

# Note décomposition en ondelettes complexes

Leo Davy

April 28, 2022

## 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

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

avec  $\xi_x$  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  $\psi$ , vu que par hypothèse le signal est isotrope, et localement homogène, afin de diminuer l'impact du bruit  $\xi$ , 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.

## 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  $\psi_r$  et  $\psi_c$  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  $\psi_r$ ). 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 ( $\rho_j$ ) et la phase ( $\omega_j$ ) 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  $\omega$  mesure bien un angle dans la texture de la manière suivante. On suppose  $v_c/v_r$  constant et donc  $\omega_j$  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  $\psi_c$  décroît plus vite que celle associée à  $\psi_r$ ; 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  $\psi_r$ . Cela correspond probablement à ce qu'on attend, car  $h_c > h_r$  signifie qu'il y a plus d'information locale suivant  $\psi_r$  que  $\psi_c$ .

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  $\psi_r$  et  $\psi_c$  capturent bien de l'information dans des directions distinctes. Pour cela, la stratégie semble être de choisir des ondelettes  $\psi_r$  et  $\psi_c$

qui sont obtenues par transformée de Fourier d'ondelettes  $\phi_r$  et  $\phi_c$ . 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  $\Omega$  est le support de  $\phi_r$ , 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  $\Omega$ . Il reste à voir quelles propriétés on souhaite avoir entre  $\psi_r$  et  $\psi_c$ . On pourra par exemple chercher à avoir que  $\rho_j$  soit une quantité anisotrope, au sens où si  $X^\theta$  est une rotation par  $\theta$  d'un signal  $X$ , alors leur coefficients complexes d'ondelettes ont même module. Cela forcerait l'information sur l'anisotropie à être contenue dans  $\omega_j$  (peut-être que c'est trop restrictif et que l'on perdrait toute l'information sur l'anisotropie).

Quelques références

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

## 3 Sketching et $K$ -means

### 3.1 $K$ -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 texture<sup>1</sup>  $\{(j,k) \mid \Omega_t\} = \Omega_t$ , pour  $t = 0, \dots, K$  on a

$$l_{j,k} \sim v_t + jh_t + \epsilon \iff (j,k) \in \Omega_t \quad (12)$$

où  $\epsilon$  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

$$\sum_{(j,k)} \min_{t=1, \dots, T} \|l_{j,k} - v_t - jh_t\|. \quad (13)$$

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

---

<sup>1</sup>On 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.

faire une régression affine pour récupérer la relation d'échelle entre les coefficients<sup>2</sup>. 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

$$S : x = (J, k_x) \mapsto \hat{t}_x := \arg \min_t \|l_{J,k_x} - v_t - Jh_t\| \quad (14)$$

ou une formule analogue (avec  $k_x$  un indice de position aussi proche que possible de  $x$  à l'échelle  $J$ ). Je considère<sup>3</sup> 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.

## 3.2 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

$$y = Ax \quad (15)$$

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 probabilité<sup>4</sup>. On mesure alors le sketch, idéalement,

$$z = \mathcal{A}(X) = \mathbb{E}(\Phi(X)) \quad (16)$$

avec  $\Phi$  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

$$\tilde{z} = \mathcal{A}(\mathcal{X}) = \frac{1}{N} \sum_{i=1}^N \Phi(x_i). \quad (17)$$

---

<sup>2</sup>C'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  $J > 3$ .

<sup>3</sup>Et c'est certainement ni vrai, ni optimal.

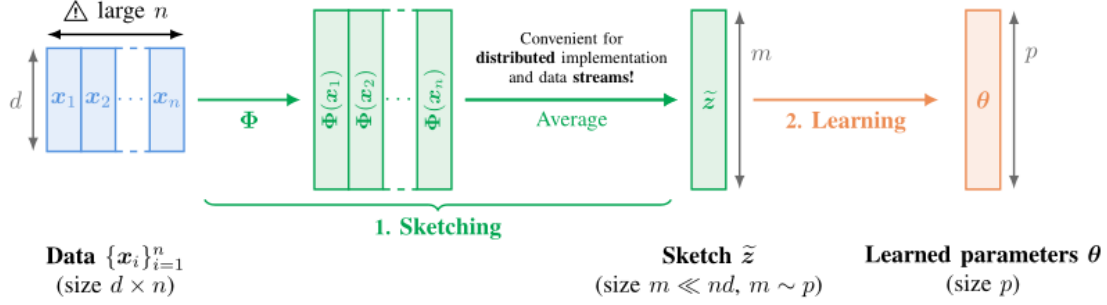
<sup>4</sup>C'est à dire que pour deux variables aléatoires  $X_1, X_2$  et  $\alpha_1, \alpha_2$  des réels positifs et tels que  $\alpha_1 + \alpha_2 = 1$ , on peut définir la somme des densités de probabilité comme  $d\mathbb{P} = \alpha_1 d\mathbb{P}_{X_1} + \alpha_2 d\mathbb{P}_{X_2}$  (qui est 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.

L'intérêt est que l'on a

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N \Phi(x_i) \rightarrow_{w.p.1} \mathcal{A}(X) \quad (18)$$

par la loi forte des grands nombres.



Un avantage majeur, est que l'on est libre sur le choix de la fonction  $\Phi$  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  $\Phi$ , 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 une classe de (densité de) variables aléatoires, la quelle a le sketch le plus proche du sketch des données, de type

$$\arg \min_{d\mathbb{P}_X} \|\Phi(\mathcal{X}) - \mathbb{E}_{d\mathbb{P}_X}(\Phi)\|. \quad (19)$$

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  $\Phi$  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.

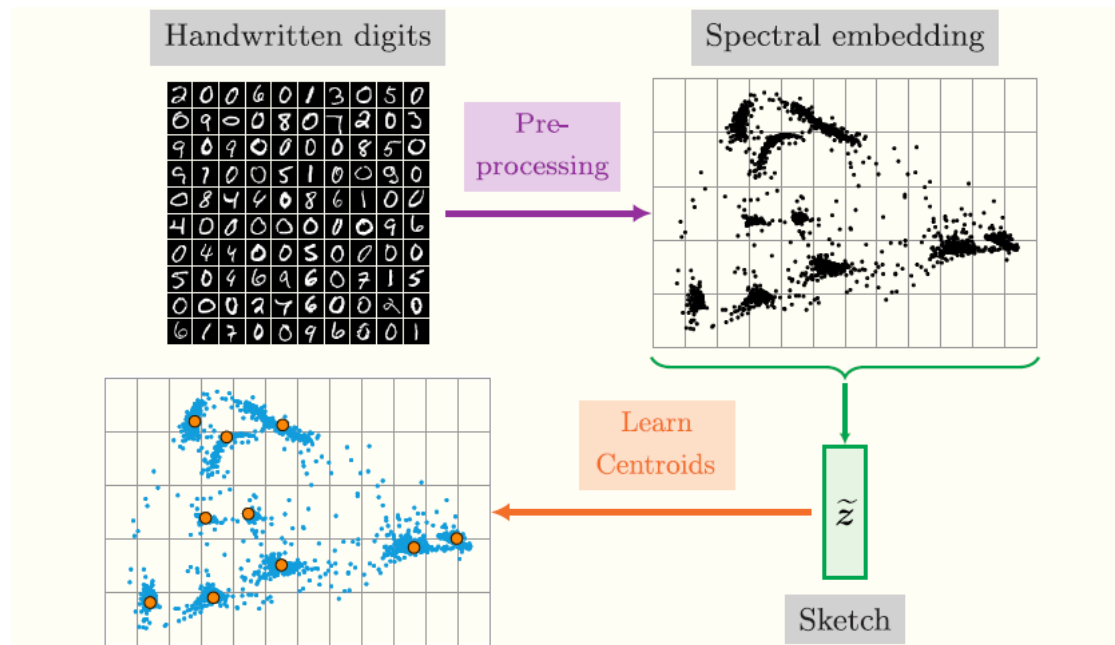


Fig. 5: Clustering of handwritten digits via compressive  $k$ -means. First, SIFT descriptors are extracted from each image. Then a similarity graph is constructed to obtain a so-called *spectral embedding* of the dataset (using the first eigenvectors of the Laplacian of the graph). The spectral features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

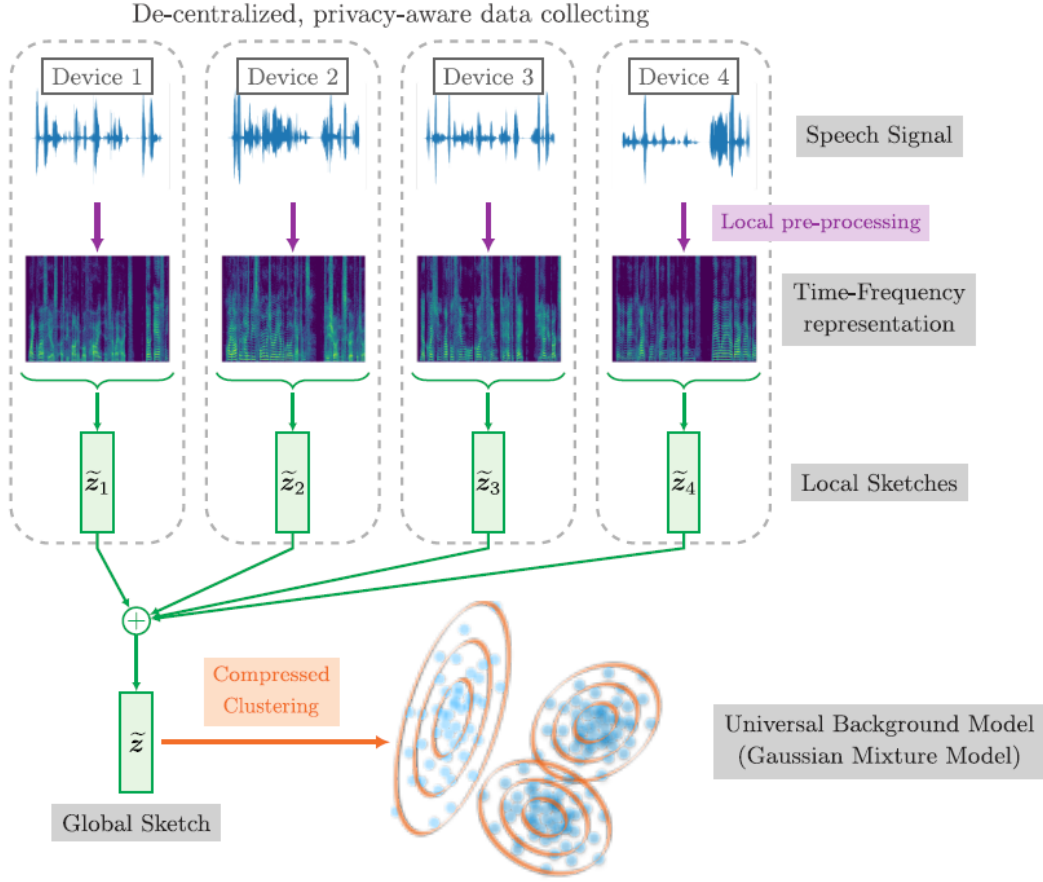


Fig. 3: Speaker verification via compressive learning. Unlabeled training speech data is collected in a decentralized manner, pre-processed, and locally sketched. The local sketches are then merged into a global sketch, from which a Universal Background Model is learned.

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

$$||\tilde{z} - \sum_t \alpha_t \Phi(v_t, h_t)|| \quad (20)$$

où les  $\alpha_t$  sont tous positifs, et de somme égale à 1. Moralement,  $\alpha_t$  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.

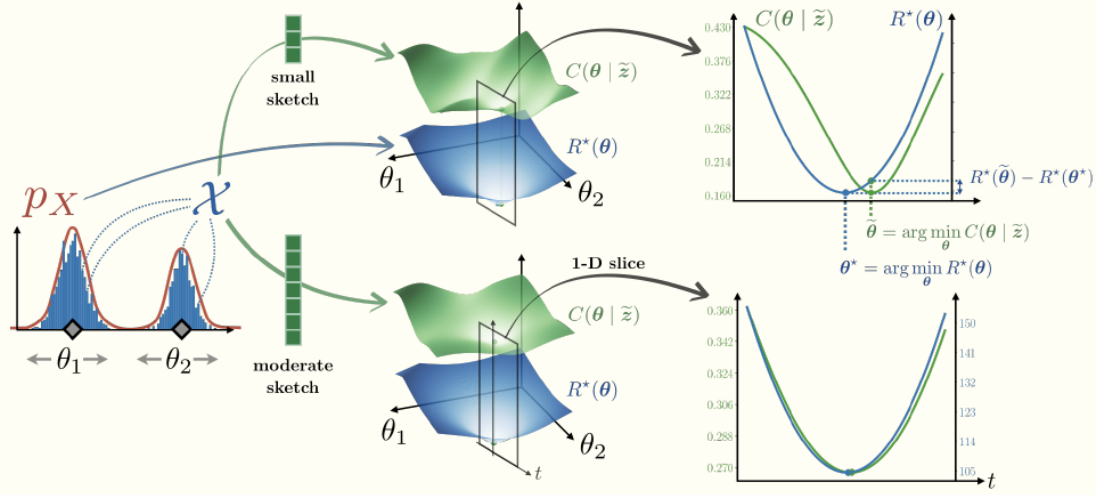


Fig. 10:  $k$ -means clustering of one-dimensional data  $\mathcal{X}$ , showing cost function  $C(\theta | \tilde{z})$ , risk  $R^*(\theta)$ , ideal estimate  $\theta^*$ , compressive learning estimate  $\hat{\theta}$ , and excess risk  $R^*(\hat{\theta}) - R^*(\theta^*)$ , for a sketch of small dimension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

### 3.3 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

$$\Phi : \mathbb{R} \rightarrow \mathbb{C}^m \quad (21)$$

$$x \mapsto (e^{-i\omega_s x})_{s=1}^m. \quad (22)$$

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

$$\tilde{z}_j = \frac{1}{N_j} \sum_{k=1}^{N_j} \Phi(l_{j,k}). \quad (23)$$

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)

### 3.3.1 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  $\alpha_t, \hat{l}_t$  (avec  $\sum_t \alpha_t = 1$ ) minimisant

$$\|\tilde{z}_j - \sum_t \alpha_t \Phi(\hat{l}_t)\|. \quad (24)$$

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 une<sup>5</sup> 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.

### 3.3.2 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

$$\Phi_J : \mathbb{R}^J \rightarrow \mathbb{C}^{Jm} \quad (25)$$

$$(x_1, \dots, x_J) \mapsto (\Phi(x_1), \dots, \Phi(x_J)). \quad (26)$$

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  $(h, v)$ :

$$\tilde{\Phi} : \mathbb{R}_+^2 \rightarrow \mathbb{C}^{Jm} \quad (27)$$

$$(h, v) \mapsto \Phi_J(v + j_1 h, v + j_2 h, \dots, v + Jh). \quad (28)$$

---

<sup>5</sup>En 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 20. Commençons par expliciter  $\tilde{\Phi}$  à une fréquence  $\omega$ ,

$$\tilde{\Phi}(h, v)(\omega) = (e^{-i\omega(v+j_1h)}, \dots, e^{-i\omega(v+Jh)}) = e^{-i\omega v} (e^{-i\omega j_1h}, \dots, e^{-i\omega Jh}) \quad (29)$$

et ainsi le problème d'optimisation porte donc sur

$$\|(\frac{1}{N_1} \sum_k \Phi(l_{j_1,k}), \dots, \frac{1}{N_J} \sum_k \Phi(l_{J,k})) - \sum_t \alpha_t \tilde{\Phi}(h_t, v_t)\|_{\mathbb{R}^{mJ}} \quad (30)$$

où on minimise sur tous les  $\alpha_t, h_t, v_t$  admissibles<sup>6</sup> et  $m$  est le nombre de fréquences que l'on considère. Un problème équivalent est

$$\sum_j \|\frac{1}{N_j} \sum_k \Phi(l_{j,k}) - \sum_t \alpha_t \Phi(jh_t + v_t)\|_{\mathbb{R}^m}^2 = \sum_j \sum_s^m |\tilde{z}_j(\omega_s) - \sum_t \alpha_t e^{-i(v_t+jh_t)\omega_s}|^2 \quad (31)$$

où l'on minimise sur les mêmes paramètres.

### 3.3.3 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,

$$\hat{t}_x = \arg \min_t \sum_j (l_{j,k_{j,x}} - v_t^* - jh_t^*)^2 \quad (32)$$

ou bien

$$\hat{t}_x = \arg \min_t \|\Phi_J(l_{j_1,k_{j_1,x}}, \dots, l_{J,k_{J,x}}) - \tilde{\Phi}(h_t^*, v_t^*)\|_{\mathbb{R}^{Jm}} \quad (33)$$

(je crois qu'il ne faut pas mettre le  $\alpha_t^*$  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](https://github.com/DavyL/sketching_slides/blob/main/main.pdf)

---

<sup>6</sup>Tous positifs,  $h_t < 1$  et  $\sum_{t=1}^T \alpha_t = 1$ .