Développements

Kevin Quirin

2011-2012

Table des matières

1 Mathématiques

1.1 Théorèmes d'Abel angulaire et taubérien faible

Gourdon (Analyse)

Remarques :

- Niels Henrik Abel, 1802-1829. Norvégien, mort de la tuberculose.
- Alfred Tauber, 1866-1942. Autrichien, mort dans le camp de Theresienstadt.

Théorème 1.1 : d'Abel angulaire
Soit $\sum a_n z^n$ une série entière de rayon de convergence 1, telle que $\sum a_n$ converge. On note f la somme de cette série sur le disque unité.

Pour $\theta_0 \in \left[0, \frac{\pi}{2}\right)$, on pose

$$\Delta_{\theta_0} = \{z \in \mathbb{C} \mid |z| < 1 \text{ et } \exists \rho > 0,\ \exists \theta \in [-\theta_0, \theta_0],\ z = 1 - \rho e^{i\theta}\}.$$

Alors

$$\lim_{\substack{z \to 1 \\ z \in \Delta_{\theta_0}}} f(z) = \sum_{n=0}^{\infty} a_n.$$

Démonstration. On note $S = \sum a_n$, $S_n = \sum_{k=0}^{n} a_k$ et $R_n = S - S_n$. Effectuons une transformation d'Abel :

$$\begin{aligned}
\left(\sum_{n=0}^{N} a_n z^n\right) - S_n &= \sum_{n=1}^{N} (R_{n-1} - R_n)(z^n - 1) \\
&= \sum_{n=0}^{N-1} R_n (z^{n+1} - 1) - \sum_{n=1}^{N} R_n (z^n - 1) \\
&= \sum_{n=0}^{N-1} R_n (z^{n+1} - z^n) - R_N (z^N - 1) \\
&= (z-1) \sum_{n=0}^{N-1} R_n z^n - R_N (z^N - 1)
\end{aligned}$$

Quand $N \to \infty$, on en déduit

$$f(z) - S = (z-1) \sum_{n=0}^{\infty} R_n z^n.$$

Prenons un $\varepsilon > 0$. Comme $\sum a_n$ converge, il existe $N \in \mathbb{N}$ tel que pour tout $n > N$, $|R_n| < \varepsilon$.

On a donc

$$\begin{aligned}
|f(z) - S| &\leqslant |z-1| \left| \sum_{n=0}^{N} R_n z^n \right| + \varepsilon |z-1| \left(\sum_{n=N+1}^{\infty} |z|^n \right) \\
&\leqslant |z-1| \left| \sum_{n=0}^{N} R_n z^n \right| + \varepsilon \frac{|z-1|}{1-|z|}
\end{aligned}$$

Soit $z \in \Delta_{\theta_0}$, qu'on écrit donc $z = 1 - \rho e^{i\theta}$. On a $|z|^2 = 1 - 2\rho\cos\theta + \rho^2$, et si $\rho \leqslant \cos\theta_0$, on a

$$\begin{aligned}\frac{|z-1|}{1-|z|} &= (1+|z|)\frac{|1-z|}{1-|z|^2} \\ &= \frac{\rho}{2\rho\cos\theta - \rho^2}(1+|z|) \\ &\leqslant \frac{2}{2\cos\theta - \rho} \\ &\leqslant \frac{2}{2\cos\theta_0 - \cos\theta_0} \\ &\leqslant \frac{2}{\cos\theta_0}\end{aligned}$$

Ainsi, si $|z-1|$ est assez petit, on a le résultat. □

REMARQUE – On note que la réciproque est fausse. On peut par exemple considérer $a_n = (-1)^n$. On a une réciproque si $a_n = o(1/n)$:

Théorème 1.2 : taubérien faible
Soit $\sum a_n z^n$ une série entière de rayon de convergence 1, avec $a_n = o(1/n)$. On note f la somme de la série, et on suppose

$$\exists S \in \mathbb{C},\ \lim_{\substack{x\to 1\\ x<1}} f(x) = S.$$

Alors

$$\sum a_n \text{ converge, et } \sum_{n=0}^{\infty} a_n = S.$$

Démonstration. On note pour tout n S_n la somme partielle de rang n de $\sum a_n$. On a alors, pour $n \in \mathbb{N}$ et $x \in (0,1)$:

$$S_n - f(x) = \sum_{k=1}^{n} a_n(1-x^k) - \sum_{k=n+1}^{\infty} a_k x^k.$$

Pour $x \in (0,1)$, on a $1 - x^k \leqslant k(1-x)$, et donc

$$\begin{aligned}|S_n - f(x)| &\leqslant (1-x)\sum_{k=1}^{n} k|a_k| + \sum_{k=n+1}^{\infty} \frac{k|a_k|}{n}x^k \\ &\leqslant (1-x)Mn + \frac{\sup_{k>n} k|a_k|}{n(1-x)}\end{aligned}$$

où M majore $(k|a_k|)$.

Soit maintentant $\varepsilon \in (0,1)$. On a

$$\left|S_n - f\left(1 - \frac{\varepsilon}{n}\right)\right| \leqslant M\varepsilon + \frac{\sup_{k>n} k|a_k|}{\varepsilon}.$$

Comme $ka_k \to 0$, on peut choisir un rang N_0 assez grand pour qu'à partir de ce rang, $\sup_{k>N_0} k|a_k| < \varepsilon^2$, et on a alors pour $n > N_0$

$$\left|S_n - f\left(1 - \frac{\varepsilon}{n}\right)\right| \leqslant M\varepsilon + \varepsilon.$$

Enfin, comme $f(x) \to S$ quand $x \to 1^-$, on a un rang N_1 tel que $|f(1-\varepsilon/n) - S| < \varepsilon$, d'où

$$|S_n - S| \leqslant (M+2)\varepsilon.$$

□

REMARQUE – On notera qu'il existe un théorème Taubérien fort (*a.k.a.* théorème taubérien d'Hardy-Littlewood) où il suffit de supposer $a_n = \mathscr{O}(1/n)$. La preuve fait appel au théorème de Weierstraß.

1.2 Simplicité de $\mathfrak{A}_n$

Perrin (n.d.)

Théorème 1.3
Pour tout $n \geqslant 5$, le groupe alterné $\mathfrak{A}_n$ est simple.

Démonstration. On commence par le cas $n = 5$:

Lemme 1.4
Le groupe $\mathfrak{A}_5$ est simple.

Démonstration. On classe les éléments de $\mathfrak{A}_5$ selon leur ordre. On a :
– 15 éléments d'ordre 2 (les produits de deux transpositions) ;
– 20 éléments d'ordre 3 (les 3-cycles) ;
– 24 éléments d'ordre 5 (les 5-cycles).
Chacune de ces classes est une classe de conjugaison.

Soit donc $H \lhd \mathfrak{A}_5$, non trivial. Si H contient un élément d'odre 3 (resp. 2, resp. 5), il les contient tous.

Comme $|H| > 1$, il contient au moins une de ces classes. Mais ni 16, ni 21 ni 25 ne divise 60. Donc il en contient au moins 2, et donc $|H| \geqslant 36$. Donc $H = \mathfrak{A}_5$. ◊

Soit maintenant $n > 5$. On pose $E = [\![1,]\!]$, et on se donne $H \lhd \mathfrak{A}_n$ non trivial.

Soit $\sigma \in H$, $\sigma \neq id$. On a donc

$$\exists a \in E,\ b = \sigma(a) \neq a.$$

Comme $n > 5$, il existe c différent de $a, b, \sigma(b)$: soit τ le 3-cycle $(a\ c\ b)$. On pose

$$\rho = \tau\sigma\tau^{-1}\sigma^{-1} = (a\ c\ b)(\sigma(a)\ \sigma(b)\ \sigma(c)).$$

L'ensemble $F = \{a, b, c, \sigma(a), \sigma(b), \sigma(c)\}$ a au plus 5 éléments ($b = \sigma(a)$). Quitte à en rajouter, on peut supposer qu'il en a exactement 5.

On a $\rho(F) = F$ et $\rho_{|E\setminus F} = id_{|E\setminus F}$. On note enfin que $\rho \neq id$, car $\rho(b) = \tau\sigma(b) \neq b$ car $\sigma(b) \neq \tau^{-1}(b) = c$.

On plonge $\mathfrak{A}(F) \cong \mathfrak{A}_5$ dans $\mathfrak{A}_n$ par l'application

$$i : \begin{array}{ccc} \mathfrak{A}(F) & \longrightarrow & \mathfrak{A}_n \\ u & \longmapsto & \overline{u} \ : \ \begin{array}{l} \overline{u}_{|F} = u \\ \overline{u}_{|E\setminus F} = id_{|E\setminus F} \end{array} \end{array}.$$

On pose $H_0 = H \cap \mathfrak{A}(F)$. $\rho_{|F} \in H_0$, et H_0 est clairement distingué dans $\mathfrak{A}(F)$. $\mathfrak{A}(F)$ est simple, et donc $H_0 = \mathfrak{A}(F)$.

Pour terminer la preuve, on a maintenant besoin du

Lemme 1.5
Les 3-cycles engendrent $\mathfrak{A}_n$.

Démonstration. Comme les produits d'un nombre pair de transpositions engendrent $\mathfrak{A}_n$, il suffit de montrer que

ces produits peuvent être écrit avec des 3-cycles :

$$(a\ b)(b\ c) = (a\ b\ c)$$
$$(a\ b)(a\ c) = (a\ c\ b)$$
$$(a\ b)(c\ d) = (a\ b)(a\ c)(a\ c)(c\ d) = (a\ c\ b)(a\ c\ d)$$

Voilà. ◊

Soit u un 3-cycle de $\mathfrak{A}(F)$. Alors $\overline{(u)}$ est un 3-cycle dans H. Comme les 3-cycles sont tous conjugués dans $\mathfrak{A}_n$, H contient tous les 3-cycles, et donc on peut conclure par le lemme. □

1.3 Théorème de Banach-Steinhaus ; applications aux séries de Fourier

Gourdon (Analyse)

Remarques :

- Stefan Banach. Polonais. 1892-1945. Mort d'un cancer du poumon.
- Hugo Steinhaus. Polonais. 1887-1972. A découvert Banach dans un parc.

Théorème 1.6 : de Banach-Steinhaus
Soient E un espace de Banach, et F un espace vectoriel normé. Soit $H \subset \mathscr{L}_c(E,F)$. Alors une et une seule des propositions suivantes est vraie :

(i) $(|||f|||)_{f\in H}$ *est borné ;*

(ii) $\exists x \in E,\ \sup\limits_{f\in H} \|f(x)\| = \infty$

Démonstration. On rappelle le théorème de Baire :

Théorème 1.7 : de Baire
*Dans un espace métrique complet, toute intersection dénombrable d'ouverts dense est dense (*i.e *toute réunion dénombrable de fermés d'intérieurs vides est d'intérieur vide).*

Posons pour tout $k \in \mathbb{N}$:

$$\Omega_k = \{x \in E \mid \sup_{f\in H} \|f(x)\| > k\}.$$

Par continuité des f et de la norme, les Ω_k sont des ouverts.

Si tous les Ω_k sont denses, alors par théorème de Baire, leur intersection l'est aussi, et donc est non vide :

$$\exists x \in E,\ \sup_{f\in H} \|f(x)\| = \infty.$$

Sinon, il existe k tel que Ω_k ne soit pas dense :

$$\exists x_0 \in E,\ \exists \rho > 0,\ B(x_0,\rho) \cap \Omega_k = \emptyset.$$

On a donc, pour tout $x \in B(x_0,\rho)$, $\|f(x)\| \leqslant k$, et donc :

$$\begin{aligned} \forall x \in B(0,\rho),\ \forall f \in H,\ \|f(x)\| &= \|f(x+x_0) - f(x_0)\| \\ &\leqslant \|f(x+x_0)\| + \|f(x_0)\| \\ &\leqslant 2k \end{aligned}$$

D'où le résultat. □

On note maintenant $\mathscr{C}_{2\pi}$ l'ensemble des fonctions 2π-périodiques de $\mathbb{R}$ dans $\mathbb{C}$, muni de la norme $\|f\| = \sup\{|f(t)| \mid -\pi \leqslant t \leqslant \pi\}$.

Pour $f \in \mathscr{C}_{2\pi}$ et $p \in \mathbb{Z}$, on note

$$c_p(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) e^{-ipt} \mathrm{d}t$$

le p-ième coefficient de Fourier de f.

On pose enfin pour tout $n \in \mathbb{N}^*$:

$$\ell_n : \begin{array}{ccc} \mathscr{C}_{2\pi} & \longrightarrow & \mathbb{C} \\ f & \longmapsto & \sum_{p=-n}^{n} c_p(f) \end{array} .$$

Théorème 1.8
Il existe une fonction continue dont la série de Fourier diverge.

Démonstration. On utilise la relation classique suivante :

Lemme 1.9
Pour tous n et t :

$$\sum_{p=-n}^{n} e^{-ipt} = \frac{\sin((2n+1)t/2)}{\sin(t/2)}.$$

On a donc, pour tout $f \in \mathscr{C}_{2\pi}$,

$$\ell_n(f) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sin((2n+1)t/2)}{\sin(t/2)} f(t) \mathrm{d}t.$$

Donc ℓ_n est continue, et

$$|||\ell_n||| \leqslant \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\sin((2n+1)t/2)}{\sin(t/2)} \right| \mathrm{d}t.$$

Montrons qu'on a en fait égalité.

Posons pour cela pour tout $\varepsilon > 0$,

$$f_\varepsilon : \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{C} \\ t & \longmapsto & \frac{\sin((2n+1)t/2)}{|\sin((2n+1)t/2)|+\varepsilon} \end{array} .$$

Pour tout $\varepsilon > 0$ et pour tout $t \in \mathbb{R}$, on a $|f_\varepsilon(t)| < 1$, et f_ε est continue.

Par théorème de convergence dominée, on a donc

$$\lim_{\varepsilon \to 0} \|\ell_n(f_\varepsilon)\| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\sin((2n+1)t/2)}{\sin(t/2)} \mathrm{d}t.$$

On a donc :

$$|||\ell_n||| = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\sin((2n+1)t/2)}{\sin(t/2)} \right| \mathrm{d}t.$$

Comme $|\sin(t/2)| \leqslant |t/2|$ pour tout t, on a donc

$$\begin{aligned} |||\ell_n||| &\geqslant \frac{1}{\pi} \int_{-\pi}^{\pi} \left| \frac{\sin((2n+1)t/2)}{t} \right| \mathrm{d}t \\ &\geqslant \frac{2}{\pi} \int_{0}^{(2n+1)\pi/2} \left| \frac{\sin u}{u} \right| \mathrm{d}u \end{aligned}$$

Donc, comme $\frac{\sin u}{u}$ n'est pas intégrable, on a

$$\lim_{n\to\infty} |||\ell_n||| = \infty.$$

Par théorème de Banach-Steinhaus, comme $\mathscr{C}_{2\pi}$ est complet, il existe f tel que $\sup_{n>0} |\ell_n(f)| = \infty$, c'est-à-dire que la série de Fourier de f diverge en 0. □

1.4 Nombres de Bell

Francinou *et al.* (Algèbre 1)

Remarque – Eric Temple Bell. Écossais/Américain. 1883-1960.

On appelle, pour $n \in \mathbb{N}^*$, le *n-ième nombre de Bell* B_n le nombre de partitions de $[\![1, n]\!]$. Alors

Théorème 1.10
On a pour tout $n \in \mathbb{N}^*$:

$$B_n = \frac{1}{e} \sum_{p=0}^{\infty} \frac{n^p}{p!}.$$

Démonstration.

Lemme 1.11
On pose $B_0 = 1$. On a alors la formule de récurrence

$$B_{n+1} = \sum_{k=0}^{n} C_n^k B_k.$$

Démonstration. Soit $k \in [\![0, n]\!]$. On appelle k l'ensemble des partitions de $[\![1, n+1]\!]$ pour lesquelles le sous-ensemble contenant $n+1$ est de cardinal $k+1$.

Il faut donc "choisir" k éléments dans $[\![1, n]\!]$, puis choisir une partition des $n-k$ éléments restants. On a donc

$$\operatorname{Card} E_k = C_n^k B_{n-k}.$$

Comme les E_i forment une partition de l'ensemble des partitions de $[\![1, n+1]\!]$, on a

$$B_{n+1} = \sum_{k=0}^{n} C_n^k B_{n-k},$$

et on a le résultat par un changement d'indice. ◊

Posons

$$f(z) = \sum_{n=0}^{\infty} \frac{B_n}{n!} z^n.$$

Lemme 1.12
Le rayon de convergence R de la série définissant f n'est pas nul. De plus :

$$f(z) = e^{e^z - 1}.$$

Démonstration. On commence par montrer par récurrence $B_n \leqslant n!$, grâce à la relation de récurrence trouvée plus tôt.

On a donc $R \geqslant 1$. Notons $D = [-R, R]$. On a pour $z \in D$:

$$f'(z) = \sum_{n=1}^{\infty} \frac{B_n}{(n-1)!} z^{n-1} = \sum_{n=0}^{\infty} \frac{B_{n+1}}{n!} z^n.$$

On a donc

$$f'(z) = \sum_{n=0}^{\infty} \frac{1}{n!} \left(\sum_{k=0}^{n} C_n^k B_k \right) z^n = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{\infty} \frac{B_k}{k!(n-k)!} \right) z^n.$$

On reconnait bien là le produit de Cauchy des séries f et exp, et on a donc l'équation différentielle

$$f' = f \cdot \exp .$$

On en déduit que f est de la forme $f(z) = Ce^{e^z}$. Comme $f(0) = 1$, on a le résultat. ◊

On a donc, en redéveloppant

$$e^{e^z} = \sum_{n=0}^{\infty} \sum_{k=0}^{\infty} \frac{(nz)^k}{n!k!}.$$

En posant $u_{n,k} = \frac{(nz)^k}{n!k!}$, on a

$$\sum_{k=0}^{\infty} |u_{n,k}| = \frac{e^{|nz|}}{n!},$$

et

$$\sum_{n=0}^{\infty} \frac{e^{|nz|}}{n!} = e^{e^{|}z|}.$$

La série double et sommable, et donc on peut intervertir la sommation. On a donc

$$f(z) = \frac{1}{e} \sum_{k=0}^{\infty} \left(\sum_{n=0}^{\infty} \frac{n^k}{k!} \right) \frac{z^k}{k!}.$$

Par unicité du développement en série entière, on a bien le résultat. □

1.5 Décomposition de Bruhat

Francinou *et al.* (Algèbre 1)

Remarque – François George René Bruhat. Français. 1929-2007.

Définition 1.13
Un drapeau d'un espace vectoriel de dimension finie E *est une suite finie strictement croissante pour l'inclusion de sous-espaces vectoriels de* E*, le premier étant l'espace nul et le dernier* E *tout entier.*

Notation – On notera :
– T_s l'ensemble des matrices triangulaires supérieures inversibles ;
– P_σ la matrice de permutation associée à la permutation σ ;
– $\mathscr{D}$ l'ensemble des drapeaux d'un espace vectoriel.

Théorème 1.14
On a :

$$GL_n(\mathbb{K}) = \bigsqcup_{\sigma \in \mathfrak{S}_n} T_s \sigma T_s.$$

Démonstration. Soit $A \in GL_n(\mathbb{K})$. Posons quelques notations :
– $E_{i,j}$ est la matrice dont tous les coefficients sont nuls, sauf celui d'indices i, j ;
– pour $i \neq j$ et $\lambda \in \mathbb{K}$, $T_{i,j}(\lambda) = Id + \lambda E_{i,j}$;
– pour tout i et tout $\alpha \neq 0$, $D_i(\alpha) = Id + (\alpha - 1)E_{i,j}$.
Multiplier A à droite par $T_{i,j}(\lambda)$ revient à faire l'opération sur les colonnes $C_i \leftarrow C_i + \lambda C_j$, et à gauche revient à faire l'opération sur les lignes $L_i \leftarrow L_i + \lambda L_j$.

Multiplier A à droite par $D_i(\alpha)$ revient à faire l'opération sur les colonnes $C_i \leftarrow \alpha C_i$, et à droite revient à faire l'opération sur les lignes $L_i \leftarrow \alpha L_i$.

On applique l'algorithme suivant :

Soit i_1 le plus grand indice k tel que $a_{k,1} \neq 0$. On fait les opérations $L_i \leftarrow L_i - \frac{a_{i,1}}{a_{i_1,1}} L_{i_1}$ pour $i \neq i_1$ et $C_j \leftarrow C_j - \frac{a_{i_1,j}}{a_{i_1,1}} C_1$.

Cela ne revient qu'à multiplier à gauche et à droite par des matrices de T_s. On termine par $C_1 \leftarrow \frac{1}{a_{i_1,1}} C_1$ pour être dans la situation :

$$\begin{pmatrix} 0 & & & \\ 0 & & & \\ \vdots & & & \\ 0 & & & \\ 1 & 0 & 0 & \cdots \\ 0 & & & \\ \vdots & & & \\ 0 & & & \end{pmatrix}$$

On prend ensuite i_2 le plus grand indice k tel que $a_{k,2} \neq 0$. Notons que $i_2 \neq i_1$. Par les mêmes opérations, on annule les coefficients de la colonne 2 et de la ligne i_2 : cela ne modifie par les 0 de la colonne 1 et de la ligne i_1.

On itère le procédé, et on obtient une matrice de permutation P_σ, où σ est la permutation définie par

$(1\ i_1\ i_{i_1}\ \cdots)$.

On a donc une décomposition $A = T_1 P_\sigma T_2$, $T_1, T_2 \in T_s$.

Supposons qu'on ait deux décompositions $T_1 P_\sigma = P_\tau T_2$.

Alors $T_2 = P_{\tau^{-1}} T_1 \P_\sigma$. Supposons $\sigma \neq \tau$: il existe i tel que $\sigma(i) < \tau(i)$.

Le coefficient i, i de T_2 est non nul car T_2 inversible, et l'égalité précédente nous donne l'égalité :

$$T_2(i,i) = T_1(\tau(i), \sigma(i)) = 0,$$

d'où une contradiction. □

Théorème 1.15
L'action de $GL_n(\mathbb{K})$ sur $\mathscr{D} \times \mathscr{D}$ possède $n!$ orbites.

Démonstration. $GL_n(\mathbb{K})$ agit à gauche sur $\mathscr{D}$, de façon transitive. Le stabilisateur du drapeau canonique est T_s, et donc $\mathscr{D}$ est en bijection avec le quotient $GL_n(\mathbb{K})/T_s$.

Soit $(A, B) \in GL_n(\mathbb{K})/T_s \times GL_n(\mathbb{K})/T_s$. Alors

$$\begin{aligned}(A,B) &\sim A(I_n, A^{-1}B) \\ &\sim A(I_n, T_1 P_\sigma T_2) \\ &\sim A T_1 (I_n, P_\sigma).\end{aligned}$$

Donc chaque orbite a un élément de la forme (I_n, P_σ). Supposons qu'il existe $\sigma, \tau \in \mathfrak{S}_n$ dans une même orbite.

Alors il existe $A \in GL_n(\mathbb{K})$ telle que $(I_n, P_\sigma) = (A, AP_\tau)$.

Alors $A \in T_s$, et donc $\exists T \in T_s,\ AP_\tau = P_\sigma S$. Par décomposition de Bruhat, $\sigma = \tau$, ce qui est faux par hypothèse.

Donc on a une bijection entre les orbites et les permutations, d'où les $n!$ orbites. □

1.6 Lemme de Burnside

Francinou *et al.* (Algèbre 2)

Remarque – William Burnside. Anglais. 1952-1927.

Définition 1.16
On dit qu'un groupe G est d'exposant fini si $\exists k \in \mathbb{N},\ \forall g \in G,\ g^k = e$.

Si G est d'exposant fini, on appelle exposant de G *le plus petit k qui convienne, et si G n'est pas d'exposant fini, l'exposant de G est $+\infty$.*

Historique : En 1902, Burnside se demande si un groupe de type fini d'exposant fini est nécessairement fini. Ce n'est qu'en 1975 qu'on a pu répondre par la négative, par un contre-exemple.

Théorème 1.17
Soit G un sous-groupe de $GL_n(\mathbb{C})$.

Alors G est fini si et seulement si *G est d'exposant fini.*

Notation – On notera $\omega(g)$ l'ordre d'un élément d'un groupe.

Démonstration. Le sens réciproque est évident, on peut considérer par exemple $k = \prod_{g\in G} \omega(g)$.

Pour le sens direct, commençons par montrer un lemme sur les matrices nilpotentes :

Lemme 1.18
Soit $A \in \mathscr{M}_n(\mathbb{C})$ telle que $\forall k \in \mathbb{N}^*,\ \mathrm{Tr}(A^k) = 0$. Alors A est nilpotente.

Démonstration. Supposons que A n'est pas nilpotente. Alors A possède des valeurs propres non nulles $\lambda_1, \dots, \lambda_r$ de multiplicités respectives $n_1, \dots, n_r$.

L'hypothèse nous donne donc (quitte à trigonaliser la matrice) :

$$\forall k \in \mathbb{N}^*,\ n_1\lambda_1^k + \cdots + n_r\lambda_r^k = 0.$$

En écrivant ces relations pour $k = 1, \dots, r$, on obtient que $(n_1, \dots, n_r)$ est une solution non nulle du système

$$\begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & \lambda_r \\ \lambda_1^2 & \lambda_2^2 & \cdots & \lambda_r^2 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^r & \lambda_2^r & \cdots & \lambda_r^r \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix} = 0$$

Or le déterminant de la matrice du système est non nul car les λ_i sont tous distincts par hypothèse (matrice de Vandermonde). Donc $(n_1, \dots, n_r) = 0$, ce qui est une contradiction. ◊

Soit $(M_i)_{1\leqslant i\leqslant m}$ une base de $\mathrm{Vect}(G)$, où tous les M_i sont dans le groupe G.

Introduisons la fonction

$$f : \begin{array}{ccc} G & \longrightarrow & \mathbb{C}^m \\ A & \longmapsto & (\mathrm{Tr}(AM_i))_{1\leqslant i\leqslant m} \end{array}.$$

Lemme 1.19
Si $f(A) = f(B)$, alors $AB^{-1} - I_n$ est nilpotente.

De plus, f est injective.

Démonstration. Comme (M_i) est une base, et que Tr est linéaire, on a $\mathrm{Tr}(AM) = \mathrm{Tr}(BM)$ pour tout M dans $\mathrm{Vect}(G)$, et en particulier pour tout M dans G.

Soit $D = AB^{-1} \in G$. Alors pour tout $k \geqslant 1$:

$$\begin{aligned}\mathrm{Tr}(D^k) &= \mathrm{Tr}\left(AB^{-1}D^{k-1}\right)\\ &= \mathrm{Tr}\left(BB^{-1}D^{k-1}\right)\\ &= \mathrm{Tr}\left(D^{k-1}\right)\end{aligned}$$

Une récurrence immédiate nous donne donc $\mathrm{Tr}\left(D^k\right) = \mathrm{Tr}\left(I_n\right) = n$, et donc :

$$\begin{aligned}\mathrm{Tr}\left((D-I_n)^k\right) &= \mathrm{Tr}\left(\sum_{j=0}^{k} C_k^j (-1)^j D^{k-j}\right)\\ &= n\sum_{j=0}^{k} C_k^j (-1)^j\\ &= n(1-1)^k\\ &= 0\end{aligned}$$

Le lemme 1.18 nous permet d'affirmer que $D - I_n$ est nilpotente.

Montrons que toute matrice de G est diagonalisable. Soit $N < \infty$ l'exposant de G. Alors le polynôme $X^N - I_n$ annule toutes les matrices de G. Comme il est scindé à racines simples, les matrices de G sont diagonalisables.

La matrice D précédente est donc diagonalisable, et donc $D - I_n$ aussi. Comme elle est aussi nilpotente, elle est nulle, et donc $AB^{-1} - I_n = 0$, d'où $A = B$.

Donc f est injective. ◊

Étant donnée la définition de f, on peut affirmer que $\mathrm{Im}(f) \subseteq T^m$, où $T = \{\mathrm{Tr}(A) \mid A \in G\}$.

On a vu que toute matrice de G est annulée par $X^N - I_n$, donc les valeurs propres des matrices de G sont toutes des racines N-ième de l'unité. Donc l'ensemble des traces possibles T est fini.

f est donc une application injective de G dans un ensemble fini, et donc G est nécessairement fini.

□

Remarque – Il y a plusieurs façon de démontrer le lemme 1.18 dans le bouquin. Cependant, celle qui utilise les polynômes symétriques élémentaires et formules de Newton demande de maîtriser celles-ci, et la récurrence sur la dimension est moche (comme toute récurrence sur la dimension).

1.7 Théorème de Cauchy-Lipschitz

Schwartz (Analyse 2)

Remarques :

- Louis Augustin, baron Cauchy. Français. 1789-1857. Meurt d'un rhume.
- Rudolf Lipschitz. Allemand, 1832-1903.

Dans la suite, on considère le problème de Cauchy :

$$\begin{cases} y' = L(t,y) \\ y(t_0) = y_0 \end{cases} \tag{E}$$

où L est continue de $|a,b| \times \mathbb{R}^n$, $|a,b|$ intervalle (ouvert, fermé ou semi-ouvert) de $\mathbb{R}$.

Théorème 1.20 : de Cauchy-Lipschitz
On suppose la fonction L continue et globalement k-lipschitzienne en sa seconde variable.

Alors l'équation (E) admet une unique solution f sur $|a,b|$.

Démonstration. Partons d'une fonction f_0, et construisons une suite de fonction $(f_n)_n$ par récurrence :

$$f_{n+1}(t) = y_0 + \int_{t_0}^{t} L(\xi, f_n(t))\mathrm{d}\xi. \tag{1}$$

Lemme 1.21
La suite (f_n) vérifie, pour tout n :

$$\begin{cases} \|f_{n+1}(t) - f_n(t)\| \leqslant C\frac{k^n|t-t_0|^n}{n!} \\ \text{où } C \leqslant \sup_{t_0\leqslant\xi\leqslant t} \|f_1(\xi) - f_0(\xi)\| \end{cases}$$

Démonstration. Le résultat est évident pour $n = 0$.

Supposons-le vrai pour $n \geqslant 0$.

Alors :

$$f_{n+2} - f_{n+1} = \int_{t_0}^{t} (L(\xi, f_{n+1}(t)) - L(\xi, f_n(t)))\,\mathrm{d}\xi,$$

d'où

$$\begin{aligned} \|f_{n+2} - f_{n+1}\| &\leqslant \int_{t_0}^{t} \|L(\xi, f_{n+1}(t)) - L(\xi, f_n(t))\|\,\mathrm{d}\xi \\ &\leqslant k\int_{t_0}^{t} \|f_{n+1}(t) - f_n(t)\| \qquad \text{car } L \text{ lipschitzienne} \\ &\leqslant \frac{Ck^{n+1}}{n!}\int_{t_0}^{t} |\xi - t_0|^n \mathrm{d}\xi \qquad \text{par hypothèse de récurrence} \\ &\leqslant C\frac{k^{n+1}|t-t_0|^{n+1}}{(n+1)!} \end{aligned}$$

◊

Ce lemme nous permet d'affirmer que la suite (f_n) est de Cauchy uniformément sur tout compact, et donc converge uniformément sur tout compact, vers une fonction f.

La continuité uniforme nous autorise à passer à la limite dans la définition des f_n, d'où :

$$f(t) = y_0 + \int_{t_0}^{t} L(\xi, f(\xi))\mathrm{d}\xi,$$

ce qui est équivalent à (E).

On peut montrer le lemme suivant par récurrence sur n :

Lemme 1.22
Soient f_0 et g_0 deux fonctions, et (f_n) et (g_n) les suites définies par la relation (1) à partir des données initiales f_0 et g_0 respectivement.

Alors pour tout n :

$$\|f_n(t) - g_n(t)\| \leqslant C\frac{k^n|t-t_0|^n}{n!}.$$

Ce lemme nous montre que la limite f est indépendant du choix de f_0.

Soit $\hat{f}$ une autre solution de (E). En considérant la suite définie par $f_0 = \hat{f}$ et la relation (1), on obtient $f_n = \hat{f}$, et donc $\hat{f} = f$.

D'où l'unicité. □

Exemple – Si L n'est pas lipschitzienne, on perd l'unicité : on peut par exemple prendre $L(t, y) = 3y^{2/3}$.

L n'est pas lipschitzienne au voisinage de 0, et si $c_1 < 0 < c_2$, la fonction définie par

$$y(t) = \begin{cases} (t-c_1)^3 & \text{pour } t \leqslant c_1 \\ 0 & \text{pour } c_1 < t < c_2 \\ (t-c_2)^3 & \text{pour } c_2 \leqslant t \end{cases}$$

est solution de $y' = L(t, y)$ avec $y(0) = 0$.

Remarque – Cette version de la démonstration ne réutilise pas le théorème du point fixe de Picard. En fait, on peut adapter pour l'utiliser (magouille, ô magouille) pour la leçon en question, en montrant directement :

Si $\Phi : g \mapsto y_0 + \int_{t_0}^{\cdot} L(\xi, g(\xi))\mathrm{d}\xi$, alors Φ vérifie pout tout p :

$$\|\Phi^p(g_1)(t) - \Phi^p(g_2)(t)\| \leqslant \frac{k^p|t-t_0|^p}{p!}\|g_1 - g_2\|_\infty.$$

Donc pour p assez grand, Φ^p est contractante et donc admet un unique point fixe. (en fait, c'est là que c'est pénible si a ou b est ∞ : il faut construire la suite et la faire converger sur tout compact).

Remarque – On montre ici le théorème avec L *globalement* lipschitzienne. On peut mentionner que le cas *localement* lipschitzien se fait de la même manière, avec un cylindre de sécurité pour pas sortir des ensembles de définition des fonctions.

1.8 Une version faible du théorème de Dirichlet

Francinou *et al.* (Algèbre 1)

Remarque – Johann Peter Gustav Lejeune Dirichlet. Allemand. 1805-1859.

On cherche à montrer le théorème suivant :

Théorème 1.23 : de la progression arithmétique de Dirichlet, version faible
Pour tout $n \geqslant 1$, il existe une infinité de nombres premiers de la forme $\lambda n + 1$, $\lambda \in \mathbb{N}$.

On note maintenant $\Phi_1 = X - 1$ et pour tout $n \geqslant 2$,

$$\Phi_n = \prod_{\substack{k=1 \\ k \wedge n = 1}}^{n} \left(X - e^{\frac{2ik\pi}{n}} \right).$$

Lemme 1.24
Les Φ_n sont à coefficients entiers.

Démonstration. Commençons par montrer que

$$X^n - 1 = \prod_{d|n} \Phi_d.$$

On sait que U_n l'ensemble des racines n-ièmes de l'unité est réunion disjointe des P_d (racines d-ièmes primitives de l'unité) pour $d|n$, d'où

$$\begin{aligned} X^n - 1 &= \prod_{\xi \in U_n} (X - \xi) \\ &= \prod_{d|n} \left(\prod_{\xi \in P_d} (X - \xi) \right) \\ &= \prod_{d|n} \Phi_d \end{aligned}$$

On peut maintenant montrer le lemme par récurrence sur n :
$n = 1$: Par définition, c'est vrai.
$n \geqslant 2$: Supposons le lemme vrai pour tout $k < n$.
Alors Φ_n est le quotient de la division de $A = X^n - 1$ par $B = \prod_{d|n,\ d \neq n} \Phi_d$. Par hypothèse de récurrence, B est à coefficients entiers, et unitaire par définition. Or on a :

Lemme 1.25
Si $A, B \in \mathbb{Z}[X]$, avec B non nul unitaire, alors le quotient et le reste de la division de A par B dans $\mathbb{C}$ sont à coefficients entiers.

Donc Φ_n est à coefficients entiers.

$\Diamond$

Soit $a \in \mathbb{Z}$. Montrons que tout nombre premier p qui divise $\Phi_n(a)$ mais aucun des $\Phi_d(a)$ pour d diviseur strict de n est de la forme $\lambda n + 1$.

Par hypothèse, p divise donc aussi $a^n - 1$, et donc, dans $\left(\mathbb{Z}/p\mathbb{Z}\right)^*$, l'ordre de $\overline{a}$ divise n.

Soit d un diviseur strict de n. On a alors dans $\mathbb{Z}/p\mathbb{Z}$

$$\overline{a}^d - 1 = \prod_{d'|d} \overline{\Phi_{d'}(a)}.$$

Or par hypothèse, aucun des $\overline{\Phi_{d'}(a)}$ n'est nul et donc $\overline{a}^d \neq 1$.

Donc l'ordre de $\overline{a}$ est n, qui doit diviser $p-1$ par théorème de Lagrange. Donc p est bien congru à 1 modulo n.

Supposons qu'il n'existe qu'un nombre fini de nombres premiers congrus à 1 modulo n, soient $p_1, \ldots, p_q$.

Si on trouve a et p comme précédemment, on saura que p est congru à 1 modulo n. Afin d'éviter les cas $p = p_i$, on remplace n par $N = np_1 \cdots p_q$. En effet

$$p \equiv 1 \mod N \implies p \neq p_i \text{ et } p \equiv 1 \mod n.$$

Notons $B = \prod_{d|n,\ d\neq n} \Phi_d$. On cherche $a \in \mathbb{Z}$ et p premier tels que $p \mid \Phi_N(a)$ mais $p \nmid B(a)$.

Comme B et ϕ_N n'ont pas de racines communes dans $\mathbb{C}[X]$, ils sont premiers entre eux dans $\mathbb{C}[X]$, et donc dans $\mathbb{Q}[X]$ (ils sont à coefficients rationnels).

Il existe donc, par théorème de Bézout, U et V dans $\mathbb{Q}[X]$ tels que

$$U\Phi_N + VB = 1.$$

On peut trouver a tel que $U' = aU$ et $V' = aV$ soient à coefficients entiers , et comme $\Phi_N \neq 0, -1, 1$, on peut choisir a tel que $\Phi_N(a) \neq 0, -1, 1$. On a donc

$$a = U'(a)\Phi_N(a) + V'(a)B(a).$$

Soit p divisant $\Phi_N(a)$. Alors p divise $a^N - 1$, et donc a est premier avec p ($\overline{a}^N = 1$, et donc $\overline{a}$ inversible dans $\mathbb{Z}/p\mathbb{Z}$).

Donc par la relation précédente, p ne divise pas $B(a)$.

D'où le résultat voulu.

1.9 Caractérisation du dual de $\mathscr{M}_n(\mathbb{K})$

Francinou *et al.* (Algèbre 1)

On cherche ici à déterminer les formes linéaires de $\mathscr{M}_n(\mathbb{K})$.

Théorème 1.26
L'application

$$f : \begin{array}{ccc} \mathscr{M}_n(\mathbb{K}) & \longrightarrow & \mathscr{M}_n(\mathbb{K})^* \\ A & \longmapsto & f_A \ : \ X \mapsto \mathrm{Tr}(AX) \end{array}$$

réalise un isomorphisme entre $\mathscr{M}_n(\mathbb{K})$ et son dual.

Démonstration. On note $(E_{i,j})_{i,j}$ la base canonique de $\mathscr{M}_n(\mathbb{K})$.

f est clairement linéaire, et les espaces sont de même dimension finie. Montrons donc l'injectivité de f.

Soit A telle que $f_A = 0$. On a alors, pour tous i_0, j_0 :

$$\begin{aligned} \mathrm{Tr}(AE_{i_0,j_0}) &= \mathrm{Tr}\left(\sum_{i,j} a_{i,j}E_{i,j}E_{i_0,j_0}\right) \\ &= \sum_{i=1}^{n} a_{i,i_0}\,\mathrm{Tr}(E_{i,j_0}) \text{ car } E_{i,j}E_{k,l} = \delta_{j,k}E_{i,l} \\ &= a_{j_0,i_0} \\ &= 0 \text{ par hypothèse} \end{aligned}$$

Finalement, $A = 0$. □

Essayons maintenant, parmi toutes ces formes linéaires, de caractériser la plus connue, la *trace*.

Théorème 1.27
Soit $g \in \mathscr{M}_n(\mathbb{K})^*$ vérifiant $g(XY) = g(YX)$ pour toutes matrices X et Y. Alors

$$\exists \lambda \in \mathbb{K},\ \forall X \in \mathscr{M}_n(\mathbb{K}),\ g(X) = \lambda\,\mathrm{Tr}(X).$$

Démonstration. D'après le théorème précédent, il existe donc une matrice A telle que $g(X) = \mathrm{Tr}(AX)$.

L'hypothèse nous donne donc

$$\mathrm{Tr}(AXY) = \mathrm{Tr}(AYX).$$

Or les propriétés de la trace nous permettent d'écrire :

$$\mathrm{Tr}(AYX) = \mathrm{Tr}(XAY).$$

Finalement, on a

$$\mathrm{Tr}((AX - XA)Y) = 0,$$

et ce pour tout matrice Y.

En réutilisant l'isomorphisme précédent, on a donc $AX = XA$, et comme il est connu que le centre de $\mathscr{L}(E)$ est l'ensemble des homothéties, A en est donc une. □

Il est maintenant temps d'utiliser la correspondance forme linéaire $\leftrightarrow$ hyperplan.

Théorème 1.28
Si $n \geqslant 2$, alors tout hyperplan de $\mathscr{M}_n(\mathbb{K})$ rencontre $GL_n(\mathbb{K})$.

Démonstration. Soit donc H un hyperplan de $\mathscr{M}_n(\mathbb{K})$, et soit φ la forme linéaire associée. Il existe donc une matrice A telle que pour toute matrice X, on ait $\varphi(X) = \mathrm{Tr}(AX)$.

On cherche donc une matrice inversible, telle que $\mathrm{Tr}(AX)$ soit nulle.

Pour simplifier le problème, notons r le rang de A. A est donc équivalente à J_r : $PAQ = J_r$, où P et Q sont inversibles.

On a donc, pour toute matrice X,

$$\mathrm{Tr}(AX) = \mathrm{Tr}(PJ_rQX) = \mathrm{Tr}(J_rQXP).$$

Si on trouve Y inversible telle que $\mathrm{Tr}(J_rY)$ soit de trace nulle, on a gagné (on pose $X = Q^{-1}YP^{-1}$ qui reste à la fois dans $GL_n(\mathbb{K})$ et dans l'hyperplan H (on vérifie $\mathrm{Tr}(AX) = 0$)).

Pour cela, on peut par exemple poser

$$Y = \begin{bmatrix} 0 & 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \ddots & & & 0 \\ 0 & 1 & 0 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 & 0 \end{bmatrix}$$

et vérifier que J_rY a sa diagonale nulle, donc sa trace aussi. □

1.10 Théorème des extrema liés

Gourdon (Analyse)

Remarque – Ce théorème est dû à Lagrange.

Théorème 1.29
Soient $f, g_1, \dots, g_r$: $U \subseteq \mathbb{R}^n \to \mathbb{R}$ des fonctions de classe $\mathscr{C}^1$, U ouvert de $\mathbb{R}^n$.

Soit $\Gamma = \{x \in U \mid g_1(x) = \cdots = g_r(x) = 0\}$.

Si $f|_\Gamma$ admet un extremum relatif en $a \in \Gamma$, et si les formes linéaires $\mathrm{d}g_1(a), \dots, \mathrm{d}g_r(a)$ sont linéairements indépendantes, alors il existe des réels $\lambda_1, \dots, \lambda_r$ (les multiplicateurs de Lagrange) tels que :

$$\mathrm{d}f(a) = \sum_{i=1}^{r} \lambda_i \mathrm{d}g_i(a).$$

Démonstration. Remarques :

- On a nécessairement $r \leqslant n$: sinon, $(\mathrm{d}g_i(a))$ ne pourrait pas être libre ($\dim(\mathbb{R}^n)^* = n$).
- Si $r = n$, alors $(\mathrm{d}g_i(a))$ est une base, et donc le théorème est évident. On suppose donc $r < n$.

Posons $s = n - r$, et identifions $\mathbb{R}^n = \mathbb{R}^s \times \mathbb{R}^n$. On écrit donc $a = (\alpha, \beta)$.

Comme la famille $(\mathrm{d}g_i(a))$ est libre, la matrice

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_1}{\partial x_s}(a) & \frac{\partial g_1}{\partial y_1}(a) & \cdots & \frac{\partial g_1}{\partial y_r}(a) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_r}{\partial x_1}(a) & \cdots & \frac{\partial g_r}{\partial x_s}(a) & \frac{\partial g_r}{\partial y_1}(a) & \cdots & \frac{\partial g_r}{\partial y_r}(a) \end{bmatrix}$$

est de rang r : on peut donc en extraire une sous-matrice $r \times r$ inversible. Quitte à renommer les variables, on peut supposer que c'est

$$\begin{bmatrix} \frac{\partial g_1}{\partial y_1}(a) & \cdots & \frac{\partial g_1}{\partial y_r}(a) \\ \vdots & & \vdots \\ \frac{\partial g_r}{\partial y_1}(a) & \cdots & \frac{\partial g_r}{\partial y_r}(a) \end{bmatrix}$$

Par théorème des fonctions implicites, on peut trouver
– U' voisinage ouvert de α dans $\mathbb{R}^s$;
– Ω voisinage ouvert de a dans $\mathbb{R}^n$;
– $\varphi = (\varphi_1, \dots, \varphi_r)$: $U' \to \mathbb{R}^r$ de classe $\mathscr{C}^1$ telle que

$$(g(x, y) = 0,\ x \in U',\ (x, y) \in \Omega) \Leftrightarrow (y = \varphi(x)).$$

En clair, les éléments de Γ au voisinage de a s'écrivent $(x, \varphi(x))$.

Soit h : $x \mapsto f(x, \varphi(x))$. La fonction h admet donc un extremum local en $x = a$, et donc, pour tout $i \in [\![1, s]\!]$:

$$0 = \frac{\partial h}{\partial x_i}(a) = \frac{\partial f}{\partial x_i}(a) + \sum_{j=1}^{r} \frac{\partial \varphi_j}{\partial x_i}(\alpha) \frac{\partial f}{\partial y_j}(a).$$

De même, on peut écrire les dérivées partielles de $g(x, \varphi(x))$ par rapport aux x_i : pour tout $k \in [\![1, r]\!]$, pour tout $i \in [\![1, s]\!]$:

$$0 = \frac{\partial g_k}{\partial x_i}(a) + \sum_{j=1}^{r} \frac{\partial \varphi_j}{\partial x_i}(\alpha) \frac{\partial g_k}{\partial y_j}(a).$$

Finalement, les s premiers vecteurs colonnes de la matrice

$$M = \begin{bmatrix} \frac{\partial f}{\partial x_1}(a) & \cdots & \frac{\partial f}{\partial x_s}(a) & \frac{\partial f}{\partial y_1}(a) & \cdots & \frac{\partial f}{\partial y_r}(a) \\ \frac{\partial g_1}{\partial x_1}(a) & \cdots & \frac{\partial g_1}{\partial x_s}(a) & \frac{\partial g_1}{\partial y_1}(a) & \cdots & \frac{\partial g_1}{\partial y_r}(a) \\ \vdots & & \vdots & \vdots & & \vdots \\ \frac{\partial g_r}{\partial x_1}(a) & \cdots & \frac{\partial g_r}{\partial x_s}(a) & \frac{\partial g_r}{\partial y_1}(a) & \cdots & \frac{\partial g_r}{\partial y_r}(a) \end{bmatrix}$$

sont combinaison linéaire des r derniers, et donc $\mathrm{Rk}(M) \leqslant r$. Donc, les $r + 1$ vecteurs lignes de la matrice forment une famille liée, d'où l'existence de coefficients μ_i non tous nuls tels que :

$$\mu_0 \mathrm{d}f(a) + \sum_{i=1}^{r} \mu_i \mathrm{d}g_i(a) = 0.$$

Comme $(\mathrm{d}g_i(a))$, est libre, il est impossible que $\mu_0 = 0$, et donc on a le résultat. □

1.11 Intégrale de Fresnel

Gourdon (Analyse)

Remarque – Augustin Jean Fresnel. Français. 1788-1827. Meurt de la tuberculose.

On pose dans la suite :

$$F(t) = \iint_{[0,t]^2} e^{i(x^2+y^2)} \mathrm{d}x\mathrm{d}y$$
$$f(t) = \int_0^t e^{ix^2} \mathrm{d}x$$
$$I(T) = \frac{1}{T} F(t) \mathrm{d}t$$

On cherche à montrer :

Théorème 1.30
L'intégrale de Fresnel vaut :

$$\int_0^\infty e^{ix^2} \mathrm{d}x = \frac{\sqrt{\pi}}{2} e^{i\pi/4}.$$

Démonstration. Il est assez clair que $F(t) = f(t)^2$. Calculons-la d'une autre manière.
$[0,t]^2$ est symétrique par rapport à la première bissectrice, et donc, en posant $\Delta_t = \{(x,y) \in \mathbb{R}^2 \mid 0 \leqslant y \leqslant x \leqslant t\}$:

$$F(t) = 2 \iint_{\Delta_t} e^{i(x^2+y^2)} \mathrm{d}x\mathrm{d}y.$$

La vision de $x^2 + y^2$ donne très envie de passer en polaire. Δ_t devient alors

$$K_t = \{(r,\theta) \in \mathbb{R}^+ \times [0, \frac{\pi}{4}] \mid 0 \leqslant r\cos\theta \leqslant t\}.$$

On a donc, par changement de variables :

$$F(t) = \iint_{K_t} e^{ir^2} r \mathrm{d}r\mathrm{d}\theta.$$

Par théorème de Fubini (on est sur un compact) :

$$\begin{aligned} F(t) &= 2\int_0^{\pi/4} \left(\int_0^{t/\cos\theta} e^{ir^2} r \mathrm{d}r \right) \mathrm{d}\theta \\ &= \int_0^{\pi/4} \frac{1}{i} \left(\exp\left(i\frac{t^2}{\cos^2\theta} \right) - 1 \right) \\ &= \frac{i\pi}{4} - i\int_0^{\pi/4} \exp\left(i\frac{t^2}{\cos^2\theta} \right) \mathrm{d}\theta \end{aligned}$$

On a donc, en utilisant à nouveau le théorème de Fubini :

$$I(t) = \frac{i\pi}{4} - \frac{i}{T} \int_0^{\pi/4} \cos(\theta) f\left(\frac{T}{\cos\theta} \right).$$

Lemme 1.31
$\varphi = \int_0^\infty e^{ix^2}\mathrm{d}x$ *est semi-convergente.*

Démonstration. On pose $u = x^2$. Alors :

$$\varphi = \frac{1}{2}\int_0^\infty e^{iu}u^{-1/2}\mathrm{d}u$$

qui est semi-convergente (faire une intégration par parties). ◊

f est donc bornée sur $\mathbb{R}^+$, et donc I converge vers $\frac{i\pi}{4}$ en $+\infty$.

Par ailleurs, on a

$$I(T) = \frac{1}{T}\int_0^T f(t)^2\mathrm{d}t,$$

, et $f(t)^2$ converge vers φ^2 en $+\infty$, donc, par théorème de Césaro, I converge vers φ^2 en $+\infty$.

Finalement, $\varphi = \pm\frac{\sqrt{\pi}}{4}e^{i\pi/4}$. Il ne nous reste qu'à trouver le signe. Pour cela, regardons

$$s = \mathrm{Im}(\varphi) = \frac{1}{2}\int_0^\infty \sin(x)x^{-1/2}\mathrm{d}x.$$

On peut écrire

$$\begin{aligned} s &= \sum \int_{2n\pi}^{2(n+1)\pi} \frac{\sin u}{2\sqrt{u}}\mathrm{d}u \\ &= \sum \int_{2n\pi}^{2(n+1)\pi} \frac{\sin u}{2}\left(\frac{1}{\sqrt{u}} - \frac{1}{\sqrt{u+\pi}}\right)\mathrm{d}y \end{aligned}$$

et chaque terme de la somme est positif.

D'où le résultat.

□

1.12 Théorème de Gauß pour les polygones réguliers constructibles

Carrega (n.d.)

Remarque – Carl Friedrich Gauß. Allemand. 1777-1855.

Prérequis : Théorème de Wantzel.

On rappelle que le polygone régulier à n côtés est dit *constructible* si l'angle de mesure $\frac{2\pi}{n}$ est constructible, *i.e* si $e^{\frac{2\pi}{n}}$ est constructible.

On notera dans la suite $P(n)$ "le" polygone régulier à n côtés (on le suppose inscrit dans le cercle unité, avec un sommet coïncidant avec 1).

Théorème 1.32 : de Gauß, 1801
Les polygones réguliers constructibles sont ceux donc le nombre de côtés n est de la forme 2^α où $\alpha > 1$, ou de la forme $2^\alpha p_1 \cdots p_r$, $\alpha, i \in \mathbb{N}$, et les p_i sont des nombres premiers de Fermat distincts.

Démonstration. On commence par montrer qu'on peut se ramener à des nombres premiers.

Lemme 1.33
Si m et n sont premiers entre eux, alors $P(mn)$ est constructible si et seulement si $P(n)$ et $P(m)$ sont constructibles.

Démonstration. Le sens direct est évident :

$$\frac{2\pi}{n} = m\frac{2\pi}{mn} \text{ et } \frac{2\pi}{m} = n\frac{2\pi}{mn},$$

et il est facile de construit le multiple d'un angle constructible.

Pour le sens réciproque, on utilise l'identité de Bezout :

$$\exists \lambda, \mu \in \mathbb{Z},\ \lambda n + \mu m = 1.$$

On en déduit

$$\frac{2\pi}{mn} = \lambda\frac{2\pi}{n} + \mu\frac{2\pi}{m},$$

et on sait construire les produits et sommes d'angles constructibles. ◊

Par récurrence, on en déduit

Lemme 1.34
Si $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$, alors $P(n)$ est constructible si et seulement si *$P(p_1^{\alpha_1}), \ldots, P(p_k^{\alpha_k})$ sont constructibles.*

Il nous reste maintenant à caractériser les p_i et α_i qui conviennent.

Lemme 1.35
Soit $\alpha \in \mathbb{N}^*$. Alors :

(i) $P(2^\alpha)$ est constructible.

(ii) Si p est un nombre premier impair, alors $P(p^\alpha)$ est constructible si et seulement si $\alpha = 1$ *et p est un nombre de Fermat (*i.e $p = 1 + 2^{(2^\beta)}$*).*

Démonstration. On se fixe $\alpha \in \mathbb{N}^*$.

(i) On sait construire la bissectrice d'un angle, et donc ce point est évident.

(ii) Soit p premier impair. On pose $q = p^\alpha$.

Supposons que $P(q)$ est constructible. Alors par définition, $e^{\frac{2\pi}{q}}$ est constructible, et donc par théorème de Wantzel :

$$\exists m \in \mathbb{N},\ \left[\mathbb{Q}\left(e^{\frac{2\pi}{q}}\right) : \mathbb{Q}\right] = 2^m.$$

On appelle $\omega = e^{\frac{2\pi}{q}}$. ω est une racine q-ième de l'unité, et donc son polynôme minimal est le q-ième polynôme cyclotomique Φ_q, qui est de degré $p^{\alpha-1}(p-1)$, et donc

$$[\mathbb{Q}(\omega) : \mathbb{Q}] = p^{\alpha-1}(p-1).$$

On en déduit

$$p^{\alpha-1}(p-1) = 2^m,$$

et comme p est premier impair, on en déduit donc $\alpha = 1$ et $p = 2^m$.

Il nous reste à montrer que $m+1$ est nécessairement une puissance de 2. On écrit $m+1 = \lambda 2^\beta$, avec λ entier non nul impair et β entier.

On a donc

$$p = 1 + \left(2^{(2^\beta)}\right)^\lambda.$$

Comme λ est impair, $X+1$ divise $X^\lambda+1$, et donc $1+2^{(2^\beta)}$ divise p. Comme p est premier, on a le résultat.

Montrons maintenant le sens réciproque : soit $p = 1 + 2^n$ un nombre premier de Fermat.

On pose $\omega = e^{2i\pi/p}$, et $K = \mathbb{Q}(\omega)$. p étant premier, le degré de l'extension K sur $\mathbb{Q}$ est $p-1$. Une base de K est

$$\{1, \omega, \omega^2, \ldots, \omega^{p-2}\}.$$

Soit G le groupe des automorphismes de K (ils laissent $\mathbb{Q}$ invariants). Si $g \in G$, alors, g est entièrement déterminé par $g(\omega)$. Comme $g(\omega)$ reste une racine p-ième de l'unité, différente de 1, et réciproquement, toutes ces applications sont des automorphismes de K.

G est donc cyclique, d'ordre $p-1 = 2^n$. On peut ainsi se fixer un générateur de G, soit g.

Posons pour $0 \leqslant i \leqslant n$

$$K_i := \left\{z \in K \mid g^{2^i}(z) = z\right\}.$$

On a donc une tour d'extensions :

$$K_0 \subseteq K_1 \subseteq \cdots \subseteq K_n = K.$$

Il est assez clair que $\mathbb{Q} \subseteq K_0$, et si $z \in K_0$, alors on peut l'écrire sous la forme

$$z = \lambda_0\omega + \lambda_1 g(\omega) + \cdots + \lambda_{p-2} g^{p-2}(\omega),$$

et en applicant g, on obtient $z = -\lambda_0 \in \mathbb{Q}$.

Donc $K_0 = \mathbb{Q}$.

Montrons que toutes les inclusions sont strictes, par exemple $K_O \subset K_1$. On pose

$$z = \omega + g^2(\omega) + \cdots + g^{2^n-2}(\omega).$$

Alors $g^2(z) = z$ car $g^{2^n}(\omega) = \omega$, mais $g(z) \neq z$ par unicité de l'écriture de z dans la base $\{g^h(\omega) \mid 0 \leqslant h \leqslant p-2\}$.

On a donc une tour d'extensions strictes

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_n = K.$$

Comme $[K : \mathbb{Q}] = 2^n$, on obtient

$$[K_n : K_{n-1}] \cdots [K_1 : K_0] = 2^n,$$

et comme chaque facteur est différent de 1, on a nécessairement pour tout i :

$$[K_{i+1} : K_i] = 2.$$

Par théorème de Wantzel, ω est donc bien constructible.

$\Diamond$

□

1.13 Méthode du gradient à pas optimal

Ramis *et al.* (n.d.)

Le but de cette méthode est de déterminer une valeur approchée de la solution d'une équation linéaire du type $Ax = b$.

On note que la solution de ce système est aussi l'unique minimum de la fonction quadratique $f(x) = \frac{1}{2}(Ax|x) - (b|x)$.

On utilise l'algorithme suivant, avec :
– A une matrice $N \times N$ symétrique définie positive ;
– b un vecteur colonne de taille N ;
– TOL un seuil de tolérance.
Alors, on fait :

(i) on choisit $x_0 \in \mathbb{R}^N$, et on pose $r_0 = b - Ax_0$. On pose $n = 0$.

(ii) tant que $\|r_n\| \geqslant \|r_0\| TOL$ on prend
- $\alpha_{n+1} = \dfrac{\|r_n\|^2}{(Ar_n|r_n)}$
- x_{n+1}
- $r_{n+1} = b - Ax_{n+1}$
- $n \leftarrow n + 1$

Remarque – En fait, r_n est la *direction de plus grande pente*, et α_n est le *pas optimal*, c'est-à-dire vérifiant :

$$f(x_n + \alpha_{n+1} r_n) = \min_{\alpha > 0} f(x_n + \alpha r_n).$$

On a alors :

Théorème 1.36
L'algorithme du gradient à pas optimal converge vers la solution $\overline{x}$ de $Ax = b$, et pour tout n :

$$\|x_n - \overline{x}\| \leqslant \left(\frac{\mathrm{Cond}(A) - 1}{\mathrm{Cond}(A) + 1}\right)^n \sqrt{\mathrm{Cond}(A)} \|x_0 - \overline{x}\|.$$

Démonstration. Commençons par noter que

$$\begin{aligned} f(x) &= \frac{1}{2}(Ax|x) - (b|x) \\ &= \frac{1}{2}(A(x - \overline{x})|x - \overline{x}) - \frac{1}{2}(A\overline{x}|\overline{x}) \end{aligned}$$

On pose donc pour tout u de $\mathbb{R}^N$ $\|u\|_A = \sqrt{(Au|u)}$ qui définit une norme sur $\mathbb{R}^N$. Avec cette notation, on a :

$$f(x) = \frac{1}{2}\|x - \overline{x}\|_A^2 - \frac{1}{2}(A\overline{x}|\overline{x}).$$

Tentons de majorer le rapport $\dfrac{\|x_{n+1} - \overline{x}\|_A}{\|x_n - \overline{x}\|_A}$...

Par définition de l'algorithme, on a

$$\begin{aligned}\|x_{n+1}-\overline{x}\|_A^2 &= \|x_n+\alpha_{n+1}r_n-\overline{x}\|_A^2\\ &= \|(A(x_n-\overline{x})+\alpha_{n+1}Ar_n|(x_n-\overline{x})+\alpha_{n+1}r_n)\\ &= \|x_n-\overline{x}\|_A^2+\alpha_{n+1}(Ar_n|x_n-\overline{x})+\alpha_{n+1}(A(x_n-\overline{x})|r_n)+\alpha_{n+1}^2\|r_n\|_A^2\\ &= \|x_n-\overline{x}\|_A^2-2\frac{\|r_n\|^2}{\|r_n\|_A^2}\|r_n\|^2+\frac{\|r_n\|^4}{\|r_n\|_A^2}\\ &= \|x_n-\overline{x}\|_A^2-\frac{\|r_n\|^4}{\|r_n\|_A^2}\end{aligned}$$

D'où

$$\frac{\|x_{n+1}-\overline{x}\|_A^2}{\|x_n-\overline{x}\|_A^2}=1-\sigma,$$

où

$$\sigma=\frac{\|r_n\|^4}{\|r_n\|_{A^{-1}}^2\|r_n\|_A^2}.$$

Remarque – Ici, on a montré la convergence.

On se trouve dans l'hypothèse de l'inégalité de Kantorovitch, et donc :

$$\begin{aligned}1-\sigma &\leqslant 1-4\frac{\lambda_m\lambda_M}{(\lambda_m+\lambda_M)^2}\\ &= \frac{\lambda_m^2+\lambda_M^2+2\lambda_m\lambda_M-4\lambda_m\lambda_M}{(\lambda_m+\lambda_M)^2}\\ &= \left(\frac{\lambda_m-\lambda_M}{\lambda_m+\lambda_M}\right)^2\\ &= \left(\frac{\mathrm{Cond}(A)-1}{\mathrm{Cond}(A)+1}\right)^2\end{aligned}$$

Une récurrence immédiate nous donne :

$$\|x_n-\overline{x}\|_A\leqslant\left(\frac{\mathrm{Cond}(A)-1}{\mathrm{Cond}(A)+1}\right)^n\|x_0-\overline{x}\|_A.$$

Pour obtenir l'inégalité souhaitée, on utilise l'équivalence des normes :

$$\lambda_m\|x\|^2\leqslant\|x\|_A^2\leqslant\lambda_M\|x\|^2.$$

On a donc

$$\|x_0-\overline{x}\|_A\leqslant\sqrt{\mathrm{Cond}(A)}\|x_0-\overline{x}\|$$

et

$$\|x_n-\overline{x}\|_A\geqslant\sqrt{\mathrm{Cond}(A)}^{-1}\|x_n-\overline{x}\|.$$

□

1.14 Ellipsoïde de John-Lœwner

Francinou *et al.* (Algèbre 3)

Remarques :

- Fritz John. Allemand. 1910-1994.
- ?

Théorème 1.37 : de John-Lœwner
Soit K un compact d'intérieur non vide de $\mathbb{R}^n$. Il existe un unique ellipsoïde centré en O de volume minimal contenant K.

Démonstration. On munit $\mathbb{R}^n$ de sa structure euclidienne usuelle. Un ellipsoïde plein centré en O a une équation du type $q(x) \leqslant 1$ où $q \in Q^{++}$ (*i.e* l'ensemble des formes quadratiques définies positives).

On note $\mathscr{E}_q = \{x \in \mathbb{R}^n \mid q(x) \leqslant 1\}$ l'ellipsoide associé à $q \in Q^{++}$.

Lemme 1.38
Le volume de $\mathscr{E}_q$ est

$$V_q = \frac{V_0}{\sqrt{\det(q)}},$$

où V_0 est le volume de la boule unité pour la norme euclidienne canonique.

Démonstration. Dans une certaine base orthogonale, q est de la forme

$$q(x) = \sum_{i=1}^{n} a_i x_i^2.$$

q est définie positive, donc tous les a_i sont strictement positifs.

On a donc

$$V_q = \int_{\sum a_i x_i^2 \leqslant 1} \mathrm{d}x_1 \dots \mathrm{d}x_n.$$

On effectue le changement de variables $t_i = \sqrt{a_i} x_i$ pour $i \in [\![1, n]\!]$. C'est un $\mathscr{C}^1$-difféomorphisme de jacobien $\frac{1}{\sqrt{a_1 \dots a_n}}$, et on obtient donc

$$V_q = \int_{\sum t_i^2 \leqslant 1} \frac{\mathrm{d}t_1 \dots \mathrm{d}t_n}{\sqrt{a_1 \dots a_n}}.$$

De plus, le déterminant de q est indépendant de la base choisie, et vaut $a_1 \dots a_n$. ◊

Par le lemme précédent, le problème consiste maintenant à montrer qu'il existe une unique forme quadratique définie positive $q \in Q^{++}$ telle que $D(q)$ soit maximal, et telle que $\forall x \in K,\ q(x) \leqslant 1$.

On munit l'ensemble des formes quadratiques Q de la norme

$$N(q) = \sup_{||x|| \leqslant 1} |q(x)|.$$

On pose $\mathscr{A} = \{q \in Q^+ \mid \forall x \in K, q(x) \leqslant 1\}$. On remarque que si $q \in \mathscr{A}$ est définie positive, alors $K \subset \mathscr{E}_q$. On va donc chercher à maximiser $D(q)$ sur ce domaine.

Lemme 1.39
$\mathscr{A}$ est un compact non vide de Q.

Démonstration. $\mathscr{A}$ est non vide : K est compact, donc il est borné : soit M tel que $\forall x \in K,\ ||x|| \leqslant M$.
En posant $q_1(x) = \frac{||x||^2}{M^2}$, on a $q_1 \in Q^{++} \subset Q^+$, et pour tout x de K, $q_1(x) \leqslant 1$.
Donc $q_1 \in \mathscr{A}$, et donc $\mathscr{A}$ est non vide.
$\mathscr{A}$ est fermé : On remarque que la convergence dans Q implique la convergence faible ; en effet, si q_n converge dans Q vers q, alors

$$\forall x \in \mathbb{R}^n,\ |q_n(x) - q(x)| \leqslant N(q_n - q)||x||^2.$$

Par suite : $q(x) = \lim q_n(x) \geqslant 0$ et $q(x) = \lim q_n(x) \leqslant 1$.
Donc $q \in \mathscr{A}$.
$\mathscr{A}$ est borné : K est d'intérieur non vide, donc K contient une boule centrée en a de rayon r.
Soit $q \in \mathscr{A}$. Si $||x|| \leqslant r$, alors $a + x \in K$, et donc $q(a + x) \leqslant 1$. Alors

$$\begin{aligned}\sqrt{q(x)} &= \sqrt{q(x + a - a)} \\ &\leqslant \sqrt{q(x + a)} + \sqrt{q(-a)} \qquad \text{par Minkowski} \\ &\leqslant 1 + 1 \\ &\leqslant 2\end{aligned}$$

Donc $q(x) \leqslant 4$.
Si $||x|| \leqslant 1$, on a

$$\begin{aligned}|q(x)| &= q(x) \\ &= \frac{1}{r^2} q(rx) \\ &\leqslant \frac{4}{r^2}\end{aligned}$$

Donc $N(q) \leqslant \frac{4}{r^2}$, et donc $\mathscr{A}$ est borné.

◊

Ainsi, $\det : \begin{array}{ccc} \mathscr{A} & \longrightarrow & \mathbb{R}^+ \\ q & \longmapsto & \det(q) \end{array}$ est continue, et donc elle est majorée, et atteint son maximum sur $\mathscr{A}$, en q_0.

On a vu que $q_1 \in \mathscr{A}$, et $q_1 \in Q^{++}$, donc $\det(q_0) \geqslant \det(q_1) > 0$.

Donc $q_0 \in Q^{++}$.

Il reste maintenant à montrer l'unicité.

Lemme 1.40
$\mathscr{A}$ est convexe.

Démonstration. Si q et q' sont dans $\mathscr{A}$, et $\lambda \in [0, 1]$:
– $\forall x \in \mathbb{R}^n,\ (\lambda q + (1 - \lambda q')(x) = \lambda q(x) + (1 - \lambda) q'(x) \geqslant 0$.
– $\forall x \in K,\ \lambda q + (1 - \lambda) q')(x) \leqslant \lambda + 1 - \lambda \leqslant 1$.

◊

Supposons qu'il existe $q \in \mathscr{A}$ tel que $D(q) = D(q_0)$ et $q \neq q_0$.

Notons S et S_0 les matrices de q et q_0 dans la base canonique de $\mathbb{R}^n$.

Par convexité de $\mathscr{A}$, $\frac{1}{2}(q+q_0) \in \mathscr{A}$, et on a :

$$\begin{aligned}\det(\frac{1}{2}(q+q_0) &= \det(\frac{1}{2}(S+S_0) \\ &> (\det S)^{1/2}(\det S)^{1/2} \qquad \text{par lemme 1.41} \\ &\geqslant \det S_0 \\ &\geqslant \det q_0\end{aligned}$$

ce qui contredit la maximalité de $\det q_0$.

Lemme 1.41
Soient A et B dans $\mathscr{S}_n^{++}(\mathbb{R})$, $\alpha, \beta \in R^+$ tels que $\alpha+\beta=1$. Alors

$$\det(\alpha A + \beta B) \geqslant (\det A)^\alpha (\det B)^\beta.$$

De plus, si $A \neq B$, alors l'inégalité est stricte.

Démonstration. Par théorème de pseudo-réduction simultanée, il existe une matrice P inversible et $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ avec les $\lambda_i \in \mathbb{R}^+$ telles que $A = {}^tPP$ et $B = {}^tPDP$.

Donc

$$\begin{aligned}(\det A)^\alpha (\det B)^\beta &= \det P^2 (\det D)^\beta \text{ et} \\ \det(\alpha A + \beta B) &= \det P^2 \det(\alpha I_n + \beta D).\end{aligned}$$

On veut montrer que $\det(\alpha I_n + \beta D) \geqslant (\det D)^\beta$, ce qui équivaut à $\prod_{i=1}^n (\alpha + \beta\lambda_i) \geqslant \left(\prod_{i=1}^n \lambda_i\right)^\beta$, ou encore à $\sum_{i=1}^n \log(\alpha+\beta\lambda_i) \geqslant \beta \sum_{i=1}^n \log \lambda_i$.

Or pour tout i de 1 à n :

$$\begin{aligned}\log(\alpha+\beta\lambda_i) &\geqslant \alpha \log(1) + \beta \log(\lambda_i) \qquad \text{par concavité du log} \\ &\geqslant \beta \log(\lambda_i)\end{aligned}$$

On a le résultat en sommant sur i.

Si $A \neq B$, un des λ_i est différent de 1.

Donc, si $\alpha \in (0,1)$, la stricte concavité de log donne une inégalité stricte.

◊

□

Référence : Oraux X-ENS, Francinou, Gianella, Nicolas.

Leçons : 123, 131, 137, 203, 219, 229

1.15 Théorème de stabilité de Lyapunov

Rouvière (n.d.)

Remarque – Aleksandr Mikhailovich Lyapunov. Russe. 1857-1918. Se tire une balle dans la tête suite au décès de sa femme atteinte de tuberculose.

Dans ce développement, on cherche à faire passer la propriété d'attractivité de l'origine d'un système linéarisé au système perturbé. Plus précisemment :

Théorème 1.42
On considère le système différentiel

$$\begin{cases} y' = f(y) \\ y(0) = x \end{cases} \tag{$\star$}$$

où $f : \mathbb{R}^n \to \mathbb{R}^n$ *est une fonction de classe* $\mathscr{C}^1$*, vérifiant* $f(0) = 0$*.*

Si toutes les valeurs propres de la matrice $\mathrm{d}f(0)$ *sont de parties réelles négatives, alors l'origine est un point d'équilibre attractif du système,* i.e *si* x *est assez voisin de* 0*, alors la solution* $y(t)$ *tend exponentiellement vers* 0 *lorsque* t *tend vers l'infini.*

Démonstration. On considère le système linéarisé au voisinage de 0 :

$$\begin{cases} z' = Az \\ z(0) = x \end{cases} \tag{$\star\star$}$$

Toutes les valeurs propres de A sont donc de partie réelle négative.

Lemme 1.43
Notons $\lambda_1, \dots, \lambda_k$ *les valeurs propres distinctes de* A*. Alors il existe un polynôme* P *tel que, pour tous* $t \in \mathbb{R}$ *et* $x \in \mathbb{R}^n$

$$\|e^{tA}x\| \leqslant P(|t|) \left(\sum_{j=1}^{k} e^{t\Re(\lambda_j)} \right) \|x\|.$$

Démonstration. On a, par théorème des noyaux, une décomposition unique de $x \in \mathbb{R}^n$ en $x = x_1 + \cdots + x_k$, où $x_j \in E_j := \ker(A - \lambda_j I)^{m_j}$.

Chaque sous-espace caractéristique E_j est stable par A, et donc pour chaque j :

$$e^{tA}x_j = e^{t\lambda_j} e^{t(A-\lambda_j I)} x_j \qquad = e^{t\lambda_j} \left(\sum_{p=0}^{m_j - 1} \frac{t^p}{p!} (A - \lambda_j I)^p \right) x_j$$

et donc

$$\begin{aligned} \|e^{tA}x_j\| &\leqslant e^{t\Re(\lambda_j)} C_j (1 + |t|)^{m_j - 1} \|x_j\| \\ &\leqslant C e^{t\Re(\lambda_j)} (1 + |t|)^{n-1} \|x_j\| \end{aligned}$$

En sommant et par inégalité triangulaire, on obtient donc

$$\begin{aligned}\|e^{tA}x\| &\leqslant \sum_{j=1}^{k} \|e^{tA}x_j\| \\ &= C(1+|t|)^{n-1}\left(\sum_{j=1}^{k} e^{t\Re(\lambda_j)}\right)\max_j \|x_j\|\end{aligned}$$

L'équivalence des normes permet donc de conclure. ◊

La solution du système $(\star\star)$ est $z(t) = e^{tA}x$. Le lemme précédent nous dit tout de suite que

$$\|z(t)\| \leqslant Ce^{-at}\|x\|$$

pour un certain $a > 0$.

Posons

$$b(x,y) = \int_0^\infty (e^{tA}x \cdot e^{tA}y)\mathrm{d}t.$$

L'inégalité précédente nous donne la convergence absolue de l'intégrale, et $q(x) := b(x,x)$ est une forme quadratique définie positive.

Lemme 1.44
On a pour tout x :

$$(\operatorname{grad} q(x) \cdot Ax) = 2b(x,Ax) = -\|x\|^2.$$

Démonstration. Pour $x, y \in \mathbb{R}^n$ et $t \in \mathbb{R}$, on a

$$q(x+ty) = q(x) + 2tb(x,y) + t^2 q(y).$$

En dérivant en $t = 0$, on obtient

$$\mathrm{d}q(x)y = 2b(x,y).$$

On a donc

$$(\operatorname{grad} q(x) \cdot Ax) = 2b(x,Ax).$$

Remarquons que

$$2b(x,Ax) = \int_0^\infty 2(e^{tA}x \cdot e^{tA}Ax)\mathrm{d}t,$$

et que l'intégrande est la dérivée par rapport à t de $(e^{tA}x \cdot e^{tA}x)$.

On a donc, en utilisant la majoration de $e^{tA}x$, le résultat. ◊

On en déduit, en posant $r(y) = f(y) - Ay$:

$$\begin{aligned}q(y)' &= \mathrm{d}q(y)y' \\ &= 2b(y,y') \\ &= 2b(y,Ay) + 2b(y,r(y)) \\ &= -\|y\|^2 + 2b(y,r(y))\end{aligned}$$

On cherche à "éliminer" le terme contenant $r(y)$, qui est tout petit. Plus précisemment, on a par Cauchy-Swcharz (on utilise ici la norme $\sqrt{q}$) :

$$b(y, r(y)) \leqslant \sqrt{q(y)}\sqrt{q(r(y))}.$$

On a par définition, $r(y) = f(y) - f(0) - \mathrm{d}f(0)y$, et donc pour tout $\varepsilon > 0$, il existe un $\alpha > 0$ tel que

$$(q(y) \leqslant \alpha) \Longrightarrow \left(\sqrt{q(r(y))} \leqslant \varepsilon\sqrt{q(y)}\right).$$

Comme les normes $\|\cdot\|$ et $\sqrt{q}$ sont équivalentes, on a en prenant ε assez petit l'existence d'un $\beta > 0$ tel que

$$q(y)' \leqslant -\beta q(y).$$

Vérifions que si $q(x) < \alpha$, alors $q(y(t))$ reste inférieur à α aussi.

Sinon, il existe un plus petit temps $t_0 > 0$ tel que $q(y(t_0)) = \alpha$. Et l'inégalité précédente nous donne

$$q(y)'(t_0) \leqslant -\beta\alpha < 0,$$

et donc $q(y)(t) > \alpha$ pour t proche de t_0, $t < t_0$. D'où une contradiction.

Donc, si $q(x) < \alpha$, on a

$$q(y)' \leqslant -\beta q(y).$$

On résout l'inéquation :

$$\left(e^{\beta t}q(y)\right)' = e^{\beta t}\left(q(y)' + \beta q(y)\right) \leqslant 0,$$

ce qui entraîne, comme $y(0) = x$, pour tout $t \geqslant 0$:

$$q(y(t)) \leqslant e^{-\beta t}q(x).$$

D'où le résultat.

□

1.16 Primalité des nombres de Mersenne

Saux Picart & Rannou (n.d.)

Remarque – Marin Mersenne. Français. 1588-1648. Après avoir bu trop d'eau froide, il tombe malade, et meurt des saignées qu'on lui a faites.

On appelle *nombres de Mersenne* les

$$M_q = 2^q - 1$$

pour $q \in \mathbb{N}$

On a d'abord le lemme :

Lemme 1.45
Si M_q est un nombre premier, alors q est premier.

Démonstration. Si q n'est pas premier, $q = mn$, avec $m, n > 2$.

Et alors $M_q = 2^{mn} - 1$ qui est divisible par $2^n - 1$. □

On a une caractérisation :

Théorème 1.46
Pour tout nombre premier impair q :

$$M_q \text{ est premier} \iff \left(2+\sqrt{3}\right)^{2^{q-1}} \equiv -1 \mod M_q.$$

On remarque qu'il faut se placer dans un corps où 3 admet une racine carrée. Dans la suite, on explicitera : on prendra $\mathbb{F}_{M_q}$ ou une de ses extensions.

Démonstration du sens direct.

Lemme 1.47
Pour tout entier k non nul, M_{2k+1} est congru à 7 modulo 12.

Démonstration. Par récurrence :
$(k = 1)$: On a bien $2^{2\times 1+1} = 7$.
$(k \to k+1)$: On a modulo 12 :

$$\begin{aligned} 2^{2(k+1)+1} - 1 &\equiv 4 \times 2^{2k+1} - 1 \\ &\equiv \left(2^{2k+1} - 1\right) \times 4 + 3 \\ &\equiv 7 \times 4 + 3 \\ &\equiv 7 \end{aligned}$$

◊

Donc, pour tout q impair, $M_q \equiv 7 \mod 12$.

Montrons maintenant que 3 n'est pas résidu quadratique modulo M_q.

Pour cela, on montre le

Lemme 1.48
3 est résidu quadratique modulo un entier premier p si, et seulement si $p \equiv \pm 1 \mod 12$.

Démonstration. Par la loi de réciprocité quadratique, on a :

$$\left(\frac{p}{3}\right)\left(\frac{3}{p}\right) = (-1)^{\frac{p-1}{2}}.$$

Ainsi, par définition du symbole de Legendre :

$$3 \text{ résidu modulo } p \Leftrightarrow \left(\frac{p}{3}\right) = (-1)^{\frac{p-1}{2}}.$$

On remarque que le seul carré non nul de $\mathbb{F}_3$ est 1, et donc 3 est résidu quadratique modulo p *si, et seulement si* l'une des conditions est vérifiée :

(i) $p \equiv 1 \mod 3$ et $\frac{p-1}{2}$ est pair.

(ii) $p \equiv 2 \mod 3$ et $\frac{p-1}{2}$ est impair.

Dans le premier cas, p est congru à 1 modulo 3 et 4, et donc modulo 12.

Dans le second cas, p est congru à 2 modulo 3, et 3 modulo 4, et donc par théorème chinois, à -1 modulo 12. $\Diamond$

Comme M_q n'est congru ni à 1, ni à -1 modulo 12, 3 n'est pas résidu quadratique modulo M_q. $X^2 - 3$ est donc irréductible sur $\mathbb{F}_{M_q}$, et donc $\mathscr{A} = \mathbb{F}_{M_q}[X]/(X^2 - 3)$ est un corps, et on note la classe de X dans $\mathscr{A}$ $\sqrt{3}$.

On remarque de plus que $2^{q+1} \equiv 2 \mod M_q$, et donc 2 admet une racine carrée $\sqrt{2} := 2^{\frac{q+1}{2}}$.

On définit les quantités

$$\rho = \frac{1+\sqrt{3}}{\sqrt{2}} \text{ et } \overline{\rho} = \frac{1-\sqrt{3}}{\sqrt{2}}.$$

On montre facilement que $\rho^2 = 2 + \sqrt{3}$ et $\overline{\rho}\rho = -1$.

De plus, on remarque que comme $\sqrt{3}$ n'est pas résidu quadratique modulo M_q, par petit théorème de Fermat :

$$\left(\sqrt{3}\right)^{M_q} = 3^{\frac{M_q-1}{2}}\sqrt{3} = -\sqrt{3}.$$

Comme $\mathscr{A}$ est de caractéristique M_q, on a par morphisme de Frobénius :

$$\left(a + b\sqrt{3}\right)^{M_q} = a - b\sqrt{3}.$$

De même, on a $\rho^{M_q} = \overline{\rho}$ Comme $\mathscr{A}$ est de caractéristique M_q, on a par morphisme de Frobénius :

$$\left(a + b\sqrt{3}\right)^{M_q} = a - b\sqrt{3}.$$

De même, on a $\sqrt{2}^{M_q} = \sqrt{2}$, et donc $\rho^{M_q} = \overline{\rho}$.

On multiplie à gauche et à droite, et on obtient :

$$\left(2+\sqrt{3}\right)^{2^{q-1}} = \left(2+\sqrt{3}\right)^{\frac{M_q+1}{2}} = -1.$$

$\square$

Démonstration du sens réciproque. On note dans la suite $\mathbb{Z}_n$ l'anneau $\mathbb{Z}/n\mathbb{Z}$.

On note encore $\mathscr{A}$ une extension de $\mathbb{Z}_{M_q}$ contenant une racine de 3 : plus précisemment, si $\mathbb{Z}_{M_q}$ contient une racine de 3, on prend $\mathscr{A} = \mathbb{Z}_{M_q}$, et sinon on prend $\mathscr{A} = \mathbb{Z}_{M_q}[X]/(X^2 - 3)$.

On suppose M_q non premier, et on appelle p un de ses diviseurs premiers.

p est donc un diviseur de 0 dans $\mathscr{A}$, et a fortiori n'est pas inversible. Il est donc contenu dans un idéal maximal $\mathscr{M}$ de $\mathscr{A}$.

Alors $\mathscr{A}/\mathscr{M}$ est un corps, de caractéristique p (p non nul dans $\mathscr{M}$).

On appelle α (resp. β) la classe de $2+\sqrt{3}$ (resp. $2-\sqrt{3}$) dans $\mathscr{A}/\mathscr{M}$.

Notre hypothèse s'écrit donc $\alpha^{2^{q-1}} \equiv -1 \mod M_q$, et on en déduit que α est d'ordre 2^q dans $\mathscr{A}/\mathscr{M}$.

On pose maintenant $Q = (X-\alpha)(X-\beta) = X^2 - 4X + 1$. C'est un polynôme à coefficient dans le corps premier de $\mathscr{A}/\mathscr{M}$, $\mathbb{F}_p$.

Donc, comme α est racine de Q, α^p aussi, et donc $\alpha^p = \alpha$ ou $\alpha^p = \beta$.

Dans le premier cas, comme l'ordre de α est 2^q, 2^q divise $p-1$. Or p divise $M_q = 2^q - 1$, donc $p < 2^q$. D'où une contradiction.

Dans le second cas, $\alpha^p = \beta = \alpha^{-1} = \alpha^{M_q}$. On a alors $p \equiv 2^q - 1 \mod 2^q$, et ceci impose $p = M_q$. Encore une contradiction.

□

Remarque – On peut citer un corollaire direct de ce théorème :

Théorème 1.49 : Test de Lehmer-Lucas
On définit la suite $(L_n) \in \mathbb{Z}_{M_q}^{\mathbb{N}}$ *par*

$$L_0 = 4 \text{ et } L_{n+1} = L_n^2 - 2 \mod M_q.$$

Alors on a :

$$M_q \text{ premier} \iff L_{q-2} \equiv 0 \mod M_q.$$

Cet algorithme permet de calculer directement dans $\mathbb{Z}_{M_q}$ plutôt que dans une extension. Au final, il est de complexité $\mathscr{O}(q^3)$ (on peut accélerer un peu avec la transformée de Fourier discrète).

1.17 Théorème de Molien

Leichtman (n.d.)

Soit E un espace vectoriel de dimension finie n, et soit G un sous-groupe fini de $GL(E)$.

On considère l'action de G sur $A := \mathbb{C}[X_1, \dots, X_n]$ définie par

$$\forall g \in G,\ \forall P \in A,\ g \cdot A := P(g^*(X_1, \dots, X_n)),$$

et $\sigma\ :\ G \to \S(A)$ le morphisme associé.

On définit de plus A_k comme l'ensemble des polynômes de A homogènes de degré k, et A_k^G l'espace vectoriel des invariants sous l'action de G sur A_k, *i.e* :

$$A_k^G = \{P \in A_k \mid \forall g \in G,\ g \cdot P = P\}.$$

Lemme 1.50
σ définit bien une action, qui de plus est linéaire et à valeurs dans $\mathrm{Aut}(A)$*. De plus, pour tout $k \in \mathbb{N}$, pour tout $g \in G$, $\sigma(g)$ induit un automorphisme de A_k.*

Démonstration. Soient $g, h \in G$. Alors

$$\begin{aligned}\forall P \in A,\ (g \circ h) \cdot P &= P((g \circ h)^*(X_1, \dots, X_n)) \\ &= P(h^* \circ g^*(X_1, \dots, X_n)) \\ &= g \cdot h \cdot P\end{aligned}$$

Donc σ définit bien une action, qui est clairement linéaire, et comme tous les éléments de G sont inversibles, les $\sigma(G)$ aussi.

Soit $k \in \mathbb{N}$, soit $g \in G$. Il est clair que $\sigma(G)(A_k) \subseteq A_k$, et comme A_k est de dimension finie et que $\sigma(g)$ est injective, $\sigma(g)$ est bien un isomorphisme. □

On appelera donc g_k l'automorphisme de A_k induit par $\sigma(g)$.

On notera dans la suite $a_k = \dim A_k$ et $a_k(G) = \dim A_k^G$. On cherche à montrer le théorème :

Théorème 1.51 : de Molien
On a égalité des séries formelles :

$$\frac{1}{\mathrm{Card}(G)} \sum_{g \in G} \frac{1}{\det(I - gX)} = \sum_{k=0}^{\infty} a_k(G) X^k.$$

Démonstration. On commence par des lemmes :

Lemme 1.52
On a l'égalité des séries formelles :

$$\frac{1}{(1-X)^n} = \sum_{k=1}^{\infty} a_k X^k.$$

Démonstration. Une base de A_k est $\left\{X_1^{i_1} \times \cdots \times X_n^{i_n} \mid (i_1, \ldots, i_n) \in \mathbb{N}^n \text{ et } i_1 + \cdots + i_n = k\right\}$, et donc

$$a_k = \text{Card}\{(i_1, \ldots, i_n) \in \mathbb{N}^n \mid \sum i_j = k\}.$$

On sait que $\left(\sum_{p=0}^{\infty} X^p\right) = \left(\frac{1}{1-z}\right)^n$, et donc en calculant *explicitement* le produit de Cauchy n fois, on montre que a_k est le coefficient de z^k dans la série formelle $\left(\frac{1}{1-z}\right)^n$. ◊

Lemme 1.53
Pour tout g de G, on a l'égalité des séries formelles :

$$\frac{1}{\det(I - gX)} = \sum_{k=0}^{\infty} \text{Tr}(g_k) X^k.$$

Démonstration. On commence par remarquer que tous les éléments de G sont diagonalisables sur $\mathbb{C}$: $X^{\text{Card}\, G} - 1$ qui est scindé à racines simples annule tous les éléments de G.

Soit $g \in G$, qu'on diagonalise en ugu^{-1}. On a alors :

$$\sigma_{|A_k}(ugu^{-1}) = \sigma_{|A_k}(u)\sigma_{|A_k}(g)\sigma_{|A_k}(u^{-1})$$

et donc :

$$\text{Tr}(g_k) = \text{Tr}(\sigma_{|A_k}(ugu^{-1})).$$

On peut donc se ramener au cas où g est diagonale, $g = \text{diag}(\lambda_1, \ldots, \lambda_n)$.

On a donc :

$$\begin{aligned}
\frac{1}{\det(I - Xg)} &= \prod_{i=1}^{n} \frac{1}{1 - \lambda_i X} \\
&= \prod_{i=1}^{n} \left(\sum_{p=1}^{\infty} \lambda_i^p X^p\right) \\
&= \sum_{p=0}^{\infty} v_p X^p
\end{aligned}$$

où

$$v_p = \sum_{k_1 + \cdots + k_n = p} \lambda_1^{k_1} \cdots \lambda_n^{k_n}.$$

De plus,

$$g_p(X_1^{k_1} \cdots X_n^{k_n}) = (\lambda_1^{k_1} \cdots \lambda_n^{k_n}) X_1^{k_1} \cdots X_n^{k_n}.$$

On a donc $\text{Tr}(g_p) = v_p$, et donc on peut identifier les coefficients dans les séries formelles précédentes. ◊

Lemme 1.54
Soit $\varphi : G \to GL(V)$ un morphisme. On note

$$V^G = \bigcap_{g \in G} \ker(id - \varphi(g)i).$$

Alors

$$\dim V^G = \frac{1}{\operatorname{Card} G} \sum_{g \in G} \operatorname{Tr}(\varphi(g)).$$

Démonstration. On définit l'endomorphisme $p_G = \frac{1}{\operatorname{Card} G} \sum_{g \in G} \operatorname{Tr}(\varphi(g))$. On veut montrer que $p_G(V) = V^G$.

On a pour tous $v \in V$ et $h \in G$:

$$\begin{aligned} \varphi(h)(p_G(v)) &= \frac{1}{\operatorname{Card} G} \sum_{g \in G} \varphi(h)\varphi(g)(v) \\ &= p_G(v) \end{aligned}$$

On en déduit la première inclusion $p_G(V) \subseteq V^G$.

Réciproquement, si $v \in V^G$, on a $p_G(v) = v$, d'où l'inclusion réciproque.

Si $v \in V$, alors $p_G(v) \in V^G$, et par le calcul précédent, $p_G \circ p_G(x) = x$, et donc p_G est un projecteur, d'image V^G.

On a alors $\dim V^G = \operatorname{Tr}(p_G)$, d'où le résultat cherché par linéarité de la trace. ◊

On peut enfin conclure :

On pose

$$\varphi : \begin{array}{ccc} G & \longrightarrow & GL(A_k) \\ g & \longmapsto & g_k \end{array}$$

Par le lemme précédent, on a

$$\dim A_k^G = a_k(G) = \frac{1}{\operatorname{Card} G} \sum_{g \in G} \operatorname{Tr}(g_k).$$

En sommant pour $g \in G$, on a le résultat. □

1.18 Lemme de Morse

Rouvière (n.d.)

Remarque – Marston Morse. Américain. 1892-1967.

Théorème 1.55 : Lemme de Morse
Soit $f : U \to \mathbb{R}$ une fonction de classe $\mathscr{C}^3$ sur un ouvert U de $\mathbb{R}^n$ contenant l'origine. On suppose que 0 est un point critique quadratique non dégénéré de f, i.e $\mathrm{d}f(0) = 0$ *et $\mathrm{d}^2 f(0)$ non dégénérée de signature $(p, n-p)$.*

Alors il existe un $\mathscr{C}^1$-difféomorphisme $x \mapsto u = \varphi(x)$ entre deux voisinages de l'origine dans $\mathbb{R}^n$ tel que $\varphi(0) = 0$ et

$$f(x) = f(0) + u_1^2 + \cdots + u_p^2 - u_{p+1}^2 - \cdots - u_n^2.$$

Démonstration. On écrit la formule de Taylor avec reste intégral au premier ordre pour f, pour $x \in V_0$ voisinage de 0 :

$$f(x) = f(0) + {}^t Q(x)x,$$

où

$$Q(x) = \int_0^1 (1-t)\mathrm{d}^2 f(tx)\mathrm{d}t.$$

Par théorème de dérivation sous intégrale (f est supposée de classe $\mathscr{C}^3$), Q est de classe $\mathscr{C}^1$.

Le théorème de réduction $\mathscr{C}^1$ des formes quadratiques nous donne l'existence d'une fonction $M : V_0 \to GL_n(\mathbb{R})$ de classe $\mathscr{C}^1$ telle que

$$Q(x) = {}^t M(x)Q(0)M(x).$$

On en déduit, en notant $y = M(x)x$:

$$f(x) = f(0) + {}^t yQ(0)y.$$

Or la signature de $Q(0)$ est la même que celle de $\mathrm{d}^2 f(0)$: $(p, n-p)$. Donc par changement de base :

$$\exists A \in GL_n(\mathbb{R}),\ Q(0) = {}^t A\,\mathrm{diag}(I_p, I_{n-p})A.$$

Finalement,

$$f(x) = f(0) + {}^t(Ay)Q(0)(Ay).$$

Pour conclure, en posant $u = \varphi(x) = AM(x)x$, on a le résultat souhaité :

$$f(x) = f(0) + u_1^2 + \cdots + u_p^2 - u_{p+1}^2 - \cdots - u_n^2.$$

Montrons que φ est un $\mathscr{C}^1$-difféomorphisme.

Soit $\psi : x \to M(x)x$.

On a alors $\mathrm{d}\psi(0) = M(0)$, avec $M(0)$ inversible. Donc ψ un est $\mathscr{C}^1$-difféomorphisme au voisinage de 0. Comme A est aussi inversible, φ l'est aussi. □

Théorème 1.56 : Réduction $\mathscr{C}^1$ des formes quadratiques
On note S l'espace des matrices symétriques de taille $n \times n$. Pour $A_0 \in S$ inversible fixé, on pose

$$\alpha : \begin{array}{ccc} \mathscr{M}_n(\mathbb{R}) & \longrightarrow & S \\ M & \longmapsto & {}^t M A_0 M \end{array}.$$

Alors il existe un voisinage V de A_0 dans S, et une application $\begin{array}{ccc} V & \longrightarrow & GL_n(\mathbb{R}) \\ A & \longmapsto & M \end{array}$ *de classe $\mathscr{C}^1$ tel que $A = {}^tMA_0M$ pour tout A de V.*

Démonstration. L'application α est polynomiale donc de classe $\mathscr{C}^1$. Calculons sa différentielle :

$$\begin{aligned} \alpha(I+H) - \alpha(I) &= {}^tHA_0 + A_0H + {}^tHA_0H \\ &= {}^t(A_0H) + A_0H + \mathscr{O}(\|H\|^2) \end{aligned}$$

Donc

$$\mathrm{d}\alpha(I)H = {}^t(A_0H) + A_0H.$$

Le noyau de $\mathrm{d}\alpha(I)$ est donc $\{H \in \mathscr{M}_n(\mathbb{R}) \mid A_0H \text{ antisymétrique}\}$, de dimension $n(n-1)/2$. Donc l'image est de dimension $n(n+1)/2 = \dim S$, et donc $\mathrm{d}\alpha(I)$ est surjective.

On note maintenant α la restriction de α à $F = \{H \in \mathscr{M}_n(\mathbb{R}) \mid A_0H \in S\}$.

Alors α est toujours de classe $\mathscr{C}^1$, mais maintenant $\mathrm{d}\alpha(I)$ est inversible.

Par théorème d'inversion locale, on a un voisinage U de I dans F (quitte à restreindre U, on le choisit inclu dans $GL_n(\mathbb{R})$) avec α $\mathscr{C}^1$-difféomorphisme de U sur $V = \alpha(U)$.

Donc pour tout A de V, il existe un unique $M \in U$ tel que $A = \alpha(M) = {}^tMA_0M$. □

Remarque – Ce théorème nous dit que pour une forme quadratique non dégénérée, les formes quadratiques assez proche lui sont équivalentes.

1.19 Théorème de Müntz

Gourdon (Analyse)

Remarque – Herman Müntz. Allemand. 1884-1956.

Leçons :
– Déterminants
Prérequis :
– Distance à un sous-espace en fonction de déterminants de Gram
– Déterminants de Cauchy
– Théorème de Stone-Weierstraß

Théorème 1.57
Soit $(\mathscr{C}, ||\cdot||_2)$ l'espace des fonctions continues sur $[0,1]$, mnui de la norme de L^2. Soit $(\alpha_n)_{n\in\mathbb{N}}$ une suite strictement croissante de réels positifs.

Alors on a l'équivalence :

(i) $V := \text{Vect}\left((x \mapsto x^{\alpha_n})_{n\in\mathbb{N}}\right)$ *est dense dans* $\mathscr{C}$.

(ii) La série $\displaystyle\sum_{n\in\mathbb{N}} \frac{1}{\alpha_n}$ *diverge.*

Démonstration. Par souci de clarté, on notera simplement x^a la fonction $x \mapsto x^a$.

Pour $m \in \mathbb{R}^+$, on note $\Delta_N(m)$ la distance de x^m à $\text{Vect}\,(x^{\alpha_1}, \dots, x^{\alpha_N})$. On a donc, avec les déterminants de Gram :

$$\Delta_N(m)^2 = \frac{\text{Gram}\,(x^m, x^{\alpha_1}, \dots, x^{\alpha_N})}{\text{Gram}\,(x^{\alpha_1}, \dots, x^{\alpha_N})}.$$

Or, pour tous a et b, on a $\langle x^a, x^b\rangle = \frac{1}{a+b+1}$.

Donc :

$$\text{Gram}\,(x^m, x^{\alpha_1}, \dots, x^{\alpha_N}) = \begin{vmatrix} \frac{1}{2m+1} & \frac{1}{m+\alpha_1+1} & \cdots & \frac{1}{m+\alpha_N+1} \\ \frac{1}{m+\alpha_1+1} & \frac{1}{2\alpha_1+1} & \cdots & \frac{1}{\alpha_1+\alpha_N+1} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{m+\alpha_N+1} & \frac{1}{\alpha_N+\alpha_1+1} & \cdots & \frac{1}{2\alpha_N+1} \end{vmatrix}$$

On a donc, d'après le calcul du déterminant de Cauchy :

$$\text{Gram}\,(x^m, x^{\alpha_1}, \dots, x^{\alpha_N}) = \frac{\left(\prod_{i<j}(\alpha_i - \alpha_j)^2\right)\left(\prod_i(\alpha_i - m)^2\right)}{(2m+1)\left(\prod_{i,j}(\alpha_i + \alpha_j + 1)\right)\left(\prod_i(\alpha_i + m + 1)^2\right)}$$

et, de la meme façon :

$$\text{Gram}\,(x^{\alpha_1}, \dots, x^{\alpha_N}) = \frac{\prod_{i<j}(\alpha_i - \alpha_j)^2}{\prod_{i<j}(\alpha_i + \alpha_j + 1)}.$$

Finalement, on a l'expression de la distance cherchée :

$$\Delta_N(m) = \frac{1}{\sqrt{2m+1}} \prod_{i=1}^{N} \left| \frac{\alpha_i - m}{\alpha_i + m + 1} \right|.$$

Supposons que V est dense dans $\mathscr{C}$. Alors, pour tout $m \in \mathbb{R}^+$, on a $\Delta_N(m)$ tend vers 0, *i.e* $\sum_n u_n$ diverge, avec $u_n = \log\left(\frac{\alpha_n - m}{\alpha_n + m + 1}\right)$. Soit $m \in \mathbb{R}^+$.

Deux cas se présentent :
– Si $(\alpha_n)_n$ est majorée, $\sum \frac{1}{\alpha_n}$ diverge.
– Sinon, il existe un rang N_0 pour lequel $\alpha_n > m$ pour tout $n \geqslant N_0$. Alors

$$u_n \sim -\frac{2m+1}{\alpha_n}, \qquad (\star)$$

et comme $\sum_n u_n$ diverge, $\sum \frac{1}{\alpha_n}$ diverge aussi.
Réciproquement, supposons que $\sum \frac{1}{\alpha_n}$ diverge. En montrant que $x^k \in \overline{V}$ pour tout entier k, on aura le résultat par théorème d'approximation de Stone-Weierstraß. Soit donc $k \in \mathbb{N}$.

Si $\alpha_n \to \infty$, alors l'équivalent $(\star)$ suffit pour conclure.

Sinon :

$$\begin{aligned}
\prod_{i=1}^{N} \frac{|\alpha_i - m|}{\alpha_i + m + 1} &\leqslant \prod_{i=1}^{N} \frac{\alpha_i + m}{\alpha_i + m + 1} \\
&= \prod_{i=1}^{N} \left(1 - \frac{1}{\alpha_i + m + 1}\right) \\
&\leqslant \left(1 - \frac{1}{\sup_i \alpha_i + m + 1}\right)^n \\
&\to 0
\end{aligned}$$

□

Référence : Xavier Gourdon, Analyse, pp 286-287.

Remarque – Il y a deux erreurs dans le Gourdon :
– page 287, tout en haut : $\langle x^a, x^b \rangle = \frac{1}{a+b\boxed{+1}}$.
– page 287, dans l'expression du déterminant, il manque un carré :

$$\mathrm{Gram}\,(x^m, x^{\alpha_1}, \ldots, x^{\alpha_N}) = \frac{\left(\prod_{i<j}(\alpha_i - \alpha_j)^2\right)\left(\prod_i (\alpha_i - m)^{\boxed{2}}\right)}{(2m+1)\left(\prod_{i,j}(\alpha_i + \alpha_j + 1)\right)\left(\prod_i (\alpha_i + m + 1)^2\right)}$$

1.20 Méthode de Newton pour les polynômes

Chambert-Loir & Fermigier (n.d.)

Remarque – Sir Isaac Newton. Anglais. 1643-1727.

Soit $\xi_1 < \ldots < \xi_r$ des réels, et $m_1 \ldots, m_r$ des entiers non nuls. On considère le polynôme

$$P = \prod_{i=1}^{r} (X - \xi_i)^{m_i}.$$

On choisit un $x_0 > \xi_r$, et on considère la suite $(x_n)_n$ définie par

$$x_{n+1} = f(x_n),$$

où

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

Alors la suite (x_n) converge vers ξ_r, et on a une estimation de la vitesse de convergence :
– si $m_r = 1$, alors pour tout $c > 0$,

$$|x_n - \xi_r| = o(c^n).$$

– si $m_r > 1$, alors il existe $c > 0$ telle que

$$|x_n - \xi_r| \sim c \left(1 - \frac{1}{m_r}\right)^n.$$

Démonstration. On reconnait en P'/P la dérivée logarithmique de P, et donc on a

$$\frac{P'(x)}{P(x)} = \sum_{i=1}^{r} \frac{m_i}{x_n - \xi_i}.$$

La relation $x_{n+1} = f(x_n)$ se réécrit donc

$$x_{n+1} = x_n - \left(\sum_{i=1}^{r} \frac{m_i}{x_n - \xi_i} \right)^{-1}.$$

Donc, si $x_n > \xi_r$, alors $x_{n+1} < x_n$. Il ne nous reste donc qu'à montrer que $x_{n+1} > \xi_r$ pour avoir la décroissance de (x_n).

On peut prolonger f à $I := [\xi_r, \infty)$ (ξ_r est une racine de P' d'ordre 1 de moins que de P, et donc $P(\xi_r)/P'(\xi_r) = 0$) en posant $f(\xi_r) = \xi_r$.

f est dérivable sur I :

$$\begin{aligned} f'(x) &= 1 - 1 + \frac{P(x)P''(x)}{P'(x)^2} \\ &= \frac{PP''}{P'^2}(x) \end{aligned}$$

Comme les zéros de P sont dans $[\xi_1, \xi_r]$, il en est de même pour les zéros de P' et P'' (théorème de Gauß-Lucas), et donc (les coefficients sont positifs) f est strictement croissante sur I.

On a donc, pour $x_n > \xi_r$:

$$\xi_r = f(\xi_r) < f(x_n) = x_{n+1}.$$

Finalement, la suite (x_n) est décroissante et minorée, et donc converge, nécessairement vers un point fixe de f : ξ_r.

Intéressons-nous maintenant à la vitesse de convergence de la suite.

Commençons par calculer $f'(x)$ plus précisemment. On a vu

$$\frac{P'}{P}(x) = \sum_i \frac{m_i}{x - \xi_i},$$

et donc en dérivant :

$$\frac{PP'' - P'^2}{P^2}(x) = -\sum_i \frac{m_i}{x - \xi_i}^2.$$

Donc

$$f'(x) = \frac{PP''}{P'^2}(x) = 1 - \left(\sum_i \frac{m_i}{x - \xi_i}\right)^{-2} \left(\sum_i \frac{m_i}{(x - \xi_i)^2}\right).$$

On peut en déduire

$$\lim_{x \to \xi_r} f'(x) = 1 - \frac{1}{m_r}.$$

Regardons d'abord le cas $m_r = 1$. On a donc $f'(\xi_r) = 0$. Donc, par théorème de Taylor-Lagrange, il existe un y_n tel que $f(x_n) - f(\xi_r) = (x_n - \xi_r)f'(y_n)$.

On a donc $f'(y_n) \xrightarrow[n\to\infty]{} 0$, et donc, si $c > 0$, l'existence d'un rang à partir duquel $|f'(y_n)| < c$.

Donc $|x_{n+1} - \xi_r| \leqslant c|x_n - \xi_r|$, et une récurrence immédiate nous donne

$$|x_n - \xi_r| = \mathcal{O}(c^n).$$

Quitte à réécrire l'égalité avec un $c' < c$, on peut remplacer $\mathcal{O}$ par o.

Il nous reste le cas $m_r > 1$. On a donc $f'(\xi_r) = 1 - \frac{1}{m_r} \in (0, 1)$. On peut à nouveau appliquer la formule de Taylor-Lagrange :

$$\exists y_n \in (\xi_r, x_n),\ x_{n+1} - \xi_r = f'(y_n)(x_n - \xi_r).$$

En passant au log :

$$\log(x_{n+1} - \xi_r) - \log(x_n - \xi_r) = \log f'(y_n) \underset{n}{\to} \log f'(\xi_r).$$

Ainsi, par théorème de Cesáro, on a

$$\log(x_n - \xi_r) \sim n \log f'(\xi_r).$$

Cependant, il est interdit de prendre l'exponentielle pour conclure. Montrons donc que $\log(x_n - \xi_r) - n \log(\xi_r)$ converge.

On peut à nouveau appliquer la formule de Taylor-Lagrange, à l'ordre 2 :

$$\exists z_n \in (\xi_r, x_n),\ x_{n+1} - \xi_r = f'(\xi_r)(x_n - \xi_r) + \frac{f''(z_n)}{2}(x_n - \xi_r)^2.$$

On a en particulier

$$\varepsilon_n := \frac{x_{n+1} - \xi_r}{f'(\xi_r)(x_n - \xi_r)} - 1 = \mathscr{O}(x_n - \xi_r).$$

Comme on a vu $\log(x_n - \xi_r) \sim n \log(f'(\xi_r))$, donc il existe une constante $f'(\xi_r) < c < 1$ telle que $|x_n - \xi_r| = \mathscr{O}(c^n)$.

On a donc $\log(1+\varepsilon_n) = \mathscr{O}(c^n)$, et comme la série de terme général c^n converge, celle de terme général $\log(x_{n+1} - \xi_r) - \log(x_n - \xi_r) - \log f'(\xi_r)$ aussi. Par série téléscopique, c'est exactement dire que $\log(x_n - \xi_r) - n \log(f'(\xi_r))$ converge, vers un certain λ.

On conclut :

$$x_n - \xi_r \sim e^{\lambda} f'(\xi_r)^n.$$

□

1.21 Formule de Poisson ; calcul de $\sum n^{-2}$

Gourdon (Analyse)Zuily & Queffélec (n.d.)

Remarque – Siméon Denis Poisson. Français. 1781-1840.

Si $f \in L^1(\mathbb{R})$, on définira sa transformée de Fourier par :

$$\hat{f}(x) = \int_{\mathbb{R}} e^{-2i\pi xt} f(t)\mathrm{d}t.$$

On a alors :

Théorème 1.58 : Formule sommatoire de Poisson
Soit $F \in L^1(\mathbb{R}) \cap \mathscr{C}^0(\mathbb{R})$. On suppose que :

$$\exists M > 0,\ \exists \alpha > 1,\ \forall x \in \mathbb{R},\ |F(x)| \leqslant M(1+|x|)^{-\alpha}$$

et

$$\sum_{n=-\infty}^{\infty} |\hat{F}(n)| < \infty.$$

Alors on a :

$$\sum_{n=-\infty}^{\infty} F(n) = \sum_{n=-\infty}^{\infty} \hat{F}(n).$$

Démonstration. On définit la fonction f par

$$f(x) = \sum_{n=-\infty}^{\infty} F(x+n).$$

Lemme 1.59
La série définissant f est normalement convergente.

Démonstration. Soit $A > 0$, et soit $|x| \leqslant A$. Si $|n| \geqslant 2A$, alors

$$\begin{aligned} |x+n| &\geqslant |n| - |x| \\ &\geqslant |n| - A \\ &\geqslant \frac{|n|}{2} \end{aligned}$$

donc

$$|F(x+n)| \leqslant M\left(1 + \frac{|n|}{2}\right)^{-\alpha},$$

qui est le terme général d'une série convergente. ◊

F est continue par hypothèse, donc f l'est aussi, et f vérifie la relation :

$$\begin{aligned} f(x+1) &= \sum_{n\in\mathbb{Z}} F(x+n+1) \\ &= f(x) \end{aligned}$$

f est 1-périodique : on peut calculer ses coefficients de Fourier.

$$\begin{aligned}
c_m(f) &= \int_0^1 f(t)e^{-2i\pi mt}\mathrm{d}t \\
&= \int_0^1 \sum_{n\in\mathbb{Z}} F(t+n)e^{-2i\pi mt}\mathrm{d}t \\
&= \sum_{n\in\mathbb{Z}} \int_0^1 F(t+n)e^{-2i\pi mt}\mathrm{d}t \text{ par convergence normale} \\
&= \sum_{n\in\mathbb{Z}} \int_n^{n+1} F(t)e^{-2i\pi nt}\mathrm{d}t \text{ par changement de variable } t \mapsto t-n \\
&= \int_{-\infty}^{\infty} F(t)e^{-2i\pi nt} \\
&= \hat{F}(n)
\end{aligned}$$

L'hypothèse $\sum_{n\in\mathbb{Z}} |\hat{F}(n)| < \infty$ nous donne donc

$$\sum_{n\in\mathbb{Z}} |c_m(f)| < \infty.$$

f est donc développable en série de Fourier :

$$\begin{aligned}
f(x) &= \sum_{n\in\mathbb{Z}} c_m(f)e^{2i\pi mx} \\
&= \sum_{n\in\mathbb{Z}} \hat{F}(m)e^{2i\pi mx}
\end{aligned}$$

D'où

$$\sum_{n\in\mathbb{Z}} F(x+n) = \sum_{n\in\mathbb{Z}} \hat{F}(m)e^{2i\pi mx}.$$

En $x = 0$, on a le résultat. □

Proposition 1.60
Soit $a > 0$. On a :

$$\sum_{n\in\mathbb{Z}} \frac{1}{n^2+a^2} = \frac{\pi}{a}\coth(\pi a).$$

Démonstration. On pose $f(t) = e^{-2\pi a|t|}$.

On a alors

$$\begin{aligned}
\hat{f}(x) &= \int_{\mathbb{R}} e^{-2i\pi tx} f(t)\mathrm{d}t \\
&= \int_{\mathbb{R}} e^{-2\pi(tx+a|t|)}\mathrm{d}t \\
&= \frac{a}{\pi(a^2+x^2)}
\end{aligned}$$

Notre fonction f vérifie les hypothèses du théorème, et donc :

$$\begin{aligned}\frac{a}{\pi}\sum_{n\in\mathbb{Z}}\frac{1}{a^2+n^2} &= \sum_{n\in\mathbb{Z}} e^{-2\pi a|n|}\\ &= 2\sum_{n\in\mathbb{N}} e^{-2\pi an} - 1\\ &= \frac{1+e^{-2\pi a}}{1-e^{-2\pi a}}\\ &= \coth(a)\end{aligned}$$

□

Corollaire 1.61
On retrouve

$$\sum_{n>0}\frac{1}{n^2} = \frac{\pi^2}{6}.$$

Démonstration. On a pour tout n :

$$\frac{1}{a^2+n^2} \geqslant \frac{1}{n^2},$$

et donc par théorème de convergence dominée, on a :

$$\lim_{a\to 0}\sum_{n>0}\frac{1}{a^2+n^2} = \sum_{n>0}\frac{1}{n^2}$$

Or on a

$$\sum_{n\in\mathbb{Z}}\frac{1}{a^2+n^2} = \frac{1}{a^2} + 2\sum_{n>0}\frac{1}{a^2+n^2}.$$

On a donc

$$\sum_{n>0}\frac{1}{a^2+n^2} = \frac{\pi}{2a}\coth(\pi a) - \frac{1}{2a^2}$$

On a le développement limité de coth :

$$\coth(x) = \frac{1}{x} + \frac{x}{3} + o(x).$$

D'où

$$\begin{aligned}\sum_{n>0}\frac{1}{a^2+n^2} &= \frac{\pi}{2a}\frac{1}{\pi a} + \frac{\pi}{2a}\frac{\pi a}{3} - \frac{1}{2a^2} + o(1)\\ &= \frac{\pi^2}{6} + o(1)\end{aligned}$$

Quand $a \to 0$, on a le résultat. □

1.22 Probabilité que deux nombres soient premiers entre eux

Francinou *et al.* (Algèbre 1)

Prérequis :
– fonction de Möbius
– formule du crible

On rappelle la définition de la fonction de Möbius :

Définition 1.62
La fonction de Möbius est la fonction $\mu : \mathbb{N}^* \to \mathbb{Z}$ *définie par :*
– $\mu(1) = 1$
– $\mu(p_1 \cdots p_r) = (-1)^r$ *si les* p_i *sont des nombres premiers distincts*
– $\mu(n) = 0$ *sinon (si* n *est divisible par le carré d'un nombre premier).*

On rappelle de plus la formule du crible :

Proposition 1.63 : Formule du crible
Soient $E_1, \dots, E_k$ *des ensembles finis. Alors :*

$$\operatorname{Card}\left(\bigcup_{i=1}^{k} E_i\right) = \sum_{\emptyset \neq I \subseteq [\![1,k]\!]} (-1)^{1+\operatorname{Card} I} \operatorname{Card}\left(\bigcap_{i \in I} E_i\right).$$

On note aussi r_n la probabilité que deux entiers choisis au hasard dans $[\![1, n]\!]$ soient premiers entre eux.

On a alors :

Théorème 1.64
On a $\lim_{n\to\infty} r_n = \dfrac{6}{\pi^2}$.

Démonstration. Appellons, pour tout $n \geqslant 1$,

$$A_n = \{(a, b) \in [\![1, n]\!]^2 \mid a \wedge b = 1\}.$$

On a donc $r_n = \frac{\operatorname{Card} A_n}{n^2}$.

On note $p_1, \dots, p_k$ la liste des nombres premiers inférieurs à n, et $U_i = \{(a, b) \in [\![1, n]\!]^2 \mid p_i | a \text{ et } p_i | b\}$.

On a alors l'identité :

$$A_n = \left(\bigcup_{i=1}^{k} U_i\right)^c.$$

Lemme 1.65
On a

$$\operatorname{Card} A_n = \sum_{d=1}^{n} \mu(d) E\left(\frac{n}{d}\right)^2.$$

Démonstration. Soit $I \subseteq [\![1,k]\!]$ non vide. Alors le cardinal de l'intersection $\bigcap_{i\in I} U_i$ est exactement égal au nombre de couples de multiples strictement positifs de $\prod_{i\in I} p_i$ inférieurs à n :

$$\operatorname{Card}\left(\bigcap_{i\in I} U_i\right) = E\left(\frac{n}{\prod_{i\in I} p_i}\right)^2.$$

On peut donc utiliser la formule du crible :

$$\begin{aligned}\operatorname{Card}\left(\bigcup_{i=1}^{k} U_i\right) &= \sum_{\emptyset\neq I\subseteq[\![1,k]\!]} (-1)^{1+\operatorname{Card} I} \operatorname{Card}\left(\bigcap_{i\in I} U_i\right)\\ &= \sum_{\emptyset\neq I\subseteq[\![1,k]\!]} (-1)^{1+\operatorname{Card} I} E\left(\frac{n}{\prod_{i\in I} p_i}\right)^2\end{aligned}$$

Donc on a :

$$\begin{aligned}\operatorname{Card} A_n &= n^2 - \sum_{\emptyset\neq I\subseteq[\![1,k]\!]} (-1)^{1+\operatorname{Card} I} E\left(\frac{n}{\prod_{i\in I} p_i}\right)^2\\ &= \sum_{d=1}^{n} \mu(d) E\left(\frac{n}{d}\right)^2\end{aligned}$$

En effet : on veut ne garder dans la somme que les produits de nombres premiers distincts, d'où le $\mu(d)$ pour "enlever" les autres, et n^2 correspond à $d = 1$.

D'où le résultat. ◊

On peut en déduire immédiatement que

$$r_n = \frac{1}{n^2}\sum_{d=1}^{n} \mu(d) E\left(\frac{n}{d}\right)^2.$$

L'intuition nous indique ici de remplacer le terme $\frac{1}{n^2}E\left(\frac{n}{d}\right)^2$ par son équivalent $\frac{1}{d^2}$ (on commence à voir apparaître $\zeta(2)\ldots$).

On estime la différence entre les deux sommes :

$$\left|r_n - \sum_{d=1}^{n} \frac{\mu(d)}{d^2}\right| = \left|\sum_{d=1}^{n} \mu(d)\left(\frac{1}{n^2}E\left(\frac{n}{d}\right)^2 - \frac{1}{d^2}\right)\right|.$$

On remarque que $E(n/d) > n/d - 1$, et donc on a :

$$\frac{1}{n^2} - \frac{2}{dn} < \frac{1}{n^2}E\left(\frac{n}{d}\right)^2 - \frac{1}{d^2} \leqslant 0.$$

Donc, par inégalité triangulaire :

$$\begin{aligned}\left|r_n - \sum_{d=1}^{n} \frac{\mu(d)}{d^2}\right| &\leqslant \sum_{d=1}^{n} \left(\frac{2}{dn} + \frac{1}{n^2}\right) \\ &\leqslant \frac{2}{n}\sum_{d=1}^{n} \frac{1}{d} + \frac{1}{n} \\ &= \mathscr{O}\left(\frac{\log n}{n}\right)\end{aligned}$$

Ainsi, on a l'identité :

$$\lim_{n\to\infty} r_n = \sum_{d=1}^{\infty} \frac{\mu(d)}{d^2}.$$

Calculons cette somme. Pour cela, calculons à tout hasard

$$\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} \times \sum_{n=1}^{\infty} \frac{1}{n^2}.$$

Les deux séries convergent absolument, et donc, par théorème de Fubini :

$$\begin{aligned}\left(\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2}\right)\left(\sum_{n=1}^{\infty} \frac{1}{n^2}\right) &= \sum_{d,n\geqslant 1} \frac{\mu(d)}{(dn)^2} \\ &= \sum_{d\geqslant 1,\ d|k} \frac{\mu(d)}{k^2} \\ &= \sum_{k\geqslant 1}\sum_{d|k} \frac{\mu(d)}{p^2} \\ &= \sum_{k\geqslant 1} \frac{1}{k^2}\left(\sum_{d|k} \mu(d)\right)\end{aligned}$$

Il ne nous manque ici plus qu'un petit lemme sur la fonction de Möbius :

Lemme 1.66
On a

$$\sum_{d|n} \mu(d) = \begin{cases} 1 \text{ si } n = 1 \\ 0 \text{ si } n \geqslant 2. \end{cases}$$

Démonstration. Notons $S(n)$ la somme considérée. Il est clair que $S(1) = 1$. Soit donc $n \geqslant 2$, et considérons sa décomposition en facteurs premiers :

$$n = \prod_{i=1}^{k} p_i^{\alpha_i},$$

où les p_i sont des nombres premiers distincts, et α_i des entiers strictements positifs.

Les seuls termes non nuls dans $S(n)$ sont des produits de p_i, sans multiplicités. Pour chaque j, n a exactement

C_k^j diviseurs de cette forme, produits de i p_j. D'où

$$\begin{aligned} S(n) &= \sum_{j=0}^{k} C_k^j (-1)^j \\ &= (1-1)^k \\ &= 0 \end{aligned}$$

$\Diamond$

On conclut avec le lemme, et la bien connue valeur de $\zeta(2)$. $\square$

1.23 Action du groupe modulaire sur le demi-plan de Poincaré

Alessandri (n.d.)

On appelle *demi-plan de Poincaré* l'ensemble $P = \{z \in \mathbb{C} \mid \Im(z) > 0\}$.

On fait agir $SL_2(\mathbb{Z})$ sur P par l'action :

$$\forall A \in SL_2(\mathbb{Z}),\ \forall z \in P,\ A * z = \frac{az+b}{cz+d}.$$

On appelle en particulier S et T les matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \text{ et } T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

Finalement, on appelle *groupe modulaire* le groupe

$$PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/_{\{\pm I\}}.$$

On fait agir ce groupe sur P de la même manière que $SL_2(\mathbb{Z})$. On remarquera que cette action est fidèle (seul le neutre fixe tous les points).

On cherche à montrer

Théorème 1.67
Le groupe $SL_2(\mathbb{Z})$ est engendré par S et T.

Démonstration. On appelle D l'ensemble

$$D = \left\{ z \in P \mid |z| \geqslant 1 \text{ et } |\Re(z)| \leqslant \frac{1}{2} \right\},$$

et G le sous-groupe de $SL_2(\mathbb{Z})$ engendré par S et T.

On cherche donc à montrer que $G = SL_2(\mathbb{Z})$. Commençons par étudier D.

Lemme 1.68
On fait agir G sur P via l'action de $SL_2(\mathbb{Z})$. Alors pour cette action de G, toutes les orbites rencontrent D.

Démonstration. On se fixe $z \in P$. On va construire un point de P qui est dans D à partir de z ; pour cela, on va faire "monter" z, puis le translater.

On remarque qu'il n'y a qu'un nombre fini de couples $(c,d) \in \mathbb{Z}^2$ tels que $|cz+d| \leqslant 1$. En effet

$$|c|\Im(z) = |\Im(cz+d)| \leqslant |cz+d| \leqslant 1,$$

et donc $|c| \leqslant 1/\Im(z)$ et $|d| \leqslant 1 + |c||z|$.

On a, pour $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ dans G,

$$\Im(A * z) = \frac{\Im(z)}{|cz+d|^2}.$$

Comme le nombre de couples (c,d) vérifiant $\Im(A*z) \geqslant \Im(z)$ est fini, il existe $A_1 \in G$ telle que $\Im(A_1 * z)$ soit maximal. On appelle $z_1 = A_1 * z$.

Soit n la partie entière de $\Re(z_1) + 1/2$.

Comme pour tout u de P, on a $T * u = u + 1$, on a :

$$-\frac{1}{2} \leqslant \Re(z_1) - n \leqslant \Re(z_1 - n) \leqslant \Re(T^{-n} * z_1) \leqslant \frac{1}{2},$$

avec $\Im(T^{-n} * z_1) = \Im(z_1)$.

En posant $z_2 = T^{-n} * z_1$), il ne nous reste plus qu'à montrer $|z_2| \geqslant 1$.

Sinon, $\Im(S * z_2) = \frac{\Im(z_2)}{|z_2|^2}$, ce qui contredit la maximalité de $\Im(z_2)$. ◊

On chercher maintenant à caractériser, pour $z \in D$ fixé, les matrices A de $SL_2(\mathbb{Z})$ telles que $A * z \in D$.

Fixons-nous donc $z \in D$, et $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ dans $SL_2(\mathbb{Z})$ tels que $A * z \in D$.

Lemme 1.69
A est une matrice de G.

Démonstration. On peut se restreindre au cas où $\Im(A*z) \geqslant \Im(z)$, *i.e* $|cz+d| \leqslant 1$: sinon $\Im(A*z) < \Im(A^{-1}*(A*z))$ et donc on peut faire la même étude pour A^{-1}.

D'après les calculs qu'on a fait plus tôt, on a

$$|c| \leqslant \frac{1}{\Im(z)} \leqslant \frac{2}{\sqrt{3}} < 2$$

et donc $c \in \{-1, 0, 1\}$.

Cas $c = 0$: On a $\det(A) = ad = 1$, et donc, quitte à changer A et $-A$ (ce qui ne change pas la valeur de $A * z$), on a $a = d = 1$, et donc $A * z = z + b$. Regardons où est z dans D.
- Si $|\Re(z)| < \frac{1}{2}$, alors $b = 0$ et donc $A = \pm I$.
- Si $|\Re(z)| = -\frac{1}{2}$, alors $b = 0$ ou 1, et donc $A = \pm I$ ou T.
- Si $|\Re(z)| = \frac{1}{2}$, alors $b = 0$ ou -1, et donc $A = \pm I$ ou T^{-1}.

Cas $c = 1$: La condition $|z + d| \leqslant 1$ n'offre que trois possibilités (dessiner D) :

(i) $d = 0$;

(ii) $z = j$ et $d = 1$;

(iii) $z = -\frac{1}{j}$ et $d = -1$.

(i) On a $\det(A) = -b = 1$, et $A * z = a - \frac{1}{z}$.
Comme de plus, $|z| \leqslant 1$ et $z \in D$, nécessairement $|z| = 1$, et donc $-\frac{1}{z}$ et le symétrique de z par rapport à l'axe des imaginaires. $a - \frac{1}{z}$ est donc dans D seulement si
- $a = 0$
- $z = j$ et $a = -1$
- $z = -\frac{1}{j}$ et $a = 1$

Ces trois possibilités mènent à
- $A = S$
- $A = \begin{pmatrix} -1 & -1 \\ 1 & 0 \end{pmatrix} = (ST)^2$
- $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = TS.$

(ii) $\det(A) = a - b = 1$ et

$$A * z = \frac{aj + (a-1)}{j+1} = a + j,$$

qui est dans D seulement si $a = 0$ ou $a = 1$, ce qui mène à

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix} = ST \text{ ou } A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = TST.$$

(iii) On conclut de la même façon

$$A = \begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix} = (TS)^2 \text{ ou } A = \begin{pmatrix} -1 & 0 \\ 1 & -1 \end{pmatrix} = TST.$$

On se ramène à $c = 1$ en changeant A en $-A$.

$\Diamond$

On peut maintenant terminer la preuve :

Le cas où z est dans l'intérieur de D ne se rencontre que dans le cas $c = 0$, et impose $A = \pm I$.

Soit A une matrice de $SL_2(\mathbb{Z})$. Notons $z' = A * z \in P$. Alors par le lemme 1.68, il existe $B \in G$ telle que $B * z' \in D$, *i.e* $(BA) * z \in D$.

On a donc $BA = \pm I$, *i.e* $A = \pm B^{-1}$. Comme $-I = S^2$ est dans G, on a bien $A \in G$, et par suite $G = SL_2(\mathbb{Z})$.

$\square$

1.24 Sous-groupes compacts de GL_n

Alessandri (n.d.)

Dans la suite, G sera un sous-groupe compact de $GL_n(\mathbb{R})$. Le théorème à démontrer est :

Théorème 1.70
Il existe un produit scalaire euclidien $(\cdot \mid \cdot)$ sur $\mathbb{R}^n$ de forme quadratique associée q, tel que G soit inclu dans l'ellipsoïde associé à q :

$$G \subseteq \mathscr{O}(q).$$

Cas $c = -1$: *Démonstration.* Soit $\mathscr{V}$ un espace vectoriel de dimension n. Commençons par montrer un lemme sur l'enveloppe convexe d'un compact :

Lemme 1.71
L'enveloppe convexe d'un compact de $\mathscr{V}$ est compacte.

Démonstration. Soit K un compact de $\mathscr{V}$ non vide, et soit $\mathrm{Conv}(K)$ son enveloppe convexe. On pose pour $n \in \mathbb{N}^*$:

$$\beta_n = \left\{ \sum_{i=1}^{n} \lambda_i x_i \;\middle|\; \sum_{i=1}^{n} \lambda_i = 1 \text{ et } \forall i,\ (\lambda_i, x_i) \in \mathbb{R}^+ \times K \right\}.$$

Le lemme de Caratheodory nous donne $\mathrm{Conv}(K) = \beta_{n+1}$.

On pose $\Lambda_{n+1} = \left\{ (\lambda_i)_{1 \leqslant i \leqslant n} \in \mathbb{R}^{n+1} \;\middle|\; \sum_{i=1}^{n+1} \lambda_i = 1 \right\}$, et

$$P : \begin{array}{ccc} \Lambda_{n+1} \times K^{n+1} & \longrightarrow & \mathrm{Conv}(K) \\ (\lambda, x) & \longmapsto & \sum_{i=1}^{n+1} \lambda_i x_i \end{array}.$$

P est polynomiale donc continue, et le lemme de Carathéodory nous montre qu'elle est surjective. Λ_{n+1} est compact dans $\mathbb{R}^{n+1}$, et donc $\Lambda_{n+1} \times K^{n+1}$ est compact aussi.

Donc $\mathrm{Conv}(K)$ est l'image d'un compact par une application continue, donc est compact. ◊

Soit K un compact convexe de $\mathscr{V}$, et soit V dans $GL_n(\mathbb{R})$ tel que K soit stable par V.

Alors :

Lemme 1.72
V admet un point fixe dans K.

Démonstration. On choisit un x_0 dans K, et on pose

$$x_k = \frac{1}{k+1} \sum_{i=0}^{k} V^i(x_0).$$

K est stable par V, donc $x_k \in K$ pour tout $k \in \mathbb{N}$. Comme K est compact, on peut donc supposer que x_k

converge, à extraction près, vers un élément $x \in K$. On a donc

$$Vx_k = x_k + \underbrace{\frac{1}{k+1}(V^{k+1}x_0 - x_0)}_{:=\varepsilon_k}.$$

K est compact donc borné, et donc ε_k tend vers 0 avec k.

Donc, en passant à la limite, on a $Vx = x$. ◊

Soit maintenant $\mathscr{G}$ un sous-groupe compact de $GL(\mathscr{V})$. On pose, pour tout x de $\mathscr{V}$, $\nu(x) = \max_{U\in\mathscr{G}} \|Ux\|$, où $\|\cdot\|$ est une norme euclidienne sur $\mathscr{V}$. Alors

Lemme 1.73
ν est une norme $\mathscr{G}$-invariante, i.e *si $U \in \mathscr{G}$, alors $\nu(Ux) = \nu(x)$.*

Démonstration. ν est bien définie par compacité, est $\mathscr{G}$-invariante car $\mathscr{G}$ est un groupe, et est bien homogène et définie positive. Montrons l'inégalité triangulaire : Soient x et y dans $\mathscr{V}$, et soit u_0 dans $\mathscr{G}$ tel que $\nu(x+y) = N(u_0(x+y))$. Alors

$$\begin{aligned}\nu(x+y) &= \|u_0(x) + u_0(y)\| \\ &\leqslant \|u_0(x)\| + \|u_0(y)\| \\ &\leqslant \nu(x) + \nu(y)\end{aligned}$$

Donc ν est bien une norme $\mathscr{G}$-invariante sur $\mathscr{V}$.

On remarque que le cas d'égalité dans l'inégalité triangulaire équivaut à celui de $\|\cdot\|$, *i.e* la dépendance positive des deux vecteurs. ◊

On peut maintenant établir :

Lemme 1.74
On suppose que tous les éléments de $\mathscr{G}$ laissent K stable. Alors il existe un point fixe dans K commun à tous les éléments de $\mathscr{G}$.

Démonstration. Pour $U \in \mathscr{G}$, on pose $F_U = \{x \in K \mid Ux = x\}$. On cherche un élement dans $\bigcap_{U\in\mathscr{G}} F_U$.

(F_U) est une famille de fermés dans le compact K, il suffit de montrer que les intersections finies de F_U sont non vides :

$$\forall p \in \mathbb{N}^*,\ \forall U_1, \ldots, U_p \in \mathscr{G},\ \bigcap_{i=1}^{p} F_{U_K}.$$

Soient donc $p \in \mathbb{N}^*$ et $U_1, \ldots, U_p \in \mathscr{G}$. On pose $V = \frac{1}{p}\sum_{i=1}^{p} U_i$. K est convexe, et donc est laissé stable par V, et donc V admet un point fixe x dans K. On a alors

$$\begin{aligned}\nu(x) &= \nu(Vx) \\ &\leqslant \frac{1}{p}\sum_{i=1}^{p} \nu(U_k x) \\ &\leqslant \nu(x)\end{aligned}$$

D'après le cas d'égalité dans l'inégalité triangulaire, tous les $U_k x$ sont positivements liés, et valent même x.

Donc l'intersection finie est non vide, d'où le lemme. ◊

Passons maintenant à la preuve :

On considère l'application

$$\rho : \begin{array}{ccc} G & \longrightarrow & GL(\mathscr{S}_n) \\ X & \longmapsto & S \mapsto {}^tXSX \end{array} .$$

ρ est alors un morphisme de groupes, continu car polynomial en les coefficients de la matrice.

$\mathscr{G} = \rho(G)$ est donc un sous groupe compact de $GL(\mathscr{S}_n)$ (par continuité de ρ et compacité de G).

G est compact, donc $\{{}^tMM \mid M \in G\}$ est compact non vide dans le convexe $\mathscr{S}_n^{++}$, et donc son enveloppe convexe K aussi.

K est clairement $\mathscr{G}$-stable, et donc il existe un élément $S \in K$ fixe par tous les éléments de $\mathscr{G}$.

Donc $S \in \mathscr{S}_n^{++}$, et $\forall A \in G,\ {}^tASA = S$.

Donc, en notant $(\cdot \mid \cdot)$ le produit scalaire sur $\mathbb{R}^n$ défini par $(x \mid y) = \langle x, Sy\rangle$ et q la forme quadratique associée, on a pour $M \in G$:

$$\begin{aligned} \forall x \in \mathbb{R}^n,\ q(Mx) &= (Mx \mid SMx) \\ &= (x \mid {}^tMSMx) \\ &= (x \mid Sx) \\ &= q(x) \end{aligned}$$

Donc $M \in \mathscr{O}(q)$.

Donc $G \subseteq \mathscr{O}(q)$. □

2 Informatique

2.1 Arbres binaires de recherche optimaux

Cormen *et al.* (n.d.)

On s'intéresse ici à optimiser les arbres binaires de recherche quand on connait la probabilité pour chaque clef d'être recherchée.

On se donne n clefs distinctes triées dans l'ordre croissant $k_1 < \cdots < k_n$, où pour tout i, k_i a une probabilité p_i d'être recherchée.

On rajoute aussi $n+1$ clefs factices $d_0, \ldots, d_n$ représentant les recherches qui ne sont aucun des k_i.

d_0 représente toutes les valeurs inférieures à k_1, d_n celle supérieures à k_n, et d_i celles comprises entre k_i et k_{i+1}. On se donne de plus une probabilité q_i que la recherche soit dans d_i.

Les nœuds internes de l'arbre seront les k_i, et les feuilles les d_i.

On a donc

$$\sum_{i=1}^{n} p_i + \sum_{i=0}^{n} q_i = 1.$$

Le coût moyen d'une recherche dans notre arbre sera

$$1 + \sum_{i=1}^{n} \text{Prof}(k_i)p_i + \sum_{i=0}^{n} \text{Prof}(d_i)q_i.$$

Un arbre binaire de recherche sera dit *optimal* si ce coût moyen est minimal. On abrège en "ABRO".

La recherche d'un ABRO en listant tous les arbres possibles a une complexité bien trop grande, on va donc essayer un algorithme de programmation dynamique.

On remarque pour cela que tout sous-arbre de notre arbre *doit* être optimal : sinon, on le remplace par un optimal, faisant ainsi baisser le coût total.

Le sous-problème à résoudre est donc, étant donnés $i \geqslant 1$ et $i - 1 \leqslant j \leqslant n$ un ABRO contenant les clefs $k_i, \ldots, k_j$.

Soit $e[i, j]$ le coût d'un arbre contenant $k_i, \ldots, k_j$.

Si $j = i - 1$, alors il n'y a que la clef factice d_{i-1} dans l'arbre : le coût moyen est q_{i-1}.

Sinon, $j \geqslant i$. On choisit un nœud k_r, $i \leqslant r \leqslant j$, et on construit l'arbre de racine k_r. Le sous-arbre gauche de k_r contiendra les clefs $k_i, \ldots, k_{r-1}$, et le sous arbre droit les clefs $k_{r+1}, \ldots, k_j$.

Quand on fait "descendre" un sous-arbre, on augmente la profondeur de ce sous-arbre de 1, on donc on augmente le coût de

$$w(i, j) = \sum_{k=i}^{j} p_k + \sum_{k=i-1}^{j} q_k.$$

On obtient donc la formule :

$$e[i, j] = p_r + e[i, r - 1] + w(i, r - 1) + e[r + 1, j] + w(r + 1, j).$$

On remarque que

$$w(i, j) = w(i, r - 1) + p_r + w(r + 1, j),$$

et donc

$$e[i, j] = e[i, r - 1] + e[r + 1, j] + w(i, j).$$

Cette formule est valide si on a choisi la bonne racine pour notre sous-arbre. On optimise donc :

$$e[i, j] = \begin{cases} q_{i-1} & \text{si } j = i - 1 \\ \min_{i \leqslant r \leqslant j} e[i, r - 1] + e[r + 1, j] + w(i, j) & \text{si } i \leqslant j \end{cases}$$

Pour garder une trace de notre construction, on définit $racine[i, j]$ comme l'indice de la racine optimale pour le sous-arbre contenant $k_i, \ldots, k_j$.

On va maintenant utiliser la programmation dynamique pour calculer $e[1, n]$.

Remarque – Plutôt que de calculer tous les $w(i, j)$ à chaque fois, on crée un tableau défini par

$$\begin{cases} w(i, i - 1) = q_{i-1} \\ w(i, j) = w(i, j - 1) + p_j + q_j \end{cases}$$

On a donc l'algorithme suivant :

Les trois boucles imbriquées nous donnent une complexité en $\mathcal{O}(n^3)$.

Algorithm 1 ABRO

Préconditions: p, q probabilités, n nombre de clefs

```
pour i = 1 à n + 1 faire
        e[i, i − 1] ← q_{i−1}
        w[i, i − 1] ← q_{i−1}
ruop
pour ℓ = 1 à n faire
        pour i = 1 à n − ℓ + 1 faire
                j ← i + ℓ − 1
                e[i, j] ← ∞
                w[i, j] ← w[i, j − 1] + p_j + q_j
                pour r = i à j faire
                        t ← e[i, r − 1] + e[r + 1, j] + w[i, j]
                        si t < e[i, j] alors
                                e[i, j] ← t
                                racine[i, j] ← r
                        is
                ruop
        ruop
ruop

retourner  e et racine
```

2.2 Fonction d'Ackermann

Cori & Lascar (tome 2)

Historique : Wilhem Ackermann, alors élève de Hilbert, proposa en 1928 un exemple de fonction récursive, non récursive primitive, à trois variables :

$$A(m, n, p) = m \to n \to p = m \uparrow^p n.$$

C'est en fait une fonction semblable, mais à seulement deux variables, proposée par Rózsa Péter qu'on appelle aujourd'hui *fonction d'Ackermann*.

Définition 2.1
On définit la fonction d'Ackermann ξ *ainsi :*
– *Pour tout entier* x, $\xi(0, x) = 2^x$.
– *Pour tout entier* y, $\xi(y, 0) = 1$.
– *Pour tous entiers* x *et* y, $\xi(y + 1, x + 1) = \xi(y, \xi(y + 1, x))$.
Pour chaque entier n, *on note* ξ_n *la fonction* $x \mapsto \xi(n, x)$, *et on a alors :*

$$\begin{cases} \xi_0(x) = 2^x \\ \xi_n(0) = 1 \\ \xi_n(x + 1) = \xi_{n-1}(\xi_n(x)) \end{cases}$$

On voit avec la définition récursive des ξ_n que pour tout entier n, la fonction ξ_n est primitive récursive (par procédé de récurrence, sur n).

En revanche, cela n'implique pas que ξ est primitive récursive, et d'ailleurs :

Théorème 2.2
La fonction d'Ackermann ξ n'est pas primitive récursive.

Démonstration. On va montrer plein de lemmes techniques pour finalement montrer la fonction ξ domine toutes les fonctions primitives récursives.

Lemme 2.3
Pour tous n et x entiers,

$$\xi_n(x) > x.$$

Démonstration. Montrons-le par récurrence sur n.

Soit donc ϕ_n la proposition « Pour tout entier x, $\xi_n(x) > x$ ».

- Il est assez clair que ϕ_0.
- Soit n un entier. Supposons ϕ_{n-1}, et montrons ϕ_n. Pour cela, faisons une récurrence sur x.
  Soit donc ψ_x la proposition « $\xi_n(x) > x$ ».
  - ψ_0 est toujours clair.
  - Soit x un entier. Supposons ψ_{x-1} et montrons ψ_x.
    On a par définition $\xi_n(x) = \xi_{n-1}(\xi_n(x-1))$, et donc, par ϕ_{n-1} on a $\xi_n(x) > \xi_n(x-1)$, c'est-à-dire

$$\xi_n(x) \geqslant \xi_n(x-1) + 1.$$

    D'après ψ_{x-1}, $\xi_n(x-1) > x-1$, d'où ψ_x
  On a donc ϕ_n.

Donc, par récurrence, on a le lemme. ◊

Lemme 2.4
Pour tout n, ξ_n est strictement croissante.

Démonstration. $\xi_0 = \lambda x.2^x$, donc ξ_0 est croissante.

Pour $n \geqslant 1$, on a la formule $\xi_n(x+1) = \xi_{n-1}(\xi_n(x))$, et donc par le lemme 2.3, $\xi_n(x+1) > \xi_n(x)$. ◊

Lemme 2.5
Pour tout $n \geqslant 1$, pour tout entier x, on a :

$$\xi_n(x) \geqslant \xi_{n-1}(x).$$

Démonstration. Par récurrence sur x :

- Pour $x = 0$, ouais.
- Pour l'hérédité :
  On a vu $\xi_n(x) \geqslant x+1$, donc par croissance de ξ_{n-1}, on a

$$\xi_{n-1}(\xi_n(x)) \geqslant \xi_{n-1}(x+1).$$

  Or par définition, $\xi_{n-1}(\xi_n(x)) = \xi_n(x+1)$, d'où le résultat.

◊

Définition 2.6
On note dans la suite ξ_n^k la fonction ξ_n itérée k fois.

Lemme 2.7
Les fonctions ξ_n^k sont strictement croissantes. De plus pour tous entiers m, n, k, h, x :

$$\begin{aligned} \xi_n^k(x) &< \xi_n^{k+1}(x) \\ \xi_n^k(x) &\geqslant x \\ \xi_n^k \circ \xi_n^h &= \xi_n^{k+h} \\ \text{Si } m \leqslant n,\ \xi_m^k(x) &\leqslant \xi_n^k(x) \end{aligned}$$

Démonstration. C'est facile par récurrence. ◊

Définition 2.8
Soient $f : \mathbb{N} \to \mathbb{N}$ strictement croissante, et $g : \mathbb{N}^p \to \mathbb{N}$.

On dit que f domine *g si pour tout $(x_1, \ldots, x_p)$ sauf un nombre fini,*

$$g(x_1, \ldots, x_p) \leqslant f(\sup(x_1, \ldots, x_p)).$$

On pose alors

$$C_n = \{g \mid \textit{il existe } k \textit{ tel que } g \textit{ soit dominée par } \xi_n^k\}.$$

On voit facilement que certaines fonctions sont dans C_0 : les projections, les constantes, la fonction successeur, les fonctions sup, l'addition, la multiplication par un entier.

On remarque aussi que $\xi_n \in C_n$, et si $f, g : \mathbb{N}^p \to \mathbb{N}$ telles $f(x_1, \ldots, x_p) \leqslant g(x_1, \ldots, x_p)$ pour tout $(x_1, \ldots x_p)$, alors $g \in C_n$ implique $f \in C_n$.

Si on montre que $C := \bigcup C_n$ est stable par composition et récurrence, alors C contiendra donc toutes les fonctions primitives récursives.

Lemme 2.9
Pour tout n, C_n est clos par composition.

Démonstration. Dans cette preuve, les inégalités seront pour tout uple, sauf un nombre fini.

Soient $f_1, \ldots, f_m : \mathbb{N}^p \to \mathbb{N}$ et $g : \mathbb{N}^m \to \mathbb{N}$ des fonctions de C_n.

Montrons que $g(f_1, \ldots, f_m)$ est dans C_n aussi.

On a l'existence de $k, k_1, \ldots, k_m$ tels que

$$\begin{aligned} g(y_1, \ldots, y_m) &\leqslant \xi_n^k(\sup(y_1, \ldots, y_m)) \\ f_i(x_1, \ldots, x_p) &\leqslant \xi_n^{k_i}(\sup(x_1, \ldots, x_p)) \end{aligned}$$

Posons $h = \sup(k_1, \ldots, k_m)$.

On a alors par le lemme 2.7 :

$$g(f_1(x_1, \ldots, x_p), \ldots, f_m(x_1, \ldots, x_p)) \leqslant \xi_n^k(\xi_n^h(\sup(x_1, \ldots, x_p))),$$

d'où le résultat avec le lemme 2.7 à nouveau. ◊

Lemme 2.10
Pour tous entiers n, k et x, on a :

$$\xi_n^k(x) \leqslant \xi_{n+1}(x+k).$$

Démonstration. Par récurrence sur k. ◊

Lemme 2.11
Si $g : \mathbb{N}^p \to \mathbb{N}$ et $h : \mathbb{N}^{p+2} \to \mathbb{N}$ sont dans C_n, alors f définie par récurrence à partir de g et h est dans C_{n+1}.

Démonstration. Comme dans la preuve précédente, les inégalités seront pour tout uple sauf un nombre fini.

Les hypothèses sont :

$$g(x_1, \ldots, x_p) \leqslant \xi_n^{k_1}(\sup(x_1, \ldots, x_p))$$
$$h(x_1, \ldots, x_p, y, z) \leqslant \xi_n^{k_2}(\sup(x_1, \ldots, x_p, y, z))$$

Montrons par récurrence sur y la propriété $\Phi_y =$« $f(x_1, \ldots, x_p, y) \leqslant \xi_n^{k_1+yk_2}(\sup(x_1, \ldots, x_p, y))$ ».
– Φ_0 est évident.
– Supposons Φ_y et montrons Φ_{y+1}.
On a

$$\begin{aligned} f(x_1, \ldots, x_p, y+1) &= h(x_1, \ldots, x_p, y, f(x_1, \ldots, x_p, y)) \\ &\leqslant \xi_n^{k_2}(\sup(x_1, \ldots, x_p, y, f(x_1, \ldots, x_p, y))) \\ &\leqslant \xi_n^{k_2}(\xi_n^{k_1+yk_2}(\sup(x_1, \ldots, x_p, y))) \text{ par le lemme 2.7 et hypothèse de récurrence} \end{aligned}$$

D'où Φ_{y+1}, et donc Φ_y pour tout y.
En appliquant le lemme 2.10, on a :

$$f(x_1, \ldots, x_p, y) \leqslant \xi_{n+1}(\sup(x_1 \ldots, x_p, y) + k_1 + yk_2).$$

La fonction $\lambda x_1 x_2 \cdots x_p y.\xi_{n+1}(\sup(x_1, \ldots, x_p, y) + k_1 + yk_2)$ est composée de fonctions de C_{n+1}, donc est C_{n+1}, et donc f aussi. ◊

Corollaire 2.12
C contient toutes les fonctions récursives primitives.

On peut maintenant montrer que ξ n'est pas primitive récursive.

Sinon, $\lambda x.\xi(x, 2x)$ l'est aussi, donc est dans C :

Il existe n, k tels que pour tout x sauf un nombre fini, $\xi(x, 2x) \leqslant \xi_n^k(x)$.

Donc on a $\xi(x, 2x) \leqslant \xi_{n+1}(x+k)$ par le lemme 2.10.

De plus, pour x assez grand :

$$\begin{aligned}\xi_{n+1}(x+k) &< \xi_{n+1}(2x)\\ &< \xi_x(2x)\\ &= \xi(x,2x)\end{aligned}$$

D'où une contradiction. □

2.3 Approximation du problème du voyageur de commerce

Cormen *et al.* (n.d.)

On rappelle le problème du voyageur de commerce (PVC dans la suite) :
- On se donne un graphe non orienté complet $G = (S, A)$.
- Chaque arête $(u, v) \in A$ a un coût entier positif ou nul $c(u, v)$.
- On cherche un cycle hamiltonien dans G qui a un coût minimal.

Dans la vraie vie (!), il semble naturel d'avoir l'inégalité triangulaire, *i.e* pour des arêtes $(u, v), (v, w) \in A$, alors

$$c(u, w) \leqslant c(u, v) + c(v, w).$$

Avec cette hypothèse, on appelle le PVC le PVC Euclidien. Le PVCE reste NP-complet, mais on peut trouver un bon algorithme d'approximation.

On rappelle l'algorithme de Prim pour trouver un arbre couvrant de poids minimal, qu'on utilisera comme sous-routine dans notre algorithme.

Algorithm 2 ACM-Prim

Préconditions: G graphe, w fonction de coût, r sommet de départ
pour $u \in S[G]$ **faire**
clé$[u] \leftarrow \infty$
$\pi[u] \leftarrow$ Nil
ruop
clé$[r] \leftarrow 0$
$F \leftarrow S[G]$
tant que $F \neq \emptyset$ **faire**
$u \leftarrow$ Extraire-Min(F)
pour $v \in Adj[u]$ **faire**
si $v \in F$ et $w(u, v) <$ clé$[v]$ **alors**
$\pi[v] \leftarrow u$
clé$[v] \leftarrow w(u, v)$
is
ruop
euq tnat

C'est un algorithme glouton : on construit à chaque étape un grape $G_E = (S, E)$, en augmentant E à chaque fois par l'arête de poids minimal rejoingnant un sommet isolé de S.

Il a une complexité, en représentant le graphe par sa matrice d'adjacence, en $\Theta(S^2)$.

On utilise cet algorithme pour notre algorithme d'approximation :

Algorithm 3 PVCE-Approché

Préconditions: G graphe, c fonction de coût
Choisir r qui fera office de racine
Calculer T=ACM-Prim(G, c, r)
Calculer L liste des sommets visités dans le parcours préfixe de T

retourner cycle hamiltonien H qui visite les sommets dans l'ordre de L

On donne un exemple, pour la distance euclidienne :

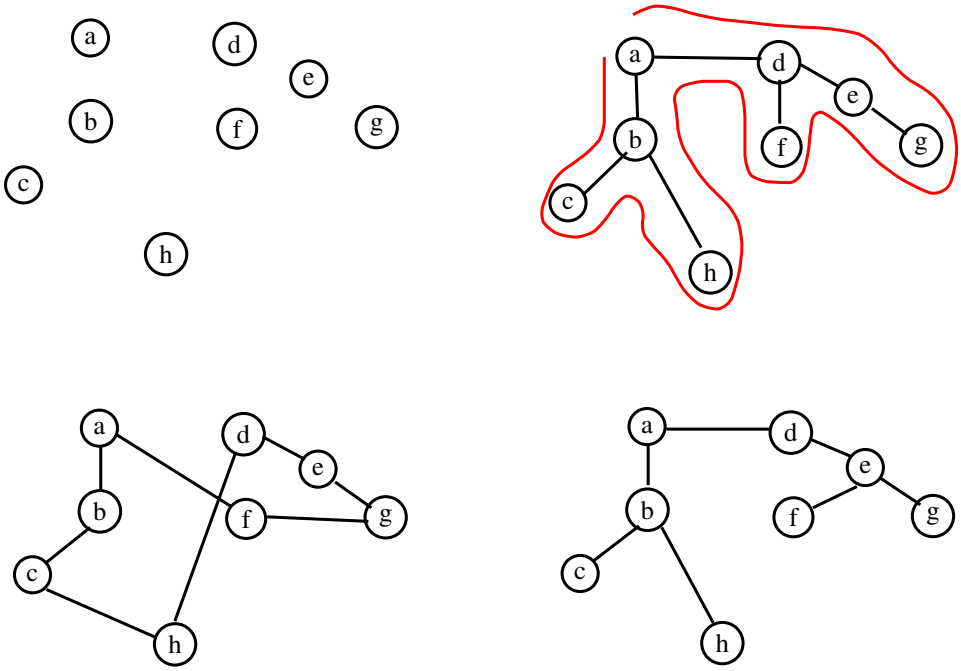


FIGURE 1 – Un exemple

La partie la plus complexe est l'appel à ACM-Prim, et donc la complexité est en $\Theta(S^2)$.

On a alors :

Théorème 2.13
PVCE-Approché est un algorithme d'approximation avec une garantie de performance 2 à temps polynomial pour le PVCE.

Démonstration. On a déjà vu que PVCE-Approché s'exécute en temps polynomial.

Soit H^* un chemin hamiltonien optimal pour un graphe donné.

L'arbre couvrant s'obtient en supprimant des arêtes dans une tournée, et donc il est assez clair que

$$c(T) \leqslant c(H^*).$$

Un parcours *complet* de T liste les sommets quand ils sont visités, pour la première fois ou après la visite d'un sous-arbre. On note W ce parcours. Pour l'exemple, on a

$$W = (a, b, c, b, h, a, d, f, e, g, e, d, a).$$

Ce parcours traverse chaque arête deux fois exactement, et donc $c(W) = 2c(T)$. On en déduit

$$c(W) \leqslant 2c(H^*).$$

Cependant, W n'est en général pas un cycle hamiltonien (les sommets peuvent être visités plusieurs fois), il faut donc le modifier. C'est ici qu'on utilise l'inégalité triangulaire : on peut enlever de W les sommets déjà visités, sans augmenter le coût. Pour l'exemple, W devient

$$W = (a, b, c, h, d, e, f, g).$$

W est donc la même liste que celle obtenue par parcours préfixe de T. On note H le cycle correspondant au parcours préfixe : c'est bien un cycle hamiltonien, qui est calculé par PCVE-Approché. Comme on a supprimé des sommets de W, on a clairement

$$c(H) \leqslant c(W).$$

Finalement, on a $c(H) \leqslant 2c(H^*)$, et donc on a le résultat. □

Remarque – On note que sans l'hypothèse d'inégalité triangulaire, on peut montrer qu'il n'existe pas l'algorithme d'approximation polynomial.

2.4 Automate des occurences

On cherche à retrouver un mot x dans un texte t. On note dans la suite A l'alphabet sur lequel est écrit le texte t.

x apparaît dans t *si et seulement si* $t \in A^*xA^*$, et donc le problème revient à trouver les préfixes de t dans A^*x.

On peut facilement construire un automate non déterministe qui convient, mais le non-determinisme est réd-hibitoire. On va plutôt chercher à construire l'automate minimal de A^*x, qui a lui aussi $|x|+1$ états.

On définit la fonction f_x par

$$f_x : \begin{array}{ccc} A^* & \longrightarrow & A^* \\ u & \longmapsto & \text{le plus long suffixe de } u \text{ qui est préfixe de } x. \end{array}.$$

On considère l'automate $\mathscr{A} = (Q, A, \delta, \{\varepsilon\}, \{x\})$ où Q est l'ensemble des préfixes de x, et δ est définie par

$$\delta(p, a) = f_x(pa).$$

Théorème 2.14
$\mathscr{A}$ est l'automate minimal reconnaissant $L = A^*x$.

Démonstration. On sait que les états de l'automate minimal sont les $u^{-1}L$. On va donc montrer que pour tout mot u sur A :

$$u^{-1}L = f_x(u)^{-1}L.$$

Soit donc $u \in A^*$, et soit u' tel que $u = u'f_x(u)$. On a alors

$$\begin{aligned} u^{-1}L &= f_x(u)^{-1}u'^{-1}L \\ &\supset f_x(u)^{-1}L \end{aligned}$$

Réciproquement, soit $w \in u^{-1}L$. Par définition, $uw \in L$, et donc il existe v tel que $uw = vx$. Deux cas se présentent :
Cas 1 : x est un suffixe de w. On a alors $w \in L$ et donc $f_x(u)w \in L$ et donc $w \in f_x(u)^{-1}L$.
Cas 2 : w est un suffixe de x. Soit z tel que $x = zw$. On a alors $u = vz$, et donc z est un suffixe de u et un préfixe de x. Par définition de $f_x(u)$, il existe donc y tel que $f_x(u) = yz$.
On a alors $f_x(u)w = yzw = yx$ et donc $w \in f_x(u)^{-1}L$.
L'ensemble des états de l'automate minimal est donc $\{f_x(u)^{-1}L \mid u \in A^*\}$ qui est exactement $\{p^{-1}L \mid p \text{ préfixe de } x\}$.

Il nous reste donc à montrer que pour deux préfixe p et q de x, si $q^{-1}L = p^{-1}L$, alors $p = q$.

Pour cela, on suppose par exemple que q est préfixe de p : $p = qp'$.

On a alors $x = pp'' = qp'p''$.

Donc $p'' \in p^{-1}L$, et on en déduit par hypothèse $p'' \in q^{-1}L$. Il existe ainsi v tel que $qp'' = vx = vqp'p''$.

Par simplification, on a $v = p' = \varepsilon$, et donc $p = q$. □

Il ne reste plus qu'à faire passer le texte dans l'automate, et vérifier à chaque étape si on est dans un état final.

L'algorithme se fait donc en temps au pire $\mathscr{O}(|t|)$, et l'automate $\mathscr{A}$ est stocké en espace $\mathscr{O}(|A|(|x|+1))$.

On cache cependant une partie de la complexité dans le calcul de f_x. En fait, on peut la calculer facilement (en $\mathscr{O}(|x|)$), grâce à la relation

$$f_x(pa) = \begin{cases} pa & \text{si } pa \text{ est préfixe de } x \\ Bord(pa) & \text{sinon} \end{cases}$$

où $Bord(u)$ est défini par le plus long mot distinct de u qui est à la fois préfixe et suffixe de u.
On peut calculer les bords avec la relation, pour un mot u et une lettre a :

$$Bord(ua) = \begin{cases} Bord(x)a & \text{si } Bord(x)a \text{ est un préfixe de } x \\ Bord(Bord(x)a) & \text{sinon} \end{cases}$$

2.5 Théorème de Cook

Carton (n.d.)

Théorème 2.15 : de Cook, 1971
Le problème SAT *est NP-complet.*

Démonstration. Soit A un problème de NP, et soit $\mathscr{M}$ une machine de Turing non-déterministe qui décide A en temps polynomial.

Pour chaque entrée w, on va construire une formule φ_w qui sera satisfiable *si et seulement si* $\mathscr{M}$ accepte w.

On note $n = |w|$. On peut supposer que chaque calcul acceptant sur w est de longueur exactement n^k (quitte à rajouter des transitions inutiles).

La machine utilise donc au plus n^k cellules de sa bande de travail, et donc les configurations sont de longueur au plus n^k : de même, on les prendra de longueur exactement n^k, quitte à rajouter des symboles blancs.

On les note dans un tableau :

Conf.	0	1	2	3	$\cdots$	n^k
$C_0 =$	q_0	w_1	w_2	w_3	$\cdots$	$\sharp$
$C_1 =$	w'_1	q_1	w_2	w_3	$\cdots$	$\sharp$
$C_2 =$	w'_1	w_2	q_2	w_3	$\cdots$	$\sharp$
$C_2 =$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\sharp$
$\vdots$						$\vdots$
$C_{n^k} =$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$

On va donc coder une formule φ_w qui code l'existence d'un tel tableau.

On définit les variables $x_{i,j,a}$ pour $i, j \in [\![0, n^k]\!]$ et a symbole de $A = \Gamma \cup Q$ qui codent le fait que la variable a se trouve dans la case i, j. Il y a $|A|n^{2k+2}$ telles variables.

On décompose notre formule φ_w en quatre formules φ_0, φ_1, φ_2 et φ_3, qui vont chacune coder une propriété du tableau.

φ_0 : Cette formule code le fait que chaque case du tableau contient un et un seul symbole de A :

$$\varphi_0 = \bigwedge_{0 \leqslant i,j \leqslant n^k} \left[\left(\bigvee_{a \in A} x_{i,j,a} \right) \wedge \left(\bigwedge_{a \neq a' \in A} (\overline{x}_{i,j,a} \vee \overline{x}_{i,j,a'} \right) \right].$$

φ_1 : Cette formule code le fait que la première ligne du tableau est bien $q_0 w$:

$$\varphi_1 = \left(\bigwedge_{0 \leqslant i \leqslant n} x_{0,i,w_i} \right) \wedge \left(\bigwedge_{n+1 \leqslant i \leqslant n^k} x_{0,i,\sharp} \right).$$

φ_2 : Cette formule assure que chaque ligne est obtenue en appliquant une transition valide de $\mathscr{M}$.
Il suffit de remarquer que la valeur d'une case (i, j) ne dépend que des trois cases au-dessus $(i-1, j-1)$, $(i-1, j)$ et $(i-1, j+1)$.
Si dans ces trois cases se trouvent des symboles de bande, alors le contenu de la case (i, j) est le même qu'en $(i-1, j)$.
Si l'état de la configuration se trouve en $(i-1, j)$, alors l'état de C_i se trouve en $(i, j-1)$ ou $(i, j+1)$.
Donc, il suffit bien de regarder les "fenêtres" de taille 2×3 du tableau. L'ensemble des fenêtres possibles ne dépend que de A et des transitions de $\mathscr{M}$, et donc ne dépend pas de la taille de l'entrée n.

Le fait que chaque fenêtre du tableau corresponde bien à une transition s'écrit donc comme une conjonction pour $0 \leqslant i, j \leqslant n^k$ de disjonctions des fenêtres possibles, ce qui est polynomial en n.

φ_3 : Cette formule code le fait que $\mathscr{M}$ accepte w, *i.e* qu'au moins une des cases de la dernière ligne contient un état final :

$$\varphi_3 = \bigvee_{q \in F} \left(\bigvee_{0 \leqslant j \leqslant n^k} x_{n^k, j, q} \right).$$

□

2.6 Algorithme de Dijkstra

Cormen *et al.* (n.d.)

L'algorithme de Dijkstra est un algorithme de calcul de plus court chemin à origine unique dans un graphe orienté pondéré $G = (S, A)$ où les poids des arcs sont positifs ou nuls.

On se fixe un sommet de départ $s \in S$.

On note pour chaque sommet v du graphe $d[v]$ un majorant de la distance de s à v, et $\pi[v]$ le prédecesseur de v dans le chemin de s à v considéré.

On initialise d à ∞ pour $v \neq s$ et $d[s] = 0$, et π à Nil.

On appelle *relâchement* d'un arc (u, v) la mise à jour de $d[v]$ et $\pi[v]$ si passer par u diminue la valeur de $d[v]$.

Algorithm 4 Relâcher

Préconditions: (u, v) arc
  si $d[v] > d[u] + w(u, v)$ **alors**
    $d[v] \leftarrow d[u] + w(u, v)$
    $\pi[v] \leftarrow u$
  is

On donne l'algorithme de Dijkstra :

Algorithm 5 Dijkstra

Préconditions: G graphe, w fonction de poids, s sommet de départ
  Initialiser d et π
  $E \leftarrow \emptyset$
  $F \leftarrow S[G]$
  tant que $F \neq \emptyset$ **faire**
    $u \leftarrow$Extraire-Min(F)
    $E \leftarrow E \cup \{u\}$
    pour $v \in Adj[u]$ **faire**
      Relâcher(u, v, w)
    ruop
  euq tnat

Théorème 2.16
En exécutant Dijkstra sur un graphe G, pour tous les sommets u, on a $d[u]$ est la longueur du chemin de longueur minimal entre s et u.

Démonstration. On utilise l'invariant de boucle

IdB =« Au début de chaque itération de la boucle **tant que**, $d[v]$ est la longueur du plus court chemin de s à v pour tout $v \in E$. »

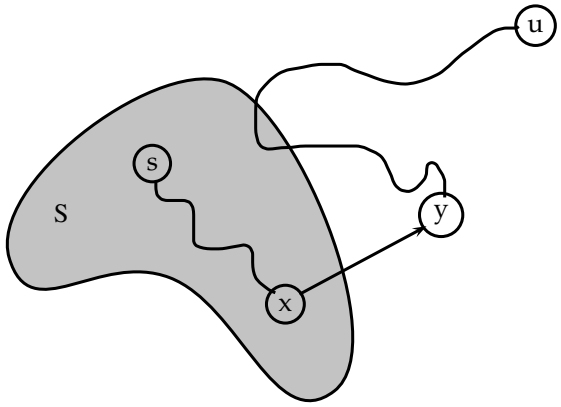
IdB est clairement vérifié au début de l'algorithme, car E est vide.

Par l'absurde, supposons que u est le premier sommet ajouté à E vérifiant $d[u] \neq \delta(s, u)$.

u est nécessairement différent de s (car $d[s] = \delta(s, s)$ à tout moment), et il existe un chemin de s à u (sinon $d[u] = \infty = \delta(s, u)$), soit p le plus court.

On se place juste avant avoir ajouté u à E. On a donc $u \in S \backslash E$. Soit y le premier sommet de p dans $S \backslash E$, et soit $x \in E$ son prédécesseur.

On est donc dans la situation suivante



On a $x \in E$, et donc par hypothèse sur u, on a $d[x] = \delta(s, x)$ quand x est ajouté à E, et comme on a relâché (x, y) à ce moment, on a alors $d[y] = \delta(s, y)$.

Comme y est avant u dans p, il est clair que $d[y] = \delta(s, y) \leqslant \delta(s, u)$.

Or $d[u] \geqslant \delta(s, u)$, on en tire donc $d[y] \leqslant d[u]$.

Mais les sommets y et u sont dans $S \backslash E$, et donc $d[u] \leqslant d[y]$.

On a donc $d[y] = \delta(s, y) = \delta(s, u) = d[u]$, ce qui est une contradiction.

L'algorithme termine bien car $|F|$ diminue strictement à chaque étape.

À la fin, on a $E = S$, donc $d[u] = \delta(s, u)$ pour tous les sommets du graphe.

□

L'algorithme de Dijkstra est donc correct et terminant.

Regardons sa complexité.

On utilise trois opérations sur les files de priorité : Insérer, Extraire-Min et Diminuer-Clé. Regardons tout d'abord la complexité en fonction de ces opérations.

- Insérer est appelée une fois par sommet.
- Extraire-Min est appelée aussi une fois par sommet.
- Pour chaque sommet v, on regarde exactement une fois chaque arc de la liste $Adj[v]$. Au pire, on aura donc $|A|$ opérations Diminuer-Clé.

La complexité de l'algorithme est donc $\mathcal{O}(i|S| + e|S| + d|A|)$ en fonction de ces trois opérations.

Maintenant, chacune de ces opérations demandera une complexité dépendant de la structure de donnée utilisée pour implémenter la file de priorité.

Si on numérote les sommets de 1 à $|S|$ et qu'on gère un tableau, les opérations Insérer et Diminuer-Clé prennent un temps (1), et Extraire-Min prend un temps $\mathcal{O}(|S|)$. On a donc une complexité finale de $\mathcal{O}(|S|^2 + |A|) = \mathcal{O}(|S|^2)$.

Si le graphe est peu dense (par exemple $|A| = o(|S|^2 / \lg(|S|))$), on peut implément la file de priorité par un tas min binaire.

– Extraire-Min prend alors un temps $\mathcal{O}(\lg(|S|))$
– Diminuer-Clé prend un temps $\mathcal{O}(\lg(|S|))$
– La création du tas prend $\mathcal{O}(S)$

On a donc une complexité de $\mathcal{O}(|A| \lg |S|)$.

2.7 Hachage parfait

Cormen *et al.* (n.d.)

On considère que l'ensemble des clefs est fixé.

L'idée est ici d'utiliser des tables de hachage à deux niveaux :
- Le premier niveau est identique au hachage par chainage : on hache n clefs dans m cases grâce à une fonction de hachage h bien choisie dans une famille de fonctions de hachage universelles.
- Plutôt que de créer une liste chaînée des éléments hachés en case j, on utilise une seconde table de hachage S_j associée à une fonction de hachage h_j.

En choisissant bien h_j, on peut assurer qu'il n'y aura pas de collision dans ce second niveau.

Pour cela, on supposera que la taille m_j de la table S_j est n_j^2, où n_j est le nombre de clefs hachées en case j.

On choisit h dans $\mathscr{H}_{p,m}$ (où p est une nombre premier plus grand que n'importe quelle clef), et $h_j \in \mathscr{H}_{p,m_j}$.

Montrons d'abord qu'il n'y a pas de collision au second niveau :

Théorème 2.17
Supposons qu'on stocke n clefs dans une table de hachage de taille $m = n^2$, par une fonction h choisie aléatoirement dans une classe de fonctions de hachage universelles.

Alors la probabilité qu'il y ait une ou plusieurs collisions est moins de $\frac{1}{2}$.

Démonstration. On a C_n^2 paires de clefs qui peuvent entrer en collision, et chaque paire entre en collision avec une probabilité $\frac{1}{m}$.

Soit X la variable aléatoire qui compte le nombre de collisions. On a :

$$\begin{aligned}\mathbb{E}X &= C_n^2 \cdot \frac{1}{n^2} \\ &= \frac{n^2 - n}{2} \cdot \frac{1}{n^2} \\ &< \frac{1}{2}\end{aligned}$$

On a alors par inégalité de Markov :

$$\mathbb{P}(\{X \geqslant 1\}) \leqslant \mathbb{E}X < \frac{1}{2}.$$

□

En choisissant au hasard nos fonctions de hachage, on a donc de grandes chances de tomber rapidement sur des fonctions sans collision.

Regardons maintenant l'espace occupé par cette double table.

Théorème 2.18
Supposons qu'on stocke n clefs dans une table de hachage de taille $m = n$ par une fonction de hachage choisie aléatoirement dans une classe de fonctions de hachage universelles. Alors :

$$\mathbb{E}\left(\sum_{j=0}^{m-1} N_j^2\right) < 2n,$$

où N_j est la variable aléatoire qui compte le nombre de clefs hachées en case j.

Démonstration. On utilise l'identité $a^2 = a + 2C_a^2$.

$$\begin{aligned}
\mathbb{E}\left(\sum_{j=0}^{m-1} N_j^2\right) &= \mathbb{E}\left(\sum_{j=0}^{m-1} N_j + 2C_{N_j}^2\right) \\
&= \mathbb{E}\left(\sum_{j=0}^{m-1} N_j\right) + 2\mathbb{E}\left(\sum_{j=0}^{m-1} C_{N_j}^2\right) \\
&= \mathbb{E}n + 2\mathbb{E}\left(\sum_{j=0}^{m-1} C_{N_j}^2\right) \\
&= n + 2\mathbb{E}\left(\sum_{j=0}^{m-1} C_{N_j}^2\right)
\end{aligned}$$

La quantité $\mathbb{E}\left(\sum_{j=0}^{m-1} C_{N_j}^2\right)$ représente le nombre total de paires de clefs qui entrent en collision. Par propriété du hachage universel, la somme vaut au plus :

$$\begin{aligned}
C_n^2 \frac{1}{m} &= \frac{n(n-1)}{2m} \\
&= \frac{n-1}{2}
\end{aligned}$$

Donc on a

$$\begin{aligned}
\mathbb{E}\left(\sum_{j=0}^{m-1} N_j^2\right) &\leqslant n + 2\frac{n-1}{2} \\
&= 2n - 1 < 2n
\end{aligned}$$

□

Corollaire 2.19
Supposons qu'on stocke n clefs dans une table de hachage de taille $m = n$ par une fonction de hachage choisie aléatoirement dans une classe de fonctions de hachage universelles.

On choisit les tailles des tables secondaires comme $m_j = n_j^2$.

Alors l'espace mémoire utilisé pour toutes les tables secondaires dans le hachage parfait est en moyenne inférieur à $2n$.

Démonstration. $m_j = n_j^2$, donc c'est immédiat. □

Corollaire 2.20
Supposons qu'on stocke n clefs dans une table de hachage de taille $m = n$ par une fonction de hachage choisie aléatoirement dans une classe de fonctions de hachage universelles.

On choisit les tailles des tables secondaires comme $m_j = n_j^2$.

Alors, la probabilité que l'espace mémoire utilisé pour les tables secondaires soit supérieur ou égal à $4n$ est moins de $\frac{1}{2}$.

Démonstration. On applique l'inégalité de Markov au corollaire précédent :

$$\begin{aligned}\mathbb{P}\left(\left\{\sum_{j=0}^{m-1} m_j \geqslant 4n\right\}\right) &\leqslant \frac{\mathbb{E}\left(\sum_{j=0}^{m-1} m_j\right)}{4n} \\ &< \frac{2n}{4n} \\ &= \frac{1}{2}\end{aligned}$$

□

Donc, en testant des fonctions de hachage aléatoirement dans une famille universelle, on en trouvera rapidement une qui utilise un espace de stockage raisonnable.

2.8 Theorème de Higman

Petazzoni (n.d.)

On définit la relation $<$ sur les mots par : si $u = u_1 \dots u_n$ et $v = v_1 \dots v_m$, on a $u < v$ *si et seulement si* il existe une application strictement croissante f de $[\![1, n]\!]$ sur $[\![1, m]\!]$ telle que $u_i = v_{f(i)}$.

Pour un mot w et un langage L, on notera L_w l'ensemble des sur-mots de w, *i.e*

$$L_w = \{v \in \Sigma^* \mid w < v\}.$$

Pour un langage L, on notera $\hat{L}$ l'ensemble des sur-mots des mots de L, et $SM(L)$ l'ensemble des sous-mots des mots de L.

Enfin, on appellera *antichaîne* sur Σ un langage L dont les mots ne sont pas comparables deux à deux pour $<$.

Théorème 2.21 : du sous-mot
Si $L \in \mathfrak{p}(\Sigma^*)$, alors $SM(L)$ est rationnel.

Démonstration. La démonstration repose sur le :

Lemme 2.22 : de Higman
Toute antichaîne sur Σ pour $<$ est finie.

On se fixe un langage L quelconque.

Lemme 2.23
$\hat{L}$ est rationnel.

Démonstration. On considère $L_\downarrow$ l'ensemble des mots de L minimaux pour $<$. Alors nécessairement, $L_\downarrow$ est une antichaîne, donc est finie.

Pour tout mot u de $\hat{L}$, il existe un mot $w \in L_\downarrow$ tel que $w < u$, et donc finalement :

$$\hat{L} = \bigcup_{w \in L_\downarrow} L_w.$$

Chaque L_w est rationnel ($L_w = \Sigma^* w_1 \Sigma^* \cdots \Sigma^* w_n \Sigma^*$), et donc par stabilité des langages rationnels par union finie, on a $\hat{L}$ rationnel. ◊

Maintenant, considérons $K = \Sigma^* \backslash SM(L)$. On va montrer que $K = \hat{K}$, et on aura alors la rationnalité de K, et donc de $SM(L)$ par clotûre des langages rationnels par complémentarisation.

L'inclusion $K \subseteq \hat{K}$ est triviale.

Soit donc $x \in \hat{K}$. Il existe $w \in K$ tel que $w < x$. Alors si $x \in SM(L)$, alors w aussi, ce qui est une contradiction.

D'où le résultat. □

Montrons le lemme de Higman :

Démonstration. On suppose que A est une antichaîne infinie sur Σ. On note alors

$$q = |\Sigma \text{ et } n = \min_{u \in A} |u|.$$

On suppose de plus que q et n sont minimaux : les antichaînes sur un alphabet plus petit que Σ sont finies, et celles dont le mot de longueur minimale est plus court que n aussi.

1. On remarque que $q > 1$ et $n > 1$

   Démonstration. Si $q = 1$, alors toute antichaîne ne contient qu'un élément.
   Si $n = 1$, alors u de longueur 1 est une lettre, et donc $A \backslash \{u\}$ antichaîne infinie sur $\Sigma \backslash \{u\}$ de taille $q - 1 \leftarrow$ contradiction avec la minimalité de q. ◊

2. On fixe u de longueur n. On définit A' comme l'ensemble des mots de A qui sont sur-mots de $u[1, n-1]$.
   On a alors $A \backslash A'$ fini, donc A' infini.

   Démonstration. Le langage $(A \backslash A') \cup \{u[1, n-1]\}$ est une antichaîne sur Σ, avec un mot minimal de longueur $n - 1$. Par minimalité de n, cette antichaîne est finie. ◊

3. On énumère A' : $A' = \{v_1, v_2, \ldots, v_i \ldots\}$, et on écrit chaque v_i comme

   $$v_i = z_{i,1} u_1 z_{i,2} u_2 \cdots z_{i,n-1} u_{n-1} z_{i,n},$$

   où pour $j \leqslant n - 1$, $z_{i,j} \in (\Sigma \backslash \{u_j\})^*$.
   Alors $z_{i,n} \in (\Sigma \backslash \{u_n\})^*$.

   Démonstration. Sinon, on retrouve u dans v_i, et donc $u < v_i$ ce qui est impossible par définition d'antichaîne. ◊

4. On définit $Z_j = \{z_{i,j} \mid i \in \mathbb{N}\}$.
   Alors Z_j a un nombre fini d'éléments minimaux pour $<$.

   Démonstration. L'ensemble des éléments minimaux de Z_j est une antichaîne, sur un alphabet de cardinale plus petit strictement que q, donc est fini. ◊

5. Si Z_j est infini, alors à extraction près, on a pour tout j :

   $$z_{i,j} < z_{i+1,j}.$$

   Démonstration. Soit $s(1)$ le plus petit indice tel qu'il n'y ait plus d'éléments minimaux : $z_{s(1),j}$ n'est pas minimal, et donc $\exists s(2) > s(1)$ tel que $z_{s(1),j} < z_{s(2),j}$.
   En itérant le procédé, on construit la fonction s sur $\mathbb{N}$. ◊

6. L'un au moins des Z_j est infini.

   Démonstration. On a $A' = Z_1 u_1 Z_2 u_2 \cdots Z_{n-1} u_{n-1} Z_n$, et A' infini. ◊

7. En applicant les extractions correspondant aux Z_j infinis, on va se retrouver avec des mots tous comparables entre eux. D'où une contradiction avec le fait que A' est une antichaîne. D'où le résultat.

□

2.9 Langage de pile d'un automate à pile

Carton (n.d.)

Théorème 2.24
On considère un automate à pile $\mathscr{A} = (\Sigma, \Gamma, Q, \delta, \gamma_0, F)$.

Soit $H = \{\gamma \in \Gamma^* \mid \exists f \in \Sigma^*,\ \exists q \in Q,\ (q_0, \gamma_0) \xrightarrow{f} (q, \gamma)\}$ *le langage de pile de* $\mathscr{A}$.

Alors H *est rationnel.*

Démonstration. On pose $A_n = \{a_1, \dots, a_n, \overline{a_1}, \dots, \overline{a_n}\}$ et on définit une relation $\to$ sur A_n^* par :

$$w \to w' \Leftrightarrow \exists i,\ \exists u, v \in A_n^*,\ w = u a_i \overline{a_i} v \text{ et } w' = uv.$$

On remarque que la relation $\to$ est noetherienne (il n'y a pas de chaîne infinie, car la longueur des mots est strictement décroissante), et confluente (si $x \to^* y$ et $x \to^* y'$, alors il existe z tel que $y \to^* z$ et $y' \to^* z$).

Donc pour tout w, il existe un unique mot $\rho(w)$ minimal pour $\to$.

On a alors le lemme :

Lemme 2.25

$$\{w \mid \rho(w) = \varepsilon\} = D_n^*.$$

Démonstration. "$\subseteq$" : Soit w tel que $\rho(w) = \varepsilon$, et soit k tel que $w \to^k \varepsilon$.
Si $k = 0$, alors $w = \varepsilon$ et donc $w \in D_n^*$.
Soit $k \in \mathbb{N}$ tel que pour tout w, $w \to^k \varepsilon \Rightarrow w \in D_n^*$. Soit w tel que $w \to^{k+1} \varepsilon$. Alors il existe w' de la forme uv de taille $2k$ tel que $w' \to^k \varepsilon$, et tel que $w = u a_i \overline{a_i} v$. Alors $w \in D_n^*$.
"$\supseteq$" : Évident par définition de $\to$ (une récurrence sur la longueur convaincra les sceptiques).

$\Diamond$

On définit maintenant la substitution σ sur A_n^* par $\sigma(a) = D_n^* a D_n^*$.

On a alors le lemme :

Lemme 2.26
Pour tout $w \in A_n^*$,

$$\sigma(w) = \{w' \mid w' \to^* w\}.$$

Démonstration. "$\subseteq$" : Soit $w = w_1 \dots w_k$. Alors $\sigma(w) = D_n^* w_1 D_n^* \dots D_n^* w_k D_n *$.
Soit $w' \in \sigma(w)$. Alors, par le lemme précédent, $w' \to^* w$.
"$\supseteq$" : Par récurrence sur la longueur de la réduction $w' \to^p w$:
Le cas $p = 0$ est trivial.
Soit p tel que tout mot w' vérifiant $w' \to^p w$ soit dans $\sigma(w)$. Soit w' tel que $w' \to^{p+1} w$.
Alors il existe $\hat{w}$ vérifiant
- $w' \to \hat{w}$
- $\hat{w} \to^p w$
- $\hat{w} = uv$

– $w' = ua_i\overline{a_i}v$
Par hypothèse de récurrence, $\hat{w}$ est de la forme $d_1w_1 \dots d_kw_kd_{k+1}$, avec les d_i dans D_n^*.
Alors u est de la forme $d_1w_1 \dots d_j$ (resp. $d_1w_1 \dots d_jw_j$) et v de la forme $w_j \dots w_kd_{k+1}$ (resp. $d_{j+1} \dots w_kd_{k+1}$).
w' est donc de la forme
$d_1w_1 \dots d_ja_i\overline{a_i}w_j \dots w_kd_{k+1}$ (resp. $d_1w_1 \dots d_jw_ja_i\overline{a_i}d_{j+1} \dots w_kd_{k+1}$), ce qui conclut la récurrence car $d_ja_i\overline{a_i}$ (resp. $a_i\overline{a_i}d_{j+1}$) est dans D_n^*.

◊

Lemme 2.27
Pour tout rationnel K sur A_n^*, le langage $\rho(K) = \{\rho(w) \mid w \in K\}$ est rationnel.

Démonstration. Un mot w est dans $\rho(K)$ si et seulement si il est irréductible pour $\rightarrow$, et si il existe $w' \in K$, tel que $w' \rightarrow^* w$.

Donc

$$\rho(K) = \sigma^{-1}(K) \setminus \left(\sum_{i=1}^{n} A_n^* a_i \overline{a_i} A_n^* \right).$$

Par stabilité des langages rationnels, $\rho(K)$ est rationnel. ◊

On va maintenant montrer que le miroir $\tilde{H}$ de H est rationnel, ce qui impliquera que H est rationnel.

On suppose que dans l'automate $\mathscr{A}$, l'alphabet de pile est $\Gamma = \{a_1, \dots, a_n\}$.

Pour une transition $\tau = q, \gamma \xrightarrow{y} q', h$ dans $\mathscr{A}$, on définit $\mu(\tau) = \overline{\gamma}\tilde{h}$, que l'on étend à δ^*.

On dit que deux transitions τ et τ' sont *consécutives* si l'état d'arrivée de τ est le même que l'état de départ de τ'.

On appelle K l'ensemble des suites de transitions de $\mathscr{A}$ consécutives.

K est clairement rationnel, car reconnu par l'automate $\mathscr{A}$ auquel on enlève la pile.

On a alors

$$\tilde{H} = \rho(\gamma_0\mu(K)) \cap \Gamma^*.$$

Par le lemme précédent, et stabilité des langages rationnels, $\tilde{H}$ est rationnel.

□

Références : Carton, Langages formels, calculabilité et complexité.

Leçons : 909, 911

2.10 Plus longue sous-séquence commune

Cormen *et al.* (n.d.)

Étant donnés deux mots $X = x_1 \cdots x_m$ et $Y = y_1 \cdots y_n$ sur un alphabet Σ, on cherche à savoir quel est le plus long sous-mot commun à X et Y.

La solution naïve est d'énumérer tous les sous-mots de X, et de regarder ensuite s'ils sont aussi sous-mots de Y, mais cette solution est inefficace : il y a 2^m sous-mots de X.

NOTATION – Si U est un mot de longueur p, on note U_{p-j} le préfixe de U de longueur $p - j$.

Un algorithme plus efficace pourra être déduit du théorème de structure suivant :

Théorème 2.28 : de structure des plus longues sous-séquences communes
Supposons que $Z = z_1 \cdots z_k$ est une plus longue sous-séquence commune à X et Y. Alors :

(i) Si $x_m = y_n$, alors $z_k = x_m (= y_n)$ et Z_{k-1} est une plus longue sous-séquence commune à X_{m-1} et Y_{n-1}.

(ii) Si $x_m \neq y_n$ alors

$$(z_k \neq x_m) \Rightarrow Z \text{ est une plus longue sous-séquence commune à } X_{m-1} \text{ et } Y.$$

(iii) Si $x_m \neq y_n$ alors

$$(z_k \neq y_n) \Rightarrow Z \text{ est une plus longue sous-séquence commune à } X \text{ et } Y_{n-1}.$$

Démonstration. C'est plutôt facile :

(i) $x_m = y_n$. Si $z_k \neq x_m$, alors Zx_m est un sous-mot commun à X et Y, ce qui contredit l'hypothèse de maximalité. D'où $z_k = x_m = y_n$.
De plus, Z_{k-1} est un sous-mot commun à X_{n-1} et Y_{m-1}. Supposons qu'il existe un sous-mot commun W de longueur plus grande que $k - 1$.
Alors, Wx_m est de longueur plus grande que k, et Wx_m est un sous-mot commun à X et Y, d'où une contradiction.
Donc Z_{k-1} est une plus longue sous-séquence commune à X_{m-1} et Y_{n-1}.

(ii) $x_m \neq y_n$. Supposons $z_k \neq x_m$. Alors Z est un sous-mot de X_{m-1}, et donc *a fortiori un sous-mot commun à X_{m-1} et Y.*
Supposons qu'il existe un sous-mot W commun à X_{m-1} et Y, de longueur plus grande. Alors W est a fortiori un sous-mot commun à X et Y, ce qui contredit l'hypothèse de maximalité de Z.

(iii) *cf* (ii)

□

Connaissant la plus longue sous-séquence commune au mot vide et n'importe quel mot, on peut en déduire une formule de récurrence sur la longueur de la plus longue sous-séquence commune de X et Y. Si on note $c[i, j]$ la longueur de la plus longue sous-séquence commune à X_i et Y_j, on a :

$$c[i,j] = \begin{cases} 0 & \text{si } i = 0 \text{ ou } j = 0 \\ c[i-1, j-1] + 1 & \text{si } i, j > 0 \text{ et } x_i = y_j \\ \max(c[i, j-1], c[i-1, j]) & \text{si } i, j > 0 \text{ et } x_i \neq y_j \end{cases} \tag{2}$$

Cette formule pourrait nous donner un algorithme récursif pour calculer $c[m, n]$, mais sa complexité serait à nouveau exponentielle. On va plutôt utiliser un algorithme de *programmation dynamique*.

Algorithm 6 Longueur-PLSC

Préconditions: X, Y mots sur Σ

```
m ← Longueur(X)
n ← Longueur(Y)
pour i = 1 to m faire
        c[i, 0] ← 0
ruop
pour j = 0 to n faire
        c[0, j] ← 0
ruop
pour i = 1 to m faire
        pour j = 1 to n faire
                si x_i = y_j alors
                        c[i, j] ← c[i − 1, j − 1] + 1
                        b[i, j] ← " ↖ "
                sinon si c[i − 1, j] ⩾ c[i, j − 1] alors
                        c[i, j] ← c[i − 1, j]
                        b[i, j] ← " ↑ "
                sinon
                        c[i, j] ← c[i, j − 1]
                        b[i, j] ← " ← "
                is
        ruop
ruop

retourner c et b
```

Cet algorithme va calculer les uns après les autres les valeurs des $c[i, j]$, en utilisant la formule de récurrence, en remplissant le tableau par lignes.

La matrice b sert à noter "d'où on vient" pour calculer la valeur $c[i, j]$ d'une case. Grâce à ce tableau, on peut donc retrouver la plus longue sous-séquence commune à X et Y par l'algorithme :

Algorithm 7 PLSC

Préconditions: X mot sur Σ, b matrice calculée par Longueur-PLSC, i, j indices.

```
si i = 0 ou j = 0 alors

        afficher ()
is
si b[i, j] = " ↖ " alors
        PLSC(b, X, i − 1, j − 1)

        afficher x_i
sinon si b[i, j] = " ↑ " alors
        PLSC(b, X, i − 1, j)
sinon
        PLSC(b, X, i, j − 1)
is
```

Théorème 2.29

La procédure Longueur-PLSC a une complexité en $\mathcal{O}(mn)$, et la procédure PLSC a une complexité en $\mathcal{O}(m+n)$.

Démonstration. Dans Longueur-PLSC, on a deux boucles "pour" imbriquées, et dans ces boucles des opérations en $\mathcal{O}(1)$. D'où la complexité.

Dans PLSC, on réalise un "chemin direct" dans notre matrice $(i+1) \times (j+1)$, et donc ce chemin ne peut avoir plus de $i+j$ étapes. □

Exemple – On cherche la plus longue sous-séquence commune aux chaînes `BDCABA` et `ABCBDAB`. Le tableau construit par Longueur-PLSC est :

	j	0	1	2	3	4	5	6
i		y_j	B	D	C	A	B	A
0	x_i	0	0	0	0	0	0	0
1	A	0	0 ↑	0 ↑	0 ↑	1 ↖	1 ←	1 ↖
2	B	0	1 ↖	1 ←	1 ←	1 ↑	2 ↖	2 ←
3	C	0	1 ↑	1 ↑	2 ↖	2 ←	2 ↑	2 ↑
4	B	0	1 ↖	1 ↑	2 ↑	2 ↑	3 ↖	3 ←
5	D	0	1 ↑	2 ↖	2 ↑	2 ↑	3 ↑	3 ↑
6	A	0	1 ↑	2 ↑	2 ↑	3 ↖	3 ↑	4 ↖
7	B	0	1 ↖	2 ↑	2 ↑	3 ↑	4 ↖	4 ↑

On suit les flèches (chemin grisé) et on note les lettres correspondant aux flèches obliques : `BCBA`.

2.11 Complétude de la méthode de résolution

Stern (n.d.)

Rappelons la méthode de résolution pour un ensemble $\mathfrak{F}$ de formules closes :

1. on met chaque formule sous forme prénexe ;
2. on skolemise ces formules ;
3. on distribue les quantificateurs.

On a maintenant un ensemble de clauses, qu'on note $\{C_1, \dots, C_r\}$. On renomme les variables de chaque clauses pour que

$$\forall i \neq j,\ Var(C_i) \cap Var(C_j) = \emptyset.$$

On cherche à appliquer la règle de résolution :

Définition 2.30 : règle de résolution
Soient C, C_1, C_2 trois clauses. On dit que C est une résolvante *de C_1 et C_2 s'il existe $S_1 \subset C_1$ et $S_2 \subset C_2$ deux ensembles de littéraux tels que :*
– *S_1 et $\neg S_2$ sont unifiables par σ unificateur principal ;*
– *$C = ((C_1 \backslash S_1) \cup (C_2 \backslash S_2))\sigma$.*

Un *arbre de résolution* est un arbre dont les feuilles sont étiquetées par des clauses de Σ, et chaque nœud a deux fils dont il est une résolvante.

Si la racine de l'arbre est la clause vide $\square$, alors on dit que l'arbre est un arbre de *réfutation*.

On se donne un langage L qui a au moins un symbole de constante. On rappelle la définition d'un modèle de Herbrand (Jacques Herbrand, 1908 - 1931) :

Définition 2.31
On appelle modèle de Herbrand $\mathfrak{H}$ *une réalisation de L de domaine H qui a les propriétés suivantes :*
– *H est l'ensemble de tous les termes clos de L ;*
– *chaque constante est interprétée par elle-même ;*
– *l'interprétation d'une fonction f d'arité n est la fonction $t_1, \dots, t_n \mapsto f(t_1, \dots, t_n)$, où les t_i sont des termes ;*
– *à chaque formule atomique close $R(t_1, \dots, t_n)$ on associe une* variable de Herbrand, *$p[R(t_1, \dots, t_n)]$.*
On associe alors à la distribution de vérité σ le modèle de Herbrand $\mathfrak{H}(\sigma)$ en interprétant R par

$$R^{\mathfrak{H}(\sigma)} = \{(t_1, \dots, t_n) \mid \sigma(p[R(t_1, \dots, t_n)]) = 1\}.$$

Définition 2.32
À toute formule du calcul propositionnel F sur les variables de Herbrand, on associe une formule close sans quantificateur $\Phi(F)$ définie par induction, en partant de

$$\Phi(p[R(t_1, \dots, t_n)]) = R(t_1, \dots, t_n).$$

On a alors clairement $Val(F, \sigma) = Val(\Phi(F), \mathfrak{A}(\sigma))$.

On définit maintenant une particularisation :

Définition 2.33

Soit F une formule close. On appelle particularisation *de F toute formule qui s'écrit $F(x_1/t_1, \ldots, x_n/t_n)$ où les t_i sont des termes clos.*

On a alors le :

Théorème 2.34 : de Herbrand
Soit Σ un ensemble de formules closes universelles. Alors l'une ou l'autre des propositions est vraie :
– *il existe un modèle de Herbrand qui satisfait Σ ;*
– *il existe un ensemble fini de particularisations de formules de Σ qui est contradictoire.*

Démonstration. On remplace chaque formule de Σ par l'ensemble de ses particularisations. On obtient ainsi un ensemble Σ' de formules closes. Par Φ, cet ensemble provient d'un ensemble de formules du calcul propositionnel :

$$\Sigma' = \{\Phi(F) \mid F \in \Sigma_0\}.$$

S'il existe σ qui satisfait Σ_0, alors le modèle de Herbrand $\mathfrak{A}(\sigma)$ satisfait toutes les formules de Σ' et donc de Σ.

Si Σ_0 est contradictoire, alors par compacité du calcul propositionnel, on a un sous-ensemble fini de Σ_0 qui est contradictoire, et par Φ, on a un sous-ensemble fini de Σ' qui est contradictoire. □

Théorème 2.35 : de complétude de la méthode de résolution
Soit Σ un ensemble de clauses contradictoire. Alors il existe un arbre de résolution qui réfute Σ.

Démonstration. Notons Σ' l'ensemble des particularisations des clauses de Σ.

Par le théorème de Herbrand, Σ' est contradictoire. Via l'application Φ, on a un ensemble de formules propositionnelles contradictoire.

On a donc un arbre de réfutation au sens propositionnel qui réfute Σ'. On va remplacer chaque nœud c de l'arbre par une clause C dont elle est une particularisation. Pour les feuilles, la construction est évidente. Pour les nœuds internes, on utilise le

Lemme 2.36 : de relèvement
Soient C_1 et C_2 des clauses au sens du calcul des prédicats, et soient c_1 et c_2 des clauses propsitionnelles qui en sont respectivement des particularisations. Alors pour toute résolvante c de c_1 et c_2, il existe une résolvante C de C_1 et C_2 donc c est une particularisation.

Démonstration. Quitte à renommer les variables, on peut supposer que les variables $x_1, \ldots, x_n$ de C_1 et $y_1, \ldots, y_p$ de C_2 sont distinctes.

c_1 et c_2 étant des particularisations, il existe des termes $t_1, \ldots, t_n, t'_1, \ldots, t'_p$ tels que c_1 et c_2 proviennent de C_1 et C_2 par la substitution

$$\tau = (x_1/t_1, \ldots, y_p/t'_p).$$

Soit u le littéral qui intervient dans la coupure de c_1 et c_2. On définit S_1 (resp. S_2) l'ensemble L des littéraux de C_1 (resp. C_2) tels que $L\tau$ conduise à u (resp. $\overline{u}$).

Alors $S_1 \cup \neg S_2$ est unifiable par τ. On choisit un unificateur principal σ : on a $\tau = \sigma\theta$.

c est obtenue par l'action de τ sur l'ensemble de littéraux

$$(C_1 \backslash S_1) \cup (C_2 \backslash S_2),$$

c'est-à-dire par l'action de θ sur la résolvante de C_1 et C_2

$$C = ((C_1 \backslash S_1) \cup (C_2 \backslash S_2))\sigma.$$

Comme θ élimine toutes les variables de C, c est la particularisation de C obtenue en remplaçant chaque variable x par $x\theta$. ◊

□

2.12 Problème de séparation par automate

Floyd & Biegel (n.d.)

Prérequis : *SAT est NP-complet.*

Alice et Bob jouent à un jeu :
– Alice pense à un langage L, et doit le faire deviner à Bob.
– Alice donne à Bob un ensemble de mots de L, S, et un ensemble de mot qui ne sont pas dans L, T.
– Bob doit trouver, à partir des exemples de S et des contre-exemples de T le langage L.
Le problème de séparation par automate consiste à chercher l'existence, pour des langages S et T finis et un entier k donnés, un automate fini déterministe à k états qui accepte tous les mots de S, et refuse tous les mots de T.

Théorème 2.37
Le problème de séparation par automate (PSA) est NP-complet.

Démonstration. Tout d'abord, PSA est de classe NP. En effet, on peut en temps polynomial construire de façon non déterministe un automate, et le tester sur les mots de S et T.

Montrons donc que PSA est NP-dur. Nous allons pour cela réduire SAT à PSA : on suppose pouvoir résoudre PSA en temps polynomial, et on en déduit un algorothme de résolution de SAT en temps polynomial.

Soit donc F une formule du calcul propositionel sous forme normale conjonctive. On note $C_0, \dots, C_{m-1}$ les clauses de F, et $x_0, \dots, x_{v-1}$ les variables de F.

Dans la suite on notera $\ell_0, \dots, \ell_{2v-1}$ les littéraux $x_0, \dots, x_{v-1}, \overline{x_0}, \dots, \overline{x_{v-1}}$.

On pose aussi $k = 2v + m$.

On va construire deux ensembles S et T tels que si P est un automate fini déterministe qui sépare S et T, alors on peut déduire polynomialement de P une valuation qui valide F.

Notre automate va en fait choisir, pour chaque clause de F, un littéral qui doit être mis à VRAI, tout en s'assurant qu'un littéral et sa négation ne soit pas mis à VRAI en même temps.

(1) On pose
– $S_1 = \{\varepsilon, a^k\}$
– $T_1 = \{a_i \mid 0 < i < k\}$.
Comme P a k états, on les appelera dans l'ordre $C_0, \dots, C_{m-1}, \ell_0, \dots, \ell_{2v-1}$. C_0 est l'unique état initial, et l'unique état final.

(2) On veut que les arêtes issues d'un état C_i et étiquetée par b aillent vers un littéral de C_i. On pose alors :

$$T_2 = \{a^i b a^j \mid 0 \leqslant i < m,\ 0 \leqslant j < k\} \backslash \{a^i b a^{2v-j} \mid \ell_j \text{ est un littéral de } C_i\}.$$

Lemme 2.38
Si P rejette aussi les mots de T_2, alors les arêtes issues d'un état C_i étiquetées par b vont vers un littéral de C_i.

Démonstration. Supposons qu'une arête étiquetée par b aille de C_i à C_j.
Soit alors $u = a^i b a^{k-j}$. Alors u est accepté par P et appartient à T_2, ce qui est impossible.
Donc les arêtes partant de C_i étiquetée par b vont nécessairement vers un ℓ_j.
La lecture de $a^i b$ va de C_0 à ℓ_j, et la lecture ensuite de a^{2v-j} revient en C_0, et donc $a^i b a^{2v-j}$ est accepté par P, et donc ℓ_j est un littéral de C_i. ◊

(3) On veut que toutes les arêtes issues de ℓ_j étiquetées par b aillent soit vers C_0 (on dit alors que ℓ_j est véridique), soit vers $\overline{x_0}$.
On pose alors
$$T_3 = \{a^{m+j}ba^h \mid 0 \leqslant j < 2v,\ 0 < h < k,\ h \neq v\}.$$

Lemme 2.39
Un automate qui rejette aussi T_3 respecte la condition ci-dessus.

Démonstration. Soit P qui rejette aussi T_3. Supposons qu'il existe une arête étiquetée par b qui va de ℓ_j à un état q autre que C_0 ou $\overline{x_0}$.
Si $q = C_w,\ w \neq 0$, alors le mot $u = a^{m+j}b$ arrive en C_w, et donc le mot $a^{m+j}ba^{m-w+2v}$ est accepté $\rightarrow$ NON.
Si $q = \ell_w,\ w \neq v$, alors le mot $u = a^{m+j}b$ arrive en ℓ_w, et donc $a^{m+j}ba^{2v-w}$ est accepté $\rightarrow$ NON. $\Diamond$

(4) On veut interdire que les littéraux x_j et $\overline{x_j}$ soit véridiques en même temps. On pose donc
$$T_4 = \{a^{m+j}ba^{m+v+j}b \mid 0 \leqslant j < v\}.$$

Lemme 2.40
Si P refuse aussi T_4, alors P satisfait l'hypothèse ci-dessus.

Démonstration. Soit P qui refuse aussi T_4. Supposons que $x_j \xrightarrow{b} C_0$ et $\overline{x_j} \xrightarrow{b} C_0$.
Alors $a^{m+j}b$ arrive en C_0, et $a^{m+v+j}b$ aussi, et donc la concaténation des deux est acceptée $\rightarrow$ NON. $\Diamond$

(5) Enfin, on veut que l'arête étiquetée par b issue de C_i (voir point (2)) aille vers un littéral véridique. On pose :
$$S_5 = \{a^ibb \mid 0 \leqslant i < m\}.$$

On prend alors $S = S_1 \cup S_5$ et $T = T_1 \cup T_2 \cup T_3 \cup T_4$.

Si P à k états accepte S et rejette T, alors on construit la valuation :

$$\rho(x_i) = \begin{cases} \text{vrai} & \text{si } (x_j \xrightarrow{b} C_0) \text{ arête de } P \\ \text{faux} & \text{sinon} \end{cases}$$

Alors la valuation ρ satisfait F.

□

2.13 Arithmétique de Presburger

Carton (n.d.)

Définition 2.41
On dit qu'une théorie logique est décidable *s'il est décidable de savoir si une formule est vraie.*

Théorème 2.42
La théorie au premier ordre des entiers munis de l'addition est décidable.

Remarque – La théorie T au premier ordre des entiers munis de l'addition est définie par la syntaxe :
– $(x = y) \in T$
– $(x + y = z) \in T$
– Si $\phi, \varphi \in T$ alors $\phi \vee \varphi \in T$, $\phi \wedge \varphi \in T$.
– Si $\phi \in T$ alors $\forall x\ \phi$, $\exists x\ \phi$, $\neg\phi$ sont dans T.

Démonstration. On se donne une formule φ sous forme prénexe :

$$\varphi = Q_1 x_1 Q_2 x_2 \cdots Q_n x_n,$$

où les Q_i sont des quantificateurs, et ψ formule sans quantificateur.

On pose pour tout k

$$\varphi_k = Q_{k+1} x_{k+1} \cdots Q_n x_n \psi,$$

avec $\varphi_0 = \varphi$ et $\varphi_n = \psi$.

On suppose de plus que $x_1, \ldots, x_k$ sont libres dans φ_k.

On va montrer par récurrence sur $n - k$ que l'ensemble des k-uplets qui satisfont φ_k est rationnel.

Les k-uplets d'entiers seront représentés comme des mots sur $\Sigma_k = \{0, 1\}^k$, quitte à rajouter des 0 en début de mot pour qu'ils soient tous de la même longueur.

Soit

$$X_k = \{(n_1, \ldots, n_k) \mid \varphi_k(n_1, \ldots, n_k) \text{ est vraie}\}.$$

On va construire par récurrence sur $n - k$ un automate $\mathscr{A}_k$ qui accepte les k-uplets de X_k. Ainsi, la vérité de φ est équivalente à la non-vacuité du langage de $\mathscr{A}_0$.

Commençons par définir $\mathscr{A}_n$: il doit accepter les n-uplets qui satisfont ψ, qui est composée de combinaisons booléennes de formules du type $x_i = x_j$ ou $x_i + x_j = x_k$. Par clôture des langages rationnels par opérations booléennes, il suffit de construire un automate pour chacune des formules.

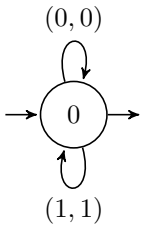


Figure 2 – Automate de l'égalité

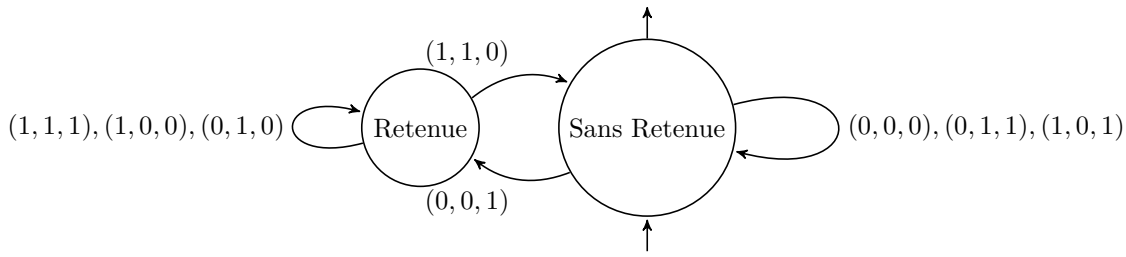


FIGURE 3 – Automate de l'addition

Pour l'hérédité, on écrit $\varphi_k = Q_{k+1}x_{k+1}\varphi_{k+1}$, et on traite les cas où Q_{k+1} est un quantificateur universel ou existentiel.

Si c'est $\forall$, alors on peut se ramener au second cas en écrivant $\varphi_k = \neg\exists x_{k+1}\neg\varphi_{k+1}$ et conclure par clôture par complémentation.

Si c'est $\exists$, alors on projette en oubliant la dernière composante Σ_{k+1} sur Σ_k par π_k. On construit alors $\mathscr{A}_k$ à partir de $\mathscr{A}_{k+1}$. Ils ont le même ensemble d'états, et si on a une transition $p \xrightarrow{x} q$ dans $\mathscr{A}_{k+1}$, alors on met la transition $p \xrightarrow{\pi_k(x)} q$ dans $\mathscr{A}_k$.

Les états finals de $\mathscr{A}_k$ sont les mêmes que ceux de $\mathscr{A}_{k+1}$, et les états initiaux sont ceux de $\mathscr{A}_{k+1}$ et ceux qui peuvent être atteint dans $\mathscr{A}_{k+1}$ en lisant $(0,\ldots,0) \in \Sigma_k$ (pour prendre en compte le fait qu'il faut rajouter de zéros au début des mots).

$\mathscr{A}_k$ va deviner le x_{k+1} qui convient de façon non déterministe. □

2.14 Exemple de programme PROLOG

Stern (n.d.)

Définition 2.43
On appelle clause de Horn *une clause dont* au plus *un littéral est positif.*

Si exactement *un littéral est positif, on parle de clause définie.*

Si aucun *littéral n'est positif, on parle de but.*

On a alors

Proposition 2.44
Soit Σ un ensemble de clauses de Horn, et soit Σ_0 le sous-ensemble formé des clauses définies. Alors :
– *Σ_0 n'est pas contradictoire ;*
– *si Σ est contradictoire, il existe un but G de Σ tel que $\Sigma_0 \cup G$ soit contradictoire.*

Définition 2.45
*On dit qu'un arbre de résolution est de type LD (*Linéaire et Défini*) s'il se compose d'une branche étiquetée de la feuille vers la racine par des buts $G_0, \ldots, G_m$ et de nœuds terminaux étiquetés par des clauses définies qui sont fils d'éléments de cette branche, $C_0, \ldots, C_{m-1}$.*

Si $C_0, \ldots, C_{m-1} \in \Sigma$ et $G_m = \square$, on dit que l'arbre réfute G_0 à partir de Σ.

Théorème 2.46
Soient Σ un ensemble de clauses définies et G un but. Si $\Sigma \cup \{G\}$ est contradictoire, il existe un arbre de réfutation de type LD qui réfute G à partir de Σ.

De plus, on peut remarquer que la recherche d'un arbre de réfutation LD donne explicitement un contre-exemple. C'est cette propriété qui est utilisée par PROLOG.

PROLOG construit en fait un arbre de réfutation LD par *effacement des buts.*

On opère sur des suites de littéraux positifs $Q_1 \ldots Q_n$, appelées *clauses ordonnées.*

S'il existe une variante d'une règle $L \to P_1 \ldots P_k$, où L s'unifie à Q_1 par l'unificateur principal σ, on dit que $P_1\sigma \ldots P_k\sigma Q_2\sigma \ldots Q_n\sigma$ est obtenue par *effacement* à partir de $Q_1 \ldots Q_n$.

Si la règle est un fait ($k = 0$), on obtient donc $Q_2\sigma \ldots Q_n\sigma$: on a effacé Q_1.

Le choix de la règle à unifier se fait de façon non déterministe. On contruit un *arbre d'effacement de la façon suivante* :

On note $R_1, \ldots, R_\ell$ les règles dans l'ordre où elles sont écrites, et $Q_1 \ldots Q_n$ le but.

Chaque nœud de l'arbre correspond au but courant, et a autant de fils que de règle qui peuvent s'unifier avec le premier but, dans l'ordre de l'indice.

Regardons maintenant le programme PROLOG suivant :

```
homme(jacques).
homme(julien).
```

```
homme(aymeric).
homme(françois).
homme(didier).
femme(brigitte).
femme(martine).
femme(vanessa).
parent(jacques,julien).
parent(jacques,aymeric).
parent(brigitte,julien).
parent(brigitte,aymeric).
parent(martine,françois).
parent(didier,vanessa).
soeur(martine,brigitte).
soeur(brigitte,martine).
soeur(martine,didier).
soeur(brigitte,didier).

fils(X,Y) :- parent(Y,X), homme(X).
fille(X,Y) :- parent(Y,X), femme(X).
frere(X,Y) :- soeur(Y,X), femme(X).
cousin(X,Y) :- fils(X,T), soeur(T,Z), parent(Z,Y).
cousin(X,Y) :- fils(X,T), frere(T,Z), parent(Z,Y).
cousine(X,Y) :- fille(X,T), frere(T,Z), parent(Z,Y).
cousine(X,Y) :- fille(X,T), soeur(T,Z), parent(Z,Y).
```

On pose la question (le but) à PROLOG :

```
cousin(françois,A).
```

PROLOG va donc construire l'arbre d'effacement.

On commence par chercher à unifier `cousin(françois,A)` avec une règle. Les deux règles possibles sont

```
cousin(X,Y) :- fils(X,T), soeur(T,Z), parent(Z,Y).}
cousin(X,Y) :- fils(X,T), frere(T,Z), parent(Z,Y).}
```

dans cet ordre là, avec l'unificateur `X` $\rightarrow$ `françois`.

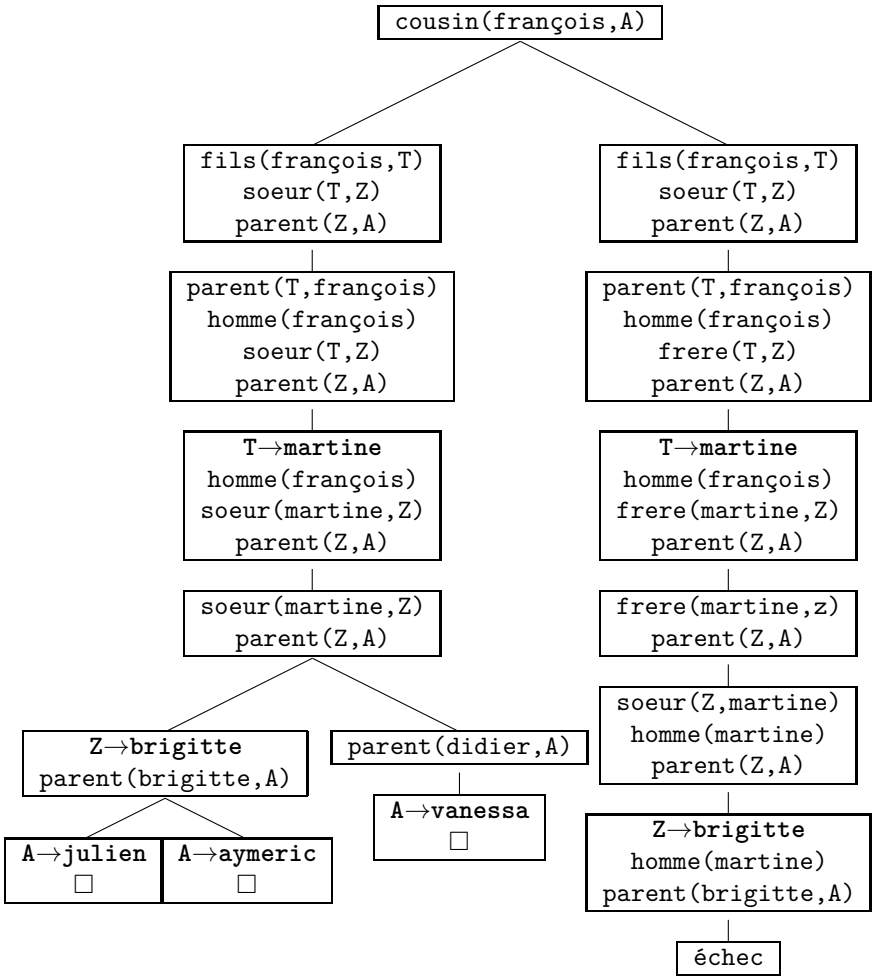
Le fils gauche de l'arbre est donc

```
fils(françois,T).
soeur(T,Z).
parent(Z,A).
```

et le fils droit

```
fils(françois,T).
frere(T,Z).
parent(Z,A).
```

On continue ainsi de suite pour obtenir l'arbre d'effacement :



2.15 Caractérisation des ensembles récursivement énumérables

Cori & Lascar (tome 2)

Théorème 2.47
Soit $A \subset \mathbb{N}$. Sont équivalentes :

(i) A est récursivement énumérable ;

(ii) A est l'image d'une fonction récursive primitive ou $A = \emptyset$;

(iii) A est l'image d'une fonction partielle récursive ;

(iv) A est la projection d'un sous-ensemble récursif primitif $B \subseteq \mathbb{N}^2$.

Démonstration. On fait une démonstration circulaire :
(i) → (ii) Soit $\psi : A \to \mathbb{N}$ partielle récursive.
On suppose $A \neq \emptyset$, et on prend $n \in A$.
On énumère les machines de Turing $(M_i)_i$, et on fixe i tel que M_i calcule ψ. On note de plus

$$B(i,t,x) = \text{« La machine } M_i \text{ s'arrête sur } x \text{ en } t \text{ étapes »}.$$

On a alors

$$x \in A \Leftrightarrow \exists t,\ B(i,t,x).$$

On énumère de façon primitive récursive (et avec une réciproque primitive récursive) $\mathbb{N}^2 : (t_k, x_k)_k$.
Alors $\{k \in \mathbb{N} \mid B(i,t_k,x_k)\}$ est primitif récursif.
On a alors $A = Im(f)$ où f est la fonction primitive récursive

$$f : \begin{array}{ccc} \mathbb{N} & \longrightarrow & A \\ k & \longmapsto & \begin{cases} x_k & \text{si } B(i,t_k,x_k) \\ n & \text{sinon} \end{cases} \end{array}$$

(ii) → (iii) Si A est l'image d'une fonction primitive récursive, A est *a fortiori* l'image d'une fonction μ-récursive partielle.
Si A est vide, alors $A = Im(\lambda y.\mu x(x+1=0))$.
(iii) → (iv) Soit f μ-récursive partielle telle que $A = Im(f)$.
Soit M_i une machine de Turing calculant f. On définit le prédicat

$$C(i,t,x,a) = \text{« La machine } M_i \text{ s'arrête en } t \text{ étapes sur } a \text{ en partant de } x \text{ »}.$$

On pose $G = \{(k,a) \mid C(i,t_k,x_k,a\}$ qui est primitif récursif.
A est alors la projection de G sur la seconde variable.
(iv) → (i) Soit B récursif primitif tel que A soit la projection de B :

$$A = \{x \in \mathbb{N} \mid \exists y \in \mathbb{N},\ (x,y) \in B\}.$$

On pose $f(x) = \mu y((x,y) \in B)$. f est partielle μ-récursive, et $x \in Dom(f)$ *si et seulement si* il existe y tel que $(x,y) \in B$ *si et seulement si* $x \in A$.
Donc A est récursivement énumérable.

□

2.16 Théorème de Rice

Wolper (n.d.)

On définit dans la suite deux langages :

Définition 2.48
On note (w_i) et (M_j) des énumérations des mots et des machines de Turing, et on pose
– $L_0 := \{w \mid w = w_i \text{ et } M_i \text{ n'accepte pas } w_i\}$.
– $LU := \{< M, w > \mid M \text{ accepte } w\}$
On note $\overline{L_0}$ et $\overline{LU}$ les complémentaires.

Théorème 2.49
Tout propriété non triviale des langages récursivement énumérables est indécidable.

Démonstration. On va montrer que les langages L_0, $\overline{L_0}$ et LU sont indécidables, puis, pour une propriété non triviale sur les langages récursivement énumérables P, on montrera que P est indécidable par une réduction à partir de LU.

Lemme 2.50
L_0 est indécidable.

Démonstration. Supposons L_0 décidable : il existe une machine de Turing qui l'accepte, soit M_k. On a alors :
– si M_k accepte w_k, alors $w_k \notin L_0$ par définition de $L_0 \rightarrow$ contradiction.
– si M_k n'accepte pas w_k, alors $w_k \in L_0 \rightarrow$ contradiction.

◊

Lemme 2.51
$\overline{L_0}$ est indécidable.

Démonstration. Si $\overline{L_0}$ était décidable, alors L_0 aussi. ◊

Lemme 2.52
LU est indécidable.

Démonstration. On fait une réduction à partir de $\overline{L_0}$.

Supposons donc LU décidable. Considérons l'algorithme suivant, prenant en entrée un mot w :
– on détermine i tel que $w = w_i$;
– on détermine M_i ;
– on applique la procédure de décision pour LU à $< M_i, w_i >$: si le résultat est positif, on accepte w, sinon on le rejette.
Alors cet algorithme décide $\overline{L_0} \rightarrow$ contradiction. ◊

Soit maintenant P une propriété non triviale sur les langages récursivement énumérables.

On peut supposer que le langage vide ne vérifie pas P (sinon, on considère $\overline{P}$).

Comme P est non triviale, il existe une machine de Turing M_p qui accepte un langage vérifiant P.

Pour une instance $< M, w >$ de LU, on construit une machine M' qui a le comportement suivant :
– M' simule l'exécution de M sur w, sans tenir compte du mot d'entrée x ;
– si M accepte w, elle simule M_p sur x ;
– si M n'accepte pas w (rejet ou exécution infinie), M' n'accepte aucun mot.
On a alors : $\mathscr{L}(M')$ vérifie P si et seulement si $< M, w > \in LU$. □

2.17 Comparaison tri fusion/tri rapide

Cormen *et al.* (n.d.) et Froidevaux *et al.* (n.d.)

Les algorithmes de tri par fusion et par tri rapide sont deux algorithmes utilisant le paradigme *diviser pour régner*. Nous allons voir les différences entre ces deux algorithmes.

2.17.1 Tri fusion

Le schéma de l'algorithme est :
- *Diviser* la séquence de n éléments en deux séquences de taille $E(n/2)$;
- *Appliquer récursivement* le tri fusion sur les deux sous-séquences ;
- *Fusionner* les deux sous-séquences triées en une seule séquence triée.

La procédure de fusion se fait de la façon suivante :
- on place les deux séquences dans deux piles A et B, avec en bas de chaque pile une valeur ∞, et on prend une pile vide C
- on compare les têtes des deux listes, et on dépile la plus petite valeur dans la pile C
- on arête quand les têtes des piles A et B sont ∞

Remarque – La pile C est "à l'envers". Il suffit de la retourner, ou de modifier légèrement l'algorithme pour rempiler dans le bon sens.

Calculons la complexité de cet algorithme.

Soit $T(2^n)$ le temps d'exécution pour une séquence d'entrée de taille 2^n.

- L'étape *Diviser* se fait en temps constant.
- L'étape *Appliquer récursivement* se fait en temps $2T(2^{n-1})$.
- L'étape *Fusionner* se fait linéairement ($2^n - 1$ pour 2^n éléments).

Finalement, on a

$$T(n) = \begin{cases} \mathcal{O}(1) & \text{si } n = 0 \\ 2T(2^{n-1}) + 2^n - 1 & \text{si } n > 0 \end{cases}$$

On a donc

$$\begin{aligned} T(2^n) &= 2T(2^{n-1}) + 2^n - 1 \\ 2T(2^{n-1}) &= 2^2T(2^{n-2}) + 2^n - 2 \\ \vdots &= \vdots \\ 2^{n-1}T(2^2) &= 2^nT(2^0) + 2^n - 2^{n-1} \end{aligned}$$

En sommant toutes ces égalités, on trouve

$$T(2^n) = 1 + 2^n(n-1).$$

Finalement, si r est quelconque, on peut l'encadrer entre deux puissances de 2, et on trouve

$$T(r) = \mathcal{O}(r \log r).$$

2.17.2 Tri fusion

Le schéma de l'algorithme :

– *Diviser* le tableau $A[p\cdots r]$ en deux tableaux non vides $A[p\cdots q-1]$ et $A[q+1\cdots r]$ vérifiant

$$\forall e \in A[p\cdots q-1],\ \forall f \in A[q+1\cdots r],\ e \leqslant x \leqslant f$$

où $x = A[p]$ est placé en position $A[q]$;
– *Appliquer récursivement* le tri rapide aux deux sous-tableaux ;
– *Fusionner* : le tri se fait sur place, il n'y a rien à faire.
Regardons la complexité.

Dans le pire cas, le tableau est déjà trié.
– *Diviser* est linéaire ;
– *Appliquer récursivement* est appliqué n
On arrive à une complexité de $\mathcal{O}(n^2)$.

En moyenne, on doit faire certaines hypothèses : on suppose que les n éléments à trier sont distincts, dans $[\![1, n]\!]$; on suppose que la probabilité pour chaque élément d'être le pivot est $1/n$.

Notons C la complexité en moyenne du tri rapide.

On a $C(0) = C(1) = 0$.

Soit $C_k(n)$ le coût moyen quand le pivot est k. On a alors

$$C(n) = \frac{1}{n}\sum_{k=1}^{n} C_k(n).$$

On peut calculer $C_k(n)$:

$$C_k(n) = \underbrace{n+1}_{\text{partitionnement}} + C(k-1) + C(n-k).$$

On a donc

$$\begin{aligned} C(n) &= \frac{1}{n}\sum_{k=1}^{n} [n+1+C(k-1)+C(n-k)] \\ &= n+1+\frac{2}{n}\sum_{k=1}^{n-1} C(k) \end{aligned}$$

On calcule donc

$$nC(n) = n(n+1) + 2\sum_{k=1}^{n-1} C(k)$$

et

$$(n+1)C(n+1) = (n+1)(n+2) + 2\sum_{k=1}^{n} C(k).$$

En soustrayant, on obtient

$$(n+1)C(n+1) - nC(n) = 2(n+1) + 2C(n),$$

soit

$$\frac{C(n+1)}{n+2} - \frac{C(n)}{n+1} = \frac{2}{n+2}.$$

En itérant, on a

$$\frac{C(n)}{n+1} - \frac{C(1)}{2} = \sum_{k=1}^{n-1} \frac{2}{k+2},$$

soit

$$C(n) = (n+2)2\sum_{k=3} n+1\frac{1}{k} = \mathscr{O}(n\log n).$$

On note que la constante cachée dans le $\mathscr{O}$ est approximativement 2γ, qui est assez petite.

2.17.3 Conclusion

Le tri fusion est toujours optimal, au pire et en moyenne, mais ne se fait pas en place.

Le tri rapide a un pire cas moins performant, mais en moyenne est optimal avec une petite constante. Il se fait en place.

2.18 Algorithme d'unification

Nour *et al.* (n.d.)

On cherche à unifier un ensemble d'équations.

On prend E un ensemble d'équations, et on construit deux suites E_n et σ_n par récurrence, avec $E_0 = E$ et $\sigma_0 = id$.

Puis :

1. Si $E_n = E' \sqcup \{f(u_1, \dots, u_q) \sim g(v_1, \dots, v_p\}$, alors
   - Si $f = g$, alors $E_{n+1} := E' \cup \{u_1 \sim v_1, \dots, u_p \sim v_p\}$ et $\sigma_{n+1} := \sigma_n$.
   - Si $f \neq g$, alors renvoyer "Échec clash".
2. Si $E_n = E' \sqcup \{x \sim x\}$, on pose $E_{n+1} := E'$ et $\sigma_{n+1} := \sigma_n$.
3. Si $E_n = E' \sqcup \{x \sim u\}$ (ou $\{u \sim x\}$), avec $u \neq x$, alors
   - Si la variable x n'apparaît pas dans u, on pose $E_{n+1} := E'[x := u]$ et $\sigma_{n+1} := [x := u] \circ \sigma_n$.
   - Sinon, renvoyer "Échec occur-check".

Remarques :

- L'algorithme n'est pas déterministe ; cependant, l'ordre des opérations n'a pas d'importance.
- La complexité est exponentielle.

Proposition 2.53
L'algorithme ci-dessus termine toujours, soit par un échec, soit avec $E_n = \emptyset$.

Démonstration. On note respectivement a_n, b_n, c_n le nombre de variables, de symboles de fonction, d'équations dans E_n, et $f(n) = (a_n, b_n, c_n)$. On montre que si $E_n \neq \emptyset$ et pas d'échec à l'étape n, $f(n+1) < f(n)$ avec $<$ l'ordre lexicographique. On regarde suivant les cas de l'algorithme :

	a_n	b_n	c_n
Cas 1	$=$	$\searrow$	
Cas 2	$\searrow$ ou $=$	$=$	$\searrow$
Cas 3	$\searrow$		

□

Théorème 2.54
Si l'algorithme termine avec $E_n = \emptyset$, alors σ_n est le mgu *de E. S'il échoue par clash ou occur-check, alors E n'a pas d'unificateur.*

Lemme 2.55
Soient x une variable, u un terme, et σ une substitution. Si $x[\sigma] = u[\sigma]$, alors $\sigma = \sigma \circ [x := u]$.

Démonstration. On pose $\sigma' = \sigma \circ [x := u]$. Alors

$$x[\sigma'] = x[x := u][\sigma] = u[\sigma] = x[\sigma],$$

et si $y \neq x$, alors

$$y[\sigma'] = y[x := u][\sigma] = y[\sigma].$$

□

Démonstration. On montre tout d'abord.

$$H_n = \text{« À l'étape } n \text{ de l'algorithme, } \sigma \text{ unifie } E \textit{ si, et seulement si } \text{il existe } \sigma' \text{ qui unifie } E_n \text{ et telle que } \sigma = \sigma' \circ \sigma_n \text{. »}$$

Si H_n est vraie pour tout n, alors, soit m la dernière étape de l'algorithme
- Si $E_m = \emptyset$, l'identité unifie E_m, et H_m nous dit alors que σ_m unifie E, et que c'est bien le *mgu*.
- Si l'algorithme échoue par clash : E_m contient une équation de la forme $f(x_1, \ldots, x_p) \sim g(y_1, \ldots, y_q)$ avec $f \neq g$, et donc E_m n'est pas unifiable. Par H_m, E n'est pas unifiable non plus.
- Si l'algorithme échoue par occur-check : il y a dans E_m une équation de la forme $x \sim u$, x variable de u, $x \neq u$. Alors pour toute substitution σ, la taille de $x[\sigma]$ est toujours strictement inférieure à $u[\sigma]$.
  E_m n'est donc pas unifiable, et par H_m, E non plus.
  Il ne nous reste plus qu'à montrer :

Lemme 2.56
Pour tout n tel que (E_n, σ_n) est défini, H_n.

Démonstration. On a $\sigma_0 = id$ et $E_0 = E$, et donc H_0 est triviale.
Soit donc n tel que (E_n, σ_n) et (E_{n+1}, σ_{n+1}) soient définis.
Montrons alors

$$\exists \sigma',\ \sigma = \sigma' \circ \sigma_n \text{ et } \sigma' \text{ unifie } E_n$$
$$\iff$$
$$\exists \sigma'',\ \sigma = \sigma'' \circ \sigma_{n+1} \text{ et } \sigma'' \text{ unifie } E_{n+1}$$

Regardons chaque cas de l'algorithme :

1. $E_n = E' \sqcup \{f(u_1, \ldots, u_n) \sim f(v_1, \ldots, v_n)\}$, et donc $E_{n+1} = E' \sqcup \{u_1 \sim v_1, \ldots, u_n \sim v_n\}$ et $\sigma_n = \sigma_{n+1}$.
   Dans ce cas, on voie bien que l'équivalence est vérifiée.
2. $E_n = E' \sqcup \{x \sim x\}$, et donc $E_{n+1} = E'$ et $\sigma_{n+1} = \sigma_n$.
   Ici encore, l'équivalence est vérifiée.
3. $E_n = E' \sqcup \{x \sim u\}$ ou $\{u \sim x\}$ avec $x \neq u$ et x n'est pas une variable de u, et donc $E_{n+1} = E'[x := u]$ et $\sigma_{n+1} = \sigma_n$.
   Alors
   ($\Leftarrow$) On prend $\sigma' = \sigma'' \circ [x := u]$.
   ($\Rightarrow$) Si σ' unifie E_n, alors $x[\sigma'] = u[\sigma']$. Par le lemme 2.55, on a $\sigma' = \sigma' \circ [x := u]$, et donc on a le résultat en posant $\sigma'' = \sigma'$.

$\Diamond$

$\square$

Références

Alessandri, Michel. *Thèmes de géométrie. groupes en situation géométrique.*

Carrega, Jean-Claude. *Théorie des corps ; la règle et le compas.*

Carton, Olivier. *Langages formels, calculabilité et complexité.*

Chambert-Loir, Antoine, & Fermigier, Stéfane. *Analyse 2, exercices.*

Cori, René, & Lascar, Daniel. tome 2. *Logique mathématique, tome 2.*

Cormen, Thomas H., Leiserson, Charles E., Rivest, Ronald L., & Stein, Clifford. *Introduction à l'algorithmique.*

Floyd, Robert, & Biegel, Richard. *Le langage des machines.*

Francinou, Serge, Gianella, Hervé, & Nicolas, Serge. Algèbre 1. *Oraux x-ens, algèbre 1.*

Francinou, Serge, Gianella, Hervé, & Nicolas, Serge. Algèbre 2. *Oraux x-ens, algèbre 2.*

Francinou, Serge, Gianella, Hervé, & Nicolas, Serge. Algèbre 3. *Oraux x-ens, algèbre 3.*

Froidevaux, Christine, Gaudel, Marie-Claude, & Soria, Michèle. *Types de données et algorithmes.*

Gourdon, Xavier. Analyse. *Les maths en tete, analyse.*

Leichtman, Éric. *Exercices corrigés de mathématiques posés à l'oral des concours de polytechniqie et des ens, algèbre et géormétrie.*

Nour, K., David, R., & Raffalli, C. *Introduction à la logique : Théorie de la démonstration. cours et exercices corrigés.*

Perrin, Daniel. *Cours d'algèbre.*

Petazzoni, Bruno. *Seize problèmes d'informatique : Avec corrigés détaillés et programmes en caml.*

Ramis, Jean-Pierre, Warusfel, André, & Moulin, François. *Cours de mathématiques pures et appliquées.* Vol. 1, Algèbre et Géométrie.

Rouvière, François. *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation.*

Saux Picart, Philippe, & Rannou, Eric. *Cours de calcul formel. corps finis, systèmes polynomiaux, applications.*

Schwartz, Laurent. Analyse 2. *Analyse 2.*

Stern, Jacques. *Fondements mathématiques de l'informatique.*

Wolper, Pierre. *Introduction à la calculabilité.*

Zuily, Claude, & Queffélec, Hervé. *Analyse pour l'agrégation.*