Source separation based on second order statistics — an algebraic approach

Ulf Lindgren
Department of Applied Electronics
Chalmers University of Technology
S-412 96 Göteborg, Sweden
lindgren@ae.chalmers.se

Alle-Jan van der Veen
Department of Electrical Engineering
Delft University of Technology
2628 CD Delft, The Netherlands
allejan@cas.et.tudelft.nl

Two unknown non-white stochastic sources (e.g. speech signals) are dynamically mixed by an unknown multipath channel and subsequently measured by two sensors. The objective is to construct an inverse filter that separates the two signals, based only on their independence. It is known that, under certain conditions, second-order statistics provide sufficient information to identify the filter. In contrast to the usual cost function optimization techniques, we propose an algorithm that computes the filter coefficients algebraically, *using linear algebra techniques such as the singular value decomposition.*

Keywords: *stochastic signal separation*

1. Introduction

We consider the problem of separating two mutually uncorrelated non-white stochastic sources jointly received over two unknown multipath channels. A number of papers have been published in this context, under various assumptions on the signals or the channels, and using various techniques; see [2–5, 7–9, 12, 13] and the references therein. Techniques may broadly be classified as (a) block-methods based on high-order statistics (second and fourth-order cumulants), (b) adaptive methods based on optimization of a blind cost function (or nonlinear contrast function), (c) maximum-likelihood estimation, presuming the source distributions are known. In many cases, a limited scenario with only scalar mixtures is considered.

The algorithm proposed in this paper is a block-method based on second-order statistics of the measurement data only. The parameters of the inverse filter are to be found such that the resulting filtered output signals $y_1(t)$ and $y_2(t)$ have zero cross-covariance function. Assuming a certain filter structure, the resulting conditions take the form of bilinear equations. The usual approach at this point is to set up a cost function whose minimum coincides with the solution of the equations, and to apply a stochastic gradient or Newton-type search algorithm to find the minimum [5]. Our main point is the observation that the equations can also be solved *algebraically*, via a singular value decomposition (SVD). This gives an exact solution to the problem in case the covariance data is exact. With estimated covariances, a subsequent step is needed, in which we have to find a linear combination of a collection of matrices such that the result has rank 1. A similar problem arose in the context of separation of constant-modulus signals [11].

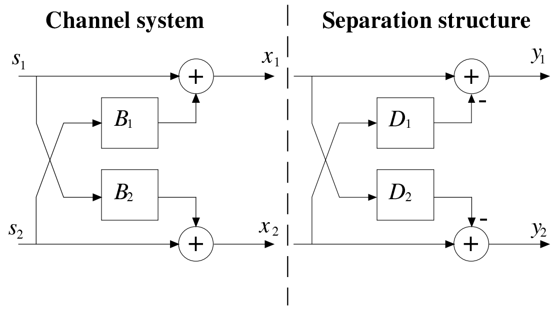

Figure 1. Separation scenario

2. Problem formulation

The data model that we consider in this paper is depicted in figure 1. The source signals are $s_1(n)$ and $s_2(n)$, which are linearly filtered white noise processes $\xi_1(n)$, $\xi_2(n)$. We make the following assumptions:

C1: $\xi_1(n)$ and $\xi_2(n)$ are realizations of mutually uncorrelated identically distributed sequences with non-zero variance and zero mean.

C2: $s_1(n)$ and $s_2(n)$ are generated by convolving $\xi_1(n)$ and $\xi_2(n)$ with two different asymptotically stable rational filters.

[0]IEEE 7-th workshop on SSAP, Corfu (Greece), June 1996.

The source signals are measured via an unknown multichannel, with outputs $x_1(n)$ and $x_2(n)$. The structure of the channel is supposed to consist of a single direct path for the transfer of s_1 to x_1 and s_2 to x_2, and short FIR multipaths $B_1(q^{-1})$ and $B_2(q^{-1})$ for the crosscoupling s_1 to x_2 and s_2 to x_1. The objective is to retrieve $s_1(n)$, $s_2(n)$ from $x_1(n)$, $x_2(n)$. This can be done in a two step procedure where step one is a separation and step two a post-filtering: (1) from x_1, x_2, find $y_1(n) = Hs_1(n)$ and $y_2(n) = Hs_2(n)$, where $H(q^{-1})$ is some FIR filter; (2) inverse filter the sequences $Hs_1(n)$ and $Hs_2(n)$ with $H^{-1}(q^{-1})$ to retrieve s_1 and s_2. Here, we focus on the first step: the actual signal separation.

The separation structure is a direct feedforward filter as depicted in figure 1, where $D_1(q^{-1})$ and $D_2(q^{-1})$ are adaptive FIR filters. When $D_1 = B_1$, $D_2 = B_2$, the separation structure is equal to the channel inverse times the filter $H(q^{-1}) = 1-D_1(q^{-1})D_2(q^{-1})$, in which case the filter outputs y_1, y_2 are equal to Hs_1, Hs_2. More generally,

$$\begin{bmatrix} y_1^{(\theta)} \\ y_2^{(\theta)} \end{bmatrix} = \begin{bmatrix} 1 & -D_1 \\ -D_2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1-B_2D_1 & B_1-D_1 \\ B_2-D_2 & 1-B_1D_2 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix} \qquad (1)$$

where $\theta = [d_{10} \ \ldots \ d_{1U-1} \ d_{20} \ \ldots \ d_{2V-1}]^T = \begin{bmatrix} \mathbf{d}_1^T & \mathbf{d}_2^T \end{bmatrix}^T$ is the parameter vector of the separation structure. To enable separation, the filter lengths U, V of D_1 and D_2 should be at least as large as the channel lengths, L_1 and L_2. This is only possible if the natural assumption

C3: $L_1 \leq U$ and $L_2 \leq V$

is introduced. Condition C3 assures that the separation structure is in the model class.

In order to recover the sources we require that

C4: $H(q^{-1})$ is minimum phase.

This is natural since $H(q^{-1})$ has a stable inverse only if it is minimum phase. The condition C4 is fulfilled if $|B_1(e^{j\omega})B_2(e^{j\omega})| < 1$ for all $\omega \in [0, 2\pi]$, cf. [1].

3. An algebraic separation algorithm

The proposed algorithm is based on finding the coefficients θ of the separation filter such that the filter outputs y_1 and y_2 are mutually uncorrelated. Let $R_{y_1y_2}^{\theta}(l) = E(y_1(n)y_2(n-l))$ be the cross-correlation between the filtered signals. We will only force independence with respect to second order statistics, i.e. the cross-correlation of y_1 and y_2 is equal to zero for a selected number of $(2L+1)$ lags [7]:

$$R_{y_1y_2}^{\theta}(l) = 0\,, \qquad -L \leq l \leq L\,. \qquad (2)$$

The cross-correlation $R_{y_1y_2}^{(\theta)}(l)$ is, under assumption C1 and C2, given in terms of the measured data x_1, x_2 as

$$\begin{aligned} R_{y_1y_2}^{\theta}(l) =& R_{x_1x_2}(l) - \mathbf{d}_1^T\mathbf{r}_{x_2x_2}(l) - \mathbf{d}_2^T\mathbf{r}_{x_1x_1}(l) + \\ &+ \mathbf{d}_1^T\mathbf{R}_{x_2x_1}(l)\mathbf{d}_2, \end{aligned} \qquad (3)$$

where

$$\mathbf{r}_{x_1x_1}(l) = [R_{x_1x_1}(l) \ \ldots \ R_{x_1x_1}(l+V-1)]^T \qquad (4)$$

$$\mathbf{r}_{x_2x_2}(l) = [R_{x_2x_2}(l) \ \ldots \ R_{x_2x_2}(l-U+1)]^T \qquad (5)$$

$$\mathbf{R}_{x_2x_1}(l) = [\mathbf{r}_{x_2x_1}(l) \ \ldots \ \mathbf{r}_{x_2x_1}(l+V-1)] \qquad (6)$$

$$\mathbf{r}_{x_2x_1}(l) = [R_{x_2x_1}(l) \ \ldots \ R_{x_2x_1}(l-U+1)]^T. \qquad (7)$$

Thus, the separation problem reduces to solving a system of bilinear equations. In [5] it is proven that there are at least as many equations as unknowns under conditions C1-3 with the exception of the static channel and white sources. By adding the condition C4, this identification problem becomes parameter identifiable (apart from static channels), cf. [6].

The equations (3) with left hand side equal to zero can be solved iteratively, in conjunction with a criterion, by means of gradient minimization techniques, cf. [5]. However, since such techniques are usually bothered by local minima and require accurate initial points, it is interesting also to consider an exact solution of the equations, as follows.

The idea is to rewrite the bilinear equations (3) in matrix form, using Kronecker products to collect all unknowns into a single (structured) parameter vector. This produces

$$\begin{bmatrix} R_{y_1y_2}(-L) \\ \vdots \\ R_{y_1y_2}(L) \end{bmatrix} = \mathbf{P} \begin{bmatrix} \mathbf{d}_2 \otimes \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_1 \\ 1 \end{bmatrix} = \mathbf{0}\,, \qquad (8)$$

where '$\otimes$' is the Kronecker product, and

$$\mathbf{P} = \begin{bmatrix} \operatorname{vec}(\mathbf{R}_{x_2x_1}(-L))^T & \mathbf{r}_{x_1x_1}^T(-L) & \mathbf{r}_{x_2x_2}^T(-L) & R_{x_1x_2}(-L) \\ \vdots & \vdots & \vdots & \vdots \\ \operatorname{vec}(\mathbf{R}_{x_2x_1}(L))^T & \mathbf{r}_{x_1x_1}^T(L) & \mathbf{r}_{x_2x_2}^T(L) & R_{x_1x_2}(L) \end{bmatrix}.$$

'vec' denotes the vectoring operation which stacks all columns of a matrix into a single column. Thus, the problem is equivalent to finding a vector with a certain structure in the null space of the data matrix $\mathbf{P}$. This null space can be determined, or estimated, by a singular value decomposition of $\mathbf{P}$. Thus let a basis for the null space be given by $\{\mathbf{v}_1, \cdots, \mathbf{v}_\delta\}$, where δ is the dimension of the null space. Since the precise basis is arbitrary, the problem is to find a linear combination of these vectors such that we obtain a vector with the required structure, i.e. to find $\lambda_1, \cdots, \lambda_\delta$ such that

$$\lambda_1 \mathbf{v}_1 + \cdots + \lambda_\delta \mathbf{v}_\delta = \begin{bmatrix} \mathbf{d}_2 \otimes \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_1 \\ 1 \end{bmatrix}. \tag{9}$$

To make this equivalent problem more tractable, we move from vectors to matrices. For a vector $\mathbf{x}$ partitioned as

$$\begin{aligned} \mathbf{x} &= [\mathbf{x}_1^T \quad \mathbf{x}_2^T \quad \mathbf{x}_3^T \quad x_4]^T \qquad (10) \\ &= [x_{11} \cdots x_{1,UV} \; x_{21} \cdots x_{2V} \; x_{31} \cdots x_{3U} \; x_4]^T \end{aligned}$$

define the operator

$$\mathrm{mat}(\mathbf{x}) := \begin{bmatrix} \mathrm{vec}^{-1}(\mathbf{x}_1) & \mathbf{x}_3 \\ \mathbf{x}_2^T & x_4 \end{bmatrix} \tag{11}$$

where $\mathrm{vec}(\mathbf{M})$ is a vectorization of the matrix $\mathbf{M}$ and $\mathrm{vec}(\mathrm{vec}^{-1}(\mathbf{x}_1)) = \mathbf{x}_1$. Note that

$$\mathrm{mat}(\begin{bmatrix} \mathbf{d}_2 \otimes \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_1 \\ 1 \end{bmatrix}) = \begin{bmatrix} \mathbf{d}_1 \mathbf{d}_2^T & \mathbf{d}_1 \\ \mathbf{d}_2^T & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{d}_1 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{d}_2^T & 1 \end{bmatrix}.$$

Denote $\mathbf{V}_1 = \mathrm{mat}(\mathbf{v}_1), \cdots, \mathbf{V}_\delta = \mathrm{mat}(\mathbf{v}_\delta)$. Equation (9) is equivalent to finding $\lambda_1, \cdots, \lambda_\delta$ such that

$$\mathbf{V}_\lambda := \lambda_1 \mathbf{V}_1 + \cdots + \lambda_\delta \mathbf{V}_\delta = \begin{bmatrix} \mathbf{d}_1 \\ 1 \end{bmatrix} \begin{bmatrix} \mathbf{d}_2^T & 1 \end{bmatrix}. \tag{12}$$

Basically, we have to select λ_k's such that the resulting linear combination of matrices $\mathbf{V}_\lambda$ is rank 1, in which case it can always be scaled and factored into the required structure.

What is the value of δ? At first sight, given enough conditions (lags) we would expect $\delta = 1$, since the solution to the separation problem is usually unique. However, the Toeplitz structure of $\mathbf{R}_{x2x1}$ adds extra vectors to the null space of $\mathbf{P}$: certain columns of $\mathbf{P}$ are duplicated, which reduces its rank. The number of repeated entries in the Toeplitz matrix is $UV - (U + V - 1) = (U - 1)(V - 1)$, so that we expect $\delta = 1 + (U - 1)(V - 1)$. The resulting null space basis also has structure: e.g. for $U = 3, V = 3$, a possible matrix basis is of the form

$$\{\mathbf{V}_1, \cdots, \mathbf{V}_5\} = \{ \left[\begin{array}{ccc|c} 1 & & & \\ & -1 & & \\ & & 0 & \\ \hline & & & 0 \end{array}\right], \left[\begin{array}{ccc|c} 0 & & & \\ & 1 & & \\ & & -1 & \\ \hline & & & 0 \end{array}\right], \left[\begin{array}{ccc|c} 0 & 1 & & \\ & 0 & -1 & \\ & & 0 & \\ \hline & & & 0 \end{array}\right], \left[\begin{array}{ccc|c} 0 & & & \\ 1 & 0 & & \\ & -1 & 0 & \\ \hline & & & 0 \end{array}\right], \left[\begin{array}{ccc|c} * & * & * & \\ * & 0 & 0 & \mathbf{d}_1 \\ * & 0 & 0 & \\ \hline & \mathbf{d}_2^T & & 1 \end{array}\right] \}. \tag{13}$$

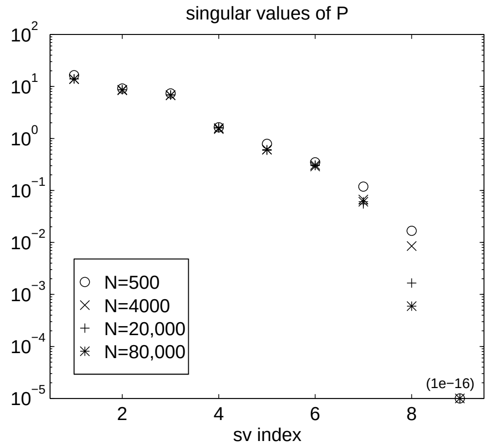

Figure 2. Singular values of $\mathbf{P}$

If we simply remove the duplicate columns of $\mathbf{P}$, then the 'trivial' null space solutions ($\mathbf{V}_1, \cdots, \mathbf{V}_4$) are suppressed. Only one vector in the null space is left, corresponding to $\mathbf{V}_5$ in (13). Hence, estimates of $\mathbf{d}_1$, $\mathbf{d}_2$ are immediately available, even without solving the rank-1 problem.

The above is true only for perfect knowledge of the covariance lags, i.e. for an infinite amount of data. In actuality, the estimates of these lags converge only slowly to their true values, and the null space is not well-determined. For accuracy reasons it is usually necessary to overestimate the value of δ, and actually search for $\lambda_1, \cdots, \lambda_\delta$ that produces $\mathbf{V}_\lambda$ in (12) that is as close to rank 1 as possible. This is reminiscent of the problem considered (and solved) in [11], where it is shown how a simultaneous diagonalization of (square) matrices $\mathbf{V}_1, \cdots, \mathbf{V}_\delta$ provides good estimates of the λ_k's. The simulation results reported in section 4 are based on a blunt application of this diagonalization algorithm, followed by a few steps of an optimization routine to improve the λ_k's. Although the results are reasonable, it should be remarked that the diagonalization algorithm is theoretically not well motivated for this application, because unlike the case in [11], we now expect only one solution $[\lambda_1 \; \cdots \; \lambda_\delta]$, rather than δ independent solutions. This means that the $\mathbf{V}_k$ need not be simultaneously diagonalizable.

4. Simulations

We investigate the performance of the algorithm by simulation. In accordance with conditions C1 and C2, the source signals $s_1(n)$ and $s_2(n)$ are generated by filtering two mutually uncorrelated white Gaussian noise sequences through two autoregressive filters. One filter has a complex pole pair at radius 0.8 and angle $\pm\pi/4$; the other filter has a radius of 0.8 and angle $\pm 3\pi/4$. The channel in this simulation consists

Method	N	*Mean*				*Variance* $\times 10^{-3}$			
		d_{10}	d_{11}	d_{20}	d_{21}	d_{10}	d_{11}	d_{20}	d_{21}
True/CRB	500	0.5	-0.1	0.7	0.3	0.134	0.130	0.104	0.104
	2000					0.034	0.033	0.026	0.026
	4000					0.017	0.016	0.013	0.013
Algebraic	500	0.499	-0.098	0.701	0.299	0.691	0.719	2.15	1.88
	2000	0.502	-0.098	0.697	0.297	0.162	0.157	0.480	0.481
	4000	0.500	-0.100	0.700	0.299	0.086	0.086	0.232	0.229
Recursive	500	0.500	-0.100	0.697	0.302	0.718	0.582	1.21	2.57
	2000	0.500	-0.100	0.700	0.300	0.043	0.048	0.032	0.035
	4000	0.500	-0.100	0.700	0.300	0.017	0.016	0.016	0.016
Weinstein	500	0.748	0.074	0.748	0.419	1677	873	1677	254
	2000	0.651	0.024	0.651	0.480	62.3	276	623	47.5
	4000	0.668	0.055	0.668	0.493	1334	104	1334	101

Table 1. Mean value and variance of the estimated filter coefficients

of two filters $B_1(q^{-1}) = 0.5 - 0.1q^{-1}$ and $B_2(q^{-1}) = 0.7 + 0.3q^{-1}$. The correlation matrices (4)–(7) are estimated from $N = 500$, 2000, and 4000 samples of x_1, x_2. We took $L = 4$ lags into account, which gives a total of 9 equations for 4 unknowns. The Cramér-Rao Bound (CRB) for this scenario is derived as $N\text{Var}_\theta = [0.067, 0.065, 0.052, 0.052]$, cf. [10].

A total of 200 independent runs were performed for each sample size. The estimated mean value and parameter variance for the present algebraic algorithm are given in table 1, along with two other algorithms. The "recursive" algorithm is basically a stochastic Newton search algorithm based on [5], and the "Weinstein" algorithm is the one found in [13].

For the algebraic algorithm, theoretically $\delta = 2$, but we have used $\delta = 3$ because even for $N = 4000$ there is no clear gap between the large and small singular values, as is illustrated in figure 2. Even so, the algebraic algorithm performed less good than the recursive method and did not reach the CRB. Given exact (rather than estimated) covariance data, $\mathbf{P}$ does have precisely two zero singular values, and the algorithm did produce the exact solution.

It is known that the "Weinstein" algorithm cannot separate the sources unless $b_{10} = b_{20}$, and this shows up in the results. For a scenario where $b_{10} = b_{20}$ the algorithm works, but yields estimates with a higher variance than the other two algorithms. Note that the variance for "Weinstein" is larger for $N = 4000$ than for $N = 2000$. This is due to large deviations for some, typically two, parameter trajectories.

References

[1] H. Broman, U. Lindgren, H. Sahlin, and P. Stoica. Source separation: A TITO system identification approach. *Submitted to IEEE Trans. Signal Proc.*, 1995.

[2] J. Cardoso, A. Belouchrani, and B. Laheld. A new composite criterion for adaptive and iterative blind source separation. In *Proc. IEEE ICASSP*, pages IV:273–276, 1994.

[3] C. Jutten and J. Herault. Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing*, 24:1–10, June 1991.

[4] R. Lambert. A new method for source separation. In *Proc. IEEE ICASSP*, pages 2116–2119, 1995.

[5] U. Lindgren. Signal separation. Technical report, No. 201L (licentiate dissertation), Chalmers University of Technology, Göteborg, Sweden, 1995.

[6] U. Lindgren and H. Broman. Source separation: Using a criterion based on second order statistics. Technical report, Technical Report CTH-TE-36, Chalmers University of Technology, Göteborg, Sweden, 1995.

[7] U. Lindgren, T. Wigren, and H. Broman. On local convergence of a class of blind separation algorithms. *IEEE Trans. Signal Proc.*, 43:3054–3058, Dec. 1995.

[8] P. Loubaton and P. Regalia. Blind deconvolution of multivariate signals using adaptive FIR lossless filters. In J. V. e.a., editor, *Signal Processing VI: Proc. EUSIPCO-92*, pages 1061–1064 vol.2. Elsevier, 1992.

[9] M. Nájar, M. Lagunas, and I. Bonet. Blind wideband source separation. In *Proc. IEEE ICASSP*, pages IV:65–68, 1994.

[10] H. Sahlin and U. Lindgren. "The Asymptotic Cramér-Rao Lower Bound for Blind Signal Separation". To appear in "The Proceedings of the 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing", Corfu, Greece, 1996.

[11] A. J. van der Veen and A. Paulraj. An analytical constant modulus algorithm. *IEEE Trans. Signal Proc.*, 44(5), May 1996.

[12] S. Van Gerven and D. Van Compernolle. On the use of decorrelation in scalar signal separation. In *Proc. IEEE ICASSP*, pages III:57–60, 1994.

[13] E. Weinstein, M. Feder, and A. V. Oppenheim. Multichannel signal separation by decorrelation. *IEEE Trans. Speech Audio Processing*, 1:405–413, Oct. 1993.