

Compression, ondelettes et algorithmes afférents

Pascal SZACHERSKI

8 janvier 2008

TABLE DES MATIÈRES

1 Transformée de Fourier	5
I Bases mathématiques	5
II La transformée de Fourier	6
III Convolution et Fourier	7
2 Signaux discrets et signaux périodiques	9
I Echantillonnage	9
II Aliasing	10
III Les signaux finis	11
3 Filtrage	13
I Le filtrage en pratique	13
II Le filtrage pour le débruitage	14
III Le filtrage pour MP3 (MPEG layer 3)	14
4 Traitement du son et de la parole	15
I Spectrogrammes et applications	15
I.1 Pitch Shifting et Time Stretching	15
I.2 Algorithme du vocodeur de phase (phase vocoder)	15
II Modélisation de la parole, modèle LPC	16
III Applications du LPC	18
5 Approximation et estimation	19
I Bases et représentations	19
II Approximation linéaire	20
III L'approximation non-linéaire	21
IV Estimation dans du bruit	22
IV.1 Bayésien -VS- minimax	22
IV.2 Estimateurs diagonaux et seuillages	23
6 Transformée en ondelettes orthogonales	25
I La transformée de Haar	25
I.1 Les ondelettes de Haar dans un cadre discret sur un intervalle	28
I.2 Calcul des coefficients	30
II Les autres bases d'ondelettes	31

Introduction

- (a) Qu'est-ce qu'un signal ? un son, une image, un courant électrique, un flot de bits sont des signaux. On peut diviser les signaux types :
- les signaux analogiques : un son, une photo argentique peuvent être représentés par des fonctions $\mathbb{R} \rightarrow \mathbb{R}$ ou de $\mathbb{R}^2 \rightarrow \mathbb{R}^3$. Ces signaux prennent une infinité de valeurs et sont définis en tout point de temps ou de l'espace.
 - les signaux digitaux : une image numérique, un fichier MP3, un flot de bits sont représentés par des fonctions de $\mathbb{N} \rightarrow \{0;1\}^{\mathbb{N}}$. Ces signaux sont définies sur un ensemble discret. Ils peuvent prendre un nombre fini de valeurs.
- Pourquoi passe-t-on de l'un à l'autre ? Comment traite-t-on ces différents signaux ?
- (b) Qu'est-ce qu'un son ? Il s'agit d'une variation de la pression de l'air. Quand un objet bouge, il fait bouger l'air et donc il y a une pression de l'air. L'oreille mesure cette pression et c'est le *cerveau* qui interprète et reconstruit le son. (exemple : Deux sons qui sont en contre-phase s'annule alors que chaque son pour soi peut être fort voire insupportable.) Quand on visualise un son, on visualise en effet la position de la membrane (de l'enceinte p.ex.).
- (c) Qu'est-ce qu'une image ? On a ici trois images monochrome (rouge, vert, bleu) superposées. Pourquoi trois couleurs ? Parce que l'œil humain possède trois récepteurs, sensibles à ces couleurs. Représentation par une fonction $I : [0, 1]^2 \rightarrow \mathbb{R}, (x, y) \mapsto I(x, y)$.
- (d) Pourquoi analyser un signal ? \rightarrow compression, cryptage, nettoyage, transformation, ...

I Bases mathématiques

Definition [1.I.1] On dit que $f : \mathbb{R} \rightarrow \mathbb{R}$ est intégrable sur $]a, b[$ si $\int_a^b |f(t)| dt < +\infty$.

Exemple Toute fonction continue sur $[a, b]$ est intégrable sur $]a, b[$. ($f : x \mapsto \frac{1}{x}$ sur $[1, 2]$).

Contre-exemple $f : x \mapsto \frac{1}{x}$ sur $]0, 1[$ n'est pas intégrable.

Definition [1.I.2] On dit que f est intégrable sur \mathbb{R} ou $f \in L^1(\mathbb{R})$ si $\lim_{a \rightarrow \infty} \int_{-a}^a |f(t)| dt < +\infty$.

Exemple $f : x \mapsto \frac{1}{1+x^2}, f : x \mapsto e^{-x^2}, f : x \mapsto \begin{cases} e^{-x} & x > 0 \\ 0 & x \leq 0 \end{cases}$.

Contre-exemple $f : x \mapsto x, f : x \mapsto \frac{1}{1+|x|}$ ne sont pas intégrables.

Definition [1.I.3] On dit que $f \in L^p(\mathbb{R}), p \geq 1$, si et seulement si $\lim_{a \rightarrow \infty} \int_{-a}^a |f(t)|^p dt < +\infty$.

Definition [1.I.4] Une **distribution** est une application linéaire qui à une fonction continue associe un nombre réel.

Exemple $D : f \mapsto f(5)$. (valeur)

$$D : f \mapsto \int_0^1 f(t) dt. \text{ (moyenne)}$$

$$D : f \mapsto \int_0^1 t^2 f(t) dt. \text{ (moyenne pondérée)}$$

Definition [1.I.5] On note $\delta_u(t)$ la distribution appelée **Dirac** qui à une fonction continue associe $f(u)$ et on note : $\delta_u : f \mapsto f(u)$. $f(u) = \int_{-\infty}^{+\infty} f(t)\delta_u(t) dt$. On note $\delta_0(t) =: \delta(t)$.

Definition [1.I.6] On appelle **produit de convolution** de f et de g la fonction h définie par

$$h(t) := f * g(t) := \int_{-\infty}^{+\infty} f(u)g(t-u) du.$$

Remarque $f * g(t) = g * f(t)$.

Definition [1.I.7] $e^{ix} = \cos x + i \sin x$.

II La transformée de Fourier

La transformée de Fourier est un outil d'analyse de signaux. Elle fournit une représentation fréquentielle de l'information.

Definition [1.II.1] Si f est intégrable, la transformée de Fourier \hat{f} est définie par

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt \in \mathbb{C}.$$

$\hat{f}(\omega)$ mesure combien il y a d'oscillation dans f à la fréquence ω .

Théorème [1.II.2] (Théorème de reconstruction) Si $f \in L^1(\mathbb{R})$ et si $\hat{f} \in L^1(\mathbb{R})$ alors

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} d\omega.$$

Le module de la transformée de Fourier $\hat{f}(\omega)$ correspond à l'amplitude de la fréquence, la phase au décalage de la sinusöide.

Il est impossible, à partir de $|\hat{f}(\omega)|$, de savoir quand une fréquence est apparue et quand elle disparaît.

Propriétés [1.II.3] (Propriétés essentielles de la transformée de Fourier) On a les propriétés suivantes :

Fonctions $\in L^1(\mathbb{R})$	transformée de Fourier
$f(t)$	$\hat{f}(\omega)$
$\hat{f}(t)$	$2\pi f(-\omega)$
$f * g(t)$	$\hat{f}(\omega) \cdot \hat{g}(\omega)$
$f(t)g(t)$	$\frac{1}{2\pi} \hat{f} * \hat{g}(\omega)$
$f(t-u)$	$e^{iu\omega} \hat{f}(\omega)$
$e^{i\xi t} f(t)$	$\hat{f}(\omega - \xi)$
$f(\frac{t}{s})$	$ s \hat{f}(s\omega)$
$f^{(p)}(t)$	$(i\omega)^p \hat{f}(\omega)$
$(-it)^p f(t)$	$\hat{f}^{(p)}(\omega)$
$\overline{f}(t)$	$\overline{\hat{f}(-\omega)}$
$f(t) \in \mathbb{R}$	$\hat{f}(-\omega) = \overline{\hat{f}(\omega)}$

(où pour $z = x + iy$, on a $\overline{z} = x - iy$ la conjuguée complexe).

Exemple [1.II.4] (1) Si $f(t) = \mathbb{1}_{[-T, T]}$, alors $\hat{f}(\omega) = \int_{-T}^T e^{-i\omega t} dt = 2 \frac{\sin(T\omega)}{\omega}$.

(2) Si $\hat{h}(\omega) = \mathbb{1}_{[-\nu, \nu]}$, alors $h(t) = \frac{\sin(\nu t)}{\pi t}$ ($\notin L^1$, mais $\in L^2$).

(3) Si $f(x) = e^{-t^2}$ alors $\hat{f}(\omega) = \sqrt{\pi} e^{-\frac{\omega^2}{4}}$.

(4) Si $f(t) = \delta_u(t)$, alors $\hat{f}(\omega) = e^{-i\omega u}$.

(5) Si $f(t) = \sum_{n=-\infty}^{\infty} \delta_{nT}(t)$, alors

$$\hat{f}(\omega) = \sum_{n=-\infty}^{\infty} e^{-inT\omega} = \frac{2\pi}{T} \sum_{k=-\infty}^{\infty} \delta\left(\omega - \frac{2\pi k}{T}\right).$$

III Convolution et Fourier

Les opérations classiques du traitement du signal comme la transmission de signaux, l'élimination de bruits stationnaires sont implémentées par des opérations linéaires stationnaires.

La stationnarité d'un opérateur L signifie que si l'entrée $f(t)$ est retardée de τ ($f_\tau(t) := f(t - \tau)$), alors la sortie est également retardée de τ : $g(t) = Lf(t) \Rightarrow g(t - \tau) = Lf_\tau(t)$.

Théorème [1.III.1] Si L est stationnaire, $\exists h(t)$ telle que $Lf(t) = \int_{-\infty}^{+\infty} f(t)h(t-u) du = f * h(t)$ avec $h(t) = L\delta(t)$. $h(t)$ est appelée la réponse impulsionnelle.

Propriétés [1.III.2] (Propriétés de la convolution) – *commutativité* : $h * f(t) = f * h(t)$.

– *dérivation* : $\frac{d(f*h)}{dt} = \frac{df}{dt} * h = f * \frac{dh}{dt}$.

– *convolution par un Dirac* : $f * \delta_\tau = f(t - \tau)$.

– *Les exponentielles complexes sont les vecteurs propres de l'opération de convolution* : $L e^{i\omega t} = \hat{h}(\omega) e^{i\omega t}$.

Théorème [1.III.3] Si $f \in L^1(\mathbb{R})$, $h \in L^1(\mathbb{R})$, alors $g := h * f \in L^1(\mathbb{R})$ et

$$\hat{g}(\omega) = \hat{h}(\omega)\hat{f}(\omega).$$

Théorème [1.III.4] (Théorème de Plancherel) Si $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, alors

$$\int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2 d\omega.$$

(Alors, l'énergie de la transformée égale celle du signal à un facteur près.)

Signaux discrets et signaux périodiques

I Echantillonnage

Un signal discret peut se représenter comme une somme de Diracs. A tout échantillon $f(nT)$ on associe un Dirac $f(nT)\delta(t - nT)$ situé en $t = nT$.

Ainsi un échantillonnage uniforme de f correspond à une somme pondérée de Diracs :

$$f_d(t) = \sum_{n=-\infty}^{\infty} f(nT)\delta(t - nT).$$

La transformée de Fourier de f_d est une série de Fourier :

$$\hat{f}_d(\omega) = \sum_{n=-\infty}^{\infty} f(nT) e^{-inT\omega}.$$

Proposition [2.I.1] La transformée de Fourier d'un signal discret obtenu par échantillonnage de f à intervalles de T vaut

$$\hat{f}_d(\omega) = \frac{1}{T} \sum_{k=-\infty}^{\infty} \hat{f}\left(\omega - \frac{2k\pi}{T}\right).$$

L'échantillonnage de f à intervalles T rend sa transformée de Fourier périodique de période $\frac{2\pi}{T}$ par sommation de ses translatées $\hat{f}\left(\omega - \frac{2k\pi}{T}\right)$.

Par symétrie on peut remarquer qu'une périodisation de la fonction induit une discretisation de la transformée de Fourier.

Proposition [2.I.2] La transformée de Fourier d'une fonction périodique f_p obtenue par périodisation de f à intervalles T vaut

$$\hat{f}_p(\omega) = \sum_{n=-\infty}^{\infty} f\left(\frac{2\pi n}{T}\right) \delta_{\frac{2\pi n}{T}}(\omega).$$

La formule de reconstruction devient dans ce cas

$$\begin{aligned} f_p(t) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}_p(\omega) e^{i\omega t} d\omega \\ &= \sum_{n=-\infty}^{\infty} \hat{f}\left(\frac{2\pi n}{T}\right) e^{i\frac{2\pi n}{T}t}. \end{aligned}$$

On déduit de la proposition [2.I.1] le

Théorème [2.I.3] (Théorème de Shannon-Whittaker) Si $\text{supp } \hat{f} \subseteq \left[-\frac{\pi}{T}; \frac{\pi}{T}\right]$ alors

$$f(t) = \sum_{n=-\infty}^{\infty} f(nT)h_T(t - nT) = h_T * f_d \tag{SW}$$

avec $h_T(t) = \frac{\sin \frac{\pi t}{T}}{\frac{\pi t}{T}}$.

Ce théorème assure, sous une condition sur \hat{f} , qu'on peut reconstruire $f(t) \forall t \in \mathbb{R}$ à partir des valeurs échantillonnées. Ce théorème impose que le support de f n'a pas de variation brutale entre deux échantillons consécutifs.

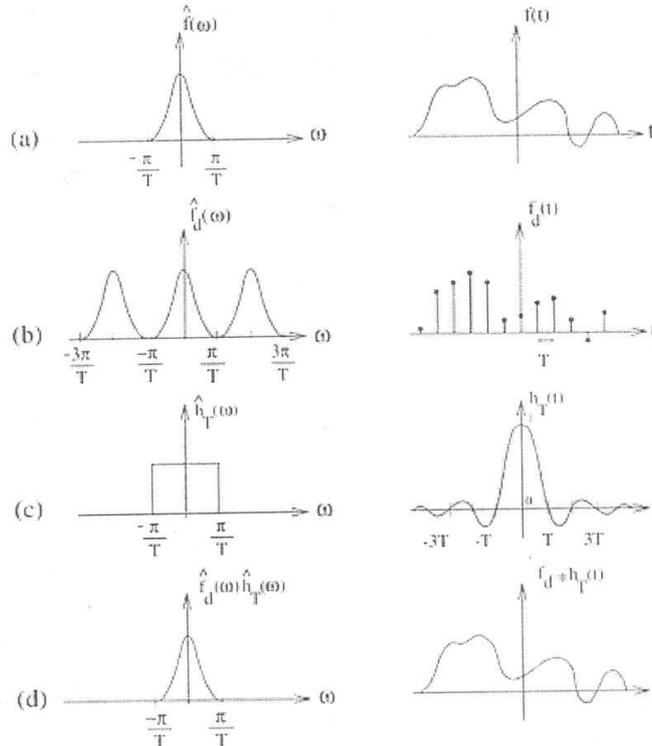


FIG. 2.1 – Echantillonnage respectant le théorème de Shannon. (de : “Une exploration des signaux en ondelettes” de Stéphane Mallat).

II Aliasing

Le pas d'échantillonnage T est souvent imposé par des contraintes sur les capacités de calcul, de mémoire ou d'acquisition et en général $\text{supp } \hat{f}$ n'est pas inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$. Dans ce cas, la formule (SW) ne redonne pas f . On peut analyser l'erreur résultante et déterminer un filtrage qui permet de la réduire.

Le repliement des composantes haute fréquence sur un intervalle à basse fréquence s'appelle **aliasing**.

Suppression de l'aliasing

Pour utiliser le théorème d'échantillonnage f est approximée par le plus proche signal \tilde{f} dont la transformée de Fourier est à support dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$.

$$\begin{aligned} \|f - \tilde{f}\|_2^2 &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega) - \hat{\tilde{f}}(\omega)|^2 d\omega \\ &= \frac{1}{2\pi} \left(\int_{|\omega| > \frac{\pi}{T}} |\hat{f}(\omega)|^2 d\omega + \int_{|\omega| \leq \frac{\pi}{T}} |\hat{f}(\omega) - \hat{\tilde{f}}(\omega)|^2 d\omega \right). \end{aligned}$$

La distance minimale est obtenue pour $\hat{\tilde{f}}(\omega) = \hat{f}(\omega) \cdot \mathbb{1}_{[-\frac{\pi}{T}, \frac{\pi}{T}]} = \frac{1}{T} \hat{h}(\omega) \hat{f}(\omega)$.

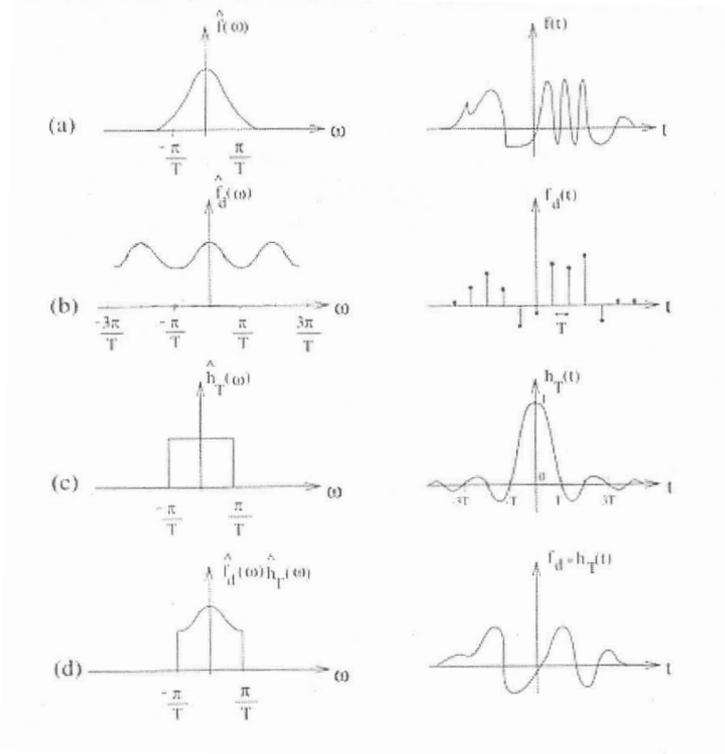


FIG. 2.2 – Echantillonnage ne respectant pas le théorème de Shannon. (de : “Une exploration des signaux en ondelettes” de Stéphane Mallat).

III Les signaux finis

Petit rappel : $f * g[n] = \sum_{k=-\infty}^{\infty} f[k]g[n - k].$

En pratique, f n’est connue que sur un support fini, disons $0 \leq n \leq N$. Il est nécessaire de modifier les convolutions pour tenir compte des effets bords en $n = 0$ et $n = N - 1$. Donc on redéfinit la transformée de Fourier pour les suites finies en vue des calculs symétriques.

Definition [2.III.1] (la convolution circulaire) Soient \tilde{f} et \tilde{h} deux signaux constitués de N échantillons. Le produit de convolution circulaire est définie par

$$\tilde{f} \circledast \tilde{h}[n] := \sum_{p=0}^{N-1} \tilde{f}[p]\tilde{h}[n - p]$$

où $\forall k \in \mathbb{Z}, \tilde{f}[k] = \tilde{f}[k \bmod N], \tilde{h}[k] = \tilde{h}[k \bmod N].$

Les valeurs propres de l’opération de la convolution circulaire

$$L_h f[n] = f \circledast h[n]$$

correspondent aux coefficients de Fourier de h , c’est-à-dire

$$\hat{h}[k] = \sum_{p=0}^{N-1} h[p] e^{-\frac{2\pi i p k}{N}}.$$

Ces valeurs propres sont associées aux vecteurs propres

$$e_k[n] = e^{\frac{2\pi i n k}{N}}.$$

Théorème [2.III.2] La famille $\left\{ e_k[n] = e^{\frac{2\pi i k n}{N}} \mid 0 \leq k \leq N \right\}$ est une base orthogonale (même orthonormée) de l’espace des signaux de période N .

Le filtrage est une opération classique en traitement du signal qui consiste à effectuer la convolution d'un signal d'entrée f par un signal h appelé filtre.

La convolution en espace correspond à une multiplication en Fourier.

- On parle de **filtre passe-bas** pour une filtre h tel que $\hat{h}(\omega) \approx 1$ dans les basses fréquences et $\hat{h}(\omega) \approx 0$ pour les hautes fréquences.
- Un **filtre passe-haut** vérifie $\hat{h}(\omega) \approx 0$ pour les basses fréquences et $\hat{h}(\omega) \approx 1$ pour les hautes fréquences.
- Un **filtre passe-bande** est une filtre h tel que $\hat{h}(\omega) \approx 1$ pour $\omega \in [\omega_1, \omega_2]$ et $\hat{h}(\omega) \approx 0$ pour $\omega \notin [\omega_1, \omega_2]$.

Exemple – Le filtre passe-bas idéal $\hat{h}(\omega) = \mathbb{1}_{\{|\omega| < \omega_1\}}$ est utilisé comme filtre anti-repliement pour supprimer l'aliasing.

- Etant donné un signal f discret et deux suites de coefficients (a_k) et (b_l) , le signal g obtenu par

$$\sum_{k=0}^K a_k f[n-k] = \sum_{l=0}^M b_l g[n-l]$$

est obtenu par filtrage ; on parle de **filtre récursif**.

$$\hat{h}(\omega) = \frac{\hat{g}(\omega)}{\hat{f}(\omega)} = \frac{\sum_{k=0}^K a_k e^{-ik\omega}}{\sum_{l=0}^M b_l e^{-il\omega}}$$

Ce type de filtre est utilisé dans le traitement de la parole.

I Le filtrage en pratique

Si on veut filtrer un signal long comme un signal audio (44 kHz signifie 44 000 échantillons par seconde), on ne considère pas le signal dans son ensemble. On le segmente de manière à étudier des signaux stationnaires sur lesquels la transformée de Fourier a une signification. Pour un signal audio standard, les trames durent de l'ordre de 30 ms.

On aurait envie de découper le signal brutalement à l'aide de fenêtres rectangulaires. Ce découpage altère le contenu fréquentiel du signal. En effet, si on multiplie f par une fonction $\mathbb{1}_{[T_1, T_2]}$, on convolve sa transformée de Fourier par un sinus cardinal (sinc). En particulier, on crée des discontinuités aux bords des trames ce qui crée artificiellement des hautes fréquences.

On doit s'efforcer de fenêtrer le signal en altérant le moins possible les fréquences. Pour cela, on fenêtre le signal avec des fenêtres régulières et en particulier continues aux bords. Suivant les applications, on utilise différentes fenêtres : *Hamming*, *Blackman*, *Gaussienne* et surtout *Hanning* qui est définie par

$$g[m] = \cos^2\left(\frac{2\pi m}{N}\right) \mathbb{1}_{[-N, N]}(m).$$

(cf. figure 3.1). (Dans MATLAB on retrouve ces filtres dans le paquetage *Wavelab*.)

Pour traiter un signal, on le multiplie p.ex. par des fenêtres Hanning translattée de $\frac{N}{2}$, $\frac{N}{4}$ ou $\frac{N}{8}$ (cf. figure 3.2). On filtre ensuite chaque trame à l'aide d'un filtre h .

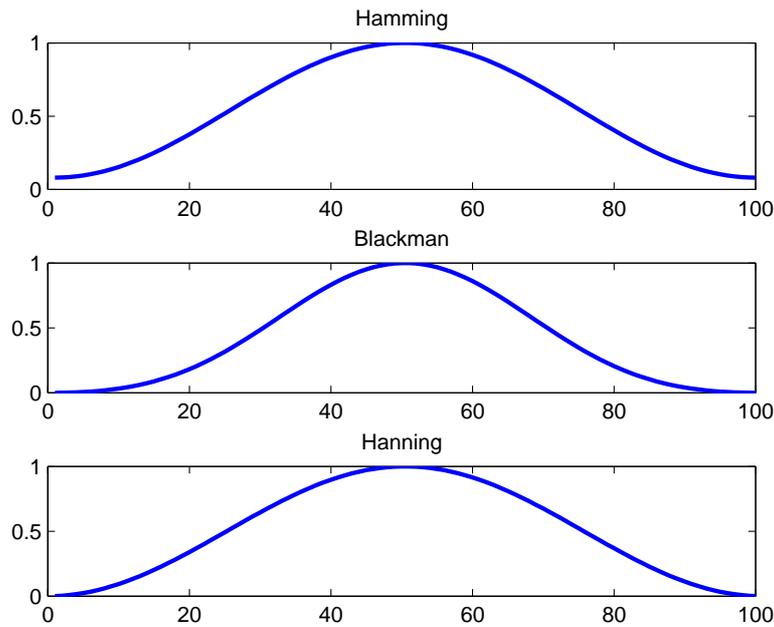


FIG. 3.1 – Les fenêtres Hamming, Blackman et Hanning

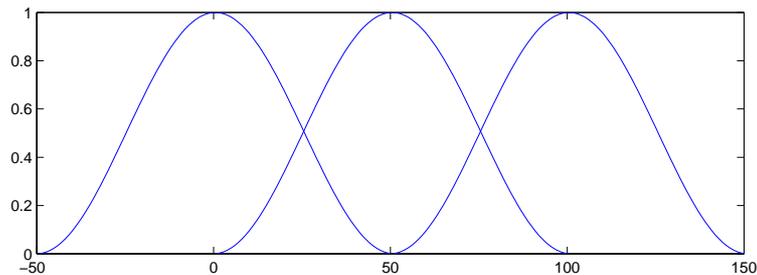


FIG. 3.2 – Les fenêtres de Hanning et deux de ses translattées

II Le fitrage pour le débruitage

Un fitrage peut servir à séparer deux signaux qui sont envoyés en même temps. Les radios émettent toutes avec le même type d'ondes ; c'est la fréquence qui change d'une radio à l'autre.

Pour retrouver le signal émis par une radio donnée, on filtre le signal autour d'une fréquence donnée.

III Le fitrage pour MP3 (MPEG layer 3)

L'oreille capte les fréquences de 20 Hz à 20 000 Hz avec une sensibilité accrue entre 1 000 Hz et 5 000 Hz. C'est-à-dire qu'à une puissance donnée en décibel ($\log_2 \left(\frac{I}{I_0} \right)^2$), un son est inaudible à 100 Hz et audible à 4 000 Hz.

La courbe de *Fletcher-Manson* décrit la sensibilité de l'oreille humaine en fonction des fréquences.

L'idée de base du MP3 est la découpage du son en 32 bandes de fréquence et de coder sur plus ou moins de bits ces fréquences.

- On fenêtré le signal sur des trames d'environ 30 ms.
- On filtre le signal avec 32 filtres qui séparent les différentes bandes de fréquence. On obtient alors 32 signaux.
- On sous-échantillonne chaque $h_k * f$ tous les 32 échantillons.
- On quantifie chaque bande de fréquence en fonction de la sensibilité de l'oreille.

I Spectrogrammes et applications

Quand on souhaite traiter un signal non-stationnaire, mais qui est stationnaire localement (par morceaux) on ne calcule pas directement une transformée de Fourier. On segmente le signal en le fenêtrant correctement et on calcule des transformées de Fourier locales. Un spectrogramme est une représentation 2D d'un signal en temps et en fréquence qui représente l'évolution fréquentielle (on dit aussi spectrale) du module de la transformée de Fourier. Il est composé d'une succession de transformées de Fourier (*fft*) de différentes fenêtres. Les fenêtres peuvent se recouvrir fortement ou pas. Un spectrogramme est défini par une longueur de fenêtre, un fenêtrage est un intervalle entre le début de chaque fenêtre. C'est un outil d'analyse du son. Plus les fenêtres sont longues, meilleure est la résolution fréquentielle ; plus elles sont courtes, meilleure est la résolution temporelle.

I.1 Pitch Shifting et Time Stretching

Le spectrogramme est un outil simple permettant de faire subir des modifications à un signal. Le pitch shifting (décalage de fréquences) et le time stretching (compression ou dilatation temporelle) sont deux opérations élémentaires duales : Si on sait faire le time stretching on peut faire du pitch shifting directement en modifiant la vitesse de lecture du signal de sortie.

Une approche simple pour faire du time stretching consiste à interpoler le spectrogramme : Considérons que le son est fenêtré avec des fenêtres se recouvrant à moitié. Si on veut obtenir un son deux fois plus rapide, on reconstruit le signal en omettant une fenêtre sur deux. Cette opération ne modifie que légèrement le contenu fréquentiel du signal. Elle revient à reconstruire un son à partir d'un spectrogramme dont on a extrait une fenêtre sur deux.

!!! Attention à la phase !

Si on veut un son deux fois plus long, on va créer de nouvelles fenêtres pour les glisser entre les fenêtres existantes. Ces fenêtres doivent avoir un contenu fréquentiel proche de leur deux voisines. Pour cela, on interpole le spectrogramme de manière linéaire.

On fait la moyenne des modules de la transformée de Fourier et on reconstruit les fenêtres synthétiques par inversion de la transformée de Fourier et par un calcul adéquat de la phase.

De manière générale, on peut ralentir ou accélérer un son d'un facteur α quelconque en interpolant le spectrogramme de manière à passer de N fenêtres à αN fenêtres.

I.2 Algorithme du vocodeur de phase (phase vocoder)

- On fenêtré le signal avec des fenêtres de Hanning qui se recouvrent à moitié (ou à $\frac{3}{4}$).
- On calcule le spectrogramme avec une interpolation linéaire du module.
- On calcule la phase et on l'interpole de manière à conserver la variation de phase.
- On reconstruit le signal par *fft inverse* et sommation des fenêtres

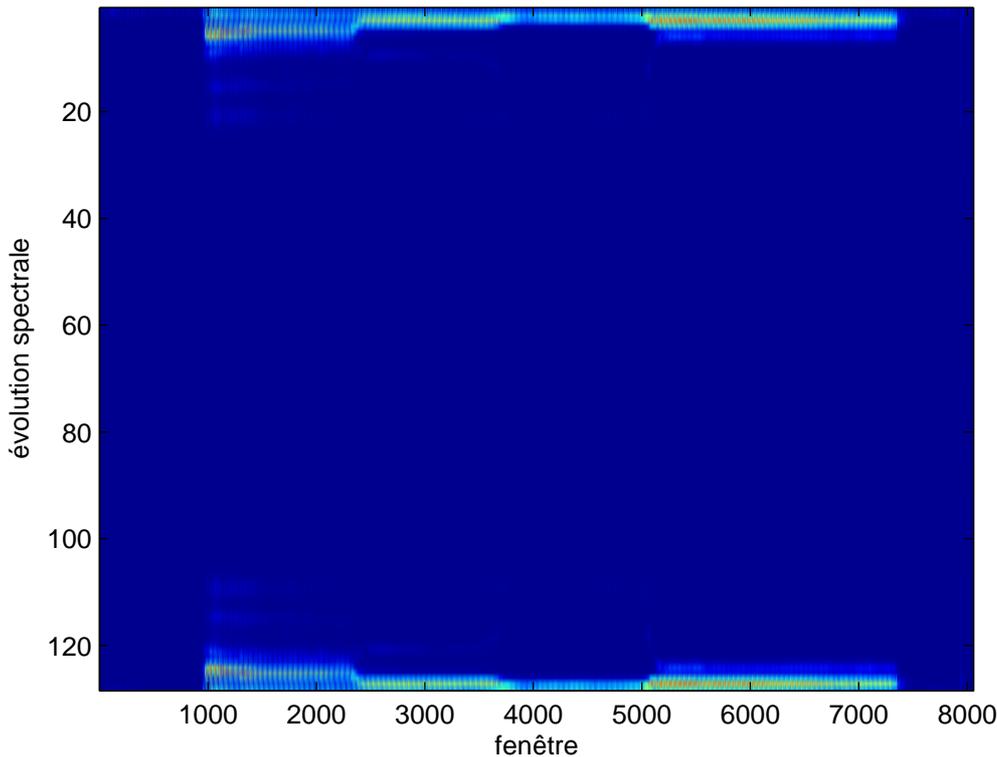


FIG. 4.1 – Spectrogramme du son aeio.wav avec une taille de fenêtre $N = 128$. Un spectrogramme est toujours symétrique !

II Modélisation de la parole, modèle LPC

La production de la parole se fait en trois étapes : Les poumons compressent l'air qui est envoyé à travers la trachée. Cet air passe dans le larynx qui est composé d'un système de cartilage et de muscles incluant les cordes vocales. Le larynx produit alors un signal d'excitation qui se propage le long du conduit vocal qui produit l'articulation de la parole.

Ce mécanisme de production de la parole conduit à un modèle mathématique de la parole. Si on modélise le conduit vocal comme une succession de p tubes (cylindres) de diamètres différents et que l'on calcule l'équation de propagation des ondes à travers le conduit, on montre que le conduit vocal se comporte comme un filtre auto-regressif, c'est à dire

$$e[n] = \sum_{k=0}^p a_k g[n-k] \quad a_k \neq 0 \forall k$$

où g est le signal de sortie et e l'excitation d'entrée.

$$\hat{g}(\omega) = \hat{h}(\omega)\hat{e}(\omega)$$

avec

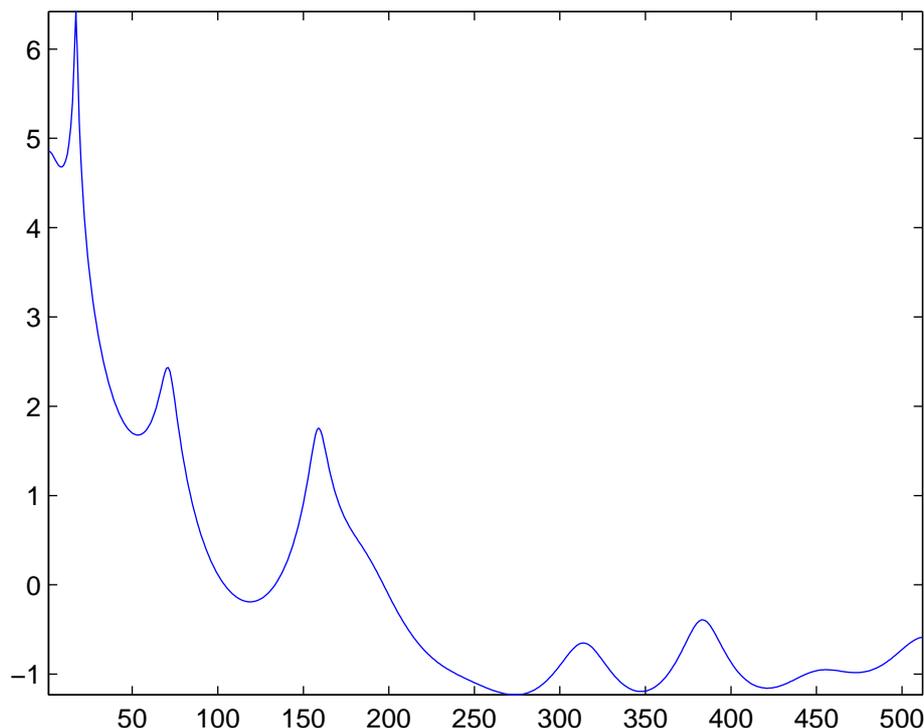
$$\hat{h}(\omega) = \frac{1}{a_0 + a_1 e^{-i\omega} + \dots + a_p e^{ip\omega}}$$

($\hat{h}(\omega)$ s'écrit aussi $\hat{h}(e^{i\omega})$)

Quitte à renormaliser on pourra prendre $a_0 = 1$.

Le filtre h amplifie les fréquences associées aux racines du polynôme

$$P(z) = a_0 + a_1 z + \dots + a_p z^p.$$


 FIG. 4.2 – Formants d’une fenêtre du son aeio .wav ayant pris $p = 20$

En effet,

$$P(\mathbf{e}^{i\omega}) = a_p \prod_{k=1}^p (\mathbf{e}^{i\omega} - r_k \mathbf{e}^{i\omega k}).$$

Les pics de module de la réponse fréquentielle $|\hat{h}(\mathbf{e}^{i\omega})|$ sont appelés **formants**. Les formants sont d’autant plus prononcés que la racine dans l’angle associé est proche du cercle unité.

A une structure de conduit vocal on associe un filtre h et donc des formants. Ils caractérisent le type de son produit, un a , un e , un s et une personne. Une méthode de reconnaissance de locuteur ou de parole consiste en l’identification des formants.

La plupart du temps, on a entre deux et quatre formants qui sont significatifs. La modélisation de la voix comme filtrage d’une excitation par un filtre autoregressif est une manière de déterminer ces formants. L’excitation est modélisé de deux manières selon si le son est voisé, c’est à dire si on utilise les cordes vocales, ou si le son est non-voisé. Les voyelles sont des sons typiquement voisés, en revanche le s , le f , le ch sont typiquement non-voisés.

Pour les signaux voisés on utilise un modèle de *trains de Diracs*. La fréquence du train de Diracs définit la hauteur du son et les formants sa nature.

Quand l’oreille entend un a , c’est l’enveloppe du a qu’elle reconstruit à partir du *spectre du raie*.

L’excitation des sons non-voisés est modélisé par un bruit blanc¹. Le signal de sortie est alors un bruit coloré².

Ce modèle de paroles correspond à une prédiction linéaire du signal g au point n à partir des échantillons précédents. En effet,

$$e[n] = \sum_{k=0}^p a_k g[n-k],$$

¹bruit blanc : variable aléatoire avec des entrées indépendantes dont sa *fft* est également aléatoire

²bruit coloré : variable aléatoire dont les entrées dépendent l’une de l’autre. Sa *fft* se situe principalement dans les basses fréquences.

on a

$$g[n] = \sum_{k=1}^n (-a_k)g[n-k] + e[n].$$

On peut interpréter $\tilde{g}[n] = \sum_{k=1}^n (-a_k)g[n-k]$ comme une estimation de $g[n]$ à partir des valeurs précédentes et l'excitation comme une erreur de prédiction.

Le codage d'un signal de voix par un filtre autoregressif et une excitation est appelé *LPC* (Linear Prediction Coding). Les $(a_k)_{k \leq p}$ sont les coefficients LPC associés au signal g et à l'entier p .

Pour les déterminer, on minimise l'erreur de prédiction linéaire. On cherche les a_k qui minimisent

$$\varepsilon^2 = \sum_{n=-\infty}^{+\infty} \left| g[n] - \sum_{k=1}^p (-a_k)g[n-k] \right|^2$$

Il existe un cadre probabiliste qui justifie ce choix des a_k quand l'excitation e est un processus aléatoire, gaussien et blanc.

En pratique, cette minimisation, qui correspond à l'inversion d'un système linéaire, donne des coefficients LPC tout à fait corrects dans les cas réels voisés et non-voisés.

III Applications du LPC

- Une application classique du modèle LPC est la *compression* de signaux de parole. Au lieu de coder une trame de 30 ms, on calcule entre 20 et 50 coefficients LPC, on estime l'excitation qu'on code sous forme d'un bruit (une variance) si le son est non-voisé ou d'un train de Diracs (pitch, gain) et on ne transmet ainsi qu'un nombre réduit de l'information. \Rightarrow CELP (Code Excited Linear Predictive).
- reconnaissance vocale ;
- synthèse vocale ;
- déclilage de signaux.

I Bases et représentations

Une des manières classiques d'analyser et de traiter (compression, débruiter) un signal est d'utiliser une représentation dans une famille de fonctions.

Si $f \in H$ et $\overline{\text{Vect}(\psi_i)_{i \in \mathbb{N}}} = H$ alors $\exists (\lambda_n)_{n \in \mathbb{N}} \subseteq \mathbb{C}^{\mathbb{N}}$ telle que

$$f = \sum_{n=0}^{+\infty} \lambda_n \psi_n,$$

On
écrira aussi
 $\langle (\psi_i)_{i \in \mathbb{N}} \rangle =$
 $\text{Vect}(\psi_i)_{i \in \mathbb{N}}$.

c'est à dire

$$\lim_{N \rightarrow \infty} \left\| f - \sum_{n=0}^N \lambda_n \psi_n \right\|_H = 0.$$

Dans la suite, on considèra que H est un espace de Hilbert.

Definition Espace de Hilbert : espace vectoriel normé complet dont la norme est issue d'un produit scalaire. Soit $\langle \cdot, \cdot \rangle_H$ le produit scalaire de H , $\|f\|_H = \sqrt{\langle f, f \rangle}$.

La famille $(\psi_i)_{i \in \mathbb{N}}$ peut avoir de *bonnes* propriétés ou non, elle peut être libre ou liée, orthogonale ou non, normée ou non. Si la famille $(\psi_i)_{i \in \mathbb{N}}$ est libre, alors les coefficients $(\lambda_n)_{n \in \mathbb{N}}$ sont uniques. Si elle est liée, les coefficients $(\lambda_n)_{n \in \mathbb{N}}$ peuvent ne pas l'être.

Si la famille $(\psi_i)_{i \in \mathbb{N}}$ est une base orthonormée, i. e. $\langle \psi_i, \psi_j \rangle = \delta_{ij}$, alors la représentation est unique et les coefficients $(\lambda_n)_{n \in \mathbb{N}}$ se calculent par un simple produit scalaire :

$$\begin{aligned} \lambda_n &= \langle f, \psi_n \rangle \\ \implies f &= \sum_{n=0}^{\infty} \langle f, \psi_n \rangle \psi_n \end{aligned} \quad \text{(formule d'inversion)}$$

Si $(\psi_n)_{n \in \mathbb{N}}$ est une BON on a le théorème de Pythagore :

$$\sum_{n=0}^{\infty} |\lambda_n|^2 = \|f\|_H^2$$

où dans la plupart du temps on pose $H = L^2(\mathbb{R})$, $\langle f, g \rangle = \int_{\mathbb{R}} f \bar{g}$, donc on identifie $\|\cdot\|_H$ avec $\|\cdot\|_2$.

Exemple Fourier discret, Diracs (\rightarrow Plancherel). Ondelettes.

Definition [5.I.1] Une **frame** est une famille de fonctions $(\psi_n)_{n \in \mathbb{N}}$ telle que $\exists A, B > 0 : \forall f \in H$ on a

$$A \cdot \sum_{n \in \mathbb{N}} |\lambda_n|^2 \leq \|f\|_2^2 \leq B \cdot \sum_{n \in \mathbb{N}} |\lambda_n|^2$$

où $\lambda_n = \langle \psi_n, f \rangle$.

Remarque [5.I.2] Une frame (répère oblique) n'est pas nécessairement libre. Si elle l'est, on parle de **base de Riesz**

Si la famille est liée on dit aussi qu'elle est redondante. On utilise des représentations en effectuant des opérations simples sur les coefficients. On peut modifier la fonction en multipliant chaque coefficient par un facteur a_n dépendant ou non de λ_n . Dans un certain nombre d'applications telles que la compression ou le débruitage, on conserve certains coefficients et on met les autres à zéro.

De manière générale il est souvent intéressant de trouver des représentations, c'est à dire des familles de fonctions $(\psi_i)_{i \in \mathbb{N}}$ et de calculer les coefficients qui assurent une représentation compacte de la fonction.

II Approximation linéaire

On se place dans le cas où $(\psi_i)_{i \in \mathbb{N}}$ est une BON. Pour réaliser une approximation d'une fonction on peut ne conserver que certains coefficients λ_i . Si $f = \sum_{n \in \mathbb{N}} \lambda_n \psi_n$, on note

$$f_I = \sum_{i \in I} \lambda_i \psi_i.$$

Si I est défini indépendamment de f , on dit que f_I est une **approximation linéaire** de f .

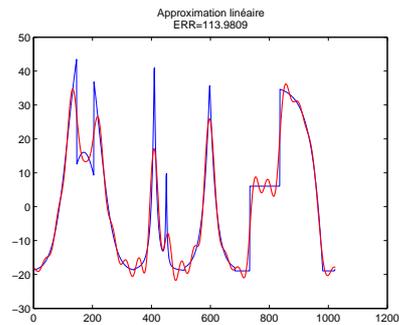
Exemple [5.II.1] Si f est définie sur N points, f_M est définie par $f_M(n) = \begin{cases} f(n) & \text{si } n \leq M \\ 0 & \text{sinon} \end{cases}$. f_M est une approximation linéaire dans la base de Diracs.

L'approximation obtenue en gardant les M plus basses fréquences est aussi une approximation linéaire.

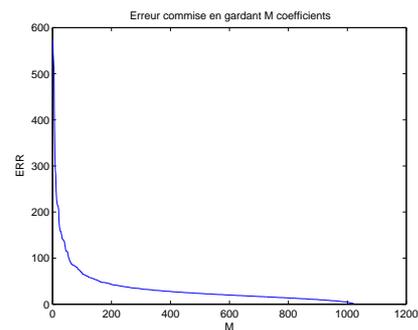
L'erreur commise par une telle approximation est

$$\|f - f_I\|_2^2 = \left\| \sum_{j \notin I} \lambda_j \psi_j \right\|_2^2 \stackrel{\text{BON}}{=} \sum_{j \notin I} |\lambda_j|^2.$$

Si on peut estimer $\sum_{j \notin I} |\lambda_j|^2$, on peut estimer l'erreur commise. Si on considère le cas classique où $I = \{i \leq M\}$ correspond aux M premiers indices, $\sum_{j \notin I} |\lambda_j|^2$ dépend de la décroissance des coefficients.



Approximation linéaire d'une fonction régulière par morceaux avec $M = 50$ coefficients de Fourier de basse fréquence ci-dessus et l'erreur commise en omettant M coefficients de Fourier de basse fréquence pour cette fonction ci-dessous.



Théorème [5.II.2] $\forall s > \frac{1}{2}, \exists A, B > 0$ tels que, si $\sum_{m=0}^{\infty} |m|^{2s} |\lambda_m|^2 < +\infty$, alors

$$\epsilon_I[M] = \|f - f_N\|^2 = \mathbf{o}(M^{-2s}).$$

En particulier, s'il existe c et ϵ tels que $\forall m, |\lambda_m| < \frac{c}{|m|^{s+\frac{1}{2}+\epsilon}}$, alors

$$\epsilon_I[M] = \mathbf{o}(M^{-2s}).$$

Proposition [5.II.3] Si $f \in L^2[0; 1]$ est à support inclus dans $]0; 1[$ et que f est n fois continûment différentiable, alors – si on note f_M l'approximation de f définie par $f_M(t) = \sum_{|m| < \frac{M}{2}} \langle f(u), e^{2i\pi mu} \rangle e^{2i\pi mt} - \forall \varepsilon > 0,$

$$\varepsilon_l[M] = \|f - f_M\|^2 = \mathbf{o}(M^{-2n-1+\varepsilon}).$$

Cette décroissance de l'erreur d'approximation linéaire est en fait directement liée à la régularité de la fonction.

Ainsi, si $f = c \cdot \mathbb{1}_{[0; \frac{1}{2}]}$,

$$\left| \langle f(u), e^{2i\pi mu} \rangle \right| = \begin{cases} 0 & \text{si } n \neq 0 \text{ est pair} \\ \frac{c}{\pi|m|} & \text{si } n \text{ est impair} \end{cases}$$

et donc $\varepsilon_l[M] \sim c^2 M^{-1}$.

Ainsi, si f est discontinue on ne peut espérer une décroissance de l'erreur d'approximation linéaire meilleure que $\varepsilon_l[M] = \mathbf{O}(M^{-1})$.

Nous verrons plus tard qu'on a un résultat similaire en base d'ondelettes.

III L'approximation non-linéaire

Si on veut minimiser l'erreur commise par l'approximation dans une BON il n'est pas toujours intéressant de considérer les premiers coefficients de la décomposition.

Ainsi, si l'erreur commise $\|f - f_M\|^2 = \sum_{j \notin M} |\lambda_j|^2$ et que l'on veut rendre l'erreur aussi petite que possible, on choisira l'ensemble I correspondant aux plus grands coefficients. Ainsi, à un entier M on associe l'ensemble des indices I_M correspondant aux plus grands produits scalaires en valeurs absolues. Les vecteurs $(\psi_i)_{i \in I}$ sont les vecteurs les plus corrélés à f .

L'erreur d'une approximation non-linéaire

$$\varepsilon_n[M] = \|f - f_{I_M}\|^2$$

est donc plus petite que l'erreur d'approximation linéaire obtenue en choisissant les M vecteurs d'approximation indépendamment de f .

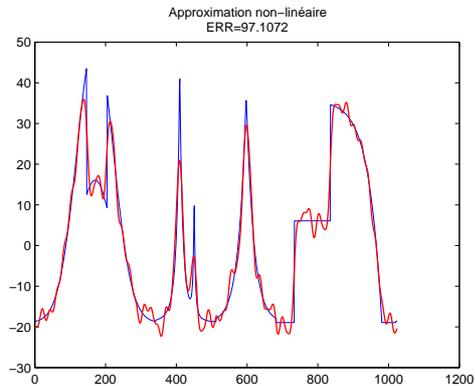
Si on classe les coefficients $|\langle f, \psi_i \rangle| = |\lambda_i|$ par ordre décroissant et qu'on note $f_B^r[k] = \langle f, \psi_{n_k} \rangle = \lambda_{n_k}$ le coefficient de rang k , on a

$$|f_B^r[k]| \geq |f_B^r[k+1]| \quad \forall k > 0.$$

L'approximation non-linéaire optimale est alors

$$f_{I_M} = \sum_{k=1}^M f_B^r[k] \psi_{n_k}.$$

Cette approximation peut s'obtenir en appliquant une fonction de seuillage. (Exemple : Dans MATLAB avec `fs = f .* (abs(f) > T)`).



Approximation non-linéaire d'une fonction régulière par morceaux en gardant les $M = 50$ plus grands (en valeur absolue) coefficients de Fourier.

Théorème [5.III.1] Si $s > \frac{1}{2}$ et si $\exists c > 0$ tels que $|f_B^r[k]| \leq \frac{c}{k^s}$, alors

$$\varepsilon_n[M] \leq \frac{c^2}{2s-1} M^{1-2s}. \tag{5.1}$$

Si $\varepsilon_n[M]$ vérifie (5.1) alors

$$|f_B^r[k]| \leq \left(1 - \frac{1}{2s}\right)^{-s} c k^{-s}.$$

Proposition [5.III.2] Si f a un nombre fini de discontinuités sur $[0; 1]$ et est uniformément lipschitz α entre ses discontinuités et si \mathcal{B} est une base d'ondelettes ayant $q > \alpha$ moments nuls alors

$$\varepsilon_n[M] = \mathbf{O}(M^{2\alpha}).$$

Remarque [5.III.3] L'exemple de la fonction indicatrice montre qu'on ne peut rien espérer de tel dans la base de Fourier où $\varepsilon_l[M] \sim \varepsilon_n[M]$.

IV Estimation dans du bruit

IV.1 Bayésien -VS- minimax

Un signal $f[n]$ de taille N est contaminé par l'addition d'un bruit. Le bruit est modélisé par la réalisation d'un processus aléatoire $W[n]$ dont la distribution de probabilité est connue. Les données mesurées sont

$$X[n] = f[n] + W[n].$$

Le signal est estimé en transformant les données bruitées X à l'aide d'un opérateur de décision D . L'estimation résultante est $\tilde{F} = DX$.

L'objectif est de minimiser l'erreur d'estimation mesurée à l'aide d'une fonction de coût.

Par simplicité en mesure cette erreur par la distance euclidienne.

Le risque de l'estimation \tilde{F} de F est calculé en moyenne par rapport à la distribution du bruit W :

$$r(D, f) = E(\|f - DX\|_2^2)$$

où E est l'espérance probabiliste.

L'optimisation de l'opérateur de décision D dépend de l'information *a priori* disponible sur le signal.

Si on dispose de l'ensemble Θ où vit la fonction f ainsi que de la loi de probabilité «d'apparition de la fonction f » dans l'ensemble Θ , on peut effectuer une approche **Bayésienne** : Dans cette approche, f est considérée comme la réalisation d'un processus aléatoire F dont la probabilité π est connue *a priori*. Dans ces cas-là :

$$X[n] = F[n] + W[n].$$

On suppose que F est indépendante de W . A partir des lois de F et W , on détermine la probabilité conditionnelle de F connaissant X , appelée distribution postérieure. C'est cette distribution postérieure qui permet de construire un opérateur de décision D dont le risque est mesuré par

$$r(D, \pi) = E_\pi(r(D, F)).$$

On cherchera un opérateur D qui minimise ce risque.

Problème : On connaît rarement la loi π !

Si on dispose uniquement de l'ensemble Θ , on se place dans le cadre *minimax*. On cherche à estimer $f \in \Theta$ à partir des données bruitées $X[n] = f[n] + W[n]$. Le risque de l'estimateur $\tilde{F} = DX$ est mesuré relativement à la distribution du bruit :

$$r(D, f) = E(\|DX - f\|^2).$$

Pour contrôler le risque sur Θ on essaie de minimiser le risque maximum

$$r(D, \Theta) = \sup_{f \in \Theta} E(\|DX - f\|^2).$$

Le risque minimax est la borne inférieure calculée sur tous les estimateurs linéaires et non-linéaires :

$$r_n(\Theta) = \inf_D r(D, \Theta).$$

On va chercher des estimateurs "simples" dont le risque est aussi proche que possible du risque minimax.

IV.2 Estimateurs diagonaux et seuillages

Si on décompose des données bruitées $X = f + W$ sur une BON \mathcal{B} , on obtient

$$X_n = \langle X, b_n \rangle = \langle f, b_n \rangle + \langle W, b_n \rangle =: f_n + W_n, \quad b_n \in \mathcal{B}.$$

Si W est un bruit blanc gaussien de moyenne nulle et de variance σ^2 , $\forall n, m \in \mathbb{N}$, W_n et W_m sont des réalisations d'un bruit blanc gaussien de moyenne nulle et de variance σ^2 . De plus, W_n et W_m sont indépendants si $n \neq m$. Le bruit blanc gaussien est bruit blanc gaussien dans toutes les BON.

Un estimateur diagonal estime *indépendamment* chaque f_m avec des fonctions d_m de X_m :

$$\tilde{F} = DX = \sum_{m=0}^{N-1} d_m(X_m)b_m.$$

On considère une classe de signaux centrés en 0 et on pose $D(0) = 0$ et par conséquent $d_m(0) = 0$ et on peut donc écrire $d_m(X_m) = a_m X_m$ où a_m dépend de X_m . L'opérateur est linéaire quand a_m est une constante indépendante de X_m .

On veut choisir a_m pour minimiser $r(D, f) = E(\|f - \tilde{F}\|^2) = \sum_{m=0}^{N-1} E(|f_m - a_m X_m|^2)$.

Or, $E(|f_m - a_m X_m|^2) = |f_m|^2 (1 - a_m)^2 + \sigma^2 a_m^2$. Ce risque est minimal pour $a_m = \frac{f_m^2}{f_m^2 + \sigma^2}$ et le risque vaut

$$\begin{aligned} r_{\text{inf}}(f) &= E(\|f - \tilde{F}\|^2) \\ &= \sum_{m=0}^{N-1} \frac{f_m^2 \sigma^2}{f_m^2 + \sigma^2}. \end{aligned}$$

En pratique, le facteur d'atténuation a_m ne peut être calculé car il dépend de f_m que l'on ne connaît pas. $r_{\text{inf}}(f)$ est donc une borne inférieure qui n'est pas atteignable.

On dit que ce risque a été obtenu avec un *oracle*.

Projection par oracle

Le projecteur non-linéaire qui minimise le risque $E(|f_m - a_m X_m|^2)$ est défini par

$$a_m = \begin{cases} 1 & \text{si } |f_m| \geq \sigma \\ 0 & \text{si } |f_m| < \sigma \end{cases}$$

Comme a_m dépend de f_m , cet estimateur ne peut être implémenté. Il utilise également un oracle. Le risque de cet estimateur est

$$\begin{aligned} r_p(f) &= E(\|f - \tilde{F}\|^2) \\ &= \sum_{m=0}^{N-1} \min(|f_m|^2, \sigma^2). \end{aligned}$$

Rappel : $\forall x, y \in \mathbb{R}^+$, $\min(x, y) \geq \frac{xy}{x+y} \geq \frac{1}{2} \min(x, y)$.

On a

$$r_p(f) \geq r_{\text{inf}}(f) \geq \frac{1}{2} r_p(f).$$

Soit M le nombre de coefficients tels que $|f_m| > \sigma$. L'approximation non-linéaire de f par ses M vecteurs de plus forte amplitude est

$$f_M = \sum_{|f_m| > \sigma} f_m b_m.$$

(NB : f_M est une fonction, f_m un coefficient !)

L'erreur est

$$\varepsilon_m[M] = \|f - f_M\|^2 = \sum_{|f_m| < \sigma} |f_m|^2.$$

Le risque de la projection oracle peut se réécrire

$$\begin{aligned} r_p(f) &= \sum_{m=0}^{N-1} \min(|f_m|^2, \sigma^2) \\ &= M\sigma^2 + \sum_{|f_m| < \sigma} |f_m|^2 \\ &= M\sigma^2 + \varepsilon_m[M]. \end{aligned}$$

Proposition [5.IV.1] Si $\varepsilon_m[M] \sim c^2 M^{1-2s}$, alors

$$r_p(f) \sim c^{\frac{1}{s}} \sigma^{2-\frac{1}{s}}.$$

Seuillage dur

Un estimateur par seuillage dur s'implémente par

$$d_m(x) = \rho_T(x) = \begin{cases} x & \text{si } |x| \geq T \\ 0 & \text{si } |x| < T \end{cases}.$$

L'opérateur D associé est alors un projecteur non-linéaire. Le risque de ce seuillage est

$$\begin{aligned} r_t(f) &= r(D, f) \\ &= \sum_{m=0}^{N-1} E(|f_m - \rho_T(X_m)|^2) \end{aligned}$$

$$\text{où } |f_m - \rho_T(X_m)|^2 = \begin{cases} |W_m|^2 & \text{si } |X_m| \geq T \\ |f_m|^2 & \text{si } |X_m| < T \end{cases}.$$

Le risque de ce projecteur est supérieur au risque du projecteur par oracle :

$$r_t(f) \geq r_p(f) = \sum_{m=0}^{N-1} \min(|f_m|^2, \sigma^2).$$

Théorème [5.IV.2] (Donoho-Johnstone) Soit $T = \sigma\sqrt{2 \ln N}$. Le risque d'un seuillage dur vérifie $\forall N \geq 4$

$$r_t(f) \leq (2 \ln N + 1)(\sigma^2 + r_p(f))$$

où N est la taille du signal.

Le seuillage en pratique

- (a) En pratique on utilise un seuil inférieur à $\sigma\sqrt{2 \ln_2 N}$, par exemple un peu supérieur à 3σ .
- (b) Si la base qu'on utilise n'est pas invariante par translation, on moyennera les signaux seuillés dans les bases translattées.

p.ex. base
d'ondelettes.
bruit reste
evlmt. ds
une base, ms
p-ê pas ds
une autre!
bruit pt ê
divisé par 3
ou 4!

Transformée en ondelettes orthogonales

Introduction

Les ondelettes et les bases d'ondelettes orthonormées en particulier sont le fruit de la rencontre des ingénieurs et des mathématiciens. La première base d'ondelettes (base de Haar) date de 1910. C'est le développement des données numériques et leur traitement par des ingénieurs astucieux d'un côté et le concept mathématique d'analyse multi-échelle d'un autre côté qui font naître les ondelettes.

Il existe une infinité de bases d'ondelettes.

Les ondelettes sont des fonctions définies sur \mathbb{R} qui forment une base orthonormée de $L^2(\mathbb{R})$ pour le produit scalaire usuel :

$$\langle f, g \rangle = \int_{-\infty}^{+\infty} f(t) \overline{g(t)} dt.$$

On note $\mathcal{B} = \{\psi_i\}_{i \in \mathbb{N}}$. Une famille est orthonormée si et seulement si $\langle \psi_i, \psi_j \rangle = \delta_{ij}$.

Les ondelettes forment des bases de $L^2(\mathbb{R})$.

$$\iff \forall f \in L^2(\mathbb{R}) \exists a_n \in \ell^2(\mathbb{R}) \text{ tels que } \forall \varepsilon > 0 \exists N(\varepsilon) \in \mathbb{N} \text{ tel que } \left\| f - \sum_{n=0}^{N(\varepsilon)} a_n \psi_n \right\|_2 \leq \varepsilon.$$

$$\iff \forall f \in L^2(\mathbb{R}) \exists a_n \in \ell^2(\mathbb{R}) \text{ tels que } \lim_{N \rightarrow \infty} \left\| f - \sum_{n=0}^N a_n \psi_n \right\|_2 = 0.$$

On écrira $f = \sum_{n=0}^{\infty} a_n \psi_n$.

Chacune de ces bases contient un nombre infini d'ondelettes. Les bases d'ondelettes ont de nombreux points en commun :

- Elles permettent une analyse multi-échelle du signal (des fonctions) et toutes les fonctions de la base sont obtenues par translation et dilatation d'une «ondelette mère». Cette ondelette mère définit entièrement la base d'ondelettes.
- Chaque base est associée à un algorithme discret (et rapide) de transformée en ondelettes des signaux digitaux.
- La plupart d'entre elles sont «localisées» à la fois en temps *et* en fréquence ! Elles assurent une représentation compacte de fonctions irrégulières par morceaux. (→ un plus par rapport à Fourier !)

I La transformée de Haar

On note V_j l'espace des fonctions de $L^2(\mathbb{R})$ telles que $\forall f \in V_j, f$ est constante sur les intervalles $[2^j k, 2^j(k+1)[\forall k \in \mathbb{Z}$.

Par définition de V_j ,

- $\forall j \in \mathbb{Z}, V_{j+1} \subseteq V_j$. (emboîtement)
- $\forall j \in \mathbb{Z} : f(t) \in V_j \iff f(\frac{t}{2}) \in V_{j+1}$.

– $\forall j, k \in \mathbb{Z} : f(t) \in V_j \iff f(t - 2^j k) \in V_j$.

On dit que V_j est un espace d'approximation à l'échelle de 2^j . En effet, toute fonction f de $L^2(\mathbb{R})$ peut être approchée par une fonction $f_j \in V_j$ en prenant la moyenne de f sur les intervalle $[2^j k, 2^j(k+1)[$. f_j est la projection de f sur V_j .

$$f_j := P_{V_j}(f) := \arg \min_{g \in V_j} \|f - g\|_2^2.$$

On dit que $\bigcup_{j \in \mathbb{Z}} V_j$ est *dense* dans $L^2(\mathbb{R})$, c'est à dire

$$\forall f \in L^2(\mathbb{R}) \forall \varepsilon > 0 \exists j \in \mathbb{Z} \text{ tel que } \|f - P_{V_j}(f)\|_2^2 \leq \varepsilon.$$

On a

$$V_j = \left\langle \left\{ \phi_{j,k} = \frac{1}{\sqrt{2^j}} \mathbb{1}_{[2^j k, 2^j(k+1)[}(x) \mid k \in \mathbb{Z} \right\} \right\rangle.$$

V_j est ainsi engendré par des fonctions $\phi_{j,k}$, toutes translatées d'une fonction $\phi_{j,0} = \frac{1}{\sqrt{2^j}} \mathbb{1}_{[0, 2^j[}(x)$. Toutes les fonction $\phi_j := \phi_{j,0}$ sont elles-même obtenues par dilatation d'une fonction d'un facteur 2^j d'une même fonction

$$\phi(x) = \mathbb{1}_{[0,1[}(x).$$

Les fonctions $\phi_{j,k}$ sont appelées fonctions d'échelle, la fonction ϕ est appelée fonction d'échelle mère.

Les fonction $\phi_{j,k}$ sont les briques de base permettant de construire une approximation d'une fonction $f \in L^2(\mathbb{R})$ à l'échelle j . Comme les fonctions d'échelle sont orthogonales à l'intérieure d'une même échelle, la projection de f sur V_j se calcule simplement :

$$P_{V_j}(f) = \sum_{k \in \mathbb{Z}} \langle f, \phi_{j,k} \rangle \phi_{j,k}$$

avec $\langle f, \phi_{j,k} \rangle = \int_{-\infty}^{+\infty} f(t) \phi_{j,k}(t) dt = \frac{1}{\sqrt{2^j}} \int_{2^j k}^{2^j(k+1)} f(t) dt.$

Les différentes fonctions f_j fournissent des approximations plus ou moins détaillées de f ; en particulier, f_{j-1} contient plus de détails que f_j car on peut reconstruire f_j directement à partir de f_{j-1} .

La famille $\{\phi_{j,k}\}_{(j,k) \in \mathbb{Z}^2}$ fournit une représentation complète mais redondante des fonctions de $L^2(\mathbb{R})$.

Pour obtenir une meilleure représentation des fonctions de $L^2(\mathbb{R})$ et de V_j , on introduit la notion d'espace de détails :

Comme pour tout j , $V_j \subseteq V_{j-1}$, on peut définir W_j par

$$V_{j-1} = V_j \oplus^\perp W_j.$$

$W_j \perp W_{j-1}$ W_j est l'espace de détails à l'échelle 2^j ; W_j est le complémentaire orthogonal de V_j dans V_{j-1} .

W_j contient les informations qui permettent de passer d'une approximation f_j de f à une approximation f_{j-1} plus fine.

W_j est un espace des fonctions

- constantes sur les intervalles de la forme $[2^{j-1}k, 2^{j-1}(k+1)[$ car $W_j \subseteq V_{j-1}$
- et de moyenne nulle sur les intervalle $[2^j k, 2^j(k+1)[$; en effet, $\forall \psi \in W_j$ et $\forall \phi_{j,k} \in V_j$, $\langle \psi, \phi_{j,k} \rangle = 0$.

$$\implies \text{Or, } \int_{-\infty}^{+\infty} \psi(x) \phi_{j,k}(x) dx = \frac{1}{\sqrt{2^j}} \int_{2^j k}^{2^j(k+1)} \psi(x) dx.$$

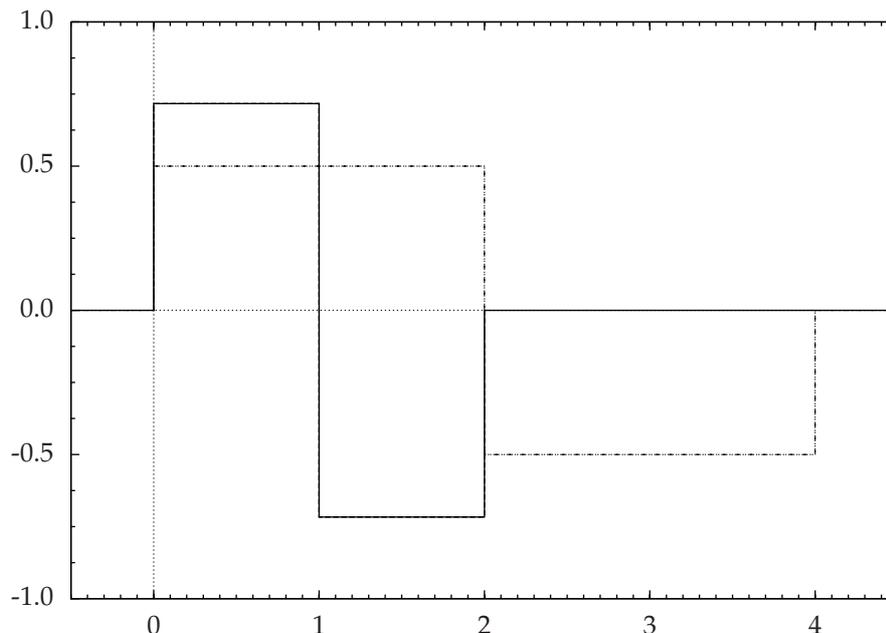
L'espace W_j est engendré par des fonctions $\psi_{j,k}$ appelées *ondelettes* : $W_j = \left\langle \left\{ \psi_{j,k} \mid k \in \mathbb{Z} \right\} \right\rangle$ où

$$\psi_{j,k} = \begin{cases} \frac{1}{\sqrt{2^j}} & \text{si } x \in [2^{j-1}k, 2^{j-1}(k+1)[\\ -\frac{1}{\sqrt{2^j}} & \text{si } x \in [2^{j-1}(k+1), 2^{j-1}(k+2)[\end{cases}$$

Notons au passage que $\forall k, j \in \mathbb{Z}$

$$\phi_{j,k} = \frac{1}{\sqrt{2}} (\phi_{j-1,2k} + \phi_{j-1,2k+1}) \quad \text{et} \quad (6.1)$$

$$\psi_{j,k} = \frac{1}{\sqrt{2}} (\phi_{j-1,2k} - \phi_{j-1,2k+1}) \quad (6.2)$$


 FIG. 6.1 – $\psi_{0,1}$ [—] et $\psi_{1,1}$ [---]

pour les ondelettes de Haar.

Les ondelettes sont obtenues par translation et dilatation d'une ondelette mère :

$$\phi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \phi\left(\frac{x-2^j k}{2^j}\right) \quad (6.3)$$

$$\psi_{j,k}(x) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{x-2^j k}{2^j}\right). \quad (6.4)$$

Les $(\psi_{j,k})_{k \in \mathbb{Z}}$ forment une BON de l'espace de détails W_j :

$$P_{W_j}(f) = \sum_{k \in \mathbb{Z}} \langle f, \psi_{j,k} \rangle \psi_{j,k}.$$

Remarquons que pour tout j entier négatif,

$$\begin{aligned} V_{j-1} &= W_j \oplus^\perp V_j \\ &= W_j \oplus^\perp W_{j+1} \oplus^\perp V_{j+1} \\ &= W_j \oplus^\perp W_{j+1} \oplus^\perp W_{j+1} \oplus^\perp V_{j+2} \\ &= \dots \\ &= \bigoplus_{j \leq l \leq 0}^\perp W_l \oplus^\perp V_0. \end{aligned}$$

Un élément de V_{j-1} peut ainsi s'écrire comme somme d'une approximation grossière sur V_0 et de détails de différente taille correspondant aux espaces W_l , $j \leq l \leq 0$.

$$\forall f \in L^2(\mathbb{R}), P_{V_j}(f) = P_{V_0}(f) + \sum_{l=j+1}^0 P_{W_l}(f).$$

Comme la projection de f sur W_l est $P_{W_l}(f) = \sum_{k \in \mathbb{Z}} \langle f, \psi_{l,k} \rangle \psi_{l,k}$, on obtient que

$$P_{W_j}(f) = P_{V_0}(f) + \sum_{\substack{k \in \mathbb{Z} \\ j+1 \leq l \leq 0}} \langle f, \psi_{l,k} \rangle \psi_{l,k}.$$

Les ondelettes fournissent une représentation multi-échelle d'une fonction.

La base de Fourier décompose une fonction en ses différentes composantes fréquentielles ; une base d'ondelettes décompose une fonction en détails à différentes échelles.

Les fonctions constantes par morceaux admettent un grand nombre de coefficients nuls dans une base d'ondelettes de Haar.

A chaque échelle, le nombre de coefficients non-nuls est au plus égal au nombre de discontinuités.

Les fonctions continues peuvent s'approcher par des fonctions constantes par morceaux et ont donc un "petit nombre" de coefficients de Haar élevé. Cette base concentre ainsi l'énergie de la fonction sur un petit nombre de coefficients ce qui est intéressant pour la compression ou pour le débruitage. (cf. figure 6.2, page 29)

I.1 Les ondelettes de Haar dans un cadre discret sur un intervalle

Un signal digital (x_1, \dots, x_N) suite de nombre peut être considéré comme un échantillon d'une fonction continue. On peut aussi considérer la suite comme une moyenne locale du signal :

$$x_k = \frac{1}{\tau} \int_{k\tau}^{(k+1)\tau} f(x) dx.$$

Ce modèle est raisonnable pour l'image où la valeur d'un pixel correspond à un nombre de photons captés par une cellule photoélectrique. On peut ainsi voir la suite (x_1, \dots, x_N) comme les coefficients de la projection d'une fonction f sur un espace V_j :

$$P_{V_j}(f) = \sum_{k=1}^N x_k \phi_{j,k} = \sum_{k=1}^N \langle f, \phi_{j,k} \rangle \phi_{j,k}$$

où $N = 2^j$.

On peut donc calculer les coefficients de cette fonction de $V_j = V_0 \oplus^\perp \bigoplus_{j+1 \leq l \leq 0}^\perp W_l$ dans la base d'ondelettes associée, c'est à dire calculer les $a_{k,l}$ tels que

$$\sum_{i=1}^N x_i \phi_{i,k} = \sum_k b_{0,k} \phi_{0,k} + \sum_{j+1 \leq l \leq 0} a_{l,k} \psi_{l,k}.$$

Comme on travaille toujours sur des signaux finis, on périodise le signal observé en dehors du support du signal pour simplifier les calculs. On prendra toujours une période T dyadique ($T = 2^{-j_0}$, $-j \in \mathbb{N}$) pour simplifier l'algorithme.

On observe alors qu'au dessus d'une échelle j notée $j = 0$, $\phi_{l,0}(x) = \text{const} \forall l \leq 0$.

On a ainsi une seule fonction d'échelle à échelle maximale qui correspond à la moyenne de la fonction.

On a :

- 1 fonction d'échelle à l'échelle 0.
- 2 fonctions d'échelle à l'échelle -1.
- 4 fonctions d'échelle à l'échelle -2.
- ...
- 2^{-j} fonctions d'échelle à l'échelle j .

Ainsi la transformée en ondelettes d'un signal de 2^{-j_0} éléments, correspondant à 2^{-j_0} coefficients de fonctions d'échelle, comprend 2^{-j_0} éléments.

- 1 coefficient de fonction d'échelle à l'échelle 0.
- 1 coefficient d'ondelettes à l'échelle 0.
- 2 coefficients d'ondelettes à l'échelle -1.
- 4 coefficients d'ondelettes à l'échelle -2.
- ...
- 2^{-j_0-1} coefficients d'ondelettes à l'échelle $j_0 + 1$.

approx. non-lin.!

"chargement de l'image" du haut en bas, défiler

"chargement de l'image" plan par plan, construction itérative

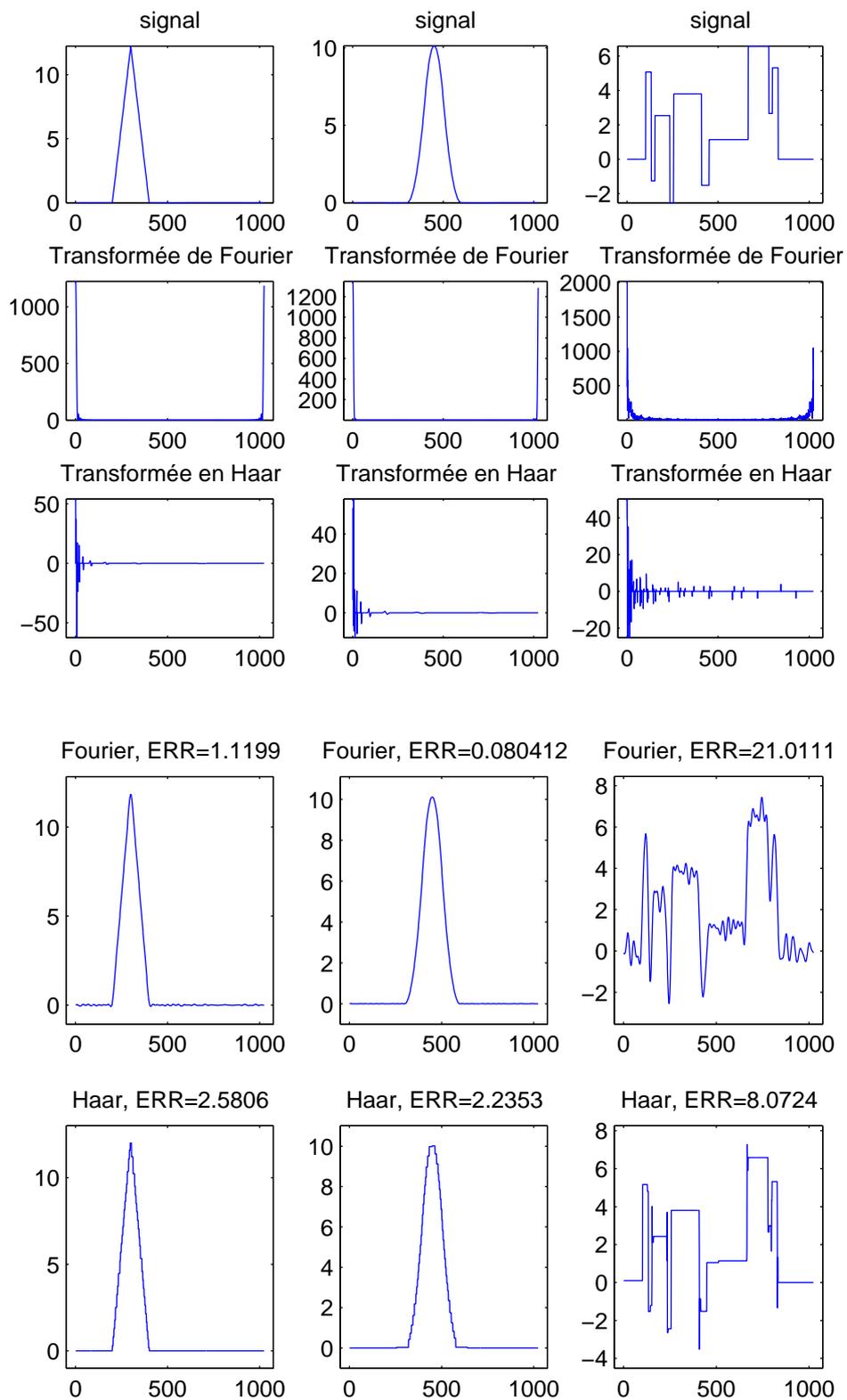


FIG. 6.2 – Trois signaux et leurs transformées en Fourier et en ondelettes de Haar, puis leur reconstructions non-linéaires avec $M = 64$ coefficients.

I.2 Calcul des coefficients

On dispose de $f = \sum_{k=0}^{N-1} x_k \phi_{j_0, k}$, $N = 2^{-j_0}$.

f est un élément de $V_{j_0} = V_{j_0+1} \oplus^\perp W_{j_0+1}$; on va l'écrire comme la somme d'un élément de V_{j_0+1} et d'un élément de W_{j_0+1} :

$$\begin{aligned} f &= x_0 \phi_{j_0, 0} + x_1 \phi_{j_0, 1} + x_2 \phi_{j_0, 2} + x_3 \phi_{j_0, 3} + \dots + x_{N-2} \phi_{j_0, N-2} + x_{N-1} \phi_{j_0, N-1} \\ &= \frac{x_0 + x_1}{2} (\phi_{j_0, 0} + \phi_{j_0, 1}) + \frac{x_0 - x_1}{2} (\phi_{j_0, 1} - \phi_{j_0, 0}) + \frac{x_2 + x_3}{2} (\phi_{j_0, 2} + \phi_{j_0, 3}) + \frac{x_2 - x_3}{2} (\phi_{j_0, 2} - \phi_{j_0, 3}) + \dots + \frac{x_{N-2} + x_{N-1}}{2} (\phi_{j_0, N-2} + \phi_{j_0, N-1}) + \frac{x_{N-2} - x_{N-1}}{2} (\phi_{j_0, N-2} - \phi_{j_0, N-1}) \\ &\stackrel{(6.1), (6.2)}{=} \frac{x_0 + x_1}{\sqrt{2}} \phi_{j_0-1, 0} + \frac{x_0 - x_1}{\sqrt{2}} \phi_{j_0-1, 1} + \frac{x_2 + x_3}{\sqrt{2}} \phi_{j_0-1, 2} + \frac{x_2 - x_3}{\sqrt{2}} \phi_{j_0-1, 3} + \dots + \frac{x_{N-2} + x_{N-1}}{\sqrt{2}} \phi_{j_0-1, N/2-1} + \frac{x_{N-2} - x_{N-1}}{\sqrt{2}} \phi_{j_0-1, N/2-1} \end{aligned}$$

Algorithme [6.I.1] (Algorithme de Mallat, Transformée en ondelettes rapide) On note h le filtre périodique ayant deux composantes non-nulles :

$$h[N] = h[0] = \frac{1}{\sqrt{2}}, \quad h[N-1] = \frac{1}{\sqrt{2}}$$

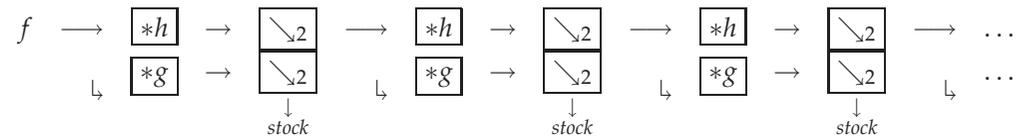
et g le filtre

$$g[N] = g[0] = \frac{1}{\sqrt{2}}, \quad g[N-1] = -\frac{1}{\sqrt{2}}.$$

On a

$$\begin{aligned} f * h &= \frac{1}{\sqrt{2}} (x_0 + x_1, x_1 + x_2, x_2 + x_3, \dots, x_{N-1} + x_N, x_N + x_1), \\ f * g &= \frac{1}{\sqrt{2}} (x_0 - x_1, x_1 - x_2, x_2 - x_3, \dots, x_{N-1} - x_N, x_N - x_1). \end{aligned}$$

Le calcul de la transformée en ondelettes peut être représenté grâce au schéma suivant :



La reconstruction se fait quasiment de la même manière : Au lieu de sous-échantillonner (indiqué par \searrow_2), on intercale des 0 entre les données stockées, on fait la convolution avec les filtres, on additionne. Puis on recommence jusqu'à ce qu'on arrive à la fonction reconstruite.

II Les autres bases d'ondelettes

Les bases de Haar sont adaptées pour concentrer l'énergie des fonctions constantes par morceaux. Pour des fonctions qui ont une régularité locale plus grande, on a construit d'autres bases d'ondelettes. Toutes les bases orthogonales d'ondelettes sont construites à partir d'une ondelette mère :

$$\forall j, k : \psi_{j,l}(x) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{x - 2^j k}{2^j} \right).$$

On définit de même les espaces de l'approximation V_j et les espaces de détails W_j :

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

Comme $V_{j+1} \subset V_j$, on a que

$$\begin{aligned} \phi_{j+1,k} &= \sum_l h_{l,k} \phi_{j,l} \\ &= \sum_l h[2k - l] \phi_{j,l} \end{aligned}$$

et comme $W_{j+1} \subset V_j$,

$$\psi_{j+1,l} = \sum_l g[2k - l] \phi_{j,l}.$$

Les coefficients $h[l]$ et $g[l]$ varient selon la base d'ondelettes. À chaque base est associé un couple (h, g) . Ces filtres sont ceux qui interviennent dans l'algorithme de Mallat [6.I.1], page 30. Ce sont eux qui déterminent la transformée en ondelettes. Quand on modifie les propriétés des filtres h et g , on modifie les propriétés de la base d'ondelettes.

Théorème [6.II.1] (Théorème de Mallat-Meyer) Si $\hat{h}(\omega)$ est 2π -périodique et continûment différentiable dans un voisinage de 0 et que

- (1) $\hat{h}(0) = \sqrt{2}$,
- (2) $\forall \omega \in \mathbb{R}, \left| \hat{h}(\omega) \right|^2 + \left| \hat{h}(\omega + \pi) \right|^2 = 2$ et
- (3) $\inf \left\{ \left| \hat{h}(\omega) \right| \mid \omega \in \left[-\frac{\pi}{2}; \frac{\pi}{2} \right] \right\} > 0$,

alors

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

est la transformée de Fourier d'une fonction d'échelle $\phi \in L^2(\mathbb{R})$.

Soit g le filtre défini par

$$\hat{g}(\omega) = e^{-i\omega} \overline{\hat{h}(\omega + \pi)}.$$

L'ondelette ψ définie par

$$\hat{\psi}(\omega) = \frac{1}{\sqrt{2}} \hat{g}\left(\frac{\omega}{2}\right) \hat{\phi}\left(\frac{\omega}{2}\right)$$

engendre une base d'ondelettes de $L^2(\mathbb{R})$:

$$\left\{ \psi_{j,k}(t) = \frac{1}{\sqrt{2^j}} \psi \left(\frac{t - 2^j k}{2^j} \right) \mid j, k \in \mathbb{Z} \right\}.$$

Definition [6.II.2] Une fonction f a $n + 1$ moments nuls, $n \in \mathbb{N} \cup \{0\}$, si et seulement si

$$\int x^k f(x) dx = 0 \quad \forall k \leq n.$$

Théorème [6.II.3] (Théorème d'approximation) Si f est C^α par morceaux et si ψ est une ondelette avec $p > \alpha + 1$ moments nuls, alors l'approximation f_M de f calculée avec les M plus grands coefficients d'ondelettes vérifient

$$\|f - f_M\|^2 \leq cM^{-\alpha}.$$

valable pour
les signaux
1D!!