Bayesian sampling of structured noise covariance matrix for hyperspectral imagery

Nicolas Dobigeon and Jean-Yves Tourneret
E-mail : {Nicolas.Dobigeon, Jean-Yves.Tourneret}@enseeiht.fr

TECHNICAL REPORT – 2008, December
University of Toulouse, IRIT/INP-ENSEEIHT
2 rue Camichel, BP 7122, 31071 Toulouse cedex 7, France

I. Linear mixing model

Consider P pixels of an hyperspectral image acquired in L spectral bands. According to the linear mixing model (LMM) described, for instance, in [1], the L-spectrum $\mathbf{y}_p = [y_{p,1}, \ldots, y_{p,L}]^T$ of the pth pixel ($p = 1, \ldots, P$) is assumed to be a linear combination of R spectra $\mathbf{m}_r$ corrupted by an additive Gaussian noise:

$$\mathbf{y}_p = \sum_{r=1}^{R} \mathbf{m}_r a_{p,r} + \mathbf{n}_p, \tag{1}$$

where $\mathbf{m}_r = [m_{r,1}, \ldots, m_{r,L}]^T$ denotes the spectrum of the rth material, $a_{p,r}$ is the fraction of the rth material in the pth observation, R is the number of materials, L is the number of available spectral bands and P is the number of observations (pixels). Considering all pixels, standard matrix notations yield:

$$\mathbf{Y} = \mathbf{M}\mathbf{A} + \mathbf{N}, \tag{2}$$

where

$$\begin{aligned} \mathbf{Y} &= [\mathbf{y}_1, \ldots, \mathbf{y}_P], \qquad \mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_R], \\ \mathbf{A} &= [\mathbf{a}_1, \ldots, \mathbf{a}_P], \qquad \mathbf{N} = [\mathbf{n}_1, \ldots, \mathbf{n}_P]. \end{aligned} \tag{3}$$

Moreover, in (1), $\mathbf{n}_p = [n_{p,1}, \ldots, n_{p,L}]^T$ is an additive noise sequence which is assumed to be a zero-mean Gaussian sequence with covariance matrix $\mathbf{\Sigma}_{\text{n}}$:

$$\mathbf{n}_p \sim \mathcal{N}(\mathbf{0}_L, \mathbf{\Sigma}_{\text{n}}). \tag{4}$$

In a lot of works (see [2]–[4] for instance), the noise covariance matrix $\mathbf{\Sigma}_{\mathrm{n}}$ has been assumed to be proportional to the identity matrix $\mathbf{\Sigma}_{\mathrm{n}} = \sigma^2 \mathbf{I}_L$, where $\mathbf{I}_L$ is the identity matrix of dimension $L \times L$, leading to an independent and identically distributed (i.i.d.) noise sequence. However this assumption may be not always ensured. As an example, for the NASA sensor boarded on the EO-1 satellite, $K = 3$ different spectral ranges ($550-700$nm, $700-1225$nm and $1225-2125$nm) with different SNRs have to be considered [5]. In this case, the noise covariance matrix can be decomposed as follows:

$$\mathbf{\Sigma}_{\mathrm{n}} = \begin{pmatrix} \sigma_1^2 \mathbf{I}_{L_1} & & & \\ & \sigma_2^2 \mathbf{I}_{L_2} & & \\ & & \ddots & \\ & & & \sigma_K^2 \mathbf{I}_{L_K} \end{pmatrix}, \tag{5}$$

where σ_k^2 and L_k are the noise level and the number of bands in the kth spectral range, respectively (with $L = \sum_{k=1}^{K} L_k$). In this report, we show how a Bayesian model as well as Gibbs sampling strategy that can be used to estimate the noise level in each spectral range.

II. Hierarchical Bayesian model

A. Likelihood

The linear mixing model defined in (1) and the statistical properties (4) of the noise vector $\mathbf{n}_p$ result in a conditionally Gaussian distribution for the observation of the pth pixel: $\mathbf{y}_p | \mathbf{\Sigma}_{\mathrm{n}} \sim \mathcal{N}(\mathbf{M}\mathbf{a}_p, \mathbf{\Sigma}_{\mathrm{n}})$. Therefore, the likelihood function of $\mathbf{y}_p$ can be expressed as:

$$f\left(\mathbf{y}_p \middle| \mathbf{\Sigma}_{\mathrm{n}}\right) = (2\pi)^{-\frac{L}{2}} \left(\det \mathbf{\Sigma}_{\mathrm{n}}\right)^{-\frac{1}{2}} \exp\left[-\frac{1}{2} \left\|\mathbf{y}_p - \mathbf{M}\mathbf{a}_p\right\|^2_{\mathbf{\Sigma}_{\mathrm{n}}^{-1}}\right], \tag{6}$$

where $\|\cdot\|_{\mathbf{P}}$ is the Euclidean norm associated to the $\mathbf{P}$-scalar product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{P}} = \mathbf{x}^T \mathbf{P} \mathbf{y}. \tag{7}$$

Assuming the independence between the noise sequences $\mathbf{n}_p$ ($p = 1, \ldots, P$), the likelihood function of all the observations $\mathbf{Y}$ is:

$$f\left(\mathbf{Y} \middle| \mathbf{\Sigma}_{\mathrm{n}}\right) = \prod_{p=1}^{P} f\left(\mathbf{y}_p \middle| \mathbf{\Sigma}_{\mathrm{n}}\right). \tag{8}$$

B. Noise variance prior

Conjugate prior distributions are chosen for each noise level σ_k^2 ($k = 1, \ldots, K$):

$$\sigma_k^2 \left| \nu_k, \gamma_k \right. \sim \mathcal{IG}\left(\frac{\nu_k}{2}, \frac{\gamma_k}{2}\right), \tag{9}$$

where $\mathcal{IG}\left(\frac{\nu_k}{2}, \frac{\gamma_k}{2}\right)$ denotes the inverse-gamma distribution with parameters $\frac{\nu_k}{2}$ and $\frac{\gamma_k}{2}$. As in previous works ([6] and [7]), the hyperparameter ν_k will be fixed to $\nu_k = 2$. On the other hand, γ_k will be random and adjustable hyperparameters, whose prior distributions are defined below. Assuming the parameters σ_k^2 ($p = 1, \ldots, K$) *a priori* independent, the joint prior distribution for $\boldsymbol{\sigma} = [\sigma_1^2, \ldots, \sigma_K^2]^T$ is:

$$f\left(\boldsymbol{\sigma}|\boldsymbol{\gamma}\right) = \prod_{k=1}^{K} f\left(\sigma_k^2|\gamma_k\right), \tag{10}$$

with $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]^T$

C. Prior distribution for hyperparameters γ_k

The prior for γ_k is a non-informative Jeffreys' prior [8] which reflects the lack of knowledge regarding this hyperparameter:

$$f\left(\gamma_k\right) \propto \frac{1}{\gamma_k} \mathbf{1}_{\mathbb{R}^+}\left(\gamma_k\right). \tag{11}$$

Assuming the parameters γ_k ($p = 1, \ldots, K$) *a priori* independent, the joint prior distribution for $\boldsymbol{\gamma}$ is:

$$f\left(\boldsymbol{\gamma}\right) = \prod_{k=1}^{K} f\left(\gamma_k\right). \tag{12}$$

D. Posterior distribution

The posterior distribution of the unknown parameter vector $\boldsymbol{\sigma}$ can be computed from marginalization using the following hierarchical structure:

$$f(\boldsymbol{\sigma}|\mathbf{Y}) = \int f(\boldsymbol{\sigma}, \boldsymbol{\gamma}|\mathbf{Y}) d\boldsymbol{\gamma} \propto f\left(\mathbf{Y}|\boldsymbol{\sigma}\right) \int \prod_{k=1}^{K} f\left(\sigma_k^2|\gamma_k\right) f\left(\gamma_k\right) d\gamma_k, \tag{13}$$

$f\left(\mathbf{Y}\middle|\boldsymbol{\sigma}\right)$ and $f\left(\gamma_k\right)$ are defined in (8) and (11) respectively. This hierarchical structure allows one to integrate out the hyperparameter vector $\boldsymbol{\gamma}$ from the joint distribution $f\left(\boldsymbol{\sigma}, \boldsymbol{\gamma}|\mathbf{Y}\right)$, yielding:

$$f\left(\boldsymbol{\sigma}|\mathbf{Y}\right) \propto \prod_{k=1}^{K} \left[\left(\frac{1}{\sigma_k^2}\right)^{\frac{PL_k}{2}+1} \exp\left(-\frac{1}{2\sigma_k^2} \sum_{p=1}^{P} \sum_{l \in \mathcal{L}_k} \epsilon_{l,p}^2\right) \right]. \tag{14}$$

where $e_{l,p}$ is the lth component of the vector $\mathbf{e}_p = \mathbf{y}_p - \mathbf{M}\mathbf{a}_p$. The subsets $\mathcal{L}_k$ define the band indexes in each spectral range, i.e., $\mathcal{L}_1 = \{1, \ldots, L_1\}$, $\mathcal{L}_2 = \{L_1 + 1, \ldots, L_1 + L_2\}$ and $\mathcal{L}_3 = \{L_1 + L_2 + 1, \ldots, L_1 + L_2 + L_3\}$ in the example above.

III. Gibbs sampling of the noise levels

Within a Gibbs sampling scheme such as the algorithm presented in [9], generating samples $\widetilde{\boldsymbol{\sigma}}^{(t)}$ according to the posterior distribution $f(\boldsymbol{\sigma}|\mathbf{Y})$ can be decomposed into 3 steps:

1) $\sigma_1^2|\mathbf{Y} \sim f(\sigma_1^2|\mathbf{Y})$,
2) $\sigma_2^2|\mathbf{Y} \sim f(\sigma_2^2|\mathbf{Y})$,
3) ...,
4) $\sigma_K^2|\mathbf{Y} \sim f(\sigma_K^2|\mathbf{Y})$.

Indeed, the conditional distributions of the noise levels in the K spectral ranges are the following inverse-Gamma distributions:

$$\sigma_k^2|\mathbf{Y} \sim \mathcal{IG}\left(\frac{PL_k}{2}, \frac{1}{2}\sum_{p=1}^{P}\sum_{l\in\mathcal{L}_k} e_{l,p}^2\right). \tag{15}$$

References

[1] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Processing Magazine*, pp. 44–57, Jan. 2002.

[2] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. and Remote Sensing*, vol. 32, no. 4, pp. 779–785, July 1994.

[3] C.-I Chang, "Further results on relationship between spectral unmixing and subspace projection," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, no. 3, pp. 1030–1032, May 1998.

[4] C.-I Chang, X.-L. Zhao, M. L. G. Althouse, and J. J. Pan, "Least squares subspace projection approach to mixed pixel classification for hyperspectral images," *IEEE Trans. Geosci. and Remote Sensing*, vol. 36, no. 3, pp. 898–912, May 1998.

[5] J. S. Pearlman, P. S. Barry, C. C. Segal, J. Shepanski, D. Beiso, and S. L. Carman, "Hyperion, a space-based imaging spectrometer," *IEEE Trans. Geosci. and Remote Sensing*, vol. 41, no. 6, pp. 1160–1173, June 2003.

[6] E. Punskaya, C. Andrieu, A. Doucet, and W. Fitzgerald, "Bayesian curve fitting using MCMC with applications to signal segmentation," *IEEE Trans. Signal Processing*, vol. 50, no. 3, pp. 747–758, March 2002.

[7] N. Dobigeon, J.-Y. Tourneret, and M. Davy, "Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach," *IEEE Trans. Signal Processing*, vol. 55, no. 4, pp. 1251–1263, April 2007.

[8] H. Jeffreys, *Theory of Probability*, 3rd ed. London: Oxford University Press, 1961.

[9] N. Dobigeon, J.-Y. Tourneret, and C.-I Chang, "Semi-supervised linear spectral unmixing using a hierarchical Bayesian model for hyperspectral imagery," *IEEE Trans. Signal Processing*, vol. 56, no. 7, pp. 2684–2695, July 2008.