BLIND SOURCE SEPARATION OF CONVOLVED SOURCES BY JOINT APPROXIMATE DIAGONALIZATION OF CROSS-SPECTRAL DENSITY MATRICES

Kamran Rahbar and James P. Reilly

Electrical & Computer Eng.
Mcmaster University, Hamilton, Ontario, Canada
Email: reilly@mcmaster.ca, kamran@reverb.crl.mcmaster.ca

ABSTRACT

In this paper we present a new method for separating non-stationary sources from their convolutive mixtures based on approximate joint diagonalization of the observed signals' cross-spectral density matrices. Several blind source separation (BSS) algorithms have been proposed which use approximate joint diagonalization of a set of scalar matrices to estimate the instantaneous mixing matrix. We extend the concept of approximate joint diagonalization to estimate MIMO FIR channels. Based on this estimate we then design a separating network which will recover the original sources up to only a permutation and scaling ambiguity for minimum phase channels. We eliminate the commonly experienced problem of arbitrary scaling and permutation at each frequency bin, by optimizing the cost function directly with respect to the time-domain channel variables. We demonstrate the performance of the algorithm by computer simulations using real speech data. Speech samples are available at: http://sparky.mcmaster.ca/SSP/telephony_kamran.htm.

1. INTRODUCTION

In a blind source separation (BSS) problem the objective is to separate independent sources that are mixed through an unknown mixing environment (channel) where no information is available about the sources nor the environment. An example is the room environment when different speakers are talking simultaneously. So far many algorithms have been proposed to solve the BSS problem. Most of them are directed toward the simpler case of *instantaneous* mixing, i.e., the case when the observed signals are a linear combination of sources and no time delays are involved [1],[2], [3],[4]. A more challenging problem which is closer to the practical situation is when the mixing environment is *convolutive*, i.e. the observed signals are a linear mixture of the sources and their delayed versions [5], [6], [7] [8]. Most of methods for convolutive BSS use a backward model; i.e., they try to estimate a separating network, rather than directly estimate the channel, based on some measure of the observed signal, (e.g., [6], [8]) or the output signals (e.g. [5]). In this work we propose a new method to estimate the MIMO FIR channel $\mathbf{H}(z)$ by direct use of the observed signals. The sources then can be separated (or even recovered) based on this estimated channel. In [9], [2], [10] the authors use joint approximate diagonalization of scalar data matrices to estimate the *instantaneous* mixing matrix. In this paper we extend the concept of joint diagonalization to rational matrix functions, for the *convolutive* mixing case. The MIMO FIR channel is estimated by joint approximate diagonalization of the cross-spectral density of the observed signals at different time lags. Our method is simi-

lar to [6] and [11] in this respect, in that all of them use second–order time–varying cross–spectral information, assuming that the sources are non-stationary. The primary differences between the proposed method and the previous methods are 1) we estimate the channel directly, and 2) we eliminate the common difficult problem of arbitrary permutation and scaling at each frequency bin by directly optimizing a new cost function (formulated in frequency domain) with respect to the channel coefficients. Also the previous methods separate the sources only up to a *filtered* version of original sources. With the proposed method, under the assumption that the order of channel is known, it is possible to recover the sources up to a *scaled* version of the original sources. Notice that in the case of separation, the separation system output is a permuted and *filtered* version of the sources, while in the case of recovery, the outputs are only a *scaled* and permuted version of the sources. Numerical simulation using real speech data have been presented to demonstrate the performance of the algorithm.

2. PROBLEM FORMULATION

Assume that we have a source signal vector $\mathbf{s}(n)$ consisting of N sources $\mathbf{s}(n) = (s_1(n), s_2(n), \cdots, s_N(n))^T$ where the source signals are real, zero mean, non-stationary, where the cross- spectral matrix of sources $\mathbf{D}(\omega, m)$ for any given frequency ω and time epoch m is diagonal. Now let's assume that $\mathbf{s}(n)$ is the input to a causal system with transfer function $\mathbf{H}(z^{-1})$, which is a square polynomial matrix of order L, the general form for which is given by:

$$\mathbf{H}(z^{-1}) = \mathbf{H}_0 + \mathbf{H}_1 z^{-1} + \cdots + \mathbf{H}_L z^{-L} \qquad (1)$$

where $\mathbf{H}_i$ $(i = 0, \cdots, L)$ are real square matrices. The output of the systems is given by

$$\mathbf{x}(n) = \sum_{i=0}^{L} \mathbf{H}_i \mathbf{s}(n - i) \qquad (2)$$

The objective of the BSS problem is, given only the observed signals, separate (or recover) the original sources. This is accomplished by means of a separating system whose transfer function we denote by $\mathbf{B}(z^{-1})$. If we denote the estimated channel by $\hat{\mathbf{H}}(z^{-1})$, then to recover the sources we can set $\mathbf{B}(z^{-1}) = \hat{\mathbf{H}}^{-1}(z^{-1})$ where $\hat{\mathbf{H}}^{-1}(z^{-1})$ is the inverse of $\hat{\mathbf{H}}(z^{-1})$ and is given by [12]:

$$\hat{\mathbf{H}}^{-1}(z^{-1}) = \frac{\operatorname{adj}\hat{\mathbf{H}}(z^{-1})}{\det\hat{\mathbf{H}}(z^{-1})} = \frac{\operatorname{adj}\hat{\mathbf{H}}(z^{-1})}{\mathcal{F}(z^{-1})} \qquad (3)$$

where $\mathrm{adj}(\cdot)$ denotes the *adjoint* matrix. Notice that for $\hat{\mathbf{H}}(z^{-1})$ to be causal and stable $\mathcal{F}(z^{-1})$ should be minimum phase. For cases when $\mathcal{F}(z^{-1})$ is not minimum phase it can always be factorized as $\mathcal{F}(z^{-1}) = \mathcal{F}_{min}(z^{-1})\mathcal{F}_a(z^{-1})$ where $\mathcal{F}_{min}(z^{-1})$ is a minimum phase filter and $\mathcal{F}_a(z^{-1})$ is an all-pass filter. If we use $\mathcal{F}_{min}(z^{-1})$ instead of $\mathcal{F}(z^{-1})$ in (3) then the separating system is always stable but the sources can only be recovered up to a scaling and phase ambiguity of the original sources. Note that for minimum phase channels, $\mathcal{F}(z^{-1}) = \mathcal{F}_{min}(z^{-1})$ and complete source recovery is possible. If we are only interested in separation of sources then we can always set $\mathbf{B}(z^{-1}) = \mathrm{adj}\hat{\mathbf{H}}(z^{-1})$. Notice that in this case $\mathbf{B}(z^{-1})$ is a polynomial matrix with order of $(N-1)(L-1)+1$ where L is the order of $\hat{\mathbf{H}}(z^{-1})$.

3. POLYNOMIAL APPROXIMATE JOINT DIAGONALIZATION (PAJD)

Blind source separation and joint diagonalization are closely related, as discussed in [13],[3],[10]. The idea behind joint diagonalization is the existence of a set of known matrices $\mathbf{P}_1, \cdots, \mathbf{P}_M \in \mathbb{C}^{\mathbb{N} \times \mathbb{N}}$ such that:

$$\mathbf{P}_m = \mathbf{A}\boldsymbol{\Lambda}_m\mathbf{A}^H \qquad m = 1, 2, \cdots, M \tag{4}$$

where $\mathbf{A}$ is an unknown matrix and $\boldsymbol{\Lambda}_m$ are some unknown diagonal matrices. Notice that in the case where an exact estimate of $\mathbf{P}_m$ is not available, the above problem becomes an approximate joint diagonalization of a set of estimates: $\hat{\mathbf{P}}_1, \cdots, \hat{\mathbf{P}}_M$. In [9] the authors propose a solution assuming an orthogonal structure for $\mathbf{A}$, while [10] and [14] present non-orthogonal joint diagonalization methods. In what follows, we extend the concept of joint diagonalization using finite order polynomial matrices. To this end, consider rational, square matrix functions $\mathbf{P}_1(z^{-1}), \cdots, \mathbf{P}_M(z^{-1})$ where $\mathbf{P}_m(\omega) \geq 0$ for $z = e^{j\omega}$ and for all $\omega \in [0, \pi]$ and $m = 1, 2, \cdots, M$. We have:

$$\mathbf{P}_m(z^{-1}) = \mathbf{H}(z^{-1})\mathbf{D}_m(z^{-1})\mathbf{H}(z^{-1})^H \tag{5}$$

for $\omega \in [0, \pi]$ and $m = 0, \cdots, M-1$. $\mathbf{H}(z^{-1})$ is a unknown polynomial matrix with a known order L and $\mathbf{D}(z^{-1})$ is a unknown diagonal matrix with diagonal elements which are rational functions of z^{-1}. In situations where only an estimate $\hat{\mathbf{P}}_m(z^{-1})$ is available, the problem becomes one of finding estimates $\hat{\mathbf{H}}(z^{-1})$ and $\hat{\mathbf{D}}_1(z^{-1}), \cdots, \hat{\mathbf{D}}_M(z^{-1})$ such that (5) approximately holds.

Note that we can directly apply the above formulation to the convolutive BSS problem. $\mathbf{P}_m(z^{-1})$ can be considered the cross spectral density of observed signal at epoch m. Then $\mathbf{H}(z^{-1})$ is the unknown transfer function of channel and $\mathbf{D}_m(z^{-1})$ is the cross-spectral density of the sources at epoch m. Notice the assumption of uncorrelated sources implies that $\mathbf{D}_m(z^{-1})$ is diagonal. In next section we propose a criterion to estimate $\mathbf{H}(z^{-1})$ using the formulation given in (5).

4. ALGORITHM

To solve (5) we propose following criterion:

$$\Gamma = \int_0^{\pi} \sum_{m=0}^{M-1} \| \mathbf{F}(\omega, m) \|_F^2 \, d\omega \tag{6}$$

where

$$\mathbf{F}(\omega, m) = \hat{\mathbf{P}}_m(\omega) - \sum_{\alpha=0, \beta=0}^{L} \hat{\mathbf{H}}_\alpha \hat{\mathbf{D}}_m(\omega)\hat{\mathbf{H}}_\beta^T e^{(-j[\alpha-\beta]\omega)} \tag{7}$$

By discretizing the frequency range $[0, \pi]$ to K equidistant sample points $\omega_k = \dfrac{\pi k}{K}$ the integral in (6) can be replaced by a summation and we have:

$$\Gamma = \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \| \mathbf{F}(\omega_k, m) \|_F^2 \tag{8}$$

We minimize the criterion using a conjugate gradient method by optimizing with respect to $\hat{\mathbf{H}}_i$ and $\hat{\mathbf{D}}_m(\omega_k)$. To do so, we need first to calculate the gradient of Γ with respect to $\hat{\mathbf{H}}_i$ and $\hat{\mathbf{D}}_m(\omega)$. Notice that $\mathbf{F}(\omega_k, m)$ is a complex matrix function and in general can be written as:

$$\mathbf{F}(\omega_k, m) = \mathbf{F}_R(\omega_k, m) + j\mathbf{F}_I(\omega_k, m) \tag{9}$$

where $\mathbf{F}_R(\omega_k, m)$ and $\mathbf{F}_I(\omega_k, m)$ are respectively the real part and imaginary part of the $\mathbf{F}(\omega_k, m)$. Substituting (9) into (8) and using the identity $\|\mathbf{A}\|_F^2 = Tr(\mathbf{A}\mathbf{A}^H)$ for any matrix $\mathbf{A}$ we can write:

$$\begin{aligned}
\Gamma &= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} Tr(\mathbf{F}(\omega_k, m)\mathbf{F}^H(\omega_k, m)) \\
&= \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} Tr(\mathbf{F}_R(\omega_k, m)\mathbf{F}_R^T(\omega_k, m) + \\
&\quad \mathbf{F}_I(\omega_k, m)\mathbf{F}_I^T(\omega_k, m))
\end{aligned} \tag{10}$$

Taking the derivative of (10) with respect to $\hat{\mathbf{H}}_i$ we can write:

$$\begin{aligned}
\frac{\partial \Gamma}{\partial \hat{\mathbf{H}}_i} &= -4 \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \mathbf{F}_R(\omega_k, m)\mathbf{G}_R^i(\omega_k, m) \\
&\quad + \mathbf{F}_I(\omega_k, m)\mathbf{G}_I^i(\omega_k, m)
\end{aligned} \tag{11}$$

where $\mathbf{G}_R^i(\omega_k, m)$ and $\mathbf{G}_I^i(\omega_k, m)$ are respectively the real part and imaginary parts of

$$\mathbf{G}^i(\omega_k, m) = \sum_{\alpha=0}^{L} \hat{\mathbf{H}}_\alpha \hat{\mathbf{D}}_m(\omega_k) \exp(-j[i-\alpha]\omega_k) \tag{12}$$

To find the derivative of Γ with respect to $\mathbf{D}_m(\omega_k)$ we first rewrite the criterion (8) as:

$$\Gamma = \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \|\hat{\mathbf{P}}_m(\omega_k) - \hat{\mathbf{H}}(\omega_k)\hat{\mathbf{D}}_m(\omega_k)\hat{\mathbf{H}}^H(\omega_k)\|_F^2 \tag{13}$$

Suppose $\mathbf{d}_m(\omega_k)$ is a vector containing the diagonal values of $\hat{\mathbf{D}}_m(\omega_k)$. Then we can write (13) as:

$$\Gamma = \sum_{m=0}^{M-1} \sum_{k=0}^{K-1} \| \mathbf{p}_m(\omega_k) - \mathbf{B}(\omega_k)\mathbf{d}_m(\omega_k) \|^2 \tag{14}$$

where $\mathbf{p}_m(\omega_k) = vec(\hat{\mathbf{P}}_m(\omega_k))$ and

$$\mathbf{B}(\omega_k) = (\mathbf{H}^*(\omega_k) \otimes \mathbf{1}) \odot (\mathbf{1} \otimes \mathbf{H}(\omega_k)). \tag{15}$$

Here * represents the conjugate operation and $\mathbf{1}$ is $1 \times M$ vector of ones. The derivative of (14) with respect to $\mathbf{d}_m(\omega_k)$ can be easily found to be:

$$\frac{\partial \Gamma}{\partial \mathbf{d}_m(\omega_k)} = -2[\mathbf{p}_m(\omega_k) - \mathbf{B}(\omega_k)\mathbf{d}_m(\omega_k)]^H \mathbf{B}(\omega_k) \quad (16)$$

4.1. Conjugate Gradient Algorithm

Using the gradient information obtained in last section we use the conjugate gradient algorithm to minimize (8) with respect to $\hat{\mathbf{H}}_i$ and $\mathbf{d}_m(\omega_k)$. The Conjugate Gradient method has the advantage that, while it uses only gradient information, it has a faster convergence rate than gradient descent [15], [4]. The search direction for the conjugate gradient method at each step is calculated using a linear combination of the gradient of the cost function at the current step and the search direction at the previous step. For our problem, the adaptation rule for $\hat{\mathbf{H}}_i$ using conjugate gradient is given by:

$$\hat{\mathbf{H}}_i^{k+1} = \hat{\mathbf{H}}_i^k + \alpha^k \Upsilon_i^k \quad (17)$$

where α^k is the step size and

$$\Upsilon^k = -\boldsymbol{\Phi}_i^k + \beta_i^k \Upsilon_i^{k-1}. \quad (18)$$

Here $\boldsymbol{\Phi}_i^k = \dfrac{\partial \Gamma}{\partial \mathbf{H}_i^k}$ and

$$\beta_k^i = \frac{Tr[\boldsymbol{\Phi}_i^k \boldsymbol{\Phi}_i^{k^T}]}{Tr[\boldsymbol{\Phi}_i^{k-1} \boldsymbol{\Phi}_i^{k-1^T}]} \quad (19)$$

Likewise the adaptation rule for $\mathbf{d}_m(\omega_l)$ can be written as:

$$\mathbf{d}_m^{k+1}(\omega_l) = \mathbf{d}_m^k(\omega_l) + \alpha^k \boldsymbol{\theta}_m^k(\omega_l) \quad (20)$$

where

$$\boldsymbol{\theta}_m^k(\omega_l) = -\boldsymbol{\psi}_m^k(\omega_l) + \gamma_m^k \boldsymbol{\theta}_m^{k-1}(\omega_l) \quad (21)$$

Here $\boldsymbol{\psi}_m^k(\omega_l) = \dfrac{\partial \Gamma}{\partial \mathbf{d}_m^k(\omega_l)}$ and

$$\gamma_m^k = \frac{\boldsymbol{\theta}_m^{k^T} \boldsymbol{\theta}_m^k}{\boldsymbol{\theta}_m^{k-1^T} \boldsymbol{\theta}_m^{k-1}} \quad (22)$$

where in all the equations $i = 0, \cdots, L, \quad m = 0, \cdots, M-1$, $l = 0, \cdots, K-1$ and k is the iteration number.

5. SIMULATION RESULTS

To demonstrate the performance of the algorithm we perform numerical simulations using real speech data. For sources, we use two segments of speech, each 2.2 seconds in duration sampled at 8.0 kHz. The speech is from two different females reading different sentences. We specified the channel $\mathbf{H}(z^{-1})$ to be of order of $L = 7$ with elements given as:

$$H_{11}(z^{-1}) = 1 + 0.8z^{-1} + 0.7z^{-2} + 0.4z^{-3} + 0.3z^{-4} + 0.25z^{-5} + 0.2z^{-6} + 0.15z^{-7}$$

$$H_{12}(z^{-1}) = 0.6 + 0.5z^{-1} + 0.5z^{-2} + 0.4z^{-3} + 0.3z^{-4} + 0.2z^{-5} + 0.25z^{-6} + 0.1z^{-7}$$

$$H_{21}(z^{-1}) = 0.5 + 0.5z^{-1} + 0.4z^{-2} + 0.35z^{-3} + 0.3z^{-4} + 0.3z^{-5} + 0.2z^{-6} + 0.1z^{-7}$$

$$H_{22}(z^{-1}) = 1 + 0.9z^{-1} + 0.8z^{-2} + 0.6z^{-3} + 0.4z^{-4} + 0.35z^{-5} + 0.3z^{-6} + 0.15z^{-7}$$

We then generated the observed signals by convolving the channel with the sources. To estimate the cross spectral density of observed signal, we first divide a block of the observed signal of size L_x into M frames of size L_m. Notice that the frame length should be chosen such that the signals within the frame stays nearly stationary. We then calculate the cross spectral density at each frame using standard spectral estimation methods. For this simulation we used the well–known Welsh method. In each frame we divide the sequence into J overlapping subsequences of length L_J and then calculate the FFT of each subsequence after applying a Hamming window. Note that L_J should be much greater than the channel length L. The resulting cross-spectral is given by:

$$\hat{\mathbf{P}}(\omega, m) = \frac{1}{J} \sum_{i=0}^{J-1} \mathbf{x}_{im}(\omega)\mathbf{x}_{im}^H(\omega) \quad m = 0, \cdots, M-1 \quad (23)$$

where $\mathbf{x}_{im}(\omega)$ is the FFT of the ith windowed subsequence in the mth frame. These estimated cross-spectral matrices were then used as inputs to our algorithm. The algorithm converged after 700 to 800 iterations.

The sources were then separated using the estimated $\hat{\mathbf{H}}(z^{-1})$. As indicated previously, this can be done by using the adjoint of $\hat{\mathbf{H}}(z^{-1})$ and convolving it with observed signals. The results are shown in figure (1). The first row of figures represents the original sources $s_1(n)$ and $s_2(n)$, the second row shows the observed signals $x_1(n)$ and $x_2(n)$, and the third row represents the outputs which are the result of the convolution of the observed signals with the adjoint of the estimated channel. As can be seen, the sources have been separated. However, we notice that (by observing the envelope of the separated signals) the outputs are not recovered, since they are a filtered version of the original sources. To recover the sources, we equalize the outputs using a truncated version of the minimum phase inverse of $\det\hat{\mathbf{H}}(z^{-1})$. The resulting recovered sources are shown in the last row of figure(1). By comparing the envelopes it can be seen that the signals in the last row are less distorted than the original sources. These qualitative results have been verified by listening tests.

To find a quantitative measure of the separation performance, we measured the signal to interference ratio (SIR) through the combined mixing and separation systems appearing at each separated output, in dB. The results are shown in table (1). Notice that the equalization process results in some degradation of the separation performance. Also to quantify equalization performance, we used following formula:

$$\kappa_{ij} = 10log_{10}\left(\frac{\|\mathbf{c}_{ij}\|_\infty^2}{\|\mathbf{c}_{ij}\|_2^2 - \|\mathbf{c}_{ij}\|_\infty^2}\right) \quad (24)$$

where $\mathbf{c}_{ij}$ is the impulse response of combined system from source i to output j, measured by applying a delta-function to one input of the combined system while the other inputs were set to zero. The results are tabulated in table 2. The computational time for this simulation was approximately 0.1s for each iteration using a 1.3G Pentium-4 machine in the Matlab environment to run the algorithm. The computational bottleneck of the algorithm is mainly due to calculating the gradient of $\hat{\mathbf{H}}_i$ given by equations (11) and (12). This can be greatly improved by using the FFT method to calculate $\mathbf{G}_i$ in (12).

Output SIR	SIR1	SIR2
Before Equalizing	24.12 dB	19.0324 dB
After Equalizing	12.75 dB	9.43 dB

Table 1: SIR for separated only and separated and equalized outputs

κ_{ij}	κ_{11}	κ_{22}
Before Equalizing	-5.4690 dB	-6.55 dB
After Equalizing	5.543 dB	6.8329 dB

Table 2: Equalization performance for separated only and separated and equalized outputs

6. CONCLUSION

In this paper we have presented a new method for blind source separation of convolutive mixtures using joint diagonalization of the observed signals' cross-spectral density matrices. We assume that the sources are non-stationary and the channel is an FIR matrix with known order. The proposed method uses the adjoint of the estimated channel, as the demixing system, to separate the sources. In this case the separated outputs are a filtered version of the original sources. We then showed that we can recover the sources by equalizing the separated outputs, using a stable inverse of the determinant of the channel matrix. We verified the performance of the proposed method using real speech data, and through qualitative and quantitative measurements.
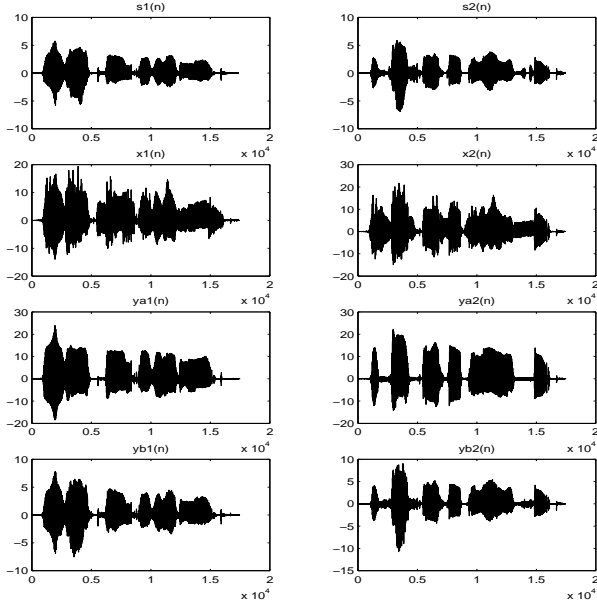

Figure 1: Results of separation: original sources $s1(n)$ and $s2(n)$, observed signals $x1(n)$ and $x2(n)$, separated outputs $ya1(n)$ and $ya2(n)$, separated and equalized outputs $yb1(n)$ and $yb2(n)$

8. REFERENCES

[1] P. Comon, "Independent component analysis, A new concept?," *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] D. Pham and J. Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources," in *Proc. ICA pp.187-190*, June 2000.

[3] A. Belouchrani, K. Meraim, and J. Cardoso, "A blind source separation technique using second order statistics," *Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.

[4] K. Rahbar and J. Reilly, "Geometric optimization methods for blind source separation of signals," in *Proc. ICA pp.375-380*, June 2000.

[5] J. Reilly and L. Mendoza, "Blind source separation for convolutive mixing environments using spatial-temporal processing," in *Proc. ICASSP vol. 3 pp. 1437-1440*, March 1999.

[6] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, vol. 8, pp. 320–327, May. 2000.

[7] D. Yellin and E. Weinstein, "Multichannel signal separation: methods and analysis," *Signal Processing*, vol. 44, pp. 106–118, Jan. 1996.

[8] H. Sahlin and H. Broman, "MIMO signal separation for fir channels: a criterion and performance analysis," *Signal Processing*, vol. 48, pp. 642–649, March. 2000.

[9] J. Cardoso and A. S. Soulomiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, pp. 161–164, no. 1 1996.

[10] A. Yeredor, "Approximate joint diagonalization using non-orthogonal matrices," in *Proc. ICA pp.33-38*, June 2000.

[11] H. Wu and J. Principe, "Simultaneous diagnolization in frequency domain (sdif) for source separation," in *Proc. ICA pp.245-250*, Jan. 1999.

[12] P. Vaidyanathan, *Multirate Systems and Filter Banks*. New Jersey: Prentice Hall, 1993.

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

[14] D. Pham, "Joint approximate diagonalization of positive definite hermitian matrices ," in *Tech. Report*, (Grenoble University), 2000.

[15] D. Bertsekas, *Nonlinear Programming*. Belmont Mass.: Athena Scientific, second ed., 1999.