

Note décomposition en ondelettes complexes

Leo Davy

May 6, 2022

1 Ondelettes complexes et Dual Tree Complex Wavelet Transform

1.1 Modèle standard

On se donne une base d'ondelettes $(\psi_{j,k})$ et on considère les coefficients d'ondelettes

$$d_j(x) = \langle X, \psi_{j,k_x} \rangle = X \star \psi_j(x) \quad (1)$$

d'un signal $x \mapsto X(x)$. Afin de gagner en invariance par translation (locale) (?) on s'intéresse plutôt aux (log-)leaders en prenant au lieu de $d_j(x)$ le supremum des valeurs absolues des coefficients de tous les coefficients dans le voisinage de k_x à toutes les échelles les plus fines. Pour alléger les notations, je confonds les notations pour les coefficients d'ondelettes et les leaders. On a ainsi

$$d_j(x) \sim v(x) 2^{jh(x)} \xi_x \stackrel{\mathbb{E}}{=} v(x) 2^{jh(x)} \quad (2)$$

avec ξ_x un bruit multiplicatif, centré en 1, log-normal si on veut un bruit gaussien sur les logs-leaders, qu'on ignore par la suite et on considère les égalités en espérance. En passant au log on est ramené à une régression linéaire. Par contre, l'analyse est isotrope parce que au début de l'analyse en dimension d , on a $2^d - 1$ ondelettes ψ , vu que par hypothèse le signal est isotrope, et localement homogène, afin de diminuer l'impact du bruit ξ , on considère chaque coefficient d'ondelette de façon uniforme (par rapport aux directions). On n'a donc pas d'espoir de récupérer des informations spécifiques aux directions.

1.2 Modèle complexe

A la place d'une ondelette réelle, on peut considérer une ondelette complexe pour faire l'analyse, de la forme :

$$\psi = \text{Re}(\psi) + i \text{Im}(\psi) = \psi_r + i \psi_c \quad (3)$$

où ψ_r et ψ_c sont des ondelettes réelles. En particulier, elles peuvent vérifier

$$d_j^{r,c}(x) = v_{r,c}(x) 2^{jh_{r,c}(x)} \quad (4)$$

où l'équation est à considérer soit pour r , soit pour c (par exemple, h_r désigne la régularité au sens de la quantité de multirésolution associée à l'analyse par l'ondelette réelle ψ_r). On se retrouve donc avec des coefficients complexes d'ondelette

$$d_j(x) = d_j^r(x) + id_j^c(x) = v_r(x)2^{jh_r(x)} + iv_c(x)2^{jh_c(x)}. \quad (5)$$

On peut alors réécrire l'expression précédente, en ignorant la dépendance en x , sous la forme

$$d_j = \sqrt{v_r^2 2^{2jh_r} + v_c^2 2^{2jh_c}} e^{i \arctan\left(\frac{v_c 2^{jh_c}}{v_r 2^{jh_r}}\right)} = \rho_j e^{i\omega_j} \quad (6)$$

avec

$$\rho_j(x) := \sqrt{v_r(x)^2 2^{2jh_r(x)} + v_c(x)^2 2^{2jh_c(x)}} \quad (7)$$

et

$$\omega_j(x) = \arctan\left(\frac{v_c(x)}{v_r(x)} 2^{j(h_c(x) - h_r(x))}\right). \quad (8)$$

Que l'on peut aussi réécrire sous la forme

$$d_j(x) = \rho_j(x) \cos(\omega_j(x)) + i\rho_j(x) \sin(\omega_j(x)). \quad (9)$$

En particulier, on peut remarquer

1. que le module (ρ_j) et la phase (ω_j) de d_j sont indépendants (l'expression de l'une des deux quantités ne fait pas intervenir l'autre).
2. $\log(\tan(\omega_j)) = \log(d_j^c/d_j^r)$ permet une régression linéaire de $\log \frac{v_c}{v_r}$ et $h_c - h_r$.
3. Comme d_j provient d'une analyse en ondelettes, on peut être aussi supposer que $\rho_j \sim v 2^{jh}$ et estimer $\log(v)$ et h aussi par régression linéaire de $\log(\rho_j)$
4. Aussi, il y a sûrement un lien entre les deux quantités précédentes qui pourrait permettre de coupler les estimations.

On peut donner une intuition au fait que ω mesure bien un angle dans la texture de la manière suivante. On suppose v_c/v_r constant et donc ω_j ne dépend que de $2^{j(h_c - h_r)}$. Dans un premier temps on suppose $h_c \gg h_r$, c'est à dire que la quantité multirésolution associée à l'ondelette ψ_c décroît plus vite que celle associée à ψ_r ; ou bien, plus simplement, X est plus régulier suivant l'ondelette c que l'ondelette r .

Donc quand on considère les échelles les plus fines ($j \rightarrow -\infty$) on a $\omega_j \rightarrow \arctan(0)$, d'où, aux échelles les plus fines, la phase disparaît et toute l'information est capturée par l'ondelette ψ_r . Cela correspond probablement à ce qu'on attend, car $h_c > h_r$ signifie qu'il y a plus d'information locale suivant ψ_r que ψ_c .

En considérant $h_c \ll h_r$ on peut appliquer le même raisonnement et obtenir que l'information est capturée seulement par l'ondelette h_c .

Il reste donc à s'assurer que ψ_r et ψ_c capturent bien de l'information dans des directions distinctes. Pour cela, la stratégie semble être de choisir des ondelettes ψ_r et ψ_c

qui sont obtenues par transformée de Fourier d'ondelettes ϕ_r et ϕ_c . C'est à dire, par exemple,

$$\psi_r(t) = \hat{\phi}_r(t). \quad (10)$$

Et ainsi, un signal anisotrope $X(t)$ qui a une densité spectrale de la forme $|\hat{X}(\zeta)| \leq \frac{1}{\|\zeta\|^{h(\zeta)}}$, où $\zeta \mapsto h(\zeta)$ est une fonction qui ne dépend que de l'angle, qui donne la régularité directionnelle (pour des directions spectrales).

Donc, si, par exemple si Ω est le support de ϕ_r , on a

$$d_{j=0}^r = \langle X, \psi_r \rangle = \langle X, \hat{\phi}_r \rangle = \langle \hat{X}, \phi_r \rangle \quad (11)$$

et donc d_0^r ne dépend que des directions (en Fourier) de X supportées dans Ω . Il reste à voir quelles propriétés on souhaite avoir entre ψ_r et ψ_c . On pourra par exemple chercher à avoir que ρ_j soit une quantité anisotrope, au sens où si X^θ est une rotation par θ d'un signal X , alors leur coefficients complexes d'ondelettes ont même module. Cela forcerait l'information sur l'anisotropie à être contenue dans ω_j (peut-être que c'est trop restrictif et que l'on perdrait toute l'information sur l'anisotropie).

1.2.1 Une version plus précise

Ci-dessus tout a été noté avec des coefficients d'ondelettes $d_{j,k}(X) = \langle X, \psi_{j,k} \rangle$, qui ne vérifient pas en général le modèle 4 contrairement aux leaders

$$l_{j,k} = \sup_{\lambda_{j',k'} \subset 3\lambda_{j,k}} 2^{j-j'} |d_{j',k'}| \quad (12)$$

où les $\lambda_{j,k}$ sont des cubes $[2^j k, 2^j(k+1))$ pour des signaux en une dimension (et la multiplication se fait par homothétie). En dimension d ,

$$l_{j,k} := l^{p=\infty}(j,k) = \left(\sum_{\lambda_{j',k'} \subset 3\lambda_{j,k}} 2^{-d(j-j')} \|d_{j,k}\|_{\ell^{p=\infty}(2^d-1)} \right) = \sup_{\lambda_{j',k'} \subset 3\lambda_{j,k}} \sup_i 2^{d(j-j')} |d_{j',k'}^{(i)}| \quad (13)$$

où on passe par la définition des p -leaders et $d_{j,k} = (d_{j,k}^{(i=1)}, \dots, d_{j,k}^{(i=2^d-1)})$ sont les coefficients d'ondelettes en dimension d .

Un problème qui apparait peut être avec cette définition est que l'on perd complètement l'information sur l'ondelette dont provient le leader choisi, alors que chaque ondelette i peut avoir des propriétés directionnelles particulières. Une solution est peut-être de considérer des leaders pour chaque i (mais l'un des problèmes est que l'on a un nombre de leaders exponentiel avec la dimension) :

$$l_{j,k}^{(i)} = \sup_{\lambda_{j',k'} \subset 3\lambda_{j,k}} 2^{d(j-j')} |d_{j',k'}^{(i)}|. \quad (14)$$

1.3 Choix des ondelettes complexes

On s'intéresse maintenant au choix des ondelettes complexes sous la forme 3:

$$\psi = \psi_r + i\psi_c. \quad (15)$$

Les premières relations que l'on souhaite vérifier sont les relations d'échelles (écrites correctement avec les leaders cette fois-ci):

$$l_{j,k}^{r,c} = \sup_{\lambda_{j',k'} \subset 3\lambda_{j,k}} 2^{d(j-j')} |d_{j',k'}^{r,c}| \sim v_{j,k} 2^{jh_{j,k}^{r,c}} \quad (16)$$

ainsi que

$$l_{j,k} = \sup_{\lambda_{j',k'} \subset 3\lambda_{j,k}} 2^{d(j-j')} |\langle X, \psi_{j',k'} \rangle| = \sup 2^{d(j-j')} \sqrt{(d_{j',k'}^r)^2 + (d_{j',k'}^c)^2} \sim v_{j,k} 2^{jh_{j,k}} \quad (17)$$

qui sont vérifiées dès que ψ_r et ψ_c ont suffisamment de moments nuls (et de régularité). On souhaite aussi faire une analyse locale, donc ψ_r et ψ_c doivent être *localisées* (support compact).

Une autre propriété que l'on veut probablement est l'*invariance par translation locale* pour autant de coefficients d'ondelettes que possible.

On peut aussi remarquer que les leaders complexes 17 "classiques" ne correspondent pas directement aux coefficients que l'on peut utiliser pour le modèle de régression écrit au début 6:

$$\tilde{l}_{j,k} = |\sqrt{(l_{j,k}^r)^2 + (l_{j,k}^c)^2} e^{i \arctan\left(\frac{l_{j,k}^r}{l_{j,k}^c}\right)}|. \quad (18)$$

Il serait appréciable que les deux quantités coïncident.

Aussi, on s'intéresse à l'orientation, on voudrait donc que les ondelettes $\psi^{(i)}$ aient des propriétés directionnelles.

Une solution partielle à ces problèmes¹, inspirés de la transformée de Fourier pour les signaux réels consiste à prendre

$$\psi = \psi_r + i\psi_c \quad (19)$$

avec ψ_r paire, et ψ_i impaire, toutes deux réelles. De plus, si ψ_r et ψ_c sont déphasées de $\pi/2$ alors ψ est un signal analytique. Par contre, dans ce cas là, ψ ne peut pas être à support compact.

1.4 Dual Tree Complex Wavelet Transform (DTCWT)

Les DTCWT semblent être une option afin de contourner ce problème analyticit /support fini en utilisant une représentation (en ondelettes) redondante. L'id e est de choisir ψ approximativement analytique, ou de fa on  quivalente, telle que ψ_c soit approximativement la transform e de Hilbert de ψ_r .

¹cf r f sur Dual Tree Complex Wavelet, partie "Trouble in paradise : four problems with real wavelets"

On utilise donc une paire d'ondelettes réelles, et on calcule la transformée en ondelettes pour chacune de ces ondelettes, l'une donne la partie réelle de l'ondelette complexe, et l'autre la partie imaginaire de l'ondelette complexe.

On peut calculer ces coefficients à partir de deux banques de filtres en parallèle, avec r_0, r_1 (resp c_0, c_1) les filtres passe-bas et passe-haut associés à l'ondelette réelle (resp. complexe).

Une condition simple pour avoir la propriété de quasi-analyticité dans le cas discret est que les filtres passe bas soient décalés d'un demi échantillon

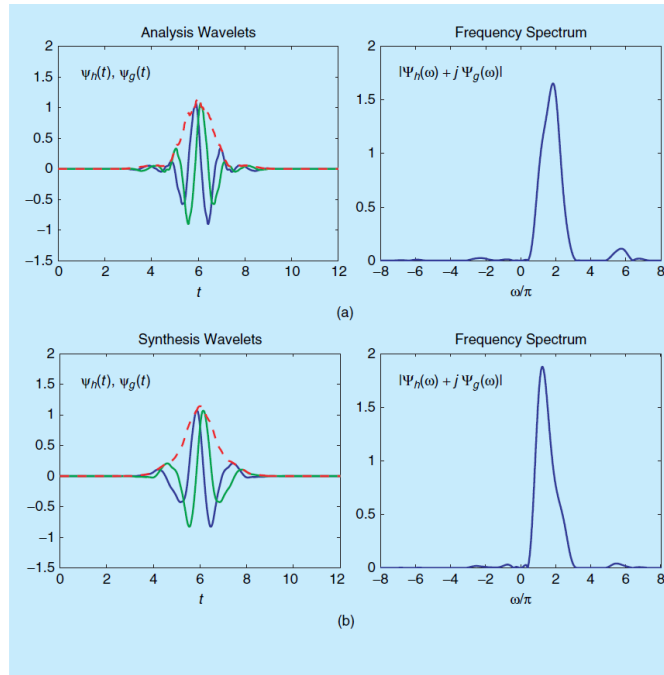
$$c_0(k) = r_0(k - 0.5). \quad (20)$$

Il semble donc que l'on cherche une paire de filtres à support fini vérifiant cette condition et dont les ondelettes associées ont suffisamment de moments nuls (et de régularité).

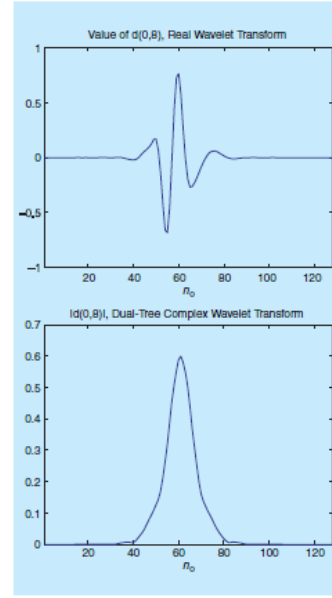
En dimension 2, on calcule 3 coefficients d'ondelette complexe par point, les parties réelles et imaginaires de chacun de ces coefficients est calculé à partir de deux ondelettes réelles séparables (mais aucune des six ondelettes n'est séparable), donc le calcul reste rapide. Il est aussi possible d'augmenter la sensibilité par rapport à l'orientation, en utilisant la Real Oriented 2D Dual Tree Transform, qui donne 6 coefficients complexes par point.

Quelques références

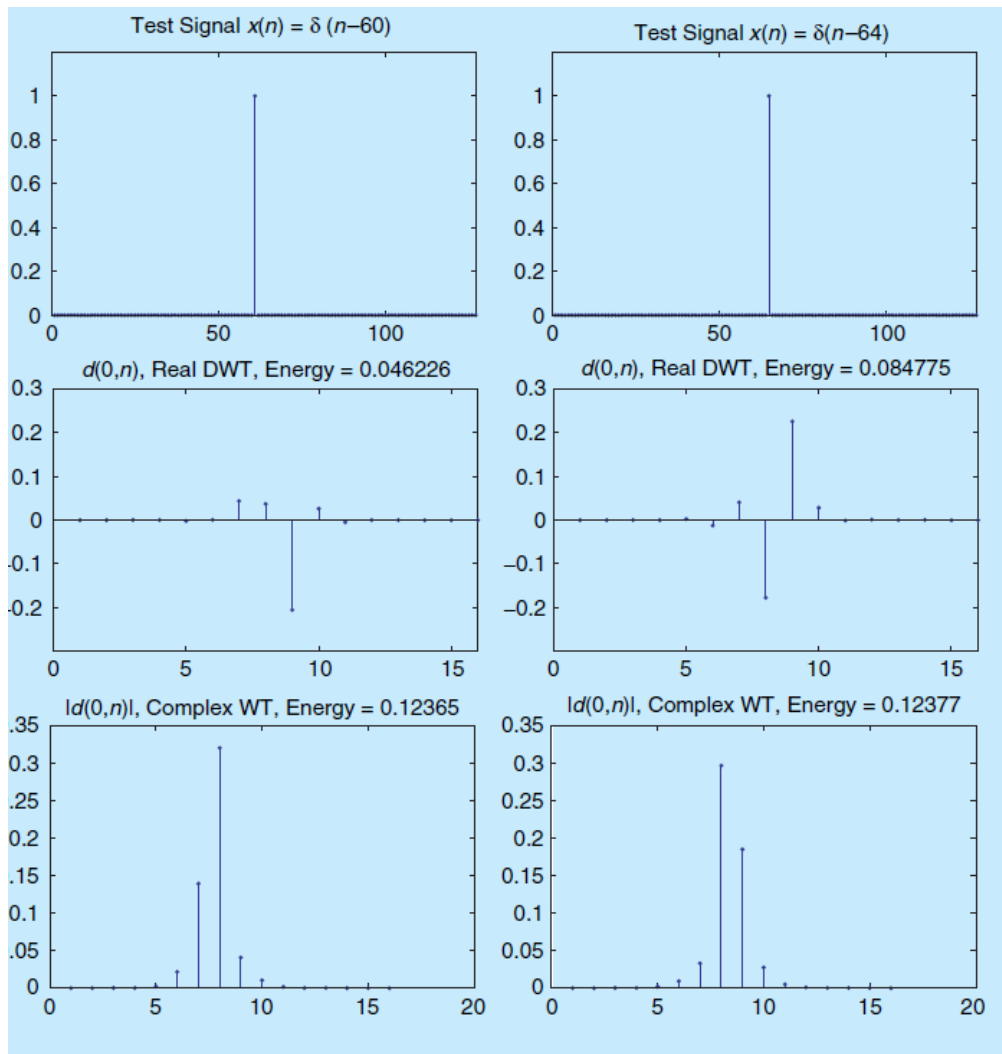
1. Sur Dual Tree Complex Wavelet Transform <https://ieeexplore.ieee.org/document/1550194>

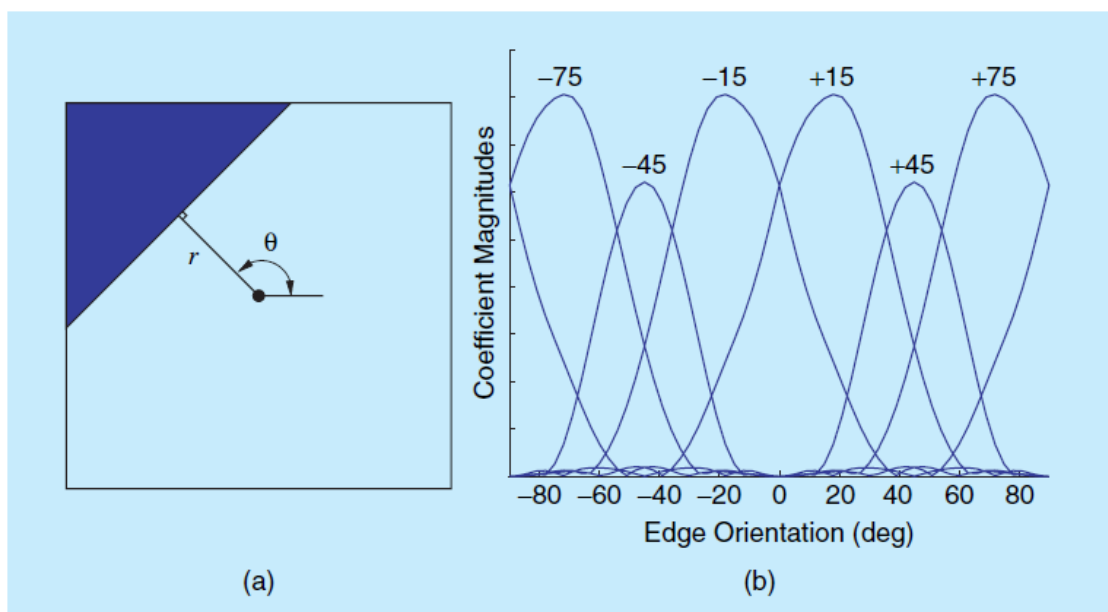


[FIG10] Common factor complex wavelet corresponding to a set of biorthogonal dual-tree filters [88].

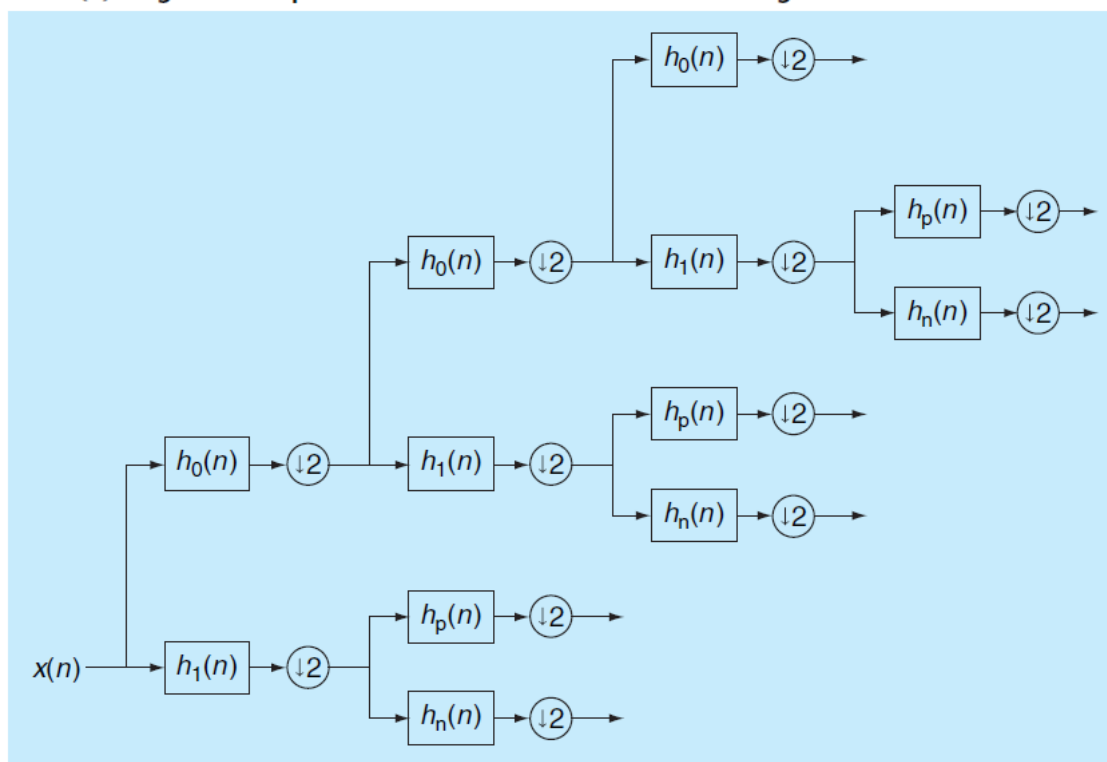


[FIG1] In the neighborhood of an edge, the real DWT produces both large and small wavelet coefficients. In contrast, the

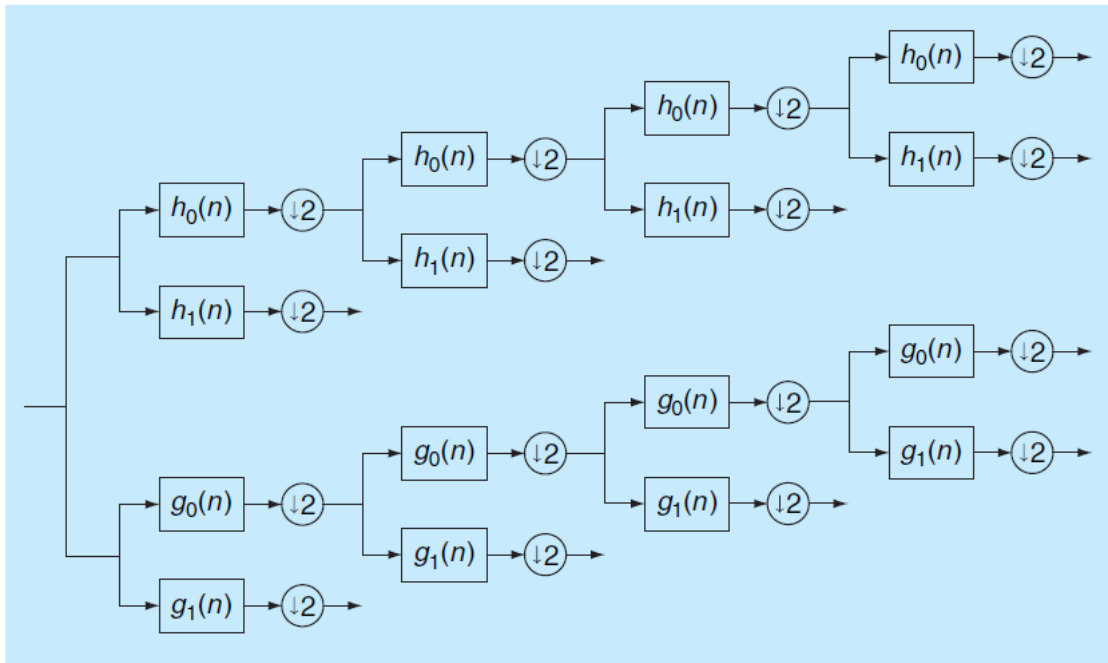




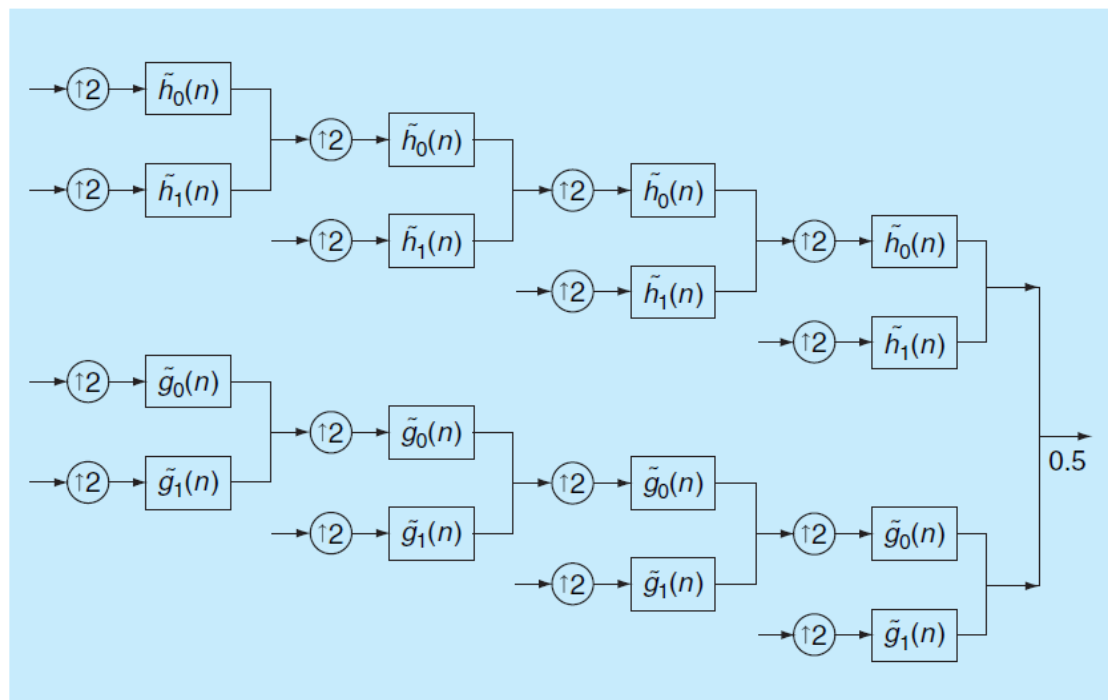
[FIG21] (a) Image segment with an edge singularity at orientation θ and offset from center r . (b) Magnitude responses of the CWT coefficients of this segment as a function of θ .



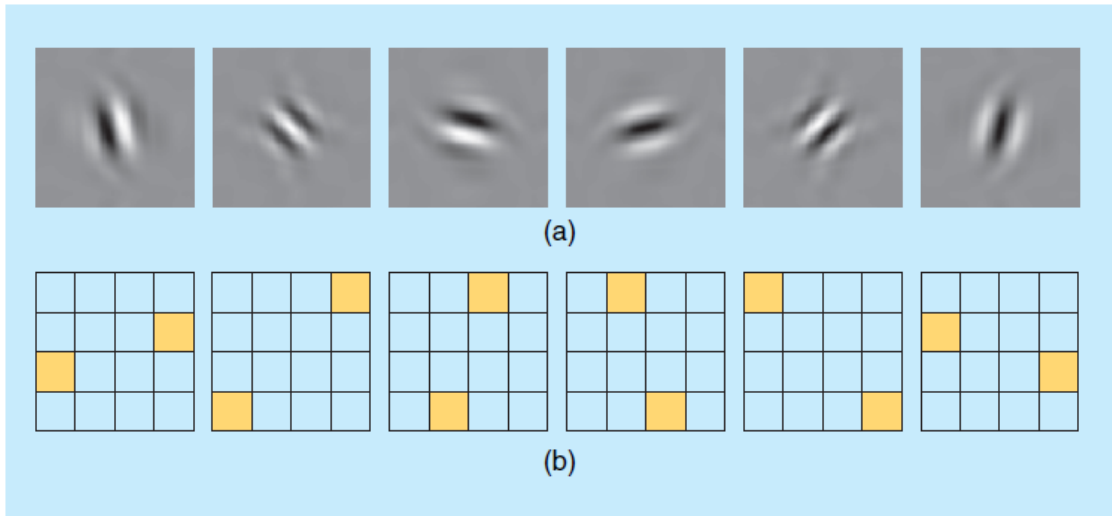
[FIG5] Analysis FB for the DWT with invertible complex post-filtering.



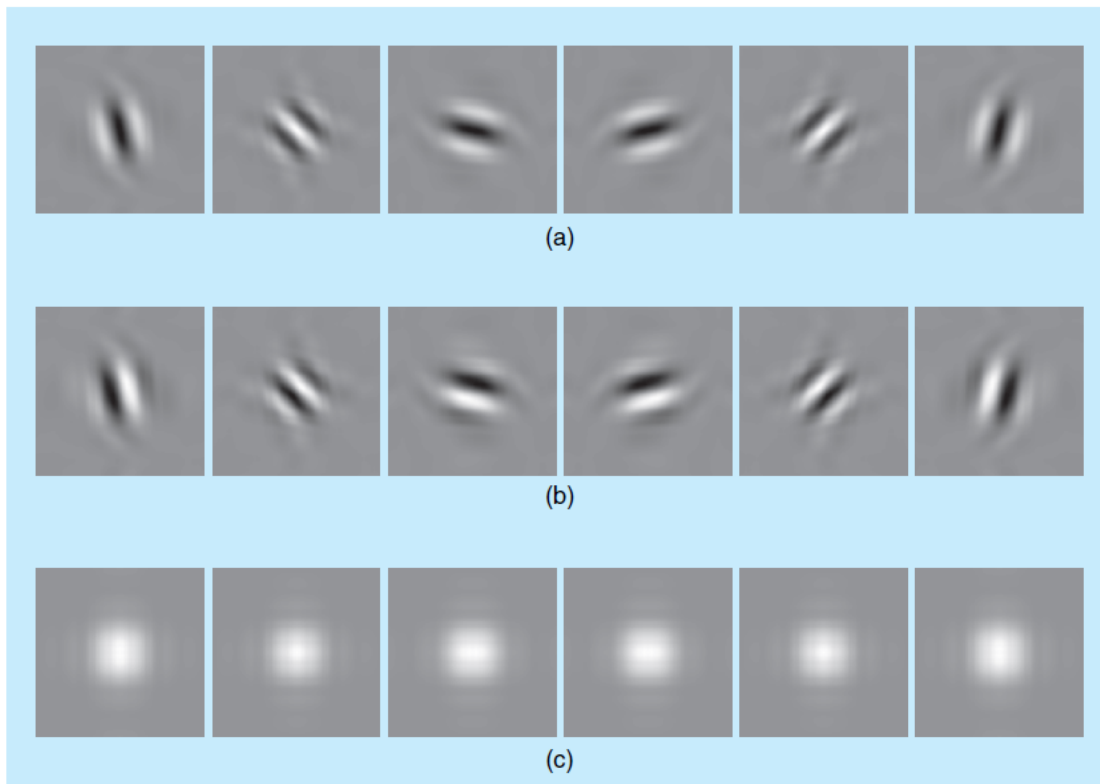
[FIG7] Analysis FB for the dual-tree discrete CWT.



[FIG8] Synthesis FB for the dual-tree CWT.



[FIG15] Typical wavelets associated with the real oriented 2-D dual-tree wavelet transform. (a) illustrates the wavelets in the space domain; (b) illustrates the (idealized) support of the Fourier spectrum of each wavelet in the 2-D frequency plane. The absence of the checkerboard phenomenon is observed in both the space and frequency domains.



[FIG16] Typical wavelets associated with the oriented 2-D dual-tree CWT. (a) illustrates the real part of each complex wavelet; (b) illustrates the imaginary part; and (c) illustrates the magnitude.

2 Sketching et K -means

2.1 K -means pour segmenter

Ici on voit le problème d'un point de vue très statistique. On a pour données les (logs)-leaders $(l_{j,k})$, et par hypothèse, sur chaque texture² $\{(j,k) \in \Omega_t\} = \Omega_t$, pour $t = 0, \dots, K$ on a

$$l_{j,k} \sim v_t + jh_t + \epsilon \iff (j,k) \in \Omega_t \quad (21)$$

où ϵ est un bruit (gaussien pour faciliter les calculs et l'identification), qui pourrait dépendre de la texture t .

On peut reformuler ce problème en K -means, avec $K = T$ le nombre de textures, donc on cherche T paires de points $(v_t, h_t)_{t=1}^T$ qui minimisent

$$\sum_{(j,k)} \min_{t=1, \dots, T} \|l_{j,k} - v_t - jh_t\|. \quad (22)$$

Une version alternative, pourrait être d'ignorer la relation d'échelle, et de chercher TJ points $(s_{j,t})$ (où J est le nombre d'échelles) en résolvant un k -means par échelle. L'avantage est qu'on a alors un problème k -means plus classique, et ensuite on peut faire une régression affine pour récupérer la relation d'échelle entre les coefficients³. Je pense qu'une reformulation un peu plus agréable de l'équation pourrait tout de même être faite.

Une fois que l'on a obtenu les T -means, on a un moyen d'identifier chaque point à une texture, en définissant la segmentation $S : \{image\} \rightarrow \{textures\}$ par

$$S : x = (J, k_x) \mapsto \hat{t}_x := \arg \min_t \|l_{J,k_x} - v_t - Jh_t\| \quad (23)$$

ou une formule analogue (avec k_x un indice de position aussi proche que possible de x à l'échelle J). Je considère⁴ ici que l'information la plus fine est à l'échelle J et c'est la plus précise. Une version pondérée à travers les échelles pourrait peut être donner de meilleurs résultats.

2.2 Sketching

Moralement on cherche à appliquer les idées du Compressed Sensing mais au lieu de faire des problèmes inverses sur des vecteurs, on le fait sur des distributions linéaires.

En compressed sensing, on a un opérateur linéaire de mesure A (incohérent/aléatoire), que l'on applique à un signal (inconnu) x , qui nous donne une mesure y . Donc on a

$$y = Ax \quad (24)$$

²On désigne une texture par t , on peut se restreindre ici à $t \in \{0, 1\}$, et on identifie les éléments de chaque texture à leur position multi-échelle.

³C'est peut être pas évident à faire parce que il faut réussir à mettre en correspondance les indices de textures t entre les différentes échelles. Par contre je pense que c'est faisable dès que le nombre d'échelle est non négligeable $J > 3$.

⁴Et c'est certainement ni vrai, ni optimal.

et on cherche à retrouver x . Sous des hypothèses de parcimonie (dans une base) de x , alors on peut reconstruire (si on connaît la base) x . La reconstruction est garantie pour un problème compliqué (décomposition atomique), et à nouveau avec une hypothèse de parcimonie, le problème devient un problème pénalisé (en norme 1) qui est "facile".

En sketching, on cherche à reconstruire la distribution \mathbb{P}_X d'une variable aléatoire X , et on se donne un opérateur \mathcal{A} linéaire par rapport aux densités de probabilité⁵. On mesure alors le sketch, idéalement,

$$z = \mathcal{A}(X) = \mathbb{E}(\Phi(X)) \quad (25)$$

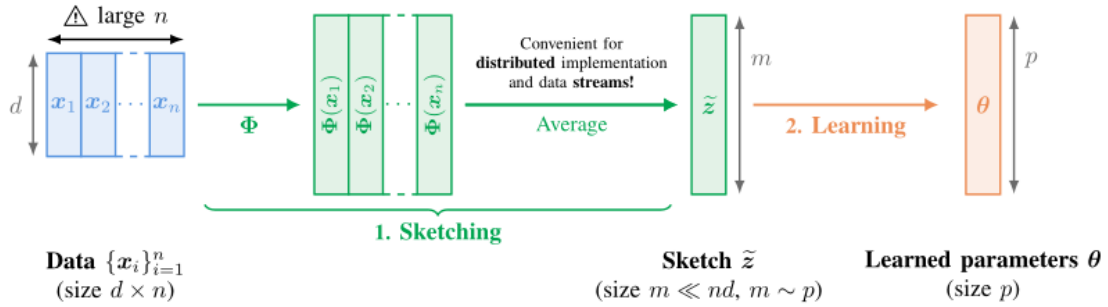
avec Φ une fonction de la variable aléatoire X . En pratique, on n'a pas accès à X , mais seulement à un échantillon $\mathcal{X} = \{x_1, \dots, x_N\}$, et on définit donc

$$\tilde{z} = \mathcal{A}(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^N \Phi(x_i). \quad (26)$$

L'intérêt est que l'on a

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \rightarrow_{w.p.1} \mathcal{A}(X) \quad (27)$$

par la loi forte des grands nombres.



Un avantage majeur, est que l'on est libre sur le choix de la fonction Φ et l'équation précédente reste vraie. Par exemple en prenant $\Phi : x \mapsto x^k$, $k \geq 0$, on a que le sketch de nos données converge vers le k -ième moment de la variable aléatoire (centrée) X . Pour cela, on peut parler de moment généralisé pour la fonction Φ , ou le sketch, mais en fait on peut aller au delà des monomes.

Afin de retrouver quelle variable aléatoire a généré les données, on va chercher, parmi

⁵C'est à dire que pour deux variables aléatoires X_1, X_2 et α_1, α_2 des réels positifs et tels que $\alpha_1 + \alpha_2 = 1$, on peut définir la somme des densités de probabilité comme $d\mathbb{P} = \alpha_1 d\mathbb{P}_{X_1} + \alpha_2 d\mathbb{P}_{X_2}$ (qui est bien une densité de probabilité). Alors \mathcal{A} est linéaire si pour toutes variables aléatoires et coefficients comme précédemment, on a $\mathcal{A}(d\mathbb{P}) = \mathcal{A}(\alpha_1 d\mathbb{P}_1 + \alpha_2 d\mathbb{P}_2) = \alpha_1 \mathcal{A}(d\mathbb{P}_1) + \alpha_2 \mathcal{A}(d\mathbb{P}_2)$.

On peut remarquer que pour n'importe quelle fonction f , on a que $\mathbb{E} \circ f$ définit une application linéaire en ce sens.

une classe de (densité de) variables aléatoires, la quelle a le sketch le plus proche du sketch des données, de type

$$\arg \min_{d\mathbb{P}_X} \|\Phi(\mathcal{X}) - \mathbb{E}_{d\mathbb{P}_X}(\Phi)\|. \quad (28)$$

On peut par exemple faire de la PCA (linéaire) en prenant un moment généralisé $\Phi : x \mapsto x^T x$, le sketch donne la matrice de covariance empirique. Afin de retrouver les composantes principales, il suffit de calculer la décomposition en valeurs singulière du sketch. On peut aussi, avec d'autres Φ résoudre des problèmes type Least squares, Gaussian Mixture Modeling, K -means (cf références). Quelque chose d'important à remarquer est que le sketch ne donne généralement pas la solution du problème qui est en train d'être résolu. Mais par contre, une fois qu'on a un sketch de nos données, on peut calculer le problème inverse posé.

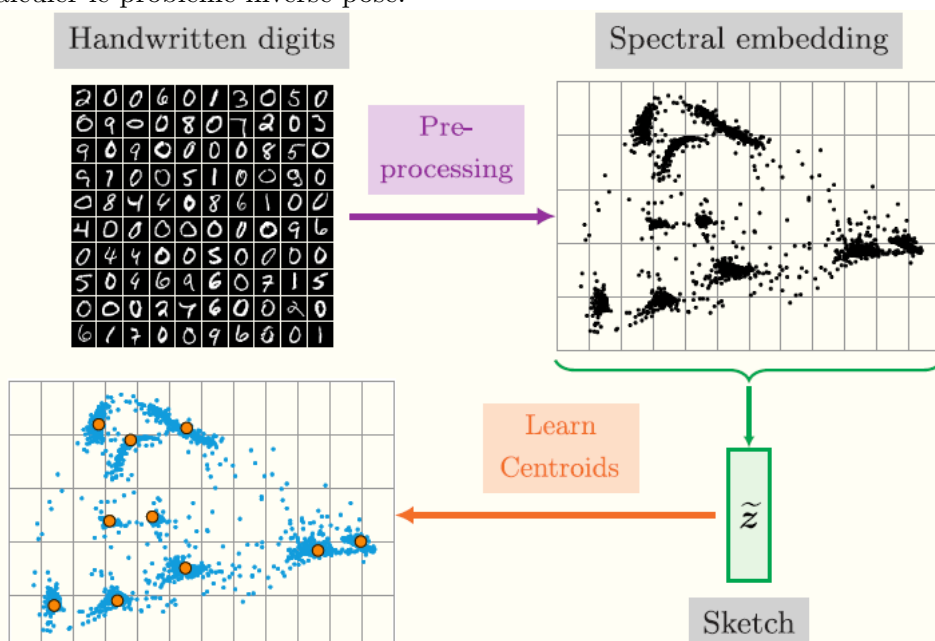


Fig. 5: Clustering of handwritten digits via compressive k -means. First, SIFT descriptors are extracted from each image. Then a similarity graph is constructed to obtain a so-called *spectral embedding* of the dataset (using the first eigenvectors of the Laplacian of the graph). The spectral features are aggregated into a sketch, from which centroids are extracted.

Un autre avantage est que la taille du sketch ne dépend pas du nombre de données à disposition. Ainsi, augmenter le nombre de données a seulement un coût pour calculer le sketch (ce qui est généralement très facile), mais par contre le coût de résolution du problème inverse reste constant (et devient même de mieux en mieux posé).

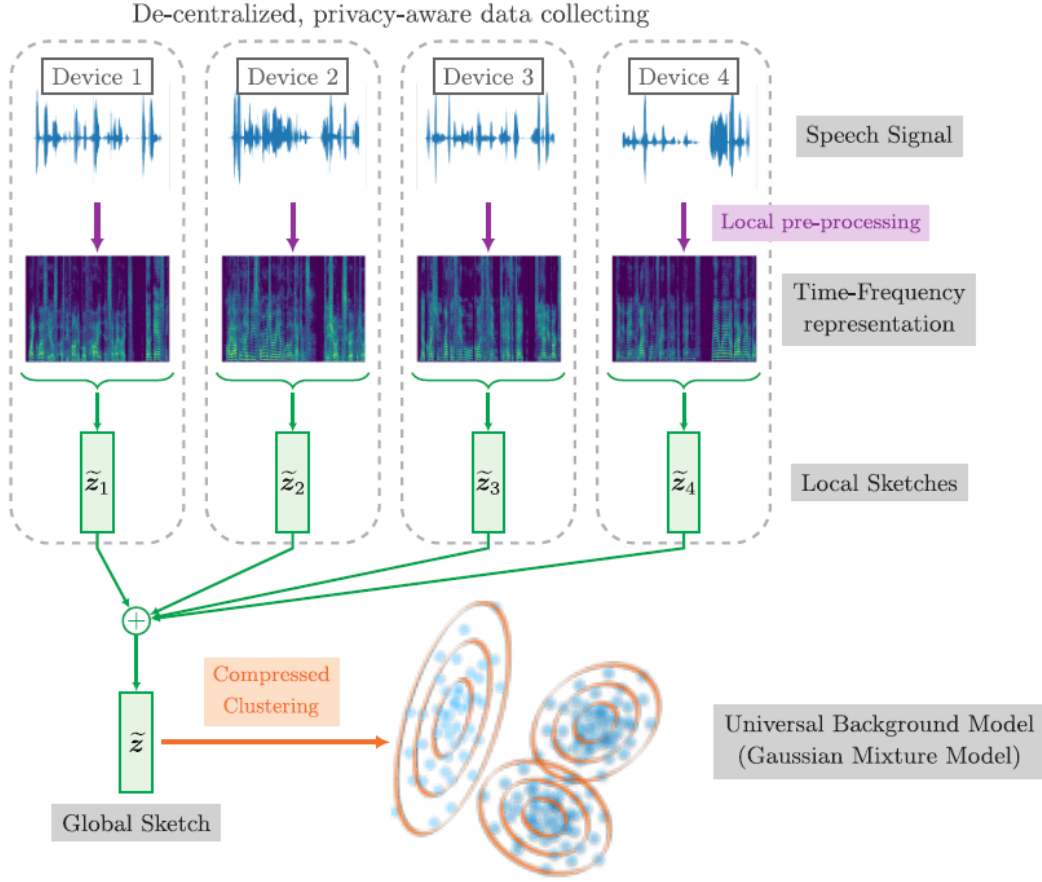


Fig. 3: Speaker verification via compressive learning. Unlabeled training speech data is collected in a decentralized manner, pre-processed, and locally sketched. The local sketches are then merged into a global sketch, from which a Universal Background Model is learned.

Pour résoudre K -means, différents choix de sketch peuvent être utilisés, par exemple une binning map (qui correspond à un histogramme sur une partition de l'ensemble image de la variable aléatoire). Une autre alternative consiste à regarder les moments de Fourier aléatoire. En fait dans ce cadre, ça correspond juste à calculer la fonction génératrice de la distribution empirique, en passant à la limite on trouve bien avec proba 1 la fonction génératrice de la vraie distribution en tout point. On peut donc calculer le sketch en un certain nombre de fréquences, et ensuite on cherche à trouver une distribution qui a les mêmes valeurs en chaque fréquence afin de retrouver la distribution initiale. Ainsi, pour résoudre K -means, on cherche à trouver les K (paires de) points (v_t, h_t) , qui minimisent

$$\|\tilde{z} - \sum_t \alpha_t \Phi(v_t, h_t)\| \quad (29)$$

où les α_t sont tous positifs, et de somme égale à 1. Moralement, α_t correspond au poids de la texture t dans l'image d'origine, et $\Phi(v_t, h_t)$ correspond à la "signature" (une sorte d'échantillon spectral) d'une texture t . On verra plus bas comment expliciter un tel

sketch.

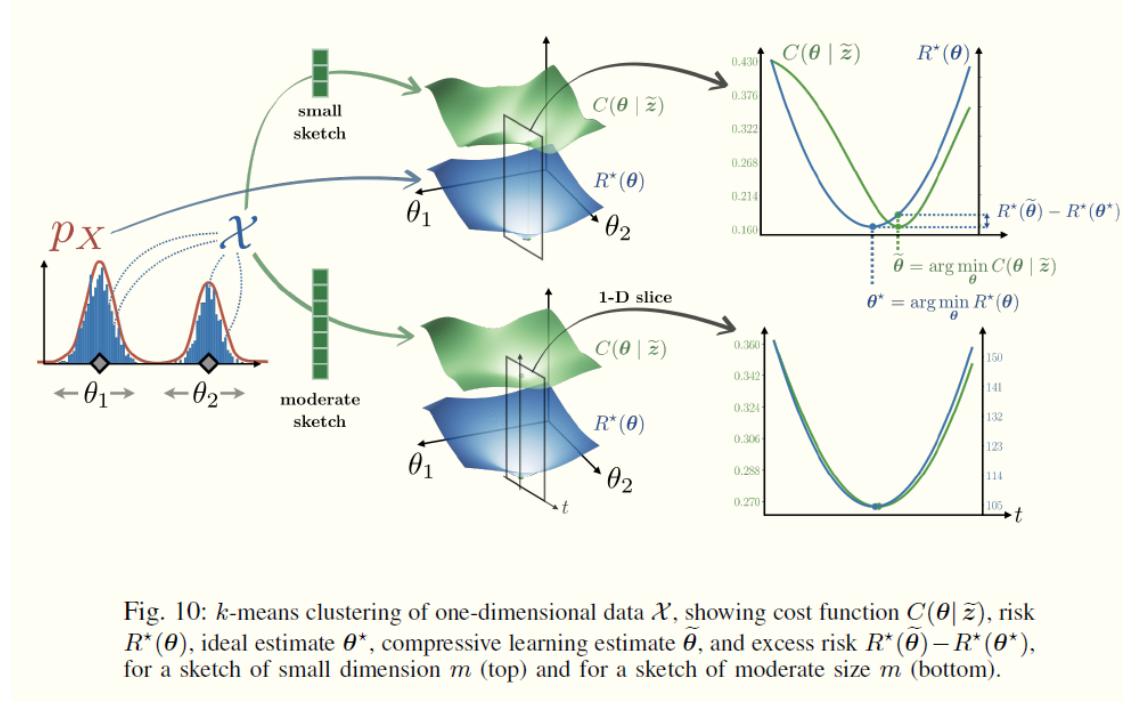


Fig. 10: k -means clustering of one-dimensional data \mathcal{X} , showing cost function $C(\theta | \tilde{z})$, risk $R^*(\theta)$, ideal estimate θ^* , compressive learning estimate $\hat{\theta}$, and excess risk $R^*(\hat{\theta}) - R^*(\theta^*)$, for a sketch of small dimension m (top) and for a sketch of moderate size m (bottom).

2.3 Sketching, K -means et segmentation

On considère les logs-leaders $l = (l_j)_j = ((l_{j,k})_k)_j$ à travers les échelles et les positions (chacun de ces coefficients correspondant à un échantillon suivant une certaine loi). Dans un modèle T -means (sous réserve qu'il soit correct), à chaque échelle j , alors il existe $u_j \in \{\hat{l}_1, \dots, \hat{l}_T\}^{N_j}$ qui minimise $\|l_j - u\|$, parmi tous les choix possibles de u qui ne prennent que T valeurs distinctes (avec N_j le nombre total de coefficients de position à l'échelle j).

On peut par exemple considérer un sketch basé sur du Fourier aléatoire en prenant une suite de fréquence $\omega_1, \dots, \omega_m$ (des nombres réels) et on pose

$$\Phi : \mathbb{R} \rightarrow \mathbb{C}^m \quad (30)$$

$$x \mapsto (e^{-i\omega_s x})_{s=1}^m. \quad (31)$$

Ce qui nous donne par exemple le sketch des logs-leaders à l'échelle j

$$\tilde{z}_j = \frac{1}{N_j} \sum_{k=1}^{N_j} \Phi(l_{j,k}). \quad (32)$$

En résumé

1. On considère que chaque leader d'une image est un échantillon d'une variable aléatoire paramétrée uniquement par la texture à laquelle le point appartient, et qui vérifie une relation d'échelle

2. On a donc, à chaque échelle, T distributions de probabilité que l'on cherche à estimer, et entre les échelles il existe une relation simple qui lie des distributions à travers les échelles
3. On a donc accès à la distribution empirique des leaders, vue comme échantillon d'une somme de T -distributions
4. On suppose que les T -means de notre distribution permettent d'identifier les textures
5. On calcule le sketch pris sur toute l'image
6. A partir du sketch on résoud un problème inverse pour retrouver les T -means (et possiblement leur relation inter-échelle)

2.3.1 Estimation échelle par échelle

Pour l'instant on va considérer le problème à une échelle fixée, c'est à dire que l'on va d'abord chercher à estimer les centroides des logs-leaders de chaque texture et pas les coefficients de variance et régularité. Pour cela on peut chercher à trouver les α_t, \hat{l}_t (avec $\sum_t \alpha_t = 1$) minimisant

$$\|\tilde{z}_j - \sum_t \alpha_t \Phi(\hat{l}_t)\|. \quad (33)$$

Si on résoud ce problème à toutes les échelles, on obtient au final J vecteurs de taille T et afin d'identifier les textures, on peut dire que J leaders représentent la même texture (chacun choisi à une échelle différente des autres parmi les leaders obtenus par T -means), si il existe une⁶ droite qui passe parmi ces points. Cependant ce problème n'est peut-être pas tout à fait évident, on peut donc considérer le problème à travers les échelles directement.

2.3.2 Estimation jointe à travers les échelles

On va ici définir un sketch global à travers les échelles en utilisant la concatenation des sketches \tilde{z}_j , ce qui nous donne

$$\Phi_J : \mathbb{R}^J \rightarrow \mathbb{C}^{Jm} \quad (34)$$

$$(x_1, \dots, x_J) \mapsto (\Phi(x_1), \dots, \Phi(x_J)). \quad (35)$$

On a donc le sketch correspondant $\tilde{z} = (\tilde{z}_1, \dots, \tilde{z}_J)$. En utilisant l'hypothèse sur la relation d'échelle, on peut transformer le sketch global en un sketch qui ne dépend que de deux paramètres (h, v) :

$$\tilde{\Phi} : \mathbb{R}_+^2 \rightarrow \mathbb{C}^{Jm} \quad (36)$$

$$(h, v) \mapsto \Phi_J(v + j_1 h, v + j_2 h, \dots, v + Jh). \quad (37)$$

⁶En raison de la relation multi-échelle que l'on ne capture pas directement, plusieurs droites peuvent passer par un même point, par contre deux droites se croisent au plus une fois.(...)

On peut alors poser le problème d'optimisation sous la forme 29. Commençons par expliciter $\tilde{\Phi}$ à une fréquence ω ,

$$\tilde{\Phi}(h, v)(\omega) = (e^{-i\omega(v+j_1h)}, \dots, e^{-i\omega(v+Jh)}) = e^{-i\omega v} (e^{-i\omega j_1h}, \dots, e^{-i\omega Jh}) \quad (38)$$

et ainsi le problème d'optimisation porte donc sur

$$\|(\frac{1}{N_1} \sum_k \Phi(l_{j_1,k}), \dots, \frac{1}{N_J} \sum_k \Phi(l_{J,k})) - \sum_t \alpha_t \tilde{\Phi}(h_t, v_t)\|_{\mathbb{R}^{mJ}} \quad (39)$$

où on minimise sur tous les α_t, h_t, v_t admissibles⁷ et m est le nombre de fréquences que l'on considère. Un problème équivalent est

$$\sum_j \|\frac{1}{N_j} \sum_k \Phi(l_{j,k}) - \sum_t \alpha_t \Phi(jh_t + v_t)\|_{\mathbb{R}^m}^2 = \sum_j \sum_s^m |\tilde{z}_j(\omega_s) - \sum_t \alpha_t e^{-i(v_t+jh_t)\omega_s}|^2 \quad (40)$$

où l'on minimise sur les mêmes paramètres.

2.3.3 Segmentation

Après résolution on a donc $(\alpha_t^*, h_t^*, v_t^*)_t$ qui minimise l'un des problèmes ci-dessus. Maintenant, il reste à faire la segmentation, c'est à dire assigner à un point (ou un leader) "ses" coefficients h_t^*, v_t^* . On peut penser à plusieurs méthodes, par exemple choisir la texture t qui minimise l'erreur quadratique à travers les échelles entre la relation affine $v_t^* + jh_t^*$, ou bien la même chose à travers les sketch pris seulement dans un voisinage de la position considérée. C'est à dire,

$$\hat{t}_x = \arg \min_t \sum_j (l_{j,k_{j,x}} - v_t^* - jh_t^*)^2 \quad (41)$$

ou bien

$$\hat{t}_x = \arg \min_t \|\Phi_J(l_{j_1,k_{j_1,x}}, \dots, l_{J,k_{J,x}}) - \tilde{\Phi}(h_t^*, v_t^*)\|_{\mathbb{R}^{Jm}} \quad (42)$$

(je crois qu'il ne faut pas mettre le α_t^* ci-dessus).

Quelques références :

1. Référence principale : <https://arxiv.org/abs/2008.01839>
2. Résolution de K -means par sketching : <https://hal.inria.fr/hal-01386077v3/document>
3. Quelques slides : https://github.com/DavyL/sketching_slides/blob/main/main.pdf

⁷Tous positifs, $h_t < 1$ et $\sum_{t=1}^T \alpha_t = 1$.