Converging Evidence of Linear Independent Components in EEG

Lucas Parra[1], and Paul Sajda[2]

[1] Adaptive Image and Signal Processing, Sarnoff Corporation, Princeton, NJ, USA

[2] Department of Biomedical Engineering, Columbia University, NY, USA

***Abstract*— Blind source separation (BSS) has been proposed as a method to analyze multi-channel electroencephalography (EEG) data. A basic issue in applying BSS algorithms is the validity of the independence assumption. In this paper we investigate whether EEG can be considered to be a linear combination of independent sources. Linear BSS can be obtained under the assumptions of non-Gaussian, non-stationary, or non-white independent sources. If the linear independence hypothesis is violated these three different conditions will not necessarily lead to the same result. We show, using 64 channel EEG data, that different algorithms which incorporate the three different assumptions lead to the same results, thus supporting the linear independence hypothesis.**

keywords: Blind source separation (BSS), electroencephalography (EEG), non-Gaussian, non-white, non-stationary.

I. Introduction

Problematic in electroencephalography (EEG) analysis is the non-invertibility of the imaging problem, i.e. a small number of sensors measures a linear combination of a multitude of neuronal and non-neuronal sources. Methods which attempt to invert the sensor readings and recover the underlying sources will be strongly affected by the regularization assumptions that are imposed. This is a well known problem of source reconstruction and localization in EEG. Recently, alternative methods have been proposed that circumvent the notion of localization and source reconstruction altogether and simply look for projections of the EEG data with "interesting" properties, such as independence [1], [2], or maximum discriminability between experimental conditions [3].

For example, blind sources separation (BSS) has been proposed as a method to find statistically independent linear projections of the measured EEG signals[1]. Statistical independence is inherently difficult to assess, typically requiring additional assumptions on the statistics of the sources. Therefore, various BSS algorithms have been used to extract independent projections. For instance, Jung et al. [1] assume that the histogram of a source has long tails (sparsity), and Tang et al. [2] assume that the sources have a colored spectrum. Yet other algorithms are available that assume non-stationary sources [4] or non-Gaussian sources [5]. If the linear mixture assumption is correct, along with the corresponding statistical properties, then all algorithms should in principle give the same result. If any of the conditions is violated, e.g. sparsity, or non-stationarity, it is not clear that the algorithms will give the same results. If, furthermore, the assumption of linearity or independence is not met it is not clear that any of the methods will give even a meaningful result, let alone the same result.

In this paper we investigate the hypothesis that EEG can be thought of as linear combination of independent sources. We apply three different methods which assume that the sources are either non-stationary, non-Gaussian, or non-white. We find that in fact all methods give the same results in support of the linear independence hypothesis. Note that this does not constitute a formal proof. However, we point out that in other domains such as independent components of images the different algorithms do not give the same projections of the data [4].

In the following we review the conditions for source separation that can be derived for non-white, non-stationary, and non-Gaussian sources [6]. These assumptions appear to be well met by EEG data: (1) In the course of an experiment different areas in the brain become active and the resulting EEG activity changes over time (non-stationary). (2) There exists oscillatory activity in different bands as well as slow changes observed through trial averaging in event related potentials. These have long been considered distinct "components" of the EEG spectrum (non-white). (3) The changes in magnitude over time tend to result in non-Gaussian statistics [4].

For these three conditions the solution for the separation problem is given by the simultaneously diagonalization the covariance matrix of the observations and additional cross-statistics whose form depends upon the particular assumptions.

II. Blind source separation based on independence

The problem of recovering sources from their linear mixtures without knowledge of the mixing process has been widely studied. In its simplest form it can be expressed as the problem of identifying the factorization of the N-dimensional observations $\mathbf{x}(t)$ into a mixing channel $\mathbf{A}$ and M-dimensional sources $\mathbf{s}(t)$,

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

In encephalography the unknown matrix $\mathbf{A}$ represents the coupling of a source with each sensor. Sensor geometry as well as the tissue properties will affect its values[2]. The term

[1]Here, the term "source" should not be taken literally in the sense of a localized neuronal source of electrical activity that is observed as skull surface potential differences. The term "source" is simply used because it is common terminology in the context of blind source separation. It should be interpreted as "component" or projection. Indeed, there is nothing that prevents these BSS methods from extracting projections of the data that correspond to extended or disjoint cortical areas

[2]In EEG the relevant property is impedance of the tissue and the precise anatomy; in magneto-encephalography (MEG) magnetic permeability and sensor location and orientation; in functional near infrared imaging (fNIR) absorption and reflection coefficients of the tissue.

blind source separation is frequently used to indicate that no precise knowledge is available on the channel $\mathbf{A}$, nor the sources $\mathbf{s}(t)$. Instead, only general statistical assumptions on the sources or the structure of the channel are made. A large body of work exists for the case that one can assume statistically independent sources. The resulting factorization is known as Independent Component Analysis (ICA) [7]. ICA makes no assumptions on the temporal structure of the sources. In this work we also consider assumptions on the statistical structure of neighboring samples in which cases separation is obtained also for decorrelated sources.

We begin by noting that the matrix $\mathbf{A}$ explains various cross-statistics of the observations $\mathbf{x}(t)$ as an expansion of the corresponding diagonal cross-statistics of the sources $\mathbf{s}(t)$. An obvious example is the time averaged covariance matrix, $\mathbf{R_x} = \sum_t E[\mathbf{x}(t)\mathbf{x}^H(t)]$,[3]

$$\mathbf{R_x} = \mathbf{A}\mathbf{R_s}\mathbf{A}^H \,, \tag{2}$$

where $\mathbf{R_s}$ is diagonal if we assume independent or decorrelated sources. $\mathbf{A}^H$ denotes the complex transpose of $\mathbf{A}$. In the following section we highlight that for non-Gaussian, non-stationary, or non-white sources there exists, in addition to the covariance matrix, other cross-statistics $\mathbf{Q_s}$ which have the same diagonalization property, namely,

$$\mathbf{Q_x} = \mathbf{A}\mathbf{Q_s}\mathbf{A}^H \,. \tag{3}$$

Note that these two conditions alone are already sufficient for source separation. To recover the sources from the observation $\mathbf{x}(t)$ we must find an inverse matrix $\mathbf{W}$ such that $\mathbf{W}^H\mathbf{A} = \mathbf{I}$. In this case we have,

$$\mathbf{s}(t) = \mathbf{W}^H\mathbf{A}\mathbf{s}(t) = \mathbf{W}^H\mathbf{x}(t) \,. \tag{4}$$

After multiplying equations (2) and (3) with $\mathbf{W}$ and equation (3) with $\mathbf{Q_s}^{-1}$ we can combine them to obtain,

$$\mathbf{R_x}\mathbf{W} = \mathbf{Q_x}\mathbf{W}\Lambda \tag{5}$$

where by assumption, $\Lambda = \mathbf{R_s}\mathbf{Q_s}^{-1}$, is a diagonal matrix. This represents a generalized eigenvalue equation that fully determines the unmixing matrix $\mathbf{W}^H$. This of course assumes nonzero diagonal values for $\mathbf{Q_s}$. Equation (5) specifies N column vectors corresponding to at most $M = N$ sources. If $\mathbf{A}$ is of rank $M < N$ only the first M eigenvectors will represent genuine sources while the remaining $N - M$ eigenvectors span the subspace orthogonal to $\mathbf{A}$. This formulation combines therefore subspace analysis and separation in one step.

Incidentally, note that if we choose, $\mathbf{Q_x} = \mathbf{I}$, and assume instead an orthogonal mixing, in fact orthonormal if we set $\mathbf{Q_s} = \mathbf{I}$, the generalized eigenvalue equation reduces to a conventional eigenvalue equation. The solutions are referred to as the Principal Components of the observations $\mathbf{x}$.

In general, however, the mixing $\mathbf{A}$ and the solution for $\mathbf{W}$ are not orthogonal. In the following section we describe several common statistical assumptions used in BSS and show how they lead to different diagonal cross-statistics $\mathbf{Q}$.

[3]The exponent H stand for the hermitian transpose. We use this notation to allow complex pair values as measured in MEG.

III. Statistical Assumptions and the Form of Q

The independence assumption gives a set of conditions on the statistics of recovered sources. All cross-moments of independent variables factor, i.e.

$$E[s_i^u(t)s_j^v(t+\tau)] = E[s_i^u(t)]E[s_j^{*v}(t+\tau)] \,,\; i \neq j \,, \tag{6}$$

where $E[\,.\,]$ represents the mathematical expectation. With (4) these equations define for each choice of $\{u, v, n, \tau\}$ a set of conditions on the coefficients of $\mathbf{W}$ and the observable statistics of $\mathbf{x}(t)$. With a sufficient number of such conditions the unknown parameters of $\mathbf{W}$ can be identified.[4] Depending on the choice this implies that in addition to independence the sources are assumed to be either *non-stationary*, *non-white*, or *non-Gaussian* as discussed in the next three sections.

A. Non-stationary Sources

First, consider second order statistics, $u + v = 2$, and non-stationary sources. The covariance of the observations varies with the time t,

$$\begin{aligned} \mathbf{R_x}(t) &= E[\mathbf{x}(t)\mathbf{x}^H(t)] \\ &= \mathbf{A}E[\mathbf{s}(t)\mathbf{s}^H(t)]\mathbf{A}^H = \mathbf{A}\mathbf{R_s}(t)\mathbf{A}^H \,. \end{aligned} \tag{7}$$

Without restriction we assume zero mean signals. For zero mean signals equation (6) implies that $\mathbf{R_s}(t)$ is diagonal. Therefore, $\mathbf{A}$ is a transformation that expands the diagonal covariance of the sources into the observed covariance at all times. In particular, the sum over time leads to equations (2) regardless of stationarity properties of the signals. Setting, $\mathbf{Q}_x = \mathbf{R_x}(t)$, for any time t, or linear combination of times, will give the diagonal cross-statistics (3) required for the generalized eigenvalue equation (5).

More generally, equation (7) specifies for each t a set of $N(N-1)/2$ conditions on the NM unknowns in the matrix $\mathbf{A}$. The unmixing matrix can be identified by simultaneously diagonalizing multiple covariance matrices estimated over different stationarity times. In the square case, $N = M$, when using the generalized eigenvalue formulation, the N^2 parameters are critically determined by the N^2 conditions in (5). To avoid the resulting sensitivity to estimation errors in the covariances $\mathbf{R_x}(t)$ it is beneficial to simultaneously diagonalize more than two matrices. This is discussed in detail in [8].

B. Non-White Sources

For non-white sources (non-zero autocorrelation) one can use second order statistics in the form of cross-correlations for different time lags τ:

$$\begin{aligned} \mathbf{R_x}(\tau) &= E[\mathbf{x}(t)\mathbf{x}^H(t+\tau)] \\ &= \mathbf{A}E[\mathbf{s}(t)\mathbf{s}^H(t+\tau)]\mathbf{A}^H = \mathbf{A}\mathbf{R_s}(l)\mathbf{A}^H \,. \end{aligned} \tag{8}$$

[4]Note however in (4) that any scaling and permutation that is applied to the coordinates of $\mathbf{s}$ can be compensated by applying the inverse scales and permutations to the rows of $\mathbf{W}$. Conditions (6) do not resolve that ambiguity.

Here we assume that the signals are stationary such that the estimation is independent of t, or equivalently, that the expectation $E[\,.\,]$ includes a time average. Again, (6) implies that $\mathbf{R_s}(\tau)$ is diagonal with the auto-correlation coefficients for lag τ on its diagonal. Equation (8) has the same structure as (3) giving us for any choice of τ, or linear combinations thereof, the required diagonal cross-statistics, $\mathbf{Q_x} = \mathbf{R_x}(\tau)$, to obtain the generalized eigenvalue solution. The identification of mixing channels using eigenvalue equations was first proposed for simultaneous diagonalization of cross-correlations [9]. Time lags τ provide new information if the source signals have distinct auto-correlations. Simultaneous diagonalization for more than two lags has been previously presented, for example in [10].

C. Non-Gaussian Sources

For stationary and white sources different t and τ do not provide any new information. In that case (6) reduces to,

$$E[s_i^u s_j^v] = E[s_i^u]E[s_j^{*v}]\,, \quad i \neq j\,. \tag{9}$$

To obtain sufficient conditions one must include more than second order statistics of the data ($u + m \geq 2$). Consider for example 4th order cumulants expressed in terms of 4th order moments:

$$\begin{aligned} Cum(s_i, s_j^*, s_k, s_l^*) = {} & E[s_i s_j^* s_k s_l^*] - E[s_i s_j^*]E[s_k s_l^*] \\ & - E[s_i s_k]E[s_j^* s_l^*] - E[s_i s_l^*]E[s_j^* s_k]\,. \end{aligned} \tag{10}$$

For Gaussian distributions all 4th order cumulants (10) vanish [11]. In the following we assume non-zero diagonal terms and require therefore non-Gaussian sources. It is straightforward to show using (9) that for independent variables the off-diagonal terms vanish, $i \neq j$: $Cum(s_i, s_j^*, s_k, s_l^*) = 0$, for any k, l, i.e. the 4th order cumulants are diagonal in i, j for given k, l. Any linear combination of these diagonal terms is also diagonal. Following the discussion in [5] we define such a linear combination with coefficients, $\mathbf{M} = \{m_{lk}\}$,

$$c_{ij}(\mathbf{M}) = \sum_{kl} Cum(s_i, s_j^*, s_k, s_l^*) m_{lk}\,. \tag{11}$$

With equation (10) and covariance, $\mathbf{R_s} = E[\mathbf{s}\mathbf{s}^H]$, one can write in matrix notation:

$$\begin{aligned} \mathbf{C_s}(\mathbf{M}) = {} & E[\mathbf{s}^H \mathbf{M} \mathbf{s}\, \mathbf{s}\mathbf{s}^H] - \mathbf{R_s} \mathrm{Trace}\,(\mathbf{M}\mathbf{R_s}) \\ & - E[\mathbf{s}\mathbf{s}^T]\mathbf{M}^T E[\mathbf{s}^* \mathbf{s}^H] - \mathbf{R_s}\mathbf{M}\mathbf{R_s}\,. \end{aligned} \tag{12}$$

We have added the index $\mathbf{s}$ to differentiate from an equivalent definition for the observations $\mathbf{x}$. Using the identity $\mathbf{I}$ this reads:

$$\begin{aligned} \mathbf{C_x}(\mathbf{I}) = {} & E[\mathbf{x}^H \mathbf{x}\, \mathbf{x}\mathbf{x}^H] - \mathbf{R_x} \mathrm{Trace}\,(\mathbf{R_x}) \\ & - E[\mathbf{x}\mathbf{x}^T] E[\mathbf{x}^* \mathbf{x}^H] - \mathbf{R_x}\mathbf{R_x}\,. \end{aligned} \tag{13}$$

By inserting (1) into (13) it is easy to see that,

$$\mathbf{C_x}(\mathbf{I}) = \mathbf{A}\mathbf{C_s}(\mathbf{A}^H\mathbf{A})\mathbf{A}^H\,. \tag{14}$$

Since $\mathbf{C_s}(\mathbf{M})$ is diagonal for any $\mathbf{M}$, it is also diagonal for $\mathbf{M} = \mathbf{A}^H\mathbf{A}$. We find therefore that $\mathbf{A}$ expands the diagonal fourth order statistic to give the corresponding observable fourth order statistic $\mathbf{Q_x}(\mathbf{I})$. This again gives us the required diagonal cross-statistics (3) for the generalized eigenvalue decomposition. This method is instructive but very sensitive to estimation errors and the spread of kurtosis of the individual sources. For robust estimation simultaneous diagonalization using multiple $\mathbf{M}s$ is recommended [5].

IV. Results

Results for real mixtures of EEG signals are shown in figure 1. This data was collected as part of an error-related negativity (ERN) experiment (for details see [3]). To obtain robust estimates of the source directions we simultaneously diagonalized five or more cross-statistics for a given conditions using the diagonalization algorithm by Cardoso and Souloumiac [12]. A segment of 0.8 seconds of data taken from approximately 200 trials was used as a window before and after a visual stimulus was presented. We recover only the 8 strongest components by setting $M = 8$. The result is shown in figure 1. First note that for each of the three different statistical assumptions nearly identical sources are recovered, as evidenced by the similarity in the scalp plots (columns in $\mathbf{A}$) and trial averaged time course of $\mathbf{s}(t)$. This is true for the data from 7 different subjects.

The spatial distribution and time course of the first component indicate a motor activity and somatosensory response. This is distributed over the motor cortex bilaterally, as subjects respond in this experiment with left and right hand button press. It is strongest at about 300-400ms after stimulus onset, which corresponds to the approximate response time of this subject. The second component (occipital) represents the response to the visual stimulus. The fourth component has a fronto-central activity distribution indicative of the hypothesized origin of the ERN in the anterior cingulate [13]. The remaining components remain open for interpretation.

V. Conclusion

In this paper we investigated whether EEG can be seen as a linear combination of independent components. If this assumption is not valid the different algorithms presented for extracting independent components will not give the same results. The EEG components we obtain are consistent with the linear independence hypothesis, and the corresponding assumptions that these components are non-stationary, non-white, and non-Gaussian. This suggests that BSS is an appropriate method for analysis of EEG data.

VI. Acknowledgments

We would like to thank Clay Spence and Craig Fancourt for their help and discussions. This work was supported by the Defense Advanced Research Project Agency (DARPA) under contract N00014-01-C-0482. P.S. was also supported

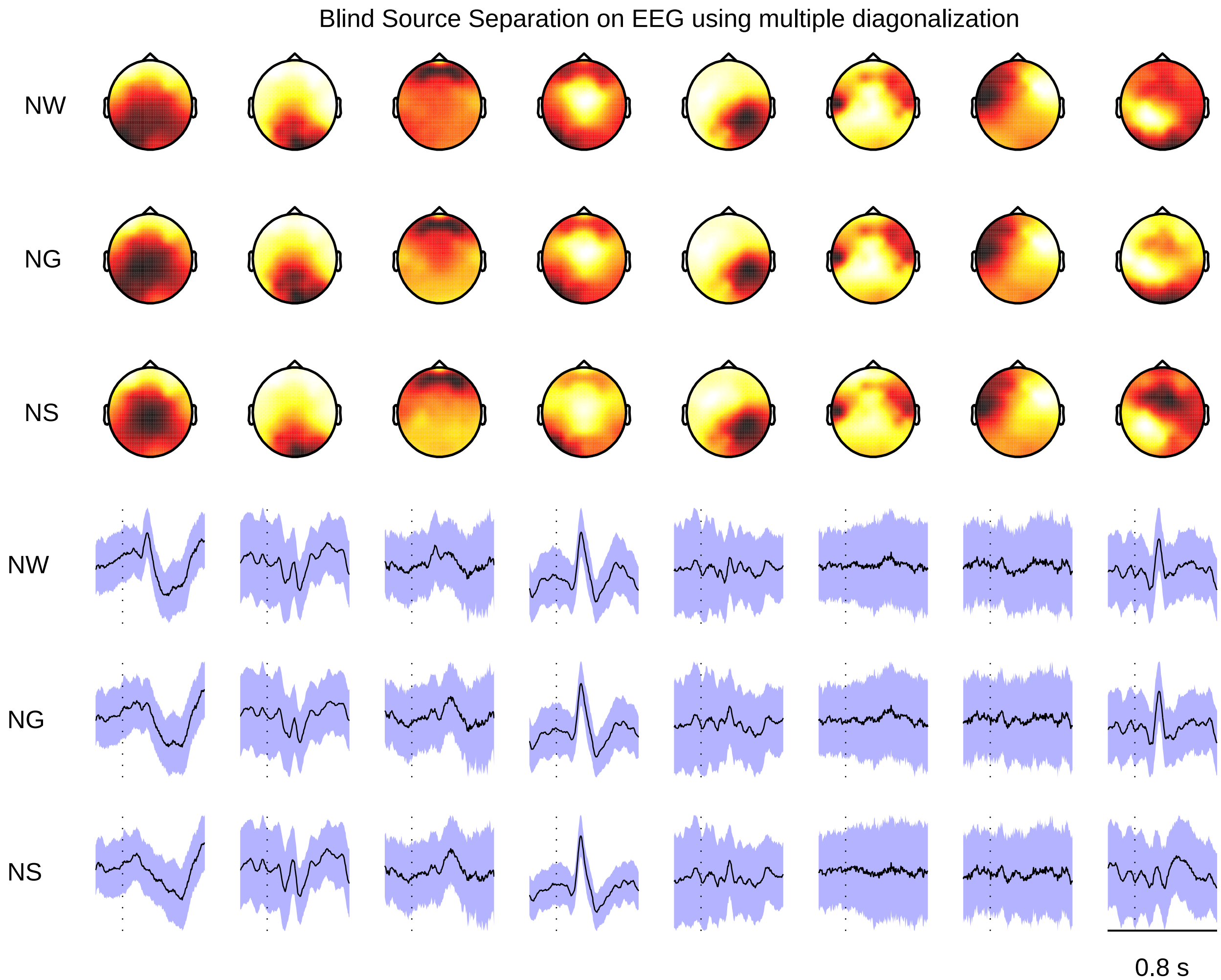

Fig. 1. EEG sensor projections using the different assumptions of non-white (NW), non-stationary (NS) and non-Gaussian (NG) sources. Top three rows show the coupling coefficients (i-th column in $\mathbf{A}$). The bottom three rows show the stimulus locked trail average (solid line) and standard deviation (shaded area) for the recovered components $\mathbf{s}(t)$. Dotted line indicates visual stimulus onset.

by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under grant N00014-01-1-0625 as well as an NSF CAREER Award (BES-0133804).

References

[1] T.-P. Jung, S. Makeig, M. Westerfield, J. Townsend, E. Courchesne, and T. J. Sejnowski, "Independent component analysis of single-trial event-related potentials," *Human Brain Mapping*, vol. 14, no. 3, pp. 168–185, 2001.

[2] A. C. Tang, B. A. Pearlmutter, N. A. Malaszenko, D. B. Phung, and B. C. Reeb, "Independent components of magnetoencephalography: Localization," *Neural Computation*, vol. 14, no. 8, pp. 1827–1858, 2002.

[3] L. Parra, C. Alvino, A. Tang, B. Pearlmutter, N. Yeung, A. Osman, and P. Sajda, "Linear spatial integration for single trial detection in encephalography," *NeuroImage*, vol. 17, no. 1, pp. 223–230, 2002.

[4] L. ParraV and C. Spence, "Separation of non-stationary natural signals," in *Independent Components Analysis, Principles and Practice*, S. Roberts and R. Everson, Eds. Canbridge University Press, 2001, pp. 135–157.

[5] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[6] J.-F. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry," in *Proc. of the ICA 2001 workshop*, San Diego, December 2001.

[7] P. Comon, "Independent comonent analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[8] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of non stationary sources," *IEEE Trans. Signal Processing*, vol. 49, no. 9, pp. 1937–1848, 2001.

[9] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Phys. Rev. Lett.*, vol. 72, no. 23, pp. 3634–3637, 1994.

[10] A. Belouchrani, A. Meraim, J.-F. Cardoso, and E. Moulines, "Second order blind separation of correlated sources," in *Proc. Int. Conf. on Digital Signal Processing*, Cyprus, 1993, pp. 346–351.

[11] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.

[12] J.-F. Cardoso and A. Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM J. Mat. Anal. Appl.*, vol. 17, no. 1, pp. 161–164, Jan. 1996.

[13] S. Dehaene, M. Posner, and D. Tucker, "Localization of a neural system for error detection and compensation," *Psychologic. Sci.*, vol. 5, p. 303 305, 1994.