
COUP D'ŒIL SUR L'ANALYSE DE FOURIER

par

Jean-Pierre Kahane

Embrasser l'analyse de Fourier d'un coup d'œil est hors de ma portée. Je promènerai le regard sur l'histoire et sur certains des termes en usage. L'histoire est ancienne et elle est instructive. Je commencerai par Platon.

Chez Platon, les mathématiques comprennent cinq parties : les nombres, les figures planes, les solides, la musique et l'astronomie. La musique et l'astronomie sont comme deux sœurs, liées par la musique des sphères. Et le programme astronomique de Platon est de rendre compte du mouvement des astres errants, les planètes, pour le rendre conforme à l'harmonie du monde et à la dignité des dieux.

Le système de Ptolémée a réalisé ce programme, par une superposition de mouvements circulaires uniformes, le grand cycle et les épicycles. Et quoique le système de Ptolémée soit abandonné, la décomposition d'un mouvement en mouvements périodiques et l'utilisation des fonctions trigonométriques a été une constante de l'astronomie jusqu'à nos jours. Aujourd'hui l'astrophysique double et dépasse l'astronomie dans l'approche des phénomènes cosmiques. Mais la perle de l'astronomie qu'est la découverte des exoplanètes repose sur l'analyse de Fourier des spectres émis par les étoiles.

Quant à la musique, son lien aux mathématiques est multiple. Le principal est la décomposition d'un son en harmoniques et sa reconstitution comme somme d'harmoniques. C'est au sens propre l'analyse et la synthèse harmoniques. A cela s'attachent les gammes, la distinction des timbres, les cordes vibrantes et la grande controverse qui au 18^e siècle a opposé Daniel Bernoulli à d'Alembert, Euler et Lagrange sur la possibilité de représenter une fonction par une série trigonométrique. Aujourd'hui l'analyse par ondelettes complète ou supprime l'usage des fonctions trigonométriques pour tenir compte des

échelles de temps aussi bien que des fréquences, et le champ de l'analyse de Fourier s'étend par ses méthodes comme par ses usages.

Fourier, selon Riemann, est le premier à avoir compris complètement la nature des séries trigonométriques, en associant ce que j'appellerai l'analyse, les formules intégrales donnant les coefficients, et la synthèse, la représentation d'une fonction par une série trigonométrique. C'est à cause de Riemann que nous parlons aujourd'hui de séries de Fourier. En France, Fourier a été longtemps méconnu. Arago, dans son éloge de Fourier, vante grandement le savant et le politique, explique l'importance de sa théorie analytique de la chaleur, mais ne dit pas un mot des séries trigonométriques, c'est-à-dire de l'outil que Fourier a forgé pour calculer des solutions des équations intégrales de la chaleur. Fourier y accordait une grande importance et, contrairement à une idée reçue, il a développé cet outil en véritable théorie, exemples, applications et démonstrations, avec une grande rigueur. Mais il s'est heurté à l'incompréhension persistante de Lagrange, et deux pages dans les manuscrits de Lagrange, que j'ai consultées et commentées, confirment que Fourier avait raison contre Lagrange. Mais Lagrange était à l'époque de Fourier le plus respecté des mathématiciens français, et son jugement négatif sur Fourier a traversé les siècles. Dans l'édition que je possède d'Encyclopedia universalis il n'y a pas d'article sur Fourier.

Fourier attachait une portée universelle à ses formules. Il précisait bien, et le premier, que la donnée d'une fonction était celle de son domaine de définition en même temps que d'une loi ou une figure ; et, sans utiliser ces termes, il distinguait soigneusement les intervalles ouverts et les intervalles fermés. Mais, après avoir multiplié les exemples et donné quelques preuves, il s'était aventuré à dire que toute fonction était sujette à cette analyse et représentable par une série trigonométrique qui converge vers la fonction. Littéralement c'est faux. Pour appliquer les formules intégrales, il faut que la fonction soit intégrable ; et la convergence des séries est un sujet difficile . Les premiers pas pour éclaircir la question sont dus à Dirichlet, avec le premier théorème général de convergence et le premier exemple de fonction non intégrable dans la conception de l'époque. Tous les concepts d'intégration, à commencer par l'intégrale de Riemann, sont liés aux formules de Fourier et à leurs conditions de validité. Au delà de Riemann, on pense à Lebesgue, Denjoy, Laurent Schwartz. Quant à la convergence, elle est mise en question par les contre-exemples : une fonction continue dont la série de Fourier diverge en un point (du Bois-Reymond) ou même sur un ensemble donné de mesure nulle (Kahane-Katznelson), et une

fonction intégrable au sens de Lebesgue dont les série de Fourier diverge partout (Kolmogorov). Le théorème de convergence de Carleson, inattendu à son époque (1966), dit que la convergence a lieu presque partout quand la fonction est de carré intégrable ; on peut améliorer cette condition, sans parvenir bien sûr aux fonctions intégrables.

De même que le concept d'intégrale, c'est le concept de série qui est en cause. Pourquoi s'attacher à la convergence, qui a d'ailleurs plusieurs significations dès qu'on passe à plusieurs variables et à des séries multiples, alors qu'il y a des procédés de sommation utilisables ? Le tournant est pris avec le théorème de Féjer de 1900 : pour les fonctions continues, les moyennes arithmétiques des sommes partielles convergent. Elles convergent même uniformément. La convergence dans les espaces fonctionnels apparaît peu après ; les fonctions sont des points, les sommes partielles d'autres points, et l'espace des fonctions de carré intégrable s'impose à l'attention avec la formule de Parseval, établie dans ce cadre par Fatou, et surtout le théorème de Riesz-Fischer, qui établit l'isomorphisme isométrique de L^2 et de ℓ^2 par transformation de Fourier.

Fischer et Frédéric Riesz, indépendamment, ont pensé à une géométrisation des espaces de fonctions, et le lemme fondamental pour la preuve de leur théorème est le même ; nous l'exprimons aujourd'hui en disant que L^2 est complet. Mais cette formulation ne date que de Banach dans sa théorie des opérations linéaires de 1930. Il fallait jusque là une longue phrase pour le dire. C'est un exemple où les définitions, tardives, viennent exprimer le suc d'une méthode, avant de servir de base à de nouveaux développements.

Autre exemple, toujours tiré de l'analyse de Fourier. Au lieu de L^2 et ℓ^2 , Norbert Wiener s'est attaché à L^1 et ℓ^1 . Leurs transformées de Fourier sont un champ d'étude toujours ouvert, avec des applications surprenantes en théorie du signal (Donoho, Candès etc). Les problèmes d'analyse et de synthèse y prennent un aspect différent, plus algébrique : dans l'algèbre des fonctions sommes de séries trigonométriques absolument convergentes, les fonctions nulles sur un ensemble donné forment un idéal fermé ; y en a-t-il d'autres ? C'est bien la cas, comme Malliavin l'a montré en 1959. Le point de départ est l'algèbre de Wiener, qui se voit soit comme l'algèbre multiplicative des fonctions sommes de séries trigonométriques absolument convergentes, soit comme algèbre de convolution ℓ^1 . La dernière expression est plus rapide, et la notion de convolution, si fondamentale en analyse, la rend très parlante. Mais en 1930, et même quand Laurent Schwartz a élaboré sa théorie des distributions, on ne parlait pas encore de convolution ; c'était *Faltung* en allemand, produit de composition en français. Dans le traité de Widder « Laplace transforms », qui

date de 1941, c'est comme « Stieltjes transforms » que sont introduites les convolutions de mesures, avec une note en bas de page qui indique le terme de convolution comme une timide nouveauté parallèlement au terme de Faltung, aussi utilisé en anglais. La convolution apparaît partout, mais c'est Wiener qui a dégagé son caractère fondamental, et c'est pourquoi Faltung est devenu pour un temps l'expression de la notion. Aujourd'hui il est raisonnable de placer la convolution au départ d'un cours d'analyse de Fourier.

S'agissant du vocabulaire, comment situer l'analyse harmonique? C'est un champ largement ouvert sur l'ensemble des mathématiques. Historiquement, on vient de voir ses relations avec les équations différentielles de la mécanique céleste, avec les équations aux dérivées partielles des cordes vibrantes et de la chaleur, avec la théorie des fonctions d'une variable réelle, et on a évoqué ses relations actuelles avec la statistique et le traitement des données. Le lien aux probabilités est ancien et profond. Le cadre de l'analyse harmonique commutative est celui des groupes abéliens localement compacts, et la dualité entre ces groupes est exprimée par la transformation de Fourier. La thèse de Tate a montré l'importance de cette approche en théorie des nombres. L'analyse harmonique non commutative est liée aux groupes non commutatifs et à leurs représentations. L'analyse harmonique abstraite part de la théorie de Gelfand des anneaux normés, autrement dits algèbres de Banach. L'analyse de Fourier classique traite de transformations intégrales, transformation de Fourier d'abord, et aussi intégrales singulières. Ses usages dans toutes les sciences ont été multipliés par la transformation de Fourier rapide puis par les ondelettes et leurs variantes. Il est clair que toute définition de l'analyse harmonique en serait une limitation injustifiée. Mais outre son étendue, on peut repérer des questions, des méthodes, des théories qui, elles, peuvent être identifiées et formalisées.

Les formules de Fourier sont l'exemple de base. Leur énoncé par Fourier exprime leur généralité, de façon formellement incorrecte. Elles ne constituent pas un théorème, mais elles ont engendré des théorèmes et permis de définir d'importantes notions. Mieux qu'un théorème, elles ont constitué un programme, à savoir, donner des conditions de leur validité. Elles ont ensuite constitué un paradigme pour tous les développements orthogonaux. Une raison de leur succès est sans doute qu'elles établissent un pont entre deux classes d'objets, en l'occurrence des fonctions et des suites. La dualité de Fourier est un modèle de traduction d'un langage dans un autre, un trait commun à d'autres grands programmes.

Le coup d'œil pourrait se poursuivre, aussi bien sur l'histoire que sur l'actualité de l'analyse de Fourier. Ce sera pour une part l'objet des exposés oraux.

Jean-Pierre Kahane 20.04.2011

J.-P. KAHANE, Laboratoire de Mathématique d'Orsay, UMR 8628, Bâtiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex

E-mail : `Jean-Pierre.Kahane@math.u-psud.fr`

STATISTIQUE ET FOURIER, DANS L'HISTOIRE ET AUJOURD'HUI

par

Jean-Pierre Kahane

Résumé. Ce sera une promenade historique, dans laquelle je pense m'attarder sur Fourier, mais qui aura comme point de départ la méthode des moindres carrés et comme point d'arrivée l'échantillonnage parcimonieux, alias *compressed sensing*.

Statistique et Fourier sont deux sujets familiers dans les mathématiques de la biologie. Je ne compte pas vous apprendre grand chose sur les aspects actuels qui vous intéressent directement. C'est plutôt moi qui devrais m'en instruire auprès de vous. Cependant j'ai près de 60 ans d'expérience de recherche en analyse de Fourier et en probabilités, et je me suis instruit en statistique plus récemment, dans les années 80, pour en enseigner des rudiments aux étudiants de première année en biologie à Orsay. Depuis une dizaine d'années, je vois la statistique faire émerger de beaux problèmes et de beaux résultats sur les séries de Fourier, ma spécialité principale. Et depuis longtemps je m'intéresse à l'histoire des sciences en général, et en particulier à celle des mathématiques. Comme je vous l'ai annoncé je compte donc faire avec vous une promenade historique ; le personnage principal sera Fourier, mais je parlerai aussi de ses aînés, Legendre et Laplace, et de ses successeurs, parmi lesquels Riemann et Lebesgue. Dans la foison des théories et des résultats du 20^e siècle j'insisterai sur deux aspects : ce qui concerne ℓ^2 , l'espace des suites de carré sommable, qui a été mis en évidence par Frédéric Riesz et Ernst Fischer, et ce qui concerne ℓ^1 , l'espace des suites sommables, qui a été découvert par Norbert Wiener. La statistique sera l'occasion de voir ℓ^2 et ℓ^1 s'introduire naturellement, avec les moindres carrés de Legendre et compagnie, et le *compressed sensing* d'Emmanuel Candès, David Donoho et compagnie.

Trame d'un exposé fait à l'Institut Jacques Monod pour l'équipe Pakdaman le 10 mars 2011.

Pour ne pas risquer de manquer de temps en m'attardant en route, je commence par la fin du voyage : le théorème de Candès, Romberg et Tao de 2004 sur la reconstitution d'un signal ou d'une image à partir de données apparemment insuffisantes. En théorie du signal, on enseigne que pour reconstituer une fonction dont la transformée de Fourier est portée par une bande de largeur L on a besoin d'un échantillonnage sur une progression arithmétique de raison a/L , a dépendant de la normalisation. En limitant la fonction à un intervalle de longueur H , on a donc besoin de HL/a points. Or voici le théorème :

On sait qu'un signal est constitué de T pics, sans connaître leur position, et on a accès à sa transformée de Fourier. Il s'agit de reconstituer le signal à partir de la transformée de Fourier de la manière la plus économique. On identifie le signal à une fonction f portée par T points du groupe cyclique à N points, N choisi assez grand. Ce groupe cyclique \mathbb{Z}_N que je note G , est isomorphe à son dual \hat{G} , le groupe des fréquences, sur lequel est définie la transformée de Fourier \hat{f} . On choisit au hasard avec la probabilité naturelle W fréquences w et on considère les $\hat{f}(w)$. Supposons

$$T < CW/\log N \quad \text{avec} \quad C = 1/23(M + 1).$$

Alors on peut reconstituer f à partir des $\hat{f}(w)$ avec une probabilité $1 - c/N^M$, où c est une constante absolue. Et la reconstitution se fait par un procédé d'analyse convexe, à savoir en minimisant la norme ℓ^1 des fonctions définies sur G dont la transformée de Fourier coïncide avec \hat{f} sur les W fréquences choisies. On voit l'économie considérable : l'utilisation classique de la transformée de Fourier nécessite N points, et l'échantillonnage aléatoire le réduit à un multiple de $T \log N$.

Rappelons que la norme ℓ^1 d'une fonction h définie sur G est la somme des valeurs absolues des $h(g)$, g parcourant G . C'est bien une fonction convexe de h . Il y a aujourd'hui des procédés variés pour traiter de tels problèmes de minimisation, et je dirai un mot de leur histoire.

Remontons à 200 ans en arrière, plus précisément en 1805. Legendre a écrit un mémoire intitulé « Nouvelles méthodes pour la détermination des orbites des comètes » où il utilise un procédé qu'il décrit soigneusement dans un appendice de quatre pages intitulé « Sur la méthode des moindres carrés », et il applique ensuite le même procédé à la mesure des degrés du méridien terrestre, un sujet de grand intérêt à l'époque.

Cet appendice est lumineux. En voici le début :

« Dans la plupart des questions où il s'agit de tirer des mesures données par l'observateur les résultats les plus exacts qu'elles peuvent offrir, on est presque toujours conduit à un système d'équations de la forme

$$E = a + bx + cy + fz + \text{etc.}$$

dans lesquelles a, b, c, f , etc. sont des coefficients connus, qui varient d'une équation à l'autre, et x, y, z , etc. sont des inconnues qu'il faut déterminer par la condition que la valeur de E se réduise, pour chaque équation, à une quantité ou nulle ou très petite. »

Je résume la suite. En général, il y a plus d'équations que d'inconnues, et il n'est pas possible d'annuler toutes les erreurs E, E', E'' etc. Le procédé consiste à rendre minimum la somme des carrés des erreurs. En dérivant la somme des carrés des erreurs par rapport à chacune des variables, on obtient des formes affines en nombre égal à celui des variables, et en les annulant on peut calculer ces variables, et aussi les erreurs E, E', E'' etc. Legendre précise qu'« on aura soin d'abrégier tous les calculs, tant des multiplications que de la résolution, en n'admettant dans chaque opération que le nombre de chiffres entiers ou décimaux que peut exiger le degré d'approximation dont la question est susceptible. »

Si parmi les erreurs ainsi calculées il en est d'aberrantes, on rejette les équations correspondantes et on reprend le calcul, dont la majeure partie (les multiplications) aura déjà été faite.

Au prix de la multiplication des coefficients et de la résolution d'un système d'équations linéaires on a donc un procédé général et efficace pour réaliser le programme indiqué par Legendre. Le procédé prend une forme remarquable quand il n'y a qu'une variable x , qui représente un point sur la droite ou dans le plan ou l'espace, et des observations x', x'', x''' etc. Le minimum de la somme des carrés distances au point x des points x', x'', x''' etc. est atteint au centre de gravité de x', x'', x''' etc. Comme le dit Legendre en conclusion,

« on voit donc que la méthode des moindres carrés fait connaître, en quelque sorte, le centre autour duquel viennent se ranger tous les résultats fournis par l'expérience, de manière à s'en écarter le moins qu'il est possible. »

Il y eut une querelle de priorité entre Legendre et Gauss sur la méthode des moindres carrés. Gauss l'a publiée plus tard et sous certaines hypothèses sur la loi des erreurs (en gros, qu'elles soient gaussiennes) et il l'a utilisée sans la publier avant 1805. Mais il semble bien que Legendre, et d'autres, l'aient aussi utilisée sans la formaliser avant 1805. Stigler, dans son histoire de la statistique, consacre son premier chapitre à la méthode des moindres carrés

exposée par Legendre (en donnant en illustration l'image d'un autre Legendre, cela est une toute autre histoire), avec un commentaire enthousiaste :

« For stark clarity of exposition the presentation is unsurpassed ; it must be counted as one of the clearest and most elegant introductions to a new statistical method in the history of statistics. »

Et surtout Stigler situe la méthode parmi celles qui avaient cours précédemment, à savoir les combinaisons ad hoc des équations pour se ramener à autant d'équations que d'inconnues. Il peut valoir la peine d'examiner ces solutions *ad hoc*. Pour autant que j'en puisse juger, elles combinent les seconds membres avec des coefficients qui sont toujours 1, 0 ou -1 . L'avantage est d'éviter les multiplications. L'inconvénient est de faire des choix ad hoc au vu des équations. Maintenant qu'il est d'usage de faire des choix aléatoires, on peut se demander si des choix aléatoires de coefficients 1, 0 et -1 , suivant une loi ne dépendant que du nombre d'équations et du nombre d'inconnues, ou plus simplement des choix aléatoires de 1 et -1 avec la probabilité naturelle, ne donneraient pas des solutions presque aussi bonnes avec, par répétition du procédé, une estimation assez correcte de l'erreur possible. Même s'il s'agit du centre de gravité, le choix aléatoire des coefficients 1 et -1 à somme nulle donne une estimation rapide de l'erreur possible, au lieu du calcul de la variance.

Ce qui manquait à l'exposé de Legendre était justement l'estimation des erreurs possibles sur les quantités calculées. La liaison avec les probabilités allait être faite par Gauss, et reprise par Laplace dès 1810 dans son « mémoire sur les intégrales définies et leur application aux probabilités » ; les applications comprennent la théorie des erreurs, celle de la combinaison des observations (dont je viens de parler), et la méthode des moindres carrés, dont Laplace déclare :

« Cette méthode, qui jusqu'à présent ne présentait que l'avantage de fournir sans aucun tâtonnement les équations finales nécessaires pour corriger les éléments, donne en même temps les corrections les plus précises. »

Bien avant Gauss, Laplace avait en main tous les éléments pour lier les probabilités et la méthode des moindres carrés. L'apparition de la fonction de Gauss et le calcul de son intégrale sur la droite apparaît chez lui dès 1774, dans son « mémoire sur la probabilité des causes par les évènements ». En effet, la loi d'une grandeur x quand on connaît la loi des erreurs possibles et les résultats d'observation x', x'', x''' est bien un calcul de probabilités des causes quand on connaît les évènements. Laplace en donne la forme générale

$$\varphi(x' - x)\varphi(x'' - x)\varphi(x''' - x)$$

quand $\varphi(\cdot)$ est la loi des erreurs. Le cas particulier où φ est la fonction de Gauss (Gauss est né en 1777) est aussi un cas limite, et la méthode des moindres carrés avec expression des corrections aurait pu apparaître dès ce moment, mais ce n'a pas été le cas. Laplace attribuait une grande valeur à la théorie des probabilités, il avait contribué au calcul de grandeurs à partir d'observations en combinant des équations, mais le lien entre les deux lui a échappé. Il a été découvert par Gauss avant que Laplace l'expose en 1810 puis dans son grand traité « Théorie analytique des probabilités ».

(Ici si j'ai le temps, ses travaux des années 1780 sur le rapport des naissances de garçons et de filles ; le livre d'Eric Brian de 2007 doit faire le point actuel sur la question).

La méthode des moindres carrés avec estimation des corrections était-elle pratiquée avant d'être parfaitement formalisée ? On peut se poser la question en lisant l'exposé très pédagogique qu'en fait Fourier dans deux mémoires tardifs, de 1826 et 1828, dans le second desquels il mentionne en passant l'avoir appliquée lorsqu'il était en Egypte, donc avant 1800, pour la détermination de la hauteur de la pyramide de Kéops. Ces mémoires ne contiennent pas de démonstrations, mais des énoncés et des applications, et ils sont destinés à des statisticiens. Depuis 1815, après avoir été démis de ses fonctions de préfet de l'Isère par Napoléon au retour de l'île d'Elbe, il s'était installé à Paris et avait été chargé, grâce au comte de Chabrol son ancien élève, de la direction des services de statistique du département de la Seine.

Puis, après son élection tardive à l'Académie des sciences en 1817, il faisait figure de connaisseur en statistique. Son influence peut se mesurer par l'audience du prix Montyon de statistique dont il avait été le rapporteur pour l'adoption comme prix de l'académie, et plus encore par le témoignage de Quételet qui lui doit, dit-il, ses connaissances en probabilités.

C'est le moment de parler de la transformation de Fourier. A la suite des séries de Fourier, elle apparaît clairement dans l'ouvrage principal de Fourier, la Théorie analytique de la chaleur, qui sous sa forme achevée a été publiée en 1822. Elle était déjà là dans le mémoire de Fourier de 1811 qui lui a valu le Prix de l'Académie des sciences (ou plutôt, les académies royales ayant été supprimées par la Révolution, de la première classe de l'Institut de France) et qui n'avait pas été publié à l'époque. On y trouve la formule d'inversion et de nombreuses applications. Mais là encore, il y a des antécédents. Le plus important est la présentation par Clairaut de la transformation de Fourier discrète, celle dont il est question au début de mon exposé, dans son mémoire de 1754 sur l'orbite apparente du soleil ; naturellement, ni Clairaut ni personne ne

pouvait soupçonner alors l'importance de la notion, et l'influence de Clairaut sur Fourier et ses successeurs semble avoir été nulle.

(Un mot sur Fourier, si cela ne prend pas trop de temps. 1768-1830. Collège militaire d'Auxerre, bénédictins. Arago. Bezout, Clairaut. 16 ans 1/2, professeur. Eq alg, Legendre. Eglise. 1789.1794 école normale. Polytechnique. Egypte. Isère :marais etc, description Egypte, chaleur. Les obstacles passés et durables : séries de Fourier, inégalités (analyse convexe), philosophie. Arago, Jacobi. Effet de serre.)

Aujourd'hui le nom de Fourier est utilisé principalement pour ce qui faisait le plus question à son époque, les séries de Fourier, les intégrales de Fourier, la transformation de Fourier, et plus généralement l'analyse de Fourier. C'est à Riemann que l'on doit l'appellation de séries de Fourier pour les séries trigonométriques dont les coefficients sont donnés par les formules intégrales connues comme formules de Fourier. Dans sa thèse sur les fonctions représentables comme sommes de séries trigonométriques Riemann fait une belle étude historique du sujet, il en montre les antécédents, puis met en évidence le couplage entre les formules intégrales donnant les coefficients et l'expression de la fonction comme somme de série trigonométrique, en déclarant que Fourier est le premier à avoir compris la portée de ce couplage. Aujourd'hui nous pouvons exprimer ce couplage comme celui de l'analyse et de la synthèse. Fourier avait insisté sur sa valeur universelle, et là un tempérament est nécessaire : la théorie ne peut s'appliquer qu'à des fonctions intégrables, et alors la convergence de la série pose problème.

Ainsi on doit considérer les formules de Fourier, celles qui expriment les coefficients à partir de la fonction et celle qui exprime la fonction à partir des coefficients, comme un programme plutôt qu'un énoncé. Comme énoncé, il a amené à la recherche de contre-exemples instructifs. Mais comme programme, il est à l'origine de la formalisation de la notion d'intégrale, avec l'intégrale de Riemann, puis celle de Lebesgue, puis celle de Denjoy, puis des distributions de Laurent Schwartz. Il est aussi à l'origine des différents moyens d'exprimer la représentation de la fonction, ou du signal, comme somme d'une série trigonométrique : les procédés de sommation au lieu de la seule considération des sommes partielles, et les espaces fonctionnels dans lesquels les fonctions apparaissent comme des points.

Parmi ces espaces fonctionnels le premier et le plus important est l'espace de Hilbert. On peut à volonté le considérer comme un ℓ^2 , espace des suites de carré sommable, ou comme un L^2 , espace des fonctions de carré intégrable. On doit préciser quels sont les indices pour les suites, et de quelles fonctions

il s'agit. Le cas classique, qui fait l'objet du théorème de Riesz-Fischer, est relatif à $\ell^2(\mathbb{Z})$ et $L^2(\mathbb{T})$, \mathbb{T} étant le cercle \mathbb{R}/\mathbb{Z} . L'isomorphisme est assuré par la transformation de Fourier : c'est la correspondance entre une fonction périodique (définie sur \mathbb{T}) et la suite de ses coefficients de Fourier. Voici des points de repère chronologiques. L'intégrale de Lebesgue date de 1901. L'égalité de Parseval, ainsi nommée par Fatou, se trouve dans sa thèse de 1906 : elle exprime que la transformation de Fourier réalise une isométrie de L^2 vers ℓ^2 , mais non que cette isométrie est surjective. En 1907, Frédéric Riesz et Ernst Fischer, indépendamment ou plutôt concurremment, ont publié deux séries de notes aux Comptes rendus de l'Académie des sciences établissant le théorème et jetant les bases de ce que Fischer a appelé une géométrie des fonctions. La priorité n'est pas claire, et on peut aussi bien parler du théorème de Fischer-Riesz. La clé pour passer de Parseval à Riesz-Fischer est le fait que L^2 est un espace complet. Et ce dernier fait montre la supériorité de l'intégrale de Lebesgue sur celle de Riemann.

Cependant raconter l'histoire ainsi est un renversement de son cours réel. En 1907, on ne disposait pas des symboles L^2 ni ℓ^2 . La notion d'espace normé n'existait pas encore, et celle d'espace normé complet encore moins. Il fallait de longues périphrases pour parler de suites de Cauchy, et même pour énoncer le théorème. C'est un cas typique où l'outil de la démonstration, le fait que L^2 est complet, est devenu le théorème principal, et que pour l'énoncer en trois mots on a eu besoin de définir les termes aujourd'hui classiques de L^2 et d'espace complet : l'énoncé sous cette forme n'apparaît qu'en 1930 dans la Théorie des opérations linéaires de Banach. En bref, le lemme est devenu théorème et sa substance est passée dans les définitions.

Le cadre général est apparu au cours des années 1930. C'est celui des groupes abéliens localement compacts et de leur dualité. Ainsi $L^2(\mathbb{T})$ et $\ell^2(\mathbb{Z})$ sont en dualité, de même que $L^2(\mathbb{R})$ avec lui-même ou $\ell^2(\mathbb{Z}/N\mathbb{Z})$ avec lui-même. La transformée de Fourier réalise un isomorphisme isométrique entre $L^2(G)$ et $L^2(\hat{G})$, G et \hat{G} étant deux groupes abéliens localement compacts en dualité. La notation ℓ au lieu de L est utilisée lorsque le groupe est discret.

L'aspect géométrique est traduit aussi par la terminologie des bases orthonormales et des systèmes orthogonaux. Avec leur ancêtre, la méthode des moindres carrés, et tous les développements d'analyse liés aux systèmes orthogonaux, les suites de carré sommable ont une place de choix dans l'univers mathématique.

Et cependant il y a plus naturel encore que L^2 et ℓ^2 . C'est L^1 et ℓ^1 . En termes classiques, les transformées de Fourier fournissent les séries de Fourier-Lebesgue (les coefficients sont donnés par les formules de Fourier avec comme intégrale l'intégrale de Lebesgue), appelées aussi simplement séries de Fourier, et d'autre part les fonctions continues sommes de séries trigonométriques absolument convergentes, qui constituent l'algèbre de Wiener. Le point essentiel, mis en évidence par Wiener, est que ces espaces L^1 et ℓ^1 sont des algèbres pour la convolution. Par transformation de Fourier, convolution et multiplication s'échangent. La convolution se trouve partout en analyse, mais en tant que notion c'est l'une des dernières nées : le terme de convolution ne s'est imposé que dans la seconde moitié du 20^e siècle. Au moins Wiener en avait fait sous le nom de Faltung un merveilleux usage, qui a servi de modèle à plusieurs égards à la théorie des distributions de Schwartz.

L'algèbre de Wiener a été un grand sujet de l'analyse au cours des années 1950 et 1960, et j'ai collaboré à cette époque à la découverte de ses secrets. Puis d'autres sujets sont passés au premier plan. Il est remarquable pour moi que les problèmes actuels de théorie du signal et de traitement des données donnent un nouveau départ à ce secteur de l'analyse de Fourier.

Wiener est connu à bien des titres. Le pense ici au mouvement brownien et aux séries de Fourier à coefficients aléatoires, un autre sujet sur lequel j'ai travaillé dans la tradition de Paley, Wiener et Zygmund. Le théorème de Candelès, Romberg et Tao par lequel j'ai débuté cet exposé marie la transformation de Fourier discrète, l'espace ℓ^1 et les probabilités. C'est donc de plusieurs manières un héritage de Wiener.

Il est bon de conclure en revenant à ce théorème pour en indiquer le contexte et d'éventuels prolongements.

L'article de CRT donne en introduction l'exposé d'une expérience numérique qui donne d'excellents résultats et qui viole l'usage établi par le théorème d'échantillonnage de Shannon. Il s'agit d'une figure appelée le « fantôme de Logan-Shepp », constituée par quelques plages en forme d'ellipses qui empiètent les unes sur les autres, et qui est un test classique pour les méthodes d'analyse du signal. On regarde sa transformée de Fourier sur 22 demi-droites issues de l'origine et régulièrement réparties, puis les valeurs qu'elle prend sur une vingtaine de points sur chaque demi-droite, au total 512 points.

On reconstitue exactement la figure par un procédé d'analyse convexe à partir des valeurs prises par la transformée de Fourier sur ces points, qui sont espacés les uns des autres d'au moins $\pi/22$, alors que le théorème d'échantillonnage exigerait pour être appliqué que l'espacement des points voisins

soit $2\pi/N$ avec $N = 1000$ environ. Cet exemple n'est pas encore parfaitement compris, mais il repose sur la recherche du minimum dans l'algèbre de Wiener des normes des fonctions qui extrapolent la restriction à un ensemble convexe de la transformée de Fourier du signal considéré. Le cadre des groupes cycliques d'ordre N pour G et \hat{G} se prête bien à l'étude théorique, et l'échantillonnage aléatoire aboutit à l'énoncé que j'ai indiqué.

L'échantillonnage aléatoire est essentiel dans la démonstration, et il est également très intéressant en pratique. En expérimentant son utilisation on constate dans des cas simples que le choix de W , l'effectif de l'échantillon, peut être rendu notablement plus petit que la valeur fixée par le théorème. L'article signale les résultats d'expériences allant dans ce sens.

Quel est le champ d'application de ces méthodes ? Au contraire de la méthode des moindres carrés, qui consistait à gérer un ensemble de données surabondantes, il s'agit ici de restreindre les observations au minimum indispensable, un peu plus de T si l'on sait que le signal est constitué de T pics au plus.

L'utilisation de l'analyse convexe, qui rejoint ce que Fourier appelait l'analyse indéterminée, c'est à dire l'étude des inégalités et leur interprétation géométrique, s'est introduit progressivement dans ce domaine. Les normes ℓ^1 se sont avérées d'usage commode, lorsque la nature des problèmes suggérait la norme ℓ^0 , c'est-à-dire le nombre de termes non nuls. Un exemple typique se trouve dans l'article de Donoho et Huo de 2001 dont la motivation est la reconstruction à partir d'un développement mixte en ondelettes et en Fourier, ou en ondelettes et en bandelettes, d'une image comprenant des plages et des contours. Le cadre général est celui déjà dressé en 1993 par Mallat et Zhang : on dispose d'un « dictionnaire » constitué de plusieurs bases orthonormales, et il s'agit de l'utiliser pour obtenir de meilleurs résultats qu'avec une seule base. Dans l'article DZ, l'étude théorique de départ concerne les groupes cycliques d'ordre N , G et \hat{G} . Toute fonction définie sur G peut se traduire par Fourier en une fonction définie sur \hat{G} , et réciproquement. Il y a donc plusieurs représentations possibles pour la même fonction, en exprimant une partie sur G et une partie sur \hat{G} . Une représentation économique utilise le minimum de points de G et de \hat{G} , soit T points de G et W points de \hat{G} ; sa norme dans ℓ^0 est alors $T + W$; et on peut espérer obtenir la meilleure représentation en minimisant la norme ℓ^0 . C'est bien le cas lorsque $T + W < \sqrt{N}$, ce n'est plus lorsqu'il y a égalité (description de l'exemple ? c'est la fonction indicatrice du sous-groupe d'ordre n quand $N = n^2$, qui, à un facteur multiplicatif près, est sa propre transformée de Fourier : le minimum de la norme ℓ^0 est atteint de deux

manières). Mais chercher le minimum dans ℓ^0 est impraticable. Le minimum dans ℓ^1 est praticable et donne le bon résultat quand $T + W < \frac{1}{2}\sqrt{N}$.

Personnages

Clairaut 1713-1765

Lagrange 1736-1813

Laplace 1749-1827

Legendre 1736-1813

Fourier 1768-1830

Gauss 1777-1855

Riemann 1826-1866

Lebesgue 1876-1941

F. Riesz 1880-1948

E. Fischer

N. Wiener 1894-1964

Yves Meyer

Stephane Mallat

David Donoho

Emmanuel Candès

JPK

(Aujourd'hui Fourier = séries, intégrales, transformation, analyse. Le couple de formules et Riemann. Les intégrales, Lebesgue. La théorie L^2 , Fatou-parseval, Riesz Fischer, les systèmes orthonormaux, la géométrisation de l'analyse.

La théorie L^1 ou ℓ^1 , Wiener. L'aléatoire en analyse, Paley-Zygmund et Wiener. Pseudopériodicité. Comment retrouver la fonction sachant que sa série de Fourier est lacunaire ?

L'expérience signalée dans Candès etc., qui viole Shannon. Commentaire sur le théorème, les essais numériques.

Retour sur Legendre : surabondance, ℓ^2 . Ici beaucoup de paramètres cachés, insuffisance d'information, ℓ^1 .

Origine : ℓ^0 , retenir les plus grands coefficients. Adaptation à un dictionnaire surabondant, matching pursuit, cf Donoho et Basis pursuit.)

J.-P. KAHANE, Laboratoire de Mathématique d'Orsay, UMR 8628, Bâtiment 425, Faculté des Sciences d'Orsay, Université Paris-Sud 11, F-91405 Orsay Cedex
E-mail : Jean-Pierre.Kahane@math.u-psud.fr

Restrictions et prolongements dans l'algèbre de Wiener

Jean-Pierre Kahane

Introduction

Les travaux sur l'échantillonnage parcimonieux (compressed sensing) en théorie du signal ont mis en relief un très intéressant problème d'analyse de Fourier. En me référant aux notations d'Emmanuel Candès dans [2], je le présenterai de la manière suivante.

G est un groupe abélien localement compact et \hat{G} son groupe dual. Soit (S) une partie de $L^1(G)$, et K une partie de \hat{G} . A toute $x \in L^1(G)$ on associe sa transformée de Fourier \hat{x} , qui appartient à l'algèbre de Wiener $A(\hat{G})(= \mathcal{FL}^1(G))$, et la restriction \hat{x} à K , $\hat{x}|_K$. Les $y \in L^1(G)$ tels que $\hat{y}|_K = \hat{x}|_K$ constituent un convexe fermé dans $L^1(G)$, et les \hat{y} , prolongements de $\hat{x}|_K$ dans $A(\hat{G})$, constituent un convexe fermé dans $A(\hat{G})$. Il s'agit de donner une condition sur (S) et K garantissant que, si $x \in (S)$, \hat{x} est le prolongement minimal unique de $\hat{x}|_K$ dans $A(\hat{G})$:

$$(1) \quad \|\hat{x}_0\|_{A(\hat{G})} = \inf\{\|\hat{y}\|_{A(\hat{G})} : \hat{y}|_K = \hat{x}|_K\} \iff \hat{x}_0 = \hat{x}.$$

Dans [2] et dans les travaux qui précèdent, de Candès, Romberg et Tao [3] et antérieurement, sur un sujet voisin, de Donoho et Huo [5], G et \hat{G} sont le groupe cyclique $\mathbb{Z}/N\mathbb{Z}$, N étant un entier très grand. Le groupe G est une représentation discrète des temps et le groupe \hat{G} des fréquences. En désignant par S le support du signal x , (S) est traduit par une condition sur S , en l'occurrence dans [2] et [3] que le cardinal de S , $|S|$, est majoré par un nombre T petit par rapport à N :

$$(2) \quad |S| = \|x\|_{\ell^\infty} \leq T \ll N$$

(la notation $\|\cdot\|_{\ell^\infty}$ a justement (2) pour définition). K est un ensemble de fréquences, de cardinal $|K|$. Le théorème principal de [2] et [3] est que, si l'on

fixe $|K|$ et que l'on choisit K aléatoirement avec la probabilité naturelle, on retrouve le signal x par le procédé (1) avec une probabilité voisine de 1 si $|K|$ est assez grand par rapport à $T \log N$, à savoir que

$$(3) \quad |K| \geq CT \log N$$

avec $C \geq 23(M + 1)$ entraîne

$$(4) \quad P((1)) \geq 1 - cN^{-M},$$

où c désigne une constante absolue.

Ainsi, si l'on sait que le signal x est porté par T points, avec $T \ll N$, on est loin d'avoir besoin des N coefficients de Fourier de x pour le reconstituer : en prenant $C = 100$ dans (3), $|K|$ coefficients choisis au hasard suffisent pratiquement.

Les titre de [3] et [5] mettent en avant ce que les auteurs appellent « uncertainty principles », c'est-à-dire le fait (incontestablement certain) que les supports de x et de \hat{x} ne peuvent pas être petits simultanément. Les relations de Heisenberg, d'où est tiré le vocable de « principe d'incertitude », expriment simplement que l'on ne peut pas localiser simultanément un signal en temps et en fréquence. En physique, elles signifient que la représentation des particules par des points matériels est inadéquate à l'échelle quantique. En mathématiques, elles constituent un théorème de localisation sur transformées de Fourier. Pour $G = \mathbb{Z}/N\mathbb{Z}$, \hat{G} est une autre copie de $\mathbb{Z}/N\mathbb{Z}$, et le théorème de localisation s'exprime ainsi :

$$(5) \quad \|x\|_{\ell^o} \|\hat{x}\|_{\ell^o} \geq N.$$

La démonstration, que j'indique en appendice, donne des informations supplémentaires en fonction des propriétés arithmétiques de N . On exprime souvent le théorème de localisation dans les $\mathbb{Z}/N\mathbb{Z}$ par une conséquence de (5), à savoir

$$(6) \quad \|x\|_{\ell^o} + \|\hat{x}\|_{\ell^o} \geq 2\sqrt{N}.$$

L'égalité dans (6) a lieu lorsque N est un carré parfait et que les supports de x et de \hat{x} sont des translatés du sous-groupe cyclique d'ordre \sqrt{N} du groupe $\mathbb{Z}/N\mathbb{Z}$, en particulier quand x et \hat{x} sont le « peigne de Dirac » d'ordre \sqrt{N} , divisé par \sqrt{N} .

Dans [5], Donoho et Huo tirent parti de cet exemple comme limitation naturelle pour la question qu'ils posent de la représentation d'un signal la plus économique possible en temps et en fréquences, c'est-à-dire l'expression de x , fonction définie sur $G = \mathbb{Z}/N\mathbb{Z}$, sous la forme

$$(7) \quad x = y + z, \quad \|y\|_{\ell^o} + \|\hat{z}\|_{\ell^o} \text{ minimum.}$$

Quand N est un carré parfait et que

$$(8) \quad \|x\|_{\ell^o} = \|\hat{x}\|_{\ell^o} = \sqrt{N},$$

on peut choisir dans (7) $x = y$ ou $x = z$: la représentation la plus économique n'est pas définie. Donoho et Huo montrent qu'elle est bien définie lorsque x est de la forme

$$(9) \quad x = y + z, \quad \|y\|_{\ell^o} + \|\hat{z}\|_{\ell^o} < \sqrt{N}$$

et que, sous l'hypothèse additionnelle

$$(10) \quad x = y + z, \quad \|y\|_{\ell^o} + \|\hat{z}\|_{\ell^o} < \frac{1}{2}\sqrt{N},$$

on obtient la solution (7) en remplaçant la condition

$$(11) \quad \|y\|_{\ell^o} + \|\hat{z}\|_{\ell^o} \text{ minimum,}$$

pratiquement ingérable, par la condition

$$(12) \quad \|y\|_{\ell^1} + \|\hat{z}\|_{\ell^1} \text{ minimum,}$$

qui se prête au calcul.

Candès, Romberg et Tao utilisent leur théorème sur le prolongement minimal (formules (3) et (4)) pour proposer une amélioration considérable du théorème de localisation (6) sous la forme d'une proposition hautement probable : si l'on choisit au hasard une partie S de $G = \mathbb{Z}/N\mathbb{Z}$ et une partie S' de \hat{G} telles que

$$(13) \quad |S| + |S'| < \gamma \frac{N}{\sqrt{\log N}},$$

alors, avec une probabilité aussi voisine de 1 que l'on veut par choix de γ , il n'existe aucune fonction x définie sur G à support dans S telle que \hat{x} ait son support dans S' .

Par cet aperçu très partiel des travaux de Donoho et Huo de 2001 et de Candès, Romberg et Tao de 2006, on voit la portée en théorie du signal de la méthode de minimisation dans ℓ^1 . Cette méthode n'est pas nouvelle, et son histoire est retracée brièvement dans [3] p. 494. Les initiateurs, au cours des années 1980, semblent avoir été des géophysiciens. La minimisation des normes ℓ^1 est une question d'analyse convexe dont on peut trouver les prolégomènes dans ce que Fourier appelait l'analyse indéterminée [6]. La réduction de ce sujet d'analyse convexe à des techniques de programmation linéaire a été exposée en 1998 par Chen, Donoho et Saunders dans [4]. La minimisation de normes ℓ^1 est aujourd'hui un outil important de la théorie du signal.

Je vais m'attacher aux conditions de validité de (1) dans un contexte très différent : la reconstruction d'une fonction à spectre lacunaire à partir de sa restriction à un intervalle par prolongement minimal dans l'algèbre de Wiener.

Reconstruction des fonctions à spectre lacunaire

Dans cette section on s'écarte de l'analyse de Fourier discrète. Le groupe G sera \mathbb{Z}^ν ($\nu = 1, 2, \dots$) et \hat{G} sera \mathbb{T}^ν . On s'intéresse aux fonctions $\hat{x} \in A(\mathbb{T}^\nu)$,

$$(14) \quad \hat{x}(t) = \sum_{n \in S} x(n)e(n.t), \quad \|\hat{x}\|_A = \sum_n |x(n)| < \infty$$

($t = (t_1, \dots, t_\nu)$, $n = (n_1, \dots, n_\nu)$, $n.t = n_1 t_1 + \dots + n_\nu t_\nu$, $e(u) = \exp(2\pi i u)$) dont le spectre S est lacunaire au sens que deux points distincts de S sont toujours à une distance $\geq d$:

$$(15) \quad \inf_{s \in S, s' \in S, s' \neq s} |s' - s| \geq d$$

$|\cdot|$ représente la distance euclidienne, et le premier membre de (15) s'appelle le « pas » de S . Ici K sera une boule fermée de rayon r dans \mathbb{T}^ν ($0 < r < \frac{1}{2}$). Le problème posé dans l'introduction est de donner une condition sur d et r (ν étant donné) garantissant que \hat{x} est le prolongement minimal unique dans $A(\mathbb{T}^\nu)$ de sa restriction à K , $\hat{x}|_K$, c'est-à-dire que la formule (1) est valable.

Pour faciliter la compréhension, limitons-nous d'abord au cas $\nu = 1$. Alors K est un intervalle, et comme la solution du problème est invariante par les translations sur \mathbb{T} , on peut choisir $K = [-r, r]$. Ecrivons (14) sous la forme

$$(16) \quad \hat{x} = \sum_{n \in S} x_n e_n, \quad \|\hat{x}\|_A = \sum_n |x_n| < \infty.$$

(1) signifie que, quelle que soit $\hat{z} \in A(\mathbb{T})$ nulle sur K et $\neq 0$, soit

$$(17) \quad \hat{z} = \sum_{n \in \mathbb{Z}} z_n e_n, \quad 0 < \sum_n |z_n| < \infty, \quad z|_K = 0,$$

on a

$$(18) \quad \sum_{n \in S} |x_n + z_n| + \sum_{n \notin S} |z_n| > \sum_{n \in S} |x_n|.$$

Dire que (18) a lieu pour toute suite $(x_n) \in \ell^1(S)$, c'est dire que

$$(19) \quad \sum_{n \notin S} |z_n| > \sum_{n \in S} |z_n|$$

(on a utilisé $\sup_{x_n} (|x_n| - |x_n + z_n|) = |z_n|$; c'est le même calcul qu'en [3] p. 494).

On cherche donc une condition sur d et r pour que (17) implique (19).

Il y a bien des façons d'aborder le problème. Voici celle qui m'a donné les meilleurs résultats (loin sans doute d'être optimaux). On multiplie \hat{z} , qui est nulle sur K , par une fonction $\hat{\varphi}$ qui est presque entièrement portée par K . Le résultat est voisin de O , et il en est de même de sa transformée de Fourier, convolution de z et de φ . Si φ est presque entièrement portée par l'intervalle $[-d, d]$, la quasi-nullité de la convolution $z * \varphi$ est capable d'entraîner (19).

Il est commode de partir de la fonction de Gauss

$$(20) \quad \gamma(t) = e^{-\pi t^2} \quad (t \in \mathbb{R}),$$

de poser

$$(21) \quad \gamma_a(t) = a\gamma(at) \quad (a \gg \frac{1}{r})$$

et de réduire γ_a modulo 1 sous la forme

$$(22) \quad \gamma_a^*(t) = \sum_{m \in \mathbb{Z}} \gamma_a(t - m) = \sum_{n \in \mathbb{Z}} \gamma\left(\frac{n}{a}\right) e_n(t).$$

C'est γ_a^* qui nous servira de fonction $\hat{\varphi}$. Comme elle est paire, et décroissante et convexe sur l'intervalle $[r, \frac{1}{2}]$, sa restriction à $\mathbb{T} \setminus K$ a une norme $\leq 2\gamma_a^*(r)$ dans l'algèbre quotient $A(\mathbb{T} \setminus K)$, d'où résulte que

$$(23) \quad \|\hat{z}\gamma_a^*\|_{A(\mathbb{T})} \leq 2\|\hat{z}\|_{A(\mathbb{T})}\gamma_a^*(r).$$

Pour chaque $s \in S$, intégrons sur \mathbb{T} le produit $\hat{z}\gamma_a^*e_{-s}$:

$$(24) \quad \int \hat{z}\gamma_a^*e_{-s} = \sum_n z_n \gamma\left(\frac{n-s}{a}\right) = z_s + \sum_{n \neq s} z_n \gamma\left(\frac{n-s}{a}\right)$$

Multiplions par $\varepsilon_s = \frac{|z_s|}{z_s}$ et faisons la somme sur S :

$$(25) \quad \sum_s \varepsilon_s \int \hat{z}\gamma_a^*e_{-s} = \sum_{s \in S} |z_s| + \sum_{n \in \mathbb{Z}, s \in S, n \neq s} \varepsilon_s z_n \gamma\left(\frac{n-s}{a}\right).$$

Dans cette dernière somme, distinguons les $n \in S$, que nous appellerons s' , et les $n \notin S$, et remarquons que le premier membre de (25) est majoré en valeur absolue par celui de (23). On obtient

$$(26) \quad \left| \sum_{s \in S} |z_s| + \sum_{s' \in S} A(s') z_{s'} + \sum_{n \notin S} B(n) z_n \right| \leq 2 \|\hat{z}\|_{A(\mathbb{T})} \gamma_a^*(r)$$

avec

$$(27) \quad \begin{cases} A(s') = \sum_{s \in S \setminus \{s'\}} \varepsilon_s \gamma\left(\frac{s'-s}{a}\right) \\ B(n) = \sum_{s \in S} \varepsilon_s \gamma\left(\frac{n-s}{a}\right) \end{cases}$$

et (26) entraîne

$$(28) \quad \sum_{n \notin S} |z_n| (|B(n)| + 2\gamma_a^*(r)) \geq \sum_{s \in S} |z_s| (1 - |A(s)| - 2\gamma_a^*(r)).$$

En vue d'obtenir (19), supposons

$$(29) \quad \frac{1}{r} < a < d \quad \text{et} \quad a \geq 10$$

et majorons $|B(n)|$, $\gamma_a^*(r)$ et $|A(s)|$.

Quand $n \notin S$, soit h la distance de n à S . Alors

$$(30) \quad \begin{aligned} |B(n)| &\leq \gamma\left(\frac{h}{a}\right) + \gamma\left(\frac{d-h}{a}\right) + 2 \sum_{j=1}^{\infty} \gamma\left(\frac{jd}{a}\right) \\ &\leq \gamma\left(\frac{1}{a}\right) + \gamma\left(\frac{d}{2a}\right) + 2 \sum_{j=1}^{\infty} \gamma\left(\frac{jd}{a}\right) \\ &\leq 1 - \left(\frac{3\pi}{4a^2}\right) + 2 \exp\left(-\pi \frac{d^2}{4a^2}\right). \end{aligned}$$

Pour $s \in S$,

$$(31) \quad |A(s)| \leq 2 \sum_{j=1}^{\infty} \gamma\left(\frac{jd}{a}\right) \leq \exp\left(-\pi \frac{d^2}{4a^2}\right).$$

Enfin

$$(32) \quad 2\gamma_a^*(r) \leq 3a \exp(-\pi a^2 r^2).$$

Supposons

$$(33) \quad r > \frac{2}{a} \sqrt{\log a};$$

alors

$$(34) \quad 2\gamma_a^*(r) \leq \frac{\pi}{8a^2}.$$

Supposons

$$(35) \quad d > 2a\sqrt{\log a};$$

alors

$$(36) \quad \begin{cases} |A(s)| & \leq \frac{\pi}{4a^2} \\ |B(n)| & \leq 1 - \frac{2\pi}{4a^2} \end{cases}$$

et (19) est bien établie.

L'une des conditions

$$(37) \quad d > \frac{5}{r} \log \frac{1}{r} \quad \left(r < \frac{1}{10}\right)$$

ou

$$(38) \quad r > \frac{5}{d} \log d \quad (d > 10)$$

entraîne donc, en choisissant $a = \sqrt{\frac{d}{r}}$, ce qui implique (29), que pour toutes les fonctions \hat{z} vérifiant (17) on a (19).

L'étude faite en dimension 1 se traduit sans peine en dimension $\nu > 1$, au prix d'une augmentation des constantes.

On prend pour K la boule $|t| < r$ et on remplace γ , γ_a et γ_a^* par

$$(39) \quad \gamma_\nu(t) = e^{-\pi|t|^2} \quad (t \in \mathbb{R}^\nu)$$

$$(40) \quad \gamma_{a,\nu}(t) = a^\nu \gamma(at) \quad (a \gg \frac{1}{r})$$

$$(41) \quad \begin{aligned} \gamma_{a,\nu}^*(t) &= \sum_{m \in \mathbb{Z}^\nu} \gamma_{a,\nu}(t - m) \\ &= \sum_{n \in \mathbb{Z}^\nu} \gamma_\nu\left(\frac{n}{a}\right) e(n.t) \end{aligned}$$

La formule (28) reste valable en remplaçant $2\gamma_a^*(r)$ par $\|\gamma_{a,\nu}^*\|_{A(\mathbb{T}^\nu \setminus K)}$, et $A(s)$ et $B(n)$ par $A_\nu(s)$ et $B_\nu(n)$, définis comme en (27) avec γ_ν au lieu de γ . La majoration de $\|\gamma_{a,\nu}^*\|_{A(\mathbb{T}^\nu \setminus K)}$ peut se faire ici en estimant les dérivées d'ordre $\leq \nu$ de $\gamma_{a,\nu}^*$ sur $\mathbb{T}^\nu \setminus K$, et au lieu de (32) on a

$$(42) \quad \|\gamma_{a,\nu}^*\|_{A(\mathbb{T}^\nu \setminus K)} \leq C_\nu a^\nu \exp(-\pi a^2 r^2), \quad C_\nu < \nu !$$

La majoration de $|B_\nu(n)|$ se fait comme en (30), en distinguant le point s le plus voisin de n (ou l'un de ces points s'il y en a plusieurs) et les points de S dans la boule de centre s et de rayon $\frac{d}{2}$, dont le nombre est $\leq 2^\nu$. Celle de $|A_\nu(s)|$ est analogue, en plus simple. On obtient

$$(43) \quad \begin{cases} |B(n)| &\leq 1 - \frac{3\pi}{4a^2} + 2^{\nu+1} \exp\left(-\pi \frac{d^2}{4a^2}\right) \\ |A(s)| &\leq 2^{\nu+1} \exp\left(-\pi \frac{d^2}{4a^2}\right), \end{cases}$$

d'où finalement l'énoncé suivant.

THÉORÈME.— *Soit $\hat{G} = \mathbb{T}^\nu$ ($\nu = 1, 2, \dots$). Pour que toute fonction $\hat{x} \in A(\hat{G})$, dont le pas du spectre est d ($d \geq 10$) s'obtienne à partir de sa restriction $\hat{x}|_K$ à une boule K de rayon r comme prolongement minimal de $\hat{x}|_K$ dans $A(\hat{G})$, il suffit que l'une des conditions suivantes soit réalisée :*

$$(44) \quad r > \frac{5}{d} \sqrt{\nu} \log(\nu d)$$

$$(45) \quad d > \frac{5}{r} \sqrt{\log \frac{1}{r} + \nu - 1}$$

(Pour $\nu = 1$, ce sont les formules (38 et (37)).

La constante 5 peut être abaissée quand d est grand, et doit être ajustée pour couvrir tous les cas.

Discussion et commentaires

Je commencerai par le cas $\nu = 1$.

1. On peut transcrire le théorème dans le contexte donné dans l'introduction, où G et \hat{G} sont le groupe cyclique $\mathbb{Z}/N\mathbb{Z}$, N très grand, G étant une représentation discrète des temps et \hat{G} des fréquences. Alors x est un signal porté par des temps séparés par des intervalle de longueur $\geq d$, et K est une bande de fréquences de largeur $\geq 2rN$ (si l'on représente \hat{G} par les racines N -ièmes de l'unité, K est la partie de \hat{G} contenue dans un arc du cercle unité). On obtient l'énoncé que voici :

Si le support S du signal x (porté par $\mathbb{Z}/N\mathbb{Z}$) a un pas $\geq d$, c'est-à-dire si

$$(46) \quad \inf_{s \in S, s' \in S, s' \neq s} |s' - s| \geq d \quad (d \geq 10)$$

on peut obtenir x à partir de $|K|$ fréquences consécutives (dans $\mathbb{Z}/N\mathbb{Z}$) par le procédé indiqué en (1), dès que

$$(47) \quad |K| \geq 10 \frac{N}{d} \sqrt{\log d}$$

Le défaut de cet énoncé par rapport à celui de Candès, Romberg et Tao est dans l'hypothèse faite sur S : (46) implique $|S| \leq \frac{N}{d}$, mais $|S| \leq T$ n'implique rien sur d . Cependant, si T est donné et si S s'obtient en disposant T points au hasard sur $\mathbb{Z}/N\mathbb{Z}$, on a $\frac{N}{d} > \frac{10}{T^2}$ avec une probabilité $> 0,95$, donc il est probable que

$$(48) \quad |K| \geq 100T^2 \sqrt{\log T}$$

suffit pour que (1) soit valable. L'avantage est qu'il s'agit de fréquences consécutives, c'est-à-dire d'une bande de fréquences.

Cela serait à comparer aux résultats connus.

2. On a beaucoup étudié les séries trigonométriques lacunaires et les fonctions qu'elles représentent :

$$(49) \quad \hat{x}(t) = \sum_{n \in S} x_n e_n(t) = \sum_{n \in S} x_n e^{2\pi i n t} \quad (t \in \mathbb{T}).$$

C'est de ces séries et fonctions que nous nous sommes occupés à partir de (16), sous la condition $\sum |x_n| < \infty$. On peut les étudier sous d'autres conditions, en particulier $\sum |x_n|^2 < \infty$.

En gros, si S est lacunaire, beaucoup de propriétés de \hat{x} se lisent sur les restrictions $\hat{x}|_K$, où K est un ensemble convenable. Je me limiterai ici au cas des intervalles.

Première propriété : la nullité. Quand est-il vrai, sous l'hypothèse (49), que $\hat{x}|_K = 0$ entraîne $\hat{x} = 0$? Autrement dit, que \hat{x} soit bien défini par les valeurs qu'elle prend sur K ? Cela se traduit lorsque $K = [-r, r]$ par une question sur les fonctions entières de type exponentiel $\leq \pi(1-r)$ bornées (ou L^1 , ou L^2, \dots) sur la droite réelle, $X(z)$ ($z \in \mathbb{C}$) : à quelle condition leur nullité sur $\mathbb{Z} \setminus S$ entraîne-t-elle leur nullité? Plus simplement, S étant donnée, quelle est la borne inférieure des r tels qu'il en soit ainsi? La réponse est fournie par la théorie de Beurling et Malliavin [1, 10]. Associons à toute suite $\Lambda \subset \mathbb{Z}$ sa fonction de décompte $n(t)$, dont la variation sur un intervalle égale de nombre de points de Λ sur cet intervalle, et convenons de dire que Λ admet Δ comme densité- BM si

$$(50) \quad \int_{\mathbb{R}} \frac{|n(t) - \Delta t|}{1 + t^2} dt < \infty.$$

La densité- BM intérieure de S se définit naturellement comme la borne supérieure des densités- BM des suites Λ contenues dans S . C'est justement la borne inférieure dans $|K|$ (longueur de K) tels que $\hat{x}|_K = 0$ entraîne $\hat{x} = 0$ ($K = [-r, r]$).

En particulier, pour que \hat{x} soit bien définie par les valeurs qu'elle prend sur n'importe quel intervalle, si petit soit-il, il faut et il suffit que S ne contienne aucune suite Λ ayant une densité- $BM > 0$. Il n'est pas nécessaire que S soit très dispersée; il suffit qu'elle ait de grandes lacunes.

Seconde propriété : l'appartenance à l'algèbre de Wiener. Rappelons que $A(\mathbb{T}) = \mathcal{F}\ell^1(\mathbb{Z})$ et que $A(K)$ est l'algèbre quotient de $A(\mathbb{T})$ par l'idéal constitué par les fonctions nulles sur K , autrement dit, l'algèbre des restrictions à K des fonctions $\in A(\mathbb{T})$. On considère les fonctions \hat{x} de (49) avec $x \in \ell^1(\mathbb{Z})$. Quelle est la borne inférieure des longueurs des intervalles K tels que $\hat{x}|_K \in A(K)$ implique $\hat{x} \in A(\mathbb{T})$? Réponse : c'est la densité uniforme extérieure de S , c'est-à-dire la borne inférieure des densités D des suites Λ telles que

$$(51) \quad n(t) = Dt + O(1) \quad (t \rightarrow \pm\infty)$$

et qui contiennent S ; on exprime (51) en disant que Λ a D pour densité uniforme [8].

Variante : l'appartenance aux espaces $A^{\alpha,r}(\mathbb{T}) = \mathcal{F}\ell^{\alpha,r}(\mathbb{Z})$. La définition de $\ell^{\alpha,r}(\mathbb{Z})$ ($1 \leq \alpha \leq \infty$, $r \in \mathbb{R}$) est

$$(52) \quad \begin{aligned} \ell^{\alpha,r}(\mathbb{Z}) &= \left\{ x : \sum_{n \in \mathbb{Z}} |x_n|^\alpha (1 + |n|)^{r\alpha} < \infty \right\} \text{ si } 1 \leq \alpha < \infty \\ &= \left\{ x : \sup_{n \in \mathbb{Z}} |x_n| (1 + |n|)^r < \infty \right\} \text{ si } \alpha = \infty \end{aligned}$$

En général, $A^{\alpha,r}(\mathbb{T})$ est un espace de distributions de Schwartz, dont on sait définir les restrictions à des intervalles ouverts J ; ces restrictions constituent l'espace $A^{\alpha,r}(J)$. Question : quelle est la borne inférieure des r tels que $\hat{x}|_J \in A^{\alpha,r}(J)$ ($J =]-r, r[$) entraîne $\hat{x} \in A^{\alpha,r}(\mathbb{T})$? Réponse : elle est la même que précédemment, indépendante de α et r .

Pour $r = 0$ et $\alpha = 1$, on retrouve $A(\mathbb{T})$. Pour $r = 0$ et $\alpha = 2$, on trouve $L^2(\mathbb{T})$, et c'est l'origine de la théorie. Sans se borner à des $\Lambda \subset \mathbb{Z}$ mais en considérant des Λ réelles « régulières » au sens que leur pas $\inf_{\lambda \in \Lambda, \lambda' \in \Lambda, \lambda \neq \lambda'} |\lambda' - \lambda|$ est > 0 , Paley et Wiener ont appelé pseudopériode attachée à Λ la borne inférieure des $\ell > 0$ tels que les normes dans $L^2(J)$ des fonctions

$$(53) \quad f(t) = \sum x_\lambda e^{i\lambda t}$$

soient équivalentes pour tous les intervalles J de longueur ℓ . Au facteur 2π près, la valeur de la pseudopériode est la densité uniforme extérieure de Λ [7].

Troisième propriété : la continuité. C'est la question la plus naturelle, mais la plus difficile. Je renvoie à l'étude d'Yves Meyer [11] pour les meilleurs résultats connus.

Quatrième propriété : l'analyticit . Sous une forme un peu diff rente, c'est la premi re question qui a  t  examin e (Fabry en 1896, P lya en 1942), mais la derni re ayant eu sa solution compl tement explicit e [10, II p. 52] [8, p. 199]. Quelle est la borne inf rieure des longueurs $|J|$ telles que, si \hat{x} est analytique sur J , \hat{x} est analytique partout sur \mathbb{T} ? C'est la densit  maximale de S ,   savoir la densit  ext rieure quand, au lieu de (50) et de (51), la densit  D d'une suite Λ est d finie par

$$(54) \quad n(t) = Dt + o(t) \quad (t \longrightarrow \pm\infty).$$

Pólya a montré que la densité maximale est atteinte : il existe une suite contenant S dont la densité (au sens de (54)) est la densité maximale de S .

Lorsque le pas de S est $\geq d$, toutes les densités extérieures de S sont $\geq \frac{1}{d}$. On pourrait donc espérer reconstruire \hat{x} à partir de $\hat{x}|_K$ ($K = [-r, r]$) quand $|K| = 2r > \frac{1}{d}$. Nous avons vu (formule (44)) que c'est possible quand $|K| > \frac{10}{d} \sqrt{\log d}$. On peut améliorer le résultat en diminuant le second membre, mais pas au dessous de $\frac{1}{d}$.

Aucune densité n'est adaptée à la reconstruction d'une fonction \hat{x} appartenant à $A(\mathbb{T})$ à partir d'une restriction $\hat{x}|_K$ par prolongement minimal dans $A(\mathbb{T})$. En effet, soit $\hat{z} \in A(\mathbb{T})$ une fonction nulle sur K . Soit \hat{x} une somme partielle de sa série de Fourier telle que $\|\hat{z} - \hat{x}\|_{A(\mathbb{T})} < \|\hat{x}\|_{A(\mathbb{T})}$; alors le procédé (1) échoue pour \hat{x} , et le support S de x est fini, donc de densité nulle. L'hypothèse naturelle pour cette reconstruction est bien que la pas de S soit convenablement minoré. Le théorème de Candès, Romberg et Tao a le grand mérite de prendre pour hypothèse une majoration de $|S|$ sans rien supposer de son pas; mais il faut alors renoncer à prendre pour K un intervalle, et choisir pour K un ensemble fini aléatoire.

Quittons le cas $\nu = 1$.

2. Pour $\nu \geq 2$, on est loin d'avoir des réponses aussi précises. Par exemple, on ne sait pas pour quelles valeurs de r l'hypothèse

$$(55) \quad \inf_{s \in S, s' \in S, s' \neq s} |s' - s| \geq d,$$

représentant la norme euclidienne, entraîne l'équivalence des normes $\|\hat{x}\|_{L^2(B)}$ pour toutes les boules B de rayon r . La relation entre la géométrie de S et celle des K associés est un problème intéressant, amorcé dans [7].

Appendice. Sur le principe de localisation de Tao.

G et \hat{G} étant deux groupes cycliques d'ordre N en dualité, x une fonction définie sur G à valeurs complexes et \hat{x} sa transformée de Fourier, il s'agit d'établir que

$$(56) \quad \|x\|_{\ell^0} \|\hat{x}\|_{\ell^0} \geq N.$$

Voici la preuve de Tao, ingénieuse et facile.

Soit S le support de x , et $|S| = \|x\|_{\ell^0} = s \neq 0$. Rappelons que

$$\hat{x}(t) = \sum_{n \in S} x_n e^{2\pi i n t / N}.$$

Supposons $\hat{x}(t_1) = \hat{x}(t_2) = \dots = \hat{x}(t_s) = 0$ avec t_1, t_2, \dots, t_s en progression arithmétique, tous distincts modulo N . Les x_n ($n \in S$) sont solutions d'un système linéaire homogène dont le déterminant, du type Van der Monde, est $\neq 0$, donc ils sont tous nuls, contrairement à l'hypothèse. Ainsi, sous l'hypothèse $\|x\|_{\ell^0} = s > 0$, \hat{x} ne peut pas s'annuler en s points formant une progression arithmétique.

Les intervalles d'entiers contigus au support \hat{S} de \hat{x} ont donc au plus s éléments, ce qui entraîne $s|\hat{S}| \geq N$, c'est-à-dire (56).

On a égalité dans (56) lorsque S et \hat{S} sont des translatés de sous-groupes cycliques d'ordre s et \hat{s} , de G et \hat{G} , avec $s\hat{s} = N$.

Lorsque N est premier, les $N - 1$ premiers multiples d'un $t \neq 0 \pmod{N}$ sont tous distincts dans \hat{G} , donc les parties de \hat{G} ayant $N - 1$ éléments sont engendrables comme progressions arithmétiques. Par conséquent, si $s < N$, on a $\hat{s} = N$. Ce fait est utilisé dans [3].

Références

- [1] A. BEURLING and P. MALLIAVIN.— *On Fourier transforms of measures with compact support*, Acta Mathematica 107, 291–309.
- [2] Emmanuel J. CANDÈS.— *Compressive sampling*, Proceedings of the International Congress of Mathematicians, Madrid 2006.
- [3] E.J. CANDÈS, J. ROMBERG and T. TAO.— *Robust Uncertainty Principles : Exact Signal Reconstruction From Highly Incomplete Frequency Information*, IEEE Transactions on Information Theory 52, 2 (2006), 489–509.
- [4] S.S. CHEN, D.L. DONOHO and M.A. SAUNDERS.— *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. vol. 20 n° 1 (1998) 33–61.
- [5] D.L. DONOHO and X. HUO.— *Uncertainty Principles and Ideal Atomic decomposition*, IEEE Transactions on Information Theory 47, 7 (2001), 2845–2862.
- [6] J. FOURIER.— *Histoire de l'Académie pour 1823 p. 28 et pour 1824 p. 47*, dans Œuvres, volume II, pp. 321–328.
- [7] J.-P. KAHANE.— *Pseudo-périodicité et séries de Fourier lacunaires*, Annales scientifiques de l'École Normale Supérieure 79 (1962), 93–150.
- [8] J.-P. KAHANE.— *Séries de Fourier absolument convergentes*, Ergebnisse der Mathematik 50, Berlin, Springer-Verlag 1970.
- [9] J.-P. KAHANE et P.-G. LEMARIÉ-RIEUSSET .— *Séries de Fourier et ondelettes*, Paris Cassini 1998.
- [10] P. KOOSIS.— *The logarithmic integral*, vol. I, II Cambridge University Press 1988–92.
- [11] Y. MEYER.— *Trois problèmes sur les sommes trigonométriques*, Astérisque 1 (1973) 1–85.
- [12] R.E.A.C. PALEY and N. WIENER.— *Fourier transforms in the complex domain*, American Math. Society New-York 1934.

Jean-Pierre Kahane,
Laboratoire de Mathématiques
Université Paris-Sud à Orsay
avril, 2011
Jean-Pierre.Kahane@math.u-psud.fr