Notes et commentaires au sujet des conférences de S. Mallat du Collège de France (2021)

Modèles multi-échelles et réseaux de neurones convolutifs

J.E Campagne *

Janv. 2021; rév. 31 mars 2024

*Si vous avez des remarques/suggestions veuillez les adresser à `jeaneric DOT campagne AT gmail DOT com`

Table des matières

1. Avant-propos

Avertissement: Dans la suite vous trouverez mes notes au style libre prises au fil de l'eau et remises en forme avec quelques commentaires ("ndje" ou bien sections dédiées). Il est clair que des erreurs peuvent s'être glissées et je m'en excuse par avance. Vous pouvez utiliser l'adresse mail donnée en page de garde pour me les adresser. Je vous souhaite une bonne lecture. Veuillez noter également que sur le site associé à ses cours S. Mallat donne en libre accès des chapitres de son livre "A Wavelet Tour of Signal Processing", 3ème edition.

Cette année 2021 c'est la quatrième du cycle de la chaire de la Science des Données de S. Mallat, le thème en est: **Régularité, Approximation et Parcimonie**.

Toujours sous la menace de la COVID-19, les cours reprennent avec un présence sur place modérée.

2. Séance du 13 Janv.

2.1 Introduction du triangle "Régularité, Approximation, Parcimonie"

Tout d'abord présentons le thème de cette année à savoir "***les représentations parcimonieuses***". Si les années précédentes nous avons étudié les *réseaux de neurones profonds* avec leurs applications où nous avons mis en avant certes leurs performances empiriques mais aussi pour ce qui nous concerne une certain manque de support mathématiques pour les comprendre vraiment, cette année nous revenons à une partie du cœur du **Traitement de données**. Si on note $x(u) \in \mathbb{R}^d$ le signal qui nous intéresse (son, image, série temporelle,...), des thèmes classiques en **Traitement du signal** sont:

— l'**Approximation** de ce signal. En effet, on peut vouloir transmettre ce signal avec le moins de bits possible (Th. de l'**Information**) pour obtenir $\tilde{x} \in \mathbb{R}^m$, et on essaye de quantifier l'*erreur* commise (*distorsion* du signal), par exemple à travers une norme $\|x - \tilde{x}\|$. Plus précisément, l'approximation dont il s'agit est à **basse dimension** c'est-à-dire $m \ll d$ car par exemple on veut faire de la **compression** du signal à transmettre.

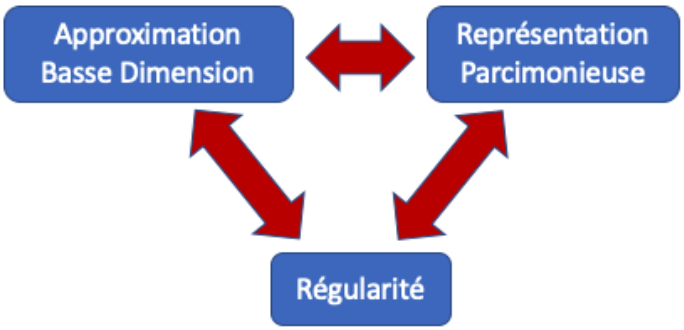

FIGURE 1 – Le triangle RAP: « Régularité, Approximation, Parcimonie ».

— le **Débruitage**. Dans ce cadre x est "contaminé" par un bruit/erreur ε et on essaye de trouver un moyen d'éliminer cette nuisance. Et si le signal peut se représenter sous une forme parcimonieuse alors que le bruit ne le peut, alors on verra que l'on a un moyen de s'en sortir et de quantifier l'erreur effectuée.
— et enfin les **Problèmes Inverses** sur lesquels nous reviendrons.

Dans toutes ces thématiques l'enjeu est de pouvoir récupérer x le plus propre possible. Un autre grand domaine est l'**Analyse**, très lié à ce que l'on nomme l'**Apprentissage Statistique** qui a pour thème de répondre à la question: comment obtenir y à partir de x? c'est-à-dire, que l'on **cherche une fonction** f, telle que $y = f(x)$. Dans ce cadre, on y range les thèmes de:

— la **Classification**: ex. trouver si telle ou telle image est celle d'un chat, d'un bateau... où si tel locuteur est Mme ou M X. Dans ce cas y est un indicateur de classe (entier).
— la **Régression** où dans ce cas y est une quantité continue. Ex. si x est la répartition des atomes d'une molécule, y est son énergie minimale.

La fonction f est l'objet sous-jacent, et on se pose la question de savoir si l'on peut la représenter avec un nombre minimum d'éléments/paramètres à des fins d'apprentissage efficace.

Donc dans ces deux grands thèmes, on va se poser le problème de l'*Approximation en basse dimension* qui est reliée aux *Représentations parcimonieuses*. Ce faisant, on va rencontrer une troisième notion: c'est la **Régularité**. Ces trois notions sont intimement reliées (Fig. 1). Par exemple, quand on prend les représentations parcimonieuses, l'objet d'étude (signal x ou fonction f) est pris dans son ensemble et l'on veut le représenter dans

une "base" avec très peu de coefficients non nuls. En pratique cependant, on ne peut pas penser ces représentations sans la notion d'approximation. En effet, le choix d'annuler des coefficients se fait avec un critère de qualité d'approximation: minimiser l'erreur effectuée. Et *in fine*, quand on découvre des représentations parcimonieuses, on découvre par la même des formes de régularité du signal et la structuration sous-jacente.

L'inter-dépendance des ces trois notions est le sujet du cours de cette année. Du coté des applications, on commencera par les réseaux de neurones, puis on passera au traitement du signal. On illustrera le triangle "RAP"

— tout d'abord dans le **domaine linéaire**. Bien entendu, on retrouvera toute l'analyse harmonique de Fourier. C'est une brique essentielle à bien savoir manipuler qui est nécessaire pour comprendre la suite. On abordera les régularités de Sobolev etc.
— et on passera au **non-linéaire** pour comprendre pourquoi c'est fondamental.

Et il faut avoir à l'esprit qu'à chaque fois que l'on introduit de nouveaux outils, on peut revisiter l'ensemble des notions du triangle RAP: quelles sont les structures qui sont mises en lumière, quels sont les théorèmes d'approximation et les représentations parcimonieuses associées.

Il est clair que le thème de la *parcimonie* n'est pas nouveau. On peut par exemple remonté au *rasoir d'Ockam*[1]. Ce principe philosophique s'applique également en science et consiste, en gros, à éliminer toutes les explications qui sont superflues. On peut également remonter à Aristote qui juge une première démonstration meilleure qu'une seconde, si la première utilise moins d'hypothèses que la seconde. On pourrait continuer à rechercher des usages de cette notion de parcimonie dans les philosophies/sciences au fil des âges. Ce principe d'hypothèse minimale est au cœur de la démarche newtonienne de la construction de modèles qui progressivement se complexifient au fur et à mesure de l'avancée dans la compréhension des phénomènes physiques, et non pas à la recherche de la Vérité avec un grand "V"[2]. Ce que l'on peut en retirer pour le cas qui nous occupe ici et mainte-

1. Guillaume d'Ockham (v. 1285 -1347): philosophe anglais du XIVe siècle, représentant de la scolastique nominaliste qui critique la possibilité d'une *démonstration* de l'existence divine. En cela il est opposé aux thèses de St Thomas d'Acquin qui quant à lui fait une synthèse entre la théologie catholique et la philosophie d'Aristote.

2. NDJE: Isaac Newton est sous l'influence à la fois de Francis Bacon (1561-1626) qui a développé une théorie empiriste de la connaissance, et de Robert Boyle (1627-91) considéré comme le père de la philosophie naturelle moderne. La "philosophie expérimentale" d'inspiration baconienne est tout à fait dans l'air du temps à la Royal Society de Londres. Donc, si les travaux de Newton sont exceptionnels ce n'est pas tant pour l'usage d'une nouvelle méthode révolutionnaire. Cependant, il serait trop long

nant, c'est que l'on a des "mesures" (x) qu'il nous faut expliquer à partir de systèmes de représentations les plus parcimonieux possibles.

Encore quelques autres points pour donner des éléments justifiant l'usage de la parcimonie. Un aspect empirique plutôt d'un penchant tiré de la philosophie anglaise, est le fait que l'on va éviter les "sur-apprentisages": en gros si le nombre d'hypothèses est trop grand face au nombre de mesures, il sera d'autant plus facile d'en donner une explication. Un autre point de vue concerne les erreurs de mesure: minimiser l'erreur de prédiction va être un compromis entre une erreur de modèle, le *biais*, et une *variance* statistique. En compression de données, on a aussi un compromis entre la qualité du signal et le nombre de bits d'information utilisés. Enfin, la parcimonie peut être un guide dans la sélection d'hypothèses pour ne retenir que celles qui ont la plus grande densité d'information. Ce point sera abordé dans le cours à travers la Théorie de l'Information et le concept d'Entropie.

Notons enfin en considérant l'aspect esthétique très présent notamment en mathématiques, on peut se demander: peut-on faire de la parcimonie un *a priori* érigé en principe absolu [3]? on peut citer par exemple qu'en biologie la simplicité n'est pas forcément de mise. Mais dans ce contexte, il est aussi important de comprendre dans quelle situation le système biologique évolue: est-ce que la "simplicité" satisfait l'ensemble des contraintes auxquelles fait face le système (ex. minimisation de l'énergie, adaptation à un éventuel prédateur, etc)? On se rend compte donc assez vite que poser la question de la simplicité/parcimonie n'est possible que pour des systèmes isolés. Dans le cours, on ne se posera que des questions qui sont bien posées.

d'expliquer ici la fameuse maxime *"hypotheses non fingo"* ("je ne fais pas d'hypothèse") qui mêlerait les aspects théologiques de son temps.

3. Notons que le courant de la philosophie analytique émergeant par les travaux de Gottlob Frege (1848-1925), Bertrand Russell (1872-1970) et Ludwig Wittgenstein (1889-1951) formule la science comme un ensemble d'énoncés dont il s'agit de trouver la structure logique et la signification, et dans ce contexte la parcimonie joue un rôle dans la sélection des signes par exemple.

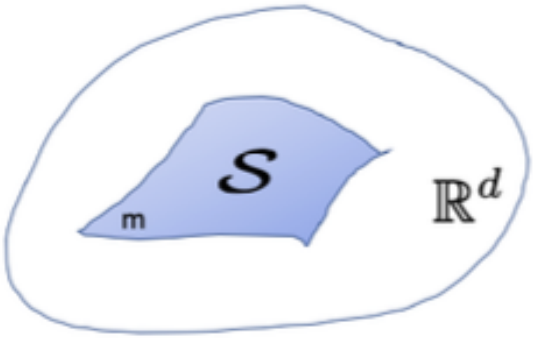

FIGURE 2 – La variété dans laquelle évolue les données $\mathcal{S}$ est paramétrisable par m coefficients.

2.2 Illustrations brève du triangle RAP

2.2.1 Traitement du signal

Dans le cas de la **compression**, on va représenter $x \in \mathbb{R}^d$ par $\Phi(x) \in \mathbb{R}^m$ et on cherche à ce que $m \ll d$, c'est-à-dire que l'information contenue dans le message x peut être réduite à m bits. Ce que cela signifie en sous-jacent, c'est que le signal (les données) n'évolue pas de manière aléatoire dans $\mathbb{R}^d$ mais plutôt sur une variété $\mathcal{S}$ qui est peut être incluse dans $\mathbb{R}^m$ tout du moins qui elle est paramétrisable par m coefficients (Fig. 2). En quelque sorte $\Phi(x)$ est une coordonnée locale de $x \in \mathcal{S}$. Et dans ce contexte, retrouver des structures qui composent le signal/la mesure va aider.

Une fois que l'on comprend que x est contraint de par sa structure à évoluer sur $\mathcal{S}$, alors le *bruitage* consiste à sortir le signal de la surface $\mathcal{S}$. Et une idée alors de **débruitage** est de reprojeter $x+\varepsilon$ sur $\mathcal{S}$ (Fig. 3). Bien entendu il y a une erreur de débruitage ε' mais qui est bien inférieure à ε grâce à cette projection. Les complications viendront de ce que l'espace sous-jacent n'est pas forcément linéaire et dans ce cas il faudra user de projection non-linéaire également. Mais ce que l'on constate est que plus l'espace dans lequel évolue x est petit (c'est-à-dire la dimensionalité de $\mathcal{S}$), plus la suppression du bruit est efficace: plus la représentation est parcimieuse et/ou en basse dimension plus l'efficacité est grande. Un aspect qui nous fait rentrer dans le triangle RAP, c'est que la surface $\mathcal{S}$ ne peut être qu'un modèle, une approximation du lieu géométrique de l'ensemble des x. Donc, de nouveau

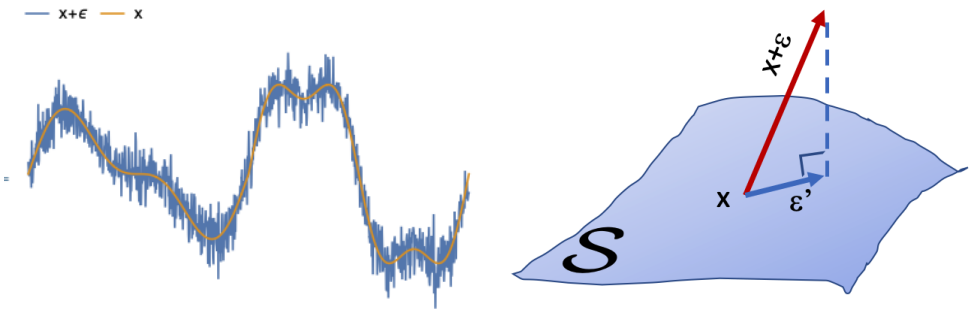

FIGURE 3 – Signal x et le bruit ε et une forme de débruitage par projection orthogonale sur la surface $\mathcal{S}$.

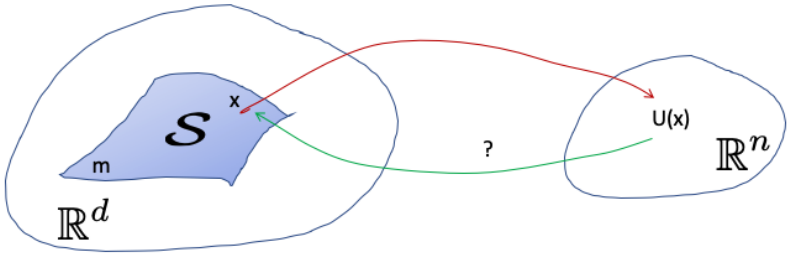

FIGURE 4 – Problème Inverse: à partir des mesures sur x, notées $U(x)$ peut-on retrouver x?

on retrouve deux types d'erreur: une sur le type de modèle car le signal n'évolue pas exactement sur $\mathcal{S}$ et l'autre sur la projection qui laisse un bruit résiduel.

Le troisième type de problèmes à certains égards bien plus important auxquels j'ai fait mention, se sont les **Problèmes Inverses** (Fig. 4). Ici, ce dont on dispose ce n'est pas x mais une mesure de x, notée $U(x) \in \mathbb{R}^n$, où n est le nombre de paramètres mesurés tel que $n \leq d$:

$$x \in \mathbb{R}^d \xrightarrow{U} U(x) \in \mathbb{R}^n \qquad n \leq d \tag{1}$$

Or, dans ce contexte l'opérateur U n'est pas *inversible* (sinon la solution est simple à trouver). Pour s'en sortir, il faut une information *a priori* sur x, en particulier que x se trouve sur une surface $\mathcal{S}$ contenue dans $\mathbb{R}^d$. Car alors on peut tenter une inversion, mais il faut que m le nombre de paramètres qui caractérisent la variété $\mathcal{S}$ soit plus grand que

n, la dimension de l'espace dans lequel évolue $U(x)$. On parlera alors d'inversion de la *restriction* de l'opérateur U à $\mathcal{S}$. Et si la surface $\mathcal{S}$ n'est pas linéaire, quand bien même l'opérateur U l'est (ex. moyenne de mesures), l'inversion est non-linéaire. Ce qui engendre l'utilisation d'algorithmes et de mathématiques beaucoup plus sophistiqués alors que l'on a un opérateur linéaire.

2.2.2 Apprentissage statistique

Dans ce domaine, on se pose la question de trouver une fonction f qui donne la réponse y si on la sollicite par une entrée x: chercher f telle que $y = f(x)$. Que cela soit pour un problème de *classification* (y un entier, ou vecteur d'entiers) ou bien de *régression* (y un réel ou un vecteur de réels). Mettons que $x \in [0,1]^d$ et $y \in \mathbb{R}$, l'espace dans lequel évolue la fonction f est colossal. On peut faire quelques hypothèses, ex. conservation de l'énergie, et f alors appartient à l'espace des fonctions de carré sommable:

$$L^2([0,1]^d) = \left\{ f / \int_{[0,1]^d} |f(x)|^2 dx < \infty \right\} \tag{2}$$

Dans ce cas l'espace peut être muni d'un produit scalaire quasi-euclidien, espace préhilbertien (dimension infinie [4]), ce qui permet de définir une norme entre les fonctions. Ceci dit l'espace $L^2([0,1]^d)$ est tout aussi énorme et pour trouver f il faut user de techniques qui sont assez proches des Problèmes Inverses.

En particulier, à partir d'un lot de données $\{x_i\}_{i \leq n}$ si l'on connait les réponses correspondantes $\{y_i = f(x_i)\}_{i \leq n}$ on est dans le cas d'un **Apprentissage Supervisé**, aussi qualifié de problème d'interpolation. Cependant, il va falloir mettre en œuvre des hypothèses très fortes sur la classe des fonctions de f et disposer de suffisamment d'échantillons (n) pour parvenir à déterminer f en dimension infinie. D'autant plus que l'on est face au fameux problème de l*a malédiction de la dimensionalité* [5].

Si l'on se place du coté des algorithmes, un réseau de neurones à 1 couche cachée à m neurones (Fig. 5) a 3 opérations bien distinctes:

— une *opération linéaire* par l'action d'une matrice $W_{m,d}$ qui peut être vue comme un produit scalaire selon m vecteurs: $W_x = \{\langle x, e_p \rangle\}_{p \leq m}$

4. En dimension finie, c'est un espace euclidien
5. Voir les cours de 2018 et 2019 par exemple.

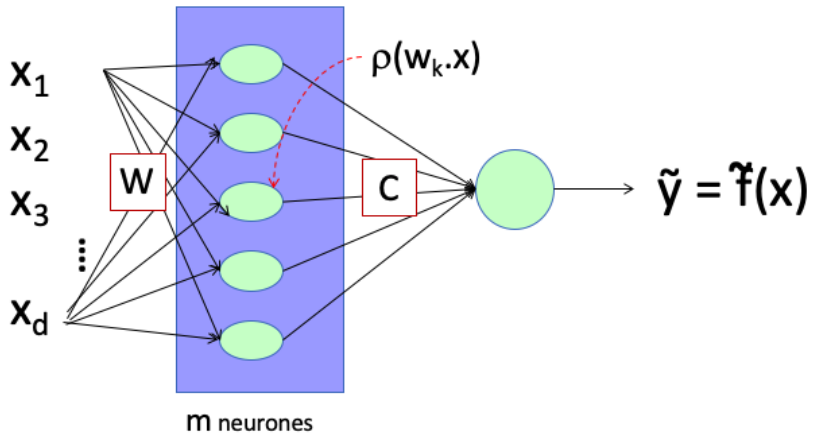

FIGURE 5 – Réseau à 1 couche cachée.

— une n*on-linéarité ponctuelle* ρ, comme un *rectificateur* défini par $\rho(a) = \max(a, 0)$ et il y a d'autres choix possible.
— et au final un *classificateur linéaire* C qui dans le cadre d'une régression est de dimension $(m, 1)$, ou (m, K) pour une classification entre K-classes.

Finalement, on peut écrire en introduisant les biais b_w et b_c au niveau des deux combinaisons linéaires:

$$\boxed{\tilde{f}(x) = C\rho(W.x + b_w) + b_c} \tag{3}$$

ou dans le cas où y est un nombre unique

$$\tilde{y} = \tilde{f}(x) = \sum_p C_p \ \rho(\langle x, e_p \rangle + b_w) + b_c \tag{4}$$

Si l'on oublie les biais pour simplifier les notations, on se rend compte que la réponse $\tilde{y}$ est une combinaison linéaire de fonctions élémentaires:

$$\tilde{y} = \sum_p C_p \ g_p(x) \qquad\qquad g_p(x) = \rho(\langle x, e_p \rangle) \tag{5}$$

C'est-à-dire que pour représenter la fonction f, on a construit *un modèle linéaire simple en relativement basse dimension* (m) grâce aux **fonctions élémentaires** $\{g_p\}_p$ qui sont basées sur l'enchainement d'un **produit scalaire** et d'**une non-linéarité**.

Pour la classification à K classes, y est un label de classe (ex. $y = 1, \ldots, K$) et ce que l'on cherche à approximer c'est $\log p(y|x)$ (logarithme de la probabilité de y sachant x). Car alors, on peut disposer du *classificateur bayésien* qui dit que le meilleur choix pour y est celui pour lequel la probabilité $p(y|x)$ est la plus grande. On peut voir ce type de problème comme K problèmes de régression sur lesquels on applique un max pour obtenir y. Donc dans un premier temps, on ne fera pas de distinction entre un problème de régression (pure) et un problème de classification.

Finalement, faire de l'apprentissage avec un réseau de neurones parcourt le triangle RAP: comprendre de combien de neurones j'ai besoin selon la régularité de la fonction, qu'est-ce que cela va me donner en termes d'approximation de la réponse et peut-être découvrira-t'on que si ça marche c'est que les matrices W et C sont creuses, ce qui est un aspect de la parcimonie. Cependant, est-ce qu'avec 1 couche cachée on arrive à faire le programme auquel on s'attaque? la réponse est en général, non, sauf à devoir prendre un nombre de neurones dans la couche caché colossal (voir le *Théorème d'Universalité d'un réseau à 1-couche cachée* du cours de 2019). Mais en pratique, on constate qu'il y a des cas où cela marche plutôt bien. Qu'est-ce à dire? ça veut dire que f a de la **structure**! Et pourquoi f en a-t'elle répond en quelque sorte à la question de la **régularité** de f.

Quand on aborde des cas où les réseaux à 1 couche ne fonctionne pas, on se tourne vers les **réseaux de neurones profonds** et alors on est contraint à sortir du cadre linéaire. En effet, un réseau profond peut se visualiser comme sur la figure 6 que l'on peut écrire comme une **cascade d'opérateurs** (on oublie les biais ici):

$$\boxed{f(x) = C\ \rho_J W_J \ldots \rho_2 W_2\ \rho_1 W_1 x = C\Phi(x)} \tag{6}$$

dont le résultat est **très non-linéaire**. De nouveau comprendre les réseaux de neurones profonds revient à passer en revue le triangle RAP: quelles sont les régularités des fonctions apprises? quelles sont les structures apprises? la parcimonie joue-t'elle un rôle? a-t'on des théorèmes qui guident le jugement sur les erreurs commises? etc. La difficulté majeure vient du fait que l'on se pose ces questions dans un cadre hautement non-linéaire et en très grande dimension. Et l'on comprend que pour appréhender le non-linéaire, il faut d'abord comprendre ce qu'il se passe quand on passe du linéaire au non-linéaire. Pourquoi est-ce nécessaire d'un coté, mais pourquoi cela vaut le challenge d'un autre coté. Un des résultats est que l'on accède en non-linéaire à des représentations parcimonieuses d'une

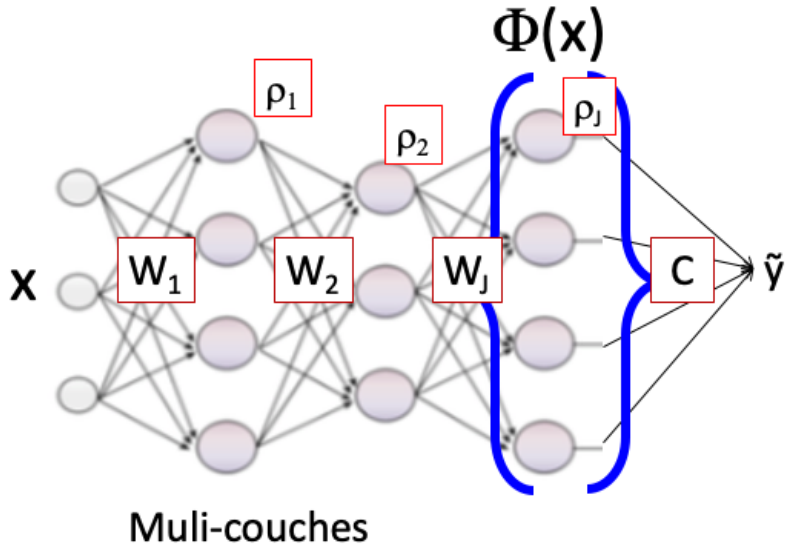

FIGURE 6 – Réseau multi-couches.

qualité bien plus puissante si elles reflètent la régularité sous-jacente du problème. Il y a des cas où le non-linéaire ne fait pas mieux que le linéaire, cependant en général ça marche mieux en non-linéaire et l'on définit des classes de régularités différentes: les variétés sur lesquelles évoluent x sont courbes.

2.3 Plan du cours

Les considérations dans la section précédente vont être approfondies:

— **Le linéaire**[6]: on va donc regarder des approximations de x (traitement de données/du signal) soit de f (apprentissage/analyse) par des projections dans des espaces linéaires. Immédiatement, le premier outil que l'on va rencontrer c'est l'**Analyse de Fourier** (Analyse Harmonique) dès que l'on a un peu de structure. Dans ce cadre, le triangle RAP est complètement compris. La régularité est considérée du point de vue de la décroissance des coefficients de Fourier (esp. de **Sobolev**, etc), on a des théorèmes d'approximation et cela amène naturellement à des représentations parcimonieuses.

6. NDJE: il y a beaucoup de matériel sur cette thématique dans les cours précédents (2018-20).

Cependant, dans certain cas linéaires, on ne connait pas la base de la représentation, et donc on fait appel à des recherches par **Analyse en Composantes Principales** (PCA). On ne reviendra pas sur les algorithmes mais on montrera le lien avec l'Analyse de Fourier, et quelles en sont les limites. Pour ce qui concerne les réseaux de neurones à 1 couche cachée, on va revisiter le **Théorème d'Universalité** qui n'est pas mystérieux, et on reverra sa limitation en grande dimension. Les systèmes de représentations que l'on utilisera seront des **bases** la plupart du temps orthonormales.

— **Le non-linéaire** [7]: on verra également comment faire de l'approximation dans des bases également, avec en particulier la notion de **seuillage** (adaptatif). Cependant, il va falloir trouver de "bonnes" bases, et en particulier celles des analyses de Fourier ou en PCA sont très mauvaises (cf. elles ne sont pas faites pour cela). Ainsi, on sera amené à revoir les **Analyses Multi-résolutions** avec des **bases orthonormales d'Ondelettes**. Et nous verrons que les algorithmes rapides de Transf. en Ondelettes (le pendant de la FFT pour les Ondelettes) ressemblent étrangement à la structure des réseaux de neurones profonds. A partir de là, on va revisiter tout le triangle RAP avec les mêmes notions qu'en linéaire mais dans un cadre non-linéaire: les représentations parcimonieuses seront différentes, on fera des approximations en basse dimension avec des algorithmes différents, ce qui conduira à des classes de régularité différentes qui seront décrites dans le cadre d'espaces plus généraux que Sobolev, à savoir **les espaces de Besov** [8], où les signaux au lieu d'être uniformément réguliers peuvent avoir des singularités donc sont plus complexes [9]. Avec ces outils, on peut appréhender par exemple des images avec des contours, qui possèdent donc des structures.

— la **Théorie de l'Information**. On y arrive dès que l'on veut relier les notions du triangle RAP à des modèles. On le verra dans le cadre de la Compression car alors ce qui importe c'est le nombre de bits et non le nombre de paramètres: la différence? un bit c'est comme son nom l'indique un nombre binaire 0 ou 1, alors qu'un paramètre est en général un réel dont il faut en principe un nombre infini de bit pour le coder. L'enjeu en sous-jacent est la **stabilité** car il faut trouver des

7. NDJE: voir les cours précédents concernant par ex. les Ondelettes.

8. De Oleg Vladimirovich Besov (1933-), mathématicien russe.

9. la mesure des singularités est l'index de l'espace. et la distribution de Dirac est membre de certains espaces de Besov

approximations stables par petits défauts de transmission de bits par exemple. Bien entendu, on verra que le nombre de bits qui permettent de coder une information est relié à l'**Entropie**. C'est la base de la théorie de Claude Elwood Shannon (1916-2001).
On remarque qu'en très grande dimension les processus se concentrent dans des espaces très petits au regard de l'espace initial, se sont les ensembles *typiques* (nom consacré) dont l'Entropie en donne la taille. On verra comment on peut alors accéder à des codes optimaux de compression avec des applications. En particulier on verra des **codes de compression** d'images avec deux standards JPEG et JPEG2000: le premier fait appel essentiellement aux bases de Fourier, et le second aux bases d'Ondelettes. La seconde application sera le **débruitage**, c'est à la fois un problème pratique mais aussi il permet d'identifier l'espace dans lequel évolue x.
On verra des aspects linéaires et non-linéaires du débruitage avec les modèles sous-jacents: l'approche **bayésienne** et l'approche **minimax**. Brièvement, pour représenter des données, il y a une approche purement *déterministe*[10] qui impose un *a priori* qui se résume par : on sait que x appartient à un ensemble $\Theta \in \mathbb{R}^d$. Dans ce cadre, on peut espérer avoir une erreur globale sur l'ensemble Θ la plus petite possible: donc on veut minimiser l'erreur maximale que l'on peut avoir si x parcourt tout l'espace Θ. Ainsi on voit apparaitre la notion de `minimax`:

$$\min \max_{x \in \Theta} \tag{7}$$

En sous-jacent, les modèles bayésiens sont *probabilistes*, ce qui pourrait paraître paradoxal car qui dirait proba. dirait que l'on a une certaine incertitude. En fait, c'est vraiment le contraire, car si l'on dispose d'un modèle probabiliste c'est que l'on dispose d'énormément d'information pour construire la probabilité que x se trouve à tel ou tel endroit de l'espace Θ ($p(x)$). Cependant, en pratique on a quasiment jamais accès au $p(x)$. C'est pourquoi on utilise des modèles `minimax` pour obtenir des résultats rigoureux, car l'idée est de prendre le "cas du pire".

Avant de conclure cette section, envisageons un point sur lequel il faut attirer l'atten-

10. NDJE: voir une discussion sur bayésien vs déterminisme dans le cours de 2019.

tion. Pour le moment, on a parlé pour l'espace dans lequel évolue x en termes de surfaces, et de variétés. En très grande dimension, cf. $\mathbb{R}^d$ avec $d \gg 1$, la surface par elle-même est de très grande dimension aussi, et mathématiquement on va la caractériser comme un *processus aléatoire*. En tout état de cause, on n'est pas dans le cas d'un espace de dimension 3 dans lequel on projetterait le signal sur une surface de dimension 2, et clairement les propriétés en dimension quasi-infinie (voire infinie) ne sont pas les mêmes qu'en basse dimension.

3. Séance du 20 Janv.

Dans cette séance, nous allons toucher du doigt que le triangle Régularité, Approximation et Parcimonie (RAP) se décline différemment si on se place dans un cadre **linéaire** ou un cadre **non-linéaire**. Deux types d'objets vont être utilisés, soit des **données** au sens large que l'on note $x(u)$ indexées par u comme par exemple le temps en 1D, la positions des pixels d'une image en 2D etc, soit une **fonction** f qui répond à la question $y = f(x)$. Donc selon le domaine, l'objet pour lequel on voudra une approximation qui bénéficiera d'une bonne représentation parcimonieuse selon sa régularité, sera donc soit x soit f, et à chaque fois il faudra bien se poser la question de savoir quel est l'objet d'étude.

Ceci dit dans la plupart des cas du cours on se fera la main sur $x(u)$ vue comme une fonction de u et pour les réseaux de neurones on reviendra à la notation f.

3.1 Un problème simple (cadre linéaire)

Soit la fonction $x(u)$ régulière dont le graphe est celui de la figure 7, et le problème que l'on se pose est de représenter cette fonction avec *le moins de paramètres* possible. Ce qui peut venir à l'esprit c'est d'échantillonner régulièrement cette fonction que l'on note $\{x(nT)\}_{n \leq N}$ et faire une interpolation régulière $\tilde{x}(u)$ entre ces valeurs. L'erreur commise dans l'approximation de x par $\tilde{x}$ peut être quantifiée selon l'écart quadratique (norme $L2$):

$$\|x - \tilde{x}\|^2 = \int |x(u) - \tilde{x}(u)|^2 \, du \tag{8}$$

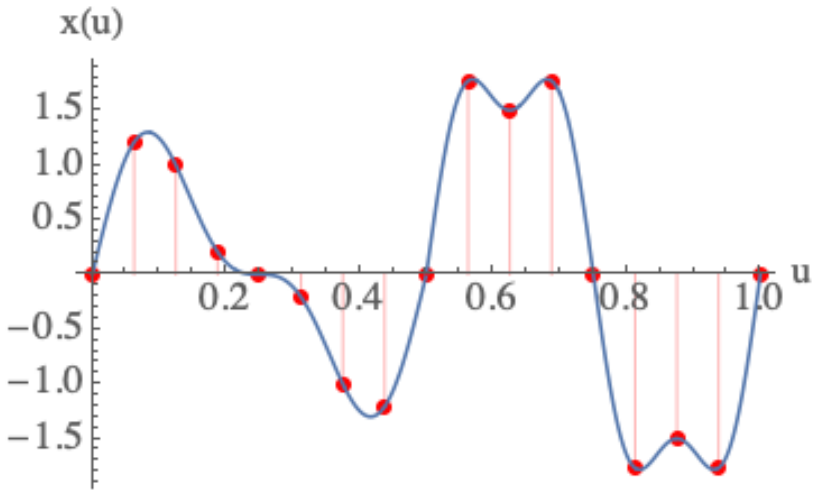

FIGURE 7 – Une fonction $x(u)$ régulière et un échantillonnage régulièrement espacé.

Ce type d'approximation est *linéaire*, attention on ne dit pas que l'on utilise des polynômes de degré 1, cela veut dire que, si on note $\tilde{x}_1 \sim x_1$ le fait que $\tilde{x}_1$ est une approximation de x_1 alors par combinaison linéaire

$$\left.\begin{array}{l} \tilde{x}_1 \sim x_1 \\ \\ \tilde{x}_2 \sim x_2 \end{array}\right\} \Rightarrow \lambda_1\tilde{x}_1 + \lambda_2\tilde{x}_2 \sim \lambda_1 x_1 + \lambda_2 x_2 \tag{9}$$

On est dans un cadre linéaire, et l'approximation $\tilde{x}$ est caractérisée par $M = 1/T$ paramètres (si le support est $[0, 1]$), elle se trouve donc dans un espace V_M de dimension $\dim(V_M) = M$. Ainsi, on a trouvé une projection $\tilde{x}$ de la fonction x dans cet espace V_M. Mais on veut minimiser l'erreur quadratique Eq. 8:

$$\underset{\tilde{x}\in V_M}{\text{Min}} \ \|x - \tilde{x}\|^2 \tag{10}$$

Or, on sait que la solution à ce problème de minimisation conduit à la projection orthogonale de x sur l'espace linéaire V_M (Fig. 8):

$$\tilde{x} = P_{V_M}\ x \tag{11}$$

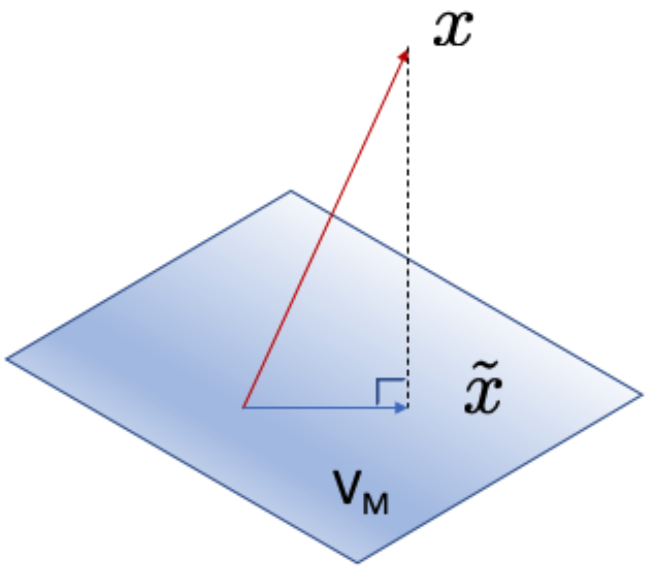

FIGURE 8 – Illustration d'une approximation linéaire $\tilde{x}$ comme le résultat de la projection orthogonale de x sur l'espace linéaire V_M.

Ceci est un résultat général dans le cas linéaire: **l'approximation qui minimise l'erreur quadratique est la projection orthogonale sur l'espace linéaire considéré.** Notons au passage que dans le triangle RAP, on se trouve du coté de l'Approximation.

Comment calcule-t'on une projection orthogonale? une manière simple est d'utiliser une base orthonormale de l'espace global de dimension d, laquelle existe toujours (rappel ici on est en dimension finie). Soit [11]

$$\mathcal{B} = \{e_i\}_{i \leq d} \qquad \text{tq.} \quad \langle e_i, e_j \rangle = \delta_{ij}^K = \delta^D[i-j] \tag{12}$$

On peut réarranger la base pour que

$$\forall \tilde{x} \in V_M, \quad \tilde{x} = \sum_{i=1}^{M} \alpha_i e_i \tag{13}$$

11. NDJE: δ^K est le symbole de Kronecker, et δ^D celui de Dirac. S. Mallat utilise la notation de Dirac donc par la suite la notation δ^D omettra le D.

La projection orthogonale de x sur V_M est alors [12]

$$\tilde{x} = \sum_{i=1}^{M} \langle x, e_i \rangle e_i \tag{14}$$

Maintenant comment obtenir des bases qui s'adaptent au fait que M peut varier? on peut d'abord se placer dans l'espace des fonctions à support $[0, 1]$ et de carré sommable $L^2([0, 1])$, notons que l'on peut se placer également dans l'espace des fonctions à support dans $\mathbb{R}$ de carré sommable $L^2(\mathbb{R})$. Ce sont des espaces pour lesquels on peut définir un produit scalaire entre 2 fonctions:

$$\langle x, \tilde{x} \rangle = \int_{[0,1] \text{ ou } \mathbb{R}} x(u)\tilde{x}^*(u)du \tag{15}$$

Dans ce type d'espace de *dimension infinie*, on construit une base orthonormale $\mathcal{B} = \{e_n\}_{n\in\mathbb{N}}$ telle que

$$\langle e_i, e_j \rangle = \delta[i - j] \tag{16}$$

et on a le résultat suivant sur l'erreur des projections

$$\forall x \in L^2, \quad \lim_{M\to\infty} \left\| x - \sum_{i=1}^{M} \langle x, e_i \rangle e_i \right\|^2 = 0 \tag{17}$$

Comme en dimension finie, l'espace de projection de dimension M (finie) V_M est généré par les M premiers vecteurs de la base $\{e_i\}_{i\leq M}$, quant à l'erreur elle se projette dans l'espace complémentaire:

$$x - P_{V_M} x = \sum_{i>M} \langle x, e_i \rangle \ e_i \tag{18}$$

L'avantage d'utiliser une base orthonormale, c'est que les erreurs se calculent facilement. En effet, c'est cette orthonormalité qui permet d'écrire:

$$\|x - P_{V_M} x\|^2 = (x - P_{V_M} x).(x - P_{V_M} x) = \sum_{i>M} |\langle x, e_i \rangle|^2 = \varepsilon_\ell \tag{19}$$

12. NDJE: *hint*: on écrit $\|x - \tilde{x}\|^2$ comme le produit scalaire de $x - \tilde{x}$ par lui-même, on écrit une décomposition de $\tilde{x}$ sur la base de V_M constituée par les M premiers vecteurs de $\mathcal{B}$. Ainsi on a une forme quadratique pour chaque α_i qui est minimale ssi $\alpha_i = \langle x, e_i \rangle$.

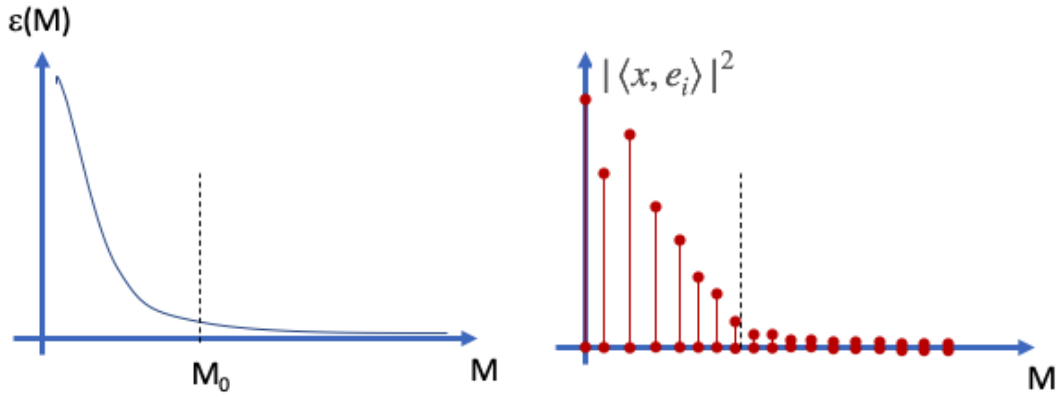

FIGURE 9 – (gauche): On aimerait que l'erreur de l'approximation (linéaire) puisse décroitre suffisamment rapidement pour pouvoir mettre le seuil M_0 assez bas. Cela se traduit par une contrainte sur l'énergie stockée dans les premiers coefficients (droite).

On voudrait que l'erreur d'approximation, notée ε_ℓ (ℓ rappelant "linéaire"), soit la plus petite possible. Bien entendu, on voit sur son expression précédente qu'elle dépend de M (plus M est grand, plus petite elle est), et d'ailleurs on sait que

$$\varepsilon_\ell(M) \xrightarrow[M\to\infty]{} 0 \tag{20}$$

Bien entendu, on aimerait (Fig. 9 gauche) que la décroissance vers 0 soit la plus rapide possible afin de pouvoir mettre un seuil M_0 pas trop grand et tel que le reste de la série $\sum_{i>M_0}$ soit une petite quantité. On peut formuler la demande comme étant que l'énergie dans tous les produits scalaires au-delà de M_0 soit petite, mais on sait que pour n'importe quelle base orthonormale, on a la loi de conservation de l'énergie de x qui se traduit par

$$\|x\|^2 = \sum_{i=1}^{\infty} |\langle x, e_i\rangle|^2 \tag{21}$$

donc il faut que **l'énergie des premiers coefficients soit la plus importante** (Fig. 9 droite) et donc que la décroissance [13] de la suite $(|\langle x, e_n\rangle|^2)_n$ vers 0 soit rapide:

$$|\langle x, e_n\rangle|^2 \xrightarrow[n\to\infty]{\text{rapide}} 0 \tag{22}$$

13. Notons que la suite est convergente car $\|x\|^2$ est finie.

Bien qu'est-ce à dire? la première chose que l'on a mis en lumière c'est l'usage **d'une représentation** dans la base orthonormale $\mathcal{B}$. Et la contrainte sur les produits scalaires $|\langle x, e_n \rangle|$ à décroitre rapidement, c'est la notion de **parcimonie**, car dans la représentation il n'y a qu'un **petit nombre de coefficients** qui supporte l'énergie du signal, si l'on veut pouvoir en faire une **approximation linéaire**. **Et donc procéder à une projection orthogonale revient à mettre à 0 les coefficients au-delà d'un rang** M_0. Or, la clé pour être capable de réaliser cette approximation sans faire trop d'erreur, c'est que l'on a fait un *a priori* sur la **régularité de la fonction** $x(u)$.

Quelles sont les questions que l'on peut se poser dans le cadre **linéaire**:

— quelle est cette notion de régularité?

— quels sont les espaces optimaux V_M d'approximation?

— et finalement quelle est la base orthonormale $\mathcal{B}$ qui réalise la meilleure représentation?

En répondant à ces questions on va trouver l'analyse de Fourier, les espaces de Sobolev, et on verra le problème de la malédiction de la dimensionalité.

3.2 Un problème moins simple (cadre non-linéaire)

L'exemple de la section précédente est certes générique mais ne remplit pas, loin sans faut, tous les cas de figures même pour des fonctions de carré sommable. Par exemple, considérons le cas à 1 dimension de la figure 10. En 2D, dans une image, on pourrait penser au cas où d'un pixel à son voisin, on passe rapidement d'un extrême à l'autre de l'échelle des nuances de gris, à cause des contours de tel ou tel objet sur un fond uni... Souvent **l'information "pertinente" se trouve dans les discontinuités**.

Comme pour le cas linéaire, on peut commencer par faire un échantillonnage régulier à M échantillons (Fig. 10 gauche). On obtient une interpolation régulière $\tilde{x}$ de la fonction x, et $\tilde{x}$ est le résultat d'une projection orthogonale sur une espace linéaire de dimension M (V_M). Ceci dit $\tilde{x}$ est prise parmi les fonctions régulières, donc **les erreurs sont principalement concentrées aux endroits des singularités**. Donc, il faut changer le lieu des échantillons, certes en garder M, mais les concentrer au niveau des singularités. Ainsi **l'échantillonnage s'adapte** au cas par cas selon x (Fig. 10 droite). L'approximation qui fait passer de x à $\tilde{x}$ est alors **non-linéaire**: si j'ai une fonction x_1 qui est approximée par $\tilde{x}_1$, et une autre fonction x_2 elle approximée par $\tilde{x}_2$, la fonction $\alpha x_1 + \beta x_2$ n'est

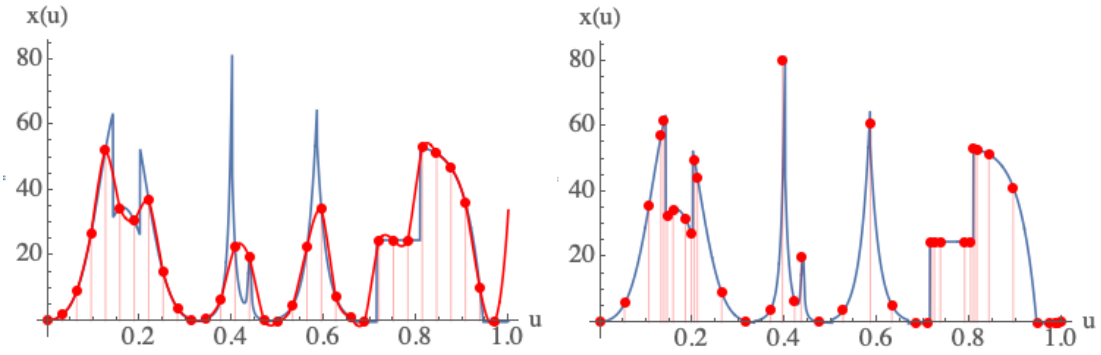

FIGURE 10 – Exemple de l'utilisation d'un échantillonnage régulier à gauche, et adaptatif à droite qui permet de mieux cerner les singularités de la fonction sous-jacente.

pas approximée par $\alpha\tilde{x}_1 + \beta\tilde{x}_2$, à cause de l'adaptation de l'échantillonnage pour chaque fonction (ici M reste constant).

Soit, mais comment choisir le bon échantillonnage? certes on va essayer de localiser les singularités mais comment? et si on y arrive comment répartir les M échantillons en proportion des singularités? et il nous faut le faire pour n'importe quelle fonction singulière (au moins pour une classe d'entres elles). Un angle d'attaque qui a été la clé d'entrée pour résoudre ces problèmes n'est pas très différent de celui du problème précédent. Partons d'une base orthogonale $\mathcal{B}$, on sait qu'alors x se décompose selon

$$x = \sum_{i=0}^{\infty} \langle x, e_i \rangle e_i \tag{23}$$

Maintenant, l'approximation $\tilde{x}$ à M paramètres est aussi décomposable sur la base

$$\tilde{x} = \sum_{i \in \mathcal{S}} \langle x, e_i \rangle e_i \tag{24}$$

le point est que on va faire le choix de $\mathcal{S}$ avec $|S| = M$. **C'est une somme partielle mais pas uniquement avec les M premiers coefficients.**

Comment va-t'on les choisir? ce que l'on veut c'est minimiser l'erreur quadratique, en l'occurrence bénéficiant de la base orthonormale, on a facilement une estimation de

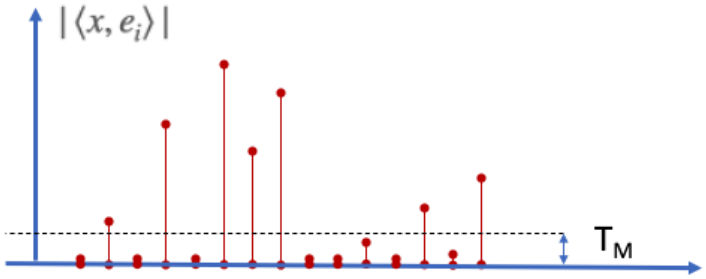

FIGURE 11 – Exemple de seuillage des produits scalaires pour obtenir un échantillonnage adaptatif.

cette erreur:

$$\|x - \tilde{x}\|^2 = \sum_{i \notin \mathcal{S}} |\langle x, e_i \rangle|^2 \tag{25}$$

Donc le problème revient à *minimiser cette erreur sous contrainte* que l'on a M paramètres. Mais pour que l'erreur soit petite, il faut que $\mathcal{S}$ soit l'ensemble des M plus grands coefficients:

$$\mathcal{S}(x) = \{i \in \mathbb{N} \ / \ |\langle x, e_i \rangle| \text{ les } M \ \text{ plus gds}\} \tag{26}$$

En fait, il faudrait pouvoir ordonner les produits scalaires $|\langle x, e_i \rangle|$ du plus grand au plus petit et prendre les M premiers. Nous verrons comment procéder, mais on peut reformuler comment obtenir $\mathcal{S}(x)$, en introduisant un **seuillage**, noté T_M qui assure que l'on obtient bien M coefficients et pour lequel (Fig. 11):

$$\mathcal{S}(x) = \{i \in \mathbb{N} \ / \ |\langle x, e_i \rangle| \geq T_M\} \tag{27}$$

La simplicité de la mise en œuvre de ce schéma bien qu'il soit non-linéaire vient du fait que la **base $\mathcal{B}$ est orthonormale**. Mais de quelle base parle-t'on? car dans le cas présent il s'agit de pouvoir décrire des fonctions ayant des singularités. Au passage, nous verrons que la base optimale dans le cadre linéaire est celle de Fourier, alors qu'elle n'est pas du tout capable de s'adapter aux singularités des fonctions dont on parle en non-linéaire. Donc, il faudra trouver autre chose, et nous verrons que les **bases d'Ondelettes** remplissent le contrat. Maintenant, est-ce que cela vaut le coup? ou autrement dit **l'erreur**

d'approximation, une fois que l'on passe à un échantillonnage adaptatif, diminue-t'elle franchement? la réponse à cette question dépendra de **la régularité des fonctions**. La notion de régularité, ici en non-linéaire, est beaucoup plus large. Et finalement, les fonctions du type de celle de la figure 10 ne sont pas si "irrégulières" que cela, car malgré leurs discontinuités ici ou là, finalement il n'y en a peu. Des fonctions franchement irrégulières sont par exemple celles qui décrivent un mouvement brownien où en tout point, on a une singularité.

Donc, on parcourra le triangle RAP dans le contexte non-linéaire pour découvrir de nouvelles bases, de nouvelles régularités et de nouveaux théorèmes d'approximation. Sachant que passer du linéaire au non-linéaire se fait quand on passe d'un réseau à 1 couche cachée à un réseau de neurones profond.

Dans la suite on va d'abord se concentrer sur le linéaire car il est fondamental pour comprendre le non-linéaire en ce sens que l'on recyclera les mêmes idées en les adaptant. Et plus précisément on va attaquer le triangle RAP du point de vue de la régularité.

3.3 Qu'est qu'une fonction régulière (linéaire)?

A partir du point de vue de la **régularité**, on se pose les questions suivantes: peut-on construire les approximations, construire des bases orthonormales, savoir si elles sont optimales ou pas, etc.

Dans un premier temps, prenons $x(u)$ comme une série temporelle. On peut par exemple appréhender sa régularité à travers ses dérivées. Si x est différentiable, on sait qu'elle est déjà continue et que ses variations sont régulières, et si de plus sa dérivée première est plus petite qu'une constante

$$\left|\frac{dx(u)}{du}\right| \leq C \tag{28}$$

alors on peut se dire que la fonction $x(u)$ est plutôt régulière. Si on veut des fonctions encore plus régulières, il faut des fonctions dont les dérivées d'ordre supérieur existent, et qu'elles soient également bornées:

$$\forall k \leq n \quad \left|\frac{d^k x(u)}{du^k}\right| \leq C \tag{29}$$

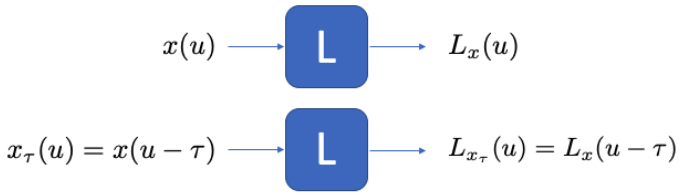

FIGURE 12 – Opérateur covariant/équivariant par translation. L'opérateur de dérivée fait parti de ce type d'opérateurs.

Cette façon de procéder est la façon la plus naturelle. **Et bien juste à partir de cette vision de la régularité, on peut dérouler tout le triangle RAP**.

Dans un premier temps, on se pose la question: qu'est-ce qu'un opérateur de dérivée? on sait qu'à $x \in L^2$, il donne la dérivée dx/du et que celle-ci est bornée. Pour le comprendre, on va le **diagonaliser**. Mais faisons une remarque qui fera lien avec le cours de 2020 [14], l'opérateur dérivé fait parti d'une large classe d'**opérateurs covariant/équivariant par rapport à la translation**:

$$\begin{aligned} g.x(u) = x(u-\tau) \Rightarrow D_u(g.x(u)) = \frac{dx(u-\tau)}{du} = \frac{dx}{du}(u-\tau) = g.D_u(x(u)) \\ \Rightarrow D_u(g.x) = g.D_u(x) \end{aligned} \tag{30}$$

(nb. un opérateur invariant par translation satisfait $f(g.x) = f(x)$). On peut représenter cet opérateur covariant par translation selon le schéma de la figure 12. Dans la suite, on utilisera uniquement le terme *covariant* bien que cela soit l'*équivariance* dont il s'agit. Notons que cette propriété est assez naturelle en traitement du signal, si l'on veut transmettre une série temporelle, on veut des opérateurs qui gardent l'enchainement temporel des valeurs. Donc un décalage à l'entrée doit se traduire par le même décalage à la sortie. Concernant la dérivée, cela revient à constater que la dérivée d'un signal translaté dans le temps, est elle même translatée dans le temps (avec le même décalage temporel).

Bien, maintenant la fonction $x(u)$ peut être vue comme une somme de Dirac:

$$x(u) = \int x(v)\delta(u-v)dv \tag{31}$$

14. NDJE: D'ailleurs la Sec. 6 du cours de 2020 peut servir de complément.

FIGURE 13 – L'opérateur linéaire covariant par translation noté L a pour vecteur propre les fonctions $e^{i\omega u}$ dont les valeurs propres associées sont les valeurs de sa Fonction de Transfert estimée en ω (cf. $\hat{h}(\omega)$).

Sans faire de théorie des distributions, on se rappelle que le Dirac est la limite de fonctions dont la "masse" se concentre en un point (intégrale est égale à 1 tout en ayant un support qui tend vers 0). Si on applique un opérateur L covariant par translation, si on a un minimum de régularité pour inverser l'ordre entre l'intégrale et l'action de l'opérateur L:

$$L.x(u) = \int x(v)L.[\delta(u-v)]dv = \int x(v)(L.\delta)(u-v)dv \tag{32}$$

En traitement du signal la fonction $L.\delta(u) = h(u)$ est la **réponse impulsionnelle de l'opérateur** L et donc par covariance de L on a:

$$\boxed{L.x(u) = \int x(v)h(u-v)dv = (x*h)(u) = (h*x)(u) = \int x(u-v)h(v)dv} \tag{33}$$

qui se traduit par le fait que **l'action d'un opérateur linéaire covariant par translation** L **sur** x **est une convolution de** x **par la réponse impulsionnelle de l'opérateur**. Donc, l'opérateur dérivée fait partie de la classe des opérateurs convolutionnels. Quels sont ses vecteurs propres? Prenons à l'entrée de l'opérateur la fonction exponentielle oscillante $e^{i\omega u}$, il vient

$$\boxed{L[e^{i\omega u}] = \int e^{i\omega(u-v)}h(v)dv = e^{i\omega u}\int h(v)e^{-i\omega v}dv = e^{i\omega u}\ \hat{h}(\omega)} \tag{34}$$

Donc, primo $e^{i\omega u}$ **est un vecteur propre** de l'opérateur linéaire L et secundo la valeur propre associée est $\hat{h}(\omega)$ c'est-à-dire **la transformée de Fourier de la réponse impulsionnelle de l'opérateur**, soit **la fonction de transfert**. On peut le schématiser comme sur la figure 13. Notons que selon la valeur de $\hat{h}(\omega)$ on peut mettre en évidence des phénomènes de résonance. Ainsi, point important à retenir: **la Transformée de Fourier permet de**

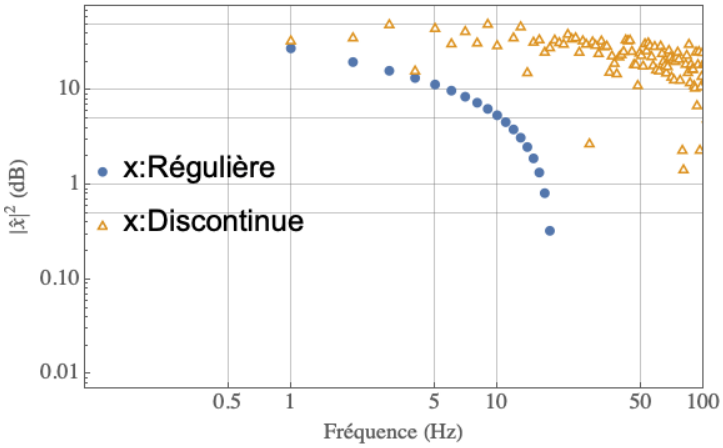

FIGURE 14 – Évolution des coefficients de Fourier des deux fonctions étudiées dans les sections 3.1 (régulière/cadre linéaire) et 3.2 (discontinue/cadre non-linéaire).

diagonaliser les opérateurs de convolution.

3.4 Analyse de Fourier

Envisageons de revoir quelques résultats sur l'**Analyse de Fourier** [15]. C'est un chapitre de mathématiques qui s'est clos grosso-modo dans les années 1960 avec les derniers théorèmes de convergence des intégrales et des séries de Fourier.

On définit la Transformée de Fourier de x selon

$$\hat{x}(\omega) = \int_{-\infty}^{+\infty} x(u) e^{-i\omega u} du \tag{35}$$

pour s'assurer que l'intégrale a un sens, on peut se limiter aux fonctions de $L^1(\mathbb{R})$ pour lesquelles:

$$\int |x(u)| du < \infty \tag{36}$$

Maintenant l'interprétation de $\hat{x}(\omega)$: c'est le résultat de la corrélation entre la fonction $x(u)$ et les sinusoïdes dont ω fixe la fréquence d'oscillation. Si la fonction x est régulière,

15. On peut également revoir ce thème dans les cours des années précédentes tellement il est fondamental de le maitriser.

elle oscille peu et donc ses coefficients $\hat{x}(\omega)$ sont grands pour des ω petits. A contrario, plus x a des variations rapides, plus les coefficients "hautes fréquences" seront importants. Ainsi, les basses fréquences rendent compte si une fonction est régulière. Ou en d'autres termes, **la décroissance des coefficients de Fourier rend compte de la régularité de la fonction**. Une illustration est donnée sur la figure 14 avec les deux fonctions utilisées dans les sections 3.1 et 3.2. La convergence vers 0 des coefficients de Fourier de la fonction régulière est bien plus rapide.

Pour le premier théorème [16] on va supposer que la fonction $\hat{x}(\omega)$ est aussi $L^1(\mathbb{R})$ c'est-à-dire que l'on n'a pas trop de composantes à hautes fréquences.

Théorème 1 *Si $\hat{x} \in L^1(\mathbb{R})$ alors on a une formule d'inversion qui permet de retrouver la fonction x à partir de sa transformée de Fourier:*

$$x(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{x}(\omega) e^{i\omega u} d\omega \tag{37}$$

Notons qu'il faut se rendre compte de cette opération de recombinaison des sinusoïdes en les pondérant par les coefficients de Fourier n'est pas évidente. Car les sinusoïdes sont délocalisées et il faut une pondération très précise pour redonner par exemple une fonction essentiellement régulière sauf peut-être un peu plus agitée sur une faible portion de son support (voir par ex. la fonction de la Figure 22 du cours de 2020). Une conséquence:

$$|x(u)| = \frac{1}{2\pi} \left| \int_{-\infty}^{+\infty} \hat{x}(\omega) e^{i\omega u} d\omega \right| \leq \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{x}(\omega)| d\omega \tag{38}$$

Comme $\hat{x} \in L^1(\mathbb{R})$ alors le membre de droite est borné et donc il en est de même pour x, elle est de **variations bornées**, et

$$\|x\|_\infty = \sup_{u \in \mathbb{R}} x(u) \leq \frac{1}{2\pi} \int |\hat{x}(\omega)| d\omega \tag{39}$$

Le second théorème découle de l'intention qui a été portée pour utiliser la transformée de Fourier, à savoir la convolution:

16. Des démonstrations sont fournies dans les notes associées au cours que fournit S. Mallat.

Théorème 2

$$g(u) = (x * h)(u) \rightarrow \hat{g}(\omega) = \hat{x}(\omega)\hat{h}(\omega)$$

Ce qui revient à dire que le produit de convolution est diagonalisé en Fourier. En effet, l'opérateur de convolution $Lx(u) = (x * h)(u)$ est linéaire et ***covariant* par translation** car

$$\begin{aligned} L[g_\tau.x(u)] &= L[x(u-\tau)] = L[f(u)] = \int f(v)h(u-v)dv \\ &= \int x(v-\tau)h(u-\tau-v+\tau)dv = \int x(v)h(u-\tau-v)dv = Lx(u-\tau) \\ &= g_\tau.(L_x(u)) \end{aligned}$$

donc il est diagonalisable par les exponentielles complexes comme on l'a montré précédemment. Donc

$$g(u) = (x * h)(u) = Lx(u) = \frac{1}{2\pi}\int \hat{x}(\omega)L[e^{i\omega u}]d\omega = \int \hat{x}(\omega)\hat{h}(\omega)e^{i\omega u}d\omega$$

d'où le résultat en identifiant $\hat{x}(\omega)\hat{h}(\omega)$ comme le coefficient de Fourier $g(\omega)$.

Le troisième théorème (fondamental) c'est la formule de Plancherel qui reflète que la Transformation de Fourier conserve les angles (à une facteur prés):

Théorème 3

$$\langle x_1, x_2\rangle = \int x_1(u)x_2^*(u)du = \frac{1}{2\pi}\int \widehat{x_1}(\omega)\widehat{x_2}(\omega)d\omega = \frac{1}{2\pi}\langle \widehat{x_1}, \widehat{x_2}\rangle \tag{40}$$

Une conséquence si $x_1 = x_2$ donne une relation (Parseval) qui est une conservation d'énergie:

$$\|x\|^2 = \|\hat{x}\|^2 \Leftrightarrow \int |x(u)|^2du = \frac{1}{2\pi}\int |\hat{x}(\omega)|^2d\omega \tag{41}$$

Donc, si $x \in L^2(\mathbb{R})$ alors sa transformée de Fourier est aussi d'énergie finie, $\hat{x} \in L^2(\mathbb{R})$, **la TF préserve l'espace** $L^2(\mathbb{R})$ et nous la définirons principalement dans cet espace.

NDJE. D'autres propriétés de la TF sont utiles pour la suite, voir par exemple le cours de 2018.

3.5 L'opérateur dérivée: régularité de Sobolev

Revenons à la dérivée et tout d'abord: quels sont ses vecteurs propres? comme elle est covariante par translation, d'après l'étude de la section précédente, alors ses vecteurs propres sont les $e^{i\omega u}$. La valeur propre est simplement $i\omega$, ou en termes de fonction de transfert de l'opérateur dérivée est:

$$\hat{h}_{d/du}(\omega) = i\omega \tag{42}$$

Si on veut dire maintenant que la dérivée d'ordre k de x est bornée, à savoir

$$\left\| \frac{d^k x(u)}{du^k} \right\|_\infty \leq C \tag{43}$$

alors une manière de procéder consiste à imposer que la TF de la dérivée d'ordre k soit intégrable (voir Eq. 39), mais alors, sachant que:

$$\frac{d^k x(u)}{du^k} \xrightarrow{TF} (i\omega)^k \hat{x}(\omega) \tag{44}$$

cela revient à dire qu'on doit imposer la condition

$$\int |\omega|^k |\hat{x}(\omega)| d\omega = C \tag{45}$$

C'est-à-dire que la condition précédente dans le domaine de Fourier *implique* la condition sur la dérivée dans l'espace réel. On entrevoit **une condition de régularité** de la fonction x, car si les coefficients de Fourier décroissent suffisamment rapidement pour que, multipliés par ω^k, l'intégrale reste finie, alors cela donne une contrainte sur les variations de la dérivée d'ordre k.

Peut-on avoir une équivalence? pour cela faisons appel à **Parseval/Plancherel** à savoir on va remplacer la condition "la dérivée d'ordre k est bornée" par la condition "la dérivée d'ordre k est de carré sommable", à savoir

$$\int \left| \frac{d^k x(u)}{du^k} \right|^2 du = \frac{1}{2\pi} \int ||\omega|^k \hat{x}(\omega)|^2 d\omega = \frac{1}{2\pi} \int |\omega|^{2k} |\hat{x}(\omega)|^2 d\omega \leq \infty \tag{46}$$

Donc là il y a une équivalence entre le fait que si l'intégrale de droite converge, alors la

dérivée d'ordre k est de carré sommable. C'est **la régularité ou la dérivabilité au sens de Sobolev**. En fait, l'exposant peut être étendu au cas d'un réel positif, et la condition devient alors

$$s \in \mathbb{R}^+, \ \int |\omega|^{2s} |\hat{x}(\omega)|^2 d\omega \leq \infty \tag{47}$$

Si l'intégrale converge, on dira que la fonction est dérivable *"s fois"* avec s réel positif. On généralise par là la notion de dérivée, et surtout **en pratique en traitement du signal, on ne calcule jamais les dérivées** mais on va dans le domaine de Fourier pour étudier **la décroissance des coefficients**. Maintenant, si la TF décroit suffisamment rapidement, on peut mettre une coupure à plus basse fréquence, et ne garder qu'un petit nombre de coefficients, donc avoir une **représentation parcimonieuse**. Le seul point à éclaircir encore c'est que pour le moment, si on reste en continu, mettre un seuil sur ω certes on se restreint aux basses fréquences, mais on a toujours un nombre infini de fréquences. Idem pour le calcul d'intégral de Fourier, il nous faudrait un nombre infini de fréquences. Or, tout le **but du traitement du signal** que l'on veut mettre en place c'est de **manipuler le moins de paramètres possible**.

3.6 Le passage du continu au discret

Une remarque que l'on peut faire, c'est qu'en pratique le support du signal $x(u)$ est fini, et on le renormalise pour qu'il soit $u \in [0, 1]$. Et en sous-jacent si j'ai besoin de l'étendre au-delà, on considérera $x(u)$ *périodique* (ou on l'étend à une fonction périodique). Donc, on considère des signaux dans $L^2([0, 1])$. On peut refaire toute l'analyse de Fourier précédente. Les briques de base sont les sinusoïdes, avec cette fois la contrainte d'avoir une période de 1 donc: $\omega = 2\pi n$ pour tout les entiers. On a alors le théorème suivant

Théorème 4

$$\mathcal{B} = \left\{ e_n(u) = e^{i2\pi nu}, \forall n \in \mathbb{Z} \right\} \tag{48}$$

*$\mathcal{B}$ est une **base orthonormale** de $L^2([0, 1])$ qui est le (fameux) résultat sur les séries de Fourier:*

$$\forall x \in L^2([0, 1]), \ x = \sum_{n \in \mathbb{Z}} \langle x, e_n \rangle e_n$$

avec

$$\langle x, e_n \rangle = \int_0^1 x(u) e^{-i2\pi nu} du = \hat{x}(2\pi n)$$

c'est-à-dire le coefficient de Fourier pris à la ***fréquence discrète*** $2\pi n$*, on procède à un échantillonnage dans l'espace de Fourier. La formule de Plancherel/Parseval devient*

$$\|x\|^2 = \sum_n |\langle x, e_n \rangle|^2 = \sum_n |\hat{x}(2\pi n)|^2$$

3.7 Le cas multi-dimensionnel

Avant d'utiliser la base orthonormale pour déployer l'analyse du triangle RAP, on va faire un détour par le multi-dimensionnel pour redéfinir la régularité de Sobolev. En apparence passer de 1D à une dimension quelconque semble "banal" à savoir les résultats se transposent très bien, mais il va y avoir un grain de sable: **les résultats d'approximation vont devenir très mauvais**. C'est la malédiction de la dimensionalité que l'on va quantifier très précisément avec Sobolev.

Donc, mettons que l'on dispose d'une base orthonormale $\{e_n(u), \forall n \in \mathbb{Z}\}$, avec $u \in [0,1]$ ([17]), et on aimerait une base orthonormale avec $u = (u_1, \ldots, u_q) \in [0,1]^q$. La méthode est simple, on effectue un produit séparable:

Théorème 5 *Si* $\{e_n(u), \forall n \in \mathbb{Z}\}$ *est une base orthonormale de* $L^2([0,1])$ *alors*

$$\Big\{ e_n(u) = (e_{n_1}(u_1) \ldots e_{n_q}(u_q)), n = (n_1, \ldots, n_q) \in \mathbb{Z}^q, u = (u_1, \ldots, u_q) \in [0,1]^q \Big\}$$

est une base orthonormale de $L^2([0,1]^q)$*.*

La démonstration procède en 2 étapes: la première consiste à démontrer que les nouveaux vecteurs (e_n) sont orthonormaux; pour la seconde étape il faut montrer que toute fonction de $L^2([0,1]^q)$ se décompose sur la dite base. Pour cette dernière étape, on peut déjà remarquer que c'est le cas pour une fonction qui est séparable, c'est-à-dire une fonction qui s'écrit selon le produit $g_1(u_1) g_2(u_2) \ldots g_q(u_q)$, et ensuite on peut montrer que toute

17. On peut également faire la même chose sur $\mathbb{R}$.

fonction de $L^2([0,1]^q)$ peut être approximée par une famille de fonctions constantes sur un petit cube de $[0,1]^q$ (en 1D on ferait des échelons).

L'important est que l'on peut ainsi étendre la transformée de Fourier en n'importe quelle dimension, car alors la base orthonormale de $L^2([0,1]^q)$ est donnée par

$$\mathcal{B} = \left\{ e_n(u) = e^{i2\pi n.u}, \forall n \in \mathbb{Z}^q, u \in [0,1]^q \right\} \tag{49}$$

avec $n.u = \sum_{k=1}^{q} n_k u_k$.

Maintenant le programme à suivre: on va pouvoir voir comment s'exprime la notion de régularité, savoir à quelle vitesse les coefficients décroissent, et quelle est la parcimonie en jeu. Nous verrons également que dans le cas linéaire on ne peut pas faire mieux pour cette notion de régularité (cf. reliée à PCA). Tout d'abord on va revisiter la régularité en dimension q via la régularité de Sobolev qui est **équivalente** à ce que $|\langle x, e_n \rangle|$ décroisse rapidement, ce qui donne une **équivalence entre régularité et parcimonie**. Ensuite on démontre une **équivalence** entre la **représentation parcimonieuse** et la qualité de l'**approximation** car on va voir l'équivalence entre la vitesse de décroissance de $|\langle x, e_n \rangle|$ et la vitesse de convergence de l'erreur d'approximation $\|x - P_{V_M} x\|^2 = \varepsilon_M$ quand M tend vers l'infini. Donc on aura des équivalences qui relient les 3 notions du triangle RAP.

La conséquence est que quand on abordera les réseaux de neurones, l'objet n'est plus $x(u)$ mais $f(x)$, c'est-à-dire que la variable est x (une image par ex. qui dépend de u avec $q = 2$) dont la dimensionalité explose (ex. d le nombre de pixels, le nombre d'échantillons en temps, etc). Donc, on verra que la décroissance de l'erreur est lente, parce que la parcimonie est mauvaise, laquelle est due à une régularité non-adaptée au problème (même si on contraignait la dérivée d'ordre 100 de f) qui est non-linéaire.

4. Séance du 27 Janv.

Durant cette séance nous allons revisiter le triangle RAP dans un **cadre linéaire** en dimension multiple. On aborde l'étude en entrant par la *régularité* des fonctions en grande dimension. Ce faisant nous allons retrouver la **base de Fourier** étant donné le cadre linéaire, et nous obtiendrons une *équivalence entre régularité et parcimonie* dans cette base. Ensuite nous aborderons *l'équivalence entre parcimonie et approximation*, étant donné que

la représentation en Fourier nous donnera une représentation avec peu de coefficients, on pourra alors faire des *approximation en basse dimension*. Dans un second temps, on se posera la question dans un cadre linéaire, quelle est la base "optimale"? on se servira alors de la **décomposition en composantes principales** (PCA), c'est-à-dire la **base de Karhunen-Loève** et on retrouvera Fourier dès que l'on a une invariance par translation (stationnarité). Ensuite, on se posera la question de savoir si on peut faire mieux quand on passe en *non-linéaire* et en particulier on se penchera sur les performances des *réseaux de neurones à 1 couche cachée*. Ces derniers peuvent être vus à la fois dans le cadre linéaire et nous retrouverons le **théorème d'universalité**, mais également en non-linéaire avec les **espaces de Barron**[18]. Mais ces études malheureusement ne donnent pas les réponses sur les performances des algorithmes en pratique.

Petit rappel: on utilise par la suite essentiellement la notation $x(u)$ mais quand on se placera en grande dimension on notera $f(x)$. Ainsi en basse dimension u est la variable sous-jacente d'une série temporelle ou d'une image par exemple, mais en grande dimension c'est bien x qui devient la variable sous-jacente dans les problèmes du type $y = f(x)$, et si on prend une image x sa dimensionnalité est alors le nombre de pixels.

4.1 La régularité d'une fonction en dimension multiple

La régularité d'une fonction dans $L^2([0,1]^q)$ en dimension $q > 1$, peut être vue d'une manière classique, en regardant les dérivées partielles. Ainsi, on veut contrôler la dérivée de $x(u)$ dans n'importe quelle direction v, cf. $\|v\| = 1$

$$\frac{\partial x(u)}{\partial v} = v.\nabla_u x(u) \tag{50}$$

L'opérateur de dérivée partielle $\partial/\partial v$ est *linéaire et covariant par translation* donc diagonalisable dans une *base de Fourier*. En effet, en prenant un élément de la base (Eq. 48)

$$\frac{\partial e^{i2\pi n.u}}{\partial v} = v.\nabla_u e^{i2\pi n.u} = (i2\pi)\ v.n\ e^{i2\pi n.u} \tag{51}$$

Maintenant, on veut vouloir contrôler n'importe quelle ordre de dérivée de x par

18. Andrew R. Barron prof. à Yale Univ.

rapport à une direction quelconque v, c'est-à-dire qu'elle soit de carré intégrable [19]

$$\left\|\frac{\partial^p x}{\partial v^p}\right\|_2 = \int_{[0,1]^q} \left|\frac{\partial^p x(u)}{\partial v^p}\right|^2 du \tag{52}$$

et voir comment cela s'exprime dans la base de Fourier. Prenons la transformée de Fourier de l'équation 50 ($\omega = 2\pi n$):

$$\widehat{\frac{\partial x}{\partial v}}(2\pi n) = (i2\pi)(v.n)\hat{x}(2\pi n) \tag{53}$$

et en généralisant par itération

$$\widehat{\frac{\partial^p x}{\partial v^p}}(2\pi n) = (i2\pi)^p (v.n)^p \hat{x}(2\pi n) \tag{54}$$

Donc, la condition de régularité via le contrôle des dérivées partielles s'écrit alors

$$\left\|\frac{\partial^p x}{\partial v^p}\right\|_2 \leq C \overset{Parseval}{\Longleftrightarrow} \sum_{n\in\mathbb{Z}^q} |\hat{x}(2\pi n)|^2 |i2\pi|^{2p} |v.n|^{2p} \leq C \tag{55}$$

Le point clé ici pour traduire la régularité du domaine réel (dérivées) dans le domaine de Fourier (coefficient de -) est bien le fait que l'opérateur de dérivation (partielle) est diagonal dans la base de Fourier. Si on lit la condition, cela dit que quand n devient grand il faut que $|\hat{x}(2\pi n)|^2$ décroissent suffisamment rapidement. Mais cette condition doit être vérifiée pour tout vecteur unitaire $v \in \mathbb{R}^q$. Le coefficient $|v.n|^{2p}$ obtient sa valeur max quand v est colinéaire à n, ainsi la condition devient

$$\sum_{n\in\mathbb{Z}^q} |\hat{x}(2\pi n)|^2 |i2\pi|^{2p} |n|^{2p} \leq C \tag{56}$$

donc finalement la régularité indépendamment de la direction impose l'équivalence

$$\boxed{\left\|\frac{\partial^p x}{\partial v^p}\right\|_2 \leq C \Longleftrightarrow \sum_{n\in\mathbb{Z}^q} |\langle x, e_n\rangle|^2 |n|^{2p} \leq C} \tag{57}$$

qui est la **régularité de Sobolev de degré** p.

19. $\|z\|_2$ est la notation pour la norme L^2 de z.

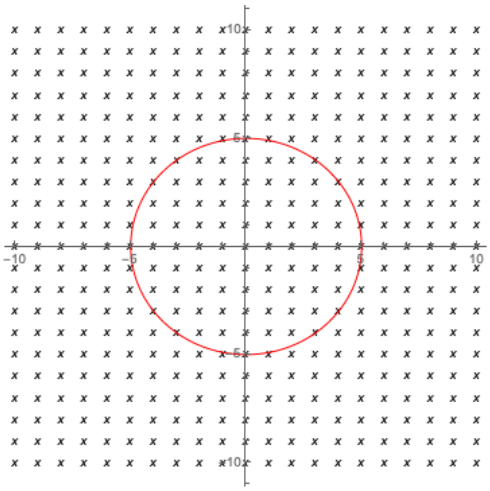

FIGURE 15 – Restriction à basse fréquence de la décomposition de $x(u)$ en 2D.

4.2 Approximation linéaire

4.2.1 Décroissances de l'erreur et des coefficients de Fourier

Quelle est la différence de l'équation 57 par rapport au cas à 1 dimension? factuellement, c'est que $n \in \mathbb{Z}^q$, mais cela va avoir des conséquences car n est dans une grille bien plus grande. Une approximation de $x(u)$ dans la base orthonormale est une somme partielle (tronquée). Si on est en dimension 1

$$x_M(u) = \sum_{n=1}^{M} \langle x, e_n \rangle e_n \tag{58}$$

et l'erreur est donnée par

$$\|x - x_M\|^2 = \sum_{n>M} |\langle x, e_n \rangle|^2 \tag{59}$$

En dimension quelconque, l'index n est un vecteur de $\mathbb{Z}^q$. On peut restreindre sa norme, laquelle donne les composantes à basse fréquence qui sélectionnent les plus grands coefficients (Fig. 15). Ceci peut s'écrire de la façon suivante

$$x_M(u) = \sum_{|n| \leq R_M} \langle x, e_n \rangle e_n \tag{60}$$

et l'erreur devient

$$\|x - x_M\|^2 = \sum_{|n|>R_M} |\langle x, e_n\rangle|^2 \tag{61}$$

Le rayon de la sphère R_M sélectionne M coefficients. Or, en dimension q, le volume d'une boule B_M de rayon R_M est donné par (supposons q paire par simplification)

$$V(B_M) = \pi^{q/2} R_M^q / (q/2)! \tag{62}$$

Chaque point de la grille de $\mathbb{Z}^q$ correspond à un petit hyper-cube et donc à peu de chose prés $V(B_M) \approx M$ et donc

$$R_M = M^{1/q} \sqrt{q} \tilde{\gamma}_q (1 + O(\log q/q)) \equiv M^{1/q} \gamma_q \tag{63}$$

(avec $\tilde{\gamma}_q = 1/\sqrt{2e\pi} \approx 0.25$)

Maintenant, l'idée est d'étudier comment l'erreur Eq. 61 se comporte en grande dimension. En fait le cœur du problème va venir du constat suivant: quand q devient grand, pour que R_M devienne non négligeable pour avoir une bonne approximation, alors M doit croître beaucoup. La condition

$$\varepsilon(M) = \|x - x_M\|^2 = \sum_{|n|>M^{1/q}\gamma_q} |\langle x, e_n\rangle|^2 \tag{64}$$

comment est-elle transcrite en termes de coefficients de Fourier (Eq. 57)? pour relier les deux notions, il y a le théorème suivant qui va relier la décroissance des coefficients de Fourier et la décroissance de l'erreur:

Théorème 6 *Si on suppose une base orthonormale quelconque (c'est-à-dire pas uniquement celle de Fourier), alors on dispose de l'équivalence suivante*

$$\sum_{n\in\mathbb{Z}^q} |\langle x, e_n\rangle|^2 |n|^{2p} \leq C \Leftrightarrow \sum_{M=1}^{+\infty} \varepsilon(M) M^{\frac{2p}{q}-1} \leq C\gamma_q' \tag{65}$$

Cela dit quoi: si $|\langle x, e_n\rangle|$ décroit suffisamment rapidement alors c'est équivalent à dire que $\varepsilon(M) = o(M^{-2p/q})$.

Démonstration 6. Considérons le membre de droite qui contraint la vitesse de décroissance des erreurs d'approximation en remplaçant l'erreur par son expression en fonction des coefficients de Fourier en dehors de la boule:

$$\begin{aligned} A &= \sum_{M=1}^{+\infty} \left(\sum_{|n|>M^{1/q}\gamma_q} |\langle x, e_n\rangle|^2 \right) M^{2p/q-1} \\ &= \sum_{n\in\mathbb{Z}^q} |\langle x, e_n\rangle|^2 \sum_{M=1}^{|n|^q\gamma_q^{-q}} M^{2p/q-1} \sim \sum_{n\in\mathbb{Z}^q} |\langle x, e_n\rangle|^2 \frac{|n|^{2p}\gamma_q^{-2p}}{2p/q} \\ &\sim \gamma_q' \times \sum_{n\in\mathbb{Z}^q} |\langle x, e_n\rangle|^2 |n|^{2p} \end{aligned}$$

On a utilisé

$$\int_1^{a+1} u^s du \leq \sum_{M=1}^{a} M^s \leq \int_0^a u^s du = \frac{a^{s+1}}{s+1}$$

qui donne un équivalent de la somme quand a tend vers l'infini (et $s < 0$). Donc, A en tant que terme de droite de l'équivalence est proportionnelle au terme de gauche, ce qui donne bien la relation entre la décroissance de l'erreur et celle des coefficients de Fourier. ■

C'est un résultat très fort car de la vitesse de décroissance de l'erreur, on déduit la forme de régularité de la fonction. Et si par exemple, on sait que la fonction est 2 fois dérivable mais pas trois fois, alors la décroissance de l'erreur est certes vraie pour $p = 2$ mais pas pour $p = 3$, et encore une fois la décroissance donne exactement l'ordre de régularité. C'est un résultat abondamment utilisé en théorie de l'Approximation pour montrer qu'il y a des équivalences entre des formes de régularité de type Sobolev et des formes d'approximations linéaires que cela soit dans des bases de Fourier ou bien des bases équivalentes.

4.2.2 Malédiction de la dimensionalité

Comme on l'a déjà fait remarquer, la décroissance de l'erreur satisfait le fait que

$$\varepsilon(M) = o(M^{-2p/q}) \tag{66}$$

et ceci est illustré sur la figure 16. Si par exemple, on veut atteindre une erreur ε alors le

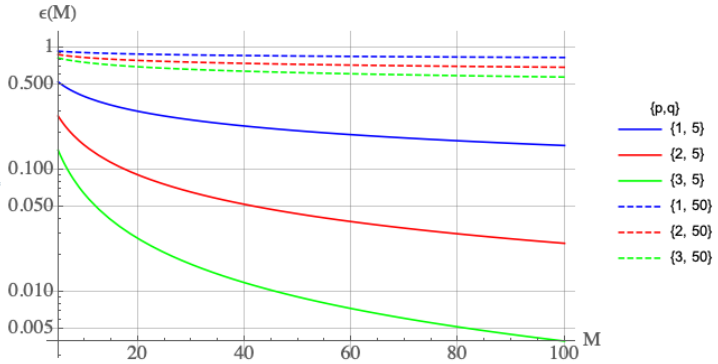

FIGURE 16 – Exemples de décroissance de $\varepsilon(M)$ selon les valeurs de p l'ordre de la dérivée et q la dimension de l'espace.

nombre de coefficients M doit satisfaire en gros

$$M \sim \varepsilon^{-q/2p} \tag{67}$$

ce qui signifie que **ce nombre doit augmenter exponentiellement avec la dimensionalité** q. Et c'est bien là le problème car la fonction a beau être par exemple 100 fois continument dérivable, en grande dimension M va exploser (raisonnement dans lequel x est la variable de $f(x)$ auquel cas $q \sim 10^{4-6}$). **Ainsi, dès que la dimension augmente sensiblement, dans le cadre linéaire on devient rapidement limité**. Mais c'est vrai que si l'on traite des séries temporelles $q = 1$ là pas de soucis, pour des images où $q = 2$ ça ira mais déjà un peu moins bien.

4.2.3 Le filtre basse-fréquence

On obtient une approximation linéaire x_M en prenant une série partielle du signal $x(u)$:

$$x_M(u) = \sum_{|n| \leq R_M} \langle x, e_n \rangle e_n \tag{68}$$

Les coefficients de Fourier de cette approximation sont ceux de x mais uniquement à basse fréquence, c'est-à-dire que l'on a la relation simple

$$\widehat{x_M}(2\pi n) = \hat{x}(2\pi n)\ \mathbb{1}_{|n|\leq R_M} \tag{69}$$

Or l'opérateur $\mathbb{1}_{|n|\leq R_M}$ vaut soit 1 soit 0, il est diagonale en Fourier donc c'est une convolution, et en fait on peut voir comme dans le cas à 1 dimension, x_M **comme la convolution de** x **avec un noyau de Dirichlet ou filtre passe-bas** h_M **qui vaut 1 dans la boule** B_M:

$$x_M = x * h_M \tag{70}$$

Ainsi, on retrouve l'idée à 1 dimension qui indique que si une fonction est régulière, on élimine les hautes fréquences pour obtenir une approximation linéaire. Ce résultat est utilisé dans les réseaux de neurones à 1 couche cachée.

4.3 Découvrir la bonne base: Apprentissage Non Supervisé

Quand on part de données, on n'a en général pas d'idée *a priori* de la régularité de la fonction sous-jacente. Faute de régularité, on va entrer dans le triangle RAP par le biais du coté approximation-parcimonie. Ainsi, on se pose la question de connaître à partir des données uniquement l'approximation linéaire x_M qui permet d'obtenir l'erreur la plus petite possible.

Qui dit approximation linéaire, dit projection de x sur un espace linéaire V_M (cf. Fig. 8) pour obtenir x_M. Donc la question peut être reformulée selon: quel est l'espace linéaire V_M qui permet d'approximer les signaux x afin de minimiser $\varepsilon(M)$? la première remarque est que si je n'ai qu'un seul signal x, on peut choisir n'importe quel hyperplan qui contient x et cela donne trivialement $\varepsilon(M) = 0$. Donc, il faut plusieurs signaux $\{x_i\}_{i\leq N}$ d'un espace Ω pour lesquels on va faire de **l'Apprentissage Non Supervisé** pour obtenir le meilleur espace d'approximation V_M.

Dans un premier temps, le signal "générique" $x(u)$ est de dimension finie $\mathbb{R}^d$, u prend d valeurs, cependant x est un *vecteur aléatoire* dans $\mathbb{R}^d$. Et il nous faut établir la mesure de l'erreur d'approximation, laquelle doit être minimisée pour tout $x \in \Omega$ pour obtenir le

V_M optimal. Ainsi, on aimerait obtenir

$$\underset{x_M \in V_M}{\text{Min}} \; E(\|x - x_M\|^2) \tag{71}$$

Une notion fondamentale quand on utilise **l'erreur quadratique**, c'est que la minimisation ne va dépendre que d'une seule chose: la **covariance**. Faisons un petit rappel pour fixer les notations

$$E[x] = \mu \in \mathbb{R}^d, \qquad E[(x(u) - E[x(u)])(x(u') - E[x(u')])^*] = \mathbf{K}(u, u') \in \mathbb{R}^{d \times d} \tag{72}$$

et $\mathbf{K}$ est une matrice hermitienne définie positive.

Comment est reliée la matrice de covariance $\mathbf{K}$ à l'erreur d'approximation avec un vecteur aléatoire. En fait, $\mathbf{K}$ va caractériser complètement les combinaisons linéaires. Une combinaison linéaire s'écrit comme le produit scalaire de x avec $z \in \mathbb{R}^d$ un vecteur déterministe (c'est-à-dire fixe par rapport à l'aléa de x)

$$\langle x, z \rangle = \sum_{u=1}^{d} x(u) z(u) \tag{73}$$

Si on considère une autre combinaison linéaire $\langle x, z' \rangle$ et l'on veut la corréler avec la précédente, alors (mettons pour simplifier les notations que $E(x) = 0$, cf. on considère le vecteur x dont on soustrait la moyenne μ)

$$\begin{aligned} E[\langle x, z \rangle \langle x, z' \rangle^*] &= E[\sum_u x(u) z^*(u) \times \sum_{u'} x^*(u') z'(u')] = \sum_{u,u'} z^*(u) z'(u') E[x(u) x^*(u')] \\ &= \sum_{u,u'} z^*(u) z'(u') \mathbf{K}(u, u') \\ &= \langle K z', z \rangle = z^T . \mathbf{K} z' \end{aligned} \tag{74}$$

Donc toute la caractérisation de la corrélation est donnée par la matrice $\mathbf{K}$. Notons au passage que si $z = z'$:

$$E[|\langle x, z \rangle|^2] = \langle K z, z \rangle \tag{75}$$

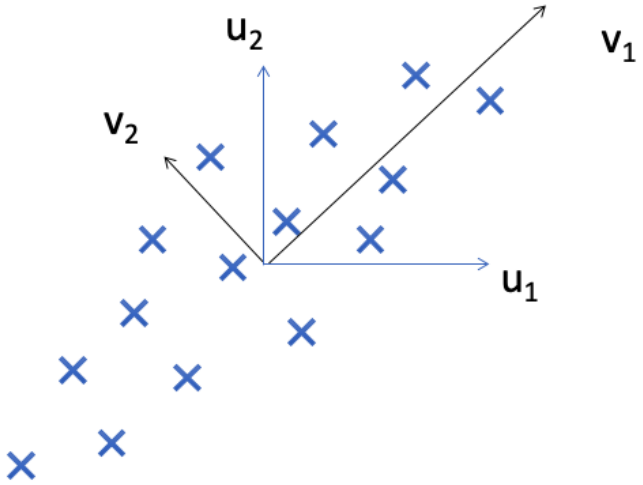

FIGURE 17 – Illustration des axes principaux de la matrice de covariance.

Donc, considérons à présent l'erreur (en espérance) de l'approximation

$$E(\|x - x_M\|^2) = E(\sum_{|n|>R_M} |\langle x, e_n\rangle|^2) = \sum_{|n|>R_M} \langle Ke_n, e_n\rangle \tag{76}$$

On constate qu'une fois la base fixée, **l'erreur ne dépend que de la matrice de covariance**. La meilleure base va dépendre des propriétés de cette matrice de covariance. Notons qu'une interprétation géométrique de $\mathbf{K}$ est donnée par **les axes principaux** d'une population de vecteurs x (Fig. 17). L'axe de plus grande variabilité est celui du vecteur propre ayant la plus grande des valeurs propres.

Voici un théorème très utilisé en analyse de données et en analyse fonctionnelle car il vous dit pourquoi la base de Fourier est optimale très souvent:

Théorème 7 *Pour tout M, $E(\|x - x_M\|^2 = \sum_{|n|>R_M} E(|\langle x, e_n\rangle|^2)$ est minimum si les éléments de la base $\{e_n\}_{n\leq d}$ diagonalisent la matrice de covariance $\mathbf{K}$ associée à x, avec les valeurs propres*

$$\lambda_n = \langle Ke_n, e_n\rangle \geq 0$$

qui sont ordonnées par ordre décroissant. Une autre façon d'exprimer la même chose

est que $E(\|x - P_{V_M} x\|^2)$ est minimum si V_M est généré par les M premiers vecteurs propres de la matrice de covariance $\mathbf{K}$.

Démonstration 7. Pour commencer la démonstration appliquons le Th. de Pythagore sachant que x_M est la projection orthogonale de x sur l'hyper-plan V_M:

$$E(\|x - x_M\|^2) = E(\|x\|^2) - E(\|x_M\|^2)$$

En contexte, on veut minimiser l'erreur d'approximation vis-à-vis de tous les x_M possibles, ce qui revient à maximiser $E(\|x_M\|^2)$. Or,

$$E(\|x_M\|^2) = \sum_{|n| \leq R_M} E(|\langle x, e_n \rangle|^2)$$

En fait, la condition $|n| \leq R_M$ définit les M vecteurs de la base $\{e_n\}_{n \leq d}$, donc on peut écrire la somme selon $\sum_{n=1}^{M}$. Prenons pour base les vecteurs qui diagonalisent la matrice $\mathbf{K}$ avec (λ_n) le spectre de valeurs propres, que l'on note $\{\bar{e}_n\}_{n \leq d}$. Il s'agit de **la base de Karhunen-Loève** [20] définie telle que

$$\langle K\bar{e}_n, \bar{e}_n \rangle = \lambda_n \geq 0 \quad \text{(décroissant)} \qquad\qquad \langle K\bar{e}_n, \bar{e}_{n'} \rangle = 0 \quad (n \neq n')$$

Donc, l'idée est de montrer que $E(\|x_M\|^2)$ est plus grand dans la base de Karhunen-Loève, donc il faut exprimer e_n dans cette base particulière:

$$e_n = \sum_{k=1}^{d} \langle e_n, \bar{e}_k \rangle \bar{e}_k$$

20. de Kari Karhunen (1915-92) mathématicien finlandais, et Michel Loève (1907-79) mathématicien franco-américain.

Ainsi, il vient

$$\begin{aligned} E(|\langle x, e_n\rangle|^2) &= E(|\sum_{k=1}^{d}\langle e_n, \bar{e}_k\rangle\langle x, \bar{e}_k\rangle|^2) \\ &= \sum_{k,k'}\langle e_n, \bar{e}_k\rangle\langle e_n, \bar{e}_{k'}\rangle^* E(\langle x, \bar{e}_k\rangle\langle x, \bar{e}_{k'}\rangle^*) = \sum_{k,k'}\langle e_n, \bar{e}_k\rangle\langle e_n, \bar{e}_{k'}\rangle^*\langle K\bar{e}_k, \bar{e}_{k'}\rangle \\ &= \sum_{k}\lambda_k|\langle e_n, \bar{e}_k\rangle|^2 \end{aligned}$$

Donc, l'expression de $E(\|x_M\|^2)$ donne

$$E(\|x_M\|^2) = \sum_{n=1}^{M}\sum_{k=1}^{d}\lambda_k|\langle e_n, \bar{e}_k\rangle|^2 = \sum_{k=1}^{d}\lambda_k\left(\sum_{n=1}^{M}|\langle e_n, \bar{e}_k\rangle|^2\right)$$

et l'on veut maximiser cette expression par le choix de la base $\{e_n\}_{n\leq M}$ de V_M (cf. les degrés de liberté du problème). Or,

$$0 \leq c_k = \sum_{n=1}^{M}|\langle e_n, \bar{e}_k\rangle|^2 \leq \sum_{n=1}^{d}|\langle e_n, \bar{e}_k\rangle|^2 = \|\bar{e}_k\|^2 = 1$$

et d'autre part

$$\sum_{k=1}^{d} c_k = \sum_{n=1}^{M}\sum_{k=1}^{d}|\langle e_n, \bar{e}_k\rangle|^2 = \sum_{n=1}^{M}\|e_n\|^2 = M$$

Donc, les $\{c_k\}_{k\leq d}$ sont des nombres compris dans $[0, 1]$ et dont la somme est égale à M. Maintenant les $\{\lambda_k\}_{k\leq d}$ sont positives ou nulles et ordonnées par ordre décroissant. Donc pour maximiser $E(\|x_M\|^2)$ il suffit d'affecter les M premiers c_k à la valeur 1 et les autres à 0:

$$c_k = \begin{cases} 1 & k \leq M \\ 0 & k = M+1, \ldots, d \end{cases}$$

Ainsi,

$$\forall k \leq M,\ \sum_{n=1}^{M}|\langle e_n, \bar{e}_k\rangle|^2 = 1$$

donc les vecteurs orthonormés $\{e_n\}_{n\leq M}$ sont dans l'espace engendré par les M premiers vecteurs de la base de Karhunen-Loève, et finalement il en est de même de V_M. ■

Ce théorème est très important en pratique bien entendu, et aussi parce que l'on se rend compte que le problème d'approximation devient un problème de diagonalisation d'opérateur. Qui plus est en pratique également, on trouve très souvent la notion de signaux stationnaires et on va voir alors que l'on peut deviner la base orthonormale.

4.4 Signaux stationnaires

Soit des Processus Aléatoires Stationnaires au second ordre [21], et $p(x)$ la densité de probabilité de $x \in \mathbb{R}^d$, que devient la densité de probabilité si le signal est translaté:

$$x_\tau(u) = x(u-\tau) \Rightarrow p(x_\tau)? \tag{77}$$

Dans ce contexte, on dit que le **processus est stationnaire** [22] si

$$\forall k, \alpha,\ p(x(u_1), x(u_2), \ldots, x(u_k)) = p(x(u_1-\alpha), x(u_2-\alpha), \ldots, x(u_k-\alpha)) \tag{78}$$

Pour conséquence la moyenne du signal est une constante. C'est le cas par exemple en imagerie ou prise de son quand il n'y a pas de point de référence. Également,

$$E[f(x(u_1), x(u_1+\tau))] = E[f(x(u_1-\alpha), x(u_1-\alpha+\tau))] = E[f(x(0), x(\tau))]$$

qui est une fonction de τ uniquement, cela a pour conséquence si on prend $f(x,y) = xy - \mu$ que la matrice de covariance a la propriété suivante

$$\forall u, u',\ K(u,u') = K(u-u') \tag{79}$$

21. On affaiblit un peu la notion de stationnarité pour faciliter l'étude.
22. hypothèse forte

Ainsi quand on applique l'opérateur K d'un processus stationnaire sur une fonction g par exemple, on a

$$K.g(u) = \sum_{u'} g(u')K(u,u') = \sum_{u'} g(u')K(u-u') = (K * g)(u) \tag{80}$$

c'est une **opération de convolution** par K.

Or, un opérateur de convolution est **diagonalisé dans une base de Fourier**. Donc, dès que l'on a affaire à un **processus stationnaire, la base de Karhunen-Loève est la base de Fourier**, à savoir que

$$\left\{\bar{e}_k(u) = e^{i2\pi ku/d}\right\}_{1\leq u\leq d} \tag{81}$$

si on est dans un espace discret à d valeurs pour lequel on utilise une périodisation modulo d de u pour définir la translation (nb. les fréquences sont $2\pi k/d$).

Si on revient sur la notion de régularité uniforme, si par exemple la dérivée du signal est bornée, il en est de même pour le signal translaté, et cela définit des classes de fonctions invariantes par translation. **Et comme la base optimale de Karhunen-Loève est identique à celle de Fourier dans ce cas, alors on ne pourra jamais faire mieux dans le cadre d'une approximation linéaire.**

Ainsi, on a l'optimum de ce que l'on peut faire en linéaire, mais pour autant est-ce satisfaisant? prenons une fonction $g(u)$ avec $u \in [1,d]$ et un signal quelconque x est défini à partir de g selon (v.a : variable aléatoire) (Fig. 18)

$$x_\tau(u) = g((u-\tau) \bmod [d]), \quad \tau \in \{1,\ldots,d\} \text{ v.a unifome} \tag{82}$$

Le signal x est un processus aléatoire et étant donné que τ a une loi uniforme, on a autant de chance de "voir" x ou x_τ, donc le processus est stationnaire. Que vaut la matrice $\mathbf{K}$? tout d'abord remarquons que l'espérance de x n'est autre que la moyenne de g (invariant par translation):

$$E[x_\tau(u)] = \sum_{\tau=1}^{d} p(\tau)g((u-\tau)\bmod[d]) = \frac{1}{d}\sum_{\tau=1}^{d} g((u-\tau)\bmod[d]) = \frac{1}{d}\sum_{u'=1}^{d} g(u') \tag{83}$$

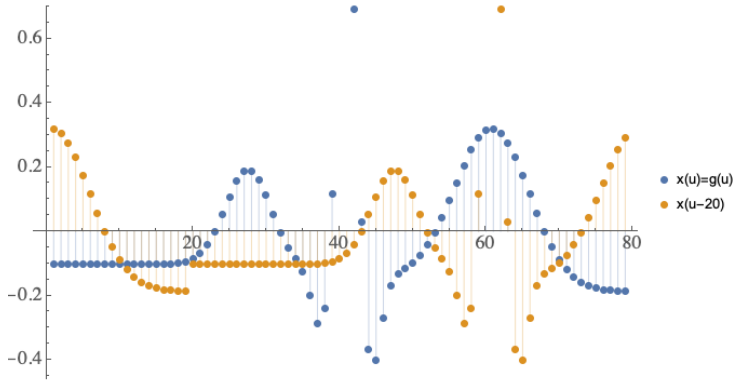

FIGURE 18 – Exemple de $g(u)$ et une version translatée périodisée (nb. la moyenne de g est nulle).

Mettons que cette moyenne soit nulle pour simplifier les calculs, il vient alors:

$$\begin{aligned} E[x_\tau(u)x_\tau(u')] &= \sum_{\tau=1}^{d} p(\tau)\ g((u-\tau)\text{mod}[d])\ g((u'-\tau)\text{mod}[d]) \\ &= \frac{1}{d}\sum_{\tau=1}^{d} g((u-\tau)\text{mod}[d])\ g((u'-\tau)\text{mod}[d]) \\ &= \frac{1}{d}\sum_{u''=1}^{d} g(u'')g(u''-(u-u')))\text{mod}[d]) = K(u-u') \end{aligned} \tag{84}$$

Donc, on a bien un signal stationnaire (Fig. 19) avec une matrice de covariance qui ne dépend que de la variable $u-u'$. Ainsi, via le théorème 7, on tombe sur la base de Fourier pour effectuer une approximation linéaire basse fréquence. Or, le signal $g(u)$ peut avoir des discontinuités et donc il n'est pas toujours légitime de procéder à une approximation basse fréquence de ce dernier. On a envie de passer par une approximation non-linéaire qui va adapter l'échantillonnage à la régularité de $g(u)$. Et du point de vue de la base optimale, on ne prendra pas forcément les M premiers vecteurs, mais il va falloir choisir astucieusement les vecteurs. Ce que l'on verra alors c'est que le théorème 7 n'est plus juste. Nous verrons qu'il faut redéfinir d'autres classes de régularité, certes plus complexes mais cela vaudra le coup pour certains types de problèmes. Nous verrons aussi que si l'on a un regard purement "linéaire" pour analyser des réseaux de neurones à 1 couche, tout semble simple et on ne fait pas nettement mieux que la base de Fourier (ou PCA), mais dès que

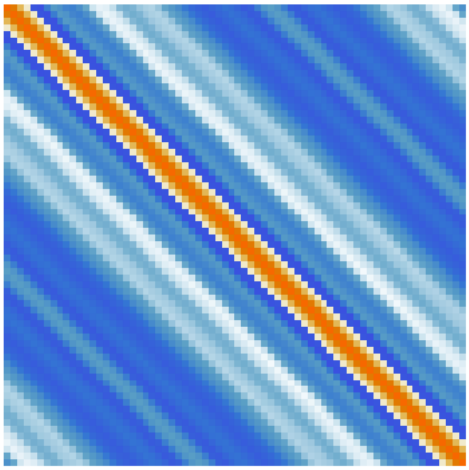

FIGURE 19 – La matrice $K(u, u')$ du signal (Fig. 18) qui a bien la structure en bandes qui signifie que la matrice ne dépend que de la variable $u - u'$.

l'on a un regard "non-linéaire" les théorèmes deviennent plus compliqués à interpréter.

5. Séance du 3 Févr.

Durant cette séance, nous allons parcourir le triangle RAP d'un point de vue *non-linéaire*. Rappelons qu'en *linéaire*, obtenir une approximation en basse dimension au cœur de l'analyse de données est équivalent à se placer dans une représentation parcimonieuse qui concentre l'information sur peu de coefficients. Cette notion de parcimonie est équivalente à la forme de régularité de la fonction sous-jacente. Dans le cas linéaire également, pour la classe de fonctions invariantes par translation, on a vu que la base optimale est celle de Fourier. D'une manière générale, on a aussi vu que la meilleure base est celle qui diagonalise la matrice de covariance, à savoir la base de Karhunen-Loève, laquelle est identique à celle de Fourier pour les processus stationnaires.

La question qui se pose à présent est de savoir si on peut faire mieux en prenant un point de vue *non-linéaire*? A la fin de la section précédente, on a pointé du doigt que l'on peut entrevoir une amélioration dans le cas où la fonction sous-jacente présente des discontinuités, c'est-à-dire qu'elle n'est que *régulière par morceaux* comme par exemple

la fonction de la figure 10. Une approximation linéaire consisterait à considérer la fonction *uniformément régulière*, ce qui suggère d'effectuer un échantillonnage régulier et ce qui revient à ne considérer que des basses fréquences. Ce faisant, ce point de vue limite d'autant la qualité de l'approximation au voisinage d'éventuelles discontinuités. Dans ce cas, il va falloir s'adapter au cas par cas. Mais comment le faire génériquement ou automatiquement? Si on considère un *échantillonnage adaptatif*, on prend le parti que la régularité de la fonction n'est pas uniforme partout, c'est-à-dire qu'il peut y avoir des singularités. Notons que ces discontinuités/singularités recèlent de l'information importante, par exemple elles fournissent des indications des contours d'objets en 2D, ou bien des attaques de morceaux en musique. Donc, la structure des signaux se trouve dans les *composantes hautes fréquences*. La question devient alors: peut-on capturer les discontinuités en ayant une représentation parcimonieuse qui permet d'obtenir une approximation de basse dimension de meilleure qualité que celle obtenue en linéaire?

5.1 Représentation parcimonieuse non-linéaire

On se place dans le cadre de la recherche d'une base orthonormale, c'est-à-dire que le signal x se décompose selon

$$x = \sum_{n=1}^{\infty} \langle x, e_n \rangle e_n \tag{85}$$

Dans le cas **linéaire**, l'approximation se fait en considérant la somme partielle tronquée aux M **premiers coefficients**. En Fourier cela correspond aux basses fréquences. Dans le cas **non-linéaire**, on va procéder **au choix des** M **coefficients**, ainsi on peut écrire:

$$x_M = \sum_{n \in I_M} \langle x, e_n \rangle e_n \tag{86}$$

avec I_M un ensemble des index n qui vont dépendre de x (tout en gardant $|I_M| = M$). Comme on a une base orthonormale, l'erreur est simple à formuler à savoir

$$\varepsilon_M = \|x - x_M\|^2 = \| \sum_{n \notin I_M} \langle x, e_n \rangle e_n \|^2 = \sum_{n \notin I_M} |\langle x, e_n \rangle|^2 \tag{87}$$

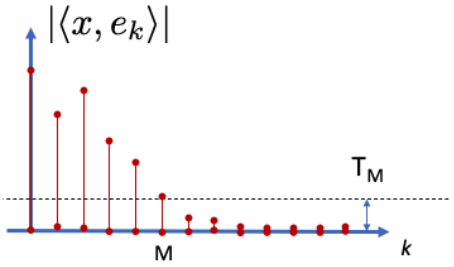

FIGURE 20 – Sélection des M premiers produits scalaires une fois ordonnés par ordre décroissant, ce qui définit le seuil T_M.

Le choix de I_M se fait dans l'optique de minimiser cette erreur d'approximation, et l'on en vient naturellement à la définition

$$I_M = \{n \ / \ |\langle x, e_n \rangle| \geq T_M\}, \qquad \text{avec } \mathrm{T_M} \text{ tq. } |I_M| = M \tag{88}$$

ce qui permet d'écrire

$$\varepsilon_M = \sum_{|\langle x, e_n \rangle| < T_M} |\langle x, e_n \rangle|^2 \tag{89}$$

La question est de savoir si ε_M calculée de la sorte est bien plus petite que celle que l'on aurait obtenu par une approche linéaire? Nous allons voir que la réponse est "oui" si la base donne une représentation parcimonieuse. Donc, Approximation et Parcimonie sont bien de nouveau deux notions intimement reliées.

La parcimonie se manifeste par la décroissance des produits scalaires mais une fois ordonnés par ordre décroissant:

$$|\langle x, e_k \rangle| \geq |\langle x, e_{k+1} \rangle| \tag{90}$$

Comme illustré sur la figure 20 l'erreur correspond à la somme des carrés des termes pour lesquels $k > M$ ou dont l'intensité est en deçà du seuil T_M. Ainsi,

$$\varepsilon_M = \sum_{k=M+1}^{\infty} |\langle x, e_k \rangle|^2 \tag{91}$$

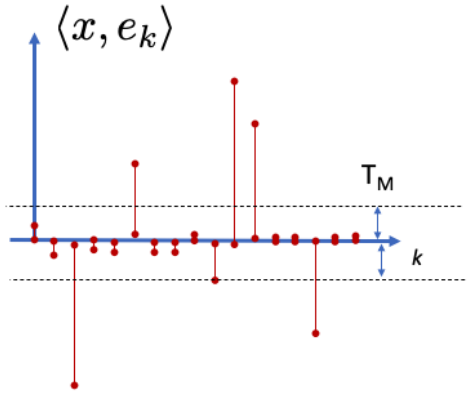

FIGURE 21 – Sélection par le biais du seuil T_M des produits scalaires non ordonnés.

Ceci dit, la parcimonie au sens non-linéaire peut tout aussi bien se représenter sans faire appel à un tri des coefficients par ordre décroissant, on peut le voir comme sur la figure 21 où **les coefficients émergeant au dessus du seuil T_M sont peu nombreux** et sont **localisés n'importe où** selon le signal x considéré. De plus, l'erreur résiduelle décroit vers 0 si le seuil est abaissé. A quelle vitesse décroit-elle?

5.1.1 Vitesse de décroissance de l'erreur non-linéaire

Comme dans le cas linéaire (Th. 6), on peut contraindre la vitesse de décroissance de l'erreur

Théorème 8 *Si on ordonne les produits scalaires par ordre décroissant, et p est un ordre de dérivée, si*[a]

$$\sum_{k=1}^{\infty} |\langle x, e_k \rangle|^2 k^{2p} < \infty \tag{92}$$

ce qui revient à dire $|\langle x, e_k \rangle|^2 = o(k^{-2p-1})$ et ce qui caractérise la vitesse de décroissance des produits scalaires, alors

$$\sum_{M=1}^{\infty} \varepsilon_M M^{2p-1} < \infty \tag{93}$$

donc cela revient à dire que $\varepsilon_M = o(M^{-2p})$. En fait il y a équivalence entre les deux vitesses de décroissance.

a. Notons au passage que la technique de caractérisation de la vitesse de décroissance par la contrainte de la convergence d'une série est assez générale.

Dans le cas linéaire, on a l'index $k \in \mathbb{Z}^q$ ce qui donne $\varepsilon_M = o(M^{-2p/q})$, dans le cas du théorème ci-dessus $k \in \mathbb{Z}$, sinon la démonstration est en tout point identique. Donc, on a obtenu une équivalence entre les notions d'Approximation de basse dimension et Représentation parcimonieuse du triangle RAP. Cependant, même s'il y a des similitudes entre le cas *linéaire* et le cas *non-linéaire*, pour ce dernier **le résultat est obtenu en ordonnant les produits scalaires selon leur importance, ce qui dépend incidemment du signal x.** Cependant, on aimerait pouvoir obtenir une contrainte indépendante de l'ordre des coefficients car cela collerait mieux au cas pratique.

5.1.2 Parcimonie et norme ℓ^α

Il y a un théorème important qui est au cœur des mathématiques de l'Approximation et à la base de beaucoup d'algorithmes, qui stipule que la vitesse de décroissance de coefficients ordonnées est caractérisée/capturée par les normes ℓ^α. C'est-à-dire qu'au lieu de manipuler la norme L^2 dans une base orthonormée par exemple via

$$\sum |\langle x, e_n\rangle|^2 = \|x\|^2$$

on va s'intéresser à des normes définies par

$$\left(\sum |\langle x, e_n\rangle|^\alpha\right)^{1/\alpha}$$

avec en particulier $\alpha < 2$.

Théorème 9 *Si pour $\alpha < 2$*

$$C_\alpha = \left(\sum_{q=1}^{\infty} |\langle x, e_q\rangle|^\alpha\right)^{1/\alpha} < \infty \tag{94}$$

alors en ordonnant les produits scalaires, celui de rang k satisfait une contrainte de décroissance, et il en découle une loi de décroissance de l'erreur:

$$|\langle x, e_k\rangle| \leq C_\alpha \; k^{-1/\alpha} \qquad \text{et} \qquad \varepsilon_M \leq \frac{C_\alpha^2}{\frac{2}{\alpha}-1} M^{-\left(\frac{2}{\alpha}-1\right)} \tag{95}$$

Même si on peut obtenir mieux en remplaçant les bornes par des petits o, ***il est important de constater que c'est la norme*** ℓ^α ***de*** x***, c'est-à-dire*** C_α***, qui contrôle à la fois la décroissance des produits scalaires ordonnés et celle de l'erreur*** ε_M***.***

Démonstration 9. On part de la définition de C_α avec une base orthonormée $\{e_n\}$ et un signal x:

$$C_\alpha^\alpha = \sum_{n=1}^{\infty} |\langle x, e_n\rangle|^\alpha$$

puis on effectue un tri des produits scalaires selon un ordre décroissant ce qui opère un réarrangement des indices que l'on rend visible par l'usage de nouveaux indices n_q. Ainsi,

$$C_\alpha^\alpha = \sum_{q=1}^{\infty} |\langle x, e_{n_q}\rangle|^\alpha$$

Or, on peut tronquer la série aux k premiers termes pour minorer la série, il vient

$$C_\alpha^\alpha \geq \sum_{q=1}^{k} |\langle x, e_{n_q}\rangle|^\alpha \geq k|\langle x, e_{n_k}\rangle|^\alpha$$

Donc autrement dit le coefficient de rang k satisfait la contrainte

$$|\langle x, e_{n_k}\rangle| \leq C_\alpha k^{-1/\alpha}$$

ce que l'on voulait démontrer [23]. ■

Ce théorème nous indique que plus α **est petit, plus rapide est la décroissance des**

23. nb. pour obtenir la version avec le petit "o", S. Mallat nous invite à prendre pour majorant non pas la somme de $q = 1$ à k mais de $k/2$ à k qui va tendre vers 0.

produits scalaires ainsi que celle de l'erreur. Donc, on voudra considérer des normes ℓ^α avec des α aussi petit que possible. dès que l'on obtient des contraintes sur les normes ℓ^α d'un signal alors on sent la parcimonie poindre son nez.

Maintenant, tant que $\alpha \geq 1$ alors C_α est une *fonction convexe* des produits scalaires, et ce n'est plus le cas dès que $\alpha < 1$. Pourquoi est-ce important? la raison en est la suivante: imaginons que l'on ne connaisse pas la base $\{e_n\}$, il faut la découvrir or dans ce cas on va vouloir minimiser la norme ℓ^α qui devient une sorte de fonction de coût. Et dans ce contexte, avoir une fonction convexe permet d'utiliser des algorithmes d'optimisation. Donc en pratique, on travaille avec α dans l'intervalle $[1, 2)$ et le plus petit possible, d'où **l'usage de la norme** ℓ^1 dans les problèmes d'optimisation tout en garantissant une sparsité [24] de la représentation.

5.2 Application aux réseaux de neurones à 1 couche cachée

Dans cette partie, nous allons mettre en pratique tout ce que nous avons vu jusqu'à présent pour étudier le problème de classification/régression de type $f(x) = y$ avec un réseau de neurones à 1 couche cachée. Nous allons constater deux écueils qui nous feront changer radicalement de point de vue.

Pour ce type de problème rappelons que l'on veut approximer la fonction f qui à x fait correspondre y soit de type entier pour la classification, soit de type réel pour la régression. Ici x est la variable et son domaine est soit $\mathbb{R}^d$ ou bien $[0, 1]^d$ pour des signaux bornés, mais surtout avec $d \approx 10^{4-6}$. Jusqu'à présent l'objet d'étude était $x(u)$ avec u en très basse dimension q ($u \in [0, 1]^q$) et on sait que lorsque q commence à augmenter significativement, on est confronté à la malédiction de la dimensionalité [25].

Chaque neurone m d'un réseau à M neurones cachés (Fig. 5) calcule tout d'abord un produit scalaire avec un vecteur w_m selon

$$x.w_m = \sum_{u=1}^{d} x(u)w_m(u) \tag{96}$$

24. Sparsity en anglais est la traduction de parcimonie.
25. ex. voir Cours de 2018, 2019

puis on applique une non-linéarité ρ (ReLU, sigmoïde, tangente hyperbolique, cosinus, etc.) et un biais b_m:

$$\rho(x.w_m + b_m) \in \mathbb{R} \tag{97}$$

et enfin on recombine linéairement les M non-linéarités pour obtenir $\tilde{f}$

$$\tilde{f}(x) = \sum_{m=1}^{M} \alpha_m \rho(x.w_m + b_m) \tag{98}$$

En l'occurrence, on prend pour parti (c'est une information *a priori*, thème du cours de 2020) de décomposer $\tilde{f}$ sur une famille de fonctions $\{\rho(x.w_m + b_m)\}_{m \leq M}$. On effectue **une projection de f dans l'espace engendré par cette famille de fonctions**. Et la question qui vient immédiatement: quelle est la taille de l'erreur commise?

Pour répondre à cette question, on va considérer l'erreur quadratique, par exemple si les signaux sont dans $[0,1]^d$ (ex. les valeurs des pixels d'une image):

$$\|f - \tilde{f}\|^2 = \int_{x \in [0,1]^d} |f(x) - \tilde{f}(x)|^2 \, dx \tag{99}$$

L'idée est de montrer qu'en utilisant ce que l'on a vu jusqu'à présent, nous allons pouvoir apporter les réponses "usuelles" sur le sujet.

5.2.1 Approximation universelle (point de vue linéaire)

Un premier résultat [26] est celui du théorème d'approximation universelle qui stipule que moyennant l'utilisation de non linéarités non-polynomiales, l'erreur d'approximation tend vers 0 quand M tend vers l'infini. Ce théorème a été démontré et raffiné essentiellement dans les années 1988-92 par la communauté des mathématiciens spécialistes de l'approximation [27] par des techniques bien connues car c'est un résultat essentiellement d'approximation linéaire. En effet, d'où vient le fait que l'erreur tende vers 0?

Pour nous éclairer, prenons le point de vue de l'*approximation linéaire*. On se pose la question de savoir s'il existe une famille de fonctions génériques qui va être suffisante pour approximer n'importe quelle fonction f quand le nombre de composantes (M) tend

26. Voir le cours de 2019 pour une version de la démonstration.
27. On peut citer: George Cybenko, Kurt Hornik , et Allan Pinkus, etc.

vers l'infini. Mais nous avons vu: *qui dit famille de fonctions en linéaire, dit Fourier*. Donc, la question est de savoir comment construire une base de Fourier à partir de la famille $\{\rho(x.w_m + b_m)\}$? Dans un premier temps, prenons comme non linéarité la fonction $\rho(a) = e^{ia}$. Il vient alors

$$\tilde{f}(x) = \sum_{m=1}^{M} \alpha_m e^{i(x.w_m+b_m)} = \sum_{m=1}^{M} \alpha_m e^{i\,b_m}\ e^{i\,x.w_m} \tag{100}$$

ce qui ressemble à une série de Fourier. Maintenant, on doit apprendre les poids w_m (point de vue du réseau), mais en Fourier cela correspond aux fréquences de la décomposition. Or, pour peu que la fonction sous-jacente f ait un peu de *régularité*, ce que suppose tous les théorèmes qui traitent de ce sujet dans ce point de vue linéaire, alors on se concentre sur une *approximation basse-fréquence* (Sec. 4.2). Si on se place dans le cas où $x \in [0,1]^d$ alors $w_n = 2\pi n$ avec $n \in \mathbb{Z}^d$ (Sec. 3.7). Il vient

$$\tilde{f}(x) = \sum_{m=1}^{M} \alpha_m e^{i\,b_m}\ e^{i2\pi\,x.m} \tag{101}$$

Et comme on a une base orthonormale de Fourier, il vient

$$\alpha_m e^{i\,b_m} = \langle f(x), e^{i2\pi\,x.m}\rangle = \int_{x\in[0,1]^d} f(x) e^{-i2\pi\,x.m}\ dx = \hat{f}(2\pi m) \tag{102}$$

De plus, on a la garantie d'avoir l'approximation (linéaire) minimum puisque la base de Fourier est optimale dans ce cas. Nous savons que f est réelle [28] donc $\hat{f}^*(\omega) = \hat{f}(-\omega)$ et

$$\tilde{f}(x) = \sum_{m=1}^{M} \alpha_m \cos(w_m.x + b_m) \tag{103}$$

Maintenant, on sait que si $f \in L^2([0,1]^d)$, donc dès que f est d'énergie finie ce qui est une hypothèse raisonnable, alors un résultat sur les séries de Fourier nous dit que (sauf accident)

$$\lim_{M\to\infty} |f(x) - \sum_{m=1}^{M} \alpha_m \cos(w_m.x + b_m)|^2 = 0 \tag{104}$$

Donc, le théorème d'approximation universelle est à-peu-prés basé sur les notions que

28. Disons que l'on se place dans ce cas de figure qui n'est pas restrictif pour les cas pratiques, traitement du son, d'images etc.

nous avons vu précédemment.

Le "à-peu-prés" signifie par exemple que le théorème marche pour d'autres types de non-linéarités que l'exponentielle complexe fussent-elles non-polynomiales. Pourquoi? On aimerait *faire un changement de base en 1D* entre les familles $\{\rho(w.x+b)\}_{w,b}$ et $\{\cos(w.x+b)\}_{w,b}$ avec ρ un ReLU, une sigmoïde, etc. Par exemple, entre le ReLU et le cosinus, il suffit de montrer que le cosinus peut s'approximer avec des fonctions linéaires par morceaux sur des pas suffisamment fins (ex. voir Cours de 2019 Sec. 5.3.2). Autre chose, la convergence citée ci-dessus est au sens de la norme L^2, alors que les théorèmes arrivent à des convergences uniformes (c'est-à-dire en norme "sup" sur un espace dans lequel x évolue), mais ce sont des raffinements sur **une dominante lourde de l'approximation linéaire et de la base de Fourier**.

Cependant, avoir utilisé la base de Fourier nous dit que si on veut une convergence rapide, on doit imposer à la classe de fonctions f des contraintes de régularité. Par exemple, si f appartient à un espace de Sobolev de degré p (Sec. 3.5), c'est-à-dire que toutes les dérivées d'ordre p sont d'énergie finie, alors (Th. 6)

$$\|\tilde{f}_M - f\| = o(M^{\frac{-2p}{d}}) \tag{105}$$

Mais examinons de plus près ce résultat: il nous dit que même si les fonctions sont très régulières (p grand) ce qui tend à accélérer la convergence, le fait que d peut être très grand ($d \approx 10^{4-6}$) réduit considérablement la vitesse de convergence sauf à devoir augmenter exponentiellement le nombre de neurones cachés (M). **On se retrouve devant la malédiction de la dimensionalité, et surtout on n'arrive pas à comprendre dans ce cadre les résultats tout à fait remarquables des réseaux de neurones.** Mais ce n'est pas la fin de l'histoire...

5.2.2 Le point de vue non-linéaire

Ce qui a été remarqué dans les années 1992 (ex. A. Barron) c'est que le problème de trouver une approximation $\tilde{f}$ n'est pas à considérer pour une classe de fonctions avec une base générique, mais il faut particulariser, pour ne pas dire **adapter le raisonnement**

à la fonction spécifique f qui nous intéresse pour le problème posé [29]. Il est vain donc de chercher le cas général, alors que le cas particulier est le seul qui nous intéresse au moins dans un premier temps, et pour peu que l'on opte pour le point de vue "non-linéaire".

Il nous faut **adapter la base de projection**, donc dans le contexte, il nous faut adapter les (w_n, b_n) vis-à-vis de l'objet à approximer, c'est-à-dire à f. Dans le cours précédent, nous avions trouvé une manière d'adapter la décomposition à x en sélectionnant les produits scalaires $\langle x, e_n \rangle$ au-delà d'un seuil $T_M(x)$ choisi pour n'avoir que M composantes. Que faire maintenant dans notre problème $f(x) = y$?

On va se servir de la relation entre parcimonie (approximation basse dimension) et les normes ℓ^α (Sec. 5.1.2). Car que faut-il envisager pour **battre la malédiction de la dimensionalité**? Il nous faut trouver une base dans laquelle les produits scalaires $|\langle f, e_k \rangle|$ décroissent d'une manière ordonnée selon $1/k^\alpha$ (voire $\alpha = 1$ pour la garantir la convexité), car alors d'après le théorème 9 on pourra déduite que l'erreur sera contrainte par

$$\|f - \tilde{f}_M\|^2 \leq \frac{C_\alpha^2(f)}{\frac{2}{\alpha} - 1} M^{-\left(\frac{2}{\alpha} - 1\right)} \tag{106}$$

c'est-à-dire que **la vitesse de convergence ne dépendra plus de la dimension** d de la variable x. Donc, ce qu'il faut obtenir pour que ce résultat se mette en place, est que **la norme ℓ^α de f soit bornée**.

Le premier article qui a mis ce schéma en pratique date de 1993 (A. Barron [30]) dont le résultat peut se mettre sous la forme suivante. Première remarque, la base $\{e_n\}$ prise par A. Barron est celle de Fourier avec $e_n = e^{i2\pi n}$, deuxièmement au lieu de considérer des fonctions avec une régularité de Sobolev pour laquelle on serait contraint à la malédiction de la dimensionalité, envisageons **un nouveau type de régularité (*les espaces de Barron*)**.

29. NDJE: ce qui sera le problème plus tard c'est d'expliquer pourquoi des poids appris pour reconnaître des chats/chiens sont tout à fait pertinents pour reconnaitre des bateaux/voitures, c'est-à-dire expliquer malgré tout une forme de généricité des fonctions apprises par les réseaux de neurones convolutionnels.

30. https://www.researchgate.net/publication/3078296_Barron_AE_Universal_approximation_bounds_for_superpositions_of_a_sigmoidal_function_IEEE_Trans_on_Information_Theory_39_930-945

Théorème 10 *(**A Barron 1993**)*
Si la norme ℓ^α de f est finie et prenons $\alpha = 1$, et si l'on considère la base de Fourier ($e_n = e^{i2\pi n}$) telle que

$$\sum_n |\langle f, e_n \rangle| \leq C_f \tag{107}$$

alors le réseau de neurones à M neurones est capable d'avoir une approximation $\tilde{f}_M$ telle que

$$\|f - \tilde{f}_M\|^2 \leq \frac{C_f^2}{M} \tag{108}$$

Remarquons alors que **la dimension d n'apparait plus**, et nous avons ouvert une brèche dans le mur de la dimensionalité. Par la suite on a pu jouer sur le paramètre α ($f \in \mathcal{B}^\alpha$) et sortir des résultats que l'on comprend bien en ayant vu les notions dans les cours précédents. **Donc, finie la malédiction de la dimensionalité, mais pour autant est-ce que cela explique les résultats des réseaux de neurones (convolutionels)?**

Le problème sous-jacent que l'on se pose face à ce résultat est le suivant: est-ce que la classe des fonctions $f \in \mathcal{B}^\alpha$ reflète bien la classe de fonctions que l'on rencontre en pratique et que les réseaux de neurones approximent bien? **Autrement dit, le résultat vaut pour une classe de fonctions, mais la dite classe est-elle représentative des fonctions auxquelles nous sommes en pratique confrontés (classifications d'images, de sons, analyse textuelle, régression etc.) et que les réseaux approximent très bien?** Or, malheureusement **la réponse est négative**, et la communauté s'en est rendu compte assez clairement à tel point qu'il y a une nette divergence entre ce type de théorème de mathématiques et ce qui est mis en pratique tous les jours par celles et ceux qui utilisent/mettent en œuvre des réseaux de neurones. Alors qu'est-ce à dire? Attention le théorème de Barron est juste ainsi que ceux qui lui ont emboité le pas, il n'y a pas de doute la dessus, cependant il faut se rendre compte pourquoi ce n'est pas (encore) la bonne approche.

5.2.3 Un nouveau point de vue: l'approche bayésienne

Essayons de se faire une idée de pourquoi l'approche à-la-Barron n'épuise pas le sujet. Si l'on prend le cas de la classification d'images (chien, chat, bateau, machine à

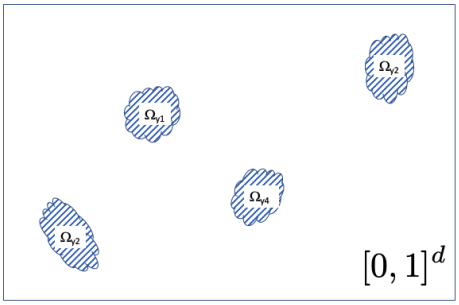

FIGURE 22 – Le volume des classes des objets Ω_y est tout petit par rapport à l'espace dans lequel évolue x

café...) mais le problème est le même avec des sons, quand on prend une image dans 99,9% des cas (pas 100% pour laisser un peu d'erreur humaine) il n'y a pas d'ambiguïté. Donc, aux différents x (images) on associe des index de classes y, et l'on tente de résoudre $y = f(x)$. On peut représenter les membres de la classe y, selon

$$\Omega_y = \{x/\ f(x) = y\} \tag{109}$$

Quand on change y, on se rend compte que le volume des Ω_y est tout petit par rapport à la taille de l'espace dans lequel évolue x (Fig. 22). C'est-à-dire qu'une image prise au hasard n'a rien à voir avec une quelconque image de chien, chat, bateau, machine à café ou que sais-je. Une image d'un bruit blanc n'a par définition aucune structure. Donc, les fonctions qui tentent d'approximer $f(x) = y$ doivent le faire bien dans un tout petit espace des images. Attention, même petits face à la dimension totale de l'espace, les ilots Ω_y peuvent être tout de même de grande dimension, c'est-à-dire que l'on ne peut pas décomposer $f(x) = y$ en petits problèmes $f_k(x) = y_k$ avec $x \in \Omega_{y_k}$ de basse dimension. Donc, les théorèmes qui tentent d'analyser (contraindre) les fonctions dans tout l'espace $[0,1]^d$, en fait se posent une question qui n'a pas vraiment lieu d'être, car ce qu'il faut analyser c'est la restriction de f au support des Ω_y. En effet, *primo* c'est le support des données, c'est-à-dire que l'on peut tenter un apprentissage, et *secundo* c'est également le support des prédictions où l'on veut que l'approximation soit bonne.

L'idée de la démarche n'est pas de regarder des fonctions plus ou moins bien régulières sur l'ensemble de l'espace $[0,1]^d$, mais plutôt se concentrer sur la *géométrie* et

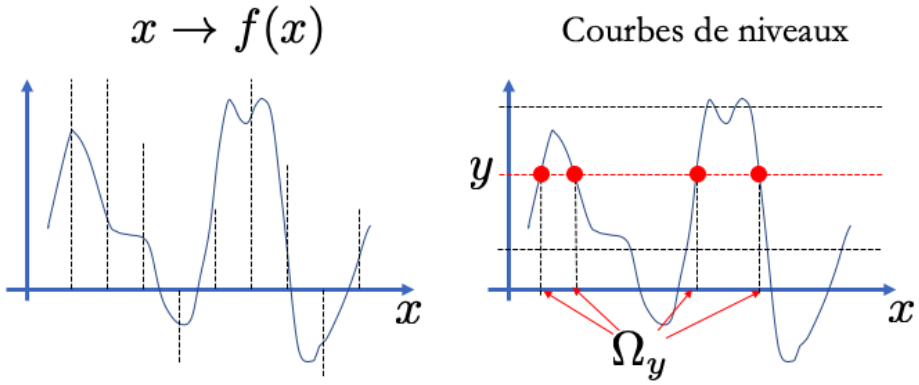

FIGURE 23 – Les deux points de vue du problème $y = f(x)$: soit on attaque le problème d'un point de vue où à chaque valeur de x on veut connaître $f(x)$ un peu d'une manière indépendante du fait que x soit un signal pertinent (schéma de gauche), soit on étudie les courbes de niveau de f (schéma de droite) et on se concentre sur les valeurs de x qui comptent au final.

le *lieu* des ilots Ω_y qui peuvent être de grande dimension. C'est exactement **le point de vue des algorithmes et le point de vue bayésien**. Attention, ce n'est pas une histoire de *probabiliste versus déterministe* qui va différentier le point de vue bayésien du précédent. L'idée est au lieu de demander "*pour n'importe quelle valeur de x, quelle est la valeur de la fonction $f(x)$?*", on va étudier la fonction à travers ses *lignes de niveaux* $f = cte$ (Fig. 23). Ces points de vue différents sur un sujet se retrouvent quand on regarde l'intégrale de Riemann (premier point de vue) et l'intégrale de Lebesgue (second point de vue).

Le point de vue bayésien est le suivant: je veux connaître le label y sachant que je connais x. Donc, l'objet est la probabilité conditionnelle $p(y|x)$, et dans le cas de la classification, pour un x_0 en question on prend le y dont la probabilité $p(y|x_0)$ est maximum. C'est le **maximum de vraisemblance** (voir discussion Cours 2018 Sec. 7.2.1). Or, le théorème de Bayes nous dit que

$$p(y|x) = \frac{p(x|y)p(y)}{p(x)} \tag{110}$$

et l'on veut trouver le maximum sur y, c'est-à-dire que la valeur approchée $\tilde{y}$ est obtenue

par

$$\tilde{y} = \underset{y}{\operatorname{argmax}} \frac{p(x|y)p(y)}{p(x)} = \underset{y}{\operatorname{argmax}}\ p(x|y)p(y) \tag{111}$$

avec $p(y)$ le *prior* qui est une information *a priori* sur l'occurrence des classes. Dans la suite, mettons que toutes les classes sont équiprobables (les volumes des Ω_y sont identiques) alors on peut laisser tomber $p(y)$ pour trouver $\tilde{y}$. Enfin, on peut maximiser le logarithme ce qui donne

$$\tilde{y} = \underset{y}{\operatorname{argmax}}\ \log p(x|y) \tag{112}$$

Donc, dans le point de vue bayésien, **il faut modéliser non pas** $f(x)$ **mais** $\log p(x|y)$ **ou** $p(x|y)$ ce qui revient à se demander connaissant le label y quel est le lieu (géométrique) de x dans l'espace $[0,1]^d$? Ce qui est justement le lieu de Ω_y. Or, en grande dimension il y a **des phénomènes de concentration** où la probabilité $p(x|y)$ est vraiment maximale sur de petits domaines Ω_y, et la plupart des images n'appartiennent à aucune classe faute de structuration. En fait, techniquement ce qu'il est plutôt pertinent d'étudier c'est la différence de probabilité entre 2 classes

$$\log p(x|y) - \log p(x|y') \tag{113}$$

car si cette différence est positive alors on attribue le label y, sinon c'est le label y' qui est la réponse.

Ce faisant, l'approche bayésienne est vraiment différente car ce n'est pas f qui est l'objet sur lequel on va réfléchir (ex. forme de régularité) et procéder à une analyse harmonique par exemple, car ce qui est important de se demander c'est où est son support et comment le caractériser. En fait, à travers $p(x|y)$, c'est l**a modélisation de** x **dans chacune des classes** qui est en sous-jacent. Ainsi, en grande dimension le problème de classification et celui de la modélisation de x sont essentiellement les mêmes, ou plus précisément **il faut s'attaquer au problème de modéliser les spécificités** de x dans Ω_y vis-à-vis des spécificités de x dans $\Omega_{y'}$. Et donc, même quand on se place dans **un problème de classification/régression** de type $y = f(x)$ où dans un premier temps on s'était penché sur f, **en définitive on retombe sur l'étude de** $x(u)$. D'où l'approfondissement de l'analyse du signal qui va nous occuper. On montrera comment **le point de vue non-linéaire va nous servir pour effectuer ces modélisations**. Cependant avant cela faisons le point sur les résultats obtenus jusqu'à présent.

5.2.4 Petit bilan

Ce que l'on a vu jusqu'à présent, c'est que l'on peut obtenir:

— soit des *approximations linéaires non-adaptatives* où essentiellement on effectue des projections sur des espaces linéaires pour lesquels la meilleure base est celle obtenue par PCA qui pour des processus stationnaires donne la base de Fourier;

— soit des *approximations non-linéaires adaptatives*, et pour que cela marche il faut s'assurer que dans la base choisie les normes ℓ^α soient finies pour garantir la décroissance de l'erreur d'approximation.

Finalement ces deux points de vue sont effectifs en basse dimension, mais dès que la dimension croit on se retrouve devant **la malédiction de la dimensionalité**, c'est-à-dire que même le théorème de Barron n'est finalement pas adapté aux cas pratiques. Il faut adopter un autre point de vue qui nous ramène à l'étude de x. Car la façon dont cette barrière va pouvoir être levée, n'est pas tant que la fonction f soit extraordinairement régulière sur $[0,1]^d$, mais plutôt que **son support**, c'est-à-dire les lieux où il est intéressant d'avoir une bonne approximation, **soit très concentré**. **Ainsi, l'enjeu est de caractériser ce support**, certes restreint mais pas de basse dimension, et donc on va revenir sur le triangle RAP pour **comprendre cette notion de régularité**.

5.3 Théorie de l'Information. Bases d'Ondelettes

On s'est rendu compte au cours de l'étude d'un réseau de neurones à 1 couche qu'il nous faut revenir à l'analyse du signal x même dans le cas de classification/régression. Nous allons nous placer dans une optique non-linéaire, car la **modélisation du signal** qui va en résulter sera de bien meilleure qualité que l'analyse de Fourier. Dans ce cadre, on va essayer de comprendre quel type de base va nous permettre de faire beaucoup mieux. En particulier, on va se poser la question de la **compression** à travers la **Théorie de l'Information** laquelle explique précisément les **phénomènes de concentration** rencontrés dans la section précédente.

Pour aborder le sujet nous allons repartir du triangle RAP. Rappelons que dans le cadre linéaire, nous étions partis de la notion de *régularité* d'une fonction (Sec. 3.3): on a pu caractériser une fonction *uniformément régulière* par le biais de ses dérivées covariantes par translation, lesquelles sont diagonalisables dans la base de Fourier, ce qui

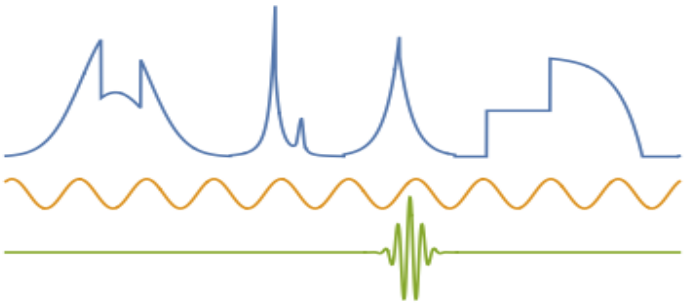

FIGURE 24 – Différents points de vue de l'analyse d'une fonction: soit le cadre linéaire qui prend pour partie une régularité uniforme ce qui donne l'analyse de Fourier avec des sinusoïdes délocalisées en temps/espace, soit le cadre non-linéaire qui étudie des fonctions non uniformément régulières avec l'analyse par Ondelettes qui opère une analyse locale des transitoires/discontinuités.

a permis de dérouler le discours sur la *parcimonie* et *l'approximation en basse dimension*. Donc, **le point de départ en linéaire est la notion de régularité uniforme**. Cependant, comme on l'a vu précédemment les signaux qui recèlent des *features* intéressantes sont ceux qui présentent des discontinuités, des transitoires: ex. les contours, les changements de rythmes, l'attaque d'une note de musique, etc. S. Mallat nous relate que si l'on change dans la trame musicale de quelques secondes d'une note produite par un violon, ne serait-ce que les 50 premières millisecondes pour la remplacer par le début de la même note mais produite par un piano, alors notre perception est totalement changée. Ce qui signifie que la perception du son est très fortement influencée par les discontinuités du signal, en l'occurence l'attaque de la note par un violon ou un piano. Donc, il faut s'intéresser à une forme de **régularité par morceaux** qui soit capable de représenter une large classe de fonctions vraiment intéressantes pour les enjeux qui nous concernent.

Pour se faire, il nous faut trouver une manière de localiser l'analyse des transitoires temporels en 1D (ou spatiaux en 2D, etc), alors que dans le cadre de Fourier les cosinus certes sont localisés en fréquences mais totalement délocalisés en temps. Ainsi, il nous faut des sinusoïdes locales à savoir des **ondelettes**[31] (Fig. 24).

31. Voir les cours de 2018 et 2020.

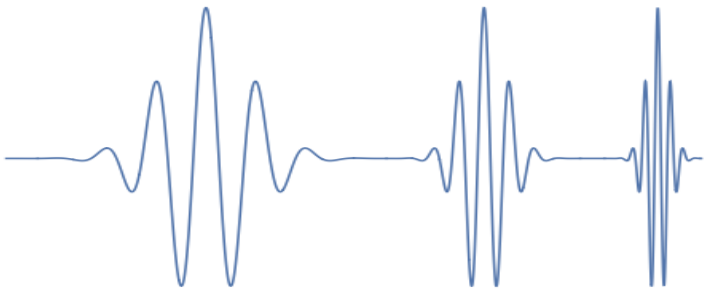

FIGURE 25 – Illustration des opérations de translation et changement d'échelle appliquée à une ondelette.

5.3.1 Analyse par Ondelettes

Une ondelette $\psi(u)$ est une fonction oscillante sur un support fini et comme elle est localisée dans l'espace, il faut pouvoir la déformer non seulement en la translatant, mais aussi en taille pour pouvoir s'ajuster avec le lieu et la taille du transitoire (Fig. 25). Ainsi, à partir de $\psi(u)$ on introduit les deux transformations en définissant la famille $\{\psi_{v,s}\}$ (ici en 1D)

$$\boxed{\psi_{v,s}(u) = \frac{1}{\sqrt{s}}\psi\left(\frac{u-v}{s}\right)} \tag{114}$$

Notons au passage la normalisation $1/\sqrt{s}$ pour obtenir une base orthonormée:

$$\|\psi_{v,s}\|^2 = \int |\psi_{v,s}(u)|^2 \, du = \|\psi\|^2 \tag{115}$$

On impose que l'ondelette oscille ce qui se traduit par la contrainte

$$\int \psi(u)du = 0 \tag{116}$$

Maintenant, comme on s'intéresse à l'analyse de la régularité locale du signal x par $\psi_{v,s}$, on définit à présent la **Transformée en Ondelettes** selon (les ondelettes sont prises

réelles ici):

$$W_x(v,s) = \langle x, \psi_{v,s} \rangle = \int x(u)\psi_{v,s}(u)\ du = \int x(u)\frac{1}{\sqrt{s}}\psi\left(\frac{u-v}{s}\right)\ du \tag{117}$$

$$= \int x(u)\tilde{\psi}_s(v-u) = (x * \tilde{\psi}_s)(v) \tag{118}$$

avec

$$\tilde{\psi}_s(u) = \frac{1}{\sqrt{s}}\psi\left(-\frac{u}{s}\right) \tag{119}$$

Ainsi, la Transformée en Ondelettes peut être vue soit comme une opération de **produit scalaire** de x avec $\psi_{v,s}$, soit comme une **convolution** entre x et le **filtre** $\tilde{\psi}_s$ [32].

Cette opération mesure **la variation locale de x au voisinage de v sur un support proportionnel à** s. Donc, vis-à-vis du triangle RAP, nous allons essayé de capturer la régularité de x, et plus précisément une régularité non-uniforme, et pour ce faire on va passer par la parcimonie en trouvant une base adaptée pour la capture d'irrégularités locales (voire de discontinuités locales).

En fait, on n'a pas tellement de choix pour construire la base, car on doit disposer d'une famille de fonctions localisées et adaptables à l'échelle/taille des irrégularités, et donc pour ainsi dire mécaniquement on trouve la transformée en ondelettes avec les bases associées. Rappelons que, s'agissant de la transformée de Fourier ou la transformée en Ondelettes, ce ne sont pas des outils que l'on choisit parmi d'autres: la base de Fourier est "la" base du cadre linéaire avec en sous-jacent l'invariance par translation, tout comme pour étudier des phénomènes transitoires on tombe sur les bases d'ondelettes.

Cependant pour aller jusqu'au bout de l'analyse du triangle RAP, il va falloir définir des approximations de basses dimensions. Or, ce type d'approximation est facile à faire dès lors que l'on a des bases orthogonales. Donc, le point critique est la construction de telles bases à partir des ondelettes. Et avant même cela, il faut démontrer que l'on capture bien la régularité des fonctions avec ces ondelettes.

32. nb. Dans le cours de 2020 le point de vue était celui des filtres convolutionnels donc la normalisation choisie était $1/s$, dans le cours de 2018 le facteur était $1/\sqrt{s}$ car l'aspect "produit scalaire" était privilégié.

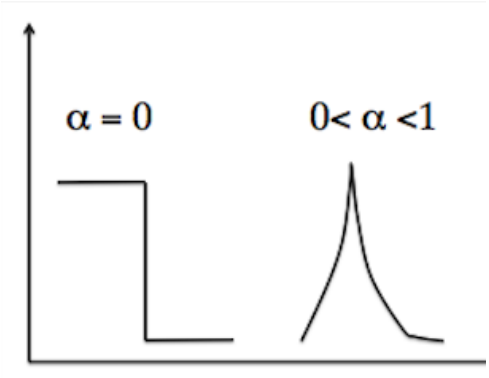

FIGURE 26 – Illustration de fonctions Lipschitz α selon la valeur de α.

5.3.2 Régularité locale de Lipschitz et décroissance des coefficients d'ondelettes

Pour ce qui concerne la régularité locale, nous allons l'aborder au sens de Lipschitz [33]:

Définition 1 *Une fonction $x(u)$ est Lipschitz α en un point v si $\exists C > 0$ telle que*

$$|x(u) - x(v)| < C|u - v|^{\alpha} \tag{120}$$

La figure 26 illustre des évolutions locales de fonctions Lipschitz $0 \leq \alpha < 1$ (nb. pour $\alpha > 1$ la fonction est localement constante, et pour un mouvement brownien on a $\alpha = 1/2 - \varepsilon$). Si on travaille sur un intervalle, on peut rendre la notion uniforme selon la définition suivante:

Définition 2 *Une fonction $x(u)$ est uniforme Lipschitz α sur un intervalle I si $\exists C > 0$ telle que*

$$\forall (u, v) \in I, \ |x(u) - x(v)| < C|u - v|^{\alpha} \tag{121}$$

Ce qu'il faut réaliser, c'est que la valeur de α peut être une constante dans le cas uniforme, mais surtout elle peut changer localement et donc nous renseigner sur la **régularité locale**

33. NDJE: Rudolph O. S. Lipschitz (1832-1903), dont Otto Ludwig Hölder (1859-1937) a établi l'extension que S. Mallat utilise.

de la fonction. Maintenant, peut-on relier la régularité lipschitzienne à la Transformée en Ondelettes?

Dans le cas de Fourier, on a été capable de relier la régularité d'une fonction à la vitesse de décroissance des produits scalaires $|\langle x, e_n\rangle|$, donc on a regardé l'évolution des produits scalaires quand la fréquence ($\omega_n = 2\pi n$) augmente. Avec les ondelettes, on va se placer autour d'un point v, et on va également augmenter la "fréquence" d'oscillation, or **l'équivalent de ω_n c'est $1/s$, c'est-à-dire le changement d'échelle**, en effet

$$\widehat{\psi_{v,s}}(\omega) = \sqrt{s}e^{-i\omega v}\hat{\psi}(s\omega) \tag{122}$$

donc si $|\hat{\psi}(\omega)|$ a un maximum en ω_0, $|\widehat{\psi_{v,s}}|$ l'a en ω_0/s. Ainsi en faisant tendre s vers 0, on devrait pouvoir apprécier la vitesse de décroissance des produits scalaires à haute fréquence et en déduire la régularité Lipschitz α de la fonction. En effet, on a le théorème suivant:

Théorème 11 *Si x est Lipschitz α en v alors $\exists C$ telle que*

$$\forall s,\ |\langle x, \psi_{v,s}\rangle| < Cs^{\alpha+1/2} \tag{123}$$

Ce théorème nous dit que plus la fonction est régulière, c'est-à-dire α proche de 1, plus la décroissance des produits scalaires $|\langle x, \psi_{v,s}\rangle|$ est grande quand $s \to 0$ ou bien à haute fréquence. On retrouve, mais autrement formulée, l'idée que nous avions avec Fourier. La différence principale est que la **vitesse de décroissance est analysée en un point précis (v)**, tandis que dans le cas de Fourier la décroissance est analysée sur un intervalle entier, et cela à tel point que si dans l'intervalle il n'y a qu'une unique discontinuité, la décroissance est fixe en $1/\omega$, et ne renseigne en rien sur le fait que la fonction peut être 100 fois dérivable en dehors de cet accident. **Avec la Transformée en Ondelettes on dispose d'un outil tel un microscope à la recherche des (ir)régularités locales**.

Le théorème 11 est-il une équivalence? En fait il y a plusieurs théorèmes. Par exemple, *il y a équivalence si par exemple on a une fonction uniformément Lipschitz α sur un intervalle arbitrairement petit.* Cependant, si on insiste sur le fait que l'on a une régularité ponctuelle, alors la réponse est que l'équivalence est "presque vraie" mais il faut

changer la borne en y apportant une correction logarithmique car il peut y a voir des accidents pour les fractales. Il s'agit d'un résultat de 1990 en "micro-localisation" (voir S. Jaffard [34]). Cependant, le théorème 11 avec la condition suffisante nous suffit.

Démonstration 11. En fait pour rétablir le résultat que la condition Lipschitz α ponctuelle donne une contrainte sur le produit scalaire, il suffit de l'écrire

$$\langle x, \psi_{v,s} \rangle = \int x(u) \frac{1}{\sqrt{s}} \psi \left(\frac{u-v}{s} \right) \, du \tag{124}$$

or on sait que l'intégrale de l'ondelette ψ est nulle, il en est de même pour $\psi_{v,s}(u)$ prise comme fonction de u, donc

$$\begin{aligned} |\langle x, \psi_{v,s} \rangle| &= \left| \int (x(u) - x(v)) \frac{1}{\sqrt{s}} \psi \left(\frac{u-v}{s} \right) \, du \right| \\ &\leq \int |x(u) - x(v)| \frac{1}{\sqrt{s}} \left| \psi \left(\frac{u-v}{s} \right) \right| \, du \\ &\leq C \int |u-v|^\alpha \frac{1}{\sqrt{s}} \left| \psi \left(\frac{u-v}{s} \right) \right| \, du \\ &\leq C \int |su'|^\alpha \sqrt{s} |\psi(u')| du' \\ &\leq C s^{\alpha+1/2} \int |u|^\alpha |\psi(u)| du \end{aligned}$$

Or, l'ondelette ψ étant localisée, l'intégrale de droite est une constante, donc le produit scalaire est bien borné comme l'indique le théorème. ■

Notez bien que la démonstration tient au fait que l'ondelette est *oscillante* et *localisée*, et que la compréhension du théorème va bien au delà d'un simple changement de variable: si la fonction est Lipschitz α alors les incréments autour de v sont multipliés par s^α ce qui est reflété par la contrainte sur les coefficients en ondelettes, c'est-à-dire les produits scalaires.

34. Par exemple le théorème 3.1 dans http://www.ens-lyon.fr/DI/wp-content/uploads/2009/07/Jaffard-IC2.pdf.

6. Séance du 10 Févr.

A la fin de la séance dernière, nous avons vu la façon dont la *régularité Lipschitz α locale* du signal peut être contenue dans les coefficients de la transformation en ondelettes. Ainsi, on a un moyen de quantifier la régularité du signal x, et s'il n'a pas trop de discontinuités, alors nous allons pouvoir construire des *représentations parcimonieuses* avec des *bases orthonormales d'ondelettes*. Ensuite, à partir de ces représentations, on va pouvoir développer des *approximations en basse dimension*, ce qui complétera l'analyse du triangle RAP que nous avons entrepris dans le *cadre non-linéaire* pour s'adapter au signal x et comprendre les phénomènes de concentration des supports des lignes de niveaux $f(x) = y$ en grande dimension.

6.1 Régularité Lipschitz α et scalogramme

Soit donc le signal $x(u)$, si l'on veut connaître la régularité au voisinage de v_0, on peut l'approximer avec la meilleure approximation polynomiale et étudier l'erreur d'approximation.

Définition 3 *(**Lipschitz** α)*
Soit $x(u)$ si

$$\exists C > 0 \text{ tq. } \forall u \quad |x(u) - p_{v_0}(u)| \leq C|u - v_0|^\alpha \tag{125}$$

avec $m - 1 \leq \alpha \leq m$ et p_{v_0} un polynôme de degré $m - 1$, alors on dira que x est une fonction Lipschitz α en v_0.

Ainsi, **on peut voir la régularité locale d'une fonction comme l'erreur d'une approximation polynomiale**. On retrouve le résultat du développement de Taylor: si x est m fois dérivable alors le reste de Taylor est un $o((u - v_0)^m)$. La régularité de Lipschitz est une extension au cas où α est un réel. Maintenant, si $x(u)$ est une fonction bien régulière, α est relativement grand, l'approximation polynomiale est d'autant plus effective, et donc on peut approximer x avec peu de paramètres. Mais, si α change en fonction de v_0, comment peut-on obtenir malgré tout une représentation parcimonieuse?

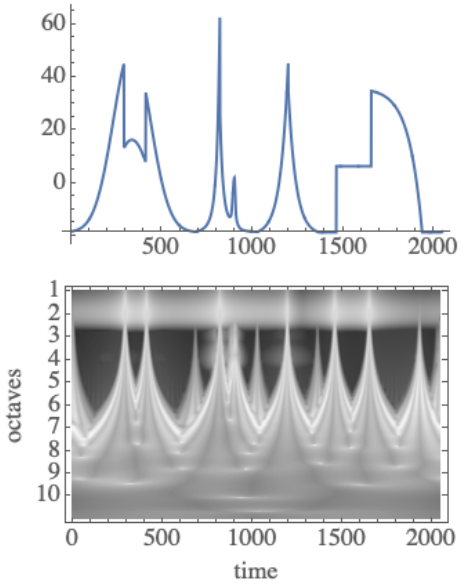

FIGURE 27 – Exemple de Scalogramme (bas) issu de la transformation en ondelettes du signal du schéma en haut. En abscisse la variable u (ex. temps), en ordonnée l'échelle $s = s_0\, 2^j\, 2^{n/Q}$ avec j l'octave, n la "voix" et $Q = 16$ le nombre de voix par octave, et s_0 la plus petite échelle de l'ondelette. L'intensité des gris donne la valeur de $|\log(W_x(v,s))|$ sur une échelle où 1 correspond au noir et 0 au blanc.

A la dernière séance, nous avons vu le théorème 11 qui permet d'encoder la régularité Lipschitz α dans la décroissance des coefficients d'ondelettes. Sur la figure 27 est montré un exemple de signal $x(u)$ avec des irrégularités à plusieurs endroits, et le résultat de la transformation en ondelettes où l'on a porté en intensité de couleur $|\log(W_x(v,s))|$ selon v en abscisse et l'échelle s en ordonnée selon $s = s_0\, 2^j\, 2^{n/Q}$ (octave j, voice/voix n) avec $Q = 16$ le nombre de voix par octave, et s_0 la plus petite échelle de l'ondelette. C'est ce qu'on appelle un **scalogramme**. Les hautes fréquences (petites échelles) sont en haut, et les basses fréquences (grandes échelles) sont en bas. On voit très nettement l'effet des discontinuités du signal. Pour mémoire, l'ondelette utilisée est celle de Morlet réelle [35].

Rappelons que l'ondelette ψ est au minimum de moyenne nulle (Eq. 116) donc son

35. Il s'agit de la fonction $\psi(x) = \pi^{-1/4} e^{-x^2/2}(\cos(\pi x(2/\log 2)^{1/2}) - e^{-\pi^2/\log 2})$ où la constante est là pour assurer l'intégrale nulle, et dont la TF donne $\hat{\psi}(\omega) = 2^{3/2}\pi^{1/4} e^{-\omega^2/2} e^{-\pi^2/\log 2} \sinh^2(\omega\pi/\sqrt{\log 4})$.

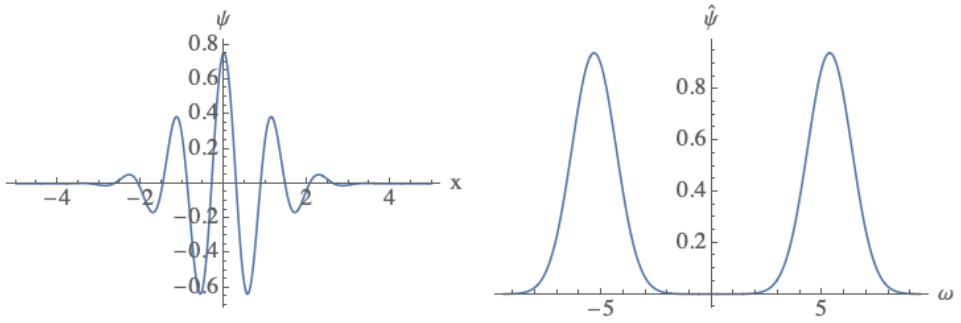

FIGURE 28 – L'ondelette ψ est de moyenne nulle, c'est donc un filtre passe bande. Illustration avec l'ondelette de Morlet réelle.

spectre de Fourier est celui d'un **filtre passe-bande** (Fig. 28). On va de plus imposer à ψ d'avoir m **moments nuls**, à savoir

$$\forall k,\ 0 \leq k < m, \quad \int \psi(u) u^k \, du = 0 \tag{126}$$

Par exemple, l'ondelette de Morlet réelle citée ci-dessus a 1 moment nul en extra. Donc, si ψ a m moments nuls, alors il vient naturellement que pour tout polynôme $p(x)$ de degré $d < m$

$$\int \psi(u) p(u) du = 0 \tag{127}$$

Pourquoi est-ce important d'utiliser de telles ondelettes? La raison en est qu'ainsi l'ondelette "ignore" la partie polynomiale de $x(u)$, et elle n'est sensible qu'à *l'erreur d'approximation* polynomiale. Si le coefficient d'ondelette W_x est petit, c'est le signe d'une petite erreur et vice-versa. Donc, dans le scalogramme de la figure 27 on ne voit apparaitre que les coefficients de grande valeur qui sont le reflet d'une interférence cohérente de l'ondelette avec le signal x au voisinage v et pour une échelle s ($\psi_{s,v}$). Plus l'échelle augmente plus l'ondelette se dilate et elle délocalise les discontinuités d'où l'apparition de "cônes" qui s'évasent aux endroits des valeurs de v où il y a une discontinuité. Ceci dit si le signal $x(u)$ possède des discontinuités presque partout alors le scalogramme est pour ainsi dire rempli de cônes comme sur la figure 29, et sa lecture devient difficile mais **la figure reflète à toutes les échelles la régularité/l'irrégularité du signal** $x(u)$. Donc dans un premier temps, nous allons analyser les propriétés du scalogramme, et dans un second temps nous tenterons de concentrer le maximum d'information sur un minimum de coefficients, ce qui

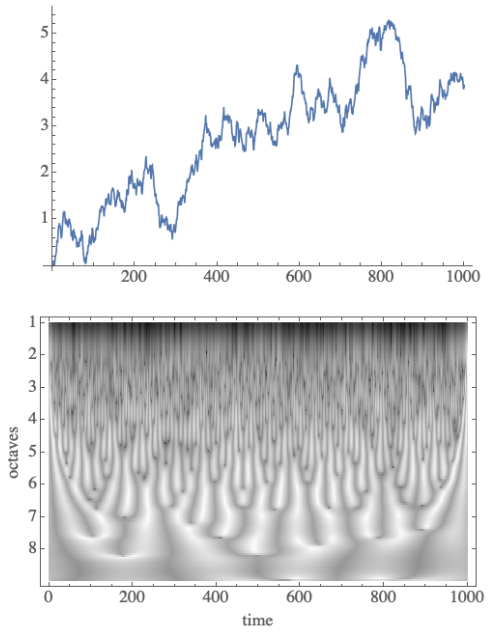

FIGURE 29 – Scalogramme d'un processus de type mouvement brownien.

donnera lieu à un échantillonnage du scalogramme qui aboutira aux bases orthonormales d'ondelettes.

6.2 Approfondissement de l'étude du scalogramme

Revenons sur le théorème 11 pour en donner une version plus précise. Il s'agit d'une version qui a été démontrée par S. Jaffard et qui caractérise vraiment l**a régularité ponctuelle de la fonction** et qui met en lumière les fameux **cônes**, la voici:

Théorème 12 ***(S. Jaffard)***
Soit une ondelette ψ a m moments nuls. Si x est Lipschitz $\alpha \leq m$ en un point v_0 alors $\exists C > 0$ telle que

$$|W_x(v,s)| \leq Cs^{\alpha+1/2}\left(1+\left|\frac{v-v_0}{s}\right|^{\alpha}\right) \tag{128}$$

et ***inversement*** [a]*, si* $\alpha' > \alpha$ *on a*

$$|W_x(v,s)| \leq Cs^{\alpha+1/2}\left(1+\left|\frac{v-v_0}{s}\right|^{\alpha'}\right) \tag{129}$$

alors x *est Lipschitz* α *en* v_0*.*

a. notez la différence de l'exposant au niveau du cône.

Ce résultat nous indique que lorsque s tend vers 0 (c'est-à-dire les hautes fréquences) alors la décroissance dépend de α. Sur la figure 27 à grande échelle la fonction a des variations douces entre son minimum et son maximum, et les coefficients sont assez uniformes le long de l'axe u, par contre plus l'échelle diminue, plus la localisation des discontinuités se fait jour. Le cône est défini par $|v-v_0|/s \leq 1$ car alors la vitesse de décroissance est dominée par $s^{\alpha+1/2}$, et la puissance α de $|v-v_0|/s$ vient contrôler des cas de figure de singularité de type $\sin(1/u)$. Voyons comment se démontre la première partie du théorème, la seconde plus technique ne sera pas abordée ici [36].

Démonstration 12.
Exprimons le coefficient $W_x(v,s)$, nous avons:

$$W_x(v,s) = \int x(u)\frac{1}{\sqrt{s}}\psi\left(\frac{u-v}{s}\right)\,du$$

Or, le signal $x(u)$ est approximé par un polynôme $p_{v_0}(u)$ de degré $m-1$, donc étant donné que ψ à m moments nuls alors $\int p_{v_0}(u)\psi(u)du = 0$. Notons que cette propriété se transmet à $\psi_{v,s}(u)$ simplement par changement de variable. Alors avec la même logique que la démonstration du théorème 11, on peut écrire

$$W_x(v,s) = \int (x(u)-p_{v_0}(u))\frac{1}{\sqrt{s}}\psi\left(\frac{u-v}{s}\right)\,du$$

36. voir Prop. 3.2 du document cité en footnote précédemment à propos de S. Jaffard, ainsi que les liens sur les Chapitres du livre de S. Mallat.

qui traduit le fait que l'ondelette n'est pas sensible à la régularité polynomiale de x. Ainsi,

$$\begin{aligned}|W_x(v,s)| &\leq \int |(x(u)-p_{v_0}(u))|\frac{1}{\sqrt{s}}\left|\psi\left(\frac{u-v}{s}\right)\right| \, du \\ &\leq C\int |u-v_0|^\alpha|\frac{1}{\sqrt{s}}\left|\psi\left(\frac{u-v}{s}\right)\right| \, du \\ &\leq Cs^{1/2}\int |su'+v-v_0|^\alpha|\psi(u')| \, du'\end{aligned}$$

Or $\forall a, b$ on peut montrer que $|a+b|^\alpha \leq 2^\alpha(|a|^\alpha+|b|^\alpha)$ donc

$$\begin{aligned}|W_x(v,s)| &\leq 2^\alpha Cs^{1/2}\int (|su'|^\alpha+|v-v_0|^\alpha)|\psi(u')| \, du' \\ &\leq 2^\alpha Cs^{1/2+\alpha}\left(\int |u'|^\alpha|\psi(u')| \, du' + \left|\frac{v-v_0}{s}\right|^\alpha \int |\psi(u')| \, du'\right)\end{aligned}$$

Il apparaît deux intégrales dépendant de l'ondelette ψ, ce sont deux constantes dont on peut prendre le max et le mettre en facteur, ce qui termine la démonstration. ■

Pour démontrer la "réciproque", il faut partir des coefficients d'ondelettes W_x et remonter à la fonction x, ce qui se fait en premier lieu en montrant que **la Transformée en Ondelettes est inversible**. Nous verrons ce point fondamental par la suite (Voir également le cours de 2020).

Donc, le théorème nous donne une possibilité d'interpréter le scalogramme, cependant il faut bien reconnaître que cette image est loin d'être une représentation parcimonieuse car nous sommes partis d'un signal 1D pour obtenir une image 2D, et la question se pose de savoir si l'on peut comprimer l'information. Jean Morlet et Alex Grossmann ont joué un rôle de premier plan dans le développement de la Transformation Continue en Ondelettes. J. Morlet (1931-2007) ingénieur chez Elf-Aquitaine étudiait les couches géologiques par l'analyse d'ondes sismiques, tandis que A. Grossmann (1930-2019) physicien franco-croate y a vu des parallèles avec les états cohérents en physique quantique. C'est alors qu'une pluralité de mathématiciens de différents domaines ont convergé pour étudier la Transformée en Ondelettes qui analyse un signal à différentes échelles (Voir les cours des années précédentes). En particulier, comment peut-on discrétiser la représentation?

6.3 Vers une représentation parcimonieuse: une double discrétisation

6.3.1 Discrétisation des échelles

Dans un premier temps fixons nous une échelle discrète des échelles [37] avec $s = 2^j$. Nous allons montrer que moyennant une condition sur l'ondelette alors se fixer ce type de discrétisation suffit, c'est-à-dire que disposant des coefficients $W_x(v, 2^j)$ on peut retrouver x. Rappelons nous que $W_x(v, 2^j)$ peut être vu comme la convolution avec le filtre $\tilde{\psi}_{2^j}$ (Eq. 119). Or qui dit convolution, dit Transformation de Fourier, ainsi

$$\widehat{W_x}(\omega, 2^j) = \widehat{x}(\omega)\ \widehat{\tilde{\psi}_{2^j}}(\omega) \tag{130}$$

avec

$$\widehat{\tilde{\psi}_{2^j}}(\omega) = \sqrt{2^j}\ \hat{\psi}^*(2^j\omega) \tag{131}$$

La question devient alors peut-on reconstruire $\widehat{x}(\omega)$ à partir des coefficients $\widehat{W_x}(\omega, 2^j)$? Cela n'est possible que si le spectre de Fourier est totalement couvert par les filtres $\hat{\psi}(2^j\omega)$. Pour l'ondelette de base ($j = 0$), le filtre est celui d'un passe-bande. Ainsi, il suffit que les supports des filtres dilatés/contractés se recouvrent de proche en proche pour que si l'on prend tous les $j \in \mathbb{Z}$ alors l'ensemble du spectre de Fourier soit couvert, c'est-à-dire sans trou. Une illustration de l'évolution des filtres $\hat{\psi}(2^j\omega)$ pour différentes valeurs de j [38] est donnée sur la figure 30.

Le théorème suivant nous permet de formaliser le propos [39]

Théorème 13 *Si on dispose de la condition de Littlewood-Paley suivante*

$$\forall\omega,\ \sum_{j\in\mathbb{Z}} |\hat{\psi}(2^j\omega)|^2 = 1 \tag{132}$$

37. Cette discrétisation est sous-jacente dans les images des scalogrammes présentés sur les figures 27 et 29, car en pratique, on ne dispose que d'un échantillon fini de valeurs de $x(u)$.

38. Il s'agit de l'ondelette $\psi_\sigma(u) = 2\pi^{-1/4}/\sqrt{3\sigma}(1 - t^2/\sigma^2)e^{-1/2(t/\sigma)^2}$ dont la transformée de Fourier est donnée par $\hat{\psi}_\sigma(\omega) = 2\sqrt{2/3}\pi^{1/4}\sigma^{5/2}\omega^2 e^{-1/2(\sigma\omega)^2}$. Pour l'illustration $\sigma = 2$.

39. Voir Cours 2018 et 2020.

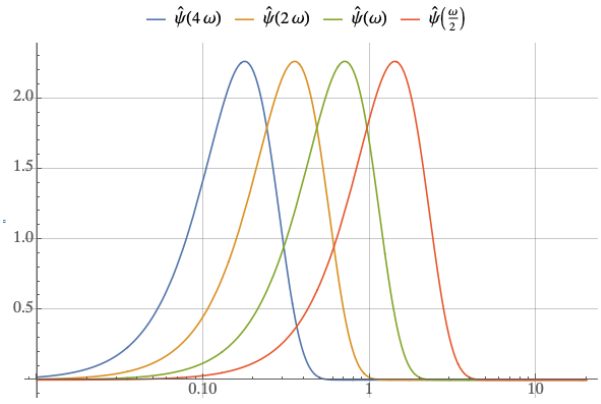

FIGURE 30 – Illustration de l'évolution du support du filtre passe-bande donné par $\hat{\psi}(2^j\omega)$: pour $s = 2^j > 1$ le support se déplace vers les basses fréquences, et pour $s < 1$ le support se déplace vers les hautes fréquences.

alors

$$x(u) = \sum_{j\in\mathbb{Z}} 2^{-j}(\widehat{W_x}(\omega, 2^j) * \psi_{2^j})(u) \tag{133}$$

La démonstration est immédiate comme d'habitude en utilisant la transformée de Fourier du produit de convolution. Ainsi, **moyennant de bien couvrir le spectre de Fourier du signal, on peut se restreindre à n'utiliser que des échelles dyadiques** $s = 2^j$ $(j \in \mathbb{Z})$. Mais on veut faire plus, à savoir discrétiser l'axe u, c'est-à-dire **procéder à un échantillonnage** du signal.

6.3.2 Discrétisation de la variable "espace"

Quand on écrit $W_x(v, 2^j)$, la variable v de genre espace ou temps est continue, comment peut-on la discrétiser? Utilisons à nouveau la formulation en termes de filtrage du signal x, à savoir

$$W_x(v, 2^j) = (x * \tilde{\psi}_s)(v) \tag{134}$$

et la question revient à se demander comment choisir judicieusement les valeurs de v pour ne perdre aucune information? A-t'on une intuition sur le sujet? La réponse est oui car la taille de l'ondelette selon v est proportionnelle à 2^j, donc il faut recourir à un

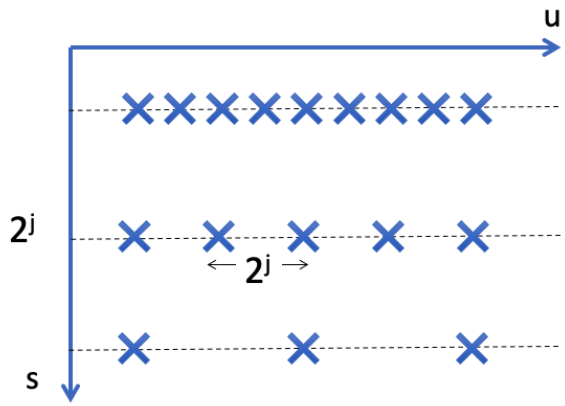

FIGURE 31 – Échantillonnage optimal selon l'axe "temps/espace" en accord avec la discrétisation dyadique des échelles.

échantillonnage également proportionnel à 2^j pour pouvoir bien couvrir tout le support de x. En fait le résultat est plus précis car on va montrer que le pas d'échantillonnage est égal à 2^j et donc les échantillons sont $v_n = 2^j n$ (Fig. 31). Donc, on disposera des coefficients $W_x(2^j n, 2^j)$ obtenus à l'aide d'ondelettes $\psi_{2^j n, 2^j}$, mais pour alléger les notations on écrira $\psi_{j,n}$. Si on utilise le point de vue "produit scalaire", on sait que

$$W_x(2^j n, 2^j) = \langle x, \psi_{j,n} \rangle \tag{135}$$

et la question est de savoir si la famille de fonctions $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ est **une base orthonormale** par exemple de $L^2(\mathbb{R})$ ou $L^2([0,1]))$ selon le problème? car alors on pourra reconstruire le signal.

6.3.3 Bases orthonormales?

La question de trouver des bases orthonormales de ce type n'est pas nouvelle, puisque **Alfred Haar** (1885-1933), mathématicien hongrois, en donne un exemple en 1909 (Fig. 32). On peut se rendre compte assez facilement que le produit scalaire de deux ondelettes de Haar est nul, et donc il vient assez naturellement que la famille $\{\psi_{j,n}^{Haar}\}_{(j,n)\in\mathbb{Z}^2}$ est une base orthonormale de $L^2(\mathbb{R})$.

Dans les années 1948-49, **Claude Shannon** (1916-2001) et **Harry Nyquist** (1889-

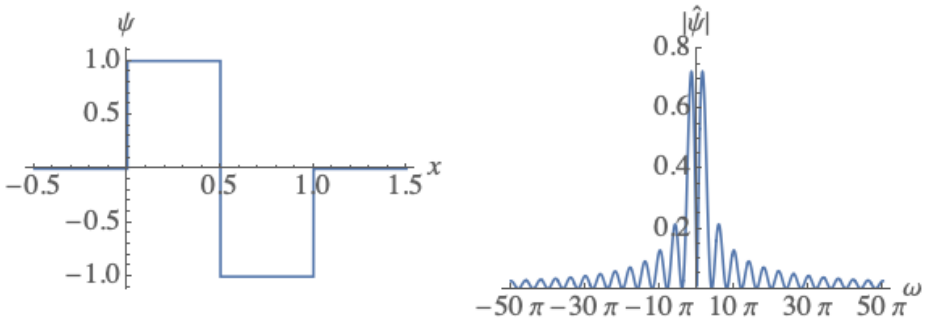

FIGURE 32 – Ondelette ψ de Haar (également "Db1" dans la famille des ondelettes d'I. Daubechies) construite dans l'espace réel. Décroissance selon $1/\omega$ dans l'espace de Fourier.

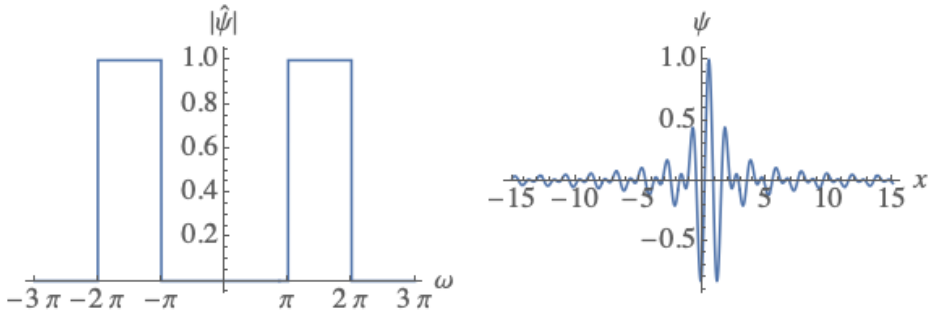

FIGURE 33 – Ondelette de Shannon construite dans l'espace de Fourier. Décroissance en $1/u$ dans l'espace réel.

1976) démontrent un théorème connu [40] sous le nom de *théorème d'échantillonnage de Shannon*. Même si Shannon ne fait pas mention d'une quelconque ondelette, il utilise un **filtre passe-bande parfait** (Fig. 33). On s'aperçoit que les filtres $\hat{\psi}_j^{Sha}(\omega) \propto \hat{\psi}^{Sha}(2^j\omega)$ ont des supports $[2^{-j}\pi, 2^{-j+1}\pi]$ (idem du coté des fréquences négatives). Il est donc facile d'obtenir la condition de Littlewood-Paley (Th. 13). Ainsi, on peut reconstruire le signal, ce que démontre d'une autre manière le théorème d'échantillonnage.

Donc, il y avait ces deux exemples connus depuis très longtemps, et la question qui

40. *Whittaker–Nyquist–Kotelnikov–Shannon* si on veut être un peu plus exhaustif sur l'historicité du théorème.

resta longtemps sans réponse est de savoir s'il existe d'autres types de fonctions qui ont ces propriétés. Cependant, les deux exemples présentent des inconvénients: l'ondelette de Haar est discontinue et n'a pas de moments nuls au-delà de son intégrale nulle donc ne peut absorber des régularités polynomiales; tandis que celle de Shannon dans l'espace réel [41] décroit en $1/u$, ce qui n'est pas une bonne indication de localisation spatiale. On aimerait avoir en fait des ondelettes bien localisées spatialement, bien régulières avec un nombre suffisant de moments nuls, et cerise sur le gâteau également localisées en Fourier. Notez que l'on ne peut localiser aussi bien dans les 2 espaces à cause du **Principe d'Incertitude** (Voir cours de 2020).

Il faut dire qu'il semblait impossible de satisfaire toutes ces contraintes. En fait, Roger Balian et Francis Low, deux physiciens théoriciens, avaient démontré qu'il n'était pas possible de construire des bases orthonormales de $L^2(\mathbb{R})$ de la forme $g_{m,n}(u) = e^{2\pi imu}g(u-n)$ avec $(m,n) \in \mathbb{Z}$ à la fois localisées dans l'espace réel et dans l'espace de Fourier. Ce fût alors une surprise totale, et le résultat remarquable de **Yves Meyer** en 1986 de trouver une telle famille de fonctions tout en voulant démontrer que ce n'était pas possible!

L'ondelette d'Y. Meyer est à la fois C^∞ et à décroissance rapide. Elle est construite dans l'espace de Fourier selon

$$\hat{\psi}^{Meyer}(\omega) = \begin{cases} \frac{1}{\sqrt{2\pi}} e^{i\omega/2} \sin\left[\frac{\pi}{2}\nu\left(\frac{3}{2\pi}|\omega|-1\right)\right] & \frac{2\pi}{3} \leq |\omega| \leq \frac{4\pi}{3} \\ \frac{1}{\sqrt{2\pi}} e^{i\omega/2} \cos\left[\frac{\pi}{2}\nu\left(\frac{3}{4\pi}|\omega|-1\right)\right] & \frac{4\pi}{3} \leq |\omega| \leq \frac{3\pi}{3} \\ 0 & \text{ailleurs} \end{cases} \tag{136}$$

avec $\nu(x)$ une fonction C^k ou C^∞ satisfaisant

$$\nu(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x \geq 1 \end{cases} \qquad \text{et} \qquad \nu(x) + \nu(1-x) = 1 \tag{137}$$

Un exemple est montré sur la figure 34.

A la suite de ce résultat, il s'en est suivi l'élaboration d'un cadre conceptuel mathématique qui a permis de mettre tous les résultats ensemble et de construire de nouvelles

41. Selon la définition de la TF, on trouve $\psi^{Sha}(u) = \mathrm{sinc}(t/2) - 2\,\mathrm{sinc}(t)$ avec $t = (2u-1)\pi$.

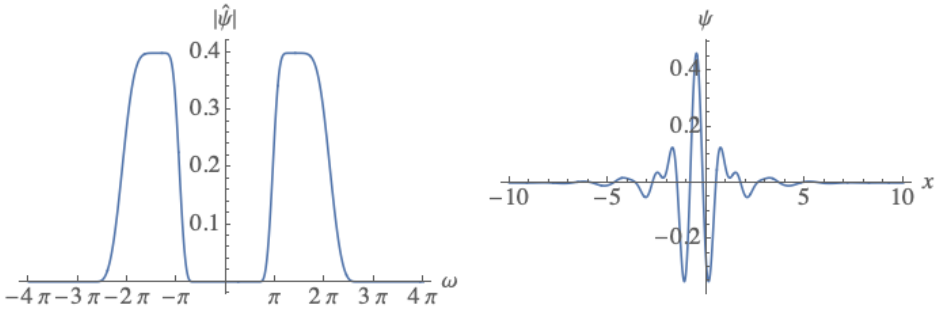

FIGURE 34 – Exemple d'une ondelette de Meyer dont la construction se fait dans l'espace de Fourier. Notez sa décroissance rapide dans l'espace réel tout en étant localisée dans l'espace de Fourier.

bases orthonormales satisfaisant aux différents critères de régularité et localisation dans l'espace réel et dans celui de Fourier. Il s'agit de l'**Analyse multirésolution** [42].

Cependant afin d'élaborer ce cadre des **bases d'ondelettes orthonormales**, il nous faut introduire des résultats en traitement du signal. En premier lieu le **théorème d'échantillonnage de Shannon**, mais sous une forme plus générale que celle dont on a l'habitude. Sa généralisation nous conduira aux **multirésolutions**. L'idée générale est la suivante: **les ondelettes fournissent des détails de la fonction** analysée, et donc on peut améliorer l'approximation de cette fonction en agrégeant progressivement des détails de plus en plus fins. Et l'analyse multirésolution conduit naturellement dans le domaine des **approximations multigrilles** que l'on retrouve en analyse numérique et probabilité. On aboutira à la construction des bases orthonormales et nous ferons le lien avec les **algorithmes de bancs de filtres** dont on s'est aperçu qu'ils sont à **la base des réseaux de neurones profonds** sauf que ces derniers incluent une non-linéarité fondamentale (voir Cours 2020). Enfin, nous arriverons aux approximations de basse dimension et la notion de parcimonie qui était la motivation initiale. Rappelons nous: si une fonction est régulière alors on applique l'analyse harmonique de Fourier, si par contre la fonction a des discontinuités locales, il faut se tourner vers les analyses multirésolutions.

42. Voir Cours 2018 pour une introduction AMR ou MRA et celui de 2020 pour un autre aspect.

6.4 Théorème d'échantillonnage de Shannon

Tout d'abord mettons en place le résultat suivant qui nous dit que **l'échantillonnage d'un signal est équivalent dans l'espace de Fourier à une périodisation**:

Théorème 14 *Si $\{x(nT)\}_n$ est un échantillonnage du signal $x(u)$ $(u \in \mathbb{R})$ alors la série de Fourier*

$$\sum_{n\in\mathbb{Z}} x(nT)e^{-inT\omega} = \frac{1}{T}\sum_{k\in\mathbb{Z}} \hat{x}\left(\omega - \frac{2k\pi}{T}\right) \tag{138}$$

C'est un théorème fondamental du traitement du signal. Sa démonstration peut être vue comme une conséquence de la *formule sommatoire de Poisson*[43].

Démonstration 14. Appelons $\hat{a}(\omega)$ la fonction du membre de droite de l'égalité, elle est $2\pi/T$-périodique donc on peut l'écrire selon

$$\hat{a}(\omega) = \sum_{n\in\mathbb{Z}} a(n)e^{-inT\omega}$$

et il nous faut démontrer que les $a(n)$ sont les $x(nT)$ du membre de gauche de l'égalité. Les $a(n)$ sont donnés selon

$$\begin{aligned}
a(n) &= \frac{T}{2\pi}\int_0^{2\pi/T} \hat{a}(\omega)e^{inT\omega}\, d\omega \\
&= \frac{T}{2\pi}\int_0^{2\pi/T} \frac{1}{T}\sum_{k\in\mathbb{Z}} \hat{x}\left(\omega - \frac{2k\pi}{T}\right) e^{inT\omega}\, d\omega \\
&= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \int_0^{2\pi/T} \hat{x}\left(\omega - \frac{2k\pi}{T}\right) e^{inT\omega}\, d\omega \\
&= \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \int_{2k\pi/T}^{(2k+1)\pi/T} \hat{x}(\omega')e^{inT\omega'}\, d\omega' \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{x}(\omega')e^{inT\omega'}\, d\omega' = x(nT)
\end{aligned}$$

43. NDJE: On peut formuler l'égalité selon la définition de la TF prise dans ce document comme $\sum_{n\in\mathbb{Z}} f(nT) = 1/T\sum_{k\in\mathbb{Z}} \hat{f}(2\pi k/T)$ qui relie les échantillonnages dans l'espace réel et l'espace de Fourier. Également on peut remarquer que $e^{-inT\omega} = \int_{-\infty}^{\infty} \delta(u-nT)du$ et on peut donc travailler sur l'égalité de deux TF mais en considérant des distributions, en particulier $T\sum_n \delta(u-nT) = \sum_k e^{i2k\pi u/T}$.

(nb. l'inversion somme-intégrable peut être effectuée si la somme $\sum_{k\in\mathbb{Z}} \hat{x}\,(\omega - 2k\pi/T)$ converge uniformément, ce qui demande quelques hypothèses de régularité $x(u)$ que l'on suppose vérifiées.) ■

Le théorème d'échantillonnage nous permet de reconstruire $x(u)$ à partir des $x(nT)$. Or, on peut se convaincre facilement que si la fonction est très irrégulière entre les échantillons, on ne peut assurément pas reconstruire le signal. Il faut donc des **hypothèses de régularité**, or cela implique automatiquement une **contrainte sur la décroissance des coefficients** de Fourier. Imaginons alors que les supports des fonctions $\hat{x}\left(\omega - \frac{2k\pi}{T}\right)$ **ne se recouvrent pas** (***pas d'aliasing***), c'est-à-dire que le support de $\hat{x}(\omega)$ soit compris dans $[-\pi/T, +\pi/T]$ ce qui est une contrainte forte, alors on peut faire du **filtrage basse fréquence**, ce qui dans le domaine spatial/temporel dit faire une **convolution** avec une fonction ***sinus cardinal***.

Théorème 15 *(**Shannon**)*
Si le support de $\hat{x}(\omega)$ est inclus dans $[-\pi/T, +\pi/T]$ alors

$$x(u) = \sum_{n\in\mathbb{Z}} x(nT)\,\phi_T(u - nT) \tag{139}$$

avec

$$\phi_T(u) = \frac{\sin(\pi u/T)}{\pi u/T} = \mathrm{sinc}(\pi u/T) \tag{140}$$

dont la Transformée de Fourier est le filtre passe-bas idéal

$$\widehat{\phi_T}(\omega) = T\ \mathbf{1}_{[-\pi/T,+\pi/T]}(\omega) \tag{141}$$

NDJE: Une proposition de démonstration peut se développer de la sorte en s'appuyant sur le théorème 14. Prenons la TF du membre de droite, il vient

$$\sum_{n\in\mathbb{Z}} x(nT)e^{-i\omega nT}\widehat{\phi_T}(\omega) = \sum_{k\in\mathbb{Z}} \hat{x}\left(\omega - \frac{2k\pi}{T}\right)\mathbf{1}_{[-\pi/T,+\pi/T]}(\omega)$$

Or, comme le support de $\hat{x}(\omega)$ est inclus dans $[-\pi/T, +\pi/T]$ il en est de même pour $\omega - 2k\pi/T$, ce qui contraint mécaniquement à ce que $k = 0$. Ainsi, on retrouve $\hat{x}(\omega)$ la TF du membre de gauche.

Regardons ce théorème classique sous un angle différent. Il nous dit que si le support

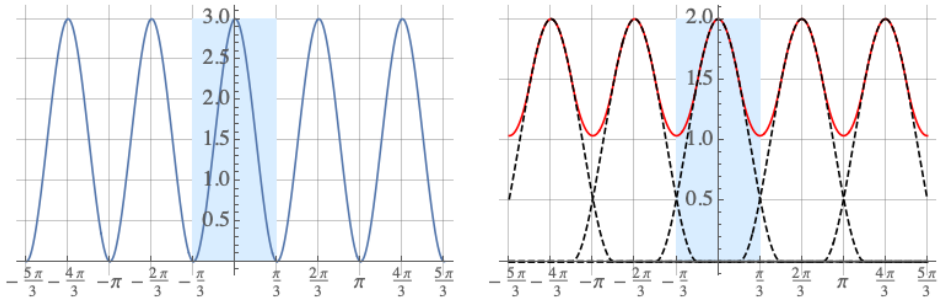

FIGURE 35 – Spectres de Fourier avec les réplications tous les $2k\pi/T$ avec à gauche un cas de figure où il n'y a pas d'aliasing, contrairement au cas de figure du schéma de droite (ici $T = 3$). La bande de base est illustrée par le rectangle bleu.

est contenu dans $[-\pi/T, +\pi/T]$ alors tout se passe bien. Mais que se passe-t'il dès lors que nous ne sommes pas dans ce cas? Comme illustré sur la figure 35 (droite), une fois que l'on coupe en ligne de base $\omega \in [-\pi/T, \pi/T]$ alors le spectre de Fourier ne correspond plus à celui de $x(u)$, et on ne peut donc reconstruire le signal. En traitement du signal, afin d'éviter l'aliasing, on commence par **préfiltrer le signal** $x(u)$ dans la bande $[-\pi/T, +\pi/T]$ à l'aide du filtre passe-bas idéal $\phi_T(u)$:

$$x \xrightarrow{\text{préfiltrage}} (x * \phi_T)(u) = x_T(u) \tag{142}$$

Ainsi on obtient un signal $x_T(u)$ qui n'a plus de composantes hautes fréquences et plus précisément:

$$\widehat{x_T}(\omega) = T\ \hat{x}(\omega)\ \mathbf{1}_{[-\pi/T, +\pi/T]}(\omega) \tag{143}$$

c'est-à-dire que le support de $\widehat{x_T}$ est bien inclus dans l'intervalle $[-\pi/T, +\pi/T]$. Ainsi, on évite bien l'aliasing comme illustré sur la figure 36.

Ce procédé de préfilrage basse fréquence est un **exemple d'approximation linéaire**. En fait, on fait en sorte que le signal soit approximé par un élément d'un espace linéaire défini par

$$V_T = \{x \ / \ \text{Suppport}\ \hat{x} \subset [-\pi/T, +\pi/T]\} \tag{144}$$

Il s'agit en fait de l'utilisation **d'une interpolation** à l'aide des fonctions $\phi_T(u)$. Sur la figure 37 à gauche se trouve en bleu le signal $x(u)$, un échantillonnage avec $T = 1/2$

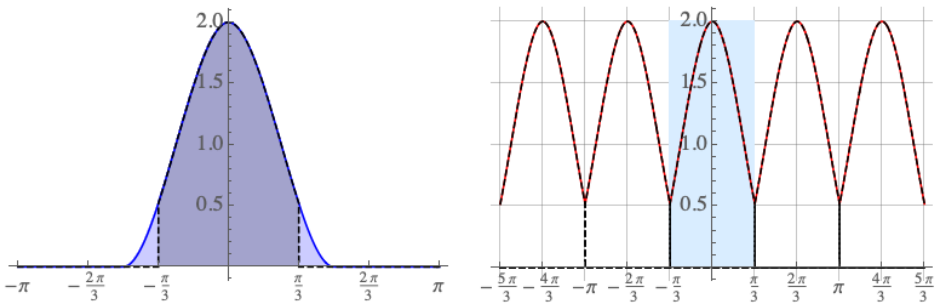

FIGURE 36 – Pour éviter l'aliasing, on préfiltre le signal dans la bande $[-\pi/T, +\pi/T]$.

et la contribution $x(nT)\phi_T(u - nT)$ pour $nT = 7$. On note au passage que les zéros de $\phi_T(u - nT)$ sont $u_k = kT$ avec $k \in \mathbb{Z}^*$ donc quand on somme les différentes contributions on obtient une fonction qui passe par tous les points (rouge) de l'échantillonnage $x(nT)$. Le résultat pour $T = 1/2$ est donné par la courbe en vert pointillé sur la figure 37 à droite. Pour illustrer l'effet du pas d'échantillonnage, l'approximation avec $T = 1$ est également en orange pointillé.

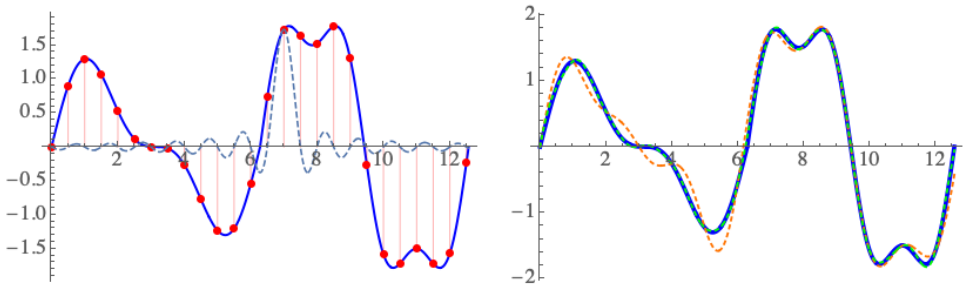

FIGURE 37 – Illustration de la formule d'interpolation Eq. 139 du théorème de Shannon (Th. 15). En bleu le signal $x(u)$ échantillonné avec $T = 1/2$ (points rouges). La contribution $x(nT)\phi_T(u - nT)$ pour $nT = 7$ est montrée sur la figure de gauche, tandis que l'approximation qui résulte de la somme de toutes les contributions est donnée sur la figure de droite en pointillés verts (difficile à distinguer de la courbe bleue). L'approximation de moins bonne qualité obtenue avec $T = 1$ est également donnée en pointillés orange.

En fait la famille de fonctions $\{\phi_{T,n}(u) = \phi_T(u - nT)\}_{n\in\mathbb{Z}}$ est une **base orthogonale**

de V_T. Primo, toutes les fonctions sont membres de V_T car leur support en Fourier est celui de $\widehat{\phi_T}$ donc compris dans $[-\pi/T, +\pi/T]$. Secundo, le produit scalaire entre $\phi_{T,n}$ et $\phi_{T,m}$ est vu par Plancherel comme l'intégrale du produit des TF de ces deux fonctions, à savoir

$$\langle \phi_{T,n}, \phi_{T,m} \rangle = \int \widehat{\phi_{T,n}}(\omega) \widehat{\phi_{T,m}}^*(\omega)\ d\omega \propto \int_{-\pi/T}^{\pi/T} e^{-i\omega T(n-m)}\ d\omega \propto \delta(n-m)$$

Ayant cette base de V_T, on sait que **la meilleure approximation linéaire est la projection orthogonale** de x sur cet espace (Sec. 3.1), ce qui donne naturellement le développement

$$P_{V_T} x(u) = \sum_{n \in \mathbb{Z}} \frac{1}{T} \langle x, \phi_{T,n} \rangle\ \phi_T(u - nT) \tag{145}$$

Si $x \in V_T$ alors x est égal à sa projection orthogonale, ce qui est l'essence du théorème d'échantillonnage. Notons que si l'on rapproche les deux développements sur la base de $\{\phi_{T,n}\}$ (Eq. 139), on se rend compte que [44]

$$\langle x, \phi_{T,n} \rangle = T\ x(nT) \tag{146}$$

Ainsi, on a mis en avant une autre façon d'exprimer **le théorème d'échantillonnage de Shannon qui est un cas très particulier d'approximation linéaire** où le filtre est celui d'un passe-bas idéal. Ce que l'on va voir, c'est que l'on peut **généraliser ce résultat en prenant d'autres types de filtres**. Et la raison pour laquelle on a envie de changer de filtre, c'est que celui de Shannon est discontinu en Fourier, donc ne décroit qu'en $1/u$ dans l'espace réel, ce qui ne permet pas du tout une bonne localisation des irrégularités du signal $x(u)$ que l'on recherche.

Donc, on va généraliser l'espace V_T de telle sorte que l'on va définir des espaces emboités. L'analogie se voit bien sur une image, où la capture des structures essentielles se fait à travers l'utilisation de versions de l'image initiale prises à différentes résolutions. Donc, comment modifier la résolution de l'image tout en gardant ces informations pertinentes. Ce schéma est au cœur des **Analyses Multirésolutions** d'une part et d'autre part se retrouve dans les techniques de résolutions d'équations différentielles (usage des **Multigrilles**). En effet, afin de trouver la solution d'ED, il est parfois plus efficace de

44. où $T = \int \phi_T^2(u) du$.

commencer par obtenir une approximation en basse résolution et de progressivement la raffiner par les détails à hautes fréquences. Or, utiliser des échelles variables nous amène à définir des échantillonnage à pas variables. Il nous faudra poser un cadre mathématique pour comprendre ces structurations multiéchelles. Dans ce cadre, les Ondelettes vont trouver naturellement leur place, car les coefficients d'ondelettes sont les fameux détails de l'image ou des solutions d'une ED. De plus, si l'image est régulière par morceaux, alors **la représentation est parcimonieuse** car les plus grands coefficients sont localisés au niveau des contours. Dans ce schéma, nous verrons également que **des algorithmes en cascade de filtres** arrivent naturellement. Lesquels sont au cœur des réseaux de neurones profonds.

7. Séance du 17 Févr.

Durant cette séance, nous allons construire les bases d'Ondelettes à partir des Analyses Multirésolutions. Ce sujet cher à S. Mallat a beaucoup d'applications en mathématiques et va nous servir pour modéliser les données $x(u)$ afin de comprendre les phénomènes de concentration dans les problèmes $f(x) = y$ qui sont au cœur des réseaux de neurones. Donc, pour mémoire, la Transformée en Ondelettes permet de faire une analyse de la régularité d'une fonction dans un plan "espace-échelle" ou "temps-échelle" via ce qu'on appelle le *scalogramme*. Les coefficients d'Ondelettes permettent de caractériser *la régularité locale* du signal x en n'importe quel point v. Les grands coefficients se trouvent au voisinage des singularités.

Cependant, on a vu que l'on peut contracter la représentation pour la rendre parcimonieuse en échantillonnant tout d'abord l'espace des échelles ($s = 2^j$) pour ne retenir que les coefficients $W(u, 2^j)$, car si l'on dispose d'une ondelette dont les versions dilatées/contractées couvrent bien l'espace de Fourier (condition de Littlewood-Paley du Th. 13) alors on peut retrouver le signal. Nous avions également trouvé par un argument intuitif que l'on peut vraisemblablement opérer un échantillonnage de l'espace réel ($v_n = n2^j$). C'est-à-dire qu'on se demande si la famille des fonctions

$$\left\{ \psi_{j,n}(u) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{u - 2^j n}{2^j} \right) \right\}_{(j,n)\in\mathbb{Z}^2} \tag{147}$$

est une **base orthonormale** de l'espace L^2 des signaux $x(u)$. Si une/de telle(s) bases(s)

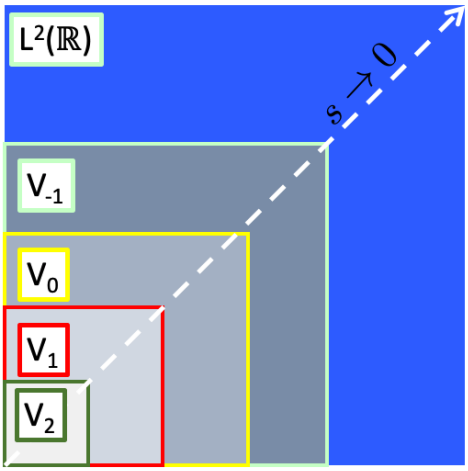

FIGURE 38 – Famille des espaces linéaires V_i emboités.

existe(nt) alors on peut appliquer le schéma précédemment utilisé d'éliminer les petits coefficients d'Ondelettes afin d'obtenir une *approximation non-linéaire de basse dimension*.

Donc le problème auquel on s'attaque est de construire de telles bases orthonormales en utilisant les multi-résolutions. Les idées sont venues en particulier du domaine de la vision par ordinateur en partie à cause de la taille mémoire limitée dans les années 80. En effet, la question était de savoir s'il y avait possibilité de réduire au maximum la taille de l'image sans perdre l'information essentielle contenue dans l'image d'origine, et de procéder à une agrégation progressive des détails afin de raffiner la perception de l'objet d'étude. Donc dans un premier temps, nous allons définir la projection du signal sur des grilles aux résolutions multiples, et ensuite par complément on ajoutera les détails. Ce faisant nous allons découvrir les bases orthonormales en chemin.

7.1 Multirésolutions

7.1.1 La définition

Il nous faut définir les projections aux différentes échelles qui sont des espaces linéaires emboités (Fig. 38) [45].

Définition 4 *(**Multirésolution**)*
Soit la famille d'espaces linéaires $\{V_j\}_{j\in\mathbb{Z}}$ indexés par j relié à l'échelle 2^j, elle est appelée une multirésolution si elle satisfait les propriétés suivantes:

i) on perd de la résolution en passant de l'échelle 2^j à 2^{j+1}, c'est-à-dire

$$V_{j+1} \subset V_j \tag{148}$$

ii) on relie V_j et V_{j+1} selon l'équivalence

$$x(u) \in V_j \Leftrightarrow x(u/2) \in V_{j+1} \tag{149}$$

on dilate $x(u)$ dans un sens et on le contracte dans l'autre.

iii) Si par ailleurs, on approxime $x(u)$ par sa projection orthogonale sur V_j, on aimerait que lorsque la résolution est infinie ($s \to 0$, ou $j \to -\infty$) on puisse reconstruire complètement le signal, c'est-à-dire en d'autres termes

$$\lim_{j\to-\infty} \|x - P_{V_j}x\| = 0, \quad \forall x \in L^2(\mathbb{R}) \tag{150}$$

iv) A l'inverse [a], si $j \to +\infty$ ou l'échelle devient infinie, alors on perd en résolution et on ne peut reconstruire le signal

$$\lim_{j\to+\infty} P_{V_j}x = 0 \tag{151}$$

v) Finalement, il nous faut rajouter une dernière propriété qui tient au fait que

45. Pour les lecteurs des notes 2018: pour la section 6.6 il faut faire attention à la façon dont les indices j sont arrangés car ils sont à l'opposé de la définition utilisée ici. C'est une petite gymnastique à laquelle il faut prêter gare dans les publications également.

l'on construit des approximations sur des grilles, et lorsque l'on translate $x \in V_j$ on ne change pas la résolution, il y a une invariance par translation. Donc,

$$\exists \phi \in V_0 \ / \ \{\phi(x-n)\}_{n\in\mathbb{Z}} \text{ soit une base orthonormale de } V_0 \tag{152}$$

a. On peut écrire ces deux propriétés selon: $\cup_{j\in\mathbb{Z}} V_j = L^2(\mathbb{R})$ et $\cap_{j\in\mathbb{Z}} V_j = 0$.

La propriété (v) peut être étendue à l'ensemble des espaces V_j par le lemme suivant:

Lemme 1 *$\forall j \in \mathbb{Z}$, alors*

$$\left\{\phi_{j,n}(u) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{u-2^j n}{2^j}\right)\right\}_{n\in\mathbb{Z}} \tag{153}$$

est une base orthonormale de V_j.

Démonstration 1. Dans un premier temps on peut faire remarquer (prop. (ii) et récursion) que $x(u) \in V_0 \Leftrightarrow x(u/2^j) \in V_j$, donc les $\phi_{j,n}$ sont bien des fonctions de V_j. Inversement $x(u) \in V_j \Leftrightarrow x(2^j u) \in V_0$, donc on peut le mettre sous la forme

$$x(2^j u) = \sum_{n\in\mathbb{Z}} \alpha_n \phi(u-n)$$

et naturellement

$$x(u) = \sum_{n\in\mathbb{Z}} \alpha_n \phi(2^{-j}u - n)$$

Enfin l'orthonormalité des $\phi(2^{-j}u-n)$ s'obtient facilement en se transportant dans V_0 selon

$$\begin{aligned}\langle \phi_{j,m}, \phi_{j,n}\rangle &= \int \phi_{j,m}(u)\phi^*_{j,n}(u)du = \int \phi(2^{-j}u-m)\phi^*(2^{-j}u-n)\ 2^{-j}du \\ &= \int \phi(v-m)\phi^*(v-n)dv = \delta[m-n]\end{aligned}$$

■

7.1.2 Quelques exemples de multirésolutions

Pour définir la fonction ϕ qui est une pierre angulaire de la définition d'une multirésolution, nous allons revenir au théorème d'échantillonnage. Soit donc

$$\phi(u) = \mathbf{1}_{[0,1[}(u) \tag{154}$$

Nous verrons que ϕ est reliée à l'ondelette de Haar. Il est assez clair que la famille $\{\phi(u-n)\}_{n\in\mathbb{Z}}$ est une base orthonormale de l'espace linéaire V_0 des fonctions constantes par morceaux à pas de 1. De même $\phi_j(u) = \phi(u/2^j)$ est la fonction $\mathbf{1}_{[0,2^j[}(u)$, et la famille associée est une base orthonormale de l'espace V_j des fonctions constantes par morceaux à pas de 2^j, c'est-à-dire constante sur les intervalles $[2^j n, 2^j(n+1)[$. Dans ce cadre, construire une approximation de x à l'échelle 2^j c'est donc obtenir l'approximation par une fonction en escalier sur des intervalle de taille 2^j. Un exemple est donné sur la figure 39(gauche).

Un autre exemple nous est fourni cette fois en définissant $\phi(u)$ à travers son spectre de Fourier comme pour le théorème de Shannon 15:

$$\hat{\phi}(\omega) = \mathbf{1}_{[-\pi,\pi]}(\omega) \tag{155}$$

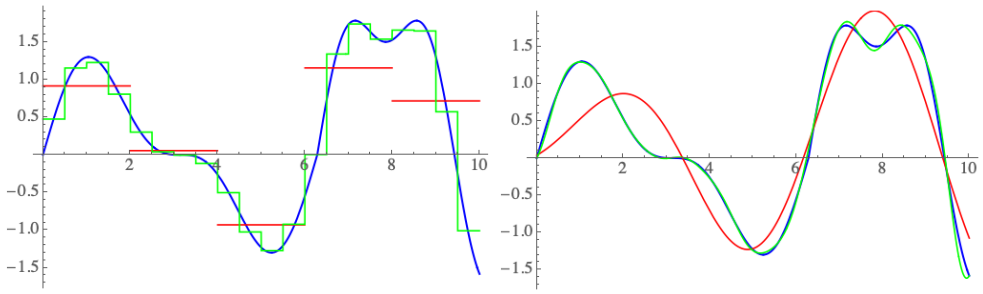

FIGURE 39 – Exemples d'analyse multirésolution avec $\phi(u) = \mathbf{1}_{[0,1[}(u)$ (gauche) et $\phi(u) = \text{sinc}(\pi u)$ (droite): en rouge $j = 1$, en vert $j = -1$ et en bleu la fonction $x(u)$.

Dans l'espace réel, on trouve que $\phi(u) = \text{sinc}(\pi u)$, et la décomposition de $x(u)$ correspond à

$$x(u) = \sum_n \alpha_n \phi(u - n) \tag{156}$$

L'on sait que l'on a $\alpha_n = x(n)$ d'après Shannon et maintenant on peut faire agir le paramètre d'échelle pour obtenir des approximations à différentes résolutions. En Fourier, $\hat{\phi}_j(\omega) = \mathbf{1}_{[-2^{-j}\pi, 2^{-j}\pi]}(\omega)$, donc quand $j \to -\infty$ on couvre de plus en plus les hautes fréquences, et l'on obtient des approximations de meilleure qualité. Un exemple est donné sur la figure 39(droite).

Les deux exemples précédents sont des cas extrêmes: dans le premier la fonction ϕ est discontinue en espace, dans le second c'est son spectre de Fourier qui est discontinu. On aimerait pourtant avoir plus de régularité dans les deux espaces. Or, déjà à première vue on peut se dire que l'approximation constante par morceaux est très grossière, et on pourrait opter pour une approximation au moins linéaire par morceaux, voire polynomiale par morceaux. Donc, on se dit que certainement on peut obtenir des multirésolutions de meilleure qualité. Ce que l'on va voir, c'est qu'à chaque **multirésolution**, on obtient une **base d'ondelettes** dont la construction est uniquement basée sur celle de **filtres discrets** et en particulier (ou avant tout) celui associé à la fonction ϕ. Et finalement, on trouvera des **algorithmes de filtrage-sous-échantillonnage** que l'on voit également apparaître dans les réseaux de neurones.

7.2 Bancs de filtres

La construction de filtres et plus précisément de *bancs de filtres*, à savoir des réseaux de filtres passe-bandes, est un sujet traditionnel en traitement du signal afin de séparer le signal d'entrée en plusieurs composants. D'ailleurs, la théorie des ondelettes est venue en quelque sorte assoir un savoir empirique sur la construction des bancs de filtres.

Dans un premier temps posons nous la question: comment construit-on une approximation P_{V_j}? C'est une projection orthogonale dans la base de V_j, il vient alors

$$P_{V_j} x = \sum_{n \in \mathbb{Z}} \langle x, \phi_{j,n} \rangle \phi_{j,n} \tag{157}$$

Or, le produit scalaire de x avec $\phi_{j,n}$ peut être vu comme une convolution. Introduisons

la fonction

$$\tilde{\phi}_j(u) = \frac{1}{\sqrt{2^j}}\phi\left(-\frac{u}{2^j}\right) = \phi_j(-u) \tag{158}$$

Ainsi, on peut interpréter la projection sur V_j comme d'abord un filtrage basse fréquence avec $\tilde{\phi}_j$ suivi d'une interpolation avec $\phi_{j,n}$:

$$\boxed{P_{V_j}x = \overbrace{\sum_{n\in\mathbb{Z}} \underbrace{(x * \tilde{\phi}_j)(2^j n)}_{\text{filtrage}}\ \phi_{j,n}}^{\text{interpolation}}} \tag{159}$$

Finalement, on peut généraliser immédiatement le théorème de Shannon [46]:

Théorème 16 *(échantillonnage multirésolution)*
Si $x \in V_j$ d'une multirésolution alors la projection orthogonale $P_{V_j}x = x$ (et vice-versa), et donc

$$x(u) = \sum_{n\in\mathbb{Z}} (x * \tilde{\phi}_j)(2^j n)\ \phi_j(u - 2^j n) \tag{160}$$

Le théorème de Shannon devient un cas particulier dans la cas où $\hat{\phi}(\omega)$ est le filtre passe-bas idéal sur $[-\pi, \pi]$. Cependant, ce qui va nous intéresser c'est le passage entre deux grilles de résolutions différentes, et c'est là où les ondelettes vont intervenir. Ce faisant, ce sont les deux propriétés (i) et (ii) d'une multirésoluton sur lesquelles ont va se pencher pour faire émerger les ondelettes.

Fixons $j = 0$, d'après la propriété (i), on sait que $V_1 \subset V_0$ et $\phi(u) \in V_0$, donc d'après (ii) $\phi(u/2) \in V_1$, donc $\phi(u/2) \in V_0$. Or, comme les $\{\phi(u-n)\}_n$ sont une base de V_0 on peut décomposer $\phi(u/2)$ sur cette base [47]

$$\boxed{\frac{1}{\sqrt{2}}\phi\left(\frac{u}{2}\right) = \sum_{n\in\mathbb{Z}} h(n)\phi(u-n)} \tag{161}$$

46. NDJE: Notons qu'ici c'est la même fonction ϕ qui sert pour le filtrage et l'interpolation. On peut généraliser ce schéma en faisant jouer deux fonctions distinctes.
47. NDJE: dans la littérature, on trouve le terme de *scaling relation*.

et les $h(n)$ sont obtenus en utilisant l'orthogonalité de la base:

$$h(n) = \langle \frac{1}{\sqrt{2}} \phi \left(\frac{u}{2} \right), \phi(u-n) \rangle \tag{162}$$

Donc, si on connait ϕ, on connait les coefficients $h(n)$. **Mais le plus intéressant par certain côté est que l'on peut se donner $h(n)$ et construire ϕ.** Passons en Fourier, il vient

$$\sqrt{2}\hat{\phi}(2\omega) = \left(\sum_{n \in \mathbb{Z}} h(n) e^{-in\omega} \right) \hat{\phi}(\omega) = \hat{h}(\omega)\hat{\phi}(\omega) \tag{163}$$

où l'on reconnait la série de Fourier associée à h. Il vient alors

$$\sqrt{2}\hat{\phi}(2\omega) = \hat{h}(\omega)\hat{\phi}(\omega) \Longleftrightarrow \hat{\phi}(\omega) = \frac{\hat{h}(\omega/2)}{\sqrt{2}} \hat{\phi}(\omega/2) \tag{164}$$

On a envie de réitérer le processus, ce qui permet d'écrire

$$\hat{\phi}(\omega) = \hat{\phi}(2^{-J}\omega) \prod_{p=1}^{J} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} \tag{165}$$

Maintenant, si on fait tendre J vers l'infini, on s'aperçoit que $\hat{\phi}(\omega)$ est **entièrement déterminée par le filtre** $\hat{h}(\omega)$:

$$\boxed{\hat{\phi}(\omega) = \hat{\phi}(0) \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}}} \tag{166}$$

Quelles sont les propriétés de $\hat{h}(\omega)$? Rappelons que la famille $\{\phi(u-n)\}_{n \in \mathbb{Z}}$ est une base orthonormale et qu'elle sert à construire une multirésolution, tout ça pour dire que le filtre $\hat{h}$ doit être assez spécial quand même. En fait, on a envie de renverser le problème: quelles sont les propriétés de h qui vont bien, pour qu'ensuite ϕ donne une multirésolution. Ce faisant, on va retrouver des propriétés utilisées en théorie du signal, à savoir *les filtres miroirs*. Voici le théorème:

Théorème 17 *(filtre h)*
Si ϕ définit une multirésolution (Def. 4), et si on définit $h(n)$ selon

$$h(n) = \langle \frac{1}{\sqrt{2}} \phi\left(\frac{u}{2}\right), \phi(u-n) \rangle = \langle \phi_1, \phi_{0,n} \rangle \tag{167}$$

alors la transformée de Fourier de h satisfait les relations

$$|\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2 \qquad \text{et} \quad \hat{h}(0) = \sqrt{2} \tag{168}$$

Ces propriétés sont illustrées sur la figure 40.

Inversement, si $\hat{h}$ satisfait les relations ci-dessus et si

$$\hat{h}(\omega) > 0 \qquad \forall \omega \in [-\pi/2, \pi/2] \tag{169}$$

alors

$$\hat{\phi}(\omega) = \prod_{p=1}^{+\infty} \frac{\hat{h}(2^{-p}\omega)}{\sqrt{2}} \tag{170}$$

est la transformée de Fourier d'une fonction ϕ qui définit une multirésolution.

Démonstration 17. Moyennant [48] que $\hat{\phi}(0) \neq 0$, alors on tire facilement que $\hat{h}(0) = \sqrt{2}$ à partir de Eq. 164. La première des deux propriétés 168 est la plus cruciale et avait été mise en évidence par M.J.T Smith et T.P Barnwell au cours des années 80 dans le contexte des bancs de filtres. Rappelons nous que l'on veut aboutir à une base orthonormale avec $\phi(u)$ translatée. Ainsi, on veut

$$\langle \phi(u), \phi(u-n) \rangle = \delta[n] \tag{171}$$

ce qui par le biais de la convolution avec $\tilde{\phi}$ (Eq. 158 avec $j = 0$) donne

$$(\phi * \tilde{\phi})(n) = \delta[n] \tag{172}$$

48. NDJE: Un argument est donné plus loin.

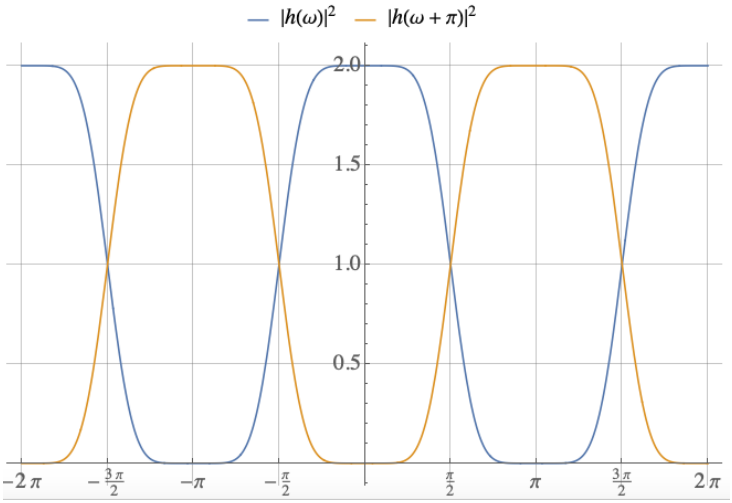

FIGURE 40 – Exemple d'un filtre $\hat{h}(\omega)$ qui satisfait les relations Eqs. 168. En l'occurrence, il s'agit d'un filtre d'une multirésolution de Daubechies d'ordre 4. *Nb. selon les librairies le filtre inclut ou non le facteur* $\sqrt{2}$.

On se sert à présent du théorème 14 avec $T = 1$, on a

$$\sum_{n\in\mathbb{Z}} x(n)e^{-in\omega} = \sum_{k\in\mathbb{Z}} \hat{x}\left(\omega - 2k\pi\right) \tag{173}$$

qui pour $x = \phi * \tilde{\phi}$ donne, dans le cas où ϕ est une fonction réelle, $\hat{x} = |\phi|^2$ et la relation suivante

$$\boxed{\sum_{k\in\mathbb{Z}} |\hat{\phi}(\omega - 2k\pi)|^2 = 1} \tag{174}$$

C'est la condition nécessaire et suffisante pour avoir une base orthonormale. Mais alors utilisons la relation 164, en remplaçant il vient

$$\sum_{k\in\mathbb{Z}} |\hat{\phi}(\omega - 2k\pi)|^2 = \frac{1}{2}\sum_{k\in\mathbb{Z}} |\hat{h}(\omega/2 - k\pi)\hat{\phi}(\omega/2 - k\pi)|^2 = 1 \tag{175}$$

Donc, en séparant les k pairs et impairs et en utilisant le fait que $\hat{h}(\omega)$ est 2π-périodique

(à savoir une série de Fourier), on peut écrire

$$\begin{aligned}
&\sum_{k\in\mathbb{Z}} |\hat{h}(\omega/2-k\pi)|^2|\hat{\phi}(\omega/2-k\pi)|^2 \\
&\quad= \sum_{p\in\mathbb{Z}} |\hat{h}(\omega/2-2p\pi)|^2|\hat{\phi}(\omega/2-2p\pi)|^2 + \sum_{p\in\mathbb{Z}} |\hat{h}(\omega/2-(2p+1)\pi)|^2|\hat{\phi}(\omega/2-(2p+1)\pi)|^2 \\
&\quad= \sum_{p\in\mathbb{Z}} |\hat{h}(\omega/2)|^2|\hat{\phi}(\omega/2-2p\pi)|^2 + \sum_{p\in\mathbb{Z}} |\hat{h}(\omega/2-\pi)|^2|\hat{\phi}(\omega/2-(2p+1)\pi)|^2 \\
&\quad= |\hat{h}(\omega/2)|^2 \underbrace{\sum_{p\in\mathbb{Z}} |\hat{\phi}(\omega/2-2p\pi)|^2} + |\hat{h}(\omega/2-\pi)|^2 \underbrace{\sum_{p\in\mathbb{Z}} |\hat{\phi}(\omega/2-(2p+1)\pi)|^2}
\end{aligned}$$

Or en utilisant la relation 174 on se convainc que les deux sommes indiquées par les accolades sont égales à 1. Donc,

$$|\hat{h}(\omega/2)|^2 + |\hat{h}(\omega/2-\pi)|^2 = 2$$

ce qui achève de démontrer la relation, car cela vaut pour tout ω et on peut utiliser la 2π-périodicité de $\hat{h}$ pour faire apparaitre le $+\pi$ dans le second terme.

Pour la réciproque, il faut se pencher sur ce qu'il se passe en $\omega = 0$. Si $\hat{\phi}(0) = 0$ alors on a un filtre passe-bande. Mais, souvenons-nous de la condition sur ψ, quand j devient de plus en plus négatif, la bande du spectre se déplace vers les hautes fréquences dépeuplant alors les basses fréquences (voir l'illustration sur la figure 30). Or, concernant ϕ, cela n'est pas possible car la projection $P_{V_j}x$ doit satisfaire la propriété (iii) de la multirésolution (Déf. 4), à savoir que $P_{V_j}x$ doit converger vers x pour tout élément de $L^2(\mathbb{R})$, ce qui est une contradiction dés lors que l'on dépeuple les basses fréquences. Ainsi,

$$\hat{\phi}(0) \neq 0$$

La démonstration complète de la réciproque est dans les notes que S. Mallat joint à son cours. Elle commence par les relations sur $\hat{h}(\omega)$ pour démontrer que le produit 170 a un sens et donne bien une fonction ϕ qui a la propriété 174 pour donner une base orthonormale. ■

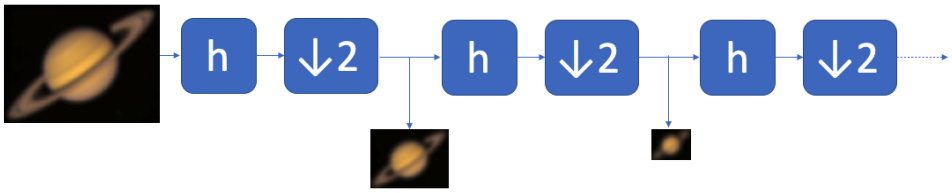

FIGURE 41 – Algorithme de Burt: cascade d'un filtre $\hat{h}(\omega)$ passe-bas pour éviter l'aliasing suivi d'un sous-échantillonnage d'un facteur 2.

7.3 Algorithmes en bancs de filtres (I)

Une fois que l'on a une fonction $\hat{h}(\omega)$ qui satisfait les relations 174, ou bien la version dans l'espace réel à travers les $h(n)$, c'est-à-dire

$$\hat{h}(\omega) = \sum_{n\in\mathbb{Z}} h(n)e^{-in\omega} \tag{176}$$

on aimerait voir comment cela va nous servir en pratique. En fait des algorithmes existaient en traitement d'images, notamment celui de Peter Burt [49]. Il s'agissait de cascader un filtre passe-bas h pour éviter l'aliasing (Sec. 6.4) suivi d'un sous-échantillonnage d'un facteur 2 (Fig. 41). Le problème qui se posait à l'époque, était de réaliser ces opérations en temps réel ce qui nécessitait de prendre des filtres de petite taille pour effectuer les convolutions rapidement. Les filtres en question ne pouvaient satisfaire les propriétés ci-dessus mentionnées, cependant la cascade d'opérations était en place.

Maintenant, la projection orthogonale de x dans V_j se fait selon le théorème 16

$$x(u) = \sum_{n\in\mathbb{Z}} a_j[n]\phi_{j,n}(u) \tag{177}$$

On passe donc de la variable continue u à la variable n discrète, c'est-à-dire de $x(u)$ à $a_j[n]$

$$a_j[n] = \langle x, \phi_{j,n}\rangle = (x * \tilde{\phi})(2^j n) \tag{178}$$

et on ne manipule plus que **des séquences discrètes**. Quel est le lien alors entre $a_j[n]$ et

49. Burt-Adelson pyramid image processing.

$a_{j+1}[n]$? La relation se sent bien dès lors que l'on a la structure emboitée des espaces V_j: connaissant la projection de x dans V_j, on peut la projeter dans l'espace V_{j+1} et ainsi de suite. Pour obtenir la projection dans V_{j+1}, il nous faut connaitre les produits scalaires avec les $\phi_{j+1,n}$, or

$$\phi_{j+1}(u) = \sum_n \langle \phi_{j+1}, \phi_{j,n} \rangle \phi_{j,n}(u) \tag{179}$$

avec

$$\begin{aligned} \langle \phi_{j+1}, \phi_{j,n} \rangle &= \int \frac{1}{\sqrt{2^{j+1}}} \phi\left(\frac{u}{2^{j+1}}\right) \frac{1}{\sqrt{2^j}} \phi\left(\frac{u - 2^j n}{2^j}\right) \, du \\ &= \frac{1}{\sqrt{2}} \int \phi\left(\frac{u}{2}\right) \phi(u - n) \, du \\ &= h(n) \end{aligned} \tag{180}$$

où la dernière égalité est obtenue à l'aide du théorème 17. Ainsi, on obtient la relation

$$\boxed{\phi_{j+1}(u) = \sum_{n \in \mathbb{Z}} h(n) \phi_{j,n}(u)} \tag{181}$$

Que vaut à présent $\phi_{j+1,n}(u)$? Il vient

$$\begin{aligned} \phi_{j+1,p}(u) = \phi_{j+1}(u - 2^{j+1}p) &= \sum_{n \in \mathbb{Z}} h(n) \phi_{j,n}(u - 2^{j+1}p) \\ &= \sum_{n \in \mathbb{Z}} h(n) \frac{1}{\sqrt{2^j}} \phi\left(\frac{u - 2^j(2p) - 2^j n}{2^j}\right) \end{aligned} \tag{182}$$

et donc on obtient la relation

$$\boxed{\phi_{j+1,p}(u) = \sum_{n \in \mathbb{Z}} h(n) \phi_{j,n+2p}(u)} \tag{183}$$

Nous pouvons à présent calculer $a_{j+1}[p]$:

$$a_{j+1}[p] = \langle x, \phi_{j+1,p} \rangle = \sum_{n \in \mathbb{Z}} h(n) \langle x, \phi_{j,n+2p} \rangle \tag{184}$$

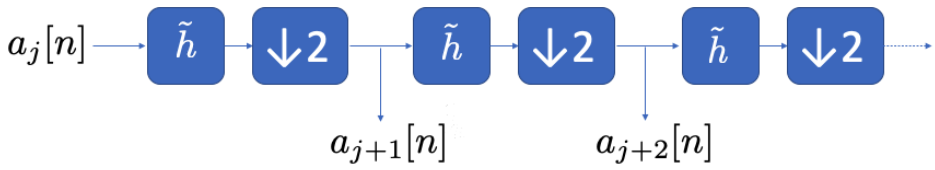

FIGURE 42 – Algorithme du passage des $a_j[n]$ aux $a_{j+1}[n]$: cascade d'un filtre discret $\tilde{h}$ passe-bas suivi d'un sous-échantillonnage d'un facteur 2. On peut noter les similitudes et différences par rapport à l'algorithme de Burt de la figure 41.

et ainsi obtenir

$$\boxed{a_{j+1}[p] = \sum_{n\in\mathbb{Z}} h(n)a_j[n+2p] = \sum_{n\in\mathbb{Z}} h(n-2p)a_j[n] = (a_j * \tilde{h})(2p)} \tag{185}$$

où on a mis en évidence la convolution avec le filtre $\tilde{h}[n] = h[-n]$ (le pendant de $\tilde{\phi}$) pris en $2p$, c'est-à-dire que l'on a bien **une cascade de filtrage-sous-échantillonnage** (Fig. 42).

7.3.1 Exemple avec la multirésolution de Haar

Reprenons la fonction ϕ indicatrice de $[0,1]$ qui constituait notre premier exemple de multirésolution (Sec. 7.1.2). Les fonctions translatées $\phi(u-n)$ sont donc des indicatrices

$$\phi(u-n) = \mathbf{1}_{[n,n+1]}(u) \tag{186}$$

de même la fonction ϕ dilatée d'un facteur 2 est aussi une indicatrice

$$\frac{1}{\sqrt{2}}\phi\left(\frac{u}{2}\right) = \frac{1}{\sqrt{2}}\mathbf{1}_{[0,2]}(u) \tag{187}$$

Ainsi via le théorème 17, on définit le filtre discret h comme produit scalaire entre $\phi(u-n)$ et $\frac{1}{\sqrt{2}}\phi\left(\frac{u}{2}\right)$, ce qui donne

$$h[n] = \begin{cases} 1/\sqrt{2} & n = 0,1 \\ 0 & \text{ailleurs} \end{cases} \tag{188}$$

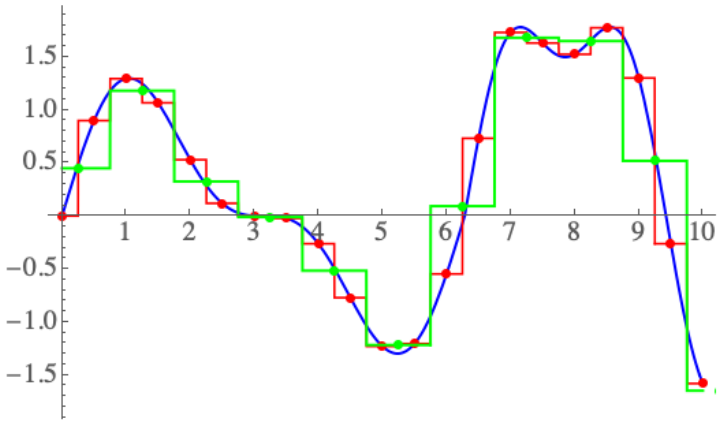

FIGURE 43 – Illustration de l'algorithme d'échantillonnage avec la fonction $\phi(u)$ de Haar: premièrement on échantillonne le signal pour obtenir une approximation (rouge) sur une grille V_0, puis dans un second temps par filtrage-sous-échantillonnage on obtient une approximation sur une grille 2 fois plus large V_1.

Maintenant, d'après le théorème 16, soit le signal $x(u)$ avec u une variable continue. Ce signal en pratique est échantillonné sur une grille de pas fixe notée V_0 et on dispose alors des coefficients $a_0[n] = x(n)$. Ensuite, l'algorithme donne les coefficients $a_1[p]$

$$a_1[p] = \frac{a_0[2p] + a_0[2p+1]}{\sqrt{2}} \tag{189}$$

et la projection sur V_1 du signal devient

$$P_{V_1}x(u) = \sum_p a_1[p]\phi_1(u-p) = \sum_p \frac{a_0[2p] + a_0[2p+1]}{2}\phi\left(\frac{u}{2} - p\right) \tag{190}$$

Notons que l'on procède *in fine* à la moyenne 2-à-2 des coefficients pour passer d'une grille à l'autre. Une illustration de cet algorithme est montrée sur la figure 43. Ce qui va changer par la suite ce sont d'une part le support de h et ses valeurs, mais l'algorithme restera identique.

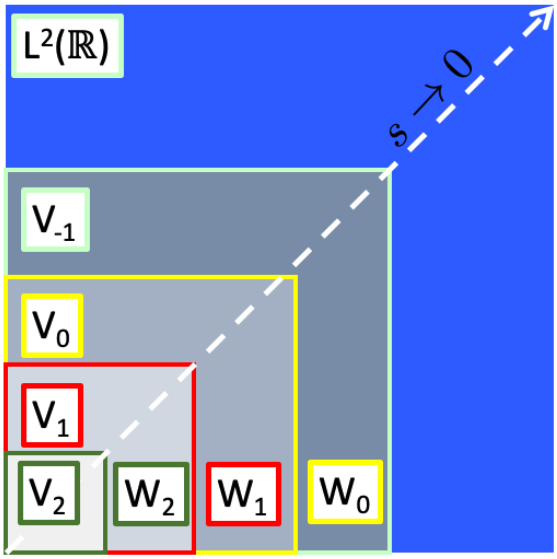

FIGURE 44 – Espaces V_j et W_j reliés par la relation de complémentarité Eq. 191. Ceci vient compléter la figure 38.

7.4 Lien avec les bases d'Ondelettes

L'algorithme en bancs de filtres étudié à la section précédente donne des *moyennages successifs*, mais ce n'est pas de la sorte que l'on peut construire des représentations parcimonieuses où l'on recherche plutôt à obtenir des 0. Remarquons que lors du passage de V_j à V_{j+1}, en réduisant l'information on a perdu par moyennage les détails du signal présents dans V_j. Si on veut mettre en évidence ces détails, il suffit non pas de projeter dans V_{j+1} mais dans l'espace complémentaire [50] W_{j+1} (Fig. 44):

$$V_j = W_{j+1} \oplus V_{j+1} \tag{191}$$

On connait une base orthonormale dans V_j et V_{j+1}, il nous faut donc construire une base orthogonale dans W_{j+1}. Cependant, remarquons que la relation précédente appelle à une

50. nb. petit rappel pour les lecteurs du Cours de 2018, les échelles sont indexées par $-j$, c'est-à-dire $s = 2^{-j}$.

récurrence

$$V_j = \bigoplus_{p=j+1}^{J} W_p \oplus V_J \tag{192}$$

Or si $J \to +\infty$, V_J tend vers l'ensemble vide, tandis que si $j \to -\infty$ alors V_j tend vers tout l'espace $L^2(\mathbb{R})$, donc d'une certaine mesure

$$L^2(\mathbb{R}) = \bigoplus_{j=-\infty}^{+\infty} W_j \tag{193}$$

et tous les espaces W_j sont orthogonaux entres-eux. Ainsi obtenir une base orthogonale de $L^2(\mathbb{R})$ se fait en groupant les bases orthogonales de tous les espaces W_j. C'est là où l'ondelette ψ va jouer son rôle pour connecter les ensembles W_j entres-eux.

Théorème 18 *(filtre g)*
Soit $g(n)$ un filtre obtenu à partir du filtre h (Th. 17) selon

$$g(n) = (-1)^{1-n} h(1-n) \tag{194}$$

Soit également la fonction $\psi(u)$ obtenue à partir de ce filtre g et de la fonction ϕ (laquelle est aussi obtenue à partir de h)

$$\frac{1}{\sqrt{2}}\psi\left(\frac{u}{2}\right) = \sum_{n\in\mathbb{Z}} g(n)\phi(u-n) \tag{195}$$

Alors $\forall j$, la famille

$$\left\{\psi_{j,n}(u) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{u-2^j n}{2^j}\right)\right\}_{n\in\mathbb{Z}} \tag{196}$$

est une base orthonormale de W_j. Et la famille $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ est une base orthonormale de $L^2(\mathbb{R})$.

On a ainsi une généralisation de Haar et cela à partir de la donnée du filtre h. Il est remarquable qu'il soit possible de construire des bases d'ondelettes qui sortent du cadre de ce théorème, mais les ondelettes sont pathologiques et à décroissance très très lente qui ne les rendent pas du tout attractives et utilisables en pratique. Donc, on peut en pratique voir ce théorème comme une condition nécessaire et suffisante pour obtenir des bases d'ondelettes orthonormales.

Démonstration 18. Dans un premier temps, si $\psi(u) \in W_1 \subset V_0$ le complémentaire orthogonal de V_1 dans V_0, et comme V_0 a comme base orthonormale $\{\phi_n\}_{n\in\mathbb{Z}}$ alors on peut décomposer $\psi(u/2)$ selon

$$\frac{1}{\sqrt{2}}\psi\left(\frac{u}{2}\right) = \sum_{n\in\mathbb{Z}} g(n)\phi(u-n) \tag{197}$$

avec $g(n)$ inconnu pour le moment mais on sait que

$$g(n) = \langle \frac{1}{\sqrt{2}}\psi\left(\frac{u}{2}\right), \phi(u-n)\rangle \tag{198}$$

On veut également une base orthonormale [51] de W_0 à partir des $\{\psi_n\}_{n\in\mathbb{Z}}$ c'est-à-dire

$$\langle \psi(u), \psi(u-n)\rangle = \langle \psi, \psi_n\rangle = \delta[n] \Leftrightarrow (\psi * \tilde{\psi})(n) = \delta[n] \tag{199}$$

donc d'après le même raisonnement que pour ϕ (Eq. 174), on obtient la relation

$$\boxed{\sum_{k\in\mathbb{Z}} |\hat{\psi}(\omega - 2k\pi)|^2 = 1} \tag{200}$$

Mais on veut aussi que W_0, le complémentaire de V_0 dans V_{-1}, soit orthogonale à V_0, donc la famille $\{\psi_n\}_{n\in\mathbb{Z}}$ doit être orthogonale à la famille $\{\phi_n\}_{n\in\mathbb{Z}}$. Cette dernière contrainte, dans le domaine de Fourier, donne la relation suivante:

$$\boxed{\sum_{k\in\mathbb{Z}} \hat{\phi}^*(\omega - 2k\pi)\hat{\psi}(\omega - 2k\pi) = 0} \tag{201}$$

NDJE: Une démonstration possible de cette relation est la suivante. On veut que $\forall n \in \mathbb{Z}$

$$\langle \psi(u), \phi(u-n)\rangle = 0 = (\psi * \tilde{\phi})(n)$$

Si on pose $x = \psi * \tilde{\phi}$ *sa transformée de Fourier est* $\hat{x} = \hat{\phi}^*\hat{\psi}$ *(ϕ est réelle). Donc d'après le théorème 14, on obtient directement la formule ci-dessus.*

Maintenant, les deux relations 200 et 201 sont équivalentes aux deux relations sui-

51. nb. on aurait pu écrire que les $\{\psi_{1,n}\}_{n\in\mathbb{Z}}$ forment une base orthonormale de W_1.

vantes:

$$\boxed{\begin{aligned} |\hat{g}(\omega)|^2 + |\hat{g}(\omega+\pi)|^2 &= 2 \quad (202)\\ \hat{g}(\omega)\hat{h}^*(\omega) + \hat{g}(\omega+\pi)\hat{h}^*(\omega+\pi) &= 0 \quad (203)\end{aligned}}$$

NDJE: Une démonstration possible est la suivante. On procède de la même façon que la démonstration du théorème 17 après avoir fait remarquer que la transformée de Fourier de la relation qui définit ψ s'écrit

$$\sqrt{2}\hat{\psi}(2\omega) = \hat{g}(\omega)\hat{\phi}(\omega)$$

avec $\hat{g}(\omega)$ la fonction 2π-périodique correspondant à la série de Fourier avec les coefficients $g(n)$

$$\hat{g}(\omega) = \sum_{n\in\mathbb{Z}} g(n)e^{-in\omega}$$

Donc, ensuite à partir de $\sum_k |\hat{\psi}(\omega - 2k\pi)|^2 = 1$, il vient

$$\sum_{k\in\mathbb{Z}} |\hat{g}(\omega/2 - k\pi)|^2 |\hat{\phi}((\omega/2 - k\pi))|^2 = 2$$

Et en utilisant l'astuce de scinder la somme sur k, en une somme sur les k-pairs et une somme sur les k-impairs, en se servant du fait que $\hat{g}(\omega)$ est 2π-périodique, on obtient bien

$$|\hat{g}(\omega/2)|^2 + |\hat{g}(\omega/2+\pi)|^2 = 2$$

pour tout ω. Pour la seconde relation, partant de la relation 201, il vient

$$\sum_{k\in\mathbb{Z}} \hat{h}^*(\omega/2 - k\pi)\hat{g}(\omega/2 - k\pi)|\hat{\phi}((\omega/2 - k\pi))|^2 = 0$$

On peut poser $\hat{x}(\omega) = \hat{h}^*(\omega)\hat{g}(\omega)$, c'est une fonction 2π-périodique et par la même astuce de séparation de la somme sur k, on aboutit à $\forall\omega$

$$\hat{g}(\omega/2)\hat{h}^*(\omega/2) + \hat{g}(\omega/2+\pi)\hat{h}^*(\omega/2+\pi) = 0$$

ce qui donne la relation escomptée.

Enfin, une solution pour $\hat{g}(\omega)$ est de choisir

$$\boxed{\hat{g}(\omega) = e^{-i\omega}\hat{h}^*(\omega+\pi)} \quad (204)$$

NDJE: Voici une possible argumentation. En prenant donc cette relation, il vient alors

$$|\hat{g}(\omega)|^2 + |\hat{g}(\omega+\pi)|^2 = |\hat{h}(\omega+\pi)|^2 + |\hat{h}(\omega+2\pi)|^2$$

Or en utilisant le fait que $\hat{h}(\omega)$ est 2π-périodique et satisfait la propriété du théorème 17 alors il vient bien

$$|\hat{g}(\omega)|^2 + |\hat{g}(\omega+\pi)|^2 = 2$$

Quant à la seconde relation (Eq. 203), évaluons le membre de gauche en utilisant la 2π-périodicité de $\hat{h}(\omega)$

$$\begin{aligned}\hat{g}(\omega)\hat{h}^*(\omega) + \hat{g}(\omega+\pi)\hat{h}^*(\omega+\pi) = \\ e^{-i\omega}\hat{h}^*(\omega+\pi)\hat{h}^*(\omega) + e^{-i(\omega+\pi)}\hat{h}^*(\omega+2\pi)\hat{h}^*(\omega+\pi) \\ = e^{-i\omega}(\hat{h}^*(\omega+\pi)\hat{h}^*(\omega) - \hat{h}^*(\omega)\hat{h}^*(\omega+\pi)) = 0\end{aligned}$$

Enfin, ayant l'expression de $\hat{g}(\omega)$ en fonction de $\hat{h}(\omega)$, on peut l'écrire [52]

$$\begin{aligned}\hat{g}(\omega) &= e^{-i\omega}\sum_{n\in\mathbb{Z}} h^*(n)e^{i(\omega+\pi)n} = \sum_{n\in\mathbb{Z}}(-1)^n h^*(n)e^{-i\omega(1-n)} \\ &= \sum_{m\in\mathbb{Z}}(-1)^{1-m}h^*(1-m)e^{-i\omega m} \qquad (205)\end{aligned}$$

d'où l'identification des $g(n)$ donne bien

$$g(n) = (-1)^{1-n}h^*(1-n) \qquad (206)$$

et le fait que $\phi(u)$ soit réel induit que $h(n)$ le soit également, ce qui finit la démonstration. ■

52. NDJE: dans le Cours de 2018, on peut trouver une autre relation car il y a un degré de latitude qui est une phase.

8. Séance du 3 Mars

Nous avons vu à la dernière séance comment à partir d'une fonction $\phi(x)$ ou de son filtre $h[n]$ (passe-bas), on peut construire des approximations en basse dimension à des échelles successives du signal $x(u)$ en les projetant sur les espaces V_j:

$$P_{V_j}x(u) = \sum_{n\in\mathbb{Z}} a_j[n]\phi_{j,n}(u) \tag{207}$$

avec $a_j[n] = \langle x, \phi_{j,n}\rangle$. De même, afin d'introduire des détails qui sont perdus lors du changement d'échelle par l'approximation basse fréquence, on peut affiner en utilisant la complémentarité entre les espaces V_j et W_j telle que (Fig. 44)

$$V_{j-1} = V_j \oplus W_j \tag{208}$$

et ainsi

$$P_{V_{j-1}}x = P_{V_j}x \oplus P_{W_j}x \tag{209}$$

On a pu alors introduire une ondelette ψ et son filtre passe-bande $g[n]$ associés. Nous avons également obtenu les relations qu'entretiennent les filtres h et g (Th. 17).

8.1 Quelques exemples de bases orthonormales

On avait vu quelques exemples d'ondelettes dans les sections 6.3.3 et 7.1.2, qu'en est-il des filtres associés?

Pour l'ondelette de **Haar** cela a été fait dans la section 7.3.1 avec une application numérique, petit rappel [53]

$$h^{Haar}[n] = \begin{cases} 1/\sqrt{2} & n = 0, 1 \\ 0 & \text{ailleurs} \end{cases} \tag{210}$$

53. *attention à la normalisation* $\hat{h}(0) = \sqrt{2} = \sum_{n\in\mathbb{Z}} h[n]$ *qui peut surprendre. D'autre part selon les librairies les conventions peuvent changer.*

ce qui donne pour le filtre de $\phi(u) = \mathbf{1}_{[0,1]}(u)$:

$$g^{Haar}[n] = \begin{cases} g[0] = -h[1],\ g[1] = h[0] \\ 0 \text{ ailleurs} \end{cases} \tag{211}$$

Ainsi, si $h[n]$ correspond à une moyenne d'échantillons 2-à-2, $g[n]$ correspond quant à lui à leur différence. De plus

$$\psi(u) = \sqrt{2} \sum_{n \in \mathbb{Z}} g[n] \phi(2u - n) \tag{212}$$

donc

$$\begin{aligned} \psi^{Haar}(u) &= -\mathbf{1}_{[0,1]}(2u) + \mathbf{1}_{[0,1]}(2u - 1) \\ &= -\mathbf{1}_{[0,1/2]}(u) + \mathbf{1}_{[1/2,1]}(u) \end{aligned} \tag{213}$$

ce qui correspond au signe prés à la forme de la figure 32.

Comme second exemple, on peut prendre la base de **Shannon**. Le spectre de Fourier de ψ est un passe-bande idéal (Fig. 33). On peut l'obtenir dans le cadre des multirésolutions, en définissant la fonction 2π-périodique passe-bas idéal

$$\hat{h}^{Shan}(\omega) = \mathbf{1}_{[-\pi/2,+\pi/2]}(\omega) \tag{214}$$

à partir de laquelle on définit la fonction 2π-périodique $\hat{g}(\omega) = e^{-i\omega} h^*(\omega + \pi)$ qui devient un passe-bande idéal:

$$\hat{g}^{Shan}(\omega) = e^{-i\omega} \left\{ \mathbf{1}_{[-\pi,-\pi/2]}(\omega) + \mathbf{1}_{[\pi/2,\pi]}(\omega) \right\} \tag{215}$$

Ainsi, le filtre $\hat{h}^{Shan}$ sélectionne les basses fréquences entre $-\pi/2$ et $\pi/2$ tandis que le filtre $\hat{g}^{Shan}$ les exclut.

Maintenant, on se pose la question de savoir quelles sont les ondelettes, ou les filtres, ou encore les multirésolutions qui sont intéressantes, car nous avons vu les problèmes *a priori* de celles de Haar et Shannon qui sont discontinues, soit dans l'espace réel soit dans l'espace de Fourier. Cette question s'est posée à la fois du coté de la communauté du Traitement du Signal et du coté des mathématiques, avec en particulier le travail de la mathématicienne américaine d'origine belge **Ingrid Daubechies** qui a étudié comment

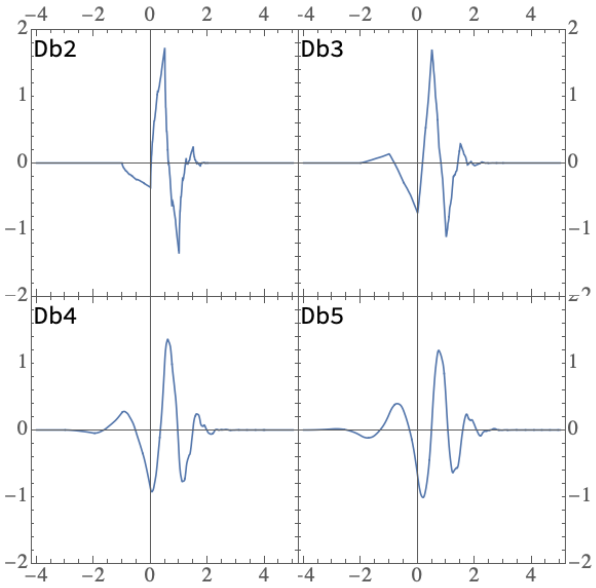

FIGURE 45 – Illustrations d'ondelettes d'I. Daubechies "Dbm" avec m le nombre de moments nuls.

obtenir des ondelettes "optimales".

Ce que l'on voudrait, ce sont des ψ à **support le plus petit possible**. Et on veut pouvoir également apprécier la régularité Lipschitz-α des signaux $x(u)$ via leur écart à une représentation polynomiale (Th. 3). Donc, il faut aussi que l'ondelette ψ soit "transparente" vis-à-vis des polynômes d'un certain degré m, ce qui se traduit par le fait que l'ondelette ait m **moments nuls** (Eq. 126). On sait en effet d'après le théorème 12 que la décroissance des coefficients d'ondelettes $|W_x(v,s)|$ donne accès à l'ordre α de la régularité de x, c'est-à-dire $|\langle x, \psi_j \rangle| < C2^{j(\alpha+1/2)}$.

Nous avons donc à satisfaire 2 contraintes: est-ce possible? La réponse est oui. Si ψ a m moments nuls et à support compact, alors $\hat{\psi}(\omega) = O(\omega^m)$ au voisinage de 0 [54]. Autrement dit, imposer des moments nuls, impose la façon dont le filtre passe-bande $\hat{\psi}(\omega)$ s'écrase plus ou moins au voisinage de 0. De plus $\hat{g}(\omega) = O(\omega^m)$ qui vient du fait que $\sqrt{2}\hat{\psi}(2\omega) = \hat{g}(\omega)\hat{\phi}(\omega)$ et $\hat{\phi}(\omega)$ ne s'annule pas en 0. Il se trouve que la réciproque est

54. NDJE: disons qu'avec les mains on peut se servir des dérivées de la TF de ψ et montrer que $\partial^{(n)}\hat{\psi}(\omega) = 0$ pour tout $n \leq m$, et ensuite on peut utiliser le développement de Taylor de $\hat{\psi}(\omega)$.

aussi vraie, que l'on peut résumer selon:

Propriété 1

$$\psi(u)\ m \text{ moments nuls } + \text{ support compact} \Leftrightarrow \hat{\psi}(\omega) = O(\omega^m) \Leftrightarrow \hat{g}(\omega) = O(\omega^m)$$

Donc, pour imposer des moments nuls à $\psi(u)$ nous procéderons à la construction ou à la vérification que le filtre $\hat{g}(\omega)$ a les propriétés de bien s'écraser au voisinage de 0. Et nous utiliserons la propriété suivante

Propriété 2

$$\psi(u) \text{ a un support compact} \Leftrightarrow h[n] \text{ a un support compact}$$

Cette propriété se déduit du raisonnement suivant: Si $\phi(u)$ est à support compact, par la définition de $h[n]$ comme produit scalaire entre $\phi(u)$ et $\phi(2u-n)$ on en déduit qu'il n'y a qu'un nombre fini de n pour lesquels $h[n] \neq 0$; ensuite d'après la relation entre $g[n]$ et $h[n]$ on en déduit que $g[n]$ est aussi non nul pour un nombre fini de n; et finalement la relation entre $\psi(u)$, $\phi(2u-n)$ et $g[n]$ assure que ψ est à support compact. Enfin, si $h[n]$ n'est pas à support compact, alors $\phi(u/2)$ et $\phi(u-n)$ ont un recouvrement non nul, donc $\phi(u)$ ne peut être à support compact.

En fait, ce qu'a démontré I. Daubechies est que l'on ne peut pas satisfaire toutes les propriétés que l'on aimerait en même temps: en particulier **si on veut augmenter le nombre de moments nuls alors le support de ψ (et ϕ) doit croître**. Ainsi, satisfaire la volonté de localisation spatiale de ψ et la régularité du filtre $\hat{\psi}$ au voisinage de 0 sont en balance, et donc il y a une optimisation à faire.

Propriété 3

Si ψ a m moments nuls et définit une base d'ondelettes orthonormales, alors le $Supp(\psi) \geq 2m-1$, et les filtres h et g ont une taille $2m$.

(nb. le cas $m = 1$ redonne l'ondelette de Haar qui n'a en effet qu'un seul moment nul à savoir que son intégrale est nulle, mais ne peut annuler un monôme de degré 1.) Quelques

ondelettes d'I. Daubechies sont montrées sur la figure 45, sachant que l'ondelette de Haar fait partie de la famille des "Dbm" avec m le nombre de moments nuls (pour Haar/"Db1", l'ondelette n'a qu'un seul moment nul). On note que plus m augmente plus l'ondelette est régulière, en plus d'avoir son support augmenté.

8.2 Algorithmes en bancs de filtres (II): DWT/IDWT

Dans la section 7.3, nous avions développé un algorithme en cascade de filtrage-sous-échantillonnage à l'aide du filtre passe-bas $h[n]$ (Fig. 41) afin d'obtenir des approximations successives en basse dimension $P_{V_j}x$. L'idée maintenant est d'introduire [55] les détails de x via l'application du filtre passe-bande $g[n]$ pour obtenir les $P_{W_j}x$. Un exemple générique de décomposition qui nous servira de support est donné sur la figure 46.

Donc, l'idée est de partir d'un échantillonnage $a_L[n]$ du signal $x(u)$ qui constitue l'approximation à une certaine échelle de référence L (le nombre d'échantillons est $N_s = 2^L$ et l'intervalle d'échantillonnage $\Delta_u \propto 1/N_s$), c'est-à-dire

$$P_{V_L}x(u) = \sum_n a_L[n]\ \phi_{L,n}(u) \tag{216}$$

A partir de ces échantillons, on va pouvoir reconstruire les coefficients d'ondelettes aux échelles spatiales plus grandes $2\Delta u$, $2^2\Delta u$, jusqu'à une échelle limite où la taille du support de l'approximation est du même ordre de grandeur de la fonction échantillonnée (moyennant les effets de bord). Comment procède-t'on?

Si on connait $P_{V_{j-1}}x$, on sait que

$$\begin{aligned} P_{V_{j-1}}x &= P_{V_j}x + P_{W_j}x \\ \Rightarrow \sum_n a_{j-1}[n]\phi_{j-1,n}(u) &= \sum_n a_j[n]\phi_{j,n}(u) + \sum_n d_j[n]\psi_{j,n}(u) \end{aligned} \tag{217}$$

Or, nous connaissons la décomposition de $\phi_{j,n}$ sur la base des $\phi_{j-1,m}$ (Eq. 183) et on obtient

$$\langle \phi_{j-1,m}, \phi_{j,n} \rangle = h[m-2n] \tag{218}$$

55. NDJE: voir également le Cours 2018 Sec. 6.6.0.4.

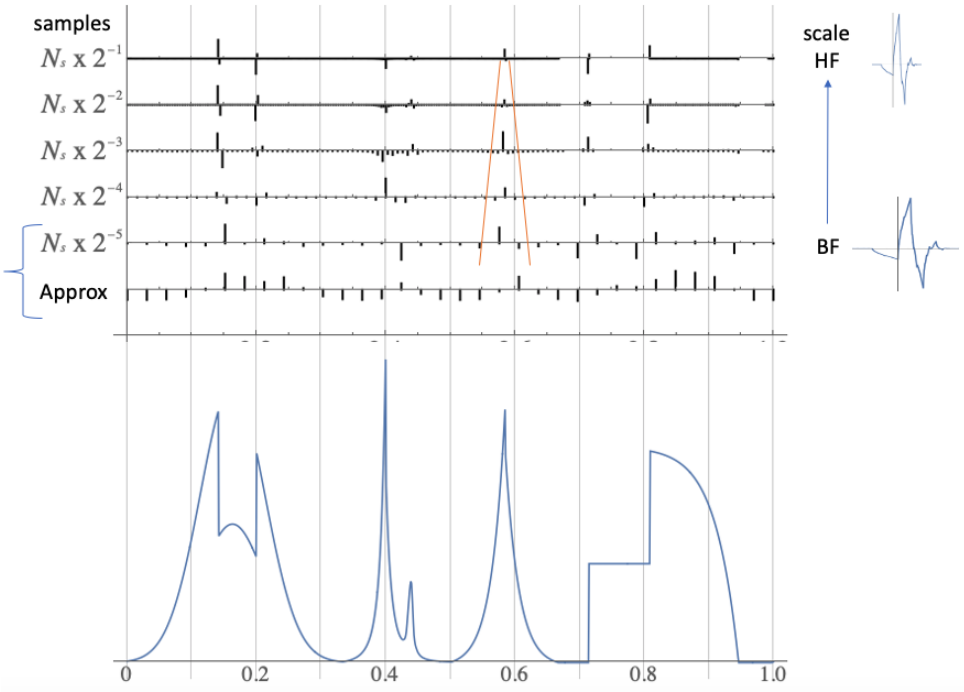

FIGURE 46 – De bas en haut: la fonction échantillonnée pour donner $2^S = 1024$ échantillons ($S = 10$), puis par décomposition successive avec l'ondelette de "Db2" on aboutit aux coefficients d'approximation basse-fréquence au nombre de $2^{S-5} = 32$ et aux coefficients de détails à la même échelle, puis on dispose des coefficients de détails de plus en plus de hautes fréquences jusqu'à aboutir à ceux au nombre de $2^{S-1} = 512$. Notons que la plupart des coefficients de détails sont quasiment nuls, et que les coefficients de détails les plus importants se concentrent aux discontinuités de la fonction d'origine dans le cône du Th. 12 (orange). Les échelles verticales ne sont pas les mêmes pour chaque lots de coefficients. *Nb. la présentation au vidéo-projecteur de S. Mallat est dans l'ordre inverse pour l'agencement des coefficients de détails allant de haut en bas, des basses fréquences aux hautes fréquences* .

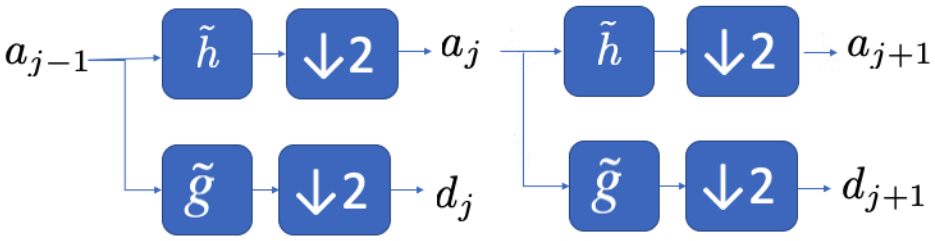

FIGURE 47 – A partir des coefficients d'approximation a_{j-1}, on obtient les coefficients d'approximation a_j et de détails d_j (Eqs. 219, 221). Puis, on peut cascader l'algorithme à partir des a_j, et ainsi de suite.

Comme les $\phi_{j,n}$ et $\psi_{j,m}$ sont orthogonaux, il vient alors en procédant au produit scalaire avec $\phi_{j,n}$ (en anonymisant l'indice n)

$$\boxed{a_j[n] = \sum_{m\in\mathbb{Z}} h[m-2n]a_{j-1}[m] = (a * \tilde{h})[2n]} \tag{219}$$

avec $\tilde{h}[n] = h[-n]$. De même, on peut montrer sur le même schéma développé dans la section 7.3 que

$$\boxed{\psi_{j+1,p}(u) = \sum_{n\in\mathbb{Z}} g[n]\phi_{j,n+2p}(u)} \tag{220}$$

et donc les coefficients de détails se déduisent des a_{j-1} selon

$$\boxed{d_j[n] = \sum_{m\in\mathbb{Z}} g[m-2n]a_{j-1}[m] = (a * \tilde{g})[2n]} \tag{221}$$

La cellule élémentaire de l'algorithme est présentée sur la figure 47, et l'algorithme DWT (Discret Wavelet Transform) sur la figure 48.

Donc, d'un point de vue algorithmique, le passage des a_{j-1} aux (a_j, d_j) se fait avec les filtres h et g, ce qui rend bien entendu attractif les filtres à support fini de Daubechies, mais il faut voir qu'en sous-jacent on calcule des produits scalaires avec des ondelettes qui fixent toutes les propriétés de ces coefficients (ex. la parcimonie). Quel est le coût calcul de la transformation DWT? A chaque cellule de la décomposition de a_{j-1} en (a_j, d_j), on a un nombre constant d'opérations si les filtres sont à support finit $(2m)$, à savoir $2m$

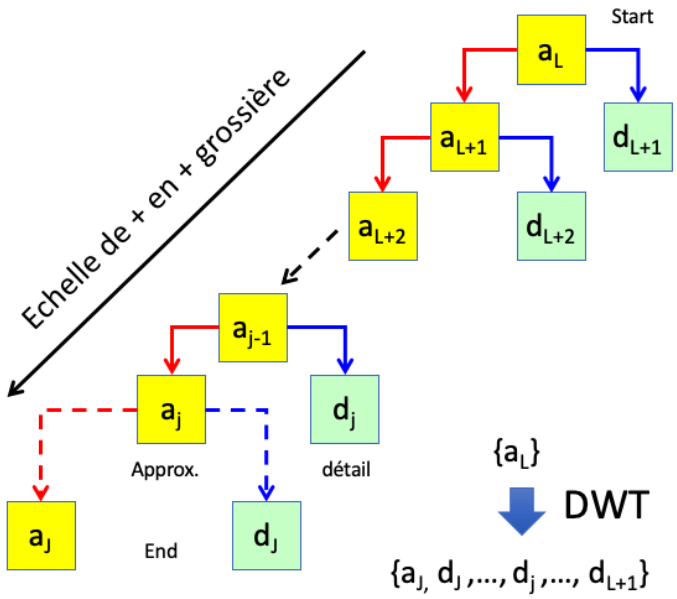

FIGURE 48 – Schéma d'une décomposition complète en ondelettes (Discret Wavelet Transform).

multiplications et $2m$ additions par coefficient, or ce nombre de coefficients est divisé par 2 à chaque étape. Donc, finalement si on part de N échantillons à l'échelle L et que l'on décompose jusqu'à l'échelle J plus grossière, le nombre d'opérations est égal à

$$\sum_{j=1}^{L-J} N2^{-j}(4m) = 4mN(1-2^{J-L}) \leq 4mN \tag{222}$$

C'est-à-dire que **le nombre d'opérations pour la DWT est linéaire en** N, donc plus rapide que la FFT qui est $O(N \log N)$, et plus les filtres sont à support petit plus rapide est la DWT.

L'algorithme est inversible (Inverse Discrete Wavelet Transform, ou IDWT) et la structure emboitée et les bases orthonormales fournissent les formules de synthèse. A partir de la relation 217 et des produits scalaires

$$\langle \phi_{j-1,m}, \phi_{j,n} \rangle = h[m-2n] \qquad \langle \phi_{j-1,m}, \psi_{j,n} \rangle = g[m-2n] \tag{223}$$

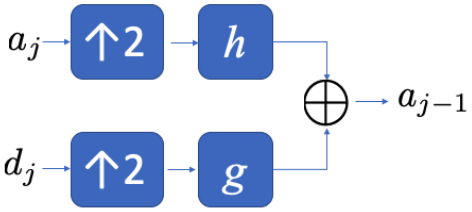

FIGURE 49 – A partir des coefficients a_j et d_j, on peut reconstruire les coefficients a_{j-1} suivant les formules Eqs. 224, 226, la dernière faisant apparaitre le grossissement d'un facteur 2 des tailles des coefficients a_j et d_j par insertion de 0 avant le filtrage.

On obtient la relation suivante qui permet de construire une approximation à l'échelle $j-1$ plus fine (ou de plus haute fréqence)

$$\boxed{a_{j-1}[n] = \sum_{m\in\mathbb{Z}} (a_j[m]h[n-2m] + d_j[m]g[n-2m])} \tag{224}$$

Afin de faire ressortir une opération de convolution, à cause du $2m$ dans les filtres, on va redéfinir les a_j et d_j en insérant des 0 selon:

$$\check{a}_{j-1}[n'] = \begin{cases} a_{j-1}[n] & n' = 2n \\ 0 & n' = 2n+1 \end{cases} \tag{225}$$

idem pour $\check{d}_{j-1}$. Ainsi on peut écrire

$$\boxed{a_{j-1}[n] = (\check{a}_j * h)[n] + (\check{d}_j * g)[n]} \tag{226}$$

Le schéma de la cellule élémentaire de l'algorithme IDWT est donné sur la figure 49. Maintenant, on se rend compte en itérant le processus que pour reconstruire l'approximation d'un certain ordre, on a besoin de a_J et d_J puis uniquement les coefficients de détails d_{J-1}, d_{J-2}, etc. Les coefficients a_{J-1}, a_{J-2} etc ne sont que des intermédiaires de calcul. L'algorithme complet de l'IDWT est présenté sur la figure 50.

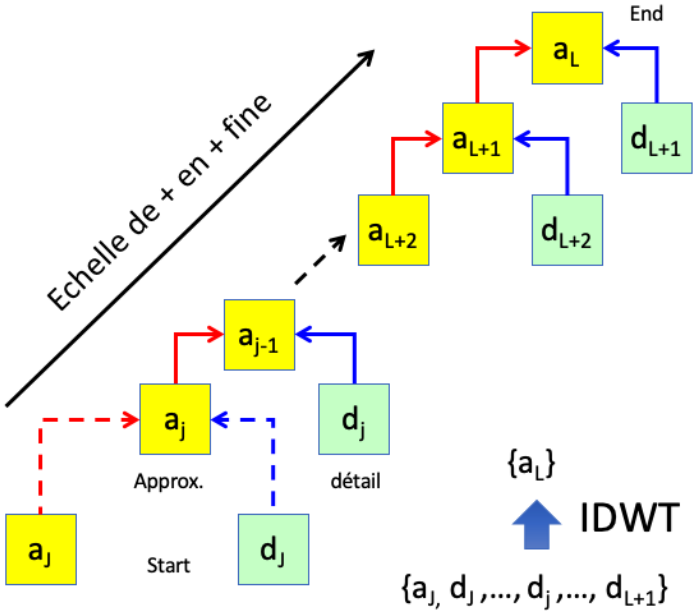

FIGURE 50 – Schéma d'une synthèse complète en ondelettes (Inverse Discret Wavelet Transform).

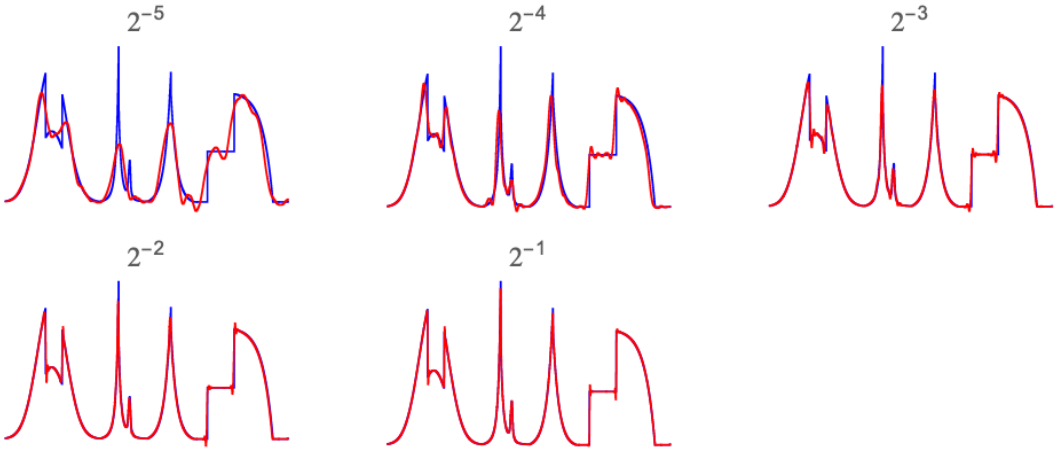

FIGURE 51 – Approximations de type $P_{V_j}x$ du signal x (Fig. 46) obtenues avec les différents coefficients a_j depuis $j = J$ jusqu'à $j = L + 1$.

8.3 Approximations du signal: expérimentation

Avec les coefficients $\{a_J, d_J, d_{J-1}, \ldots, d_{L+1}\}$, on peut construire différents types d'approximations du signal x. Par exemple, on peut éliminer tous les coefficients des hautes fréquences (les d_j) pour obtenir des **approximations linéaires** basses fréquences $P_{V_j}x$ avec les coefficients a_j. Un exemple est montré sur la figure 51 avec la fonction de la figure 46.

Seulement si on zoome sur l'une de ces approximations linéaires (Fig. 52), on se rend compte que les erreurs se situent au niveau des discontinuités car on a lissé le signal, et on observe aussi des petites oscillations résiduelles qui sont le fait du phénomène de Gibbs [56].

Comment faire mieux? En se souvenant de la section 5.1, on sait que si l'on veut

56. Le "phénomène de Gibbs" (de Josiah Willard Gibbs, physicien, 1839-1903): il s'agit d'un phénomène de non convergence uniforme des séries de Fourier mis en évidence par Henry Wilbraham en 1848, puis discuté par Albert Michelson (prix Nobel de Physique, 1852-31) en 1898 ainsi que Gibbs dans la revue *Nature*. Entre parenthèse, la machine de Michelson (https://www.youtube.com/watch?v=NAsM30MAHLg&feature=em-comments) ne pouvait mettre en évidence ce phénomène malgré la légende qui perdure. Ceci dit ce n'est qu'en 1906 que le mathématicien américain Maxime Bôcher (1867-18) en donna une explication claire et le baptisa du nom de "phénomène de Gibbs" souvent également cité Wilbraham-Gibbs surtout via la constante éponyme. Par la suite, on a étendu la définition de ce type d'erreur d'approximation linéaire "basse fréquence".

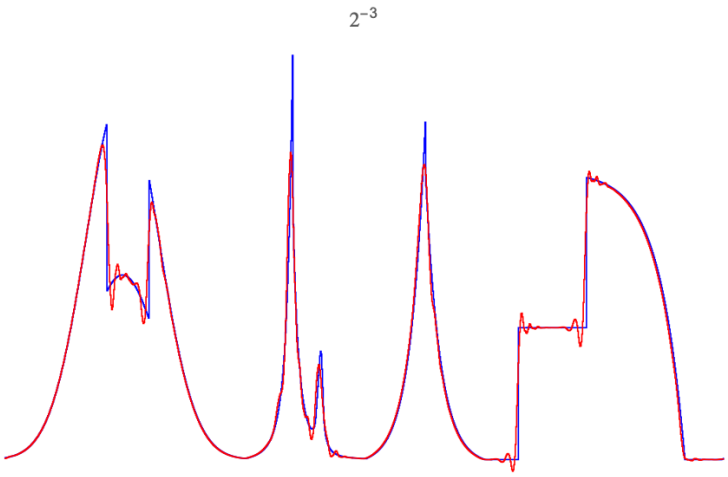

FIGURE 52 – Extrait de la figure 51 pour montrer que les erreurs de l'approximation basse fréquence se situent au niveau des discontinuités qui engendrent un phénomène de Gibbs.

une **approximation non-linéaire** de qualité, il faut pouvoir **s'adapter à la fonction/signal sous-jacent**, par exemple en ne gardant que **les plus grands coefficients de la décomposition**. Ce que l'on peut faire par exemple, c'est prendre la décomposition DWT qui donne l'approximation linéaire comme celle notée "2^{-5}" sur la figure 51 qui correspond à $J = 5$, puis de ne considérer dans la liste $\{a_J, d_J, d_{J-1}, \ldots, d_{L+1}\}$ que les coefficients les plus importants pour en avoir que 128 au total qui correspond à peu près au nombre de coefficients de l'approximation linéaire "2^{-3}" (Fig. 52). Le résultat est donné sur la figure 53 et montre la nette amélioration de l'approximation (nb. en montant à 150 coefficients on ne voit plus les petites oscillations de Gibbs). Il faut se rappeler que l'on part de 1024 échantillons de la fonction/signal au départ. En termes de qualité d'approximation avec le même nombre de coefficients, on a:

$$\frac{\|f - f_{\text{app. lin.}}\|^2}{\|f\|^2} = 1.06\ 10^{-2} \qquad \frac{\|f - f_{\text{app. n-lin.}}\|^2}{\|f\|^2} = 6.95\ 10^{-4}$$

Donc, on a pu obtenir une représentation parcimonieuse du signal $x(u)$ et obtenir une bonne approximation non-linéaire qui surpasse l'approximation linéaire de plus d'un facteur 10.

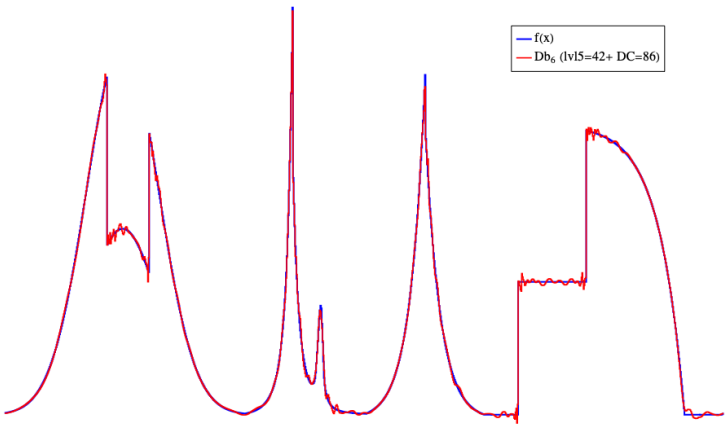

FIGURE 53 – En bleu foncé la fonction/signal sous-jacent et en rouge l'approximation non-linéaire qui est obtenue avec les 128 coefficients les plus importants de la décomposition DWT. Cette approximation non-linéaire est à comparer avec l'approximation linéaire de la figure 52 qui a à peu près le même nombre de coefficients.

8.4 Ondelettes en 2D

Le traitement d'images par ondelettes se fait tout aussi bien qu'en 1D à une subtilité prés. Deux exemples sont montrés sur la figure 54. On a besoin de 3 ondelettes ψ, pourquoi? A priori, en 2D la variable u a deux composantes (u_1, u_2), on pourrait être tenté alors de faire un produit séparable avec la base 1D des $\{\psi_{j,n}\}_{j,n}$: $\{\psi_{j_1,n_1}(u_1)\psi_{j_2,n_2}(u_2)\}_{j_1,n_1,j_2,n_2}$. L'inconvénient majeur de ce type de traitement est de privilégier les 2 directions u_1 et u_2 avec des échelles dissociées 2^{j_1} et 2^{j_2}, alors que l'on a envie d'avoir le même comportement de la décomposition si l'image est tournée. Donc, on a envie d'avoir des supports d'ondelettes associées à l'échelle 2^j dans les 2 directions. En fait, il faut revenir à l'idée originelle des multirésolutions: l'image est d'abord approximée sur une grille 2D régulière, et progressivement on réduit la résolution par sous-échantillonnage et l'on regarde comment incorporer les détails pour remonter la chaine en sens inverse.

On peut séparer les échelles des variables à j fixé pour obtenir une approximation basse fréquence en utilisant $\{\phi_{j,n_1}(u_1)\phi_{j,n_2}(u_2)\}_{j,n_1,n_2}$ qui définit une base de l'espace V_j d'approximation sur une grille 2D d'échelle 2^j. Et de même qu'en 1D, on peut calculer des approximations de plus en plus grossières $P_{V_{j+1}}x$, $P_{V_{j+2}}x$, etc. Et maintenant tout comme

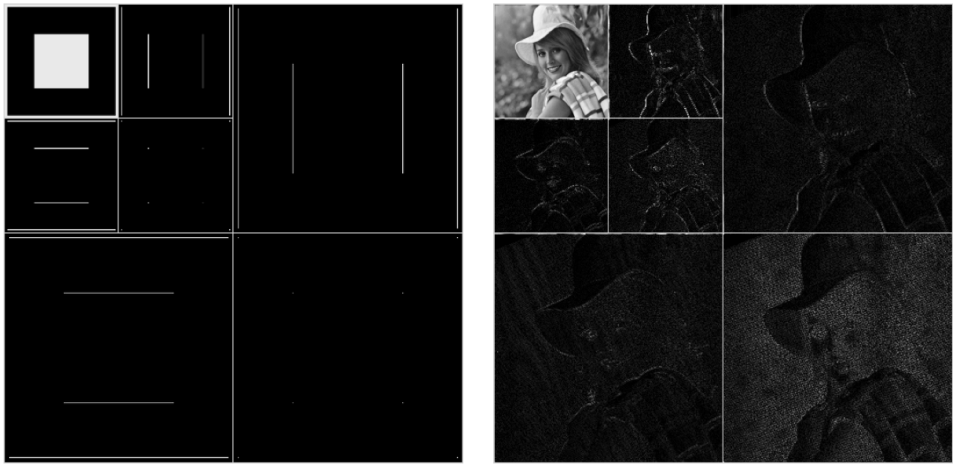

FIGURE 54 – Exemples de décomposition en ondelettes de deux images 512×512: tout d'abord l'image est décomposée en une imagette constituant l'approximation basse-fréquence de taille 256×256 et 3 imagettes de "détails" à la même échelle sensibles aux discontinuités verticales (en haut à droite), horizontales (en bas à gauche) et dans les deux directions (en bas à droite). Puis l'imagette basse-fréquence est de nouveau décomposée comme l'image d'origine, on obtient de nouveau 4 imagettes 128×128, et ainsi de suite. Tout en haut à gauche on a donc l'imagette 128×128 d'approximation linéaire la plus grossière de l'image d'origine. On a donc une pyramide d'images, là où en 1D on a une cascade de signaux.

en 1D, les détails vont apparaitre dans la transformation qui fait passer de $P_{V_j}x$ à $P_{V_{j-1}}x$. Pour ce la, il faut pouvoir décomposer V_{j-1} comme une somme directe de V_j et d'un espace complémentaire W_j (Eq. 208). On peut se placer dans l'espace des fréquences de Fourier (Fig. 55) et se convaincre de la nécessité de 3 types d'ondelettes pour compléter l'espace des fréquences entre la zone couverte par l'approximation sur V_{j-1} (plus grossière donc couvrant une plus petite zone autour de l'origine) et celle sur V_j.

Concrètement, on définit les 3 ondelettes qui couvrent les 3 zones de Fourier de la façon suivante:

$$\begin{cases} \psi^1(u_1,u_2) &= \psi(u_1)\phi(u_2) \\ \psi^2(u_1,u_2) &= \phi(u_1)\psi(u_2) \\ \psi^3(u_1,u_2) &= \psi(u_1)\psi(u_2) \end{cases} \tag{227}$$

Comme on doit faire cela à toutes les échelles, on va dilater/contracter et translater ces trois ondelettes selon

$$\psi^k_{j,n}(u) = \frac{1}{2^j}\psi^k\left(\frac{u-2^j n}{2^j}\right) \qquad u=(u_1,u_2),\ n=(n_1,n_2) \tag{228}$$

Ce que l'on démontre c'est que **la famille** $\left\{\psi^k_{j,n}\right\}_{n\in\mathbb{Z}^2}$ **est une base orthonormale de** W_j. De même,

$$\phi(u_1,u_2) = \phi(u_1)\phi(u_2) \tag{229}$$

également dilater/contracter et translater définit une famille $\{\phi_{j,n}\}_{n\in\mathbb{Z}^2}$ **qui est une base orthonormale de** V_j. Et par complément récursif on peut obtenir **une base orthonormale de** $L^2(\mathbb{R}^2)$ **avec la famille** $\left\{\psi^k_{j,n}\right\}_{j\in\mathbb{Z},n\in\mathbb{Z}^2}$.

Du point de vue algorithmique, la cascade de filtrage-sous-échantillonnage est globalement conservée, il faut par contre étendre les coefficients de détails qui maintenant couvrent 3 directions:

$$d_j[n] = (\langle x,\psi^1_{j,n}\rangle, \langle x,\psi^2_{j,n}\rangle, \langle x,\psi^3_{j,n}\rangle) \tag{230}$$

tandis que les coefficients de basse fréquence (a_j) sont égaux formellement à

$$a_j[n] = \langle x,\phi_{j,n}\rangle \tag{231}$$

avec cette fois $\phi_{j,n}$ agissant sur les deux directions (u_1,u_2).

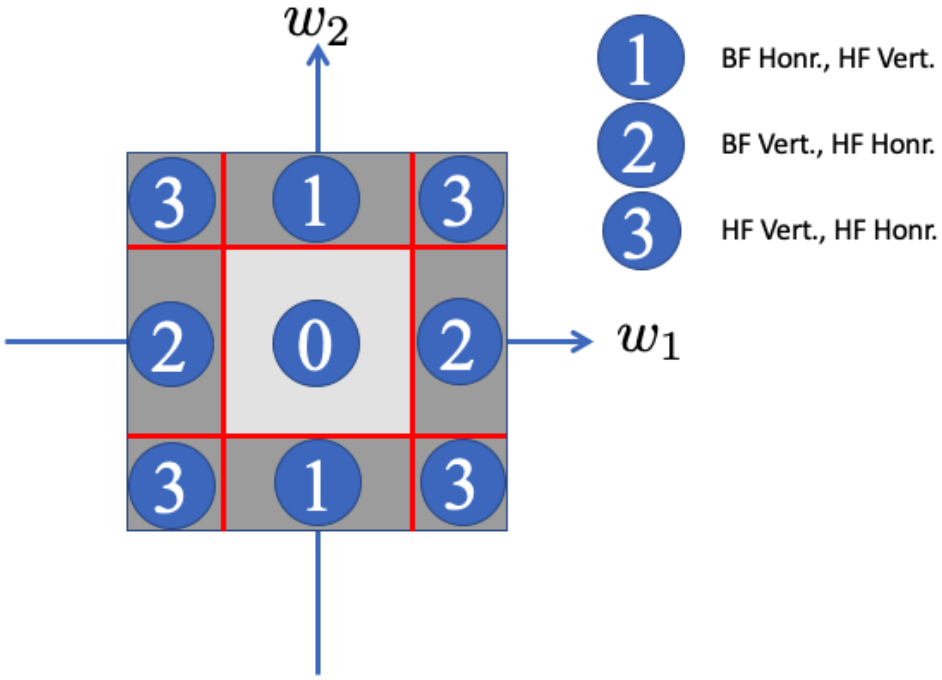

FIGURE 55 – Découpage du plan de Fourier 2D quand on passe d'une approximation V_j (carré gris-foncé) à une approximation plus grossière V_{j-1} (carré gris-clair): il faut compléter cette dernière zone par 3 types de zones numérotées de 1 à 3 correspondant à des couples de fréquences qui respectivement détectent: des contours *horizontaux* soit rapidement variables selon l'axe vertical donc de hautes fréquences verticales, des contours *verticaux* soit rapidement variables selon l'axe horizontal donc de hautes fréquences horizontales, et enfin des contours qui comportent les deux types de variations.

FIGURE 56 – Encart d'une image originale 512×512, au milieu une approximation linéaire qui réduit d'un facteur 16 la taille de l'image, et à droite une approximation non-linéaire également avec 1/16 des coefficients initiaux en partant d'un décomposition qui réduit d'un facteur 32 et complète avec des coefficients de détails.

On peut tout comme en 1D, obtenir des approximations d'une image linéaires et non-linéaires bien plus performantes comme illustré sur la figure 56.

9. Séance du 10 Mars

9.1 Résumé des notions développées dans les séances précédentes

Au long des différentes séances, nous avons abordé le triangle RAP qui traite des problématiques de *Régularité* qui conditionnent les **Approximations** en basse dimension au cœur du traitement de données et le lien avec les *Représentations Parcimonieuses*. On a montré qu'il y avait des équivalences entre ces trois notions, et que l'on peut les voir avec deux points de vue *linéaire* versus *non-linéaire*. Dans l'approche linéaire, on a revu que les approximations obtenues par projection sur des espaces linéaires sont associées à des formes de régularité qui s'expriment en particulier à travers la base de Fourier pour des problèmes invariants par translation; tandis que l'approche non-linéaire donne lieu à des approximations qui sont des projections sur des unions d'espaces linéaires (MRA) et où on sélectionne en s'adaptant au cas par cas les coefficients les plus représentatifs dans des bases orthonormales d'Ondelettes.

Petit rappel, en linéaire la notion de régularité de la fonction f est intimement reliée

à la décroissance des coefficients de Fourier $|\hat{f}(\omega)|$. Cette régularité est globale, c'est-à-dire que la moindre discontinuité locale gouverne la décroissance de $|\hat{f}(\omega)|$, donc *primo* dégrade l'approximation de basse dimension, et *secundo* "masque" en quelque sorte le fait que la fonction puisse être très régulière en dehors de ce point singulier. La question qui se pose alors est peut-on faire mieux et comment? La réponse est oui, il faut pouvoir localiser l'information de ces discontinuités/transitoires. Notons qu'il n'est pas anecdotique/superflu de faire cela car ces singularités/transitoires sont porteuses de sens (ex. détection de contours, d'attaque de note, etc).

On a vu également que ces approches linéaires/non-linéaires nous donnent des points d'entrée pour essayer de comprendre les réseaux de neurones à 1 couche cachée à M neurones (Sec. 5.2). En particulier, l'approche linéaire permet de comprendre le *Théorème d'Universalité* par analogie avec un développement en Séries de Fourier et un changement de bases que l'on peut schématiser par la formule "cosinus vers ReLU". Et on comprend également que ce théorème est vain, car on démontre que pour des fonctions régulières on a [57]

$$f \in L^2 \Rightarrow \lim_{M\to\infty} \|f - f_M\| = 0 \qquad \text{(Universalité)} \tag{232}$$

$$f \in H^\alpha \Rightarrow \|f - f_M\| = o(M^{-\alpha/d}) \qquad \text{(Malédiction)} \tag{233}$$

Dans le cadre non-linéaire, on a vu également l'approche de A. Barron qui consiste à s'adapter pour chaque fonction f afin d'obtenir une représentation parcimonieuse et ne prendre que les coefficients les plus importants pour définir l'approximation de basse dimension. Cette démarche semble consistante pour saisir qu'en effet dans le cadre d'un réseau de neurones, ce dernier va être entrainé pour répondre à une question précise: reconnaitre une image de chat parmi des images de chiens, de tasses à café etc, reconnaitre le son d'un piano parmi celui d'un violon, d'une harpe, etc. En particulier, on utilise la norme ℓ^p ($p < 2$) avec préférentiellement $p = 1$ pour la parcimonie, si on s'assure que l'on contrôle les coefficients de Fourier de la fonction alors (Th. 10)

$$\sum_{n\in\mathbb{Z}^d} |\langle f(x), e_n(x)\rangle|^p \leq \infty \Rightarrow \|f - f_M\| = o(M^{-2/p+1}) \qquad \text{(Indép. de } d\text{)} \tag{234}$$

il n'y a plus de malédiction de la dimensionalité, la convergence est indépendante de la

57. H^α est une espace de Sobolev associé au facteur α soit "l'ordre de dérivation".

dimension d. Cependant, de nouveau ce théorème est vain. Pourquoi? La raison en est simple: le théorème nous dit que tout se passe bien si la fonction est sparse en Fourier. Or, les images de chats, chiens, etc ou les trames musicales ou les spectres des voix ne sont pas sparse du tout en Fourier, et on n'explique pas du tout par cette approche les performances des réseaux de neurones profonds.

Néanmoins, les techniques adaptatives en basse dimension sont beaucoup plus puissantes que les analyses de Fourier pour détecter les transitoires. Ainsi, nous avons passé en revue la mise en œuvre de multirésolutions et trouver les bases orthonormales d'Ondelettes (Sec. 7.1). On a revu le triangle RAP en se posant la question de l'extension de la notion de régularité au sens de Sobolev. Le point clé est de capturer des singularités localisées [58]. Comme dans le cas de Fourier, on est amené à calculer les corrélations entre la fonction et des ondelettes dilatées/contractées et translatées de moyenne nulle ($\psi_{u,s}$, $\int \psi(u)du = 0$), ce sont les coefficients d'ondelettes $W_f(u,s)$. L'intensité de ces coefficients indique dans le plan (u,s) à la fois où (u_0) et sur quelle échelle (s_0) la fonction varie. On a vu que si l'on impose que l'ondelette ψ a des moments nuls, alors elle ignore des parties polynomiales du signal et l'intensité des coefficients W_f reflètent les écarts à cette forme polynomiale.

Cependant, afin *primo* d'obtenir des approximations de basse dimension et *secundo* de comprendre le lien entre régularité et décroissance des coefficients W_f, on a mis en place des représentations qui permettent une reconstruction du signal par échantillonnage sur le plan (u,s): $s = 2^j$ et $u = n2^j$ avec $(j,n) \in \mathbb{Z}^2$. On aboutit alors aux bases orthonormales d'Ondelettes $\{\psi_{j,n}(u) = 2^{-j/2}\psi(2^{-j}u - n)\}_{(j,n)\in\mathbb{Z}^2}$ qui permettent une généralisation du théorème d'échantillonnage de Shannon (Th. 16). Dans le plan de Fourier, les $\hat{\psi}_j(\omega) = \hat{\psi}(2^j\omega)$ définissent des filtres passe-bandes plus ou moins dilatés dans lesquels le spectre $\hat{f}(\omega)$ est analysé. Le point clé pour que la reconstruction de f soit possible, est que l'ensemble des filtres doit couvrir intégralement le plan de Fourier (condition de *Littlewood-Paley*) à savoir $\sum_{j\in\mathbb{Z}} |\hat{\psi}_j(\omega)|^2 = 1$.

Maintenant, historiquement on connaissait les ondelettes de Haar et de Shannon (Sec. 7.1.2) qui sont certes des bases orthonormales mais sont soit discontinue dans l'espace réel pour la première, soit un passe-bande idéal pour la seconde ce qui dans l'espace réel se traduit par une décroissance lente en $1/u$. La possibilité d'obtenir des solutions à

58. nb. tout en se restreignant néanmoins aux cas où le nombre de singularités n'est pas grand

décroissance rapide et régulières a été longtemps jugée infaisable. Mais nous avons vu que l'on peut construire de telles bases orthonormales. En premier lieu avec une ondelette ψ à la fois C^∞ et à décroissance rapide d'Y. Meyer (Sec. 6.3.3), puis avec des ondelettes construites à partir d'ensembles emboités V_j fournissant une approximation linéaire du signal $P_{V_j}f$ à une échelle 2^j et dont les bases orthonormales se déduisent par transformation d'échelles (S. Mallat/Y. Meyer). A partir de là, des ondelettes à *support compact* et possédant un certain nombre de *moment nuls* ont pu être constuites (I. Daubechies, Sec. 8.1). Dans ce cadre, les ondelettes ψ_j permettent de capturer les détails du signal f à une échelle 2^j. Ces détails viennent compléter l'approximation de basse dimension $P_{V_j}f$ qui est l'essence de la décomposition

$$P_{V_{j-1}f} = P_{V_jf} + P_{W_jf} \tag{235}$$

On démontre alors qu'il existe une fonction ϕ (*scaling function*) qui permet d'obtenir une base orthonormale de V_j, à savoir $\{\phi_{j,n}(u) = 2^{-j/2}\phi(2^{-j}u - n)\}_{n\in\mathbb{Z}}$, et qu'il existe une fonction ψ (l'ondelette) qui donne une base orthonormale de W_j, à savoir $\{\psi_{j,n}(u) = 2^{-j/2}\psi(2^{-j}u-n)\}_{n\in\mathbb{Z}}$. Et finalement la réunion de toutes les bases $\{\psi_{j,n}\}_{(j,n)\in\mathbb{Z}^2}$ forme une base orthonormée de l'espace $L^2(\mathbb{R})$. Les deux fonctions ϕ et ψ sont reliées par une relation vue au cours du théorème 18 (Sec. 7.4). Le point clé sous-jacent est que la *scaling function* ϕ et l'ondelette ψ sont déterminées par les propriétés de deux filtres 2π-périodiques $h(\omega)$ et $g(\omega)$ respectivement

$$\hat{\phi}(\omega) = \frac{1}{\sqrt{2}}\prod_{p=1}^{\infty}\hat{h}(2^{-p}\omega) \qquad\qquad \hat{\psi}(\omega) = \frac{1}{\sqrt{2}}\hat{g}(\omega/2)\hat{\phi}(\omega/2) \tag{236}$$

et $h(\omega)$ et $g(\omega)$ sont reliés l'un à l'autre par

$$\hat{g}(\omega) = e^{-i\omega}\hat{h}^*(\omega+\pi) \tag{237}$$

avec $\hat{h}$ satisfaisant

$$\forall\omega\in[0,\pi/2],\ \hat{h}(\omega)\neq 0 \qquad\qquad |\hat{h}(\omega)|^2 + |\hat{h}(\omega+\pi)|^2 = 2 \tag{238}$$

Ainsi, l'ondelette ψ est le résultat d'une cascade de filtres passe-bas à différentes échelles suivie d'un filtre passe-bande. A partir de la base d'ondelettes de L^2, on peut projeter

n'importe quelle fonction f selon

$$f = \sum_{(j,n)\in\mathbb{Z}^2} \langle f, \psi_{j,n}\rangle \psi_{j,n} \tag{239}$$

et obtenir les coefficients d'ondelettes $W_f(j,n) = \langle f, \psi_{j,n}\rangle$. Ces coefficients d'ondelettes sont essentiellement nuls (représentation parcimonieuse) sauf au niveau des singularités/transitions du signal.

Maintenant, ce qu'il a été important d'un point de vue pratique (au moins), c'est la mise au point d'algorithmes par S. Mallat de transformations rapides DWT et son inverse IDWT, dont la complexité est $O(N)$ soit plus rapides que la FFT qui est en $O(N \log N)$(Sec. 8.2). Ces algorithmes sont uniquement basés sur les propriétés des coefficients de Fourier, $h[n]$ et $g[n]$ ($n \in \mathbb{Z}$), des filtres $\hat{h}(\omega)$ et $\hat{g}(\omega)$. On réalise une cascade de filtrages-sous-échantillonnages lors d'une DWT (Fig. 47), et une cascade inverse dans l'IDWT (Fig. 49). Cependant, en sous-jacent c'est bien de la structure emboitée des ensembles V_j et de leurs complémentaires W_j qu'émane la structure de ces cascades. Lors de la décomposition (DWT), les coefficients obtenus via le filtre h sont des composantes à basse-fréquence du signal tandis que les coefficients obtenus via le filtre g sont des composantes à haute-fréquence.

Ce schéma a été étendu (Y. Meyer) en 2D pour traiter des images avec cependant une subtilité (Sec. 8.4). En effet, à partir de la scaling function ϕ et l'ondelette ψ, pour obtenir une base orthonormale de $L^2(\mathbb{R}^2)$, il nous faut à présent définir 3 ondelettes $\{\psi^k(u_1,u_2)\}_{k\leq 3}$. Ces trois ondelettes sont obtenues pour couvrir des zones de l'espace de Fourier (Fig. 55) qui sont sensibles à des transitoires soit horizontales, soit verticales, soit ayant les deux types de variations. En dimension D, il nous faut $2^D - 1$ ondelettes. Pour compléter la décomposition, il nous faut définir $\phi(u_1,u_2)$ comme simple produit de $\phi(u_1)\phi(u_2)$. Ainsi, on peut mettre en place un algorithme pyramidal rapide (Fig. 54) où cette fois la partie haute-fréquence a trois composantes. Tout comme en 1D, l'essentiel des coefficients d'ondelettes sont nuls, et seuls sont significatifs ceux qui signalent un transitoire dans son domaine de sélection.

Finalement, on peut relier l'analyse de la régularité du signal au comportement des coefficients d'ondelettes. Si on considère la régularité locale selon Lipschitz-α/Hölder,

c'est-à-dire par exemple en u_0

$$|f(u) - f(u_0)| \leq C|u - u_0|^\alpha \tag{240}$$

selon la valeur de α on obtient différents types de comportements (Sec. 5.3.2). Une façon de voir la régularité Lipschitz-α et de quantifier l'amplitude de l'incrément par rapport à $u = u_0$, c'est-à-dire

$$|f(u_0(1+s)) - f(u_0)| \leq C'|s|^\alpha \tag{241}$$

qui donne un scaling en $|s|^\alpha$. Donc, on relie la propriété de régularité à celle du comportement du signal lors d'une dilatation/contraction, ce qui rend assez intuitif que la transformation en ondelettes soit bien adaptée pour capturer ce type de régularité. Et en effet nous avons formulé le théorème (S. Jaffard, Th. 12) que l'on peut écrire dans le contexte des grilles dyadiques comme:

$$\text{Si } f \text{ Lipschitz}{-}\alpha \text{ en } u_0 \Rightarrow |\langle f, \psi_{j,n}\rangle| \leq C2^{(\alpha+1/2)j}(1 + |n - 2^{-j}u_0|^\alpha) \tag{242}$$

et surtout on a inversement si pour $\alpha' < \alpha$ on a

$$|\langle f, \psi_{j,n}\rangle| \leq C2^{(\alpha+1/2)j}(1 + |n - 2^{-j}u_0|^{\alpha'}) \tag{243}$$

alors f est Lipschitz-α en u_0. Ainsi, on comprend que les coefficients les plus importants sont localisés à l'échelle 2^j dans le cône où les supports des ondelettes ψ_j ont un recouvrement non nul avec les singularités du signal. Par exemple, on sait que aller vers les hautes fréquences c'est faire tendre s vers 0 donc j vers $-\infty$, et donc plus α est grand plus la décroissance des coefficients est rapide. On comprend alors les décompositions en ondelettes telle que celle de la figure 46. En passant, si la fonction est bornée, ce qui correspond à $\alpha = 0$, les coefficients d'ondelettes décroissent au moins en $2^{j/2}$ (rappel: $s \to 0$ vers les htes fréquences, $\Leftrightarrow j \to -\infty$).

A partir de ces décompositions, on peut concevoir des approximations du signal f. Tout d'abord, on peut ne conserver que les approximations basses fréquences (*approximation linéaire*) qui correspondent aux projections sur les espaces V_j, à savoir les $P_{V_j}f$ comme illustré sur la figure 51. Alors on fixe la taille de la grille dans laquelle on projette le signal, et c'est à rapprocher de la stratégie que l'on utilise quand on veut calculer une meilleure base par exemple en analyse PCA. Mais, on a constaté le même type de

problèmes qu'en Fourier, à savoir des oscillations de Gibbs aux endroits des singularités/transitions du signal, car on utilise que les basses fréquences. Pour faire mieux, on a eu recours à la description parcimonieuse qui s'adapte (*approximation non-linéaire*) à la fonction f en requérant de ne garder que les coefficients d'ondelettes supérieurs à un seuil T:

$$f_T = \sum_{|\langle f,\psi_{j,n}\rangle|>T} \langle f, \psi_{j,n}\rangle \psi_{j,n} \tag{244}$$

Ce faisant, on a pu vérifier que cela soit en 1D (Fig. 52) ou en 2D (Fig. 56) que l'on peut reconstruire la fonction/image originale avec ses détails fins en ne conservant que 10% environ de l'ensemble des coefficients d'ondelettes.

9.2 Amélioration quantitative du passage au non-linéaire

Considérons le théorème suivant [59]

Théorème 19 *(cadre linéaire)*
Si $x \in C^\alpha[0,1]$ (Lipschitz-α), et si on ne garde que M coefficients "basse fréquence", alors $\exists C$ tq.

$$\varepsilon_\ell^{1D} = \|x - x_M\|^2 \leq C M^{-2\alpha} \tag{245}$$

Démonstration 19.
Comme on est dans le cadre linéaire, l'approximation x_M s'obtient en projetant x dans un espace linéaire V_L, à savoir $P_{V_L} x$. Il faut seulement adapter la taille L pour n'avoir que M coefficients. Or,

$$P_{V_L} x = \sum_{n \leq M} \langle x, \phi_{L,n}\rangle \phi_{L,n} \tag{246}$$

avec

$$\phi_{L,n}(u) = \frac{1}{\sqrt{2^L}} \phi(2^{-L}u - n) \tag{247}$$

Comme le support de x est $[0,1]$, alors [60] $n \in [0, 2^{-L}]$. Donc, $M = 2^{-L}$.

59. NDJE: attention si dans la section précédente la notation de la fonction est $f(u)$ pour coller aux slides de S. Mallat en séance, ici on reprend la notation $x(u)$.
60. attention $L < 0$ puisque 2^L est le pas d'échantillonnage sur $[0,1]$, et on prend $2^L \gg 1$.

Maintenant concernant l'erreur, comme on a la décomposition suivante de $L^2(\mathbb{R})$ (Sec. 7.4)

$$L^2(\mathbb{R}) = V_L \bigoplus_{j=-\infty}^{L} W_j$$

on peut formuler l'expression de l'erreur de ne retenir que la projection sur V_L

$$\|x - x_M\|^2 = \sum_{j=-\infty}^{L} \sum_{n=0}^{2^{-j}} |\langle x, \psi_{j,n} \rangle|^2 \tag{248}$$

Or, on sait que les coefficients d'ondelettes décroissent selon

$$|\langle x, \psi_{j,n} \rangle| \leq C 2^{j(\alpha+1/2)} \tag{249}$$

Donc,

$$\|x - x_M\|^2 \leq C^2 \sum_{j=-\infty}^{L} \sum_{n=0}^{2^{-j}} 2^{(2\alpha+1)j} = C'(2^{2\alpha L}) = C' M^{-2\alpha} \tag{250}$$

■

C'est un résultat identique que l'on a obtenu en Fourier (Th. 6) dans le cas d'une fonction régulière. Mais ce qui nous intéresse, ce sont les cas où la fonction est irrégulière.

Il est intéressant d'étudier le cas 1D, car il y a **aucun coût à coder des singularités**. Mettons que le signal a Q singularités $u_1, \ldots, u_Q$ dont on ne connait pas la localisation. La question est: combien de coefficients d'ondelettes vont être affectés par ces singularités? Supposons que x est C^α sur chaque intervalle $]u_k, u_{k+1}[$. A l'intérieur de ces intervalles, l'ondelette va être "transparente" et on va pouvoir appliquer le théorème précédent. Là où les coefficients vont être grands c'est aux abords des discontinuités (c'est-à-dire dans le "cône" dont on a parlé plus avant). Or, si l'ondelette a un support de taille finie, on peut se convaincre que pour toute échelle 2^j, le nombre d'ondelettes translatées dont le support va contenir une singularité est constant, notons le K. Ainsi, le nombre total de coefficients affectés est $QK \times N_s$ avec N_s le nombre d'échelles s que l'on conserve. Or, ce dernier est égal à $|L'|$ si on coupe la décomposition à une échelle L' à hautes fréquences.

Donc, le nombre de coefficients concernés par les singularités est

$$M_1 = QK|L'|$$

Que vaut L'? Si le signal est au moins borné sur son support (singularités comprises) ce qui correspond à $\alpha = 0$, alors les coefficients d'ondelettes à l'échelle 2^j décroissent au moins en $2^{j/2}$. Mais si on supprime les échelles $j \in]-\infty, L']$, on fait une erreur que l'on peut quantifier

$$\sum_{j=-\infty}^{L'} (QK)C^2 2^j = C^2 QK 2^{L'+1} = C' 2^{L'} \tag{251}$$

Or, on voudrait que cette précision soit du même ordre que celle du cas linéaire, c'est-à-dire $C'2^{L'} \approx M^{-2\alpha}$, ce qui donne $|L'| \sim 2\alpha \log M \ll M$. Donc le nombre total de coefficients est égal à ceux concernés par l'approximation linéaire et ceux concernés par la présence des singularités donc c'est un nombre de l'ordre de

$$M_{tot} \sim M + C'' \log M$$

Donc, en augmentant très peu le nombre de coefficients, l'erreur en présence de singularités est du même ordre que l'erreur sans singularité.

Finalement, on aboutit au théorème suivant

Théorème 20 *(cadre non-linéaire)*
Si $x \in C^\alpha$ sur chaque intervalle $]t_k, t_{k+1}[$ $(k = 1, \ldots, K)$, en prenant les M plus grands coefficients d'ondelettes, alors $\exists C > 0$ tq.

$$\varepsilon_{n\ell}^{1D} = \|x - x_M\|^2 \leq C M^{-2\alpha} \tag{252}$$

Donc, quand on approxime une fonction avec singularités dans un cadre linéaire, ce qui va gouverner la décroissance de l'erreur, c'est le α le plus petit dans l'intervalle $[0, 1]$ et on se retrouve pratiquement à prendre $\alpha \approx 0$ (fonction bornée) à cause des singularités en escalier qui donnerait une décroissance en $1/M$ (prop. non démontrée); tandis qu'avec une approche non-linéaire on n'est plus limité par cette décroissance lente, le α peut être plus grand. Ainsi, on comprend pourquoi une approche non-linéaire en 1D gagne beaucoup par rapport à une approche linéaire.

Par contre, **en plus grande dimension le phénomène change complètement**. Si on reprend le théorème 19, quand on calcule le nombre de coefficients, il faut se placer sur une grille, et en 2D on a $M = 2^{-2L}$. Toujours, en 2D, on sait que la normalisation des ondelettes passe de $1/\sqrt{2^j}$ à $1/2^j$ pour normaliser la norme L^2. Alors, la condition sur les coefficients d'ondelettes est modifiée en

$$|\langle x, \psi_{j,n}\rangle| \leq C 2^{j(\alpha+1)} \tag{253}$$

et si on regarde l'erreur linéaire, on doit considérer non pas 2^{-j} coefficients n mais 2^{-2j} coefficients (n_1, n_2), ainsi

$$\varepsilon_\ell^{2D} = \|x - x_M\|^2 \leq C^2 \sum_{j=-\infty}^{L} 2^{-2j} 2^{2j(\alpha+1)} \leq C' 2^{2\alpha L} = C' M^{-\alpha} \tag{254}$$

On généralise en dimension q facilement, et on voit apparaître alors que

$$\varepsilon_\ell^{2D} \leq C' M^{-2\alpha/q} \qquad \text{(sans singularité)} \tag{255}$$

Bon, si on prend des fonctions régulières C^α sur $[0,1]^2$ (2D) on ne perd pas beaucoup [61].

Mais si l'on prend une fonction uniquement C^α par morceaux, c'est-à-dire avec des singularités, là il y a un changement car même l'approximation non-linéaire va être limitée. Prenons comme exemple l'image du carré (Fig. 54) et plus précisément un bord horizontal du carré blanc, et reprenons le cheminement du raisonnement en 1D en identifiant les différences. Le nombre d'ondelettes qui vont être sensibles au bord est de l'ordre de:

$$M_j = \ell/2^j \times K \tag{256}$$

avec ℓ la taille du bord, 2^j est le nombre de translations le long du bord d'une ondelette à cette échelle, et K le nombre (constant) d'ondelettes concernées dans la direction verticale (raisonnement 1D). Maintenant, la fonction a une marche d'escalier bornée, donc $\alpha = 0$ ce qui donne une contrainte sur les coefficients d'ondelettes (Eq. 253)

$$|\langle x, \psi_{j,n}\rangle| \leq C 2^j \tag{257}$$

61. En grande dimension, on aurait déjà un problème...

Donc, l'erreur linéaire en présence de ce bord est de l'ordre de

$$\varepsilon_{n\ell}^{2D} \leq \sum_{j=-\infty}^{L'} M_j C^2 2^{2j} = (\ell K C^2) 2^{L'} \tag{258}$$

Mais, $|L'|$ est toujours d'ordre $\log M$, cependant $M = 2^{-2L}$ donc $|L'| \sim 2|L|$. Ainsi, on obtient

$$\varepsilon_{n\ell}^{2D} \leq C' M^{-1} \qquad \text{(avec singularité)} \tag{259}$$

Ce que ça dit, c'est que même si l'image est C^∞ (ex. à l'intérieur/extérieur du carré blanc), **l'erreur non-linéaire est dominée par les discontinuités**. Plus on augmente la dimension q, plus la surface des discontinuités augmente, et donc il va falloir augmenter le nombre de coefficients pour obtenir une bonne approximation.

9.3 La compression

On se place à présent du point de vue de la quantité d'information qu'il nous faut garder pour bien reconstruire une image. D'un point de vue naïf, on peut se dire que l'on peut coder chaque pixel de l'image soit avec par exemple 8bits (256 niveaux de gris), ou bien 1bit (B & W). En passant de 8bits à 1bit, on gagne évidemment un facteur 8 en taille, mais la qualité s'en ressent, et on a perdu beaucoup d'information (Fig. 57). Mais on peut faire mieux, l'idée est la suivante, on se place dans la base orthonormale d'ondelettes et ce sont les coefficients d'ondelettes que l'on va compresser (fonction $Q(x)$) sur peu de bits:

$$\tilde{x} = \sum_{j,n} Q(\langle x, \psi_{j,n} \rangle) \psi_{j,n} \tag{260}$$

Si maintenant, on regarde la distribution des valeurs des coefficients d'ondelettes (les détails seulement) sur la figure 58, on constate que les distributions sont certes piquées à 0 (illustration de la parcimonie), mais selon la complexité des textures l'étalement est plus ou moins important.

Donc pour coder, il est naturel de se placer dans la base d'ondelettes qui donne des représentations parcimonieuses avec beaucoup de 0, et on va dans un premier temps ne coder que les coefficients non nuls. Pour ce faire, on utilise un quantificateur à pas Δ fixe

FIGURE 57 – A gauche image dont la valeur d'un pixel est codée sur 8bits (1byte) et à droite la même image où le codage se fait sur 1bit.

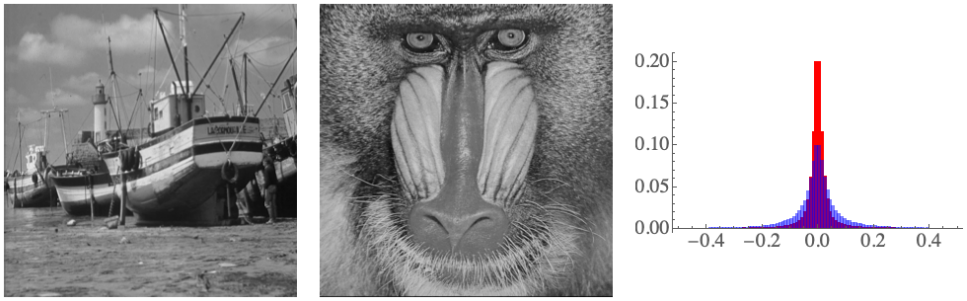

FIGURE 58 – A gauche et au milieu deux images. A droite, en rouge l'histogramme normalisé des coefficients d'ondelettes (de détails seulement) pour l'image du bateau, en bleu pour celle du singe montrant plus de textures.

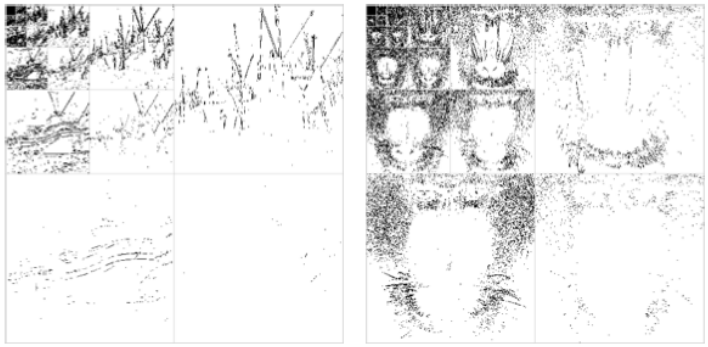

FIGURE 59 – Pour les deux images de la figure 58, on a procédé à la binarisation Eq. 262 avec les paramètres $a = 5$ et $p = 15$ du quantificateur Eq. 261 (nb. la distribution des coefficients de détails ont des queues assez longues). Les coefficients dont le résultat est 0 apparaissent en blanc sur les images. La décomposition en ondelettes est effectuée avec "Db2".

dont la définition est la suivante: si la valeur x prend des valeurs dans $[-a, a]$ et si cet intervalle est découpé en $n = 2p + 1$ boîtes, alors

$$Q_\Delta(x) = k\Delta \qquad \Delta = 2a/n \qquad k = \lfloor x/\Delta + 1/2 \rfloor \in [\![-p, p]\!] \tag{261}$$

c'est-à-dire que l'on affecte à x la valeur centrale de la k-ième boîte. Donc, un coefficient dont la valeur est comprise dans l'intervalle $[-\Delta/2, \Delta/2]$ est affecté à la valeur 0. Ceci est particulièrement intéressant si on prend en compte l'histogramme des valeurs des coefficients de détails (Fig. 58). Ainsi, on construit une carte binaire $b[j, n]$ telle que

$$b[j, n] = \begin{cases} 0 & \text{si } Q_\Delta(\langle x, \psi_{j,n} \rangle) = 0 \\ 1 & \text{sinon} \end{cases} \tag{262}$$

Un exemple est donné sur la figure 59. Si maintenant, on affecte une valeur aux coefficients non nuls selon le quantificateur simple ci-dessus, on obtient pour les paramètres de la figure 59, des images reconstruites de la figure 60. Pour apprécier, la différence entre l'image d'origine et deux reconstructions avec deux valeurs de p, la première celle des figures 59, et 60, et l'autre 4 fois plus grande (bin 4 fois plus petits), on montre les zooms des images reconstruites et ceux des images d'origines, sur la figure 61. Le Standard JPEG2000

FIGURE 60 – Images reconstruites à partir de la quantification des coefficients d'ondelettes avec les paramètres de la figure 59. On voit l'effet de la quantification.

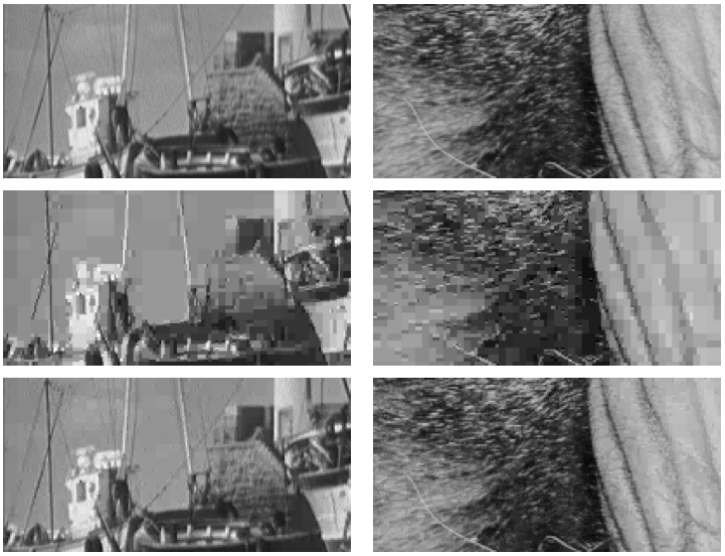

FIGURE 61 – Zoom: en haut les images d'origines, au milieu des images de la figure 60, et en bas des images reconstruites avec une quantification qui 4 fois plus de bins.

procède un peu de la sorte avec des raffinements qu'il serait trop long de mentionner ici (voir le livre de S. Mallat). Ceci dit le message est qu'avec un facteur de compression de 40 (c'est-à-dire que l'on passe de 8bits/pixel à 0.2bit/pixel) alors on reconstruit fidèlement l'image. Bien entendu, si l'on diminue trop le nombre bit/pixel, il se produit ce que l'on observe sur l'image 60, c'est-à-dire que l'on commence à perdre des détails. Typiquement, une image peut être comprimée d'un facteur 50-100 sans trop introduire d'artefacts. Pour des compressions plus importantes, on n'y arrive pas avec ce type de techniques.

Ceci dit, en quoi ce que l'on observe sur ces images renvoit-il à l'approximation non-linéaire? L'erreur d'approximation de l'image x par le codage $\tilde{x}$ (Eq. 260) peut être décrite dans la base orthonormale des ondelettes, et si l'on prend le quantificateur simple Q_Δ, il vient:

$$\begin{aligned}\|x-\tilde{x}\|^2 &= \sum_{j,n} |\langle x,\psi_{j,n}\rangle - Q_\Delta(\langle x,\psi_{j,n}\rangle)|^2 \\ &= \sum_{|\langle x,\psi_{j,n}\rangle)|\leq\Delta/2} |\langle x,\psi_{j,n}\rangle|^2 + \sum_{|\langle x,\psi_{j,n}\rangle)|>\Delta/2} (\Delta/2)^2 \end{aligned} \tag{263}$$

La première somme n'est autre que l'erreur faite quand on ne garde que les M plus grands coefficients; de plus le nombre de coefficients qui ne sont pas mis à 0 par la quantification (la seconde partie de la somme), sont au nombre de M justement par le même argument. Donc, la distorsion D due au codage est encadrée par

$$\varepsilon_{n\ell}(M) \leq D = \|x-\tilde{x}\|^2 \leq \varepsilon_{n\ell}(M) + M\Delta^2/4 \tag{264}$$

Ainsi, cette distorsion D a deux composantes: une composante non-linéaire due aux coefficients que l'on ne retient pas par le seuillage induit par la largeur du bin "0", et une composante qui est due à la quantification des coefficients qui sont retenus par l'approximation non-linéaire.

Si on suppose que les coefficients d'ondelettes sont sparses (la représentation est parcimonieuse), ce qui est suggéré par les histogrammes de la figure 58, que l'on peut traduire par une norme ℓ^p (Sec. 5.1.2) bornée, ex.

$$\sum_{j,n} |\langle x,\psi_{j,n}\rangle)|^p \leq C_p^p \tag{265}$$

alors le coefficient d'ondelette de rang k (Th. 9) satisfait

$$|\langle x, \psi_k \rangle)| \leq C_p^p k^{-1/p} \tag{266}$$

Le nombre M correspond au nombre de coefficients dont la valeur absolue est plus grande de $\Delta/2$, donc

$$\Delta/2 = C_p^p M^{-1/p} \tag{267}$$

et d'après le même théorème

$$\varepsilon_{n\ell}(M) \leq \frac{C_p^{2p}}{2/p-1} M^{1-2/p} \tag{268}$$

Donc

$$D = \|x - \tilde{x}\|^2 \leq \frac{C_p^{2p}}{2/p-1} M^{1-2/p} + M C_p^{2p} M^{-2/p} = \frac{C_p^{2p}}{1-p/2} M^{1-2/p} \tag{269}$$

Ainsi, l'erreur de quantification et l'erreur non-linéaire sont de même grandeur, ce qui donne finalement que $D = O(M^{1-2/p})$. Ce que cela dit c'est que **l'erreur de codage dépend essentiellement de l'approximation non-linéaire de basse dimension**. Cela est dû au fait que le nombre de coefficients mis à 0, qui contribuent à l'erreur non-linéaire, est très grand à cause de la parcimonie. Et le nombre de bits qu'il va falloir pour coder l'information a deux composantes proportionnelles l'une et l'autre à M dont la première vient de la localisation des coefficients nuls, et l'autre du codage des amplitudes des coefficients non-nuls. La chose importante à retenir est que **pour ces algorithmes de codage, il faut obtenir une représentation parcimonieuse**.

L'an prochain nous aborderons la Théorie de l'Information afin de comprendre le fonctionnement des réseaux de neurones. En particulier, en grande dimension on a un atout, le théorème central limite, qui nous indique que si l'on somme des variables indépendantes, on converge vers la moyenne. C'est un phénomène qui nous indique que l'information est concentrée dans des sous-parties de l'espace. Et le nombre de bits qui la code est défini par l'Entropie.