Blind Separation of Multiple Speakers in a Multipath Environment

Russell H. Lambert

TRW Space & Electronics Group
One Space Park
Redondo Beach, CA 90278

Anthony J. Bell

The Salk Institute
Computational Neurobiology Lab
La Jolla, CA 92037

ABSTRACT

We relate information theoretic blind learning methods (infomax) and Bussgang blind equalization methods. The multipath extension of blind source separation methods can be seen in the frequency domain using FIR matrix algebra (matrices of finite impulse response filters). Three forms of Bussgang algorithms are given. The blind serial update method of Cardoso and Laheld is related to the infomax objective of Bell and Sejnowski. The application emphasis is on speech separation. We demonstrate the robustness and power of the new techniques by blindly separating speech signals recorded in a multipath environment.

1. INTRODUCTION

We make an important connection between the information theoretic infomax blind learning methods of Bell and Sejnowski [2], Cardoso [5] and the Bussgang blind equalization methods of [8, 3, 6, 4]. Multiple input and multiple output linear systems are considered in an inverse system estimation problem which does not have access to the input data. We are interested in estimating input data by means of the inverse system. This is a source separation problem with multipath mixtures. A two input and two output system would be written as

$$\underline{\mathbf{H}} = \begin{bmatrix} h_{11} & h_{21} \\ h_{12} & h_{22} \end{bmatrix}. \tag{1}$$

The h_{ij}'s are FIR filters which each represent an acoustic multi-path transfer function from source i to sensor j. Referring to Figure 1, a two-sensor, two-source problem can be written element wise:

$$y_1 = x_1 * h_{11} + x_2 * h_{21} \tag{2}$$
$$y_2 = x_1 * h_{12} + x_2 * h_{22}. \tag{3}$$

A vector $\mathbf{x}$ of source signals passes through unknown system $\underline{\mathbf{H}}$. We equalize the outputs of the system $\mathbf{y}$ using an estimate of the inverse system $\underline{\hat{\mathbf{W}}}$ to get $\hat{\mathbf{x}}$. We assume knowledge of the source pdfs.

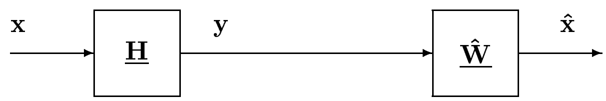

Figure 1: General multichannel channel system.

The linear systems in this problem can be thought of as "FIR matrices" or simply a matrix with FIR (finite impulse response) filters as elements. Linear algebra techniques generally apply to these systems with convolution (or multiplication in the frequency domain) replacing scalar multiplication of matrix elements [10].

We begin with the detailed presentation of more traditional stochastic gradient algorithms, moving to finite difference approximation algorithms which we call serial update forms. We present frequency domain Bussgang forms, some of which offer improved performance over the traditional (stochastic gradient) blind Bussgang form.

2. TRADITIONAL BUSSGANG MULTICHANNEL ALGORITHMS

Given access to N sensors with an assumed number of sources L less than or equal to N, all with unknown direct and cross channels as in (1), we wish to recover all of the unknown sources. We are given only the probability density functions (pdf) of the non-Gaussian and independent sources.

Given sources of known pdfs, for the ith source we can use a traditional blind equalization technique to find the optimal cost function $J_i = E|\hat{x}_i - g_i(\hat{x}_i)|^2$ needed for single channel blind equalization. A multichannel cost function can be made as the simple sum of these single channel blind cost functions:

$$J = J_1 + J_2 + \ldots + J_N, \tag{4}$$

the general form being rooted in Wiener filter theory.

The multichannel Wiener cost function is (see [10])

$$J = \text{tr } E\{(\mathbf{x}-\hat{\mathbf{x}})(\mathbf{x}-\hat{\mathbf{x}})^H\} \quad \text{MLMS} \tag{5}$$

and the blind form is readily seen

$$J = \text{tr } E\{(\hat{\mathbf{x}}-\mathbf{g})(\hat{\mathbf{x}}-\mathbf{g})^H\} \quad \text{MBLMS,} \tag{6}$$

where $\mathbf{g} = [g_1(\hat{x}_1)\ g_2(\hat{x}_2)\ \cdots\ g_N(\hat{x}_N)]^T$, and $g_i(\cdot)$ is the Bussgang nonlinearity– the log derivative of the pdf of x_i.

We present adaptive forms: multichannel (training-based) LMS (MLMS), multichannel blind LMS (MBLMS) and the fast, blind extension (MBRLS). Equation (5) is the cost function for MLMS and MBLMS uses equation (6). The updates are

$$\underline{\hat{\mathbf{W}}} = \underline{\hat{\mathbf{W}}} + \mu(\hat{\mathbf{x}}-\mathbf{x})\underline{\mathbf{y}}^* \quad \text{MLMS} \tag{7}$$

$$\underline{\hat{\mathbf{W}}} = \underline{\hat{\mathbf{W}}} + \mu(\hat{\mathbf{x}}-\mathbf{g})\underline{\mathbf{y}}^* \quad \text{MBLMS} \tag{8}$$

$$\underline{\hat{\mathbf{W}}} = \underline{\hat{\mathbf{W}}} + R^{-1}(\hat{\mathbf{x}}-\mathbf{g})\underline{\mathbf{y}}^* \quad \text{MBRLS.} \tag{9}$$

2.1. STOCHASTIC GRADIENT VERSUS FINITE DIFFERENCE APPROXIMATION

The traditional forms of LMS adaptive algorithms are stochastic gradient methods. The original stochastic gradient idea was presented in a pioneering paper by Robbins and Monro 1951 [11]. One year later, Kiefer and Wolfowitz presented what came to be know as the finite difference approximation method [9]. A simplified form for this type of update is

$$\underline{\hat{\mathbf{W}}} = \underline{\hat{\mathbf{W}}}(1+\mu J). \tag{10}$$

This serial update enjoys eigenvalue disparity robustness because the input data Y does not directly appear. We discuss and present powerful and robust serial update forms of this type.

2.2. FIR MATRICES AS THE WIENER SOLUTION- BATCH (NON ADAPTIVE) SOLUTION OF EQUATION (5)

The batch solution which minimizes this multichannel cost function uses the traditional least square procedure of forming the data matrix and solving for the estimate of the inverse filter:

$$\mathbf{Y}_i = \begin{bmatrix} y_i(1) & y_i(2) & \cdots & y_i(T) & 0 & \cdots & 0 \\ 0 & y_i(1) & y_i(2) & \cdots & y_i(T) & \cdots & 0 \\ . & & . & & & & . \\ . & & & . & & & . \\ . & & & & . & & . \\ 0 & & \cdots & y_i(1) & y_i(2) & \cdots & y_i(T) \end{bmatrix} \tag{11}$$

$$\mathbf{Y} = [\mathbf{Y}_1\ \ \mathbf{Y}_2\ \ \mathbf{Y}_3\ \ \cdots\ \ \mathbf{Y}_N] \tag{12}$$

$$\mathbf{x_i} = [\ x_i(1)\ x_i(2)\ \cdots\ x_i(T)\ ]^T \tag{13}$$

$$\mathbf{X} = [\mathbf{x_1}\ \ \mathbf{x_2}\ \ \mathbf{x_3}\ \ \cdots\ \ \mathbf{x_N}]^H \tag{14}$$

$$\underline{\mathbf{W}} = (\mathbf{Y}\mathbf{Y}^H)^{-1}\mathbf{Y}\mathbf{X}^H, \tag{15}$$

where N is the number of input and output channels and T is the number of data samples collected at the sensors. Here $\mathbf{R} = \mathbf{Y}\mathbf{Y}^H$ is block toeplitz and $\underline{\mathbf{W}}$ is an FIR matrix.

3. THREE FORMS OF DIRECT BUSSGANG ALGORITHM COSTS (DBAC)

Through the use of the FIR matrix algebra, the Bussgang property represented in the frequency domain uncovers useful new forms.

The Bussgang property holds for finite variance iid data. It only strictly holds at convergence, when the estimate of the inverse channel is good and the output data $\hat{x}$ is again iid. It is therefore a good way to monitor convergence of an algorithm for constructing a convergent cost function. Bellini [3] discusses how the Bussgang property can be used to directly obtain the update equation for a Bussgang equalizer (or separator). We start with the standard system of

$$x \rightarrow H \rightarrow y \rightarrow W \rightarrow \hat{x},$$

in the frequency domain,

$$Y = XH \tag{16}$$

$$\hat{X} = YW \tag{17}$$

and the Bussgang property which says that the autocorrelation of the output $\hat{x}$ is equal to the cross correlation of $\hat{x}$ and $g(\hat{x})$, where the nonlinearity is of a special form (the log derivative of the pdf of x).

$$E\{\hat{x}_{i+k}\hat{x}_i\} = E\{\hat{x}_{i+k}g(\hat{x}_i)\} \quad \text{Bussgang Property} \tag{18}$$

$$E\{\hat{X}^*\hat{X}\} = E\{\hat{X}^*\text{fft}\{g(\underline{\hat{x}})\}\} \quad \text{Frequency Domain} \tag{19}$$

3.1. FORM 1: BLIND WIENER COST

Divide both sides of (19) by W^* (or $\underline{\mathbf{W}}^H$ with conjugate transpose for the multichannel case).

$$E\{Y^*\hat{X}\} = E\{Y^*\text{fft}\{g(\underline{\hat{x}})\}\} \quad \text{Bussgang Form 1} \tag{20}$$

This gives the "Direct Bussgang Algorithm Cost" update form 1:

$$W = W + \mu(X - \text{fft}\{g(\underline{\hat{x}})\})Y^*, \quad \text{BLMS} \tag{21}$$

or the traditional Bussgang result in the time domain i.e. $(\hat{x} - g(\hat{x}))\underline{y}^*$.

3.2. FORM 2: INFOMAX/DIRECT MINIMUM ENTROPY DECONVOLUTION

Using the basic relations $Y = XH$ and $\hat{X} = YW$, express the left-hand side of (20) in X.

$$WH^*HE\{X^*X\} = E\{Y^*\text{fft}\{g(\underline{\hat{x}})\}\} \tag{22}$$

Since the source data is independent and $H = W^{-1}$, we get

$$\frac{1}{W^*} = E\{Y^*\text{fft}\{g(\underline{\hat{x}})\}\} \quad \text{Bussgang Form 2} \tag{23}$$

This gives the direct Bussgang update form 2:

$$W = W + \mu(\frac{1}{W^*} - \text{fft}\{g(\underline{\hat{x}})\})Y^*, \quad \text{INFOMAX or DMED} \tag{24}$$

or the DMED (Direct Minimum Entropy Deconvolution) [10] result.

3.3. FORM 3: DIRECT BUSSGANG ALGORITHM COST (NEW FORM)

Starting with the Bussgang property itself (in the frequency domain):

$$E\{\hat{X}^*\hat{X}\} = E\{\hat{X}^*\text{fft}\{g(\underline{\hat{x}})\}\}, \tag{25}$$

defining $R = E\{\hat{X}^*\hat{X}\}$ and $R_g = E\{\hat{X}^*\text{fft}\{g(\underline{\hat{x}})\}\}$ gives the direct Bussgang update form 3:

$$W = W + \mu(R - R_g) \quad \text{DBAC form 3} \tag{26}$$

R_g is also the discrete Fourier transform of

$$E\{\hat{x}_n g(\hat{x}_{n+k})\},$$

and $g(\cdot) = -E|x|^2 \frac{p_x^{'}(x)}{p_x(x)}$.

4. EQUIVARIENT BLIND SERIAL UPDATE METHODS

The Cardoso and Laheld EASI blind serial update cost is,

$$J_{EASI-BSU}(R, R_s) = \frac{(1-R)}{E|x|^2} + \frac{(R_s - R_s^H)}{E|x|^s}$$

Blind Algorithm	Cost
DBAC3 and DBAC3-BSU	$R - R_g$
EASI-BSU	$(1-R) + (R_g - R_g^*)$
AMARI-BSU	$\underline{\mathbf{1}} - \underline{\mathbf{R}}_{\mathbf{g}}$
DMED	$E\{\log\|W\|\ \Phi_x(k_{\hat{x}})\}$
BLMS	$E\{\log\frac{\Phi_x(k_{\hat{x}})}{\Phi_G(k_{\hat{x}})}\}$

Table 1: Table of FIR polynomial matrix Bussgang related algorithm cost functions in the frequency domain.

Blind Alg.	Update
DBAC3	$\Delta W = \mu(R - R_g)$
DBAC3-BSU	$\Delta W = \mu(R - R_g)W$
EASI-BSU	$\Delta W = \mu((1-R) + (R_g - R_g^*))W$
AMARI-BSU [1]	$\Delta W = \mu(1 - R_g)W$
DMED	$\Delta W = \mu(\frac{1}{W^*} - \text{fft}\{g(\hat{x})\}Y^*)$
BLMS	$\Delta W = \mu(\hat{\mathbf{X}} - \text{fft}\{g(\hat{x})\})Y^*$

Table 2: Table of FIR polynomial Bussgang related algorithm updates.

where R is the discrete Fourier transform of $E\{\hat{x}_n\hat{x}_{n+k}\}$ and R_s is the discrete Fourier transform of $E\{\hat{x}_n g(\hat{x}_{n+k})\}$ where $g(\cdot)$ can be of the form $|\hat{x}|^{s-2}\hat{x}$.

The update is (again in the frequency domain)

$$W = W - \mu(\frac{(1-R)}{E|x|^2} + \frac{(R_s - R_s^H)}{E|x|^s})W$$

and enjoys the equivarient/uniform convergence property [5] as well as eigenvalue disparity robustness.

5. MULTICHANNEL BLIND ALGORITHMS

The frequency domain representation is suited to multipath generalizations of vector/matrix algebra. With the algorithms listed in table 1, we can make the extension to matrix forms. Care must given to right/left multiply issues and conjugate $\rightarrow$ conjugate transpose.

6. TESTS

Speech was modeled as Laplace distributed and thirty seconds of data was presented to the adaptive algorithm using a freq. domain overlap and save method as in [7]. A two microphone recording was made in a room approximately 11' by 10' with two persons talking simultaneously. An inverse FIR matrix of dimensions 2 by 2 by 2048 was used with the Amari infomax method. It was noted that all the serial update methods had similar performance. Microphone separation was 1.5'. The separation obtained was approximately 12-15 dB.

Another test was performed using one speech source and one Gaussian source (as presented in [5]). These sources were mixed on the computer using FIR matrices of size 2 by 2 by 256. This artificial mixture could be separated in less than 10 seconds as is demonstrated in figure 3 and in the audio file included on the ICASSP cdrom publication.

The audio of the processed result is presented to show the separation obtained as shown in figures 2 and 3.

7. CONCLUSION

We have shown that the serial update form of Cardoso and Laheld [5] for blind learning is related to the Infomax objective of Bell and Sejnowski [2]. We have presented FIR matrices, the Bussgang property, and frequency domain finite difference approximation/serial update forms as powerful tools for separation of multipath mixtures.

8. REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems 8, MIT Press*, 1996.

[2] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, Vol. 7, pp. 1129–1159, 1995.

[3] S. Bellini, "Bussgang techniques for blind deconvolution and equalization," in *Blind Deconvolution* (S. Haykin, ed.), pp. 8–52, Englewood Cliffs, New Jersey: Prentice Hall, 1994.

[4] A. Benveniste, M. Metivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations*, Berlin: Springer-Verlag, 1987.

[5] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation.." submitted to the IEEE Trans. on S.P., 1994.

[6] D. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis II*, pp. 565–608, New York: Academic Press, 1981.

[7] E. R. Ferrara, "Fast Implementation of LMS Adaptive Filters," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-28, No. 4, pp. 474–475, 1980.

[8] W. C. Gray, *Variable Norm Deconvolution*, Ph.D. dissertation, Stanford, 1979.

[9] J. Kiefer and J. Wolfowitz, "Stochastic estimation of the modulus of a regression function," *Ann. Math. Stat.*, Vol. 23, pp. 462–466, 1952.

[10] R. H. Lambert, *Multichannel Blind Deconvolution: FIR Matrix Algebra and Separation of Multipath Mixtures*, Ph.D. dissertation, University of Southern California, 1996.

[11] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Stat.*, Vol. 22, pp. 4oo–407, 1951.

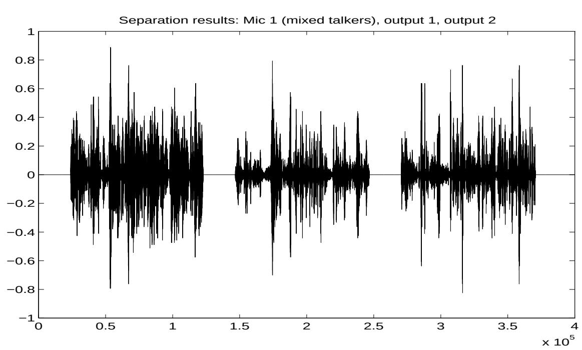

Figure 2: Separation results of true acoustic test.

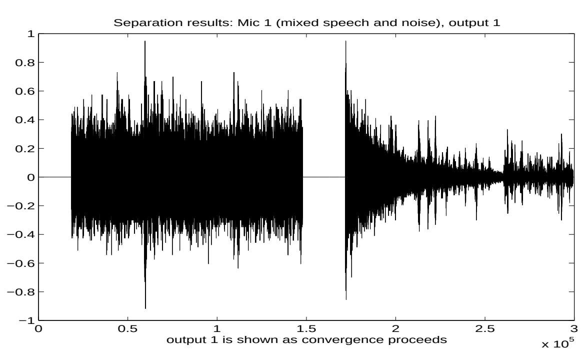

Figure 3: Separation results of speech and Gaussian source.