Noise-adjusted non orthogonal linear projections for hyperspectral data analysis

M. Lennon[1,2], G. Mercier[1]

[1]École Nationale Supérieure des Télécommunications de Bretagne - Département ITI
Technopôle de Brest-Iroise - BP 832 - 29285 Brest Cedex - France

[2]SARL AvelMor - Technopôle de Brest-Iroise - Place Nicolas Copernic - 29280 Plouzané - France

Tel: (33) (0)2 29 00 10 61; Fax: (33) (0)2 29 00 10 98; Email: marc.lennon@enst-bretagne.fr

Abstract— **Independent Component Analysis (ICA) and Projection Pursuit (PP) are non orthogonal linear projection methods useful for dimensionality reduction of hyperspectral data cubes, in many cases more interesting than the standard Principal Component Analysis (PCA) but unfortunately not very robust to the noise. In this paper, the spatial correlation information is taken into account in order to improve the performances of both methods, following the ideas behind the so-called Noise-Adjusted Principal Component Analysis (NAPCA). This leads to the construction of two robust non orthogonal linear projection methods, respectively called Noise-Adjusted Independant Component Analysis (NAICA) and Noise-adjusted Projection Pursuit (NAPP).**

I. INTRODUCTION

In order to analyze data such as high dimensionality hyperspectral images, a dimensionality reduction procedure is usually used as a preprocessing step in order to reduce the dimension of the support vectorial space, thus increasing the capabilities of pattern recognition algorithms in a lower dimension subspace [1].

Principal Component Analysis (PCA) is one of the techniques commonly used for this purpose. PCA is however only optimal for a multidimensional gaussian structure, which usually does not follows the structure of real multispectral or hyperspectral data. Independant Component Analysis (ICA) [2] and Projection Pursuit (PP) [3] are both techniques which have been introduced for increasing the capacities of PCA in the dimensionality reduction procedure of hyperspectral images [4] [5].

Both techniques are attractive, mainly because they allow the constraint of orthogonality of PCA to be relaxed, thus preserving a maximum amount of information in the projection subspace. However, ICA and PP are very sensitive to the noise, generally leading to noisier transformed components than the components computed from a PCA on real data.

Following the ideas of the so called "Noise-adjusted Principal Component Analysis" (NAPCA) [6] (or equivalent "Minimum Noise Fraction Transform" MNF [7]), we propose to adjust both ICA and PP to the noise by first separating the signal and noise subspaces thanks to the spatial structure of the signal, then performing the transformations in the signal subspace. The process leads to the signal to noise ratio (SNR) of the computed components to be largely increased in the projection subspace. Orthogonality constraint of PCA is then relaxed, while robustness to the noise is ensured.

II. ORIGINAL NON ORTHOGONAL LINEAR PROJECTIONS

A. Independent Component Analysis

ICA is a multivariate data analysis process designed by signal processing research teams in the field of blind source separation [2]. It consists in projecting random vectors in a space where the new components are statistically independent, while PCA restricts the projected components to be uncorrelated. ICA, based on the concept of statistical independence which takes into account higher order moments, thus has stronger statistical properties than PCA, which only takes into account the second order statistics of the data. The method is based on the optimization of an independence criterion between the observed signals and leads to a non orthogonal linear projection of the data.

The model of ICA states as: $\mathbf{x} = A\mathbf{s}$; with $\mathbf{x}$: vector of observed signals; A: scalar matrix of mixing coefficients; $\mathbf{s}$: vector of source signals.

The goal of ICA is to find the linear transformation $W = A^{-1}$ so that the sources $\hat{\mathbf{s}} = W\mathbf{x}$ can be estimated from the observed vector $\mathbf{x}$ by optimizing a statistical independence criterion.

It should be noticed that a strong relationship exists between statistical independence and non gaussianity. The Central Limit Theorem states that the sum of N independent random variables tends in law towards a gaussian distribution when N tends to infinity. The corrolary of this theorem states that the sum of two independent random variables has a "more gaussian" distribution than any of the initial random variables. The estimation of the independent components thus consists in finding the weights of the matrix W so that the vector $\hat{\mathbf{s}} = W\mathbf{x}$ has the "least gaussian" components. Hence, the computing strategy consists in finding projections by optimizing a practical gaussianity measure.

Fast-ICA algorithm [8] has been adapted to hyperspectral data analysis and used in our simulations [4].

B. Projection Pursuit

PP is multivariate data analysis process originally designed with the aim to vizualise and understand a high dimensional data structure in a space of dimension lower than 3 [3]. It consists in finding "interesting projections" according to a quantitative index to be defined. Following the corollary of the Central Limit Theorem which states that any linear combination of random variables leads to a more gaussian distribution than the original ones, the projections revealing the least gaussian distributions have been shown to be the most interesting ones [9]. It should be noticed that PP is closely related to ICA which also searches for least gaussian projections. However, a major difference exists in the sense that ICA simultaneously searches for the directions where all the projected components are the "most independent" and hence does not privilegiate any direction among others. In opposition, PP sequentially searches for monodimensional least gaussian projections. Once a direction is found, data are projected onto the subspace orthogonal to the latter and so on. PP thus has ordering capabilities, while ICA does not. As for ICA, Fast-ICA algorithm [8], has been used in our simulations. However, unless Fast-ICA used for ICA, Fast-ICA used for PP sequentially finds monodimensional projections by using a deflation scheme instead of a global optimization scheme.

III. Noise-Adjusted projections

A. Noise-Adjusted Independent Component Analysis

Following the idea behind NAPCA [7], we propose to add a noise component $\mathbf{N}(\mathbf{x})$ uncorrelated with the signal component $\mathbf{S}(\mathbf{x})$ to form the n-dimensional image model $\mathbf{I}(\mathbf{x})$ ($\mathbf{x}$ represents the 2-D spatial position vector):

$$\mathbf{I}(\mathbf{x}) = A\mathbf{S}(\mathbf{x}) + \mathbf{N}(\mathbf{x}) \text{ with : } \mathbf{I}^T(\mathbf{x}) = \{I_1(\mathbf{x}), ..., I_n(\mathbf{x})\}$$

The noise covariance matrix needs to be estimated and is then diagonalized in order to compute the transformation matrix allowing the noise to be spectrally uncorrelated with unit variance (identity noise covariance matrix). This transformation is applied to the data, to adjust the data to the noise. ICA being equivalent to searching for privilegiate directions in the joint distributions space, this preprocessing step precisely consists in removing the directions induced by the noise. Noise being now spherical, only the signal subspace contributes to the search of the projection directions by ICA. The practical method rises from those concepts:
1. Estimation of the noise covariance matrix $R_{\mathbf{N}}$. If we assume that spatial autocorrelation of the signal is high compared to the noise's one, the noise component $\mathbf{N}$ can be estimated by computing the mean of the differences in the horizontal and vertical neighbours in the original image $\mathbf{I}$:

$$\mathbf{N}_{i,j} = \frac{1}{2}((\mathbf{I}_{i,j} - \mathbf{I}_{i,j+1}) + (\mathbf{I}_{i,j} - \mathbf{I}_{i+1,j}))$$

From $\mathbf{N}_{i,j}$, the covariance matrix $R_{\mathbf{N}}$ is computed. This estimator is very simple and should be computed on homogeneous regions of the image so that the assumption of strong signal autocorrelation with respect to the noise's one is respected. More robust noise estimators can also be used.
2. Diagonalization of $R_{\mathbf{N}}$. The eigenvectors matrix E and the eigenvalues matrix $\Delta_{\mathbf{N}}$ of $R_{\mathbf{N}}$ are computed so that:

$$E^T R_{\mathbf{N}} E = \Delta_{\mathbf{N}}$$

3. Computation of the transformation matrix F :

$$F = E\Delta_{\mathbf{N}}^{-1/2}$$

with $F^T R_{\mathbf{N}} F = Id$ and $F^T F = \Delta_{\mathbf{N}}^{-1}$ (Id : Identity matrix)
4. Noise-Adjustment of the data:

$$\mathbf{I}_{adj} = F^T(\mathbf{I} - \mathbf{m}) \quad \text{with:} \quad \mathbf{m} : \text{mean vector of } \mathbf{I}$$

5. ICA on the noise-adjusted data: $\mathbf{I}_{NAICA} = W\mathbf{I}_{adj}$ with W: transformation matrix from ICA.
It should be noticed that if the dimension is reduced, NAICA is equivalent to remove the noise subspace, and to carry out an ICA in the signal subspace. This shows the potential robustness of the method to the noise.

B. Noise-Adjusted Projection Pursuit

The same image model as the one used in NAICA, including an additive noise component uncorrelated with the signal component, is used in NAPP. The strategy is also the same: data are first noise-adjusted thanks to the diagonalization of the noise covariance matrix. Least gaussian directions are then searched by PP in the noise-adjusted data. As for NAICA, noise being spherical, only the signal subspace contributes to the search of the projection directions by PP. The practical method rises from those concepts:
1. Estimation of the noise covariance matrix $R_{\mathbf{N}}$.
2. Diagonalization of $R_{\mathbf{N}}$ and computation of the transformation matrix F.
3. Noise-Adjustment of the data: $\mathbf{I}_{adj} = F^T(\mathbf{I} - \mathbf{m})$
4. PP on the noise-adjusted data: $\mathbf{I}_{NAPP} = W\mathbf{I}_{adj}$ with W: transformation matrix from PP.
It should be noticed that if the dimension is reduced and as for NAICA, NAPP is equivalent to remove the noise subspace, and to carry out a PP in the signal subspace. This shows the potential robustness of the method to the noise.

IV. Qualitative comparison of the methods

A 2 meters ground resolution hyperspectral image from the airborne CASI sensor including 17 spectral bands from 445 to 895 nm has been used for simulation purposes (false-colour RGB composition on Fig. 1a). In order to test the capabilities of the methods, a multidimensional, gaussian (centered, relative large variance compared with the dynamics of the image), spatially white and spectrally correlated noise has been added to the image (false-colour RGB composition on Fig. 1b). False-colour RGB compositions of 3-dimensional projections from the 4 methods (ICA, NAICA, PP, NAPP) are shown on Fig. 1c to Fig. 1f. It should be noticed that both ICA and PP are very sensitive to the noise and that SNR of images from NAICA and NAPP are largely increased.

V. Quantitative comparison of the methods

Performances of projections are evaluated thanks to a statistical supervised bayesian (gaussian) classification carried out in 2 to 10-dimensional projection subspaces. The same original image as in the previous paragraph is used (without the noise). 10 classes are defined from ground truth (GT) and 3 sets of different training samples are selected: 10 samples per class, 20 samples per class, totality of the samples available. Fig. 2 shows the evolution of the classification rates after each method. The Hughes phenomenon [1] strongly appears: if the training samples data set is too small, the classification rate falls when the dimension of the subspace increases; The classification rate presents an asymptotic behavior only when the training samples data set is significant with respect to the dimension of the space. ICA and PP lead to low classification rates, mainly because of the noisy projected components. Noise-adjustment contributes to increase largely the classification rates. It should be noted that ICA and PP, even if they lead to different projected components, have the same joint probability density function, except for a rotation. In the case of the bayesian classifier used here, they thus lead to the same classification results. The same applies to the pair constituted by NAICA and NAPP.

VI. Conclusions

Noise-adjustment of non orthogonal linear projection methods has been proposed. Comparisons showed that NAICA and NAPP lead to higher SNR new components than original ICA and PP, and are moreover able to capture non orthogonal structures in a hyperspectral data cube, unless PCA ou NAPCA. They are thus very useful for many further image processing tasks.

Acknowledgments

Authors would like to thank Dr. Laurence Hubert-Moy from COSTEL for her generous offer of the CASI data.

References

[1] D. Landgrebe, "Information extraction principles and methods for multispectral and hyperspectral image data," in *Information processing for remote sensing*, C.H. Chen, Ed., pp. 3–38. World Scientific Publishig Co. Inc., USA, 1999.

[2] C. Jutten and J. Hrault, "Blind separation of sources, part i: an adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1–10, 1991.

[3] J.H. Friedman and J.W. Tuckey, "A projection pursuit algorithm for exploratory data analysis," *IEEE Transactions on Computers*, vol. c-23, no. 9, pp. 881–889, 1974.

[4] M. Lennon, G. Mercier, M.C. Mouchot, and L. Hubert-Moy, "Independent component analysis as a tool for the dimensionality reduction and the representation of hyperspectral images," in *Proceedings of IGARSS 2001, Sydney, Australia*, 2001.

[5] A. Ifarraguerri and C.-I. Chang, "Unsupervised hyperspectral image analysis with projection pursuit," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 6, pp. 2529–2538, 2000.

[6] J.B. Lee, S. Woodyatt, and M. Berman, "Enhancement of high spectral resolution remote-sensing data by a noise-adjusted principal component transform," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295–304, 1990.

[7] A.A. Green, M. Berman, P. Switzer, and M.D. Craig, "A tranformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, no. 1, pp. 65–74, 1988.

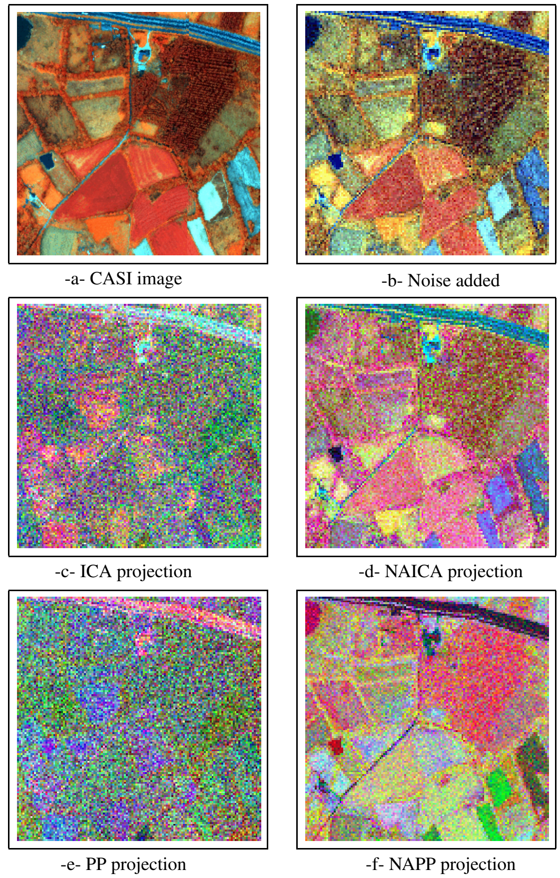

Fig.1 Qualititative comparison of the projection methods

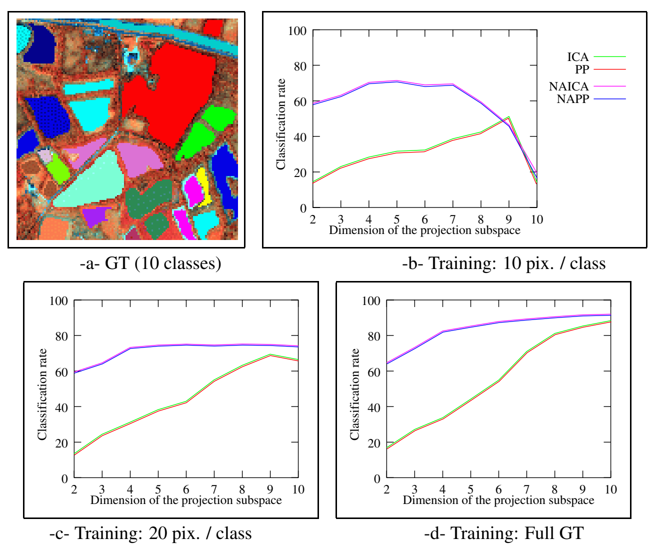

Fig. 2. Quantitative comparison of the projection methods

[8] A. Hyvrinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[9] J.H. Friedman, "Exploratory projection poursuit," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 249–266, 1987.

0-7803-7930-6/$17.00 (C) 2003 IEEE