

Compendium de formules utiles pour l'informatique et l'analyse des algorithmes

Steven Pigeon
Professeur
Département de mathématiques,
informatique et génie
UQAR

Compendium de formules
utiles pour l'informatique
et l'analyse des algorithmes

Compendium de formules utiles pour l'informatique et l'analyse des algorithmes

par

Steven Pigeon, professeur

Département de mathématiques,
informatique et génie
UQAR

Steven Pigeon
Département de mathématiques, informatique, et génie
Université du Québec à Rimouski
Rimouski, Québec
G5L 3A1
steven_pigeon@uqar.ca

Le document a été typographié par l'auteur avec la police libre URW-Garamond par Mathdesign et le système L^AT_EX. Les illustrations, sauf indications contraires, sont de la main de l'auteur. La photo « *Les moustaches de Piskounov* » de la couverture est © Steven Pigeon.

V 0.5 Janvier 2020

005.7~~xxxx~~
QA76.~~xx~~ 2020
ISBN 978-0-0000-0000-2

© 2020 Steven Pigeon

Tous droits de traduction totale ou partielle réservés pour tous les pays. La reproduction d'un extrait quelconque de ce livre, sauf selon les dispositions prévue par la loi, par quelque procédé que ce soit, tant électronique que mécanique, en particulier par photocopie, est interdite sans l'autorisation écrite de l'auteur.

*Je ne me suis jamais privé de donner mon temps aux sciences
Par la science j'ai dénoué les quelques noeuds d'obscur secrets
Après soixante-douze années de réflexion sans jour de trêve
Mon ignorance, je la sais...*

Omar Khayyam
Le Ruba'iyat

Table des matières

Table des matières	i
Table des notations	v
Conventions typographiques	vii
1 Exposants et logarithmes	1
1.1 Exposants	1
1.1.1 Lois	1
1.1.2 Algorithmes	2
1.2 Logarithmes	3
1.2.1 Lois	3
1.2.2 Algorithmes	5
1.3 Remarques bibliographiques	6
2 Différentielles et intégrales	7
2.1 limites	7
2.1.1 Lois	7
2.2 Dérivées	8
2.2.1 Notation	8
2.2.2 Lois	9
2.3 Intégrales	10
2.3.1 Lois	10
2.4 Démonstrations	12
2.4.1 Règles de dérivées	12
2.4.2 Règles d'intégrales	12
2.5 Remarques bibliographiques	12
3 Sommes et produits	15
3.1 Propriétés des sommes	15
3.1.1 Notation	15
3.1.2 Lois	16
3.2 Sommes de Puissances	18
3.2.1 Lois	18
3.2.2 Formules générales	19
3.3 Sommes de progressions	20
3.3.1 Lois	20
3.4 Propriétés des produits	24
3.4.1 Notation	24
3.4.2 Lois	24
3.5 Remarques bibliographiques	26
4 Nombres Complexes	27
4.1 Forme cartésienne	27

TABLE DES MATIÈRES

4.1.1	Opérations et lois	28
4.2	Forme polaire	30
4.2.1	Opérations et lois	30
4.3	Remarques bibliographiques	31
5	Polynômes et racines	35
5.1	Définitions	35
5.1.1	Structure et représentation	35
5.1.2	Vocabulaire et notation	36
5.2	Évaluation	37
5.2.1	Méthode de Horner	38
5.2.2	Évaluation factorisée	38
5.2.3	Formes spéciales	38
5.2.4	Évaluation parallèle	40
5.3	Racines	41
5.3.1	Réduction à la forme homogène	41
5.3.2	Degrés 0 et 1	42
5.3.3	Second degré	42
5.3.3.1	Méthode du discriminant	43
5.3.3.2	Trouver les racines : compléter le carré	43
5.3.3.3	Trouver les racines : décalage à l'ordonnée	45
5.3.3.4	Trouver les racines : factorisation	47
5.3.4	Cubique	48
5.3.4.1	Trouver les racines : compléter le cube	49
5.3.4.2	Trouver les racines : décalage à l'ordonnée	51
5.3.5	Degrés supérieurs et méthodes numériques	53
5.3.5.1	Méthode de la sécante	53
5.3.5.2	Méthode de Newton	57
5.3.5.3	Méthode des points fixes	60
5.3.5.4	Remarques sur les méthodes numériques	63
5.4	Factorisation	65
5.4.0.1	Identités remarquables	65
5.4.0.2	Factorisations connues	66
5.5	Remarques bibliographiques	67
6	Vecteurs et matrices	71
6.1	Vecteurs et matrices	71
6.2	Notation	71
6.3	Vecteurs	72
6.3.1	Opérations	72
6.3.1.1	Addition/Soustraction	72
6.3.1.2	Produit avec un scalaire	73
6.3.1.3	Produit scalaire	73
6.3.1.4	Produit dyadique	74
6.3.1.5	Produit croisé	75
6.4	Matrices	76
6.4.1	Opérations	77
6.4.1.1	Addition/Soustraction	77
6.4.1.2	Produit avec un scalaire	77
6.4.1.3	Produit avec un vecteur	77
6.4.1.4	Produit avec une autre matrice	78
6.4.1.5	Transposées	79
6.4.1.6	Inverses	79
6.5	Normes	80
6.6	Déterminants	82
6.6.1	Calcul du déterminant	83
6.7	Projections	83

6.7.1	Projection vecteur contre vecteur	83
6.7.2	Projection vecteur contre un plan	84
6.8	Systèmes d'équations	86
6.8.1	Rang, déterminant, et solvabilité	88
6.8.2	Résoudre les systèmes d'équations	88
6.8.2.1	Systèmes homogènes	89
6.8.2.2	Systèmes sous-déterminés et pseudo-inverse à droite	90
6.8.2.3	Systèmes sur-déterminés et pseudo-inverse à gauche	92
6.8.2.4	Régressions	94
6.9	Dérivées	95
6.9.1	Vecteurs	95
6.9.2	Matrices	96
6.9.3	Normes	97
6.10	Remarques bibliographiques	97
7	Constantes	99
7.1	Constantes	99
7.1.1	Constante d'Euler-Mascheroni	99
7.1.2	Logarithme naturel de 2	100
7.1.3	Inverse de racine carrée de 2	101
7.1.4	Racine carrée de 2	102
7.1.5	Le nombre d'or	102
7.1.6	Racine de 2 Pi	103
7.1.7	e	104
7.1.8	Pi	105
7.2	Remarques bibliographiques	106
8	Combinatoire	111
8.1	Combinatoire	111
8.2	Factorielles	111
8.3	Coefficients binomiaux & cie.	112
8.4	Remarques bibliographiques	113
9	Astuces en vrac	115
9.1	Astuces en vrac	115
9.2	Tests de divisibilité	115
9.2.1	Divisibilité par 2 ou 2^n	115
9.2.2	Divisibilité par 3 ou 6	116
9.2.3	Divisibilité par 5, 5^k , 10^k	117
9.2.4	Divisibilité par 7	117
9.2.5	Divisibilité par 9	118
9.3	Remarques bibliographiques	118
Bibliographie		121
Index		129

TABLE DES MATIÈRES

Table des notations

\mathbb{N}	Les nombres naturels, $\{1, 2, 3, \dots\}$
\mathbb{N}_0	Les naturels incluant le zéro, $\{0, 1, 2, 3, \dots\}$
\mathbb{Z}	Les entiers, $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$
\mathbb{Z}^*	Les entiers non négatifs, équivalent à \mathbb{N}_0
\mathbb{Z}^+	Les entiers positifs, équivalent à \mathbb{N}
\mathbb{Q}, \mathbb{Q}'	Les rationnels, les irrationnels
\mathbb{R}	Les réels
\mathbb{R}^*	Les réels non négatifs
\mathbb{R}^+	Les réels (strictement) positifs
$ x $	Valeur absolue
$[x]$	Le <i>plancher</i> de x : le plus grand entier plus petit ou égal à x
$\lceil x \rceil$	Le <i>plafond</i> de x : le plus petit entier plus grand ou égal à x
$\{x_i\}_{i=a}^b$	La séquence $\{x_a, x_{a+1}, \dots, x_{b-1}, x_b\}$
\perp	Séquence vide, mot vide
$ A $	Cardinalité de l'ensemble A , longueur de l'objet A
$A \times B$	Si A et B sont des ensembles, produit cartésien
A^n	Si A est un ensemble, la puissance cartésienne
A^*	$\{\perp\} \cup A \cup A^2 \cup A^3 \cup \dots$
$A \cup B$	Union de A et B
$A \cap B$	Intersection de A et B
$A \subseteq B, A \subset B$	Inclusion, inclusion stricte
$A^c, \complement_U A$	Complément de A , complément de A dans U
$P(S)$	L'ensemble de toutes les parties d'un ensemble
$\ln x$	Logarithme naturel (base e) de x
$\log_{10} x$	Logarithme à base 10 de x
$\lg x$	Logarithme à base 2 de x
$\log x$	Logarithme à base indifférente

$\exists, \exists!, \nexists$	Quantificateurs existentiels : il existe, il n'existe qu'un seul, il n'existe aucun
\forall	Quantificateur universel : pour tous
$a \ll b$	a est beaucoup plus petit que b
$a \gg b$	a est beaucoup plus grand que b
$a \sim b$	a est de la forme b , a est semblable à b
$a \pm b$	$a + b$ ou $a - b$. Lorsqu'en paire $a \pm b$, $c \mp d$, on considère les choix complémentaires : $a + b$ et $c - d$ ou $a - b$ et $c + d$.
$\binom{n}{m}$	Nombre de façons de choisir m parmi n , $\binom{n}{m} = \frac{n!}{m!(n-m)!}$
$\binom{n}{n_1 n_2 \dots n_k}$	Nombre de façons de choisir n_1, n_2, \dots, n_k symboles indistinguables parmi n
ssi	Si, et seulement si.
γ	La constante d'Euler-Mascheroni, $\gamma = 0.577215664901532\dots$
e	La constante d'Euler, $e = 2.718281828459045\dots$
π	La constante d'Archimède, $\pi = 3.141592653589793\dots$

Conventions typographiques

Le texte normal est typographié avec la police libre URW-Garamond par Mathdesign. Les programmes et les éléments de programmation sont typographiés avec la police ASCII de Syropoulos et Nickalls, qui rappelle une police de terminal. Les éléments ordinaires sont typographiés ainsi alors que les mots-clefs du langage sont typographiés en gras, comme pour **while**, **if**, **for**, etc.

Les librairies utilisées par les programmes sont tirés de la librairie standard. Par convention, les librairies introduites par `#include <librairie>` font partie du système et sont en principe disponibles en même temps que le compilateur C++. Les librairies introduites par des guillemets, comme par exemple `#include "librairie"` sont les librairies fournies par l'utilisateur, c'est-à-dire par le programme plutôt que par la librairie standard.

Les références, notées par exemple [126], vous amènent à la fin de l'ouvrage où elles sont présentées en ordre alphabétique d'auteurs, et non en ordre d'apparition.

Les démonstrations se trouvent presque toujours dans un encadré intitulé comme tel. Pour une première lecture ou pour un cours moins avancé, ces démonstrations peuvent être omises, mais nous invitons tout de même le lecteur à s'y attarder. Comme pour les démonstrations, les exemples sont encadrées.

1

Exposants et logarithmes

Sommaire. Dans ce chapitre, nous considérons les lois qui gouvernent les exposants et les logarithmes (qui sont, en quelque sorte, l'inverse des exposants). Nous expliquerons aussi comment calculer les uns et les autres. Nous terminerons avec les remarques bibliographiques.

1.1 Exposants

1.1.1 Lois

1. $x^0 = 1$, mais 0^0 est indéterminé.
2. $x^a = \underbrace{x \cdot x \cdots x \cdot x}_{a \text{ fois}}$.
3. $x^{-a} = \frac{1}{x^a}$ et $\frac{1}{x^{-b}} = x^b$.
4. $a = e^{\ln a}$ (et $a = b^{\log_b a}$).
5. $x^a x^b = x^{a+b}$.
6. $(x^a)^b = x^{ab}$.
7. $\left(\frac{x}{y}\right)^a = \frac{x^a}{y^a}$.
8. $\frac{x^a}{x^b} = x^{a-b} = \begin{cases} 1 & \text{si } a = b \text{ car } x^0 = 1 \\ x^{a-b} & \text{si } a > b \\ \frac{1}{x^{b-a}} & \text{si } a < b \end{cases}$.

9. $\sqrt[b]{x} = x^{\frac{1}{b}}$ et donc $\sqrt[b]{x^a} = (x^a)^{\frac{1}{b}} = x^{\frac{a}{b}}$.

10. $\sqrt[b]{\frac{y}{x}} = \frac{\sqrt[b]{y}}{\sqrt[b]{x}} = \frac{y^{\frac{1}{b}}}{x^{\frac{1}{b}}}$.

11. $x^a + x^b = x^a \left(1 + \frac{x^b}{x^a}\right) = x^a (1 + x^{b-a})$.

12. $a^x = (e^{\ln a})^x = e^{x \ln a}$ (donc on a aussi $a^x = (b^{\log_b a})^x = b^{x \log_b a}$).

1.1.2 Algorithmes

Pour calculer e^x , on peut utiliser

$$e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n.$$

Si l'exposant est entier, il existe un algorithme pour calculer x^n en $O(\log_2 n)$ étapes, car la décomposition binaire de l'exposant nous donne quelles puissances participent au produit. Par exemple, x^{25} peut s'écrire comme $x^{25} = (1 \cdot x^{16})(1 \cdot x^8)(0 \cdot x^4)(0 \cdot x^2)(1 \cdot x^1) = x^{16}x^8x$, et on peut vérifier que les coefficients des puissances correspondent bien à la décomposition binaire de l'exposant, $25_{10} = 11001_2$. De plus, les puissances de x se calculent avec des mises au carré successives, comme pour $(x^8)^2 = x^{16}$. Puisqu'il y a $O(\lg n)$ bits dans l'exposant n , l'algorithme calcule l'exponentiation en $O(\lg n)$ multiplications.

```

bigint fast_expo(bigint x, bigint e)
{
    bigint p=1;

    while (e)
    {
        if (e&1) p*=x;
        x*=x;
        e/=2;
    }

    return p;
}

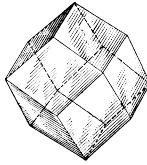
```

Lorsque l'exposant est quelconque, nous pouvons l'approximer comme un nombre rationnel, et calculer $x^y \approx x^{\frac{a}{b}} = \sqrt[b]{x^a}$. Sinon, nous pouvons toujours utiliser l'identité

$$x^y = \sum_{i=0}^{\infty} \frac{(y \ln x)^i}{i!} = 1 + y \ln x + \frac{(y \ln x)^2}{2} + \frac{(y \ln x)^3}{6} + \frac{(y \ln x)^4}{24} + \dots$$

ou encore

$$x^y = \lim_{n \rightarrow \infty} \left(1 + \frac{y \ln x}{n} \right)^n.$$



1.2 Logarithmes

1.2.1 Lois

Dans tous les cas ci-dessous, nous supposons la base strictement positive (donc plus grande que 0) et différente de 1 :

1. $\ln a$ est le logarithme naturel, base e , de a .
2. $\lg a$ est le logarithme à base de 2 de a .
3. $\log a$ est logarithme de a , mais la notation est ambiguë. Parfois c'est le logarithme à base 10, parfois à base e . Souvent, c'est le logarithme à la base osef.
4. $\log_{10} a$ est le logarithme à base 10 de a .
5. $\log_b a$ est le logarithme à base b de a .
6. $\log_b 1 = 0$.
7. $\log_b b = 1$.
8. $\log_b a = \frac{\ln a}{\ln b}$, ou encore $\log_b a = \frac{\log_d a}{\log_d b} = \frac{\frac{\ln a}{\ln d}}{\frac{\ln b}{\ln d}} = \frac{\ln a \ln d}{\ln d \ln b} = \frac{\ln a}{\ln b}$.
9. $a = e^{\ln a}$ (et $a = b^{\log_b a}$).
10. $\log_b x^y = y \log_b x$.
11. $\log_b \sqrt{x} = \log_b x^{\frac{1}{2}} = \frac{1}{2} \log_b x$.

$$12. \log_b \sqrt[y]{x} = \log_b x^{\frac{1}{y}} = \frac{1}{y} \log_b x.$$

$$13. \log_b xy = \log_b x + \log_b y.$$

$$14. \log_b \frac{x}{y} = \log_b x - \log_b y.$$

$$15. \frac{\log_a x}{\log_a y} = \frac{\log_b x}{\log_b y} = \frac{\ln x}{\ln y}.$$

Démonstration 1.2.1.

Montrons la première égalité, ce qui nous donnera automagiquement la seconde. Certes :

$$\begin{aligned} \frac{\log_a x}{\log_a y} &= \frac{\log_b x}{\log_b y} \\ \frac{\frac{\ln x}{\ln a}}{\frac{\ln y}{\ln a}} &= \frac{\frac{\ln x}{\ln b}}{\frac{\ln y}{\ln b}} \\ \frac{\ln x}{\ln a} \frac{\ln a}{\ln y} &= \frac{\ln x}{\ln b} \frac{\ln b}{\ln a} \\ \frac{\ln x}{\ln y} &= \frac{\ln x}{\ln y}. \end{aligned}$$

□

$$16. \log_b(x+y) = \log_b x \left(1 + \frac{y}{x}\right) = \log_b x + \log_b \left(1 + \frac{y}{x}\right). \text{ Si } x \text{ est beaucoup plus grand que } y, \\ \text{ alors } 1 + \frac{y}{x} \approx 1, \text{ et } \log_b \left(1 + \frac{y}{x}\right) \approx 0.$$

$$17. \log_b(-a) = i\pi \log_b e + \log_b a \text{ (généralisation des logarithmes aux nombres négatifs).}$$

Démonstration 1.2.2.

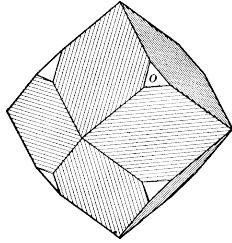
Par la formule d'Euler, $e^{i\pi} + 1 = 0$, nous avons $e^{i\pi} = -1$, et donc $\log_b(-1) = \log_b e^{i\pi} = i\pi \log_b e$ (avec le cas spécial $\log_b e = 1$ lorsque $b = e$ (voir règle 7 ci dessus)). Nous avons donc

$$\log_b(-a) = \log_b(-1)(a) = \log_b(-1) + \log_b(a) = i\pi + \log_b a.$$

□

$$18. \log_b(x+iy) = \log_b \sqrt{x^2+y^2} + i \tan^{-1}(x,y), \text{ où } \tan^{-1}(x,y) \text{ est l'arctangente de la pente } \frac{y}{x}.$$

$$19. \log_b(-ai) = \log_b a - i \frac{\pi}{2} \log_b e \text{ (généralisation aux nombres complexes; voir ch. 4).}$$



1.2.2 Algorithmes

Si $|x| \ll 1$, alors nous pouvons utiliser la série

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots = \sum_{i=1}^{\infty} (-1)^{i+1} \frac{(x-1)^i}{i},$$

tandis que pour $x \approx z$,

$$\ln x = \ln z + \frac{x-z}{z} - \frac{(x-z)^2}{z^2} + \frac{(x-z)^3}{z^3} - \frac{(x-z)^4}{z^4} + \dots = \ln z + \sum_{i=1}^{\infty} (-1)^i \frac{(x-z)^i}{iz^i},$$

où z est choisi pour être pratique¹. Si on remarque que puisque $x \approx z$, on peut écrire $x = z + y$, la dernière formule se simplifie grandement :

$$\begin{aligned} \ln x &= \ln z + y = \ln z + \sum_{i=1}^{\infty} (-1)^i \frac{(x-z)^i}{iz^i} \\ &= \ln z + \sum_{i=1}^{\infty} (-1)^i \frac{((z+y)-z)^i}{iz^i} \\ &= \ln z + \sum_{i=1}^{\infty} (-1)^i \frac{y^i}{iz^i}. \end{aligned}$$

Mal pris, on peut toujours utiliser l'approximation

$$\ln z \approx n(\sqrt[n]{x} - 1),$$

avec $n \ll \text{assez grand}$.

Démonstration 1.2.3. $\ln z \approx n(\sqrt[n]{x} - 1)$ pour $n \ll \text{assez grand}$. Sachant que

$$z = e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n,$$

1. On supposera qu'il a une forme pratique qui se prête bien à une astuce quelconque pour faciliter le calcul de $\ln z$.

calculer $\ln z$ revient à isoler x dans l'éq. ci-dessus. Voyons :

$$\begin{aligned} z &= \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n \\ \lim_{n \rightarrow \infty} \sqrt[n]{z} &= \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right) \\ \lim_{n \rightarrow \infty} \sqrt[n]{z} &= \lim_{n \rightarrow \infty} 1 + \lim_{n \rightarrow \infty} \frac{x}{n} \\ \lim_{n \rightarrow \infty} \sqrt[n]{z} - 1 &= \lim_{n \rightarrow \infty} \frac{x}{n} \\ \lim_{n \rightarrow \infty} n(\sqrt[n]{z} - 1) &= \lim_{n \rightarrow \infty} x \\ \lim_{n \rightarrow \infty} n(\sqrt[n]{z} - 1) &= x, \end{aligned}$$

établissant $\ln z = x \approx n(\sqrt[n]{z} - 1)$ lorsque n est « assez grand ». □

1.3 Remarques bibliographiques

L'algorithme d'exponentiation rapide par mises au carré successives n'est pas nouveau. En effet, Pingala (vers 300–200 av. J.-C.), dans son *Chandahśāstra* (le *Chandah sūtra*), présente une procédure très proche [29, p. 76] et l'idée se retrouve dans le monde Arabe au X^e siècle [139]. Pour en savoir plus sur l'invention des logarithmes sans lire Napier en latin [113], on pourra lire *John Napier and the Invention of Logarithms* [69] ou encore les *Memoirs of John Napier of Merchiston* [110]. Pour un traitement mathématique, il existe plusieurs références [19, 101].

*
* * *

Le problème du calcul efficace et numériquement stable des fonctions spéciales comme l'exponentiation, les logarithmes et les fonctions trigonométriques n'est pas entièrement réglé. Le calcul en matériel est encore sujet d'explorations [120]. On s'intéresse depuis longtemps à l'approximation de fonction avec des erreurs bornées [158], et on s'y intéresse encore [164]. Le problème d'obtenir des approximations ou même des méthodes « exactes » numériquement stables est vaste et riche [5, 6, 27, 65, 76, 132, 167, 169, 170].

2

Différentielles et intégrales

Sommaire. Dans ce chapitre, nous présentons les lois qui s'appliquent au limites, aux dérivées et aux intégrales usuelles. Il existe évidemment des listes bien plus longues où nous trouvons aussi les intégrales et les dérivées avec des formes trigonométriques, mais celles-ci nous intéressent assez rarement en informatique. Nous terminerons, comme à l'habitude, avec des remarques bibliographiques.

2.1 limites

2.1.1 Lois

1. $\lim_{x \rightarrow c} af(x) = a \lim_{x \rightarrow c} f(x).$
2. $\lim_{x \rightarrow c} (f(x) \pm g(x)) = \lim_{x \rightarrow c} f(x) \pm \lim_{x \rightarrow c} g(x).$
3. $\lim_{x \rightarrow c} (f(x)g(x)) = \left(\lim_{x \rightarrow c} f(x) \right) \left(\lim_{x \rightarrow c} g(x) \right).$
4. $\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \frac{\lim_{x \rightarrow c} f(x)}{\lim_{x \rightarrow c} g(x)}.$
5. $\lim_{x \rightarrow c} f(x)^n = \left(\lim_{x \rightarrow c} f(x) \right)^n.$

6. Loi de l'Hôpital : $\lim_{x \rightarrow c} \frac{f(x)}{g(x)} = \lim_{x \rightarrow c} \frac{f'(x)}{g'(x)} = \lim_{x \rightarrow c} \frac{\frac{\partial}{\partial x} f(x)}{\frac{\partial}{\partial x} g(x)}$, en autant que $\frac{\partial}{\partial x} g(x) \neq 0$.

La loi de l'Hôpital est particulièrement utile pour les formes indéterminées $\frac{0}{0}$ et $\frac{\infty}{\infty}$, mais elle s'applique à chaque fois que la limite existe. Elle n'est pas définie si $g'(x) = 0$, car nous avons une division par zéro, ce qui est interdit, même à l'Hôpital.

2.2 Dérivées

2.2.1 Notation

Les dérivées sont notées de plusieurs façons :

1. $f'(x)$, $\frac{d}{dx} f(x)$ ou $\frac{df}{dx}$, la dérivée de f par rapport à la variable x .
2. $\frac{\partial}{\partial x} f(x)$ ou $\frac{\partial f}{\partial x}$, la dérivée partielle de f par rapport à l'une des variables, x .
3. $f''(x)$, $\frac{d^2}{dx^2} f(x)$ ou $\frac{d^2 f}{dx^2}$, la dérivée seconde.
4. $\frac{\partial^2}{\partial x^2} f(x)$ ou $\frac{\partial^2 f}{\partial x^2}$, pour la dérivée partielle seconde.
5. On trouve parfois $f'''(x)$, $f^{IV}(x)$, $f^{(n)}(x)$, etc., pour les dérivées d'ordre supérieurs, plutôt que la notation plus habituelle $\frac{d^n}{dx^n} f(x)$ ou $\frac{d^n f}{dx^n}$, ou encore $\frac{\partial^n}{\partial x^n} f(x)$ ou $\frac{\partial^n f}{\partial x^n}$ lorsqu'il s'agit de dérivées partielles.
6. Les physiciens utilisent parfois la « notation de Newton », \dot{x} , \ddot{x} , etc., pour les « fluxions » des fonctions. On supposera alors que \dot{x} signifie quelque chose comme $\frac{d}{dt} x(t)$, à déterminer selon le contexte.

2.2.2 Lois

1. Si a est une constante, $\frac{da}{dx} = 0$.
2. $\frac{dx}{dx} = 1$.
3. $\frac{\partial}{\partial x} au = a \frac{\partial u}{\partial x}$.
4. $\frac{\partial}{\partial x}(u + v) = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial x}$.
5. $\frac{\partial}{\partial x} uv = u \frac{\partial v}{\partial x} + v \frac{\partial u}{\partial x}$.
6. $\frac{\partial}{\partial x} uvw = uv \frac{\partial w}{\partial x} + uw \frac{\partial v}{\partial x} + vw \frac{\partial u}{\partial x}$.
7. $\frac{\partial}{\partial x} \frac{u}{v} = \frac{v \frac{\partial u}{\partial x} - u \frac{\partial v}{\partial x}}{v^2} = \frac{1}{v} \frac{\partial u}{\partial x} - \frac{u}{v^2} \frac{\partial v}{\partial x}$.
8. $\frac{\partial}{\partial x} u^n = n u^{n-1} \frac{\partial u}{\partial x}$.
9. $\frac{\partial}{\partial x} \frac{1}{u} = -\frac{1}{u^2} \frac{\partial u}{\partial x}$.
10. $\frac{\partial}{\partial x} \sqrt{u} = \frac{1}{2\sqrt{u}} \frac{\partial u}{\partial x}$.
11. $\frac{\partial}{\partial x} \frac{1}{u^n} = -\frac{n}{u^{n+1}} \frac{\partial u}{\partial x}$.
12. $\frac{\partial}{\partial x} f(u) = \left(\frac{\partial}{\partial u} f(u) \right) \frac{\partial u}{\partial x}$, la « règle de la chaîne ».
13. $\frac{\partial}{\partial x} \ln u = \frac{1}{u} \frac{\partial u}{\partial x}$.
14. $\frac{\partial}{\partial x} \log_a u = \frac{1}{\ln a} \frac{1}{u} \frac{\partial u}{\partial x} = \frac{1}{u \ln a} \frac{\partial u}{\partial x}$.
15. Si a est une constante, $\frac{\partial}{\partial x} a^u = a^u \ln a \frac{\partial u}{\partial x}$

16. Le cas particulier $a = e$, $\frac{\partial}{\partial x} e^u = e^u \frac{\partial u}{\partial x}$.

17. $\frac{\partial}{\partial x} \frac{u^n}{v^m} = \frac{u^{n-1}}{v^{m-1}} \left(nv \frac{\partial u}{\partial x} - mu \frac{\partial v}{\partial x} \right)$.

18. $\frac{\partial}{\partial x} u^v = vu^{v-1} \frac{\partial u}{\partial x} + u^v \ln u \frac{\partial v}{\partial x}$.

19. $\frac{\partial^2}{\partial x^2} f(u) = \frac{\partial}{\partial u} f(u) \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2}{\partial u^2} f(u) \left(\frac{\partial u}{\partial x} \right)^2$.

20. Les lois de Leibniz

(a) $\frac{\partial}{\partial b} \int_a^b f(x) \partial x = f(b)$, avec a constante.

(b) $\frac{\partial}{\partial a} \int_a^b f(x) \partial x = -f(a)$, avec b constante.

(c) $\frac{\partial}{\partial a} \int_b^c f(x, a) \partial x = \int_b^c \frac{\partial}{\partial a} f(x, a) \partial x - f(b, a) \frac{db}{da} + f(c, a) \frac{\partial c}{\partial a}$.

2.3 Intégrales

2.3.1 Lois

1. $\int a \partial x = a \int \partial x = ax$.

2. $\int af(x) \partial x = a \int f(x) \partial x$.

3. $\int (u + v) \partial x = \int u \partial x + \int v \partial x$.

4. $\int u \partial v = u \int \partial v - \int v \partial u = uv - \int v \partial u$, « par parties »
 variante : $\int u \frac{dv}{dx} \partial x = uv - \int v \frac{du}{dx} \partial x$.

5. $\int x^n dx = \frac{1}{n+1} x^{n+1}$, si $n \neq -1$. Si $n = -1$, $\int \frac{1}{x} dx = \ln x$.

6. $\int \frac{\frac{d}{dx} f(x)}{f(x)} dx = \ln f(x)$.

7. $\int b^{ax} dx = \frac{b^{ax}}{a \ln b}$.

Avec les cas spéciaux :

(a) $b = e$, $\int e^{ax} dx = \frac{1}{a} e^{ax}$,

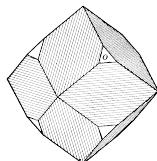
(b) and $b = e, a = 1$, $\int e^x dx = e^x$.

8. $\int \log_a x dx = \frac{1}{\ln a} \int \ln x dx = \frac{x \ln x - x}{\ln a}$.

9. $\int a^x \ln a dx = \ln a \int a^x dx = a^x$.

10. Si $n \neq -1$, $\int (a + bx)^n dx = \frac{(a + bx)^{n+1}}{(n+1)b}$.

11. $\int x^m (a + bx)^n dx = \frac{x^{m+1}(a + bx)^n}{m+n+1} + \frac{an}{m+n-1} \int x^m (a + bx)^{n-1} dx$.

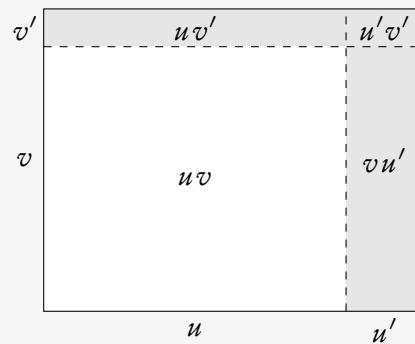


2.4 Démonstrations

Démontrons quelques unes des lois ci-dessus.

2.4.1 Règles de dérivées

Démonstration 2.4.1. Montrons que la règle pour la dérivée d'un produit est en fait seulement une approximation. Si on considère le produit uv comme une surface, on peut, avec les dérivées u' et v' , faire le schéma suivant :



On y voit clairement que la dérivée de cette surface est $uv' + u'v + u'v'$, mais comme $u'v' \approx 0$ (puisque petit fois petit, c'est « petit au carré »), on peut approximer, et obtenir $(uv)' = uv' + u'v$. \square

Démonstration 2.4.2. La dérivée de $f(x)g(x)$ est $(f(x)+f'(x))(g(x)+g'(x)) - f(x)g(x)$, soit $f(x)g'(x) + f'(x)g(x) + f'(x)g'(x)$. Or, si on considère que $f'(x)g'(x)$ est négligeable, il ne nous reste que $f(x)g'(x) + f'(x)g(x)$, ce qui établit la formule. \square

2.4.2 Règles d'intégrales

2.5 Remarques bibliographiques

En 1694, Guillaume François Antoine, marquis de L'Hospital (1661–1704), s'attache les services de Johann (Jean) Bernoulli (1667–1748) avec la condition expresse d'être le premier à prendre connaissance de toute découverte faite par ce dernier, et d'en faire l'usage qu'il lui plaira. À peine

deux ans plus tard, en 1696, le marquis fait paraître — d'abord anonymement, par modestie sans doute — le premier livre en français sur le calcul différentiel selon la méthode de Leibniz, *Analyse des Infiniment Petits pour l'Intelligence des Lignes Courbes*. Or, voilà ! Scandale ! Bien que le marquis ait gracieusement reconnu « devoir beaucoup aux lumières de M^{rs} Bernoulli, surtout à celles du jeune présentement professeur à Groningue » [102, p. XIV], Bernoulli se juge lésé et se plaint à qui veut l'entendre que nombre de nouveautés tirées du livre, voire le livre complet, lui sont dus. Toutefois, la relative modération avec laquelle il formule ses doléances après la mort de L'Hospital et les différentes versions des évènements, plus ou moins enjolivées, jettent un doute sur l'affaire [160], mais il n'en demeure pas moins que les contributions de Bernoulli sont importantes.

3

Sommes et Produits

Sommaire. Ce chapitre se consacre aux sommes et aux produits discrets, en insistant sur les formes que l'on rencontre le plus souvent dans l'analyse des algorithmes. Certaines des identités les plus intéressantes seront aussi démontrées. Nous finirons avec les remarques bibliographiques.

3.1 Propriétés des sommes

3.1.1 Notation

La notation

$$\sum_{i=a}^b f(i)$$

exprime la somme des valeurs que prend $f(i)$ en faisant varier i , l'indice, de a à b , inclusivement, par incrément de 1. C'est-à-dire :

$$\sum_{i=a}^b f(i) = f(a) + f(a+1) + f(a+2) + \cdots + f(b-2) + f(b-1) + f(b).$$

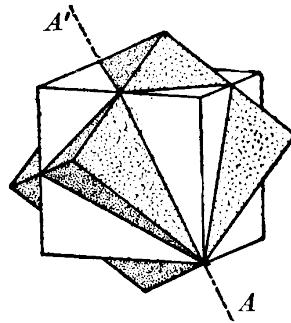
La notation est souvent étendue pour exprimer différemment comment l'indice prend ses valeurs. Par exemple,

$$\sum_{x \in X} g(x)$$

exprime que l'indice x prend tour à tour la valeur de chaque élément de l'ensemble X . La somme peut être imbriquée :

$$\sum_{i=a}^b \sum_{j=c}^d h(i, j) = \sum_{i=a}^b \left(\sum_{j=c}^d h(i, j) \right).$$

Ici, pour chaque valeur de i , j va varier de c à d , inclusivement. C'est l'équivalent de deux boucles imbriquées dans votre langage de programmation préféré.



3.1.2 Lois

Avec la notation usuelle :

1. $\sum_{i=a}^b f(i) = \sum_{i=d}^b f(i) - \sum_{i=d}^{a-1} f(i)$, pour $d < a$.
2. $\sum_{i=a}^b f(i) = \sum_{i=a}^c f(i) + \sum_{i=c+1}^b f(i)$, pour $a \leq c < b$.
3. $\sum_{i=a}^b f(i) = \sum_{j=0}^{b-a} f(j+a)$ (« glissement » des indices.)
4. $\sum_{i=a}^b cf(i) = c \sum_{i=a}^b f(i)$ (mise en évidence, distributivité.)
5. $\sum_{i=a}^b f(i) \pm g(i) = \sum_{i=a}^b f(i) \pm \sum_{i=a}^b g(i)$.
6. $\sum_{i=a}^b \sum_{j=c}^d f(i, j) = \sum_{j=c}^d \sum_{i=a}^b f(i, j)$. (*Hoisting*; renversement des indices.)

$$7. \sum_{i=a}^b \sum_{j=c}^d f(i)g(j) = \left(\sum_{i=a}^b f(i) \right) \left(\sum_{j=c}^d g(j) \right).$$

Démonstration 3.1.1.

$$\begin{aligned} \sum_{i=a}^b \sum_{j=c}^d f(i)g(j) &= \sum_{i=a}^b f(i) \sum_{j=c}^d g(j) \\ &= \sum_{i=a}^b f(i) \left(\sum_{j=c}^d g(j) \right) \\ &= \left(\sum_{j=c}^d g(j) \right) \sum_{i=a}^b f(i) \\ &= \left(\sum_{j=c}^d g(j) \right) \left(\sum_{i=a}^b f(i) \right) \\ &= \left(\sum_{i=a}^b f(i) \right) \left(\sum_{j=c}^d g(j) \right). \end{aligned}$$

□

Avec la notation ensembliste :

1. $\sum_{x \in X} f(x)$, la somme sur les éléments de l'ensemble X .
2. $\sum_{u \in X \cup Y} f(u) = \sum_{x \in X} f(x) + \sum_{y \in Y} f(y) - \sum_{v \in X \cap Y} f(v)$ (inclusion/exclusion.)
3. $\sum_{u \in X \setminus Y} f(u) = \sum_{x \in X} f(x) - \sum_{u \in X \cap Y} f(u)$.
4. $\sum_{u \in X \cap Y} f(u) = \sum_{x \in X} f(x) + \sum_{y \in Y} f(y) - \sum_{u \in X \cup Y} f(u)$.

Sommes, logarithmes, et produits :

1. $\sum_{i=a}^b \log f(i) = \log \left(\prod_{i=a}^b f(i) \right)$.
2. $c^{\sum_{i=a}^b f(i)} = \prod_{i=a}^b c^{f(i)}$, pour une constante c .

3.2 Sommes de Puissances

3.2.1 Lois

0. $\sum_{i=1}^n i^0 = n.$
1. $\sum_{i=1}^n i = \frac{1}{2}n(n+1) = \frac{n^2}{2} + \frac{n}{2}.$
2. $\sum_{i=1}^n i^2 = \frac{1}{6}n(n+1)(2n+1) = \frac{n^3}{3} + \frac{n^2}{2} + \frac{n}{6}.$
3. $\sum_{i=1}^n i^3 = \frac{1}{4}n^2(n+1)^2 = \frac{n^4}{4} + \frac{n^3}{2} + \frac{n^2}{4}.$
4. $\sum_{i=1}^n i^4 = \frac{1}{30}n(n+1)(2n+1)(3n^2+3n-1) = \frac{n^5}{5} + \frac{n^4}{2} + \frac{n^3}{3} - \frac{n}{30}.$
5. $\sum_{i=1}^n i^5 = \frac{1}{12}n^2(n+1)^2(2n^2+2n-1) = \frac{n^6}{6} + \frac{n^5}{2} + \frac{n^4}{12} - \frac{n^2}{12}.$
6. $\sum_{i=1}^n i^6 = \frac{1}{42}n(n+1)(3n^4+6n^3-3n+1) = \frac{n^7}{7} + \frac{n^6}{2} + \frac{n^5}{2} - \frac{n^3}{6} + \frac{n}{42}.$
7. $\sum_{i=1}^n i^7 = \frac{1}{24}n^2(n+1)^2(3n^4+6n^3-n^2-4n+2) = \frac{n^8}{8} + \frac{n^7}{2} + \frac{7n^6}{12} - \frac{7n^4}{24} + \frac{n^2}{12}.$
8. $\begin{aligned} \sum_{i=1}^n i^8 &= \frac{1}{90}n(n+1)(2n+1)(5n^6+15n^5+5n^4-15n^3-n^2+9n-3) \\ &= \frac{n^9}{9} + \frac{n^8}{2} + \frac{2n^7}{3} - \frac{n^5}{15} + \frac{2n^3}{9} - \frac{n}{30}. \end{aligned}$
9. $\begin{aligned} \sum_{i=1}^n i^9 &= \frac{1}{20}n^2(n+1)^2(n^2+n-1)(2n^4+4n^3-n^2-3n+3) \\ &= \frac{n^{10}}{10} + \frac{n^9}{2} + \frac{3n^8}{4} - \frac{7n^6}{10} + \frac{n^4}{2} - \frac{3n^2}{20}. \end{aligned}$
10. $\begin{aligned} \sum_{i=1}^n i^{10} &= \frac{1}{66}n(n+1)(2n+1)(n^2+n-1)(3n^6+9n^5+2n^4-11n^3+3n^2+10n-5) \\ &= \frac{n^{11}}{11} + \frac{n^{10}}{2} + \frac{5n^9}{6} - n^7 + n^5 - \frac{n^3}{2} + \frac{5n}{66}. \end{aligned}$

3.2.2 Formules générales

1. Formule de Piskounov (voir § 3.5) :

$$\sum_{i=1}^n i^k = \frac{1}{k+1} \left((n+1)^{k+1} - \sum_{i=0}^{k-1} \binom{k+1}{i} \sum_{j=0}^n i^j - 1 \right).$$

Démonstration 3.2.1. La démonstration repose sur un cas particulier du théorème binomial. En général nous avons

$$(x+y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i},$$

mais nous utiliserons le cas particulier

$$(x+1)^n = \sum_{i=0}^n \binom{n}{i} x^i 1^{n-i} = \sum_{i=0}^n \binom{n}{i} x^i$$

pour obtenir une somme télescopique qui met en évidence la somme désirée, $\sum_{i=1}^n i^k$. Pour la k^{e} puissance, nous avons, en additionnant par ligne *et* par colonne :

$$\begin{aligned} 1^{k+1} - 0^{k+1} &= 1 \\ 2^{k+1} - 1^{k+1} &= \sum_{i=0}^k \binom{k+1}{i} 1^i \\ 3^{k+1} - 2^{k+1} &= \sum_{i=0}^k \binom{k+1}{i} 2^i \\ &\vdots \quad = \quad \vdots \\ (n+1)^{k+1} - n^{k+1} &= \sum_{i=0}^k \binom{k+1}{i} n^i \\ &\parallel \quad \parallel \\ (n+1)^{k+1} &= \sum_{i=0}^k \sum_{j=1}^n \binom{k+1}{i} j^i + 1 \end{aligned}$$

Réarrangeons les termes :

$$\begin{aligned} (n+1)^{k+1} &= \sum_{i=0}^k \sum_{j=1}^n \binom{k+1}{i} j^i + 1 \\ (n+1)^{k+1} - 1 &= \sum_{i=0}^k \binom{k+1}{i} \sum_{j=1}^n j^i \\ (n+1)^{k+1} - 1 &= \sum_{i=0}^{k-1} \binom{k+1}{i} \sum_{j=1}^n j^i + \underbrace{\binom{k+1}{k} \sum_{j=1}^n j^k}_{\text{le morceau intéressant}} \end{aligned}$$

Enfin,

$$\binom{k+1}{k} \sum_{j=1}^n j^k = (n+1)^{k+1} - \sum_{i=0}^{k-1} \binom{k+1}{i} \sum_{j=1}^n j^i - 1$$

$$\sum_{j=1}^n j^k = \frac{1}{k+1} \left((n+1)^{k+1} - \sum_{i=0}^{k-1} \binom{k+1}{i} \sum_{j=1}^n j^i - 1 \right),$$

ce qui établit le résultat puisque $\binom{k+1}{k} = k+1$.

□

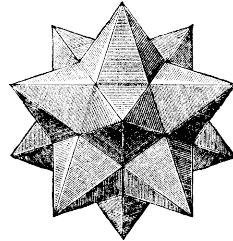
2. Formule de Faulhaber :

$$\sum_{i=1}^n i^k = \frac{1}{k+1} \sum_{i=1}^{k+1} (-1)^{\delta_{ik}} \binom{k+1}{i} B_{k+1-i} n^i,$$

où $d_{ij} = 1$ si $i = j$ et 0 autrement, et les B_j sont les nombres de Bernoulli.

3. Formule très approximative :

$$\sum_{i=1}^n i^k \approx \frac{n^{k+1}}{k+1} + \frac{n^k}{2}.$$



3.3 Sommes de progressions

3.3.1 Lois

1. Somme d'une progression arithmétique $a_1, a_1 + d, a_1 + 2d, \dots$

$$\sum_{i=1}^n a_n = \sum_{i=1}^n (a_1 + (i-1)d) = \frac{1}{2} n (2a_1 + (n-1)d) = \frac{1}{2} n (a_1 + a_n).$$

Démonstration 3.3.1. Sachant que $a_i = a_1 + (i - 1)d$, nous avons

$$\begin{aligned}\sum_{i=1}^n a_i &= \sum_{i=1}^n a_1 + (i - 1)d \\ &= a_1 n + d \sum_{i=1}^n (i - 1) \\ &= a_1 n + d \left(\sum_{i=1}^n i - \sum_{i=1}^n 1 \right) \\ &= a_1 n + d \left(\frac{1}{2} n(n+1) - n \right) \\ &= a_1 n + d \frac{1}{2} n(n-1) \\ &= \frac{1}{2} n(2a_1 + (n-1)d).\end{aligned}$$

Mais

$$\begin{aligned}\frac{1}{2} n(2a_1 + (n-1)d) &= \frac{1}{2} n(a_1 + a_1 + (n-1)d) \\ &= \frac{1}{2} n(a_1 + a_n).\end{aligned}$$

□

2. Somme d'une progression géométrique a, a^2, a^3, \dots

$$\sum_{i=1}^n a^i = \frac{a^n - a}{a - 1} = \frac{a(a^n - 1)}{a - 1}.$$

et

$$\begin{aligned}\sum_{i=0}^n a^i &= 1 + \sum_{i=1}^n a^i \\ &= \frac{a(a^n - 1)}{a - 1} + 1 \\ &= \frac{a^{n+1} - 1 + (a - 1)}{a - 1} \\ &= \frac{a^{n+1} - 1}{a - 1}.\end{aligned}$$

Démonstration 3.3.2. Posons

$$t_n = \sum_{i=1}^n a^i.$$

Alors

$$\begin{aligned}
 at_n - t_n &= a \sum_{i=1}^n a^i - \sum_{i=1}^n a^i \\
 t_n(a-1) &= \sum_{i=1}^n a^{i+1} - \sum_{i=1}^n a^i \\
 &= (a^2 + a^3 + \cdots + a^n + a^{n+1}) - (a + a^2 + \cdots + a^n) \\
 &= (a^{n+1} + a^n + \cdots + a^3 + a^2) - (a^n + \cdots + a^2 + a) \\
 &= a^{n+1} + a^n + \cdots + a^3 + a^2 - a^n - \cdots - a^2 - a \\
 &= a^{n+1} + a^n - a^n + \cdots + a^2 - a^2 - a \\
 &= a^{n+1} - a
 \end{aligned}$$

Enfin

$$\begin{aligned}
 t_n(a-1) &= a^{n+1} - a \\
 t_n &= \frac{a^{n+1} - a}{a - 1} = \frac{a(a^n - 1)}{a - 1}.
 \end{aligned}$$

□

Cas spéciaux :

- $a = 2, i = 0 : \sum_{i=0}^n 2^i = \frac{a^{n+1} - 1}{a - 1} = \frac{2^{n+1} - 1}{2 - 1} = 2^{n+1} - 1.$
- $a = 2, i = 1 : \sum_{i=1}^n 2^i = \frac{a(a^n - 1)}{a - 1} = \frac{2(2^n - 1)}{2 - 1} = 2^{n+1} - 2.$
- $a = \frac{1}{2}, i = 0 : \sum_{i=0}^n \left(\frac{1}{2}\right)^i = \frac{a^{n+1} - 1}{a - 1} = \frac{\left(\frac{1}{2}\right)^{n+1} - 1}{\frac{1}{2} - 1} = 2 \left(1 - \left(\frac{1}{2}\right)^{n+1}\right) = 2 - \left(\frac{1}{2}\right)^n.$
Autrement dit, $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 2.$
- $a = \frac{1}{2}, i = 1 : \sum_{i=1}^n \left(\frac{1}{2}\right)^i = \left(\sum_{i=0}^n \left(\frac{1}{2}\right)^i\right) - 1 = 2 - \left(\frac{1}{2}\right)^n - 1 = 1 - \left(\frac{1}{2}\right)^n.$
Autrement dit, $\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots = 1.$
- $a = \frac{1}{k}, i = 0 :$

$$\sum_{i=1}^n \left(\frac{1}{k}\right)^i = \frac{\left(\frac{1}{k}\right)^{n+1} - 1}{\left(\frac{1}{k}\right) - 1} = \frac{1 - \left(\frac{1}{k}\right)^{n+1}}{1 - \left(\frac{1}{k}\right)} = \frac{1 - \left(\frac{1}{k}\right)^{n+1}}{\frac{k-1}{k}} = \frac{k}{k-1} \left(1 - \left(\frac{1}{k}\right)^{n+1}\right).$$

3. Escalier de Gabriel (*Gabriel's staircase*). Pour $a \neq 0$,

$$\sum_{i=1}^n ia^i = \frac{a^{n+1}(n(a-1)-1)+a}{(a-1)^2}.$$

et $\sum_{i=0}^n ia^i = \sum_{i=1}^n ia^i$, puisque $0a^0 = 0$.

Démonstration 3.3.3. Posons

$$g_n = \sum_{i=1}^n ia^i.$$

Encore une fois,

$$\begin{aligned} ag_n - g_n &= a \sum_{i=1}^n ia^i - \sum_{i=1}^n ia^i \\ g_n(a-1) &= \sum_{i=1}^n ia^{i+1} - \sum_{i=1}^n ia^i. \end{aligned}$$

En déroulant les sommes,

$$\begin{aligned} g_n(a-1) &= (a^2 + 2a^3 + \cdots + na^{n+1}) - (a + 2a^2 + \cdots + na^n) \\ &= (na^{n+1} + \cdots + 2a^3 + a^2) - (na^n + \cdots + 2a^2 + a) \\ &= na^{n+1} + (n-1)a^n - na^n + \cdots + a^2 - 2a^2 - a \\ &= na^{n+1} - a - a^2 - \cdots - a^n \\ &= na^{n+1} - (a + a^2 + \cdots + a^n) \\ &= na^{n+1} - \frac{a^{n+1} - a}{a - 1} \\ &= \frac{na^{n+1}(a-1) - a^n + 1 + a}{(a-1)} \\ &= \frac{a^{n+1}(n(a-1)-1)+a}{(a-1)} \end{aligned}$$

Finalement,

$$\begin{aligned} g_n(a-1) &= \frac{a^{n+1}(n(a-1)-1)+a}{(a-1)} \\ g_n &= \frac{a^{n+1}(n(a-1)-1)+a}{(a-1)^2}. \end{aligned}$$

□

3.4 Propriétés des produits

3.4.1 Notation

La notation

$$\prod_{i=a}^b f(i)$$

exprime le produit des valeurs que prend $f(i)$ en faisant varier i , l'indice, de a à b , inclusivement, par incrément de 1. C'est-à-dire :

$$\prod_{i=a}^b f(i) = f(a) \times f(a+1) \times f(a+2) \times \cdots \times f(b-2) \times f(b-1) \times f(b).$$

La notation est souvent étendue pour exprimer différemment comment l'indice prend ses valeurs. Par exemple,

$$\prod_{x \in X} g(x)$$

exprime que l'indice x prend tour à tour la valeur de chaque élément de l'ensemble X .

Les produits peuvent être imbriqués :

$$\prod_{i=a}^b \prod_{j=c}^d h(i, j) = \prod_{i=a}^b \left(\prod_{j=c}^d h(i, j) \right).$$

Ici aussi, pour chaque valeur de i, j va varier de c à d , inclusivement.

3.4.2 Lois

Avec la notation usuelle :

1. $\prod_{i=a}^b f(i) = \frac{\prod_{i=d}^b f(i)}{\prod_{i=d}^{a-1} f(i)}$, pour $d < a$.
2. $\prod_{i=a}^b f(i) = \left(\prod_{i=a}^c f(i) \right) \left(\prod_{i=c+1}^b f(i) \right)$, pour $a \leq c < b$.
3. $\prod_{i=a}^b f(i) = \prod_{j=0}^{b-a} f(j+a)$ (« glissement » des indices.)

4. $\prod_{i=a}^b f(i)g(i) = \left(\prod_{i=a}^b f(i)\right) \left(\prod_{i=a}^b g(i)\right).$
5. $\prod_{i=a}^b \prod_{j=c}^d f(i,j) = \prod_{j=c}^d \prod_{i=a}^b f(i,j). \text{ (Hoisting; renversement des indices.)}$

Avec la notation ensembliste :

1. $\prod_{x \in X} f(x)$, le produit sur les éléments de l'ensemble X .
2. $\prod_{u \in X \cup Y} f(u) = \frac{(\prod_{x \in X} f(x))(\prod_{y \in Y} f(y))}{\prod_{v \in X \cap Y} f(v)}$ (inclusion/exclusion.)
3. $\prod_{u \in X \setminus Y} f(u) = \frac{\prod_{x \in X} f(x)}{\prod_{u \in X \cap Y} f(u)}.$
4. $\prod_{u \in X \cap Y} f(u) = \frac{(\sum_{x \in X} f(x))(\sum_{y \in Y} f(y))}{\sum_{u \in X \cup Y} f(u)}.$

Produits, factorielles, et symboles de Pochhammer :

1. $\prod_{i=1}^n i = 1 \times 2 \times 3 \times \cdots \times (n-1) \times n = n! = \Gamma(n+1).$
2. $\prod_{i=a}^b i = a \times (a+1) \times (a+2) \times \cdots \times (b-1) \times b = \frac{\prod_{i=1}^b i}{\prod_{i=1}^{a-1} i} = \frac{b!}{(a-1)!}.$
3. $\prod_{i=0}^{a-1} (n+i) = n \times (n+1) \times (n+2) \times \cdots \times (n+a-1) = n^{\bar{a}} = n^{(a)}$, où $n^{\bar{a}}$ et $n^{(a)}$ sont deux notations pour la *factorielle ascendante*, et $n^{(a)}$ est le *symbole de Pochhammer* (voir § 8.4).
4. $\prod_{i=0}^{a-1} (n-i) = n \times (n-1) \times (n-2) \times \cdots \times (n-a+1) = n^{\underline{a}} = (n)_a$, où $n^{\underline{a}}$ et $(n)_a$ sont deux notations pour la *factorielle descendante*.
5. $\prod_{i=1}^n a^i = a^{\sum_{i=1}^n i} = a^{\frac{1}{2}n(n+1)}.$

3.5 Remarques bibliographiques

La notation \sum a été introduite par Joseph Fourier (1768–1830) [28, 48]. Fort pratique, cette notation s'impose rapidement [61]. En effet, calquée sur le s allongé des intégrales, \int , symbole d'une somme, le sigma majuscule, Σ , aussi une espèce de s, devient sa contrepartie discrète. Le grand « pi », \prod , le p majuscule grec, pour le produit a été introduit par Jacobi en 1829 [73].

*
* * *

Piskounov ne donne pas directement la formule générale pour la sommation de puissances, mais seulement une esquisse pour un cas particulier ($k = 2$) entre deux exercices [128, p. 69]. C'est encore un cas d'une preuve laissée au soin du lecteur! Le supplice vous est épargné ici, la démonstration complète se trouvant à la § 3.2.2. La formule de Faulhaber (1580–1635) paraît pour la première fois en 1631 [43]. Knuth analyse le résultat de Faulhaber et propose quelques formules équivalentes basées sur les coefficients binomiaux [87]. Pour en savoir plus sur le problème du calcul des sommes de puissances, consultez aussi Beardon [11] et Derby [31].

4

Nombres Complexes

Sommaire. Ce chapitre présente très rapidement les nombres complexes. Nous ne les rencontrons que rarement dans l'analyse des algorithmes, beaucoup plus en traitement de signal; leur connaissance est cependant essentielle. Nous terminons ce chapitre avec les remarques bibliographiques.

4.1 Forme cartésienne

Dans sa forme la plus simple, un nombre complexe s'exprime par la somme d'une partie réelle et d'une partie imaginaire :

$$z = a + i b$$

où $a, b \in \mathbb{R}$ et $i = \sqrt{-1}$. Ce n'est pas toujours i qui est utilisée, $\sqrt{-1}$ est parfois notée j — surtout dans les ouvrages de physique ou d'ingénierie. Dans leur forme cartésienne, les nombres complexes peuvent être interprétés comme des vecteurs liés à l'origine, comme le montre la fig. 4.1.1. Nous discuterons des valeurs de θ et r à la section 4.2.

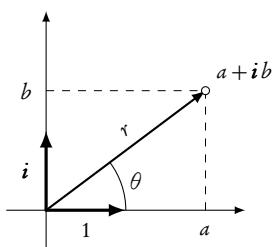


Figure 4.1.1 — Le plan complexe.

4.1.1 Opérations et lois

Les opérations les plus importantes sont :

- Le changement de signe : $-(a + ib) = -a - ib$.
- La conjugaison complexe : $\overline{a + ib} = a - ib$.
- Le module (ou valeur absolue) : $|a + ib| = \sqrt{a^2 + b^2}$.
- La partie réelle : $\operatorname{Re}(a + ib) = a$, parfois notée $\Re(z)$.
- La partie imaginaire : $\operatorname{Im}(a + ib) = b$, parfois on trouve la notation $\Im(z)$.
- L'addition : $(a + ib) + (c + id) = (a + c) + i(b + d)$.
- La multiplication : $(a + ib)(c + id) = (ac - bd) + i(bc + ad)$.

Démonstration 4.1.1. Nous utiliserons la distributivité de la multiplication et le fait que $i^2 = -1$:

$$\begin{aligned} (a + ib)(c + id) &= ac + iad + ibc + i^2bd \\ &= ac + iad + ibc + (-1)bd \\ &= ac - bd + i(ad + bc). \end{aligned}$$

□

- La division : $\frac{a + ib}{c + id} = \frac{ac + bd}{c^2 + d^2} + i \frac{bc - ad}{c^2 + d^2} = \frac{(a + ib)(c - id)}{c^2 + d^2}$.

Démonstration 4.1.2. Pour cette démonstration, il nous faudra résoudre un système d'équations, mais nous démontrerons l'inverse ce faisant. Nous cherchons l'inverse $x + iy$ tel que $(a + ib)(x + iy) = 1$, soit encore

$$(a + ib)(x + iy) = (ax - by) + i(ay + bx) = 1,$$

ce qui nous donne deux équations en deux inconnues,

$$\begin{aligned} ax - by &= 1 \\ ay + bx &= 0. \end{aligned}$$

On peut réécrire ce système d'équations comme

$$\begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

soit donc

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} &= \begin{bmatrix} a & -b \\ b & a \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{a^2 + b^2} \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{a^2 + b^2} \begin{bmatrix} a \\ -b \end{bmatrix} \end{aligned}$$

ce qui nous donne $x = \frac{a}{a^2 + b^2}$ et $y = \frac{-b}{a^2 + b^2}$. L'inverse est donc

$$x + iy = \frac{a}{a^2 + b^2} + i \frac{-b}{a^2 + b^2} = \frac{a - ib}{a^2 + b^2}$$

Enfin, puisque l'inverse de $c + id$ est donné par $\frac{c - id}{c^2 + d^2}$,

$$\begin{aligned} \frac{a + ib}{c + id} &= (a + ib)(c + id)^{-1} \\ &= (a + ib) \frac{c - id}{c^2 + d^2} \\ &= \frac{(a + ib)(c - id)}{c^2 + d^2} \\ &= \frac{ac + bd}{c^2 + d^2} + i \frac{bc - ad}{c^2 + d^2}. \end{aligned}$$

□

- L'inverse : $(a + ib)^{-1} = \frac{1}{a + ib} = \frac{a - ib}{a^2 + b^2}$. Donc pour un nombre complexe z , $z^{-1} = \frac{\bar{z}}{|z|^2}$.
- L'argument, $\arg(a + ib) = \tan^{-1} \frac{b}{a}$.

- La racine carrée, $\sqrt{a+ib} = \pm \sqrt{\frac{a \pm \sqrt{a^2+b^2}}{2}} \pm i \frac{b}{\sqrt{2}\sqrt{a \pm \sqrt{a^2+b^2}}}$, où les signes en couleur (\pm) sont les mêmes.

Démonstration 4.1.3.

On cherche $c+id$ tel que $(c+id)^2 = a+ib$, ce qui est équivalent à poser $c+id = \sqrt{a+ib}$. Or, $(c+id)^2 = c^2-d^2+i2cd = a+ib$ nous donne deux équations :

$$\begin{aligned} c^2 - d^2 &= a \\ 2cd &= b. \end{aligned}$$

En posant $u = \sqrt{\frac{a \pm \sqrt{a^2+b^2}}{2}}$, on trouve $c = \pm \frac{u}{\sqrt{2}}$ et $d = \pm \frac{b}{\sqrt{2}u}$. Enfin,

$$\sqrt{a+ib} = c+id = \pm \sqrt{\frac{a \pm \sqrt{a^2+b^2}}{2}} \pm i \frac{b}{\sqrt{2}\sqrt{a \pm \sqrt{a^2+b^2}}}.$$

□

4.2 Forme polaire

Si nous revenons à la fig. 4.1.1, nous voyons qu'un nombre complexe $a+ib$ peut aussi être exprimé en fonction d'un angle θ et d'une longueur r . Le nombre complexe prend alors la forme

$$z = re^{i\theta},$$

où $e = 2.7182818284\dots$ est la constante d'Euler, et $e^{ix} = \cos(x) + i \sin(x)$ la formule d'Euler.

4.2.1 Opérations et lois

Pour convertir d'une forme à l'autre :

- Pour un nombre complexe $z = x+iy$, $re^{i\theta}$ est tel que

- $r = |z| = \sqrt{x^2 + y^2}$,
- $\theta = \arg(z) = \tan^{-1} \frac{y}{x}$.

- Pour un nombre complexe $z = re^{i\theta}$, $x + iy$ est tel que
 - $x = r \cos \theta$,
 - $y = r \sin \theta$.

Pour $z = re^{i\theta}$, les opérations deviennent

- Le changement de signe : $-z = -re^{i\theta}$.
- La conjugaison complexe : $\overline{re^{i\theta}} = re^{-i\theta}$.
- Le module est $|z| = r$.
- La partie réelle : $\operatorname{Re}(re^{i\theta}) = r \cos \theta$.
- La partie imaginaire : $\operatorname{Im}(re^{i\theta}) = r \sin \theta$.
- L'argument est $\arg(z) = \theta$.
- La multiplication : $(r_1 e^{i\theta_1})(r_2 e^{i\theta_2}) = r_1 r_2 e^{i(\theta_1 + \theta_2)}$.
- La division : $\frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}$.
- L'addition/soustraction¹ : $r_1 e^{i\theta_1} \pm r_2 e^{i\theta_2}$.

4.3 Remarques bibliographiques

Les nombres complexes, en commençant par $\sqrt{-1}$, ont une histoire très intéressante [109]. Il faut se rappeler que le zéro a été difficilement accepté, que les nombres négatifs ont longtemps été considérés absurdes, alors vous devinez comment le chemin vers l'acceptation des nombres complexes a été sinueux ! Nonobstant, les nombres complexes sont maintenant au cœur de la science, du traitement de signal [30, 66, 77, 91, 147, 159] à la physique quantique [14, 83].

*
* * *

Le calcul efficace des opérations sur les nombres complexes est donc primordial pour un grand nombre d'applications. La démonstration 4.1.1 nous montre que le produit de deux nombres

1. L'addition/soustraction se simplifie qu'à travers les conversions de polaire vers cartésien, faire l'opération, puis convertir de nouveau vers polaire.

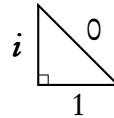


Figure 4.3.1 — Les nombres complexes sont parfois contre-intuitifs.

complexes demande quatre produits et deux additions (car dans l'ordinateur, le nombre $a + ib$ sera représenté comme la paire (a, b) , éliminant le besoin de multiplier par i et l'addition entre la partie réelle et la partie imaginaire). Ungar (vers 1963, par cité [88, p. 706]) propose de calculer

$$(a + ib)(c + id) = (ac - bd) + i((a + b)(c + d) - ac - bd)$$

qui ne demande plus que trois produits, mais demande 5 additions/soustractions. Cette méthode commence à être intéressante lorsque le coût de la multiplication est plus de trois fois celui d'une addition/soustraction.

*
* * *

La fig. 4.3.1 nous montre le « théorème » de Pythagore dans le plan complexe. Évidemment, mesurer la longueur de l'hypoténuse comme étant $i^2 + 1^2 = -1 + 1 = 0$ est fautif. La longueur de i est donnée par $|i| = \sqrt{0^2 + 1^2}$, puisque $i = 0 + 1i$. Il faut donc réécrire le théorème comme $|c| = \sqrt{|a|^2 + |b|^2}$.

*
* * *

Les nombres complexes sont en effet très contre-intuitifs et démontrent souvent des comportements surprenants. Par exemple, l'ensemble de Mandelbrot [97, 98] est l'ensemble de tous les points c du plan complexe tels que, à partir de $z_0 = 0$, la récurrence

$$z_t = z_{t-1}^2 + c$$

converge. Nous calculons un grand nombre d'itérations, et si nous avons toujours que $|z_t| \leq 2$, alors nous déterminons que c est membre de l'ensemble (avec une bonne probabilité). De fort jolies images, comme la fig. 4.3.2, peuvent être générées avec une assignation arbitraire de couleurs à partir du nombre d'itérations t nécessaires pour atteindre la décision, à savoir si le nombre c est membre de l'ensemble ou non. Ces images furent en vogue un certain moment (1985–1995, environ) [32, 44, 107, 133], mais sont le prélude à des mathématiques bien plus intéressantes qui posent des questions fondamentales sur la nature même d'une dimension géométrique et la complexité issue de règles simples [41, 42, 99, 143, 165].

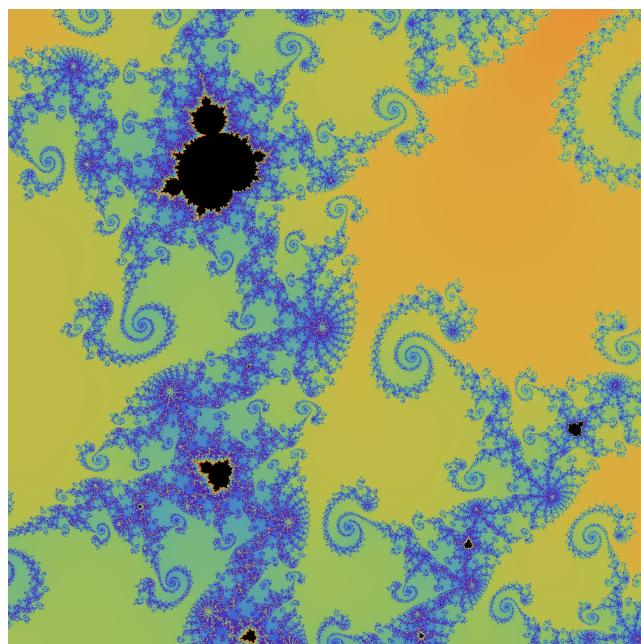


Figure 4.3.2 — L'ensemble de Mandelbrot, détail aux environs de $-0.7703 + i0.1112$, échelle 0.001.

5

Polynômes et racines

Sommaire. Dans ce chapitre, nous nous intéressons aux polynômes : leur structure, leur évaluation et leurs racines. Nous présenterons les solutions exactes au problème des racines pour les polynômes de degrés inférieurs à quatre et considérerons ensuite les méthodes numériques pour trouver les racines pour les polynômes et les expressions plus compliquées. Nous discuterons aussi brièvement de la factorisation. Nous terminerons, à chaque fois, sur les remarques bibliographiques.

5.1 Définitions

5.1.1 Structure et représentation

Un polynôme est une expression de la forme

$$\alpha_n x^n + \alpha_{n-1} x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0, \quad (5.1.1)$$

où $n \geq 0$, est un entier et où les $\{\alpha_i\}_{i=0}^n$ sont les *coefficients*. La valeur de n donne le *degré* du polynôme. On ne suppose rien de spécial sur les coefficients : ils peuvent être positifs, négatifs, ou nuls — sauf pour α_n qui est nécessairement différent de zéro. Les $\{\alpha_i\}_{i=0}^n$ sont, comme les x , tirés d'un anneau A ¹. Il importe peu que nous commençons par α_0 ou par $\alpha_n x^n$, mais l'ordre le plus courant est présenté par l'éq. (5.1.1), et pour un degré n , l'expression comporte $n + 1$ termes ; un polynôme de degré zéro est possible et ne contient qu'une constante, α_0 .

1. Un anneau est un ensemble muni de deux opérateurs, associés à $+$ et \times , qui commutatifs et tels que \times est distributif sur $+$. De plus, il existe un élément neutre pour chaque opération, un élément opposé pour $+$. Typiquement, nous considérons \mathbb{R} ou \mathbb{C} .

On peut exprimer un polynôme comme un produit scalaire :

$$\alpha^T x = (\alpha_n, \alpha_{n-1}, \dots, \alpha_2, \alpha_1, \alpha_0)^T (x^n, x^{n-1}, \dots, x^2, x, 1) = \alpha_n x^n + \alpha_{n-1} x^{n-1} + \dots + \alpha_2 x^2 + \alpha_1 x + \alpha_0.$$

Les polynômes se généralisent à un nombre quelconque de variables. En deux variables, l'expression devient

$$\alpha_{nm} x^n y^m + \alpha_{n,m-1} x^n y^{m-1} + \alpha_{n-1,m} x^{n-1} y^m + \dots + \alpha_{1,0} x + \alpha_{0,1} y + \alpha_{00} = \sum_{i=0}^n \sum_{j=0}^m x^i y^j, \quad (5.1.2)$$

et contient $(n+1)(m+1)$ termes. Le polynôme de l'éq. (5.1.2) est de degré $n+m$. Cette dernière formulation donne l'expression générale

$$\sum_{i=0}^{n_1} \sum_{j=0}^{n_2} \dots \sum_{k=0}^{n_m} \alpha_{i,j,\dots,k} x_1^i x_2^j \dots x_m^k,$$

dont l'expansion contient (au plus) $(n_1+1)(n_2+1)\dots(n_m+1)$ termes, et dont le degré est $n_1+n_2+\dots+n_m$.

5.1.2 Vocabulaire et notation

Clarifions quelques points de vocabulaire et de notation :

- Un monôme¹ est un « polynôme » avec un seul terme, de degré quelconque ax^k .
- Un binôme est un polynôme contenant deux termes, de la forme générale $ax^k + bx^l$.
- Un trinôme (sans surprise) contient trois termes, en général de degrés quelconques, $ax^k + bx^l + cx^m$. Souvent, nous nous intéresserons au cas spécial $ax^2 + bx + c$, le trinôme de second degré; qu'on dit aussi souvent « quadratique ».
- Ensuite, nous aurons polynôme pour le cas général avec plusieurs monômes.
- Les polynômes sont parfois notés de façon abrégée par $p(x)$, $p_d(x)$ pour un polynôme de degré d en x , etc.

1. Étymologiquement, -nôme vient du grec *nomos*, « partie », « division », voire même « province ». Par euphonie, nous avons *monôme* plutôt que *mononôme*.

- Il existe aussi des noms pour des formes spécifiques, plus rares, comme « quadrinôme », etc.
- On dira qu'un polynôme qu'il est tenu (anglais *sparse*, même parfois *fewnomial* [154]) lorsque le nombre de termes est petit par rapport à son degré; et dense sinon. Il y a évidemment un flou artistique dans cette définition; mais on peut deviner qu'un polynôme avec un petit degré ne peut être que dense.
- Un polynôme est *irréductible* lorsqu'on ne peut le factoriser en demeurant dans le même ensemble que celui dont sont tirés les coefficients. On peut quand même trouver une factorisation en étendant l'ensemble considéré (avec des radicaux, ou des nombres complexes, etc.). Nous en discuterons à nouveau à la § 5.4.
- Une racine¹ d'un polynôme est une valeur de la variable qui rend le polynôme nul. Autrement dit, x est une racine ssi $p(x) = 0$.

5.2 Évaluation

Comme nous utiliserons souvent les polynômes (pour interpoler comme pour tracer des courbes grâce aux différents types de *splines*), il devient intéressant de développer des méthodes pour les évaluer efficacement et, autant que possible, d'une façon numériquement stable. Si nous considérons un polynôme quelconque

$$\alpha_n x^n + \alpha_{n-1} x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha + 0,$$

une évaluation naïve nous mène à $O(n^2)$ multiplications. En effet, le terme $\alpha_j x^j$ demande $j - 1$ multiplications pour le calcul de x^j plus une pour le calcul de $\alpha_j \times x^j$. Comme j varie de 0 (on peut voir α_0 comme étant $\alpha_0 x^0$) à n , nous avons

$$\sum_{j=0}^n j = \sum_{j=1}^n = \frac{1}{2} n(n+1)$$

multiplications, c'est bien $O(n^2)$. C'est donc très coûteux dès que n est un peu grand, et bien inefficace si on a l'intention d'évaluer le polynôme un grand nombre de fois. On pourrait penser à utiliser l'exponentiation rapide qui permet de calculer x^k en $O(\lg k)$ multiplications². Si on calcule chaque

1. Un terme que l'on doit à notre ami Al-Khuwārizmī (c. 780–c. 850), un mathématicien perse, auteur du *Kitāb al-mukhtaṣar fi hisāb al-jabr wa-l-muqābala* (l'abrégué du calcul par la restauration et la comparaison), livre dont le titre nous a donné le mot *algèbre*. Une corruption de son nom nous a donné *algorithme*.

2. La décomposition binaire de l'exposant (entier) nous donne quelles puissances participent au produit. Par exemple, x^{25} peut s'écrire comme $x^{25} = x^{16}x^8x = (1 \cdot x^{16})(1 \cdot x^8)(0 \cdot x^4)(0 \cdot x^2)(1 \cdot x^1)$, et on peut vérifier que les coefficients des puissances correspondent bien à la décomposition binaire de l'exposant : $25_{10} = 11001_2$. De plus, les puissances de x se calculent avec des mises au carré successives, comme pour $(x^8)^2 = x^{16}$. Puisqu'il y a $O(\lg \alpha)$ bits dans l'exposant α , l'algorithme calcule l'exponentiation en $O(\lg \alpha)$ multiplications. Voyez la § 1.1.2 pour plus de détail.

monôme indépendemment, nous avons

$$\sum_{j=1}^n \lg j = \lg \prod_{j=1}^n j = \lg n!$$

produits. C'est beaucoup mieux que l'approche naïve en $O(n^2)$ puisque nous sommes maintenant en $O(n \lg n)$ — il suffit de savoir que $n! \approx \sqrt{2\pi n}(n/e)^n$ (par la formule de Stirling) et d'en tirer le logarithme. Cependant, en calculant chaque monôme indépendemment, nous nous privons encore d'un partage des calculs, puisqu'*a priori*, c'est plus facile de calculer x^k à partir de x^{k-1} qu'à partir de seulement x . Cela pourrait nous donner l'idée d'utiliser un « produit roulant », en commençant à x , puis en calculant x^2 , puis $x^3 = x^2 \times x$, et ainsi de suite jusqu'à $x^n = x^{n-1} \times x$. Cette stratégie nous amène au temps linéaire. Voyons : le premier terme, a_0 , ne demande aucun produit, le second, $a_1 x$ demande un seul produit, mais tous les autres en demandent deux (un pour calculer $x^j = x^{j-1} \times x$ et un autre pour calculer $a_j x^j$), ce qui nous fait $1 + 2(n - 1) = 2n - 1$ produits. C'est donc $O(n)$, mais un « gros » $O(n)$ puisque nous avons cette constante 2 qui multiplie la complexité.

5.2.1 Méthode de Horner

Enfin, si nous remarquons que nous pouvons écrire

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = (\cdots ((a_n x + a_{n-1}) x + a_{n-2}) x + \cdots + a_1) x + a_0,$$

nous obtenons un calcul en exactement n produits. C'est la forme de Horner [72], et si le polynôme n'a rien de spécial, la méthode est optimale [121].

5.2.2 Évaluation factorisée

Si par chance on dispose des $n + 1$ racines du polynôme, on peut exploiter le fait que

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = (x + r_0)(x + r_1)(x + r_2) \cdots (x + r_n)$$

pour calculer la valeur en n produits. Évidemment, si on ne dispose pas déjà des racines (soit qu'on ne les a pas de façon naturelle par le problème considéré, soit que le degré du polynôme est élevé et/ou irréductible) la méthode n'est guère pratique.

5.2.3 Formes spéciales

Si notre polynôme est de la forme

$$x^n + x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1,$$

on peut utiliser l'identité remarquable

$$x^n + x^{n-1} + x^{n-2} + \cdots + x^2 + x + 1 = \frac{x^{n+1} - 1}{x - 1},$$

que l'on peut vérifier grâce à une simple division polynomiale. Cette identité nous permet de calculer le polynôme en $O(\lg n)$ multiplication plus une division.

Si le polynôme est unitaire (on dit aussi normal), le premier coefficient est 1, c'est-à-dire que le polynôme est de la forme

$$p(x) = x^n + a_{n-1}x^{n-1} + a_{n-2}x^{n-2} + \cdots + a_2x^2 + a_1x + a_0.$$

Si de plus, le degré est de la forme $n = 2^k - 1$, on peut le décomposer en

$$p(x) = (x^{2^{k-1}} + a)q(x) + r(x),$$

et on applique récursivement la décomposition aux deux polynômes de degrés moindres $q(x)$ et $r(x)$. C'est la méthode de Belaga [12, 16].

Exemple 5.2.1.

Méthode de Belaga. Soit le polynôme unitaire $x^3 + ax^2 + bx + c$, dont le degré est de la forme $2^k - 1$, puisque $2^2 - 1 = 3$. On décompose :

$$x^3 + ax^2 + bx + c = (x^2 + \mu_1)(x + \mu_2) + (x + \mu_3).$$

En distribuant, on trouve que

$$\begin{aligned} a &= \mu_2 \\ b &= \mu_1 + 1 \\ c &= \mu_1\mu_2 + \mu_3 \end{aligned}$$

et (après un peu de travail) on trouve

$$\begin{aligned} \mu_1 &= b - 1 \\ \mu_2 &= a \\ \mu_3 &= c - a(b - 1), \end{aligned}$$

soit donc enfin

$$x^3 + ax^2 + bx + c = (x^2 + (b - 1))(x + a) + (x + (c - a(b - 1))).$$

On évalue le polynôme en deux multiplications seulement!

□

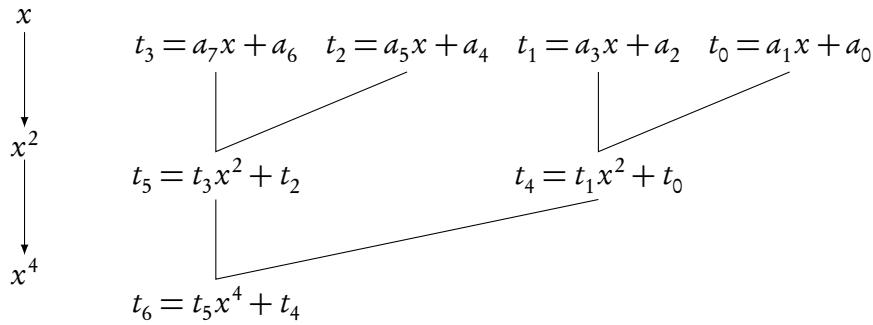


Figure 5.2.1 — Méthode d’Estrin pour l’évaluation parallèle des polynômes. Ici, le polynôme de degré 7, $a_7x^7 + a_6x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$.

Ce dernier exemple ne semble pas très convaincant, mais il faut voir que les expressions impliquant les constantes sont elles aussi constantes, et donc, précalculées. De plus, Belaga montre que le nombre espéré de multiplications pour évaluer un polynôme de degré $2^k - 1$ est au plus $\frac{1}{2}(n - 3) + \lg(n + 1)$ ce qui, étant donné n assez grand, devient très intéressant.

5.2.4 Évaluation parallèle

Enfin, dans le monde du futur, les ordinateurs sont capables d’exécution parallèle (soit en distribuant les calculs sur plusieurs cœurs, soit en exploitant des instructions de type SIMD¹), et nous devrions exploiter ce parallélisme pour évaluer les polynômes aussi rapidement que possible.

Estrin pose l’hypothèse d’un processeur sans limite de matériel où les calculs peuvent être réalisés grâce à une structure parallèle arborescente [39, 108]. On peut comprendre la méthode proposée par Estrin comme une généralisation parallèle du truc décrit à la note en bas de la page 37 (et à la § 1.1.2). La fig. 5.2.1 montre l’arbre des calculs effectués par la méthode d’Estrin. On décompose le polynôme en binômes, qui sont tous évalués en parallèle (en haut dans la figure). Ces binômes sont combinés pour donner des quadrinômes, qui sont combinés pour donner des octonômes, et ainsi de suite jusqu’à ce que l’expression combine toutes les valeurs. En parallèle, on calcule, à chaque étape, la valeur de x , de x^2 , de x^4 pour mettre les résultats à l’échelle. Le nombre de produits est toujours linéaire, mais la profondeur maximale est $O(\lg n)$, et comme nous exploitons le parallélisme, c’est la profondeur qui domine le temps de calcul !

*
* * *

1. SIMD, pour *single instruction, multiple data*, un type de parallélisme qui permet de multiples instances de la même instruction qui s’exécutent simultanément, mais en utilisant chacune des données différentes. Les processeurs modernes possèdent des jeux d’instructions SIMD riches et variés.

Remarquons que même si nous n'utilisons pas explicitement le parallélisme grâce aux instructions SIMD, un bon compilateur pourra paralléliser automagiquement une implémentation séquentielle de l'algorithme d'Estrin. Si notre compilateur n'est pas assez astucieux, nous pourrons toujours avoir recours aux primitives SIMD intrinsèques ou, pire, à l'écriture de la routine en assembleur.

5.3 Racines

Dans cette section, nous nous intéresserons au problème de trouver les zéros des polynômes, que l'on nomme *racines*. Nous cherchons donc les valeurs de x quelles que le polynôme $p(x)$ vaut zéro. Pour trouver ces racines, nous devons savoir, étant donné le degré du polynôme, combien il y en a.

Théorème 5.3.1.

Théorème fondamental de l'algèbre (aussi *théorème de d'Alembert-Gauss*). Tout polynôme à coefficients complexes de degré supérieur à 1 admet au moins une racine dans \mathbb{C} . \square

Une autre façon, équivalente, d'exprimer le théorème, c'est de dire que tout polynôme de degré n possède n racines dans \mathbb{C} , pas toutes nécessairement distinctes. Parfois, les racines sont dans \mathbb{R} , car $\mathbb{R} \subset \mathbb{C}$. Autrement dit, les n racines sont dans \mathbb{R} ou dans \mathbb{C} , et pas nécessairement distinctes.

5.3.1 Réduction à la forme homogène

Lorsqu'un polynôme se présente sous la forme

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = d$$

avec d , une constante différente de zéro, il suffit de le réécrire comme

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 - d = 0$$

soit encore

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a'_0 = 0$$

(en posant $a'_0 = a_0 - d$) pour le ramener à la forme homogène.

5.3.2 Degrés 0 et 1

Par le théorème fondamental de l'algèbre, un polynôme de degré 0, seulement composé d'une constante a_0 , aura zéro racine. Ce n'est donc pas un cas très intéressant à considérer.

Les polynômes de degré 1, de la forme

$$a_1x + a_0 = 0$$

que l'on retrouve plus souvent notés $ax + b$ ou encore $mx + b$, correspondent à l'équation d'une droite¹. Dans la forme $mx + b$, m est la pente (on trouve aussi parfois *coefficent directeur*) et b est l'ordonnée à l'origine, c'est-à-dire à quelle hauteur la droite coupe l'axe des y lorsque $x = 0$. Cependant, nous nous intéressons à l'endroit où la droite coupe l'abscisse, les x , lorsque $y = 0$. Il nous suffit d'isoler x :

$$\begin{aligned} ax + b &= 0 \\ ax &= -b \\ x &= -\frac{b}{a}. \end{aligned}$$

5.3.3 Second degré

Alors que la résolution des équations de degrés 0 et 1 est triviale, trouver les racines d'un polynôme de degré 2 demande un peu plus de travail. La solution générale pour un polynôme de la forme

$$ax^2 + bx + c = 0$$

est donnée par

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}, \quad (5.3.1)$$

formule qui est parfois nommée simplement « la quadratique ». Les valeurs que x peut prendre sont en général complexes (c'est-à-dire tirées de \mathbb{C}) mais souvent on considère que les solutions « intéressantes » sont réelles (ce qui ne pose pas de problème puisque $\mathbb{R} \subset \mathbb{C}$). Or, comme les solutions réelles nous intéressent plus souvent, on propose souvent l'analyse par la « méthode du discriminant », mais on peut aussi « compléter le carré », décaler à l'origine ou encore factoriser le polynôme.

1. On trouve aussi la forme *cartésienne*, $ax + by + c = 0$, qui se ramène à la forme désirée en isolant y , soit $y = -\frac{a}{b}x - \frac{c}{b}$, qui est bien de la forme $y = mx + b$.

5.3.3.1 Méthode du discriminant

Pour un polynôme

$$ax^2 + bx + c = 0$$

on pose le *discriminant* $\Delta = b^2 - 4ac$. Si...

- $\Delta > 0$, nous avons deux racines réelles et distinctes,

$$x = \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b + \sqrt{b^2 - 4ac}}{2a}$$

et

$$x = \frac{-b - \sqrt{\Delta}}{2a} = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

- $\Delta = 0$, nous n'avons qu'une seule solution, $x = -\frac{b}{2a}$, mais de multiplicité deux (ou, autrement dit, il y a deux racines de même valeur).
- $\Delta < 0$, au secondaire, on nous aura dit « il n'y a pas de solution », mais en fait, les deux racines seront complexes et conjuguées l'une de l'autre (voir § 4.1.1). Il est vrai qu'en général, ces solutions nous apparaissent souvent moins intéressantes, mais elles sont néanmoins valides.

5.3.3.2 Trouver les racines : compléter le carré

L'éq. (5.3.1) donne bien les racines d'un polynôme de second degré, mais elle semble « sortie de nulle part ». Dans cette section, nous allons dériver ce résultat, pas à pas.

On réduit d'abord notre polynôme de second degré

$$ax^2 + bx + c = 0$$

à la forme « monique » :

$$x^2 + \frac{b}{a}x + \frac{c}{a} = 0,$$

ce qui ne change pas la position des zéros, puisque ce n'est qu'une mise à l'échelle verticale. Nous allons compléter le carré pour obtenir un terme en x^2 seulement (par un raisonnement géométrique montré à la fig. 5.3.1), en remarquant qu'on peut écrire

$$x^2 + bx + c = \left(x + \frac{1}{2}b\right)^2 - \frac{1}{4}b^2 + c = 0,$$

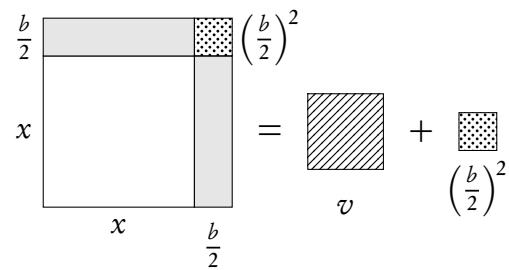
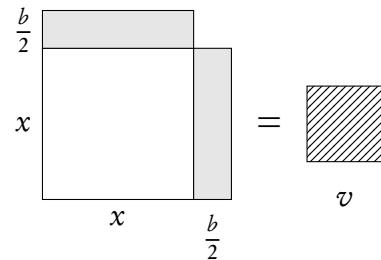
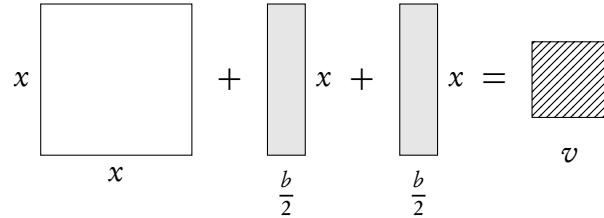
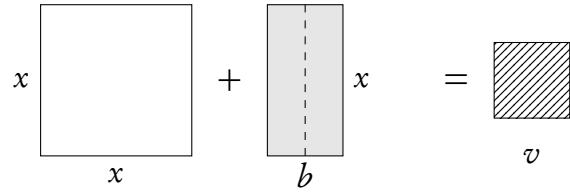


Figure 5.3.1 — Le raisonnement géométrique derrière la complémentation du carré. De haut en bas, la réorganisation de l'équation.

car on a bien $\left(x + \frac{1}{2}b\right)^2 - \frac{1}{4}b^2 + c = \left(x^2 + bx + \frac{1}{4}b^2\right) - \frac{1}{4}b^2 + c$.

Appliqué à la forme monique, on obtient

$$x^2 + \frac{b}{a}x + \frac{c}{a} = \left(x + \frac{1}{2}\frac{b}{a}\right)^2 - \frac{1}{4}\left(\frac{b}{a}\right)^2 + \frac{c}{a} = 0.$$

Il ne nous reste plus qu'à isoler x :

$$\begin{aligned}
 & \left(x + \frac{1}{2} \frac{b}{a} \right)^2 - \frac{1}{4} \left(\frac{b}{a} \right)^2 + \frac{c}{a} = 0 \\
 & \left(x + \frac{1}{2} \frac{b}{a} \right)^2 - \left(\frac{1}{2} \frac{b}{a} \right)^2 + \frac{c}{a} = 0 \\
 & \left(x + \frac{1}{2} \frac{b}{a} \right)^2 = -\frac{c}{a} + \left(\frac{1}{2} \frac{b}{a} \right)^2 \\
 & \left(x + \frac{b}{2a} \right)^2 = \frac{-4ac + b^2}{4a^2} \\
 & x + \frac{b}{2a} = \pm \sqrt{\frac{-4ac + b^2}{4a^2}} \\
 & x + \frac{b}{2a} = \pm \frac{\sqrt{b^2 - 4ac}}{2a} \\
 & x = -\frac{b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} \\
 & x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},
 \end{aligned}$$

ce qui est bien l'éq. (5.3.1). On pourra vérifier qu'on a bien

$$\left(x - \frac{-b - \sqrt{b^2 - 4ac}}{2a} \right) \left(x - \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right) = ax^2 + bx + c,$$

ce qui termine cette démonstration. ■

5.3.3.3 Trouver les racines : décalage à l'ordonnée

Une autre façon de dériver l'éq. (5.3.1) consiste à réduire le cas général d'un polynôme de second degré à un cas plus restrictif, plus facile à résoudre — et, il me semble, de façon bien plus intuitive que la complétion de carrés. Pour ce faire, nous allons ramener l'équation quadratique à l'ordonnée, comme nous le montre la fig. 5.3.2. Pour ramener la quadratique à l'ordonnée, il faut déplacer son sommet à $x = 0$, et donc, commencer par trouver où est ce sommet.

Considérons encore une fois le polynôme $ax^2 + bx + c$. Une quadratique n'a qu'un minimum/maximum global, et le calcul différentiel nous dit que ce minimum/maximum est tel qu'à ce point, la dérivée est zéro. Il nous faut donc résoudre

$$\frac{\partial}{\partial x} ax^2 + bx + c = 2ax + b = 0.$$

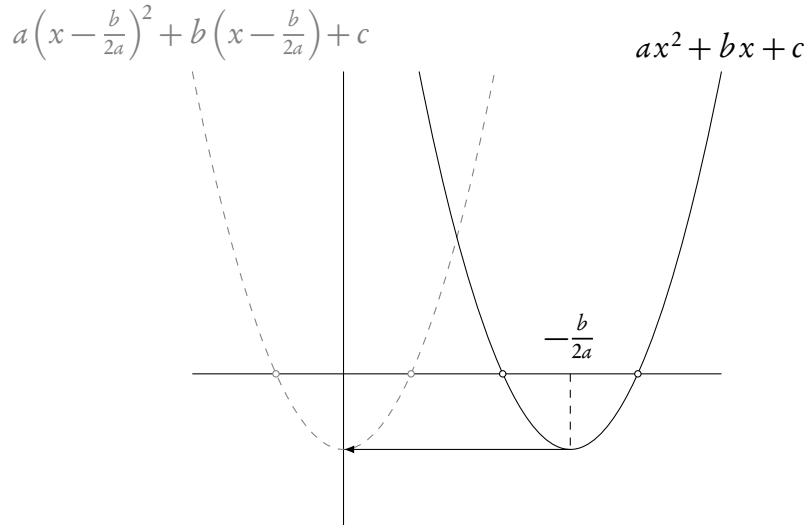


Figure 5.3.2 — Une quadratique décalée à l'ordonnée.

On trouve $x = -\frac{b}{2a}$. On ramène à la quadratique l'ordonnée en appliquant le décalage

$$a\left(x - \frac{b}{2a}\right)^2 + b\left(x - \frac{b}{2a}\right) + c = ax^2 - \frac{b^2}{4a} + c$$

puis on isole x :

$$\begin{aligned} ax^2 - \frac{b^2}{4a} + c &= 0 \\ ax^2 &= \frac{b^2}{4a} - c \\ x^2 &= \frac{b^2}{4a^2} - \frac{c}{a} \\ x^2 &= \frac{b^2 - 4ac}{4a^2} \\ x &= \pm \sqrt{\frac{b^2 - 4ac}{4a^2}} \\ x &= \pm \frac{\sqrt{b^2 - 4ac}}{2a}, \end{aligned}$$

ce qui nous donne les racines de la quadratique décalée sur l'ordonnée. Pour revenir à la quadratique originale, il faut inverser le décalage, et on obtient bien l'expression attendue,

$$x = \frac{-b}{2a} \pm \frac{\sqrt{b^2 - 4ac}}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

ce qui termine cette démonstration. ■

5.3.3.4 Trouver les racines : factorisation

Si par chance on peut trouver facilement les facteurs du polynôme, on obtient, par le fait même, les racines. Pour trouver les facteurs, on peut « forcer » la décomposition du polynôme

$$\begin{aligned} ax^2 + bx + c &= (u_1x + u_2)(u_3x + u_4) \\ &= u_1u_3x^2 + u_1u_4x + u_2u_3x + u_2u_4 \\ &= u_1u_3x^2 + (u_1u_4 + u_2u_3)x + u_2u_4 \end{aligned}$$

ce qui nous donne les trois équations

$$\begin{aligned} a &= u_1u_3 \\ b &= u_1u_4 + u_2u_3 \\ c &= u_2u_4 \end{aligned}$$

mais en *quatre* inconnues : il y aura un des u_i « libre », qu'on devra choisir arbitrairement.

Exemple 5.3.1.

$6x^2 + 7x - 5$. Nous avons $a = 6$ et comme $a = u_1u_3$, nous avons $a = 6 = 2 \times 3$ ou 3×2 . Posons $u_1 = 3$ et $u_3 = 2$. Nous avons $c = -5$ et comme $c = u_2u_4$, nous avons -1×5 ou 5×-1 . Il nous faut aussi $b = 7$, et comme $b = u_1u_4 + u_2u_3$, et que nous avons posé $u_1 = 3$ et $u_3 = 2$, nous avons $b = 3u_4 + 2u_2$; et poser $u_2 = 5$ et $u_4 = -1$ nous donne bien $b = u_1u_4 + u_2u_3 = 3(-1) + 2(5) = 7$. Puisque $u_1 = 3$, $u_2 = 5$, $u_3 = -2$, $u_4 = -1$, soit donc

$$6x^2 + 7x - 5 = (3x + 5)(2x - 1)$$

ce qui nous donne les racines

$$\begin{aligned} 6x^2 + 7x - 5 &= (3x + 5)(2x - 1) \\ &= 3\left(x + \frac{5}{3}\right)2\left(x - \frac{1}{2}\right) \\ &= 6\left(x + \frac{5}{3}\right)\left(x - \frac{1}{2}\right). \end{aligned}$$

Les racines sont donc $-\frac{5}{3}$ et $\frac{1}{2}$. □

Ce dernier exemple suggère donc qu'on peut mettre en évidence le coefficient du x^2 , c'est-à-dire utiliser le fait que

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x + \frac{c}{a}\right)$$

pour obtenir une factorisation (qu'on espère) plus facile,

$$ax^2 + bx + c = a \left(x^2 + \frac{b}{a}x + \frac{c}{a} \right) = a(x^2 + b'x + c') = a(x + u_1)(x + u_2),$$

ce qui nous donne plus que deux équations en deux inconnues :

$$\begin{aligned} u_1 + u_2 &= b' = \frac{b}{a} \\ u_1 u_2 &= c' = \frac{c}{a} \end{aligned}$$

qu'il nous faut résoudre par votre méthode préférée. S'il s'adonne que

$$u_1 = \frac{b \pm \sqrt{b^2 - 4ac}}{2a} = b' \pm \sqrt{b'^2 - 4c'}$$

et

$$u_2 = \frac{b \mp \sqrt{b^2 - 4ac}}{2a} = b' \mp \sqrt{b'^2 - 4c'}$$

(qui sont de signes opposés aux racines) se simplifient joliment, on trouve une factorisation sympathique en même temps que les racines.

5.3.4 Cubique

La résolution des équations cubiques est surprenamment compliquée si on la compare aux équations quadratiques. Le cas le plus général,

$$ax^3 + bx^2 + cx + d = 0,$$

demande de poser

$$\Delta_0 = b^2 - 3ac$$

et

$$\Delta_1 = 2b^3 - 9abc + 27a^2d,$$

puis

$$C = \sqrt[3]{\frac{\Delta_1 \pm \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}.$$

Avec $\rho = \frac{-1 \pm \sqrt{-3}}{2}$, une troisième racine de l'unité¹. Alors les trois racines, pas nécessairement distinctes ni toutes réelles, sont données par

$$x_k = -\frac{1}{3a} \left(b + \rho^k C + \frac{\Delta_0}{\rho^k C} \right),$$

que l'on obtient en posant $k = 0, 1, 2$.

1. Les n^{es} racines de l'unité sont des nombres tels que portés à la puissance n , ils valent 1.

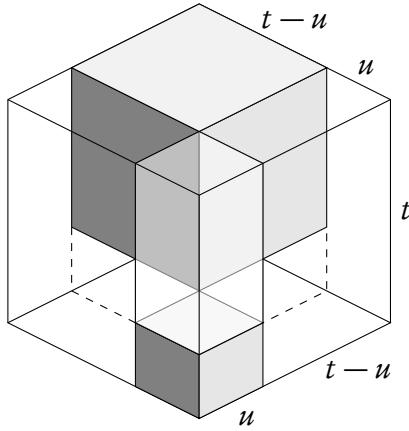


Figure 5.3.3 — Méthode de Cardan pour compléter le cube et résoudre l'équation cubique réduite $x^3 + mx = n$.

5.3.4.1 Trouver les racines : compléter le cube

Il est possible de résoudre une forme amoindrie de l'équation cubique, $x^3 + mx = n$ (ou dans la forme homogène $x^3 + mx - n = 0$), grâce à une construction due à Cardan, similaire dans le principe à la complétion de carré, mais cette fois-ci étendue au cube. Voyez la fig. 5.3.3. Le volume du cube de la fig. 5.3.3 est

$$t^3 = u^3 + (t - u)^3 + 2tu(t - u) + u^2(t - u) + u(t - u)^2,$$

où chacun des termes correspond à un des morceaux du cube complété. En simplifiant complètement, on peut réécrire l'équation précédente comme

$$t^3 - u^3 = (t - u)^3 + 3tu(t - u)$$

soit encore

$$(t - u)^3 + 3tu(t - u) = t^3 - u^3 \quad (5.3.2)$$

Cette équation a la forme $x^3 + mx = n$. En y substituant $x = t - u$, l'éq. (5.3.2) devient

$$x^3 + 3txu = t^3 - u^3,$$

ce qui nous permet de poser

$$m = 3tu$$

$$n = t^3 - u^3$$

Ce qui nous donne deux équations en deux inconnues, ce qui nous permet de trouver ce que valent t et u , puisqu'on suppose m et n donnés. On trouve rapidement $u = \frac{m}{3t}$, et donc

$n = t^3 - u^3 = t^3 - \frac{m^3}{27t^3}$, d'où il faut isoler t . En multipliant par t^3 des deux côtés,

$$\begin{aligned} n &= t^3 - \frac{m^3}{27t^3} \\ nt^3 &= t^6 - \frac{m^3}{27t^3}t^3 \\ &= t^6 - \frac{m^3}{27} \\ t^6 - nt^3 - \frac{m^3}{27} &= 0, \end{aligned}$$

nous arrivons à une forme quadratique. Certes, si on applique les parenthèses, cela devient évident :

$$(t^3)^2 - n(t^3) - \frac{m^3}{27} = 0,$$

il suffit de poser $x = t^3$ et nous avons une quadratique bien ordinaire, avec $a = 1$, $b = -n$ et $c = -\frac{m^3}{27}$.

En utilisant l'éq. 5.3.1, nous trouvons

$$t^3 = \frac{n \pm \sqrt{n^2 + \frac{4m^3}{27}}}{2} = \frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{4m^3}{27}}$$

et donc

$$t = \sqrt[3]{\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{4m^3}{27}}}.$$

Maintenant, puisque $u^3 = t^3 - n$,

$$\begin{aligned} u^3 &= t^3 - n \\ &= \frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{4m^3}{27}} - n, \end{aligned}$$

soit donc

$$u = \sqrt[3]{-\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{4m^3}{27}}}.$$

Enfin ! comme $x = t - u$, comme notre substitution le posait,

$$x = t - u = \sqrt[3]{\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{4m^3}{27}}} - \sqrt[3]{-\frac{n}{2} \pm \sqrt{\frac{n^2}{4} + \frac{4m^3}{27}}}. \quad (5.3.3)$$

Nous obtenons les solutions en variant \pm . ■

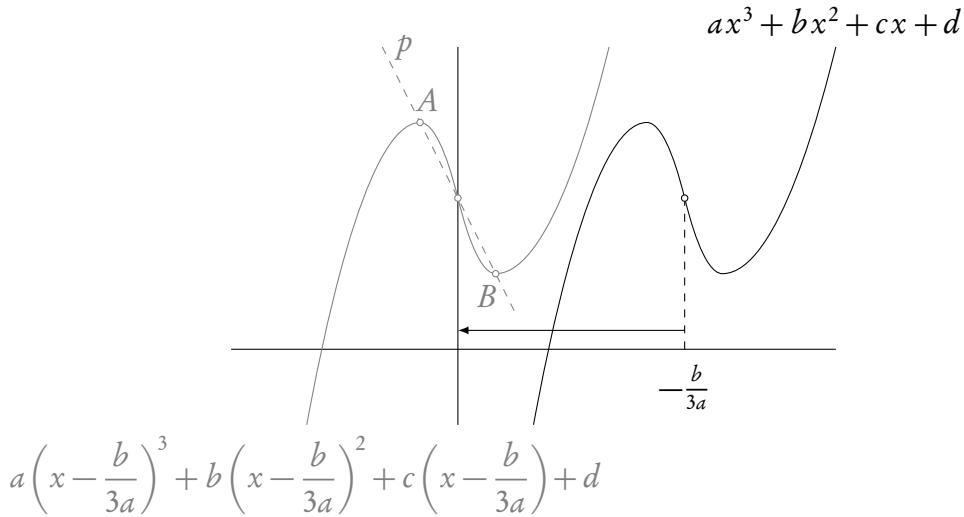


Figure 5.3.4 — Une cubique décalée à l'ordonnée par son point d'inflexion.

5.3.4.2 Trouver les racines : décalage à l'ordonnée

La première étape pour ramener une équation cubique à l'ordonnée, c'est de remarquer qu'elle est toujours symétrique par rapport à son point d'inflexion — rassurez-vous, nous n'en ferons pas la preuve ici. Or, le point d'inflexion, c'est où la dérivée seconde est nulle. Nous devons résoudre

$$\frac{\partial^2}{\partial x^2}ax^3 + bx^2 + cx + d = 6ax + 2b = 0.$$

On trouve $x = -\frac{b}{3a}$. On ramène le point d'inflexion à l'ordonnée pour obtenir

$$a\left(x - \frac{b}{3a}\right)^3 + b\left(x - \frac{b}{3a}\right)^2 + c\left(x - \frac{b}{3a}\right) + d = 0,$$

qu'on simplifie pour obtenir

$$ax^3 + \left(c - \frac{b^2}{3a}\right)x^2 + d - \frac{bc}{3a} + \frac{2b^3}{27a^3} = 0,$$

que l'on peut encore réduire en divisant par a (ce qui n'affecte que l'échelle verticale de la courbe) pour obtenir

$$x^3 + \left(\frac{-b^2 + 3ac}{3a^2}\right)x + \left(\frac{d}{a} - \frac{bc}{3a^2} + \frac{2b^3}{27a^3}\right),$$

qui est de la forme $x^3 + px + q$ (ce qui est la même chose que $x^3 + mx = n$, que nous avons utilisé dans la section précédente), avec

$$p = \frac{-b^2 + 3ac}{3a^2}$$

et

$$q = \frac{d}{a} - \frac{bc}{3a^2} + \frac{2b^3}{27a^3}.$$

Pour connaître le nombre de racines qu’admet l’équation, nous allons voir si les points A et B (dans la fig. 5.3.4) sont du même côté ou de côtés différents de l’abscisse. Heureusement, ils ne sont pas trop difficiles à trouver pour l’équation ramenée à l’ordonnée puisqu’ils sont des minimums/maximums locaux, que l’on trouve en cherchant où la dérivée (première) est zéro. Et la dérivée d’une équation cubique n’est qu’une quadratique, que nous pourrons résoudre très simplement ! Certes, il faut isoler x dans

$$\frac{\partial}{\partial x}x^3 + px + q = 3x^2 + p = 0,$$

où on trouve $x^2 = -\frac{p}{3}$, c’est-à-dire

$$x = \pm \sqrt{-\frac{p}{3}}.$$

Il suffit d’évaluer la cubique à ces deux points pour déterminer leur signe, et savoir combien de racines nous avons. Si les points sont de côtés opposés par rapport à l’abscisse, nous avons trois racines, du même côté, une seule, et si un point est sur l’abscisse tandis que l’autre est d’un côté ou de l’autre, deux.

Nous pouvons aussi tester autrement. Nous aurons

- une solution réelle si $\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 > 0$,
- deux solutions réelles si $\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 = 0$,
- trois solutions réelles si $\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3 < 0$.

*
* * *

Pour trouver les zéros, nous poursuivons avec la méthode de Cardan car nous avons montré que ramener la cubique à l’ordonnée nous donne le cas spécial réduit $x^3 + px + q = 0$. En complétant le cube, et en substituant cette fois p et q dans l’équation (plutôt que n et m), nous trouvons les racines. Nous pouvons aussi utiliser le théorème binomial :

$$\begin{aligned}(u + v)^3 &= u^3 + 3u^2v + 3uv^2 + v^3 \\ &= u^3 + 3uv(u + v) + v^3\end{aligned}$$

qu’on réorganise pour obtenir

$$(u + v)^3 - 3uv(u + v) - (u^3 + v^3) = 0.$$

En posant $x = u + v$, nous obtenons encore la forme $x^3 + px + q = 0$, avec le système d'équations

$$\begin{aligned}x &= u + v \\ uv &= -\frac{p}{3} \\ u^3 + v^3 &= -q,\end{aligned}$$

qu'il faut résoudre pour u et v (encore une fois!). On trouve

$$u = v = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

soit enfin les racines à

$$x = u + v = \sqrt[3]{-\frac{q}{2} \pm \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}} + \sqrt[3]{-\frac{q}{2} \mp \sqrt{\left(\frac{q}{2}\right)^2 + \left(\frac{p}{3}\right)^3}}$$

Comme nous avons $u = v$, les choix de \pm/\mp nous donnent trois jeux de solutions :

- — dans v , + dans u (ou vice-versa, puisque $u = v$);
- — dans les deux;
- + dans les deux.

On les ramène à la cubique initiale en rappliquant le décalage $-\frac{b}{3a}$. ■

5.3.5 Degrés supérieurs et méthodes numériques

La section précédente sur les racines de l'équation cubique vous a sûrement convaincu que trouver les racines des polynômes ne va qu'en empirant avec le degré du polynôme — et, à partir du quatrième degré, ce n'est même plus toujours possible d'exprimer les racines avec seulement les opérateurs de base et les racines [3]. Il nous faut donc des méthodes qui ne dépendent pas trop de la forme du polynôme, voire même applicables à une fonction arbitraire.

5.3.5.1 Méthode de la sécante

La méthode de la sécante utilise une estimation de la pente de la fonction pour trouver où la fonction croise l'abscisse. L'estimation de la pente se fait grâce à une *sécante*, une droite passant par deux points distincts de la fonction. Cette droite est prolongée de façon à croiser l'abscisse, et c'est ce lieu de croisement qui est l'estimation du zéro de la fonction. Évidemment, on ne tombe que rarement sur le zéro du premier coup, et on devra raffiner progressivement notre solution.

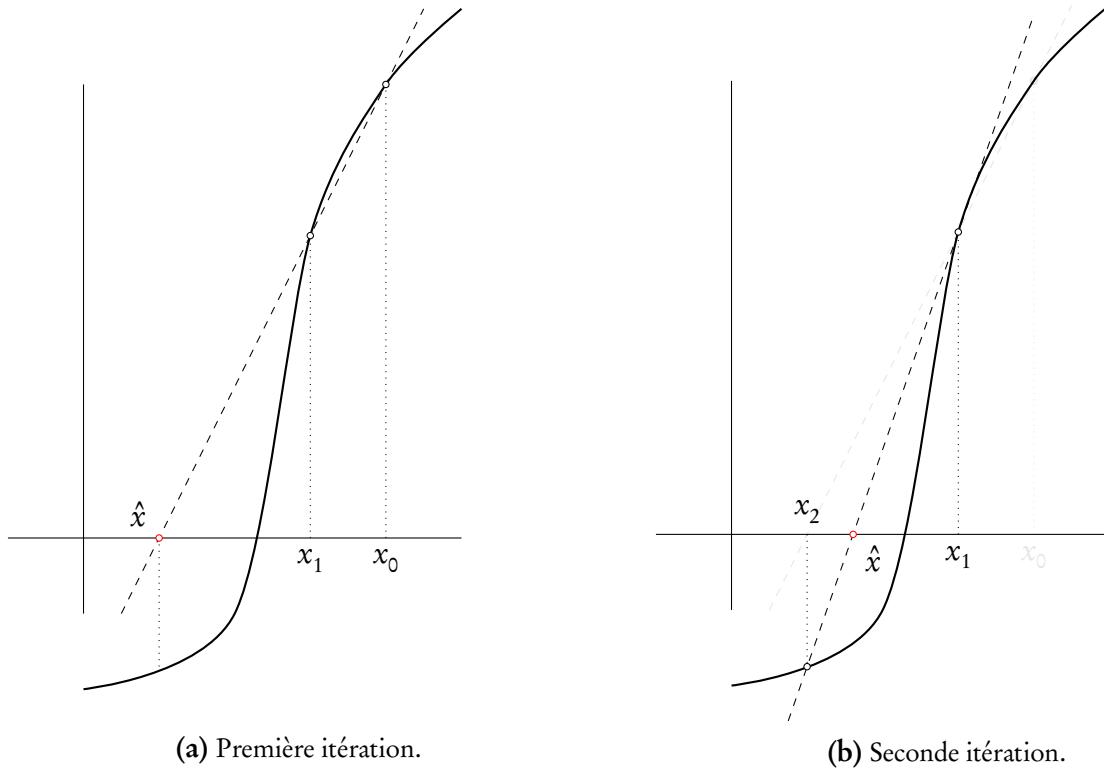


Figure 5.3.5 — Méthode de la sécante

La méthode procède donc ainsi : on choisit deux points, x_0 et x_1 , que l'on espère près du zéro de la fonction et préférablement (mais pas obligatoirement) tels que $f(x_0)$ et $f(x_1)$ sont de différents côtés de l'abscisse — c'est-à-dire tels que leurs signes sont contraires. On prolonge la droite passant par x_0 et x_1 et on trouve \hat{x} , le point où elle coupe l'abscisse. On vérifie alors le signe de $f(\hat{x})$. Si le signe de $f(\hat{x})$ est le même que le signe de $f(x_0)$, alors le croisement est entre \hat{x} et x_1 , et x_0 prend la valeur \hat{x} , sinon, c'est x_1 qui prend la valeur \hat{x} . On répète jusqu'à ce que $f(\hat{x})=0$ (ou encore $|f(\hat{x})|\leq \varepsilon$ pour un seuil ε). La fig. 5.3.5 montre deux itérations de cette procédure.

Détaillons cette méthode mathématiquement. La pente de la sécante est

$$m = \frac{y_1 - y_0}{x_1 - x_0} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}.$$

La sécante passant par x_0 et x_1 coupe l'axe des x au point \hat{x} tel que

$$f(x_1) - m(x_1 - \hat{x}) = 0,$$

et pour trouver \hat{x} , il suffit de l'isoler :

$$\begin{aligned} f(x_1) - m(x_1 - \hat{x}) &= 0 \\ f(x_1) &= m(x_1 - \hat{x}) \\ \frac{f(x_1)}{m} &= x_1 - \hat{x} \\ \frac{f(x_1)}{m} - x_1 &= -\hat{x} \\ \hat{x} &= x_1 - \frac{f(x_1)}{m}, \end{aligned}$$

et donc

$$\hat{x} = x_1 - f(x_1) \frac{x_1 - x_0}{f(x_1) - f(x_0)}, \quad (5.3.4)$$

qu'il ne reste plus qu'à simplifier selon la forme spécifique de $f(x)$.

*
* * *

Et si nous nous intéressons à autre chose que $f(x) = 0$? Il est peut-être intéressant de trouver x tel que $f(x) = y$ pour un y désiré. Voyons!

La pente est toujours $m = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$, mais maintenant nous cherchons \hat{x} tel que

$$f(x_1) - m(x_1 - \hat{x}) = y,$$

avec un y désiré. Encore une fois, isolons \hat{x} :

$$\begin{aligned} f(x_1) - m(x_1 - \hat{x}) &= y \\ f(x_1) - y &= m(x_1 - \hat{x}) \\ \frac{f(x_1) - y}{m} &= x_1 - \hat{x} \\ \frac{f(x_1) - y}{m} - x_1 &= -\hat{x} \\ \hat{x} &= x_1 - \frac{f(x_1) - y}{m}, \end{aligned}$$

soit donc

$$\hat{x} = x_1 - (f(x_1) - y) \frac{x_1 - x_0}{f(x_1) - f(x_0)}, \quad (5.3.5)$$

une expression fort peu différente de l'éq. (5.3.4)!

Exemple 5.3.2.

Racine carrée. Considérons l'application de la méthode de la sécante pour la résolution de l'équation $x^2 = n$. Il s'agit d'utiliser l'éq. (5.3.5) et de simplifier en fonction de la forme spécifique de notre équation. Nous avons

$$\begin{aligned}
 f(x_1) - m(x_1 - \hat{x}) &= y \\
 x_1^2 - \frac{x_1^2 - x_0^2}{x_1 - x_0}(x_1 - \hat{x}) &= y \\
 x_1^2 - \frac{(x_1 - x_0)(x_1 + x_0)}{x_1 - x_0}(x_1 - \hat{x}) &= y \\
 x_1^2 - (x_1 + x_0)(x_1 - \hat{x}) &= y - x_1^2 \\
 -(x_1 + x_0)(x_1 - \hat{x}) &= y - x_1^2 \\
 x_1 - \hat{x} &= -\frac{y - x_1^2}{x_1 + x_0} \\
 -\hat{x} &= -x_1 - \frac{y - x_1^2}{x_1 + x_0} \\
 \hat{x} &= x_1 + \frac{y - x_1^2}{x_1 + x_0}.
 \end{aligned}$$

L'algorithme devient, en posant $x_0 = n/2$, $x_1 = n/4$ et $\varepsilon = 10^{-6}$:

Répéter

$$\begin{aligned}
 \hat{x} &= x_1 + \frac{y - x_1^2}{x_1 + x_0} \\
 x_0 &= x_1 \\
 x_1 &= \hat{x}
 \end{aligned}$$

jusqu'à ce que $|x_0 - x_1| \leq \varepsilon$. Retourner x_1 .

En posant $n = 17$, l'algorithme produit :

t	x_0	x_1	\hat{x}	$\sqrt{n} - x$
0	8.5	4.25	4.1666...	-0.04356104...
1	4.25	4.1666...	4.1237...	-0.00065675...
2	4.1666...	4.1237...	4.1231...	$O(10^{-6})$
3	4.1237...	4.1231...	4.1231...	$O(10^{-10})$
4	4.1231...	4.1231...	4.1231...	$O(10^{-16})$

On voit que x_0 et x_1 convergent rapidement l'un vers l'autre et vers la solution. \square

*
* *

La méthode de la sécante a l'avantage de ne pas demander de connaître $f(x)$ dans le détail, et ainsi se prête bien à une implémentation qui prend cette fonction en argument. Cependant, elle demande quand même que les valeurs x_0 et x_1 soient assez près l'une de l'autre et que la fonction $f(x)$ ne varie pas de façon hystérique entre x_0 et x_1 , c'est-à-dire que la fonction $f(x)$ soit « lisse » localement. On remarque aussi que l'ordre de x_0 et x_1 importe peu. En effet, que nous ayons $x_0 < x_1$ ou $x_0 > x_1$, l'algorithme fonctionnera. On aurait l'intuition qu'il faille quand même avoir $x_0 \neq x_1$ pour s'éviter une division par zéro dans le calcul de la pente, mais les éqs (5.3.4) et (5.3.5) nous demandent plutôt de nous méfier de $f(x_0) = f(x_1)$!

Il faudra toutefois que le choix initial pour x_0 et x_1 ne soit pas trop loin du zéro (ou du y désiré) pour que l'algorithme fonctionne correctement. Mal choisis, la sécante peut s'éloigner plutôt que se rapprocher de la solution désirée. Le choix initial influe aussi sur la solution trouvée : il peut s'avérer difficile de trouver toutes les solutions.

5.3.5.2 Méthode de Newton

Approximer la pente par une sécante nous dispense de deux choses : de l'une, de connaître la fonction dont on veut trouver la racine, de l'autre de connaître le calcul différentiel. Or, s'il est vrai que calculer la pente de la sécante n'est qu'une question d'algèbre, pour calculer la dérivée, il faut d'abord inventer le calcul différentiel (voir § 5.5). Si l'on dispose du calcul différentiel, nous n'avons plus besoin d'estimer la pente : nous pouvons la connaître exactement avec la dérivée !

La méthode de Newton est très semblable dans son principe à la méthode de la sécante. On commence avec une première estimation du zéro, x_0 . On calcule la tangente à $f(x_0)$ (c'est-à-dire la dérivée $f'(x_0)$) et on la prolonge jusqu'à l'abscisse, et on trouve le point x_1 où elle la coupe. Ce point devient la nouvelle estimation du zéro. On répète jusqu'à ce que nous trouvions $f(x_t) = 0$ (ou suffisamment près, soit $|f(x_t)| \leq \varepsilon$, pour un seuil ε). La fig. 5.3.6 montre quelques itérations de la méthode.

Précisons la méthode mathématiquement. Soit $f(x)$, la fonction qui nous intéresse. Sa dérivée est donnée par

$$f'(x) = \frac{\partial f(x)}{\partial x}.$$

Prolonger la tangente en $f(x_t)$ jusqu'à l'abscisse revient à résoudre

$$f(x_t) - f'(x_t)(x_t - x_{t+1}) = 0$$

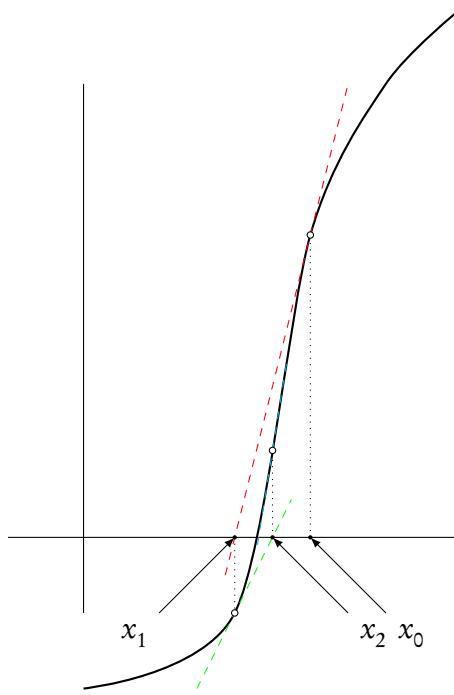


Figure 5.3.6 — La méthode de Newton. Les itérations sont codées par couleurs : rouge pour la première, verte pour la seconde, et bleu pour la troisième. Déjà, à la troisième itération, à cette résolution, l'approximation et la solution exacte se confondent.

pour x_{t+1} . Isolons x_{t+1} :

$$\begin{aligned} f(x_t) - f'(x_t)(x_t - x_{t+1}) &= 0 \\ f(x_t) &= f'(x_t)(x_t - x_{t+1}) \\ \frac{f(x_t)}{f'(x_t)} &= x_t - x_{t+1} \\ \frac{f(x_t)}{f'(x_t)} - x_t &= -x_{t+1} \end{aligned}$$

soit donc

$$x_{t+1} = x_t - \frac{f(x_t)}{f'(x_t)}. \quad (5.3.6)$$

Cette forme rappelle fortement l'éq. (5.3.4), et ce n'est pas accidentel : c'est en fait sa limite lorsque $|x_0 - x_1| \rightarrow 0$. L'éq. (5.3.6) est très simple et fonctionne bien numériquement tant que $|f'(x_t)|$ n'est pas zéro (et même pas trop près de zéro).

Exemple 5.3.3.

Racine carrée. Utilisons la méthode de Newton pour trouver la racine carrée d'un nombre n . Commençons par poser le problème sous la forme $f(x) = 0$. Comme on veut $x^2 = n$, il suffit de poser $f(x) = x^2 - n$ et de chercher x tel que $f(x) = 0$. On trouve

$$\frac{\partial f(x)}{\partial x} = \frac{\partial}{\partial x}(x^2 - n) = 2x.$$

On substitue $f(x) = x^2 - n$ et $f'(x) = 2x$ dans l'éq. (5.3.6) :

$$\begin{aligned} x_{t+1} &= x_t - \frac{f(x_t)}{f'(x_t)} \\ &= x_t - \frac{x_t^2 - n}{2x_t} \\ &= \frac{x_t}{2} - \frac{n}{2x_t} \\ x_{t+1} &= \frac{1}{2} \left(x_t - \frac{n}{x_t} \right). \end{aligned}$$

Nous avons l'itération et voyons comment elle se comporte numériquement avec $n = 17$, et $x_0 = \frac{n}{2} = \frac{17}{2}$:

t	x_t	$\sqrt{n} - x$
0	8.5	-4.37689437...
1	5.25	-1.12689437...
2	4.244047619...	-0.12094199...
3	4.124828859...	-0.00172323...
4	4.123105986...	-0.00000035...

La vraie valeur étant $\sqrt{17} = 4.1231056256176605498\dots$, on voit que l'algorithme converge très rapidement. Après 5 itérations, l'erreur est déjà $O(10^{-7})$. L'algorithme dépend cependant de la valeur initiale choisie. Si « par chance » on commence avec $x_0 = 4$, on obtient plutôt :

t	x_t	$\sqrt{n} - x$
0	4	0.123105625...
1	4.125	-0.00189437...
2	4.123106060...	$O(10^{-7})$
3	4.123105625...	$O(10^{-14})$
4	4.123105625...	$O(10^{-29})$

□

L'exemple montre que la méthode de Newton, lorsqu'on utilise un bon point de départ, converge extrêmement rapidement. Nous ne le démontrerons pas ici, mais l'algorithme de Newton possède une *convergence quadratique*, c'est-à-dire (si on simplifie un peu) que le nombre de chiffres significatifs double à chaque itération !

5.3.5.3 Méthode des points fixes

La méthode des points fixes (ou parfois du point fixe) est un algorithme qui est utilisé pour résoudre les problèmes de la forme générale

$$x = g(x),$$

pour une fonction $g(x)$ qui nous intéresse. La valeur x est dite *point fixe* de la fonction $g(x)$, ou encore l'*équilibre* de $g(x)$. Il est vrai qu'avec les polynômes, ce qui nous intéresse généralement, c'est de trouver les racines, c'est-à-dire résoudre un problème de la forme

$$f(x) = 0,$$

mais il est facile de transformer ce problème en problème des points fixes. Puisqu'on veut résoudre pour x plutôt que zéro, on ajoute x des deux côtés de l'équation

$$\begin{aligned} f(x) &= 0 \\ f(x) + x &= x \\ x &= f(x) + x \\ x &= g(x), \end{aligned}$$

ce qui pose notre problème sous une forme cromulente pour la résolution par la méthode des points fixes.

La méthode en elle-même est simple : on commence avec une estimation x_0 d'un des points fixes de la fonction, et on itère $x_{t+1} = g(x_t)$ jusqu'à convergence (c'est-à-dire où on trouve $x_t = x = g(x) = g(x_t)$ ou encore que $|x - g(x)| \leq \varepsilon$ pour un seuil ε) ou jusqu'à ce que la solution s'éloigne, c'est-à-dire diverge, ce que l'on détecte en comptant le nombre maximal d'itérations. Autrement dit : on pose ε , un seuil (qu'on assimilera à l'erreur tolérable), n le nombre maximal d'itérations, et x_0 une estimation initiale du point fixe. Soit donc :

On répète

$$x_{t+1} = g(x_t)$$

jusqu'à ce que $t \geq n$ ou qu'on ait

$$\frac{|x_{t+1} - x_t|}{|x_{t+1}|} \leq \varepsilon,$$

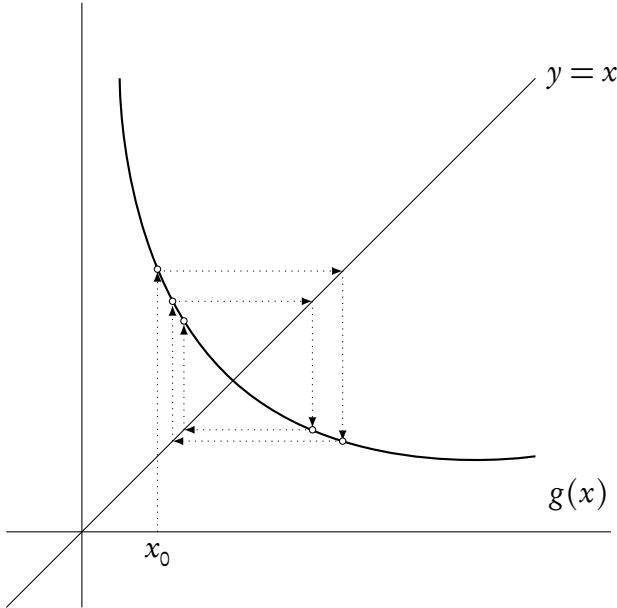


Figure 5.3.7 — La méthode des points fixes.

ce qui nous donnera x_t comme solution.

Une interprétation géométrique de la méthode est montrée à la fig. 5.3.7. À partir de notre estimation initiale x_0 , on calcule $x_1 = g(x_0)$, ce qui nous donne le prochain x à examiner. Une interprétation possible de cette manipulation, c'est de projeter *horizontalement* la solution $g(x_t)$ contre l'axe $y = x$ (sur lequel seront obligatoirement les points fixes, puisque $x = g(x)$ c'est $x = y = g(x)$) puis *verticalement* contre l'abscisse, l'axe des x , ce qui nous donne le prochain x à essayer.

La fig. 5.3.7 montre une spirale contractante, ce qui indiquerait que, pour cette fonction, la solution convergera vers un point fixe. Nous n'entrerons pas dans les détails des critères de convergence (vous pourrez les trouver ici [22, 46]), mais il suffit que la dérivée de $g(x)$ soit telle que $|g'(x)| < 1$. Si on a $|g'(x)| > 1$, la méthode diverge. La méthode est instable si $|g'(x)| = 1$ pour les x successifs.

De plus,

- Si $|g'(x)| < 1$ mais $g'(x) \neq 0$, la méthode converge linéairement ;
- Si $g'(x) = 0$ mais $g''(x) \neq 0$, la méthode a une convergence quadratique ;
- Si $g'(x) = g''(x) = 0$, mais que $g'''(x) \neq 0$, la méthode a une convergence d'ordre trois¹.
- Si les n premières dérivées sont nulles mais la $n + 1^{\text{e}}$ ne l'est pas, la convergence est d'ordre n .

1. On pourrait dire une « convergence cubique », mais le terme n'est pas utilisé dans la littérature.

Exemple 5.3.4.

Racine d'un polynôme. Posons le polynôme $x^2 - 3x - 5$ dont on veut trouver une racine. Bien qu'il soit facile d'utiliser l'équation quadratique pour trouver les racines, procédons grâce à la méthode des points fixes. Commençons par transformer le problème de $x^2 - 3x - 5 = 0$ en une forme compatible avec la méthode des points fixes.

Une première façon pourrait être d'isoler le x du terme $-3x$:

$$\begin{aligned} x^2 - 3x - 5 &= 0 \\ x^2 - 5 &= 3x \\ x &= \frac{x^2 - 5}{3}. \end{aligned}$$

Résoudre le problème original avec cette formule équivalente garantit la convergence puisque sa dérivée est inférieure à 1 :

$$\frac{\partial}{\partial x} \left(\frac{x^2 - 5}{3} \right) = \frac{2}{3}x,$$

et $\left| \frac{2}{3}x \right| < 1$.

Une seconde façon pourrait être de scinder le terme en x^2 :

$$\begin{aligned} x^2 - 3x - 5 &= 0 \\ x(x - 3) - 5 &= 0 \\ x(x - 3)5 & \\ x &= \frac{5}{x - 3}. \end{aligned}$$

Cette transformation aussi assure la convergence, car nous avons bien

$$\frac{\partial}{\partial x} \left(\frac{5}{x - 3} \right) = -\frac{5}{(x - 3)^2}.$$

Cette version converge plus rapidement que la précédente car la dérivée tends vers zéro : la convergence sera (au moins) quadratique !

On constate donc qu'il y a plus d'une façon de ramener le problème à la forme générale $x = g(x)$ à résoudre pour x , mais que certaines formes sont plus intéressantes que d'autres pour la convergence. \square

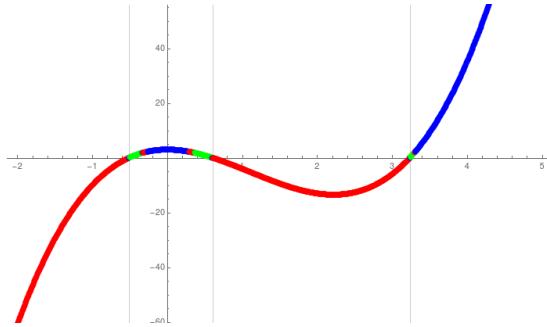


Figure 5.3.8 — La méthode de Newton et ses bassins d'attraction.

Exemple 5.3.5.

Racine carrée. Voyons si nous pouvons utiliser la méthode des points fixes pour trouver la racine carrée d'un nombre. Nous avons donc l'équation $x^2 = n$ à résoudre. Nous commençons par la transformer en problème de points fixes :

$$\begin{aligned} x^2 &= n \\ x^2 + x &= n + x \\ x^2 + x - n &= x \\ x &= x^2 + x - n. \end{aligned}$$

Or, après quelques itérations, on voit que même en choisissant judicieusement x_0 , les x_i divergent rapidement. Sans surprise, puisque

$$\frac{\partial}{\partial x} (x^2 + x - n) = 2x$$

Donc, pour la plupart des valeurs de x , nous avons $|g'(x)| > 1$: la méthode des points fixes ne peut pas converger. \square

5.3.5.4 Remarques sur les méthodes numériques

Les méthodes numériques pour trouver les racines des polynômes — plus généralement, des solutions aux équations, ou systèmes d'équations — sont souvent sujettes au choix du point de départ, même quand le problème paraît simple.

Prenons pour exemple une équation cubique que nous voudrions résoudre numériquement avec la méthode de Newton (prétendons que nous ne connaissons pas les solutions exactes données à la § 5.3.4). Tout dépendant du point de départ, nous pourrons trouver l'une ou l'autre racine. Voyez

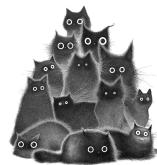
la fig. 5.3.8. Dans la figure, les points de départ sont codés en trois couleurs, qui correspondent aux racines qu'ils atteindront, de gauche à droite, les trois racines étant codées par les couleurs rouge, vert et bleu, respectivement. Ces plages de couleurs montrent les *bassins d'attraction* des racines, c'est-à-dire tous les points de départ qui mènent à la racine r_i . On voit que cela semble fort capricieux.

Les méthodes qui sont basées sur le gradient (la dérivée, la sécante, ou tout autre type d'estimation de la pente de la fonction) sont sensibles au point de départ parce que la pente pourrait faire en sorte que le prochain point examiné soit extrêmement loin de la solution. En effet, imaginez que par malchance vous choisissiez un point sur la fonction tel que la dérivée en ce point est zéro : l'intersection entre le prolongement de la pente et l'abscisse n'existe pas ! Si la dérivée est seulement *proche* de zéro, alors l'intersection entre le prolongement et l'abscisse sera quand même possiblement très éloigné, ce qui rendra la convergence beaucoup plus lente dans le meilleur des cas (et forcera une divergence sinon).

Il faut donc déjà avoir une bonne idée de la racine (ou de la solution) avant d'avoir recours à un algorithme de résolution numérique. Un bon estimé de départ n'est cependant pas toujours facile à obtenir. Traditionnellement, on utilise $x_0 = \frac{n}{2}$ pour trouver les racines carrées grâce à l'algorithme de Newton ou de la sécante. Ce cas particulier fonctionne bien parce que la fonction est convexe (elle a une belle forme de U), mais la fig. 5.3.8 nous montre que dès que la fonction est quelque peu sinuose, les choses se compliquent.



Il faut aussi se méfier de la précision réduite des calculs réalisés par ordinateur. Sauf si on utilise un logiciel comme *Maple* qui gère des nombres en précision arbitraires ou *Mathematica* qui peut les manipuler symboliquement, on se retrouve bien souvent avec une approximation rationnelle des réels, quelque chose comme `float` et `double` qui n'offrent qu'un nombre limité de chiffres précis. Les problèmes sont variés : impossibilité de représenter des nombres trop petits (mais différents de zéro) ou trop grands (mais pas infini), artefacts étranges avec les opérations arithmétiques (par exemple $a + b = a$ même si ni a ni b ne sont infinis ou zéro), etc. Le problème de la stabilité numérique des méthodes est bien étudié, mais pas très bien maîtrisé [5, 6, 120, 132].



5.4 Factorisation

La factorisation des polynômes est en général un problème difficile et il faudrait y consacrer plus que cette section. Le but d'une factorisation complète est de transformer un polynôme de façon à ce que nous ayons

$$a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 = (x + r_0)(x + r_1)(x + r_2) \cdots (x + r_n),$$

Une formule que nous avons déjà rencontrée à la § 5.2.2. Or, obtenir les racines, nous l'avons vu, en général, est plutôt difficile. Nous devons alors user d'astuces ou profiter d'une forme particulière du polynôme pour en simplifier la factorisation, sans compter que certains polynômes sont dits *irréductibles*.

Définition 5.4.1.

Un polynôme est dit *irréductible* s'il ne peut être factorisé en (au moins) deux polynômes (de degrés moindres) non constants. \square

Sous-entendu ici est que les coefficients des polynômes facteurs demeurent dans le même anneau (le même ensemble) que les coefficients du polynôme original. Ainsi, même si on peut factoriser $x^2 - x - 1 = (x + \frac{1}{2}(1 + \sqrt{5}))(x + \frac{1}{2}(1 - \sqrt{5}))$, on considérera tout de même $x^2 - x - 1$ comme irréductible puisque les coefficients des polynômes facteurs (dans \mathbb{R}) ne sont pas du même ensemble que les coefficients du polynôme original (dans \mathbb{Z}), et surtout que $\mathbb{R} \supset \mathbb{Z}$.

De plus, le théorème Abel-Ruffini [3] nous dit qu'en général, les polynômes de degrés $n \geq 4$ n'auront pas tous (et même la plupart) de factorisation algébrique (dont les coefficients s'expriment avec $+, -, \times, \div$ et $\sqrt{}$). Il faudra donc nous en remettre aux cas spéciaux et aux « identités remarquables ».

5.4.0.1 Identités remarquables

Voici quelques identités remarquables

- Polynômes du second degré :

$$\begin{aligned} ax^2 + bx + c &= \left(x + \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right) \left(x + \frac{-b - \sqrt{b^2 - 4ac}}{2a} \right) \\ ax^2 + c &= \left(x - i\sqrt{\frac{c}{a}} \right) \left(x + i\sqrt{\frac{c}{a}} \right) \\ ax^2 + bx &= x(ax + b) = ax \left(x + \frac{b}{a} \right). \end{aligned}$$

- Polynômes au carré :

$$(a + b)^2 = a^2 + b^2 + 2ab$$

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2(ab + ac + bc)$$

$$(a + b + c + d)^2 = a^2 + b^2 + c^2 + d^2 + 2(ab + ac + ad + bc + bd + cd)$$

Ici, on voit le motif émerger : d'abord la somme des carrés des variables, puis les $C(n, 2)$ combinaisons de variables, deux fois.

- Identité de Brahmagupta¹ :

$$(a^2 - nb^2)(c^2 - nd^2) = (ac + nbd)^2 - (ad + bc)^2.$$

- Identité de Sophie Germain² :

$$\begin{aligned} x^4 + 4y^4 &= (x^2 + 2y^2)^2 - 4x^2y^2 \\ &= (x^2 + 2y^2 - 2xy)(x^2 + 2y^2 + 2xy) \end{aligned}$$

- Identités de Lagrange :

$$(a^2 + b^2)(x^2 + y^2) = (ax + by)^2 + (ay - bx)^2$$

$$(a^2 + b^2 + c^2)(x^2 + y^2 + z^2) = (ax + by + cz)^2 + (ay - bx)^2 + (ax - cx)^2 + (bx - cy)^2$$

etc.

5.4.0.2 Factorisations connues

Sommes et différences de mêmes puissances :

- Sommes de mêmes puissances :

$$a + b = a + b \text{ (irréductible)}$$

$$a^2 + b^2 = a^2 + b^2 \text{ (irréductible dans } \mathbb{R}, (a + ib)(a - ib) \text{ dans } \mathbb{C})$$

$$a^3 + b^3 = (a + b)(a^2 - ab + b^2)$$

$$a^4 + b^4 = a^4 + b^4 \text{ (irréductible dans } \mathbb{Q}, (a^2 + \sqrt{2}ab + b^2)(a^2 - \sqrt{2}ab + b^2) \text{ sinon)}$$

$$a^5 + b^5 = (a + b)(a^4 - a^4b + a^2b^2 - ab^3 + b^4)$$

$$a^6 + b^6 = (a^2 + b^2)(a^4 - a^2b^2 + b^4)$$

$$a^7 + b^7 = (a + b)(a^6 - a^5b + a^4b^2 - a^3b^4 - ab^5 + b^6)$$

$$a^8 + b^8 = a^8 + b^8 \text{ (irréductible dans } \mathbb{Q}).$$

1. Brahmagupta (598–670) est un mathématicien et astronome indien.

2. Sophie Germain (1776–1831) est une mathématicienne française qui a dû faire une George Sand d'elle même et publier sous le pseudonyme d'Antoine Augustin Leblanc (ou Le Blanc) pour être prise au sérieux.

- Différences de mêmes puissances :

$$\begin{aligned}
 a - b &= a - b \text{ (irréductible)} \\
 a^2 - b^2 &= (a - b)(a + b) \\
 a^3 - b^3 &= (a - b)(a^2 + ab + b^2) \\
 a^4 - b^4 &= (a^2 - b^2)(a^2 + b^2) \\
 &\quad = (a - b)(a + b)(a^2 + b^2) \\
 a^5 - b^5 &= (a - b)(a^4 + a^3b + a^2b^2 + ab^3 + b^4) \\
 a^6 - b^6 &= (a^2 - b^2)(a^4 + a^2b^2 + b^4) \\
 &\quad = (a - b)(a + b)(a^2 - ab + b^2)(a^2 + ab + b^2) \\
 a^7 - b^7 &= (a - b)(a^6 + a^5b + a^4b^2 + a^3b^3 + a^2b^4 + ab^5 + b^6) \\
 a^8 - b^8 &= (a^2 - b^2)(a^2 + b^2)(a^4 + b^4) \\
 &\quad = (a - b)(a + b)(a^2 + b^2)(a^4 + b^4)
 \end{aligned}$$

Certaines formes se simplifient parfois encore un peu, même si ce n'est pas sous la forme d'une série de produits. Par exemple, $a^2 + ab + b^2$ peut se décomposer comme $(a + b)^2 - ab$.

5.5 Remarques bibliographiques

L'histoire de l'équation cubique remonte à la plus haute antiquité [34, 62]. D'abord avec Ménechme¹ (vers 380–320 av. J.-C.) qui montre que les sections coniques sont liées aux équations cubiques, puis avec Diophante d'Alexandrie (peut-être au II^e ou III^e siècle?) qui résout quelques cas spéciaux. Puis les savants arabes, vers l'an 1000, s'y intéresseront à leur tour : Abu Hafar al Hazin propose une solution à l'aide des sections coniques tandis qu'Abul Gud résout le cas spécial $x^3 - x^2 - x + 1 = 0$. Omar al Hay de Chorassan (c. 1079) pensait, quant à lui, qu'il était impossible de résoudre algébriquement les équations cubiques.

Le prochain chapitre de l'histoire se déroule dans l'Italie de la renaissance. Scipio Ferro († 1526) résout le cas spécial $x^3 + mx = n$. Nicolo de Brescia (c. 1506–1557), mieux connu comme Tartaglia (litt. « le bègue »), résout (vers 1530?) partiellement le cas $x^3 + px^2 = q$. Un étudiant de Ferro, un certain Floridas, dit aussi connaître cette forme (par l'enseignement de son maître). S'en suit un concours de qui pissoit le plus loin contre Tartaglia : celui qui connaît les vraies solutions des 30 problèmes proposés remportera le « duel » mathématique ! Finalement, c'est Tartaglia qui les résoudra

1. En grec, *Μένακμος*, Mēnaikmos.

tous tandis que Floridas n'en réussira aucun. Vers 1541, Tartaglia trouve la solution générale aux équations de la forme amoindrie $x^3 + px^2 = q$.

C'est Cardan qui résoudra le cas général (en utilisant la méthode présentée à la § 5.3.3.2). Connu en français comme Jérôme Cardan, Gerolamo Cardano (1501–1576) est un joyeux drille. Il déshérite un de ses fils pour avoir volé son grand-père, tandis que l'autre est décapité pour avoir empoisonné sa femme. Il quitte Pavie pour s'envier à Bologne (et y vivre une relation homosexuelle avec un de ses étudiants), a maille à partir avec l'inquisition... et quand la vie prenait un tour un peu trop agréable, il s'infligeait divers sévices pour se rappeler que la vie n'est que souffrance¹. Par ailleurs joueur compulsif, on lui doit le premier traité de probabilité [21]. Par la suite, Viète (1522–1565) et Descartes (1596–1650) arriveront à des solutions de type trigonométrique pour les équations cubiques [62, 117].

*
* * *

Il existe une littérature assez touffue sur l'évaluation des polynômes, et chacun y a été de sa proposition et de son analyse. Chose curieuse, beaucoup de ces développements datent des années 1960 [33, 36, 85].

*
* * *

Compléter le carré, ou le cube, nous apparaissent aujourd'hui comme des raisonnements tarabiscotés alors qu'il « suffit » de raisonner algébriquement pour arriver aux mêmes conclusions. Dunham [34, p. 144] suppose que c'est l'effet combiné de l'héritage grec — pour lesquels *tout* était géométrie — et de l'« algèbre rhétorique », seules façons d'exprimer les mathématiques au milieu du XVI^e siècle. L'algèbre rhétorique nous donne des formules (pour nous parfaitement opaques) du genre « du cube de la chose on soustrait la moitié de l'excédant de l'autre chose » — bonne chance pour deviner que c'est $x^3 - \frac{1}{2}(b-x)$.

*
* * *

Récemment, en décembre 2019, on a vu sur différents réseaux sociaux la nouvelle selon laquelle un professeur de mathématique avait inventé une nouvelle méthode géniale pour résoudre la

1. Comme se pincer au sang ou se tordre les doigts [34, p. 136–137]. J'ai lu ailleurs, mais sans retrouver la référence, qu'il se plantait un bâton dans l'œil jusqu'à ce qu'il en voit des étoiles, ou que ça saigne. *Good times indeed.*

quadratique [95]. C'est une variation sur la méthode par décalage (que nous avons présentée à la § 5.3.3.3), mais la méthode était connue depuis au moins les années 1990 [142]. Malgré le battage médiatique, Loh a reconnu la précédence de Savage et de façon fort gracieuse par ailleurs.

*
* * *

La méthode de la sécante prend son origine dans une technique beaucoup plus ancienne, la *regula falsi* (ou *fausse position*). Connue depuis la très haute antiquité, la *regula falsi* simple consiste à résoudre les problèmes linéaires par une règle de trois, et les problèmes non linéaires par la *regula falsi* double, soit une interpolation linéaire [51–53, 80, 81, 106].

*
* * *

Nous avons fait remarqué, à la § 5.3.5.2 que pour utiliser la dérivée (ou le gradient) d'une fonction, il fallait bien commencer par inventer le calcul différentiel. On ne sait trop encore aujourd'hui si on le doit à Newton ou à Leibniz. Leibniz publia bien avant Newton, en 1684 et 1686 [10, 151], tandis que ce dernier attendit 1693 et 1704 [20, 116]. Newton accusa Leibniz de l'avoir honteusement plagié, mais tous sont d'accord que Leibniz n'avait pas besoin de Newton pour inventer non seulement le calcul différentiel mais aussi une bien meilleure notation [57]. En effet, les intérêts de Leibniz étaient pointus et variés : politique, philosophie, physique, voir même les machines à calculer [9, 45, 47, 103, 105]. Ce qui, par ailleurs, ne les a pas empêchés de se chamailler comme des chiffonniers [10, 13, 64].

C'est peut-être le caractère austère et peu avenant de Newton qui a inspiré Gotlib d'en faire sa tête de Turc, lui faisant recevoir pommes, pots de fleurs, enclumes, chats, clefs anglaises, déchets divers, sofas, etc., sur la tronche (voir, par exemple, [60, p. 121]).



© Dargaud, 2018

6

Vecteurs et matrices

Sommaire. Nous présentons cette fois l'arithmétique des scalaires, vecteurs et matrices. Nous discuterons (un peu plus longuement) de la résolution de systèmes d'équations, en insistant sur les inverses, dont, en particulier, les pseudo-inverses.

6.1 Vecteurs et matrices

6.2 Notation

- a, b, c, \dots , (les « petites » lettres) dénoteront les constantes scalaires (réelles ou complexes).
- u, v, x, y, \dots , (les « grosses » lettres) dénoteront les vecteurs. Ils ont parfois noté en gras, \mathbf{x} , ou avec une flèche, $\vec{x}, \overrightarrow{x}$, voire même \underline{x} [84, 152]. À moins d'indications contraires, un vecteur est un vecteur-colonne.
- A, B, X , etc., dénoteront des matrices, et pour une matrice A qui est $m \times n$ (m rangées de n colonnes),
 - a_{ij} , avec $1 \leq i \leq m$ et $1 \leq j \leq n$, est un *coefficent* scalaire (réel ou complexe) de la matrice A .
 - a_{i*} , avec $1 \leq i \leq m$, est la i^{e} rangée de la matrice, un vecteur-rangée.
 - a_{*j} , avec $1 \leq j \leq n$, est la j^{e} colonne de la matrice, un vecteur-colonne.

- La transposée, notée par un T en exposant, s'applique aux vecteurs comme aux matrices :
 - x^T dénote un vecteur transposé ; si x est un vecteur-colonne, x^T est un vecteur-rangée ; de même, si x est un vecteur-rangée, x^T représente un vecteur-colonne.
 - A^T dénote une matrice transposée. Si A est $m \times n$, A^T est $n \times m$, et telle que $a_{ij}^T = a_{ji}$.
- L'inverse (multiplicatif) d'une matrice A (carré et non singulière) est noté A^{-1} , est tel que $A^{-1}A = AA^{-1} = I$, où I est la matrice identité (de même dimension que A et dont les coefficients de la diagonale sont 1, et zéro partout ailleurs). L'inverse ne s'applique habituellement pas aux vecteurs¹.

6.3 Vecteurs

Un vecteur u est un k -tuple de scalaires (réels ou complexes), noté

$$u = (u_1, u_2, \dots, u_k).$$

On dira qu'un vecteur contenant k scalaires est en k dimensions (ou de dimension k). Graphiquement, en deux ou trois dimensions, un vecteur peut être représenté comme une flèche dessinée sur un plan ou dans un volume², mais on peut les concevoir de façon plus abstraite.

6.3.1 Opérations

6.3.1.1 Addition/Soustraction

Pour deux vecteurs de même dimension u et v , nous avons

$$\begin{aligned} u \pm v &= (u_1, u_2, \dots, u_k) \pm (v_1, v_2, \dots, v_k) \\ &= (u_1 \pm v_1, u_2 \pm v_2, \dots, u_k \pm v_k). \end{aligned}$$

Nous avons donc que $u + v = v + u$, mais $u - v \neq v - u$.

1. Si on veut seulement trouver un vecteur inverse sous le produit scalaire (ce à quoi nous ne nous intéresserons pas), on pourra toujours utiliser $\frac{x}{\|x\|^2}$ comme inverse, et on vérifiera que $\frac{x^T x}{\|x\|^2} = 1$.

2. En quatre dimensions et plus, notre intuition spatiale nous est que de peu de secours. Comme dit Geoffrey Hinton : « *To deal with a 14-dimensional space, visualize a 3D space and say 'fourteen' to yourself very loudly* » (Pour manipuler un espace en quatorze dimensions, visualisez un espace en trois dimensions et dites-vous 'quatorze dimensions' très fort).

6.3.1.2 Produit avec un scalaire

Pour un vecteur u et une constante scalaire a , on a que

$$au = a(u_1, u_2, \dots, u_k) = (au_1, au_2, \dots, au_k).$$

6.3.1.3 Produit scalaire

Le produit scalaire (on dit aussi « *inner product* » en anglais) entre deux vecteurs-colonnes de même dimension est donné par

$$u^T v = (u_1 \ u_2 \ u_3 \ \cdots \ u_k) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_k \end{pmatrix} = \sum_{i=1}^k u_i v_i.$$

Le produit scalaire est

- Commutatif : $u^T v = v^T u$.
- Commutatif avec un scalaire : $au = ua$.
- Distributif : $u^T(v + w) = u^T v + u^T w$.
- Distributif avec un scalaire : $a(u + v) = au + av$.
- Associatif avec un scalaire : $au^T b v = ab u^T v$.
- Non-associatif : en général, $u^T(v^T w) \neq (u^T v)w$.

Il possède aussi des propriétés géométriques intéressantes :

- Le produit scalaire est nul ($= 0$) lorsque les vecteurs sont orthogonaux, c'est-à-dire $u^T v = 0$ ssi u et v sont perpendiculaires.
- Le produit scalaire nous informe sur l'angle que forment deux vecteurs u et v dans le plan qui les contient¹ :

$$u^T v = \|u\| \|v\| \cos \theta,$$

où $\|u\| = \sqrt{u^T u}$ est la longueur de u . En isolant θ , on a donc que

$$\cos \theta = \frac{u^T v}{\|u\| \|v\|}.$$

1. Deux vecteurs, peu importe le nombre de dimensions sont toujours dans le même (hyper)plan.

Démonstration 6.3.1. *Formule de l'angle.* Pour deux vecteurs u et v , posons

$$\begin{aligned} u_x &= \|u\| \cos \theta_u, \\ u_y &= \|u\| \sin \theta_u, \end{aligned}$$

et

$$\begin{aligned} v_x &= \|v\| \cos \theta_v, \\ v_y &= \|v\| \sin \theta_v, \end{aligned}$$

avec les angles θ_u et θ_v relatifs au plan contenant les vecteurs u et v . Nous avons alors dans ce plan que

$$\begin{aligned} u^T v &= (\|u\| \cos \theta_u, \|u\| \sin \theta_u) \begin{pmatrix} \|v\| \cos \theta_v \\ \|v\| \sin \theta_v \end{pmatrix} \\ &= \|u\| \|v\| (\cos \theta_u, \sin \theta_u) \begin{pmatrix} \cos \theta_v \\ \sin \theta_v \end{pmatrix} \\ &= \|u\| \|v\| (\cos \theta_u \cos \theta_v + \sin \theta_u \sin \theta_v) \\ &= \|u\| \|v\| \cos(\theta_u - \theta_v) \\ &= \|u\| \|v\| \cos \theta. \end{aligned}$$

□

- Le produit scalaire permet la projection d'un vecteur contre un autre. Voyez la section 6.7.

6.3.1.4 Produit dyadique

Le produit dyadique (« *outer product* » en anglais) de deux vecteurs u et v (possiblement de dimensions différentes) est tel que

$$u \otimes v = uv^T = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix} (v_1, v_2, \dots, v_p) = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_p \\ u_2 v_1 & \cdots & \cdots & u_2 v_p \\ \vdots & & & \vdots \\ u_k v_1 & u_k v_2 & \cdots & u_k v_p \end{pmatrix}.$$

Le produit dyadique possède les propriétés suivantes :

- Sous la transposée : $(u \otimes v)^T = v \otimes u$.

- Distributivité : $u \otimes (v + w) = (u \otimes v) + (u \otimes w)$ et $(u + v) \otimes w = (u \otimes w) + (v \otimes w)$.
- Multiplication par une constante scalaire : $c(u \otimes v) = (cu) \otimes v = u \otimes (cv)$.
- Associativité : $(u \otimes v) \otimes w = u \otimes (v \otimes w)$.

6.3.1.5 Produit croisé

Contrairement aux autres types de produits, le produit croisé (ou produit vectoriel) ne fonctionne qu'en trois dimensions. Nous avons déjà fait remarqué que deux vecteurs — peu importe le nombre de dimensions — sont toujours coplanaires. Le produit croisé calcule un vecteur perpendiculaire au plan contenant les deux vecteurs u et v , et ce vecteur est de longueur égale à l'aire du parallélogramme dont les côtés (non parallèles) sont donnés par u et v (nous y reviendrons à la § 6.6).

Le produit croisé (*cross product* en anglais) est donné par

$$u \times v = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = \mathbf{i} \begin{vmatrix} u_2 & u_3 \\ v_2 & v_3 \end{vmatrix} + \mathbf{j} \begin{vmatrix} u_1 & u_3 \\ v_1 & v_3 \end{vmatrix} + \mathbf{k} \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix} \quad (6.3.1)$$

où \mathbf{i} , \mathbf{j} , et \mathbf{k} sont les vecteurs de base standard (que l'on peut comprendre comme les vecteurs dans la direction des x , des y et des z dans un référentiel ordinaire), et la notation $|\cdots|$ désigne le déterminant (voir la § 6.6). On peut aussi écrire le produit croisé comme le produit matrice/vecteur

$$u \times v = \begin{pmatrix} \mathbf{i} \\ \mathbf{j} \\ \mathbf{k} \end{pmatrix} = \begin{pmatrix} 0 & -u_3 & u_2 \\ u_3 & 0 & -u_1 \\ -u_2 & u_1 & 0 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} = \begin{pmatrix} u_2 v_3 - u_3 v_2 \\ u_1 v_3 - u_3 v_1 \\ u_1 v_2 - u_2 v_1 \end{pmatrix}.$$

Le produit croisé est tel que

- $u \times u = 0$,
- $u \times v = -v \times u$ (il n'est donc pas tout à fait commutatif),
- $u \times (v + w) = (u \times v) + (u \times w)$,
- pour une constante a , $a u \times v = u \times av = a(u \times v)$.

*
* * *

Mais dans quelle direction pointe le vecteur issu de $u \times v$? Nous pouvons utiliser l'astuce de la « règle de la main droite ». Si on place u sur l'index, v sur le majeur, alors le pouce pointe dans la direction $u \times v$, comme nous le montre la fig. 6.3.1.

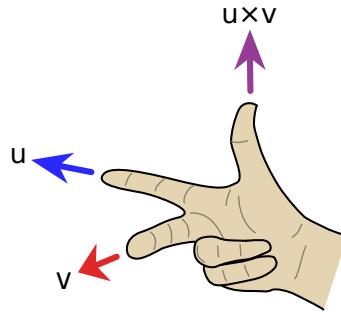


Figure 6.3.1 — La règle de la main droite. Image : Modifiée de Wikipedia.

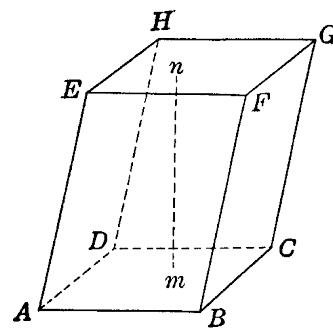
6.4 Matrices

Une matrice est un tableau de scalaires (réels ou complexes) de m rangées de n colonnes (on dira une matrice $m \times n$), de la forme générale

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1,n-1} & a_{1n} \\ a_{21} & a_{22} & \dots & & \dots & a_{2n} \\ \vdots & & & & & \vdots \\ a_{m1} & a_{m2} & \dots & \dots & a_{m,n-1} & a_{mn} \end{pmatrix}$$

où

- Le coefficient a_{ij} , avec $1 \leq i \leq m$ et $1 \leq j \leq n$, est l'élément scalaire à la rangée i et la colonne j de la matrice. Il est habituel de noter un élément a_{ij} sans virgule pour séparer les indices lorsqu'il n'y a pas d'ambiguïté possible, mais de les séparer par une virgule lorsque c'est nécessaire, par exemple $a_{m-1,n-2}$.
- a_{i*} , avec $1 \leq i \leq m$, est la i^e rangée de la matrice, un vecteur-rangée.
- a_{*j} , avec $1 \leq j \leq n$, est la j^e colonne de la matrice, un vecteur-colonne.



6.4.1 Opérations

6.4.1.1 Addition/Soustraction

Pour deux matrices A et B de même dimensions, nous avons

$$\begin{aligned} A \pm B &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix} \pm \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & \cdots & \cdots & b_{2n} \\ \vdots & & & \vdots \\ b_{m1} & \cdots & \cdots & b_{mn} \end{pmatrix} \\ &= \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & \cdots & \cdots & a_{2n} + b_{2n} \\ \vdots & & & \vdots \\ a_{m1} + b_{m1} & \cdots & \cdots & a_{mn} + b_{mn} \end{pmatrix}, \end{aligned}$$

c'est donc une addition/soustraction entrée-par-entrée, comme pour les vecteurs.

6.4.1.2 Produit avec un scalaire

Pour une matrice A , $m \times n$ et une constante scalaire c , nous avons

$$cA = c \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1n} \\ ca_{21} & \cdots & \cdots & ca_{2n} \\ \vdots & & & \vdots \\ ca_{m1} & \cdots & \cdots & ca_{mn} \end{pmatrix}.$$

6.4.1.3 Produit avec un vecteur

Pour une matrice A , $m \times n$ et un vecteur-colonne v de dimension n ,

$$Av = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} a_{1*}v \\ a_{2*}v \\ \vdots \\ a_{m*}v \end{pmatrix},$$

où $a_{i*}v$ est le produit scalaire de la i^{e} rangée de A et le vecteur v . Il faut donc que le nombre de colonnes dans la matrice ($m \times n$) soit le même que le nombre de rangées dans le vecteur-colonne (n). Le vecteur qui en résulte est de dimension m .

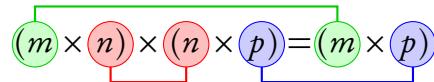
6.4.1.4 Produit avec une autre matrice

Soient une matrice A , $m \times n$ et une matrice B , $n \times p$. Alors

$$\begin{aligned} AB &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1p} \\ b_{21} & \cdots & \cdots & b_{2p} \\ \vdots & & & \vdots \\ b_{n1} & \cdots & \cdots & b_{np} \end{pmatrix} \\ &= \begin{pmatrix} a_{1*} b_{*1} & a_{1*} b_{*2} & \cdots & a_{1*} b_{*p} \\ a_{2*} b_{*1} & \cdots & \cdots & a_{2*} b_{*p} \\ \vdots & & & \vdots \\ a_{m*} b_{*1} & \cdots & \cdots & a_{m*} b_{*p} \end{pmatrix}, \end{aligned}$$

où $a_{i*} b_{*j}$ est le produit scalaire entre la i^{e} rangée de la matrice A et la j^{e} colonne de la matrice B .

Pour que le produit soit valide, il faut que les matrices soient compatibles : si A est $m \times n$, B doit être $n \times p$, et le résultat est une matrice $m \times p$, c'est-à-dire :



De plus, si

- x est un vecteur-colonne de dimension n ,
- A est une matrice $m \times n$,
- B est une matrice $p \times q$,

Nous aurons que

- Ax est $(m \times n) \times (n \times 1) = (m \times 1)$,
- $x^T A^T$ est $(1 \times n) \times (n \times m) = (1 \times m)$,
- AB est $(m \times n) \times (p \times q) = (m \times q)$, ssi $n = p$,
- ABx est $(m \times n) \times (p \times q) \times (n \times 1) = (m \times q) \times (n \times 1) = (m \times 1)$, ssi $n = p$ et $n = q$,

Le produit de matrice est

- Associatif : $ABC = A(BC) = (AB)C$,
- Non-commutatif : en général $AB \neq BA$,
- Distributif : $A(B + C) = AB + AC$, $(B + C)A = BA + CA$.

6.4.1.5 Transposées

Pour une matrice A de dimensions $m \times n$, la transposée est donnée par

$$A^T = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & \cdots & & \cdots & a_{2n} \\ a_{31} & \cdots & & \cdots & a_{3n} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{m-1,n} & a_{mn} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} & a_{31} & \cdots & a_{m1} \\ a_{12} & \cdots & & \cdots & a_{m2} \\ a_{13} & \cdots & & \cdots & a_{m3} \\ \vdots & & & & \vdots \\ a_{1n} & a_{2n} & \cdots & \cdots & a_{mn} \end{pmatrix}$$

Donc, pour une matrice A , A^T et telle que $a_{ij}^T = a_{ji}$.

Par ailleurs,

- La transposée d'un produit $(AB)^T$ est B^TA^T .
- La matrice adjointe, notée A^H (on dira alors aussi hermitienne), A^* ou même \bar{A} , est une matrice transposée dont les éléments sont aussi conjugués (voir § 4.1.1).

6.4.1.6 Inverses

L'inverse d'une matrice A ($m \times n$), noté A^{-1} , est une matrice $n \times m$ telle que

- Si la matrice A est « haute » ($m \times n$ avec $m > n$), il existe un inverse à gauche, $A^{-1}A = I$.
- Si la matrice A est « large » ($m \times n$ avec $m < n$), il existe un inverse à droite, $AA^{-1} = I$.
- Si la matrice A est carrée, les inverses coïncident et $A^{-1}A = AA^{-1} = I$.

Si la matrice A est carrée, on pourra calculer l'inverse grâce à la méthode d'élimination de Gauss

Jordan, qui transforme le système $A : I$ en $I : A^{-1}$ par des transformations élémentaires :

$$\begin{aligned} [AI] &= \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} & 1 & 0 & \cdots & 0 \\ a_{21} & \cdots & \cdots & a_{2n} & 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{m1} & \cdots & \cdots & a_{mn} & 0 & 0 & \cdots & 1 \end{pmatrix} \\ &\downarrow \\ &\text{transformations élémentaires} \\ &\downarrow \\ [IA^{-1}] &= \begin{pmatrix} 1 & 0 & \cdots & 0 & b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & 1 & \cdots & 0 & b_{21} & \cdots & \cdots & b_{2n} \\ \vdots & \ddots & \vdots & \vdots & \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 & b_{m1} & \cdots & \cdots & b_{mn} \end{pmatrix}. \end{aligned}$$

Nous verrons comment calculer les inverses à droite et à gauche dans la section sur la résolution des systèmes d'équations, la section 6.8.

Les inverses obéissent aux lois suivantes :

- $(A^{-1})^{-1} = A$,
- $(kA)^{-1} = k^{-1}A^{-1}$,
- $(A^T)^{-1} = (A^{-1})^T$,
- $(AB)^{-1} = B^{-1}A^{-1}$.

6.5 Normes

Une norme, intuitivement, correspond à une « longueur » ou au moins un ordre de grandeur. Une norme, typiquement notée $\|x\|$, parfois $|x|$, doit avoir les trois propriétés suivantes :

- $\|x\| \geq 0$,
- $\|kx\| = \|k\| \|x\|$ pour une constante scalaire k ,
- $\|x + y\| \leq \|x\| + \|y\|$, l'inégalité triangulaire.

On a naturellement plusieurs fonctions qui respectent ces trois conditions. Pour un vecteur x en k dimension, on trouve les normes :

- L_1 , notée $\|x\|_1$, calculée par $\|x_1\| = \sum_{i=1}^k |x_i|$, où $|x_j|$ est la valeur absolue de x_j . La norme L_1 est parfois appelée distance de Manhattan, en référence au schéma des rues bien perpendiculaires dans cette grande ville. En anglais, on trouve *taxicab distance* et *taxicab geometry* [90].
- L_2 , ou distance euclidienne, t. q. $\|x\|_2 = \sqrt{x^T x} = \sqrt{\sum_{i=1}^k x_i^2}$.
- L_p , la norme p ou encore la p -norme, parfois aussi notée L^p , est donnée par $\|x\|_p = \sqrt[p]{\sum_{i=1}^k x_i^p}$.
- L_∞ , la « norme infinie » ou norme-max (ou *max-norm*, *uniform norm*, en anglais), donnée par $\|x\|_\infty = \lim_{p \rightarrow \infty} \|x\|_p = \lim_{p \rightarrow \infty} \sqrt[p]{\sum_{i=1}^k x_i^p} = \max_i x_i$.

*
* * *

Pour des matrices, les normes doivent respecter les propriétés suivantes :

- $\|A\| \geq 0$,
- $\|kA\| = \|k\| \|A\|$ pour une constante scalaire k ,
- $\|A + B\| \leq \|A\| + \|B\|$, l'inégalité triangulaire,
- $\|AB\| \leq \|A\| \|B\|$.

De façon analogue aux vecteurs, nous aurons les normes usuelles suivantes :

- La norme absolue, $\|A\|_1 = \max_j \sum_{i=1}^n |a_{ij}|$.
- La norme spectrale, $\|A\|_2$, qui est la racine carrée de la plus grande valeur propre de $A^H A$.
- La norme infinie, $\|A\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|$.

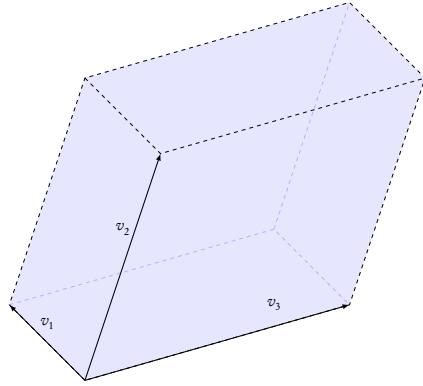


Figure 6.6.1 — Déterminant et volume du parallélépipède généré.

- La norme Hilbert-Schmidt : $\|A\|_{HS} = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}$.
- La norme de Frobenius, $\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$.

6.6 Déterminants

Noté $|A|$, ou $\det(A)$, le déterminant de la matrice carrée A correspond, à un signe près au volume du parallélépipède généré par les vecteurs (rangées ou colonnes) de la matrice A (et peu importe le nombre de dimensions). La fig. 6.6.1 montre le volume généré par trois vecteurs contenus dans une matrice A .

Lorsque le déterminant est nul, cela implique qu'au moins une dimension est nulle, et que le « volume » devient un (hyper)plan de volume zéro.

Le déterminant obéit aux lois suivantes :

- $\det(I) = |I| = 1$.
- $|A^T| = |A|$.
- $|A^{-1}| = |A|^{-1} = \frac{1}{|A|}$.
- $|AB| = |A||B|$.
- $|cA| = c^n |A|$ pour une constante c et une matrice $n \times n$.
- $|D| = \prod_{i=1}^n d_{ii}$ pour une matrice diagonale D .

6.6.1 Calcul du déterminant

Le calcul du déterminant est en général assez pénible. Cependant, pour de petites matrices, nous avons des cas spéciaux. Le cas spécial

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc, \quad (6.6.1)$$

peut être réutilisé pour calculer le déterminant 3×3 :

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \begin{vmatrix} d & e \\ g & h \end{vmatrix}, \quad (6.6.2)$$

ou la méthode de Sarrus, qui ne fonctionne qu'en 3×3 où les produits positifs sont indiqués en vert, les négatifs, en rouge [141] :

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \begin{matrix} \nearrow & \times & \nearrow \\ \nearrow & \times & \nearrow \\ \nearrow & \times & \nearrow \end{matrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix} \begin{matrix} \nearrow & \nearrow \\ \nearrow & \nearrow \\ \nearrow & \nearrow \end{matrix} e = aei + bf g + cdb - gec - hfa - idb. \quad (6.6.3)$$

On peut arriver au résultat de l'éq. (6.6.3) en simplifiant joyeusement et vigoureusement l'éq. (6.6.2). Il s'agit donc plutôt d'un truc pour se souvenir du calcul. Pour 4×4 , on aura

$$\begin{vmatrix} a & b & c & d \\ e & f & g & h \\ i & j & k & l \\ m & n & p & q \end{vmatrix} = a \begin{vmatrix} f & g & h \\ j & k & l \\ n & p & q \end{vmatrix} - b \begin{vmatrix} e & g & h \\ i & k & l \\ m & p & q \end{vmatrix} + c \begin{vmatrix} e & f & h \\ i & j & l \\ m & n & q \end{vmatrix} - d \begin{vmatrix} e & f & g \\ i & j & k \\ m & n & p \end{vmatrix}.$$

De plus, il existe plusieurs algorithmes pour calculer le déterminant de matrices spéciales (diagonales, triangulaires supérieures, etc.).

6.7 Projections

6.7.1 Projection vecteur contre vecteur

Pour calculer la projection orthogonale d'un vecteur u sur un vecteur v , il faut trouver un vecteur dans la même direction que v , mais de longueur proportionnelle au cosinus de l'angle formé par les deux vecteurs (voir fig. 6.7.1).

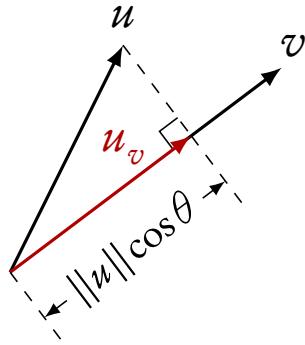


Figure 6.7.1 — Projection d'un vecteur contre un autre vecteur.

Sachant que

$$u^T v = \|u\| \|v\| \cos \theta,$$

on peut trouver que u_v (u -projeté-sur- v) est donné par

$$\begin{aligned} u_v &= \frac{v}{\|v\|} \|u\| \cos \theta \\ &= \frac{v}{\|v\|} \|u\| \frac{u^T v}{\|u\| \|v\|} \\ &= \frac{v}{\|v\|} \frac{u^T v}{\|v\|} \\ &= u^T v \frac{v}{\|v\|^2} \\ &= u^T v \frac{v}{v^T v}. \end{aligned}$$

Où $\frac{v}{\|v\|}$ est un vecteur de longueur 1 qui pointe dans la même direction que v . On remarquera par ailleurs qu'on a pris soin de ne pas mélanger scalaires et vecteurs, et ainsi le résultat ne se simplifie pas plus.

6.7.2 Projection vecteur contre un plan

On veut projeter un vecteur v contre plan. Le vecteur v « habite » un espace vectoriel V et le plan un sous-espace $S \subseteq V$ (plus souvent qu'autrement, nous aurons strictement inclus, $S \subset V$, si on veut projeter contre un (hyper)plan).

La projection d'un vecteur v contre un plan S est donnée par

$$v_S = S(S^T S)^{-1} S^T v. \quad (6.7.1)$$

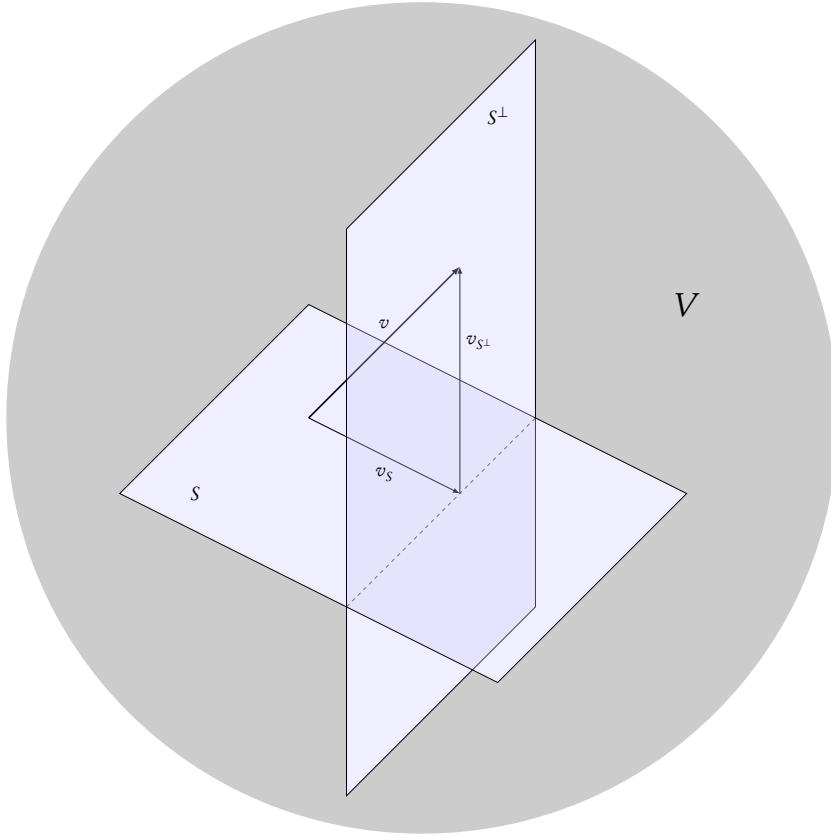


Figure 6.7.2 — Projection d'un vecteur v contre un (hyper)plan S .

Démonstration 6.7.1.

Démonstration de l'éq. (6.7.1). Supposons que nous voulions projeter un vecteur $v \in V$ sur un plan $S \subseteq V$ (plus souvent, sans perte de généralité, $S \subset V$).

Notons S^\perp , le complément perpendiculaire de S dans V . Pour un vecteur $v \in V$, une partie du vecteur est dans S et une partie dans S^\perp , soit donc

$$v = v_S + v_{S^\perp},$$

où $v_S \in S$ et $v_{S^\perp} \in S^\perp$. Donc, en réarrangeant,

$$\begin{aligned} v_{S^\perp} &= v - v_S \\ &= v - Sy, \end{aligned}$$

pour S , la matrice qui décrit le sous-espace et y un vecteur qu'on voudra trouver (c'est-à-dire que Sy est v_S). On veut trouver Sy perpendiculaire à

S^\perp , donc on veut résoudre pour y :

$$\begin{aligned} S^T(v - Sy) &= 0 \\ S^T v - S^T S y &= 0 \\ S^T S y &= S^T v \\ y &= (S^T S)^{-1} S^T v. \end{aligned}$$

Enfin, puisque $v_S = Sy$, nous avons $v_S = S(S^T S)^{-1} S^T v$. □

6.8 Systèmes d'équations

Une des applications intéressantes de l'algèbre linéaire est la résolution de systèmes d'équations linéaires par des méthodes systématiques. Commençons par établir le vocabulaire.

Définition 6.8.1.

On appelle *système de m équations en n inconnues* tout ensemble S de la forme

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

où les a_{ij} sont les coefficients, les x_j les variables et les b_i des constantes scalaires (réelles ou complexes). Si les b_i sont tous nuls, nous dirons que le système est *homogène*. Si les b_i ne sont pas tous nuls, le système sera *non homogène*. □

Définition 6.8.2.

Soit S , un système d'équations linéaires en m équations et n inconnues. Nous dirons que $x_1^*, x_2^*, \dots, x_n^*$ est une solution à S ssi toutes les équations du système sont vérifiées en posant $x_1 = x_1^*$, $x_2 = x_2^*$, ..., $x_n = x_n^*$. □

Définition 6.8.3.

L'ensemble de tous les n -tuplets qui sont solution au système S forment l'*ensemble-solution* de S . □

Définition 6.8.4.

Deux systèmes d'équations de m équations en n inconnues sont *équivalents* ssi ils ont le même ensemble solution. \square

*

* *

Si on a un système d'inéquations, on peut le transformer en système d'équations en ajoutant des variables d'écart (en anglais, *slack variable*, en allemand, *schlupfvariable*!). Par exemple, le système d'inéquations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &\leq b_1 \\ a_{21}x_1 + a_{22}x_2 &\geq b_2 \end{aligned}$$

devient un système d'équations par l'introduction de variables d'écart,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + s_1 &= b_1 \\ a_{21}x_1 + a_{22}x_2 + s_2 &= b_2 \end{aligned}$$

*

* *

On peut exprimer les systèmes d'équations linéaires avec le formalisme des matrices et des vecteurs, en remarquant que le système

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

peut être réécrit comme

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix},$$

c'est-à-dire l'équation matrice/vecteur $Ax = b$, qu'on peut résoudre en isolant x .

6.8.1 Rang, déterminant, et solvabilité

Pour une résolution par élimination « classique » (c'est-à-dire résoudre $Ax = b$ par $x = A^{-1}b$), il faut que la matrice soit carrée et de déterminant non nul (car si le déterminant est zéro, le volume généré par la matrice est nul, et ne peut pas « remplir » l'espace, et détermine tout au plus un (hyper)plan ; voir fig. 6.6.1). Si la matrice A n'est pas carrée, ou si elle est singulière d'une façon ou d'une autre, nous pourrons toujours résoudre, au moins approximativement, le système d'équations.

6.8.2 Résoudre les systèmes d'équations

Pour résoudre un système d'équations par élimination, nous pouvons calculer l'inverse de la matrice par l'élimination de Gauss-Jordan, ou encore transformer le système en une forme « échelon réduite » grâce aux *transformations élémentaires* qui préservent l'équivalence des systèmes d'équations. Ces transformations élémentaires sont

- une permutation de deux équations du système,
- la multiplication d'une équation du système par une constante non nulle,
- l'addition d'un multiple d'une équation du système à une autre.

Et on les utilisera pour trouver la solution.

Théorème 6.8.1.

On peut, par une série appropriée de transformations élémentaires, amener un système d'équation en m équations en n inconnues à la forme échelon, c'est-à-dire la forme

$$\begin{aligned} x_1 + c_{12}x_2 + c_{13}x_3 + \cdots + c_{1n}x_n &= d_1 \\ x_2 + c_{23}x_3 + \cdots + c_{2n}x_n &= d_2 \\ &\vdots && \vdots \\ x_{r-1} + c_{r-1,n}x_r &= d_{r-1} \\ x_r &= d_r \\ 0 &= 0 \\ &\vdots && \vdots \\ 0 &= 0 \end{aligned}$$

où $r \leq m$ détermine le *rang* du système. Par une autre série de transformations élémentaires, on peut ramener la forme échelon à

la forme échelon réduite, de la forme

$$\begin{aligned} x_1 &= e_1 \\ x_2 &= e_2 \\ x_3 &= e_3 \\ \vdots &= \vdots \\ x_r &= e_r \end{aligned}$$

ce qui nous donne la solution au système d'équations.

□

Évidemment, trouver les transformations élémentaires minimales n'est pas complètement trivial. Habituellement, nous procéderons par étapes. En choisissant un *pivot*, c'est-à-dire une rangée et une colonne en particulier, nous allons faire en sorte que les coefficients dans une colonne soient zéro partout, sauf pour une rangée. Nous choisissons donc une rangée de la matrice où le coefficient dans la j^{e} colonne est numériquement pratique (près de un; pas trop grand, en évitant les valeurs extrêmement petites ou extrêmement grandes pour des raisons de stabilité numérique) et nous additionnons ses multiples sur toutes les autres rangées de façon à mettre les coefficients de la j^{e} colonne à zéro partout, et à 1 sur la rangée choisie. Cette stratégie est équivalente à calculer l'inverse de la matrice.

Cependant, la résolution de systèmes d'équations n'est pas toujours aussi facile. Le système peut être homogène (de la forme $Ax = 0$), sous-déterminé (avec moins d'équations que de variables) ou sur-déterminé (avec plus d'équations que de variables). Voyons comment les résoudre.

6.8.2.1 Systèmes homogènes

Dans le cas particulier où le système est homogène (et que le rang du système correspond au nombre d'équations, soit $r = m = n$), la seule solution au système

$$Ax = 0$$

est $x = 0$!

Souvent, même si la solution $x = 0$ est exacte, nous la trouverons assez peu intéressante. Plutôt que de résoudre directement le système $Ax = 0$, nous allons résoudre un système qui impose des conditions sur x , par exemple, grâce à une régularisation de Tikhonov [156, 157], c'est-à-dire qu'on tentera de minimiser

$$\|Ax - b\|^2 + \lambda\|Bx\|^2,$$

où la matrice B impose des conditions sur x , et λ est un multiplicateur de Lagrange, un paramètre à ajuster qui balance les deux parties de l'équation. On pourra aussi imposer n'importe quelle condition, par exemple

$$\|Ax\|^2 + \lambda(1 - x^T x)^2,$$

pour chercher un x de norme 1. La résolution de ce type de systèmes est présentée à la prochaine section.

6.8.2.2 Systèmes sous-déterminés et pseudo-inverse à droite

Dans le cas d'un système d'équations sous-déterminé (on dit aussi sous-spécifié) nous aurons moins d'équations que d'inconnues, c'est-à-dire $m < n$, et donc cela nous laisse (au moins) $n - m$ variables libres (car le rang est au plus $r = m$). Au mieux, nous pourrons déterminer m des n variables, laissant les $n - m$ autres prendre potentiellement une infinité de valeur. Mais parmi cet infinité de solutions, laquelle choisir ?

Comme nous ne pouvons pas résoudre simplement en calculant l'inverse de la matrice, nous allons rajouter un terme au système sous-déterminé

$$Ax = b \tag{6.8.1}$$

pour obtenir (arbitrairement) une solution près de zéro (zéro en n dimensions, donc le vecteur nul). Cette régularisation prend la forme

$$S = \Lambda^T \|Ax - b\|^2 + \|x\|^2, \tag{6.8.2}$$

où Λ est une matrice. Or, cette équation étant quadratique, elle détermine une fonction convexe¹, et on pourra résoudre

$$\frac{\partial S}{\partial x} = 0,$$

pour x . Faisons donc (avec les règles présentées à la § 6.9)! On trouve

$$\begin{aligned} \frac{\partial S}{\partial x} &= \Lambda^T A + 2x^T = 0 \\ x^T &= -\frac{1}{2}\Lambda^T A \\ x &= -\frac{1}{2}A^T \Lambda. \end{aligned} \tag{6.8.3}$$

1. Ici « convexe » inclut aussi le cas concave ; l'important c'est que la fonction ait un seul minimum/maximum et que les dérivées soient nulles à ce point.

Cependant, A est $m \times n$ et de rang (au plus) $r = m$. Nous remarquons (c'est important) que AA^T comme A^TA sont des matrices carrées de rang $r = m$ (si A est de rang $r = m$) et donc réversibles. Nous exploiterons cette observation.

En substituant l'éq. (6.8.3) dans l'éq. (6.8.1), nous trouvons

$$\begin{aligned} Ax &= b \\ A\left(-\frac{1}{2}A^T\Lambda\right) &= b \\ -\frac{1}{2}AA^T\Lambda &= b \\ -\Lambda &= 2(AA^T)^{-1}b \\ \Lambda &= -2(AA^T)^{-1}b. \end{aligned} \tag{6.8.4}$$

Nous avons maintenant trouvé Λ . Forts de ce résultats, replaçons le résultat de l'éq. (6.8.4) dans l'éq. (6.8.3) :

$$\begin{aligned} x &= -\frac{1}{2}A^T\Lambda \\ x &= -\frac{1}{2}A^T(-2(AA^T)^{-1}b) \\ x &= A^T(AA^T)^{-1}b \end{aligned} \tag{6.8.5}$$

ce qui nous donne notre solution pour x . Cependant, le membre de droite de l'éq. (6.8.5) est bien plus intéressant qu'il n'y paraît. En effet, si nous repartons de l'éq. (6.8.1),

$$\begin{aligned} Ax &= b \\ A(A^T(AA^T)^{-1}b) &= b \\ \underbrace{AA^T(AA^T)^{-1}}_{\approx I}b &= b \end{aligned} \tag{6.8.6}$$

Et donc, $A^T(AA^T)^{-1}$ est un inverse de A , puisque $AA^T(AA^T)^{-1} \approx I$. Plus spécifiquement $A^T(AA^T)^{-1}$ est le *pseudo-inverse à droite* de la matrice sous-déterminée A .

Remarquons que le pseudo-inverse n'est pas forcément l'inverse exact. C'est le « moins pire inverse » aux moindres carrés, puisqu'il minimise l'éq. (6.8.2). Lorsque A est carrée et de rang maximal, le pseudo-inverse à droite et l'inverse « ordinaire » coïncident.

On trouve parfois l'équation $x = A^T(AA^T)^{-1}b$ écrite comme $AA^Tx = A^Tb$, et on parlera de l'« équation normale » (surtout en statistiques). Pourquoi « normale » ? Parce que géométriquement, la solution est le point zéro (en n dimensions) projeté orthogonalement contre le (hyper)plan qui

contient les solutions, ce qui donne une normale au plan. Ou à l'inverse, on peut voir la solution comme le point contenu dans le (hyper)plan le plus près de l'origine, qui est en quelque sorte la plus petite solution possible. Revisitez la § 6.7.2 sur les projections de vecteurs contre un plan. Voyez-vous le parallèle ?

6.8.2.3 Systèmes sur-déterminés et pseudo-inverse à gauche

Dans le cas d'un système d'équations sur-déterminé (on dit aussi sur-spécifié) nous aurons plus d'équations que de variables. Cela nous laisse plusieurs possibilités :

- Les équations sont contradictoires (le système contient deux équations du genre $x_1 = 3$ et $x_1 = 5$) et le système n'a aucune solution.
- Les équations sont cohérentes et permettent de ramener le système à une forme échelon de rang n où seules les n premières rangées sont non nulles (et donc les $m - n$ dernières lignes du système sont de la forme $0 = 0$). Nous avons une seule solution.
- Les équations sont cohérentes mais ramènent à une forme échelon de rang moins que n ; nous avons une infinité de solution.

Si nous avons des équations incohérentes (ou contradictoires) le système n'admet pas de solution exacte, mais pourrait bien admettre une solution approximative qui ne viole pas trop les équations. Comme pour les systèmes sous-spécifiés, nous chercherons une solution qui minimise l'erreur au carré moyenne, c'est-à-dire qui minimise $\|Ax - b\|^2$.

*
* *

Cependant, avant de nous lancer dans la dérivation de la solution en différenciant l'erreur par rapport à x , nous pouvons remarquer que A , la matrice $m \times n$ avec $m > n$ (et parfois $m \gg n$), a n colonnes (puisque x est un vecteur-colonne de dimension n), et que $A^T A$ est $n \times n$, ce qui nous permet une dérivation algébrique qui respecte toutes les dimensions des matrices et vecteurs (i.e., dont la forme reste toujours valide) :

$$\begin{aligned}
 Ax &= b \\
 A^T Ax &= A^T b \\
 \underbrace{(A^T A)^{-1}(A^T A)}_{\approx I} x &= (A^T A)^{-1} A^T b \\
 \tilde{x} &= (A^T A)^{-1} A^T b.
 \end{aligned} \tag{6.8.7}$$

Cela nous donne $(A^T A)^{-1} A^T$ comme pseudo-inverse à gauche de A , puisque $(A^T A)^{-1}(A^T A) \approx I$. On pourra vérifier à loisir que la géométrie des matrices est respectée dans les manipulations de l'éq. (6.8.7).

Ne s'agit-il là que d'un « truc d'algébriste » ? Non, car nous pouvons dériver le même résultat avec la dérivée de la fonction-objectif $\|Ax - b\|^2$. Voyons comment. Posons le problème comme un problème de minimisation, cette fois

$$S = \|Ax - b\|^2.$$

Puisque le problème est convexe, on résout

$$\frac{\partial}{\partial x} \|Ax - b\|^2 = 0$$

pour x . On trouve

$$\frac{\partial}{\partial x} \|Ax - b\|^2 = 2A^T(Ax - b) \quad (6.8.8)$$

puis

$$\begin{aligned} A^T(Ax - b) &= 0 \\ A^T Ax - A^T b &= 0 \\ A^T Ax &= A^T b \\ (A^T A)^{-1}(A^T A)x &= (A^T A)^{-1}A^T b \\ \tilde{x} &= (A^T A)^{-1}A^T b, \end{aligned}$$

ce qui confirme le résultat précédent obtenu par manipulations algébriques.

Ainsi donc, $(A^T A)^{-1} A^T$ est le pseudo-inverse à gauche de A .

*
* * *

Remarquons enfin que lorsque A est carré et de rang n , les trois inverses coïncident (si A n'est pas carrée, les inverses sont différents). Certes :

$$\begin{aligned} (A^T A)^{-1} A^T &\stackrel{?}{=} A^T (A A^T)^{-1} \\ \underbrace{A^{-1} (\underbrace{A^T}_{I})^{-1} A^T}_{I} &= \underbrace{A^T (A^T)^{-1} A^{-1}}_{I} \\ A^{-1} &= A^{-1}. \end{aligned}$$

Cette dernière démonstration ne fonctionne que si A est carrée et A^{-1} existe.

6.8.2.4 Régressions

Les systèmes sur-déterminés correspondent de façon naturelle à un autre problème que vous avez sûrement déjà rencontré, la régression. Dans un problème de régression, nous avons une fonction $f(x; \theta)$ (possiblement horriblement compliquée) dont on souhaite ajuster les paramètres θ de façon à ce que, pour un nuage de points $\{(x_i, y_i)\}_{i=1}^m$, nous ayons partout $f(x_i; \theta) \approx y_i$, tout en minimisant globalement l'erreur. Lorsque la fonction $f(x; \theta)$ est linéaire en θ (les θ sont les variables, et les (x_i, y_i) sont constants), on peut utiliser le pseudo-inverse à gauche pour trouver les paramètres, donnant ainsi un ajustement optimal aux moindres carrés.

Considérons un cas particulier, une fonction $f(x; \alpha)$, un polynôme de degré n , c'est-à-dire de la forme

$$\alpha_n x^n + \alpha_{n-1} x^{n-1} + \cdots + \alpha_2 x^2 + \alpha_1 x + \alpha_0 = y.$$

En utilisant tous les points, $\{(x_i, y_i)\}_{i=1}^m$, (avec $m \geq n+1$, puisqu'un polynôme de degré d est déterminé de façon unique par $d+1$ points — c'est le théorème d'unisolvance), nous obtenons un système d'équations de la forme

$$\begin{aligned} \alpha_n x_1^n + \alpha_{n-1} x_1^{n-1} + \cdots + \alpha_2 x_1^2 + \alpha_1 x_1 + \alpha_0 &= y_1 \\ \alpha_n x_2^n + \alpha_{n-1} x_2^{n-1} + \cdots + \alpha_2 x_2^2 + \alpha_1 x_2 + \alpha_0 &= y_2 \\ &\vdots \\ \alpha_n x_m^n + \alpha_{n-1} x_m^{n-1} + \cdots + \alpha_2 x_m^2 + \alpha_1 x_m + \alpha_0 &= y_m \end{aligned}$$

que l'on peut réécrire comme

$$\begin{pmatrix} x_1^n & x_1^{n-1} & x_1^{n-2} & \cdots & x_1 & 1 \\ x_2^n & x_2^{n-1} & x_2^{n-2} & \cdots & x_2 & 1 \\ \vdots & & & & \vdots & \\ x_m^n & x_m^{n-1} & x_m^{n-2} & \cdots & x_m & 1 \end{pmatrix} \begin{pmatrix} \alpha_n \\ \alpha_{n-1} \\ \alpha_{n-2} \\ \vdots \\ \alpha_1 \\ \alpha_0 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

C'est donc un système de la forme $X\alpha = y$ (donc même modèle que $Ax = b$, aux noms des variables près). Notons au passage que la matrice X a une forme particulière, avec des 1 dans la dernière colonne. C'est une matrice de Vandermonde [163], pour laquelle il existe des algorithmes plus efficaces que l'élimination de Gauss-Jordan lorsqu'elle est carrée [58, 161], mais dans le cas de la régression, la matrice n'est pas carrée, mais « haute » avec $m \gg n+1$. Nous utilisons donc le pseudo-inverse à gauche,

$$\begin{aligned} X\alpha &= y \\ \tilde{\alpha} &= (X^T X)^{-1} X^T y, \end{aligned}$$

ce qui nous donne directement les coefficients optimaux $\tilde{a}_n, \tilde{a}_{n-1}, \dots, \tilde{a}_0$ pour notre polynôme de degré n passant (approximativement) par m points.

6.9 Dérivées

À la section précédente, la section 6.8.2, nous avons vu que pour résoudre certains systèmes d'équations, nous avons eu recours aux dérivées d'équations linéaires.

6.9.1 Vecteurs

Soient donc a , un scalaire, et u, v des vecteurs. Alors

- $\frac{\partial}{\partial a} au = u$, car

$$\begin{aligned}\frac{\partial}{\partial a} au &= \frac{\partial}{\partial a}(au_1, au_2, \dots, au_k) \\ &= \left(\frac{\partial}{\partial a} au_1, \frac{\partial}{\partial a} au_2, \dots, \frac{\partial}{\partial a} au_k\right) \\ &= (u_1, u_2, \dots, u_k) \\ &= u.\end{aligned}$$

- $\frac{\partial}{\partial u} au = \vec{a}$, car

$$\begin{aligned}\frac{\partial}{\partial u} au &= \frac{\partial}{\partial u}(au_1, au_2, \dots, au_k) \\ &= \left(\frac{\partial}{\partial u_1} au_1, \frac{\partial}{\partial u_2} au_2, \dots, \frac{\partial}{\partial u_k} au_k\right) \\ &= (a, a, \dots, a) \\ &= \vec{a}.\end{aligned}$$

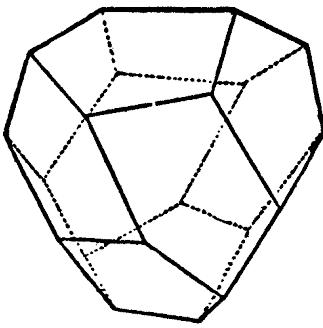
- $\frac{\partial}{\partial u_i} au = (0, \dots, a, \dots, 0)$, car

$$\begin{aligned}\frac{\partial}{\partial u_i} au &= \frac{\partial}{\partial u_i}(au_1, au_2, \dots, au_k) \\ &= \left(\frac{\partial}{\partial u_i} au_1, \frac{\partial}{\partial u_i} au_2, \dots, \frac{\partial}{\partial u_i} au_k\right) \\ &= (0, \dots, a, \dots, 0),\end{aligned}$$

avec le a en i^e position.

- $\frac{\partial}{\partial u} u^T v = v$, car $u^T v = \sum_{i=1}^k u_i v_i$, et donc
$$\begin{aligned}\frac{\partial}{\partial u} u^T v &= \left(\frac{\partial}{\partial u_1} u^T v, \frac{\partial}{\partial u_2} u^T v, \dots, \frac{\partial}{\partial u_k} u^T v \right) \\ &= (v_1, v_2, \dots, v_k) \\ &= v.\end{aligned}$$

De même façon, $\frac{\partial}{\partial v} u^T v = u$.
- $\frac{\partial}{\partial u_i} u^T v = v_i$ puisque $u^T v = \sum_{i=1}^k u_i v_i$, et $\frac{\partial}{\partial u_i} \sum_{j=1}^k u_j v_j = v_i$.



6.9.2 Matrices

Soient A , une matrice $m \times n$, B une matrice $n \times p$, x un vecteur-colonne de dimension n , et c une constante.

Nous avons

- $\frac{\partial}{\partial x_i} Ax = a_{*i}$. Nous avons bien que

$$\frac{\partial}{\partial x_i} Ax = \frac{\partial}{\partial x_i} \begin{pmatrix} a_{1*}x \\ a_{2*}x \\ \vdots \\ a_{m*}x \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_i} a_{1*}x \\ \frac{\partial}{\partial x_i} a_{2*}x \\ \vdots \\ \frac{\partial}{\partial x_i} a_{m*}x \end{pmatrix} = \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{mi} \end{pmatrix} = a_{*i},$$

où $a_{*i}x$ est le produit scalaire entre la i^{e} rangée de la matrice A et le vecteur x .

- $\frac{\partial}{\partial x} Ax = A$.

- $\frac{\partial}{\partial a_{ij}} Ax = (0, \dots, x_j, \dots)$ avec x_j en i^e position. Vérifions :

$$\frac{\partial}{\partial a_{ij}} Ax = \frac{\partial}{\partial a_{ij}} \begin{pmatrix} a_{1*}x \\ a_{2*}x \\ \vdots \\ a_{m*}x \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial a_{ij}} a_{1*}x \\ \frac{\partial}{\partial a_{ij}} a_{2*}x \\ \vdots \\ \frac{\partial}{\partial a_{ij}} a_{m*}x \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ x_j \\ \vdots \\ 0 \end{pmatrix},$$

car a_{ij} n'intervient que dans le produit $a_{i*}x$, dans la i^e rangée, et a_{ij} multiplie x_j .

6.9.3 Normes

Soient x , un vecteur, et A une matrice, tous deux compatibles. Alors

- $\frac{\partial}{\partial x} \|x\|^2 = 2x$. Certes :

$$\begin{aligned} \frac{\partial}{\partial x} \|x\|^2 &= \frac{\partial}{\partial x} x^T x = \frac{\partial}{\partial x} \sum_{i=1}^k x_i^2 \\ &= (\frac{\partial}{\partial x_1} \sum_{i=1}^k x_i^2, \frac{\partial}{\partial x_2} \sum_{i=1}^k x_i^2, \dots, \frac{\partial}{\partial x_k} \sum_{i=1}^k x_i^2) \\ &= (2x_1, 2x_2, \dots, 2x_k) \\ &= 2(x_1, x_2, \dots, x_k) \\ &= 2x. \end{aligned}$$

- $\frac{\partial}{\partial x} \|Ax\|^2 = 2A^T Ax$. Voyons :

$$\begin{aligned} \frac{\partial}{\partial x} \|Ax\|^2 &= \frac{\partial}{\partial x} (Ax)^T (Ax) = \left(\frac{\partial}{\partial x} (Ax)^T \right) Ax + (Ax)^T \left(\frac{\partial}{\partial x} (Ax) \right) \\ &= A^T Ax + (Ax)^T A \\ &= 2A^T Ax. \end{aligned}$$

Ce qui explique l'éq. (6.8.8).

6.10 Remarques bibliographiques

Le produit croisé a été introduit indépendamment par Josiah Willard Gibbs (1839–1903) et Oliver Heaviside (1850–1925) en 1881. C'est ce pendant William Kingdon Clifford (1845–1879) qui a

introduit les vocables « produit scalaire » et « produit vectoriel » pour distinguer $u^T v$ et $u \times v$ [24,25]. En anglais, c'est *cross product* qui est le plus souvent rencontré, bien que *vector product* soit aussi fréquent.

*
* * *

Les déterminants ne sont pas une invention récente. Déjà, Gabriel Cramer (Français, 1704–1752) les utilise en 1750 [23], peut-être devancé par Maclaurin (oui, le Maclaurin des séries) dont certains lui accordent la préséance en 1748, peut-être même 1729 [68,96]. D'autres, au contraire, confirment Cramer comme premier inventeur [89].

7

Constantes

Sommaire. Dans l'étude des algorithmes, il n'est pas rare de rencontrer e , π , $\sqrt{2}$, la constante d'Euler-Mascheroni γ , $\ln 2$, $\sqrt{2\pi}$, $\frac{1}{\sqrt{2}}$, etc... Nous donnons ici la valeur de ces constantes, des approximations rationnelles, et quelques séries pour les calculer.

7.1 Constantes

7.1.1 $\gamma = 0.57721566490153286061\dots$

La constante d'Euler-Mascheroni, γ , apparaît souvent en compagnie de $n!$, des sommes de logarithmes, et dans l'expression des nombres harmoniques. La voici avec 200 chiffres après le point :

$$\begin{aligned}\gamma = & 0.57721566490153286065120900824024310421 \\& 5933593992359880576723488486772677766467 \\& 0936947063291746749514631447249807082480 \\& 9605040144865428362241739976449235362535 \\& 0033374293733773767394279259525824709491\dots\end{aligned}$$

De bonnes approximations sont données par

$$\gamma \approx \frac{228}{395} \quad \text{et} \quad \gamma \approx \frac{9\,062\,634}{15\,700\,603},$$

lesquelles donnent des erreurs $\varepsilon = O(10^{-7})$ et $\varepsilon = O(10^{-17})$, respectivement, bien suffisantes pour float et double. La constante γ est définie par

$$\gamma = \lim_{n \rightarrow \infty} \left(\sum_{i=1}^{\infty} \frac{1}{i} \right) - \ln n .$$

Sa fraction continue¹ est donnée par

$$\gamma = 0 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{3 + \cfrac{1}{13 + \dots}}}}}}}}}$$

Les autres coefficients sont donnés, selon la notation habituelle pour les fractions continues, par $[0; 1, 1, 2, 1, 2, 1, 4, 3, 13, 5, 1, 1, 8, 1, 2, 4, 1, 1, 40, \dots]$ (on sépare l'unité des autres coefficients par un ;).

7.1.2 $\ln 2 = 0.69314718055994530942\dots$

La constante $\ln 2$ n'est jamais bien loin lorsqu'on a des logarithmes à base 2. À 200 chiffres de précision, on a

$$\begin{aligned} \ln 2 &= 0.6931471805599453094172321214581765680755 \\ &\quad 0013436025525412068000949339362196969471 \\ &\quad 5605863326996418687542001481020570685733 \\ &\quad 6855202357581305570326707516350759619307 \\ &\quad 2757082837143519030703862389167347112335\dots \end{aligned}$$

De bonnes approximations sont

$$\ln 2 \approx \frac{445}{642} \quad \varepsilon = O(10^{-7}) \quad \text{et} \quad \ln 2 \approx \frac{5\,278\,668}{7\,615\,537} \quad \varepsilon = O(10^{-15}).$$

Nous pouvons calculer

$$\ln 2 = 1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \frac{1}{5} - \dots$$

C'est un cas spécial de la formule

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots$$

1. Les fractions continues sont d'autant plus intéressantes qu'elles donnent la meilleure approximation de toutes les fractions ayant un dénominateur plus petit ou égal. Par contre, pour un même nombre de *chiffres*, elles ne sont pas nécessairement les plus précises.

où il suffit de poser $x = 1$.

Sa fraction continue est

$$\ln 2 = 0 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{6 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \dots}}}}}}}}}}$$

soit $[0; 1, 2, 3, 1, 6, 3, 1, 1, 2, 1, 1, 1, 3, 10, 1, 1, 1, 2, \dots]$.

$$7.1.3 \quad \frac{1}{\sqrt{2}} = 70710678118654752440\dots$$

À 200 chiffres de précision, on a

$$\begin{aligned} \frac{1}{\sqrt{2}} &= 0.7071067811865475244008443621048490392848 \\ &\quad 3593768847403658833986899536623923105351 \\ &\quad 9425193767163820786367506923115456148512 \\ &\quad 4624180279253686063220607485499679157066 \\ &\quad 1133296375279637789997525057639103028573\dots \end{aligned}$$

Nous pouvons approximer par

$$\frac{1}{\sqrt{2}} \approx \frac{408}{577} \quad \varepsilon = O(10^{-6}) \quad \text{et} \quad \frac{1}{\sqrt{2}} \approx \frac{38\,613\,965}{54\,608\,393} \quad \varepsilon = O(10^{-16}).$$

Nous pouvons le calculer grâce aux algorithmes pour $\sqrt{2}$ (voir la section suivante). Sa fraction continue est

$$\frac{1}{\sqrt{2}} = 0 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \dots}}}}}$$

soit $[0; 1, 2, 2, 2, \dots]$.

7.1.4 $\sqrt{2} = 1.4142135623730950488\dots$

La racine carrée de 2, notée $\sqrt{2}$, nous donne les 200 premières décimales

$$\begin{aligned}\sqrt{2} &= 1.4142135623730950488016887242096980785696 \\ &\quad 7187537694807317667973799073247846210703 \\ &\quad 8850387534327641572735013846230912297024 \\ &\quad 9248360558507372126441214970999358314132 \\ &\quad 2266592750559275579995050115278206057147\dots\end{aligned}$$

Les approximations

$$\sqrt{2} \approx \frac{577}{408} \quad \varepsilon = O(10^{-6}) \quad \text{et} \quad \sqrt{2} \approx \frac{54\,608\,393}{38\,613\,965} \quad \varepsilon = O(10^{-16}).$$

sont les réciproques des approximations pour $\frac{1}{\sqrt{2}}$ (qui l'eût cru?). On peut la calculer grâce à la série de produits

$$\sqrt{2} = \prod_{i=0}^{\infty} \left(1 + \frac{1}{4i+1}\right) \left(1 - \frac{1}{4i+3}\right) = \left(1 + \frac{1}{1}\right) \left(1 - \frac{1}{3}\right) \left(1 + \frac{1}{5}\right) \left(1 - \frac{1}{7}\right) \dots$$

Sa fraction continue est

$$\sqrt{2} = 0 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \cfrac{1}{2 + \dots}}}}}$$

c'est-à-dire $[0; 2, 2, 2, \dots]$.

7.1.5 $\phi = 1.6180339887498948482\dots$

Le nombre d'or, noté ϕ est défini comme étant la racine positive du polynôme $x^2 - x - 1$, et vaut exactement

$$\phi = \frac{1 + \sqrt{5}}{2}$$

ce qui nous donne

$$\begin{aligned}\phi &= 1.6180339887498948482045868343656381177203 \\ &\quad 0917980576286213544862270526046281890244 \\ &\quad 9707207204189391137484754088075386891752 \\ &\quad 1266338622235369317931800607667263544333 \\ &\quad 8908659593958290563832266131992829026788\dots\end{aligned}$$

Nous pourrons utiliser les approximations

$$\phi \approx \frac{987}{610} \quad \varepsilon = O(10^{-6}) \quad \text{et} \quad \phi \approx \frac{63\,245\,986}{39\,088\,169} \quad \varepsilon = O(10^{-16}).$$

On pourra le calculer directement à partir de son expression exacte $\frac{1+\sqrt{5}}{2}$, ou grâce à la curieuse série

$$\phi = \sqrt{1 + \dots}}}}}}}$$

ou encore grâce à sa fraction continue, peut-être la plus simple de toutes :

$$\phi = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \dots}}}},$$

soit donc $[1; 1, 1, 1, \dots]$.

7.1.6 $\sqrt{2\pi} = 2.5066282746310005024\dots$

La constante $\sqrt{2\pi}$ apparaît, entre autres, dans la formule de Stirling pour la factorielle,

$$n! = \sqrt{2\pi n} \left(\frac{n}{e} \right)^n \left(1 + \frac{1}{12} + \frac{1}{288n^2} - \frac{139}{51840n^3} - \dots \right)$$

Les 200 premières décimales sont

$$\begin{aligned} \sqrt{2\pi} &= 2.5066282746310005024157652848110452530069 \\ &\quad 8674060993831662992357634229365460784197 \\ &\quad 4946595838378057266116009972665203879644 \\ &\quad 8663236181267361809578556559661479319134 \\ &\quad 3548045812373113732780431301993880264715\dots \end{aligned}$$

et on remarque que $\frac{5}{2}$ est déjà une bonne approximation. On peut quand même utiliser

$$\sqrt{2\pi} \approx \frac{1702}{679} \quad \varepsilon = O(10^{-7}) \quad \text{et} \quad \sqrt{2\pi} \approx \frac{13\,268\,734}{5\,293\,459} \quad \varepsilon = O(10^{-14}).$$

Sa fraction continue est

$$\sqrt{2\pi} = 2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{37}{44 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{9 + \cfrac{1}{1 + \dots}}}}}}}}}}$$

ou encore $[2; 1, 1, 37, 4, 1, 1, 1, 1, 9, 1, 1, 2, 8, 6, 1, 2, 2, 1, 3, \dots]$.

7.1.7 $e = 2.7182818284590452354\dots$

La constante e apparaît souvent de façon inattendue, mais est surtout connue comme étant la base des logarithmes naturels. À 200 décimales, nous avons

$$\begin{aligned} e &= 2.7182818284590452353602874713526624977572 \\ &\quad 4709369995957496696762772407663035354759 \\ &\quad 4571382178525166427427466391932003059921 \\ &\quad 8174135966290435729003342952605956307381 \\ &\quad 3232862794349076323382988075319525101901\dots \end{aligned}$$

Le découpage $2.7/1828/1828/45/90/45\dots$ permet de retenir 15 chiffres.

Les approximations

$$e \approx \frac{2721}{1001} \quad \varepsilon = O(10^{-7}) \quad \text{et} \quad e \approx \frac{28\,245\,729}{10\,391\,649} \quad \varepsilon = O(10^{-16}),$$

sont tout à fait cromulentes. On peut aussi utiliser

$$e \approx 2 + \frac{1}{2} + \frac{1}{5} + \frac{1}{55} + \frac{1}{9999} = \frac{271801}{99990} \quad \varepsilon = O(10^{-10})$$

On peut, de plus, le calculer en observant que

$$e = \lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n,$$

mais comme n « infini » n'est pas très pratique, il suffit d'un n très grand pour s'en approcher. L'approximation panumérique¹ de Sabey,

$$e \approx \left(1 + 9^{-4^{6 \times 7}}\right)^{3^{2^{85}}}$$

1. De *pan*, tous, et *numérique*, pour les chiffres. En anglais on trouve *pandigital* (qui ne signifie *pas* tous les doigts). C'est donc une approximation où tous les chiffres apparaissent exactement une fois.

nous donne 18 trillions de décimales ($\approx 1.8 \times 10^{25}$). En effet, vous pouvez vérifier que $9^{4^{6 \times 7}} = 9^{4^{42}} = 9^{2^{84}} = 3^{2 \times 2^{84}} = 3^{2^{85}} \approx 10^{1.8 \times 10^{25}}$. Nous pouvons aussi utiliser la série

$$e = 1 + 1 + \frac{1}{2} + \frac{1}{6} + \frac{1}{24} + \dots$$

un cas spécial de la série

$$e^x = \sum_{i=0}^{\infty} \frac{x^i}{i!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \frac{x^4}{24} + \dots$$

où $x = 1$.

Sa fraction continue est $[2; 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1, 12, 1, 1, \dots]$, c'est-à-dire

$$e = 2 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{4 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{6 + \cfrac{1}{1 + \dots}}}}}}}}$$

7.1.8 $\pi = 3.1415926535897932385\dots$

Célèbre entre toute, la constante d'Archimède, π , en a fasciné plus d'un et surgit là où on l'attend le moins. Bien que certains se soient amusés à calculer π à plusieurs zillions de décimales, en voici déjà 200 :

$$\begin{aligned} \pi = & 3.1415926535897932384626433832795028841971 \\ & 6939937510582097494459230781640628620899 \\ & 8628034825342117067982148086513282306647 \\ & 0938446095505822317253594081284811174502 \\ & 8410270193852110555964462294895493038196\dots \end{aligned}$$

Les approximations

$$\pi \approx \frac{355}{113} \quad \varepsilon = O(10^{-7}) \quad \text{et} \quad \pi \approx \frac{80\,143\,857}{25\,510\,582} \quad \varepsilon = O(10^{-16}),$$

seront des plus utiles.

La fraction continue de π est « rugueuse » : $[3; 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, 2, 2, 2, \dots]$. C'est donc

$$\pi = 3 + \cfrac{1}{7 + \cfrac{1}{15 + \cfrac{1}{1 + \cfrac{1}{292 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{2 + \cfrac{1}{1 + \cfrac{1}{3 + \dots}}}}}}}}}$$

7.2 Remarques bibliographiques

La constante γ apparaît pour la première fois dans un article d'Euler publié en 1740 [40, 168], mais avec l'approximation 0.577218. Pour en savoir plus sur γ , consultez [67].

*
* * *

Hippase de Métaponte (c. 500 av. J.-C.) aurait été le premier à démontrer que $\sqrt{2}$ est *incommensurable*, c'est-à-dire irrationnel [50]. La conception grecque des nombres, en particulier dans la philosophie pythagoricienne, n'admettait jusqu'alors que les nombres naturels et les ratios de naturels. Tout ce qui n'était ni un nombre naturel ni un ratio de nombres naturel n'était pas, à leur sens, un nombre. Or voilà, ce nombre, $\sqrt{2}$, la longueur de la diagonale d'un carré, existe pourtant bel et bien... Quel pavé dans la mare pythagoricienne ! Le choc fut profond et l'odieux porté sur Hippase : selon une légende, il fut simplement expulsé de l'académie, selon une autre il périt en mer par la colère des dieux [18, 155]. D'autres diront que le châtiment s'est abattu sur lui pour avoir révélé la construction du dodécaèdre régulier (ou peut-être simplement du pentagone) [124].

Quel que fût le sort réservé à Hippase, il n'en demeure pas moins que $\sqrt{2}$ était déjà connu depuis longtemps. Les Babyloniens en connaissaient la valeur avec précision, car une tablette de terre cuite datant d'entre 1800 et 1600 av. J.-C. (la tablette YBC 7289 dans la collection babylonienne de Yale) donne la très bonne approximation $1 + \frac{24}{60} + \frac{51}{3600} + \frac{10}{216000}$, dont nous obtenons six chiffres décimaux exacts avec 1.41421296 [134]. Les Babyloniens obtenaient les racines carrées par un algorithme itératif basé sur la complétion de carrés [49, 138].

*
* * *

On retrouve $\sqrt{2}$ dans la définition du format de papier A, qui décrit, entre autres, les feuilles de taille A4 (qui correspondent plus ou moins au *US letter* et au format du cahier d'écolier). En 1786,

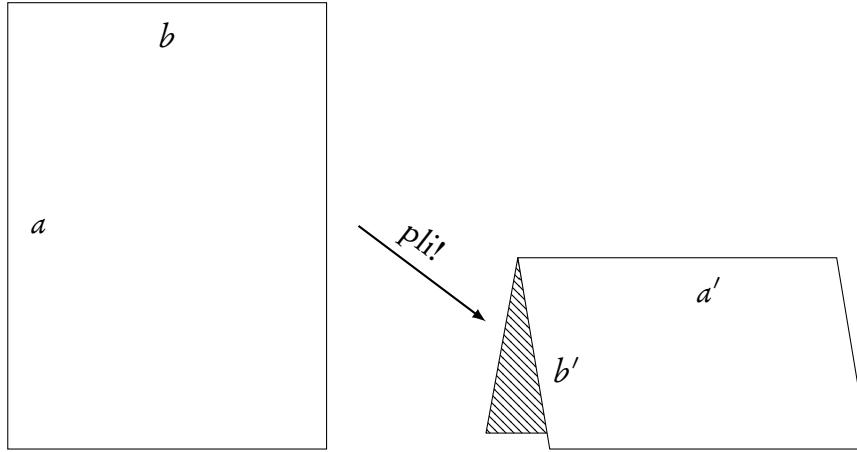


Figure 7.2.1 — Une feuille de côtés a et b produira un pli de même proportions a' et b' si, et seulement si, $\frac{a}{b} = \sqrt{2}$.

Lichtenberg, dans une correspondance à son collègue Beckman, fait remarquer les propriétés utiles de $\sqrt{2}$ [78]. En effet, il montre que la relation

$$\frac{a}{b} = \frac{2b}{a},$$

qui survient, par exemple, lorsqu'on plie une feuille de papier en deux dans la direction perpendiculaire à son côté le plus long (fig. 7.2.1), n'est satisfaite que par la « fraction »

$$\frac{a}{b} = \sqrt{2}.$$

Cette propriété, croit-on, est idéale pour la papeterie! En effet, dès 1798, le gouvernement français stipule une première loi (n° 2136) instaurant le format. C'est aujourd'hui un standard ISO [1].

*
* * *

Le nombre d'or, $\phi = \frac{1+\sqrt{5}}{2}$, ne cesse de fasciner et d'entretenir toutes sortes de mythes et de légendes. Ce serait le mathématicien Mark Barr, qui, au début du XX^e siècle, aurait inventé la notation ϕ en l'honneur de Phidias [94, p. 5–6]. Phidias (c. 480–430 av. J.-C.), architecte réputé, aurait, dit-on, bâti le Parthénon à Athènes en utilisant ça et là — et sans modération — les ratios de ϕ et ϕ^{-1} pour lui garantir des proportions harmonieuses — alors qu'il s'avère que le Parthénon utilise plutôt les proportions 3 : 2 [92, Ch. 5]. Pour plusieurs, ϕ est la *divine proportion*, manifestation suprême de l'harmonie universelle, et on la retrouverait partout en architecture, peinture, photographie et même dans la nature [131], menant cependant souvent à toutes sortes de dérives numérologico-mystiques [55]. D'autres, plus prosaïques, estiment la réputation de ϕ largement surfaite et considèrent les « preuves » de façon beaucoup plus critique [114, 115].

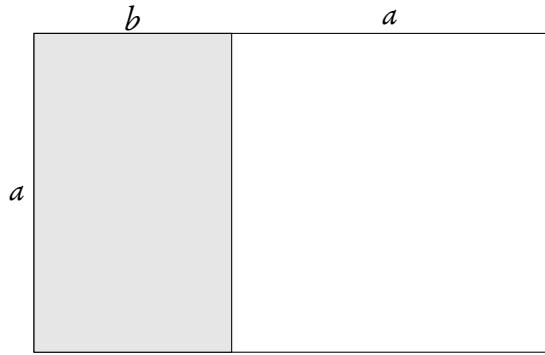


Figure 7.2.2 — Une feuille de côtés $a+b$ et a produira un pli de même proportions a et b si, et seulement si, $\frac{a}{b} = \phi$.

Si on laisse les délires numéologiques de côté, la « section dorée » résout l'équation

$$\frac{a}{b} = \frac{a+b}{a},$$

ce qui correspond géométriquement à la fig. 7.2.2. Trouvons la valeur de la proportion $\frac{a}{b}$:

$$\begin{aligned} \frac{a+b}{a} &= \frac{a}{b} \\ 1 + \frac{b}{a} &= \frac{a}{b} \\ \frac{a}{b} \left(1 + \frac{b}{a}\right) &= \frac{a}{b} \left(1 + \frac{a}{b}\right) \\ \frac{a}{b} + 1 &= \left(\frac{a}{b}\right)^2 \\ 0 &= \left(\frac{a}{b}\right)^2 - \frac{a}{b} - 1. \end{aligned}$$

C'est la forme $x^2 - x - 1 = 0$, et, grâce à l'équation quadratique, on trouve $x_0 = \frac{a}{b} = \frac{1 \pm \sqrt{5}}{2}$ (mais il faut bien voir que de parler de fraction, ici, c'est un abus, puisque ϕ est irrationnel).

*
* *
*

La formule de Stirling [71, 149], serait, selon Dutka [35] en fait due à de Moivre, le premier à s'intéresser au calcul de $n!$ pour de grandes valeurs de n . La formule de Stirling n'est pas la seule que l'on puisse trouver pour approximer $n!$, ni d'ailleurs celle qui converge le plus rapidement [111, 112, 135]. Quoi qu'il en soit, l'intérêt se porte surtout sur la fonction Gamma, notée $\Gamma(n)$, une généralisation de la factorielle où n peut être quelconque, et pas seulement entier [7, 162].

*
* *

Historiquement, l'apparition de e est lié au calcul des intérêts [100]. Disons que nous ayons un taux annuel d'intérêt r et une somme initiale S_0 . Au terme d'un an, nous aurons $S_1 = S_0(1+r)$ en banque. Au terme de la deuxième année, $S_2 = S_1(1+r) = S_0(1+r)^2$. Clairement, après a années, $S_a = S_0(1+r)^a$. Imaginons maintenant que le versement annuel ne nous convient plus et que nous le voulions plutôt m fois par an. Le calcul de l'intérêt devient donc, au t^{e} versement, $S_t = S_0(1 + \frac{r}{m})^t$, soit donc $S_m = S_0(1 + \frac{r}{m})^m$ au terme de la première année. Si on considère le cas spécial $r = 1$, et qu'on considère des versements toujours plus fréquents, on remarque que

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$$

et, en général,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n = e^x. \quad (7.2.1)$$

Nous retrouvons donc e dès qu'il est question de croissance exponentielle — intérêts comme épidémies. Si on définit le logarithme comme l'inverse de l'éq. (7.2.1), on comprend mieux pourquoi nous parlons de « logarithme naturel » lorsque nous considérons les logarithmes en base e .

*
* *

Les fractions continues sont encore un sujet de recherche à peu près actif [15, 17, 26, 79, 82, 118, 125, 130, 137]. Elles possèdent plusieurs propriétés intéressantes, surtout en termes de convergence et d'approximation. En effet, si on tronque une fraction continue à un certain point, la fraction qui résulte est la meilleure approximation possible étant donné un dénominateur du même ordre de magnitude. Ici, nous avons donné les fractions continues pour nos constantes choisies, mais dans la variété « descendante », où l'expansion se fait au dénominateur. Mais pourrait-on concevoir des fractions continues « ascendantes » avec l'expansion au numérateur? Mais oui! Engels [37] propose de faire justement cela! Par exemple, pour π , on trouverait

$$\pi = 3 + \cfrac{1}{8 + \cfrac{1}{17 + \cfrac{1}{300 + \cfrac{1}{1991 + \cdots}}}},$$

où les dénominateurs sont donnés par la séquence A006784 [146].

*
* *

Le nombre π est connu depuis longtemps. Les anciens Égyptiens, déjà, sans par ailleurs vraiment reconnaître π comme une constante universelle, utilisaient l'approximation $\pi \approx \frac{256}{81}$ dans le papyrus Rhind (qui daterait d'environ 1545 av. J.-C., mais qui serait une copie d'un original datant de 1840–1800 av. J.-C.) [56, 74, 75, 136]. Nous ne savons pas de façon certaine comment ils sont arrivés à cette approximation, mais les spéculations vont bon train [38, 54, 63, 127, 153, 166]. Il faudra attendre Archimète, au III^e siècle av. J.-C. pour trouver $3\frac{1}{7} < \pi < 3\frac{10}{71}$, grâce à la mesure des périmètres de deux ennécatahexagones réguliers (96 côtés), l'un circonscrit, l'autre excrit. La moyenne des deux périmètres s'approche de π et donne l'inégalité ci-dessus. C'est Zu Chongzhi, au V^e siècle qui trouvera la très bonne approximation $\pi \approx \frac{355}{113}$ [93, 104].

*

* *

Enfin, pour nous, faibles mortels, qui, incapables de se rappeler de centaines de chiffres de notre constante préférée, π , il y a toujours la *piphilologie*, la « science » de la construction de trucs mnémotechniques pour se souvenir d'un grand nombre de décimales de notre constante préférée. Par exemple, le « poème »

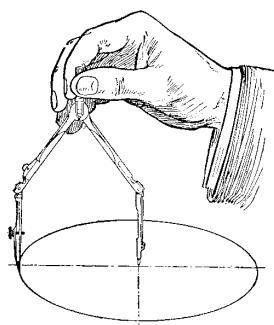
3. 1 4 1 5 9 2 6 5 3 5
Que j'aime à faire apprendre un nombre utile aux sages!

8 9 7 9
Immortel Archimède, artiste ingénieur,

3 2 3 8 4 5 2 6
Qui, de ton jugement, peut priser la valeur?

etc.

est plutôt facile à retenir et donne déjà 22 décimales.



8

Combinatoire

Sommaire. Factorielles, coefficients binomiaux et autres fonctions combinatoires sont au menu.

8.1 Combinatoire

Ce chapitre sera court mais intense. Nous regarderons comment calculer diverses fonctions combinatoires comme la factorielle, les coefficients binomiaux, multinomiaux et des approximations asymptotiques.

8.2 Factorielles

Produits, factorielles, et symboles de Pochhamer :

1. La factorielle, $n! = \Gamma(n+1) = 1 \times 2 \times 3 \times \cdots \times (n-1) \times n$.

(a) $n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n$ (voir § 8.4).

(b) $\ln n! \approx \left(n + \frac{1}{2}\right) \ln n - n + \ln \sqrt{2\pi}$.

(c) $\ln \Gamma(n) \approx \left(n - \frac{1}{2}\right) \ln n - n + \ln \sqrt{2\pi}$.

2. La factorielle ascendante, $n^{\bar{k}} = n \times (n+1) \times (n+2) \times \cdots \times (n+k-1) = \frac{(n+k-1)!}{(n-1)!}$.

3. La factorielle descendante, $n^{\underline{k}} = n \times (n-1) \times (n-2) \times \cdots \times (n-k+1) = \frac{n!}{(n-k)!}$.

De plus,

$$(a) \quad x^{\bar{n}} = (x+n)^{\underline{n}},$$

$$(b) \quad x^{\underline{n}} = (x-n)^{\bar{n}}.$$

8.3 Coefficients binomiaux & cie.

1. $n!$ est le nombre d'ordres distincts dans lesquels on peut arranger n objets.

2. Une k -permutation, $P(n, k) = \frac{n!}{(n-k)!}$, est le nombre de façons de prendre k objets parmi n , lorsqu'on conserve l'ordre;

3. Le coefficient binomial, $\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{P(n, k)}{k!}$, noté parfois $C(n, k)$, est le nombre de façons de prendre k objets parmi n mais en négligeant l'ordre.

(a) Il respecte la somme $\sum_{i=0}^n \binom{n}{i} = 2^n$.

(b) Il donne aussi les coefficients dans l'expansion du binôme $(x+y)^n$, soit

$$(x+y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i} = \sum_{i=0}^n \binom{n}{i} x^{n-i} y^i.$$

(c) Il admet la symétrie $\binom{n}{k} = \binom{n}{n-k}$.

(d) Les récurrences

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$$

et

$$\binom{n+1}{k} = \binom{n}{k-1} + \binom{n}{k}$$

permettent de le calculer (et de générer le triangle de Pascal).

4. Le nombre de combinaisons avec remise est donné par $\binom{n}{k} = \binom{n+k-1}{k}$.

5. Le coefficient multinomial,

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n}{n_1! n_2! \dots n_k!}$$

donne le nombre de permutations de n objets parmi lesquels n_1, n_2, \dots, n_k sont indistincts entre eux, et où $n = n_1 + n_2 + \dots + n_k$.

8.4 Remarques bibliographiques

Les notations pour les factorielles ascendantes et descendantes sont loin d'être standard et forment, pour ainsi dire, un joyeux carnaval. Alors que Pochhammer introduit la notation $(p)_q$ (pour autre chose complètement, $\binom{p}{q}$, le coefficient binomial) [129], elle est rapidement réutilisée. Steffesen utilise $z^{(n)}$ pour noter la factorielle descendante, $z^{\underline{n}}$ et $z^{(-n)}$ pour l'ascendante, $z^{\bar{n}}$ [148, p. 8]. D'autres auteurs utilisent $(z)_n$ pour la factorielle ascendante [4]. Knuth, quant à lui, propose les notations $z^{\bar{a}}$ et $z^{\underline{a}}$ pour les factorielles ascendantes et descendantes [86], généralement bien moins ambiguës — en effet, que représente $c^{(a+b)}$? Cette notation est reprise par d'autres auteurs [119].

*
* * *

La fonction gamma, $\Gamma(z)$, est une extension de la factorielle pour accommoder les valeurs réelles, y compris négatives. Emil Artin lui a consacré un sympathique opuscule [8], tandis d'autres se contentent de présenter des façons de la calculer, ou au moins de l'approximer : voyez [4, ch. 6] et [119, p. 139–142]. Si vous voulez approximer $n!$, vous avez plusieurs choix. Par exemple, la formule de Stirling [144, 145, 149],

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \dots\right)$$

est exacte, mais est souvent tronquée à seulement

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n,$$

ce qui donne quand même une bonne précision relative. Plusieurs proposent des approximations tronquées avec des corrections, par exemple Hodgman qui propose

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n-1}\right).$$

qui est en fait un peu moins bonne que la série tronquée en $1 + \frac{1}{12n}$ [70, p. 326]. Parmi les formules courtes meilleures que celles de Stirling tronquées, on trouve celle de Gosper [59],

$$n! \approx \sqrt{2\pi \left(n + \frac{1}{6}\right)} \left(\frac{n}{e}\right)^n,$$

qui est elle-même moins bien que celle que je propose¹, à savoir

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n \left(1 + \frac{1}{12n - \frac{1}{2}}\right).$$

1. Il faudra que je fasse un papier avec ça, éventuellement.

9

Astuces en vrac

Sommaire.

9.1 Astuces en vrac

9.2 Tests de divisibilité

Pour vérifier rapidement si un nombre se divise par 2, 3, 5, 7, nous avons heureusement des méthodes relativement simples. Évidemment, certains tests sont plus simples que d'autres (tester si un nombre est divisible par 1 ou deux est plutôt simple) tandis que d'autres sont plus laborieux.

9.2.1 Divisibilité par 2 ou 2^n

Pour savoir si un nombre est divisible par 2 il suffit de vérifier s'il est pair (et donc termine en 0, 2, 4, 6 ou 8). Pour les autres puissances :

- Il suffit de vérifier que le nombre formé par les deux derniers chiffres soit divisible par 4. On peut écrire un nombre n comme $n = 100k + r$, et puisque 100 est un multiple de 4, $100k$ est aussi un multiple de 4; il faut que r soit aussi divisible par 4 pour que $100k + r$ soit divisible par 4 (car avec $r = 4s$, on peut écrire $100k + r = 4 \cdot 25 \cdot k + 4 \cdot s = 4(25k + s)$).

- Il faut que le nombre formé par les trois derniers chiffres soit divisibles par 8. On peut écrire $n = 1000k + r$, et puisque $1000 = 8 \times 125$, les multiples de 1000 sont aussi des multiples de 8, il faut donc que r soit aussi visible par 8 pour que $1000k + r$ soit divisible par 8.
- Il suffit de vérifier que le nombre formé par les quatre derniers chiffres soit divisible par 16. On peut écrire un nombre comme $n = 10000k + r$, et puisque $10000 = 625 \times 16$, les multiples de 10000 sont aussi des multiples de 16, il faut donc que le reste soit aussi divisible par 16.
- Il suffit de vérifier que le nombre formé par les cinq derniers chiffres soit divisible par 32. Posons $n = 100000k + r$, et puisque $100000 = 3125 \times 32$, les multiples de 100000 sont des multiples de 32, il faut donc que le reste soit aussi divisible par 32.

Les tests se simplifient grandement lorsque les nombres sont exprimés en binaire :

- Par 2 : (Au moins) le dernier bit est zéro,
- Par 4 : (Au moins) les deux derniers bits sont zéro,
- ...
- Par 2^n : (Au moins) les n derniers bits sont zéro.

9.2.2 Divisibilité par 3 ou 6

Si la somme des chiffres est divisible par trois, le nombre l'est aussi — et ça s'applique récursivement. Par exemple, 531441 donne $5 + 3 + 1 + 4 + 4 + 1 = 18 \rightarrow 1 + 8 = 9$. Donc, 531441 est divisible par 3 car 9 est un multiple de 3 ($3^{12} = 531441$)!

Démonstration 9.2.1. *Démonstration par congruences.* Nous utiliserons la distributivité de modulo sur l'addition, soit donc $(a + b) \bmod m \equiv ((a \bmod m) + (b \bmod m)) \bmod m$, et sur le produit, $ab \bmod m \equiv (a \bmod m)(b \bmod m) \bmod m$.

Posons $n = a_n 10^n + a_{n-1} 10^{n-1} + \dots + a_1 10 + a_0$. Sachant que $10^k \pmod{3} \equiv 1$ (puisque $10^k = 99\dots9 + 1$, et que $99\dots9 \bmod 3 \equiv 0$), on

peut écrire

$$\begin{aligned}
 n \pmod{3} &\equiv a_n 10^n + a_{n-1} 10^{n-1} + \dots + a_1 10 + a_0 \\
 &\equiv ((a_n \pmod{3})(10^n \pmod{3}) + \dots + (a_0 \pmod{3})) \pmod{3} \\
 &\equiv (a_n \pmod{3} + a_{n-1} \pmod{3} + \dots + a_0 \pmod{3}) \pmod{3} \\
 &\equiv (a_n + a_{n-1} + \dots + a_1 + a_0) \pmod{3}.
 \end{aligned}$$

Et si $n \pmod{3} \equiv 0$, alors n est un multiple de 3. □

Démonstration 9.2.2. *Démonstration 2.* Une variante où nous passons les propriétés de modulo sous le tapis :

$$\begin{aligned}
 n &= a_n 10^n + a_{n-1} 10^{n-1} + \dots + a_1 10 + a_0 \\
 &= a_n(10^n - 1 + 1) + a_{n-1}(10^n - 1 + 1) + \dots + a_1(9 + 1) + a_0 \\
 &= a_n(99\dots9 + 1) + a_{n-1}(9\dots9 + 1) + \dots + a_2(99 + 1) + a_1(9 + 1) + a_0 \\
 &= \underbrace{a_n 999\dots9 + a_{n-1} 99\dots9 + a_2 99 + a_1 9}_{(\equiv 0 \pmod{3}) \text{ puisque tous sont multiples de 9, donc de 3}} + a_n + a_{n-1} + \dots + a_1 + a_0 \\
 &\equiv a_n + a_{n-1} + \dots + a_1 + a_0 \pmod{3}.
 \end{aligned}$$

□

9.2.3 Divisibilité par 5, 5^k , 10 k

Si le nombre termine en 0 ou 5, il est divisible par 5 — s'il termine en 0, c'est un multiple de 10. S'il termine en k zéros, il est divisible par 10 k . Pour 5 k , on commence par remarquer que 10 k peut s'écrire comme 10 k = 2 k 5 k . Ainsi, on peut écrire $n = c2^k5^k + r$, et n se divise par 5 k si le reste, le nombre formé par les k derniers chiffres, se divise par 5 k .

9.2.4 Divisibilité par 7

Si la différence entre le double du dernier chiffre et le reste du nombre est divisible par 7, alors le nombre l'est aussi.

Démonstration 9.2.3. *Test de divisibilité par 7.* Posons par hypothèse $n = 7k = a_n^n + a_{n-1}10^{n-1} + \dots + a_0$ pour $k \in \mathbb{Z}$. Alors,

$$\begin{aligned} n &= 10(a_n 10^{n-1} + a_{n-1} 10^{n-2} + \dots + a_2 10 + a_1) + a_0 \\ &= 10(a_n 10^{n-1} + a_{n-1} 10^{n-2} + \dots + a_1) + 20a_0 - 20a_0 + a_0 \\ &= 10(a_n 10^{n-1} + a_{n-1} 10^{n-2} + \dots + a_1 - 2a_0) + 21a_0 \\ 7k - 21a_0 &= 10(a_n 10^{n-1} + a_{n-1} 10^{n-2} + \dots + a_1), \end{aligned}$$

mais $7k - 21a_0 = 7(k - 3a_0)$ est forcément un multiple de 7; le facteur 10 n'y contribue en rien, et donc il faut

$$\underbrace{a_n 10^{n-1} + \dots + a_1}_{\text{le « reste du nombre »}} - 2a_0 \equiv 0 \pmod{7}$$

pour que le nombre soit divisible par 7. □

Démonstration 9.2.4. *Test de divisibilité par 7.* Un entier n peut être écrit comme $n = 10a + b$ (avec $0 \leq b < 10$). Comme $2 \cdot 10 \equiv -1 \pmod{7}$, on a que $10(a - 2b) \equiv n \pmod{7}$. Or, comme 10 n'est pas un multiple de 7, c'est $(a - 2b) \equiv n \pmod{7}$ qui doit être un multiple de 7; c'est-à-dire que $(a - 2b)$ doit être divisible par 7 pour que n le soit aussi. □

9.2.5 Divisibilité par 9

Il suffit que la somme des chiffres soit divisible par 9. La démonstration reprend la structure de la preuve du test de divisibilité par 3.

9.3 Remarques bibliographiques

Il existe un grand nombre de tests de divisibilité : de 2 à 30, et pour les nombres premiers comme 41, 43, etc. Cependant, plusieurs de ces tests sont compliqués, et reviennent, en termes du nombre d'opérations, à faire une division, et ne sont donc que d'une utilité très relative. Les tests de 2 à 9 sont sûrement plus utiles qu'un test pour 43, d'autant plus que 2, 3, 5 et 7 suffisent à tester correctement

$$1 - \left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{5}\right)\left(1 - \frac{1}{7}\right) = \frac{27}{35} \approx 77\%,$$

des nombres, un résultat que l'on obtient par inclusion/exclusion. On peut atteindre $\approx 80\%$ si on ajoute 11 et 13, mais il faudra aller jusqu'au 55^e nombre premier (257) pour tester environ 90% des nombres, et au 7397^e (75029) pour tester un peu plus de 95% des nombres! Aussi bien faire la division!

C'est Blaise Pascal, dans *De numeris multiplicibus* (1665), qui a proposé le test pour la division par 7, grâce à « technique du ruban » [122, 123, 150]. Cette technique, basée sur les congruences $(\text{mod } n)$, se généralise à un diviseur quelconque, et s'implémente assez bien avec des automates fini déterminisites [140].

*
* * *

Bibliographie

La bibliographie qui suit présente les références en ordre alphabétique des auteurs. En principe, aucun des ouvrages cités n'est un « introuvable ». Vous pourrez trouver les livres plus anciens grâce au projet de numérisation de l'*Internet Archive* [2]. Cette vaste collection comprend des ouvrages très variés, autant littéraires que scientifiques, numérisés à bonne résolution dans des formats pratiques comme PDF et DjVu. Pour plusieurs autres références, surtout les références aux articles relativement récents, nous vous invitons à vérifier auprès de votre institution quelles ententes ont été établies auprès des éditeurs. Les journaux de l'IEEE, de l'ACM et d'Elsevier sont souvent accessibles sans frais par l'intermédiaire de votre établissement scolaire, collégial ou universitaire. Plusieurs auteurs auront aussi, comme vous le découvrirez si vous utilisez un engin de recherche, mis à disposition leurs publications sur leurs pages personnelles. D'autres références, enfin, contiendront des URL qu'il vous suffira de suivre pour atteindre l'information.

- [1] ISO 216 :2007 *Writing Paper and Certain Classes of Printed Matter - Trimmed Sizes - A and B Series, and Indication of Machine Direction* — Rapport technique, ISO (2007).
- [2] *eBooks and Texts : Internet Archive* — <https://archive.org/details/texts> (2015).
- [3] N. H. Abel — *Beweis der Unmöglichkeit, algebraische Gleichungen von höheren Graden als dem vierten allgemein aufzulösen* — Journal für die reine und angewandte Mathematik, vol. 1 (1826), p. 65–84.
- [4] Milton Abramowitz, Irene A. Stegun (éds) — *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables* — Dover (1972).
- [5] Forman S. Acton — *Real Computing made Real* — Prometheus Books (1995).
- [6] Forman S. Acton — *Numerical Methods that (Usually) Work* — The American Mathematical Association (1997).
- [7] Emil Artin — *Einführung in die Theorie der Gammafunktion* — n° 1 dans la collection *Hamburger Mathematische Einzelschriften*, B. G. Teubner, Leipzig (1931).
- [8] Emil Artin — *The Gamma Function* — Holt, Rinehart and Winston (1964).
- [9] Stan Augarten — *Bit by Bit : An Illustrated History of Computers* — Ticknor & Fields (1984).
- [10] Jason Socrates Bardi — *The Calculus Wars : Newton, Leibniz, and The Greatest Mathematical Clash of All Time* — Thunder's Mouth Press (2006).
- [11] A. F. Beardon — *Sums of Powers of Integers* — American Mathematical Monthly, vol. 103, n° 3 (1996), p. 201–213.
- [12] E. G. Belaga — *On Computing Polynomials in One Variable with Initial Preconditioning of the Coefficients* — Problemi Kibernetiki, vol. 5 (1961), p. 7–15. En Russe.

- [13] Brian E. Blank — *The Calculus Wars, Reviewed* — Notices of the AMS, vol. 56, n° 5 (mai 2009), p. 602–610.
- [14] David Bohm — *Quantum Theory* — Dover (1989).
- [15] Amnon Bracha — *Application of Continued Fractions for Fast Evaluation of Certain Functions on a Digital Computer* — Rapport technique n° UIUCDCS-R-72-510, Dept. Computer Science, University of Illinois at Urbana Champaign (mars 1972).
- [16] Gilles Brassard, Paul Bratley — *Algorithmique : conception et analyse* — Masson / Presses de l'Université de Montréal (1987).
- [17] Claude Brezinski — *History of Continued Fractions and Padé Approximants* — Springer-Verlag (1980).
- [18] Luc Brisson, Alain-Philippe Segonds — *Jamblique : Vie de Pythagore* — Les Belles Lettres (2011).
- [19] Ian Bruce — *Napier's Logarithms* — American Journal of Physics, vol. 68, n° 2 (2000), p. 148–154.
- [20] Florent Cajori — *A History of Mathematical Notations (The Two Volumes Bound as One Edition)* — Dover (1993).
- [21] Gerolamo Cardano — *The Book of Games of Chances : The 16th-Century Treatise on Probability translated by Sydney Henry Gould* — Holt, Rinehart and Winston (1961).
- [22] Amel Chaabouni, Arezki Mohammedi — *Applications des mathématiques : la boîte à outils – bases théoriques, exemples et exercices résolus* — Presses polytechniques et universitaire romandes (2018).
- [23] Gabriel Cramer — *Introduction à l'analyse des lignes courbes algébriques* — Chez les frères Cramer et Cl. Philibert (1750).
- [24] Michael J. Crowe — *A History of Vector Analysis : The Evolution of the Idea of a Vectorial System* — Dover (1967).
- [25] Michael J. Crowe — *A History of Vector Analysis* — Rapport technique, University of Louisville (2002) (Notes de cours?).
- [26] Annie Cuyt, Vigdis Brevik Petersen, Brigitte Verdonk, Haakon Waadeland, William B. Jones, *et al.* — *Handbook of Continued Fractions for Special Functions* — Springer (2008).
- [27] Germund Dahlquist, Åke Björck — *Numerical Methods* — Dover (2003).
- [28] Gaston Darboux (éd.) — *Oeuvres de Fourier, tome second* — Gauthier-Villars et fils (1890).
- [29] Bibhutibhusan Datta, Avadhesh Narayan Singh — *History of Hindu Mathematics : A Source Book, Part I* — Motilal Barnasi Das (1935).
- [30] Harry F. Davis — *Fourier Series and Orthogonal Functions* — Dover (1963).
- [31] Nigel Derby — *A Search for Sums of Powers* — The Mathematical Gazette, vol. 99, n° 546 (2015), p. 416–421.
- [32] Robert L. Devaney, Linda Keen (éds) — *Chaos and Fractals : The Mathematics Behind the Computer Graphics* — n° 39 dans la collection *Procs. of Symposia in Applied Mathematics*, American Mathematical Society (1989).
- [33] W. S. Dorn — *Generalizations of Horner's Rule for Polynomial Evaluation* — IBM Journal of Research and Development, vol. 6, n° 2 (avril 1962), p. 239–245.
- [34] William Dunham — *Journey Through Genius : The Great Theorems of Mathematics* — Penguin Books (1990).
- [35] Jacques Dutka — *The Early History of the Factorial Function* — Archive for History of Exact Sciences, vol. 43, n° 3 (1991), p. 225–249.

- [36] S. H. Eisman — *Polynomial Evaluation Revisited* — C. of the ACM, vol. 6, n° 7 (1963), p. 384–385.
- [37] Friedrich Engel — *Entwicklung der Zahlen nach Strammbrüechen* — dans *Verhandlungen der 52, Versammlung deutscher Philologen und Schulmaenner in Marburg*, (1913), p. 190–191.
- [38] Hermann Engels — *Quadrature of the Circle in Ancient Egypt* — Historia Mathematica, vol. 4, n° 2 (1977), p. 137–140.
- [39] Gerald Estrin — *Organization of Computer Systems — The Fixed plus Variable Structure Computer* — dans *Procs. Western Joint IRE-AIEE-ACM Computer Conference*, (mai 1960), p. 33–40.
- [40] Leonard Euler — *De Progressionibus Harmonicis Observationes* — Comentarii Academiæ Scientiarum Petropolitanæ, vol. 7 (1740), p. 416–421.
- [41] Kenneth Falconer — *Fractals : A Very Short Introduction* — Oxford University Press (2013).
- [42] Kenneth Falconer — *Fractal Geometry : Mathematical Foundations and Applications* — 3^e éd., John Wiley & Sons (2014).
- [43] Johann Faulhabern — *Academia Algebrae, darinnen die miraculosische Inventiones, zur den höchsten Cossin weiters continuirt und profitiert werden* — Johann Reñling Verlag (1631).
- [44] Michael Field, Martin Golubitsky — *Symmetry in Chaos : A Search for Pattern in mathematics, Art, and Nature* — Oxford University Press (1992).
- [45] Jean-Paul Flad — *Les trois premières machines à calculer : Schickard (1623), Pascal (1642), Leibniz (1673)* — n° 93 dans la collection *Les conférences du Palais de la Découverte*, Série D, Palais de la Découverte (1963).
- [46] André Fortin — *Analyse numérique pour ingénieurs* — 2^{de} éd., Presses Internationales Polytechniques (2001).
- [47] A. Foucher de Careil — *Oeuvres de leibniz, publiées pour la première fois d'après les manuscrits originaux, avec notes et introductions, 7 vols.* — Firmin Didot Frères (1859–1875).
- [48] Jean-Baptiste Joseph Fourier — *Refroidissement séculaire du globe terrestre* — dans *Bulletin des sciences par la société philomathique de Paris*, Série 3, vol. 7, (avril 1820), p. 58–70.
- [49] David Fowler, Eleanor Robson — *Square Root Approximations in Old Babylonian Mathematics : YBC 7289 in Context* — Historia Mathematica, vol. 25 (1998), p. 366–378.
- [50] Kurt von Fritz — *The Discovery of Incommensurability by Hippasus of Metapontum* — Annals of Mathematics, Second Series, vol. 46, n° 2 (1945), p. 242–264.
- [51] Jérôme Gavin, Alain Schärlig — *Longtemps avant l'algèbre : la fausse position* — Presses polytechniques et universitaire romandes (2012).
- [52] Jérôme Gavin, Alain Schärlig — *Sept pères du calcul écrit, des chiffres romains aux chiffres arabes : 799–1202–1619* — Presses polytechniques et universitaire romandes (2018).
- [53] Jérôme Gavin, Alain Schärlig — *Schreyber alias Grammateus : de la « fausse position » aux timides débuts de l'algèbre* — BibNum, n° 575 (mai 2019), . <https://journals.openedition.org/bibnum/575>.
- [54] Paulus Gerdes — *Three Alternate Methods of Obtaining the Ancient Egyptian Formula for the Area of a Circle* — Historia Mathematica, vol. 12, n° 3 (1985), p. 216–268.
- [55] Matila Costiescu Ghyska — *Le Nombre d'Or : rites et rythmes pythagoriciens dans le développement de la civilisation occidentale* — Gallimard (1931).
- [56] Richard J. Gillings — *Mathematics in the Time of the Pharaohs* — MIT Press (1972).
- [57] James Gleick — *Isaac Newton* — Pantheon Books (2003).

- [58] Gene H. Golub, Charles F. van Loan — *Matrix Computation* — 3^e éd., Johns Hopkins University Press (1996).
- [59] R. William Gosper, Jr. — *Decision Procedure for Indefinite Hypergeometric Summation* — Procs. National Academy of Science USA, vol. 75, n° 1 (1978), p. 42–46.
- [60] Gotlib — *Rubrique-à-brac : l'intégrale* — Dargaud (2018).
- [61] Ronald L. Graham, Donald E. Knuth, Oren Patashnik — *Concrete Mathematics* — Addison-Wesley (1994).
- [62] Lucye Guilbeau — *The History of the Solution of the Cubic Equation* — Mathematical News Letters, vol. 5, n° 4 (décembre 1826), p. 8–12.
- [63] Michel Guillemot — *À propos de la « géométrie égyptienne des figures »* — Sciences & Techniques en perspectives, vol. 21 (1992), p. 125–146.
- [64] Alfred Rupert Hall — *Philosophers at War : The Quarrel between Newton and Leibniz* — Cambridge University Press (2002).
- [65] Richard W. Hamming — *Numerical Methods for Scientists and Engineers* — 2^{de} éd., Dover (1986).
- [66] G. H. Hardy, W. W. Rogosinski — *Fourier Series* — Dover (1999).
- [67] Julian Havil — *Gamma : Exploring Euler's Constant* — Princeton University Press (2003).
- [68] Bruce A. Hedman — *An Earlier Date for "Cramer's Rule"* — Historia Mathematica, vol. 26, n° 4 (1999), p. 365–368.
- [69] E. W. Hobson — *John Napier and the Invention of Logarithms, 1614* — Cambridge University Press (1914).
- [70] Charles D. Hodgman, et al. (éds) — *CRC Standard Mathematical Tables* — 10^e éd., Chemical Rubber Publishing Company (1955).
- [71] Francis Holliday — *The Differential Method : Or, A Treatise Concerning Summation and Interpolation of Infinite Series, by James Stirling* — E. Cave, Londres (1749).
- [72] W. G. Horner — *A New Method of Solving Numerical Equations of All Orders, by Continuous Approximations* — Philosophical Transactions of the Royal Society of London, vol. 109 (1819), p. 308–335.
- [73] Carolo Gustavo Iacobo Iacobi — *Fundamenta Nova Theoriæ Functionon Ellipticarum* — Regiomonti Sumtibus Fratrum Bornträger (1829). (Karl Gustav Jacob Jacobi).
- [74] Annette Imhausen — *Ägyptische algorithmen : Eine untersuchung zur den mittelägyptischen mathematischen aufgabentexten* — Harrassowitz Verlag.
- [75] Annette Imhausen — *Egyptian Mathematics* — dans Victor J. Katz (éd.), *The Mathematics of Egypt, Mesopotamia, China, India, and Islam : A Sourcebook*, Princeton University Press (2007), p. 7–57.
- [76] Eugene Isaacson, Herbert Bishop Keller — *Analysis of Numerical Methods* — Dover (1994).
- [77] Dunham Jackson — *Fourier Series and Orthogonal Polynomials* — Dover (2004).
- [78] Ulrich Joost, Albrecht Schöne (éds).
- [79] Oleg Karpenkov — *Geometry of Continued Fractions* — Springer-Verlag (2013).
- [80] Viktor J. Katz (éd.) — *The Mathematics of Egypt, Mesopotamia, China, India, and Islam : A Sourcebook* — Princeton University Press (2007).
- [81] Viktor J. Katz — *A History of Mathematics, an Introduction* — Addison-Wesley (2009).

- [82] A. Y. Khinchin — *Continued Fractions* — University of Chicago Press (1964).
- [83] A. Y. Khinchin — *Mathematical Foundations of Quantum Statistics* — Dover (1998).
- [84] Heinrich Kirchauer — *Photolithography Simulation* — Thèse de doctorat, Fakultät für Elektrotechnik, Technische Universität Wien (1998).
- [85] Donald E. Knuth — *Evaluation of Polynomials by Computer* — C. of the ACM, vol. 5, n° 12 (1962), p. 595–599.
- [86] Donald E. Knuth — *Two Notes on Notation* — American Mathematical Monthly, vol. 99, n° 546 (1992), p. 403–422.
- [87] Donald E. Knuth — *Johann Faulhaber and Sums of Powers* — Mathematics of Computation, vol. 61, n° 203 (1993), p. 277–294.
- [88] Donald E. Knuth — *The Art of Computer Programming Vol II: Seminumerical Algorithms* — Addison-Wesley Longman (1998).
- [89] A. A. Kosinski — *Cramer's Rule is due to Cramer* — Mathematical Magazine, vol. 74, n° 4 (2001), p. 310–312.
- [90] Eugene F. Krause — *Taxicab Geometry* — Dover (1986).
- [91] T. W. Körner — *Fourier Analysis* — Cambridge University Press (1992).
- [92] Geoff Lehman, Michael Weinman — *The Parthenon and Liberal Education* — State University of New York Press (2018).
- [93] Yan Li, Shiran Du — *Chinese Mathematics : A Concise History* — Clarendon Press, Oxford (1984).
- [94] Mario Livio — *The Golden Ratio : The Story of Phi, the World's Most Astonishing Number* — Broadway Books (2002).
- [95] Po-Shen Loh — *A simple Proof of the Quadratic Formula* — ArXiv preprint (décembre 2019).
- [96] Colin Maclaurin — *A Treatise of Algebra, in Three Parts* — A Millar and J. Nourse (1748).
- [97] Benoît B. Mandelbrot — *The Fractal Geometry of Nature* — W. H. Freeman and Company (1983).
- [98] Benoît B. Mandelbrot — *Les objets fractals : forme, hasard et dimension* — Champs, Flammarion (1995).
- [99] Benoît B. Mandelbrot — *Multifractals and $1/f$ Noise : Wild Self-Affinity in Physics (1963–1976)* — Springer (1999).
- [100] Eli Maor — *e : The Story of a Number* — Princeton Univerty Press (1994).
- [101] Eli Maor — *e : The Story of a Number* — Princeton University Press (2009).
- [102] M. le Marquis de l'Hospital — *Analyse des infiniment petits, pour l'intelligence des lignes courbes* — 2^{de} éd., François Montalant, Paris (1716).
- [103] Ernst Martin — *The Calculating Machines : Their History and Development* — MIT Press (1992). Translated and edited by P. A. Kidwell and M. R. Williams.
- [104] Jean-Claude Martzloff — *Histoire des mathématiques chinoises* — Masson (1987).
- [105] John Theodore Merz — *Leibniz* — William Blackwood and Sons (1883).
- [106] Uta C. Merzbach, Carl B. Boyer — *History of Mathematics* — John Wiley & Sons (2001).
- [107] Florence Messineo — *Le monde fascinant des objets fractals* — Ellipses (2015).
- [108] Jean-Michel Muller — *Elementary Functions : Algorithms and Implementation* — 3^e éd., Birkhäuser (2016).

- [109] Paul J. Nahim — *An Imaginary Tale : the story of $\sqrt{-1}$* — Princeton University Press (1998).
- [110] Mark Napier — *Memoirs of John Napier of Merchiston* — William Blackwood (1834).
- [111] Gergő Nemes — *New Asymptotic Expansion for the Gamma Function* — Archiv der Mathematik, vol. 95, n° 2 (2010), p. 161–169.
- [112] Gergő Nemes — *On the Coefficients of the Asymptotic Expansion of $n!$* — Journal of Integer Sequences, vol. 13, n° 6 (2010), p. 1–5. Article 10.6.6.
- [113] Ioanne Nepero — *Mirifici Logarithmorum Canonis Descriptio* — Andreæ Hart (1614).
- [114] Margerite Neveux — *Le nombre d'or est une affabulation* — La Recherche, n° 387 (2005), p. 26.
- [115] Margerite Neveux, Herbert E. Huntley — *Le nombre d'or : Radiographie d'un mythe, suivi de La Divine Proportion* — n° S108 dans la collection Points Sciences, Éditions du Seuil (2014).
- [116] Isaac Newton — *Opticks or A Treatise of the Reflections, Refractions, Inflections, & the Colors of Light* — Dover (1952).
- [117] R. W. D. Nickalls — *A New Approach to Solving the Cubic. Cardan's Solution Revealed* — The Mathematical Gazette, vol. 77, n° 480 (novembre 1993), p. 354–359.
- [118] Carl D. Olds, Andrew M. Rockett, Peter Szüsz — *Continued Fractions* — Random House (1963).
- [119] Frank W. J. Olver, Daniel W. Lozier, Ronald F. Boisvert, Charles W. Clark — *NIST Handbook of Mathematical Functions* — Dover/Cambridge University Press (2010).
- [120] Amos R. Omondi — *Computer Arithmetic : Algorithms, Architecture and Implementation* — Prentice Hall Series in Computer Science, Prentice Hall (1994).
- [121] V. Ya. Pan — *Methods of Computing Values of Polynomials* — Russian Mathematical Surveys, vol. 21 (1966), p. 105–136.
- [122] Blaise Pascal — *De numeris multiplicibus ex sola characterum numericorum additione agnoscendis* — dans *Traité du Triangle Arithmétique, avec quelques autres petit traitez sur la mesme matière*, Guillaume Desprez (1665), p. 42–48.
- [123] Blaise Pascal — *De numeris multiplicibus ex sola characterum numericorum additione agnoscendis* (1654) — dans Léon Brunschvigg, Pierre Boutroux (éds), *Œuvres de Blaise Pascal, publiées suivant l'ordre chronologique, avec documents complémentaires, introductions et notes*, vol. III, 2^{de} éd., Hachette (1923), p. 313–339.
- [124] Jean-Luc Périllié — *La découverte des incommensurables et le vertige de l'infini* — Cahiers philosophiques, n° 91 (2002), p. 9–29.
- [125] Oskar Perron — *Die Lehre van den Kettenbrüchen* — B. G. Teubner (1913).
- [126] Steven Pigeon — *Contributions à la compression de données* — Thèse de doctorat, Département d'informatique et de recherche opérationnelle, Faculté des Arts et des Sciences, Université de Montréal (décembre 2001).
- [127] Steven Pigeon — *Quadrature in ancient egypt, revisited* — dans *Procs. 7th European Congress of Mathematics*, (juillet 2016), p. CS-A-12.
- [128] Nikolaï Piskounov — *Calcul différentiel et Intégral, Tome 1* — 8^e éd., Éditions MIR (1978).
- [129] L. A. Pochhammer — *Ueber hypergeometrische Funktionen nter Ordnung* — Journal für die reine un angewandte Mathematik, n° 71 , p. 316–352.
- [130] William Porras — *Reduced Order Models via Continued Fractions Applied to Control Systems* — (1980). Peut-être une « thèse de B.sc »?

- [131] Alfred S. Posamentier, Ingmar Lehman — *The (Fabulous) Fibonacci Numbers* — Prometheus Books (2007).
- [132] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery — *Numerical Recipes in C : The Art of Scientific Computing* — 2^{de} éd., Cambridge University Press (1998).
- [133] Diogo Queiroz-Condé, Jean Chaline, Jacques Dubois — *Le monde des fractales : la nature trans-échelle* — Ellipses (2015).
- [134] Benoît Rittaud — *Le fabuleux destin de $\sqrt{2}$* — Le Pommier (2006).
- [135] Herbert Robbins — *A Remark on Stirling's Formula* — The American Mathematical Monthly, vol. 62, n° 1 (1955), p. 26–29.
- [136] Gay Robins, Charles Shute — *The Rhind Mathematical Papyrus : an ancient Egyptian text* — British Museum Publications (1987).
- [137] Andrew M. Rockett, Peter Szűsz — *Continued Fractions* — World Scientific Press (1992).
- [138] Peter Strom Rudman — *How Mathematics Happened : The First 50,000 Years* — Prometheus Books (2006).
- [139] A. S. Saidan — *The Arithmetic of Al-Uqlidisi : The Story of Hindu-Arabic Arithmetic as told in the Kitāb al-Fusūl fī al-Hisāb al-Hindī* — D. Reidel Publishing Co. (1978).
- [140] Jacques Sakarovitch — *Éléments de théorie des automates* — Vuibert (2003).
- [141] Frédéric Sarrus — *Nouvelle méthode pour la résolution des équations numériques* — Chez Bachelier (1933).
- [142] John Savage — *Factoring Quadratics* — The Mathematical Teacher, vol. 82, n° 1 (janvier 1989), p. 35–36.
- [143] Manfred Schroeder — *Fractals, Chaos, Power Laws : Minutes from an Infinite Paradise* — Dover (1999).
- [144] Neil J. A. Sloane — *A001163 : Stirling's formula : numerators of asymptotic series for Gamma function* — <http://oeis.org/A001163>.
- [145] Neil J. A. Sloane — *A001164 : Stirling's formula : denominators of asymptotic series for Gamma function* — <http://oeis.org/A001164>.
- [146] Neil J. A. Sloane — *A006784 : Engel expansion of Pi (Formerly M4475)* — <http://oeis.org/A006784>.
- [147] Ian N. Sneddon — *Fourier Transforms* — Dover (1995).
- [148] J. F. Steffensen — *Interpolation* — 2^{de} éd., Dover (2006).
- [149] James Stirling — *Methodus Differentialis, sive Tractatus de Summatione et Interpolatione Serierum Infinitarum* — G. Strahan, Londres (1730).
- [150] Roland Stowasser — *A Textbook Chapter from an Idea of Pascal* — For the Learning of Mathematics, vol. 3, n° 2 (novembre 1982), p. 25–30.
- [151] D. J. Struik (éd.) — *A Source Book in Mathematics, 1200–1800* — Harvard University Press (1969).
- [152] Tilo Strutz — *Bilddatenkompression : Grundlagen, Codierung, JPEG, MPEG, Wavelets* — 2^{de} éd., Friedrich Vieweg & Sohn / Studium Technik (2002).
- [153] Vasilii Vasil'evič Struve, Boris Turaev — *Mathematischer Papyrus des Staatlichen Museums der Schönen Künste in Moskau* — Quellen und Studien zur Geschichte der Mathematik, (1930), p. 45–49.
- [154] Bernd Sturmfels — *Solving Systems of Polynomials Equations* — American Mathematical Society (2002).

- [155] Thomas Taylor — *The Life of Pythagoras by Iamblichus* — The Theosophical Publishing House (1918).
- [156] Andrey N. Tikhonov, Vasily Y. Arsenin, Fritz John (éds) — *Solution of Ill-Posed Problems* — V. H. Winston & Sons (1977).
- [157] Andrey N. Tikhonov, A. V. Goncharsky, V. V. Stepanov, A. G. Yagola — *Numerical Methods for the Solution of Ill-Posed Problems* — Springer Science+Business Media (1995).
- [158] A. F. Timan — *Theory of Approximation of Functions of a Real Variable* — Pergamon Press (1963).
- [159] Georgi P. Tolstov — *Fourier Series* — Dover (1976).
- [160] Clifford Truesdell — *The New Bernoulli Edition* — ISIS, vol. 49, n° 1 (1958), p. 54–62.
- [161] L. Richard Turner — *Inverse of the Vandermonde Matrix with Applications* — Rapport technique n° TN D-3547, NASA (août 1966).
- [162] Mitsuru Uchiyama — *The Principal Inverse of the Gamma Function* — Proceedings of the American Mathematical Society, vol. 140, n° 4 (2012), p. 1343–1348.
- [163] Alexandre-Théophile Vandermonde — *Mémoire sur la résolution des équations* — Histoire de l'académie des Sciences, (1771), p. 365–416.
- [164] Vijay V. Vazirani — *Approximation Algorithms* — Springer (2003).
- [165] Tamás Vicsek — *Fractal Growth Phenomena* — World Scientific (1992).
- [166] Kurt Vogel — *Vorgriechische Mathematik 1 : Vorgeschichte und Ägypten* — Herman Schroedel Verlag K. G. (1958).
- [167] Henry S. Warren, Jr. — *Hacker's Delight* — 2^{de} éd., Addison-Wesley (2013).
- [168] Eric W. Weisstein — *Euler Mascheroni Constant* — <http://mathworld.wolfram.com/Euler-MascheroniConstant.html>.
- [169] David M. Young, Robert Todd Gregory — *A Survey of Numerical Mathematics, in Two Volumes, Vol I* — Dover (1988).
- [170] David M. Young, Robert Todd Gregory — *A Survey of Numerical Mathematics, in Two Volumes, Vol II* — Dover (1988).

Index

Les noms propres, les noms des algorithmes, et les concepts se retrouvent ici dans le même index. Les titres de livres apparaissent en italiques. Notez que les entrées sont indexées relativement au début des paragraphes qui abordent le sujet en question.

- Égyptiens, 110
- Symboles ~ $\Gamma(n)$, 108 ~ γ , 106 ~ ϕ , 107 ~ π , 110 ~ $\sqrt{2}$, 106
- 96-gone, 110
- A ~ Format, 106
- Algèbre ~ Rhétorique, 68
- Algorithm ~ Babylonien pour calculer \sqrt{x} , 106
- Archimète, 110
- Assembleur, 41
- Athènes, 107
- Babyloniens ~ Algorithme pour calculer \sqrt{x} , 106 ~ Et $\sqrt{2}$, 106
- Barr, Mark, 107
- Belaga ~ Algorithme de, 39–40
- Binôme, 36
- Brescia, Nicolo de, 67
- Calcul Différentiel ~ Invention du, 69
- Calcul différentiel, 57
- Cardan, voir Cardano
- Cardano, Gerolamo, 49, 68
- Chiffonniers, 69
- Clifford, William Kingdon, 97
- Cramer, Gabriel, 98
- Croisé ~ Produit, 75, 97
- Cubique, 48 ~ Compléter le cube, 49–50 ~ Décalage à l'ordonnée, 51–53 ~ Histoire de l'équation, 67
- Définition ~ Polynôme, 35–36
- Dérivée ~ Matrices, 96–97 ~ Matrices et vecteurs, 97 ~ Norme, 97 ~ Vecteur, 95–96
- Dérivées ~ Matrices et vecteurs, 95
- Descartes, René, 68
- Déterminant, 82–83, 98 ~ Méthode de Sarrus, 83
- Divine proportion, 107
- Dodécaèdre ~ régulier, construction du, 106
- Droite, 42
- Dyadique ~ Produit, 74–75
- Ennéacontahexagone, 110
- Équation ~ Et inéquations, 87
- Équation ~ Linéaire, 42
- Estrin, Gerald, 40
- Euler ~ Léonard, 106
- Évaluation ~ De polynômes, 37–41 ~ De polynômes, parallèle, 40–41
- Exponentiation ~ Rapide, 37
- Factorielle ~ Approximation par la fonction de Stirling, 108
- Factorisation ~ De polynômes, connues, 66–67 ~ Des polynômes, 65–67
- Fausse position, 69
- Ferro, Scipio, 67
- Fewnomial, 37
- Fibonacci ~ Nombre de, 107
- Floridas, 67
- Fonction ~ Gamma, 108
- Format ~ A, 106
- Formule ~ De Stirling, 108
- Fraction ~ continue, 109
- Gamma ~ Fonction, 108
- Géométrique ~ Raisonnement, 68
- Gibbs, Josiah Willard, 97
- Gotlib, Marcel, 69
- Heaviside, Oliver, 97
- Hippase de Métaponte, 106
- Histoire ~ De la notation ϕ pour le nombre d'or, 107 ~ Équation cubique, 67
- Homogène ~ Forme d'un polyôme, 41 ~ Système, 89
- Horner ~ Méthode de, 38
- Identités remarquables ~ Polyômes, 65–66
- Inclusion/exclusion, 118
- Inéquation ~ Et équations, 87
- Inquisition, 68
- Inverse ~ Matrice, 79–80
- Irréductible ~ Polynôme, 37
- Irrationnel ~ Preuve que $\sqrt{2}$ est, 106
- Leibniz, Gottfried Wilhelm, 69

Loh, Po-Shen, 68

* * *

MacLaurin, Colin, 98

Matrice ~ Addition/soustraction, 77 ~ Inverse, 79–80 ~ Norme, 81–82 ~ Notation, 71 ~ Produit avec un scalaire, 77 ~ Produit avec un vecteur, 77 ~ Produit avec une matrice, 78 ~ Transposée, 72, 79

Méthode ~ De la sécante, 53–57 ~ De Newton, 57–60 ~ Des points fixes, 60–63 ~ Numériques, remarques, 63

Mnémotechnique ~ Trucs, 110

Monôme, 36

* * *

Newton ~ Invention du calcul différentiel, 57 ~ Méthode de, 57–60

Newton, Isaac, 69

Nombre ~ D'or, 107 ~ De Fibonacci, 107 ~ Premier, 118

Norme, 80–82 ~ Matrice, 81–82 ~ Vecteur, 81

Notation ~ Matrices, 71 ~ Polynôme, 36–37 ~ Vecteurs, 71

* * *

Or ~ Nombre d', 107

* * *

Papyrus ~ Rhind, 110

Parthénon, 107

Pascal ~ Blaise, 119

Phidias, 107

Piphilologie, 110

Polynôme ~ Évaluation par la méthode de Horner, 38 ~ Binôme, 36 ~ De second degré, 42–48 ~ Définitions, 35–36 ~ Degré 0, 42 ~ Degré 1, 42 ~ Dense, 37 ~ Du troisième degré, 48–53 ~ Évaluation,

37–41 ~ Évaluation parallèle, 40–41 ~ Factorisation, 65–67 ~ Factorisations connues, 66–67 ~ Forme homogène, 41 ~ Identités remarquables, 65–66 ~ Irréductible, 37 ~ Monôme, 36 ~ Notation, 36–37 ~ Quadrinôme, 37 ~ Racine, 37, 41–64 ~ Tenu, 37 ~ Trinôme, 36 ~ Vocabulaire, 36–37

Polynômes ~ Évaluation avec factorisation, 38 ~ Évaluation, formes spéciales, 38

Position ~ Méthode de la fausse, 69

Premier ~ Nombre, 118

Produit ~ Croisé, 75, 97 ~ Dyadique, 74–75 ~ Scalaire, 73–74

Projection ~ Vecteur contre plan, 84–86 ~ Vecteur contre vecteur, 83–84

Pseudo-inverse ~ À droite, 90–92 ~ À gauche, 92–94

* * *

Quadratique, 42 ~ Compléter le carré, 43–45 ~ Décalage à l'ordonnée, 45–46 ~ Factorisation, 47–48 ~ Méthode du déterminant, 43 ~ Nouvelles méthodes, 68

Quadrinôme, 37

* * *

Régression, 94–95

Racine ~ Polynôme, 37, 41–64

Rang ~ D'un système, 88

Réduction ~ D'un polynôme à la forme homogène, 41

Regula falsi, 69

Remarques ~ Sur les méthodes numériques, 63

Rhétorique ~ Algèbre, 68

Rhind ~ Papyrus, 110

Ruban ~ Technique du, 119

* * *

Sarrus ~ Méthode de, 83

Savage, John, 68

Scalaire ~ Produit, 73–74 ~ Produit avec un, 73

Sécante ~ Méthode de la, 53–57

SIMD, 41

Sous-déterminé ~ Système, 90–92

Sparse ~ Polynôme, 37

Stirling ~ Formule de, 108 ~ James, 108

Sur-déterminé ~ Système, 92–93

Système ~ D'équations, 86–95 ~ Homogène, 89 ~ Résolution des systèmes d'équations, 88–95 ~ Rang d'un, 88 ~ Sous-déterminé, 90–92 ~ Sur-déterminé, 92–93

* * *

Tablette ~ YBC 7289, 106

Tartaglia, 67

Technique ~ du ruban, 119

Tenu ~ Polynôme, 37

Transposée ~ Matrice, 72, 79 ~ Vecteur, 72

Trinôme, 36

Truc ~ Mnémotechnique, 110

* * *

Vecteur ~ Addition/soustraction, 72 ~ Norme, 81 ~ Notation, 71, 72 ~ Opérations sur les, 72–75 ~ Produit avec un scalaire, 73 ~ Produit avec une matrice, 77 ~ Produit croisé, 75 ~ Produit dyadique, 74–75 ~ Produit scalaire, 73–74 ~ Projection contre un plan, 84–86 ~ Projection contre un vecteur, 83 ~ Transposée, 72

Viète, François, 68

* * *

YBC 7289, 106

Compendium de formules utiles pour l'informatique et l'analyse des algorithmes

LOREM IPSUM dolor sit amet, consectetur adipiscing elit. Aliquam consequat vel sapien in sodales. Mauris tincidunt pulvinar enim ut dictum. Sed lectus massa, maximus a est non, gravida tincidunt mi. Non, ce n'est pas un message secret. Oui, c'est exprès. Mauris molestie fringilla nunc vitae tincidunt. Nam imperdunt mauris volutpat, consectetur orci nec, rhoncus orci. Vestibulum vestibulum vel tellus vel lobortis. Suspendisse fermentum malesuada purus malesuada posuere. Nullam aliquet cursus ex, et lacinia ante lobortis quis. Sed vel posuere leo.

