VARIATIONAL BLIND SOURCE SEPARATION TOOLBOX AND ITS APPLICATION TO HYPERSPECTRAL IMAGE DATA

Ondřej Tichý and Václav Šmídl

Institute of Information Theory and Automation,
Czech Academy of Sciences,
Pod Vodarenskou vezi 4, Prague 8, Czech republic,
otichy@utia.cas.cz, smidl@utia.cas.cz

ABSTRACT

The task of blind source separation (BSS) is to decompose sources that are observed only via their linear combination with unknown weights. The separation is possible when additional assumptions on the initial sources are given. Different assumptions yield different separation algorithms. Since we are primarily concerned with noisy observations, we follow the Variational Bayes approach and define noise properties and assumptions on the sources by prior probability distributions. Due to properties of the Variational Bayes algorithm, the resulting inference algorithm is very similar for many different source assumptions. This allows us to build a modular toolbox, where it is easy to code different assumptions as different modules. By using different modules, we obtain different BSS algorithms. The potential of this open-source toolbox is demonstrated on separation of hyperspectral image data. The MATLAB implementation of the toolbox is available for download.

***Index Terms*—** Blind Source Separation, Variational Bayes Method, Sparse Prior, Hyperspectral Image

1. INTRODUCTION

The task of blind source separation (BSS) is to recover original signal sources that are observed only via their superposition with unknown weights. Typical examples of this task with audio data is the famous cocktail party problem [7] where many speaker are recorded by many microphones. This task is also common in image decomposition, where the same source images are observed via their weighted sum with dynamically changing weights [11]. The sources, may also have the role of low rank representation of high dimensional signal. This is useful e.g. in hyperspectral imaging [1, 2] where the same region is acquired by a sensor with a large number of spectral bands. The analysis of hyperspectral image data aims at identifying a small number of materials depending on the spectra. Noise level of different spectral images can vary significantly.

Many BSS algorithms were proposed in the signal processing literature, with various assumptions on the noise, source signal, mixing weights or the preselected number of sources. However, application of the general purpose algorithm on a particular task may not be straightforward. For example, the independent component analysis (ICA) is not suitable for hyperspectral imaging since its assumptions are not valid [8] here. Another issue of hyperspectral images is the variable quality of the spectral bands, which is not equal as assumed by many methods [1]. A potential solution is to assume sparsity of each source image and source spectrum.

In this paper, we propose a variational blind source separation algorithm with sparse priors on both source images and weighting coefficients of the spectral bands. Since these assumptions are common in blind source separation algorithms, we do not design a single method, but develop only a new module into a universal Variational BSS toolbox. This is possible due to the Variational Bayes (VB) methodology [9] which provides estimates of the model parameters in the form of marginal posterior probability densities. A unified algorithm can be derived for different representations of the image sources, image weights, and noise properties. These can be written as different modules that interact via variational message passing [14]. The number of potential combinations of different assumptions yields a wide range of BSS algorithms with different properties. In this paper, we focus on comparison of assumptions of isotropic priors [9] and sparse priors (via automatic relevance determination (ARD) principle) [3]. Each of these priors is a different module, which can be switched on or off for the image sources or the image weights. The Matlab implementation of the toolbox with these (and many other modules) is available for download.

The flexibility of the toolbox is demonstrated on data from hyperspectral imaging for hyperspectral unmixing [2]. The results of the Variational BSS toolbox with different choices of the prior is compared with the state of the art BSS algorithms such as the non-negative matrix factorization (NMF) [6] or non-negative projection algorithm (SNPA) [5].

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

2. GENERAL MODEL FOR BLIND SOURCE SEPARATION

In the image separation task, the observation data are considered to be vectors, $\mathbf{d}_j \in \mathbf{R}^{p\times 1}$, $j = 1,\dots n$, where n is the number of images and p is the number of pixels. The images are stored columnwise. The observation vector is assumed to be a result of superposition of r columnwise images, $\mathbf{a}_k \in \mathbf{R}^{p\times 1}$, $k = 1,\dots,r$, weighted by their specific weights in each recorded image, $x_{j,k}$,

$$\mathbf{d}_j = \sum_{k=1}^{r} \mathbf{a}_k x_{j,k} + \mathbf{e}_j, \qquad (1)$$

where recorded image $\mathbf{d}_j$ is corrupted by noise term $\mathbf{e}_j$ of the same size. Typically, the number of sources r is much lower than the number of images, n, or pixels, p.

The task of subsequent analysis is to estimate source images $\mathbf{a}_k$ and their weights $\mathbf{x}_k$, $k = 1,\dots,r$, from the data $D = [\mathbf{d}_1,\dots,\mathbf{d}_n]$. Since the problem (1) has infinitely many solutions, we make restrictive assumptions on model parameters using prior distributions.

2.1. Noise Prior

Here, we assume that all noise elements, $e_{i,j}$, $i = 1,\dots,p$, $j = 1,\dots,n$, are independent, identically distributed with unknown common variance ω^{-1} as follows

$$f(e_{i,j}|\omega) = \mathcal{N}_{e_{i,j}}(0,\omega^{-1}), \qquad (2)$$

where $\mathcal{N}(.,.)$ denotes normal distribution. This model is known as the isotropic Gaussian noise model [13]. The model is accompanied with the prior model of the precision parameter ω as

$$f(\omega) = \mathcal{G}_\omega(\vartheta_0,\rho_0), \qquad (3)$$

where $\mathcal{G}(.,.)$ denotes gamma distribution with selected prior parameters ϑ_0,ρ_0. The prior model of ω is graphically demonstrated in Figure 1, right.

2.2. Source Weights Priors

2.2.1. Isotropic Prior with Positivity

Here, we assume the prior model for each source weights vector $\mathbf{x}_k$, $k = 1,\dots,r$, as [7, 9]:

$$f(\mathbf{x}_k|\upsilon_k) = t\mathcal{N}_{\mathbf{x}_k}(\mathbf{0}_{n,1},\upsilon_k^{-1}I_n,[0,\infty]), \qquad (4)$$

where $\mathbf{0}_{n,1}$ denotes zero vector of the given size and I_n denotes identity matrix of the given size. The parameter υ_k models precision of the kth source. It is common for all observed data $\mathbf{d}_j$, hence we call this prior isotropic. The prior for the parameter υ_k is chosen as conjugate to (4) as $f(\upsilon_k) = \mathcal{G}_{\upsilon_k}(\alpha_0,\beta_0)$ where while α_0,β_0 are known prior parameters.

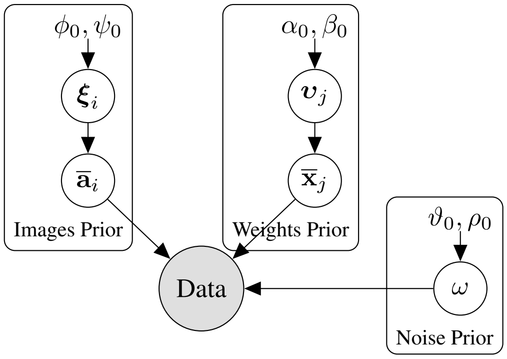

Fig. 1. Graphical model of blind source separation for a specific choice of priors: isotropic Gaussian noise and sparse priors for source images and spectral weights.

Positivity of the weights is enforced by truncated support of the prior. The effect of the truncation has impact on the moments of the distribution, see Appendix B. Without truncation to positive support, this kind of prior encourages orthogonal solution and leads to results similar to the principal component analysis [9].

2.2.2. Sparse Prior

An undesired effect of the positivity restriction on the isotropic prior is that the expected value of the posterior can not approach zero. This causes an undesired artifacts when the true weights are sparse. Better prior for sparse signals is based on the automatic relevance determination (ARD) principle [4, 12]. Specifically, prior on the source weight, $x_{j,k}$ is

$$f(x_{j,k}|\upsilon_{j,k}) = t\mathcal{N}_{x_{j,k}}(0,\upsilon_{j,k}^{-1},[0,\infty]), \qquad (5)$$

$$f(\upsilon_{j,k}) = \mathcal{G}_{\upsilon_{j,k}}(\alpha_0,\beta_0), \qquad (6)$$

where $\upsilon_{j,k}$ is unknown precision parameter to be estimated together with weight $x_{j,k}$ while α_0,β_0 are selected prior parameters. In effect, the prior allows for sparse solution, since variance of the zero terms also approaches zero. This prior model is graphically demonstrated in Figure 1, middle.

2.2.3. Other priors

When the weights have mutual correlations, it is possible to design more complex priors [11] with the same modular properties. However, this module will not be used for the hyperspectral data.

2.3. Source Images Priors

Equivalent priors can be designed for the image sources.

2.3.1. *Isotropic Prior with positivity*

Here, we assume the prior for each source image $\mathbf{a}_k$, $k = 1, \ldots, r$, as

$$f(\mathbf{a}_k|\xi_k) = t\mathcal{N}_{\mathbf{a}_k}(\mathbf{0}_{p,1}, \xi_k^{-1} I_p, [0,1]), \tag{7}$$

The parameter ξ_k is unknown prior precision of the kth source image in the model. The precision is common for all pixels in the image. The prior for unknown precision is $f(\xi_k) = \mathcal{G}_{\xi_k}(\phi_0, \psi_0)$. The truncation of this prior is between zero and one. This helps to reduce the scaling ambiguity of the multiplicative decomposition [9].

2.3.2. *Sparse Prior*

The source images can contain empty areas, where pixels should be equal to zero. This can be again achieved by the ARD prior. In the case of source images it is:

$$f(a_{i,k}|\xi_{i,k}) = t\mathcal{N}_{a_{i,k}}(0, \xi_{i,k}^{-1}, [0,1]), \tag{8}$$

$$f(\xi_{i,k}) = \mathcal{G}_{\xi_{i,k}}(\phi_0, \psi_0), \tag{9}$$

with selected prior parameters ϕ_0, ψ_0. The prior model of images is graphically demonstrated in Figure 1, left.

3. VARIATIONAL BAYES METHOD FOR BSS

Bayesian estimation of all unknown parameters of the BSS models from previous Section require evaluation of joint posterior densities. However, this is analytically intractable and an approximate evaluation is required in practice. We use the Variational Bayes (VB) approach [7, 9] which seeks the best posterior in the form of conditionally independent factors. For the elementary BSS model, where matrices $A = [\mathbf{a}_1, \ldots, \mathbf{a}_k]$ and $X = [\mathbf{x}_1, \ldots, \mathbf{x}_k]$, it would be:

$$f(A, X, \omega|D) \approx f(A|D) f(X|D) f(\omega|D). \tag{10}$$

The best approximation of this form in the sense of Kullback-Leibler divergence can be found analytically, and the result is known as the Variational Bayes [9] (or ensemble learning [7]). Following this methodology, the posterior distributions are

$$\begin{aligned} \tilde{f}(A|D) &\propto \exp\left(\mathrm{E}_{\tilde{f}(X|D)\tilde{f}(\omega|D)}\left[\ln\left(f(A, X, \omega, D)\right)\right]\right), \\ \tilde{f}(X|D) &\propto \exp\left(\mathrm{E}_{\tilde{f}(A|D)\tilde{f}(\omega|D)}\left[\ln\left(f(A, X, \omega, D)\right)\right]\right), \\ \tilde{f}(\omega|D) &\propto \exp\left(\mathrm{E}_{\tilde{f}(A|D)\tilde{f}(X|D)}\left[\ln\left(f(A, X, \omega, D)\right)\right]\right), \end{aligned} \tag{11}$$

where symbol $\propto$ means up to normalizing constant, $\mathrm{E}_f(.)$ means expected value of an argument with respect to distribution f. When the prior has hyper parameters (e.g. υ_k in (4)) it becomes an additional factor $\tilde{f}(\upsilon_k|D)$ in the conditional independence assumption (10) with analogical posterior to those in (11).

Algorithm 1 Variational BSS algorithm for general BSS model (1).

```
1. Initialization:
    (a) Set all prior parameters (subscripted by 0) to $10^{-10}$.
    (b) Set initial values for $\widehat{\overline{\mathbf{a}}_i}, \widehat{\overline{\mathbf{a}}_i^T\overline{\mathbf{a}}_i}$, for all $i$, and initial values of hyper-parameters on $A$.
    (c) Set initial values of $\widehat{\overline{\mathbf{x}}_j}, \widehat{\overline{\mathbf{x}}_j^T\overline{\mathbf{x}}_j}$ for all $j$, and initial values of hyper-parameters on $X$.
    (d) Set the initial number of sources $r_{max}$.
2. Iterate until convergence is reached using computation of shaping parameters (Appendix A) of:
    (a) Source images, $\tilde{f}(A|D)$, i.e. $\mu_{\overline{\mathbf{a}}_i}, \Sigma_{\overline{\mathbf{a}}_i}$, and their hyper-parameters for all $i = 1, \ldots, p$.
    (b) Source weights, $\tilde{f}(X|D)$, i.e. $\mu_{\overline{\mathbf{x}}_j}, \Sigma_{\overline{\mathbf{x}}_j}$ and their hyper-parameters for all $j = 1, \ldots, n$.
    (c) Noise parameters, $\tilde{f}(\omega|D)$, i.e. $\vartheta, \rho$.
3. Report estimates of source images and their weights.
```

The posterior distributions obtained using the VB method for the considered priors are:

$$\tilde{f}(\omega|D) = \mathcal{G}_\omega(\vartheta, \rho), \tag{12}$$

$$\tilde{f}(X|D) = \prod_{j=1}^{n} \tilde{f}(\overline{\mathbf{x}}_j|D) = \prod_{j=1}^{n} t\mathcal{N}_{\overline{\mathbf{x}}_j}(\mu_{\overline{\mathbf{x}}_j}, \Sigma_{\overline{\mathbf{x}}_j}, [0, \infty]), \tag{13}$$

$$\tilde{f}(A|D) = \prod_{i=1}^{p} \tilde{f}(\overline{\mathbf{a}}_i|D) = \prod_{i=1}^{p} t\mathcal{N}_{\overline{\mathbf{a}}_i}(\mu_{\overline{\mathbf{a}}_i}, \Sigma_{\overline{\mathbf{a}}_i}, [0,1]), \tag{14}$$

where shaping parameters $\vartheta, \rho, \mu_{\overline{\mathbf{x}}_j}, \Sigma_{\overline{\mathbf{x}}_j}, \mu_{\overline{\mathbf{a}}_i}, \Sigma_{\overline{\mathbf{a}}_i}$ and moments required to evaluate them are defined in Appendix A. Here, the bar symbol over a lower case letter denotes row vector of the matrix denoted by capital letter. The form of these posterior is common for all prior models in Section 2.

Specific variants of this algorithm arise for different prior models by addition of different hyper-parameters, e.g. ξ_k for isotropic model, and $\xi_{i,k}$ for the sparse model of images A. Note from Fig. 1, that these hyper-parameters influence only the posterior $\tilde{f}(A, D)$. Thus different modules for computation of $\tilde{f}(A, D)$ can be written with universal interface, following general algorithm given in Algorithm 1.

For example, module for the sparse prior on A, adds posterior $\tilde{f}(\xi_{i,k}|D) = \mathcal{G}_{\xi_{i,k}}(\phi_{i,k}, \psi_{i,k})$, with moments given in Appendix A. However, these are considered to be internal variables of the module evaluating posterior distribution $\tilde{f}(A|D)$. Replacement of the posterior by another module is thus almost trivial.

3.1. The Variational BSS Toolbox

The Variational BSS toolbox implements the basic Algorithm 1 for several modules. Both the isotropic and sparse priors are

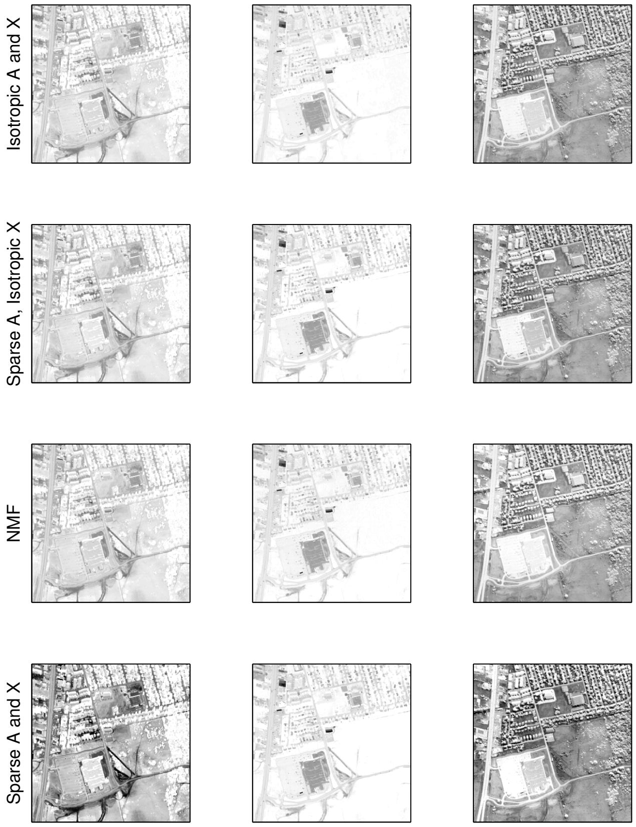

Fig. 2. Comparison of three most significant source image estimates from hyperspectral image data obtained by the compared methods. Each row contains images from one method.

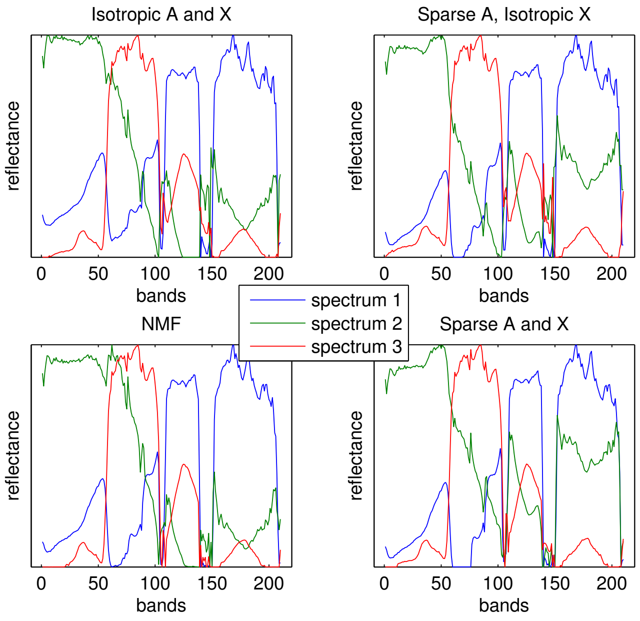

Fig. 3. Comparison of three most significant sources weights estimates from hyperspectral image data obtained by the compared methods.

supported, plus additional modules for more complex priors [10]. It is implemented in MATLAB and can be downloaded from `http://www.utia.cz/AS/softwaretools/image_sequences`. No additional toolboxes are required. The algorithm allows to choose:

- prior model of source images with positivity switched on or off,
- prior model of source weights including deconvolution models [11], with positivity on different quantities switched on or off,
- the maximum number of sources r_{max} , or heuristic for estimation of the number of relevant sources [11].

The toolbox is under open-source license and extensions of the toolbox are very welcome.

4. APPLICATION TO HYPERSPECTRAL IMAGE DATA

We demonstrate the flexibility of the proposed Variational BSS toolbox on hyperspectral image data. We will analyze the URBAN dataset[1] from hyper-spectral digital imagery collection experiment (HYDICE) where the number of spectral bands is $n = 210$ and the size of images is 307 × 307 pixels, $p = 307^2$. Note that this dataset is highly corrupted by noise, some of the bands do not contain any signal. We will apply 5 algorithms on this data: (i) isotropic priors of both, images and weights, i.e. Sections 2.2.1 and 2.3.1, (ii) sparse model of images while the model of weights remain isotropic, i.e. Sections 2.2.1 and 2.3.2, (iii) sparse model of both, images and weights, Sections 2.2.2 and 2.3.2. These algorithms will be compared with state of the arts algorithms: (iv) non-negative matrix factorization (NMF) [6] and (v) successive nonnegative projection algorithm (SNPA) [5]; however, since SNPA can not cope with noisy data, the noisy bands were removed manually to obtain comparable results.

All algorithms run with preselected number of sources to 6. The results are given in Figure 2 using 3 meaningful source images from each algorithm in each row while the other 3 source images are observation of noise presented in the original dataset. The resulting images are accompanied with spectra for each source and each algorithm given in Figure 3.

We conjecture that the best results are provided by our proposed algorithm with sparse priors on both, images and spectral weights. In estimated images, we reach better contrast in comparison with other methods. In estimated spectra, we achieve suppression of contribution from completely noisy bands 104–109 an 139–151, due to sparsity priors on the source weights.

[1] http://www.agc.army.mil/

5. DISCUSSION AND CONCLUSION

The variational Bayes (VB) approach to the blind source separation problem is studied in this paper. Since the problem is in general ill-posed, additional assumptions are required to obtain a solution. Here, these are formalized using probability distributions. We note that for many priors, the VB algorithm can be generalized to a common message passing algorithm. This algorithm has been implemented for several priors and is available in the form of Matlab toolbox. The number of combinations of possible assumptions in the toolbox is high, yielding algorithms that have not been tested yet.

Flexibility of the toolbox was demonstrated on hyperspectral image data from hyper-spectral digital imagery collection experiment (HYDICE). This data is corrupted by severe noise in many spectral bands. We show that these artifacts can be automatically suppressed by using sparsity prior on the source weights. This can be achieved using the sparsity module for weights of the toolbox. The results of the proposed method improve over the results of other state-of-the-art approaches.

The MATLAB toolbox is freely available for downloads under open-source license.

Acknowledgment This work was supported by the Czech Science Foundation, grant No. 13-29225S.

A. POSTERIOR ESTIMATES

The shaping parameters of posterior distributions (12)–(14) are derived as $\vartheta = \vartheta_0 + \frac{np}{2}$, $\rho = \rho_0 + \frac{1}{2}\mathrm{tr}\left(DD^T - \widehat{A}\widehat{X}^T D^T\right) + \frac{1}{2}\mathrm{tr}\left(-D\widehat{X}\widehat{A}^T\right) + \frac{1}{2}\mathrm{tr}\left(\widehat{A^T A}\widehat{X^T X}\right)$, $\Sigma_{\overline{\mathbf{a}}_i}^{-1} = \widehat{\omega}\sum_{j=1}^{n}(\widehat{\overline{\mathbf{x}}_j^T\overline{\mathbf{x}}_j}) + \mathrm{diag}(\widehat{\Xi}_i)$, $\mu_{\overline{\mathbf{a}}_i} = \Sigma_{\overline{\mathbf{a}}_i}\widehat{\omega}\sum_{j=1}^{n}(\widehat{\overline{\mathbf{x}}_j}d_{i,j})$, $\phi_i = \phi_0 + \frac{1}{2}\mathbf{1}_{r,1}$, $\psi_i = \psi_0 + \frac{1}{2}\mathrm{diag}(\widehat{\overline{\mathbf{a}}_i^T\overline{\mathbf{a}}_i})$, $\Sigma_{\overline{\mathbf{x}}_j}^{-1} = \widehat{\omega}\sum_{i=1}^{p}(\widehat{\overline{\mathbf{a}}_i^T\overline{\mathbf{a}}_i}) + \mathrm{diag}(\widehat{\upsilon_j})$, $\mu_{\overline{\mathbf{x}}_j} = \Sigma_{\overline{\mathbf{x}}_j}\widehat{\omega}\sum_{i=1}^{p}(\widehat{\overline{\mathbf{a}}}_i d_{i,j})$, $\alpha_j = \alpha_0 + \frac{1}{2}\mathbf{1}_{r,1}$, $\beta_j = \beta_0 + \frac{1}{2}\mathrm{diag}(\widehat{\overline{\mathbf{x}}_j^T\overline{\mathbf{x}}_j})$ with associate moments of standard distributions forms as $\widehat{\omega} = \frac{\vartheta}{\rho}$, $\widehat{\overline{\mathbf{a}}}_i = \mu_{\overline{\mathbf{a}}_i}$, $\widehat{\overline{\mathbf{a}}_i^T\overline{\mathbf{a}}_i} = \mu_{\overline{\mathbf{a}}_i}^T\mu_{\overline{\mathbf{a}}_i} + \Sigma_{\overline{\mathbf{a}}_i}$, $\widehat{\xi_{i,k}} = \phi_{i,k}\psi_{i,k}^{-1}$, $\widehat{\overline{\mathbf{x}}_j} = \mu_{\overline{\mathbf{x}}_j}$, $\widehat{\overline{\mathbf{x}}_j^T\overline{\mathbf{x}}_j} = \mu_{\overline{\mathbf{x}}_j}^T\mu_{\overline{\mathbf{x}}_j} + \Sigma_{\overline{\mathbf{x}}_j}$, $\widehat{\upsilon_{j,k}} = \alpha_{j,k}\beta_{j,k}^{-1}$.

B. TRUNCATED NORMAL DISTRIBUTION

Truncated normal distribution $t\mathcal{N}()$ of a scalar x on interval $[a;b]$ is $x \sim t\mathcal{N}(\mu,\sigma,[a,b]) = \frac{\sqrt{2}\exp((x-\mu)^2)}{\sqrt{\pi\sigma}(erf(\beta)-erf(\alpha))}\chi_{[a,b]}(x)$, where $\alpha = \frac{a-\mu}{\sqrt{2\sigma}}$, $\beta = \frac{b-\mu}{\sqrt{2\sigma}}$, function $\chi_{[a,b]}(x)$ is defined as $\chi_{[a,b]}(x) = 1$ if $x \in [a,b]$ and $\chi_{[a,b]}(x) = 0$ otherwise. $\mathrm{erf}()$ is the error function defined as $\mathrm{erf}(t) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-u^2}\mathrm{d}u$. The moments are $\widehat{x} = \mu - \sqrt{\sigma}\frac{\sqrt{2}[\exp(-\beta^2)-\exp(-\alpha^2)]}{\sqrt{\pi}(\mathrm{erf}(\beta)-\mathrm{erf}(\alpha))}$ and $\widehat{x^2} = \sigma + \mu\widehat{x} - \sqrt{\sigma}\frac{\sqrt{2}[b\exp(-\beta^2)-a\exp(-\alpha^2)]}{\sqrt{\pi}(\mathrm{erf}(\beta)-\mathrm{erf}(\alpha))}$.

REFERENCES

[1] N. Bali and A. Mohammad-Djafari. Bayesian approach with hidden Markov modeling and mean field approximation for hyperspectral data analysis. *Image Processing, IEEE Transactions on*, 17(2):217–225, 2008.

[2] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(2):354–379, 2012.

[3] C.M. Bishop. Variational principal components. *IET Conference Proceedings*, pages 509–514(5), 1999.

[4] C.M. Bishop and M.E. Tipping. Variational relevance vector machines. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 46–53, 2000.

[5] N. Gillis. Successive nonnegative projection algorithm for robust nonnegative blind source separation. *SIAM Journal on Imaging Sciences*, 7(2):1420–1450, 2014.

[6] D.D. Lee and H.S. Seung. Algorithms for non-negative matrix factorization. In *Advances in neural information processing systems*, pages 556–562, 2001.

[7] J.W. Miskin. *Ensemble learning for independent component analysis*. PhD thesis, University of Cambridge, 2000.

[8] J.M.P. Nascimento and J.M. Bioucas Dias. Does independent component analysis play a role in unmixing hyperspectral data? *Geoscience and Remote Sensing, IEEE Transactions on*, 43(1):175–187, 2005.

[9] V. Šmídl and A. Quinn. *The Variational Bayes Method in Signal Processing*. Springer, 2006.

[10] O. Tichý. *Bayesian Blind Source Separation in Dynamic Medical Imaging*. PhD thesis, Czech Technical University, Faculty of Nuclear Sciences and Physical Engineering, 2015.

[11] O. Tichý and V. Šmídl. Bayesian blind separation and deconvolution of dynamic image sequences using sparsity priors. *Medical Imaging, IEEE Transaction on*, 34(1):258–266, January 2015.

[12] M.E. Tipping. Sparse Bayesian learning and the relevance vector machine. *The journal of machine learning research*, 1:211–244, 2001.

[13] M.E. Tipping and C.M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):611–622, 1999.

[14] J. Winn and C.M. Bishop. Variational message passing. *The J. of Machine Learning Research*, 6:661–694, 2005.