```
--------------------------------------------------------------------------------
|         U n i v e r s i t é  L i l l e  1  -  U F R  d e  M a t h é m a t i q u e s         |
|         L i c e n c e  S e m e s t r e  6  -  2 0 0 5 / 2 0 0 6  -  M A T H  3 1 2         |
--------------------------------------------------------------------------------
```

Séries de Fourier et Espaces de Hilbert

Compléments de cours

17 mai 2006

Il s'agit d'un cours sur un semestre en Licence de Mathématiques, troisième année.

L'objectif a été de développer les notions relatives au théorème de Riesz-Fischer et à l'égalité de Fatou-Parseval pour les séries de Fourier. Le cours sert donc aussi d'introduction aux espaces de Hilbert et d'approfondissement en théorie de la mesure et de l'intégration.

Le présent document comporte une quinzaine d'annexes, qui pour la plupart apportent des compléments au cours, en particulier des compléments d'approfondissement en théorie de l'intégration. Ce polycopié n'est donc absolument pas une transcription du cours ; on pourra au besoin se faire une idée de la première moitié du cours effectif via l'annexe intitulée << Résumé de Cours >>. La deuxième moitié a porté sur les notions initiales liées aux espaces de Hilbert (bases orthonormées et projections orthogonales). Cette deuxième moitié du cours n'est pas abordée ici et n'est présente que dans les feuilles d'exercices disponibles sur mon site.

Les annexes H-I-J-K-L-M-N et Q ont été rédigées dans le but de faciliter l'acquisition par l'étudiant d'une culture suffisamment approfondie en théorie de la mesure et de l'intégration pour être en mesure de suivre avec aisance les cours de niveau M1 (analyse fonctionnelle, Fourier et EDP, etc...).

L'annexe A portant sur l'intégrale de Riemann pourra apporter aux lecteurs intéressés une discussion moins sommaire que celle dont elle fait dorénavant l'objet dans les facultés.

Novembre 2009 : améliorations dans cette fiche sur l'intégrale de Riemann avec incorporation d'extraits du Mémoire de ce dernier.

Jean-François Burnol.

http://jf.burnol.free.fr/ens.html

page suivante, s.v.p.

L'intégrale de Riemann

version du 17 mai 2006. Quelques améliorations en novembre 2009.

Je commence par une présentation de l'intégrale de Riemann que j'avais écrite à l'origine pour un cours de première année (!) en 2003 (on était pourtant déjà au XXI$^{\text{e}}$ siècle). J'ai laissé le texte à peu près tel quel. Un mot s'impose pour exprimer mon désarroi en ce qui concerne l'état actuel de l'enseignement de l'intégration en France dans le supérieur (je ne donnerai que les initiales « c.p. » de ce qui, hélas, donne le ton dans notre cher Pays. Mais pas à Lille en 2003 !). On laisse croire aux gens que l'intégrale de Riemann ne s'applique qu'aux fonctions continues par morceaux. On ne parle presque plus de sommes de Riemann, et je ne suis même pas sûr que l'on prouve leur convergence lorsque le pas tend vers zéro. Il paraît que cela remonte à 1995. On admet l'utilisation, sans démonstration, du théorème de la convergence dominée. On ne donne plus de problème nécessitant de trouver et de prouver de la convergence uniforme, apprentissage fondamental pour toute pensée autonome. Du coup lorsque les étudiants entendent parler de l'intégrale de Lebesgue, pour eux c'est la même théorie (car une théorie est définie par ses théorèmes ; à ce propos dans le temps la Mathématique prouvait ses Théorèmes), la seule différence étant qu'elle s'applique à plus de fonctions. Au final on a détruit et Riemann et Lebesgue.

Ici je donne sur une quinzaine de pages un exposé raisonnablement complet de la théorie de l'intégrale de Riemann : ne manquent principalement que les formules de la moyenne et les changements de variable, ce qui pour ces dernières prendrait à peine une page permettant d'insister sur le fait qu'il y a bien DEUX types de changement de variable, bref même cela est très mal fait dans la littérature, mais bon pas envie de m'y arrêter aujourd'hui.

Si l'on tient absolument à utiliser la convergence dominée, il FAUT la prouver. C'est ce que j'ai fait dans un texte qui est disponible sur Internet(*), par une méthode abordable au niveau L1/L2.

Ici, j'ai ajouté à la suite de mon exposé de première année le théorème de Lebesgue qui caractérise les fonctions intégrables au sens de Riemann. Un énoncé très proche avait déjà été donné par Riemann, et je reproduis en annexe en allemand et en traduction son texte. C'est dans le cadre de son Mémoire sur les séries trigonométriques qu'il présenta son approche à l'intégration : *Ueber die Darstellbarkeit einer Funktion durch eine trigonometrische Reihe*, 1854,

(*) http://jf.burnol.free.fr/convergencedominee.pdf parution posthume en 1867. Ce texte a été écrit à une époque où la notion même de nombres réels n'avait pas été bien clarifiée.

Voici les paroles très claires de J.-P. Kahane à ce sujet : *Comme nous le voyons, des notions aussi importantes que les réunions dénombrables, les ensembles fermés, la compacité des ensembles fermés bornés et les ensembles de mesure de Lebesgue nulle étaient appelées par le problème et par la condition donnée par Riemann.* (*Séries de Fourier et ondelettes*, J.-P. Kahane et P. G. Lemarié. éd. Cassini, 1998.) Vous retrouverez dans ce dernier livre une large partie du Mémoire de Riemann, plus étendue que les extraits que je vous propose ici en annexe.

** DÉBUT DU RAPPEL DE PREMIÈRE ANNÉE **

Bernhard Riemann, 1826-1866, a introduit plusieurs idées révolutionnaires en mathématiques et a aussi pensé profondément les rapports entre mathématiques et modélisation du monde naturel. Son héritage immensément vaste inclut entre autres choses une approche à la géométrie utilisée par Einstein pour la théorie de la gravitation, et de magnifiques découvertes sur les nombres premiers.

Définitions et premières propriétés principales

Soit f une fonction à valeurs réelles, définie sur un intervalle borné $[a,b]$, $-\infty < a < b < +\infty$. Nous imposerons à la fonction f d'être bornée. (*)

On commence par quelques définitions : une subdivision $\mathcal{A}$ est la donnée d'un nombre fini de points $a_0 = a$, a_1, a_2, ..., $a_N = b$ avec $a_j \leq a_{j+1}$ pour $0 \leq j < N$. On peut combiner deux subdivisions, ce qui en donne une troisième plus « fine ». Le j-ième intervalle de la subdivision est $I_j = [a_{j-1}, a_j]$. Définissons les nombres réels m_j et M_j par(**)

$$m_j = \inf_{x \in I_j} f(x) \qquad M_j = \sup_{x \in I_j} f(x)$$

La somme de Darboux inférieure $S_-(f, \mathcal{A})$ associée à $\mathcal{A}$ est la quantité :

$$S_-(f,\mathcal{A}) = \sum_{1 \leq j \leq N} (a_j - a_{j-1}) m_j$$

(*) la théorie de Riemann procède en deux temps : d'abord on ne considère que les fonctions bornées sur les intervalles bornés; ensuite on a une deuxième définition (intégrales « généralisées »), par une limite, si l'intervalle est infini, ou si la fonction n'est pas bornée en a^+ ou en b^-.

(**) On aurait pu prendre à leur place des infima ou suprema sur des intervalles ouverts, cela ne change rien au final, rendrait plus facile certaines considérations, un peu moins d'autres. On aurait aussi pu imposer aux points a_j la condition $a_j < a_{j+1}$

La somme de Darboux supérieure $S_+(f,\mathcal{A})$ est :

$$S_+(f,\mathcal{A}) = \sum_{1\le j\le N} (a_j - a_{j-1})M_j$$

On se fait en passant la petite remarque que si $a_{j-1} = a_j$ alors il n'y pas de contribution de l'intervalle I_j aux sommes de Darboux. Plus essentiel est l'observation que toujours : $S_-(f,\mathcal{A}) \le S_+(f,\mathcal{A})$. On montrera bien mieux plus loin avec le Lemme crucial :

$$\forall\mathcal{A}, \forall\mathcal{B} \quad S_-(f,\mathcal{A}) \le S_+(f,\mathcal{B})$$

Admettons-le pour l'instant. On définit une intégrale inférieure $I_-(f)$ et une intégrale supérieure $I_+(f)$ via :

$$\sup_{\text{tous les } \mathcal{A}} S_-(f,\mathcal{A}) = I_-(f) \le I_+(f) = \inf_{\text{tous les } \mathcal{B}} S_+(f,\mathcal{B})$$

On dira que la fonction f est *intégrable au sens de Riemann* (ou plus simplement R-intégrable) sur l'intervalle $[a,b]$ si $I_-(f) = I_+(f)$. On appelle intégrale de Riemann de f la valeur commune, que l'on note temporairement ici $I(f)$. Ensuite on adoptera la notation plus usuelle $\int_a^b f(x)dx$ (nota bene : la lettre x peut être remplacée par n'importe quelle autre : $\int_a^b f(x)dx = \int_a^b f(y)dy = \int_a^b f(u)du = \dots$).

Nous donnerons plus loin les preuves des affirmations suivantes :
- Toute fonction monotone est intégrable au sens de Riemann.
- Toute fonction continue est intégrable au sens de Riemann.
- Si f est intégrable au sens de Riemann sur $[a,b]$ alors f est intégrable au sens de Riemann sur tout intervalle $[c,d] \subset [a,b]$.
- Si la fonction f est intégrable au sens de Riemann sur $[a,b]$ et sur $[b,c]$ alors elle est intégrable au sens de Riemann sur $[a,c]$ ($a < b < c$). De plus on a :

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

Le *pas* noté $\delta(\mathcal{A})$ d'une subdivision $\mathcal{A}$ est la valeur maximale des $a_j - a_{j-1}$, $1 \le j \le N$. En outre on désigne par

$$\Delta(\mathcal{A}) = S_+(f,\mathcal{A}) - S_-(f,\mathcal{A})$$

l'écart entre la somme inférieure et la somme supérieure. Clairement, si $\Delta(\mathcal{A}) < \epsilon$ pour un $\mathcal{A}$ alors $I_+(f) - I_-(f) < \epsilon$.

Théorème : *Si* f *est R-intégrable alors quel que soit* $\epsilon > 0$ *il existe* $\delta > 0$ *tel que :*

$$\delta(\mathcal{A}) \le \delta \quad \Longrightarrow \quad I(f) - \epsilon \le S_-(f,\mathcal{A}) \le I(f) \le S_+(f,\mathcal{A}) \le I(f) + \epsilon$$

La démonstration n'est pas immédiate : nous la verrons plus loin.

À toute subdivision $\mathcal{A}$ et tout choix subordonné $\underline{\xi} = (\xi_j)_{1\leq j\leq N}$ de points ξ_j, c'est-à-dire $\xi_j \in I_j$ pour $1 \leq j \leq N$, on associe la somme de Riemann :

$$S(f, \mathcal{A}, \underline{\xi}) = \sum_{1\leq j\leq N} (a_j - a_{j-1}) f(\xi_j)$$

On a toujours :

$$S_-(f, \mathcal{A}) \leq S(f, \mathcal{A}, \underline{\xi}) \leq S_+(f, \mathcal{A})$$

Soit $\epsilon > 0$ et soit $\delta > 0$ ayant la propriété du Théorème précédent pour ce ϵ. Si $\delta(\mathcal{A}) \leq \delta$ alors $S(f, \mathcal{A}, \underline{\xi})$, qui est entre $S_-(f, \mathcal{A})$ et $S_+(f, \mathcal{A})$, appartient à l'intervalle $[I(f) - \epsilon, I(f) + \epsilon]$. Donc :

$$\delta(\mathcal{A}) \leq \delta \Longrightarrow \left| S(f, \mathcal{A}, \underline{\xi}) - I(f) \right| \leq \epsilon$$

On a ainsi le théoréme suivant :

Théorème : *Soit* f *intégrable au sens de Riemann. Pour toute suite de subdivisions* $\mathcal{A}^{(n)}$*, de pas* $\delta(\mathcal{A}^{(n)})$ *tendant vers zéro, et tout choix arbitraire de points subordonnés* $\underline{\xi}^{(n)}$*, les sommes de Riemann associées convergent vers l'intégrale de Riemann de* f *:*

$$\lim_{n\to\infty} \delta(\mathcal{A}^{(n)}) = 0 \quad \Longrightarrow \quad \lim_{n\to\infty} S(f, \mathcal{A}^{(n)}, \underline{\xi}^{(n)}) = I(f)$$

D'ailleurs, Riemann définit précisément la R-intégrabilité comme étant cette propriété d'existence d'une limite pour les sommes de Riemann lorsque le pas tend vers zéro. L'énoncé permet de comprendre la notation $\int_a^b f(x)dx$: le symbole $\int$ est une lettre S stylisée, comme dans << Somme >>, le dx représente les accroissements $a_j - a_{j-1}$, et $\int_a^b f(x)dx$ est donc là pour rappeler la somme de Riemann $\sum_j f(\xi_j)(a_j - a_{j-1})$, $\xi_j \in [a_{j-1}, a_j]$ (certes la notation existait avant la notion de somme de Riemann !).

En particulier on a, si f est R-intégrable :

$$\int_a^b f(x)dx = \lim_{N\to\infty} \frac{b-a}{N} \left(f(a) + f(a + \frac{b-a}{N}) + f\left(a + 2\frac{b-a}{N}\right) + \cdots + f\left(a + (N-1)\frac{b-a}{N}\right) \right)$$

car on a utilisé ici les subdivisions << équidistantes >> dont le pas $\frac{b-a}{N}$ tend vers zéro lorsque N tend vers l'infini, en prenant comme point d'évaluation dans I_j le point a_{j-1} qui est son extrémité gauche.

Exemple et contre-exemple

Si on prend sur $[0, 1]$ la fonction considérée pour la première fois par Dirichlet qui vaut 1 si $x \in \mathbb{Q}$ et 0 sinon, alors les sommes de Riemann ci-dessus valent toutes 1. Mais si entre $a_{j-1} = \frac{j-1}{N}$ et $a_j = \frac{j}{N}$ on choisit ξ_j irrationnel, ce qui est toujours possible (par exemple $\xi_j = \frac{j-1}{N} + \frac{\sqrt{2}-1}{N}$), on obtient d'autres sommes, valant toutes zéro. Cette fonction bizarre n'est donc pas intégrable au sens de Riemann.

Par contre si on modifie un peu l'exemple de Dirichlet on posant $f(x) = 0$ si $x \notin \mathbb{Q}$ et $f(\frac{p}{q}) = \frac{1}{q}$ pour toute fraction irréductible $\frac{p}{q}$, $q \geq 1$, alors il n'est pas trop difficile de voir que f est bien intégrable au sens de Riemann, et d'ailleurs avec $\int_0^1 f(x)\,dx = 0$, bien que $f \geq 0$ et qu'il existe des x partout denses avec $f(x) > 0$. Vous pourrez prouver que f est continue en tout $x \notin \mathbb{Q}$ et discontinue en tout $x \in \mathbb{Q}$. En fait $\lim_{y \to x} f(y) = 0$ pour tout x.

Démonstrations

Regardons ce qui se passe lorsque l'on ajoute un point α à une subdivision $\mathcal{A}$ pour obtenir la subdivision $\mathcal{A}'$. Si α coïncide avec l'un des a_j on n'a rien changé. Sinon on a $a_{j-1} < \alpha < a_j$ pour l'un des j. La somme de Darboux inférieure associée à $\mathcal{A}'$ diffère de celle associée à $\mathcal{A}$ par le fait que $(\alpha - a_{j-1}) \inf_{a_{j-1} \leq x \leq \alpha} f(x) + (a_j - \alpha) \inf_{\alpha \leq x \leq a_j} f(x)$ remplace $(a_j - a_{j-1})m_j$. Or certainement $\inf_{a_{j-1} \leq x \leq \alpha} f(x) \geq m_j$ puisque $m_j = \inf_{a_{j-1} \leq x \leq a_j} f(x)$. De même $\inf_{\alpha \leq x \leq a_j} f(x) \geq m_j$. Donc $S_-(f, \mathcal{A}') \geq S_-(f, \mathcal{A})$. Si on itère on obtient que si $\mathcal{C}$ est obtenue à partir de $\mathcal{A}$ en ajoutant un nombre fini de points alors nécessairement $S_-(f, \mathcal{C}) \geq S_-(f, \mathcal{A})$. Par contre on se convainc que c'est le contraire qui se passe pour les sommes supérieures : $S_+(f, \mathcal{C}) \leq S_+(f, \mathcal{A})$.

Maintenant soient $\mathcal{A}$ et $\mathcal{B}$ deux subdivisions quelconques. Formons $\mathcal{C}$ en combinant les points utilisés dans $\mathcal{A}$ et dans $\mathcal{B}$. Comme $\mathcal{C}$ est << plus fine >> que (ou égale à) $\mathcal{A}$ on a $S_-(f, \mathcal{C}) \geq S_-(f, \mathcal{A})$ car on peut imaginer passer de $\mathcal{A}$ à $\mathcal{C}$ en lui ajoutant un par un des points supplémentaires distincts des précédents. Comme $\mathcal{C}$ est << plus fine >> que $\mathcal{B}$ on a $S_+(f, \mathcal{C}) \leq S_+(f, \mathcal{B})$. Ainsi $S_-(f, \mathcal{A}) \leq S_-(f, \mathcal{C}) \leq S_+(f, \mathcal{C}) \leq S_+(f, \mathcal{B})$ et donc :

$$\forall \mathcal{A}, \mathcal{B} \quad S_-(f, \mathcal{A}) \leq S_+(f, \mathcal{B})$$

Soit alors E le sous-ensemble de $\mathbb{R}$ de toutes les valeurs possibles pour les sommes inférieures $S_-(f, \mathcal{A})$ associées à toutes les subdivisions possibles $\mathcal{A}$ et soit F le sous-ensemble de $\mathbb{R}$ de toutes les valeurs possibles pour les sommes supérieures $S_+(f, \mathcal{B})$ associées à toutes les subdivisions possibles $\mathcal{B}$. Tout élément de F est un majorant de E donc au moins égal à $\sup E$. Ce nombre $\sup E$ est donc un minorant de F donc au plus égal à $\inf F$. En définissant $I_-(f) = \sup E$ et $I_+(f) = \inf F$ on a donc :

$$I_-(f) \leq I_+(f)$$

Soit f une fonction R-intégrable. Elle est donc bornée par hypothèse, on notera $M = \sup f$ et $m = \inf f$. Réexaminons ce qui se passe lorsque l'on ajoute un point α à une subdivision $\mathcal{A}$ pour obtenir $\mathcal{A}'$. On a :

$$S_-(f,\mathcal{A}') - S_-(f,\mathcal{A}) = (\alpha - a_{j-1}) \inf_{a_{j-1}\leq x\leq\alpha} f(x) + (a_j - \alpha) \inf_{\alpha\leq x\leq a_j} f(x) - (a_j - a_{j-1})m_j$$

On minore m_j par m et on majore les deux autres inf par M, ce qui donne :

$$(0 \leq\ ) \ S_-(f,\mathcal{A}') - S_-(f,\mathcal{A}) \leq (a_j - a_{j-1})(M-m) \leq \delta(\mathcal{A})(M-m)$$

Supposons que l'on passe alors de $\mathcal{A}$ à $\mathcal{B}$ en K étapes, on aura (puisque $\delta(\mathcal{A})$ majore les pas de chacune des subdivisions intermédiaires entre $\mathcal{A}$ et $\mathcal{B}$) :

$$S_-(f,\mathcal{B}) \leq S_-(f,\mathcal{A}) + K\delta(\mathcal{A})(M-m)$$

On prouve de même

$$S_+(f,\mathcal{B}) \geq S_+(f,\mathcal{A}) - K\delta(\mathcal{A})(M-m)$$

et donc

$$\Delta(\mathcal{B}) \geq \Delta(\mathcal{A}) - 2K\delta(\mathcal{A})(M-m)$$

lorsque $\mathcal{B}$ est obtenu en ajoutant au plus K points à $\mathcal{A}$.

Soit $\epsilon > 0$. Il existe $\mathcal{C}_1$ avec $I(f) - \epsilon \leq S_-(f,\mathcal{C}_1) \leq I(f)$ et $\mathcal{C}_2$ avec $I(f) \leq S_+(f,\mathcal{C}_2) \leq I(f)+\epsilon$. En combinant $\mathcal{C}_1$ et $\mathcal{C}_2$ en une seule subdivision plus fine $\mathcal{C}$ on aura donc $I(f)-\epsilon \leq S_-(f,\mathcal{C}) \leq S_+(f,\mathcal{C}) \leq I(f)+\epsilon$. Soit K le nombre de points de $\mathcal{C}$. Soit $\mathcal{A}$ quelconque et formons $\mathcal{B}$ en ajoutant à $\mathcal{A}$ les K points de $\mathcal{C}$. On aura

$$\Delta(\mathcal{A}) \leq \Delta(\mathcal{B}) + 2K(M-m)\delta(A) \leq 2\epsilon + 2K(M-m)\delta(A)$$

On a utilisé l'astuce que comme $\mathcal{B}$ est plus fine que $\mathcal{C}$ on a $\Delta(\mathcal{B}) \leq \Delta(\mathcal{C}) \leq 2\epsilon$. En prenant $\delta > 0$ suffisamment petit on peut donc imposer $\Delta(\mathcal{A}) \leq 3\epsilon$ pour tout $\mathcal{A}$ avec $\delta(\mathcal{A}) \leq \delta$, ce qu'il fallait montrer au 3 près qui n'est pas important (on prendra le δ qui marche pour $\epsilon/3$). On remarque que le point crucial dans cette preuve c'est que l'entier K ne dépend que de ϵ, via le choix de $\mathcal{C}$, ce qui est complètement indépendant de $\mathcal{A}$. Le théorème de convergence lorsque le pas tend vers zéro est démontré.

Fonctions monotones

Montrons que toute f croissante sur l'intervalle $[a,b]$ est R-intégrable. Elle est bornée puisque $\forall x\ f(a) \leq f(x) \leq f(b)$. Prenons la subdivision $\mathcal{A}_N$ à pas constant $\frac{b-a}{N}$. Dans chaque I_j comme f est croissante on aura $m_j = f(a_{j-1})$ et $M_j = f(a_j)$ de sorte que :

$$S_-(f,\mathcal{A}_N) = \frac{b-a}{N}\left(f(a) + f\left(a+\frac{b-a}{N}\right) + \cdots + f\left(a+(N-1)\frac{b-a}{N}\right)\right)$$

$$S_+(f,\mathcal{A}_N) = \frac{b-a}{N}\left(f\left(a+\frac{b-a}{N}\right) + \cdots + f\left(a+(N-1)\frac{b-a}{N}\right) + f(b)\right)$$

Donc $\Delta(\mathcal{A}_N) = \frac{b-a}{N}(f(b) - f(a))$. Ainsi $\forall N \geq 1 \quad I_+(f) - I_-(f) \leq \frac{b-a}{N}(f(b) - f(a))$ et en faisant tendre N vers $+\infty$ on obtient $I_+(f) = I_-(f)$. Donc f est bien intégrable au sens de Riemann. On procéderait de même pour f décroissante.

On montrerait alors si l'on voulait que toute fonction $f = g - k$ différence de deux fonctions croissantes sur $[a,b]$ est R-intégrable. Cette remarque anodine pourrait être approfondie car les fonctions obtenues de cette manière ont, comme l'a montré Jordan, une autre caractérisation : celle d'être de « variation bornée ». Et les fonctions monotones par morceaux sur $[a,b]$, ainsi que les fonctions de classe C^1 par morceaux, sont de variation bornée.

Fonctions continues : première approche

Soit f continue sur $[a,b]$ et soit $\epsilon > 0$. Si pour tout x on a $|f(x) - f(a)| < \epsilon$ on pose $a_1 = b$. Sinon il existe x avec $|f(x) - f(a)| \geq \epsilon$ et par le théorème des valeurs intermédiaires il existe y avec $|f(y) - f(a)| = \epsilon$. On prend a_1 égal à l'infimum de tous ces y. En utilisant la continuité de f on constate que a_1 vérifie $|f(a_1) - f(a)| = \epsilon$. Donc $a_1 > a$ et pour tout $x \in [a,a_1]$ on a $|f(x) - f(a)| \leq \epsilon$. Remarquons dès maintenant que $|f(x) - f(y)| \leq 2\epsilon$ si x et y sont tous deux dans $[a,a_1]$. Si $a_1 < b$ on réitère à partir de a_1, obtenant ainsi a_2, a_3, ... Cela peut-il continuer indéfiniment ? Non, car sinon on aurait une suite croissante (a_j) donc convergente. Soit L sa limite. Comme f est continue et que $|f(a_{j+1}) - f(a_j)| = \epsilon$ on obtient en passant à la limite $|f(L) - f(L)| = \epsilon$. Contradiction. Donc au bout d'un nombre fini d'étapes on finit par avoir $a_N = b$. Remarquons que cela permet de voir que la fonction continue f est bornée sur l'intervalle $[a,b]$ puisque $\forall x\, |f(x) - f(a)| \leq N\epsilon$.

Les points a_j définissent une subdivision $\mathcal{A}$. Majorons $\Delta(\mathcal{A})$: dans chaque intervalle I_j de la subdivision on a $f(a_{j-1}) - \epsilon \leq f(x) \leq f(a_{j-1}) + \epsilon$, donc $f(a_{j-1}) - \epsilon \leq m_j \leq M_j \leq f(a_{j-1}) + \epsilon$ donc $M_j - m_j \leq 2\epsilon$. Ainsi :

$$\Delta(\mathcal{A}) \leq 2(b-a)\epsilon$$

Cela prouve $I_+(f) - I_-(f) \leq 2(b-a)\epsilon$ mais comme $\epsilon > 0$ est arbitraire, c'est que $I_+(f) = I_-(f)$. La fonction continue f est bien intégrable au sens de Riemann.

Fonctions continues : continuité uniforme

On peut utiliser notre construction pour montrer que la fonction f est *uniformément continue* :

$$\forall\epsilon > 0\ \exists\delta > 0 \quad \forall x,y \in [a,b] \quad |x-y| \leq \delta \Longrightarrow |f(x) - f(y)| \leq \epsilon$$

Il suffira de considérer les points a_j associés à $\frac{1}{3}\epsilon$. Soit $\lambda = \min_j(a_j - a_{j-1})$, qui est > 0. Posons $\delta = \frac{1}{2}\lambda$. Si $0 \leq y - x \leq \delta$, il n'y a aucun ou un seul indice j avec $x \leq a_j \leq y$. Dans le premier cas on a, comme vu précédemment, $|f(x) - f(y)| \leq 2\frac{1}{3}\epsilon$, et dans le deuxième $|f(x) - f(y)| \leq |f(x) - f(a_j)| + |f(a_j) - f(y)| \leq 2\frac{1}{3}\epsilon + \frac{1}{3}\epsilon = \epsilon$.

Il y a une formulation équivalente de la continuité uniforme : si $x_n - y_n \to 0$ alors $f(x_n) - f(y_n) \to 0$. C'est un bon exercice que de passer d'une formulation à l'autre. La formulation avec les suites est utile pour montrer qu'une fonction n'est pas uniformément continue : par exemple avec $x_n = \frac{1}{n}$ et $y_n = \frac{2}{n}$ on voit que $f(x) = \frac{1}{x}$ sur l'intervalle $]0,1]$ n'est pas uniformément continue. De même la fonction $x\sin(x)$ sur l'intervalle fermé mais non borné $[0,+\infty[$ n'est pas uniformémement continue : on prendra $x_n = n\pi$ et $y_n = (n+\frac{1}{n})\pi$.

Réitérons cet énoncé fondamental : *toute fonction continue sur un intervalle fermé borné* $[a,b]$ *est uniformément continue,* mais ce n'est plus nécessairement le cas si l'intervalle n'est pas un segment.

Une fois connue la continuité uniforme on peut en déduire (c'est ce que les gens font habituellement) la R-intégrabilité. En effet notons

$$\omega_N = \sup\{|f(u) - f(v)| \ : \quad |u - v| \leq \frac{b-a}{N},\ u, v \in [a,b]\} \ .$$

On a bien sûr $\omega_N \leq 2\sup|f|$. Si on prend la subdivision régulière de pas $(b-a)/N$, l'écart Δ_N entre sommes supérieure et inférieure est majoré par $(b-a)\omega_N$. Or la continuité uniforme équivaut justement au fait que $\lim\omega_N = 0$. Donc f est R-intégrable.

Démontrons à nouveau cette fameuse continuité uniforme. La suite ω_N est décroissante. Si elle ne tend pas vers zéro, elle est minorée par un $\epsilon > 0$. Donc, on peut prendre pour chaque $N \geq 1$ un couple (u_N, v_N) avec $|u_N - v_N| \leq \frac{b-a}{N}$ et $|f(u_N) - f(v_N)| \geq \frac{1}{2}\epsilon$. On utilise alors le théorème de Bolzano-Weierstrass qui affirme qu'il existe un $x \in [a,b]$ et une suite extraite (u_{N_k}) de limite x. Donc $\lim v_{N_k} = x$ aussi puis $\lim|f(u_{N_k}) - f(v_{N_k})| = |f(x) - f(x)| = 0$, par continuité de f au point x. Contradiction.

Relation de Chasles, linéarité, positivité

Nous avons donc presque tout démontré de ce qui était annoncé. Supposons que f soit R-intégrable sur $[a,b]$ et soit $[c,d] \subset [a,b]$. Certainement f est à nouveau bornée sur $[c,d]$. Soit $\epsilon > 0$ et soit $\mathcal{A}$ une subdivision de $[a,b]$ telle que l'écart entre sa somme supérieure et sa somme inférieure soit au plus ϵ. On peut ajouter les points c et d à $\mathcal{A}$ ce qui ne peut que diminuer cet écart. Finalement soit $\mathcal{B}$ la subdivision de $[c,d]$ obtenue en ne retenant des points de $\mathcal{A}$ que ceux dans cet intervalle. L'écart entre la somme inférieure et la somme supérieure pour $\mathcal{B}$ sur l'intervalle $[c,d]$ est majoré par l'écart pour $\mathcal{A}$ sur $[a,b]$ donc par ϵ. Comme ϵ est arbitraire les intégrales inférieure et supérieure de f sur $[c,d]$ coïncident, ce qu'il fallait montrer.

Si on suppose que f est R-intégrable sur $[a,b]$ et sur $[b,c]$ alors d'abord elle est clairement aussi bornée sur $[a,c]$ $(a < b < c)$. Puis en prenant une subdivision de $[a,c]$ contenant le point b et suffisamment fine dans chacun des sous-intervalles $[a,b]$ et $[b,c]$ on rend l'écart entre sommes inférieure et supérieure sur $[a,c]$ arbitrairement petit. Donc f est R-intégrable sur $[a,c]$ tout entier. En ce qui concerne la formule $\int_a^c f(x)dx = \int_a^b f(x)dx + \int_b^c f(x)dx$

pour a < b < c il suffit de prendre des sommes de Riemann pour des subdivisions contenant le point intermédiaire b et de passer à la limite lorsque que le pas tend vers zéro.

Chasles : On conviendra que $\int_a^a f(x)dx = 0$ quelque soit f (définie au point a). Et on posera $\int_b^a f(x)dx = -\int_a^b f(x)dx$ si $b \geq a$. On a alors pour tout a, b, c, contenus dans un intervalle où f est R-intégrable, quel que soit l'ordre :

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

ce qui constitue la Relation de Chasles.

Linéarité : Si f est R-intégrable alors il est à peu près évident que pour tout $\lambda \in \mathbb{R}$ la fonction λf est R-intégrable et $\int_a^b \lambda f(x)dx = \lambda \int_a^b f(x)dx$. De plus, si f et g sont R-intégrables alors $f+g$ l'est aussi. Elle est certainement bornée. Ensuite soit $\epsilon > 0$. Prenons $\mathcal{A}$ avec $\Delta_f(\mathcal{A}) \leq \epsilon$ (notation auto-explicative) et $\mathcal{B}$ avec $\Delta_g(\mathcal{B}) \leq \epsilon$. Les combinant en une seule subdivision $\mathcal{C}$ on aura à la fois $\Delta_f(\mathcal{C}) \leq \epsilon$ et $\Delta_g(\mathcal{C}) \leq \epsilon$. Sur le j-ième intervalle on a $m_j(f) \leq f(x) \leq M_j(f)$ et $m_j(g) \leq g(x) \leq M_j(g)$ donc $m_j(f) + m_j(g) \leq f(x) + g(x) \leq M_j(f) + M_j(g)$ donc $m_j(f+g) \geq m_j(f) + m_j(g)$ et $M_j(f+g) \leq M_j(f) + M_j(g)$ donc $M_j(f+g) - m_j(f+g) \leq (M_j(f) - m_j(f)) + (M_j(g) - m_j(g))$ donc $\Delta_{f+g}(\mathcal{C}) \leq \Delta_f(\mathcal{C}) + \Delta_g(\mathcal{C}) \leq 2\epsilon$. Comme ϵ est arbitraire c'est que $f+g$ est R-intégrable. En utilisant une somme de Riemann on obtient :

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

Positivité et Monotonie : Si f est R-intégrable et à valeurs positives ou nulles alors son intégrale est positive ou nulle (propriété de p o s i t i v i t é) : immédiat en écrivant l'intégrale comme une limite de sommes de Riemann. Plus généralement :

$$\forall x\ f(x) \geq g(x) \quad \Longrightarrow \int_a^b f(x)dx \geq \int_a^b g(x)dx$$

On appelle cela la propriété de m o n o t o n i e de l'intégrale de Riemann. Attention : $a \leq b$ dans cette inégalité ! On utilise souvent la conséquence (attention ici aussi $a \leq b$) :

$$\forall x \quad m \leq f(x) \leq M \quad \Longrightarrow m(b-a) \leq \int_a^b f(x)dx \leq M(b-a)$$

Valeurs absolues : Si f est R-intégrable alors $|f|$ l'est aussi. Preuve : On va utiliser pour cela l'inégalité $\left||x| - |y|\right| \leq |x-y|$ et aussi le fait que pour toute fonction bornée (prouvez-le !) :

$$\sup_{\alpha \leq x,y \leq \beta} |f(x) - f(y)| = (\sup_{\alpha \leq x \leq \beta} f(x)) - (\inf_{\alpha \leq x \leq \beta} f(x))$$

Vous en déduirez

$$\sup_{\alpha \leq x \leq \beta} |f(x)| - \inf_{\alpha \leq x \leq \beta} |f(x)| \leq \sup_{\alpha \leq x \leq \beta} f(x) - \inf_{\alpha \leq x \leq \beta} f(x)$$

puis pour toute subdivision $\Delta_{|f|}(\mathcal{A}) \leq \Delta_f(\mathcal{A})$. Conclure.

Comme la fonction $|f| - f$ est à valeurs positives ou nulles on a $\int_a^b |f(x)|dx \geq \int_a^b f(x)dx$ et comme la fonction $|f| + f$ est à valeurs positives ou nulles on a $\int_a^b |f(x)|dx \geq -\int_a^b f(x)dx$ d'où :

$$a \leq b \Longrightarrow \left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx$$

La formule à utiliser si on ne sait pas $a \leq b$ est : $\left|\int_a^b f(x)dx\right| \leq \left|\int_a^b |f(x)|dx\right|$.

Fonctions à valeurs complexes

Si f est à valeurs complexes, on l'écrit $f = u + iv$ avec u et v ses parties réelle et imaginaire. On dira que f est R-intégrable si u et v le sont et, bien sûr, on posera : $\int_a^b f(x)\,dx = \int_a^b u(x)\,dx + i\int_a^b v(x)\,dx$. L'inégalité :

$$a \leq b \Longrightarrow \left|\int_a^b f(x)dx\right| \leq \int_a^b |f(x)|dx$$

est alors valable. Pour la preuve, on commence d'abord par montrer que l'intégrale est $\mathbb{C}$-linéaire :

$$\int_a^b z\,f(t)\,dt = z\int_a^b f(t)\,dt$$

Ceci se vérifie en développant $(x+iy)(u+iv) = xu - yv + i(xv+yu)$, etc. ... Prenons ensuite w un nombre complexe de module 1, avec $w\int_a^b f(t)\,dt \in [0,+\infty[$. Alors $|\int_a^b f(t)\,dt| = |w\int_a^b f(t)\,dt| = |\mathrm{Re}(w\int_a^b f(t)\,dt)| = |\mathrm{Re}(\int_a^b wf(t)\,dt)| = |\int_a^b \mathrm{Re}(wf(t))\,dt| \leq \int_a^b |\mathrm{Re}(wf(t))|\,dt \leq \int_a^b |wf(t)|\,dt = \int_a^b |f(t)|\,dt$. Où l'on a utilisé entre autres la monotonie de l'intégrale.

Cependant on a triché car on a admis comme un fait évident que $|f|$ était R-intégrable. Pour le prouver, il suffit d'observer la chose suivante :

$$\Big||f(x)| - |f(y)|\Big| \leq |f(x) - f(y)| \leq |u(x) - u(y)| + |v(x) - v(y)|$$

Donc, si x et y sont dans l'intervalle I_j d'une subdivision :

$$\Big||f(x)| - |f(y)|\Big| \leq M_j(u) - m_j(u) + M_j(v) - m_j(v)$$

Comme on l'a déjà fait remarquer on a : $M_j(|f|) - m_j(|f|) = \sup_{x,y\in I_j}\Big||f(x)| - |f(y)|\Big|$. Au final, pour toute subdivision :

$$\Delta_{|f|}(\mathcal{A}) \leq \Delta_u(\mathcal{A}) + \Delta_v(\mathcal{A})$$

Donc $|f|$ est bien R-intégrable si f l'est.

Produits

Théorème : *Si* f *et* g *sont toutes deux R-intégrables sur l'intervalle* $[a, b]$ *alors il en est de même de* fg.

Tout d'abord ce produit est bien borné. Ensuite, nous commençons par observer qu'il suffit de montrer que la fonction $x \mapsto (f(x) + C)g(x)$ est R-intégrable, pour une constante C choisie arbitrairement. En effet $f(x)g(x) = (f(x) + C)g(x) - Cg(x)$ et on sait déjà que toute combinaison linéaire de fonctions R-intégrables est R-intégrable. On prendra C de sorte que $\forall x\ f(x) + C \geq 0$. De même il suffit de montrer que $(f(x) + C)(g(x) + D)$ est R-intégrable, avec un D quelconque : on le prendra de sorte que $\forall x\ g(x) + D \geq 0$. On notera que tout cela est possible parce que par hypothèse les fonctions f et g sont bornées (ce qui fait partie des conditions pour être R-intégrable). En fin de compte, quitte à remplacer f par $f + C$ et g par $g + D$, on pourra supposer $f \geq 0$ et $g \geq 0$. Sur le j^e intervalle d'une subdivision $\mathcal{A}$ quelconque :

$$0 \leq m_j(f) \leq f(x) \leq M_j(f) \qquad 0 \leq m_j(g) \leq g(x) \leq M_j(g)$$

Ainsi (notez bien que c'est grâce aux $0 \leq \ldots$ que cela est valable) :

$$m_j(f)m_j(g) \leq f(x)g(x) \leq M_j(f)M_j(g)$$

d'où :

$$m_j(f)m_j(g) \leq m_j(fg) \leq M_j(fg) \leq M_j(f)M_j(g)$$

Or $M_j(f)M_j(g) - m_j(f)m_j(g) = (M_j(f) - m_j(f))M_j(g) + m_j(f)(M_j(g) - m_j(g))$ donc

$$0 \leq M_j(fg) - m_j(fg) \leq (M_j(f) - m_j(f)) \sup(g) + (M_j(g) - m_j(g)) \sup(f)$$

On a, bien sûr, noté sup(f) et sup(g) les bornes supérieures respectives de f et de g sur $[a, b]$. Si on prend maintenant $\mathcal{A}$ de sorte que $\Delta_f(\mathcal{A}) \leq \epsilon$ et $\Delta_g(\mathcal{A}) \leq \epsilon$, on en déduit $\Delta_{fg}(\mathcal{A}) \leq (\sup(f) + \sup(g))\epsilon$. Cela montre que l'on peut choisir $\mathcal{A}$ de sorte à rendre $\Delta_{fg}(\mathcal{A})$ arbitrairement petit : autrement dit on a prouvé la R-intégrabilité de la fonction fg.

Divers

Si on modifie f en un nombre fini de points elle reste R-intégrable et $\int_a^b f(x)dx$ reste identique ! Je vous laisse cette affirmation comme un excellent exercice. Je laisse également l'énoncé qui suit en exercice :

Théorème : *si* f *est bornée sur* $[a, b]$ *et R-intégrable sur chaque* $[a + \eta, b]$ *(*$\eta > 0$*) alors elle est R-intégrable sur* $[a, b]$.

Fonctions en escalier

On dit que f est en escalier si on peut trouver une subdivision $\mathcal{A}$ telle que f soit constante sur chaque $]x_{j-1}, x_j[$, $1 \leq j \leq N$. Les valeurs de f aux x_j sont arbitraires. Par la relation de Chasles, la fonction f est intégrable au sens de Riemann et $I(f) = \int_a^b f(x)dx = \sum_j (x_j - x_{j-1})f(\xi_j)$ où ξ_j est choisi arbitraire dans $]x_{j-1}, x_j[$.

Soit f une fonction R-intégrable, soit $\mathcal{A}$ une subdivision quelconque. On considère la fonction en escalier U qui vaut $m_j = \inf_{[x_{j-1},x_j]} f(x)$ sur $]x_{j-1}, x_j[$ et vaut $\inf_{[a,b]} f$ aux points x_j. Cette fonction en escalier U vérifie $\forall x\ U(x) \leq f(x)$. De plus $\int_a^b U(x)dx$ est exactement identique avec la somme de Darboux inférieure $S_-(\mathcal{A}, f)$. De même si on définit V comme valant $M_j = \sup_{[x_{j-1},x_j]} f(x)$ sur $]x_{j-1}, x_j[$ et $\sup_{[a,b]} f$ aux points x_j, alors $\forall x\ f(x) \leq V(x)$ et $S_+(\mathcal{A}, f) = \int_a^b V(x)\, dx$. On peut donc, si f est R-intégrable, trouver pour tout $\epsilon > 0$ donné U et V en escaliers telles que $\forall x\ U(x) \leq f(x) \leq V(x)$ et $\int_a^b (V(x) - U(x))\, dx \leq \epsilon$.

Réciproquement, si $U \leq f$ alors certainement les sommes de Darboux inférieures pour U minorent celles pour f donc $\int_a^b U(x)dx \leq I_-(f)$, et si par ailleurs $V \geq f$ alors les sommes de Darboux supérieures pour V majorent celles pour f donc $\int_a^b V(x)dx \geq I_+(f)$. Donc $I_+(f) - I_-(f) \leq \int_a^b V(x)dx - \int_a^b U(x)dx$ à chaque fois que $U \leq f \leq V$. Si on peut rendre $\int_a^b (V(x) - U(x))\, dx$ plus petit que tout $\epsilon > 0$ c'est donc que f est R-intégrable. On a donc caractérisé les fonctions R-intégrales (réelles) comme étant les fonctions que l'on peut encadrer par deux fonctions en escalier U et V de sorte que $\int_a^b (V(x) - U(x))\, dx$ soit arbitrairement petit.

Limites uniformes

Théorème : *si f est la limite uniforme d'une suite de fonctions f_n R-intégrables alors elle est R-intégrable et* $\int_a^b f(t)\, dt = \lim \int_a^b f_n(t)\, dt$.

En effet pour tout $\epsilon > 0$ et pour $N \gg 1$ on a $f_N - \epsilon \leq f \leq f_N + \epsilon$ et $\int_a^b (f_N(x) + \epsilon) - (f_N(x) - \epsilon)\, dx = 2(b - a)\epsilon$ est arbitrairement petit. On imite alors la méthode de preuve de la section précédente. Et $\left|\int_a^b f(t)\, dt - \int_a^b f_n(t)\, dt\right| \leq (b - a) \sup_t |f(t) - f_n(t)|$.

Composition avec une fonction continue

Supposons que $f : [a, b] \dashrightarrow [m, M]$ soit R-intégrable. Soit $g : [m, M] \dashrightarrow \mathbb{R}$ continue.

Théorème : *la fonction composée* $g \circ f$ *est R-intégrable.*

Soit $\epsilon > 0$. En utilisant la continuité uniforme de g on peut montrer qu'il existe k continue, affine par morceaux, avec $|k(t) - g(t)| \leq \epsilon$ sur $[m, M]$. Il existe une constante C avec $|k(t) - k(u)| \leq C|t - u|$ pour tous les t, u. Un petit instant de réflexion montre qu'alors $M_j(k \circ f) - m_j(k \circ f) \leq C(M_j(f) - m_j(f))$ pour toute subdivision, donc $\Delta(k \circ f) \leq C\Delta(f)$. Donc $k \circ f$ est R-intégrable.

Mais $|k(f(x)) - g(f(x))| \leq \epsilon$ pour tout x, et $\epsilon > 0$ est arbitraire. Donc $g \circ f$ est R-intégrable. Je laisse au lecteur enthousiaste le cas où f est à valeurs complexes !

Les théorèmes fondamentaux du Calcul

La fonction $\mathbb{1}_{[a,x]}(t)$ qui vaut 1 pour $a \leq t \leq x$ et 0 pour $t > x$ (avec x fixé, dans l'intervalle $[a, b]$), est R-intégrable, c'est une fonction en escalier. Pour toute fontion f R-intégrable sur $[a, b]$, sa restriction à $[a, x]$ est aussi R-intégrable et $\int_a^x f(t)dt = \int_a^b \mathbb{1}_{[a,x]}(t)f(t)dt$.

Théorème : *si f est R-intégrable sur $[a, b]$, alors l'intégrale indéfinie $x \mapsto \int_a^x f(t)dt$ est une fonction continue de x.*(*)

En effet, en la notant F(x), on a $|F(x) - F(y)| \leq (\sup|f|)|x - y|$. Elle est donc même Lipschitzienne. Le Théorème suivant est peut-être plus fondamental encore : (*)

Théorème : *Si f est R-intégrable*(**) *sur $[a, b]$ et si x_0 est un point de continuité de f alors la fonction*

$$F(x) = \int_a^x f(t)dt$$

est dérivable au point x_0 et on a :

$$F'(x_0) = f(x_0)$$

J'en laisse la démonstration en exercice, tellement cette preuve est importante ! Son corollaire (existence de primitives) est souvent appelé << Théorème fondamental du calcul >> :

Théorème fondamental du Calcul : *Si f est continue sur $[a, b]$ alors la fonction*

$$F(x) = \int_a^x f(t)dt$$

est dérivable sur $[a, b]$ et on a :

$$\forall x \in [a, b] \quad F'(x) = f(x)$$

(*) bon exercice pour M312 : cela vaut aussi pour f Lebesgue-intégrable.

(*) M312 : si f est intégrable au sens de Lebesgue sur $[a, b]$ alors la fonction $x \mapsto F(x) = \int_a^x f(t)dt$ a la propriété d'être presque partout dérivable avec $F'(x) = f(x)$. Ce théorème de Lebesgue est, certainement, l'un des plus importants de la théorie de l'intégration. Nous en donnerons une preuve dans une autre annexe au Cours.

(**) exercice pour M312 : cela vaut aussi pour f L-intégrable

Autrement dit, lorsque f est continue :

$$\frac{d}{dx}\int_a^x f(t)dt = f(x)$$

On notera soigneusement le signe moins qui apparaît lorsque l'on dérive par rapport à l'autre borne d'intégration :

$$\frac{d}{dx}\int_x^b f(t)dt = -f(x)$$

Avec les définitions à l'oeuvre dans la formule de Chasles, ces relations sont vraies que x soit inférieur ou supérieur à a ou à b, sous la contrainte bien sûr que f soit définie et continue sur tout l'intervalle allant de x à l'autre borne fixée.

Nous voyons donc que toute fonction continue admet une primitive. Cela peut être utilisé pour construire la fonction logarithme : $\log(x) = \int_1^x \frac{1}{t}dt$.... Mais comme vous le savez pour avoir passé des nuits à faire des centaines de calculs d'intégrales, on utilise la plupart du temps ce théorème dans le sens contraire : pour calculer l'intégrale d'une fonction on en recherche une primitive, par exemple en consultant des tables de dérivées. La formule fondamentale est alors :

$$\text{si } F' = f : \quad \int_a^b f(t)dt = [F(t)]_a^b = F(b) - F(a)$$

Pour la preuve à partir du Théorème fondamental, on remarque que la formule est certainement vraie pour la primitive particulière $F(x) = \int_a^x f(t)dt$, et donc pour toute primitive.

Cet argument suppose la fonction f continue. En fait la formule est valable sous la seule hypothèse que $f = F'$ est Riemann-intégrable :

Deuxième Théorème fondamental du Calcul : *si* f *est R-intégrable sur* [a, b] *et si elle admet une primitive* F *alors :*

$$\int_a^b f(t)dt = [F(t)]_a^b = F(b) - F(a)$$

De manière équivalente : si la fonction dérivable F *a une dérivée* $f = F'$ *qui est R-intégrable alors la formule vaut.*

Preuve : soit $N \geq 1$ et posons $x_j = a + j\frac{b-a}{N}$ pour $0 \leq j \leq N$. On écrit $F(b) - F(a) = (F(x_1) - F(x_0)) + (F(x_2) - F(x_1)) + \cdots + (F(x_N) - F(x_{N-1}))$, et on applique le théorème des accroissements finis à chaque terme ; on voit donc apparaître une somme de Riemann pour $f = F'$. On fait tendre N vers l'infini, et on utilise l'hypothèse que f est R-intégrable pour conclure.

** FIN DU RAPPEL DE PREMIÈRE ANNÉE **

Un mot bref : si F est dérivable sur $[a,b]$ et si $f = F'$ est Lebesgue-intégrable alors $\int_a^b f(t)dt = [F(t)]_a^b = F(b) - F(a)$. Voir le livre de Rudin, << Analyse réelle et complexe >> pour une démonstration.

Par ailleurs il a manqué au présent exposé les deux formules de changement de variable. Et les formules de la moyenne. Et l'intégration par parties...

Caractérisation des fonctions R-intégrables

Il est assez phénoménal que le premier théorème donné par Riemann, immédiatement après la définition de son intégrale, fut le résultat subtil qui caractérise parmi les fonctions bornées celles qui sont intégrables. Je reproduis en annexe le texte de Riemann et sa traduction. Ici je vais montrer l'énoncé ultérieur (1903) dû à Lebesgue :

> Pour qu'une fonction bornée soit R-intégrable il est nécessaire et suffisant que l'ensemble de ses points de discontinuité soit de mesure nulle.

On travaille avec une fonction bornée f sur un intervalle borné $[a,b]$. Notons A l'ensemble des points de discontinuité de f. L'oscillation $\omega(f,I)$ de f sur un intervalle I (intersecté avec $[a,b]$, et dans la suite ce genre de détail sera laissé tacitement à l'entendement du lecteur) est $\sup_{u,v\in I}|f(u) - f(v)|$. Soit $\omega_f(x)$ l'oscillation au point x : $\omega_f(x) = \lim_{r\to 0^+}\omega(f,]x-r,x+r[)$, limite qui existe par monotonie en r. Il est clair (si si) que f est continue au point x si et seulement si $\omega_f(x) = 0$. Soit $A_n = \{x : \omega_f(x) \geq \frac{1}{n}\}$. L'ensemble A des points de discontinuité de f est la réunion des A_n.

Supposons f R-intégrable. Soit $n \geq 1$ fixé. Soit N grand, et considérons la partition de $[a,b]$ en les $I_k =]a + (b-a)\frac{k-1}{N}, a + (b-a)\frac{k}{N}[$, $1 \leq k \leq N$ et les points $a + (b-a)\frac{k}{N}$, $0 \leq k \leq N$. Il y a M indices k pour lesquels I_k rencontrent A_n. L'écart sur I_k entre l'infimum et le supremum de f est au moins $\frac{1}{n}$. L'écart $\Delta(N)$ entre les sommes (de Darboux) supérieure et inférieure pour cette subdivision est au moins $M\frac{1}{n}\frac{b-a}{N}$. La somme $M\frac{b-a}{N}$ des longueurs de ces M intervalles est donc majorée par $n\Delta(N)$. L'ensemble A_n à un nombre fini de points près est inclus dans cette union d'intervalles donc sa mesure extérieure est au plus $n\Delta(N)$. Lorsque N tend vers l'infini, $\Delta(N)$ tend vers zéro. Donc A_n est de mesure nulle. Donc, A est de mesure nulle.

Nous montrons la réciproque. On suppose que A est de mesure nulle. C'est donc aussi le cas de chaque A_n. Notons tout d'abord que A_n est fermé, comme conséquence du fait suivant laissé en exercice : $\omega_f(x) \geq \limsup_{y\to x}\omega_f(y)$.

Ensuite, pour tout fermé B négligeable et tout $\epsilon > 0$, on peut trouver un nombre *fini* d'intervalles ouverts J_k *disjoints* recouvrant B et avec $\sum_k |J_k| \leq \epsilon$. En effet on a en tout cas par définition une collection dénombrable J_k d'intervalles ouverts recouvrant B et avec $\sum_k |J_k| \leq \epsilon$. Tout d'abord on élimine tous les J_k d'intersection vide avec B. Puis si $J_2 \cap J_1 \neq \emptyset$, on remplace J_1 par $J_1 \cup J_2$, ce qui ne fait que diminuer la somme des longueurs. On fait cela avec tous les $k \geq 2$. S'il subsiste un J_k, k minimal, on itère le processus avec lui jouant le rôle de J_1. Etc. Au final les J_k sont disjoints (on reconnaît les composantes connexes de l'ouvert initial $\bigcup J_k$). Montrons qu'il n'y en a pas plus qu'un nombre fini K. Sinon, on choisit $x_k \in B \cap J_k$. Par Bolzano-Weierstrass la suite (x_k) admet une valeur d'adhérence x, qui est donc dans B. Il y a précisément un unique J_n qui contient ce x. Les x_k avec $k \neq n$ ne peuvent pas être dans J_n car $J_n \cap J_k = \emptyset$. Contradiction.

On applique ce qui précède à $B = A_n$. On peut supposer les J_k ordonnés de la gauche vers la droite, et s'il y a des extrémités communes on les fusionne. Soit X le complémentaire de l'union de ces J_k, qui est donc une union finie d'intervalles fermés disjoints. Notons génériquement I_p ces intervalles fermés dont la réunion fait X. S'il existe un ou plusieurs des I_p avec $\omega(f, I_p) > \frac{2}{n}$ alors on les divise en deux. Et on recommence. Si l'on peut continuer indéfiniment, c'est qu'il existe dans X des $u_n < v_n$ avec $|u_n - v_n| \to 0$ et $|f(u_n) - f(v_n)| \geq \frac{2}{n}$. Quitte à passer à une suite extraite on aura une limite x, nécessairement dans X, et $\omega_f(x)$ sera au moins $\frac{2}{n}$, ce qui contredit $x \notin A_n$. Donc, au final on a recouvert X par un nombre fini d'intervalles fermés I_p avec $\omega(f, I_p) \leq \frac{2}{n}$. Pour la subdivision de $[a, b]$ formée par les extrémités de tous nos intervalles, la différence entre les sommes inférieure et supérieure sera au plus $2 \sup |f| \epsilon + \frac{2}{n}(b - a)$, donc au plus $(2 \sup |f| + 1)\epsilon$ si l'on prend n suffisamment grand. Cela montre que f est intégrable au sens de Riemann.

F o n c t i o n s r é g l é e s

Les fonctions en escalier sont R-intégrables et les limites uniformes de fonctions R-intégrables sont R-intégrables. On dira que f est << réglée >> si elle est la limite uniforme d'une suite de fonctions en escalier. Toute fonction réglée est donc R-intégrable.

T h é o r è m e : *pour qu'une fonction* f *sur* $[a, b]$ *soit réglée il est nécessaire et suffisant qu'elle ait en tout point une limite à droite et une limite à gauche.*

Preuve : montrons que la condition est nécessaire. Soit $x < b$, soit $\epsilon > 0$ et soit U en escalier avec $\forall y \quad |f(y) - U(y)| \leq \epsilon$. Il existe $\eta > 0$ tel que U soit constante sur $]x, x + \eta]$. Donc $|f(y) - f(y')| \leq 2\epsilon$ pour y et y' dans $]x, x + \eta]$. Par le critère de Cauchy la limite $\lim_{\substack{y \to x \\ y > x}} f(y)$ existe. Idem pour les limites à gauche si $x > a$.

La condition est suffisante : Soit $\epsilon > 0$. Pour chaque x, soit $\eta_x > 0$ tel que $|f(y) - f(y')| \leq \epsilon$ est vrai si $x - \eta_x \leq y < y' < x$ et aussi si $x < y < y' \leq x + \eta_x$. Le compact $[a, b]$ est recouvert par un nombre fini de tels intervalles $]x_j - \eta_j, x_j + \eta_j[$

($\cap[a,b]$). Considérons tous ceux parmi les points $x_j-\eta_j, x_j, x_j+\eta_j$ qui sont dans $[a,b]$, ils forment avec a et b une subdivision de $[a,b]$. Supposons que le segment $[y_1, y_2]$ ne rencontre aucun de ces points. Si l'on choisit x_j de sorte que $y_1 \in]x_j-\eta_j, x_j+\eta_j[$, alors soit $y_1 < x_j$ et alors $x_j-\eta_j < y_1 < y_2 < x_j$, soit $y_1 > x_j$ et alors $x_j+\eta_j > y_2 > y_1 > x_j$. Dans tous les cas on aura $|f(y_1) - f(y_2)| \leq \epsilon$. Par conséquent si dans chacun des intervalles ouverts successifs de la subdivision nous prenons un point ξ arbitraire et définissons la fonction en escalier U comme valant $f(\xi)$ sur cet intervalle, et si nous posons en outre $U(y) = f(y)$ pour les points décidant de la subdivision, on aura $|U(y) - f(y)| \leq \epsilon$ pour tout $y \in [a,b]$.

En prenant $\epsilon = \frac{1}{n}$ et les U_n ainsi construites on obtient une suite de fonctions en escalier, convergeant uniformément sur $[a,b]$ vers f.

Lemme de Riemann-Lebesgue, étages et escaliers

Exercice : en utilisant que pour tout $\epsilon > 0$ on peut trouver g en escalier avec $\int_a^b |f(x) - g(x)|dx < \epsilon$ prouver le Lemme de Riemann :

$$\lim_{\lambda \to \pm\infty} \int_a^b e^{i\lambda x} f(x)dx = 0$$

pour toute fonction f R-intégrable.

Dans la théorie de Lebesgue, presque par définition, on peut trouver des ϕ et ψ étagées encadrant f mesurable bornée et telles que $\int_a^b(\psi(x) - \phi(x))dx \leq \epsilon$. Mais les fonctions étagées, combinaisons linéaires finies des fonctions indicatrices de parties mesurables sont plus générales que les fonctions en escalier. La surprise c'est que comme nous le voyons en cours, bien qu'il soit faux (puisque cela est caractéristique des fonctions R-intégrables) que l'on puisse toujours encadrer f Lebesgue mesurable bornée par des fonctions en escalier proches au sens L^1 il n'en reste pas moins vrai que l'on peut trouver des fonctions en escalier U avec $\int_a^b |f(x) - U(x)|dx$ arbitrairement petit. Cela permet d'étendre le Lemme de Riemann aux fonctions Lebesgue-intégrables. Dernier mot pour les étudiants de Lebesgue : par pitié, ne confondez pas mesurabilité et intégrabilité !

Voici maintenant, d'abord en allemand, puis en traduction, des extraits du Mémoire d'Habilitation de Riemann, 1854, parution posthume en 1867. Sont concernés :

1. la notion de série absolument convergente et le théorème sur les séries semi-convergentes,
2. la définition de l'intégrale de Riemann,
3. la caractérisation donnée par Riemann des fonctions intégrables.

E x t r a i t s d u M é m o i r e d e R i e m a n n

Ueber die Darstellbarkeit einer Function durch eine trigonometrische Reihe.(*)

Bernhard Riemann

[Aus dem dreizehnten Bande der Abhandlungen der Königlichen Gesellschaft der Wissenschaften zu Göttingen.](**)

[⋮]

3.

Erst in Januar 1829 erschien im Journal von *Crelle*(***) eine Abhandlung von *Dirichlet*, worin für Functionen, die durchgehends eine Integration zulassen und nicht unendlich viele Maxima und Minima haben, die Frage ihrer Darstellbarkeit durch trigonometrische Reihen in aller Strenge entschieden wurde.

Die Erkenntniss des zur Lösung dieser Aufgabe einzuschlagenden Weges ergab sich ihm aus der Einsicht, dass die unendliche Reihen in zwei wesentlich verschiedene Klassen zerfallen, je nachdem sie, wenn man sämmtliche Glieder positiv macht, convergent bleiben oder nicht. In den ersteren können die Glieder beliebig versetzt werden, der Werth der letzteren dagegen ist von der Ordnung der Glieder abhängig. In der That, bezeichnet man in einer Reihe zweiter Klasse die positiven Glieder der Reihe nach durch

$$a_1, a_2, a_3, \ldots,$$

die negativen durch

$$-b_1, -b_2, -b_3, \ldots,$$

so ist klar, dass sowohl $\sum a$, als $\sum b$ unendlich sein müssen; denn wären beide endlich, so würde die Reihe auch nach Gleichmachung der Zeichen convergiren; wäre aber *eine* unendlich, so würde die Reihe divergiren. Offenbar kann nun die Reihe durch geeignete Anordnung der Glieder einen beliebig gegebenen Werth C erhalten. Denn nimmt man abwechselnd so lange positive Glieder der Reihe, bis

(*) Transcription récupérée sur http://www.maths.tcd.ie/pub/HistMath/People/Riemann/Papers.html

(**) Diese Abhandlung ist im Jahre 1854 von dem Verfasser behufs seiner Habilitation an der Universität zu Göttingen der philosophischen Facultät eingereicht. Wiewohl der Verfasser ihre Veröffentlichung, wie es scheint, nicht beabsichtigt hat, so wird doch die hiermit erfolgende Herausgabe derselben in gänzlich ungeänderter Form sowohl durch das hohe Interesse des Gegenstandes an sich als durch die in ihr niedergelegte Behandlungsweise der wichtigsten Principien der Infinitesimal-Analysis wohl hinlänglich gerechtfertigt erscheinen.
Braunschweig, im Juli 1867. R. Dedekind.

(***) Bd. IV. p. 157.

ihr Werth grösser als C wird, und so lange negative, bis ihr Werth kleiner als C wird, so wird die Abweichung von C nie mehr betragen, als der Werth des dem letzten Zeichenwechsel voraufgehenden Gliedes. Da nun sowohl die Grössen a, als die Grössen b mit wachsendem Index zuletzt unendlich klein werden, so werden auch die Abweichungen von C, wenn man in der Reihe nur hinreichend weit fortgeht, beliebig klein werden, d. h. die Reihe wird gegen C convergiren.

Nur auf die Reihen erster Klasse sind die Gesetze endlicher Summen anwendbar; nur sie können wirklich als Inbegriff ihrer Glieder betrachtet werden, die Reihen der zweiten Klasse nicht; ein Umstand, welcher von den Mathematikern des vorigen Jahrhunderts übersehen wurde, hauptsächlich wohl aus dem Grunde, weil die Reihen, welche nach steigenden Potenzen einer veränderlichen Grösse fortschreiten, allgemein zu reden (d. h. einzelne Werthe dieser Grösse ausgenommen), zur ersten Klasse gehören.

[⋮]

Ueber den Begriff eines bestimmten Integrals und den Umfang seiner Gültigkeit.

4.

Die Unbestimmtheit, welche noch in einigen Fundamentalpunkten der Lehre von den bestimmten Integralen herrscht, nöthigt uns, Einiges voraufzuschicken über den Begriff eines bestimmten Integrals und den Umfang seiner Gültigkeit.

Also zuerst: Was hat man unter $\int_a^b f(x)\,dx$ zu verstehen?

Um dieses festzusetzen, nehmen wir zwischen a und b der Grösse nach auf einander folgend, eine Reihe von Werthen $x_1, x_2, \ldots, x_{n-1}$ an und bezeichnen der Kürze wegen $x_1 - a$ durch δ_1, $x_2 - x_1$ durch $\delta_2, \ldots$, $b - x_{n-1}$ durch δ_n und durch ε einen positiven ächten Bruch. Es wird alsdann der Werth der Summe

$$\begin{aligned} S &= \delta_1 f(a + \varepsilon_1\delta_1) + \delta_2 f(x_1 + \varepsilon_2\delta_2) + \delta_3 f(x_2 + \varepsilon_3\delta_3) + \cdots \\ &\quad + \delta_n f(x_{n-1} + \varepsilon_n\delta_n) \end{aligned}$$

von der Wahl der Intervalle δ und der Grössen ε abhängen. Hat sie nun die Eigenschaft, wie auch δ und ε gewählt werden mögen, sich einer festen Grenze A unendlich zu nähern, sobald sämmtliche δ unendlich klein werden, so heisst dieser Werth $\int_a^b f(x)\,dx$.

[⋮]

5.

Untersuchen wir jetzt zweitens den Umfang der Gültigkeit dieses Begriffs oder die Frage: in welchen Fällen lässt eine Function eine Integration zu und in welchen nicht?

Wir betrachten zunächst den Integralbegriff im engern Sinne, d. h. wir setzen voraus, dass die Summe S, wenn sämmtliche δ unendlich klein werden, convergirt. Bezeichnen wir also die grösste Schwankung der Function zwischen a und x_1, d. h. den Unterschied ihres grössten und kleinsten Werthes in diesem Intervalle, durch D_1, zwischen x_1 und x_2 durch $D_2 \ldots$, zwischen x_{n-1} und b durch D_n, so muss

$$\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$$

mit den Grössen δ unendlich klein werden. Wir nehmen ferner an, dass, so lange sämmtliche δ kleiner als d bleiben, der grösste Werth, den diese Summe erhalten kann, Δ sei; Δ wird alsdann eine Function von d sein, welche mit d immer abnimmt und mit dieser Grösse unendlich klein wird. Ist nun die Gesammtgrösse der Intervalle, in welchen die Schwankungen grösser als σ sind, = s, so wird der Beitrag dieser Intervalle zur Summe $\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$ offenbar $\geq \sigma s$. Man hat daher

$$\sigma s \leq \delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n \leq \Delta, \text{ folglich } s \leq \frac{\Delta}{\sigma}.$$

$\frac{\Delta}{\sigma}$ kann nun, wenn σ gegeben ist, immer durch geeignete Wahl von d beliebig klein gemacht werden; dasselbe gilt daher von s, und es ergiebt sich also:

Damit die Summe S, wenn sämmtliche δ unendlich klein werden, convergirt, ist ausser der Endlichkeit der Function f(x) noch erforderlich, dass die Gesammtgrösse der Intervalle, in welchen die Schwankungen > σ sind, was auch σ sei, durch geeignete Wahl von d beliebig klein gemacht werden kann.

Dieser Satz lässt sich auch umkehren:

Wenn die Function f(x) immer endlich ist, und bei unendlichem Abnehmen sämmtlicher Grössen δ die Gesammtgrösse s der Intervalle, in welchen die Schwankungen der Function f(x) grösser, als eine gegebene Grösse σ, sind, stets zuletzt unendlich klein wird, so convergirt die Summe S, wenn sämmtliche δ unendlich klein werden.

Denn diejenigen Intervalle, in welchen die Schwankungen > σ sind, liefern zur Summe $\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$ einen Beitrag, kleiner als s, multiplicirt in die grösste Schwankung der Function zwischen a und b, welche (n. V.) endlich ist; die übrigen Intervalle einen Beitrag < $\sigma(b - a)$. Offenbar kann man nun erst σ beliebig klein annehmen und dann immer noch die Grösse der Intervalle (n. V.) so bestimmen, dass auch s beliebig klein wird, wodurch der Summe $\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$ jede beliebige Kleinheit gegeben, und folglich der Werth der Summe S in beliebig enge Grenzen eingeschlossen werden kann.

Wir haben also Bedingungen gefunden, welche nothwendig und hinreichend sind, damit die Summe S bei unendlichem Abnehmen der Grössen δ convergire und also im engern Sinne von einem Integrale der Function f(x) zwischen a und b die Rede sein könne.

Wird nun der Integralbegriff wie oben erweitert, [...] an die Stelle der Bedingung, dass die Function immer endlich sei, aber tritt die Bedingung, dass die Function *nur* bei Annäherung des Arguments an *einzelne* Werthe unendlich werde, und dass sich ein bestimmter Grenzwerth ergebe, wenn die Grenzen der Integration diesen Werthen unendlich genähert werden.

Sur la possibilité de représenter une fonction par une série trigonométrique.

Bernhard Riemann

[Traduction publiée dans le *Bulletin des Sciences mathém. et astron.*, tome V ; juillet 1873] (*) (**)

[⋮]

§III.

En janvier 1829, parut dans le *Journal de Crelle* (***) un Mémoire de Dirichlet, où la possibilité de la représentation par les séries trigonométriques se trouvait établie en toute rigueur pour les fonctions qui sont, en général, susceptibles d'intégration, et qui ne présentent pas une infinité de maxima et de minima.

Il arriva à la découverte du chemin à suivre pour obtenir la solution de ce problème, par la considération que les séries infinies se partagent en deux classes, suivant qu'elles restent ou non convergentes, lorsqu'on rend leurs termes tous positifs. Dans les premières, les termes peuvent être intervertis d'une manière quelconque ; dans les autres, au contraire, la valeur dépend de l'ordre des termes. Si l'on désigne, en effet, dans une série de seconde classe, les termes positifs successifs par

$$a_1, a_2, a_3, \ldots,$$

et les termes négatifs par

$$-b_1, -b_2, -b_3, \ldots,$$

il est clair que $\sum a$, ainsi que $\sum b$ doit être infinie ; car, si ces deux sommes étaient finies l'une et l'autre, la série serait encore convergente

(*) Ce Mémoire a été présenté par l'auteur, en 1854, à la Faculté de Philosophie pour son habilitation à l'Université de Göttingue. Bien que l'auteur ne semble pas l'avoir destiné à la publicité, l'impression de ce travail sans aucun changement de forme paraîtra suffisamment justifiée tant par l'intérêt considérable qui s'attache au sujet, que par la manière dont y sont traités les principes les plus importants de l'Analyse infinitésimale.

Brunswick, juillet 1867. R. Dedekind.

(**) Petites modifications très mineures de la traduction. JFB

(***) Bd. IV. p. 157.

lorsqu'on donnerait à tous les termes le même signe ; si une seule était infinie, la sére serait divergente. Il est clair maintenant que la série, en plaçant les termes dans un ordre convenable, pourra prendre une valeur donnée C ; car, si l'on prend alternativement des termes positifs de la série jusqu'à ce que sa valeur soit plus grande que C, puis des termes négatifs jusqu'à ce que sa valeur soit moindre que C, la différence entre cette valeur et C ne surpassera jamais la valeur du terme qui précède le dernier changement de signe. Or les quantités a, aussi bien que les quantités b, finissant toujours par devenir infiniment petites pour des valeurs croissantes de l'indice, les écarts entre la somme de la série et C deviendront encore infiniment petits, lorsqu'on prolongera assez loin la série, c'est-à-dire que la série converge vers C.

C'est aux seules séries de la première classe que l'on peut appliquer les lois des sommes finies ; elles seules peuvent être considérées comme l'ensemble de leurs termes ; celles de la seconde classe ne le peuvent pas : circonstance qui avait échappé aux mathématiciens du siècle dernier, principalement par la raison que les séries qui procèdent suivant les puissances ascendantes d'une variable appartiennent, généralement parlant (c'est-à-dire à l'exception de certaines valeurs particulières de cette variable), à la première classe.

[⋮]

Sur la notion de l'intégrale définie, et l'étendue de son domaine de validité.

§IV.

L'incertitude qui règne encore sur quelques points fondamentaux de la théorie des intégrales définies nous oblige à placer ici quelques remarques sur la notion de l'intégrale définie, et sur la généralité dont elle est susceptible.

Et d'abord que doit-on entendre par $\int_a^b f(x)\,dx$?

Pour répondre à cette question, prenons entre a et b une série de valeurs $x_1, x_2, \ldots, x_{n-1}$, rangées par ordre de grandeur, depuis a jusqu'à b, et désignons pour abréger $x_1 - a$ par δ_1, $x_2 - x_1$ par $\delta_2, \ldots$, $b - x_{n-1}$ par δ_n ; soit, en outre, ε une fraction positive strictement comprise entre zéro et un. Il est clair que la valeur de la somme

$$\begin{aligned} S \;=\; & \delta_1 f(a+\varepsilon_1\delta_1) + \delta_2 f(x_1+\varepsilon_2\delta_2) + \delta_3 f(x_2+\varepsilon_3\delta_3) + \cdots \\ & + \delta_n f(x_{n-1}+\varepsilon_n\delta_n) \end{aligned}$$

dépendra du choix des intervalles δ et des fractions ε. Si elle a la propriété, de quelque manière que les δ et les ε puissent être choisis, de s'approcher indéfiniment d'une limite fixe A, quand les δ

tendent tous vers zéro, cette limite est la valeur de $\int_a^b f(x)\,dx$.

[⋮]

§V.

Recherchons maintenant l'étendue et la limite de la définition précédente, et posons-nous cette question : dans quels cas une fonction est-elle susceptible d'intégration ? dans quels cas ne l'est-elle pas ?

Considérons d'abord la définition de l'intégrale dans son sens le plus étroit, c'est-à-dire supposons que la fonction ne devienne pas infinie, et que la somme S converge, quand tous les δ tendent vers zéro. Désignons la plus grande oscillation de la fonction entre a et x_1, c'est-à-dire la différence entre sa plus grande et sa plus petite valeur dans cet intervalle par D_1 ;(*) de même, les plus grandes oscillations entre x_1 et x_2 par $D_2, \ldots$, entre x_{n-1} et b par D_n ; alors la somme

$$\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$$

doit devenir infiniment petite avec les quantités δ. Supposons que la plus grande valeur que cette somme puisse prendre,(**) quand tous les δ sont plus petits que d soit Δ ; Δ sera alors une fonction de d, diminuant et devenant infiniment petite avec d. Maintenant, si la somme totale des intervalles pour lesquels les oscillations sont plus grandes qu'une quantité σ est = s, la contribution de ces intervalles à la somme $\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$ sera évidemment $\geq \sigma s$. On aura donc

$$\sigma s \leq \delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n \leq \Delta, \text{ d'où } s \leq \frac{\Delta}{\sigma}.$$

$\frac{\Delta}{\sigma}$ peut d'ailleurs, si σ est donné, être rendu infiniment petit par un choix convenable de d ; il en sera donc de même de s, et l'on peut énoncer la proposition suivante :

Pour que la somme S *converge, quand tous les* δ *deviennent infiniment petits, il faut non seulement que la fonction demeure finie, mais encore que la somme totale des intervalles pour lesquels les oscillations sont >* σ*, quelque soit* σ*, puisse être rendue infiniment petite par un choix convenable de* d.

Cette proposition admet une réciproque :

Si la fonction f(x) *est toujours finie, et si, par le décroissement indéfini de toutes les quantités* δ*, la grandeur totale* s *des intervalles dans lesquels les oscillations de la fonction sont plus grandes qu'une quantité donnée* σ

(*) [jfb] On va comprendre cela comme la différence entre le sup et le inf, ceux-ci n'étant pas nécessairement atteints. La fonction est supposée bornée et intégrable au sens du paragraphe précédent.

(**) [id.] Ici encore on pensera à un sup.

peut toujours être rendue infiniment petite, la somme S *converge quand tous les* δ *tendent vers zéro.*

Car ces intervalles, dans lesquels les oscillations sont > σ, apportent à la somme $\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$ une contribution plus petite que s multiplié par la plus grande oscillation de la fonction entre a et b, oscillation qui est finie (p. hyp.) : les autres intervalles donnent dans la somme une partie $< \sigma(b - a)$;(*) on peut prendre évidemment σ aussi petit qu'on le veut, et alors (p. hyp.) on peut déterminer la grandeur des intervalles de telle manière que s soit aussi petit qu'on le veut.

On peut donc rendre $\delta_1 D_1 + \delta_2 D_2 + \cdots + \delta_n D_n$ aussi petit qu'on le veut, et, par suite, renfermer la somme S entre les limites aussi rapprochées qu'on le voudra.

Nous avons donc trouvé les conditions qui sont nécessaires et suffisantes pour que la somme S converge, quand les intervalles δ tendent vers zéro, et, par suite, pour qu'il puisse être question, dans le sens restreint, de l'intégrale de la fonction f(x) entre les limites a et b.

Si l'on étend, comme nous l'avons indiqué plus haut, la notion d'intégrale aux cas où la fonction devient infinie, [. . .], il faudra faire intervenir [la condition] suivante : que la fonction ne devienne infinie que lorsque son argument s'approche de certaines valeurs particulières, et que l'on obtienne une limite parfaitement déterminée, quand les limites des intégrations s'approchent indéfiniment de ces valeurs.

(*) [id.] lire $\leq \sigma(b - a)$.

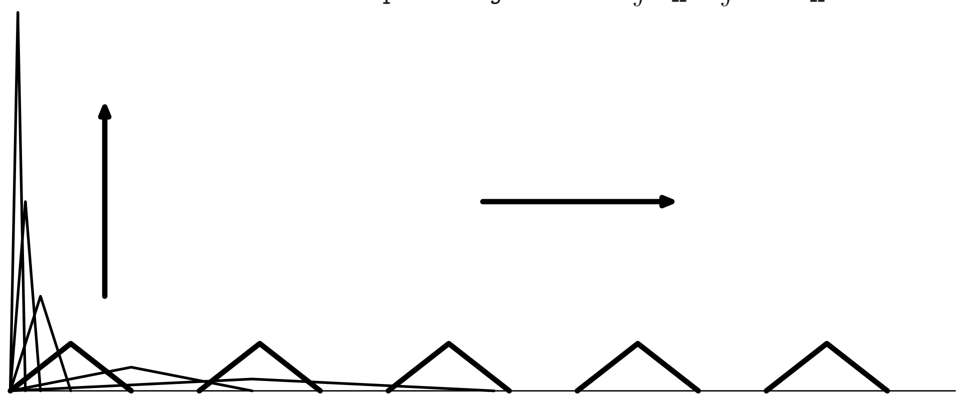
De Riemann à Lebesgue

Considérons des fonctions f_n sur l'intervalle $[0,1]$, intégrables au sens de Riemann. On suppose de plus qu'en tout x la limite $f(x) = \lim_{n\to\infty} f_n(x)$ existe. Comment pouvons-nous démontrer le Théorème qui affirme :

$$\lim_{n\to\infty}\int_0^1 f_n(x)dx = \int_0^1 \lim_{n\to\infty} f_n(x)dx \quad ?$$

Il y a plusieurs problèmes. D'abord ce << théorème >> est FAUX, comme le montre la figure suivante avec des fonctions de plus en plus effilées (en tout x on a $\lim f_n(x) = 0$ mais $\int_0^1 f_n(x)dx$ ne dépend pas de n) ou, sur $[0,\infty[$, de plus en plus aplaties. D'ailleurs il est alors encore plus facile de ne PAS avoir $\lim_{n\to\infty}\int_0^\infty f_n(x)dx = \int_0^\infty \lim_{n\to\infty} f_n(x)dx$ car il suffit de prendre $f_n(x) = f(x-n)$ avec f à support compact et d'intégrale non nulle, ce qui est aussi illustré sur la figure.

FIG. 1 - On n'a pas toujours $\lim\int f_n = \int \lim f_n$



Le deuxième problème c'est que la fonction limite n'est pas forcément intégrable au sens de Riemann : l'un des exemples les plus simples est de prendre une énumération $q_1, q_2, q_3\ldots$ des rationnels de $[0,1]$ et de prendre pour f_n la fonction indicatrice de l'ensemble fini $\{q_1,\ldots,q_n\}$. La fonction limite est la fonction indicatrice des rationnels de $[0,1]$ et elle n'est pas R-intégrable.(*)

(*) la double limite de fonctions continues $\lim_{N\to\infty}\lim_{M\to\infty}\sqrt[M]{|\sin(\pi N!x)|}$ est la fonction indicatrice des irrationnels. MAIS : il est impossible de trouver des fonctions continues $g_n(x)$ avec la limite simple $\lim g_n(x) = 1$ pour x irrationnel et 0 sinon. Je connais une preuve, mais elle utilise un théorème avancé de Topologie, le théorème de Baire.

Bon, alors ajoutons d'emblée comme hypothèse que f est R-intégrable. Et pour interdire le comportement exhibé par la figure 1, on va imposer la condition très naturelle :

$$\exists C \quad \forall n \forall x \quad |f_n(x)| \leq C$$

Je répète la question : est-il vrai et si c'est vrai comment prouve-t-on

$$\lim_{n \to \infty} \int_0^1 f_n(x)dx = \int_0^1 \lim_{n \to \infty} f_n(x)dx \quad ?$$

Cela vous paraît difficile ? Des pédagogues bien intentionnés vous ont dit que vous n'aviez qu'à admettre cette << convergence dominée >> ? Et si votre vie en dépendait ? On rigole moins maintenant ! Heureusement, je suis là.

D'abord je remplace C par 2C et les f_n par $f_n - f$ et je me ramène à montrer :

$$(\forall x \quad \lim f_n(x) = 0) \quad \Longrightarrow \lim \int_0^1 f_n(x)dx = 0 \,,$$

toujours sous l'hypothèse que les f_n sont uniformément bornées bien sûr. Cela serait chouette si la convergence était uniforme ! Et bien, on n'a qu'à faire en sorte qu'elle le soit ! Fixons un $\epsilon > 0$ et considérons les ensembles

$$X_N(\epsilon) = \{x | \ \forall n \geq N \quad |f_n(x)| \leq \epsilon\} \,.$$

Un très bref instant de réflexion montre que $X_1(\epsilon) \subset X_2(\epsilon) \subset X_3(\epsilon) \subset \dots$, et que de plus $\bigcup_N X_N(\epsilon) = [0,1]$. Je vais donc choisir N suffisamment grand de sorte que le complémentaire de $X_N(\epsilon)$ soit << petit >>, disons pour être précis de mesure au plus ϵ. J'écris alors :

$$n \geq N \Longrightarrow \int_0^1 |f_n(x)|dx = \int_{X_N(\epsilon)} |f_n(x)|dx + \int_{[0,1]\setminus X_N(\epsilon)} |f_n(x)|dx \leq \epsilon + M\epsilon \,,$$

où j'ai noté $M = \sup_n \sup_x |f_n(x)|$ qui est $< \infty$ par hypothèse. Le premier ϵ c'est parce que les $|f_n(x)|$ sont $\leq \epsilon$ et le deuxième epsilon c'est parce que le complémentaire de $X_N(\epsilon)$ a une mesure au plus ϵ. On a fini. On a prouvé :

$$\lim \int_0^1 |f_n(x)|dx = 0 \quad \text{et donc aussi} \quad \lim \int_0^1 f_n(x)dx = 0$$

C'est de la g n o g n o t t e cette convergence dominée !

Où est l'arnaque ? Il n'y en a pas... cette preuve est vraiment une preuve, totalement correcte, mais, et c'est un grand MAIS, nous avons utilisé une notion de << mesure >> pour des parties $X \subset [0,1]$ assez peu contrôlables et donc en apparence assez quelconques. Peut-on associer à t o u t sous-ensemble de $[0,1]$ une mesure de sorte qu'un certain nombre de propriétés naturelles soient vérifiées (l'additivité dénombrable(*) étant ce que j'entends par << propriété naturelle >>) ? C'est là le hic, et ça donne un peu mal à la tête : la réponse dépend des axiomes que l'on utilise pour faire des mathématiques ! Si

(*) il semble que c'est Émile Borel qui le premier a mis en avant cette notion fondamentale.

l'on admet la validité de l'axiome du choix, on peut construire des parties non-mesurables dans $[0,1]$. Mais les ensembles $X_N(\epsilon)$ que nous utilisons ne sont pas si arbitraires que cela : ils appartiennent (grâce au théorème de Riemann qui nous dit que l'ensemble des points de discontinuité d'une fonction R-intégrable a une mesure (extérieure) nulle) à la tribu engendrée par les intervalles et par les ensembles négligeables (une notion que l'on peut définir totalement intrinsèquement en disant que A est négligeable si pour tout ϵ on peut recouvrir A par une collection dénombrable d'intervalles ouverts dont la somme des longueurs est $\leq \epsilon$). Donc le plan est clair : définir la notion de tribu, définir la tribu Borélienne, la tribu de Lebesgue qui est celle qui en plus des Boréliens contient tous les négligeables, et, grand finale, prouver l'existence d'une mesure μ, c'est-à-dire d'une fonction d'ensembles à valeurs positives vérifiant deux ou trois axiomes évidents, dont le plus signifiant est l'additivité dénombrable, avec $\mu(]a,b[) = b - a$, $\mu([0,1]) = 1$, l'appeler << mesure de Lebesgue >>, définir la notion de $\int_A f(x)d\mu(x)$ lorsque A est Lebesgue-mesurable et f R-intégrable, ce qui quasiment oblige en fait à définir $\int_{[0,1]} f(x)d\mu(x)$ pour les fonctions f mesurables quelconques, prouver des trucs utiles comme $|\int f d\mu| \leq \int |f| d\mu$, et puis prouver aussi que l'intégrale de Lebesgue donne le même résultat que celle de Riemann, et voilà, après ces quelques petites broutilles la preuve que j'ai donnée est une vraie preuve.

Le théorème de la convergence dominée

Venons-en maintenant à l'énoncé complet. On oublie les fonctions R-intégrables, on travaille avec les fonctions Lebesgues mesurables (c'est-à-dire les fonctions f tel que $f^{-1}(]u,v])$ est dans la tribu de Lebesgue pour tout u, v). On a g à valeurs positives ou nulles, intégrable : $\int_0^1 g(x)dx < \infty$. On suppose(*) $\forall n \forall x\ |f_n(x)| \leq g(x)$ et $\forall x\ \lim f_n(x) = f(x)$. Tout d'abord on démontre que f est Lebesgue-mesurable, je l'admets ici. Elle est intégrable puisque majorée par g. Quitte à remplacer g par 2g et f_n par $f_n - f$, il suffit d'établir :

$$(\forall x \quad \lim f_n(x) = 0) \quad \Longrightarrow \lim \int_0^1 |f_n(x)| dx = 0\ ,$$

toujours sous l'hypothèse bien sûr qu'elles sont bornées par une fonction intégrable fixe g (le double de notre ancien g). Fixons un $\epsilon > 0$ et considérons les ensembles

$$X_N(\epsilon) = \{x|\ \forall n \geq N \quad |f_n(x)| \leq \epsilon\}\ .$$

Un très bref instant de réflexion montre que $X_1(\epsilon) \subset X_2(\epsilon) \subset X_3(\epsilon) \subset \ldots$, et que de plus $\bigcup_N X_N(\epsilon) = [0,1]$. Je vais donc choisir N suffisamment grand de sorte que le complémentaire $Y_N(\epsilon)$ de $X_N(\epsilon)$ soit << petit >>, et pour être précis disons :

$$\int_{Y_N(\epsilon)} g(x)dx \leq \epsilon\ .$$

(*) on pourrait ne supposer cela que pour presque tout x et définir f(x) = 0 pour les autres.

J'écris alors :

$$n \geq N \Longrightarrow \int_0^1 |f_n(x)|dx = \int_{X_N(\epsilon)} |f_n(x)|dx + \int_{Y_N(\epsilon)} |f_n(x)|dx \leq \epsilon + \int_{Y_N(\epsilon)} g(x)dx \leq 2\epsilon\,,$$

On a fini. On a prouvé :

$$\forall \epsilon > 0 \exists N \ \ n \geq N \Longrightarrow \int_0^1 |f_n(x)|dx \leq \epsilon$$

C'est de la g n o g n o t t e cette convergence dominée !

Mais comment savons-nous que nous pouvons choisir N tel que $\int_{Y_N(\epsilon)} g(x)dx \leq \epsilon$ soit vrai ? pour cela nous avons besoin de savoir que si g est une fonction intégrable positive alors ses intégrales sur des ensembles de petite mesure sont petites, uniformément. Autrement dit nous voulons(*) :

$$\boxed{\begin{array}{c} \text{si} \quad \displaystyle\int_0^1 |f(x)|dx < \infty \quad \text{alors} \\[2ex] \forall \epsilon > 0 \ \ \exists \delta > 0 \quad |A| \leq \delta \Longrightarrow \displaystyle\int_A |f(x)|dx \leq \epsilon \end{array}}$$

Si vous connaissez le théorème de la convergence dominée mais que vous ne connaissez pas ce théorème c'est dommage. Ce théorème est très important à connaître(**), et nous venons de voir combien il est facile d'en déduire le théorème de la convergence dominée. Alors, prouvons-le. On raisonne par l'absurde. Si il était faux pour f (que l'on peut dorénavant remplacer par $|f|$, donc $f \geq 0$) alors il existerait $\epsilon > 0$ tel que pour tout $\delta > 0$, disons $\delta_n = \frac{1}{2^n}$, on puisse trouver A_n avec $|A_n| \leq \delta_n$ mais $\int_{A_n} f(x)dx \geq \epsilon$. Considérons $B_n = \bigcup_{m \geq n} A_m$, de sorte que $B_1 \supset B_2 \supset B_3 \supset \dots$ et soit $B = \bigcap_{n \geq 1} B_n$.

Remarquons que $|B_n| \leq \sum_{m \geq n} \frac{1}{2^m} = \frac{2}{2^n}$, donc $|B| \leq \frac{2}{2^n}$ pour tout $n \geq 1$ et donc B est de mesure nulle.

Soit $C_n = B_n \setminus B$. On a donc $\int_{C_n} f(x)dx = \int_{B_n} f(x)dx \geq \int_{A_n} f(x)dx \geq \epsilon$. Soit $D_n = C_n \setminus C_{n+1}$. Comme $\bigcap_{n \geq 1} C_n = \emptyset$, on a une partition dénombrable de C_n en $\bigcup_{m \geq n} D_m$ et donc $\int_{C_n} f(x)dx = \sum_{m \geq n} \int_{D_m} f(x)dx = \sum_{m \geq n} u_m$ avec $u_m = \int_{D_m} f(x)dx$. Mais pour toute série convergente $\sum_{m=1}^{\infty} u_m$ on a $\lim_{N \to \infty} \sum_{k \geq N} u_k = 0$. Cela donne une contradiction puisque $\int_{C_n} f(x)dx \geq \epsilon$ pour tout n.

Nous avons eu besoin d'utiliser : si des parties mesurables D_m d i s j o i n t e s sont données alors pour toute fonction intégrable positive f on a $\int_D f(x)dx =$

(*) comme il est usuel on a noté $|A|$ la mesure de Lebesgue de A.

(**) malheureusement l'incompétence notoire de votre professeur se traduit par son incapacité à vous dire le nom, si il en existe un, de ce théorème.

$\sum_m \int_{D_m} f(x)dx$ pour $D = \bigcup_m D_m$. On traite d'abord le cas d'un ensemble d'indices m fini, ce qui se ramène au cas de deux ensembles D_1 et D_2. Mais alors $\int_D f(x)dx = \int_{[0,1]} f(x)(\mathbb{1}_{D_1}(x) + \mathbb{1}_{D_2}(x))dx$ et on s'est ramené à la linéarité de l'intégrale que je ne redémontre pas ici. Revenons au cas dénombrable. En tout cas $D \supset D_1 \cup \dots \cup D_N$ donc $\int_D f(x)dx \geq \sum_{1\leq m\leq N} \int_{D_m} f(x)dx$ donc $\int_D f(x)dx \geq \sum_{m=1}^{\infty} \int_{D_m} f(x)dx$. Cela donne une première inégalité.

Rappelons que par définition $\int_D f(x)dx = \sup \int \phi(x)\mathbb{1}_D(x)dx$, le supremum étant pris sur les fonctions étagées avec $0 \leq \phi \leq f$. Supposons le théorème vrai pour les ϕ. Alors $\int_D \phi(x)dx = \sum_m \int_{D_m} \phi(x)dx \leq \sum_m \int_{D_m} f(x)dx$ pour toute $\phi \leq f$ par monotonie de l'intégrale. Cela donne l'inégalité dans l'autre sens.

Donc il suffit de prouver la formule pour $f = \phi$ une fonction étagée ce qui se ramène à $f = \mathbb{1}_E$ avec E une partie mesurable. La formule dit alors $|E \cap D| = \sum_m |E \cap D_m|$ ce qui n'est pas autre chose que la sigma-additivité pour la partition de $E \cap D$ en les $E \cap D_m$. Voilà on a tout démontré.

le théorème de la convergence dominée vaut aussi en supposant convergence presque partout au lieu de partout je vous en laisse la vérification.

Le théorème de la convergence monotone

On suppose $0 \leq f_1 \leq f_2 \leq \dots$ et on pose $f(x) = \lim f_n(x) \in [0, +\infty]$. La théorie de l'intégrale définit une valeur pour $\int f(x)dx$ pour toute telle f (les x avec $f(x) = \infty$ forment une partie mesurable). Le théorème de la convergence monotone affirme

$$\lim \int_0^1 f_n(x)dx = \int_0^1 f(x)dx$$

Prouvons-le.

Certainement on a $\leq$ puisque $f_n \leq f$. Si $\int_0^1 f(x)dx < \infty$ on est dans le cadre du théorème de la convergence dominée ! donc la formule ne peut-être fausse que si $\int_0^1 f(x)dx = \infty$. Il s'agit ainsi de prouver que forcément $\lim \int_0^1 f_n(x)dx = +\infty$ si $\int_0^1 f(x)dx = \infty$, ou de manière équivalente que si les $\int_0^1 f_n(x)dx$ sont bornés alors $\int_0^1 f(x)dx < \infty$. Soit ϕ étagée avec $0 \leq \phi \leq f$ et posons $g_n = \min(f_n, \phi)$. On a $\lim g_n(x) = \phi(x)$ pour tout x(*). De plus $g_n \leq \phi$, donc par convergence dominée on a $\int_0^1 \phi(x)dx = \lim \int_0^1 g_n(x)dx \leq \lim \int_0^1 f_n(x)dx$. Donc si ces intégrales sont majorées par $M < \infty$, alors il en est de même pour $\int_0^1 \phi(x)dx$. Cela prouve par définition que f est intégrable (et aussi que $\int_0^1 f(x)dx \leq M$, donc $\int_0^1 f(x)dx \leq \sup_n \int_0^1 f_n(x)dx = \lim \int_0^1 f_n(x)dx$ et comme l'inégalité dans l'autre sens est évidente on a égalité ; mais on savait déjà car par $\int_0^1 f(x)dx < \infty$ on a convergence dominée). Le théorème de la convergence monotone est établi.

Si l'on réexamine notre preuve on en déduit le corollaire extrêmement utile

(*) vous êtes des grands maintenant, alors j'espère que vous suivez.

suivant :

> si les $f_n(x)$ forment en tout (ou presque tout) x une suite croissante positive et si de plus les intégrales $\int_0^1 f_n(x)dx$ sont bornées alors en presque tout x les $f_n(x)$ convergent vers une limite finie. Ou encore : si les $u_n(x) \geq 0$ sont telles que
>
> $$\sum_{n=0}^{\infty} \int_0^1 u_n(x)dx < \infty$$
>
> alors la somme infinie $\sum_{n=0}^{\infty} u_n(x)$ est presque partout finie (et bien sûr
>
> $$\sum_{n=0}^{\infty} \int_0^1 u_n(x)dx = \int_0^1 \sum_{n=0}^{\infty} u_n(x)dx\ ,$$
>
> mais cela on le sait même en cas d'une somme divergente.)

En effet, pour tout $N \in \mathbb{N}$ on a $\int_0^1 f(x)dx \geq N|\{x|f(x) = +\infty\}|$, donc si l'intégrale est finie, alors $\{x|f(x) = +\infty\}$ est de mesure nulle.

Espaces mesurés, Lemme de Fatou, etc...

En fait c'est uniquement pour des raisons typographiques que je me suis limité à $[0,1]$ et à la mesure de Lebesgue dx, nos preuves de la convergence dominée et de la convergence monotone s'appliquent mutatis mutandis au cas général d'un espace mesuré (X,μ) quelconque. Je vous laisse vous amuser à expliciter ce que cela donne pour $X = \mathbb{N}$ ou $\mathbb{Z}$, avec μ la mesure de dénombrement (et la tribu maximale égale à $\mathcal{P}(\mathbb{N})$ ou $\mathcal{P}(\mathbb{Z})$.)

Pour Fatou je ne peux pas faire tout le boulot, et je n'ai pas envie d'entamer une nouvelle page. Je vous laisse le soin de vous assurer que vous connaissez l'énoncé et une preuve du fameux Lemme de Fatou. Si si, c'est très utile.

Le théorème de Riesz-Fischer

Théorème (1907, indépendamment par Ernst Fischer et Frigyes Riesz) : *Soit* c_n, $n \in \mathbb{Z}$ *des nombres complexes vérifiant la condition*

$$\sum_{-\infty}^{+\infty} |c_n|^2 < \infty$$

Alors il existe une fonction de carré intégrable $f \in L^2(0, 2\pi; \frac{dx}{2\pi})$ *dont les coefficients de Fourier* $c_n(f)$ *sont les* c_n : $\forall n \in \mathbb{Z} \quad c_n(f) = c_n$.

Preuve : il suffit de savoir le faire lorsque l'on suppose les c_n nuls pour $n < 0$, car dans le cas général on obtiendra f par $f_1 + \overline{f_2}$ avec $c_n(f_1) = c_n$ pour $n \geq 0$, $= 0$ pour $n < 0$ et $c_n(f_2) = \overline{c_{-n}}$ pour $n > 0$, $= 0$ pour $n \leq 0$. Posons

$$F_M(x) = \sum_{n=0}^{n=M} c_n e^{inx}$$

et remarquons par Cauchy-Schwarz, et par les << relations d'orthogonalité >>, pour $N \geq M$: $\int_0^{2\pi} |F_N(x) - F_M(x)| \frac{dx}{2\pi} \leq \left(\int_0^{2\pi} |F_N(x) - F_M(x)|^2 \frac{dx}{2\pi}\right)^{\frac{1}{2}} = \left(\sum_{M<n\leq N} |c_n|^2\right)^{\frac{1}{2}}$.

Comme $\sum_{n=0}^{\infty} |c_n|^2 < \infty$ on peut trouver N_1 avec $\sum_{n=N_1+1}^{\infty} |c_n|^2 \leq \frac{1}{4}$. Puis on a $N_2 > N_1$ avec $\sum_{n=N_2+1}^{\infty} |c_n|^2 \leq \frac{1}{16}$. Plus généralement on construit une suite strictement croissante (N_k), $1 \leq k$ avec pour chaque k, $\sum_{n=N_k+1}^{\infty} |c_n|^2 \leq \frac{1}{4^k}$.

On pose alors $g_0(x) = F_{N_1}(x)$, $g_1(x) = F_{N_2}(x) - F_{N_1}(x)$, etc..., $g_k(x) = F_{N_{k+1}}(x) - F_{N_k}(x)$. On a donc

$$F_{N_{K+1}}(x) = g_0(x) + g_1(x) + \cdots + g_K(x)$$

Définissons aussi

$$G_K(x) = |g_0(x)| + |g_1(x)| + \cdots + |g_K(x)|$$

et même

$$G(x) = \sum_{k=0}^{\infty} |g_k(x)| = \lim_{K \to \infty} G_K(x)$$

Bien sûr la valeur $+\infty$ n'est pas exclue a priori pour G(x), mais on va montrer que cela ne peut se produire que sur un ensemble de mesure nulle. Il suffit pour cela de montrer $\int_0^{2\pi} G(x) \frac{dx}{2\pi} < \infty$. Par le théorème de la convergence monotone on a

$$\int_0^{2\pi} G(x) \frac{dx}{2\pi} = \lim_{K \to \infty} \int_0^{2\pi} G_K(x) \frac{dx}{2\pi}$$

Par construction et par Cauchy-Schwarz on a pour $k \geq 1$, $\int_0^{2\pi} |g_k(x)| \frac{dx}{2\pi} \leq \left(\frac{1}{4^k}\right)^{\frac{1}{2}} = \frac{1}{2^k}$, donc $\int_0^{2\pi} G_K(x) \frac{dx}{2\pi} \leq \int_0^{2\pi} |g_0(x)| \frac{dx}{2\pi} + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^K} \leq \int_0^{2\pi} |g_0(x)| \frac{dx}{2\pi} + 1$.

Donc effectivement $\int_0^{2\pi} G(x) \frac{dx}{2\pi} < \infty$, et du coup le lieu des x avec $G(x) = +\infty$ est de mesure nulle. Pour ceux là posons $F(x) = 0$. Pour les autres la série $\sum_{k=0}^{\infty} g_k(x)$ est absolument convergente et on pose $F(x) = \sum_{k=0}^{\infty} g_k(x)$. On a donc pour tous les x, $|F(x)| \leq G(x)$. On va montrer $\int_0^{2\pi} G(x)^2 \frac{dx}{2\pi} < \infty$ et donc on saura alors que F est de carré intégrable. À nouveau par le théorème de la convergence monotone on a

$$\int_0^{2\pi} G(x)^2 \frac{dx}{2\pi} = \lim_{K \to \infty} \int_0^{2\pi} G_K(x)^2 \frac{dx}{2\pi}$$

Par la propriété de norme de la norme L^2, on a

$$\begin{aligned}
\left(\int_0^{2\pi} \Big(\sum_{0 \leq k \leq K} |g_k(x)|\Big)^2 \frac{dx}{2\pi}\right)^{\frac{1}{2}} &\leq \sum_{0 \leq k \leq K} \left(\int_0^{2\pi} |g_k(x)|^2 \frac{dx}{2\pi}\right)^{\frac{1}{2}} \\
&\leq \left(\int_0^{2\pi} |g_0(x)|^2 \frac{dx}{2\pi}\right)^{\frac{1}{2}} + \frac{1}{2} + \frac{1}{4} + \cdots + \frac{1}{2^K} \\
&\leq \left(\int_0^{2\pi} |g_0(x)|^2 \frac{dx}{2\pi}\right)^{\frac{1}{2}} + 1
\end{aligned}$$

Donc la suite croissante $\int_0^{2\pi} G_K(x)^2 \frac{dx}{2\pi}$ est majorée, elle a donc une limite finie, et ainsi on a bien prouvé que G, donc aussi F est de carré intégrable.

Finalement comme $F(x)e^{-inx} = \lim_{K \to \infty} \sum_{0 \leq k \leq K} g_k(x) e^{-inx}$ pour presque tout x et que pour tout x on a $\left|\sum_{0 \leq k \leq K} g_k(x) e^{-inx}\right| \leq G(x)$ et que G est intégrable on peut appliquer le théorème de la convergence dominée et en déduire $c_n(F) = \lim_{K \to \infty} c_n(F_{N_{K+1}})$, $F_{N_{K+1}} = \sum_{0 \leq k \leq K} g_k$. Si $n < 0$ tous les $c_n(F_{N_{K+1}})$ sont nuls, et si $n \geq 0$, tous les $c_n(F_{N_{K+1}})$ sont égaux, pour K suffisamment grand, au c_n dont on est parti au départ. Le théorème de Riesz-Fischer est démontré.

Théorème : *soit f une fonction intégrable mais qui n'est pas de carré intégrable. Alors* $\sum_{n \in \mathbb{Z}} |c_n(f)|^2 = +\infty$.

Preuve : sinon, soit F la fonction de carré intégrable qui est donnée par le théorème de Riesz-Fischer. Elle vérifie donc $c_n(f) = c_n(F)$ pour tous les $n \in \mathbb{Z}$. Comme F est de carré intégrable elle est intégrable, et on invoque alors le théorème d'unicité dans L^1 (démontré dans un cours précédent). On en déduit que $f = F$ dans L^1 donc presque partout, donc que f est de carré intégrable. Contradiction. L'égalité de Bessel-Parseval est donc toujours valable.

Université Lille 1 - UFR de Mathématiques
Licence Semestre 6 - 2005/2006 - MATH 312

Résumé du Cours (I)

Je récapitule le contenu des six ou sept premières semaines de Cours, sans trop entrer dans le détail, peut-être en laissant de côté certaines choses. C'est juste pour aider à avoir les idées claires à ce stade du semestre.

Au départ il y a les formules de Fourier pour des coefficients $a_n(f)$, et $b_n(f)$, ou encore $c_n(f)$ que l'on peut associer à une fonction 2π-périodique et intégrable. Inscrivez-ici ces formules :

On peut motiver le fait que les formules de Fourier doivent être ce qu'elles sont en considérant un polynôme trigonométrique $P(x) = \frac{1}{2}a_0 + \sum_{1\leq n\leq N}(a_n\cos(nx) + b_n\sin(nx))$ et en constatant qu'elles redonnent bien les coefficients a_n et b_n (et des coefficients nuls pour $n > N$; aussi on a la convention $a_{-n} = a_n$ et $b_{-n} = -b_n$). Derrière ces calculs se trouvent des << relations d'orthogonalité >> pour les fonctions 2π-périodiques $\cos(nx)$, $\sin(nx)$, $e_n(x) = e^{inx}$, relations qui sont aussi à l'oeuvre dans l'importante identité

$$\frac{1}{2\pi}\int_0^{2\pi}|P(x)|^2\,dx = \sum_{-N\leq n\leq +N}|c_n|^2 = \frac{|a_0|^2}{4} + \frac{1}{2}\sum_{1\leq n\leq N}(|a_n|^2 + |b_n|^2)$$

Preuve (écrire petit) :

Dès maintenant on se familiarise avec le << produit scalaire hermitien >> $(f,g) = \frac{1}{2\pi}\int_0^{2\pi} f(x)\overline{g(x)}\,dx$ et ses propriétés élémentaires. L'inégalité de

Cauchy-Schwarz, dont on peut donner plusieurs preuves, est :

$$|(f,g)| \leq \sqrt{(f,f)}\sqrt{(g,g)}$$

Preuve :

Elle est fondamentale. Les notations $\|f\|_1 = \frac{1}{2\pi}\int_0^{2\pi}|f(x)|\,dx$ et $\|f\|_2 = \sqrt{(f,f)}$ sont constamment utilisées. On a : $\|\alpha f\|_1 = |\alpha|\|f\|_1$, $\|\alpha f\|_2 = |\alpha|\|f\|_2$.

$$\|f+g\|_1 \leq \|f\|_1 + \|g\|_1 \qquad \|f+g\|_2 \leq \|f\|_2 + \|g\|_2 \qquad \|f\|_1 \leq \|f\|_2$$

Preuves (écrire petit) :

À propos de Cauchy-Schwarz (valable pour tout espace mesuré), on note que l'on peut rajouter une étape intermédiaire $|(f,g)| \leq (|f|,|g|) \leq \|f\|_2\|g\|_2$, puisque $\|f\|_2 = \|\,|f|\,\|_2$, $\|g\|_2 = \|\,|g|\,\|_2$. Attention au fait que l'inégalité $\|f\|_1 \leq \|f\|_2$, c'est parce que notre espace mesuré est de masse totale 1. Sur $\mathbb{R}$ tout entier pour la mesure dx, ça ne marche pas. Mais sur $[0,2\pi]$ pour $\frac{dx}{2\pi}$ toute fonction de carré intégrable est intégrable (la réciproque étant fausse).

À propos, j'utilise dès maintenant le vocabulaire de l'intégration selon Lebesgue, puisque, bien que l'on puisse faire beaucoup de choses avec l'intégrale de Riemann, les résultats centraux du cours (théorème de Riesz-Fischer, complétude de l'espace L^2) nécessitent les outils de l'intégrale de Lebesgue.

Par ailleurs, pour f une fonction intégrable 2π-periodique, de coefficients de Fourier $c_n(f)$, $n \in \mathbb{Z}$, on écrit parfois :

$$f(x) \sim \sum_{-\infty}^{\infty} c_n(f)\, e^{inx}\,,$$

le symbole $\sim$ ne signifiant ici pas une égalité, ou une équivalence, mais simplement rappelant que la série trigonométrique à droite est définie à partir de la fonction à gauche. Je mentionne tout de suite que l'un de nos théorèmes principaux est réciproquement que la fonction à gauche est déterminée, à l'égalité presque partout près, par sa série de Fourier (théorème d'unicité dans L^1). En ce qui concerne la validité ponctuelle :

$$f(x) \stackrel{?}{=} \lim_{N \to \infty} \sum_{-N \leq n \leq +N} c_n(f)\, e^{inx} ,$$

la toute première question est déjà de se demander si cela change quelque chose si l'on prend, par exemple $\sum_{-N\leq n\leq +N+1}$ au lieu de $\sum_{-N\leq n\leq +N}$. Non cela ne change rien, grâce au :

<u>Lemme de Riemann-Lebesgue</u> : pour toute fonction intégrable on a $c_n(f) \to 0$ pour $n \to +\infty$ et pour $n \to -\infty$. Plus généralement, pour tout intervalle $I \subset \mathbb{R}$ et toute fonction intégrable f par rapport à dx sur I on a $\lim_{\lambda \to +\infty} \int_I f(x)\cos(\lambda x)\,dx = 0$ et $\lim_{\lambda \to +\infty} \int_I f(x)\sin(\lambda x)\,dx = 0$.

Ce théorème fondamental est très fondamental. Il est intimement lié à la << densité L^1 >> des fonctions en escalier. Il suffit de considérer f à valeurs réelles. Dans le cas d'un intervalle borné, et d'une fonction f Riemann-intégrable on a, presque par définition de l'intégrale de Riemann, pour tout $\epsilon > 0$ donné à l'avance, l'existence de U et V en escalier avec $U \leq f \leq V$ et $\int_I (V(x) - U(x))\,dx \leq \epsilon$. Dans le cas général on a en tout cas, c'est l'un des théorèmes subtils démontrés en cours, l'existence de U en escalier avec $\int_I |f(x) - U(x)|\,dx \leq \epsilon$. J'y reviens plus loin. Pour toute fonction en escalier U, un calcul explicite montre $\int_I U(x) \frac{\sin}{\cos}(\lambda x)\,dx = \mathcal{O}(\lambda^{-1})$. Par une méthode expliquée en cours, on peut alors prouver le Lemme de Riemann-Lebesgue pour toute f intégrable sur l'intervalle I. À la place des fonctions en escalier, on peut utiliser les fonctions de classe C^1 à support compact, mais montrer leur << densité >> n'est pas plus simple.

Dans le cas des coefficients de Fourier $c_n(f)$ on peut aussi obtenir $|c_n(f)| \to 0$ comme conséquence de l'inégalité de Bessel, que nous voyons plus loin, qui s'applique lorsque f est de carré intégrable, en particulier lorsque f est Riemann-intégrable.

Je reviens au problème de la convergence ponctuelle des sommes partielles $S_N(f)(x) = \sum_{-N\leq n\leq +N} c_n(f)\, e^{inx}$ vers f(x). La notion de <u>convolution</u> est extrêmement utile dans cette discussion. En effet on a la formule :

$$S_N(f)(x) = (D_N * f)(x)$$

avec D_N le <u>noyau de Dirichlet</u>. Fejér a vu qu'il était extrêmement utile de s'intéresser aux moyennes de Cesàro $\Sigma_N = \frac{1}{N}(S_0 + S_1 + \ldots S_{N-1})$, qui sont donc aussi des convolutions :

$$\Sigma_N(f)(x) = (F_N * f)(x)$$

Inscrire ici les formules pour D_N, pour F_N, et pour la notion de convolution $f * g$:

En utilisant en particulier que D_N et F_N sont paires, on a aussi, pour tout $L \in \mathbb{C}$:

$$S_N(f)(x) - L = \frac{1}{\pi}\int_0^\pi D_N(t)\left(\frac{f(x+t)+f(x-t)}{2} - L\right) dt$$

$$\Sigma_N(f)(x) - L = \frac{1}{\pi}\int_0^\pi F_N(t)\left(\frac{f(x+t)+f(x-t)}{2} - L\right) dt$$

On a alors deux théorèmes, le premier que l'on a envie d'appeler théorème de Dirichlet, bien que le vrai théorème de Dirichlet soit autre chose, et le second qui est un célèbre théorème de Fejér :

Théorème : si g est une fonction intégrable sur $]0,\pi]$ avec existence de la limite L = g(0+) à droite en 0 et existence d'une dérivée à droite en 0 : $\lim_{h \to 0^+}(g(h) - L)/h$, ou une autre condition plus faible, comme la condition de Dini $\int_0^\pi \frac{|g(h)-L|}{h}\, dh < \infty$,(*) alors

$$g(0^+) = L = \lim_{N \to \infty} \frac{1}{\pi}\int_0^\pi D_N(t) g(t)\, dt$$

Conséquence, ce que l'on appelle théorème de Dirichlet dans les livres, mais encore une fois, ce n'est pas le théorème démontré par Dirichlet : si f est de classe C^1 par morceaux alors en tout x on a la convergence ponctuelle :

$$\lim_{N \to \infty} S_N(f)(x) = \frac{f(x^+) + f(x^-)}{2}$$

Théorème (Fejér) : si g est une fonction intégrable sur $]0,\pi]$ avec existence de la limite $L = g(0+)$ à droite en 0 alors

$$g(0^+) = L = \lim_{N \to \infty} \frac{1}{\pi}\int_0^\pi F_N(t) g(t)\, dt$$

(*) il ne peut y avoir au plus qu'un L qui vérifie la condition de Dini, et un tel L peut exister sans que g(0+) n'existe. Les intégrales ont pour limite L même si g(0+) n'existe pas.

En ce qui concerne la convergence ponctuelle, si f est simplement supposée intégrable, si $l = \lim S_N(f)(x)$ existe, alors par le théorème de Cesàro on a aussi $l = \lim \Sigma_N(f)(x)$ et donc par le théorème de Fejér, si on suppose de plus que x est un point de continuité de f alors on a nécessairement $l = f(x)$.

Par ailleurs on a aussi la version uniforme du théorème de Fejér : si f est continue 2π-périodique alors les polynômes trigonométriques $\Sigma_N(f)$ convergent uniformément vers f pour $N \to \infty$.

On a aussi la version uniforme de la convergence des $S_N(f)$: si f est de classe C^1 par morceaux ET si f est de plus continue, alors sa série de Fourier est normalement convergente vers f : $\sum_{n\in\mathbb{Z}} |c_n(f)| < \infty$, $\forall x \; f(x) = \sum_{n\in\mathbb{Z}} c_n(f)\, e^{inx}$. Cette convergence normale se prouve en utilisant le fait que la fonction continue par morceaux f' (avec un nombre fini de points de discontinuité) est de carré intégrable donc vérifie l'inégalité de Bessel, un sujet vers lequel nous nous tournons maintenant.

Pour toute fonction f de carré intégrable on a l'Inégalité de Bessel

$$\sum_{n\in\mathbb{Z}} |c_n(f)|^2 \leq \frac{1}{2\pi}\int_0^{2\pi} |f(x)|^2\, dx$$

Preuve : montrez

$$\sum_{-N\leq n\leq +N} |c_n(f)|^2 + \frac{1}{2\pi}\int_0^{2\pi} |f(x) - S_N(f)(x)|^2\, dx = \frac{1}{2\pi}\int_0^{2\pi} |f(x)|^2\, dx$$

Dans le contexte de la preuve de l'inégalité de Bessel, on établit une propriété fondamentale des sommes partielles $S_N(f)$ de la série de Fourier de f : parmi tous les polynômes trigonométriques de degrés au plus N, l'unique polynôme P qui minimise $\|f - P\|_2$ est $P = S_N(f)$.

L'inégalité de Bessel est en fait une égalité, dite de Bessel-Parseval, ou de Parseval tout court. Pour la preuve il faut donc montrer $\|f - S_N(f)\|_2 \to 0$. Comme dans la preuve du Lemme de Riemann-Lebesgue, le point crucial est la densité (ici au sens de la norme L^2) de fonctions pour lesquelles on sait montrer cela directement. Par exemple pour les fonctions continues, on a $\|g - \Sigma_N(g)\|_2 \dashrightarrow 0$ car $\Sigma_N(g)$ converge uniformément vers la fonction continue g sur $[0, 2\pi]$ et par ailleurs par la propriété de meilleure approximation on a $\|g - S_N(g)\|_2 \leq \|g - \Sigma_N(g)\|_2$. Donc pour prouver que $\|f - S_N(f)\|_2 \dashrightarrow 0$ pour toutes les fonctions de carrés intégrables, il suffit de montrer que pour toute fonction f de carré intégrable et pour tout $\epsilon > 0$ on peut trouver une fonction continue g (2π-périodique) avec $\|f - g\|_2 \leq \epsilon$. Et pour cela il suffit de trouver k en escalier avec $\|f - k\|_2 \leq \frac{1}{2}\epsilon$ car il est facile de trouver g continue avec $\|k - g\|_2 \leq \frac{1}{2}\epsilon$.

Nous avons prouvé que les fonctions en escalier étaient denses au sens L^2 dans l'espace $L^2(0, 2\pi; \frac{dx}{2\pi})$ en montrant que les parties mesurables A pour lesquelles les fonctions indicatrices $\mathbb{1}_A$ sont arbitrairement bien approchables par des fonctions en escalier au sens L^2 forment une tribu. Notre preuve repose sur l'emploi des célèbres << identités Burnoliennes >>.

Remarque : dans certains livres on construit l'espace L^2 en complétant (comme on complète $\mathbb{Q}$ pour construire $\mathbb{R}$) l'espace vectoriel des fonctions continues pour la norme $\|\cdot\|_2$. Dans cette approche la densité des fonctions continues ou des fonctions en escalier est donc une trivialité. Mais, encore faut-il alors identifier l'espace complet obtenu avec l'ensemble des classes d'équivalence pour l'égalité presque partout des fonctions Lebesgue ou Borel-mesurables et de carrés intégrables. Dans certains livres la discussion de cela n'est pas assez satisfaisante. Une fois connue la densité des fonctions en escalier on a l'important :

Théorème de continuité des translations : pour f intégrable on a $\lim_{h \dashrightarrow 0} \|f - f_h\|_1 = 0$ avec $f_h(x) = f(x - h)$. (idem pour f de carré intégrable et $\|\cdot\|_2$.)

Lorsque nous aurons le vocabulaire adapté nous aurons l'énoncé plus complet qui dit que l'application $h \dashrightarrow f_h$ est continue. En combinant le théorème de Fejér et le théorème de continuité des translations, on a montré, pour f intégrable $\lim \|f - \Sigma_N(f)\|_1 = 0$ (théorème de Fejér au sens L^1). Il en résulte :

Théorème d'unicité : si deux fonctions intégrables ont la même série de Fourier elles sont égales presque partout.

Enfin, nous avons récemment prouvé un autre théorème essentiel, en utilisant les théorèmes de l'intégration selon Lebesgue :

Théorème de Riesz-Fischer : si les nombres complexes c_n vérifient $\sum_{n\in\mathbb{Z}} |c_n|^2 < \infty$ alors il existe une fonction f de carré intégrable avec $\forall n \in \mathbb{Z} \quad c_n(f) = c_n$.

Il y a donc une bijection entre l'espace petit $l^2(\mathbb{Z})$ des suites indexées par $\mathbb{Z}$ et de carrés sommables et l'espace $L^2(0, 2\pi; \frac{dx}{2\pi})$. En ce moment nous complétons (sic) cette discussion en prouvant que ces deux espaces, munis des formes sesquilinéaires naturelles et de la norme correspondante, sont complets en tant qu'espaces métriques. En fait c'est souvent cela que l'on appelle << Théorème de Riesz-Fischer >>.

Les espaces L^1 et L^2 sont complets

Mode << télé-enseignement >> (no comments). v1 : 17 mars ; rev. 4 avril.

L'espace vectoriel $L^1(0, 2\pi; \frac{dx}{2\pi})$ est l'espace des fonctions mesurables intégrables sur l'intervalle $[0, 2\pi]$, à l'équivalence pour l'égalité presque partout près. On peut aussi voir ces (classes de) fonctions comme des fonctions périodiques (imposer $f(2\pi) = f(0)$ est possible puisque les singletons sont de mesure nulle). On peut avoir besoin de l'espace $\mathcal{L}^1(0, 2\pi; \frac{dx}{2\pi})$ des fonctions mesurables intégrables, encore que là il faudrait être plus précis dans la notation : $\mathcal{L}^1([0, 2\pi]; \frac{dx}{2\pi})$, ou $\mathcal{L}^1([0, 2\pi[; \frac{dx}{2\pi})$ ou encore $\mathcal{L}^1(]0, 2\pi[; \frac{dx}{2\pi})$ ou encore l'espace $\mathcal{L}^{1,\mathrm{per}}$ des fonctions 2π-périodiques mesurables, intégrables sur une période, ce sont à chaque fois des choses un peu différentes. Mais bon en général on travaille avec l'espace $L^1(0, 2\pi; \frac{dx}{2\pi})$ des classes d'équivalence car on veut pouvoir dire $\|f\|_1 = 0 \Longrightarrow f = 0$. Allez, je suis totalement honnête, on travaille dans cet espace de classes d'équivalence, mais un symbole comme f en général fait référence à une fonction, pas à une classe d'équivalence. Et on rajoute des p.p. (<< presque partout >>) à toutes les égalités, ou inégalités. Sauf que parfois on oublie ; ou alors on décide soudainement sans prévenir que ça y est, le f c'est plus une fonction, c'est une classe d'équivalence. Voilà ma politique en fait : je veux pouvoir dire << ce qui prouve $f_1(x) = f_2(x)$ p.p. donc $f_1 = f_2$ dans L^1. >> En effet j'ai énormément de mal à me dire que cos(x) par exemple, c'est plusieurs choses différentes suivant que je le vois dans L^1 ou dans L^2, ou dans $\mathcal{C}^{\infty,\mathrm{per}}$. Je préfère voir cos(x) comme une chose qui a plusieurs aspects. Donc je veux pouvoir dire $\cos \in L^1$ et aussi $\cos \in L^2$ (et aussi $\cos \in C^{\infty,\mathrm{per}}$). Dans le même temps je ne veux pas non plus considérer L^2 comme réellement un sous-ensemble de L^1, j'ai toujours en tête plutôt la flèche injective (grâce à Cauchy-Schwarz) $L^2 \hookrightarrow L^1$. Donc il y a L^2 d'un côté, L^1 de l'autre, et le premier s'injecte dans le second mais je ne veux pas vraiment le voir comme un sous-ensemble du second. Bref, il y a un petit aspect artistique dans tout cela qui en fait tout son charme.

J'ai déjà mentionné $L^2(0, 2\pi; \frac{dx}{2\pi})$. Je rappelle que (c'est lié à Cauchy-Schwarz) on a la norme $\|\cdot\|_2$. En posant $d(f, g) = \|f - g\|_2$ on définit une notion de distance sur l'espace vectoriel L^2, qui devient un espace métrique (on dit que l'on a un espace vectoriel normé). De même L^1 est un espace métrique. Dans tout espace métrique on a la notion de suite de Cauchy, et si toute suite de Cauchy converge on dit que l'espace métrique est complet. Vous savez tout cela.

Théorème : *les espaces vectoriels normés* $L^1(0, 2\pi; \frac{dx}{2\pi})$ *et* $L^2(0, 2\pi; \frac{dx}{2\pi})$ *sont complets en tant qu'espaces métriques.*

Un espace vectoriel normé complet est appelé << espace de Banach >>. Souvent on dit que le théorème précédent (ou le suivant) est le << Théorème de Riesz-Fischer >>. À cette époque ils ne disposaient pas de notre vocabulaire (la thèse de Banach c'est vers 1920.) Dans le cas de L^2 la norme est associée à une forme sesquilinéaire (f,g) et on parle alors d'<< espace de Hilbert >>.(*)

Théorème : *$L^2(0,2\pi;\frac{dx}{2\pi})$ est un Hilbert et $L^1(0,2\pi;\frac{dx}{2\pi})$ est un Banach.*

Il y a peu de théorèmes aussi importants en Mathématiques. Dans la suite du semestre nous allons étudier plusieurs notions liées aux espaces de Hilbert.

Théorème : *l'espace petit $l^2(\mathbb{Z})$ est un espace de Hilbert.*

En fait par l'identité de Parseval (Fatou, est-ce dans sa thèse de 1906 ou plus tard, à vérifier) et le théorème de Riesz-Fischer de 1907, il y a une bijection entre $L^2(0,2\pi;\frac{dx}{2\pi})$ et $l^2(\mathbb{Z})$ compatible aux normes. Pour prouver que $L^2(0,2\pi;\frac{dx}{2\pi})$ est complet on pourrait se contenter de prouver que $l^2(\mathbb{Z})$ est complet, ce qui est plus facile, car n'utilisant pas les théorèmes de l'intégrale de Lebesgue. Mais comme je veux aussi prouver que L^1 est complet, je ne prends pas ce raccourci. Comme je vois que la page s'amenuise dangereusement FAITES L^2 VOUS-MÊMES ! (cf. technique de l'annexe Riesz-Fischer).

Preuve que $L^1(0,2\pi;\frac{dx}{2\pi})$ est complet : je considère une suite de Cauchy (f_n), $n = 1,2\ldots$. Je rappelle que cela veut dire $\lim_{N\to\infty}\sup_{n,m\geq N}\|f_n - f_m\|_1 = 0$. Il y a un exercice classique dans les espaces métriques qui demande de montrer que si une suite de Cauchy a une suite extraite convergente elle est elle-même convergente. Donc on va juste se contenter de construire une suite extraite convergente. D'abord on prend $N_1 < N_2 < N_3 < \ldots$ tels que $\sup_{n,m\geq N_k}\|f_n-f_m\|_1 \leq \frac{1}{2^k}$. On pose(**) alors $g_0(x) = f_{N_1}(x)$, $g_1(x) = f_{N_2}(x) - f_{N_1}(x)$, etc..., $g_k(x) = f_{N_{k+1}}(x) - f_{N_k}(x)$. On a donc

$$f_{N_{K+1}}(x) = g_0(x) + g_1(x) + \cdots + g_K(x)$$

Définissons aussi $G_K(x) = |g_0(x)| + |g_1(x)| + \cdots + |g_K(x)|$ et même $G(x) = \sum_{k=0}^{\infty}|g_k(x)|$. Par le théorème de la convergence monotone $\int_0^{2\pi}G(x)\,\frac{dx}{2\pi} \leq \|g_0\|_1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots < +\infty$. Du coup le lieu des x avec $G(x) = +\infty$ est de mesure nulle. Pour ceux là posons $F(x) = 0$. Pour les autres la série $\sum_{k=0}^{\infty}g_k(x)$ est absolument convergente et on pose $F(x) = \sum_{k=0}^{\infty}g_k(x)$. On a donc pour tous les x, $|F(x)| \leq G(x)$ donc $F \in L^1$. De plus pour presque tout x on a $F(x) = f_{N_{K+1}}(x) + \sum_{j=K+1}^{\infty}g_j(x)$ donc presque partout(***) :

$$|F(x) - f_{N_{K+1}}(x)| \leq \sum_{j=K+1}^{\infty}|g_j(x)|$$

et donc (convergence monotone) $\|F - f_{N_{K+1}}\|_1 \leq \frac{1}{2^K}$. C'est fini. Ok ?

(*) l'article fondateur de Hilbert date de 1906 ; mais, bizarrement, ce n'est qu'en 1929 que la notion d'<< espace de Hilbert >> est présentée axiomatiquement par von Neumann. Les travaux de Riesz, Weyl, Carleman sur la théorie spectrale des opérateurs dans les Hilbert concernaient des contextes plus concrets d'opérateurs intégraux et d'espaces de fonctions.

(**) petite subtilité : on a imposé à chaque f_n d'être partout finie. Pourquoi ?

(***) en fait partout, pourquoi ?

Les espaces L^p

Dans ce Cours, notre intérêt principal est dans l'étude des espaces normés $L^1(0,2\pi;\frac{dx}{2\pi})$ et $L^2(0,2\pi;\frac{dx}{2\pi})$. Sous l'influence de mathématiciens tels que Hardy et Riesz dans la première moitié du vingtième siècle, la boîte à outils de l'Analyse s'est élargie à la considération des variantes que sont les espaces << L^p >>. Nous n'aurons pas vraiment l'occasion de voir les espaces L^p en action, ce que nous ferons ici simplement c'est de les définir et de prouver qu'ils sont complets. Leur existence même repose sur des inégalités importantes qui généralisent Cauchy-Schwarz, les inégalités de Hölder et de Minkowski : nous commencerons par cela. Ensuite nous prouvons qu'ils sont des espaces normés complets (espaces de Banach). Il y a de plus un important théorème de dualité, mais je ne suis pas assez motivé aujourd'hui pour en rédiger une preuve. Mais en tout cas, cela sera fait à un moment ou un autre, au moins dans le cas particulier de l'espace de Hilbert L^2.

Dans cette feuille je travaille en toute généralité sur un ensemble X, avec une tribu et une mesure μ. Je ne suppose pas que μ soit une mesure finie, donc par exemple on peut avoir $X = \mathbb{R}$ et $d\mu = dx$, et bien sûr il y aussi notre exemple favori $X = \mathbb{R}/2\pi\mathbb{Z}$ et $d\mu = \frac{dx}{2\pi}$ dans ce cas X est de masse totale 1 et on aura des inclusions naturelles $L^{p_2} \subset L^{p_1}$ pour $p_2 > p_1$, qui généralisent $L^2 \subset L^1$. Mais pour $X = \mathbb{R}$, $d\mu = dx$, il n'y a pas de telles inclusions.

Sous-jacent à toute notre discussion, il y a la notion de convexité ; et on pourrait beaucoup développer ce thème (théorie des espaces localement convexes). Je me contente ici de vous rappeler que lorsqu'une fonction f vérifie $f'' \geq 0$ alors $f(ux + vy) \leq uf(x) + vf(y)$ pour $u + v = 1$, $0 \leq u \leq 1$. On va prendre $f(x) = e^x$, et $u = \frac{1}{p}$ avec $p > 1$. On écrit aussi $v = 1 - u = \frac{1}{q}$ avec $q > 1$ et on dit que (p,q) est une paire d'exposants conjugués. Si on voulait étendre à $p = 1$ il faudrait alors $q = \infty$ pour que $\frac{1}{p} + \frac{1}{q} = 1$. Mais justement ce cas limite se trouve avoir dans le contexte des espaces L^p un comportement différent donc nous nous occuperons ici seulement du cas $1 < p < \infty$, $1 < q < \infty$. Pour $p = 2$ l'exposant conjugué est $q = 2$.

En posant $e^x = a^p$, $e^y = b^q$, donc $ux = \frac{1}{p}x = \log(a)$, $vy = \log(b)$, on obtient, pour $0 < a, b < +\infty$:

$$ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q \qquad (0 \leq a \leq +\infty, 0 \leq b \leq +\infty, p > 1, q > 1, \frac{1}{p} + \frac{1}{q} = 1)$$

Les cas $a = 0$ ou $+\infty$ ou $b = 0$ ou $+\infty$ nécessitent un examen particulier, et on vérifie que cela marche (avec la convention $0.\infty = 0$ toujours utilisée en théorie de l'intégration). Il est important de vérifier ces cas limites car nous voulons appliquer cette inégalité aux valeurs $a = |f(x)|$ et $b = |g(x)|$ prises par les valeurs absolues (modules) de deux fonctions mesurables sur X, donc certainement nous ne voulons pas exclure la valeur 0, et aussi il est commode de ne pas exclure la valeur $+\infty$. En intégrant nous obtenons

$$\int_X |f(x)g(x)|d\mu(x) \leq \frac{1}{p}\int_X |f(x)|^p d\mu(x) + \frac{1}{q}\int_X |g(x)|^q d\mu(x)$$

Supposons $f \in \mathcal{L}^p$, c'est-à-dire $\int_X |f(x)|^p d\mu(x) < \infty$ et $g \in \mathcal{L}^q$, c'est-à-dire $\int_X |g(x)|^q d\mu(x) < \infty$. Nous voyons en tout cas que $\int_X |f(x)g(x)|d\mu(x) < \infty$, donc $fg \in \mathcal{L}_1$. Plus précisément posons $\|f\|_p = \left(\int_X |f(x)|^p d\mu(x)\right)^{1/p}$ et idem pour $\|g\|_q$. Remarquez l'homogénéité :(*) $\|\alpha \cdot f\|_p = |\alpha|\|f\|_p$, $\|\beta \cdot g\|_q = |\beta|\|g\|_q$. Remplaçons f par $f_1 = \frac{f}{\|f\|_p}$, bien sûr cela nous ne pouvons le faire que si f n'est pas presque partout nulle, et aussi si g n'est pas presque partout nulle, remplaçons g par $g_1 = \frac{g}{\|g\|_q}$. Vérifiez que $\|f_1\|_p = 1$ et $\|g_1\|_q = 1$, donc par l'inégalité ci-dessus $\|f_1 g_1\|_1 \leq \frac{1}{p} + \frac{1}{q} = 1$, donc :

Inégalité de Hölder : $$\int_X |f(x)g(x)|d\mu(x) \leq \left(\int_X |f(x)|^p \, d\mu(x)\right)^{\frac{1}{p}} \left(\int_X |g(x)|^q \, d\mu(x)\right)^{\frac{1}{q}}$$

Si f ou g est presque partout nulle, alors le produit fg l'est aussi, le terme de gauche vaut 0 et l'inégalité est valable. Si $\|f\|_p = \infty$ et si g n'est pas presque partout nulle le terme de droite vaut $+\infty$ et l'inégalité est aussi valable. Bref, l'inégalité est toujours valable. Mais bien sûr c'est principalement pour $\|f\|_p < \infty$, et $\|g\|_q < \infty$ qu'elle est intéressante, et alors on a donc $fg \in \mathcal{L}^1$ et on peut écrire :

$$\left|\int_X f(x)g(x)d\mu(x)\right| \leq \int_X |f(x)g(x)|d\mu(x) \leq \|f\|_p\|g\|_q < +\infty$$

Remarquez que pour $p = q = 2$ on retrouve l'inégalité de Cauchy-Schwarz. À partir de l'inégalité de Hölder, on va prouver l'inégalité de Minkowski :

Inégalité de Minkowski : $\|f_1 + f_2\|_p \leq \|f_1\|_p + \|f_2\|_p \qquad (\|f_1\|_p < \infty, \|f_2\|_p < \infty)$

À propos, pour que $f_1 + f_2$ existe on doit exclure des cas comme $+\infty + (-\infty)$; pour cela on peut supposer $f_1, f_2 \geq 0$, ou encore supposer que f_1 et f_2 sont toutes deux presque partout finies. Avec $\|f_1\|_p < \infty$, $\|f_2\|_p < \infty$ on est bien dans ce deuxième cas de figure. Ensuite, $|f_1 + f_2| \leq |f_1| + |f_2|$ et $\|f_1\|_p = \|\,|f_1|\,\|_p$, $\|f_2\|_p = \|\,|f_2|\,\|_p$, donc on peut pour la preuve remplacer f_1 par $|f_1|$, f_2 par $|f_2|$, ce qui revient à supposer $f_1 \geq 0$, $f_2 \geq 0$. Ensuite on écrit :(**)

$$\int_X f_1 \cdot (f_1 + f_2)^{p-1} d\mu(x) \leq \|f_1\|_p \left(\int_X (f_1 + f_2)^{qp-q} \, d\mu(x)\right)^{\frac{1}{q}}$$

(*) la règle $0.\infty = 0$ fait que cela marche aussi si $\|f\|_p = +\infty$.
(**) remarque : $qp - q = p$.

$$\int_X f_2\cdot(f_1+f_2)^{p-1}d\mu(x) \le \|f_2\|_p\left(\int_X (f_1+f_2)^{qp-q}\,d\mu(x)\right)^{\frac{1}{q}}$$

$$\int_X (f_1+f_2)^p d\mu(x) \le (\|f_1\|_p+\|f_2\|_p)\left(\int_X (f_1+f_2)^p\,d\mu(x)\right)^{\frac{1}{q}}$$

Ici il y a un piège dans la rédaction. SI ON SUPPOSE $\int_X (f_1+f_2)^p\,d\mu(x) < \infty$, ALORS on peut diviser des deux côtés et on obtient effectivement :

$$\|f_1+f_2\|_p = \left(\int_X (f_1+f_2)^p\,d\mu(x)\right)^{1-\frac{1}{q}} \le \|f_1\|_p + \|f_2\|_p$$

Mais il faut D'ABORD justifier $\int_X (f_1+f_2)^p\,d\mu(x) < \infty$ (toujours sous l'hypothèse $f_1, f_2 \ge 0$). Par exemple : $(f_1+f_2)^p \le (2\max(f_1,f_2))^p = 2^p\max(f_1^p, f_2^p) \le 2^p(f_1^p+f_2^p)$ donc effectivement $\int_X (f_1+f_2)^p\,d\mu(x) < \infty$ si $\int_X f_1^p\,d\mu(x) < \infty$ et $\int_X f_2^p\,d\mu(x) < \infty$.

Remarque : si je ne m'abuse les vrais inégalités de Hölder et de Minkowski portent sur des sommes finies (on les retrouve en prenant pour X un ensemble fini muni de la mesure de dénombrement.) C'est F. Riesz qui a prouvé la version avec des intégrales (probablement d'abord pour $X = \mathbb{R}$ ou un intervalle, et $d\mu = dx$). L'inégalité de Minkowski donne $\|\alpha f + \beta g\|_p \le |\alpha|\|f\|_p + |\beta|\|g\|_p$, et en définissant L^p comme le quotient de $\mathcal{L}^p$ pour la relation d'équivalence donnée par l'égalité presque partout on obtient un espace vectoriel normé. Aussi, on a par récurrence $\|f_1+f_2+\cdots+f_N\|_p \le \|f_1\|_p+\cdots+\|f_N\|_p$. Nous allons montrer :

Théorème : *$L^p(X,\mu)$ est un espace de Banach.*

Autrement dit L^p est complet comme espace métrique. La preuve est très semblable à celle que nous avons donnée pour L^1 et laissée en exercice pour L^2. Je considère une suite de Cauchy (f_n), $n = 1, 2 \ldots$. Je rappelle que cela veut dire $\lim_{N\to\infty}\sup_{n,m\ge N}\|f_n - f_m\|_p = 0$. Il y a un exercice classique dans les espaces métriques qui demande de montrer que si une suite de Cauchy a une suite extraite convergente elle est elle-même convergente. Donc on va juste se contenter de construire une suite extraite convergente. D'abord on prend $N_1 < N_2 < N_3 < \ldots$ tels que $\sup_{n,m\ge N_k}\|f_n - f_m\|_p \le \frac{1}{2^k}$. On pose(*) alors $g_0(x) = f_{N_1}(x)$, $g_1(x) = f_{N_2}(x) - f_{N_1}(x)$, etc..., $g_k(x) = f_{N_{k+1}}(x) - f_{N_k}(x)$. On a donc

$$f_{N_{K+1}}(x) = g_0(x) + g_1(x) + \cdots + g_K(x)$$

Définissons aussi $G_K(x) = |g_0(x)| + |g_1(x)| + \cdots + |g_K(x)|$ et même $G(x) = \sum_{k=0}^\infty |g_k(x)|$. Par le théorème de la convergence monotone on a :

$$\int_X G(x)^p\,d\mu(x) = \lim_{K\to\infty}\int_X (\sum_{k=0}^K |g_k(x)|)^p\,d\mu(x)$$

et par l'inégalité de Minkowski :

$$\int_X (\sum_{k=0}^K |g_k(x)|)^p\,d\mu(x) \le (\|g_0\|_p + \frac{1}{2} + \cdots + \frac{1}{2^K})^p < (\|g_0\|_p + 1)^p$$

(*) petite subtilité : on a imposé à chaque f_n d'être partout finie. Pourquoi ?

Donc $\int_X G(x)^p d\mu(x) \leq (\|g_0\|_p + 1)^p < \infty$ et du coup le lieu des x avec $G(x) = +\infty$ est de mesure nulle. Pour ceux là posons $F(x) = 0$. Pour les autres la série $\sum_{k=0}^{\infty} g_k(x)$ est absolument convergente et on pose $F(x) = \sum_{k=0}^{\infty} g_k(x)$. On a donc pour tous les x, $|F(x)| \leq G(x)$ donc $F \in L^p$, puisque l'on a vu $G \in L^p$. De plus pour presque tout x on a $F(x) = f_{N_{K+1}}(x) + \sum_{j=K+1}^{\infty} g_j(x)$ donc presque partout(**) :

$$|F(x) - f_{N_{K+1}}(x)| \leq \sum_{j=K+1}^{\infty} |g_j(x)|$$

et donc par convergence monotone et par Minkowski :

$$\begin{aligned}\int_X |F(x) - f_{N_{K+1}}(x)|^p d\mu(x) &\leq \lim_{L \to \infty} \int_X (\sum_{j=K+1}^{L} |g_j(x)|)^p d\mu(x) \\ &\leq \lim_{L \to \infty} (\|g_{K+1}\|_p + \cdots + \|g_L\|_p)^p \leq (\frac{1}{2^K})^p\end{aligned}$$

Donc $\|F - f_{N_{K+1}}\|_p \leq \frac{1}{2^K}$ et ainsi F est la limite de la suite f_{N_k} au sens L^p, donc aussi la limite de la suite f_n puisque celle-ci est une suite de Cauchy.

Cette preuve nous a fourni une information très utile : de toute suite (f_n) convergente dans L^p, de limite F au sens L^p, on peut extraire une sous-suite f_{n_k} de sorte que pour presque tout x on a convergence ponctuelle $\lim f_{n_k}(x) = F(x)$. En effet la fonction F de la preuve ci-dessus est forcément identique (dans L^p, donc au sens du presque partout) avec celle du présent paragraphe car dans un espace métrique une suite ne peut avoir qu'une seule limite. En général on ne peut pas dire plus, mais dans le cas particulier des séries de Fourier, Carleson (prix Abel 2006) a démontré (1965-1966) que la série de Fourier d'une fonction f de $L^2(\mathbb{R}/2\pi\mathbb{Z}; \frac{dx}{2\pi})$ non seulement converge au sens L^2 vers f, on a même presque partout $\lim S_N(f)(x) = f(x)$ (vrai (Carleson-Hunt) pour $f \in L^p$, $p > 1$; faux (Kolmogorov 1922) pour certaines $f \in L^1$).

Bon, considérons pour conclure le cas particulier d'un espace de masse 1 : $\mu(X) = 1$. Alors en prenant dans l'inégalité de Hölder $f \in \mathcal{L}^p$ et $g = 1$ on obtient $\|f\|_1 \leq \|f\|_p$. Donc $\mathcal{L}^p \subset \mathcal{L}^1$ et en passant au quotient on peut si l'on veut considérer $L^p \subset L^1$ (mais comme je l'ai dit déjà je préfère voir cette inclusion comme une flèche injective). Plus généralement supposons $1 \leq p_1 < p_2$, et posons $p = \frac{p_2}{p_1}$ de sorte que $p > 1$. Soit f une fonction mesurable et appliquons l'inégalité de Hölder à $|f|^{p_1}$ et 1 pour le couple d'exposants (p, q), $\frac{1}{p} + \frac{1}{q} = 1$. On obtient $(\|f\|_{p_1})^{p_1} = \| |f|^{p_1} \|_1 \leq \| |f|^{p_1} \|_p = (\|f\|_{p_2})^{p_1}$ donc $\|f\|_{p_1} \leq \|f\|_{p_2}$. On a ainsi, pour $1 \leq p_1 < p_2$: $\mathcal{L}^{p_2} \subset \mathcal{L}^{p_1}$ et, si l'on veut, $L^{p_2} \subset L^{p_1}$.

Lorsque X n'est plus de masse totale 1 mais toujours de masse finie, ces inclusions subsistent, et sont continues pour les topologies définies par les normes $\|\cdot\|_p$, la seule différence, c'est que $\|f\|_{p_1} \leq \|f\|_{p_2}$ devient $\|f\|_{p_1} \leq \mu(X)^{\frac{1}{p_1} - \frac{1}{p_2}} \|f\|_{p_2}$.

(**) en fait partout, pourquoi ?

Le théorème de Fréchet-Riesz

Je ne redonne pas ici la discussion du cours autour de la notion de projection orthogonale. Pour la dimension finie se reporter aux exercices de la feuille 4. Pour la dimension infinie on a besoin du point suivant :

Hyper-important (prouvez-le !) : Soit V un espace hermitien et $W \subset V$ un sous-espace vectoriel. On suppose que W est, pour le produit scalaire induit de V, un espace de Hilbert. Montrer que pour tout $v \in V$ il existe un unique vecteur $w \in W$ minimisant $\|v - w\|$, et que cela équivaut à $v - w \perp W$. Ind. : prendre une suite $(w_j)_{j\geq 1}$ avec $\|v - w_j\| \to \inf_{w\in W} \|v - w\|$ et utiliser

$$\|w_j - w_k\|^2 + 4\|\frac{w_j + w_k}{2} - v\|^2 = 2\|w_j - v\|^2 + 2\|w_k - v\|^2$$

pour affirmer avec force que (w_j) est de Cauchy, donc convergente dans W, vers une limite w qui est ce que l'on recherche.

Si V lui-même est supposé être un Hilbert alors le sous-espace vectoriel $W \subset V$ le sera si et seulement si il est fermé pour la topologie de V (prouvez-cela).

Lorsque vous avez une fonction continue sur un espace topologique, le lieu de ses zéros est un fermé. Donc si $L : V \to \mathbb{C}$ est une fonction continue sur un espace de Hilbert, à valeurs dans $\mathbb{C}$ (muni de sa topologie standard) l'ensemble $W = \{u \mid L(u) = 0\}$ est fermé. Si L est aussi une application linéaire ($L(\alpha u + \beta v) = \alpha L(u) + \beta L(v)$), alors on a la garantie que le fermé W est un sous-espace vectoriel de V. Lorsque l'on a une application linéaire L de V vers $\mathbb{C}$ on dit que l'on a une forme linéaire. Quelle est la condition pour qu'une forme linéaire soit continue ? Comme $L(u_0 + v) = L(u_0) + L(v)$ la continuité de L au point u_0 découle de la continuité de L au point 0. Pour cela il faut en particulier qu'il existe un $\delta > 0$ tel que $0 < \|v\| < \delta$ implique $|L(v)| \leq 1$. Prenons $v \neq 0$ quelconque, et soit $v' = \frac{1}{2}\delta\|v\|^{-1}v$ de sorte que $\|v'\| = \frac{1}{2}\delta$. On a $|L(v')| \leq 1$ donc $|L(v)| \leq \frac{2}{\delta}\|v\|$. Réciproquement si il existe $C < \infty$ tel que $\forall v\ |L(v)| \leq C\|v\|$ alors (clair ?) L est continue en 0, donc partout. Ce que nous venons d'expliquer vaut pour tout espace vectoriel normé : une forme linéaire L est continue si et seulement si elle est bornée, c'est-à-dire si il existe $C < \infty$ tel que $\forall v\ |L(v)| \leq C\|v\|$. On posera alors $\|L\| = \sup_{\|v\|=1} |L(v)|$, qui est le plus petit C qui puisse marcher. L'ensemble V^* des formes linéaires continues sur l'espace vectoriel normé V est un espace vectoriel, et (vérifiez !) il est lui-même un espace normé par la définition que nous venons d'adopter pour $\|L\|$. C'est le dual de V.

Le Théorème de Fréchet-Riesz détermine les formes linéaires continues sur un

Hilbert V. Prenons un $u \in V$ et définissons $L(v) = (v,u)$. C'est une forme linéaire. Elle est continue, car $|L(v)| = |(v,u)| \leq \|v\|\|u\|$, donc elle est bornée et l'on a $\|L\| \leq \|u\|$. En prenant $v = \|u\|^{-1}u$, on a $\|v\| = 1$ et $L(v) = \|u\|^{-1}(u,u) = \|u\|$, donc $\|L\| \geq \|u\|$. Ainsi $\|L\| = \|u\|$. En associant à u la forme linéaire L on a une application (isométrique) de V vers V^*. Petite contrariété, cette application n'est pas linéaire, elle est conjuguée-linéaire (vous comprenez ?). Le théorème de Fréchet-Riesz affirme que cette association est bijective. L'injectivité est claire (gare à vous sinon), reste à montrer la surjectivité.

Preuve : on se donne une L, forme linéaire continue sur V et il faut trouver le u (unique ; pourquoi ?) tel que $\forall v \in V$ $L(v) = (v,u)$. Si $L \equiv 0$ on prend $u = 0$. Sinon on prend u_1 avec $L(u_1) = 1$ (why is this possible ?). Et par ailleurs on regarde le sous espace vectoriel fermé $W = \{v \in V \mid L(v) = 0\}$. On fait la projection orthogonale de u_1 sur W, cela donne un vecteur v_1. Soit $u_0 = u_1 - v_1$. On a toujours $L(u_0) = 1$. De plus nous savons que u_0 est perpendiculaire à W : $L(w) = 0 \Longrightarrow (w,u_0) = 0$. Maintenant prenons v quelconque et regardons le vecteur $w = v - L(v)u_0$. Calculons $L(w) = L(v) - L(v) = 0$. Donc $w \in W$ et donc $0 = (w,u_0) = (v,u_0) - L(v)(u_0,u_0)$. Certainement $(u_0,u_0) > 0$. Posons $u = \frac{u_0}{(u_0,u_0)}$. Alors ce u marche (ALLES KLAR ?).

Tout espace de Hilbert est donc canoniquement (anti-)isomorphe à son dual. Il est donc canoniquement isomorphe à son double-dual V^{**}.

D'une manière générale, pour tout espace vectoriel normé on a une application linéaire canonique de V vers son double dual V^{**}, car tout $v \in V$ définit une forme linéaire continue sur V^* par $L \mapsto L(v)$ (warum ?). Cette application de V vers son double-dual V^{**} est toujours injective, car si $v \in V$ est non nul, il existe une forme linéaire continue $L \in V^*$ telle que $L(v) \neq 0$. Ceci n'est pas du tout trivial, en fait c'est une partie d'un théorème de base en Analyse Fonctionnelle, le théorème de Hahn-Banach. Lorsque l'injection canonique $V \subset V^{**}$ est bijective, on dit que l'espace vectoriel normé V est réflexif. Les Hilbert sont donc réflexifs. Qu'ils soient en fait déjà (anti)-isomorphes (canoniquement) à leur dual V^* est très très spécial.

Dernière remarque : un espace vectoriel normé V se << complète >> en un Banach $V^{\sim}$ par la méthode introduite par Cantor pour constuire $\mathbb{R}$ à partir de $\mathbb{Q}$, la méthode des suites de Cauchy. Comme $V \subset V^{\sim}$ on a $V^{\sim *} \subset V^*$: cette deuxième inclusion est en fait une égalité, prouvez-le.

Vraiment dernière remarque : pour tout espace vectoriel normé V son dual V^* est automatiquement complet ; c'est toujours un Banach. Prouvez-le.

Le théorème de dualité des espaces L^p

Version provisoire du 2 avril 2006.

Dans un article de 1910 qui a profondément influencé l'Analyse mathématique du vingtième siècle, << Untersuchungen über Systeme integrierbarer Funktionen >>, Math. Annalen 69, 449-497, F. Riesz prouve (parmi d'autres choses) un résultat que l'on peut reformuler ainsi :

Théorème : *Soit* X *un espace mesuré*(*), *soit* $1 < p < \infty$, *d'exposant conjugué* q ($\frac{1}{p} + \frac{1}{q} = 1$), *alors le dual de* $L^p(X,\mu)$ *est canoniquement* $L^q(X,\mu)$ *via l'appariement* $(u,v) \mapsto \int_X u(x)v(x)d\mu(x)$.

Riesz ne considère (je crois) que le cas $X = I$, $I \subset \mathbb{R}$ un intervalle, muni de la mesure de Lebesgue $d\mu = dx$ (j'avoue, piteusement, ne jamais encore avoir consulté les Oeuvres de Riesz, à l'exception de son livre d'Analyse Fonctionnelle avec Sz-Nagy et c'est donc aussi le moment tragique de vous révéler que je ne suis nullement un expert des choses que je suis censé(**) vous enseigner ; pire je n'ai réellement réfléchi au contenu de cette feuille que ces derniers jours).

Par l'inégalité de Hölder (qui, pour les intégrales, fut en fait énoncée par Riesz dans ce même article de 1910, je crois) on sait déjà qu'effectivement chaque $v \in L^q(X,\mu)$ définit une forme linéaire continue L sur $L^p(X,\mu)$ via $L(u) = \int_X u(x)v(x)d\mu(x)$, et $\|L\| \leq \|v\|_q$. Supposons $v \neq 0$ dans L^q et prenons $u = v^{-1}|v|^q$ (pour les x avec $v(x) = 0$ on pose $u(x) = 0$). Notons (aussi pour ces x) que $|u|^p = |v|^{-p+pq} = |v|^q$ donc $u \in L^p$ et $(\|u\|_p)^p = (\|v\|_q)^q$. De plus pour ce u on constate que $L(u) = \int_X |v(x)|^q \, d\mu(x)$. Il en résulte $\|L\| \geq \frac{|L(u)|}{\|u\|_p} = \|v\|_q^q \, \|v\|_q^{-q/p} = \|v\|_q$. Donc en fait $\|L\| = \|v\|_q$.(***)

On a ainsi une injection isométrique $L^q \hookrightarrow (L^p)^*$ et il s'agit de montrer qu'elle est surjective, c'est-à-dire, pour toute forme linéaire continue L sur L^p trouver v dans L^q telle que $\forall u \in L^p \quad L(u) = \int_X uv d\mu$. Ce n'est pas (du tout)

(*) donc, muni d'une tribu et d'une mesure. Dans la notation $L^p(X,\mu)$ je ne fais figurer que la mesure, la tribu reste implicite.

(**) le suis-je vraiment ? ça n'a pas l'air d'être considéré si important que cela en fait ces jours-ci.

(***) à propos, l'inégalité de Minkowski $\|v_1+v_2\|_q \leq \|v_1\|_q+\|v_2\|_q$ peut se concevoir comme une conséquence de cela, pourquoi ?.

évident. Nous allons chercher autant que possible à imiter la preuve faite dans le cas $p = q = 2$ (théorème de Fréchet-Riesz[(****)]). En particulier un rôle important est joué par la notion de << projection >> sur un sous-espace vectoriel fermé. Cependant nous verrons que la situation est, si l'on attaque le problème frontalement, plus délicate pour $1 < p < 2$ que pour $p > 2$.

Nous pouvons supposer L non identiquement nulle, et nous notons W son noyau qui est un sous-espace vectoriel fermé de L^p. Nous prenons u_0 avec $L(u_0) = 1$, et, nous admettons (la discussion relative à ce point vient ensuite) qu'il existe $w_0 \in W$ qui a la propriété $\|u_0 - w_0\|_p = \inf_{w \in W} \|u_0 - w\|_p$.

Soit maintenant $u_1 = u_0 - w_0$, alors à nouveau $L(u_1) = 1$. Remplacer u_0 par u_1[(*)] c'est comme remplacer w_0 par 0, et on voit qu'effectivement on a :

$$\forall w \in W \quad \|u_1\|_p \leq \|u_1 - w\|_p$$

Considérons, pour $w \in W$ fixé, et $(t, x) \in \mathbb{R} \times X$:

$$f(t,x) = |u_1(x) - tw(x)|^p \qquad f(t) = \int_X f(t,x)d\mu(x) = (\|u_1 - t\,w\|_p)^p$$

Pour $x \in X$ fixé, et en t_0 tel que $u_1(x) - t_0 w(x) \neq 0$ la fonction $t \mapsto f(t,x)$ est dérivable : pour ϵ réel, $|\epsilon| \ll 1$:

$$\begin{aligned}|u_1(x) - t_0 w(x) - \epsilon w(x)|^p &= |u_1(x) - t_0 w(x)|^p (1 - \epsilon\, 2\, \Re \frac{w(x)}{u_1(x) - t_0 w(x)} + \mathcal{O}(\epsilon^2))^{p/2} \\ &= |u_1(x) - t_0 w(x)|^p (1 - \epsilon\, p\, \Re \frac{w(x)}{u_1(x) - t_0 w(x)} + \mathcal{O}(\epsilon^2))\end{aligned}$$

$$\begin{aligned}f'(t_0, x) &= -p|u_1(x) - t_0 w(x)|^p\, \Re \frac{w(x)}{u_1(x) - t_0 w(x)} \\ &= -p\,|u_1(x) - t_0 w(x)|^{p-1}\, \Re \left(\frac{|u_1(x) - t_0 w(x)|}{u_1(x) - t_0 w(x)} w(x) \right) \\ |f'(t_0, x)| &\leq p|u_1(x) - t_0 w(x)|^{p-1}\, |w(x)|\end{aligned}$$

Si $u_1(x) - t_0 w(x) = 0$ alors $f(t_0 + \epsilon, x) = |\epsilon|^p\, |w(x)|^p$ est dérivable en $\epsilon = 0$ de dérivée nulle ($p > 1$; aussi on a imposé à u_1 et w d'être partout finies). On peut considérer que la formule ci-dessus est encore valable. Donc $f(t,x)$ est dérivable en tout t, pour chaque x. De plus on peut, pour $-T < t < T$, majorer $|f'(t,x)|$ par $p\,(|u_1(x)| + T|w(x)|)^{p-1}|w(x)|$ qui est une fonction intégrable (pourquoi ?) indépendante de t. Par le théorème de dérivation sous le signe somme, on peut donc affirmer que la fonction $f(t)$ est une fonction dérivable, et en prenant en particulier $t_0 = 0$, on a :

$$f'(0) = -p \int_X |u_1(x)|^{p-1} \Re \left(\frac{|u_1(x)|}{u_1(x)}\, w(x) \right)\, d\mu(x)$$

[(****)] le théorème de Fréchet-Riesz détermine le dual d'un Hilbert V quelconque. Spécialisé à $V = L^2(X, \mu)$ il dit que toute forme linéaire s'écrit $u \mapsto \int_X u\overline{v}d\mu$. On remplacera v par $\overline{v}$ qui est encore dans L^2 pour voir cela comme un cas particulier pour $p = 2$ du Théorème de Dualité des L^p.

[(*)] désolé, l'annexe précédente a je crois les notations inversées.

Je rappelle qu'aux x avec $u_1(x) = 0$ on a convenu qu'il fallait comprendre que l'intégrand était nul. Maintenant par construction la fonction $t \mapsto f(t)$ a un minimum en $t = 0$, donc $f'(0) = 0$. Ainsi :

$$\forall w \in W \quad 0 = \Re\left(\int_X u_1(x)^{-1}|u_1(x)|^p\, w(x)\, d\mu(x)\right)$$

Nous pouvons toujours trouver un nombre complexe ρ de module 1 de sorte que $\rho\int_X u_1(x)^{-1}|u_1(x)|^p\, w(x)\, d\mu(x) \in \mathbb{R}$. Maintenant nous appliquons ce qui précède à ρw au lieu de w, et finalement nous concluons :

$$\forall w \in W \qquad 0 = \int_X u_1(x)^{-1}|u_1(x)|^p\, w(x)\, d\mu(x)$$

Soit $v_1 = u_1^{-1}|u_1|^p$. Cette fonction appartient à L^q et $\|v_1\|_q^q = \|u_1\|_p^p$. Soit u quelconque dans L^p, et soit $w = u - L(u)u_1$. De $L(w) = L(u) - L(u) = 0$ on a $w \in W$ et donc par ce qui précède :

$$\begin{aligned} 0 = \int_X v_1(x)w(x)\, d\mu(x) &= \int_X v_1(x)u(x)\, d\mu(x) - L(u)\int_X v_1(x)u_1(x)\, d\mu(x) \\ &= \int_X v_1(x)u(x)\, d\mu(x) - L(u)\|u_1\|_p^p \end{aligned}$$

Finalement on pose $v(x) = \|v_1\|_q^{-q}\, v_1$ et on a ce que l'on voulait :

$$\forall u \in L^p \qquad L(u) = \int_X v(x)u(x)\, d\mu(x)$$

Pour que la preuve du Théorème de Dualité de Riesz soit complète, il suffit donc de prouver le théorème suivant :(*)

Théorème de Projection : *soit* $W \subset L^p$ *un sous-espace vectoriel fermé, et soit* $u_0 \in L^p$. *Notons* $d(u_0, W) = \inf_{w\in W}\|u_0 - w\|_p$. *Alors il existe* $w_0 \in W$ *tel que* $\|u_0 - w_0\|_p = d(u_0, W)$.

Si W est de dimension finie alors il suffit de dire que la fonction continue $w \mapsto \|u_0 - w\|_p$, sur le compact $\|w\|_p \leq 2\|u_0\|_p$ atteint son minimum (si $\|w\|_p > 2\|u_0\|_p$ alors $2\|u_0\|_p < \|w\|_p \leq \|w - u_0\|_p + \|u_0\|_p$ donc $\|u_0 - w\|_p > \|u_0\|_p = \|u_0 - 0\|_p$ et ainsi l'infimum sur la boule $\{\|w\|_p \leq 2\|u_0\|_p\}$ est égal à l'infimum sur W tout entier). Mais, si W est de dimension infinie on ne peut pas procéder ainsi car F. Riesz, toujours lui, a prouvé : *dans un espace vectoriel normé* $(W, \|\cdot\|)$ *de dimension infinie la boule unité fermée n'est pas compacte.*

Comme c'est important, en voici une preuve. Si $Y \subset W$ est un sous-espace vectoriel de dimension finie, soit $x \notin Y$. Soit $d = d(x, Y) > 0$. Soit $y \in Y$ avec $\|x - y\| \leq 2d$. Soit $z = x - y$, alors $d(z, Y) = d(x, Y) = d \geq \frac{1}{2}\|z\|$ et z est non nul. Enfin soit $z' = \frac{z}{\|z\|}$. Alors $\|z'\| = 1$ et $d(z', Y) = \|z\|^{-1}d(z, Y) \geq \frac{1}{2}$. Maintenant je prouve que la boule unité n'est pas compacte. On prend z_1 de norme 1

(*) On peut remplacer << sous-espace vectoriel fermé >> par << ensemble convexe fermé >>, nous n'en avons pas besoin ici.

quelconque. Supposons connus z_1, ..., z_N. Soit Y_N l'espace vectoriel qu'ils engendrent. Il existe par ce qui précède z' de norme 1 à distance $\geq \frac{1}{2}$ de Y_N. On pose $z_{N+1} = z'$. La suite (z_n) est construite de sorte que pour tout n on a $\|z_n\| = 1$ et pour $n \neq m$ on a $\|z_n - z_m\| \geq \frac{1}{2}$. Aucune suite extraite ne peut être convergente. La boule unité n'est pas compacte.

Pour le théorème de projection il faut donc une idée autre. Dans le cas $p = 2$, la preuve classiquement proposée (il y en a de légèrement différentes), a été résumée avec indications au début de l'Annexe précédente. Elle utilise l'identité du parallélogramme (pour alléger la notation j'écris $\|f - g\|$ au lieu de $\|f - g\|_2$ etc...) :

$$\|f - g\|^2 + \|f + g\|^2 = 2\|f\|^2 + 2\|g\|^2$$

Soit $(w_j)_{j\geq 1}$ une suite dans $W \subset L^2$ avec $\|u_0 - w_j\| \dashrightarrow d(u_0, W)$. Avec $f = w_j - u_0$, $g = w_k - u_0$, on obtient :

$$\|w_j - w_k\|^2 = 4\left(\frac{\|w_j - u\|^2 + \|w_k - u\|^2}{2} - \|\frac{w_j + w_k}{2} - u_0\|^2\right)$$

Compte tenu de $\|\frac{w_j+w_k}{2} - u_0\| \geq d(u_0, W)$ il est alors facile de prouver que (w_j) est de Cauchy dans ce cas L^2. Comme L^2 est complet cette suite converge et sa limite w_0 appartient à W puisque l'on a supposé ce dernier fermé.

Pour $p \neq 2$ cela est (semble-t-il) plus compliqué pour $1 < p < 2$ que pour $p > 2$. J'explique donc d'abord comment on peut procéder pour $p > 2$. L'identité du parallélogramme n'est pas valable au niveau des normes mais elle est valable ponctuellement :

$$|f(x) - g(x)|^2 + |f(x) + g(x)|^2 = 2|f(x)|^2 + 2|g(x)|^2$$

Je vais élever cela à la puissance $\frac{p}{2}$ (comme $p > 2$, la fonction $x \mapsto x^{p/2}$ est convexe et c'est cela qui va nous aider). Une petite étude préliminaire de $(a + b)^r$ versus $a^r + b^r$ pour $r = \frac{p}{2} > 1$ fixé , et $a, b > 0$ n'est pas superflue. Soit $\phi(x) = \frac{(1+x)^r}{1+x^r}$ pour $0 \leq x \leq 1$. La dérivée logarithmique $(0 < x < 1)$ est $r\frac{1}{1+x} - r\frac{x^{r-1}}{1+x^r} = r\frac{1-x^{r-1}}{(1+x)(1+x^r)} > 0$ donc $1 \leq \phi(x) \leq 2^{r-1}$ pour $0 \leq x \leq 1$ et(*)

$$a^r + b^r \leq (a + b)^r \leq 2^{r-1}(a^r + b^r) \qquad (r > 1; 0 \leq a, b \leq +\infty)$$

On écrit alors :

$$\begin{aligned} |f(x) - g(x)|^p + |f(x) + g(x)|^p &= (|f(x) - g(x)|^2)^{p/2} + (|f(x) + g(x)|^2)^{p/2} \\ &\leq \left(|f(x) - g(x)|^2 + |f(x) + g(x)|^2\right)^{p/2} \\ &= 2^{p/2}\left(|f(x)|^2 + |g(x)|^2\right)^{p/2} \\ &\leq 2^{p-1}\left(|f(x)|^p + |g(x)|^p\right) \end{aligned}$$

(*) vous pouvez retrouver l'inégalité de gauche comme un cas particulier de Minkowski et celle de droite comme un cas particulier de Hölder.

$$|f(x) - g(x)|^p \leq 2^p \left(\frac{|f(x)|^p + |g(x)|^p}{2} - \left| \frac{f(x) + g(x)}{2} \right|^p \right)$$

$$(\|f - g\|_p)^p \leq 2^p \left(\frac{\|f\|_p^p + \|g\|_p^p}{2} - \|\frac{f + g}{2}\|_p^p \right)$$

Et on y prend $f = w_j - u_0$, $g = w_k - u_0$, $d(u_0, W) = \inf_{w \in W} \|u_0 - w\|_p$, $\|u_0 - w_j\| \dashrightarrow d(u_0, W)$, on fait la remarque que $\|\frac{w_j + w_k}{2} - u_0\|_p \geq d(u_0, W)$, et alors :

$$\|w_j - w_k\|_p \leq 2 \left(\frac{\|u_0 - w_j\|_p^p + \|u_0 - w_k\|_p^p}{2} - d(u_0, W)^p \right)^{1/p}$$

et donc (w_j) est de Cauchy. Voilà qui complète la preuve du théorème de dualité pour $p > 2$.

Le cas $1 < p < 2$ est plus difficile, par cette approche qui consiste à vouloir faire dans L^p comme dans L^2, avec le point-clé de la projection. Que cela soit encore possible pour $1 < p < 2$ a été rattaché par Clarkson (<< Uniformly convex spaces >>, Trans. AMS 40 (1936), 396-414) a une notion de << convexité uniforme >> pour L^p $(p > 1)$; cela veut dire qu'il y a des inégalités qui remplacent l'identité du parallélogramme et qui permettent de faire la preuve selon notre schéma général. Grâce à un travail de Hanner (<< On the uniform convexity of L^p and l^p >> ; Arkiv för Matematik, Band 3, nr 19, 1956, 239-244) on connaît l'inégalité optimale (en un certain sens). Cette inégalité sophistiquée, la voici pour $1 \leq p \leq 2$ (pour $p \geq 2$ l'inégalité est dans l'autre sens et n'a pas d'intérêt spécial dans le contexte de la recherche du $w_0 \in W$ minimisant $\|u_0 - w_0\|$) :

$$|\, \|f + g\|_p + \|f - g\|_p \,|^p + |\, \|f + g\|_p - \|f - g\|_p \,|^p \leq 2^p (\|f\|_p^p + \|g\|_p^p) \qquad (1 \leq p \leq 2)$$

Je vous laisse le challenge de l'utiliser pour compléter notre programme : si $(w_j)_{j \geq 1}$ est une suite dans W avec $\|u_0 - w_j\|_p \dashrightarrow d(u_0, W)$ alors elle est de Cauchy, donc convergente car L^p est complet. Sa limite $w_0 = \lim w_j$ est dans W car ce dernier est fermé, et w_0 fait l'affaire car de $w_0 = \lim w_j$ on déduit $\|u_0 - w_0\|_p = \lim \|u_0 - w_j\|_p = d(u_0, W)$ par continuité de $u \mapsto \|u\|_p$. Pour la justification de l'inégalité de Hanner et la notion d'uniforme convexité de Clarkson je vous renvoie à des livres tels que << Analyse fonctionnelle élémentaire >> de M. Willem (éd. Cassini) et << Analysis >> de E. Lieb et M. Loss (éd. AMS).

L'article de Riesz est pour $X = \mathbb{R}$ ou un intervalle ; sa méthode ne distingue pas $1 < p \leq 2$ de $p \geq 2$. Elle semble à première vue assez spécifique à $X = \mathbb{R}$ (voir le Chapitre Deux du livre de Riesz et Sz-Nagy pour une preuve probablement assez proche de celle de Riesz en 1910), mais l'ingéniosité d'homo sapiens est grande, et avec le Théorème de Radon-Nikodym, que je n'énoncerai pas ici, et la notion attenante de mesure absolument continue par rapport à une autre, on a le formalisme qui permet de suivre Riesz pour un espace mesuré X quelconque ; pour être précis disons pour les espaces mesurés σ-finis (c'est-à-dire tels que $X = \cup_{n \geq 1} X_n$ avec $\mu(X_n) < \infty$ pour chaque n). Cela donne une preuve que vous trouverez au Chapitre Six du livre de Rudin << Analyse réelle et complexe >>.

La preuve de Radon-Nikodym chez Rudin suit un procédé de von Neumann qui introduit un espace de Hilbert-Lebesgue ad-hoc (c'est surprenant puisque rien ne faisait référence a priori dans Radon-Nikodym à des fonctions de carrés intégrables). Finalement on est toujours dans une histoire de projection mais elle est faite dans un Hilbert !

L'invocation de Radon-Nikodym fournit une fonction v dont on montre après qu'elle est dans L^q. Cette étape, nous allons la court-circuiter, car nous n'avons plus qu'à faire pour $p < 2$ et comme on sait que cela marche pour $p = 2$ on va avoir un v par ce biais. Je suis en cela Lax au Chapitre Huit de son livre << Functional Analysis >> (éd. Wiley). On suppose $\mu(X) = 1$. Je laisse en exercice l'extension du résultat au cas plus général d'un X qui est σ-fini. Comme $1 < p < 2$ et $\mu(X) = 1$ on a l'inclusion $L^2 \subset L^p$ et les inégalités $\|f\|_p \leq \|f\|_2$. Donc toute forme linéaire continue L sur L^p définit par restriction une forme linéaire continue sur L^2.(*) Par le théorème de Fréchet-Riesz il existe v dans L^2 tel que $L(u) = \int_X u(x)v(x)d\mu(x)$ pour tout u dans L^2. On va montrer que v est dans L^q. Pour cela définissons pour N = 1,2... : $u_N(x) = \min(|v(x)|, N)^{q-1}\frac{\overline{v(x)}}{|v(x)|}$ si $v(x) \neq 0$ et $u_N(x) = 0$ si $v(x) = 0$. Alors $|u_N(x)| \leq N^{q-1}$ est borné donc dans L^2 et :

$$L(u_N) = \int_X \min(|v(x)|, N)^{q-1}|v(x)|d\mu(x) \geq \int_X \min(|v(x)|, N)^q\, d\mu(x)$$

Par ailleurs $|L(u_N)| \leq \|L\| \cdot \|u_N\|_p$. Or :

$$(\|u_N\|_p)^p = \int_X \min(|v(x)|, N)^{pq-p}\, d\mu(x) = \int_X \min(|v(x)|, N)^q\, d\mu(x)$$

Donc

$$\int_X \min(|v(x)|, N)^q\, d\mu(x) \leq |L(u_N)| \leq \|L\| \left(\int_X \min(|v(x)|, N)^q\, d\mu(x)\right)^{\frac{1}{p}}$$

Donc

$$\left(\int_X \min(|v(x)|, N)^q\, d\mu(x)\right)^{\frac{1}{q}} \leq \|L\|$$

On passe à la limite par le théorème de la convergence monotone et on obtient $\|v\|_q \leq \|L\|$. Donc v est bien un élément de L^q. Soit M la forme linéaire sur L^p définie par v. On a L = M sur le sous-espace $L^2 \subset L^p$ et il s'agit d'un sous-espace dense car les fonctions simples (combinaisons linéaires finies de fonctions indicatrices de parties mesurables) sont contenues dans L^2 et sont denses dans L^p (si vous avez lu et tout compris dans toutes mes annexes vous êtes suffisamment calé(e) pour justifier cela sans peine).

Donc la forme linéaire continue L sur L^p est bien de la forme

$$u \mapsto L(u) = \int_X u(x)v(x)\, d\mu(x)\ ,$$

pour un certain $v \in L^q$. Cela complète la preuve du Théorème de Dualité.

(*) dans ce qui suit la notation $\|L\|$ est la norme de L comme forme linéaire sur L^p, pas L^2.

Hellinger-Toeplitz-Landau-Riesz

Théorème : *Soit* X *un espace mesuré* σ*-fini. Soit* g *telle que* fg *est dans* L^1 *pour toute* f *dans* L^p*. Alors* g *est dans* L^q *(*$p \geq 1$*,* $\frac{1}{p} + \frac{1}{q} = 1$*).*

À propos je n'ai pas encore parlé de $L^\infty(X, \mu)$. Ses éléments sont les (classes d'équivalences de) fonctions essentiellement bornées. Une fonction mesurable g est dite essentiellement bornée si il existe $C < \infty$ tel que $\mu(\{x \mid |g(x)| > C\}) = 0$. Si c'est le cas l'ensemble des $C \geq 0$ possibles possède un plus petit élément qui est noté $\|g\|_\infty$. Cela définit une norme sur l'espace vectoriel des classes de telles fonctions, et cet espace est complet. Ces affirmations forment un très bon exercice instructif. Problème plus ambitieux : adaptez la méthode de la fin de l'annexe précédente (pour $1 < p < 2$) et prouvez que pour X sigma-fini le dual de $L^1(X, \mu)$ est $L^\infty(X, \mu)$. Prouvez aussi : *Soit* X *avec* $0 < \mu(X) < \infty$*. Pour toute fonction mesurable* f *sur* X *on a* $\|f\|_\infty = \lim_{p \to \infty} \|f\|_p$*.*

Le dual de $L^\infty(X, \mathcal{B}, \mu)$ est (en général) énorme : c'est l'espace des fonctions $m : \mathcal{B} \to \mathbb{C}$ finiment additives, de variations totales bornées ($\sup_{A\in\mathcal{B}} |m(A)| < \infty$), et absolument continues ($\mu(A) = 0 \Longrightarrow m(A) = 0$). (*cf.* Dunford-Schwartz ou Yosida)

Le théorème est vrai mais trivial pour $p = \infty$, $q = 1$ (prendre $f = 1$). Pour montrer ce qui peut ne pas aller lorsque $1 \leq p < \infty$ sans l'hypothèse << X est σ-fini >>, prenons un ensemble non vide X, prenons comme tribu $\mathcal{B} = \mathcal{P}(X)$ et comme mesure $\mu(A) = +\infty$ sauf pour $A = \emptyset$. Alors $L^p(X, \mu) = \{0\}$, pour $1 \leq p < \infty$ et $L^\infty(X, \mu)$ est l'espace des fonctions bornées. Donc, pour $1 \leq p < \infty$ n'importe quelle fonction g vérifie l'hypothèse du théorème. Pour $p > 1$ seule la fonction nulle obéit à sa conclusion, et pour $p = 1$ seules les fonctions bornées. Donc le théorème est faux pour $p > 1$ et si X est de cardinalité infinie il est aussi faux pour $p = 1$.

Comme cas particulier du théorème, prenons $X = \mathbb{N}$, tribu $\mathcal{P}(\mathbb{N})$, mesure de dénombrement. Il s'agit de prouver que si $\sum_{n\in\mathbb{N}} |x_n y_n| < \infty$ dès que $\sum_n |x_n|^p < \infty$ alors $\sum_n |y_n|^q < \infty$ (pour $p = 1$ lire : alors $\sup_n |y_n| < \infty$). Pour $p = q = 2$ cela fut établi par Hellinger et Toeplitz, et pour $p \geq 1$ (ou peut-être $p > 1$) par Landau.

Je commence par $p = 1$. Soit (y_n) une suite avec $\sup_n |y_n| = \infty$. Il y a une suite extraite avec $|y_{n_k}| > 2^k$. Posons $x_{n_k} = 2^{-k}$ et 0 pour les autres. Alors $\sum_n |x_n| < \infty$ mais $\sum_n |x_n y_n| = \infty$.

Prenons $1 < p < \infty$. Soit $(y_n)_{n\geq 0}$ une suite avec $\sum_n |y_n|^q = \infty$. On va construire (x_n) de sorte que $\sum_n |x_n|^p < \infty$ mais $\sum_n |x_n y_n| = \infty$. On peut toujours remplacer

y_0 par 1 de sorte que $y_0 \neq 0$. On pose :

$$x_n = \frac{|y_n|^{q-1}}{|y_0|^q + \cdots + |y_n|^q}$$

Alors $|x_n y_n| = |y_n|^q/(|y_0|^q + \cdots + |y_n|^q)$ et, compte tenu de $pq - p = q$, $x_n^p = |y_n|^q/(|y_0|^q + \cdots + |y_n|^q)^p$. On est donc ramené au Lemme :

Lemme : *si une série* $\sum_n u_n$ *à termes positifs ou nuls diverge alors la série de terme général* $u_n/(u_0 + \cdots + u_n)$ *diverge et la série de terme général* $u_n/(u_0 + \cdots + u_n)^p$ *converge pour* $p > 1$. *On a supposé* $u_0 > 0$ *pour assurer que les dénominateurs sont tous non nuls.*

Preuve : la somme pour $M \leq n \leq N$ des $u_n/(u_0 + \cdots + u_n)$ est minorée par $\sum_{M \leq n \leq N} u_n / \sum_{0 \leq n \leq N} u_n$. Pour M fixé quelconque, cela tend vers 1 pour $N \to \infty$. Donc $\sum_{n \geq M} u_n/(u_0 + \cdots + u_n) \geq 1$ pour tout M et donc la série de terme général $u_n/(u_0 + \cdots + u_n)$ diverge. Par contre :

$$n \geq 1 \Longrightarrow \quad \frac{u_n}{(u_0 + \cdots + u_n)^p} \leq \int_{u_0+\cdots+u_{n-1}}^{u_0+\cdots+u_n} \frac{dt}{t^p}$$

Donc $\sum_{n \geq 1} u_n/(u_0 + \cdots + u_n)^p \leq \int_{u_0}^{\infty} t^{-p}\, dt < \infty$.

Cela termine la preuve de Hellinger-Toeplitz-Landau. Je passe maintenant à la version plus générale de Riesz avec les intégrales sur un espace mesuré sigma-fini (j'imagine que Riesz a, au moins initialement, traité uniquement $\mathbb{R}$ ou un intervalle avec la mesure de Lebesgue.) On va commencer par un Lemme :

Lemme : *soit* X σ*-fini et* $p > 1$. *Si toute fonction* L^p *est* L^1 *alors* $\mu(X) < \infty$.

Preuve : supposons $\mu(X) = +\infty$. Comme X est sigma-fini on peut l'écrire comme une union disjointe de X_n avec $\mu(X_n) = u_n$, $0 < u_n < \infty$ et $\sum_n u_n = \infty$. Considérons la fonction qui vaut $(u_0 + u_1 + \cdots + u_n)^{-1}$ sur X_n. Par le Lemme précédent cette fonction est L^p et n'est pas L^1.

Maintenant soit g (partout finie) avec la propriété que gf est L^1 pour toute fonction f qui est L^p ($p > 1$ fixé). Considérons la fonction k avec $k(x) = 2^j$ lorsque $2^j \leq |g(x)| < 2^{j+1}$ ($j \in \mathbb{Z}$) et $k(x) = 0$ lorsque $g(x) = 0$. Alors $k \leq |g|$ donc k a la même propriété que g. De plus $|g| \leq 2k$, donc il suffit de montrer que k est L^q. Soit $A_j = \{x \mid k(x) = 2^j\}$. Comme X est sigma-fini, A_j aussi. Supposons $A_j \neq \emptyset$. Toute fonction sur A_j s'étend par zéro à X. Toute f sur A_j qui est L^p est telle que fk, c'est-à-dire $2^j f$ est L^1. Par le Lemme, on en déduit $\mu(A_j) < \infty$. Si il n'y a qu'un nombre fini de j avec $\mu(A_j) > 0$, $k \in L^q$ est immédiat. Sinon, on renumérote $B_0 = A_{j_0}$, etc..., ceux de mesure non nulle ($\mathbb{Z}$ est dénombrable !). Disons ensuite que B_n est de mesure $\mu_n > 0$ et que k prend sur B_n la valeur k_n. Alors par notre hypothèse dès que $\sum_{n \geq 0} |f_n|^p \mu_n < \infty$ on a $\sum_{n \geq 0} |f_n| k_n \mu_n < \infty$. Posons $x_n = f_n(\mu_n)^{1/p}$, $y_n = k_n(\mu_n)^{1/q}$. Alors dès que $\sum_{n \geq 0} |x_n|^p < \infty$ on a $\sum_{n \geq 0} |x_n y_n| < \infty$. Par ce qui précède on a $\sum_{n \geq 0} y_n^q < \infty$, donc on obtient bien $\int_X k(x)^q\, d\mu(x) < \infty$. Ceci termine la preuve pour $p > 1$. Il n'y a plus la place pour $p = 1$ et je le laisse en exercice.

Points de Lebesgue

Version du 15-04-2006, 15h00. J'avais expliqué que ce cours était aussi une occasion d'approfondir nos connaissances en théorie de l'intégrale de Lebesgue. Le plus central est que les (classes d'équivalence pour l'égalité presque partout des) fonctions intégrables au sens de Lebesgue, ou de carrés intégrables, forment des espaces vectoriels normés complets, ce que l'on appelle en général le << théorème de Riesz-Fischer >>. Mais il n'y a pas que cela. En particulier, Lebesgue a montré que pour une fonction intégrable f, qui est 2π-périodique, on a lim Σ_N(f)(x) = f(x) pour presque tout x (comme d'habitude Σ_N(f) ou σ_N(f) est la Nième somme de Fejér). C'est l'objet de l'annexe suivante. Ici je vais prouver un autre théorème de Lebesgue : lorsque f est intégrable, alors $F(x) = \int_a^x f(t)\,dt$ est presque partout dérivable avec $F'(x) = f(x)$. Le point commun entre les deux théorèmes c'est que si x est un << point de Lebesgue >> de f alors il est << ok >> pour les deux théorèmes. Et donc, le point (sic) clé c'est que presque tout point est un point de Lebesgue.

On dira que x est un point de Lebesgue de f si(*)

$$\lim_{h\to 0^+} \frac{1}{2h}\int_{[x-h,x+h]} |f(t)-f(x)|\,dt = 0$$

C'est facile de montrer que si x_0 est un point de Lebesgue de f alors $F(x) = \int_a^x f(t)\,dt$ est dérivable au point x_0 avec $F'(x_0) = f(x_0)$. Exercice pour vous. Pour les fonctions 2π-périodiques, c'est moins facile, mais vrai, que si x_0 est un point de Lebesgue de f alors lim $\Sigma_N(f)(x_0) = f(x_0)$. Voir annexe suivante. Bref, ici j'ai une fonction intégrable f sur un intervalle $[a,b]$ et je vais prouver que presque tout point est un point de Lebesgue de f. Quand faut y aller faut y aller. Attention, à partir d'ici ce n'est pas du pipeau.

Donc soit f intégrable sur $[a,b]$. On étend f par 0 à $\mathbb{R}$ tout entier. Pour tout x soit $M_f(x) = \sup_{h>0} \frac{1}{2h}\int_{[x-h,x+h]} |f(t)|\,dt$. Pour certains x on peut avoir $M_f(x) = +\infty$. En tout cas $M_f(x) = \mathcal{O}(|x|^{-1})$ pour $|x|\to\infty$. Pour tout $C > 0$ l'ensemble $A = \{x \mid M_f(x) > C\}$ est donc borné. De plus A est ouvert. Soit K un compact inclus dans A. On va prouver $|K| \leq \frac{3}{C}\int_a^b |f(t)|dt$. Une fois cela fait en écrivant A comme une union d'intervalles ouverts, etc, on aura donc $|A| \leq \frac{3}{C}\int_a^b |f(t)|dt$. On appelle ce genre de choses une << inégalité maximale >>. À cause du sup dans la définition de M_f sans doute, et donc du fait que l'on appelle M_f la

(*) si on modifie f dans sa classe d'équivalence on modifie l'ensemble des points de Lebesgue par un ensemble de mesure nulle.

<< fonction maximale >> associée à f. J'en reviens au compact K inclus dans $A = \{x \mid M_f(x) > C\}$. Notons $D(x,h) =]x-h, x+h[$. Un nombre fini de tels $D_j = D(x_j, h_j)$, $1 \leq j \leq N$ recouvrent notre compact K, avec, pour chaque j, $\int_{D_j} |f(t)|\,dt \geq C|D_j|$. On les a numérotés de sorte que $h_1 \geq h_2 \geq \ldots$. Posons $j_1 = 1$ et $E_1 = D_1$. On élimine tous les autres qui l'intersectent. Soit j_2 l'indice du premier D_j qui reste (si il y en a). Et soit $E_2 = D_{j_2}$. On élimine tous ceux qui l'intersectent. Soit j_3 l'indice du premier qui reste (si il y en a). Soit $E_3 = D_{j_3}$. Etc. ... On a donc E_1, E_2, ..., E_M, $M \leq N$, qui sont deux-à-deux disjoints. Soient E_1^*, E_2^*, etc..., les intervalles ouverts de mêmes centres, mais trois fois plus longs. Soit j un indice autre que j_1, j_2, ... Alors D_j intersecte l'un des E_k avec $j_k < j$. Donc D_j est inclus entièrement dans l'un des E_k^*. Au final K est inclus dans l'union des E_k^*, donc $|K|$ est majoré par $\sum_k |E_k^*| = 3\sum_k |E_k|$. Par ailleurs les intervalles ouverts E_k sont deux à deux disjoints donc $\sum_k \int_{E_k} |f(t)|\,dt \leq \int_{\mathbb{R}} |f(t)|\,dt$. Rappelons que pour chaque D_j on a $\int_{D_j} |f(t)|\,dt \geq C|D_j|$ donc pour chaque E_k on a $\int_{E_k} |f(t)|\,dt \geq C|E_k|$. Ainsi :

$$|K| \leq 3\sum_k |E_k| \leq \frac{3}{C}\sum_k \int_{E_k} |f(t)|\,dt \leq \frac{3}{C}\int_{\mathbb{R}} |f(t)|dt$$

L'inégalité maximale est prouvée.

Maintenant, pour tout x, posons $\omega_f(x) = \limsup_{h \to 0} \frac{1}{2h}\int_{[x-h,x+h]} |f(t) - f(x)|\,dt$ (il est possible d'avoir $\omega_f(x) = +\infty$). Si f est continue au point x alors $\omega_f(x) = 0$. Si g est une autre fonction intégrable on a $|f(t) - f(x)| \leq |f(t) - g(t) - (f(x) - g(x))| + |g(t) - g(x)|$ donc $\omega_f \leq \omega_{f-g} + \omega_g$. En particulier si g est continue en tout point alors $\omega_f \leq \omega_{f-g} \leq \omega_{f-g-(-g)} = \omega_f$, donc $\omega_f = \omega_{f-g}$.

On prend note de $\omega_f(x) \leq M_f(x) + |f(x)|$. Soit $\eta > 0$ et considérons $A_\eta = \{x \mid \omega_f(x) > \eta\}$. On a l'inclusion $A_\eta \subset B_1 \cup B_2$ avec $B_1 = \{M_f(x) > \frac{\eta}{2}\}$ et $B_2 = \{|f(x)| > \frac{\eta}{2}\}$. On a $\frac{\eta}{2}|B_2| \leq \int_{\mathbb{R}} |f(t)|\,dt$ donc $|B_2| \leq \frac{2}{\eta}\|f\|_1$. Par l'inégalité maximale on a $|B_1| \leq \frac{6}{\eta}\|f\|_1$. Donc $|A_\eta| \leq \frac{8}{\eta}\|f\|_1$. Remplaçons f par $f - g$ avec g continue à support compact. On a vu que les deux fonctions ω_f et ω_{f-g} étaient identiques. Donc ce remplacement ne modifie pas A_η. Donc en fait $|A_\eta| \leq \frac{8}{\eta}\|f - g\|_1$. Mais nous savons que nous pouvons trouver des fonctions continues g, au support borné éventuellement un peu plus grand que $[a,b]$ et rendant $\int_{\mathbb{R}} |f(t) - g(t)|\,dt$ arbitrairement petit. Donc l'ensemble des x avec $\omega_f(x) > \eta$ est de mesure nulle. Comme $\eta > 0$ est arbitraire, l'ensemble des x avec $\omega_f(x) > 0$ est de mesure nulle. Donc pour presque tout x on a $\omega_f(x) = 0$ ce qui signifie exactement que x est un point de Lebesgue de f.

Exercice : montrer que x est un point de Lebesgue de f si et seulement si pour tout $\lambda \in \mathbb{C}$ on a $\lim_{h \to 0^+} \frac{1}{2h}\int_{x-h}^{x+h} |f(t) - \lambda|\,dt = |f(x) - \lambda|$.

Exercice : justifier la mesurabilité des ensembles A_η dans la preuve ci-dessus (à vue de nez, ma conclusion est qu'il faut y réfléchir ; je vous délègue ce devoir de réfléchir, car j'ai une conception très démocratique des choses).

Fejér presque partout

Kolmogorov a construit en 1926 une fonction f intégrable 2π-périodique dont la série de Fourier est partout divergente. Cela a fortement impressionné les esprits et pendant longtemps on n'a pas su quoi penser, sous des hypothèses plus fortes, comme la continuité de f. La réponse vint en 1966, lorsque Carleson prouva que pour toute f de carré intégrable (donc, en particulier, pour toute fonction continue), la série de Fourier est presque partout convergente (déjà au dix-neuvième siècle, Du Bois-Reymond avait donné un exemple de fonction continue telle que la série de Fourier diverge en au moins un point ; prouver la convergence partout était donc exclu d'emblée). Longtemps avant le résultat de Carleson, Lebesgue avait montré un résultat de convergence sous la seule hypothèse $f \in L^1(0,2\pi;\frac{dx}{2\pi})$: pour presque tout x on a $\lim \Sigma_N(f)(x) = f(x)$. Je rappelle que $\Sigma_N(f)(x)$ est la $N^{\text{ième}}$ somme de Fejér. Donc pour l'exemple de Kolmogorov, la série de Fourier est partout divergente, mais sa moyenne de Cesàro est presque partout convergente.

Je vais prouver ici $\lim \Sigma_N(f)(x_0) = f(x_0)$ en tout point de Lebesgue de la fonction intégrable 2π-périodique f (et par l'annexe précédente presque tout point est un point de Lebesgue). Quitte à remplacer $f(x)$ par $f(x - x_0)$ on peut supposer $x_0 = 0$. Ensuite on a la formule :

$$\Sigma_N(f)(0) - f(0) = \frac{1}{\pi}\int_0^\pi \left(\frac{f(t) + f(-t)}{2} - f(0)\right) F_N(t)\,dt$$

Donc, en posant $g(t) = \left|\frac{f(t)+f(-t)}{2} - f(0)\right|$ et en rappelant $F_N(t) \geq 0$ on a :

$$|\Sigma_N(f)(0) - f(0)| \leq \frac{1}{\pi}\int_0^\pi g(t)F_N(t)\,dt = J_N$$

Notre objectif est de prouver $\lim J_N = 0$. Notre hypothèse est que 0 est un point de Lebesgue de f, et donc, en posant

$$G(u) = \frac{1}{u}\int_0^u g(t)\,dt$$

on a $G(0^+) = 0$. Sur $]0,\pi]$ G est continue et on posera $\sup_{0<u\leq\pi} G(u) = M < \infty$.

L'idée cruciale est de majorer $F_N(t) = \frac{1}{N}\frac{\sin^2(\frac{Nt}{2})}{\sin^2(\frac{t}{2})}$ par un noyau $K_N(t)$ décroissant sur $[0,\pi]$. Je propose $K_N(t) = N$ pour $0 \leq t \leq \frac{\pi}{N}$ et $K_N(t) = \frac{1}{N}\frac{\pi^2}{t^2}$ pour $\frac{\pi}{N} \leq t \leq \pi$. Vous vérifierez que $0 \leq F_N(t) \leq K_N(t)$, et que K_N est continue et décroissante. Pour future référence, je signale $\int_0^\pi K_N(t)\,dt \leq 2\pi$ comme il résulte d'un calcul très simple. En posant $k_N(t) = 0$ pour $0 \leq t < \frac{\pi}{N}$ et $k_N(t) = \frac{2}{N}\frac{\pi^2}{t^3}$ pour $\frac{\pi}{N} \leq t \leq \pi$ on a $K_N(t) = \int_t^\pi k_N(u)\, du + \frac{1}{N}$. Le point important c'est que $k_N \geq 0$.

Par Fubini, on a :

$$J_N \leq \frac{1}{\pi}\int_0^\pi g(t)K_N(t)\, dt = \frac{1}{\pi N}\int_0^\pi g(t)\, dt + \frac{1}{\pi}\iint_{0\leq t\leq u\leq\pi} g(t)k_N(u)\, dtdu$$
$$= \frac{1}{\pi N}\int_0^\pi g(t)\, dt + \frac{1}{\pi}\int_0^\pi uG(u)\, k_N(u)\, du$$

Il s'agit donc de montrer $\lim \int_0^\pi uG(u)\, k_N(u)\, du = 0$. On remarque en prenant $g \equiv 1$, $\frac{1}{\pi}\int_0^\pi K_N(t)\, dt = \frac{1}{N} + \frac{1}{\pi}\int_0^\pi u\, k_N(u)\, du$, donc $\frac{1}{\pi}\int_0^\pi u\, k_N(u)\, du \leq 2$. On fait la découpe habituelle, pour chaque $\delta \in]0, \pi[$ (et en rappelant $M = \sup_{0<u\leq\pi} G(u)$) :

$$\frac{1}{\pi}\int_0^\pi uG(u)\, k_N(u)\, du = \frac{1}{\pi}\int_0^\delta uG(u)\, k_N(u)\, du + \frac{1}{\pi}\int_\delta^\pi uG(u)\, k_N(u)\, du$$
$$\leq 2 \sup_{0<u\leq\delta} G(u) + M\,\frac{1}{\pi}\int_\delta^\pi u\, k_N(u)\, du$$

Pour $N \geq \frac{\pi}{\delta}$ on a $k_N(u) = \frac{2}{N}\frac{\pi^2}{u^3}$ sur $[\delta, \pi]$ et donc $\int_\delta^\pi u\, k_N(u)\, du \leq \frac{2}{N}\frac{\pi^2}{\delta}$. On peut donc écrire pour chaque $\delta > 0$:

$$\limsup \frac{1}{\pi}\int_0^\pi uG(u)\, k_N(u)\, du \leq 2 \sup_{0<u\leq\delta} G(u)$$

Comme $\lim_{u \to 0^+} G(u) = 0$, on en déduit alors en faisant tendre δ vers 0 :

$$\limsup \frac{1}{\pi}\int_0^\pi uG(u)\, k_N(u)\, du = 0$$

Mesures Complexes

Soit X un ensemble muni d'une tribu de sous-ensembles. Une *mesure complexe* μ est une fonction sur cette tribu, à valeurs dans $\mathbb{C}$, qui vérifie l'axiome d'additivité dénombrable : pour toute partie mesurable $A \subset X$, et pour toute partition dénombrable $A = \cup_n A_n$, $0 \leq n < N$ $(1 \leq N \leq +\infty)$ on a :

$$\mu(A) = \sum_{0 \leq n < N} \mu(A_n)$$

Lorsque $\mu(A) \in \mathbb{R}$ pour tout A on dit que μ est une *mesure signée*. À noter : une mesure positive ne peut être considérée comme une mesure signée (ou complexe) que si elle est finie ($\mu(X) < +\infty$). Les parties réelle et imaginaire d'une mesure complexe sont des mesures signées. Comme exemple (instructif) de mesure signée ou complexe, on peut prendre $X = [0,1]$ avec la tribu de Borel ou de Lebesgue et $\mu(A) = \int_A f(x)\,dx$ pour une certaine fonction $f \in L^1(0,1;dx)$. On déduit de l'axiome d'additivité les relations habituelles $\mu(\emptyset) = 0$(*), $\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$, et $\mu(X \setminus A) = \mu(X) - \mu(A)$.

De plus, si les A_n forment une partition infinie de A, on peut les permuter arbitrairement et donc la série $\sum_{n=0}^{\infty} \mu(A_n)$ converge même après un réordonnancement quelconque des indices. Par un théorème connu (...j'espère) cela n'est possible que si elle est absolument convergente. On peut donc considérer la quantité finie $\sum_{0 \leq n < N} |\mu(A_n)|$, puis prendre la borne supérieure $\nu(A)$ de toutes ces quantités pour toutes les partitions finies ou infinies dénombrables de A. Nous allons montrer que ν est une mesure positive ; puis dans un deuxième temps nous montrerons qu'elle est finie. On l'appelle << mesure des variations de μ >> et on la note $\nu = |\mu|$. On a $|\mu(A)| \leq |\mu|(A)$ pour tout A de la tribu, mais l'inégalité peut être stricte (réfléchir à l'exemple $d\mu = f(x)dx$; à votre avis qu'est-ce que $|\mu|$?). À propos, évidemment, on ne dira pas que A est μ-négligeable si elle vérifie $\mu(A) = 0$, mais seulement sous l'hypothèse plus forte $\mu(B) = 0$ pour tout $B \subset A$, ce qui d'ailleurs équivaut à $|\mu|(A) = 0$.

Montrons que ν est effectivement une mesure positive. Notons que $\nu(\emptyset) = 0$. Supposons que l'on se donne une partition dénombrable (finie ou infinie) $A = \bigcup_\lambda B_\lambda$. On veut prouver $\nu(A) = \sum_\lambda \nu(B_\lambda)$. Donnons nous une autre par-

(*) car $\mu(\emptyset) = \mu(\emptyset) + \mu(\emptyset)$; pour les mesures positives on doit rajouter explicitement l'axiome $\mu(\emptyset) = 0$. Comme pour une mesure complexe $\mu(\emptyset) \neq \infty$ a priori, on n'a pas à postuler $\mu(\emptyset) = 0$.

tition $A = \bigcup_j A_j$ de A(**). Pour chaque j on a $\mu(A_j) = \sum_\lambda \mu(A_j \cap B_\lambda)$ donc $|\mu(A_j)| \leq \sum_\lambda |\mu(A_j \cap B_\lambda)|$ donc

$$\sum_j |\mu(A_j)| \leq \sum_j \sum_\lambda |\mu(A_j \cap B_\lambda)| = \sum_\lambda \sum_j |\mu(A_j \cap B_\lambda)| \leq \sum_\lambda \nu(B_\lambda)\ ,$$

l'égalité du milieu car nous manipulons des séries à termes positifs, l'inégalité de droite par définition des $\nu(B_\lambda)$. En prenant le sup sur toutes les partitions $A = \bigcup A_j$ on obtient $\nu(A) \leq \sum_\lambda \nu(B_\lambda)$.

Pour montrer l'inégalité dans l'autre sens, nous observons tout d'abord que si $\nu(B_\lambda) = +\infty$ pour au moins un λ alors en adjoignant $A \setminus B_\lambda$ à une partition de B_λ, on obtient une partition de A et donc on obtient des partitions de A rendant les sommes des modules de μ arbitrairement grandes, donc $\nu(A) = +\infty$, et donc l'identité $\nu(A) = \sum_\lambda \nu(B_\lambda)$ vaut. On supposera donc que pour tout λ, on a $\nu(B_\lambda) < \infty$. On se donne un $\epsilon > 0$, on suppose que l'ensemble des indices λ est $\mathbb{N}$ ou $\{0, 1, \ldots, N-1\}$, on se donne pour chaque B_λ une partition dénombrable $B_\lambda = \bigcup_j B_{\lambda,j}$ avec $\sum_j |\mu(B_{\lambda,j})| \geq \nu(B_\lambda) - \frac{\epsilon}{2^\lambda}$. Prises toutes ensemble ces partitions donnent une partition (dénombrable car $\mathbb{N} \times \mathbb{N}$ est dénombrable) de A : $A = \bigcup_{\lambda,j} B_{\lambda,j}$ et la somme associée $\sum_{\lambda,j} |\mu(B_{\lambda,j})|$ est $\geq \sum_\lambda (\nu(B_\lambda) - \frac{\epsilon}{2^\lambda}) \geq (\sum_\lambda \nu(B_\lambda)) - 2\epsilon$. Donc $\nu(A) \geq \sum_\lambda \nu(B_\lambda) - 2\epsilon$ et comme $\epsilon > 0$ est arbitraire $\nu(A) \geq \sum_\lambda \nu(B_\lambda)$. Finalement on a bien prouvé

$$\nu(A) = \sum_\lambda \nu(B_\lambda)$$

pour toute partition dénombrable (finie ou infinie) $A = \bigcup_\lambda B_\lambda$ d'une partie mesurable $A \subset X$. Donc $\nu = |\mu|$ est bien une mesure positive.

On a évidemment par la définition $|\mu|(X) \leq |\mathrm{Re}(\mu)|(X) + |\mathrm{Im}(\mu)|(X)$ donc pour montrer que $|\mu|$ est finie il suffira de considérer le cas des mesures signées.

Mesures Signées

Supposons donc que μ soit à valeurs réelles. Si A est une partie mesurable, qui a été partitionnée en des A_n, on peut faire l'union des A_n avec $\mu(A_n) \geq 0$, ce qui donne un certain $B \subset A$. Alors $\sum |\mu(A_n)| = \mu(B) + |\mu(C)|$ avec $C = A \setminus B$. Comme $\mu(A) = \mu(B) + \mu(C)$ on a $|\mu(C)| \leq \mu(B) + |\mu(A)|$, et donc $2\mu(B) \geq -|\mu(A)| + \sum |\mu(A_n)|$. Donc si $\nu(A) = +\infty$ alors on peut trouver $B \subset A$ avec $\mu(B)$ arbitrairement grand, disons $\mu(B) \geq 2|\mu(A)| + 1$. Mais de $\mu(A) = \mu(B) + \mu(C)$ on déduit $|\mu(C)| \geq \mu(B) - |\mu(A)|$ donc $|\mu(C)| \geq |\mu(A)| + 1$. Et comme ν est une mesure on a $\nu(B) = +\infty$ ou $\nu(C) = +\infty$.

On peut donc par récurrence, si l'on suppose $\nu(X) = +\infty$, affirmer l'existence de $A_0 = X$, $A_1 \subset A_0$, ..., $A_{n+1} \subset A_n$,...avec pour tout n, $|\mu(A_{n+1})| \geq |\mu(A_n)| + 1$ (et $\nu(A_{n+1}) = +\infty$). Mais alors $|\mu(A_n \setminus A_{n+1})| \geq 1$ et la série $\sum_n \mu(A_n \setminus A_{n+1})$ ne

(**) en utilisant l'ensemble vide on peut toujours remplacer une partition finie par une partition infinie donc on peut supposer ici que les j parcourent $\mathbb{N}$.

peut pas être convergente ce qui contredit l'axiome d'additivité dénombrable (pour une réunion disjointe). C'est donc que l'hypothèse $\nu(X) = +\infty$ est à rejeter. On a ainsi $\nu(X) < +\infty$. Cette valeur finie s'appelle la *variation totale de μ sur* X.

Posons $\mu_1 = \frac{1}{2}(|\mu| + \mu)$ et $\mu_2 = \frac{1}{2}(|\mu| - \mu)$. Ce sont clairement deux mesures positives finies. Et $\mu = \mu_1 - \mu_2$. Toute mesure signée est donc la différence de deux mesures positives finies. Réciproquement soit σ_1 une mesure positive finie vérifiant $\sigma_1(A) \geq \mu(A)$ pour tout A. Il est immédiat que $\sigma_2(A) = \sigma_1(A) - \mu(A)$ est une mesure positive (finie). Comme $\mu = \sigma_1 - \sigma_2$, on a clairement (?) $|\mu| \leq \sigma_1 + \sigma_2$, donc en sommant $2\mu_1 \leq 2\sigma_1$, et en soustrayant, $2\mu_2 \leq 2\sigma_2$. Cela montre que μ_1 est caractérisée comme étant << LA plus petite >> mesure positive σ_1 vérifiant $\sigma_1 \geq \mu$. Et μ_2 est LA plus petite mesure positive σ_2 vérifiant $\sigma_2 \geq -\mu$. Bien sûr $|\mu|$ est LA plus petite mesure positive minorée par $\max(\mu, -\mu)$, c'est-à-dire telle que pour tout A, $|\mu|(A) \geq |\mu(A)|$.

Comme μ peut s'écrire comme la différence de deux mesures positives, on peut définir $\int_X f(x)d\mu(x)$ de manière évidente : $= \int_X f(x)d\sigma_1(x) - \int_X f(x)d\sigma_2(x)$ et petit exercice si $\sigma_1 - \sigma_2 = \sigma_3 - \sigma_4$ alors $\sigma_1 + \sigma_4 = \sigma_2 + \sigma_3$ donc . . . indépendant des choix, ou on prend directement les canoniques μ_1 et μ_2. Le choix canonique est utile pour voir que $|\int_X f(x)d\mu(x)| \leq |\int_X f(x)d\mu_1(x)| + |\int_X f(x)d\mu_2(x)| \leq \int_X |f(x)|d\mu_1(x) + \int_X |f(x)|d\mu_2(x) = \int_X |f(x)|d|\mu|(x)$.

Hahn-Jordan

Notons à nouveau $\nu = |\mu|$ et considérons l'espace de Hilbert $L^2(X, \nu)$. Alors $f \mapsto L(f) = \int_X f(x)d\mu(x)$ est une forme linéaire continue sur $L^2(X, \nu)$, par l'inégalité que nous venons d'établir (et Cauchy-Schwarz ; rappelons que $\nu(X) < \infty$). Par le théorème de Fréchet-Riesz, il existe une fonction de carré intégrable g telle que pour toute $f \in L^2(X, \nu)$ on a $\int_X f(x)d\mu(x) = \int_X f(x)g(x)d\nu(x)$. En particulier $\mu(A) = \int_A g(x)d\nu(x)$ est réel pour tout A, donc $\int_A \mathrm{Im}(g(x))d\nu(x) = 0$ pour tout A, donc $\mathrm{Im}(g) = 0$ p.p., et on peut donc prendre g à valeurs réelles.

Soit $c > 1$ (par exemple $c = 1 + \frac{1}{n}$) et soit $A = \{x \mid g(x) \geq c\}$. Alors $\mu(A) = \int_X \mathbb{1}_A(x)g(x)d\nu(x) \geq c\nu(A)$, mais $|\mu(A)| \leq \nu(A)$. Donc en fait $\nu(A) = 0$. Par union dénombrable le lieu des x avec $g(x) > 1$ est de ν-mesure nulle. De même pour < -1. Soit maintenant $0 < a < 1$ (par exemple $a = 1 - \frac{1}{n}$) et soit $A = \{x \mid |g(x)| \leq a\}$. Alors pour toute partie $B \subset A$ on a $|\mu(B)| \leq a\nu(B)$. Donc si on fait une partition dénombrable $\cup A_n$ de A on a ensuite $\sum |\mu(A_n)| \leq a \sum_n \nu(A_n) = a\nu(A)$. Donc en prenant le sup, $\nu(A) \leq a\nu(A)$. Ici encore on trouve $\nu(A) = 0$. Finalement par union dénombrable le lieu des x avec $|g(x)| < 1$ est de ν-mesure nulle. Donc quitte à modifier g on pourra supposer que g ne prend que les deux valeurs +1 et -1.

Notons $X_1 = \{x \mid g(x) = 1\}$ et $X_2 = X \setminus X_1$. Alors pour toute partie mesurable on a $\mu(A) = \int_X \mathbb{1}_A(x)g(x)d\nu(x) = \nu(A \cap X_1) - \nu(A \cap X_2)$. Dans le même temps on a $\nu(A) = \nu(A \cap X_1) + \nu(A \cap X_2)$. Donc $\mu_1(A) = \nu(A \cap X_1)$ et $\mu_2(A) = \nu(A \cap X_2)$. Donc si $A \subset X_1$ on a $\mu(A) = \mu_1(A) = \nu(A)$ et si $A \subset X_2$ on a $\mu(A) = -\mu_2(A) = -\nu(A)$. Autrement dit on a trouvé une décomposition de X en deux sous-ensembles disjoints tels que la restriction de μ au premier est une mesure positive (finie) et la restriction de μ au second est l'opposé d'une mesure positive (finie). Une telle décomposition est dite décomposition de Hahn-Jordan de la mesure signée μ. Par exemple si $d\mu = f(x)dx$, cela revient à décomposer l'intervalle $[0,1]$ en $\{x \mid f(x) \geq 0\} \cup \{x \mid f(x) < 0\}$.

Si l'on a une autre décomposition $X = X_1' \cup X_2'$ de type Hahn-Jordan alors pour tout $A \subset X_1 \cap X_2'$ on a $\mu(A) = 0$ (car à la fois positif et négatif) donc $X_1 \cap X_2'$ est μ-négligeable. De même $X_2 \cap X_1'$ est μ-négligeable. Donc la différence symétrique $X_1 \Delta X_1' = (X_1 \cap X_2') \cup (X_2 \cap X_1')$ est μ-négligeable, et aussi $X_2 \Delta X_2'$ (qui est la même chose...). La décomposition de Hahn-Jordan est donc unique à la modification près de X_1 et X_2 par des négligeables.

Exercice : soit λ une mesure positive finie ou infinie et f une fonction à valeurs réelles dans $L^1(X,\lambda)$. En posant $\mu(A) = \int_X \mathbb{1}_A(x)f(x)d\lambda(x)$ on définit une mesure signée. Montrer que $\nu = |\mu|$ est donnée par $d\nu = |f|d\lambda$. C'est très facile en considérant une construction directe d'une décomposition de Hahn-Jordan et en invoquant l'unicité de μ_1 et de μ_2.

Mesures complexes, encore

Reprenons le cas d'une mesure complexe μ. On peut l'écrire $\mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ avec quatre mesures positives finies, et donc donner un sens à $\int_X f(x)\, d\mu(x) =$ (compléter) en montrant (exercice) que cela ne dépend pas du choix des μ_i (à partir du moment où toutes les intégrales considérées sont absolument convergentes ; prendre d'abord f à valeurs réelles). Si on prend les μ_i définies par les décompositions canoniques de $\mathrm{Re}(\mu)$ et $\mathrm{Im}(\mu)$ alors certainement chacune vérifie $\mu_i \leq \nu = |\mu|$ (car $|\mathrm{Re}(\mu)| \leq |\mu|$, ..., justifier), donc on peut définir ainsi $\int_X f(x)\, d\mu(x)$ pour toute $f \in L^1(X,\nu)$.

Supposons que f soit une fonction simple $f = \sum c_j \mathbb{1}_{A_j}$, avec les A_j deux-à-deux disjoints. Alors $\int_X f d\mu = \sum c_j \mu(A_j)$, donc $|\int_X f d\mu| \leq \sum |c_j||\mu(A_j)| \leq \sum |c_j||\mu|(A_j) = \int_X |f| d|\mu|$. Dans le cas général de $f \in L^1(X,\nu)$ il existe des fonctions simples f_n qui convergent vers f dans $L^1(X,\nu)$, donc aussi dans $L^1(X,\mu_i)$ pour chacune des μ_i puisque $\mu_i \leq |\mu|$. Donc $\int_X f d\mu = \lim \int_X f_n d\mu$. Alors $|\int_X f d\mu| \leq \liminf \int_X |f_n| d|\mu| = \lim \int_X |f_n| d|\mu| = \int_X |f| d|\mu|$ car on a aussi $|f_n| \dashrightarrow |f|$ dans $L^1(X,\nu)$.

On peut alors à nouveau considérer $f \mapsto \int_X f(x)d\mu(x)$ comme une fonctionnelle linéaire continue sur l'espace de Hilbert associé à la mesure des variations $\nu = |\mu|$. Il existe donc une fonction g à valeurs complexes telle que $\mu = g\nu$. Je vous délègue la tâche de montrer que $|g| = 1$ ν-presque partout. C'est dans le même style en un peu plus difficile que dans le cas réel. Par ailleurs vous montrerez que si λ est une mesure positive (finie ou non) et si $F \in L^1(X,\lambda)$, alors (avec des notations évidentes) $F\lambda$ est une mesure complexe dont la mesure des variations est la mesure finie $|F|\lambda$.

Variation Bornée et Dirichlet-Jordan

Les hypothèses du théorème de Dirichlet sur la convergence de la série de Fourier d'une fonction (à valeurs réelles) 2π-périodique f sont que celle-ci est continue par morceaux et ne possède sur $[0, 2\pi]$ qu'un nombre fini de maxima ou de minima (locaux, et stricts). Bref, elle est monotone-continue par morceaux. Mais cette condition n'est pas stable par combinaison linéaire : par exemple une fonction C^1 peut sûrement avoir une infinité de (stricts) maxima et minima locaux ($f(x) = x^3 \sin(\frac{1}{x})$) et cependant elle est toujours la somme d'une fonction continue croissante ($\int^x \max(f'(t), 0)\,dt$) et d'une fonction continue décroissante ($\int^x \min(f'(t), 0)\,dt$). La notion de << fonction de variation bornée >>, sur un intervalle $[a, b]$, fut introduite par Camille Jordan ; elle englobe à la fois les conditions de Dirichlet et les fonctions de classe C^1 par morceaux (que l'on associe en général dans les livres de premier cycle avec le théorème de Dirichlet). De plus c'est une notion extrêmement naturelle en théorie de l'intégration puisque comme nous le verrons elle caractérise les fonctions qui (à peu de choses près) sont de la forme $x \longmapsto \int_{[a,x]} d\mu(u)$ avec μ une *mesure complexe* sur $[a, b]$ au sens de l'annexe précédente.

Soit f une fonction (à valeurs complexes) sur $[a, b]$. À tout choix d'au moins deux points $a \leq x_0 \leq x_1 \leq \cdots \leq x_N \leq b$ on peut associer $V((x_j); f) = |f(x_0) - f(x_1)| + \cdots + |f(x_{N-1}) - f(x_N)|$. Et on définit $V([a, b]; f)$ comme le supremum de toutes ces sommes finies : c'est la << variation totale >> de f sur l'intervalle $[a, b]$ (si $b = a$, posons $V([a, a]; f) = 0$; si $b > a$ on peut toujours imposer $x_i < x_{i+1}$ pour tout i dans les subdivisions). On dit que f est de variation bornée sur $[a, b]$ si $V([a, b]; f) < \infty$. Des définitions analogues imposant $x_0 > a$ ou $x_N < b$ mènent à $V(]a, b]; f)$ ou $V([a, b[; f)$ ou $V(]a, b[; f)$.

Il est évident que pour $a \leq b \leq c$ on a $V([a, c]; f) = V([a, b]; f) + V([b, c]; f)$. Attention cependant à ne pas commettre d'erreur du style $V([a, c[; f) = V([a, b[; f) + V([b, c[; f)$. La formule correcte, pour $b < c$, est $V([a, c[; f) = V([a, b]; f) + V([b, c[; f)$.

Un exemple instructif est celui d'une fonction, non nécessairement continue, mais de classe C^1 par morceaux. Vous prouverez que la variation totale est la somme d'une part de tous les (modules des) << (demi)-sauts >> aux discontinuités de f, et d'autre part de $\int_a^b |f'(t)|dt$, la fonction continue par morceaux f' étant définie à l'exception d'un nombre fini de points (à propos, quelle différence entre $f(x)$ et $\int_a^x f'(t)dt$?).

Soit donc f de variation bornée sur $[a, b]$. Remarquez en passant que f est alors bornée. De plus je vais montrer que f admet une limite à droite $f(a^+)$ en a (un raisonnement analogue montre que f a en tout point une limite à droite et une à gauche). En effet comme $V([a,b];f) < \infty$ on a aussi $V(]a,b];f) < \infty$. Soit (x_j) avec $V((x_j);f) \geq V(]a,b];f) - \epsilon$ (et $x_0 > a$). Soit $y \in]a,x_0[$. En considérant la subdivision $y_0 = y$, $y_1 = x_0$, ..., on obtient $|f(y) - f(x_0| + V((x_j);f) \leq V(]a,b];f)$. Donc $|f(y) - f(x_0)| \leq \epsilon$. Donc $|f(y) - f(z)| \leq 2\epsilon$ pour $a < y,z \leq x_0$ et le critère de Cauchy est vérifié pour l'existence de $f(a^+)$.

Je vous laisse maintenant la formule suivante comme un exercice assez facile : $V([a,b];f) = |f(a) - f(a^+)| + V(]a,b];f)$. Plus généralement

$$V([x,y];f) = |f(x) - f(x^+)| + V(]x,y[;f) + |f(y^-) - f(y)|$$

J'aurai aussi besoin plus tard de la propriété suivante :

$$\lim_{y \to x^+} V(]x,y];f) = 0$$

La technique de preuve est semblable à celle pour établir l'existence de $f(a^+)$. Soit $\epsilon > 0$ et soit (x_j) une subdivision de $]x,b]$ avec $x_0 > x$ et $V((x_j);f) \geq V(]x,b];f) - \epsilon$. Supposons maintenant $x < y \leq x_0$, et considérons une subdivision quelconque (y_k) de l'intervalle $]x,y]$. En mettant les (y_k) ensemble avec les (x_j) et en considérant la variation associée, on obtient la majoration $V((y_k);f) \leq \epsilon$. Donc $V(]x,y];f) \leq \epsilon$. Ainsi $\lim_{y \to x} V(]x,y];f) = 0$ est établi et on en déduit aussi $\lim_{y \to x} V([x,y];f) = |f(x^+) - f(x)|$. Dans le même style on a par exemple pour $a < y \leq z \leq b$: $\lim_{x \to y^-} V([x,z];f) = |f(y) - f(y^-)| + V([y,z];f)$, etc...

Il est évident qu'une fonction f est à variation bornée si et seulement si Re(f) et Im(f) le sont. Je suppose maintenant que f est *à valeurs réelles*. Considérons la fonction $x \mapsto f_1(x) = V([a,x];f)$. Elle est croissante. Définissons $f_2(x) = f_1(x) - f(x)$, et examinons pour $x < y$ la quantité $f_2(y) - f_2(x) = f_1(y) - f_1(x) - (f(y) - f(x)) = V([x,y];f) - (f(y) - f(x))$. La dernière expression est certainement ≥ 0. Donc f_2 est aussi une fonction croissante. Autrement dit $f = f_1 - f_2$ est la différence de deux fonctions croissantes.

Il est par ailleurs évident que des combinaisons linéaires de fonctions de variations bornées sont de variations bornées et que les fonctions croissantes ou décroissantes sont de variations bornées. On a donc obtenu :

Théoréme : *les fonctions à valeurs complexes de variations bornées sont exactement les fonctions de la forme* $f = f_1 - f_2 + i(f_3 - f_4)$ *avec* f_1, f_2, f_3, f_4 *croissantes.*

Remarquons que toute fonction de variation bornée est Riemann intégrable puisque les fonctions monotones sont Riemann intégrables. Aussi elle est bornée et n'a au pire qu'un nombre dénombrable de discontinuités puisque c'est le cas des fonctions monotones.

Variation bornée et mesures complexes

Vous pouvez sauter cette section dans une première lecture, et continuer avec la discussion du théorème de Dirichlet-Jordan en sautant ensuite les paragraphes parlant de mesures complexes. Le théorème précédent est fortement réminiscent de la décomposition $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$ d'une mesure complexe en une combinaison de mesures positives finies. Rappelons qu'il existe un chapitre essentiel du cours d'intégration qui montre qu'à toute fonction g croissante et continue à gauche sur $[a,b[$, on peut associer une mesure positive, dite de Lebesgue-Stieltjes, telle que $\mu([a,x[) = g(x) - g(a)$ pour $a \leq x < b$, autrement dit

$$\mu([x,y[) = g(y) - g(x)$$

pour $a \leq x \leq y < b$ (on notera que $\mu(\{a\}) = g(a^+) - g(a)$, plus généralement $\mu(\{x\}) = g(x^+) - g(x)$ pour tout x). Et à toute mesure positive μ sur $[a,b[$ on peut associer les fonctions croissantes et continues à gauche $g(x) = C + \mu([a,x[)$, avec $C = g(a)$ arbitraire (la mesure sera finie si et seulement si $g(b^-) < \infty$).

Donc, revenons à notre fonction à valeurs complexes et de variation bornée sur $[a,b]$. On l'écrit $f = f_1 - f_2 + i(f_3 - f_4)$ avec f_1, f_2, f_3, f_4 croissantes, et on remplace chaque f_j par f_j^-. Chaque fonction croissante f_j n'a au pire qu'un nombre dénombrable de points de discontinuités, et donc la fonction croissante $x \to f_j^-(x) = f_j(x^-)$ (en $x = a$, on convient que $f_j^-(a) = f_j(a)$), qui est continue *à gauche*, ne peut différer de f_j qu'en un nombre dénombrable de points. Alors $f^* = f_1^- - f_2^- + i(f_3^- - f_4^-)$ est continue à gauche et ne différe de f qu'en au pire un nombre dénombrable de points. Par construction $f^*(a) = f(a)$ et par ailleurs comme $f^*(y) = f(y)$ à un nombre dénombrable d'exceptions près et que $f^*(x) = \lim_{y \to x, y<x} f^*(y)$ on a aussi $f^*(x) = f(x^-)$ pour tout $x > a$ (rappel : on sait que $f(x^-)$ existe !), donc f^* ne dépend pas des choix faits et est égal à f^-. En considérant les mesures positives finies μ_j associées aux f_j^-, on aboutit donc à une écriture $f(x^-) = f(a) + \mu([a,x[)$ pour une certaine mesure complexe μ sur $[a,b[$. On pourra alors définir $\mu(\{b\})$ comme valant $f(b) - f(b^-)$. Réciproquement les formules $x \mapsto C + \mu([a,x[)$ et $f(b) = C + \mu([a,b])$ définissent une fonction à variation bornée sur $[a,b]$ et continue à gauche sur $[a,b[$.

Donc, toute fonction f de variation bornée sur $[a,b]$ définit de manière unique une mesure complexe sur $[a,b]$ via les formules $\mu([a,x[) = f(x^-) - f(a)$, $\mu(\{b\}) = f(b) - f(b^-)$ et f coïncide avec $f(a) + \mu([a,x[)$ en tout point où elle est continue à gauche. L'on a $\mu([a,b]) = f(b) - f(a)$. On peut aussi si l'on préfère, utiliser des formules du genre $C + \mu([a,x])$ (continue à droite) ou $C + \mu([x,b])$ (continue à gauche) ou $f(x^+) = f(b) + \mu(]x,b])$ (continue à droite). Ou encore on conviendra que $f(x) = f(a) + \mu([a,x[) + \frac{1}{2}\mu(\{x\})$ pour $a < x < b$ ce qui impose $f(x) = \frac{1}{2}(f(x^-) + f(x^+))$ en tout point de $]a,b[$, ce qui est utile en théorie des séries de Fourier.

Un dernier mot (assez long, désolé). Soit f de variation bornée sur $[a,b]$ et normalisée de manière à être en chaque x soit continue à droite, soit continue à gauche, soit à vérifier la formule $f(x) = \frac{1}{2}(f(x^-) + f(x^+))$ en tout point intérieur (ce que l'on veut en fait c'est que f(x) soit sur le segment dans le plan complexe $[f(x^-), f(x^+)]$ pour tout x de sorte que $|f(x^+) - f(x^-)| = |f(x) - f(x^-)| + |f(x^+) - f(x)|$). Soit μ la mesure de Lebesgue-Stieltjes associée, qui vérifie donc $\mu(\{a\}) = f(a^+) - f(a)$, $\mu(\{b\}) = f(b) - f(b^-)$, $\mu(]x, y[) = f(y^-) - f(x^+)$.

<u>Théorème</u> : $V([a, b]; f) = |\mu|([a, b])$ *est la variation totale de* μ. (*)

Tout d'abord si $a = x_0 < x_1 < \cdots < x_N = b$, alors $V((x_j)) \leq |f(a) - f(a^+)| + |f(a^+) - f(x_1^-)| + |f(x_1^-) - f(x_1)| + |f(x_1) - f(x_1^+)| + \ldots$. Or, par nos conventions sur f en chaque discontinuité, $|f(x_1^-) - f(x_1)| + |f(x_1) - f(x_1^+)| = |f(x_1^-) - f(x_1^+)|$, etc..., donc $V((x_j)) \leq |\mu(\{a\})| + |\mu(]a, x_1[)| + |\mu(\{x_1\})| + |\mu(]x_1, x_2[)| + \ldots$ ce qui est toujours inférieur ou égal à $|\mu|([a, b])$ par définition de ce dernier. Donc en tout cas $V([a, b]; f) \leq |\mu|([a, b])$. Le problème est de montrer l'inégalité dans l'autre sens, sachant que $|\mu|([a, b])$ est défini comme étant le supremum des sommes $\sum |\mu(A_n)|$ pour toutes les partitions dénombrables $[a, b] = \bigcup_n A_n$.

Considérons (ne me demandez pas pourquoi, je suis mon instinct, guidé par le premier mouvement de la sérénade pour cordes de Piotr Tchaikowski) la fonction croissante $g(x) = V([a, x]; f)$. Il est associé à g une mesure positive finie ν par les formules $\nu([a, x[) = g(x^-) - g(a) = g(x^-) = V([a, x[; f)$, $\nu(\{b\}) = g(b) - g(b-) = |f(b) - f(b^-)|$, $\nu([a, b]) = g(b) - g(a) = V([a, b]; f)$.

Comparons $\mu(]x, y[)$ et $\nu(]x, y[)$. Tout d'abord on a $\nu(]x, y[) = g(y^-) - g(x^+)$ et cela en fait coïncide avec $V(]x, y[; f)$ (en effet, on sait que $g(x^+) = g(x) + \lim_{z \to x^+} V([x, z]; f) = g(x) + |f(x^+) - f(x)|$ et $g(y^-) = V([a, y[; f) = g(y) - |f(y) - f(y^-)|$, donc $g(y^-) - g(x^+) = g(y) - g(x) - |f(y) - f(y^-)| - |f(x^+) - f(x)| = V([x, y]; f) - |f(y) - f(y^-)| - |f(x^+) - f(x)| = V(]x, y[; f)$). Par ailleurs $\mu(]x, y[) = f(y^-) - f(x^+)$ donc $|\mu(]x, y[)| \leq |f(y^-) - f(x^+)| \leq V(]x, y[; f)$. Ainsi

$$x < y \Longrightarrow \quad |\mu(]x, y[)| \leq \nu(]x, y[)$$

De plus on a $\nu([a, x[) = g(x^-) - g(a) = V([a, x[; f)$ et $\mu([a, x[) = f(x^-) - f(a)$ donc $|\mu([a, x[)| \leq |\nu([a, x[)|$. Et finalement $\nu(]y, b]) = g(b) - g(y^+) = g(b) - g(y) - |f(y^+) - f(y)| = V([y, b]; f) - |f(y^+) - f(y)| = V(]y, b]; f)$ et $\mu(]y, b]) = f(b) - f(y^+)$ donc aussi $|\mu(]y, b])| \leq \nu(]y, b])$. Tout ouvert U pour la topologie induite dans l'intervalle $[a, b]$ est une union disjointe dénombrable d'intervalles du type $[a, x[$, $]x, y[$, et $]y, b]$. Donc on a $|\mu(U)| \leq \nu(U)$ pour tout ouvert. Cela implique donc $|\mu(A)| \leq \nu(A)$ pour (maintenant c'est un rêve d'amour de Franz Liszt, interprété par François René Duchable. Ça change des choses qui commencent par Blo et terminent par cage) toute intersection dénombrable (que l'on peut prendre décroissante) d'ouverts. Bon, je dois réfléchir. Considérons une fonction continue à valeurs complexes $\phi(x)$ quelconque ; alors soit $\eta > 0$ et décomposons le plan complexe (merci Mozart et Anne-Sophie Mutter) en les translatés

(*) <u>Attention</u> : il aurait été plus prudent d'écrire $V(]a, b[; f) = |\mu|(]a, b[)$ (c'est équivalent car $|\mu|(\{a\}) = |\mu(\{a\})| = |f(a) - f(a^+)|$, etc...) car alors plus généralement $V(]c, d[; f) = |\mu|(]c, d[)$ pour tout $a \leq c \leq d \leq b$; Mais comme $|\mu|(\{c\}) = |f(c^+) - f(c^-)|$, tandis que $V([c, d]; f)$ ne tient compte que de $|f(c^+) - f(c)|$ on peut avoir $V([c, d]; f) < |\mu|([c, d])$.

du << carré >> $\{0 < \mathrm{Re}(z), \mathrm{Im}(z) < \eta\} \cup [0,\eta[\cup [0, i\,\eta[$. Les images réciproques de ces carrés sont chacune une intersection dénombrable d'ouverts car ϕ est continue. Donc en examinant $\int_{[a,b]} \phi(x) d\mu(x)$ avec un η-découpage des valeurs de ϕ, en faisant tendre η vers 0, j'obtiens (Toréador, toréador... l'amour, l'amour d'abord) $\left|\int_{[a,b]} \phi(x) d\mu(x)\right| \leq \int_{[a,b]} |\phi(x)| d\nu(x)$ pour toute fonction continue sur $[a,b]$, à valeurs complexes.

Petite pause Stade 2. Reprise. En utilisant que les fonctions continues sont denses dans l'espace $L^1([a,b], |\mu| + \nu)$,(*) nous pouvons alors affirmer que $\left|\int_{[a,b]} \phi(x) d\mu(x)\right| \leq \int_{[a,b]} |\phi(x)| d\nu(x)$ pour toute fonction dans cet espace par exemple pour toute fonction mesurable bornée. Je rappelle qu'il existe une fonction $h(x)$ de module 1 telle que l'on puisse écrire $d\mu(x) = h(x) d|\mu|(x)$, on utilise $\phi(x) = \overline{h(x)}$ et on obtient

$$|\mu|([a,b]) = \int_{[a,b]} d|\mu|(x) \leq \int_{[a,b]} d\nu(x) = V([a,b]; f)$$

Le Théorème est ainsi démontré.

Autrement dit ce qui a été prouvé ici c'est que lorsque μ est une mesure complexe sur $[a,b]$ sa variation totale $\|\mu\| = |\mu|([a,b])$ au sens de l'annexe précédente est aussi le supremum de toutes les sommes finies $|\mu([a, x_1[)| + |\mu([x_1, x_2[)| + \cdots + |\mu([x_{N-1}, b[)| + |\mu(\{b\})|$, pour toutes les subdivisions $a = x_0 < x_1 < \cdots < x_N = b$. Une troisième caractérisation de $\|\mu\|$ est à portée de main si vous reprenez notre technique de démonstration c'est que $\|\mu\|$ est la plus petite constante C avec $|\int_{[a,b]} \phi(x)\, d\mu(x)| \leq C \cdot \|\phi\|_\infty$ pour toutes les fonctions continues $\phi : [a,b] \dashrightarrow \mathbb{C}$. Cependant je n'en parle pas plus car cela m'entraînerait sur la pente glissante vers une discussion du fondamental << Théorème de représentation de Riesz >> (à ne pas confondre avec le théorème de Fréchet-Riesz) qui caractérise les mesures complexes comme étant les éléments du dual de l'espace de Banach des fonctions continues muni de la norme sup (de la convergence uniforme). Je dois laisser un peu de travail à mes collègues sévissant en M1 et M2, et de plus j'ai déjà réécrit au fil de ces annexes une trop grande partie du Rudin << Analyse réelle et complexe >>, si vous vous sentez intéressé(e) allez le regarder (il vous faudra surmonter votre dégoût de l'horrible typographie de l'édition française actuelle, qui induit même des erreurs dans certaines preuves ou énoncés d'exercices !).

(*) damenède ! je n'ai prouvé cela que pour dx. Bon, voyons voir, oui, je rappelle que l'idée (communément appelée << technique Burnolienne >> dans la littérature ; euh...non, pas encore, mais cela va venir), dans la démonstration du cours, c'est de montrer que les parties mesurables A pour lesquelles $\mathbb{1}_A$ est approchable (dans le cours par des fonctions en escalier, ici par des fonctions continues) forment une tribu, ensemble vide ok, complémentaire...ok, union croissante......ok si union finie, union de deux.........les << identités Burnoliennes >> font le job ici aussi ; et bien sûr cette tribu contient les intervalles ouverts, voyons, oui, ok on approxime par en dessous, convergence dominée, c'est bon. Sous-jacent à cela sans le dire il y a la propriété de << régularité >> des mesures de Lebesgue-Stieltjes, $\nu(A)$ est l'infimum des $\nu(U)$ pour U ouvert contenant A et le supremum des $\nu(K)$ pour K compact inclus dans A. Si vous voulez approfondir, direction le Rudin ou tout livre sérieux sur la théorie de l'intégration.

Dirichlet-Jordan

Théorème : *soit* f *une fonction* 2π*-périodique et de variation bornée sur* $[0,2\pi]$*. Alors en tout* x *on a*

$$\lim S_N(f)(x) = \frac{f(x^-)+f(x^+)}{2}$$

Par le théorème de localisation si on suppose seulement f de variation bornée sur un sous-intervalle [a, b] alors la conclusion vaudra pour tous les $x \in]a,b[$. Je vais faire une première preuve, qui utilise la mesure complexe μ de Lebesgue-Stieltjes associée à f et le théorème de la convergence dominée (on peut modifier la preuve pour n'utiliser que la convergence uniforme, mais à partir du moment où l'on parle de mesures, cela n'a pas grand sens de vouloir éviter l'emploi du théorème de la convergence dominée ou celui de la convergence monotone !). Dans cette preuve je vais utiliser la formule

$$\sum_{k=1}^{\infty} \frac{\sin(kx)}{k} = \frac{\pi - x}{2} \qquad (0 < x < 2\pi)$$

et surtout la propriété

$$\exists C < \infty \qquad \forall x \in \mathbb{R} \quad \forall N \in \mathbb{N} \quad \left|\sum_{k=1}^{N} \frac{\sin(kx)}{k}\right| \leq C$$

L'existence d'un tel $C < \infty$ a été traitée en exercice (voir en particulier l'exercice ??, pour une parmi de multiples méthodes possible).

Soit $L = \frac{f(x^-)+f(x^+)}{2}$ et $g(t) = \frac{f(x+t)+f(x-t)}{2} - L$ sur $]0,\pi]$, de sorte que $g(0^+) = 0$. On pose par convention $g(0) = 0$. Il s'agit de montrer que $\lim \int_0^\pi g(t)D_N(t)dt = 0$, sachant que g est de variation bornée et que $g(0^+) = 0$. Pour ceux qui ont peur des mesures complexes, sautez ce qui suit. Une variante qui utilise plutôt la deuxième formule de la moyenne est donnée ensuite ; on y utilise toujours l'existence d'une majoration uniforme pour les sommes $\sum_{k=1}^N \frac{\sin(kx)}{k}$.

L'intégrale ne change pas si je modifie g en le nombre dénombrable de points où elle est discontinue et donc je peux lui imposer d'être continue à gauche. J'écris $g(t) = \int_{[0,t[} d\mu(u)$, et je prends note que $\mu(\{0\}) = g(0^+) = 0$, donc je peux aussi écrire $g(t) = \int_{]0,t[} d\mu(u)$. J'applique Fubini (Fubini vaut pour μ_1, μ_2, μ_3, μ_4 de l'écriture $\mu = \mu_1 - \mu_2 + i(\mu_3 - \mu_4)$), soigneusement, car ici on a une mesure qui peut donner une masse aux singletons :

$$\begin{aligned}\int_{0<t<\pi} g(t)D_N(t)dt &= \iint_{0<u<t<\pi} D_N(t)\,dt\,d\mu(u)\\ &= \int_{0<u<\pi} \left(\int_u^\pi D_N(t)\,dt\right)\,d\mu(u)\end{aligned}$$

Je calcule $\int_u^\pi D_N(t)\,dt = \int_u^\pi (1+2\cos t+\cdots+2\cos Nt)\,dt = \pi - u - 2\sum_{k=1}^N \frac{\sin(ku)}{k}$. Je sais que pour chaque $u \in]0,\pi]$ cela tend vers 0 lorsque $N \to \infty$ et surtout que toutes ces expressions sont bornées indépendamment de u et de N, donc il y a convergence dominée vers 0 sur $]0,\pi]$ (ou alors vous travaillez sur $[0,\pi]$ et il y a alors convergence presque partout puisque $|\mu|(\{0\}) = 0$). Le théorème de la convergence dominée (valable pour μ car valable pour μ_1, ..., μ_4) donne alors la conclusion recherchée.

À propos de la formule $\pi - u = 2\sum_{k=1}^\infty \frac{\sin(ku)}{k}$, je signale en passant qu'elle découle immédiatement du fait que $\lim \int_u^\pi D_N(t)\,dt = 0$ pour $0 < u \leq \pi$ par application du Lemme de Riemann-Lebesgue, compte tenu de la formule $D_N(t) = \dfrac{\sin((N+\frac{1}{2})t)}{\sin(\frac{t}{2})}$. Sans même d'ailleurs à avoir à invoquer le Lemme de Riemann-Lebesgue il suffit élémentairement d'intégrer par parties pour faire apparaître un $\dfrac{1}{N+\frac{1}{2}}$!

Deuxièmes formules de la moyenne

La raison pour le pluriel apparaîtra bientôt. Je signale tout de suite que toutes ces << deuxièmes >> formules s'obtiennent très élémentairement par des intégrations par parties, lorsqu'elles sont possibles. Le substitut à l'intégration par parties, c'est de se ramener à une intégrale double, avec des mesures complexes, et d'utiliser le théorème de Fubini, comme dans la preuve que j'ai faite de Dirichlet-Jordan. Mais bon, je vais faire sans ici. Soit donc g de variation bornée sur $[a,b]$ et soit f une fonction Lebesgue intégrable quelconque. Comme g est bornée le produit $f(x)g(x)$ est intégrable. On veut majorer intelligemment $I = \int_a^b f(x)g(x)\,dx$. Prenons une subdivision $a = x_0 < x_1 < \cdots < x_N = b$ quelconque. Et considérons l'approximation :

$$S = \sum_{j=1}^N \int_{x_{j-1}}^{x_j} f(x)g(x_j)\,dx$$

Déjà, que pouvons-nous dire que $I - S$? On a :

$$|I-S| \leq \sum_{j=1}^N \left(\int_{x_{j-1}}^{x_j} |f(x)|\,dx\right) V([x_{j-1},x_j];g) \leq \left(\max_j \int_{x_{j-1}}^{x_j} |f(x)|\,dx\right) V([a,b];g)$$

Je rappelle que (pourquoi ?) la fonction $x \mapsto \int_a^x |f(t)|dt$ est continue, donc uniformément continue sur $[a,b]$. Donc on aura $S \to I$ pour tous les choix de subdivisions assurant que le pas $\max_j |x_j - x_{j-1}|$ tende vers zéro pour $N \to \infty$. Par exemple on peut prendre $x_j = a + j\frac{b-a}{N}$.

Posons maintenant $u_j = g(x_j)$, $w_j = \int_{x_j}^b f(x)\,dx$ et écrivons $\int_{x_{j-1}}^{x_j} f(x)g(x_j)\,dx = (w_{j-1} - w_j)u_j$. Ainsi (sommation d'Abel ; remarquez $w_N = 0$) :

$$S = \sum_{j=1}^N (w_{j-1} - w_j)u_j = w_0u_1 + w_1(u_2 - u_1) + \cdots + w_{N-1}(u_N - u_{N-1})$$

$$S - g(a)\int_a^b f(x)\,dx = w_0(u_1-u_0)+w_1(u_2-u_1)+\cdots+w_{N-1}(u_N-u_{N-1})$$

$$\left|S - g(a)\int_a^b f(x)\,dx\right| \le \Big(\max_{0\le j\le N}|w_j|\Big)\big(|u_1-u_0|+|u_2-u_1|+\cdots+|u_N-u_{N-1}|\big)$$
$$\le \left(\sup_{[a,b]}\left|\int_\xi^b f(x)\,dx\right|\right)V([a,b];g)$$

Nous avons donc identifié un disque fermé du plan complexe dans lequel tombe toutes les expressions S. Comme l'intégrale I est un point limite de telles S, on a donc prouvé une première version de la deuxième formule de la moyenne :

$$\int_a^b f(x)g(x)\,dx = g(a)\int_a^b f(x)\,dx + R \qquad |R| \le \left(\sup_{[a,b]}\left|\int_\xi^b f(x)\,dx\right|\right)V([a,b];g)$$

En particulier :

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \Big(|g(a)|+V([a,b];g)\Big)\sup_{[a,b]}\left|\int_\xi^b f(x)\,dx\right|$$

Je vous conseille en supposant g de classe C^1 d'intégrer par parties l'intégrale de gauche et vous retrouverez immédiatement ce résultat. Par ailleurs dans le terme de gauche on peut remplacer g(a) par $g(a^+)$ et aussi g(b) par $g(b^-)$ et on a ainsi une autre majoration :

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \Big(|g(a^+)|+V(]a,b[;g)\Big)\sup_{[a,b]}\left|\int_\xi^b f(x)\,dx\right|$$

On peut échanger dans la formule de la moyenne les rôles de a et de b :

$$\left|\int_a^b f(x)g(x)\,dx\right| \le \Big(|g(b^-)|+V(]a,b[;g)\Big)\sup_{[a,b]}\left|\int_a^\xi f(x)\,dx\right|$$

Mais passons maintenant à la variante la plus générale : on modifie la preuve plus haut en écrivant certes $\int_{x_{j-1}}^{x_j} f(x)\,dx = w_{j-1}-w_j$ mais avec $w_j = -F(x_j)$ pour $F(x) = C+\int_a^x f(t)\,dt$, $C\in\mathbb{C}$ quelconque. Le réarrangement télescopique donne :

$$S = w_0 g(x_1) + w_1(g(x_2)-g(x_1))+\cdots+w_{N-1}(g(b)-g(x_{N-1})) - w_N g(b)$$

$$S+F(a)g(a)-F(b)g(b) = w_0(g(x_1)-g(a))+w_1(g(x_2)-g(x_1))+\cdots+w_{N-1}(g(b)-g(x_{N-1}))$$

et le terme de droite est majoré par $(\sup_{[a,b]}|F(\xi)|)V([a,b];g)$. On conclut :

$$\left|\int_a^b f(x)g(x)\,dx - \Big(F(b)g(b)-F(a)g(a)\Big)\right| \le \Big(\sup_{[a,b]}\big|F(\xi)\big|\Big)\,V([a,b];g)$$

L'analogie avec la formule d'intégration par parties fait que cette inégalité est facile à mémoriser ! On peut remplacer g(a), g(b), V([a,b];g) par $g(a^+)$, $g(b^-)$ et $V(]a,b[;g)$. Avec $f(x) = e^{ix}$ par exemple le choix de $F(x) = -ie^{ix}$ peut être plus avantageux du point de vue de $\sup_\xi |F(\xi)|$ que $\int_a^x e^{it}\,dt$ ou $-\int_x^b e^{it}\,dt$.

Pour g monotone et f à valeurs réelles, il y a une version un peu plus précise et très classique. Supposons g croissante (non constante). Revenons à :

$$S - g(a)\int_a^b f(x)\,dx = w_0(g(x_1) - g(a)) + w_1(g(x_2) - g(x_1)) + \cdots + w_{N-1}(g(b) - g(x_{N-1}))\ ,$$

le terme de droite, divisé par $g(b) - g(a)$ est un barycentre des valeurs $w_j = \int_{x_j}^b f(x)\,dx$ prises par la fonction continue $\xi \dashrightarrow F(\xi) = \int_\xi^b f(x)\,dx$. Soit J l'intervalle (fermé borné) $F([a,b])$. On a donc $S \in g(a)\int_a^b f(x)\,dx + (g(b) - g(a))J$. C'est donc aussi le cas de l'intégrale I qui est une limite de tels S :

$$\exists\xi \in [a,b] \qquad \int_a^b f(x)g(x)\,dx = g(a)\int_a^b f(x)\,dx + (g(b) - g(a))\int_\xi^b f(x)\,dx$$

On peut remplacer $g(a)$ par $g(a^+)$ et $g(b)$ par $g(b^-)$.

$$\exists\xi \in [a,b] \qquad \int_a^b f(x)g(x)\,dx = g(a^+)\int_a^b f(x)\,dx + (g(b^-) - g(a^+))\int_\xi^b f(x)\,dx$$

En fait l'on peut toujours prendre $\xi \in\ ]a,b[$ (sautez ce paragraphe en première lecture). Si g est constante sur $]a,b[$ alors en fait $g(b^-) - g(a^+) = 0$ et la formule marche pour tous les ξ. Je suppose donc g non constante sur $]a,b[$. Supposons par exemple que $\xi = b$ marche. Il s'agit alors de montrer qu'il existe un $y \in\ ]a,b[$ avec $\int_y^b f(x)\,dx = 0$. Sinon, le signe est constant, disons strictement positif. Prenons $a < c < d < b$ avec $g(c) < g(d)$. Sur chaque sous-intervalle on applique la forme actuelle de la formule, ce qui donne $\int_a^c f(x)g(x)\,dx = g(a^+)\int_a^c f(x)\,dx + (g(c) - g(a^+))\int_{\xi_1}^c f(x)\,dx = g(a^+)\int_a^b f(x)\,dx + (g(c) - g(a^+))\int_{\xi_1}^b f(x)\,dx - g(c)\int_c^b f(x)\,dx$, puis $\int_c^d f(x)g(x)\,dx = g(c)\int_c^d f(x)\,dx + (g(d) - g(c))\int_{\xi_2}^d f(x)\,dx = g(c)\int_c^b f(x)\,dx + (g(d) - g(c))\int_{\xi_2}^b f(x)\,dx - g(d)\int_d^b f(x)\,dx$, enfin $\int_d^b f(x)g(x)\,dx = g(d)\int_d^b f(x)\,dx + (g(b^-) - g(d))\int_{\xi_3}^b f(x)\,dx$. En sommant on obtient : $\int_a^b f(x)g(x)\,dx = g(a^+)\int_a^b f(x)\,dx + (g(c) - g(a^+))\int_{\xi_1}^b f(x)\,dx + (g(d) - g(c))\int_{\xi_2}^b f(x)\,dx + (g(b^-) - g(d))\int_{\xi_3}^b f(x)\,dx$. Comme $\xi_2 \in [c,d] \subset\ ]a,b[$ le troisième terme est strictement positif, tandis que le deuxième et le quatrième sont positifs ou nuls. On a donc $\int_a^b f(x)g(x)\,dx > g(a^+)\int_a^b f(x)\,dx$ ce qui contredit le fait que $\xi = b$ marchait dans la formule. On montrerait d'une manière exactement semblable que si $\xi = a$ marche, c'est-à-dire si $\int_a^b f(x)g(x)\,dx = g(b^-)\int_a^b f(x)\,dx$ alors il existe $y \in]a,b[$ avec $\int_a^y f(x)\,dx = 0$ et donc aussi $\int_a^b f(x)g(x)\,dx = g(b^-)\int_a^b f(x)\,dx + (g(a^+) - g(b^-))\int_a^y f(x)\,dx = g(a^+)\int_a^b f(x)\,dx + (g(b^-) - g(a^+))\int_y^b f(x)\,dx$.

Nous avons donc obtenu la formule célèbre (f à *valeurs réelles*, g monotone) :

$$\exists\xi \in\ ]a,b[\qquad \int_a^b f(x)g(x)\,dx = g(a^+)\int_a^\xi f(x)\,dx + g(b^-)\int_\xi^b f(x)\,dx$$

On avait g croissante non constante ; g constante est évident, et pour g décroissante on remplace g par -g. Pour g à variation bornée, ou f à valeurs complexes, on doit se contenter des inégalités établies plus haut.

Encore une fois, lorsque g est C^1, intégrez par parties, et vous aurez tout ce que peut donner la deuxième formule de la moyenne. Et en toute généralité la puissance optimale est obtenue en exprimant g avec une mesure complexe, et en utilisant Fubini, car là on a un résultat exact, pas une inégalité, ou une pseudo-égalité avec un ξ que l'on ne connaît pas. Une erreur très fréquente est d'utiliser la formule avec un ξ alors que f est à valeurs complexes, pas à valeurs réelles. L'inégalité que l'on écrit après est en général valable, mais l'étape intermédiaire était fausse.

D i r i c h l e t - J o r d a n (b i s)

Voici une formulation plus élémentaire de la preuve du théorème de Dirichlet-Jordan. On veut prouver $\lim\int_0^\pi g(t)D_N(t)dt = 0$, sachant que g est de variation bornée et que $g(0^+) = 0$. On fait la découpe usuelle pour un $\delta \in]0,\pi]$:

$$\int_0^\pi g(t)D_N(t)\,dt = \int_0^\delta g(t)D_N(t)\,dt + \int_\delta^\pi g(t)\frac{\sin((N+\frac{1}{2})t)}{\sin(\frac{t}{2})}\,dt$$

Le deuxième terme à droite tend vers 0 par le Lemme de Riemann-Lebesgue (comme $g(t)/\sin(\frac{t}{2})$ est Riemann intégrable sur $[\delta,\pi]$ il suffit d'invoquer quelque chose d'<< élémentaire >> comme l'inégalité de Bessel ou l'approximation par une fonction en escalier). On a donc

$$\forall\delta\in\,]0,\pi]\quad \limsup_{N\to\infty}\left|\int_0^\pi g(t)D_N(t)\,dt\right| \le \limsup_{N\to\infty}\left|\int_0^\delta g(t)D_N(t)\,dt\right|$$

On utilise la deuxième formule de la moyenne :

$$\left|\int_0^\delta g(t)D_N(t)\,dt\right| \le \Big(|g(0^+)| + V(]0,\delta[;g)\Big)\sup_{]0,\delta]}\left|\int_\xi^\delta D_N(t)\,dt\right|$$

Le supremum est majoré par une constante absolue C, en écrivant $\int_\xi^\delta D_N(t)\,dt = \int_\xi^\pi D_N(t)\,dt - \int_\delta^\pi D_N(t)\,dt$ et en rappelant que $\int_u^\pi D_N(t)\,dt = \pi - u - 2\sum_{k=1}^N \frac{\sin(ku)}{k}$ est borné indépendamment de u et de N. De plus $g(0^+) = 0$. Ainsi :

$$\left|\int_0^\delta g(t)D_N(t)\,dt\right| \le V(]0,\delta[;g)\cdot C$$

et a fortiori $\limsup_{N\to\infty}\left|\int_0^\delta g(t)D_N(t)\,dt\right| \le V(]0,\delta[;g)\cdot C$. Donc

$$\forall\delta\in\,]0,\pi]\quad \limsup_{N\to\infty}\left|\int_0^\pi g(t)D_N(t)\,dt\right| \le V(]0,\delta[;g)\cdot C$$

Il suffit alors de rappeler que $\lim_{\delta\to 0} V(]0,\delta];g) = 0$ pour conclure la preuve.

Dans ce contexte, je laisse en exercice $c_n(g) = \int_0^{2\pi} g(t)e^{-int}\frac{dt}{2\pi} = \mathcal{O}(\frac{1}{|n|})$ pour g à variation bornée sur $[0,2\pi]$. Plus précisément $|c_n(g)| \le \frac{V([0,2\pi];g)}{2\pi|n|}$ (attention : on impose $g(2\pi) = g(0)$... pourquoi ?). La deuxième formule de la moyenne (version générale) ou Fubini permettent de le prouver.

Continuité absolue et Radon-Nikodym

Dans l'annexe précédente nous avons caractérisé les fonctions $F(x) = C + \int_{[a,x[} d\mu(t)$ associées aux mesures complexes sur $[a, b[$ comme étant les fonctions continues à gauche et de variation bornée (<< VB >> ou << BV >>) sur $[a, b[$. Parmi les mesures complexes il y a les mesures à densité $d\mu(t) = g(t)\,dt$ avec $g \in L^1(a, b; dt)$. Les fonctions $F(x) = C + \int_a^x g(t)\,dt$ sont continues (pourquoi ?). Elles ont en fait une propriété plus forte, dite de continuité absolue (<< AC >>), qui les caractérise, c'est l'objet de cette feuille.

Si vous vous reportez à l'annexe sur la convergence dominée, vous y verrez que nous avons prouvé : $\forall \epsilon > 0$, $\exists \delta > 0$, $|A| \leq \delta \Longrightarrow \int_A |g(t)|\,dt \leq \epsilon$. En particulier si $a \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_N < b_N \leq b$ en considérant $A = \bigcup_j]a_j, b_j[$ on obtient :

$$\forall \epsilon > 0 \qquad \exists \delta > 0 \quad \forall\, a \leq a_1 < b_1 \leq a_2 < b_2 \leq \cdots \leq a_N < b_N \leq b,$$
$$\sum_j |b_j - a_j| \leq \delta \quad \Longrightarrow \quad \sum_j |F(b_j) - F(a_j)| \leq \epsilon$$

Cette condition nécessaire sur F s'appelle la continuité absolue. Le résultat difficile que nous prouverons ici c'est que cette condition est aussi suffisante pour pouvoir écrire F sous la forme $C + \int_a^x g(t)\,dt$ pour une certaine $g \in L^1(a, b; dt)$. Je rappelle que nous savons qu'alors F est p.p. dérivable avec $F' = g$. Donc, une fonction absolument continue s'écrit $F(x) = C + \int_a^x F'(t)\,dt$ en sachant que F' n'est définie que presque partout.(*)

Je commence bien sûr par montrer que F est VB. Prenons un $N \gg 1$ tel que $\delta = \frac{b-a}{N}$ marche pour $\epsilon = 1$. Soit (x_j) une subdivision quelconque, on ne peut qu'augmenter $V((x_j); F)$ en incluant les points équidistants $a + k\frac{b-a}{N}$, $0 \leq k \leq N$. La variation de cette subdivision élargie est la somme de N sous-sommes, chacune étant majorée par $\epsilon = 1$, donc en fait $V((x_j); F) \leq N$ pour toute subdivision et donc $V([a, b]; F) \leq N$. Donc F est de variation bornée et il lui est associé une mesure complexe μ. Il s'agit de prouver que μ est une mesure à densité.

Soit A de mesure de Lebesgue nulle. On peut donc, pour tout $\delta > 0$ (en particulier pour un δ correspondant à un $\epsilon > 0$ dans la propriété AC de F), compte tenu de la construction usuelle de la mesure de Lebesgue, recouvrir A par une union dénombrable d'intervalles ouverts I_n, $n \geq 1$, dont la somme des longueurs est au plus δ. Soit $J_N = I_1 \cup \cdots \cup I_N$ et $A_N = A \cap J_N$. On a $|\mu|(A_N) \leq |\mu|(J_N)$. On peut réécrire

(*) il y a un théorème en fait d'un esprit assez différent qui dit que si f est une fonction partout dérivable, et si f' est Lebesgue intégrable, alors $f(x) = C + \int_a^x f'(t)\,dt$, autrement dit, sous ces hypothèses, f est AC.

J_N comme une union finie d'intervalles ouverts disjoints $]a_k, b_k[$, $1 \leq k \leq M$ (si $b_k = b$ comprendre $,b]$ au lieu de $,b_k[$ et idem pour a). Donc $|\mu|(J_N) = \sum_k |\mu|(]a_k, b_k[) = \sum_k V(]a_k, b_k[; F)$. Comme $\sum_k |b_k - a_k| = |J_N| \leq \sum_{1\leq n\leq N} |In| \leq \delta$, si l'on prend des subdivisions quelconques des M intervalles $]a_k, b_k[$ et que l'on applique aux variations associées la propriété AC de F pour ce ϵ et δ on obtient au total $\sum_k V(]a_k, b_k[; F) \leq \epsilon$ et donc $|\mu|(J_N) \leq \epsilon$. A fortiori $|\mu|(A_N) \leq \epsilon$ et par union dénombrable $|\mu|(A) \leq \epsilon$. Finalement comme $\epsilon > 0$ est arbitraire on a donc $|\mu|(A) = 0$. Donc tout ensemble Lebesgue-négligeable est $|\mu|$-négligeable. Il se trouve que cela caractérise les mesures à densité :

R a d o n - N i k o d y m

<u>T h é o r è m e</u> : *Soit* (X, ν) *un espace mesuré avec* ν *positive finie, et* μ *une mesure complexe telle que tout* ν*-négligeable est* μ*-négligeable (*μ *est dite* ν*-absolument continue). Alors* $d\mu(x) = g(x)d\nu(x)$ *pour un certain* $g \in L^1(X; \nu)$*, qui est dit être la << dérivée de Radon-Nikodym >> de* μ *par rapport à* ν*.*

Soit $\lambda = |\mu| + \nu$ et considérons l'espace de Hilbert $L^2(X; \lambda)$. Alors $f \mapsto \int_X f(x)d|\mu|(x)$ est une forme linéaire continue, donc de la forme $\int_X f(x)G(x)d\lambda(x)$. On a $\int_A G(x)d\lambda(x) = |\mu|(A) \geq 0$ pour toute partie mesurable donc $G \geq 0$ λ-p.p., et on peut changer G pour vérifier partout $G \geq 0$. De même $\int_A G(x)d\lambda(x) \leq \int_A d\lambda(x)$, donc $G \leq 1$ p.p. et on peut imposer $G \leq 1$ partout. Enfin soit $A = \{G = 1\}$. Alors $|\mu|(A) = \lambda(A)$ donc $\nu(A) = 0$. Donc, et c'est l'unique endroit où l'on utilisera l'hypothèse, A est aussi $|\mu|$-négligeable, donc A est λ-négligeable, donc au total on peut supposer $0 \leq G < 1$ partout.

Soit f une fonction mesurable bornée, on a, pour tout A mesurable :

$$\int_A f(x)d|\mu|(x) = \int_A f(x)G(x)(d|\mu|(x) + d\nu(x))$$

On applique cela aussi à fG, fG^2, fG^3, etc..., fG^{N-1}, et l'on fait la somme télescopique ce qui donne :

$$\int_A f(x)(1 - G(x)^N)d|\mu|(x) = \int_A f(x)(G(x) + \cdots + G(x)^N)d\nu(x)$$

Supposons $f \geq 0$. À gauche on applique la convergence dominée et à droite on applique la convergence monotone. Il vient, avec $g_0(x) = \frac{G(x)}{1-G(x)}$:

$$\int_A f(x)d|\mu|(x) = \int_A f(x)g_0(x)\, d\nu(x)$$

En prenant $A = X$, $f \equiv 1$ on obtient $g_0 \in L^1(X; \nu)$. Puis via $f = \max(f, 0) - \max(-f, 0)$ on étend à toute f réelle bornée, puis évidemment à toute f mesurable complexe bornée. Nous savons qu'il existe h de module 1 avec $\mu(A) = \int_A h(x)d|\mu|(x)$ pour tout A. Donc $\mu(A) = \int_A g(x)d\nu(x)$ avec $g(x) = h(x)g_0(x) \in L^1(X; \nu)$ puisque $|h| = 1$.

Il ne serait pas difficile d'obtenir la décomposition plus générale $\mu = g\nu + \mu_S$ de <u>toute</u> mesure complexe en une partie ν-abs.cont. et une partie << ν-singulière >> ; voir un livre compétent (théorème de Lebesgue-Radon-Nikodym).

Phénomène de Gibbs

Dans les graphes suivants on a représenté pour $N = 20$ et $N = 60$ les sommes partielles $S_N(x) = \sum_{k=1}^{N} \frac{\sin kx}{k}$ de la série de Fourier de la fonction impaire 2π-périodique qui vaut $\frac{\pi - x}{2}$ pour $0 < x < 2\pi$.

Fig. 2 - Un graphe de S_{20} (les axes ne sont pas à la même échelle)

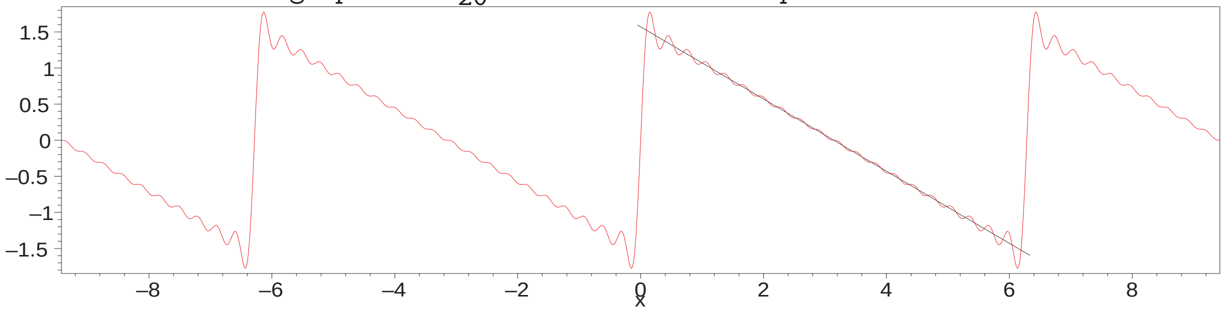
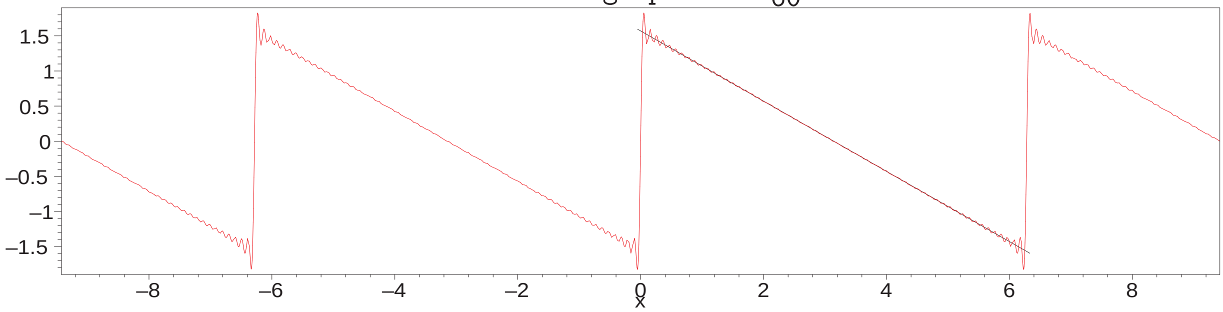


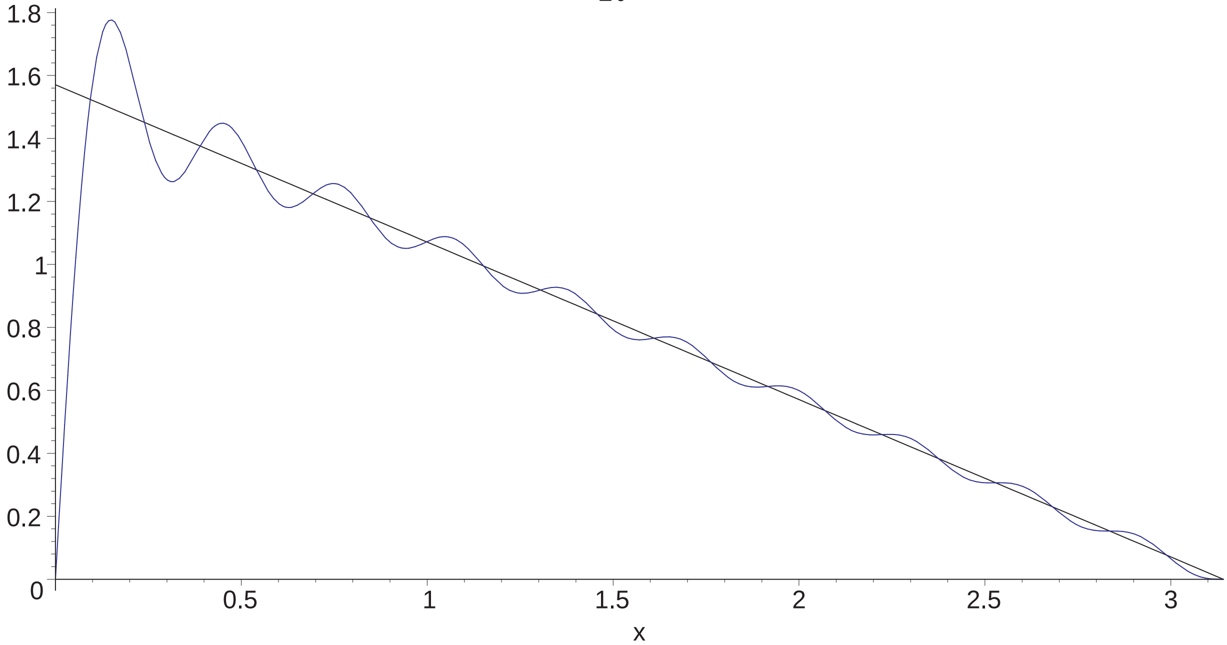
Fig. 3 - Un graphe de S_{60}



Le << phénomène de Gibbs >> saute aux yeux : les graphes des sommes partielles descendent un peu plus bas à gauche que $f(0^-) = -\frac{\pi}{2}$ et montent ensuite un peu plus haut que $f(0^+) = +\frac{\pi}{2}$. Nous verrons en fait que le premier maximum (local) du graphe de S_N (qui est en fait son maximum global) est pour $N \gg 1$ presque égal à $\int_0^\pi \frac{\sin(x)}{x} dx \approx 1{,}8519$. Ceci est supérieur d'environ 18% à $f(0^+) = \frac{\pi}{2}$ (ou plus précisément et pour l'exprimer d'une manière plus universelle l'écart à $\frac{1}{2}(f(0^+) + f(0^-))$ est supérieur d'environ 18% à $\frac{1}{2}(f(0^+) - f(0^-))$).

Il n'y a pas là un << phénomène >> mathématique bien mystérieux, et son analyse ne nécessite que des outils élémentaires. D'ailleurs bien que la contribution de Gibbs date de 1899, déjà en 1848, un dénommé Wilbraham en avait fait mention. Si la contribution de Gibbs a plus fait date, c'est pour des raisons sociologiques : il a trouvé un public qui a fait écho à son travail. Josiah Willard Gibbs (1839-1903) est célèbre pour ses travaux fondateurs en thermodynamique et physique statistique, et il est l'un des rares physiciens théoriciens de renom mondial en activité aux États-Unis au 19ième siècle. Son article de 1899 répondait à un compte rendu l'année précédente par Michelson(*) de la construction d'une machine traceuse de courbes, utilisée en particulier pour dessiner des graphes de sommes de sinus et cosinus. Testée sur des séries de Fourier de fonctions avec des discontinuités, on observait systématiquement un saut du dernier minima au premier maxima (ou le contraire) en excès de 18% par rapport à la discontinuité.

FIG. 4 - S_{20} sur $[0, \pi]$



Signalons d'abord une propriété manifeste et curieuse des sommes $S_N(x) = \sum_{1 \leq k \leq N} \frac{\sin(kx)}{k}$ c'est que visiblement pour $0 < x < \pi$ on a $S_N(x) > 0$.(*) Une surprenante démonstration par récurrence sur N est possible, la voici : d'abord on a besoin de la dérivée $S'_N(x)$. Clairement :

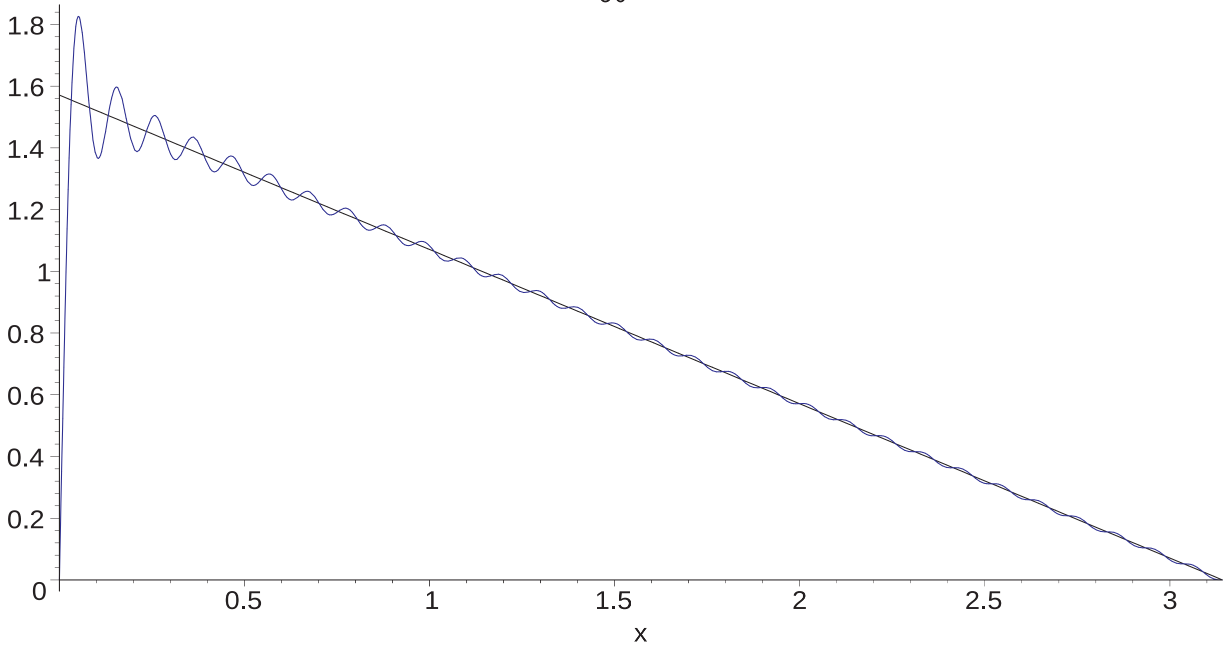
$$S'_N(x) = \cos(x) + \cos(2x) + \cdots + \cos(Nx) = \frac{D_N(x) - 1}{2} =$$
$$\frac{1}{2} \frac{\sin((N + \frac{1}{2})x) - \sin(\frac{x}{2})}{\sin(\frac{x}{2})} = \frac{\cos(\frac{N+1}{2}x) \sin(\frac{N}{2}x)}{\sin(\frac{x}{2})}$$

Pour $0 < x < 2\pi$, on a $\sin(\frac{x}{2}) > 0$, donc le signe de $S'_N(x)$ est celui de

(*) celui de la fameuse expérience d'interférométrie optique de Michelson-Morley dont le résultat a trouvé une justification dans la théorie de la relativité d'Einstein.

(*) Dans le beau livre de Bromwich << Theory of infinite series >> (2^e éd., 1926), les S_N sont étudiées dans une section intitulée << Fejér's Lemma >>. Il y a une mention d'un article de Jackson de 1911, qui contient en fait la première preuve de l'inégalité (conjecturée par Fejér) $S_N(x) > 0$ pour $0 < x < \pi$. Turán prouve un résultat plus général en 1952 et donne une réécriture << magique >> de S_N en une intégrale d'une fonction positive. Vietoris (1891-2002) fait de remarquables contributions, dont une publication à l'âge respectable de 102 ans...voir << Special functions >> de Andrews, Askey et Roy (1999).

Fig. 5 - S_{60} sur $[0,\pi]$



$\cos(\frac{N+1}{2}x)\sin(\frac{N}{2}x)$. Les facteurs $\cos(\frac{N+1}{2}x)$ et $\sin(\frac{N}{2}x)$ sont des fonctions analytiques et leurs zéros dans $]0,\pi[$ sont simples : ce sont les $t_k = \frac{\pi}{N+1} + k\frac{2\pi}{N+1} = (2k+1)\frac{\pi}{N+1}$, $0 \leq k \leq N$, et les $u_j = j\frac{2\pi}{N} = 2j\frac{\pi}{N}$, $1 \leq j \leq N-1$. Lorsque l'on voudra indiquer la dépendance par rapport à N on écrira $t_k^{(N)}$ et $u_j^{(N)}$. On ne peut avoir $t_k = u_j$ que si $2j(N+1) = (2k+1)N$. Or N et $N+1$ sont premiers entre eux donc N divise 2j. Cela n'est possible pour $1 \leq j \leq N-1$ que si $2j = N$ et alors $2k+1 = N+1$, donc $j = k = \frac{N}{2}$. Donc si $N = 2n+1$ est impair cela n'arrive pas : vous vérifierez que $0 < t_0 < u_1 < t_1 < u_2 < \cdots < u_n < t_n < \pi < t_{n+1} < u_{n+1} < \cdots < u_{N-1} < t_N < 2\pi$. Si $N = 2n$ on a à la place $0 < t_0 < u_1 < t_1 < \cdots < t_{n-1} < u_n = t_n = \pi < t_{n+1} < u_{n+1} < \cdots < u_{N-1} < t_N < 2\pi$. Dans ce cas $S_N'(x)$ a un zéro double en $x = \pi$ ce qui correspond à un point d'inflexion pour $S_N(x)$, qui traverse de manière tangente l'axe horizontal en $x = \pi$ (j'ai oublié de dire que $S_N(\pi) = 0$ pour tout N, évidemment). D'autre part $S_N'(0) = N$, donc $S_N'(x)$ est successivement positif, négatif, positif, négatif, etc. . .Donc S_N est d'abord strictement croissante, puis strictement décroissante, puis strictement croissante, etc. . .les $t_k < \pi$ sont des maxima locaux, les $u_j < \pi$ sont des minima locaux (sur l'intervalle $]\pi, 2\pi[$ c'est le contraire car $S_N(2\pi - x) = -S_N(x)$).

Supposons que $S_N(x) < 0$ soit possible sur $]0,\pi[$. Alors prenons x_0 tel que S_N y atteigne son minimum absolu (il peut y en avoir plusieurs, on en choisit un). Ce x_0 est un minimum local donc de la forme $u_j = 2j\frac{\pi}{N}$. Mais alors, miracle, $\sin(Nu_j) = 0$, donc $S_N(u_j) = S_{N-1}(u_j)$. Donc on peut obtenir une contradiction en faisant l'hypothèse de récurrence que ça marche pour S_{N-1} ! (pour $N = 1$, on a $S_1(x) = \sin(x)$; pour $N = 2$, on a $S_2(x) = \sin(x) + \sin(x)\cos(x) = \sin(x) 2\cos^2(\frac{x}{2})$). Donc en tout cas $S_N(x) \geq 0$ pour $0 < x < \pi$. Si la valeur 0 est atteinte (une ou plusieurs fois), alors à nouveau c'est en un minimum local donc de la forme

u_j et donc encore une fois $S_N(u_j) = S_{N-1}(u_j) > 0$ par hypothèse de récurrence...contradiction. Conclusion : on a prouvé :

$$0 < x < \pi \Longrightarrow \quad 0 < S_N(x)$$

Venons-en à l'analyse du phénomène de Gibbs : le $(k+1)^{\text{ème}}$ maxima local est en t_k $(0 \leq 2k < N)$, la valeur de S_N en ce point est :

$$S_N((2k+1)\frac{\pi}{N+1}) = \sum_{1 \leq j \leq N} \frac{\sin(j\frac{(2k+1)\pi}{N+1})}{j} = \sum_{1 \leq j \leq N+1} \frac{(2k+1)\pi}{N+1}\psi(j\frac{(2k+1)\pi}{N+1})$$

où l'on a machiavéliquement introduit la fonction $\psi(x) = \frac{\sin(x)}{x}$ (le point supplémentaire $j = N+1$ ne change rien à la somme). Il s'agit donc exactement d'une somme de Riemann pour cette fonction, sur l'intervalle $[0, (2k+1)\pi]$. Donc :

$$\forall k \qquad \lim S_N(t_k) = C_k = \int_0^{(2k+1)\pi} \frac{\sin(x)}{x} dx$$

En particulier le premier maximum de S_N tend (en croissant avec N d'ailleurs, car $t_0^{(N+1)} < t_0^{(N)} < u_1^{(N+1)}$, donc $S_{N+1}(t_0^{(N+1)}) > S_{N+1}(t_0^{(N)}) = S_N(t_0^{(N)})$), pour $N \to \infty$, vers

$$C_0 = \int_0^{\pi} \frac{\sin(x)}{x} dx \approx 1,8519$$

Les deuxièmes, troisièmes, etc..., maxima locaux de S_N tendent vers $C_1 = \int_0^{3\pi} \frac{\sin(x)}{x} dx$, $C_2 = \int_0^{5\pi} \frac{\sin(x)}{x} dx$, etc.... On a $1,8519 \approx C_0 > C_1 > C_2 > \ldots$ avec pour limite $\frac{\pi}{2}$.

De même les minima locaux successifs tendent, pour $N \to \infty$ vers les valeurs $D_j = \int_0^{2j\pi} \frac{\sin(x)}{x} dx$ ($D_1 = \int_0^{2\pi} \frac{\sin(x)}{x} dx \approx 1.4181$). Il est facile de voir que $D_1 < D_2 < D_3 < \ldots$ avec pour limite $\int_0^{\infty} \frac{\sin(x)}{x} dx = \frac{\pi}{2} \approx 1.5708$.

Il nous reste à montrer que S_N atteint en $t_0 = \frac{\pi}{N+1}$ non seulement un maximum local, mais même son maximum global.

Notons $M_k = S_N(t_k)$. Nous allons montrer que la suite M_k, $0 \leq k \leq \frac{N}{2}$, est (strictement) décroissante (on se limite donc aux $t_k \leq \pi$). On a :

$$M_k - M_{k+1} = -\int_{t_k}^{t_{k+1}} S_N'(x) dx$$

$$S_N'(x) = \frac{D_N(x) - 1}{2} = \frac{1}{2}\frac{\sin((N+\frac{1}{2})x)}{\sin(\frac{x}{2})} - \frac{1}{2} = -\frac{1}{2} + \frac{1}{2}\frac{\sin((N+1)x)\cos(\frac{x}{2}) - \cos((N+1)x)\sin(\frac{x}{2})}{\sin(\frac{x}{2})}$$
$$= -\frac{1}{2} - \frac{1}{2}\cos((N+1)x) + \frac{1}{2}\sin((N+1)x)\,\mathrm{cotg}(\frac{x}{2})$$

d'où il résulte, puisque l'intervalle d'intégration est de longueur $\frac{2\pi}{N+1}$:

$$M_k - M_{k+1} = +\frac{\pi}{N+1} - \frac{1}{2}\int_{t_k}^{t_{k+1}} \sin((N+1)x)\,\mathrm{cotg}(\frac{x}{2}) dx$$

On fait le changement de variable $x = t_k + \theta \frac{1}{N+1} = \frac{2k\pi+\pi+\theta}{N+1}$ de sorte que $\sin((N+1)x) = -\sin\theta$ et donc :

$$M_k - M_{k+1} = \frac{\pi}{N+1} + \frac{1}{2}\int_0^{2\pi} \sin(\theta)\,\mathrm{cotg}(\frac{x(\theta)}{2})\frac{d\theta}{N+1}$$

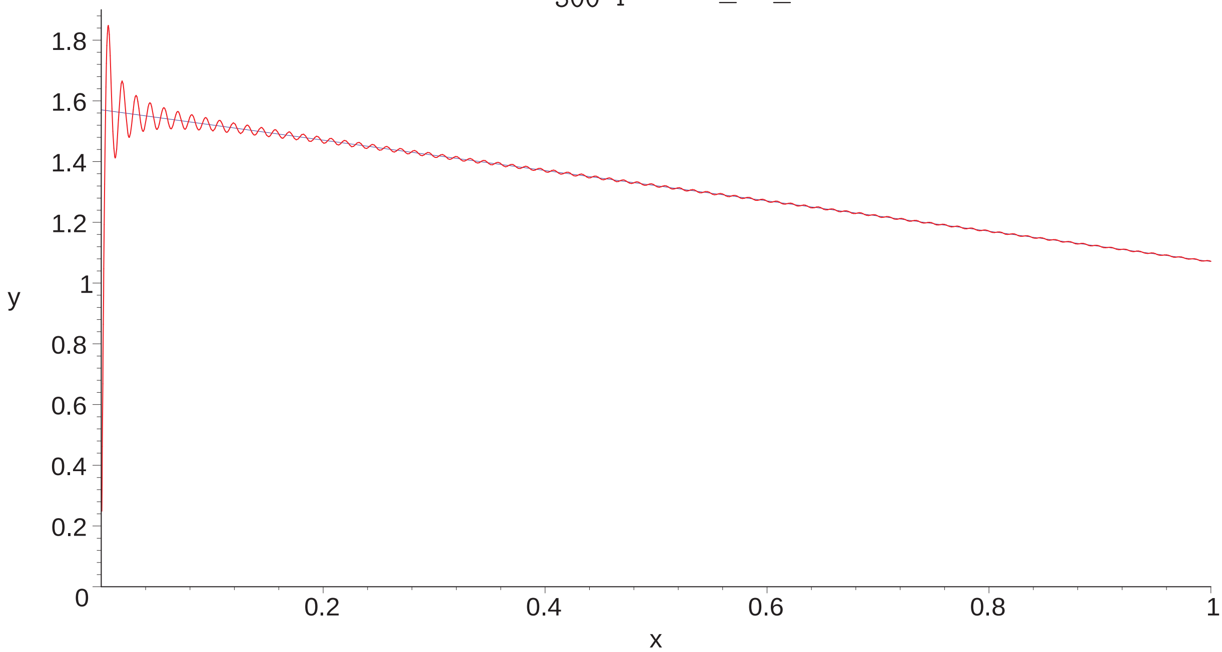
La fonction continue $\mathrm{cotg}(\frac{x}{2})$ est strictement décroissante sur $]0,\pi]$. En écrivant l'intégrale sous la forme :

$$\frac{1}{2}\int_0^{\pi} \sin(\theta)\left(\mathrm{cotg}(\frac{x(\theta)}{2}) - \mathrm{cotg}(\frac{x(\theta+\pi)}{2})\right)\frac{d\theta}{N+1}$$

on voit alors que : $M_k - M_{k+1} > \frac{\pi}{N+1}$, pour $2(k+1) \leq N$. En particulier, le premier maximum en $x = t_0$ est bien le maximum absolu de la fonction S_N sur $\mathbb{R}$ (qui est donc partout strictement inférieur à $\int_0^{\pi} \frac{\sin(x)}{x}dx$).

En ce qui concerne les minima, le graphe pour N = 60 montre qu'ils ne forment pas (sauf pour, disons, $N \lesssim 30$) une suite décroissante. Cependant ils forment une suite (finie) concave. Je me réserve la démonstration au cas où j'ai besoin de publier un dernier article mathématique avant ma mise à la retraite. De même les maxima forment en fait une suite convexe (tout cela se rapporte uniquement à l'intervalle $]0,\pi]$, tout est inversé dans $[\pi, 2\pi[$).

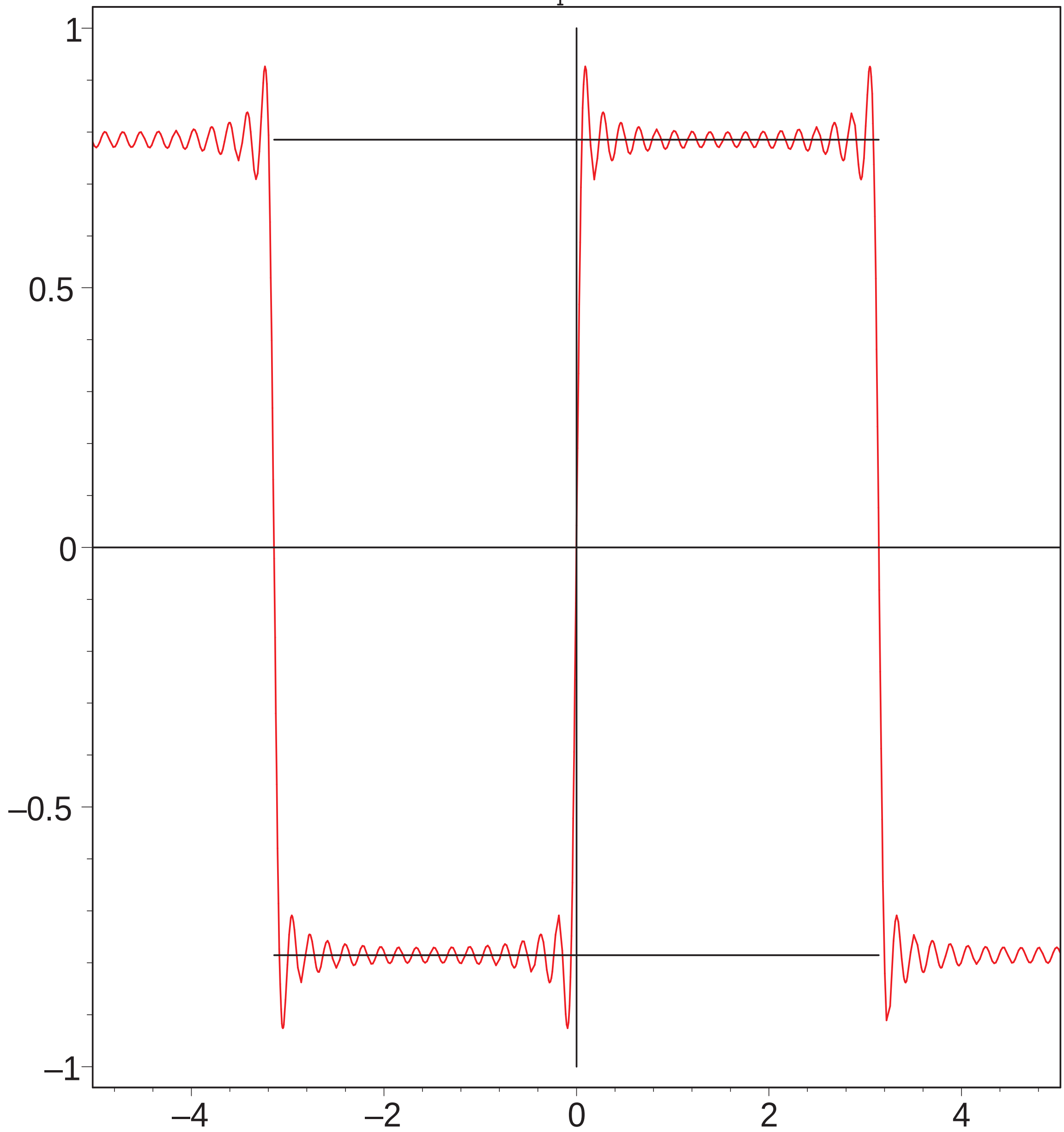
Fig. 6 - S_{500} pour $0 \leq x \leq 1$



Il n'est pas difficile de montrer que le phénomène de Gibbs vaut pour toute fonction F (raisonnable) avec un saut : après translation on peut supposer que la discontinuité est en 0 ; on soustrait de F un multiple adapté de la fonction que nous venons d'étudier en détail, pour éliminer le saut, et si le résultat est continu et C^1 par morceaux dans un voisinage de l'origine, la différence des séries de Fourier sera uniformément convergente sur un voisinage de 0, et donc cela permet de voir que les sommes partielles de la série de Fourier de F présentent aussi le phénomène de Gibbs. Détails laissés au lecteur.

À titre d'exemple, la série de Fourier correspondant au dernier graphe de cette feuille est $\sin(x) + \frac{\sin(3x)}{3} + \frac{\sin(5x)}{5} + \frac{\sin(7x)}{7} + \ldots$. Elle converge vers $\frac{\pi}{4}$ pour $0 < x < \pi$ et vers $-\frac{\pi}{4}$ pour $-\pi < x < 0$. La figure représente la somme partielle jusqu'au terme $\frac{\sin(33x)}{33}$ inclus. Ici encore on constate l'omniprésence du << phénomène >>. [obsolète : à noter que visiblement le logiciel qui a produit le graphe montre ses limites, mais j'ai eu beau essayer d'ajuster le nombre de points, la précision, etc..., je n'ai pas réussi à obtenir quelque chose de plus réaliste.] [Post-Scriptum (avril 2007) : j'ai enfin compris que je devais utiliser la v8 et non pas la v9, totalement défectueuse pour les graphiques, du logiciel que j'avais utilisé pour tracer ces graphes.]

Fig. 7 - Phénomène de Gibbs pour une fonction créneau



Les Théorèmes de Cantor-Lebesgue et Denjoy-Lusin

Si $\lim a_n = \lim b_n = 0$ alors évidemment on a $a_n \cos(nx) + b_n \sin(nx) \to 0$ pour tout x. La réciproque vaut-elle ? oui, et ce théorème est dû à Cantor (on ne suppose pas que les a_n et b_n soient des coefficients de Fourier d'une fonction). Mieux encore il suffit d'après Lebesgue de supposer $\lim a_n \cos(nx) + b_n \sin(nx) = 0$ pour les x d'un ensemble de mesure non-nulle. Par ailleurs si $\sum_{n \geq 1} |a_n \cos(nx) + b_n \sin(nx)| < \infty$ sur un ensemble de mesure > 0 alors $\sum_{n=1}^{\infty} |a_n| + |b_n| < \infty$. C'est le théorème de Denjoy-Lusin.

Cantor-Lebesgue

On suppose $a_n \cos(nx) + b_n \sin(nx) \to 0$ pour tout x. En prenant x = 0 on a immédiatement $a_n \to 0$. Compte tenu de cela on peut supposer $\forall x\ b_n \sin(nx) \to 0$. Pour la démonstration de $b_n \to 0$, raisonnons par l'absurde. Si l'on n'a pas $b_n \to 0$ c'est qu'il existe $\epsilon > 0$ et une suite extraite $c_k = b_{n_k}$ avec $|c_k| \geq \epsilon$ pour tout k. Pour des raisons qui vont devenir claires dans une seconde, on peut quitte à passer encore à une suite extraite $d_l = c_{k_l} = b_{n_{k_l}} = b_{m_l}$, supposer $m_{l+1} \geq 3m_l$. En effet il suffit de définir k_{l+1} (et donc $m_{l+1} = n_{k_{l+1}}$) par récurrence comme étant le premier $k > k_l$ tel que $n_k \geq 3n_{k_l}$.

Cela dit supposons construits des intervalles fermés emboîtés $I_1 \supset I_2 \supset \cdots \supset I_l$ dans $[0, \pi]$ tels que I_j pour $1 \leq j \leq l$ soit de longueur $\frac{2\pi}{3m_j}$ et avec $|\sin(m_j x)| \geq \frac{1}{2}$ sur I_j (je vous laisse la définition du premier d'entre eux). Dans l'intervalle I_l il y a au moins deux points consécutifs en lesquels $|\sin(m_{l+1}x)|$ s'annule : en effet si il y en avait au plus un, I_l serait de longueur strictement inférieure à $2\frac{\pi}{m_{l+1}}$. Mais on a $m_{l+1} \geq 3m_l$ par construction et la longueur de I_l est exactement $\frac{2\pi}{3m_l}$. Contradiction. Donc la fonction $|\sin(m_{l+1}x)|$ a dans I_l un intervalle entier allant d'un zéro au suivant. On définit I_{l+1} comme étant le sous-intervalle fermé de cet intervalle sur lequel $|\sin(m_{l+1}x)| \geq \frac{1}{2}$. Il est de longueur $\frac{2\pi}{3m_{l+1}}$. Le principe d'induction donne ainsi les I_l pour tout $l \geq 1$.

Les extrémités des I_l forment une paire de deux suites adjacentes, donc l'intersection de tous les I_l est non vide, et en fait égale au singleton $\{x_0\}$ avec x_0 la limite commune de ces deux suites. On peut enfin conclure : comme $\forall l\ |\sin(m_l x_0)| \geq \frac{1}{2}$, de $b_{m_l} \sin(m_l x_0) \to 0$ on déduit $b_{m_l} \to 0$. Or par construction on a $|b_{m_l}| \geq \epsilon$ pour tout l ce qui apporte la contradiction finale.

Remarquons qu'il y a en particulier dans cette preuve le principe intéressant suivant : pour montrer qu'une suite converge vers L il suffit de montrer que toute suite-extraite a une sous-suite extraite qui converge vers L. Lorsque l'on lit ce genre de choses, on n'est pas trop surpris d'apprendre que Cantor a eu de graves problèmes d'équilibre psychique par la suite...

La démonstration ci-dessus est d'une nature topologique (intervalles, compacité). De plus elle utilise la continuité de la fonction sinus, ce qui en fait n'est nullement nécessaire pour garantir la conclusion, comme nous allons le voir maintenant grâce aux outils et concepts de la mesure des ensembles. En effet je vais démontrer l'extension due à Lebesgue du théorème de Cantor : *si l'ensemble des* $x \in [0, 2\pi]$ *avec* $a_n \cos(nx) + b_n \sin(nx) \dashrightarrow 0$ *est de mesure* > 0 *alors c'est que* $\lim a_n = \lim b_n = 0$.

On ne peut plus faire la première réduction, mais écrivons $a_n \cos(nx) + b_n \sin(nx) = \delta_n \sin(n(x - x_n))$ avec $\delta_n^2 = a_n^2 + b_n^2$. Soit $\epsilon > 0$ et considérons l'ensemble $X_N(\epsilon)$ des $x \in [0, 2\pi]$ tels que $|\delta_n \sin(n(x - x_n))| \leq \epsilon$ pour tout $n \geq N$. D'une part $X_1(\epsilon) \subset X_2(\epsilon) \subset \ldots$ et d'autre part $\bigcup X_N(\epsilon)$ contient l'ensemble des x tels que $a_n \cos(nx) + b_n \sin(nx) \dashrightarrow 0$. Il existe donc en tout cas l'un des X_N qui soit de mesure > 0. On a $\forall n \geq N$ $\int_{X_N} |\delta_n \sin(n(x - x_n))| dx \leq \epsilon |X_N(\epsilon)|$. De plus on a certainement $\int_{X_N} |\sin(n(x - x_n))| dx \geq \int_{X_N} \sin^2(n(x - x_n)) dx = \frac{1}{2}|X_N| - \frac{1}{2}\int_{[0,2\pi]} (\cos(2nx_n)\cos(2nx) + \sin(2nx_n)\sin(2nx)) \mathbb{1}_{X_N}(x) dx$. Par le Théorème de Riemann-Lebesgue la deuxième intégrale tend vers 0 pour $n \dashrightarrow \infty$. Compte tenu de $|X_N(\epsilon)| > 0$ on en déduit que pour $n \gg 1$ on a $\int_{X_N} |\sin(n(x - x_n))| dx \geq \frac{1}{4}|X_N|$ et donc $|\delta_n| \leq 4\epsilon$. Ainsi $\lim\ \delta_n = 0$, d'où $\lim\ a_n = 0$ et $\lim\ b_n = 0$.

Je trouve cette preuve plus simple que la précédente (en principe comme en exécution), et de plus sous une hypothèse bien plus faible on a obtenu la même conclusion. Je signale que le théorème vaut avec n'importe quelle fonction positive g bornée (périodique, non presque partout nulle) : si $\lim b_n g(nx) = 0$ sur un ensemble de mesure > 0 alors $\lim\ b_n = 0$. La preuve donnée ici marche presque telle quelle, mais il faut raisonner autrement au niveau de l'astuce où l'on est passé de $|\sin|$ à $\sin^2$.

Le Théorème de Denjoy-Lusin

Théorème : *si l'ensemble des* $x \in [0, 2\pi]$ *tels que* $\sum_{n \geq 1} (a_n \cos(nx) + b_n \sin(nx))$ *est absolument convergente a une mesure non nulle alors* $\sum_{n \geq 1} |a_n| + |b_n| < \infty$.

Preuve : on écrit encore $a_n \cos(nx) + b_n \sin(nx) = \delta_n \sin(n(x - x_n))$. Pour $N \gg 1$ l'ensemble (dont je vous laisse le soin d'établir la mesurabilité) $X_N = \{x \in [0, 2\pi] \mid \sum_{n \geq N} |\delta_n \sin(n(x - x_n))| \leq 1\}$ sera de mesure > 0 (même technique). Par le théorème de la convergence monotone : $|X_N| \geq \int_{X_N} \sum_{n \geq N} |\delta_n \sin(n(x - x_n))| dx = \sum_{n \geq N} |\delta_n| \int_{X_N} |\sin(n(x - x_n))| dx$. Or (même technique), pour $n \gg 1$ on a $\int_{X_N} |\sin(n(x - x_n))| dx \geq \frac{1}{4}|X_N|$. La série $\sum |\delta_n|$ est ainsi convergente ce qu'il fallait prouver.

Remarque : ici encore, pour toute fonction $g \geq 0$, π-périodique, telle que $0 < \int_0^\pi g(x) dx < \infty$, si $\sum_{n=1}^\infty |b_n\ g(nx)| < \infty$ sur un ensemble de mesure > 0 alors $\sum |b_n| < \infty$ (et ceci implique $\sum_{n=1}^\infty |b_n\ g(nx)| < \infty$ pour presque tout x et même tout x si g est bornée).

```
-----------------------------------------------------------------------
|        U n i v e r s i t é  L i l l e  1  -  U F R  d e  M a t h é m a t i q u e s        |
|        L i c e n c e  S e m e s t r e  6  -  2 0 0 5 / 2 0 0 6  -  M A T H  3 1 2        |
-----------------------------------------------------------------------
```

F a m i l l e s S o m m a b l e s

version du 16 mai 2006.

La notion de famille sommable est utile, par exemple pour définir ce qu'est une base orthonormée non nécessairement dénombrable, ou pour être à l'aise avec des sommes avec des multi-indices, ou pour impressionner. Bien qu'on puisse considérer cette notion comme englobée par le formalisme général des espaces mesurés, il y a tout de même quelques petites choses qu'il est utile de connaître (pour le niveau M). Comme j'avais rédigé cela un jour de folie de l'année dernière, je le reprends à peu près tel quel.

On a un ensemble d'indices Λ et une fontion $\Lambda \dashrightarrow \mathbb{C}$, que l'on notera $\lambda \dashrightarrow u_\lambda$ et l'on se demande si l'on peut donner un sens à $\sum_{\lambda\in\Lambda} u_\lambda$. Le point crucial est que la valeur de la somme doit correspondre à la notation : donc il ne faut pas que le résultat dépende d'une énumération de Λ par $\mathbb{N}$ (si Λ est infini dénombrable).

Le support de $\lambda \dashrightarrow u_\lambda$ est l'ensemble $\Lambda_0 = \{\lambda \mid u_\lambda \neq 0\}$. Si ce support est de cardinalité finie, alors << pas de problème >> (sauf qu'il faudrait en fait énumérer Λ_0, et utiliser la commutativité et l'associativité de l'addition pour prouver que cela marche. On va dire que ce travail a été fait au collège, ou au lycée, ou en premier cycle, ou dans votre tête à un moment ou un autre,...). Plus généralement pour toute partie finie $A \subset \Lambda$, on a donc << évidemment >> une somme $S(A) = \sum_{\lambda\in A} u_\lambda$. C'est notre point de départ et on veut s'affranchir de l'hypothèse que l'ensemble des indices est fini.

Supposons dans un premier temps $\forall\lambda\ u_\lambda \geq 0$. Il n'y a pas le choix : on conviendra que $S = S(\Lambda) = \sum_{\lambda\in\Lambda} u_\lambda$ est la borne supérieure de toutes les sommes $S(A)$ pour $A \subset \Lambda$ de cardinalité finie.(*) On aura $S \in [0, +\infty]$. Supposons $S < \infty$: il y a au plus SN éléments λ tels que $u_\lambda \geq \frac{1}{N}$. Donc le support Λ_0 est au plus dénombrable (fini ou infini).

Il faudrait maintenant prouver (en se limitant pour l'instant à ces sommes de termes positifs) des choses comme $\sum_\Lambda (u_\lambda + v_\lambda) = \sum_\Lambda u_\lambda + \sum_\Lambda v_\lambda$, $\sum_{\Lambda_1\cup\Lambda_2} u_\lambda + \sum_{\Lambda_1\cap\Lambda_2} u_\lambda = \sum_{\Lambda_1} u_\lambda + \sum_{\Lambda_2} u_\lambda$, $\sum_{\Lambda_1\times\Lambda_2} u_{\lambda_1} u_{\lambda_2} = (\sum_{\Lambda_1} u_{\lambda_1})(\sum_{\Lambda_2} u_{\lambda_2})$(**) et mieux encore $\sum_{\Lambda_1\times\Lambda_2} u_{\lambda_1,\lambda_2} = \sum_{\Lambda_1}\left(\sum_{\Lambda_2} u_{\lambda_1,\lambda_2}\right)$.(***) Toutes ces choses sont des cas particuliers des théorèmes de la théorie de l'intégrale de Lebesgue, spécialisés à la mesure de dénombrement $\mu(A) = \mathrm{card}(A)$. Évidemment on peut les

(*) à ce propos je rappelle la convention indispensable : $S(\emptyset) = 0$.

(**) pourquoi FAUT-il définir $0 \cdot \infty = 0$ dans ce contexte ?

(***) donc, il faut autoriser $u_\lambda = +\infty$ dès le début.

établir directement, je délègue à votre enthousiasme cette tâche.

Prenons maintenant les u_λ complexes. Notons $S^*(A) = \sum_{\lambda\in A}|u_\lambda|$ et supposons $S^* = S^*(\Lambda) = \sum_{\lambda\in\Lambda}|u_\lambda| < \infty$. Si c'est le cas, on dit que la famille $(u_\lambda)_{\lambda\in\Lambda}$ (à valeurs réelles ou complexes) est absolument sommable.

Supposons que cela soit le cas et soit $\epsilon > 0$ et soit A choisi de sorte que $S^* - \epsilon \leq S^*(A) \leq S^*$. Soit $B \supset A$ une partie finie quelconque contenant A. Alors on voit (comment ?) $|S(B) - S(A)| \leq S^*(B) - S^*(A)$ et $0 \leq S^*(B) - S^*(A) \leq \epsilon$, donc $|S(B) - S(A)| \leq \epsilon$.

Prenons maintenant $\epsilon = \frac{1}{2^n}$ et choisissons A_n tel que $S^* - \frac{1}{2^n} \leq S^*(A_n) \leq S^*$. Quitte à remplacer A_n par $A_1 \cup \cdots \cup A_n$ nous pouvons supposer $A_n \subset A_{n+1}$. La suite de nombres réels $S(A_n)$ vérifie le critère de Cauchy puisque $|S(A_n) - S(A_m)| \leq \frac{1}{2^n}$ pour $m \geq n$ ($A_n \subset A_m$). Donc il existe $S = \lim S(A_n)$. Et en passant à la limite on a $|S(A_n) - S| \leq \frac{1}{2^n}$. Soit B une partie finie quelconque contenant A_n. Alors $|S(B) - S(A_n)| \leq \frac{1}{2^n}$ et $|S(A_n) - S| \leq \frac{1}{2^n}$ et donc $|S(B) - S| \leq \frac{2}{2^n}$. Ceci motive la définition suivante :

On dira que la famille de nombres complexes $(u_\lambda)_{\lambda\in\Lambda}$ *est sommable (on parle aussi de << convergence en vrac >>) de somme* S *si pour tout* $\epsilon > 0$ *il existe une partie finie* A *telle que pour toute partie finie* $B \supset A$ *on a* $|S(B) - S| \leq \epsilon$.

On vient d'établir qu'une famille absolument sommable est sommable.

Pour une famille sommable on a unicité de S : si un autre S' marche, avec A' pour un ϵ donné, on prend $B = A \cup A'$ car il vérifie à la fois $|S(B) - S| \leq \epsilon$ et $|S(B) - S'| \leq \epsilon$ et on en déduit $|S - S'| \leq 2\epsilon$, et ceci pour tout $\epsilon > 0$. Donc $S = S'$.

Le point surprenant c'est qu'une famille sommable est toujours absolument sommable. Il suffit (pourquoi ?) de traiter le cas d'une famille réelle. Soit A_n pour $\epsilon = \frac{1}{2^n}$ et ici encore on peut imposer $A_n \subset A_{n+1}$. Soit $\Lambda_1 = \bigcup A_n$. Si $\lambda \notin \Lambda_1$ en écrivant le singleton $\{\lambda\} = (A_n \cup \{\lambda\}) \setminus A_n$ comme différence de deux ensembles contenant chacun A_n on obtient $|u_\lambda| \leq \frac{2}{2^n}$ pour tout n, et donc $u_\lambda = 0$. Donc Λ_1 contient le support Λ_0 (qui est ainsi au plus dénombrable). Si Λ_1 est de cardinalité finie alors le support est aussi de cardinalité finie, et en particulier la famille est absolument sommable. Supposons donc Λ_1 infini. On peut alors choisir une bijection $\phi : \mathbb{N} \dashrightarrow \Lambda_1$ de sorte que les A_n (on peut imposer $A_1 \neq \emptyset$) correspondent à une chaîne croissante $F_n \subset F_{n+1}$ dans $\mathbb{N}$ avec $F_n = \{0, \ldots, N_n\}$, $N_1 \leq N_2 \ldots\ldots$, $N_n = \mathrm{card}(A_n) - 1$, $N_n \dashrightarrow \infty$. Notons $v_k = u_{\phi(k)}$.

Montrons que la série $\sum_{k=0}^{\infty} v_k$ est convergente. Nous le prouvons via le critère de Cauchy : supposons $m > N_n$, alors pour tout $p \geq m$ on peut écrire $v_m + \cdots + v_p = S(B_p) - S(B_{m-1})$ avec la notation $B_k = \phi(\{0, \ldots, k\})$. Donc pour $N_n < m \leq p$, on a $|\sum_{m\leq k\leq p} v_k| \leq |S(B_p) - S(B_{m-1})| \leq |S(B_p) - S| + |S - S(B_{m-1})| \leq \frac{2}{2^n}$ puisque B_p et

B_{m-1} contiennent A_n. (*) Donc $\sum_{k=0}^{\infty} v_k$ est convergente.

Si $\sum_{k=0}^{\infty} v_k$ n'est pas absolument convergente alors $\sum_{k=0}^{\infty} \max(v_k, 0) = +\infty$(*) et donc $\sum_{k>N}^{\infty} \max(v_k, 0) = +\infty$ pour tout N et donc pour tout N on peut trouver $M > N$ avec $\sum_{N<k\leq M} \max(v_k, 0) \geq 1$. Prenons en particulier $N = N_2$ et définissons la partie finie B comme étant la réunion de A_2 et de ceux des λ qui correspondent aux k avec $N_2 < k \leq M$ et $v_k > 0$. Alors $S(B) - S(A_2) = \sum_{N_2<k\leq M} \max(v_k, 0) \geq 1$ ce qui est une contradiction (pourquoi ?).

La série $\sum_{k=0}^{\infty} v_k$ est donc absolument convergente. La famille $(v_k)_{k\in\mathbb{N}}$ est ainsi absolument sommable et donc aussi $(u_\lambda)_{\lambda\in\Lambda_1}$ qui lui est en bijection. Comme $u_\lambda = 0$ pour $\lambda \notin \Lambda_1$ on peut affirmer que $(u_\lambda)_{\lambda\in\Lambda}$ est absolument sommable. On a donc prouvé en partie :

<u>Théorème</u> : *Toute famille sommable $(u_\lambda)_{\lambda\in\Lambda}$ est absolument sommable et son support est au plus dénombrable. La somme $\sum_{\lambda\in\Lambda} u_\lambda$ au sens des familles sommables coïncide avec $\int_\Lambda u_\lambda d\mu(\lambda)$ au sens de la théorie de l'intégrale sur les espaces mesurés, avec ici μ la mesure de dénombrement sur la tribu maximale $\mathcal{P}(\Lambda)$ de toutes les parties de Λ.*

Pour la dernière affirmation, il suffit de traiter le cas réel. Puis, compte tenu que l'on sait qu'il y a absolue sommabilité, en écrivant $u_\lambda = \max(u_\lambda, 0) - (-\min(u_\lambda, 0))$ on se ramène par linéarité (que j'ai laissée en exercice pour les familles sommables, et qui est du cours d'intégration pour un espace mesuré général) au cas $\forall\lambda\ u_\lambda \geq 0$. Mais alors notre définition par la borne supérieure des sommes finies coïncident exactement avec la définition par la borne supérieure des intégrales de fonctions étagées, puisqu'une fonction étagée positive majorée par u n'est pas autre chose qu'un choix d'un ensemble fini A d'indices λ et de nombres réels v_λ pour $\lambda \in A$ avec $0 \leq v_\lambda \leq u_\lambda$. La somme la plus grande est pour $v_\lambda = u_\lambda$ autrement dit c'est S(A). Donc la définition via la théorie de l'intégrale redonne la définition via les bornes supérieures des S(A).

On peut donc considérer à ce stade que l'on a complètement réduit la théorie des familles sommables à celle de l'intégrale. En particulier on a Fubini-Tonelli etc...

Cependant il y a bien encore une ou deux choses intéressantes à dire. Soit à nouveau Λ un ensemble quelconque indexant des nombres complexes u_λ.

<u>Théorème</u> : *Si à chaque fois que l'on a une chaîne croissante $A_1 \subset A_2 \subset \ldots$ de parties finies de Λ les sommes $S(A_n)$ convergent vers une certaine limite (finie) alors la famille $(u_\lambda)_{\lambda\in\Lambda}$ est (absolument) sommable. La réciproque est vraie.*

Comme on n'a pas dit que $\bigcup A_n$ doit contenir tout le support Λ_0 de la famille

(*) de plus $\sum_{k=0}^{\infty} v_k = \lim \sum_{k=0}^{N_n} v_k = \lim S(A_n) = S$.

(*) pourquoi ?

les limites en question ne sont pas forcément toutes égales à $\sum_{\lambda\in\Lambda} u_\lambda$.

Montrons d'abord la réciproque. Posons $\Lambda_1 = \bigcup A_n$. La famille $(u_\lambda)_{\lambda\in\Lambda}$ est absolument sommable donc aussi $(u_\lambda)_{\lambda\in\Lambda_1}$. Soit $\epsilon > 0$ et soit $A \subset \Lambda_1$ fini tel que $\sum_{\Lambda_1\setminus A} |u_\lambda| \leq \epsilon$. Comme $A \subset \bigcup A_n$ est fini, et que les A_n forment une chaîne croissante, il existe N tel que pour $n \geq N$ on a $A \subset A_n$. Alors $|S(A_n) - S(A)| \leq \epsilon$ et donc $|S(A_n) - S(A_m)| \leq 2\epsilon$ pour $N \leq n \leq m$ et le critère de Cauchy est vérifié. La suite $S(A_n)$ est ainsi prouvée convergente.

Montrons maintenant dans le sens direct. On peut supposer la famille réelle. On raisonne par l'absurde. Si l'on avait $\sum_\Lambda |u_\lambda| = +\infty$, on aurait $\sum_\Lambda \max(u_\lambda, 0) = +\infty$, ou $\sum_\Lambda -\min(u_\lambda, 0) = +\infty$ (ou les deux). Supposons le premier cas, le deuxième se traitant de manière semblable. On peut alors construire des sommes finies parmi les $u_\lambda > 0$, arbitrairement grandes ; disons qu'elles sont indexées par des ensembles B_1, B_2, ..., on pose alors $A_1 = B_1$, $A_2 = B_1 \cup B_2$..., et par construction on a $S(A_1) \leq S(A_2) \leq \cdots \leq S(A_n) \dashrightarrow +\infty$. Le théorème est démontré.

Avec cela il pourrait sembler que l'on ait dit tout ce que l'on avait à dire. Et non ! c'est compter sans Riemann qui des décennies avant que l'on se mette à roucouler avec les impressionnantes familles sommables avait démontré la seule chose réellement spectaculaire, et cette seule chose n'est même pas incorporée à la théorie des familles sommables !

Le théorème de Riemann c'est que si la série réelle $\sum_{k=0}^\infty v_k$ est une série convergente mais pas absolument convergente(*) alors pour tout nombre réel L on peut trouver une bijection ϕ de $\mathbb{N}$ telle que $L = \sum_{k=0}^\infty v_{\phi(k)}$.(**)

Je vais vous laisser y réfléchir. La preuve de Riemann occupe dix lignes sans aucune formule mathématique si ce n'est C (pour notre L), a (pour les $v_k \geq 0$), et b (pour les $-v_k \geq 0$).

(*) je rappelle que l'on dit alors qu'elle est semi-convergente, et que la répartition entre séries divergentes, séries absolument convergentes, et séries semi-convergentes est l'un des piliers de toute l'analyse mathématique.

(**) mais pour la « plupart » des bijections la série sera divergente, en fait.