Variational Inference for Nonparametric Subspace Dictionary Learning with Hierarchical Beta Process

Shaoyang Li, Xiaoming Tao, Jianhua Lu, *Fellow, IEEE*

Department of Electronic Engineering, Tsinghua University, Beijing, China
Tsinghua National Laboratory for Information Science and Technology (TNList)
Email: lishaoyang12@mails.tsinghua.edu.cn

Introduction

Nonparametric Bayesian models have been implemented in dictionary learning. However, for signal samples from multiple subspaces, existing methods only learn one uniform dictionary and thus are not optimal for representing the *subspace structures*. To address this issue, we first utilize a combination of Dirichlet process and hierarchical Beta process as priors to *infer the latent subspace number and dictionary dimension automatically*; second, to derive tractable variational inference, we modify the priors with the Sethuraman's construction and further employ the multinomial approximation. Experimental results indicate that our approach can achieve a set of nonparametric subspace dictionaries, while showing performance enhancements in the tasks of image denoising.

Nonparametric Bayesian Model

The conventional dictionary learning framework:

$$\boldsymbol{x}_i = \mathbf{D}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i$$

The basic model for *subspace dictionary learning*:

$$\boldsymbol{x}_i = \mathbf{D}_{c(i)}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i$$

where $c(i)$ denotes the subspace index of signal sample $\boldsymbol{x}_i$

In nonparametric Bayesian framework, our model includes 3 parts:

(1) *Sparse representation* using subspace dictionaries:

$$\begin{aligned}
&\boldsymbol{x}_i \sim \mathcal{N}(\mathbf{D}_{c(i)}\boldsymbol{w}_i, \alpha_{c(i)}^{-1}\mathbf{I}_P),\\
&\mathbf{D}_{c(i)} \sim \textstyle\prod_{k=1}^{K}\mathcal{N}(\mathbf{0}, \frac{1}{P}\mathbf{I}_P),\\
&\boldsymbol{w}_i = \boldsymbol{z}_i \odot \boldsymbol{s}_i,\ \ \boldsymbol{s}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_K),\\
&\boldsymbol{z}_i|\boldsymbol{\pi}_c \sim \textstyle\prod_{k=1}^{K}\mathrm{Bernoulli}(\pi_{ck}),\ \forall\ i : c(i) = c,
\end{aligned}$$

(2) Modeling *subspace number* with Dirichlet process:

$$\begin{aligned}
&G = \textstyle\sum_{c=1}^{\infty}\xi_c\delta_{G_c} \sim \mathcal{DP}(\eta, \{G_c\}_{c=1}^{\infty}),\\
&c(i) \sim \mathrm{Multinomial}(\boldsymbol{\xi}),\\
&\boldsymbol{\xi} \sim \mathrm{GEM}(\eta),\ \ \eta \sim \mathrm{Gamma}(a, b),\\
&\xi_c = \rho_c\textstyle\prod_{l=1}^{c-1}(1-\rho_l),\ \ \rho_c \sim \mathrm{Beta}(1, \eta)
\end{aligned}$$

Note: By placing a Dirichlet process over the multiple dictionaries, the number of underlying subspaces can be *automatically learned*.

(3) Describing *subspace correlation* with hierarchical Beta process

$$\begin{aligned}
&H = \textstyle\sum_{t=1}^{\infty}\upsilon_t\delta_{\phi_t} \sim \mathcal{BP}(\lambda, H_0),\ \ \phi_t \overset{\text{iid}}{\sim} H_0,\\
&G_c = \textstyle\sum_{k=1}^{\infty}\pi_{ck}\delta_{\varphi_{ck}} \overset{\text{iid}}{\sim} \mathcal{BP}(\gamma_c, H),\\
&\varphi_{ck} = \phi_{u_{ck}},\ \ u_{ck} \sim \mathrm{Multinomial}(\boldsymbol{\upsilon}),\\
&\pi_{ck} = \textstyle\prod_{m=1}^{k}\varpi_{cm},\ \ \varpi_{cm} \sim \mathrm{Beta}(\gamma_c, 1),\\
&\upsilon_t = \textstyle\prod_{j=1}^{t}\beta_j,\ \ \beta_j \sim \mathrm{Beta}(\lambda, 1).
\end{aligned}$$

Note: By placing a hierarchical Beta process over the subspace dictionaries, the correlations among them can also be described. In contrast to the conventional HBP priors, we utilize a *Sethuraman's stick-breaking construction* to achieve closed-form inference.

Variational Inference

We focus on variational inference procedures for the proposed DP-HBP-based model by updating factorized variational distributions $\mathcal{Q}$ to *minimize their KL divergence*, which is equivalent to the maximization of the marginal likelihood lower bound $\mathcal{L}$

$$\mathcal{L} = \mathbb{E}_{\mathcal{Q}}[p(\mathcal{X}, \mathcal{M}|\mathcal{H})] + \mathbb{H}[\mathcal{Q}],$$

To maximize it, we derive a *coordinate ascent inference* algorithm:

(1) Coordinate update for the Dirichlet process:

$$\begin{aligned}
\bar{\xi}_c \propto \exp\{&\mathbb{E}\left[-\tfrac{P}{2}\ln(2\pi\alpha_c) - \alpha_c\|\boldsymbol{x}_i - \mathbf{D}_c\boldsymbol{w}_i\|_2^2/2\right]\\
&+\mathbb{E}[\ln\rho_c + \textstyle\sum_{l=1}^{c-1}\ln(1-\rho_l)]\},\\
q(\rho_c) = \mathrm{Beta}&\left(1 + \textstyle\sum_{i=1}^{N}\bar{\xi}_c, \eta + \sum_{i=1}^{N}\sum_{l=c+1}^{C}\bar{\xi}_l\right),\\
q(\eta) = \mathrm{Gamma}&\left(a + C - 1, b - \textstyle\sum_{c=1}^{C-1}\mathbb{E}[\ln(1-\rho_c)]\right),
\end{aligned}$$

(2) Coordinate update for the hierarchical Beta process:

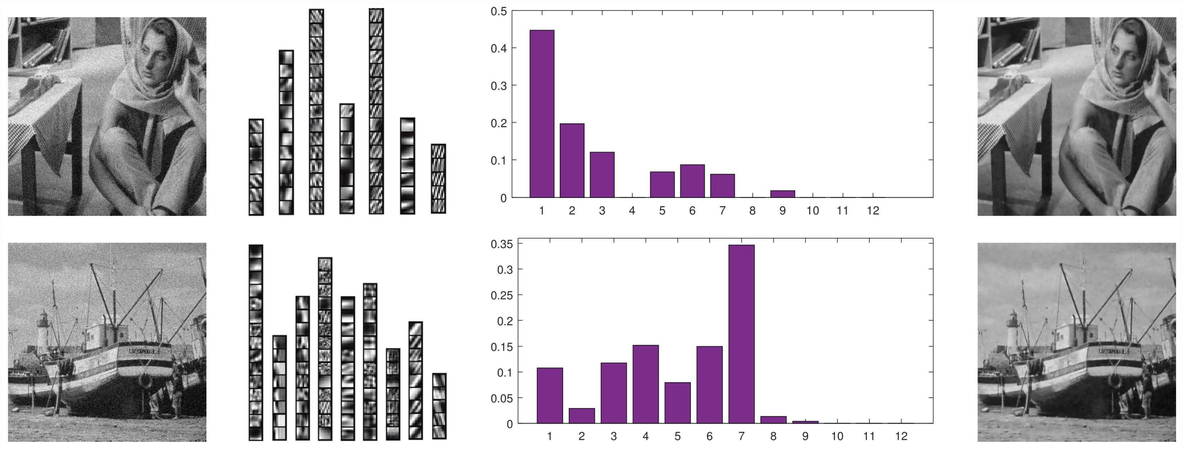
Note: To infer HBP parameters, we need to *evaluate an expectation term* which is a byproduct of lower bound and has not a closed form. Here we resort to *the multinomial approximation* to lower bond it.

$$\begin{aligned}
&\mathbb{E}_{\boldsymbol{\varpi}_c}\left[\ln\left(1 - \textstyle\prod_{m=1}^{k}\varpi_{cm}\right)\right]\\
&= \mathbb{E}_{\boldsymbol{\varpi}_c}\left[\ln\left(\textstyle\sum_{y=1}^{k}q_k(y)\frac{(1-\varpi_{cy})\prod_{m=1}^{y-1}\varpi_{cm}}{q_k(y)}\right)\right]\\
&\geq \mathbb{E}_{\boldsymbol{\varpi}_c}\mathbb{E}_y\left[\ln(1-\varpi_{cy}) + \textstyle\sum_{m=1}^{y-1}\ln\varpi_{cm}\right] + \mathbb{H}(q_k).
\end{aligned}$$

Note: For brevity, details of updated distributions is omitted here.

Experimental Results

We evaluate the proposed variational inference for nonparametric subspace dictionary learning (NSDL) in *image denoising tasks*.



(a) Noisy image ($\sigma = 25$) (b) Subspace dictionaries (c) Usage probability vs. Subspace index (d) Denoised Image

Comparisons of denoising PSNR as a function of noise deviation:

σ	Image	Barbara	Lena	Boats	House	Peppers
10	TV	29.77	32.71	31.64	33.76	32.40
	K-SVD	33.96	34.87	33.13	35.43	32.99
	BPFA	34.32	35.37	33.54	35.81	34.15
	NSDL	**34.50**	**35.57**	**33.70**	**35.98**	**34.31**
15	TV	27.49	30.96	29.79	31.89	30.44
	K-SVD	31.72	33.01	31.38	33.56	31.25
	BPFA	32.40	33.58	31.71	34.16	32.14
	NSDL	**32.69**	**33.84**	**31.96**	**34.45**	**32.44**
20	TV	26.01	29.84	28.50	30.76	29.25
	K-SVD	30.16	31.53	29.87	32.61	29.53
	BPFA	30.95	32.27	30.39	33.16	30.83
	NSDL	**31.17**	**32.46**	**30.74**	**33.47**	**31.18**
25	TV	25.07	28.87	27.57	29.96	28.26
	K-SVD	28.80	30.48	28.91	31.51	28.35
	BPFA	29.71	31.28	29.36	32.01	29.72
	NSDL	**30.03**	**31.45**	**30.12**	**32.33**	**30.09**

(1) Learned atoms of different subspace dictionaries are *grouped by textures*;
(2) *Dictionary number and dimension* are inferred;
(3) The proposed method provides improvement in the *image denoising task*.