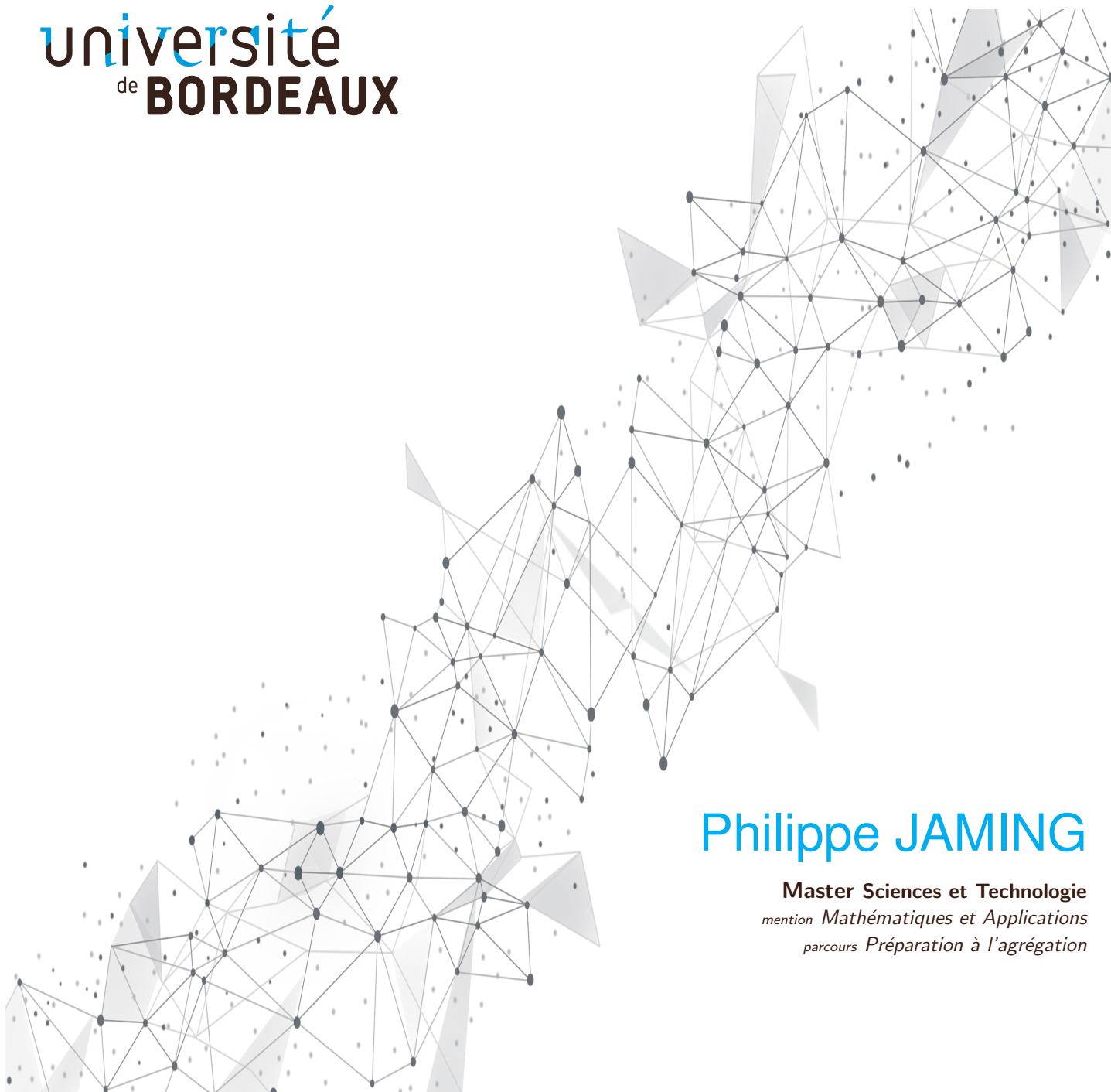




Philippe JAMING

Master Sciences et Technologie
mention Mathématiques et Applications
parcours Préparation à l'agrégation

Cours de M2
Analyse pour l'agrégation

Suites et séries de fonctions.
Séries entières.

Philippe Jaming

Institut de Mathématiques de Bordeaux UMR 5251

351 Cours de la Libération

F 33405 Talence cedex, France

Philippe.Jaming@math.u-bordeaux.fr

<http://www.math.u-bordeaux.fr/~pjaming/>

Table des matières

partie 1. Suites et séries de fonctions	1
Chapitre 1. Convergence des suites de fonctions continues	3
1. Convergence simple et convergence uniforme	3
1.1. Définitions	3
1.2. Convergence simple et convergence uniforme	5
1.3. Convergence des séries de fonctions continues	12
1.3.1. Adaptation des résultats sur les suites aux séries	12
1.3.2. L'équation de Cauchy et sa stabilité	14
1.3.3. La fonction de Weierstrass : une fonction continue nulle-part dérivable	16
1.3.4. Une seconde fonction continue nulle-part dérivable	19
1.3.5. L'escalier de Cantor	20
2. Compacité dans $\mathcal{C}(K)$ et applications	21
2.1. Théorème de Helly	21
2.2. Théorème d'Ascoli-Arzelà	23
2.3. Lemme de Dini	25
2.4. Applications en probabilités : convergence des fonctions de répartition	26
2.5. Applications en probabilités : distance de Wasserstein	29
2.6. Application en géométrie : existence de géodésiques	32
2.7. Application en théorie des nombres	34
3. Densité dans $\mathcal{C}(K)$, K compact	37
3.1. Première méthode : par approximation de l'unité	37
3.2. Seconde méthode : les polynômes de Bernstein	39
3.3. Troisième méthode : le théorème de Stone-Weierstrass	47
3.4. Interpolation de Lagrange et phénomène de Runge	50
3.5. Le théorème des moments de Hausdorff	52
Index	57

Première partie

Suites et séries de fonctions

Convergence des suites de fonctions continues

1. Convergence simple et convergence uniforme

1.1. Définitions. Avant de nous consacrer plus spécifiquement aux suites de fonctions, commençons par les définitions de deux modes de convergence :

Definition 1.1

Soit X un ensemble et (Y, d_Y) un espace métrique. Pour chaque $n \geq 0$, on considère une fonction $f_n : X \rightarrow Y$ de sorte que $(f_n)_{n \geq 0}$ soit une suite de fonctions sur X à valeurs dans Y . Soit $f : X \rightarrow Y$ une fonction.

On dit que la suite $(f_n)_{n \geq 0}$ *converge simplement vers f sur X* si, pour tout $x \in X$, $f_n(x) \rightarrow f(x)$, c'est-à-dire si, pour tout $x \in X$, pour tout $\varepsilon > 0$, il existe $N = N(x, \varepsilon)$ tel que, si $n \geq N$, alors $d_Y(f_n(x), f(x)) < \varepsilon$.

On dit que la suite $(f_n)_{n \geq 0}$ *converge uniformément vers f sur X* si $\sup_{x \in X} d_Y(f_n(x), f(x)) \rightarrow 0$, c'est-à-dire si, pour tout $\varepsilon > 0$, il existe $N = N(\varepsilon)$ tel que, si $n \geq N$, alors pour tout $x \in X$, $d_Y(f_n(x), f(x)) < \varepsilon$.

Évidemment, la convergence uniforme sur X implique la convergence simple. De plus, si $X' \subset X$, la convergence simple (resp. uniforme) sur X implique la convergence simple (resp. uniforme) sur X' alors que la réciproque est évidemment fausse.

Il est essentiel de toujours préciser les ensembles sur lesquels ces convergences ont lieu.

Lorsque cela n'est pas fait, il est implicite que ces convergences ont lieu sur l'ensemble sur lequel les fonctions sont définies.

Pour l'instant, nous n'avons besoin que de peu de propriétés sur X et sur Y :

- une distance sur Y pour pouvoir définir la convergence dans Y ,
- X est simplement un ensemble.

Par exemple

– si $X = \mathbb{N}$, une fonction $f_n : X \rightarrow Y$ est une suite $(f_n(m))_{m \geq 0}$ de Y et la convergence simple est la convergence terme à terme : $f_n(m) \rightarrow f(m)$ pour tout m . Lorsque la suite f_n est bornée, f aussi et la convergence uniforme est la convergence dans $\ell^\infty(\mathbb{N}, Y)$.

– Si $X \subset \mathcal{B}$ avec $(\Omega, \mathcal{B}, \mu)$ un espace mesuré et $\mu(\Omega \setminus X) = 0$ alors la convergence simple sur X est la convergence μ -presque partout. La convergence dans $L^\infty(\Omega, \mathcal{B}, \mu)$ est alors la convergence uniforme sur une partie $X \subset \Omega$ avec $\mu(\Omega \setminus X) = 0$.

Dans la suite, nous allons nous intéresser à des suites de fonctions continues. Pour cela, X doit également être un espace métrique. Afin d'éviter des technicités inutiles au niveau de l'agrégation, nous allons maintenant nous focaliser sur le cas suivant

- X est soit un compact $X = K$ soit $X = \mathbb{R}^d$,
- Y est un espace vectoriel de dimension finie. En choisissant une base, tous nos résultats peuvent se réduire à $Y = \mathbb{R}$ ou $Y = \mathbb{C}$. Sauf mention contraire, nous allons toujours considérer $Y = \mathbb{R}$.

Nous allons en outre introduire les espaces suivants

- K un compact, $\mathcal{C}(K) = \mathcal{C}(K, \mathbb{R}) = \{f : K \rightarrow \mathbb{R}, \text{ continue}\}$; et $\mathcal{C}(K, \mathbb{C}) = \{f : K \rightarrow \mathbb{C} \text{ continue}\}$;

- $\mathcal{C}_b(\mathbb{R}^d) = \mathcal{C}_b(\mathbb{R}^d, \mathbb{R}) = \{f : \mathbb{R}^d \rightarrow \mathbb{R}, \text{ continue et bornée}\};$
- $\mathcal{C}_0(\mathbb{R}^d) = \mathcal{C}_0(\mathbb{R}^d, \mathbb{R}) = \{f : \mathbb{R}^d \rightarrow \mathbb{R}, \text{ continue et t.q. } \lim_{|x| \rightarrow +\infty} f(x) = 0\}.$

Lorsque nous voudrions considérer des fonctions à valeurs complexes, nous noterons les espaces correspondants $\mathcal{C}(K, \mathbb{C}), \mathcal{C}_b(K, \mathbb{C}), \mathcal{C}_0(\mathbb{R}^d, \mathbb{C})$. On a bien sûr $\mathcal{C}_0(\mathbb{R}^d) \subset \mathcal{C}_b(\mathbb{R}^d)$. Comme une combinaison linéaire de fonctions continues (resp. bornées ou tendant vers 0 à l'infini) est continue (resp. bornée ou tendant vers 0 à l'infini), ces espaces sont des espaces vectoriels.

Les fonctions de ces espaces sont de plus *bornées*.

Lemme 1.2

Soit K un compact, pour $f \in \mathcal{C}(K)$ définissons

$$\|f\|_\infty = \sup_{x \in K} |f(x)|.$$

Alors $(\mathcal{C}(K), \|\cdot\|_\infty)$ est un espace normé.

De même, $\|f\|_\infty = \sup_{x \in \mathbb{R}^d} |f(x)|$ est une norme sur $\mathcal{C}_0(\mathbb{R}^d)$ et $\mathcal{C}_b(\mathbb{R}^d)$.

DÉMONSTRATION. Tout d'abord, $\|f\|_\infty \geq 0$ et si $\|f\|_\infty = 0$ alors pour tout $x \in K$ $|f(x)| \leq \|f\|_\infty = 0$ donc $f = 0$.

Ensuite, si $f, g \in \mathcal{C}(K)$ et $x \in K$ alors $|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty$. En oubliant la première inégalité et en passant au sup sur $x \in K$ à gauche, on obtient $\|f + g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$.

Si $\lambda \in \mathbb{R}$ et $f \in \mathcal{C}(I)$ alors $|\lambda f(x)| = |\lambda| |f(x)| \leq |\lambda| \|f\|_\infty$ et, en passant au sup à gauche, $\|\lambda f\|_\infty \leq |\lambda| \|f\|_\infty$. Ensuite, si $f = 0$, $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty = 0$. Enfin, si $f \neq 0$, on prend $\varepsilon > 0$ et x_0 tel que $|f(x_0)| \geq (1 - \varepsilon) \|f\|_\infty$ alors $|\lambda f(x_0)| \geq (1 - \varepsilon) |\lambda| \|f\|_\infty$ donc $\|\lambda f\|_\infty \geq (1 - \varepsilon) |\lambda| \|f\|_\infty$. Comme ε est arbitraire, $\|\lambda f\|_\infty \geq |\lambda| \|f\|_\infty$.

Les cas $\mathcal{C}_0(\mathbb{R}^d)$ et $\mathcal{C}_b(\mathbb{R}^d)$ sont identiques. ■

Remarque 1.3

Dans la suite, sauf précision contraire, ces espaces seront munis de cette norme. L'observation clé est la suivante :

si $f_n, f \in \mathcal{C}(K)$ alors $f_n \rightarrow f$ uniformément sur K signifie précisément que $f_n \rightarrow f$ dans $\mathcal{C}(K)$.

De même, si $f_n, f \in \mathcal{C}_b(\mathbb{R}^d)$ (ou $\mathcal{C}_0(\mathbb{R}^d)$) alors $f_n \rightarrow f$ uniformément sur $\mathbb{R}^d$ signifie précisément que $f_n \rightarrow f$ dans $\mathcal{C}_b(\mathbb{R}^d)$.

Terminons par deux exemples. Le premier montre que convergence simple et uniforme diffèrent :

Exemple 1.4

Considérons $f, f_n : [0, 1] \rightarrow \mathbb{C}$ défini par $f_n(x) = x^n$ et $f(x) := \begin{cases} 0 & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x = 1 \end{cases}$.

Alors, pour tout $x \in [0, 1]$ $f_n(x) \rightarrow f(x)$, i.e. $f_n \rightarrow f$ simplement sur $[0, 1]$. Par ailleurs

$$\sup_{x \in [0, 1]} |f_n(x) - f(x)| = 1$$

donc f_n ne converge pas uniformément vers f sur $[0, 1]$.

Par contre, si $0 \leq a < 1$, alors $\sup_{x \in [0, a]} |f_n(x) - f(x)| = a^n \rightarrow 0$ donc $f_n \rightarrow f$ uniformément sur $[0, a]$.

Notez que dans cet exemple, bien que chaque f_n soit continue, sa limite simple sur $[0, 1]$ ne l'est pas alors que sa limite uniforme sur $[0, a]$ l'est. Cet exemple sera donc également un exemple pour les théorèmes de ci-dessous.

Exemple 1.5

Soit f une fonction continue sur $[0, 1]$. Alors f est uniformément continue sur $[0, 1]$. Pour tout $n \geq 1$, il existe alors un entier $k_n > 0$ tel que, si $|x - y| \leq \frac{1}{k_n}$ alors $|f(x) - f(y)| < \frac{1}{n}$. Pour $k = 0, \dots, k_n - 1$, on choisit alors $\xi_{k,n} \in \left[\frac{k}{k_n}, \frac{k+1}{k_n}\right]$. On définit enfin une fonction $f_n : [0, 1] \rightarrow \mathbb{C}$ par $f_n(x) = f(\xi_{k,n})$ pour $x \in [k/k_n, (k+1)/k_n]$, $k = 0, \dots, k_n - 1$ et $f_n(1) = f(1)$. On a alors, pour tout $x \in [0, 1]$, $|f(x) - f_n(x)| < \frac{1}{n}$ donc $f_n \rightarrow f$ uniformément sur $[0, 1]$.

Bien sûr, f_n n'est en général pas continue, il ne s'agit donc pas d'une convergence dans $\mathcal{C}([0, 1])$. Cette construction est à la base de l'intégration de Riemann.

Notons enfin que la convergence uniforme sur $\mathbb{R}^d$ pour des fonctions non bornées est difficile à obtenir (et souvent inutile) :

Exemple 1.6

Soit P_n une suite de polynômes et supposons que P_n converge uniformément sur $\mathbb{R}$, alors il existe N et un polynôme P tel que, pour $n \geq N$, $P - P_n = \alpha_n$, un polynôme constant et de plus $\alpha_n \rightarrow \alpha$.

En effet, si (P_n) converge uniformément sur $\mathbb{R}$, elle est de Cauchy, en particulier, il existe N tel que, pour $n \geq N$, $\sup_{x \in \mathbb{R}} |P_n(x) - P_N(x)| \leq 1$. Mais alors $P_n - P_N$ est un polynôme borné, donc constant. On pose alors $P = P_N$ et $\alpha_n = P_n - P$. Comme P_n est de Cauchy, $P_n - P$ aussi donc α_n est de Cauchy dans $\mathbb{R}$ donc converge.

1.2. Convergence simple et convergence uniforme. Nous allons maintenant nous intéresser au passage à la limite de la régularité des (f_n) . L'exemple 1.4 montre que la limite simple ne suffit pas.

Théorème 1.7

Soit K un compact. Soit $(f_n)_{n \geq 0} \subset \mathcal{C}(K)$ (resp. $\mathcal{C}_b(K)$) une suite de fonctions continues $K \rightarrow \mathbb{R}$ et $f : K \rightarrow \mathbb{R}$. Si (f_n) converge uniformément vers f sur K alors f est continue sur K i.e. $f \in \mathcal{C}(K)$.

De même, si $(f_n)_{n \geq 0} \subset \mathcal{C}_b(K)$ (resp. $\mathcal{C}_0(\mathbb{R}^d)$) converge uniformément vers f sur $\mathbb{R}^d$ alors $f \in \mathcal{C}_b(\mathbb{R}^d)$ (resp. $f \in \mathcal{C}_0(\mathbb{R}^d)$). En particulier, $\mathcal{C}_0(\mathbb{R}^d)$ est un sous-espace fermé de $\mathcal{C}_b(\mathbb{R}^d)$.

DÉMONSTRATION. En effet, soit $\varepsilon > 0$ et $x_0 \in D$. Il existe N tel que, pour tout $n \geq N$ et tout $x \in K$, $|f_n(x) - f(x)| \leq \varepsilon$. Comme f_n est continue en x_0 , il existe un voisinage V de x_0 dans K tel que si $x \in V$, alors $|f_N(x) - f_N(x_0)| < \varepsilon$. Mais alors

$$|f(x) - f(x_0)| \leq |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)| \leq 3\varepsilon$$

et f est bien continue en x_0 .

Supposons maintenant que $f_n \in \mathcal{C}_b(\mathbb{R}^d)$ converge uniformément vers f sur $\mathbb{R}^d$ alors elle converge uniformément vers f sur tout compact K donc, avec la première partie, f est continue sur K . Comme K est arbitraire, f est continue sur $\mathbb{R}^d$. De plus, en prenant $\varepsilon = 1$ dans la définition de la convergence uniforme, il existe N tel que $\|f - f_N\|_\infty \leq 1$ et

donc, avec l'inégalité triangulaire, $\|f\|_\infty \leq \|f - f_N\|_\infty + \|f_N\|_\infty \leq 1 + \|f_N\|_\infty$ donc f est bien bornée.

Enfin, si $(f_n)_{n \geq 0} \subset \mathcal{C}_0(\mathbb{R}^d) \subset \mathcal{C}_b(\mathbb{R}^d)$ converge uniformément vers f alors f est continue sur $\mathbb{R}^d$ avec la partie précédente. De plus, si $\varepsilon > 0$, il existe N tel que $\|f - f_N\|_\infty \leq \varepsilon$. Comme $f_N \in \mathcal{C}_0(\mathbb{R}^d)$, il existe $R > 0$ tel que, si $\|x\| \geq R$, alors $|f_N(x)| \leq \varepsilon$. Mais alors, pour $\|x\| \geq R$, $|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| \leq \|f - f_N\|_\infty + |f_N(x)| \leq 2\varepsilon$. Ainsi $f \in \mathcal{C}_0(\mathbb{R}^d)$.

Cela montre bien que $\mathcal{C}_0(\mathbb{R}^d)$ est fermé dans $\mathcal{C}_b(\mathbb{R}^d)$: si une suite de $\mathcal{C}_0(\mathbb{R}^d)$ converge dans $\mathcal{C}_b(\mathbb{R}^d)$ alors sa limite est dans $\mathcal{C}_0(\mathbb{R}^d)$. ■

Notons que la même démonstration s'applique pour les fonctions définies sur un espace métrique (pour pouvoir définir la continuité) et à valeurs dans un espace métrique (puisque seule l'inégalité triangulaire a été utilisée).

Théorème 1.8

Soit K un compact. Les espaces $\mathcal{C}(K), \mathcal{C}_b(\mathbb{R}^d), \mathcal{C}_0(\mathbb{R}^d)$ sont des espaces de Banach. En particulier, une suite $(f_n) \in \mathcal{C}(K)$ converge uniformément sur K si et seulement si elle est de Cauchy pour $\|\cdot\|_\infty$ c'est-à-dire si et seulement si, pour tout $\varepsilon > 0$, il existe N tel que, si $m, n \geq N$, alors pour tous $x \in K$, $|f_n(x) - f_m(x)| < \varepsilon$.

DÉMONSTRATION. Soit $(f_n)_{n \geq 0} \subset \mathcal{C}(K)$ une suite de Cauchy. Tout d'abord, pour tout $\varepsilon > 0$, il existe N_ε tel que, si $n, m \geq N_\varepsilon$, alors, pour tout $x \in K$, $|f_n(x) - f_m(x)| \leq \varepsilon$. En particulier, pour chaque $x \in K$, $(f_n(x))_{n \geq 0}$ est une suite de Cauchy dans $\mathbb{R}$. Comme $\mathbb{R}$ est complet, cette suite converge et on note $f(x)$ sa limite. Ainsi (f_n) converge *simplement* vers f sur K .

Mais comme, pour tout $n \geq N_\varepsilon$ et tout $x \in K$, pour tout $m \geq N_\varepsilon$, $|f_n(x) - f_m(x)| \leq \varepsilon$, en passant à la limite quand $m \rightarrow +\infty$, on obtient $|f_n(x) - f(x)| \leq \varepsilon$. Mais ceci est précisément la définition de (f_n) converge *uniformément* vers f sur K , ou encore que $\|f_n - f\|_\infty \rightarrow 0$. Enfin, le théorème précédent nous dit que $f \in \mathcal{C}(K)$. On a donc bien montré qu'il existe $f \in \mathcal{C}(K)$ tel que $f_n \rightarrow f$ pour la norme $\|\cdot\|_\infty$ et $\mathcal{C}(K)$ est donc complet.

La démonstration est identique pour $\mathcal{C}_b(\mathbb{R}^d)$ et pour $\mathcal{C}_0(\mathbb{R}^d)$. ■

Nous allons maintenant munir K ou $\mathbb{R}^d$ d'une mesure μ et regarder si

$$\lim \int f_n(x) d\mu(x) = \int \lim f_n(x) d\mu(x).$$

Le théorème de convergence dominée fournit évidemment le critère le plus efficace pour établir un tel résultat, mais il requière des intégrales absolument convergentes. De plus, il nécessite une théorie de l'intégration assez évoluée. Nous allons ici nous intéresser à ce qui peut être obtenu uniquement avec l'intégration de Riemann.

Tout d'abord, voici deux exemples qui doivent nous appeler à la prudence :

Exemple 1.9

Soit $(f_n)_{n \geq 2}$ la suite de fonctions continues et affines par morceaux sur $[0, 1]$ définie par $f_n(x) = n^2x$ sur $[0, 1/n]$, $f_n(x) = 2n - n^2x$ sur $[1/n, 2/n]$ et $f_n(x) = 0$ sur $[2/n, 1]$. Alors $f_n(x) \rightarrow 0$ simplement sur $[0, 1]$ mais

$$\int_0^1 f_n(x) dx = 1 \neq 0 = \int_0^1 f(x) dx.$$

Exemple 1.10

- (1) Soit $(f_n)_{n \geq 1}$ la suite de fonctions continues sur $[0, +\infty)$ définies par $f_n(x) = \frac{1}{n} \left(1 - \frac{x}{n}\right)$ si $x < n$ et $f_n(x) = 0$ si $x > n$. Alors $\|f_n\|_\infty = \frac{1}{n} \rightarrow 0$ i.e. (f_n) tend uniformément vers 0 alors que

$$\int_0^{+\infty} f_n(x) dx = \frac{1}{2} \rightarrow \frac{1}{2} \neq 0 = \int_0^{+\infty} f(x) dx.$$

- (2) Soit $(f_n)_{n \geq 1}$ la suite de fonctions définies sur $[0, +\infty[$ par $f_n(x) = \frac{1}{n} e^{-x/n}$; alors $|f_n(x)| \leq \frac{1}{n} \rightarrow 0$ donc $f_n \rightarrow 0$ uniformément sur $[0, +\infty[$ mais

$$\int_0^{+\infty} \frac{1}{n} e^{-x/n} dx = \int_0^{+\infty} e^{-t} dt = 1 \neq \int_0^{+\infty} 0 dx.$$

- (3) Même si $\int_0^{+\infty} f(x) dx = +\infty$, ce n'est pas forcément le cas, mais pour cela, il faut une suite qui change de signe. Par exemple, on prend $f(x) = \frac{1}{1+x}$ sur $[0, +\infty)$ et f_n une fonction continue telle que

$$- f_n(x) = f(x) \text{ sur } [0, n] \text{ de sorte que } f_n(n) = \frac{1}{1+n},$$

$$- \text{ affine sur } [n, n+1] \text{ telle que } f_n(n+1) = -\frac{1}{1+n}, \text{ c'est-à-dire } f_n(x) = \frac{1-2(x-n)}{n+1} \text{ et ainsi } \int_n^{n+1} f_n(x) dx = 0,$$

$$- \text{ on choisit } a = 2(n+1) \ln(1+n) \text{ de sorte que le triangle de hauteur } \frac{1}{n+1} \text{ de base } a \text{ ait pour aire } \int_0^n f_n(x) dx = \ln(1+n) \text{ et on prend } f_n \text{ affine sur}$$

$$[n+1, n+1+a] \text{ telle que } f_n(n+1) = -\frac{1}{n+1} \text{ et } f_n(n+1+a) = 0 \text{ donc } f_n(x) = \frac{1}{a(n+1)}(x - a - n - 1),$$

$$- \text{ et enfin } f_n(x) = 0 \text{ pour } x > n+1+a.$$

Très clairement $\|f_n - f\|_\infty \leq \sup_{x \geq n} |f(x)| + \sup_{x \geq n} |f_n(x)| \leq \frac{2}{n+1}$ donc $f_n \rightarrow f$ uniformément mais

$$\lim \int_0^{+\infty} f_n(x) dx = 0 \neq +\infty = \int_0^{+\infty} f(x) dx.$$

Théorème 1.11 (Intégration des suites de fonctions)

Soit K un compact de $\mathbb{R}^d$, alors la forme linéaire $I : \mathcal{C}(K) \rightarrow \mathbb{R}$ définie par

$$I(f) = \int_K f(x) dx$$

est continue. En particulier, si $(f_n) \subset \mathcal{C}(K)$ et $f_n \rightarrow f$ uniformément sur K alors

$$\int_K f_n(x) dx \rightarrow \int_K f(x) dx.$$

Notez que $\int_K f(x) dx$ est a priori une intégrale de Lebesgue, mais si $\mathbb{1}_K$ est Riemann intégrable, alors c'est une intégrale de Riemann. On peut bien sûr se restreindre à des géométries sympatiques, par exemple des convexes.

DÉMONSTRATION. I est bien définie, linéaire et continue pour $\|\cdot\|_\infty$ puisque

$$|I(f)| = \left| \int_K f(x) dx \right| \leq \int_K |f(x)| dx \leq \left(\int_K 1 dx \right) \|f\|_\infty.$$

Par continuité, si $f_n \rightarrow f$ alors $I(f_n) \rightarrow I(f)$. ■

Exemple 1.12

Considérons $I_n = \int_0^1 (1+x^n)^{1/n} dx$. Notez qu'on ne peut pas calculer cette intégrale directement, et que le changement de variable $t = x^n$ puis $u = \frac{t}{1+t}$ nous donne

$$I_n = \frac{1}{n} \int_0^1 (1+t)^{\frac{1}{n}} t^{\frac{1}{n}-1} dt = \frac{1}{n} \int_0^{1/2} u^{\frac{1}{n}-1} (1-u)^{-1-\frac{2}{n}} du \frac{1}{n} B\left(\frac{1}{n}, -\frac{2}{n}\right)$$

où $B_x(\alpha, \beta) = \int_0^x u^{\alpha-1} (1-u)^{\beta-1} du$ est la fonction bêta incomplète. On peut également montrer que $I_n = {}_2F_1\left(-\frac{1}{n}, \frac{1}{n}, 1 + \frac{1}{n}; -1\right)$. Toutefois, ces expressions ne permettent pas de donner la limite de I_n aisément.

Posons alors $f_n(x) = (1+x^n)^{1/n}$ et observons que, pour $x \in [0, 1]$, $1 \leq f_n(x) \leq 2^{1/n}$ ou encore $\sup_{x \in [0,1]} |f_n(x) - 1| \leq 2^{1/n} - 1 \rightarrow 0$. Ainsi, la suite (f_n) converge uniformément vers la fonction constante égale à 1 sur $[0, 1]$.

Le théorème de passage à la limite sous l'intégrale donne alors

$$\lim_{n \rightarrow \infty} I_n = \int_0^1 \lim_{n \rightarrow \infty} (1+x^n)^{1/n} dx = \int_0^1 1 dx = 1.$$

Par conséquent, $\lim_{n \rightarrow \infty} \int_0^1 (1+x^n)^{1/n} dx = 1$.

La dernière partie s'écrit

$$\lim \int_K f_n(x) dx = \int_K \lim f_n(x) dx$$

où la limite à l'extérieur de l'intégrale est une limite d'une suite réelle et celle à l'intérieur de l'intégrale est une limite uniforme. Ce résultat peut être obtenu avec le théorème de convergence dominée : la suite f_n converge dans $\mathcal{C}(K)$ donc simplement et de plus, cette suite est bornée par une constante M et M est intégrable sur K .

Remarque 1.13

L'importance de ce résultat est que si K a une géométrie un peu régulière, par exemple, un intervalle, un disque, un polygone régulier... alors $\int_K f(x) dx$ est défini comme une intégrale de Riemann. Par rapport au théorème de convergence dominée, on a échangé la construction d'un majorant dominant la suite par une convergence uniforme.

La clé ici est que K muni de la mesure de Lebesgue est de mesure finie, on a donc le même raisonnement en remplaçant K par $\mathbb{R}^d$ (ou certains espaces plus généraux) à condition de remplacer la mesure de Lebesgue par une mesure finie.

Notons enfin que dans notre contre-exemple ci-dessus, la suite f_n prenait des valeurs positives et négatives. Si $f_n \geq 0$ converge uniformément vers f sur $[0, +\infty)$ et si $\int_0^{+\infty} f(x) dx = +\infty$ alors $\int_0^{+\infty} f_n(x) dx \rightarrow +\infty$.

En effet, pour tout $M > 0$, il existe A, B tel que $\int_A^B f(x) dx \geq M$ et $\int_A^B f_n(x) dx \rightarrow \int_A^B f(x) dx$ donc, pour n assez grand

$$\int_0^{+\infty} f_n(x) dx \geq \int_A^B f_n(x) dx \geq \frac{M}{2}$$

d'où le résultat.

Pour les deux résultats suivants, nous allons nous restreindre à $K = [a, b]$ un intervalle fermé borné de $\mathbb{R}$ ($-\infty < a < b < +\infty$) :

Corollaire 1.14 (Théorème des primitives des suites de fonctions)

Soit $I = [a, b] \subset \mathbb{R}$ un intervalle borné et $x_0 \in I$. Alors l'application $\Pi : \mathcal{C}(I) \rightarrow \mathcal{C}(I)$ définie par

$$\Pi[f](x) = \int_{x_0}^x f(t) dt, \quad x \in [a, b]$$

est continue.

En particulier, si

- $(f_n)_{n \geq 0} \subset \mathcal{C}(I)$ converge uniformément sur I vers une fonction $f : I \rightarrow \mathbb{R}$
- $(\alpha_n)_{n \geq 0}$ une suite qui converge vers $\alpha \in \mathbb{R}$
- et si F_n, F sont définies sur $[a, b]$ par

$$F_n(x) = \alpha_n + \int_{x_0}^x f_n(t) dt \quad \text{et} \quad F(x) = \alpha + \int_{x_0}^x f(t) dt$$

alors $F_n \rightarrow F$ uniformément sur I .

DÉMONSTRATION. Tout d'abord, $\Pi[f]$ est bien définie, est continue et $f \rightarrow \Pi[f]$ est linéaire. Comme

$$\left| \int_{x_0}^x f(t) dt \right| \leq \int_{[x_0, x]} |f(t)| dt \leq \int_a^b |f(t)| dt \leq (b-a) \|f\|_\infty,$$

Π est continue.

Mais alors $\|\Pi[f_n] - \Pi[f]\|_\infty \rightarrow 0$ puisque $f_n \rightarrow f$ uniformément sur $[a, b]$. Comme $F_n(x) = \alpha_n + \Pi[f_n](x)$ et $F(x) = \alpha + \Pi[f](x)$

$$|F_n(x) - F(x)| \leq |\alpha_n - \alpha| + |\Pi[f_n](x) - \Pi[f](x)| \leq |\alpha_n - \alpha| + \|\Pi[f_n] - \Pi[f]\|_\infty$$

donc $\|F_n - F\|_\infty \rightarrow 0$ i.e. $F_n \rightarrow F$ uniformément sur $[a, b]$. ■

Remarque 1.15

Dans les hypothèses du théorème, $\alpha_n = F_n(x_0)$ et $\alpha = F(x_0)$, on suppose donc $F_n(x) \rightarrow F(x)$ pour le seul x_0 . Cela est utile quand on veut définir F par ce procédé. Lorsque cela est possible, il est en général préférable d'identifier F , i.e. de montrer que $F_n \rightarrow F$ simplement sur I .

On peut étendre ce résultat sans difficulté à des intégrales le long de segments en dimension plus grande.

On peut maintenant reformuler ce corollaire en terme de dérivation :

Corollaire 1.16 (Théorème de dérivation des suites de fonctions)

Soit $I \subset \mathbb{R}$ un intervalle. Soit $(F_n)_{n \geq 0}$ une suite de fonctions $I \rightarrow \mathbb{C}$ de classe $\mathcal{C}^1$.

On suppose que

– la suite des dérivées $(F'_n)_{n \geq 0}$ converge uniformément sur tout intervalle compact $J \subset I$ vers une fonction $\varphi : I \rightarrow \mathbb{C}$.

– il existe $x_0 \in I$ tel que $F_n(x_0)$ a une limite.

Alors il existe une fonction F de classe $\mathcal{C}^1$ sur I telle que $(F_n)_{n \geq 0}$ converge uniformément vers F sur tout compact $J \subset I$. De plus $F' = \varphi$.

DÉMONSTRATION. En effet, il suffit de poser $a = \lim F_n(x_0)$ et $F(x) = a + \int_{x_0}^x \varphi(t) dt$ et d'appliquer le corollaire précédent à $f_n = F'_n$ et $f = \varphi$. ■

On peut étendre ce résultat aux fonctions de plusieurs variables.

Exemple 1.17

Soit $(f_n)_{n \geq 1}$ la suite de fonctions définies sur $[0, 1]$ par $f_n(x) = \sqrt{x + \frac{1}{n}}$ et soit f défini par $f(x) = \sqrt{x}$ sur $[0, 1]$. Alors $f_n \rightarrow f$ uniformément sur $[0, 1]$ puisque

$$\begin{aligned} 0 \leq f_n(x) - f(x) &= \sqrt{x + \frac{1}{n}} - \sqrt{x} = \frac{(\sqrt{x + \frac{1}{n}} - \sqrt{x})(\sqrt{x + \frac{1}{n}} + \sqrt{x})}{\sqrt{x + \frac{1}{n}} + \sqrt{x}} \\ &= \frac{1}{n} \frac{1}{\sqrt{x + \frac{1}{n}} + \sqrt{x}} \leq \frac{1}{n} \frac{1}{\sqrt{\frac{1}{n}}} = \frac{1}{\sqrt{n}} \rightarrow 0. \end{aligned}$$

Mais f_n est dérivable alors que f ne l'est pas. Notez qu'elle n'est même pas lipschitzienne.

Exemple 1.18

On a $\left(1 + \frac{x}{n}\right)^n \rightarrow e^x$ uniformément sur tout intervalle compact.

Posons $F_n(x) = n \ln \left(1 + \frac{x}{n}\right)$ et $F(x) = x$. L'équivalent classique $\ln(1+t) \sim t$ en 0 nous donne immédiatement que $F_n \rightarrow F$ simplement sur $\mathbb{R}$. De plus

$$F'_n(x) = \frac{1}{1 + \frac{x}{n}} \rightarrow 1.$$

Cette convergence est uniforme sur tout compact. Il suffit de considérer $[-A, A]$ (un compact étant borné). Alors, si $n \geq 2A$, $\left|1 + \frac{x}{n}\right| \geq 1 - \frac{A}{n} \geq \frac{1}{2}$ et donc

$$|F'_n(x) - 1| = \frac{\frac{|x|}{n}}{\left|1 + \frac{x}{n}\right|} \leq \frac{2A}{n} \rightarrow 0$$

D'après le théorème de dérivation, on a donc $F_n \rightarrow F$ uniformément sur $[-A, A]$. Enfin, toujours en fixant A , on en déduit que la suite F_n est bornée sur $[-A, A]$ puisque convergence dans $\mathcal{C}([-A, A])$. Il existe donc $B \geq A$ tel que $|F_n(x)| \leq B$ quand $x \in [-A, A]$. Mais la fonction exponentielle est continue sur $[-B, B]$ donc uniformément continue. Soit alors $\varepsilon > 0$, il existe $\eta > 0$ tel que, si $u, v \in [-B, B]$ sont tels que $|u - v| \leq \eta$ alors $|e^u - e^v| \leq \varepsilon$. De plus, comme $F_n(x) \rightarrow x$ uniformément sur $[-A, A]$, il existe N tel que, si $n \geq N$, $|F_n(x) - x| \leq \eta$. En prenant $u = F_n(x)$ et $v = x$, on en déduit donc que $|e^{F_n(x)} - e^x| \leq \varepsilon$. Mais cela nous dit précisément que, si $n \geq N$ et $x \in [-A, A]$, alors

$$\left| \left(1 + \frac{x}{n}\right)^n - e^x \right| \leq \varepsilon.$$

C'est la convergence uniforme cherchée.

La dernière partie de l'exemple s'adapte pour montrer que si (f_n) est une suite de fonctions continues sur $[-A, A]$ telle que $f_n \rightarrow f$ uniformément sur $[-A, A]$ et si φ est une fonction continue sur $\overline{\bigcup f_n([-A, A])}$ alors $\varphi(f_n) \rightarrow \varphi(f)$ uniformément sur $[-A, A]$.

Ce résultat se reformule comme suit

Théorème 1.19

Soit $I \subset \mathbb{R}$ un intervalle fermé borné et $\mathcal{C}^1(I)$ l'espace vectoriel des fonctions de classe $\mathcal{C}^1$ sur I qu'on munit de la norme

$$\|f\|_{\mathcal{C}^1} = \|f\|_{\infty} + \|f'\|_{\infty}.$$

Alors $\mathcal{C}^1$ muni de cette norme est complet.

De plus, si $x_0 \in I$, alors

$$\|f\| = |f(x_0)| + \|f'\|_{\infty}$$

définit une norme équivalente à $\|\cdot\|_{\mathcal{C}^1}$.

DÉMONSTRATION. Tout d'abord, comme $f \mapsto f'$ est linéaire, il est clair que $\|\cdot\|_{\mathcal{C}^1}$ est bien une norme sur $\mathcal{C}^1(I)$ et de plus $\|f\| \leq \|f\|_{\mathcal{C}^1}$.

Si $f \in \mathcal{C}^1$, elle est continue sur $I = [a, b]$ compact donc $|f|$ aussi et atteint donc son maximum en un point x_m . On écrit alors

$$\begin{aligned} \|f\|_{\infty} &= |f(x_m)| = \left| f(x_0) + \int_{x_0}^{x_m} f'(t) dt \right| \\ &\leq |f(x_0)| + \int_{[x_0, x_m]} |f'(t)| dt \\ &\leq |f(x_0)| + |x_m - x_0| \|f'\|_{\infty} \leq |f(x_0)| + (b - a) \|f'\|_{\infty} \end{aligned}$$

et au final $\|f\|_{\mathcal{C}^1} \leq \|f\|_{\infty} + \|f'\|_{\infty} \leq (1 + (b - a)) \|f\|$.

Soit $(f_k)_{k \geq 0} \subset \mathcal{C}^1(I)$ une suite de Cauchy pour la norme $\|f\|$. Alors (f'_k) est de Cauchy dans $\mathcal{C}(I)$, elle converge donc vers une fonction continue g . De plus, $(f_k(x_0))$ est de Cauchy dans $\mathbb{R}$ donc converge vers un nombre α . Enfin

$$f_k(x) = f_k(x_0) + \int_{x_0}^x f'_k(t) dt \rightarrow f(x) := \alpha + \int_{x_0}^x g(t) dt.$$

Ainsi $\alpha = f(x_0)$, f est de classe $\mathcal{C}^1$ et $f' = g$. Enfin, $f_k \rightarrow f$ converge simplement. En considérant la norme $\|\cdot\|_{\mathcal{C}^1} \geq \|\cdot\|_{\infty}$, f_k converge uniformément, par unicité de la limite $f_k \rightarrow f$ uniformément et $f'_k \rightarrow g = f'$ uniformément, on a donc bien $f_k \rightarrow f$ en norme $\|\cdot\|_{\mathcal{C}^1}$. ■

1.3. Convergence des séries de fonctions continues.

1.3.1. *Adaptation des résultats sur les suites aux séries.*

Tous les résultats sur les suites de fonctions s'adaptent immédiatement aux séries de fonctions. Pour faire un peu plus court, nous allons nous contenter du cas $\mathcal{C}(K)$. Les adaptations à $\mathcal{C}_b(\mathbb{R}^d)$ sont immédiates.

Definition 1.20

Soit K un compact. soit $(f_k)_{k \geq 0} \subset \mathcal{C}(K)$ et $S_n = \sum_{k=0}^n f_k$. On dit que

- la série $\sum f_k$ converge simplement sur K si la suite de fonctions $(S_n)_{n \geq 0}$ a une limite simple sur K qu'on note $\sum_{k=0}^{+\infty} f_k$. Pour $x \in K$, celle-ci est donnée par la série (convergente) $\sum_{k=0}^{+\infty} f_k(x)$.
- la série $\sum_{k=0}^{+\infty} f_k$ converge uniformément sur K vers $\sum_{k=0}^{+\infty} f_k$.
- la série $\sum_{k=0}^{+\infty} f_k$ converge normalement sur K si la série numérique $\sum_{k=0}^{+\infty} \|f_k\|_\infty$ converge.

Il suffit maintenant de reformuler les résultats de convergence sur les suites de fonctions en considérant la suite des sommes partielles. En premier lieu, comme $\mathcal{C}(K)$ est complet, la convergence uniforme étant la convergence dans $\mathcal{C}(K)$, on en déduit immédiatement le théorème suivant :

Théorème 1.21 (Critère de Cauchy des séries de fonctions continues)

Soit $(f_k)_{k \geq 0}$ une suite de fonctions continues sur un compact K . Alors la série de fonctions $\sum_{k=0}^{\infty} f_k$ converge uniformément sur K si (et seulement si), pour tout $\varepsilon > 0$, il existe N tel que, si $n \geq m \geq N$, alors

$$\sup_{x \in K} \left| \sum_{k=m}^n f_k(x) \right| < \varepsilon.$$

Dans ce cas, $\sum_{k=0}^{\infty} f_k$ est continue.

Enfin, si $\sum_{k=0}^{\infty} f_k$ converge simplement sur K , on peut considérer la suite des restes $R_N = S - S_N = \sum_{k=N+1}^{\infty} f_k$. Alors la série converge uniformément si et seulement si la suite des restes converge uniformément vers 0.

Un corollaire immédiat est le suivant :

Corollaire 1.22

Soit $(f_k)_{k \geq 0}$ une suite de fonctions continues sur un compact K . Si la série $\sum_{k=0}^{\infty} f_k$ converge normalement sur K , alors elle converge uniformément sur K .

C'est un fait général dans un Banach (qui est d'ailleurs équivalent à la complétude dans un espace vectoriel normé). La démonstration est exactement la même que le fait qu'une série (numérique) absolument convergente est convergente :

DÉMONSTRATION. Soit $\varepsilon > 0$, le critère de Cauchy des séries numériques implique qu'il existe N tel que, si $m, n \geq N$, $\sum_{k=m}^n \|f_k\|_{\infty} < \varepsilon$. Mais alors, avec l'inégalité triangulaire,

$\left\| \sum_{k=m}^n f_k \right\|_{\infty} \leq \sum_{k=m}^n \|f_k\|_{\infty} < \varepsilon$. Ainsi, la série $\sum_{k=0}^{+\infty} f_k$ vérifie le critère de Cauchy, donc converge uniformément sur K . ■

Il nous reste à reformuler les résultats de dérivation et d'intégration. Pour simplifier, nous ne les considérons que pour des fonctions $[a, b] \rightarrow \mathbb{R}$.

Théorème 1.23

Soit $(f_n)_{n \geq 0}$ une suite de fonctions continues $[a, b] \rightarrow \mathbb{R}$. Alors

– si $\sum_{n=0}^{+\infty} f_n$ converge uniformément sur $[a, b]$ alors cette somme est continue et, de plus

$$\int_a^b \left(\sum_{n=0}^{+\infty} f_n(x) \right) dx = \sum_{n=0}^{+\infty} \int_a^b f_n(x) dx;$$

– si les f_n sont de classe $\mathcal{C}^1$, s'il existe x_0 tel que $\sum_{n=0}^{+\infty} f_n(x_0)$ converge et si $\sum_{n=0}^{+\infty} f'_n$

converge uniformément sur $[a, b]$ alors $\sum_{n=0}^{+\infty} f_n$ est de classe $\mathcal{C}^1$ sur $[a, b]$ et

$$\frac{d}{dx} \left(\sum_{n=0}^{+\infty} f_n(x) \right) = \sum_{n=0}^{+\infty} \left(\frac{df}{dx}(x) \right).$$

Le critère des séries alternées permet d'estimer des restes. On en déduit le résultat suivant :

Théorème 1.24 (Critère de Leibniz uniforme)

Soit K un compact. Soit $(f_k)_{k \geq 0} \subset \mathcal{C}(K)$ une suite de fonctions positives et telles que, pour tout $x \in K$, $(f_k(x))_{k \geq 0}$ décroît i.e. $0 \leq f_{k+1}(x) \leq f_k(x)$.

Si $f_k \rightarrow 0$ simplement sur K alors la série $\sum_{k \geq 0} (-1)^k f_k$ converge simplement sur K .

Si $f_k \rightarrow 0$ uniformément sur K alors la série $\sum_{k \geq 0} (-1)^k f_k$ converge uniformément sur K .

DÉMONSTRATION. La convergence simple résulte directement du critère des séries alternées appliqué à la série réelle $\sum_{k=0}^{+\infty} (-1)^k f_k(x)$ en chaque $x \in K$.

Maintenant que cette série converge, on peut introduire $R_N(x) = \sum_{k=N+1}^{+\infty} (-1)^k f_k(x)$. Toujours avec le critère des séries alternées, on a l'estimation du reste $|R_N(x)| \leq f_{N+1}(x)$ donc $\|R_N\|_\infty \leq \|f_{N+1}\|_\infty$. Ainsi $R_N \rightarrow 0$ uniformément sur K si $f_k \rightarrow 0$ uniformément sur K et, comme vu plus haut, cela implique que la série $\sum_{k=0}^{+\infty} (-1)^k f_k$ converge uniformément sur K . ■

Avec ce critère, il est facile de montrer que la convergence uniforme n'implique pas la convergence normale :

Exemple 1.25 (Une série uniformément mais pas normalement convergente)

La série de fonctions $\sum_{n=1}^{+\infty} \frac{(-1)^n}{n+x}$ converge uniformément sur $[0, +\infty[$, mais elle ne converge pas normalement sur cet intervalle.

Pour $n \geq 1$ et $x \geq 0$, posons $f_n(x) = \frac{1}{n+x}$. Pour tout $x \geq 0$, la suite $(f_n(x))_{n \geq 1}$ est décroissante et tend vers 0. De plus,

$$\sup_{x \geq 0} f_n(x) = f_n(0) = \frac{1}{n} \rightarrow 0.$$

Le critère de Leibniz uniforme montre donc que la série

$$\sum_{n=1}^{+\infty} (-1)^n f_n(x) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n+x}$$

converge uniformément sur $[0, +\infty[$.

En revanche, pour tout $n \geq 1$,

$$\left\| \frac{(-1)^n}{n+x} \right\|_\infty = \sup_{x \geq 0} \frac{1}{n+x} = \frac{1}{n}$$

(le sup étant atteint en $x = 0$). Par conséquent, la série

$$\sum_{n=1}^{+\infty} \left\| \frac{(-1)^n}{n+x} \right\|_\infty$$

diverge et la série $\sum_{n=1}^{+\infty} \frac{(-1)^n}{n+x}$ ne converge pas normalement sur $[0, +\infty[$.

Nous allons maintenant utiliser cela pour construire des fonctions aux propriétés exotiques.

1.3.2. *L'équation de Cauchy et sa stabilité.*

Definition 1.26 (Fonctions additive et équation de Cauchy)

Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est dite *additive* si, elle vérifie l'équation de Cauchy pour tous $x, y \in \mathbb{R}$, $f(x+y) = f(x) + f(y)$.

Exemple 1.27

Une fonction linéaire, $f(x) = cx$ est évidemment additive.

Il existe des exemples, nécessairement non triviaux, de fonctions additives qui ne sont pas linéaires. En voici une construction qui repose sur l'existence de bases de Hamel : on prend $\alpha \in \mathbb{R} \setminus \mathbb{Q}$ un irrationnel. Ainsi $1, \alpha$ sont linéairement indépendants sur $\mathbb{Q}$. Une des conséquences classiques de l'axiome du choix est qu'on peut compléter $\{1, \alpha\}$ en une base $\mathcal{B}$ du $\mathbb{Q}$ -espace vectoriel $\mathbb{R}$. On pose alors $f(\alpha) = 1$ et $f(b) = 0$ si $b \in \mathcal{B} \setminus \{\alpha\}$ puis on prolonge f à $\mathbb{R}$ par $\mathbb{Q}$ -linéarité : tout $x \in \mathbb{R}$ s'écrit de façon unique $x = \sum_{j=1}^n x_j b_j$ avec $n \in \mathbb{N}$ et $b_1, \dots, b_n \in \mathcal{B}$ et $x_1, \dots, x_n \in \mathbb{Q}$. On pose alors

$$f(x) = \sum_{j=1}^n x_j f(b_j) = \begin{cases} x_j & \text{s'il existe } j \text{ t.q. } b_j = \alpha \\ 0 & \text{sinon.} \end{cases}$$

Mais alors f n'est pas de la forme $f(x) = cx$ avec $c \in \mathbb{R}$ puisqu'on aurait alors $0 = f(1) = c$ donc $f(x) = 0$ pour tout x .

On peut montrer que cette fonction n'est pas Lebesgue mesurable.

Le résultat le plus simple dans cette direction est le suivant :

Proposition 1.28

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction continue et additive. Alors il existe $c \in \mathbb{R}$ tel que $f(x) = cx$.

DÉMONSTRATION. En effet, une récurrence simple montre que, pour tout $x \in \mathbb{R}$ et $n \in \mathbb{N}^*$, $f(nx) = nf(x)$. On pose alors $c = f(1)$,

– comme $f(0) = f(0 + 0) = 2f(0)$ donc $f(0) = 0$ puis $0 = f(x - x) = f(x) + f(-x)$ donc $f(-x) = -f(x)$, donc $f(nx) = nf(x)$ pour $x \in \mathbb{R}$, $n \in \mathbb{Z}$. En particulier, si $n \in \mathbb{Z}$, $f(n) = cn$.

– si $x = \frac{m}{n} \in \mathbb{Q}$, on a $f(nx) = nf(x)$ donc $f\left(\frac{m}{n}\right) = \frac{1}{n}f(m) = \frac{m}{n}c$ i.e. on a $f(x) = cx$ pour $x \in \mathbb{Q}$.

– si $x \in \mathbb{R}$, on prend une suite $\frac{m_k}{n_k} \in \mathbb{Q}$ telle que $\frac{m_k}{n_k} \rightarrow x$. On a alors $c \frac{m_k}{n_k} = f\left(\frac{m_k}{n_k}\right)$.

Par continuité, en passant à la limite, on obtient $cx = f(x)$. ■

Théorème 1.29 (Stabilité de Hyers–Ulam de l'équation de Cauchy)

Soit $\varepsilon > 0$ et soit $f : \mathbb{R} \rightarrow \mathbb{R}$ une fonction telle que, pour tous $x, y \in \mathbb{R}$, $|f(x + y) - f(x) - f(y)| \leq \varepsilon$. Alors il existe une unique fonction additive $g : \mathbb{R} \rightarrow \mathbb{R}$ telle que, pour tout $x \in \mathbb{R}$, $|f(x) - g(x)| \leq \varepsilon$.

Si, de plus, f est continue, alors il existe $c \in \mathbb{R}$ tel que, pour tout $x \in \mathbb{R}$, $|f(x) - cx| \leq \varepsilon$.

Remarque 1.30

Ce résultat montre qu'une fonction qui satisfait l'équation de Cauchy $f(x + y) = f(x) + f(y)$ à une erreur uniformément bornée près est elle-même uniformément proche d'une véritable solution de cette équation. La preuve repose directement sur la convergence normale d'une série de fonctions.

DÉMONSTRATION. Pour $n \geq 0$, posons $f_n(x) = \frac{1}{2^n} f(2^n x)$.

En prenant $y = x$ dans l'hypothèse, on obtient $|f(2x) - 2f(x)| \leq \varepsilon$. Par conséquent, pour tout $x \in \mathbb{R}$,

$$|f_{n+1}(x) - f_n(x)| = \frac{1}{2^{n+1}} |f(2^{n+1}x) - 2f(2^n x)| \leq \frac{\varepsilon}{2^{n+1}}.$$

Ainsi $\|f_{n+1} - f_n\|_\infty \leq \frac{\varepsilon}{2^{n+1}}$ donc la série de fonctions $\sum_{n=0}^{+\infty} (f_{n+1} - f_n)$ converge normalement donc uniformément sur $\mathbb{R}$. La suite (f_n) converge donc uniformément sur $\mathbb{R}$ vers une fonction $g : \mathbb{R} \rightarrow \mathbb{R}$. Ici on utilise le fait que $f_{N+1} - f_0 = \sum_{n=0}^N (f_{n+1} - f_n)$. Remarquons de plus que si f est continue, f_n aussi donc g aussi.

De plus

$$\begin{aligned} |g(x) - f(x)| &= |g(x) - f_0(x)| = \left| \sum_{n=0}^{+\infty} (f_{n+1}(x) - f_n(x)) \right| \\ &\leq \sum_{n=0}^{+\infty} |f_{n+1}(x) - f_n(x)| \leq \sum_{n=0}^{+\infty} \frac{\varepsilon}{2^{n+1}} = \varepsilon \end{aligned}$$

Ainsi $|g(x) - f(x)| \leq \varepsilon$.

Il faut donc montrer que g est additive. Une fois cela fait, si f est continue, g aussi donc g est de la forme $g(x) = cx$ ce qui donnera la dernière partie du théorème.

Soient donc $x, y \in \mathbb{R}$. On a

$$|f_n(x+y) - f_n(x) - f_n(y)| = \frac{1}{2^n} |f(2^n(x+y)) - f(2^n x) - f(2^n y)| \leq \frac{\varepsilon}{2^n}.$$

En faisant tendre n vers $+\infty$ et en utilisant la convergence simple de f_n vers g , on obtient $g(x+y) = g(x) + g(y)$ qui est donc additive.

Il reste à montrer l'unicité. Soient g et h deux fonctions additives telles que

$$|f(x) - g(x)| \leq C \quad \text{et} \quad |f(x) - h(x)| \leq C \quad \text{pour tout } x \in \mathbb{R}$$

pour une certaine constante $C > 0$. Alors $|g - h| \leq |g - f| + |f - h| \leq 2C$ donc $u = g - h$ est bornée. Elle est également additive. Mais alors, pour tout entier $m \geq 1$ et tout réel x , $|u(mx)| = m|u(x)| \leq 2C$ donc, en faisant $m \rightarrow +\infty$, $u(x) = 0$ et $g = h$. ■

1.3.3. La fonction de Weierstrass : une fonction continue nulle-part dérivable.

Théorème 1.31 (Fonction W de Weierstrass)

Soit $0 < a < 1$ et $b > 1$ tel que $ab > 1$. Soit f la série de fonctions définie sur $\mathbb{R}$ par

$$f(x) = \sum_{k=0}^{+\infty} a^k \cos(2\pi b^k x).$$

Alors, cette série converge uniformément sur $\mathbb{R}$, sa somme est continue sur $\mathbb{R}$ et nulle-part dérivable.

Lorsque b est un entier, ceci est une série de Fourier lacunaire.

Lemme 1.32

Soit $b > 1$ et $u \in \mathcal{S}(\mathbb{R})$ une fonction dont la transformée de Fourier est à support inclus dans $]b^{-1}, b[$ avec $\hat{u}(1) = 1$. Soit f une fonction continue bornée sur $\mathbb{R}$. Pour

$s > 0, t \in \mathbb{R}$

$$\mathcal{W}_u[f](s, t) = \int_{\mathbb{R}} f(x) u(s(x - t)) \, dx.$$

Si f est dérivable en t , alors, $b^{2j} \mathcal{W}_u[f](b^j, t) \rightarrow 0$.

En partant d'une fonction v de classe $\mathcal{C}^\infty$ à support $[b^{-1}, b]$ avec $v(1) = 1$ (on a construit de telles fonctions dans le chapitre sur les fonctions à une variable), et en remarquant que v est dans la classe de Schwarz, il suffit de prendre pour u la transformée de Fourier inverse de v .

On remarque ensuite que $\mathcal{W}_u[f]$ est bien définie lorsque f est à croissance au plus polynomiale $|f(x)| \leq C(1 + |x|)^N$. En effet, $x \rightarrow u(s(x - t))$ est également dans la classe de Schwarz, il existe donc C' tel que $|u(s(x - t))| \leq C'(1 + |x|)^{-N-2}$ et alors

$$\int_{\mathbb{R}} |f(x) u(s(x - t))| \, dx \leq \int_{\mathbb{R}} \frac{CC'}{(1 + |x|)^2} \, dx < +\infty.$$

Enfin $f \rightarrow \mathcal{W}_u[f]$ est linéaire sur l'ensemble des fonctions à croissance au plus polynomiale.

La notation $\mathcal{W}_u$ provient de *wavelet* (ondelette) et non de Weierstrass.

DÉMONSTRATION DU LEMME 1.32. Tout d'abord, remarquons que $u \in \mathcal{S}(\mathbb{R})$ et donc que sa transformée de Fourier est également dans $\mathcal{S}(\mathbb{R})$. Ensuite, si f est dérivable en t , son développement limité à l'ordre 1 s'écrit

$$f(x) = f(t) + (x - t)f'(t) + (x - t)R(x - t)$$

avec $R(s) \rightarrow 0$ quand $s \rightarrow 0$. Notons que $R(0) = 0$ et, pour $s \neq 0$,

$$R(s) = \frac{f(t + s) - f(t) - sf'(t)}{s}.$$

Ainsi R est une fonction continue (la formule donne la continuité en dehors de 0) et bornée puisque pour $|s| \geq 1$, $|R(s)| \leq 2\|f\|_\infty + |f'(t)|$.

La linéarité de $\mathcal{W}_u$ donne alors

$$\begin{aligned} \mathcal{W}_u[f](b^j, t) &= f(t) \int_{\mathbb{R}} u(b^j(x - t)) \, dx + f'(t) \int_{\mathbb{R}} (x - t) u(b^j(x - t)) \, dx \\ &\quad + \int_{\mathbb{R}} (x - t) R(x - t) u(b^j(x - t)) \, dx. \end{aligned}$$

Mais

$$b^{2j} \int_{\mathbb{R}} u(b^j(x - t)) \, dx = b^j \int_{\mathbb{R}} u(s) \, ds = b^j \hat{u}(0) = 0$$

par construction de u , et

$$b^{2j} \int_{\mathbb{R}} (x - t) u(b^j(x - t)) \, dx = b^{j-1} \int_{\mathbb{R}} su(s) \, ds = b^{j-1} 2i\pi \hat{u}'(0) = 0$$

où on utilise le fait que la dérivée de la transformée de Fourier d'une fonction $\varphi \in \mathcal{B}(\mathbb{R})$ est la transformée de Fourier de $-2i\pi x\varphi(x)$. Il nous reste à montrer que

$$b^{2j} \int_{\mathbb{R}} (x - t) R(x - t) u(b^j(x - t)) \, dx = b^{2j} \int_{\mathbb{R}} s R(s) u(b^j s) \, ds \rightarrow 0.$$

Fixons donc $\varepsilon > 0$. Il existe $\eta > 0$ tel que, si $|s| < \eta$, alors $|R(s)| < \varepsilon$. Mais alors

$$\left| b^{2j} \int_{|s| \leq \eta} R(s) su(b^j s) \, ds \right| \leq \varepsilon \int_{|t| < \eta b^j} |tu(t)| \, dt \leq \varepsilon \|tu\|_1.$$

D'autre part,

$$\begin{aligned} \left| b^{2j} \int_{|s| \geq \eta} R(s) s u(b^j s) \, ds \right| &= \left| b^{-j} \int_{|s| \geq \eta} \frac{R(s)}{s^2} (b^j s)^3 u(b^j s) \, ds \right| \\ &\leq \frac{\|s^3 u\|_\infty \|R\|_\infty}{b^j} \int_{|s| \geq \eta} \frac{1}{s^2} \, ds \rightarrow 0 \end{aligned}$$

quand $j \rightarrow \infty$. Il existe donc J tel que, si $j \geq J$,

$$\left| b^{2j} \int_{\mathbb{R}} R(s) s u(b^j s) \, ds \right| \leq (\|tu\|_1 + 1)\varepsilon$$

d'où la convergence $b^{2j} \mathcal{W}_u[f](b^j, t) \rightarrow 0$ lorsque f est dérivable en t . ■

DÉMONSTRATION DU THÉORÈME DE WEIERSTRASS. Tout d'abord, comme $|a^k \cos(2\pi b^k x)| \leq a^k$ et que, pour $0 < a < 1$, $\sum a^k$ converge. La série $\sum a^k \cos(2\pi b^k x)$ est donc normalement convergente sur $\mathbb{R}$, donc uniformément convergente. Comme $x \rightarrow a^k \cos(2\pi b^k x)$ est continue, la somme f de cette série est continue.

On considère ensuite

$$W_u[f](s, t) = \int_{\mathbb{R}} \sum_{k=0}^{+\infty} a^k \cos(2\pi b^k x) u(s(x-t)) \, dx = \sum_{k=0}^{+\infty} a^k \int_{\mathbb{R}} \cos(2\pi b^k x) u(s(x-t)) \, dx.$$

Pour justifier cette interversion, il convient d'être prudent : la série converge certes normalement sur $\mathbb{R}$, mais l'intégration n'est pas sur un intervalle borné. Toutefois, le théorème de convergence dominée permet de conclure aisément :

$$- \text{ on a convergence ponctuelle, } \sum_{k=0}^N a^k \cos(b^k x) u(s(x-t)) \rightarrow \sum_{k=0}^{+\infty} a^k \cos(b^k x) u(s(x-t))$$

pour tout $x \in \mathbb{R}$;

$$- \text{ on a la majoration } \sum_{k=0}^N |a^k \cos(2\pi b^k x) u(s(x-t))| \leq \sum_{k=0}^{+\infty} a^k |u(s(x-t))| \leq \frac{|u(s(x-t))|}{1-a} \in$$

$L^1(\mathbb{R})$.

On remarque ensuite que, si k, j sont des entiers,

$$\begin{aligned} &\int_{\mathbb{R}} \cos(2\pi b^k x) u(b^j(x-t)) \, dx \\ &= \frac{1}{2} \int_{\mathbb{R}} u(b^j(x-t)) e^{2i\pi b^k x} \, dx + \frac{1}{2} \int_{\mathbb{R}} u(b^j(x-t)) e^{-2i\pi b^k x} \, dx \\ &= \frac{e^{2i\pi b^{k-j}t}}{2b^j} \int_{\mathbb{R}} u(y) e^{2i\pi b^{k-j}y} \, dy + \frac{e^{-2i\pi b^{k-j}t}}{2b^j} \int_{\mathbb{R}} u(y) e^{-2i\pi b^{k-j}y} \, dy \\ &= \frac{1}{2b^j} (e^{2i\pi b^{k-j}t} \widehat{u}(-b^{k-j}) + e^{-2i\pi b^{k-j}t} \widehat{u}(b^{k-j})). \end{aligned}$$

Mais $\widehat{u}$ est à support inclus dans $]1/b, b[$, $\widehat{u}(-b^{k-j}) = 0$ alors que $\widehat{u}(b^{k-j}) = 0$ si $k \neq j$ et $\widehat{u}(b^{k-k}) = \widehat{u}(1) = 1$. Ainsi

$$b^{2j} W_u[f](b^j, t) = \sum_{k=0}^{+\infty} a^k \int_{\mathbb{R}} \cos(2\pi b^k x) u(b^j(x-t)) \, dx = \frac{(ab)^j}{2} e^{-2i\pi t}.$$

Comme nous avons supposé que $ab > 1$, cette quantité ne tend pas vers 0. D'après le lemme 1.32, f n'est pas dérivable en t . ■

Corollaire 1.33

L'ensemble des fonctions continues nulle-part dérivable est dense dans $\mathcal{C}([0, 1])$.

DÉMONSTRATION. Soit $\varphi \in \mathcal{C}([0, 1])$ et $\varepsilon > 0$. Soit f la fonction de Weierstrass construite précédemment, restreinte à $[0, 1]$. Comme $\varphi - f$ est continue sur $[0, 1]$ et que les polynômes sont denses dans $\mathcal{C}([0, 1])$ (voir plus loin pour la démonstration), il existe un polynôme P tel que $\|\varphi - (f + P)\|_\infty = \|(\varphi - f) - P\|_\infty \leq \varepsilon$. Comme P est dérivable partout, $\psi = f + P$ est encore dérivable nulle-part (sinon $f = \psi - P$ serait dérivable aux mêmes points que ψ). ψ fournit l'approximation à ε -près de φ cherchée. ■

1.3.4. Une seconde fonction continue nulle-part dérivable.

Nous allons maintenant donner une construction un peu plus simple, mais qui n'est plus une série de Fourier d'une seconde fonction continue nulle-part dérivable :

SECONDE CONSTRUCTION D'UNE FONCTION CONTINUE NULLE-PART DÉRIVABLE.

La construction part d'une fonction g continue, paire, 4-périodique, définie par $g(x) = 1 - x$ sur $[0, 2]$ (et donc $g(x) = 1 - |x|$ sur $[-2, 2]$ et $g(x) = x - 3$ sur $[2, 4]$). Ainsi, g est une fonction continue, affine par morceaux, dont le graphe relie les points $(2k, (-1)^k)$, $k \in \mathbb{Z}$. En particulier, $|g| \leq 1$. Remarquons aussi que g est 1-lipschitzienne : $|g(x) - g(y)| \leq |x - y|$. En effet, si $|x - y| \geq 2$ alors $|g(x) - g(y)| \leq 2 \leq |x - y|$. Supposons donc que $|x - y| \leq 2$ et, quitte à échanger x et y , que $x \leq y$ donc $x \leq y \leq x + 2$. Soit alors N tel que $4N - 2 \leq x \leq 4N + 2$. Par 4-périodicité, on peut remplacer x, y par $x - 4N, y - 4N$ et supposer que $N = 0$ donc que $-2 \leq x \leq 2$. On distingue alors plusieurs cas

- si $-2 \leq x, y \leq 2$ alors $|g(x) - g(y)| = |(1 - |x|) - (1 - |y|)| = ||x| - |y|| \leq |x - y|$
- sinon $0 \leq x \leq 2$ et $2 \leq y \leq 4$ et

$$|g(x) - g(y)| = |1 - x - (y - 3)| = |2 - x - (y - 2)| = ||x - 2| - |y - 2|| \leq |(x - 2) - (y - 2)| = |x - y|.$$

On définit ensuite f par

$$f(x) = \sum_{k=1}^{+\infty} \frac{1}{2^k} g(2^{2^k} x).$$

Tout d'abord, comme $\sup_{x \in \mathbb{R}} |g(x)| = 1$, cette série est normalement convergente dans $\mathcal{C}_b(\mathbb{R})$ (l'espace de Banach des fonctions continues bornées sur $\mathbb{R}$)

$$= \sum_{k=1}^{+\infty} \frac{1}{2^k} \sup_{x \in \mathbb{R}} |g(2^{2^k} x)| = \sum_{k=1}^{+\infty} \frac{1}{2^k} = 1 < +\infty$$

donc uniformément convergente sur $\mathbb{R}$ et donc f est bien une fonction continue bornée.

Fixons maintenant $x \in \mathbb{R}$ et construisons une suite (h_n) telle que $h_n \rightarrow 0$ et

$$\left| \frac{f(x + h_n) - f(x)}{h_n} \right| \rightarrow +\infty.$$

Pour ce x fixé, pour chaque n , il existe N tel que $2^{2^n} x \in [2N, 2N + 2]$ et on choisit $\varepsilon_n = \pm 1$ de sorte que $2^{2^n} x + \varepsilon_n \in [2N, 2N + 2]$ i.e. $\varepsilon_n = 1$ si $2^{2^n} x \in [2N, 2N + 1]$ et $\varepsilon_n = -1$ sinon. On pose alors $h_n = \varepsilon_n 2^{-2^n}$ et on observe que

- si $k > n$, $|2^{2^k} h_n| = |2^{2^k - 2^n}| = 4^{2^{k-1} - 2^n}$ et est donc un multiple de 4. Par 4-périodicité de g on a donc

$$g(2^{2^k} (x + h_n)) - g(2^{2^k} x) = 0.$$

- si $k = n$,

$$|g(2^{2^k} (x + h_n)) - g(2^{2^k} x)| = |g(2^{2^n} x + \varepsilon_n) - g(2^{2^n} x)| = |2^{2^n} x + \varepsilon_n - 2^{2^n} x| = 1$$

en utilisant que g est affine de pente 1 sur $[2n, 2N + 2]$.

- si $k < n$, en utilisant le fait que g soit 1-Lipschitzienne,

$$|g(2^{2^k} (x + h_n)) - g(2^{2^k} x)| \leq |2^{2^k} h_n| \leq 2^{2^{k-1} - 2^n} = 2^{-2^{n-1}}.$$

Ainsi

$$\sum_{k=1}^{n-1} \frac{1}{2^k} |g(2^{2^k}(x+h_n)) - g(2^{2^k}x)| \leq 2^{-2^{n-1}} \sum_{k=1}^{n-1} \frac{1}{2^k} \leq 2^{-2^{n-1}}.$$

On en déduit que

$$\begin{aligned} \left| \frac{f(x+h_n) - f(x)}{h_n} \right| &= 2^{2^n} \left| \sum_{k=1}^{+\infty} \frac{1}{2^k} (g(2^{2^k}(x+h_n)) - g(2^{2^k}x)) \right| \\ &\geq -2^{2^n} \sum_{k=1}^{n-1} \frac{1}{2^k} |g(2^{2^k}(x+h_n)) - g(2^{2^k}x)| \\ &\quad + \frac{2^{2^n}}{2^n} |g(2^{2^n}(x+h_n)) - g(2^{2^n}x)| - 2^{2^n} \sum_{k=n+1}^{+\infty} \frac{1}{2^k} |g(2^{2^k}(x+h_n)) - g(2^{2^k}x)| \\ &\geq -2^{2^n-2^{n-1}} + 2^{2^n} \frac{1}{2^n} = 2^{2^n-n}(1-2^{-2^{n-1}}) \rightarrow +\infty. \end{aligned}$$

Ainsi f n'est pas dérivable en x . ■

1.3.5. L'escalier de Cantor.

Construisons maintenant une fonction continue, non-constante dont la dérivée existe presque partout et est nulle presque partout.

Supposons que f soit une fonction $[0,1] \rightarrow \mathbb{R}$, dérivable en tout point de $[0,1]$ et telle que $f'(x) = 0$ pour tout $x \in [0,1]$. Une telle fonction est de classe $\mathcal{C}^1$ (continue car dérivable et la dérivée est clairement continue). D'après l'inégalité des accroissements finis, pour tout $x \in [0,1]$,

$$|f(x) - f(0)| \leq |x - 0| \sup_{t \in [0,1]} |f'(t)| = 0$$

donc $f(x) = f(0)$ i.e. f est constante.

Notons que la conclusion reste vraie si on suppose seulement que $f'(x) = 0$ pour tout $x \in [0,1] \setminus \{t_j\}_{j \in \mathbb{N}}$ i.e. si la dérivée est nulle sauf en un ensemble dénombrable de points. De même, si on suppose que f est Lipschitzienne, alors elle est constante.

Théorème 1.34 (Escalier du diable ou de Cantor)

Il existe une fonction $f : [0,1] \rightarrow \mathbb{R}$ continue, croissante avec $f(0) = 0$ et $f(1) = 1$ et telle que f' existe et vaut 0 presque partout sur $[0,1]$.

Remarque 1.35

Il faut noter qu'ici le théorème fondamental de l'analyse n'est pas vérifié (il faut que f soit un peu mieux : *absolument continue*). De même, la dérivée usuelle, bien qu'existant presque partout, n'est pas la dérivée au sens des distributions.

DÉMONSTRATION. Cette construction est basée sur la construction de l'ensemble de Cantor triadique.

On commence par poser $I_{0,0} = [0,1]$ et $f_0(x) = x$. Ensuite, on subdivise $I_{0,0}$ en trois intervalles $I_{1,0} = [0,1/3]$, $I_{1,1} = [1/3,2/3]$ et $I_{1,2} = [2/3,1]$ et on définit f_1 la fonction continue sur $[0,1]$, affine sur chaque $I_{1,j}$ et telle que $f_1(0) = f_0(0)$, $f_1(1) = f_0(1)$, $f_1(1/3) = f_1(2/3) = \frac{f_0(1) + f_0(2)}{2} = \frac{1}{2}$. Notez que f_1 est constante sur $I_{1,1}$.

On itère ensuite cette construction : une fois f_n construite, continue et affine sur chaque $I_{n,k} = [k3^{-n}, (k+1)3^{-n}]$, $k = 0, \dots, 3^n - 1$, on définit alors f_{n+1} en découpant chaque $I_{n,k}$ en trois intervalles : $I_{n+1,3k}$, $I_{n+1,3k+1}$, $I_{n+1,3k+2}$ et en posant $f_{n+1}(3k3^{-n-1}) =$

$$f_{n+1}(k3^{-n}) = f_n(k3^{-n}), f_{n+1}((3k+3)3^{-n-1}) = f_{n+1}((k+1)3^{-n}) = f_n((k+1)3^{-n}) \text{ et } f_{n+1}((3k+1)3^{-n-1}) = f_{n+1}((3k+2)3^{-n-1}) = \frac{f_n(k3^{-n}) + f_n((k+1)3^{-n})}{2}.$$

Notez qu'avec cette construction,

- comme f_n est croissante, il en va de même de f_{n+1} .
- Si f_n est constante sur $I_{n,k} = I_{n+1,3k} \cup I_{n+1,3k+1} \cup I_{n+1,3k+2}$ alors $f_{n+1} = f_n$ sur $I_{n,k}$ et y est constante.
- Si f_n n'est pas constante sur $I_{n,k}$ alors f_{n+1} est constante sur $I_{n+1,3k+1}$ et

$$\begin{aligned} \sup_{I_{n,k}} |f_n - f_{n+1}| &= f_{n+1}((3k+1)3^{-n-1}) - f_n((3k+1)3^{-n-1}) \\ &= f_{n+1}((3k+2)3^{-n-1}) - f_n((3k+2)3^{-n-1}) \\ &= \frac{f_n((k+1)3^{-n}) - f_n(k3^{-n})}{2}. \end{aligned}$$

Une récurrence immédiate montre que cette quantité est $\leq 2^{-n}$. Ainsi $\|f_{n+1} - f_n\|_\infty \leq 2^{-n}$. La série $\sum_{n \geq 1} (f_{n+1} - f_n)$ est donc normalement convergente, donc uniformément convergente sur $[0, 1]$. Mais ses sommes partielles sont les f_n qui convergent donc uniformément vers une fonction f .

Notons maintenant A_n la réunion (disjointe à un ensemble fini près) des intervalles $I_{n,k}$ sur lesquels f_n n'est pas constante, on écrit $A_n = \bigcup_{k \in F_n} I_{n,k}$. On vient de voir que $A_{n+1} \subset A_n$ et que, de chaque intervalle $I_{n,k}$, $k \in F_n$ on retire le tiers central $I_{n+1,3k+1}$ pour former A_{n+1} : $A_{n+1} = \bigcup_{k \in 3F_n \cup (3F_n+2)} I_{n+1,k}$. Donc $|A_{n+1}| = \frac{2}{3}|A_n|$ et comme $A_0 = 1$

on en déduit que $|A_n| = (2/3)^n$ puis que $\left| \bigcap_{n \geq 0} A_n \right| = 0$.

Mais f_n étant constante sur $D_n = [0, 1] \setminus A_n$ et $f = f_n$ sur D_n qui est une réunion d'intervalles ouverts, f est dérivable sur D_n et $f' = 0$ sur cet ensemble. Il en résulte que f est dérivable sur $\bigcup_{n \geq 0} D_n$ avec $f' = 0$ sur cet ensemble. Comme $\left| \bigcup_{n \geq 0} D_n \right| = 1 - \left| \bigcap_{n \geq 0} A_n \right| = 1$, on obtient le résultat annoncé. ■

2. Compacité dans $\mathcal{C}(K)$ et applications

2.1. Théorème de Helly.

Théorème 2.1 (Théorème de sélection de Helly)

Soit $(f_n)_{n \geq 1}$ une suite de fonctions croissantes sur un intervalle I (borné ou non). On suppose que la suite est uniformément bornée, c'est-à-dire qu'il existe une constante $M > 0$ telle que, pour tout $n \geq 1$ et tout $x \in [a, b]$, $|f_n(x)| \leq M$. Alors il existe une sous-suite $(f_{n_k})_{k \geq 1}$ et une fonction croissante $f : [a, b] \rightarrow [-M, M]$ telles que, pour tout $x \in [a, b]$, $f_{n_k}(x) \rightarrow f(x)$.

En d'autres termes, toute suite bornée de fonctions croissantes sur $[a, b]$ admet une sous-suite simplement convergente.

Nous donnerons plus loin, dans le théorème de Prohorov ?? une application aux probabilités.

DÉMONSTRATION. La démonstration se fait en deux étapes. On commence par obtenir la convergence sur un sous-ensemble dénombrable dense de I , puis on utilise la monotonie pour obtenir la convergence en tout point de continuité de la fonction limite. Enfin, on effectue une dernière extraction aux points de discontinuité.

Quitte à composer les f_n par un homéomorphisme $I \rightarrow [0, 1]$, on peut supposer que $I = [0, 1]$.

Étape 1. *Convergence sur un ensemble dense.*

Soit $\mathcal{D} = \mathbb{Q} \cap [0, 1]$, qui est dénombrable dense dans $[0, 1]$ et contient 0 et 1. On énumère $\mathcal{D} = \{x_1, x_2, \dots\}$.

La suite numérique $(f_n(x_1))_n$ est bornée dans $[-M, M]$. D'après le théorème de Bolzano-Weierstrass, on peut en extraire une sous-suite $(f_{n_k^{(1)}})_k$ telle que $f_{n_k^{(1)}}(x_1)$ converge lorsque $k \rightarrow +\infty$.

De cette sous-suite, on extrait une nouvelle sous-suite $(f_{n_k^{(2)}})_k$ telle que $f_{n_k^{(2)}}(x_2)$ converge. En poursuivant ainsi, on construit une famille de sous-suites emboîtées telle que, pour tout $j \geq 1$, $(f_{n_k^{(j)}}(x_i))_k$ converge pour tout $i \leq j$.

On considère alors la sous-suite diagonale $g_k = f_{n_k^{(k)}}$. Pour tout $x_j \in \mathcal{D}$, la suite $(g_k(x_j))_k$ converge. Ainsi, pour $x \in \mathcal{D}$, on peut poser $g(x) = \lim_{k \rightarrow +\infty} g_k(x)$.

Comme chaque fonction g_k est croissante, si $x, y \in \mathcal{D}$ avec $x < y$, alors $-M \leq g_k(x) \leq g_k(y) \leq M$. En passant à la limite, $-M \leq g(x) \leq g(y) \leq M$. La fonction g est donc croissante et bornée sur $\mathcal{D}$.

Notons enfin que, si $x \in]0, 1[$, et $t \in \mathbb{Q} \cap [0, x[$, $g(t) \leq g(1) \leq M$ donc $\lim_{t \rightarrow x^-} g(t) := \sup\{g(t) : t \in \mathbb{Q} \cap]0, x[\}$ existe. De même, $\lim_{t \rightarrow x^+} g(t) := \inf\{g(t) : t \in \mathbb{Q} \cap]x, 1[\}$ existe. Comme g est croissante, on a bien entendu $-M \leq \lim_{t \rightarrow x^-} g(t) \leq \lim_{t \rightarrow x^+} g(t) \leq M$.

Étape 2. *Construction de la fonction limite.*

On commence par poser $f(0) = \lim g_n(0)$ et $f(1) = \lim g_n(1)$. Soit maintenant

$$D = \{x \in]0, 1[: \lim_{x \rightarrow x^-} g(x) < \lim_{x \rightarrow x^+} g(x)\}.$$

Tout d'abord, comme g est croissante, $\lim_{x \rightarrow x^-} g(x)$ existe puisque c'est la limite d'une suite croissante majorée et de même $\lim_{x \rightarrow x^+} g(x)$ existe. De plus, si $x \in]0, 1[\setminus D$, i.e. si $\ell^- := \lim_{x \rightarrow x^-} g(x)$ et $\ell^+ := \lim_{x \rightarrow x^+} g(x)$, sont égales, $\ell = \ell^- = \ell^+$, alors, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que si $u \in \mathbb{Q} \cap [x - \eta, x]$, $\ell - \varepsilon = \ell^- - \varepsilon < g(u) \leq g(x)$ et si $v \in \mathbb{Q} \cap [x, x + \eta]$, $g(x) \leq g(v) < \ell^+ + \varepsilon = \ell + \varepsilon$. Comme $g_n(u) \rightarrow g(u)$ et $g_n(v) \rightarrow g(v)$ il existe N tel que, si $n \geq N$

$$\ell - 2\varepsilon < g(u) - \varepsilon \leq g_n(u) \leq g_n(x) \leq g_n(v) \leq g(v) + \varepsilon < \ell + 2\varepsilon$$

donc $g_n(x) \rightarrow \ell$. Ainsi, pour tout $x \in [0, 1] \setminus D$, $f(x) := \lim g_n(x)$ existe.

Si $D = \emptyset$, on a fini. Sinon, on pose $D_n = \{x \in]0, 1[: \lim_{x \rightarrow x^+} g(x) \geq \lim_{t \rightarrow x^-} g(t) + \frac{1}{n}\}$ de sorte que $D = \bigcup_{n \geq 1} D_n$. Mais si $D_n \neq \emptyset$ et $x_1 < \dots < x_N \in D_N$ alors g a (au moins) N

sauts de hauteur au moins $\frac{1}{n}$ donc $2M \geq \max g - \min g \geq \frac{N}{n}$ et ainsi D_n a au plus $2Mn$ éléments. Comme D_n est fini, D est une réunion au plus dénombrable d'ensembles finis et est donc au plus dénombrable.

On recommence maintenant la première étape avec g_n au lieu de f_n et D au lieu de $\mathcal{D}$. Il existe donc une suite extraite (g_{n_k}) de (g_n) , telle que, pour tout $x \in D$, $f(x) := \lim g_{n_k}(x)$ existe. Bien entendu, pour $x \in [0, 1] \setminus D$ on a aussi $f(x) := \lim g_n(x) = \lim g_{n_k}(x)$. Enfin, comme (g_{n_k}) est extraite de (g_n) qui est extraite de (f_n) , (g_{n_k}) est extraite de (f_n) et répond donc à la question.

Comme une limite simple de fonctions croissantes est croissante, f est bien croissante. ■

Remarque 2.2

Une première extraction diagonale suffit pour obtenir la convergence en tout point de continuité de la fonction limite. La dernière extraction n'est nécessaire que si l'on veut obtenir la convergence en tout point du segment.

Remarque 2.3

Dans certaines applications, particulièrement pour les limites de fonctions de répartition de variables aléatoires, on se contente de la convergence aux points de continuité de la limite. Ainsi, le principe de sélection de Helly peut s'écrire comme suit :

Soit (f_n) une suite de fonctions croissantes bornées sur un intervalle I . Alors il existe une fonction croissante $f : I \rightarrow \mathbb{R}$, bornée, continue à droite, telle que, si f est continue en x alors $f_n(x) \rightarrow f(x)$.

En effet, on commence à construire f par le théorème de Helly. C'est une fonction croissante, elle a donc un ensemble au plus dénombrable de points de discontinuité. On définit ensuite sa régularisée à droite, $\tilde{f}(x) = \inf_{y>x} f(y)$ (et $\tilde{f}(b) = f(b)$ si $I = [a, b]$).

Cette fonction est bornée puisque $-M \leq f(y) \leq M$ pour tout $y > x$ donc $-M \leq \tilde{f}(x) \leq M$.

Ensuite, si $y > x_1 > z > x_2$ alors $f(y) \geq f(x_1) \geq f(z) \geq f(x_2)$ donc $\tilde{f}(x_1) \geq \tilde{f}(z) \geq f(x_2)$ en passant à l'inf sur les y , puis en passant à l'inf sur les z , $\tilde{f}(x_1) \geq \tilde{f}(x_2)$ et ainsi $\tilde{f}$ est croissante.

De plus, si f est continue en x , comme f est croissante, $\tilde{f}(x) = \inf_{y>x} f(y) = \lim_{y \rightarrow x^+} f(y) = f(x)$.

Enfin, si $x_m \rightarrow x^+$, alors $\tilde{f}(x) \leq \tilde{f}(x_m)$ donc $\tilde{f}(x) \leq \liminf \tilde{f}(x_m)$. De plus, si $y > x$, il existe M tel que si $m \geq M$, $y > x_m$. Mais alors $\tilde{f}(x_m) \leq f(y)$ et par suite $\limsup \tilde{f}(x_m) \leq f(y)$. Mais ensuite, en faisant $y \rightarrow x^+$ $\limsup \tilde{f}(x_m) \leq \tilde{f}(x)$. Ainsi $\limsup \tilde{f}(x_m) = \liminf \tilde{f}(x_m) = \lim f(x_m) = \tilde{f}(x)$. Comme la suite (x_m) est arbitraire, $\tilde{f}$ est bien continue à droite en x .

Comme les fonctions sont croissantes, si I est bornée, il suffit que les 2 valeurs extrêmes $f_n(a), f_n(b)$ soient bornées. Toute la suite est alors bornée.

Corollaire 2.4

Soit (f_n) une suite de fonctions croissantes sur $[a, b]$ telle que

$$\sup_{n \geq 1} (|f_n(a)| + |f_n(b)|) < +\infty.$$

Alors (f_n) possède une sous-suite convergeant simplement sur $[a, b]$ vers une fonction croissante.

2.2. Théorème d'Ascoli-Arzelà.**Théorème 2.5 (Ascoli-Arzelà)**

Soit K un espace métrique compact et $(f_n)_{n \geq 1}$ une suite de fonctions continues de K dans $\mathbb{R}$.

On suppose que :

- (1) la suite (f_n) est uniformément bornée, c'est-à-dire qu'il existe $M > 0$ tel que $|f_n(x)| \leq M$ pour tout $n \geq 1$ et tout $x \in K$;
- (2) la suite (f_n) est équicontinue, c'est-à-dire que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que si $x, y \in K$ sont tels que $d(x, y) < \delta$ alors $|f_n(x) - f_n(y)| < \varepsilon$ pour tout $n \geq 1$.

Alors il existe une sous-suite de (f_n) qui converge uniformément sur K vers une fonction continue.

De manière équivalente, une partie $\mathcal{F} \subset C(K, \mathbb{R})$ est relativement compacte pour la norme uniforme si et seulement si elle est uniformément bornée et équicontinue.

Montrons d'abord l'implication principale.

DÉMONSTRATION DE LA CONDITION SUFFISANTE. Comme K est un espace métrique compact, il est séparable. Il existe donc une partie dénombrable dense $D = \{x_1, x_2, \dots\} \subset K$.

La suite réelle $(f_n(x_1))_n$ est bornée. D'après le théorème de Bolzano–Weierstrass, on peut en extraire une sous-suite $(f_{\varphi_1(n)})_n$ telle que $(f_{\varphi_1(n)}(x_1))_n$ converge.

De cette sous-suite, on extrait une nouvelle sous-suite $(f_{\varphi_2(n)})_n$ telle que $(f_{\varphi_2(n)}(x_2))_n$ converge. Cette nouvelle sous-suite converge encore en x_1 , puisqu'elle est extraite de la précédente.

En poursuivant ce procédé, on construit une famille de sous-suites emboîtées telle que la p -ième sous-suite converge en chacun des points $x_1, \dots, x_p$.

Considérons alors la sous-suite diagonale $g_n = f_{\varphi_n(n)}$. Pour tout $p \geq 1$, la suite réelle $(g_n(x_p))_n$ converge. Ainsi, (g_n) converge simplement sur la partie dense D .

Montrons que (g_n) est uniformément de Cauchy sur K . Soit $\varepsilon > 0$. Par équicontinuité, il existe $\delta > 0$ tel que si $d(x, y) < \delta$ alors $|g_n(x) - g_n(y)| < \varepsilon$ pour tout $n \geq 1$.

La famille de boules ouvertes $(B(x_p, \delta))_{p \geq 1}$ recouvre K , puisque D est dense. Par compacité de K , on peut en extraire un sous-recouvrement fini : $K \subset \bigcup_{p=1}^r B(x_p, \delta)$.

Pour chaque $p \in \{1, \dots, r\}$, la suite $(g_n(x_p))_n$ converge, donc elle est de Cauchy. Comme il n'y a qu'un nombre fini de points x_p , il existe $N \geq 1$ tel que, pour tous $m, n \geq N$ et tout $p \in \{1, \dots, r\}$,

$$|g_n(x_p) - g_m(x_p)| < \varepsilon.$$

Soit maintenant $x \in K$. Il existe $p \in \{1, \dots, r\}$ tel que $d(x, x_p) < \delta$. On a alors, pour tous $m, n \geq N$,

$$|g_n(x) - g_m(x)| \leq |g_n(x) - g_n(x_p)| + |g_n(x_p) - g_m(x_p)| + |g_m(x_p) - g_m(x)| \leq 3\varepsilon.$$

La suite (g_n) est donc de Cauchy dans $C(K, \mathbb{R})$, muni de la norme uniforme.

Comme $C(K, \mathbb{R})$ est complet, il existe une fonction $g \in C(K, \mathbb{R})$ telle que $\|g_n - g\|_\infty \rightarrow 0$. La suite (f_n) possède donc une sous-suite qui converge uniformément sur K . ■

DÉMONSTRATION DE LA RÉCIPROQUE Supposons que $\mathcal{F} \subset C(K, \mathbb{R})$ soit relativement compacte pour la norme uniforme.

La fermeture $\overline{\mathcal{F}}$ est compacte, donc bornée. Il existe ainsi $M > 0$ tel que $\|f\|_\infty \leq M$ pour toute $f \in \mathcal{F}$. La famille $\mathcal{F}$ est donc uniformément bornée.

Montrons qu'elle est équicontinue. Soit $\varepsilon > 0$. La compacité de $\overline{\mathcal{F}}$ permet de la recouvrir par un nombre fini de boules de rayon ε . Il existe donc $g_1, \dots, g_r \in C(K, \mathbb{R})$ telles que, pour toute $f \in \mathcal{F}$, il existe $j \in \{1, \dots, r\}$ vérifiant $\|f - g_j\|_\infty < \varepsilon$.

Chaque fonction g_j est uniformément continue sur le compact K . Comme les fonctions g_j sont en nombre fini, il existe $\delta > 0$ tel que, si $d(x, y) < \delta$ alors $|g_j(x) - g_j(y)| < \varepsilon$ pour tout $j \in \{1, \dots, r\}$.

Soit $f \in \mathcal{F}$, et choisissons j tel que $\|f - g_j\|_\infty < \varepsilon$. Si $d(x, y) < \delta$, alors

$$|f(x) - f(y)| \leq |f(x) - g_j(x)| + |g_j(x) - g_j(y)| + |g_j(y) - f(y)| \leq 3\varepsilon$$

La famille $\mathcal{F}$ est donc équicontinue. ■

La façon la plus simple d'obtenir une famille équicontinue est d'obtenir une famille équi-lipschitzienne, c'est-à-dire que toutes les fonctions de $\mathcal{F}$ sont M -lipschitziennes, $|f(x) - f(y)| \leq Md(x, y)$. Alors si $\varepsilon > 0$, en prenant $\eta = \varepsilon/M$, si $d(x, y) < \eta$, $|f(x) - f(y)| < \varepsilon$. Voici un exemple :

Exemple 2.6

Soient $a < b$ et $B = \{f \in C^1([a, b]); \|f\|_{C^1} \leq 1\}$, où, par exemple, $\|f\|_{C^1} = \|f\|_\infty + \|f'\|_\infty$. Alors B est relativement compacte dans $C([a, b])$, muni de la norme uniforme. En d'autres termes, de toute suite (f_n) de B , on peut extraire une sous-suite qui converge uniformément sur $[a, b]$.

Cela s'exprime parfois sous la forme : *L'injection canonique*

$$C^1([a, b]) \longrightarrow C([a, b]), \quad f \longmapsto f,$$

est un opérateur compact.

En effet, la famille B est uniformément bornée puisque, si $f \in B$, $\|f\|_\infty \leq \|f\|_{C^1} \leq 1$. La famille B est également équicontinue : si $f \in B$ alors $\|f'\|_\infty \leq \|f\|_{C^1} \leq 1$ donc, si $x, y \in [a, b]$, le théorème des accroissements finis implique que

$$|f(x) - f(y)| \leq \|f'\|_\infty |x - y| \leq |x - y|.$$

Ainsi, pour tout $\varepsilon > 0$, en prenant $\delta = \varepsilon$, si $x, y \in [a, b]$ sont tels que $|x - y| < \delta$ alors pour toute $f \in B$, $|f(x) - f(y)| < \varepsilon$.

Comme $[a, b]$ est compact, le théorème d'Ascoli–Arzelà montre que B est relativement compacte dans $C([a, b])$.

Notez que ce qu'on montre vraiment est qu'un ensemble borné dans $\mathcal{C}(K)$ de fonctions K -Lipschitziennes et relativement compacte. Cet ensemble est en fait compact puisqu'il est fermé : si $(f_n)_{n \geq 0}$ est une suite de fonctions K -Lipschitziennes qui converge simplement vers f alors f est encore K -Lipschitzienne. Il suffit en effet de passer à la limite dans l'inégalité $|f_k(x) - f_k(y)| \leq Kd(x, y)$ pour obtenir $|f(x) - f(y)| \leq Kd(x, y)$. La boule unité de $\mathcal{C}^1([a, b])$ n'est en revanche pas fermée dans $\mathcal{C}([a, b])$ puisqu'une limite uniforme de fonctions $\mathcal{C}^1$ peut ne pas être $\mathcal{C}^1$.

2.3. Lemme de Dini.

Théorème 2.7 (Lemme de Dini)

Soit K un espace compact et (f_n) une suite de fonctions continues de K dans $\mathbb{R}$.

On suppose que

- $f_n(x) \rightarrow f(x)$ pour tout $x \in K$,
- la limite f est continue,
- la suite (f_n) est monotone, c'est-à-dire que $f_n \leq f_{n+1}$ pour tout n (ou bien $f_n \geq f_{n+1}$ pour tout n).

Alors la convergence de (f_n) vers f est uniforme sur K .

Ce théorème est particulièrement intéressant car il montre que la monotonie remplace l'équi-continuité dans ce contexte. Il peut être comparé au théorème d'Ascoli–Arzelà :

- Ascoli–Arzelà : équicontinuité + compacité de l'ensemble de départ $\Rightarrow$ convergence uniforme (après extraction).

— Dini : monotonie + continuité de la limite + compacité de l'ensemble de départ $\Rightarrow$ convergence uniforme de toute la suite.

Un exemple d'utilisation du lemme de Dini est donné dans la première étape de la démonstration du théorème de Stone-Weierstrass 3.14.

DÉMONSTRATION. On traite le cas où (f_n) est croissante, le cas décroissant étant analogue.

Posons $g_n = f - f_n$. Les fonctions g_n sont continues sur K , positives, et $g_{n+1} \leq g_n$. De plus, $g_n(x) \rightarrow 0$ pour tout $x \in K$.

Supposons par l'absurde que la convergence ne soit pas uniforme. Alors il existe $\varepsilon > 0$ tel que, pour tout n , on puisse choisir $x_n \in K$ vérifiant $g_n(x_n) \geq \varepsilon$. Comme K est compact, quitte à extraire une sous-suite, on peut supposer $x_n \rightarrow x \in K$.

Fixons maintenant un entier m . Pour tout $n \geq m$, la décroissance de (g_n) donne $g_m(x_n) \geq g_n(x_n) \geq \varepsilon$. En faisant tendre n vers l'infini et en utilisant la continuité de g_m , on obtient $g_m(x) \geq \varepsilon$. Cette inégalité est vraie pour tout m . Or $g_m(x) \rightarrow 0$, cela est donc impossible.

Ainsi, la convergence est nécessairement uniforme. ■

Comme les sommes partielles d'une série à termes positifs sont croissantes, on en déduit immédiatement la version suivante du lemme de Dini pour les séries :

Corollaire 2.8 (Lemme de Dini pour les séries à termes positifs)

Soit K un espace compact et $(f_n)_{n \geq 0}$ une suite de fonctions *positives* de K dans $\mathbb{R}$. On suppose que

— pour tout $x \in K$, la série $\sum_{n=0}^{+\infty} f_n(x)$ converge,

— la somme $\sum_{n=0}^{+\infty} f_n$ est continue,

Alors la série $\sum_{n=0}^{+\infty} f_n$ converge uniformément sur K .

Exemple 2.9

Notons que la continuité de la limite est essentielle : la série à termes positifs $\sum_{n=0}^{+\infty} (1-x)x^n$ converge simplement sur $[0,1]$ vers la fonction f donnée par

$$f(x) = \begin{cases} 1 & \text{si } 0 \leq x < 1 \\ 0 & \text{si } x = 1 \end{cases} \quad (\text{c'est une série géométrique})$$
 mais, comme f n'est pas continue alors que $x \rightarrow (1-x)x^n$ l'est, la convergence ne peut être uniforme.

Exemple 2.10

Le lemme de Dini peut également être faux lorsqu'on considère des fonctions sur autre chose qu'un compact. Par exemple, si on définit f_n par $f_n(x) = 1 - e^{-x/n}$ alors $f_n \rightarrow 0$ simplement sur $[0, +\infty)$ qui est continue et $f_{n+1} \leq f_n$. Pourtant la convergence n'est pas uniforme sur $[0, +\infty)$ puisque $\sup_{x \in [0, +\infty)} |f_n(x)| = 1$.

2.4. Applications en probabilités : convergence des fonctions de répartition.

Le résultat suivant remplace la croissance ponctuelle par la croissance des fonctions.

Théorème 2.11 (Second lemme de Dini)

Soit $(f_n)_{n \geq 1}$ une suite de fonctions *croissantes* sur un intervalle borné $[a, b]$. On suppose que (f_n) converge simplement sur $[a, b]$ vers une fonction continue f . Alors la convergence de (f_n) vers f est uniforme sur $[a, b]$.

Notez que les f_n n'ont pas besoin d'être continus, seule f doit l'être.

DÉMONSTRATION. Comme f est la limite simple de fonctions croissantes, elle est elle-même croissante. De plus, puisqu'elle est continue sur le compact $[a, b]$, elle est uniformément continue. Soit alors $\varepsilon > 0$ et $a = x_0 < x_1 < \dots < x_p = b$ tels que, pour tout $i \in \{1, \dots, p\}$, $f(x_i) - f(x_{i-1}) < \varepsilon$.

Comme $f_n(x_i) \rightarrow f(x_i)$ pour tout $i \in \{0, \dots, p\}$, et comme il n'y a qu'un nombre fini de points x_i , il existe $N \in \mathbb{N}$ tel que, pour tout $n \geq N$ et tout $i \in \{0, \dots, p\}$, $|f_n(x_i) - f(x_i)| < \varepsilon$.

Fixons $n \geq N$ et $x \in [a, b]$. Il existe $i \in \{1, \dots, p\}$ tel que $x_{i-1} \leq x \leq x_i$. Puisque f_n et f sont croissantes, on a $f_n(x_{i-1}) \leq f_n(x) \leq f_n(x_i)$ et $f(x_{i-1}) \leq f(x) \leq f(x_i)$. On en déduit

$$f_n(x) - f(x) \leq f_n(x_i) - f(x_{i-1}) = f_n(x_i) - f(x_i) + f(x_i) - f(x_{i-1}) < 2\varepsilon.$$

De même,

$$f(x) - f_n(x) \leq f(x_i) - f_n(x_{i-1}) = f(x_i) - f(x_{i-1}) + f(x_{i-1}) - f_n(x_{i-1}) < 2\varepsilon.$$

Ainsi, $|f_n(x) - f(x)| < 2\varepsilon$ pour tout $x \in [a, b]$ et tout $n \geq N$. Par conséquent, $\|f_n - f\|_\infty \leq 2\varepsilon$ pour tout $n \geq N$, ce qui démontre que $f_n \rightarrow f$ uniformément sur $[a, b]$. ■

Rappelons qu'une fonction de répartition est une fonction définie sur $\mathbb{R}$, continue à droite, croissante telle que $\lim_{x \rightarrow -\infty} F(x) = 0$, $\lim_{x \rightarrow +\infty} F(x) = 1$.

Théorème 2.12 (Théorème de Pólya)

Soit (F_n) une suite de fonctions de répartition et soit F une fonction de répartition *continue* sur $\mathbb{R}$. On suppose que $F_n(x) \rightarrow F(x)$ pour tout $x \in \mathbb{R}$. Alors la convergence est uniforme sur $\mathbb{R}$, c'est-à-dire

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0.$$

DÉMONSTRATION. Soit $\varepsilon > 0$. Comme F est une fonction de répartition, il existe donc $A < B$ tels que $F(A) < \varepsilon$ et $1 - F(B) < \varepsilon$.

Sur le segment compact $[A, B]$, les fonctions F_n sont croissantes et convergent simplement vers la fonction continue F . D'après le second lemme de Dini, la convergence est uniforme sur $[A, B]$. Il existe donc N_0 tel que, pour tout $n \geq N_0$,

$$\sup_{x \in [A, B]} |F_n(x) - F(x)| < \varepsilon.$$

Il reste à contrôler les deux intervalles $(-\infty, A]$ et $[B, +\infty)$. Comme $F_n(A) \rightarrow F(A)$, il existe $N_1 \geq N_0$ tel que, pour tout $n \geq N_1$, $|F_n(A) - F(A)| < \varepsilon$. Par conséquent, $F_n(A) < F(A) + \varepsilon < 2\varepsilon$. Par ailleurs, comme F_n et F sont croissantes sur $(-\infty, A]$, pour $x \leq A$, $0 \leq F_n(x) \leq F_n(A)$ et $0 \leq F(x) \leq F(A)$. Ainsi,

$$|F_n(x) - F(x)| \leq F_n(x) + F(x) \leq F_n(A) + F(A) < 3\varepsilon.$$

De même, comme $F_n(B) \rightarrow F(B)$, il existe $N_2 \geq N_1$ tel que, pour tout $n \geq N_2$, $|F_n(B) - F(B)| < \varepsilon$. Puisque $1 - F(B) < \varepsilon$, on obtient $1 - F_n(B) \leq 1 - F(B) + |F_n(B) - F(B)| < 2\varepsilon$. Mais alors, pour $x \geq B$, $F_n(B) \leq F_n(x) \leq 1$ et $F(B) \leq F(x) \leq 1$ et, par

conséquent,

$$|F_n(x) - F(x)| = |F_n(x) - 1 + 1 - F(x)| \leq 1 - F_n(x) + 1 - F(x) \leq 1 - F_n(B) + 1 - F(B) < 3\varepsilon.$$

Au final, si $n \geq N_2$, pour tout $x \in \mathbb{R}$, $|F_n(x) - F(x)| < 3\varepsilon$ donc $F_n \rightarrow F$ uniformément sur $\mathbb{R}$. ■

Rappelons que si $X_n \rightarrow X$ en loi et si la fonction de répartition F_X de X est continue alors les fonctions de répartition F_{X_n} de X_n convergent simplement vers F_X . Le théorème de Pólya nous dit que cette convergence est en fait uniforme sur $\mathbb{R}$.

Le théorème de Helly a lui aussi une application en probabilités.

Definition 2.13 (Suite de variables aléatoires tendue)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires réelles. On dit que (X_n) est tendue si, pour tout $\varepsilon > 0$, il existe M tel que $\sup_n \mathbb{P}(|X_n| \geq M) \leq \varepsilon$.

Théorème 2.14 (Prohorov)

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires.

- (i) Si $(X_n) \xrightarrow{\mathcal{L}} X$ alors (X_n) est tendue.
- (ii) Si (X_n) est tendue, alors il existe une sous-suite $(X_{n_k})_k$ qui converge en loi.

Remarque 2.15

On peut évidemment également écrire ce théorème en termes de mesures de probabilités. Si $(\mu_n)_{n \geq 0}$ est une suite de mesures de probabilités sur $\mathbb{R}$ telles que, pour tout $\varepsilon > 0$, il existe M tel que $\sup_n \mu_n(\mathbb{R} \setminus [-M, M]) \leq \varepsilon$ alors il existe une mesure de probabilités μ sur $\mathbb{R}$ et une sous-suite telle que, pour toute $\varphi \in \mathcal{C}_b(\mathbb{R})$,

$$\int_{\mathbb{R}} \varphi(x) d\mu_n(x) \rightarrow \int_{\mathbb{R}} \varphi(x) d\mu(x).$$

C'est par exemple le cas dès que les mesures μ sont à support dans un compact fixe $[A, B]$. Dans ce cas, en prenant φ supportée en-dehors de $[A, B]$, on obtient $\int_{\mathbb{R}} \varphi(x) d\mu(x) = 0$ et ainsi μ est également à support dans $[A, B]$. Ensuite, on n'a plus besoin de supposer que φ soit bornée puisqu'en prenant χ une fonction continue à support compact telle que $\chi = 1$ sur $[A, b]$,

$$\int_{\mathbb{R}} \varphi(x) d\mu_n(x) = \int_{\mathbb{R}} \varphi(x)\chi(x) d\mu_n(x) \rightarrow \int_{\mathbb{R}} \varphi(x)\chi(x) d\mu(x) = \int_{\mathbb{R}} \varphi(x) d\mu(x).$$

DÉMONSTRATION. Soit F_n (resp. F) la fonction de répartition de X_n (resp. F).

Tout d'abord, F est croissante, donc elle n'a qu'un nombre au plus dénombrable de points de discontinuité. De plus $F(x) \rightarrow 0$ quand $x \rightarrow -\infty$ et $F(x) \rightarrow 1$ quand $x \rightarrow +\infty$. Pour tout $\varepsilon > 0$, il existe alors R tel que $R, -R$ soient des points de continuité de F et tels que $F(-R) < \varepsilon$ $F(R) > 1 - \varepsilon$. La convergence en loi $X_n \rightarrow X$ et la continuité de F en $\pm R$ implique que $F_n(\pm R) \rightarrow F(\pm R)$. Il existe donc N tel que, si $n \geq N$, $F_n(-R) < 2\varepsilon$ et $F_n(R) > 1 - 2\varepsilon$. Mais alors, pour tout $M > R$ et $n > M$, $F(-M), F_n(-M) < 2\varepsilon$ et $F(M), F_n(M) > 1 - 2\varepsilon$. D'autre part, $\sup_{n \leq N} F_n(-M), \sup_{n \leq N} (1 - F_n(M)) \rightarrow 0$ quand $M \rightarrow +\infty$. Ainsi, on peut choisir M assez grand pour que $\mathbb{P}(|X_n| \geq M) \leq 4\varepsilon$ pour tout n . La suite est bien tendue.

Montrons maintenant l'autre direction. Les fonctions F_n sont croissantes et vérifient $0 \leq F_n(x) \leq 1$ pour tout $x \in \mathbb{R}$. Le théorème de sélection de Helly 2.1 fournit donc une sous-suite $(F_{n_k})_{k \geq 1}$ et une fonction croissante $F : \mathbb{R} \rightarrow [0, 1]$ telles que $F_{n_k}(x) \rightarrow F(x)$

en tout point de continuité de F . On peut en outre supposer que F est continue à droite (voir la remarque 2.2).

Ensuite, Comme la suite (X_n) est tendue, pour tout $\varepsilon > 0$, il existe R tel que $\sup_{n \geq 1} \mathbb{P}(|X_n| > R) < \varepsilon$. Considérons alors un point de continuité S de F tel que $S < -R$. On a

$$F_{n_k}(S) = \mathbb{P}(X_{n_k} \leq S) \leq \mathbb{P}(|X_{n_k}| > R) < \varepsilon.$$

En faisant tendre k vers $+\infty$, il vient $0 \leq F(S) \leq \varepsilon$. Mais comme F est monotone, pour tout $x < -S$, $F(x) < \varepsilon$ i.e. $F(x) \rightarrow 0$ quand $x \rightarrow -\infty$.

De même, considérons un point de continuité S de F tel que $S > R$. On a

$$F_{n_k}(S) = \mathbb{P}(X_{n_k} \leq S) \geq \mathbb{P}(|X_{n_k}| \leq R) \geq 1 - \varepsilon.$$

En passant à la limite, $F(S) \geq 1 - \varepsilon$ puis, pour tout $x > S$, $1 \geq F(x) \geq 1 - \varepsilon$ donc $F(x) \rightarrow 1$ quand $x \rightarrow +\infty$.

En résumé, F est une fonction croissante, continue à droite en tout point, telle que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$. Une telle fonction est la fonction de répartition d'une mesure de probabilité μ .

Enfin, $F_{n_k}(x) \rightarrow F(x)$ en tout point de continuité de F . Par le critère de convergence en loi en termes de fonctions de répartition, on obtient $\mathcal{L}(X_{n_k}) \rightarrow \mu$ i.e. la sous-suite (X_{n_k}) converge en loi. ■

Remarque 2.16

L'hypothèse de tension est essentielle. Par exemple, si $X_n = n$ presque sûrement, alors $F_n(x) \rightarrow 0$ pour tout $x \in \mathbb{R}$, mais la fonction limite identiquement nulle n'est pas une fonction de répartition (la masse s'échappe vers $+\infty$).

2.5. Applications en probabilités : distance de Wasserstein.

Definition 2.17 (Distance de Wasserstein)

Soit (K, d) un espace métrique compact et soient μ et ν deux mesures de probabilité boréliennes sur K . On note $\Pi(\mu, \nu)$ l'ensemble des mesures de probabilité π sur $K \times K$ dont les marginales sont respectivement μ et ν , c'est-à-dire telles que

$$\pi(A \times K) = \mu(A) \quad \text{et} \quad \pi(K \times A) = \nu(A)$$

pour tout borélien $A \subset K$.

La distance de Wasserstein d'ordre 1 entre μ et ν est définie par

$$W_1(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \int_{K \times K} d(x, y) d\pi(x, y).$$

Une mesure $\pi \in \Pi(\mu, \nu)$ est appelée un *plan de transport*. La quantité

$$\int_{K \times K} d(x, y) d\pi(x, y)$$

représente le coût moyen du transport de la masse distribuée selon μ vers la masse distribuée selon ν , lorsque le coût du déplacement de x vers y est égal à $d(x, y)$.

Théorème 2.18 (Dualité de Kantorovich–Rubinstein)

Soit (K, d) un espace métrique compact et soient μ, ν deux mesures de probabilité boréliennes sur K . Alors

$$W_1(\mu, \nu) = \sup_{\text{Lip}(f) \leq 1} \left(\int_K f d\mu - \int_K f d\nu \right),$$

où le supremum est pris sur l'ensemble des fonctions réelles 1-lipschitziennes sur K .

La formule précédente donne une description *duale* de la distance de Wasserstein. La définition initiale consiste à minimiser un coût sur l'ensemble des plans de transport, tandis que la formule duale consiste à maximiser une forme linéaire sur une famille de fonctions lipschitziennes.

L'inégalité la plus immédiate de la dualité s'obtient de la manière suivante. Soit $f : K \rightarrow \mathbb{R}$ une fonction 1-lipschitzienne et soit $\pi \in \Pi(\mu, \nu)$. Comme les marges de π sont μ et ν , on a

$$\int_K f d\mu - \int_K f d\nu = \int_{K \times K} (f(x) - f(y)) d\pi(x, y).$$

Or $f(x) - f(y) \leq |f(x) - f(y)| \leq d(x, y)$, donc

$$\int_K f d\mu - \int_K f d\nu \leq \int_{K \times K} d(x, y) d\pi(x, y).$$

En prenant l'infimum sur $\pi \in \Pi(\mu, \nu)$, on obtient

$$\int_K f d\mu - \int_K f d\nu \leq W_1(\mu, \nu).$$

Puis, en prenant le supremum sur les fonctions 1-lipschitziennes,

$$\sup_{\text{Lip}(f) \leq 1} \left(\int_K f d\mu - \int_K f d\nu \right) \leq W_1(\mu, \nu).$$

L'inégalité réciproque constitue la partie profonde du théorème de Kantorovich–Rubinstein. Elle est admise ici. L'objectif de cette section est de montrer que ce supéum est atteint :

Proposition 2.19 (Existence d'un potentiel optimal)

Sous les hypothèses précédentes, il existe une fonction $f : K \rightarrow \mathbb{R}$, 1-lipschitzienne, telle que

$$W_1(\mu, \nu) = \int_K f d\mu - \int_K f d\nu.$$

Une telle fonction est appelée un *potentiel de Kantorovich optimal*.

DÉMONSTRATION DE LA PROPOSITION. Fixons un point $x_0 \in K$. Comme μ et ν sont deux mesures de probabilité, elles ont même masse totale. Par conséquent, pour toute constante $c \in \mathbb{R}$,

$$\int_K (f + c) d\mu - \int_K (f + c) d\nu = \int_K f d\mu - \int_K f d\nu.$$

On peut donc normaliser les fonctions admissibles en imposant $f(x_0) = 0$.

Considérons l'ensemble

$$\mathcal{F} = \{f \in \mathcal{C}(K) : \text{Lip}(f) \leq 1 \text{ et } f(x_0) = 0\}.$$

On a donc

$$\sup_{\text{Lip}(f) \leq 1} \left(\int_K f d\mu - \int_K f d\nu \right) = \sup_{f \in \mathcal{F}} \left(\int_K f d\mu - \int_K f d\nu \right).$$

Notre objectif est donc de montrer que ce suprémum est atteint. Pour cela, nous allons montrer que $\mathcal{F}$ est compact dans $\mathcal{C}(K)$ muni de la norme uniforme. Par ailleurs, la forme linéaire

$$\Phi(f) = \int_K f d\mu - \int_K f d\nu$$

est continue pour la norme uniforme : pour tous $f, g \in \mathcal{C}(K)$,

$$|\Phi(f) - \Phi(g)| \leq \int_K |f - g| d\mu + \int_K |f - g| d\nu \leq 2\|f - g\|_\infty.$$

Une fois qu'on aura montré que $\mathcal{F}$ est compact, la fonction continue Φ atteindra alors son maximum sur ce compact. Il existe ainsi $f_* \in \mathcal{F}$ tel que

$$\Phi(f_*) = \max_{f \in \mathcal{F}} \Phi(f).$$

Il ne nous reste donc plus qu'à montrer que $\mathcal{F}$ est compact dans $\mathcal{C}(K)$. Tout d'abord, cette famille est équicontinue. En effet, pour tous $x, y \in K$ et toute fonction $f \in \mathcal{F}$, $|f(x) - f(y)| \leq d(x, y)$. Elle est donc équi-lipschitzienne donc équicontinue.

Elle est également uniformément bornée. Pour tout $x \in K$ et tout $f \in \mathcal{F}$, on a

$$|f(x)| = |f(x) - f(x_0)| \leq d(x, x_0) \leq \text{diam}(K).$$

D'après le théorème d'Ascoli–Arzelà, l'ensemble $\mathcal{F}$ est donc relativement compact dans $\mathcal{C}(K)$.

Montrons qu'il est fermé. Soit (f_n) une suite d'éléments de $\mathcal{F}$ convergeant uniformément vers une fonction $f \in \mathcal{C}(K)$. On a $f(x_0) = \lim_{n \rightarrow \infty} f_n(x_0) = 0$. De plus, pour tous $x, y \in K$,

$$|f(x) - f(y)| = \lim_{n \rightarrow \infty} |f_n(x) - f_n(y)| \leq d(x, y).$$

Ainsi, f est 1-lipschitzienne et appartient à $\mathcal{F}$. L'ensemble $\mathcal{F}$ est donc fermé. Étant relativement compact et fermé dans $\mathcal{C}(K)$, il est compact. ■

Remarque 2.20

La démonstration de l'existence du potentiel optimal n'établit pas à elle seule la dualité de Kantorovich–Rubinstein. Elle montre que, dès lors que la formule duale est connue, le supremum qui y apparaît est en réalité un maximum.

L'intervention d'Ascoli–Arzelà permet ainsi de transformer

$$\sup_{\text{Lip}(f) \leq 1} \left(\int_K f d\mu - \int_K f d\nu \right)$$

en une quantité effectivement atteinte par une fonction 1-lipschitzienne.

Exemple 2.21

Prenons $K \subset \mathbb{R}$ et deux mesures de Dirac $\mu = \delta_x$ et $\nu = \delta_y$. Il n'existe qu'un seul plan de transport entre δ_x et δ_y , à savoir $\pi = \delta_{(x,y)}$. Par conséquent, $W_1(\delta_x, \delta_y) = |x - y|$. La formule duale donne également

$$W_1(\delta_x, \delta_y) = \sup_{\text{Lip}(f) \leq 1} (f(x) - f(y)).$$

Comme f est 1-lipschitzienne, $f(x) - f(y) \leq |x - y|$. Cette borne est atteinte, par exemple, par la fonction $f(t) = |t - y|$, pour laquelle $f(x) - f(y) = |x - y|$. Ainsi,

$$W_1(\delta_x, \delta_y) = |x - y|.$$

Remarque 2.22

Dans le cas général d'un espace métrique non compact, la formule de Kantorovich–Rubinstein reste valable pour des mesures de probabilité ayant un moment d'ordre

1 fini, c'est-à-dire vérifiant, pour un point x_0 fixé,

$$\int_K d(x, x_0) d\mu(x) < +\infty, \quad \int_K d(x, x_0) d\nu(x) < +\infty.$$

La démonstration de l'existence d'un potentiel optimal demande alors davantage de précautions, car la famille des fonctions lipschitziennes normalisées n'est plus nécessairement uniformément bornée sur tout l'espace.

2.6. Application en géométrie : existence de géodésiques.

Definition 2.23 (Longueur d'une courbe)

Soit (X, d) un espace métrique. Soit $\gamma : [0, T] \rightarrow X$ une courbe continue. Sa longueur est définie par

$$L(\gamma) = \sup_{\mathcal{P}} \sum_{j=1}^m d(\gamma(t_{j-1}), \gamma(t_j)),$$

où le supremum porte sur toutes les subdivisions $\mathcal{P} : 0 = t_0 < t_1 < \dots < t_m = T$. La courbe γ est dite rectifiable lorsque $L(\gamma) < +\infty$.

Lemme 2.24 (Reparamétrisation des courbes rectifiables)

Soit $\gamma : [0, 1] \rightarrow X$ une courbe rectifiable. Alors il existe $\varphi : [0, 1] \rightarrow [0, 1]$ continue telle que $\hat{\gamma} = \gamma \circ \varphi$ est encore une courbe rectifiable avec $L(\hat{\gamma}) = L(\gamma)$ et de plus

$$d(\hat{\gamma}(t), \hat{\gamma}(s)) \leq L(\gamma)|t - s|.$$

DÉMONSTRATION. La deuxième assertion entraîne immédiatement que $L(\tilde{\gamma}) = L(\gamma)$, il suffit donc de construire φ . Pour cela, on considère γ_t la restriction de γ à $[0, t]$ et on remarque que γ_t est encore rectifiable (il suffit de compléter la partition de $[0, t]$ en ajoutant 1). De plus $\ell(t) = L(\gamma_t)$ est une fonction croissante, $\ell(0) = L(\gamma_0) = 0$, $\ell(1) = L(\gamma)$.

Remarquons ensuite que si $\ell(s_1) = \ell(s_2)$ alors $L(\gamma_{s_1}) - L(\gamma_{s_2}) = 0$ donc γ est constante sur $[s_1, s_2]$. On peut donc définir sans ambiguïté $\tilde{\gamma} : [0, L(\gamma)] \rightarrow X$ par $\tilde{\gamma}(v) = \gamma(t)$ si $\ell(t) = v$. Mais alors si $u, v \in [0, L(\gamma)]$ et s, t sont tels que $\tilde{\gamma}(u) = \gamma(s)$, $\tilde{\gamma}(v) = \gamma(t)$

$$d(\tilde{\gamma}(u), \tilde{\gamma}(v)) = d(\gamma(s), \gamma(t)) \leq L(\gamma|_{[s, t]}) = |\ell(t) - \ell(s)| = |v - u|.$$

Par suite $\tilde{\gamma}$, qu'on appelle *la paramétrisation par longueur d'arc*, est une fonction 1-lipschitzienne. On peut se ramener à $[0, 1]$ en posant $\hat{\gamma}(t) = \tilde{\gamma}(L(\gamma)t)$. ■

Théorème 2.25 (Existence d'une courbe de longueur minimale)

Soit (X, d) un espace métrique compact et soient $x, y \in X$. On suppose qu'il existe au moins une courbe rectifiable joignant x à y . Posons

$$\ell = \inf \{L(\gamma) : \gamma \text{ est une courbe rectifiable joignant } x \text{ à } y\}.$$

Alors il existe une courbe rectifiable $\gamma : [0, 1] \rightarrow X$ telle que $\gamma(0) = x$, $\gamma(1) = y$, et $L(\gamma) = \ell$.

DÉMONSTRATION. Par définition de ℓ , il existe une suite de courbes rectifiables $\gamma_n : [0, 1] \rightarrow X$ joignant x à y et telle que $L(\gamma_n) \rightarrow \ell$. Quitte à éliminer un nombre fini de termes de la suite (γ_n) , on peut supposer que $L(\gamma_n) \leq \ell + 1$ pour tout n .

On reparamètre chaque courbe γ_n proportionnellement à sa longueur et donc supposer que, pour tous $s, t \in [0, 1]$,

$$d(\gamma_n(s), \gamma_n(t)) \leq L(\gamma_n)|t - s|.$$

Ainsi,

$$d(\gamma_n(s), \gamma_n(t)) \leq (\ell + 1)|t - s|.$$

La famille (γ_n) est donc équilipschitzienne, et en particulier équicontinue.

Comme toutes les fonctions γ_n prennent leurs valeurs dans l'espace compact X , le théorème d'Ascoli–Arzelà fournit une sous-suite, encore notée (γ_n) , qui converge uniformément vers une application continue $\gamma : [0, 1] \rightarrow X$.

La convergence uniforme implique $\gamma(0) = \lim_{n \rightarrow \infty} \gamma_n(0) = x$ et de même $\gamma(1) = y$. La courbe γ joint donc x à y .

Montrons maintenant que $L(\gamma) \leq \ell$. Fixons une subdivision $0 = t_0 < t_1 < \dots < t_m = 1$. Pour tout j , la convergence uniforme donne $\gamma_n(t_j) \rightarrow \gamma(t_j)$. Par continuité de la distance,

$$d(\gamma_n(t_{j-1}), \gamma_n(t_j)) \rightarrow d(\gamma(t_{j-1}), \gamma(t_j)).$$

Par conséquent,

$$\sum_{j=1}^m d(\gamma(t_{j-1}), \gamma(t_j)) = \lim_{n \rightarrow \infty} \sum_{j=1}^m d(\gamma_n(t_{j-1}), \gamma_n(t_j)).$$

Or, pour tout n ,

$$\sum_{j=1}^m d(\gamma_n(t_{j-1}), \gamma_n(t_j)) \leq L(\gamma_n).$$

On en déduit

$$\sum_{j=1}^m d(\gamma(t_{j-1}), \gamma(t_j)) \leq \liminf_{n \rightarrow \infty} L(\gamma_n) = \ell.$$

Comme cette inégalité vaut pour toute subdivision, on obtient $L(\gamma) \leq \ell$.

Mais γ est une courbe rectifiable joignant x à y . Par définition de ℓ , $L(\gamma) \geq \ell$ donc, au final $L(\gamma) = \ell$. ■

Remarque 2.26

Le point essentiel de la démonstration est la semi-continuité inférieure de la longueur pour la convergence uniforme :

$$L(\gamma) \leq \liminf_{n \rightarrow \infty} L(\gamma_n).$$

Le théorème d'Ascoli–Arzelà permet d'extraire une limite uniforme d'une suite minimisante de courbes.

Corollaire 2.27

Supposons de plus que (X, d) soit un espace de longueur, c'est-à-dire que, pour tous $x, y \in X$,

$$d(x, y) = \inf \{L(\gamma) : \gamma \text{ est une courbe rectifiable joignant } x \text{ à } y\}.$$

Alors, pour tous $x, y \in X$, il existe une courbe rectifiable $\gamma : [0, 1] \rightarrow X$ telle que $\gamma(0) = x$, $\gamma(1) = y$ et $L(\gamma) = d(x, y)$.

DÉMONSTRATION. Il suffit d'appliquer le théorème précédent, puisque l'infimum des longueurs des courbes joignant x à y est égal à $d(x, y)$. ■

Remarque 2.28

Après un paramétrage proportionnel à la longueur, une courbe minimisante γ vérifie $d(\gamma(s), \gamma(t)) = |t - s| d(x, y)$, pour $s, t \in [0, 1]$. Une telle courbe est appelée une *géodésique minimisante*.

2.7. Application en théorie des nombres.**Definition 2.29 (Fonction de Möbius)**

La fonction de Möbius $\mu : \mathbb{N}^* \rightarrow \{-1, 0, 1\}$ est définie par $\mu(1) = 1$, et, pour $n \geq 2$,

$$\mu(n) = \begin{cases} 0, & \text{si } n \text{ est divisible par le carré d'un nombre premier,} \\ (-1)^r, & \text{si } n \text{ est le produit de } r \text{ nombres premiers distincts.} \end{cases}$$

Lemme 2.30

Pour tout entier $n \geq 1$,

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

DÉMONSTRATION. Écrivons $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, où $p_1, \dots, p_r$ sont les facteurs premiers distincts de n . Les seuls diviseurs d de n pour lesquels $\mu(d) \neq 0$ sont les produits de certains des nombres premiers $p_1, \dots, p_r$. On obtient donc

$$\sum_{d|n} \mu(d) = \sum_{J \subset \{1, \dots, r\}} (-1)^{|J|} = (1 - 1)^r.$$

Cette quantité vaut 0 si $r \geq 1$, c'est-à-dire si $n \geq 2$, et vaut 1 si $n = 1$. ■

En appliquant ce lemme à l'entier (p, q) , on obtient :

Corollaire 2.31

Pour tous entiers positifs p et q ,

$$\mathbf{1}_{(p,q)=1} = \sum_{d|(p,q)} \mu(d),$$

où $\mathbf{1}_{(p,q)=1}$ vaut 1 si p et q sont premiers entre eux et 0 sinon.

Théorème 2.32 (Répartition uniforme des fractions de Farey)

Pour $N \geq 1$ et $x \in [0, 1]$, posons

$$A_N(x) = |\{(p, q) \in \mathbb{N}^* \times \mathbb{N}^* : 1 \leq q \leq N, \quad 1 \leq p \leq xq, \quad (p, q) = 1\}|$$

le nombre de fractions réduites $0 \leq \frac{p}{q} \leq x$ dont le dénominateur est $q \leq N$. Alors

$$A_N(x) = \frac{3}{\pi^2} x N^2 + O(N \log N), \text{ uniformément pour } x \in [0, 1].$$

En particulier,

$$\sup_{x \in [0, 1]} \left| \frac{A_N(x)}{N^2} - \frac{3}{\pi^2} x \right| \rightarrow 0.$$

DÉMONSTRATION. En utilisant l'identité

$$\mathbf{1}_{(p,q)=1} = \sum_{d|(p,q)} \mu(d),$$

on obtient

$$A_N(x) = \sum_{q \leq N} \sum_{1 \leq p \leq xq} \mathbf{1}_{(p,q)=1} = \sum_{q \leq N} \sum_{1 \leq p \leq xq} \sum_{\substack{d|p \\ d|q}} \mu(d).$$

Toutes les sommes étant finies, on peut changer leur ordre :

$$A_N(x) = \sum_{d \leq N} \mu(d) \sum_{\substack{q \leq N \\ d|q}} \sum_{\substack{1 \leq p \leq xq \\ d|p}} 1.$$

Écrivons $q = dm$, $p = dr$. Les conditions deviennent $m \leq \frac{N}{d}$, et $r \leq xm$. Par conséquent,

$$A_N(x) = \sum_{d \leq N} \mu(d) \sum_{m \leq N/d} \lfloor xm \rfloor.$$

Pour tout réel u , on a $\lfloor u \rfloor = u + O(1)$, où la constante implicite est absolue. Ainsi, pour tout entier $M \geq 1$,

$$\sum_{m \leq M} \lfloor xm \rfloor = x \sum_{m \leq M} m + O(M),$$

uniformément pour $x \in [0, 1]$. Or

$$\sum_{m=1}^M m = \frac{M(M+1)}{2} = \frac{M^2}{2} + O(M).$$

On en déduit

$$\sum_{m \leq M} \lfloor xm \rfloor = \frac{x}{2} M^2 + O(M),$$

uniformément pour $x \in [0, 1]$.

En prenant $M = \left\lfloor \frac{N}{d} \right\rfloor$, il vient

$$A_N(x) = \frac{x}{2} \sum_{d \leq N} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor^2 + O \left(\sum_{d \leq N} \left\lfloor \frac{N}{d} \right\rfloor \right).$$

D'une part,

$$\sum_{d \leq N} \left\lfloor \frac{N}{d} \right\rfloor \leq N \sum_{d \leq N} \frac{1}{d} = O(N \log N).$$

Ainsi,

$$A_N(x) = \frac{x}{2} \sum_{d \leq N} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor^2 + O(N \log N),$$

uniformément en x .

D'autre part, $\left\lfloor \frac{N}{d} \right\rfloor = \frac{N}{d} + O(1)$, donc

$$\left\lfloor \frac{N}{d} \right\rfloor^2 = \frac{N^2}{d^2} + O \left(\frac{N}{d} \right) + O(1).$$

Par conséquent,

$$\sum_{d \leq N} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor^2 = N^2 \sum_{d \leq N} \frac{\mu(d)}{d^2} + O \left(N \sum_{d \leq N} \frac{1}{d} \right) + O(N).$$

Il vient donc

$$\sum_{d \leq N} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor^2 = N^2 \sum_{d \leq N} \frac{\mu(d)}{d^2} + O(N \log N).$$

On utilise maintenant l'identité classique

$$\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}.$$

De plus,

$$\left| \sum_{d > N} \frac{\mu(d)}{d^2} \right| \leq \sum_{d > N} \frac{1}{d^2} = O\left(\frac{1}{N}\right).$$

Ainsi,

$$\sum_{d \leq N} \frac{\mu(d)}{d^2} = \frac{6}{\pi^2} + O\left(\frac{1}{N}\right).$$

On en déduit

$$\sum_{d \leq N} \mu(d) \left\lfloor \frac{N}{d} \right\rfloor^2 = \frac{6}{\pi^2} N^2 + O(N \log N).$$

Finalement,

$$A_N(x) = \frac{x}{2} \left(\frac{6}{\pi^2} N^2 + O(N \log N) \right) + O(N \log N),$$

uniformément pour $x \in [0, 1]$. Comme $0 \leq x \leq 1$, on obtient

$$A_N(x) = \frac{3}{\pi^2} x N^2 + O(N \log N),$$

uniformément en x .

En divisant par N^2 , il vient

$$\sup_{x \in [0, 1]} \left| \frac{A_N(x)}{N^2} - \frac{3}{\pi^2} x \right| = O\left(\frac{\log N}{N}\right),$$

ce qui achève la démonstration. ■

Definition 2.33 (Suite de Farey)

La suite de Farey d'ordre N , notée $\mathcal{F}_N$, est l'ensemble ordonné des fractions irréductibles

$$\mathcal{F}_N = \left\{ \frac{p}{q} \in [0, 1], \quad 0 \leq p \leq q \leq N \right\}.$$

Corollaire 2.34

On a

$$|\mathcal{F}_N| = 1 + \sum_{q=1}^N \varphi(q) = \frac{3}{\pi^2} N^2 + O(N \log N),$$

où φ désigne l'indicatrice d'Euler.

Si l'on pose

$$F_N(x) = \frac{\#\{r \in \mathcal{F}_N : r \leq x\}}{|\mathcal{F}_N|},$$

pour $x \in [0, 1]$, alors $F_N(x) \rightarrow x$ uniformément sur $[0, 1]$.

DÉMONSTRATION. Pour $x \in [0, 1]$, les fractions strictement positives de $\mathcal{F}_N$ qui sont inférieures ou égales à x sont exactement celles comptées par $A_N(x)$. En tenant compte de la fraction $0/1$, on a

$$|\{r \in \mathcal{F}_N : r \leq x\}| = 1 + A_N(x).$$

En particulier,

$$|\mathcal{F}_N| = 1 + A_N(1) = 1 + \sum_{q=1}^N \varphi(q).$$

D'après le théorème précédent, $A_N(x) = \frac{3}{\pi^2}xN^2 + O(N \log N)$ uniformément en x , et

$$|\mathcal{F}_N| = \frac{3}{\pi^2}N^2 + O(N \log N).$$

Par conséquent,

$$F_N(x) = \frac{\frac{3}{\pi^2}xN^2 + O(N \log N)}{\frac{3}{\pi^2}N^2 + O(N \log N)} + O\left(\frac{1}{N^2}\right),$$

uniformément en $x \in [0, 1]$. Il s'ensuit que

$$\sup_{x \in [0, 1]} |F_N(x) - x| = O\left(\frac{\log N}{N}\right) \rightarrow 0$$

comme annoncé. ■

Remarque 2.35

Le corollaire signifie que les fractions de Farey deviennent uniformément réparties dans $[0, 1]$. Les fonctions F_N sont des fonctions en escalier, croissantes, qui convergent uniformément vers la fonction identité, $F(x) = x$.

3. Densité dan/s $\mathcal{C}(K)$, K compact

3.1. Première méthode : par approximation de l'unité.

Théorème 3.1

Soient $-\infty < a < b < +\infty$ deux réels, alors les polynômes sont denses dans $\mathcal{C}([a, b])$. Ainsi, pour toute fonction f continue sur $[a, b]$, il existe une suite de polynômes P_n telle que $\sup_{t \in [a, b]} |f(t) - P_n(t)| \rightarrow 0$.

La première démonstration est due à Weierstrass *

DÉMONSTRATION. Nous commençons par rappeler que si $\gamma(t) = e^{-\pi t^2}$ alors $\gamma_\lambda(t) = \frac{1}{\lambda} \gamma\left(\frac{t}{\lambda}\right)$ est une approximation de l'unité donc que si $f \in \mathcal{C}_0(\mathbb{R})$, alors $f * \gamma_\lambda \rightarrow f$ uniformément quand $\lambda \rightarrow 0$.

Soit alors $f \in \mathcal{C}([a, b])$. Commençons par prolonger f en une fonction g de $\mathcal{C}_0(\mathbb{R})$ avec $\|g\|_\infty \leq \|f\|_\infty$ comme suit :

$$g(x) = \begin{cases} f(x) & \text{si } x \in [a, b] \\ f(a)(x - a + 1) & \text{si } x \in [a - 1, a] \\ f(b)(b + 1 - x) & \text{si } x \in [b, b + 1] \\ 0 & \text{sinon} \end{cases}.$$

*. Weierstrass, Über die analytische Darstellbarkeit sogenannter willkürlicher Functionen einer reellen Veränderlichen, Verl. d. Kgl. Akad. d. Wiss. Berlin **2**(1885) 633–639.

Ainsi $\gamma_\lambda * g \rightarrow g$ uniformément sur $\mathbb{R}$. Soit donc $\varepsilon > 0$ et λ tel que $\sup_{t \in \mathbb{R}} |\gamma_\lambda * g(t) - g(t)| \leq \varepsilon$. En particulier, pour $t \in [a, b]$, $|\gamma_\lambda * g(t) - f(t)| \leq \varepsilon$.

Nous allons maintenant approcher $\gamma_\lambda * g(t)$ par des polynômes. Pour cela, remarquons que

$$\gamma_\lambda * g(t) = \int_{a-1}^{b+1} \frac{1}{\lambda} g(s) e^{-\pi(s-t)^2/\lambda^2} ds.$$

Mais, si $t \in [a, b]$ et $s \in [a-1, b+1]$ alors $s-t \in [-(b-a)-1, (b-a)+1]$. Comme le rayon de convergence de la série exponentielle est infini, il existe N tel que, pour tout $u \in [-(b-a)-1, (b-a)+1]$,

$$\left| e^{-\pi u^2/\lambda^2} - \sum_{k=0}^N (-1)^k \frac{\pi^k}{\lambda^{2k} k!} u^{2k} \right| \leq \varepsilon.$$

Mais alors, pour $t \in [a, b]$

$$\begin{aligned} & \left| \int_{a-1}^{b+1} \frac{1}{\lambda} g(s) e^{-\pi(s-t)^2/\lambda^2} ds - \int_{a-1}^{b+1} \frac{1}{\lambda} g(s) \sum_{k=0}^N (-1)^k \frac{\pi^k}{\lambda^{2k} k!} (s-t)^{2k} ds \right| \\ & \leq \int_{a-1}^{b+1} \frac{1}{\lambda} |g(s)| \left| e^{-\pi(s-t)^2/\lambda^2} - \sum_{k=0}^N (-1)^k \frac{\pi^k}{\lambda^{2k} k!} (s-t)^{2k} \right| ds \\ & \leq \varepsilon \int_{a-1}^{b+1} \frac{1}{\lambda} |g(s)| ds \leq \frac{(b-a+2)\|f\|_\infty}{\lambda} \varepsilon. \end{aligned}$$

On conclue en remarquant que

$$\int_{a-1}^{b+1} \frac{1}{\lambda} g(s) \sum_{k=0}^N (-1)^k \frac{\pi^k}{\lambda^{2k} k!} (s-t)^{2k} ds$$

est un polynôme de degré $2N$ en t . ■

Notez que la convergence de $\gamma_\lambda * g$ vers g ne requière pas d'arguments de densité (on est dans $\mathcal{C}_0(\mathbb{R})$ ici, pas dans L^p).

Une démonstration un peu plus directe est possible.[†]

DÉMONSTRATION. On suppose sans perte de généralité que $a = \frac{1}{4}, b = \frac{3}{4}$ en remplaçant $f(x)$ par $f(\frac{1}{4} + \frac{1}{2(b-a)}(x-a))$ puis on prolonge f en une fonction continue sur $\mathbb{R}$, à support $[0, 1]$ et qui soit affine sur $[0, 1/4]$ et sur $[3/4, 1]$. On suppose aussi que $f \neq 0$.

On pose alors $J_n = \int_{-1}^1 (1-u^2)^n du$ et

$$P_n(x) = \frac{1}{J_n} \int_0^1 f(t) (1-(t-x)^2)^n dt.$$

En développant $(1-(x-t)^2)^n$ on voit immédiatement que P_n est un polynôme de degré $2n$. Par ailleurs on remarque que si $0 \leq x \leq 1$ alors $f = 0$ sur $[-1+x, 0]$ et sur $[1, 1+x]$ donc

$$P_n(x) = \frac{1}{J_n} \int_{-1+x}^{1+x} f(t) (1-(t-x)^2)^n dt = \frac{1}{J_n} \int_{-1}^1 f(t+x) (1-t^2)^n dt$$

avec un changement de variable. Il en résulte que

$$P_n(x) - f(x) = \frac{1}{J_n} \int_{-1}^1 (f(t+x) - f(x)) (1-t^2)^n dt.$$

[†] D. Jackson *A Proof of Weierstrass's Theorem* The American Mathematical Monthly, **41** (1934), 309-312 qui simplifie légèrement E. Landau, *Über die Approximation einer stetigen Funktion durch eine ganze rationale Funktion*, Rendiconti del Circolo Matematico di Palermo, vol. 25 (1908), pp. 337-345.

Soit maintenant $\varepsilon > 0$. Comme f est uniformément continue sur $\mathbb{R}$, il existe $\eta > 0$ tel que, si $|t| \leq \eta$, $|f(t+x) - f(x)| \leq \varepsilon$ pour tout $x \in \mathbb{R}$.

Par ailleurs, $|f(t+x) - f(x)| \leq 2\|f\|_\infty$ donc, pour $|t| \geq \eta$, $1 \leq t^2/\eta^2$ et alors $|f(t+x) - f(x)| \leq 2\|f\|_\infty t^2/\eta^2$. En additionnant les deux estimation, on obtient

$$|f(t+x) - f(x)| \leq \varepsilon + \frac{2\|f\|_\infty}{\eta^2} t^2.$$

Il en résulte que

$$\begin{aligned} |P_n(x) - f(x)| &= \frac{1}{J_n} \int_{-1}^1 \left(\varepsilon + \frac{2\|f\|_\infty}{\eta^2} t^2 \right) (1-t^2)^n dt \\ &= \varepsilon + \frac{2\|f\|_\infty}{\eta^2} \frac{1}{J_n} \int_{-1}^1 t^2 (1-t^2)^n dt. \end{aligned}$$

Notons $K_n = \int_{-1}^1 t^2 (1-t^2)^n dt$ et intégrons par parties

$$K_n = \frac{1}{2} \int_{-1}^1 t(1-t^2)^n 2t dt = \frac{1}{2(n+1)} \int_{-1}^1 (1-t^2)^{n+1} dt = \frac{J_{n+1}}{2(n+1)}.$$

Enfin, pour $t \in [-1, 1]$, $(1-t^2)^{n+1} \leq (1-t^2)^n$ donc $J_{n+1} \leq J_n$ donc $\frac{K_n}{J_n} \leq \frac{1}{2(n+1)}$. Il en résulte que

$$|P_n(x) - f(x)| \leq \varepsilon + \frac{\|f\|_\infty}{\eta^2} \frac{1}{n+1} \leq 2\varepsilon$$

pour peu que $n+1 \geq \frac{\|f\|_\infty}{\eta^2 \varepsilon}$. ■

Nous laissons en exercice de montrer qu'on peut démontrer par la même méthode que

$$\frac{1}{H_n} \int_{-\pi}^{\pi} f(t) \cos^{2n} \frac{t-x}{2} dt, \quad H_n = \int_{-\pi}^{\pi} \cos^{2n} \frac{t-x}{2} dt$$

est une suite de polynômes trigonométriques qui converge uniformément vers f .

3.2. Seconde méthode : les polynômes de Bernstein.

Definition 3.2 (Polynômes de Bernstein)

Soit f une fonction continue sur $[0, 1]$, le *polynôme de Bernstein* associé à f est

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

où les $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ sont les coefficients binomiaux.

Théorème 3.3 (Bernstein)

Pour $f \in \mathcal{C}([0, 1])$, $B_n f \rightarrow f$ uniformément.

DÉMONSTRATION. Pour commencer, il convient de calculer $B_n f$ quand $f_0(t) = 1$, $f_1(t) = t$ et $f_2(t) = t^2$:

$$B_n f_0(x) = \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} = (x+1-x)^n = 1$$

avec la formule du binôme de Newton. On écrit alors

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

et en dérivant par rapport à x

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

donc

$$\sum_{k=0}^n \binom{n}{k} \frac{k}{n} x^k y^{n-k} = x(x + y)^{n-1}.$$

En prenant $y = 1 - x$ on trouve $B_n f_1(x) = x$ (pour $n \geq 1$ car $B_0 f_1 = 0$). Pour obtenir $B_n f_2$ on redérive et on obtient, pour $n \geq 2$

$$\sum_{k=0}^n \binom{n}{k} \frac{k^2}{n^2} x^k y^{n-k} = \frac{x}{n} \frac{\partial}{\partial x} [x(x + y)^{n-1}] = \frac{x}{n} ((x + y)^{n-1} + (n-1)x(x + y)^{n-2}).$$

En prenant $y = 1 - x$ on trouve

$$B_n f_2(x) = \frac{x}{n} (1 + (n-1)x) = x^2 + \frac{1}{n} x(1-x).$$

On constate que

$$(3.1) \quad B_n f_2(x) - f_2(x) = \frac{1}{n} x(1-x)$$

en particulier $B_n f_2 \rightarrow f_2$ uniformément sur $[0, 1]$.

Soit maintenant $f \in \mathcal{C}([0, 1])$. On utilise le fait que $B_n f_0 = 1$ pour écrire

$$B_n f(x) - f(x) = B_n f(x) - f(x) B_n f_0(x) = \sum_{k=0}^n \left[f\left(\frac{k}{n}\right) - f(x) \right] \binom{n}{k} x^k (1-x)^{n-k}$$

donc pour $x \in [0, 1]$,

$$(3.2) \quad |B_n f(x) - f(x)| \leq \sum_{k=0}^n \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k}.$$

Soit $\varepsilon > 0$. Il existe $\eta > 0$ tel que, si $|y - x| \leq \eta$ alors $|f(y) - f(x)| \leq \varepsilon$. (Remarquons pour la suite que η ne dépend pas de x). Soit

$$I = \{k \in \{0, \dots, n\} : |\frac{k}{n} - x| \leq \eta\} \quad \text{et} \quad J = \{0, \dots, n\} \setminus I.$$

On a

$$(3.3) \quad \begin{aligned} \sum_{k \in I} \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} &\leq \varepsilon \sum_{k \in I} \binom{n}{k} x^k (1-x)^{n-k} \\ &\leq \varepsilon \sum_{k=0}^n \binom{n}{k} x^k (1-x)^{n-k} = \varepsilon. \end{aligned}$$

Par ailleurs, si $k \in J$, alors $|\frac{k}{n} - x| \geq \eta$ que nous allons écrire $1 \leq \frac{1}{\eta^2} (\frac{k}{n} - x)^2$. D'autre part, $|f(\frac{k}{n}) - f(x)| \leq 2\|f\|_\infty$. Il en résulte que

$$\begin{aligned}
 (3.4) \quad \sum_{k \in J} \left| f\left(\frac{k}{n}\right) - f(x) \right| \binom{n}{k} x^k (1-x)^{n-k} &\leq \sum_{k \in J} 2\|f\|_\infty \binom{n}{k} \frac{1}{\eta^2} \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} \\
 &\leq \frac{2\|f\|_\infty}{\eta^2} \sum_{k=0}^n \binom{n}{k} \left[\left(\frac{k}{n}\right)^2 - 2x\frac{k}{n} + x^2 \right] x^k (1-x)^{n-k} \\
 &= \frac{2\|f\|_\infty}{\eta^2} [B_n f_2(x) - 2x B_n f_1(x) + x^2 B_n f_0(x)] \\
 &= \frac{2\|f\|_\infty}{\eta^2 n} x(1-x) \leq \frac{\|f\|_\infty}{2\eta^2 n}
 \end{aligned}$$

avec (3.1). Il suffit alors de prendre N tel que $\frac{\|f\|_\infty}{2\eta^2 N} \leq \varepsilon$ et on obtient

$$|B_n f(x) - f(x)| \leq 2\varepsilon \quad \text{pour } n \geq N.$$

Notons que, comme η est indépendant de x , N aussi, la convergence est donc uniforme. ■

En analysant plus finement la démonstration, on peut en déduire le résultat suivant. On définit trois fonctions $[0, 1] \rightarrow \mathbb{R}$ par $e_0(x) = 1$, $e_1(x) = x$ et $e_2(x) = x^2$. La première étape consiste à montrer que $B_n e_j \rightarrow e_j$ uniformément. On remarque aussi que $|B_n f| \leq B_n |f|$. Une petite modification de la démonstration précédente nous donne le résultat suivant :

Corollaire 3.4 (Théorème de Bohman–Korovkin)

Soit $(T_n)_{n \geq 1}$ une suite d'applications linéaires $\mathcal{C}([0, 1]) \rightarrow \mathcal{C}([0, 1])$ telles que

- si $f \in \mathcal{C}([0, 1])$ est telle que $f \geq 0$, alors $T_n f \geq 0$ (souvent abrégé en $T_n \geq 0$) ;
- pour $j = 0, 1, 2$, $T_n e_j \rightarrow e_j$ uniformément sur $[0, 1]$.

Alors $T_n f \rightarrow f$ uniformément pour tout $f \in \mathcal{C}([0, 1])$.

DÉMONSTRATION. On commence par écrire, pour $f \in \mathcal{C}([0, 1])$,

$$\begin{aligned}
 |T_n f(x) - f(x)| &\leq |T_n f(x) - f(x) T_n e_0(x)| + |f(x) T_n e_0(x) - f(x) e_0(x)| \\
 &= |T_n f(x) - T_n [f(x) e_0](x)| + |f(x)| |T_n e_0(x) - e_0(x)| \\
 &\leq |T_n [f - f(x) e_0](x)| + \|f\|_\infty \|T_n e_0 - e_0\|_\infty.
 \end{aligned}$$

Comme $\|T_n e_0 - e_0\|_\infty \rightarrow 0$, il suffit donc de montrer que $\sup_{x \in [0, 1]} |T_n [f - f(x) e_0](x)| \rightarrow 0$.

Pour cela, comme $f \in \mathcal{C}([0, 1])$ est uniformément continue, pour tout $\varepsilon > 0$, il existe $\eta > 0$ tel que, si $|x - t| \leq \eta$, $|f(x) - f(t)| \leq \varepsilon$. On en déduit que, pour tout $t, x \in [0, 1]$,

$$|f(t) - f(x)| \leq \varepsilon + \frac{2\|f\|_\infty}{\eta^2} (t - x)^2.$$

Pour voir cela, on voit que chaque terme de droite major l'un des cas $|t - x| < \eta$ et $|t - x| \geq \eta$.

Nous allons maintenant fixer x et faire agir T_n dans la variable t . Cette inégalité peut s'écrire

$$|f - f(x) e_0| \leq \varepsilon e_0 + \frac{2\|f\|_\infty}{\eta^2} (e_2 - 2x e_1 + x^2 e_0).$$

La positivité et la linéarité de T_n implique que si $\varphi \leq \psi$ alors $T_n\varphi \leq T_n\psi$ d'où on déduit que $|T_n\varphi(t)| \leq T_n[|\varphi|](t)$. En particulier

$$\begin{aligned} |T_n[f - f(x)e_0](t)| &\leq T_n[|f - f(x)e_0|](t) \\ &\leq \varepsilon T_n e_0(t) + \frac{2\|f\|_\infty}{\eta^2} (T_n[e_2](t) - 2xT_n[e_1](t) + x^2T_n[e_0](t)). \end{aligned}$$

En évaluant en x on obtient donc

(3.5)

$$\sup_{x \in [0,1]} |T_n[f - f(x)e_0](x)| \leq \varepsilon \|T_n e_0\|_\infty + \frac{2\|f\|_\infty}{\eta^2} \sup_{x \in [0,1]} |T_n[e_2](x) - 2xT_n[e_1](x) + x^2T_n[e_0](x)|.$$

Mais, d'une part $T_n e_0$ converge en norme $\|\cdot\|_\infty$, cette suite est bornée, $\|T_n e_0\|_\infty \leq M$. D'autre part, on observe ensuite que, quand $t = x$, $e_2(x) - 2xe_1(x) + x^2e_0(x) = x^2 - 2x^2 + x^2 = 0$ donc, pour $0 \leq x \leq 1$;

$$\begin{aligned} |T_n[e_2](x) - 2xT_n[e_1](x) + x^2T_n[e_0](x)| \\ = |T_n[e_2](x) - e_2(x) - 2x(T_n[e_1](x) - e_1(x)) + x^2(T_n[e_0](x) - e_0(x))| \\ \leq \|T_n[e_2] - e_2\|_\infty + 2\|T_n[e_1] - e_1\|_\infty + \|T_n[e_0] - e_0\|_\infty \rightarrow 0. \end{aligned}$$

Ainsi, il existe N tel que, si $n \geq N$,

$$\sup_{x \in [0,1]} |T_n[e_2](x) - 2xT_n[e_1](x) + x^2T_n[e_0](x)| \leq \varepsilon.$$

En injectant cela dans (3.5), on obtient

$$\sup_{x \in [0,1]} |T_n[f - f(x)e_0](x)| \leq \left(M + \frac{2\|f\|_\infty}{\eta^2}\right) \varepsilon.$$

On a donc bien $\sup_{x \in [0,1]} |T_n[f - f(x)e_0](x)| \rightarrow 0$ donc $T_n f \rightarrow f$ uniformément. ■

Ce théorème admet aussi une version *trigonométrique*. Cette fois-ci on se place sur $\mathcal{C}(\mathbb{T}) = \{f \in \mathcal{C}_b(\mathbb{R}), f \text{ } 2\pi\text{-périodique}\}$. On pose alors $f_0(t) = 1$, $f_1(t) = \sin t$ et $f_2(t) = \cos t$.

Théorème 3.5 (Théorème de Korovkin périodique)

Soit $(T_n)_{n \geq 1}$ une suite d'applications linéaires $\mathcal{C}(\mathbb{T}) \rightarrow \mathcal{C}(\mathbb{T})$ telles que

- si $f \in \mathcal{C}(\mathbb{T})$ est telle que $f \geq 0$, alors $T_n f \geq 0$;
- pour $j = 0, 1, 2$, $T_n f_j \rightarrow f_j$ uniformément sur $\mathbb{R}$.

Alors $T_n f \rightarrow f$ uniformément sur $\mathbb{R}$ pour tout $f \in \mathcal{C}(\mathbb{T})$.

DÉMONSTRATION. La démonstration est une réécriture immédiate de la précédente. Il faut utiliser qu'une fonction continue périodique est uniformément continue. La principale différence est qu'on écrit maintenant

$$|f(t) - f(x)| \leq \varepsilon + \frac{2M}{1 - \cos \delta} (1 - \cos(t - x)).$$

Ensuite $1 - \cos(t - x) \geq 0$ et $1 - \cos(t - x) = f_0 - \sin(x)f_1(t) - \cos(x)f_2(t)$. ■

Revenons aux polynômes de Bernstein.

Remarque 3.6

Il faut remarquer que les polynômes de Bernstein ne sont pas des polynômes d'interpolation de f . On a en général $B_n f(j/n) \neq f(j/n)$ (bien que $B_n f(j/n)$ soit une approximation de $f(j/n)$ si n est assez grand).

On pourrait se demander pourquoi ne pas considérer les polynômes d'interpolation de Lagrange. Il se trouve que la suite ainsi construite ne converge pas uniformément vers f . Cela peut se montrer à l'aide du théorème de Banach-Steinhaus.

Remarque 3.7

La démonstration ci-dessus donne un résultat un peu plus précis. On note

$$\omega(\eta) = \sup_{x,y \in [0,1], |x-y| \leq \eta} |f(x) - f(y)|$$

le module de continuité de f . Par uniforme continuité, $\omega(\delta) \rightarrow 0$ quand $\delta \rightarrow 0$. En ré-examinant la démonstration ci-dessus, on réalise qu'on a en fait démontré dans (3.3)-(3.4) que

$$\sup_{x \in [0,1]} |B_n f(x) - f(x)| \leq \omega(\eta) + \frac{\|f\|_\infty}{2\eta^2 n}.$$

Par exemple, si f est Lipschitzienne, $|f(x+y) - f(x)| \leq \text{Lip}(f)|y|$ donc $\omega(\eta) \leq \text{Lip}(f)\eta$. On peut alors optimiser l'inégalité précédente

$$\sup_{x \in [0,1]} |B_n f(x) - f(x)| \leq \inf_{\eta > 0} \left(\text{Lip}(f)\eta + \frac{\|f\|_\infty}{2\eta^2 n} \right) = \frac{3}{2} \text{Lip}(f)^{2/3} \|f\|_\infty^{1/3} n^{-1/3}.$$

On peut faire un peu mieux et obtenir une décroissance en $n^{-1/2}$ et montrer qu'elle est optimale.

DÉMONSTRATION PROBABILISTE DU THÉORÈME DE BERNSTEIN. Les polynômes de Bernstein ont une interprétation probabiliste. Soient $X_1, \dots, X_n$, n variables aléatoires de Bernoulli de paramètre x indépendantes *i.e.* n tirages de pile ou face avec une pièce où la probabilité d'avoir pile est x , $\mathbb{P}[X_j = 1] = x$. Soit $S_n = X_1 + \dots + X_n$ le nombre de piles.

On considère alors une fonction continue $f : [0, 1] \rightarrow \mathbb{R}$. On convient que le gain G du joueur est $f(m/n)$ si $S = m$. Alors l'espérance de la variable aléatoire G est

$$\mathbb{E}[G] = \mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] = B_n f(x).$$

C'est d'ailleurs l'interprétation donnée par S.N. Bernstein.[‡] La loi (faible) des grands nombres nous dit que $S_n/n \rightarrow x$ donc intuitivement, on devrait avoir $\mathbb{E}[f(\frac{S_n}{n})] \rightarrow \mathbb{E}[f(x)] = f(x)$ (c'est l'espérance d'une constante).

La démonstration ci-dessus se reformule entièrement en termes probabilistes. On veut estimer

$$|f(x) - B_n f(x)| = \left| \mathbb{E}[f(x)] - \mathbb{E} \left[f \left(\frac{S_n}{n} \right) \right] \right| \leq \mathbb{E} \left[\left| f(x) - f \left(\frac{S_n}{n} \right) \right| \right].$$

On écrit alors, pour $\eta > 0$,

$$\left| f(x) - f \left(\frac{S_n}{n} \right) \right| \leq \omega(\eta) \mathbf{1}_{|x - S_n/n| \leq \eta} + 2\|f\|_\infty \mathbf{1}_{|x - S_n/n| > \eta}$$

donc

$$|f(x) - B_n f(x)| \leq \omega(\eta) \mathbb{P} \left[\left| x - \frac{S_n}{n} \right| \leq \eta \right] + 2\|f\|_\infty \mathbb{P} \left[\left| x - \frac{S_n}{n} \right| > \eta \right].$$

Pour le premier terme, on utilise qu'une probabilité est toujours ≤ 1 . Pour le deuxième, on va utiliser l'inégalité de Tchebichev. Pour cela, remarquons que S_n étant la somme de

[‡]. S.N. Bernstein *Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités* Communications de la Société Mathématique de Kharkov 13 (1913).

n variables de Bernoulli $\mathcal{B}(x)$, $\mathbb{E}\left[\frac{S_n}{n}\right] = x$, et $Var\left(\frac{S_n}{n}\right) = \frac{Var(S_n)}{n^2} = \frac{x(1-x)}{n}$. Ainsi

$$\mathbb{P}\left[\left|x - \frac{S_n}{n}\right| > \eta\right] \leq \frac{x(1-x)}{n\eta^2} \leq \frac{1}{4n\eta^2}.$$

En regroupant les deux, on obtient

$$|f(x) - B_nf(x)| \leq \omega(\eta) + \frac{\|f\|_\infty}{2n\eta^2}.$$

Comme précédemment, si $\varepsilon > 0$, on fixe d'abord η pour que $\omega(\eta) < \varepsilon$ puis N pour que, si $n \geq N$, $\frac{\|f\|_\infty}{2n\eta^2} \leq \varepsilon$ et alors $|f(x) - B_nf(x)| \leq 2\varepsilon$. ■

Il se trouve qu'on peut faire un peu mieux. Pour cela, commençons par un lemme simple :

Lemme 3.8

Soit $f : \mathbb{R} \rightarrow \mathbb{R}$ et ω son module de continuité. Alors $\omega(\eta + \delta) \leq \omega(\eta) + \omega(\delta)$.
En particulier, pour $\lambda > 0$, $\omega(\lambda\eta) \leq (\lambda + 1)\omega(\eta)$.

DÉMONSTRATION. La seconde partie du lemme se déduit de la première puisque pour λ entier, on a $\omega(\lambda\eta) \leq \lambda\omega(\eta)$. Quand λ n'est pas entier, on utilise la croissance de ω pour écrire

$$\omega(\lambda\eta) \leq \omega(\lceil \lambda \rceil \eta) \leq \lceil \lambda \rceil \omega(\eta) \leq (\lambda + 1)\omega(\eta).$$

Montrons donc la première inégalité. Soient x, y tels que $|x - y| \leq \eta + \delta$. Si $|x - y| \leq \eta$ on a bien sûr

$$|f(x) - f(y)| \leq \omega(\eta) \leq \omega(\eta) + \omega(\delta).$$

Ainsi, il suffit de supposer que $\eta \leq |x - y| \leq \eta + \delta$. Il existe alors $z \in [x, y]$ tel que $|x - z| = \eta$ et $|y - z| \leq \delta$. Plus précisément, si $x < y$, on prend $z = x + \eta$ sinon on prend $z = x - \eta$. Donc

$$|f(x) - f(y)| \leq |f(x) - f(z)| + |f(z) - f(y)| \leq \omega(\eta) + \omega(\delta).$$

On en déduit l'inégalité souhaitée. ■

Théorème 3.9

Soit $f \in \mathcal{C}([0, 1])$, alors $\|f - B_nf\|_\infty \leq \frac{3}{2}\omega(n^{-1/2})$.

DÉMONSTRATION. On a maintenant

$$\omega\left(\left|x - \frac{S_n}{n}\right|\right) = \omega\left(\sqrt{n}\left|x - \frac{S_n}{n}\right|n^{-1/2}\right) \leq \left(\sqrt{n}\left|x - \frac{S_n}{n}\right| + 1\right)\omega(n^{-1/2}).$$

On repart alors de

$$\begin{aligned} |f(x) - B_nf(x)| &\leq \mathbb{E}\left[\left|f(x) - f\left(\frac{S_n}{n}\right)\right|\right] \leq \mathbb{E}\left[\omega\left(\left|x - \frac{S_n}{n}\right|\right)\right] \\ &\leq \omega(n^{-1/2})\left(\sqrt{n}\mathbb{E}\left[\left|x - \frac{S_n}{n}\right|\right] + 1\right) \end{aligned}$$

On utilise ensuite Cauchy-Schwarz ($\mathbb{E}[|f|^p] = \|f\|_p^p$ donc $\mathbb{E}[|f|]^2 \leq \mathbb{E}[|f|^2]$)

$$\begin{aligned} \mathbb{E}\left[\left|x - \frac{S_n}{n}\right|\right]^2 &\leq \mathbb{E}\left[\left|x - \frac{S_n}{n}\right|^2\right] = \mathbb{E}\left[\left(\frac{S_n}{n} - \mathbb{E}\left[\frac{S_n}{n}\right]\right)^2\right] \\ &= Var\left(\frac{S_n}{n}\right) = \frac{x(1-x)}{n} \leq \frac{1}{4n} \end{aligned}$$

avec le calcul précédent. On en déduit que

$$|f(x) - B_n f(x)| \leq \omega(n^{-1/2}) \left(\sqrt{n} \frac{1}{\sqrt{4n}} + 1 \right) = \frac{3}{2} \omega(n^{-1/2}).$$

■

Remarque 3.10

Notons maintenant que si f est Lipschitzienne, on obtient $|f(x) - B_n f(x)| \leq \frac{3 \text{Lip}(f)}{2n^{1/2}}$.

Pour démontrer que cette estimation en $n^{1/2}$ est optimale, nous allons utiliser un outil important en probabilités et en analyse :

Théorème 3.11 (Inégalité de Khintchine)

Soient $r_1, \dots, r_n$ des variables aléatoires indépendantes de Rademacher *i.e.* $\mathbb{P}[r_j = 1] = \mathbb{P}[r_j = -1] = \frac{1}{2}$. Soit $a_1, \dots, a_n \in \mathbb{R}$ et $f = \sum_{j=1}^n a_j r_j$ une combinaison linéaire des r_j . Alors

$$\|f\|_2 = \mathbb{E}[|f|^2]^{1/2} \leq \sqrt{e} \mathbb{E}[|f|] = \|f\|_1.$$

Remarque 3.12

Considérons l'espace vectoriel de dimension finie $E_n = \text{Vect}(r_1, \dots, r_n)$. On peut munir cet espace des normes $\|f\|_1$ ou $\|f\|_2$ et ces normes sont évidemment équivalentes. L'inégalité de Cauchy-Schwarz nous garantit que $\|f\|_1 \leq \|f\|_2$. Ce que nous dit l'inégalité de Khintchine, c'est que la constante qui apparaît dans l'inégalité inverse $\|f\|_1 \geq C\|f\|_2$ est indépendante de la dimension (et égale à $e^{-1/2}$).

DÉMONSTRATION. Tout d'abord, remarquons que

$$\mathbb{E}[r_j] = +1 \mathbb{P}[r_j = 1] - 1 \mathbb{P}[r_j = -1] = 0.$$

On note encore $E = \text{Vect}(r_1, \dots, r_n)$ et on munit cet espace de la norme $\|f\|_2$. Il faut remarquer que $r_1, \dots, r_n$ est une base orthonormée de E puisque, par indépendance,

$$\mathbb{E}[r_i r_j] = \begin{cases} \mathbb{E}[r_i] \mathbb{E}[r_j] = 0 & \text{si } i \neq j \\ \mathbb{E}[r_i^2] = \mathbb{E}[1] = 1 & \text{si } i = j \end{cases}.$$

Soit alors $f = \sum_{j=1}^n a_j r_j$. On veut montrer que $\|f\|_2 \leq \sqrt{e} \|f\|_1$. Par homogénéité, il suffit

de montrer cela quand $\|f\|_2 = 1$, c'est-à-dire quand $\sum_{j=1}^n a_j^2 = \|f\|_2^2 = 1$.

Posons $g = \prod_{j=1}^n (1 + ia_j r_j)$ de sorte que

$$\begin{aligned} |g| &= \prod_{j=1}^n |1 + ia_j r_j| = \prod_{j=1}^n \sqrt{1 + (a_j r_j)^2} = \prod_{j=1}^n \sqrt{1 + a_j^2} \\ &\leq \prod_{j=1}^n \sqrt{e^{a_j^2}} = \exp\left(\frac{1}{2} \sum_{j=1}^n a_j^2\right) = \sqrt{e} \end{aligned}$$

où on a utilisé le fait que $1 + u \leq e^u$. En d'autres termes $\|g\|_\infty \leq \sqrt{e}$.

Remarquons maintenant que, par indépendance

$$\begin{aligned} \mathbb{E}[r_k g] &= \mathbb{E}\left[r_k (1 + ia_k r_k) \prod_{j \in \{1, \dots, n\} \setminus \{k\}} (1 + ia_j r_j)\right] \\ &= \mathbb{E}[r_k (1 + ia_k r_k)] \prod_{j \in \{1, \dots, n\} \setminus \{k\}} \mathbb{E}[1 + ia_j r_j]. \end{aligned}$$

Mais $\mathbb{E}[r_k (1 + ia_k r_k)] = \mathbb{E}[r_k] + ia_k \mathbb{E}[r_k^2] = ia_k$ alors que $\mathbb{E}[1 + ia_j r_j] = 1 + ia_j \mathbb{E}[r_j] = 1$. Ainsi $\mathbb{E}[r_k g] = ia_k$. Mais alors

$$\mathbb{E}[fg] = \sum_{k=1}^n a_k \mathbb{E}[r_k g] = i \sum_{k=1}^n a_k^2 = i.$$

On conclue alors avec

$$1 = |\mathbb{E}[fg]| \leq \mathbb{E}[|fg|] = \|fg\|_1 \leq \|f\|_1 \|g\|_\infty \leq \sqrt{e} \|f\|_1$$

qui est bien ce qu'on voulait démontrer. ■

Proposition 3.13

Il existe une fonction f Lipschitzienne sur $[0, 1]$ telle que $\sup_{x \in [0, 1]} |f(x) - B_n f(x)| \geq \frac{1}{2\sqrt{en}}$.

Ainsi, l'estimation que nous avons eu dans le théorème de Bernstein est optimale.

DÉMONSTRATION. La fonction que nous considérons est simplement $f(x) = |x - 1/2|$ et on utilise

$$\sup_{x \in [0, 1]} |f(x) - B_n f(x)| \geq |f(1/2) - B_n f(1/2)| = B_n f(1/2) = \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right]$$

où $S_n = X_1 + \dots + X_n$ est la somme de n variables aléatoires X_i de Bernoulli de paramètre $1/2$. Mais

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \mathbb{E}\left[\left|\frac{S_n}{n} - \frac{1}{2}\right|\right] = \mathbb{E}\left[\left|\frac{S_n}{n} - \frac{1}{2}\right|\right] = \frac{1}{2n} \mathbb{E}[|2S_n - n|].$$

Maintenant remarquons que $2S_n - n = \sum_{j=1}^n r_j$ avec $r_j = 2X_j - 1$. Remarquons que r_j est

une variable aléatoire de Rademacher puisque si $X_j = 1$, $r_j = 1$ et si $X_j = 0$, $r_j = -1$ et ces deux valeurs sont prises avec probabilité $1/2$ chacune. Ainsi, avec l'inégalité de Khinchine, $\mathbb{E}[|2S_n - n|] \geq e^{-1/2} \|2S_n - n\|_2 = (n/e)^{1/2}$. En regroupant toutes les estimations, on trouve

$$\sup_{x \in [0, 1]} |f(x) - B_n f(x)| \geq \frac{1}{2n} \left(\frac{n}{e}\right)^{1/2}$$

qui est bien l'estimation voulue. ■

3.3. Troisième méthode : le théorème de Stone-Weierstrass. Nous allons maintenant démontrer un théorème qui implique la densité des polynômes dans $\mathcal{C}([0, 1])$:

Théorème 3.14 (Stone-Weierstrass)

Soit (X, d) un espace métrique compact. Soit $\mathcal{D} \subset \mathcal{C}(X, \mathbb{R})$ un sous-espace vectoriel tel que

- (i) $\mathcal{D}$ contient les fonctions constantes ;
- (ii) $\mathcal{D}$ est stable par multiplication : si $f, g \in \mathcal{D}$ alors $fg \in \mathcal{D}$ (on dit que $\mathcal{D}$ est une algèbre) ;
- (iii) $\mathcal{D}$ sépare les points : si $x, y \in X$ avec $x \neq y$, alors il existe $f \in \mathcal{D}$ tel que $f(x) \neq f(y)$.

Alors $\mathcal{D}$ est dense dans $\mathcal{C}(X, \mathbb{R})$.

Nous donnerons plus loin la version pour les fonctions à valeurs complexes qui nécessite une condition supplémentaire.

Corollaire 3.15

Les polynômes réels sont denses dans $\mathcal{C}([0, 1], \mathbb{R})$.

DÉMONSTRATION DU COROLLAIRE. Les polynômes forment clairement une algèbre. Ils séparent aussi les points car si $x \neq y$, et $P(t) = t$, alors $P(x) \neq P(y)$. ■

DÉMONSTRATION DU THÉORÈME. La démonstration du théorème se fait en plusieurs étapes. Mais avant de débiter la démonstration, remarquons que si $\mathcal{D}$ vérifie les hypothèses du théorème de Stone-Weierstrass, son adhérence $\bar{\mathcal{D}}$ aussi. Nous allons donc supposer que $\mathcal{D}$ est également *fermé* et montrer qu'alors $\mathcal{D} = \mathcal{C}(X, \mathbb{R})$. La première étape consiste à construire une suite de polynômes qui approchent uniformément $|x|$ sur $[-1, 1]$

Étape 1. Il existe une suite de polynômes P_n telle que $\sup_{x \in [-1, 1]} |P_n(x) - |x|| \rightarrow 0$.

En effet, on considère la suite de polynômes définie par $P_0(t) = 0$ et

$$(3.6) \quad P_{n+1}(t) = P_n(t) + \frac{1}{2}(t^2 - P_n(t)^2).$$

Une récurrence immédiate montre que P_n est bien un polynôme pair. De plus, si $0 \leq P_n(t) \leq |t|$ (qui est vérifié pour $n = 0$) alors, d'une part (3.6) montre que

$$P_{n+1}(t) \geq P_n(t) \geq 0$$

et d'autre part

$$\begin{aligned} |t| - P_{n+1}(t) &= |t| - P_n(t) - \frac{1}{2}(|t| - P_n(t))(|t| + P_n(t)) \\ &= (|t| - P_n(t)) \left(1 - \frac{1}{2}(|t| + P_n(t)) \right) \\ &\geq (|t| - P_n(t))(1 - |t|) \geq 0. \end{aligned}$$

Ainsi, pour t fixé, $(P_n(t))_{n \in \mathbb{N}}$ est une suite croissante et majorée. Cette suite est donc convergente. Déterminons sa limite (simple pour l'instant) $u(t)$. Avec la formule de récurrence, elle vérifie

$$u(t) = u(t) + \frac{1}{2}(t^2 - u(t)^2)$$

donc $u(t)^2 = t^2$ et comme $P_n(t) \geq 0$, $u(t) \geq 0$ donc $u(t) = |t|$.

Remarquons ensuite que $P'_0 \geq 0$ et que, si $P'_n \geq 0$ alors, pour $0 \leq t \leq 1$,

$$P'_{n+1}(t) = P'_n(t) + t - P_n(t)P'_n(t) \geq P'_n(t) + 0 - tP'_n(t) \geq (1-t)P'_n(t) \geq 0$$

donc P_n est croissante sur $[0, 1]$.

On peut donc appliquer le théorème de Dini sur $[0, 1]$: (P_n) est une suite monotone de fonctions monotone, dont la limite est continue sur le compact $[0, 1]$. Le théorème de Dini nous garantit que la convergence est uniforme sur $[0, 1]$. Par parité, elle est donc uniforme sur $[-1, 1]$.

Remarque 3.16

Il est assez difficile de démontrer que P_n converge uniformément vers $|x|$ sans utiliser le lemme de Dini.

Si on ne veut pas utiliser le lemme de Dini, il vaut mieux utiliser une autre suite pour laquelle la démonstration est un peu plus facile (mais plus longue). On considère d'abord $Q_0 = 0$ et $Q_{n+1}(t) = \frac{1}{2}(Q_n^2(t) + t)$ (en particulier $Q_1 = t/2$). On voit assez facilement que

(i) Q_n est un polynôme, pour $t \in [0, 1]$, $0 \leq Q_n(t) \leq 1$, $Q'_{n+1} = Q_n Q'_n + 1/2 \geq 0$ donc Q_n est croissant ;

(ii) on a $Q_{n+2} - Q_{n+1} = \frac{1}{2}(Q_{n+1} - Q_n)(Q_{n+1} + Q_n)$ d'où on déduit que $Q_{n+1} - Q_n \geq 0$ sur $[0, 1]$ puis que $Q_{n+1} - Q_n$ est croissant sur $[0, 1]$.

Il en résulte d'abord que, à $t \in [0, 1]$ fixé, $Q_n(t)$ est croissant et majoré, donc converge vers un $u(t)$ qui vérifie $u(t) = \frac{1}{2}(u^2(t) + t)$ donc, comme $u(t) \leq 1$, $u(t) = 1 - \sqrt{1-t}$. Ensuite, comme $Q_{n+1} - Q_n$ est croissant, pour $t \in [0, 1]$

$$0 \leq Q_{n+1}(t) - Q_n(t) \leq Q_{n+1}(1) - Q_n(1)$$

puis

$$\begin{aligned} 0 &\leq Q_{n+k+1}(t) - Q_n(t) \\ &= Q_{n+k+1}(t) - Q_{n+k}(t) + Q_{n+k}(t) \cdots - Q_{n+1}(t) + Q_{n+1}(t) - Q_n(t) \\ &\leq Q_{n+k+1}(1) - Q_n(1) \end{aligned}$$

et enfin, en faisant $k \rightarrow +\infty$

$$0 \leq u(t) - Q_n(t) \leq u(1) - Q_n(1).$$

Comme $Q_n(1) \rightarrow u(1)$, on en déduit que la convergence est uniforme :

$$Q_n(t) \rightarrow 1 - \sqrt{1-t} \quad \text{uniformément sur } [0, 1].$$

On pose alors

$$P_n(t) = 1 - Q_n(1 - t^2) \rightarrow 1 - (1 - \sqrt{1 - (1 - t^2)}) = \sqrt{t^2} = |t|$$

uniformément en $t^2 \in [0, 1]$ c'est-à-dire en $t \in [-1, 1]$.

Étape 2. Si $f, g \in \mathcal{D}$ alors $|f|$, $\max(f, g)$ et $\min(f, g)$ appartiennent à $\mathcal{D}$.

Remarquons d'abord que si $\mathcal{D}$ est une algèbre, alors pour tout polynôme P et tout $f \in \mathcal{D}$, on a $P(f) \in \mathcal{D}$. En particulier, pour la suite P_n construite à l'étape précédente, si $f \neq 0$, $P_n(f/\|f\|_\infty) \in \mathcal{D}$. Mais $-1 \leq f/\|f\|_\infty \leq 1$. Ainsi, si on fixe $\varepsilon > 0$, il existe N tel que si $n \geq N$, $\sup_{t \in [-1, 1]} |P_n(t) - |t|| \leq \varepsilon$ donc

$$\sup_{x \in X} \left| P_n \left(\frac{f}{\|f\|_\infty} \right) - \frac{|f|}{\|f\|_\infty} \right| \leq \varepsilon.$$

En d'autres termes $P_n \left(\frac{f}{\|f\|_\infty} \right) \rightarrow \frac{|f|}{\|f\|_\infty}$ dans $\mathcal{C}(X)$ donc $\frac{|f|}{\|f\|_\infty} \in \bar{\mathcal{D}} = \mathcal{D}$. Enfin $\mathcal{D}$ étant un espace vectoriel, $|f| = \|f\|_\infty \frac{|f|}{\|f\|_\infty} \in \mathcal{D}$. Évidemment, si $f = 0$, $f \in \mathcal{D}$.

Enfin, on écrit

$$\max(f, g) = \frac{1}{2}(f + g + |f - g|) \quad \text{et} \quad \min(f, g) = \frac{1}{2}(f + g - |f - g|)$$

qui sont donc toutes deux dans $\mathcal{D}$.

En itérant la propriété, si $(g_j)_{j \in J} \subset \mathcal{D}$ est une famille finie d'éléments de $\mathcal{D}$ alors $\max_{j \in J} g_j, \min_{j \in J} g_j \in \mathcal{D}$.

Étape 3. Si $x, y \in X$ avec $x \neq y$ et $\alpha, \beta \in \mathbb{R}$ avec $\alpha \neq \beta$, alors il existe $f \in \mathcal{D}$ tel que $f(x) = \alpha$ et $f(y) = \beta$.

On a supposé qu'il existait $h \in \mathcal{D}$ tel que si $h(x) \neq h(y)$. On définit alors

$$f(t) = \alpha + (\beta - \alpha) \frac{h(t) - h(x)}{h(y) - h(x)}.$$

Comme $\mathcal{D}$ est un espace vectoriel qui contient les constantes, $f \in \mathcal{D}$.

Étape 4. Soit $f \in \mathcal{C}(X, \mathbb{R})$, $\varepsilon > 0$ et $x_0 \in X$. Il existe $g \in \mathcal{D}$ tel que $g(x_0) = f(x_0)$ et, pour tout $x \in X$, $g(x) < f(x) + \varepsilon$.

Pour $y \in X$, $y \neq x_0$, l'étape précédente nous fournit une fonction $g_y \in \mathcal{D}$ telle que $g_y(x_0) = f(x_0)$ et $g_y(y) = f(y)$ (si $f(x_0) = f(y)$ il suffit de prendre $g_y \equiv f(x_0)$ qui est bien dans $\mathcal{D}$). Comme g_y et f sont toutes deux continues, les ensembles

$$U_y = \{x \in X : g_y(x) < f(x) + \varepsilon\}$$

sont ouverts. De plus $x \in U_x$, on a donc un recouvrement ouvert $\bigcup_{y \in X} U_y$ de X qui est compact, il existe donc un ensemble fini $F \subset X$ tel que $\bigcup_{y \in F} U_y = X$. On pose donc $g = \min_{y \in F} g_y \in \mathcal{D}$ et on remarque que

$$g(x_0) = \min_{y \in F} g_y(x_0) = \min_{y \in F} f(x_0) = f(x_0)$$

et si $x \in X$, il existe $y \in F$ tel que $x \in U_y$ donc $g_y(x) < f(x) + \varepsilon$ donc $g(x) = \min_{z \in F} g_z(x) \leq g_y(x) < f(x) + \varepsilon$.

Étape 5. Soit $f \in \mathcal{C}(X, \mathbb{R})$, $\varepsilon > 0$. Il existe $g \in \mathcal{D}$ tel que, pour tout $x \in X$, $f(x) - \varepsilon < g(x) < f(x) + \varepsilon$.

Ceci permet bien évidemment de conclure puisque ceci nous dit que

$$\mathcal{C}(X, \mathbb{R}) = \bar{\mathcal{D}} = \mathcal{D}.$$

Cette étape est une répétition de la précédente où on remplace le min par un max et l'étape 3 par l'étape 4. En effet, pour tout $y \in X$, l'étape 4 nous fournit une fonction g_y telle que $g_y(y) = f(y)$ et pour tout $x \in X$, $g_y(x) < f(x) + \varepsilon$. On considère alors

$$U_y = \{x \in X : g_y(x) > f(x) - \varepsilon\}$$

qui est un ouvert contenant y . Donc $\bigcup_{y \in X} U_y$ de X qui est compact, il existe donc un ensemble fini $F \subset X$ tel que $\bigcup_{y \in F} U_y = X$. On pose donc $g = \max_{y \in F} g_y \in \bar{\mathcal{D}}$. Pour conclure, on prend $x \in X$, il existe $y \in F$ tel que $x \in U_y$ donc

$$g(x) = \max_{z \in F} g_z(x) \geq g_y(x) > f(x) - \varepsilon$$

et par ailleurs pour tout $z \in F$, $g_z(x) < f(x) + \varepsilon$ donc, comme F est fini $g(x) = \max_{z \in F} g_z(x) < f(x) + \varepsilon$. ■

Remarque 3.17

Pour la densité dans $\mathcal{C}(X, \mathbb{C})$, il faut ajouter la condition que $\mathcal{D}$ est stable par conjugaison complexe (si $f \in \mathcal{D}$, alors $\bar{f}$ aussi) et remarquer qu'alors $\Re(\mathcal{D}), \text{Im}(\mathcal{D}) \subset \mathcal{D}$ et sont denses dans $\mathcal{C}(X, \mathbb{R})$ donc $\Re(\mathcal{D}) + i\text{Im}(\mathcal{D})$ est dense dans $\mathcal{C}(X, \mathbb{C})$.

3.4. Interpolation de Lagrange et phénomène de Runge. Le théorème de Weierstrass affirme que toute fonction continue sur $[-1, 1]$ peut être approchée uniformément par des polynômes. Il ne dit cependant pas que les polynômes d'interpolation associés à une famille donnée de nœuds convergent vers la fonction. Rappelons que pour $f \in \mathcal{C}([-1, 1])$, si $x_0 < x_1 < \dots < x_n \in [-1, 1]$, le polynôme d'interpolation de Lagrange en ces nœuds est donné par

$$L_n f(x) = \sum_{j=0}^n f(x_j) \prod_{k \neq j} \frac{x - x_k}{x_j - x_k}.$$

La question qu'on pose ici est de faire tendre le nombre de nœuds vers l'infini tout en faisant tendre l'écart maximal entre deux nœuds vers 0.

L'exemple suivant, dû à Runge,[§] montre que l'interpolation polynomiale aux nœuds équidistants peut même diverger pour une fonction analytique au voisinage de l'intervalle réel.

Théorème 3.18 (Phénomène de Runge)

Pour $n \geq 1$, considérons les $n + 1$ nœuds équidistants $x_{k,n} = -1 + \frac{2k}{n}$, pour $k = 0, \dots, n$, et notons $L_n f$ le polynôme de degré au plus n qui interpole f en ces points. Pour la fonction $f : [-1, 1] \rightarrow \mathbb{R}$ donnée par $f(x) = \frac{1}{1+25x^2}$, la suite $(L_n f)$ ne converge pas uniformément vers f sur $[-1, 1]$. Plus précisément,

$$\left| L_{2m+1} f \left(1 - \frac{1}{2m+1} \right) - f \left(1 - \frac{1}{2m+1} \right) \right| \rightarrow +\infty.$$

En particulier, $\|L_{2m+1} f - f\|_\infty \not\rightarrow 0$.

DÉMONSTRATION DE LA PROPOSITION. Pour $a > 0$, il sera commode d'introduire $f_a(x) = \frac{1}{a^2+x^2}$ de sorte que $f(x) = \frac{1}{25} f_{1/5}(x)$.

Posons $\omega_n(z) = \prod_{k=0}^n (z - x_{k,n})$ et supposons désormais que $n = 2m + 1$ soit impair. Le polynôme ω_n est alors pair, puisque les nœuds sont symétriques par rapport à l'origine.

Comme f_a est paire et que l'ensemble des nœuds est symétrique, son polynôme d'interpolation $L_n f_a$ est également pair. Le polynôme $Q_n(z) = 1 - (z^2 + a^2)L_n f_a(z)$ s'annule en chacun des points $x_{0,n}, \dots, x_{n,n}$. Il est donc divisible par ω_n . Par ailleurs, $\deg Q_n \leq n + 2$. Le quotient Q_n/ω_n est donc de degré au plus un. Puisque Q_n et ω_n sont pairs, ce quotient est constant. Il existe ainsi une constante c_n telle que

$$1 - (z^2 + a^2)L_n f_a(z) = c_n \omega_n(z).$$

En prenant $z = ia$, on obtient $1 = c_n \omega_n(ia)$, i.e. $c_n = \frac{1}{\omega_n(ia)}$.

§. C. Runge, *Über empirische Funktionen und die Interpolation zwischen äquidistanten Ordinaten*, Zeitschrift für Mathematik und Physik **46** (1901), 224–243. La démonstration ici est inspirée de J. F. Epperson, *On the Runge example*, Amer. Math. Monthly **94** (1987), no. 4, 329–341.

Nous avons donc la formule exacte

$$(3.7) \quad f_a(x) - L_n f_a(x) = \frac{\omega_n(x)}{(x^2 + a^2)\omega_n(ia)}.$$

Évaluons cette expression au point $y_n = 1 - \frac{1}{n}$, qui est le milieu des deux derniers nœuds $x_{n-1,n} = 1 - \frac{2}{n}$ et $x_{n,n} = 1$. On a

$$|\omega_n(y_n)| = \prod_{k=0}^n \left| 1 - \frac{1}{n} + 1 - \frac{2k}{n} \right| = \frac{1}{n^{n+1}} \prod_{k=0}^n |2n - 1 - 2k|.$$

Comme n est impair, les nombres apparaissant dans le dernier produit sont $2n - 1, 2n - 3, \dots, 3, 1, -1$, d'où

$$(3.8) \quad |\omega_n(y_n)| = \frac{(2n - 1)!!}{n^{n+1}} = \frac{(2n)!}{2^n n! n^{n+1}}.$$

La formule de Stirling donne

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad (2n)! \sim \sqrt{4\pi n} \left(\frac{2n}{e}\right)^{2n}.$$

On en déduit

$$(3.9) \quad |\omega_n(y_n)|^{1/n} \longrightarrow \frac{2}{e}.$$

D'autre part, $|\omega_n(ia)| = \prod_{k=0}^n \sqrt{a^2 + x_{k,n}^2}$. Par conséquent,

$$\frac{1}{n} \log |\omega_n(ia)| = \frac{1}{2n} \sum_{k=0}^n \log(a^2 + x_{k,n}^2).$$

Cette somme est une somme de Riemann. Comme les points $x_{k,n}$ sont équidistants dans $[-1, 1]$, on obtient

$$\frac{1}{n} \log |\omega_n(ia)| \rightarrow \frac{1}{4} \int_{-1}^1 \log(a^2 + x^2) dx = \frac{1}{2} \int_0^1 \log(a^2 + x^2) dx.$$

Or, en intégrant par parties

$$\begin{aligned} \int_0^1 \log(a^2 + x^2) dx &= \left[x \log(a^2 + x^2) \right]_0^1 - \int_0^1 \frac{2x^2}{x^2 + a^2} dx \\ &= \left[x \log(a^2 + x^2) - 2x + 2a \arctan\left(\frac{x}{a}\right) \right]_0^1 \\ &= \log(1 + a^2) - 2 + 2a \arctan\left(\frac{1}{a}\right). \end{aligned}$$

Ainsi,

$$(3.10) \quad |\omega_n(ia)|^{1/n} \longrightarrow \sqrt{1 + a^2} \exp\left(-1 + a \arctan\left(\frac{1}{a}\right)\right).$$

En combinant (3.9) et (3.10), il vient

$$(3.11) \quad \left(\frac{|\omega_n(y_n)|}{|\omega_n(ia)|} \right)^{1/n} \longrightarrow \rho(a),$$

où

$$\rho(a) = \frac{2 \exp\left(-a \arctan\left(\frac{1}{a}\right)\right)}{\sqrt{1 + a^2}}.$$

Prenons maintenant $a = 1/5$. Alors

$$\rho\left(\frac{1}{5}\right) = \frac{2 \exp\left(-\frac{1}{5} \arctan 5\right)}{\sqrt{26/25}} \simeq 1.49 > 1.$$

Il résulte de (3.11) que

$$\frac{|\omega_n(y_n)|}{|\omega_n(i/5)|} \longrightarrow +\infty$$

exponentiellement vite le long des entiers impairs.

Enfin, la formule exacte (3.7) donne

$$|f_{1/5}(y_n) - L_n f_{1/5}(y_n)| = \frac{|\omega_n(y_n)|}{\left(y_n^2 + \frac{1}{25}\right) |\omega_n(i/5)|}.$$

Comme $y_n^2 + \frac{1}{25} \longrightarrow \frac{26}{25}$, on en déduit $|f_{1/5}(y_n) - L_n f_{1/5}(y_n)| \longrightarrow +\infty$. Puisque $f = \frac{1}{25} f_{1/5}$, le même résultat vaut pour la fonction de Runge : $|f(y_n) - L_n f(y_n)| \longrightarrow +\infty$. En particulier, $\|L_n f - f\|_\infty \longrightarrow +\infty$ le long des entiers impairs. ■

Remarque 3.19

Le phénomène ne contredit pas le théorème de Weierstrass. Celui-ci affirme que, pour toute fonction continue f , il existe une suite de polynômes convergeant uniformément vers f . Il ne garantit nullement que les polynômes d'interpolation de Lagrange associés aux nœuds équidistants possèdent cette propriété.

Remarque 3.20

L'origine de la divergence est visible dans la formule

$$f_a(x) - L_n f_a(x) = \frac{\omega_n(x)}{(x^2 + a^2) \omega_n(ia)}.$$

Les singularités complexes les plus proches de f_a sont les pôles $\pm ia$. Lorsque ceux-ci sont suffisamment proches de l'intervalle réel, le dénominateur $\omega_n(ia)$ décroît, relativement à certaines valeurs de $\omega_n(x)$ proches des extrémités, à une vitesse qui provoque une croissance exponentielle de l'erreur.

Remarque 3.21

L'utilisation de nœuds de Tchebychev, plus concentrés aux extrémités de l'intervalle, supprime ce comportement pour les fonctions analytiques dans un voisinage approprié de $[-1, 1]$.

Toutefois, le théorème de Faber nous dit que quels que soient les nœuds, il existe une fonction dont les interpolants de Lagrange ne convergent pas uniformément. Erdős et Vértesi ont même montré qu'il existait une fonction continue dont les interpolants de Lagrange divergeaient presque partout.

3.5. Le théorème des moments de Hausdorff. Le théorème suivant caractérise les suites qui sont les moments d'une mesure positive sur l'intervalle $[0, 1]$.

Théorème 3.22 (Hausdorff)

Soit $(a_n)_{n \geq 0}$ une suite réelle. Les assertions suivantes sont équivalentes.

(1) Il existe une mesure positive finie μ sur $[0, 1]$ telle que, pour tout $n \geq 0$,

$$a_n = \int_0^1 x^n d\mu(x).$$

(2) Pour tous les entiers $n, k \geq 0$, $(-1)^k \Delta^k a_n \geq 0$, où Δ désigne l'opérateur de différences finies

$$\Delta a_n = a_{n+1} - a_n, \quad \Delta^2 a_n = \Delta(\Delta a)_n = a_{n+2} - 2a_{n+1} + a_n \dots$$

DÉMONSTRATION. Commençons par (1) $\implies$ (2).

On suppose donc que $a_n = \int_0^1 x^n d\mu(x)$. Mais alors

$$\Delta a_n = a_{n+1} - a_n = \int_0^1 x^n (x - 1) d\mu(x).$$

Supposons qu'on ait montré que

$$\Delta^k a_n = \int_0^1 x^n (x - 1)^k d\mu(x).$$

Alors

$$\begin{aligned} \Delta^{k+1} a_n &= \Delta(\Delta^k a_n) = \int_0^1 x^{n+1} (x - 1)^k d\mu - \int_0^1 x^n (x - 1)^k d\mu \\ &= \int_0^1 x^n (x - 1)^{k+1} d\mu. \end{aligned}$$

Ainsi, on a montré par récurrence que cette formule était vraie à tout rang.

Il en résulte que,

$$(-1)^k \Delta^k a_n = \int_0^1 x^n (1 - x)^k d\mu(x) \geq 0,$$

puisque l'intégrande est positif.

Montrons maintenant (2) $\implies$ (1).

Remarquons que $a_k = \Delta^0 a_k \geq 0$. Ensuite, si $a_0 = 0$, alors comme $\Delta a_n = a_{n+1} - a_n \leq 0$, (a_n) est une suite positive décroissante donc est la suite nulle. Il suffit donc de prendre $\mu = 0$. On va donc supposer que $a_0 > 0$.

Pour chaque entier N , considérons la mesure discrète sur $[0, 1]$ définie par

$$\mu_N = \sum_{j=0}^N \binom{N}{j} (-1)^{N-j} \Delta^{N-j} a_j \delta_{j/N},$$

où δ_x désigne la masse de Dirac en x .

Nous avons supposé que $(-1)^{N-j} \Delta^{N-j} a_j \geq 0$, de sorte que μ_N est une mesure positive.

Introduisons maintenant l'opérateur de décalage $S a_j = a_{j+1}$ de sorte que $\Delta = S - I$. Comme S, Δ commutent, la formule de Newton donne

$$I = I^N = (S - \Delta)^N = \sum_{j=0}^N \binom{N}{j} (-1)^{N-j} \Delta^{N-j} S^j.$$

En partie, la masse totale de μ_N est

$$\mu_N([0, 1]) = \sum_{j=0}^N \binom{N}{j} (-1)^{N-j} \Delta^{N-j} a_j = (S - \Delta)^N a_0 = a_0.$$

Ensuite, pour $x \in \mathbb{R}$, $k \in \mathbb{N}$ définissons $(x)_k = x(x-1)\cdots(x-k+1)$. Remarquons que, si $N \geq k$,

$$\frac{(j)_k}{(N)_k} \binom{N}{j} = \binom{N-k}{j-k}.$$

En particulier,

$$\begin{aligned} \int_0^1 \frac{(Nx)_k}{(N)_k} d\mu(x) &= \sum_{j=0}^N \binom{N}{j} (-1)^{N-j} \Delta^{N-j} a_j \frac{(j)_k}{(N)_k} \\ &= \sum_{j=k}^N \binom{N-k}{j-k} (-1)^{N-j} \Delta^{N-j} a_j \\ &= \sum_{\ell=0}^{N-k} \binom{N-k}{\ell} (-1)^{N-k-\ell} \Delta^{N-k-\ell} a_{k+\ell} = a_k. \end{aligned}$$

Posons maintenant $p_{N,j} = \binom{N}{j} (-1)^{N-j} \Delta^{N-j} a_j$ de sorte que ceci s'écrive

$$\sum_{j=0}^N (j)_k p_{N,j} = (N)_k a_k.$$

Fixons maintenant $k \geq 0$. L'identité classique faisant intervenir les nombres de Stirling de seconde espèce donne

$$j^k = \sum_{\ell=0}^k \left\{ \begin{matrix} k \\ \ell \end{matrix} \right\} (j)_\ell.$$

La seule chose dont nous avons besoin est l'existence d'une telle formule avec $\left\{ \begin{matrix} k \\ k \end{matrix} \right\} = 1$ qui peut s'établir facilement. Par conséquent,

$$\begin{aligned} \int_0^1 x^k d\mu_N(x) &= \sum_{j=0}^N \left(\frac{j}{N} \right)^k p_{N,j} = \frac{1}{N^k} \sum_{j=0}^N j^k p_{N,j} = \frac{1}{N^k} \sum_{j=0}^N \sum_{\ell=0}^k \left\{ \begin{matrix} k \\ \ell \end{matrix} \right\} (j)_\ell p_{N,j} \\ &= \sum_{\ell=0}^k \left\{ \begin{matrix} k \\ \ell \end{matrix} \right\} \frac{1}{N^k} \sum_{j=0}^N (j)_\ell p_{N,j} = \sum_{\ell=0}^k \left\{ \begin{matrix} k \\ \ell \end{matrix} \right\} \frac{(N)_\ell}{N^k} a_\ell. \end{aligned}$$

On fait maintenant tendre N vers $+\infty$, avec k fixé. Pour $\ell < k$,

$$\frac{(N)_\ell}{N^k} = O(N^{\ell-k}) \longrightarrow 0,$$

tandis que

$$\frac{(N)_k}{N^k} = \prod_{j=0}^{k-1} \left(1 - \frac{j}{N} \right) \longrightarrow 1.$$

De plus, $\left\{ \begin{matrix} k \\ k \end{matrix} \right\} = 1$ de sorte que

$$\int_0^1 x^k d\mu_N(x) \longrightarrow a_k.$$

Remarque 3.23

Nous ne passons pas à la limite sous une intégrale ici car la mesure dépend elle aussi de N .

La famille (μ_N) est bornée dans l'espace des mesures positives sur $[0, 1]$. Le théorème de Prokhorov (ou plutôt la remarque ??) permet d'extraire une sous-suite (μ_{N_ℓ}) convergeant vaguement vers une mesure positive μ . En passant à la limite,

$$\int_0^1 x^k d\mu(x) = \lim_{\ell \rightarrow \infty} \int_0^1 x^k d\mu_{N_\ell}(x) = a_k.$$

La mesure μ possède donc les moments prescrits. ■

Notez que la mesure μ est unique puisque

$$\int_0^1 x^k d\mu(x) = \int_0^1 x^k d\nu(x)$$

implique

$$\int_0^1 P(x) d\mu(x) = \int_0^1 P(y) d\nu(x).$$

Mais alors, si $f \in \mathcal{C}([0, 1])$, en prenant pour $P = B_n(f)$ (par exemple), et en passant à la limite (remarquer que $|B_n(f)| \leq \|f\|_\infty$ et utiliser la convergence dominée), on obtient

$$\int_0^1 f(x) d\mu(x) = \int_0^1 f(y) d\nu(x)$$

soit $\mu = \nu$.

Remarque 3.24

Ce résultat montre que les moments sur $[0, 1]$ sont entièrement caractérisés par une condition de positivité des différences finies. Il constitue le prototype des problèmes classiques des moments, étudiés notamment par Stieltjes et Hamburger. Pour en savoir plus sur le problème des moments, on pourra consulter les ouvrages suivants : J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943.

N. I. Akhiezer, *The Classical Moment Problem*, Oliver & Boyd, 1965.

Théorème 3.25 (Cas positif du théorème de représentation de Riesz)

Soit $L : \mathcal{C}([0, 1]) \rightarrow \mathbb{R}$ une forme linéaire positive, c'est-à-dire telle que $Lf \geq 0$ si $f \geq 0$. Alors il existe une unique mesure borélienne positive finie μ sur $[0, 1]$ telle que, pour tout $f \in \mathcal{C}([0, 1])$,

$$L(f) = \int_0^1 f(x) d\mu(x).$$

De plus, $\mu([0, 1]) = L(1)$, en particulier, μ est une mesure de probabilités si $L(1) = 1$.

Notez qu'une forme linéaire positive est nécessairement continue sur $\mathcal{C}([0, 1])$ puisque, si $f \in \mathcal{C}([0, 1])$, on a $-\|f\|_\infty 1 \leq f \leq \|f\|_\infty 1$ donc, par positivité de L , $-\|f\|_\infty L(1) \leq L(f) \leq \|f\|_\infty L(1)$ soit $|L(f)| \leq L(1)\|f\|_\infty$. En particulier, si $L(1) = 0$ alors $L = 0$.

DÉMONSTRATION. Si $L = 0$, il suffit de prendre $\mu = 0$. On peut donc supposer que $L(1) > 0$. Posons alors $a_n = \frac{L(x^n)}{L(1)}$ pour $n \geq 0$ de sorte que $a_0 = 1$.

La même récurrence immédiate que dans le théorème de Hausdorff montre alors que $(-1)^k \Delta^k a_n = \frac{1}{L(1)} L(x^n(1-x)^k) \geq 0$, puisque $x^n(1-x)^k \geq 0$ sur $[0, 1]$.

La suite $(a_n)_{n \geq 0}$ est donc complètement monotone. Le théorème des moments de Hausdorff fournit alors une probabilité ν sur $[0, 1]$ telle que, pour $n \geq 0$,

$$a_n = \int_0^1 x^n d\nu(x).$$

En posant $\mu = L(1)\nu$, on obtient

$$L(x^n) = \int_0^1 x^n d\mu(x)$$

pour tout $n \geq 0$. Par linéarité,

$$L(P) = \int_0^1 P(x) d\mu(x)$$

pour tout polynôme P .

Soit maintenant $f \in C([0, 1])$. D'après le théorème de Weierstrass, il existe une suite de polynômes (P_j) telle que $\|P_j - f\|_\infty \rightarrow 0$. Par continuité de L , $L(P_j) \rightarrow L(f)$.

D'autre part,

$$\left| \int_0^1 P_j d\mu - \int_0^1 f d\mu \right| \leq \mu([0, 1]) \|P_j - f\|_\infty \rightarrow 0.$$

On obtient donc en passant à la limite

$$L(f) = \int_0^1 f d\mu.$$

Enfin, l'unicité de μ découle à nouveau du théorème de Weierstrass (ou de Bernstein) comme pour le théorème des moments de Hausdorff.

Enfin, en appliquant la représentation à la fonction constante 1,

$$\mu([0, 1]) = \int_0^1 1 d\mu = L(1)$$

ce qui montre que μ est une mesure de probabilité si $L(1) = 1$. ■

Remarque 3.26

Cette démonstration relie directement la positivité d'une forme linéaire sur $C([0, 1])$ au théorème des moments de Hausdorff, puis utilise le théorème de Weierstrass pour passer des polynômes aux fonctions continues.

Index

- Équation
 - Cauchy, [14](#)
- Convergence
 - simple, [3](#)
 - uniforme, [3](#)
- Courbe
 - géodésique, [34](#)
 - longueur, [32](#)
 - rectifiable, [32](#)
- Critère
 - de Leibniz, [13](#)
- Distance
 - Wasserstein, [29](#)
- Fonction
 - W de Weierstrass, [16](#)
 - additive, [14](#)
 - Möbius, [34](#)
- Inégalité
 - Khintchine, [45](#)
- Lemme
 - Dini, [25](#)
- Phénomène de Runge, [50](#)
- Polynôme
 - Bernstein, [39](#)
- Suite
 - Farey, [36](#)
- Suites de fonctions
 - dérivation, [10](#)
 - intégration, [7](#)
 - primitives, [9](#)
- Théorème
 - Ascoli–Arzelà, [23](#)
 - Bernstein, [39](#)
 - Bohman–Korovkin, [41](#)
 - Farey, [34](#)
 - Hausdorff, [52](#)
 - Hyers–Ulam, [15](#)
 - Korovkin (périodique), [42](#)
 - Pólya, [27](#)
 - Prohorov, [28](#)
 - représentation de Riesz, [55](#)
 - Weierstrass (fonction continue nulle-part dérivable), [16](#)