Simultaneous Diagonalization for Blind Signal Processing

Bijan Afsari and P. S. Krishnaprasad

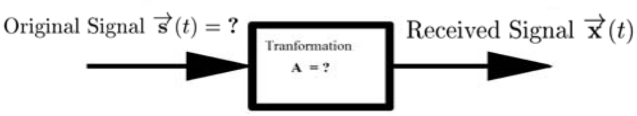
Introduction

What is Blind Signal Processing (BSP)?

The ultimate goal for BSP is to recover a transformed signal using "the least amount of *a priori* information" about the transformation or the original signal. Many practical problems can be formulated in this general context.

Applications:

Blind channel equalization for wireless communications, Biomedical signal processing, Financial data analysis, Sound Localization and Separation, …



The setting of BSP problem

Approach:

Define a cost function such that its minimization gives the transformation or the unknown signal. The cost function is a non-linear function derived based on: Information Theory, Statistics or some properties of the signal. The problem is then rephrased as :

$$\text{Min}_{\{A,\vec{s}\}} J_{\vec{x}}(A, \vec{s})$$

Other ideas:

In defining J some important ideas are:

- Take advantage of diversity, non-gaussianity, non-stationarity and statistical independence ,
- Use Nonlinear functions of data

Blind Source Separation (BSS)

$$\vec{x} = A_{n\times n}\vec{s} + \vec{n}$$

- A simplified instance of BSP is the BSS problem : n statistically independent sources are mixed through the matrix A and contaminated with Gaussian noise **n**, the BSS problem is to estimate the sources, mixing matrix A or its inverse B only by observing samples of **x**.
- In the noiseless case BSS is equivalent to restoring independence, i.e. if the elements of B**x** are independent then separation is achieved.
- Cumulants are statistical measures of independence. The elements of the fourth order cumulant tensor of **x** (assuming zero mean for **x**) are:

$$Cum(x_i,x_j,x_k,x_l) = E\{x_ix_jx_kx_l\} - E\{x_ix_j\}E\{x_kx_l\} - E\{x_ix_k\}E\{x_jx_l\} - E\{x_jx_k\}E\{x_ix_l\}$$

- If we form matrix slices of the cumulant tensor of **x** by fixing two indices k,l we obtain a set of matrices that are diagonalizable (in the congruence manner)j by the un-mixing matrix B .
- Finding the un-mixing matrix is equivalent to the problem of finding a matrix B that simultaneously or jointly diagonalizes a collection of cumulant matrix slices.

Simultaneous Matrix Diagonalization

- If $\{C_1, \dots, C_N\}$ is a set of symmetric matrices to be simultaneously diagonalized by B (e.g. fourth order cumulant matrix slices), define a cost function for simultaneous diagonalization as:

$$J(B) = \sum_{i=1}^{N} \|BC_iB^T - \text{diag}(BC_iB^T)\|_F^2$$

- Unless B is orthogonal this cost function is a not suitable for simultaneous diagonalization, because it can be reduced by diagonal matrices which do not result in separation or independence.

A Gradient Based Simultaneous Diagonalization Algorithm:

- If B is orthogonal we use the gradient on the Riemannian manifold of orthogonal matrices to minimize J.
- If B is non-orthogonal we project the gradient to a suitable sub-space of the tangent space of the Riemannian manifold of non-singular matrices GL(n) so that the cost is not reduced by diagonal matrices.
- The gradient minimization of J(B) with respect to suitable Riemannian metrics result in multiplicative updates of B:

$$B_{k+1} = (I - \mu\Delta_k)B_k, \quad B_0 = I$$

where μ is fixed or variable step size

- If B is orthogonal then Δ_k is skew-symmetric and is defined as:

$$\Delta_k = -\sum_{i=1}^{N}[\text{diag}(B_kC_iB_k^T), B_kC_iB_k^T], \text{ where} [X,Y] = XY - YX$$

- If B is not orthogonal Δ_k has zero trace and it can be defined as:

$$\Delta_k = \Psi_k - \frac{tr(\Psi_k)}{n} \quad \text{or} \quad \Delta_k = \Psi_k - \text{diag}(\Psi_k)$$ where Ψ_k is defined as:

$$\Psi_k = \sum_{i=1}^{N}(B_kC_iB_k^T - \text{diag}(B_kC_iB_k^T))B_kC_iB_k^T$$

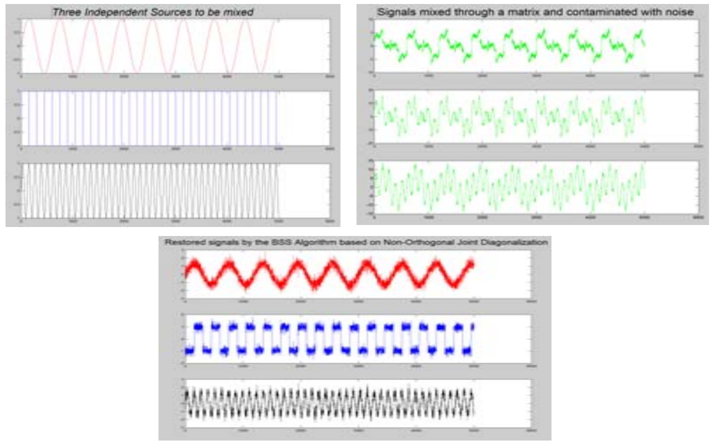
- For small μ, B_k will have determinant close to 1.

A Class of BSS Algorithms Based on Simultaneous Diagonalization

1. Find a whitening matrix W, a square root of the covariance matrix of x, and compute y=Wx
2. Find $\{C_1, \dots, C_N\}$ the set of fourth order cumulant matrix slices of y.
3. Apply the algorithm above for non-orthogonal B.
4. Compute z=By=BWx

- This Algorithm has the advantage that it makes effective use of both second order statistics (which has little variance) and fourth order cumulant which is blind to Gaussian noise but has higher variance

An Example



References:

[1] B. Afsari, P. S. Krishnaprasad, "Some Gradient Based Joint Diagonalization Methods for ICA", Submitted to 5th International Conference on Independent Component Analysis and Blind Signal Separation

Research supported in part by Army Research Office under the ODDR&E MURI01 Program Grant No. DAAD19-01-1-0465 to the Center for Communicating Networked Control Systems (through Boston University).