Penalty Function Approach for Constrained Convolutive Blind Source Separation

Wenwu Wang[1], Jonathon A. Chambers[1], and Saeid Sanei[2]

[1] Cardiff School of Engineering, Cardiff University
Queen's Building, Cardiff, CF24 0YF, UK
wenwu.wang@ieee.org,chambersj@cf.ac.uk
[2] Division of Engineering, King's College London
Strand, London, WC2R 2LS, UK
saeid.sanei@kcl.ac.uk

Abstract. A new approach for convolutive blind source separation (BSS) using penalty functions is proposed in this paper. Motivated by nonlinear programming techniques for the constrained optimization problem, it converts the convolutive BSS into a joint diagonalization problem with unconstrained optimization. Theoretical analyses together with numerical evaluations reveal that the proposed method not only improves the separation performance by significantly reducing the effect of large errors within the elements of covariance matrices at low frequency bins and removes the degenerate solution induced by a null unmixing matrix, but also provides an unified framework to constrained BSS.

1 Introduction

Among open issues in BSS, recovering the independent unknown sources from their linear convolutive mixtures remains a challenging problem. To address this problem, we focus on the operation in the frequency domain [2]-[5] rather than the approaches developed in the time domain (see [1] for example), due to its simpler implementation and better convergence performance. Using a discrete Fourier transformation (DFT), a time-domain linear convolutive BSS model can be transformed into the frequency domain [2], i.e., $\mathbf{X}(\omega, k) = \mathbf{H}(\omega)\mathbf{S}(\omega, k) + \mathbf{V}(\omega, k)$, where $\mathbf{S}(\omega, k)$ and $\mathbf{X}(\omega, t)$ are the time-frequency vectors of the N source signals and the M observed signals respectively ($M \geq N$), k is the discrete time index. The objective of BSS is to find $\mathbf{W}(\omega)$ which is a weighted pseudo-inverse of $\mathbf{H}(\omega)$, so that the elements of estimated sources $\mathbf{Y}(\omega, k)$ are mutually independent, where $\mathbf{Y}(\omega, k) = \mathbf{W}(\omega)\mathbf{X}(\omega, k)$. To this end, we exploit the statistical nonstationarity of signals by using the following criterion [4]

$$\mathcal{J}(\mathbf{W}(\omega)) = \arg\min_{\mathbf{W}} \sum\nolimits_{\omega=1}^{T} \sum\nolimits_{k=1}^{K} \mathcal{F}(\mathbf{W})(\omega, k), \tag{1}$$

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

C.G. Puntonet and A. Prieto (Eds.): ICA 2004, LNCS 3195, pp. 661–668, 2004.

times, i.e., $\mathbf{R}_Y(\omega,k) = \mathbf{W}(\omega)[\mathbf{R}_X(\omega,k) - \mathbf{R}_V(\omega,k)]\mathbf{W}^H(\omega)$, where $\mathbf{R}_X(\omega,k)$ and $\mathbf{R}_V(\omega,k)$ are respectively the covariance matrices of $\mathbf{X}(\omega,k)$ and $\mathbf{V}(\omega,k)$, and $(\cdot)^H$ denotes the Hermitian transpose operator. Minimization of this criterion is equivalent to joint diagonalization of $\mathbf{R}_Y(\omega,k)$ for all time blocks k, $k = 1,\ldots,K$, that is, $\mathbf{R}_Y(\omega,k)$ will become a diagonal matrix $\mathbf{\Lambda}_C(\omega,k)$ due to the independence assumption [4].

However, there exists degenerate effect at low frequency bins induced by the large errors within the elements of covariance matrices (see more details in Section 4). Moreover, a null unmixing matrix $\mathbf{W}(\omega)$ also minimizes the criterion and potentially leads to a degenerate solution. In this paper, we propose a new approach based upon penalty functions, which is motivated by nonlinear programming techniques for constrained optimization. Essentially, we reformulate of the constrained BSS discussed in Section 2 as an unconstrained optimization problem using penalty functions. We will show that this approach provides an effective way of overcoming the aforementioned problems and a framework of unifying the joint diagonalization with unitary and non-unitary constraint. The remainder of this paper is organized as follows. Constrained BSS problem is briefly discussed in Section 2. The penalty function approach is introduced in Section 3, which includes its mathematical formulation, convergence behavior, numerical stability, and algorithm summary. The experimental results and the conclusion are respectively given in Section 4 and Section 5.

2 Constrained Blind Source Separation

Although BSS employs the least possible information pertaining to the sources and the mixing system, there exists useful information in practice for developing various effective algorithms to separate the mixtures, such as the geometrically constrained parameter space with $\|\mathbf{w}\| = 1$ exploited in [12], orthonormal constraint on $\mathbf{W}(k)$ i.e., $\mathbf{W}(k)\mathbf{W}^T(k) = \mathbf{I}$ used in [11] and [13], a non-holonomic constraint on $\mathbf{W}(k)$ maintained by a natural gradient procedure in [15], the source geometric information constraint exploited in [17] and a non-negative constraint in [16]. The orthonormal constraint has also been addressed as the optimization problem on the *Stiefel manifold* or *Grassman manifold* in [14] [13]. A recent contribution in [10] justifies that imposing an appropriate constraint on the separation matrix $\mathbf{W}(k)$ or the estimated source signals with special structure provides meaningful information to develop a more effective BSS solution for practical applications.

3 Penalty Function Approach

Effectively, a constrained BSS problem can be reformulated as the following equality constrained optimization problem,

$$P_1: \quad \min \mathcal{J}(\mathbf{W}(\omega)) \qquad s.t. \quad \mathbf{g}(\mathbf{W}) = \mathbf{0} \tag{2}$$

where $\mathbf{g}(\mathbf{W}) = [g_1(\mathbf{W}), g_2(\mathbf{W}), \cdots, g_r(\mathbf{W})]^T$: $\mathbb{C}^{N\times M} \rightarrow \mathbb{R}^r$ denotes the possible constraints, $\mathcal{J}: \mathbb{C}^{N\times M} \rightarrow \mathbb{R}^1$, and $r \geq 1$ indicates there may exist more than one constraint. In the BSS context, $\mathcal{J}(\mathbf{W}(\omega))$ denotes the various joint diagonalization criteria (1), and $\mathbf{g}(\mathbf{W})$ represents various constraints such as unitary constraint $\mathbf{W}\mathbf{W}^H = \mathbf{I}$ or non-unitary constraint $\mathbf{W}\mathbf{W}^H \neq \mathbf{I}$. To convert (2) into an unconstrained optimization problem, we have to define suitable penalty functions since it is unlikely to find a generic penalty function 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: (i) $\mathcal{U}_q(\mathbf{W}) = 0, \forall\, \mathbf{W} \in \mathcal{W},\ q \in \mathbb{N}$; (ii) $0 < \mathcal{U}_q(\mathbf{W}) < \mathcal{U}_{q+1}(\mathbf{W}), \forall\, \mathbf{W} \notin \mathcal{W},\ q \in \mathbb{N}$; (iii) $\mathcal{U}_q(\mathbf{W}) \to \infty,\ as\ \ q \to \infty, \forall\, \mathbf{W} \notin \mathcal{W}$.

Fig. 1 shows a typical example of such a function. According to Definition 1, it is straightforward to show that a function $\mathcal{U}_q(\mathbf{W}) : \mathbb{C}^{N\times M} \to \mathbb{R}$ defined as follows forms a sequence of exterior penalty functions for the set $\mathcal{W}$,

$$\mathcal{U}_q(\mathbf{W}) \triangleq \zeta_q \left\|\mathbf{g}(\mathbf{W})\right\|_b^\gamma \tag{3}$$

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

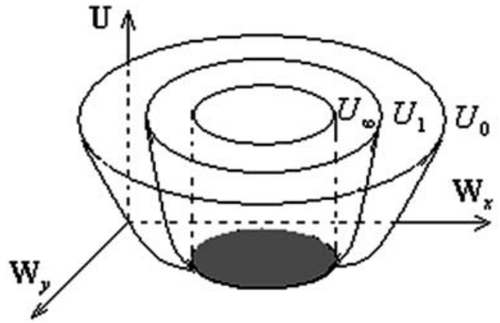

Fig. 1. $\mathcal{U}_i(\mathbf{W})$ $(i = 0, 1, \cdots, \infty)$ are typical exterior penalty functions, where $\mathcal{U}_0(\mathbf{W}) < \mathcal{U}_1(\mathbf{W}) < \cdots < \mathcal{U}_\infty(\mathbf{W})$ and the shadow area denotes the subset $\mathcal{W}$.

After incorporating penalty functions, the new cost function becomes,

$$P_2 : \quad \mathfrak{J}(\mathbf{W}(\omega)) = \mathcal{J}(\mathbf{W}(\omega)) + \boldsymbol{\kappa}^T\mathbf{U}(\mathbf{W}(\omega)), \tag{4}$$

where $\mathbf{U}(\mathbf{W}(\omega)) = [\mathcal{U}_1(\mathbf{W}(\omega)), \cdots, \mathcal{U}_r(\mathbf{W}(\omega))]^T$, whose elements take the form (3) which can be designed properly so that $\mathbf{W}(\omega) \neq 0$, $\mathcal{J}(\mathbf{W}(\omega))$ can be the form of (1), and $\boldsymbol{\kappa} = [\kappa_1, \cdots, \kappa_r]^T$ $(\kappa_i \geq 0)$ are the weighted factors.

3.1 Convergence Behavior

The separation problem is thereby converted into an unconstrained optimization problem using joint diagonalization, i.e., $\min \mathfrak{J}(\mathbf{W}(\omega))$. The equivalence between (4) and (2), together with their critical points obey the following theorems.

Theorem 1: Let the set $\mathcal{W}$ be a closed subset of $\mathbb{C}^{N\times M}$ which satisfies $\mathbf{g}(\mathbf{W}) = \mathbf{0}$, the set $\mathbf{B}(\hat{\mathbf{W}}, \rho)$ be denoted by $\{\mathbf{W} \in \mathbb{C}^{N\times M} | \left\| \mathbf{W} - \hat{\mathbf{W}} \right\|_F \leq \rho\}$, where $\rho > 0$. Suppose that the following assumptions are satisfied: (a) If there exists a point $\mathbf{W}^*$ such that the level set $\{\mathbf{W} \in \mathbb{C}^{N\times M} | \quad f(\mathbf{W}) \leq f(\mathbf{W}^*)\}$ is compact, and (b) there exists an optimal solution $\hat{\mathbf{W}}$ for problem (2) such that for $\forall \rho > 0$, $\mathbf{B}(\hat{\mathbf{W}}, \rho) \cap \mathcal{W}$ is not empty. Then: (i) For any given $i \in \mathbb{N}$, let $\mathbf{W}_i$ be an optimal solution to problem P_2 in (4) at the ith trial. Then any accumulation point $\hat{\mathbf{W}}$ of $\mathbf{W}_i (i = 0 \rightarrow \infty)$, is an optimal solution to problem P_1 in (2). (ii) For every $i \in \mathbb{N}$, let $\mathbf{W}_i$ be a strict local minimizer for problem P_2 in (4) at the ith trial, so that for some $\rho_i > 0$, $f_i(\mathbf{W}_i) < f_i(\mathbf{W})$ for all $\mathbf{W} \in \mathbf{B}(\mathbf{W}_i, \rho_i) = \{\mathbf{W} \in \mathbb{C}^{N\times M} | \quad \|\mathbf{W} - \mathbf{W}_i\|_F \leq \rho\}$. If $\hat{\mathbf{W}}$ is an accumulation point of $\mathbf{W}_i (i = 0 \rightarrow \infty)$, and there exists a $\rho > 0$, such that $\rho_i \geq \rho$, for all $i \in \mathbb{N}$, then $\hat{\mathbf{W}}$ is a local minimizer for the problem P_1 in (2).

The proof of this theorem is omitted due to the limited space. It is worth noting that the assumption (a) in Theorem 1 is to ensure that problem (2) has a solution and the assumption (b) is to ensure that the closure of the set $\mathbf{B}(\hat{\mathbf{W}}, \rho) \cap \mathcal{W}$ contains an optimal solution to problem (2). The theorem implies that only given large enough penalty parameters, the new criterion (4) holds the same global and local properties as that without the penalty term. In practical situations, however, this means that the choice of the initial values of the penalty parameters has an important effect on the overall optimization accuracy and efficiency. Too small values will violate major constraints, and too large values may create an ill-conditioned computation problem [9]. This fact can also be observed from the eigenvalue structure of its Hessian matrix demonstrated in Section 3.2.

3.2 Numerical Equivalence and Stability

Assuming that $\mathfrak{J}(\mathbf{W})$ is twice-differentiable and calculating the perturbation matrix $\mathbf{\Delta}$ of $\mathbf{W}$, we have the following Hessian matrix

$$\nabla^2 \mathfrak{J}(\mathbf{W}) \triangleq \nabla^2 \mathcal{F}(\mathbf{W}) + \kappa \frac{\partial \mathcal{U}(\mathbf{W})}{\partial \mathbf{W}^*} \nabla^2 g_i(\mathbf{W}) + \kappa \frac{\partial^2 \mathcal{U}(\mathbf{W})}{\partial \mathbf{W}^*} \nabla g_i(\mathbf{W}) \nabla g_i(\mathbf{W})^T \quad (5)$$

The conditions of *Theorem 1* indicate that as $\kappa \rightarrow \infty$, $\mathbf{W}$ 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 $\frac{\partial \mathcal{U}(\hat{\mathbf{W}})}{\partial \mathbf{W}^*} + \sum \bar{\lambda}_i \nabla g_i(\mathbf{W}) = \mathbf{0}$ [7]. This means $\kappa \frac{\partial \mathcal{U}(\mathbf{W})}{\partial \mathbf{W}^*} \rightarrow \bar{\lambda}_i$ as $\mathbf{W} \rightarrow \hat{\mathbf{W}}$. The first two terms in (5) approach the Hessian of $\mathcal{F}(\mathbf{W}) + \sum \bar{\lambda}_i g_i(\mathbf{W})$. Considering the last term in (5), it can be shown that as $\kappa \rightarrow \infty$, $\nabla^2 \mathfrak{J}(\mathbf{W})$ has some eigenvalues approaching ∞, and others 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 small to prevent the ill-conditioned computation problem induced by large eigenvalues with a trade-off of having a lower convergence rate. Such a theoretical analysis is verified in section 4.

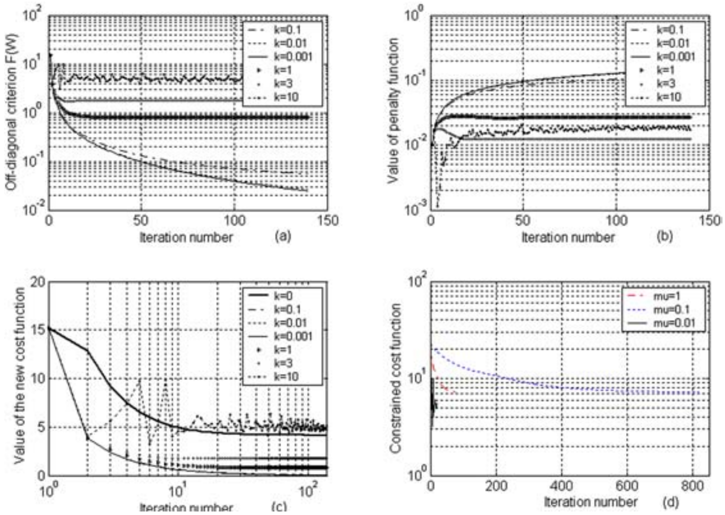

Fig. 2. Convergence behavior of the penalty function approach.

3.3 Approach Summary

Based on the discussions given in the above sections, the proposed algorithm by incorporating penalty functions is conducted as following steps (using the steepest descent gradient adaptation):

1). Initialize parameters N, M, D, T, K, $\mathbf{W}_0$, α, ξ, ς, IRN, $\mathbf{W}_0(\omega)$;

2). Convert the input mixtures $\mathbf{x}(n)$ to $\mathbf{X}(\omega, n)$; calculate the cross-power spectrum matrix $\hat{\mathbf{R}}_X(\omega, k) = \frac{1}{D}\sum_{m=0}^{D-1} \mathbf{X}(\omega, Dk+m)\mathbf{X}^H(\omega, Dk+m)$;

3). Calculate the cost function and update gradient:

- for $i = 1$ to IRN
  * Update $\mu_{J_M}(\omega) = \alpha/(\sum_{k=1}^{K} ||\mathbf{R}_X(\omega, k)||_F^2)$, and $\mu_{J_C}(\omega) = \xi/(\varsigma + \sum_{k=1}^{K} \left\| \frac{\partial J_C(\mathbf{W})(\omega,k)}{\partial \mathbf{W}^*(\omega)} \right\|_F)$ respectively;
  * Update $\mathbf{W}(\omega) \leftarrow \mathbf{W}(\omega) + \mu(\mu_{J_M} \frac{\partial \mathcal{J}}{\partial \mathbf{W}^*(\omega)} + \mu_{J_C} \frac{\partial \mathcal{U}}{\partial \mathbf{W}^*(\omega)})$;
  * Update $\mathfrak{J}_i(\mathbf{W}(\omega))$ using (4);
  * if $(\mathfrak{J}_i(\mathbf{W}(\omega)) > \mathfrak{J}_{i-1}(\mathbf{W}(\omega)))$ break;
- end

4). Solve permutation problem $\mathbf{W}_{new}(\omega) \leftarrow \mathcal{P}(\mathbf{W}(\omega))$, where $\mathcal{P}$ is a function dealing with permutation operation (refer to [4]);

5). Calculate $\mathbf{Y}(\omega, k) = \mathbf{W}(\omega)\mathbf{X}(\omega, k)$ and reconstruct the time domain signals $\mathbf{y}(n) = IDFT(\mathbf{Y}(\omega, k))$;

6). Calculate the performance index, e.g., signal to interference ratio (SIR) [4].

7). End.

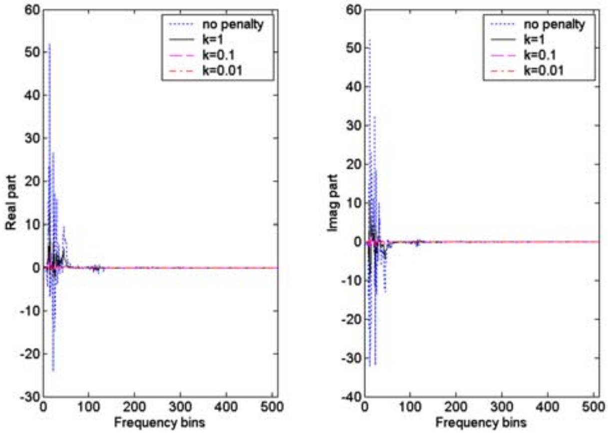

Fig. 3. Comparison of the off-diagonal elements of the cross-correlation matrices $\mathbf{R}_Y(\omega, k)$ at each frequency bin between the proposed method and that in [4] ($\kappa = 0$).

4 Numerical Examples

To examine the proposed method, we use an exterior penalty function with the form of $\|diag[\mathbf{W}(\omega) - \mathbf{I}]\|_F^2$ [6], and a variant of gradient adaptation $\kappa diag[\mathbf{W}(\omega) - \mathbf{I}]\mathbf{W}(\omega)$. A system with two inputs and two outputs (TITO) is considered for simplicity, that is, $N = M = 2$. Two real speech signals are used in the following experiments, which are available from [19]. In the first experiment, 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 $H_{22}(z) = 0.8-0.1z^{-1}$ [18]. Other parameters are set to be $T = 1024$, $K = 5$, $D = 7$, $\alpha = 1$, $\varsigma = 0.05$, $\xi = 0.2$, $\mathbf{W}_0(\omega) = \mathbf{I}$, and $\mu = 1$. We applied the short term FFT to the separation matrix and the cross-correlation of the input data. Fig. 2 show the convergence behavior by incorporating penalty functions. Fig 2 (a)-(c) indicate that, when increasing the penalty coefficient κ, not only the constraint is approached more quickly, but also the cost function converges faster. However, it is also observed that a large penalty κ (e.g. $\kappa = 10$) introduces the ill-conditional problem under a common step size. Such effect can be properly removed by reducing the step size, see Fig. 2 (d), where κ is fixed to be 10, but μ is changing (The adaptation stops when a threshold is satisfied). Theoretically, due to the independence assumption, the cross-correlation of the output signals should approach zero. Fig. 3 demonstrates that it is true at most frequency bins, however with exception for the low frequency bins. From Fig. 3, we see that this effect can be significantly reduced by using penalty functions.

In the second experiment, the proposed joint diagonalization method is compared with other two joint diagonalization criteria [4] [8]. The mixtures are obtained from the simulated room environment, which was implemented by a *roommix* function available from [20]. The room is assumed to be a $10m \times 10m \times 10m$ cube with wall reflections computed up to the fifth order. The position matrices

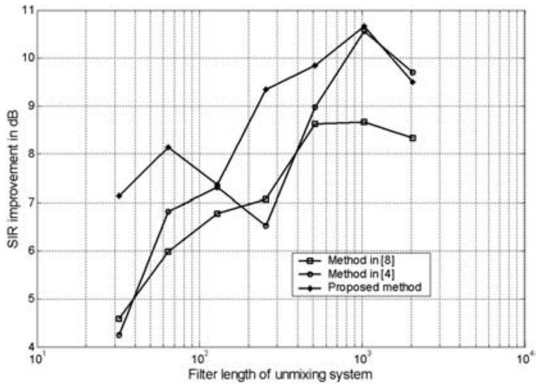

Fig. 4. SIR measurement for a simulated room environment with high reverberance.

of two sources and two sensors are respectively [2 2 5;8 2 5] and [3 8 5;7 8 5]. The SIR plot in Fig. 4 shows that: (i) incorporating a suitable penalty function can increase the SIR which indicates a better separation performance; (ii) the separation quality increases with the increasing filter length of the separation system; (iii) exploiting spectral continuity of the separation matrix (the proposed method and that in [4]) may have superior performance to the method (e.g., [8]) which considers the separation at each frequency bin independently.

5 Conclusion

The penalty function based joint diagonalization approach for frequency domain BSS has been presented. Its convergence behavior and numerical stability have also been discussed. Experimental evaluation indicates that the proposed approach improves the convergence performance as compared with cross-power spectrum based method, and significantly reduces the degenerate effect existing on the lower frequency bins therefore improves its separation performance. This approach also provides a unifying view to constrained BSS which is useful to develop suitable BSS algorithms using optimization techniques.

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. 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*, Nara, Japan, Apr. 1-4, 2003.
6. 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.
7. M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear Programming Theory and Algorithms*, 2nd ed. John Wiley & Sons Inc., 1993.
8. N. Murata, S. Ikeda, and A. Ziehe. "An approach to blind source separation based on temporal structure of speech signals." *Neurocomputing*, vol. 41, pp. 1-24, 2001.
9. A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, Wiley, 1993.
10. A. Cichocki and P. Georgiev, "Blind source separation algorithms with matrix constraints", *IEICE Trans. on Fundamentals of Elect. Comm. and Computer Science*, vol. E86-A, pp. 522-531, Mar. 2003.
11. J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017-3030, Dec. 1996.
12. S. C. Douglas, S. Amari, and S.-Y. Kung, "On gradient adaptation with unit norm constraints," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1843-1847, June 2000.
13. S. C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. on Neural Networks*, vol. 11 no. 6, pp. 1490-1497, June 2000.
14. J. H. Manton, "Optimisation algorithms exploiting unitary constraints," *IEEE Trans. Signal Processing*, vol. 50, pp. 635–650, Mar. 2002.
15. S. Amari, T. P. Chen and A. Cichocki, "Nonholonomic orthogonal learning algorithms for blind source separation," *Neural Computation*, vol. 12, pp. 1463-1484, 2000.
16. M. D. Plumbley, "Algorithms for non-negative independent component analysis," *IEEE Transactions on Neural Networks*, vol. 14 no. 3, pp. 534- 543, May 2003.
17. L. Parra and C. Alvino, "Geometric Source Separation: Merging convolutive source separation with geometric beamforming", *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 6, pp. 352-362, Sept. 2002.
18. T. W. Lee, A. J. Bell, and R. Lambert, "Blind separation of delayed and convolved sources", *Advances in neural information processing systems 9*, MIT Press, Cambridge MA, pp. 758—764, 1997.
19. J. Anemüller, http://medi.uni-oldenburg.de/members/ane.
20. Westner, http://www.media.mit.edu/˜westner.