294

A BUSSGANG-TYPE ALGORITHM FOR BLIND SIGNAL SEPARATION

Seungjin CHOI and Youngki LYU

School of Electrical and Electronics Engineering
Chungbuk National University, KOREA
Tel: (0431)261-2421
Fax: (0431)275-6139
Email: schoi@engine.chungbuk.ac.kr

ABSTRACT

This paper presents a new computationally efficient adaptive algorithm for blind signal separation, which is able to recover the narrowband source signals in the presence of cochannel interference without a prior knowledge of array manifold. We derive a new blind signal separation algorithm using the Natural gradient [1] from an information-theoretic approach. The resulting algorithm has the Bussgang property which has been widely used in blind equalization [12]. Extensive computer simulation results confirm the validity and high performance of the proposed algorithm.

1. INTRODUCTION

Blind signal separation is a fundamental problem encountered in many applications such as multiuser communications, array processing, sonar, and image processing. The task of blind signal separation is to calculate the possibly scaled estimates of (unknown) source signals from their instantaneous mixtures without the knowledge of mixing nor source signals. Typical examples are the extraction of multiple signals-of-interest [14] from the outputs of an array of sensors when cochannel interference is dominant channel impairment or the recovery of transmitted symbols from the outputs of a bank of matched filters in Code Division Multiple Access (CDMA) systems [5].

In blind beamforming or multiuser communications, an array of sensors (or a bank of matched filters in CDMA systems) provides an m dimensional observation vector $\mathbf{x}(t)$ which is modeled as a linear instantaneous mixture of n dimensional vector of source signals $\mathbf{s}(t)$. This mathematical formulation is described as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), \tag{1}$$

where $\mathbf{A}$ is an $(m \times n)$ mixing matrix. The additive white Gaussian noise is represented by $\mathbf{v}(t)$ which is assumed to statistically independent of sources $\mathbf{s}(t)$. We should mention that the number of sensors m is greater than or equal to the number of sources.

For an array of sensors that are uniformly spaced (assuming the antenna elements are omnidirectional), the mixing matrix $\mathbf{A}$ is given by

$$\mathbf{A} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{-j\phi_1} & \cdots & e^{-j\phi_n} \\ \vdots & \vdots & \vdots \\ e^{-j(m-1)\phi_1} & \cdots & e^{-j(m-1)\phi_n} \end{bmatrix}, \tag{2}$$

where $\phi_k = 2\pi(d/\lambda)\sin(\theta_k)$, d is the interelement spacing, λ is the wavelength of the sources, and θ_k is the angle of arrival (AOA) of the kth source. The columns of the matrix $\mathbf{A}$ are known as direction vectors or steering vectors because they indicate the response of the array to a narrowband source estimating from a particular direction.

In order to recover the source signals $\mathbf{s}(t)$, we design an adaptive system $\mathbf{W}(t)$ whose output $\mathbf{y}(t)$ is described as

$$\mathbf{y}(t) = \mathbf{W}(t)\mathbf{x}(t). \tag{3}$$

In the context of blind signal separation, it is desirable to update the demixing system $\mathbf{W}(t)$ such that the global system $\mathbf{G}(t) = \mathbf{W}(t)\mathbf{A}$ converges to the generalized permutation matrix as $t \to \infty$, i.e., the steady-state value $\mathbf{G}$ becomes

$$\mathbf{G} = \mathbf{P}\mathbf{\Lambda}, \tag{4}$$

where $\mathbf{P}$ is the permutation matrix and $\mathbf{\Lambda}$ is nonsingular diagonal matrix.

In contrast to the conventional approaches to signal separation, blind signal separation is based on the statistical independence of sources. Since Jutten and Herault's proposal [13], a variety of algorithms have been developed [10, 3, 2, 4, 9, 8, ?, 6]. Recently signal separation algorithms based on the Constant Modulus (CM) criterion [16] have been developed [14, 5].

In this paper, we focus on communication signals which are sub-Gaussian signals (negative kurtosis) and symmetrically distributed. Exploiting these, we derive a Bussgang-type blind signal separation algorithm.

2. MODEL ASSUMPTIONS AND SEPARATION PRINCIPLE

We make the following assumptions throughout this paper:

AS1: The mixing matrix $\mathbf{A}$ is of full column rank.

AS2: Elements of the source vector $\mathbf{s}(t)$ are mutually independent.

AS3: At most, one source signal is Gaussian. (the rest of them are non-Gaussian)

AS4: Each element of source vector $\mathbf{s}(t)$ is zero mean with non-zero variance.

AS5: All source signals $\{s_i(t)\}$ are sub-Gaussian, i.e., their kurtosis are negative.

AS6: All source signals $\{s_i(t)\}$ are symmetrically distributed, i.e., their skewness are zero.

Let us consider a linear mapping from source signals $\mathbf{s}(t)$ to the demixing system output $\mathbf{y}(t)$ in terms of the global transformation $\mathbf{G} = \mathbf{WA}$ ($\mathbf{W}$ is the steady-state value of $\mathbf{W}(t)$),

$$\mathbf{y}(t) = \mathbf{G}\mathbf{s}(t). \quad (5)$$

The vector of source signals, $\mathbf{s}(t)$ consists of statistically independent non-Gaussian signals. The blind signal separation boils down to force all components of $\mathbf{y}(t)$ as independent as possible.

The elements of the measurement data $\mathbf{x}(t)$ are not statistically independent since source signal vector $\mathbf{s}(t)$ is linearly transformed. To recover the source signals, we should transform the measurement data $\mathbf{x}(t)$ into $\mathbf{y}(t)$ whose elements are statistically independent. Thus blind signal separation is called sometimes "independent component analysis". As an optimization criterion, we choose Kullback-Leibler divergence which is an asymmetric measure between two different distributions. Our risk function $R(\mathbf{W})$ is given by

$$\begin{aligned} R(\mathbf{W}) &= K[p(\mathbf{y}) \| \prod_{i=1}^{n} p_i(y_i)] \\ &= \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^{n} p_i(y_i)} d\mathbf{y} \\ &= E\{\log p(\mathbf{y})\} - \sum_{i=1}^{n} E\{\log p_i(y_i)\}, \quad (6) \end{aligned}$$

where $p(\mathbf{y})$ is joint probability density of $\mathbf{y}(t)$ and $p_i(y_i)$ is the marginal probability density of $y_i(t)$. $E\{\cdot\}$ denotes the statistical expectation. The Kullback-Leibler divergence $K[p(\mathbf{y}) \| \prod_{i=1}^{n} p_i(y_i)]$ is nothing but mutual information $I(\mathbf{y})$. This is always greater than or equal to zero [11]. The minimum occurs if and only if all components of $\mathbf{y}(t)$ are statistically independent.

3. THE ADAPTATION ALGORITHM

For the sake of simplicity, we consider the case where the number of sensors are equal to the number of sources, i.e. $m = n$, however the algorithm can be easily extended to the case where $m > n$ (not $m < n$). Additive noise $\mathbf{v}(t)$ is neglected. Consider a demixing system $\mathbf{W}(t)$ whose output $\mathbf{y}(t)$ is described as

$$\mathbf{y}(t) = \mathbf{W}(t)\mathbf{x}(t). \quad (7)$$

The joint probability density of the observation $\mathbf{x}(t)$ and the joint probability density of the demixing system output $\mathbf{y}(t)$ has the following relation,

$$p(\mathbf{y}) = \frac{p(\mathbf{x})}{|\det \mathbf{W}|}, \quad (8)$$

where the det denotes the determinant of a matrix. With this relation, the risk function (6) can be written as

$$R(\mathbf{W}) = -H(\mathbf{x}) - \log|\det\mathbf{W}| + \sum_{i=1}^{n} H(y_i), \quad (9)$$

where $H(\mathbf{x})$ is the joint (differential) entropy of the mixture vector $\mathbf{x}(t)$ and $H(y_i)$ is the marginal entropy of $y_i(t)$.

Note that the risk function (9) depends on the probability distribution $p_i(\cdot)$, however, a prior knowledge of marginal distribution of sources is not available. Thus, we approximate the marginal probability density function of y_i by Gram-Charlier expansion [15] (up to the 4th-order cumulant). Then the marginal entropy $H(y_i)$ can be approximated by (see [7] for the detailed derivation)

$$H(y_i) = \frac{1}{2}\log 2\pi e - \frac{\kappa_{4,i}^2}{2 \cdot 4!} + \frac{1}{16}\kappa_{4,i}^3, \quad (10)$$

where $\kappa_{4,i}$ is the kurtosis of y_i defined by

$$\kappa_{4,i} = E\{|y_i|^4\} - 2E\{|y_i|^2\} - E\{y_i^2\}. \quad (11)$$

Note that the approximation of $H(\mathbf{y}_i)$ is made for standardized variables (i.e., zero mean and unit variance). Since optimization has to be done under the constraint

$E\{|y_i|^2\} = 1$, we add Lagrange multipliers ρ_i in the risk function (6), then the resulting risk function $R(\mathbf{W})$ is

$$R(\mathbf{W}) = -\log|\det \mathbf{W}| + \sum_{i=1}^{n}(\frac{\kappa_{4,i}^3}{16} - \frac{\kappa_{4,i}^2}{2\cdot 4!}) + \frac{1}{2}\sum_{i=1}^{n}\rho_i(E\{|y_i|^2\} - 1). \quad (12)$$

We apply the natural stochastic gradient descent [1] which has been shown to find steepest descent direction in Riemannian manifold (see [7] for detailed derivation). Then, the updating algorithm for $\mathbf{W}$ has the form

$$\Delta\mathbf{W} = \eta\{\mathbf{I} + \mathbf{\Psi}\mathbf{y}\mathbf{y}^H - \mathbf{\Gamma}f(\mathbf{y})\mathbf{y}^H\}\mathbf{W}, \quad (13)$$

where $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ are nonsingular diagonal matrices whose ith diagonal elements are γ_i and $\psi_i = \gamma_i - \rho_i$, respectively. The γ_i is defined by

$$\gamma_i = \kappa_{4,i}\{\frac{3}{4}\kappa_{4,i} - \frac{1}{6}\}, \quad (14)$$

The $\eta > 0$ is a learning rate. The nonlinear function $f(\mathbf{y})$ is a elementwise function which is given by

$$f(\mathbf{y}) = [f_1(y_1), \cdots, f_n(y_n)]^T = [|y_1|^2 y_1, \cdots, |y_n|^2 y_n]^T. \quad (15)$$

The adaptation algorithm (13) can be viewed as Bussgang-type algorithm [12] in spatial domain. For example, if we choose both $\mathbf{\Gamma} = \mathbf{\Psi} = \mathbf{I}$, then stationary points of the averaged version of (13) satisfy

$$E\{f_i(y_i)y_j^*\} = E\{y_i y_j^*\}, \quad \text{for } i \neq j, \quad (16)$$

where $*$ denotes the complex-conjugate. Bussgang-type algorithms have been widely used for blind equalization in digital communications.

4. COMPUTER SIMULATION

In this computer simulation, we considered three QPSK source signals and five sensors which are linearly equispaced with half wavelength separation. Five sensor output signals are mixtures of three QPSK source signals. Each source signals were assumed to have unit variance. The angles of arrival were given by $\theta_1 = -10°, \theta_2 = 45°, \theta_3 = 70°$.

The number of source signals are estimated first by decomposing the covariance matrix of sensor output signals via SVD. The adaptive algorithm (13) was applied with constant learning rate $\eta = .005$. The matrices $\mathbf{\Gamma}$ and $\mathbf{\Psi}$ were set as $\mathbf{\Gamma} = \mathbf{I}$, $\mathbf{\Psi} = .1\mathbf{I}$. Additive white Gaussian noise was added by SNR=20dB.

As an performance measure, the following performance index PI was used. It is defined by

$$PI = \sum_{i=1}^{n}\{(\sum_{k=1}^{n}\frac{|g_{ik}|^2}{\max_j |g_{ij}|^2} - 1) + (\sum_{k=1}^{n}\frac{|g_{ki}|^2}{\max_j |g_{ji}|^2} - 1)\}, \quad (17)$$

where g_{ij} is the (i,j)th element of the global system matrix $\mathbf{G}$ and $\max_j g_{ij}$ represents the maximum value among the elements in the ith row vector of $\mathbf{G}$, $\max_j g_{ji}$ does the maximum value among the elements in the ith column vector of $\mathbf{G}$. When perfect signal separation is carried out, the performance index PI is zero. In practice, it is very small number.

Figure 1 shows five sensor output signals which are linear instantaneous mixtures of three source signals. Three recovered signals are plotted in Figure 2 over the duration [3000,5000]. Performance index is shown in Figure 3.

5. CONCLUSIONS

We have presented a new computationally efficient blind signal separation algorithm. The algorithm has been derived from an information-theoretic approach using the natural gradient. The proposed algorithm was successfully applied to the problem of blind beamforming.

6. REFERENCES

[1] S. Amari. Natural gradient works efficiently in learning. *Neural Computation*, 10:251–276, 1998.

[2] S. Amari, A. Cichocki, and H. H. Yang. A new learning algorithm for blind signal separation. In D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, editors, *Advances in Neural Information Processing Systems*, volume 8, pages 757–763. MIT press, 1996.

[3] A. Bell and T. Sejnowski. An information maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129–1159, 1995.

[4] J. F. Cardoso and B. H. Laheld. Equivariant adaptive source separation. *IEEE Trans. Signal Processing*, 44:3017–3030, 1996.

[5] L. Castedo, C. J. Escudero, and A. Dapena. A blind signal separation method for multiuser communications. *IEEE Trans. Signal Processing*, 45:1343–1348, 1997.

[6] S. Choi. Neural learning algorithms for independent component analysis. *Journal of IEEE Korea Council*, 2(1):24–33, 1998.

[7] S. Choi. Adaptive blind signal separation for multiuser communications: An information-theoretic approach. *Journal of Electrical Engineering and Information Science*, 1998, to appear.

[8] S. Choi and A. Cichocki. A linear feedforward neural network with lateral feedback connections for blind source separation. In *IEEE Signal Processing Workshop on Higher-order Statistics*, pages 349–353, Banff, Canada, 1997.

[9] S. Choi, R. Liu, and A. Cichocki. A spurious equlibria-free learning algorithm for the blind separation of non-zero skewness signals. *Neural Processing Letters*, 7:61–68, 1998.

[10] A. Cichocki, I. Sabala, S. Choi, B. Orsier, and R. Szupiluk. Self-adaptive independent component analysis for sub-gaussian and super-gaussian mixtures with unknown number of source signals. In *International Symposium on Nonlinear Theory and Applications*, pages 731–734, 1997.

[11] A. Cichocki, R. Unbehauen, and E. Rummert. Robust learning algorithm for blind separation of signals. *Electronics Letters*, 43(17):1386–1387, 1994.

[12] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiely & Sons, New York, 1991.

[13] S. Haykin eds. *Blind Deconvolution*. Prentice-Hall, Englewood Cliffs, NJ, 1994.

[14] C. Jutten and J. Herault. Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, 1991.

[15] J. J. Shynk and R. P. Gooch. The constant modulus arrary for cochannel signal copy and direction finding. *IEEE Trans. Signal Processing*, 44:652–660, 1996.

[16] A. Stuart and J. K. Ord. *Kendall's Advanced Theory of Statistics*. Edward Arnold, 1994.

[17] J. R. Treichler and B. G. Agee. A new approach to multipath correction of constant modulus signals. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 31:459–472, 1983.

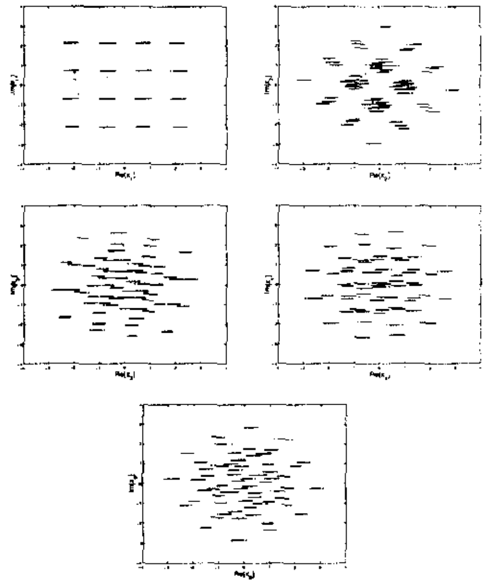

Figure 1: Five different sensor output signals, $x_1(t), \ldots, x_5(t)$.

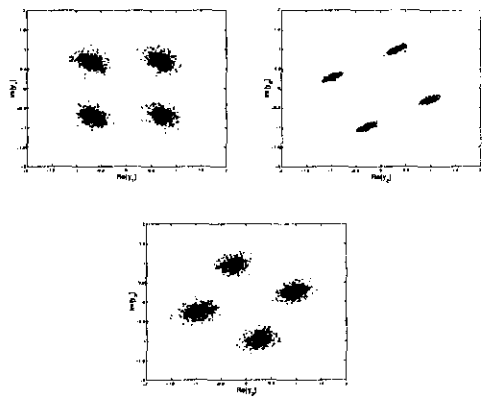

Figure 2: Three recovered signals, $y_1(t)$, $y_2(t)$, and $y_3(t)$.

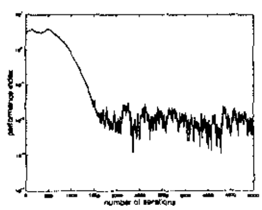

Figure 3: The performance of the adaptation algorithm.