Traitement du signal

Stéphane Mallat

Table des Matières

Chapitre 1

Introduction

Initialement appliqué aux télécommunications, le traitement du signal se retrouve à présent dans tous les domaines nécessitant d'analyser et transformer de l'information numérique. La manipulation de données obtenues par capteurs bio-médicaux, lors d'expériences physiques ou biologiques, sont aussi des problèmes de traitement du signal. Le téléphone, la radio et la télévision ont motivé l'élaboration d'algorithmes de filtrage linéaires permettant de coder des sons ou des images, de les transmettre, et de supprimer certains bruits de transmission. Le chapitre 2 introduit le filtrage analogique avec une application à la transmission par modulation d'amplitude.

Lorsque la rapidité de calcul le permet, le filtrage par circuits d'électronique analogique est remplacé par des calculs numériques sur des signaux digitaux. Le calcul digital est en effet plus fiable et offre une flexibilité algorithmique bien plus importante. C'est pourquoi les disques et cassettes analogiques ont récemment été remplacés par les disques compacts numériques et les cassettes digitales. La conversion analogique-digitale est étudiée dans le chapitre 3 ainsi que l'extension des opérateurs de filtrage aux signaux discrets. L'introduction de la transformée de Fourier discrète rapide par Cooley et Tuckey en 1965 a fait de l'analyse de Fourier un outil algorithmique puissant qui se retrouve dans la plupart des calculs rapides de traitement du signal.

Lorsque l'on veut décrire les propriétés d'une classe de signaux, comme un même son prononcé par différents locuteurs, il est utile de se placer dans un cadre probabiliste. La variété des signaux d'une telle classe peut en effet être caractérisée par un processus aléatoire. La modélisation de signaux par procéssus stationnaires est introduite dans le chapitre 4, où nous étudions une application pour la suppression de bruits additifs.

Les problèmes les plus difficiles du traitement du signal sont souvent liés au traitement de l'information. Comment extraire l'information utile d'un signal ? La reconnaissance de la parole a motivé un grand nombre de travaux dans ce domaine. La performance des systèmes de reconnaissance de parole a progressé beaucoup plus lentement que les projections optimistes des années 50. Les algorithmes de traitement doivent s'adapter au contenu complexe du signal, et sont donc beaucoup plus sophistiqués que des filtrages linéaires homogènes. Le chapitre 5 étudie l'application des modèles autorégressifs. Extraire l'information d'un signal nécessite souvent d'analyser les évolutions temporelles de ses "composantes fréquentielles". Le chapitre 6 introduit des décompositions temps-fréquence qui représentent un signal en structures élémentaires ressemblant à des notes

de musiques, afin de plus facilement caractériser son contenu.

La notion d'information contenue dans un signal peut se formaliser par la théorie de Shannon qui la relie au nombre de bits minimum pour coder le signal. L'entropie introduite dans le chapitre 7 mesure la quantité d'information d'un signal. Le chapitre 8 montre que ces conceptes permettent d'élaborer des algorithmes de compression qui suppriment la redondance interne d'un signal afin de le représenter avec un nombre de bits réduit. De tels algorithmes augmentent considérablement les capacités de stockage, et permettent de transmettre des signaux à travers des canaux à débits réduits. Par exemple, la transmission d'images sur Internet utilise l'algorithme de compression JPEG qui est décrit dans ce dernier chapitre.

Chapitre 2

Traitement du signal analogique

Le traitement du signal analogique repose essentiellement sur l'utilisation d'opérateurs linéaires qui modifient les propriétés d'un signal de façon homogène dans le temps. La transformées de Fourier diagonalise ces opérateurs et apparaît donc comme le principal outil d'analyse mathématique. Nous étudions la synthèse de filtres homogènes et une application à la transmission par modulation d'amplitude.

2.1 Filtrage linéaire homogène

Un signal analogique mono-dimensionnel est une fonction $f(t)$ d'une variable continue $t \in \mathbb{R}$, que nous supposerons être le temps. De nombreux algorithmes de traitement du signal tels que la transmission par modulation d'amplitude, le débruitage de signaux stationnaires ou le codage par prédiction, s'implémentent avec des opérateurs linéaires homogènes dans le temps.

L'homogénéité temporelle d'un opérateur L signifie que si l'entrée $f_\tau(t) = f(t-\tau)$ est retardée par $\tau \in \mathbb{R}$ alors la sortie $L[f_\tau(t)]$ est aussi retardée par τ

$$L[f(t)] = g(t) \Rightarrow L[f_\tau(t)] = g(t-\tau). \tag{2.1}$$

Pour garantir une stabilité numérique, nous supposons aussi que L a une continuité faible. Pour tout t, la sortie $L[f](t)$ est peu perturbée si $f(t)$ est un signal régulier qui est légèrement modifié. Cette continuité peut être formalisée dans le cadre de la théorie des distributions [4].

2.1.1 Dirac

Un Dirac est une masse ponctuelle qui est souvent utilisée pour simplifier les calculs. C'est une "distribution" $\delta(t)$ dont le support est réduit au point $t = 0$ et d'intégrale unité, si bien que pour toute fonction continue $f(t)$

$$\int_{-\infty}^{+\infty} f(u)\delta(u)du = f(0).$$

La théorie des distributions [4] définie formellement cette intégrale comme une forme linéaire qui associe a toute fonction sa valeur en $t = 0$. Nous nous contentons ici

d'une définition plus intuitive d'un Dirac comme limite de "bosses" qui sont contractées indéfiniment.

Soit $\phi(t)$ une fonction continue à support dans $[-1,1]$ et de masse unité

$$\int_{-\infty}^{+\infty} \phi(u)du = 1. \tag{2.2}$$

La fonction $\phi_s(t) = \frac{1}{s}\phi(\frac{t}{s})$ a un support dans $[-s,s]$. Avec le changement de variable $t' = \frac{t}{s}$ on montre que

$$\int_{-\infty}^{+\infty} \phi_s(u)du = \int_{-\infty}^{+\infty} \frac{1}{s}\phi(\frac{u}{s})du = 1.$$

Soit $f(t)$ une fonction continue, on vérifie facilement que

$$\lim_{s\to 0} \int_{-\infty}^{+\infty} f(u)\phi_s(u)du = f(0).$$

Un Dirac se définit par

$$\delta(t) = \lim_{s\to 0} \phi_s(t),$$

où la limite doit être formellement prise au sens où pour toute fonction continue $f(t)$

$$\lim_{s\to 0} \int_{-\infty}^{+\infty} f(u)\phi_s(u)du = \int_{-\infty}^{+\infty} f(u)\delta(u)du = f(0). \tag{2.3}$$

Un Dirac $\delta(t)$ n'est pas une fonction mais la théorie des distributions montre que cela se manipule comme une fonction dans les calculs. On oubliera donc son statut de distribution par la suite. Ainsi, on peut définir un Dirac translaté en τ par

$$\delta(t-\tau) = \lim_{s\to 0} \phi_s(t-\tau),$$

et on montre que pour toute fonction continue $f(t)$

$$\int_{-\infty}^{+\infty} f(u)\delta(u-\tau)du = f(\tau). \tag{2.4}$$

On peut cependant éviter d'utiliser une limite et déduire simplement ce resultat de (2.2) par un changement de variable $u' = u + \tau$.

Un Dirac est symétrique $\delta(t) = \delta(-t)$ car

$$\int_{-\infty}^{+\infty} f(u)\delta(-u)du = \int_{-\infty}^{+\infty} f(-u)\delta(u)du = f(0).$$

On peut donc réécrire (2.4) comme une décomposition de $f(t)$ en une somme de Diracs translatés en différents points

$$f(t) = \int_{-\infty}^{+\infty} f(u)\delta(t-u)du.$$

2.1.2 Réponse impulsionnelle

En décomposant un signal comme une somme de Diracs translatés, nous montrons que tout opérateur homogène peut s'écrire comme un produit de convolution. On a vu que

$$f(t) = \int_{-\infty}^{+\infty} f(u)\delta(t-u)du.$$

La continuité et la linéarité de L montrent que

$$L[f(t)] = \int_{-\infty}^{+\infty} f(u)L[\delta(t-u)]du.$$

Soit $h(t)$ la réponse de L pour une impulsion $\delta(t)$

$$h(t) = L[\delta(t)].$$

L'homogénéité temporelle implique $L[\delta(t-u)] = h(t-u)$ et donc

$$L[f(t)] = \int_{-\infty}^{+\infty} f(u)h(t-u)du = \int_{-\infty}^{+\infty} h(u)f(t-u)du.$$

On note le produit de convolution de h avec f

$$L[f(t)] = h \star f(t) = \int_{-\infty}^{+\infty} h(u)f(t-u)du.$$

Un opérateur linéaire homogène se calcule donc par un produit de convolution avec la réponse impulsionnelle.

On rappelle quelques propriétés importantes du produit de convolution :

- Commutativité
$$f \star h(t) = h \star f(t). \tag{2.5}$$
- La convolution de $f(t)$ avec un Dirac translaté $\delta_\tau(t) = \delta(t-\tau)$ translate $f(t)$ par τ
$$f \star \delta_\tau(t) = \int_{-\infty}^{+\infty} f(t-u)\delta_\tau(u)du = f(t-\tau). \tag{2.6}$$

Stabilité et causalité Un filtre est *causal* si et seulement si $L[f(t)]$ ne dépend que des valeurs $f(u)$ pour $u < t$. Comme

$$L[f(t)] = \int_{-\infty}^{+\infty} h(u)f(t-u)du,$$

cela signifie que $h(u) = 0$ pour $u < 0$. On dit alors que $h(t)$ est une fonction *causale*. On exprime souvent les fonctions causale comme un produit avec une fonction échelon

$$h(t) = h(t)\gamma(t)$$

avec

$$\gamma(t) = \begin{cases} 0 & \text{si } t < 0 \\ 1 & \text{si } t \geq 0 \end{cases} \tag{2.7}$$

Lorsque $f(t)$ est bornée on veut garantir que $L[f(t)]$ est aussi bornée. On dit alors que le filtre L et $h(t)$ sont *stables*. Comme

$$|L[f(t)]| \leq \sup_{u \in \mathbb{R}} |f(u)| \int_{-\infty}^{+\infty} |h(u)| du, \tag{2.8}$$

il suffit que $h \in \mathbf{L^1}(\mathbb{R})$

$$\int_{-\infty}^{+\infty} |h(u)| du < +\infty.$$

On vérifie (exercice) que si $h(t)$ est une fonction definie pour presque tout $t \in \mathbb{R}$ la condition $h \in \mathbf{L^1}(\mathbb{R})$ est aussi nécessaire.

Exemples

- Un système *d'amplification* par λ et de *délai* par τ est défini par

$$L[f(t)] = \lambda f(t - \tau).$$

La réponse impulsionnelle de ce filtre est

$$h(t) = \lambda \delta(t - \tau).$$

- La *moyenne uniforme* de $f(t)$ dans un voisinage de taille T est

$$L[f(t)] = \frac{1}{T} \int_{t-\frac{T}{2}}^{t+\frac{T}{2}} f(u) du.$$

Cette intégrale peut être réécrite comme un produit de convolution avec

$$h(t) = \begin{cases} \frac{1}{T} & \text{si } t \in [-\frac{T}{2}, \frac{T}{2}] \\ 0 & \text{si } |t| > \frac{T}{2} \end{cases}$$

- Une *moyenne pondérée* correspond à une réponse impulsionnelle $h(t)$ telle que

$$\int_{-\infty}^{+\infty} h(u)\, du = 1.$$

L'intégrale

$$L[f(t)] = \int_{-\infty}^{+\infty} h(u) f(t - u) du$$

peut être interprétée comme une moyenne pondérée par $h(u)$ de $f(u)$ au voisinage de t. Si $f(t) = c$ alors on vérifie que $L[f(t)] = c$. On verra comment optimiser le choix de $h(u)$ pour enlever au mieux les fluctuations irrégulières de $f(t)$ dues à un bruit de mesure.

2.1.3 Fonction de transfert

Les exponentielles complexes $e^{i\omega t}$ sont les vecteurs propres des opérateurs de convolution. En effet

$$L[e^{i\omega t}] = \int_{-\infty}^{+\infty} h(u)e^{i(t-u)\omega}\ du,$$

ce qui nous donne

$$L[e^{i\omega t}] = e^{it\omega}\int_{-\infty}^{+\infty} h(u)e^{-i\omega u}du = e^{it\omega}\hat{h}(\omega),$$

avec pour valeur propre

$$\hat{h}(\omega) = \int_{-\infty}^{+\infty} h(u)e^{-i\omega u}\ du.$$

La fonction $\hat{h}(\omega)$ est la transformée de Fourier de $h(t)$ et est appelée fonction de transfert du filtre. Les exponentielles étant les vecteurs propres d'un système linéaire homogène, il est tentant d'essayer d'exprimer le signal $f(t)$ comme somme d'exponentielles complexes, de façon à facilement calculer la réponse du filtre. L'analyse de Fourier prouve qu'une telle décomposition est possible, en imposant des conditions très faibles sur $f(t)$.

2.2 Analyse de Fourier

2.2.1 Transformée de Fourier dans $\mathbf{L^1}(\mathbb{R})$

Pour s'assurer que l'intégrale de Fourier

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t)e^{-i\omega t}dt \tag{2.9}$$

existe, on suppose que $f(t)$ est intégrable $f(t) \in \mathbf{L^1}(\mathbb{R})$. Cela nous permet d'étudier ses propriétés principales avant de l'étendre à d'autres classes de fonctions.

Lorsque $f(t) \in \mathbf{L^1}(\mathbb{R})$,

$$|\hat{f}(\omega)| \leq \int_{-\infty}^{+\infty} |f(t)|dt. \tag{2.10}$$

La transformée de Fourier est alors bornée et l'on peut vérifier que $\hat{f}(\omega)$ est continue (exercice). Si $\hat{f}(\omega) \in \mathbf{L^1}(\mathbb{R})$, on peut prouver [3] que l'opérateur de Fourier s'inverse et que

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)e^{i\omega t}d\omega. \tag{2.11}$$

La transformée de Fourier $\hat{f}(\omega)$ peut donc être interprétée comme l'amplitude de la composante sinusoïdale $e^{i\omega t}$ de fréquence ω dans $f(t)$. Au lieu de décrire $f(t)$ par ses valeurs à chaque instant, la transformée de Fourier donne une description de $f(t)$ en somme de sinusoïdes totalement délocalisées dans le temps.

Regularité Les composantes irrégulières de $f(t)$ sont reconstruites par les sinusoïdes $e^{i\omega t}$ qui oscillent rapidement et donc de hautes fréquences ω. Si la transformée de Fourier

$\hat{f}(\omega)$ décroît rapidement, cela signifie donc que $f(t)$ est une fonction régulière. Cette propriété se formalise en montrant que si

$$\int_{-\infty}^{+\infty} |\hat{f}(\omega)|(|\omega|^p + 1) < +\infty,$$

alors $f(t)$ est p fois continûment dérivable. Pour démontrer ce résultat, on prouve d'abord que la transformée de Fourier de $\frac{d^p f(t)}{dt^p} \in \mathbf{L^1}(\mathbb{R})$ est $(i\omega)^p \hat{f}(\omega)$ (exercice). On utilise ensuite le fait que si $\hat{g}(\omega) \in \mathbf{L^1}(\mathbb{R})$ alors $g(t)$ est bornée et continue, ce qui se montre en utilisant (2.11). L'équivalence entre régularité temporelle et décroissance du module de la transformée de Fourier est particulièrement importante pour analyser les propriétés d'un signal $f(t)$.

Pour les applications au traitement du signal, le résultat le plus important est le théorème de convolution.

Théorème 2.1 (Convolution) *Soit $f(t) \in \mathbf{L^1}(\mathbb{R})$ et $h(t) \in \mathbf{L^1}(\mathbb{R})$. La fonction $g(t) = h \star f(t) \in \mathbf{L^1}(\mathbb{R})$ et sa transformée de Fourier est*

$$\hat{g}(\omega) = \hat{h}(\omega)\hat{f}(\omega). \tag{2.12}$$

Démonstration

$$\hat{g}(\omega) = \int_{-\infty}^{+\infty} e^{-it\omega} \left\{ \int_{-\infty}^{+\infty} f(t-u)h(u)du \right\} dt.$$

Comme $|f(t-u)||h(u)|$ est sommable dans $\mathbb{R}^2$, on peut appliquer le théorème de Fubini et le changement de variable $(t,u) \to (v = t-u, u)$ nous donne

$$\begin{aligned} \hat{g}(\omega) &= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} e^{-i(u+v)\omega} f(v)h(u)dudv \\ &= \left\{ \int_{-\infty}^{+\infty} e^{-iv\omega} f(v)dv \right\} \left\{ \int_{-\infty}^{+\infty} e^{-iu\omega} h(u)du \right\}, \end{aligned}$$

ce qui vérifie (2.12). □

Filtrage Le théorème de convolution prouve que la transformée de Fourier d'un filtrage $L[f](t) = f \star h(t)$ est $\hat{f}(\omega)\hat{h}(\omega)$. La formule de reconstruction

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) e^{i\omega t} d\omega. \tag{2.13}$$

appliquée à $L[f](t)$ implique donc

$$L[f](t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{h}(\omega)\hat{f}(\omega) e^{i\omega t} d\omega. \tag{2.14}$$

Chaque composante fréquentielle $e^{i\omega t}$ de $f(t)$ d'amplitude $\hat{f}(\omega)$ est donc amplifiée ou atténuée par $\hat{h}(\omega)$. Ceci n'est pas surprenant puisque nous avons déjà prouvé que les exponentielles $e^{i\omega t}$ sont des vecteurs propres d'une convolution. Si $\hat{h}(\omega) = 0$ pour $\omega \in [\omega_1, \omega_2]$, les composantes fréquentielles de $f(t)$ pour $\omega \in [\omega_1, \omega_2]$ sont annulées par l'opérateur L, d'où l'appellation "filtre".

Propriétés générales Le tableau qui suit résume certaines propriétés importantes de la tranformée de Fourier. Les démonstrations se font le plus souvent par un simple changement de variable dans l'intégrale de Fourier.

Propriété	**Fonction**	**Transformée de Fourier**	
	$f(t)$	$\hat{f}(\omega)$	
Inverse	$\hat{f}(t)$	$2\pi f(-\omega)$	(2.15)
Convolution	$f_1 \star f_2$	$\hat{f}_1(\omega)\hat{f}_2(\omega)$	(2.16)
Multiplication	$f_1(t)f_2(t)$	$\frac{1}{2\pi}\hat{f}_1 \star \hat{f}_2(\omega)$	(2.17)
Translation	$f(t-t_0)$	$e^{-it_0\omega}\hat{f}(\omega)$	(2.18)
Modulation	$e^{i\omega_0 t}f(t)$	$\hat{f}(\omega-\omega_0)$	(2.19)
Dilatation	$f(at)$	$\frac{1}{\lvert a\rvert}\hat{f}(\frac{\omega}{a})$	(2.20)
Différentiation	$\frac{d^p f(t)}{dt^p}$	$(i\omega)^p\hat{f}(\omega)$	(2.21)
Multiplication Polynômiale	$(-it)^p f(t)$	$\frac{d^p\hat{f}(\omega)}{d\omega^p}$	(2.22)
Complexe Conjugué	$f^*(t)$	$\hat{f}^*(-\omega)$	(2.23)
Symétrie Hermitienne	$f(t) = \mathrm{Re}f(t)$	$\hat{f}(-\omega) = \hat{f}^*(\omega)$	(2.24)
Composante Réelle	$\mathrm{Re}f(t)$	$\frac{\hat{f}(\omega)+\hat{f}^*(-\omega)}{2}$	(2.25)
Composante Imaginaire	$\mathbf{Im}f(t)$	$\frac{\hat{f}(\omega)-\hat{f}^*(-\omega)}{2i}$	(2.26)
Composante Paire	$\frac{f(t)+f^*(-t)}{2}$	$\mathrm{Re}\hat{f}(\omega)$	(2.27)
Composante Impaire	$\frac{f(t)-f^*(-t)}{2}$	$\mathbf{Im}\hat{f}(\omega)$	(2.28)

2.2.2 Transformée de Fourier dans $\mathbf{L^2}(\mathbb{R})$ et Diracs

Plutôt que de travailler dans $\mathbf{L^1}(\mathbb{R})$, il est souvent plus facile de considérer les signaux comme des éléments de l'espace de Hilbert $\mathbf{L^2}(\mathbb{R})$ car on a alors accès à toutes les facilités données par l'existence d'un produit scalaire. Le produit scalaire de $f(t) \in \mathbf{L^2}(\mathbb{R})$ et $g(t) \in \mathbf{L^2}(\mathbb{R})$ est défini par

$$< f, g >= \int_{-\infty}^{+\infty} f(t)g^*(t)dt,$$

et la norme

$$\|f\|^2 =< f, f >= \int_{-\infty}^{+\infty} |f(t)|^2 dt.$$

Pour facilement travailler dans cette structure Hilbertienne, il nous faut y définir la transformée de Fourier. Cela pose un problème car l'intégrale de Fourier (2.9) d'une fonction

de carré intégrable n'est pas toujours convergente.

Conservation d'énergie La transformée de Fourier s'étend à partir de $\mathbf{L^1}(\mathbb{R})$ par un argument de densité qui utilise le fait qu'à une constante près, la norme et les angles dans $\mathbf{L^2}(\mathbb{R})$ ne sont pas modifiés par transformée de Fourier.

Théorème 2.2 *Soient $f(t)$ et $h(t)$ dans $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$,*

$$< f, h >= \int_{-\infty}^{+\infty} f(t)h^*(t)dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{h}^*(\omega)d\omega = \frac{1}{2\pi} < \hat{f}, \hat{h} > . \tag{2.29}$$

Pour $h = f$ on obtient la formule de Plancherel

$$||f||^2 = \int_{-\infty}^{+\infty} |f(t)|^2 dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega)|^2 d\omega = \frac{1}{2\pi}||\hat{f}||^2. \tag{2.30}$$

Démonstration de (2.29)
Prenons $g(t) = f \star \tilde{h}(t)$ avec $\tilde{h}(t) = h^*(-t)$. Le théorème de convolution en (2.12) et (2.23) montre que $\hat{g}(\omega) = \hat{f}(\omega)\hat{h}^*(\omega)$. La formule de reconstruction (2.11) appliquée à $g(0)$ nous donne

$$\int_{-\infty}^{+\infty} f(t)h^*(t)dt = g(0) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{g}(\omega)d\omega = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \hat{f}(\omega)\hat{h}^*(\omega)d\omega. \tag{2.31}$$

□

Extension dans $\mathbf{L^2}(\mathbb{R})$ Pour définir la transformée de Fourier de $f(t) \in \mathbf{L^2}(\mathbb{R})$, on construit une famille $\{f_n\}_{n\in\mathbb{Z}}$ de fonctions dans $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$ qui convergent vers f

$$\lim_{n\to+\infty} \|f - f_n\| = 0.$$

Ceci est possible car $\mathbf{L^1}(\mathbb{R}) \cap \mathbf{L^2}(\mathbb{R})$ est dense dans $\mathbf{L^2}(\mathbb{R})$. La famille $\{f_n(t)\}_{n\in\mathbb{Z}}$ est une suite de Cauchy, et donc $\|f_n - f_p\|$ est arbitrairement petit pour n et p suffisamment grands. Comme $f_n(t) \in \mathbf{L^2}(\mathbb{R})$, la formule de Plancherel montre que $\hat{f}_n(\omega) \in \mathbf{L^2}(\mathbb{R})$. De plus, $\{\hat{f}_n(\omega)\}_{n\in\mathbb{Z}}$ est aussi une suite de Cauchy car

$$\|\hat{f}_n - \hat{f}_p\| = \sqrt{2\pi}\|f_n - f_p\|$$

est arbitrairement petit pour n et p suffisamment grands. Comme toute suite de Cauchy converge dans $\mathbf{L^2}(\mathbb{R})$, il existe $\hat{f}(\omega) \in \mathbf{L^2}(\mathbb{R})$ tel que

$$\lim_{n\to+\infty} \|\hat{f} - \hat{f}_n\| = 0.$$

Par définition, $\hat{f}(\omega)$ est la tranformée de Fourier de $f(t)$. On peut alors vérifier que le théorème de convolution, la formule de Parseval et les propriétés (2.15-2.28) restent valables dans $\mathbf{L^2}(\mathbb{R})$.

Dirac La transformée de Fourier d'un Dirac peut simplement se calculer en se souvenant que pour toute fonction continue $f(t)$

$$\int_{-\infty}^{+\infty} f(t)\delta(t)dt = f(0).$$

La transformee de Fourier de $\delta(t)$ est donc

$$\int_{-\infty}^{+\infty} e^{-i\omega t}\delta(t)dt = 1.$$

C'est la fonction constante égale à 1. La théorie des distributions [4] montre que l'on peut definir la transformée de Fourier pour toute distribution tempérée.

2.2.3 Exemples de transformée de Fourier

• Une *Gaussienne* $f(t) = e^{-t^2}$ étant une fonction de la classe de Schwartz, sa transformée de Fourier est aussi une fonction $\mathbf{C}^\infty$ à décroissance rapide. Pour calculer sa transformée de Fourier, on montre par une intégration par partie (exercice) que sa transformée de Fourier

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} e^{-t^2}e^{-i\omega t}dt$$

satisfait l'équation différentielle

$$2\hat{f}'(\omega) + \omega\hat{f}(\omega) = 0.$$

La solution est une gaussienne

$$\hat{f}(\omega) = Ke^{-\omega^2/4},$$

et comme

$$\hat{f}(0) = \int_{-\infty}^{+\infty} e^{-t^2}dt = \sqrt{\pi},$$

$$\hat{f}(\omega) = \sqrt{\pi}e^{-\omega^2/4}. \tag{2.32}$$

• La *fonction indicatrice* $f(t) = 1_{[-T,T]}(t)$ est discontinue et donc a une transformée de Fourier qui n'est pas dans $\mathbf{L^1}(\mathbb{R})$ mais qui est dans $\mathbf{L^2}(\mathbb{R})$

$$\hat{f}(\omega) = \int_{-T}^{T} e^{-i\omega t}dt = \frac{2\sin(T\omega)}{\omega}. \tag{2.33}$$

• Le *sinus cardinal* $f(t) = \frac{\sin \pi t}{\pi t}$ est dans $\mathbf{L^2}(\mathbb{R})$ mais pas dans $\mathbf{L^1}(\mathbb{R})$. Sa transformée de Fourier se déduit de (2.33) grâce à la propriété de symétrie (2.15)

$$\hat{f}(\omega) = 1_{[-\pi,\pi]}(\omega). \tag{2.34}$$

• Un *Dirac* translaté $\delta_\tau(t) = \delta(t-\tau)$ a une transformée de Fourier qui se calcul directement

$$\hat{\delta}_\tau(\omega) = \int_{-\infty}^{+\infty} \delta_\tau(t)e^{-i\omega t}dt = e^{-i\omega\tau}. \tag{2.35}$$

• La *valeur principale* $f(t) = \mathrm{vp}\frac{1}{\pi t}$ définit par convolution la transformée de Hilbert

$$\mathcal{H}[g](t) = g \star \mathrm{vp}\frac{1}{\pi t} = \frac{1}{\pi}\int_{-\infty}^{+\infty} g(u)\mathrm{vp}\frac{1}{t-u}du. \tag{2.36}$$

On calcule la transformée de Fourier de $\mathrm{vp}\frac{1}{\pi t}$ en observant que

$$tf(t) = \frac{1}{\pi},$$

ce qui se traduit en Fourier grâce à (2.22) par

$$\frac{d\hat{f}(\omega)}{d\omega} = -2i\delta(\omega).$$

Donc

$$\hat{f}(\omega) = -i\ \mathrm{sign}(\omega) + c \tag{2.37}$$

avec

$$\mathrm{sign}(\omega) = \begin{cases} 1 & \text{si } \omega > 0 \\ 0 & \text{si } \omega = 0 \\ -1 & \text{si } \omega < 0 \end{cases}$$

Comme $f(t)$ est réelle antisymétrique, sa transformée de Fourier est imaginaire pure antisymétrique ce qui prouve que $c = 0$.

• Le *peigne de Dirac*

$$c(t) = \sum_{n=-\infty}^{+\infty} \delta(t - nT) \tag{2.38}$$

est une distribution dont la transformée de Fourier se déduit de (2.35)

$$\hat{c}(\omega) = \sum_{n=-\infty}^{+\infty} e^{-inT\omega}. \tag{2.39}$$

La formule de Poisson prouve que $\hat{c}(\omega)$ est aussi égal à un peigne de Dirac dont l'espacement est $\frac{2\pi}{T}$.

Théorème 2.3 (Formule de Poisson)

$$\sum_{n=-\infty}^{+\infty} e^{-inT\omega} = \frac{2\pi}{T}\sum_{k=-\infty}^{+\infty} \delta(\omega - \frac{2\pi k}{T}). \tag{2.40}$$

Démonstration Comme $C(\omega) = \sum_{n=-\infty}^{+\infty} e^{-inT\omega}$ est $2\pi/T$ périodique, il suffit de prouver que sa restriction à $[-\pi/T, \pi/T]$ est égale à $2\pi/T\delta(\omega)$. Pour tout $\phi(\omega) \in \mathbf{C}_0^\infty$ dont le support est inclus dans $[-\pi/T, \pi/T]$, on veut montrer que

$$< C, \phi >= \lim_{N\to+\infty} \int_{-\infty}^{+\infty} \sum_{n=-N}^{+N} e^{-inT\omega}\phi(\omega)d\omega = \frac{2\pi}{T}\phi(0). \tag{2.41}$$

La somme de cette série géométrique est

$$\sum_{n=-N}^{+N} e^{-inT\omega} = \frac{\sin[(N+1/2)T\omega]}{\sin[T\omega/2]}. \tag{2.42}$$

Donc

$$< C, \phi >= \lim_{N\to+\infty} \frac{2\pi}{T} \int_{-\pi/T}^{+\pi/T} \frac{\sin[(N+1/2)T\omega]}{\pi\omega} \frac{T\omega/2}{\sin[T\omega/2]} \phi(\omega) d\omega. \tag{2.43}$$

Pour $|\omega| < \pi/T$, on définit

$$\hat{\psi}(\omega) = \phi(\omega) \frac{T\omega/2}{\sin[T\omega/2]}$$

tandis que $\hat{\psi}(\omega) = 0$ si $|\omega| > \pi/T$. Cette fonction est la transformée de Fourier de $\psi(t) \in \mathbf{L^2}(\mathbb{R})$. Comme $2\sin(a\omega)/\omega$ est la transformée de Fourier de $1_{[-a,a]}(t)$, la formule de Parseval (2.29) implique

$$< C, \phi >= \frac{2\pi}{T} \int_{-\infty}^{+\infty} \frac{\sin[(N+1/2)T\omega]}{\pi\omega} \hat{\psi}(\omega) d\omega = \frac{2\pi}{T} \int_{-(N+1/2)T}^{(N+1/2)T} \psi(t) dt. \tag{2.44}$$

Lorsque N tend vers $+\infty$ l'intégral converge vers $\hat{\psi}(0) = \phi(0)$. □

2.3 Synthèse de filtres

2.3.1 Filtres passe-bandes

La transformée de Fourier d'un signal filtré $g(t) = f \star h(t)$ est

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

De nombreuses applications nécessitent d'isoler les composantes du signal dans différentes bandes de fréquences.

Un filtre passe-bas idéal a une fonction de transfert définie par

$$\hat{h}_0(\omega) = \begin{cases} 1 & \text{si } |\omega| \leq \omega_c \\ 0 & \text{si } |\omega| > \omega_c \end{cases} \tag{2.45}$$

Il élimine donc toutes les fréquences de $\hat{f}(\omega)$ au delà de ω_c. On déduit de (2.34) que la réponse impulsionnelle de ce filtre est

$$h_0(t) = \frac{\sin(\omega_c t)}{\pi t}.$$

Ce filtre passe-bas idéal est ni causal ni stable. Le paragraphe suivant explique comme l'approximer avec un système physiquement réalisable.

Un filtre passe-bande réel a une fonction de transfert qui supprime toute composante fréquentielle en dehors de deux intervalles symétriques par rapport à $\omega = 0$

$$\hat{h}_1(\omega) = \begin{cases} 1 & \text{si } |\omega| \in [\omega_0 - \omega_c, \omega_0 + \omega_c] \\ 0 & \text{ailleurs} \end{cases} \tag{2.46}$$

Un tel filtre peut se déduire d'un filtre passe-bas. Comme

$$\hat{h}_1(\omega) = \hat{h}_0(\omega - \omega_0) + \hat{h}_0(\omega + \omega_0).$$

Comme la transformée de Fourier de $h_0(t)e^{i\omega_0 t}$ est $\hat{h}_0(\omega - \omega_0)$ on déduit que

$$h_1(t) = 2\cos(\omega_0 t)\ h_0(t) = 2\cos(\omega_0 t)\ \frac{\sin \omega_c t}{\pi t}.$$

Ce filtre est généralement approximé par un filtre causal et stable, en remplaçant $h_0(t)$ par une approximation stable et causale.

2.3.2 Filtrage par circuits électroniques

Un filtrage linéaire analogique est le plus souvent implémenté avec un circuit électronique. Le signal $f(t)$ est représenté par une différence de potentiel $u(t) = f(t)$ appliquée à l'entrée du circuit. Pour certaines réponses impulsionnelles $h(t)$, nous allons montrer que l'on peut configurer le circuit de façon à ce que la différence de potentiel $v(t)$ à la sortie soit égale au produit de convolution $v(t) = u \star h(t)$ (voire figure 2.1).

Les circuits VLSI analogiques sont essentiellement composés de résistances, de capacités, et d'amplificateurs opérationnels, construits avec des transistors. Les inductances ne sont pas utilisées car elles demandent trop de place sur le silicium, mais elles sont remplacées par des circuits équivalents. Ce type de circuit relie les différences de potentiels à l'entrée et à la sortie par une équation différentielle à coefficients constants

$$a_N \frac{d^N v(t)}{dt^N} + ... + a_1 \frac{dv(t)}{dt} + a_0 v(t) = b_M \frac{d^M u(t)}{dt^M} + ... + b_1 \frac{du(t)}{dt} + b_0 u(t). \tag{2.47}$$

On suppose que $u(t)$ est un signal causal, $u(t) = 0$ pour $t < 0$, et l'on veut calculer la solution $v(t)$ de cette équation différentielle. Cette solution dépend des conditions initiales à la sortie du circuit spécifiées par $\{\frac{d^k v(0)}{dt^k}\}_{0 \leq k < N}$. Nous supposerons que le circuit est initialement au repos ce qui signifie que toutes ces dérivées sont nulles. La sortie $v(t)$ est alors reliée a $u(t)$ par un opérateur linéaire homogène dont nous calculons la fonction de transfert.

Fonction de transfert La propriété de différentiation (2.21) permet de calculer la transformée de Fourier de chaque côté de l'égalité (2.47)

$$\begin{aligned} a_N (i\omega)^N \hat{v}(\omega) + ... + a_1 (i\omega)\hat{v}(\omega) + a_0 \hat{v}(\omega) \quad &= \\ b_M (i\omega)^M \hat{u}(\omega) + ... + b_1 (i\omega)\hat{u}(\omega) + b_0 \hat{u}(\omega). & \end{aligned}$$

La fonction de transfert est donc

$$\hat{h}(\omega) = \frac{\hat{v}(\omega)}{\hat{u}(\omega)} = \frac{b_M(i\omega)^M + ... + b_1(i\omega) + b_0}{a_N(i\omega)^N + ... + a_1(i\omega) + a_0}. \tag{2.48}$$

Cette fonction de transfert est aussi appelée l'impédance du circuit.

Dans le cas d'un circuit électronique, on a $N \geq M$, car $|\hat{h}(\omega)|$ ne peut pas tendre ver $+\infty$ à haute fréquences. La sortie du circuit initialement au repos peut s'écrire

$$v(t) = \int_{-\infty}^{+\infty} h(\tau)u(t-\tau)d\tau = \int_{0}^{+\infty} h(\tau)u(t-\tau)d\tau,$$

car la réponse impulsionnelle $h(t)$ est causale.

Exemple Le circuit RC avec amplification de la figure 2.1 est un exemple particulièrement simple qui relie l'entrée et la sortie par l'équation

$$RC\frac{dv(t)}{dt} + v(t) = (1 + \frac{R_2}{R_1})u(t).$$

L'impédance est donc

$$\hat{h}(\omega) = \frac{1 + \frac{R_2}{R_1}}{1 + RCi\omega}.$$

One peut vérifier (exercice) que la réponse impulsionnelle du filtrage homogène est causale et s'exprime à partir de la fonction échelon (2.7) par

$$h(t) = \frac{1}{RC}\,(1 + \frac{R_2}{R_1})\,e^{-\frac{t}{RC}}\,\gamma(t).$$

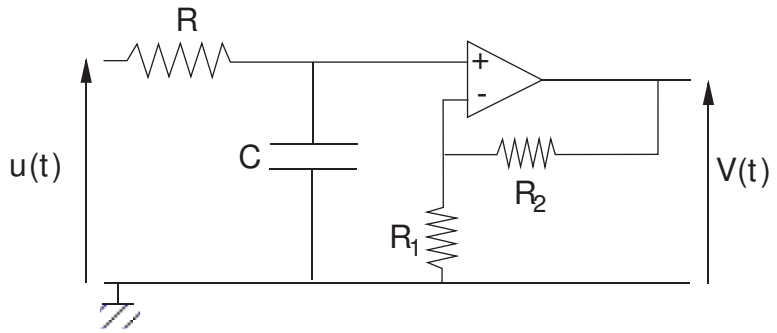

Figure 2.1: Circuit RC avec un amplificateur opérationnel

2.3.3 Approximations par filtres rationnels

Nous avons vu qu'un circuit électronique implémente un filtre dont la fonction de transfert est une fonction rationnelle de $i\omega$

$$\hat{h}(\omega) = \frac{N(i\omega)}{D(i\omega)}, \tag{2.49}$$

où $N(u)$ et $D(u)$ sont des polynômes à coefficients réels. On peut démontrer (exercice) que le filtre est causal et stable si et seulement si $D(s)$ est un polynôme dont les racines ont des parties réelles strictement négatives. Par ailleurs on montre aussi que si $P(\omega)$ est une fonction rationnelle de $i\omega$ avec $P(\omega) \geq 0$ pour tout $\omega \in \mathbb{R}$, alors il existe une fonction de transfert rationelle, correspondant à un filtre causal et stable, qui satisfait

$$|\hat{h}(\omega)|^2 = P(\omega).$$

Un filtre passe-bas idéal

$$\hat{h}_0(\omega) = 1_{[-\omega_c,\omega_c]}(\omega)$$

n'est pas réalisable par un circuit électrique car sa fonction de transfert n'est pas rationnelle. Le nombre d'éléments (résistances, capacités, amplificateurs) nécessaires pour implémenter une fonction de transfert rationnelle $\hat{h}(\omega)$ est proportionnel au degré du dénominateur. Pour limiter la complexité du circuit, on veut donc approximer $|\hat{h}_0(\omega)|^2$ par une fonction rationnelle de faible degré, tout en minimisant l'erreur d'approximation. L'erreur d'approximation est définie par un gabarit illustré par la figure 2.2, qui spécifie l'amplitude maximum des oscillations de $|\hat{h}(\omega)|^2$ dans les bandes de passage et d'atténuation ainsi que la largeur de la bande de transition. Le problème est donc de trouver des fonctions rationnelles de degré le plus faible possible, qui satisfont les contraintes de gabarit imposées par une application. Les polynômes de Butterworth ou de Chebyshev ont des propriétés particulièrement bien adaptées à ce type d'approximation.

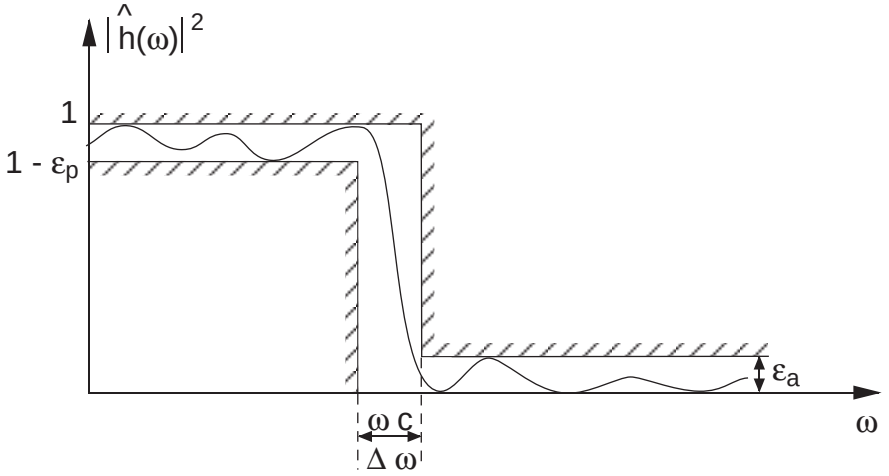

Figure 2.2: Le gabarit d'un filtre spécifie l'amplitude maximum des oscillations ϵ_p et ϵ_a dans les bandes de passage et d'atténuation du filtre, ainsi que la largeur $\Delta\omega$ de la bande de transition

Filtres de Butterworth Un filtre de Butterworth d'ordre n est défini par

$$|\hat{h}_n^b(\omega)|^2 = \frac{1}{1 + (\omega/\omega_c)^{2n}}. \tag{2.50}$$

Plus n augmente plus le filtre est plat au voisinage de $\omega = 0$ car les $2n-1$ premières dérivées de $|\hat{h}(\omega)|^2$ sont nulles en $\omega = 0$. A la fréquence de coupure ω_c, $|\hat{h}_n^b(\omega_c)|^2 = \frac{1}{2}$. Ces

filtres convergent vers le filtre passe-bas idéal (2.45), au sens òu pour tout $\omega \in \mathbb{R} - \{\omega_c\}$

$$\lim_{n\to+\infty} |\hat{h}_n^b(\omega)|^2 = |\hat{h}_0(\omega)|^2.$$

Filtres de Chebyshev Les filtres de Chebyshev ne sont pas plats au voisinage de $\omega = 0$ mais ils ont des oscillations de tailles constantes dans la bande de passage. A degré égal, ils ont une bande de transition plus faible que les filtres de Butterworth. Ils sont définis par

$$|\hat{h}_n^c(\omega)|^2 = \frac{1}{1+\epsilon^2 C_n^2(\omega/\omega_c)},$$

où $C_n(\omega)$ est un polynôme de Chebyshev de degré n qui peut s'écrire

$$C_n(\omega) = \begin{cases} \cos(n\cos^{-1}\omega) & \text{si } 0 \le |\omega| \le 1 \\ \cosh(n\cosh^{-1}\omega) & \text{si } |\omega| \ge 1 \end{cases} \tag{2.51}$$

Ces polynômes peuvent aussi être caractérisés par récurrence

$$C_0(\omega) = 1 \quad , \quad C_1(\omega) = \omega,$$

$$C_{n+1}(\omega) = 2\omega C_n(\omega) - C_{n-1}(\omega).$$

Pour $|\omega| < 1$, $|C_n(\omega)|^2$ oscille régulièrement entre 0 et 1 tandis que lorsque $|\omega| > 1$, le cosinus hyperbolique augmente de façon monotone. En conséquence $|\hat{h}_n^c(\omega)|^2$ oscille entre 1 et $\frac{1}{1+\epsilon^2}$ lorsque $0 \le |\omega|/\omega_c \le 1$. Lorsque $|\omega|/\omega_c \ge 1$ alors $|\hat{h}_n^c(\omega)|^2$ a une décroissance monotone vers 0.

Il existe d'autres familles de fonctions rationnelles utilisées pour l'approximation du filtre passe-bas idéal et le choix d'une approximation pour une application est un art qui dépend du type de distortions que l'on peut admettre.

2.4 Modulation d'amplitude

A travers un canal unique de transmission il est souvent nécessaire de transmettre plusieurs signaux simultanément, comme par exemple des émissions de radio ou des conversations téléphoniques. Lorsque ces signaux peuvent être bien approximés par des fonctions dont la transformée de Fourier est à support compact, la modulation d'amplitude permet de multiplexer ces signaux pour les transmettre en même temps. L'audition n'étant essentiellement sensible qu'à des sons entre 300Hz et 3300Hz, on peut limiter les sons par filtrage passe-bas à l'intervalle de fréquence $[-3300, 3300]$, lors de leur transmission téléphonique ou radio.

On suppose que les signaux $\{f_n\}_{0\le n<N}$ que l'on veut multiplexer ont tous une transformée de Fourier dont le support est inclu dans $[-b, b]$. La modulation d'amplitude permet de multiplexer ces N signaux en un seul signal dont la bande de fréquence est N fois plus grande. Pour cela on transforme chaque signal $f_n(t)$ en un signal modulé $f_n^m(t)$ dont la transformée de Fourier a un support égal à $[-\omega_n - b, -\omega_n] \cup [\omega_n, b + \omega_n]$. En

choisissant $\omega_n = nb$, le support de $f_n^m(\omega)$ n'intersecte pas le support de $f_p^m(\omega)$ si $n \neq p$. A partir du signal multiplexé

$$M(t) = \sum_{n=0}^{N-1} f_n^m(t) \tag{2.52}$$

on peut récupérer chaque signal f_n^m par filtrage passe-bande, puis on reconstruit $f_n(t)$ par démodulation. Les paragraphes suivants expliquent le calcul de ces différentes étapes.

2.4.1 Signal analytique et transformée de Hilbert

Les signaux réels ayant une transformée de Fourier à symétrie hermitienne, leur support est symétrique par rapport à $\omega = 0$. On veut donc transposer les fréquences de $f_n(t)$ de l'intervalle $[-b, b]$ à un double intervalle symétrique $[-\omega_n - b, -\omega_n] \cup [\omega_n, \omega_n + b]$. Pour cela nous considérons séparément les fréquences positives et négatives de $f_n(t)$, comme l'illustre la figure 2.3.

Comme $\hat{f}_n(\omega) = \hat{f}_n^*(-\omega)$, $\hat{f}_n(\omega)$ est entièrement caractérisé par sa restriction à $\omega > 0$ donnée par

$$\hat{f}_n^z(\omega) = \begin{cases} 2\hat{f}_n(\omega) & \text{si } \omega > 0 \\ \hat{f}_n(0) & \text{si } \omega = 0 \\ 0 & \text{si } \omega < 0 \end{cases} \tag{2.53}$$

La fonction $f_n^z(t)$ est appelée partie analytique de $f_n(t)$ car on peut démontrer qu'elle admet une extension analytique dans la partie supérieure du plan complexe que l'on construit grâce à la transformée de Laplace [Bony]. Les propriétés (2.25, 2.26) de la transformée de Fourier montrent que la transformée de Fourier de la partie réelle de $f_n^z(t)$ est

$$\frac{\hat{f}_n^z(\omega) + \hat{f}_n^{z*}(-\omega)}{2} = \hat{f}_n(\omega),$$

et donc que $\mathrm{Re} f_n^z(t) = f_n(t)$. De même la transformée de Fourier de sa partie imaginaire est

$$\frac{\hat{f}_n^z(\omega) - \hat{f}_n^{z*}(-\omega)}{2i} = -i \ \mathrm{sign}(\omega) \hat{f}_n(\omega). \tag{2.54}$$

Nous avons vu en (2.37) que $-i \ \mathrm{sign}(\omega)$ est la fonction de transfert du filtre de Hilbert. La partie imaginaire de $f_n^z(t)$ est donc la transformée de Hilbert de $f(t)$

$$\mathbf{Im} f_n^z(t) = \mathcal{H}[f_n](t) = f_n \star \mathrm{vp} \frac{1}{\pi t} = \frac{1}{\pi} \int_{-\infty}^{+\infty} f_n(u) \ \mathrm{vp} \frac{1}{t-u} du. \tag{2.55}$$

Modulation d'amplitude Pour transposer en fréquence $\hat{f}_n(\omega)$ et obtenir une fonction $\hat{f}_n^m(\omega)$ dont le support est $[-b - \omega_n, -\omega_n] \cup [\omega_n, \omega_n + b]$, on décale de ω_n les fréquences positives $\hat{f}_n^z(\omega)$ et de $-\omega_n$ les fréquences négatives $\hat{f}_n^z(-\omega)$ (figure 2.3)

$$\hat{f}_n^m(\omega) = \frac{\hat{f}_n^z(\omega - \omega_n) + \hat{f}_n^{z*}(-\omega - \omega_n)}{2}. \tag{2.56}$$

On calcule avec (2.25) et (2.19) la transformée de Fourier inverse du côté droit de cette équation, ce qui nous donne

$$f_n^m(t) = \mathrm{Re}[f_n^z(t)e^{i\omega_n t}]. \tag{2.57}$$

Comme $f_n^z(t) = f_n(t) + i\mathcal{H}[f_n](t)$, la modulation d'amplitude s'exprime à partir de la transformée de Hilbert par

$$f_n^m(t) = f_n(t)cos(\omega_n t) - \mathcal{H}[f_n](t)sin(\omega_n t). \tag{2.58}$$

Figure 2.3: Multiplexage par modulation d'amplitude

2.4.2 Démodulation et détection synchrone

Le signal multiplexé $M(t)$ (2.52) est la somme de signaux modules $f_n^m(t)$ dont les supports fréquentiels ne s'intersectent pas. On définit un filtre passe-bande dont le support est le même que celui de $\hat{f}_n^m(\omega)$

$$\hat{h}_n(\omega) = \begin{cases} 1 & \text{si} \quad |\omega| \in [\omega_n, \omega_n + b] \\ 0 & \text{ailleurs} \end{cases} \tag{2.59}$$

On a alors

$$f_n^m(t) = M \star h_n(t).$$

Le signal $f_n(t)$ se reconstruit à partir de $f_n^m(t)$ en supprimant la modulation d'amplitude. L'équation (2.58) montre que

$$g_n(t) = 2f_n^m(t)cos(\omega_n t) = f_n(t) + f_n(t)cos(2\omega_n t) - \mathcal{H}[f_n](t)sin(2\omega_n t), \tag{2.60}$$

ce qui s'écrit en Fourier

$$\begin{aligned} \hat{g}_n(\omega) = \hat{f}_n(\omega) &+ \frac{\hat{f}_n(\omega - 2\omega_n) + \hat{f}_n(\omega + 2\omega_n)}{2} \\ &+ \frac{\hat{\mathcal{H}[f_n]}(\omega - 2\omega_n) - \hat{\mathcal{H}[f_n]}(\omega + 2\omega_n)}{2i}. \end{aligned} \tag{2.61}$$

Comme le support de $\hat{f}_n(\omega)$ et de $\hat{\mathcal{H}[f_n]}(\omega)$ est $[-b, b]$ et que $\omega_n > b$, on sépare $\hat{f}(\omega)$ des autres composantes fréquentielles avec la fonction de transfert

$$\hat{h}_0(\omega) = \begin{cases} 1 & \text{si } |\omega| \leq b \\ 0 & \text{ailleurs} \end{cases}. \tag{2.62}$$

On déduit de (2.61) que

$$f_n(t) = g_n \star h_0(t) = (2f_n^m(u)cos(\omega_n u) \star h_0(u))(t).$$

Chapitre 3

Traitement du signal discret

Le traitement du signal discret a pris son essor dans les années 70 grâce à l'apparition des microprocesseurs et à l'utilisation de la transformée de Fourier rapide. Il remplace progressivement le traitement du signal analogique dans la majorité des applications telles que l'enregistrement digital, la télévision, le traitement de la parole et de l'image. Le calcul informatique permet la mise en place d'algorithmes nettement plus sophistiqués et plus précis que le calcul analogique dont la fiabilité est limitée par les bruits de circuits et les erreurs de calibrage des composants électroniques. Le traitement du signal analogique reste cependant beaucoup plus rapide ce qui est fondamental pour certaines applications en temps réel.

Les signaux étant le plus souvent d'origine analogique, nous étudions la conversion analogique-digitale et les conditions permettant d'effectuer la transformation inverse. Le filtrage homogène est étendu au calcul discret et nous introduisons le calcul rapide par transformée de Fourier discrète.

3.1 Conversion analogique-digitale

L'approche la plus simple pour discrétiser une fonction $f(t)$ est d'effectuer un échantillonnage avec un intervalle T uniforme, en enregistrant les valeurs $\{f(nT)\}_{n\in\mathbb{Z}}$. Pour effectuer la transformation inverse, nous étudions l'existence d'algorithmes d'interpolation permettant de reconstruire $f(t)$ à partir de ses échantillons.

3.1.1 Echantillonnage

Pour traiter les signaux discrets dans le même cadre que les signaux analogiques, nous les représentons par des distributions de Dirac. Un échantillon $f(nT)$ est représenté par un Dirac d'amplitude $f(nT)$ centré en nT. L'échantillonnage uniforme de $f(t)$ correspond à la distribution

$$f_d(t) = \sum_{n=-\infty}^{+\infty} f(nT)\delta(t-nT). \tag{3.1}$$

Puisque $f(nT)\delta(t-nT) = f(t)\delta(t-nT)$,

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

Un échantillonnage uniforme est donc obtenu par multiplication avec le peigne de Dirac

$$c(t) = \sum_{n=-\infty}^{+\infty} \delta(t-nT). \tag{3.2}$$

Les propriétés de cet échantillonnage s'étudient plus facilement dans le domaine de Fourier. Si $\{f(nT)\}_{n\in\mathbb{Z}}$ est borné, $f_d(t)$ est une distribution tempérée [3] dont la transformée de Fourier $\hat{f}_d(\omega)$ est bien définie. La transformée de Fourier de $\delta(t-nT)$ étant $e^{-inT\omega}$, on déduit de (3.1) que $\hat{f}_d(\omega)$ est une série de Fourier $\frac{2\pi}{T}$ périodique

$$\hat{f}_d(\omega) = \sum_{n=0}^{+\infty} f(nT) e^{-inT\omega}. \tag{3.3}$$

Pour comprendre comment reconstruire $f(t)$ à partir de ses échantillons, nous exprimons $\hat{f}_d(\omega)$ en fonction de $\hat{f}(\omega)$. Comme $f_d(t) = f(t)c(t)$, sa transformée de Fourier peut aussi s'écrire

$$\hat{f}_d(\omega) = \frac{1}{2\pi} \hat{f} \star \hat{c}(\omega).$$

La formule de Poisson (2.40) prouve que la transformée de Fourier du peigne de Dirac $c(t)$ est

$$\hat{c}(\omega) = \frac{2\pi}{T} \sum_{k=-\infty}^{+\infty} \delta(\omega - \frac{2\pi k}{T}). \tag{3.4}$$

Comme $\hat{f} \star \delta(\omega - \frac{2\pi k}{T}) = \hat{f}(\omega - \frac{2\pi k}{T})$,

$$\hat{f}_d(\omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} \hat{f}(\omega - \frac{2k\pi}{T}). \tag{3.5}$$

Echantillonner un signal est donc équivalent à une périodisation de sa transformée de Fourier, obtenue en additionnant les translatées $\hat{f}(\omega - \frac{2k\pi}{T})$. Le théorème de Nyquist donne une condition suffisante sur le support de $\hat{f}(\omega)$ pour reconstruire $f(t)$ à partir des échantillons $f(nT)$. Cette condition garantit que $f(t)$ n'a pas d'oscillations violentes entre chaque paire d'échantillons.

Théorème 3.1 (Nyquist) *Soit $f(t)$ un signal dont la transformée de Fourier $\hat{f}(\omega)$ a un support inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$. Alors $f(t)$ peut être reconstruite en interpolant ses échantillons*

$$f(t) = \sum_{n=-\infty}^{+\infty} f(nT) h_T(t-nT), \tag{3.6}$$

avec

$$h_T(t) = \mathrm{sinc}(\frac{\pi t}{T}) = \frac{\sin \frac{\pi t}{T}}{\frac{\pi t}{T}}. \tag{3.7}$$

Démonstration
Comme le support de $\hat{f}(\omega)$ est inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$, si $n \neq 0$ le support de $\hat{f}(\omega - \frac{2n\pi}{T})$ n'intersecte pas le support de $\hat{f}(\omega)$. En conséquence (3.5) prouve que

$$\hat{f}_d(\omega) = \frac{\hat{f}(\omega)}{T} \quad \text{si} \quad |\omega| \leq \frac{\pi}{T}. \tag{3.8}$$

Soit $\hat{h}_T(\omega)$ la fonction de transfert d'un filtre passe-bas idéal

$$\hat{h}_T(\omega) = \begin{cases} T & \text{si } |\omega| \leq \frac{\pi}{T} \\ 0 & \text{si } |\omega| > 0 \end{cases} \tag{3.9}$$

et dont la réponse impulsionnelle $h_T(t)$ est donnée par (3.7). On déduit de (3.8) que

$$\hat{f}(\omega) = \hat{h}_T(\omega)\hat{f}_d(\omega)$$

ce qui se traduit en variable de temps par

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

□

Le théorème d'échantillonnage de Nyquist donne une condition nécessaire pour reconstruire un signal à partir de ses échantillons étant donnée une information à priori sur son support fréquentiel. L'échantillonnage et l'interpolation sont illustrés par la figure 3.1, dans les domaines temporels et fréquentiels. D'autres caractérisations peuvent être obtenues en imposant des contraintes différentes sur $f(t)$.

3.1.2 Repliement spectral

Si le support de $\hat{f}(\omega)$ n'est pas inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$, la formule d'interpolation (3.6) ne reconstruit pas $f(t)$. Nous étudions les propriétés de l'erreur de reconstruction ainsi qu'une procédure de filtrage pour la réduire.

Recouvrement fréquentiel La transformée de Fourier de $h_T \star f_d(t)$ a un support inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$ et donc ne peut être égale à $f(t)$ dont la transformée de Fourier a un support qui s'étend au-delà de $[-\frac{\pi}{T}, \frac{\pi}{T}]$. Nous avons vu que

$$\hat{f}_d(\omega) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} \hat{f}(\omega - \frac{2k\pi}{T}). \tag{3.10}$$

Lorsque le support de $\hat{f}(\omega)$ n'est pas inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$, pour certaines fréquences $\omega \in [-\frac{\pi}{T}, \frac{\pi}{T}]$ il existe des entiers $k \neq 0$ pour lesquels $\hat{f}(\omega - \frac{2k\pi}{T}) \neq 0$ (voire figure 3.2). Dans ce cas, $\hat{f}_d(\omega)$ est la somme de $\hat{f}(\omega)$ plus certaines composantes de hautes fréquences $\hat{f}(\omega - \frac{2k\pi}{T})$. La valeur de $\hat{f}_d(\omega)\hat{h}_T(\omega)$ peut donc être très différente de $\hat{f}(\omega)$ même lorsque $\omega \in [-\frac{\pi}{T}, \frac{\pi}{T}]$.

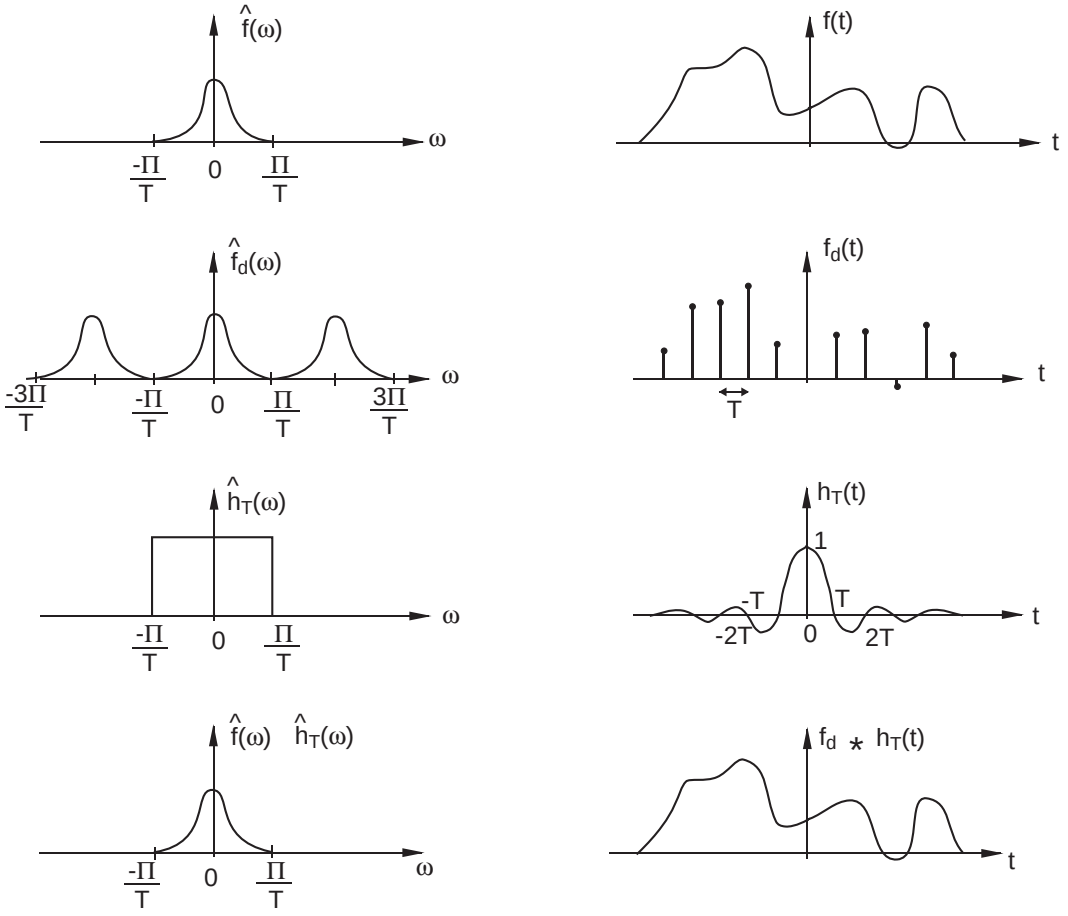

Figure 3.1: Echantillonnage et interpolation dans les domaines temporels et fréquentiels

Considérons par exemple le signal

$$f(t) = \cos(\omega_0 t) = \frac{e^{i\omega_0 t} + e^{-i\omega_0 t}}{2}$$

avec $\frac{2\pi}{T} > \omega_0 > \frac{\pi}{T}$. Sa transformée de Fourier étant

$$\hat{f}(\omega) = \pi\Big(\delta(\omega - \omega_0) + \delta(\omega + \omega_0)\Big),$$

la périodisation (3.5) nous donne

$$\hat{f}_d(\omega) = \frac{\pi}{T} \sum_{k=-\infty}^{+\infty} \Big(\delta(\omega - \omega_0 - \frac{2k\pi}{T}) + \delta(\omega + \omega_0 - \frac{2k\pi}{T})\Big).$$

Les seules composantes dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$ sont $\delta(\omega - \omega_0 + \frac{2\pi}{T}) + \delta(\omega + \omega_0 - \frac{2\pi}{T})$ donc après filtrage par le filtre passe-bas $h_T(\omega)$, on obtient

$$f_d \star h_T(t) = \cos\Big((\frac{2\pi}{T} - \omega_0)t\Big).$$

Le repliement spectral réduit la fréquence du cosinus de ω_0 à $\frac{2\pi}{T} - \omega_0 \in [-\frac{\pi}{T}, \frac{\pi}{T}]$. Ce repli fréquentiel s'observe lorsque l'on filme un mouvement trop rapide avec un nombre insuffisant d'images par seconde. Une roue de voiture tournant à grande vitesse apparaît comme tournant beaucoup plus lentement dans le film.

Préfiltrage Supposons que le pas d'échantillonnage est limité à une valeur T par des contraintes de temps calcul ou de mémoire et que $\omega_0 > \frac{\pi}{T}$. A défaut de reconstruire exactement $f(t)$, on veut récupérer la meilleure approximation de $f(t)$ par interpolation d'un échantillonnage avec $h_T(t)$. Une telle interpolation est une convolution avec $h_T(t)$ et a donc une transformée de Fourier dont le support est inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$. Soit $\mathbf{V}$ l'espace des fonctions dont les transformées de Fourier ont un support inclus dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$. La fonction de $\mathbf{V}$ qui est la plus proche de $f(t)$ est la projection orthogonale $P_{\mathbf{V}} f(t)$ de $f(t)$ dans $\mathbf{V}$ qui minimise

$$\|f - P_{\mathbf{V}} f\|^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} |\hat{f}(\omega) - \hat{P}_{\mathbf{V}} f(\omega)|^2 d\omega. \tag{3.11}$$

Comme $P_{\mathbf{V}} f(t) \in \mathbf{V}$, le support de sa transformée de Fourier $\hat{P}_{\mathbf{V}} f(\omega)$ est incluse dans $[-\frac{\pi}{T}, \frac{\pi}{T}]$. La distance (3.11) est minimisée si

$$\hat{P_{\mathbf{V}} f}(\omega) = \hat{f}(\omega) \quad \text{pour } |\omega| \leq \tfrac{\pi}{T}.$$

La projection orthogonale est donc obtenue par le filtrage linéaire

$$P_{\mathbf{V}} f(t) = \frac{1}{T} f \star h_T(t) \tag{3.12}$$

qui enlève toute composante fréquentielle au delà de la fréquence d'échantillonnage $\frac{\pi}{T}$. Puisque $P_{\mathbf{V}}f \in \mathbf{V}$, le théorème de Nyquist prouve que

$$P_{\mathbf{V}}f(t) = \sum_{n=-\infty}^{+\infty} P_{\mathbf{V}}f(nT)h_T(t-nT).$$

On calcule la projection orthogonale de $f(t)$ sur $\mathbf{V}$ en préfiltrant $f(t)$ avec (3.12) et cette projection orthogonale est reconstruite à partir de son échantillonnage uniforme. Un convertisseur analogique digital est donc composé d'un filtre qui limite la bande de fréquence du signal à $[-\frac{\pi}{T}, \frac{\pi}{T}]$ suivi d'un échantillonnage uniforme avec intervalles T. En pratique, l'implémentation par circuit électronique nécessite d'approximer le filtre passe-bas idéal $h_T(t)$ par un filtre réalisable (par exemple Butterworth ou Chebyshev).

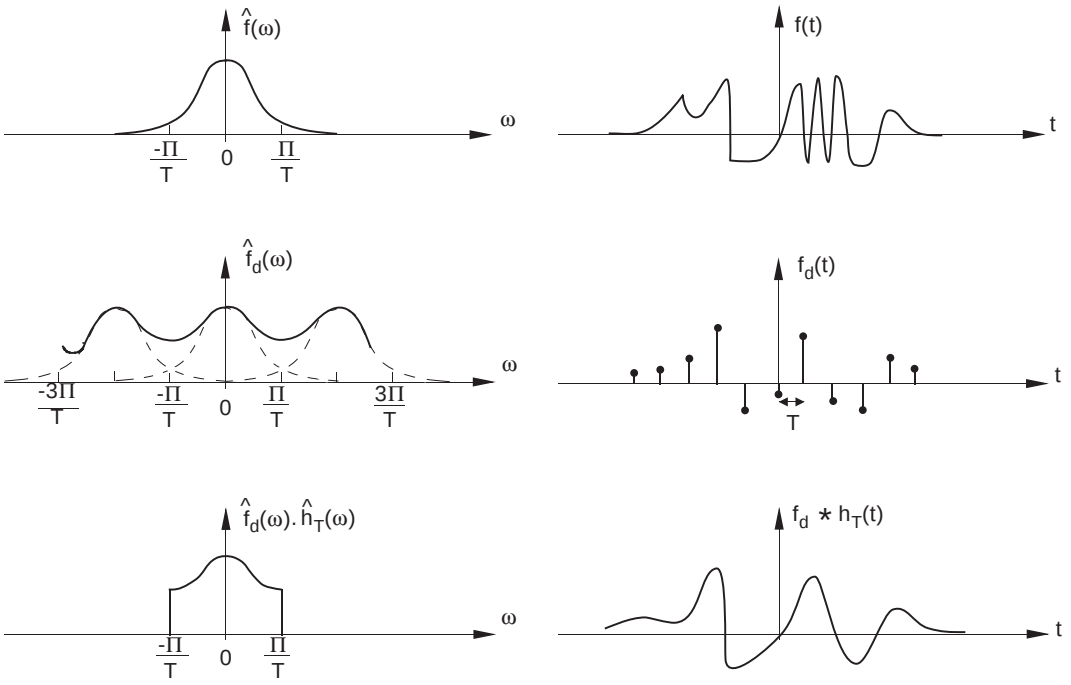

Figure 3.2: Cette figure illustre le recouvrement spectral créé par un pas d'échantillonnage trop grand. Le signal reconstruit $f_d * h_T(t)$ peut être très différent de $f(t)$

3.2 Filtrage discret homogène

Les opérateurs analogiques de filtrage linéaire homogène s'étendent aux signaux discrets en remplaçant les intégrales par des sommes discrètes. La transformée de Fourier est alors remplacée par les séries de Fourier. Les propriétés des filtres discrets s'analysent souvent plus facilement avec la transformée en z qui étend les séries de Fourier à tout le plan complexe. Pour simplifier les notations, nous supposons que l'intervalle d'échantillonnage est $T = 1$ et les échantillons d'un signal discret sont notés $f[n]$.

3.2.1 Convolutions discrètes

Dans le cas discret, l'homogénéité temporelle se limite à des translations sur la grille d'échantillonnage. Un opérateur linéaire discret L est homogène dans le temps si et seulement si pour tout $f[n]$ et tout décalage $f_p[n] = f[n-p]$ avec $p \in \mathbb{Z}$

$$L f_p[n] = Lf[n-p].$$

Réponse impulsionnelle On note $\delta[n]$ le Dirac discret

$$\delta[n] = \begin{cases} 1 & \text{si } n = 0 \\ 0 & \text{si } n \neq 0 \end{cases}. \tag{3.13}$$

Tout signal $f[n]$ peut être décomposé comme somme de Diracs translatés

$$f[n] = \sum_{p=-\infty}^{+\infty} f[p]\delta[n-p].$$

Soit $L\delta[n] = h[n]$ la réponse impulsionnelle de cet opérateur. La linéarité et l'invariance temporelle impliquent

$$Lf[n] = \sum_{p=-\infty}^{+\infty} f[p]h[n-p] = f \star h[n].$$

Un opérateur linéaire homogène est donc un produit de convolution discret.

Stabilité et causalité Un filtre discret L est *causal* si et seulement si $Lf[p]$ ne dépend que des valeurs de $f[n]$ pour $n \leq p$. Cela implique donc que $h[n] = 0$ si $n < 0$. La réponse impulsionnelle $h[n]$ est causale. On représente souvent un signal causal grâce à la fonction de Heavyside discrète

$$\gamma[n] = \begin{cases} 1 & \text{si } n \geq 0 \\ 0 & \text{si } n < 0 \end{cases} \tag{3.14}$$

car $h[n] = h[n]\gamma[n]$.

Pour qu'un signal d'entrée borné produise un signal de sortie borné il suffit que

$$\sum_{n=-\infty}^{+\infty} |h[n]| < +\infty, \tag{3.15}$$

car

$$|Lf[n]| \leq \sup_{n\in\mathbb{Z}} |f[n]| \sum_{k=-\infty}^{+\infty} |h[k]|.$$

On peut vérifier (exercice) que cette condition suffisante est aussi nécessaire. On dit alors que le filtre et la réponse impulsionnelle sont *stables*.

Fonction de transfert Comme dans le cas continu, les vecteurs propres de ces opérateurs de convolutions sont des exponentielles complexes $e_\omega[k] = e^{i\omega k}$,

$$Le_\omega[n] = \sum_{k=-\infty}^{+\infty} e^{i\omega(n-k)} h[k] = e^{i\omega n} \sum_{k=-\infty}^{+\infty} h[k] e^{-i\omega k}. \tag{3.16}$$

Les valeurs propres correspondantes sont donc obtenues par la série de Fourier

$$\hat{h}(e^{i\omega}) = \sum_{k=-\infty}^{+\infty} h[k] e^{-i\omega k}, \tag{3.17}$$

que l'on appelle fonction de transfert du filtre.

3.2.2 Séries de Fourier

La transformée de Fourier d'un signal discret $f[n]$ est définie par

$$\hat{f}(e^{i\omega}) = \sum_{k=-\infty}^{+\infty} f[k] e^{-i\omega k}. \tag{3.18}$$

C'est la transformée de Fourier de sa représentation par somme de Dirac

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

Toutes les propriétés de la transformée de Fourier (2.15-2.28) restent donc valables si $f_d(t)$ est une distribution tempérée, ce qui est le cas si $|f[n]|$ est borné.

On peut aussi démontrer [3] que la famille $\{\mathrm{e}^{ik\omega}\}_{k\in\mathbb{Z}}$ est une base orthonormale de $\mathbf{L^2}[-\pi,\pi]$ muni du produit scalaire

$$< a(\omega), b(\omega) >= \frac{1}{2\pi} \int_{-\pi}^{\pi} a(\omega) b^*(\omega) d\omega.$$

Si $f[n] \in \mathbf{l^2}(\mathbb{Z})$, la série (3.18) peut alors s'interpréter comme la décomposition de $\hat{f}(\mathrm{e}^{i\omega}) \in \mathbf{L^2}[0, 2\pi]$ dans cette base orthonormale. Les coefficients de décomposition sont obtenus par produit scalaire

$$f[n] =< \hat{f}(\mathrm{e}^{i\omega}), e^{-i\omega n} >= \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(e^{i\omega}) e^{i\omega n} d\omega \ , \tag{3.19}$$

et l'on obtient des formules de Parseval

$$\sum_{n=-\infty}^{+\infty} f[n]\, g^*[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{f}(e^{i\omega})\, \hat{g}(\mathrm{e}^{i\omega})\, d\omega \tag{3.20}$$

et de Plancherel

$$\sum_{n=-\infty}^{+\infty} |f[n]|^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\hat{f}(e^{i\omega})|^2 d\omega. \tag{3.21}$$

Filtrage discret Les exponentielles complexes étant les vecteurs propres des opérateurs de convolution discrète, il en résulte le théorème suivant.

Théorème 3.2 (Convolution discrète) *Soient $f[n]$ et $h[n]$ deux signaux dans $\mathbf{l^2}(\mathbb{Z})$. La transformée de Fourier de $g[n] = f \star h[n]$ est*

$$\hat{g}(e^{i\omega}) = \hat{f}(e^{i\omega})\hat{h}(e^{i\omega}). \tag{3.22}$$

La démonstration est formellement identique à la démonstration du théorème 2.1 si on remplace les intégrales par des sommes discrètes et que l'on suppose que $f[n]$ et $h[n]$ sont dans $\mathbf{l^1}(\mathbb{Z})$. Le même résultat dans $\mathbf{l^2}(\mathbb{Z})$ s'obtient par un argument de densité.

La formule de reconstruction (3.19) montre qu'un signal filtré s'écrit

$$f \star h[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{h}(\mathrm{e}^{i\omega})\hat{f}(e^{i\omega})e^{i\omega n}d\omega. \tag{3.23}$$

La fonction de transfert $\hat{h}(\mathrm{e}^{i\omega})$ amplifie ou atténue les composantes fréquentielles $\hat{f}(\mathrm{e}^{i\omega})$ de $f[n]$ dans l'intervalle de fréquence $[-\pi, \pi]$.

On vérifie de même qu'une multiplication temporelle est équivalente à une convolution dans le domaine fréquentiel. Si $g[n] = f[n]w[n]$ alors

$$\hat{g}(\mathrm{e}^{i\omega}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{f}(e^{iu})\hat{w}(e^{i(\omega-u)})du.$$

Exemple
La moyenne discrète uniforme définie par

$$Lf[n] = \frac{1}{2N+1}\sum_{p=-N}^{+N} f[n-p],$$

est un filtre dont la réponse impulsionnelle est

$$h[n] = \begin{cases} \frac{1}{2N+1} & \text{si } -N \leq n \leq N \\ 0 & \text{si } |n| > N \end{cases} \tag{3.24}$$

La fonction de transfert est la série de Fourier

$$\hat{h}(\mathrm{e}^{i\omega}) = \frac{1}{2N+1}\sum_{n=-N}^{+N} e^{-in\omega} = \frac{1}{2N+1}\frac{\sin(N+\frac{1}{2})\omega}{\sin\omega/2}.$$

Ce filtre atténue surtout les fréquences au-delà de $2\pi/(2N+1)$.

3.2.3 Sélection fréquentielle idéale

La fonction de transfert d'un filtre discret étant 2π périodique, elle est spécifiée sur l'intervalle $[-\pi, \pi]$. La fonction de transfert du filtre discret passe-bas idéal est définie pour $\omega \in [-\pi, \pi]$ par

$$\hat{h}_0(e^{i\omega}) = \begin{cases} 1 & \text{si } |\omega| \leq \omega_c \\ 0 & \text{si } |\omega| > \omega_c \end{cases} \tag{3.25}$$

Sa réponse impulsionnelle calculée grâce à l'intégrale (3.19) est

$$h_0[n] = \frac{\sin \omega_c n}{\pi n}. \tag{3.26}$$

C'est l'échantillonnage uniforme de la fonction de transfert d'un filtre analogique passe-bas idéal.

La fonction transfert d'un filtre passe-bande discret idéal est

$$\hat{h}_1(\mathrm{e}^{i\omega}) = \begin{cases} 1 & \text{si } |\omega| \in [\omega_0 - \omega_c, \omega_0 + \omega_c] \\ 0 & \text{ailleurs} \end{cases} \tag{3.27}$$

Comme $\hat{h}_1(\mathrm{e}^{i\omega}) = \hat{h}_0(e^{i(\omega-\omega_0)}) + \hat{h}_0(e^{i(\omega+\omega_0)})$, on peut en déduire que sa réponse impulsionnelle est

$$h_1[n] = 2\cos(\omega_0 n)\ h_0[n].$$

La convolution discrète d'un signal $f[n]$ avec un filtre passe-bas ou passe-bande idéal ne peut se calculer exactement avec un nombre fini d'opérations. Il est donc nécessaire d'approximer ces filtres par des opérateurs de convolutions qui se calculent en temps fini.

3.3 Synthèse de filtres discrets

Lors de la synthèse de filtres discrets, tout comme dans le cas analogique, on impose des conditions d'atténuation sur le gain du filtre $|\hat{h}(\mathrm{e}^{i\omega})|$. Le problème est d'obtenir des filtres tels que $|\hat{h}(\mathrm{e}^{i\omega})|$ satisfasse aux conditions d'atténuation et dont la structure permette de calculer les convolutions discrètes avec le moins d'opérations possibles.

3.3.1 Filtres récursifs

Pour effectuer des calculs numériques, on utilise une classe de filtres pour lesquels la convolution discrète se calcule avec un nombre fini d'opérations par échantillon. La sortie $g[n] = Lf[n]$ est reliée à $f[n]$ par une équation de différences

$$\sum_{k=0}^{N} a_k g[n-k] = \sum_{k=0}^{M} b_k f[n-k], \tag{3.28}$$

où a_k et b_k sont des réels et $a_0 \neq 0$. Donc

$$g[n] = \frac{1}{a_0}\left(\sum_{k=0}^{M} b_k f[n-k] - \sum_{k=1}^{N} a_k g[n-k]\right)$$

se calcule à partir de son passé et de $f[n]$ avec $N + M$ multiplications et additions.

Etant donné un signal causal $f[n]$, le calcul de $g[n]$ nécessite la connaissance de "conditions initiales", par exemple N valeurs consécutives de $g[n]$. Si l'on impose que $g[n] = 0$ pour $-N \leq n < 0$, alors $g[n]$ est entièrement caractérisé pour tout $n \in \mathbb{Z}$. L'opérateur L est alors un filtre linéaire homogène causal.

Si $N = 0$ alors le fitre a une réponse impulsionnelle $h[n]$ finie de taille M

$$g[n] = \sum_{k=0}^{M} \frac{b_k}{a_0} f[n-k] = h \star f[n].$$

Si $M = 0$, on dit que le filtre est autorégressif

$$g[n] = \frac{b_0}{a_0} f[n] - \sum_{k=1}^{N} \frac{a_k}{a_p} g[n-k].$$

Fonction de transfert Pour caractériser la classe des opérateurs de convolutions L qui satisfont (3.28), nous évaluons la condition imposée sur la fonction de transfert en calculant la transformée de Fourier de chaque côté de l'égalité (3.28). Si $\hat{f}(\mathrm{e}^{i\omega})$ est la transformée de Fourier de $f[n]$ alors la transformée de Fourier de $f[n-k]$ est $e^{-ik\omega}\hat{f}(\mathrm{e}^{i\omega})$. La transformée de Fourier de (3.28) est donc

$$\sum_{k=0}^{N} a_k\, e^{-ik\omega} \hat{g}(\mathrm{e}^{i\omega}) = \sum_{k=0}^{M} b_k\, e^{-ik\omega} \hat{f}(\mathrm{e}^{i\omega}),$$

d'où l'on déduit que

$$\hat{h}(\mathrm{e}^{i\omega}) = \frac{\hat{g}(\mathrm{e}^{i\omega})}{\hat{f}(\mathrm{e}^{i\omega})} = \frac{\sum_{k=0}^{M} b_k\, e^{-ik\omega}}{\sum_{k=0}^{N} a_k\, e^{-ik\omega}}. \tag{3.29}$$

La fonction de transfert d'un filtre récursif est donc un rapport de polynômes en $e^{-i\omega}$.

Les propriétés du module et de la phase s'analysent plus facilement en calulant les pôles d_k et les zéros c_k de la fonction rationnelle (3.29)

$$\hat{h}(\mathrm{e}^{i\omega}) = \frac{b_0}{a_0} \frac{\prod_{k=1}^{M}(1 - c_k e^{-i\omega})}{\prod_{k=1}^{N}(1 - d_k e^{-i\omega})}.$$

Le module de la transformée de Fourier est donc

$$|\hat{h}(\mathrm{e}^{i\omega})| = \frac{|b_0|}{|a_0|} \frac{\prod_{k=1}^{M}|1 - c_k e^{-i\omega}|}{\prod_{k=1}^{N}|1 - d_k e^{-i\omega}|}.$$

L'amplitude de la fonction de transfert est le plus souvent calculée en décibels (db) qui mesurent

$$20 \log_{10} |\hat{h}(\mathrm{e}^{i\omega})| = 10 \log_{10} \frac{|b_0|^2}{|a_0|^2} + \sum_{k=1}^{M} 10 \log_{10} |1 - c_k e^{-i\omega}|^2 - \sum_{k=1}^{N} 10 \log_{10} |1 - d_k e^{-i\omega}|^2.$$

Les pôles et les zéros ne se distinguent donc que par un changement de signe. La phase complexe de $\hat{h}(\mathrm{e}^{i\omega})$ se mesure de même par

$$\arg\hat{h}(\mathrm{e}^{i\omega}) = \arg\frac{b_0}{a_0} + \sum_{k=1}^{M} \arg(1 - c_k e^{-i\omega}) - \sum_{k=1}^{N} \arg(1 - d_k e^{-i\omega}).$$

Exemple Prenons le cas d'un pôle ou d'un zéro situé en $re^{i\theta}$ et étudions le module et la phase de $(1 - re^{i\theta}e^{-i\omega})$.

$$10\log_{10}|1 - re^{i\theta}e^{-i\omega}|^2 = 10\log_{10}[1 + r^2 - 2r\cos(\omega - \theta)].$$

Le module est donc minimum pour $\omega = \theta$ où il vaut $20\log_{10}|1 - r|$ et maximum en $\omega = \theta + \pi$ où il vaut $20\log_{10}|1 + r|$. Suivant que ce facteur est un pôle ou un zéro, il produit une atténuation ou une amplification au voisinage de $\omega = \theta$. La phase complexe est

$$\arg\hat{h}(\mathrm{e}^{i\omega}) = \arctan\left[\frac{r\sin(\omega - \theta)}{1 - r\cos(\omega - \theta)}\right].$$

3.3.2 Transformée en z

Pour étudier plus facilement les propriétés des fonctions de transfert des filtres discrets, et en particulier des filtres récursifs, on introduit la transformée en z qui étend la série de Fourier

$$\hat{h}(\mathrm{e}^{i\omega}) = \sum_{n=-\infty}^{+\infty} h[n]\mathrm{e}^{-in\omega} \tag{3.30}$$

à tout le plan complexe $z \in \mathbb{C}$, avec la série de Laurent

$$\hat{h}(z) = \sum_{n=-\infty}^{+\infty} h[n]z^{-n}. \tag{3.31}$$

Anneau de convergence On dit que la série de Laurent $\hat{h}(z)$ est convergente si

$$\sum_{n=-\infty}^{+\infty} |h[n]|\,|z|^{-n} < +\infty.$$

Le domaine de convergence ne dépend que de $|z|$ et est donc isotrope. La proposition suivante montre que le domaine de convergence est un anneau dans le plan complexe.

Proposition 3.1 *Il existe ρ_1 et ρ_2 tels que $\hat{h}(z)$ est convergente pour $\rho_1 < |z| < \rho_2$ et divergente pour $|z| < \rho_1$ ou $|z| > \rho_2$. On note $A(\hat{h})$ l'intervale de $|z|$ sur lequel $\hat{h}(z)$ est convergente.*

La démonstration est laissée en exercice. Dans le cas où la transformée en z est convergente pour $|z| = 1$, la transformée de Fourier est égale à la restriction de la transformée en z au cercle unité du plan complexe.

Stabilité et causalité Le domaine de convergence (absolu) de la transformée en z dépend des propriétés de causalité et de stabilité du filtre. Le filtre est causal si $h[n] = 0$ pour $n < 0$ d'où l'on déduit que si $\hat{h}(z)$ converge pour $|z| = \rho$ alors il converge pour $|z| \geq \rho$. L'anneau de convergence s'étend donc à l'infini ($\rho_2 = +\infty$).

Le filtre est stable si et seulement si

$$\sum_{n=0}^{+\infty} |h[n]| < +\infty.$$

Cela signifie que l'anneau de convergence contient $|z| = 1$. Si le filtre est causal et stable, on déduit donc que $\hat{h}(z)$ est convergente pour $|z| \geq 1$.

Inverse La transformée en z peut s'inverser mais le calcul de $h[k]$ à partir de $\hat{h}(z)$ dépend du domain de convergence choisi. La formule générale d'inversion se fait par une intégrale de Cauchy qui calcule $h[k]$ en integrant $\hat{h}(z)$ le long d'un contour inclu dans l'anneau de convergence. Dans le cas où l'anneau de convergence inclu le cercle unité, cette intégrale peut se faire le long du cercle unité, auquel cas on obtient la transformée de Fourier inverse

$$h[k] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{h}(\mathrm{e}^{i\omega}) e^{ik\omega} d\omega.$$

Pour montrer que $h[k]$ ne dépend pas seulement de $\hat{h}(z)$ mais aussi du domaine de convergence choisi, prenons par exemple

$$\hat{h}(z) = \frac{1}{1 - az^{-1}}.$$

La réponse impulsionnelle correspondant à la région de convergence à l'extérieur du cercle $|z| = a$ est causale et se calcule par un dévelopement en série de $\frac{1}{1-x}$

$$\hat{h}(z) = \sum_{n=0}^{+\infty} a^n z^{-n},$$

d'où $h[n] = a^n \gamma[n]$. Pour que la région de convergence soit $|z| < a$ on réécrit

$$\hat{h}(z) = \frac{-a^{-1}z}{1 - a^{-1}z}.$$

En utilisant la décomposition en série de $\frac{1}{1-x}$ on obtient une réponse impulsionnelle anti-causale

$$h[n] = \begin{cases} -a^n & \text{si } n \leq -1 \\ 0 & \text{si } n \geq 0 \end{cases}$$

Exemples On utilise généralement un filtre causal, ce qui impose que l'anneau de convergence s'étende à l'infini.
• Si $h[n] = \delta[n-k]$ alors

$$\hat{h}(z) = z^{-k} \tag{3.32}$$

et $A(\hat{h}) =]0, +\infty[$.
• Si $h[n] = a^n \gamma[n]$ alors

$$\hat{h}(z) = \frac{1}{1 - az^{-1}} \tag{3.33}$$

$A(\hat{h}) =]|a|, +\infty[$.
• Si $h[n] = na^n \gamma[n]$ alors

$$\hat{h}(z) = \frac{az^{-1}}{1 - az^{-1}}$$

$A(\hat{h}) =]|a|, +\infty[$.

Convolution Toutes les propriétés de la transformée de Fourier s'étendent directement à la transformée en z. En particulier, si $g[n] = f \star h[n]$ alors la transformée en z de $g[n]$ est le produit

$$\hat{g}(z) = \hat{f}(z)\hat{h}(z)$$

et son anneau de convergence est

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

Filtres récursifs Nous avons vu en (3.29) que la fonction de transfert d'un filtre récursif est une fonction rationnelle. Sa tranformée en z peut donc s'écrire

$$\hat{h}(z) = \frac{\sum_{k=0}^{M} b_k z^{-k}}{\sum_{k=0}^{N} a_k z^{-k}}. \tag{3.34}$$

La réponse impulsionnelle causale $h[n]$ se calcule facilement en décomposant $\hat{h}(z)$ en éléments simples. Si $\hat{h}(z)$ a des pôles simples situés en d_k, on peut montrer par identification des coefficients (exercice) qu'il peut s'écrire sous la forme

$$\hat{h}(z) = \sum_{r=0}^{M-N} B_r z^{-r} + \sum_{k=0}^{N} \frac{A_k}{1 - d_k z^{-1}}.$$

Le filtre causal correspondant à une réponse impulsionelle qui se calcule avec (3.32) et (3.33)

$$h[n] = \sum_{r=0}^{M-N} B_r \delta[n-r] + \sum_{k=0}^{N} A_k (d_k)^n \gamma[n].$$

Dans le cas de pôles multiples, la décomposition fractionnelle s'étend avec des puissances aux dénominateurs des fractions. On distingue les filtres à réponse impulsionnelle finie

dont la transformée en z est un polynôme en z^{-1} (N=0) et les filtres à réponse impulsionnelle infinie pour lesquels $N > 0$.

On observe que la réponse impulsionelle $h[n]$ est causale et stable si et seulement si pour tout k, $|d_k| < 1$. Cela signifie donc que tous les pôles de $\hat{h}(z)$ ont un module plus petit que 1.

3.3.3 Approximation de filtres sélectifs en fréquence

Tout comme pour la synthèse de filtres analogiques, on approxime un filtre passe-bas idéal (3.25) par un filtre récursif dont la fonction de transfert satisfait les conditions imposés par un gabarit qui limite les oscillations dans la bande passante et la bande d'atténuation (voir figure 2.2). La technique de synthèse la plus courante est de transformer un filtre passe-bas analogique rationnel

$$\hat{h}_a(\omega) = \frac{N(i\omega)}{D(i\omega)}$$

en un filtre discret récursif par un changement de variable

$$i\omega = F(\mathrm{e}^{i\omega}),$$

où F est une fonction rationnelle de $\mathrm{e}^{i\omega}$ qui envoie $]-\pi, \pi[$ sur $]-\infty, +\infty[$. La fonction de transfert

$$\hat{h}_d(\mathrm{e}^{i\omega}) = \frac{N(F(\mathrm{e}^{i\omega}))}{D(F(\mathrm{e}^{i\omega}))}$$

est une fonction rationnelle de $\mathrm{e}^{i\omega}$ et donc la fonction de transfert d'un filtre discret récursif. Le changement de variable $F(\mathrm{e}^{i\omega})$ qui associe $]-\pi, \pi[$ à l'axe réel $\mathbb{R}$ doit être aussi "régulier" que possible pour ne pas trop modifier les propriétés de la fonction de transfert $\hat{h}(\omega)$. On utilise souvent l'application

$$F(\mathrm{e}^{i\omega}) = \frac{2}{T}\tan(\frac{\omega}{2}) = \frac{2}{T}\frac{1-e^{-i\omega}}{1+e^{-i\omega}}.$$

Le facteur T est un paramètre de dilatation qui peut être ajusté arbitrairement.

Par exemple, on peut construire un filtre passe-bas dont la fréquence de coupure est en ω_c à partir d'un filtre analogique de Butterworth (2.50). On obtient

$$|\hat{h}(\mathrm{e}^{i\omega})|^2 = \frac{1}{1+\left(\frac{\tan(\omega/2)}{\tan(\omega_c/2)}\right)^{2N}}.$$

L'ordre N du filtre doit être adapté aux conditions imposées par le gabarit du filtre passe-bas.

3.3.4 Factorisation spectrale

Lors de la synthèse d'un filtre récursif, une fois que l'on a calculé $|\hat{h}(\mathrm{e}^{i\omega})|^2$ pour satisfaire les conditions d'amplification ou d'atténuation, il reste à adapter la phase pour que $\hat{h}(\mathrm{e}^{i\omega})$ soit un filtre causal et stable. Le module est donné par

$$|\hat{h}(\mathrm{e}^{i\omega})|^2 = \hat{h}(\mathrm{e}^{i\omega})\hat{h}^*(\mathrm{e}^{i\omega}) = \hat{h}(z)\hat{h}^*(1/z^*),$$

avec $z = e^{i\omega}$. Pour un filtre récursif,

$$\hat{h}(z) = \frac{b_0}{a_0} \frac{\prod_{k=1}^{M}(1 - c_k z^{-1})}{\prod_{k=1}^{N}(1 - d_k z^{-1})}.$$

et

$$C(z) = \hat{h}(z)\hat{h}^*(1/z^*) = \frac{|b_0|^2}{|a_0|^2} \frac{\prod_{k=1}^{M}(1 - c_k z^{-1})(1 - c_k^* z)}{\prod_{k=1}^{N}(1 - d_k z^{-1})(1 - d_k^* z)}.$$

La donnée de $|\hat{h}(e^{i\omega})|^2$ impose la position des zéros et des pôles de $C(z)$. Les zéros et les pôles de $C(z)$ vont par paires $(c_k, 1/c_k^*)$ et $(d_k, 1/d_k^*)$. Pour chaque paire, il y a un élément dans le cercle unité et l'autre à l'extérieur, à moins qu'ils ne soient confondus sur le cercle unité. On peut construire $\hat{h}(z)$ en choisissant arbitrairement un pôle et un zéro dans chaque paire. Pour que $\hat{h}(z)$ soit la transformée en z d'un système stable et causal, nous avons vu que tous les pôles doivent être strictement dans le cercle unité. Cela laisse libre le choix des zéros. Un choix particulier des zéros est de les prendre tous dans le cercle unité. Un filtre dont les zéros et les pôles sont dans le cercle unité est appelé filtre à phase minimale.

Filtre inverse Le filtre inverse d'un filtre h est le filtre h_i tel que pour tout $f[n]$

$$f \star h \star h_i[n] = f[n].$$

Cela signifie que les zones de convergence de $\hat{h}(z)$ et de $\hat{h}_i(z)$ s'intersectent et que sur ce domaine

$$\hat{h}(z)\hat{h}_i(z) = 1.$$

Le filtre inverse d'un filtre à phase minimale est stable et causal. En effet, les pôles de $\hat{h}_i(z)$ sont les zéros de $\hat{h}(z)$ et inversement. Or, pour que $\hat{h}_i(z)$ soit stable et causal, il faut et il suffit que ses pôles soient dans le cercle unité et donc que les zéros de $\hat{h}(z)$ soient dans le cercle unité.

3.4 Signaux finis

Nous avons supposé jusqu'à présent que nos signaux discrets $f[n]$ sont définis pour tout $n \in \mathbb{Z}$. Le plus souvent, $f[n]$ est connu sur un domaine fini, disons $0 \leq n < N$. Il faut donc adapter les calculs de convolutions en tenant compte des effets de bord en $n = 0$ et $n = N - 1$. Par ailleurs, pour utiliser la transformée de Fourier comme outil de calcul numérique, il faut pouvoir la redéfinir sur des signaux discrets finis. Ces deux problèmes sont résolus en périodisant les signaux finis. L'algorithme de transformée de Fourier rapide est décrit avec une application au calcul rapide des convolutions.

3.4.1 Convolutions circulaires

Soient $\tilde{f}[n]$ et $\tilde{h}[n]$ des signaux de N échantillons. Pour calculer la convolution

$$\tilde{f} \star \tilde{h}[n] = \sum_{p=-\infty}^{+\infty} \tilde{f}[p]\tilde{h}[n-p]$$

pour $0 \leq n < N$, il nous faut connaître $\tilde{f}[n]$ et $\tilde{h}[n]$ au-delà de $0 \leq n < N$. Une approche possible est d'étendre $\tilde{f}[n]$ et $\tilde{h}[n]$ avec une périodisation sur N échantillons

$$f[n] = \tilde{f}[n \text{ modulo } N] \quad , \quad h[n] = \tilde{h}[n \text{ modulo } N].$$

La convolution circulaire de ces deux signaux de période N est réduite à une somme sur leur période

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

Les vecteurs propres d'un opérateur de convolution circulaire

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

sont les exponentielles discrètes $e_k[n] = e^{\frac{i2\pi k}{N}n}$. En effet,

$$Le_k[n] = e^{\frac{i2\pi k}{N}n} \sum_{p=0}^{N-1} h[p]e^{\frac{-i2\pi k}{N}p},$$

et les valeurs propres sont données par la transformée de Fourier discrète de $h[n]$

$$\hat{h}[k] = \sum_{p=0}^{N-1} h[p]e^{\frac{-i2\pi k}{N}p}.$$

3.4.2 Transformée de Fourier discrète

L'espace des signaux discrets de période N est de dimension N et l'on note le produit scalaire

$$< f, g >= \sum_{n=0}^{N-1} f[n]g^*[n]. \tag{3.35}$$

Le théorème suivant démontre que tout signal de période N peut s'écrire comme une transformée de Fourier discrète.

Théorème 3.3 *La famille d'exponentielles discrètes* $(e_k[n])_{0 \leq k < N}$

$$e_k[n] = e^{\frac{i2\pi k}{N}n}, \tag{3.36}$$

est une base orthogonale de l'espace des signaux de période N.

Pour prouver ce théorème, il suffit de montrer que cette famille de N vecteurs est orthogonale (exercice). Comme l'espace est de dimension N, c'est donc une base de l'espace. Tout signal $f[n]$ de période N peut se décomposer dans cette base

$$f[n] = \sum_{k=0}^{N-1} \frac{< f, e_k >}{\|e_k\|^2} e_k[n]. \tag{3.37}$$

La transformée de Fourier discrète de $f[n]$ est

$$\hat{f}[k] =< f, e_k >= \sum_{n=0}^{N-1} f[n] e^{\frac{-i2\pi n}{N}k}. \tag{3.38}$$

Comme $\|e_k[n]\|^2 = N$,

$$f[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}[k] e^{\frac{i2\pi k}{N}n}. \tag{3.39}$$

L'orthogonalité implique une formule de Plancherel

$$\sum_{n=0}^{N-1} |f[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |\hat{f}[k]|^2. \tag{3.40}$$

Effets de bord La transformée de Fourier discrète d'un signal de période N se calcule à partir des valeurs de $f[n]$ pour $0 \leq n < N$. Pourquoi se soucier du fait que ce soit un signal de période N plutôt qu'un signal de N échantillons ? La somme de Fourier (3.38) définit un signal de période N pour lequel l'échantillon $f[0]$ étant le même que $f[N]$ se retrouve placé à côté de $f[N-1]$. Si $f[0]$ et $f[N-1]$ sont très différents, cela induit une transition brutale dans le signal périodisé qui se traduit par l'apparition de coefficients de Fourier de relativement grande amplitude aux hautes fréquences. Par exemple, le signal apparemment régulier $f[n] = n$ pour $0 \leq n < N$ a une transition brutale en $n = pN$ pour $p \in \mathbb{Z}$, une fois périodisé. Cette transition apparaît dans sa série de Fourier.

Filtrage fini Comme $e^{\frac{i2\pi k}{N}n}$ sont les vecteurs propres des opérateurs de convolution circulaire, on déduit un théorème de convolution.

Théorème 3.4 (Convolution Circulaire) *La convolution circulaire $g[n] = f \circledast h[n]$ est un signal de période N dont la transformée de Fourier discrète est*

$$\hat{g}[k] = \hat{f}[k]\hat{h}[k] \tag{3.41}$$

La démonstration de ce théorème est identique à la démonstration des deux théorèmes de convolution précédents et laissée en exercice. Ce théorème montre que la convolution circulaire est un filtrage fréquentiel. Il ouvre aussi la porte au calcul rapide de convolutions en utilisant la transformée de Fourier rapide.

3.4.3 Transformée de Fourier rapide

Pour un signal $f[n]$ de N points, le calcul direct de la transformée de Fourier discrète

$$\hat{f}[k] = \sum_{n=0}^{N-1} f[n] e^{\frac{-i2\pi k}{N}n}, \tag{3.42}$$

pour $0 \leq k < N$, demande $O(N^2)$ multiplications et additions. Il est cependant possible de réduire le nombre d'opérations à $O(N \log_2 N)$ en réorganisant les calculs. Lorsque k est pair, on regroupe les termes n et $n + \frac{N}{2}$

$$\hat{f}[2k] = \sum_{n=0}^{\frac{N}{2}-1} (f[n] + f[n + N/2]) e^{\frac{-i2\pi k}{\frac{N}{2}} n}. \tag{3.43}$$

Lorsque k est impair, le même regroupement devient

$$\hat{f}[2k+1] = \sum_{n=0}^{\frac{N}{2}-1} e^{\frac{-i2\pi}{N} n} (f[n] - f[n + \frac{N}{2}]) e^{\frac{-i2\pi k}{\frac{N}{2}} n}. \tag{3.44}$$

L'équation (3.43) montre que les fréquences paires sont obtenues en calculant la transformée de Fourier discrète du signal de période $\frac{N}{2}$

$$f_p[n] = f[n] + f[n + \frac{N}{2}],$$

tandis que (3.44) permet de calculer les fréquences impaires par la transformée de Fourier discrète du signal de période $\frac{N}{2}$

$$f_i[n] = e^{\frac{-i2\pi}{N} n} (f[n] - f[n + \frac{N}{2}]).$$

Complexité Une transformée de Fourier d'un signal de taille N s'obtient donc en calculant deux transformées de Fourier de signaux de taille $\frac{N}{2}$. Le signal $f[n]$ étant complexe, le calcul des signaux $f_p[n]$ et $f_i[n]$ demande N additions complexes et $\frac{N}{2}$ multiplications complexes, ce qui fait $3N$ additions et $2N$ multiplications réelles. Soit $C(N)$ le nombre d'opérations d'une transformée de Fourier rapide d'un signal de période N. On a donc

$$C(N) = 2C(\frac{N}{2}) + KN, \tag{3.45}$$

avec $K = 5$. La transformée de Fourier d'un signal d'un seul point étant égale à lui-même, $C(1) = 0$. Avec le changement de variable $l = \log_2 N$ et de fonction $T(l) = C(N)/N$, on déduit de (3.45) que

$$T(l) = T(l-1) + K.$$

Comme $T(0) = 0$, $T(l) = Kl$ et donc

$$C(N) = KN \log_2(N).$$

Il existe différents algorithmes de transformée de Fourier rapide qui décomposent une transformée de Fourier de taille N en deux transformées de Fourier de taille $\frac{N}{2}$ plus $O(N)$ opérations. L'algorithme le plus performant à ce jour est un peu plus compliqué que celui que nous venons de décrire, mais ne demande que $N \log_2 N$ multiplications et $3N \log_2 N$ additions.

La transformée de Fourier inverse se calcule à partir de l'algorithme rapide en observant que

$$f^*[n] = \frac{1}{N} \sum_{k=0}^{N-1} \hat{f}^*[k] e^{\frac{-i2\pi kn}{N}}. \tag{3.46}$$

La tranformée de Fourier discrète inverse est donc obtenue en calculant la transformée de Fourier directe du complexe conjugué du signal, et en calculant le complexe conjugué du résultat.

3.4.4 Convolution rapide

L'algorithme rapide de la transformée de Fourier discrète permet d'utiliser le théorème 3.4 pour calculer efficacement les convolutions discrètes de deux signaux à support fini. Soient $f[n]$ et $h[n]$ deux signaux ayant des échantillons non nuls pour $0 \leq n < M$. Le signal causal

$$g[n] = f \star h[n] = \sum_{k=-\infty}^{+\infty} f[k]h[n-k], \tag{3.47}$$

n'a de valeurs non-nulles que pour $0 \leq n < 2M$. Le calcul direct de ce produit de convolution, en évaluant la somme (3.47), demande $O(M^2)$ additions et multiplications. Le théorème de convolution circulaire 3.4 suggère un procédé plus rapide basé sur des convolutions circulaires.

Pour réduire le calcul de la convolution non-circulaire (3.47) à une convolution circulaire, on définit deux signaux de période $2M$

$$a[n] = \begin{cases} f[n] & \text{si } 0 \leq n < M \\ 0 & \text{si } M \leq n < 2M \end{cases} \tag{3.48}$$

$$b[n] = \begin{cases} h[n] & \text{si } 0 \leq n < M \\ 0 & \text{si } M \leq n < 2M \end{cases} \tag{3.49}$$

Il est maintenant facile de vérifier que la convolution circulaire

$$c[n] = a \circledast b[n]$$

satisfait

$$c[n] = g[n] \quad \text{pour } 0 \leq n < 2M. \tag{3.50}$$

On peut donc calculer les valeurs non-nulles de $g[n]$ en calculant les transformées de Fourier discrètes de $a[n]$ et $b[n]$, en les multipliant et en calculant la transformée de Fourier inverse du résultat. En utilisant l'algorithme de transformée de Fourier rapide, ce calcul demande un total de $O(M \log_2 M)$ additions et multiplications, au lieu de $O(M^2)$.

Cet algorithme rapide de calcul de convolutions est utilisé pour la convolution de signaux par des filtres de réponse impulsionnelle finie. Si le signal $f[n]$ a N points non-nuls et le filtre $h[n]$ a L échantillons non-nuls, l'algorithme de convolution peut être modifié pour calculer la convolution $h \circledast f[n]$ en $O(Nlog_2 L)$ opérations (exercice).

Chapitre 4

Traitement du signal aléatoire

Pour analyser les propriétés d'une classe de signaux, tels que des signaux de paroles en général ou le son "s" prononcé par différents locuteurs, on utilise une modélisation par des processus stochastiques qui reflète les propriétés communes de ces signaux. Cette modélisation permet de séparer le signal d'un bruit dont les caractéristiques stochastiques sont différentes, de coder efficacement les signaux d'une classe, ou même de les identifier. Nous ne considérons ici que des signaux réels stationnaires, dont les propriétés ne changent pas au cours du temps.

L'hypothèse de stationnarité nous ramène dans le champ de la transformée de Fourier, qui permet de définir la puissance spectrale d'un processus. Nous nous concentrons sur les propriétés du second ordre des processus stationnaires, dont la modélisation se réduit à la recherche d'un filtre paramétré. Nous étudions une application au débruitage de signaux par filtrage linéaire.

4.1 Processus stationnaires au sens large

Un processus discret à valeurs réelles est une suite de variables aléatoires $\{X[n]\}_{n\in\mathbb{Z}}$ définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathrm{P})$. Ce processus est caractérisé par la loi de probabilité

$$P(X[n_1] \in [a_1, b_1], ..., X[n_k] \in [a_k, b_k])$$

de tout sous-ensemble de k variables aléatoires, pour tout intervalle $[a_i, b_i]$ avec $1 \leq i \leq k$.

La réalisation d'un tel processus est un signal discret $x[n]$ qui donne les valeurs prises par chaque variable aléatoire $X[n]$ lors d'une observation. On modélise une classe de signaux par un processus aléatoire dont les réalisations correspondent à l'ensemble de tous les signaux de cette classe. Par exemple, des enregistrements de température en différents points de la terre peuvent être modélisés par un processus discret. Une réalisation de ce processus est l'ensemble des enregistrements de températures à un moment donné.

Stationnarité Le processus est strictement stationnaire si sa loi de probabilité ne change pas avec un décalage temporel. Pour tout $p \in \mathbb{Z}$ et tout sous-ensemble de k variables aléatoires, la loi de probabilité de $\{X[n_1], ..., X[n_k]\}$ est la même que celle de

$\{X[n_1+p], ..., X[n_k+p]\}$. Cela implique que la moyenne ne dépend pas de l'instant

$$\mu[n] = E\{X[n]\} = \mu, \tag{4.1}$$

et que l'autocovariance ne dépend que de la différence de position

$$\mathrm{Cov}(X[n], X[m]) = E\{(X[n]-\mu[n])(X[m]-\mu[m])\} = R_X[n-m]. \tag{4.2}$$

En traitement du signal, on observe le plus souvent une seule réalisation, à partir de laquelle on veut estimer certains paramètres du processus sous-jacent. La pauvreté de cette information numérique nous limite généralement à l'étude de la moyenne et de l'autocovariance. Puisque l'on n'étudie que les moments d'ordre 2 du processus, au lieu d'imposer que le processus soit strictement stationnaire, nous supposons simplement que la moyenne et la covariance satisfont aux propriétés de stationnarité (4.1) et (4.2). Si de plus la variance du processus est finie

$$E\{|X[n]|^2\} < +\infty, \tag{4.3}$$

on dit que le processus est stationnaire au sens large (SSL).

Exemple Un processus $X[n]$ est *gaussien* si et seulement si pour tout k et n_1, ..., n_k, le vecteur de k variables aléatoires $(X[n_1], ..., X[n_k])$ est gaussien. On rappelle [8] qu'un vecteur de p variables aléatoires $(X_1, ..., X_p)$ est gaussien si et seulement si la densité de probabilité peut s'écrire sous la forme

$$p(x_1, ..., x_p) = \frac{1}{(2\pi)^{p/2}|\mathbf{R}|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{m})^t \mathbf{R}^{-1}(\mathbf{x}-\mathbf{m})\right],$$

où $\mathbf{x} = (x_1, ..., x_p)$ est le vecteur de valeurs, $\mathbf{m} = (\mu_1, ..., \mu_p)$ est le vecteur de moyennes avec $\mu_i = E\{X_i\}$, $\mathbf{R} = (\mathrm{Cov}(X_n, X_m))_{1\le n\le p, 1\le m\le p}$ est la matrice de covariance des p variables aléatoires, $\mathbf{R}^{-1}$ son inverse et $|\mathbf{R}|$ son déterminant. Un processus gaussien est donc entièrement défini par sa moyenne et sa covariance. Si $X[n]$ est gaussien et stationnaire au sens large (SSL), on vérifie facilement qu'il est aussi stationnaire au sens strict.

En l'absence d'information sur les moments d'ordre supérieur à 2, on suppose souvent par défaut que le processus étudié est gaussien. Les processus gaussiens apparaissent dans de nombreux phénomènes physiques à cause du théorème de limite centrale [8] qui démontre que la somme de variables aléatoires indépendantes converge vers une variable aléatoire gaussienne.

4.1.1 Estimation de la moyenne et de l'autocovariance

Si le processus est stationnaire, on peut tenter d'estimer la moyenne μ et la fonction d'autocorrélation à partir de moyennes temporelles d'une réalisation.

Moyenne L'estimateur classique de la moyenne à partir de N échantillons d'une réalisation de $X[n]$ se calcul avec une *moyenne empirique* définie par

$$\tilde{\mu} = \frac{1}{N} \sum_{n=0}^{N-1} X[n].$$

Cet estimateur est *non biaisé* puisque

$$E\{\tilde{\mu}\} = \mu.$$

Lorsque N augmente il est nécessaire que la variance de $\tilde{\mu}$ décroisse vers zéro, de façon à améliorer l'estimation de μ. On dit alors que le processus est ergodique pour la moyenne. La proposition suivante donne une condition sur l'autocovariance pour que cette propriété soit satisfaite.

Proposition 4.1 *Le processus est ergodique pour la moyenne si et seulement si*

$$\lim_{N\to+\infty} E\{(\mu - \tilde{\mu})^2\} = \lim_{N\to+\infty} \frac{1}{N} \sum_{l=-N+1}^{N-1} (1 - \frac{|l|}{N}) R_X[l] = 0.$$

Démonstration La variance de $\tilde{\mu}$ est

$$\begin{aligned}
E\left\{(\tilde{\mu} - \mu)^2\right\} &= \frac{1}{N^2} E\left\{ \left(\sum_{n=0}^{N-1} (X[n] - \mu) \right)^2 \right\} && (4.4)\\
&= \frac{1}{N^2} E\left\{ \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} (X[n] - \mu)(X[m] - \mu) \right\} && (4.5)\\
&= \frac{1}{N^2} \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} R_X[n-m] && (4.6)\\
&= \frac{1}{N} \sum_{l=-N+1}^{N-1} (1 - \frac{|l|}{N}) R_X[l]. && (4.7)
\end{aligned}$$

□

La fonction d'autocovariance doit avoir une décroissance suffisamment rapide ce qui signifie que les corrélations longue-portée sont faibles. Le cas le plus favorable correspond à des variables $X[n]$ qui sont deux à deux décorrélées, si bien que $R_X[l] = R_X[0]\,\delta[l]$. On a alors

$$E\{(\mu - \tilde{\mu})^2\} = \frac{R_X[0]}{N}.$$

Autocovariance Pour tout k, l'autocovariance $R_X[k]$ de $X[n]$ peut aussi s'estimer avec une moyenne empirique sur les valeurs d'une réalisations:

$$\tilde{R}_X[k] = \frac{1}{N} \sum_{n=0}^{N-1-|k|} (X[n] - \tilde{\mu})\,(X[n+|k|] - \tilde{\mu}). \tag{4.8}$$

Si on remplace $\tilde{\mu}$ par la vraie moyenne μ, on s'aperçoit que cet estimateur est biaisé

$$E\{\tilde{R}_X[k]\} = \frac{N-|k|}{N} R_X[k].$$

Le biais

$$E\{\tilde{R}_X[k]\} - R_X[k] = \frac{k\, R_X[k]}{N}$$

est toujours petit si $R_X[k]$ décroît rapidement lorsque k augmente. Dans le cas d'un processus Gaussien, on peut montrer [Priestley] que si $R_X[k] \in \mathbf{l^2}(\mathbb{Z})$ alors la variance de $\tilde{R}_X[k]$ est en $O(\frac{1}{N})$. L'erreur quadratique moyenne de l'estimateur est égale à la somme du biais au carré et de la variance

$$E\{|\tilde{R}_X[k] - R_X[k]|^2\} = \left|E\{\tilde{R}_X[k]\} - R_X[k]\right|^2 + E\{|\tilde{R}_X[k] - E\{\tilde{R}_X[k]\}|^2\}.$$

Si $R_X[k] = O(\frac{1}{\sqrt{k}})$ on en déduit que $E\{|\tilde{R}_X[k] - R_X[k]|^2\} = O(\frac{1}{N})$.

On peut définir un estimateur non biaisé

$$\tilde{R}_X[k] = \frac{1}{N-|k|} \sum_{n=0}^{N-1-|k|} (X[n] - \tilde{\mu})\,(X[n+|k|] - \tilde{\mu}) \tag{4.9}$$

mais la variance de cet estimateur est en $O(\frac{1}{N-|k|})$. Cette variance est grande lorsque k est de l'ordre de N. Si $R_X[k] = O(\frac{1}{\sqrt{k}})$ et k est de l'ordre de N, l'erreur quadratique moyenne de l'estimateur biasé est plus faible que celle de l'estimateur non biaisé. On utilise donc plutôt l'estimateur biaisé (4.8) que l'estimateur (4.9).

4.1.2 Opérateur de covariance

Le traitement du signal aléatoire utilise le plus souvent des combinaisons linéaires de variables aléatoires correspondant aux valeurs d'un processus à différents instants. Nous montrons que l'autocovariance de $X[n]$

$$R_X[n,m] = \mathrm{Cov}(X[n], X[m]) = E\{(X[n] - E\{X[n]\})\,(X[m] - E\{X[m]\})\}$$

permet de facilement calculer la covariance de combinaisons linéaires des $X[n]$.

Soient

$$A = \sum_{n=-\infty}^{+\infty} a[n]X[n] \quad \text{et} \quad B = \sum_{n=-\infty}^{+\infty} b[n]X[n].$$

Les moyennes de ces variables aléatoires étant

$$E\{A\} = \sum_{n=-\infty}^{+\infty} a[n]E\{X[n]\} \quad \text{et} \quad E\{B\} = \sum_{n=-\infty}^{+\infty} b[n]E\{X[n]\}$$

on déduit

$$\mathrm{Cov}(A,B) = E\left\{ \sum_{n=-\infty}^{+\infty} a[n](X[n] - E\{X[n]\}) \sum_{m=-\infty}^{+\infty} b[m](X[m] - E\{X[m]\}) \right\}$$

et donc

$$\mathrm{Cov}(A,B) = \sum_{n=-\infty}^{+\infty} a[n] \sum_{m=-\infty}^{+\infty} b[m] R_X[n,m]. \tag{4.10}$$

Soit C l'opérateur symétrique de covariance défini par

$$Cb[n] = \sum_{m=-\infty}^{+\infty} b[m] R_X[n,m], \tag{4.11}$$

on peut réécrire (4.10) comme un produit scalaire dans $\mathbf{l^2}(\mathbb{Z})$

$$\mathrm{Cov}(A,B) = \quad < a, Cb > \quad .$$

L'opérateur de covariance est symétrique et positif car

$$< a, Ca >= \mathrm{Cov}(A,A) \geq 0.$$

4.1.3 Puissance spectrale

L'autocovariance de $X[n]$ est caractérisée par les vecteurs propres et valeurs propres de C. Si $X[n]$ est stationnaire au sens large

$$R_X[n,m] = R_X[n-m],$$

donc

$$Ca[n] = \sum_{m=-\infty}^{+\infty} a[m] R_X[n-m] = a \star R_X[n]$$

est un opérateur de convolution discret. Nous avons vu dans le paragraphe 3.2.1 que les vecteurs propres d'une convolution discrète sont les exponentielles $e_\omega[n] = \mathrm{e}^{-i\omega}$

$$Ce_\omega[n] = \hat{R}_X(e^{i\omega}) e_\omega[n].$$

Les valeurs propres associées sont positives et données par la série de Fourier

$$\hat{R}_X(\mathrm{e}^{i\omega}) = \sum_{n=-\infty}^{+\infty} R_X[n] \mathrm{e}^{-in\omega} \geq 0.$$

La fonction $\hat{R}_X(\mathrm{e}^{i\omega})$ est appelée *puissance spectrale* du processus car nous verrons que cela mesure l'énergie moyenne du processus par unité de fréquence. La puissance spectrale caractérise complètement l'autocovariance du processus que l'on retrouve par transformée de Fourier inverse (3.19). En particulier, la variance du processus est

$$\sigma^2 = R_X[0] = \frac{1}{2\pi}\int_0^{2\pi} \hat{R}_X(\mathrm{e}^{i\omega})d\omega. \tag{4.12}$$

Exemple On appelle *bruit blanc* tout processus SSL dont les valeurs à des instants différents sont décorrélés, ce qui signifie que

$$R_X[n] = \sigma^2\delta[n].$$

La puissance spectrale d'un bruit blanc est donc constante

$$\hat{R}_X(\mathrm{e}^{i\omega}) = \sigma^2.$$

4.1.4 Filtrage homogène

Une convolution discrète étant homogène dans le temps, on peut s'attendre à ce que cela ne modifie pas la stationnarité d'un processus. Le théorème suivant relie la puissance spectrale d'un processus filtré à la puissance spectrale originale.

Théorème 4.1 *Soient $h[n]$ et $g[n]$ les réponses impulsionnelles de deux filtres dans $\mathbf{l^2}(\mathbb{Z})$, dont les fonctions de transfert sont bornées. Soit $X[n]$ un processus SSL tel que $R_X[n] \in \mathbf{l^1}(\mathbb{Z})$. Les deux processus*

$$Y[k] = h \star X[k] = \sum_{n=-\infty}^{+\infty} h[k-n]X[n]$$

et

$$Z[l] = g \star X[l] = \sum_{m=-\infty}^{+\infty} g[l-m]X[m]$$

sont stationnaires au sens large et

$$\mathrm{Cov}(Y[k], Z[l]) = h \star \tilde{g} \star R_X[k-l], \tag{4.13}$$

avec $\tilde{g}[n] = g[-n]$. La puissance spectrale de $Y[n]$ est

$$\hat{R}_Y(\mathrm{e}^{i\omega}) = |\hat{h}(\mathrm{e}^{i\omega})|^2 \hat{R}_X(\mathrm{e}^{i\omega}). \tag{4.14}$$

Démonstration
En identifiant $Y[n]$ et $Z[m]$ respectivement à A et B dans (4.10) on obtient

$$\mathrm{Cov}(Y[k], Z[l]) = \sum_{n=-\infty}^{+\infty} h[k-n] \sum_{m=-\infty}^{+\infty} g[l-m]R_X[n-m].$$

Le changement de variables $(n, m) \to (k - n', l + m')$ donne

$$\mathrm{Cov}(Y[k], Z[l]) = \sum_{n=-\infty}^{+\infty} h[n'] \sum_{m=-\infty}^{+\infty} \tilde{g}[m'] R_X[k - l - n' - m]$$

qui est équivalent à (4.13).

Pour montrer que $Y[n]$ et $Z[n]$ sont stationnaires au sens large, on vérifie d'abord que leur moyennes sont constantes et finies. Concentrons nous sur $Y[n]$

$$E\{Y[k]\} = \sum_{n=-\infty}^{+\infty} h[k - n] E\{X[n]\} = E\{X[n]\}\hat{h}(1) < +\infty.$$

On calcule sa covariance en prenant $Z = Y$ dans (4.13)

$$\mathrm{Cov}(Y[k], Y[l]) = R_Y[k - l] = h \star \tilde{h} \star R_X[k - l], \tag{4.15}$$

ce qui démontre sa stationnarité au sense large. Comme la transformée de Fourier de $h \star \tilde{h}[n]$ est $|\hat{h}(\mathrm{e}^{i\omega})|^2$, la puissance spectrale (4.14) de Y se déduit de (4.15) par le théorème de convolution discret. □

Ce théorème montre que la notion de filtrage par convolution reste valable pour les processus stationnaires puisque les composantes fréquentielles du processus filtré sont atténuées ou amplifiées par $|\hat{h}(\mathrm{e}^{i\omega})|^2$.

Densité d'énergie On appelle $\hat{R}_X(e^{i\xi})$ puissance spectrale car on peut l'interpréter comme une densité d'énergie le long de l'axe des fréquences.

Soit $h^\xi_\Delta[n]$ le filtre réel dont la fonction de transfert est

$$\hat{h}^\xi_\Delta(\mathrm{e}^{i\omega}) = \begin{cases} \frac{1}{\pi\Delta} & \text{si } ||\omega| - \xi| \leq \Delta/2 \\ 0 & \text{si } ||\omega| - \xi| > \Delta/2 \end{cases}$$

L'énergie de ce filtre est normalisée

$$\|h^\xi_\Delta\|^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |\hat{h}^\xi_\Delta(\mathrm{e}^{i\omega})|^2 d\omega = 1.$$

Soit

$$X^\xi_\Delta[n] = X \star h^\xi_\Delta[n]$$

le processus obtenu en ne gardant que les composantes fréquentielles dans le voisinage de ξ. Le résultat (4.14) du théorème 4.1 permet de montrer que la variance de ce processus est

$$\begin{aligned} E\{|X^\xi_\Delta[n]|^2\} &= R_{X^\xi_\Delta}[0] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \hat{R}_{X^\xi_\Delta}(e^{i\omega}) d\omega \\ &= \frac{1}{2} \int_{||\omega| - \xi| \leq \frac{\Delta}{2}} \frac{\hat{R}_X(e^{i\omega})}{\Delta} d\omega. \end{aligned}$$

Comme $R_X[n] \in \mathbf{l^1}(\mathbb{Z})$, $\hat{R}_X(\mathrm{e}^{i\omega})$ est continue et $\hat{R}_X(e^{i\xi}) = \hat{R}_X(e^{-i\xi})$ si bien que

$$\lim_{\Delta \to 0} R_{X_\Delta^\xi}[0] = \hat{R}_X(e^{i\xi}).$$

La puissance spectrale à la fréquence ξ est donc proportionnelle à la densité d'énergie du processus filtré autour de la fréquence ξ.

4.2 Filtrage de Wiener

Lors d'une mesure ou durant une transmission, le signal est souvent contaminé par un bruit additif. Une partie du bruit peut cependant être éliminé grâce à un estimateur que l'on optimise en fonction des propriétés du signal et du bruit.

Le signal et le bruit sont considérés comme des réalisations de deux processus réels stationnaires, respectivement $X[n]$ et $B[n]$. Les données bruitées sont

$$D[n] = X[n] + B[n].$$

On suppose que les valeurs du signal $X[n]$ et du bruit $B[k]$ sont indépendantes pour tout k, n. Dans la suite, on prend $E\{X[n]\} = 0$, ce qui peut être obtenu en soustrayant $E\{X[n]\}$ de $D[n]$, pour se replacer dans les conditions où le signal a une moyenne nulle.

Soit $\tilde{X} = L\,D$ l'estimateur du signal X calculé à partir des données D avec un opérateur L. Pour tout n on veut minimiser *l'erreur quadratique moyenne* $E\{(X[n] - \tilde{X}[n])^2\}$. On sait [8] que l'estimateur optimal $\tilde{X}[n]$ de $X[n]$ à partir des données bruitées $\{D[k]\}_{k\in\mathbb{Z}}$, qui minimise l'espérance quadratique de l'erreur

$$E\{(X[n] - \tilde{X}[n])^2\} \tag{4.16}$$

est l'espérance conditionnelle

$$\tilde{X}[n] = LD[n] = E\{X[n] \;/\; D[k]\, k \in \mathbb{Z}\}. \tag{4.17}$$

L'espérance conditionnelle (4.17) est souvent une fonction non-linéaire compliquée des données $D[k]$, qu'il est difficile de calculer. Le filtrage de Wiener simplifie le problème imposant que L soit un opérateur linéaire.

Estimation linéaire Le filtre de Wiener calcule le meilleur estimateur $\tilde{X}[n]$ qui soit combinaison linéaire des $\{D[k]\}_{k\in\mathbb{Z}}$ et qui minimise (4.16). Le théorème suivant montre que l'erreur obtenue est décorrélée des données.

Théorème 4.2 *Un estimateur linéaire $\tilde{A}$ d'une variable aléatoire A:*

$$\tilde{A} = \sum_{k=-\infty}^{+\infty} a[k]\, D[k]$$

minimise $\epsilon(a) = E\{(A - \tilde{A})^2\}$ si et seulement si

$$E\{(A - \tilde{A})\, D[k]\} = 0 \quad \textit{pour tout } k \in \mathbb{Z}\ . \tag{4.18}$$

L'erreur minimum est

$$\epsilon(a) = E\{A^2\} - \sum_{k=-\infty}^{+\infty} a[k]\, E\{A\, D[k]\} \ . \tag{4.19}$$

Démonstration On calcule

$$\epsilon(a) = E\left\{ \Big(A - \sum_{k=-\infty}^{+\infty} a[k]\, D[k]\Big) \Big(A - \sum_{k=-\infty}^{+\infty} a[k]\, D[k]\Big) \right\}. \tag{4.20}$$

Au minimum,

$$\frac{\partial \epsilon(a)}{\partial a[n]} = -2E\left\{ \Big(A - \sum_{k=-\infty}^{+\infty} a[k]\, D[k]\Big)\, D[n] \right\} = 0,$$

ce qui démontre (4.18). Comme

$$\frac{\partial^2 \epsilon(a)}{\partial a[n]^2} = 2\, E\{|D[n]|^2\} \geq 0,$$

ce point critique est un minimum de la forme quadratique (4.20). On montre que $\epsilon(a) = E\{(A - \tilde{A})\, A\}$ en en utilisant (4.18) dans (4.20), d'où l'on déduit (4.19).

□

Appliqué à $A = X[n]$, le théorème 4.2 montre qu'un estimateur linéaire $\tilde{X}[n]$ minimise $E\{(X[n] - \tilde{X}[n])^2\}$ si et seulement si

$$E\{(X[n] - \tilde{X}[n])\, D[k]\} = 0 \quad \text{pour tout } n, k. \tag{4.21}$$

Le théorème suivant montre que l'estimateur linéaire optimal $\tilde{X}$ qui minimise l'erreur moyenne se calcule avec un filtre dont la fonction de transfert est spécifiée en fonction de la puissance spectrale du signal $\hat{R}_X(\mathrm{e}^{i\omega})$ et du bruit $\hat{R}_B(\mathrm{e}^{i\omega})$.

Théorème 4.3 (Wiener) *L'estimateur linéaire qui minimise l'erreur quadratique moyenne est* $\tilde{X} = D \star h$ *avec*

$$\hat{h}(\mathrm{e}^{i\omega}) = \frac{\hat{R}_X(\mathrm{e}^{i\omega})}{\hat{R}_X(\mathrm{e}^{i\omega}) + \hat{R}_B(\mathrm{e}^{i\omega})} \ . \tag{4.22}$$

L'erreur minimum résultante est

$$\epsilon = E\Big\{(X[n] - \tilde{X}[n])^2\Big\} = \frac{1}{2\pi} \int_0^{2\pi} \frac{\hat{R}_X(\mathrm{e}^{i\omega})\, \hat{R}_B(\mathrm{e}^{i\omega})}{\hat{R}_X(\mathrm{e}^{i\omega}) + \hat{R}_B(\mathrm{e}^{i\omega})}\, d\omega \ . \tag{4.23}$$

Démonstration Soit $\tilde{X}[n]$ un estimateur linéaire de $X[n]$:

$$\tilde{X}[n] = \sum_{l=-\infty}^{+\infty} h[n,l]\, D[l]. \tag{4.24}$$

La condition de non-corrélation (4.21) signifie que pour tout n, k

$$E\{X[n]\,D[k]\} = E\{\tilde{X}[n]\,D[k]\} = \sum_{l=-\infty}^{+\infty} h[n,l]\,E\{D[l]\,D[k]\}.$$

Comme $D[k] = X[k] + B[k]$ et $E\{X[n]\,B[k]\} = 0$, on en déduit que

$$E\{X[n]\,X[k]\} = \sum_{l=-\infty}^{+\infty} h[n,l]\,\Big(E\{X[l]\,X[k]\} + E\{B[l]\,B[k]\}\Big). \tag{4.25}$$

Le signal et le bruit étant stationnaire $E\{X[n]\,X[k]\} = R_X[n-k]$ et $E\{B[n]\,B[k]\} = R_B[n-k]$ donc $h[n,k]$ ne dépend que de $n-k$ et peut s'écrire $h[n,k] = h[n-k]$. On déduit de (4.24) que $\tilde{X} = D \star h$ est un processus stationnaire. Avec les changements de variables $m = n-k$ et $p = n-l$ la relation (4.25) se réécrit

$$R_X[m] = \sum_{p=-\infty}^{+\infty} h[p]\,\Big(R_X[m-p] + R_B[m-p]\Big)$$

et donc

$$R_X[m] = (R_X + R_B) \star h[m]\ .$$

En calculant la transformée de Fourier de chaque membre de cette égalité on en déduit (4.22).

L'erreur de l'estimation se calcule avec (4.19) pour $A = X[n]$ et $a[k] = h[n-l]$

$$\epsilon = E\{X^2[n]\} - \sum_{l=-\infty}^{+\infty} h[n-l]\,E\{X[n]\,D[l]\}\ .$$

Comme $D[l] = X[l] + B[l]$ et que $E\{X[n]B[l]\} = 0$

$$\epsilon = R_X[0] - \sum_{l=-\infty}^{+\infty} h[n-l]\,R_X[n-l] = R_X[0] - \sum_{l=-\infty}^{+\infty} h[l]\,R_X[l]\ .$$

Cette relation s'exprime en fonction des transformées de Fourier de R_X et de h en utilisant la formule de Parseval (3.20)

$$\epsilon = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{R}_X(e^{i\omega})\,d\omega - \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{R}_X(e^{i\omega})\,\hat{h}(\mathrm{e}^{i\omega})\,d\omega\ ,$$

d'où l'on déduit (4.23) en insérant (4.22). □

Ce théorème démontre que la meilleur estimation linéaire est obtenu par filtrage. La fonction de transfert $\hat{h}(\omega)$ est proche de 1 lorsque le rapport signal sur bruit $\frac{\hat{R}_X(\omega)}{\hat{R}_B(\omega)}$ est grand à la fréquence ω. Lorsque ce rapport ce réduit, la valeur de $\hat{h}(\omega)$ tend vers 0 afin d'éliminer les composantes fréquentielles où le bruit domine le signal.

Si $X[n]$ est un vecteur gaussien et $B[n]$ est un bruit blanc gaussien, alors l'estimateur linéaire optimal est aussi optimal parmi les estimateurs non linéaires. En effet, deux variables aléatoires conjointement gaussiennes sont indépendantes si et seulement si elles sont décorrélées [8]. Comme $X[n] - \tilde{X}[n]$ est conjointement gaussien avec $D[k]$, la non-corrélation (4.21) signifie que $X[n] - \tilde{X}[n]$ et $D[k]$ sont indépendants pour tout k, n. Dans ce cas, on peut montrer que que $\tilde{X}$ est égale à l'estimateur optimal (4.17).

La valeur numérique de l'erreur est souvent exprimée par le *rapport signal sur bruit* (SNR pour *Signal to Noise Ratio*), qui se mesure en décibels par

$$SNR_{\text{db}} = 10 \log_{10} \left(\frac{E\{X^2[n]\}}{E\{(X[n] - \tilde{X}[n])^2\}} \right). \tag{4.26}$$

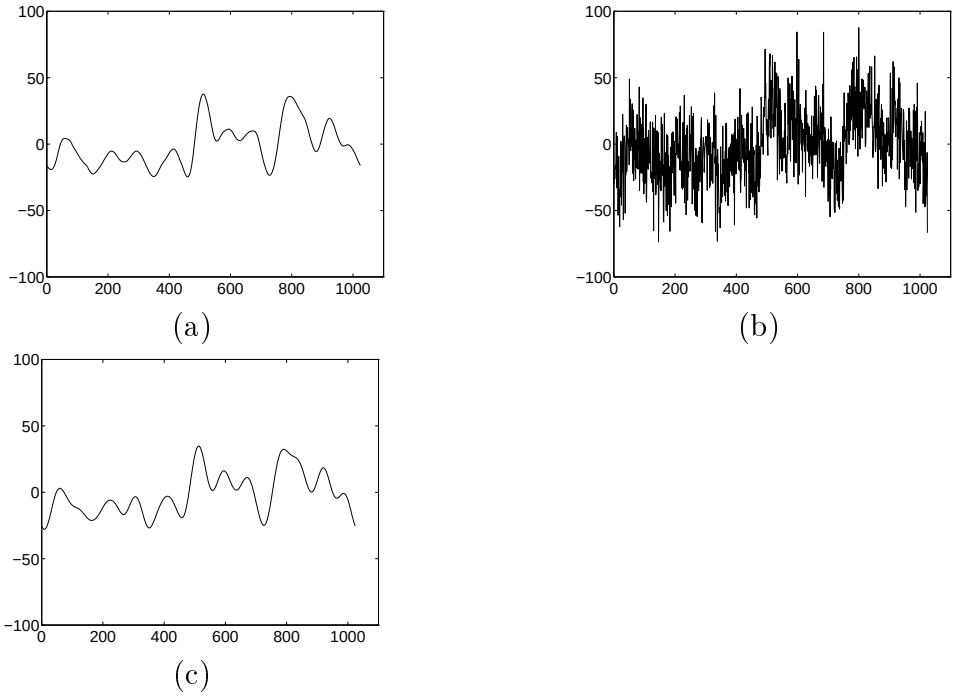

Figure 4.1: (a): Réalisation d'un processus gaussien X. (b): Signal bruité obtenu en ajoutant un bruit blanc gaussien (SNR = -0,48db). (c): Estimation de Wiener $\tilde{X}$ (SNR = 15,2db).

Exemple La figure 4.1(a) montre une réalisation d'un processus gaussien X obtenu par un filtrage passe-bas d'un bruit blanc gaussien B_0 de variance α_0^2:

$$X[n] = B_0 \star g[n],$$

avec

$$g[n] = \cos^2 \left(\frac{\pi n}{2K} \right) \mathbf{1}_{[-K,K]}[n] \ .$$

Le théorème 4.1 montre que

$$\hat{R}_X(\mathrm{e}^{i\omega}) = \hat{R}_B(\mathrm{e}^{i\omega}) \, |\hat{g}(\mathrm{e}^{i\omega})|^2 = \alpha_0^2 \, |\hat{g}(\mathrm{e}^{i\omega})|^2.$$

Le signal bruité D de la figure 4.1(b) est contaminé par un bruit blanc B de variance σ^2. L'estimation de la figure 4.1(c) est calculée par le filtre de Wiener (4.22):

$$\hat{h}(\mathrm{e}^{i\omega}) = \frac{\alpha_0^2 \, |\hat{g}(\mathrm{e}^{i\omega})|^2}{\alpha_0^2 \, |\hat{g}(\mathrm{e}^{i\omega})|^2 + \sigma^2}.$$

C'est un filtre passe-bas qui moyenne le signal bruité afin d'éliminer les fluctuations du bruit blanc. Dans ce cas le signal X et le bruit B étant conjointement gaussiens, le filtre de Wiener est aussi optimal parmi tous les estimateurs non-linéaires.

La figure 4.2 montre un second example où les réalisations d'un processus non gaussien X sont des signaux réguliers par morceaux. Si l'on connait la puissance spectrale de X, on peut alors calculer le filtre de Wiener, ce qui a été fait en (c). On voit que le filtre de Wiener laisse du bruit dans les régions régulières du signal. Un lissage plus fort aurait dégradé davantage les discontinuités du signal ce qui aurait augmenté l'erreur de l'estimation. Dans ce cas, un estimateur non-linéaire effectuant un lissage adaptatif peut nettement réduire l'erreur.

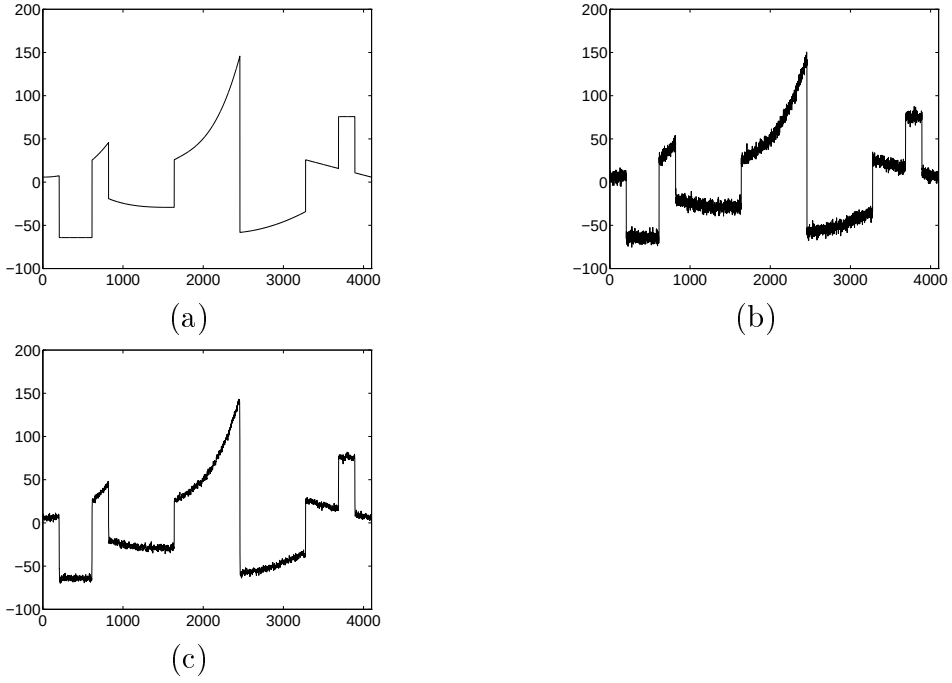

Figure 4.2: (a): Réalisation d'un processus non gaussien X. (b): Signal bruité contaminé par un bruit blanc gaussien (SNR = 21,9db). (c): Estimation de Wiener $\tilde{X}$ (SNR= 25,9db).

Chapitre 5

Traitement de la Parole

Le traitement de la parole nécessite une analyse physiologique des mécanismes de production, ce qui permet de comprendre les propriétés particulières de ce type de signal. Cela motive notament l'utilisation de modèles autorégressifs. L'identification des paramètres de ces modèles se fait par régression linéaire. Outre les applications à la reconnaissance de la parole, ces modèles permettent de coder efficacement les signaux de parole pour la téléphonie cellulaire.

5.1 Modélisation du signal de parole

5.1.1 Production

La production de la parole se fait en trois étapes. Les poumons compressent de l'air qui est envoyé à travers la trachée. Cet air passe par le larynx qui est composé d'un système de cartilages et de muscles incluant les cordes vocales. Le larynx produit alors un signal d'excitation qui se propage à travers le conduit vocal. C'est la déformation du conduit vocal qui produit l'articulation de la parole. Les éléments principaux de cette articulation sont la langue, les lèvres et la machoire inférieure. La figure 5.1 illustre l'appareil phonatoire.

Excitation Le larynx peut produire des signaux d'excitation différents. Les sons voisés tels que les voyelles sont produits par vibration des cordes vocales. L'air est forcé à travers les cordes vocales qui vibrent comme les lèvres d'un trompettiste. Cela produit un train de quasi-impulsions, illustré en figure 5.2, qui est envoyé dans le conduit vocal. La fréquence des répétitions, appelée "pitch", est essentiellement contrôlée par la tension des cordes vocales. Elle correspond à la fréquence fondamentale (hauteur) du son. Dans le cas de la voix parlée, le pitch est typiquement entre 100Hz et 300Hz. Une soprano peut cependant augmenter cette fréquence jusqu'à 3600Hz.

Pour un chuchotement, les cordes vocales ne vibrent pas mais laissent un passage étroit entre les cartilage du larynx, qui envoie un air turbulent dans le conduit vocal. Cet air turbulent peut être modélisé par un bruit Gaussien, dont la puissance spectrale a un large support fréquentiel.

Figure 5.1: Système phonatoire [6].

Figure 5.2: Train de quasi-impulsions émises par les cordes vocales

.

Articulation Le conduit vocal donne l'articulation au son qui caractérise chaque phonème. Nous avons vu que pour un son voisé, le larynx émet un train d'onde riche en harmoniques qui est filtré par le conduit vocal. Ce conduit vocal a des résonances appelées "formants". La déformation du conduit vocal déplace les fréquences de résonance, ce qui permet de former toutes les voyelles et certaines consonnes. Pour des sons non-voisés, le conduit vocal peut aussi effectuer des constrictions qui produit des fricatives ou des sons chuintés telles que le [s] et le [ch]. Les sons plosifs sont produits par une fermeture du conduit vocal ce qui crée une occlusion. Le relâchement brutal de cette occlusion produit alors une plosive telle que [p] ou [t]. L'articulation du son est aussi affectée par l'ouverture de la cavité nasale. Des catégorisations très détaillées des différents sons de parole ont été faites par les phonéticiens.

5.1.2 Conduit vocal

Le conduit vocal peut se modéliser comme la juxtaposition de plusieurs cylindres de même longueur égale à Δ mais de diamètres variables, comme l'illustre la figure 5.3. Chaque cylindre est un système linéaire avec en entrée une onde directe $f_{n-1}(t)$ mesurant le débit du flot d'air qui passe à l'entrée du cylindre par unité de temps, et une onde inverse $b_{n-1}(t)$. En sortie, on a onde directe $f_n(t)$ et une onde inverse $b_n(t)$ qui est reliée à l'entrée

Figure 5.3: Le conduit vocal peut se modéliser comme une succession de cylindres de même longueur Δ ayant des diamètres différents.

par une matrice T_n qui dépend de l'impédance acoustique des cylindres $n-1$ et n

$$\begin{pmatrix} f_n(t) \\ b_n(t) \end{pmatrix} = T_n \begin{pmatrix} f_{n-1}(t) \\ b_{n-1}(t) \end{pmatrix}$$

Si l'on cascade la réponse de chaque cylindre, on obtient

$$\begin{pmatrix} f_p(t) \\ b_p(t) \end{pmatrix} = T \begin{pmatrix} f_0(t) \\ b_0(t) \end{pmatrix}$$

avec

$$T = \prod_{n=1}^{p} T_n.$$

On discrétise ce système avec un pas d'échantillonnage de $\frac{\Delta}{c}$ qui est le temps de propagation de l'onde acoustique dans chaque cylindre. Le signal d'entrée est le signal discret

$$f_0[n] = f_0(\frac{n\Delta}{c}) - b_0(\frac{n\Delta}{c}),$$

tandis que le signal de sortie est

$$f_p[n] = f_p(\frac{n\Delta}{c}) - b_p(\frac{n\Delta}{c}).$$

En écrivant les équations de propagation des ondes et les conservations de débit et de pression au travers des jonctions des cylindres, on peut calculer la fonction de transfert qui relie la transformée en z de $f_0[n]$ et de $f_p[n]$ à partir des diamètres de chacun des cylindres

$$\frac{\hat{f_p}(z)}{\hat{f_0}(z)} = \hat{h}(z).$$

En l'absence de perte le long du système, on peut montrer que $\hat{h}(z)$ a N pôles mais n'a pas de zéros. C'est donc un filtre autorégressif qui peut s'écrire

$$\hat{h}(z) = \frac{1}{a_0 + a_1 z^{-1} + ... + a_N z^{-N}}.$$

Cette condition reste valable tant que le conduit vocal peut être représenté par un seul tube sans embranchement. On néglige donc l'influence introduite par le conduit nasal

Figure 5.4: La figure de gauche donne la position des 8 pôles d'un filtre autorégressif tandis que la figure de droite donne le module $|\hat{h}(\mathrm{e}^{i\omega})|_{\mathrm{db}}$.

ainsi que les pertes d'énergie dues aux vibrations des parois du conduit vocal et aux frictions.

Formants Ce modèle simplifié montre que le conduit vocal peut être représenté par un filtre autorégressif dont les paramètres $a[k]$ dépendent de la configuration du conduit vocal. Ce filtre étant causal et stable, nous savons que les pôles de $\hat{h}(z)$ sont tous de module plus petit que 1. Les pôles de $\hat{h}(z)$ qui sont proches du cercle unité produisent un pic dans le module de la réponse fréquentielle $|\hat{h}(\mathrm{e}^{i\omega})|$. Ces pics de fréquence sont appelés formants. Ils ont une importance particulière pour la reconnaissance des sons produits. Plus le zéro est proche du cercle unité, plus le formant est prononcé. Les formants de plus grande amplitude sont généralement les 2 premiers, apparaissant aux plus basses fréquences. La figure 5.4 donne un exemple de filtre autorégressif ayant 8 pôles dont la position est indiquée dans le plan complexe. La réponse fréquentielle $|\hat{h}(\mathrm{e}^{i\omega})|_{\mathrm{db}}$ est donnée à droite.

5.1.3 Excitation

Sons voisés En mode vibratoire, les cordes vocales émettent un train d'onde qui peut s'écrire

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

où $g(t)$ est une onde élémentaire dont le support est petit devant T (voire figure 5.2). On peut supposer que cette onde est indépendante de la forme du conduit vocal. En

appliquant la formule de Poisson (2.40), on calcule la transformée de Fourier de $f(t)$

$$\hat{f}(\omega) = \hat{g}(\omega)\frac{2\pi}{T}\sum_{n=-\infty}^{+\infty}\delta(\omega - \frac{2n\pi}{T}) = \frac{2\pi}{T}\sum_{n=-\infty}^{+\infty}\hat{g}(\frac{2n\pi}{T})\delta(\omega - \frac{2n\pi}{T}).$$

C'est une succession d'harmoniques dont l'enveloppe est égale a $\hat{g}(\omega)$.

On a vu que le conduit vocal est équivalent à un filtre linéaire de fonction de transfert $\hat{h}(\omega)$. La transformée de Fourier du son émis est donc

$$\hat{f}(\omega)\hat{h}(\omega) = \hat{h}(\omega)\hat{g}(\omega)\frac{2\pi}{T}\sum_{n=-\infty}^{+\infty}\delta(\omega - \frac{2n\pi}{T}).$$

Cette réponse peut être modélisée comme un train d'impulsions de Diracs qui passe à travers un filtre dont la fonction de transfert est spécifiée par l'enveloppe totale $\hat{h}(\omega)\hat{g}(\omega)$.

On sait que la discrétisation de l'onde se propageant dans le conduit vocal peut se modéliser par un filtrage autorégressif. La discrétisation de $g(t)$ peut de même être approximée par un filtre autorégressif. Un signal de parole voisé discrétisé se modélise donc par un train d'impulsion de Diracs discrets $\sum_{k=-\infty}^{+\infty}\delta[n-kT]$ filtré par un filtre autorégressif qui dépend à la fois du conduit vocal et de la forme de l'impulsion $g(t)$ produite par les cordes vocales.

Sons non voisés Les sons non voisés sont produits par un signal turbulent émis par le larynx qui est ensuite modifié par le conduit vocal. La discrétisation de ce signal turbulent peut être modélisée par un processus Gaussien stationnaire $Y[n]$ dont la puissance spectrale $\hat{R}_Y(\mathrm{e}^{i\omega})$ à une énergie qui est répartie sur une large bande de fréquence. Un tel processus peut aussi s'ecrire comme un bruit blanc Gaussien $B[n]$ filtré par un filtre $g[n]$

$$Y[n] = B \star g[n].$$

Le théorème 4.1 prouve que la puissance spectrale de $Y[n]$ est reliée a $\hat{g}(\mathrm{e}^{i\omega})$ par

$$\hat{R}_Y(\mathrm{e}^{i\omega}) = |\hat{g}(\mathrm{e}^{i\omega})|^2.$$

Nous avons vu que le conduit vocal se comporte comme un filtre AR de réponse impulsionelle $h[n]$. Le son produit est alors modélisé par le processus

$$X[n] = Y \star h[n] = B \star g \star h[n].$$

Un tel signal discrétisé se modélise donc par un bruit blanc discret $B[n]$ filtré par $g \star h[n]$ dont la fonction de transfert est $\hat{g}(z)\hat{h}(z)$. En général, $\hat{g}(z)$ peut avoir des zéros. Ces zéros sont negligés car leur importance perceptuelle est secondaire à côte des pôles. On modélise donc $\hat{g}(z)\hat{h}(z)$ par un seul filtre autorégressif.

Modèle synthétique Suivant que le son est voisé ou non, le signal de parole peut se modéliser comme un train d'impulsion ou comme la réalisation d'un bruit blanc filtré par un filtre autorégressif dont les caractéristiques dépendent du son prononcé. Dans le cas

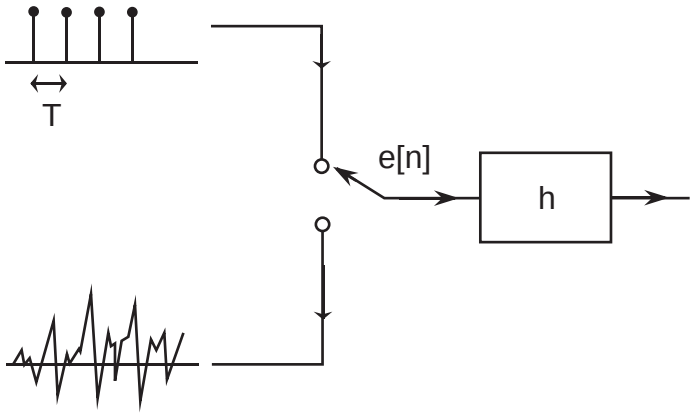

Figure 5.5: Modélisation d'un son de parole par une excitation périodique ou aléatoire, filtrée par un filtre autorégressif.

d'un son voisé, la période des impulsions (pitch) est un paramètre qui doit être déterminé. Ce modèle est illustré par la figure 5.5.

Stationnarité Le conduit vocal ne se comporte comme un filtre linéaire homogène que sur des intervalles de temps relativement petits. Sur une durée plus longue, un signal de parole est non stationnaire puisque les sons changent au cours du temps. La taille des intervalles sur lesquels le son peut être approximé par une excitation modulée par un filtre homogène dépend de la nature du son. Pour une voyelle, cette approximation est valable sur environ 10^{-2} seconde alors que beaucoup de sons de consonnes telles que les plosives ne restent pas stationnaires sur cette durée. La variation de ces intervalles de stationnarité est l'une des difficultés du traitement de la parole.

5.2 Processus autorégressifs

Un son non voisés est modélisé comme étant un bruits blanc filtré par un filtre autorégressif stable. Le processus résultant est appellé *processus autorégressif*. Cette classe de processus est souvent utilisée en traitement du signal car les paramètres d'un modèle autorégressif peuvent facilement être calculés à partir de l'autocovariance.

Un processus autorégressif $X[n]$ s'écrit

$$X[n] = B \star h[n] \tag{5.1}$$

où $B[n]$ est un bruit blanc de variance σ^2 et h est un filtre autorégressif normalisé dont la fonction de transfert s'écrit

$$\hat{h}(z) = \frac{1}{1 - a_1 z^{-1} - ... - a_p z^{-p}}.$$

On suppose que ce filtre est causal et stable et donc que ses poles sont inclus dans le cercle unité. Le théorème 4.1 montre que $X[n]$ est un processus stationnaire dont la puissance spectrale est

$$\hat{R}_X(\mathrm{e}^{i\omega}) = \hat{R}_B(\mathrm{e}^{i\omega})\,|\hat{h}(\mathrm{e}^{i\omega})|^2 = \frac{\sigma^2}{|1 - \sum_{k=1}^{N} a_k e^{-ik\omega}|^2},$$

car $\hat{R}_B(\mathrm{e}^{i\omega}) = \sigma^2$.

On a vu en (3.34) qu'un filtre autorégressif relie l'entrée et la sortie par une équation récurrente (3.28). Comme $X[n] = B \star h[n]$ on obtient

$$X[n] - \sum_{k=1}^{N} a_k X[n-k] = B[n]. \tag{5.2}$$

Cette équation est une régression linéaire de $X[n]$ sur N valeurs passées $\{X[n-k]\}_{1\leq k\leq N}$, et l'erreur $B[n]$ est l'*innovation* au temps n, non prévue par la régression.

Pour identifier un processus autorégressif à partir d'une réalisation il nous faut calculer les constantes de régression $\{a_k\}_{1\leq k\leq N}$. Comme X est stationaire, on peut estimer sa covariance $R_X[k]$ à partir d'une réalisation avec la somme empirique (4.8). Or nous allons montrer que l'on peut calculer les constantes de régression en résolvant un système linéaire faisant intervenir R_X.

L'autocovariance se calcule directement à partir de l'équation récurrente

$$X[n] - \sum_{k=1}^{N} a_k\, X[n-k] = B[n]. \tag{5.3}$$

En multipliant par $X[n-l]$ de chaque côté de (5.3) et en calculant l'espérance on obtient

$$E\{X[n]X[n-l]\} - E\left\{\sum_{k=1}^{N} a_k\, X[n-k]X[n-l]\right\} = E\{B[n]X[n-l]\}.$$

Comme h est un filtre causal, $X[n-k] = B \star h[n-k]$ est une combinaison linéaire de $\{B[l]\}_{l\leq n-k}$. Par ailleur B étant un bruit blanc, $E\{B[n]X[n-l]\} = 0$ si $l > 0$ et donc

$$R_X[l] - \sum_{k=1}^{N} a_k R_X[l-k] = 0. \tag{5.4}$$

Si l'on réécrit (5.4) pour $1 \leq l \leq N$, on obtient le système de N équations de Yule-Walker qui s'écrit sous forme matricielle

$$\begin{pmatrix} R_X[0] & R_X[1] & \dots & R_X[N-1] \\ R_X[-1] & R_X[0] & \dots & R_X[N-2] \\ . & . & \dots & . \\ . & . & \dots & . \\ . & . & \dots & . \\ R_X[-N+1] & R_X[-N+2] & \dots & R_X[0] \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_N \end{pmatrix} = \begin{pmatrix} R_X[1] \\ R_X[2] \\ . \\ . \\ . \\ R_X[N] \end{pmatrix}. \tag{5.5}$$

Si cette matrice symétrique n'est pas singulière alors le vecteur de coefficients $(a_k)_{1\leq k\leq N}$ s'obtient à partir de $R_X[n]$ en inversant la matrice. L'algorithme rapide de Levinson-Durbin permet d'effectuer ce calcul avec $O(N^2)$ opérations [1].

On peut calculer la covariance R_X en fonction des a_k en résolvant l'équation récurrente

$$R_X[l] - \sum_{k=1}^{N} a_k R_X[l-k] = 0.$$

On peut vérifier que la solution sécrit

$$R_X[l] = \sum_{k=0}^{N-1} \lambda_k \, (c_k)^l,$$

où les c_k sont les racines de l'équation

$$1 - \sum_{k=1}^{N} a_k \, z^{-k} = 0 \ .$$

Les c_k sont les pôles de $\hat{h}(z)$ et comme h est stable, $|c_k| < 1$ pour tout $0 \leq k < N$. On en déduit l'autocovariance décroit exponentiellement $\lim_{l \to +\infty} R_X[l] = 0$.

La variance du bruit blanc $B[n]$ se calcule à partir de (5.3) en observant que comme $B[n]$ est indépendant de $X[n-l]$

$$\sigma^2 = E\{B[n]^2\} = E\{B[n]X[n]\} = R_X[0] - \sum_{k=1}^{N} a_k \, R_X[k]. \tag{5.6}$$

Comme les coefficients a_k ne dépendent que de $R_X[l]$, la variance du bruit blanc est aussi entièrement définie par $R_X[l]$.

5.3 Estimation d'un modèle de parole

Nous avons expliqué qu'un signal de parole peut localement être approximé par une excitation $e[n]$ filtrée par un filtre AR dont les paramètres dépendent du son prononcé. Pour des applications de reconnaissance et de codage, on veut identifier l'excitation ainsi que les paramètres du filtre. On isole une portion d'un signal de parole $f[n]$ en le multipliant avec une fenêtre $w[n]$ de taille P centrée en un instant pP

$$x[n] = f[n]w[n - pP],$$

comme le montre la figure 5.6. On prend P suffisamment petit pour que le son isolé puisse être considéré comme stationnaire. Par exemple, la fenêtre de Hamming est définie par

$$w[n] = \begin{cases} 0.54 + 0.46 \cos(\frac{2\pi n}{P}) & \text{si } |n| < \frac{P}{2} \\ 0 & \text{sinon} \end{cases}$$

Le signal résultant $x[n]$ possède au plus P coefficients non-nuls. Dans un modèle de parole, $x[n]$ est produit par une excitation $e[n]$ filtrée par un filtre AR dont la fonction de transfert est renormalisée pour s'écrire

$$\hat{h}(z) = \frac{1}{1 - a_1 z^{-1} - ... - a_N z^{-N}}.$$

Figure 5.6: Des portions du signal de parole sont isolées par des fenêtres, qui couvrent des intervalles de temps où le signal peut être considéré comme stationnaire.

Calcul de l'excitation De nombreuses techniques ont été développées pour déterminer le voisement et le pitch d'un son. Une approche particulièrement simple est basée sur la somme des différences du signal à intervalles k variables

$$d[k] = \frac{1}{P} \sum_n |x[n] - x[n-k]|. \tag{5.7}$$

Si le signal est voisé et que la fréquence des impulsions est T alors $d[k]$ a un minimum a $k = T$. Lorsque le minimum de $d[k]$ n'est pas suffisamment bas, on en déduit que le son est non voisé. Cet algorithme a l'avantage de ne nécessiter aucune multiplication et donc de s'implémenter très rapidement.

5.3.1 Régression linéaire

Pour identifier tous les paramètres d'un son, il nous faut estimer les paramètres a_k du filtre AR qui spécifient les propriétés du conduit vocal. Lorsque l'excitation est un bruit blanc, le signal est la réalisation d'un processus autorégressif $X[n]$ et les équations de Yule-Walker (5.5) permettent de calculer les paramètres a_k à partir de l'autocovariance R_X. Nous allons retrouver ce résultat d'un point de vue déterministe par un calcul de régression linéaire.

Le signal $x[n]$ satisfait l'équation récurrente

$$x[n] - \sum_{k=1}^{N} a_k x[n-k] = e[n].$$

On peut interpréter

$$\tilde{x}[n] = \sum_{k=1}^{N} a_k \, x[n-k] \tag{5.8}$$

comme une estimation de $x[n]$ à partir de N valeurs passées, auquel cas l'erreur d'estimation n'est autre que l'excitation

$$x[n] - \tilde{x}[n] = e[n].$$

Si $e[n]$ est la réalisation d'un bruit blanc, donc non corrélée d'un échantillon à l'autre, chaque $e[n]$ peut être considéré comme l'innovation apportée par l'excitation relativement

à la prédiction de $x[n]$ par son passé. Ce point de vue permet de poser l'identification des paramètres a_k du filtre AR comme un problème de prédiction linéaire. Etant donné un signal $x[n]$, on veut calculer les coefficients de régression tels que l'estimation (5.8) génère une erreur

$$\sum_{n=-\infty}^{+\infty} (x[n] - \tilde{x}[n])^2 = \sum_{n=-\infty}^{+\infty} e^2[n] \tag{5.9}$$

qui est minimum. Pour effectuer ce calcul on introduit l'autocorrélation empirique du signal $x[n]$

$$r_x[k] = \sum_{n=-\infty}^{+\infty} x[n-k]x[n]. \tag{5.10}$$

Cette somme s'étend seulement sur P valeurs car $x[n]$ a au plus P valeurs consécutives non nulles. Le théorème suivant caractérise les coefficients de régression de $\tilde{x}[n]$.

Théorème 5.1 (Prédiction linéaire) *Les coefficients de régression $\{a_k\}_{1\leq k\leq N}$ du signal $\tilde{x}[n] = \sum_{k=1}^{N} a_k\, x[n-k]$ qui minimise*

$$\epsilon = \sum_{n=-\infty}^{+\infty} |x[n] - \tilde{x}[n]|^2$$

sont solutions du système

$$\begin{pmatrix} r_x[0] & r_x[1] & ... & r_x[N] \\ r_x[-1] & r_x[0] & ... & r_x[N-1] \\ . & . & ... & . \\ . & . & ... & . \\ . & . & ... & . \\ r_x[-N] & r_x[-N+1] & ... & r_x[0] \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ . \\ . \\ . \\ a_N \end{pmatrix} = \begin{pmatrix} r_x[1] \\ r_x[2] \\ . \\ . \\ . \\ r_x[N] \end{pmatrix}. \tag{5.11}$$

L'erreur résultante est

$$\epsilon = r_x[0] - \sum_{k=1}^{N} a_k\, r_x[k]. \tag{5.12}$$

Démonstration La démonstration se fait par une interprétation géométrique du problème de minimisation Le signal $x[n]$ a une énergie finie et donc appartient à l'espace $\mathbf{l^2}(\mathbb{Z})$ muni de la norme

$$\|x[n]\|^2 = \sum_{n=-\infty}^{+\infty} |x[n]|^2$$

et du produit scalaire

$$< x[n], y[n] >= \sum_{n=-\infty}^{+\infty} x[n]\, y[n].$$

Le vecteur $\tilde{x}[n]$ est une combinaison linéaire des vecteurs $\{x_k[n] = x[n-k]\}_{1\leq k\leq N}$ et appartient donc à l'espace $\mathbf{V}$ généré par ces N vecteurs. Minimiser l'erreur de prédiction

(5.9) revient donc trouver un vecteur $\tilde{x}[n]$ de $\mathbf{V}$ qui minimise $\|x - \tilde{x}\|^2$. Le théorème de projection prouve que ce vecteur est la projection orthogonale de $\tilde{x}$ dans $\mathbf{V}$. C'est donc un vecteur tel que $x - \tilde{x}$ est orthogonal à tous les vecteurs de $\mathbf{V}$ et en particulier aux vecteurs $\{x[n-k]\}_{1 \leq k \leq N}$

$$< x[n] - \tilde{x}[n], x[n-k] >= \sum_{n=-\infty}^{+\infty} (x[n] - \tilde{x}[n]) x[n-k] = 0.$$

En insérant (5.8) ces équations se réécrivent

$$\sum_{n=-\infty}^{+\infty} x[n]x[n-k] - \sum_{1=0}^{N} a[p] \sum_{n=-\infty}^{+\infty} x[n-p]x[n-k] = 0. \tag{5.13}$$

En insérant (5.10) on obtient

$$\sum_{p=1}^{N} a[p] r_x[k-p] = r_x[k] \quad , \quad \text{pour } 1 \leq k \leq N,$$

qui correspond au système de N équations à N inconnues (5.11).

Comme $x[n] - \tilde{x}[n]$ est orthogonal à tout vecteur dans $\mathbf{V}$ et donc à $\tilde{x}[n]$, l'énergie de l'erreur peut s'écrire

$$\epsilon =< x[n] - \tilde{x}[n], x[n] - \tilde{x}[n] >=< x[n] - \tilde{x}[n], x[n] > .$$

En remplaçant $\tilde{x}[n]$ par son expression (5.8), on obtient

$$\epsilon = \sum_{n=-\infty}^{+\infty} x[n]x[n] - \sum_{p=1}^{N} a[p] \sum_{n=-\infty}^{+\infty} x[n-p]x[n]$$

et donc (5.12) □

Filtre AR pour sons non-voisés Un son non-voisé est modelisé par bruit blanc traversant un filtre AR. En comparant le système de Yule-Walker (5.5) et le système (5.11), on s'aperçoit que ces équations sont identiques lorsque l'on remplace l'autocorrélation $R_X[k]$ du processus $X[n]$ par l'autocorrélation empirique $r_x[k]$ du signal $x[n]$. Si $e[n]$ est une réalisation du bruit blanc $B[n]$ alors $x[n]$ est une réalisation du processus autorégressif $X[n]$. L'autocorrélation empirique $r_x[k]$ peut donc s'interpréter comme une estimation de la véritable autocorrélation $R_X[k]$ du processus. Les équations de prédictions linéaires (5.11) sont donc une approximation des équations de Yule-Walker (5.5), qui permettent d'estimer les coefficients $a[k]$ du filtre AR. La résolution du systéme (5.11) peut se faire en utilisant l'algorithme rapide de Levinson-Durbin qui nécessite $O(N^2)$ opérations [1].

Filtre AR pour son voisé Lorsque le son est voisé, l'algorithme de régression linéaire donne aussi une bonne estimation du filtre AR. La justification est cependant plus compliquée et nous n'en donnons qu'une explication superficielle. Le signal de parole $f[n]$ est

construit en filtrant un train d'impulsion

$$e[n] = \sum_{k=-\infty}^{+\infty} \delta[n - kT]$$

avec un filtre AR $h[n]$ et $x[n]$ est obtenu en multipliant $f[n]$ par la fenêtre $w[n - pP]$

$$x[n] = w[n - pP]\ (e \star h)[n].$$

On peut en déduire que le spectre $\hat{x}(\mathrm{e}^{i\omega})$ est la somme de composantes harmoniques situées autour des fréquences $\omega = \frac{2k\pi}{T}$ dont l'amplitude est proportionnelle a $\hat{h}(e^{\frac{2k\pi}{T}})$. Pour vérifier que le filtre obtenu par régression linéaire est proche du filtre $h[n]$ on montre que l'optimisation des coefficients a_k de la régression linéaire calcule un filtre autorégressif qui interpole approximativement les valeurs $\hat{h}(e^{\frac{2k\pi}{T}})$. Comme le filtre $\hat{h}(\mathrm{e}^{i\omega})$ est lui-même autorégressif, on en déduit que la régression linéaire optimale calcule une approximation de ce filtre.

5.3.2 Compression par prédiction linéaire

Pour la téléphonie et en particulier les téléphones cellulaires, le débit d'information est limité par la gamme de fréquences utilisable pour la transmission. Au contraire, la demande augmente constamment, ce qui nécessite de transmettre toujours plus de conversations. Une solution est de comprimer le signal de parole pour augmenter le nombre de conversations sous contrainte d'un débit fixe. La qualité du signal de parole peut être dégradée mais le codage doit maintenir une bonne intelligibilité des sons prononcés. Pour des forts taux de compression, le codage par prédiction linéaire est actuellement la technique la plus efficace.

Le standard LPC-10 demande 2400bits/s pour coder un signal de parole échantillonné à 8kHz. Le signal de parole est divisé sur des fenêtres de $P = 180$ échantillons. Un filtre AR d'ordre $N = 10$ est calculé pour chaque fenêtre, à partir de l'autocorrélation du signal, par régression linéaire. Le voisement et le pitch sont déterminés en testant l'amplitude des différences (5.7) à intervalles variables. Pour les signaux voisés, on code aussi l'intervalle T du pitch, en quantifiant uniformément $\log T$. Les algorithmes de quantification sont présentés dans le paragraphe 7.2.

La quantificaton des coefficients a_k va déplacer les pôles du filtre AR, qui risquent de sortir du cercle unité. Le filtre résultant est alors instable. Pour guarantir la stabilité du filtre AR, on quantifie plutot un ensemble de N paramètres, appellés coefficients de réflexion $\{K_m\}_{1 \leq m \leq N}$ [1]. Ces coefficients charactérisent les valeurs des $\{a_k\}_{1 \leq k \leq N}$ et on peut vérifier que le filtre AR est stable si et seulement si $|K_m| < 1$ pour $1 \leq m \leq N$. Il suffit donc de s'assurer que les valeurs quantifiées des K_m restent plus petites que 1 pour obtenir un filtre AR stable.

A la réception, on restaure un signal dans chaque fenêtre de 180 échantillons en utilisant les paramètres du code. Si l'excitation est codée comme étant un bruit blanc, elle est reproduite avec un générateur de nombres aléatoires. Sinon, on génère un train d'impulsions séparées par un intervalle T dont la valeur a été codée. Cette excitation est ensuite filtrée par le filtre AR dont les coefficients de réflection ont été transmis.

La qualité de ce code peut être améliorée en reproduisant plus fidèlement l'excitation $e[n]$. Au lieu de coder cette excitation comme un bruit blanc ou un train d'impulsions, des techniques de quantifications vectorielles permettent de restaurer des propriétés importantes de cette excitation de façon a synthétiser des voix de meilleure qualité. Ces codes sont appelés "Coded Excited Linear Predictive Filters" (CELP).

5.3.3 Reconnaissance de la parole

La découverte dans les années 50 des propriétés acoustiques de la parole ainsi que de la structure des formants a ouvert la possibilité d'automatiser la reconnaissance de la parole. Plus de 40 ans plus tard, le problème se révèle bien plus difficile qu'on ne s'y attendait. On distingue plusieurs types de problèmes. La reconnaissance de mots isolés séparés par une pause, la détection de mots appartenant à un vocabulaire limité dans un flot continu de parole, et la reconnaissance de parole sans pose. Les difficultés de la reconnaissance de parole ont diverses origines.

• Variations mono-locuteur. La prononciation est souvent déformée. Par exemple, le "et" peut être réduit à un simple grognement. La prononciation d'un phonème est aussi affectée par le son avant et après. Enfin, le débit de parole peut varier de façon considérable.

• Variations multi-locuteurs. Par exemple, pour les sons voisés, la position des 2 premiers formants varient d'un locuteur à l'autre.

• Ambiguïté des sons. Les variables acoustiques ne spécifient pas toujours de façon unique les variables phonétiques. Il est souvent nécessaire d'utiliser des informations complémentaires provenant de la structure du langage.

• Bruits et interférences. Un son de parole est souvent superposé avec d'autres sons provenant éventuellement d'une autre conversation, qu'il est nécessaire d'éliminer lors de la reconnaissance.

Reconnaissance de mots isolés La localisation fréquentielle des formant donne des indications essentielles pour reconnaître les phonèmes et donc les mots isolés qu'ils composent. Ces formants peuvent être identifiés à partir de la position des pôles du filtre autorégressif associé, comme on l'a expliqué au paragraphe 5.1.2. Cependant, l'utilisation des formants n'est souvent pas suffisante pour obtenir un bon taux de reconnaissance. On améliore la reconnaissance en mesurant les probabilités de transition d'un son à un autre. A partir d'un modèle basé sur des chaînes de Markov, on cherche alors le mot qui a une probabilité conditionnelle maximum, étant donné le signal observé.

Le débit de parole étant un paramètre non contrôlé, il est aussi nécessaire de renormaliser le temps pour faire correspondre le son avec les structures de référence utilisées pour la reconnaissance. Cela peut se faire avec des dilatations et compressions arbitraires en essayant d'optimiser un critère de correspondance. D'autres algorithmes effectuent des dilatations temporelles locales, guidées par les structures du son.

Il existe actuellement des softwares commerciaux effectuant de la reconnaissance de mots isolés. Pour la parole continue, la fiabilité n'est pas suffisante pour que cela soit utilisé comme interface standard avec des systèmes d'information tel qu'un ordinateur.

Chapitre 6

Analyse Temps-Fréquence

En écoutant de la musique, nous percevons clairement les variations temporelles des "fréquences" sonores. On met en évidence les propriétés temporelles et fréquentielles des sons grâce à la transformée de Fourier à fenêtre, qui décompose les signaux en fonctions élémentaires bien concentrées en temps et en fréquence. La mesure des variations temporelles des "fréquences instantanées" est une application importante, qui illustre les limitations imposées par le principe d'incertitude de Heisenberg.

6.1 Transformée de Fourier à fenêtre

On peut classifier les sons suivant leurs propriétés fréquentielles. Par exemple, les fréquences de résonance du conduit vocale produisent des "formants" qui caractérisent les voyelles. La transformée de Fourier ne peut pas être utilisée car

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) \mathrm{e}^{-i\omega t} dt$$

dépend des valeurs de $f(t)$ à tout instant. Pour différencier des sons produits successivement, on définit une transformée de Fourier à fenêtre qui sépare les différentes composantes du signal grâce à une fenêtre translatée.

La fenêtre $g(t)$ est une fonction paire dont le support est concentré au voisinage de 0, que l'on normalise

$$\|g\|^2 = \int_{-\infty}^{+\infty} |g(t)|^2 \, dt = 1 \ .$$

La transformée de Fourier à fenêtre au voisinage de u, à la fréquence ξ est définie par

$$Sf(u,\xi) = \int_{-\infty}^{+\infty} f(t) \, g(t-u) \, \mathrm{e}^{-i\xi t} dt.$$

Localisation temps-fréquence On définit

$$g_{u,\xi}(t) = g(t-u) \, \mathrm{e}^{i\xi t}$$

qui peut être interprété comme une "note de musique" localisée au voisinage de $t = u$, et autour de la fréquence ξ. La transformée de Fourier à fenêtre mesure la corrélation entre le signal $f(t)$ et cette note élémentaire

$$Sf(u,\xi) = \int_{-\infty}^{+\infty} f(t) g^*_{u,\xi}(t) dt. \tag{6.1}$$

Comme g est paire, $g_{u,\xi}(t) = \mathrm{e}^{i\xi t} g(t-u)$ en centré sur u. Le produit $f\, g_{u,\xi}$ isole donc les composantes de f au voisinage de u, et $Sf(u,\xi)$ ne dépend que des propriétés de f dans ce voisinage. L'étalement en temps de $g_{u,\xi}$ est indépendant de u et de ξ:

$$\sigma_t^2 = \int_{-\infty}^{+\infty} (t-u)^2 \, |g_{u,\xi}(t)|^2 \, dt = \int_{-\infty}^{+\infty} t^2 \, |g(t)|^2 \, dt. \tag{6.2}$$

Si l'on applique la formule de Parseval (2.29) à (6.1), on obtient une intégrale fréquentielle

$$Sf(u,\xi) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \, \hat{g}^*_{u,\xi}(\omega) \, d\omega.$$

La valeur $Sf(u,\xi)$ ne dépend donc que du comportement de $\hat{f}(\omega)$ dans le domaine fréquentiel où $\hat{g}^*_{u,\xi}(\omega)$ n'est pas négligeable. La transformée de Fourier $\hat{g}$ de g est réelle et symétrique car g est réelle et symétrique. En utilisant les propriétés (2.18) et (2.19), on montre que la transformée de Fourier de $g_{u,\xi}(t) = g(t-u)e^{i\xi t}$ peut s'écrire

$$\hat{g}_{u,\xi}(\omega) = \mathrm{e}^{-iu(\omega-\xi)} \, \hat{g}(\omega - \xi).$$

C'est donc une fonction centrée à la fréquence $\omega = \xi$. Son étalement fréquentiel autour de ξ vaut

$$\sigma_\omega^2 = \frac{1}{2\pi} \int_{-\infty}^{+\infty} (\omega - \xi)^2 \, |\hat{g}_{u,\xi}(\omega)| \, d\omega = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \omega^2 \, |\hat{g}(\omega)| \, d\omega. \tag{6.3}$$

Il est indépendant de u et de ξ. Dans un plan temps-fréquence (t,ω), on représente $g_{u,\xi}$ par une *boîte de Heisenberg* de taille $\sigma_t \times \sigma_\omega$, centrée en (u,ξ), comme on peut le voir sur la figure 6.1. La taille de cette boîte ne dépend pas de (u,ξ), ce qui veut dire que la résolution de la transformée de Fourier fenêtrée reste constante sur tout le plan temps-fréquence.

Incertitude de Heisenberg Pour mesurer les composantes de f et de $\hat{f}$ dans des petits voisinages de u et ξ il faut construire une fenêtre $g(t)$ qui est bien localisée dans le temps, et dont l'énergie de la transformée de Fourier est concentrée dans un petit domaine fréquentiel. Le Dirac $g(t) = \delta(t)$ a un support ponctuel $t = 0$, mais sa transformée de Fourier $\hat{\delta}(\omega) = 1$ a une énergie qui est distribuée uniformément sur toutes les fréquences. On sait que $|\hat{g}(\omega)|$ décroit rapidement dans les hautes fréquences seulement si $g(t)$ est une fonction qui varie régulièrement. L'énergie de g est donc nécessairement répartie sur un domaine temporel relativement large.

Pour réduire l'étalement temporel de g, on peut opérer un changement d'échelle de temps d'un facteur $s < 1$ sans changer son énergie totale, soit

$$g(t) = \frac{1}{\sqrt{s}} \, g_0\left(\frac{t}{s}\right) \quad \text{avec} \quad \|g\|^2 = \|g_0\|^2.$$

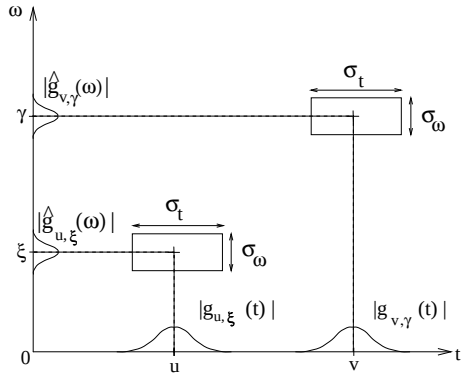

Figure 6.1: Les boîtes de Heisenberg de deux atomes de Fourier fenêtrés $g_{u,\xi}$ and $g_{\nu,\gamma}$.

La transformée de Fourier $\hat{g}(\omega) = \sqrt{s}\,\hat{g}_0(s\omega)$ est dilatée d'un facteur $1/s$, et on perd donc en fréquentiel ce qu'on a gagné en temporel. On voit apparaître un compromis entre la localisation en temps et celle en fréquence.

Les concentrations en temps en en fréquences sont limitées par le principe d'incertitude d'Heisenberg. Ce principe d'incertitude à une interprétation, particulièrement importante en mécanique quantique, comme une incertitude sur la position et l'impulsion d'une particule libre. Plus σ_t et σ_ω sont grandes, plus on a d'incertitude sur la position et l'impulsion de la particule libre. Le théorème suivant montre que le produit $\sigma_t \times \sigma_\omega$ ne peut être arbitrairement petit.

Théorème 6.1 (Incertitude de Heisenberg) *On suppose que $g \in \mathbf{L^2}(\mathbb{R})$ est une fonction centrée en 0 et dont la transformée de Fourier est aussi centrée en 0:*

$$\int_{-\infty}^{+\infty} t\,|g(t)|^2\,dt = \int_{-\infty}^{+\infty} \omega\,|\hat{g}(\omega)|^2\,d\omega = 0\ .$$

Alors les variances definies en (6.2,6.3) satisfont

$$\sigma_t^2\,\sigma_\omega^2 \geq \frac{1}{4}. \tag{6.4}$$

Cette inégalité est une égalité si et seulement si il existe $(a,b) \in \mathbb{C}^2$ tel que

$$g(t) = a\,\mathrm{e}^{-bt^2}. \tag{6.5}$$

Démonstration La preuve suivante, due à Weyl, suppose que $\lim_{|t|\to+\infty} \sqrt{t}g(t) = 0$, mais le théorème est vrai pour tout $g \in \mathbf{L^2}(\mathbb{R})$. Remarquons que

$$\sigma_t^2\,\sigma_\omega^2 = \frac{1}{2\pi\|g\|^4}\int_{-\infty}^{+\infty} |t\,g(t)|^2\,dt \int_{-\infty}^{+\infty} |\omega\,\hat{g}(\omega)|^2\,d\omega. \tag{6.6}$$

Comme $i\omega\hat{g}(\omega)$ est la transformée de Fourier de $g'(t)$, l'identité de Plancherel (2.30) appliquée à $i\omega\hat{g}(\omega)$ donne

$$\sigma_t^2\sigma_\omega^2 = \frac{1}{\|g\|^4}\int_{-\infty}^{+\infty} |t\,g(t)|^2\,dt \int_{-\infty}^{+\infty} |g'(t)|^2\,dt. \tag{6.7}$$

L'inégalité de Schwarz implique

$$\begin{aligned}\sigma_t^2\sigma_\omega^2 &\geq \frac{1}{\|g\|^4}\left[\int_{-\infty}^{+\infty} |t\,g'(t)\,g^*(t)|\,dt\right]^2 \\ &\geq \frac{1}{\|g\|^4}\left[\int_{-\infty}^{+\infty} \frac{t}{2}\left[g'(t)\,g^*(t)+g'^*(t)\,g(t)\right]dt\right]^2 \\ &\geq \frac{1}{4\|g\|^4}\left[\int_{-\infty}^{+\infty} t\,(|g(t)|^2)'\,dt\right]^2.\end{aligned}$$

Comme $\lim_{|t|\to+\infty}\sqrt{t}\,g(t)=0$, on obtient, après intégration par parties

$$\sigma_t^2\sigma_\omega^2 \geq \frac{1}{4\|g\|^4}\left[\int_{-\infty}^{+\infty}|g(t)|^2\,dt\right]^2 = \frac{1}{4}. \tag{6.8}$$

Pour atteindre l'égalité, il faut que l'inégalité de Schwarz appliquée à (6.7) soit elle-même une égalité. Cela implique qu'il existe $b\in\mathbb{C}$ tel que

$$g'(t) = -2\,b\,t\,g(t). \tag{6.9}$$

Il existe donc $a\in\mathbb{C}$ tel que $g(t)=a\,\mathrm{e}^{-bt^2}$. Les inégalités suivantes dans la preuve sont alors des égalités, ce qui fait qu'on atteint effectivement le minorant. □

En mécanique quantique, ce théorème montre qu'on ne peut arbitrairement réduire l'incertitude à la fois sur la position et sur l'impulsion d'une particule libre. Les gaussiennes (6.5) ont une localisation minimale à la fois en temps et en fréquences.

Spectrogramme On peut associer à la transfromée de Fourier à fenêtre une densité d'énergie qu'on appelle *spectrogramme*, et qu'on note P_S:

$$P_Sf(u,\xi) = |Sf(u,\xi)|^2 = \left|\int_{-\infty}^{+\infty} f(t)\,g(t-u)\,\mathrm{e}^{-i\xi t}\,dt\right|^2. \tag{6.10}$$

Il mesure l'énergie de f et de $\hat{f}$ dans le voisinage temps-fréquence ou l'énergie de $g_{u,\xi}$ est concentrée.

Exemples

1. Une sinusoïde $f(t)=\mathrm{e}^{i\xi_0 t}$, dont la transformée de Fourier est le Dirac $\hat{f}(\omega)=2\pi\delta(\omega-\xi_0)$ a pour transformée de Fourier à fenêtre

   $$Sf(u,\xi) = \mathrm{e}^{-iu(\xi-\xi_0)}\,\hat{g}(\xi-\xi_0).$$

   Son énergie est répartie sur l'intervalle fréquentiel $[\xi_0-\frac{\sigma_\omega}{2},\xi_0+\frac{\sigma_\omega}{2}]$.

2. La transformée de Fourier fenêtrée d'un Dirac $f(t)=\delta(t-u_0)$ vaut

   $$Sf(u,\xi) = \mathrm{e}^{-i\xi u_0}\,g(u_0-u).$$

   Son énergie est localisée dans l'intervalle temporel $[u_0-\frac{\sigma_t}{2},u_0+\frac{\sigma_t}{2}]$.

3. Dans la figure 6.2, on voit le spectrogramme d'un signal qui a une composantes dont la "fréquence instantanée" augmente linéairement dans le temps (chirp linéaire) et une seconde composante dont la fréquence décroit de façon quadratique dans le temps (chirp quadratique). S'ajoute à cela deux gaussiennes modulées. On a calculé le spectrogramme avec une fenêtre gaussienne dilatée d'un facteur $s = 0,05$. Le chirp linéaire a des coefficients de grande amplitude le long de la trajectoire de sa fréquence instantanée. Le chirp quadratique donne des grands coefficients le long d'une parabole. Les deux gaussiennes modulées donnent deux taches fréquentielles à haute et basse fréquence, en $u = 0,5$ et $u = 0,87$.

4. La figure 6.3 montre le spectrogramme du son "greasy" dont le graphique est donné au-dessus. L'amplitude de $|Sf(u,\xi)|^2$ est d'autant plus grande que l'image du spectrogramme est sombre. Le "ea" tout comme le "y" sont des sons voisés dont les formants apparaissent clairement sur le spectrogramme. Le "s" est un son non-voisé dont l'énergie est diffusé en hautes fréquences.

Complétude et stabilité Lorsque les coordonnées temps-fréquence (u,ξ) parcourent $\mathbb{R}^2$, les boîtes de Heisenberg des atomes $g_{u,\xi}$ recouvrent tout le plan temps-fréquence. On peut donc s'attendre à pouvoir reconstituer f à partir de sa transformée de Fourier à fenêtre $Sf(u,\xi)$. Le théorème suivant nous fournit une formule de reconstruction et montre qu'on a conservation de l'énergie.

Théorème 6.2 *Si* $f \in \mathbf{L^2}(\mathbb{R})$ *alors*

$$f(t) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi)\, g(t-u)\, \mathrm{e}^{i\xi t}\, d\xi\, du \tag{6.11}$$

et

$$\int_{-\infty}^{+\infty} |f(t)|^2\, dt = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |Sf(u,\xi)|^2\, d\xi\, du. \tag{6.12}$$

Démonstration On commence par la preuve de la formule de reconstruction (6.11). Appliquons la formule de Fourier Parseval (2.29) à l'intégrale (6.11) en la variable u. On calcule la transformée de Fourier de $f_\xi(u) = Sf(u,\xi)$ en u en remarquant que

$$Sf(u,\xi) = \mathrm{e}^{-iu\xi} \int_{-\infty}^{+\infty} f(t)\, g(t-u)\, \mathrm{e}^{i\xi(u-t)}\, dt = \mathrm{e}^{-iu\xi}\, f \star g_\xi(u), \tag{6.13}$$

avec $g_\xi(t) = g(t)\mathrm{e}^{i\xi t}$, car $g(t) = g(-t)$. Sa transformée de Fourier vaut donc

$$\hat{f}_\xi(\omega) = \hat{f}(\omega+\xi)\, \hat{g}_\xi(\omega+\xi) = \hat{f}(\omega+\xi)\, \hat{g}(\omega).$$

La transformée de Fourier de $g(t-u)$ en u vaut $\hat{g}(\omega)\mathrm{e}^{-it\omega}$. On a donc

$$\begin{aligned} \frac{1}{2\pi} \left(\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} Sf(u,\xi)\, g(t-u)\, \mathrm{e}^{i\xi t}\, du \right) d\xi \quad &= \\ \frac{1}{2\pi} \int_{-\infty}^{+\infty} \left(\frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega+\xi)\, |\hat{g}(\omega)|^2\, \mathrm{e}^{it(\omega+\xi)}\, d\omega \right) d\xi \quad &. \end{aligned}$$

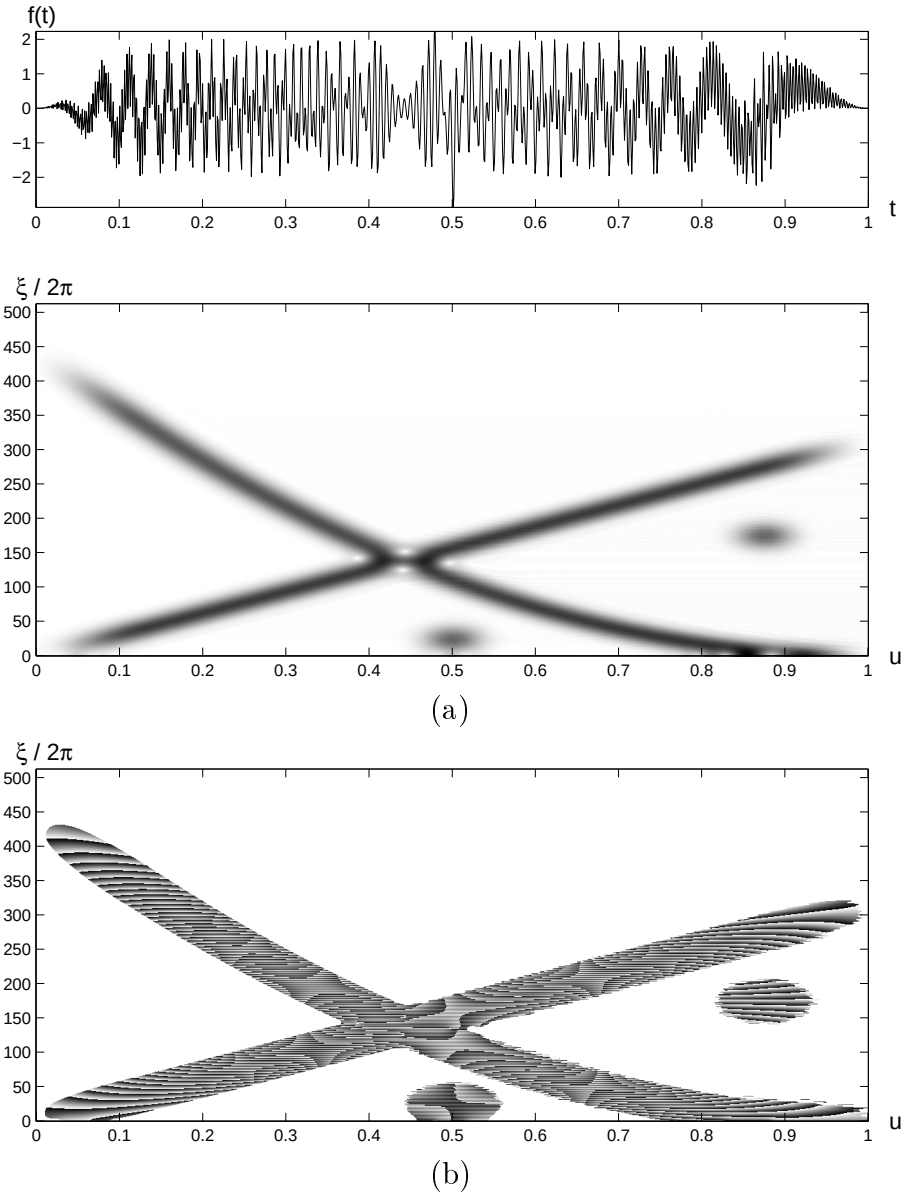

Figure 6.2: Le signal comprend un chirp linéaire de fréquence croissante, un chirp quadratique de fréquence décroissante, et deux gaussiennes modulées situées en $t = 0,5$ et $t = 0,87$. (a) Spectrogramme $P_S f(u, \xi)$. Les axes horizontaux et verticaux correspondent respectivement au temps u et à la fréquence ξ. Les points sombres correspondent à des coefficients de grande amplitude. (b) Phase complexe de $Sf(u, \xi)$ dans les régions où le module de $P_S f(u, \xi)$ est non nul.

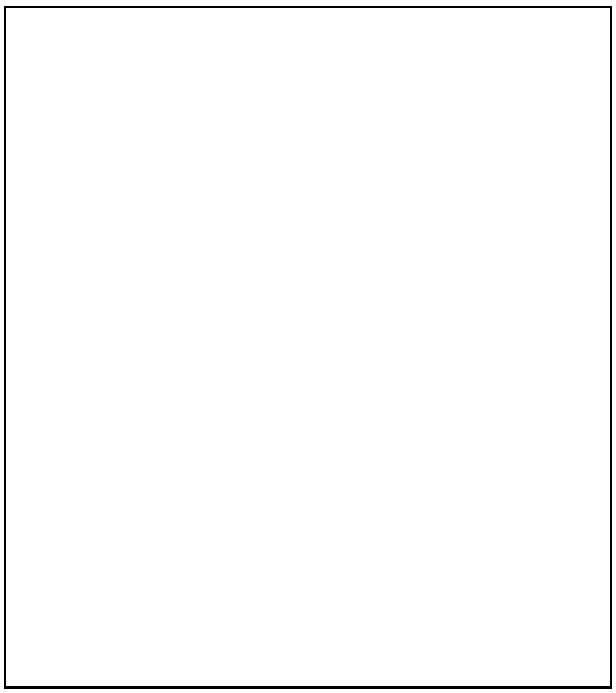

Figure 6.3: Le graphique du dessus correspond au son “greasy” enregistré à 8kHz. Son spectrogramme $|Sf(u,\xi)|^2$ est montré au-dessous, dans le plan temps-fréquence.

Si on peut appliquer le théorème de Fubini pour changer l'ordre d'intégration. Le théorème de transformée de Fourier inverse montre que

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

Comme $\frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{g}(\omega)|^2\,d\omega = 1$, on en déduit (6.11). Si $\hat{f} \notin \mathbf{L^1}(\mathbb{R})$, on démontre la formule à partir de là grâce à un argument de densité.

Occupons nous maintenant de la conservation de l'énergie (6.12). Comme la transformée de Fourier en u de $Sf(u,\xi)$ est $\hat{f}(\omega+\xi)\,\hat{g}(\omega)$, on obtient, en appliquant la formule de Plancherel au membre de droite de (6.12):

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} |Sf(u,\xi)|^2\,du\,d\xi = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{2\pi}\int_{-\infty}^{+\infty} |\hat{f}(\omega+\xi)\,\hat{g}(\omega)|^2\,d\omega\,d\xi.$$

On peut appliquer le théorème de Fubini, et la formule de Plancherel montre que

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

ce qui implique (6.12). □

On peut réécrire la formule de reconstruction (6.11) sous la forme

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \langle f, g_{u,\xi}\rangle\, g_{u,\xi}(t)\,d\xi\,du. \tag{6.14}$$

Cette formule ressemble à celle d'une décomposition sur une base orthogonale, mais ce n'est pas le cas, car la famille $\{g_{u,\xi}\}_{u,\xi\in\mathbb{R}^2}$ est largement redondante dans $\mathbf{L^2}(\mathbb{R})$. La seconde identité (6.12) justifie qu'on interprète le spectrogramme $P_Sf(u,\xi) = |Sf(u,\xi)|^2$ comme une densité d'énergie, car son intégrale en temps-fréquence est égale à l'énergie du signal.

Discrétisation La discrétisation et le calcul rapide de la transformée de Fourier à fenêtre relève des mêmes idées que la discrétisation de la transformée de Fourier classique, décrite précédemment dans le paragraphe 3.4. On considère des signaux discrets de période N. On prend comme fenêtre $g[n]$ un signal discret symétrique et de période N et de norme unité $\|g\| = 1$. On définit les atomes de Fourier fenêtrés discrets par

$$g_{m,l}[n] = g[n-m]\,\mathrm{e}^{\frac{i2\pi ln}{N}}.$$

La transformée de Fourier discrète $g_{m,l}$ a pour valeurs

$$\hat{g}_{m,l}[k] = \hat{g}[k-l]\,\mathrm{e}^{\frac{-i2\pi m(k-l)}{N}}.$$

La transformée de Fourier fenêtrée discrète d'un signal de période N est

$$Sf[m,l] = \langle f, g_{m,l}\rangle = \sum_{n=0}^{N-1} f[n]\,g[n-m]\,\mathrm{e}^{\frac{-i2\pi ln}{N}}, \tag{6.15}$$

Pour chaque $0 \leq m < N$, on calcule $Sf[m,l]$ pour $0 \leq l < N$ par transformée de Fourier discrète sur $f[n]g[n-m]$. On réalise ce calcul au moyen de N FFT de taille N, ce qui donne un total de $O(N^2 \log_2 N)$ opérations. C'est cet algorithme qui a servi pour le calcul des figures 6.2 et 6.3.

6.2 Fréquence instantanée

Dans un morceau de musique, on distingue plusieurs fréquences variant dans le temps. Il reste à définire la notion de fréquence instantanée. Afin d'estimer plusieurs fréquences instantanées, on sépare les composantes fréquentielles à l'aide d'une transformée de résolution suffisante en fréquence, mais également en temps, de manière à pouvoir réaliser des mesures variables dans le temps. On étudie la mesure des fréquences instantanées par des transformées de Fourier fenêtrées.

6.2.1 Fréquence instantanée analytique

Un cosinus modulé

$$f(t) = a\cos(w_0 t + \phi_0) = a\cos\phi(t)$$

a une fréquence ω_0 égale à la dérivée de la phase $\phi(t) = w_0 t + \phi_0$. Afin de généraliser cette notion, on écrit les signaux réels f comme ayant une amplitude a et un phase ϕ variant dans le temps:

$$f(t) = a(t)\,\cos\phi(t) \quad \text{avec} \quad a(t) \geq 0\ . \tag{6.16}$$

On définit *la fréquence instantanée* comme la dérivée de la phase:

$$\omega(t) = \phi'(t) \geq 0\ .$$

On peut se ramener à une dérivée positive en jouant sur le signe de $\phi(t)$. Il convient néanmoins d'être prudent car il existe de nombreuses valeurs pour $a(t)$ et de $\phi(t)$, et $\omega(t)$ n'est donc pas défini de manière unique pour un f donné.

On obtient une décompostion particulière de type (6.16) en calculant la partie analytique f_a de f, dont la transformée de Fourier est définie par

$$\hat{f}_a(\omega) = \begin{cases} 2\,\hat{f}(\omega) & \text{si } \omega \geq 0 \\ 0 & \text{si } \omega < 0 \end{cases}\ . \tag{6.17}$$

On dit que le signal complexe $f_a(t)$ est analytique car on peut démontrer qu'il a une extension analytique sur le demi plan complexe supérieur. Par ailleurs on peut verifier que $f = \mathrm{Re}[f_a]$ car $\hat{f}(\omega) = \frac{1}{2}(\hat{f}_a(\omega) + \hat{f}_a^*(-\omega))$.

On peut représenter f_a en séparant le module de la phase complexe

$$f_a(t) = a(t)\,\mathrm{e}^{i\phi(t)}\ .$$

On en déduit

$$f(t) = a(t)\,\cos\phi(t).$$

On dira que $a(t)$ est l'amplitude *analytique* de $f(t)$, et $\phi'(t)$ sa fréquence analytique instantanée; elles sont définies de manière unique.

Exemple

1. Si $f(t) = a(t)\,\cos(\omega_0 t + \phi_0)$, alors

$$\hat{f}(\omega) = \frac{1}{2}\left(\mathrm{e}^{i\phi_0}\,\hat{a}(\omega - \omega_0) + \mathrm{e}^{-i\phi_0}\,\hat{a}(\omega + \omega_0)\right).$$

Si les variations de $a(t)$ sont lentes en comparaison de la période $\frac{2\pi}{\omega_0}$, ce qu'on peut obtenir en forçant le support de $\hat{a}(\omega)$ à être dans $[-\omega_0, \omega_0]$, alors

$$\hat{f}_a(\omega) = \mathrm{e}^{i\phi_0}\, \hat{a}(\omega - \omega_0)$$

d'où $f_a(t) = a(t)\, \mathrm{e}^{i(\omega_0 t + \phi_0)}$.

Si f est un signal constitué de la somme de deux sinusoïdes:

$$f(t) = a\cos(\omega_1 t) + a\cos(\omega_2 t),$$

alors

$$f_a(t) = a\, \mathrm{e}^{i\omega_1 t} + a\, \mathrm{e}^{i\omega_2 t} = a\, \cos\left(\frac{\omega_1 - \omega_2}{2}\, t\right)\, \mathrm{e}^{i\frac{\omega_1+\omega_2}{2}t}.$$

La fréquence instantanée vaut $\phi'(t) = \frac{\omega_1+\omega_2}{2}$ et l'amplitude instanée est

$$a(t) = a\, \left|\cos\left(\frac{\omega_1 - \omega_2}{2}\, t\right)\right|.$$

Ce résultat n'est pas satisfaisant parce qu'il ne montre pas que le signal est composé de deux sinusoïdes de même amplitude. On a obtenu une fréquence moyenne. Nous allons expliquer dans la section qui suit comment mesurer les fréquences instantanées de plusieurs composantes spectrales en les séparant à l'aide de transformées de Fourier à fenêtre.

Modulation de fréquence En communications, on peut transmettre l'information à travers son amplitude $a(t)$ (modulation d'amplitude) ou sa fréquence instantanée $\phi'(t)$ (modulation de fréquence). La modulation de fréquence est plus robuste en présence de bruits blancs gaussiens additifs. De plus, elle résiste mieux aux interférences entre chemins multiples, qui détruisent l'information d'amplitude. Une modulation de fréquence envoie un message $m(t)$ à travers un signal

$$f(t) = a\cos\phi(t) \quad \text{with} \quad \phi'(t) = \omega_0 + k\, m(t).$$

La largeur de bande de f est proportionnelle à k. On ajuste cette constante en fonction des bruits de transmission et de la bande passante disponible. A la réception, on récupère le message $m(t)$ grâce à une démodulation de fréquence qui calcule la fréquence instantanée $\phi'(t)$.

Modèles de sons additifs On peux modéliser les sons musicaux et les phonèmes comme des somme de *partielles* sinusoïdales:

$$f(t) = \sum_{k=1}^{K} f_k(t) = \sum_{k=1}^{K} a_k(t)\cos\phi_k(t)\ , \tag{6.18}$$

où a_k et ϕ'_k sont lentement variables. De telles décompositions sont utilisées pour reconnaître des formes et modifier des sons. Le paragraphe 6.2.2 explique comment calculer

les a_k et les fréquences instantanées ϕ'_k de chaque partielle, dont on déduit la phase ϕ_k par intégration.

Pour comprimer de son f d'un facteur α dans le temps, et sans modifier les valeurs des ϕ'_k et des a_k, on synthétise

$$g(t) = \sum_{k=1}^{K} a_k(\alpha\, t)\, \cos\Big(\frac{1}{\alpha}\, \phi_k(\alpha\, t)\Big). \tag{6.19}$$

Les partielles de g en $t = \alpha\, t_0$ et les partielles de f en $t = t_0$ ont les mêmes amplitudes et les mêmes fréquences instantanées. Pour $\alpha > 1$, le son g est plus court que f tout en étant perçu comme ayant le même "contenu fréquentiel" que f.

On opère un déplacement fréquentiel en multipliant chaque phase par une constante α:

$$g(t) = \sum_{k=1}^{K} b_k(t)\, \cos\Big(\alpha\, \phi_k(t)\Big). \tag{6.20}$$

La fréquence instantanée de chaque partielle vaut maintenant $\alpha\, \phi'_k(t)$. Pour calculer les nouvelles amplitudes $b_k(t)$, on utilise un modèle de résonnance, qui suppose que ces amplitudes sont les échantillons d'une enveloppe fréquentielle lisse $F(t, \omega)$:

$$a_k(t) = F\Big(t, \phi'_k(t)\Big) \quad \text{et} \quad b_k(t) = F\Big(t, \alpha\, \phi'_k(t)\Big)\ .$$

En traitement de la parole, cette enveloppe est composé de plusieurs formants. Il est fonction du type de phonème qui a été prononcé. Comme $F(t, \omega)$ est une fonction régulière de ω, on calcule son amplitude en $\omega = \alpha\, \phi'_k(t)$ par interpolation des valeurs $a_k(t)$ correspondant à $\omega = \phi'_k(t)$.

6.2.2 Crêtes de transformée de Fourier à fenêtre

Le spectrogramme $P_S f(u, \xi) = |Sf(u, \xi)|^2$ mesure l'énergie de f dans un voisinage temps-fréquence de (u, ξ). L'algorithme de crêtes calcule les fréquences instantanées à partir des maxima locaux de $P_S f(u, \xi)$ [14].

On calcule la transformée de Fourier à fenêtre à l'aide d'une fenêtre symétrique $g(t) = g(-t)$ de support $[-\frac{s}{2}, \frac{s}{2}]$. La transformée de Fourier $\hat{g}$ de g est une fonction symétrique réelle avec $|\hat{g}(\omega)| \leq \hat{g}(0)$ pour tout $\omega \in \mathbb{R}$. On normalise g de manière à avoir $\|g\| = 1$. On définit la largeur de bande $\Delta\omega$ de $\hat{g}$ par

$$|\hat{g}(\omega)| \ll |\hat{g}(0)| \quad \text{pour} \quad |\omega| \geq \Delta\omega. \tag{6.21}$$

La valeur de $\Delta\omega$ est de l'ordre de $1/s$ si g est une fenètre réglière, car le support temporel de g est de taille s.

Les atomes de Fourier correspondants sont

$$g_{u,\xi}(t) = g(t - u)\, \mathrm{e}^{i\xi t}.$$

La transformée de Fourier à fenêtre est alors

$$Sf(u, \xi) \ = \ \int_{-\infty}^{+\infty} f(t)\, g(t - u)\, \mathrm{e}^{-i\xi t}\, dt. \tag{6.22}$$

Le théorème suivant exprime $Sf(u,\xi)$ en fonction de la fréquence instantanée de f. Les termes d'erreur de cette formule peuvent être controlés avec un dévelopement à un ordre supérieur [7].

Théorème 6.3 *Soit $f(t) = a(t)\cos\phi(t)$. Si les variations de $a(t)$ et $\phi'(t)$ sont négligeables sur le support $[u-s/2, u+s/2]$ de $g(t-u)$ et que $\phi'(u) \geq \Delta\omega$ alors pour tout $\xi \geq 0$ on a*

$$Sf(u,\xi) \approx \frac{1}{2}\, a(u)\, \mathrm{e}^{i(\phi(u)-\xi u)}\, \hat{g}\Big(\xi - \phi'(u)\Big). \tag{6.23}$$

Démonstration Remarquons que

$$\begin{aligned} Sf(u,\xi) &= \int_{-\infty}^{+\infty} a(t)\, \cos\phi(t)\, g(t-u)\, \mathrm{e}^{-i\xi t}\, dt \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t)\, \left(\mathrm{e}^{i\phi(t)} + \mathrm{e}^{-i\phi(t)}\right)\, g(t-u)\, \mathrm{e}^{-i\xi t}\, dt \\ &= I(\phi) + I(-\phi). \end{aligned}$$

Commençons par étudier

$$\begin{aligned} I(\phi) &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t)\, \mathrm{e}^{i\phi(t)}\, g(t-u)\, \mathrm{e}^{-i\xi t}\, dt \\ &= \frac{1}{2}\int_{-\infty}^{+\infty} a(t+u)\, \mathrm{e}^{i\phi(t+u)}\, g(t)\, \mathrm{e}^{-i\xi(t+u)}\, dt. \end{aligned}$$

Comme $a(t+u) \approx a(u)$ et $\phi(t+u) \approx \phi(u) + t\phi'(u)$ pour $|t| \leq s/2$ il vient

$$I(\phi) \approx \frac{a(u)\, \mathrm{e}^{i(\phi(u)-\xi u)}}{2} \int_{-\infty}^{+\infty} g(t)\, \mathrm{e}^{-it(\xi-\phi'(u))}\, dt$$

et donc

$$I(\phi) \approx \frac{1}{2}\, a(u)\, \mathrm{e}^{i(\phi(u)-\xi u)}\, \hat{g}\Big(\xi - \phi'(u)\Big). \tag{6.24}$$

De même

$$I(-\phi) \approx \frac{1}{2}\, a(u)\, \mathrm{e}^{i(-\phi(u)-\xi u)}\, \hat{g}\Big(\xi + \phi'(u)\Big).$$

Comme $\xi \geq 0$, $\xi + \phi'(u) \geq \Delta\omega$ donc $\hat{g}\Big(\xi + \phi'(u)\Big) \ll 1$. On peut donc négliger $I(-\phi)$ et l'on déduit (6.23) de (6.24). □

Points de crête Supposons que $a(t)$ et $\phi'(t)$ ont des variations negligeables sur les intervalles de taille s, et que $\phi'(t) \geq \Delta\omega$. Comme $|\hat{g}(\omega)|$ est maximal en $\omega = 0$, (6.23) montre que, pour chaque u, le spectrogramme $|Sf(u,\xi)|^2$ est maximal en $\xi(u) = \phi'(u)$. Les points correspondants $(u, \xi(u))$ du plan temps-fréquence sont appelés des *crêtes*. Aux points de crête, (6.23) devient

$$Sf(u,\xi) = \frac{1}{2}\, a(u)\, \mathrm{e}^{i(\phi(u)-\xi u)}\, \hat{g}(0) \tag{6.25}$$

La fréquence de crête donne la fréquence instantanée $\xi(u) = \phi'(u)$, et on calcule $\xi(u) = \phi'(u)$ avec

$$a(u) = \frac{2\left|Sf\Big(u, \xi(u)\Big)\right|}{|\hat{g}(0)|}. \tag{6.26}$$

Soit $\Phi_S(u,\xi)$ la phase complexe de $Sf(u,\xi)$. L'expression (6.25) montre que les points de crête sont également des points où la phase est stationnaire:

$$\frac{\partial \Phi_S(u,\xi)}{\partial u} = \phi'(u) - \xi = 0.$$

La vérification de la stationnarité de la phase permet de mieux situer les crêtes.

Fréquences multiples Lorsque le signal contient plusieurs lignes spectrales de fréquences suffisamment distinctes, la transformée de Fourier fenêtrée en sépare les diverses composantes, et les crêtes permettent de détecter l'évolution temporelle de chacune d'entre elles. Considérons

$$f(t) = a_1(t)\cos\phi_1(t) + a_2(t)\cos\phi_2(t),$$

où $a_k(t)$ et $\phi_k'(t)$ sont à variations petites sur les intervalles de largeur s, et où on suppose $\phi_k'(t) \geq \Delta\omega$. Comme la transformée de Fourier fenêtrée est linéaire, on applique (6.23) à chaque composante spectrale:

$$\begin{aligned} Sf(u,\xi) &= \frac{1}{2}\, a_1(u)\, \hat{g}\Big(\xi - \phi_1'(u)\Big)\, \mathrm{e}^{i(\phi_1(u) - \xi u)} \\ &\quad + \frac{1}{2}\, a_2(u)\, \hat{g}\Big(\xi - \phi_2'(u)\Big)\, \mathrm{e}^{i(\phi_2(u) - \xi u)}. \end{aligned} \tag{6.27}$$

On arrive à séparer les composantes spectrales si, pour tout u

$$|\hat{g}\Big(\phi_1'(u) - \phi_2'(u)\Big)| \ll |\hat{g}(0)|, \tag{6.28}$$

ce qui signifie que la différence entre les fréquences est plus grande que la largeur de bande de $\hat{g}(\omega)$:

$$|\phi_1'(u) - \phi_2'(u)| \geq \Delta\omega. \tag{6.29}$$

Dans ce cas, on peut négliger, pour $\xi = \phi_1'(u)$, le second terme de (6.27), et le premier terme engendre un point de crête permettant de reconstituer $\phi_1'(u)$ et $a_1(u)$ en utilisant (6.26). De même, on peut négliger le premier terme lorsque $\xi = \phi_2'(u)$, et on obtient un second point de crête caractérisant $\phi_2'(u)$ et $a_2(u)$. Les points de crête sont distribués sur les deux lignes temps-fréquence $\xi(u) = \phi_1'(u)$ et $\xi(u) = \phi_2'(u)$. Ce résultat demeure valable pour un nombre quelconque de composantes spectrales instationnaires, tant que la distance entre deux fréquences instantanées vérifie (6.29). Lorsque les lignes spectrales sont trop proches, elles créent des interférences qui détruisent la structure de crêtes.

En général, on ne connait pas le nombre de fréquences instantanées. On détecte alors tous les maxima locaux de $|Sf(u,\xi)|^2$ dont la phase est stationnaire $\frac{\partial \Phi_S(u,\xi)}{\partial u} = \phi'(u) - \xi = 0$. Ces points définissent des courbes dans le plan (u,ξ) qui sont les crêtes de la transformée de Fourier fenêtrée. On supprime souvent les crêtes de faible amplitude

$a(u)$ parce qu'elles peuvent être des artifacts provenant de variations bruitées, ou des "ombres" d'autres fréquences instantanées, dûes aux lobes latéraux de $\hat{g}(\omega)$.

Dans la figure 6.4, on peut voir les crêtes obtenues à partir du module et de la phase de la transformée de Fourier fenêtrée de la figure 6.2. Pour $t \in [0,4\,,\,0,5]$, les fréquences intantanées sont trop proches et la résolution en fréquence de la fenêtre ne permet pas de les séparer. En conséquence, les crêtes y détectent une fréquence instantanée moyenne.

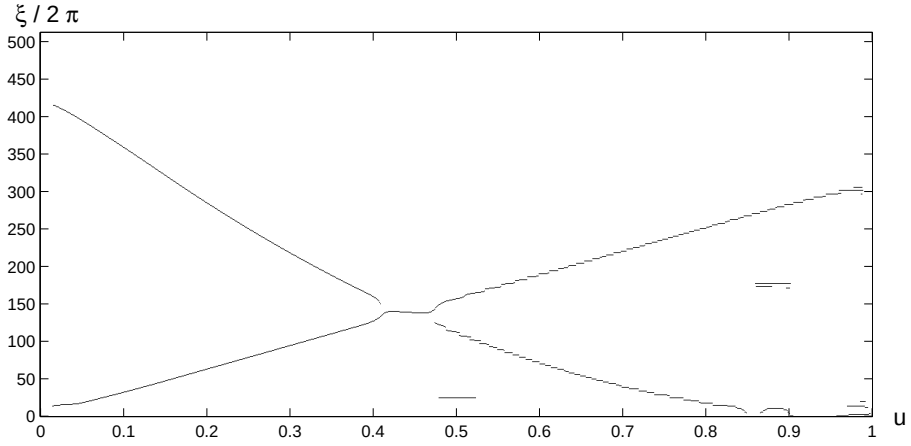

Figure 6.4: Crêtes de plus grande amplitude calculées à partir du spectrogramme de la figure 6.2. Ces crêtes donnent les fréquences instantanées des chirps linéaires et quadratiques, et des transitoires à basse et haute fréquence en $t = 0,5$ et $t = 0,87$.

Choix de la fenêtre La mesure des fréquences instantanées aux points de crête a été validée seulement lorsque la taille s de la fenêtre g est suffisamment petite pour que $a(t)$ et $\phi'(t)$ soient approximativement constant sur $[u - s/2, u + s/2]$. D'un autre côté, il faut que la largeur de bande $\Delta\omega \sim 1/s$ soit suffisamment petite pour séparer les composantes spectrales succesives en (6.29). Le choix de l'échelle s de la fenêtre doit donc réaliser un compromis entre ces deux contraintes.

Exemples

1. La somme de deux chirps linéaires parallèles

$$f(t) = a_1 \cos(bt^2 + ct) + a_2 \cos(bt^2) \tag{6.30}$$

a deux fréquences instantanées $\phi_1'(t) = 2bt + c$ et $\phi_2'(t) = 2bt$. On voit un exemple numérique en figure 6.5. La fenêtre g a une résolution fréquentielle suffisamment fine pour séparer les deux chirps si

$$|\phi_1'(t) - \phi_2'(t)| = |c| \geq \Delta\omega. \tag{6.31}$$

Son support temporel est assez fin comparé à la variation temporelle des chirps si la fréquence instantanée $\phi'(t)$ est quasiment constante sur $[u - s/2, u + s/2]$, ce que l'on obtient si

$$s^2\,|\phi_1''(u)| = s^2\,|\phi_2''(u)| = 2\,b\,s^2 \ll 1. \tag{6.32}$$

Les conditions (6.31) et (6.32) montrent qu'on peut trouver une fenêtre g adéquate si et seulement si

$$\frac{c}{\sqrt{b}} \gg s\,\Delta\omega \sim 1\ . \tag{6.33}$$

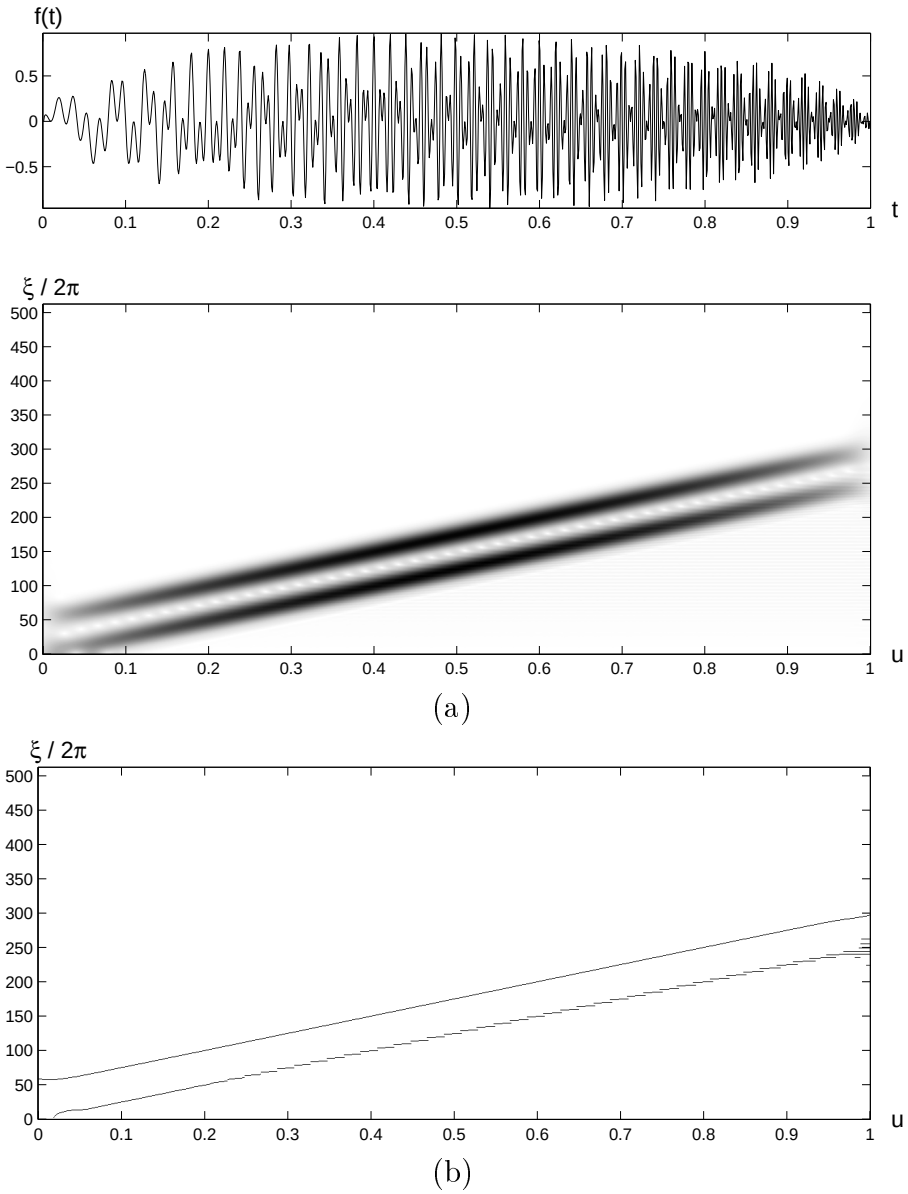

Figure 6.5: Somme de deux "chirps" linéaires prallèles. (a): Spectrogramme $P_S f(u,\xi) = |Sf(u,\xi)|^2$. (b): Crêtes calculées à partir du spectrogramme.

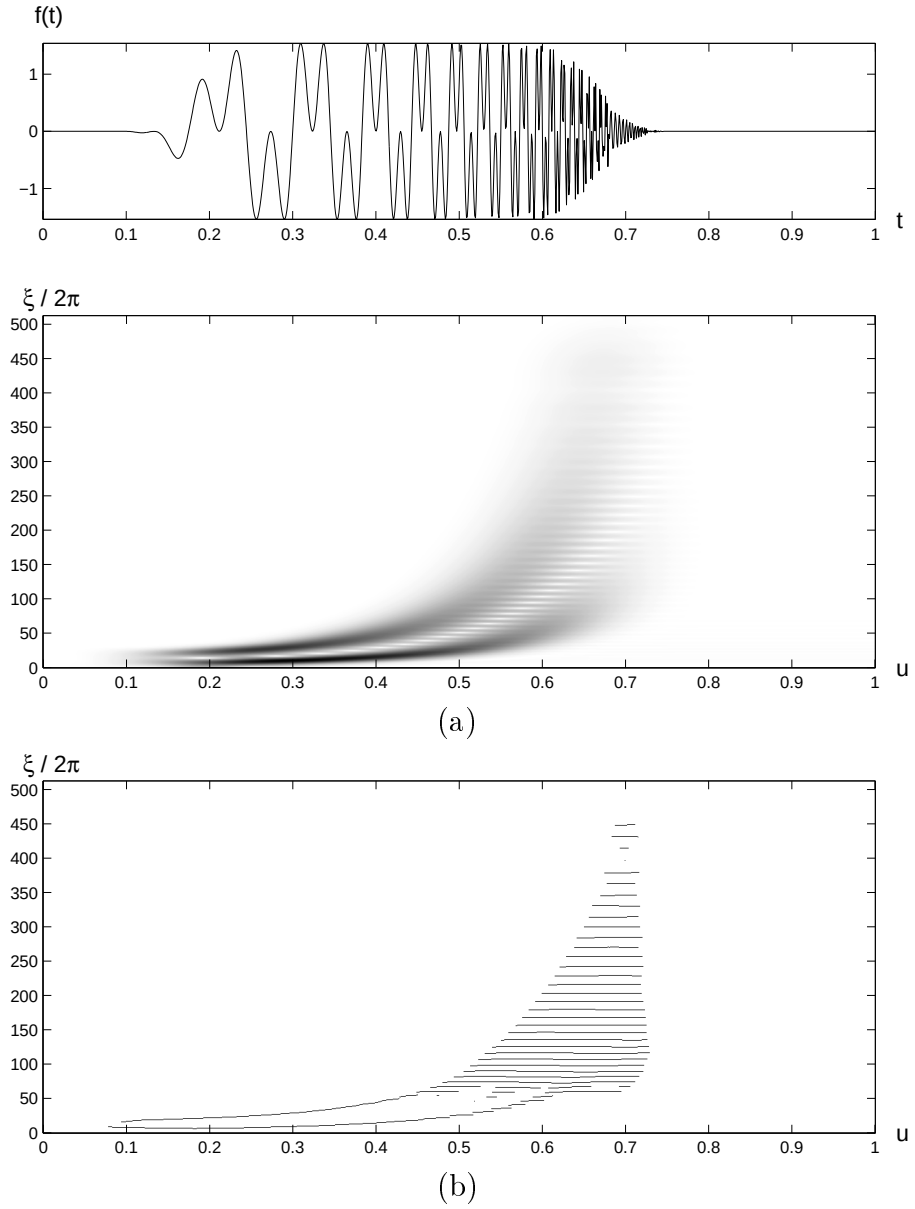

Figure 6.6: Somme de deux chirps hyperboliques. (a): Spectrogramme $P_S f(u, \xi)$. (b): Crêtes calculées à partir du spectrogramme.

Les chirps linéaires de la figure 6.5 vérifient (6.33). On a calculé leurs crêtes en utilisant une fenêtre gaussienne avec $s = 0,05$.

2. Le chirp hyperbolique

$$f(t) = \cos\left(\frac{\alpha}{\beta - t}\right)$$

pour $0 \leq t < \beta$ a une fréquence instantanée

$$\phi'(t) = \frac{\alpha}{(\beta - t)^2},$$

à variation rapide quand t est proche de β. Les fréquences instantanées des chirps hyperboliques vont de 0 à $+\infty$ en un temps fini. Cette propriété est particulièrement utile aux radars. Ces chirps sont également produits par les sonars de navigation des chauves-souris [14].

On ne peut pas estimer les fréquences instantanées des chirps hyperboliques par une transformée de Fourier fenêtrée parce que, pour une taille de fenêtre donnée, la fréquence instantanée varie trop rapidement dans les hautes fréquences. Lorsque u est assez proche de β, on ne peut considérer que la fréquence instantanée $\phi'(t)$ est constante sur le support $[u - s/2, u + s/2]$ de $g(t - u)$ car

$$s^2|\phi''(u)| = \frac{s^2\alpha}{(\beta - u)^3} > 1.$$

En figure 6.6, on voit un signal composé de la somme de deux chirps hyperboliques:

$$f(t) = a_1 \cos\left(\frac{\alpha_1}{\beta_1 - t}\right) + a_2 \cos\left(\frac{\alpha_2}{\beta_2 - t}\right), \tag{6.34}$$

avec $\beta_1 = 0,68$ et $\beta_2 = 0,72$. Au début du signal, les deux chirps ont des fréquences instantanées voisines qui sont séparées par la transformée de Fourier fenêtrée, qu'on a calculée avec une fenêtre large. Quand on se rapproche de β_1 ou de β_2, la fréquence instantanée varie trop rapidement en regard de la taille de la fenêtre. Les crêtes correspondantes ne permettent plus de suivre ces fréquences instantanées.

Chapitre 7

Information et Codage

La complexité d'une suite de symboles peut se mesurer par la taille minimum d'un code permettant de reconstruire cette suite. La théorie de l'information de Shannon montre que le nombre de bit moyen pour coder chaque symbole dépend de l'entropie du processus aléatoire sous-jacent. Pour coder des suites de nombre réels, il est necessaire de les approximer avec une quantification avant d'effectuer un codage entropique. L'optimisation de cette quantification est étudiée. Ces résultats donnent les bases mathématiques et algorithmiques permettant de comprimer des signaux audios ou des images.

7.1 Complexité et entropie

La théorie de l'information définie la complexité d'une série numérique en évaluant la taille des codes permettant de reproduire cette série. Les fondations de cette théorie sont mises en place par Shannon en 1948, qui modélise des séries numériques comme des réalisations d'un processus aléatoire. Il démontre alors l'existence d'une complexité intrinsèque associée à tout processus aléatoire, qu'il appelle *entropie*. En 1965, Kolmogorov introduit une définition plus générale de la complexité d'une série numérique, comme étant la longueur minimum du programme binaire permettant de reproduire cette série avec un ordinateur. Le modèle d'ordinateur est une machine de Turing ayant un nombre fini d'états. Cette définition n'a pas recours à un modèle probabiliste mais est plus délicate à manipuler mathématiquement. Nous suivront donc ici l'approche de Shannon qui donne des résultats suffisament précis pour la plupart des problèmes de traitement du signal.

7.1.1 Suites typiques

Considérons des suites de symboles de taille n prenant leurs valeurs dans un alphabet $A = \{a_k\}_{1 \leq k \leq K}$ de taille K. L'approche probabiliste de Shannon modélise ces séquences de symboles comme étant les valeurs prises par des variables aléatoires $X_1\, X_2 \ldots X_n$. Pour simplifier l'analyse, nous nous placerons dans le cas le plus simple où les X_i sont des variables aléatoires indépendantes et de même loi. On note

$$p(a_k) = \Pr\{X_i = a_k\}.$$

Comme les variables X_i sont indépendantes, la probabilité d'une suite de valeurs est:

$$p(x_1, \ldots, x_n) = \Pr\{X_1 = x_1, \ldots, X_n = x_n\} = \prod_{k=1}^{n} \Pr\{X_k = x_k\} = \prod_{k=1}^{n} p(x_k).$$

On peut définir $p(X_1, .. X_n) = \prod_{k=1}^{n} p(X_k)$ qui est une variable aléatoire donnant la probabilité d'une suite de valeurs tirée au hasard. Le théorème suivant montre que pour n fixé et suffisament grand, alors pour la plupart des tirages, le log de cette probabilité est presque constante et égale à l'entropie

$$H = -\sum_{k=1}^{K} p(a_k) \log_2 p(a_k) = -E\{\log_2 p(X_i)\}.$$

L'entropie H peut s'interpréter comme l'incertitude moyenne sur les valeurs que prennent les variables aléatoire X_i. On peut vérifier que

$$0 \leq H \leq \log_2 K.$$

L'entropie est maximum, $H = \log_2 K$, si $p(a_k) = \frac{1}{K}$ pour $1 \leq k \leq K$. Il y a en effet une incertitude maximum sur les valeurs prises par X_i. L'entropie est minimum, $H = 0$, si l'un des symboles a_k a une probabilité 1. On connaît alors à l'avance la valeur de X_i.

Théorème 7.1 *Si les X_i sont des variables aléatoires indépendantes et de même probabilité $p(x)$ alors*

$$-\frac{1}{n} \log_2 p(X_1, ..., X_n) \quad \textit{tend vers } H \textit{ avec une probabilité } 1$$

lorsque n tend vers $+\infty$.

Démonstration On calcule

$$-\frac{1}{n} \log_2 p(X_1, \ldots, X_n) = -\frac{1}{n} \sum_{i=1}^{n} \log_2 p(X_i) \ .$$

Comme les X_i sont indépendants, les $\log_2 p(X_i)$ sont aussi des variables aléatoires indépendentes. En appliquant la loi forte des grands nombres [8] on démontre que $-\frac{1}{n} \sum_{i=1}^{n} \log_2 p(X_i)$ tend vers $-E\{\log p(X_i)\} = H$ lorsque n tend vers $+\infty$, avec probabilité 1. □

Bien qu'a priori $X_1, \ldots, X_n$ puisse prendre des valeurs quelconques dans l'ensemble A^n des vecteurs de symboles de taille n, ce théorème permet de montrer qu'il y a une probabilité presque 1 pour que ce vecteur soit une suite typique appartenant à un ensemble beaucoup plus petit. On appelle *ensemble typique* T_ϵ^n relativement à $p(x)$ l'ensemble des suites $(x_1, ..., x_n) \in A^n$ telles que

$$2^{-n(H+\epsilon)} \leq p(x_1, ..., x_n) \leq 2^{-n(H-\epsilon)}. \tag{7.1}$$

On note $|T_\epsilon^n|$ le cardinal de T_ϵ^n. Le théorème suivant montre que $|T_\epsilon^n|$ est de l'ordre de 2^{nH}, et que toutes les suites typiques ont une probabilité presque égale à 2^{-nH}.

Proposition 7.1 (Ensembles typiques) *Soit $\epsilon > 0$.*
1. Si $(x_1, ..., x_n) \in T_\epsilon^n$ alors

$$H - \epsilon \leq -\frac{1}{n} \log_2 p(x_1, ..., x_n) \leq H + \epsilon. \tag{7.2}$$

2. Lorsque n est suffisament grand

$$\Pr\{(X_1, ..., X_n) \in T_\epsilon^n\} > 1 - \epsilon. \tag{7.3}$$

3. Lorsque n est suffisament grand

$$2^{n(H-\epsilon)} \leq |T_\epsilon^n| \leq 2^{n(H+\epsilon)}. \tag{7.4}$$

Démonstration La propriété (7.2) est une conséquence directe de la définition (7.1) de T_ϵ^n.

L'inégalité (7.3) se déduit du théorème 7.1 qui montre que pour tout $\epsilon > 0$ et $\delta > 0$ il existe n_0 tel que pour tout $n \geq n_0$

$$\Pr\left\{\left|-\frac{1}{n}\log_2 p(X_1, ..., X_n) - H\right| < \epsilon\right\} > 1 - \delta.$$

En prenant $\delta = \epsilon$ on obtient (7.3).

On note $\vec{x} = (x_1, ..., x_n)$,

$$\begin{aligned} 1 &= \sum_{\vec{x} \in A^n} p(\vec{x}) \geq \sum_{\vec{x} \in T_\epsilon^n} p(\vec{x}) \\ &\geq \sum_{\vec{x} \in T_\epsilon^n} 2^{-n(H+\epsilon)} = |T_\epsilon^n|\, 2^{-n(H+\epsilon)}, \end{aligned}$$

ce qui démontre l'inégalité (7.4) à droite.

Lorsque n est suffisament grand, on a montré en (7.3) que

$$\begin{aligned} 1 - \epsilon &< \Pr\{(X_1, ..., X_n) \in T_\epsilon^n\} \\ &\leq \sum_{x \in T_\epsilon^n} 2^{-n(H-\epsilon)} = |T_\epsilon^n|\, 2^{-n(H-\epsilon)}, \end{aligned}$$

ce qui démontre l'inégalité (7.4) à gauche. □

Codage On peut effectuer un codage "ϵ-typique" des valeurs de $X_1 ..., X_n$ qui utilise des mots binaires plus courts pour coder les séquences typiques qui sont les plus probables. Comme il y a moins de $2^{n(H+\epsilon)}$ éléments dans T_ϵ^n, ces éléments peuvent être indéxés par des mots binaires de $\lfloor n(H+\epsilon) \rfloor + 1$ bits, où $\lfloor x \rfloor$ est le plus grand entier inférieur à x. Comme il y a K^n élements dans A, les éléments qui n'appartiennent pas à T_ϵ^n peuvent être indéxés par des mots binaires de $\lfloor \log_2 K \rfloor + 1$ bits. Afin de savoir si $x \in T_\epsilon^n$ on rajoute un 0 au début de son code binaire, qui est donc de longueur $\lfloor n(H+\epsilon) \rfloor + 2$. Si $x \notin T_\epsilon^n$ on rajoute un 1 au début de son code binaire, dont la taille est donc $\lfloor \log_2 K \rfloor + 2$. On note R le nombre moyen de bits pour coder chaque symbole d'une sequence $X_1, ..., X_n$.

Théorème 7.2 *Il existe $C > 0$ tel que pour tout $\epsilon > 0$, et n suffisament grand, le nombre moyen R de bits par symbole d'un codage ϵ-typique satisfait*

$$R \leq H + C\,\epsilon.$$

Démonstration On note $\vec{X} = (X_1, ..., X_n)$, $\vec{x} = (x_1, ..., x_n)$. Soit $l(x_i)$ la longueur du mot binaire utilisé par un code typique pour coder x_i. Le nombre total de bits pour coder $\vec{x}$ est

$$l(\vec{x}) = \sum_{i=1}^{n} l(x_i)\ .$$

Le nombre total moyen de bits par symbole est donc

$$\begin{aligned} n\,R &= E\{l(\vec{X})\} = \sum_{\vec{x}\in A^n} l(\vec{x})\,p(\vec{x}) = \sum_{\vec{x}\in T_\epsilon^n} l(\vec{x})\,p(\vec{x}) + \sum_{\vec{x}\notin T_\epsilon^n} l(\vec{x})\,p(\vec{x}) \\ &\leq \Pr\{\vec{X}\in T_\epsilon^n\}\left(\lfloor n(H+\epsilon)\rfloor + 2\right) + \Pr\{\vec{X}\notin T_\epsilon^n\}\left(\lfloor n\log_2 K\rfloor + 2\right) \\ &\leq n(H+\epsilon) + 2 + \epsilon(n\log_2 K + 2) \leq n\,H + n\,C\,\epsilon \end{aligned}$$

avec $C = 5 + \log_2 K$ pour $n \geq 1/\epsilon$. □

Ce théorème démontre que l'on peut constuire un code dont le nombre de bit moyen par pixel est arbitrairement près de l'entropie H. Par ailleurs, on peut montrer que tout code nécessite un nombre moyen de bit par symbole $R \geq H$. Le paragraphe suivant démontre ce résultat pour les codes par blocs.

7.1.2 Codage entropique

Nous considerons dans un premier temps les codes instantanés, qui définissent un code binaire w_k pour chaque symbole a_k de l'alphabet A. Cela permet de décoder symbole par symbole toute séquence $x_1, ..., x_n$. Si $\log_2 K$ est un entier, chaque symbole a_k peut être codé par un mot binaire de $\lfloor \log_2 K\rfloor + 1$ bits. Ce code peut cependant être amélioré en utilisant des mots binaires plus courts pour des symboles qui apparaissent plus souvent.

Soit l_k la longeur du code binaire w_k associée à a_k. Le nombre moyen de bits nécessaires pour coder les symboles d'une suite de variables aléatoires $X_1 \ldots X_n$ de même probabilité $p(x)$ est

$$R = \sum_{k=1}^{K} l_k\,p(a_k). \tag{7.5}$$

Le but est de trouver un code instantané qui soit décodable et qui minimise R.

Condition de préfixe Un code instané n'est pas toujours uniquement décodable. Par exemple, le code qui associe à $\{a_k\}_{1\leq k\leq 4}$ les mots binaires

$$\{w_1 = 0\ ,\ w_2 = 10\ ,\ w_3 = 110\ ,\ w_4 = 101\}$$

n'est pas décodable de façon unique. La suite 1010 peut soit correspondre à $w_2\ w_2$ ou à $w_4\ w_1$. La condition de préfixe impose qu'aucun mot binaire n'est le début d'un autre mot

binaire. Cette condition est clairement nécessaire et suffisante pour garantir que toute suite de mots binaires se décode de façon unique. Dans l'exemple précédent, w_2 est le préfixe de w_4. Le code suivant

$$\{w_1 = 0\ ,\ w_2 = 10\ ,\ w_3 = 110\ ,\ w_4 = 111\}$$

satisfait la condition de préfixe.

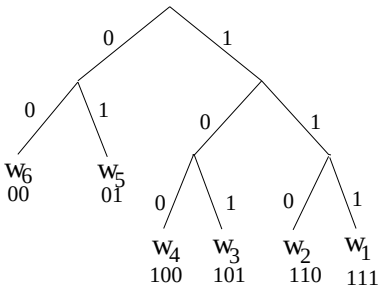

Figure 7.1: Arbre binaire d'un code de 6 symboles, qui satisfait la condition de préfixe. Le mot binaire w_k de chaque feuille est indiqué au-dessous.

Un code qui satisfait la condition de préfixe peut être associé à un arbre binaire, dont les K feuilles correspondent aux symboles $\{a_k\}_{1\leq k\leq K}$. Cette représentation est utile pour construire le code qui minimise le nombre de bits moyen R. Les branches de gauche et de droite de l'arbre binaire sont respectivement codées par 0 et 1. La figure 7.1 montre un exemple pour un code de 6 symboles. Le mot binaire w_k associé au symbole a_k est la succession de 0 et de 1 correspondant aux branches de gauche et de droite, le long du chemin de la racine de l'arbre à la feuille correspondant à a_k. Le code binaire généré par un tel arbre satisfait toujours la condition de préfixe. En effet, w_m est un préfixe de w_k si et seulement si a_m est un ancêtre de a_k dans l'arbre binaire. Ceci n'est pas possible puisque les deux symboles correspondent à des feuilles de l'arbre. Inversement, tout code préfixe peut être représenté par un tel arbre binaire. La longueur l_k du mot binaire w_k est la profondeur de la feuille a_k dans l'arbre binaire. L'optimisation d'un code de préfixe est donc équivalente à la construction d'un arbre binaire optimal qui distribue les profondeur des feuilles de façon à minimiser (7.5).

Entropie de Shannon Le théorème de Shannon prouve que le nombre moyen de bit R par symbole est plus grand que l'entropie.

Théorème 7.3 (Shannon) *On suppose que les symboles $\{a_k\}_{1\leq k\leq K}$ apparaissent avec la distribution de probabilité $\{p(a_k)\}_{1\leq k\leq K}$. Le nombre moyen R de bit d'un code ayant la propriété du préfixe satisfait*

$$R \geq H = -\sum_{k=1}^{K} p(a_k)\log_2 p(a_k). \tag{7.6}$$

Il existe un code ayant la propriété du préfixe tel que

$$R \leq H + 1. \tag{7.7}$$

Démonstration Ce théorème de Shannon se démontre à partir de l'inégalité de Kraft.

Lemme 7.1 (Inégalité de Kraft) *Tout code ayant la propriété du préfixe satisfait*

$$\sum_{k=1}^{K} 2^{-l_k} \leq 1. \tag{7.8}$$

Inversement, si $\{l_k\}_{1\leq k\leq K}$ sont des entiers positifs tels que l'inégalité (7.8) est satisfaite alors il existe un code de mots binaires $\{w_k\}_{1\leq k\leq K}$ de longueurs $\{l_k\}_{1\leq k\leq K}$ et qui satisfait la condition de préfixe.

Pour démontrer (7.8) on associe un arbre binaire au code considéré. Chaque l_k correspond à un noeud de l'arbre à une profondeur l_k qui dépend du mot binaire w_k. Soit

$$m = \max\{l_1, l_2, ..., l_K\}. \tag{7.9}$$

On considére l'arbre binaire complet dont toutes les feuilles sont à la profondeur m. On note T_k le sous-arbre issu du noeud correspondant au mot binaire w_k. Ce sous arbre a une profondeur $m - l_k$ et contient donc 2^{m-l_k} noeud au niveau m, comme l'illustre la figure 7.2. Comme il y a 2^m noeud à la profondeur m de l'arbre binaire complet et que la propriété du préfixe implique que tous les sous arbres $T_1, ..., T_K$ sont distincts, on déduit que

$$\sum_{k=1}^{K} 2^{m-l_k} \leq 2^m,$$

d'où (7.8).

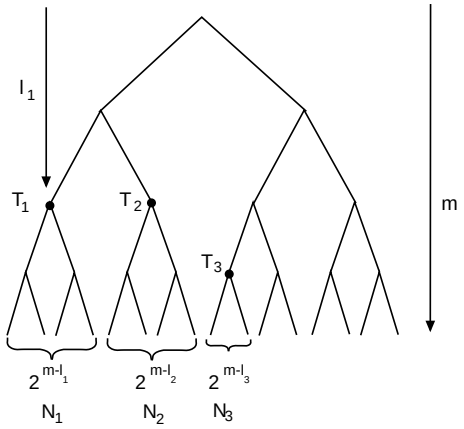

Figure 7.2: Disposition des sous arbres T_i dont la racine est à la profondeur l_i dans l'arbre d'un code de préfixe.

Inversement, on considère $\{l_k\}_{1\leq k\leq K}$ satisfaisant (7.8) avec $l_1 \leq l_2 \leq ... \leq l_K$ et $m = \max\{l_1, l_2, ..., l_K\}$. On définit les ensembles N_1 des 2^{m-l_1} premiers noeud au niveau m sur la gauche de l'arbre, puis N_2 l'ensemble des 2^{m-l_2} noeuds suivants et ainsi de suite comme l'indique la figure 7.2. Les noeuds des ensembles N_k sont les noeuds terminaux

de sous-arbres T_k qui sont disjoints. On associe à la racine de l'arbre T_k qui est à la profondeur l_k le mot binaire w_k. Cela définit un code qui satisfait la condition du préfixe où chaque mot a la longueur l_k voulue. Cela termine la démonstration du lemme.

Pour démontrer les deux inégalités (7.6) et (7.7) du théorème, on considère la minimisation de

$$R = \sum_{k=1}^{K} p(a_k)\, l_k$$

sous la contrainte de Kraft

$$\sum_{k=1}^{K} 2^{-l_k} \leq 1.$$

Dans un premier temps, nous supposons que l_k peut être un réel quelconque. Le minimum se calcule en utilisant un multiplicateur de Lagrange λ et en minimisant

$$J = \sum_{k=1}^{K} p(a_k) l_k + \lambda \sum_{k=1}^{K} 2^{-l_k}.$$

L'annulation de la dérivee par rapport à l_k donne

$$\frac{\partial J}{\partial l_k} = p(a_k) - \lambda\, 2^{-l_k}\, \log_{\mathrm{e}} 2 = 0.$$

Le minimum est obtenu pour $\sum_{k=1}^{K} 2^{-l_k} = 1$ et comme $\sum_{k=1}^{K} p(a_k) = 1$ on obtient $\lambda = 1/\log_{\mathrm{e}} 2$. La longueur optimale minimisant R est donc

$$l_k = -\log_2 p(a_k),$$

et

$$R = \sum_{k=1}^{K} p(a_k)\, l_k = -\sum_{k=1}^{K} p(a_k) \log_2 p(a_k) = H.$$

Pour garantir que l_k est entier, on choisit

$$l_k = \lceil -\log_2 p(a_k) \rceil$$

où $\lceil x \rceil$ est la plus petite valeur entière supérieure à x. Cela correspond au code de Shannon. Comme $l_k \geq -\log_2 p(a_k)$, l'inégalité de Kraft est satisfaite puisque

$$\sum_{k=1}^{K} 2^{-l_k} \leq \sum_{k=1}^{K} 2^{\log_2 p(a_k)} = 1.$$

Il existe donc un code préfixe dont les mots de code ont une longueur l_k. Pour ce code

$$\sum_{k=1}^{K} p(a_k) l_k \leq \sum_{k=1}^{K} p(a_k)(-\log_2 p(a_k) + 1) = H + 1.$$

□

Codage par blocs L'ajout de 1 bit dans l'inégalité (7.7) vient du fait que $-\log_2 p_i$ n'est pas nécessairement un entier alors que la longueur d'un mot binaire doit être un entier. On peut construire des codes tels que R est plus proche de H en répartissant ce bit supplémentaire sur un bloc de n éléments. Au lieu de faire un codage instantané, symbole par symbole, on code d'un coup le bloc de symboles $\vec{X} = X_1, \ldots, X_n$, qui peut être considéré comme une variable aléatoire à valeurs dans l'alphabet A^n de taille K^n. A tout bloc de symboles $\vec{a} \in A^n$ on associe un mot binaire de longuer $l(\vec{a})$. Le nombre de bits R par symbole pour un tel code par bloc est

$$R = \frac{1}{n} \sum_{\vec{a} \in A^n} p(\vec{a})\, l(\vec{a}).$$

Proposition 7.2 *Le nombre moyen R de bit d'un code par bloc de taille n ayant la propriété du préfixe satisfait*

$$R \geq H = -\sum_{k=1}^{K} p(a_k) \log_2 p(a_k). \tag{7.10}$$

Il existe un code par blocs de taille n ayant la propriété du préfixe tel que

$$R \leq H + \frac{1}{n}. \tag{7.11}$$

Démonstration L'entropie associée à $\vec{X}$ est

$$\vec{H} = \sum_{\vec{x} \in A^n} p(\vec{x})\, \log_2 p(\vec{x}).$$

Comme les variables aléatoires X_i sont indépendantes

$$p(\vec{x}) = p(x_1, ..., x_n) = \prod_{i=1}^{n} p(x_i)\ .$$

On démontre par récurrence sur n que $\vec{H} = nH$. Soit $\vec{R}$ le nombre de bits moyen pour coder les n symboles $\vec{X}$. Le théorème de Shannon 7.3 montre que $\vec{R} \geq \vec{H}$ et qu'il existe un code par bloc tel que $\vec{R} \leq \vec{H} + 1$. On déduit donc (7.10,7.11) pour $R = \frac{\vec{R}}{n}$, qui est le nombre de bits moyen par symbole. □

Ce théorème démontre que des codes par blocs utilisent un nombre moyen de bits par symbole qui tendent vers l'entropie lorsque la taille du bloc augmente.

Code de Huffman L'algorithme de Huffman est un algorithme de programmation dynamique qui construit de bas en haut un arbre correspondant à un code préfixe et qui minimise

$$R = \sum_{k=1}^{K} p(a_k)\, l_k. \tag{7.12}$$

Nous ordonnons $\{a_k\}_{1\leq k\leq K}$ pour que $p(a_k) \leq p(a_{k+1})$. Pour minimiser (7.12) les symboles de plus petites probabilités doivent être associés aux mot binaires w_k de longueur maximale, ce qui correspond à un noeud au bas de l'arbre. Nous commençons donc par représenter les deux symboles de plus petite probabilité a_1 et a_2 comme les enfants d'un noeud commun. Ce noeud peut être interprété comme un symbole $a_{1,2}$ correspondant à "a_1 ou a_2" et dont la probabilité est $p(a_1) + p(a_2)$. La proposition suivante prouve que l'on peut itérer ce regroupement élémentaire et construire un code optimal.

Proposition 7.3 *On considère K symboles avec leurs probabilités ordonnées en ordre croissant: $p(a_k) \leq p(a_{k+1})$. On regroupe les deux symboles a_1 et a_2 de probabilité minimum en un seul symbole $a_{1,2}$ de probabilité*

$$p(a_{1,2}) = p(a_1) + p(a_2).$$

Un arbre correspondant à un code préfixe optimal pour les K symboles se construit à partir d'un arbre de code préfixe optimal pour les $K-1$ symboles $\{a_{1,2}\} \cup \{a_k\}_{3\leq k\leq K}$, en divisant la feuille de $a_{1,2}$ en deux noeuds correspondant à a_1 et a_2.

La démonstration de cette proposition se trouve dans [2]. Cette proposition réduit la construction d'un code optimal de K symboles à la construction d'un code optimal pour les $K-1$ symboles. Le code de Huffman itère $K-1$ fois ce regroupement et fait progressivement pousser l'arbre d'un code de préfixe optimal depuis le bas jusqu'en haut. Le Théorème 7.3 de Shannon prouve que

$$H \leq R \leq H + 1. \tag{7.13}$$

Exemple Les probabilités des $\{a_k\}_{1\leq k\leq 6}$ sont

$$\{p(a_k)\}_{1\leq k\leq 6} = \{0.05\ ,\ 0.1\ ,\ 0.1\ ,\ 0.15\ ,\ 0.2\ ,\ 0.4\}. \tag{7.14}$$

La figure 7.3 donne l'arbre binaire construit avec l'algorithme de Huffman. Les symboles a_1 et a_2 sont regroupés en un symbole $a_{1,2}$ de probabilité $p(a_{1,2}) = p(a_1) + p(a_2) = 0.15$. A l'itération suivante, les symboles de plus basse probabilité sont $p(a_3) = 0.1$ et $p(a_{1,2}) = 0.15$. On regroupe donc $a_{1,2}$ et a_3 en un symbole $a_{1,2,3}$ dont la probabilité est 0.25. Les deux symboles de probabilités les plus faibles sont alors a_4 et a_5 qui sont regroupés en $a_{4,5}$ de probabilité 0.35. On regroupe ensuite $a_{4,5}$ et $a_{1,2,3}$ pour obtenir un symbole $a_{1,2,3,4,5}$ de probabilité 0.6 qui est finalement regroupé avec a_6, ce qui finit le code, comme l'illustre l'arbre de la figure 7.3. Le nombre moyen de bits obtenu par ce code est $R = 2.35$ alors que l'entropie est $H = 2.28$.

Sensibilité au bruit Un code de Huffman est plus compact qu'un code de taille fixe $\log_2 K$ mais est aussi plus sensible au bruit. Pour un code de taille constante, une erreur de transmission d'un bit modifie seulement la valeur d'un symbole. Au contraire, une erreur d'un bit dans un code de taille variable peut modifier toute la suite des symboles. Lors de transmissions bruitées, de telles erreurs peuvent se produire. Il est alors nécessaire d'utiliser un code correcteur qui introduit une légère redondance de facon à identifier les erreurs.

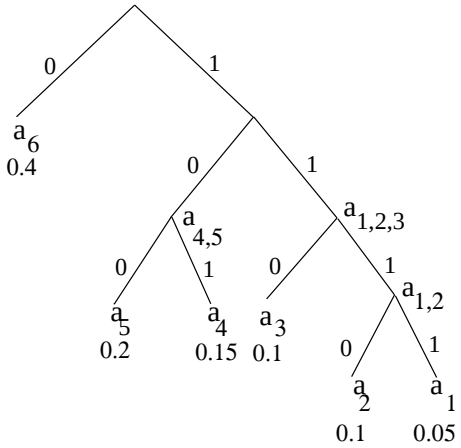

Figure 7.3: Arbre correspondant au code de Huffman pour une source dont les probabilités sont données par (7.14).

7.2 Quantification scalaire

Si une variable aléatoire X prend des valeurs réelles quelconques, on ne peut pas obtenir un code exact de taille finie. Il est alors nécessaire d'approximer X par $\tilde{X}$ qui prend ses valeurs dans un alphabet fini, et l'erreur résultante est

$$D = E\{|X - \tilde{X}|^2\}.$$

Un quantificateur scalaire décompose l'axe réel en K intervalles $\{[y_{k-1}, y_k]\}_{1 \le k \le K}$ de tailles variables, avec $y_0 = -\infty$ et $y_K = +\infty$. Le quantificateur associe à tout $x \in [y_{k-1}, y_k]$ une valeur $Q(x) = a_k$. Si les K niveaux de quantification $\{a_k\}_{1 \le k \le K}$ sont fixés a priori, pour minimiser $|x - Q(x)| = |x - a_k|$, il faut que la quantification associe à x son niveau de quantification a_k le plus proche. On doit alors choisir des intervalles de quantification qui satisfont

$$y_k = \frac{a_k + a_{k+1}}{2} \tag{7.15}$$

Quantification haute résolution Soit $p(x)$ la densité de probabilité de X. On note $\tilde{X} = Q(X)$ la variable quantifiée. L'erreur quadratique moyenne est

$$D = E\{(X - \tilde{X})^2\} = \int_{-\infty}^{+\infty} |x - Q(x)|^2 p(x) dx. \tag{7.16}$$

On dit que le quantificateur a une haute résolution si $p(x)$ peut être approximé par une constante sur tout intervalle de quantification $[y_{k-1}, y_k]$. La taille de ces intervalles est $\Delta_k = y_k - y_{k-1}$. L'hypothèse de haute résolution implique que

$$p(x) = \frac{p_k}{\Delta_k} \quad \text{pour } x \in [y_{k-1}, y_k], \tag{7.17}$$

avec

$$p_k = \Pr\{X \in [y_{k-1}, y_k]\} = \Pr\{\tilde{X} = a_k\}.$$

La proposition suivant calcule l'erreur D sous cette hypothèse.

Proposition 7.4 *Pour un quantificateur de haute résolution sur des intervalles* $[y_{k-1}, y_k]$, *l'erreur* D *minimum obtenue en optimisant la position des niveaux* $\{a_k\}_{0 \leq k \leq K}$ *est*

$$D = \frac{1}{12} \sum_{k=1}^{K} p_k \, \Delta_k^2. \tag{7.18}$$

Démonstration Comme $Q(x) = a_k$ si $x \in [y_{k-1}, y_k)$, on peut reécrire (7.16)

$$D = \sum_{k=1}^{K} \int_{y_{k-1}}^{y_k} (x - a_k)^2 p(x) dx.$$

En remplaçant $p(x)$ par son expression (7.17) on a

$$D = \sum_{k=1}^{K} \frac{p_k}{\Delta_k} \int_{y_{k-1}}^{y_k} (x - a_k)^2 dx. \tag{7.19}$$

Cette erreur est minimum pour $a_k = \frac{1}{2}(y_k + y_{k-1})$, et l'intégration donne (7.18). □

Quantification uniforme Le quantificateur uniforme est un cas particulier important où tous les intervalles de quantification sont de même taille

$$y_k - y_{k-1} = \Delta \quad \text{pour } 1 \leq k \leq K.$$

L'erreur quadratique moyenne (7.18) devient

$$D = \frac{\Delta^2}{12} \sum_{k=1}^{K} p_k = \frac{\Delta^2}{12}. \tag{7.20}$$

Elle est indépendante de la distribution de probabilité $p(x)$ de la source.

Quantification optimale On veut optimiser le quantificateur pour minimiser le nombre de bits nécessaires pour coder les valeurs quantifiées $\tilde{X}$, étant donnée une distortion D admissible. Le théorème de Shannon 7.3 prouve que la valeur moyenne minimum de bits nécessaire pour coder $\tilde{X}$ est supérieure à l'entropie H de la variable aléatoire $\tilde{X}$. Comme le code de Huffman donne un résultat proche de cette entropie, il nous faut minimiser l'entropie H pour D fixe.

La source quantifiée $\tilde{X}$ prend K valeurs différentes $\{a_k\}_{1 \leq k \leq K}$ avec probabilités $\{p_k\}_{1 \leq k \leq K}$. L'entropie du signal quantifié est donc

$$H = -\sum_{k=1}^{K} p_k \, \log_2 p_k.$$

On définit l'entropie différentielle de la variable aléatoire X à valeurs réelles

$$H_d = -\int_{-\infty}^{+\infty} p(x) \log_2 p(x) dx. \tag{7.21}$$

Le théorème suivant montre que pour un quantificateur de haute résolution produisant une erreur D, l'entropie est minimum lorsque le quantificateur est uniforme.

Théorème 7.4 *L'entropie de tout quantificateur de haute résolution satisfait*

$$H \geq H_d - \frac{1}{2}\log_2(12D). \tag{7.22}$$

Le minimum est atteint si et seulement si Q est un quantificateur uniforme.

Démonstration Pour un quantificateur de haute résolution $p(x)$ est approximativement constant sur $[y_{k-1}, y_k]$ et donc

$$p_k = \int_{y_{k-1}}^{y_k} p(x)\,dx = p_k \Delta_k$$

avec $\Delta_k = y_k - y_{k-1}$. Donc

$$\begin{aligned} H &= -\sum_{k=1}^{K} p_k \log_2(p(a_k)\,\Delta_k) \\ &= -\sum_{k=1}^{K} \int_{y_{k-1}}^{y_k} p(x) \log_2 p(a_k)\,dx - \sum_{k=1}^{K} p_k \log_2 \Delta_k \\ &= H_d - \frac{1}{2}\sum_{k=1}^{K} p_k \log_2 \Delta_k^2, \end{aligned}$$

car $p(x) = p(a_k)$ pour $x \in [y_{k-1}, y_k]$. Pour toute fonction concave $\phi(x)$, l'inégalité de Jensen montre que pour tout $\sum_{k=1}^{K} p_k = 1$ et $\{a_k\}_{1\leq k\leq K}$ alors

$$\sum_{k=1}^{N} p_k \phi(a_k) \leq \phi(\sum_{k=1}^{N} p_k a_k). \tag{7.23}$$

Si $\phi(x)$ est strictement concave, l'inégalité devient une égalité si et seulement si tous les a_k sont égaux lorsque $p_k \neq 0$. Comme $\log_2(x)$ est strictement concave, (7.18) montre que

$$\frac{1}{2}\sum_{k=1}^{N} p_k \log_2 \Delta_k^2 \leq \frac{1}{2}\log_2 \sum_{k=1}^{N} p_k \Delta_k^2 = \frac{1}{2}\log_2(12D).$$

On en déduit donc que

$$H \geq H_d - \frac{1}{2}\log_2(12D).$$

Cette inégalité devient une égalité si et seulement si tous les Δ_k sont égaux, ce qui correspond à un quantificateur uniforme. □

Ce théorème montre que pour un quantificateur haute résolution le nombre minimum de bits $R = H$ est obtenu pour un quantificateur uniforme et

$$R = H_d - \frac{1}{2}\log_2(12D). \tag{7.24}$$

La distortion en fonction du nombre de bits est donc

$$D(R) = \frac{1}{12}\, 2^{2H_d}\, 2^{-2R}.$$

Ce résultat est important pour optimiser la compression de signaux étudiée dans le chapitre suivant.

Chapitre 8

Compression de Signaux

La compression de signaux s'apparente à la deshydratation d'un litre de jus d'orange en quelques grammes de poudre concentrée. Le goût de la boisson orange restituée est similaire au jus d'orange mais a souvent perdu de sa subtilité. En traitement du signal, nous sommes plus interessés par des sons ou des images, mais l'on rencontre le même conflit entre qualité et compression. Minimiser la dégradation pour un taux de compression donné est le but des algorithmes de codage. Les applications principales concernent le stockage des données et la transmission à travers des canaux à débit limité.

Nous étudions les algorithmes de codage qui décomposent le signal sur une base orthogonale et approximent efficacement les coefficients de décomposition. Ce type de codage est actuellement le plus performant pour restituer des signaux audios ou des images de bonne qualité.

8.1 Codage compact

La performance ultime d'un algorithme de codage est mesurée par un "score d'opinion moyen". Pour un taux de compression donné, la qualité des signaux codés est évaluée par plusieurs personnes et calibrés selon une procédure précise. De telles évaluations sont longues à faire et les résultats difficiles à interpréter mathématiquement. La qualité d'un algorithme de codage est donc le plus souvent optimisée avec une distance qui a une forme analytique simple et qui tient compte partiellement de notre sensibilité visuelle ou auditive. La distance euclidienne, bien que relativement grossière d'un point de vue perceptuel, a l'avantage d'être facile à manipuler analytiquement.

8.1.1 Etat de l'art

Parole Le codage de la parole est particulièrement important pour la téléphonie où il peut être de qualité médiocre tout en maintenant une bonne intelligibilité. Un signal de parole par téléphone est limité à la bande de fréquence 200-3400 Hz et est échantillonné à 8kHz. Un "Pulse Code Modulation (PCM)" qui quantifie chaque échantillon sur 8 bits produit un code de 64kb/s (64 10^3 bits par seconde). Ceci peut être considérablement réduit en supprimant certaines composantes redondantes de la parole.

Nous avons vu dans le paragraphe 5.1 que la production d'un signal de parole est bien comprise. Des filtres autorégressifs excités par des trains d'impulsions ou un bruit blanc Gaussien permettent de restaurer un signal intelligible à partir de peu de paramètres. Ces codes d'analyse-synthèse, tel que le standard LPC-10 décrit dans le paragraphe 5.3.2, produisent un signal de parole intelligible à 2,4kb/s.

Audio Les signaux audios peuvent inclure de la parole mais aussi de la musique et n'importe quel type de son. Ils sont donc beaucoup plus difficiles à modéliser que la parole. Sur un disque compact, le signal audio est limité à un maximum de 20kHz. Il est échantillonné à 44.1kHz et chaque échantillon est codé sur 16bits. Le débit du code PCM résultant est donc de 706kb/s. Le signal audio d'un disque compact ou d'une cassette digitale doit être codé sans distortion auditive.

Les codeurs par transformée orthogonale qui décomposent les signaux sur des bases locales en temps et en fréquence sont parmi les plus performants pour les signaux audios, car ils ne nécessitent pas la mise en place d'un modèle. Une qualité de disque compact est obtenue par des codeurs nécessitant 128kb/s. Avec 64kb/s les dégradations sont à peine audibles. Ces algorithmes sont particulièrement importants pour les CD-ROM en multimédia.

Images Une image couleur est composée de trois canaux d'intensité dans le rouge, le vert et le bleu. Chacune de ces images a typiquement 500 par 500 pixels, qui sont codés sur 8 bits (256 niveaux de gris). Un canal téléphonique digital ISDN a un débit de 64kb/s. Il faut donc environ 2 minutes pour transmettre une telle image.

Les images tout comme les signaux audios incluent le plus souvent des structures de type différent qu'il est difficile de modéliser. Courament, les algorithmes de compression les plus efficaces sont basés sur des transformées orthogonales utilisant des bases de cosinus locaux ou des bases d'ondelettes. Avec moins de 1 bit/pixel, ces codes reproduisent une image de qualité visuelle presque parfaite. A 0.25 bit/pixel, l'image reste de bonne qualité visuelle et peut être transmise en 4 secondes sur une ligne téléphonique digitale.

8.1.2 Codage dans une base orthogonale

Un code orthogonal décompose le signal dans une base orthogonale bien choisie de façon à optimiser la compression des coefficients de décomposition. La classe des signaux codés est representée par un vecteur aléatoire $Y[n]$ de taille N. On décompose $Y[n]$ sur une base orthogonale $\{g_m[n]\}_{0 \leq n < N}$

$$Y[n] = \sum_{m=0}^{N-1} A[m]\, g_m[n].$$

Les coefficients de décomposition $A[m]$ sont des variables aléatoires

$$A[m] = < Y[n], g_m[n] > = \sum_{n=0}^{N-1} Y[n]\, g_m^*[n].$$

Si $A[m]$ n'est pas de moyenne nulle, on code $A[m] - E\{A[m]\}$ et on mémorise la valeur moyenne $E\{A[m]\}$. Nous supposerons par la suite que $E\{A[m]\} = 0$.

Quantification Les valeurs réelles $\{A[m]\}_{0 \leq m < N}$ doivent être approximées avec une précision finie pour construire un code de taille finie. Une quantification scalaire approxime chaque $A[m]$ individuellement. Si les coefficients $A[m]$ ont une forte dépendance, le taux de compression peut être amélioré avec une quantification vectorielle qui regroupe les coefficients en blocs. Cette approche nécessite cependant plus de calculs. Si la base $\{g_m\}_{0 \leq m < N}$ est choisie de façon à ce que les coefficients $A[m]$ soient presque indépendants, l'amélioration d'une quantification vectorielle est marginale.

Chaque $A[m]$ est une variable aléatoire dont la valeur quantifiée est $\tilde{A}[m] = Q_m(A[m])$ Le signal quantifié reconstruit est

$$\tilde{Y}[n] = \sum_{m=0}^{N-1} \tilde{A}[m]\, g_m[n].$$

Comme la base est orthogonale

$$\|Y - \tilde{Y}\|^2 = \sum_{m=0}^{N-1} |A[m] - \tilde{A}[m]|^2$$

et donc la valeur moyenne de l'erreur est

$$E\{\|Y - \tilde{Y}\|^2\} = \sum_{m=0}^{N-1} E\{|A[m] - \tilde{A}[m]|^2\}.$$

Si l'on note

$$D_m = E\{|A[m] - \tilde{A}[m]|^2\},$$

l'erreur totale devient

$$D = \sum_{m=0}^{N-1} D_m.$$

Soit R_m le nombre moyen de bits pour coder $\tilde{A}[m]$. Pour une quantification à haute résolution, le théorème 7.4 montre que si D_m est fixé alors on minimise R_m avec une quantification scalaire uniforme. On note Δ_m la taille des intervales de quantification. On a montré en (7.20) que

$$D_m = \frac{\Delta_m^2}{12}.$$

et (7.22) prouve que

$$R_m = H_d(X) - \frac{1}{2}\log_2(12 D_m) = H_d(X) - \log_2 \Delta_m \ ,$$

où $H_d(X)$ est l'entropie différentielle de X définie par (7.21).

Allocation de bits Si l'on fixe l'erreur totale D, il nous faut optimiser le choix de $\{\Delta_m\}_{0\leq m<N}$ afin de minimiser le nombre total de bits

$$R = \sum_{m=0}^{N-1} R_m.$$

Soit $\bar{R} = \frac{R}{N}$ le nombre moyen de bits par coefficient. Le théorème suivant montre que le codage est optimisé lorsque tous les Δ_m sont égaux.

Théorème 8.1 *Pour une quantification haute résolution et une erreur totale D, on minimize $\bar{R}$ avec*

$$\Delta_m^2 = \frac{12\,D}{N} \quad \textit{pour } 0 \leq m < N, \tag{8.1}$$

auquel cas

$$D(\bar{R}) = \frac{N}{12}\, 2^{2\overline{H}_d}\, 2^{-2\bar{R}}\ , \tag{8.2}$$

où $\overline{H}_d$ est l'entropie différentielle moyenne

$$\overline{H}_d = \frac{1}{N} \sum_{m=0}^{N-1} H_d(A[m]).$$

Démonstration Pour une quantification haute résolution uniforme, (7.24) montre que

$$R_m = H_d(A[m]) - \frac{1}{2}\log_2(12\,D_m).$$

Donc

$$\bar{R} = \frac{1}{N}\sum_{m=0}^{N-1} R_m = \frac{1}{N}\sum_{m=0}^{N-1} H_d(A[m]) - \frac{1}{N}\sum_{m=0}^{N-1}\frac{1}{2}\log_2(12\,D_m). \tag{8.3}$$

Minimiser $\bar{R}$ revient à minimiser $\sum_{m=0}^{N-1}\log_2(12D_m)$. En appliquant l'inégalité de Jensen (7.23) à la fonction concave $\phi(x) = \log_2(x)$ pour $p_k = \frac{1}{N}$ on obtient

$$\frac{1}{N}\sum_{m=0}^{N-1}\log_2(12\,D_m) \leq \log_2\left(\frac{12}{N}\sum_{m=0}^{N-1} D_m\right) = \log_2\left(\frac{12\,D}{N}\right).$$

Cette inégalité est une égalité si et seulement si tous les D_m sont égaux. Donc $\frac{\Delta_m^2}{12} = D_m = \frac{D}{N}$, ce qui prouve (8.1). On déduit aussi de (8.3) que

$$\bar{R} = \frac{1}{N}\sum_{m=0}^{N-1} H_d(A[m]) - \frac{1}{2}\log_2\left(\frac{12\,D}{N}\right)$$

ce qui implique (8.2). □

Ce théorème montre que le codage est optimisé en introduisant la même erreur moyenne $D_m = \frac{\Delta_m^2}{12} = \frac{D}{N}$ dans la direction de chaque vecteur g_m de la base. Le nombre moyen de bits R_m pour coder $A[m]$ dépend alors seulement de l'entropie différentielle:

$$R_m = H_d(A[m]) - \frac{1}{2}\log_2\left(\frac{12D}{N}\right). \tag{8.4}$$

On note σ_m^2 la variance de $A[m]$, et $\hat{A}[m] = \frac{1}{\sigma_m} A[m]$ la variable aléatoire normalisée de variance 1. Un changement de variable dans l'intégrale de l'entropie différentielle montre que

$$H_d(A[m]) = H_d(\hat{A}[m]) + \log_2 \sigma_m \ .$$

L'allocation de bit optimal R_m donné par (8.4) peut donc devenir négative si la variance σ_m est trop petite, ce qui n'est clairement pas une solution admissible. En pratique R_m doit être un entier positif mais imposer cette contrainte supplémentaire ne permet pas de faire un calcul analytique simple. La formule (8.4) n'est donc utilisable que si les valeurs $\{R_m\}_{0\leq m<N}$ sont positives et on les approxime alors par les entiers les plus proches.

Normes quadratiques pondérées Nous avons mentionné qu'une erreur quadratique souvent ne mesure pas bien l'erreur perçue pour des images ou des signaux audios. Lorsque les vecteurs g_m sont bien localisés en temps/espace et en fréquence, on peut améliorer cette norme en pondérant les erreurs par des poids qui dépendent de la fréquence, afin de se rapprocher de notre sensibilité auditive/visuelle, qui varie avec la fréquence du signal. Une norme pondérée est définie par

$$D = \sum_{m=0}^{N-1} \frac{D_m}{w_m^2}, \tag{8.5}$$

où $\{w_m^2\}_{0\leq m<N}$ sont des constantes.

Le théorème 8.1 s'applique à une norme pondérée en observant que

$$D = \sum_{m=0}^{N-1} D_m^w,$$

où $D_m^w = \frac{D_m}{w_m^2}$ est l'erreur de quantification de $A^w[m] = \frac{A[m]}{w_m}$. Le théorème 8.1 prouve que l'allocation de bits optimale est obtenue en quantifiant uniformément tous les $A^w[m]$ avec des intervales de même tailles Δ. Cela implique que les coefficients $A[m]$ sont uniformément quantifiés avec des intervales de taille $\Delta_m = \Delta\, w_m$, et donc $D_m = \frac{w_m^2 D}{N}$. Comme on s'y attendait, en augmentant les poids w_m on augment l'erreur dans la direction de g_m. La quantification uniform Q_{Δ_m} à intervales Δ_m est souvent calculée avec un quantificateur Q qui associe a tout nombre réel l'entier le plus proche:

$$Q_{\Delta_m}\Big(A[m]\Big) = \Delta_m\, Q\left(\frac{A[m]}{\Delta_m}\right) = \Delta\, w_m\, Q\left(\frac{A[m]}{\Delta\, w_m}\right). \tag{8.6}$$

8.2 Bases de cosinus locaux

Choix de la base Le codage d'une classe de signaux $Y[n]$ dans une base orthogonale $\{g_m[n]\}_{1\leq m\leq N}$ est d'autant meilleur que la base supprime les corrélations entre les coefficients $Y[n]$. La base $\{g_m[n]\}_{1\leq m\leq N}$ est donc choisie de façon à obtenir des coefficients $A[m] =< Y, g_m >$ aussi décorrelés que possible. De même il est souvent désirable d'obtenir des coefficients $A[m]$ qui ont une forte probabilité d'être proches de zéro. De tels coefficients sont en effet annulés par la quantification et efficacement codés par en séquences. Les bases de cosinus décrites dans ce paragraphe sont bien adaptées pour le codage des signaux audios et des images. L'existence d'algorithmes de calcul rapide basés sur la transformée de Fourier rapide permet d'utiliser ces bases pour du codage en temps réel.

Bases de Cosinus Le codage de signaux réels $f[n]$ defini pour $0 \leq n < N$ se fait plutôt sur des bases de cosinus que sur des bases de Fourier discrètes $\{\frac{1}{\sqrt{N}} e^{\frac{i2\pi kn}{N}}\}_{0\leq k<N}$. En effet nous avons vu dans le paragraphe 3.4.2 que la décompositon d'un signal de taille N dans une base de Fourier discrète effectue une périodisation de $f[n]$ sur N échantillons. Si $f[0] \neq f[N-1]$, ce signal périodique a des transitions brutales en $n = 0$ et $n = N-1$ ce qui produit des coefficients de Fourier de large amplitude. La quantification de ces coefficients de large amplitude crée des erreurs aux bords que l'on veut éviter. On montre que la base de cosinus spécifiée par le Théorème 8.2 possède essentiellement les mêmes propriétés qu'une base de Fourier discrète, mais ne produit pas des coefficients de large amplitude par effet de bord. Cette base de cosine est basée sur une extension $\tilde{f}[n]$ périodique de $f[n]$, qui évite l'introduction de transition brutale aux bords.

Le signal $f[n]$ défini sur $0 \leq n < N$ est étendu par symétrie par rapport à $-\frac{1}{2}$ en un signal $\tilde{f}[n]$ de taille $2N$

$$\tilde{f}[n] = \begin{cases} f[n] & \text{for } 0 \leq n \leq N \\ f[-n-1] & \text{for } -N \leq n \leq -1 \end{cases} \tag{8.7}$$

La symétrie évite d'introduire une transition brutale lors de la periodisation sur $2N$ coefficients car $\tilde{f}[0] = \tilde{f}[2N-1]$. La transformée de Fourier discrète de taille $2N$ décompose $\tilde{f}[n]$ comme une somme de sinus et de cosinus

$$\tilde{f}[n] = \sum_{k=0}^{N-1} a_k \cos\Big[\frac{2k\pi}{2N}(n+\frac{1}{2})\Big] + \sum_{k=0}^{N-1} b_k \sin\Big[\frac{2k\pi}{2N}(n+\frac{1}{2})\Big].$$

Comme $\tilde{f}[n]$ est symétrique par rapport à $-\frac{1}{2}$, on déduit que $b_k = 0$ pour tout $0 \leq k < N$. De plus $f[n] = \tilde{f}[n]$ pour $0 \leq n < N$, ce qui prouve que tout signal $f \in \mathbb{R}^N$ peut s'écrire comme une somme de ces cosinus. On vérifie aussi facilement que ces cosinus sont orthogonaux dans $\mathbb{R}^N$. On obtient donc le théorème suivant.

Théorème 8.2 *La famille de cosinus discrets*

$$\left\{ c_k[n] = \lambda_k \sqrt{\frac{2}{N}} \cos\Big[\frac{k\pi}{N}(n+\frac{1}{2})\Big] \right\}_{0\leq k<N},$$

avec

$$\lambda_k = \begin{cases} \frac{1}{\sqrt{2}} & si\ k = 0 \\ 1 & sinon \end{cases} \tag{8.8}$$

est une base orthonormale de $\mathbb{R}^N$.

Les produits scalaires

$$\hat{f}_c[k] =< f[n], c_k[n] >= \lambda_k \sqrt{\frac{2}{N}} \sum_{n=0}^{N-1} f[n] \cos\Big[\frac{k\pi}{N}(n+\frac{1}{2})\Big] \tag{8.9}$$

définissent la transformée en cosinus de $f[n]$. Comme ces cosinus forment une base orthonormale

$$f[n] = \sum_{k=0}^{N-1} < f, c_k > \ c_k[n] = \sum_{k=0}^{N-1} \hat{f}_c[k]\, c_k[n]. \tag{8.10}$$

En modifiant l'algorithme de transformée de Fourier rapide, on peut calculer les coefficients $\{\hat{f}_c[k]\}_{0\leq k<N}$ avec $O(N \log N)$ opérations [7]. De même la reconstruction (8.10) s'obtient avec $O(N \log N)$ opérations.

Localisation temporelle Lorsque le signal inclut des structures variées à différents instants, il est préférable de séparer ces composantes avec des fenêtres temporelles et d'effectuer une transformée en cosinus à l'intérieur de ces fenêtres. Cette approche est similaire à la transformée de Fourier à fenêtre que nous avons étudiée dans le chapitre 6.

Un signal de taille N peut etre séparé en $\frac{N}{M}$ composantes de tailles M en utilisant des fenêtres rectangulaires de taille M

$$g[n] = \begin{cases} 1 & \text{si } 0 \leq n < M \\ 0 & \text{sinon} \end{cases}$$

Clairement

$$f[n] = \sum_{p=1}^{\frac{N}{M}} g[n-pM] f[n].$$

Le produit $g[n-pM]f[n]$ est la restriction de $f[n]$ pour $pM \leq n < (p+1)M$. On peut décomposer cette restriction sur une base de de cosinus de taille M obtenue en translatant la famille

$$\left\{ c_k[n] = \lambda_k \sqrt{\frac{2}{M}} \cos\Big[\frac{k\pi}{M}(n+\frac{1}{2})\Big] \right\}_{0\leq k<M}.$$

Chaque partie $g[n-pM]f[n]$ se décompose sur la famille $\{c_k[n-pM]g[n-pM]\}_{0\leq k<M}$ restreinte à $pM \leq n < (p+1)M$. Cela revient à décomposer $f[n]$ sur une base orthogonale de taille N composée de $\frac{N}{M}$ fenêtres de taille M, modulées par des cosinus

$$\Big\{ g_{p,k}[n] = g[n-pM] c_k[n-pM] \Big\}_{0\leq k<M, 0\leq p<\frac{N}{M}}. \tag{8.11}$$

Cette famille est une base orthonormale de l'espace des signaux de taille N.

Bases bidimensionnelles La proposition suivante montre que des bases orthogonales d'images $f[n,m]$ de taille N^2 peuvent se construire par un produit séparable de bases orthogonales de signaux mono-dimensionnels $f[n]$ de taille N.

Proposition 8.1 *Si* $\{e_k[n]\}_{0\leq k<N}$ *est une base orthonormale de l'espace des signaux de taille* N *alors*

$$\{e_{k,j}[n,m] = e_k[n]e_j[m]\}_{0\leq k<N, 0\leq j<N}$$

est une base orthogonale de l'espace des images $f[n,m]$ *de taille* N^2.

Démonstration Il suffit pour cela de montrer que ces N^2 vecteurs sont orthogonaux. En effet

$$\begin{aligned} < e_{k,j}, e_{k',j'} > &= \sum_{n=0}^{N-1}\sum_{m=0}^{N-1} e_{k,j}[n,m]e^*_{k',j'}[n,m] \\ &= \sum_{n=0}^{N-1} e_k[n]e^*_{k'}[n] \sum_{m=0}^{N-1} e_j[m]e^*_{j'}[m]. \end{aligned}$$

Ces produits scalaires sont donc tous nuls si $k \neq k'$ et $j \neq j'$ car $\{e_k[n]\}_{1\leq k\leq N}$ est une base orthogonale. □

Cette proposition nous permet de construire des bases de cosinus locaux d'images par un produit séparable de bases de cosinus locaux (8.11) de signaux monodimensionnels. La famille de N^2 fenêtres modulées

$$\{g_{k,j}[n-pM, m-qM] = g[n-pM]g[m-qM]c_k[n-pM]c_j[m-qM]\}_{0\leq k,l<M, 0\leq p,q\leq \frac{N}{M}} \tag{8.12}$$

est une base orthogonale de l'espace des images $f[n,m]$ de taille N^2. Cette base décompose l'image en $\frac{N^2}{M^2}$ carrés de taille M. Elle décompose ensuite la restriction de l'image sur chaque carré de M^2 pixels dans une base de cosinus.

8.3 Codage perceptuel

Les codes par transformée orthogonale sont particulièrement efficaces pour de larges classes de signaux pour lesquelles on ne peut définir des modèles de production. C'est le cas pour les signaux audios ou les images. Le choix de la base et la quantification des coefficients doivent cependant être adaptés à la perception humaine de façon à produire des erreurs qui introduisent peu de dégradations perceptuelles.

8.3.1 Codage audio

Un disque compact mémorise un son audio de haute qualité avec un échantillonnage a 44.1 kHz. Les échantillons sont quantifiés uniformément sur 16 bits ce qui produit un "Pulse Code Modulation" de 706kb/s. Un code "transparent" peut introduire des erreurs

numériques mais ces erreurs doivent rester inaudibles pour un auditeur "moyen". De forts taux de compression sont obtenus en adaptant la quantification aux propriétés de masquage auditif.

Masquage auditif Une petite erreur de quantification n'est pas entendue si elle est additionnée à un signal qui a une forte énergie dans la même bande de fréquence. Cet effet de masquage se fait à l'intérieur de bandes de fréquence "critiques" de la forme $[\omega_c - \frac{\Delta\omega}{2}, \omega_c + \frac{\Delta\omega}{2}]$, qui ont été mesurées par des expériences de psycho-physiologie auditive. Un fort signal dont la transformée de Fourier a un support contenu dans la bande de fréquence $[\omega_c - \frac{\Delta\omega}{2}, \omega_c + \frac{\Delta\omega}{2}]$ décroît la sensibilité d'un auditeur pour d'autres composantes qui sont à l'intérieur de cette bande de fréquence. Dans l'intervalle de fréquences $[0, 20\text{kHz}]$, il y a approximativement 25 bandes critiques dont la largeur $\Delta\omega$ et la fréquence centrale ω_c satisfont

$$\Delta\omega \approx \begin{cases} 100 & \text{for } \omega_c \leq 700 \\ 0.15\omega_c & \text{for } 700 \leq \omega_c \leq 20\ 000 \end{cases} \tag{8.13}$$

Quantification adaptative Le signal $f[n]$ est décomposé dans une base de cosinus locaux

$$\left\{ g_{p,k}[n] = g[n - pM] c_k[n - pM] \right\}_{0 \leq k < M, 0 \leq p < \frac{N}{M}}$$

construits avec une fenêtre $g[n]$ couvrant généralement $M = 1024$ échantillons. Par une transformée de Fourier rapide, pour chaque composante du signal $f[n]g[n - pM]$ de M échantillons, on calcule l'énergie en fréquence dans les bandes critiques $[\omega_c - \frac{\Delta\omega}{2}, \omega_c + \frac{\Delta\omega}{2}]$. Cette énergie permet de calculer le niveau de masquage et donc l'amplitude maximum Δ des erreurs de quantification qui ne sont pas audibles dans cette bande de fréquence. On quantifie alors uniformement avec un pas Δ chaque produit scalaire $< f, g_{p,k} >$ pour des cosinus locaux dont la fréquence se trouve à l'intérieur de la bande critique $[\omega_c - \frac{\Delta\omega}{2}, \omega_c + \frac{\Delta\omega}{2}]$. Cet algorithme introduit dans chaque bande critique des erreurs de quantification qui sont au-dessous du niveau d'audition.

MUSICAM L'algorithme MUSICAM (Masking-pattern Universal Subband Integrated Coding and Multiplexing) utilisé par le standard MPEG-I est le plus simple des codeurs perceptuels. Il décompose le signal en 32 bandes de fréquences de tailles égales dont la largeur est de 750Hz. Cette décomposition est très semblable à une décomposition dans une base de cosinus locaux. Chaque $8\ 10^{-3}$ seconde la quantification est adaptée dans chaque bande de fréquence pour tenir compte des propriétés de masquage du signal. Ce système comprime des signaux audios jusqu'à 128 kb/s sans introduire d'erreur audible.

8.3.2 Codage d'images par JPEG

Le standard JPEG est le plus couramment utilisé pour la compression d'images. L'image est décomposée dans une base séparable de cosinus locaux, en utilisant des fenêtres de $M = 8$ par 8 pixels

$$\{ g_{k,j}[n - pM, m - qM] = g[n - pM] g[m - qM] c_k[n - pM] c_j[n - qM] \}_{0 \leq k,j < M, 0 \leq p,q \leq \frac{N}{M}} .$$

Cela signifie que l'image est décomposée en carrés de 8 par 8 pixels et que chacun de ces carrés est décomposé dans une base séparable de cosinus. Dans chaque fenêtre, les 64 coefficients de décomposition sont quantifiés uniformément. Dans les zones où l'image est régulière, les cosinus de hautes fréquences génèrent des petits coefficients, qui sont annulés par la quantification. Pour coder efficacement la position de ces coefficients nuls, on utilise un code par séquences.

Codage des zéros Enregistrer la position des coefficients nuls correspond au codage d'une source binaire égale à 1 lorsque le coefficient est non-nul et 0 lorsque qu'il est nul. On peut utiliser la redondance de cette source de 0 et de 1 en codant les valeurs sous forme de séquences. Un code "run-length" enregistre la taille Z des séquences successives de 0 et la taille I des sequences successives de 1. Les variables aléatoires Z et I sont ensuite enregistrées avec un code entropique. Un code "run-length" est un algorithme de codage vectoriel dont on peut démontrer qu'il est optimal lorsque la séquence de 0 et de 1 est produite par une chaîne de Markov d'ordre 1. Ce type de codage binaire est utilisé pour la transmission de fax.

Chaque bloc de 64 coefficients en cosinus est parcouru en zig-zag, comme l'illustre la figure 8.1. Sur ce parcours, on enregistre la taille des séquences de 0 et de 1 correspondant aux coefficients quantifiés à zero ou pas.

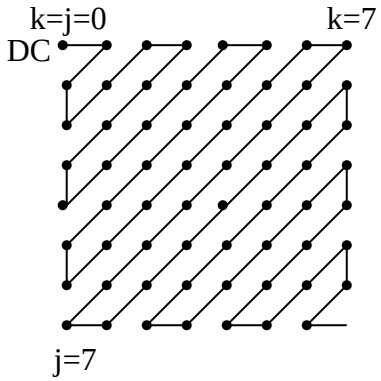

Figure 8.1: Sur une fenêtre de 64 coefficients en cosinus, la fréquence zero (DC) est en haut à gauche. Le codage des zéro effectue un parcours en zig-zag depuis les basses vers les hautes fréquences.

Sur chaque bloc, le vecteur de fréquence zéro $g_{0,0}[n - pM, m - qM]$ a une valeur constante. Le coefficient correspondant est donc proportionel à l'intensité moyenne de l'image sur le bloc. Ces coefficients sont codées séparéments car ils sont fortement correlées d'un bloc à l'autre. Plutôt que de quantifier individuellement les fréquences nulles ($k = j = 0$), on code la différence des coefficients de fréquence nulles, entre un bloc de l'image et le suivant qui se trouve à sa droite.

Perception visuelle La sensibilité visuelle dépend de la fréquence et de l'orientation du stimulus. Les propriétés de masquage sont aussi importantes en vision que pour l'audition mais sont beaucoup plus compliquées à modéliser. Des expériences psycho-physiologiques montrent qu'un stimulus dont la transformée de Fourier est localisée dans une bande de

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	108	103	77
24	35	55	64	81	194	113	92
49	64	78	87	103	121	120	101
72	92	95	98	121	100	103	99

Table 8.1: Matrice de poids $w_{k,j}$ utilisés pour la quantification des coefficients correspondant à chaque bloc de cosinus $g_{k,j}$. Les fréquences augmentent de gauche à droite et de haut en bas.

fréquence étroite produit un effet de masquage sur une bande de fréquence de l'ordre de 1 à 1.5 octave. Ce masquage dépend de la fréquence et de l'orientation du signal mais aussi d'autres paramètres d'intensité, de couleur et de texture. Cette complexité rend beaucoup plus difficile l'utilisation des propriétés de masquage en vision. Les codeurs adaptent donc plutôt l'erreur de quantification suivant une sensibilité "moyenne" dans chaque bande de fréquence, sans utiliser les propriétés de masquage Nous sommes typiquement moins sensibles aux oscillations de hautes fréquences qu'aux variations de basses fréquences.

Pour minimiser la dégration visuelle des images codées, JPEG effectue une quantification avec des intervales de quantification dont les valeurs sont proportionelles à des poids qui sont calculés grace des expériences psychophysiques. Ceci revient à optimiser l'erreur pondérée (8.5). La table 8.1 est un exemple de matrice de 8 par 8 poids qui peuvent être utilisés par JPEG. Les poids des fréquences les plus basses, qui apparaissent en haut à gauche de la table 8.1, sont 10 fois plus petit qu'aux plus hautes fréquences, qui apparaissent en bas à droite.

Qualité de compression Pour $\bar{R} \in [0,75\,,\,1]$bit/pixel, les images codées avec JPEG sont visuellement quasiment parfaites. L'algorithme JPEG est souvent utilisé avec $\bar{R} \in [0,5\,,\,1]$. Lorsque $\bar{R} \in [0,2\,,\,0,5]$bit/pixel, la figure 8.2 montre que les images restorées à partir d'un code JPEG sont de qualité modérée. A fort taux de compression, on voit apparaître les blocs de 8 par 8 pixels sur lesquels la transformée en cosinus est calculée. Les performances de l'algorithme de compression JPEG ont récemment été améliorées en utilisant une base orthonormale différente, appellée base d'ondelettes [7]. Le nouveau standard de la compression d'images utilisera donc ces nouvelles bases orthonormales.

image originale

1,28 bits/pixel

1,09 bits/pixel

0,4 bits/pixel

0,27 bits/pixel

0,18 bits/pixel

Figure 8.2: Images comprimées avec l'algorithme JPEG.

Bibliographie

[1] P. Brémaud, "Signaux aléatoires pour le traitement du signal et les communications", *Collection: Cours de l'Ecole Polytechnique, Ellipses*, 1993.

[2] P. Brémaud, "Introduction aux probabilités", Springer Verlag, Berlin.

[3] J.M. Bony, "Cours d'Analyse", *Ecole Polytechnique*, 1994.

[4] J.M. Bony, "Méthodes mathématiques pour les sciences physiques," *Ecole Polytechnique*, 1995.

[5] J.F. Genat, "Synthèse et traitement des sons en temps réel", *Ecole Polytechnique*, 1994.

[6] J.F Genat et A. Karar, "Introduction à l'analyse et synthèse de la parole", *Ecole Polytechnique*, 1994.

[7] S. Mallat, "A wavelet tour of signal processing", *Academic Press*, 1998.

[8] J. Neveu, "Introduction aux Probabilités", *Ecole Polytechnique*, 1994.

[9] A. Oppenheim et R. Schafer, "Discrete-time signal processing", *Prentice Hall*, 1989.

[10] T. Parsons, "Voice and speech processing", *Mc-Graw Hill*, 1987.

[11] A. Papoulis, "Signal analysis", *Mc Graw Hill*, 1977.

[12] M.B. Priestley, "Spectral analysis and time series", *Academic Press*, 1981.

[13] Y. Thomas, "Signaux et systèmes linéaires", *Masson*, 1994.

[14] B. Torrésani, "Analyse continue par ondelettes", *CNRS editions*, 1995.