A Data-Derived Quadratic Independence Measure for Adaptive Blind Source Recovery in Practical Applications

Khurram Waheed and Fathi M. Salam
Circuits, Systems and Artificial Neural Networks Laboratory
Michigan State University
East Lansing, MI 48824-1226

Abstract

We present a novel performance index to measure the statistical independence of data sequences and apply it to the framework of blind source recovery (BSR) that includes blind source separation, deconvolution and equalization. This performance index is capable of measuring the mutual independence of data sequences directly from the data. This information theoretic; Quadratic Independence Measure (QIM) is derived based on Renyi's quadratic entropy estimated by a finite data length Parzen window using a gaussian kernel. Simulation results are presented to validate the performance of the proposed benchmark and compare it with other established benchmarks.

1. Introduction

Stochastic blind signal processing tasks such as blind source separation, deconvolution and equalization have been the focus of wide interest during the last two decades. This wide interest is due to several attractive and diverse application domains for these kinds of problems that include blind classification of topics in an internet chatroom, communication systems, audio and acoustics, finance and marketing, astronomy and physics, bio-medical, space and geo-exploratory applications. However, a daunting task remains to reliably classify the performance of a number of proposed algorithms especially in a fashion that makes them suitable for practical implementation. Most of the presented research on the topic relies on performance benchmarks that require knowledge of either the original source signals or the mixing environment transfer function. Both these quantities are otherwise assumed to be unknown (no precise knowledge of the above mentioned quantities is actually what makes these tasks blind). Therefore, these performance benchmarks render themselves useful for synthetic simulations only. In a practical situation, where we only observe or record a sequence of observation vectors, precise knowledge of the original signals or the environment transfer function is never available. Compounding this situation with the inherent indeterminacies of permutation and scaling in the blind signal recovery problem, the dilemma of having an unknown number of sources, the unknown order of filters required say for a deconvolution type of setup etc. severely limits the capability to determine the quality of signal separation algorithms in practical situations.

We propose and demonstrate the use of a new benchmark to determine mutual independence of a batch of signals for such circumstances. Unlike most other simulation performance benchmarks employed for BSR, this benchmark can be computed directly from the output data of the BSR network. It may also be applied directly to the observed mixture data in order to quantify the transformation effect of the BSR network. This novel benchmark is based on Renyi's Mutual Information, see Principe et. al [3,4]. In the context of BSR, the use of this independence performance benchmark ensues the debate on the correspondence between statistical independence and source recovery.

In this paper, we will first discuss a number of performance benchmarks proposed by different researchers. We will focus on the a priori knowledge required for their computation and point out their drawbacks when applied to the area of blind source recovery. Later we will provide an overview of Renyi's quadratic entropy and how to utilize it for estimation of signal independence. This is followed by a discussion on how to apply QIM to BSR type problems. At the end we present simulated results and draw our conclusions from this study.

2. Survey of BSR Performance Benchmarks

In this section we will provide an overview of several performance benchmarks that have been employed by various researchers for the problem of blind source recovery.

2.1 Signal to Noise Ratio (SNR)

Signal to noise ratio is a well-known communication benchmark, which defines the ratio between the desired signal and the unwanted noise/distortion powers usually expressed in dBs.

$$SNR = 10\log {P_S}/{P_n} \tag{2.1}$$

where P_S – represents the signal power, and

P_n – represents the noise and/or unwanted signal power

This benchmark is applicable in situations where the desired signal is known and is therefore not very suitable for BSR problems where no such reference signal exists.

2.2 Multichannel Intersymbol Interference (MISI)

MISI is an extension of the communication benchmark inter-symbol interference, which quantifies the smearing of a communication channel by an adjacent channel due to their overlapping temporal and/or spectral characteristics. ISI is useful as a dispersion measure as it is insensitive to the overall gain and mean group delay. In the case of BSR type problems, MISI is a measure to determine the "distance" between the global, i.e., the combined mixing and demixing network transfer functions and an identity impulse transfer function. This is a measure of the global diagonalization and permutation as achieved by the demixing network [9 and the references therein]

$$MISI_k = \sum_{i=1}^{N} \frac{\left|\sum_j \sum_p \left|G_{pij}\right| - \max_{p,j}\left|G_{pij}\right|\right|}{\max_{p,j}\left|G_{pij}\right|} + \sum_{j=1}^{N} \frac{\left|\sum_i \sum_p \left|G_{pij}\right| - \max_{p,i}\left|G_{pij}\right|\right|}{\max_{p,i}\left|G_{pij}\right|} \tag{2.2}$$

where for a linear case

$G(z) = H(z) * \bar{H}(z)$ - represents Global Transfer Function,

$\bar{H}(z) = [A_e, B_e, C_e, D_e]$ – Transfer Function of Environment

$H(z) = [A, B, C, D]$ – Transfer Function of Network

This is the performance benchmark of choice for most of the research on BSR. However, the main drawback of this benchmark is that it requires knowledge of the mixing environment, which is available offline only in synthetic simulation environments and hence cannot be employed in practical blind scenarios.

Another variation of the above performance measure uses a squared version of the global transfer function matrix G_{pij}, i.e., the performance measure relies on L_2 norm instead of the L_1 norm used above. This technique has an additional inherent flaw that small error terms will be de-emphasized.

2.3 Maximum signal to interference plus noise ratio (SINRM)

This measure uses the maximum signal to interference plus noise ratio as a measure to determine separation. Interference in source k is constituted by all sources j such that $j \neq k$. Then SINR of source k at output i of separator W is defined as [5]

$$SINR_k(w_i) = \sigma_k^2 \left| w_i^H a_k \right|^2 \Big/ w_i^H R_{vk} w_i \qquad (2.3)$$

where

w_i - is the i^{th} column of separator matrix W

a_k - is the k^{th} column of mixing matrix A

σ_k^2 - is the input power of source k

R_{vk} - is the temporal mean (for nonstationary sources) of the correlation matrix of the noise plus interference for source k, defined by

$$R_{vk} = R_x - \sigma_k^2 a_k a_k^H \qquad (2.4)$$

Again this measure requires knowledge of mixing transfer function, which is unknown in blind problems.

2.4 Mean Squared Error (MSE)

Mean squared error is probably the best-known adaptation measure. MSE uses the L2 norm of the error as a measure to determine the convergence of an algorithm. In the absence of a desired reference to compute an explicit error measure as in the BSR case, this error may be computed as the difference between the current and the previous value of an adapted parameter or using a probabilistic score function. Although using MSE one can determine whether an algorithm has reached its steady state but there is no way to determine whether the algorithm has converged to the desired solution or a spurious local minimum, and as such is not suitable for quantification of BSR algorithm. For any parameter γ, MSE is given by

$$MSE(\gamma) = \begin{cases} \left|\gamma(k) - \gamma(k-1)\right|^2 \\ \left|\gamma(k) - \psi(\gamma(k))\right|^2 \end{cases} \qquad (2.5)$$

Some other performance indices proposed in [6] for synthetic audio environments are given below

2.5 Distortion Measure

The distortion in the j^{th} separated output can be defined as

$$D_j = 10\log\left(E\left\{ \left(m_{j,s_j} - \alpha_j y_j \right)^2 \right\} \Big/ E\left\{ \left(m_{j,s_j}, sj \right)^2 \right\} \right) \qquad (2.6)$$

where

$\alpha_j = E\left\{ m_{j,s_j}^2 \right\} \Big/ E\left\{ y_j^2 \right\}$ - is a scaling factor

$E\{.\}$ - is the statistical expectation operator

y_j - is the corresponding output for the j^{th} source, and

m_{i,s_j} - is the contribution of the j^{th} source to the i^{th} microphone.

2.6 Separation Measure

The quality of separation for the j^{th} separated output can be defined as

$$S_j = 10\log\left(E\left\{ \left(y_{j,s_j} \right)^2 \right\} \Big/ E\left\{ \left(\sum_{i \neq j} y_{j,s_i} \right)^2 \right\} \right) \qquad (2.7)$$

where

y_{j,s_i} - represents the jth output when only ith source is active.

The last two benchmarks can also be represented in the frequency domain and are applicable for synthetic evaluations only, where the original source signals are known and their contribution to an observed mixture & output can be estimated.

3. Proposed Independence Benchmark

We present our proposed general benchmark for determination of statistical independence based on Renyi's quadratic mutual information. We present some salient features of this measure and use of Parzen Window method for its estimation. For a more detailed discussion on Renyi's quadratic mutual information, see [1,4]. The framework was applied to the problem of blind source separation by Principe et. al. for the development of a generalized information theoretic learning framework; where the learning of the demixing structure can be derived directly from data. However, the resulting update laws are computationally very expensive. We are instead proposing to utilize a finite data-length quadratic entropic measure as a performance benchmark for achieving statistical independence.

3.1 Renyi's Entropy

Renyi's entropy definition is derived from his proposed theory of means [1] and is given by

$$H = \varphi^{-1}\left(\sum_{k=1}^{N} p_k \varphi(I(p_k)) \right) \qquad (3.1)$$

where

$\varphi(.)$ - is a continuous and strictly monotonic function subclass of Kolmogorov-Nagumo functions. To satisfy the constraints of an information measure

$$\varphi(x) = \begin{cases} x & \text{Shannon's Entropy} \\ 2^{(1-\alpha)x}; & \text{Renyi's Entropy} \end{cases} \qquad (3.2)$$

$I(p_k)$ - any information measure, e.g., $I(p_k) = -\log(p_k)$ is Hartley's information measure.

Simplifying the above relation, we have

$$H_{R_\alpha} = \frac{1}{1-\alpha}\log\left(\sum_{k=1}^{N} p_k^{\alpha}\right); \alpha > 0, \alpha \neq 1 \quad (3.3)$$

Renyi's quadratic entropy is the special case for $\alpha = 2$, i.e.,

$$H_{R_2} = -\log\left(\sum_{k=1}^{N} p_k^2\right) \quad (3.4)$$

3.2 Parzen Window Estimator

For the purpose of estimating the squared probability density for the quadratic mutual information, a Parzen kernel estimator [2] is employed. The Parzen estimator takes the form of a convolution of an estimator kernel with the observations. Using this estimator, the pdf $p_y(z)$ of a random vector $y \in \Re^m$ is given by

$$\hat{p}_y(z) = \frac{1}{N}\sum_{i=1}^{N} \kappa(z - y_i) \quad (3.5)$$

where

$y_i \in \Re^m$ is the i^{th} observation vector

$\kappa(.)$ is a kernel function that satisfies the conditions for a pdf. There are several such well-known kernel functions, e.g., the gaussian kernel, the cauchy kernel or the uniform kernel [2]. For our implementation we choose the Gaussian kernel with covariance matrix Σ, i.e.,

$$\kappa(z) = G(z,\Sigma) = \frac{1}{(2\pi)^{M/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{z^T \Sigma^{-1} z}{2}\right) \quad (3.6)$$

This choice is motivated by an integral property of the Gaussian kernel, which results in an efficient and exact computation of Renyi's quadratic entropy, for the case of 2 gaussian kernels

$$\begin{aligned} &\lim_{N\to\infty} \sum_{k=1}^{N} G(z_k - y_i, \Sigma_1) G(z_k - y_j, \Sigma_2) \\ &= \int_{-\infty}^{\infty} G(z - y_i, \Sigma_1) G(z - y_j, \Sigma_2) dz = G(y_i - y_j, \Sigma_1 + \Sigma_2) \end{aligned} \quad (3.7)$$

where

$y_i \in \Re^m$ and $y_j \in \Re^m$ are two data vectors in the space

Σ_1 and Σ_2 are covariance matrices for corresponding Gaussian kernels.

3.3 Quadratic Independence Measure (QIM)

The proposed performance benchmark is based on the Cauchy-Schwartz Inequality. For L_2 norm, the inequality is

$$\|x\|^2 \|y\|^2 \geq \left(x^T y\right)^2 \quad (3.8)$$

Using the above inequality, we can "approximate" the Kullback-Lieblar divergence by taking the logarithm of both sides of (3.8) and rearranging

$$\log \frac{\|x\|^2 \|y\|^2}{\left(x^T y\right)^2} \geq 0 \quad (3.9)$$

For two pdfs $f(x)$ and $g(x)$, the above inequality can be used as a divergence measure between the two pdfs. The relation is given by

$$P_{cs}(f,g) = \log \frac{\left(\int_{-\infty}^{\infty} f(x)^2 dx\right)\left(\int_{-\infty}^{\infty} g(x)^2 dx\right)}{\left(\int_{-\infty}^{\infty} f(x) g(x) dx\right)^2} \quad (3.10)$$

with the following properties

- $0 \leq P_{cs}(.) \leq 1$
- $P_{cs}(.) = 0$ if and only if $f(x) = g(x)$
- $P_{cs}(f,g) = P_{cs}(g,f)$

expressing in terms of quadratic entropy

$$P_{cs}(f,g) = 2H_{R_2}\left(\sqrt{f(x)g(x)}\right) - \left(H_{R_2}\left(f(x)\right) + H_{R_2}\left(g(x)\right)\right) \quad (3.11)$$

In order to apply this criterion for the measurement of independence in the source separation framework, let

$f(x)$ represent the joint probability density estimate of the signal vector using Parzen estimator

$g(x)$ represent the product of marginal probability density estimates using the Parzen estimator

For k observations of an output vector $Y \in \Re^N$, the above relation can be expressed as

$$P_{cs}(y) = -\log\left[\mathrm{M}_{CI}(Y)\right] \quad (3.12)$$

where

$\mathrm{M}_{CI}(Y)$ -represents the normalized cross information measure

$$\mathrm{M}_{CI}(Y) = \frac{\mathrm{M}_{UCIP}(Y)^2}{\mathrm{M}_{JIP}(Y)\prod_{i=1}^{N} \mathrm{M}_{MIP}(Y_i)} \quad (3.13)$$

$\mathrm{M}_{JIP}(Y)$ - represents the joint information measure

$$\mathrm{M}_{JIP}(Y) = \frac{1}{k^2}\sum_{i=1}^{k}\sum_{j=1}^{k}\left(\prod_{l=1}^{N} G(Y_{il} - Y_{jl}, 2\sigma^2 I)\right) \quad (3.14)$$

$M_{PCIP}(Y_j, Y_l)$ - represents the partial marginal information measure

$$\mathrm{M}_{PCIP}(Y_j, Y_l) = \frac{1}{k}\sum_{i=1}^{k}\left(G(Y_{jl} - Y_{il}, 2\sigma^2 I)\right), l = 1 \ldots N \quad (3.15)$$

$\mathrm{M}_{MIP}(Y_i)$ - represents the marginal information measure

$$\mathrm{M}_{MIP}(Y_i) = \frac{1}{k}\sum_{j=1}^{k}\left(\mathrm{M}_{PCIP}(Y_j, Y_l)\right), l = 1 \ldots N \quad (3.16)$$

$\mathrm{M}_{UCIP}(Y)$ - represents unnormalized cross information measure

$$\mathrm{M}_{UCIP}(Y) = \frac{1}{k}\sum_{j=1}^{k}\left(\prod_{l=1}^{N} \mathrm{M}_{PCIP}(Y_j, Y_l)\right) \quad (3.17)$$

In case of linear instantaneous and convolutive mixtures, separability can be defined as a combination of output independence and diagonalization of the global transfer function [7,8,9]. The diagonalization measure ensures that all sources are extracted and that the algorithm did not converge to a spurious solution in the data space, where estimated data sources are independent but different from the original sources. In a completely blind scenario, there is no known criterion to determine this diagonalization without explicit knowledge of the mixing transfer function. However, for a finite known number of network outputs practically one can verify that all the separated sources are distinct and independent. This ensures separation of all non-gaussian source distributions, whereas for a gaussian source this ensures separation from other distributions only, while the reverberative effects might still be present.

The effectiveness of this performance measure depends on the length as well as the probability distributions of the constituents in an observation sequence. In Fig. 1, we present a graph for QIM performance measure as a function of length of observed data. The data has three constituents that include speech, communication signals, and gaussian noise.

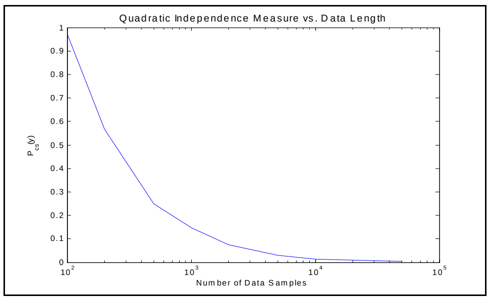


Figure 1. Effect of length of observation sequence on Quadratic Independence Measure (QIM)

4. Simulations Results

Due to space limitation, we present only one simulation case. The mixing environment is given by

$$\sum_{j=0}^{m-1} A_i m(k-i) = \sum_{i=0}^{n-1} B_i s(k-i) + v(k) \qquad (4.1)$$

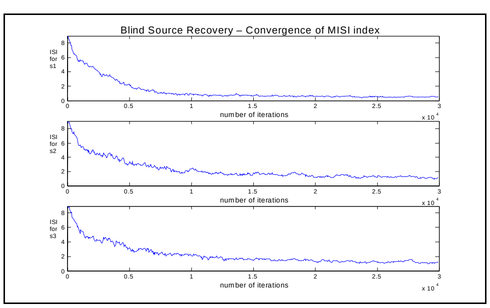
where

$$A_0 = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, A_1 = \begin{bmatrix} 0.5 & 0.8 & -0.7 \\ 0.8 & 0.3 & -0.2 \\ -0.1 & -0.5 & 0.4 \end{bmatrix}, A_2 = \begin{bmatrix} 0.06 & 0.4 & -0.5 \\ 0.16 & -0.1 & -0.4 \\ -0.3 & -0.06 & 0.3 \end{bmatrix}$$

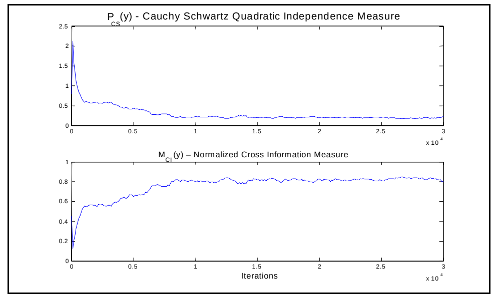
$$B_0 = \begin{bmatrix} 1 & 0.6 & 0.8 \\ 0.3 & 1 & 0.1 \\ 0.6 & -0.8 & 1 \end{bmatrix}, B_1 = \begin{bmatrix} 0.5 & 0.5 & 0.6 \\ -0.3 & 0.2 & -0.3 \\ -0.2 & -0.43 & 0.6 \end{bmatrix}, B_2 = \begin{bmatrix} .125 & 0.06 & 0.2 \\ -0.1 & 0 & 0.4 \\ 0.08 & -0.13 & 0.3 \end{bmatrix} \qquad (4.2)$$

$v(k)$ - Additive Gaussian noise

The feedforward separation results for each channel using MISI and QIM (computed with a batch of 1000 samples) are shown below



(a)



(b)

Figure 2. (a) Convergence of MISI performance Index (b) Corresponding performance of QIM performance Index

5. Conclusions

We have extensively verified the proposed independence benchmark for BSR problems that include both minimum phase and non-minimum phase systems. Due to the normalized nature of QIM, it is suitable for performance comparison of demixing achieved by different algorithms. The primary inhibition is that computationally the benchmark has a quadratic relationship with the length of the data set. Therefore for practical situations, it is recommended to use either the QIM measure with a shorter data set or another computationally inexpensive measure such as MSE to determine the convergence of the algorithm. Once the algorithm has converged, a larger data chunk may be used to determine the achieved performance level using QIM.

6. References

[1] Renyi A., “*On measures of Entropy and Information*”, Selected papers of Alfred Renyi, vol. 2. (Budapest Akademia Kiado) 1976.

[2] Parzen E., “*On the estimation of a probability density function and the mode*”, Ann. Math. Stat., vol. 33, 1962, pp. 1065

[3] Xu, D. “*Energy, entropy and information potential in neurocomputing*”, Ph.D dissertation 1998, University of Florida.

[4] Principe J., Xu D., and Fisher J. ,”*Information-Theoretic Learning*”, Chapter 7, Unsupervised Adaptive Filtering, vol. I Blind Source Separation. John Wiley & Sons, 1999.

[5] Chevalier P. “*On the performance of higher order blind source separation methods*,” in IEEE-ATHOS workshop on HOS, Begur Spain, 12-14 June 1995, pp. 30-34.

[6] Schobben D., Torkkola K. & Smaragdis P., “*Evaluation of Signal Separation Methods*”, in Proc. of 1st Intl. workshop on ICA and BSS, Aussois, France, January 11-15, 1999, pp. 261-266.

[7] Cardoso J. F., “*Entropic Contrasts for Source Separation: Geometry and Stability*”, Chapter 4, Unsupervised Adaptive Filtering, vol. I Blind Source Separation. John Wiley & Sons, 1999.

[8] Common P., “*Independent Component Analysis, a new concept* ?”, Signal Processing, Special issue on High-Order Statistics, September 1996, vol. 53, no. 2, pp. 93-107.

[9] Salam F. M., Waheed K.,”*State-space feedforward and feedback structures for blind source recovery”*; in 3rd Intl. Conf. on ICA and BSS, December 9-12, 2001 - San Diego, California, pp. 248-253