Journal of Machine Learning Research 4 (2003) 1261-1269 Submitted 10/02; Published 12/03

# Blind Source Separation via Generalized Eigenvalue Decomposition

**Lucas Parra** PARRA@CCNY.CUNY.EDU
*Department of Biomedical Engineering*
*The City College of New York*
*New York NY 10031, USA*

**Paul Sajda** PS629@COLUMBIA.EDU
*Department of Biomedical Engineering*
*Columbia University*
*New York NY, 10027, USA*

**Editors:** Te-Won Lee, Jean-François Cardoso, Erkki Oja and Shun-ichi Amari

## Abstract

In this short note we highlight the fact that linear blind source separation can be formulated as a generalized eigenvalue decomposition under the assumptions of non-Gaussian, non-stationary, or non-white independent sources. The solution for the unmixing matrix is given by the generalized eigenvectors that simultaneously diagonalize the covariance matrix of the observations and an additional symmetric matrix whose form depends upon the particular assumptions. The method critically determines the mixture coefficients and is therefore not robust to estimation errors. However it provides a rather general and unified solution that summarizes the conditions for successful blind source separation. To demonstrate the method, which can be implemented in two lines of matlab code, we present results for artificial mixtures of speech and real mixtures of electroencephalography (EEG) data, showing that the same sources are recovered under the various assumptions.

**Keywords:** blind source separation, generalized eigenvalue decomposition, non-Gaussian, non-white, non-stationary

## 1. Introduction

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

$$\mathbf{X} = \mathbf{A}\mathbf{S}\,. \tag{1}$$

The $T$ columns of the matrices $\mathbf{X}$ and $\mathbf{S}$ represent multiple samples. Often the samples in the data have a specific ordering such as consecutive samples in time domain signals or neighboring pixels in images. Without loss of generality we consider each column of $\mathbf{X}$ and $\mathbf{S}$ as a sample in time. In this case we can write Equation (1) as,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)\,. \tag{2}$$

The term *blind source separation* (BSS) is frequently used to indicate that no precise knowledge is available of either the channel $\mathbf{A}$ or 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) and was first introduced by Comon (1994). ICA makes no assumptions on the temporal structure of the sources. In this paper we consider additional assumptions related to the statistical structure of neighboring samples. In these cases separation is also obtained 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, $\mathbf{R_x} = \sum_t E[\mathbf{x}(t)\mathbf{x}^H(t)]$,

$$\mathbf{R_x} = \mathbf{A}\mathbf{R_s}\mathbf{A}^H\,, \tag{3}$$

where $\mathbf{R_s}$ is diagonal if we assume independent or decorrelated sources. In this paper we consider the general case of complex-valued variables, and $\mathbf{A}^H$ denotes the Hermitian 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,

$$\mathbf{Q_x} = \mathbf{A}\mathbf{Q_s}\mathbf{A}^H\,. \tag{4}$$

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,

$$\mathbf{s}(t) = \mathbf{W}^H\mathbf{A}\mathbf{s}(t) = \mathbf{W}^H\mathbf{x}(t)\,. \tag{5}$$

Let us further assume nonzero diagonal values for $\mathbf{Q_s}$. After multiplying Equations (3) and (4) with $\mathbf{W}$ and Equation (4) with $\mathbf{Q_s}^{-1}$ we combine them to obtain,

$$\mathbf{R_x}\mathbf{W} = \mathbf{Q_x}\mathbf{W}\Lambda\,, \tag{6}$$

where by assumption, $\Lambda = \mathbf{R_s}\mathbf{Q_s}^{-1}$, is a diagonal matrix. This constitutes a generalized eigenvalue equation, where the eigenvalues represent the ratio of the individual source statistics measured by the diagonals of $\mathbf{R_s}$ and $\mathbf{Q_s}$. For distinct eigenvalues Equation (6) fully determines the unmixing matrix $\mathbf{W}^H$ specifying $N$ column vectors corresponding to at most $M = N$ sources. As with any eigenvalue problem the order and scale of these eigenvectors is arbitrary. Hence, the recovered sources are arbitrary up to scale and permutations. This is also reflected in (5), where 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 columns of $\mathbf{W}$. A common choice to resolve these ambiguities is to scale the eigenvectors to unit norm, and to sort them by the magnitude of their generalized eigenvalues. For identical eigenvalues the corresponding sources are determined only up to rotations in the space spanned by the respective columns of $\mathbf{W}$. 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 therefore combines subspace analysis and separation into a single step. It does not, however, address the case of more sources than observations, i.e. $M > N$.

Incidentally, note that if we choose, $\mathbf{Q} = \mathbf{I}$, regardless of the observed statistic, Equation (4) reduces to the assumption of an orthonormal mixing, and the generalized eigenvalue equation reduces to a conventional eigenvalue equation. The solutions are often referred to as the Principal Components of the observations $\mathbf{x}$.

In general, 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 $\mathbf{Q}$. A summary of the different assumptions and choices for $\mathbf{Q}$ is given in Table 1. We also show experimental results for signals that simultaneously satisfy the various assumptions. The results demonstrate that the method recovers the same set of underlying sources for different forms of $\mathbf{Q}$.

## 2. 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.

$$E[s_i^u(t)s_j^{*v}(t+\tau)] = E[s_i^u(t)]E[s_j^{*v}(t+\tau)], \quad i \neq j, \tag{7}$$

where $E[\,.\,]$ represents the mathematical expectation, and $^*$ is the complex conjugate. With (5) these equations define for each choice of $\{u, v, t, \tau\}$ 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 up to scale and permutation. 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.

### 2.1 Non-Stationary Sources

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

$$\mathbf{R_x}(t) = E[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{A}E[\mathbf{s}(t)\mathbf{s}^H(t)]\mathbf{A}^H = \mathbf{A}\mathbf{R_s}(t)\mathbf{A}^H\,. \tag{8}$$

Without restriction we assume throughout this paper zero mean signals.[1] For zero mean signals Equation (7) 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 Equation (3) 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 (4) required for the generalized eigenvalue Equation (6). Note that we have assumed non-stationary power. For sources that are spectrally non-stationary, but maintain a constant power profile, this approach is insufficient.

More generally, Equation (8) 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 (6). 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 by Pham and Cardoso (2001).

1. The mean $E[\mathbf{x}(t)]$ can always be subtracted after estimating it with the same estimation procedure that is used for the correlations.

### 2.2 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 $\tau$:

$$\mathbf{R_x}(\tau) = E[\mathbf{x}(t)\mathbf{x}^H(t+\tau)] = \mathbf{A}E[\mathbf{s}(t)\mathbf{s}^H(t+\tau)]\mathbf{A}^H = \mathbf{A}\mathbf{R_s}(\tau)\mathbf{A}^H . \tag{9}$$

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, (7) implies that $\mathbf{R_s}(\tau)$ is diagonal with the auto-correlation coefficients for lag $\tau$ on its diagonal. Equation (9) has the same structure as (4) giving us for any choice of $\tau$, 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 by Molgedey and Schuster (1994) who suggested simultaneous diagonalization of cross-correlations. Time lags $\tau$ provide new information if the source signals have distinct auto-correlations. Simultaneous diagonalization for more than two lags has been previously presented (Belouchrani et al., 1997).

### 2.3 Non-Gaussian Sources

For stationary and white sources different $t$ and $\tau$ do not provide any new information. In that case (7) reduces to,

$$E[s_i^u s_j^{*v}] = E[s_i^u]E[s_j^{*v}] , \quad i \neq j . \tag{10}$$

To gain 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:

$$Cum(s_i, s_j^*, s_k, s_l^*) = E[s_i s_j^* s_k s_l^*] - E[s_i s_j^*]E[s_k s_l^*] - E[s_i s_k]E[s_j^* s_l^*] - E[s_i s_l^*]E[s_j^* s_k] . \tag{11}$$

For Gaussian distributions all 4th order cumulants (11) vanish (Papoulis, 1991). In the following we assume non-zero diagonal terms and require therefore non-Gaussian sources. It is straightforward to show using (10) 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 by Cardoso and Souloumiac (1993) we define such a linear combination with coefficients, $\mathbf{M} = \{m_{lk}\}$,

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

With Equation (11) and covariance, $\mathbf{R_s} = E[\mathbf{s}\mathbf{s}^H]$, one can write in matrix notation:

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

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:

$$\mathbf{C_x}(\mathbf{I}) = E[\mathbf{x}^H\mathbf{x}\,\mathbf{x}\mathbf{x}^H] - \mathbf{R_x}\mathrm{Trace}\,(\mathbf{R_x}) - E[\mathbf{x}\mathbf{x}^T]E[\mathbf{x}^*\mathbf{x}^H] - \mathbf{R_x}\mathbf{R_x} . \tag{12}$$

By inserting (2) into (12) it is easy to see that,

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

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 therefore find that $\mathbf{A}$ expands the diagonal 4th order statistics to give the corresponding observable 4th order statistic $\mathbf{Q_x}(\mathbf{I})$. This again gives us the required diagonal cross-statistics (4) 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 (Cardoso and Souloumiac, 1993).

## 3. Experimental Results

We first demonstrate, for an artificial mixture, that if a signal satisfies the various statistical assumptions, the different choices for $\mathbf{Q}$ result in the same unmixing. Figure 1 shows an example where two speech signals were artificially mixed with a random mixing matrix $\mathbf{A}$. $10^5$ samples were used in this experiment. Speech satisfies all three statistical assumptions, namely it is non-stationary, non-white and non-Gaussian. The results show that the recovered source orientations are independent, and equivalent, for all three choices of $\mathbf{Q}$.

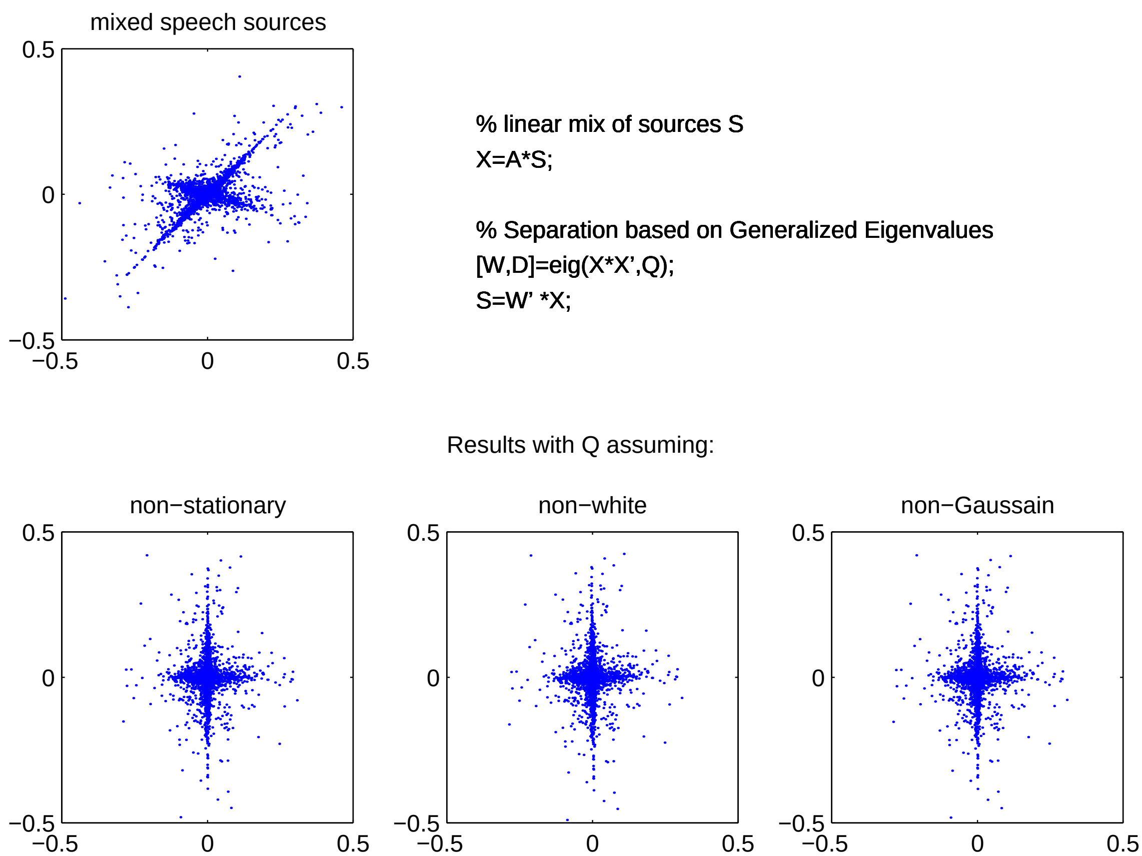


Figure 1: Results for recovery of two speech sources from an instantaneous linear mixture. (top) Scatterplot of mixture and MATLAB code used for mixing/unmixing. (bottom) Scatterplots showing recovered sources for different choices of Q.

Table 1: Summary of procedures for blind source separation using generalized eigenvalue decomposition. Given a $N \times T$ matrix $\mathbf{X}$ containing $T$ samples of $N$ sensor readings generated by, $\mathbf{X} = \mathbf{AS}$, the sources $\mathbf{S}$ are recovered with the MATLAB code: `[W,D]=eig(X*X',Q); S=W'*X;`

| **Assuming sources are** | **Use** | **MATLAB code** | **Details** | **Simple version of** | **References** |
|---|---|---|---|---|---|
| non-stationary and decorrelated | $\mathbf{Q_x} = \mathbf{R_x}(t) = E[\mathbf{x}(t)\mathbf{x}^H(t)]$ | `Q=X(:,1:t)*X(:,1:t)';` | Q is the covariance computed for a separate period of stationarity. Use $t$ in the order of magnitude of the stationarity time of the signal. | simultaneous decorrelation | Molgedey and Schuster (1994), Parra and Spence (May 2000) |
| non-white and decorrelated | $\mathbf{Q} = \mathbf{R_x}(\tau) = E[\mathbf{x}(t)\mathbf{x}^H(t+\tau)]$ | `Q=X(:,1:T-tau)* X(:,tau+1:T)'+ X(:,tau+1:T) * X(:,1:T-tau)';` | Q is the symmetric cross-correlation for time delayed $\tau$. Use *tau* with non-zero autocorrelation in the sources. | simultaneous decorrelation | Weinstein et al. (1993), Parra and Spence (May 2000) |
| non-Gaussian and independent | $\mathbf{Q} = \sum_k Cum(s_i, s_j, s_k, s_k) = E[\mathbf{x}^H\mathbf{x}\mathbf{x}\mathbf{x}^H] - \mathbf{R_x}\mathrm{Trace}(\mathbf{R_x}) - E[\mathbf{x}\mathbf{x}^T]E[\mathbf{x}^*\mathbf{x}^H] - \mathbf{R_x}\mathbf{R_x}$ | `Q=((ones(N,1) * sum(abs(X).^2)).*X)*X' -X*X'*trace(X*X')/T -(X*X.')*conj(X*X.')/T -X*X'*X*X'/T;` | Q is the sum over 4th order cumulants. | ICA | Cardoso and Souloumiac (1993) |
| decorrelated and mixing matrix is orthogonal | $\mathbf{Q} = \mathbf{I}$ | `Q=eye(N);` | Q is the identity matrix. The method reduces to a standard eigenvalue decomposition. | PCA | any linear algebra textbook |

Results for real mixtures of EEG signals are shown in Figure 2. This data was collected as part of an error-related negativity (ERN) experiment (for details see Parra et al., 2002). To obtain robust estimates of the source directions we simultaneously diagonalized five or more cross-statistics, for a given condition, using the diagonalization algorithm by Cardoso and Souloumiac (1996).[2] Sources with the largest generalized eigenvalues are shown in Figure 2. First note that for each of the three different statistical assumptions the same sources are recovered, as evidenced by the similarity in the scalp plots and their averaged time courses. It is clear that the same eight sources were selected among a posible 64 as having the largest generalized eigenvalue (with exeption of the last source in the non-stationary case). In addition, the spatial distribution of the first and fourth sources are readily identified as visual response (occipital) and and ERN (fronto-central) respectively. Fronto-central localization is indicative of the hypothesized origin of the ERN in the anterior cingulate (Dehaene et al., 1994). The consistent results, using different assumptions on the source statistics in combination with their functional neuroanatomical interpretation, is a validation of this approach. We note that others have attempted to recover EEG sources using a supervised method,[3] which attempts to jointly diagonalize spatial sensor covariances for two different conditions, for example left and right motor imagery. This method, termed "common spatial patterns" (CSP) by Ramoser et al. (2000) can be seen as another example of the generalized eigenvalue decomposition, with the matrices $\mathbf{R_x}$ and $\mathbf{Q_x}$ representing the covariances for the two different conditions.

## 4. Conclusion

In this paper we formulate the problem of BSS as one of solving a generalized eigenvalue problem, where one of the matrices is the covariance matrix of the observations and the other is chosen based on the underlying statistical assumptions on the sources. This view unifies various approaches in simultaneous decorrelation and ICA, together with PCA and supervised methods such as CSP. Though in some cases the most straightforward implementation is not robust (e.g. see Table 1), we believe that it is a simple framework for understanding and comparing the various approaches, as well as a method for verifying the underlying statistical assumptions.

## 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 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).

2. For an alternative diagonalization algorithm see also Yeredor (2002).
3. The method is supervised in that one constructs the covariance matrices to diagonalize given the assignment of a window in time ($t_i$) to one of the two conditions ($t_1$ or $t_2$).

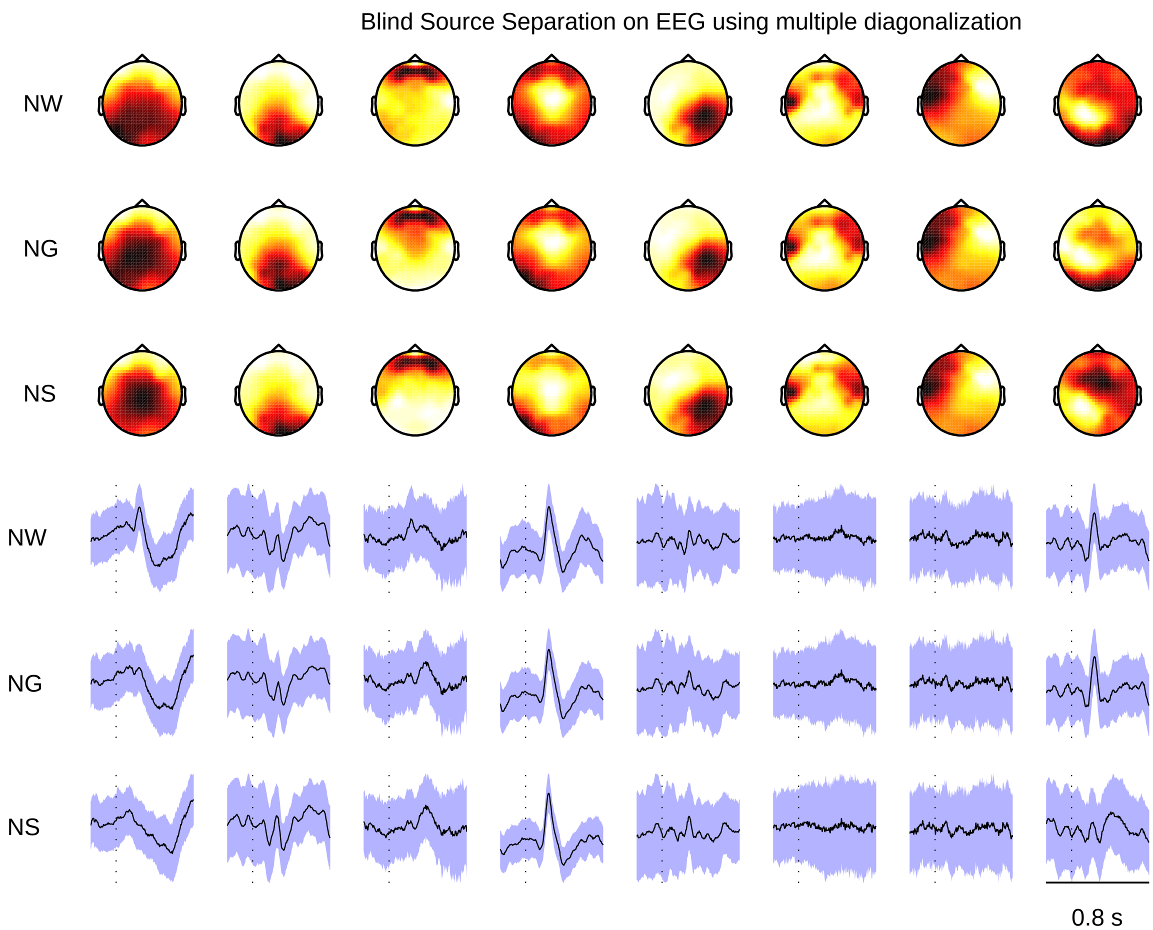


Figure 2: Recovered EEG sources using the different assumptions of non-white (NW), non-stationary (NS) and non-Gaussian sources (NG). Eight sources with the largest magnitude generalized eigenvalues are shown. Sources have been sorted from left to right to match their spatial distributions. Top three rows show sensor scalp plots (columns of **A**). The bottom three rows show trial averaged time course (solid line) and standard deviation (shaded area). Dotted line indicates time of visual stimulus presentation.

## References

A. Belouchrani, A.K. Meraim, J.-F. Cardoso, and E. Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Signal Processing*, 45(2):434–444, February 1997.

J.-F. Cardoso and A. Souloumiac. Blind beamforming for non Gaussian signals. *IEE Proceedings-F*, 140(6):362–370, December 1993.

J.-F. Cardoso and A. Souloumiac. Jacobi angles for simultaneous diagonalization. *SIAM J. Mat. Anal. Appl.*, 17(1):161–164, January 1996.

P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, 1994.

S. Dehaene, M. Posner, and D. Tucker. Localization of a neural system for error detection ad compensation. *Psychologic. Sci.*, 5:303–305, 1994.

L. Molgedey and H. G. Schuster. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.*, 72(23):3634–3637, 1994.

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

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*, 17:223–230, 2002.

L. Parra and C. Spence. Convolutive blind source separation of non-stationary sources. *IEEE Trans. on Speech and Audio Processing*, pages 320–327, May 2000.

D.-T. Pham and J.-F. Cardoso. Blind separation of instantaneous mixtures of non stationary sources. *IEEE Trans. Signal Processing*, 49(9):1937–1848, 2001.

H. Ramoser, J. Mueller-Gerking, and G. Pfurtscheller. Optimal spatial filtering of single trial EEG during imagined hand movements. *IEEE Trans. Rehab. Engng.*, 8(4):441–446, 2000.

E. Weinstein, M. Feder, and A.V. Oppenheim. Multi-channel signal separation by decorrelation. *IEEE Trans. Speech Audio Processing*, 1(4):405–413, Apr. 1993.

A. Yeredor. Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. *IEEE Trans. on Signal Processing*, 50(7):1545–1553, July 2002.