BLIND SIGNAL SEPARATION WITH A PROJECTION PURSUIT INDEX

A. Sarajedini, P.M. Chau
Department of Electrical and Computer Engineering
University of California, San Diego
La Jolla, California, 92093, USA

ABSTRACT

Blind Signal Separation (BSS) is a powerful technique for separation of mixed signals with weak assumptions on the incoming signals. The objectives of BSS are analogous to the objectives of Exploratory Projection Pursuit which is widely used in the statistical community for finding structure in high dimensional data sets. In this paper, we adapt Exploratory Projection pursuit for BSS. First, we introduce Exploratory Projection Pursuit and the associated projection pursuit index (PPI). We adapt the PPI for application to BSS. We also investigate the order of approximation required to achieve satisfactory separation using the PPI, and compare its performance to a maximum-likelihood BSS technique using a Gram-Charlier Expansion.

INTRODUCTION

Signal Model

We make the standard assumptions on the signal model for bind signal separation, i.e. independently distributed signals are linearly mixed by a mixing matrix A:

$$x = As + n \qquad (1)$$

where n is AWGN. We assume real signals, real noise, and a real mixing matrix. We will vary the noise in our simulations, but we ignore the noise for the purposes of analysis.

Exploratory Projection Pursuit

We will give a brief introduction to Exploratory Projection Pursuit. Full details of the algorithm are discussed in [4]. Projection pursuit is a statistical method for function approximation (and density estimation) using a representation as a product of one dimensional functions:

$$f(x) = \prod_{n=1}^{N} g(\omega_n^T x) \qquad (2)$$

As with blind signal separation, projection pursuit first does a sphering of the data. This consists of determining the sample covariance matrix of the incoming data, the SVD of the covariance matrix into VDV^T where V is the matrix of eigenvectors and D is the diagonal matrix of eigenvalues. When these are applied to the incoming data, X (which is here assumed to have mean zero):

$$Z = D^{-\frac{1}{2}} V^T X \qquad (3)$$

the resulting data Z has unit variance in all projection directions. As shown by Harroy-Lacoume [5], this recovers the column space eigenvectors and eigenvalues of the mixing matrix A, but the matrix of row space eigenvectors (a unitary matrix) cannot be recovered from the correlation matrix.

The criterion projection pursuit uses to find approximation directions (‘interesting’ directions) is the distance of the one dimensional density estimate in a particular direction from the Gaussian density. In previous methods [3], this distance from Gaussianity was measured using a Kullback-Leibler divergence which amounted to maximum-likelihood density estimation. The PPI measures distance from Gaussianity in a different way: by passing the data through an error function transformation and estimating the resulting data density using Legendre polynomials. Because the Legendre polynomials are only applicable over the range [-1,1] and to remove any multiplicative Gaussian factor in the density, the data is passed through an error function nonlinearity prior to being estimated:

$$R = 2\phi(\alpha^T Z) - 1 \qquad (4)$$

where ϕ is the error function (Gaussian cumulative distribution function) and α is a vector along which the data density is estimated. The ϕ function forces all the data in the range [-∞, ∞] to lie in the range [0,1], and divides the data density by a Gaussian density thus removing any Gaussian component of the data density. The index used to determine interesting directions is the following:

$$I(\alpha) = \tfrac{1}{2} \sum_{j=1}^{J} (2j+1) E_R^2 (P_j(R)) \qquad (5)$$

where P_j is the jth Legendre polynomial and J is the number of terms in the Legendre expansion of the density. The motivation for using the Legendre polynomials lies in their ease of computation. The recursion equations for the jth Legendre polynomial are given by:

$$
\begin{aligned}
P_0(R) &= 1 \\
P_1(R) &= R \\
P_j(R) &= \frac{1}{j}((2j-1)RP_{j-1}(R) - (j-1)P_{j-2}(R))
\end{aligned} \tag{6}
$$

The equations for the derivative are also straightforward if one wanted to do gradient descent. Friedman uses gradient descent as well as coarse search to find 'interesting directions'. However, we found gradient descent to be problematic for BSS and avoided its use altogether. The reasons are discussed in the Body of this paper.

Blind signal separation

Blind signal separation is the recovery of signals that have been linearly combined with a mixing matrix A as in the signal model above. A is assumed nonsingular. As in projection pursuit, the signal model is often modified by taking the correlation matrix of the incoming data, determining its SVD, and projecting the data onto the signal subspace and scaling so that the resulting data has unit variance in all directions. The SVD of the mixing matrix can be recovered in this way except for a unitary matrix, producing the following model of the projected and scaled incoming data:

$$
\begin{aligned}
z &= Us + Bn \\
B &= D^{-\frac{1}{2}}V^T
\end{aligned} \tag{7}
$$

where U is a unitary matrix. Standard methods for recovering the unitary transformation are based on density estimation using cumulants [3,5], cumulant matching [2], density estimation using kernels [7], temporal processing with joint diagonalization of correlation matrices [1]. The density estimation methods assume i.i.d. sources and write the joint pdf as a product of one dimensional pdf's. The one dimensional pdf's are estimated as linear combinations of Hermite polynomials where the polynomial coefficients are cumulants of the data.

PROJECTION PURSUIT INDEX FOR BSS

Modifications of PPI for BSS

Since we know the projection directions of the sources are orthogonal, and each source direction should correspond to an 'interesting' direction in the PPI sense, we modify the PPI to include both directions simultaneously. We restrict ourselves to 2 independently distributed real input signals and a real mixing matrices so that the unitary matrix may be parametrized as the orthonormal transformation in the following way:

$$
U = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \tag{8}
$$

We modify the PPI to the sum of the projection pursuit index in orthonormal directions:

$$
\begin{aligned}
I(\alpha) &= \tfrac{1}{2}\sum_{j=1}^{J}(2j+1)E_{R_1}^2(P_j(R_1)) \\
&\quad + \tfrac{1}{2}\sum_{j=1}^{J}(2j+1)E_{R_2}^2(P_j(R_2)) \\
R_1 &= 2\phi(\cos(\theta)z_1 - \sin(\theta)z_2) - 1 \\
R_2 &= 2\phi(\sin(\theta)z_1 + \cos(\theta)z_2) - 1
\end{aligned} \tag{9}
$$

Since the 1-d densities in the orthogonal directions are both farthest away from Gaussian (relative to other pairs of orthogonal directions), this composite index should be more robust than the PPI in either projection direction alone.

Interpretation of PPI BSS relative to Edgeworth expansion

The application of the Gaussian transformation to the data in each projection direction improves robustness to outliers, forces the data into the range [-1,1] so that the Legendre polynomials may be used, and divides the output density by a Gaussian density causing the estimate to approach the Edgeworth expansion. The Edgeworth expansion is the expansion around a Gaussian density using cumulants. The estimating density is assumed to have the form:

$$
\begin{aligned}
f_X(x;\kappa) = \phi(x;\kappa)\Big\{&1 + \kappa^{i,j,k}h_{ijk}(x)/3! + \\
&(\kappa^{i,j,k,l}h_{ijkl}(x)/4! + \kappa^{i,j,k}\kappa^{l,m,n}h_{ijklmn}(x)/6!) + ..\Big\}
\end{aligned} \tag{10}
$$

where ϕ is a Gaussian density, κ are cumulants and h are Hermite Polynomials. The notation follows that of [6]. Now, the data passed through the Gaussian distribution function in the projection pursuit technique has the form:

$$
\rho(R) = \rho_x(\vartheta^{-1}(R))/\phi(\vartheta^{-1}(R)) \tag{11}
$$

where

$$
\vartheta^{-1}(R) = U(\phi^{-1}((R+1)/2)) \tag{12}
$$

so that if the incoming data is assumed to have the density $f_x(x,\kappa)$, then

$$
\begin{aligned}
f_X(\vartheta^{-1}(R);\kappa) = \Big\{&1 + \kappa^{i,j,k}h_{ijk}(\vartheta^{-1}(R))/3! + \\
&(\kappa^{i,j,k,l}h_{ijkl}(\vartheta^{-1}(R))/4! + \kappa^{i,j,k}\kappa^{l,m,n}h_{ijklmn}(\vartheta^{-1}(R))/6!) + ...\Big\}
\end{aligned} \tag{13}
$$

and this is the density that is estimated using Legendre polynomials. Thus, the density that Edgeworth attempts to approximate using Hermite polynomials, PPI BSS estimates using Legendre polynomials. Since the recursion equation for Legendre polynomials involves only other Legendre polynomials, it is easy to compute. This gives an interpretation of the relation between the Projection pursuit technique and the standard Edgeworth (or related Gram-Charlier) expansion (see [6]).

SIMULATIONS

Procedure

We tested the projection pursuit index against the first maximum-likelihood estimator of Harroy-Lacoume as presented in [5]

$$\theta_{ML} = \tfrac{1}{4}\arctan\left(\frac{\sum_i \rho_i^4 \sin(4\varphi_i)}{\sum_i \rho_i^4 \cos(4\varphi_i)}\right) \qquad (14)$$

Both estimators were used to determine the rotation angle of data generated according to the product of 2 one-dimensional distributions and passed through a mixing matrix. Laplace and uniform distributions were used to generate the data. We also varied the following parameters to determine their effect on the estimation accuracy: number of samples of the data, rotation angle, order of the ppi estimate, resolution of the ppi search, and noise level. All simulations were done in Matlab.

As mentioned in the introduction, we encountered difficulty in doing gradient descent to find the rotation angle. Because even high dimensional BSS can be reduced to compositions of 2 dimensional rotations, we opted to do BSS over two dimensions at a time. This restricted the search space of rotation angles to $[-\pi/4,\pi/4]$. The search space being so small, we could reasonably carry out brute force search of the PPI at the rotation angles over the range of interest. We did this intelligently using the following algorithm analogous to Comon[3]:

```
determine PPI at θ=0
for k=1:n
        if PPI at θ+π/2^(k+1) is > PPI at θ
                θ= θ+π/2^(k+1);
        else if PPI at θ-π/2^(k+1) is > PPI at θ
                θ= θ-π/2^(k+1);
        end
end
```

this does a binary search of the range space producing an estimate of the rotation angle to within $\pi/2^{k+1}$ in k steps.

Results

Below, we show the estimation accuracy of the rotation angle estimate as a function of the order used to determine the PPI. We carried out this analysis for both uniform and Laplace densities as done in [5]. As expected, for densities far from Gaussian, such as the uniform, a much larger order is required to obtain good angle estimates. The Laplacian density, being much closer to the Gaussian, required lower order Legendre polynomials to obtain good estimates.

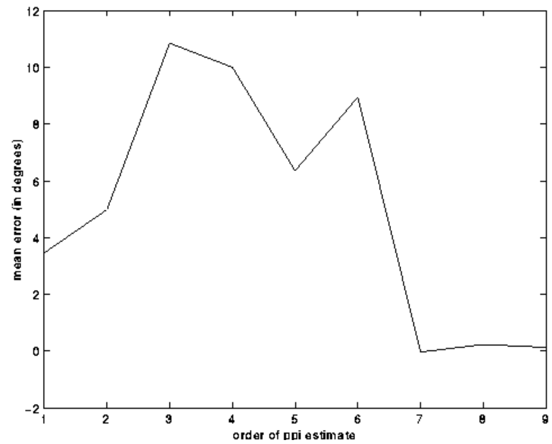

Figure 1: Error in rotation angle as a function of order of Legendre polynomial approximation for uniform density.

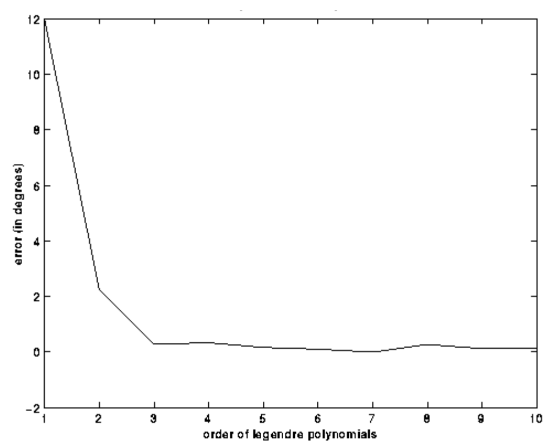

Figure 2: Error in rotation angle as a function of order of Legendre Polynomial Approximation for Laplacian Density.

This is consistent with the observations of Friedman et. al. [4] who claim that the ability to find interesting directions is not critically dependent on the order J as long as J is in the range [4,8] and sample sizes are not small.

We also compared the PPI to the Gram-Charlier estimator (as proposed by Harroy-Lacoume) in estimating the rotation angle. The bias in the estimate as a function of the rotation angle are shown in graphs below. The data were distributed according to the joint uniform and Laplacian densities mentioned above. This data was taken at an SNR of 10db and 5000 samples, but the same catastrophic increase in error occurs for the Gram-Charlier estimate at signal SNR's of 6dB and 20dB and sample sizes of 500 and 2000 samples at a rotation angle of approximately 23 degrees. The PPI performs uniformly well across all rotation angles.

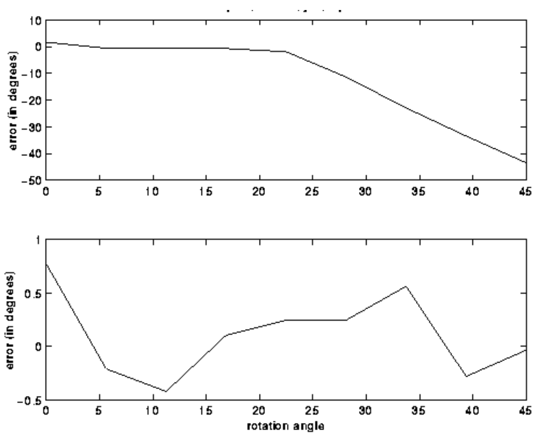

Figure 3: Error in rotation angle estimate as a function of rotation angle for Gram-Charlier method (top graph) and PPI method (bottom graph). The Gram-Charlier method has catastrophic increase in error for rotation angle> 23 degrees.

DISCUSSION

Comparison of PPI with Gram-Charlier expansion

The PPI has several advantages over log-likelihood with the Gram-Charlier expansion. In particular, the Gram-Charlier estimator is limited to null skewness and kurtosis within the range [0,4] ([5], p170). The PPI is not limited in this way. This may explain the failure of the Gram-Charlier estimate at large rotation angles (i.e.>23 degrees).

In fact, although Harroy and Lacoume claim they are using the Gram-Charlier expansion, their result also holds for the Edgeworth expansion for a symmetric density (such as we are using here). The Edgeworth expansion is simply a Gram-Charlier expansion except that terms that converge at the same rate are grouped together. But in the Edgeworth expansion, the 3^{rd} order cumulants and 5^{th} order cumulants vanish for symmetric pdf's so the remaining terms in the Edgeworth expansion [6]vanish. Thus, in the symmetric pdf case, the Gram-Charlier and Edgeworth expansions would produce the same estimate.

CONCLUSION

We have presented Exploratory Projection Pursuit technique and adapted it to Blind Signal Separation. We have also explored its connection with the maximum-likelihood using the Gram-Charlier expansion. We have tested the technique on some standard densities to study the order required for good rotation angle estimates. We have also compared its performance to a standard rotation angle estimator.

REFERENCES

[1] Belouchrani, A.; Abed-Meraim, K.; Cardoso, J.-F.; Moulines, E. "A blind source separation technique using second-order statistics". IEEE Transactions on Signal Processing, vol.45, (no.2), IEEE, Feb. 1997. p.434-44.

[2] Cardoso, J.-F.; Bose, S.; Friedlander, B. "On optimal source separation based on second and fourth order cumulants". Proceedings. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing, Corfu, Greece, 24-26 June 1996.) Los Alamitos, CA, USA: IEEE Comput. Soc. Press, 1996. p.198-201

[3] Comon, P. "Independent component analysis, a new concept?" Signal Processing, vol.36, (no.3), April 1994. p.287-314.

[4] Friedman, J. H., "Exploratory Projection Pursuit", Journal of The American Statistical Association, March 1987, v. 82, n.397, pp.249-66.

[5] Harroy, F.; Lacoume, J.-L. "Maximum likelihood estimators and Cramer-Rao bounds in source separation". Signal Processing, vol.55, (no.2), Elsevier for EURASIP, Dec. 1996. p.167-77.

[6] McCullagh, P., Tensor methods in statistics. London ; New York : Chapman and Hall, 1987.

[7] Pham, D. T. "Blind separation of instantaneous mixture of sources via an independent component analysis". IEEE Transactions on Signal Processing, vol.44, (no.11), IEEE, Nov. 1996. p.2768-79.