INDIAN BUFFET PROCESS DICTIONARY LEARNING FOR IMAGE INPAINTING

Hong Phuong Dang, Pierre Chainais

Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL
Centre de Recherche en Informatique Signal et Automatique de Lille, F-59000 Lille, France
hong_phuong.dang, pierre.chainais@ec-lille.fr

ABSTRACT

Ill-posed inverse problems call for adapted models to define relevant solutions. Dictionary learning for sparse representation is often an efficient approach. In many methods, the size of the dictionary is fixed in advance and the noise level as well as regularization parameters need some tuning. Indian Buffet process dictionary learning (IBP-DL) is a Bayesian non parametric approach which permits to learn a dictionary with an adapted number of atoms. The noise and sparsity levels are also inferred so that the proposed approach is really non parametric: no parameters tuning is needed. This work adapts IBP-DL to the problem of image inpainting by proposing an accelerated collapsed Gibbs sampler. Experimental results illustrate the relevance of this approach.

Index Terms— sparse representations, dictionary learning, inverse problems, Indian Buffet Process

1. INTRODUCTION

Many works have proposed to use sparse representations [1] to solve ill-posed inverse problems. Such representations can be inferred by *dictionary learning (DL)* from a set of reference signals. Many DL methods are based on solving an optimization problem where sparsity is typically promoted by an L0 or L1 penalty terms on the set of encoding coefficients. Such approaches can be numerically very efficient but may suffer from some limitations. They often call for some parameter tuning depending on the noise level which also needs to be estimated. Moreover, the size of the redundant dictionary is generally fixed in advance [1, 2] even though there has been some tentative to overcome such a constraint [3–6]. In [7], a Bayesian DL method implemented using Gibbs sampling is proposed, where sparsity is promoted through an adapted Beta-Bernoulli prior to enforce many coefficients to zero. This corresponds to a parametric approximation of the *Indian Buffet Process (IBP)* [8] where one works with a (large) fixed number of atoms. In a previous work [9] we have proposed the IBP-DL that is really a *Bayesian non parametric (BNP)* approach for denoising thanks to the use of the IBP prior which both promotes sparsity and deal with an adaptive number of atoms. IBP-DL starts from an empty dictionary, except the constant atom to deal with the DC component apart, and learns a potentially infinite number of atoms. This approach does not need to tune any parameter: the noise level and the sparsity level are inferred as well.

The present contribution describes how to use IBP-DL for image inpainting, beyond basic denoising. We derive a collapsed Gibbs sampling algorithm and its accelerated version [10], which have not been proposed yet in the inpainting case, up to our knowledge. Numerical experiments illustrate the relevance of the proposed approach.

Section 2 recalls on the problem of dictionary learning. Section 3 recalls on the IBP prior. Section 4 presents the IBP-DL model for inpainting and the sampling algorithm for inference. Section 6 illustrates the relevance of our DL approach on image inpainting experiments with respect to other methods. Section 7 gathers conclusions and prospects.

2. DICTIONARY LEARNING (DL)

Let $\mathbf{Y}\in \mathbb{R}^{P\times N}$ a set of N column vectors $\mathbf{y}_i$ representing a patch of size $\sqrt{P}\times\sqrt{P}$ (e.g 8×8), in lexicographic order. $\mathbf{H}$ is the observation operator of patches $\mathbf{X}\in\mathbb{R}^{P\times N}$ in the initial image, e.g. a binary mask in the inpainting problem. Patches are represented by the coefficients $\mathbf{W}$=$[\mathbf{w}_1, ..., \mathbf{w}_N]\in\mathbb{R}^{K\times N}$ of their representation in a dictionary $\mathbf{D}$=$[\mathbf{d}_1, ..., \mathbf{d}_K]\in\mathbb{R}^{P\times K}$ with K atoms so that

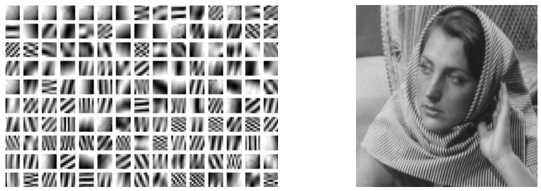
$$\begin{array}{l}\mathbf{Y} = \mathbf{HX} + \boldsymbol{\varepsilon}\\ \mathbf{X} = \mathbf{DW}\end{array} \quad (1)$$

where $\boldsymbol{\varepsilon}\in\mathbb{R}^{P\times N}$ is a white Gaussian noise. Each $\mathbf{x}_i$ is described by $\mathbf{x}_i = \mathbf{D}\mathbf{w}_i$ where $\mathbf{w}_i$ is sparse. Various approaches have been proposed [1] to solve this problem by an alternate optimization on $\mathbf{D}$ and $\mathbf{W}$ for fixed size dictionaries, e.g. with 256 or 512 atoms [2]. Sparsity is typically imposed through a L0 or L1-penalty. Note that the weight of the regularization term is important and depends on the noise level $\sigma_{\boldsymbol{\varepsilon}}$.

In the Bayesian framework, the problem is translated in a Gaussian likelihood according to the model (1). The prior

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

978-1-4673-7802-4/16/$31.00 ©2016 IEEE



(a)Dictionary : 150 atoms (b) A segment of Barbara

Fig. 1. IBP-DL dictionary of 150 atoms on a segment of size 256×256 of Barbara grayscale image

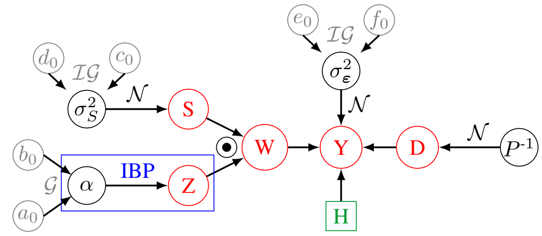


Fig. 2. Graphical model of IBP-DL for inpainting.

$p(\mathbf{D},\mathbf{W},\sigma_\varepsilon)$ plays the role of regularization and the joint posterior distribution writes:

$$p(\mathbf{D},\mathbf{W},\sigma_\varepsilon \mid \mathbf{Y},\mathbf{H}) \propto p(\mathbf{Y} \mid \mathbf{H},\mathbf{D},\mathbf{W},\sigma_\varepsilon)p(\mathbf{D},\mathbf{W},\sigma_\varepsilon) \quad (2)$$

Using Gibbs sampling for inference, the problem is solved by sampling alternately $\mathbf{D}$, $\mathbf{W}$ and σ_ε. In BNP framework, the dictionary can be learnt without setting the size in advance and no parameter tuning is necessary. To deal both with the sparsity constraint and the desirable adaptive number of atoms, we use a Indian Buffet Process (IBP) prior [9] that can be seen as a generalization of a Bernoulli prior on the support of representation as in a Bernoulli-Gaussian model. Fig 1 shows the dictionary and the reconstruction of a segment on Barbara obtained by IBP-DL. The purpose of this paper is to deal with the problem of inpainting, that is when there are missing pixels at random locations, in the presence of noise.

3. INDIAN BUFFET PROCESS (IBP)

The IBP [8] emulates a distribution over binary matrices $\mathbf{Z}$ in which the number of lines is potentially infinite. It provides a relevant BNP prior to deal with problems of latent feature analysis. Each binary coefficient $\mathbf{Z}(k,i)$ indicates whether the observation $\mathbf{y}_i$ uses feature $\mathbf{d}_k$. The classic generative process of IBP is as follows. N customers (observations) enter an Indian restaurant and select dishes (atoms) from a potentially infinite buffet. The first customer tries Poisson(α) dishes. Recursively, the i-th customer takes portions from previously-selected dish k with probability m_k/i, where m_k is the number of previous customers who selected dish k before him; then he also tries Poisson(α/i) new dishes, which corresponds to adding new lines to matrix $\mathbf{Z}$. In [8], a construction of the IBP as the infinite-limit construction of a finite model with K_f features is also proposed. In the finite model, each feature k is assigned to a probability π_k which is sampled from Beta(α/K_f,1). Then, $\mathbf{Z}(k,i)\sim$Bernoulli(π_k): each customer samples independently a dish of the other customers. Taking the limit $K_f \to \infty$ yields the IBP. It appears that the ordering of the observations does not influence the distribution of $\mathbf{Z}$. This property is called exchangeability. The behaviour of IBP is governed by a concentration parameter α, which controls the expected total number of features $\mathbb{E}[K] \simeq \alpha \log N$. Note that the IBP prior both penalizes the actual number K of active rows in $\mathbf{Z}$ (the size of the dictionary later on) and promotes sparsity.

4. IBP-DL MODEL FOR INPAINTING

The IBP-DL model for inpainting is $\forall 1 \leq i \leq N$:

$$\mathbf{y}_i = \mathbf{H}_i\mathbf{D}\mathbf{x}_i + \varepsilon_i, \quad (3)$$
$$\mathbf{x}_i = \mathbf{D}\mathbf{w}_i, \text{ with } \mathbf{w}_i = \mathbf{z}_i \odot \mathbf{s}_i, \quad (4)$$
$$\mathbf{d}_k \sim \mathcal{N}(0, P^{-1}\mathbb{I}_P), \quad (5)$$
$$\mathbf{Z} \sim IBP(\alpha), \quad (6)$$
$$\mathbf{s}_i \sim \mathcal{N}(0, \sigma_s^2\mathbb{I}_K), \quad (7)$$
$$\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2\mathbb{I}_P). \quad (8)$$

where $\odot$ is the Hadamard product. Fig. 2 shows the graphical model. The presence of the operator $\mathbf{H}_i$ makes the model more general than the model used for simple denoising [9]. In this paper, we propose to illustrate this approach with an application to the inpainting problem where $\mathbf{H}_i$ is a binary diagonal matrix of size $P \times P$ corresponding to a binary mask: zeros indicate missing pixels. The model applies to any inverse problem where $\mathbf{H}_i$ is a linear operator, e.g. deblurring, compressed sensing... The observation matrix $\mathbf{Y}$ contains N column vectors $\mathbf{y}_i \in \mathbb{R}^P$. The representation coefficients are defined as $\mathbf{w}_i = \mathbf{z}_i \odot \mathbf{s}_i$, in the spirit of a Bernoulli-Gaussian model. The sparsity of $\mathbf{W}$ is induced by the sparsity of $\mathbf{Z}$ thanks to the IBP prior. The vector $\mathbf{z}_i \in \{0,1\}^K$ encodes which of the K atoms of $\mathbf{D}$ are used to represent $\mathbf{y}_i$; $\mathbf{s}_i \in \mathbb{R}^K$ represents the coefficients used for this representation. We emphasize that this model deals with a potentially infinite number of atoms $\mathbf{d}_k$ so that the size of the dictionary is not limited a priori. The IBP prior both penalizes the number K of (useful) atoms and promotes sparsity. Except for σ_D^2 that is fixed to $1/P$, conjugate priors are used for parameters $\boldsymbol{\theta}$=($\sigma_S^2, \sigma_\varepsilon^2, \alpha$): vague inverse Gamma distributions for variances with very small hyperparameters (c_0=d_0=e_0=f_0=10^{-6})

$\mathcal{G}(x;a,b) = x^{a-1}b^a \exp(-bx)/\Gamma(a)$ pour $x > 0$
$\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$: Gaussian distribution with expectation $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.

are used for σ_ε^2, σ_S^2, and a $\mathcal{G}(1,1)$ for α associated to a Poisson law in the IBP. No important parameter is tuned.

5. MCMC ALGORITHMS FOR INFERENCE

```
Init. : K=0, Z=∅, D=∅, α=1, σ_D^2=P^{-1}, σ_S^2=1, σ_ε
Result: D ∈ R^{P×K}, Z ∈ {0;1}^{K×P}, S ∈ R^{K×P}, σ_ε
for each iteration t
    Use information form according to (10)
    for data i=1:N
        Remove influence of data i via eq.(11)
        m_{-i} ∈ N^{K×1} ← ∑Z(:, −i)
        for k ∈ k_used ← find(m_{-i} ≠ 0)
            Infer Z(k, i)
        Infer new atoms
        Restore influence of data i via eq.(11)
    for atoms k=1:K
        Sample d_k eq. (13)
        Sample s_k eq. (14)
    Sample σ_S, σ_ε, α see [9]
```

Algorithm 1: Pseudo-algorithm of accelerated Gibbs sampling for IBP-DL method.

In this section, we briefly describe the strategy to sample the posterior $P(\mathbf{D}, \mathbf{S}, \mathbf{Z}, \boldsymbol{\theta} | \mathbf{Y}, \mathbf{H})$ (see Algo.1). We derive a collapsed Gibbs sampler for $\mathbf{Z}$ and then propose an accelerated version [10] using sufficient statistics. The inference of new atoms is processed in a proper manner thanks to a Metropolis-Hastings step.

Sampling Z. Sampling from an IBP goes in 2 steps: 1) update the $z_{ki} = \mathbf{Z}(k, i)$ for 'active' atoms k such that $m_{k,-i} > 0$ (the number of patches $\neq i$ using $\mathbf{d}_k$); 2) add new rows to $\mathbf{Z}$ (will create new atoms in dictionary $\mathbf{D}$).

Collapsed Gibbs sampling (CGS). One can integrate out $\mathbf{D}$ to obtain a collapsed likelihood to compute the weights of the Bernoulli distributions used to sample $\mathbf{Z}(k, i)$. Due to the presence of mask matrices $\mathbf{H}_i$, integration must be carried out with respect to the rows of $\mathbf{D}$ in contrast with the denoising case [8, 9]. Let $\{\mathbf{F}_\ell\}$ the set of binary diagonal matrices of size N for ℓ=1,...,P such that $\mathbf{F}_\ell(i, i)$ indicates whether pixel at location ℓ in patch i is observed or not so that $\mathbf{F}_\ell(i, i)$=$\mathbf{H}_i(\ell, \ell)$=$H_{i,\ell}$. Let $\mathbf{M}_\ell = (\mathbf{W}\mathbf{F}_\ell\mathbf{F}_\ell^T\mathbf{W}^T + \frac{\sigma_\varepsilon^2}{\sigma_D^2}\mathbb{I}_K)^{-1}$ and $\mathbf{Y}_\ell = \mathbf{Y}(\ell, :)$. One obtains the marginalized likelihood:

$$p(\mathbf{Y} \mid \{\mathbf{F}_\ell\}, \mathbf{Z}, \mathbf{S}, \sigma_\varepsilon^2, \sigma_D^2) = \frac{1}{(2\pi)^{\|\mathbf{Y}\|_0/2}\sigma_\varepsilon^{\|\mathbf{Y}\|_0 - KP}\sigma_D^{KP}}$$
$$\prod_{\ell=1}^{P} |\mathbf{M}_\ell|^{1/2} \exp\left[-\frac{1}{2\sigma_\varepsilon^2}(\mathbf{Y}_\ell(\mathbb{I} - \mathbf{F}_\ell^T\mathbf{W}^T\mathbf{M}_\ell\mathbf{W}\mathbf{F}_\ell)\mathbf{Y}_\ell^T\right]$$

Accelerated Gibbs sampling (AGS). One limitation of CGS is its computational cost due to the evaluation of the collapsed likelihood leading to a complexity per-iteration of $O(N^3(K^2 + KP))$. The AGS [10] can reduce the complexity to $O(N(K^2 + KP))$ by proposing to maintain the $\mathbf{D}$

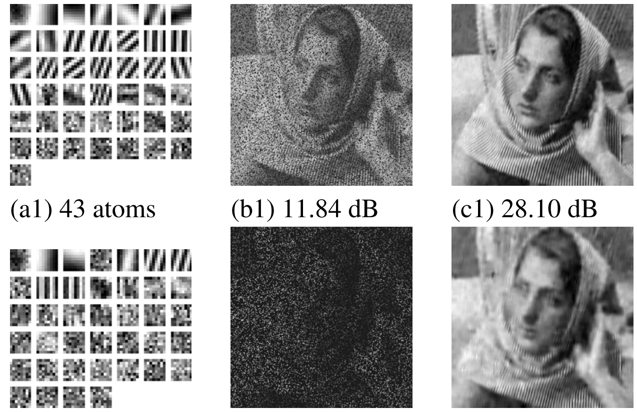

(a1) 43 atoms (b1) 11.84 dB (c1) 28.10 dB
(a2) 39 atoms (b2) 6.37 dB (c2) 23.74 dB

Fig. 3. IBP-DL restorations of a Barbara segment. From left to right are the IBP-DL dictionary, observed image, restored image. From top to bottom are restoration of masked noisy (σ_ε=25) image with 20% missing pixels, 80% missing pixels.

posterior instead of integrating out $\mathbf{D}$ entirely. The observations $\mathbf{Y}$ and the coefficients matrix $\mathbf{W}$ can split in two parts, one for the observation i, the other for the rest so that

$$P(z_{ki} = 1 \mid \mathbf{Y}, \mathbf{H}, \mathbf{W}, \sigma_D, \sigma_\varepsilon, \alpha) \propto \frac{m_{k,-i}}{N} \times \tag{9}$$
$$\int p(\mathbf{y}_i \mid \mathbf{H}_i, \mathbf{w}_i, \mathbf{D}) p(\mathbf{Y}_{-i} \mid \mathbf{H}_{\neq i}, \mathbf{W}_{-i}, \mathbf{D}) p(\mathbf{D} \mid \sigma_D) d\mathbf{D}$$
$$\propto \frac{m_{k,-i}}{N} \int p(\mathbf{y}_i \mid \mathbf{H}_i, \mathbf{D}, \mathbf{w}_i) \prod_{\ell=1}^{P} p(\mathbf{D}(\ell, :) \mid \mathbf{F}_\ell, \mathbf{W}_{-i}, \sigma_D) d\mathbf{D}$$

One can show that the posterior of $\mathbf{D}$ is a gaussian distribution with expectation $\boldsymbol{\mu}_{\mathbf{D}\ell}$ and covariance $\boldsymbol{\Sigma}_{\mathbf{D}\ell}$.

One may rather use *sufficient statistics* (information form)

$$\begin{aligned} g_{\mathbf{D}\ell} &= \boldsymbol{\Sigma}_{\mathbf{D}\ell}^{-1} = (1/\sigma_\varepsilon^2)\mathbf{M}_\ell^{-1} \\ h_{\mathbf{D}\ell} &= \boldsymbol{\mu}_{\mathbf{D}\ell} g_{\mathbf{D}\ell} = (1/\sigma_\varepsilon^2)\mathbf{Y}(\ell, :)\mathbf{F}_\ell^T\mathbf{W}^T \end{aligned} \tag{10}$$

to make it easy to deal with the influence of one individual observation i apart. Indeed, one can define

$$\begin{aligned} g_{\mathbf{D}\ell,\pm i} &= g_{\mathbf{D}\ell} \pm \sigma_\varepsilon^{-2} H_{i,\ell}\, \mathbf{w}_i\mathbf{w}_i^{\mathrm{T}} \\ h_{\mathbf{D}\ell,\pm i} &= h_{\mathbf{D}\ell} \pm \sigma_\varepsilon^{-2} H_{i,\ell}\, y_i(\ell)\, \mathbf{w}_i^{\mathrm{T}} \end{aligned} \tag{11}$$

as well as the corresponding $\boldsymbol{\mu}_{\mathbf{D}\ell,\pm i}$ and $\boldsymbol{\Sigma}_{\mathbf{D}\ell,\pm i}$. Since the likelihood is Gaussian, the integral in (9) is proportional to

$$\prod_{\ell=1}^{P} \mathcal{N}(y_i(\ell); H_{i,\ell}\boldsymbol{\mu}_{\mathbf{D}\ell,-i}\mathbf{w}_i, H_{i,\ell}\mathbf{w}_i^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{D}\ell,-i}\mathbf{w}_i + \sigma_\varepsilon^2) \tag{12}$$

At each iteration on data i, one uses (11) to remove/restore the influence of data i on the posterior distribution of z_{ki}.

Sampling new atoms. Following [11], a Metropolis-Hastings algorithm is used to sample the number k_{new} of new atoms as well as associated coefficients $\mathbf{S}_{new}$. This is carried out by dealing with rows of $\mathbf{Z}$ such that $m_{k,-i}$=0 that is when an

σ_{ε} ╲ Missing		80%	50%	20%	0%
0	BPFA	26.87	35.60	40.12	42.94
	IBP-DL	57 - **27.49**	47 - **35.40**	40 - **38.87**	150 - **44.94**
15	BPFA	25.17	29.31	29.93	32.14
	IBP-DL	62 - **25.28**	58 - **28.90**	45 - **30.68**	121 - **31.87**
25	BPFA	23.49	26.79	24.10	29.30
	IBP-DL	39 - **23.74**	52 - **26.54**	43 - **28.10**	67 - **28.90**

Table 1. Restoration results of a Barbara grayscale segment. The top part of each cell is the PSNR (dB) used BPFA with 256 atoms to restore the image. The bottom part of each cell, from left to right are the IBP-DL dictionary size K and the restoration PSNR (dB): K - PSNR.

atom is not used (z_{ki}=0 $\forall i$) or is used by patch i only (z_{ki}=0 $\forall j \neq i$), what we call *singletons*. To sample k_{new} amounts to sample the number of singletons: the activation of a new atom creates a new singleton. Let k_{sing} the number of singletons in matrix $\mathbf{Z}$ and $\mathbf{S}_{sing}$ the corresponding coefficients. Let $k_{prop} \in \mathbb{N}$ a proposal for the new number of singletons and $\mathbf{S}_{prop}$ the corresponding coefficients. One proposes a move $\zeta_{sing} = \{k_{sing}, \mathbf{S}_{sing}\} \rightarrow \zeta_{prop} = \{k_{prop}, \mathbf{S}_{prop}\}$ with probability $p(\zeta_{prop}) = P_K(k_{prop})p_S(S_{prop})$. We use the prior distributions to propose ζ_{prop} so that the acceptance ratio is a likelihood ratio, see (12).

Then $\mathbf{D}$ and $\mathbf{S}$ are sampled according to Gaussian distributions (see Appendix for detailed expressions):

$$p(\mathbf{d}_k|\mathbf{Y}, \mathbf{H}, \mathbf{Z}, \mathbf{S}, \mathbf{D}_{-k}, \boldsymbol{\theta}) = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{d}_k}, \boldsymbol{\Sigma}_{\mathbf{d}_k}) \quad (13)$$

$$p(s_{ki}|\mathbf{Y}, \mathbf{H}_i, \mathbf{D}, \mathbf{Z}, \mathbf{S}_{k,-i}, \boldsymbol{\theta}) = \mathcal{N}(\mu_{s_{ki}}, \Sigma_{s_{ki}}) \quad (14)$$

Sampling of σ_{ε}, σ_S, α goes as usual for denoising [9].

6. EXPERIMENTAL RESULTS

Denoising results have been presented in previous work [9]: IBP-DL performs as fine as other DL approaches while being non parametric. Note that the estimation error on the noise level is maximum 10% in this case. Here we focus on the more difficult problem of inpainting (missing pixels).

Fig. 3 displays some inpainting results on a segment (256×256) of Barbara grayscale image. IBP-DL is trained from a set of 62001 overlapping patches. Table 1 displays several restoration results by using IBP-DL compared to those of BPFA [7]. As a reminder, the BPFA method belongs to the same bayesian family. Despite a connection with the Indian Buffet Process, BPFA is not really a non-parametric approach: it is a parametric approximation of the IBP because it works with a fixed number of atoms in advance. The initial size of the BPFA dictionary depends on the size of the image. A subset of atoms is used in the end, see [7] for more details. In this experiment, the BPFA method fixes the maximal size of the dictionary to K=256 in advance. Then, the size of BPFA dictionaries is slightly smaller than 256, typically between 200 and 256 atoms.

A first basic remark is that IBP-DL brings a significant improvement over the restoration by the DC atom taken as a minimal reference (PSNR = 22 dB). Note that the reconstruction of the original image without noise by using IBP-DL is very accurate since one gets PSNR=44.97 dB with 150 atoms. In the worst case of 80% missing data and σ_{ε}=25, the restoration by IBP-DL yields a PSNR of 23.74 dB with 39 atoms. In all cases, an important observation is that IBP-DL restores the image with an adapted rather small number of atoms, $39 \leq K \leq 150$. For 80% missing data without noise one gets $\text{PSNR}_{\text{BPFA}}$= 26.87 dB using 256 atoms while $\text{PSNR}_{\text{IBP-DL}}$=27.49 dB with 57 atoms; for 50% missing data and σ_{ε}=25, $\text{PSNR}_{\text{BPFA}}$=26.79 dB and $\text{PSNR}_{\text{IBP-DL}}$=26.54 dB with K=52. IBP-DL performances are comparable to BPFA [7]. This confirms the efficiency of inferred IBP-DL dictionaries to solve an inverse problem such as inpainting. IBP-DL even improves on BPFA [7] and our observations support the interest of a really non parametric approach that is more adaptive to the actual content of the image. As a really non parametric approach, a natural limitation of our method is its increased computational cost due to the atom addition step. An accelerated version of the algorithm reduces the complexity to $O(N(K^2 + KP))$ that is still larger than the $O(K(N + P))$ complexity of BPFA with an initial number K of atoms.

7. CONCLUSION

The paper presents IBP-DL, an Indian Bufffet process dictionary learning approach for image inpainting (missing pixels) in the presence of noise. We have derived the corresponding collapsed and accelerated Gibbs samplers. The IBP-DL approach learns a dictionary of adaptive size. The resulting dictionary is often relatively small but restoration performances are at least similar to those of other methods. We emphasize that the IBP-DL approach also samples other parameters such as the noise and sparsity levels. IBP-DL appears as a really *non parametric* method with no parameter tuning. Note that the same method applies to compressive sensing. Future work will explore other inverse problems and other inference methods for scalability.

A. DETAILED EXPRESSIONS TO SAMPLE D AND S

$$\boldsymbol{\Sigma}_{\mathbf{d}_k} = (\sigma_D^{-2}\mathbb{I}_P + \sigma_{\varepsilon}^{-2}\sum_{i=1}^{N} w_{ki}^2 \mathbf{H}_i^T \mathbf{H}_i)^{-1}$$

$$\boldsymbol{\mu}_{\mathbf{d}_k} = \sigma_{\varepsilon}^{-2}\boldsymbol{\Sigma}_{\mathbf{d}_k}\sum_{i=1}^{N} w_{ki}(\mathbf{H}_i^T\mathbf{y}_i - \mathbf{H}_i^T\mathbf{H}_i\sum_{j\neq k}^{K}\mathbf{d}_j w_{ji})$$

$$z_{ki} = 1 \Rightarrow \begin{cases} \Sigma_{s_{ki}} = (\sigma_{\varepsilon}^{-2}\mathbf{d}_k^T\mathbf{H}_i^T\mathbf{H}_i\mathbf{d}_k + \sigma_S^{-2})^{-1} \\ \mu_{s_{ki}} = \sigma_{\varepsilon}^{-2}\Sigma_{s_{ki}}\mathbf{d}_k^T(\mathbf{H}_i^T\mathbf{y}_i - \mathbf{H}_i^T\mathbf{H}_i\sum_{j\neq k}^{K}\mathbf{d}_j w_{ji}) \end{cases}$$

$$z_{ki} = 0 \Rightarrow \Sigma_{s_{ki}} = \sigma_S^2 \ ; \mu_{s_{ki}} = 0$$

8. REFERENCES

[1] I. Tosic and P. Frossard, "Dictionary learning : What is the right representation for my signal," *IEEE Signal Process. Mag.*, vol. 28, pp. 27–38, 2011.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, pp. 4311–4322, 2006.

[3] R. Mazhar and P.D. Gader, "EK-SVD: Optimized dictionary design for sparse representations," in *Proc. of ICPR*, 2008.

[4] J. Feng, L. Song, X. Yang, and W. Zhang, "Sub clustering k-svd: Size variable dictionary learning for sparse representations," in *ICIP*, 2009, pp. 2149–2152.

[5] C. Rusu and B. Dumitrescu, "Stagewise k-svd to design efficient dictionaries for sparse representations," *IEEE Signal Process. Lett.*, vol. 19, pp. 631–634, 2012.

[6] M. Marsousi, K. Abhari, P. Babyn, and J. Alirezaie, "An adaptive approach to learn overcomplete dictionaries with efficient numbers of elements," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3272–3283, 2014.

[7] M. Zhou *et al*, "Nonparametric bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, pp. 130–144, 2012.

[8] T.L. Griffiths and Z. Ghahramani, "The Indian Buffet Process: An introduction and review," *J. Mach. Learn. Res.*, vol. 12, pp. 1185–1224, 2011.

[9] H-P. Dang and P. Chainais, "A Bayesian non parametric approach to learn dictionaries with adapted numbers of atoms," in *IEEE Int. Workshop on MLSP*, 2015, pp. 1–6.

[10] F. Doshi-Velez and Z. Ghahramani, "Accelerated sampling for the indian buffet process," in *ICML*, 2009, pp. 273–280.

[11] Knowles and Ghahramani, "Nonparametric Bayesian sparse factor models with application to gene expression modeling," *The Annals of Applied Statistics*, vol. 5, pp. 1534–1552, 2011.