

Ondelettes dyadiques et nouvelles représentations multirésolution

Avec l'explosion de la diversité des modes de consommation de contenus multimédia, il est souvent souhaitable de disposer d'un même média dans des résolutions et des qualités différentes. Les exemples sont légion : sur un site web, les imageries nous donnent un aperçu des images en grand format et permettent ainsi une présélection rapide du contenu. De même, on souhaitera disposer d'un morceau de musique numérisé de haute qualité lors d'un concert et on lui préférera une version prenant moins d'espace lors d'une écoute distraite sur baladeur.

En pratique, on utilise souvent plusieurs fichiers pour représenter différentes versions d'un même contenu multimédia. Cette solution coûteuse en bande passante n'est cependant pas satisfaisante. La scalabilité est la caractéristique d'un objet ou d'un algorithme à être représentable sur plusieurs niveaux de résolution ou de qualité. Nous justifions et motivons dans la section 1.1 comment cette propriété peut être mise à profit pour créer un *seul* flux compressé capable de représenter plusieurs versions d'un même contenu multimédia.

L'analyse multirésolution et la transformée en ondelettes sont des outils mathématiques capables de fournir une représentation scalable d'un signal. Tout d'abord, nous rappelons dans la section 1.2 les fondements mathématiques de ces outils. Nous décrivons ensuite comment la transformée en ondelettes peut être utilisée conjointement avec un schéma de codage emboîté afin de fournir un algorithme de compression d'image scalable et performant.

Il existe cependant d'autres transformations capables de fournir une représentation scalable d'un signal. La structure lifting, rappelée en section 1.3 permet d'étendre la théorie des ondelettes dans un cadre non-linéaire et autorise simplement la construction de transformées non-linéaires et inversibles. Nous étudions ensuite pourquoi les bases d'ondelettes séparables 2D ne sont pas bien adaptées à la représentation des images. Nous dressons alors un inventaire d'ondelettes géométriques fixes ou adaptatives appartenant à la famille des *Xlets*, permettent de contourner ce problème. Nous présentons enfin d'autres représentations multirésolution récentes, ayant pour but de fournir des bases mieux adaptées à la description scalable de signaux discontinus.

1.1 Introduction aux représentations multirésolution

Sur la scalabilité

Le terme scalabilité est un néologisme directement emprunté de l'anglais *scalability* qui peut approximativement être traduit par le terme échelonnabilité. La scalabilité est la

caractéristique d'un objet ou d'un signal à être représentable sur plusieurs niveaux de résolution ou de qualité. Une transformation sera ainsi dite scalable si elle est en mesure de représenter un signal sur plusieurs niveaux de résolution ou de qualité.

La notion de scalabilité est en fait très générale et il existe plusieurs types de scalabilité. Dans le cas d'un signal monodimensionnel, on parlera de scalabilité en résolution pour désigner le fait qu'un signal puisse être décrit par un nombre variable d'échantillons. Dans le cas d'une image, la scalabilité spatiale qualifie la propriété de pouvoir représenter une image sur plusieurs niveaux de résolution spatiale, comme illustré en Fig. 1.1.



FIG. 1.1 – Scalabilité spatiale. Exemples de facteurs de résolution dyadiques obtenus avec le codec scalable JPEG-2000.

Il est aussi possible de représenter un signal sur différents niveaux de qualité, où chaque échantillon ou coefficient peut être décrit avec une précision plus ou moins grande. On parlera dans ce cas de scalabilité en qualité. La Fig. 1.2 montre un exemple de scalabilité en qualité où chaque point de l'image est décrit avec plus ou moins de précision en fonction du débit qui lui est accordé.



FIG. 1.2 – Scalabilité en qualité. Exemples de différentes qualités obtenues avec le codec JPEG-2000 lors du décodage à différents débits, exprimés en bits par pixel.

Il existe d'autres types de scalabilité : dans le cas d'une séquence vidéo, on parlera de scalabilité temporelle pour désigner la propriété de pouvoir la représenter à plusieurs fréquences temporelles, exprimées en nombre d'images par seconde. D'autres types de scalabilité peuvent être définis comme la scalabilité en complexité, en objets ou en délai mais nous ne les aborderons pas dans ce document.

Motivation et cas d'utilisation

Avec l'explosion des applications multimédia et le besoin croissant de diffusion de contenu à destination de récepteurs hétérogènes, la scalabilité est devenue indispensable dans la conception d'un schéma de compression d'image ou de codage vidéo. Cette propriété permet ainsi de pouvoir diffuser un *unique* flux vidéo compressé, capable d'être adapté par les nœuds d'un réseau ou d'être décodé par une grande variété de récepteurs.

Il existe de nombreux cas d'utilisation nécessitant une description scalable et parcimonieuse d'un contenu multimédia, relevant pour la plupart du domaine de l'adaptation de contenu. Par exemple, les images présentes sur Internet sont souvent disponibles sous deux voire trois résolutions (aperçu *thumbnail*, résolution moyenne et haute résolution) en fonction de la façon dont elles sont visualisées. De plus, il est souvent nécessaire de posséder un morceau de musique compressé sous plusieurs débits, en fonction de la qualité désirée et de la place disponible. Enfin, les opérateurs commerciaux de diffusion de contenus multimédia ont tout intérêt à utiliser un format scalable. Un opérateur de téléphonie mobile pourra ainsi diffuser un flux vidéo TV destiné à un parc hétérogène de récepteurs dont les écrans sont de tailles différentes.

De plus, la scalabilité est une propriété très utile lors de la diffusion de contenu multimédia dans un environnement enclin aux erreurs de transmissions, comme les réseaux IP sans fil. En effet, elle permet l'adaptation du débit du flux compressé en fonction de la capacité du canal, susceptible de varier selon les conditions de transmission, et augmente la robustesse d'un schéma de codage en cas de pertes, d'erreurs ou d'encombrements.

Ces nombreux cas d'utilisation poussent depuis quelques années les organismes internationaux de normalisation ITU, JPEG et MPEG à concevoir des algorithmes de compression d'images et de codage vidéo *scalables*. La norme de compression d'images fixes JPEG-2000, scalable en résolution et en qualité a ainsi été normalisée en 2000. Le futur algorithme de codage vidéo scalable SVC décrit dans la section 2.1.4, est quant à lui en cours de normalisation. On trouvera dans l'appel à proposition [6] qui a précédé sa création, les nombreuses motivations industrielles et cas d'utilisation auxquels il répond.

Techniques classiques de description scalable

Comment créer une représentation scalable d'un signal ? On peut tout d'abord penser naïvement à une solution de type Simulcast. C'est une stratégie brutale qui consiste simplement à proposer *simultanément* plusieurs versions du même contenu multimédia. C'est en effet une forme de description scalable mais qui est loin d'être parcimonieuse.

Une autre classe de solutions existe : les schémas de codage prédictif en couches. Dans ce type de stratégie, chaque version du contenu multimédia constitue une couche et un mécanisme existe pour permettre la prédiction d'une couche à partir d'une autre, réduisant ainsi la redondance comparée à une stratégie Simulcast. Un exemple simple de schéma de codage prédictif en couches peut être imaginé par un format de musique compressé, où chaque couche représente une qualité différente, obtenue par différence avec la couche de qualité inférieure. Cependant, l'efficacité de ce type de représentation repose principalement sur l'opérateur de prédiction utilisé entre couches et est susceptible de chuter si le nombre de couches est trop important. Enfin, la scalabilité offerte par les schémas en couches est grossière et statique : seules les couches disponibles peuvent fournir une version du contenu compressé et ces dernières doivent être connues au moment de l'encodage.

La transformée en ondelettes est par construction même, une transformation capable de donner une représentation scalable d'un signal. En effet, les coefficients issus de la transformée en ondelettes donnent une représentation du signal sur plusieurs niveaux de résolution, du plus grossier au plus fin. Nous rappelons dans la section suivante les bases de l'analyse par ondelettes.

1.2 Ondelettes dyadiques

L'analyse multirésolution par ondelettes d'un signal, d'une image ou plus généralement d'une fonction $f \in L^2(\mathbb{R}^N)$ consiste en sa projection sur des bases de fonctions, donnant des approximations de moins en moins fines de la fonction originale. Nous rappelons tout d'abord dans cette section le concept d'analyse multirésolution par ondelettes et voyons ensuite comment un algorithme de transformée rapide en ondelettes peut être mis en œuvre sous forme de banc de filtres. Après avoir établi un panorama d'ondelettes classiquement utilisées en compression d'image, nous décrivons alors plusieurs algorithmes de codage de coefficients d'ondelettes utilisés dans ce domaine.

1.2.1 Bases d'ondelettes

Analyse multirésolution

La construction d'espaces multirésolution aptes à représenter plus ou moins grossièrement une fonction $f \in L^2(\mathbb{R})$ a été proposée par Mallat et Meyer [79, 80, 89] pour fournir un cadre formel permettant l'analyse d'une fonction f sur plusieurs niveaux de résolution. On définit une *approximation multirésolution* comme une suite de sous-espaces vectoriels fermés $\{V_j\}_{j \in \mathbb{Z}}$ de $L^2(\mathbb{R})$ emboîtés selon la relation :

$$\emptyset \subset \dots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \dots \subset L^2(\mathbb{R}) \quad (1.1)$$

La projection d'une fonction $f \in L^2(\mathbb{R})$ sur un espace V_j représente alors une *approximation* de f au niveau de résolution j . Du fait de l'emboîtement des espaces $\{V_j\}$, l'approximation de niveau j sera nécessairement plus précise que celle du niveau $j + 1$ car l'espace V_j dispose de plus de fonctions que l'espace V_{j+1} pour représenter f .

On suppose alors l'existence d'une fonction $\phi \in L^2(\mathbb{R})$, appelée *fonction d'échelle* ou *ondelette père* telle que ses translatées $\{t \mapsto \phi(t-k)\}_{k \in \mathbb{Z}}$ forment une base orthonormale de V_0 . On affirme enfin que les fonctions de V_{j+1} sont obtenues par dilatation d'un facteur 2 des fonctions de V_j selon la relation :

$$\forall j \in \mathbb{Z}, t \mapsto f(t) \in V_j \Leftrightarrow t \mapsto f\left(\frac{t}{2}\right) \in V_{j+1} \quad (1.2)$$

permettant ainsi de caractériser intuitivement les propriétés de l'analyse multirésolution et de supputer que l'approximation de f sur V_{j+1} est deux fois plus grossière que celle sur V_j . On peut alors introduire la notion d'échelle et définir la projection de f sur V_j comme l'approximation de f à l'échelle 2^j , où j est le niveau de résolution.

L'utilisation de la relation de dilatation (1.2) nous permet alors d'affirmer que les fonctions $\{\phi_{j,k}\}_{k \in \mathbb{Z}}$ obtenues par dilatations et translations de ϕ et définies par :

$$\phi_{j,k} = t \mapsto \frac{1}{2^{j/2}} \phi\left(\frac{t}{2^j} - k\right), \quad k \in \mathbb{Z} \quad (1.3)$$

forment une base orthonormale de V_j .

Bases d'ondelettes orthogonales

La relation d'emboîtement implique que les projections de f sur V_j sont de plus en plus grossières, au fur et à mesure que j croît. La différence entre l'approximation sur V_j et celle sur V_{j+1} représente ainsi l'information de détail perdue par incrémentation du niveau de résolution j . Il est cependant possible de définir l'espace de détail W_{j+1} contenant les fonctions nécessaires à représenter cette information perdue, en utilisant l'opérateur de sommation directe \oplus de sous-espaces vectoriels :

$$V_j = V_{j+1} \oplus W_{j+1} \quad (1.4)$$

On peut alors montrer l'existence d'une fonction ψ appelée *ondelette mère* telle que ses translatées $\{t \mapsto \psi(t - k)\}_{k \in \mathbb{Z}}$ forment une base orthonormale de W_0 . On montre de même que les fonctions $\{\psi_{j,k}\}_{k \in \mathbb{Z}}$ définies par :

$$\psi_{j,k} = t \mapsto \frac{1}{2^{j/2}} \psi\left(\frac{t}{2^j} - k\right), \quad k \in \mathbb{Z} \quad (1.5)$$

forment une base orthonormale de W_j . Enfin, en exploitant les conditions limites de l'analyse multirésolution, on conclut que l'ensemble des fonctions $\{\psi_{j,k}\}_{(j,k) \in \mathbb{Z}^2}$ forme une base d'ondelettes orthogonales de $L^2(\mathbb{R})$.

Si f est une fonction discrète alors pour toute fonction ϕ , il existe un niveau de résolution j suffisamment petit tel que f appartienne à V_j . On peut donc traduire le niveau de résolution et fixer $j = 0$ pour que f appartienne à V_0 . La transformée en ondelettes d'une fonction $f \in V_0$ sur n niveaux est alors définie comme la projection de cette fonction sur les espaces V_n et $\{W_j\}_{1 \leq j \leq n}$ car $V_0 = V_n \oplus \left[\bigoplus_{j=1}^n W_j \right]$. Les coefficients de projection sur V_j sont notés $a_j[k]$ et nommés *coefficients d'approximation* tandis que ceux sur W_j sont notés $d_j[k]$ et nommés *coefficients d'ondelette* ou *coefficients de détail*. On a alors :

$$a_j[k] = \langle f, \phi_{j,k} \rangle \quad (1.6)$$

$$d_j[k] = \langle f, \psi_{j,k} \rangle \quad (1.7)$$

où $\langle \cdot, \cdot \rangle$ représente le produit scalaire dans $L^2(\mathbb{R})$.

Ces relations nous permettent de calculer explicitement les coefficients de la transformée en ondelettes de f sur n niveaux. Cependant, l'intégration sur \mathbb{R} qu'elles nécessitent les rendent très lourdes à utiliser. Nous verrons dans la section 1.2.2 qu'il est possible de construire un algorithme de calcul rapide des coefficients $a_j[k]$ et $d_j[k]$.

Enfin, on remarquera que la transformée en ondelettes est une application linéaire, inversible et orthogonale. C'est donc une isométrie qui préserve la norme ℓ_2 , c'est à dire l'énergie d'un signal. On a alors $\sum_k a_j[k]^2 = \sum_k a_{j+1}[k]^2 + d_{j+1}[k]^2$.

1.2.2 Ondelettes et bancs de filtres

Filtres miroirs conjugués

L'espace V_{j+1} étant un sous-espace vectoriel de V_j , les fonctions de V_{j+1} peuvent être écrites comme une combinaison linéaire de fonctions de V_j . On peut donc exprimer la fonction $t \mapsto \frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right)$ appartenant à V_0 comme une combinaison linéaire des fonctions $\{t \mapsto \phi(t - k)\}_{k \in \mathbb{Z}}$ en introduisant la suite $h_0[k]$, $k \in \mathbb{Z}$:

$$\frac{1}{\sqrt{2}} \phi\left(\frac{t}{2}\right) = \sum_{k=-\infty}^{\infty} h_0[k] \phi(t - k) \quad (1.8)$$

De même, l'espace \mathbf{W}_{j+1} étant un sous-espace vectoriel de \mathbf{V}_j , il est possible de définir la fonction $t \mapsto \frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right)$ comme une combinaison linéaire des fonctions $\{t \mapsto \phi(t-k)\}_{k \in \mathbb{Z}}$ en introduisant la suite $h_1[k], k \in \mathbb{Z}$:

$$\frac{1}{\sqrt{2}}\psi\left(\frac{t}{2}\right) = \sum_{k=-\infty}^{\infty} h_1[k]\phi(t-k) \quad (1.9)$$

Les relations (1.8) et (1.9) sont aussi appelées *équations à deux échelles*. Enfin, la définition des suites h_0 et h_1 permet de montrer [79] qu'une condition suffisante assurant l'existence de ψ peut s'exprimer par la relation suivante, aussi appelée condition d'orthogonalité :

$$h_1[n] = (-1)^{1-n}h_0[1-n] \quad (1.10)$$

Transformée en ondelettes rapide

Mallat a montré [79] l'existence d'équations liant les coefficients d'approximation $a_j[k]$ et les coefficients d'ondelettes $d_j[k]$ obtenus entre deux niveaux de résolution consécutifs. En effet, par combinaison de (1.6), (1.7), (1.8) et (1.9), on vérifie aisément que :

$$a_{j+1}[k] = \sum_{n=-\infty}^{\infty} h_0[n-2k]a_j[n] = a_j \star \bar{h}_0[2k] \quad (1.11)$$

$$d_{j+1}[k] = \sum_{n=-\infty}^{\infty} h_1[n-2k]a_j[n] = a_j \star \bar{h}_1[2k] \quad (1.12)$$

où \star est le produit de convolution et \bar{h} dénote le retournement temporel du filtre h , où pour tout n , $\bar{h}[n] = h[-n]$. Les relations (1.11) et (1.12) permettent ainsi de calculer les coefficients de projection $a_{j+1}[k]$ et $d_{j+1}[k]$ à partir des seuls coefficients $a_j[k]$ et des suites h_0 et h_1 précédemment définies.

La présence du produit de convolution nous suggère l'utilisation d'un opérateur de filtrage, classiquement utilisé en traitement du signal. Les équations (1.11) et (1.12) font ainsi le lien entre la transformée en ondelettes définie précédemment comme la projection d'un signal dans les espaces \mathbf{V}_j et $\{\mathbf{W}_j\}_{1 \leq j \leq n}$ et son interprétation en termes de bancs de filtres. Les séquences $h_0[k]$ et $h_1[k]$ peuvent alors s'identifier respectivement aux réponses impulsionnelles d'un filtre passe-bas et d'un filtre passe-haut d'un banc d'analyse.

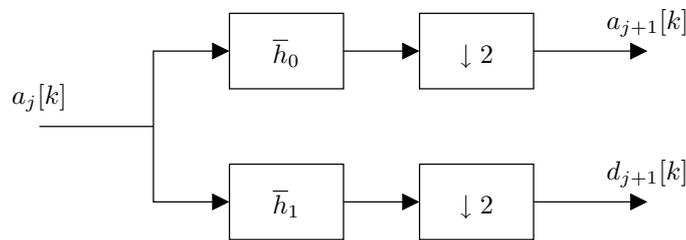


FIG. 1.3 – Banc de filtres d'analyse en quadrature miroir.

Il est alors envisageable de construire un algorithme de calcul rapide des coefficients $a_{j+1}[k]$ et $d_{j+1}[k]$ par filtrage des coefficients $a_j[k]$, d'une part par le filtre \bar{h}_0 et d'autre part par le filtre \bar{h}_1 , suivi par la décimation d'un facteur 2. Cette dernière opération, aussi appelée sous-échantillonnage d'un facteur 2, est notée $\lfloor \downarrow 2$ et consiste à se débarrasser d'un

coefficient sur deux. L'algorithme peut être représenté par un banc de filtres d'analyse en quadrature miroir [155] et est illustré par la Fig. 1.3.

Ce banc de filtres d'analyse permet ainsi l'implémentation effective de la transformée en ondelettes rapide. En effet, en supposant connus les filtres h_0 et h_1 , la décomposition en ondelettes d'un signal x d'une longueur de n échantillons consiste à initialiser¹ $a_0[k] = x[k]$ et à utiliser le banc de filtres. On obtient alors les sous-bandes a_1 et d_1 , comportant chacune $n/2$ échantillons. Ces signaux sont les coefficients de la projection de x sur les espaces V_1 et W_1 . L'analyse multirésolution sur un nombre supérieur de niveaux s'obtient par la décomposition successive des signaux a_j et donc par une mise en cascade du banc de filtres jusqu'au niveau j_{max} désiré, comme illustré par la Fig. 1.4.

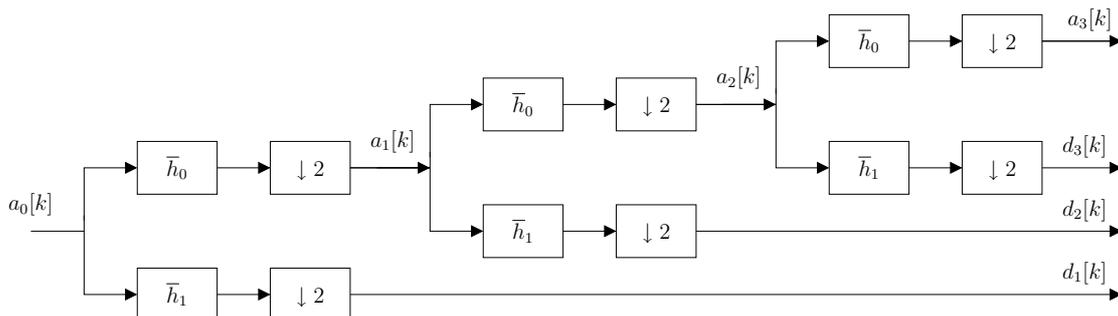


FIG. 1.4 – Banc de filtres d'analyse assurant une décomposition en ondelettes sur $j_{max} = 3$ niveaux de résolution. Elle correspond à une projection sur V_3, W_3, W_2 et W_1 .

L'algorithme présenté permet ainsi de calculer rapidement la transformée en ondelettes d'un signal donné, sous la connaissance des filtres h_0 et h_1 . Sa complexité est de $\mathcal{O}(n)$ opérations élémentaires où n est la taille du signal, rivalisant ainsi avec la transformée de Fourier rapide. Cependant, sa mise en œuvre sur un signal fini nécessite généralement l'utilisation d'une convolution périodique qui crée des coefficients d'ondelettes de large amplitude sur les bords du signal et nuit ainsi légèrement à l'efficacité de décorrélacion. Les ondelettes de bords [79] et surtout la structure lifting présentée en section 1.3.1 permet de s'affranchir simplement de ce problème.

Reconstruction par transformée inverse

Comme pour la transformée directe décrite par les équations (1.11) et (1.12), il est possible de montrer la relation suivante, utile pour la reconstruction du signal original :

$$\begin{aligned} a_j[p] &= \sum_{n=-\infty}^{\infty} h_0[p-2n]a_{j+1}[n] + \sum_{n=-\infty}^{\infty} h_1[p-2n]d_{j+1}[n] \\ &= \hat{a}_{j+1} \star h_0 + \hat{d}_{j+1} \star h_1 \end{aligned} \quad (1.13)$$

où \hat{h} est le signal résultant du suréchantillonnage de h d'un facteur 2. Cette opération consiste en l'introduction de zéros entre les échantillons du signal d'origine : elle est définie pour tout n par $\hat{h}[2n] = h[n]$ et $\hat{h}[2n+1] = 0$, et se note $[\uparrow 2]$.

Cette relation nous permet alors de construire un algorithme rapide de reconstruction du signal a_j à partir de ses coefficients d'approximation $a_{j+1}[k]$ et de ses coefficients

¹Nous ne détaillerons pas ici les conditions sur le signal continu sous lesquelles cette relation est vraie.

d'ondelette $d_{j+1}[k]$ du niveau supérieur. En prenant $a_0[k] = x[k]$, on montre ainsi qu'on peut reconstruire parfaitement le signal $x[k]$ à partir de ses coefficients $a_1[k]$ et $d_1[k]$. L'algorithme de reconstruction peut être représenté par un banc de filtres, nommé banc de filtres de synthèse et est illustré en Fig. 1.5.

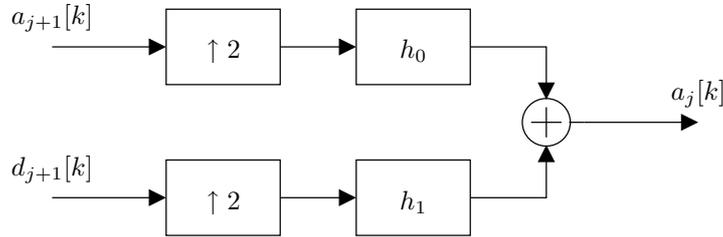


FIG. 1.5 – Banc de filtres de synthèse.

Nous avons donc mis en évidence l'existence d'algorithmes rapides de transformation en ondelettes et de reconstruction sous formes de banc de filtres. Leur mise en œuvre nécessitent la seule connaissance du filtre passe-bas h_0 , le filtre passe-haut h_1 étant obtenu grâce à la condition d'orthogonalité (1.10).

Cas 2D et ondelettes séparables

Les résultats précédents ont été obtenus pour des fonctions monodimensionnelles appartenant à $L^2(\mathbb{R})$. L'extension aux dimensions supérieures peut se faire simplement par utilisation du produit tensoriel de bases d'ondelettes monodimensionnelles, conduisant à des bases d'ondelettes séparables. Dans le cas 2D, la transformée rapide en ondelettes consiste alors en une transformée en ondelettes 1D des colonnes puis des lignes de l'image (ou inversement). Un exemple d'une telle décomposition est illustré en Fig. 1.11.

Bases d'ondelettes adaptées à la compression d'un signal

Comment construire une base d'ondelettes adaptée à la compression d'un signal ? Il existe en effet de nombreuses fonctions d'échelles ϕ et donc d'ondelettes mère ψ , vérifiant les propriétés nécessaires à la construction de bases d'ondelettes. En fait, nous nous intéressons à la compression d'images et de séquences vidéos et souhaitons disposer ainsi d'une base permettant une représentation *parcimonieuse* d'un signal, c'est à dire donnant peu de coefficients d'ondelettes de grande amplitude. Il est donc souhaitable d'imposer des contraintes sur l'ondelette mère ψ afin de favoriser son aptitude à décorréler ce type de signaux.

Tout d'abord, il est fortement souhaitable que l'ondelette ψ soit à support fini. Ceci permet en effet la mise en œuvre simple de la transformée en ondelettes rapide. La symétrie de l'ondelette est aussi un critère important en compression d'images, permettant de donner un poids équivalent aux pixels lors de leur traitement et de préserver la linéarité de la phase.

Enfin, il est utile que l'ondelette possède un grand nombre de moments nuls. Ce paramètre important caractérise l'aptitude d'une ondelette à approximer les polynômes. On dit que ψ possède N moments nuls si et seulement si :

$$\forall 0 \leq n < N \quad \int_{\mathbb{R}} \psi(t) t^n dt = 0 \quad (1.14)$$

Ceci signifie que ψ est orthogonale à tout polynôme de degré inférieur ou égal à $N - 1$. Ainsi, si f est un signal localement polynomial de degré inférieur ou égal à $N - 1$, alors les coefficients de détail $d_j[k]$ résultant de la transformation en ondelettes seront localement nuls. Comme une image est bien modélisée par des fonctions polynomiales par morceaux, sa transformée en ondelettes avec un nombre suffisant de moments nuls est susceptible de contenir de nombreux coefficients d'ondelettes proches de zéro, correspondants aux régions où l'image présente un comportement polynomial.

Banc de filtres à reconstruction parfaite et ondelettes biorthogonales

Nous avons abordé dans les sections précédentes le cas des ondelettes orthogonales liées par la condition (1.10). La seule connaissance du filtre h_0 nous a ainsi permis de mettre en œuvre un algorithme de transformée rapide en ondelettes. Est-il cependant possible de construire une transformée par bancs de filtres plus générale en omettant cette condition ?

Considérons la structure décrite en Fig. 1.6, combinant un banc de filtres d'analyse h_0 et h_1 et un banc de synthèse dont les réponses impulsionnelles sont \tilde{h}_0 et \tilde{h}_1 . Ces quatre filtres sont volontairement supposés indépendants. On s'intéresse aux conditions nécessaires et suffisantes, nommées conditions de reconstruction parfaite, liant ces filtres et assurant que le signal reconstruit \tilde{x} soit strictement égal au signal d'entrée x .

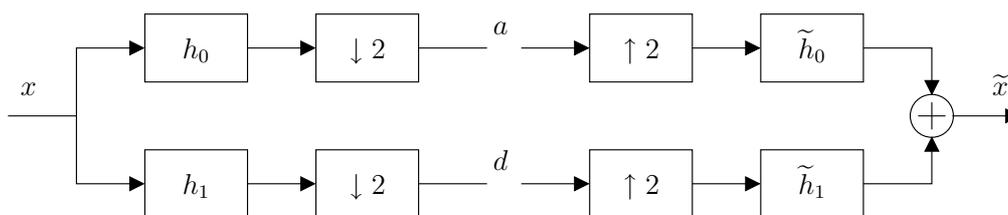


FIG. 1.6 – Banc de filtres d'analyse-synthèse.

L'utilisation de la transformée de Fourier permet de formuler simplement le problème. En notant $\hat{h}(f)$ la transformée de Fourier en fréquence normalisée f du filtre h , on montre que les conditions de reconstruction parfaite s'écrivent dans le domaine fréquentiel :

$$\hat{h}_0(f)\hat{\tilde{h}}_0(f) + \hat{h}_1(f)\hat{\tilde{h}}_1(f) = 2 \quad (1.15)$$

$$\hat{h}_0(f + 1/2)\hat{\tilde{h}}_0(f) + \hat{h}_1(f + 1/2)\hat{\tilde{h}}_1(f) = 0 \quad (1.16)$$

Ces équations imposent donc des conditions sur les réponses fréquentielles des filtres h_0 , h_1 , \tilde{h}_0 et \tilde{h}_1 , et autorisent leur construction uniquement dans le domaine de Fourier. Les conditions de reconstruction parfaite peuvent cependant s'écrire aussi dans le domaine temporel sous la forme concise suivante, en utilisant l'opérateur de Kronecker δ :

$$\forall 0 \leq i, j \leq 1, \forall n \in \mathbb{Z}, \sum_k h_i[k]\tilde{h}_j[k - 2n] = \delta_n \delta_{i-j} \quad (1.17)$$

En supposant la connaissance des filtres passe-bas d'analyse h_0 et de synthèse \tilde{h}_0 , les conditions de reconstruction parfaite imposent les coefficients des filtres passe-haut :

$$h_1[n] = (-1)^{1-n}\tilde{h}_0[1 - n] \quad (1.18)$$

$$\tilde{h}_1[n] = (-1)^{1-n}h_0[1 - n] \quad (1.19)$$

et permettent ainsi de caractériser entièrement la transformée par ses filtres passe-bas d'analyse h_0 et de synthèse \tilde{h}_0 . De plus, ces conditions impliquent que les familles $\{h_0[k-2n], h_1[k-2n]\}_{n \in \mathbb{Z}}$ et $\{\tilde{h}_0[k-2n], \tilde{h}_1[k-2n]\}_{n \in \mathbb{Z}}$ soient biorthogonales entre elles, ce qui permet de donner une interprétation en termes d'ondelettes au banc de filtres à reconstruction parfaite. Elle revient à lever les contraintes d'orthogonalité imposées aux bases V_j et W_j à les remplacer par des contraintes de biorthogonalité. Ces dernières conduisent alors à l'introduction de bases duales \tilde{V}_j et \tilde{W}_j et de leur fonctions duales $\tilde{\phi}$ et $\tilde{\psi}$ associées, utilisées lors de la reconstruction. On parlera alors d'ondelettes biorthogonales et de transformée en ondelettes biorthogonales.

1.2.3 Panorama d'ondelettes dyadiques utilisées en codage d'image

Il existe de nombreuses ondelettes dyadiques décrites dans la littérature [79, 109] (Spline, Shannon-Nyquist, Daubechies, etc...) utilisées en codage, débruitage ou analyse de signaux. Nous présentons ici quelques ondelettes couramment utilisées en codage d'image.

Ondelette de Haar

L'ondelette de Haar est une ondelette orthogonale symétrique possédant un seul moment nul $N = 1$ et un support de $p = 2$ échantillons. Les coefficients de la réponse impulsionnelle de son filtre passe-bas h_0 sont présentés dans la partie gauche du Tab. 1.1. Daubechies a montré que c'est la seule ondelette orthogonale et symétrique correspondant à un banc de filtres à réponse impulsionnelle finie. De part sa simplicité, cette ondelette illustrée en Fig. 1.7 est assez utilisée en codage d'image.

n	$h_0[n]$	n	$h_0[n]$
0	0.70710678118655	0	0.48296291314483
1	0.70710678118655	1	0.83651630373771
		2	0.22414386804192
		3	-0.12940952255095

TAB. 1.1 – Coefficients de la réponse impulsionnelle du filtre passe-bas d'analyse $h_0[n]$ associé à l'ondelette de Haar (gauche) et à l'ondelette de Daubechies-4 (droite).

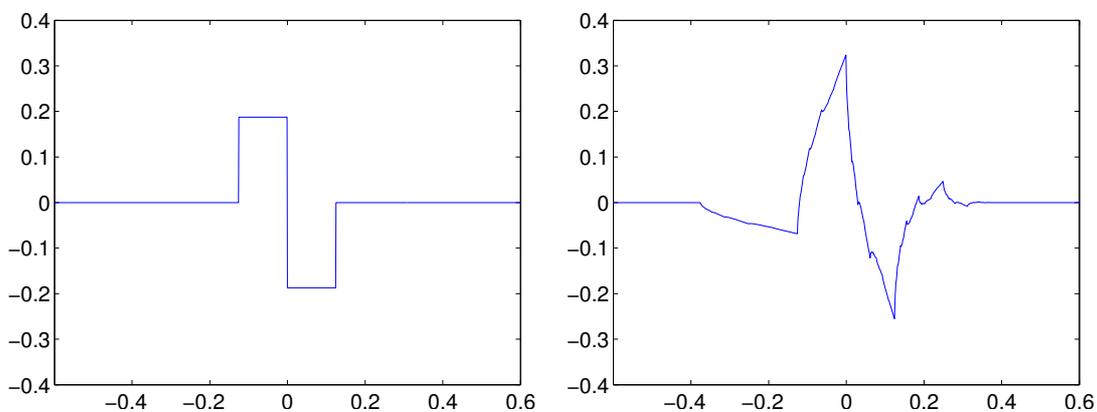


FIG. 1.7 – Ondelette de Haar (gauche) et ondelette Daubechies-4 (droite).

Ondelette Daubechies-4

L'ondelette Daubechies-4 fait partie de la famille des ondelettes orthogonales de Daubechies, possédant un support de $2N$ échantillons pour N moments nuls. L'ondelette Daubechies-4 possède donc $N = 2$ moments nuls et un support de $p = 4$ échantillons. Les coefficients de la réponse impulsionnelle de son filtre passe-bas h_0 sont dressés dans la partie droite du Tab. 1.1. On peut montrer que les ondelettes de cette famille ont une largeur de support minimale et un déphasage minimal pour un nombre de moments nuls donné. Elles sont pourtant peu utilisées en codage car elles sont fortement asymétriques et très irrégulières. Cependant, l'ondelette Daubechies-4 illustrée en Fig. 1.7 est courte, orthogonale et suffisamment régulière pour susciter un intérêt en codage.

Ondelettes biorthogonales 5/3

Les ondelettes biorthogonales 5/3 font partie de la famille des ondelettes biorthogonales symétriques de Cohen-Daubechies-Feauveau (CDF) [79]. Elles sont dénommées ainsi car la largeur du support de leurs filtres passe-bas, détaillés dans le Tab. 1.2, est de $p = 5$ échantillons à l'analyse et $\tilde{p} = 3$ à la synthèse. De plus, elles possèdent $N = \tilde{N} = 2$ moments nuls. De part leur relative simplicité et la symétrie qu'elles offrent, les ondelettes 5/3 présentées en Fig. 1.8 sont assez utilisées en codage d'image.

Les ondelettes de cette famille sont aussi dénommées CDF (N, \tilde{N}) , où N désigne le nombre de moments nuls de l'ondelette d'analyse ψ et \tilde{N} son équivalent à la synthèse. Comme pour les ondelettes de Daubechies, il est possible de montrer que les ondelettes CDF ont un support minimal pour un nombre de moments nuls (N, \tilde{N}) donnés.

n	$h_0[n]$	$\tilde{h}_0[n]$
0	1.06066017177982	0.70710678118655
1	0.35355339059327	0.35355339059327
2	-0.17677669529664	

TAB. 1.2 – Coefficients des réponses impulsionnelles symétriques des filtres passe-bas d'analyse $h_0[n]$ et de synthèse $\tilde{h}_0[n]$ associés aux ondelettes CDF 5/3.

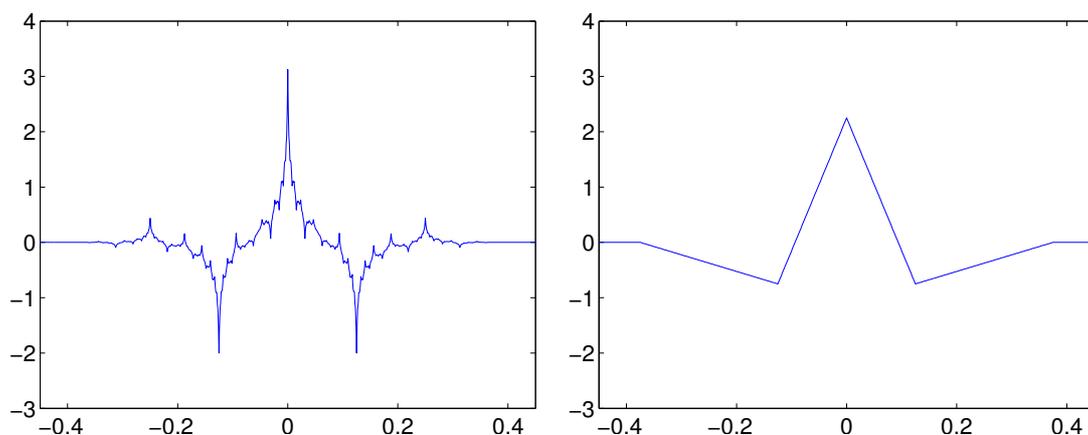


FIG. 1.8 – Ondelette CDF 5/3 d'analyse ψ et sa duale $\tilde{\psi}$.

Ondelettes biorthogonales 9/7

Tout comme les ondelettes 5/3, les ondelettes biorthogonales 9/7 font partie de la famille des ondelettes biorthogonales symétriques CDF. Les filtres passe-bas associés aux ondelettes 9/7 possèdent ainsi $p = 9$ coefficients à l'analyse, $p = 7$ coefficients à la synthèse et sont décrits dans le Tab. 1.3. Les ondelettes biorthogonales 9/7 sont illustrées en Fig. 1.9 et possèdent $N = 4$ moments nuls à l'analyse et $\tilde{N} = 4$ à la synthèse.

Les ondelettes 9/7 possèdent un grand nombre de moments nuls pour un support relativement court. Elles sont de plus symétriques et très proches de l'orthogonalité. C'est une caractéristique importante en codage qui lui permet d'assurer que l'erreur de reconstruction soit très proche de l'erreur de quantification, en terme d'erreur quadratique moyenne. Antonini et Barlaud furent les premiers [18] à montrer la supériorité de la transformée en ondelettes biorthogonale 9/7 pour la décorrélation d'images naturelles. Elle est depuis très utilisée en codage d'image [122, 159] et est utilisée par le codec JPEG-2000 [5, 139]. Une étude assez complète des propriétés théoriques des ondelettes biorthogonales 5/3 et 9/7 est présentée dans [153].

n	$h_0[n]$	$\tilde{h}_0[n]$
0	0.85269867900940	0.78848561640566
1	0.37740285561265	0.41809227322221
2	-0.11062440441842	-0.04068941760956
3	-0.02384946501938	-0.06453888262894
4	0.03782845550699	

TAB. 1.3 – Coefficients des réponses impulsionnelles symétriques des filtres passe-bas d'analyse $h_0[n]$ et de synthèse $\tilde{h}_0[n]$ associés à l'ondelette CDF 9/7.

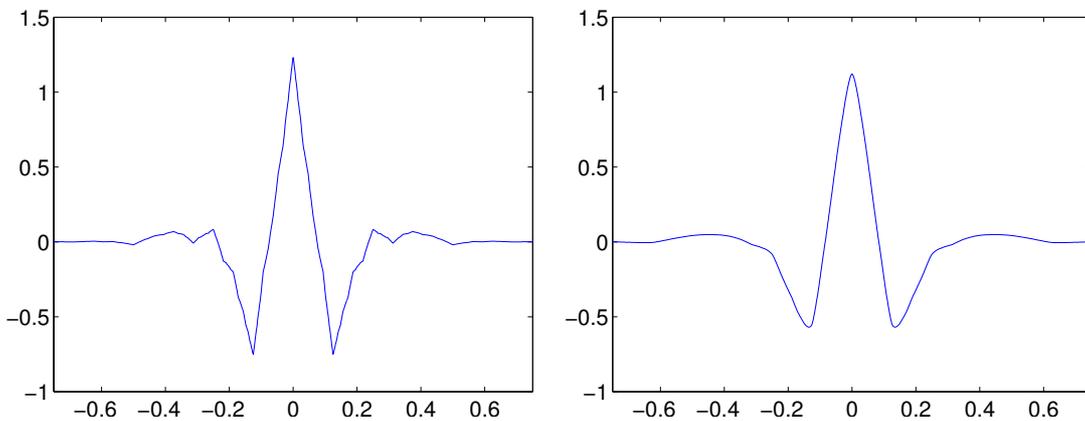


FIG. 1.9 – Ondelette CDF 9/7 d'analyse ψ et sa duale $\tilde{\psi}$.

1.2.4 Compression d'image par transformée en ondelettes

Décomposition en ondelettes séparables

La transformation de signaux multidimensionnels peut être réalisé de façon séparable par transformations successives de l'image sur ses dimensions. Ainsi, la décomposition

en ondelettes des lignes puis des colonnes d'une image permet d'obtenir sa décomposition 2D sur un puis sur plusieurs niveaux, comme illustré en Fig. 1.10.

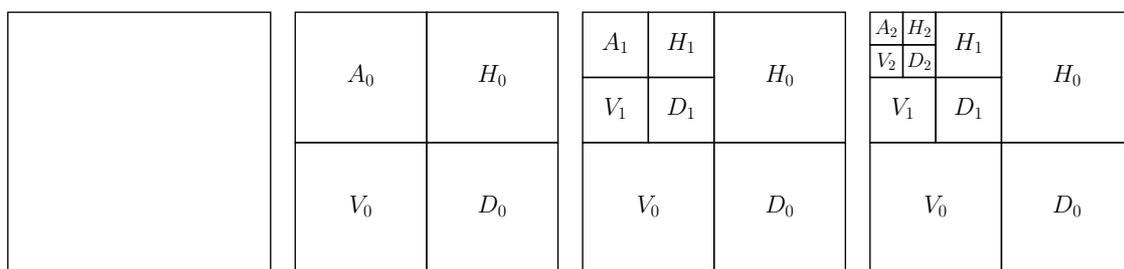


FIG. 1.10 – Décomposition successive d'une image en ondelettes sur trois niveaux. On remarquera la disposition pyramidale des sous-bandes d'approximation A_k de niveau k et des sous-bandes de détail horizontal H_k , vertical V_k et diagonal D_k .

L'image originale est tout d'abord décomposée en une sous-bande d'approximation A_0 et en sous-bandes de détails H_0 , V_0 et D_0 , correspondant respectivement aux détails horizontaux, verticaux et diagonaux. La décomposition successive des sous-bandes d'approximation A_k permet d'obtenir alors l'analyse multirésolution de l'image, qui se présente sous une forme pyramidale. La Fig. 1.11 illustre les sous-bandes obtenues par la décomposition en ondelettes 9/7 séparable de l'image *Lena* sur 3 niveaux (le contraste de l'image a été augmenté afin que les coefficients de détail soient visibles).

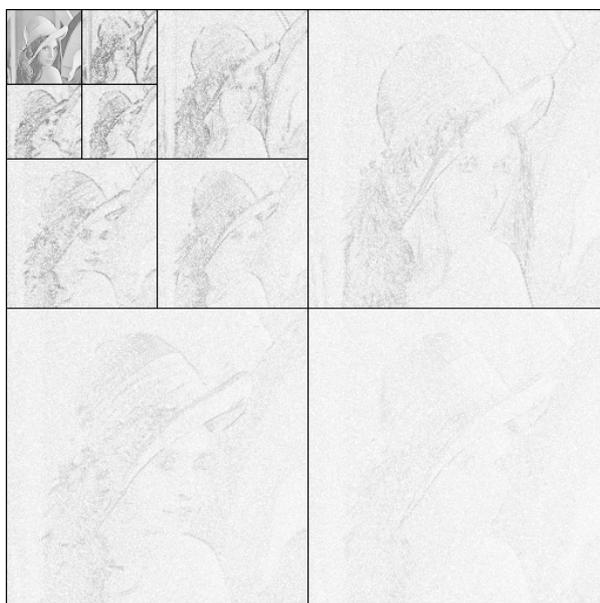


FIG. 1.11 – Décomposition en ondelettes séparables biorthogonales 9/7 de *Lena* sur 3 niveaux de résolution.

Nous détaillons dans les sous-sections suivantes les principaux algorithmes utilisés pour encoder efficacement et de façon scalable les coefficients issus de la décomposition en ondelettes d'une image.

EZW

L'algorithme de codage emboîté par ondelettes à arbres de zéros EZW (*Embedded Zero-tree Wavelet*) a été proposé par Shapiro [130] en 1993. Il permet un codage efficace des coefficients d'ondelettes tout en assurant une scalabilité spatiale dyadique et une scalabilité fine en qualité. L'algorithme consiste en un codage progressif par plans de bits (*bitplanes*) des coefficients de la pyramide de décomposition spatiale, en tenant compte de leur dépendance hiérarchique entre niveaux de résolutions, comme illustré par la Fig. 1.12.

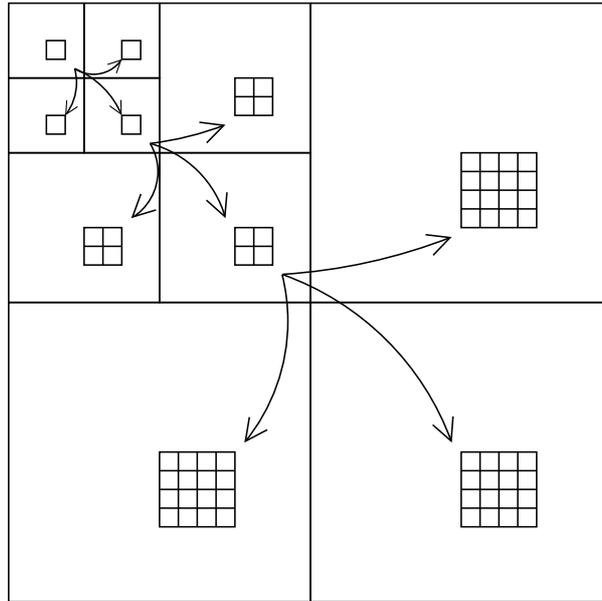


FIG. 1.12 – Prise en compte par le codec emboîté EZW de la dépendance hiérarchique spatiale des coefficients d'ondelettes entre niveaux de résolutions.

La pyramide 2D est encodée de façon progressive par plans de bits, par l'algorithme suivant. On considère à l'initialisation un seuil $T = 2^n$ où n est une valeur suffisamment grande pour que $T + T/2$ soit supérieur à l'amplitude maximale de tous les coefficients de la pyramide. Chaque itération i est composée de deux passes : une passe de description des coefficients significatifs et une passe de raffinement des valeurs des coefficients significatifs. La pyramide est parcourue sous-bande par sous-bande, en commençant par la sous-bande d'approximation et en progressant vers les hauts niveaux de résolution. Durant la première passe, chaque coefficient x de la pyramide est codé ou pas par un symbole et est rajouté au flux binaire. Tout d'abord, si $|x| > T$ alors le coefficient est déclaré comme significatif et est codé par le symbole P si il est positif et par le symbole N sinon. Si au contraire $|x| < T$, le coefficient n'est pas déclaré comme significatif et deux cas se présentent. Si x possède des descendants dans la pyramide dont la valeur absolue est supérieure à T alors il est codé par le symbole I . Sinon, il est codé par le symbole Z , comme *Zerotree root*, nœud racine d'un arbre de zéros, et aucun de ses descendants ne sera codé durant cette passe. Durant la passe de raffinement, les $(n - i)$ -ième bits de poids forts de tous les coefficients significatifs sont encodés. Le seuil T est alors divisé par 2 et on relance une itération, en ne considérant durant la première passe que les coefficients déclarés comme non-significatifs. Le flux binaire ainsi formé constitue un fichier aisément décompressable, scalable en résolution spatiale et en qualité. La scalabilité en qualité est

obtenue par troncation des plans de bits de poids faible tandis que la scalabilité spatiale est assurée par la suppression des zones codant les sous-bandes de détail.

Le schéma de codage EZW est simple et très efficace. Il offre des courbes de débit-distorsion meilleures que le codec JPEG, produit des images de qualité visiblement supérieure et offre une scalabilité spatiale et en débit.

SPIHT

L'algorithme SPIHT (*Set Partitioning In Hierarchical Tree*), proposé par Said et Pearlman [122] est une amélioration du schéma de codage EZW. Il repose sur les mêmes concepts : codage progressif par plans de bits et utilisation des dépendances hiérarchiques qu'entretiennent les coefficients d'une pyramide de décomposition 2D. L'algorithme est cependant plus sophistiqué : contrairement à l'algorithme EZW qui n'utilise qu'un seul ensemble décrivant la signifiante des coefficients, SPIHT met en jeu une liste des ensembles insignifiants (LIS), une liste des coefficients insignifiants (LIP) et une liste des coefficients significatifs (LSP). Tout comme EZW, SPIHT utilise une passe de description des coefficients significatifs et une passe de raffinement. Enfin, du fait de sa meilleure modélisation de la signifiante des coefficients, l'algorithme SPIHT offre une meilleure efficacité de codage que EZW.

EZBC

Le codec EZBC [60] de Hsiang et Woods (*Embedded ZeroBlocks with Context modeling*) est basé sur le codec SPIHT en utilisant un codeur arithmétique contextuel, prenant en compte le voisinage du coefficient courant. L'algorithme EZBC est ainsi capable de prendre en compte les relations de dépendance statistique existant entre les coefficients d'une même sous-bande. Il offre une efficacité de codage supérieure au codec SPIHT et est à la base du schéma de codage vidéo MC-EZBC décrit dans le chapitre suivant.

EBCOT et la norme JPEG-2000

Le codec EBCOT (*Embedded Block Coding with Optimized Truncation*) est issu des travaux de Taubman [138] mais n'appartient pas à la famille des schémas de codage à arbre de zéros. Dans EBCOT, chaque sous-bande est tout d'abord découpée en petits blocs indépendants : les *codeblocks*. Ces derniers sont alors codés de façon progressive, par plans de bits et au moyen d'un codeur arithmétique contextuel. Enfin, une procédure d'optimisation débit-distorsion est utilisée pour déterminer le choix optimal des *codeblocks* à conserver pour assurer une qualité maximale pour un ensemble de débits donnés et connus à l'encodage : les points de troncatures. Les débits intermédiaires restent toutefois accessibles mais n'offriront peut-être pas une qualité de reconstruction optimale. Le codec EBCOT est très performant et offre une efficacité de codage comparable à EZBC. L'algorithme utilisé dans la norme JPEG-2000 [139] est largement basé sur ce codec.

Codeurs morphologiques - EMDC

Après avoir observé la tendance qu'ont les coefficients d'ondelettes à s'agglomérer près des contours d'une image, Servetto et Ramchandran [129] ont utilisé un opérateur morphologique de dilatation au sein d'un codeur d'images fixes, afin de modéliser ces agglutinations. Les auteurs obtiennent des résultats expérimentaux comparables avec le

codeur emboîté SPIHT et observent une amélioration visuelle notable des images décodées. Dans la continuation de ces travaux, Lazzaroni, Leonardi et Signoroni ont récemment proposé le schéma de codage d'image EMDC (*Embedded Morphological Dilation Coding*) [71], combinant les avantages de l'approche morphologique et des codeurs à arbre de zéros. Cet algorithme possède une efficacité de codage supérieure aux codeurs emboîtés et ouvre des perspectives intéressantes sur l'utilisation de modèles morphologiques pour mieux capturer l'information géométrique contenue dans les images.

1.3 Nouvelles représentations multirésolution

L'analyse multirésolution par ondelettes exposée dans la section précédente permet d'obtenir une représentation scalable d'un signal. Cependant, cette représentation est nécessairement linéaire et ceci constitue un inconvénient majeur dans la description de signaux réels comme les images naturelles où les discontinuités, contours et autres singularités sont nombreux. Afin de pallier à ce problème, de nouvelles décompositions multirésolution mieux adaptées à la représentation de tels signaux ont été introduites.

Nous présentons tout d'abord le schéma *lifting* de Sweldens. C'est une structure de décomposition multirésolution toujours inversible, capable de représenter n'importe quelle transformée en ondelettes dyadique basée sur des bancs de filtres à réponse impulsionnelle finie. De plus, elle autorise la construction de transformées multirésolution non-linéaires, de façon très naturelle. La structure *lifting* est à la base de nombreuses décompositions multirésolution non-linéaires dont on donnera quelques exemples.

De plus, bien que les ondelettes soient des outils adaptés à la description des discontinuités de signaux monodimensionnels, cette propriété n'est plus vraie pour des dimensions supérieures. Les ondelettes séparables sont en effet isotropes et ne peuvent pas capturer par exemple la régularité présente le long d'un contour d'une image. De nombreuses constructions adaptées aux images ont été proposées pour tenir compte de ce problème, nommées ondelettes géométriques, que nous décrivons dans la suite du document. Nous aborderons enfin d'autres représentations multirésolution récentes, palliant à certaines faiblesses inhérentes aux bases d'ondelettes.

1.3.1 Structure *lifting*

La formulation en banc de filtres et l'algorithme de transformée en ondelettes rapide, décrits dans la section 1.2.2, permettent une réalisation effective de la transformée en ondelettes discrète. Cependant, il n'est pas aisé de construire et de déterminer les filtres passe-bas h_0 et passe-haut h_1 mis en jeu dans la transformée à partir des équations de reconstruction parfaite (1.15) et (1.16).

C'est en cherchant à améliorer les propriétés d'un banc de filtres à reconstruction parfaite par l'ajout de nouveaux étages que Sweldens [136] a mis en évidence la structure dite en *lifting*. Cette structure possède le triple avantage de pouvoir représenter les transformées en ondelettes dyadiques, de toujours assurer une reconstruction parfaite et d'être suffisamment flexible pour construire de nouvelles transformées.

Définition

On appelle une structure en *lifting* dyadique, un banc dans lequel un signal d'entrée $x[k]$ est tout d'abord séparé en deux composantes : généralement, sa composante paire

$x[2k]$ et sa composante impaire $x[2k + 1]$. Grâce à un nombre fini d'étages, on ajoute alors successivement sur une des composantes le résultat d'un opérateur appliqué sur l'autre composante. Ces opérateurs se nomment généralement prédicteur ou opérateur de *prédiction* et opérateur de *mise à jour*, selon le type d'opération qu'ils effectuent sur le signal. Les composantes finales sont alors multipliées par des constantes arbitraires.

Dans le cas particulier d'une structure lifting à deux étages, utilisant une étape de séparation en composantes polyphases ("lazy wavelets") du signal d'entrée, un opérateur de prédiction suivi d'un opérateur de mise à jour, on obtient le schéma d'analyse présenté en Fig. 1.13, en utilisant le signal $a_j[k] = x[k]$. L'ensemble des opérations effectuées peut alors s'écrire sous la forme suivante :

$$\begin{aligned} d_{j+1}^0[k] &= a_j[2k + 1] + P(\{a_j[2k]\}_{k \in \mathbb{Z}}) \\ a_{j+1}^0[k] &= a_j[2k] + U(\{d_{j+1}^0[k]\}_{k \in \mathbb{Z}}) \\ d_{j+1}[k] &= \alpha d_{j+1}^0[k] \\ a_{j+1}[k] &= \beta a_{j+1}^0[k] \end{aligned}$$

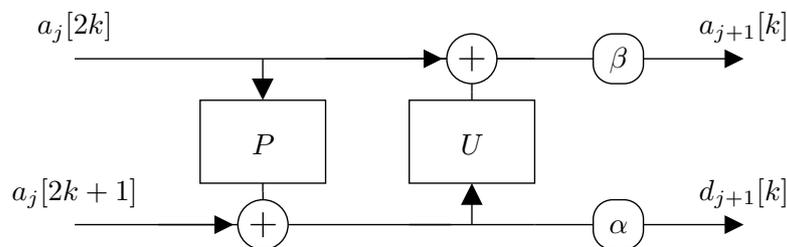


FIG. 1.13 – Structure d'analyse en lifting à deux étages.

Dans cette structure, le signal d'entrée a_j est tout d'abord séparé en deux composantes, nommées aussi sous-bandes : la composante paire $a_j[2k]$ et la composante impaire $a_j[2k + 1]$. On ajoute alors à la sous-bande impaire le résultat de l'opérateur de prédiction P appliqué à la composante paire, donnant ainsi le signal $d_{j+1}^0[k]$. L'opérateur P est généralement choisi de façon à minimiser l'amplitude de la sous-bande de détail et se borne ainsi à prédire la sous-bande impaire à partir de la sous-bande paire.

Le signal $d_{j+1}^0[k]$, aussi nommé résidu de prédiction, est alors mis à jour au moyen de l'opérateur U et est ajouté à la sous-bande $a_j[2k]$ pour obtenir le signal $a_{j+1}^0[k]$. Enfin, les signaux $a_{j+1}^0[k]$ et $d_{j+1}^0[k]$ sont multipliés par les réels β et α . On obtient alors respectivement la sous-bande d'approximation $a_{j+1}[k]$ et la sous-bande de détail $d_{j+1}[k]$.

Propriétés

Le lien entre la structure lifting et la transformée en ondelettes est simple. Daubechies a montré [43] que toute transformée en ondelettes dont les filtres sont à réponse impulsionnelle finie, peut être factorisée sous forme lifting avec un nombre fini d'étages de prédiction et de mise à jour. Ceci justifie ainsi l'utilisation des notations a_j et d_j introduites dans la section 1.2.1 pour désigner les coefficients de la transformée en ondelettes d'un signal. La transformée en ondelettes d'un signal peut donc être réalisée par la structure d'analyse en lifting de la Fig. 1.13. On remarquera ainsi son analogie avec le banc de filtres d'analyse de la Fig. 1.3, présenté dans la section 1.2.2.

Le théorème de factorisation de Daubechies est donc un outil puissant permettant de lier la formulation en banc de filtres à une formulation de type lifting. Ce théorème est de plus constructif et propose un algorithme permettant d'obtenir explicitement les opérateurs de prédiction P et de mise à jour U à partir des filtres h_0 et h_1 mis en jeu dans un banc de filtres d'analyse. Cet algorithme est basé sur la factorisation des transformées en Z de h_0 et h_1 par division euclidienne dans $\mathbb{R}[z, z^{-1}]$. À titre d'exemple, nous donnons en fin de section les structures lifting des transformées en ondelettes classiquement utilisées en codage d'images énoncées précédemment dans la section 1.3.1.

Cependant, l'intérêt principal de la structure en lifting réside dans la propriété suivante : quels que soient les opérateurs de prédiction et de mise à jour utilisés, la transformation par schéma lifting est inversible et on peut retrouver le signal original a_j à partir de ses composantes a_{j+1} et d_{j+1} . En effet, l'inversion du schéma d'analyse se réalise par un simple renversement des étapes et une inversion des signes, comme illustré par la Fig. 1.14 et les équations suivantes :

$$\begin{aligned} a_{j+1}^0[k] &= a_{j+1}[k]/\beta \\ d_{j+1}^0[k] &= d_{j+1}[k]/\alpha \\ a_j[2k] &= a_{j+1}^0[k] - U(\{d_{j+1}^0[k]\}_{k \in \mathbb{Z}}) \\ a_j[2k+1] &= d_{j+1}^0[k] - P(\{a_j[2k]\}_{k \in \mathbb{Z}}) \end{aligned}$$

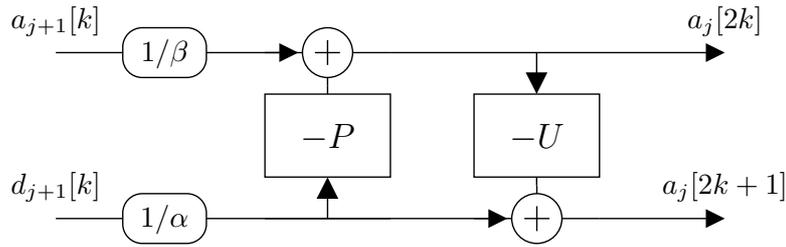


FIG. 1.14 – Structure de synthèse en lifting.

Il est important de remarquer que les opérateurs de prédiction P et de mise à jour U n'ont pas besoin d'être linéaires ni même inversibles pour que le schéma soit inversible. C'est une propriété importante qui permet la construction simple de transformations multirésolution inversibles, non nécessairement linéaires et qui ne sont donc pas représentables par des bases d'ondelettes classiques ou des structures en banc de filtres. Par exemple, le schéma de codage vidéo $t + 2D$ abordé dans cette thèse utilise des images à la place d'échantillons et met en œuvre des opérateurs non-linéaires faisant appel à la compensation de mouvement et à d'autres traitements spatio-temporels.

La structure en lifting possède d'autres avantages par rapport à une structure en banc de filtres. La gestion des effets de bords lors de la transformation est par exemple facilitée. En effet, il suffit simplement de replier symétriquement les opérateurs de prédiction et de mise à jour lors du traitement des échantillons situés aux bords du signal. De plus, le schéma lifting permet de minimiser la complexité [43] de la mise en œuvre d'une transformation en ondelettes. On peut ainsi montrer que la structure lifting nécessite toujours moins de multiplications et d'additions qu'une structure en banc de filtres.

La construction explicite des opérateurs de prédiction P et de mise à jour U est plus problématique. Il est ainsi clair que dans un schéma à deux étages, le prédicteur P sera

choisi de façon à minimiser l'amplitude des coefficients de détail. Il est plus délicat de justifier le choix d'un opérateur de mise à jour car de nombreux critères rentrent en jeu. Un exemple de construction explicite d'une structure lifting adaptée à la compression d'images avec perte est donné dans [56] : en fonction des propriétés statistiques de l'image, l'opérateur de prédiction est choisi de façon à minimiser la variance des coefficients de détail et l'opérateur de mise à jour est conçu pour minimiser l'erreur de reconstruction. On remarquera cependant qu'il est difficile dans le cas général de donner une interprétation aux opérateurs P et U dans une structure lifting possédant un nombre d'étages supérieur à deux.

Généralisations

La structure en lifting peut se généraliser de nombreuses façons. Il est tout d'abord possible d'utiliser un nombre M de composantes polyphases supérieur à deux. Ceci permet la construction de représentations multirésolutions fournissant des facteurs de scalabilité d'un rapport M . De plus, le processus de séparation (*split*) d'un signal en composantes polyphases n'est pas nécessairement basé sur la parité. Ce processus peut être complètement arbitraire, pourvu qu'il soit inversible. Il est ainsi possible d'utiliser une séparation en quinconce 2-bandes non-séparable pour la décomposition d'images, minimisant ainsi le nombre de sous-bandes à traiter.

De plus, il est possible d'utiliser une autre fonction que l'addition \oplus lors du mélange d'une sous-bande avec le résultat de l'opérateur appliqué sur l'autre sous-bande, pourvu que la fonction soit inversible. Les travaux de Piella [111, 115] et de Solé [131] envisagent ainsi à la place de l'addition des fonctions non-linéaires, mettant en œuvre un seuillage et des tests conditionnels. De même, les opérateurs finaux de multiplications peuvent être changés tant qu'ils restent inversibles. La structure reste dite en lifting si elle peut être inversée par renversement des étapes qui la composent.

Exemple de construction *ad-hoc*

Sweldens donne [136] un exemple de la construction *ad-hoc* d'une transformation en ondelettes sous forme lifting. Supposons un échantillon impair x_{2t+1} , il est raisonnable de prédire sa valeur par la moyenne de ses deux voisins pairs $(x_{2t} + x_{2t+2})/2$; on peut donc prendre comme opérateur de prédiction $P = -(x_{2t} + x_{2t+2})/2$. Les coefficients de détail h_t seront donc donnés par :

$$h_t = x_{2t+1} - (x_{2t} + x_{2t+2})/2 \quad (1.20)$$

L'opérateur de mise à jour est choisi de manière similaire et symétrique ; on souhaite ainsi qu'il soit de la forme $U = \xi(h_{t-1} + h_t)/2$. Si on impose de plus que la moyenne courante du signal d'entrée $\sum_t x_t$ soit égale à la moyenne du signal d'approximation $\sum_t l_t$, on obtient $\xi = 1/4$. Les coefficients d'approximation l_t seront donc donnés par :

$$l_t = x_{2t} + (h_{t-1} + h_t)/4 \quad (1.21)$$

La structure ainsi construite et décrite par les équations (1.20) et (1.21) est, aux coefficients multiplicatifs $\alpha = 1/\sqrt{2}$ et $\beta = \sqrt{2}$ près, la transformation biorthogonale en ondelettes 5/3 décrite dans la section 1.2.3.

Exemple de structure en lifting d'ondelettes dyadiques

Cette section énumère les structures en lifting des transformées en ondelettes classiquement utilisées en codage d'image et présentées dans le panorama de la section 1.2.3. La notation est simplifiée et on omet l'indice de résolution j en ne décrivant qu'un seul niveau de décomposition, les autres niveaux étant obtenus par des décompositions subséquentes. Le signal d'entrée a_j est alors noté x et les sous-bandes résultantes d'approximation a_{j+1} et de détail d_{j+1} sont notées respectivement l et h .

La structure lifting de la transformée de Haar est très simple : elle revient à calculer la moyenne et la différence des deux échantillons x_{2t} et x_{2t+1} . Elle s'exprime par :

$$h_t^0 = x_{2t+1} - x_{2t} \quad (\text{P})$$

$$l_t^0 = x_{2t} + \frac{1}{2}h_t^0 \quad (\text{U})$$

$$h_t = \frac{1}{\sqrt{2}}h_t^0 \quad (\text{S1})$$

$$l_t = \sqrt{2}l_t^0 \quad (\text{S2})$$

La structure en lifting de la décomposition en ondelettes Daubechies-4 est plus complexe et nécessite trois étages. Elle s'écrit :

$$h_t^0 = x_{2t+1} - \sqrt{3}x_{2t} \quad (\text{P1})$$

$$l_t^0 = x_{2t} + \frac{\sqrt{3}}{4}h_t^0 + \frac{\sqrt{3}-2}{4}h_{t+1}^0 \quad (\text{U})$$

$$h_t^1 = h_t^0 + l_{t-1}^0 \quad (\text{P2})$$

$$h_t = \frac{\sqrt{3}-1}{\sqrt{2}}h_t^1 \quad (\text{S1})$$

$$l_t = \frac{\sqrt{3}+1}{\sqrt{2}}l_t^0 \quad (\text{S2})$$

La structure en lifting de la transformation en ondelette 5/3 correspond intuitivement à la prédiction d'un échantillon pair par la moyenne de ces voisins et au rehaussement des échantillons impairs afin de préserver la moyenne courante du signal. Plus précisément, elle s'exprime par les relations suivantes :

$$h_t^0 = x_{2t+1} - \frac{1}{2}(x_{2t} + x_{2t+2}) \quad (\text{P})$$

$$l_t^0 = x_{2t} + \frac{1}{4}(h_{t-1}^0 + h_t^0) \quad (\text{U})$$

$$h_t = \frac{1}{\sqrt{2}}h_t^0 \quad (\text{S1})$$

$$l_t = \sqrt{2}l_t^0 \quad (\text{S2})$$

Enfin, la structure lifting de la transformée biorthogonale 9/7 est composée de quatre étages : deux opérateurs de prédictions et deux opérateurs de mise à jour. On trouvera sa forme explicite dans [43].

Ondelettes entières

L'introduction d'un opérateur non-linéaire d'arrondi dans la structure lifting d'une transformée en ondelettes permet d'obtenir aisément une nouvelle transformée entière, c'est à dire à valeurs dans \mathbb{Z} . Cette décomposition entière est scalable, inversible et dispose de propriétés proches de la transformée originale. Ce type de construction est particulièrement utile en codage d'image sans perte car les coefficients issus de la transformation sont entiers et ne nécessitent pas de quantification. Les travaux de Calderbanks [28] ont consisté en l'étude de transformées entières donnant la meilleure efficacité de codage en compression d'image sans perte. Ses résultats ont conclu sur la supériorité de la transformée 5/3 entière, reprise ensuite par la norme JPEG-2000 et décrite simplement par les relations :

$$h_t = x_{2t+1} - \lfloor (x_{2t} + x_{2t+2})/2 + 1/2 \rfloor \quad (\text{P})$$

$$l_t = x_{2t} + \lfloor (h_{t-1} + h_t)/4 + 1/2 \rfloor \quad (\text{U})$$

où $\lfloor \cdot \rfloor$ désigne l'opérateur d'arrondi à l'entier inférieur.

1.3.2 Ondelettes géométriques non-adaptatives

Les bases d'ondelettes séparables utilisées pour décomposer les images possèdent un support carré indéformable et sont isotropes. Pour cette raison, elles ne peuvent pas représenter de manière optimale des régions comportant des contours ou des singularités locales. De nombreuses transformées en ondelettes anisotropes, aptes à la représentation optimale d'images comportant des contours, ont été proposées afin de pallier à cet inconvénient. Nous nous proposons dans cette section de décrire quelques ondelettes géométriques non-adaptatives, qui ont la particularité de posséder une base fixe et indépendante de l'image qu'elles représentent.

Ridgelets

Les ondelettes sont efficaces pour représenter des fonctions continues par morceaux et pour capturer des singularités *ponctuelles* dans le cas monodimensionnel. Ceci n'est pas vrai dans le cas 2D, où l'utilisation de bases d'ondelettes séparables ne permet pas de saisir la régularité présente le long des contours. La représentation par Ridgelets a été proposée par Candès et Donoho [31] pour apporter une solution à ce problème et permet de capturer efficacement la régularité présente le long de contours rectilignes. Elle repose sur l'utilisation de la transformée de Radon qui permet de représenter une image $f \in L^2(\mathbb{R}^2)$ de façon bijective dans le domaine polaire :

$$R_f(\theta, t) = \int_{\mathbb{R}^2} f(x, y) \delta(x \cos(\theta) + y \sin(\theta) - t) dx dy$$

Les ridges $R_f(\theta, t)$ représentent le résultat de la projection radiale de f sur la droite d'équation $x \cos(\theta) + y \sin(\theta) = t$. La transformée de Radon est donc capable de transformer les singularités rectilignes présentes dans une image en singularités ponctuelles. La transformée en Ridgelets s'obtient alors en appliquant une transformée en ondelettes 1D sur le long des ridges $R_f(\theta, \cdot)$ en utilisant la variable d'intégration t . La décomposition d'une image $f \in L^2(\mathbb{R}^2)$ en Ridgelets s'écrit alors :

$$RT_f(a, b, \theta) = \frac{1}{\sqrt{a}} \int_{\mathbb{R}} \psi\left(\frac{t-b}{a}\right) R_f(\theta, t) dt$$

où a est un facteur d'échelle, b un paramètre de translation, θ l'angle de projection et ψ une ondelette. Elle peut s'écrire aussi :

$$RT_f(a, b, \theta) = \int_{\mathbb{R}^2} \psi_{a,b,\theta}(x, y) f(x, y) dx dy$$

$$\text{avec } \psi_{a,b,\theta}(x, y) = \frac{1}{\sqrt{a}} \psi \left[\frac{x \cos(\theta) + y \sin(\theta) - b}{a} \right]$$

et correspond ainsi à la projection de l'image f sur la base fixe $\{\psi_{a,b,\theta}\}_{a,b,\theta}$.

La transformée en Ridgelets est cependant définie dans le domaine continu et n'est pas directement utilisable en compression d'images. Cependant, on peut facilement la discrétiser et en déduire une trame : cela conduit à une description redondante, permettant d'assurer la reconstruction parfaite. On notera enfin les travaux récents de Do et Vetterli [46] proposant une transformée en Ridgelets discrète, inversible et non-redondante. Bien qu'intéressante, la transformée en Ridgelets n'est cependant adaptée qu'aux images présentant des discontinuités le long de contours rectilignes : elle n'est donc pas optimale pour la représentation d'images naturelles, comportant de nombreuses singularités ponctuelles, rectilignes et curvilignes.

Curvelets

La transformée en Ridgelets n'est adaptée qu'aux images présentant des discontinuités le long de contours rectilignes et son intérêt est limité dans le cas d'images naturelles. Cependant, il est clair qu'une image possède *localement* des contours rectilignes : c'est l'idée de la transformation en Curvelets introduite par Candès et Donoho [30, 48]. Elle se décrit en deux étapes : le support de l'image est tout d'abord partitionné en carrés de taille variable avec recouvrement, pour éviter les effets de bords, et ces carrés sont alors décomposés par une analyse en Ridgelets discrète.

Durant cette transformée, les contours non capturés par l'analyse en ondelettes séparables se retrouvent dans les sous-bandes de détail. Un partitionnement suffisamment fin des sous-bandes permet alors d'obtenir des blocs où ces contours forment des lignes droites et sont donc adaptés à l'analyse en Ridgelets. La transformée en Curvelets est inversible mais redondante car l'analyse en Ridgelets discrète sous-jacente est réalisée au moyen d'une FFT du plan polaire, nécessitant plus de points que ceux disponibles dans la grille rectangulaire. Starck [134] montre que son utilisation donne de bons résultats en débruitage d'images.

Contourlets

Les Contourlets sont des ondelettes géométriques non-adaptatives issues des résultats des travaux de Do et Vetterli [47]. Les bases de Contourlets possèdent un grand nombre d'orientations différentes et sont aptes à saisir la régularité présente le long des contours d'une image. De plus, contrairement aux Curvelets, elles sont définies directement dans le domaine discret et leur mise en œuvre en est grandement facilitée. La transformée en Contourlets s'effectue au moyen d'une pyramide Laplacienne redondante et d'un banc de filtres directionnels, construit au travers d'un arbre de décomposition binaire de l étages. Ce dernier utilise des filtres en éventail pour séparer l'information horizontale et verticale et des opérateurs de cisaillement (*shearing*) pour varier les orientations obtenues. La combinaison de ces outils permet de construire une base de Contourlets offrant 2^l

orientations possibles. La transformation en Contourlets est inversible mais la pyramide Laplacienne sous-jacente crée cependant une petite redondance (allant jusqu'à 33 %) qui réduit l'intérêt de son utilisation en compression d'image fixe.

Des travaux plus récents sur les CRISP-contourlets [76] permettent de s'affranchir de ce problème en proposant une transformée en Contourlets inversible et non-redondante mais aucun résultat expérimental de compression n'est présenté. Une méthode hybride combinant une décomposition en Contourlets à l'échelle fine suivie d'une décomposition en ondelettes pour les résolutions grossières a été proposée par Chapellier [33] et offre une efficacité de codage supérieure à la transformée 9/7 séparable dans les bas débits.

1.3.3 Ondelettes géométriques adaptatives

Plutôt que d'utiliser une base fixe, de nombreuses constructions font appel à une base dont les fonctions sont choisies pour s'adapter au mieux à une image donnée. On parle alors d'ondelettes géométriques adaptatives. Le point commun de ces décompositions réside dans une étape d'estimation préalable de la géométrie de l'image (par triangulation, détection de contours, estimation de régularité...) avant de procéder à la décomposition.

Bandelettes

Le flux géométrique d'une image est défini comme un champ indiquant les directions où les variations d'intensités sont régulières. Les bases de Bandelettes, dues à Le Pennec et Mallat [72] sont obtenues par déformation d'une base d'ondelettes selon le flux géométrique local. Les Bandelettes sont cependant construites dans le domaine continu et leur mise en œuvre dans le cas discret est complexe. De plus, le modèle de flux géométrique retenu n'est pas scalable : bien qu'elle offre des résultats expérimentaux en compression d'image convaincants, la transformée en Bandelettes ne semble pas optimale.

Une nouvelle génération de Bandelettes, dues à Peyré [110], sont directement construites dans le domaine discret. La transformation en Bandelettes discrète est définie par l'algorithme suivant. On réalise d'abord une décomposition 2D en ondelettes séparables. Les sous-bandes résultantes sont alors partitionnées en sous-blocs, de façon à isoler les contours orientés. Les sous-blocs sont à leur tour décomposés par une transformée de Haar orientée dans la direction d'angle $k\pi$, où k est un nombre rationnel choisi de façon à minimiser l'amplitude des coefficients résultants. Les paramètres k de chaque sous-bloc décrivent en fait la géométrie de l'image et sont codés à part. Les résultats expérimentaux observés en compression d'image lors de l'utilisation de la transformée en Bandelettes discrète sont satisfaisants.

Représentation adaptée aux contours

Un autre schéma de décomposition non-linéaire et adapté aux contours a été proposé par Cohen et Matei [42, 87]. Il consiste en la construction d'un opérateur de prédiction non-linéaire, capable d'agrandir une image d'un facteur deux. L'utilisation de cet opérateur au sein d'une pyramide Laplacienne permet alors d'obtenir une décomposition multirésolution d'une image. La prédiction est effectuée au moyen d'un ensemble limité de *stencils*, illustrés en Fig. 1.15, qui sont des supports de prédiction de formes variées. Ces *stencils* sont ainsi en mesure de prédire les quatre petits pixels à partir du support de prédiction. Le choix des *stencils* de prédiction est fait à chaque pixel lors d'une étape

préalable de détection de contours et sont codés dans une carte de décision. Cette approche donne de bons résultats sur les images synthétiques et sur les images naturelles à contours nets. Cependant, l'efficacité de cette représentation reste très tributaire de la pertinence et du coût de la carte de décisions.

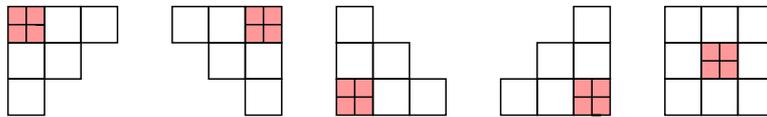


FIG. 1.15 – Stencils : supports de prédiction utilisés dans la représentation de Cohen et Matei. Les pixels de support sont en clair tandis que ceux qui seront prédits sont grisés.

Ondelettes orientées

Afin de représenter au mieux les contours, Chappelier [32] a proposé un schéma de décomposition adaptatif original mettant en œuvre des ondelettes géométriques orientées. Après une décomposition polyphase de l'image en quinconce, les pixels sont prédits soit par rapport à leurs voisins horizontaux, soit par rapport à leurs voisins verticaux, comme illustré sur la Fig. 1.16. La mise à jour est alors faite en fonction de l'état de connexité du pixel. Une carte de décision externe est utilisée pour mémoriser les décisions binaires prises à chaque pixel lors de la prédiction et une procédure d'optimisation débit-distorsion est utilisée pour minimiser son coût. Les résultats expérimentaux obtenus sont bons sur des images présentant de forts contours horizontaux et verticaux comme *Barbara* et sont satisfaisants sur d'autres images naturelles.

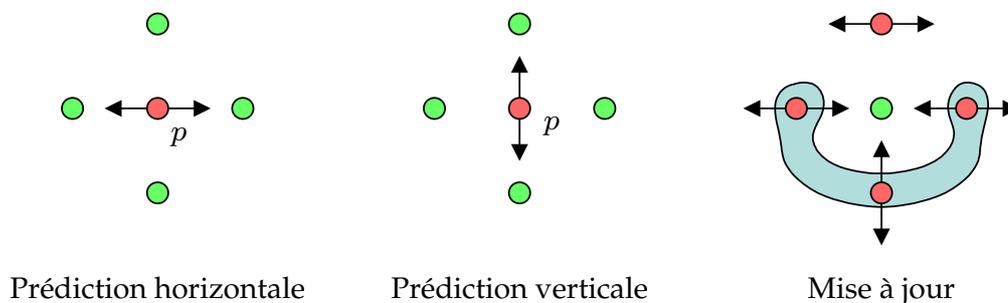


FIG. 1.16 – Étapes de prédiction adaptative et de mise à jour utilisées dans la décomposition en ondelettes orientées de Chappelier.

1.3.4 Autres représentations

L'analyse multirésolution par ondelettes possède certaines faiblesses inhérentes aux bases d'ondelettes même : la transformée en ondelettes n'est ainsi pas invariante par translation et a tendance à créer des coefficients de détail interdépendants entre eux sur plusieurs niveaux de résolution. Nous décrivons ci-après quelques représentations permettant de pallier à ces inconvénients.

Représentation en arbre dual

La transformation en ondelettes classique n'est pas invariante par translation. Cette caractéristique est gênante dans plusieurs applications, par exemple lors de la détection de mouvements dans le domaine transformé, comme décrit en section 2.2.5. La représentation en arbre dual, décrite par Kingsbury [67] permet de s'affranchir de ce problème en proposant une transformée redondante approximativement invariante par translation.

Elle consiste d'une part en la décomposition classique d'un signal en ondelettes et d'autre part, en la décomposition de ce même signal avec les mêmes filtres mais déphasés d'un échantillon. Dans le cas monodimensionnel, cette représentation est redondante d'un facteur 2 et peut être assimilée à une transformation en ondelettes *complexes* où la partie réelle est fournie par la décomposition en ondelettes et la partie imaginaire est donnée par les coefficients issus de la décomposition déphasée.

Footprints

Les discontinuités représentent souvent une partie importante de l'information véhiculée par un signal, en particulier dans le cas des images. Lors d'une décomposition en ondelettes, elles engendrent cependant la création de coefficients de large amplitude présents à tous les niveaux des différentes sous-bandes de détail. Comme illustré sur la Fig. 1.11, ces coefficients entretiennent une interdépendance mutuelle et cette caractéristique n'est pas souhaitable en compression de signal.

Les Footprints, introduites par Dragotti et Vetterli [49], sont des bases redondantes de fonctions capables d'engendrer n'importe quel signal fini polynomial par morceaux, de taille N et de degré maximal D . Elles sont obtenues par la décomposition en ondelettes d'un ensemble de fonctions polynomiales élémentaires de degré $d = 0, 1, \dots, D$ et présentant une discontinuité à l'indice $k = 0, 1, \dots, N - 1$, formant ainsi une base redondante de $N \times (D + 1)$ fonctions. La décomposition en Footprints est alors réalisée par un algorithme spécifique rapide de poursuite adaptative dans cette base. Les résultats obtenus sur des applications de débruitage et de compression de signaux sont très satisfaisants.

1.4 Conclusion

La transformation en ondelettes est un outil capable de donner une représentation multirésolution et parcimonieuse d'un signal monodimensionnel. Dans le cas de signaux multidimensionnels comme les images, il est possible de construire des bases d'ondelettes séparables par produit tensoriel. Cependant, ces dernières possèdent un support carré indéformable et sont isotropes : pour ces raisons, elles ne peuvent pas représenter de manière optimale les régions d'une image comportant des contours ou des singularités locales. Afin de pallier à cet inconvénient, de nombreuses bases d'ondelettes anisotropes (Curvelets, Contourlets, Bandelettes, Ondelettes orientées...) ont alors été proposées pour permettre une représentation plus économique des images.

En parallèle de ces travaux, la découverte de la structure lifting a permis de construire simplement des transformées multirésolution, toujours inversibles et autorisant la mise en œuvre d'opérateurs non-linéaires capables de saisir les singularités d'un signal. De plus, la structure lifting est aisément extensible au cas multidimensionnel et constitue un cadre idéal pour concevoir des transformées scalables capables de fournir une représentation économique d'une séquence vidéo.

