Blind Signal Separation of Convolutive Mixtures: A Time-Domain Joint-Diagonalization Approach

Marcel Joho

Phonak Hearing Systems, Champaign, IL, USA
joho@ieee.org

Abstract. We address the blind source separation (BSS) problem for the convolutive mixing case. Second-order statistical methods are employed assuming the source signals are non-stationary and possibly also non-white. The proposed algorithm is based on a joint-diagonalization approach, where we search for a single polynomial matrix that jointly diagonalizes a set of measured spatiotemporal correlation matrices. In contrast to most other algorithms based on similar concepts, we define the underlying cost function entirely in the time-domain. Furthermore, we present an efficient implementation of the proposed algorithm which is based on fast convolution techniques.

1 Introduction

1.1 Problem formulation

Signal mixing The system setup is described as follows: M_s unknown mutually uncorrelated *source signals* s_m are filtered and mixed by an unknown time-invariant finite-length causal *convolutive mixing system* $\mathbf{A}^{M_\mathrm{x}\times M_\mathrm{s}} \triangleq \{\mathbf{A}_n\}_{n=0}^{N_\mathrm{a}}$ resulting in M_x measurable sensor signals x_m. The source- and sensor-signals are stacked in vectors, $\mathbf{s}$ and $\mathbf{x}$, respectively. For simplicity we neglect any additive noise components. Hence, the convolutive mixing process is described as $\mathbf{x}=\mathbf{A}*\mathbf{s}$ where

$$\mathbf{x}(t) = (\mathbf{A} * \mathbf{s})(t) = \sum\nolimits_{n=0}^{N_\mathrm{a}} \mathbf{A}_n\, \mathbf{s}(t-n) \tag{1}$$

or, written in the z-domain,

$$\mathbf{x}(z) \triangleq \sum\nolimits_{t} \mathbf{x}(t)\, z^{-t} = \mathbf{A}(z)\, \mathbf{s}(z)\,. \tag{2}$$

Signal separation The M_x sensor signals x_m are mixed and filtered with a finite-length non-causal convolutive separation system $\mathbf{W}^{M_\mathrm{u}\times M_\mathrm{x}} \triangleq \{\mathbf{W}_n\}_{n=-N_\mathrm{w}}^{N_\mathrm{w}}$ resulting in M_u output signals u_m. The separation process is described as $\mathbf{u}=\mathbf{W}*\mathbf{x}=\mathbf{W}*\mathbf{A}*\mathbf{s}$, or written in the z-domain

$$\mathbf{u}(z) = \mathbf{W}(z)\, \mathbf{x}(z) = \mathbf{W}(z)\mathbf{A}(z)\, \mathbf{s}(z)\,. \tag{3}$$

The objective of the blind-source-separation problem for the convolutive mixing case is to find a $\mathbf{W}(z)$ such that the global system can be written as

$$\mathbf{G}(z) = \mathbf{W}(z)\,\mathbf{A}(z) = \mathbf{P}\,\mathbf{D}(z) \tag{4}$$

where $\mathbf{D}(z)$ is a diagonal polynomial matrix and $\mathbf{P}$ is a permutation matrix. In the following we assume that $M_\mathrm{u} = M_\mathrm{s} \leq M_\mathrm{x}$, and that the source signals and mixing system can be complex valued. Depending on $\mathbf{A}(z)$, M_x, and M_s perfect separation is possible for a finite N_w.

1.2 Mathematical preliminaries

Basic notation The notation used throughout this paper is the following: Vectors are written in lower case, matrices in upper case. Matrix and vector transpose, complex conjugation and Hermitian transpose are denoted by $(.)^T$, $(.)^*$, and $(.)^H \triangleq ((.)^*)^T$, respectively. The sample index is denoted by t. The identity matrix is denoted by $\mathbf{I}$, a vector or a matrix containing only zeros by $\mathbf{0}$. $E\{.\}$ denotes the expectation operator. The Frobenius norm and the trace of a matrix are denoted by $\|.\|_F$ and $\mathrm{tr}\{.\}$, respectively. $\mathrm{diag}(\,\mathbf{A}\,)$ zeros the off-diagonal elements of $\mathbf{A}$ and

$$\mathrm{off}(\,\mathbf{A}\,) \triangleq \mathbf{A} - \mathrm{diag}(\,\mathbf{A}\,) \tag{5}$$

zeros the diagonal elements of $\mathbf{A}$. The extension for polynomial matrices is defined straightforwardly as $\mathrm{off}(\,\mathbf{A}(z)\,) \triangleq \mathbf{A}(z) - \mathrm{diag}(\,\mathbf{A}(z)\,) = \sum_n \mathrm{off}(\,\mathbf{A}_n\,)z^{-n}$. Linear convolution between two sequences is denoted by $*$. Furthermore, we define

$$\mathbf{A}^\dagger(z) \triangleq \mathbf{A}^H(1/z^*) = \sum\nolimits_n \mathbf{A}_n^H z^{+n}\,. \tag{6}$$

Signals We use the following notation: $x_m(t)$ denotes the value of the signal x_m at discrete time t and $x_m \triangleq \{x_m(t)\}$ denotes the time series of signal x_m. Furthermore, we define $\mathbf{x}(t) \triangleq (x_1(1), \ldots, x_M(t))^T$ and $\mathbf{x} \triangleq (x_1, \ldots, x_M)^T = \{\mathbf{x}(t)\}$. The spatiotemporal correlation matrix between two signal vectors $\mathbf{u}$ and $\mathbf{x}$, and the corresponding z-transform of the correlation sequence are defined as

$$\mathbf{R_{ux}}(\tau;t) \triangleq E\big\{\mathbf{u}(t)\,\mathbf{x}^H(t-\tau)\big\} \tag{7}$$

$$\mathbf{R_{ux}}(z;t) \triangleq \sum\nolimits_{\tau=-\infty}^{\infty} \mathbf{R_{ux}}(\tau;t)z^{-\tau}\,, \tag{8}$$

respectively. For stationary signals we have $\mathbf{R_{ux}}(\tau;t) = \mathbf{R_{ux}}(\tau)$ and, hence, $\mathbf{R_{ux}}(z;t) = \mathbf{R_{ux}}(z)$.

Frobenius norm In the following we, will make use of some concepts from functional analysis [1]: Let $\mathbb{M}$ be the inner product space of complex matrixes. Given two matrices $\mathbf{A}$ and $\mathbf{B}$ with $\mathbf{A}, \mathbf{B} \in \mathbb{M}$, we define the *scalar product* of two matrices as $\langle \mathbf{A}, \mathbf{B} \rangle \triangleq \mathrm{tr}\{\mathbf{A}\mathbf{B}^H\}$. The induced norm is equivalent to the *Frobenius norm*, i.e. $\|\mathbf{A}\|_F \triangleq \sqrt{\langle \mathbf{A}, \mathbf{A} \rangle}$. Norms provide a convenient way to measure a distance between two matrices, as they induce a *metric* defined as $d(\mathbf{A}, \mathbf{B}) \triangleq \|\mathbf{A} - \mathbf{B}\|_F$.

Frobenius norm for polynomial matrices We can extend the definition of the Frobenius norm to polynomial matrices. Let $\mathbb{P}$ be the inner product space of complex polynomial matrixes. Let $\mathbf{A}(z) \triangleq \sum_n \mathbf{A}_n z^{-n}$ and $\mathbf{B}(z) \triangleq \sum_n \mathbf{B}_n z^{-n}$ be two matrix polynomials or Laurent series, i.e., their coefficients are complex matrices. If $\mathbf{A}(z)$ or $\mathbf{B}(z)$ have *finite energy*, i.e., $\sum_n \|\mathbf{A}_n\|_F^2 < \infty$ or $\sum_n \|\mathbf{B}_n\|_F^2 < \infty$, we can define the following *inner product*

$$\langle \mathbf{A}(z), \mathbf{B}(z) \rangle_{\mathcal{F}} \triangleq \sum\nolimits_n \langle \mathbf{A}_n, \mathbf{B}_n \rangle = \sum\nolimits_n \mathrm{tr}\{\mathbf{A}_n \, \mathbf{B}_n^H\} \,. \tag{9}$$

The inner product $\langle ., . \rangle_{\mathcal{F}}$ defines an induced norm on $\mathbb{P}$ given by

$$\|\mathbf{A}(z)\|_{\mathcal{F}} \triangleq \sqrt{\langle \mathbf{A}(z), \mathbf{A}(z) \rangle_{\mathcal{F}}} = \sqrt{\sum\nolimits_n \|\mathbf{A}_n\|_F^2} \tag{10}$$

and a *metric* on $\mathbb{P}$ induced by the norm $d(\mathbf{A}(z), \mathbf{B}(z)) \triangleq \|\mathbf{A}(z) - \mathbf{B}(z)\|_{\mathcal{F}}$. It is not very difficult to show that the definitions of $\langle ., . \rangle_{\mathcal{F}}$ and $\|.\|_{\mathcal{F}}$ fulfill the properties of scalar products and norms [1], respectively. The induced metric $d(\mathbf{A}(z), \mathbf{B}(z)) \triangleq \|\mathbf{A}(z) - \mathbf{B}(z)\|_{\mathcal{F}}$ allows us to measure the "distance" between two polynomial matrices $\mathbf{A}(z)$ and $\mathbf{B}(z)$. In our case, we will use $d(., .)$ to measure the distance between two spatiotemporal correlation matrices.

2 A joint-diagonalization approach

2.1 Correlation matrices

Stationary source signals Assuming that the source signals s_m are stationary, the input spatiotemporal correlation matrix $\mathbf{R}_{\mathbf{xx}}(z)$ of the mixing process (2) is

$$\mathbf{R}_{\mathbf{xx}}(z) = \mathbf{A}(z)\mathbf{R}_{\mathbf{ss}}(z)\mathbf{A}^{\dagger}(z) \,. \tag{11}$$

The output correlation matrix $\mathbf{R}_{\mathbf{uu}}(z)$ of the separation process (3) is

$$\mathbf{R}_{\mathbf{uu}}(z) = \mathbf{W}(z)\mathbf{R}_{\mathbf{xx}}(z)\mathbf{W}^{\dagger}(z) = \mathbf{W}(z)\mathbf{A}(z)\mathbf{R}_{\mathbf{ss}}(z)\mathbf{A}^{\dagger}(z)\mathbf{W}^{\dagger}(z) \,. \tag{12}$$

Extension to block-wise stationary source signals If we relax the stationary assumption and assume that the source signals are non-stationary, but block-wise stationary, then Eq. (11) changes to

$$\mathbf{R}_{\mathbf{xx}}(z; t_p) = \mathbf{A}(z)\mathbf{R}_{\mathbf{ss}}(z; t_p)\mathbf{A}^{\dagger}(z) \tag{13}$$

where t_p denotes the center of the pth *snapshot* of $\mathbf{R}_{\mathbf{ss}}(z; t_p)$ and $\mathbf{R}_{\mathbf{xx}}(z; t_p)$. Since the output correlation matrix depends now also on t_p, (12) becomes

$$\mathbf{R}_{\mathbf{uu}}(z; t_p) = \mathbf{W}(z)\mathbf{R}_{\mathbf{xx}}(z; t_p)\mathbf{W}^{\dagger}(z) = \mathbf{W}(z)\mathbf{A}(z)\mathbf{R}_{\mathbf{ss}}(z; t_p)\mathbf{A}^{\dagger}(z)\mathbf{W}^{\dagger}(z) \tag{14}$$

assuming $\mathbf{A}(z)$ and $\mathbf{W}(z)$ are time-invariant. Since we assume that the source signals s_m are mutually uncorrelated for all t, $\mathbf{R}_{\mathbf{ss}}(z; t_p)$ has a diagonal structure for every snapshot. In the special case where all source signals are also white, then $\mathbf{R}_{\mathbf{ss}}(z; t_p) = \mathbf{R}_{\mathbf{ss}}(0)$. However, we do not require that the source signals need to be white.

2.2 Cost function

Non-blind cost function In the blind source separation setup, the source signals s_m are unknown. Let us assume for the moment that $\mathbf{R_{ss}}(z;t_p)$ is known for P *snapshots* at t_p ($p{=}1..P$). In this case a possible cost function for the (non-blind) source separation task is (recall that $\|\,\mathbf{R_{uu}} - \mathbf{R_{ss}}\,\|_{\mathcal{F}} = d(\mathbf{R_{uu}}, \mathbf{R_{ss}})$)

$$\mathcal{J}_0(\mathbf{W}(z)) \triangleq \sum_{p=1}^{P} \mathcal{J}_0'(t_p) = \sum_{p=1}^{P} \|\,\mathbf{R_{uu}}(z;t_p) - \mathbf{R_{ss}}(z;t_p)\,\|_{\mathcal{F}}^2 \tag{15}$$

$$= \sum_{p} \|\,\mathbf{W}(z)\mathbf{R_{xx}}(z;t_p)\mathbf{W}^{\dagger}(z) - \mathbf{R_{ss}}(z;t_p)\,\|_{\mathcal{F}}^2 . \tag{16}$$

which obviously has a global minimum for $\mathbf{W}(z) = \mathbf{A}^{-1}(z)$.

Blind cost function In the *blind signal separation* (BSS) problem we do not know the true source correlation matrices $\mathbf{R_{ss}}(z;t_p)$. Hence, we need to replace them by some estimates $\hat{\mathbf{R}}_{\mathbf{ss}}(z;t_p)$ in order to still use the cost function (15). Since we assume that the source signals s_m are mutually uncorrelated, we also assume that $\mathbf{R_{ss}}(z;t_p)$ has a diagonal structure. Therefore a possible choice is

$$\hat{\mathbf{R}}_{\mathbf{ss}}(z;t_p) = \operatorname{diag}\big(\,\mathbf{R_{uu}}(z;t_p)\,\big) . \tag{17}$$

With this choice, we pretend that the diagonal entries of $\mathbf{R_{uu}}(z;t_p)$ coincide with those of $\mathbf{R_{ss}}(z;t_p)$ and simply ignore the nonzero off-diagonal elements of $\mathbf{R_{uu}}(z;t_p)$. The estimate (17) is consistent with the assumption of $\mathbf{R_{ss}}(z;t_p)$ having a diagonal structure. Inserting (17) into (15) yields the *blind* cost function

$$\mathcal{J}_1(\mathbf{W}(z)) \triangleq \sum_{p=1}^{P} \mathcal{J}_1'(t_p) \triangleq \sum_{p=1}^{P} \|\ \operatorname{off}\big(\,\mathbf{R_{uu}}(z;t_p)\,\big)\,\|_{\mathcal{F}}^2 \tag{18}$$

$$= \sum_{p} \|\ \operatorname{off}\big(\,\mathbf{W}(z)\mathbf{R_{xx}}(z;t_p)\mathbf{W}^{\dagger}(z)\,\big)\,\|_{\mathcal{F}}^2 . \tag{19}$$

The cost function (19) attains its global minimum for a polynomial matrix $\mathbf{W}(z)$ which jointly diagonalizes all input correlation matrices $\mathbf{R_{xx}}(z;t_p)$. Because of our assumptions, the global minimum of (18) is, in fact, zero: Inserting (4) into (14) gives $\mathbf{R_{uu}}(z;t_p) = \mathbf{P}\,\mathbf{D}(z)\,\mathbf{R_{ss}}(z;t_p)\,\mathbf{D}^{\dagger}(z)\mathbf{P}^T$ which has a diagonal structure. In order to prevent the trivial solution $\mathbf{W}(z) \equiv \mathbf{0}$, which obviously minimizes (18) as well, we need to impose some additional constraints on $\mathbf{W}(z)$.

The optimization problem defined in (18) subject to some constraints, is referred to as a *joint-diagonalization problem.* In our case, we wish to find a polynomial matrix $\mathbf{W}(z)$ that jointly diagonalizes all products $\mathbf{W}(z)\mathbf{R_{xx}}(z;t_p)\mathbf{W}^{\dagger}(z)$. In fact, the cost function (18) can be seen, as the straightforward polynomial extention of a cost function commonly used in blind source separation for the instantaneous mixing case, see [2,3]. On the other hand, by setting $z = e^{j\omega}$ and evaluating ω at discrete frequency bins ω_i, (18) turns into a the cost function used in [4, 5]. There the assumption has been made that the cost function in each frequency bin can be decoupled from the other frequency bins and therefore treated separately. Algorithms which treat each bin separately seem to have decent bin-wise convergence properties. Unfortunately, as it has been reported in the literature, the bin-wise decoupling of the adaptation also leads to a bin-wise permutation ambiguity of the separated source signals, which is commonly known in the context of blind source separation as *the permutation problem.*

2.3 Iterative algorithm

Derivation of the gradient In order to minimize the cost function $\mathcal{J}_1$ we will use a steepest-descent algorithm. To this end, we need to derive the gradient of $\mathcal{J}_1$ with respect to the filter coefficients $\mathbf{W}_r$. We reformulate (18) in a similar way as carried out in [3]:

$$\begin{aligned}\mathcal{J}_1 &= \sum_p \left\| \mathbf{R_{uu}}(z;t_p) - \operatorname{diag}(\, \mathbf{R_{uu}}(z;t_p)\,) \right\|_{\mathcal{F}}^2 & (20)\\ &= \sum_p \left\| \mathbf{R_{uu}}(z;t_p) \right\|_{\mathcal{F}}^2 - \sum_p \left\| \operatorname{diag}(\, \mathbf{R_{uu}}(z)\,) \right\|_{\mathcal{F}}^2 \triangleq \mathcal{J}_1^{(a)} - \mathcal{J}_1^{(b)}\,. & (21)\end{aligned}$$

Hereby we exploited $\langle \mathbf{R_{uu}}(z;t_p), \operatorname{diag}(\, \mathbf{R_{uu}}(z;t_p)\,)\rangle_{\mathcal{F}} = \| \operatorname{diag}(\, \mathbf{R_{uu}}(z;t_p)\,)\|_{\mathcal{F}}^2$. We derive the gradient for the two terms in (21) separately. By using the definition (10), we obtain after a few steps the two gradients

$$\nabla_{\mathbf{W}_r} \mathcal{J}_1^{(a)} = 4 \sum_p \sum_\tau \mathbf{R_{uu}}(\tau;t_p)\, \mathbf{R_{ux}}(r-\tau;t_p) \tag{22}$$

$$\nabla_{\mathbf{W}_r} \mathcal{J}_1^{(b)} = 4 \sum_p \sum_\tau \operatorname{diag}(\, \mathbf{R_{uu}}(\tau;t_p)\,)\, \mathbf{R_{ux}}(r-\tau;t_p)\,. \tag{23}$$

By combining (22) and (23), we obtain the overall gradient

$$\begin{aligned}\nabla_{\mathbf{W}_r} \mathcal{J}_1 &= \nabla_{\mathbf{W}_r} \mathcal{J}_1^{(a)} - \nabla_{\mathbf{W}_r} \mathcal{J}_1^{(b)} \\ &= 4 \sum_p \sum_\tau \operatorname{off}\left(\, \mathbf{R_{uu}}(\tau;t_p)\,\right) \mathbf{R_{ux}}(r-\tau;t_p)\,. & (24)\end{aligned}$$

Update equation In the following, we consider only a finite interval, $\tau \in [-\tau_\mathrm{e}, \tau_\mathrm{e}]$, of $\mathbf{R_{uu}}(\tau;t_p)$ in the cost function (18). The slightly modified gradient (24) is then used to obtain the following time-domain update equation

$$\mathbf{W}_r[k+1] = \mathbf{W}_r[k] - 4\mu \sum_{p=1}^{P} \sum_{\tau=-\tau_\mathrm{e}}^{\tau_\mathrm{e}} \operatorname{off}\left(\, \mathbf{R_{uu}}(\tau;t_p)[k]\,\right) \mathbf{R_{ux}}(r-\tau;t_p)[k] \tag{25}$$

where $[k]$ denotes the kth iteration and

$$\mathbf{R_{ux}}(\tau;t_p)[k] = \sum_m \mathbf{W}_m[k]\, \mathbf{R_{xx}}(\tau-m;t_p) \tag{26}$$

$$\mathbf{R_{uu}}(\tau;t_p)[k] = \sum_m \sum_n \mathbf{W}_m[k]\, \mathbf{R_{xx}}(\tau-m+n;t_p)\, \mathbf{W}_n^H[k]\,. \tag{27}$$

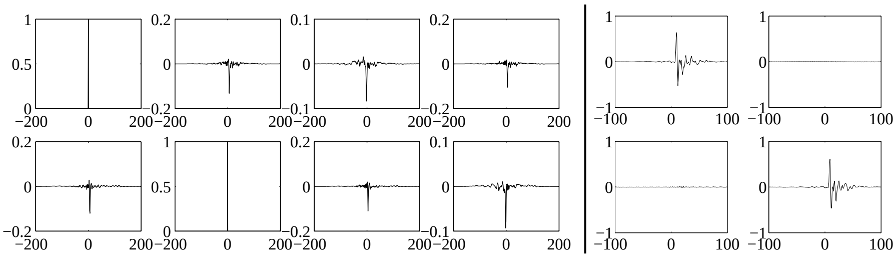

Fig. 1. Impulse responses of the 2×4 demixing filter $\mathbf{W}(z)$ (left) and the 2×2 global system $\mathbf{G}(z) = \mathbf{W}(z)\mathbf{A}(z)$ (right) after convergence.

Constraints In order to prevent the algorithm of converging to the trivial solution $\mathbf{W}(z) \equiv \mathbf{0}$, additional constraints need to be imposed on $\{\mathbf{W}_r\}$ during the adaptation. The most common ones are to constrain $\|\mathbf{W}(z)\|_{\mathcal{F}} \equiv 1$ or $\mathrm{diag}(\,\mathbf{W}(z)\,) \equiv \mathbf{I}$ (sometimes referred to as the *minimum distortion principle* [6]).

3 Efficient implementation in the frequency domain

In Fig. 2 we present an efficient implementation of the proposed joint-diagonalization algorithm for the convolutive mixing case. The algorithm works also for complex source signals and complex filter coefficients. Since (25), (26), and (27) are multichannel convolutional sums, we can compute them efficiently in the frequency domain by applying fast convolution techniques. Note that this procedure does not change the underlying cost function or the time-domain update equation if applied properly. The derivation and notation of the vectors are based on the same concepts as described in [7, Chapter 3 & Appendix F]. Even though the proposed algorithm is a major contribution of this paper, a detailed derivation is not possible at this point, due to lack of space.

4 Simulation example

To verify the performance of the proposed algorithm, we setup an artificial mixing system with $M_\mathrm{s} = 2$ source signals and $M_\mathrm{x} = 4$ sensors. The mixing system $\mathbf{A}(z)$ is extracted from real measured HRTFs (head-related transfer functions) and have length $N_\mathrm{a} = 100$. The correlation matrices $\mathbf{R}_{\mathbf{ss}}(z; t_p)$ of $P = 3$ snapshots are generated artificially to be diagonal matrices with diagonal elements $(\mathbf{R}_{\mathbf{ss}}(z; t_p))_{m,m} = b(z; t_p)\, b^{\dagger}(z; t_p)$ where $b(z; t_p)$ are randomly chosen filters of length 20. This setup simulates the case where the source signals are non-stationary *and* non-white. The input correlation matrices are computed as $\mathbf{R}_{\mathbf{xx}}(z; t_p) = \mathbf{A}(z)\mathbf{R}_{\mathbf{ss}}(z; t_p)\mathbf{A}^{\dagger}(z)$. This artificial generation of $\{\mathbf{R}_{\mathbf{ss}}(z; t_p)\}$ guarantees that the global minimum of $\mathcal{J}_1$ is, in fact, zero. The demixing system $\mathbf{W}(z)$ is a 2×4 matrix where each filter has length 199, ($N_\mathrm{w} = 99$). The impulse responses of $\mathbf{W}(z)$ and $\mathbf{G}(z)$ after convergence are shown in Fig. 1. From the vanishing off-diagonal impulse responses of $\mathbf{G}(z)$ it is clearly seen, that the proposed algorithm can perform almost perfect signal separation. Fig. 1 also indicates that the proposed algorithm does not suffer any permutation problem.

5 Conclusions

Many BSS algorithms for the convolutive mixing case are straightforward extensions of an instantaneous-mixing-case algorithm in the sense that the chosen cost function and corresponding update rule are applied independently in every frequency bin. This approach usually causes a so-called permutation problem. Our approach differs insofar that we define a single global cost function which penalizes all cross-correlations over all time-lags. Even though the update equation for the demixing system is derived in the time domain, most of the computation is carried out in the frequency domain. We would like to point out, that our main motivation to go into frequency domain was because of computational efficiency, similar to [8], and not to decouple the update equations, as done in [4, 5]. Consequently, the proposed algorithm does not suffer from a so-called permutation problem, likewise to related pure time-domain algorithms described in [6, 9, 10].

References

1. E. Kreyszig, *Introductory Functional Analysis with Applications.* John Wiley & Sons, 1978.
2. A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
3. M. Joho and K. Rahbar, “Joint diagonalization of correlation matrices by using Newton methods with application to blind signal separation,” in *Proc. SAM*, Rosslyn, VA, Aug. 4–6, 2002, pp. 403–407.
4. L. Parra and C. Spence, “Convolutive blind separation of non-stationary sources,” *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, 2000.
5. K. Rahbar and J. P. Reilly, “Blind source separation algorithm for MIMO convolutive mixtures,” in *Proc. ICA*, San Diego, CA, Dec. 9–12, 2001, pp. 224–229.
6. K. Matsuoka and S. Nakashima, “Minimal distortion principle for blind source separation,” in *Proc. ICA*, San Diego, CA, Dec. 9–12, 2001, pp. 927–932.
7. M. Joho, “A systematic approach to adaptive algorithms for multichannel system identification, inverse modeling, and blind identification,” Ph.D. dissertation, ETH Zürich, Dec. 2000. http://e-collection.ethbib.ethz.ch/show?type=diss&nr=13783
8. M. Joho and P. Schniter, “Frequency-domain realization of a multichannel blind deconvolution algorithm based on the natural gradient,” in *Proc. ICA*, Nara, Japan, Apr. 1–4, 2003, pp. 543–548.
9. D. W. E. Schobben, “Efficient adaptive multi-channel concepts in acoustics: Blind signal separation and echo cancellation,” Ph.D. dissertation, Technical University Eindhoven, 1999.
10. B. S. Krongold and D. L. Jones, “Blind source separation of nonstationary convolutively mixed signals,” in *Proc. SSAP*, Pocono Manor, PA, Aug. 14–16, 2000, pp. 53–57.

FCONVBSS-JD

Definitions:

$$\tilde{\mathbf{P}}_N \triangleq \begin{bmatrix} \mathbf{I}_{N+1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{C-2N-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_N \end{bmatrix}$$

$$\tilde{\mathbf{w}}_{mn}[k] \triangleq (w_{mn,0}[k], \ldots, w_{mn,N_\mathrm{w}}[k], 0, \ldots, 0, w_{mn,-N_\mathrm{w}}[k], \ldots, w_{mn,-1}[k])^T$$

Initialization ($\forall\, m, n, p$):

$$\tilde{\mathbf{w}}_{mn}[0] := \begin{cases} (1, 0, \ldots, 0)^T & \text{for } m = n \\ (0, 0, \ldots, 0)^T & \text{for } m \neq n \end{cases}$$

$$\bar{\mathbf{w}}_{mn}[0] := \mathrm{FFT}(\, \tilde{\mathbf{w}}_{mn}[0]\,)$$

$$r^{(p)}_{x_m x_n}(\tau) := E\{x_m(t_p)\, x_n^*(t_p - \tau)\} \qquad \text{for } \tau \in \{-\tau_\mathrm{xx}, \ldots, \tau_\mathrm{xx}\}$$

$$\tilde{\mathbf{r}}^{(p)}_{x_m x_n} := (r^{(p)}_{x_m x_n}(0), \ldots, r^{(p)}_{x_m x_n}(\tau_\mathrm{xx}), 0, \ldots, 0, r^{(p)}_{x_m x_n}(-\tau_\mathrm{xx}), \ldots, r^{(p)}_{x_m x_n}(-1))^T$$

$$\bar{\mathbf{r}}^{(p)}_{x_m x_n} := \mathrm{FFT}(\, \tilde{\mathbf{r}}^{(p)}_{x_m x_n}\,)$$

For each loop k do ($\forall\, m, n, p$):

$$\bar{\mathbf{r}}^{(p)}_{u_m x_n}[k] := \sum\nolimits_{l=1}^{M_\mathrm{x}} \bar{\mathbf{w}}_{ml}[k] \odot \bar{\mathbf{r}}^{(p)}_{x_l x_n}$$

$$\bar{\mathbf{r}}^{(p)}_{u_m u_n}[k] := \sum\nolimits_{l=1}^{M_\mathrm{x}} \bar{\mathbf{w}}^*_{nl}[k] \odot \bar{\mathbf{r}}^{(p)}_{u_m x_l}[k]$$

$$\tilde{\mathbf{r}}^{(p)}_{u_m u_n}[k] := \mathrm{IFFT}(\, \bar{\mathbf{r}}^{(p)}_{u_m u_n}[k]\,)$$

$$\tilde{\mathbf{e}}^{(p)}_{mn}[k] := \begin{cases} \mathbf{0} & \text{for } m = n \\ \tilde{\mathbf{P}}_{\tau_\mathrm{e}}\, \tilde{\mathbf{r}}^{(p)}_{u_m u_n}[k] & \text{for } m \neq n \end{cases}$$

$$\bar{\mathbf{e}}^{(p)}_{mn}[k] := \mathrm{FFT}(\, \tilde{\mathbf{e}}^{(p)}_{mn}[k]\,)$$

$$\mathit{\Delta}\bar{\mathbf{w}}_{mn}[k] := \sum\nolimits_{p=1}^{P} \sum\nolimits_{l=1}^{M_\mathrm{u}} \bar{\mathbf{e}}^{(p)}_{ml}[k] \odot \bar{\mathbf{r}}^{(p)}_{u_l x_n}[k]$$

$$\bar{\mathbf{w}}'_{mn}[k+1] := \begin{cases} \bar{\mathbf{w}}_{mn}[k] & \text{for } m = n \\ \bar{\mathbf{w}}_{mn}[k] - \mu \cdot \mathit{\Delta}\bar{\mathbf{w}}_{mn}[k] & \text{for } m \neq n \end{cases}$$

$$\bar{\mathbf{w}}_{mn}[k+1] := \mathrm{FFT}\left(\tilde{\mathbf{P}}_{N_\mathrm{w}}\ \mathrm{IFFT}\left(\bar{\mathbf{w}}'_{mn}[k+1]\right)\right)$$

Fig. 2. FCONVBSS-JD : Frequency-domain implementation of CONVBSS-JD All vectors have length C, which is also the FFT size. Since the linear convolutions are embedded in cyclic convolutions, the projection matrices $\tilde{\mathbf{P}}$ are necessary to extract only the linear-convolution part. In order to prevent circular wrap-around effects affecting the updates, the FFT size needs to be chosen large enough. The notation and concept behind the arrangement of the vector elements are taken from [7, Chapter 3 & Appendix F].