\sqrt{2}$ et $c_2 = -1/\sqrt{2}$.

En conclusion, on a trouvé deux orbitales moléculaires π . La première orbitale moléculaire a une énergie plus basse que l'énergie d'une orbitale p_z isolée, $E_1 = \alpha + \beta < \alpha$ (car β est négatif) et a pour expression :

$$\psi_\pi = \frac{1}{\sqrt{2}} (\psi_{p_{z,1}} + \psi_{p_{z,2}}).$$

Il s'agit de l'orbitale π liante. La deuxième orbitale moléculaire a une énergie plus haute que l'énergie d'une orbitale p_z isolée, $E_2 = \alpha - \beta > \alpha$ (car β est négatif) et a pour expression :

$$\psi_{\pi^*} = \frac{1}{\sqrt{2}} (\psi_{p_{z,1}} - \psi_{p_{z,2}}).$$

Il s'agit de l'orbitale π^* antiliante.

2.2 Diagonalisation d'une matrice

Définition (Matrice diagonalisable). Une matrice carrée \mathbf{A} de taille $n \times n$ est dite diagonalisable si \mathbf{A} possède n vecteurs propres linéairement indépendants (c'est-à-dire qu'aucun vecteur propre ne peut s'exprimer comme combinaison linéaire des autres).

Attention, toutes les matrices ne sont pas diagonalisables.

Une matrice diagonalisable peut être exprimée en fonction d'une matrice diagonale comme le montre le théorème suivant.

Théorème (Diagonalisation d'une matrice). Soit une matrice carrée \mathbf{A} de taille $n \times n$ diagonalisable. On peut écrire \mathbf{A} sous la forme

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1},$$

où \mathbf{D} est la matrice diagonale de taille $n \times n$ contenant sur la diagonale les valeurs propres de \mathbf{A} (et des 0 partout ailleurs)

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

et \mathbf{P} est une matrice inversible de taille $n \times n$ contenant sur ses colonnes n vecteurs propres $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n$ de \mathbf{A} correspondant aux valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_n$

$$\mathbf{P} = \begin{pmatrix} | & | & & | \\ | & | & \cdots & | \\ \mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_n \\ | & | & & | \end{pmatrix}.$$

On dit que la matrice \mathbf{A} est *semblable* à la matrice diagonale \mathbf{D} , et \mathbf{P} s'appelle la *matrice de passage*.

L'avantage d'écrire la matrice \mathbf{A} sous la forme $\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1}$ et que cela permet d'exploiter la simplicité de la matrice diagonale \mathbf{D} pour faire des calculs avec la matrice \mathbf{A} . Par

exemple, la trace de \mathbf{A} peut s'écrire comme

$$\begin{aligned}\operatorname{tr}(\mathbf{A}) &= \operatorname{tr}(\mathbf{P} \mathbf{D} \mathbf{P}^{-1}) \\ &= \operatorname{tr}(\mathbf{D} \mathbf{P}^{-1} \mathbf{P}) \\ &= \operatorname{tr}(\mathbf{D} \mathbf{I}) \\ &= \operatorname{tr}(\mathbf{D}) \\ &= \sum_{i=1}^n \lambda_i,\end{aligned}$$

où nous avons utilisé que $\operatorname{tr}(\mathbf{P}(\mathbf{D} \mathbf{P}^{-1})) = \operatorname{tr}((\mathbf{D} \mathbf{P}^{-1}) \mathbf{P})$. La trace de \mathbf{A} est donc égal à la trace de \mathbf{D} qui est la somme des valeurs propres. De même, le déterminant de \mathbf{A} devient

$$\begin{aligned}\det(\mathbf{A}) &= \det(\mathbf{P} \mathbf{D} \mathbf{P}^{-1}) \\ &= \det(\mathbf{D} \mathbf{P}^{-1} \mathbf{P}) \\ &= \det(\mathbf{D} \mathbf{I}) \\ &= \det(\mathbf{D}) \\ &= \lambda_1 \lambda_2 \cdots \lambda_n,\end{aligned}$$

où nous avons utilisé que $\det(\mathbf{P}(\mathbf{D} \mathbf{P}^{-1})) = \det((\mathbf{D} \mathbf{P}^{-1}) \mathbf{P})$ et que $\mathbf{P}^{-1} \mathbf{P} = \mathbf{I}$. Le déterminant de \mathbf{A} est donc égal au déterminant de \mathbf{D} qui est le produit des valeurs propres. On voit donc que, si toutes les valeurs propres de \mathbf{A} sont non nulles, alors le déterminant \mathbf{A} est non nul et la matrice \mathbf{A} est inversible. Dans ce cas, l'inverse de \mathbf{A} peut s'écrire

$$\begin{aligned}\mathbf{A}^{-1} &= (\mathbf{P} \mathbf{D} \mathbf{P}^{-1})^{-1} \\ &= \mathbf{P} \mathbf{D}^{-1} \mathbf{P}^{-1},\end{aligned}$$

où \mathbf{D}^{-1} est la matrice diagonale contenant sur la diagonale les inverses des valeurs propres

$$\mathbf{D}^{-1} = \begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1/\lambda_n \end{pmatrix}.$$

Par ailleurs, on peut calculer la puissance $k^{\text{ème}}$ de \mathbf{A} , où k est un entier, comme

$$\begin{aligned}\mathbf{A}^k &= \underbrace{\mathbf{A} \mathbf{A} \cdots \mathbf{A}}_{k \text{ fois}} \\ &= \mathbf{P} \mathbf{D} \mathbf{P}^{-1} \mathbf{P} \mathbf{D} \mathbf{P}^{-1} \cdots \mathbf{P} \mathbf{D} \mathbf{P}^{-1} \\ &= \mathbf{P} \underbrace{\mathbf{D} \mathbf{D} \cdots \mathbf{D}}_{k \text{ fois}} \mathbf{P}^{-1} \\ &= \mathbf{P} \mathbf{D}^k \mathbf{P}^{-1},\end{aligned}$$

où la puissance $k^{\text{ème}}$ de \mathbf{D} est simplement la matrice diagonale contenant sur la diagonale les puissances $k^{\text{ème}}$ des valeurs propres

$$\mathbf{D}^k = \begin{pmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{pmatrix}.$$

Pour k grand, cette expression pour \mathbf{A}^k est très pratique car elle évite de devoir calculer un grand nombre de produits matriciels $\mathbf{A} \mathbf{A} \cdots \mathbf{A}$. Plus généralement, pour une fonction $f : x \mapsto f(x)$ d'une variable, suffisamment régulière, on peut définir la matrice $f(\mathbf{A})$ comme

$$f(\mathbf{A}) = \mathbf{P} f(\mathbf{D}) \mathbf{P}^{-1},$$

où $f(\mathbf{D})$ est la matrice diagonale contenant sur la diagonale $f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)$

$$f(\mathbf{D}) = \begin{pmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & f(\lambda_n) \end{pmatrix}.$$

Il est donc très utile de savoir si une matrice est diagonalisable. Le théorème suivant donne un critère simple.

Théorème (Théorème spectral). Soit une matrice carrée \mathbf{A} de taille $n \times n$. Si \mathbf{A} est une matrice hermitienne, c'est-à-dire égale à sa matrice adjointe ($\mathbf{A} = \mathbf{A}^\dagger$), alors \mathbf{A} est diagonalisable. De plus, toutes les valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_n$ sont des nombres réels (partie imaginaire nulle) et on peut toujours trouver des vecteurs propres orthonormés de sorte que la matrice de passage \mathbf{P} soit une matrice unitaire, c'est-à-dire son inverse est égale à sa matrice adjointe ($\mathbf{P}^{-1} = \mathbf{P}^\dagger$). On a donc

$$\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^\dagger.$$

Pour une matrice \mathbf{A} dont tous les éléments sont des nombres réels, rappelons que la condition d'être hermitienne est identique à être symétrique ($\mathbf{A} = \mathbf{A}^T$). La matrice de passage \mathbf{P} peut alors être choisie comme une matrice orthogonale ($\mathbf{P}^{-1} = \mathbf{P}^T$) et on a : $\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^T$.

Ce théorème est extrêmement utile car on a très souvent à faire des matrices hermitiennes dans les applications.

Exemple. (*Chimie quantique*). La matrice hamiltonienne \mathbf{H} du modèle de Hückel (de n'importe quelle molécule) est une matrice hermitienne (car réelle symétrique). Cela garantit que les valeurs propres correspondant à l'énergie des orbitales soient des nombres réels et qu'on puisse toujours trouver des vecteurs propres orthonormés correspondant aux orbitales moléculaires. Cela dépasse en fait le simple modèle de Hückel et une caractéristique centrale de la mécanique quantique.

3 Applications linéaires

3.1 Représentation matricielle des applications linéaires

Une très grande utilité des matrices vient du fait qu'elles représentent des applications linéaires.

Définition (Application linéaire de \mathbb{R}^n dans \mathbb{R}^m). Une application (ou fonction) linéaire \vec{f} de \mathbb{R}^n dans \mathbb{R}^m est une fonction vectorielle

$$\begin{aligned} \vec{f}: \mathbb{R}^n &\rightarrow \mathbb{R}^m \\ \vec{x} &\mapsto \vec{f}(\vec{x}) \end{aligned}$$

satisfaisant la propriété de linéarité

$$\vec{f}(c\vec{x} + d\vec{y}) = c\vec{f}(\vec{x}) + d\vec{f}(\vec{y}),$$

pour tous $\vec{x}, \vec{y} \in \mathbb{R}^n$ et $c, d \in \mathbb{R}$.

Si on se donne une base de vecteurs de \mathbb{R}^n

$$B = (\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n),$$

alors on peut décomposer n'importe quel vecteur $\vec{x} \in \mathbb{R}^n$ dans la base B

$$\vec{x} = x_1\vec{e}_1 + x_2\vec{e}_2 + \dots + x_n\vec{e}_n,$$

où les nombres réels x_1, x_2, \dots, x_n sont les composantes du vecteur \vec{x} dans la base B . On dira que le vecteur \vec{x} est représenté dans la base B par le vecteur colonne \mathbf{X} de composantes x_1, x_2, \dots, x_n et on écrira

$$\vec{x} \equiv \mathbf{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Inversement, à tout vecteur colonne \mathbf{X} à n composantes réelles, après avoir choisi une base B , on peut associer un vecteur $\vec{x} \in \mathbb{R}^n$. De manière similaire, si on se donne une base de vecteurs de \mathbb{R}^m

$$B' = (\vec{e}'_1, \vec{e}'_2, \dots, \vec{e}'_m)$$

alors on peut décomposer le vecteur $\vec{x}' = \vec{f}(\vec{x}) \in \mathbb{R}^m$ dans la base B'

$$\vec{x}' = x'_1\vec{e}'_1 + x'_2\vec{e}'_2 + \dots + x'_m\vec{e}'_m,$$

et le vecteur \vec{x}' est alors représenté dans la base B' par le vecteur colonne \mathbf{X}' de composantes x'_1, x'_2, \dots, x'_m

$$\vec{x}' \equiv \mathbf{X}' = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{pmatrix}.$$

Les vecteurs sont donc représentés dans une base par des vecteurs colonnes (ou matrices à une colonne). Le théorème suivant montre que les applications linéaires peuvent être représentées par des matrices.

Théorème (Représentation matricielle d'une application linéaire). Soit \vec{f} une application linéaire de \mathbb{R}^n dans \mathbb{R}^m , \vec{x} un vecteur de \mathbb{R}^n représenté par le vecteur colonne \mathbf{X} dans une base B , et \vec{x}' un vecteur de \mathbb{R}^m représenté par le vecteur colonne \mathbf{X}' dans une base B' . L'équation vectorielle $\vec{x}' = \vec{f}(\vec{x})$ peut être écrite sous forme matricielle :

$$\vec{x}' = \vec{f}(\vec{x}) \iff \mathbf{X}' = \mathbf{A} \mathbf{X},$$

où \mathbf{A} est une matrice (unique) de taille $m \times n$ représentant l'application linéaire \vec{f} dans les bases B et B' .

Inversement, à toute matrice \mathbf{A} de taille $m \times n$, après avoir choisi les bases B et B' , on peut associer une unique application linéaire \vec{f} de \mathbb{R}^n dans \mathbb{R}^m . Dans le cas où $m = n$, une application linéaire de \mathbb{R}^n dans \mathbb{R}^n peut être représentée dans une base $B = (\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n)$ de \mathbb{R}^n par une matrice carrée \mathbf{A} de taille $n \times n$.

Cette équivalence entre matrices et applications linéaires signifie que tous les concepts définis pour les matrices peuvent aussi être définis pour les applications linéaires, comme par exemple le concept de valeur propre et de vecteur propre.

Définition (Valeur propre et vecteur propre d'une application linéaire). Soit \vec{f} une application linéaire de \mathbb{R}^n dans \mathbb{R}^n . On dit qu'un vecteur $\vec{v} \in \mathbb{R}^n$ non nul ($\vec{v} \neq \vec{0}$) est un vecteur propre de \vec{f} associée à la valeur propre λ (nombre réel ou complexe) si

$$\vec{f}(\vec{v}) = \lambda \vec{v}.$$

Si, dans une base $B = (\vec{e}_1, \vec{e}_2, \dots, \vec{e}_n)$ de \mathbb{R}^n , on représente l'application linéaire \vec{f} par la matrice carrée \mathbf{A} et le vecteur \vec{v} par le vecteur colonne \mathbf{V} , on a alors l'équivalence :

$$\vec{f}(\vec{v}) = \lambda \vec{v} \iff \mathbf{A} \mathbf{V} = \lambda \mathbf{V}.$$

Si \mathbf{A} est diagonalisable alors on peut écrire $\mathbf{A} = \mathbf{P} \mathbf{D} \mathbf{P}^{-1}$ où \mathbf{D} est la matrice diagonale des valeurs propres $\lambda_1, \lambda_2, \dots, \lambda_n$ et \mathbf{P} est la matrice de passage contenant sur ses

colonnes des vecteurs propres $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n$ de la matrice \mathbf{A} . On peut alors écrire :

$$\begin{aligned} \vec{x}' = \vec{f}(\vec{x}) &\iff \mathbf{X}' = \mathbf{A} \mathbf{X} \\ &\iff \mathbf{X}' = \mathbf{P} \mathbf{D} \mathbf{P}^{-1} \mathbf{X} \\ &\iff \mathbf{P}^{-1} \mathbf{X}' = \mathbf{D} \mathbf{P}^{-1} \mathbf{X} \\ &\iff \mathbf{Y}' = \mathbf{D} \mathbf{Y}. \end{aligned}$$

Les vecteurs colonne $\mathbf{Y} = \mathbf{P}^{-1} \mathbf{X}$ et $\mathbf{Y}' = \mathbf{P}^{-1} \mathbf{X}'$ représentent les vecteurs \vec{x} et \vec{x}' dans la nouvelle base $B_p = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n)$ de \mathbb{R}^n composée des vecteurs propres de \vec{f} représentés par $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_n$. Dans cette nouvelle base B_p , l'application linéaire \vec{f} est représentée par la matrice diagonale \mathbf{D} . Il est donc beaucoup plus facile de calculer l'action de \vec{f} sur un vecteur si on se place dans la base de vecteurs propres de \vec{f} . Il est important de noter que les matrices \mathbf{A} et \mathbf{D} représentent la même application linéaire \vec{f} mais dans des bases différentes. On dit que l'on a effectué un changement de base de l'ancienne base B à la nouvelle base B_p . La matrice \mathbf{P}^{-1} assure le changement de B à B_p , alors que la matrice \mathbf{P} assure le changement inverse de B_p à B . L'action de l'application linéaire \vec{f} est beaucoup plus simple à calculer

3.2 Opérations de symétrie d'une molécule

Un exemple d'applications linéaires est donné par les opérations de symétrie d'une molécule. Ce sont des transformations géométriques de \mathbb{R}^3 dans \mathbb{R}^3 qui laissent inchangée la molécule. Il y a cinq types d'opérations de symétrie dite ponctuelles que l'on rencontre dans les molécules : l'identité, les rotations, les réflexions, l'inversion et les rotations impropres. Nous allons donner les représentations matricielles dans la base canonique $B = (\vec{i}, \vec{j}, \vec{k})$ de \mathbb{R}^3 de quelques exemples d'opérations de symétrie.

Pour cela, on fixe un repère $(O, \vec{i}, \vec{j}, \vec{k})$ et un point M , et on regarde comment le vecteur \overrightarrow{OM} se transforme par l'opération de symétrie \vec{f}

$$\overrightarrow{OM} = x\vec{i} + y\vec{j} + z\vec{k} \longrightarrow \vec{f}(\overrightarrow{OM}) = \overrightarrow{OM}' = x'\vec{i} + y'\vec{j} + z'\vec{k}.$$

Identité

L'identité est la transformation géométrique qui laisse échangé le point M :

$$\begin{cases} x' = x \\ y' = y \\ z' = z \end{cases}.$$

En représentation matricielle, on a donc :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

Rotations

Une rotation est une transformation géométrique définie par un axe et un angle. Par exemple, pour la rotation d'axe (Oz) et d'angle θ , les coordonnées du point M' s'expriment en fonction de celles du point M par :

$$\begin{cases} x' = (\cos \theta)x - (\sin \theta)y \\ y' = (\sin \theta)x + (\cos \theta)y \\ z' = z \end{cases} .$$

En représentation matricielle, on a donc :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

Réflexions

Une réflexion est une transformation géométrique correspondant à une symétrie par rapport à un plan. Par exemple, pour la réflexion par rapport au plan (Oxy) , les coordonnées du point M' sont :

$$\begin{cases} x' = x \\ y' = y \\ z' = -z \end{cases} .$$

En représentation matricielle, on a donc :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

Inversion

L'inversion correspond à l'opération de symétrie par rapport au centre du repère :

$$\begin{cases} x' = -x \\ y' = -y \\ z' = -z \end{cases} .$$

En représentation matricielle, on a donc :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

Rotations impropres (ou roto-inversions)

Une rotation impropre est une transformation géométrique correspondant à une rotation suivie d'une réflexion par rapport au plan perpendiculaire à l'axe de rotation.

Par exemple, pour la rotation impropre d'axe (Oz) et d'angle θ , les coordonnées du point M' sont :

$$\begin{cases} x' = (\cos \theta)x - (\sin \theta)y \\ y' = (\sin \theta)x + (\cos \theta)y \\ z' = -z \end{cases} .$$

En représentation matricielle, on a donc :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} .$$

Exemple. (*Opérations de symétrie de la molécule d'eau*). La molécule d'eau dans sa géométrie d'équilibre a quatre opérations de symétrie : l'identité, la rotation d'axe Oz (axe bissecteur de l'angle de liaison) et d'angle π , la réflexion par rapport au plan (Oxz) (plan de la molécule) et la réflexion par rapport au plan (Oyz) (plan bissecteur de la molécule).

Troisième partie

Des données aux modèles

Chapitre III.1

Modèles de régression

En sciences, il arrive très souvent d'avoir des données et de vouloir les représenter au mieux par une fonction. On utilise pour faire ça des méthodes de régression. Dans ce chapitre, nous allons expliquer la méthode de régression linéaire pour approcher les données par une droite et, plus généralement, la méthode de régression polynomiale pour approcher les données par un polynôme.

1 Idée générale des méthodes de régression

Supposons que nous disposions de données sous la forme de N couples de points

$$(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_N, y_N),$$

issus de mesures expérimentales ou de calculs compliqués. Souvent x est une variable que l'on fixe, et y est une variable que l'on mesure. Pour fixer les idées, prenons l'exemple suivant avec $N = 11$ couples de points :

x	0	1	2	3	4	5	6	7	8	9	10
y	10	8	15	15	30	40	44	60	80	85	110

Ces données sont tracées sur la figure III.1.1.

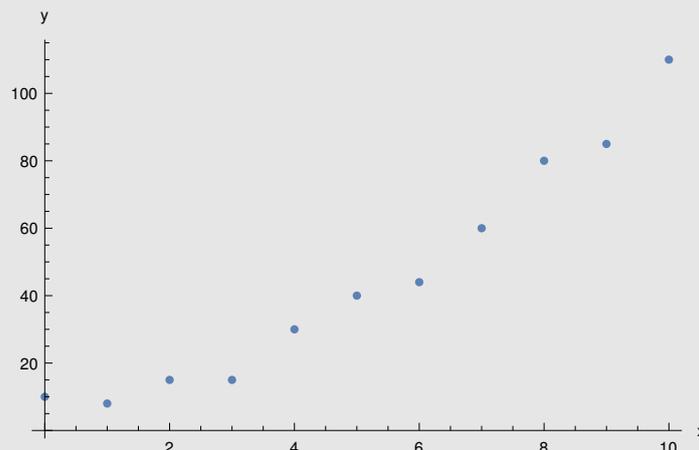


FIGURE III.1.1 – Représentation graphique des données.

Dans le problème de la régression, on cherche à trouver une fonction $f : x \mapsto f(x)$, si possible simple, qui permet d'approcher au mieux les données

$$y \approx f(x).$$

On parle aussi d'ajustement de courbe ou, en anglais, de « fitting ».

Approcher les données par une fonction a deux intérêts :

- On détermine ainsi la relation (approximative) existant entre les variables x et y , c'est-à-dire une loi empirique qui permet de rationaliser les données.
- On est alors capable de prédire la valeur de y pour une valeur de x pour laquelle on ne dispose pas de données, ou inversement.

Nous traiterons uniquement les cas simples où l'on choisit pour f :

- une fonction affine (une droite) $f(x) = ax + b$ où a et b sont les paramètres à déterminer ;
- un polynôme de plus haut degré, comme par exemple une parabole $f(x) = ax^2 + bx + c$ où a , b et c sont les paramètres à déterminer.

Les méthodes de régression sont néanmoins beaucoup plus générales que ça. On peut faire des régressions avec des fonctions beaucoup plus compliquées. Par exemple, on peut utiliser une fonction très compliquée de plusieurs variables donnée sous la forme d'un réseau de neurones artificiels (codé dans un ordinateur), et qui contient alors beaucoup de paramètres à déterminer. Il faut déterminer tous ces paramètres avec beaucoup de données. C'est le domaine du « machine learning » ou apprentissage automatique, qui est une branche de l'intelligence artificielle. Ce domaine se développe actuellement rapidement. En chimie par exemple, avec ces techniques et beaucoup de données, on peut essayer d'« apprendre » la fonction reliant la géométrie d'une molécule (natures et positions des atomes) à une propriété physicochimie donnée (par exemple, son enthalpie de formation) afin de pouvoir faire des prédictions rapides sur des nouvelles molécules.

2 Régression linéaire

2.1 Description de la méthode de régression linéaire

Dans la régression linéaire, on choisit d'approcher les données par une droite, c'est-à-dire une fonction $f : x \mapsto f(x)$ affine

$$f(x) = ax + b,$$

avec a et b deux paramètres réels à déterminer. La méthode la plus courante pour déterminer ces paramètres est la méthode des moindres carrés.

Pour une valeur x_i dans les données, la valeur de y associée prédite par le modèle est $f(x_i) = ax_i + b$ alors que la valeur réelle dans les données est y_i . Pour ce point, l'erreur commise sur y est donc $(f(x_i) - y_i)$. Dans la méthode des moindres carrés, on choisit les paramètres a et b qui minimisent la somme des carrés de ces erreurs

$$F(a, b) = \sum_{i=1}^N (f(x_i) - y_i)^2.$$

On prend les carrés des erreurs car on n'est pas intéressé par les signes de ces erreurs mais uniquement par leurs valeurs absolues. Il s'agit donc d'un problème d'optimisation (sans contraintes) : celui de chercher le point (a_0, b_0) correspondant au minimum de la fonction

$$F : \mathbb{R}^2 \rightarrow \mathbb{R}$$

$$(a, b) \mapsto F(a, b) = \sum_{i=1}^N (a x_i + b - y_i)^2.$$

2.2 Détermination des paramètres a_0 et b_0

On cherche donc (a_0, b_0) comme un point stationnaire de F , c'est-à-dire un point où les dérivées partielles premières de F sont nulles

$$\begin{cases} \frac{\partial F}{\partial a}(a_0, b_0) = 0 \\ \frac{\partial F}{\partial b}(a_0, b_0) = 0, \end{cases}$$

ce qui donne, après calcul des dérivées partielles,

$$\begin{cases} \sum_{i=1}^N 2x_i (a_0 x_i + b_0 - y_i) = 0 \\ \sum_{i=1}^N 2(a_0 x_i + b_0 - y_i) = 0. \end{cases}$$

Il s'agit d'un système de deux équations à deux inconnues (a_0 et b_0). Écrivons le plus clairement en factorisant par a_0 et b_0

$$\begin{cases} \left(\sum_{i=1}^N x_i^2 \right) a_0 + \left(\sum_{i=1}^N x_i \right) b_0 = \sum_{i=1}^N x_i y_i \\ \left(\sum_{i=1}^N x_i \right) a_0 + \left(\sum_{i=1}^N 1 \right) b_0 = \sum_{i=1}^N y_i, \end{cases}$$

En utilisant $\sum_{i=1}^N 1 = N$ et en introduisant les quantités suivantes, que l'on peut calculer facilement à partir des données,

$$S_x = \sum_{i=1}^N x_i, \quad S_y = \sum_{i=1}^N y_i, \quad S_{xx} = \sum_{i=1}^N x_i^2, \quad S_{xy} = \sum_{i=1}^N x_i y_i,$$

on peut mettre finalement le système sous la forme compacte

$$\begin{cases} S_{xx} a_0 + S_x b_0 = S_{xy} \\ S_x a_0 + N b_0 = S_y. \end{cases}$$

Utilisons par exemple la méthode matricielle pour résoudre ce système d'équations. Sous forme matricielle, le système d'équations se réécrit comme

$$\begin{pmatrix} S_{xx} & S_x \\ S_x & N \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \end{pmatrix} = \begin{pmatrix} S_{xy} \\ S_y \end{pmatrix}.$$

Le déterminant de la matrice 2×2 du système d'équations vaut $S_{xx}N - S_x^2$. Si celui-ci est non nul, on peut introduire l'inverse de la matrice, ce qui permet de résoudre le système d'équations sous la forme

$$\begin{pmatrix} a_0 \\ b_0 \end{pmatrix} = \begin{pmatrix} S_{xx} & S_x \\ S_x & N \end{pmatrix}^{-1} \begin{pmatrix} S_{xy} \\ S_y \end{pmatrix},$$

où l'inverse de la matrice 2×2 est

$$\begin{pmatrix} S_{xx} & S_x \\ S_x & N \end{pmatrix}^{-1} = \frac{1}{S_{xx}N - S_x^2} \begin{pmatrix} N & -S_x \\ -S_x & S_{xx} \end{pmatrix}.$$

On arrive donc aux expressions de a_0 et b_0 :

$$\begin{cases} a_0 = \frac{NS_{xy} - S_x S_y}{S_{xx}N - S_x^2} \\ b_0 = \frac{-S_x S_{xy} + S_{xx} S_y}{S_{xx}N - S_x^2}. \end{cases}$$

On pourrait vérifier en calculant les dérivées partielles secondes de F (mais on ne le fera pas) que ce point stationnaire (a_0, b_0) correspond bien à un minimum de la fonction F .

Appliquons ces expressions à l'exemple de la partie 1. On a $N = 11$ et on calcule $S_x = 55$, $S_y = 497$, $S_{xx} = 385$, $S_{xy} = 3592$. Cela conduit à $a_0 = 1107/110 \approx 10.064$ et $b_0 = -113/22 \approx -5.136$. On a donc trouvé par régression linéaire l'expression de la droite approchant au mieux les données :

$$f(x) = 10.064 x - 5.136.$$

Cette droite est tracée sur la figure III.1.2.

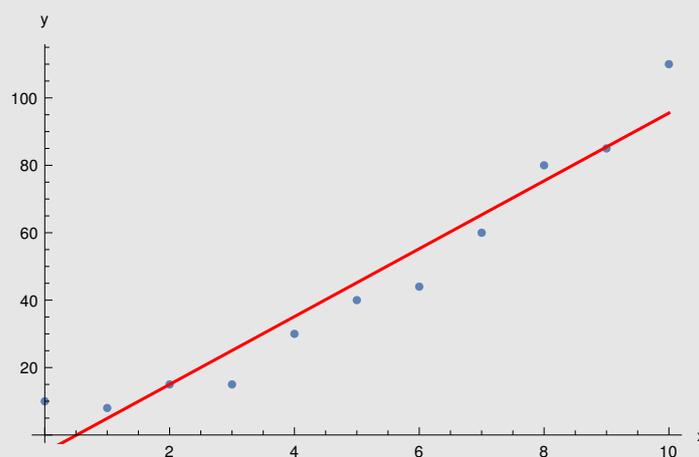


FIGURE III.1.2 – Droite déterminée par régression linéaire approchant au mieux les données.

2.3 Coefficient de détermination R^2

Pour mesurer la qualité de l'approximation des données par la fonction f de régression, on utilise très souvent le coefficient de détermination, noté R^2 , et défini par

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{\sum_{i=1}^N (y_i - \bar{y})^2},$$

où \bar{y} est la moyenne des valeurs y_i définie par

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i.$$

Le coefficient de détermination R^2 est toujours compris entre 0 et 1. Dans la définition de R^2 , le terme $\sum_{i=1}^N (y_i - f(x_i))^2 / \sum_{i=1}^N (y_i - \bar{y})^2$ est le ratio entre la variance des erreurs sur y_i due au modèle de régression et la variance totale des valeurs y_i .

- Si ce terme est petit, cela signifie que le modèle explique bien les variations des valeurs y_i , et alors R^2 est proche de 1. La qualité de la régression est bonne. Dans la limite idéale où les erreurs sur y_i sont toutes nulles (c'est-à-dire que la fonction f passe exactement par tous les points) alors on a la valeur maximale $R^2 = 1$.
- Si ce terme est grand, cela signifie que le modèle n'explique pas bien les variations des valeurs y_i , et alors R^2 est éloigné de 1. La qualité de la régression est mauvaise, et on aurait intérêt à essayer de modéliser les données avec une autre fonction f .

Une explication graphique du coefficient de détermination R^2 est donnée dans la figure III.1.3.

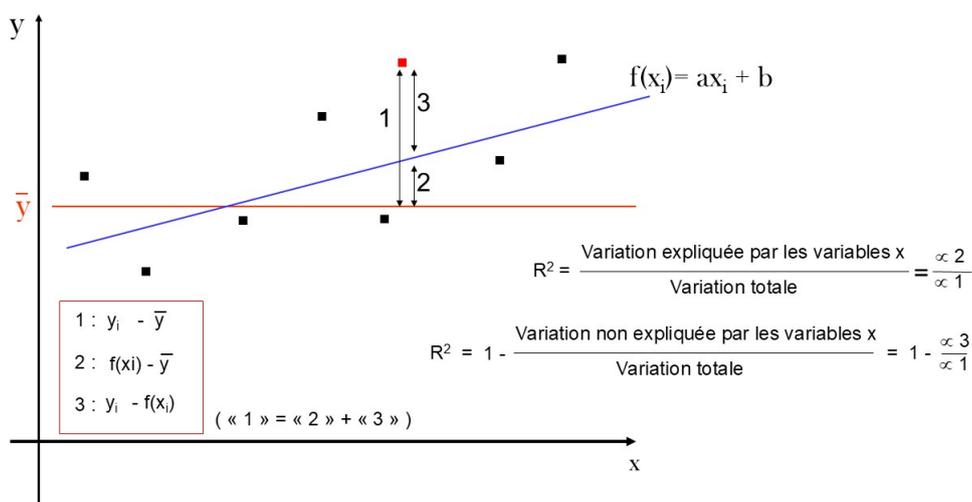


FIGURE III.1.3 – Explication graphique du coefficient de détermination R^2 .

Pour l'exemple de la régression linéaire effectuée dans la partie précédente, on calcule $\bar{y} \approx 45.182$, $\sum_{i=1}^N (y_i - f(x_i))^2 \approx 779.191$ et $\sum_{i=1}^N (y_i - \bar{y})^2 \approx 11919.6$, ce qui donne un coefficient de détermination de $R^2 \approx 0.935$. Il s'agit d'une régression de qualité acceptable mais pas exceptionnelle.

2.4 Application à la cinétique chimique

À 773 K, le cyclopropane se transforme en propène. La mesure de la concentration en cyclopropane c en fonction du temps t donne les résultats suivants :

t (s)	0	300	600	900	1200	1800	2400	3000
c (mmol/L)	1.70	1.20	0.90	0.73	0.68	0.46	0.31	0.25

Si on fait l'hypothèse que cette réaction a une cinétique d'ordre 1 avec une constante de vitesse k , on devrait avoir la loi d'évolution de la concentration suivante :

$$-\frac{dc(t)}{dt} = k c(t) \implies \ln c(t) = -k t + \ln c(0).$$

Pour vérifier cette hypothèse, on peut tracer $\ln c$ en fonction de t et effectuer une régression linéaire.

Le calcul de $\ln c$ donne :

t (s)	0	300	600	900	1200	1800	2400	3000
$\ln c$	0.5306	0.1823	-0.1054	-0.3147	-0.3857	-0.7765	-1.1712	-1.3862

Utilisons les mêmes notations utilisées que précédemment : $x = t$ et $y = \ln c$. On veut déterminer par régression linéaire la fonction affine $f(x) = a_0 x + b_0$ approchant au mieux

les données : $y \approx f(x)$. Le nombre de points est $N = 8$. On calcule $S_x = 10200$, $S_y \approx -3.4268$, $S_{xx} = 20700000$, $S_{xy} \approx -9122.0285$. Les paramètres a_0 et b_0 recherchés sont déterminés par le système d'équations

$$\begin{cases} 20700000 a_0 + 10200 b_0 = -9122.0285 \\ 10200 a_0 + 8 b_0 = -3.4268. \end{cases}$$

Cela donne $a_0 = -0.0006177$ et $b_0 = 0.3592$. Revenant maintenant aux notations du problème, la meilleure droite approchant au mieux les données est donc :

$$\ln c = -0.0006177 t + 0.3592.$$

On a donc une estimation de la constante de vitesse $k \approx 0.0006177 \text{ s}^{-1}$ et du logarithme de la concentration initiale $\ln c(0) \approx 0.3592$, c'est-à-dire $c(0) \approx 1.4321 \text{ mmol/L}$.

La droite de régression est comparée aux données sur la figure III.1.4.

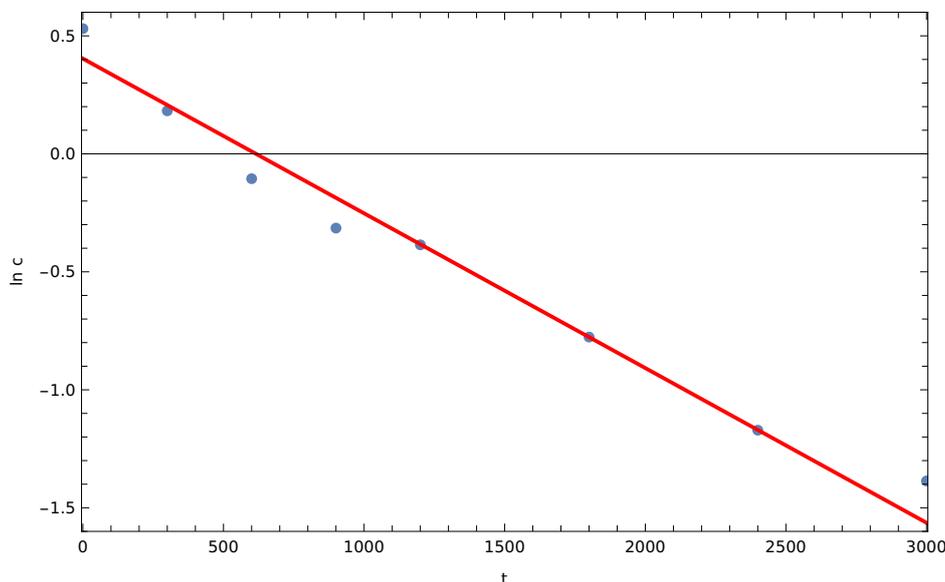


FIGURE III.1.4 – Droite de régression $\ln c = -0.0006177 t + 0.3592$ comparée aux données.

Graphiquement, la régression apparaît comme étant assez bonne. Pour confirmer cela de manière plus quantitative, calculons à présent le coefficient de détermination R^2 . On calcule $\bar{y} \approx -0.4283$, $\sum_{i=1}^N (y_i - f(x_i))^2 \approx 0.0779$ et $\sum_{i=1}^N (y_i - \bar{y})^2 \approx 3.0023$, ce qui donne $R^2 \approx 0.9741$. Il s'agit bien d'une régression de bonne qualité.

On conclut que les données sont en accord avec un modèle cinétique d'ordre 1.

3 Régression polynomiale

3.1 Description de la méthode de régression polynomiale de degré 2

Dans la régression polynomiale, on choisit d'approcher les données par un polynôme d'un certain degré. Nous traiterons ici le cas d'un polynôme de degré 2, c'est-à-dire une parabole. La fonction $f : x \mapsto f(x)$ est alors

$$f(x) = ax^2 + bx + c,$$

avec a , b et c trois paramètres réels à déterminer.

Nous utilisons ici aussi la méthode des moindres carrés, c'est-à-dire que l'on choisit les paramètres a , b et c qui minimisent la somme des carrés des erreurs entre le modèle et les données. On cherche donc le point (a_0, b_0, c_0) correspondant au minimum de la fonction de trois variables

$$F : \mathbb{R}^3 \rightarrow \mathbb{R}$$

$$(a, b, c) \mapsto F(a, b, c) = \sum_{i=1}^N (a x_i^2 + b x_i + c - y_i)^2.$$

3.2 Détermination des paramètres a_0 , b_0 et c_0

On cherche donc (a_0, b_0, c_0) comme un point stationnaire de F , c'est-à-dire un point où les dérivées partielles premières de F sont nulles

$$\begin{cases} \frac{\partial F}{\partial a}(a_0, b_0, c_0) = 0 \\ \frac{\partial F}{\partial b}(a_0, b_0, c_0) = 0 \\ \frac{\partial F}{\partial c}(a_0, b_0, c_0) = 0, \end{cases}$$

ce qui donne, après calcul des dérivées partielles,

$$\begin{cases} \sum_{i=1}^N 2x_i^2 (a_0 x_i^2 + b_0 x_i + c_0 - y_i) = 0 \\ \sum_{i=1}^N 2x_i (a_0 x_i^2 + b_0 x_i + c_0 - y_i) = 0 \\ \sum_{i=1}^N 2 (a_0 x_i^2 + b_0 x_i + c_0 - y_i) = 0. \end{cases}$$

Il s'agit d'un système de trois équations à trois inconnues (a_0 , b_0 et c_0). Écrivons le plus clairement en factorisant par a_0 , b_0 et c_0

$$\left\{ \begin{array}{l} \left(\sum_{i=1}^N x_i^4 \right) a_0 + \left(\sum_{i=1}^N x_i^3 \right) b_0 + \left(\sum_{i=1}^N x_i^2 \right) c_0 = \sum_{i=1}^N x_i^2 y_i \\ \left(\sum_{i=1}^N x_i^3 \right) a_0 + \left(\sum_{i=1}^N x_i^2 \right) b_0 + \left(\sum_{i=1}^N x_i \right) c_0 = \sum_{i=1}^N x_i y_i \\ \left(\sum_{i=1}^N x_i^2 \right) a_0 + \left(\sum_{i=1}^N x_i \right) b_0 + \left(\sum_{i=1}^N 1 \right) c_0 = \sum_{i=1}^N y_i, \end{array} \right.$$

En utilisant $\sum_{i=1}^N 1 = N$ et en introduisant les quantités suivantes, que l'on peut calculer facilement à partir des données,

$$\begin{array}{l} S_x = \sum_{i=1}^N x_i, \quad S_y = \sum_{i=1}^N y_i, \quad S_{x^2} = \sum_{i=1}^N x_i^2, \quad S_{xy} = \sum_{i=1}^N x_i y_i, \\ S_{x^3} = \sum_{i=1}^N x_i^3, \quad S_{x^2 y} = \sum_{i=1}^N x_i^2 y_i, \quad S_{x^4} = \sum_{i=1}^N x_i^4, \end{array}$$

on peut mettre finalement le système sous la forme compacte

$$\left\{ \begin{array}{l} S_{x^4} a_0 + S_{x^3} b_0 + S_{x^2} c_0 = S_{x^2 y} \\ S_{x^3} a_0 + S_{x^2} b_0 + S_x c_0 = S_{xy} \\ S_{x^2} a_0 + S_x b_0 + N c_0 = S_y. \end{array} \right.$$

Utilisons par exemple la méthode matricielle pour résoudre ce système d'équations. Sous forme matricielle, le système d'équations se réécrit comme

$$\begin{pmatrix} S_{x^4} & S_{x^3} & S_{x^2} \\ S_{x^3} & S_{x^2} & S_x \\ S_{x^2} & S_x & N \end{pmatrix} \begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} S_{x^2 y} \\ S_{xy} \\ S_y \end{pmatrix}.$$

Si le déterminant de la matrice 3×3 est non nul, on peut introduire l'inverse de la matrice, ce qui permet de résoudre le système d'équations sous la forme

$$\begin{pmatrix} a_0 \\ b_0 \\ c_0 \end{pmatrix} = \begin{pmatrix} S_{x^4} & S_{x^3} & S_{x^2} \\ S_{x^3} & S_{x^2} & S_x \\ S_{x^2} & S_x & N \end{pmatrix}^{-1} \begin{pmatrix} S_{x^2 y} \\ S_{xy} \\ S_y \end{pmatrix},$$

où l'inverse de la matrice peut être calculé (de préférence avec l'aide d'un ordinateur !) pour chaque application après avoir remplacé les expressions par des valeurs numériques.

Encore une fois, on pourrait vérifier en calculant les dérivées partielles secondes de F que ce point stationnaire (a_0, b_0, c_0) correspond bien à un minimum de la fonction F .

Appliquons ces expressions à l'exemple de la partie 1. On a $N = 11$ et on calcule $S_x = 55$, $S_y = 497$, $S_{x^2} = 385$, $S_{xy} = 3592$, $S_{x^3} = 3025$, $S_{x^2y} = 29212$, $S_{x^4} = 25333$. Cela conduit à $a_0 = 249/286 \approx 0.871$, $b_0 = 1941/1430 \approx 1.357$ et $c_0 = 103/13 \approx 7.923$. On a donc trouvé par régression polynomiale l'expression de la parabole approchant au mieux les données :

$$f(x) = 0.871 x^2 + 1.357 x + 7.923.$$

Cette parabole est tracée sur la figure III.1.5.

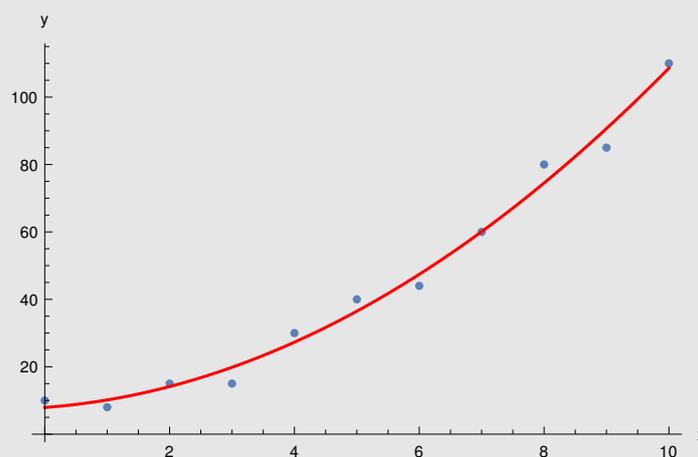


FIGURE III.1.5 – Parabole déterminée par régression polynomiale approchant au mieux les données.

Pour mesurer la qualité de la régression, on peut calculer le coefficient de détermination R^2 (la définition reste la même que celle introduite plus haut). On trouve $R^2 \approx 0.989$. Cette valeur beaucoup plus proche de 1 que celle obtenue avec la régression linéaire confirme qu'une fonction polynomiale de degré 2 est un bien meilleur modèle pour nos données.

Chapitre III.2

Modèles physicochimiques

Un des buts de la science en général, et de la physicochimie en particulier, est d'arriver à décrire les phénomènes observés à l'aide de modèles mathématiques, qui permettent de rationaliser et de prédire les observations. Très souvent, ces modèles prennent la forme d'équations différentielles impliquant des fonctions dépendant du temps et de l'espace. Ces équations différentielles décrivent alors des relations simples existantes entre les variations temporelles et/ou spatiales de ces fonctions. On cherche alors à résoudre ces équations différentielles pour faire des prédictions. Dans ce chapitre, nous allons voir plusieurs modèles différentiels avec des applications à la cinétique chimique, à la spectroscopie, aux phénomènes de diffusion et à la chimie quantique.

1 Modèles différentiels d'ordre 1

Nous nous limiterons aux équations différentielles linéaires.

1.1 Équation différentielle linéaire d'ordre 1 sans second membre

Définition (Équation différentielle linéaire d'ordre 1 sans second membre). Soit $I \subseteq \mathbb{R}$ un intervalle et $a, b : I \rightarrow \mathbb{R}$ deux fonctions continues, avec a ne s'annulant pas sur I . On appelle équation différentielle d'ordre 1 sans second membre (ou homogène) l'équation de la forme :

$$a(t)y'(t) + b(t)y(t) = 0, \quad (\text{III.2.1})$$

où l'inconnue est une fonction $y : t \mapsto y(t)$ dérivable sur I .

Remarques :

- Il s'agit d'une équation différentielle *linéaire* car elle ne contient que des termes proportionnels à $y(t)$ et à sa dérivée. Une équation différentielle qui contient par

exemple un terme en $y(t)^2$ est appelée *non linéaire*. En général, les équations différentielles non linéaires sont beaucoup plus difficiles à résoudre.

- Il s'agit d'une équation différentielle *d'ordre 1* car elle ne contient que la dérivée première y' de y .

On suppose que la fonction y ne s'annule pas sur I . On peut déterminer la solution générale en écrivant l'équation différentielle sous la forme

$$a(t) \frac{dy(t)}{dt} + b(t)y(t) = 0 \Leftrightarrow \frac{1}{y(t)} \frac{dy(t)}{dt} = -\frac{b(t)}{a(t)} \Leftrightarrow \frac{d \ln |y(t)|}{dt} = -\frac{b(t)}{a(t)},$$

où l'on a utilisé que la dérivée de $\ln |y|$ est $(\ln |y|)' = y'/y$. En passant aux primitives, on a :

$$\ln |y(t)| = -\int \frac{b(t)}{a(t)} dt + K,$$

où $\int b(t)/a(t) dt$ est une primitive de b/a et $K \in \mathbb{R}$ est une constante arbitraire. En prenant l'exponentielle de chacun des membres de cette équation, on arrive à :

$$|y(t)| = e^{-\int \frac{b(t)}{a(t)} dt + K} = e^K e^{-\int \frac{b(t)}{a(t)} dt}.$$

Puisque l'on a supposé que y ne s'annule pas, y est soit toujours positif soit toujours négatif, et on a donc :

$$y(t) = \pm e^K e^{-\int \frac{b(t)}{a(t)} dt} = C e^{-\int \frac{b(t)}{a(t)} dt},$$

où $C = \pm e^K$ est une constante arbitraire.

On peut montrer, même sans supposer que y ne s'annule pas, que toutes les solutions de l'équation différentielle ont cette forme, ce qui conduit au théorème suivant.

Théorème (Solution générale d'une équation différentielle linéaire d'ordre 1 sans second membre). La solution générale de l'équation différentielle linéaire d'ordre 1 sans second membre (III.2.1) est la fonction $t \mapsto y(t)$ de la forme :

$$y(t) = C e^{-\int \frac{b(t)}{a(t)} dt}, \quad (\text{III.2.2})$$

où $C \in \mathbb{R}$ est une constante arbitraire.

En pratique, la constante arbitraire C peut être fixée en imposant une condition initiale (si t représente le temps) ou condition aux limites (si t est une variable d'espace), par exemple $y(0) = y_0$ où y_0 est la valeur supposée connue prise par la fonction y en $t = 0$.

1.2 Modèle cinétique d'une réaction chimique unidirectionnelle $A \longrightarrow B$

Pour une réaction chimique unidirectionnelle $A \longrightarrow B$ de cinétique d'ordre 1 (comme la transformation du cyclopropane en propène), la concentration $[A](t)$ du réactif A au temps t satisfait l'équation différentielle :

$$-\frac{d[A](t)}{dt} = k [A](t),$$

où k est la constante de vitesse de la réaction. Il s'agit d'une équation différentielle linéaire d'ordre 1 sans second membre. La solution générale est :

$$[A](t) = C e^{-kt},$$

où $C \in \mathbb{R}$ est une constante arbitraire. Si on impose la condition initiale $[A](0) = [A]_0$, où $[A]_0$ est la concentration de A à $t = 0$, alors on a $C = [A]_0$, et la solution satisfaisant cette condition initiale est donc :

$$[A](t) = [A]_0 e^{-kt}.$$

1.3 Équation différentielle linéaire d'ordre 1 avec second membre

Définition (Équation différentielle linéaire d'ordre 1 avec second membre). Soit $I \subseteq \mathbb{R}$ un intervalle et $a, b, f : I \rightarrow \mathbb{R}$ trois fonctions continues, avec a ne s'annulant pas sur I . On appelle équation différentielle d'ordre 1 avec second membre (ou inhomogène) l'équation de la forme :

$$a(t)y'(t) + b(t)y(t) = f(t), \quad (\text{III.2.3})$$

où l'inconnue est une fonction $y : t \mapsto y(t)$ dérivable sur I .

Théorème (Solution générale d'une équation différentielle linéaire d'ordre 1 avec second membre). La solution générale de l'équation différentielle linéaire d'ordre 1 avec second membre (III.2.3) est la fonction $t \mapsto y(t)$ de la forme :

$$y(t) = y_{g,ssm}(t) + y_{p,asm}(t),$$

où $y_{g,ssm}$ est la solution générale de l'équation différentielle sans second membre (ssm) et $y_{p,asm}$ est une solution particulière de l'équation différentielle avec second membre (asm).

La solution générale de l'équation sans second membre $y_{g,ssm}$ (c'est-à-dire avec $f = 0$) est la solution donnée dans l'équation (III.2.2) : $y_{g,ssm}(t) = C e^{-\int \frac{b(t)}{a(t)} dt}$ avec $C \in \mathbb{R}$ arbitraire. Une solution particulière de l'équation avec second membre $y_{p,asm}$ est n'importe quelle solution satisfaisant l'équation $a(t)y'_{p,asm}(t) + b(t)y_{p,asm}(t) = f(t)$. Pour déterminer une telle solution particulière $y_{p,asm}$, soit on arrive à en deviner une facilement, soit on peut appliquer la méthode systématique de variation de la constante. Cette dernière méthode consiste à chercher une solution particulière sous la forme $y_{p,asm}(t) = C(t) e^{-\int \frac{b(t)}{a(t)} dt}$ où C est une fonction dérivable. En injectant $y_{p,asm}(t)$ dans l'équation différentielle avec second membre, on détermine alors la fonction C .

1.4 Modèle cinétique d'une réaction chimique bidirectionnelle $A \rightleftharpoons B$

Pour une réaction chimique bidirectionnelle $A \rightleftharpoons B$ de cinétique d'ordre 1, la dérivée de la concentration de A par rapport au temps est donnée par :

$$-\frac{d[A](t)}{dt} = k_+ [A](t) - k_- [B](t),$$

où k_+ est la constante de vitesse dans le sens $A \rightarrow B$ et k_- est la constante de vitesse dans le sens $A \leftarrow B$. Par ailleurs, par conservation de la matière, la concentration totale $c_{tot} = [A](t) + [B](t)$ est indépendante du temps. On voit donc que la concentration de A satisfait l'équation différentielle suivante :

$$-\frac{d[A](t)}{dt} = k_+ [A](t) - k_- (c_{tot} - [A](t)),$$

ou

$$\frac{d[A](t)}{dt} + (k_+ + k_-)[A](t) = c_{tot} k_-.$$

Il s'agit d'une équation différentielle linéaire d'ordre 1 avec second membre. La solution générale est :

$$[A](t) = [A]_{g,ssm}(t) + [A]_{p,asm}(t),$$

où $[A]_{g,ssm}$ est la solution générale de l'équation sans second membre est

$$[A]_{g,ssm}(t) = C e^{-(k_+ + k_-)t},$$

où $C \in \mathbb{R}$ est une constante arbitraire, et $[A]_{p,asm}$ est une solution particulière de l'équation avec second membre. Par exemple, on peut essayer de trouver une solution $[A]_{p,asm}(t)$ ayant une dérivée nulle, $d[A]_{p,asm}(t)/dt = 0$, pour tout t . Cette solution doit donc satisfaire l'équation :

$$(k_+ + k_-)[A]_{p,asm}(t) = c_{tot} k_-,$$

ce qui conduit à

$$[A]_{\text{p,asm}}(t) = \frac{c_{\text{tot}} k_-}{k_+ + k_-},$$

dont la dérivée par rapport au temps est bien nulle. La solution générale de l'équation avec second membre est donc :

$$[A](t) = C e^{-(k_+ + k_-)t} + \frac{c_{\text{tot}} k_-}{k_+ + k_-}.$$

On peut déterminer la constante C en imposant la condition initiale $[A](0) = [A]_0$, ce qui donne

$$[A]_0 = C + \frac{c_{\text{tot}} k_-}{k_+ + k_-} \implies C = [A]_0 - \frac{c_{\text{tot}} k_-}{k_+ + k_-}.$$

2 Modèles différentiels d'ordre 2

Nous nous limiterons aux équations différentielles linéaires d'ordre 2 à coefficients constants.

2.1 Équation différentielle linéaire d'ordre 2 à coefficients constants sans second membre

Définition (Équation différentielle linéaire d'ordre 2 à coefficients constants sans second membre). Soit trois constantes $a, b, c \in \mathbb{R}$, avec $a \neq 0$. On appelle équation différentielle d'ordre 2 à coefficients constants sans second membre (ou homogène) l'équation de la forme :

$$a y''(t) + b y'(t) + c y(t) = 0, \quad (\text{III.2.4})$$

où l'inconnue est une fonction $y : t \mapsto y(t)$ dérivable sur I .

Théorème (Solution générale d'une équation différentielle linéaire d'ordre 2 à coefficients constants sans second membre). L'équation caractéristique associée à l'équation différentielle est :

$$ar^2 + br + c = 0.$$

1. Si l'équation caractéristique a deux racines réelles r_1 et r_2 , alors la solution générale de l'équation différentielle est :

$$y(t) = C_1 e^{r_1 t} + C_2 e^{r_2 t} \text{ avec } C_1, C_2 \in \mathbb{R}.$$

2. Si l'équation caractéristique a une racine double $r = r_1 = r_2$, alors la solution générale de l'équation différentielle est :

$$y(t) = (C_1 + C_2 t) e^{rt} \text{ avec } C_1, C_2 \in \mathbb{R}.$$

3. Si l'équation caractéristique a deux racines complexes conjuguées $r_1 = \alpha + i\beta$ and $r_2 = \alpha - i\beta$, alors la solution générale de l'équation différentielle est :

$$y(t) = (C_1 \cos(\beta t) + C_2 \sin(\beta t)) e^{\alpha t} \text{ avec } C_1, C_2 \in \mathbb{R}.$$

Pour fixer les deux constantes arbitraires C_1 et C_2 , on peut par exemple imposer une condition initiale sur y et sur y' , c'est-à-dire $y(0) = y_0$ et $y'(0) = d_0$ où y_0 et d_0 sont les valeurs prises par y et y' en $t = 0$.

2.2 Modèle de l'oscillateur harmonique

On considère le mouvement de vibration de deux atomes de masses m_1 et m_2 reliés par une liaison chimique. En première approximation, on peut décrire ce phénomène avec la mécanique classique (de Newton) en modélisant la liaison chimique comme un ressort de constante de raideur k . En ce plaçant dans le repère du centre de masse, le principe fondamental de la dynamique (ou deuxième loi de Newton) s'écrit :

$$\mu x''(t) = -kx(t),$$

où $\mu = m_1 m_2 / (m_1 + m_2)$ est la masse réduite et $x(t)$ est la distance entre les deux atomes au temps t . Dans cette équation, $x''(t)$ correspond à l'accélération et $-kx(t)$ est la force de rappel du ressort. On peut réécrire cette équation sous la forme :

$$\mu x''(t) + kx(t) = 0,$$

où l'on voit qu'il s'agit d'une équation différentielle linéaire d'ordre 2 à coefficients constants. L'équation caractéristique associée

$$\mu r^2 + k = 0,$$

a deux racines complexes conjuguées $r_1 = i\omega_0$ et $r_2 = -i\omega_0$ où $\omega_0 = \sqrt{k/\mu}$. La solution générale est donc :

$$x(t) = C_1 \cos(\omega_0 t) + C_2 \sin(\omega_0 t) \quad \text{avec } C_1, C_2 \in \mathbb{R}.$$

Si on choisit par exemple les conditions initiales $x(0) = x_0$ et $x'(0) = 0$, on a :

$$x(0) = C_1 = x_0,$$

et

$$x'(t) = -C_1 \omega_0 \sin(\omega_0 t) + C_2 \omega_0 \cos(\omega_0 t) \implies x'(0) = C_2 \omega_0 = 0 \implies C_2 = 0.$$

On arrive donc à la solution :

$$x(t) = x_0 \cos(\omega_0 t).$$

Il s'agit d'un mouvement oscillatoire de période $T = 2\pi/\omega_0$. La quantité $\nu_0 = 1/T = \omega_0/(2\pi)$ s'appelle la fréquence propre du système. La quantité ω_0 est la pulsation (éviter de la dénommer fréquence, c'est un abus de langage) propre du système. Ce modèle s'appelle le modèle de l'oscillateur harmonique. C'est un modèle célèbre que l'on retrouve dans de très nombreuses situations en physique et chimie.

2.3 Équation différentielle linéaire d'ordre 2 à coefficients constants avec second membre

Définition (Équation différentielle linéaire d'ordre 2 à coefficients constants avec second membre). Soit trois constantes $a, b, c \in \mathbb{R}$, avec $a \neq 0$, et une fonction $f : I \rightarrow \mathbb{R}$ continue sur un intervalle I . On appelle équation différentielle d'ordre 2 à coefficients constants avec second membre (ou inhomogène) l'équation de la forme :

$$a y''(t) + b y'(t) + c y(t) = f(t), \quad (\text{III.2.5})$$

où l'inconnue est une fonction $y : t \mapsto y(t)$ dérivable sur I .

Théorème (Solution générale d'une équation différentielle linéaire d'ordre 2 à coefficients constants avec second membre). La solution générale de l'équation différentielle linéaire d'ordre 2 à coefficients constants avec second membre (III.2.5) est la fonction $t \mapsto y(t)$ de la forme :

$$y(t) = y_{g,ssm}(t) + y_{p,asm}(t),$$

où $y_{g,ssm}$ est la solution générale de l'équation différentielle sans second membre (ssm) et $y_{p,asm}$ est une solution particulière de l'équation différentielle avec second membre (asm).

2.4 Modèle de l'oscillateur harmonique forcé

Reprenons l'exemple de l'oscillateur harmonique et ajoutons une force extérieure oscillatoire $f(t) = F \cos(\omega t)$ où F est l'amplitude de la force et ω est sa pulsation. Le principe fondamental de la dynamique s'écrit maintenant :

$$\mu x''(t) = -kx(t) + f(t),$$

ou

$$\mu x''(t) + kx(t) = f(t).$$

Il s'agit d'une équation différentielle linéaire d'ordre 2 à coefficients constants avec second membre. La solution générale est de la forme :

$$x(t) = x_{g,ssm}(t) + x_{p,asm}(t),$$

où $x_{g,ssm}$ est la solution générale de l'équation sans second membre

$$x_{g,ssm}(t) = C_1 \cos(\omega_0 t) + C_2 \sin(\omega_0 t),$$

avec $\omega_0 = \sqrt{k/\mu}$ et $C_1, C_2 \in \mathbb{R}$, et $x_{p,asm}$ est une solution particulière de l'équation avec second membre. On peut essayer de trouver une solution particulière de la forme $x_{p,asm}(t) = A \cos(\omega t)$ où A est une constante à déterminer. En injectant $x_{p,asm}(t)$ dans l'équation différentielle avec second membre, on a :

$$-\mu A \omega^2 \cos(\omega t) + kA \cos(\omega t) = F \cos(\omega t),$$

ce qui conduit à

$$-\mu A \omega^2 + kA = F \implies A = \frac{F}{k - \mu \omega^2} = \frac{F/\mu}{\omega_0^2 - \omega^2}.$$

La solution générale de l'équation différentielle avec second membre est donc :

$$x(t) = C_1 \cos(\omega_0 t) + C_2 \sin(\omega_0 t) + \frac{F/\mu}{\omega_0^2 - \omega^2} \cos(\omega t).$$

Si on impose les conditions initiales $x(0) = x_0$ et $x'(0) = 0$, on a :

$$x(0) = C_1 + \frac{F/\mu}{\omega_0^2 - \omega^2} = x_0 \implies C_1 = x_0 - \frac{F/\mu}{\omega_0^2 - \omega^2},$$

et

$$x'(t) = -C_1\omega_0 \sin(\omega_0 t) + C_2\omega_0 \cos(\omega_0 t) - \frac{F/\mu}{\omega_0^2 - \omega^2} \omega \sin(\omega t)$$

$$\implies x'(0) = C_2\omega_0 = 0 \implies C_2 = 0.$$

On arrive donc à la solution :

$$x(t) = \left(x_0 - \frac{F/\mu}{\omega_0^2 - \omega^2} \right) \cos(\omega_0 t) + \frac{F/\mu}{\omega_0^2 - \omega^2} \cos(\omega t),$$

ou

$$x(t) = x_0 \cos(\omega_0 t) + \frac{F/\mu}{\omega_0^2 - \omega^2} (\cos(\omega t) - \cos(\omega_0 t)).$$

Le mouvement total est une superposition d'un mouvement oscillatoire de pulsation ω_0 et d'un autre mouvement oscillatoire de pulsation ω .

Étudions ce qui se passe quand la pulsation de l'excitation externe se rapproche de la pulsation propre du système, $\omega \rightarrow \omega_0$. Le développement limité de $\cos(\omega t)$ quand $\omega \rightarrow \omega_0$ à l'ordre 1 en $(\omega - \omega_0)$ est

$$\cos(\omega t) = \cos(\omega_0 t) - t \sin(\omega_0 t)(\omega - \omega_0) + o(\omega - \omega_0),$$

ce qui conduit

$$x(t) = x_0 \cos(\omega_0 t) - \frac{(F/\mu)(\omega - \omega_0)}{\omega_0^2 - \omega^2} t \sin(\omega_0 t) + o(\omega - \omega_0)$$

$$= x_0 \cos(\omega_0 t) + \frac{(F/\mu)}{\omega_0 + \omega} t \sin(\omega_0 t) + o(\omega - \omega_0).$$

Le deuxième terme dépend du temps en $t \sin(\omega_0 t)$. Donc, $x(t)$ a des oscillations dont l'amplitude augmente avec le temps. Dans la limite d'un grand temps $t \rightarrow \infty$, l'amplitude diverge : $|x(t)| \rightarrow \infty$. Il s'agit du phénomène de résonance : lorsque l'on excite un oscillateur avec une fréquence égale à sa fréquence propre, on obtient des oscillations de grande amplitude. Dans les systèmes mécaniques réels, l'amplitude des oscillations ne va pas jusqu'à l'infini grâce à des phénomènes de dissipation de l'énergie que nous n'avons pas inclus dans le modèle.

Pour le cas d'une vibration d'une liaison chimique entre deux atomes, on peut penser la force externe comme provenant d'un champ électrique oscillant de pulsation ω . Quand ω est égale à la pulsation propre ω_0 de vibration de la liaison, il y a effectivement un phénomène de résonance. On montre en mécanique quantique que cela correspond à l'absorption d'un photon (de pulsation ω) qui augmente le niveau de vibration de la liaison. Ce phénomène est la base de la spectroscopie infrarouge.

3 Modèles différentiels aux dérivées partielles

Les modèles différentiels précédents impliquaient uniquement des fonctions d'une seule variable. En pratique, en science, on a souvent des fonctions de plusieurs variables et les modèles correspondants impliquent les dérivées partielles de ces fonctions. Nous allons voir deux modèles différentiels aux dérivées partielles importants pour la phy-

sicochimie : l'équation de la diffusion et l'équation de Schrödinger.

3.1 Équation de la diffusion

Le phénomène de diffusion se réfère à la tendance d'un grand ensemble de particules de se déplacer des régions de grandes concentrations vers les régions de basses concentrations (ce qui est une conséquence de la deuxième loi de la thermodynamique). Par exemple, il peut s'agir de la diffusion de molécules de parfum dans l'air ou de la diffusion d'ions dans l'eau.

Si le nombre de particules est suffisamment grand, alors la densité volumique $\rho(x, y, z, t)$ de particules et le vecteur de densité de courant de particules $\vec{j}(x, y, z, t)$ au point (x, y, z) et au temps t suit deux lois établies par Adolf Eugene Fick en 1855 :

$$\begin{aligned} \text{Première loi de Fick : } \vec{j}(x, y, z, t) &= -D \vec{\nabla} \rho(x, y, z, t), \\ \text{Deuxième loi de Fick : } -\vec{\nabla} \cdot \vec{j}(x, y, z, t) &= \frac{\partial \rho(x, y, z, t)}{\partial t}, \end{aligned}$$

où D est une constante appelée coefficient de diffusion. Le coefficient de diffusion D dépend de la nature des particules qui diffusent, du milieu dans lequel se fait la diffusion, et de la température. Le vecteur de densité de courant de particules $\vec{j}(x, y, z, t)$ correspond au produit de la densité $\rho(x, y, z, t)$ de particules avec le vecteur vitesse moyenne des particules \vec{V} . Le flux de particules Φ_S qui traverse une surface S par unité de temps est égal à $\vec{j}(x, y, z, t) \cdot \vec{n}$ où \vec{n} est le vecteur unitaire normal à la surface S .

La combinaison des deux lois de Fick aboutit à un modèle différentiel aux dérivées partielles : l'équation de la diffusion.

Définition (Équation de la diffusion). Soit une fonction $\rho : (x, y, z, t) \mapsto \rho(x, y, z, t)$ des coordonnées d'espace x, y, z et du temps t , représentant la densité d'un grand ensemble de particules. L'équation de la diffusion est :

$$\frac{\partial \rho}{\partial t}(x, y, z, t) = D \Delta \rho(x, y, z, t),$$

Dans l'équation de la diffusion, rappelons que Δ est l'opérateur laplacien :

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}.$$

Écrite explicitement avec toutes les dérivées partielles, l'équation de la diffusion prend la forme :

$$\frac{\partial \rho}{\partial t}(x, y, z, t) = D \left(\frac{\partial^2 \rho}{\partial x^2}(x, y, z, t) + \frac{\partial^2 \rho}{\partial y^2}(x, y, z, t) + \frac{\partial^2 \rho}{\partial z^2}(x, y, z, t) \right).$$

Connaissant une densité de particules initiale, $\rho(x, y, z, 0)$, la résolution de l'équation de la diffusion permet de prédire la densité $\rho(x, y, z, t)$ à un temps ultérieur. Cependant, on ne sait pas résoudre exactement l'équation de la diffusion pour n'importe quelle densité initiale.

Pour simplifier, considérons la diffusion de particules suivant uniquement l'axe x . On peut oublier les coordonnées y et z , et l'équation de diffusion devient :

$$\frac{\partial \rho}{\partial t}(x, t) = D \frac{\partial^2 \rho}{\partial x^2}(x, t).$$

Supposons que la densité initiale en $t = 0$ soit donnée par une fonction gaussienne centrée en $x = 0$ et d'écart-type σ_0 :

$$\rho(x, 0) = \frac{N}{\sqrt{2\pi\sigma_0^2}} e^{-x^2/(2\sigma_0^2)},$$

où N est le nombre de particules et le préfacteur a été choisi de sorte que $\int_{-\infty}^{\infty} \rho(x, 0) dx = N$. Après diffusion, la densité au temps t est encore donnée par une fonction gaussienne,

$$\rho(x, t) = \frac{N}{\sqrt{2\pi\sigma(t)^2}} e^{-x^2/(2\sigma(t)^2)},$$

mais avec un nouveau écart-type $\sigma(t)$ donné par

$$\sigma(t) = \sqrt{\sigma_0^2 + 2Dt}.$$

L'intégrale de la densité au temps t donne toujours le même nombre de particules : $\int_{-\infty}^{\infty} \rho(x, t) dx = N$. Comme exercice, nous invitons le lecteur de calculer les dérivées partielles $\partial \rho(x, t) / \partial t$ et $\partial^2 \rho(x, t) / \partial x^2$ et de vérifier que $\rho(x, t)$ satisfait bien l'équation de la diffusion. On voit que l'écart-type $\sigma(t)$, donnant une mesure de l'étalement spatial de la densité en particules, croît avec le temps en \sqrt{t} . L'évolution temporelle de la densité lors de la diffusion est illustrée sur la Figure III.2.1.

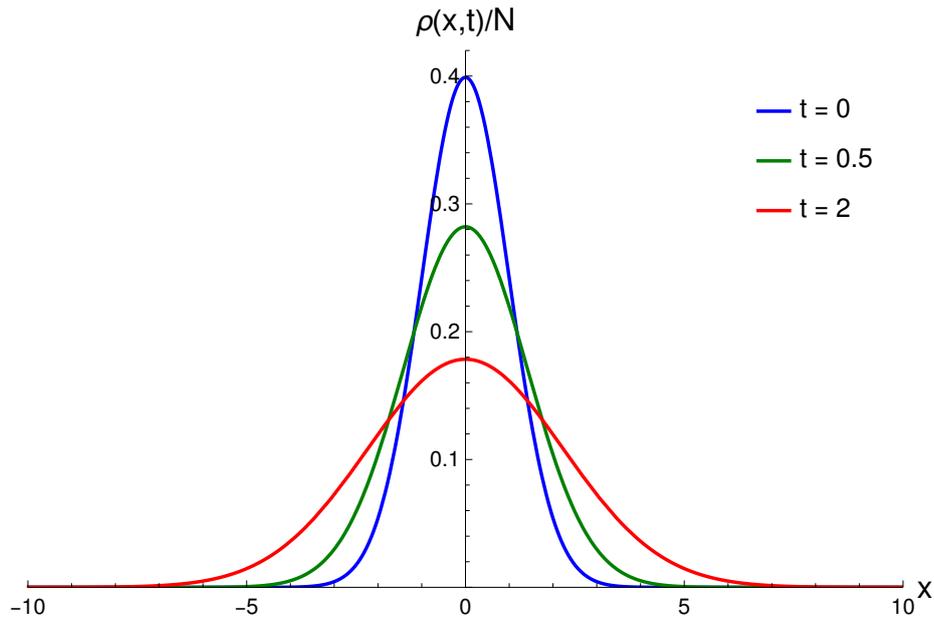


FIGURE III.2.1 – Densité gaussienne de particules $\rho(x,t)/N = (1/\sqrt{2\pi\sigma(t)^2})e^{-x^2/(2\sigma(t)^2)}$ avec $\sigma(t) = \sqrt{\sigma_0^2 + 2Dt}$ diffusant au cours du temps t , pour $\sigma_0 = 1$ et $D = 1$.

Si on imagine inverser l'axe du temps, c'est-à-dire transformer t en $-t$, on observerait une concentration de la densité au lieu d'un étalement. L'évolution de la densité n'est pas symétrique lors d'un renversement du temps : $\rho(x, -t) \neq \rho(x, t)$. La diffusion est un phénomène irréversible. De manière générale, l'évolution spontanée de systèmes macroscopiques, c'est-à-dire contenant un très grand nombre de particules (de l'ordre du nombre d'Avogadro $N \sim 10^{23}$), est irréversible. Ceci correspond à la deuxième loi de la thermodynamique qui stipule que l'évolution spontanée d'un système macroscopique isolé se fait de façon irréversible dans le sens de l'augmentation de l'entropie.

3.2 Équation de Schrödinger

Un électron est une particule élémentaire dont la description précise nécessite la mécanique quantique. En mécanique quantique, un électron est décrit par une fonction d'onde $\Psi : (x, y, z, t) \mapsto \Psi(x, y, z, t)$ dépendant des coordonnées d'espace x, y, z et du temps t , et qui peut prendre des valeurs complexes, c'est-à-dire $\Psi(x, y, z, t) \in \mathbb{C}$. La signification physique de la fonction d'onde Ψ est que son module au carré $\rho(x, y, z, t) = |\Psi(x, y, z, t)|^2$ représente la densité de probabilité de détecter l'électron au point (x, y, z) si on mesure sa position au temps t . La fonction d'onde d'un électron suit un modèle différentiel aux dérivées partielles très célèbre : l'équation de Schrödinger (1925).

Définition (Équation de Schrödinger). Soit la fonction d'onde $\Psi : (x, y, z, t) \mapsto \Psi(x, y, z, t)$ d'un électron dépendant des coordonnées d'espace x, y, z et du temps t , et une fonction d'énergie potentielle $V : (x, y, z) \mapsto V(x, y, z)$. L'équation de Schrödinger est :

$$i\hbar \frac{\partial \Psi}{\partial t}(x, y, z, t) = -\frac{\hbar^2}{2m_e} \Delta \Psi(x, y, z, t) + V(x, y, z) \Psi(x, y, z, t), \quad (\text{III.2.6})$$

où $\hbar = h/(2\pi)$ est la constante de Planck réduite et m_e est la masse de l'électron.

Le terme $-\hbar^2/(2m_e)\Delta\Psi(x, y, z, t)$ est relié à l'énergie cinétique de l'électron, alors que le terme $V(x, y, z)\Psi(x, y, z, t)$ est relié à l'énergie potentielle de l'électron. Connaissant une fonction d'onde initiale, $\Psi(x, y, z, 0)$, la résolution de l'équation de Schrödinger permet de d'obtenir la fonction d'onde $\Psi(x, y, z, t)$ à un temps ultérieur. Cependant, comme pour l'équation de diffusion, on ne sait pas résoudre exactement l'équation de Schrödinger pour n'importe quelle fonction d'onde initiale et énergie potentielle $V(x, y, z)$.

En chimie quantique, on recherche le plus souvent des fonctions d'onde sous la forme

$$\Psi(x, y, z, t) = \psi(x, y, z)e^{-iEt/\hbar},$$

où E est un nombre réel représentant l'énergie de l'électron, $e^{-iEt/\hbar}$ s'appelle un facteur de phase, et ψ est une fonction d'onde ne dépendant pas de t . Ces fonctions d'onde Ψ sont dites stationnaires car la densité de probabilité est indépendante du temps

$$\rho(x, y, z, t) = |\Psi(x, y, z, t)|^2 = |\psi(x, y, z)|^2,$$

où nous avons utilisé que le facteur de phase a un module égale à un : $|e^{-iEt/\hbar}| = 1$. En injectant cette forme de Ψ dans chacun des membres de l'équation de Schrödinger (III.2.6), on a :

$$i\hbar \frac{\partial \Psi}{\partial t}(x, y, z, t) = E\psi(x, y, z)e^{-iEt/\hbar}$$

et

$$\begin{aligned} & -\frac{\hbar^2}{2m_e} \Delta \Psi(x, y, z, t) + V(x, y, z) \Psi(x, y, z, t) \\ & = \left(-\frac{\hbar^2}{2m_e} \Delta \psi(x, y, z) + V(x, y, z) \psi(x, y, z) \right) e^{-iEt/\hbar}. \end{aligned}$$

On peut donc simplifier le facteur de phase $e^{-iEt/\hbar}$ et on arrive à l'équation de Schrödinger indépendante du temps

$$\left(-\frac{\hbar^2}{2m_e} \Delta \psi(x, y, z) + V(x, y, z) \psi(x, y, z) \right) = E\psi(x, y, z).$$

Pour les atomes ou les molécules, une fonction d'onde ψ qui satisfait l'équation de Schrödinger indépendante du temps s'appelle une orbitale. La résolution de l'équation de Schrödinger indépendante du temps permet de déterminer toutes les orbitales ψ et leurs énergies associées E .

Par exemple, pour l'atome hydrogène, l'énergie potentielle de l'électron due à son attraction électrostatique avec le proton est

$$V(x, y, z) = \frac{-e^2}{4\pi\epsilon_0 r},$$

où e est la charge élémentaire, ϵ_0 est la permittivité du vide et $r = \sqrt{x^2 + y^2 + z^2}$ est la distance entre l'électron et le proton. L'orbitale de plus basse énergie est l'orbitale 1s qui a pour expression :

$$\psi_{1s}(x, y, z) = \frac{1}{\sqrt{\pi}a_0^{3/2}} e^{-\sqrt{x^2+y^2+z^2}/a_0},$$

où $a_0 = 4\pi\epsilon_0\hbar^2/(m_e e^2)$ est le rayon de Bohr. Comme exercice, nous invitons le lecteur de vérifier que la fonction ψ_{1s} satisfait bien l'équation de Schrödinger indépendante du temps et de déterminer l'énergie E correspondante.