

# AN EXTENDED ACDC ALGORITHM FOR THE BLIND ESTIMATION OF CONVOLUTIVE MIXING SYSTEMS

*Alfred Mertins*

University of Oldenburg  
Institute of Physics  
26111 Oldenburg, Germany  
alfred.mertins@uni-oldenburg.de

*Iain Russel*

University of Wollongong  
School of Elec., Comp., and Tele. Eng.  
Wollongong, NSW 2522, Australia  
iain@titr.uow.edu.au

## ABSTRACT

This paper presents an extension of the ACDC algorithm introduced by Yeredor for the instantaneous mixing problem to the more general convolutive mixing problem with non-white sources. Further assumptions made on the source signals are their mutual statistical independence, nonstationarity and smoothness of their power spectra. The algorithm iterates the estimation of the mixing system (AC step) and the source statistics (DC step) until convergence is achieved. The proposed algorithm operates in the frequency domain, but unlike most frequency domain algorithms, it carries out some of the operations jointly for all frequencies. This allows us to overcome frequency dependent permutation and scaling problems.

## 1. INTRODUCTION

Blind system identification and blind source separation refer to estimating the system functions of linear multiple-input multiple-output systems and their inverses, respectively, from observations made only at the system output. Apart from some basic assumptions about the source statistics, the input signals are considered to be unknown. Overviews of the topic can, for example, be found in [1, 2].

We consider a linear mixing process where  $N$  source signals  $s_1(n), \dots, s_N(n)$  are mixed in a convolutive manner into  $M \geq N$  observable signals  $x_1(n), \dots, x_M(n)$ . This operation can be expressed in matrix notation as

$$\mathbf{x}(n) = \sum_{m=0}^{P-1} \mathbf{H}(m)\mathbf{s}(n-m) \quad (1)$$

with  $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^T$ ,  $\mathbf{x}(n) = [x_1(n), \dots, x_M(n)]^T$ , and  $[\mathbf{H}(n)]_{i,j} = h_{i,j}(n)$ . The terms  $h_{i,j}(n)$  denote the impulse responses from input  $j$  to output  $i$  of the mixing system. The aim is to identify the impulse responses  $h_{i,j}(n)$  on the basis of the observed signals  $x_1(n), \dots, x_M(n)$  up to the well known scaling and permutation ambiguities that are inherent to all blind identification and separation algorithms [1, 2]. The assumptions made about the input signals are their mutual statistical

independence and quasi-stationarity over short periods of time, but nonstationarity in a more general sense, similar to [3–7], and smoothness of their power spectra. These assumptions are for example well justified for independent speech signals. In particular, the assumption of nonstationarity of the sources allows us to solve the problem based on second-order instead of higher-order statistics [2, 3, 5–7].

Approaches to solve the above mentioned blind identification problem can be divided into time and frequency domain methods. In this work, we use a frequency domain approach to transfer the convolutive time-domain mixing process (1) into instantaneous mixing processes in the frequency domain. Let  $\mathbf{x}(\omega)$ ,  $\mathbf{H}(\omega)$ , and  $\mathbf{s}(\omega)$  denote the Fourier transforms of the sequences  $\mathbf{x}(n)$ ,  $\mathbf{H}(n)$ , and  $\mathbf{s}(n)$ , respectively. Assuming that a total number of  $K$  frequencies  $\omega_k = 2\pi k/K$ ,  $k = 0, 1, \dots, K-1$  are observed, we may replace (1) with  $K$  instantaneous mixing processes of the form

$$\mathbf{x}(\omega_k) = \mathbf{H}(\omega_k)\mathbf{s}(\omega_k). \quad (2)$$

The aim is to find estimates  $\hat{\mathbf{H}}(\omega_k)$  such that the remaining ambiguities can be expressed as

$$\hat{\mathbf{H}}(\omega) = \mathbf{H}(\omega)\mathbf{P}\mathbf{D}(\omega) \quad \forall \omega \quad (3)$$

where  $\mathbf{P}$  is a permutation matrix and  $\mathbf{D}(\omega)$  is a (possibly frequency dependent) diagonal scaling matrix. Thus, we essentially look at a setting that is similar to the one in [6]. Differences between our approach and the one in [6] are that we assume colored instead of white source signals and that we use a different approach to determine the unknown mixing system. Our method can be seen as an extension of the ACDC algorithm that was introduced in [7] for instantaneous mixing. We transfer the ACDC algorithm to the frequency domain and optimize some of the involved parameters for all frequencies simultaneously. This joint optimization allows us to overcome the frequency dependent scaling and permutation ambiguity problems that occur with all frequency-domain approaches.

A few comments on the notation. Vectors and matrices are printed in boldface. The superscript  $\{\cdot\}^H$  means transposition and complex conjugation of a matrix or vector. The superscript  $\{\cdot\}^+$  denotes the pseudoinverse.  $E\{\cdot\}$  means

the expectation operation.  $\|\cdot\|_F$  is the Frobenius norm of a matrix. The symbols  $\otimes$  and  $\odot$  denote the Kronecker and Hadamard products, respectively. The term  $v = \text{diag}[u]$  denotes the formation of a diagonal matrix  $v$  from a set of values  $u$  as well as forming a column vector  $v$  from the diagonal elements of a matrix  $u$ . If the argument is a set of matrices, then the result is a block diagonal matrix.

## 2. OBJECTIVE FUNCTION AND ADAPTATION PROCESS

We assume the signals  $x_j(n)$  to be observed during  $T$  different time epochs and rewrite the model (2) as

$$\mathbf{x}(\omega_k, t) = \mathbf{H}(\omega_k) \mathbf{s}(\omega_k, t), \quad t = 1, 2, \dots, T. \quad (4)$$

Given the observations  $\mathbf{x}(\omega_k, t)$  it is straightforward to find estimates for the cross-power spectral density matrices  $\mathbf{R}_{\omega_k, t} = E\{\mathbf{x}(\omega_k, t)\mathbf{x}^H(\omega_k, t)\}$ , and based on the model (4), these can be described as

$$\mathbf{R}_{\omega_k, t} = \mathbf{H}(\omega_k) \mathbf{\Lambda}_{\omega_k, t} \mathbf{H}^H(\omega_k) \quad (5)$$

where  $\mathbf{\Lambda}_{\omega_k, t}$  are the cross-power spectral density matrices of the nonstationary input processes  $\mathbf{s}(\omega_k, t)$ . Because the different sources are assumed to be statistically independent, the matrices  $\mathbf{\Lambda}_{\omega_k, t}$  are diagonal.

The criterion to find estimates for the mixing systems  $\mathbf{H}(\omega_k)$  and the unknown input power spectra  $\mathbf{\Lambda}_{\omega_k, t}$  is defined as

$$\text{minimize } C = \sum_{k=0}^{K-1} C_{\omega_k} \quad (6)$$

with

$$C_{\omega_k} = \sum_{t=1}^T \left\| \mathbf{R}_{\omega_k, t} - \mathbf{H}(\omega_k) \mathbf{\Lambda}_{\omega_k, t} \mathbf{H}^H(\omega_k) \right\|_F^2. \quad (7)$$

The minimization involves two steps that are repeated until convergence. First, we carry out a so-called AC step, where the criterion is minimized with regard to individual columns of  $\mathbf{H}(\omega_k)$ , while matrices  $\mathbf{\Lambda}_{\omega_k, t}$  remain constant (cf. [7]). This step is carried out repeatedly for all columns of  $\mathbf{H}(\omega_k)$ , until convergence. In the second step, the DC step,  $C$  is minimized with respect to  $\mathbf{\Lambda}_{\omega_k, t}$ . Then, another AC step is carried out, and so on, until the final minimum is reached. In order to minimize the effect of different scale factors and permutations at different frequencies, a projection procedure is included in the AC step that ensures that the identified time-domain impulse responses do not significantly exceed a maximum, pre-determined length.

### AC Step (Part 1)

In this step, we minimize  $C_{\omega_k}$  with respect to the  $\ell$ th column of  $\mathbf{H}(\omega_k)$  for each frequency  $\omega_k$  separately. Using

the equality

$$\mathbf{H}(\omega_k) \mathbf{\Lambda}_{\omega_k, t} \mathbf{H}^H(\omega_k) = \sum_{n=1}^N \lambda_n^{(\omega_k, t)} \mathbf{h}_{\omega_k, n} \mathbf{h}_{\omega_k, n}^H, \quad (8)$$

where  $\mathbf{h}_{\omega_k, n}$  is the  $n$ th column of  $\mathbf{H}(\omega_k)$ ,  $\lambda_n^{(\omega_k, t)} = [\mathbf{\Lambda}_{\omega_k, t}]_{n, n}$  are the diagonal elements of  $\mathbf{\Lambda}_{\omega_k, t}$ , and

$$\tilde{\mathbf{R}}_{\omega_k, t}^{(\ell)} = \mathbf{R}_{\omega_k, t} - \sum_{n=1, n \neq \ell}^N \lambda_n^{(\omega_k, t)} \mathbf{h}_{\omega_k, n} \mathbf{h}_{\omega_k, n}^H, \quad (9)$$

we can write

$$C_{\omega_k} = \sum_{k=1}^K \left\| \tilde{\mathbf{R}}_{\omega_k, t}^{(\ell)} - \lambda_\ell^{(\omega_k, t)} \mathbf{h}_{\omega_k, \ell} \mathbf{h}_{\omega_k, \ell}^H \right\|_F^2. \quad (10)$$

Similar to [7] for instantaneous mixing, this criterion can be rewritten as

$$C_{\omega_k} = \tilde{C}_{\omega_k} - 2\mathbf{h}_{\omega_k, \ell}^H \mathbf{P}_{\omega_k, \ell} \mathbf{h}_{\omega_k, \ell} + p_{\omega_k, \ell} (\mathbf{h}_{\omega_k, \ell}^H \mathbf{h}_{\omega_k, \ell})^2 \quad (11)$$

with

$$\mathbf{P}_{\omega_k, \ell} = \frac{1}{2} \sum_{t=1}^T \lambda_\ell^{(\omega_k, t)} [\tilde{\mathbf{R}}_{\omega_k, t}^{(\ell)H} + \tilde{\mathbf{R}}_{\omega_k, t}^{(\ell)}] \quad (12)$$

and

$$p_{\omega_k, \ell} = \sum_{t=1}^T [\lambda_\ell^{(\omega_k, t)}]^2 \quad (13)$$

The optimal vector  $\mathbf{h}_{\omega_k, \ell}$  is given by

$$\mathbf{h}_{\omega_k, \ell} = b_{\omega_k, \ell} \boldsymbol{\beta}_{\omega_k, \ell} \quad (14)$$

where  $\boldsymbol{\beta}_{\omega_k, \ell}$  is the unit-norm eigenvector of  $\mathbf{P}_{\omega_k, \ell}$  that corresponds to the largest positive eigenvalue  $\mu_{\omega_k, \ell}$  [7]. The optimal prefactors  $b_{\omega_k, \ell}$ , computed separately for each frequency, are given by  $b_{\omega_k, \ell} = (\mu_{\omega_k, \ell} / p_{\omega_k, \ell})^{1/2}$ .

### AC Step (Part 2)

Minimizing the objective criterion for each frequency separately does not allow us to resolve any of the permutation and scale ambiguities. Therefore, at this stage, we employ a projection technique that is similar to the one in [6] in order to jointly compute the prefactors  $b_{\omega_k, \ell}$  that result in time domain responses  $h_{i, j}(n)$  of given, arbitrary length  $P$ . The vectors  $\boldsymbol{\beta}_{\omega_k, \ell}$  remain the ones computed in Part 1 of the AC step. Note that concentrating  $h_{i, j}(n)$  in time will also yield smooth frequency responses  $H_{i, j}(\omega)$ , which are characteristic for most real-world mixing systems.

Let

$$\mathbf{B}_\ell(n) = [e^{j\omega_0 n} \boldsymbol{\beta}_{\omega_0, \ell}, e^{j\omega_1 n} \boldsymbol{\beta}_{\omega_1, \ell}, \dots, e^{j\omega_{K-1} n} \boldsymbol{\beta}_{\omega_{K-1}, \ell}]. \quad (15)$$

Then, the time-domain impulse responses  $\mathbf{h}_\ell(n) = [h_{1, \ell}(n), \dots, h_{M, \ell}(n)]^T$  that correspond to  $\mathbf{h}_{\omega_k, \ell}$  are given by

$$\mathbf{h}_\ell(n) = \frac{1}{K} \mathbf{B}_\ell(n) \boldsymbol{\alpha}_\ell \quad (16)$$

with  $\alpha_\ell = [b_{\omega_0, \ell}, b_{\omega_1, \ell}, \dots, b_{\omega_{K-1}, \ell}]^T$ . The vector  $\alpha_\ell$  that maximizes

$$\sum_{n=0}^{P-1} \mathbf{h}_\ell^H(n) \mathbf{h}_\ell(n) \quad \text{subject to} \quad \sum_{n=0}^{K-1} \mathbf{h}_\ell^H(n) \mathbf{h}_\ell(n) = 1$$

is the one that maximizes  $\alpha_\ell^H \Psi \alpha_\ell$  subject to  $\alpha_\ell^H \Omega \alpha_\ell = 1$  with

$$\Psi = \sum_{n=0}^{P-1} \mathbf{B}_\ell^H(n) \mathbf{B}_\ell(n), \quad \Omega = \sum_{n=0}^{K-1} \mathbf{B}_\ell^H(n) \mathbf{B}_\ell(n).$$

This optimal vector  $\alpha_\ell$  is given by the eigenvector that corresponds to the largest eigenvalue  $\rho$  of the generalized eigenvalue problem  $\Psi \alpha_\ell = \rho \Omega \alpha_\ell$ , normalized such that  $\alpha_\ell^H \Omega \alpha_\ell = 1$ .

### DC Step

In this step, we minimize  $C$  with respect to  $\Lambda_{\omega_k, t}$ . We first rewrite the criterion for each  $t$  and frequency  $\omega_k$  as a squared Euclidean norm of a difference vector (cf. [7]):

$$C_{\omega_k, t} = \|\mathbf{a}_{\omega_k, t} - \mathcal{H}_{\omega_k} \boldsymbol{\lambda}_{\omega_k, t}\|_2^2 \quad (17)$$

with

$$\begin{aligned} \boldsymbol{\lambda}_{\omega_k, t} &= \text{diag}[\Lambda_{\omega_k, t}], \\ \mathbf{a}_{\omega_k, t} &= \text{vec}(\mathbf{R}_{\omega_k, t}), \\ \mathcal{H}_{\omega_k} &= (\mathbf{H}^*(\omega_k) \otimes \mathbf{1}) \odot (\mathbf{1} \otimes \mathbf{H}(\omega_k)) \end{aligned} \quad (18)$$

Computing the optimal vectors  $\boldsymbol{\lambda}_{\omega_k, t}$  for each frequency separately is straightforward:  $\boldsymbol{\lambda}_{\omega_k, t} = \mathcal{H}_{\omega_k}^+ \mathbf{a}_{\omega_k, t}$ . However, such an approach has the drawback that the large degree of freedom may make the tradeoff between the spectral properties of the sources and the mixing system too easy to allow for an accurate estimation of the true underlying random processes on the basis of a finite number of observations. Therefore, algorithms like the one in [6] simplify the source modeling to white sources and absorb all coloration into the mixing system. In practice, however, one often has some *a priori* knowledge about the source signals, which could be exploited during the blind identification process. In the following, we assume that the power density spectra  $\lambda_n^{(\omega_k, t)}$  are smooth functions of frequency and that spectral samples  $\lambda_n^{(\omega_k, t)}$ ,  $k = 1, 2, \dots, K$  can be well approximated in the form

$$\boldsymbol{\lambda}_n^{(t)} = \mathbf{B} \mathbf{v}_n^{(t)} \quad (19)$$

with  $\boldsymbol{\lambda}_n^{(t)} = [\lambda_n^{(\omega_0, t)}, \lambda_n^{(\omega_1, t)}, \dots, \lambda_n^{(\omega_{K-1}, t)}]^T$ , where  $\mathbf{B}$  is a  $T \times Q$  matrix with  $Q < T$  whose columns contain appropriate smooth basis functions.

We now consider the simultaneous optimization of all unknown values  $\lambda_n^{(\omega_k, t)}$  for a given  $t$  using the approximation (19). For this, we first define the cost function

$$C_t = \sum_{k=0}^{K-1} C_{\omega_k, t} = \|\mathbf{a}_t - \mathcal{H} \boldsymbol{\lambda}_t\|_2^2 \quad (20)$$

where

$$\boldsymbol{\lambda}_t = \begin{bmatrix} \lambda_{\omega_0, t} \\ \vdots \\ \lambda_{\omega_{K-1}, t} \end{bmatrix}, \quad \mathbf{a}_t = \begin{bmatrix} \mathbf{a}_{\omega_0, t} \\ \vdots \\ \mathbf{a}_{\omega_{K-1}, t} \end{bmatrix},$$

$$\mathcal{H} = \text{diag}[\mathcal{H}_{\omega_0}, \mathcal{H}_{\omega_1}, \dots, \mathcal{H}_{\omega_{K-1}}].$$

Equation (19) can be rewritten as

$$\boldsymbol{\lambda}_t = \bar{\mathbf{B}} \mathbf{v}_t, \quad \bar{\mathbf{B}} = [\mathbf{B} \otimes \mathbf{I}_{N \times N}], \quad (21)$$

and hence we can write

$$C_t = \|\mathbf{a}_t - \bar{\mathcal{H}} \mathbf{v}_t\|_2^2 \quad (22)$$

with  $\bar{\mathcal{H}} = \mathcal{H} \bar{\mathbf{B}}$ . The vector  $\mathbf{v}_t$  that minimizes (22) is given by

$$\mathbf{v}_t = \bar{\mathcal{H}}^+ \mathbf{a}_t. \quad (23)$$

Given  $\mathbf{v}_t$ , the vector  $\boldsymbol{\lambda}_t$ , containing all values  $\lambda_n^{(\omega_k, t)}$  required to set up the matrices  $\Lambda_{\omega_k, t}$  for the next AC step, are found from (21). Depending on the basis  $\mathbf{B}$  used, it is not assured that all  $\lambda_n^{(\omega_k, t)}$  turn out positive. Therefore, we include one more step where any negative values  $\lambda_n^{(\omega_k, t)}$  are set to zero.

### 3. EXPERIMENTAL RESULTS

We consider the case where two nonstationary, colored source signals are mixed with a two-input two-output mixing system. The autocorrelation sequences of the sources were randomly generated for each time epoch as  $r_{ss}^{(i, t)}(m) = \sigma_{i, t}^2 c_{i, t}(m) * c_{i, t}(-m)$  where  $\sigma_{i, t}^2$  are uniform random variables and  $c_{i, t}(m)$  are length-3 sequences of real-valued Gaussian random variables. The mixing system was chosen as

$$\mathbf{H}(0) = \frac{1}{2} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{H}(1) = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

This is a paraunitary system, and its inverse, delayed to become causal, is given by  $\mathbf{G}(0) = \mathbf{H}(1)$ ,  $\mathbf{G}(1) = \mathbf{H}(0)$ . For the cascade of both systems we have  $\mathbf{C}(n) = \sum_{m=0}^1 \mathbf{G}(m) \mathbf{H}(n-m) = \delta_{n,1} \mathbf{I}$ .

In the tests,  $K = 64$  frequency points were considered. The basis sequences included in the columns of matrix  $\mathbf{B}$  for the DC step were chosen as

$$B_{i, j} = \gamma_j \cos\left(\frac{2\pi i j}{T}\right), \quad \gamma_j = \begin{cases} \sqrt{2/T}, & j \in \{0, \frac{T}{2}\} \\ \sqrt{1/T}, & \text{otherwise} \end{cases} \quad (24)$$

with  $j = 0, 1, \dots, J$  and  $J = 4$ . Note that this exactly describes  $J$ th order moving average source modeling.

Initial values for the estimated mixing system were randomly generated by adding Gaussian random variables with

standard deviation  $\sigma = 0.1$  to the coefficients of the true system. Fig. 1 shows the convergence behavior for 20 different starting points and the same input statistics. As the examples show, in most cases the value of the objective function decreased to extremely low values. Only three times out of 20 the algorithm got trapped in a local minimum with a relatively high value for  $C$ .

In all cases where the final value  $C$  was below  $-200$  dB the power spectra of the sources were perfectly estimated, and also the mixing system was identified up to permutations and scaling. For cases where the final value was around  $-14$  dB the estimates were close to the true values. For one of these cases, four pairs of the identified source power spectra are depicted in Fig. 2 together with the true ones. As can be seen, even for these cases the estimated source power spectra are very close to the true ones. Also the cascades of the estimated mixing systems and the inverse to the true one were near-perfect, as can be seen in the following example:

$$C(0) = \begin{bmatrix} 0.0007 & 0.0002 \\ 0.0007 & 0.0002 \end{bmatrix}$$

$$C(1) = \begin{bmatrix} -1 & 0.0008 \\ 0.0009 & -1 \end{bmatrix}$$

$$C(2) = \begin{bmatrix} 0.0012 & -0.0002 \\ -0.0012 & 0.0002 \end{bmatrix}$$

#### 4. CONCLUSIONS

A new method for the blind estimation of convolutive mixing systems in the presence of colored sources has been presented. The results show that the algorithm estimates both the source spectra and the mixing system with little variance if it converges to a low value of the objective function. Further work will be directed toward automatic initialization of the algorithm and optimization for recorded data such as speech and audio.

#### 5. REFERENCES

- [1] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
- [2] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing*, Wiley, Chichester, 2002.
- [3] L. Parra and C. Spence, "Convolutive blind source separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 10, pp. 320–327, May 2000.
- [4] J. Anemüller and B. Kollmeier, "Amplitude modulation decorrelation for convolutive blind source separation,"

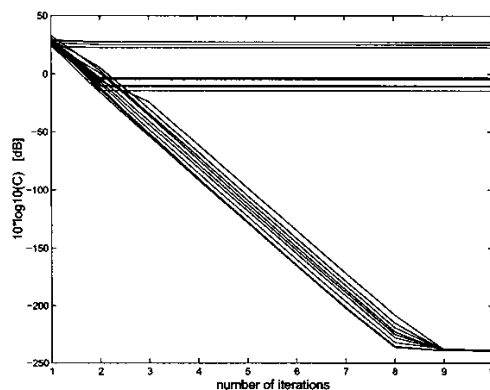


Fig. 1. Value of objective function on a logarithmic scale versus the number of full iterations.

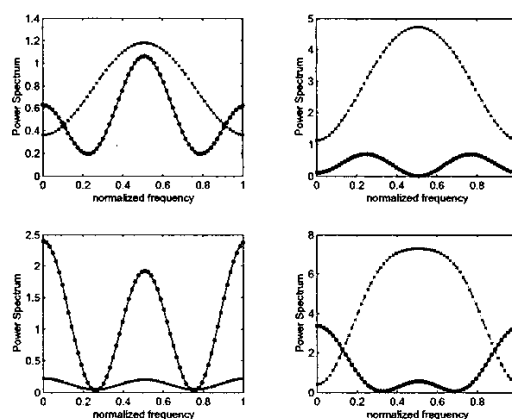


Fig. 2. Source power spectral densities (psd's) and their estimates. Legend: — psd1; - - - psd2; o estimate for psd1; x estimate for psd2.

in *2nd IEEE Int. Workshop Indep. Compon. Anal. Signal Separation*, Helsinki, Finland, June 2000, pp. 215–220.

- [5] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 49, no. 9, pp. 1837–1848, Sept. 2001.
- [6] K. Rahbar, J. P. Reilly, and J. H. Manton, "A frequency domain approach to blind identification of MIMO FIR systems driven by quasi-stationary signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Orlando, FL, USA, May 2002, vol. 2, pp. 1717–1720.
- [7] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, no. 7, pp. 1545–1553, July 2002.