

Parcimonie, Ondelettes, *-lettres

B. Torrèsani

LATP, Université de Provence, Marseille, France

EPIT, Porquerolles, Mai 2008

Parcimonie, une vieille idée

Parcimonie: représenter et traiter un objet en utilisant un volume minimum de données.

Ca pourrait commencer comme ça: une conversation au Bar de la marine, Porquerolles, le 28 Mai 2008:

Parcimonie, une vieille idée

Cet exposé: complément au cours sur les méthodes à noyau: la parcimonie par le versant est.

Complément à l'antépénultième diapo de S. Canu: la notion de parcimonie s'est déclinée systématiquement dans le monde statistique depuis (au moins) les années 1990, suivant Coifman, Donoho et d'autres.

Il s'agit d'une notion un peu différente de parcimonie.

Objectif

Décrire des méthodes fournissant des décompositions de fonctions, signaux, images,...

$$f = \sum_{k=1}^K \alpha_k \phi_k$$

comme combinaisons linéaires **parcimonieuses** (c'est à dire avec K petit) d'**atomes** ϕ_k .

- 1 Parcimonie et ses conséquences
- 2 Atomes
- 3 Comment ?

- 1 Introduction: analyse, compression et parcimonie
 - Analyse du signal
 - Compression des signaux et des images
 - Séparation de sources
- 2 Gaborettes, ondelettes,... et *-lettres
 - Fourier et Gabor
 - Ondelettes
 - Le cas des images: ondelettes, mdct, brosettes, curvelettes,...
 - Autres *-lettres
 - Apprenti-lettres
- 3 Comment générer de la parcimonie
 - Comment mesurer la parcimonie
 - Bibliothèque de bases: algorithme de meilleure base
 - Approches variationnelles
 - Au delà de la parcimonie
- 4 Conclusions

Analyse du signal

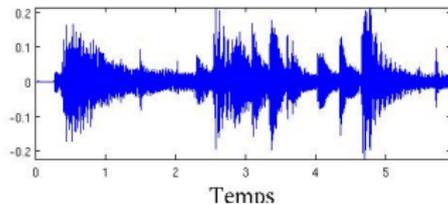
Analyser un signal: extraire des informations, simplifier.

Notion classique: transformation. Pour analyser un signal, lui associer un signal transformé, dans un autre espace, dans lequel l'information sera plus facilement lisible.

Exemple: la partition musicale encode un son complexe en une série de symboles facilement compréhensibles.

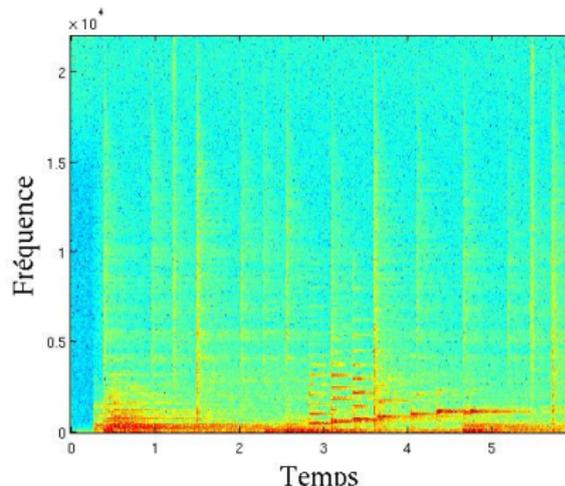
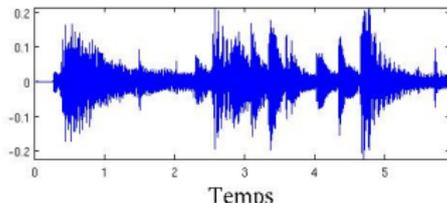
Analyse du signal (2)

Partition “mathématique”: diagramme temps-fréquence parcimonieux



Analyse du signal (2)

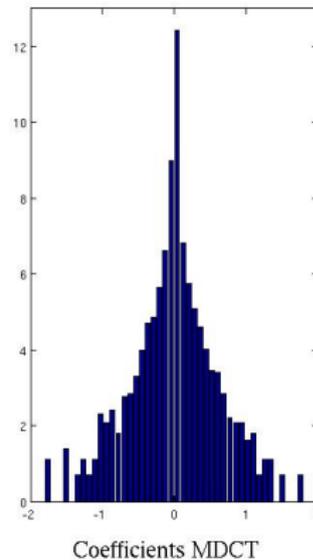
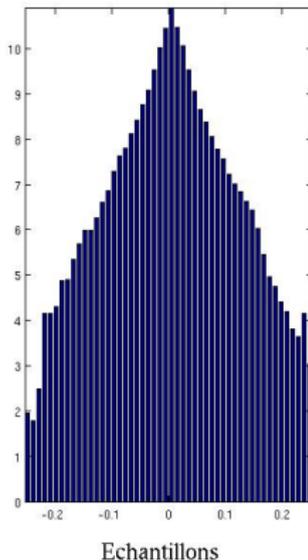
Partition “mathématique”: diagramme temps-fréquence parcimonieux



Cette transformation “concentre” l’information pertinente sur un petit nombre de “coefficients” significatifs.

De plus, elle est inversible: pas de perte d’information.

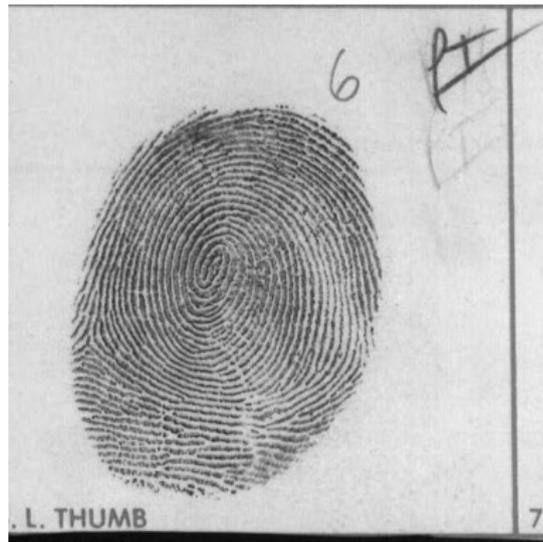
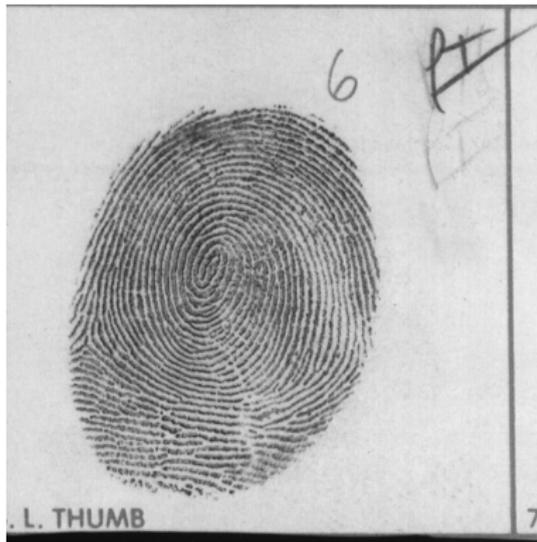
Compression des signaux



Compression avec pertes: force à zéro les coefficients les plus petits.

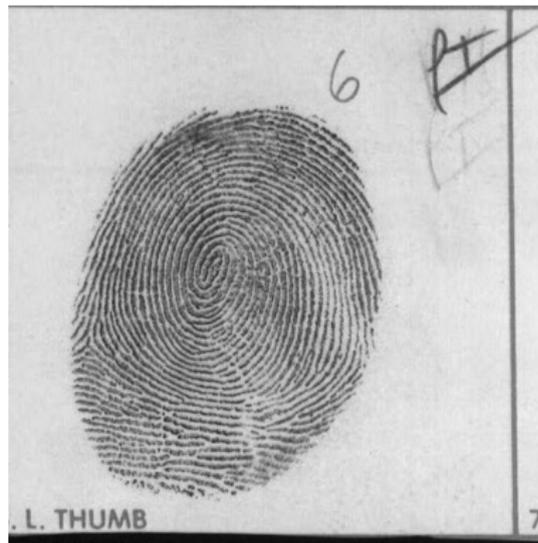
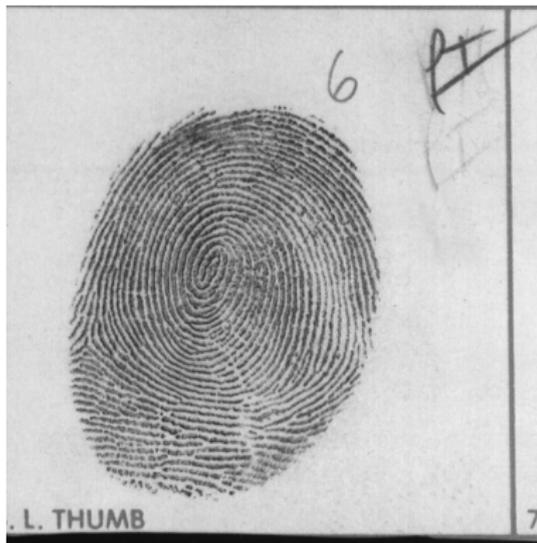
Compression des images

Question: Quelle différence y a-t-il entre ces deux images ?



Compression des images

Question: Quelle différence y a-t-il entre ces deux images ?



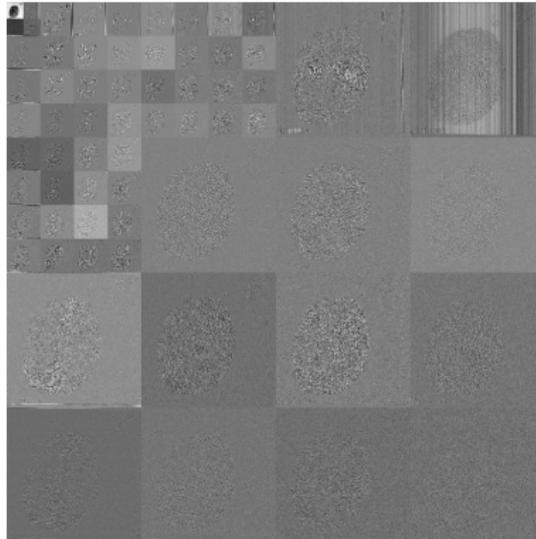
Réponse: Celle de droite pèse 18 fois moins lourd que celle de gauche.

Compression des images (2)

Comment ? quantification d'une transformée bien choisie, qui remplace l'image par une série d'imagettes plus petites, qui en capturent les *détails fins* avec précision.

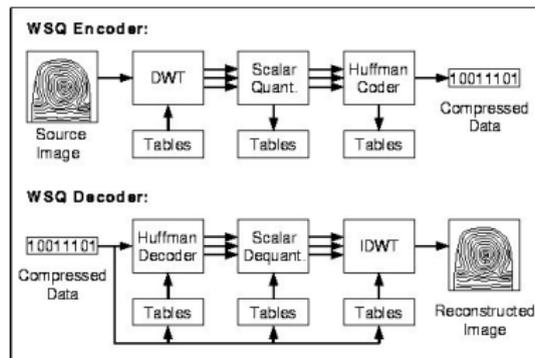
Compression des images (2)

Comment ? quantification d'une transformée bien choisie, qui remplace l'image par une série d'imagettes plus petites, qui en capturent les *détails fins* avec précision.



Compression des images (3)

Plus de détails sur le codeur: transformation, suivie d'une quantification, puis d'un codage binaire.



Points clés:

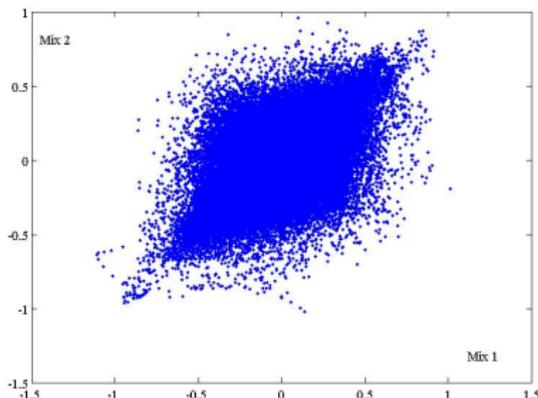
- La transformation est inversible
- Elle génère de la parcimonie

Séparation aveugle de sources pour les nuls

Exemple: mélange d'orgue et xylophone: mélange 1, mélange 2

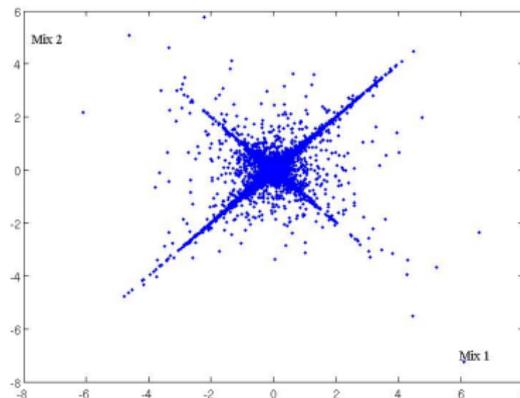
$$\begin{aligned}m_1[t] &= a_{11}x_1[t] + a_{12}x_2[t] & t = 0, \dots, T-1 \\m_2[t] &= a_{21}x_1[t] + a_{22}x_2[t] & t = 0, \dots, T-1,\end{aligned}$$

avec x_1 , x_2 et la matrice A inconnus



Séparation aveugle de sources pour les nuls

Une transformation (isométrique) bien choisie rend les deux signaux parcimonieux, et permet leur séparation



reconstruction orgue

Séparation aveugle de sources pour les pros

Pourquoi ? Notons $T : \mathbb{R}^L \rightarrow \mathbb{R}^L$ la transformation qui “parcimonise”.

- Tx_1 possède $n_1 \ll L$ coefficients significatifs, ou “allumés”, et $L - n_1$ coefficients “éteints”
- Tx_2 possède $n_2 \ll L$ coefficients allumés

Séparation aveugle de sources pour les pros

Pourquoi ? Notons $T : \mathbb{R}^L \rightarrow \mathbb{R}^L$ la transformation qui “parcimonise”.

- Tx_1 possède $n_1 \ll L$ coefficients significatifs, ou “allumés”, et $L - n_1$ coefficients “éteints”
- Tx_2 possède $n_2 \ll L$ coefficients allumés
- La probabilité pour qu’un même coefficient soit allumé à la fois dans Tx_1 et dans Tx_2 est minimale.
- Tm_1 et Tm_2 ont à peu près $n_1 + n_2$ coefficients allumés.

Donc, dans Tm_1 et Tm_2 on peut faire un bon algorithme qui sépare les coefficients allumés à cause de x_1 et ceux qui sont allumés à cause de x_2 , et les attribuer à leur signal “père”.

Parcimonie et réduction de dimension

La notion de parcimonie est cousine de la notion de réduction de dimension, mais pas identique.

Réduction de dimension: étant donné un espace \mathcal{H} , on réduit la dimension en projetant les éléments de \mathcal{H} sur un sous-espace \mathcal{H}_N , de dimension N plus faible. Si on a une base orthonormale $\mathcal{B}_N = \{\phi_1, \dots, \phi_N\}$ de \mathcal{H}_N , la projection s'écrit

$$f_N^{(L)} = \sum_{k=1}^N \langle f, \phi_k \rangle \phi_k$$

En pratique, on recherche le meilleur sous-espace de dimension N , par exemple par des techniques de type ACP.

Parcimonie et réduction de dimension

Approche parcimonieuse: on recherche la meilleure approximation de f

$$f_N^{(NL)} = \sum_{k=1}^N \alpha_{n_k} \phi_{n_k}$$

que l'on puisse obtenir avec N vecteurs de base $\phi_{n_1}, \dots, \phi_{n_N}$

Par exemple, on peut chercher à minimiser

$$\|f - f_N^{(NL)}\|$$

Parcimonie et réduction de dimension

Remarque

- Pour N fixé, l'approximation parcimonieuse est toujours au moins aussi bonne que l'autre: $\|f - f_N^{(NL)}\| \leq \|f - f_N^{(L)}\|$.
- Dans certains espaces fonctionnels, et pour de "bonnes" bases, les vitesses de convergence diffèrent.
- La parcimonie a un prix: dans une perspective de codage, il faut aussi coder les "adresses" n_1, \dots, n_N des atomes "allumés".
- Si $f_N^{(NL)}$ et $g_N^{(NL)}$ sont deux approximations parcimonieuses à N termes de f et g , $f_N^{(NL)} + g_N^{(NL)}$ n'est pas une approximation à N termes de $f + g$ (utile pour la séparation de sources).

- 1 Introduction: analyse, compression et parcimonie
 - Analyse du signal
 - Compression des signaux et des images
 - Séparation de sources
- 2 Gaborettes, ondelettes,... et *-lettres
 - Fourier et Gabor
 - Ondelettes
 - Le cas des images: ondelettes, mdct, brosettes, curvelettes,...
 - Autres *-lettres
 - Apprenti-lettres
- 3 Comment générer de la parcimonie
 - Comment mesurer la parcimonie
 - Bibliothèque de bases: algorithme de meilleure base
 - Approches variationnelles
 - Au delà de la parcimonie
- 4 Conclusions

De Fourier à Gabor: Fourier

Analyse de Fourier: décomposer des signaux comme superpositions de sinusoides:

En dimension finie: les vecteurs $\epsilon_k \in \mathbb{C}^L$ définis par

$$\epsilon_k[n] = \frac{1}{\sqrt{L}} e^{2i\pi kn/L}$$

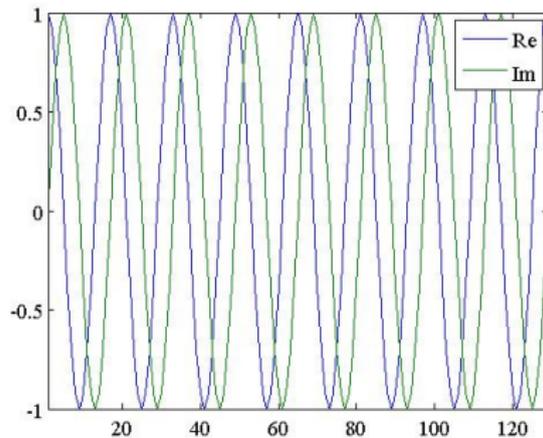
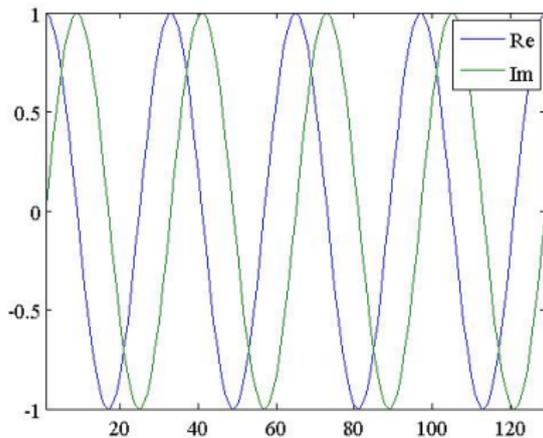
forment une base orthonormale de \mathbb{C}^L .

TFF: $\hat{f}[k] = \sum_{n=0}^{L-1} f[n] e^{-2i\pi kn/L}$

Inversion: $f[n] = \frac{1}{L} \sum_{k=0}^{L-1} \hat{f}[k] e^{2i\pi kn/L}$

Parseval: $\|\hat{f}\|^2 = \sum_{k=0}^{L-1} |\hat{f}[k]|^2 = L \sum_{n=0}^{L-1} |f[n]|^2 = L \|f\|^2$

De Fourier à Gabor: Fourier



De Fourier à Gabor: Fourier

Alternatives: utilisation de cosinus au lieu d'exponentielles complexes.
Par exemple, les vecteurs $u_k \in \mathbb{C}^L$ définis par

$$u_k[n] = \sqrt{\frac{2}{L}} \cos\left(\pi \left(k + \frac{1}{2}\right) n\right)$$

forment eux aussi une base de \mathbb{C}^L (ou \mathbb{R}^L).
La transformation associée

$$x \in \mathbb{C}^L \longmapsto \{\langle x, u_k \rangle, k = 0, \dots, L-1\} \in \mathbb{C}^L$$

est appelée **DCT-IV** (voir wikipedia pour la définition des 7 autres versions de la DCT).

De Fourier à Gabor: Fourier

En dimension infinie:

- Transformation de Fourier Discrète: pour $s \in \ell^2(\mathbb{Z})$,

$$\hat{s}(\nu) = \sum_{n=-\infty}^{\infty} s[n]e^{-2i\pi\nu n}$$

- Coefficients de Fourier: pour $f \in L^2([0, 1])$,

$$c_n(f) = \int_0^1 f(t)e^{-2i\pi n t} dt$$

(remarque: si $s \in \ell^2(\mathbb{Z})$, $s[n] = c_{-n}(\hat{s})$).

- Intégrale de Fourier:

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

De Fourier à Gabor: Fourier

Fourier et régularité:

La vitesse de décroissance des coefficients de Fourier d'une fonction est directement liée à la régularité de celle-ci.

Par exemple, si $f \in C^k([0, 1])$,

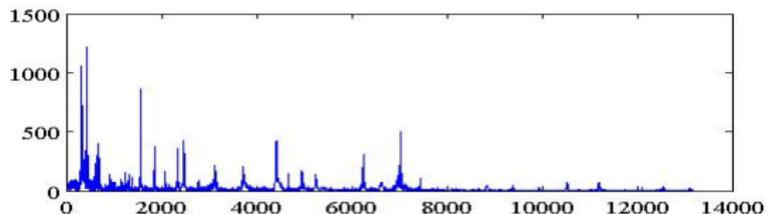
$$|c_n(f)| \leq C|n|^{-k}$$

pour une certaine constante C .

De Fourier à Gabor: Fourier

L'analyse de Fourier s'avère insuffisante dès lors que les caractéristiques fréquentielles du signal varient avec le temps (on parle d'*instationnarité*)

- Instationnarité "lente": les caractéristiques fréquentielles varient avec le temps
- Instationnarité rapide (transitoire): singularités,...
- ...



Transformée de Fourier du son précédent

De Fourier à Gabor: MDCT

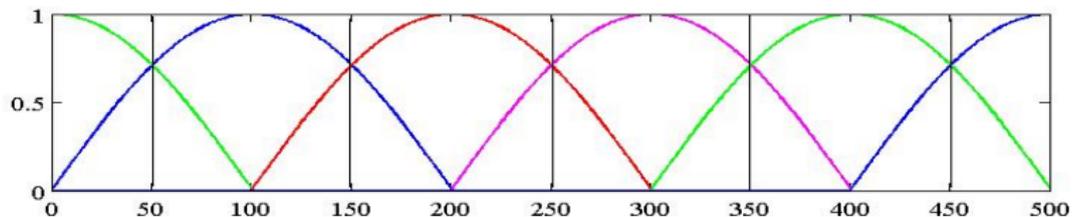
MDCT: version localisée de la transformation DCT-IV. En dimension L , soit N un diviseur de L .

- \mathbb{Z}_L est segmenté en $K = L/N$ intervalles entiers de longueur N
- Pour tout $k = 0, \dots, K - 1$, soit $w_k \in \mathbb{C}^L$ telle que
 - $w_k[t] = 0$ pour $t < (k - 1/2)N$ et $t > (k + 3/2)N$.
 - $w_k[kL + \tau] = w_{k+1}[kL - \tau] \forall \tau = 1 - L/2, \dots, L/2 - 1$
 - $w_k[kL + \tau]^2 + w_{k+1}[kL + \tau]^2 = 1 \forall \tau = 1 - L/2, \dots, L/2 - 1$
- On note $u_{kn} \in \mathbb{C}^L$ les vecteurs définis par leurs composantes

$$u_{kn}[t] = \sqrt{\frac{2}{N}} w_k[t] \cos \left(\pi \left(n + \frac{1}{2} \right) (t - kN) \right)$$

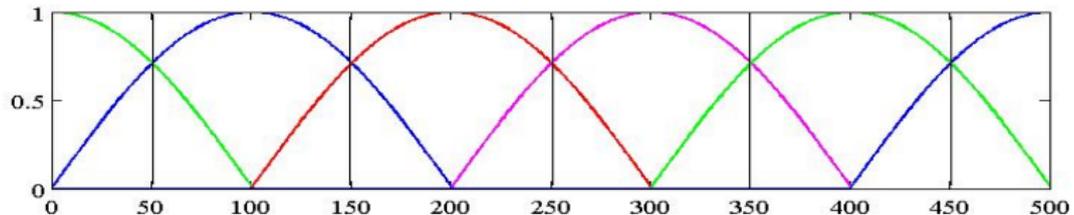
De Fourier à Gabor: MDCT

Les voilà (les fenêtres):



De Fourier à Gabor: MDCT

Les voilà (les fenêtres):



Théorème

Sous ces conditions, la famille de fenêtres modulées

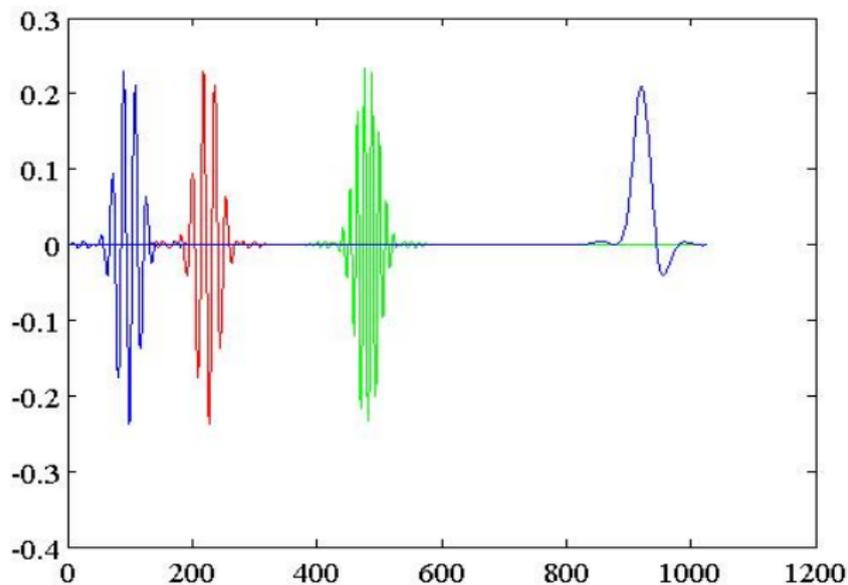
$\{u_{kn}, k = 1, \dots, K, n = 0, \dots, N - 1\}$ *est une base orthonormale de \mathbb{C}^L .*

Remarque

On construit de même des bases orthonormales de $L^2(\mathbb{R})$, $\ell^2(\mathbb{Z})$ et $L^2([a, b])$, sans difficulté.

De Fourier à Gabor: MDCT

Exemples d'atomes MDCT



De Fourier à Gabor: MDCT

Sous Matlab/Octave, la toolbox Ltfat (P. Soendergaard) fournit une implémentation efficace de la MDCT... et de multiples variantes

LTFAT

```
> ltfatstart  
> x=wavread('ydkm.wav');  
> L = length(x)  
> NbFreq = 1024;  
> g = wilorth(NbFreq,L);  
> c = mdct(x,g,NbFreq)  
> imagesc(log(abs(c)));  
> set(gca,'ydir','normal');
```

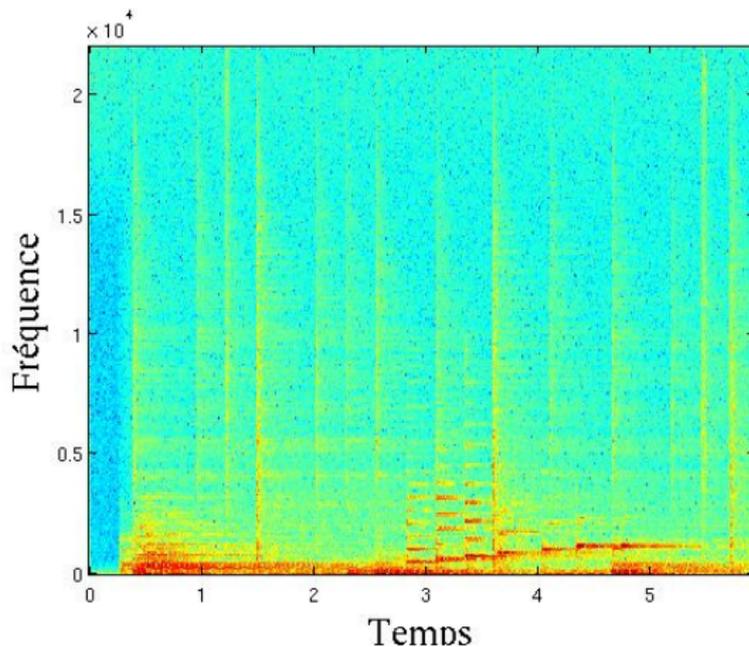
Taille des fenêtres

Génération de la fenêtre

Calcul de la MDCT

Représentation graphique

De Fourier à Gabor: MDCT



De Fourier à Gabor: MDCT

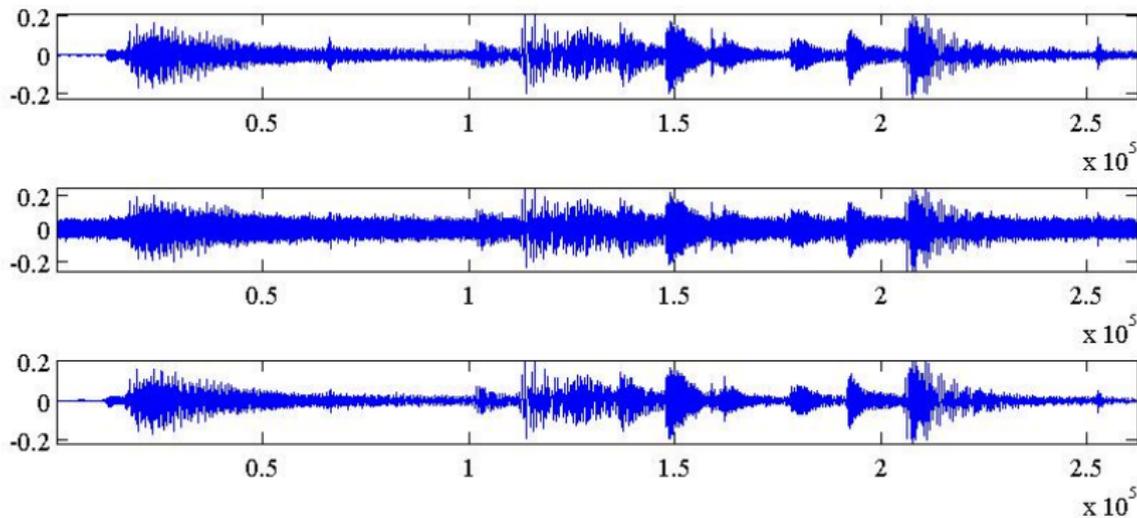
Inversion et Application: débruitage:

LTFAT

```
> nx = x + 0.5*std(x)*randn(size(x));           Ajout de bruit
> nc = mdct(nx,g,NbFreq);
> Threshold = 0.06;                             une valeur raisonnable
> SignifMap = abs(nc) > Threshold;
> nc2 = nc.*SignifMap;                          Seuillage
> rec = imdct(nc2,g,NbFreq);                    MDCT inverse
> norm(x-nx)/norm(x)                            réponse: 0.4998
> norm(x-rec)/norm(x)                          réponse: 0.1639
```

Dans les cas simples (bruit blanc Gaussien, de variance connue ou bien estimée), on peut obtenir une estimation pour la *valeur raisonnable*...

De Fourier à Gabor: MDCT



De Fourier à Gabor: Gabor

Les conditions sur les fenêtres sont en fait assez restrictives. Dans un cadre plus général, on peut être amené à s'affranchir de l'hypothèse "base" pour utiliser de meilleures "fonctions de base".

Atomes de Gabor: copies modulées et translatées d'une fenêtre de référence.

$$g_{kn}[t] = e^{2i\pi n\nu_0(t-kb_0)/L} g[t - kb_0], \quad k = 0, \dots, K-1, n = 0, \dots, N-1$$

où ν_0 et b_0 sont des constantes (diviseurs de L), et où $K = L/b_0$ et $N = L/\nu_0$.

Etant donné $f \in \mathbb{C}^L$, la famille de coefficients

$$\mathcal{V}_g f[k, n] = \langle f, g_{kn} \rangle = \sum_{t=0}^{L-1} f[t] \bar{g}[t - kb_0] e^{-2i\pi n\nu_0(t-kb_0)/L}$$

forme une **transformée de Fourier à fenêtre** (si $b_0 = \nu_0 = 1$) ou une **transformée de Gabor** de f .

De Fourier à Gabor: Gabor

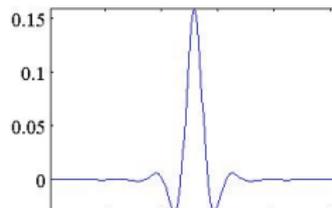
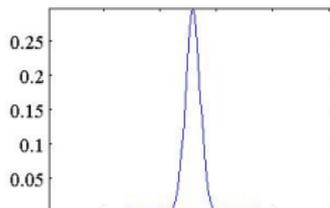
Il n'existe pas de (bonne) base d'atomes de Gabor, mais sous certaines conditions sur g , b_0 et ν_0 on peut montrer que la famille $\{g_{kn}, k = 0, \dots, K-1, n = 0, \dots, N-1\}$ est un **repère** de \mathbb{C}^L , c'est à dire que la transformation

$$f \in \mathbb{C}^L \longmapsto \{\langle f, g_{kn} \rangle, k = 0, \dots, K-1, n = 0, \dots, N-1\} \in \mathbb{C}^{KN}$$

est inversible.

Dans ce cas, il existe une autre fenêtre $\tilde{g} \in \mathbb{C}^L$ (non unique), appelée fenêtre duale, telle que

$$f = \sum_{k,n} \langle f, g_{kn} \rangle \tilde{g}_{kn} = \sum_{k,n} \langle f, \tilde{g}_{kn} \rangle g_{kn}$$



De Fourier à Gabor: Gabor

Dans la toolbox Ltfat: la fonction `candual` génère une fenêtre duale canonique d'une fenêtre d'analyse g donnée, pour b_0 et ν_0 fixés.

Syntaxe: `gamma = candual(g, b_0, N)` (avec $N > b_0$).

LTFAT

```
> g = pgauss(256);                               Fenêtre Gaussienne
> plot(fftshift(g));
> gamma = candual(g,4,16);                         Calcul d'une fenêtre duale
> plot(fftshift(gamma));
> gamma = candual(g,4,8);                          Calcul d'une fenêtre duale
> plot(fftshift(gamma));
```

Remarque: plus N s'approche de b_0 , plus la fenêtre duale est différente de la fenêtre de départ. Pour $N = b_0$, la fenêtre duale n'existe plus.

De Fourier à Gabor: Gabor

Transformation de Gabor:

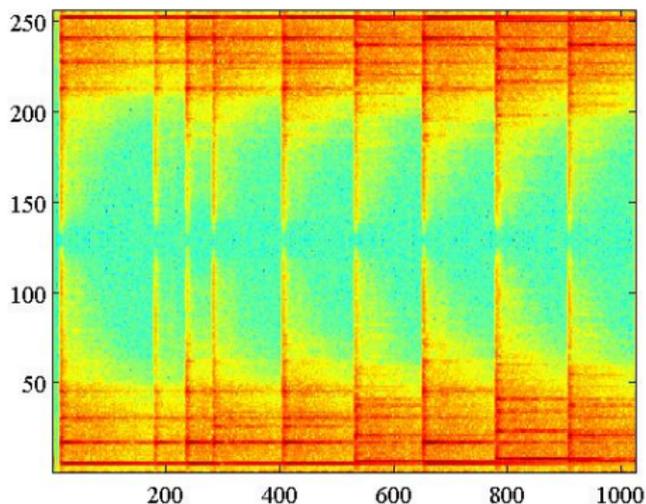
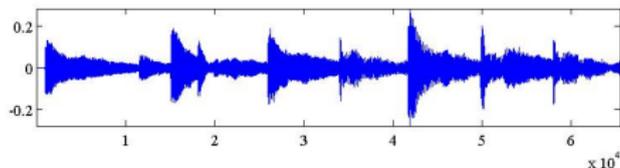
LTFAT

```
> x = xylo;                                xylophone  
> g = pgauss(length(x));                  Fenêtre Gaussienne  
> TimeStep = 64;                          Pas d'échantillonnage temporel  
> NbFreq = 256;                            Nombre de valeurs de la fréquence  
> gc = dgt(x,g,TimeStep,NbFreq);          Calcul de la transformée  
> imagesc(abs(gc));                        Représentation graphique  
> set(gca,'ydir','normal');
```

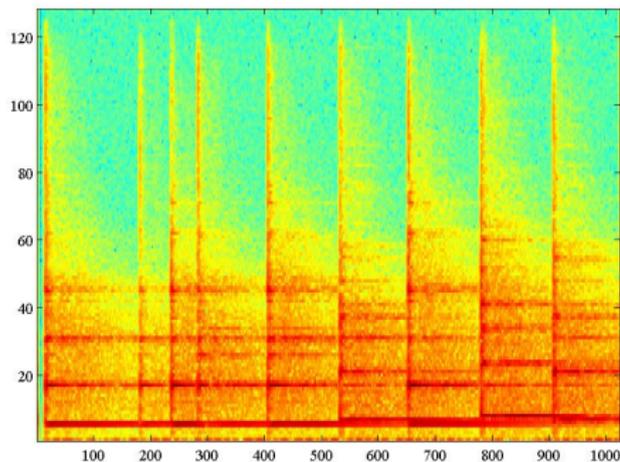
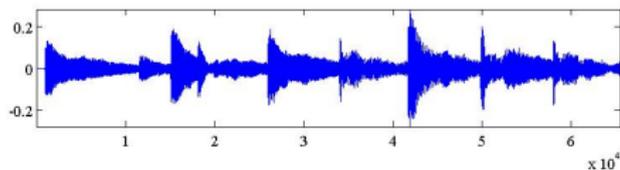
Remarque

La représentation graphique représente les fréquences positives et négatives.

De Fourier à Gabor: Gabor



De Fourier à Gabor: Gabor



De Fourier à Gabor

Transformation inverse:

il faut tout d'abord générer la fenêtre duale γ .

LTFAT

```
> gamma = candual(g,TimeStep,NbFreq);           fenêtre duale  
> xrec = idgt(tgc,gamma,TimeStep);  
> norm(rec-xylo)/norm(xylo)                   réponse: 3.9856e-16
```

De là, on peut jouer les mêmes jeux que précédemment: par exemple, sélection des coefficients les plus significatifs et inversion.

De Fourier à Gabor

Transformation inverse:

il faut tout d'abord générer la fenêtre duale γ .

LTFAT

```
> gamma = candual(g, TimeStep, NbFreq);           fenêtre duale  
> xrec = idgt(tgc, gamma, TimeStep);  
> norm(rec-xylo)/norm(xylo)                       réponse: 3.9856e-16
```

De là, on peut jouer les mêmes jeux que précédemment: par exemple, sélection des coefficients les plus significatifs et inversion.

Remarque

Le fait que les atomes de Gabor ne forment pas une base mais un repère complique l'opération de seuillage...

Intermède: repère dans un Hilbert

Définition (Duffin & Schaeffer)

Dans un espace de Hilbert \mathcal{H} , une famille de vecteurs $\{\phi_\lambda, \lambda \in \Lambda\}$ est un repère de \mathcal{H} si il existe deux constantes A, B telles que pour tout $f \in \mathcal{H}$, on ait

$$0 < A\|f\|^2 \leq \sum_{\lambda \in \Lambda} |\langle f, \phi_\lambda \rangle|^2 \leq B\|f\|^2 < \infty .$$

Etant donné un repère $\{\phi_\lambda, \lambda \in \Lambda\}$ de \mathcal{H} , on introduit

- L'opérateur d'analyse $U : f \in \mathcal{H} \rightarrow \{\langle f, \phi_\lambda \rangle, \lambda \in \Lambda\} \in \ell^2(\Lambda)$
- L'opérateur de synthèse $U^* : \alpha \in \ell^2(\Lambda) \rightarrow \sum_{\lambda \in \Lambda} \alpha_\lambda \phi_\lambda \in \mathcal{H}$
- L'opérateur de repère (borné, inversible à inverse borné)

$$\mathcal{R} = U^*U : f \in \mathcal{H} \mapsto \sum_{\lambda \in \Lambda} \langle f, \phi_\lambda \rangle \phi_\lambda$$

Intermède: repère dans un Hilbert

Remarque

- 1 En dimension finie, la notion de repère coïncide avec la notion de famille génératrice.
- 2 Etant donné un repère $\{\phi_\lambda, \lambda \in \Lambda\}$ de \mathcal{H} , en notant $\tilde{\phi}_\lambda = \mathcal{R}^{-1}\phi_\lambda$ la famille $\{\tilde{\phi}_\lambda, \lambda \in \Lambda\}$ est aussi un repère de \mathcal{H} , appelé **repère dual canonique**. On a

$$f = \sum_{\lambda \in \Lambda} \langle f, \tilde{\phi}_\lambda \rangle \phi_\lambda = \sum_{\lambda \in \Lambda} \langle f, \phi_\lambda \rangle \tilde{\phi}_\lambda$$

- 3 On peut montrer que la famille de vecteurs $\gamma_\lambda = \mathcal{R}^{-1/2}\phi_\lambda$, $\lambda \in \Lambda$ est un repère **strict**, au sens où ses deux bornes A et B sont égales. On a alors

$$f = \sum_{\lambda \in \Lambda} \langle f, \gamma_\lambda \rangle \gamma_\lambda$$

Intermède: repère dans un Hilbert

Remarque

- 1 Lorsque le repère n'est pas une base, tout vecteur f admet une infinité de décompositions de la forme

$$f = \sum_{\lambda \in \Lambda} \alpha_{\lambda} \phi_{\lambda} .$$

Ceci permet de se poser le problème de la meilleure décomposition, dans un sens à définir (parcimonie ?).

- 2 Si \mathcal{H} est un espace de fonctions, les noyaux

$$k(t, s) = \sum_{\lambda \in \Lambda} \phi_{\lambda}(t) \overline{\phi_{\lambda}(s)} \quad \tilde{k}(t, s) = \sum_{\lambda \in \Lambda} \tilde{\phi}_{\lambda}(t) \overline{\phi_{\lambda}(s)}$$

sont "moralement" des noyaux reproduisants de \mathcal{H} .

Ondelettes

Dans le cas des bases MDCT, ou des Gaborettes, les tailles de fenêtres sont fixées une fois pour toutes.

Les **ondelettes** (Grossmann & Morlet, 1983) sont des fonctions de taille variable, qui s'adaptent aux détails du vecteur analysé: partant d'une ondelette de base $\psi(t)$, on considère des copies dilatées et translatées

$$\psi_{(b,a)}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right)$$

ou, pour $b_0, a_0 \in \mathbb{R}^+$,

$$\psi_{jk}(t) = a_0^{-j/2} \psi\left(a_0^{-j} t - kb_0\right)$$

Ondelettes

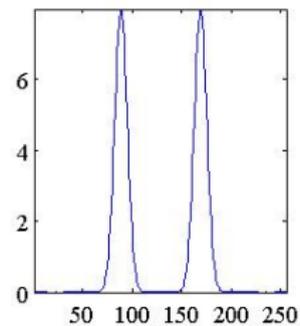
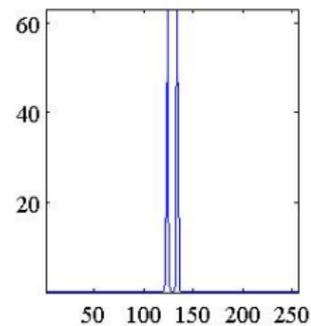
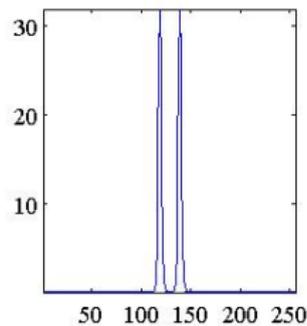
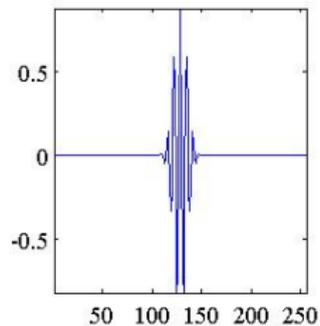
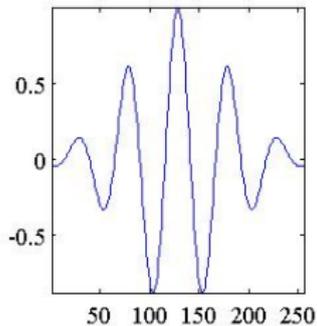
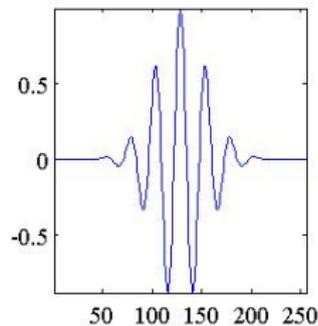
De nouveau, il est possible de montrer que pour des ψ , b_0 et a_0 bien choisis, la famille de fonctions $\{\psi_{jk}, j, k \in \mathbb{Z}\}$ est un **repère** de $L^2(\mathbb{R})$.

Pour tout $f \in L^2(\mathbb{R})$, on peut écrire

$$\begin{aligned} f &= \sum_{j,k} \langle f, \psi_{jk} \rangle \tilde{\psi}_{jk} \\ &= \sum_{j,k} \langle f, \tilde{\psi}_{jk} \rangle \psi_{jk} \end{aligned}$$

où $\{\tilde{\psi}_{jk}, j, k \in \mathbb{Z}\}$ est un **repère dual**.

Ondelettes



Ondelettes

Qu'apportent les ondelettes ?

- Moments nuls: $\int t^m \psi(t) dt = 0$ pour $m = 0, \dots, M - 1$: les ondelettes sont aveugles aux polynômes de degré $\leq M - 1$. Ceci génère de la parcimonie.
- La capacité de “zoomer” sur des portions où le signal varie rapidement (par exemples, des singularités).
- Mathématiquement, des approximations non-linéaires optimales dans certains espaces fonctionnels (qu'on pense être de bons modèles pour certains types de signaux).

Ondelettes

Le rôle des oscillations et des moments nuls: soit ψ une ondelette à M moments nuls:

$$\int t^m \psi(t) dt = 0, \quad m = 0, \dots, M-1$$

Si $f \in C^{M-1}(\mathbb{R})$, on développe au voisinage de t_0

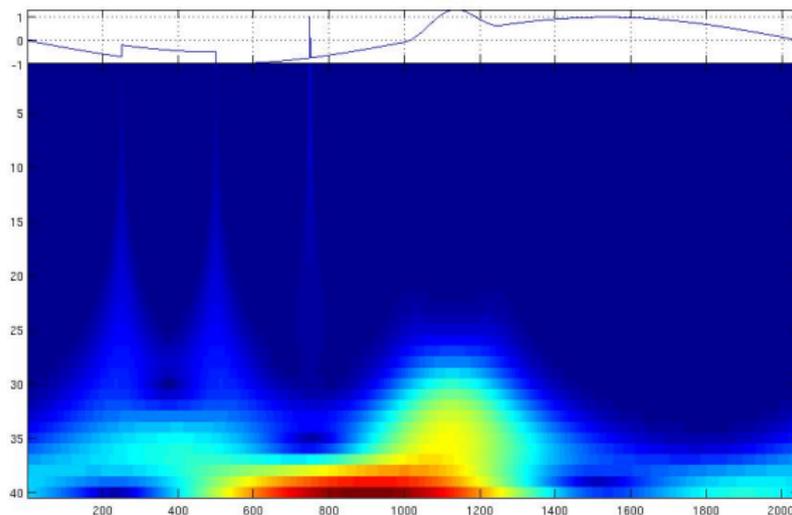
$$f(t) = f(t_0) + (t-t_0)f'(t_0) + \dots + \frac{1}{(M-1)!}(t-t_0)^{M-1}f^{(M-1)}(t_0) + r(t)$$

et

$$\begin{aligned} \langle f, \psi_{(b,a)} \rangle &= \int_{-\infty}^{\infty} f(t) \psi_{(b,a)}(t) dt \\ &= f(t_0) \int_{-\infty}^{\infty} \psi_{(b,a)}(t) dt + f'(t_0) \int_{-\infty}^{\infty} (t-t_0) \psi_{(b,a)}(t) dt + \dots + \\ &= \int_{-\infty}^{\infty} r(t) \psi_{(b,a)}(t) dt \end{aligned}$$

Ondelettes

Une image de coefficients d'ondelettes dans un cas d'école



Ondelettes: bases

Théorème (Mallat, Meyer)

Il existe $\psi \in L^2(\mathbb{R})$ telle que la collection des ondelettes dilatées et translattées ψ_{jk} définies par

$$\psi_{jk}(t) = 2^{-j/2} \psi(2^{-j}t - k)$$

soit une base orthonormée de $L^2(\mathbb{R})$: $\forall f \in L^2(\mathbb{R})$,

$$f = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk}, \quad \text{avec } d_{jk} = \langle f, \psi_{jk} \rangle.$$

Ondelettes: bases (2)

Interprétation: à j correspond l'échelle 2^j

- Pour j_0 fixé,

$$S_{j_0} f = \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk}$$

est une **approximation** de f , à une échelle approximativement égale à 2^{j_0} .

Ondelettes: bases (2)

Interprétation: à j correspond l'échelle 2^j

- Pour j_0 fixé,

$$S_{j_0} f = \sum_{j=j_0}^{\infty} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk}$$

est une **approximation** de f , à une échelle approximativement égale à 2^{j_0} .

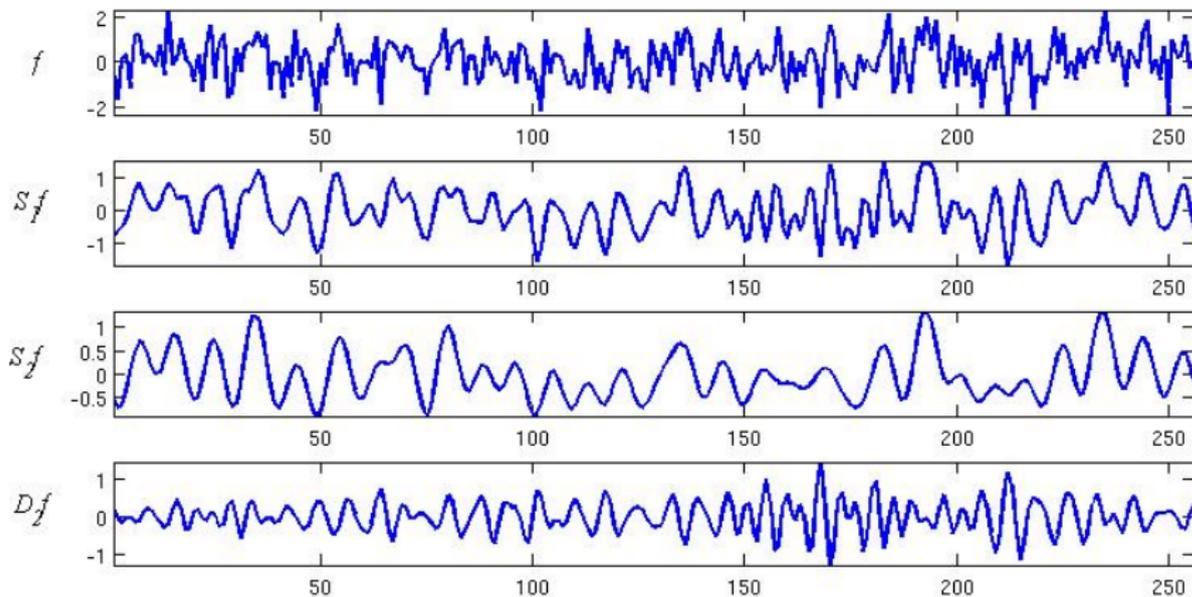
- Pour j_0 fixé, la composante

$$D_{j_0} f = \sum_{k=-\infty}^{\infty} d_{j_0 k} \psi_{j_0 k}$$

représente les **détails** de f de taille approximativement égale à 2^{j_0} ,
i.e. visibles dans $S_{01} f$ mais plus dans $S_{j_0+1} f$.

Ondelettes: bases (3)

Approximations et détails



Ondelettes: multirésolution et passage au discret

Cette logique d'**approximation et détails** peut être poussée plus loin.

Une **analyse multirésolution** est basée sur l'existence simultanée d'une **ondelette** ψ et d'une **fonction d'échelle** ϕ , telles que $\forall f \in L^2(\mathbb{R})$,

$$D_j f = \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk}, \quad S_j f = \sum_{k=-\infty}^{\infty} s_{jk} \phi_{jk},$$

où

$$d_{jk} = \langle f, \psi_{jk} \rangle, \quad s_{jk} = \langle f, \phi_{jk} \rangle.$$

Les coefficients s sont les coefficients d'approximation (ou d'échelle), et les coefficients d sont les coefficients de détail (ou d'ondelettes).

Ondelettes: multirésolution et passage au discret

On écrit, $\forall f \in L^2(\mathbb{R})$

$$\begin{aligned} f &= S_{j_0} f + \sum_{j=-\infty}^{j_0} D_j f \\ &= \sum_{k=-\infty}^{\infty} s_{j_0 k} \phi_{j_0 k} + \sum_{j=-\infty}^{j_0} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk} \end{aligned}$$

Ondelettes: multirésolution et passage au discret

On écrit, $\forall f \in L^2(\mathbb{R})$

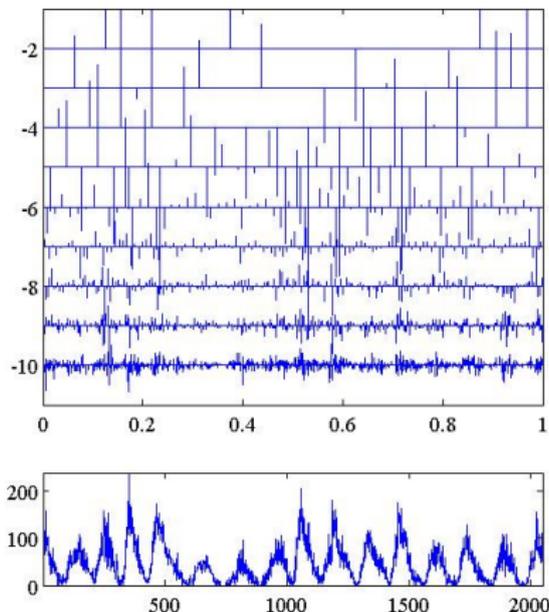
$$\begin{aligned} f &= S_{j_0} f + \sum_{j=-\infty}^{j_0} D_j f \\ &= \sum_{k=-\infty}^{\infty} s_{j_0 k} \phi_{j_0 k} + \sum_{j=-\infty}^{j_0} \sum_{k=-\infty}^{\infty} d_{jk} \psi_{jk} \end{aligned}$$

- $S_j f$ est la projection de f sur un **espace d'approximation**, engendré par les translatées ϕ_{jk} de ϕ_{j_0} , noté \mathcal{V}_j .
- $D_j f$ est la projection de f sur un **espace de détails**, engendré par les translatées ψ_{jk} de ψ_{j_0} , noté \mathcal{W}_j .

$$\mathcal{V}_{j_1} = \mathcal{V}_{j_0} \oplus \bigoplus_{j=j_1}^{j_0-1} \mathcal{W}_j$$

Ondelettes: multirésolution et passage au discret

Exemple: taches solaires



Ondelettes: multirésolution et passage au discret

Dans le cadre multirésolution, l'ondelette et la fonction d'échelle sont complètement caractérisées par deux filtres numériques (deux suites), notés g et h , et appelés **filtres miroir en quadrature** (qmf).

Ondelettes: multirésolution et passage au discret

Dans le cadre multirésolution, l'ondelette et la fonction d'échelle sont complètement caractérisées par deux filtres numériques (deux suites), notés g et h , et appelés **filtres miroir en quadrature** (qmf).

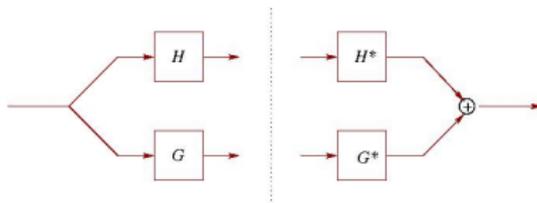
On leur associe deux opérateurs H et G (filtrage suivi d'un sous-échantillonnage) agissant sur les suites de coefficients

$$H : \mathbf{x} \in \ell^2(\mathbb{Z}) \mapsto H\mathbf{x} \in \ell^2(2\mathbb{Z}) : \quad (H\mathbf{x})_k = \sum_{\ell} h_{\ell} x_{2k-\ell}$$

$$G : \mathbf{x} \in \ell^2(\mathbb{Z}) \mapsto G\mathbf{x} \in \ell^2(2\mathbb{Z}) : \quad (G\mathbf{x})_k = \sum_{\ell} g_{\ell} x_{2k-\ell} \cdot$$

qui satisfont la **condition de reconstruction parfaite**

$$H^*H + G^*G = 1$$



Ondelettes: multirésolution et passage au discret

Algorithme:

Décomposition: $\mathbf{s}_j = H\mathbf{s}_{j-1}$, et $\mathbf{d}_j = G\mathbf{s}_{j-1}$,

$$s_{jk} = \sum_{\ell} h_{\ell} s_{j-1, 2k-\ell}, \quad d_{jk} = \sum_{\ell} g_{\ell} s_{j-1, 2k-\ell}.$$

Reconstruction: $\mathbf{s}_{j-1} = H^*\mathbf{s}_j + G^*\mathbf{d}_j$.

$$s_{j-1, k} = \sum_{\ell} \bar{h}_{2\ell-k} s_{j\ell} + \sum_{\ell} \bar{g}_{2\ell-k} d_{j\ell}$$

Complexité: $O(N)$.

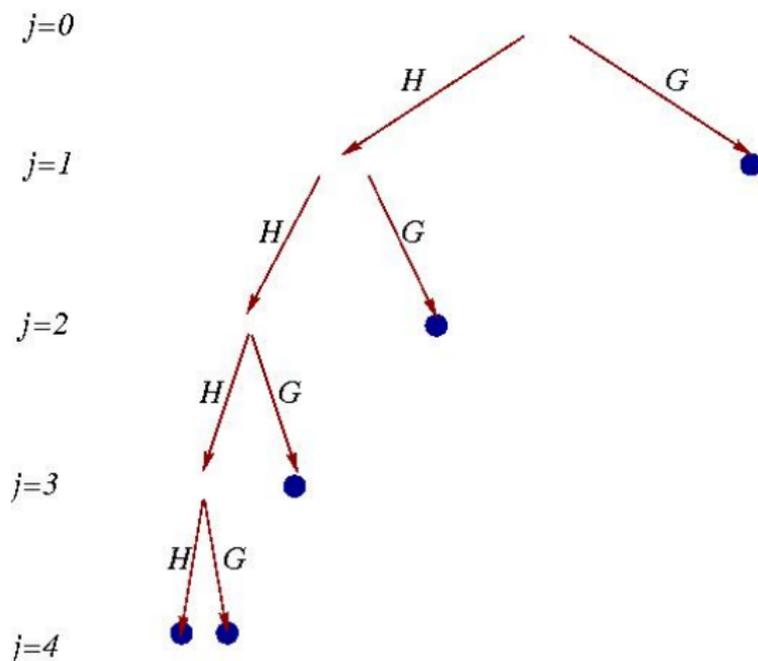
Ondelettes: multirésolution et passage au discret

Les deux filtres QMF h et g caractérisent complètement l'ondelette et la fonction d'échelle.

Il existe une multitude de filtres QMF, référencés en fonction de leurs propriétés.

- Ondelettes "splines": correspondent à des espaces d'approximation de fonctions splines. L'ondelette ψ et la fonction d'échelle ϕ ont une transformée de Fourier à support borné.
- Ondelettes de Daubechies: à support compact.
- Ondelettes de Littlewood-Paley, Coiflettes, ondelettes de Vaidyanathan,...
- Généralisations (ondelettes biorthogonales, multi-ondelettes, ...).

Ondelettes: multirésolution et passage au discret



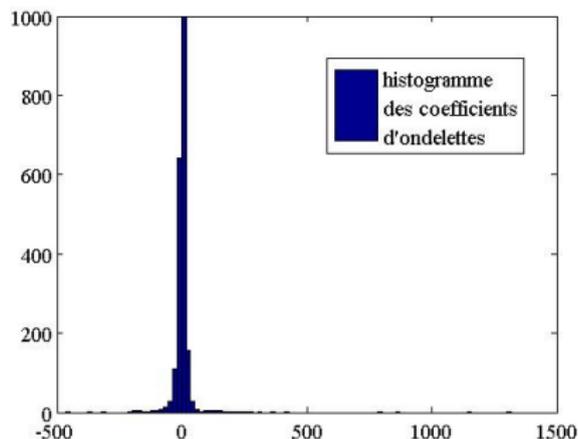
Ondelettes: multirésolution et passage au discret

La toolbox WaveLab implémente (entre autres choses) de multiples versions de la décomposition en ondelettes, notamment FWT_PO: transformée orthogonale périodisée.

Wavelab

```
> qmf = MakeONFilter('Daubechies',4);  
> sig = ReadSignal('Sunspots');  
> depth = 2;  
> wc = FWT_PO(sig,depth,g);  
> PlotWaveCoeff(wc,depth);  
> hist(wc,100);
```

Ondelettes: multirésolution et passage au discret



La distribution des coefficients est “piquée” autour de la valeur zéro. Ceci suggère de seuiller les coefficients pour simplifier la transformée.

Ondelettes: multirésolution et passage au discret

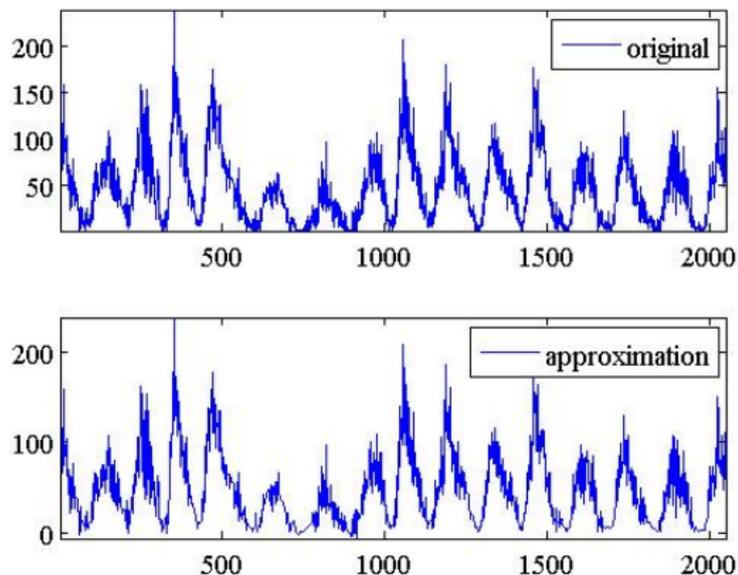
Seuillage:

Wavelab

```
> rec = IWT_P0(wc,depth,qmf);  
> threshold = 10; au pif !  
> SignifMap = abs(wc) > threshold;  
> length(find(SignifMap)) réponse: 37%  
> rec2 = IWT_P0(wc.*SignifMap,depth,qmf);  
> norm(rec2-rec1)/norm(rec1); réponse: 6%  
> subplot(2,1,1);plot(rec);axis tight;  
> subplot(2,1,2);plot(rec2);axis tight;
```

Ondelettes: multirésolution et passage au discret

Qualité de la reconstruction:



le cas 2D: ondelettes et MDCT

Produit tensoriel: La façon la plus simple de générer des bases, repères, dictionnaires 2D est de prendre des produits tensoriels de bases, repères,... 1D.

Par exemple: Dans $\mathbb{C}^{L^2} = \mathbb{C}^L \times \mathbb{C}^L$, partant d'une base $\{\phi_k, k = 0, \dots, L - 1\}$ de \mathbb{C}^L , la famille $\{\phi_{k;\ell}, k, \ell = 0, \dots, L - 1\}$ définie par

$$\phi_{k;\ell}[m, n] = \phi_k[m]\phi_\ell[n]$$

est une base de \mathbb{C}^{L^2} appelée base "produit tensoriel".

le cas 2D: ondelettes et MDCT

Application: codage d'images: Le codeur JPEG utilise un produit tensoriel de deux bases MDCT ($N=8$) pour coder les images.

Problème: existence de directions privilégiées dans l'image, qui se traduisent par des effets de "bloc" à bas débit.

Ce problème est partiellement corrigé par l'utilisation d'ondelettes 2D (JPEG2000).

Alternative: ondelettes "directionnelles" (Murenzi, 1986).

le cas 2D: ondelettes directionnelles

Pour éviter les directions privilégiées: on introduit un degré de liberté angulaire:

$$\psi_{(b,a,\theta)}(x) = \frac{1}{a} \psi \left(r_{-\theta} \frac{x-b}{a} \right)$$

où

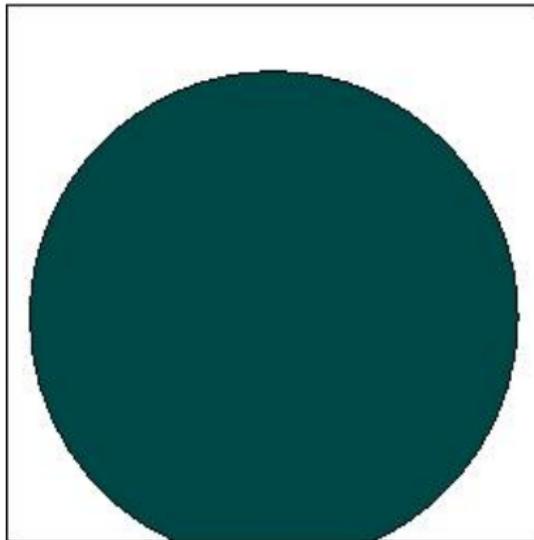
$$r_{-\theta} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

est la matrice de rotation d'angle $-\theta$.

L'ondelette oscille dans une direction; en variant θ , on varie la direction des oscillations, et donc la direction dans laquelle l'ondelette détecte des variations rapides.

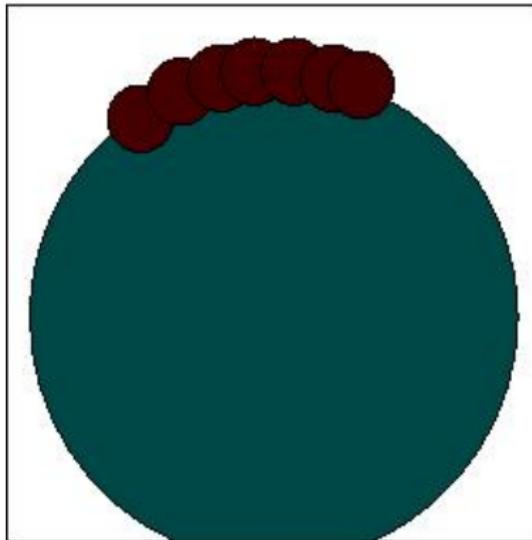
le cas 2D: ondelettes directionnelles

Les ondelettes directionnelles ne génèrent pas de parcimonie pour les images "cartoon": la dilatation n'est pas une bonne opération pour cela.



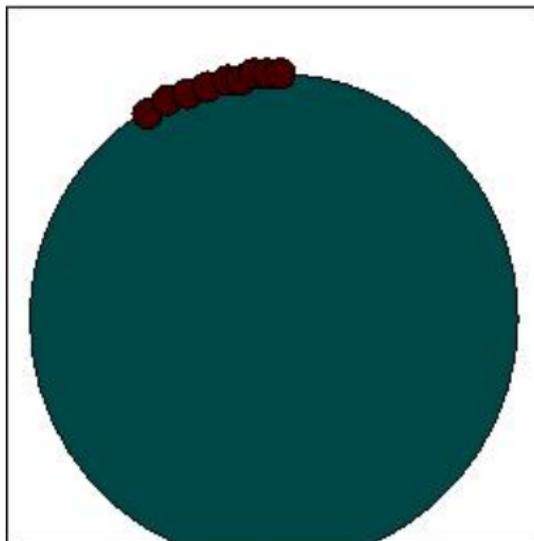
le cas 2D: ondelettes directionnelles

Les ondelettes directionnelles ne génèrent pas de parcimonie pour les images "cartoon": la dilatation n'est pas une bonne opération pour cela.



le cas 2D: ondelettes directionnelles

Les ondelettes directionnelles ne génèrent pas de parcimonie pour les images "cartoon": la dilatation n'est pas une bonne opération pour cela.



le cas 2D: curvelettes

- La vitesse de convergence de l'approximation non-linéaire une image constante par morceaux avec bords C^1 n'est pas la vitesse optimale.
- **Curvelettes:** (Candès, Donoho, Guo, Demanet) remplacent la dilatation usuelle par une dilatation anisotrope:

$$\psi_a(x_1, x_2) = \psi\left(\frac{x_1}{a}, \frac{x_2}{\sqrt{a}}\right), \quad \psi_{\mathbf{b}, a, \theta}(\mathbf{x}) = \psi_a(r_{-\theta}(\mathbf{x} - \mathbf{b}))$$

Résultat: amélioration de la vitesse de convergence d'approximations parcimonieuses pour des images présentant des singularités sur des contours.

Autres *-lettres

Depuis quelques années, les *-lettres envahissent le monde scientifique:

- 1D: ondelettes, Gaborlettres, chirplettes, ...
- 2D: brushlets (brossettes ?), ridgelets, beamlets, curvelets (courbettes ?), shearlets (ciselettes ?), Fresnelets, bandelettes,...
- ...

Chaque nouvelle construction est destinée à obtenir des approximations parcimonieuses les plus efficaces possibles pour un type ou un autre de situation.

Plusieurs situations: base, repère, ou dictionnaire

Apprenti-lettres

Idée: apprendre un dictionnaire d'atomes à partir d'un ensemble de K signaux $f^{(1)}, \dots, f^{(K)}$

Régression parcimonieuse:

$$\min_{\alpha} \left[\|f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}\|^2 + \mu \Psi[\alpha] \right]$$

où Ψ pénalise les suites de coefficients non parcimonieuses.

Apprenti-lettres

Idée: apprendre un dictionnaire d'atomes à partir d'un ensemble de K signaux $f^{(1)}, \dots, f^{(K)}$

Régression parcimonieuse:

$$\min_{\alpha} \left[\|f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}\|^2 + \mu \Psi[\alpha] \right]$$

où Ψ pénalise les suites de coefficients non parcimonieuses.

Apprentissage de dictionnaire:

$$\min_{\alpha, \phi} \frac{1}{K} \sum_{k=1}^K \left[\|f^{(k)} - \sum_{\lambda} \alpha_{\lambda}^{(k)} \phi_{\lambda}\|^2 + \mu \Psi[\alpha^{(k)}] \right]$$

Problème cousin de la séparation de source. Algorithmes: K -SVD (Aharon et al), neurosciences (Olshausen & Field, Lewicki,...)

- 1 Introduction: analyse, compression et parcimonie
 - Analyse du signal
 - Compression des signaux et des images
 - Séparation de sources
- 2 Gaborettes, ondelettes,... et *-lettres
 - Fourier et Gabor
 - Ondelettes
 - Le cas des images: ondelettes, mdct, brosettes, curvelettes,...
 - Autres *-lettres
 - Apprenti-lettres
- 3 Comment générer de la parcimonie
 - Comment mesurer la parcimonie
 - Bibliothèque de bases: algorithme de meilleure base
 - Approches variationnelles
 - Au delà de la parcimonie
- 4 Conclusions

Généralités

Objectif: étant donné un dictionnaire (repère, union de bases,...), construire une méthode fournissant un développement le plus parcimonieux possible des signaux par rapport à ce dictionnaire.

- Notion de parcimonie
- Formulation du problème: maximisation d'un critère de parcimonie donné.
- Algorithme

Comment mesurer la parcimonie

Mesures de **diversité**: pour un vecteur normalisé $f \in \mathbb{C}^L$

- Normes ℓ^p , pour $p < 2$:

$$\|f\|_p = \left(\sum_{n=0}^{L-1} |f_n|^p \right)^{1/p}$$

- Entropie de Shannon

$$S(f) = - \sum_{n=0}^{L-1} |f_n|^2 \log_2(|f_n|^2)$$

- Entropies de Renyi

$$R_\alpha(f) = \frac{1}{1-\alpha} \log_2(\|f\|_\alpha^\alpha)$$

Ces quantités sont d'autant plus grandes que f est diffus (peu concentré).

Comment mesurer la parcimonie (2)

Objectif: parmi un grand nombre de décompositions possibles d'un signal (a priori sur une famille redondante, repère ou dictionnaire), sélectionner la plus parcimonieuse.

- L'approche "best basis"
- Les approches "greedy"
- Les approches variationnelles
- Les approches probabilistes
- ...

L'approche "best basis"

Ingrédients de base: (Coifman et Wickerhauser)

- 1 Une grande **bibliothèque structurée** de bases orthonormales de \mathbb{C}^N , et si possible des algorithmes rapides de décomposition.
- 2 Un critère de parcimonie additif: entropie, norme ℓ^p , ...

Recherche exhaustive de la base qui produit la décomposition la plus parcimonieuse, i.e. qui optimise le critère, par programmation dynamique.

L'approche "best basis"

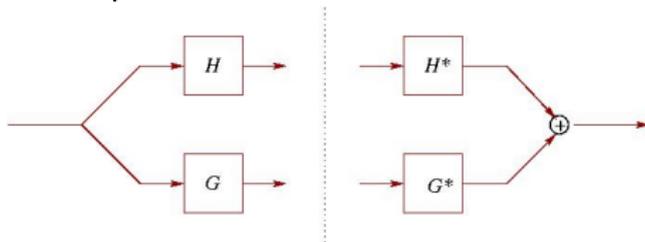
Exemple: les paquets d'ondelettes utilisent les opérateurs H et G utilisés pour les décompositions en ondelettes.

$$H : \mathbf{x} \in \ell^2(\mathbb{Z}) \mapsto H\mathbf{x} : \quad (H\mathbf{x})_k = \sum_{\ell} h_{\ell} x_{2k-\ell}$$
$$G : \mathbf{x} \in \ell^2(\mathbb{Z}) \mapsto G\mathbf{x} : \quad (G\mathbf{x})_k = \sum_{\ell} g_{\ell} x_{2k-\ell} .$$

tels que

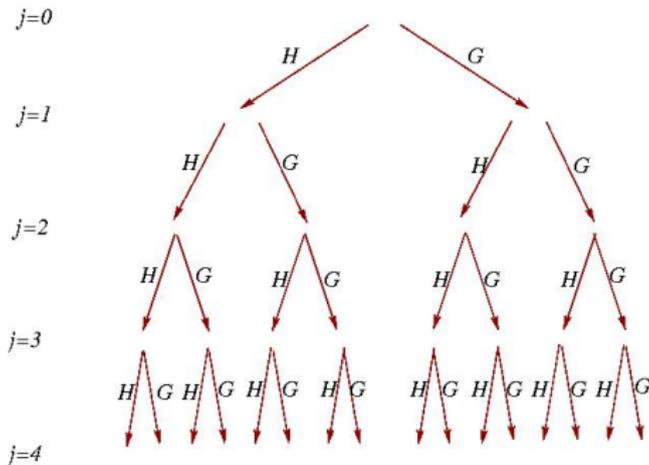
$$H^*H + G^*G = 1$$

ce que l'on schématise par le bloc suivant



La philosophie "best basis"

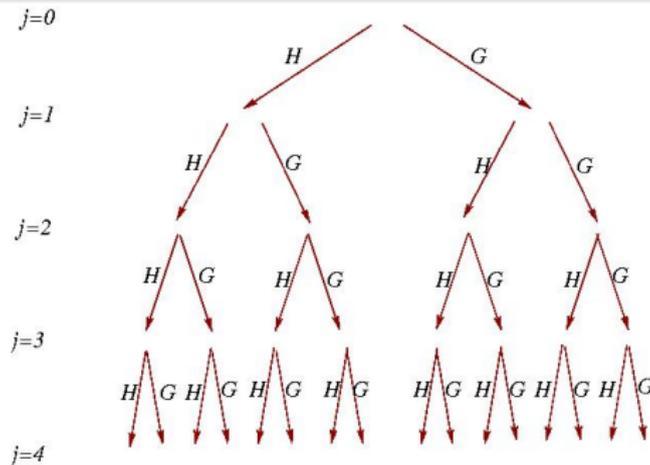
A chaque noeud de l'arbre correspond un sous-espace de l'espace de départ, équipé d'une base orthonormale. Les sous-espace sont en somme directe.



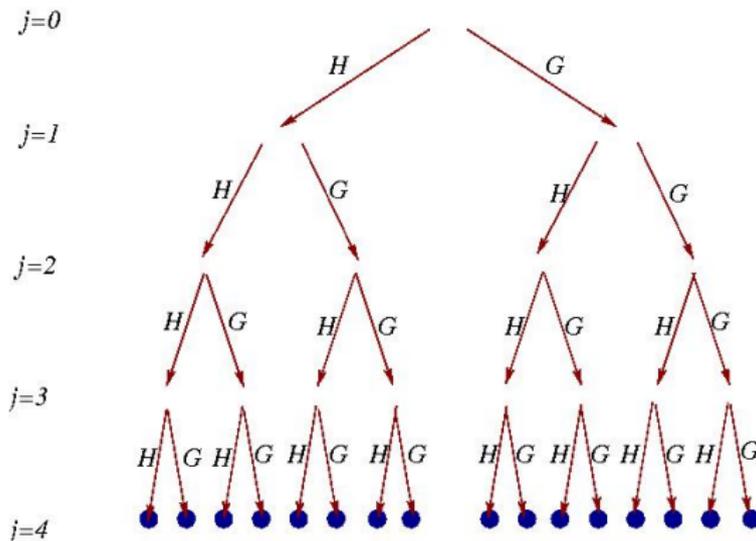
La philosophie "best basis"

Théorème

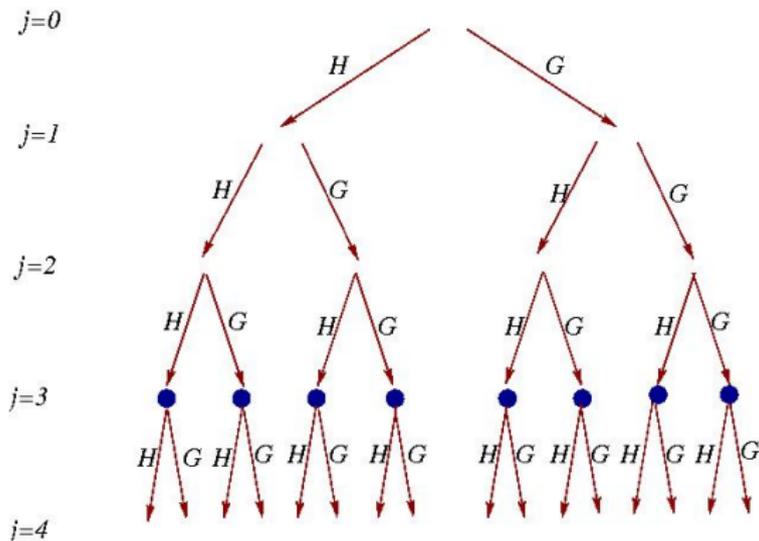
A l'ensemble des feuilles de chaque sous-arbre connexe correspond une base de $L^2(\mathbb{R})$ (ou \mathbb{C}^L).



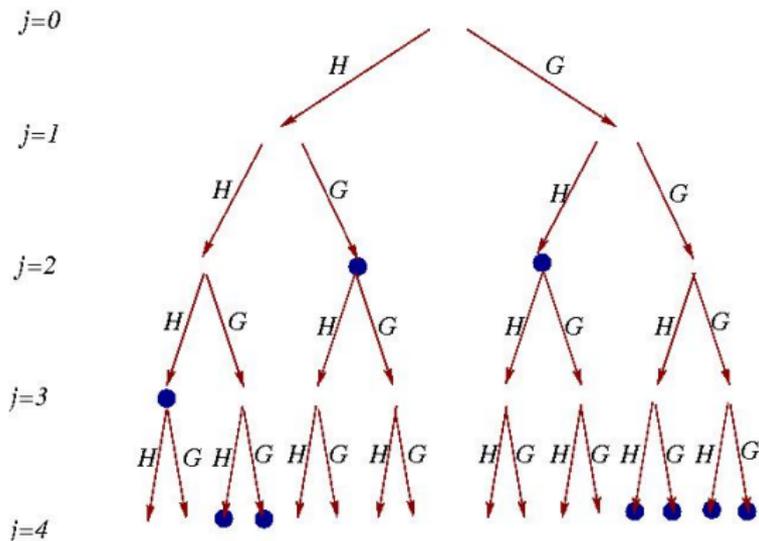
L'approche "best basis"



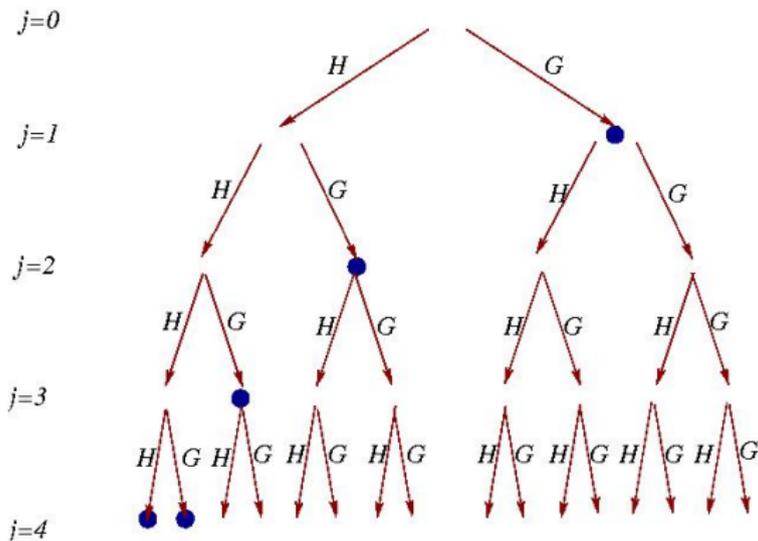
L'approche "best basis"



L'approche "best basis"



L'approche "best basis"



L'approche "best basis"

Algorithme:

- Calcul de toutes les décompositions possibles ($O(N \log N)$), sur toutes les bases de la bibliothèque.
- Partant des feuilles du bas: à chaque noeud, comparaison du critère sur le père et ses deux fils, sélection du meilleur.
- Remontée.

A la fin, à chacun des noeuds sélectionnés correspond un sous-espace et une base de celui-ci. L'union de ces bases donne la base optimale trouvée par l'algorithme.

Variante: famille hiérarchique de bases MDCT.

Approches variationnelles

Etant donné un repère $\{\phi_\lambda, \lambda \in \Lambda\}$, on cherche à exprimer $f \in \mathcal{H}$ comme combinaison linéaire sur ce repère.

1 Repère dual:

$$\min \|\alpha\|_2^2 \quad \text{sous contrainte} \quad f = \sum_{\lambda \in \Lambda} \alpha_\lambda \phi_\lambda$$

conduit à $\alpha_\lambda = \langle f, \tilde{\phi}_\lambda \rangle$, et ne génère pas de parcimonie.

2 Basis pursuit:

$$\min \|\alpha\|_1 \quad \text{sous contrainte} \quad f = \sum_{\lambda \in \Lambda} \alpha_\lambda \phi_\lambda$$

3 Basis pursuit denoising, ou LASSO:

$$\min_{\alpha} \left(\left\| f - \sum_{\lambda \in \Lambda} \alpha_\lambda \phi_\lambda \right\|_2^2 + \mu \|\alpha\|_1 \right)$$

Repère dual: pas de parcimonie

En introduisant un multiplicateur de Lagrange (vectoriel) g , le problème s'écrit

$$\min_{\alpha} \left(\sum_{\lambda} |\alpha_{\lambda}|^2 + \left\langle f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}, g \right\rangle \right)$$

L'optimisation conduit à

$$\alpha_{\lambda} = \langle g, \phi_{\lambda} \rangle ,$$

et g est déterminé par la contrainte:

$$g = \mathcal{R}^{-1} f ,$$

où $\mathcal{R} = U^* U$ est l'opérateur de repère: $\mathcal{R}f = \sum_{\lambda} \langle f, \phi_{\lambda} \rangle \phi_{\lambda}$ qui donne

$$\alpha_{\lambda} = \langle \mathcal{R}^{-1} f, \phi_{\lambda} \rangle = \langle f, \mathcal{R}^{-1} \phi_{\lambda} \rangle = \langle f, \tilde{\phi}_{\lambda} \rangle .$$

Basis pursuit

Pour générer de la parcimonie, on remplace la norme $\ell^2(\mathbb{Z})$ sur les coefficients par la norme $\ell^1(\mathbb{Z})$:

$$\min_{\alpha} \sum_{\lambda} |\alpha_{\lambda}| \quad \text{sous contrainte} \quad f = \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}$$

Le problème peut être résolu par des méthodes de type *point intérieur* (Chen, Donoho & Saunders, 1998)

Régularisation de Tikhonov: pas de parcimonie

Dans le cas “bruité”, il est nécessaire d’introduire un terme de régularisation, par exemple la norme $\ell^2(\mathbb{Z})$ de la suite de coefficients:

$$\min \left(\left\| f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda} \right\|^2 + \mu \|\alpha\|^2 \right),$$

Interprétation Bayésienne: régression dans un bruit blanc Gaussien, avec *a priori* Gaussien sur les coefficients

$$\rho(\alpha|f) = K\rho(f|\alpha)\rho(\alpha), \quad \rho(f|\alpha) \sim e^{-\|f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}\|^2}, \quad \rho(\alpha) \sim e^{-\mu\|\alpha\|^2}.$$

Régularisation de Tikhonov: pas de parcimonie

Dans le cas “bruité”, il est nécessaire d’introduire un terme de régularisation, par exemple la norme $\ell^2(\mathbb{Z})$ de la suite de coefficients:

$$\min \left(\left\| f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda} \right\|^2 + \mu \|\alpha\|^2 \right),$$

Interprétation Bayésienne: régression dans un bruit blanc Gaussien, avec *a priori* Gaussien sur les coefficients

$$\rho(\alpha|f) = K\rho(f|\alpha)\rho(\alpha), \quad \rho(f|\alpha) \sim e^{-\|f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}\|^2}, \quad \rho(\alpha) \sim e^{-\mu\|\alpha\|^2}.$$

Le problème s’écrit aussi comme

$$\min_{\alpha} \left(\|f\|^2 - 2\Re \left(\sum_{\lambda} \bar{\alpha}_{\lambda} \langle f, \phi_{\lambda} \rangle \right) + \sum_{\lambda, \lambda'} \alpha_{\lambda} \bar{\alpha}_{\lambda'} \langle \phi_{\lambda}, \phi_{\lambda'} \rangle + \mu \sum_{\lambda} |\alpha_{\lambda}|^2 \right)$$

Régularisation de Tikhonov: pas de parcimonie

Les équations normales conduisent à

$$\langle f, \phi_\lambda \rangle + \sum_{\lambda'} \alpha_{\lambda'} \langle \phi_{\lambda'}, \phi_\lambda \rangle + \mu \alpha_\lambda = 0$$

c'est à dire au problème matriciel

$$\langle f, \phi_\lambda \rangle = \sum_{\lambda'} \mathbf{G}_{\lambda\lambda'}^{(\mu)} \alpha_{\lambda'},$$

où on note \mathbf{G} la matrice de Gram du repère, i.e. $\mathbf{G}_{\lambda\lambda'} = \langle \phi_{\lambda'}, \phi_\lambda \rangle$, et

$$\mathbf{G}^{(\mu)} = \mathbf{G} + \mu \mathbf{I}$$

Remarque: Tout se simplifie évidemment lorsque le dictionnaire est une base orthogonale.

Parcimonie: LASSO et basis pursuit denoising

Pour générer la parcimonie, le terme de régularisation doit être remplacé par un terme générant de la parcimonie.

LASSO/basis pursuit denoising: utiliser la norme $\ell^1(\mathbb{Z})$.

$$\min \left(\frac{1}{2} \left\| f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda} \right\|^2 + \mu \|\alpha\|_1 \right),$$

Interprétation Bayésienne: régression dans un bruit blanc Gaussien, avec *a priori* Laplacien sur les coefficients

$$\rho(\alpha|f) = K\rho(f|\alpha)\rho(\alpha), \quad \rho(f|\alpha) \sim e^{-\|f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}\|^2/2}, \quad \rho(\alpha) \sim e^{-\mu\|\alpha\|_1}.$$

Parcimonie: LASSO et basis pursuit denoising

Optimisation: problème complexe.

- Point intérieur (Chen, Donoho & Saunders, 1998)
- Block Coordinate Relaxation (Sardy, Bruce & Tseng, 1998), basée sur la solution du problème lorsque le dictionnaire est une base orthonormale.
- Iterations de Landweber
- FOCUSS.
- ...

Parcimonie: LASSO et basis pursuit denoising

Si le dictionnaire est une base orthonormale la solution est simple: le problème s'écrit (dans le cas réel) comme

$$\min_{\alpha} \left(\|f\|^2 - \left(\sum_{\lambda} \alpha_{\lambda} \langle f, \phi_{\lambda} \rangle \right) + \sum_{\lambda} \alpha_{\lambda}^2 / 2 + \mu \sum_{\lambda} \alpha_{\lambda} \operatorname{sgn}(\alpha_{\lambda}) \right)$$

et se découple en sous-problèmes 1D

$$\min_{\alpha_{\lambda}} (\alpha_{\lambda} \langle f, \phi_{\lambda} \rangle + \alpha_{\lambda}^2 / 2 + \mu \alpha_{\lambda} \operatorname{sgn}(\alpha_{\lambda})) , \quad \forall \lambda$$

Parcimonie: LASSO et basis pursuit denoising

Si le dictionnaire est une base orthonormale la solution est simple: le problème s'écrit (dans le cas réel) comme

$$\min_{\alpha} \left(\|f\|^2 - \left(\sum_{\lambda} \alpha_{\lambda} \langle f, \phi_{\lambda} \rangle \right) + \sum_{\lambda} \alpha_{\lambda}^2 / 2 + \mu \sum_{\lambda} \alpha_{\lambda} \operatorname{sgn}(\alpha_{\lambda}) \right)$$

et se découple en sous-problèmes 1D

$$\min_{\alpha_{\lambda}} (\alpha_{\lambda} \langle f, \phi_{\lambda} \rangle + \alpha_{\lambda}^2 / 2 + \mu \alpha_{\lambda} \operatorname{sgn}(\alpha_{\lambda})) , \quad \forall \lambda$$

Solution: seuillage doux (*soft thresholding, shrinkage,...*)

$$\alpha_{\lambda} = \mathbb{S}_{\mu}(\langle f, \phi_{\lambda} \rangle) ,$$

avec

$$\mathbb{S}_{\mu}(x) = \operatorname{sgn}(x)(|x| - \mu)^+ .$$

Parcimonie: LASSO et basis pursuit denoising

Algorithme BCR: s'applique lorsque le dictionnaire $\phi_\lambda, \lambda \in \Lambda$ est une union de bases orthonormales,

$$\mathcal{D} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_M$$

et procède par estimation itérative des coefficients par rapport à l'une des bases, les autres étant fixés.

- 1 Initialisation des coefficients $\{\alpha_{m,k}, k = 0, L - 1, \dots, m = 1, \dots, M\}$
- 2 Sélection d'une des bases \mathcal{B}_m . Calcul du résidu
 $f^{(m)} = f - \sum_{m' \neq m} \sum_k \alpha_{km'} \phi_{km'}$
- 3 Résolution par seuillage doux de

$$\min_{\alpha_{mk}} \left(\frac{1}{2} \left\| f^{(m)} - \sum_k \alpha_{mk} \phi_{mk} \right\|^2 + \mu \|\alpha\|_1 \right)$$

- 4 Si la convergence n'est pas atteinte, retour à l'étape 2.

Parcimonie: LASSO et basis pursuit denoising

Choix des bases dans l'algorithme: plusieurs stratégies. La plus simple est la sélection cyclique: l'une après l'autre, en boucle.

Parcimonie: LASSO et basis pursuit denoising

Choix des bases dans l'algorithme: plusieurs stratégies. La plus simple est la sélection cyclique: l'une après l'autre, en boucle.

Théorème

L'algorithme BCR, associé à la sélection cyclique des bases, converge vers une solution du problème de basis pursuit denoising

Parcimonie: LASSO et basis pursuit denoising

Choix des bases dans l'algorithme: plusieurs stratégies. La plus simple est la sélection cyclique: l'une après l'autre, en boucle.

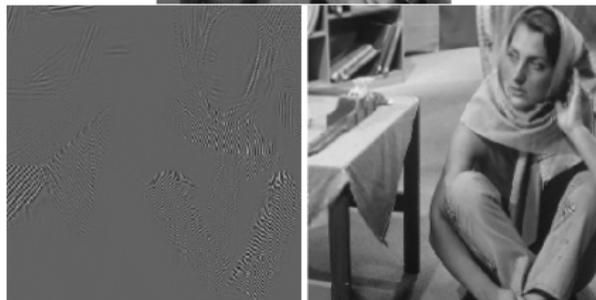
Théorème

L'algorithme BCR, associé à la sélection cyclique des bases, converge vers une solution du problème de basis pursuit denoising

Application: décompositions multicouches. Lorsque l'on cherche à décomposer un vecteur en sommes de "composantes", ou "couches", significativement différentes au sens où elles admettent une décomposition parcimonieuse dans des bases différentes.

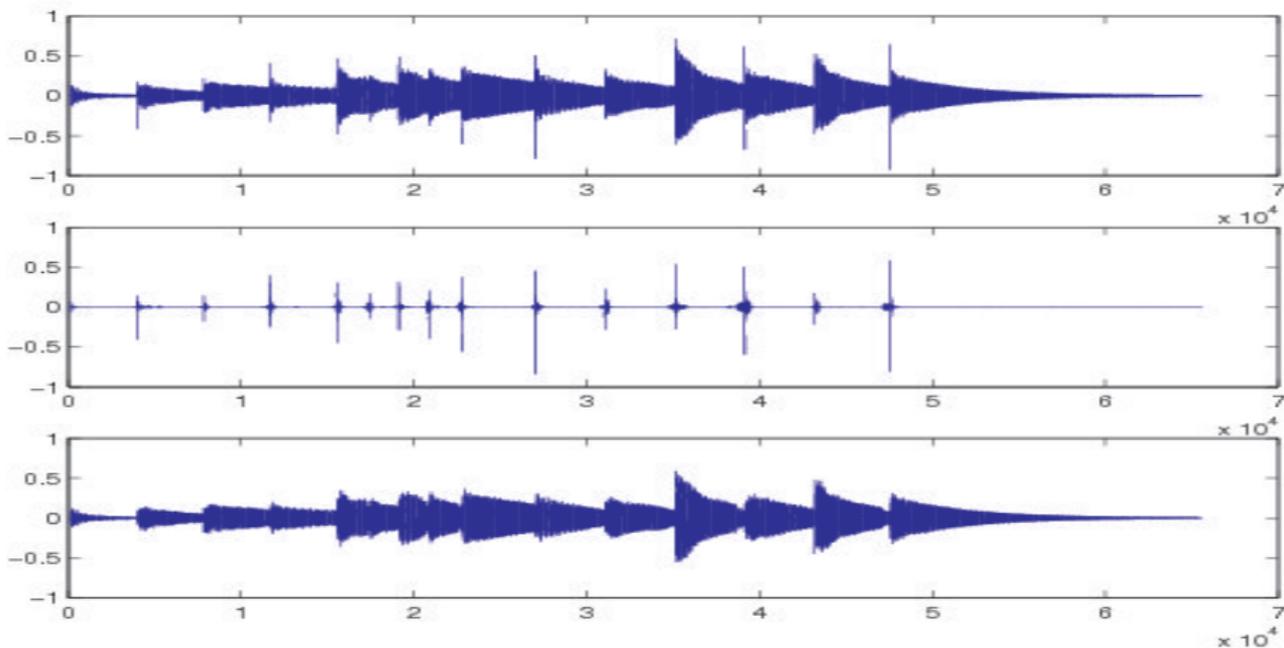
$$\min \left[\frac{1}{2} \left\| f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda} - \sum_{\rho} \beta_{\rho} \psi_{\rho} \right\|^2 + \mu \|\alpha\|_1 + \mu' \|\beta\|_1 \right]$$

Décompositions multicouches.



(D'après Elad, Starck, Querre et Donoho, 2005)

Décompositions multicouches.



(D'après Jaillet & BT, 2005)

Au delà de la parcimonie

Hypothèse sous-jacente à la majorité des méthodes parcimonieuses: dans \mathbb{R}^L , on cherche des approximations

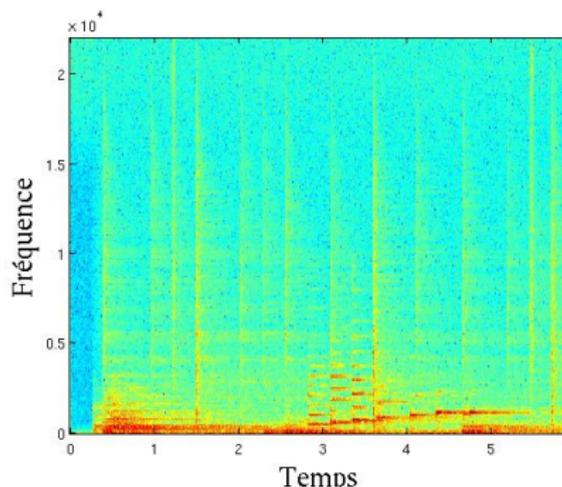
$$f \approx \sum_{\lambda \in \Lambda} \alpha_{\lambda} \phi_{\lambda}$$

où la **carte de signifiante** $\Lambda \subset \{1, \dots, L\}$ est telle que $|\Lambda| \ll L$, et les éléments de λ sont équiprobables.

Cette hypothèse est souvent peu réaliste en pratique.

Au delà de la parcimonie

Exemple avec des signaux audio:



exemple avec une carte Bernoulli

Au delà de la parcimonie

Deux pistes:

- Trouver des atomes plus réalistes (apprentissage de dictionnaire), ou grouper des atomes en molécules (R. Gribonval, L. Daudet,...).
- Introduire la modélisation dans le domaine des coefficients:
modélisation des cartes de signifiante: approches probabilistes et/ou variationnelles.

Au delà de la parcimonie

Approche variationnelle: divise la carte de coefficients en groupes $(\alpha_\lambda = \alpha_{g,m})$, et remplace la norme ℓ^1 (égalitaire) par une norme qui favorise certains comportements de groupe.

$$\min_{\alpha} \left[\|f - \sum_{\lambda} \alpha_{\lambda} \phi_{\lambda}\|^2 + \mu \|\alpha\|_{pq}^q \right]$$

où

$$\|\alpha\|_{pq}^q = \sum_g \left(\sum_m |\alpha_{g,m}|^p \right)^{q/p} \quad \text{ou} \quad \|\alpha\|_{pq}^q = \sum_m \left(\sum_g |\alpha_{g,m}|^p \right)^{q/p}$$

Au delà de la parcimonie

Par un bon choix de p et q , on peut ainsi favoriser, au choix, un comportement de groupe

- Solidaire: tous les coefficients d'un groupe s'allument comme un seul homme (group LASSO)
- Elitiste: chaque groupe n'allume que ses "champions" (SLOT).
- Variantes (cf exposé de M. Kowalski à CAP)

Conclusions

- On s'est focalisés sur une forme de parcimonie différente de la parcimonie exploitée dans les approches SVM, et des extensions.
- Cette notion produit des décompositions de signaux, images,... parcimonieuses en atomes, pas nécessairement en données.
- Pour lever cette difficulté: échantillonnage compressif (compressed sensing).
- à toi Albert !
- Merci à tous...

Quelques références

- I. Daubechies, *Ten lectures on wavelets*, NSF-CBMS lectures series.
- M. Vetterli, J. Kovacevic, *Wavelets and subband coding*, Prentice Hall.
- S. Mallat, *A wavelet tour of signal processing*, Academic Press.
- M.V. Wickerhauser, *Adapted waveform analysis, from theory to software*, CRC press
- Y. Meyer: *Ondelettes* (Dunod)
- D. Donoho, M. Vetterli, R. DeVore, I. Daubechies, Data compression and harmonic analysis, *IEEE Trans. Info. Theory* 44:6 (1998), 2435:2476

Quelques références (suite)

- S. S. Chen, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.
- A. G. Bruce, S. Sardy, and P. Tseng, "Block coordinate relaxation methods for nonparametric signal denoising," in *Proceedings of the SPIE - The International Society for Optical Engineering*, no. 3391, 1998, pp. 75–86.
- I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, *Communications on Pure and Applied Mathematics* 57 (11) (2004) 1413 – 1457.
- M. Elad, J.L. Starck, P. Querre and D.L. Donoho, *Simultaneous Cartoon and Texture Image inpainting using Morphological Component Analysis (MCA)*, Technical Report, March 2005.