CONSTRAINED INDEPENDENT COMPONENT ANALYSIS FOR HYPERSPECTRAL UNMIXING

Wei Xia, Bin Wang, and Liming Zhang

Department of Electronic Engineering, Fudan University
Shanghai 200433, China

1. INTRODUCTION

In order to analyze the hyperspectral data, it is necessary to unmix the mixed pixels into a collection of endmembers' spectra and their corresponding proportions [1]. Independent Component Analysis (ICA) [2] is a useful technique to decompose mixed data into independent components, but it assumes that the original sources are statistically independent, which is not reasonable for hyperspectral unmixing. Besides, ICA assumes that the distributing regularity of the source is stationary, but the regularity of nature signal is usually variable. Therefore, in this paper we present a novel algorithm called as Constrained Independent Component Analysis (CICA), which can overcome these limitations by proposing two improvements. Firstly, according to the characteristics of hyperspectral images, we layout an Adaptive Probability Model (APM), which is capable of describing more than one probability distributions adaptively. Secondly, Abundance Nonnegative Constraint (ANC) and Abundance Sum-to-one Constraint (ASC) are added to ICA, generating a Constrained ICA algorithm. The goal of this method is to minimize the mutual information of abundances, at the same time with ANC and ASC satisfied. It means that we solve the unmixing issue by minimizing uncorrelation instead of pursuing independence.

2. ICA FOR HYPERSPECTRAL UNMIXING

Considering the Linear Mixture Model (LMM) [1] $\mathbf{x} = \mathbf{As} + \mathbf{e}$, ICA can recovers abundance signal $\mathbf{s}$ by linear transformation $\mathbf{y} = \mathbf{Wx} = \mathbf{Us}$. $\mathbf{W}$ is gained by maximizing the independence between the components of $\mathbf{y}$. Usually, mutual information

$$I(\mathbf{y}) \overset{def}{=} D_{KL}\Big(p(\mathbf{y}) \,\|\, \prod_{i=1}^{P} p(y_i) \Big) = \int p(\mathbf{y}) \log\big(p(\mathbf{y}) / \prod_{i=1}^{P} p(y_i) \big) d\mathbf{y} \tag{1}$$

is a measure of independence. Minimizing $I(\mathbf{y})$ yield the learning approach [3]-[5]:

$$\mathbf{W} \leftarrow \mathbf{W} + \eta \Delta \mathbf{W}, \quad \Delta \mathbf{W} = (\mathbf{I} - \mathbf{f(y)}\mathbf{y}^T)\mathbf{W}, \tag{2}$$

where $\mathbf{f(y)} = [-p'(y_1)/p(y_1), \ldots, -p'(y_i)/p(y_i), \ldots, -p'(y_p)/p(y_p)]^{\mathrm{T}}$, η is update factor, and $p'(y_i)$ is the derivative of $p(y_i)$.

3. THE PROPOSED ALGORITHM

3.1. Adaptive Probability Model

$p(y_i)$ can be approximated by Pearson mixture model [6] : $p(y_i) = (1-a_i) N\left(\mu_{1(i)}, \sigma_i^2\right) + a_i N\left(\mu_{2(i)}, \sigma_i^2\right)$, where $N\left(\mu, \sigma^2\right)$ represents the normal PDF with mean μ and variance σ^2. We design an Adaptive Probability Model (APM) in which parameters a_i and $\mu_{1(i)}$ is automatically changed according to statistical characteristic of the data. Generally, the endmember is sparse since each material does not distribute over the whole imagery. For any endmembers' abundance signals, the probability of its equaling to zero is not very small, which can be considered as function $p(y_i)$ taking peak value when $y_i = 0$. So we set $\mu_{2(i)} = 0$. In addition, σ_i is not a critical parameter for APM, and it is set to 1/3 here. The learning rule of Eq. (2) becomes $\Delta \mathbf{W}\_APM = (\mathbf{I} + \mathbf{d}\mathbf{y}^T - 9\mathbf{y}\mathbf{y}^T)\mathbf{W}$, where $\mathbf{d} = 9\boldsymbol{\mu}_1 \cdot / \left[\mathbf{1} - \mathbf{a} \cdot /(\mathbf{a}-\mathbf{1}) \cdot * \exp(-9\boldsymbol{\mu}_1 \cdot * \mathbf{y} + 4.5\boldsymbol{\mu}_1 \cdot * \boldsymbol{\mu}_1)\right]$, $\mathbf{I}$ is the identity matrix, $\boldsymbol{\mu}_1 = [\mu_{1(1)}, \ldots, \mu_{1(i)}, \ldots, \mu_{1(P)}]^{\mathrm{T}}$, $\mathbf{1}$ is a vector of ones, and $\mathbf{a} = \left[a_1, \ldots, a_i, \ldots, a_P\right]^{\mathrm{T}}$. Now we determine a_i and $\mu_{1(i)}$ by analyzing the cumulants of the data. the mean, skewness and kurtosis of y can be computed by

$$mean(y_i) = (1-a_i)\mu_{1(i)}, \quad skewness(y_i) = (1-a_i)(2a_i - 1)\mu_{1(i)}^3 / \kappa_2^{3/2}, \quad kurtosis(y_i) = (1-a_i)(1-a_i)\mu_{1(i)}^4 (6a_i^2 - 6a_i + 1) / \kappa_2^{3/2}, \tag{3}$$

where $\kappa_2 = \left(a(1-a)\mu_1^2 + \sigma^2\right)$. Therefore, $kurtosis(y_i) < 0$ if $a_i \in \left(0.5 - \sqrt{3}/6,\ 0.5 + \sqrt{3}/6\right)$, and $kurtosis(y_i) > 0$ if $a_i \in (0,\ 0.5 - \sqrt{3}/6)\ or\ (0.5 + \sqrt{3}/6,\ 1)$. So we calculate $kurtosis(y_i)$, and set a_i =0.6443 if $kurtosis(y_i) < 0$, or set a_i =0.8943 if $kurtosis(y_i) > 0$. After a_i is obtained, we calculate $mean(y_i)$, and set $\mu_{1(i)} = mean(y_i)/(1-a_i)$.

3.2. Abundance Nonnegative Constraint

In order to meet ANC, we design an ANC object function for the estimated abundance signal **y** of any pixel

$$J\_ANC(\mathbf{y}) = \sum_{i=1}^{P} \left(f(y_i) + \left|f(y_i)\right|\right)/2, \tag{4}$$

where $f(x)$ may be any function which can satisfy the qualification: $f(x) = 0\ if\ x \in [0,\ 1]$ and $f(x) > 0\ if\ x \notin [0,\ 1]$. Here we choose $f(x) = \left((x-0.5)^{2b} - 0.5^{2b}\right)/2b, b \in \mathbb{N}_+$. If $b = 1$, calculating the natural gradient of $J\_ANC(\mathbf{y})$ yields:

$$\Delta\mathbf{W}\_ANC = -\left(\partial J\_ANC(\mathbf{y})/\partial\mathbf{W}\right)\mathbf{W}^{\mathrm{T}}\mathbf{W} = -\mathbf{m}\mathbf{x}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}, \tag{5}$$

where the ith element of $\mathbf{m}$ is $m_i = (y_i - 0.5)\ if\ y_i \notin [0,\ 1]$, or $m_i = 0$ otherwise.

3.3. Abundance Sum-to-one Constraint

We give ASC objective function as

$$J\_ASC(\mathbf{y}) = \left(\sum_{i=1}^{P} y_i - 1\right)^{2c} / 2c\,, \tag{6}$$

where $c \in \mathbb{N}_+$. We can choose $c = 1$, and the ASC update rule of **W** can be given under the gradient ascent rule

$$\Delta\mathbf{W}\_ASC = -\left(\partial J\_ASC(\mathbf{y})/\partial\mathbf{W}\right)\mathbf{W}^{\mathrm{T}}\mathbf{W} = -\mathbf{1}\left(\sum_{i=1}^{P} y_i - 1\right)\mathbf{x}^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}\mathbf{W}\,. \tag{7}$$

3.4. The procedure of Constrained ICA

The whole objective function of CICA is $J(\mathbf{y}) = I(\mathbf{y}) + \eta_1 J\_ANC(\mathbf{y}) + \eta_2 J\_ASC(\mathbf{y})$. Based on the above analysis, the update rule of **W** can be expressed as

$$\Delta\mathbf{W} = \Delta\mathbf{W}\_APM + \eta_1\Delta\mathbf{W}\_ANC + \eta_2\Delta\mathbf{W}\_ASC\,, \tag{8}$$

After obtaining **W** by Eq. (8), the abundance estimation is solved by **y=Wx**.

4. CONCLUSION

In this paper, we proposed a new approach based on ICA for hyperspectral unmixing. Generally, ICA cannot solve the unmixing problem well since the independence assumption violates ANC and ASC. Our algorithm overcomes this drawback by considering not only the independence of sources, but also the ANC and ASC of abundance. Standard ICA algorithms only consider independence as the object function, but the CICA performs otherwise. The CICA includes ANC and ASC constraint to accord with the hyperspectral-data reality, and it performs under a probability mode which is adaptive. The adaptive probability model is reliable and reasonable because it is designed based on the sparseness property of hyperspectral imagery. The experimental results on simulated images and real hyperspectral data show that the proposed method gives good results. Further, we should point out that the proposed CICA can separate different endmembers using very few iterations.

5. REFERENCES

[1] C.-I Chang, *Hyperspectral imaging: techniques for spectral detection and classification*. New York: Plenum, 2003.

[2] J. Nascimento and J. Bioucas-Dias, "Does independent component analysis play a role in unmixing hyperspectral data?" *IEEE Transactions on Geoscience and Remote Sensing*, vol.43, no.4, pp.175-187, April, 2005.

[3] T. Lee, M. Girolami, & T. Sejnowski, "Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources," *Neural Computation*, 11,417-441, 1999.

[4] J. Bell，Terrence J. Sejnowski. "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, No. 6, pp. 1129-1159, November 1995.

[5] Shun-Ichi Amari, "Natural gradient works efficiently in learning," *Neural Computation*, v.10 n.2, pp. 251-276, Feb. 15, 1998.

[6] K. Pearson, *Contributions to the theory of mathematical evolution*. Phil. Trans. R. Soc. Lond., 185, 71–110, 1894.