RENCONTRE DES MATHÉMATIQUES APPLIQUÉES ET DU TRAITEMENT DU SIGNAL

par

Stéphane Mallat

1. Au-delà de Fourier

La transformée de Fourier a longtemps dominé le traitement du signal, laissant peu de place à de nouveaux outils mathématiques. Jusqu'aux années 70, les signaux étaient le plus souvent des enregistrements de paroles ou de sons, modélisés par des réalisations de processus gaussiens. En conséquence, les algorithmes linéaires étaient considérés comme optimaux. En ajoutant une hypothèse de stationnarité, on se retrouve dans le pré carré des opérateurs de convolution, qui sont diagonalisés par la transformée de Fourier.

La situation a complètement changé avec le développement du traitement d'images dans les années 80. Les images sont mal modélisées par des processus gaussiens, et les composantes transitoires telles que les contours sont souvent plus importantes que les propriétés stationnaires. L'utilisation d'algorithmes non-linéaires est devenue nécessaire, ce qui a ouvert le traitement du signal aux mathématiques modernes. Au-delà des applications classiques pour la transmission, le codage et la restauration, le traitement du signal s'est aussi engagé dans le traitement de l'information, notamment pour la reconnaissance vocale et la vision par ordinateur.

La construction de représentations compactes est à la base de nombreux problèmes de traitement du signal et de l'information. Une représentation compacte caractérise l'approximation d'un signal (fonction) avec peu de paramètres, qui peuvent être des coefficients de décomposition dans une base ou dans un “dictionnaire” plus large. Des traitements non-linéaires complexes peuvent souvent s'implémenter par des algorithmes simples et rapides dans de telles représentations. L'utilisation de représentations compactes est un outil que l'on retrouve dans de nombreuses branches des mathématiques. Nous montrons que l'étude d'applications à la compression d'images, au débruitage et à la modélisation stochastique de signaux nécessite d'utiliser des résultats récents en théorie de l'approximation, en analyse harmonique, en théorie des opérateurs, en statistique et en probabilités.

2. Représentations compactes

L'élaboration de représentations compactes est clairement utile pour le codage de signaux, mais elle est aussi nécessaire pour réduire la complexité des problèmes de classification et d'identification pour des signaux de grandes tailles.

2.1. Modèles d'images. — L'interprétation bayésienne du monde considère un signal $f(x)$ comme la réalisation d'un processus stochastique $F(x)$. La mesure de l'erreur introduite par un algorithme se fait en moyenne relativement à la distribution de probabilité de F. Les images "naturelles" sont des réalisations de processus non-gaussiens, et il n'y a pas encore de modèle stochastique qui incorpore la diversité de scènes complexes incluant contours et textures, telle que dans l'image 1(a). Cela explique l'utilisation de modèles plus pauvres mais plus réalistes, qui représentent les signaux comme des fonctions appartenant à un ensemble $\mathcal{S}$ de $\mathbf{L^2}[0,1]^d$, sans imposer de condition a priori sur leur distribution de probabilité dans cet ensemble. Pour un problème particulier, on essaye alors de minimiser l'erreur maximum pour les signaux dans $\mathcal{S}$, ce que l'on appelle un problème *minimax*. La discrétisation d'un signal f avec N échantillons est simplement une projection dans un sous-espace de dimension N.

De nombreuses images, dont l'image 1(a), ont une variation totale bornée. Sur $[0,1]$ la variation totale de $f(x)$ mesure la somme des amplitudes de ses oscillations

$$\|f\|_{TV} = \int |f'(x)|\, dx < +\infty \ .$$

La variation totale d'une image sur $[0,1]^2$ est définie par

$$\|f\|_{TV} = \iint |\vec{\nabla} f(x)|\, dx \leq C \ .$$

Cette norme a une interprétation géométrique simple, basée sur les courbes de niveaux de l'image. Soit

$$\Omega_t = \{(x,y) \in \mathbb{R}^2 \quad : \quad f(x,y) > t\} \ .$$

Si $H^1(\partial\Omega_t)$ est la mesure de Hausdorff mono-dimensionnelle de la frontière de Ω_t alors

$$\|f\|_{TV} = \int_{-\infty}^{+\infty} H^1(\partial\Omega_t)\, dt. \tag{1}$$

Une image à variation bornée a aussi une amplitude bornée, et appartient donc à un ensemble du type

$$(2) \qquad \mathcal{S}_{\mathbf{BV}} = \{f \ : \ \|f\|_{TV} = \iint |\vec{\nabla} f(x)|\, dx \leq C \ , \ \|f\|_\infty = \sup_{x\in[0,1]^2} |f(x)| \leq C\} \ .$$

Les images dans $\mathcal{S}_{\mathbf{BV}}$ ont des courbes de niveau qui sont presque toutes de longueur finie. Bien que simple, ce modèle est suffisant pour illustrer les idées principales et les difficultés pour construire des représentations d'images. Des classes plus restreintes d'images, telles que des textures homogènes, sont mieux modélisées par des processus aléatoires tels que des champs de Markov, qui sont étudiés dans la section 4.

2.2. Représentations par approximations. — Une représentation compacte de $f \in \mathbf{L^2}[0,1]^d$ peut s'obtenir en tronquant sa décomposition dans une base orthonormale $\mathcal{B} = \{g_m\}_{m\in\mathbb{N}}$

$$f = \sum_{m=0}^{+\infty} \langle f, g_m\rangle \, g_m \ .$$

Comprendre la performance de représentations compactes dans une base est un sujet central de la théorie de l'approximation. Nous expliquons pourquoi il est nécessaire d'utiliser des représentations non-linéaires, mais une introduction plus complète est donnée dans [**13**].

Une approximation linéaire de f à partir de M produits scalaires $\langle f, g_m\rangle$ est une projection orthogonale dans l'espace $\mathbf{V}_M$ généré par M vecteurs de $\mathcal{B}$, disons les M premiers,

$$f_M = P_{\mathbf{V}_M} f = \sum_{m=0}^{M-1} \langle f, g_m\rangle \, g_m \ .$$

L'erreur d'approximation maximum sur un ensemble de signaux $\mathcal{S}$ est

$$\epsilon_l(\mathcal{S}, M) = \sup_{f\in\mathcal{S}} \|f - f_M\|^2 = \sup_{f\in\mathcal{S}} \sum_{m=M}^{+\infty} |\langle f, g_m\rangle|^2.$$

Une telle représentation est efficace si $\epsilon_l(\mathcal{S}, M)$ a une décroissance rapide lorsque M augmente, et donc si $|\langle f, g_m\rangle|$ a une décroissance rapide lorsque m croît. Cela dépend du choix de $\mathcal{B}$ relativement à $\mathcal{S}$. Par exemple, les fonctions uniformément régulières sont bien approximées par M sinusoïdes de basses fréquences dans une base de Fourier $\{\mathrm{e}^{i2\pi mx}\}_{m\in\mathbb{Z}}$ de $\mathbf{L^2}[0,1]$. Si $\mathcal{S}$ est inclus dans une boule d'un espace de Sobolev $\mathbf{W}^s[0,1]$ de fonctions de période 1 alors la décroissance des coefficients de Fourier aux hautes fréquences implique que $\epsilon_l(\mathcal{S}, M) = O(M^{-2s})$ [**13**]. Les fonctions à variation bornée peuvent avoir des discontinuités, et ne sont pas bien approximées dans une base de Fourier. On démontre que dans une boule $\mathcal{S}_{\mathbf{BV}}$ de fonctions à variation bornée, la décroissance la plus rapide qui puisse être obtenue dans une base $\mathcal{B}$ est $\epsilon_l(\mathcal{S}_{\mathbf{BV}}, M) \sim M^{-1}$ **[13]**.

Pour améliorer ce résultat, il est nécessaire d'utiliser une représentation plus adaptée au signal, qui projette f sur M vecteurs selectionnés en fonction de f

$$f_M = \sum_{m \in I_M} \langle f, g_m \rangle \, g_m \, . \tag{3}$$

Comme $\|f - f_M\|^2 = \sum_{m \notin I_M} |\langle f, g_m \rangle|^2$, la meilleure approximation est obtenue en sélectionnant dans I_M les M vecteurs dont l'amplitude des produits scalaires $|\langle f, g_m \rangle|$ est maximum. Cette approximation dépend de $2M$ paramètres, les M indexes dans I_M et les M coefficients $\{\langle f, g_m \rangle\}_{m \in I_M}$. Soit $c_k = \langle f, g_{m_k} \rangle$ le coefficient dont l'amplitude est de rang k

$$|c_k| \geq |c_{k+1}| \quad \text{pour } k \geq 1 \, .$$

L'erreur d'approximation est

$$\|f - f_M\|^2 = \sum_{k=M+1}^{+\infty} |c_k|^2$$

et

$$\epsilon_n(\mathcal{S}, M) = \sup_{f \in \mathcal{S}} \|f - f_M\|^2 \, .$$

Cette erreur dépend de la vitesse de décroissance des coefficients ordonnés $|c_k|$ pour les signaux dans $\mathcal{S}$. Dans la base $\mathcal{B}$, une boule $w\mathbf{l^p}$ de rayon C est définie par

$$\mathcal{S}_{wl^p} = \{f \; : \; |c_k| = |\langle f, g_{m_k} \rangle| \leq C \, k^{-1/p}\} \, . \tag{4}$$

On vérifie facilement que $\mathcal{S} \subset \mathcal{S}_{wl^p}$ pour un $C > 0$ et $p < 2$ si et seulement si $\epsilon_n(\mathcal{S}, M) = O(M^{1-\frac{2}{p}})$. La difficulté principale de l'approximation non-linéaire est de calculer le p minimum et la base correspondante $\mathcal{B}$ telle que $\mathcal{S} \subset \mathcal{S}_{wl^p}$. Une telle base est dite *optimale* pour l'approximation non-linéaire de $\mathcal{S}$. Les bases inconditionnelles sont des exemples de bases optimales.

Une base orthonormale $\mathcal{B}$ est une base *inconditionnelle* d'un sous-espace de Banach $\mathbf{B} \subset \mathbf{L^2}[0,1]^d$ s'il existe A tel que pour toute suite de signes $s_m \in \{-1, 1\}$ et $f \in \mathbf{B}$

$$\|f\|_{\mathbf{B}} = \left\| \sum_{m=0}^{+\infty} \langle f, g_m \rangle \, g_m \right\|_{\mathbf{B}} \leq A \left\| \sum_{m=0}^{+\infty} s_m \, \langle f, g_m \rangle \, g_m \right\|_{\mathbf{B}} \, .$$

Le fait que $\|f\|_{\mathbf{B}} < +\infty$ ne dépend donc que des amplitudes $|\langle f, g_m \rangle|$, et peut être relié à une condition sur la décroissance des coefficients ordonnés. On démontre

[13] que si $\mathcal{B}$ est une base inconditionnelle de $\mathbf{B}$ alors c'est une base optimale pour l'approximation non-linéaire d'une boule $\mathcal{S} = \{f \ : \ \|f\|_{\mathbf{B}} \leq C\}$ de $\mathbf{B}$.

2.3. Grille adaptative par ondeletttes. — Les bases d'ondelettes ont d'importantes applications en mathématiques et en traitement du signal car elles permettent de construire des représentations compactes pour de larges classes de fonctions. Les premières bases d'ondelettes orthonormales de $\mathbf{L^2}(\mathbb{R})$ furent introduites par Strömberg et Meyer **[25]**. L'interprétation par multirésolution des bases d'ondelettes donne un cadre général pour construire pratiquement toutes les ondelettes qui génèrent une base orthonormale de $\mathbf{L^2}(\mathbb{R})$ **[19]**. Cela permet aussi d'introduire un algorithme rapide qui nécessite que $O(N)$ opérations pour calculer N coefficients en ondelettes **[22]**. Daubechies **[9]** a découvert des ondelettes à support compact, et les bases résultantes ont été adaptées à $\mathbf{L^2}[0,1]^d$. Nous résumons ici les résultats principaux.

Une bases orthonormale d'ondelettes de $\mathbf{L^2}[0,1]$ est une famille de fonctions

$$\mathcal{B} = \left(\{\phi_{l,n}\}_{0 \leq n < 2^l} \ \cup \ \{\psi_{j,n}\}_{j \geq l, 0 \leq n < 2^j} \right) .$$

A la résolution 2^l, les *fonctions d'échelles* $\{\phi_{l,n}\}_{0 \leq n < 2^l}$ génèrent un espace $\mathbf{V}_l$ qui inclut les polynômes de degré q, pour un certain $q \geq 0$. Les ondelettes $\psi_{j,n}$ aux hautes résolutions $2^j > 2^l$ sont donc orthognales aux polynômes de degré q. Les ondelettes $\psi_{j,n}$ dont le support est inclus dans $]0,1[$ sont obtenues par dilatation et translation d'une ondelette "mère" ψ :

$$\psi_{j,n}(t) = \sqrt{2^j}\, \psi(2^j t - n) \ .$$

Les ondelettes de bord sont modifiées pour maintenir leur support à l'intérieur de $[0,1]$.

Une approximation linéaire de f à partir de $M = 2^J > 2^l$ ondelettes et fonctions d'échelle est calculée en gardant tous les coefficients aux résolutions $2^j < 2^J$:

$$f_M = \sum_{n=0}^{2^l} \langle f, \phi_{l,n} \rangle \, \phi_{l,n} + \sum_{j=l}^{J-1} \sum_{n=0}^{2^j} \langle f, \psi_{j,n} \rangle \, \psi_{j,n}.$$

La première somme donne une approximation grossière de f à la résolution 2^l, et chaque somme partielle $\sum_{n=0}^{2^j} \langle f, \psi_{j,n} \rangle \, \psi_{j,n}$ apporte les "détails" qui améliorent cette approximation de la résolution 2^j à la résolution 2^{j+1}. Si f est continue, cette approximation linéaire à la résolution 2^J est essentiellement équivalente à une approximation sur une grille uniforme calculée en interpolant les échantillons $\{f(2^{-J}n)\}_{0 \leq n < 2^J}$. Comme une approximation de Fourier linéaire, cette approximation est précise si et seulement si f est uniformément régulière. On obtient de mauvaises appproximations pour des fonctions ayant des singularités, telles que les fonctions à variations bornées.

Une approximation non-linéaire par ondelettes garde les M coefficients d'amplitude maximum. L'amplitude de $|\langle f, \psi_{j,n}\rangle|$ dépend de la régularité locale de f. Supposons que l'ondelette mère ψ soit $\mathbf{C}^{q+1}$ et qu'elle soit orthogonale aux polynômes de degré q. On démontre alors **[25]** que f est uniformément Lipschitz $\alpha < q+1$ sur un intervalle $[a,b]$ si et seulement s'il existe $A > 0$ tel que pour tout $\psi_{j,n}$ dont le support est inclus dans $[a,b]$ (modulo des problèmes de bord)

$$|\langle f, \psi_{j,n}\rangle| \leq A\, 2^{-(\alpha+1/2)j} \ .$$

Dans les domaines où la régularité Lipschitzienne α est grande, $|\langle f, \psi_{j,n}\rangle|$ a une décroissance rapide lorsque la résolution 2^j augmente. A haute résolution 2^j, les grands coefficients apparaissent dans le voisinage des singularités, où $0 \leq \alpha < 1$. Un plus grand nombre de coefficients en ondelettes est gardé au voisinage des singularités. Cette approximation est équivalente à un échantillonnage sur grille adaptative dont la résolution est augmentée au voisinage des singularités.

L'impact des bases d'ondelettes en analyse fonctionnelle vient du fait qu'elle sont des bases inconditionnelles d'une large famille d'espaces, appelés espaces de *Besov*. Bien que l'espace $\mathbf{BV}$ de fonctions à variations bornées n'admette aucune base inconditionnelle, $\mathbf{BV}$ peut être encadré par deux espaces de Besov. Cela permet de démontrer que les bases d'ondelettes sont optimales pour approximer les fonctions d'une boule $\mathcal{S}_{\mathbf{BV}}$. On peut montrer que l'ensemble $\mathcal{S}_{\mathbf{BV}}$ est inclus dans la boule $w\mathbf{l^p}$ (4) pour $p = 2/3$ mais pas pour $p < 2/3$ **[12]**. Donc $\epsilon_n(\mathcal{S}_{\mathbf{BV}}, M) = O(M^{1-2/p}) = O(M^{-2})$. Lorsque M augmente, la décroissance asymptotique de $\epsilon_n(\mathcal{S}_{\mathbf{BV}}, M)$ est donc plus rapide que l'erreur de n'importe quelle approximation linéaire utilisant M paramètres, qui décroît au mieux comme M^{-1}.

En deux dimensions, les bases d'ondelettes sont construites avec 3 ondelettes mères, ψ^k pour $1 \leq k \leq 3$, qui sont dilatées et translatées

$$\psi^k_{j,n}(x) = \psi^k_{j,(n_1,n_2)}(x_1,x_2) = 2^j\, \psi^k(2^j x_1 - n_1, 2^j x_2 - n_2) \ .$$

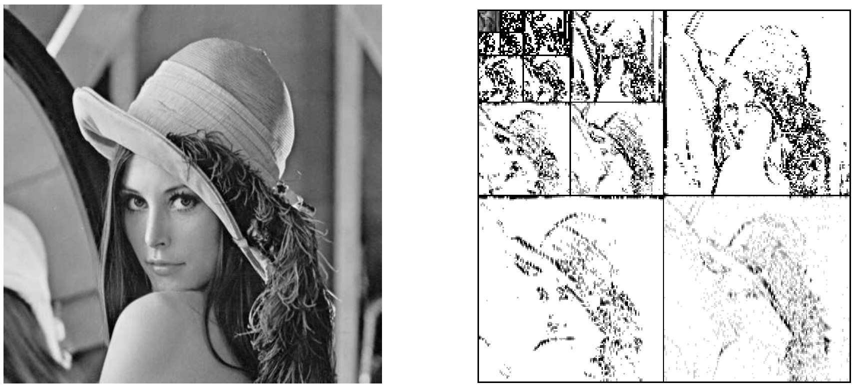
Des modifications appropriées sont faites aux bords de telle sorte que le support soit maintenu dans $[0,1]^2$. Une ondelette $\psi^k_{j,n}$ a un support carré de taille proportionnelle à 2^{-j}, et centré près de $2^{-j}n = (2^{-j}n_1, 2^{-j}n_2)$. Une base orthonormale d'ondelettes de $\mathbf{L^2}[0,1]^2$ est obtenue en incluant des fonctions d'échelles qui engendrent un espace de basse résolution

$$\mathcal{B} = \Big(\{\phi_{l,n}\}_{2^{-l}n \in [0,1)^2} \ \cup \ \{\psi^k_{j,n}\}_{j \geq l\,,\, 2^{-j}n \in [0,1)^2\,,\, 1 \leq k \leq 3} \Big) \ . \tag{5}$$

Une image discrète est un tableau de N^2 points (pixels), avec $N = 512$ pour l'image 1(a). La base d'ondelettes (5) se discrétise pour définir une base orthonormale d'images ayant N^2 pixels. Les coefficients en ondelettes de l'image 1(a) sont montrés en 1(b). Chaque sous-image donne la valeur de

$$\{|\langle f, \psi_{j,n}^k\rangle|\}_{2^{-j}n\in[0,1)^2}$$

pour j et k fixés. Le nombre de coefficients en ondelettes pour chaque sous-image est 2^{2j}. Le module de chaque coefficient $|\langle f, \psi_{j,n}^k\rangle|$ est représenté par un point, qui est d'autant plus sombre que cette valeur est grande. Ces sous-images sont organisées par triplets, correspondant aux indexes $1 \leq k \leq 3$. Les ondelettes pour $k = 1, 2, 3$ sont sensibles aux variations de l'image selon différentes orientations. La plupart des points sont blancs, ce qui signifie que la majorité des coefficients d'ondelettes sont quasiment nuls. Les quelques grands coefficients sont localisés là où l'intensité de l'image varie brutalement, le long d'un contour ou sur une texture.



(a) (b)

FIGURE 1. (a) : Image originale f. (b) : Amplitude des coefficients $|\langle f, \psi_{j,n}^k\rangle|$ dans une base d'ondelettes. Les sous-images correspondent à des résolutions 2^j et des orientations k différentes (voir texte).

Une approximation linéaire à partir de $M = 2^{2J}$ ondelettes est calculée en gardant tous les coefficients aux résolutions $2^j < 2^J$. Cette approximation sur une grille uniforme est particulièrement inefficace pour des images ayant des discontinuités, ce qui est souvent le cas. Pour une boule $\mathcal{S}_{\mathbf{BV}}$ d'images à variations bornées (2), on démontre que $\epsilon_l(\mathcal{S}_{\mathbf{BV}}, M) = A > 0$. Le maximum de l'erreur d'approximation ne décroît pas vers zéro lorsque M augmente.

Une approximation non-linéaire est bien plus efficace car elle garde les coefficients en ondelettes près des singularités (1). En particulier, on peut montrer que $\mathcal{S}_{\mathbf{BV}}$ est inclus dans une boule $w\mathbf{l}^1$ [**4**] et donc que $\epsilon_n(\mathcal{S}_{\mathbf{BV}}, M) = O(M^{-1})$. Une grille adaptative d'ondelettes donne une bien meilleure approximation qu'une grille uniforme, et aucune autre base orthonormale ne peut améliorer la vitesse de décroissance de l'erreur d'approximation sur une boule $\mathcal{S}_{\mathbf{BV}}$.

2.4. Compression de Signaux. — La transmission rapide de signaux au travers de canaux à débit limité comme Internet, et le stockage de signaux de grande taille, nécessitent d'effectuer une compression en introduisant une erreur aussi faible que possible. Le codage efficace d'un signal avec un nombre minimum de bits utilise des représentations compactes.

Les ingénieurs en traitement du signal n'ont pas attendu que les mathématiciens étudient les propriétés des approximations non-linéaires pour coder efficacement des sons et des images dans des bases orthonormales. C'est en 1986 que fut implémenté le premier algorithme de codage d'images dans une base d'ondelettes [**34**], avant même que ces bases ne soient étudiées en mathématiques. En effet, l'algorithme rapide de décomposition dans une base d'ondelettes s'implémente avec un banc de filtres, qui fut initialement introduit en traitement du signal pour *multiplexer* des signaux (agréger plusieurs signaux en un seul) [**7**]. Une théorie des bancs de filtres discrets a été élaborée en traitement du signal [**33**], mais ce n'est que plus tard que la connection avec les bases orthonormales d'ondelettes fut établie [**19**]. Bien que les mathématiques soient arrivées plus tard, l'analyse des performances de ces codeurs d'images demande d'utiliser des résultats récents de la théorie de l'approximation, et ceux-ci ouvrent des directions nouvelles pour des améliorations potentielles.

Les signaux dans $\mathcal{S}$ sont discrétisés et approximés à la résolution N, ce qui signifie qu'ils appartiennent à un espace de dimension N. Un codage par transformée décompose f dans une base orthonormale $\mathcal{B} = \{g_m\}_{0\leq m<N}$

$$f = \sum_{m=0}^{N-1} \langle f, g_m\rangle\, g_m \,,$$

et approxime chaque coefficient $\langle f, g_m\rangle$ par une valeur *quantifiée*, qui est codée sur le moins de bits possibles. Un quantificateur uniforme de pas Δ approxime chaque $x \in \mathbb{R}$ par $Q(x) = k\Delta$, avec $k \in \mathbb{Z}$ et $|x - Q(x)| \leq \Delta/2$. Le signal quantifié résultant est

$$\tilde{f} = \sum_{m=0}^{N-1} Q(\langle f, g_m\rangle)\, g_m \,.$$

Le problème est de minimiser la *distorsion* maximum $d(\mathcal{S}, R) = \sup_{f\in\mathcal{S}} \|f - \tilde{f}\|^2$ pour un nombre maximum de bits R alloués pour coder $\tilde{f}$.

Le calcul de la distorsion d'un codage par transformée est reliée à un problème d'approximation non-linéaire. Soit M le nombre de coefficients plus grands que $\Delta/2$

(a) (b)

FIGURE 2. (a) : Image codée avec 0.25 bits/pixel, en quantifiant les coefficients en ondelettes de l'image 1. (b) : Image codée avec 0.125 bits/pixel.

et f_M l'approximation non-linéaire de f à partir de ces M coefficients

$$f_M = \sum_{|\langle f, g_m\rangle| > \Delta/2} \langle f, g_m\rangle\, g_m \, .$$

Comme $Q(x) = 0$ quand $|x| < \Delta/2$, et que $|x - Q(x)| \leq \Delta/2$, la distorsion est

$$d(f,R) = \|f - \tilde{f}\|^2 \leq \|f - f_M\|^2 + M\, \frac{\Delta^2}{4} \, . \tag{6}$$

Supposons que $\mathcal{S}$ soit inclus dans une boule $\mathcal{S}_{wl^p}$ (4) de rayon C, avec $p < 2$, si bien que $\|f - f_M\| = O(M^{1-2/p})$. On note $M_0 = C^p (\Delta/2)^{-p}$. Comme $|c_k| = |\langle f, g_{m_k}\rangle| \leq C\, k^{-1/p}$, nécessairement $M \leq M_0$. On vérifie aussi que

$$d(\mathcal{S}, R) = \sup_{f \in \mathcal{S}} d(f,R) \leq \sup_{f \in \mathcal{S}} \|f - f_{M_0}\|^2 + M_0\, \frac{\Delta^2}{4} \; = O(M_0^{-2/p+1}) \, . \tag{7}$$

La distorsion totale dépend donc essentiellement de l'erreur d'approximation non-linéaire.

Pour optimiser le codage par transformée, il faut minimiser le nombre de bits R nécessaires pour coder les N valeurs $\{Q(\langle f, g_m\rangle)\}_{0 \leq m < N}$ lorsque $f \in \mathcal{S}$. Pour de forts taux de compression, $N \gg M_0 \geq M$, auquel cas une grande proportion $\frac{N-M}{N}$ des coefficients est annulée par la quantification. Un codage entropique utilise des mots binaires plus courts pour représenter les coefficients qui apparaissent le plus souvent. Sachant que $\mathcal{S} \subset \mathcal{S}_{wl^p}$, on peut construire un tel code, qui nécessite au plus $R \sim M_0 \log_2 \frac{M_0}{N}$ bits **[22]**. On déduit de (7) que $d(\mathcal{S}, R) = O(R^{1-2/q})$ pour tout $q > p$.

La décroissance de $d(\mathcal{S}, R)$ est maximisée dans une base $\mathcal{B}$ qui est optimale pour l'approximation non-linéaire de $\mathcal{S}$, car une telle base minimise l'exposant p tel que $\mathcal{S} \subset \mathcal{S}_{wl^p}$. En particulier, les bases d'ondelettes sont optimales pour les images à variations bornées et l'exposant minimum est $p = 1$. Les figures 2(a,b) sont les images $\tilde{f}$ obtenues en quantifiant les coefficients en ondelettes de la figure 1(b). Elles sont codées respectivement avec $\frac{R}{N} = 0.25$ bits/pixel et 0.125 bits/pixel, en optimisant le codage des coefficients nuls [**30**]. L'image originale 1(a) est codée avec 8 bits/pixel, donc les taux de compression sont respectivement 32 et 64. Pour 0.25 bits/pixel, l'erreur de distorsion n'est pratiquement pas visible mais devient apparente pour 0.125 bits/pixel.

Le choix de la base dépend entièrement de la nature des signaux dans $\mathcal{S}$. Pour des sons, des bases différentes doivent être choisies afin d'approximer efficacement les structures oscillantes de durées variables. La figure 3 montre l'enregistrement du mot "greasy". Le standard Dolby pour la compression audio utilise une base de cosinus locale. Une telle base est construite à partir d'une fonction paire $w(t)$, que l'on appelle une *fenêtre*, dont le support est inclus dans $[-l, l]$ et qui est translatée afin de couvrir l'axe temporel :

$$\sum_{p=-\infty}^{+\infty} |w(t - p\,l)|^2 = 1.$$

Malvar [**23**], Coifman et Meyer [**5**] ont démontré qu'en imposant d'autres propriétés de symétrie sur $w(t)$, la multiplication par des fonctions cosinus génère une base orthonormale de $\mathbf{L^2}(\mathbb{R})$

$$\left\{ g_{p,k}(t) = \frac{1}{\sqrt{l}}\, w(t - p\,l)\, \cos\Big(\pi k\, (l^{-1} t - p)\Big) \right\}_{k \in \mathbb{N}, p \in \mathbb{Z}}. \tag{8}$$

Comme dans le cas des images, la performance d'un codage audio dans une telle base dépend de la capacité à approximer les sons avec peu de vecteurs de cette base. Cependant, les meilleures mesures de distorsion audio ne sont pas des normes $\mathbf{L^2}$. Des techniques de masquage sont utilisées par les ingénieurs pour introduire des erreurs de quantification qui sont au-dessous des seuils de perception [**22**], mais celles-ci sont bien plus difficiles à comprendre mathématiquement.

2.5. Géométrie et représentations adaptatives. — Les bases d'ondelettes sont optimales pour représenter la classe générale des images à variations bornées, mais de meilleures approximations peuvent être obtenues en utilisant la régularité géométrique des images. La formule de variation totale (1) montre que les lignes de niveau d'une image à variation bornée ont des longueurs presque toujours finies. Cependant, cela n'impose aucune condition sur la régularité de ces lignes de niveau. Dans

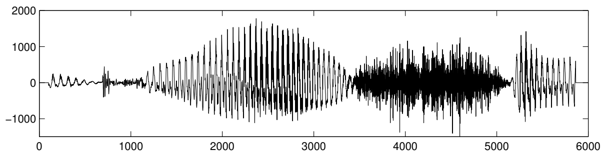


FIGURE 3. Enregistrement du son "greasy" échantillonné à 16kHz.

l'image 1(a), les contours sont le plus souvent des courbes régulières dans le plan de l'image, ayant une faible courbure. Comprendre comment utiliser cette régularité est fondamental pour le traitement d'images. Cela a notamment motivé de nombreux travaux utilisant des équations différentielles partielles non-linéaires pour modifier la courbure des lignes de niveau dans les images [**1, 28, 31**]. Cette nouvelle branche du traitement d'images mathématique a donné lieu à des applications intéressantes pour le débruitage de la segmentation d'images. Cependant, nous ne suivrons pas cette ligne de pensée, qui n'utilise pas de facon explicite une représentation compacte des images.

Pour comprendre l'importance de la régularité géométrique, nous considèrerons un exemple simple, $f = \mathbf{1}_\Omega$, qui est la fonction indicatrice d'un ensemble Ω. La frontière $\partial\Omega$ de Ω est une courbe différentiable de longueur finie, dont la courbure est bornée. Si le support carré de $\psi_{j,n}^k$ n'intersecte pas $\partial\Omega$ alors $\langle f, \psi_{j,n}^k\rangle = 0$. Les ondelettes $\psi_{j,n}^k$ sont translatées sur une grille uniforme d'intervalle 2^{-j} et ont un support carré de taille proportionelle à 2^{-j}, comme l'illustre la figure 4(a). A la résolution 2^j, il y a $O(2^j)$ ondelettes $\psi_{j,n}^k$ dont le support intersecte $\partial\Omega$. Les M coefficients en ondelettes d'amplitude maximum, selectionnés par une approximation non-linéaire, sont aux résolutions $2^j \leq 2^J \sim M$. Les coefficients non sélectionnés produisent une erreur $\|f - f_M\|^2 \sim M^{-1}$.

Une meilleure approximation linéaire par morceau s'obtient avec une triangulation adaptative de $[0,1]^2$, utilisant M triangles [**16**]. Comme la courbure de $\partial\Omega$ est bornée, cette frontière peut être couverte par $M/2$ triangles ayant une petite largeur proportionnelle à M^{-2} le long de la normale à $\partial\Omega$, et une longueur d'ordre M^{-1} le long de la tangente à $\partial\Omega$. L'intérieur et l'extérieur de Ω sont couverts par $M/2$ grands triangles, comme l'illustre la figure 4(b). Une fonction f_M linéaire sur chaque triangle peut approximer $f = \mathbf{1}_\Omega$ avec une erreur $\|f - f_M\|^2 = O(M^{-2})$. Cette erreur d'approximation est concentrée sur les triangles qui intersectent la frontière $\partial\Omega$ et leur faible largeur produit une erreur plus petite qu'avec des ondelettes dont les supports sont carrés. L'erreur est réduite car la représentation est adaptée aux propriétés géométriques de $\partial\Omega$.

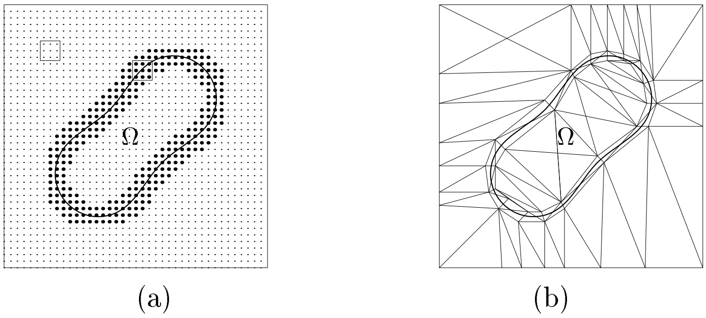
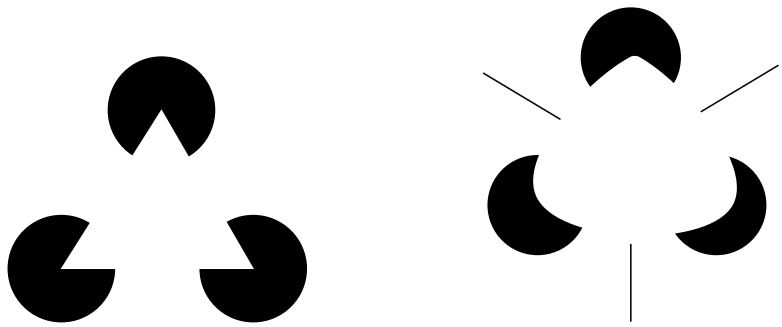


FIGURE 4. (a) : Les ondelettes $\psi_{j,n}$ sont translatées sur une grille d'intervalle 2^{-j}, et ont un support carré proportionnel à 2^{-j}. Pour $f = \mathbf{1}_\Omega$, les points les plus sombres correspondent aux ondelettes $\psi_{j,n}$ telles que $\langle f, \psi_{j,n}\rangle \neq 0$. (b) : Une approximation linéaire par morceau de $f = \mathbf{1}_\Omega$ est optimisée en choisissant le long de la frontière des triangles qui sont étroits dans la direction où f est discontinue.

Construire un pont entre les contraintes géométriques et des approximations adaptatives est un des problèmes fondamentaux du traitement d'images. Le système visuel humain est très sensible aux structures géométriques telles que des "coins" ou la régularité des contours [**24, 26**]. Les illusions de Kanizsa de la figure 5 illustrent ce phénomène. Le contour triangulaire est perceptible alors que l'intensité de l'image reste constante au milieu. Cette illusion s'explique en imposant des contraintes géométriques sur l'interprétation (modélisation) des images. Par ailleurs, on sait [**11**] que les "neurones simples" du cortex visuel effectuent une décomposition de l'image sur une famille de fonctions qui ressemblent à des ondelettes, mais qui forment une famille plus redondante qu'une base et donc offrent plus de flexibilité. Cela semble indiquer que notre cerveau traverse constamment ce pont entre l'analyse fonctionnelle et la géométrie.

S'adapter à la géométrie des images peut s'interpréter comme un exemple particulier d'un problème plus général d'approximation adaptative. Une base est une famille complète dans un espace fonctionnel, mais est souvent trop petite pour utiliser au mieux la diversité des structures apparaissant dans un signal complexe. Des approximations plus précises sont obtenues avec M vecteurs selectionnés dans un dictionnaire plus large $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$, qui peut inclure un nombre infini de bases. Cela suit la même logique, qui motive une personne à élargir son vocabulaire afin de former des phrases plus précises et plus concises. Pour la reconnaissance, il est aussi important de construire des représentations qui satisfont certains invariants relativement à des translations ou des transformations affines. Cela impose davantage



(a) (b)

FIGURE 5. Les contours apparents de ces triangles sont perceptibles dans des domaines où l'image reste uniformément blanche.

de conditions sur le dictionnaire [**20**]. Un dictionnaire pour des images peut être construit avec des ondelettes dont le support a une élongation et une orientation qui peuvent varier arbitrairement. Comme pour les triangles allongés de la figure 4(b), la famille d'ondelettes choisie peut s'adapter à la géométrie des courbes de niveau de l'image. De même, les signaux audio sont mieux approximés dans un dictionnaire de bases de cosinus locaux (8), où la taille l de la fenêtre peut être adaptée à la durée des composantes du sons : attaques, harmoniques, vibrato...

Une représentation adaptative est construite à partir d'un dictionnaire $\mathcal{D}$ en sélectionnant M vecteurs $\{g_{\gamma_k}\}_{1\leq k\leq M}$ pour approximer f avec une somme partielle

$$f_M = \sum_{k=1}^{M} \alpha_k \, g_{\gamma_k} \ .$$

En l'absence d'orthogonalité, trouver les M vecteurs qui minimisent $\|f - f_M\|$ produit une explosion combinatoire. Des algorithmes gloutons de poursuites ont été développés pour éviter cette explosion [**20**]. Ils sélectionnent les vecteurs g_{γ_k} un par un, en minimisant à chaque étape l'erreur d'approximation. La performance de ces algorithmes est cependant loin d'être optimale [**3, 13**]. Dans des dictionnaires composés de bases orthonormales qui peuvent être structurées dans des arbres, Coifman and Wickerhauser [**6**] ont développé un algorithme rapide qui effectue une sélection "resonable" bien que non-optimale des M vecteurs d'approximation. Il n'y a cependant pas de résultat de théorie de l'approximation qui permette d'analyser la performance de ces algorithmes hautement non-linéaires, ou d'améliorer leur précision.

Augmenter la taille du dictionnaire a un coût. Plus le dictionnaire est grand, plus il faut de paramètres pour caractériser l'index γ_k de chaque vecteur selectionné. Pour une erreur d'approximation fixée, en choisissant un dictionnaire trop large, on augmente le nombre total de paramètres qui caractérisent l'approximation f_M. Trouver des dictionnaires de taille optimale est un autre problème ouvert.

3. Débruitage par seuillage

La suppression d'un bruit, additionné lors de la mesure d'un signal ou durant sa transmission, est un problème important où les représentations compactes jouent un rôle crucial. Dans une base qui transforme le signal en quelques coefficients de grande amplitude plus un petit reste, une grande partie du bruit est facilement supprimée avec un seuillage qui annule les coefficients les plus petits. Une version similaire a été utilisée pour enlever le bruit des images de télévision, depuis les années 60. Cependant, ce n'est que récemment que Donoho et Johnstone **[14]** ont développé les mathématiques permettant de démontrer que ces estimateurs de seuillage sont quasiment optimaux dans des représentations compactes, ce qui a ouvert de nouvelles applications.

On note $f[n]$ les N coefficients qui caractérisent la projection de $f(x)$ sur un espace de dimension N. On mesure des données bruitées

$$D[n] = f[n] + W[n] \ , \tag{9}$$

où les coefficients $W[n]$ du bruit sont des variables aléatoires gaussiennes indépendantes. On dit alors que W est un *bruit blanc*. La figure 6(a) montre un exemple. Un estimateur F de f est calculé en appliquant un opérateur L sur les données, $F = L\, D$. Le risque de cette estimation est

$$r(L, f) = \mathsf{E}\{\|f - LD\|^2\}.$$

On veut minimiser le risque maximum pour un ensemble $\mathcal{S}$ de signaux

$$r(L, \mathcal{S}) = \sup_{f \in \mathcal{S}} r(L, f) \ .$$

Le but est de trouver un opérateur L qui approche le risque minimax

$$r_o(\mathcal{S}) = \inf_L r(L, \mathcal{S}).$$

Pour approcher l'optimum minimax, on sépare les problèmes de représentation et d'estimation. On construit d'abord une représentation en décomposant $D = f + W$ sur une base orthonormale $\mathcal{B} = \{g_m\}_{0 \leq m < N}$:

$$\langle D, g_m \rangle \ = \ \langle f, g_m \rangle + \langle W, g_m \rangle.$$

Un estimateur de seuillage est alors défini par

$$F = L_t D = \sum_{m=0}^{N-1} \theta_T(\langle D, g_m \rangle)\, g_m, \tag{10}$$

avec $\theta_T(x) = x\,\mathbf{1}_{|x|>T}$. Il annule tous les coefficients plus petits que T, et garde les autres.

Le seuil T est choisi comme étant juste au-dessus de $\max_{0\leq m<N} |\langle W, g_m\rangle|$, avec une forte probabilité, pour que $\theta_T(\langle D, g_m\rangle) = 0$ si $\langle f, g_m\rangle \approx 0$.

Comme W est un bruit gaussien de variance σ^2, dans n'importe quelle base $\mathcal{B}$, les coefficients du bruit $\langle W, g_m\rangle$ sont des variables aléatoires gaussiennes indépendantes de même variance σ^2. Soit M le nombre de coefficients tels que $|\langle f, g_m\rangle| \geq \sigma$, et f_M l'approximation non-linéaire (3) de f calculée avec ces M vecteurs. Si $T = \sigma\sqrt{2\log_e N}$ alors Donoho and Johnstone ont montré **[14]** que

$$r(L_t, f) \leq (2\log_e N + 1)\left(\|f - f_M\|^2 + (M+1)\,\sigma^2\right).$$

Cette borne supérieure est similaire à la distorsion (6) d'un codage par transformée. Le risque est réduit en choisissant une base où il y a un petit nombre M de grands coefficients d'amplitude supérieure à σ, qui produisent une petite erreur d'approximation $\|f - f_M\|$. Une fois de plus, le problème est de trouver une représentation compacte et précise. L'image 6(b) est une estimation calculée en seuillant les coefficients en ondelettes de l'image bruitée en (a).

(a) (b)

FIGURE 6. (a) : Image dégradée par l'addition d'un bruit blanc gaussien. (b) : Estimation par seuillage dans une base d'ondelettes.

La performance asymptotique d'un estimateur de seuillage est calculée tandis que la résolution N des mesures tend vers $+\infty$. Pour un ensemble $\mathcal{S}_0$ de signaux $f(x)$, on définit une base orthonormale $\mathcal{B}_0$ qui est optimale pour l'approximation non-linéaire. Si $\mathcal{S}_0$ est une boule dans un espace $\mathbf{B}$ alors on peut choisir $\mathcal{B}_0$ comme étant une base inconditionnelle de $\mathbf{B}$. L'ensemble $\mathcal{S}$ des signaux discrétisés est obtenu par une projection dans un espace de dimension N. Ces signaux sont décomposés dans une base discrétisée $\mathcal{B}$ calculée à partir de $\mathcal{B}_0$ avec la même projection. Lorsque que N augmente, on démontre **[15]** que l'estimateur de seuillage est presque optimal au sens où

$$r(L_t, \mathcal{S}) \leq O(\log N)\, r_o(\mathcal{S}) \ . \tag{11}$$

Ce résultat s'applique aux signaux discrets provenant d'espaces de Besov, décomposés dans une base d'ondelettes discrète. Il reste aussi valable pour un ensemble $\mathcal{S}_{\mathbf{BV}}$ de fonctions à variations bornées, car $\mathbf{BV}$ est encadré par deux espaces de Besov qui sont proches. Dans ce cas, lorsque N augmente, le risque de l'estimateur de seuillage a une décroissance plus rapide que le risque de n'importe quel estimateur linéaire.

L'efficacité d'un estimateur de seuillage dépend fondamentalement de la représentation. Pour utiliser les propriétés des structures complexes, telles que la régularité géométrique des contours d'une image, le seuillage doit être calculé dans des représentations plus flexibles, comme l'explique la section 2.5. L'optimalité minimax dans de telles représentations adaptatives est encore mal comprise.

4. Processus à interaction limitée

Pour de nombreux problèmes de classification, dont la reconnaissance de la parole et la discrimination de textures visuelles, le signal observé est modélisé comme étant la réalisation d'un processus qui doit être caractérisé. Ceci est particulièrement difficile car le processus est souvent non gaussien et non stationnaire, et une seule réalisation donne peu d'informations pour l'identification. Il est donc nécessaire de caractériser de tels processus avec peu de paramètres pouvant être estimés avec peu de données. Après l'étude de processus gaussiens non stationnaires, nous considérons un modèle plus général basé sur les champs de Markov.

4.1. Processus gaussiens non stationnaires. — Les processus aléatoires gaussiens donnent des modèles utiles pour de nombreuses classes de signaux, dont les enregistrements de parole. Un processus gaussien $X(t)$ de moyenne nulle pour $t \in \mathbb{R}$, est entièrement caractérisé par sa covariance

$$k(t,s) = \mathsf{E}\{X(t)\, X(s)\} \ ,$$

qui est le noyau de l'opérateur de covariance K :

$$Kf(t) = \int_{-\infty}^{+\infty} k(t,s)\, f(s)\, ds. \tag{12}$$

Pour estimer cette covariance avec peu de réalisations, il est nécessaire de réduire le nombre de coefficients qui décrivent le noyau. Cela peut être fait avec une base orthonormale $\mathcal{B} = \{g_m\}_{m\in\mathbb{Z}}$ dans laquelle la matrice

$$\langle K g_m, g_n\rangle = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} k(t,s)\, g_m(s)\, g_n(t)\, ds\, dt \tag{13}$$

a une décroissane rapide hors-diagonale. Ces valeurs sont les coefficients de décomposition de $k(t,s)$ dans une base orthonormale séparable $\{g_n(t)\, g_m(s)\}_{(n,m)\in\mathbb{Z}^2}$ de $\mathbf{L^2}(\mathbb{R}^2)$.

Trouver une représentation compacte pour cette matrice revient à approximer $k(t,s)$ avec peu de vecteurs dans une base séparable. Si les coefficients de la matrice ont une décroissance rapide hors de la diagonale, alors K est bien approximée (avec une norm sup ou de Hilbert Schmidt) par une matrice à bande étroite $\tilde{K}$ dans $\mathcal{B}$. Si cette matrice est positive, c'est la covariance d'un processus gaussien $\tilde{X}$ qui approxime X [**21**]. Comme $\tilde{K}$ est représenté par une matrice bande dans $\mathcal{B}$, pour tout $m \in \mathbb{N}$ il existe un voisinage $\mathcal{N}_m$ qui inclut un nombre fini d'entiers tels que si $n \notin \mathcal{N}_m$ alors

$$\langle \tilde{K} g_m, g_n\rangle = \mathsf{E}\{\langle \tilde{X}, g_m\rangle\, \langle \tilde{X}, g_n\rangle\} = 0.$$

Comme $\langle \tilde{X}, g_m\rangle$ et $\langle \tilde{X}, g_n\rangle$ ont une probabilité jointe gaussienne, ils sont indépendants car non corrélés. Dans la base $\mathcal{B}$, les coefficients de $\tilde{X}$ sont dépendants seulement dans de petits voisinages, ce qui est un exemple particulier de champ de Markov.

Pour étudier les propriétés de K, on l'écrit sous forme d'opérateur pseudo-différentiel [**25**]. Soit

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(s)\, \mathrm{e}^{-i\omega s}\, ds$$

la transformée de Fourier de f. Le *symbole* de l'opérateur K est

$$\beta(t,\omega) = p.v. \int_{-\infty}^{+\infty} k(t, t-s)\, \mathrm{e}^{-i\omega s}\, ds\,.$$

En appliquant la formule de Parseval à (12), on obtient

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

Par exemple, si $\beta(t,\omega) = \sum_{p=0}^{P} a_p(t)\, (i\omega)^p$ alors $K = \sum_{p=0}^{P} a_p(t)\, (\frac{d}{dt})^p$ est un opérateur différentiel avec des coefficients variables. Le processus X est stationnaire si

$k(t,s) = k(t-s)$, auquel cas $\beta(t,\omega) = \beta(\omega)$ est le spectre de K. La transformée de Fourier est donc l'outil idéal pour caractériser les processus gaussiens stationnaires. Pour les processus non-stationnaires, il faut relier les propriétés de $X(t)$ aux propriétés de $\beta(t,\omega)$, afin de déduire une base dans laquelle K peut être approximée par une matrice bande.

Les processus localement stationnaires $X(t)$ apparaissent dans de nombreux systèmes physiques, pour lesquels les mécanismes qui produisent les fluctuations aléatoires changent lentement dans le temps ou dans l'espace [**29**]. Sur de courts intervalles de temps l, de tels processus peuvent être approximés par des processus stationnaires. C'est le cas pour les signaux de parole et de nombreux sons. Sur des intervalles suffisamment courts, la gorge se comporte comme un résonateur stationnaire, qui est excité par une source de bruit stationnaire. Des processus localement stationnaires s'obtiennent en imposant qu'il existe $A > 0$ tel que pour tout $j, k \geq 0$

$$|\partial_t^k \partial_\omega^j \beta(t,\omega)| \leq A\, l^{j-k} .$$

On démontre [**21**] l'existence d'une base de cosinus locaux (8) dans laquelle l'opérateur K est bien approximé par une matrice à bande étroite. La taille l de chaque fenêtre correspond à la taille des intervalles de stationnarité. Lorsque la taille $l(t)$ des intervalles de stationnarité varie dans le temps, ce qui est le cas des signaux audio, la covariance a des propriétés plus complexes et souvent n'appartient pas aux familles classiques d'opérateurs pseudo-différentiels. Selon la régularité de $l(t)$, on peut construire des bases de cosinus locaux avec des fenêtres de longueurs adaptées, qui définissent des représentations compactes de tels opérateurs [**21**].

Les multifractales sont d'autres modèles importants de signaux, ayant des propriétés d'auto-similarité [**27**]. Parmi les nombreux exemples, mentionnons les séries économétriques telles que l'indice Dow Jones, des données physiologiques dont certains enregistrements cardiaques, les fluctuations électromagnétiques des radiations galactiques, certaines textures d'images, la densité du trafic automobile... Un mouvement brownien fractionnaire $X(t)$ d'exposant de Hurst H est un exemple canonique de processus gaussien fractal. Ses incréments sont stationnaires, et il est auto-similaire au sens où $s^{-H}X(st)$ a la même distribution de probabilité que $X(t)$, pour tout $s > 0$. Le symbole de la covariance K de X est $\beta(t,\omega) = \lambda\,|\omega|^{-2H-1}$. Cela correspond à un opérateur de Calderón-Zygmund de la première génération [**25**], dont on sait qu'il a une décroissance rapide dans une base d'ondelettes. En traitement du signal, les browniens fractionnaires sont souvent approximés par un processus $\tilde{X}$ dont la covariance $\tilde{K}$ est diagonale dans une base d'ondelettes, ce qui donne un algorithme rapide de synthèse [**27**]. Des conditions générales sur $\partial_t^k \partial_\omega^j \beta(t,\omega)$ peuvent être établies pour garantir que K ait une décroissance rapide hors-diagonale dans une base d'ondelettes [**2**]. Les processus browniens multifractionnaires sont des exemples dont les exposants de Hurst varient dans le temps : $\beta(t,\omega) = \beta_0(t)\,|\omega|^{-2H(t)-1}$. Des estimations précises de $\beta_0(t)$ et $H(t)$ s'obtiennent dans des bases d'ondelettes.

Lorsque le processus est uniformément localement stationnaire ou multifractal, la base qui comprime la matrice de covariance est connue à l'avance. Pour des processus plus complexes, cette base doit aussi être estimée, à partir de certaines informations a priori. Ceci est un problème d'approximation adaptive, similaire à celui décrit dans la section 2.5, bien que l'on approxime un opérateur plutôt qu'une fonction. Des algorithmes de recherche de meilleure base ont été introduits pour effectuer de telles approximations pour des opérateurs de covariance [**21**], mais ces techniques ne sont qu'un premier pas dans un domaine qui demande un travail de recherche plus approfondi afin de comprendre les propriétés des estimateurs statistiques résultants.

4.2. Champs de Markov sur des réprésentations adaptées. — La caractérisation et synthèse de textures visuelles est un des problèmes les plus difficiles de la vision bas-niveau. Des textures visuelles homogènes, telles que du bois, du marbre, ou toute autre, sont stationnaires mais non gaussiennes. La Figure 7 montre deux exemples. Il est nécessaire de modéliser de tels processus avec peu de paramètres afin de pouvoir les identifier à partir d'une seule réalisation. Ceci est possible car le système visuel humain peut le faire. L'importance de ce problème va bien au-delà de la discrimination des textures visuelles. En effet, de nombreuses classes de signaux telles que les séries financières ou les mesures de vitesse d'un champ turbulent, ne peuvent être approximées par des processus gaussiens. Il est nécessaire de construire un cadre suffisamment général, incluant des processus qui peuvent modéliser de tels signaux.

Les champs de Markov sont devenus un outil attractif pour modéliser les textures visuelles, grâce au travail de Geman et Geman [**17**], qui ont introduit un algorithme de relaxation stochastique pour échantillonner des distributions de Gibbs. Pour simplifier la présentation, nous nous restreignons à des vecteurs aléatoires $X(n)$, où $n \in \mathbb{Z}^d$ varie sur une grille $\mathcal{G}$ de taille N. On définit un système de voisinage $\mathcal{N} = \{\mathcal{N}_n\}_{n\in\mathcal{G}}$ tel que $n \notin \mathcal{N}_n$, et $m \in \mathcal{N}_n$ si et seulement si $n \in \mathcal{N}_m$. Pour tout $\mathcal{G}_0 \subset \mathcal{G}$, on note $X(\mathcal{G}_0)$ l'ensemble des valeurs prises par X sur $\mathcal{G}_0$. On dit que $p(X)$ est la distribution de probabilité d'un *champ de Markov aléatoire* relativement $\mathcal{N}$ si

$$p\Big(X(n) \mid X(\mathcal{G} - \{n\})\Big) = p\Big(X(n) \mid X(\mathcal{N}_n)\Big).$$

Un sous-ensemble C de $\mathcal{G}$ est appelé une clique si chaque paire d'éléments dans C sont des voisins. Soit $\mathcal{C}$ l'ensemble de toutes les cliques. Si X prend ses valeurs sur un alphabet fini alors le théorème de Hammersley-Clifford démontre que $p(X)$ est un champ de Markov si et seulement s'il peut être écrit sous la forme d'une *distribution de Gibbs* relativement à $\mathcal{N}$

$$p(X) = \frac{1}{Z} \exp\Big[-\sum_{C \in \mathcal{C}} \phi_C(X)\Big],$$

où Z est une constante de normalisation et ϕ_C est une *fonction potentielle* qui dépend seulement des valeurs de X dans la clique C. Les modèles de Markov ont des propriétés intéressantes pour la discrimination de texture et la restauration d'images, mais un succès limité à cause de la difficulté d'incorporer les interactions longues portées des pixels de certaines images. Plusieurs approches ont été introduites pour contourner ce problème, dont les techniques de renormalisation [**18**].

Mumford et Zhu [**35**] ont introduit un point de vue différent en créant des champs de Markov sur des représentations comprimées de X, plutôt que sur les échantillons $X(n)$. Soit $\mathcal{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ un dictionnaire de vecteurs, qui peut être une base orthogonale ou plus redondante. Soit $X_\gamma = \langle X, g_\gamma \rangle$. Un système de voisinage $\mathcal{N} = \{\mathcal{N}_\gamma\}_{\gamma\in\Gamma}$ est définit sur Γ. Par exemple, si $\mathcal{D} = \{\psi^k_{j,n}\}_{k,j,n}$ est une base d'ondelettes en deux dimensions, l'index $\gamma = (k, j, n)$ définit l'orientation k, la résolution 2^j et la position $2^{-j}n$ de l'ondelette. Un voisinage $\mathcal{N}_{(k,j,n)}$ inclut typiquement les ondelettes $\psi^{k'}_{j',n'}$ telles que $|j - j'| \leq 1$ et dont la position $2^{-j'}n'$ est proche de $2^{-j}n$. Les propriétés de multirésolution des bases d'ondelettes permettent de construire des champs de Markov qui incluent des interactions à courte et longue portées.

Pour construire un champ de Markov X à partir de signaux observés $\{X^{obs}_p\}_{0\leq p<P}$, on effectue des mesures moyennées sur M cliques $\{C_m\}_{0\leq m<M}$, avec des fonctions potentielles ϕ_{C_m},

$$\mu^{obs}_{C_m} = \frac{1}{P} \sum_{p=1}^{P} \phi_{C_m}(X^{obs}_p) \ .$$

Si X est stationnaire alors on calcule la moyenne empirique sur tous les ϕ_{C_m} qui effectuent les mêmes opérations en des points différents de l'image. Ces moyennes empiriques sont des estimations de $\mathsf{E}\{\phi_{C_m}\}$ pour le modèle X que l'on construit. Le plus souvent, les cliques ont au plus deux éléments $C = \{\gamma, \gamma'\}$. Des mesures de covariance correspondent à $\phi_C(X) = X_\gamma\, X_{\gamma'}$. Cependant, des fonctions potentielles différentes peuvent être utiles, telles que des moments d'ordre p

$$\phi_C(X) = |X_\gamma|^p\, |X_{\gamma'}|^p \quad \text{for } p > 0 \ . \tag{14}$$

Le *principe d'entropie maximum* suggère de choisir $p(X)$ qui maximise l'entropie

$$p(X) = \arg\,\max\{-\int p(X)\ \log p(X)\, dX\} \ .$$

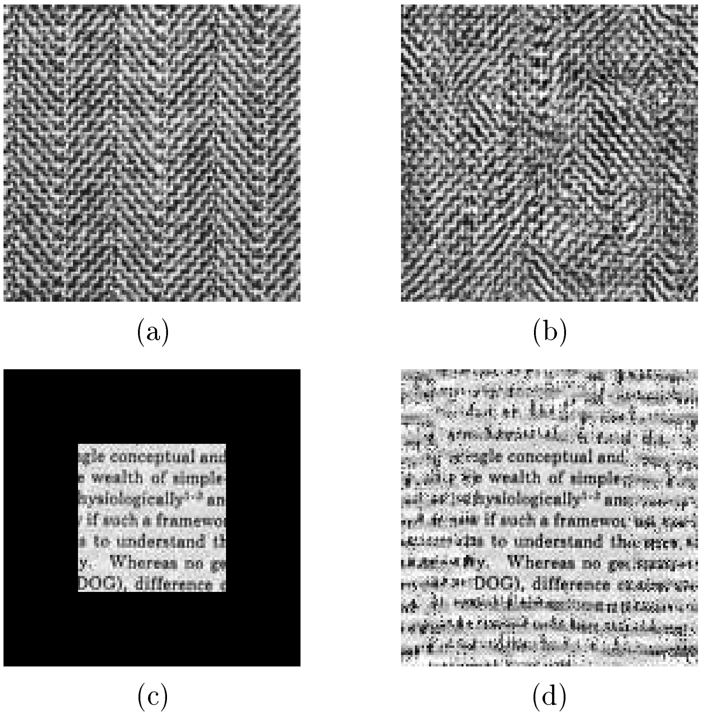


FIGURE 7. (a) : Observation d'une texture uniforme. (b) : Réalisation d'un modèle de Markov par ondelettes calculé à partir de (a). (c) : Le centre montre un exemple de texture. (d) : Le centre est identique a (c), tandis que la périphérie est la réalisation d'un modèle de Markov par ondelettes, calculé à partir de (c).

sous les contraintes

$$\mathsf{E}\{\phi_{C_m}(X)\} = \int \phi_{C_m}(X)\, p(X)\, dX = \mu^{obs}_{C_m} \quad \text{pour } 1 \leq m \leq M \ . \tag{15}$$

La distribution d'entropie maximum $p(X)$ est la plus "uniforme" étant donné les contraintes imposées par les observations $\mu^{obs}_{C_m}$. Aucune "information" supplémentaire n'est donc introduite par le modèle. La solution est calculée avec les multiplicateurs de Lagrange

$$p(X, \Lambda) = \frac{1}{Z(\Lambda)} \exp\left(- \sum_{m=1}^{M} \lambda_m\, \phi_{C_m}(X) \right) \ . \tag{16}$$

Le vecteur de paramètres $\Lambda = \{\lambda_m\}_{1 \leq m \leq M}$ est caractérisé uniquement par les contraintes (15), si les fonctions potentielles satisfont une propriété d'indépendance linéaire.

Si $\phi_{C_m}(X)$ sont des mesures de covariance alors (16) est une distribution de probabilité gaussienne, et si $\mathcal{D}$ est une base orthonormale alors Λ est calculée en inversant une matrice bande. La maximisation de l'entropie est une optimisation convexe [**17**], mais pour des fonctions potentielles générales ϕ_{C_m}, le vecteur Λ ne peut être calculé analytiquement. Les algorithmes numériques obtiennent Λ en estimant $\mathsf{E}_{p(X,\Lambda)}\{\phi_{C_m}(X)\}$ à chaque itération, puis en modifiant Λ avec une descente de gradient afin de statisfaire les conditions (15). Le calcul de $\mathsf{E}_{p(X,\Lambda)}\{\phi_{C_m}(X)\}$ est fait avec un échantillonneur de Gibbs ou un algorithme de Monté Carlo, ce qui nécessite beaucoup d'opérations.

Mumford et Zhu [**35**] et Portilla et Simoncelli [**32**], utilisent de tels champs de Markov pour construire un modèle à partir d'une seule réalisation d'une texture. Le modèle de Markov de Portilla et Simoncelli est calculé sur une base d'ondelettes, avec des contraintes sur la covariance et sur des moments (14) avec $p = 1$. La texture 7(a) est l'unique observation utilisée pour calculer les paramètres Λ du modèle, avec une hypothèse de stationnarité. La figure 7(b) montre une réalisation du modèle de Markov par ondelettes. Cette réalisation est remarquablement proche de la texture originale, au sens où elles ne peuvent pas être distinguées visuellement en moins de 10^{-1} secondes (vision préattentive). Un modèle similaire de Markov par ondelettes est calculé à partir de la texture de "texte" de la figure 7(c). L'image 7(d) inclut une réalisation de ce modèle de Markov à la périphérie de la texture (c), que l'on ne distingue pas du centre par vision "préattentive".

Les champs de Markov offrent un cadre général pour construire des processus ayant des interactions locales dans une représentation appropriée. La validité de ces modèles dépend du choix de la représentation et des fonctions potentielles ϕ_C. Comprendre comment optimiser ces deux composantes et analyser les propriétés de tels champs de Markov sur des espaces fonctionnels est un autre sujet ouvert.

Références

[1] L. Alvarez, F. Guichard, P.L. Lions and J.M. Morel. Axioms and fundamental equations of image processing. *Archieve for Rational Mechanic*, 123 :199–257, 1993.

[2] A. Benassi, S. Jaffard, and D. Roux, Elliptic Gaussian random processes *Revista Mathematica Iberoamericana* 13 :19–90, 1997.

[3] S. Chen and D. Donoho. Atomic decomposition by basis pursuit. In *SPIE International Conference on Wavelets*, San Diego, July 1995.

[4] A. Cohen, R. DeVore, P. Pertrushev, and H. Xu, Non-linear approximation and the space $BV(\mathbb{R}^2)$ *American J. of Math.*, to appear 1998.

[5] R. R. Coifman and Y. Meyer. Remarques sur l'analyse de Fourier à fenêtre. *C.R. Acad. Sci.*, pages 259–261, 1991.

[6] R. R. Coifman and M. V. Wickerhauser. Entropy-based algorithms for best basis selection. *IEEE Trans. Info. Theory*, 38(2) :713–718, March 1992.

[7] A. Croisier, D. Esteban, and C. Galand. Perfect channel splitting by use of interpolation-/decimation/tree decomposition techniques. In *Int. Conf. on Info. Sciences and Systems*, pages 443–446, Patras, Greece, August 1976.

[8] G. Cross, and A. Jain. Markov random field texture models *IEEE Trans. on PAMI*, 5 :25–39, 1983.

[9] I. Daubechies. *Ten Lectures on Wavelets.* SIAM, Philadelphia, PA, 1992.

[10] I. Daubechies Wavelets and other phase space localization methods *Proc. of ICM 94*, Birkhäuser Verlag, Switzerland 1995.

[11] J. G. Daugmann. Two-dimensional spectral analysis of cortical receptive field profile. *Vision Research*, 20 :847–856, 1980.

[12] R. DeVore, B. Jawerth, and V. Popov, Compression of wavelet decompositions *Amer. J. Math.*, 114 :737–785, 1992.

[13] R. DeVore, Nonlinear approximation *Acta Numerica*, 51–150, Cambridge University Press, 1998.

[14] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrika*, 81 :425–455, December 1994.

[15] D. Donoho, Abstract statistical estimation and modern harmonic analysis. *Proc. of ICM 94*, Birkhäuser Verlag, Switzerland 1995.

[16] N. Dyn and S. Rippa. Data-dependent triangulations for scattered data interpolation and finite element approximation. *Applied Num. Math.*, 12 :89–105, 1993.

[17] S. Geman, and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. on PAMI*, 6 :721-741, November 1984.

[18] B. Gidas. A renormalization group approach to image processing. *IEEE Trans. on PAMI*, 11 :164-180, 1989.

[19] S. Mallat. Multiresolution approximations and wavelet orthonormal bases of $\mathbf{L^2}(\mathbb{R})$. *Trans. Amer. Math. Soc.*, 315 :69–87, September 1989.

[20] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12) :3397–3415, December 1993.

[21] S. Mallat, Z. Zhang, and G. Papanicolaou. Adaptive covariance estimation of locally stationary processes. *Annals of Stat.*, 26(1) :1–47, 1997.

[22] S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, Boston, MA, 1998.

[23] H. S. Malvar. *Signal Processing with Lapped Transforms.* Artech House, Norwood, MA, 1992.

[24] D. Marr. *Vision.* W.H. Freeman and Co., San Fransisco, 1982.

[25] Y. Meyer. *Ondelettes et Opérateurs.* Hermann, Paris, 1990.

[26] D. Mumford. Mathematical Theories of Shape : Do they model perception ? *SPIE : Geometric Methods in Computer Vision*, 1570 :2–10, 1991.

[27] J. F. Muzy, E. Bacry, and A. Arneodo. The multifractal formalism revisited with wavelets. *Int. J. of Bifurcation and chaos*, 4 :245, 1994.

[28] S. Osher, and L. Rudin. Feature-oriented image enhancement using shock filters. *SIAM J. on Numerical Analysis*, 27 :919–940, 1990.

[29] M. Priestley, Evolutionary spectra and non-stationary processes. *J. Roy. Stat. Soc. Ser. B*, 27 :204–237, 1965.

[30] J. M. Shapiro. Embedded image coding using zero-trees of wavelet coefficients. *IEEE Trans. Signal Proc.*, 41(12) :3445–3462, December 1993.

[31] G. Shapiro, and A. Tannenbaum. On invariant curve evolution and image analysis. *J. of Functional Analysis*, 119(1) :79–120, 1993.

[32] E. Simoncelli., and J. Portilla. Texture characterization via second-order statistics of wavelet coefficient amplitudes. Proc. of 5^{th} IEEE Int. Conf. on Image Proc., Chicago, October 1998.

[33] M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding.* Prentice-Hall, Englewood Cliffs, NJ, 1995.

[34] J. W. Woods and S. D. O'Neil. Sub-band coding of images. *IEEE Trans. Acoust., Speech, and Signal Proc.*, 34(5) :1278–1288, May 1986.

[35] S. C. Zhu, and D. Mumford. Prior learning and Gibbs reaction-diffusion. *IEEE Trans. on PAMI*, 19 :1236–1250, November 1997.

Stéphane Mallat, Centre de Mathématiques Appliquées, Ecole Polytechnique, 91128 Palaiseau Cedex