Patrick Flandrin
CNRS & École Normale Supérieure de Lyon

"Data first", ou comment piloter l'analyse par les données

M2 de Physique — Cours 2016

http://perso.ens-lyon.fr/patrick.flandrin/enseignement.html

Table des matières

1 Introduction **4**

2 Fourier, filtrage et adaptativité **5**
2.1 Fourier 5
2.1.1 Temps continu et fréquence continue 5
2.1.2 Temps continu et fréquence discrète 6
2.1.3 Temps discret et fréquence continue 7
2.1.4 Temps discret et fréquence discrète 7
2.2 Filtrage linéaire 9
2.2.1 Définition 9
2.2.2 Interprétation 10
2.2.3 Relations entrée-sortie 11
2.3 Filtrage adapté 13
2.3.1 Rapport signal-sur-bruit 14
2.3.2 Contraste 17
2.4 Filtrage optimal 20
2.4.1 Principe d'orthogonalité 20
2.4.2 Filtrage de Wiener 20
2.4.3 Filtrage inverse 22
2.5 Filtrage adaptatif 25
2.5.1 Principe 25
2.5.2 Algorithme du gradient 27
2.5.3 Algorithme LMS 30
2.5.4 Deux exemples 31

3 Temps-fréquence, ondelettes et adaptativité **33**
3.1 Incertitude 33
3.2 Décompositions linéaires 35
3.3 Ondelettes 39
3.3.1 Incertitude "adaptée" 39
3.3.2 Transformée continue 41
3.3.3 Transformées discrètes 46
3.3.4 Quelques base d'ondelettes 55
3.4 Sélection et parcimonie 58
3.4.1 Paquets d'ondelettes et cosinus locaux 58
3.4.2 Reconstruction seuillée 59

3.4.3 Poursuite adaptative 61
3.5 Distributions quadratiques 62
3.6 Distributions dépendantes du signal 69
3.6.1 Wigner-Ville comme TFCT adaptée 70
3.6.2 Géométrie 70
3.6.3 Noyaux optimaux 71
3.6.4 Diffusion adaptative 72
3.7 Méthodes exploitant la phase du signal 75
3.7.1 Réallocation 75
3.7.2 "Synchrosqueezing" 78
3.8 Données substituts 79
3.8.1 De l'importance de la phase 79
3.8.2 Randomisation simple 80
3.8.3 Test de stationnarité 82
3.8.4 Randomisation sous contrainte 84

4 Structures propres et adaptativité **87**
4.1 Karhunen-Loève 87
4.1.1 Principe 87
4.1.2 Quelques propriétés 89
4.1.3 Exemples 92
4.2 "Singular Spectrum Analysis" 94
4.2.1 Principe 94
4.2.2 Algorithme 94
4.2.3 Quelques propriétés 94
4.2.4 Exemples 94
4.3 "Empirical Mode Decomposition" 94
4.3.1 Principe 94
4.3.2 Algorithme 95
4.3.3 Quelques propriétés 96
4.3.4 Exemples 97
4.3.5 Variations 100

1 Introduction

L'analyse de données, en incluant dans ce terme les disciplines associées comme le traitement du signal ou l'analyse des séries temporelles, joue un rôle central dans la compréhension du monde qui nous entoure, que celui-ci soit physique, biologique ou même social. Dès lors que l'accès à l'information passe par l'observation, disposer de données fiables et de méthodes efficaces pour leur analyse et leur traitement est un point de passage obligé, que ce soit en analyse exploratoire ou en modélisation.

Jusqu'à un passé récent, les données étaient des denrées le plus souvent rares et précieuses, fruits d'expérimentations coûteuses, et difficiles à stocker autant qu'à gérer lorsqu'elles étaient plus nombreuses. La situation est aujourd'hui très différente, avec un déluge de données ("big data") dû autant à la disponibilité de capteurs en tous genres, déployables facilement et bon marché, qu'à une ouverture à grande échelle ("open data") permise par les capacités nouvelles des machines et l'explosion des réseaux de communication. Il en résulte une prolifération de données qui s'accompagne d'une mise en évidence de variabilités jusqu'alors inaccessibles, rendant d'une certaine façon vain l'espoir de pouvoir disposer de méthodes "universelles" permettant de faire face à toutes les situations. Ce changement de situation quant aux données conduit à un changement de point de vue sur leur traitement, suggérant de reconsidérer les approches anciennes — dont le mode opératoire est essentiellement d'appliquer des techniques existantes à des observations — en replaçant les données au centre du jeu et en faisant en sorte qu'elles puissent "piloter" leur propre analyse.

L'ambition étant de trouver un équilibre entre une trop grande universalité (au caractère opératoire alors nécessairement limité) et une trop grande spécificité (au risque d'une trop faible capacité de généralisation), il ne s'agit bien sûr pas de tout oublier des méthodes "classiques". On s'attachera dans ce cours à voir comment, en s'appuyant sur de tels acquis (dont on rappellera les principes), il est possible d'aller au-delà en exploitant les degrés de liberté offerts au bénéfice d'un pilotage par les données, tout en présentant quelques méthodes spécifiquement dédiées à cet objectif.

2 Fourier, filtrage et adaptativité

2.1 Fourier

Parmi les différentes transformations possibles, celle de Fourier joue un rôle prépondérant. Introduite il y a un peu plus de 200 ans, elle fait aujourd'hui partie de la "boîte à outils" de presque tous les champs scientifiques et est le point de départ de nombreuses méthodes plus récentes. C'est par elle que l'on commencera.

2.1.1 Temps continu et fréquence continue

Soit un signal à temps continu $x(t)$, défini sur la droite réelle $\mathbb{R}$. On définit sa *transformée de Fourier* $\{X(f), f \in \mathbb{R}\}$ par la relation

$$X(f) = \int_{-\infty}^{+\infty} x(t)\, e^{-i2\pi ft}\, dt.$$

Ceci définit le *spectre* d'un signal, quantité à valeurs complexes fonction de la *fréquence* f. Pour les signaux absolument sommables, on peut retrouver la forme d'onde à partir du spectre grâce à la formule d'inversion

$$x(t) = \int_{-\infty}^{+\infty} X(f)\, e^{i2\pi ft}\, df,$$

ce que l'on peut encore écrire formellement

$$x(t) = \int_{-\infty}^{+\infty} \langle x, e_f \rangle\, e_f(t)\, df$$

en introduisant le produit scalaire

$$\langle x, y \rangle = \int_{-\infty}^{+\infty} x(t)\, y^*(t)\, dt$$

et la notation

$$e_f(t) = e^{i2\pi ft}.$$

L'interprétation en est que la transformation de Fourier réalise une projection sur des ondes monochromatiques, le signal analysé s'expliquant comme superposition de telles ondes, avec des poids s'identifiant aux valeurs du

spectre obtenues par projection. Les ondes monochromatiques jouent ainsi un rôle d'*atomes* pour la décomposition d'un signal.

Pour les signaux de carré sommable, la transformation de Fourier est une isométrie (relation de Plancherel) :

$$\int_{-\infty}^{+\infty} x(t)\, y^*(t)\, dt = \int_{-\infty}^{+\infty} X(f)\, Y^*(f)\, df,$$

avec la conséquence que (formule de Parseval) :

$$\int_{-\infty}^{+\infty} |x(t)|^2\, dt = \int_{-\infty}^{+\infty} |X(f)|^2\, df.$$

Cette quantité intégrée correspondant à l'*énergie* E_x du signal, on en déduit que les intégrands $|x(t)|^2$ et $|X(f)|^2$ s'identifient à des densités d'énergie appelées respectivement *puissance instantanée* et *densité spectrale d'énergie*.

2.1.2 Temps continu et fréquence discrète

Soit un signal à temps continu $x(t)$, défini sur l'intervallle $[-T/2, +T/2]$ et supposé périodique de période T. On vérifie alors que les fonctions

$$\{\psi_k(t) := e^{i2\pi kt/T}, k \in \mathbb{Z}\}$$

sont orthogonales sur $[-T/2, +T/2]$ dans la mesure où

$$\begin{aligned}
\langle \psi_k, \psi_l \rangle &= \int_{-T/2}^{+T/2} e^{i2\pi(k-l)t/T}\, dt \\
&= \left[\frac{T}{i2\pi(k-l)}\, e^{i2\pi(k-l)t/T} \right]_{-T/2}^{+T/2} \\
&= \frac{T}{\pi(k-l)}\, \frac{1}{2i}\, \left[e^{i\pi(k-l)} - e^{-i\pi(k-l)} \right] \\
&= \frac{1}{T}\, \frac{\sin \pi(k-l)}{\pi(k-l)} \\
&= \frac{1}{T}\, \delta_{k,l}
\end{aligned}$$

car $k-l \in \mathbb{Z}$. Il s'ensuit que $x(t)$ admet la décomposition (en *série de Fourier*)

$$x(t) = \sum_{k \in \mathbb{Z}} X_k\, e^{i2\pi kt/T},$$

avec

$$X_k = \frac{1}{T} \int_{-T/2}^{+T/2} x(t)\, e^{-i2\pi kt/T}\, dt.$$

2.1.3 Temps discret et fréquence continue

Le spectre $X(f)$ d'un signal à temps discret étant périodique, il est possible de le développer en série de Fourier. Supposons que le spectre "fondamental", associé au signal non-échantillonné, occupe la bande $[-B/2, +B/2]$ et que l'échantillonnage soit critique, on a la décomposition

$$X(f) = \sum_{k\in\mathbb{Z}} x_k\, e^{i2\pi kf/B},$$

avec

$$\begin{aligned} x_k &= \frac{1}{B} \int_{-B/2}^{+B/2} X(f)\, e^{-i2\pi kf/B}\, df \\ &= \frac{1}{B}\, x\left(-\frac{k}{B}\right). \end{aligned}$$

Si l'on adopte la convention selon laquelle l'échantillonnage critique ($T_e = 1/B$) est arbitrairement pris comme unité, on a $B = 1$ et, en écrivant $x[n] = x(nT_e)$ lorsque $T_e = 1$, on est conduit à la représentation

$$X(f) = \sum_n x[n]\, e^{-i2\pi nf}, |f| \leq 1/2$$

dans laquelle les échantillons $x[n]$ sont définis par

$$x[n] = \int_{-1/2}^{+1/2} X(f)\, e^{i2\pi nf}\, df.$$

Cette paire d'équations constitue une transformation de Fourier des signaux à temps discret.

2.1.4 Temps discret et fréquence discrète

Considérons maintenant un signal $x(t)$ de durée finie T. Par application du théorème d'échantillonnage dans le domaine des fréquences, son spectre

$X(f)$ est entièrement caractérisé par ses échantillons $X(k/T)$. Considérons alors l'échantillonnage temporel d'un tel signal avec une période T_e, ce qui fournit N échantillons $x[n] = x(nT_e)$ en supposant que $T = NT_e$. Il lui correspond un spectre périodique $X_e(f)$ de période $1/T_e$, mais celui-ci est encore caractérisé par ses échantillons

$$X[k] := X_e(k/T)$$

puisque, $x(t)$ étant de durée limitée, $x[n]$ l'est aussi. En tenant compte à la fois de la période $1/T_e$ de $X_e(f)$ et de la relation $T = NT_e$, on en déduit que le signal échantillonné, qui est constitué de N valeurs $x[n]$ en temps, est également décrit par N valeurs $X[k]$ en fréquence. Ces valeurs s'écrivent

$$X[k] = \int_{-\infty}^{+\infty} x_e(t)\, e^{-i2\pi kt/T}\, dt$$

et, en introduisant la définition du signal échantillonné en temps

$$x_e(t) = \sum_{n=0}^{N-1} x[n]\, \delta(t - nT_e),$$

on obtient la définition de la transformation de Fourier discrète ou série de Fourier discrète :

$$\begin{aligned} X[k] &= \int_{-\infty}^{+\infty} \sum_{n=0}^{N-1} x[n]\, \delta(t - nT_e)\, e^{-i2\pi kt/T}\, dt \\ &= \sum_{n=0}^{N-1} x[n] \int_{-\infty}^{+\infty} \delta(t - nT_e)\, e^{-i2\pi kt/T}\, dt \\ &= \sum_{n=0}^{N-1} x[n]\, e^{-i2\pi knT_e/T} \\ &= \sum_{n=0}^{N-1} x[n]\, e^{-i2\pi nk/N}. \end{aligned}$$

On en déduit que

$$\begin{aligned}\sum_{k=0}^{N-1} X[k]\, e^{i2\pi nk/N} &= \sum_{k=0}^{N-1}\sum_{m=0}^{N-1} x[m]\, e^{-i2\pi mk/N}\, e^{i2\pi nk/N} \\ &= \sum_{m=0}^{N-1} x[m] \sum_{k=0}^{N-1} e^{i2\pi(n-m)k/N} \\ &= \sum_{m=0}^{N-1} x[m]\, N\, \delta_{nm},\end{aligned}$$

d'où la formule d'inversion :

$$x[n] = \frac{1}{N}\sum_{k=0}^{N-1} X[k]\, e^{i2\pi nk/N}.$$

Il importe de noter que, si les $x[n]$ correspondent de façon exacte aux échantillons de $x(t)$ en temps, les $X[k]$ sont en fréquence des échantillons de $X_e(f)$ et ne sont donc qu'une approximation de ceux de $X(f)$. En effet, $x(t)$ étant supposé à durée limitée, $X(n)$ ne peut être à bande strictement limitée, d'où un phénomène inévitable de repliement spectral.

2.2 Filtrage linéaire

Au-delà de sa capacité à représenter une fonction dans l'espace de ses fréquences associées (on l'a présentée sous l'aspect de la fréquence temporelle associée à la variable "temps", mais elle s'applique de la même manière pour d'autres variables comme celle d'"espace"), la transformation de Fourier joue un rôle particulièrement important du fait des relations étroites qu'elles tisse avec d'autres concepts comme celui, par exemple, de *filtrage linéaire*.

2.2.1 Définition

D'une manière générale, un système linéaire transforme une entrée $x(t)$ en une sortie $y(t)$ selon une relation intégrale de la forme

$$y(t) = \int_{-\infty}^{+\infty} h(t,s)\, x(s)\, ds.$$

Un tel opérateur linéaire est appelé *filtre linéaire* s'il est en outre covariant vis-à-vis des translations temporelles, c'est-à-dire si la filtrée d'une entrée décalée est la décalée de la filtrée. Ceci implique que

$$\begin{aligned}\int_{-\infty}^{+\infty} h(t-\tau,s)\,x(s)\,ds &= \int_{-\infty}^{+\infty} h(t,s)\,x(s-\tau)\,ds \\ &= \int_{-\infty}^{+\infty} h(t,s+\tau)\,x(s)\,ds.\end{aligned}$$

Cette égalité devant être vérifiée pour tout signal, on a nécessairement, pour tout s, $h(t,s+\tau) = h(t-\tau,s)$. En particulier, le choix de la valeur $s=0$ montre que $h(t,\tau) = h(t-\tau,0)$, ce qui veut dire que le noyau d'un filtre linéaire est de la forme $h(t,s) = h_0(t-s)$.

2.2.2 Interprétation

La relation liant la sortie d'un filtre linéaire à son entrée est ainsi une convolution :

$$y(t) = \int_{-\infty}^{+\infty} h(t-s)\,x(s)\,ds$$

et la fonction $h(t)$ est appelée la *réponse impulsionnelle* du filtre du fait que

$$x(t) = \delta(t) \Rightarrow y(t) = \int_{-\infty}^{+\infty} h(t-s)\,\delta(s)\,ds = h(t).$$

Par transformation de Fourier, il est immédiat de voir que la relation entrée-sortie du filtre devient multiplicative dans le domaine fréquentiel :

$$Y(f) = H(f)\,X(f),$$

la transformée de Fourier $H(f)$ de la réponse impulsionnelle $h(t)$ étant appelée la *fonction de transfert* ou le *gain complexe* du filtre. Celui-ci mesure, en module et en phase, la modification introduite par le filtre sur le spectre du signal d'entrée. La relation multiplicative en fréquence permet une évaluation particulièrement simple de l'influence d'une cascade de filtres dans une chaîne : les gains (réels) se multiplient et les phases s'ajoutent.

Dans le cas où l'entrée d'un filtre linéaire est une exponentielle complexe de fréquence f_0, la sortie vaut alors

$$\int_{-\infty}^{+\infty} h(t-s)\,e^{i2\pi f_0 s}\,ds = H(f_0)\,e^{i2\pi f_0 t},$$

ce qui n'est autre que l'entrée, à un facteur multiplicatif près. On voit ainsi que les exponentielles complexes sont *fonctions propres* des filtres linéaires, la valeur propre associée étant le gain complexe à la fréquence correspondante. Par suite, la transformation de Fourier, qui décompose un signal sur la base de telles exponentielles complexes, est naturellement adaptée aux transformations dans les filtres linéaires.

2.2.3 Relations entrée-sortie

Dans le cas des signaux aléatoires, les densités spectrales de puissance se transforment également de façon simple dans un filtrage linéaire. Si l'on suppose que l'entrée d'un filtre de réponse impulsionnelle $h(t)$ est un signal aléatoire stationnaire $x(t)$ de moyenne m_x et de fonction de corrélation $\gamma_x(\tau)$, il est facile de voir que la moyenne m_y de la sortie vaut :

$$\begin{aligned} m_y &:= \mathbb{E}\left\{\int_{-\infty}^{+\infty} h(t-s)\,x(s)\,ds\right\} \\ &= \int_{-\infty}^{+\infty} h(t-s)\,\mathbb{E}\{x(s)\}\,ds \\ &= m_x \int_{-\infty}^{+\infty} h(t-s)\,ds \\ &= m_x\,H(0). \end{aligned}$$

Passant au second ordre, on montre (en supposant pour simplifier que $m_x = 0$) que le filtrage linaire préserve la stationnarité et que la fonction de corrélation $\gamma_y(\tau)$ de la sortie du filtre s'écrit

$$\begin{aligned} \gamma_y(\tau) &= \mathbb{E}\{y(t)y(t-\tau)\} \\ &= \mathbb{E}\left\{\iint_{-\infty}^{+\infty} h(t-u)\,h(t-\tau-v)\,x(u)\,x(v)\,du\,dv\right\} \\ &= \int_{-\infty}^{+\infty} \gamma_h(\theta-\tau)\,\gamma_x(\theta)\,d\theta, \end{aligned}$$

en notant

$$\gamma_h(\tau) := \int_{-\infty}^{+\infty} h(t)\,h(t-\tau)\,dt$$

la fonction de corrélation *déterministe* de la réponse $h(t)$ du filtre.

Par suite, en faisant usage de l'identité de Parseval-Plancherel assurant la conservation du produit scalaire par transformation de Fourier, on peut écrire de manière équivalente :

$$\gamma_x(\tau) = \int_{-\infty}^{+\infty} |H(f)|^2 \, \Gamma_x(f) \, e^{i2\pi f\tau} \, df,$$

d'où l'on déduit la relation fondamentale de transformation des densités spectrales :

$$\Gamma_y(f) = |H(f)|^2 \, \Gamma_x(f).$$

Ce résultat se généralise au cas de deux filtrages en parallèle d'un même signal par deux filtres différents :

$$\begin{aligned} y_1(t) &= \int_{-\infty}^{+\infty} h_1(t-s) \, x(s) \, ds; \\ y_2(t) &= \int_{-\infty}^{+\infty} h_2(t-s) \, x(s) \, ds, \end{aligned}$$

pour lequel un calcul analogue conduit au résultat :

$$\Gamma y_1, y_2(f) = H_1(f) \, H_2^*(f) \, \Gamma_x(f),$$

appelé parfois *formule des interférences*.

On peut tirer de cette expression deux conséquences principales :

1. Si $h_1(t)$ et $h_2(t)$ sont deux filtres *spectralement disjoints*, on a identiquement $H_1(f) \, H_2^*(f) = 0$ et l'interspectre (et donc aussi l'intercorrélation) de leurs sorties est nulle : ils ne partagent pas d'énergie (ce point rejoint la remarque faite précédemment sur la décorrélation des contributions spectrales d'un même signal à des fréquences différentes).
2. Si deux signaux sont issus d'un même troisième par filtrage linéaire, alors leur fonction de cohérence (lorsqu'elle est définie) est toujours maximale :

$$\begin{aligned} C_{y_1,y_2}(f) &= \frac{|\Gamma_{y_1,y_2}(f)|}{\sqrt{\Gamma_{y_1}(f) \, \Gamma_{y_2}(f)}} \\ &= \frac{|H_1(f) \, H_2^*(f) \, \Gamma_x(f)|}{\sqrt{|H_1(f)|^2 \, \Gamma_x(f) \, |H_2(f)|^2 \, \Gamma_x(f)}} \\ &= 1. \end{aligned}$$

Dans les cas où $h_1(t) = \delta(t)$ et où $h_2(t) = h(t)$, c'est-à-dire lorsqu'un signal se déduit d'un autre par filtrage linéaire, on voit que leur cohérence est telle que $C_{x,y}(f) = C_{x,h\star x}(f) = 1$. L'examen de la fonction de cohérence est ainsi une façon de tester, fréquence par fréquence, l'existence d'une relation de filtrage linéaire entre deux signaux.

On notera enfin que le filtrage linéaire préserve aussi la *gaussiannité*. Un filtrage linéaire revenant essentiellement à sommer les entrées, on peut se convaincre de ce résultat en se restreignant à l'analyse de la somme de deux variables aléatoires conjointement gaussiennes, que l'on peut considérer comme les valeurs de l'entrée à deux instants ($x_1 := x(t_1)$ et $x_2 := x(t_2)$). La fonction de répartition de la somme $z = x_1 + x_2$ est donnée par

$$F_Z(z) = \Pr\left\{x_1 + x_2 \leq z\right\},$$

ce qui définit dans le plan (x_1, x_2) un domaine D_z borné supérieurement par la droite $x_1 + x_2 = z$ et permet d'écrire :

$$F_Z(z) = \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x_2} p_{\mathbf{X}}(x_1, x_2)\, dx_1\, dx_2.$$

On en déduit par dérivation par rapport à z que la densité de probabilité associée a pour valeur

$$p_Z(z) = \int_{-\infty}^{+\infty} p_{\mathbf{X}}(z - x, x)\, dx$$

et, dans le cas conjointement gaussien (supposé centré, mais éventuellement corrélé) pour lequel

$$p_{\mathbf{X}}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \exp\left[-\frac{1}{2(1-r^2)}\left(\frac{x_1^2}{\sigma_1^2} - 2r\frac{x_1 x_2}{\sigma_1\sigma_2} + \frac{x_2^2}{\sigma_2^2}\right)\right],$$

il est facile de voir que l'intégrale précédente ne met en jeu que l'exponentielle d'un polynôme de degré 2 en x, ce qui assure au résultat une structure de même nature en z, caractéristique de la gaussiannité.

2.3 Filtrage adapté

Le problème de la détection optimale s'attache à décider si un signal, dont tout ou partie des caractéristiques sont connues, est présent ou non dans une observation donnée. Il ne prend de sens que lorsque l'observation en question est bruitée et si l'on dispose d'un minimum d'informations *a priori* sur les propriétés statistiques du bruit perturbateur.

2.3.1 Rapport signal-sur-bruit

Une première approche au problème de la détection peut se formuler en termes de *filtrage*, l'idée étant de "faire sortir" le signal du bruit (s'il est présent). D'une manière plus précise, on suppose que l'on dispose d'une observation de la forme

$$y(t) = s(t) + b(t),$$

où $s(t)$ est le signal, supposé connu, dont on souhaite tester la présence éventuelle et $b(t)$ est le bruit d'observation, que l'on supposera stationnaire, centré, blanc et de variance connue γ_0 :

$$\mathbb{E}\{b(t)b(s)\} = \gamma_0\,\delta(t-s)$$

de telle sorte que, sans traitement, la limite de détection est *a priori* fixée par la *valeur-crête* du signal à détecter.

Afin de quantifier la possibilité de "voir" le signal émerger du bruit après traitement, on peut introduire une mesure de *contraste* entre l'hypothèse H_0 où l'observation est formée du bruit seul et l'hypothèse H_1 où elle résulte du mélange signal + bruit :

$$d(z) := \frac{|\mathbb{E}\{z(t)|H_1\} - \mathbb{E}\{z(t)|H_0\}|}{\sqrt{\mathrm{var}\{z(t)|H_0\}}},$$

expression dans laquelle on suppose que l'on a filtré l'observation au moyen d'un filtre linéaire de réponse impulsionnelle $h(t)$.

La quantité d'intérêt devenant :

$$z(t) = \int_{-\infty}^{+\infty} h(t-u)\,y(u)\,du,$$

on peut alors se poser la question de savoir si le choix judicieux d'un tel filtre permettrait d'obtenir un contraste maximum entre les deux hypothèses H_0 et H_1. En raisonnant de façon instantanée, on a $\mathbb{E}\{z(t)|H_0\} = 0$ et

$$\begin{aligned}
\mathbb{E}\{z^2(t)|H_0\} &= \mathbb{E}\left\{\iint_{-\infty}^{+\infty} h(t-u)\,h(t-v)\,b(u)\,b(v)\,du\,dv\right\} \\
&= \iint_{-\infty}^{+\infty} h(t-u)\,h(t-v)\,\gamma_0\,\delta(u-v)\,du\,dv \\
&= \gamma_0 \int_{-\infty}^{+\infty} h^2(t-u)\,du \\
&= \gamma_0\,E_h,
\end{aligned}$$

en notant E_h l'énergie du filtre de réponse $h(t)$, d'où il suit directement que

$$\mathrm{var}\{z(t)|H_0\} = \gamma_0\, E_h.$$

Comme, par ailleurs,

$$\begin{aligned}\mathbb{E}\{z(t)|H_1\} &= \mathbb{E}\left\{\int_{-\infty}^{+\infty} h(t-u)\,[s(u)+b(u)]\,du\right\}\\ &= \int_{-\infty}^{+\infty} h(t-u)\,s(u)\,du,\end{aligned}$$

car le signal $s(t)$ est supposé certain et le bruit $b(t)$ centré, on obtient

$$d^2(z) = \frac{\left|\int_{-\infty}^{+\infty} h(t-u)\,s(u)\,du\right|^2}{\gamma_0\, E_h}.$$

Le contraste ainsi obtenu peut alors être maximisé en faisant usage de l'inégalité de Cauchy-Schwarz selon laquelle

$$\left|\int_{-\infty}^{+\infty} h(t-u)\,s(u)\,du\right|^2 \leq \int_{-\infty}^{+\infty} h^2(t-u)\,du.\int_{-\infty}^{+\infty} s^2(u)\,du = E_h\,E_s,$$

d'où le résultat central :

$$d(z) \leq \sqrt{\frac{E_s}{\gamma_0}}.$$

On voit ainsi que ce contraste maximal après traitement fait intervenir une grandeur *intégrée* (l'énergie) et non pas *instantanée* (la puissance-crête). Il y a ainsi un gain potentiel évident dans la mesure où, pour une même valeur-crête, une augmentation de la durée du signal suffira à améliorer sa possibilité de détection. À titre d'exemple, on peut prendre pour $s(t)$ la fonction :

$$s(t) = a\,\mathbf{1}_{[0,T]}(t)$$

pour laquelle

$$d(z) \leq \frac{a\sqrt{T}}{\sqrt{\gamma_0}}.$$

Ceci met en évidence le rôle conjoint joué par la durée et l'amplitude, cette dernière pouvant rester inférieure au niveau de bruit (on ne "voit" pas

le signal qui est "noyé" dans le bruit) sans que cela nuise à sa détection à condition qu'elle soit maintenue suffisamment longtemps.

Si l'on revient au cas général, la valeur maximale du contraste est atteinte dès lors que les conditions pour garantir que l'inégalité de Cauchy-Schwarz devienne une égalité sont remplies, c'est-à-dire lorsqu'il y a colinéarité entre les facteurs du produit scalaire :

$$h(t-u) \propto s(u),$$

soit encore

$$h(u) \propto s(t-u).$$

Il faut voir cette relation comme relative à la variable de temps courante u de façon à garantir une valeur maximale au contraste en sortie du filtre à l'instant t. Si l'on renverse la perspective en considérant t comme la variable de temps courante, la réponse impulsionnelle $h(t)$ n'est autre que le signal à détecter *renversé dans le temps* :

$$h(t) = s(-t)$$

et on parle alors de *filtre adapté*.

Le filtrage adapté revient à effectuer l'inter-corrélation entre l'observation et le signal à détecter pusique l'on a

$$z(t) = \int_{-\infty}^{+\infty} y(u)\, s(u-t)\, du.$$

Remarque — D'un point de vue pratique, on considère en général le signal à détecter comme défini sur un support temporel limité. Si l'on convient de noter ce support $[0, T]$ et l'on souhaite opérer le filtrage *en ligne*, la détection ne pourra être effective qu'avec un retard (connu) lié à la durée T du signal. La réponse impulsionnelle du filtre adapté prend alors la forme

$$h(t) = s(T-t)$$

et sa sortie s'écrit

$$z(t) = \int_{t-T}^{t} y(u)\, s\left(u-(t-T)\right)\, du.$$

Du fait que le bruit additif est supposé centré, la sortie du filtre adapté se comporte en moyenne comme

$$\mathbb{E}\left\{z(t)|H_1\right\} = \gamma_s(t)$$

en notant γ_s la fonction d'auto-corrélation déterministe du signal, quantité qui est par définition maximale en $t = 0$ et prenant la valeur E_s à cette date. Si l'on raffine alors le modèle d'observation initial en introduisant un retard inconnu t_0 pour le signal à détecter :

$$y(t) = s(t - t_0) + b(t),$$

il vient immédiatement que

$$\mathbb{E}\left\{z(t)|H_1\right\} = \gamma_s(t - t_0)$$

et le problème de détection se double de la possibilité d'une *estimation* du retard selon

$$\hat{t}_0 = \arg\max_t z(t),$$

la question en suspens étant de déterminer dans quelle mesure la valeur maximale de la sortie du filtre adapté est significativement au-dessus du niveau des fluctuations du bruit.

2.3.2 Contraste

Avec la notion de filtrage adapté, on a fait deux hypothèses :

1. le traitement visant à maximiser le contraste est un *filtrage linéaire*, faisant du détecteur qui en résulte un détecteur *à structure imposée* (en l'occurrence, linéaire) ;
2. le signal à détecter est noyé dans un bruit *blanc*.

On peut alors se poser la question de relâcher ces deux hypothèses, en se contentant de dire que l'observation $y(t)$ est transformée en une quantité $z(t)$ au moyen d'une opération $\mathcal{S}$:

$$z = \mathcal{S}(y)$$

dont on ne spécifie pas la nature *a priori* (détecteur *à structure libre*), conduisant au contraste

$$d(\mathcal{S}(y)) = \frac{|\mathbb{E}\left\{\mathcal{S}(y)|H_1\right\} - \mathbb{E}\left\{\mathcal{S}(y)|H_0\right\}|}{\sqrt{\mathrm{var}\left\{\mathcal{S}(y)|H_0\right\}}}.$$

En remarquant que ce contraste est invariant par toute transformation affine, c'est-à-dire que

$$d(\lambda \mathcal{S}(y) + \mu) = d(\mathcal{S}(y))$$

pour tout λ et tout μ, on peut supposer que $\mathbb{E}\{\mathcal{S}y|H_0\} = 0$ sans perte de généralité, et donc réduire l'étude du contraste à celle de

$$d(\mathcal{S}(y)) = \frac{|\mathbb{E}\{\mathcal{S}(y)|H_1\}|}{\sqrt{\mathbb{E}\{\mathcal{S}^2(y)|H_0\}}}.$$

Si l'on exprime alors le numérateur de ce contraste en fonction des seules lois de probabilités des observations suivant l'une et l'autre des hypothèses, on obtient

$$\begin{aligned}
\mathbb{E}\{\mathcal{S}(y)|H_1\} &= \int_{-\infty}^{+\infty} \mathcal{S}(y)\, p_{Y|H_1}(y|H_1)\, dy \\
&= \int_{-\infty}^{+\infty} \mathcal{S}(y)\, \frac{p_{Y|H_1}(y|H_1)}{p_{Y|H_0}(y|H_0)}\, p_{Y|H_0}(y|H_0)\, dy \\
&= \mathbb{E}\{\mathcal{S}(y)\, \Lambda(y)|H_0\}
\end{aligned}$$

en introduisant la quantité

$$\Lambda(y) := \frac{p_{Y|H_1}(y|H_1)}{p_{Y|H_0}(y|H_0)}$$

appelé *rapport de vraisemblance* ("likelihood ratio" en anglais).

En appliquant alors l'inégalité de Cauchy-Schwarz, il vient

$$|\mathbb{E}\{\mathcal{S}(y)\, \Lambda(y)|H_0\}| \leq \sqrt{\mathbb{E}\{\Lambda^2(y)|H_0\}}\sqrt{\mathbb{E}\{\mathcal{S}^2(y)|H_0\}},$$

d'où l'on déduit que le contraste est maximisé selon

$$d(\mathcal{S}(y)) \leq \sqrt{\mathbb{E}\{\Lambda^2(y)|H_0\}},$$

la borne étant atteinte lorsque

$$\mathcal{S}(y) \propto \Lambda(y),$$

c'est-à-dire lorsque le détecteur est construit en calculant le rapport de vraisemblance.

Les deux approches, filtrage adapté et rapport de vraisemblance, proviennent d'hypothèses différentes mais qui ne sont pas nécessairement sans recouvrement, permettant de retrouver les résultats de la première à partir de certaines configurations de la seconde. Ainsi, si l'on se place pour simplifier dans un contexte discret permettant de calculer les densités de probabilité mises en jeu dans le rapport de vraisemblance, on peut montrer que le résultat essentiel du filtrage adapté, à savoir la structure de *corrélation* du détecteur, est la conséquent naturelle d'une hypothèse de gaussiannité des observations.

Supposons en effet que l'observation soit de la forme

$$\left\{ \begin{array}{lcl} H_0 : y_i & = & b_i \\ H_1 : y_i & = & b_i + s_i \end{array} \right.$$

pour $i = 1, \ldots, N$, avec s_i les échantillons du signal à détecter et b_i ceux du bruit supposés tels que $\mathbb{E}\{b_i\} = 0$ et $\mathbb{E}\{b_i\, b_j\} = \sigma^2\, \delta_{ij}$.

Si l'on suppose alors que les b_i sont aussi *gaussiens*, la décorrélation se prolonge en indépendance et le rapport de vraisemblance global se factorise selon

$$\Lambda(y_1, \ldots, y_N) = \prod_{i=1}^{N} \Lambda(y_i).$$

Comme l'on a

$$p_{Y|H_1}(y|H_1) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(y_i - s_i)^2 \right\}$$

et

$$p_{Y|H_1}(y|H_0) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2} y_i^2 \right\},$$

on en déduit que

$$\begin{array}{rcl} \Lambda(y_i) & = & \exp\left\{ -\dfrac{1}{2\sigma^2} \left[(y_i - s_i)^2 - y_i^2 \right] \right\} \\ & \propto & \exp\left\{ \dfrac{y_i\, s_i}{\sigma^2} \right\}, \end{array}$$

d'où

$$\log \Lambda(y) \propto \sum_{i=1}^{N} y_i\, s_i,$$

ce qui est explicitement la forme d'une corrélation (à retard nul) entre l'observation et le signal à détecter utilisé comme référence.

2.4 Filtrage optimal

2.4.1 Principe d'orthogonalité

Supposons que l'on dispose d'observations $y(t)$ et que l'on veuille élaborer à partir d'elles une estimée *linéaire* d'une quantité d'intérêt $d(t)$ (il peut s'agir d'un *filtrage simple* au cas où l'on aurait $y(t) = x(t) + b(t)$, avec $b(t)$ un bruit additif de caractéristiques statistiques connues, et l'on voudrait estimer $x(t)$, ou encore d'une *prédiction* si l'on veut inférer la valeur $y(t + \tau)$ à partir de la connaissance de l'observation jusqu'au seul instant courant t, etc.).

Dans le cas général, $d(t)$ n'appartient pas nécessairement à l'espace des solutions engendré linéairement par les observations. Il en résulte que, faute de pouvoir répondre exactement à la question posée, il faut se contenter de trouver une solution qui soit la plus proche en un sens donné de ce qui est cherché. Un critère naturel en ce sens est de minimiser l'*erreur quadratique moyenne* entre $d(t)$ et son estimée $\hat{d}(t)$:

$$\mathbb{E}\{[d(t) - \hat{d}(t)]^2\} \rightarrow \min .$$

En s'appuyant sur le fait que l'opérateur d'espérance mathématique définit un produit scalaire permettant d'accéder à une mesure de distance entre les quantités aléatoires considérées, une interprétation géométrique simple du problème de minimisation posé conduit à retenir pour solution l'estimée associée à la projection orthogonale de $d(t)$ sur l'espace des observations (cf. Figure 1).

C'est le *principe d'orthogonalité*, stipulant que la solution de filtrage linéaire optimal (à erreur quadratique moyenne minimale) est obtenue en rendant l'erreur orthogonale aux observations au sens du produit scalaire retenu, c'est-à-dire en rendant erreur et observations décorrélées :

$$\mathbb{E}\{[d(t) - \hat{d}(t)]y(v)\} = 0.$$

2.4.2 Filtrage de Wiener

Soit $d(t)$ la quantité "désirée" et $\hat{d}(t)$ son estimée, élaborée par filtrage linéaire à partir de l'observation $y(t)$:

$$\hat{d}(t) = \int_{-\infty}^{+\infty} h(t-u)\, y(u)\, du.$$

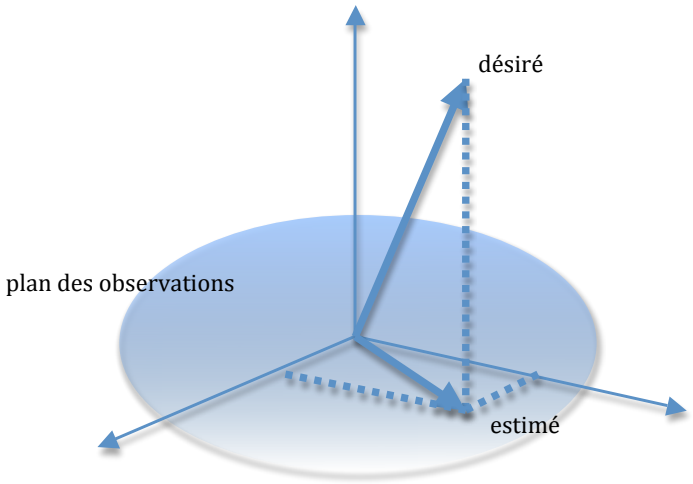

FIGURE 1 – Principe d'orthogonalité.

On appelle *filtre de Wiener* le filtre linéaire de réponse $h_*(t)$, optimal au sens où l'erreur associée à l'estimation, c'est-à-dire la différence

$$e(t) = d(t) - \hat{d}(t),$$

est telle que sa puissance est minimisée :

$$P = \mathbb{E}\{e^2(t)\} \to \min .$$

Par application du principe d'orthogonalité, la solution est obtenue lorsque cette erreur est orthogonale aux observations :

$$\mathbb{E}\{e(t)y(v)\} = 0,$$

d'où l'on déduit, en combinant cette équation et la forme choisie pour le filtre que l'on doit avoir [1] :

$$\gamma_{d,y}(\tau) = \int_{-\infty}^{+\infty} h(\tau - \theta)\, \gamma_y(\theta)\, d\theta.$$

1. On convient de noter $\gamma_{x,y}(\tau) := \mathbb{E}\{x(t)y(t-\tau)\}$ et d'adopter la notation simplifiée $\gamma_x(\tau) := \gamma_{x,x}(\tau)$.

On reconnaît une équation de convolution dans le domaine temporel, dont on sait qu'elle se transforme en produit dans l'espace de Fourier associé des fréquences, conduisant à l'expression du gain complexe du filtre optimal de Wiener donnée par :

$$H_*(f) = \frac{\Gamma_{d,y}(f)}{\Gamma_y(f)}.$$

Du fait que, par construction, l'estimée est orthogonale aux observations, l'erreur quadratique minimale P_* associée au filtre de Wiener se réduit à :

$$P_* = \mathbb{E}\{[d(t) - \hat{d}(t)]d(t)\},$$

soit encore :

$$P_* = \gamma_d(0) - \int_{-\infty}^{+\infty} h(\tau)\, \gamma_{d,y}(\tau)\, d\tau.$$

Passant dans le domaine des fréquences et utilisant la propriété d'isométrie de la transformation de Fourier, on peut ré-écrire cette dernière équation selon

$$P_* = \int_{-\infty}^{+\infty} \Gamma_d(f)\, df - \int_{-\infty}^{+\infty} H^*(f)\, \Gamma_{d,y}(f)\, df,$$

soit encore

$$P_* = \int_{-\infty}^{+\infty} \Gamma_d(f)\, \left[1 - C_{d,y}^2(f)\right]\, df,$$

où $C_{d,y}(f)$ est la fonction de cohérence définie précédemment. Dans le contexte de filtrage optimal qui nous intéresse ici, on voit donc qu'une cohérence unité a pour conséquence une puissance d'erreur identiquement nulle, ce qui est en accord avec le fait que la quantité désirée appartient à l'espace engendré linéairement à partir des observations.

2.4.3 Filtrage inverse

Un cas particulièrement important où le filtrage de Wiener peut trouver une application est celui du *filtrage inverse* dans lequel l'observation $y(t)$ est modélisée comme la version déformée d'un signal d'intérêt $x(t)$ par une *fonction d'appareil* $h(t)$ (supposée connue) à laquelle s'ajoute un bruit $b(t)$ que l'on supposera indépendant du signal, centré, stationnaire et de densité spectrale de puissance $\Gamma_b(f)$ connue :

$$y(t) = \int_{-\infty}^{+\infty} g(t-u)\, x(u)\, du + b(t).$$

Ce modèle étant posé, le filtrage inverse consiste à remonter au signal, c'est-à-dire à considérer

$$d(t) = x(t),$$

sur la base des observations $y(t)$ disponibles et des connaissances *a priori* quant à la fonction d'appareil $g(t)$ et au bruit $b(t)$. Pour construire la solution générale au problème, il faut évaluer dans un premier temps la corrélation croisée entre le signal à estimer et l'observation, ce qui donne :

$$\gamma_{d,y}(\tau) = \int_{-\infty}^{+\infty} g(\theta - \tau)\, \gamma_x(\theta)\, d\theta,$$

équation de corrélation dont la transformée de Fourier fournit directement :

$$\Gamma_{d,y}(f) = G^*(f)\Gamma_x(f).$$

Dans un deuxième temps, on calcule la densité spectrale de puissance de l'observation, ce qui se fait de manière immédiate en utilisant la relation d'entrée-sortie des filtres linéaires et l'indépendance entre signal et bruit :

$$\Gamma_y(f) = |G(f)|^2\Gamma_x(f) + \Gamma_b(f).$$

Utilisant ces deux quantités, on aboutit au résultat cherché qui s'écrit :

$$H_*(t) = \frac{G^*(f)\Gamma_x(f)}{|G(f)|^2\Gamma_x(f) + \Gamma_b(f)}$$

et peut encore se mettre sous la forme :

$$H_*(f) = \frac{1}{G(f)} \frac{1}{1 + \rho^{-1}(f)}$$

en introduisant la quantité :

$$\rho(f) := \frac{|G(f)|^2\Gamma_x(f)}{\Gamma_b(f)}$$

dont l'interprétation physique est celle d'un *rapport signal-sur-bruit*, local en fréquence.

Repartant de la forme explicite de la corrélation croisée entre $d(t) = x(t)$ et $y(t)$, on est conduit à :

$$\begin{aligned} \int_{-\infty}^{+\infty} h(\tau)\, \gamma_{d,y}(\tau)\, d\tau &= \iint_{-\infty}^{+\infty} h(\tau)\, g(\theta - \tau)\, \gamma_x(\theta)\, d\theta\, d\tau \\ &= \int_{-\infty}^{+\infty} H(f)\, G(f)\, \Gamma_x(f)\, df. \end{aligned}$$

Faisant usage de cette équation et remarquant que :

$$\gamma_x(0) = \int_{-\infty}^{+\infty} \Gamma_x(f)\, df,$$

on obtient :

$$P_* = \int_{-\infty}^{+\infty} \left(1 - \frac{|G(f)|^2 \Gamma_x(f)}{|G(f)|^2 \Gamma_x(f) + \Gamma_b(f)}\right) \Gamma_x(f)\, df,$$

expression que l'on peut simplifier en la ré-écrivant :

$$P_* = \int_{-\infty}^{+\infty} \frac{\Gamma_x(f)}{1 + \rho(f)}\, df.$$

On peut déduire des expressions générales précédentes plusieurs cas particuliers, qui en révêlent par exemple les situations limites.

Ainsi, dans le cas où le bruit devient négligeable, c'est-à-dire lorsque le rapport signal-sur-bruit $\rho(f)$ tend vers l'infini, on obtient comme attendu :

$$\lim_{\rho(f)\to\infty} H_*(f) = \frac{1}{G(f)}$$

et

$$\lim_{\rho(f)\to\infty} P_* = 0.$$

L'interprétation en est qu'en l'absence de bruit, la réponse fréquentielle du filtre inverse n'est autre que l'inverse de la fonction d'appareil, la connaissance supposée parfaite de cette dernière conduisant en outre à une erreur nulle.

À l'inverse, si l'on suppose maintenant que le bruit devient prépondérant au sens où le rapport signal-sur-bruit tend vers zéro, on obtient immédiatement que :

$$\lim_{\rho(f)\to 0} H_*(f) = 0$$

et

$$\lim_{\rho(f)\to 0} P_* = \gamma_x(0).$$

Dans ce cas où l'information utile portée par le signal est totalement noyée dans le bruit d'observation, l'effet additionnel de sa modification par l'appareil de mesure devient négligeable et la meilleure estimation se réduit à

la valeur moyenne du bruit (nulle par hypothèse), avec une erreur quadratique s'identifiant à la variance de ce bruit (valeur de la corrélation à l'origine).

On peut enfin noter que si l'on s'affranchit de l'effet d'une éventuelle fonction d'appareil et que l'on s'intéresse au seul *débruitage* d'un signal, on se place dans le cas simplifié :

$$G(f) = 1 \Rightarrow H_*(f) = \frac{\Gamma_x(f)}{\Gamma_x(f) + \Gamma_b(f)}$$

pour lequel les remarques précédentes relatives aux cas limites de rapports signal-sur-bruit nul ou infini continuent bien sûr à s'appliquer.

Dans le cas d'un signal inconnu dans un bruit inconnu, le problème est évidemment mal posé mais, si l'on dispose d'une "observation bruit seul", on peut ré-écrire la solution selon

$$H_*(f) = 1 - \frac{\Gamma_b(f)}{\Gamma_y(f)}$$

et construire une estimée de cette réponse en estimant d'une part la densité spectrale de puissance globale de l'observation et d'autre part celle du bruit à partir d'un intervalle temporel supposé "bruit seul".

2.5 Filtrage adaptatif

Le filtrage optimal au sens de Wiener repose sur deux hypothèses. La première est que les processus impliqués sont stationnaires et la seconde que les caractéristiques de second ordre sur lesquelles est construite la solution sont connues. En pratique cependant, on a souvent affaire à des situations évolutives et, de plus, il est rare que l'on en connaisse précisément les statistiques. L'objet du filtrage adaptatif est d'aborder ces questions en essayant de construire itérativement une solution sur la base de l'arrivée séquentielle d'observations, l'idée étant de converger vers la solution de Wiener dans le cas stationnaire ou de s'adapter à des évolutions en mettant à jour les estimations destinées à les suivre.

2.5.1 Principe

Afin de formuler le principe du filtrage adaptatif, il est utile de revenir au filtrage de Wiener, mais en se plaçant dans un cadre explicitement discret

et en en analysant la solution dans le domaine temporel. Pour ce faire on considère l'observation donnée par une séquence de valeurs $\{y[n], n \in \mathbb{Z}\}$ que l'on filtre par un filtre (de réponse impulsionnelle finie de longueur M, décrite par le vecteur $\mathbf{w}^T = [w_0, w_1, \dots w_{M-1}]$) pour obtenir en sortie les approximations $\hat{d}[n]$ de la grandeur désirée $d[n]$. On a ainsi

$$\hat{d}[n] = \sum_{k=0}^{M-1} w_k \, y[n-k],$$

l'erreur correspondante s'écrivant

$$e[n] = d[n] - \hat{d}[n].$$

En notant $\mathbf{y}[n]^T = [y[n], y[n-1], \dots y[n-M+1]]$ le vecteur d'état des observations au temps n, cette erreur peut se ré-écrire

$$e[n] = d[n] - \mathbf{w}^T \mathbf{y}[n].$$

Par suite, la puissance de l'erreur $P := \mathbb{E}\{e^2[n]\}$ a pour valeur

$$\begin{aligned} P &= \mathbb{E}\left\{d^2[n] - 2d[n]\mathbf{w}^T\mathbf{y}[n] + \mathbf{w}^T(\mathbf{y}[n]\mathbf{y}[n]^T)\mathbf{w}\right\} \\ &= \mathbb{E}\{d^2[n]\} - 2\mathbf{w}^T\mathbf{c}_{dy} + \mathbf{w}^T\mathbf{R}_{yy}\mathbf{w} \end{aligned}$$

en introduisant le vecteur d'inter-corrélation $\mathbf{c}_{dy} := \mathbb{E}\{d[n]\mathbf{y}[n]\}$ entre la sortie désirée et l'observation, et la matrice d'auto-corrélation $\mathbf{R}_{yy} := \mathbb{E}\{\mathbf{y}[n]\mathbf{y}[n]^T\}$ de cette même observation.

Le filtre optimal est celui qui minimise la puissance de l'erreur. Il doit donc être tel que

$$\nabla P = \mathbf{0}.$$

En dérivant composante par composante, on obtient

$$\frac{\partial}{\partial w_l} \sum_k w_k c_{dy,k} = \mathbf{c}_{dy} \Leftrightarrow \nabla \mathbf{w}^T \mathbf{c}_{dy} = \mathbf{c}_{dy}.$$

De façon analogue, on peut écrire

$$\frac{\partial}{\partial w_l} \sum_k \sum_{k'} w_k w_{k'} R_{yy,kk'} = \sum_{k'} w_{k'} R_{yy,lk'} + \sum_k w_k R_{yy,kl},$$

soit encore

$$\frac{\partial}{\partial w_l} \sum_k \sum_{k'} w_k w_{k'} R_{yy,kk'} = 2 \sum_k w_{k'} R_{yy,lk} \Rightarrow \nabla \mathbf{w}^T \mathbf{R}_{yy} \mathbf{w} = 2 \mathbf{R}_{yy} \mathbf{w}$$

en utilisant le fait que la matrice de corrélation $\mathbf{R}_{yy}$ est symétrique. Il suit que

$$\nabla P = 2 \left(\mathbf{R}_{yy} \mathbf{w} - \mathbf{c}_{dy} \right)$$

d'où la solution optimale

$$\mathbf{w}_* = \mathbf{R}_{yy}^{-1} \mathbf{c}_{dy},$$

la puissance minimale de l'erreur associée ayant pour valeur

$$P_* = \mathbb{E}\{d^2[n]\} - \mathbf{w}_*^T \mathbf{R}_{yy} \mathbf{w}_*.$$

2.5.2 Algorithme du gradient

Tel qu'il vient d'être établi, le filtre optimal de Wiener nécessite de résoudre le système de M équations à M inconnues

$$\mathbf{R}_{yy} \mathbf{w} = \mathbf{c}_{dy},$$

ce qui est très coûteux par une méthode directe ($O(M^3)$ opérations). Si l'on se souvient cependant que la solution optimale provient de la minimisation d'un coût quadratique dont le minimum est unique, on peut préférer approcher itérativement ce minimum par une descente de gradient. Le vecteur des coefficients du filtre devient ainsi lui-même une fonction du temps $\mathbf{w}[n]$ et, partant d'une solution initiale $\mathbf{w}[0]$, on opère itérativement en construisant la solution à l'instant $n+1$ à partir de celle à l'instant n en effectuant la mise à jour

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \frac{1}{2} \mu[n] \left. \nabla P \right|_{\mathbf{w}[n]}$$

dans laquelle la quantité $\mu[n]$ (pouvant être dépendante du temps) est le pas d'adaptation de l'algorithme.

Si l'on revient à l'expression du gradient de la puissance de l'erreur, on trouve que

$$\left. \nabla P \right|_{\mathbf{w}[n]} = 2 \mathbf{R}_{yy} \mathbf{w}[n] - 2 \mathbf{c}_{dy},$$

ce qui conduit à la forme explicite de l'algorithme

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu[n] \left(\mathbf{c}_{dy} - \mathbf{R}_{yy} \mathbf{w}[n] \right).$$

Le "bon" comportement de l'algorithme (convergence, si possible rapide, vers la solution exacte) dépendra à la fois de la structure de corrélation et du pas d'adaptation.

Afin d'étudier la convergence, il est commode d'introduire le vecteur d'écart $\mathbf{v}[n] := \mathbf{w}[n] - \mathbf{w}_*$ entre la valeur estimée à l'instant n et la valeur optimale (cible). Comme on sait que, à la convergence, $\mathbf{c}_{dy} = \mathbf{R}_{yy}\mathbf{w}_*$, on peut ré-écrire l'algorithme d'adaptation selon

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu[n]\left(\mathbf{R}_{yy}\left(\mathbf{w}_* - \mathbf{w}[n]\right)\right),$$

soit encore

$$\mathbf{v}[n+1] = \left(\mathbf{I} - \mu[n]\mathbf{R}_{yy}\right)\mathbf{v}[n].$$

La matrice d'auto-corrélation est diagonalisable selon $\mathbf{R}_{yy} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$, où $\mathbf{\Lambda}$ est la matrice (diagonale) des valeurs propres et $\mathbf{Q}$ la matrice (unitaire) assurant la diagonalisation. Multipliant alors l'équation précédente par $\mathbf{Q}^T$ et notant $\mathbf{r}[n] := \mathbf{Q}^T\mathbf{v}[n]$, on obtient que

$$\mathbf{Q}^T\mathbf{v}[n+1] = \mathbf{Q}^T\mathbf{v}[n] - \mu[n]\mathbf{Q}^T\mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T\mathbf{v}[n],$$

soit encore

$$\mathbf{r}[n+1] = \left(\mathbf{I} - \mu[n]\mathbf{\Lambda}\right)\mathbf{r}[n]$$

en utilisant l'unitarité de $\mathbf{Q}$.

La convergence de cette équation aux modes propres est assurée si la suite des vecteurs $\mathbf{r}[n]$ tend vers zéro, garantissant en retour qu'il en est de même pour $\mathbf{v}[n]$, et donc que $\mathbf{w}[n] \to \mathbf{w}_*$.

Dans le cas à pas constant ($\mu[n] = \mu$, indépendant de n), l'équation aux modes propres correspond à la récurrence

$$\mathbf{r}[n] = \left(\mathbf{I} - \mu\mathbf{\Lambda}\right)^n \mathbf{r}[0],$$

soit encore (la matrice $\mathbf{\Lambda}$ étant diagonale), aux M relations de récurrence

$$r_k[n] = \left(1 - \mu\lambda_k\right)^n r_k[0]\,;\; k = 0, \ldots M-1,$$

en notant λ_k la k-ième valeur propre. On en déduit que la convergence nécessite d'avoir $|1 - \mu\lambda_k| < 1$ pour tout k, soit encore $-1 < 1 - \mu\lambda_k < 1$, et donc :

$$0 < \mu < \frac{2}{\lambda_{\max}}$$

en notant $\lambda_{\max}$ la valeur propre maximale.

En ce qui concerne la vitesse de convergence, il convient d'évaluer la quantité (dépendante du temps)

$$P[n] = \mathbb{E}\{d^2[n]\} - 2\mathbf{w}[n]^T\mathbf{R}_{yy}\mathbf{w}_* + \mathbf{w}[n]^T\mathbf{R}_{yy}\mathbf{w}[n].$$

Si l'on remarque que

$$\begin{aligned}\mathbf{v}[n]^T\mathbf{R}_{yy}\mathbf{v}[n] &= (\mathbf{w}[n]-\mathbf{w}_*)^T\mathbf{R}_{yy}(\mathbf{w}[n]-\mathbf{w}_*)\\ &= \mathbf{w}^T[n]\mathbf{R}_{yy}\mathbf{w}[n] - \mathbf{w}_*^T\mathbf{R}_{yy}\mathbf{w}[n] - \mathbf{w}^T[n]\mathbf{R}_{yy}\mathbf{w}_* + \mathbf{w}_*^T\mathbf{R}_{yy}\mathbf{w}_*,\end{aligned}$$

on obtient l'égalité

$$\mathbf{v}[n]^T\mathbf{R}_{yy}\mathbf{v}[n] - \mathbf{w}_*^T\mathbf{R}_{yy}\mathbf{w}_* = \mathbf{w}^T[n]\mathbf{R}_{yy}\mathbf{w}[n] - 2\,\mathbf{w}^T[n]\mathbf{R}_{yy}\mathbf{w}_*,$$

d'où l'expression de l'erreur courante :

$$\begin{aligned}P[n] &= \mathbb{E}\{d^2[n]\} - 2\mathbf{w}[n]^T\mathbf{R}_{yy}\mathbf{w}_* + \mathbf{w}[n]^T\mathbf{R}_{yy}\mathbf{w}[n]\\ &= \mathbb{E}\{d^2[n]\} - \mathbf{w}_*^T\mathbf{R}_{yy}\mathbf{w}_* + \mathbf{v}[n]^T\mathbf{R}_{yy}\mathbf{v}[n]\\ &= P_* + \mathbf{v}[n]^T\mathbf{R}_{yy}\mathbf{v}[n]\\ &= P_* + \mathbf{r}[n]^T\mathbf{\Lambda}\mathbf{r}[n],\end{aligned}$$

la dernière égalité étant obtenue en utilisant de nouveau les modes propres donnés par $\mathbf{r}[n] = \mathbf{Q}^T\mathbf{v}[n]$.

Comme on sait que l'on a $\mathbf{r}[n] = (\mathbf{I} - \mu\,\mathbf{\Lambda})^n\,\mathbf{r}[0]$, il s'ensuit que

$$\begin{aligned}P[n] &= P_* + \mathbf{r}[0]^T(\mathbf{I}-\mu\,\mathbf{\Lambda})^n\mathbf{\Lambda}(\mathbf{I}-\mu\,\mathbf{\Lambda})^n\mathbf{r}[0]\\ &= P_* + \sum_{k=0}^{M-1}\lambda_k\,(1-\mu\lambda_k)^{2n}\,r_k^2[0].\end{aligned}$$

Dans les conditions de convergence citées précédemment, on a le résultat attendu $\lim_{n\to\infty} P_n = P_*$, et ceci quelle que soit l'initialisation. L'évolution de la puissance de l'erreur en fonction du temps définit une *courbe d'apprentissage* constituée de la superposition de M exponentielles amorties. À chacun de ces modes correspond une vitesse de convergence fixée par $(1-\mu\lambda_k)$, le mode le plus lent (resp. le plus rapide) étant lié à la valeur propre la plus petite (resp. la plus grande), la convergence étant de plus d'autant plus lente (resp. rapide) que le pas d'adaptation est plus petit (resp. grand).

On peut alors chercher à optimiser le pas en minimisant la plus grande valeur prise par la quantité $\delta_k := |1 - \mu\lambda_k|$:

$$\mu_* = \arg\min_{\mu} \max_{k} \delta_k.$$

L'analyse des cas possibles en μ conduit à

$$\begin{aligned} \mu > \mu_* \Rightarrow 1 - \mu\lambda_k < 1 - \mu_*\lambda_k \quad &\Rightarrow \quad \max \delta_k = \mu\lambda_{\max} - 1 \\ \mu < \mu_* \Rightarrow 1 - \mu\lambda_k > 1 - \mu_*\lambda_k \quad &\Rightarrow \quad \max \delta_k = 1 - \mu\lambda_{\min} \end{aligned}$$

d'où la valeur optimale

$$\mu_* = \frac{2}{\lambda_{\min} + \lambda_{\max}}$$

obtenue en égalant les valeurs maximales identifiées pour l'un et l'autre cas.

En pratique, calculer les valeurs propres extrêmes est aussi coûteux qu'inverser la matrice d'autocorrélation, ce dont l'algorithme du gradient se propose précisément de faire l'économie. Une approximation est alors possible en calculant

$$\tilde{\mu}_* = \frac{2}{\sum_{k=0}^{M-1} \lambda_k} = \frac{2}{\mathrm{trace}\{\mathbf{R}_{yy}\}}.$$

La trace est facile à calculer, l'hypothèse de stationnarité se traduisant par le caractère Toeplitz de la matrice d'autocorrélation et donc la valeur

$$\mathrm{trace}\{\mathbf{R}_{yy}\} = M\,\mathbb{E}\{y^2[n]\}.$$

2.5.3 Algorithme LMS

L'algorithme du gradient est par nature une procédure *déterministe*, qui suppose connues la matrice d'auto-corrélation $\mathbf{R}_{yy} = \mathbb{E}\{\mathbf{y}[n]\mathbf{y}[n]^T\}$ et le vecteur d'inter-corrélation $\mathbf{c}_{dy} = \mathbb{E}\{d[n]\mathbf{y}[n]\}$. Les moyennes d'ensemble mises en jeu étant en pratique inaccessibles, il convient de les estimer, ce qui peut être fait de la manière la plus élémentaire en ignorant l'espérance mathématique :

$$\begin{aligned} \mathbf{R}_{yy} \quad &\rightarrow \quad \hat{\mathbf{R}}_{yy} := \mathbf{y}[n]\mathbf{y}[n]^T; \\ \mathbf{c}_{dy} \quad &\rightarrow \quad \hat{\mathbf{c}}_{dy} := d[n]\mathbf{y}[n]. \end{aligned}$$

La relation de mise à jour, qui s'écrit alors

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \mu[n]\,\mathbf{y}[n]\left(d[n] - \mathbf{y}[n]^T\mathbf{w}[n]\right),$$

prend ainsi un caractère *aléatoire*. On parle de *gradient stochastique* ou encore d'algorithme LMS (pour "Least Mean Squares").

2.5.4 Deux exemples

Identification — On considère un signal modélisé par la relation auto-régressive

$$y[n] = \sum_{k=1}^{K} a_k \, y[n-k] + b[n],$$

où $b[n]$ est une séquence blanche. Le problème est alors d'*identifier* ce modèle, c'est-à-dire d'estimer les paramètres $a_1, a_2 \ldots a_K$. Ceci peut se résoudre classiquement hors-ligne en établissant un système linéaire (dit "de Yule-Walker") et en en trouvant la solution soit de façon directe, soit par un algorithme récursif comme celui de Levinson-Durbin.

Une autre façon de résoudre le problème est de recourir à une approche adaptative dans laquelle le filtre à estimer est $\mathbf{w} := [a_1, a_2, \ldots a_K]^T$, son action sur le vecteur courant $\mathbf{y}[n] := [y[n], y[n-1], \ldots y[n-K+1]]^T$ réalisant en fait une *prédiction linéaire* de la valeur $y[n+1]$ à l'instant suivant. On peut donc utiliser l'approche du LMS en posant $d[n] = y[n+1]$, construisant K séquences d'estimées $\hat{a}_1[n], \hat{a}_2[n], \ldots \hat{a}_K[n]$ pour les coefficients recherchés. Pour un choix convenable du pas d'adaptation, ces séquences convergent vers les valeurs vraies, l'intérêt supplémentaire étant que, si les coefficients de la régression changent au cours du temps, l'algorithme a la capacité de les suivre. Il s'agit alors de fixer un compromis entre une bonne réactivité au changement (permise par un pas grand mais conduisant à de fortes fluctuations) et un lissage des fluctuations (obtenu avec un pas faible, mais au prix d'une plus faible capacité de suivi). Un exemple est donné à la Figure 2 pour un processus d'ordre 2 dont les coefficients changent brutalement de valeurs à la moitié de l'observation.

Soustraction de bruit — On considère cette fois le cas d'un mélange de la forme $y[n] = s[n] + b[n]$, où $s[n]$ est un signal utile corrompu par un bruit additif $b[n]$. On suppose de plus que l'on dispose d'une référence $u[n]$ dite "bruit seul", séquence de valeurs corrélées au bruit $b[n]$ que l'on désire "soustraire" [2]. Le filtrage adaptatif procède cette fois d'un renversement de perspective, l'objectif étant de réaliser une estimation du bruit additif ($d[n] = b[n]$)

2. Un cas typique est celui de la mesure du rythme cardiaque fœtal *in utero*, corrompu par le rythme cardiaque de la mère dont il est possible de prendre une deuxième mesure non affectée par celle du fœtus. Un autre exemple est celui de l'enregistrement d'une conversation dans un environnement bruité mais clos (habitacle, casque), la référence bruit seul pouvant être obtenue par une mesure extérieure.

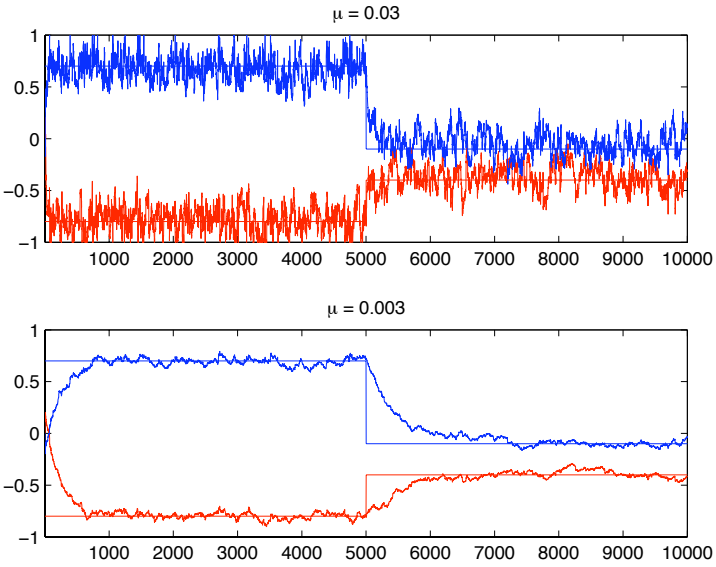

FIGURE 2 – Identification de filtre par l'algorithme LMS — Les courbes fluctuantes matérialisent l'estimation adaptative des coefficients d'un filtre auto-régressif d'ordre 2 dont les valeurs vraies sont indiquées par les valeurs constantes par morceaux (commutation à la moitié de l'observation). Une grande valeur du pas d'adaptation (diagramme du haut) permet une plus grande réactivité au changement mais fournit des estimées très variables. À l'inverse (diagramme du bas), un faible pas lisse les estimations mais requiert plus de temps pour la convergence.

sur la base d'une combinaison de valeurs de la référence bruit seul, le signal jouant en quelque sorte le rôle d'un "bruit" perturbateur dans ce contexte.

Une façon de faire est d'estimer le bruit par un filtre de réponse impulsionnelle finie de longueur M :

$$\hat{b}[n] = \sum_{k=0}^{M-1} w_k[n]\, u[n-k],$$

le signal estimé s'en déduisant selon $\hat{s}[n] = y[n] - \hat{b}[n]$, quantité jouant de plus le rôle d'erreur dans l'actualisation du filtre adaptatif relatif au bruit auxiliaire. Un exemple est donné à la Figure 3.

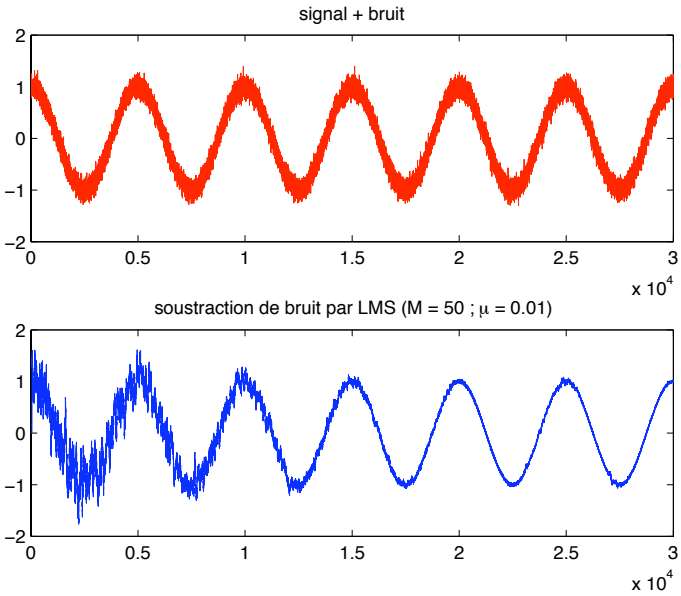

FIGURE 3 – Soustraction de bruit par l'algorithme LMS — Il est possible de soustraire une partie du bruit d'un signal bruité (diagramme du haut) si l'on dispose d'une référence "bruit seul" partiellement corrélée au bruit perturbateur. L'algorithme LMS est alors appliqué au bruit auxiliaire pour converger vers une estimation du bruit additif et, par soustraction, une estimation du signal (diagramme du bas).

3 Temps-fréquence, ondelettes et adaptativité

3.1 Incertitude

L'analyse de Fourier considère par nature les deux variables de temps et de fréquence comme exclusives, la représentation dans un domaine gommant *in fine* toute dépendance vis-à-vis de la variable de l'autre domaine. L'intuition suggère cependant qu'une représentation mixte (consistant d'une certaine manière à écrire la "partition" d'un signal sur une "portée" mathématique) devrait pouvoir exister de façon à pouvoir suivre les propriétés d'évolution du contenu spectral d'un signal.

Étant liées par une transformation de Fourier, les deux variables de temps et de fréquence (dites "canoniquement conjuguées") conduisent à des descrip-

tions individuelles qui entretiennent des relations de dépendance, dont une description conjointe traduira nécessairement les limitations.

La contrainte la plus classique est de type "incertitude" et stipule qu'un signal ne peut être arbitrairement localisé de façon simultanée en temps et en fréquence [3]. Pour le montrer, on peut convenir de mesurer l'encombrement d'un signal $x(t)$ (supposé d'énergie E_x finie, et centré en temps aussi bien qu'en fréquence) par les mesures de second ordre

$$\Delta t_x^2 := \frac{1}{E_x} \int_{-\infty}^{+\infty} t^2 \, |x(t)|^2 \, dt \quad ; \quad \Delta f_x^2 := \frac{1}{E_x} \int_{-\infty}^{+\infty} f^2 \, |X(f)|^2 \, df.$$

Si l'on introduit alors la quantité

$$I := \int_{-\infty}^{+\infty} t \, x^*(t) \, \dot{x}(t) \, dt,$$

l'inégalité de Cauchy-Schwarz permet d'établir que

$$\begin{aligned} |I|^2 &\leq \int_{-\infty}^{+\infty} t^2 \, |x(t)|^2 \, dt \int_{-\infty}^{+\infty} |\dot{x}(t)|^2 \, dt \\ &= E_x \, \Delta t_x^2 \, 4\pi^2 \int_{-\infty}^{+\infty} f^2 \, |X(f)|^2 \, df \\ &= 4\pi^2 E_x^2 \, \Delta t_x^2 \, \Delta f_x^2. \end{aligned}$$

Une intégration par parties montre par ailleurs que

$$I = \left[t \, |x(t)|^2 \right]_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \left[x^*(t) + t \, \dot{x}^*(t) \right] \, x(t) \, dt = -E_x - I^*,$$

d'où l'on déduit que

$$\mathrm{Re}\{I\} = -\frac{E_x}{2}.$$

Comme on a nécéssairement

$$|I|^2 = (\mathrm{Re}\{I\})^2 + (\mathrm{Im}\{I\})^2 \geq (\mathrm{Re}\{I\})^2 = \frac{E_x^2}{4},$$

3. Les relations d'incertitude sont associées aux noms de Werner Heisenberg (mécanique quantique, 1925) et Dennis Gabor (théorie des communications, 1946).

l'enchaînement des deux inégalités précédentes conduit au résultat central

$$\Delta t_x \, \Delta f_x \geq \frac{1}{4\pi}.$$

Il est facile de voir que la borne inférieure de cette inégalité est atteinte lorsque l'inégalité de Cauchy-Schwarz dont elle est issue se transforme en égalité, c'est-à-dire lorsque les deux termes impliqués dans le produit scalaire sont proportionnels :

$$\dot{x}(t) = k\, t\, x(t),$$

soit encore l'équation différentielle du premier ordre

$$\frac{dx}{x} = k\, t\, dt$$

dont la solution est de la forme :

$$x(t) = C\, e^{-\alpha t^2},$$

avec $\alpha > 0$ pour garantir à la solution d'être d'énergie finie. Les minimiseurs de l'incertitude temps-fréquence sont donc les gaussiennes, encore appelés "logons" de Gabor.

3.2 Décompositions linéaires

Afin d'obtenir une représentation conjointe en temps et en fréquence, une première approche, intuitive, est de généraliser la transformation de Fourier en décomposant *linéairement* un signal sur un ensemble de "briques de base" auxquelles on est en droit d'imposer de "bonnes" propriétés de localisation, en temps comme en fréquence. D'une manière plus précise, la valeur que prend un signal $x(t)$ en une date t_0 peut s'exprimer de façon équivalente selon

$$x(t_0) = \int_{-\infty}^{+\infty} x(t)\, \delta_t(t_0)\, dt = \int_{-\infty}^{+\infty} X(f)\, e_f(t_0)\, df,$$

en notant symboliquement $\delta_t(.)$ la distribution de Dirac en t et $e_f(t) := e^{i2\pi ft}$ l'exponentielle complexe de fréquence f. Si la première décomposition privilégie la description temporelle, la deuxième — dans laquelle $X(f)$ s'identifie à la transformée de Fourier du signal $x(t)$ — repose sur une interprétation duale en terme d'ondes. Ce sont bien sûr ces deux points de vue antinomiques

qu'une description mixte, en temps et en fréquence, se propose de concilier. Pour ce faire, les deux décompositions précédentes peuvent être remplacées par une troisième, intermédiaire, qui s'écrit

$$x(t_0) = \iint_{-\infty}^{+\infty} \lambda_x(t,f)\, g_{tf}(t_0)\, dt\, df.$$

Les fonctions $g_{tf}(.)$ ainsi mises en jeu permettent une transition entre les situations extrêmes précédentes : localisation parfaite en temps et nulle en fréquence lorsque $g_{tf}(.) \to \delta_t(.)$, parfaite en fréquence et nulle en temps lorsque $g_{tf}(.) \to e_f(.)$. Elles jouent en fait un rôle d'*atomes* temps-fréquence dans la mesure où on leur demande à la fois d'être des *constituants* de tout signal et de jouir de propriétés de *localisation* conjointe aussi idéales que possible ("élémentarité", dans la limite de ce qu'autorisent les inégalités temps-fréquence de type Heisenberg-Gabor). Pour qu'une telle décomposition prenne tout son sens, il faut naturellement qu'elle soit *inversible*, de telle sorte que l'on ait

$$\lambda_x(t,f) = \int_{-\infty}^{+\infty} x(t_0)\, g^*_{tf}(t_0)\, dt_0,$$

ce qui fait de $\lambda_x(t,f)$ une *représentation* temps-fréquence linéaire de $x(t)$.

Il y a bien sûr un grand arbitraire dans le choix d'une telle représentation. Une façon simple de procéder consiste à engendrer la famille des atomes $g_{tf}(.)$ par l'action d'un groupe de transformations agissant sur un élément primordial unique $g(.)$, par exemple un logon de Gabor dont on sait qu'il minimise l'encombrement temps-fréquence conjoint au sens de l'incertitude évoquée précédemment. La transformation la plus naturelle est quant à elle celle des translations en temps et en fréquence, dont est en en droit d'attendre qu'elle permette d'atteindre tout point temps-fréquence à partir de n'importe quel autre. Si l'on se place dans ce cas, on a en fait $g_{tf}(s) = g(s-t)\, e^{i2\pi fs}$, ce qui conduit à la famille des *transformées de Fourier à court-terme* (TFCT) de fenêtre $g(.)$:

$$F_x^{(g)}(t,f) := \langle x, g_{tf} \rangle = \int_{-\infty}^{+\infty} x(s)\, g^*(s-t)\, e^{-i2\pi fs}\, ds.$$

On peut tirer de cette définition trois conséquences :

1. *Admissibilité* — La fenêtre d'analyse doit être choisie de telle sorte qu'elle permette une généralisation de l'analyse-synthèse exacte de

Fourier selon [4]

$$F_x^{(g)}(t,f) = \langle x, g_{tf} \rangle$$

et

$$x(s) = \int\int_{-\infty}^{+\infty} \langle x, g_{tf} \rangle \, g_{tf}(s) \, dt \, df.$$

Dans le cas de la TFCT (où $g_{tf}(s)$ est défini par les translations temps-fréquence), on montre facilement que la condition d'admissibilité devant être vérifiée par l'atome de base $g(t)$ se réduit à $E_g = 1$. En effet, la vérification simultanée des deux relations d'analyse et de synthèse conduit à

$$\begin{aligned} x(s) &= \int\int_{-\infty}^{+\infty} \langle x, g_{tf} \rangle \, g_{tf}(s) \, dt \, df \\ &= \int_{-\infty}^{+\infty} x(s') \left(\int\int_{-\infty}^{+\infty} g_{tf}(s) \, g_{tf}^*(s') \, dt \, df \right) ds', \end{aligned}$$

soit encore, cette égalité devant être vérifiée pour tout signal $x(t)$, à la relation dite *de fermeture* :

$$\int\int_{-\infty}^{+\infty} g_{tf}(s) \, g_{tf}^*(s') \, dt \, df = \delta(s - s').$$

L'explicitation de cette condition conduit à

$$\begin{aligned} \delta(s - s') &= \int\int_{-\infty}^{+\infty} g(s-t) \, e^{i2\pi sf} \, g^*(s'-t) \, e^{-i2\pi s'f} \, dt \, df \\ &= \int_{-\infty}^{+\infty} g(s-t) \, g^*(s'-t) \, \left(\int_{-\infty}^{+\infty} e^{i2\pi(s-s')f} \, df \right) dt \\ &= \int_{-\infty}^{+\infty} g(s-t) \, g^*(s'-t) \, \delta(s-s') \, dt \\ &= E_g \, \delta(s-s'), \end{aligned}$$

d'où le résultat.

4. De manière générale, l'atome de "synthèse" ne s'identifie pas nécessairement à l'atome d'"analyse", mais on se restreindra ici au cas le plus simple où les deux sont identiques.

2. *Noyau reproduisant* — Une représentation temps-fréquence doit nécessairement porter la marque des relations d'incertitude qu'entretiennent ses variables. Dans le cas de la TFCT, ceci s'exprime par la relation de type "noyau reproduisant" :

$$F_x^{(g)}(t,f) = \iint_{-\infty}^{+\infty} K(t,f;t',f')\, F_x^{(g)}(t',f')\, dt'\, df',$$

où

$$K(t,f;t',f') := \langle g_{t'f'}, g_{tf} \rangle$$

est un noyau d'extension nécessairement non nulle dans le plan. En effet, si l'on part de la relation de synthèse

$$x(s) = \iint_{-\infty}^{+\infty} \langle x, g_{t'f'} \rangle g_{t'f'}(s)\, dt'\, df',$$

on en déduit que

$$\begin{aligned} F_x^{(g)}(t,f) &= \langle x, g_{tf} \rangle \\ &= \left\langle \iint_{-\infty}^{+\infty} \langle x, g_{t'f'} \rangle g_{t'f'}\, dt'\, df', g_{tf} \right\rangle \\ &= \iint_{-\infty}^{+\infty} \langle x, g_{t'f'} \rangle \langle g_{t'f'}, g_{tf} \rangle\, dt'\, df' \\ &= \iint_{-\infty}^{+\infty} K(t,f;t',f')\, F_x^{(g)}(t',f')\, dt'\, df'. \end{aligned}$$

La signification de cette relation est qu'une TFCT (continue) possède une redondance interne permettant d'exprimer sa valeur en n'importe quel point comme combinaison linéaire de ses autres valeurs en tous les autres points. On peut voir ceci comme l'analogue temps-fréquence de ce qui se passe dans le cas des signaux à bande limitée : les conséquences en sont de même nature, permettant d'exploiter la structure induite pour *échantillonner*, réduisant la redondance sans perte d'information.

3. *Isométrie* — On montre enfin que, tout comme pour la transformation de Fourier usuelle qui est telle que

$$\langle x, y \rangle = \langle X, Y \rangle,$$

la TFCT réalise elle aussi une isométrie, au sens où :

$$\langle x, y \rangle = \langle\!\langle F_x^{(g)}, F_y^{(g)} \rangle\!\rangle,$$

en notant

$$\langle\!\langle F, G \rangle\!\rangle := \int\!\!\int_{-\infty}^{+\infty} F(t,f)\, G^*(t,f)\, dt\, df.$$

En effet, on a

$$\begin{aligned}
\langle\!\langle F_x^{(g)}, F_y^{(g)} \rangle\!\rangle &= \int\!\!\int_{-\infty}^{+\infty} \langle x, g_{tf} \rangle \, \langle y, g_{tf} \rangle^* \, dt\, df \\
&= \int\!\!\int\!\!\int\!\!\int_{-\infty}^{+\infty} x(s)\, g_{tf}^*(s)\, y^*(s')\, g_{tf}(s')\, ds\, ds'\, dt\, df \\
&= \int\!\!\int_{-\infty}^{+\infty} x(s)\, y^*(s') \left(\int\!\!\int_{-\infty}^{+\infty} g_{tf}^*(s)\, g_{tf}(s')\, dt\, df \right) ds\, ds' \\
&= \int\!\!\int_{-\infty}^{+\infty} x(s)\, y^*(s')\, \delta(s'-s)\, ds\, ds' \\
&= \langle x, y \rangle.
\end{aligned}$$

d'où le résultat. On en déduit en particulier que

$$\int\!\!\int_{-\infty}^{+\infty} \left| F_x^{(g)}(t,f) \right|^2 \, dt\, df = E_x,$$

ce qui fait du module carré d'une TFCT (quantité que l'on appelle un *spectrogramme*) une distribution d'énergie analogue, dans le contexte temps-fréquence, à la densité spectrale d'énergie $|X(f)|^2$ basée sur la seule transformation de Fourier dans le contexte purement fréquentiel.

3.3 Ondelettes

3.3.1 Incertitude "adaptée"

Une lecture des relations d'incertitude est de dire qu'une fréquence est d'autant mieux définie qu'elle est observée sur un intervalle plus long, le minimum étant de la laisser "vivre" sur au moins une longueur d'onde. Cette observation amène à identifier une limitation dans la TFCT, une fenêtre de largeur temporelle fixe conduisant à une fréquence mal définie aux très basses

fréquences et "sur-définie" aux hautes fréquences. C'est en ce sens qu'une modification de la TFCT a été proposée par Jean Morlet à la fin des années 70, proposant d'asservir la durée d'une oscillation fenêtrée à la fréquence de celle-ci. Une façon simple d'y parvenir est de remplacer les translations en fréquence par des dilatations, soit encore de recourir au *groupe affine* dont l'action sur une forme d'onde élémentaire $\psi(.)$ s'écrit :

$$\psi(s) \longrightarrow \psi_{ta}(s) := \frac{1}{\sqrt{a}}\psi\left(\frac{s-t}{a}\right),$$

où a est un paramètre d'échelle positif. La Figure 4 illustre de façon comparée la génération d'atomes temps-fréquence par une TFCT et une TOC.

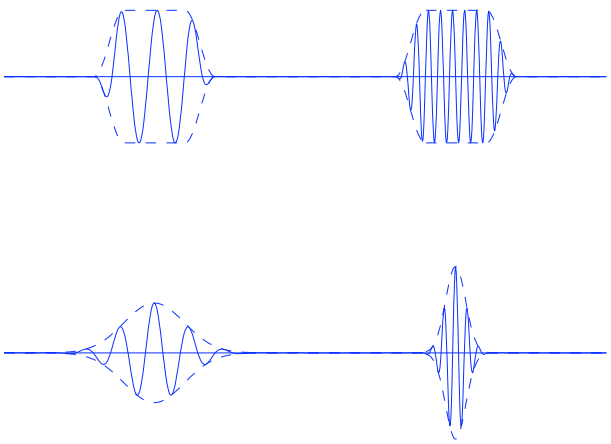

FIGURE 4 – TFCT vs. TOC. Les atomes d'une TFCT (en haut) diffèrent les uns des autres par une translation temporelle et une translation fréquentielle (modulation). Ceux d'une TOC (en bas) diffèrent par une translation temporelle et une dilatation (changement d'échelle).

3.3.2 Transformée continue

En procédant de manière parallèle à ce qui a été fait pour la TFCT, on définit ainsi la Transformée en Ondelettes Continues (TOC) selon :

$$T_x^{(\psi)}(t,a) := \langle x, \psi_{ta} \rangle = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(s)\, \psi^* \left(\frac{s-t}{a} \right) ds.$$

On en déduit de même les trois propriétés fondamentales suivantes :

1. *Admissibilité* — La TOC est inversible selon :

$$x(t) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} T_x^{(\psi)}(s,a)\, \psi_{sa}(t)\, \frac{ds\, da}{a^2}$$

sous la condition (plus restrictive que pour la TFCT) que

$$\int_0^{+\infty} |\Psi(f)|^2\, \frac{df}{f} = 1.$$

En effet, en procédant comme on l'a fait pour la TFCT, on voit que la condition de vérification simultanée des relations d'analyse et de synthèse de la TOC conduit à :

$$\begin{aligned} x(s) &= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \langle x, \psi_{ta} \rangle\, \psi_{ta}(s)\, \frac{dt\, da}{a^2} \\ &= \int_{-\infty}^{+\infty} x(s') \left(\int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi_{ta}(s)\, \psi_{ta}^*(s')\, \frac{dt\, da}{a^2} \right) ds', \end{aligned}$$

d'où la nouvelle relation de fermeture :

$$\int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi_{ta}(s)\, \psi_{ta}^*(s')\, \frac{dt\, da}{a^2} = \delta(s-s').$$

En explicitant cette relation, on obtient

$$\begin{aligned} \delta(s-s') &= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \frac{1}{\sqrt{a}} \psi \left(\frac{s-t}{a} \right) \frac{1}{\sqrt{a}} \psi^* \left(\frac{s'-t}{a} \right) \frac{dt\, da}{a^2} \\ &= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi(\theta)\, \psi^* \left(\theta - \frac{s-s'}{a} \right) \frac{d\theta\, da}{a^3}, \end{aligned}$$

soit encore

$$\delta(\tau) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi(\theta)\, \psi^*\left(\theta - \frac{\tau}{a}\right)\, \frac{d\theta\, da}{a^3}.$$

Par transformation de Fourier des deux membres, ceci est équivalent à

$$\begin{aligned}
1 &= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi(\theta)\, \psi^*\left(\theta - \frac{\tau}{a}\right)\, \frac{d\theta\, da}{a^3} \right) e^{-i2\pi f\tau}\, d\tau \\
&= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi(\theta) \left(\int_{-\infty}^{+\infty} \psi^*\left(\theta - \frac{\tau}{a}\right)\, e^{-i2\pi f\tau}\, d\tau \right) \frac{d\theta\, da}{a^3} \\
&= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \psi(\theta) \left(\int_{-\infty}^{+\infty} \psi(\theta')\, e^{i2\pi fa(\theta-\theta')}\, d\theta' \right)^* \frac{d\theta\, da}{a} \\
&= \int_{0}^{+\infty} \left| \int_{-\infty}^{+\infty} \psi(\theta)\, e^{-i2\pi af\theta}\, d\theta \right|^2 \frac{da}{a} \\
&= \int_{0}^{+\infty} |\Psi(f)|^2\, \frac{df}{f}.
\end{aligned}$$

2. *Noyau reproduisant* — La TOC possède naturellement une structure de redondance décrite par la relation

$$T_x^{(\psi)}(t,a) = \int_{-\infty}^{+\infty} \int_{0}^{+\infty} K(t,a;t',a')\, T_x^{(\psi)}(t',a')\, \frac{dt'\, da'}{a'^2},$$

avec

$$K(t,a;t',a') := \langle \psi_{t'a'}, \psi_{ta} \rangle.$$

En effet,

$$\begin{aligned}
T_x^{(\psi)}(t,a) &= \langle x, \psi_{ta} \rangle \\
&= \left\langle \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \langle x, \psi_{t'a'} \rangle \psi_{t'f'}\, \frac{dt'\, da'}{a'^2}, \psi_{ta} \right\rangle \\
&= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} \langle x, \psi_{t'a'} \rangle\, \langle \psi_{t'a'}, \psi_{ta} \rangle\, \frac{dt'\, da'}{a'^2} \\
&= \int_{-\infty}^{+\infty} \int_{0}^{+\infty} K(t,a;t',a')\, T_x^{(\psi)}(t',a')\, \frac{dt'\, da'}{a'^2}.
\end{aligned}$$

3. *Isométrie* — La TOC réalise enfin une isométrie de $L^2(\mathbb{R})$ vers $L^2(\mathbb{R}^2)$ qui s'exprime par

$$\langle\langle T_x^{(\psi)}, T_y^{(\psi)} \rangle\rangle = \langle x, y \rangle,$$

en notant cette fois

$$\langle\langle T, S \rangle\rangle := \int_{-\infty}^{+\infty} \int_0^{+\infty} T(t,a)\, S^*(t,a)\, \frac{dt\, da}{a^2},$$

et avec la conséquence que

$$\int_{-\infty}^{+\infty} \int_0^{+\infty} \left| T_x^{(\psi)}(t,a) \right|^2 \frac{dt\, da}{a^2} = E_x.$$

On notera que le module carré de la TOC (quantité que l'on appelle *scalogramme*) joue ainsi un rôle de distribution d'énergie temps-échelle [5].

Sur l'admissibilité — La condition d'admissibilité de la TOC appelle quelques commentaires. Afin que l'intégrale mise en jeu converge, il est nécessaire que la décroissance à l'infini du spectre $\Psi(f)$ de la fonction de base $\psi(t)$ soit suffisamment rapide, mais surtout que le comportement du même spectre à l'origine permette de contrebalancer la divergence possible liée à la mesure df/f. Une condition nécessaire est qu'il prenne une valeur nulle à l'origine, ce qui peut s'écrire de deux manières :

$$\Psi(0) = 0 \Leftrightarrow \int_{-\infty}^{+\infty} \psi(t)\, dt = 0.$$

En mettant l'accent sur la première expression de la condition en question, on voit que le spectre d'une ondelette admissible (nul à l'origine et à l'infini) a nécessairement un caractère passe-bande. En privilégiant la seconde, la condition stipule que, dans le domaine temporel, une ondelette admissible doit être de moyenne nulle. Dans les deux cas, la fonction de base possède la double caractéristique d'être oscillante et localisée, d'où sa dénomination d'"ondelette".

5. Tout comme pour la généralisation du spectrogramme à des classes de distributions d'énergie temps-fréquence, on peut construire des classes générales de distributions d'énergie temps-échelle.

Quelques ondelettes — La pemière ondelette à avoir été utilisée dans un cadre continu, et ceci avant même que la théorie n'en soit formalisée, est celle dite *de Morlet*, directement inspirée de l'analyse proposée précédemment par Gabor dans le cadre de la TFCT. L'argument était de se référer à l'usage d'une gaussienne à cause de sa propriété "d'incertitude minimale", en la modulant ni trop ni trop peu afin qu'elle ait à la fois un caractère oscillant mais aussi une durée suffisamment brève relativement à la fréquence de cette oscillation pour lui permettre de la distinguer d'un mode de Fourier. Il en résulte une fonction du type

$$\psi_m(t) = C\, \exp\left\{-\frac{t^2}{2\sigma^2} + i2\pi f_0 t\right\},$$

avec $\sigma = \alpha/f_0$ et, typiquement, $\alpha \approx 1$ (cf. Figure 5). Paradoxalement, une telle ondelette n'est cependant pas admissible *stricto sensu*, puisque l'on a

$$\Psi_M(0) = C\,\sqrt{2\pi}\,\sigma\, \exp\{-2\pi^2\alpha^2\} \neq 0.$$

On remédie usuellement à ce défaut en "centrant" artificiellement l'oscillation ($\psi_m(t) \to \psi_m(t) - \Psi_m(0)$), en notant que l'erreur peut en tout état de cause être rendue négligeable pour un choix adéquat de α (ainsi, une erreur relative de 10^{-6} à l'origine correspond à $\alpha = 0.837$).

Un autre choix, admissible et plus flexible, est de toujours s'appuyer sur la gaussienne, mais de recourir à des dérivées successives de celle-ci, définissant ainsi une *famille* d'ondelettes

$$\psi^{(n)}(t) := C_n\, \frac{d^n}{dt^n}\left(\exp\left\{-\frac{t^2}{2\sigma^2}\right\}\right); n \geq 1,$$

la valeur $n = 2$ fournissant ce qu'il est convenu d'appeler l'ondelette "chapeau mexicain" (cf. Figure 5).

Interprétation fréquentielle — Par application de la relation de Plancherel (conservation du produit scalaire), il est immédiat de donner de la TOC la formulation équivalente :

$$T_x^{(\psi)}(t,a) = \sqrt{a}\, \int_{-\infty}^{+\infty} X(f)\, \Psi^*(af)\, e^{i2\pi ft}\, df,$$

opérant sur le spectre $X(f)$ du signal. Ceci permet de voir la TOC, en temps que fonction du temps t, comme la sortie d'un filtre linéaire de fonction de transfert $\Psi(af)$.

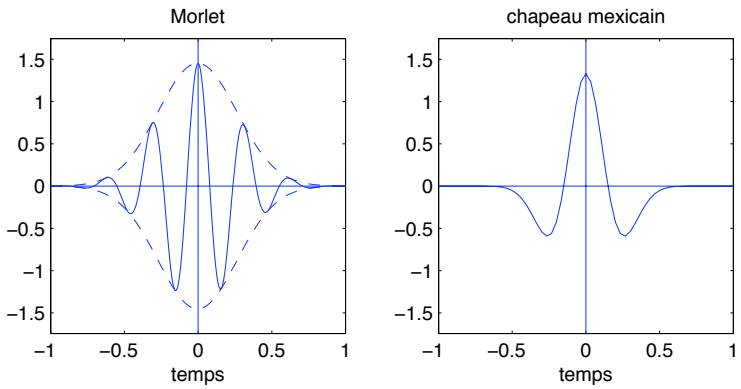

FIGURE 5 – Deux exemples d'ondelettes continues : Morlet (à gauche) et "chapeau mexicain" (à droite).

Pour une échelle de référence fixée et prise arbitrairement comme unité ($a = 1$), le cas le plus simple est celui d'une ondelette de spectre monomodal prenant sa valeur maximale à une fréquence $f_0 > 0$. Lui appliquant alors la transformation induite par le groupe affine, on voit qu'il en résultera un spectre de même forme (à un changement d'échelle et une renormalisation d'amplitude près), prenant sa valeur maximale pour la fréquence f_0/a. Ceci correspond à une valeur plus grande lorsque $a < 1$ (dilatation spectrale et compression temporelle) et plus petite lorsque $a > 1$ (compression spectrale et dilatation temporelle). Changer d'échelle permet ainsi de faire une analyse que l'on interpréter en termes temps-fréquence, moyennant l'identification formelle $f = f_0/a$. La résolution fréquentielle Δf de cette analyse est elle-même dépendante de la fréquence dans les mêmes proportions ($\Delta f/f = \text{Cte}$, analyse dite "à surtension constante") : les petites échelles privilégient une analyse temporelle fine au détriment de la résolution fréquentielle, la situation étant inversée pour les grandes échelles. Ceci est illustré en Figure 6.

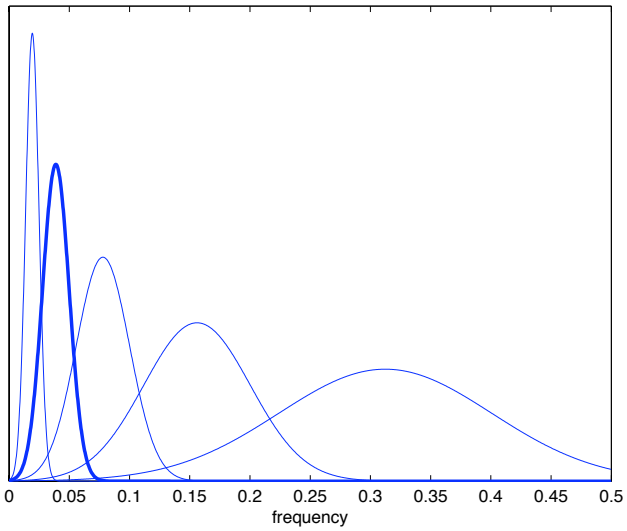

FIGURE 6 – Interprétation fréquentielle d'une TOC. Pour une échelle de référence $a = 1$ prise arbitrairement comme unité, une ondelette effectue un lissage passe-bande (trait épais). Pour des valeurs différentes de l'échelle, le filtrage est de même nature, à une dilatation en fréquence et une renormalisation en amplitude près (traits fins).

3.3.3 Transformées discrètes

Obtenir une Transformée en Ondelettes Discrètes (TOD) peut s'envisager de deux manières. La première consiste à discrétiser une TOC en réduisant sa redondance *via* la structure de son noyau reproduisant. La seconde consiste à construire directement une transformée discrète sur la base d'une approche générale dite de *multirésolution*, et sans référence directe explicite à une idée de discrétisation. On se contentera ici de donner quelques éléments essentiels relatifs à cette deuxième approche.

Analyses multirésolution et bases orthonormées — La discrétisation d'une transformation en ondelettes peut être associée, dans certains cas, à l'existence de véritables *bases orthonormées* construites sur des atomes temps-échelle bien localisés en temps et en fréquence. Ainsi, en se plaçant dans le cas d'une décomposition en échelles *dyadiques*, il est possible trouver une

ondelette $\psi(t) \in \mathbb{R}$ telle que la famille $\{\psi_{nm}(t) := 2^{m/2}\psi(2^m t - n); n, m \in \mathbb{Z}\}$ soit une base orthonormée de $L^2(\mathbb{R})$, c'est-à-dire dont les éléments sont tels que $\langle \psi_{nm}, \psi_{n'm'} \rangle = \delta_{nm}\delta_{n'm'}$, de telle sorte que l'on ait, pour tout signal $x(t) \in L^2(\mathbb{R})$,

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} \langle x, \psi_{nm} \rangle \, \psi_{nm}(t).$$

Pour ce faire, la construction repose de façon centrale sur la notion d'*analyse multirésolution* (AMR). Celle-ci formalise l'idée intuitive selon laquelle tout signal peut être construit par raffinements successifs, c'est-à-dire par l'ajout de *détails* à une *approximation*, et l'itération du processus.

D'une manière plus précise, une AMR de $L^2(\mathbb{R})$ est définie par une séquence de sous-espaces emboîtés $\ldots \subset \mathbf{V}_{-1} \subset \mathbf{V}_0 \subset \mathbf{V}_1 \subset \ldots$ tels que

1. $\bigcap_{m=-\infty}^{\infty} \mathbf{V}_m = \{0\}$;
2. $\overline{\bigcup_{m=-\infty}^{\infty} \mathbf{V}_m} = L^2(\mathbb{R})$;
3. $x(t) \in \mathbf{V}_m \Leftrightarrow x(2t) \in \mathbf{V}_{m+1}$;
4. il existe une fonction $\varphi(t)$ telle que $\{\varphi(t-n); n \in \mathbb{Z}\}$ soit une base de $\mathbf{V}_0$.

Chaque sous-espace $\mathbf{V}_m$ se trouve associé à une résolution 2^m et l'approximation d'un signal $x(t)$ à cette résolution est obtenue par projection sur ce sous-espace. Étant données les propriétés ci-dessus, une base de $\mathbf{V}_m$ peut alors se déduire de celle de $\mathbf{V}_0$ en construisant la famille des dilatées et translatées $\{\varphi_{nm}(t) := 2^{m/2}\varphi(2^m t - n); n, m \in \mathbb{Z}\}$ à partir de la seule fonction $\varphi(t)$, appelée *fonction d'échelle*. Les coefficients d'approximation $\langle x, \varphi_{nm} \rangle$ associés à l'ensemble de ces différentes bases partagent évidemment une grande quantité d'information, ce qui en fait une représentation extrêmement redondante. En reprenant l'idée d'un signal en termes d'approximations *successives*, une représentation beaucoup plus économique consiste à s'intéresser à la *différence d'information* qui existe entre deux approximations voisines, c'est-à-dire au *détail* qu'il faut ajouter à la plus grossière pour passer à la plus fine. Pour chaque espace d'approximation $\mathbf{V}_m$, ceci revient à dire que les détails appartiennent à l'espace $\mathbf{W}_m$ qui est le complément orthogonal de $\mathbf{V}_m$ dans $\mathbf{V}_{m+1}$. Ceci signifie que l'on a la relation

$$\mathbf{V}_{m+1} = \mathbf{V}_m \oplus \mathbf{W}_m,$$

d'où l'on déduit la décomposition

$$L^2(\mathbb{R}) = \bigoplus_{m=-\infty}^{\infty} \mathbf{W}_m.$$

Il suffit alors de trouver une fonction $\psi(t)$ (qui est l'*ondelette* à proprement parler) telle $\{\psi(t-n); n \in \mathbb{Z}\}$ soit une base de $\mathbf{W}_0$ pour que la collection $\{\psi_{nm}(t) := 2^{m/2}\psi(2^m t - n); n, m \in \mathbb{Z}\}$ soit une base orthonormée de $L^2(\mathbb{R})$.

Fonction d'échelle — En remarquant que, si la fonction d'échelle $\varphi(t)$ est dans $\mathbf{V}_0$, elle est également dans $\mathbf{V}_1$, il existe nécessairement une famille de coefficients $h[n]$ tels que l'on puisse écrire la relation à deux échelles

$$\varphi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} h[n]\, \varphi(2t-n),$$

avec

$$h[n] = \sqrt{2} \int_{-\infty}^{+\infty} \varphi(t)\, \varphi(2t-n)\, dt$$

et

$$\sum_{n=-\infty}^{\infty} h^2[n] = 1.$$

Parce que les translatées entières de la fonction d'échelle $\varphi(t)$ forment une base de $\mathbf{V}_0$, le *filtre discret* défini par la séquence des coefficients $h[n]$ possède des propriétés très particulières. Par transformation de Fourier de l'équation à deux échelles caractérisant $\varphi(t)$, on obtient ainsi $\Phi(f) = H(f/2)\,\Phi(f/2)$ en posant

$$H(f) := \frac{1}{\sqrt{2}} \sum_{n=-\infty}^{\infty} h[n]\, e^{i2\pi f n}.$$

Cette fonction de transfert (périodique de période 1) n'est elle-même pas quelconque puisque, par l'orthogonalité des $\varphi_{n0}(t)$ dans $\mathbf{V}_0$, on a

$$\begin{aligned}
\delta_{k0} &= \int_{-\infty}^{+\infty} \varphi(t)\, \varphi(t-k)\, dt \\
&= \int_{-\infty}^{+\infty} |\Phi(f)|^2\, e^{i2\pi f k}\, df \\
&= \int_0^1 \left(\sum_{n=-\infty}^{\infty} |\Phi(f+n)|^2 \right) e^{i2\pi f k}\, df,
\end{aligned}$$

ce qui entraîne que, pour toute fréquence f,

$$\sum_{n=-\infty}^{\infty} |\Phi(f+n)|^2 = 1.$$

De plus, en faisant apparaître dans cette équation la fonction de transfert $H(f)$, en utilisant sa périodicité et en posant $f = 2\zeta$, on obtient

$$\begin{aligned} 1 &= \sum_{n=-\infty}^{\infty} |H(\zeta + n/2)|^2 \, |\Phi(\zeta + n/2)|^2 \\ &= \sum_{k=-\infty}^{\infty} |H(\zeta + k)|^2 \, |\Phi(\zeta + k)|^2 + \sum_{k=-\infty}^{\infty} |H(\zeta + 1/2 + k)|^2 \, |\Phi(\zeta + 1/2 + k)|^2 \\ &= |H(\zeta)|^2 \sum_{k=-\infty}^{\infty} |\Phi(\zeta + k)|^2 + |H(\zeta + 1/2)|^2 \sum_{k=-\infty}^{\infty} |\Phi(\zeta + 1/2 + k)|^2, \end{aligned}$$

d'où il suit que l'on a, pour toute fréquence f, la relation dite de *symétrie en miroir*

$$|H(f)|^2 + |H(f + 1/2)|^2 = 1.$$

Ondelette — En revenant à l'ondelette $\psi(t)$ qui est dans $\mathbf{V}_1$, on peut également la caractériser par un filtre discret de coefficients $g[n]$ tels que

$$\psi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} g[n] \, \varphi(2t - n),$$

avec

$$g[n] = \sqrt{2} \int_{-\infty}^{+\infty} \psi(t) \, \varphi(2t - n) \, dt$$

et

$$\sum_{n=-\infty}^{\infty} g^2[n] = 1.$$

Avec les mêmes conventions que précédemment, on peut écrire cette fois $\Psi(f) = G(f/2) \, \Phi(f/2)$, où $G(f)$ est la fonction de transfert (1-périodique) du filtre discret $g[n]$ associé à l'ondelette $\psi(t)$.

Parce que l'espace $\mathbf{W}_0$ auquel appartient $\psi(t)$ est orthogonal à $\mathbf{V}_0$, on a nécessairement

$$\begin{aligned} 0 &= \int_{-\infty}^{+\infty} \psi(t)\,\varphi(t-k)\,dt \\ &= \int_{-\infty}^{+\infty} \Psi(f)\,\Phi^*(f)\,e^{i2\pi fk}\,df \\ &= \int_0^1 \left(\sum_{n=-\infty}^{\infty} \Psi(f+n)\,\Phi^*(f+n) \right) e^{i2\pi fk}\,df \end{aligned}$$

et donc, pour toute fréquence f,

$$\sum_{n=-\infty}^{\infty} \Psi(f+n)\,\Phi^*(f+n) = 0.$$

Procédant comme précédemment, c'est-à-dire en faisant apparaître dans cette équation les fonctions $H(f)$ et $G(f)$ et en utilisant leurs propriétés de péridodicité, on obtient finalement que

$$G(f)\,H^*(f) + G(f+1/2)\,H^*(f+1/2) = 0.$$

Dans le langage de la théorie des bancs de filtres, les filtres discrets $h[n]$ et $g[n]$ forment une paire de *filtres miroirs en quadrature*.

À $H(f)$ fixée, la solution en $G(f)$ de l'équation ci-dessus est de la forme $G(f) = \lambda(f)\,H^*(f+1/2)$, où $\lambda(f)$ est une fonction 1-périodique telle que $\lambda(f)+\lambda(f+1/2) = 0$. Un choix possible est de prendre $\lambda(f) = -\exp(i2\pi f)$, de telle sorte que

$$G(f) = e^{i2\pi(f+1/2)}\,H^*(f+1/2) \Rightarrow g[n] = (-1)^n\,h[1-n].$$

Il reste à vérifier que l'ondelette $\psi(t)$ ainsi construite est bien telle que

ses translatées entières forment une base de $\mathbf{W}_0$. Pour ce faire, on montre

$$\begin{aligned}
\sum_{n=-\infty}^{\infty} |\Psi(f+n)|^2 &= \sum_{n=-\infty}^{\infty} |G(f/2+n/2)|^2 \, |\Phi(f/2+n/2)|^2 \\
&= \sum_{n=-\infty}^{\infty} |H(f/2+n/2+1/2)|^2 \, |\Phi(f/2+n/2)|^2 \\
&= |H(f/2+1/2)|^2 \sum_{k=-\infty}^{\infty} |\Phi(f/2+k/2)|^2 \\
&\qquad + |H(f/2)|^2 \sum_{k=-\infty}^{\infty} |\Phi(f/2+1/2+k)|^2 \\
&= |H(f/2+1/2)|^2 + |H(f/2)|^2 \\
&= 1
\end{aligned}$$

et l'on en déduit que

$$\begin{aligned}
\int_{-\infty}^{+\infty} \psi(t)\,\psi(t-k)\,dt &= \int_{-\infty}^{+\infty} |\Psi(f)|^2 \, e^{i2\pi fk} \, df \\
&= \int_0^1 \left(\sum_{n=-\infty}^{\infty} |\Psi(f+n)|^2 \right) e^{i2\pi fk} \, df \\
&= \delta_{k0}.
\end{aligned}$$

Ainsi, et pour résumer, en partant d'une fonction d'échelle $\varphi(t)$ (telle que la famille $\{\varphi(t-n); n \in \mathbb{Z}\}$ soit une base de $\mathbf{V}_0$) et des coefficients $h[n]$ du filtre discret qui lui est associé, l'ondelette

$$\psi(t) = \sqrt{2} \sum_{n=-\infty}^{\infty} (-1)^n \, h[1-n] \, \varphi(2t-n)$$

est telle que la collection $\{\psi_{nm}(t) := 2^{m/2}\psi(2^m t - n); n, m \in \mathbb{Z}\}$ est une base de $L^2(\mathbb{R})$.

Passe-bas, passe-haut et passe-bande — La fonction d'échelle $\varphi(t)$ et son filtre associé $h[n]$ possèdent les caractéristiques d'un filtre *passe-bas* tandis

que l'ondelette $\psi(t)$ et son filtre associé $g[n]$ possèdent celles d'un filtre *passe-haut*. En effet, on peut écrire

$$\begin{aligned}\Phi(f) &= H(f/2)\,\Phi(f/2)\\ &= H(f/2)\,H(f/4)\,\Phi(f/4)\end{aligned}$$

et itérer la factorisation. Par suite, le passage à la limite conduit à

$$\Phi(f) = \Phi(0)\prod_{m=1}^{\infty} H(2^{-m}f),$$

ce qui impose à $\Phi(0)$ d'avoir une valeur finie non nulle. Il s'ensuit que

$$\begin{aligned}\Phi(0) &= \int_{-\infty}^{+\infty}\varphi(t)\,dt\\ &= \int_{-\infty}^{+\infty}\sqrt{2}\sum_{n=-\infty}^{\infty}h[n]\,\varphi(2t-n)\,dt\\ &= \Phi(0)\,\frac{1}{\sqrt{2}}\sum_{n=-\infty}^{\infty}h[n],\end{aligned} \tag{1}$$

soit encore $H(0) = 1$ et, par voie de conséquence, $H(1/2) = 0$. On en déduit facilement que $G(0) = 0$ (d'où $\Psi(0) = 0$) et $|G(1/2)| = 1$. Incidemment, on tire des relations de filtrage en quadrature que $|H(f)|^2 + |G(f)|^2 = 1$, ce qui, en utilisant l'extinction à l'infini de $\Phi(f)$ (due à son caractère passe-bas), entraîne que

$$\begin{aligned}|\Phi(f)|^2 &= |\Psi(2f)|^2 + |\Phi(2f)|^2\\ &= |\Psi(2f)|^2 + |\Psi(4f)|^2 + |\Phi(4f)|^2\\ &= \ldots\\ &= \sum_{m=1}^{\infty}|\Psi(2^m f)|^2\end{aligned} \tag{2}$$

pour toute fréquence f non nulle. On peut lire ce résultat comme le fait qu'appliquer l'action *passe-haut* du filtre ondelette à l'approximation résultant d'un filtrage *passe-bas* consiste globalement en une opération de filtrage *passe-bande* à chaque échelle, rejoignant en cela l'interprétation de l'ondelette qui avait été donnée dans le cas de la TOC.

Analyse-synthèse par algorithmes pyramidaux — L'introduction des filtres discrets $h[n]$ et $g[n]$ présente un intérêt majeur pour le calcul pratique des coefficients d'*approximation* $a_x[n,m] := \langle x, \varphi_{nm} \rangle$ et des coefficients de *détail* $d_x[n,m] := \langle x, \psi_{nm} \rangle$. En effet, en utilisant les relations établies précédemment, il est facile de voir que

$$\begin{aligned} a_x[n,m] &= \int_{-\infty}^{+\infty} x(t)\, 2^{m/2}\, \varphi(2^m t - n)\, dt \\ &= \int_{-\infty}^{+\infty} x(t) 2^{m/2} \left(\sqrt{2} \sum_{k=-\infty}^{\infty} h[k] \varphi(2(2^m t - n) - k) \right) dt \\ &= \sum_{k=-\infty}^{\infty} h[k] \int_{-\infty}^{+\infty} x(t) 2^{(m+1)/2} \varphi(2^{m+2} t - (k+2n)) dt \\ &= \sum_{k=-\infty}^{\infty} h[k]\, a_x[k+2n, m+1], \end{aligned}$$

soit encore

$$a_x[n,m] = \sum_{k=-\infty}^{\infty} h[k-2n]\, a_x[k, m+1].$$

De la même façon, on établit que

$$d_x[n,m] = \sum_{k=-\infty}^{\infty} g[k-2n]\, a_x[k, m+1].$$

Ainsi, on voit que les coefficients d'approximation et de détail associés à une résolution donnée peuvent se déduire, par *filtrages* suivis de *décimation*, de la donnée des coefficients d'approximation à la résolution immédiatement supérieure. Opérant de proche en proche, on dispose donc d'un algorithme rapide et récursif, qui ne met en jeu que deux filtres discrets dont on itère l'action. Un tel algorithme est dit *pyramidal*. D'un point de vue pratique, l'initialisation de l'algorithme se fait, soit en projetant le signal analysé sur $\mathbf{V}_0$ si l'on en dispose sous une forme à temps continu, soit en ramenant la séquence des échantillons dont on dispose dans ce même espace, si les données sont à temps discret.

Le schéma d'*analyse* est réversible et conduit à un algorithme dual de *synthèse*, dans lequel une approximation à une résolution donnée se déduit de l'approximation et du détail à la résolution immédiatement inférieure. Pour

établir la structure de cet algorithme, il suffit de considérer que l'approximation $x_m(t)$ d'un signal $x(t)$ à la résolution 2^m s'obtient par projection de ce dernier sur $\mathbf{V}_m$, sous-espace de $L^2(\mathbb{R})$ dont $\{\varphi_{nm}(t) = 2^{m/2}\,\varphi(2^m t - n); n \in \mathbb{Z}\}$ est une base, à m fixé. Par suite, on peut écrire

$$x_m(t) = \sum_{n=-\infty}^{\infty} a_x[n,m]\,\varphi_{nm}(t).$$

Comme on sait par ailleurs que $\mathbf{V}_{m+1} = \mathbf{V}_m \oplus \mathbf{W}_m$, où le sous-espace orthogonal $\mathbf{W}_m$ admet pour base $\{\psi_{nm}(t) = 2^{m/2}\,\psi(2^m t - n); n \in \mathbb{Z}\}$ à m fixé, on obtient que

$$x_{m+1}(t) = x_m(t) + \sum_{k=-\infty}^{\infty} d_x[k,m]\,\psi_{km}(t)$$

d'où, par projection de cette égalité sur $\varphi_{n,m+1}(t)$,

$$a_x[n,m+1] = \sum_{k=-\infty}^{\infty} a_x[k,m]\langle\varphi_{km},\varphi_{n,m+1}\rangle + \sum_{k=-\infty}^{\infty} d_x[n,m]\langle\psi_{km},\psi_{n,m+1}\rangle.$$

Il est alors facile de se convaincre que

$$\begin{aligned}\langle\varphi_{km},\varphi_{n,m+1}\rangle &= \int_{-\infty}^{+\infty} 2^{m/2}\varphi(2^m t-k)2^{(m+1)/2}\varphi(2^{m+1}t-n)dt\\ &= \sqrt{2}\int_{-\infty}^{+\infty}\varphi(t)\,\varphi(2t-(n-2k))\,dt\\ &= h[n-2k]\end{aligned}$$

et, par un raisonnement analogue, que $\langle\psi_{km},\psi_{n,m+1}\rangle = g[n-2k]$, d'où l'on déduit enfin que la relation de reconstruction cherchée s'écrit

$$a_x[n,m+1] = \sum_{k=-\infty}^{\infty} h[n-2k]\,a_x[k,m] + \sum_{k=-\infty}^{\infty} g[n-2k]\,d_x[k,m].$$

À l'inverse de l'algorithme d'analyse qui opérait par filtrages suivis de décimation, l'algorithme de synthèse opère par *interpolation* suivie de *filtrages*.

La structure d'ensemble de l'analyse-synthèse par ondelettes orthonormées est résumée en Figure 7.

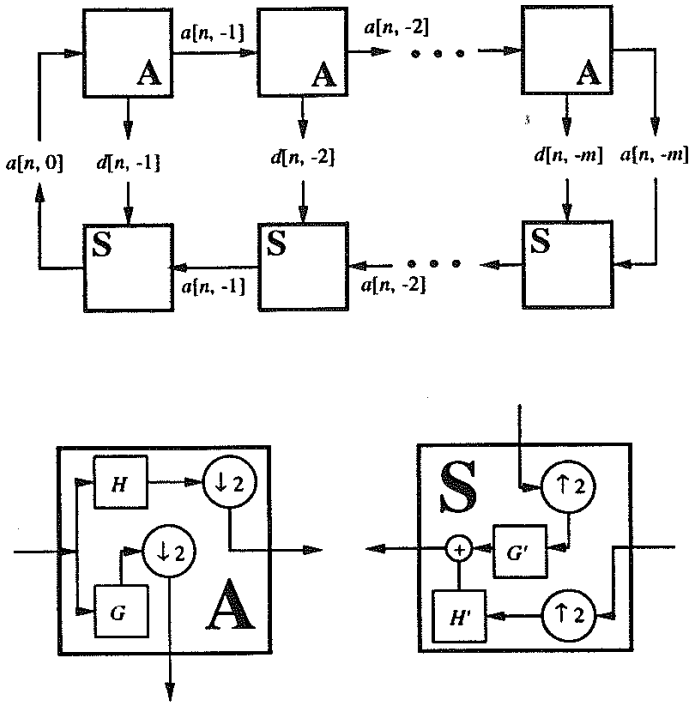

FIGURE 7 – Analyse-synthèse par ondelettes orthonormées.

3.3.4 Quelques base d'ondelettes

Haar — L'exemple le plus simple de base orthonormée d'ondelettes se construit à partir de la fonction d'échelle

$$\varphi(t) = \begin{cases} 1 & \text{si } 0 \leq t < 1 \\ 0 & \text{sinon,} \end{cases}$$

ce qui reveint à s'intéresser essentiellement aux signaux continus par morceaux. On obtient alors pour le filtre associé

$$h[n] = \sqrt{2} \int_{-\infty}^{+\infty} \varphi(t)\, \varphi(2t - n)\, dt = \begin{cases} 1/\sqrt{2} & \text{si } n = 0, 1 \\ 0 & \text{sinon,} \end{cases}$$

d'où l'on déduit que

$$g[n] = (-1)^n\, h[1-n] = \begin{cases} +1/\sqrt{2} & \text{si } n = 0 \\ -1/\sqrt{2} & \text{si } n = 1 \\ 0 & \text{sinon.} \end{cases}$$

Par suite, l'ondelette associée est telle que $\psi(t) = \varphi(2t) - \varphi(2t-1)$, soit encore

$$\psi(t) = \begin{cases} +1 & \text{si } 0 \leq t < 1/2 \\ -1 & \text{si } 1/2 \leq t < 1 \\ 0 & \text{sinon,} \end{cases}$$

où l'on reconnaît l'ondelette de Haar.

Daubechies — L'ondelette de Haar présente l'avantage d'être à support compact, mais l'inconvénient d'être peu régulière. On peut donc se poser la question de trouver des ondelettes qui possèderaient la propriété d'être à support compact (en ne mettant en jeu qu'un nombre fini de coefficients dans les filtres associés) tout en ayant de meilleures propriétés de régularité. Pour ce faire, il est naturel de partir de fonctions d'échelle qui soient elles-mêms à support compact. Une conséquence directe de ce choix est que

$$H(f) = \frac{1}{\sqrt{2}} \sum_{n=-\infty}^{\infty} h[n]\, e^{i2\pi f n}$$

devient un *polynôme trigonométrique*, 1-périodique et devant satisfaire la relation "miroir en quadrature" $|H(f)|^2 + |H(f+1/2)|^2 = 1$. Si l'on demande en outre que la solution possède une certaine régularité, le problème est davantage contraint. En effet, la régularité d'une ondelette dépend étroitement de ses propriétés de *cancellation* (c'est-à-dire du nombre de ses *moments nuls*), en ce sens que l'appartenance de $\psi(t)$ à la classe C^r est liée à la propriété

$$\int_{-\infty}^{+\infty} t^k\, \psi(t)\, dt = 0 \Leftrightarrow \frac{d^k \Psi}{df^k}(0) = 0; k = 0, \ldots r.$$

En tenant compte de la relation "miroir en quadrature", on voit que le filtre (passe-bas) $H(f)$ correspondant à une ondelette à r moments nuls possède nécessairement un zéro de multiplicité r à la fréquence discrète $f = 1/2$. Par suite, il admet une factorisation du type

$$H(f) = \left(\frac{1 + e^{i2\pi f}}{2}\right)^r L(f),$$

où $L(f)$ est une fonction 1-périodique, de classe C^r si $H(f)$ l'est. Dans le cas d'une réponse impulsionnelle finie, $L(f)$ est de plus un polynôme trigonométrique.

La relation de quadrature ne faisant intervenir en fait que le module carré de $H(f)$, on peut alors poser

$$|H(f)|^2 = (\cos^2 \pi f)^r \, P(\sin^2 \pi f),$$

où le polynôme P est solution de l'équation

$$(1-x)^r \, P(x) + x^r \, P(1-x) = 1.$$

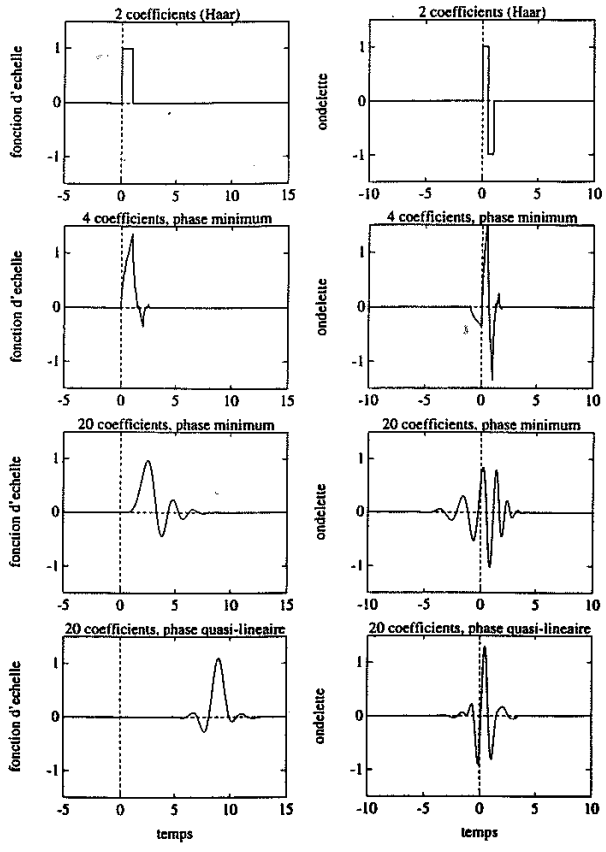

FIGURE 8 – Exemples d'ondelettes de Daubechies à support compact.

Ingrid Daubechies a montré qu'une telle solution était donnée par

$$P(x) = \sum_{k=0}^{r-1} \left(\begin{array}{c} r-1+k \\ k \end{array} \right) x^k,$$

ce qui, après factorisation spectrale de la fonction $|H(f)|^2$ résultante, apporte une solution au problème posé en définissant une ondelette à support compact à $2r$ coefficients non nuls et dont la régularité augmente linéairement avec r.

Un choix communément retenu pour la factorisation spectrale consiste à imposer que tous les zéros de la fonction de transfert soient à l'intérieur du cercle unité, ce qui correspond à la solution dite *à minimum de phase*. Des exemples d'ondelettes de Daubechies à support compact sont donnés en Figure 8. On peut remarquer que le choix le plus simple, à savoir $r = 1$, redonne l'ondelette de Haar.

3.4 Sélection et parcimonie

Lorsqu'on s'intéresse à la décomposition d'un signal à l'aide d'une TFCT ou d'une transformée en ondelettes, on fait classiquement le choix d'une fonction-gabarit unique (une fenêtre ou une ondelette), celle-ci étant laissée à la discrétion de l'utilisateur. Si l'objectif est d'obtenir une représentation compacte dans laquelle un petit nombre seulement de valeurs seraient significativement non nulles, l'intuition est d'*adapter* la fonction de référence au signal analysé, ou encore de simplifier la représentation obtenue avec une famille "quelconque" en en regroupant et/ou en en seuillant les termes. Ceci peut se décliner de plusieurs façons.

3.4.1 Paquets d'ondelettes et cosinus locaux

Le principe d'une décomposition en ondelettes orthogonales est d'itérer l'action parallèle d'un filtre passe-bas et d'un filtre passe-haut, celle-ci n'intervenant à une étape donnée que sur la sortie passe-bas de l'étape précédente. Il en résulte un découpage passe-bande du spectre du signal, dont la largeur ne peut décroître qu'en allant vers les basses fréquences. Le principe des *paquets d'ondelettes* est de généraliser cette décomposition en itérant à chaque échelle la découpe passe-haut/passe-bas sur à la fois l'approximation (filtrée passe-bas) et le détail (filtrée passe-haut). Il en résulte un arbre complet de décomposition dont la terminaison possible à un nœud quelconque

fournit autant de décompositions pouvant être attachées chacune à une base orthonormée.

3.4.2 Reconstruction seuillée

Supposons que l'on dispose d'une représentation de la forme

$$f[n] = \sum_{m=0}^{\infty} \langle f, g_m \rangle g_m[n],$$

construite par projections sur une famille orthogonales de fonctions de base $\{g_m[n], m \in \mathbb{Z}\}$ (par exemple, des ondelettes). En se restreignant aux M premiers termes, on obtient l'approximation

$$f_M[n] = \sum_{m=0}^{M-1} \langle f, g_m \rangle g_m[n],$$

avec l'erreur associée

$$e_M[n] = f[n] - f_M[n] = \sum_{m=M}^{\infty} \langle f, g_m \rangle g_m[n].$$

L'orthogonalité de la représentation garantit que

$$\|f - f_M\|^2 = \sum_{m=M}^{\infty} |\langle f, g_m \rangle|^2 ,$$

d'où il suit évidemment que l'erreur décroît lorsque M augmente, mais d'une façon qui dépend de la vitesse de décroissance des termes $|\langle f, g_m \rangle|$ en fonction de m. Ainsi, l'indexation des termes de la somme prend une importance particulière dans la mesure où, pour un même nombre de termes retenus dans l'approximation, le choix de ceux-ci pourra résulter en des erreurs très différentes. De ce point de vue, il est facile de se convaincre que prendre pour M premiers termes les M premières composantes obtenues dans l'ordre "naturel" d'une décomposition (par exemple, celui des fréquences croissantes dans le cas de Fourier) a toutes chances de conduire à une approximation s'appuyant sur beaucoup de valeurs dont le faible poids ne fait décroître que lentement l'erreur. En contraste avec cette perspective *linéaire*, il est alors possible d'envisager une contrepartie *non linéaire* (et, *de facto*, pilotée par les

données) dans laquelle seraient retenues préférentiellement les valeurs dont les poids sont les plus forts.

Pour ce faire, on convient de choisir comme approximation à M termes l'expression

$$f_M[n] = \sum_{m \in \Lambda} \langle f, g_m \rangle g_m[n],$$

en définissant l'ensemble Λ de taille $|\Lambda| = M$ (le "budget" de l'approximation) selon

$$\Lambda = \{m \,|\, |\langle f, g_m \rangle| \geq T(M)\},$$

où $T(M)$ est un *seuil* à fixer de telle sorte que

$$\|f - f_M\|^2 = \sum_{m \notin \Lambda} |\langle f, g_m \rangle|^2 \to \min .$$

Si on classe les coefficients de la décomposition par amplitude décroissante selon :

$$\left|\langle f, g_{m_{k+1}} \rangle\right| \leq \left|\langle f, g_{m_k} \rangle\right| ,$$

l'approximation devient

$$f_M[n] = \sum_{k=1}^{M} \langle f, g_{m_k} \rangle g_{m_k}[n],$$

avec, pour l'erreur associée,

$$\|f - f_M\|^2 = \sum_{k \geq M+1} \left|\langle f, g_{m_k} \rangle\right|^2 .$$

Ainsi, une décroissance des coefficients telle que

$$\left|\langle f, g_{m_{k+1}} \rangle\right| = O\left(k^{-\alpha}\right)$$

se traduit par une décroissance de l'erreur de la forme

$$\|f - f_M\|^2 = O\left(M^{1-2\alpha}\right) .$$

3.4.3 Poursuite adaptative

Le principe de la *poursuite adaptative* (ou "matching pursuit") est de considérer un signal comme représentable au moyen d'un dictionnaire redondant de formes élémentaires et de sélectionner itérativement les éléments les plus pertinents pour sa description. De manière plus précise, on se donne un dictionnaire $\mathcal{D} = \{\phi_p(t); p = 1, \dots P\}$ de P formes d'ondes, que l'on suppose complet. On projette alors le signal à décomposer $x(t)$ sur un des vecteurs $\phi_{p_0}(t)$ de ce dictionnaire, ce qui conduit à une représentation de la forme

$$x(t) = \langle x, \phi_{p_0} \rangle \, \phi_{p_0}(t) + r(t),$$

où $r(t)$ est le *résidu* de la décomposition. Celui-ci étant orthogonal au vecteur choisi $\phi_{p_0}(t)$, on a

$$\|x\|^2 = |\langle x, \phi_{p_0} \rangle|^2 + \|r\|^2$$

et, de façon à minimiser l'importance énergétique de ce résidu, le vecteur de projection à retenir doit être celui maximisant la corrélation avec le signal :

$$\phi_{p_0}(t) = \arg\max_p |\langle x, \phi_p \rangle|.$$

Ce choix étant fait, on peut considérer le résidu $r(t)$ comme un nouveau signal auquel on applique la même procédure, sélectionnant le second vecteur de $\mathcal{D}$ partageant la plus grande ressemblance (au sens de la corrélation) avec le signal analysé, et itérer. Ainsi, si l'on part de l'initialisation $r^{(0)}(t) = x(t)$, on peut écrire, pour tout ordre $m \geq 0$,

$$r^{(m)}(t) = \langle r^{(m)}, \phi_{p_m} \rangle \, \phi_{p_m}(t) + r^{(m+1)}(t),$$

en ayant choisi $\phi_{p_m}(t)$ tel que

$$\phi_{p_m}(t) = \arg\max_p |\langle r^{(m)}, \phi_p \rangle|.$$

Opérant de proche en proche, on obtient, pour une profondeur M de décomposition, la représentation

$$x(t) = \sum_{m=0}^{M-1} \langle r^{(m)}, \phi_{p_m} \rangle \, \phi_{p_m}(t) + r^{(M)}(t),$$

l'énergie totale du signal analysé se décomposant elle-même selon

$$\|x\|^2 = \sum_{m=0}^{M-1} |\langle r^{(m)}, \phi_{p_m} \rangle|^2 + \| \, r^{(M)}\|^2$$

dans la mesure où

$$\|r^{(m)}\|^2 = |\langle r^{(m)}, \phi_{p_m} \rangle|^2 + \| \, r^{(m+1)}\|^2.$$

Il suit de cette relation de conservation de l'énergie que

$$\begin{aligned} \frac{\| \, r^{(m+1)}\|^2}{\|r^{(m)}\|^2} &= 1 - \left| \left\langle \frac{r^{(m)}}{\|r^{(m)}\|}, \phi_{p_m} \right\rangle \right|^2 \\ &\leq 1 - \mu^2(r^{(m)}, \mathcal{D}), \end{aligned}$$

où $\mu(r, \mathcal{D})$ mesure la *cohérence* entre un vecteur et l'ensemble du dictionnaire :

$$\mu(r, \mathcal{D}) := \max_p \left| \left\langle \frac{r}{\|r\|}, \phi_p \right\rangle \right|.$$

Comme on a $\mu(r, \mathcal{D}) \leq 1$, il suit que la convergence de la poursuite adaptative est *exponentielle*.

Le résultat est bien sûr dépendant du choix du dictionnaire et de son adaptation au signal analysé. À la limite, un dictionnaire qui contiendrait le signal conduirait à la représentation la plus parcimonieuse qui soit, puisqu'un seul coefficient serait non nul. À l'autre extrême, un dictionnaire qui serait totalement incohérent avec le signal résultera en un grand nombre de coefficients non nuls. Il y a donc un compromis à régler entre la spécificité et la généralité d'un dictionnaire. Un choix possible est de considérer la famille à 3 paramètres

$$\psi_{t,f,a}(s) = \frac{1}{\sqrt{a}} \, \psi \left(\frac{s-t}{a} \right) \, e^{i2\pi f s},$$

combinant les caractéristiques d'une TFCT et d'une TOC.

3.5 Distributions quadratiques

Si, plus qu'une représentation reposant sur une décomposition du signal lui-même, on souhaite obtenir une distribution de son énergie, un spectogramme n'est pas la seule possibilité. En effet, il est naturel qu'une distribution d'énergie soit quadratique, mais la classe des transformations quadratiques ne se réduit pas à celle du module carré des transformées linéaires.

La question qui se pose alors est de définir une quantité $\rho_x(t,f)$ telle que

$$E_x = \iint_{-\infty}^{+\infty} \rho_x(t,f)\,dt\,df,$$

ce qui peut donner lieu à un grand nombre d'approches différentes.

Parmi celles-ci, une possibilité de construction de classes de solutions consiste à imposer une structure *a priori* très générale et à en déduire des paramétrisations plus restrictives par l'imposition progressive de contraintes jugées "naturelles". On peut partir pour cela d'une forme bilinéaire en le signal,

$$\rho_x(t,f) = \iint_{-\infty}^{+\infty} K(s,s';t,f)\,x(s)\,x^*(s')\,ds\,ds',$$

caractérisée par un noyau dépendant *a priori* de quatre variables indépendantes, avec de plus la contrainte

$$\iint_{-\infty}^{+\infty} K(s,s';t,f)\,dt\,df = \delta(s-s')$$

de façon à garantir à la forme bilinéaire de définir une distribution d'énergie. Ceci étant posé, il suffit d'imposer des contraintes additionnelles de *covariances* pour réduire l'espace des solutions admissibles. D'une manière raccourcie, ceci revient à demander que soit satisfaite l'équation

$$\rho_{\mathbf{T}x}(t,f) = (\tilde{\mathbf{T}}\rho)_x(t,f),$$

dans laquelle $\mathbf{T} : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ représente un opérateur de transformation (et $\tilde{\mathbf{T}} : L^2(\mathbb{R}^2) \to L^2(\mathbb{R}^2)$ l'opérateur associé agissant dans le plan), c'est-à-dire à imposer que la distribution recherchée "suive" le signal dans les transformations qu'il subit.

L'exemple le plus simple est celui des *translations* en temps et en fréquence, pour lesquelles le principe de covariance conduit le noyau général à prendre la forme particulière

$$K(s,s';t,f) = K_0(s-t,s'-t)\,e^{-i2\pi f(s-s')},$$

où $K_0(s,s')$ est une fonction arbitraire ne dépendant plus que de *deux* variables [6]. On obtient ainsi une classe générale de solutions, appelé *classe de*

6. Cette situation est évidemment à rapprocher de la covariance par les seules translations temporelles, qui transforme un opérateur linéaire quelconque en un *filtre* linéaire dont le noyau est convolutif et ne dépend que d'*une* variable.

Cohen et définie par :

$$C_x(t,f) := \iiint_{-\infty}^{+\infty} \varphi(\xi,\tau)\, x\left(s+\frac{\tau}{2}\right)\, x^*\left(s-\frac{\tau}{2}\right)\, e^{i2\pi[\xi(s-t)-f\tau]}\, ds\, d\xi\, d\tau,$$

où $\varphi(\xi,\tau)$ est une fonction de paramétrisation "arbitraire" dont la spécification peut résulter de contraintes additionnelles.

Une façon plus simple de ré-écrire la classe de Cohen est de la voir comme une convolution 2D, et donc comme la transformée de Fourier 2D de la quantité $\varphi(\xi,\tau)A_x(\xi,\tau)$, où

$$A_x(\xi,\tau) := \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right) x^*\left(t-\frac{\tau}{2}\right) e^{i2\pi\xi t}\, dt$$

est une transformée intrinsèque au signal, appelée *fonction d'ambiguïté* et dont l'interprétation est d'être la mesure d'une corrélation temps-fréquence au sens où (à une phase près) :

$$A_x(\xi,\tau) = \langle x, x_{\tau\xi}\rangle.$$

Ce point de vue permet d'introduire la classe de Cohen sur la base d'une heuristique très naturelle. En effet, si la transformation de Fourier met en dualité les variables de temps et de fréquence, il est également notoire que, d'un point de vue *énergétique* ou de *puissance*, elle met en dualité les concepts de *distribution d'énergie* et de *fonction de corrélation*. Ceci est matérialisé par les relations de type "Wiener-Khintchine" selon lesquelles une densité spectrale (d'énergie ou de puissance) $\Gamma_x(f)$ est image de Fourier d'une fonction de corrélation (déterministe ou aléatoire) $\gamma_x(\tau)$ selon

$$\Gamma_x(f) = \int_{-\infty}^{+\infty} \gamma_x(\tau)\, e^{-i2\pi f\tau}\, d\tau,$$

la notion même de corrélation étant relative, dans les deux cas, à une interaction entre le signal et ses translatées en temps. Du point de vue de l'estimation (dans le cas aléatoire), une méthode comme le *corrélogramme* réalise alors une transformation de Fourier *pondérée* (par une fenêtre $w(.)$) d'une estimée $\hat{\gamma}_x(\tau)$ de la fonction de corrélation aléatoire, telle qu'elle peut être fournie par une corrélation déterministe :

$$\hat{\Gamma}_x(f) = \int_{-\infty}^{+\infty} w(\tau)\, \hat{\gamma}_x(\tau)\, e^{-i2\pi f\tau}\, d\tau.$$

Le schéma évoqué ci-dessus prend tout son intérêt dans le cas des signaux stationnaires pour lesquels la description spectrale ne change pas au cours du temps. Si l'on s'intéresse par contre au cas non stationnaire et si l'on admet d'interpréter une distribution temps-fréquence comme une densité spectrale évolutive, il devient naturel, pour en construire une définition, de recourir à l'approche stationnaire en lui rajoutant une variable d'évolution. On est ainsi amené à généraliser la notion de corrélation au temps *et* à la fréquence, ce qui correspond en fait à considérer la *fonction d'ambiguïté* définie précédemment. Il est alors remarquable de constater que l'extension temps-fréquence de la procédure stationnaire (corrélation suivie d'une pondération et d'une transformation de Fourier) conduit à calculer une estimée de la forme

$$\iint_{-\infty}^{+\infty} \varphi(\xi,\tau)\, A_x(\xi,\tau)\, e^{i2\pi(\xi t+\tau f)}\, d\xi\, d\tau,$$

soit exactement la classe de Cohen.

Dans le cadre de l'approche envisagée, il n'y a donc pas de solution unique pour la définition d'une distribution d'énergie temps-fréquence, mais autant que de choix possibles de la fonction de paramétrisation $\varphi(\xi,\tau)$. On peut néanmoins faire ressortir de cette infinité potentielle au moins quelques cas particuliers.

Distribution de Wigner-Ville — La paramétrisation la plus simple que l'on puisse imaginer est $\varphi(\xi,\tau) = 1$. Il est facile de voir que ceci conduit à la distribution spécifique

$$W_x(t,f) = \int_{-\infty}^{+\infty} x\left(t+\frac{\tau}{2}\right)\, x^*\left(t-\frac{\tau}{2}\right)\, e^{-i2\pi f\tau}\, d\tau,$$

appelée *Distribution de Wigner-Ville* (DWV). Cette quantité apparaît ainsi comme étant la transformée de Fourier 2D de la fonction d'ambiguïté, justifiant que la classe de Cohen puisse s'exprimer de façon équivalente selon :

$$C_x(t,f;\varphi) = \iint_{-\infty}^{+\infty} W_x(\tau,\xi)\, \Pi(\tau-t,\xi-f)\, d\tau\, d\xi,$$

en notant $\Pi(t,f)$ la transformée de Fourier (inverse) 2D de la fonction de paramétrisation $\varphi(\xi,\tau)$.

La DWV présente un intérêt majeur en analyse temps-fréquence car elle possède un grand nombre de propriétés intéressantes. Parmi celles-ci, on peut

citer le fait qu'elle ait des "marginales correctes", c'est-à-dire que

$$\int_{-\infty}^{+\infty} W_x(t,f)\,df = |x(t)|^2 \quad ; \quad \int_{-\infty}^{+\infty} W_x(t,f)\,dt = |X(f)|^2,$$

soit encore que l'intégration relativement à une variable fournisse la densité d'énergie dans le domaine de l'autre variable.

L'expression "marginales correctes" est en fait due à l'analogie que l'on peut établir entre une DWV (supposée normalisée) et une densité de probabilité. Quoique cette analogie ne soit pas strictement correcte dans la mesure où la DWV peur prendre localement des valeurs négatives, elle permet de justifier une autre propriété selon laquelle la *fréquence instantanée* et le *retard de groupe* d'un signal peuvent s'obtenir comme moments locaux de la DWV :

$$\frac{\int_{-\infty}^{+\infty} f\,W_x(t,f)\,df}{|x(t)|^2} = f_x(t) \quad ; \quad \frac{\int_{-\infty}^{+\infty} t\,W_x(t,f)\,dt}{|X(f)|^2} = t_x(f),$$

avec

$$f_x(t) := \frac{1}{2\pi}\,\frac{d}{dt}\arg x(t) \quad ; \quad t_x(f) := -\frac{1}{2\pi}\,\frac{d}{df}\arg X(f).$$

En effet, on sait que pour une densité conjointe $p(u,v)$, la relation de Bayes la relie aux densités individuelles $p(u)$ et $p(v)$ d'une part, et aux densités conditionnelles $p(u|v)$ et $p(v|u)$ d'autre part, selon :

$$p(u,v) = p(u|v)\,p(v) = p(v|u)\,p(u).$$

Il suit que la *moyenne conditionnelle* (par exemple celle de u à v fixé) s'écrit

$$\mathbb{E}\{u|v\} = \int_{-\infty}^{+\infty} u\,p(u|v)\,du = \frac{\int_{-\infty}^{+\infty} u\,p(u,v)\,du}{p(v)},$$

ce qui est en accord, dans le cas temps-fréquence, avec l'interprétation de la fréquence instantanée comme la moyenne locale de la distribution conjointe.

Sans entrer dans les détails de toutes les propriétés intéressantes de la DWV, on pourra noter qu'elle conserve le support du signal analysé (en temps et en fréquence), qu'elle est transformée de façon naturelle par les filtrages linéaires et les modulations, et qu'elle est covariante par rapport à des opérations comme les changements d'échelle. Enfin, il est intéressant de remarquer que, dans le cas d'un "chirp linéaire"

$$c(t) = \exp\left\{i2\pi\left(\frac{\alpha}{2}t^2 + \beta t + \gamma\right)\right\},$$

on a la propriété de localisation exacte

$$W_c(t,f) = \delta\left(f - f_c(t)\right),$$

où $f_c(t) = \alpha t + \beta$ est la loi de modulation de fréquence (linéaire) du "chirp". D'une certaine façon, on pourrait ainsi penser que la DWV se comporte vis-à-vis des "chirps" linéaires comme la transformée de Fourier vis-à-vis des fréquences pures. Si ceci est vrai pour un "chirp" isolé, on reviendra plus loin sur la limitation de cette interprétation dans des cas plus généraux.

Enfin, la DWV est également "unitaire" au sens où elle vérifie la relation de pseudo-conservation du produit scalaire :

$$\iint_{-\infty}^{+\infty} W_x(t,f)\, W_y(t,f)\, dt\, df = \left| \int_{-\infty}^{+\infty} x(t)\, y^*(t)\, dt \right|^2,$$

ce que l'on peut écrire de façon plus compacte

$$\langle\!\langle W_x, W_y \rangle\!\rangle = |\langle x, y \rangle|^2$$

avec une notation évidente.

Spectrogramme — Le spectrogramme étant une distribution d'énergie quadratique, il est nécessairement dans la classe de Cohen. On peut alors montrer qu'il est associé à la paramétrisation $\varphi(\xi,\tau) = A_h^*(\xi,\tau)$, c'est-à-dire la fonction d'ambiguïté de la fenêtre à court-terme utilisée. Contrairement à la DWV, le spectrogramme ne satisfait pas les propriétés de marginales exactes, de moments locaux, de conservation de supports, de covariances générales et de localisation sur les "chirps" linéaires. Une façon de le comprendre est de voir le spectrogramme comme une version *lissée* de la DWV. En effet, le spectrogramme de fenêtre $h(t)$ peut s'écrire :

$$S_x^{(h)}(t,f) = |\langle x, \mathbf{T}_{tf} h \rangle|^2$$

en notant $\mathbf{T}_{tf}$ l'opérateur de translations en temps et en fréquence. Il suit alors de l'unitarité de la DWV que

$$S_x^{(h)}(t,f) = \langle\!\langle W_x, W_{\mathbf{T}_{tf} h} \rangle\!\rangle,$$

soit encore

$$S_x^{(h)}(t,f) = \langle\!\langle W_x, \tilde{\mathbf{T}}_{tf} W_h \rangle\!\rangle$$

en notant $\tilde{\mathbf{T}}_{tf}$ l'opérateur de translations en temps et en fréquence opérant dans le plan sur la DWV, cette égalité étant obtenue du fait que la DWV est, en tant que membre de la classe de Cohen, covariante par les translations. On a ainsi :

$$S_x^{(h)}(t,f) = \iint_{-\infty}^{+\infty} W_x(\tau,\xi)\, W_h(\tau - t, \xi - f)\, d\tau\, d\xi,$$

le caractère passe-bas de la fenêtre $h(t)$ assurant que la relation de corrélation bidimensionnelle (qui existe pour toute distribution de la classe de Cohen) met en jeu un noyau qui est lui-même passe-bas dans le plan temps-fréquence, correspondant ainsi à une opération de lissage. À titre d'exemple, le choix d'une fenêtre gaussienne permet d'écrire :

$$h(t) \propto \exp\left\{-\pi \frac{t^2}{T^2}\right\} \Rightarrow W_h(t,f) \propto \exp\left\{-2\pi\left(\frac{t^2}{T^2} + T^2 f^2\right)\right\}.$$

Par construction, un spectrogramme n'est donc pas une transformée intrinsèque au signal car il dépend d'une quantité externe que l'on peut voir comme un *instrument de mesure* : la fenêtre $h(t)$ devant être fixée par l'utilisateur. Il en résulte un effet d'interaction "objet-mesure", dont le résultat est de mélanger les propriétés que l'on cherche à mettre en évidence dans un signal avec la façon de le "voir" à travers $h(t)$. Ainsi, la relation de lissage établie plus haut montre que, à l'opposé de la DVW, un spectrogramme ne peut être localisé parfaitement sur la droite de fréquence instantanée d'un "chirp" linéaire (cf. Figure 9).

Distributions pseudo-Wigner-Ville lissées — Si, par construction, la DWV est associée dans la classe de Cohen à un noyau temps-fréquence parfaitement local ($\Pi(t,f) = \delta(t)\,\delta(f)$), le spectrogramme résulte quant à lui de l'application d'un lissage contraint par les relations d'incertitude et ne pouvant donc être réduit dans la direction d'une des variables sans être augmenté dans la direction de l'autre (une façon de le justifier est de remarquer que la DWV définissant le noyau de lissage ($\Pi(t,f) = W_h(t,f)$ a mêmes supports en temps et en fréquence que la fenêtre $h(t)$). Entre ces deux extrêmes (pas de lissage et lissage incompressible), il est néanmoins possible d'imaginer une transition de la forme :

$$\delta(t)\,\delta(f) \rightarrow g(t)\,W(f) \rightarrow W_h(t,f),$$

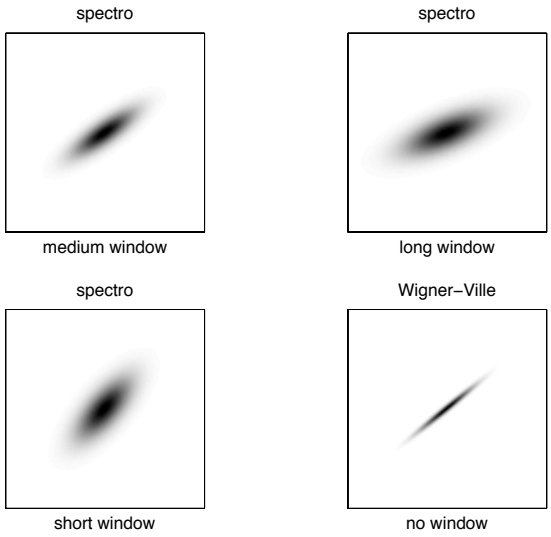

FIGURE 9 – Spectrogramme vs. Wigner-Ville. Dans le cas d'un "chirp" linéaire, la DWV se localise parfaitement sur la droite de fréquence instantanée, alors qu'un spectrogramme "étale" l'énergie du signal dans un domaine temps-fréquence qui dépend du choix de la fenêtre à court-terme utilisée.

dans laquelle le choix d'une fonction séparable ($\Pi(t,f) = g(t)\,W(f)$ permet un contrôle indépendant dans les deux directions.

La classe des distributions résultante, qui prend la forme

$$C_x(t,f;H.w) = \int_{-\infty}^{+\infty} w(s) \left[\int_{-\infty}^{+\infty} g(\tau - t)\, x\left(\tau + \frac{s}{2}\right)\, x^*\left(\tau - \frac{s}{2}\right)\, d\tau\right] e^{-i2\pi(\xi - f)s}\, ds,$$

est appelée classe des *Distributions Pseudo-Wigner-Ville Lissées* (DPWVL).

3.6 Distributions dépendantes du signal

Si l'on s'en tient à la classe de Cohen, le choix d'une paramétrisation est laissé à la discrétion de l'utilisateur en fonction d'un certain nombre de propriétés que celui-ci souhaite voir satisfaites, mais il est *a priori* fixe, quelque soit le signal analysé. Indépendamment des contraintes auxquelles elle peut être associée, la paramétrisation admet aussi une interprétation géométrique

que l'on peut rattacher à la lisibilité de la représentation résultante dans le plan.

3.6.1 Wigner-Ville comme TFCT adaptée

Une TFCT de fenêtre $h(t)$ est la filtrée linéaire du signal analysé selon l'opération

$$F_x^{(h)}(t,f) = \int_{-\infty}^{+\infty} x(s)\, h^*(s-t)\, e^{-i2\pi fs}\, ds.$$

La fenêtre $h(t)$ étant *a priori* arbitraire et le résultat dépendant de celle-ci, on peut se poser la question de laquelle choisir, à signal donné. Un argument de type "filtrage adapté" suggère alors de choisir pour fenêtre l'*image retournée en temps* du signal analysé, c'est-à-dire de fixer $h(t) = x_-(t) := x(-t)$. Ce faisant, il est immédiat de voir que

$$F_x^{(x_-)}(t,f) = \frac{1}{2}\, W_x\left(\frac{t}{2}, \frac{f}{2}\right)\, e^{-i\pi ft}.$$

Ainsi, la DWV apparaît comme étant (à une phase et une dilatation près) une "TFCT adaptée", l'opération d'adaptation présentant l'avantage de rendre la transformation *intrinsèque* au signal et l'inconvénient de lui faire perdre son caractère linéaire.

3.6.2 Géométrie

La première remarque que l'on peut faire quant aux distributions d'énergie comme celles de la classe de Cohen est, étant quadratiques par nature, elles satisfont un principe de superposition lui-même quadratique, selon lequel la représentation de la somme de deux signaux ajoute à la somme des représentations individuelles un terme d'interaction [7], à la façon dont on a l'identité élémentaire $(a+b)^2 = a^2 + b^2 + 2ab$.

Pour comprendre plus précisément le principe de création de ces termes supplémentaires, on peut revenir à l'expression de la classe de Cohen dans le plan des ambiguïtés, à savoir $\varphi(\xi,\tau)\, A_x^*(\xi,\tau)$, mettant ainsi l'accent sur une opération de *pondération* agissant sur la quantité intrinsèque au signal

7. On peut remarquer cette situation est déjà vraie pour une densité spectrale, justifiant du principe de toute méthode interférentielle.

que constitue sa fonction d'ambiguïté. cette dernière étant l'analogue temps-fréquence d'une fonction de corrélation ordinaire, elle en possède plusieurs propriétés :

1. elle est de module maximal à l'origine du plan (ce qui traduit simplement le fait qu'un signal ne peut pas être plus ressemblant à un autre signal qu'à lui-même) ;
2. une fonction de corrélation mesurant le degré de ressemblance d'un signal et des translatées, la présence éventuelle de plusieurs composantes distinctes conduit à la fois à des termes d'*auto-corrélation* (où chacun des termes mesure sa ressemblance par rapport à ses propres translatées) et d'*inter-corrélation* (où ce sont cette fois les ressemblances entre termes distincts, convenablement translatés, qui sont mesurées).

On voit ainsi que la signature "propre" des composantes se situe au voisinage de l'origine du plan, les contributions signant les interactions étant quant à elles rejetées à une distance de l'origine qui est de l'ordre de la distance temps-fréquence qui les sépare.

3.6.3 Noyaux optimaux

En se basant sur l'analyse précédente, il devient naturel, pour favoriser les termes propres par rapport aux termes interférentiels, de trouver une fonctionde paramétrisation $\varphi(\xi, \tau)$ qui soit *dépendante du signal*, en prenant comme point de départ l'interprétation des interférences dans le plan des ambiguïtés. Puisque les termes interférentiels à réduire sont, par construction, excentrés relativement à l'origine du plan des ambiguïtés, cela suggère d'imposer à la fionction cherchée de jouer un rôle de fonction de *pondération* (réelle et positive) ayant la propriété d'être radialement non croissante, c'est-à-dire d'être telle que, ses coordonnées étant mises sous forme polaire, on ait pour tout angle θ

$$f(r_1 \cos\theta, r_1 \sin\theta) \leq f(r_2 \cos\theta, r_2 \sin\theta)$$

si $r_1 \leq r_2$. Munis d'une telle contrainte, il devient possible de retenir comme *meilleure* pondération la solution du problème d'optimisation :

$$\varphi_*(\xi, \tau) = \max_{\varphi} \iint_{-\infty}^{+\infty} |\varphi(\xi, \tau)\, A_x((\xi, \tau)|^2 \; d\xi \, d\tau,$$

sous la contrainte de volume

$$\iint_{-\infty}^{+\infty} |\varphi(\xi,\tau)|^2 \, d\xi \, d\tau \leq \alpha,$$

où α est une borne fixée *a priori*. En effet, à volume fixé, la contrainte de non-croissance radiale tend à pénaliser les contributions excentrées de $|A_x((\xi,\tau)|^2$, puisque leur prise en compte revient à "perdre" la fraction de volume correspondant à l'intervalle entre auto- et inter-ambiguïtés. Ainsi, les contributions retenues préférentiellement par la procédure d'optimisation sont celles qui se concentrent autour de l'origine, c'est-à-dire celles qui sont représentatives prioritairement des termes "signal". Le choix de la borne α sur le volume contrôle évidemment le compromis entre résolution conjointe et réduction des interférences : on peut retenir comme ordre de grandeur la valeur $\alpha = 1$ qui correspond au cas du spectrogramme construit sur une fenêtre d'énergie unité. La résolution du problème d'optimisation peut prendre plusieurs formes, la plus populaire reposant sur l'emploi d'un modèle *gaussien radial* pour la fonction de paramétrisation ? Dans tous les cas, il est nécessaire de calculer explicitement la fonction d'ambiguïté complète du signal analysé, la représentation à interférences réduites s'en déduisant par transformation de Fourier inverse après pondération.

Une alternative à l'approche *globale* qui vient d'être décrite est de chiffrer l'adaptation d'un lissage à une représentation par l'intermédiaire d'une mesure de concentration *locale* dans le plan. Une telle mesure peut par exemple être donnée par le "kurtosis"

$$K(t,f;\Pi) := \frac{\iint_{-\infty}^{+\infty} |C_x(\tau,\xi;\varphi)\, w((\tau-t,\xi-f)|^4 \, d\tau \, d\xi}{\left(\iint_{-\infty}^{+\infty} |C_x(\tau,\xi;\varphi)\, w((\tau-t,\xi-f)|^2 \, d\tau \, d\xi\right)^2},$$

dans lequel $w(t,f)$ est une fonction temps-fréquence passe-bas assurant le caractère local de l'évaluation. Le lissage *localement optimal* se trouve alors défini par

$$\Pi_*(t,f) = \arg\max_{\Pi} K(t,f;\Pi).$$

3.6.4 Diffusion adaptative

Si l'on considère la *diffusion* d'une grandeur bidimensionnelle $U(v_1,v_2;\tau)$ (par exemple une concentration, dans l'espace indexé par les deux variables

v_1 et v_2) avec la condition initiale $U(v_1, v_2; 0) = U_0(v_1, v_2)$, on sait qu'elle suit l'équation de la chaleur

$$\frac{\partial U}{\partial \tau} = \Delta U,$$

dont la solution est telle que

$$U(v_1, v_2; \tau) = (G_\tau \star U_0)(v_1, v_2)$$

avec

$$G_\tau(v_1, v_2) := \frac{1}{4\pi\tau} \exp\left\{-\frac{1}{4\tau}\left(v_1^2 + v_2^2\right)\right\}.$$

En d'autres termes, au temps τ de l'évolution de l'EDP, la concentration courante est liée à la concentration initiale par une relation de *lissage* assurée par un noyau *gaussien* dont l'étalement est d'autant plus grand que τ est plus élevé, la solution tendant vers une distribution uniforme de la concentration lorsque $\tau \to \infty$.

Cette remarque a été à la base d'une approche nouvelle en traitement d'images, le rôle de la concentration étant tenu par l'intensité. La forme isotrope et "universelle" du lissage gaussien est cependant insuffisante dans ce contexte, l'idée étant de pouvoir disposer d'un lissage *adaptatif* à même d'être important dans les zones "lisses" tout en étant inhibé au niveau des discontinuités de façon à pouvoir les préserver. Une manière d'atteindre cet objectif est d'utiliser l'dentité $\Delta U = \mathrm{div}(\nabla U)$ et d'y inclure une fonction de diffusion *locale* en lieu et place de la valeur scalaire (arbitrairement prise pour unité) intervenant dans l'équation, linéaire et homogène, de base. Ceci revient à considérer une nouvelle EDP, non linéaire et inhomogène, de la forme

$$\frac{\partial U(v_1, v_2; \tau)}{\partial \tau} = \mathrm{div}\left(\eta_U(v_1, v_2)\, \nabla U(v_1, v_2; \tau)\right)$$

dans laquelle la fonction (de "conductance") $\eta_U(v_1, v_2)$ est typiquement une fonction décroissante du module du gradient.

En se souvenant que le spectrogramme (que l'on peut voir comme une "image" temps-fréquence) résulte d'un lissage gaussien particulier de la DWV, on peut imaginer d'adopter une approche analogue pour lisser cette dernière de façon adaptive, avec l'idée de se rapprocher du spectrogramme dans les zones d'interférences (au prix donc d'un lissage fort) et de préserver la localisation de la DWV dans les zones propres (donc en maintenant un faible niveau de lissage). À cette fin, la non-linéarité à introduire ne peut pas directement se calquer sur le module du gradient, une "image" temps-fréquence

étant en fait une image particulière, dont la structuration implique de fortes fluctuations dans les zones interférentielles, qu'il convient de gommer et non de préserver. Prenant en compte cette situation, une possibilité est de partir de la condition initiale

$$\rho_x(t, f; 0) = W_x(t, f)$$

et de faire évoluer celle-ci selon l'EDP

$$\frac{\partial \rho_x(t, f; \tau)}{\partial \tau} = \operatorname{div} \left(\eta_x(t, f) \, \nabla \rho_x(t, f; \tau) \right),$$

dans laquelle la conductance est une fonction décroissance du spectrogramme $S_x(t, f)$, prenant par exemple la forme

$$\eta_x(t, f) = \left[1 + \left(\frac{S_x(t, f)}{\beta} \right)^{\alpha} \right]^{-1},$$

avec α et $\beta > 0$.

Comme on l'a dit précédemment, laisser la diffusion opérer à l'infini se terminerait par une distribution uniformément étalée. Il convient donc de l'arrêter à un "temps" optimal τ_* que l'on peut fixer par un critère entropique. En effet, si l'on introduit l'entropie (de Shannon) de la distribution au temps τ selon

$$H_{\rho_x}(\tau) := - \iint_{-\infty}^{+\infty} |\rho_x(t, f; \tau)| \, \log |\rho_x(t, f; \tau)| \, dt \, df,$$

on peut se convaincre que, pour une distribution composite (attachée à un signal multicomposantes), celle-ci passe par un minimum. En effet, dans le cas d'un signal de la forme

$$x(t) = \sum_{n=1}^{N} x_n(t),$$

une distribution quadratique $\rho_x(t, f)$ peut s'écrire comme la somme d'un terme "signal" $\rho_S(t, f)$ composé de la somme des N distributions individuelles $\rho_{x_n}(t, f)$, et d'un terme "interférentiel" $\rho_I(t, f)$ fait de toutes les distributions croisées pour $m \neq n$. Ces deux termes ont des comportements différents vis-à-vis du lissage et, partant, de leur contribution à l'entropie totale. En effet, un lissage croissant :

1. étale les composantes "signal", provoquant en retour une augmentation de $H_{\rho_S}(t, f)$;

2. réduit les oscillations des termes interférentiels, faisant ainsi décroître $H_{\rho_I}(t, f)$.

Ces variations antagonistes conduisent, lorsqu'on les superpose, à un passage par un minimum que l'on peut retenir comme valeur d'arrêt.

Tout en ayant été rendue inhomogène par l'introduction d'une conductance locale dépendant du signal, l'approche précédente reste isotrope, ce qui en fait une limitation dans le cas (fréquent) de structures "filaires" associées à des modulations de fréquences. Pour dépasser cette limitation, on notera qu'il est possible de généraliser encore l'approche en introduisant un tenseur de diffusion permettant d'adapter localement l'orientation et l'élongation du lissage gaussien.

3.7 Méthodes exploitant la phase du signal

3.7.1 Réallocation

La "réallocation" (*reassignment* en anglais) est une autre technique visant à améliorer la lisibilité d'une représentation temps-fréquence quadratique en réduisant la contribution des termes interférentiels sans sacrifier la localisation des termes signal. C'est encore une technique dépendante du signal, dans la mesure où elle procède par modification d'une distribution de référence en fonction de ses propriétés locales. Pour en expliquer de façon simple le fonctionnement, on peut (quoique la technique soit applicable beaucoup plus largement) se restreindre au cas du spectrogramme et rappeler que celui-ci est lié à la DWV par la relation de lissage :

$$S_x^{(h)}(t, f) = \iint_{-\infty}^{+\infty} W_x(s, \xi)\, W_h(s - t, \xi - f)\, ds\, d\xi.$$

Ainsi, le lissage a l'avantage de réduire les termes interférentiels qui sont, par nature, oscillants, mais il a aussi l'inconvénient d'étaler les termes signal. Cette même relation de lissage montre aussi que, lorsqu'on calcule le spectrogramme d'un signal, la valeur obtenue en un point (t, f) ne peut pas être considérée comme une mesure de l'énergie du signal uniquement en ce point. Bien au contraire, elle apparaît explicitement comme la somme de toutes les valeurs de la DWV contenues dans un domaine délimité par les largeurs temporelle et fréquentielle de la fenêtre d'analyse. C'est donc toute une distribution de valeurs qui est résumée en un seul nombre, celui-ci étant affecté au centre géométrique du domaine sélectionné.

Si l'on fait une analogie avec la mécanique, ceci revient à affecter la masse totale d'un solide à son *centre géométrique*. Un tel choix est en général complètement arbitraire, sauf lorsque la distribution est uniforme. Une autre possibilité, qui semble plus pertinente, consiste à affecter la masse totale au *centre de gravité* de la distribution. C'est exactement le sens de la réallocation. À chaque point du plan temps–fréquence où le spectrogramme est calculé, deux quantités complémentaires $\hat{t}(t,f)$ et $\hat{f}(t,f)$ sont également évaluées. Elles correspondent aux coordonnées du centre de gravité des valeurs de la DWV, pondérées par la fenêtre temps-fréquence de lissage (centrée en (t,f)) :

$$\hat{t}(t,f) = \frac{1}{S_x^{(h)}(t,f)} \iint_{-\infty}^{+\infty} s\, W_x(s,\xi)\, W_h(s-t,\xi-f)\, ds\, d\xi;$$

$$\hat{f}(t,f) = \frac{1}{S_x^{(h)}(t,f)} \iint_{-\infty}^{+\infty} \xi\, W_x(s,\xi)\, W_h(s-t,\xi-f)\, ds\, d\xi.$$

C'est alors au point $(\hat{t}(t,f), \hat{f}(t,f))$ que la valeur du spectrogramme calculée en (t,f) est ré-affectée, le spectrogramme réalloué étant défini par :

$$\hat{S}_x^{(h)}(t,f) = \iint_{-\infty}^{+\infty} S_x^{(h)}(s,\xi)\, \delta\left(t - \hat{t}_x(s,\xi), f - \hat{f}_x(s,\xi)\right)\, ds\, d\xi,$$

où $\delta(x,y)$ est l'impulsion de Dirac bidimensionnelle.

Conceptuellement, la réallocation peut donc être vue comme la seconde étape d'un processus en deux phases :

1. un *lissage*, qui permet de faire disparaître les termes d'interférence oscillants, mais qui élargit en contrepartie les composantes localisées,
2. une *compression*, qui vise à reconcentrer les contributions qui ont survécu au lissage.

On peut alors vérifier aisément que la propriété de localisation parfaite sur des droites du plan temps–fréquence, satisfaite par la DWV, sera conservée après le processus de réallocation, puisque le centre de gravité d'une distribution linéaire est nécessairement sur la droite support. Plus généralement, un résultat similaire peut être obtenu lorsque le signal analysé se comporte *localement* comme un chirp, le caractère local étant lié au support temps-fréquence de la fenêtre de lissage.

D'un point de vue pratique, le calcul des centres de gravité peut se faire de façon implicite en recourant à des TFCT additionnelles. En effet, si l'on

part par exemple du numérateur de l'expression définissant $\hat{t}_x(t,f)$, le calcul développé de la quantité :

$$J(s) := \int_{-\infty}^{+\infty} s\, W_x(s,\xi)\, W_h(s-t,\xi-f)\, d\xi$$

conduit à

$$J(s) = \int_{-\infty}^{+\infty} x\left(s+\frac{\tau}{2}\right) x^*\left(s-\frac{\tau}{2}\right) h^*\left(s+\frac{\tau}{2}-t\right) h\left(s-\frac{\tau}{2}-t\right) e^{-i2\pi f\tau} d\tau,$$

d'où, après le changement de variable (de Jacobien unité)

$$s+\frac{\tau}{2} = u \quad ; \quad s-\frac{\tau}{2} = v,$$

l'expression suivante pour le numérateur :

$$\int_{-\infty}^{+\infty} s\, J(s)\, ds = \iint_{-\infty}^{+\infty} \left(\frac{u+v}{2}\right) x(u) x^*(v) h^*(u-t) h(v-t) e^{-i2\pi f(u-v)} du\, dv,$$

soit encore :

$$\begin{aligned} \int_{-\infty}^{+\infty} s\, J(s)\, ds &= \mathrm{Re}\left\{ \left[\int_{-\infty}^{+\infty} u\, x(u) h^*(u-t) e^{-i2\pi f u} du \right] F_x^{(h)*}(t,f) \right\} \\ &= \mathrm{Re}\left\{ \left[t\, F_x^{(h)}(t,f) + F_x^{(th)}(t,f) \right] F_x^{(h)*}(t,f) \right\}, \end{aligned}$$

en notant $th(t) := t.h(t)$, d'où

$$\hat{t}_x(t,f) = t + \mathrm{Re}\left\{ \frac{F_x^{(th)}(t,f)}{F_x^{(h)}(t,f)} \right\}.$$

On montrerait de la même façon que

$$\hat{f}_x(t,f) = f - \mathrm{Re}\left\{ \frac{F_x^{(h')}(t,f)}{F_x^{(h)}(t,f)} \right\}$$

en introduisant la fenêtre auxiliaire $h'(t) := (dh/dt)(t)$.

On voit ainsi que, en comparaison du spectrogramme usuel qui nécessite le calcul d'une TFCT, l'évaluation du spectrogramme réalloué met en en jeu 3 TFCT dont les valeurs sont combinées (voire de 2 seulement dans le cas d'une fenêtre gaussienne, puisqu'alors les fonctions $h'(t)$ et $th(t)$ sont proportionnelles).

3.7.2 "Synchrosqueezing"

Si la réallocation permet d'améliorer la lisibilité d'un spectrogramme en garantissant en particulier une localisation parfaite sur toute droite du plan (donc aussi bien un "chirp" linéaire qu'une fréquence pure ou une impulsion), elle n'admet pas de formule d'inversion permettant de remonter au signal analysé. La raison première en est la caractère quadratique de la représentation avant réallocation, suggérant dans un premier de réallouer directement la TFCT (dont on sait qu'elle admet une reconstruction exacte) selon la définition modifiée

$$\tilde{S}_x^{(h)}(t,f) = \left|\hat{F}_x^{(h)}(t,f)\right|^2.$$

avec

$$\hat{F}_x^{(h)}(t,f) = \iint_{-\infty}^{+\infty} F_x^{(h)}(s,\xi)\,\delta\left(t-\hat{t}_x(s,\xi), f-\hat{f}_x(s,\xi)\right)\,ds\,d\xi,$$

Ceci ne résout cependant que partiellement le problème dans la mesure où inverser la transformation nécessiterait de connaître pour chaque valeur réallouée la localisation de ses antécédents. Une façon d'y parvenir est en fait de contraindre la réallocation à n'opérer que selon une direction du plan, par exemple la direction fréquentielle afin de "suivre" l'évolution filaire des trajectoires temps-fréquence de modulations de fréquence. On montre alors qu'il existe une formule de reconstruction exacte par intégration fréquentielle seule, à condition de choisir correctement la définition de la TFCT mise en jeu dans le calcul du spectrogramme. En effet, dans la mesure où il en est le module carré, un même spectrogramme est obtenu à partir d'une TFCT définie à une phase pure près. Lorsqu'on ne s'intéresse qu'au spectrogramme, l'usage veut que l'on adopte la définition classique utilisée jusqu'ici mais, si l'on souhaite garantir la reconstruction, il est préférable de lui substituer la définition :

$$F_x'^{(h)}(t,f) := \int_{-\infty}^{+\infty} x(s)\,h^*(s-t)\,e^{-i2\pi f(s-t)}\,ds$$

grâce à laquelle on a la formule de reconstruction unidimensionnelle :

$$\begin{aligned} m_x(t) &:= \int_{-\infty}^{+\infty} F_x^{\prime(h)}(t,f)\,df \\ &= \iint_{-\infty}^{+\infty} x(s)\,h^*(s-t)\,e^{-i2\pi f(s-t)}\,ds\,df \\ &= \int_{-\infty}^{+\infty} x(s)\,h^*(s-t)\,\left(\int e^{-i2\pi f(s-t)}\,df\right)\,ds \\ &= \int_{-\infty}^{+\infty} x(s)\,h^*(s-t)\,\delta(t-s)\,ds \\ &= x(t)\,h^*(0). \end{aligned}$$

3.8 Données substituts

3.8.1 De l'importance de la phase

Le spectre d'un signal $\{x(t), t \in \mathbb{R}\}$ ou $\{x[n], n \in \mathbb{Z}\}$, tel qu'il est défini par la transformation de Fourier $X(f)$, est une quantité à valeurs complexes qui peut se caractériser par deux grandeurs : un spectre d'amplitude $|X(f)|$ et un spectre de phase $\arg X(f)$. Le spectre d'amplitude permet d'identifier les fréquences servant à décrire un signal ainsi que le poids de leurs contributions respectives, alors que le spectre de phase porte la signature de comment ces différentes fréquences sont en relation entre elles, le déphasage d'une raie spectrale pure se ramenant à un décalage temporel. Ainsi, il existe une infinité de signaux ayant même spectre d'amplitude, ces signaux pouvant avoir une structure temporelle très différente suivant la forme de leurs spectres de phase. L'exemple le plus extrême de cette situation est sans doute celui qui différencie le bruit blanc (fluctuations désordonnées à tous les temps) et l'impulsion dite "de Dirac" (signal nul partout sauf à un instant). Si l'on ne s'en tient qu'à leurs composantes constitutives individuelles en termes d'amplitudes et de fréquences, rien ne distingue ces deux situations, l'une comme l'autre se caractérisant par un spectre continu et *plat* : le bruit blanc comme l'impulsion ont tous deux une représentation de Fourier dans laquelle toutes les fréquences possibles sont présentes, avec égale amplitude. Comme illustré en Figure 10, la différence majeure est que, dans le premier cas, les phases relatives de ces différentes fréquences sont distribuées aléatoirement, l'incohérence de celles-là se traduisant par la structure désordonnée de la superposition de celles-ci. Dans le deuxième cas au contraire, il existe un instant

particulier où toutes les composantes sont “en phase”, se renforçant mutuellement dans leur superposition et créant ainsi une localisation qu’aucune des composantes individuelles ne possède.

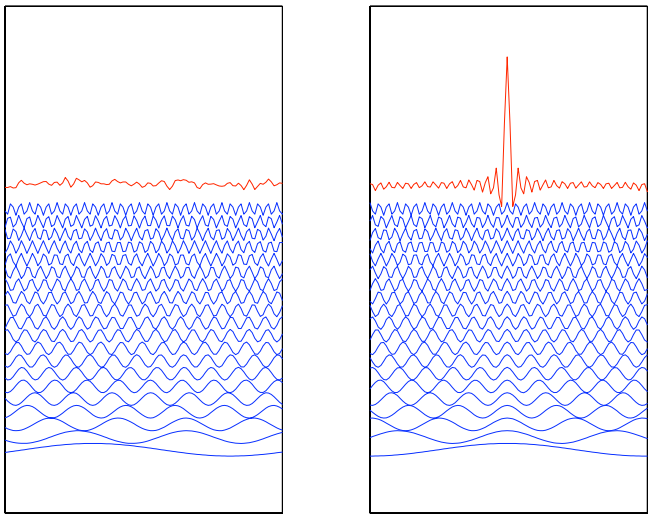

FIGURE 10 – Importance de la phase — Pour un même jeu de composantes de Fourier (fréquences d’égale amplitude), une répartition aléatoire des phases relatives conduit dans le domaine temporel à un signal de type bruit blanc (diagramme de gauche), alors que s’il existe un instant particulier où toutes les composantes sont en phase, leur superposition tend vers une impulsion de Dirac (diagramme de droite).

3.8.2 Randomisation simple

Pour un même spectre marginal sur le même intervalle d’observation, un processus *non stationnaire* diffère d’un processus *stationnaire* par une organisation structurée en temps qui se reflète dans la phase de son spectre. Un ensemble de J substituts (“surrogates”) peut alors être calculé de telle sorte que chacun d’eux ait même densité spectrale que le processus initial tout en ayant une phase désorganisée par une “randomisation”. À cette

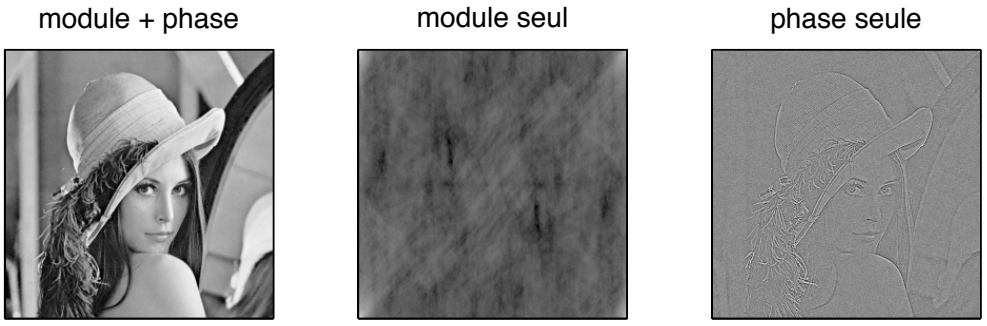

FIGURE 11 – Importance de la phase — Reconstruction.

fin, la première étape est de prendre la transformée de Fourier de l'observation, d'en conserver le module inchangé et de remplacer la phase par une séquence aléatoire, uniformément distribuée sur $[-\pi, \pi]$. La seconde étape est de prendre la transformée de Fourier inverse de ce spectre modifié. On montre alors que le substitut obtenu (que l'on peut avoir en nombre aussi grand que l'on opère de randomisations) est stationnaire au second ordre.

En effet, si l'on suppose que le signal est une observation à temps discret $(x[n], n = 1, \ldots, N)$, on peut lui associer une transformée de Fourier discrète $X[k]$ telle que

$$x[n] = \frac{1}{N} \sum_k X[k] e^{i2\pi nk/N}.$$

Exprimant $X[k]$ en termes de son module $A[k]$ et de sa phase $\Phi[k]$ selon $X[k] = A[k]e^{i\Phi[k]}$, un substitut $s[n]$ est construit à partir de $X[k]$ en remplaçant $\Phi[k]$ par une séquence de phase i.i.d. $\Psi[k]$ tirée selon une loi uniforme sur $[-\pi, \pi]$:

$$s[n] = \frac{1}{N} \sum_k A[k] e^{i\Psi[k]} e^{i2\pi nk/N},$$

d'où il suit que

$$s[n]s^*[m] = \frac{1}{N^2} \sum_k \sum_l A[k]A[l] e^{i(\Psi[k]-\Psi[l]+2\pi(nk-ml)/N)}.$$

En explicitant les contributions dans la double somme ci-dessus :

$$\sum_k \sum_l G[k,l] = \sum_k \left(G[k,k] + \sum_{l \neq k} G[k,l] \right)$$

et en évaluant les moyennes d'ensemble relatives aux quantités aléatoires, on obtient pour le premier terme :

$$\frac{1}{N^2} \sum_k \mathbb{E}\{A^2[k]\} e^{i2\pi(n-m)k/N}$$

et, pour le second :

$$\frac{1}{N^2} \sum_k \sum_{l \neq k} \mathbb{E}\{A[k]A[l]\} e^{i2\pi(nk-ml)/N} \mathbb{E}\{e^{i(\Psi[k]-\Psi[l])}\}.$$

Étant donné que les $\Psi[k]$ sont i.i.d. et uniformément distribués sur $[-\pi, \pi]$, leur différence a la distribution triangulaire :

$$\Lambda(\psi) = \frac{1}{4\pi^2}(1 - |\psi|/2\pi)$$

sur $[-2\pi, 2\pi]$, d'où :

$$\mathbb{E}\{e^{i(\Psi[k]-\Psi[l])}\} = \int_{-2\pi}^{2\pi} e^{i\psi} \Lambda(\psi) d\psi = 0.$$

Il suit que la fonction de covariance du substitut $s[n]$ se réduit au premier terme de la double somme précédente et, n'étant fonction que de la différence $n - m$, que $s[n]$ est stationnaire au second ordre [8].

3.8.3 Test de stationnarité

Une application possible de la technique des données substituts est de permettre un *test de stationnarité*.

Le terme "stationnarité" est couramment utilisé en traitement du signal et analyse de données mais, pris souvent dans une acception large, il peut correspondre à différentes qualités qui ne correspondent pas nécessairement

8. On pourrait montrer en fait que $s[n]$ est stationnaire au sens strict.

à ce que l'on appelle "stationnarité" dans les manuels. De façon classique , le concept de stationnarité se réfère à des processus stochastiques et est défini comme une invariance temporelle de propriétés statistiques ou, en d'autres termes, comme l'indépendance de ces propriétés par rapport à un temps absolu. En pratique cependant, la stationnarité est communément invoquée dans des contextes assez différents et/ou aménagée de considérations additionnelles. Comme exemple, on peut considérer les signaux de parole. Lorsqu'on les considère à des échelles de temps de plusieurs secondes, ceux-ci sont unanimement considérés comme "non stationnaires" et, en tant que tels, un très grand nombre de méthodes ont été proposées pour, par exemple, leur segmentation en zones "stationnaires". La façon dont ces zones sont considérées comme stationnaires diffère cependant significativement de la définition standard. D'une part, une échelle de temps est prise en compte, ce qui nécessite un aménagement par rapport à la définition stricte qui est supposée s'appliquer à tous les temps. D'autre part, une identification implicite entre stationnarité et périodicité est couramment faite (par exemple dans les parties voisées), ce qui constitue un autre écart à la définition standard donnée dans un cadre stochastique. De ce point de vue, la "stationnarité" apparaît comme n'étant pas un concept absolu, mais quelque chose qui n'a de sens que relativement à une échelle d'observation. Ainsi, suivant que la mesure en est faite sur un horizon temporel ou un autre, un même signal peut être considéré comme stationnaire ou pas, le principe étant de s'intéresser à la permanence éventuelle de propriétés descriptives, mais à l'intérieur d'un intervalle servant de cadre de référence. Ceci met naturellement en jeu deux échelles de temps : une *globale* fixée par l'horizon de référence et une *locale* à même de mettre en évidence des variations de caractéristiques à l'intérieur de la première. De façon à concilier dans un cadre unique les deux acceptions mentionnées plus haut de la "stationnarité", liées à un point de vue tant stochastique (comportement statistique de descripteurs comme la moyenne, la variance, etc.) que déterministe (périodicités), le concept de stationnarité relative peut être défini en termes temps-fréquence. En effet, étant donnée une observation, une représentation temps-fréquence offre un cadre unique pour caractériser l'évolution de propriétés spectrales aussi bien déterministes (comme une modulation de fréquence) qu'aléatoires (au sens d'un spectre dépendant du temps), la distribution calculée pouvant se voir indifféremment comme une caractérisation certaine ou comme l'estimée d'une quantité aléatoire. Dans l'un ou l'autre des cas, on conviendra d'appeler stationnaire, relativement à un horizon d'observation T, un signal dont le comportement spectral local

est semblable à sa caractérisation globale obtenue par marginalisation.

D'un point de vue pratique, pour un signal donné $x(t)$, on choisit une représentation temps-fréquence qui, dans le cas le plus simple, peut être un spectrogramme $S_x^{(h)}(t,f)$. La caractérisation de la stationnarité relative revient alors à comparer les spectres locaux $S_x^{(h)}(t_n,f)$ (obtenus pour une séquence de N instants t_n répartis sur l'intervalle d'observation T avec un espacement proportionné à la taille des fenêtres à court-terme) au spectre global défini par la marginalisation :

$$\langle S_x^{(h)}(t_n,f)\rangle_n = \frac{1}{N}\sum_{n=1}^{N} S_x^{(h)}(t_n,f).$$

Quelle que soit la mesure de dissimilarité retenue entre les spectres locaux et le spectre global, son calcul sur une observation unique ne donne jamais un résultat strictement nul dans le cas stationnaire, et la question est de savoir décider dans quelle mesure une valeur non nulle est significative ou liée aux fluctuations intrinsèques de l'estimation. Afin de poser cette question dans le cadre d'un test statistique, ceci revient à pouvoir disposer d'une référence caractérisant l'hypothèse nulle de stationnarité. Bien qu'on ne dispose par hypothèse que d'une observation, il est possible d'accéder à une telle référence en recourant à la technique des substituts décrite précédemment dans la mesure où il est immédiat de créer autant de substituts (stationnarisés) que l'on opére de randomisations sur la phase, ce qui permet la caractérisation d'une distribution d'ensemble de l'hypothèse nulle de stationnarité pour n'importe quel descripteur choisi en vue de comparer les propriétés locales et globales. Il devient alors possible, au vu de cette distribution, d'attacher un degré de signification à la valeur effective, unique, prise par le même descripteur pour l'observation. Un exemple simple illustrant l'interprétation temps-fréquence de la stationnarisation par randomisation de la phase spectrale est donné en Figure 12.

3.8.4 Randomisation sous contrainte

Si la randomisation de la phase spectrale stationnarise les données, elle tend aussi — par un effet de type théorème de la limite centrale — à les *gaussianniser*. Dans le cas général de données initialement non gaussiennes, les substituts s'éloignent ainsi des données de départ en perdant la structure spécifique de leur distribution d'amplitude marginale. Il est cependant pos-

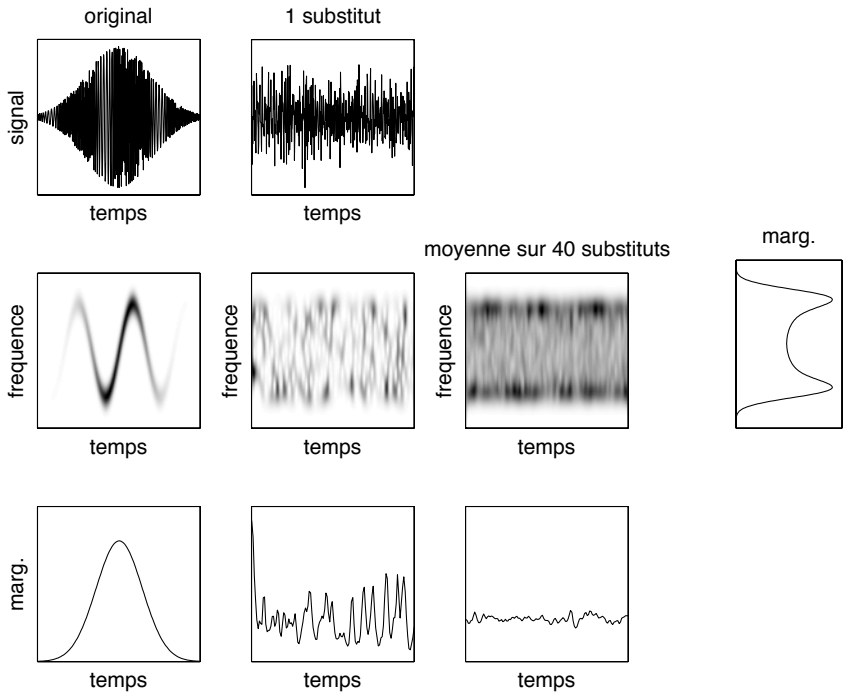

FIGURE 12 – Stationnarisation par substituts — La colonne de gauche présente un signal "non stationnaire" (en haut), son spectrogramme (au milieu) et la distribution marginale en temps de ce dernier (en bas). La deuxième colonne présente de la même façon les informations relatives à un substitut et la troisième colonne celles correspondant à une moyenne calculée sur 40 substituts du même signal. La quatrième colonne présente enfin la distribution marginale en fréquence qui, par construction, est identique pour les trois spectrogrammes. Ces différentes distributions mettent en évidence une *stationnarisation* au sens où, pour un même spectre marginal, le comportement temporel local a perdu la forte structuration du signal original.

sible de remédier à cette limitation en utilisant un algorithme amélioré qui procède de la façon itérative suivante :

— *Entrée* — On dispose de la collection $\mathbf{x} = \{x[n]; n = 0, \ldots N-1\}$ des observations et du module $A_x[k]$ de la transformée de Fourier discrète de la série associée.

— *Boucle* — On initialise $\mathbf{r}^{(1)} = \{r^{(1)}[n]; n = 0, \ldots N-1\}$, substitut obtenu "classiquement" à partir des $x[n]$. On pose $\mathbf{v} := \text{sort}(\mathbf{x})$ et, à l'itération m, on applique les 2 étapes suivantes :

1. *Projection sur le spectre* — On calcule la transformée de Fourier de $r^{(m)}[n]$, on retient la phase correspondante $\Psi^{(m)}[k]$ et on l'adjoint à l'amplitude spectrale désirée $A_x[k]$ pour calculer la transformée inverse
$$s^{(m)}[n] = \frac{1}{N} \sum_k A_x[k] e^{i\Psi^{(m)}[k]} e^{i2\pi nk/N}.$$

2. *Projection sur la distribution d'amplitude* — On ré-ordonne les valeurs de $\mathbf{s}^{(m)} = \{s^{(m)}[n]; n = 0, \ldots N-1\}$ selon l'ordre prescrit par la distribution d'amplitude désirée :
$$r^{(m+1)}[n] = v\left(\text{rank}(s^{(m)}[n])\right),$$
avec la convention que $\text{rank}(s^{(m)}[n]) = k$ si $s^{(m)}[n]$ est la k-ième plus petite valeur de $\mathbf{s}^{(m)}$.

— *Sortie* — On arrête l'itération de la boucle lorsque $\mathbf{r}^{(m)} \approx \mathbf{s}^{(m)}$, ou lorsque $\mathbf{r}^{(m+1)} \approx \mathbf{r}^{(m)}$ et/ou $\mathbf{s}^{(m+1)} \approx \mathbf{s}^{(m)}$.

4 Structures propres et adaptativité

4.1 Karhunen-Loève

4.1.1 Principe

Soit $b(t)$ une fonction aléatoire

1. définie sur un intervalle fermé $[0, T]$;
2. continue en moyenne quadratique, c'est-à-dire telle que

$$\lim_{\Delta t \to 0} \mathbb{E}\left\{[b(t+\Delta t) - b(t)]^2\right\} = 0,$$

soit encore, dans le cas stationnaire pour lequel

$$r_b(t,s) := \mathbb{E}\{b(t)\, b^*(s)\} =: \gamma_b(t-s),$$

telle que la fonction d'autocorrélation $\gamma_b(\tau)$ soit continue à l'origine :

$$\lim_{\Delta t \to 0} [\gamma_b(0) - \gamma_b(\Delta t)] = 0.$$

La *décomposition de Karhunen-Loève* se propose alors de représenter la fonction aléatoire $b(t)$ par une *combinaison linéaire de fonctions certaines dont les poids sont aléatoires*. Si l'on écrit explicitement la fonction aléatoire $b(t, \omega)$ pour signifier sa dépendance vis-à-vis des épreuves ω (ce que l'on ne fera pas dans la suite pour alléger les notations), la décomposition recherchée revient à "séparer les variables" selon :

$$b(t, \omega) = \sum_i b_i(\omega)\, \varphi_i(t).$$

La spécificité supplémentaire de la décomposition de Karhunen-Loève est d'être associée à une propriété de *double orthogonalité* garantissant :

1. la *décorrélation* entre les coefficients de la décomposition :

$$\mathbb{E}\left\{b_i\, b_j^*\right\} = \mathbb{E}\left\{|b_i|^2\right\}.\delta_{ij};$$

2. l'*orthogonalité* au sens des fonctions de $L^2([0,T])$ des fonctions de base de la décomposition :

$$\int_0^T \varphi_i(t)\, \varphi_j^*(t)\, dt = \delta_{ij}.$$

Ceci permet d'obtenir les poids b_i comme projections de $b(t)$ sur les fonctions $\varphi_i(t)$. En effet, on a :

$$\begin{aligned}\int_0^T b(t)\,\varphi_i^*(t)\,dt &= \int_0^T \left[\sum_j b_j\,\varphi_j(t)\right]\,\varphi_i^*(t)\,dt \\ &= \sum_j b_j\left[\int_0^T \varphi_j(t)\,\varphi_i^*(t)\,dt\right] \\ &= \sum_j b_j\,\delta_{ji} \\ &= b_i.\end{aligned}$$

De façon à trouver les fonctions $\varphi_i(t)$ permettant la double orthogonalité, on part de la décomposition

$$b(t) = \sum_j b_j\,\varphi_j(t)$$

et, par multiplication par b_i^*, on forme

$$\begin{aligned}b(t)\,b_i^* &= \left[\sum_j b_j\,\varphi_j(t)\right]\,b_i^* \\ &= \sum_j b_j\,b_i^*\,\varphi_i(t).\end{aligned}$$

En en prenant l'espérance mathématique, il vient

$$\begin{aligned}\mathbb{E}\left\{b(t)\,b_i^*\right\} &= \mathbb{E}\left\{\sum_j b_j\,b_i^*\,\varphi_j(t)\right\} \\ &= \sum_j \mathbb{E}\left\{b_j\,b_i^*\right\}\,\varphi_j(t) \\ &= \sum_j \mathbb{E}\left\{|b_j|^2\right\}\,\delta_{ji}\,\varphi_j(t) \\ &= \mathbb{E}\left\{|b_i|^2\right\}\,\varphi_i(t)\end{aligned}$$

si les b_i sont mutuellement décorrélés. Or, on a par ailleurs :

$$\begin{aligned}\mathbb{E}\left\{b(t)\, b_i^*\right\} &= \mathbb{E}\left\{b(t)\, \left[\int_0^T b(u)\, \varphi_i^*(u)\, du\right]^*\right\} \\ &= \int_0^T \mathbb{E}\left\{b(t)\, b^*(u)\right\}\, \varphi_i(u)\, du \\ &= \int_0^T r_b(t,u)\, \varphi_i(u)\, du.\end{aligned}$$

Par suite, les fonctions $\varphi_i(t)$ permettant la décomposition doublement orthogonale de Karhunen-Loève sont celles pour lesquelles

$$\int_0^T r_b(t,u)\, \varphi_i(u)\, du = \lambda_i\, \varphi_i(t); t \in [0,T],$$

avec

$$\lambda_i = \mathbb{E}\left\{|b_i|^2\right\}.$$

Ce sont donc les *fonctions propres* de la covariance, la puissance des coefficients associés se mesurant quant à elle par la valeur propre correspondante.

4.1.2 Quelques propriétés

Propriété 1 (résolution de l'identité). Comme

$$b_i = \int_0^T b(t)\, \varphi_i^*(t)\, dt,$$

on en déduit que

$$b_i\, \varphi_i(u) = \int_0^T b(t)\, \varphi_i^*(t)\, \varphi_i(u)\, du$$

et, par suite,

$$\begin{aligned}\sum_i b_i\, \varphi_i(u) &= b(u) \\ &= \sum_i \int_0^T b(t)\, \varphi_i^*(t)\, \varphi_i(u)\, du \\ &= \int_0^T b(t)\, \left[\sum_i \varphi_i^*(t)\, \varphi_i(u)\right] dt.\end{aligned}$$

On a donc :

$$\sum_i \varphi_i^*(t)\,\varphi_i(u) = \delta(t-u).$$

Propriété 2 (Théorème de Mercer). Sachant que

$$\lambda_i\,\varphi_i(t) = \int_0^T r_b(t,v)\,\varphi_i(v)\,dv,$$

il vient :

$$\begin{aligned}
\sum_i \lambda_i\,\varphi_i(t)\,\varphi_i^*(u) &= \sum_i \left[\int_0^T r_b(t,v)\,\varphi_i(v)\,dv\right]\,\varphi_i^*(u) \\
&= \int_0^T r_b(t,v)\,\left[\sum_i \varphi_i^*(u)\,\varphi_i(v)\right]\,dv \\
&= \int_0^T r_b(t,v)\,\delta(u-v)\,dv
\end{aligned}$$

d'après la Propriété 1. On en déduit (*Théorème de Mercer*) que :

$$r_b(t,u) = \sum_i \lambda_i\,\varphi_i(t)\,\varphi_i^*(u).$$

Propriété 3 (trace invariante). Partant du théorème de Mercer, on obtient directement

$$r_b(t,t) = \sum_i \lambda_i\,|\varphi_i(t)|^2$$

et donc

$$\begin{aligned}
\int_0^T r_b(t,t)\,dt &= \sum_i \lambda_i\,\int_0^T |\varphi_i(t)|^2\,dt \\
&= \sum_i \lambda_i,
\end{aligned}$$

soit la propriété de *trace invariante* :

$$\sum_i \lambda_i = \int_0^T r_b(t,t)\,dt.$$

On montre également que :

$$\sum_i \lambda_i^2 = \iint_0^T r_b(t,u)\,dt\,du.$$

Propriété 4 (bruit blanc stationnaire). Dans le cas particulier où $b(t)$ est *stationnaire*, l'équation de Karhunen-Loève devient :

$$\int_0^T \gamma_b(t-u)\,\varphi_i(u)\,du = \lambda_i\,\varphi_i(t).$$

Par suite, dans le cas d'un *bruit blanc* de densité spectrale de puissance γ_0, on obtient :

$$\int_0^T \gamma_0\,\delta(t-u)\,\varphi_i(u)\,du = \lambda_i\,\varphi_i(t),$$

soit encore :

$$\gamma_0\,\varphi_i(t) = \lambda_i\,\varphi_i(t).$$

Ce résultat signifie que, dans le cas d'un bruit blanc stationnaire :

1. *toute* base est de Karhunen-Loève ;
2. toutes les valeurs propres sont *égales* et ont pour valeur la densité spectrale de puissance du bruit :

$$\lambda_i = \mathbb{E}\left\{|b_i|^2\right\} = \gamma_0.$$

En conséquence, dans le cas d'un bruit *gaussien*, centré, blanc et stationnaire, la *variance* de ses coefficients de Karhunen-Loève s'identifie à la *densité spectrale de puissance* : la connaissance de cette dernière est donc suffisante pour expliciter la densité de probabilité associée.

Propriété 5 (covariance inverse). En considérant la covariance $r_b(t,u)$ comme le noyau d'un opérateur intégral, on peut envisager de lui associer le noyau $\sigma_b(t,u)$ de l'*opérateur inverse*, défini par :

$$\int_0^T \sigma_b(t,u)\,r_b(u,v)\,du = \delta(t-v).$$

On vérifie alors que :

$$\begin{aligned}\int_0^T \sigma_b(t,u)\,\varphi_i(u)\,du &= \int_0^T \sigma_b(t,u)\,\left[\frac{1}{\lambda_i}\int_0^T r_b(u,v)\,\varphi_i(v)\,dv\right]\,du\\ &= \frac{1}{\lambda_i}\int_0^T\left[\int_0^T \sigma_b(t,u)\,r_b(u,v)\,du\right]\,\varphi_i(v)\,dv\\ &= \frac{1}{\lambda_i}\int_0^T \delta(t-v)\,\varphi_i(v)\,dv\\ &= \frac{1}{\lambda_i}\,\varphi_i(v).\end{aligned}$$

Ceci signifie que l'opérateur inverse de la covariance vérifie :

$$\int_0^T \sigma_b(t,u)\,\varphi_i(u)\,du = \frac{1}{\lambda_i}\,\varphi_i(t),$$

c'est-à-dire que ses fonctions propres sont les mêmes que celles de l'opérateur de covariance, ses valeurs propres étant par contre inverses ce celles de l'opérateur direct. On en déduit en particulier (par application du Théorème de Mercer) que l'opérateur inverse de la covariance admet la décomposition :

$$\sigma_b(t,u) = \sum_i \frac{1}{\lambda_i}\,\varphi_i(t)\,\varphi_i^*(u).$$

4.1.3 Exemples

Exemple 1 (bruit à bande limitée). Supposons que le bruit $b(t)$ soit *à bande limitée* $[-B/2,+B/2]$ et *blanc* dans cette bande, c'est-à-dire tel que sa densité spectrale de puissance s'écrive :

$$\Gamma_b(f) = \gamma_0\,\mathbf{1}_{[-B/2,+B/2]}(f).$$

Par transformation inverse de Fourier, on en déduit que sa fonction d'autocorrélation (stationnaire) $\gamma_b(\tau)$ a pour valeur :

$$\gamma_b(\tau) = \gamma_0\,\frac{\sin\pi B\tau}{\pi\tau}$$

et qu'il admet une décomposition sur une base de Karhunen-Loève dont les fonctions sont solutions de l'équation propre ? :

$$\int_0^T \frac{\sin\pi B(t-u)}{\pi(t-u)}\,\varphi_i(u)\,du = \lambda_i\,\varphi_i(t)\,;\ t\in[0,T].$$

De telles fonctions sont connues sous le nom de *fonctions sphéroïdales aplaties* ("prolate spheroidal wave functions" en anglais). Celle associée à la plus grande valeur propre permet en particulier de maximiser l'énergie d'un signal à bande limitée B donnée sur un support temporel T fixé.

Exemple 2 (processus de Wiener). Un *processus de Wiener* $w(t)$ peut se voir comme la limite à temps continu d'une marche aléatoire. C'est un processus *non stationnaire* dont on peut montrer qu'il a pour fonction de covariance :

$$r_w(t,s) = \alpha \, \min(t,s) \, ; \, (t,s) \in [0,T]^2.$$

Reportant cette covariance dans l'équation de Karhunen-Loève, on obtient :

$$\alpha \int_0^t u \, \varphi_i(u) \, du + \alpha t \int_t^T \varphi_i(u) \, du = \lambda_i \, \varphi_i(t) \, ; \, t \in [0,T].$$

On vérifie alors que $\varphi_i(0) = 0$ et, en dérivant par rapport à t, on obtient :

$$\alpha \int_t^T \varphi_i(u) \, du = \lambda_i \, \dot{\varphi}_i(t).$$

On en déduit que $\dot{\varphi}_i(T) = 0$ et, en dérivant une seconde fois par rapport à t, on aboutit à l'équation différentielle :

$$\ddot{\varphi}_i(t) + \frac{\alpha}{\lambda_i} \, \varphi_i(t) = 0,$$

dont les solutions sont des fonctions circulaires de pulsations :

$$\omega_i = \sqrt{\frac{\alpha}{\lambda_i}}.$$

Des conditions aux bords trouvées précédemment, on déduit d'une part que

$$\varphi_i(0) = 0 \Rightarrow w(t) \propto \sin \omega_i t$$

et d'autre part que

$$\dot{\varphi}_i(T) = 0 \Rightarrow \omega_i T = (2i+1)\frac{\pi}{2}.$$

Imposant en outre aux fonctions solutions d'être de norme unité, on obtient finalement la décomposition de Karhunen-Loève du processus de Wiener sur $[0, T]$ qui s'écrit :

$$w(t) = \sqrt{\frac{2}{T}} \sum_i w_i \sin \frac{(2i+1)\pi t}{2T},$$

expression dans laquelle les coefficients w_i sont *décorrélés* et de variance λ_i, soit encore :

$$\mathrm{var}\{w_i\} = \frac{4\alpha\pi^2/T^2}{(2i+1)^2}.$$

4.2 "Singular Spectrum Analysis"

4.2.1 Principe

4.2.2 Algorithme

4.2.3 Quelques propriétés

4.2.4 Exemples

4.3 "Empirical Mode Decomposition"

4.3.1 Principe

La "Décomposition Modale Empirique" (ou *EMD*, pour "Empirical Mode Decomposition") est une méthode de décomposition pilotée par les données qui a été introduite par N.E. Huang à la fin des années 90, pour l'analyse de signaux non stationnaires et/ou issus de systèmes non linéaires. Dans son principe, l'EMD considère les signaux à l'échelle de leurs *oscillations locales*, sans que celles-ci soient nécessairement harmoniques au sens de Fourier. D'une manière plus précise, si l'on cherche à décrire un signal $x(t)$ entre deux extrema consécutifs (par exemple, deux minima situés aux temps t_- et t_+), on peut définir de façon heuristique une contribution "hautes fréquences" locale $\{d_1(t), t_- \leq t \leq t_+\}$, ou *détail local*, qui correspond à l'oscillation se terminant aux deux minima considérés et passant par le maximum qui existe nécessairement entre eux. Pour que la description du comportement local soit complète, il suffit d'identifier la contribution "basses fréquences" locale correspondante $a_1(t)$, ou *tendance locale*, de telle sorte que l'on ait

$$x(t) = a_1(t) + d_1(t)$$

pour $t_- \leq t \leq t_+$. Si ce point de vue est adopté pour l'ensemble des oscillations constituant le signal, la procédure peut alors être appliquée sur le résidu $a_1(t)$ formé par l'ensemble des tendances locales et considéré comme un nouveau signal, conduisant à un nouveau détail $d_2(t)$ et à un nouveau résidu $a_2[(t)$.

On obtient ainsi une décomposition dont les différents *modes* (ou IMF, pour "Intrinsic Mode Functions") $d_k(t)$ sont extraits itérativement, conduisant à une représentation du type :

$$x(t) = a_K(t) + \sum_{k=1}^{K} d_k(t)$$

pour une profondeur de décomposition K. Une telle expression rappelle dans sa forme une décomposition en ondelettes, mais la différence essentielle est que cette dernière repose de façon explicite sur l'usage de filtres (passe-haut et passe-bas) définis *a priori*, alors que l'EMD s'adapte localement au contenu fréquentiel du signal, l'oscillation "hautes fréquences" n'étant pas définie de façon fixe et prescrite, mais seulement relativement à une contribution plus lente acquérant de fait un statut "basses fréquences".

4.3.2 Algorithme

De façon plus précise, la décomposition est effectuée de manière itérative et hiérarchique, partant de l'identification de l'oscillation la plus rapide pour aller vers la plus lente. À chaque étape de l'algorithme, la tendance $a_k(t)$ et l'IMF $d_k(t)$ d'ordre $k > 1$ sont extraits de la tendance $a_{k-1}(t)$ d'ordre $k-1$ (avec l'initialisation $a_0(t) = x(t)$), la caractérisation d'un IMF reposant sur le double critère que :

1. le nombre des extrema et des passages à zéro diffère d'au plus 1 ;
2. la moyenne locale soit nulle.

La première contrainte garantit le caractère oscillant d'un IMF. La seconde est plus sévère et nécessite un traitement spécifique pour être satisfaite. La solution initiale proposée est de construire (par interpolation) deux "enveloppes" s'appuyant d'une part sur les maxima et d'autre part sur les minima, la valeur moyenne locale nulle de l'IMF se traduisant par la symétrie des enveloppes. Le calcul direct fournit un "proto-mode" qui ne possède en général pas cette symétrie : l'algorithme procède alors à une opération dite

de *tamisage* dans laquelle la même procédure est appliqué au proto-mode jusqu'à obtenir pour la moyenne une valeur jugée négligeable par l'utilisateur :

1. initialisation d'une fonction temporaire $s(t) = a_{k-1}(t)$,
2. identification de tous les extrema de $s(t)$,
3. interpolation entre les minima (resp. maxima) pour construire une enveloppe inférieure $e_{\min}(t)$ (resp. minimum $e_{\max}(t)$),
4. calcul de l'enveloppe moyenne $m(t) = \frac{e_{\min}(t)+e_{\max}(t)}{2}$,
5. extraction du résidu $s(t) = s(t) - m(t)$,
6. itération des étapes 2 à 5 jusqu'à ce que le résidu $s(t)$ ait une enveloppe moyenne "nulle",
7. obtention de l'IMF $d_k(t) = s(t)$ et de la tendance $a_k(t) = a_{k-1}(t) - s(t)$.

La structure complète est obtenue en itérant sur l'ordre k, en partant de l'initialisation $a_0(t) = x(t)$ et en arrêtant lorsque le dernier résidu est à variation monotone.

4.3.3 Quelques propriétés

L'EMD possède un certain nombre de propriétés qui en font un outil se différenciant d'autres méthodes de décomposition :

— *Adaptativité* — C'est tout d'abord une méthode véritablement adaptative, le schéma de décomposition ne s'appuyant que sur les données, sans référence à un modèle ou une structure de filtrage prescrite a priori. En ce sens, elle se distingue d'une décomposition en ondelettes — avec laquelle elle partage le principe d'une séparation en une composante rapide et une composante lente, suivie d'une itération sur la composante lente —, dans la mesure où la distinction lente vs. rapide ne résulte pas de filtrages "demi-bande" fixes.

— *Localité* — L'EMD est ensuite une méthode locale, opérant à l'échelle d'une oscillation. Ceci renforce le caractère adaptatif de la méthode, le caractère "rapide" ou "lent" d'une oscillation étant relatif (et non pas absolu comme dans le cas d'une décomposition en ondelettes). Il s'en suit une capacité a priori renforcée de suivre les non-stationnarités.

— *Oscillation* — Parce qu'elle ne prend en compte que les extrema, l'EMD est une méthode faisant davantage ressortir des *oscillations* que des *fréquences*, celles-ci étant liés de manière plus stricte à des formes harmoniques à base de sinus et cosinus. Ainsi, une oscillation

non harmonique (issue par exemple d'un système non linéaire) peut être vue comme un mode unique, là où une analyse fréquentielle (Fourier, ondelettes ou variations) mettra en jeu un nombre plus important de composantes.

— *Multirésolution* — Dans le cas de signaux à structure très désordonnée (typiquement, des bruits à large bande), on peut montrer que l'EMD fournit spontanément une décomposition dont les modes sont semblables à ceux qui résulteraient du filtrage par un banc de filtres quasi-dyadiques.

4.3.4 Exemples

Un exemple simple de décomposition par EMD est donné à la figure 13, avec les distributions temps-fréquence associées des différents modes à la figure 14 (spectrogrammes réalloués). On y voit clairement la possibilité offerte par la méthode de séparer deux composantes AM-FM alors même que leurs bandes spectrales *globales* se chevauchent, la séparation reposant sur l'identification *locale* de l'oscillation la plus rapide.

Sur la base de cette interprétation, il est facile de construire des "contre-exemples" dans lesquels l'EMD se comporte d'une manière qui n'est qu'indirectement représentative des composantes réelles d'un signal. Il en est ainsi à la figure 15, relative à l'analyse du signal de la figure 14 auquel a été ajouté en son milieu un transitoire à fréquence constante. Il est clair dans ce cas que le premier IMF extrait est amené à "sauter" du mode AM-FM de plus haute fréquence locale pour rejoindre le transitoire lorsque celui-ci est présent, dans la mesure où c'est celui-ci qui devient alors l'oscillation la plus rapide. La méthode se comporte bien comme attendu et, plutôt qu'un contre-exemple, il faut davantage y voir une indication sur les hypothèses à faire pour son utilisation et pour une interprétation directe des modes extraits.

Ainsi, d'une manière générale, l'EMD est bien adaptée aux situations où le signal analysé est de la forme AM-FM

$$x(t) = \sum_{k=1}^{K} \alpha_k(t) \cos \varphi_k(t),$$

avec les hypothèses additionnelles que

1. les K composantes *co-existent* de façon permanente sur toute la durée de l'observation, soit $\alpha_k(t) > 0$ pour tout k ;

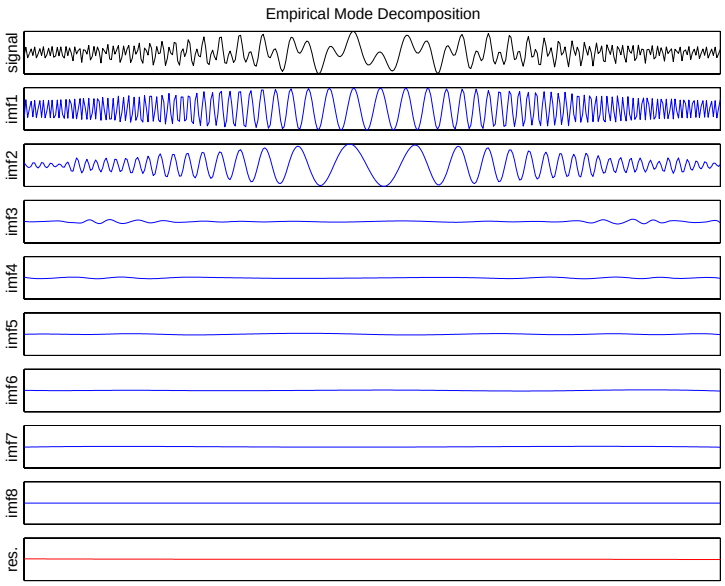

FIGURE 13 – EMD d'un signal AM-FM à deux composantes — Le signal composite (en noir) est décomposé en une collection de modes (en bleu) dont seuls les deux premiers sont significativement non nuls, auxquels s'ajoute un résidu (en rouge).

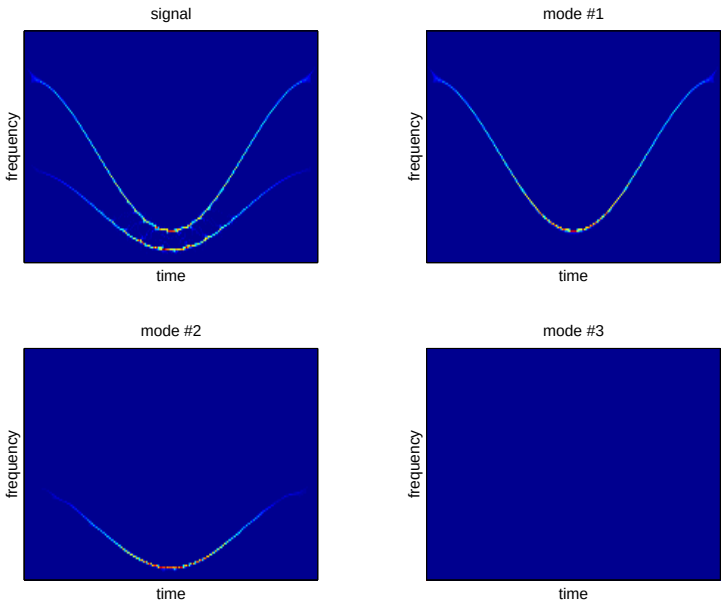

FIGURE 14 – EMD du signal de la figure 13 — Spectrogrammes réalloués du signal composite (en haut à gauche) et des trois premiers modes extraits.

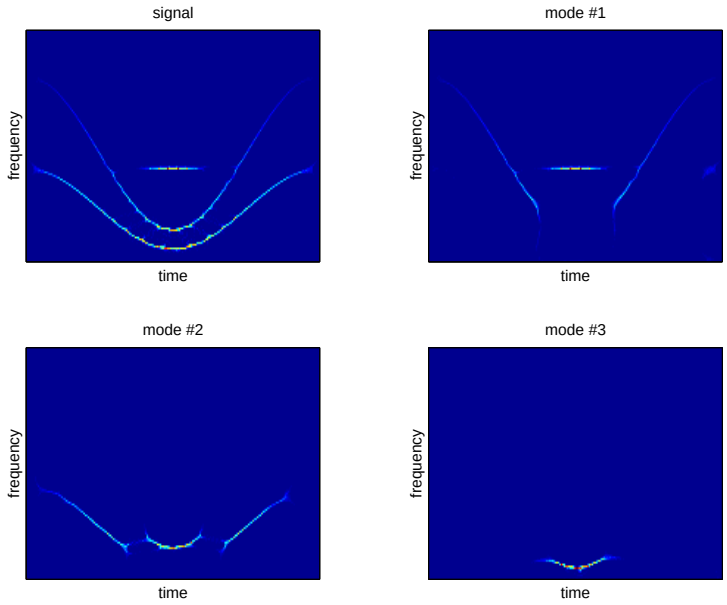

FIGURE 15 – EMD du signal de la figure 13 auquel a été rajouté en son milieu un transitoire — Spectrogrammes réalloués du signal composite (en haut à gauche) et des trois premiers modes extraits.

2. les trajectoires des fréquences instantanées $f_k(t) = \dot{\varphi}_k(t)/2\pi$ *ne se croisent pas* dans le plan, soit $f_1(t) > f_2(t) > \ldots > f_K(t)$ à tout instant t.

4.3.5 Variations

Le principe général de l'EMD étant acquis, de nombreuses variations peuvent en être proposées, parmi lesquelles on peut retenir les trois suivantes.

1. *"Ensemble EMD (EEMD)"* — Le point de départ est la reconnaissance expérimentale du fait que l'EMD est sujet à un phénomène de "mode mixing" selon lequel une composante donnée peut se répartir sur plusieurs modes extraits, de façon intermittente. Il en ainsi parce que, dans le cas par exemple d'un signal bruité, la prise en compte de la fluctuation locale d'une réalisation particulière du bruit dépendra du comportement local du signal (cf. figure 16). Une façon de remédier à cet effet est de rajouter à l'observation un niveau suffisant de bruit

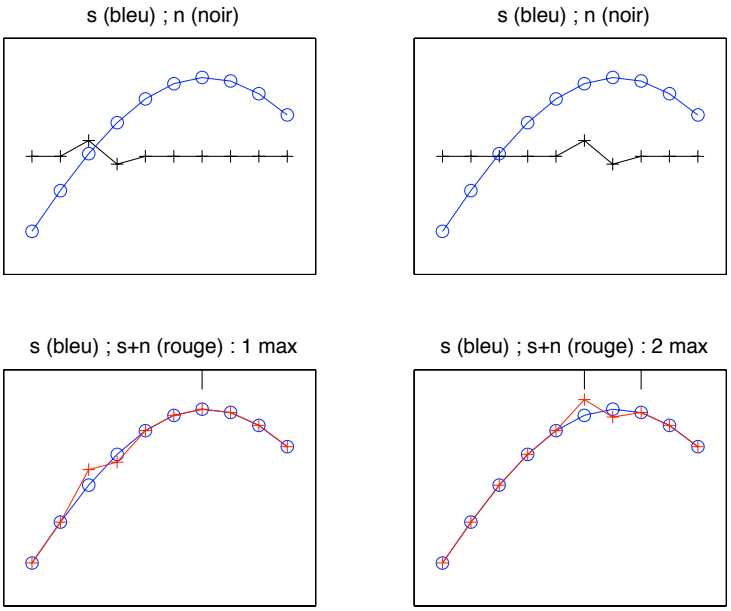

FIGURE 16 – "Mode mixing" — Suivant sa position relativement au signal s, une même perturbation n pourra créer dans le signal somme $s + n$ un nombre différent de maxima locaux (1 dans la colonne de gauche et 2 dans celle de droite). Au niveau de l'EMD correspondante, il en résultera une prise en compte de la perturbation pouvant être, soit intégrée à l'IMF "signal" (gauche), soit distincte de celui-ci (droite)

extérieur afin de contraindre les fluctuations superposées aux modes supposés lisses de se détacher de ceux-ci. En moyennant, mode par mode, un nombre suffisant de décompositions relatives à des réalisations de bruit extérieur indépendantes, il est possible d'obtenir une réduction du "mode mixing".

2. *EMD multivariée* — Une extension de l'EMD classique aux signaux bivariés (et, plus généralement, multivariés) est possible en remplaçant l'idée d'*oscillation* par celle de *rotation*. L'analogue des enveloppes inférieure et supérieure entre lesquelles le signal oscille devient alors un *tube* dont le centre local joue le rôle de moyenne. L'intérêt de cette approche est de garantir une analyse cohérente (en termes d'appartenance à un mode donné) des deux composantes d'un signal bivarié.
3. *EMD à base d'optimisation* —

Références

[1] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, 1999.

[2] N. Gershenfeld, *The Nature of Mathematical Modeling*, Cambridge Univ. Press, 1999.

[3] S. Mallat, *A Wavelet Tour of Signal Processing — The Sparse Way* (3rd ed.), Academic Press, 2009.

[4] `http://www.ens-lyon.fr/PHYSIQUE/Equipe3/ANR_ASTRES/resources.html`