Blind Source Recovery Using an Adaptive Generalized Gamma Score Function

Abu-Taleb A.S.[1] , Zayed E.M.E.[2], M. El-sayed Waheed [2], Badawy A.M. [1], Mohammed O.A.[2]
[1] Dept. of System and Bioengineering. Fuculty of Engineering,Cairo University (Egypt)
[2] Dept of Mathematics, Faculty of Science, Zagazig University (Egypt)

Abstract

This paper discusses Blind Source Recovery (BSR) of mixtures of multiple source distributions represented by state space equations using an adaptive score function. The proposed parametric score function is derived from the generalized gamma distribution model. An adaptive algorithm to determine the parameters for the proposed score function using mutual information of BSR output is also presented. The primary advantage of the proposed framework is that it renders the adaptive estimation of the demixing network to be completely blind. No a priori information about the distribution structure of the original sources is required. Simulation example verifying the proposed framework is also presented.

Keywords: *Blind Source Recovery, Blind Source Deconvolution, Blind Signal Equalization, Generalized Gamma Function, Kurtosis.*

1. Introduction

Blind Source Recovery (BSR) is an interesting autonomous (or unsupervised) stochastic adaptation problem that includes well-known adaptive signal processing problems of multi-channel Blind Source Separation (BSS), Deconvolution (BSD) and Equalization (BSE) with several potential applications [2,3,6]. The BSR problem denotes recovering original sources from environments that may include convolution, transients, and even possible nonlinearity. The performance of BSR algorithms strongly depends on the choice of an appropriate score function which appears as an element wise acting non-linearity on the output signals [2-6]. For a particular problem, the optimal score function depends on the distribution of the original source signals which are unknown in a blind scenario. In such cases, unless some assumptions about the distribution of the sources are made, any BSR algorithm will potentially be unable to deliver the desired performance. Therefore, adaptive estimation of appropriate score functions is very attractive from a practical implementation viewpoint. The proposed score function was applied to BSR in a state space framework. The state space notion provides a compact representation, capable of handling both time delayed and filtered versions of signals in an organized manner [2,3,6]. Unlike the transfer function models of standard dynamic filters, the use of the state-space can result in several generalized, equivalent and efficient internal descriptions of a system. This allows for recovery of original sources independent from environment identifiability, i.e. determining the exact (or a specific function of) parameters of the environment. There are many adaptive network solutions (representations), which succeed in recovering the original signals even in the absence of precise identifiability, termed as recoverability [2,6]. Existence and constructions of a theoretical solution to the BSR problem can be easily derived using the state space, given a structure of the environment [1,6]. Most of the mixtures encountered in practical BSR problems are from sources with a variety of non-gaussian distributions. On the other hand, most noise phenomena or unidentified sources are assumed to possess gaussian distributions. This results in practical situations, where one has to cope with multiple source distribution mixtures including gamma distributions. We describe the use of the proposed adaptive score function in the linear convolutive class of state space BSR . This paper includes seven sections. In Sec.2 we expose generalized gamma probability density function and its characteristics. The third section show how to use the state space approach in multichannel blind deconvolution. Section four contain the performance function, gradient-based rules and the derivation of an adaptive score function for generalized gamma. Section five contain simulation results, environmental model and demixing Network structure. Section six contain conclusions. Section seven represent references.

2. Generalized Gamma Probability Density Function

The generalized gamma function is a three-parameter distribution. One version of the generalized gamma distribution uses the parameters $k, \beta \, and \, \theta$. The *pdf* for this form of the generalized gamma distribution is given by,

$$f(t) = \frac{\beta}{\Gamma(k).\theta}\left(\frac{t}{\theta}\right)^{k\beta-1} e^{-\left(\frac{t}{\theta}\right)^{\beta}} \qquad (1)$$

where $\theta > 0$ is a scale parameter, $\beta > 0$ and k> 0 are shape parameters and $\Gamma(x)$ is the gamma function of x, which is defined by,

$$\Gamma(x) = \int_0^\infty s^{x-1} e^{-s} \, ds \qquad (2)$$

With this version of the distribution, however, convergence problems arise which severely limit its usefulness. Even with data sets containing 200 or more data points, the maximum likelihood estimation(MLE) methods may fail to converge. Further adding to the confusion is the fact that distributions with widely different values $k, \beta \, and \, \theta$ may appear almost identical. In order to overcome these difficulties, Weibull++ uses a “reparameterization” with parameters $\mu, \sigma \, and \, \lambda$ where,

$$\mu = \ln(\theta) + \frac{1}{\beta}\ln\left(\frac{1}{\lambda^2}\right)$$
$$\sigma = \frac{1}{\beta\sqrt{k}}, \quad \lambda = \frac{1}{\sqrt{k}} \qquad (3)$$

Where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \lambda < \infty$.

While this makes the distribution converge much more easily in computations, it does not facilitate manual manipulation of the equation. The **pdf** of the reparameterized distribution is given by:

$$f(y) = \begin{cases} \frac{|\lambda|}{\sigma.y}.\frac{1}{\Gamma\left(\frac{1}{\lambda^2}\right)}.\exp\left[\frac{\lambda.\frac{\ln(y)-\mu}{\sigma} + \ln(\frac{1}{\lambda^2}) - \exp(\lambda.\frac{\ln(y)-\mu}{\sigma})}{\lambda^2}\right] & \text{if } \lambda \neq 0 \\ \frac{1}{y\sigma\sqrt{2\pi}}\exp(-\frac{1}{2}\left(\frac{\ln(y)-\mu}{\sigma}\right)^2) & \text{if } \lambda = 0 \end{cases} \qquad (4)$$

2.1. Characteristics of the Generalized Gamma

As mentioned previously, the generalized gamma distribution includes other distributions as special cases based on the values of the parameters.

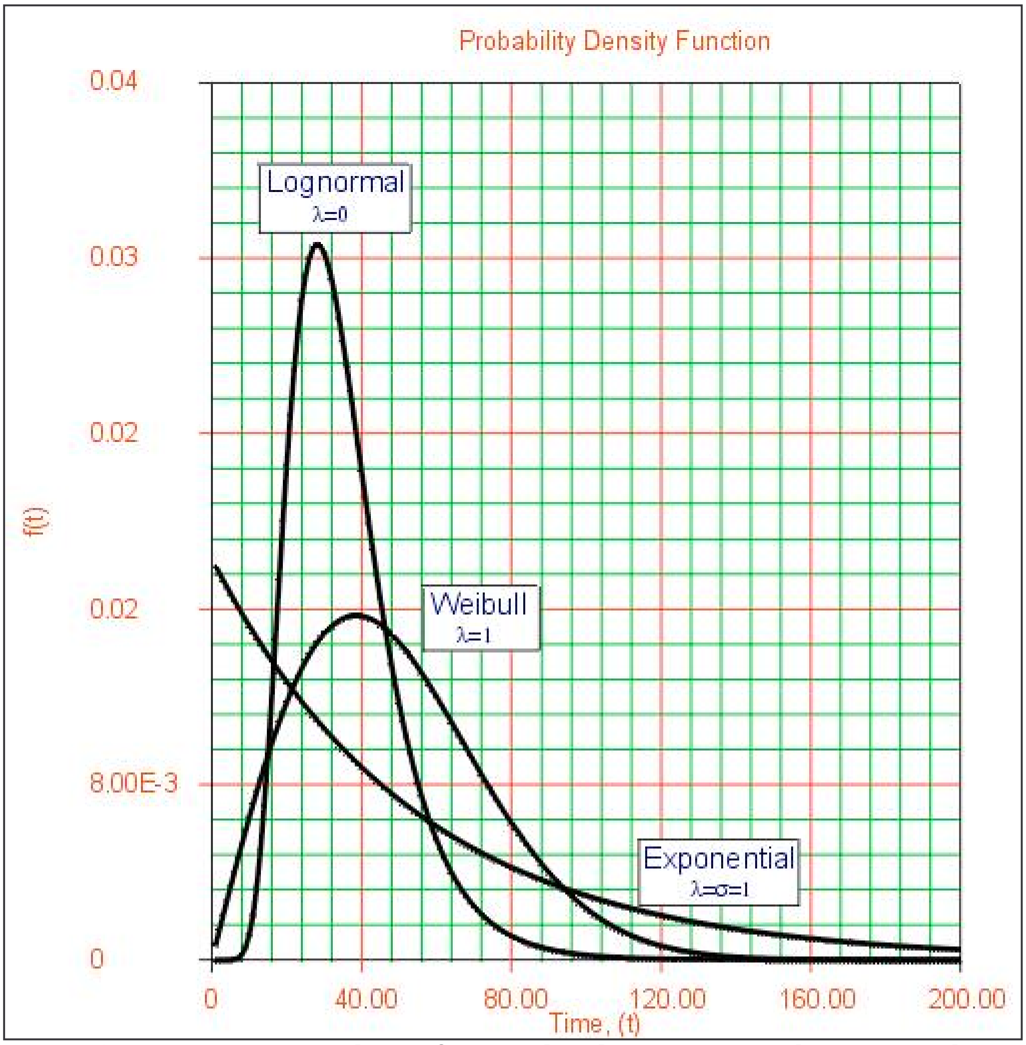

Figure 1. Some members of Generalized Gamma Distribution Family

- If λ= 1 The Weibull distribution is a special when

$$\beta = \frac{1}{\sigma} \qquad \eta = \ln(\mu) \tag{5}$$

In this case the generalized distribution has the same behavior as the Weibull for $\sigma > 1$, $\sigma = 1$ and $\sigma < 1$ ($\beta < 1$, $\beta = 1$, and $\beta > 1$ respectively).

- The exponential distribution is a special case when $\lambda = 1$ and $\sigma = 1$.
- The lognormal distribution is a special case when $\lambda = 0$.
- The gamma distribution is a special case when $\lambda = \sigma$.

By allowing λ to take negative values, the generalized gamma distribution can be further extended to include additional distributions as special cases. For example, the Frechet distribution of maxima (also known as a reciprocal Weibull) is a special case when $\lambda = -1$.

3. State space approach to multichannel blind deconvolution

Given a vector of observed signals u(k); k∈ [0,N], we wish to recover the source signals s(k) based on the assumption that the sources are statistically Independent. If we assume that the observations are convolutive version of the sources, the problem can be tackled using state space approach [6]. We consider the state space approach instead of the transfer function approach, as the state space approach can be easily extended to nonlinear mixing systems. Moreover, the state space approach not only gives an efficient internal description of the dynamic systems, but also there exist different possible equivalent state space realizations, for instance, canonical controller form [6] which allows us to find “efficient” representations of the demixer. We model the mixing environment of the MBD problem as follows:

$$\overline{x}(k+1) = \overline{A}\overline{x}(k) + \overline{B}s(k) + \overline{L}\xi_p(k) \tag{6}$$

$$u(k) = \overline{C}x(k) + \overline{D}s(k) + \theta(k) \tag{7}$$

where $x(k) \in R^d$ is the state vector of the system, $s(k) \in R^n$ is the vector of unknown source signals(assuming that they are zero-mean, I.I.D. and spatially independent), $u(k) \in R^m$ is an available vector of sensor signals, $\overline{A} \in R^{dxd}$ is a state matrix, $\overline{B} \in R^{dxn}$ is an input mixing matrix, $\overline{C} \in R^{mxd}$ is an output mixing matrix, $\overline{D} \in R^{mxn}$ is an input-output mixing matrix and $\overline{L} \in R^{dxd}$ is a noise matrix. The integer d is called the state dimension or system order.

Correspondingly, we model the demixer using a similar discrete time dynamical system:

$$\mathrm{x(k+1) = Ax(k) + Bu(k) + L}\xi_R\mathrm{(k)} \tag{8}$$

$$\mathrm{y(k) = Cx(k) + Du(k)} \tag{9}$$

where $x(k) \in R^M$ is the state vector of the separating system and the unknown state-space matrices have dimensions: $A \in R^{M\,x\,M}, B \in R^{M\,x\,m}, C \in R^{n\,x\,M}, D \in R^{n\,x\,m}$, with $M \geq d$. Here we assume both the mixer and the demixer exist, in particular, D^{-1} exists. The condition of the existence of solution in multichannel blind deconvolution is studied in [6].

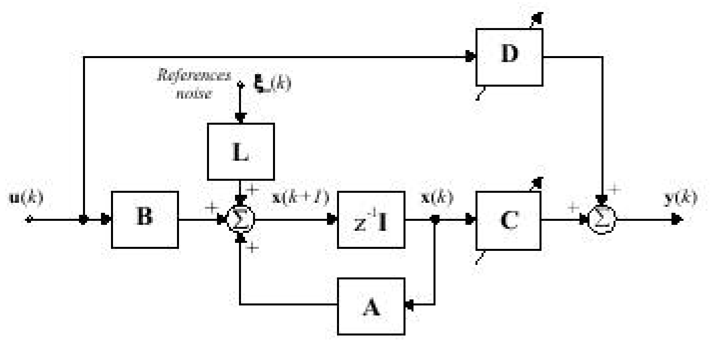

Figure2. General linear state space model for blind deconvolution

4. Performance function

We measure the dependence among the recovered sources y using mutual information. Given P(y), the probability density function (PDF) of the recovered signal vector y, the mutual information between the recovered signals can be defined as follows:

$$I(y) = \int p(y) \ln \frac{p(y)}{\prod_{k=1}^{n} p(y_k)} dy \tag{10}$$

$$=-H(y)+\sum_{q=1}^{n}H(y_q), \tag{11}$$

where $H(y) = -E[\log(P(y)]$ is the entropy of y, $H(y_q) = -E[\log(p(y_q)]$ is the marginal entropy of y_q. For simplicity, for the remaining part of this paper, the time index k is dropped if there is no risk of confusion. Observe that I(y) = 0 if and only if the components of vector y are statistically independent. Therefore I(y) is an appropriate measurement of the dependence among the recovered signals. Unfortunately, mutual information is difficult to compute explicitly, hence we use a cost function similar to [6]:

$$\max_{\theta} I(y,\theta) = \max_{\theta}\left[\log|\det(D)| - \sum_{q=1}^{n}\log p(y_q)\right] \tag{12}$$

where θ is the set of system parameters and source model parameters, which we will study in Section 4.2, det is the determinant. There exist various ways to tackle the optimization problem [6]. Here we follow the derivation of information back-propagation approach .

4.1. Gradient-based learning rules

Based on the cost function (12), we can easily obtain the following updating rules. For matrices Δ and

$$D(k+1) = D(k) + \eta(k)(I - \varphi(y)u^T D^T)D \tag{13}$$

$$C(k+1) = C(k) - \eta(k)\varphi(y)x^T \tag{14}$$

where $\varphi(y)$ is a vector nonlinearity related to the source model. This will be discussed further in Section 4.2. Note, natural gradient [1] is used in (13,14) to improve the performance of the learning process. Similarly, for matrices A and B, we have:

$$A_{ij}(k+1) = A_{ij}(k) - \eta(k)\varphi^T(y)\sum_{l=1}^{N} C_{\bullet l}\frac{\partial x_l}{\partial A_{ij}} \tag{15}$$

$$B_{iq}(k+1) = B_{iq}(k) - \eta(k)\varphi^T(y)\sum_{l=1}^{N} C_{\bullet l}\frac{\partial x_l}{\partial B_{iq}} \tag{16}$$

where $l, i, j = 1,2,...,M$, $q = 1,2,...,m$, $C_{\bullet l}$ denotes the l-th column vector of the matrix C $\frac{\partial x_l}{\partial A}, \frac{\partial x_l}{\partial B}$ can be obtained from following:

$$\frac{\partial x_l(k+1)}{\partial A_{ij}} = \sum_{m=1}^{N} A_{lm}\frac{\partial x_m(k)}{\partial A_{ij}} + \delta_{li}x_j \tag{17}$$

$$\frac{\partial x_l(k+1)}{\partial B_{iq}} = \sum_{m=1}^{N} A_{lm}\frac{\partial x_m(k)}{\partial B_{iq}} + \delta_{li}u_q \tag{18}$$

where δ_{li} is the Kronecker delta function.

4.2. Derivation of an Adaptive Score Function for generalized gamma

Due to the very parameterized nature of the generalized gamma family and its ability to model a variety of source densities of interest to blind source recovery, we derive a generalized score

function (nonlinearity) based on this family. This generalized score function inherits a nice parametric structure from its paramterized parent density.
The element-wise nonlinear score function required for the BSR problem has the definition [3,4,5,6]

$$\varphi(y) = -\frac{\partial \log p(y)}{\partial y} = \left[\varphi_1(y_1),\ \varphi_2(y_2), \dots \varphi_n(y_n)\right] \tag{19}$$

where y_i, i=1,…,n- represent the ith output of the network, $p_i(y_i)$ represent the statistical probability density function of the ith output . Using the generalized gamma family as the candidate density function we have

$$\varphi_i(y_i) = -\frac{\partial \log p_i(y_i)}{\partial y_i} = \begin{cases} \dfrac{\lambda\sigma - 1 + \sigma \exp\left(-\lambda\left(\dfrac{-\ln(y)\sigma + \mu}{\sigma}\right)\right)}{y\lambda\sigma} & if \quad \lambda \neq 0 \\ \dfrac{\sigma^2 + \ln(y) - \mu}{\sigma^2 y} & if \quad \lambda = 0 \end{cases} \tag{20}$$

$$\frac{\partial p}{\partial \lambda} = \frac{sign(\lambda)}{c_1\Gamma\left(\frac{1}{\lambda^2}\right)}\left[\exp\left(\frac{\lambda c_2 + \ln\left(\frac{1}{\lambda^2}\right) - \exp(\lambda c_2)}{\lambda^2}\right)\right] +$$

$$2\left[\frac{|\lambda|}{c_1\Gamma\left(\frac{1}{\lambda^2}\right)}\exp\left(\frac{\lambda c_2 + \ln\left(\frac{1}{\lambda^2}\right) - \exp(\lambda c_2)}{\lambda^2}\right)\right]\frac{psi\left(\frac{1}{\lambda^2}\right)}{\lambda^3} + \tag{21}$$

$$\frac{|\lambda|}{c_1\Gamma\left(\frac{1}{\lambda^2}\right)}\left[\frac{-c_2 - \frac{2}{\lambda} - c_2\exp(\lambda c_2) - \frac{2}{\lambda}\left[\ln\left(\frac{1}{\lambda^2}\right) - \exp(\lambda c_2)\right]}{\lambda^2}\right]\exp\left[\frac{\lambda c_2 + \ln\left(\frac{1}{\lambda^2}\right) - \exp(\lambda c_2)}{\lambda^2}\right]$$

where $c_1 = \sigma y$, $c_2 = \dfrac{\ln y - \mu}{\sigma}$

$$\frac{\partial p}{\partial \mu} = \frac{|\lambda|}{c_1\Gamma\left(\frac{1}{\lambda^2}\right)}\left[\frac{-1 + \exp(c_2\lambda)}{\lambda\sigma}\right]\exp\left(\frac{c_2\lambda + \ln\left(\frac{1}{\lambda^2}\right) - \exp(c_2\lambda)}{\lambda^2}\right) \tag{22}$$

$$\frac{\partial p}{\partial \sigma} = \frac{-|\lambda|}{\sigma c_2\Gamma\left(\frac{1}{\lambda^2}\right)}\exp\left(\frac{c_2\lambda + \ln\left(\frac{1}{\lambda^2}\right) - \exp(c_2\lambda)}{\lambda^2}\right) + \tag{23}$$

$$\frac{|\lambda|}{c_1\Gamma\left(\frac{1}{\lambda^2}\right)}\left(\frac{\frac{-1}{\lambda}c_2 - \frac{1}{\lambda}c_2\exp(\lambda c_2)}{\lambda^2}\right)\exp\left(\frac{\lambda c_2 + \ln\left(\frac{1}{\lambda^2}\right) + \exp(\lambda c_2)}{\lambda^2}\right)$$

and

$$\lambda(k+1)=\lambda(k)+\eta\frac{\partial l}{\partial\lambda}=\lambda(k)+\frac{1}{p}\frac{\partial p}{\partial\lambda} \quad (24)$$

$$\mu(k+1)=\mu(k)+\eta\frac{\partial l}{\partial\mu}=\mu(k)+\frac{1}{p}\frac{\partial p}{\partial\mu} \quad (25)$$

$$\sigma(k+1)=\sigma(k)+\eta\frac{\partial l}{\partial\sigma}=\sigma(k)+\frac{1}{p}\frac{\partial p}{\partial\sigma} \quad (26)$$

where an appropriate choice of λ,μ,σ makes the nonlinearity suitable to a particular source distribution.

5. Simulation Results

In this section, we present simulation results for the non-minimum phase FIR filtering environment with state space representation equivalent to the FIR model mixing environment. The demixing system is formulated as a state-pace network [2,3,5]. In all presented simulations, the source signals are chosen to possess different distributions, respectively; the score function is adapted online along with the BSR algorithm. The convergence performance of the algorithm is measured using the multi-channel intersymbol interference (MISI) benchmark. MISI is a measure of the global transfer function diagonalization and permutation as achieved by the demixing network and is defined as

$$ISI_k=\sum_{j=1}^{N}\frac{\left|\sum_j\sum_p\left|G_{pij}\right|-\max_{p,j}\left|G_{p,i,j}\right|\right|}{\max_{p,j}\left|Gpij\right|}+\sum_{j=1}^{N}\frac{\left|\sum_i\sum_p\left|G_{pij}\right|-\max_{p,i}\left|G_{p,i,j}\right|\right|}{\max_{p,i}\left|Gpij\right|} \quad (27)$$

Where G(z)-Global Transfer Function given by

$$\mathrm{G(z)}=\mathrm{H(z)}*\overline{\mathrm{H}}(\mathrm{z}) \quad (28)$$

and $\overline{\mathrm{H}}(\mathrm{z})=[\mathrm{A}_e, \mathrm{B}_e, \mathrm{C}_e, \mathrm{D}_e]$-Transfer Function of Environment

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

5. 1. Environment Model

This environment is assumed to be a 3×2 IIR filter

$$m(k)=\sum_{j=0}^{m-1}\mathrm{H}_i\,\mathrm{s(k-i)}+\mathrm{v(k)} \quad (29)$$

Where

$$H_0=\begin{bmatrix}1 & -0.2\\ 0.5 & 0.6\\ -0.4 & -0.2\end{bmatrix},\ H_1=\begin{bmatrix}1 & 0.4\\ -0.3 & 1\\ -0.25 & -0.8\end{bmatrix},\ H_2=\begin{bmatrix}-0.75 & 0.7\\ 0.2 & -0.4\\ 0.1 & 0.15\end{bmatrix}$$

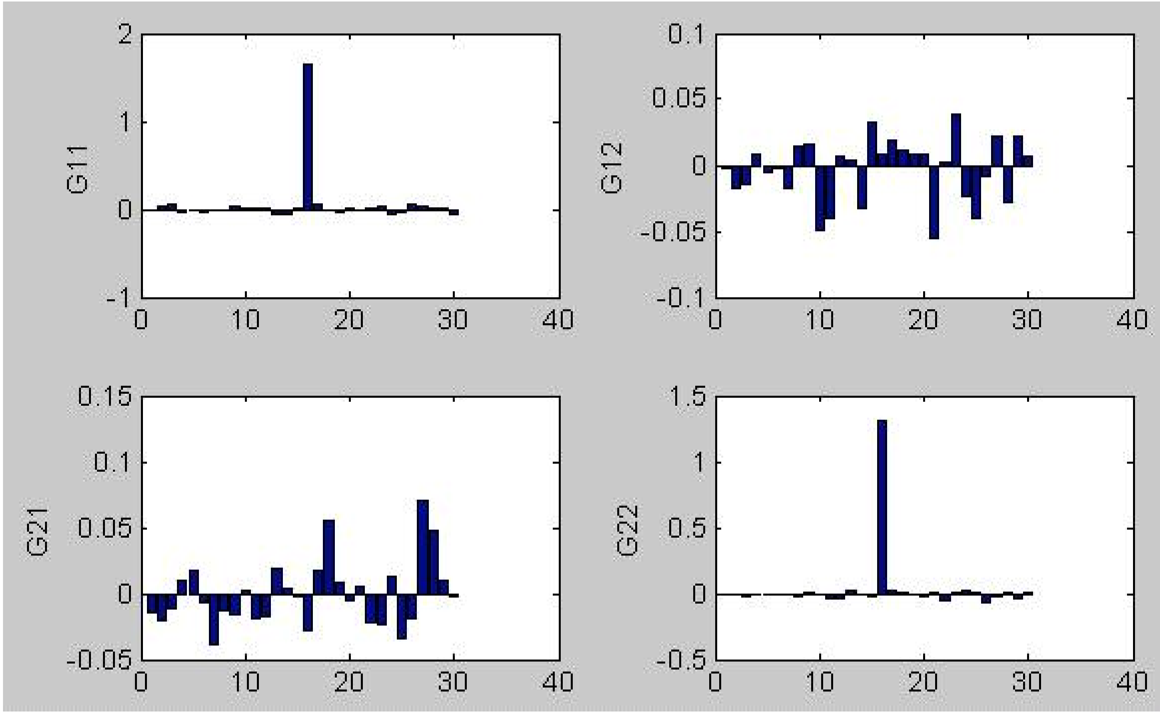

Figure 2. Global transfer function

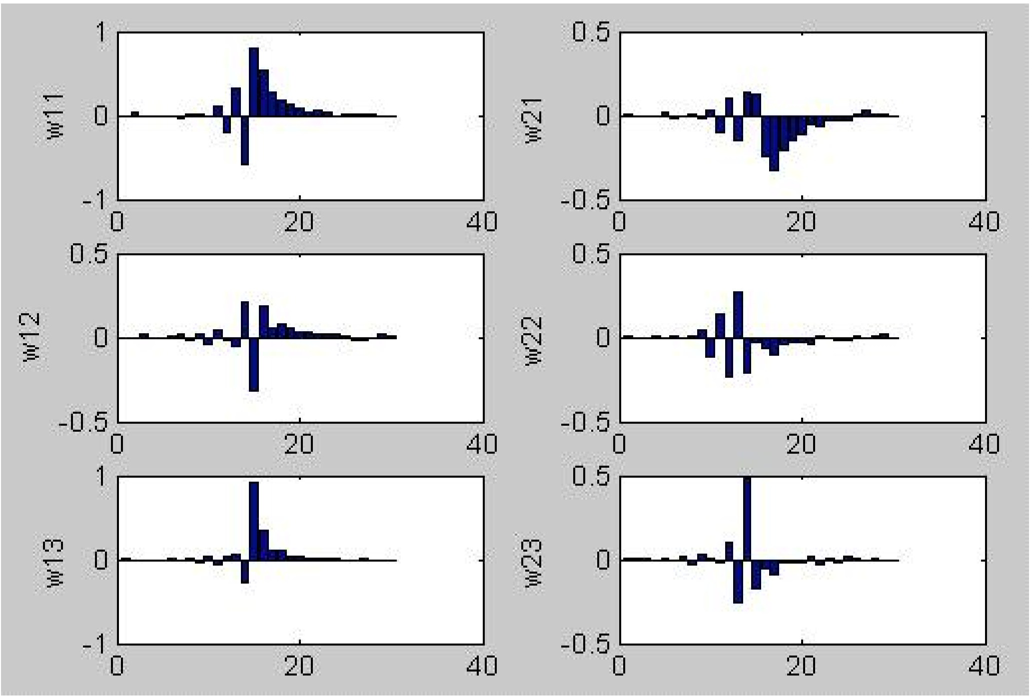

Figure 3. Estimated Demixing Network by BSR

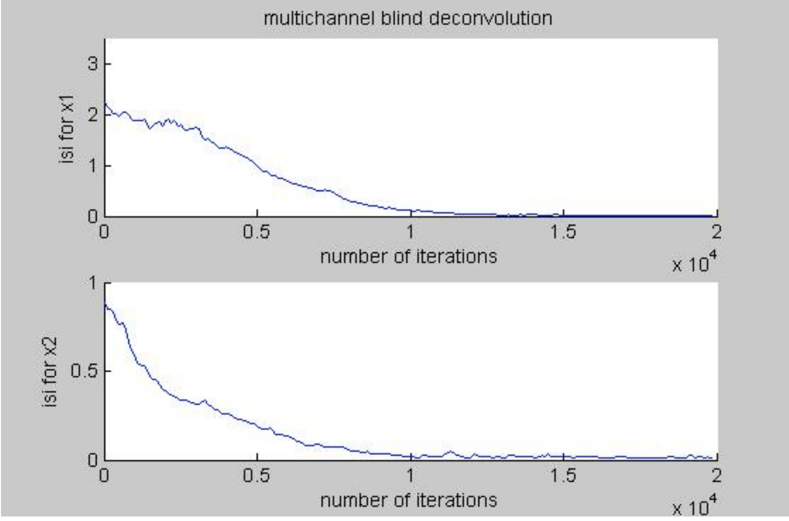

Figure 4. multi-channel intersymbol interference (MISI) benchmark

5.2. Demixing Network Structure

The theoretical inverse of this FIR mixing environment will be a 3× 2 IIR matrix filter with each element having both numerator and denominator polynomials of degree 12. For both the feedforward and the feedback demixing cases, matrix C can be initialized with very small random numbers or all zero elements, while the matrix D is chosen to be an identity matrix. The number of taps per filter was chosen to be 30.

6. Conclusions

We have proposed an adaptive score function for BSR algorithms based on the Generalized Gamma density model. Further we presented a method to adapt this score function online. This adaptive scheme for determination of a score function suited to an unknown source model has been extensively tested for convolutive mixing. A simulation example has also been presented and it is observed that the proposed framework can recover sources from a complex convolutive mixture efficiently even when all the sources have different probability structure.

References:

1. Miller J. H. & Thomas J. B. “Detectors for Discrete-Time Signals in Non-Gaussian noise” in IEEE Transactions on Information Theory, Vol IT-18, no. 2, 1972. pp. 241-250.
2. Salam F. M., Erten G. and Waheed K. “Blind Source Recovery: Algorithms for Static and Dynamic Environments”; Int’l Joint Conf. on Neural Networks, 2001- Washington, D.C.; Vol. 2, pp. 902-907.
3. Salam F. M. and Waheed K. “State Space Feedforward and Feedback Structures for Blind Source Recovery”, in 3 rd Int’l Conf. on ICA and BSS, 2001 - San Diego, CA; pp. 248-253
4. Waheed K. and Salam F. M. “State Space Blind Source Recovery For Mixtures Of Multiple Source Distributions”; in Proc. of Intl Symposium on Circuits and Systems ISCAS-2002, May 26-29, 2002 – Phoenix, Az; Vol I, Page(s) 197-200.
5. Waheed K. and Salam F. M. “Blind Source Recovery: Some Implementation and Performance Issues”; 44th IEEE Midwest Symp. On Circuits and Systems; 2001- Dayton Ohio. Vol. 2, pp. 580-583.
6. Zhang L. and Cichocki A.: “Blind deconvolution of dynamical systems: A state space approach”, Journal of Signal Processing, Vol. 4, No. 2, Mar. 2000, pp. 111-130.

Article received: 2006-03-04