Agrégation 2025 - Développements

Jonathan BADIN

Table des matières

Groupe d'ordre pq

On se donne $p < q$ avec p et q premiers. Le but de ce développement est de classifier les groupes d'ordre pq.

1. Dévissage

Montrons que si H est un sous-groupe d'un groupe fini G dont l'indice est égal au plus petit facteur premier p de $|G|$ alors H est distingué dans G. On considère l'action : $g \cdot \overline{x} = \overline{gx}$ où $\mathrm{Stab}(\overline{x}) = x^{-1}Hx \cap H$. L'équation aux classes s'écrit :

$$p = \sum_{\overline{x} \text{ rep}} |\mathrm{Orb}(\overline{x})| = 1 + \sum_{\substack{\overline{x} \text{ rep} \\ \overline{x} \neq \overline{1}}} |\mathrm{Orb}(\overline{x})|.$$

Comme $|\mathrm{Orb}(\overline{x})|$ divise $|H|$ divise $|G|$ si $|\mathrm{Orb}(\overline{x})| \neq 1$ alors $|\mathrm{Orb}(\overline{x})| \geqslant p$ ce qui contredit l'égalité. Ainsi toutes les orbites sont réduites à un point. Par la relation orbite-stabilisateur : $\forall x,\ \mathrm{Stab}(\overline{x}) = H$ i.e. $x^{-1}Hx \subset H$ donc H est distingué.

Ainsi si G est un groupe d'ordre pq pour $y \in G$ d'ordre p [1], $\langle y \rangle$ est distingué dans G.

2. Extension

Soit $x \in G$ d'ordre p [2] montrons que $\langle x \rangle$ est un complémentaire de $\langle y \rangle$.

- $\langle x \rangle \cap \langle y \rangle = \{1\}$ par le théorème de Lagrange [3]
- $\langle x \rangle \langle y \rangle$ par la formule du produit $|HK| = \frac{|H||K|}{|H \cap K|}$.

On a alors :

$$G = \langle x \rangle \ltimes \langle y \rangle \simeq \mathbf{Z}/p\mathbf{Z} \ltimes \mathbf{Z}/q\mathbf{Z}.$$

3. Analyse des produits semi-directs

On cherche à déterminer les produits semi-directs entre $\mathbf{Z}/p\mathbf{Z}$ et $\mathbf{Z}/q\mathbf{Z}$ pour cela on s'intéresse aux morphismes $\alpha : \mathbf{Z}/p\mathbf{Z} \to \mathrm{Aut}(\mathbf{Z}/q\mathbf{Z})$. D'abord les automorphismes de $\mathbf{Z}/q\mathbf{Z}$ sont de la forme $x \mapsto kx$ où k est inversible dans $\mathbf{Z}/q\mathbf{Z}$ donc ce dernier groupe est isomorphe à $(\mathbf{Z}/q\mathbf{Z})^\times$ donc cyclique d'ordre $q-1$. Ensuite comme $\mathbf{Z}/p\mathbf{Z}$ est cyclique les morphismes $\alpha : \mathbf{Z}/p\mathbf{Z} \to (\mathbf{Z}/q\mathbf{Z})^\times$ sont exactement les applications de la forme $x \mapsto a^x$ où a est un élément d'ordre divisant p de $(\mathbf{Z}/q\mathbf{Z})^\times$. Alors,

- Si p ne divise pas $q-1$ il n'existe pas d'élément d'ordre p dans $(\mathbf{Z}/q\mathbf{Z})^\times$ par Lagrange donc α est nécessairement le morphisme trivial.
- Si p divise $q-1$ comme $(\mathbf{Z}/q\mathbf{Z})^\times$ est cyclique il existe des éléments d'ordre p. Plus précisément, l'ensemble des éléments d'ordre divisant p forme un sous-groupe cyclique d'ordre p, dans ce cas si b est un générateur de ce sous-groupes on a $\alpha = \beta \circ \varphi_l$ où $\beta : x \mapsto b^x$ et $\varphi_l : x \mapsto lx$ avec $0 \leqslant l \leqslant p-1$.

Conclusion.

- Si p ne divise pas $q-1$ [4] alors il n'existe qu'un produit semi-direct de $\mathbf{Z}/p\mathbf{Z}$ par $\mathbf{Z}/q\mathbf{Z}$ qui est le produit direct isomorphe à $\mathbf{Z}/pq\mathbf{Z}$. Ainsi, dans ce cas tout sous-groupe d'ordre pq est cyclique.
- Si p divise $q-1$ [5] mise à part le produit direct il existe des produits semi-direct non triviaux et ces derniers sont tous isomorphes. Ainsi on a à isomorphisme un unique groupe d'ordre pq non

1. qui existe par Cauchy
2. qui existe par Cauchy
3. car $\langle x \rangle \cap \langle y \rangle$ est un sous-groupe de $\langle x \rangle$ donc son ordre divise p et par symétrie il divise q donc est 1.
4. ce qui arrive pour 15,35,51 etc...
5. ce qui arrive pour 21, 39, etc...

cyclique dont une présentation est : $\langle x, y \mid x^p 1, y^q = 1, xyx^{-1} = y^b \rangle$ où b est un élément d'ordre p de $(\mathbf{Z}/q\mathbf{Z})^\times$.[6]

Remarque. *1. Avec le théorème de la progression arithmétique de Dirichlet*[7] *il existe une infinité de groupes d'ordre pq non isomorphes.*

2. *Une question que soulève ce développer est s'il existe des entiers tels que tout groupe de cet ordre est cyclique ? On dira qu'un tel entier est cyclique, ce développement montre que pq $(p < q)$ est cyclique si et ssi p ne divise pas $q - 1$. Et bien plus généralement un entier n est cyclique si et seulement n est premier avec $\varphi(n)$, ce qui en terme de facteurs premiers signifie que $n = p_1 \cdots p_k$ avec les p_i distincts et que $p_i \not| p_j - 1$. Ce résultat illustre de façon significative comment les propriétés arithmétiques de l'ordre d'un groupe influe sur sa structure.*

3. *Pour terminer donnons quelques idées pour démontrer le résultat annoncé précédemment. Dans un sens si n ne vérifie pas les conditions évoquées : d'un si $n = p^2 s$ alors $\mathbf{Z}/ps\mathbf{Z} \times \mathbf{Z}/p\mathbf{Z}$ n'est pas cyclique et d'autre part si $n = pqs$ avec p et q premiers tels que $p \mid q - 1$ alors comme nous l'avons vu précédemment il existe un groupe G d'ordre pq non cyclique et alors $G \times \mathbf{Z}/s\mathbf{Z}$ est non cyclique d'ordre n. Pour la réciproque on raisonne par récurrence, on se donne G est un groupe d'ordre n vérifiant les conditions évoquées, dans ce cas tout sous-groupe H de G vérifie aussi les conditions donc est cyclique. Un point essentiel est que l'on peut choisir H distingué ce qui revient à dire que G est non simple. P.Caldero montre cela directement à partir du fait que tous les sous-groupes de G sont abéliens, D.Perrin conseille d'utiliser une conséquence au théorème du p-complément de Burnside. On peut ainsi dévisser notre groupe par H, et de même que précédemment G/H est cyclique par hypothèse de récurrence. Maintenant on peut montrer que H est central à partir des contraintes sur l'ordre : on considère le morphisme $\overline{g} \mapsto (h \mapsto ghg^{-1})$ de G/H vers $\mathrm{Aut}(H)$; ce morphisme ne peut être que trivial. Finalement H, G/H sont cycliques et H est central alors un argument classique permet de conclure que G est abélien et l'on peut conclure avec le théorème des structures des groupes abéliens finis.*

6. Pour $p = 2$ on peut prendre $b = -1$ on retrouve alors le groupe diédral D_q.
7. Ce cas peut se démontrer spécifiquement avec des polynômes cyclotomiques c'est un développement qu'on voit.

Cardinal du cône nilpotent

1. Décomposition de Fitting

Lemme. *Soit E un K-espace vectoriel de dimension $d < \infty$. Pour $u \in \mathrm{End}(E)$ il existe une unique décomposition u-stable $E = F \oplus G$ tel que $u_{|F}$ soit inversible et $u_{|G}$ nilpotent.*

- Pour une telle décomposition on a $\chi_u = \chi_{u_{|F}}\chi_{|G}$ où $\chi_{|G} = X^p$ avec $p = \dim F$. De plus 0 n'est pas racine de $\chi_{u_{|F}}$ car $u_{|F}$ est inversible donc p est la multiplicité de 0 en en tant que racine de χ_u.
- De plus par le lemme des noyaux et multiples utilisations de Cayley-Hamilton on a $E = \mathrm{Ker}Q(u) \oplus \mathrm{Ker}u^p$ où $F \subset \mathrm{Ker}Q(u)$ et $G \subset \mathrm{Ker}u^p$ donc par un argument de dimension $F = \mathrm{Ker}Q(u)$ et $G = \mathrm{Ker}u^p$.
- Ainsi si une telle décomposition on a nécessairement $F = \mathrm{Ker}Q(u)$ et $G = \mathrm{Ker}u^p$ où $\chi_u = X^p Q$ avec $X \nmid Q$. Cela montre en particulier l'unicité de la décomposition. Maintenant pour conclure il faut vérifier qu'on a bien $E = F \oplus G$ et $u_{|F}$ inversible et $u_{|G}$ nilpotent. Le premier point découle du lemme des noyaux, le second découle du fait que Q est un polynôme annulateur de $u_{|F}$ qui n'admet pas 0 comme racines.

2. Relation implicite

Avec la décomposition de Fiting on a une bijection

$$\mathrm{End}(E) \longrightarrow \{(F, G, u_F, u_G) \mid E = F \oplus G,\ u_F \in \mathrm{GL}(F) \text{ et } u_G \in \mathrm{Nil}(G)\}.$$

En notant g_d le cardinal du groupe linéaire et n_d le cardinal du cône nilpotent d'un espace vectoriel de dimension d sur $\mathbf{F}_q$ on obtient la relation :

$$q^{d^2} = \sum_{k=0}^{d} g_{d-k} n_k \#\{E = F \oplus G \mid \dim F = k \text{ et } \dim G = d-k\}.$$

Il s'agit maintenant de dénombrement les décomposition de E en deux sous-espaces de dimension fixé. On va interpréter cette ensemble comme une orbite d'une action de groupe pour se ramener par la formule orbite-stabilisateur au dénombrement du stabilisateur.

3. Dénombrement des décompositions

Le groupe $\mathrm{GL}(E)$ agit sur l'ensemble des sommes directes par : $\phi \cdot (F, G) = (\phi(F), \phi(G))$.

- Les orbites de cette action sont exactement $\{F \oplus G \mid \dim F = k\}$ pour $0 \leqslant k \leqslant n$. D'une part si ϕ est un isomorphisme on a $\dim \phi(F) = \dim F$ d'autre part si $\dim F = \dim F'$ alors considérant une base $(e_1, \cdots, e_n)$ adaptée à la décomposition $F \oplus G$ et une base $(e'_1, \cdots, e_n)'$ adaptée à la décomposition $F' \oplus G'$ on définit $\phi(e_i) = e'_i$. Comme ϕ envoie une base sur une base c'est un isomorphisme et on a par construction $(F', G') = (\phi(F), \phi(G))$.
- Un élément ϕ est dans le stabilisateur d'un couple (F, G) avec $\dim F = k$ si et seulement si dans une base adaptée à la décomposition la matrice de ϕ est de la forme :

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \qquad (A, B) \in \mathrm{GL}_k(\mathbf{F}_q) \times \mathrm{GL}_{n-k}(\mathbb{F}_q).$$

  Par la relation orbite stabilisateur on a donc :

$$\#\{E = F \oplus G \mid \dim F = k \text{ et } \dim G = d-k\} = \frac{g_d}{g_k g_{d-k}}.$$

4. Calcul

Finalement $q^{d^2} = \sum_{k=0}^{d} n_k \frac{g_d}{g_k}$ d'où :

$$n_d = q^{d^2} - \frac{g_d}{g_{d-1}} q^{(d-1)^2}.$$

On a $\frac{g_d}{g_{d-1}} = \frac{(q^d-1)\cdots(q^d-q^{d-1})}{(q^{d-1}-1)\cdots(q^{d-1}-q^{d-2})} = (q^d-1)q^{d-1}$ de sorte que :

$$\begin{aligned} n_d &= q^{d^2} - q^{d-1}(q^d-1)q^{(d-1)^2} \\ &= q^{d^2} - q^{2d-1+(d-1)^2} + q^{(d-1)^2+d-1} \\ &= q^{d(d-1)}. \end{aligned}$$

Remarque. *1. A partir de ce résultat on peut dénombrer les endomorphismes trigonalisables. En utilisant la représentation en sous-espaces caractéristiques on a une bijection entre les endomorphismes trigonalisables et $\{(F_1, n_1, \cdots, F_q, n_q)$ où $F_1 \oplus \cdots \oplus F_q = E$ et n_i nilpotent sur $F_i\}$. On en déduit que :*

$$t_n = \sum_{\substack{(n_1,\cdots,n_q) \\ n_1+\cdots+n_q=n}} \frac{g_n}{g_{n_1}\cdots g_{n_q}} q^{n_1(n_1-1)}\cdots q^{n_q(n_q-1)}$$

où $\frac{g_n}{g_{n_1}\cdots g_{n_q}}$ représente le nombre de décomposition en sommes directes de q sous-espaces de dimensions $n_1, \cdots, n_q$ généralisation naturelle du cas de deux sous-espaces vu précédemment.

2. Un autre dénombrement usuel et plus facile concerne celui des matrices diagonalisables. Dans ce cas on a une bijection avec $\{(F_1, \cdots, F_q)$ où $F_1 \oplus \cdots \oplus F_q = E\}$ d'où l'on déduit que :

$$d_n = \sum_{\substack{(n_1,\cdots,n_q) \\ n_1+\cdots+n_q=n}} \frac{g_n}{g_{n_1}\cdots g_{n_q}}.$$

Marche aléatoire sur Z/nZ

Considérons une marche aléatoire X_n sur $\mathbf{Z}/N\mathbf{Z}$ i.e. $X_n = U_1 + \cdots + U_n$ où les U_i sont des variables aléatoires indépendantes identiquement distribuées sur $\mathbf{Z}/N\mathbf{Z}$. La loi de X_n que l'on note f_n définit une fonction de $\mathbf{Z}/N\mathbf{Z}$ dans $\mathbf{C}$ égal à la convoluée de la loi des U_i correspondant à f_1 *i.e.* $f_n = f_1^{*n}$. Pour étudier f_n l'idée de ce développement est de passer en Fourier où convolution devient multiplication. Encore faut-il développer une théorie de Fourier sur $\mathbf{Z}/N\mathbf{Z}$, c'est ce qu'on fera dans un premier temps en présentant la transformée de Fourier discrète.
Pour faire les calculs on va se limiter à un exemple en considérant que $f_1 = \frac{1}{3}\delta_{-1} + \frac{1}{3}\delta_0 + \frac{1}{3}\delta_1$ c'est à dire que la marche aléatoire choisie de faire un pas dans le sens direct, de faire un pas dans le sens indirect ou de rester sur place.

1. Transformée de Fourier discrète

Pour commencer il nous faut déterminer les caractères de $\mathbf{Z}/N\mathbf{Z}$ i.e. les morphismes de groupes de $\mathbf{Z}/N\mathbf{Z} \to \mathbf{C}$. Ce sont exactement les applications de la forme :

$$\chi_k : j \mapsto \left(\mathrm{e}^{\frac{2\mathrm{i}k\pi}{N}}\right)^j.$$

Pour $f \in \mathscr{F}(\mathbf{Z}/N\mathbf{Z}, \mathbf{C})$ on définit alors sa transformée de Fourier comme :

$$\widehat{f} : k \mapsto \langle f, \chi_k \rangle = \sum_{j=0}^{N-1} f(j)\mathrm{e}^{-\frac{2\mathrm{i}\pi kj}{N}}.$$

Un calcul qu'on passe donne $\widehat{f * g} = \widehat{f} \times \widehat{g}$ ce qui nous donne $\widehat{f}_n = \widehat{f}_1^n$ où $\widehat{f}_1(k) = \frac{1+2\cos\frac{k\pi}{N}}{3}$.

2. Inversion de Fourier, expression de la loi de X_n

Les caractères sont orthogonaux pour $\langle f \mid g \rangle = \sum_{j \in \mathbf{Z}/N\mathbf{Z}} f(j)\overline{g(j)}$ car :

$$\sum_{j=0}^{N-1} \mathrm{e}^{\frac{2\mathrm{i}\pi(k-l)}{N}} = \begin{cases} 0 & \text{si} \quad k \neq l \\ N & \text{si} \quad k = l \end{cases}$$

Ils forment donc une base orthogonal[8]. On en déduit la formule d'inversion de Fourier :

$$f(j) = \frac{1}{N} \sum_{k=0}^{N-1} \widehat{f}(k)\mathrm{e}^{\frac{2\mathrm{i}\pi kj}{N}}.$$

Dans notre cas on trouve :

$$f_n(j) = \frac{1}{N} \sum_{k=0}^{N-1} \left(\frac{1 + 2\cos\frac{k\pi}{N}}{3}\right)^n \mathrm{e}^{\frac{2\mathrm{i}\pi kj}{N}}.$$

Comme $|\frac{1+2\cos\left(\frac{2\mathrm{i}\pi k}{N}\right)}{3}| < 1$ lorsque $k \in \{1, \cdots, N-1\}$ et vaut 1 sinon on a $f_n(j) \to \frac{1}{N}$. Cela signifie que la marche aléatoire X_n converge en loi vers la distribution uniforme sur $\mathbf{Z}/N\mathbf{Z}$.

3. Estimation de la vitesse de convergence

L'orthogonalité des caractères assure que $f \mapsto \frac{1}{\sqrt{N}}\widehat{f}$ est une isométrie :

$$\|f_n\|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\widehat{f}_n(k)|^2 = \frac{1}{N} \|\widehat{f}_n\|^2.$$

8. étant au nombre de N

Utilisons cette propriété pour contrôler la vitesse de convergence de f_n vers $f_\infty = \frac{1}{N}$ en norme 2. Sachant que $\widehat{f_\infty}(k) = \sum_{j=0}^{N-1} N^{-1} \mathrm{e}^{\frac{2\mathrm{i}\pi kj}{N}} = N^{-1}\delta_{0,k}$ d'après un calcul précédent on a :

$$\|f_n - f_\infty\| = \frac{1}{\sqrt{N}}\|\widehat{f_n} - \widehat{f_\infty}\| = \frac{1}{\sqrt{N}}\left(\sum_{k=1}^{N-1}\left(\frac{1+2\cos\left(\frac{2\mathrm{i}\pi k}{N}\right)}{3}\right)^{2n}\right)^{1/2} \leqslant \left|\frac{1+2\cos\left(\frac{2\mathrm{i}\pi}{N}\right)}{3}\right|^n.$$

On a ainsi convergence au moins géométrique de la loi de X_n vers la loi uniforme.

Remarque. *Ce n'est qu'un cas particulier de convergence des chaînes de Markov, ici la chaîne est irréductible apériodique et la distribution uniforme est son unique mesure invariante, c'est en fait une mesure réversible. L'intérêt de l'utilisation de l'analyse de Fourier discrète et d'obtenir des estimations de la vitesse de convergence. Notons que cette démarche correspond à la méthode général d'étude de la vitesse de convergence par analyse du spectre de la matrice de transition ici la matrice circulante C_N. En effet la transformée de Fourier est simplement la transformation qui diagonalise C_N, précisément si F_N est la matrice de $\mathscr{F}$ dans la base canonique alors $F_N^{-1}C_NF_N$ est diagonale de valeurs propres $\frac{1+2\cos\left(\frac{2\mathrm{i}\pi k}{N}\right)}{3}$.*

Polygones constructibles

Théorème. *Un nombre réel x est constructible si et seulement s'il existe une suite d'extensions :*

$$\mathbf{Q} = K_0 \hookrightarrow K_1 \hookrightarrow \cdots \hookrightarrow K_n \ni x.$$

Corollaire. *Un nombre constructible est algébrique de degré une puissance de 2.*

Ce résultat permet de résoudre négativement de nombreux problèmes géométriques : trissection de l'angle, quadrature du cercle ou duplication du cube.

Application. *Si le n-gone régulier à n côtés est constructibles alors n est le produit d'une puissance de 2 et de nombres premiers de Fermat.*

1. Analyse algébrique

Soit $M = (x, y)$ un point constructible. On a par définition une suite de parties $A_0 \subset A_1 \subset \cdots \subset A_n \ni M$ tel que A_{i+1} s'obtiennent en ajoutant un point constructible $M_{i+1} = (x_{i+1}, y_{i+1})$ à partir de A_i. Notons K_i le corps engendré par A_i. On a ainsi :

$$K_0 = \mathbf{Q} \subset K_1 \subset \cdots \subset K_n \ni M.$$

où $K_{i+1} = K_i(x_{i+1}, y_{i+1})$. On s'est ramené à comprendre l'extension d'un sous-corps K de $\mathbf{R}$ par ajout d'un élément constructible $M = (x, y)$ à partir de points de K.

- si $M = (x, y)$ est l'intersection de deux droites distinctes formé par des points de coordonnées dans K alors M est solution de deux équations $\det(\overrightarrow{A_iB_i}, \overrightarrow{A_iM}) = 0$ pour $i = 1, 2$ qui se développe en le système d'équations :
$$\begin{cases} a_1x + b_1y = c_1 \\ a_2x + b_2y = c_2 \end{cases}$$
Les paramètres de ces équations peuvent être pris à c Comme les droites sont distinctes, le point d'intersection est unique donc ce système possède une unique solution. Les formules de Cramer permettent d'exprimer cette solution comme somme et produits des coefficients a_i, b_i, c_i. Ainsi $(x, y) \in K$.
- si $M = (x, y)$ est l'intersection d'une droite est d'un cercle alors M est solution des équations $\det(\overrightarrow{A_1B_1}, \overrightarrow{A_1M}) = 0$ et $\|\overrightarrow{A_2M}\|^2 = \|\overrightarrow{A_2B_2}\|$ où $A, B, C, D \in K \times K$. En développant on obtient le système d'équations :
$$\begin{cases} a_1x + b_1y = c_1 \\ x^2 + y^2 + a_2x + b_2y = c_2 \end{cases}$$
On va alors montrer que x et y sont solutions d'une équation de degré $\leqslant 2$. Parmi a_1 et b_1 l'un de deux coefficients est non nul, sinon l'équation définie un point ou rien du tout, en tout cas pas une vraie droite. Par symétrie supposons que $a_1 \neq 0$, ainsi $x = \frac{c_1 - b_1y}{a_1}$ puis en injectant cette expression dans la seconde équation on voit que y vérifie une équation de degré 2. Ainsi $K(y)$ est une extension de degré 1 ou 2 de K. Ensuite $x \in K(y)$ donc la même conclusion vaut pour $K(x, y)$.
- si (x, y) est l'intersection de deux cercles on a un système d'équations de la forme
$$\begin{cases} x^2 + y^2 + a_1x + b_1y = c_1 \\ x^2 + y^2 + a_2x + b_2y = c_2 \end{cases}$$
à coefficients dans K, qu'on peut réécrire en soustrayant la seconde ligne à la première :
$$\begin{cases} (a_1 - a_2)x + (b_1 - b_2)y = (c_1 - c_2) \\ x^2 + y^2 + a_2x + b_2y = c_2 \end{cases}$$
On est donc ramené à un système de la forme précédente. Géométriquement on a écrit l'intersection de deux cercles comme l'intersection d'un des deux cercles avec la droite formé par les deux points d'intersections.

On conclut donc que les extensions de corps $K_i \hookrightarrow K_{i+1}$ sont de degré inférieur à 2.

2. Critère de constructibilité

Nous venons de montrer que $e^{\frac{2i\pi}{n}}$ est algébrique de degré $\varphi(n)$. On veut s'intéresser au degré de $\cos(2i\pi/n)$. Notons que :

$$d^0 \cos(2i\pi/n) \times (\mathbf{Q}(e^{\frac{2i\pi}{n}}) : \mathbf{Q}(\cos(2i\pi/n))) = \varphi(n)$$

or $(\mathbf{Q}(e^{\frac{2i\pi}{n}}) : \mathbf{Q}(\cos(2i\pi/n)))$ est degré inférieur à 2 car $e^{\frac{2i\pi}{n}}$ est solution d'une équation de degré 2 sur $\mathbf{Q}(\cos(2i\pi/n))$ à savoir $X^2 - 2\cos(2i\pi/n)X + 1$. Ainsi, si le n-gone régulier est constructible alors $\varphi(n)$ est une puissance de 2. Écrivons $n = 2^{\alpha_0} p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ la décomposition en facteurs premiers de n alors $\varphi(n) = 2^{\alpha_0 - 1} p_1^{\alpha_1 - 1}(p_1 - 1) \cdots p_k^{\alpha_k - 1}(p_k - 1)$. Ainsi pour que $\varphi(n)$ soit une puissance de 2 il faut et il suffit que les nombres premiers p_i impairs soit de multiplicité 1 et de la forme $2^m + 1$. De telles nombres premiers sont appelés nombres premiers de Fermat et on peut montrer qu'il sont nécessairement de la forme $2^{2^m} + 1$.[9] Ainsi si le polygone à n côtés est constructibles alors n est produit de puissances de 2 et de nombres premiers de Fermat.[10]

3. La question de la réciproque

On commence par montrer la réciproque au théorème de Wantzel, pour cela il faut essentiellement montrer que l'ensemble des nombres constructibles est stable par les opérations $+, \times$ et $\sqrt{}$. On utilise pour cela qu'on sait faire les constructions suivantes : perpendiculaire, parallèle et report de longueur.

- Déjà avec la construction de perpendiculaire, on obtient qu'un point $M = (x, y)$ est constructible si et seulement si les points $(x, 0)$ et $(0, y)$ le sont. Notons de plus qu'avec un compas $(x, 0)$ est constructible ssi $(0, x)$ est constructible. On en déduit que l'ensemble des nombres constructibles est de la forme $\Gamma \times \Gamma$ où Γ est l'ensemble des nombres constructibles sur la droite réelle. Dans la suite on se restreint à Γ.
- Par report de longueur il est clair que Γ est stable par addition. Aussi, avec un compas si $x \in \Gamma$ alors $-x \in \Gamma$.
- Avec le théorème de Thalès on montre que Γ est stable par produit. On peut aussi montrer que si $x \in \Gamma$ avec $x \neq 0$ alors $x^{-1} \in \Gamma$.
- Finalement avec la construction d'Audin on a la stabilité par racine carrée.

Cela montre la réciproque de Wantzel : d'abord $\mathbf{Q} \subset \Gamma$ et si $K \subset \Gamma$ et K' est une extension de K de degré 2 on peut écrire $K' = K(x)$ où x est de degré 2 sur K donc $x^2 + ax + b = 0$ puis les formules de résolutions donnent $x = \frac{-a \pm \sqrt{a^2 - 4b}}{2}$ sachant que le discriminant de l'équation est positif car admet une racine réelle par hypothèse. Maintenant pour ce qui est de la constructibilité des n-gones réguliers pour montrer la réciproque la démarche classique consiste à construire la suite d'extensions quadratique par dualité avec la théorie de Galois. On peut déjà se ramener à n premier car si n et m sont premiers entre eux et que $e^{\frac{2i\pi}{n}}$ et $e^{\frac{2i\pi}{m}}$ sont constructibles alors en écrivant $1 = an + bm$ on a $e^{\frac{2i\pi}{nm}} = (e^{\frac{2i\pi}{n}})^a (e^{\frac{2i\pi}{m}})^b$ est constructible. Maintenant pour n premier différent de le groupe de Galois de $\cos\left(\frac{2\pi}{n}\right)$ est par correspondance de Galois un sous-groupe d'ordre 2 du groupe cyclique $(\mathbf{Z}/p\mathbf{Z})^*$ qui est le groupe de Galois de $e^{\frac{2i\pi}{n}}$ donc est un groupe cyclique d'ordre $\frac{p-1}{2}$. Comme ce nombre est une puissance de 2 on peut construire une filtration avec des quotients d'ordre 2 ce qui donne le résultat voulue par correspondance de Galois.

Remarque. *Plus généralement on peut caractériser les nombres constructibles de la façon suivante : un nombre z est constructible si et seulement si la clôture normale de $\mathbf{Q}(z)$ est de degré une puissance de 2.*

9. Si $p = 2^m + 1$ est premier alors on peut voir que l'ordre de 2 modulo p est $2m$ donc $2m$ divise $p - 1 = 2^m$ donc m est une puissance de 2.

10. On connait peu de nombres premiers de Fermat, 5 précisément $3, 5, 17, 257, 65537, 4294967297, 18446744073709551617$. Le problème est que même avec un critère de primalité spécifique (test de Pépin) les temps de calcul ne permette guère d'aller loin, on a ainsi pu tester la primalité de $2^{2^m} + 1$ jusqu'à $m = 32$.

Polynômes cyclotomiques

On cherche les polynômes minimaux des racines de l'unité. Une racine n-ème de l'unité étant racine du polynôme X^n-1 c'est un nombre algébrique. Le problème devient alors de décomposer X^n-1 en facteurs irréductibles. On commence par chercher un candidat, pour cela on s'inspire des idées de la théorie de Galois qui dit qu'il faut regrouper entre eux des éléments de "même nature algébrique". Dans ce sens on utilise la partition $\mu_n(\mathbf{C})\sqcup_{d\,|\,n}\mu_d^*(\mathbf{C})$ qui conduit aux polynômes cyclotomiques :

$$X^n-1=\prod_{d\,|\,n}\Phi_d \qquad \text{où} \quad \Phi_d=\prod_{\zeta\in\mu_d^*(\mathbf{C})}(X-\zeta).$$

Montrons alors que les candidats Φ_d sont bien les facteurs irréductibles de X^n-1 sur $\mathbf{Q}$.

1. Coefficients des polynômes cyclotomiques

On commence par montrer que les polynômes cyclotomiques sont à coefficients rationnels, on va voir qu'ils sont même à coefficients entiers. On raisonne par récurrence sur le degré.

- Pour $n=1$ on a $\Phi_1=X-1$.
- Supposons le résultat pour tout entier $d<n$. Écrivons alors $X^n-1=\Phi_nA(X)$ où $A(X)=\prod_{d\,|\,n,d\neq n}\Phi_d\in\mathbf{Z}[X]$ par hypothèse de récurrence. Comme X^n-1 est unitaire, on a existence d'une division euclidienne de X^n-1 par A sur $\mathbf{Z}[X]$ qu'on écrit $X^n-1=AQ+R$. Maintenant par unicité de la division euclidienne sur $\mathbf{C}[X]$ on a $\Phi_n(X)=Q(X)\in\mathbf{Z}[X]$.

2. Irréductibilité des polynômes cyclotomiques

On ne connait de Φ_n que ses racines, on va exploiter cela en introduisant P le polynôme minimal de $\zeta:=e^{\frac{2i\pi}{n}}$ et montrons que $P=\Phi_n$. Pour cela on montre que les autres n-ème de l'unité sont racines de P c'est à dire $A(\zeta^m)=0$ dès que $m\wedge n=1$ dès lors on aura $\deg P\geqslant\varphi(n)$ et comme P divise Φ_n et sont tout deux unitaires on aura égalité. Maintenant il suffit de faire cela pour $m=p$ premier puis d'itérer le processus. Dans la suite on se donne p un nombre premier ne divisant pas n. Pour montrer que ζ^p est racine de P on va comparer P au polynôme minimal Q de ζ^p. Comme ζ est racine de $Q(X^p)$ le polynôme P divise $Q(X^p)$ dans $\mathbf{Q}[X]$. On va vouloir réduire cette égalité modulo, avant cela vérifions que nos polynômes sont à coefficients entiers et que les divisions qu'on considère sont en fait dans $\mathbf{Z}[X]$. On utilise pour cela le lemme suivant :

Lemme. *Si $A\in\mathbf{Q}[X]$ unitaire divise $C\in\mathbf{Z}[X]$ unitaire alors $A\in\mathbf{Z}[X]$ et A divise C dans $\mathbf{Z}[X]$.*

Ainsi on a $P,Q\in\mathbf{Z}[X]$ car ils divisent X^n-1. Ensuite P divise $Q(X^p)$ dans $\mathbf{Z}[X]$ car P unitaire et $Q(X^p)$ unitaire à coefficients entiers. On peut alors réduire modulo p et ont trouve que $\overline{P}$ divise $\overline{Q}^p$. On va pouvoir conclure : supposons que $P\neq Q$, alors PQ divise X^n-1 puis si π est un facteur irréductible de $\overline{P}$ alors π divise $\overline{Q}$ par le lemme d'Euclide donc π^2 divise X^n-1 dans $\mathbf{Z}/p\mathbf{Z}[X]$ or $\mathscr{D}(X^n-1)=nX^{n-1}$ n'a pas de racines communes avec X^n-1 (hypothèse p premier avec n) donc X^n-1 est sans facteur carré. Terminons par la preuve du lemme : écrivons $AB=C$ on considère a,b tels que aA et bB soient à coefficients entiers, on divise alors chacun de ses polynômes par leurs contenus i.e. le pgcd de leurs coefficients, comme A est unitaire donc B aussi, les contenus respectifs de aA et bB divise respectivement a et b de sorte qu'on peut prendre a et b afin que aA et bB soient primitifs. De la multiplicativité du contenu on en déduit que $1=c(aA)c(bB)=c(abC)=ab$ d'où le résultat.

3. Construction des polygones réguliers à la règle et au compas

Pour construire le n-gone régulier il suffit de savoir construire $e^{\frac{2i\pi}{n}}$ car alors on peut répéter la construction en prenant comme points de départ O et $e^{\frac{2i\pi}{n}}$. Et pour construire $e^{\frac{2i\pi}{n}}$ il suffit de savoir

construire $\cos\frac{2\pi}{n}$ car alors $e^{\frac{2i\pi}{n}}$ se trouve à l'extérieur du cercle unité et de la perpendiculaire à l'axe des abscisses passant par $\cos\frac{2\pi}{n}$. Maintenant le théorème de Wantzel nous dit que pour que $\cos\frac{2\pi}{n}$ soit constructible il faut que $\cos\frac{2\pi}{n}$ soit algébrique de degré une puissance de 2. On a :

$$(\mathbf{Q}(\cos\frac{2\pi}{n}):\mathbf{Q})(\mathbf{Q}(e^{\frac{2i\pi}{n}}):\mathbf{Q}(\cos\frac{2\pi}{n}))=\varphi(n)$$

or $(\mathbf{Q}(e^{\frac{2i\pi}{n}}):\mathbf{Q}(\cos\frac{2\pi}{n}))$ est de degré 1 ou 2 car $e^{\frac{2i\pi}{n}}$ est solution de $X^2-2\cos\frac{2\pi}{n}X+1=0$ ce qui traduit algébriquement la situation géométrique décrit précédemment. Somme toute $(\mathbf{Q}(\cos\frac{2\pi}{n}):\mathbf{Q})$ est une puissance de 2 si et ssi $\varphi(n)$ est une puissance de 2 or en notant $n=2^{\alpha_0}p_1^{\alpha_1}\cdots p_k^{\alpha_k}$ la décomposition en facteurs premiers de n on a $\varphi(n)=2^{\alpha_0-1}p_1^{\alpha_1-1}(p_1-1)\cdots p_k^{\alpha_k-1}(p_k-1)$. Ainsi pour que $\varphi(n)$ soit une puissance de 2 il faut et il suffit que les nombres premiers p_i impairs soit de multiplicité 1 et de la forme 2^m+1. De telles nombres premiers sont appelés nombres premiers de Fermat et on peut montrer qu'il sont nécessairement de la forme $2^{2^m}+1$. [11] Ainsi si le polygone à n côtés est constructibles alors n est produit de puissances de 2 et de nombres premiers de Fermat. [12]

Remarque.

1. *Avec la théorie (ou simplement les idées) de Galois on peut montrer la réciproque : un polygone n est constructible si et seulement si n est produit d'une puissance de* 2 *et de nombres premiers de Fermat de multiplicité* 1.
2. *Plus généralement on peut caractériser les nombres constructibles de la façon suivante : un nombre z est constructible si et seulement si la clôture normale de* $\mathbf{Q}(z)$ *est de degré une puissance de* 2.
3. *Construire explicitement à la règle et au compas ces polygone est possible en analysant les extensions de degrés* 2 *intermédiaires qui mènent à* $\mathbf{Q}(z)$. *Notons que la difficulté se situe dans la construction des polygone avec un nombre impair de côtés. Il est en comparaison relativement simple de dédoubler le nombre de côtés d'un polygone en traçant des bissectrices.*

11. Si $p=2^m+1$ est premier alors on peut voir que l'ordre de 2 modulo p est $2m$ donc $2m$ divise $p-1=2^m$ donc m est une puissance de 2.

12. On connait peu de nombres premiers de Fermat, 5 précisément $3,5,17,257,65537,4294967297,18446744073709551617$. Le problème est que même avec un critère de primalité spécifique (test de Pépin) les temps de calcul ne permette guère d'aller loin, on a ainsi pu tester la primalité de $2^{2^m}+1$ jusqu'à $m=32$.

Simplicité du groupe projectif orthogonal

On étudie la décomposition du groupe orthogonal $\mathrm{O}_n(\mathbf{R})$. Ce groupe à un premier sous-groupe $\mathrm{SO}_n(\mathbf{R})$ noyau du déterminant. Lorsque $n=2$, $\mathrm{SO}_n(\mathbf{R})$ est un groupe abélien isomorphe à $\mathbf{S}^1$, c'est connu, on va donc s'intéresser au cas $n \geqslant 3$. Dans ce cas $\mathrm{SO}_n(\mathbf{R})$ est un groupe parfait, son centre est constituée des homothéties de $\mathrm{SO}_n(\mathbf{R})$ c'est à dire $\{\mathrm{Id}\}$ si n est impair et $\{\pm\mathrm{Id}\}$ si n est pair. On note alors $\mathrm{PSO}_n(\mathbf{R})$ le quotient de $\mathrm{SO}_n(\mathbf{R})$ par son centre. On a alors épuisé nos techniques génériques de dévissage, et on peut espérer avoir un groupe simple. C'est ce que nous allons voir à une exception près. Pour cela on va utiliser que les retournements forme un système générateur [13] de $\mathrm{SO}_n(\mathbf{R})$ et que ces retournements sont tous conjugués [14] On se donne ensuite un sous-groupe distingué non trivial de $\mathrm{PSO}_n(\mathbf{R})$ qu'on peut relever en un sous-groupe N distingué dans $\mathrm{SO}_n(\mathbf{R})$ non réduit aux homothéties. On va montrer que $N = \mathrm{SO}_n(\mathbf{R})$. Pour cela on va montrer que N contient un retournement, comme les retournements sont conjugués N les contiendra tous et alors $N = \mathrm{SO}_n(\mathbf{R})$ car les retournement sont générateurs.

- Construire un retournement demande de construire un élément avec beaucoup de points fixes. On va déjà essayer de se ramener à un élément agissant trivialement sur un sous-espace de dimension $n-2$. En fait comme une isométrie positive en dimension 3 admet une droite fixe il suffit de construire un élément agissant trivialement sur un sous-espace de dimension $\geqslant n-3$. Dans le cas $n=3$ il n'y a rien à faire, maintenant pour que notre démarche de réduction fonctionne nous aurons besoin de suffisamment de place : $n \geqslant 5$.
- On va y aller à coût de commutateur ; on part d'un élément non trivial $u \in N$ puis on pose $v = [u, r]$. Pour que v possède beaucoup de points fixes il faut prendre r avec beaucoup de points fixes, on va donc prendre pour r un retournement disons de plan P. Ainsi, $v = (uru^{-1})r^{-1}$ est le produit d'un retournement de plan $u(P)$ et d'un retournement de plan P, ainsi v laisse stable $u(P)^\perp \cap P^\perp$. Par la formule de Grassman, $u(P)^\perp \cap P^\perp = (u(P)+P)^\perp$ donc est de dimension $n - \dim(u(P)+P)$. En particulier v laisse stable un sous-espace de dimension $\geqslant n-4$. On aimerait gagner une dimension pour se ramener à un sous-espace de dimension 3. On peut faire cela si u stabilise une droite de P ce qui revient à demander qu'il existe a tel que $u(a) = \pm a$.
- Récapitulons, s'il existe a tel que $u(a) = \pm a$ alors en prenant P un plan contenant a disons $P = \langle a, b\rangle$ alors v laisse stable un sous-espace F de dimension $n-3$. De plus en prenant b tel que $c = u(b)$ ne soit pas colinéaire à b ce qui est possible car u n'est pas une homothétie alors uru^{-1} est un retournement de plan $\langle a, c\rangle$ qui est donc différent de r^{-1} qui est un retournement de plan $\langle a, b\rangle$.
- Maintenant s'il n'existe pas de tel vecteur a, alors on peut utiliser la construction précédente, comme v fixe un sous-espace de dimension $n-4$ et que $n \geqslant 5$ on en déduit que v possède un point fixe. Pour être ramener au cas précédent il faut seulement vérifier que $v \neq \pm\mathrm{Id}$, ayant un point fixe il suffit de vérifier que $v \neq \mathrm{Id}$, pour cela on choisie u de sorte que $u(P) \neq P$ ce qui est possible comme u n'est pas une homothétie.
- Ainsi on s'est ramené à $v \in N$ non trivial laissant fixe un sous-espace F de dimension $\geqslant n-3$. On peut profiter de cela en se restreignant à la dimension 3, les opérations que l'on fait devront être comprise comme la trace des opérations sur $F^\perp$, l'action sur le supplémentaire étant trivial. Pour faire les choses proprement on pose $N_0 = N \cap \mathrm{SO}(F^\perp)$ où $\mathrm{SO}(F^\perp)$ est vu comme un sous-groupe de $\mathrm{SO}_n(\mathbf{R})$ par l'injection $t \mapsto (t, \mathrm{Id}_F)$.
- Les éléments de $\mathrm{SO}_3(\mathbf{R})$ sont des rotations, elles sont définies par un axe de rotation et un angle θ. Cet angle apparait dans la trace $1 + 2\cos(\theta)$. Un retournement correspond à une rotation

13. On montre que les réflexions engendre $\mathrm{SO}_n(\mathbf{R})$ de façon classique en faisant agir les réflexions pour fixer des éléments. Les éléments de $\mathrm{SO}_n(\mathbf{R})$ sont alors produit d'un nombre pair de réflexions, dans le cas $n=3$ comme l'opposé d'une transvection est un retournement on a le résultat et dans le cas général $n \geqslant 3$ on montre que le produit de deux réflexions s'écrit comme produit de deux retournement en se ramenant à un sous-espace de dimension 3 contenant l'axe des deux réflexions.

14. Cela découle de la transitivité de l'action de $\mathrm{SO}_n(\mathbf{R})$ sur les sous-espace de dimension k.

d'angle $\theta = \pi$, pour cela il suffit de disposer d'une rotation r d'angle $\frac{\pi}{n}$ car alors r^n sera un retournement c'est que nous allons montrer

- On introduit l'application :
$$T : g \in \mathrm{SO}_3(\mathbf{R}) \mapsto \mathrm{Tr}([v,g]).$$
Il faut montrer que l'image de T contient un élément de la forme $1+2\cos\left(\frac{\pi}{n}\right)$ avec $n \in \mathbf{N}^*$. Remarquons pour cela que $\mathrm{SO}_3(\mathbf{R})$ étant compact et connexe, l'image de T est un intervalle $[a,b]$. On a toujours $T(g) \leqslant 3$ avec égalité pour $g = \mathrm{Id}$ de sorte que $b = 3$. Maintenant comme $\frac{\pi}{n} \to 0$ lorsque $n \to \infty$ il suffit de montrer que $a < 3$. Or $T(g) = 3$ si et seulement si $[v,g] = \mathrm{Id}$ et v n'étant pas une homothétie il existe g ne commutant pas avec v, un tel élément a donc une image < 3 ce qui permet de conclure.

Remarque. *Dans le cas $n = 4$ on a un comportement exceptionnel :* $\mathrm{PSO}(4,\mathbf{R}) \simeq \mathrm{SO}(3,\mathbf{R}) \times \mathrm{SO}(3,\mathbf{R})$ *en particulier* $\mathrm{PSO}(4,\mathbf{R})$ *n'est pas simple. Ce résultat peut s'établir à partir du corps des quaternions. On montre d'abord un isomorphisme entre* $\mathrm{SO}(3,\mathbf{R})$ *et le groupe G des quaternions unités quotienté par $\{\pm 1\}$ via $q \mapsto (h \mapsto qhq^{-1})$. On montre ensuite un isomorphisme entre* $\mathrm{SO}(4,\mathbf{R})$ *et $G \times G$ quotienté par $\{\pm(1,1)\}$ via $(q_1,q_1) \mapsto (h \mapsto q_1 q q_2^{-1})$. On déduit de ces deux isomorphismes le résultat annoncé (voir Perrin).*

Permutations aléatoires

Lorsque l'on étudie la la distribution des cycles d'une permutation aléatoire uniforme le premier résultat qu'on dérive est la distribution conjointe des cycles (surement un item du plan). Ce résultat est très satisfaisant et semble conclure le problème seulement l'expression de la loi conjointe est assez compliqué et il n'est pas évident d'en déduire l'ensemble des quantités d'intérêt, notamment les lois marginales. Le but de ce développement est de calculer cette quantité. Pour cela on démontre préalablement un résultat général de combinatoire des ensembles, généralisation de la formule du crible de Poincaré.

1. Principe d'inclusion-exclusion

On se donne $A_1, \cdots, A_n$ des évènements, on cherche à calculer la probabilité $\mathbf{P}(E_k)$ qu'exactement k évènements parmi les A_i sont réalisés. Pour cela on calcule la somme des intersections de k évènements, mais dans ce cas on compte plusieurs fois les issues contenant dans l'intersection de $k+1$ évènements, il faut donc enlever la somme des intersection de $k+1$ évènements mais alors on enlève plusieurs fois les issues dans l'intersection de $k+2$ évènement etc... Cela amène à la formule suivante :

$$\mathbf{1}_{E_k} = \sum_{r=k}^{n} (-1)^{r-k} \binom{r}{k} \underbrace{\sum_{i_1<\cdots<i_r} \mathbf{1}_{A_{i_1}\cap\cdots\cap A_{i_r}}}_{:=\Sigma_r}.$$

Pour montrer cette égalité, on se donne un point ω et on note s le nombre d'évènements A_i auquel appartient ω. Dans ce cas le terme de gauche vaut 1 si $s = k$ et 0 sinon. Montrons qu'on obtient la même chose dans le terme de droite. Regardons pour cela la contribution de chaque Σ_r :

- si $r > s$ alors ω n'est dans aucune intersection de r évènements A_i donc la contribution est nulle :
- si $s \geqslant r$ alors dans Σ_r une intersection contenant ω correspond au choix r évènements parmi les s évènements contenant ω, la contribution est donc de $\binom{s}{r}$.

En particulier si $s < k$ alors le terme de droite est nulle, dans ce cas on a bien égalité. Ensuite si $s \geqslant k$ le terme de droite donne :

$$\sum_{r=k}^{s} (-1)^{r-k} \binom{r}{k}\binom{s}{r}.$$

En utilisant que $\binom{s}{r}\binom{r}{k} = \binom{s}{k}\binom{s-k}{r-k}$ ce qui se voit à partir de l'expression factorielle on obtient :

$$\binom{s}{k} \sum_{r=k}^{s} (-1)^{r-k} \binom{s-k}{r-k} = \binom{s}{k}(1-1)^{s-k} = \begin{cases} 0 & \text{si} \quad s \neq k \\ 1 & \text{si} \quad s = k \end{cases}$$

Ainsi les deux membres sont dans tous les cas égaux, ce qui permet de conclure. En passant à l'espérance on en déduit la forme suivante sur les probabilités :

$$\mathbf{P}(E_k) = \sum_{r=k}^{n} (-1)^{r-k} \binom{r}{k} \sum_{i_1<\cdots<i_r} \mathbf{P}(A_{i_1} \cap \cdots \cap A_{i_r}).$$

2. Application à notre problème

Considérons :

- un entier $n \geqslant 1$;
- $\sigma^{(n)}$ une permutation aléatoire uniforme sur S_n, on notera $(C_1^{(n)}, \cdots, C_n^{(n)})$ son profil ;
- un entier j entre 1 et n ;
- $\{c_1, \cdots, c_m\}$ l'ensemble des cycles de longueur j de S_n ;
- A_i l'évènement $\sigma^{(n)}$ possède le cycle c_i.

Ainsi par le point précédent :

$$\mathbf{P}(C_j^{(n)} = k) = \sum_{r=k}^{m} (-1)^{r-k} \binom{r}{k} \sum_{i_1 < \cdots < i_r} \mathbf{P}(A_{i_1} \cap \cdots \cap A_{i_r}).$$

Calculons les probabilités $\mathbf{P}(A_{i_1} \cap \cdots \cap A_{i_r})$. Si les cycles $c_{i_1}, \cdots, c_{i_r}$ ne sont pas disjoints, étant distincts, cette probabilité est nulle. Sinon elle vaut $\mathbf{1}_{jr \leqslant n} \frac{(n-jr)!}{n!}$. En effet pour construire une permutation contenant les cycles $c_1, \cdots, c_r$, ces cycles étant fixé il reste à déterminer les valeurs de la permutation sur les $(n-jr)$ entiers restants.
Il reste à compter le nombre d'ensemble de r-cycles à support disjoints. On trouve :

$$\frac{1}{r!} \frac{(n)_j}{j} \frac{(n-j)_j}{j} \cdots \frac{(n-j(r-1))_j}{j} = \frac{n!}{r!(n-jr)!j^r} \mathbf{1}_{jr \leqslant n}$$

où :

- la division par $r!$ est dû à l'absence d'ordre
- $\frac{(n)_j}{j}$ correspond au choix d'un cycle de longueur j dans un ensemble à n éléments, un j-cycle étant une j-liste invariante par permutation circulaire ;
- $\frac{(n-j)_j}{j}$ correspond au choix d'un cycle de longueur j dans un ensemble à $n-j$ éléments afin que les cycles soient disjoints
- etc...

Finalement on trouve :

$$\begin{aligned} \mathbf{P}(C_j^{(n)} = k) &= \sum_{r=k}^{m} (-1)^{r-k} \binom{r}{k} \frac{1}{r!j^r} \mathbf{1}_{jr \leqslant n} \\ &= \frac{1}{k!j^k} \sum_{r=0}^{\lceil n/j \rceil - k} (-1)^l \frac{1}{l!j^l}. \end{aligned}$$

3. Convergence

Avec l'expression précédente il apparaît que :

$$\mathbf{P}(C_j^{(n)} = k) \underset{n \to +\infty}{\longrightarrow} \frac{\mathrm{e}^{-1/j}}{k!j^k} = \mathrm{Poi}_{1/j}[k].$$

Ainsi $C_j^{(n)}$ converge en loi lorsque $n \to +\infty$ vers une Poisson de paramètre $\frac{1}{j}$. On peut de plus estimer la vitesse de convergence en variation totale ; en notant $n_j = \lceil n/j \rceil$

$$d_{\mathrm{TV}}(C_j^{(n)}, \mathrm{Poi}(j^{-1})) \sim \frac{2^{n_j+1}}{(n_j+1)!j^{n_j+1}}.$$

D'abord l'écart entre $\mathbf{P}(C_j^{(n)} = k)$ et $\mathrm{Poi}_{1/j}[k]$ est le reste d'une série alternée, on peut donc majorer et minorer à partir de ces deux premiers termes, précisément :

$$\frac{1}{j^{n_j+1}} \frac{1}{k!(n_j+1-k)!} - \frac{1}{j^{n_j+2}} \frac{1}{k!(n_j+2-k)} \leqslant \left| \mathbf{P}(C_j^{(n)} = k) - \mathrm{Poi}_{1/j}[k] \right| \leqslant \frac{1}{j^{n_j+1}} \frac{1}{k!(n_j+1-k)!}.$$

En sommant pour $k = 0, \cdots, n$ et en reconnaissant une formule du binôme on en déduit l'équivalent.

Relation de Frobenius-Zolotarev

1. Lemme de Zolotarev

1. Regardons les cycles de la permutation m_a. Il y a 0 qui est un point fixe puis les autres sont les orbites de l'action de groupe $\langle m_a\rangle \curvearrowright \mathbf{F}_q^\times$ c'est à dire de l'action par translation de $\langle a\rangle$ sur $\mathbf{F}_q^\times$. Les supports des cycles sont ainsi les classes de $\mathbf{F}_q^\times$ modulo $\langle a\rangle$, il y en a $\frac{q-1}{\mathrm{ord}(a)}$ de cardinal $\mathrm{ord}(a)$. On en déduit que :
$$\varepsilon(m_a) = \left((-1)^{\mathrm{ord}(a)}\right)^{\frac{q-1}{\mathrm{ord}(a)}}.$$
2. On va distinguer deux cas :
   - si $\mathrm{ord}(a)$ est pair alors $\varepsilon(m_a) = (-1)^{\frac{q-1}{\mathrm{ord}(a)}}$ or $a^{\frac{\mathrm{ord}(a)}{2}} = -1$ [15] donc $\varepsilon(m_a) = a^{\frac{q-1}{2}}$.
   - si $\mathrm{ord}(a)$ est impair alors $\varepsilon(m_a) = 1$ et dans ce cas on a une racine explicite de a à savoir $a^{\frac{\mathrm{ord}(a)+1}{2}}$.

2. Extension de Frobenius

- Il suffit de monter le résultat sur une partie génératrice de $\mathrm{GL}_n(\mathbf{F}_q)$: les transvections. Par conjugaison, on peut se ramener à une dilatation élémentaire. Une dilatation élémentaire u agit de la façon suivante : $u(x_1,\cdots,x_n) = u(x_1,\cdots,x_{i-1},ax_i,x_{i+1},\cdots,x_n)$ où $a \in \mathbf{F}_q^\times$. Dans ce cas $\det(u) = a$ et on cherche à montrer que $\varepsilon(u) = a^{\frac{q-1}{2}}$. En terme de permutations u correspond à $\mathrm{Id}_{\mathbb{F}_q^{i-1}} \times m_a \times \mathrm{Id}_{\mathbb{F}_q^{n-i}}$ où m_a est la multiplication par a sur $\mathbf{F}_q$.
- Soit σ une permutation sur un ensemble X et τ une permutation sur un ensemble Y ; calculons la signature de la permutation $\sigma\times\tau$. On suppose pour cela que X et Y sont munis d'un ordre et on munit $X\times Y$ de l'ordre lexicographique. Alors les couples $(x,y) < (x',y')$ sont inversés si et seulement si $(\sigma(x'),\tau(y')) < (\sigma(x),\tau(y))$ si et seulement si $\sigma(x') < \sigma(x)$ ou $\sigma(x) = \sigma(x')$ et $\tau(y) < \tau(y')$. Le premier cas comporte $\mathrm{Inv}(\sigma)\times|Y|$ possibilités, le second $|X|\times\mathrm{Inv}(\tau)$. Ainsi, $\varepsilon(\sigma\times\tau) = \varepsilon(\sigma)^{|X|}\varepsilon(\tau)^{|Y|}$.
- En appliquant le point précédent comme $\mathbf{F}_q$ est impair on a $\varepsilon(u) = \varepsilon(m_a)$ où m_a désigne la multiplication par a sur $\mathbf{F}_q$ et l'on ramené à une dimension.

On a ainsi montré le résultat et son extension par le calcul. Maintenant en prenant un peu de recul sur le problème, on peut dériver le résultat de la nature des objets et de leurs propriétés calculatoires, presque sans calculs explicites. On représente la situation avec le diagramme suivant :

Imaginer le diagramme commutatif.

Il s'agit de montrer que le morphisme $\delta : \mathbf{F}_q^* \to \{\pm 1\}$ est $a \mapsto a^{\frac{q-1}{2}}$. En fait on n'a pas vraiment le choix car $\mathbf{F}_p^*$ étant cyclique, il existe au plus deux morphismes de $\mathbf{F}_p^*$ dans $\{\pm1\}$. Or on connait au moins deux morphismes de $\mathbf{F}_q^*$ dans $\{\pm1\}$, le morphisme trivial et $a\mapsto a^{\frac{q-1}{2}}$. Pour montrer que δ est ce second morphisme il suffit donc de montrer que δ n'est pas trivial. Pour cela on exhibe un automorphisme u tel que $\varepsilon(u) = -1$. Un exemple est l'automorphisme induit sur $\mathbf{F}_q^n$ munie de la structure du corps fini à q^n éléments par la multiplication par g où g est un générateur du groupe cyclique de ce corps. Cet automorphisme agit comme un q^n-1 cycle donc est de signature $(-1)^{q^n} = -1$.

3. Application : deuxième loi complémentaire

On va utiliser le lemme de Zolotarev pour établir la deuxième loi complémentaire :
$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

15. Car dans un corps 1 n'a que deux racines 1 et -1 et par définition de l'ordre le cas 1 est exclue.

On regarde pour cela la multiplication par 2 sur $\mathbf{F}_p$, le tableau suivant décrit cette permutation : Pour calculer sa signature on regarde son nombre d'inversions. Il ne peut avoir d'inversions qu'avec $1 \leqslant i \leqslant \frac{p-1}{2}$ et $\frac{p+1}{2} \leqslant j \leqslant p-1$ dans ce cas à i fixé on a une inversion si et seulement si $\frac{p+1}{2} \leqslant j < \frac{p+1}{2} + i$. Ainsi le nombre total d'inversions est :

$$\sum_{i=1}^{\frac{p-1}{2}} i = \frac{\frac{p-1}{2}\frac{p+1}{2}}{2} = \frac{p^2-1}{8}.$$

Remarque. *Avec un peu plus de travail on peut également établir la loi de réciprocité quadratique, on donne les détails dans la suite. On se donne a et b deux nombres impairs premier entre eux, on cherche à montrer que $(a \mid b)(b \mid a) = (-1)^{\frac{a-1}{2}\frac{b-1}{2}}$. Pour cela il faut comparer les signatures des permutations m_a sur $\mathbf{Z}/b\mathbf{Z}$ et m_b sur $\mathbf{Z}/a\mathbf{Z}$. Pour manipuler ensemble ces deux permutations on va travailler sur $\mathbf{Z}/a\mathbf{Z} \times \mathbf{Z}/b\mathbf{Z}$.*

- *On a une bijection naturelle $\pi : \mathbf{Z}/ab\mathbf{Z} \to \mathbf{Z}/a\mathbf{Z} \times \mathbf{Z}/b\mathbf{Z}$ qui envoie x sur $(x \mod a, x \mod b)$. A partir de cette représentation construisons les permutations :*

  $$\sigma : (x, y) \mapsto (ax + y, y)$$
  $$\tau : (x, y) \mapsto (x, by + x).$$

  On a légèrement modifié nos permutations mais comme la translation par k dans $\mathbf{Z}/n\mathbf{Z}$ est la puissance k-ème d'un n-cycle donc est de signature $(-1)^{k(n-1)}$, cette quantité vaut 1 si n est impair de sorte qu'on a bien $\varepsilon(\sigma) = (a \mid b)$ et $\varepsilon(\tau) = (b \mid a)$. Maintenant cette modification permet facilement de calculer :

  $$\pi^{-1} \circ \sigma(x, y) = ax + y \text{ et } \pi^{-1} \circ \tau(x, y) = by + x.$$

- *Maintenant une autre représentation de $\mathbf{Z}/ab\mathbf{Z}$, la représentation de la base mixte consiste à écrire tout élément de $\mathbf{Z}/ab\mathbf{Z}$ sous la forme $ta + s$ avec $(t, s) \in [0, a[\times [0, b[$. On a ainsi une bijection entre $\mathbf{Z}/ab\mathbf{Z}$ et $[0, a[\times [0, b[$. Or, sur ce dernier ensemble on dispose de deux ordres naturelles : l'ordre lexicographique et l'ordre lexicographique inverse qui se transporte à $\mathbf{Z}/ab\mathbf{Z}$. La permutation $ax + y \mapsto by + x$ de $\mathbf{Z}/ab\mathbf{Z}$ correspond à l'unique application croissante passant d'un ordre à l'autre.*
- *Calculons donc le nombre d'inversions du passage de l'ordre lexicographique de $[0, a[\times [0, b[$ à l'ordre lexicographique inverse. Pour cela il faut dénombrement le nombre doubles couples (i, j) et (i', j') tels que $(i, j) < (i', j')$ pour l'ordre à gauche et $(i', j') < (i, j)$ pour l'autre à droite, dans le premier cas il faut que $i < i'$ ou $i = i'$ et $j < j'$ et dans le second cas $j' < j$ ou $j' = j$ et $i' < i$, les seules possibilités sont donc $i < i'$ et $j' < j$, il y a donc $\binom{a}{2}\binom{b}{2}$ possibilités, d'où le résultat.*

Structure du groupe multiplicatif de Z/n Z et applications

1. Dévissage par le théorème chinois

Écrivons $n = p_1^{\alpha_1} \cdots p_k^{\alpha_k}$ la décomposition en facteurs premiers de n. Les anneaux $\mathbf{Z}/n\mathbf{Z}$ et $\mathbf{Z}/p_1^{\alpha_1}\mathbf{Z} \times \cdots \times \mathbf{Z}/p_k^{\alpha_k}\mathbf{Z}$ sont isomorphes, en particulier on a un isomorphisme entre leurs groupes multiplicatifs :

$$(\mathbf{Z}/n\mathbf{Z})^\times \simeq (\mathbf{Z}/p_1^{\alpha_1}\mathbf{Z})^\times \times \cdots \times (\mathbf{Z}/p_k^{\alpha_k}\mathbf{Z})^\times.$$

On est ainsi ramené à étudier $(\mathbf{Z}/p^\alpha\mathbf{Z})^*$.

2. Structure de $(\mathbf{Z}/p^\alpha\mathbf{Z})^*$ pour p impair

Pour ce développement on suppose connu que $(\mathbf{Z}/p\mathbf{Z})^*$ est cyclique, c'est le fait général pour le groupe multiplicatif d'un corps fini. Plus généralement on montre que $(\mathbf{Z}/p^\alpha\mathbf{Z})^*$ est cyclique. Rappelons que son ordre est $\varphi(p^\alpha) = p^{\alpha-1}(p-1)$. Pour construire un élément d'ordre $p^{\alpha-1}(p-1)$ on construit un élément x d'ordre $p^{\alpha-1}$ et un élément y d'ordre $p-1$. Le groupe étant abélien on a alors $\mathrm{ord}(xy) = \mathrm{ppcm}(\mathrm{ord}(x), \mathrm{ord}(y)) = p^{\alpha-1}(p-1)$.

- Commençons par l'élément d'ordre $p-1$. Considérons le morphisme surjectif $(\mathbf{Z}/p^\alpha\mathbf{Z})^\times \to (\mathbf{Z}/p\mathbf{Z})^\times$ et notons z un antécédent d'un générateur de $(\mathbf{Z}/p\mathbf{Z})^*$. Cet élément est d'ordre un multiple $u(p-1)$ de p de sorte que z^u est d'ordre $p-1$.
- Pour l'élément d'ordre $p^{\alpha-1}$ on va monter que $1+p$ convient. Pour cela on montre par récurrence que pour $k \in \mathbf{N}_0$ on a $(1+p)^{p^k} = 1 + \lambda p^{k+1}$ avec $\lambda \wedge p = 1$. On utilisera le fait suivant : p divise $\binom{p}{i}$ si $1 \leqslant i \leqslant p-1$.
  - $(1+p)^p = \sum_{i=0}^p \binom{p}{i} p^i$. Pour $2 \leqslant i \leqslant p$, p^3 divise $\binom{p}{i} p^i$. Pour $i = p$ on utilise que $p \geqslant 3$ et dans les cas restants le résultat rappelé. On a ainsi :
  $$(1+p)^p = 1 + p^2 + up^3 = 1 + (1+up)p^2.$$
  - $(1+p)^{p^{k+1}} = (1+\lambda p^{k+1})^p = \sum_{i=0}^p \binom{p}{i} \lambda^i p^{i(k+1)}$. Pour $2 \leqslant i \leqslant p$, p^{k+3} divise $\binom{p}{i} p^{i(k+1)}$. Pour $i = p$ on utilise que $p \geqslant 3$ et dans les cas restants avec le résultat rappelé p^{2k+3} divise $\binom{p}{i} p^{i(k+1)}$ et $k \geqslant 1$. Ainsi,
  $$(1+p)^{p^{k+1}} = 1 + \lambda p^{k+2} + up^{k+3} = 1 + (\lambda + up)p^{k+2}.$$

  On conclut que $(1+p)^{p^{\alpha-1}} = 1 \mod p^\alpha$ donc $\mathrm{ord}(1+p)$ divise $p^{\alpha-1}$ et $(1+p)^{p^k} \neq 1 \mod p^\alpha$ pour tout $k < \alpha$ donc $\mathrm{ord}(1+p) = p^{\alpha-1}$.

3. Application aux tests de primalité

- Le test de Fermat $a^{n-1} \equiv 1 \mod n$. Pour analyse ce test on cherche à contrôler lorsque n n'est pas premier le nombre d'entiers a vérifiant la condition précédente. Elle n'est pas vérifié lorsque a n'est pas premier avec n mais si l'on choisit les a au hasard il y a peu de chance de tomber sur un entier non premier avec n. Maintenant l'ensemble des entiers a premier avec n passant le test de Fermat forme un sous-groupe de $(\mathbf{Z}/n\mathbf{Z})^*$, si ce groupe n'est pas propre au moins la moitié des entiers premiers avec n sont des témoins. Maintenant que tous les entiers premiers avec n sont des menteur revient à dire que $\exp((\mathbf{Z}/n\mathbf{Z})^*)$ divise $n-1$. Or par le théorème de structure :
  $$\exp((\mathbf{Z}/n\mathbf{Z})^*) = \mathrm{ppcm}(p_1^{\alpha-1}(p_1-1), \cdots, p_k^{\alpha_k-1}(p_k-1)).$$

  Donc cela revient à dire que $p_i^{\alpha_i-1}(p_i-1)$ divise $n-1$ pour $i = 1, \cdots, k$ ce qui revient à dire que $\alpha_1 = \cdots = \alpha_k = 1$ et $p_i - 1$ divise $n-1$ pour $i = 1, \cdots, k$.

- Pour éviter ces pathologies on peut renforcer le test en le suivant : $a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \mod n$ où le terme de droite fait apparaître le symbole de Jacobi. En élevant au carré on voit que ce test renforce celui de Fermat. Maintenant si tout a premier avec n passe le test montrons que n est premier. Sinon, c'est un nombre de Carmichael donc en particulier $n = p_1 \cdots p_k$. Supposons que $k \geqslant 2$ alors :

$$a^{\frac{n-1}{2}} \equiv \left(\frac{a}{p_1}\right) \cdots \left(\frac{a}{p_k}\right) \mod p_1.$$

  Maintenant par le théorème chinois on peut choisir a tel que a soit un carré modulo p donc $a^{\frac{n-1}{2}}$ mais prendre a tel que a_2 ne soit pas un carré modulo p_2 mais bien un carré pour les autres ce qui contredit l'égalité précédente.

Théorème des deux carrés

On s'intéresse à la représentation des entiers en somme de deux carrés. La factorisation $a^2+b^2 = (a+\mathrm{i}b)(a-\mathrm{i}b)$ invite à se placer sur l'anneau des entiers de Gauss $\mathbf{Z}[\mathrm{i}] = \{a+\mathrm{i}b \,|\, (a,b) \in \mathbf{Z}\}$ et d'étudier son arithmétique.

1. $\mathbf{Z}[\mathrm{i}]$ est euclidien

- Si $x, y \in \mathbf{Z}[\mathrm{i}]$ avec y non nul alors en notant q un entier de Gauss le plus proche de $\frac{x}{y}$ on a $\left|\frac{x}{y} - q\right| \leqslant \frac{1}{\sqrt{2}}$ puis $x = yq + r$ avec $r = \frac{x}{y} - q$ et $N(r) \leqslant \frac{1}{\sqrt{2}} N(y) < N(y)$.
- Si z est inversible dans $\mathbf{Z}[\mathrm{i}]$ on a $zz' = 1$ donc par multiplicativité de la norme $N(z) = 1$. Inversement si $N(z) = z\overline{z}$ donc si $N(z) = 1$, z est inversible dans $\mathbf{Z}[\mathrm{i}]$. Les unités de $\mathbf{Z}[\mathrm{i}]$ sont donc exactement les éléments de norme 1.

2. Factorisation des nombres premiers

Soit p un nombre premier

- Si p est réductible dans $\mathbf{Z}[\mathrm{i}]$ écrivons $p = z_1 z_2$ avec z_1, z_2 non inversibles. Alors $p^2 = N(z_1)N(z_2)$ avec $N(z_1)$ et $N(z_2)$ différents de 1 donc $p = N(z_1) = N(z_2)$. On a ainsi écrit p comme somme de deux carrés. Le problème de la factorisation des nombres premiers dans $\mathbf{Z}[\mathrm{i}]$ correspond ainsi exactement à notre problème d'écrire un nombre premier comme somme de deux carrés.
- Regardons alors à quelle condition un premier p est irréductible dans $\mathbf{Z}[\mathrm{i}]$. On sait que cela est équivalent à ce que (p) soit un idéal maximal ou encore que $\mathbf{Z}[\mathrm{i}]/(p)$ soit un corps. On simplifie [16]

$$\mathbf{Z}[\mathrm{i}]/(p) \simeq (\mathbf{Z}[X]/(X^2+1))/(p) \simeq \mathbf{Z}/p\mathbf{Z}/(X^2+1).$$

  Or comme $\mathbf{Z}/p\mathbf{Z}$ est un corps, $\mathbf{Z}/p\mathbf{Z}[X]$ est principal donc ce dernier anneau est un corps si et seulement si X^2+1 est irréductible. Comme ce polynôme est de degré 2, irréductibilité équivaut à absence de racines, or -1 est une racine dans $\mathbf{Z}/p\mathbf{Z}$ si et seulement si $p \neq 3 \mod 4$ par le critère d'Euler le cas $p = 2$ étant traité à part.
- Pour résumé p est irréductible si et ssi $p \equiv 3 \mod 4$. Ainsi, un nombre premier est somme de deux carrés si et seulement s'il n'est pas congru à 3 modulo 4.

3. Calcul d'une décomposition

La démonstration précédent peut être rendu effectif. Si l'on y revient, on dit qu'un entier $p = 1 \mod 4$ est réductible dans $\mathbf{Z}[\mathrm{i}]$ car il existe $\alpha^2 \equiv -1 \mod p$ de sorte que p divise $\alpha^2 + 1 = (\alpha+\mathrm{i})(\alpha-\mathrm{i})$ mais p aucun des deux facteurs. Ainsi, p ne saurait être irréductible car il n'est pas premier et un facteur non trivial de p est précisément $z = \mathrm{pgcd}(\alpha+\mathrm{i}, p)$. En effet z est non trivial car sinon p diviserait $\alpha - \mathrm{i}$ est non associé à p car sinon p diviserait $\alpha + \mathrm{i}$. Maintenant le calcul d'une racine de -1 modulo p se fait bien en tirant au hasard des entiers modulo p jusqu'à tomber sur un non-résidu quadratique a pour lequel $a^{(p-1)/4}$ qui d'après le critère d'Euler est une racine de -1. Comme la moitié des entiers de $\{1, \cdots, p-1\}$ sont des non-résidus quadratique, on a une chance sur deux à chaque tirage de tomber sur un non-résidu, donc en moyenne on a besoin de 2 étapes et le test s'effectue rapidement à l'aide du symbole de Jacobi. Ensuite il reste à calculer un pgcd pour lequel on utilise l'algorithme d'Euclide.

16. Formellement on peut voir ces isomorphismes comme des conséquences du troisième théorème d'isomorphisme en écrivant par exemple $(\mathbf{Z}[X]/(X^2+1)/(p)) = (\mathbf{Z}[X]/(X^2+1)/((p, X^2+1)/(X^2+1)) \simeq \mathbf{Z}[X]/(p, X^2+1)$ puis en utilisant les mêmes manipulations dans le sens inverse.

4. Entiers sommes de deux carrés

Pour terminer on va déterminer plus généralement à quelle condition un entier n peut s'écrire comme somme de deux carrés. On s'appuie pour cela sur la description des facteurs irréductibles de $\mathbf{Z}$[i] qui découle du travail précédent. En quelques mots on montre que tout facteur irréductible de $\mathbf{Z}$[i] est facteur irréductible d'un nombre premier car si z est un entier de Gauss et p un nombre premier divisant $N(z)$ alors en considérant π un facteur irréductible de p on a $\pi \mid z\overline{z}$ donc π divise z ou $\overline{z}$ c'est à dire π ou $\overline{\pi}$ divise z. Or on a vu qu'un nombre premier p est soit irréductible, soit réductible et dans ce cas $p = \pi\overline{\pi}$ avec $N(z) = p$. Un tel π est alors irréductible car si z divise π, $N(z)$ divise $N(\pi)$. Finalement les facteurs irréductibles de $\mathbf{Z}$[i] sont exactement à inversible près :

1. les nombres premiers congru à 3 modulo 4 ;
2. les entiers de Gauss de norme un nombre premier.

A partir de la description des irréductibles on peut répondre à la question. Soit n un entier que l'on cherche à décomposer en somme de deux carrés c'est à dire à écrire $n = z\overline{z}$ avec $z \in \mathbf{Z}$[i]. Si une telle écriture est possible, en décomposant z en produits d'irréductibles on voit que les nombres premiers congru à 3 modulo 4 facteur de n dans $\mathbf{Z}$ sont de valuation paire. Inversement décomposons n en produits de facteurs irréductibles dans $\mathbf{Z}$ *i.e.* $n = \prod_{i=1}^{k} p_i^2 \prod_{j=1}^{\ell} q_j$ où les p_i sont congru à 3 modulo 4 et les q_j non. Alors q_j se factorise sous la forme $q_j = \pi_j\overline{\pi_j}$ de sorte que $n = z\overline{z}$ avec

$$z = p_1 \cdots p_k \pi_1 \cdots \pi_\ell .$$

Ainsi un entier n est somme de deux carrés si et seulement pour tout facteur premier p de n congru à 3 modulo 4, la valuation de p est paire.

Remarque. *Dans l'analyse précédente on seulement utilisé l'existence d'une décomposition en facteurs irréductibles, on peut aller plus loin dans l'analyse en comptant le nombre de représentations possibles ce qui demande l'unicité. On montre alors que le nombre de façon de représenter n en somme de deux carrés est* $4(d_1(n) - d_3(n))$ *où* $d_j(n)$ *représente le nombre de diviseurs de n congru à j modulo* 4.

Procédure de construction des corps finis

Le but de ce développement est de développer une procédure permettant la réalisation informatique des corps fini. Rappelons que si q est un nombre premier, le corps fini à q^d éléments est le corps de décomposition sur $\mathbf{F}_q$ du polynôme $X^{q^d} - X$. Cependant cette définition ne rend pas les calculs explicites, il serait préférable d'écrire $\mathbf{F}_{q^d}$ comme un rupture, on sait que cela est possible d'après le théorème de l'élément primitif, maintenant on va chercher à rendre cela effectif.

1. Polynômes irréductibles sur $\mathbf{F}_q$

On s'intéresse aux polynômes irréductibles sur $\mathbf{F}_q$. Les polynômes irréductibles sont les briques élémentaires des polynômes sur $\mathbf{F}_q$ au même titre que les nombres premiers sont les briques élémentaires des entiers. Suivant cette idée on va chercher des polynômes irréductibles comme facteurs irréductibles de polynômes remarquables de même qu'on utilise des entiers remarquables comme $\binom{2n}{n}$ pour rechercher des nombres premiers. Dans le cadre des corps finis, une famille de polynômes occupe une place centrale à savoir ceux de la forme $X^{q^n} - X$. On va alors montrer le résultat suivant :

$$X^{q^n} - X = \prod_{\substack{P \text{ irr} \\ \deg(P) \mid n}} P.$$

Pour montrer ce résultat on va en quelque sorte lire la divisibilité sur l'inclusion des corps de décomposition en utilisant le théorème de structure des corps finis.

- Si P irréductible de degré d divise $X^{q^n} - X$ alors P possède une racine α dans $\mathrm{Dec}(X^{q^n} - X) = \mathbf{F}_{q^n}$ or $\mathbf{F}_q(\alpha)$ est un sous-corps de $\mathbf{F}_{q^n}$ de degré q^d donc $d \mid n$.
- Si P est irréductible de degré d divisant n alors $\mathrm{Rup}(P)$ est un corps de cardinal q^d donc s'injecte dans $\mathbf{F}_{q^n}$. Ainsi P possède une racine dans $\mathbf{F}_{q^n}$ donc P donc P et $X^{q^n} - X$ ne sont pas premier entre eux sur $\mathbf{F}_{q^n}$ donc également sur $\mathbf{F}_q$ donc P divise $X^{q^n} - X$.
- Finalement le polynôme dérivée de $X^{q^n} - X$ est -1 sur $\mathbf{F}_q$ qui est premier avec $X^{q^n} - X$ donc ce dernier est sans facteurs carrés.

On va déduire deux choses de cette formule : première des informations quantitatives sur le nombre de polynômes irréductibles de degré fixé. Deuxièmement un critère d'irréductibilité. Ces deux informations permettront de mettre en œuvre une procédure pour déterminer un polynôme irréductible de degré prescrit en tirant aléatoirement un polynôme, les informations quantitatives donneront des bornes sur le temps d'atteinte, et en testant son irréductibilité avec le test précédent.

2. Estimations du nombre de polynômes irréductibles de degré fixé

En regardant les degrés dans la formule précédente on a $q^n = \sum_{d \mid n} d I(n,q)$. Avec la formule d'inversion de Mobius on en déduit que $I(n,q) = \frac{1}{n} \sum_{d \mid n} \mu\left(\frac{n}{d}\right) q^d$. A partir de ces formules on en déduit l'encadrement suivant :

$$\begin{aligned} \frac{q^n}{n} \geqslant I(n,q) &\geqslant \frac{q^n}{n} - \frac{1}{n} \sum_{k=1}^{\lfloor \frac{n}{2} \rfloor} q^k \\ &\geqslant \frac{q^n}{n} - \frac{q^{\lfloor n/2 \rfloor + 1} - 1}{q-1} \\ &\geqslant \frac{q^n}{n} \left(1 - q^{1-n/2}\right). \end{aligned}$$

Pour $n \geqslant 3$ on a $1 - q^{1-n/2} \geqslant 1 - q^{-1/2} \geqslant 1 - \frac{1}{\sqrt{2}}$. Pour $n = 2$ la minoration est 0 donc n'apporte pas d'informations, on peut calculer directement $I(2,q) = \frac{q(q-1)}{2} \geqslant \frac{q^2}{4}$. Ces majorations montre que la proportion de polynômes irréductibles est de l'ordre de $\frac{1}{n}$ est justifie la méthode aléatoire pour

trouver un polynôme irréductible de degré donnée. On peut ainsi estimer le nombre moyens de polynômes à tirer avant d'atteindre un polynôme irréductible [17] :

$$n \geqslant \frac{q^n}{I(n,q)} \geqslant n \times \left| \begin{array}{cc} \frac{1}{1-1/\sqrt{2}} \simeq 3.14 & \text{si } n \geqslant 3 \\ 4 & \text{sinon} \end{array} \right.$$

3. Test d'irréductibilité de Rabin

Pour mettre en œuvre notre démarche il faut un test efficace d'irréductibilité. Avec la formule précédente on en déduit le test suivant. Un polynôme P sur $\mathbf{F}_q$ unitaire de degré n est irréductible si et seulement si les deux conditions suivantes sont vérifiés :

1. P divise $X^{q^n} - X$;
2. P ne divise aucun $X^{q^d} - X$ où d divise n, condition qu'on peut restreindre en P ne divise aucun $X^{q^{n_i}} - X$ où les n_i sont les diviseurs maximaux de n.

Pour mettre en œuvre ce test on effectuera tous les calculs modulo P, la complexité est alors en $O_q(n^3 \log_3 n)$ où l'on cache dans le O le coût de la complexité des opérations dans $\mathbf{F}_q$. [18]

17. sachant que l'espérance d'une loi géométrique de paramètre p est $1/p$.

18. D'abord on calcule la matrice du Frobenius sur $\mathbf{F}_q[X]/(P)$. Cela demande de calculer n puissances [19] q-ème pour un total de $O_q(n^3)$ opérations, un produit de deux polynômes de degrés inférieur n demandant avec l'algorithme naïf moins de n^2 opérations. Ensuite on calcule $X^{q^n} - X$ modulo P ce qui demande d'itérer n fois le Frobenius sur le vecteur X soit un $O_q(n^3)$ opérations. De la même façon le calcul de $X^{q^{n_i}} - X$ demande un $O_q(n_i n^3)$ opérations. Finalement le calcul du pgcd demande un $O_q(n^3)$ opérations. Pour évaluer la complexité il reste à estimer $n_1 + \cdots + n_k = n \sum_{i=1}^k p_i^{-1}$ où les p_i sont les facteurs premiers distincts de n. Si l'on suppose les p_i triés par ordre croissante alors $p_i \geqslant q_i$ où (q_i) est la suite croissante des nombres premiers. On a ainsi par le second théorème de Mertens $\sum_{i=1}^k \frac{1}{p_i} \leqslant \sum_{i=1}^k \frac{1}{q_i} = \log_2 q_k + O(1)$. Maintenant par le théorème des nombres premiers $q_k \sim k \log k$, on a une une constante C indépendante de k tel que $q_k \leqslant Ck \log k$ d'où $\log_2 q_k \leqslant O(\log_2 k)$. Finalement $p_1 \cdots p_k$ divise n donc $2^k \leqslant n$ de sorte que $k \leqslant C \log n$ puis $\sum_{i=1}^k \frac{1}{p_i} \leqslant O(\log_3 n)$.

Décomposition caractéristique

Proposition. *Soit u un endomorphisme sur un K-ev avec un polynôme annulateur $P = P_1 \cdots P_k$ où les P_i sont des polynômes premiers entre eux alors :*

$$E = \mathrm{Ker} P_1(u) \circ \cdots \circ \mathrm{Ker} P_k(u).$$

1. Lemme des noyaux

Le résultat découle essentiellement de l'écriture d'une relation de Bézout pour les polynômes $Q_i = \prod_{j \neq i} P_j$. On a $Q_1, \cdots, Q_k$ premiers entre eux dans leur ensemble donc il existe des polynômes A_i tels que $A_1 Q_1 + \cdots + A_k Q_k = 1$. En évaluant en u on obtient : $\mathrm{Id}_E = A_1(u) \circ Q_1(u) + \cdots + A_k(u) \circ Q_k(u)$. Cette relation assure que les $p_i = A_i(u) \circ Q_i(u)$ forment une famille de projecteurs :

- $p_1 + \cdots + p_k = \mathrm{Id}_E$;
- si $i \neq j$ alors $p_i \circ p_j = 0$ car $p_i \circ p_j = (A_i A_j)(u) \circ (Q_i Q_j)(u)$ or $Q_i Q_j$ annule u.

Cela assure que $(p_1, \cdots, p_k)$ est un système de projecteurs : la première condition assure que $\mathrm{Im}(p_1) + \cdots + \mathrm{Im}(p_k) = F$ quand la deuxième assure la somme est direct : si $0 = x_1 + \cdots + x_k$ alors $0 = p_i(x_i)$ mais aussi $x_i = p_i(x_i)$ donc $x_i = 0$. Terminons en montrant que $\mathrm{Im} p_i = \mathrm{Ker} P_i(u)$.

- d'un côté si $x \in \mathrm{Im} p_i$ alors $P_i(x) = Q_i(p_i(x)) = (A_i P_i Q_i)(u) \cdot x = 0$;
- de l'autre si $x \in \mathrm{Ker} P_i(u)$ alors $p_j(x) = A_j(u) \circ Q_j(u) \cdot x = 0$ donc $x = p_i(x)$.

2. Décomposition caractéristique

En appliquant ce résultat à la décomposition de π_u en facteurs irréductibles $P_1^{\alpha_1} \cdots P_k^{\alpha_k}$ on obtient la décomposition en sous-espace caractéristique. On traite maintenant du cas particulier où π_u est scindée, dans ce cas $P_i(X) = (X - \lambda_i)$ où λ_i est une valeur propre de u et l'endomorphisme induit sur $E_i = \mathrm{Ker} P_i^{\alpha_i}$ s'écrit $\lambda_i \mathrm{Id}_{E_i} + n_i$ où n_i est nilpotent d'indice inférieur à α_i. Cette écriture est particulièrement adaptée au calcul sur les endomorphismes, par exemple pour l'exponentielle. On se place sur $\mathbf{K} = \mathbf{R}$ ou $\mathbf{C}$ alors sous les hypothèses précédentes :

$$\mathrm{e}^{tu} = \sum_{i=1}^{k} \mathrm{e}^{\lambda_i t} \left(\sum_{j=0}^{\alpha_i - 1} t^j \frac{n_i^j}{j!} \right) p_i.$$

Cette formule donne à la fois un moyen pratique pour calculer l'exponentielle et un outil théorique pour son étude, par exemple son étude asymptotique. On en déduit que e^{tu} converge vers l'endomorphisme nulle si et seulement si $\mathrm{Sp}\, u \subset \{\mathrm{Re} < 0\}$.

- Si u a une valeur propre λ de partie réelle $\geqslant 0$ alors pour x un vecteur propre associé on a $\|\mathrm{e}^{tu} x\| = |\mathrm{e}^{t\lambda}| \|x\|$ ne tend pas vers 0.
- Sinon par la formule précédente : $\|\mathrm{e}^{tu}\| \leqslant \mathrm{e}^{\sigma t} P(t)$ où $\sigma = \max_{\lambda \in \mathrm{Sp}\, u} \mathrm{Re}(\lambda) < 0$ et P est un polynôme en t.

Une étude plus fine permet de montrer que e^{tu} est bornée si et ssi toute valeur propre est de partie réelle et négative et toute valeur propre de partie réelle nulle est non défective.

3. Décomposition d+n

- On a précédemment écrit $u = d + n$ avec d diagonalisable, u nilpotente. De plus d et n sont des polynômes en u. Il suffit de le voir pour d, cela découle du fait que les p_i sont des polynômes en u. En particulier d et n commutent entre eux.
- Les propriétés $u = d + n$ avec d diagonalisable, n nilpotente et $dn = nd$ assure l'unicité de la décomposition. On se donne $u = d_1 + n_1$ une autre décomposition vérifiant les mêmes propriétés.

- d_1 commute avec u : $d_1u = d_1^2 + d_1n_1 = d_1^2 + n_1d_1 = ud_1$ et de la même façon n_1 commute avec u.
- comme d et n sont des polynômes en u, d_1 commute avec d et n_1 commute avec n.
- les endomorphismes d et d_1 sont ainsi co-diagonalisable, donc $d - d_1$ est diagonalisable.
- les endomorphismes n et n_1 sont co-trigonalisable donc $n - n_1$ est trigonalisable, or comme les diagonales de n et n_1 dans une base de trigonalisation sont nulles, pour $n - n_1$ aussi donc $n - n_1$ est nilpotent.

Ainsi $d - d_1 = n_1 - n$ est diagonalisable et nilpotent donc nulle car l'unique valeur propre d'un endomorphisme nilpotent est 0.

- Cette décomposition peut s'interpréter comme une mesure de l'écart à la diagonalisibilité. Elle est utile pour étudier la diagonalisabilité d'une transformation de notre endomorphisme dans la mesure où elle fournit un critère de diagonalisibilité : partie nilpotente nulle. Par exemple pour l'exponentielle on écrit $\mathrm{e}^u = \mathrm{e}^d\mathrm{e}^n = \mathrm{e}^d + \mathrm{e}^d(\mathrm{e}^n - 1)$ cette derrière écriture correspond à la décomposition d+n de e^u d'où l'on déduit que e^u est diagonalisable si et seulement si $\mathrm{e}^n - 1 = 0$. En écrivant cette dernière expression comme un polynôme en n connaissant la forme des polynômes annulateur de n on en déduit nécessaire que $n = 0$.

Remarque. *1. Pour calculer d il suffit de connaître P tel que $P(u) = d$. Un choix consiste à prendre P solution du système de congruence : $P \equiv \lambda_i \mod (X - \lambda_i)^{\alpha_i}$. En effet si P est un tel polynôme alors $P(X) = \lambda_i + (X - \lambda_i)^{\alpha_i}Q$ de sorte que sur K_i, $P(u)\cdot x = \lambda_i x$. Ainsi, $P(u) = \sum_{i=1}^k \lambda p_i$. En fait si l'on revient à la construction du lemme des noyaux $d = \sum_{i=1}^k \lambda_i(A_iP_i)(u) = P(u)$ avec $P = \sum_{i=1}^k A_iP_i$ qui est la solution donnée par le théorème chinois au système de congruence précédent.*

2. La formulation en terme de systèmes de congruences donne une stratégie pour calculer la décomposition de Dunford d'un endomorphisme. On peut mettre en œuvre cette stratégie à la main sur des petits exemples mais cette méthode se prête mal à la l'automatisation pour une première raison : le calcul numérique des racines d'un polynôme est un problème très compliqué. Une méthode plus adaptée effectue un parallèle avec la méthode de Newton. L'endomorphisme d de la décomposition de Dunford est la solution de l'équation $R(d) = 0$ où R désigne le radical de χ_u i.e. $\frac{\chi_u}{\chi_u \wedge \chi_u'}$ le produit de ses facteurs irréductibles sans multiplicité. Partant de la donnée initiale $d_0 = d + n$ on va alors considérer le schéma suivant :

$$d_{k+1} = d_k + P'(d_k)^{-1}P(d_k).$$

Décomposition de Frobenius

Le but de ce développement est d'établir la décomposition de Frobenius qui écrit un endomorphisme en somme direct d'endomorphismes cycliques. Cette décomposition peut se voir comme un raffinement de la décomposition en sous-espaces caractéristique. Nous nous mettrons ainsi "sur un sous-espace caractéristique" c'est à dire avec un endomorphisme u de polynôme minimal la puissance P^α d'un irréductible P.

Théorème. *Il existe une décomposition en sous-espaces u-stables $E = E_1 \oplus \cdots \oplus E_s$ où l'endomorphisme induit par u sur E_i est cyclique de polynôme minimal P^{α_i} avec la suite $\alpha_1 \geqslant \cdots \geqslant \alpha_s$ ne dépend pas de la décomposition choisie.*

Corollaire. *Soit u un endomorphisme sur un K-espace vectoriel de dimension finie E. Il existe une décomposition de E en sous-espaces u-stables $E = F_1 \oplus \cdots \oplus F_r$ où l'endomorphisme induit par u sur F_i est cyclique de polynôme minimal Q_i avec $Q_r \mid \cdots \mid Q_1$. Cette suite ne dépend pas de la décomposition choisie et est appelée suite des invariants de similitudes de u.*

1. Décomposition en sous-espaces cycliques

Le tout est de construire un supplémentaire stable à un sous-espace cyclique $F = K[u]\cdot x$ afin d'amorcer une récurrence. On va construire ce sous-espace par dualité c'est à dire à partir d'un ensemble équations. On introduit φ une forme linéaire tel que :

$$\varphi(u^{p-1}(x)) = 1, \quad \varphi(u^{p-2}(x)) = \cdots = \varphi(x) = 0.$$

Considérons alors $G = \Gamma^\circ$ où $\Gamma = \text{vect}(\varphi, {}^tu\circ\varphi, \cdots, {}^tu^{p-1}\circ\varphi)$ et montrons que G est un supplémentaire de F.

- si $x \in F \cap G$ écrivons $x = \sum_{k=0}^{p-1} \lambda_k u^k(x)$ alors $0 = \varphi(x) = \lambda_{p-1}$ puis $0 = \varphi(u(x)) = \lambda_{p-2}$ et ainsi de suite donc $x = 0$.
- $(\varphi, {}^tu\circ\varphi, \cdots, {}^tu^{p-1}\circ\varphi)$ est une famille libre : si $\sum_{k=0}^{p-1} \lambda_k {}^tu^k \circ \varphi = 0$ alors en évaluant en x on a $\lambda_{p-1} = 0$ puis en évaluant en $u(x)$ on a $\lambda_{p-2} = 0$ et ainsi de suite donc $\lambda_1 = \cdots = \lambda_k = 0$. Ainsi Γ est de dimension p donc G est de dimension $n - p$. Ainsi par la formule de Grassman G est un supplémentaire de F.

On voudrait finalement vérifier que G est u-stable. Pour cela il faut et il suffit que Γ soit stable par ${}^tu$, malheureusement ce n'est pas forcément le cas. Il faudrait pour cela savoir exprimer $\varphi\circ u^p$ en fonction des $\varphi\circ u^k$ avec $k < p$ ce que rien n'assure. Pour cela on va vouloir que $p = \deg \pi_u$ de sorte que u^p s'exprime comme combinaison linéaire des u^k avec $k < p$. Cela demande à prendre x tel que $\pi_{u,x} = \pi_u$. L'existence d'un tel x est direct avec l'hypothèse que π_u est la puissance P^k d'un irréductible P. En effet $\pi_{u,x}$ divise π_u donc il existe $k_x \leqslant k$ tel que $\pi_{u,x} = P^{k_x}$ alors si l'on note $\tilde{k} = \sup k_x = \max k_x$ on a $P^{\tilde{k}}(u)\cdot x = 0$ pour tout $x \in E$ d'où $P^k(u) = 0$ puis π_u divise $P^{\tilde{k}}$ ou encore $k \leqslant \tilde{k}$ or comme $k_x \leqslant k$ pour tout x on a aussi $\tilde{k} \leqslant k$ puis l'égalité.
Prenons alors x tel que $\pi_{u,x} = \pi$ on a ainsi construit un supplémentaire à F. On peut alors répéter l'opération sur G. Notons que $\pi_{u_{|G}}$ divise $\pi_{u_{|F}}$ de sorte qu'en itérant on construit une suite de sous-espaces cycliques $F_1, \cdots, F_k$ où le polynôme minimal $P_i = P^{\alpha_i}$ de u sur F_i est tel que $\alpha_1 \geqslant \cdots \geqslant \alpha_r$.

2. Unicité de la décomposition

Nous allons montrer l'unicité des exposants dans la décomposition. Supposons donnée deux décompositions de E en sous-espaces cycliques :

$$E = \bigoplus_{i=1}^{r} F_i = \bigoplus_{i=1}^{s} G_i$$

où le polynôme minimal de u sur F_i (resp. G_i) est P^{α_i} (resp. P^{β_i}) avec $\alpha_1 \geqslant \cdots \geqslant \alpha_r$ et $\beta_1 \geqslant \cdots \geqslant \beta_s$. Nous allons montrer que les suites α et β sont égales.

1. Déjà en regardant le polynôme minimal on voit que $\alpha_1 = \beta_1 = k$.
2. Ensuite en appliquant $P^{\alpha_2}(u)$ on trouve :

$$P^{\alpha_2}(u)\cdot E = P^{\alpha_2}(u)\cdot F_1 = \bigoplus_{i=1}^{s} P^{\alpha_2}(u)\cdot G_i.$$

   Or comme les endomorphismes $u_{|F_1}$ et $u_{|G_1}$ sont cycliques de même polynôme minimal, ils sont semblables (dans une base on a la même matrice compagnon), de sorte que $\dim P^{\alpha_2}(u)\cdot F_1 = \dim P^{\alpha_2}(u)\cdot G_1$. Il en découle que $P^{\alpha_2}(u)\cdot G_j = 0$ pour $j \geqslant 2$ i.e. $\alpha_2 \geqslant \beta_2$. Par symétrie on en déduit que $\alpha_2 = \beta_2$.
3. De proche en proche on en déduit le résultat. Pour être précis on peut raisonner de la façon suivante : on note j le premier indice à partir duquel $\alpha_i \neq \beta_i$. Un tel indice existe toujours car même si les suites n'ont pas la même longueur $\sum \alpha_i = \sum \beta_j$. Alors on applique comme précédemment $P^{\alpha_i}(u)$ pour obtenir :

$$P^{\alpha_j}(u)\cdot E = \bigoplus_{i=1}^{j-1} P^{\alpha_j}(u)\cdot F_i = \bigoplus_{i=1}^{s} P^{\alpha_j}(u)\cdot G_i.$$

   Par hypothèse pour $i < j$ les endomorphismes induits par u sur F_i et G_i sont semblables donc en particulier $\dim P^{\alpha_j}(u)\cdot F_i = \dim P^{\alpha_j}(u)\cdot G_i$. On en déduit que :

$$P^{\alpha_j}(u)\cdot G_j = \cdots = P^{\alpha_j}(u)\cdot G_s.$$

   Ainsi $\alpha_j \geqslant \beta_j$ et par symétrie $\alpha_j = \beta_j$ ce qui est absurde.

3. Réécriture de la décomposition

Pour l'existence on applique le théorème précédent aux endomorphismes induit sur les sous-espaces caractéristiques. On réordonne alors les blocs en utilisant le lemme suivant : $C_{PQ} \sim C_P \times C_Q$ si $P \wedge Q = 1$ qui se montre en montrant que les deux matrices ont même polynôme minimal et même polynôme caractéristique. On détaille davantage l'unicité où nous allons faire ce réordonnement à l'envers. Écrivons $Q_i = P_1^{\alpha_{i,1}} \cdots P_s^{\alpha_{i,s}}$ la décomposition en irréductibles des Q_i. En utilisant le lemme précédent on a :

$$C_{Q_i} \sim C_{P_1^{\alpha_{i,1}}} \times \cdots C_{P_1^{\alpha_{i,s}}}.$$

En réindéxant les blocs on en déduit que $u \sim M_1 \times \cdots \times M_s$ où $M_j = C_{P_j^{\alpha_{1,j}}} \cdots C_{P_j^{\alpha_{r,j}}}$. La somme des exposants est égal à la multiplicité α_j de P_j dans χ_u donc ne dépend pas de la décomposition choisie. Si l'on note $E = F_1 \oplus \cdots \oplus F_s$ la décomposition de E sous-jacente à cette représentation matricielle on voit alors que $F_i \subset \operatorname{Ker} P_i^{\alpha_j}$ le sous-espace caractéristique de u. Avec un argument de dimension cette inclusion est en fait une égalité, on en déduit que $M_j \sum u_j$ l'endomorphisme induit par u sur le sous-espace caractéristique de sorte que la classe de similitude de M_j ne dépend pas de la décomposition choisie. Finalement le résultat précédent assure que la suite $\alpha_{1,j} \geqslant \cdots \geqslant \alpha_{r,j}$ ne dépend pas de la base de F_i choisie.

Surjectivité exponentielle de matrices et application

On sait que l'exponentielle de matrices envoie $\mathcal{M}_d(\mathbf{C})$ sur $\mathrm{GL}_d(\mathbf{C})$ mais quelle est exactement son image ? On va voir que c'est précisément l'ensemble des matrices inversibles.

1. Surjectivité de l'exponentielle de matrices

Soit $A \in \mathrm{GL}_d(\mathbf{C})$. On peut espérer que les antécédents de A par exp soit polynomiale ce qui conduit à la restriction :

$$\exp : \mathbf{C}[A] \to \mathbf{C}[A] \cap \mathrm{GL}_d(\mathbf{C}) = \mathbf{C}[A]^\times$$

où la dernière égalité résulte du fait que l'inverse d'une matrice inversible M est un polynôme M ce qui se voit par exemple avec Cayley-Hamilton. Avec cette restriction $\mathbf{C}[A]$ devient une algèbre commutative donc on pourrait utiliser sans justification la propriété de morphisme $\exp(X+Y) = \exp(X)\exp(Y)$. On va montrer que cette application est surjective, l'argument est le suivant : $\mathbf{C}[A]^\times$ est connexe et $\mathscr{E} = \exp(\mathbf{C}[A]^\times)$ est ouvert et fermé.

- On raisonne comme pour la connexité de $\mathrm{GL}_d(\mathbf{C})$: soient $X, Y \in \mathbf{C}[A]^\times$, l'application $p : z \in C \mapsto \det(Az + (1-z)B)$ est un polynôme en z non identiquement nulle (car $p(0) \neq 0$) donc admet un nombre fini de racines. Ainsi il existe un chemin $z : [0,1] \to \mathbf{C}$ tel que $\gamma(0) = 0$ et $\gamma(1) = 1$ évitant les racines de p (0 et 1 n'étant pas racines par hypothèse). Alors $z(t)X + (1-z(t))Y$ est un chemin reliant X à Y dans $\mathbf{C}[A]^\times$. Ainsi, $\mathbf{C}[A]^\times$ est connexe par arcs.
- exp est C^1 de différentielle inversible en tout point donc est un C^1-difféomorphisme local.
  - On a $\exp(X+H) = \exp(X)\exp(H) = \mathrm{e}^X(1 + H + O(\|H\|^2))$ ainsi exp est différentiable en X de différentielle $H \mapsto \mathrm{e}^X H$.
  - Maintenant exp est C^1 car $\|\mathrm{d}\exp(X) - \mathrm{d}\exp(Y)\| \leqslant \|\mathrm{e}^X - \mathrm{e}^Y\|$ et $X \mapsto \mathrm{e}^X$ est continue.
  - Ensuite en tout point X la différentielle de exp en X est inversible d'inverse $H \mapsto \mathrm{e}^{-X}H$ (commutativité à l'arrivée).

  Par le théorème d'inversion locale il existe donc un voisinage V_X de X et un voisinage W_X de e^X tel que $\exp : V_X \to W_X$ soit un C^1-difféomorphisme. En particulier $W_X \subset \mathscr{E}$.
- $\exp : (\mathbf{C}[A], +) \to (\mathbf{C}[A]^\times, \times)$ est un morphisme de groupes donc $\mathscr{E}$ est un sous-groupe de $\mathbf{C}[A]^\times$ puis on peut écrire la décomposition en classes :

  $$\mathbf{C}[A]^\times \bigsqcup_{g \text{ rep}} g\mathscr{E}.$$

  Ainsi $\mathscr{E} = \mathbf{C}[A]^\times \setminus \bigsqcup_{g \neq 1} g\mathscr{E} = \bigcap_{g \neq 1}(\mathbf{C}[A]^\times \setminus g\mathscr{E})$ or $g\mathscr{E}$ est ouvert [20] donc $\mathbf{C}[A]^\times \setminus g\mathscr{E}$ fermé puis $\mathscr{E}$ est une intersection de fermés donc est fermé.

2. Application aux systèmes périodiques

Corollaire. *Soit $A \in \mathscr{C}^0_{T-per}(\mathbf{R}, \mathrm{GL}_d(\mathbf{C}))$. Il existe $Q \in \mathscr{C}^0_{T-per}(\mathbf{R}, \mathrm{GL}_d(\mathbf{C}))$ et $B \in \mathscr{M}_d(\mathbf{C})$ tels que les solutions de $X'(t) = A(t)X(t)$ s'écrivent $Q(t)\mathrm{e}^{tB}X_0$.*

On introduit R la résolvante du système de sorte que les solutions s'écrivent $X(t) = R(t)X_0$. Si une telle décomposition existe alors $R(T) = \mathrm{e}^{TB}$. Partons de là.

- Par surjectivité de l'exponentielle de matrices il existe $B \in \mathscr{M}_d(\mathbf{C})$ tel que $R(T) = \mathrm{e}^{TB}$. Posons $Q(t) = R(t)\mathrm{e}^{-tB}$.
- Vérifions que Q est T-périodique. On a $Q(t+T) = R(t+T)\mathrm{e}^{-(t+T)B}$ or $R(t+T) = R(t)R(T)$ par la propriété de flot [21] de sorte que :

  $$Q(t+T) = R(t)R(T)\mathrm{e}^{-TB}\mathrm{e}^{-tB} = R(t)\mathrm{e}^{-tB} = Q(t).$$

20. En tant qu'image d'un ouvert par une application continue

21. La position d'une solution au temps $t+T$ est égal à la position au temps t de la solution partant en $t=0$ de la position de la première solution au temps T.

Ainsi les solutions d'un système périodique s'exprime à un terme périodique près comme les solutions d'un système linéaire. Cette écriture est notamment utile lorsque l'on cherche à analyser la stabilité d'un système périodique, on se ramène ainsi à analyser le système linéaire associée. Ainsi ;

1. 0 est AS si et seulement si $\mathrm{Sp}(B) \subset \{\mathrm{Re} < 0\}$ i.e. $\mathrm{Sp}\, R(T) \subset D(0,1)$;
2. 0 est S si et seulement si $\mathrm{Sp}\, R(T) \subset \overline{D}(0,1)$ et toute valeur propre de module 1 est non dégénérée.

Formule de Gram

1. Formule de Gram

Soit $(u_1,\cdots,u_p)$ une base F. La projection orthogonal de x sur F est l'unique vecteur $y\in F$ solution du système d'équations : $u_i\cdot(x-y)=0,\ i=1,\cdots,p$. Si l'on note $y=\sum_{j=1}^p y^j u_j$ ces équations se réécrivent $\sum_{j=1}^p y^j(u_i\cdot u_j)=u_i\cdot x,\ i=1,\cdots,p$ c'est à dire matriciellement :

$$\begin{pmatrix} u_1\cdot u_1 & \cdots & u_1\cdot u_p \\ \vdots & & \vdots \\ u_p\cdot u_1 & \cdots & u_p\cdot u_p \end{pmatrix}\begin{pmatrix} y^1 \\ \vdots \\ y^p \end{pmatrix}=\begin{pmatrix} u_1\cdot x \\ \vdots \\ u_p\cdot x \end{pmatrix}.$$

Ainsi à l'aide des formules de Cramer on peut en déduire une expression de y. Maintenant en pratique on s'intéresse plus à $\delta=\|x-y\|^2$ qui mesure la distance de x à F. Notons que

$$\delta=\langle x-y,x-y\rangle=\langle x-y,x\rangle=\|x\|^2-\langle x,y\rangle=x\cdot x-\sum_{i=1}^p y_i(u_i\cdot x).$$

On peut ajouter cette équation à nos précédentes, en voyant δ comme un nouveau paramètre, on a ainsi :

$$\begin{pmatrix} u_1\cdot u_1 & \cdots & u_1\cdot u_p & 0 \\ \vdots & & \vdots & 0 \\ u_p\cdot u_1 & \cdots & u_p\cdot u_p & 0 \\ x\cdot u_1 & \cdots & x\cdot u_p & 1 \end{pmatrix}\begin{pmatrix} y^1 \\ \vdots \\ y^p \\ \delta \end{pmatrix}=\begin{pmatrix} u_1\cdot x \\ \vdots \\ u_p\cdot x \\ x\cdot x \end{pmatrix}$$

On peut alors utiliser les formules de Cramer pour exprimer δ à l'aide du déterminant.

2. Un exemple de calcul de distance

On peut utiliser ce résultat pour évaluer la distance d'un point à un sous-espace dans certains cas où le calcul du déterminant est classique. On va développer le cas des fonctions polynomiales sur $[0,1]$ où pour le produit scalaire usuel $\langle f,g\rangle:=\int_0^1 fg$ on a $\langle x^i,x^j\rangle=(i+j+1)^{-1}$ ce qui fera apparaître des déterminants de Cauchy. On se donne ainsi des réels positifs distincts[22] $a_1,\cdots,a_n,a_{n+1}$ et l'on va monter que :

$$\delta:=d(x^{a_{n+1}},\mathrm{vect}(x^{a_1},\cdots,x^{a_n}))^2=\frac{1}{2a_{n+1}+1}\prod_{i=1}^n\left|\frac{a_i-a_{n+1}}{a_i+a_{n+1}-1}\right|^2.$$

Les calculs sont les suivant :

$$\begin{aligned}\delta^2&=\left|\frac{1}{a_i+a_j+1}\right|_{1\leqslant i,j\leqslant n+1}\left|\frac{1}{a_i+a_j+1}\right|^{-1}_{1\leqslant i,j\leqslant n}\\&=\frac{\prod_{1\leqslant i<j\leqslant n+1}(a_j-a_i)^2}{\prod_{1\leqslant i,j\leqslant n+1}(a_i+a_j+1)}\frac{\prod_{1\leqslant i,j\leqslant n}(a_i+a_j+1)}{\prod_{1\leqslant i<j\leqslant n}(a_j-a_i)^2}\\&=\frac{1}{2a_{n+1}+1}\prod_{i=1}^n\left|\frac{a_i-a_{n+1}}{a_i+a_{n+1}+1}\right|^2.\end{aligned}$$

3. Application au théorème de Muntz

On peut déduire de ce calcul à quelle condition une famille de monôme $(x^{a_n})_{n\in\mathbf{N}}$ est totale dans $L^2([0,1],\mathbf{R})$:

22. Si les réels ne sont pas distincts les manipulations suivantes font intervenir des divisions par 0 et la formule résultante est fausse.

Soit $(a_n)_{n\in\mathbf{N}}$ une suite de réels positifs distincts de limite $+\infty$, la famille $(x^{a_n})_{n\in\mathbf{N}}$ est totale dans $L^2([0,1],\mathbf{R})$ si et seulement si $\sum_{\substack{n\in\mathbf{N}\\ a_n\neq 0}} \frac{1}{a_n} = +\infty$

En effet, par le théorème de Weierstrass cette famille est totale si et seulement si x^k est dans l'adhérence de vect$(x^{a_n} : n \in \mathbf{N})$ pour tout $k \in \mathbf{N}$ ce qui est équivalent à ce que :

$$d(x^k, \text{vect}(x^{a_1}, \cdots, x^{a_n})) = \frac{1}{2k+1} \prod_{i=1}^{n} \left| \frac{a_i - k}{a_i + k + 1} \right|^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Supposons $k \notin \{a_n\}$ auquel cas le résultat est évident. Cette quantité tend vers 0 avec n si et seulement son logarithme diverge vers $-\infty$, logarithme qui est :

$$-\log(2k+1) + 2\sum_{i=1}^{n} \log\left|1 - \frac{2k+1}{a_i + k + 1}\right|.$$

Mais comme $a_n \to +\infty$ on a :

$$\log\left|1 - \frac{2k+1}{a_n + k + 1}\right| \underset{n\to\infty}{\sim} -\frac{2k+1}{a_n + k + 1} \underset{n\to\infty}{\sim} -\frac{2k+1}{a_n}$$

de sorte la divergence de la série des logarithmes précédente est équivalente à la divergence de la série $\sum_{a_n\neq 0} \frac{1}{a_n}$.

Remarque. *1. Lorsque la suite (a_i) ne tend pas vers $+\infty$ la famille est toujours totale car on peut extraire de (a_i) une sous-suite borné donc une sous-suite convergente ce qui montre que $\left|1 - \frac{2k+1}{a_i+k+1}\right|$ ne tend vers 1 de sorte que la série précédente diverge grossièrement. De plus comme $\left|\frac{a_i-k}{a_i+k+1}\right| \leqslant 1$, elle diverge vers $-\infty$.*

2. Et pour un critère de densité $\|\cdot\|_\infty$? Il faut déjà ajouter $a_0 = 0$ car $\{f \mid f(0) = 0\}$ est un sous-espace fermé de $\mathscr{C}^0([0,1])$. On se ramène alors à des fonctions nulles en 0, alors en utilisant la majoration $\|f\|_\infty \leqslant \|f'\|_2$ on va pouvoir déduire la densité par la densité de la suite des dérivées. Supposons alors $a_n \geqslant 1$ pour $n \geqslant 1$, on a applique à (x^{a_n}) le critère précédent sur sa suite dérivée. Avec l'inégalité précédente on a donc que toute fonction $f \in C^1$ est dans l'adhérence de (x^{a_n}) pour la norme infini. On a ainsi montré que sous l'hypothèse $a_0 = 1$ et $a_n \geqslant 1$ pour $n \geqslant 1$ la densité de (x^{a_n}) équivaut à la divergence de $\sum_{n=1}^{\infty} \frac{1}{a_n}$. Finalement on peut remplacer la condition $a_n \geqslant 1$ par la condition $\inf_{n\geqslant 1} a_n > 0$. On remplace pour cela la suite a_n par $\frac{a_n}{m}$ où m désigne l'infimum précédent, sur les approximation il faut alors composer avec la fonction $x \mapsto x^{1/m}$.

SVD et applications

1. Décomposition en valeurs singulières

Théorème. *Pour $A \in \mathscr{M}_{n,p}(\mathbf{R})$ il existe $(U,V) \in \mathrm{O}_n(\mathbf{R}) \times \mathrm{O}_p(\mathbf{R})$ tels que $A = U\Sigma V^\top$ avec $V = \mathrm{diag}(\sigma_1,\cdots,\sigma_r,0,\cdots,0)$ où les σ_i sont des réels strictement positifs qu'on peut prendre ordonnés $\sigma_1 \geqslant \cdots \geqslant \sigma_r > 0$.*

- Si une telle décomposition existe on a $A^\top A = V(\Sigma^\top\Sigma)V^\top$. Cette décomposition est la diagonalisation de la matrice symétrique A en base orthonormée. On part de cette idée en reprenant dans l'ordre.
- La matrice $A^\top A \in \mathscr{M}_p(\mathbf{R})$ est symétrique positive donc il existe $V \in \mathrm{O}_p(\mathbf{R})$ tel que $A^\top A = VDV^\top$ avec $D = \mathrm{diag}(d_1,\cdots,d_r,0,\cdots,0)$ où $d_1 \geqslant \cdots \geqslant d_r > 0$.
- L'entier r représente le rang de $A^\top A$ c'est donc aussi le rang de A. En particulier $r \leqslant \min(n,p)$ ce qui permet de définir en notant $\sigma_i = \sqrt{d_i}$

$$\Sigma = \begin{pmatrix} \Delta & \mathbf{0}_{r,p-r} \\ \mathbf{0}_{n-r,p} & \mathbf{0}_{n-r,p-r} \end{pmatrix} \qquad \text{où} \quad \Delta = \begin{pmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{pmatrix}.$$

  Ainsi défini on a $\Sigma^\top\Sigma = D$.
- On veut finalement construire une matrice $U \in \mathrm{O}_n(\mathbf{R})$ tel que $AV = U\Sigma$. Cela conduit aux équations suivantes sur les vecteurs colonnes :

$$Av_i = \sigma_i u_i \ (1 \leqslant i \leqslant r) \qquad Av_i = 0 (r+1 \leqslant i \leqslant p).$$

  Ces dernières équations sont déjà vérifiés car $v_i \in \ker A^\top A = \ker A$ pour $i \geqslant r+1$. Pour résoudre les autres équations on pose $u_i = \frac{1}{\sigma_i} Av_i$, cela définit les vecteurs u_i pour $1 \leqslant i \leqslant r$ vérifions qu'on a déjà une famille orthonormée :

$$\langle u_i, u_j \rangle = \frac{1}{\sigma_i\sigma_j}\langle Av_i, Av_j\rangle = \frac{1}{\sigma_i\sigma_j}\langle A^\top Av_i, v_j\rangle = \frac{\sigma_i}{\sigma_j}\langle v_i, v_j\rangle = \delta_{i,j}.$$

  Finalement il reste le choix des u_i d'indice $i \geqslant r+1$, pour cela on complète $(u_1,\cdots,u_r)$ en une base orthonormée quelconque. Ainsi la matrice $U = (u_1 | \cdots | u_n)$ est orthogonal et vérifie $AV = U\Sigma$ ce qui conclut la démonstration.

2. Approximation par des matrices de petits rangs

La caractérisation du rang par le déterminant montre que l'adhérence des matrices de rang k est l'ensemble des matrices de rang inférieur à k. Ainsi avec une matrice de rang k on ne pourra approximation aussi bien que le veut une matrice de rang strictement supérieur à k ; se pose alors le problème de l'approximation du matrice A par des matrices de rang inférieur. Ce problème est d'importance pratique, car si A est très grande, le stockage d'une telle matrice est très couteux et il sera souhaitable de réduire le poids en compressant A, mais pour ne pas trop détériorer les données il faut une bonne approximation de A.[23] En reprenant les notations précédentes avec $k < r$ on va montrer que :

Corollaire. *En norme 2 et en norme de Frobenius*

$$\min_{\substack{B \in \mathscr{M}_{n,p}(\mathbf{R}) \\ \mathrm{rg}(B)=k}} \|A - B\| = \|A - A_k\|$$

où A_k désigne la matrice $U\Sigma_k V^\top$ avec $\Sigma_k = \mathrm{diag}(\sigma_1,\cdots,\sigma_k,0,\cdots,0)$.

23. Il me semble que c'est ce que l'on fait avec les images numériques.

On commence par montrer le résultat pour la norme 2 :

- En utilisant que $\|UMV\|_2 = \|M\|_2$ si U et V sont orthogonales et que UMV est de rang $\mathrm{rg}(M)$ si U et V sont inversibles on se ramène à $A = \Sigma$ et $A_k = \Sigma_k$.
- Montrons d'abord que pour une matrice B de rang k, $\|\Sigma - B\|_2 \geqslant \sigma_{k+1}$. On utilise pour cela que $\ker B$ est de dimension $n-k$ donc $\ker B \cap \mathrm{vect}(e_1, \cdots, e_{k+1})$ est non trivial (en notant e_i les vecteurs de la base canonique). Si $x = \sum_{i=1}^{k+1} x_i e_i$ est un vecteur non nulle dans cette intersection on a :
$$\|\Sigma x - Bx\|_2^2 = \|\Sigma x\|_2^2 = \sum_{i=1}^{k+1} \sigma_i^2 |x_i|^2 \geqslant \sigma_{k+1}^2 \|x\|_2^2.$$
- Montrons ensuite que $\|\Sigma - \Sigma_k\|_2 \leqslant \sigma_{k+1}$. On a pour cela $(\Sigma - \Sigma_k)x = \sigma_{k+1}x$ en prenant $x = e_{k+1}$ ce qui conclut.

On continue avec la norme de Frobenius :

- Avec la décomposition en valeurs singulières on a $\|A\|_F^2 = \sum_{i=1}^r \sigma_i(A)$. Il s'agit donc de montrer que pour toute matrice B de rang k,
$$\|A - A_k\|_F^2 = \sum_{i=k+1}^{r} \sigma_i(A) \leqslant \sum_{i=1}^{r} \sigma_i(A - B).$$
- On va utiliser les inégalités de Weyl qui découle des formulations variationnels :
$$\sigma_{i+j-1}(A+B) \leqslant \sigma_i(A) + \sigma_j(B).$$
Si B est une matrice de rang k on a $\sigma_{k+1}(B) = 0$ donc avec $j = k+1$ on trouve :
$$\sigma_{i+k}(A - B) \leqslant \sigma_i(A).$$
Ainsi,
$$\|A - B\|_F^2 = \sum_{i=1}^{n} \sigma_i(A-B)^2 \geqslant \sum_{i=1}^{n-k} \sigma_i(A-B)^2 \geqslant \sum_{i=k+1}^{n} \sigma_i(A)^2.$$

3. Résolution d'un problème des moindres carrés

Soit $A \in \mathscr{M}_{n,p}(\mathbf{R})$ et $b \in \mathbf{R}^n$. On cherche $x \in \mathbf{R}^n$ tel que $\|Ax - b\|_2$ soit minimal. Pour cela écrivons la décomposition en valeurs singulières : $A = U\Sigma V^\top$. En posant $y = V^\top x$ et $c = U^\top b$ et en utilisant que $\|\cdot\|_2$ est invariante par l'action de matrices orthogonales on a :

$$\|Ax - b\|_2 = \|\Sigma y - c\|_2.$$

Alors les vecteurs y minimisant le dernier terme sont exactement ceux de la forme :

$$(\sigma_1^{-1} c_1, \cdots, \sigma_r^{-r} c_r, y_{r+1:p}).$$

De plus celui de norme minimal est obtenue pour $y_{r+1:p} = 0$ et correspond à $\Sigma^\dagger c$ où :

$$\Sigma^\dagger = \begin{pmatrix} \Delta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}.$$

Ainsi la solution à notre problème de norme minimal est $V\Sigma^\dagger U^\top b$.

Réduction des formes quadratiques

1. Théorème d'inertie de Sylvester

- Commençons par montrer qu'on peut se ramener à une forme quadratique non dégénérée. Pour cela on se donne F un supplémentaire quelconque de $\mathrm{Ker}q$, alors par définition de $\mathrm{Ker}q$ la somme direct $E = \mathrm{Ker}q \oplus F$ est q-orthogonal. Cela permet de se ramener à $q_{|F}$ donc à une forme quadratique non dégénérée.[24]
- La construction se fait alors par récurrence sur la dimension n de l'espace et se base sur le schéma suivant : si x est un vecteur non isotrope de q alors $\langle x\rangle \oplus_{\perp_q} \langle x\rangle^{\perp_q}$. En effet comme x est non isotrope $\langle x\rangle \cap \langle x\rangle^{\perp_q} = \{0\}$ d'où l'on déduit immédiatement que les deux espaces sont supplémentaires par les items du plan. La récurrence se rédige alors de la façon suivante, on y montre le postulat suivant : toute forme quadratique non dégénérée sur un espace de dimension n admet une base q-orthogonal. L'initialisation de la récurrence est immédiate, pour l'hérédité on se donne x un vecteur non isotrope qui existe car q est non nulle car non dégénérée, alors $\langle x\rangle \oplus \langle x\rangle^{\perp_q} = E$ puis la forme quadratique induite par q sur $\langle x\rangle^{\perp_q}$ est non dégénérée donc l'hypothèse de récurrence assure l'existence d'une base q-orthogonal de $\langle x\rangle^{\perp_q}$. En ajoutant x on obtient une base q-orthogonal de E.
- La démonstration précédente est valide dans le cas d'un corps quelconque, maintenant on va se restreindre au cas réelle. Dans une base q-orthogonal $(e_1,\cdots,e_n)$ la forme quadratique q admet en coordonnées l'expression suivante $q(x_1,\cdots,x_n) = a_1x_1^2 + \cdots + a_nx_n^2$ où les a_i sont non nuls. Le nombre t de $a_i > 0$ et le nombre s de $a_i < 0$ ne dépend pas de la décomposition choisie c'est la signature de q. On donne la caractérisation géométrique suivante :

  $$t = \max\{\dim F \ : \ q_F \text{ définie positive}\}.$$

  Quitte à réordonner les vecteurs de la base supposons que $a_1,\cdots,a_s > 0$ puis $a_{s+1},\cdots,a_{s+t} < 0$ et $a_{s+t+1},\cdots,a_n = 0$. Dans ce cas $F = \langle e_1,\cdots,e_s\rangle$ est un s.e.v de dimension s sur lequel q est définie positive. Maintenant si F est un s.e.v de dimension $s+1$ alors $F \cap \langle e_{s+1},\cdots,e_n\rangle$ est non vide, prenons x dans l'intersection alors $q(x) = \sum_{i=s+1}^{n} x_i^2 a_i q(e_i) \leqslant 0$ donc $q_{|F}$ n'est pas définie positive.
- Quitte à remplacer e_i par λe_i où $\lambda \in K^*$ on peut remplacer a_i par $a_i\lambda^2$. Cela montre qu'on peut prescrire la valeur des a_i dans un représentant de $K^\times/K^{\times 2}$. Par exemple sur $\mathbf{R}$ on peut demander aux a_i non nuls d'être égaux à ± 1.

2. Théorème spectral

On suppose que E est munie d'un produit scalaire $\langle\cdot|\cdot\rangle$. Dans ce cas on va montrer que l'on peut prendre la base simultanée q-orthogonal et orthogonal pour le produit scalaire ambiant. Le tout est de montrer qu'on peut prendre x non isotrope tel que $\langle x\rangle^{\perp_q} = \langle x\rangle^{\perp}$ où sans indice le symbole $\perp$ désigne l'orthogonalité pour le produit scalaire. Cette égalité de deux hyperplans correspond à la proportionnalité des formes linéaires dont elles sont le noyau à savoir $f(x,\cdot)$ et $\langle x|\cdot\rangle$ où f désigne la forme polaire de q. Maintenant on peut voir ce problème comme la condition d'Euler d'un problème de minimisation sous contrainte à savoir : maximiser q sur $\{x \ : \ \|x\| = 1\}$. Par compacité de la boule unité ce problème admet une unique solution ce qui fournit l'existence de x. On peut visualiser géométriquement cette démonstration avec la quadrique définie par q. Pour terminer mentionnons qu'avec ce raffinement on dispose comme précédemment d'un degré de liberté supplémentaire qu'on peut utiliser pour orthonormaliser les vecteurs de la base construite ce qui fait apparaître les valeurs propres comme coefficients diagonaux de q OU comme précédemment prescrire les coefficients diagonaux.

24. Si l'on possède une base q-orthogonal de F on complète avec une base quelconque de $\mathrm{Ker}q$ ce qui fera bien une base q-orthogonale de E.

3. Décomposition de Cholesky

On peut voir le théorème spectral comme un raffinement du théorème de Sylvester où l'on impose une forme particulière à la matrice de changement de base à savoir un changement de base orthogonal. On va voir un autre cas où l'on dispose d'un degré de liberté supplémentaire pour imposer une forme particulière à la matrice de changement de base. On suppose ainsi q anisotrope ce qui dans le cas réelle correspond à une forme quadratique définie positive ou définie négative. On va alors pouvoir construire le changement de base afin d'avoir une matrice de changement de base triangulaire supérieure. Dans la suite on note $(E_1,\cdots,E_n)$ une base préalablement.

- On peut prendre $e_1 = E_1$ car q est anisotrope.
- Ensuite il faut $e_2 \in \langle E_1,E_2\rangle$ tout en ayant $E_2 \in \langle E_1\rangle^{\perp_q}$. Cela est possible car $\dim\langle E_1,E_2\rangle \cap \langle E_1\rangle^{\perp_q} \geqslant 1$ en tant qu'intersections de sous-espaces de dimension 2 et $n-1$.
- On continue alors en prenant $e_3 \in \langle E_1,E_2,E_3\rangle \cap \langle e_1,e_2\rangle^{\perp_q}$ ce qui est possible car ce dernier espace est de dimension supérieur à 1 et tant qu'intersections de sous-espaces de dimensions 3 et $n-2$ et ainsi de suite...
- Notons de plus qu'on peut toujours changer e_i en λe_i et ainsi comme précédemment prescrire les valeurs des $q(e_i)$ non nuls dans un système de représentant de $K^\times$ modulo les carrés.

En passant en représentation matricielle on a établit la décomposition de Cholesky.

Remarque. *Cette étude montre qu'on peut ramener dans le cas d'une matrice définie positive l'équation $A = P^\top P$ à n^2 inconnues à une équation à $n(n+1)/2$ inconnues. On s'attend que cette restriction simplifie le problème et c'est effectivement le cas. L'algorithme classique pour rendre effectif le critère de Sylvester est l'algorithme de Gauss est requiert asymptotiquement entre $\frac{n^3}{3}$ et $\frac{2n^3}{3}$ opérations tandis qu'avec l'écriture de Cholesky on peut trouver un algorithme s'effectuant asymptotiquement en $\frac{n^3}{3}$ opérations.*

Disques de Gershgorin

Proposition. *Les valeurs propres d'une matrice $A \in \mathcal{M}_d(\mathbf{R})$ sont comprises dans les disques :*

$$\overline{D}_i := \overline{D}\left(a_{ii}, \sum_{j\neq i} |a_{i,j}|\right).$$

De plus dans chaque composante connexe de $\bigcup_{i=1}^d \overline{D}_i$ on compte autant de valeurs propres de A que de nombres de disques $\overline{D}_i$ dans la composante connexe.

Lemme. *Soit K un compact inclus dans un ouvert Ω. Il existe un cycle Γ inclus dans Ω disjoint de K tel que :*

$$\mathrm{Ind}(z,\Gamma) = \begin{cases} 0 & si \quad z \notin \Omega \\ 1 & si \quad z \in K \end{cases}$$

1. Disques de Gershgorin

On montre la première assertion du théorème. Soit λ une valeur propre de A et x un vecteur propre associé. De $(Ax)_i = \lambda x_i$ on tire $(a_{ii} - \lambda)x_i = \sum_{j\neq i} a_{i,j}x_j$ d'où :

$$|a_{i,i} - \lambda||x_i| \leqslant \sum_{j\neq i} |a_{i,j}||x_j| \leqslant \|x\|_\infty \sum_{j\neq i} |a_{i,j}|.$$

En prenant i tel que $|x_i|$ soit maximale on en déduit le résultat après simplification.

2. Localisation dans les disques

L'argument va consister à déformer la matrice A pour la réduire à sa diagonale et de montrer que le nombre de valeurs propres dans les composantes connexes des disques de Gershgorin de A reste constant au cours de cette déformation or à la fin les valeurs propres de la matrice sont les coefficients diagonaux de A et le résultat en résulte. Posons :

- $C_1, \cdots, C_s$ les composantes connexes de $X = \bigcup_{i=1}^d \overline{D}_i$. Comme X est fermé borné les composantes connexes également donc sont compactes.
- A_t la matrice $(1-t)A + tA_0$ où A_0 est la partie diagonale de A. Notons que les disques de Gershgorin de A_t sont contenues dans ceux de A.

Fixons une composante connexe disons C_1 et notons $n(t) = \#\mathrm{Sp}(A_t) \cap C_1$ il s'agit de montrer que n est constant. Heuristiquement cela résulte d'une forme de continuité des valeurs propres et d'un argument de connexité ; on pourra en effet déduire ce résultat d'un résultat général de continuité des valeurs propres mais plutôt on utilise une démarche mieux adaptée au problème qui utilise l'analyse complexe pour exprimer la quantité $n(t)$. On va vouloir appliquer la formule des résidus à $\frac{\chi'_{A_t}}{\chi_{A_t}}$ dont les pôles sont exactement les valeurs propres de A_t. Pour ne conserver que les valeurs propres dans C_1 il faut trouver un cycle qui entoure simplement tout point de C_1 et n'entoure aucun points des C_i. On utilise pour cela le lemme de séparation dans le plan.

1. *Construction du cycle.* On utilise $\delta = \inf_{i\neq j} d(C_i, C_j) > 0$ car ces parties étant compacts les distances sont atteintes puis on applique le lemme de séparation avec $K = C_1$ et $\Omega = C_1 + B_{\delta/2}$. On a ainsi un cycle Γ disjoint des composantes connexes tel que pour $z \in C_i$ on a $\mathrm{Ind}(z,\Gamma) = \delta_{1,i}$
2. *Application de la formule des résidus.* Comme les disques de Gershgorin de A_t sont inclus dans ceux de A quelque soit t le support de Γ ne contient aucune valeur propre de A_t. Ainsi on peut appliquer le théorème des résidus qui nous donne :

$$n(t) = \frac{1}{2\mathrm{i}\pi} \int_\Gamma \frac{\chi'_{A_t}(z)}{\chi_{A_t}(z)} \mathrm{d}z.$$

3. *Continuité et conclusion.* La fonction $(t,z) \mapsto \frac{\chi'_{A_t}(z)}{\chi_{A_t}(z)}$ est continue sur le compact $[0,1] \times \Gamma^*$ donc est borné. On a ainsi une domination et par le théorème de convergence dominée on en déduit que n est continue sur $[0,1]$. Étant à valeurs entières et $[0,1]$ étant connexe elle est constante donc en particulier $n(1) = n(0)$ ce qui est précisément le résultat qu'on cherche à démontrer.

3. Le lemme de séparation

Si le temps le permet on peut discuter de la démonstration formelle du lemme de séparation bien que ce résultat nous semble intuitivement évident. On se réfère à Rudin.

Remarque. *Ces résultats peuvent être utilisés en analyse numérique pour établir des critères d'arrêt dans des algorithmes de calcul numérique de valeurs propres comme la méthode QR.*

Par 5 points passe une conique

Soient A,B,C,D,E 5 points distincts d'un plan affine $\mathscr{E}$.

1. Il existe une conique passant par ces 5 points ;
2. Cette conique est unique si et seulement si 4 points parmi 5 sont toujours non alignés ;
3. Elle est non dégénérée si et seulement si 3 points parmi 5 sont toujours non alignés.

1. Existence de la conique

Si A,B,C,D,E sont alignés la droite passant par ces 5 points est une conique solution. Supposons maintenant A,B,C,D,E non alignés. Parmi ces 5 points, il y a en a 3 qui ne sont pas alignés et forment donc un repère affine. Quitte à numéroter nos points on suppose que A,B,C forment un repère affine. Écrivons alors l'équation d'une conique dans le repère barycentrique (A,B,C). La forme générale de cette équation est la suivante :

$$ax^2+by^2+cz^2+pxy+qyz+rxz=0.$$

Pour que A,B,C soit sur cette conique il faut et il suffit que $a=b=c=0$. Ensuite $D=(x_1,y_1,z_1)$ et $E=(x_2,y_2,z_2)$ sont sur cette conique si et seulement si :

$$\begin{cases} px_1y_1+qy_1z_1+rx_1z_1 &= 0\\ px_2y_2+qy_2z_2+rx_2z_2 &= 0 \end{cases}$$

Il s'agit de montrer qu'il existe (p,q,r) tels que ces deux équations soient satisfaites. On regarde ainsi les équations précédentes comme un système de deux équations d'inconnues (p,q,r), ce système est de rang $\leqslant 2$ donc admet une solution ce qui permet de conclure.

2. Unicité de la conique

Dire que la conique passant par A,B,C,D,E est unique revient à dire que le système précédent est de rang exactement égal à 2. En effet si le système est de rang 2 alors l'ensemble des solutions est une droite vectorielle ce qui correspond à une unique conique. Inversement si le système est de rang <2 le système possède deux solutions indépendantes qui conduisent à deux coniques différentes.
Maintenant le système est de rang <2 si et seulement si tous les mineurs d'ordre 2 sont nuls c'est à dire :

$$\begin{vmatrix} x_1y_1 & y_1z_1\\ x_2y_2 & y_2z_2 \end{vmatrix}=0,\quad \begin{vmatrix} y_1z_1 & x_1z_1\\ y_2z_2 & x_2z_2 \end{vmatrix}=0,\quad \begin{vmatrix} x_1y_1 & x_1z_1\\ x_2y_2 & x_2z_2 \end{vmatrix}=0.$$

On peut développer ces expression, par exemple $\begin{vmatrix} x_1y_1 & y_1z_1\\ x_2y_2 & y_2z_2 \end{vmatrix}=y_1y_2\begin{vmatrix} x_1 & z_1\\ x_2 & z_2 \end{vmatrix}$ pour le premier mineur ; on notera $\delta_y=\begin{vmatrix} x_1 & z_1\\ x_2 & z_2 \end{vmatrix}$ et on notera de façon équivalente δ_x et δ_z. Ainsi le système est de rang <2 si et seulement si :

$$x_1x_2\delta_x=y_1y_2\delta_y=z_1z_2\delta_z=0.$$

Maintenant dire que $\delta_x=0$ revient à dire que les vecteurs $\overrightarrow{AD}$ et $\overrightarrow{AE}$ sont liés c'est à dire que les points A,D,E sont alignés. Ainsi deux quantités parmi δ_x,δ_y et δ_z sont nuls entraine que 4 points parmi A,B,C,D,E sont alignés.
Supposons maintenant que deux δ ne sont pas nuls disons δ_x et δ_y. Dans ce cas on a $x_1x_2=0$ et $y_1y_2=0$. Prenons par exemple $x_1=0$, alors $y_1,z_1\neq 0$ car sinon le point $E=A$ or les points sont supposés distincts. Il en découle que $y_2=0$ et de même $x_2,z_2\neq 0$. De la non nullité de z_1 et z_2 il vient que $\delta_z=0$ c'est à dire que les points C,D,E sont alignés. Or comme $x_1=0$ on a D sur la droite (BC) ainsi B,C,D,E sont alignés. En fait comme $y_2=0$ on a aussi E sur la droite (AC) donc

A,B,C,D,E sont alignés ce qui contredit l'hypothèse sur les faites sur les δ, mais bon soit.
On a ainsi montré que si le système possède plusieurs solutions alors 4 points parmi A,B,C,D,E sont alignés. Inversement supposons par exemple que A,B,C,D sont alignés alors toute conique formé par la droite passant par A,B,C,D et une droite passant par E est solution, il n'y a donc pas unicité.

3. Dégénérescence de la conique

On suppose qu'il n'y a pas 4 points de A,B,C,D,E alignés, il y a donc unicité de la conique. Cette conique est dégénérée s'il est formé d'une ou deux droites, cela implique que 3 points parmi A,B,C,D,E sont alignés. Inversement si 3 points de A,B,C,D,E sont alignés disons A,B,C alors la conique formé de la droite passant par A,B,C et de la droite (DE) passe par A,B,C,D,E par unicité, c'est notre conique et elle est dégénérée.

Remarque.
1. *La démarche présenté dans ce développement est effective, il suffit de résoudre un système linéaire, on peut ainsi déterminer effectivement l'existence et l'unicité de la conique. De plus connaissant une équation de la conique, il est possible de déterminer sa nature à partir des signatures de la partie quadratique et de la conique homogénéisé.*
2. *Maintenant d'un point de vue géométrique on peut caractériser la nature de la conique. Mettons de côtes les cas dégénérées, on se donne* 4 *des* 5 *points, parmi ces quatre points passent deux paraboles. Si le 5-ème point se trouve sur l'intersection d'une parabole, la conique est cette parabole, si le point se trouve à l'extérieur de la zone d'intersection de l'intérieur des paraboles la conique est une ellipse, et à l'intérieur c'est une hyperbole.*
3. *Naturellement après 5 points on souhaiterait passer à 6. Dans ce cas on a pas toujours une conique passant par ses 6 points, ils doivent vérifier une certaine hypothèse (théorème de Pascal).*

Théorème de Riesz-Fréchet-Kolmogorov

L'objet de ce développement est de caractériser les parties relativement compactes de $L^p(\mathbf{R}^d)$. On a besoin de pour de trois hypothèses. La première hypothèse est une propriété générale des parties relativement compact, elle empêche l'explosion de la taille des éléments, la seconde appelée hypothèse de tension permet empêcher la fuite du graphe de f [25] à l'infini et la troisième hypothèse est une forme L^p de l'équicontinuité garantie la convergence dans L^p [26].

Théorème. *Pour $p \in [1,\infty)$ une partie $A \subset L^p(\mathbf{R}^d)$ est relativement compact si et seulement si :*

(a) A est borné dans L^p ;

(b) $\int_{|x|>R} |f(x)|^p \mathrm{d}x \to 0$ lorsque $R \to +\infty$ uniformément en $f \in A$;

(c) $\|\tau_h f - f\|_{L^p} \to 0$ lorsque $h \to 0$ uniformément en $f \in A$.

On va montrer ce résultat en utilisant la caractérisation suivante : *une partie A d'un espace métrique complet X est relativement compacte si et seulement pour tout $\varepsilon > 0$ on peut recouvrir A par un nombre fini de boules ouvertes de rayons ε.*

1. Condition suffisante

L'idée de la preuve suivante est de se ramener au théorème d'Ascoli en régularisant les fonctions de A. L'hypothèse c permettra de faire cela par convolution uniformément en A, l'hypothèse b. permettra alors de se ramener à un compact en contrôlant la perte uniformément en A.

- Soit (ρ_n) une approximation de l'unité avec les ρ_n positives C^∞ à support dans $B(0,1/n)$. On majore :

$$\begin{aligned}\|f - \rho_n * f\|_{L^p}^p &= \int_{\mathbf{R}^d} \left(\int_{\mathbf{R}^d} |f(x) - f(x-y)| \rho_n(y) \mathrm{d}y \right)^p \mathrm{d}x \\ &\leqslant int_{\mathbf{R}^d \times \mathbf{R}^d} |f(x) - f(x-y)|^p \rho_n(x) \mathrm{d}x \mathrm{d}y \\ &= \int_{\mathbf{R}^d} \|\tau_y f - f\|_{L^p}^p \rho_n(y) \mathrm{d}y \\ &\leqslant \sup_{|y| \leqslant 1/n} \|\tau_y f - f\|_p^p.\end{aligned}$$

  Ainsi avec l'hypothèse (c) il existe $\rho = \rho_n$ tel que $\|f - \rho * f\|_{L^p} \leqslant \varepsilon$ pour tout $f \in \mathscr{F}$.

- Soit avec l'hypothèse (b) $R > 0$ tel que $\left(\int_{|x|>R} |f(x)|^p \mathrm{d}x \right)^{1/p} \leqslant \varepsilon$ pour tout $f \in A$. Dans la suite on note K le compact $\overline{B(0,R)}$.

- On applique le théorème d'Ascoli à la famille $B = \{\tilde{f} = (f * \rho)_{|K} \mid f \in A\} \subset C^0(K, \mathbf{R})$ où :
  (a) B est borné : $\|\tilde{f}\|_{L^p} \leqslant \|f * \rho\|_{L^p} \leqslant \|f\|_{L^\infty} \|\rho\|_{L^q}$ par Hölder ;
  (b) B est équicontinue car borné dans $C^1(K, \mathbf{R})$ étant donnée que : $\partial_i (f * \rho) = f * \partial_i \rho$ et comme précédemment :
  $$\|f * \partial_i \rho\|_{L^\infty} \leqslant \|f\|_{L^p} \|\partial_i \rho\|_{L^q}.$$

  Il existe ainsi des fonctions $f_1, \cdots, f_n \in A$ tel que :

$$B \subset \bigcup_{k=1}^n B(\tilde{f_k}, \frac{\varepsilon}{|K|}) \quad \text{dans } C^0(K, \mathbf{R}).$$

- Concluons. Pour $f \in A$ soit f_k tel que $\tilde{f} \in B(\tilde{f_k}, \varepsilon)$ on a $\|f - f_k\|_{L^p} \leqslant 5\varepsilon$ à partir de :

$$f - f_k = f \mathbf{1}_{K^c} - f_k \mathbf{1}_{K^c} + \mathbf{1}_K (f - f * \rho) + \mathbf{1}_K (f * \rho - f_k * \rho) + \mathbf{1}_K (f_k * \rho - f_k)$$

25. Par exemple la suite $f_n : x \mapsto \mathbf{1}_{[n,n+1]}$ n'admet pas de valeur d'adhérence car une telle valeur d'adhérence serait nécessairement la fonction nulle qui est de masse nulle.

26. Sans cette hypothèse on quand même extraire une sous-suite convergent mais en un sens plus faible, celle de la convergence faible.

2. Condition nécessaire

Venons à la nécessité de ces conditions. Pour l'hypothèse (a) on peut prendre $\varepsilon = 1$ dans la caractérisation par relative compacité. La nécessité des autres conditions vient du fait qu'elles sont vérifiés ponctuellement, et que l'hypothèse de compacité permet de contrôler tous le monde à la précision voulue à partir d'un nombre fini d'éléments ce qui permet de passer à des bornes uniformes.

- Soit $\varepsilon > 0$. On a $A \subset \bigcup_{i=1}^{n} B(f_i, \varepsilon)$ alors :

$$\left(\int_{|x|>R} |f(x)|^p \mathrm{d}x\right)^{1/p} \leqslant \|f - f_i\|_{L^p} + \left(\int_{|x|>R} |f_i(x)|^p \mathrm{d}x\right)^{1/p} \leqslant \varepsilon + \max_{1\leqslant i\leqslant n} \left(\int_{|x|>R} |f_i(x)|^p \mathrm{d}x\right)^{1/p}.$$

  Maintenant par convergence dominée $\left(\int_{|x|>R} |f_i(x)|^p \mathrm{d}x\right)^{1/p} \to 0$ lorsque $R \to +\infty$ donc il existe $R > 0$ tel que $\left(\int_{|x|>R} |f(x)|^p \mathrm{d}x\right)^{1/p} \leqslant 2\varepsilon$ quelque soit $f \in A$.

- Par le même principe on a :

$$\|\tau_h f - f\|_p \leqslant \|\tau_h f - \tau_h f_i\|_p + \|\tau_h f_i - f_i\|_p + \|f_i - f\|_p.$$

  Comme τ_h est une isométrie on a :

$$\|\tau_h f - f\|_p \leqslant 2\varepsilon + \max_{1\leqslant i\leqslant n} \|\tau_h f_i - f_i\|_p.$$

  En invoquant la "continuité des translations" on en déduit comme précédemment le résultat.

3. Compléments

De façon analogue au rôle joué par le caractère C^1 à dérivée borné dans l'équicontinuité, l'hypothèse (c) est vérifié dès que A est une partie bornée de $W^{1,p}(R^d)$. On se contente de la dimension 1.

Application. *Toute partie bornée de $W^{1,1}(\mathbf{R})$ est de restriction relativement compacte sur $L^1(I)$ où I est un intervalle bornée.*

Preuve. Il suffit de vérifier la troisième l'hypothèse pour laquelle on utilise qu'un élément $u \in W^{1,p}(\mathbf{R})$ s'écrit comme l'intégrale de sa dérivée ce qui donne $\|\tau_h u - u\|_{L^1(I)} \leqslant |h|$. □

On peut étendre ce résultat à $W^{1,p}(\mathbf{R})$ avec $p > 1$ mais dans ce cas il découle plus simplement du théorème d'Ascoli : l'injection $W^{1,p}(\mathbf{R})$ dans $C(\bar{I})$ est compacte lorsque I est borné. Ce n'est pas le cas de l'application précédente car il existe des parties bornés de $W^{1,1}(\mathbf{R})$ qui ne sont pas compactes dans $C(]0,1[)$ - on peut pour cela faire converger une suite vers un trottoir. Maintenant en utilisant que $W^{1,1} \subset C(\bar{I})$ lorsque I borné on a $\|\tau_h u - u\|_{L^p(I)} \leqslant (2\|u\|_\infty)^{p-1}|h|$ et on peut étendre le critère de compacité de L^1 à tout L^p avec $1 \leqslant p < \infty$.

Résolution d'un problème aux limites

Le but de ce développement est de résoudre le problème aux limites :

$$\begin{cases} -u'' = g(u) \\ u(0) = u(1) = 0 \end{cases}$$

où $g : \mathbf{R} \to \mathbf{R}$ est continue admettant une primitive G majorée : $G \leqslant M$. La démarche consiste à reformuler ce problème en un problème de point critique et à montrer l'existence d'une solution par un argument de compacité.

1. Théorème de compacité faible

Soit $(x_n)_{n\in\mathbf{N}}$ une suite bornée de E on note $X_n : \phi \in E' \to \langle x_n, \phi\rangle$.

- Supposons dans un premier temps E' séparable : $E' = \overline{D}$ avec D dénombrable dense. Par extraction diagonale on peut extraire de X_n une sous-suite $X_{j(n)}$ convergeant simplement sur D.
- Comme la suite (X_n) est équicontinue on montre que $X_{j(n)}$ converge simplement sur E. En effet :

  $$|X_n(\phi) - X_m(\phi)| \leqslant 2M\|\phi - \psi\| + |X_n(\psi) - X_m(\psi)|$$

  donc pour $\varepsilon > 0$ en prenant $\psi \in D$ tel que $2M\|\phi - \psi\| \leqslant \varepsilon$ comme la suite $(X_{j(n)}(\phi))$ est de Cauchy on a N tel que pour tout $n, m \geqslant N$, $|X_{j(n)}(\phi) - X_{j(m)}(\phi)| \leqslant \varepsilon$.
- Ainsi $X_{j(n)}$ converge simplement, sa limite X est une application linéaire $E' \to \mathbf{R}$ et comme $|X_{j(n)}(\phi)| \leqslant M\|\phi\|$ l'application X est continue donc par réflexivité on a $X : \phi \mapsto \langle x, \phi\rangle$ pour un $x \in E$ ce qui prouve le résultat.
- On se passe de l'hypothèse séparable en se ramenant à $F = \overline{vect(x_n)}$ et en utilisant les deux résultats suivant :
  - un sous-espace fermé d'un espace réflexif est réflexif
  - si le dual d'un espace de Banach est séparable l'espace lui même est séparable.

2. Minimisation des fonctionnelles convexes

Il existe en tout point $x \in C$ une forme linéaire ϕ tel que :

$$f(y) \geqslant f(x) + \phi(y - x).$$

Dans le cas où f est de classe C^1 cette inégalité découle de l'inégalité des tangentes appliqués à $t \mapsto f(tx + (1-t)y)$ convexe d'une variable réelle, on a en fait $\phi = \mathrm{d}f(x)$. Dans le cas général où f est continue on utilise l'existence d'un hyperplan d'appui qui résulte du théorème de Hahn-Banach. Maintenant considérons (x_n) une suite minimisante alors la suite x_n est bornée par coercitivité de f, on peut donc extraire une sous-suite toujours noté x_n convergeant faiblement vers un vecteur x. Alors $\phi(x_n - x) \to 0$ donc $\inf f \lim_{n\to+\infty} f(x_n) \geqslant f(x)$ donc x est un minimiseur de f.

3. Application à la résolution du problème aux limites

Considérons le problème aux limites :

$$\begin{cases} -u'' + u|u|^{p-1} = f \\ u(0) = u(1) = 0 \end{cases}$$

où p est un réel strictement positif et $f \in C^0(\overline{I}, \mathbf{R})$. On écrit ce problème sous la formulation faible :

$$\int_0^1 u'v' + \int_0^1 u|u|^{p-1}v = \int_0^1 fv$$

où les inconnues u, v sont dans H_0^1. On voit cette équation comme une équation de point critique [27] pour la fonctionnelle :

$$J(u) = \int_0^1 \left(\frac{1}{2}|u'|^2 + \frac{|u|^{p+1}}{p+1} - fu \right)$$

- J est coercivite car $J(u) \geqslant c\|u\|_{H_0^1}$ d'après Poincaré ;
- J est strictement convexe par strict convexité des fonctions $x \mapsto |x|^2$ et $x \mapsto |x|^{p+1}$ (hypothèse $p > 0$).

Ainsi J possède un unique minimum ce qui montre l'existence et l'unicité d'une solution à notre problème aux limites.

27. Détaillons les calculs de la différentielle. On note $g(u) = u|u|^{p-1}$ et $G(u) = \frac{|u|^{p+1}}{p+1}$ par soucis de simplicité. On a :

$$J(u+v) - J(u) - \int_I \left(u'v' + g(u)v + fv \right) = frac12 \int_I |v'|^2 + \int_I \varepsilon(v(t))v(t)\mathrm{d}t$$

où $G(x+h) = G(x) + hg(x) + \varepsilon(h)h$ avec $\lim_{h\to 0} \varepsilon(h) = 0$ comme g est continue. Le premier terme est par l'inégalité de Poincaré un $O(\|v\|_{H_0^1}^2)$, le second se majore par $\|v\|_{H_0^1} \sqrt{|I| \sup_I |\varepsilon(v(t))|}$ or comme H_0^1 s'injecte continument dans $C^0(I)$, lorsque $\|v\|_{H_0^1} \to 0$ on a $\sup_I |\varepsilon(v(t))| \to 0$ de sorte que le terme de droite est un $o(\|v\|_{H_0^1})$.

Approximation de Bernstein

Le but de ce développement est de démontrer une version construction du théorème d'approximation de Weierstrass. On associe à une fonction continue $f : [0,1] \to \mathbf{R}$ son n-ème polynôme d'approximation de Bernstein définie comme

$$B_n f := \mathbf{E}\left(f(\overline{X}_n)\right) \in \mathbf{R}[x] \qquad \text{où} \quad \overline{X}_n \sim \frac{1}{n}\mathrm{Bin}(n,x).$$

Première en calculant l'espérance on trouve :

$$B_n f(x) = \sum_{k=0}^{n} \binom{n}{k} f\left(\frac{k}{n}\right) x^k (1-x)^{n-k}$$

de sorte que $B_n f(x)$ est bien une fonction polynomiale de degré n. Ensuite on s'attend à avoir convergence vers f en vertu :

1. la loi des grands nombres : $\overline{X}_n \xrightarrow{\mathbf{P}} x$;
2. la continuité de f qui avec le continuous mapping theorem entraine que $f(\overline{X}_n) \xrightarrow{\mathscr{L}} x$.

Ces deux points entrainent que $B_n f(x) \to f(x)$ à x fixé. Maintenant en reprenant la démonstration de ces résultats il apparaît que la convergence peut être rendu uniforme en x d'une part dans la loi des grands nombres car la variance de $\mathrm{Ber}(x)$ est uniformément borné en x par 1/4 et deuxième par l'uniforme continuité de f selon le théorème de Heine. On ne détaille pas ces points qui vont résulter de résultats plus forts, précisément on va chercher une borne sur la vitesse de convergence à savoir :

Proposition. *En notant ω le module d'uniforme continuité de f on a :*

$$\|f - B_n f\|_\infty \leqslant \frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}\right).$$

1. Convergence et vitesse de convergence

- On commence par écriture avec l'inégalité triangulaire : $|f(x) - B_n f(x)| \leqslant \mathbf{E}_x|f(x) - f(\overline{X}_n)|$. Si f étant C^1 on aurait écrit $f(x) - f(y)$ comme l'intégrale de sa dérivée pour en déduire une majoration type inégalité des accroissements finis. On va généraliser cette idée en remplaçant la somme continue, par une somme de Riemann discrète. Introduisons $\delta > 0$ puis un chemin $x = x_0 \to x_1 \to \cdots \to x_{k+1} = y$ avec $x_{i+1} = x_i + \delta$ jusqu'à l'avant dernier pas choisie pour tombé pile sur y. On écrit :
$$|f(x) - f(y)| \leqslant \sum_{i=0}^{k} |f(x_i) - f(x_{i+1})| \leqslant (k+1)\omega(\delta).$$
Or comme $k\delta \leqslant |y - x|$ on a $|f(x) - f(y)| \leqslant (1 + \delta^{-1}|y - x|)\omega(\delta)$.
- Avec l'inégalité précédente on a la majoration :
$$|f(x) - B_n f(x)| \leqslant \omega(\delta)\left(1 + \frac{1}{\delta}\mathbf{E}|\overline{X}_n - x|\right).$$
On est amené à évaluer l'écart de $\overline{X}_n$ à x en norme 1. Plutôt on va se ramener à une norme 2 par l'inégalité de Cauchy-Schwartz où l'indépendance est symbole d'orthogonalité. On a :
$$\mathbf{E}|\overline{X}_n - x| \leqslant \sqrt{\mathbf{E}|\overline{X}_n - x|^2} = \mathbf{Var}(\overline{X}_n)^{1/2} = \left(\frac{1}{n^2}\sum_{k=1}^{n}\mathbf{Var}(X_k)\right)^{1/2} = \frac{1}{\sqrt{n}}\mathbf{Var}(X_1)^{1/2}.$$

- La variance d'une Bernoulli de paramètre x est $x(1-x)$. La fonction $q : x \mapsto x(1-x)$ est une fonction de degré 2 de coefficient dominant négatif donc atteint son maximum au point d'annulation de sa dérivée $2x-1$ *i.e.* $x = \frac{1}{2}$ puis $q(x) \leqslant 1/4$ pour tout $x \in \mathbf{R}$. Ainsi,

$$|f(x) - B_n f(x)| \leqslant \omega(\delta)\left(1 + \frac{1}{2\delta\sqrt{n}}\right).$$

  Pour conclure il faut bien choisir δ, ne connaissant pas ω on ne pourra faire cela de façon optimal. On a dans une mesure envie de prendre δ petit car le terme $\omega(\delta)$ le sera d'autant plus mais on est limite par le terme $\left(1 + \frac{1}{2\delta\sqrt{n}}\right)$ qui explose lorsque $\delta \to 0$. On trouve un compromis en prenant δ afin que le terme de droite soit constant *i.e.* $\delta = n^{-1/2}$. Ainsi on a :

$$\|f - B_n f\|_\infty \leqslant \frac{3}{2}\omega\left(\frac{1}{\sqrt{n}}\right).$$

  Cette vitesse de convergence semble un accord avec l'intuition probabiliste, le théorème central limite indiqua,t que les fluctuations relatives entre $\overline{X}_n$ et sa moyenne sont de l'ordre de $\frac{1}{\sqrt{n}}$.
- Cette vitesse de convergence est optimal à une constante près on a pour $f(x) = |x - 1/2|$,

$$f(x) - B_n f(x)|_{|x=1/2} = |B_n f(1/2)| = \frac{1}{2^n}\sum_{k=0}^{n}\binom{n}{k}\left|\frac{k}{n} - \frac{1}{2}\right| = \frac{1}{2^n}\binom{n-1}{\lfloor (n-1)/2 \rfloor}.$$

  Avec l'équivalent $\binom{2k}{k} \sim \frac{4^k}{\sqrt{\pi k}}$ qu'on peut obtenir à partir de la formule de Stirling on en déduit que ce dernier terme est équivalent à $\frac{1}{\sqrt{2\pi n}}$ et pour ce choix on a $\omega(\delta) = \delta$.

2. Application à d'autres approximations polynomiales

- Notons pour un choix de nœuds d'interpolation $T_n f$ le polynôme d'interpolation de f avec n points. La question est de savoir si $T_n f$ converge uniformément vers f. En général on a des résultats négatifs car si T_n converge pour toute fonction continue, alors par le théorème de Banach-Steinhaus $\sup_{n\in\mathbf{N}} \|T_n\| = +\infty$. Par exemple pour les noeuds d'interpolation de Tchebychev on a $\|T_n\| \asymp \log(n)$, il existe donc une fonction continue tel que $T_n f$ ne converge pas uniformément vers f.
- Maintenant on a la majoration suivante :

$$\|T_n f - f\| \leqslant (1 + \|T_n\|) d(f, \mathbf{R}_n[x]).$$

  Cette majoration découle du fait que $T_n g = g$ pour tout $g \in \mathbf{R}_n[x]$ de sorte que $T_n f - f = T_n(f-g) - (f-g)$ d'où le résultat suit par définition des normes opérateurs et en optimisant sur g. Cette majoration indique que le contrepoids à la divergence de $\|T_n\|$ est la qualité de l'approximation polynomiale. Dans le cas de l'interpolation de Tchebychev en utilisant l'approximation de Bernstein on en déduit que dès lors que $\lim_{\delta\to 0}\omega(\delta)\log\frac{1}{\delta} = 0$ les polynômes d'interpolation $T_n f$ convergent vers f donc par exemple pour toute fonction f holdérienne.

Principe de la borne uniforme

Proposition. *Soient E un Banach, F un e.v.n et $(T_i)_{i\in I}$ une famille d'applications linéaires continues de E dans F. Si la famille est ponctuellement bornée* $\sup_{i\in I}\|T_i x\| < +\infty,\ \forall x \in E$ *alors elle est uniformément borné :* $\sup_{i\in I}\|T_i\| < +\infty$.

1. Démonstration du résultat

Les fermés $F_n = \{x \in E\ :\ \forall i \in I,\ \|T_i x\| \leqslant n\}$ recouvrent E lorsque n parcourt $\mathbf{N}$. Par le lemme de Baire, un de ses fermés F_n est d'intérieur non vide : $B(x_0, r) \subset F_n$ avec $r > 0$. Ainsi on a pour tout $z \in B(0,1)$, quelque soit $i \in I$, $\|T_i(x_0 + zr)\| \leqslant n$ d'où $\|T_i z\| \leqslant \frac{n+\|T_i x_0\|}{r}$ ce qui prouve le résultat.

2. Application à la construction d'objets pathologiques

Sous le cadre précédent, par contraposition si $\sup_{i\in I}\|T_i\| = +\infty$ alors il existe $x \in E$ tel que $\sup_{i\in I}\|T_i x\| = +\infty$. Ce résultat permet de montrer l'existence d'objets pathologiques dans des problèmes d'approximations. On va l'appliquer aux séries de Fourier, on a :

$$T_n : f \in C_{2\pi-\text{per}} \mapsto \frac{1}{2\pi}\int_{-\pi}^{\pi} f(x-t)D_n(t)\mathrm{d}t$$

où $D_n(t) = \frac{\sin(n+1/2)t}{\sin t/2}$ est le noyau de Dirichlet.

- Calculons les normes opérateurs, on affirme que : $\|T_n\| = \frac{1}{2\pi}\int_{-\pi}^{\pi}|D_n(t)\mathrm{d}t$. En effet,

  $$|T_n f| \leqslant \frac{1}{2\pi}\|f\|_\infty \int_{-\pi}^{\pi}|D_n(t)|\mathrm{d}t$$

  avec égalité pour $f = \mathrm{sng}(D_n)$. Seulement cette dernière fonction n'est pas continue, mais on peut l'approcher par des fonctions continues $f_\varepsilon = \frac{D_n}{|D_n|+\varepsilon}$. Ainsi, f_ε converge p.s. vers f lorsque $\varepsilon \to 0$ et $|f_\varepsilon| \leqslant 1$ et le résultat suit par convergence dominée.
- On a déjà par parité :

  $$\|T_n\| = \frac{1}{\pi}\int_0^{\pi}\left|\frac{\sin(n+1/2)t}{\sin t/2}\right|\mathrm{d}t.$$

  Sur $[0,\pi]$ on a $\frac{1}{\pi}t \leqslant \sin t/2 \leqslant t/2$ de sorte que :

  $$\|T_n\| \asymp \int_0^{\pi}\frac{|\sin(n+1/2)t|}{t}\mathrm{d}t =: I_n.$$

- On écrit :

  $$I_n = \int_0^{(2n+1)\pi/2}\frac{|\sin t|}{t}\mathrm{d}t = \sum_{k=0}^{2n}\int_{k\pi/2}^{(k+1)\pi/2}\frac{|\sin t|}{t}\mathrm{d}t.$$

  Maintenant pour $k \geqslant 1$:

  $$\frac{1}{k+1} \leqslant \int_{k\pi/2}^{(k+1)\pi/2}\frac{|\sin t|}{t}\mathrm{d}t \leqslant \frac{1}{k}.$$

  Ainsi on a :

  $$I_n = \sum_{k=1}^{2n}\frac{1}{k} + O(1) = \log(n) + O(1).$$

- On obtient finalement que $\|T_n\| \asymp \log(n)$ d'où l'on déduit l'existence d'une fonction continue dont la sa série de Fourier ne converge pas uniformément.

3. Complément sur l'analyse de la convergence des séries de Fourier

Le résultat précédent peut sembler bien négatif mais si l'on pousse un peu plus loin l'analyse précédente on va en fait voir que toute fonction un petit peu régulière est somme de sa série de Fourier au sens de la convergence uniforme. Précisément on a l'estimation élémentaire :

$$\|f - T_n f\| \leqslant (1 + \|T_n\|) d(f, \mathscr{P}_n)).$$

Cette majoration s'obtient en écrivant $f - T_n f = (f - g) - T_n(f - g)$ et en utilisant $T_n g = g$ pour $g \in \mathscr{P}_n$ l'ensemble des polynômes trigonométriques de degré inférieur à n puis en écrivant. Ainsi le contrepoids à la divergence de $\|T_n\|$ est la qualité de l'approximation trigonométrique et cette dernière correspond à très peu de choses près à la régularité de f comme le montre les théorèmes de Jackson et Bernstein. On en déduit ainsi par exemple que toute fonction vérifiant une condition de Dini-Lipschitz est limite uniforme de sa série de Fourier.

Lemme de Morse et méthode de Laplace

Le but de ce développement est d'étudier l'asymptotique $t \to +\infty$ des intégrales de la forme suivante :

$$\int_{\mathbf{R}^d} a(x)\mathrm{e}^{-tf(x)}\mathrm{d}x.$$

Lorsque $t \to +\infty$ la masse relative de cette intégrale se concentre en les points où f est minimal. Autour d'un tel point x_0 supposé non dégénérée la fonction f se comporte comme son approximation à l'ordre 2 : $f(x) - f(x_0) \approx \frac{1}{2}\mathrm{H}f(x_0)\cdot(x-x_0)^2$ ce qui nous ramène à un calcul d'intégrale gaussienne et donne une contribution de l'ordre de :

$$\frac{1}{|\mathrm{H}f(x_0)|^{1/2}} a(x_0)\mathrm{e}^{-tf(x_0)}\left(\frac{2\pi}{t}\right)^{d/2}.$$

Nous allons montrer précisément cela sous les hypothèses suivantes :

1. f est de classe C^3 possède un unique minimum global non dégénérée x_0 ;
2. pour tout voisinage V de x_0 il existe $\alpha > 0$ tel que $|f(x)| \geqslant |f(x_0)| + \alpha$ quelque soit $x \in V^c$.
3. a est continue intégrable sur $\mathbf{R}^d$.

Pour justifier l'approximation précédente, on va utiliser un changement de variable pour passer d'un terme à l'autre. Ensuite nous reviendrons à notre intégrable. Pour simplifier la présentation sans pour autant perdre en généralité nous supposerons $x_0 = 0$ et $f(0) = 0$.

1. Lemme de Morse

On montre qu'il existe un C^1 difféomorphisme local en 0 tel que $f(x) = \|\varphi(x)\|^2$.

- Écrivons la formule de Taylor avec reste intégrale à l'ordre 2 :

  $$f(x) = \frac{1}{2}x^\top Q(x)x \qquad \text{où } Q(x) = 2\int_0^1 (1-t)\mathrm{H}f(tx)\mathrm{d}x.$$

  La fonction $Q : \mathbf{R}^d \mapsto \mathrm{S}_n(\mathbf{R})$ est C^1 par dérivation sous l'intégrale.
- On a $Q(0) = \mathrm{H}f(0)$ matrice symétrique définie (hypothèse) positive (car minimum local) donc d'après le théorème d'inertie de Sylvester on peut écrire $Q(0) = P_0^\top P_0$ avec P_0 inversible. On va montrer qu'au voisinage de 0 on peut écrire $Q(x) = P(x)^\top P(x)$ où P est C^1. Pour cela on cherche à appliquer le théorème d'inversion local à $\Psi : P \mapsto P^\top P$ en P_0.
- Calculons la différentielle de Ψ en P_0 :

  $$\mathrm{d}\Psi(P_0)\cdot H = H^\top P_0 + P^\top H.$$

  Le noyau de cette différentielle est $(P_0^{-1})^\top \mathrm{A}_n(\mathbf{R})$. Ainsi,

  $$\overline{\Psi} : (P_0^{-1})^\top \mathrm{S}_n(\mathbf{R}) \to \mathrm{S}_n(\mathbf{R})$$

  est une fonction C^1 de différentielle injective en P_0 donc bijective par égalité de dimension. Ainsi, par inversion local $\overline{\Psi}$ est un C^1-difféomorphisme local en P_0.
- Posons alors $P(x) = \overline{\Psi}^{-1}(Q(x))$. Comme Q est continue, pour x dans un voisinage de 0 on a $Q(x)$ dans le domaine de définition de $\overline{\Psi}^{-1}$ donc P est bien définie au voisinage de 0, est de classe C^1 et on a bien : $Q(x) = P(x)^\top P(x)$.
- Pour terminer posons $\varphi(x) = P(x)x$ qui est bien définie au voisinage de 0, de classe C^1 et comme $\varphi(x) = P_0 x + o(1)$ par continuité de P on a $\mathrm{d}\varphi(0) = P_0$ est inversible donc par inversion local φ est un difféomorphisme local en 0.

2. Intégrales de Laplace

Comme suggéré on sépare les contributions sur un voisinage de x_0 et ailleurs. Considérons $B_\delta = B(0,\delta)$ dont l'adhérence est contenue dans le domaine de définition de φ.

- Sur B_δ on pose le changement de variable $y = \sqrt{t}\varphi(x)$ ce qui donne :

$$\int_{B_\delta} a(x)\mathrm{e}^{-tf(x)}\mathrm{d}x = \int_{t^{1/2}\varphi(B_\delta)} a(\varphi(y/\sqrt{t}))\mathrm{e}^{-\frac{1}{2}\|y\|^2}\frac{\mathrm{d}y}{t^{d/2}|\mathrm{D}\varphi\cdot\varphi^{-1}(y/\sqrt{t})|}.$$

  En faisant tendre $t \to +\infty$ on obtient l'équivalent :

$$\frac{1}{t^{d/2}}\int_{\mathbf{R}^d} a(0)\mathrm{e}^{-\frac{1}{2}\|y\|^2}\frac{\mathrm{d}y}{|\mathrm{D}\varphi(0)|} = \frac{a(0)}{|\mathrm{H}f(0)|^{1/2}}\left(\frac{2\pi}{t}\right)^{t/2}$$

  en se rappelant que $\mathrm{D}\varphi(0) = P_0$ vérifie $\mathrm{H}f(0) = P_0^\top P_0$ et en utilisant la valeur de l'intégrale de la gaussienne : $\int_{\mathbf{R}^d}\mathrm{e}^{-\frac{1}{2}\|y\|^2}\mathrm{d}y = \pi^{d/2}$. Le passage à la limite est justifié par la domination :

$$\frac{\sup_{\overline{B}_\delta}|a|}{\inf_{\overline{B}_\delta}|\mathrm{D}\varphi\circ\varphi^{-1}|}\mathrm{e}^{-\frac{1}{2}\|y\|^2}.$$

- Pour le terme restant on a par l'hypothèse 2. $|f(x)| \geqslant \alpha > 0$ pour tout $x \in B_\delta^c$ de sorte que :

$$\left|\int_{B_\delta} a(x)\mathrm{e}^{-tf(x)}\mathrm{d}x\right| \leqslant \mathrm{e}^{-t\alpha}\int|a| = o\left(\frac{1}{t^{d/2}}\right).$$

Théorème de Rademacher

Le but de ce résultat est d'établir le résultat suivant de représentation des fonctions lipschitziennes.

Théorème. *Soit I un intervalle de $\mathbf{R}$ puis $f : I \to \mathbf{R}$ une fonction lipschitzienne. Il existe une fonction $g \in L^\infty(I)$ tel que :*

$$\forall x, y \in I, \quad f(y) - f(x) = \int_x^y g(t)\mathrm{d}t.$$

Par un prolongement par continuité on peut se ramener à $I = \mathbf{R}$ ce qu'on supposera dans la suite.

- On va construire g avec le théorème de Riesz. Pour cela il nous faut une forme linéaire ; on prend :[28]

  $$T : \varphi \in \mathscr{C}_c^\infty(I) \longmapsto -\int_{\mathbf{R}} f\varphi'.$$

  Cette application est bien définie car $f\varepsilon'$ est continue et le domaine d'intégration est bien compact (c'est le support de φ).

- L'hypothèse Λ-Lipschitz se traduit de la façon suivante : $T : \mathscr{C}_c^1(\mathbf{R}) \subset L^1(\mathbf{R}) \to \mathbf{R}$ est une forme linéaire continue. En effet en écrivant la dérivée comme limite de son taux d'accroissements on a :

  $$T\phi = \lim_{h\to 0} \int_{\mathbf{R}} f(x)\frac{\phi(x+h)-\phi(x)}{h}\mathrm{d}x.$$

  L'interversion des limites est justifié par la domination $\|f\phi'\|_{L^\infty(K)}\mathbf{1}_K$ où $K = \operatorname{supp}\phi$. Alors on peut faire une "intégration par parties" avec le taux d'accroissement de la façon suivante :

  $$\begin{aligned}\int_{\mathbf{R}} f(x)\frac{\phi(x+h)-\phi(x)}{h}\mathrm{d}x &= \frac{1}{h}\left(\int_{\mathbf{R}} f(x)\phi(x+h)\mathrm{d}x - \int_{\mathbf{R}} f(x)\phi(x)\mathrm{d}x\right)\\ &= \frac{1}{h}\left(\int_{\mathbf{R}} f(x-h)\phi(x)\mathrm{d}x - \int_{\mathbf{R}} f(x)\phi(x)\mathrm{d}x\right)\\ &= \int_{\mathbf{R}} \frac{f(x-h)-f(x)}{h}\phi(x)\mathrm{d}x.\end{aligned}$$

  On en déduit finalement que :

  $$|T\phi| \leqslant \Lambda\|\phi\|_{L^1(\mathbf{R})}.$$

- On veut maintenant appliquer Riesz. Par le théorème de prolongement des applications uniformément continue, on peut prolonger T (uniquement) en une forme linéaire continue sur $L^1(\mathbf{R})$. Dès lors on ne peut appliquer Riesz sur $\mathbf{R}$ tout entier, mais on peut le faire localement car $L^2([-n,n]) \subset L^1([-n,n])$ continument par Cauchy-Schwartz. Ainsi il existe une unique fonction $g_n \in L^2(-n,n)$ tel que pour tout $\phi \in L^2(-n,n)$, $T\phi = \int_{-n}^n \phi g_n$. Par unicité de g_n on a g_{n+1} prolonge g_n de sorte que l'on peut bien définir $g \uparrow g_n$ on a dès lors :

  $$T\phi = \int_{\mathbf{R}} \phi g_n, \qquad \forall \phi \in \bigcup_{n\in\mathbf{N}} L^2([-n,n]).$$

- Montrons première que $g \in L^\infty(\mathbf{R})$, on va en fait montrer que $|g| \leqslant \Lambda$ p.p. S'il existe un borélien A non négligeable sur lequel $|g| > \Lambda$ alors :

  $$\Lambda|A| < \int_A |g| = T(\operatorname{sgn}(g)\mathbf{1}_A) \leqslant \Lambda|A|.$$

28. Naturellement on aurait pris $\phi \mapsto \int f'\phi$ mais f n'étant pas assez régulière on fait une intégration par parties. C'est en fait la dérivée de f au sens des distributions.

- Introduisons maintenant $G : x \longmapsto \int_0^x g(t)\mathrm{d}t$ et montrons que $f = G$ à une constante près. Suivant la démarche précédente on va tester l'égalité contre des fonctions tests. On a pour $\phi \in \mathscr{C}_c^1(I)$ avec le théorème de Fubini :

$$\begin{aligned}\int_{\mathbf{R}} G(x)\phi'(x)\mathrm{d}x &= \int_{\mathbf{R}} \left(\int_0^x g(t)\mathrm{d}t\right)\phi'(x)\mathrm{d}x \\ &= \int_{\mathbf{R}} g(t)\left(\int_t^{+\infty} \phi'(x)\mathrm{d}x\right)\mathrm{d}t \\ &= -\int_{\mathbf{R}} g(t)\phi(t)\mathrm{d}t = T\phi.\end{aligned}$$

  On a donc en notant $h = f - G$,

$$\int_{\mathbf{R}} h(x)\phi'(x)\mathrm{d}x = 0, \qquad \forall \phi \in C_c^1(\mathbf{R}).$$

- On a donc montré l'égalité contre l'ensemble des fonctions tests $E = \mathscr{D}(C_c^1(\mathbf{R})) = \{\psi \in C_c^0(\mathbf{R}) \mid \int_{\mathbf{R}} \psi = 0\}$. C'est un hyperplan de $C_c^0(\mathbf{R})$ noyau de la forme linéaire $\psi \mapsto \int_{\mathbf{R}} \psi$. On a donc $\mathscr{C}_c^0(\mathbf{R}) = E \oplus \mathbf{R}\chi$ où $\chi \in C_c^0(\mathbf{R})$ avec $\int_{\mathbf{R}} \chi \neq 0$ disons égal à 1. On peut ainsi écrire toute fonction $\psi \in C_c^0(\mathbf{R})$ comme $\psi = \phi' + c\chi$, la constante c étant égal à $\int \psi$. On a finalement :

$$\int h\psi = c\int h\chi \Rightarrow \int \left(h - \int h\chi\right)\psi = 0.$$

  On en déduit que $h = \int h\chi$ p.p ce qui permet de conclure.

Remarque. *1. Avec le théorème de différentiation de Lebesgue (Rudin) on en déduit qu'une fonction lipschitzienne sur* $\mathbf{R}$ *est p.p. dérivable ;*

2. De façon assez général avec une régularité inférieur à lipschitzienne ce résultat n'est plus vraie, par exemple il existe des fonctions holdériennes pour tout $0 < \alpha < 1$ *qui sont nulle part dérivable comme la fonction de Takagi (voir Zully-Queffelec).*

Méthode du gradient à pas fixe

Le but de ce développement est d'étudier la convergence d'une méthode numérique d'optimisation d'une fonction convexe : la méthode du gradient à pas fixe. Le principe de l'algorithme consiste à partir d'un point quelconque puis sachant que le gradient décrit la direction de plus forte variation, de se déplacer dans la direction opposée de ce dernier. On peut aussi interpréter cette algorithme comme une méthode de résolution de l'équation de point critique par une suite récurrente ce qui est pertinent dans le cadre des fonctions convexes.

Théorème. *Soit $f : \mathbf{R}^d \to \mathbf{R}$ une fonction de classe C^1. On définit à partir de $x_0 \in \mathbf{R}^d$ la suite $x_{k+1} = x_k - t\nabla f(x_k)$.*

1. *Si f est strictement convexe (coercitive) gradient ℓ-Lipschitz alors pour $t > 0$ suffisamment petit - on peut prendre $0 < t < 2\ell^{-2}$ - x_k converge vers l'unique minimum de f.*
2. *Si f est de classe C^2 avec $\mathrm{Sp}\,\mathrm{H}f(x) \subset [\alpha, \ell]$ où $0 < \alpha \leqslant \ell < \infty$ alors pour $t = \frac{2}{\alpha+\ell}$ on a :*

$$\|x_k - x^*\| \leqslant \left(\frac{\ell - \alpha}{\ell + \alpha}\right)^k \|x_0 - x^*\|.$$

1. Convergence

- f étant supposée strictement convexe, elle atteint son minimum en un unique point x^*. Maintenant l'idée est que :

$$f(x_{k+1}) - f(x_k) \approx \nabla f(x_k) \cdot (x_{k+1} - x_k) = -t\|\nabla f(x_k)\|^2.$$

  Pour justifier cette approximation on écrit en utilisant l'hypothèse du gradient Lipschitz :

$$\begin{aligned} f(y) - f(x) - \nabla f(x) \cdot (y - x) &= \int_0^1 \langle \nabla f(x + s(y-x)) - \nabla f(x) \,|\, y - x \rangle \mathrm{d}s \\ &\leqslant \int_0^1 \ell s \|y - x\|^2 \mathrm{d}s = \frac{\ell}{2}\|x - y\|^2. \end{aligned}$$

  Ainsi $f(x_{k+1}) \leqslant f(x_k) - t\left(1 - \frac{L^2}{2}t\right)\|\nabla f(x_k)\|^2$.

- Pour $0 < t < \frac{2}{L^2}$ le terme $\rho = t\left(1 - \frac{L^2}{2}t\right)$ est strictement positif. On en déduit que $f(x_k)$ est alors décroissante, minorée donc convergence puis :

$$0 \leftarrow f(x_{k+1}) - f(x_k) \leqslant -\rho\|\nabla f(x_k)\|^2 \leqslant 0$$

  de sorte que $\|\nabla f(x_k)\| \to 0$. Maintenant par coercitivité de f la suite (x_k) est borné et la convergence du gradient montre que toute valeur d'adhérence de (x_k) est x^* donc x_k converge vers x^*.

2. Vitesse de convergence

- Les hypothèses précédentes ne permettent pas d'établir d'avoir des estimations sur la vitesse de convergence.[29] Le point 2. renforce nos hypothèses, notons que la condition sur ℓ assure l'hypothèse gradient ℓ-lispschitz quand l'hypothèse sur α assure l'α-convexité de la fonction, mieux que la convexité strict. Bref elles généralisent les hypothèses précédentes.

29. C'est un vrai résultat, on peut montrer qu'il existe une constante $c > 0$ el que pour tout n il existe une fonction f sur un certain $\mathbf{R}^d$ vérifiant les hypothèses précédentes telle que $\|x_n - x^*\| \geqslant c\|x_0 - x^*\|$.

- Maintenant la démarche correspond à l'étude classique d'une suite récurrente. La fonction $\varphi : x \mapsto x - t\nabla f(x)$ est différentiable avec : $\mathrm{d}\varphi(x) : h \mapsto h - t\mathrm{H}f(x)\cdot h$. On a alors :

$$\begin{aligned}\mathrm{Sp}\,\mathrm{d}\varphi(x) &= \{1 - tv \,:\, v \in \mathrm{Sp}\,\mathrm{H}f(x)\}\\ &\subset \{1 - tv \,:\, v \in [\alpha, \lambda]\}.\end{aligned}$$

  Ainsi le rayon spectral de $\mathrm{d}\varphi(x)$ est inférieur à $r(t) = \max(|1 - t\alpha|, |1 - t\ell|)$ et comme $\mathrm{d}\varphi(x)$ est symétrique on a $\|\mathrm{d}\varphi(x)\|_{2\to 2} = r(t)$. Le minimum de $r(t)$ est atteint au point $t = \frac{2}{\alpha+\ell}$ où les deux termes sont égaux à $\frac{\ell-\alpha}{\ell+\alpha} < 1$. On a ensuite avec l'inégalité des accroissements finis : $\|x_{k+1} - x^*\| \leqslant \frac{\ell-\alpha}{\ell+\alpha}\|x_k - x^*\|$ de sorte qu'en itérant :

$$\|x_k - x^*\| \leqslant \left(\frac{\ell - \alpha}{\ell + \alpha}\right)^k \|x_0 - x^*\|.$$

  Ainsi la convergence est géométrique de rapport $\frac{\ell-\alpha}{\ell+\alpha}$. Ce rapport est d'autant plus petit que le rapport $\frac{\ell}{\alpha}$ est petit. On peut comprendre ce qu'il se passe sur l'exemple des fonctions quadratiques $x \mapsto x^\top Ax$. Dans ce cas α et ℓ correspond à la courbure de la surface dans les deux directions propres orthogonales, si la courbure est grande dans une direction, il faudrait prendre t petit afin d'éviter le balancement de l'autre côté dans la direction de faible variation il faudrait prendre t grand pour bien avancer. Le compromis fait sur t entraine les défauts évoqués précédemment ce qui nuit à la rapidité de la convergence pour un point x_0 "moyen".

Remarque.
1. *Suivant l'idée de résoudre l'équation de point critique, une méthode plus classique est la méthode de Newton. Le défaut de cette méthode est qu'elle demande l'évaluation et l'inversion d'une matrice de taille $d \times d$ à chaque étape ce qui peut s'avérer très couteux.*
2. *La méthode du gradient supposer qu'on sait évaluer la différentielle de la fonction. Maintenant si l'on se limite à la donnée de f on peut construire des algorithmes d'optimisation mais ces algorithmes ne donnent aucune information à nombre de pas fixés.*
3. *Sur la classe des fonctions vérifiant l'hypothèse 2, la méthode du gradient à pas fixe n'est pas optimal parmi les méthodes de degrés 1 (i.e. qui considère des évaluations de la fonction et de sa dérivée).*

Stabilité en première approximation

Le but de ce développement est d'énoncer un critère de stabilité des points d'équilibre d'une équation différentielle autonome. Dans le cas linéaire, la décomposition caractéristique permet de comprendre très précisément la stabilité du point d'équilibre. Dans le cas général on s'appuie sur le cas linéaire en utilisant qu'au voisinage du point critique la fonction supposée régulière est égal à son approximation linéaire + un petit terme correctif. Ce petit terme d'erreur demande de renforcer légèrement les hypothèses faite sur la partie linéaire, ce qui fera apparaître un cas critique où nous verrons qu'on ne peut en général rien dire.

Théorème. *Soit $f : \mathbf{R}^d \to \mathbf{R}$ de classe C^1 tel que $f(0) = 0$ avec $A = \mathrm{D}f(0)$ de spectre compris dans $\{z : \mathrm{Re}\, z < 0\}$. Dans ce cas 0 est un point d'équilibre stable i.e. pour tout $\delta > 0$ il existe $\alpha > 0$ tel que si $x_0 \in \overline{B}_\alpha$ alors l'unique solution au système différentiel vérifiant $x(0) = x_0$ est définie sur $[0, +\infty[$ et reste dans $\overline{B}_\delta$ et asymptotiquement stable i.e. $x(t) \to 0$ lorsque $t \to +\infty$.*

L'idée est que tant que la solution est proche du point d'équilibre, f sera assez bien approchée par sa partie linéaire et assurera que la solution reste proche de 0 un peu plus loin, et pas à pas on a que la solution est proche de 0 en tout temps. Mathématiquement on formalise cette idée par un argument de connexité.

- On note $f(x) = Ax + b(x)$ puis pour $\varepsilon > 0$ à déterminer δ_0 tel que $|b(x)| \leqslant \varepsilon |x|$ pour tout $\|x\| < \delta_0$.
- On se donne α à déterminer puis $x_0 \in \overline{B}_\alpha$ et $x : [0, T^+) \to \mathbf{R}^d$ une solution maximale en les temps positifs au système différentiel $x'(t) = f(x(t))$ avec la donnée de Cauchy $x(0) = x_0$.
- On fixe $0 < \delta \leqslant \delta_0$ puis on pose $I = \{0 \leqslant T < T^+ \ : \ \forall t \in [0, T],\ x(t) \in B_\delta\}$.

Montrons que I est non vide ouvert et fermé dans $[0, T^+)$. Déjà :

- Si $\alpha \leqslant \delta$ alors I contient 0 donc est non vide.
- Si $T_n \to T$ dans $[0, T^+)$ alors pour $t \in [0, T]$ si $t < T$ il existe T_n tel que $t \in [0, T_n]$ d'où $|x(t)| \leqslant \delta$ et finalement comme $T < T^+$ par hypothèse par continuité de x en T on a $|x(T)| = \lim_{n \to +\infty} |x(T_n)| \leqslant \delta$. Ainsi I est fermé.
- Soit $T \in I$ on veut montrer qu'il existe $h > 0$ tel que $T + h \in I$. On a $x'(t) = Ax(t) + b(x(t))$ qu'on réécrit $\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{e}^{-tA} x(t)) = \mathrm{e}^{-tA} b(x(t))$ qui s'intègre pour donner :

$$x(t) = \mathrm{e}^{tA} x_0 + \int_0^t \mathrm{e}^{(t-s)A} b(x(s)) \mathrm{d}s, \qquad \forall t \in [0, T^+).$$

  Avec le lemme de décroissance exponentielle il existe deux constantes $C, \sigma > 0$ tels que $\|\mathrm{e}^{tA}\| \leqslant C\mathrm{e}^{-t\sigma}$ de sorte qu'en passant aux normes :

$$\forall t \in [0, T], \quad |x(t)| \leqslant C\mathrm{e}^{-t\sigma}\alpha + C\varepsilon \int_0^t \mathrm{e}^{(t-s)\sigma} |x(s)| \mathrm{d}s.$$

  En appliquant le lemme de Gronwall à $\mathrm{e}^{t\sigma} x(t)$ on en déduit que $|x(t)| \leqslant C\alpha \mathrm{e}^{(C\varepsilon - \sigma)t}$ sur $[0, T]$. Choisissons à postériori δ_0 afin que $C\varepsilon < \sigma$ puis α tel que $C\alpha \leqslant \delta/2$. On a dans ce cas en notant $r = \sigma - C\varepsilon > 0$,

$$\forall t \in [0, T], \quad |x(t)| \leqslant \frac{1}{2} \delta \mathrm{e}^{-rt}.$$

  On en déduit par continuité de x qu'il existe $h > 0$ tel que $\|x(t)\| \leqslant \delta$ sur $[T, T + h]$, cela permet de conclure que I est ouvert. Par le lemme des bouts on en déduit que $T = T^+$ ce qui conclut à montrer la stabilité. Finalement la dernière inégalité est valable pour tout $T \geqslant 0$ d'où l'on déduit que $|x(t)| \to 0$.

Remarque. *1. Lorsque A possède une valeur propre de partie réelle nulle contrairement au cas linéaire on ne peut en générale rien dire, la solution peut être stable ou instable comme*

l'illustre l'exemple $f(x) = x^3$. Cette solution admet une unique solution maximale $x(t)$ par le théorème de Cauchy, $x_0 > 0$ par séparation des courbes intégrables on a $x(t) > 0$ en tout temps donc x est strictement croissante donc en particulier la solution n'est pas stable. Par contre si $x_0 < 0$ alors x est décroissante donc convergence vers un point équilibre donc vers 0 la solution est stable et asymptotiquement stable. En fait dans ce cas on le système est intégrable, on peut calculer explicitement sa solution mais je m'embrouille avec les signes.

2. *Dans le cas où A à une valeur propre de partie réelle > 0 on peut montrer que le système est instable, mais c'est beaucoup plus dur.*

Équation de la chaleur périodique

Le but de ce développement est d'étudier l'équation de la chaleur avec une donnée périodique. Cette équation représente par exemple la diffusion de la chaleur dans un anneau. On fait une étude fréquentielle de l'équation c'est à dire par l'analyse des coefficients de Fourier d'une éventuelle solution. Nous verrons ensuite comment retrouver la solution et ses propriétés.

Théorème. *Pour tout $f \in C_{2\pi}(\mathbf{R})$ il existe une unique fonction $u \in C_{2\pi,x}(\mathbf{R}_t^{\geqslant 0} \times \mathbf{R}_x)$ une fois dérivable en temps et deux fois en espace de dérivées continues solution de l'équation de la chaleur $\partial_t u = \partial_{xx}^2 u$ pour tout $(t,x) \in (0,+\infty) \times \mathbf{R}$ avec la donnée initiale $u_{|t=0} = f$.*

1. Analyse par les séries de Fourier

Dans la suite on fixe donc $f \in C_{2\pi}(\mathbf{R})$ et l'on se donne u une solution du problème précédent.

- Pour tout $t > 0$ comme $u(t,\cdot)$ est C^2 elle est développable en série de Fourier, on écrit :

$$u(t,x) = \sum_{n=-\infty}^{+\infty} c_n(t) \mathrm{e}^{inx} \qquad \text{avec} \quad c_n(t) = \frac{1}{2\pi} \int_0^{2\pi} u(t,x) \mathrm{e}^{-inx} \mathrm{d}x.$$

- Comme $\partial_t u(t,x)$ est continue elle est localement bornée ce qui permet d'appliquer le théorème de dérivation sous le signe intégrale à $c_n(t)$. On obtient ainsi avec l'équation de la chaleur et deux intégrations par parties :

$$\begin{aligned} c_n'(t) &= \frac{1}{2\pi} \int_0^{2\pi} \partial_t u(t,x) \mathrm{e}^{-inx} \mathrm{d}x \\ &= \frac{1}{2\pi} \int_0^{2\pi} \partial_{xx}^2 u(t,x) \mathrm{e}^{-inx} \mathrm{d}x \\ &= -n^2 c_n(t). \end{aligned}$$

  Ainsi $c_n(t)$ est solution d'une équation différentielle sur $(0,+\infty)$ qu'on résout facilement $c_n(t) = \mathrm{e}^{-n^2 t} c_n$.

- Finalement u est continue sur $R_+ \times \mathbf{R}$ donc uniforme continue sur $[0,1] \times [0,2\pi]$ par le théorème de Heine d'où l'on déduit que $\sup_{x \in [0,2\pi]} |u(t,x) - u(0,x)| \to 0$ lorsque $t \to 0^+$ de sorte que $c_n(t) \to c_n(f)$. On a ainsi montré que toute solution du problème précédent s'écrit :

$$\forall (t,x) \in (0,+\infty) \times \mathbf{R}, \quad u(t,x) = \sum_{n=-\infty}^{+\infty} c_n(f) \mathrm{e}^{inx - n^2 t} = f * K_t \qquad \text{où } K_t = \sum_{n=-\infty}^{+\infty} \mathrm{e}^{inx - n^2 t}$$

2. Synthèse de la solution

- Pour tout $a > 0$ on a sur $[a,+\infty[$ une domination de $c_n(f)\mathrm{e}^{inx - n^2 t}$ est de ses dérivées par $P(n)|c_n(f)|\mathrm{e}^{-n^2 a}$ où P est un polynôme en n. Cela montre que u est C^∞ sur $(0,+\infty) \times \mathbf{R}$ et qu'on peut dériver sous le signe somme ce qui permet de vérifier que $\partial_t u - \partial_{xx}^2 u = 0$ pour tout $(t,x) \in (0,+\infty) \times \mathbf{R}$.
- Il reste à établir que $u(t,x)$ tend vers $f(x)$ lorsque $t \to 0$. Si on veut utiliser le théorème de continuité sous la somme, il faut que $|c_n(f)|$ soit sommable ce qui n'est pas nécessairement le cas. Ça l'est si f suffisamment régulière par exemple C^1, nous avons donc résolue le problème dans ce cas.
- Pour montrer le problème dans le cas général on va utiliser qu'on a en fait résultat le problème sur une partie dense de $C_{2\pi}$. On introduit pour $t > 0$ l'opérateur :

$$A_t : f \in C_{2\pi}^0 \to K_t * f \in C_{2\pi}^0.$$

Le but est de montrer que A_t converge simplement vers Id lorsque $t \to 0$. On a montré que cela est le cas sur la partie dense $C^1_{2\pi}$. Pour en déduire le résultat général il suffit (et il suffit d'après BS) de montrer que $\|A_t\|$ est bornée au voisinage de 0. Cela découle de la majoration suivante :

$$\|A_t f - f\| \leqslant \|A_t g - g\| + 2\|A_t\|\|f - g\|.$$

Pour calculer $\|A_t\|$ on utilise une propriété analytique de l'équation de chaleur : $K_t \geqslant$ pour $t > 0$. Nous reviendrons peut être faire ce fait ultérieurement. En tout cas on en déduit que $\|A_t\| \leqslant \|K_t\|_{L^1} = c_0(K_t) = 1$.[30]

3. Principe du maximum

La propriété $K_t \geqslant 0$ est équivalente au fait que l'opérateur A_t. Cela signifie que si la donnée initiale - disons C^1 pour ne pas se mordre la queue - est positive alors la solution est positive ce qui traduit physiquement une part du 2nd principe de la thermodynamique : les transferts thermiques se font du chaud vers le froid. Mathématiquement il a plusieurs façon de voir ce fait :

1. la première plus astucieuse consiste à utiliser la formule de Poisson qui donne :

$$K_t = \sqrt{\frac{2\pi}{t}} \sum_{l \in \mathbf{Z}} \mathrm{e}^{-\frac{(x+2l\pi)^2}{4t}} > 0.$$

   En fait ce qu'il se passe derrière est que la formule de Poisson exprime la série de Fourier de la périodisé d'une fonction, or comme on peut s'y attendre le noyau de la chaleur périodique est le périodisé du noyau de la chaleur sur $\mathbf{R}$ à savoir $\mathrm{e}^{-\frac{x^2}{4t}}$.

2. la deuxième consiste à voir notre problème comme un cas particulier du principe du minimum : $\inf u(t,x) = \inf u(0,x)$. En particulier. Cela nous dit essentiellement que u ne peut attendre de minimum local en un point (t,x) avec $t > 0$ et peut se comprendre mathématiquement en voyant que les contraintes différentielles auxquelles sont soumises les minimums locaux sont difficilement compatible avec l'équation de la chaleur. Ce n'est pas vraiment le cas et pour raisonner précisément il faut perturber légèrement notre solution par exemple en considérant $v : t \mapsto u(t,x)\mathrm{e}^{-t}$.

30. En fait on a l'égalité $\|A_t\| = 1$.

Processus de Galton-Watson

Le but de ce développement est d'étudier le comportement du processus de Galton-Watson. Le processus de Galton-Watson de loi de reproduction p représente l'évolution d'une population où le nombre d'enfants de chaque individu est une variable aléatoire, précisément :

$$Z_0 = 1 \text{ p.s.}, \quad Z_{n+1} = \sum_{i=1}^{Z_n} X_{i,n}.$$

On supposera que les variables aléatoires $X_{i,n}$ sont i.i.d de loi p loi de probabilité sur $\mathbf{N}$ admettant une espérance m. En première approximation, le comportement de ce processus peut être quantifier par son espérance m suivant que $m < 1$, $m = 1$ ou $m > 1$. On s'intéressera exclusivement au cas $m = 1$ dit critique. On supposera $p_1 < 1$ auquel cas $Z_n = 1$ p.s.

Théorème. *La suite Z_n converge vers 1 p.s. De plus,*

$$\mathbf{P}(Z_n > 0) = \frac{2}{\sigma^2 n} + \frac{2}{\sigma^2}\left(\frac{2\rho}{3\sigma^4} - 1\right)\frac{\log(n)}{n^2} + o\left(\frac{\log(n)}{n^2}\right).$$

1. Observations

- Z_n est une chaîne de Markov sur $\mathbf{N}$ pour laquelle 0 est un état absorbant, les autres états sont transitoires car connectés à 0 car $p_0 > 0$; on en déduit que presque sûrement, $Z_n \to 0$ ou $Z_n \to +\infty$. On va chercher à quantifier cette alternative.
- Le fait fondamental sur le processus de Galton-Watson est que l'on connait la fonction génératrice du modèle $f_n(s) = \mathbf{E}[s^{Z_n}]$. En notant f la fonction génératrice de p on a :

$$\begin{aligned} f_n(s) &= \mathbf{E}(\mathbf{E}(s^{Z_n} \mid Z_{n-1})) \\ &= \mathbf{E}(f(s)^{Z_{n-1}}) \\ &= f_{n-1}(f(s)). \end{aligned}$$

  Comme de plus $f_0(s) = 1$ on en déduit que $f_n(s) = f^{\circ n}(s)$.
- Notons $q = \mathbf{P}(Z_n \to 0) = \mathbf{P}(\cup\{Z_n = 0\})$. Comme les évènements $\{Z_n = 0\}$ forment une suite croissante on a $q = \lim q_n$ où $q_n = \mathbf{P}(Z_n = 0) = f_n(0)$. Ainsi, q_n est suite récurrente de transition f.

2. Étude de la convergence

Pour étudier la convergence de q_n on fait une étude de fonctions sur f en commençant par déterminer ses points fixes.

- Comme $f(s) = \sum_{n=0}^{+\infty} p_n s^n$ avec $p_n \geqslant 0$ la fonction f est croissante et convexe. De plus les hypothèses assure que f est strictement croissante et strictement convexe.
- Ainsi f est au dessus de sa tangente en 1 d'équation $y = x$ donc $f(s) \geqslant s$ pour tout $s \in [0,1]$ et $f(s) > s$ pour $s \neq 1$ donc en particulier 1 est l'unique point fixe de f.
- La suite q_n est donc croissante, donc converge et sa limite ne peut être autre qu'un point fixe de f donc sa limite est 1.

On a donc montré qu'un processus de Galton-Watson critique s'éteint presque sûrement.

3. Vitesse de convergence

Dans la suite on note $u_n = 1 - q_n$.

- On suppose pour que f est de classe C^2 ou ce qui est équivalente que p a une variance finie. Un développement de Taylor en 1 donne :

$$1-f(s)=1-s+\frac{\sigma^2}{2}(1-s)^2+o(1-s)^2.$$

En inversant l'égalité précédente on obtient :

$$\frac{1}{1-f(s)}=\frac{1}{1-s}+\frac{\sigma^2}{2}+o(1)$$

On en déduit que $u_{n+1}^{-1}-u_n^{-1}\sim\frac{\sigma^2}{2}$ donc en sommant cette relation $u_n^{-1}\sim\frac{n\sigma^2}{2}$ ce qui donne :

$$u_n=\mathbf{P}(Z_n>0)\sim\frac{2}{\sigma^2 n}.$$

- Allons un grand plus loin dans le développement asymptotique. Il faut commencer par pousser le développement limité de f. Sous l'hypothèse d'un moment d'ordre 3 on a en notant $\rho=f^3(1)$,

$$1-f(s)=1-s+\frac{\sigma^2}{2}(1-s)^2-\frac{\rho}{6}(1-s)^3+o(1-s)^3.$$

Avec la même transformation :

$$\frac{1}{1-f(s)}=\frac{1}{1-s}+\frac{\sigma^2}{2}+\left(\frac{\sigma^4}{4}-\frac{\rho}{6}\right)(1-s)+o(1-s)$$

Ainsi,

$$\frac{1}{u_{n+1}}-\frac{1}{u_n}=\frac{\sigma^2}{2}+\left(\frac{\sigma^4}{4}-\frac{\rho}{6}\right)\sum_{k=1}^{n-1}(u_k+o(u_k)).$$

En utilisant le premier équivalent on trouve $\sum_{k=1}^n(u_k+o(u_k))=\frac{2}{\sigma^2}\log n+o(\log n)$ d'où :

$$\begin{aligned}u_n&=\frac{1}{\frac{n\sigma^2}{2}+\frac{2}{\sigma^2}\left(\frac{\sigma^4}{4}-\frac{\rho}{6}\right)\log n+o(\log n)}\\&=\frac{2}{\sigma^2 n}\left\{1-\frac{4}{\sigma^4}\left(\frac{\sigma^4}{4}-\frac{\rho}{6}\right)\frac{\log n}{n}+o\left(\frac{\log n}{n}\right)\right\}\\&=\frac{2}{\sigma^2 n}+\frac{2}{\sigma^2}\left(\frac{2\rho}{3\sigma^4}-1\right)\frac{\log(n)}{n^2}+o\left(\frac{\log(n)}{n^2}\right).\end{aligned}$$

Remarque. *1. Le premier équivalent donne également* $\mathbf{E}(Z_n\mid Z_n>0)\sim\frac{n\sigma^2}{2}$. *On peut montrer en renormalisant que* $\frac{2}{\sigma^2 n}Z_n\mid Z_n>0$ *converge vers une loi exponentielle standard.*

2. *Dans le cas sous-critique on montre que* $Z_n\to 0$ *p.s. Cette fois ci la convergence est au moins géométrique de rapport m. Cela résultat du résultat général des convergence des suites récurrentes avec ici on a* $f'(1)=m$. *De plus la convergence est exactement géométrique d'ordre m c'est à dire* $m^{-n}\mathbf{P}(Z_n>0)$ *converge vers une constante non nulle si f a un moment d'ordre 2, encore par un résultat général. Maintenant chose très intéressante un résultat de 1967 dû à Seneta, Vere-Jones et Heathcote établit que la convergence est exactement géométriquement d'ordre m si et seulement si* $\mathbf{E}_p(X\log^+(X))<+\infty$. *Cela donne un exemple remarquable où la convergence d'une suite récurrente n'est pas quantifier par la dérivée au point fixe de sa dynamique. De façon probabiliste on peut comprendre ce résultat en se disant que si cette condition n'est pas vérifié c'est que X admet une trop grande queue de distribution qui contribue de façon importante dans la moyenne. En bref on aura plus de familles nombreuses mais aussi plus de pères célibataires qui risque d'éteindre la population.*

3. *Dans le sur-critique la fonction génératrice possède un unique autre point fixe autre que* 1 *qui est la probabilité d'extinction du processus.*

Marche aléatoire sur $\mathbf{Z}^d$

On note $X_n = \sum_{k=1}^n U_k$ la marche aléatoire simple symétrique sur $\mathbf{Z}^d$ c'est à dire où les U_k sont des variables aléatoires i.i.d de loi $\mathcal{U}(\{\pm e_1, \cdots, \pm e_d\})$ avec $(e_1, \cdots, e_d)$ la base canonique de $\mathbf{R}^d$. Le but de ce développement est d'étudier le nombre d'états visités par la marche.

Théorème. *On introduit la variable aléatoire $C_n := \#\{X_0, \cdots, X_n\}$ comptant le nombre d'états visités au temps n. On a alors :*

- $\mathbf{E}(C_n) \sim n\alpha_d$ *où α_d est une constante non nulle pour $d \geqslant 3$;*
- $\mathbf{E}(C_n) \sim \frac{\pi n}{\log n}$ *pour $d \geqslant 2$;*
- $\mathbf{E}(C_n) \sim \frac{2\sqrt{2}}{\sqrt{\pi n}}$ *pour $d = 1$.*

- Au temps $n+1$ on a visité un état de plus si $X_{n+1} \notin \{X_0, \cdots, X_n\}$, donc par récurrence

$$C_n = 1 + \sum_{k=1}^{n} \mathbf{1}_{X_i \notin \{X_0, \cdots, X_{k-1}\}}.$$

  Intéressons nous à l'évènement $X_k \notin \{X_0, \cdots, X_{k-1}\}$. En regardant le parcours inverse $(X_k, \cdots, X_0)$ c'est à dire la marche partant de X_k et revenant en arrières, cette évènement correspond selon ce point de vue, au fait que la marche inversée ne reviennent pas en son point de départ. Or cette nouvelle à la même loi que l'ancienne par homogénéité et symétrie d'où $\mathbf{P}(X_k \notin \{X_0, \cdots, X_{k-1}\}) = \mathbf{P}(\tau \geqslant k)$ où τ désigne le temps de retour $\tau = \inf\{n \geqslant 1 \,|\, X_n = 0\}$. Nous avons donc :

$$\mathbf{E}(C_n) = 1 + \sum_{k=1}^{n} \mathbf{P}(\tau > k).$$

- On a ainsi $n^{-1}\mathbf{E}(C_n) \longrightarrow \alpha$ où $\alpha = \alpha_d := \mathbf{P}(\tau = \infty)$. Le théorème de Polya dit précisément que α_d vaut 0 pour $d \in \{1,2\}$ et est différent de 0 pour $d \geqslant 3$. Pour $d \geqslant 3$ on a le premier terme du développement asymptotique de $\mathbf{E}(C_n)$, par contre pour $d \in \{1,2\}$ on a seulement l'estimation $\mathbf{E}(C_n) = o(n)$ et l'on va chercher un équivalent. Dans le cadre du développement on se restreint à $d = 2$.
- Intéressons nous aux retours à l'origine de la marche aléatoire. Comme la somme des composantes de X_n à la parité de n, ces retours ne peuvent s'effectuer qu'à des dates paires. En conditionnant par rapport au temps de premier retour et en utilisant la propriété de Markov forte on obtient l'égalité suivante :

$$\mathbf{P}(X_{2n} = 0) = \sum_{k=1}^{n} \mathbf{P}(X_{2(n-k)} = 0)\mathbf{P}(\tau = 2k).$$

  On voit ainsi apparaître un produit de Cauchy ce qui conduit à introduire les fonctions génératrices

$$f(z) = \sum_{n=0}^{\infty} \mathbf{P}(X_{2n} = 0) z^n,$$
$$g(z) = \sum_{n=1}^{\infty} \mathbf{P}(\tau = 2n) z^n.$$

  qui permettent de réécrire l'égalité précédente sous la forme $f(z)g(z) = f(z) - 1$ d'où $g(z) = 1 - 1/f(z)$. Maintenant on s'intéresse non pas aux $\mathbf{P}(\tau = 2n)$ mais plutôt aux $\mathbf{P}(\tau > 2n)$ ce qui nous conduit à introduire une troisième série génératrice :

$$h(z) = \sum_{n=1}^{\infty} \mathbf{P}(\tau > 2n) z^n = \frac{1 - g(z)}{1 - z} = \frac{1}{(1-z)f(z)}.$$

- Intéressons nous donc à f et aux probabilités $\mathbf{P}(X_{2n}=0)$. On calcule que :

$$\mathbf{P}(X_{2n}=0)=\frac{1}{4^n}\sum_{\substack{i_1+i_2+j_1+j_2=2n\\ i_1=i_2,j_1=j_2}}\binom{n}{i_1,i_2,j_1,j_2}=\frac{1}{4^n}\sum_{n_1+n_2=2n}\frac{(2n)!}{n_1!^2n_2!^2}=\frac{1}{4^n}\binom{2n}{n}\sum_{n_1+n_2=2n}\binom{n}{n_1}\binom{n}{n_2}=\frac{1}{4^n}\binom{2n}{n}^2.$$

Ces coefficients ne sont pas ceux d'une série entière usuelle, pour récupérer de l'information sur nos coefficients on va regarder le comportement en 1 de nos séries génératrices et recourir à un théorème taubérien. On a avec la formule de Stirling $\mathbf{P}(X_{2n}=0)\sim\frac{1}{n\pi}$ donc par sommation des équivalents sur une série entière $f(x)\sim\frac{1}{\pi}\log(1-x)$ lorsque $x\to1^-$ puis :

$$h(x)\underset{x\to1^-}{\sim}\frac{\pi}{(1-x)\log(1-x)}.$$

- On applique le théorème d'Hardy-Littlewood à la série entière $\log(1-z)h(z)$ de coefficients $\sum_{k=1}^n\frac{a_k}{n-k}$ ce qui après réarrangement donne :

$$\pi+o(1)=\frac{1}{m}\sum_{k=1}^{m-1}a_kH_{m-k}$$

où l'on note $H_j=1+\frac{1}{2}+\cdots+\frac{1}{j}$ la j-ème somme harmonique. En utilisant que $H_j=\log j+O(1)$ comme $a_k=o(1)$ on a :

$$\pi+o(1)=\frac{1}{m}\sum_{k=1}^{m-1}a_k\log(m-k).$$

Maintenant l'idée est que le logarithme croissante si lentement qu'on peut le considérer constant égal à $\log m$. Précisément en utilisant la propriété de morphisme on a :

$$\pi+o(1)=\frac{\log m}{m}\sum_{k=1}^{m-1}a_k+r_m$$

où :

$$\begin{aligned}|r_m|&=\left|\frac{1}{m}\sum_{k=1}^{m-1}a_k\log\left(1-\frac{k}{m}\right)\right|\\&\leqslant o(1)+\frac{a_{m_0}}{m}\sum_{k=m_0}^{m}\left|\log\left(1-\frac{k}{m}\right)\right|\\&\leqslant o(1)+a_{m_0}\underbrace{\int_0^1|\log(1-t)|\mathrm{d}t}_{=:C<\infty}.\end{aligned}$$

On en déduit que $\limsup|r_m|\leqslant Ca_{m_0}$ puis en faisant rendre $m_0\to+\infty$ on en déduit le résultat :

$$\mathbf{E}(C_m)\sim\frac{\pi}{m\log(m)}.$$

Remarque. *1. En dimension 1 on peut utiliser la même démarche, le fait remarquable est qu'en fait les séries entières qu'on manipule sont des séries usuelles. Cela permet de déduire des expressions exactes des coefficients et ne nécessite plus le recours à la théorie taubérienne.*

2. En dimension 2 on peut montrer le résultat plus fort : $\mathbf{P}(\tau\geqslant n)\sim\frac{\pi}{\log(n)}$. *Il est plus aisé de déduire ce résultat de la forme faible vu précédemment. D'une part nous avons par décroissance de* $\log(n)a_n\geqslant\pi+o(1)$ *d'où* $\pi\geqslant\limsup_{n\to\infty}\log(n)a_n$. *De l'autre côté introduisons* $q\in\mathbf{N}^*$. *Nous*

avons $\pi = \lim_{,\to\infty} \frac{\log(qn)}{qn}\sum_{l=1}^{qn} a_k$ *or,*

$$\begin{aligned}\frac{\log(qn)}{qn}\sum_{l=1}^{qn} a_l &= \frac{1}{q}\frac{\log(n)}{N}\sum_{l=1}^{qn} a_l + o(1)\\ &\leqslant \frac{1}{q}\frac{\log(n)}{n}(qn-n)a_n + \frac{1}{q}\frac{\log(n)}{n}\sum_{l=1}^{n} a_l + o(1)\\ &= \left(1-\frac{1}{q}\right)\log(n)a_n + \frac{1}{q}\pi + o(1)\end{aligned}$$

d'où

$$\pi \leqslant \left(1-\frac{1}{q}\right)\liminf_{N\to\infty} \log(n)a_n + \frac{1}{q}\pi.$$

Cela est valable quelque soit $q \in \mathbf{N}^*$ *; en faisant tendre* $q \to \infty$ *on en déduit que*

$$\pi \leqslant \liminf_{n\to\infty} \log(n)a_n$$

donc finalement $\log(n)a_n \longrightarrow \pi$

3. *On peut montrer des résultats de convergence pour* C_n*. Toujours* $n^{-1}C_n$ *converge en probabilité vers* α*. Dans le cas* $d = 2$ *on a* $\frac{\log n}{n}C_n$ *converge en probabilité vers* π*. Ces résultats sont relativement simples à obtenir avec les équivalents précédentes de* $\mathbf{E}(C_n)$*, le gros du travail est fait (voir pour cela le document cité).*
4. *Il existe une expression explicite de la quantité* α_d *pour* $d \geqslant 3$ *à savoir*

$$\alpha_d^{-1} = \frac{d}{(2\pi)^d}\int_{[-\pi,\pi]^d} \frac{\mathrm{d}x_1\cdots\mathrm{d}x_d}{d-\cos(x_1)-\cdots-\cos(x_d)}. \tag{1}$$

En particulier dans le cas $d = 3$*,* GLASSER *et* ZUCKER *(1977) on trouvé une expression analytique*

$$\alpha_3 = \left\{\frac{\sqrt{6}}{32\pi^3}\Gamma\left(\frac{1}{24}\right)\Gamma\left(\frac{5}{24}\right)\Gamma\left(\frac{7}{24}\right)\Gamma\left(\frac{11}{24}\right)\right\}^{-1} \simeq 0.6595\cdots \tag{2}$$

Par contre pour $d \geqslant 4$ *on ne connait pas d'expression semblable pour* α_d*.*

Formule d'inversion de Fourier

1. Formule d'inversion sur la périodisé

Soit $f \in S(\mathbf{R}^d)$, on note $f_T(x) = \sum_{l\in\mathbf{Z}} f(x+Tl)$ la T-périodisé de f pour T un réel strictement positif.

- Comme $f \in S(\mathbf{R})$ on a $|f(t)| \lesssim \frac{C}{1+t^2}$ quelque soit $t \in \mathbf{R}$. En utilisant cela on a la convergence uniforme de la série précédente sur tout intervalle fermé de $\mathbf{R}$. Précisément sur $[-A,A]$ on a pour $|l| \geqslant A/T$,
$$|f(x+Tl)| \lesssim \frac{C}{1+(|l|T-A)^2}.$$
- En appliquant le point précédent à $f' \in S(\mathbf{R})$ on voit que f_T est de classe C^1.
- Le théorème de Dirichlet permet ainsi d'écrire f_T comme somme de sa série de Fourier. Calculons alors les coefficients de Fourier de f_T :

$$\begin{aligned}c_{n,T}(f_T) &= \frac{1}{T}\int_0^T f_T(x)\mathrm{e}^{-\frac{2\mathrm{i}\pi nx}{T}}\mathrm{d}x\\ &= \frac{1}{T}\sum_{l=-\infty}^{+\infty}\int_0^T f(x+Tl)\mathrm{e}^{-\frac{2\mathrm{i}\pi nx}{T}}\mathrm{d}x\\ &= \frac{1}{T}\sum_{l=-\infty}^{+\infty}\int_{Tl}^{T(l+1)} f(x)\mathrm{e}^{-\frac{2\mathrm{i}\pi nx}{T}}\mathrm{d}x\\ &= \frac{1}{T}\widehat{f}(2n\pi/T).\end{aligned}$$

On obtient ainsi :

$$\sum_{l=-\infty}^{+\infty} f(x+Tl) = \frac{1}{T}\sum_{n=-\infty}^{+\infty} \widehat{f}\left(\frac{2n\pi}{T}\right) \mathrm{e}^{\frac{2in\pi x}{T}}.$$

2. Limite $T \to +\infty$

On va faire tendre la période vers l'infini.

- Pour le premier terme on a $f_T(x) = f(x) + \sum_{l\neq 0} f(x+lT)$ or pour $l \neq 0$ on a $|f(x+Tl)| \leqslant \frac{C}{1+(|l|-1/2)^2T^2}$ si $T \geqslant 2|x|$ de sorte que par convergence.
- On reconnait dans le terme de droite, au facteur 2π près, l'approximation des rectangles de l'intégrale $\int_{\mathbf{R}} \widehat{f}(y)\mathrm{e}^{ixy}\mathrm{d}y$. En faisant tendre $T \to +\infty$ on s'attend donc à voir ce terme convergence vers cette dernière intégrale. Pour montrer cela on estime l'erreur de quadrature élémentaire, on note à x fixé $g : y \mapsto \widehat{f}(y)\mathrm{e}^{ixy}$ la fonction intégrée et $a_n = \frac{2n\pi}{T}$. On a alors :

$$\begin{aligned}
&\left|\frac{2\pi}{T}\widehat{f}\left(\frac{2n\pi}{T}\right)\mathrm{e}^{\frac{2in\pi x}{T}} - \int_{2n\pi/T}^{2(n+1)\pi/T} \widehat{f}(y)\mathrm{e}^{ixy}\mathrm{d}y\right| \\
&= \left|(a_{n+1}-a_n)g(a_n) - \int_{a_n}^{a_{n+1}} g(y)\mathrm{d}y\right| \\
&\leqslant \int_{a_n}^{a_{n+1}} |g(y)-g(a_n)|\mathrm{d}y \\
&\leqslant \frac{1}{2}\sup_{[a_n,a_{n+1}]} |g'|(a_{n+1}-a_n)^2.
\end{aligned}$$

Comme $g' \in S(\mathbf{R})$ on a $|g'(y)| \lesssim \frac{1}{1+|y|^2}$. Ce dernier terme se majore alors à une constante près par :

$$\frac{1}{1+\left(\frac{2n\pi}{T}\right)^2}\cdot\frac{1}{T^2}.$$

Ainsi on peut estimer l'erreur global :

$$\begin{aligned}
\left|\frac{2\pi}{T}\sum_{n=-\infty}^{+\infty}\widehat{f}\left(\frac{2n\pi}{T}\right)\mathrm{e}^{\frac{2in\pi x}{T}} - \int_{-\infty}^{+\infty}\widehat{f}(y)\mathrm{e}^{ixy}\mathrm{d}y\right| &\lesssim \frac{1}{T^2}\sum_{n=-\infty}^{+\infty}\frac{1}{1+\left(\frac{2n\pi}{T}\right)^2} \\
&\lesssim \frac{1}{T^2}\left(1+2\pi T\int_{-\infty}^{+\infty}\frac{\mathrm{d}x}{1+x^2}\right) \\
&\lesssim \frac{1}{T}.
\end{aligned}$$

Remarque. *1. On peut déduire de ce résultat par la formule de dualité et la densité de la classe de Schwartz la formule d'inversion générale.*

2. Présenté ici dans $\mathbf{R}$ ce résultat s'adapte à $\mathbf{R}^d$ en manipulant des séries de Fourier de plusieurs variables.

3. La formule établit dans le 1. appelée formule sommatoire de Poisson est à la base d'autres identités remarquables comme la suivante...

Théorème central limite de Linderberg

Le but de ce développement est de montrer le résultat suivant qui optimise les hypothèses du théorème limite central selon l'idée de preuve classique avec les fonctions caractéristiques. On autorise ainsi des variables aléatoires non identiquement distribués mais toujours indépendantes.

Théorème. *Soient X_k ($k \geqslant 1$) des variables aléatoires indépendantes centrées avec un moment d'ordre 2. On note $\sigma_k^2 = \mathbf{E}(X_k^2)$ puis,*

$$s_n^2 = \sum_{k=1}^{n} \sigma_k^2.$$

On suppose la condition suivante dite de Linderberg :

$$\forall \varepsilon > 0, \quad \frac{1}{s_n^2} \sum_{k=1}^{n} \mathbf{E}\left(\mathbf{1}_{|X_k| > \varepsilon s_n} |X_k|^2\right) \underset{n\to\infty}{\longrightarrow} 0.$$

Alors,

$$Z_n := \frac{1}{s_n} \sum_{k=1}^{n} X_k \xrightarrow{(d)} \mathcal{N}(0,1).$$

En réduisant on se ramène à $s_n^2 = 1$. L'idée est la suivante : on a $\varphi_{Z_n}(t) = \prod_{k=1}^n \varphi_{X_k}(t)$ puis $\varphi_{X_k}(t) \approx 1 - \frac{1}{2}\sigma_k^2 t^2 \approx \mathrm{e}^{-\frac{1}{2}\sigma_k^2 t^2}$ de sorte que $\varphi_{Z_n}(t) \approx \prod_{k=1}^n \mathrm{e}^{-\frac{1}{2}\sigma_k^2 t^2} = \mathrm{e}^{-\frac{1}{2}t^2}$. Pour justifier le résultat il faut préciser les approximations.

1. Première approximation

La première approximation consiste en un développement à l'ordre 2 de φ_{X_k} ; pour préciser le reste on utilise des formules de Taylor pour $x \mapsto \mathrm{e}^{\mathrm{i}x}$ qu'on intégrer contre $\mathbf{P}_{X_k}$.

- On a pour tout $x \in \mathbf{R}$,
$$\mathrm{e}^{\mathrm{i}x} - \sum_{k=1}^{m} \frac{\mathrm{i}^k x^k}{k!} = \frac{\mathrm{i}^{m+1}}{m!} \int_0^x (x-y)^n \mathrm{e}^{\mathrm{i}y} \mathrm{d}y.$$
On en déduit que : $\left|\mathrm{e}^{\mathrm{i}x} - \sum_{k=1}^m \frac{\mathrm{i}^k x^k}{k!}\right| \leqslant \frac{|x|^{m+1}}{(m+1)!}$. Cette approximation est efficace pour des petites valeurs de x, pour des valeurs de x plus grandes, typiquement $x \to +\infty$ le bonne ordre de grandeur n'est pas $|x|^{m+1}$ mais $|x|^m$, on préféra la borne : $\left|\mathrm{e}^{\mathrm{i}x} - \sum_{k=0}^m \frac{\mathrm{i}^k x^k}{k!}\right| \leqslant \frac{2|x|^m}{m!}$ qui s'obtient à partir de la précédente à l'ordre $n-1$ en ajoutant le terme $\frac{\mathrm{i}^m x^m}{m!}$ avec l'inégalité triangulaire.
- On va intégrer ces inégalités contre $\mathbf{P}_{X_k}$ pour $m = 2$ en découpant suivant les évènements X_k petit : $\{|X_k| \leqslant \varepsilon\}$ ou X_k grand : $\{X_k > \varepsilon\}$ en utilisant respectivement la première et la seconde inégalité ; on obtient ainsi :
$$\begin{aligned} |\varphi_{X_k}(t) - (1 - \frac{1}{2}\sigma_k^2 t^2)| &\leqslant \frac{1}{6}\mathbf{E}[|X_k|^3 \mathbf{1}_{|X_k|\leqslant\varepsilon}] + \mathbf{E}[|X_k|^2 \mathbf{1}_{|X_k|>\varepsilon}] \\ &\leqslant \frac{\varepsilon|t|^3}{6}\sigma_k^2 + \mathbf{E}[|X_k|^2 \mathbf{1}_{|X_k|>\varepsilon}]. \end{aligned}$$

2. Seconde approximation

La seconde approximation consiste en un développement à l'ordre 1 de $\mathrm{e}^{-\frac{1}{2}\sigma_k^2 t^2}$.

- On utilise le développement en série de l'exponentielle pour obtenir l'inégalité suivante $|\mathrm{e}^z - 1 - z| \leqslant |z|^2 \mathrm{e}^{|z|}$.
- On a ainsi :
$$|(1 - \frac{1}{2}\sigma_k^2 t^2) - \mathrm{e}^{-\frac{1}{2}\sigma_k^2 t^2}| \leqslant \frac{t^4}{4}\sigma_k^4 \mathrm{e}^{\frac{\sigma_k t^2}{2}} \leqslant \frac{t^4 \mathrm{e}^{\frac{1}{2}t^2}}{4}\sigma_k^4.$$

3. Répercutions des erreurs sur le produit et conclusion

La troisième approximation demande de contrôler la répercutions des erreurs sur le produit.

- On utilise l'inégalité élémentaire suivante valable pour des complexes a_i, b_i de module inférieur à 1 :
$$\left|\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i\right| \leqslant \sum_{i=1}^{n} |a_i - b_i|.$$
- On obtient ainsi :
$$\left|\varphi_{Z_n}(t) - \prod_{k=1}^{n}\left(1-\frac{1}{2}\sigma_k^2 t^2\right)\right| \leqslant \frac{\varepsilon|t|^3}{6} + \sum_{k=1}^{n} \mathbf{E}[|X_k|^2 \mathbf{1}_{|X_k|>\varepsilon}] + \frac{t^4 \mathrm{e}^{\frac{t^2}{2}}}{4}\sum_{k=1}^{n}\sigma_k^4.$$
Il nous faut contrôler les deux derniers termes, l'hypothèse assure que le second terme tend vers 0, elle assure aussi la convergence du dernier terme car :
$$\sigma_k^2 \leqslant \varepsilon^2 + \mathbf{E}[|X_k|^2 \mathbf{1}_{|X_k|>\varepsilon}] \leqslant \varepsilon^2 + \sum_{k=1}^{n} \mathbf{E}[|X_k|^2 \mathbf{1}_{|X_k|>\varepsilon}]$$
de sorte que $\max_{1\leqslant k\leqslant n} \sigma_k^2 \to 0$ puis :
$$\sum_{k=1}^{n} \sigma_k^4 \leqslant (\max_{1\leqslant k\leqslant n} \sigma_k^2) \to 0.$$

4. Une illustration

On prend $Y_k \sim \mathrm{Ber}\left(\frac{1}{k}\right)$ indépendantes puis $X_k = Y_k - \frac{1}{k}$. On a $s_n^2 = \sum_{k=1}^{n} \frac{k-1}{k^2} \sim \log(n)$ et comme $|X_k| \leqslant 2$ p.s. la condition de Linderberg est facilement vérifié et l'on a :
$$\frac{1}{\sqrt{\log(n)}} \sum_{k=1}^{n} X_k \underset{n\to+\infty}{\longrightarrow} \mathcal{N}(0,1).$$
On peut alors conclure en notant $S_n = \sum_{k=1}^{n} Y_k$ que :
$$\frac{S_n - \log(n)}{\sqrt{\log(n)}} \underset{n\to+\infty}{\longrightarrow} \mathcal{N}(0,1).$$
La variable aléatoire S_n apparaît dans diverses situations, c'est par exemple le nombre de cycles d'une permutation aléatoire de S_n.

Remarque. *On déduit en particulier de ce résultat le théorème central limite classique car alors $s_n^2 = n\sigma^2$ et la condition de Linderberg s'écrit :*
$$\frac{1}{n\sigma^2}\sum_{k=1}^{n} \mathbf{E}(\mathbf{1}_{|X_k|>\varepsilon\sqrt{n}\sigma^2}|X_k|^2) = \frac{1}{\sigma^2}\mathbf{E}(\mathbf{1}_{|X_1|>\varepsilon\sqrt{n}\sigma^2}|X_k|^2) \overset{\text{CVD}}{\longrightarrow} 0.$$

Formule des compléments

L'objectif de ce développement est de montrer que : $\forall s \in \mathbf{C}$ t.q. $\mathrm{Re}\, s \in (0,1)$,

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin \pi s}.$$

- Comme $s \mapsto \Gamma(s)\Gamma(1-s)$ est holomorphe sur $\{s \in \mathbf{C} : \mathrm{Re}\, s \in (0,1)\}$ il suffit par le principe des zéros isolés de montrer le résultat pour $s \in (0,1)$ ce qu'on supposera dans la suite.
- On commence par écrire le terme de gauche comme une intégrale double à l'aide de Fubini :

$$\Gamma(s)\Gamma(1-s) = \int_{(0,+\infty)^2} u^{s-1}\mathrm{e}^{-u} v^{-s}\mathrm{e}^{-v}\,\mathrm{d}u\,\mathrm{d}v.$$

  Simplifions cette intégrale avec le changement de variable $(x,y) = (u/v, v)$ de sorte que $u^{s-1}v^{-s} = x^{s-1}y^{s-1}y^{-s} = x^{s-1}y^{-1}$ puis :

$$\begin{aligned}\Gamma(s)\Gamma(1-s) &= \int_{(0,+\infty)^2} x^{s-1}\mathrm{e}^{-(x+1)y}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{+\infty} \frac{x^{s-1}}{(1+x)}\,\mathrm{d}x.\end{aligned}$$

  Récrivons par commodité les puissances sous forme exponentielle en posant $x = \mathrm{e}^t$,

$$\Gamma(s)\Gamma(1-s) = \int_{\mathbf{R}} \frac{\mathrm{e}^{st}}{1+\mathrm{e}^t}\,\mathrm{d}t.$$

- La dernière intégrale s'exprime à l'aide de fonctions usuelles qu'on sait méromorphes sur $\mathbf{C}$ ce qui invite à appliquer le théorème des résidus. On considère ainsi la fonction méromorphe $f : z \mapsto \frac{\mathrm{e}^{sz}}{1+\mathrm{e}^z}$ qui n'admet que des pôles simples en les $(2\mathbf{Z}+1)\mathrm{i}\pi$. On va alors intégrer sur le contour suivant rectangulaire passant par les sommets $-R, R, R+2\mathrm{i}\pi$ et $-R+2\mathrm{i}\pi$. On peut motiver le choix du contour par rapport aux propriétés de f, si $z = t+\mathrm{i}\sigma$ son module est $\frac{\mathrm{e}^{ts}}{\sqrt{1+2\mathrm{e}^t\cos(\sigma)+\mathrm{e}^{2t}}}$ donc décroit exponentielle lorsque $t \to \pm\infty$ et f possède une symétrie intéressante $f(z+2\mathrm{i}\pi) = \mathrm{e}^{2\mathrm{i}\pi s} f(z)$. Finalement f ne possède pas de pôle sur le contour et un unique à l'intérieur $\mathrm{i}\pi$ où son résidu vaut $\frac{\mathrm{e}^{s\mathrm{i}\pi}}{\mathrm{e}^{\mathrm{i}\pi}} = -\mathrm{e}^{s\mathrm{i}\pi}$.
- Le théorème des résidus donne :

$$-2\mathrm{i}\pi\mathrm{e}^{s\mathrm{i}\pi} = (1-\mathrm{e}^{2\mathrm{i}\pi s})\int_{-R}^{R} \frac{\mathrm{e}^{st}}{1+\mathrm{e}^t}\,\mathrm{d}t + \mathrm{i}\int_0^{2\pi} (f(R+\mathrm{i}\theta) - f(-R+\mathrm{i}\theta))\mathrm{d}\theta.$$

  Maintenant pour $|f(R+\mathrm{i}\theta)| \leqslant \mathrm{e}^{R(s-1)}$ et $|f(-R+\mathrm{i}\theta)| \leqslant \mathrm{e}^{-sR}$ de sorte que comme $s \in (0,1)$ la dernière intégrale tend vers 0 lorsque $R \to +\infty$. En prenant cette limite nous avons donc :

$$\int_{-\infty}^{+\infty} = \frac{2\mathrm{i}\pi\mathrm{e}^{s\mathrm{i}\pi}}{\mathrm{e}^{2\mathrm{i}\pi s}-1} = \frac{\pi}{\sin \pi s}.$$

Problème des moments

On se pose la question suivante : une mesure μ sur $\mathbf{R}$ est-elle déterminée par la donnée de ses moments (dont on suppose l'existence à tout ordre) $m_n(\mu) = \int_{\mathbf{R}} x^n \mathrm{d}\mu$? C'est à dire que si μ et ν sont deux mesures sur $\mathbf{R}$ avec les mêmes moments à tout ordre sont elles égales.

Théorème. *Soient μ et ν deux lois de probabilités sur $\mathbf{R}$ avec les mêmes moments à tout ordre.*

1. *Si μ et ν sont à support compacts alors*

1. Cas de mesures à support compact

On suppose μ et ν supportés dans $[0,1]$.

- Par le théorème d'approximation de Weierstrass il existe pour $f \in C([0,1],\mathbf{R})$ une suite f_n de polynômes tels que $f_n \xrightarrow{L^\infty} f$. Ainsi,
$$\int_0^1 f_n \mathrm{d}\mu \longrightarrow \int_0^1 f \mathrm{d}\mu.$$
Ainsi $\int f \mathrm{d}\mu = \int f \mathrm{d}\nu$ pour toute fonction $f \in C([0,1],\mathbf{R})$.
- En prenant $f(x) = \mathrm{e}^{itx}$ on en déduit que $\widehat{\mu} = \widehat{\nu}$ puis $\mu = \nu$ comme la fonction caractéristique caractérise la loi.

2. Critère d'analycité

Les moments de μ existant à tout ordre, ces derniers s'obtiennent à partir des dérivées de $\widehat{\mu}$, précisément on a $\widehat{\mu}^{(n)}(t) = \int_{\mathbf{R}} \mathrm{i}^n t^n \mathrm{e}^{\mathrm{i}tx} \mathrm{d}\mu(x)$. L'idée est que sous une bonne hypothèse de décroissance des moments, on va pouvoir montrer l'analycité de la fonction caractéristique. Ainsi si une autre mesure ν à les mêmes moments, $\widehat{\mu}$ et $\widehat{\nu}$ seront deux fonctions analytiques égales sur un voisinage de 0 donc partout égales et l'on pourra conclure comme précédemment.

Supposons que la série entière $\sum m_n t^n/n!$ ait un rayon de convergence $R > 0$. On va voir que cela implique que $\widehat{\mu}$ est développable en série entière au voisinage de tout point. Pour cela il faut comparer $\widehat{\mu}$ à son développement de Taylor, on va faire cela en écrivant une formule de Taylor sur $\mathrm{e}^{\mathrm{i}tx}$ puis en l'intégrant en x.

- Par l'inégalité de Taylor-Lagrange on a :
$$\left| \mathrm{e}^{\mathrm{i}(t+h)x} - \sum_{k=0}^{n} \frac{\mathrm{i}^k h^k x^k}{k!} \mathrm{e}^{\mathrm{i}tx} \right| \leqslant \frac{|t|^{n+1}|x|^{n+1}}{(n+1)!}.$$
On a dès lors :
$$\left| \widehat{\mu}(t+h) - \sum_{k=0}^{n} \frac{h^k}{k!} \widehat{\mu}^{(k)}(t) \right| \leqslant \frac{|t|^{n+1}}{(n+1)!} |m|_{n+1}$$
où l'on note $|m|_n = \int_{\mathbf{R}} |x|^n \mathrm{d}\mu$.
- On a par hypothèse $\frac{m_n r^n}{n!} = o(1)$ pour tout $0 \leqslant r < R$. On veut passer de m_n à $|m|_n$, déjà pour n pair il n'y a aucune différence puis pour n impair en utilisant que $|x|^{2n-1} \leqslant 1 + |x|^{2n}$ on a $|m|_{2n-1} \leqslant 1 + |m|_{2n}$. Pour $r < s < R$ on a alors pour n suffisamment grand :
$$\frac{|m|_{2n-1} r^{2n-1}}{(2n-1)!} \leqslant \frac{r^{2n-1}}{(2n-1)!} + \frac{m_{2n} s^{2n}}{(2n)!} = o(1).$$
On en déduit ainsi que pour tout h tel que $|h| < R$,
$$\widehat{\mu}(t+h) = \lim_{n \to +\infty} \sum_{k=0}^{n} \frac{h^k}{k!} \widehat{\mu}^{(k)}(t).$$
Ainsi $\widehat{\mu}$ est analytique.

- Pour conclure si ν et μ ont les mêmes moments m_n et si la série entière $\sum \frac{m_n}{n!} z^n$ a un rayon de convergence > 0 ce qui revient à dire par le critère d'Hadamard que $\liminf m_n^{1/n}/n > 0$ alors $\widehat{\mu}$ et $\widehat{\nu}$ sont analytiques, ayant les mêmes dérivées en 0 elle coïncide sur un voisinage ouvert de 0 donc sont égales par le principe des zéros isolés.

3. Méthode des moments

Si μ_n et μ sont des lois de probabilités sur $\mathbf{R}$ avec des moments à tout ordre tels que :

- $m_k(\mu_n) \to m_k(\mu)$ pour tout $k \in \mathbf{N}$;
- μ est caractérisé par ses moments.

Alors μ_n converge étroitement vers μ. On applique le critère de Prokhorof :

- $\mu_n([-A,A]) \leqslant \frac{1}{A^2} \sup_{n\in\mathbf{N}} m_2(\mu_n)$ (tension)
- si μ_{n_l} converge étroitement vers ν alors il existe $X_{n_l} \sim \mu_{n_l}$ et $X \sim \nu$ tels que $X_{n_l} \to X$ p.s. (Skorokhod) puis comme $X_{n_l}^k$ est équi-intégrable car borné dans L^2 on a par Vitalli : $\mathbf{E}(X_{n_l}^k) \to \mathbf{E}(X^k)$. Ainsi, $m_k(\nu) = \mu_k(\mu)$ pour tout k donc $\mu = \nu$.

Remarque. *1. La loi normale est caractérisé par ses moments car $|m_n| \leqslant n!$.*

2. La loi log-normale, exponentielle d'une normale, de densité de $x \mapsto \frac{1}{\sqrt{2\pi}x} \mathrm{e}^{-\frac{\log(x)^2}{2}} \mathbf{1}_{x>0}$ n'est pas caractérisé par ses moments car possède les mêmes moments que la loi de densité $g(x) = f(x)(1+\sin(2\pi\log(x))\mathbf{1}_{x>0}$. En effet,

$$\begin{aligned}\int_0^{+\infty} x^k f(x)\sin(2\pi\log(x) &= \int_0^{+\infty} x^k \frac{1}{\sqrt{2\pi}x} \mathrm{e}^{-\frac{\log(x)^2}{2}} \sin(2\pi\log(x)\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \mathrm{e}^{ky-\frac{y^2}{2}} \sin(2\pi y)\mathrm{d}y \\ &= \frac{\mathrm{e}^{\frac{1}{2}k^2}}{\sqrt{2\pi}} \int_{\mathbf{R}} \mathrm{e}^{-\frac{1}{2}z^2} \sin(2\pi z)\mathrm{d}z \\ &= 0.\end{aligned}$$

3. La méthode des moments présente certains avantages comparé à d'autres classes de fonctions de tests comme les exponentielles complexes (fonction caractéristique). Cette méthode permet notamment de traiter des problèmes avec de la dépendance entre les variables aléatoires, son principale désavantage est qu'elle induit de nombreux calculs. Par exemple elle permet d'établir le théorème d'Erdös-Kac.

Méthode des moments

Lemme. *Sélection de Helly*
Si X_n est une suite de variables aléatoires réelles de fonctions de répartitions F_n vérifiant l'hypothèse de tension :

$$\forall \varepsilon > 0,\ \exists A > 0\ t.q\ \forall n \in \mathbf{N},\ F_n(A) - F_n(-A) \geqslant 1 - \varepsilon.$$

Alors on peut extraire de X_n une sous-suite convergeant en loi.

Proposition. *Critère de Prokhorof*
Soit X_n une suite de variables aléatoires. Alors X_n converge en loi vers une variable aléatoire X si et seulement la suite X_n est tendu et si X_{n_l} converge en loi vers Y alors $\mathscr{L}(X) = \mathscr{L}(Y)$.

Application. *Méthode des moments*
Si μ_n et μ sont des lois de probabilités sur $\mathbf{R}$ avec des moments à tout ordre tels que :

- *$m_k(\mu_n) \to m_k(\mu)$ pour tout $k \in \mathbf{N}$;*
- *μ est caractérisé par ses moments.*

Alors μ_n converge étroitement vers μ.

1. Théorème de sélection de Helly

Soit X_n une suite de variables aléatoires et F_n leurs fonctions de répartitions. On veut extraire de cette suite une sous-suite convergeant en loi, ce qui revient à extraire de F_n une sous-suite convergeant vers une fonction de répartition en tout point de continuité de sa limite.

- Par extraction diagonale on peut extraire de F_n une sous-suite $F_{j(n)}$ convergeant sur une partie dénombrable D. Comme $\mathbf{R}$ est séparable on peut prendre D dense dans $\mathbf{R}$ (par exemple $D = \mathbf{Q}$). Il s'agit maintenant de voir que $F_{j(n)}$ converge partout (enfin presque) mais vers quoi.
- Notons $F(x) = \lim_{n\to+\infty} F_{j(n)}(x)$ puis posons :

$$\overline{F}(x) = \inf_{\substack{x<y \\ y\in D}} F(y).$$

  On adopte des notations distinctes car en général on a $\overline{F}(x) \neq F(x)$ pour $x \in D$. Montrons que si x est un point de continuité de F alors $F_{j(n)}(x)$ converge vers $F(x)$. On a pour $x_1 < x < x_2$ avec $x_1, x_2 \in D$;

$$\begin{array}{ccc} F_{j(n)}(x_1) \leqslant F_{j(n)}(x) & & \leqslant F_{j(n)}(x_2) \\ \downarrow & & \downarrow \\ \overline{F}(x) - \varepsilon \leqslant \overline{F}(\tilde{x}_1) \leqslant F(x_1) & & F(x_2) \leqslant \overline{F}(x_2) \leqslant \overline{F}(x) + \varepsilon \end{array}$$

  pour $\tilde{x}_1 < x_1$ et x_2 choisie assez près de x. On a donc :

$$\overline{F}(x) - \varepsilon \leqslant \liminf F_{j(n)}(x) \leqslant \limsup F_{j(n)}(x) \leqslant \overline{F}(x) + \varepsilon.$$

  Cela étant valable pour tout $\varepsilon > 0$ on en déduit le résultat.
- Finalement étudions les propriétés de $\overline{F}$:
  - $\overline{F}$ est croissante [31]
  - $\overline{F}$ est continue à droite [32]

31. Si $x \leqslant y$ alors $\overline{F}(x) \leqslant F(z)$ pour tout $y < z$ avec $z \in D$ en prenant l'inf sur z on a $\overline{F}(x) \leqslant \overline{F}(y)$.
32. Fixons x et $\varepsilon > 0$ il existe $y > x$ tel que $F(y) - \varepsilon \leqslant \overline{F}(x)$ et alors pour tout $z \in [x, y)$ on a $\overline{F}(z) - \varepsilon \leqslant \overline{F}(x) \leqslant \overline{F}(z)$.

- Pour que $\overline{F}$ soit une fonction de répartition il faut et il suffit que $\lim_{-\infty}\overline{F}=0$ et $\lim_{+\infty}\overline{F}=1$. Pour cela on a besoin de l'hypothèse supplémentaire suivante :

$$\forall \varepsilon>0,\ \exists A>0 \text{ t.q } \forall n\in\mathbf{N},\ F_n(A)-F_n(-A)\geqslant 1-\varepsilon.$$

En passant cette égalité à la limite (on peut être amené à augmenter légèrement A pour éviter une discontinuité) on a :

$$\forall \varepsilon>0,\ \exists A>0 \text{ t.q } \forall n\in\mathbf{N},\ F(-A)\leqslant \varepsilon \text{ et } F(A)\geqslant 1-\varepsilon.$$

2. Critère de Prokhorof

L'implication est relativement direct, on se concentre sur la réciproque. Soit $f\in\mathscr{C}_b(\mathbf{R},\mathbf{R})$ on veut montrer que $\mathbf{E}(f(X_n))$ converge vers $\mathbf{E}(f(X))$. Cette suite est bornée (par $\|f\|_\infty$) pour montrer sa convergence il suffit de montrer qu'elle possède une unique valeur d'adhérence. Si $\mathbf{E}(f(X_{n_l}))$ converge vers v par le théorème de Helly on peut extraire de X_{n_l} une sous-suite $X_{n_{l_k}}$ converge en loi, sa limite est nécessairement X par hypothèse, donc par définition de la convergence en loi $\mathbf{E}(f(X_{n_{l_k}}))\to\mathbf{E}(f(X))$ donc par identification $v=\mathbf{E}(f(X))$ d'où le résultat.

3. Méthode des moments

On applique le critère de Prokhorof :

- $\mu_n([-A,A])\leqslant \frac{1}{A^2}\sup_{n\in\mathbf{N}} m_2(\mu_n)$ (tension)
- si μ_{n_l} converge étroitement vers ν alors il existe $X_{n_l}\sim\mu_{n_l}$ et $X\sim\nu$ tels que $X_{n_l}\to X$ p.s. (Skorokhod) puis comme $X_{n_l}^k$ est équi-intégrable car borné dans L^2 on a par Vitalli : $\mathbf{E}(X_{n_l}^k)\to\mathbf{E}(X^k)$. Ainsi, $m_k(\nu)=\mu_k(\mu)$ pour tout k donc $\mu=\nu$.

Remarque. *1. Le critère de Prokhorof permet de déduire de nombreux autres caractérisations de la convergence en loi, avec d'autres classes de fonctions tests. Notamment il permet de déduire le théorème de Lévy fort : si $\varphi_n=\mathbf{E}(e^{itX_n})$ est une suite de fonctions caractéristiques convergeant simplement vers une fonction caractéristique φ continue en 0 (pour l'hypothèse de tension) alors φ est la fonction caractéristique d'une v.a. X et X_n converge en loi vers X.*

2. La méthode des moments présente certains avantages comparé à d'autres classes de fonctions de tests comme les exponentielles complexes (fonction caractéristique). Cette méthode permet notamment de traiter des problèmes avec de la dépendance entre les variables aléatoires, son principale désavantage est qu'elle induit de nombreux calculs. Par exemple elle permet d'établir le théorème d'Erdös-Kac.