

Une approche bayésienne non paramétrique pour l'apprentissage d'un dictionnaire de taille adaptative

Hong-Phuong DANG, Pierre CHAINAIS

Ecole Centrale Lille, CRISAL CNRS UMR 9189
CS 20048, 59651 Villeneuve d'Ascq, France

hong-phuong.dang@ec-lille.fr, pierre.chainais@ec-lille.fr

Résumé – L'apprentissage de dictionnaire pour la représentation parcimonieuse est maintenant bien connu dans le cadre de la résolution de problèmes inverses en traitement d'image. Généralement, le nombre d'atomes du dictionnaire est fixé à l'avance. Nous proposons une méthode qui apprend automatiquement un dictionnaire de taille adaptée grâce à un modèle Bayésien non paramétrique de type Buffet Indien. Le niveau de bruit est aussi estimé avec précision, de sorte que presque aucun réglage des paramètres n'est nécessaire. Les résultats comparatifs de débruitage illustrent la pertinence de la méthode proposée.

Abstract – Dictionary learning for sparse representation is well known in solving inverse problems in image processing. In general, the number of dictionary atoms is fixed in advance. We propose a dictionary learning approach that automatically learns a dictionary of adapted size thanks to a Bayesian non parametric approach : the Indian Buffet Process prior. The noise level is also accurately estimated so that nearly no parameter tuning is needed. The denoising comparative results show the relevance of the proposed method.

1 Introduction

En traitement d'image, les *problèmes inverses* (débruitage, déconvolution, inpainting, ...) sont des *problèmes mal posés* : ils peuvent avoir plusieurs solutions. Il est nécessaire d'ajouter des contraintes ou des a priori permettant de choisir un espace de représentation dans lequel on identifie une solution unique et pertinente. Depuis quelques années, les *représentations parcimonieuses* connaissent une évolution croissante en traitement du signal et des images. La parcimonie consiste à reconstruire un signal en utilisant *un petit nombre d'atomes* choisis dans un dictionnaire éventuellement redondant. Une première possibilité est de choisir le dictionnaire parmi un ensemble de dictionnaires préexistants en utilisant des fonctions mathématiques (la transformée en cosinus discrète, la transformée en ondelette ...). Néanmoins, le choix du dictionnaire reste crucial et influence beaucoup la qualité de la décomposition et de la reconstruction obtenues. Une deuxième possibilité consiste à apprendre un dictionnaire à partir d'un ensemble de signaux de référence. Plusieurs travaux proposent alors d'apprendre un dictionnaire *redondant* où le nombre d'atomes K est supérieur à la dimension P de l'espace depuis les travaux de [1].

Les méthodes d'optimisation ont beaucoup été explorées. K-SVD [2] trouve un dictionnaire optimal selon le critère de parcimonie retenu pour une image. Certaines méthodes basées sur K-SVD [3–5] suggèrent une augmentation ou une diminution de la taille du dictionnaire afin de déterminer automatiquement

le nombre *efficace* d'atomes du dictionnaire. L'apprentissage de dictionnaire en ligne pour la représentation parcimonieuse a été étudié dans [6] et a été amélioré dans [7] en ajoutant une étape de *mean-shift clustering* dans la mise à jour du dictionnaire. Un dictionnaire de taille adaptative est proposé dans [8]. Deux atomes initiaux sont récursivement améliorés visant ainsi à un compromis entre l'erreur de reconstruction et la parcimonie de la représentation. Dans ces méthodes d'optimisation, la parcimonie est typiquement favorisée par une pénalité L_0 ou L_1 sur l'ensemble des coefficients de codage.

L'apprentissage de dictionnaire par des approches bayésiennes a été étudié dans une moindre mesure. Une approche de cette famille est présentée dans [9] en introduisant une loi a priori Beta-Bernoulli sur le support des représentations pour favoriser la parcimonie. Elle fonctionne avec un nombre d'atomes fixé malgré sa connexion avec les approches bayésiennes non paramétriques. Notre contribution appartient aussi à la famille bayésienne, mais sans fixer à l'avance la taille du dictionnaire grâce à l'utilisation d'une loi a priori nommée Processus du Buffet Indien (IBP). L'échantillonnage de Gibbs est utilisé pour inférer le modèle. Le niveau de bruit est aussi estimé lors de l'inférence. Ainsi la méthode est complètement non paramétrique et le nombre de degré de liberté (le nombre d'atomes du dictionnaire) n'est pas fixé à l'avance.

La partie 2 rappelle le principe de l'apprentissage de dictionnaire. La partie 3 présente le modèle proposé et le processus Buffet Indien avant de décrire l'échantillonnage de Gibbs pour l'inférence. La partie 4 est consacrée aux résultats expérimentaux et à la discussion.

Thanks to the BNPSI ANR project no ANR-13-BS-03-0006-01 and to the Fondation Ecole Centrale Lille for funding.

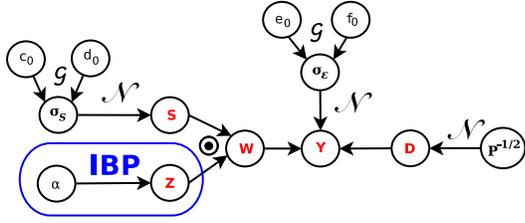


FIGURE 1 – Modèle graphique de l’IBP-DL

2 Apprentissage de dictionnaire

En traitement d’image, chaque observation est en général une imagerie (ou *patch* [10]) de taille $p \times p$ (par ex. 8×8) rangée dans un vecteur de dimension $P=p^2$ (par ex. 64) par ordre lexicographique. Soit $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_N] \in \mathbb{R}^{P \times N}$, l’ensemble des N patches extraits d’une image ou famille d’images. L’apprentissage de dictionnaire revient à factoriser la matrice des observations sous la forme du produit de deux matrices :

$$\mathbf{Y} = \mathbf{D}\mathbf{W} + \boldsymbol{\varepsilon} \quad (1)$$

où $\mathbf{D} = (\mathbf{d}_1, \dots, \mathbf{d}_K) \in \mathbb{R}^{P \times K}$ est le dictionnaire de K atomes, la matrice des coefficients est $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N] \in \mathbb{R}^{K \times N}$ et le bruit qui perturbe les données est généralement gaussien iid de variance σ_ε . Les coefficients \mathbf{w}_i encodent la représentation parcimonieuse de chaque observation \mathbf{y}_i . Le plus souvent dans la littérature, un dictionnaire de $K = 256$ ou 512 atomes est appris [2, 9, 10].

On peut aborder la question comme un problème d’optimisation jointe où la parcimonie est typiquement imposée par une pénalité L0 ou L1 (d’autres formulations sont possibles) :

$$(\mathbf{D}, \mathbf{W}) = \operatorname{argmin}_{(\mathbf{D}, \mathbf{W})} \frac{1}{2} \|\mathbf{Y} - \mathbf{D}\mathbf{W}\|_2^2 + \lambda \|\mathbf{W}\|_1 \quad (2)$$

Différentes méthodes d’optimisation alternée entre \mathbf{D} et \mathbf{W} ont été proposées pour résoudre ce problème [10]. Il est à noter que le choix du paramètre de régularisation λ en fonction du niveau de bruit σ_ε est important.

Dans le cadre bayésien, le problème s’écrit typiquement sous la forme d’une loi a posteriori :

$$p(\mathbf{D}, \mathbf{W}, \sigma_\varepsilon | \mathbf{Y}) \propto p(\mathbf{Y} | \mathbf{D}, \mathbf{W}, \sigma_\varepsilon) p(\mathbf{D}, \mathbf{W}, \sigma_\varepsilon) \quad (3)$$

La vraisemblance $p(\mathbf{Y} | \mathbf{D}, \mathbf{W}, \sigma_\varepsilon)$ est construite conformément au modèle (1) tandis que la loi a priori $p(\mathbf{D}, \mathbf{W}, \sigma_\varepsilon)$ permet de régulariser le problème. En utilisant par exemple l’échantillonnage de Gibbs pour l’inférence, le problème peut être résolu en échantillonnant alternativement :

$$p(\mathbf{W} | \mathbf{Y}, \mathbf{D}, \sigma_\varepsilon) \propto p(\mathbf{Y} | \mathbf{D}, \mathbf{W}, \sigma_\varepsilon) p(\mathbf{W}) \quad (4)$$

$$p(\mathbf{D} | \mathbf{Y}, \mathbf{W}, \sigma_\varepsilon) \propto p(\mathbf{Y} | \mathbf{D}, \mathbf{W}, \sigma_\varepsilon) p(\mathbf{D}) \quad (5)$$

$$p(\sigma_\varepsilon | \mathbf{Y}, \mathbf{W}, \sigma_\varepsilon) \propto p(\mathbf{Y} | \mathbf{D}, \mathbf{W}, \sigma_\varepsilon) p(\sigma_\varepsilon) \quad (6)$$

Dans le cadre bayésien paramétrique, la taille du dictionnaire doit être fixée avant d’inférer les paramètres du modèle. Dans la suite, nous proposons une méthode d’apprentissage sans fixer la taille du dictionnaire à l’avance en utilisant une loi a priori bayésienne *non paramétrique* [11].



FIGURE 2 – Barbara, $\sigma_\varepsilon=40$: Dictionnaire de l’IBP-DL

3 Approche proposée : IBP-DL

L’acronyme IBP-DL signifie Indian Buffet Process for Dictionary Learning. Afin de prendre en compte la question du choix du nombre de degrés de liberté (la taille du dictionnaire), nous proposons une approche Bayésienne non paramétrique (BNP). Le processus Buffet Indien (IBP) [11, 12] est proposé comme loi a priori BNP sur le support de représentation. La matrice binaire \mathbf{Z} ci-dessous contrôle ce support.

3.1 Modèle : IBP-DL

La figure 1 montre le modèle graphique de l’IBP-DL qui peut être décrit par :

$$\mathbf{y}_i = \mathbf{D}\mathbf{w}_i + \boldsymbol{\varepsilon}_i, \forall 1 \leq i \leq N \quad (7)$$

$$\mathbf{d}_k \sim \mathcal{N}(0, P^{-1}\mathbb{I}_P), \forall k \in \mathbb{N} \quad (8)$$

$$\mathbf{w}_i = \mathbf{z}_i \odot \mathbf{s}_i, \forall 1 \leq i \leq N \quad (9)$$

$$\mathbf{Z} \sim \text{IBP}(\alpha) \quad (10)$$

$$\mathbf{s}_i \sim \mathcal{N}(0, \sigma_s^2 \mathbb{I}_K), \forall 1 \leq i \leq N \quad (11)$$

$$\boldsymbol{\varepsilon}_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 \mathbb{I}_P), \forall 1 \leq i \leq N \quad (12)$$

où $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ est la distribution Gaussienne d’espérance $\boldsymbol{\mu}$ et de covariance $\boldsymbol{\Sigma}$, \mathbf{d}_k représente l’atome k de \mathbf{D} , \odot est le produit matriciel de Hadamard, \mathbb{I} représente une matrice identité. Le parcimonie de \mathbf{W} est induite par celle de la matrice binaire \mathbf{Z} grâce à l’a priori IBP (voir paragraphe suivant). Sauf pour σ_D^2 qui est fixée à $1/P$, les lois a priori conjuguées sont utilisées pour les paramètres $\boldsymbol{\theta} = (\sigma_s^2, \sigma_\varepsilon^2, \alpha)$: distributions Gamma Inverse pour les variances avec de faibles hyperparamètres ($c_0=d_0=e_0=f_0=10^{-6}$) dans (18), (19), (20) pour des hyperpriors non-informatives ; une distribution Gamma est utilisée pour α associé à une loi de Poisson dans l’IBP. Nous soulignons que la variance du bruit σ_ε^2 est aussi estimée lors de l’inférence. La figure 2 montre un exemple produit par IBP-DL pour l’image Barbara.

3.2 Processus de Buffet Indien (IBP)

La matrice binaire \mathbf{Z} est une représentation parcimonieuse dans un modèle à caractéristiques latentes (*Latent Features*). Si la donnée i a la caractéristique k alors $\mathbf{Z}(k, i)=1$, (0 si non), notons que chaque donnée peut posséder un ensemble de caractéristiques. Dans le cas où le nombre total de caractéristiques est inconnu, l’IBP permet de définir un a priori sur \mathbf{Z} de taille potentiellement infinie. L’IBP est obtenu à partir d’un processus de Bernoulli avec une loi a priori conjuguée Beta sur les paramètres. Dans la métaphore du Buffet Indien, chaque client

(donnée) i choisit d'abord le plat (caractéristique) k avec probabilité m_k/i où m_k est le nombre de fois où le plat k a été choisi par les clients précédents. Puis, ce client i choisit encore un nombre $k_{new} \sim \text{Poisson}(\alpha/i)$ de nouveaux plats. Cette étape permet d'enrichir progressivement le dictionnaire. L'IBP est caractérisée par une distribution sur les classes d'équivalence de matrices binaires [12] qui dépend de m_k le nombre d'observations utilisant l'atome k , K_+ le nombre d'atomes tels que $m_k > 0$, le nombre des données N et le paramètre $\alpha > 0$ de l'IBP. Ce paramètre α quantifie le niveau de régularisation puisque $\mathbb{E}[K_+] = \alpha \log N$. Plus α est petit, plus la régularisation est forte. En bref, IBP génère des matrices binaires *creuses* et *potentiellement infinies*. En pratique, \mathbf{Z} est de taille finie car il suffit de travailler sur les K_+ atomes.

3.3 Algorithme pour l'échantillonnage de Gibbs

Cette partie décrit l'échantillonnage de Gibbs de la loi a posteriori $p(\mathbf{D}, \mathbf{W}, \boldsymbol{\theta} | \mathbf{Y})$ (Algo. 1). Le dictionnaire est initialement vide. Notons que l'ordre de l'échantillonnage \mathbf{Z} - \mathbf{D} - \mathbf{S} - $\boldsymbol{\theta}$ est important à cause de la marginalisation par rapport à \mathbf{D} dans (14) [13]. Dans nos expériences, 9 images (512×512) sont traitées pour 2 niveaux de bruit $\sigma_\epsilon=25$ et 40. La valeur initiale de $\hat{\sigma}_\epsilon$ est fixée à $2\sigma_\epsilon$. Chaque image a 255025 patches de taille 8×8 (soit $P=64$) avec recouvrement. IBP-DL est entraîné sur $N=16129$ patches avec 50% recouvrement. La moyenne de patches est traitée à part. L'échantillonnage de $\mathbf{Z} \sim \text{IBP}(\alpha)$ se décompose en 2 étapes [11] : l'inférence de \mathbf{Z} et l'ajout de nouvelles lignes pour \mathbf{Z} correspond au choix des nouveaux atomes. 1) *Mise à jour* $z_{ki} = \mathbf{Z}(k, i)$. Nous avons $p(z_{ki} = 1 | \mathbf{Z}_{k,-i}) = m_{k,-i}/N$ avec $m_{k,-i}$ le nombre de données utilisant l'atome k sauf la donnée i . D'après la règle de Bayes :

$$p(z_{ki} | \mathbf{Y}, \mathbf{D}, \mathbf{Z}_{k,-i}, \mathbf{S}, \boldsymbol{\theta}) \propto p(\mathbf{Y} | \mathbf{D}, \mathbf{W}, \boldsymbol{\theta}) p(z_{ki} | \mathbf{Z}_{k,-i}) \quad (13)$$

La vraisemblance $p(\mathbf{Y} | \mathbf{D}, \mathbf{Z}, \mathbf{S}, \boldsymbol{\theta})$ est facilement calculée à partir du modèle de bruit gaussien. En marginalisant par rapport à \mathbf{D} , il vient :

$$p(\mathbf{Y} | \mathbf{Z}, \mathbf{S}, \sigma_\epsilon^2, \sigma_D^2) = \frac{\exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \text{tr} \left[\mathbf{Y} (\mathbb{I} - \mathbf{W}^T \mathbf{M} \mathbf{W}) \mathbf{Y}^T \right] \right\}}{(2\pi)^{NP/2} \sigma_\epsilon^{(N-K)P} \sigma_D^{KP} |\mathbf{W} \mathbf{W}^T + \frac{\sigma_\epsilon^2}{\sigma_D^2} \mathbb{I}|^{P/2}} \quad (14)$$

où $\mathbf{M} = (\mathbf{W} \mathbf{W}^T + \frac{\sigma_\epsilon^2}{\sigma_D^2} \mathbb{I})^{-1}$. Lorsque $\mathbf{Z}(k, :) = \mathbf{0} \forall i$, on supprime l'atome \mathbf{d}_k dans \mathbf{D} et la ligne $\mathbf{Z}(k, :)$.

2) *Ajout de nouveaux atomes*. Soit k_{new} le nombre de nouveaux atomes, sa loi a posteriori est

$$p(k_{new}) \propto p(\mathbf{Y} | \mathbf{D}, \mathbf{Z}^*, \mathbf{S}, \boldsymbol{\theta}) \text{Poisson}(k_{new}, \frac{\alpha}{N}) \quad (15)$$

où \mathbf{Z}^* est la matrice binaire proposée par la concaténation de \mathbf{Z} et k_{new} lignes fixées à 1 pour la donnée i et 0 sinon. En pratique, nous calculons les poids dans (15) pour $0 \leq k_{new} \leq K_{max}$. Ensuite, une loi multinomiale permet de décider k_{new} . On pose $\gamma_\epsilon = 1/\sigma_\epsilon^2$, $\gamma_D = 1/\sigma_D^2$, $\gamma_S = 1/\sigma_S^2$. L'échantillonnage de \mathbf{D} , \mathbf{W} et $\boldsymbol{\theta}$ est fait selon :

```

Init. :  $\mathbf{Z}=\emptyset, \mathbf{D}=\emptyset, \alpha=1, \sigma_D^2 = P^{-1}, \sigma_S^2 = 1, \sigma_\epsilon, K_{max}=4$ 
Résultats :  $\mathbf{D} \in \mathbb{R}^{P \times K}, \mathbf{Z} \in \{0, 1\}^{K \times P}, \mathbf{S} \in \mathbb{R}^{K \times P}, \sigma_\epsilon$ 
Pour itération  $t=1 : T$ 
   $\mathbf{Z} \sim \text{IBP}(\alpha)$ 
  Pour donnée  $i=1 : N$ 
    Pour atome  $k=1 : \text{size}(\mathbf{Z}, 1)$ 
      | Échantillonner  $\mathbf{Z}(k, i)$  selon (13), (14)
    fin
    Échantillonner  $k_{new}$  de nouveaux atomes (15)
    Compléter  $\mathbf{Z}$  avec  $k_{new}$  lignes
    Compléter  $\mathbf{S}$  avec  $k_{new}$  lignes  $\sim \mathcal{N}(0, \sigma_S^2)$ 
  fin
  Pour  $f=1 : F$  boucle interne
    Pour atome  $k=1 : \text{size}(\mathbf{Z}, 1)$ 
      |  $\mathbf{d}_k \sim \mathcal{N}(\boldsymbol{\mu}_{dk}, \boldsymbol{\Sigma}_{dk})$  (16)
      |  $\mathbf{S}(k, \mathbf{z}_k \neq 0) \sim \mathcal{N}(\boldsymbol{\mu}_{sk}, \boldsymbol{\Sigma}_{sk})$  (17)
    fin
    Échantillonner  $\sigma_S$  selon (18)
    Échantillonner  $\sigma_\epsilon$  selon (19)
  fin
  Échantillonner  $\alpha$  selon (20)
fin

```

Algorithm 1: Pseudo-algorithme de la méthode IBP-DL.

$$\mathbf{D} \begin{cases} p(\mathbf{d}_k | \mathbf{Y}, \mathbf{Z}, \mathbf{S}, \mathbf{D}_{-k}, \boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\Sigma}_{\mathbf{d}_k}, \boldsymbol{\mu}_{\mathbf{d}_k}) \\ \boldsymbol{\Sigma}_{\mathbf{d}_k} = \left(\gamma_D \mathbb{I}_p + \gamma_\epsilon \sum_{i=1}^N w_{ki}^2 \right)^{-1} \\ \boldsymbol{\mu}_{\mathbf{d}_k} = \gamma_\epsilon \boldsymbol{\Sigma}_{\mathbf{d}_k} \sum_{i=1}^N w_{ki} (\mathbf{y}_i - \mathbf{D} \mathbf{w}_i + \mathbf{d}_k w_{ki}) \end{cases} \quad (16)$$

$$\mathbf{S} \begin{cases} p(s_{ki} | \mathbf{Y}, \mathbf{D}, \mathbf{Z}, \mathbf{S}_{k,-i}, \boldsymbol{\theta}) \propto \mathcal{N}(\boldsymbol{\mu}_{s_{ki}}, \boldsymbol{\Sigma}_{s_{ki}}) \\ z_{ki} = 1 \Rightarrow \begin{cases} \boldsymbol{\Sigma}_{s_{ki}} = (\gamma_\epsilon \mathbf{d}_k^T \mathbf{d}_k + \gamma_S)^{-1} \\ \boldsymbol{\mu}_{s_{ki}} = \gamma_\epsilon \boldsymbol{\Sigma}_{s_{ki}} \mathbf{d}_k^T (\mathbf{y}_i - \mathbf{D} \mathbf{w}_i + \mathbf{d}_k s_{ki}) \end{cases} \\ z_{ki} = 0 \Rightarrow \begin{cases} \boldsymbol{\Sigma}_{s_{ki}} = \sigma_S^2 \\ \boldsymbol{\mu}_{s_{ki}} = 0 \end{cases} \end{cases} \quad (17)$$

$$\gamma_S \sim \mathcal{G} \left(c_0 + \frac{KN}{2}, d_0 + \frac{1}{2} \sum_{i=1}^N \mathbf{s}_i^T \mathbf{s}_i \right) \quad (18)$$

$$\gamma_\epsilon \sim \mathcal{G} \left(e_0 + \frac{NP}{2}, f_0 + \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^N \|\mathbf{y}_i - \mathbf{D} \mathbf{w}_i\|_{l_2}^2 \right) \quad (19)$$

$$\alpha \sim \mathcal{G} \left(1 + \sum_{i=1}^N \mathbf{z}_i^T \mathbf{z}_i, \frac{1}{1+N} \right) \quad (20)$$

Dans le modèle proposé, tous les paramètres sont échantillonnés, y compris le paramètre α de l'IBP et la variance du bruit σ_ϵ^2 .

4 Résultats et discussion

La pertinence d'IBP-DL est illustrée en comparant ses résultats de débruitage avec les autres méthodes (Tableau 1). La méthode

	Cas $\sigma_\epsilon=25$ ($\approx 20.17dB$)				Cas $\sigma_\epsilon=40$ ($\approx 16.08dB$)			
Barbara	29.06 - 116 \leq 120 \leq 126				26.27 - 52 \leq 59 \leq 65			
	29.28	29.37	27.95	28.96	26.49	26.65	25.37	26.54
Boat	28.92 - 87 \leq 91 \leq 95				26.73 - 39 \leq 43 \leq 47			
	29.04	29.13	28.49	28.11	26.85	26.93	26.38	26.12
Cameraman	28.56 - 491 \leq 668 \leq 846				26.27 - 110 \leq 120 \leq 126			
	28.98	28.81	28.09	26.90	26.57	26.72	25.50	25.45
Fingerprint	26.55 - 46 \leq 47 \leq 48				23.85 - 25 \leq 30 \leq 33			
	26.61	26.94	21.76	25.51	24.16	24.44	19.68	23.48
GoldHill	28.79 - 51 \leq 55 \leq 58				26.95 - 18 \leq 21 \leq 25			
	28.89	28.97	28.51	28.44	27.01	27.03	26.80	26.49
House	31.55 - 58 \leq 62 \leq 64				29.10 - 25 \leq 27 \leq 30			
	31.85	32.12	30.37	30.94	29.33	29.53	28.11	28.29
Lena	31.12 - 61 \leq 65 \leq 69				28.79 - 26 \leq 31 \leq 36			
	31.21	31.29	30.44	30.50	28.86	28.90	28.03	28.20
Mandrill	24.31 - 192 \leq 214 \leq 230				22.10 - 51 \leq 63 \leq 74			
	24.36	24.38	23.58	23.53	22.17	22.26	21.71	22.23
Peppers	29.41 - 117 \leq 136 \leq 148				27.07 - 44 \leq 57 \leq 63			
	29.60	29.71	28.35	28.27	27.25	27.36	26.54	26.24

TABLE 1 – Résultats de débruitage sur 9 images pour 2 niveaux de bruit $\sigma_\epsilon = 25$ et 40. Pour chaque image, les chiffres en gras sont le PSNR et la taille du dictionnaire issus d’IBP-DL. Ci-dessous, de gauche à droite : le PSNR utilisant K-SVD avec le nombre d’atomes déduit d’IBP-DL ; K-SVD avec 256 atomes ; K-SVD avec 256 atomes appris à partir de l’ensemble réduit (comme IBP-DL) ; NL-means.

de débruitage [14] utilise *Orthogonal Matching Pursuit* et le dictionnaire K-SVD¹. Les images sont débruitées en utilisant cette méthode avec les mêmes initialisations mais en remplaçant le dictionnaire et le niveau de bruit issus de IBP-DL. IBP-DL produit de meilleurs résultats que NL-means, une méthode de référence de débruitage [15]. Ses performances sont au moins comparables aux méthodes basées sur K-SVD. Le résultat de débruitage de K-SVD avec le nombre d’atomes proposé par IBP-DL est similaire à celui de K-SVD avec 256 atomes. Notons que IBP-DL propose souvent un dictionnaire de taille $K < 64$ qui n’est pas toujours un dictionnaire redondant mais la performance de débruitage reste comparable. Nous remarquons au passage la sensibilité de K-SVD à l’ensemble d’apprentissage. Le tableau 1 montre que les résultats de K-SVD entraîné sur 16129 patches comme IBP-DL sont nettement moins bons que ceux de K-SVD entraîné sur 255025 patches. Nous comparons aussi nos résultats avec DLENE [8], une méthode récente qui adapte également la taille du dictionnaire en visant un compromis entre l’erreur de reconstruction et la parcimonie. Pour Barbara avec $\sigma_\epsilon=25$, DLENE donne $\text{PSNR}_{\text{DLENE}}=28.82$ dB et $\text{PSNR}_{\text{IBP-DL}}=29.06$ dB ; pour Peppers $\text{PSNR}_{\text{DLENE}}=27.27$ dB et $\text{PSNR}_{\text{IBP-DL}}=27.07$ dB avec $\sigma_\epsilon=40$. En général, l’IBP-DL est aussi performant que DLENE. Une limite de notre algorithme est son coût de calcul dû à l’échantillonnage de Gibbs malgré l’implémentation d’une version accélérée [16] Un autre type d’inférence est envisagé pour réduire le temps de calcul. Cependant notons qu’un résultat essentiel de cette approche est l’estimation avec une bonne précision du niveau de bruit. L’erreur d’estimation varie entre 2%- 14% pour $\sigma_\epsilon=25$ et entre 0.2%- 8% pour $\sigma_\epsilon=40$. Cette estimation est un résultat essentiel et rend notre approche non paramétrique.

1. Le code Matlab de R. Rubinstein est disponible à <http://www.cs.technion.ac.il/~ronrubin/software.html>

5 Conclusion

Nous proposons l’une des seules méthodes permettant d’appréhender un dictionnaire de taille adaptative grâce à une approche bayésienne non paramétrique. Cette méthode présente de plus l’avantage de ne nécessiter pratiquement aucun ajustement de paramètre. Les performances de débruitage d’IBP-DL illustrent la pertinence des dictionnaires appris.

Références

- [1] B.A. OLSHAUSEN et D.J. FIELD : Emergence of simple-cell receptive properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [2] M. AHARON, M. ELAD et A. BRUCKSTEIN : K-SVD : An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Sig. Proc.*, 54(11):4311–4322, 2006.
- [3] R. MAZHAR et P.D. GADER : EK-SVD : Optimized dictionary design for sparse representations. *In ICPR*, pages 1–4, 2008.
- [4] Jianzhou FENG, Li SONG, Xiaokang YANG et Wenjun ZHANG : Sub clustering K-SVD : Size variable dictionary learning for sparse representations. *IEEE ICIP*, pages 2149–2152, 2009.
- [5] C. RUSU et B. DUMITRESCU : Stagewise k-svd to design efficient dictionaries for sparse representations. *IEEE Sig. Proc. Letters*, 19(10):631–634, 2012.
- [6] J. MAIRAL, F. BACH, J. PONCE et G. SAPIRO : Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.*, 11:19–60, 2010.
- [7] N. RAO et F. PORIKLI : A clustering approach to optimize online dictionary learning. *Proc. of ICASSP*, pages 1293–1296, 2012.
- [8] M. MARSOUSI, K. ABHARI, P. BABYN et J. ALIREZAIE : An adaptive approach to learn overcomplete dictionaries with efficient numbers of elements. *IEEE Trans. Sig. Proc.*, 62(12):3272–3283, 2014.
- [9] M. ZHOU, H. CHEN, J. PAISLEY, L. REN, L. LI, Z. XING, D. DUNSON, G. SAPIRO et L. CARIN : Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images. *IEEE Trans. Im. Proc.*, 21:130–144, 2012.
- [10] I. TOSIC et P. FROSSARD : Dictionary learning. *IEEE Sig. Proc. Magazine*, 28(2):27–38, 2011.
- [11] T. GRIFFITHS et Z. GHAHRAMANI : The indian buffet process : An introduction and review. *J. Mach. Learn. Res.*, 12:1185–1224, 2011.
- [12] T. GRIFFITHS et Z. GHAHRAMANI : Infinite latent feature models and the indian buffet process. *In Advances in NIPS 18*, pages 475–482. MIT Press, 2006.
- [13] D. A VAN DYK et T. PARK : Partially collapsed gibbs samplers. *J. of the American Stat. Ass.*, 103(482):790–796, 2008.
- [14] M. ELAD et M. AHARON : Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. on Im. Proc.*, 15(12):3736–3745, 2006.
- [15] A. BUADES, B. COLL et J.M. MOREL : A review of image denoising algorithms, with a new one. *Multiscale Modeling & Simulation*, 4(2):490–530, 2005.
- [16] F. DOSHI-VELEZ et Z. GHAHRAMANI : Accelerated sampling for the indian buffet process. *In Proc. of ICML*, pages 273–280. ACM, 2009.