Higher Order Statistics and Their Roles in Blind Source Separation (BSS)

Sattar B. Sadkhan[1], Nidaa A. Abbas[2], *Member, IEEE*

Abstract- **Blind Signal separation and independent component analysis are emerging techniques of Data analysis that aim to recover unobserved signals or " sources" from observed mixture.**

Such problem requires us to venture familiar second order statistics, because a penalty term involving only pair wise decorrelation would not lead to separation.

Source separation can be obtained by optimizing a contrast function, i.e., a scalar measure of sum distributional property of the output. The constant modules property is very specific, more general contrast function are based on other measure, such as entropy, mutual independence, higher order decorrelation, divergence between distribution of output and some model, etc.

The contrast function is used here can derived from maximum likelihood principle. The basic BSS model can be treated in several directions, considering for instance, more sensors than sources, noisy observation, and complex signal and mixture, or obtains the standard narrow band array processing / beaminforming model. Another extension is to consider convolution mixture: this result in multi channel blind deconvolution problem. These extensions are of practical importance.

Sometimes the researches are restricted to simplest model (i.e., real signal as many sensors as sources, nonconvolutive mixture, noise free observation) because its capture the essence of the BSS problem. Normally the BSS approach answers the following questions:-

- **When is source separation possible?**
- **To what extent can the source signal be recovered?**
- **What are the properties of the source signal allowing for partial or complete blind recovery ?.**

The aim of this paper is to analyze some of the operations that have been recently developed to address the blind signal (source) separation based on statistical principles and parameters.

Index Term- Applied Statistics, Blind deconvolution and equalization, blind separation of signals, independent component analysis, higher order statistics, learning rate, Principal Component analysis.

S. B. Sadkhan with the University of Babylon, Iraq (phone: +9647801884154; e-mail: drengsattar@ ieee.org).
N. A. Abbas , with University of Babylon, Iraq. She is now with the Department of Computer Science, (phone: +9647801805510, e-mail: drnidaa_muhsin@ieee.org).

I. Introduction

Moments and cumulants are widely used in scientific disciplines that deal with data, random variables or stochastic processes. They are well known tools that can be used to quantify certain statistical properties of the probability distribution like location (first moment) and scale (second moment). The definition is given by [1],

$$\left.\begin{aligned} \mu &= \frac{1}{N}\sum_{i=0}^{N-1} x_i \\ \sigma^2 &= \frac{1}{N-1}\sum_{i=0}^{N-1}(x_i-\mu)^2 \end{aligned}\right] \qquad (1)$$

where N is the total number of samples, x_i are the values in the signal, μ, σ^2 are the mean and variance respectively (some times the mean value can represented by the name of the value with carrying small bar, such as $\overline{Y}$, $\overline{X}$). In practise we have set of probability distribution samples and compute the estimates of these moments. However, for higher order moments these estimates become increasingly dominated by outliers, by which we mean the samples which are far away from the mean. Especially for heavy tailed distributions, this implies that these estimates have high variance and are generally unsuitable to measure properties of the distribution.

An undesirable property of moments is the fact that lower order moments can have dominating influence on the value of higher order moments. For instance, when the mean is large it will have dominating effect on the second order moment,

$$E[x^2] = E[x]^2 + E[x - E[x]]^2 \qquad (2)$$

where E[.] denotes the expectation operator. The second term which measures the variation around the mean, i.e. the variance, is much more suitable statistic for scale than the second order moment. This process of subtracting lower order information can be continued to higher order statistics. Well known higher order cumulants are skewness (third order) measuring asymmetry and kurtosis (fourth order) measuring ”peakiness” of the probability distribution [1].

Many statistical methods and techniques use moments and cumulants because of their convenient properties. For instance they follow easy transformation rules under affine transformations. Examples in the machine learning literature there are certain algorithms for Independent Components Analysis (ICA) [2,3,4]. Well known drawback of this algorithm is their sensitivity to outliers in the data. Thus, there is a need to define robust cumulants which are relatively insensitive to outliers but retain most of the convenient properties that moments and cumulants enjoy.

In Blind Source Separation (BSS), multiple observations acquired by an array of sensors are processed in order to recover the initial multiple source signals. The term blind refers to the fact that the source signals are not observed and no information is available about the mixture [5].

The above problem which is related to foundation of the latent structure in high dimensional data. The term latent means hidden, unknown or unobserved; the term structure refers to some regularities in the data; high dimensional may be tens or tens of thousands of dimensions, depending on the situation; and data is any information that can be transformed into numerical values, most often represented as a matrix of multidimensional observations where each dimension corresponds to a variable whose value can be measured. One of the important points in this subject is to answer the question: "what are the data contains?", to form a simple representation of a large data set that is difficult to analyze as such, and to present the data in a form that is understandable to a human observer[6,7,8].

The main method for analyzing latent structure in the data is Independent Component Analysis. ICA is a computational method for separating a multivariate signal into additive subcomponents supposing the mutual statistical independence of the non-Gaussian source signals.

The independence assumption is correct in most cases so the blind ICA separation of a mixed signal gives very good results. It is also used for signals that are not supposed to be generated by a mixing for analysis purposes. The statistical method finds the independent components (factors, latent variables or sources) by maximizing the statistical independence of the estimated components. Non-Gaussianity, motivated by the central limit theorem, is one method for measuring the independence of the components. Non-Gaussianity can be measured, for instance, by kurtosis or approximations of negentropy.

Typical algorithms for ICA use centering, whitening and dimensionality reduction as preprocessing steps in order to simplify and reduce the complexity of the problem for the actual iterative algorithm. Whitening and dimension reduction can be achieved with Principal Component Analysis (PCA). Algorithms for ICA include infomax, FastICA and JADE, but there are many others also [9,10,11,12].

In the other hand, Principal Components Analysis (PCA) is a technique that can be used to simplify a dataset; more formally it is a linear transformation that chooses a new coordinate system for the data set such that the greatest variance by any projection of the data set comes to lie on the first axis (then called the first principal component), the second greatest variance on the second axis, and so on. PCA can be used for reducing dimensionality in a dataset while retaining those characteristics of the dataset that contribute most to its variance by eliminating the later principal components (by a more or less heuristic decision). These characteristics may be the "most important", but this is not necessarily the case, depending on the application.

Some algorithms utilize second-order (SO) statistics as the classical PCA in factor analysis. In contrast, ICA attempts to restore the independence of outputs using higher order statistics. The consequence is that the indeterminacy is reduced so that ICA allows blind identification of the static mixture, and transmitted sources can eventually be extracted [13,14]

More precisely, the ICA concept relies on the core assumptions that: -

i) Sources should be independent in some way. Additionally, when a contrast functional is sought to be maximized.

ii) the mixture has to be overdetermined, which means that there should be at most as many sources as sensors [15]. In fact, there must exist a linear source separator [6].

Since the first paper related to higher order (HO) BSS, published in 1985 [16], many concepts and algorithms have come out. For instance, the ICA concept was proposed a few years later, as well as the maximization of a fourth-order (FO) contrast criterion (subsequently referred to as COM2) [6]. At the same time, a matrix approach was developed in [7] and gave rise to the joint diagonalization (JAD) [17]. A few years later, Hyvarinen *et al.* developed the FastICA method: first for signals with values in the real field [18] and later for complex signals [10], using the fixed-point algorithm to maximize an FO contrast. This algorithm is of deflation type, as is that of Delfosse *et al.* [18], and must extract one source at a time, although some versions of FastICA extract all sources simultaneously. In addition, Comon proposed a simple solution named COM1 in [19], to the maximization of another FO contrast function previously published in [20,21,22]. Another algorithm of interest is second order (SO) blind identification (SOBI), based only on SO statistics, developed independently by several authors in the 1990s and addressed in depth later in [14].

The aim of this paper is to analyze some of the operations that have been recently developed to address the blind signal (source) separation based on statistical principles and parameters.

This paper is organized as follows. Section two introduces the higher order statistics (HOS). Section three introduces the BSS problem. Section four defines the PCA and ICA in detail. Section five gives the statistical properties of adaptive algorithm for blind separation. Section six introduces the adaptive algorithm for blind deconvolution. Section seven provides the conclusions.

II. Higher Order Statistics (HOS)

In recent years the field of HOS has continued its expansion, and applications have been found in fields as diverse as economics, speech processing, seismic data processing, plasma physics and optics. Many signal processing conferences (ICASSP, EUSIPCO) now have sessions specifically for HOS, and an IEEE Signal Processing Workshop on HOS has been held every two years since 1989 [1,4].

HOS measures are extensions of second-order measures (such as the autocorrelation function and power spectrum) to higher orders. The second-order measures work fine if the signal has a Gaussian (Normal) probability density function, but as mentioned above, many real-life signals are non-Gaussian.

A. *Higher Order Moments (3rd -skewness)*

For univariate data $Y_1, Y_2, ..., Y_N$, the formula for skewness is:

$$skewness = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^3}{(N-1)\sigma^3} \qquad (3)$$

where $\bar{Y}$ is the mean value, σ is the standard deviation, and N is the number of data samples. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero. Data skewed to the left are said to be negatively skewed; the mean and median are to the left of the mode. Data skewed to the right are said to be positively skewed; the mean and median are to the right of the mode, Fig.(1) show the skewness [4].

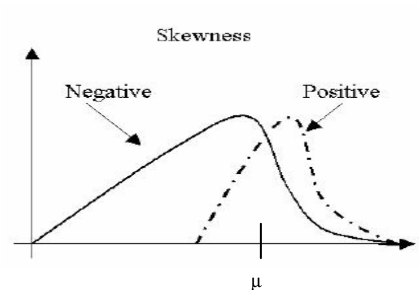


Fig. (1). Skewness.

B. *Higher Order Moments (4th-kurtosis)*

For univariate data $Y_1, Y_2, ..., Y_N$, the formula for kurtosis is:

$$kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)\sigma^4} \qquad (4)$$

where $\bar{Y}$ is the mean value, σ is the standard deviation, and N is the number of data samples. Fig. (2) show the kurtosis.

The kurtosis for a standard normal distribution is three. For this reason, some sources use the following definition of kurtosis [21]:

$$kurtosis = \frac{\sum_{i=1}^{N}(Y_i - \bar{Y})^4}{(N-1)\sigma^4} - 3 \qquad (5)$$

Kurtosis can be either positive or negative. Random variables that have a negative Kurtosis are called subGaussian, and those with positive Kurtosis are called superGaussian. In statistical literature, the corresponding expression platykurtic and leptokurtic are also used.

SuperGaussian random variables have typically a “spiky” PDF with heavy tails, i.e., the PDF is relatively large at zero and at large values of the variable, will being small for intermediate values. On the other hand, subGaussian random variables have typically a “flat” PDF, which is rather constant near zero, and very small for large values of the variables, as shown in Fig. (2) [23].

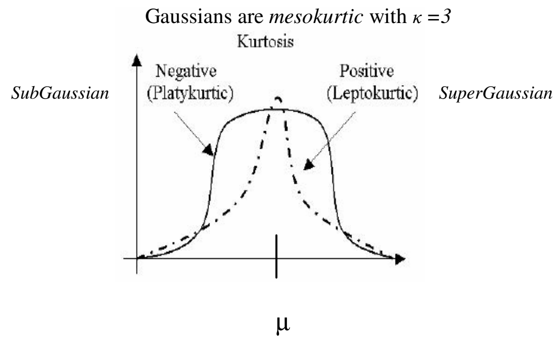


Fig.(2). Kurtosis

Example (1)

The following example shows histograms for 10,000 random numbers generated from a normal, a double exponential, a Cauchy, and a Weibull distribution which show the measure of skewness and kurtosis [4] as in Fig. (3).

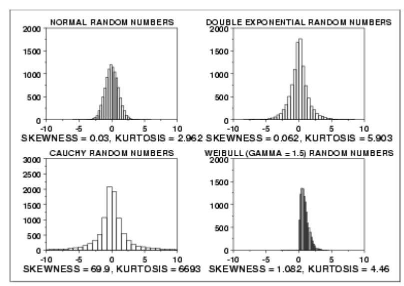


Fig. (3). an example show the measure of skewness ad kurtosis

III. Blind Signal Separation (BSS)

Blind signal separation, also known as blind source separation, is the separation of a set of signals from a set of mixed signals, without the aid of information (or with very little information) about the nature of the signals.

Blind signal separation relies on the following assumption: The source signals are non-redundant. For example, the signals may be mutually statistically indepndent or decorrelated.

Blind signal separation thus separates a set of signals into a set of other signals, such that the regularity of each resulting signal is maximized, and the regularity between the signals is minimized (i.e. statistical independence is maximized).

The separation of independent sources from mixed observed data is a fundamental and challenging signal processing problem [9,24]. In many practical situations, one or more desired signals need to be recovered from the mixtures only. A typical example is speech recordings made in an acoustic environment in the presence of background noise and/or competing speakers. The task of Blind Signal Separation (BSS) is that of recovering unknown source signals from sensor signals described by:

$$x(t)=As(t) \qquad (6)$$

where $x(t)=[x_1,x_2,...,x_n]^T$ is an available $n\times 1$ sensor vector, $s(t)=[s_1,s_2,...,s_n]^T$ $n\times 1$ unknown source vector having stochastic independent and zero-mean non-Gaussian elements $s_i(t)$, and A is a $n \times n$ unknown full-rank and non singular mixing matrix. The BSS problem consists in recovering the source vector s(t) using only the observed data x(t), the assumption of independence between the entries of the input vector s(t) and possibly some a priori information about the probability distribution of the inputs. Statistical independence means that given one of the source signals, nothing can be estimated or predicted about any other source signals. Fig. (4) Shows an example of eq. (6).

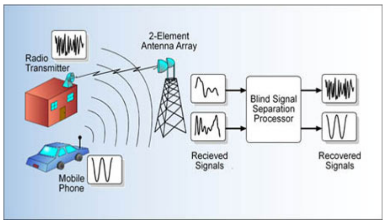


Fig. (4). Example of Mixing Model.

This model in Eq. (6) is instantaneous (or memoryless) because the mixing matrix contains fixed elements, and also noise-free.

If noise is included in the model, it can be treated as an additional source signal or as measurement noise. In this case the model becomes:-

$$x(t) = As(t) + n(t) \quad (7)$$

where the noise vector $n(t)$ is of dimension $n \times 1$. The mixing matrix may be constant, or can be variable with the time index t. In the time-varying case, A becomes $A(t)$ [10,13].

In "multichannel blind deconvolution" or "blind equalization", the *n*- dimensional vector of received signals $x(t)$ is assumed to be produced from the *m*-dimensional vector of source signals using "z -domain" mixture model: -

$$x(z) = A(z)s(z) \quad (8)$$

In this case, the mixture is said to be a "convolutive mixture", (i.e., the channel has some memory effect) [10].

A. Instantaneous Linear Mixtures of Signals

In order to recover the original source signals from the observed mixtures, we use a simple linear separating system [8]: -

$$y(t) = Bx(t) \quad (9)$$

where $y(t) = [y_1(t),..., y_n(t)]^T$ is an estimate $s(t)$, and B is a $n \times n$ (assume $n = m$) separating matrix.

B.Convolutive Mixtures of Signals

A simple Finite Impulse Response (*FIR*) feedback architecture is combined with a second order cost function and gradient descent learning to separate two speech signals. The process is blind in that nothing is known about the sources or the mixing process. The conditions under which it is possible to separate multiple signals are given. Spatial diversity information, which exploits only the structure between multiple sensors, is employed to separate instantaneous mixtures and a combination of spatial and spectral diversity information is used to separate convolutive mixtures. The mixing process is assumed to be linear and time-invariant and the demixing process is linear.

The overall two-source, two-observation system $(m=2, n=2)$ for a feedback architecture [25] is shown in Fig. (5).

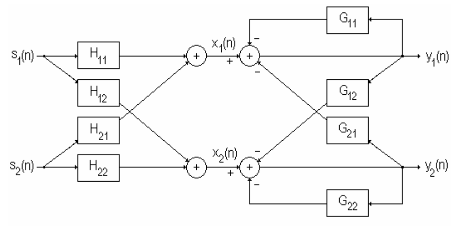


Fig. (5). Block diagram of overall system (including both mixing $x(z) = H(z)s(z)$ and demixing sub-blocks) When $(m = 2, n = 2)$ and $y(z) = x(z)[I + G(z)z^{-1}]^{-1}$.

where $s(z)$ is the $m \times 1$ source vector in the $z-$domain, $H(z)$ is the $n \times m$ mixture matrix, and $x(z)$ is the $n \times 1$ observation vector and $G(z)$ is the $m \times n$ demixing matrix . Each element of $G(z)$ is an *FIR* filter, hence the reason for the name *FIR* feedback.

IV. PRINCIPAL COMPONENTS ANALYSIS (PCA) AND INDEPENDENT COMPONENTS ANALYSIS (ICA)

Principal Components Analysis (PCA) is a technique used to reduce multidimensional data sets to lower dimensions for analysis. Depending on the field of application. It is also named the discrete Karhunen-Loève transform, the Hotelling transform or proper orthogonal decomposition (POD) [1].

PCA is mostly used as a tool in exploratory data analysis and for making predictive models. PCA involves the calculation of the eigenvalue decomposition or Singular value decomposition of a data set, usually after mean centering the data for each attribute. The PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system. PCA is theoretically the optimum transform for a given data in least square terms.

In PCA an observed vector x is first centered by removing its mean. Then the vector is transformed by a linear transformation into a new vector, possibly of lower dimension, whose elements are uncorrelated with each other. The linear transformation is found by computing the "eigenvalues decomposition" of the covariance matrix, which for zero-mean vectors is the correlation matrix $E\{xx^T\}$ of the data. The eigenvectors of $E\{xx^T\}$ form a new coordinate system in which the data are presented.

The decorrelating process is called whitening or sphering. This can be accomplished by scaling the vector elements by the inverses of the eigenvalues of the correlation matrix. The whitened data have the form:

$$\tilde{x}(t) = D^{-1/2} E^{T} x(t) \qquad (10)$$

where $\tilde{x}(t)$ is the whitened data vector, D is a diagonal matrix containing the eigenvalues of the correlation matrix and E contains the corresponding eigenvectors of the correlation matrix as its columns [16].

Independent Components Analysis (ICA) is a statistical and computational technique for revealing hidden factors (latent) that underlies sets of random variables, measurements, or signals.

ICA defines a generative model for the observed multivariate data, which is typically given as a large database of samples. In the model, the data variables are assumed to be linear or nonlinear mixtures of some unknown latent variables, and the mixing system is also unknown. The latent variables are assumed nonGaussian and mutually independent and they are called the independent components (IC) of the observed data. These independent components, also called sources or factors, can be found by ICA [9].

Several assumptions are required for successful blind separation using ICA method, they are [6, 9, 11, 12]: -

- The sources are "statistically independent" of one another. This assumption is very important and a common one for all the algorithms of blind separation.
- The channel can be instantaneous or convolutive and the matrix A is assumed to be invertible.
- The number of sensors n is greater than or equal to the number of the sources m. This is necessary assumption in most existing algorithms. However, it has been shown that in some applications, the number of sources can be greater than the number of sensors.
- At most one source is normally distributed. This valid assumption only for the noise-free model given in Eq. (6).
- The mixing matrix A is full rank.
- Sources are zero mean and stationary.
- The noise $n(t)$ is white and Gaussian noise.

Although robust moments and cumulants can potentially find applications in a broad range of scientific disciplines, we will illustrate their usefulness by showing how they can be employed to improve algorithms for independent components analysis (ICA). The objective of ICA is to find a new basis for which the data distribution factorizes into a product of independent one dimensional marginal distribution. To achieve this, removes first and second order statistics from the data by shifting the sample mean to the origin and sphering the sample covariance to be the identity matrix. These operations render the data decorrelated but higher order dependencies may still remain. It can be shown [3] that if an independent basis exists, it must be a rotation away from the basis in which the data is decorrelated, i.e.

$$x_{ica} = O\, x_{decor} \qquad (11)$$

where O is a rotation. One approach to find O is to propose a contrast function that, when maximized, returns a basis onto which the data distribution is a product of independent marginal distributions. Various contrast functions have been proposed, e.g. the negentropy [6] and the mutual information [2]. All contrast functions share the property that they depend on the marginal distributions which need to be estimated from the data. Naturally, the Edgeworth expansion [6,4] and the GramCharlier expansion [2] have been proposed for this purpose. This turns these contrast functions into functions of moments or cumulants. However, to obtain reliable estimates one needs to include cumulants of up to fourth order. It has been observed frequently that in the presence of outliers these cumulants often become unreliable.

Example (2) [26]: -

Let us take two sources; each one has ten discrete samples, as shown in Fig. (6) given below, and the data as follows: -

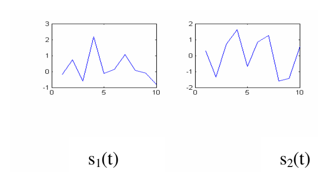


Fig. (6). Original data of the two sources.

$s_1(t) = [-0.18671, 0.72579, -0.58832, 2.1832, -0.1364, 0.11393, 1.0668, 0.059281, -0.095648, -0.83235]$

$s_2(t) = [0.29441, -1.3362, 0.71432, 1.6236, -0.69178, 0.858, 1.254, -1.5937, -1.441, 0.57115]$

Mixing process

Let use choose uniform randomly distributed mixing matrix (its elements values are bounded between 1 and -1) for the given two sources and two mixing sensors, as follows: -

$$A = \begin{bmatrix} -0.90026 & -0.21369 \\ 0.53772 & 0.028035 \end{bmatrix}$$

Then using Eq. (12), the resulting mixing signals as shown in Fig. (7), and the data are: -

$x_1(t) = [0.10518, -0.36787, 0.377, -2.3124, 0.27062, -0.28591, -1.2284, 0.28718, 0.39403, 0.62728]$

$x_2(t) = [-0.092144, 0.35281, -0.29633, 1.2195, -0.09274, 0.085317, 0.6088, -0.012803, -0.091831, -0.43156]$

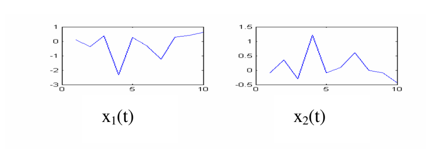


Fig. (7). Mixed Data.

Whitening Process using PCA

This process calculates the necessary two whitening vectors to be used in separation algorithms. Using Eq. (16), the resulting whitening vectors as shown in Fig. (8) and the data are: -

$v_1(t) = [0.39534, 1.7124, 1.0502, 0.018475, 0.54197, 0.70937, 0.43103, 1.4999, 1.2558, 1.0833]$

$v_2(t) = [-0.13926, 0.50168, 0.48294, 2.6763, 0.28931, 0.2996, 1.403, 0.26609, 0.40059, 0.7741]$

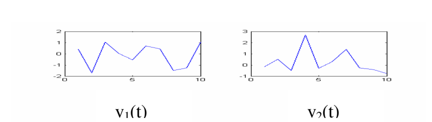
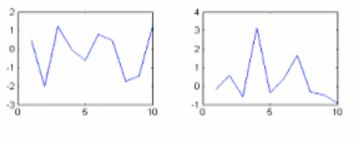


Fig. (8). Whitening Data.

By applying one of the BSS algorithm [27,28] in example above the result is shown in Fig. (9).



$y_1(t)$ $y_2(t)$

Fig. (9). Demixed signal.

V. Statistical Properties of Adaptive Algorithm for Blind Separation

Some of adaptive algorithms are efficient in giving accurate estimators and some are convergent. Let

$$p(s)=\prod_{i=1}^{n} p_i(s_i) \qquad (12)$$

be the true probability density function (PDF) of the source signals s_i and $p_i(s_i)$ is the probability of signal samples. Then, the PDF of $x=As$ is written in terms of $B=A^{-1}$ as

$$p_x(x;B,p)=\det(B)\,p(Bx) \qquad (13)$$

Given a series of observations $x(1),\ldots, x(T)$, it is a statistical problem to estimate the true B. This problem is "ill-posed" in the sense that the statistical model in Eq.(13) includes not only the parameters B, which we want to know, but also n unknown functions $p(s)$, $i= 1,\ldots, n$.

A statistical model is said to be semiparametric when it includes extra unknown parameters of infinite dimensions. Since the unknown functions are of infinite dimensions, this brings some difficulties for estimating B [10].

A.Estimating Functions

By design, all valid contrast functions reach their minima at a separating point when the model holds; in this sense, no one is better than another. In practice, however, contrasts are only estimated from a finite data set. Sample-based contrasts depend not on the distribution of y but on its sample distribution. Estimating from a finite data set introduces stochastic errors depending on the available samples and also on the contrast function. Thus, a statistical characterization of the minima of sample-based contrast functions is needed and will provide a basis for comparing contrast functions [10,29].

A learning algorithm is easily obtained from an estimating function as: -

$$W(k+1)=W(k)+\eta(k)F[y(k)]W(k) \qquad (14)$$

where $W(k)$ is demixing matrix after using whitening process and $\eta(k)$ is a learning rate at time k.

An important problem is to find such an estimating function which gives a good performance. An estimating function F is said to be inadmissible when there exists an estimating function *F* which gives a better estimator than F does for any probability distributions. We need to obtain the class of admissibly estimating functions [8].

B.Information Geometry and its Role in Analysis of Such Statistical Problem

Information geometry is particularly useful for analyzing this type of problem. When it is applied to the present problem we can obtain all the set of estimating functions. It includes the Fisher efficient one, which is asymptotically the best one. However, the best choice of F (estimating function) again depends on the unknown *P*, thus we need to use an adaptive method. The following important results are obtained by applying information geometry:

1)The off-diagonal components $f_{ij}(y,W)$, $i\neq j$, of an admissible estimating function has the form

$$f_{ij}(y,W)=\alpha f_i(y_i)y_i+\beta f_i(y_i)y_i \qquad (15)$$

where α and β are suitably chosen constants or variable parameters.

2)The diagonal part $f_{ii}(y,W)$ can be arbitrarily assigned.

Most learning algorithms have been derived heuristically, one might further try to obtain better rules by searching for an extended class of estimating functions such as $f(y)g(y)^T$ or more general ones. However, this is not admissible, and we can find a better function for noiseless case in the class of:

$$F(y)=\alpha f(y)y^T+\beta y f(y)^T \qquad (16)$$

It should be noted that F(y) and K (w) F(y) give the same estimating equations, where K(w) is a linear operator. Therefore, F and KF are equivalent when we estimate w by batch processing. Two learning rules: -

$$W(k+1)=W(k)+\eta(k)F(y) \qquad (17)$$

$$W(k+1)=W(k)+\eta(k)K(w)F(y) \qquad (18)$$

have different dynamical characteristics, although their equilibria are the same. The universally convergent algorithm uses the inverse of the Hessian as K(w), so that the convergence is guaranteed.

Summarizing, all the adaptive learning algorithms with the equivariant properties for blind separation of sources can be written in the general form using estimating functions [8,29].

VI Adaptive Algorithms for (Blind Deconvolution)

A.Learning Algorithms in the Frequency Domain

When the observed signals x(k) are time-delayed multi-path mixtures as in

$$x(k) = \sum_{p=-\infty}^{\infty} H_p s(k-p) \qquad (19)$$

where H_p is an $(m \times n)$-dimensional matrix of mixing coefficients at lag p.

By using a truncated version of a doubly-infinite multichannel equalizer of the form: -

$$y(k) = \sum_{p=-\infty}^{\infty} W_p(k) x(k-p) \qquad (20)$$

where $y(k) = [y_1(k),...y_n(k)]^T$ is an n-dimensional vector of the output signals which are to be estimators of the source signals, and $\{W_p(k),\ -\infty \le p \le \infty\}$ is a. sequence of $(m \times n)$-dimensional coefficient matrices .We need spatial separation and temporal decomposition in order to extract the source signals s(k). A simple way of extending the blind source separation algorithms is to use frequency domain techniques. By using the Fourier Transform, in eq. (21) and (22) are represented by: -

$$x(w) = H(w)s(w) \qquad (21)$$
$$Y(w) = w(w)x(w) \qquad (22)$$

where w denotes the frequency and H, W are the linear matrix expressions [8,29].

B.Adaptive Algorithm in Time Domain for Multi-Input Multi-Output for Blind Deconvolution

We discuss the natural gradient algorithm for adapting W(z, k) in the convolutive model

$$x(k) = H(z)[s(k)] \qquad (23)$$
$$y(k) = W(z, k)[x(k)] = T(z, k)[s(k)] \qquad (24)$$

where

$$W(z,k) = \sum_{p=-\infty}^{\infty} w_p(k) z^{-p}$$

$$H(z) = \sum_{p=-\infty}^{\infty} H_{p^{z^{-p}}}$$

$$T(z,k)=W(z,k)H(z)$$

For the multichannel deconvolution and equalization task, we assume that the sources $\{s_i(k)\}$ are independent identical distribution (i.i.d.) and that both H(z) and W(z, k) are stable with no zero eigenvalues on the unit circle |z|=1 in complex plane of z. We assume that the number of sources m equals the number of the sensors n and that all signals and coefficients are real valued [17,30].

C. Adaptive Learning Rules for SISO and SIMO Blind Equalization

For SISO blind equalization, the following adaptive learning algorithms: -

1-filtered-regressor (FR) algorithm

$$W(k+1) = W(k) - \eta(k) f(y(k-L)) u_k^* \qquad (25)$$

where $w(k) = [w_0(k),...w_L(k)]^T$, $y(k) = \sum_{p=0}^{L} w_p(k) x(k-p)$, and

$u_k = [u(k),..u(k-L)]^T$ *with* $u(k) = \sum_{p=0}^{L} w_{L-p}^*(k) y(k-p)$

2-Extended Blind Separation (EBS) Algorithm

$$W(k+1) = W(k) + \eta(k) F(y_k] W(k) \qquad (26)$$

where $y_k = [y(k),...y(k-L)]^T$, and the $m \times m$ matrix $F[y_k]$ can take one of the following forms: -

$$F^T[y_k] = \Lambda(k) - f(y_k) y_k^H \qquad (27)$$

$$F^T[y_k] = \Lambda(k) - \alpha_1(k) y_k y_k^H - \alpha_2(k) f(y_k) y_k^H + \alpha_3(k) y_k f(y_k^H) \qquad (28)$$

where $\Lambda(k)$ is a diagonal positive definite matrix, e.g., $\Lambda(k) = I$ or $\Lambda(k) = diag\{f(y_k) y_k^H$ and $\alpha_i \ge 0$ are suitable nonnegative parameters [3].

D.Adaptive Optimality of learning rate in the Learning Algorithms

The problem of optimal updating of the learning rate (step size) is a key problem encountered in all the learning algorithms. Many of the research works related to this problem are devoted to batch and / or supervised algorithms. Various techniques like the conjugate gradient, quasi-Newton, and Kalman filter methods have been applied. However, relatively little work has been devoted to this problem for on-line adaptive unsupervised algorithms [17].

VII. CONCLUSIONS

In this paper, we have reviewed adaptive blind signal processing with higher order statistics. Learning adaptive algorithms are mathematically justified and their properties are briefly analyzed.

Any Gaussian signal is completely characterised by its mean and variance. Consequently the HOS of Gaussian signals are either zero (e.g. the third-order moment of a Gaussian signal is zero), or contain redundant information. Many signals encountered in practice have non-zero HOS, and many measurement noises are Gaussian, and so in principle the HOS are less affected by Gaussian background noise than the second-order measures. (e.g. the power spectrum of a deterministic signal plus Gaussian noise is very different from the power spectrum of the signal alone. However the bispectrum of the signal + noise is, at least in principle, the same as that of the signal).

Due to wide interest in this fascinating area of research, further developments (are expected) of computationally efficient separation, deconvolution, equalization, self adaptive or self-organized systems with robust on-line algorithms for many real word applications like wireless communication, the "cocktail party" problem, speech and image recognition, intelligent analysis of medical signals and image, feature extraction, ect.

REFERENCES

[1] Welling M., "Robust Higher Order Statistics", School of Information and Computer Science University of California Irvine, 2001.
[2] S. Amari, A. Cichocki, and H.H. Yang, "A new algorithm for blind signal separation", Advances in Neural Information Processing Systems, 8:757–763, 1996.
[3] A.J. Bell and T.J. Sejnowski, "The independent components of natural scenes are edge filters", *Vision Research*, 37:3327–3338, 1997.
[4] J.F. Cardoso, "High-order contrast for independent component analysis", *Neural Computation*, 11:157–192, 1999.
[5] J.F Cardoso, "Source Separation Using Higher Order Moments", In Proc. ICASSP'89, PP. 2109-2112, 1989.
[6] P. Comon, "Independent component analysis, A new concept ?," Signal Process., vol. 36, no. 3, pp. 287–314, Apr. 1994.
[7] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F*, vol. 140, no. 6, pp.362–370, Dec. 1993.
[8] Amari, S-I and Cichocki, A., "Adaptive Blind Signal processing-Neural Network Approaches", Proceedings of the IEEE, Vol. 86, No., 10, October 1998.
[9] Hyvarinen A., "What is Independent Component Analysis?", Definition, Available at http://cis.hut.fi/projects/ compneuro/whatisica .html/, 2001.
[10] Cardoso,J-F, "Blind Signal Separation: Statistical Principles", Proceeding of the IEEE, Vol. 86, No. 10, October 1998.
[11] Matei B., "A Review of Independent Component Analysis", Available at http://rutgers.edu/rin/resea/Tutorial, PP. 1-28, 1999.
[12] Stone J.V. and Porrill J., "Independent Component Analysis and Projection Pursuit: A Tutorial Introduction", Available at http://www.shef.ac.uk/psychology/stone/, PP. 1-23, 1998.
[13] Hyvarinen A. and Oja E., "Independent Component Analysis: Algorithms and Applications", Neural Networks, Vol. 13, PP. 411-430, 2000.
[14] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics", IEEE Trans. Signal Process., vol. 45, no. 2, pp. 434–444, Feb. 1997.
[15] Oja E., "A Simplified Neuron Model as a Principal Component Analyzer", Journal Math.Bio.15, PP. 267-273, 1982.
[16] J. Herault, C. Jutten, and B. Ans, "Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé," in Proc. Dixième Colloque sur le Traitement du Signal et des Images, Nice, France, Sep. 1985, pp. 1017–1022.
[17] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," SIAM J. MatrixAnal. Appl., vol. 17, no. 1, pp. 161–164, 1996.
[18] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," in *Signal Process.*, 1995, vol. 45, pp.59–83.
[19] P. Comon, "From source separation to blind equalization, contrast-based approaches", in Proc. Int. Conf. Image Signal Process., Agadir, Morocco, May 3–5, 2001, pp. 20–32.
[20] P. Comon and E. Moreau, "Improved contrast dedicated to blind separation in communications," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., Munich, Germany, Apr. 20–24, 1997, pp. 3453–3456.
[21] E. Moreau and O. Macchi, "New self-adaptive algorithms for source separation based on contrast functions," in *Proc. IEEE Signal Processing Workshop Higher Order Statistics*, Lake Tahoe, CA, Jun. 1993,pp. 215–219.
[22] N. Thirion and E. Moreau, "New criteria for blind signal separation", in *Proc. IEEE* Workshop Statistical Signal Array Process., Philadelphia, PA, Aug. 2000, pp. 344–348.
[23] Mansour A., Jutten C., and Ohnishi N., "Kurtosis: Definition and Properties", Fusion'98 Int. Conf., PP. 40-46, 1998.
[24] Mitra S.k., "Digital Signal Processing- A computer-Based Approach", McGraw-Hill, 1998.
[25] Hild K., "Blind Source Separation", Internal Report, March 20, 2000.
[26] Muhsin A. N., "A Comparison Among Adaptive ICA Algorithms for Blind Speech Signals Separation: Cocktail Party Problem", Ph.D. Dissertation, University of Technology, Department of Computer Science, 2006.
[27] A. Hyvarinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Comput.*, vol. 9, no. 7, pp. 1483–1492,1997.
[28] E. Bingham and H. Hyvarinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 1, pp. 1–8, 2000. 3642 IEEE Transactions On Signal Processing, Vol. 53, No. 10, Oct. 2005
[29] Mackay D.J., "Information Theory, Inference, and Learning Algorithms",

Copyright Cambridge University Press 2003, 6.0 Ed., Cambridge, England, 2003.
[30] Enescu M., "Adaptive Methods for Blind Equalization and Signal Separation in MIMO Systems", Ph.D. Dissertation, Helsinki University of Technology, Department of Electrical and Communications Engineering, Signal Processing Laboratory, 2002.

Dr. Eng Sattar Bader Sadkhan , Chief Scientific Researcher Science 1997, Technology Expert Since 2003, Chairman of IEEE IRAQ SECTION and (currently) Postgraduate Lecturer in Babylon University.

Expert in : Wireless Digital Communication and Information Security.

Diploma in Radar Equipments Repairing (1970-1974)- from Wireless and Radar training Center – B.Sc. (1974-1978) Baghdad – IRAQ – Electrical and Electronic Engineering. M.Sc. (1979-1981) in VAAZ Academy- BRNO – CZECH Republic – Wireless Communication Engineering. Ph.D. (1981-1984) in VAAZ Academy- BRNO- CZECH Republic- Detection of Digital Modulated Signals. Diploma in Cryptography in 1988 – From Switzerland.

Since 1991 up to 2003, in Communication Research and Development Center –Institute of Research and Development. I am engaged as Researcher in this Center and I obtained the director position of this center in 2000 till 8 April 2003.

Postgraduate Lecturer, Director of Cultural Relation and Scientific Affairs, Director of Research Centers of Babylon Studies,

Editor-in-Chief, for the International Scientific Journal of Advancement in Computing Technology, (IJACT), in South Korea, since beginning of 2009.

http://www.aicit.org/ijact/

Member of (80) Scientific and Editing Committees of International Scientific Journals, and International Conferences (Arabic and International).

Advisor of more than (100) **Postgraduate theses** (M.Sc. And Ph.D.).

Publishing more than (160) papers in (Arabic, and International) Academicals journals and conferences.

Dr. Nidaa A. Muhsin (M'08) was born in Basraha, Iraq in 1968. She received the B.Sc. in Computer Science from Basraha University, Iraq in 1990, and M.Sc. in computer science from University of Babylon, Iraq in 1999, and Ph.D. in Computer Science from the University of Technology, Iraq in 2006.

She is currently a lecturer in Computer Science Department in University of Babylon. Her main research interests in adaptive signal processing, ICA, data compression, speech scrambling, Steganography, steganalysis.