PENALTY FUNCTION BASED JOINT DIAGONALIZATION APPROACH FOR CONVOLUTIVE CONSTRAINED BSS OF NONSTATIONARY SIGNALS

Wenwu Wang[†], *Jonathon Chambers*[†] *and Saeid Sanei*[‡]

† Cardiff School of Engineering, Cardiff University, CF24 0YF, U.K.
Emails: wangw2@cf.ac.uk, chambersj@cf.ac.uk
‡ Dept. of Electronic Engineering, King's College London, WC2R 2LS, U.K.
Email: saeid.sanei@kcl.ac.uk

ABSTRACT

In this paper, we address convolutive blind source separation (BSS) of speech signals in the frequency domain and explicitly exploit the second order statistics (SOS) of nonstationary signals. Based on certain constraints on the BSS solution, we propose to reformulate the problem as an unconstrained optimization problem by using nonlinear programming techniques. The proposed algorithm therefore utilizes penalty functions with the cross-power spectrum based criterion and thereby converts the task into a joint diagonalization problem with unconstrained optimization. Using this approach, not only can the degenerate solution induced by a null unmixing matrix and the over-learning effect existing at low frequency bins be automatically removed, but a unifying view to joint diagonalization with unitary or non-unitary constraint is provided. Numerical experiments verify the validity of the proposed approach.

1. INTRODUCTION

Among open issues in BSS, recovering the independent unknown sources from their linear convolutive mixtures remains a challenging problem. Other than the approaches conventionally developed in the time domain (see [1] and the reference therein), we focus on the operation in the frequency domain in this paper, on the basis of its simpler implementation and better convergence performance [2]-[6]. The representative separation criterion used in the frequency domain is the cross-power spectrum based cost function [4]. However, there are two drawbacks of this criterion, one is the over-learning effect (i.e., large errors of the off-diagonal elements of the covariance matrix) existing at low frequency bins, and the other is the degenerate solution induced by $\mathbf{W}(\omega) = 0$, which also minimizes the criterion.

On the other hand, it has been shown that an appropriate constraint on the separation matrix $\mathbf{W}(k)$ or estimated source signals with special structure, such as invariant norm, orthogonality, geometry information or non-negativity, provides meaningful information to develop a more effective BSS solution, especially for real world signals and practical problems [7]. We will show that the constrained separation problem and the application to convolutive mixtures can be reformulated as an optimization problem with penalty functions using nonlinear programming techniques. Following this, not only can the previous two downsides be removed, but also some current approaches with either unitary constraint or non-unitary constraint have a unifying viewpoint based on such an observation.

This work was supported by the EPSRC of the U.K.

2. SECOND ORDER NONSTATIONARITY

Assume that N source signals are recorded by M microphones (here we are particularly interested in acoustic applications), where $M \geq N$. The output of the j-th microphone is modeled as a weighted sum of convolutions of the source signals corrupted by additive noise, that is

$$x_j(n) = \sum_{i=1}^{N} \sum_{p=0}^{P-1} h_{jip} s_i(n-p) + v_j(n), \tag{1}$$

where h_{jip} is the P-point impulse response from source i to microphone j ($j = 1, \cdots, M$), s_i is the signal from source i, x_j is the signal received by microphone j, v_j is the additive noise, and n is the discrete time index. All signals are assumed zero mean. Using a T-point windowed discrete Fourier transformation (DFT), a time-domain linear convolutive BSS model can be transformed into the frequency domain [2], that is

$$\mathbf{X}(\omega, k) = \mathbf{H}(\omega)\mathbf{S}(\omega, k) + \mathbf{V}(\omega, k) \tag{2}$$

where $\mathbf{S}(\omega, k) = [S_1(\omega, k), \cdots, S_N(\omega, k)]^T$ and $\mathbf{X}(\omega, t) = [X_1(\omega, k), \cdots, X_M(\omega, k)]^T$ are the time-frequency representations of the source signals and the observed signals respectively, k is the discrete time index, N and M are the number of sources and sensors ($M \geq N$). Let $\mathbf{W}(\omega)$ be a weighted pseudo-inverse of $\mathbf{H}(\omega)$, then $\mathbf{Y}(\omega, k) = \mathbf{W}(\omega)\mathbf{X}(\omega, k)$, where $\mathbf{Y}(\omega, k)$ is the time-frequency representation of the estimated source signals. The objective of BSS is to recover *independent* $Y_i(\omega, k)$ ($i = 1, \cdots, N$) from $\mathbf{X}(\omega, k)$, by estimating $\mathbf{W}(\omega)$ using an appropriate criterion.

Non-stationarity of speech signals can be generated in various ways, e.g. variation of the vocal tract filter and glottis, and even detected through higher-order moments. Here we resort to the cross-power spectrum of the output signals at multiple times, i.e.

$$\mathbf{R}_Y(\omega, k) = \mathbf{W}(\omega)[\mathbf{R}_X(\omega, k) - \mathbf{R}_V(\omega, k)]\mathbf{W}^H(\omega) \tag{3}$$

where $\mathbf{R}_X(\omega, k)$ is the covariance matrix of $\mathbf{X}(\omega, k)$, and $\mathbf{R}_V(\omega, k)$ is the covariance matrix of $\mathbf{V}(\omega, k)$ (referring to [4] for its estimation) , and $(\cdot)^H$ denotes the Hermitian transpose operator. A necessary condition to exploit the

SOS conditions for nonstationary signals is to ensure that the $Y_i(\omega, k)$ are mutually uncorrelated. To this end, we need to find a $\mathbf{W}(\omega)$ that (at least approximately) jointly diagonalizes these matrices simultaneously for all time blocks $k, k = 1, \ldots, K$, $\mathbf{R}_Y(\omega, k) \to \mathbf{\Lambda}_C(\omega, k)$, where $\mathbf{\Lambda}_C(\omega, k)$ is a diagonal matrix. An effective criterion for joint diagonalization of $\mathbf{R}_Y(\omega, k)$ is to minimize the following off-diagonal loss function (denoted by problem P_0) [4],

$$P_0: \quad \mathcal{J}(\mathbf{W}(\omega)) = \arg\min_{\mathbf{W}} \sum_{\omega=1}^{T} \sum_{k=1}^{K} \mathcal{F}(\mathbf{W})(\omega, k) \tag{4}$$

where $\mathcal{F}(\mathbf{W}) = \|\mathbf{R}_Y(\omega, k) - diag[\mathbf{R}_Y(\omega, k)]\|_F^2$, $diag(\cdot)$ is an operator which zeros the off-diagonal elements of a matrix, and $||\cdot||_F^2$ is the squared Frobenius norm.

3. PENALTY FUNCTION BASED JOINT DIAGONALIZATION APPROACH

Joint diagonalization is an effective and robust way to employ the average statistical property of signals in the BSS context. Other than the off-diagonal criterion in (4), there are alternatives such as higher order statistics based criterion [9] and the log likelihood based criterion [10]. However, the various joint diagonalization criteria consider the case either with orthogonal (unitary) constraints $\mathbf{W}\mathbf{W}^H = \mathbf{I}$ or with non-orthogonal (non-unitary) constraints $\mathbf{W}\mathbf{W}^H \neq \mathbf{I}$. These constraints have been traditionally addressed as a prewhitening process or optimization on the *Stiefel manifold* for unitary constraint and no *hard whitening* for non-unitary constraint. Effectively, these criteria can be reformulated as the following equality constraint optimization problem,

$$P_1: \quad \min \mathcal{J}(\mathbf{W}(\omega)) \qquad s.t. \quad \mathbf{g}(\mathbf{W}) = \mathbf{0} \tag{5}$$

where $\mathbf{g}(\mathbf{W}) = [g_1(\mathbf{W}), \cdots, g_r(\mathbf{W})]^T : \mathbb{C}^{N\times M} \to \mathbb{R}^r$ denotes the possible constraints, $\mathcal{J} : \mathbb{C}^{N\times M} \to \mathbb{R}^1$, and $r \geq 1$ indicates there may exist more than one constraint. It is unlikely that a generic penalty function exists which is optimal for all constrained optimization problems. Regarding the equality constraint, we introduce a class of exterior penalty functions given as follows:

Definition 1: Let $\mathcal{W}$ be a closed subset of $\mathbb{C}^{N\times M}$. A sequence of continuous functions $\mathcal{U}_q(\mathbf{W}) : \mathbb{C}^{N\times M} \to \mathbb{R}^1$, $q \in \mathbb{N}$, is a sequence of exterior penalty functions for the set $\mathcal{Z}$ if the following three conditions are satisfied:

$$\mathcal{U}_q(\mathbf{W}) = 0, \ \forall\ \mathbf{W} \in \mathcal{W}, \ q \in \mathbb{N}, \tag{6}$$

$$0 < \mathcal{U}_q(\mathbf{W}) < \mathcal{U}_{q+1}(\mathbf{W}), \ \forall\ \mathbf{W} \notin \mathcal{W}, \ q \in \mathbb{N}, \tag{7}$$

$$\mathcal{U}_q(\mathbf{W}) \to \infty, \ \textbf{as} \quad q \to \infty, \forall\ \mathbf{W} \notin \mathcal{W}, \tag{8}$$

Fig. 1 represents a typical example of such a function. Therefore, it is straightforward to follow that a function $\mathcal{U}_q(\mathbf{W}) : \mathbb{C}^{N\times M} \to \mathbb{R}$ defined in (9) forms a sequence of exterior penalty functions for the set $\mathcal{W}$,

$$\mathcal{U}_q(\mathbf{W}) \triangleq \zeta_q \|\mathbf{g}(\mathbf{W})\|_b^\gamma \tag{9}$$

where $q \in \mathbb{N}$, $\gamma \geq 1$, $\zeta_{q+1} > \zeta_q > 0$ and $\zeta_q \to \infty$, as $q \to \infty$, where $b = 1, 2,$or ∞.

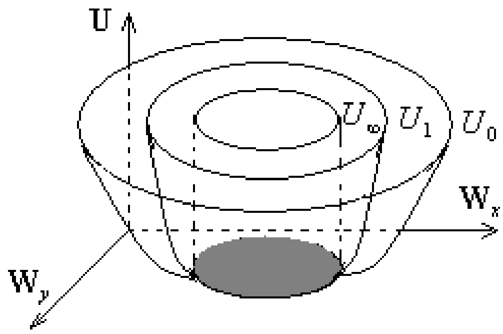

Figure 1: $\mathcal{U}_i(\mathbf{W})$ ($i = 0, 1, \cdots, \infty$) are typical exterior penalty functions, where $\mathcal{U}_0(\mathbf{W}) < \mathcal{U}_1(\mathbf{W}) < \cdots < \mathcal{U}_\infty(\mathbf{W})$ and the shadow area denotes the subset $\mathcal{W}$.

Using a factor vector $\boldsymbol{\kappa}$ to combine the exterior penalty functions (9), our novel general cost function becomes,

$$P_2: \quad \mathfrak{J}(\mathbf{W}(\omega)) = \mathcal{J}(\mathbf{W}(\omega)) + \boldsymbol{\kappa}^T \mathbf{U}(\mathbf{W}(\omega)), \tag{10}$$

where $\mathbf{U}(\mathbf{W}(\omega)) = [\mathcal{U}_1(\mathbf{W}(\omega)), \cdots, \mathcal{U}_r(\mathbf{W}(\omega))]^T$ is a set of penalty functions with desired properties, e.g. $\mathbf{W}(\omega) \neq 0$, $\mathcal{J}(\mathbf{W}(\omega))$ represents an original joint diagonalization criterion, such as in (4), and $\boldsymbol{\kappa} = [\kappa_1, \cdots, \kappa_r]^T$ ($\kappa_i \geq 0$) are the weighting factors. The separation problem is thereby converted into a new unconstrained optimization problem, i.e., $\arg\min_{\mathbf{W}} \mathfrak{J}(\mathbf{W}(\omega))$.

Assuming that $\boldsymbol{\sigma}(\omega) = (\sigma_1(\omega), \cdots, \sigma_r(\omega))^T$ is the vector of perturbations of $\mathbf{g}(\mathbf{W})$, the penalty functions are in the form of (9), e.g. $\mathcal{U}_i(\mathbf{W}(\omega)) = \|\mathbf{g}(\mathbf{W})\|^2$, and the penalty functions share the same parameter κ in (10), then minimization of the criterion with the equality constraint $\mathbf{g}(\mathbf{W}) = \mathbf{0}$ is equivalent to

$$\begin{aligned}
& \min_{\mathbf{W}} \{\mathcal{J}(\mathbf{W}(\omega)) + \kappa \|\mathbf{g}(\mathbf{W})\|^2\} \\
= \ & \min_{\mathbf{W},\boldsymbol{\sigma}} \{\mathcal{J}(\mathbf{W}(\omega)) + \kappa \|\boldsymbol{\sigma}(\omega)\|^2\} \\
= \ & \min_{\boldsymbol{\sigma}} \{\kappa \|\boldsymbol{\sigma}(\omega)\|^2\} + \min_{\mathbf{W}} \{\mathcal{J}(\mathbf{W}(\omega)) : \mathbf{g}(\mathbf{W}) = \boldsymbol{\sigma}\} \\
= \ & \min_{\boldsymbol{\sigma}} \{\kappa \|\boldsymbol{\sigma}(\omega)\|^2 + \vartheta(\boldsymbol{\sigma})\}
\end{aligned} \tag{11}$$

where $\vartheta(\boldsymbol{\sigma})$ is the perturbation function defined as the optimal value function for the equality constraint problem. Equation (11) implies that by adding the term $\kappa \|\boldsymbol{\sigma}(\omega)\|^2$, an attempt is made to convexify $\vartheta(\boldsymbol{\sigma})$ as κ increases, i.e., as $\kappa \to \infty$, the perturbation value $\boldsymbol{\sigma}$ approaches zero. This indicates that the constraint can be satisfied while the criterion is minimized simultaneously. However, it is unlikely in practical situations due to the numerical problem and a finite penalty may generate a satisfactory and well-understood solution (see section 4). It should be noted that, due to the limited space, we have omitted proofs of the given theorems and theoretical analysis of the proposed approach (We leave some more details in [8]).

Under the penalty function based framework, either a unitary or a non-unitary constraint problem can be deemed as an example of a penalized unconstrained optimization problem, whereas the forms of the penalty functions can be designed accordingly. Assuming that the cost function $\mathfrak{J}(\mathbf{W})$ is twice-differentiable and calculating the perturbation matrix $\mathbf{\Delta}$ of $\mathbf{W}$, we have the following Hessian matrix

$$\nabla^2\mathfrak{J}(\mathbf{W}) \triangleq \nabla^2\mathcal{F}(\mathbf{W}) + \kappa\frac{\partial\mathcal{U}(\mathbf{W})}{\partial\mathbf{W}^*}\nabla^2 g_i(\mathbf{W})$$
$$+\kappa\frac{\partial^2\mathcal{U}(\mathbf{W})}{\partial\mathbf{W}^*}\nabla g_i(\mathbf{W})\nabla g_i(\mathbf{W})^T \qquad (12)$$

It follows that as $\kappa \to \infty$, $\mathbf{W}$ will approach the optimum $\hat{\mathbf{W}}$. If $\hat{\mathbf{W}}$ is a regular solution to the constrained problem, then there exists unique Lagrangian multipliers $\bar{\lambda}_i$ such that $\frac{\partial\mathcal{U}(\hat{\mathbf{W}})}{\partial\mathbf{W}^*} + \sum\bar{\lambda}_i\nabla g_i(\mathbf{W}) = \mathbf{0}$ [12]. This means $\kappa\frac{\partial\mathcal{U}(\mathbf{W})}{\partial\mathbf{W}^*} \to \bar{\lambda}_i$ as $\mathbf{W} \to \hat{\mathbf{W}}$. The first two terms in (12) approach the Hessian of the Lagrangian function $L(\mathbf{W}) = \mathcal{F}(\mathbf{W}) + \sum\bar{\lambda}_i g_i(\mathbf{W})$. Considering the last term in (12), it is straightforward to show that as $\kappa \to \infty$, $\nabla^2\mathfrak{J}(\mathbf{W})$ has some eigenvalues approaching ∞, depending on the number of the constraints, and the other eigenvalues approach finite value. The infinite eigenvalues will lead to an ill-conditioned computation problem. Let ϵ be the step size in the adaptation, then in the presence of nonlinear equality constraints, the direction $\mathbf{\Delta}$ may cause any reduction of $\mathcal{F}(\mathbf{W}+\epsilon\mathbf{\Delta})$ to be shifted by $\kappa\mathcal{U}(\mathbf{W}+\epsilon\mathbf{\Delta})$. This requires the step size to be very small to prevent the ill-conditioned computation problem induced by large eigenvalues with a trade-off of having a lower convergence rate. This is verified by simulations in section 4. The eigenstructure provides the guideline of the selection of penalty parameters in practical applications.

4. NUMERICAL EXPERIMENTS

To evaluate the performance of the proposed method, we use an exterior penalty function $\|diag[\mathbf{W}(\omega) - \mathbf{I}]\|_F^2$, and a variant of gradient adaptation based on $\Delta\mathbf{W}(\omega) = \kappa diag[\mathbf{W}(\omega) - \mathbf{I}]\mathbf{W}(\omega)$ [8] [11]. We use the filter length constraint method as in [4] to address the permutation problem which allows us to compare the performance of the proposed method with that in [4] (another SOS based joint diagonalization approach). A system with two inputs and two outputs (TITO) ($N = M = 2$) is considered for simplicity. Two real speech signals are used in the following experiments, which are available from [13]. We artificially mix the two sources by a non-minimum phase system with $H_{11}(z) = 1+1.0z^{-1} - 0.75z^{-2}$, $H_{12}(z) = 0.5z^{-5} + 0.3z^{-6} + 0.2z^{-7}$, $H_{21}(z) = -0.7z^{-5} - 0.3z^{-6} - 0.2z^{-7}$, and $H_{22}(z) = 0.8 - 0.1z^{-1}$. Other parameters are set to be $D = 7$, $T = 1024$, $K = 5$, $\mathbf{W}_0(\omega) = \mathbf{I}$, and the step size in gradient adaptation is $\mu = 1$. We applied the short FFT to the separation matrix and the cross-correlation of the input data. Fig. 2 shows the convergence behavior by incorporating penalty functions, where we can clearly see that by increasing the penalty coefficient κ, we can not only approach the constraint in a quicker way, but also attain a better convergence performance. However, from Fig. 2, we also see that a large penalty κ (e.g. $\kappa = 10$) will introduce the ill-conditional calculation under a common step size. This effect can be removed by reducing the step size. Theoretically, due to the independence assumption, the cross-correlation of the output signals in (3) should approximately approach zero. Fig. 3 (a) (the imaginary part is not plotted due to its similar behavior) shows that it is true at most frequency bins, however there exists over-learning at very low frequency bins. From Fig. 3 (b)-(d), we see that the over-learning effect can be significantly reduced using penalty functions. We further resort to the mean square error $MSE(dB) = 10\log_{10}\{\frac{1}{N}\sum_{i=1}^{N} E\left[|y_i(k) - s_i(k)|^2\right]\}$ and set $\mu = 0.06$ (This allows the selection of a larger penalty as compared to former simulations). Other parameters are the same as the previous experiments. The estimation error is plotted in Fig. 4 (upper) in dB scale. From this simulation, we can clearly see that the separation performance is improved with an increasing penalty, and can reach up to 16dB when $\kappa = 10$. However, with further increasing of the penalties, the separation per-

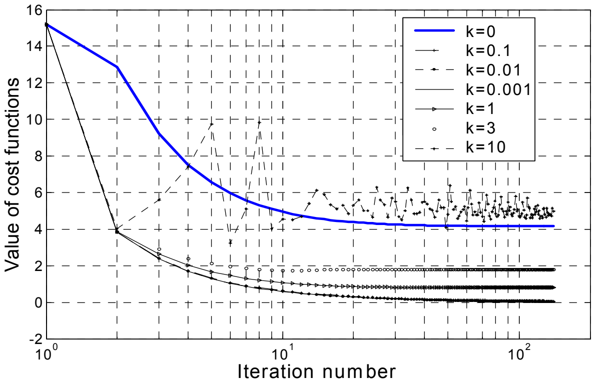

Figure 2: Convergence performance of the new criterion (10) ($\kappa \neq 0$) and the criterion (4) ($\kappa = 0$, cf [4]).

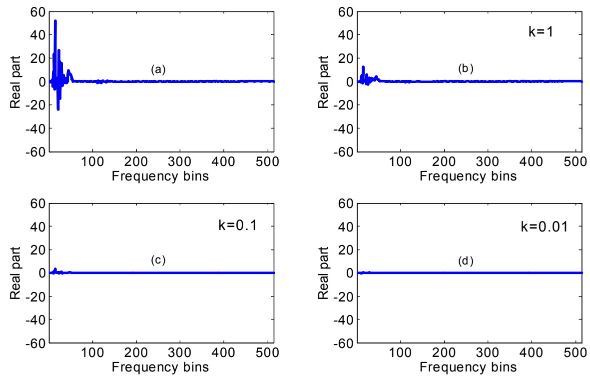

Figure 3: The values (real part) of the off-diagonal elements of the cross-correlation matrices $\mathbf{R}_Y(\omega, k)$ at each frequency bin. (a) corresponds to the criterion (4) (cf [4]); (b)-(d) corresponds to the proposed criterion (10).

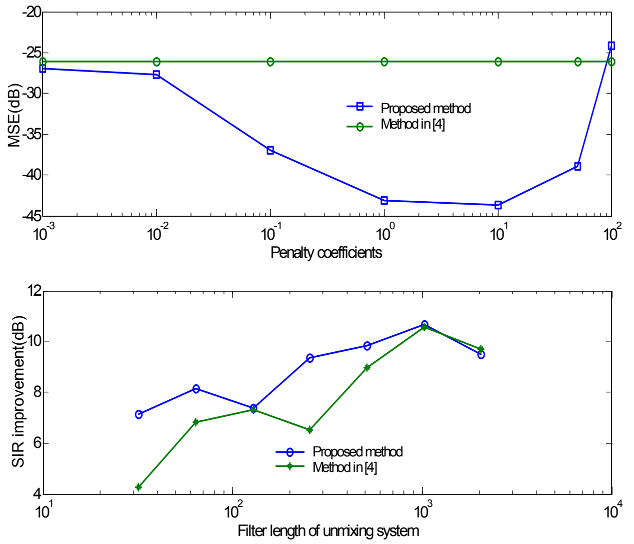

Figure 4: Performance measurement using MSE (upper) for artificially mixing system and SIR improvement (below) for a simulated highly reverberant room environment.

formance may degrade due to the fluctuation.

Another method to quantify the performance is using signal to interference ratio (SIR) as defined in [4],

$$SIR[H,s] = 10\log\frac{\sum_\omega \sum_i |H_{ii}(\omega)|^2 \left\langle |s_i(\omega)|^2 \right\rangle}{\sum_\omega \sum_{i\neq j} |H_{ij}(\omega)|^2 \left\langle |s_j(\omega)|^2 \right\rangle}$$

In this experiment, we use the *roommix* function available from [14] by Westner to simulate a highly reverberant environment. The simulated room is assumed to be a $10m \times 10m \times 10m$ cube. Wall reflections are computed up to the fifth order, and an attenuation by a factor of two is assumed at each wall bounce. We set the position matrices of two sources and two sensors respectively as [2 2 5; 8 2 5], [3 8 5; 7 8 5]. The parameters are set the same as in the last simulation. The SIR is plotted in Fig. 4 (below), which indicates that the separation quality increases with the filter length of the separation system. The performance is highly related to the filter length and it is especially obvious when the filter length becomes long. Fig. 4 (below) also indicates that incorporating a suitable penalty can increase the SIR. Additionally, the penalty function may change the local minima which can be observed from Fig. 4 (below), as the SIR plot is not *smooth* and the increased amplitude is not consistent between the two methods.

5. CONCLUSION

A new joint diagonalization criterion for separating convolutive mixtures of nonstationary source signals in the frequency domain has been presented. Using the cross-power spectrum and nonstationarity of speech signals, penalty functions are accommodated within the conventional criterion. This automatically removes the degenerate solution induced by the null unmixing matrix and over-learning effect at low frequency bins, and hence improves the separation performance. The new criterion transforms the separation problem into a joint diagonalization problem with unconstrained optimization which provides a way of unifying the unitary and non-unitary constraint joint diagonalization methods. The MSE and SIR measurement show that, with a suitable function, the proposed approach has superior separation performance for the convolutive mixtures of speech signals.

REFERENCES

[1] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. John Wiley, Chichester, Apr. 2002.

[2] P. Smaragdis, “Blind separation of convolved mixtures in the frequency domain,” *Neurocomputing*, vol.22, pp. 21–34, 1998.

[3] K. Rahbar and J. Reilly,"Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," *Proc. ICASSP*, May, 2001.

[4] L. Parra and C. Spence, "Convolutive blind source separation of nonstationary sources," *IEEE Trans. on Speech and Audio Proc.*, pp. 320–327, May 2000.

[5] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Speech Audio Processing*, vol. 11, no. 2, pp. 109-116, Mar. 2003.

[6] W. Wang, J. A. Chambers, and S. Sanei, "A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain," *Proc. ICA 2003*, Nara, Japan, Apr. 1-4, 2003.

[7] A. Cichocki and P. Georgiev, "Blind source separation algorithms with matrix constraints," *IEICE Trans. Fundamentals*, vol. E86-A, no. 1, Jan. 2003.

[8] W. Wang, J. A. Chambers, and S. Sanei, "Penalty function based joint diagonalization approach for convolutive blind separation of nonstationary sources," *IEEE Trans. Sig. Proc.*, submitted in Sept. 2003, subject to revision in Jan. 2004.

[9] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proc. F*, vol. 140, pp. 362-370, 1993.

[10] D. T. Pham, "Joint approximate diagonalization of positive definite Hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, pp. 1136-1152, 2001.

[11] M. Joho and H. Mathis, “Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation,” *Proc. SAM*, Rosslyn, VA, 4-6, Aug. 2002.

[12] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming Theory and Algorithms*, 2nd ed. Wiley, 1993.

[13] J. Anemüller. http://medi.uni-oldenburg.de/ members/ane.

[14] Westner. http://www.media.mit.edu/~westner