TV-SOBI: AN EXPANSION OF SOBI FOR LINEARLY TIME-VARYING MIXTURES

Arie Yeredor

Department of Electrical Engineering - Systems
Tel-Aviv University, Tel-Aviv, 69978, Israel
e-mail: arie@eng.tau.ac.il

ABSTRACT

We address the problem of blind separation of linear instantaneous mixtures of stationary, mutually uncorrelated sources with distinct spectra, where the mixing matrix is time-varying (rendering the observations non-stationary). The time variation of the mixing is parameterized using the assumption of linear time-dependence. Relying on second-order statistics in the framework of Second-Order Blind Identification (SOBI Algorithm, Belouchrani et al., 1997), we offer an expansion of SOBI, such that the further parameterization is properly accommodated. In actual situations of slowly varying mixtures, such first-order parameterization thereof enables the estimation of the necessary statistics over longer observation intervals than those possible with classical static methods (essentially zero-order approximations). The prolonged validity of the approximate model is thus utilized in improving the statistical stability of the estimates. In this paper we identify the "raw" statistics that need to be estimated from the data, and then propose several approaches for the estimation of the mixing model parameters, pointing out some trade-offs involved. We demonstrate the enhanced performance relative to conventional SOBI when applied to time-varying mixtures.

1. INTRODUCTION

Traditionally, the blind source separation (BSS) problem addresses so-called "static" mixture models, whereby the term "static" encompasses two distinct assumptions - namely that the mixtures are both time-invariant and instantaneous. For non-static mixtures, considerable work has been addressed towards mixtures that are not instantaneous (namely, convolutive mixtures). However, the possibility of time-variations of (instantaneous) mixtures has seen little explicit treatment. It is common practice to rely on the adaptive form of most static BSS algorithms (see, e.g., [1]) to be able to track slow changes in the mixing parameters.

Nevertheless, adaptive algorithms rely on an explicit or implicit memory span, which can often be tuned by proper setting of some effective adaptation "step-size". When time-variations in the true mixing coefficients are slow enough, the step-size can be set to a relatively small value, which in turn implies a longer memory span. In such cases the necessary statistics are in effect estimated using a considerable amount of history data. Consequently, at the cost of slow tracking ability, the statistical stability of the estimation is improved. Conversely, when the changes in the true mixing coefficient are faster, the use of a long memory span can introduce bias into the estimation process. Therefore, the adaptation step-size has to be increased, which effectively results in improved tracking at the cost of decreased statistical stability.

A similar effect is introduced by time-varying mixtures into the framework of batch processing BSS algorithms, where the estimation of the mixing parameters, as well as the demixing, are performed on a given data block. When such algorithms assume a static model, they can accommodate time-varying mixtures with acceptable performance only up to a certain extent of variation rate. Faster variation rates pose an upper limit on the block-length that can be used under the false assumption of a constant mixing. Consequently, when the block length is reduced, the statistical stability of the estimation is degraded.

In this paper we propose further parameterization for time-varying mixtures, involving an explicit term for linear variation in time. Using proper estimation of the necessary statistics, we present an algorithm for the estimation of the time-variation parameters in addition to the mixing parameters. We then apply the respectively estimated time-varying demixing, and demonstrate an improved overall Interference to Signal Ratio (ISR).

While in practical situations the time variations may not obey the presumed linear time-dependence model, such a model nonetheless offers a first-order approximation thereof. Often this approximation is superior at least to the implied zero-order approximations used when applying algorithms that are originally intended for truly static mixtures.

Although this concept of linear modeling of time-varying mixtures can be applied to various BSS algorithms, we chose to demonstrate it under the framework of second order demixing of stationary source signals with distinct spectra (the static version is the well-known Second Order Blind Identification (SOBI) algorithm, by Belouchrani et al., [2]).

2. PROBLEM FORMULATION

Let $\boldsymbol{s}[n] = [s_1[n]s_2[n]...s_M[n]]^T$ be a vector of M zero-mean Wide-Sense Stationary (WSS) mutually uncorrelated source signals with unknown but distinct spectra. By the term "distinct" spectra we refer to the property that no pair of signals in the set have spectra (or autocorrelation sequences) that are merely scaled versions of one another.

A general time-varying (noiseless) mixture model would be

$$\boldsymbol{x}[n] = \boldsymbol{A}[n]\boldsymbol{s}[n] \quad (1)$$

where $\boldsymbol{x}[n] = [x_1[n]x_2[n]...x_M[n]]^T$ are the M observed signals.

We now assume that the variation of $\boldsymbol{A}[n]$ in time is linear, according to the following model:

$$\boldsymbol{A}[n] = (\boldsymbol{I} - n\boldsymbol{\mathcal{E}})\boldsymbol{A}_0 \quad (2)$$

where $\boldsymbol{I}$ is the $M \times M$ identity matrix, $\boldsymbol{A}_0$ is the mixing matrix at time $n = 0$, and the matrix $\boldsymbol{\mathcal{E}}$ reflects the time-dependence parameters, and will be referred to as the "relative rate" matrix. Note that this formulation involves no loss of generality with respect to a linear time change in $\boldsymbol{A}[n]$, provided that $\boldsymbol{A}_0$ is nonsingular. Thus, if $\boldsymbol{A}[n]$ is to change linearly in n from $\boldsymbol{A}_0$ at time $n = 0$ to some $\boldsymbol{A}_N$ at time $n = N$, this would be attained by setting $\boldsymbol{\mathcal{E}} = \frac{1}{N}(\boldsymbol{A}_N\boldsymbol{A}_0^{-1} - \boldsymbol{I})$.

When slow variations are involved, $\boldsymbol{\mathcal{E}}$ is usually a small matrix ($\boldsymbol{\mathcal{E}} << \boldsymbol{I}$), and we chose to use the formulation of (2), which is reminiscent of the notion of a "relative" or "natural" gradient ([3, 4, 5]), in order to facilitate further derivations. In addition, for further simplification, and for reasons to be revealed later on, we shall assume (without loss of generality) that the observation interval is $[-N : N]$, symmetrically positioned around $n = 0$. To conclude, the model is given by:

$$\boldsymbol{x}[n] = (\boldsymbol{I} + n\boldsymbol{\mathcal{E}})\boldsymbol{A}_0\boldsymbol{s}[n] \quad , \quad n = -N, -N+1, ...N \quad (3)$$

3. DERIVATION OF THE ALGORITHMS

Although the source signals are WSS, the observations are obviously nonstationary, due to the time-varying mixture. Let us denote the following:

$$\begin{aligned}\boldsymbol{R}_x[n,l] &\triangleq E\left[\boldsymbol{x}[n+l]\boldsymbol{x}^T[n]\right] \\ &= E\left[(\boldsymbol{I} + (n+l)\boldsymbol{\mathcal{E}})\boldsymbol{A}_0\boldsymbol{s}[n+l]\boldsymbol{s}^T[n]\boldsymbol{A}_0^T(\boldsymbol{I} + n\boldsymbol{\mathcal{E}})^T\right] \\ &= (\boldsymbol{I} + (n+l)\boldsymbol{\mathcal{E}})\boldsymbol{A}_0\boldsymbol{R}_s[l]\boldsymbol{A}_0^T(\boldsymbol{I} + n\boldsymbol{\mathcal{E}})^T \quad (4)\end{aligned}$$

where $\boldsymbol{R}_s[l] = E[\boldsymbol{s}[n+l]\boldsymbol{s}^T[n]]$ are the source signal's diagonal autocorrelation matrices at lag l, which we shall from now on denote, for convenience, as $\boldsymbol{\Lambda}_l$. For additional convenience we shall also denote $\boldsymbol{R}_x[n,l]$ as $\boldsymbol{R}_l[n]$. We then have

$$\begin{aligned}\boldsymbol{R}_l[n] &= \boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T \quad (5)\\ &+ (\boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T + \boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T)n \\ &+ (\boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T)n^2 \\ &+ (\boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T)l.\end{aligned}$$

We now make some approximating assumption, namely that for most values of n in the interval $[-N : N]$, the term which depends linearly on l (the last term in (5)) is negligible with respect to the term that depends linearly on n (the second term in (5)) for all l in which the correlations are nonzero. This is equivalent to assuming that the observation interval's length $(2N + 1)$ is much larger than the correlation lengths of the source signals, or, in our context, that the change of the mixing parameters over intervals of the order of the signals' correlation lengths are negligible relative to the change over the observation length. Thus we can write the following approximate expression for $\boldsymbol{R}_l[n]$:

$$\boldsymbol{R}_l[n] \approx \boldsymbol{R}_l^{(0)} + \boldsymbol{R}_l^{(1)}n + \boldsymbol{R}_l^{(2)}n^2 \quad (6)$$

where

$$\boldsymbol{R}_l^{(0)} \triangleq \boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T, \quad (7a)$$

$$\boldsymbol{R}_l^{(1)} \triangleq \boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T + \boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T = \boldsymbol{\mathcal{E}}\boldsymbol{R}_l^{(0)} + \boldsymbol{R}_l^{(0)}\boldsymbol{\mathcal{E}}^T \quad (7b)$$

and

$$\boldsymbol{R}_l^{(2)} \triangleq \boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T = \boldsymbol{\mathcal{E}}\boldsymbol{R}_l^{(0)}\boldsymbol{\mathcal{E}}^T. \quad (7c)$$

3.1. Estimating $\boldsymbol{R}_l^{(0)}$, $\boldsymbol{R}_l^{(1)}$ and $\boldsymbol{R}_l^{(2)}$

We now wish to estimate the unknown matrices $\boldsymbol{R}_l^{(0)}$, $\boldsymbol{R}_l^{(1)}$ and $\boldsymbol{R}_l^{(2)}$ from the available data $\boldsymbol{x}[-N]$ to $\boldsymbol{x}[N]$. We shall take the following linear least-squares estimation approach: Since the (i,j)-th element of $\boldsymbol{R}_l[n]$ is the expected value of the product $x_i[n+l]x_j[n]$, this product can be regarded as a "noisy" measurement of its expected value (with zero-mean noise). We can therefore arrange these samples in the following manner, applying the model specified by (6) casted as a linear least squares model in the unknown parameters:

$$\boldsymbol{y}(l,i,j) \triangleq \begin{bmatrix} x_i[-N+l]x_j[1] \\ \vdots \\ x_i[1+l]x_j[1] \\ x_i[2+l]x_j[2] \\ \vdots \\ x_i[N+l]x_j[N] \end{bmatrix} \approx \begin{bmatrix} \boldsymbol{R}_l[-N](i,j) \\ \vdots \\ \boldsymbol{R}_l[1](i,j) \\ \boldsymbol{R}_l[2](i,j) \\ \vdots \\ \boldsymbol{R}_l[N](i,j) \end{bmatrix}$$

$$\approx \begin{bmatrix} 1 & -N & N^2 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & N & N^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{R}_l^{(0)}(i,j) \\ \boldsymbol{R}_l^{(1)}(i,j) \\ \boldsymbol{R}_l^{(2)}(i,j) \end{bmatrix} \triangleq \boldsymbol{H}\boldsymbol{\theta}(l,i,j) \tag{8}$$

where $\boldsymbol{\theta}(l,i,j)$ denotes the unknown parameters, namely the (i,j)-th elements of the three matrices $\boldsymbol{R}_l^{(0)}$, $\boldsymbol{R}_l^{(1)}$ and $\boldsymbol{R}_l^{(2)}$, and therefore needs to be estimated for each $1 \le i,j \le M$ and for all desired lags, say $0 \le l \le L$, where L is some maximum lag to be used. It is also assumed implicitly in (8), that $N+L$ samples are actually available, so that end effects are mitigated at the cost of not exploiting all the available samples for the shorter lags. Assuming $L << N$, the associated loss is quite negligible.

The LS estimator is then given by

$$\hat{\boldsymbol{\theta}}(l,i,j) = (\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{y}(l,i,j)$$
$$= \begin{bmatrix} S_0 & 0 & S_2 \\ 0 & S_2 & 0 \\ S_2 & 0 & S_4 \end{bmatrix}^{-1} \cdot \sum_{n=1}^{N} \begin{bmatrix} x_i[n+l]x_j[n] \\ nx_i[n+l]x_j[n] \\ n^2x_i[n+l]x_j[n] \end{bmatrix}, \tag{9}$$

where $S_p \triangleq \sum_{n=-N}^{N} n^p$, so that $S_0 = 2N+1$, $S_2 = \frac{1}{3}N(N+1)(2N+1)$ and $S_4 = \frac{1}{15}N(N+1)(2N+1)(3N^2+3N-1)$.

When the relative rate matrix $\boldsymbol{\mathcal{E}}$ is particularly small and the observation length N is not particularly large, so that we have $N\boldsymbol{\mathcal{E}} << \boldsymbol{I}$, it may become desirable to neglect the quadratic term $(\boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T)n^2$ in (5), namely to assume that $\boldsymbol{R}_l^{(2)}$ is practically zero. In that way, over-parameterization is avoided, and the variance in estimating $\boldsymbol{R}_l^{(0)}$ and $\boldsymbol{R}_l^{(1)}$ can be reduced at the cost of a negligible bias. The parameters vector $\boldsymbol{\theta}(l,i,j)$ would then be reduced to a two-parameters vector, denoted $\bar{\boldsymbol{\theta}}(l,i,j)$, and, eliminating the third column of $\boldsymbol{H}$ accordingly, the LS estimate of $\bar{\boldsymbol{\theta}}(l,i,j)$ would be given by:

$$\hat{\bar{\boldsymbol{\theta}}}(l,i,j) = \begin{bmatrix} S_0 & 0 \\ 0 & S_2 \end{bmatrix}^{-1} \cdot \sum_{n=1}^{N} \begin{bmatrix} x_i[n+l]x_j[n] \\ nx_i[n+l]x_j[n] \end{bmatrix}. \tag{10}$$

Naturally, in this case, due to the symmetric observation interval, the estimate of $\boldsymbol{R}_l^{(0)}$ reduces to the conventional correlation estimate. This guarantees the positive-definiteness of the estimate $\hat{\boldsymbol{R}}_0^{(0)}$, which in turn guarantees the existence of a "whitening matrix" (to be discussed later). Since, in general, this is not guaranteed with non-symmetric observation intervals, we chose to use the symmetric formulation.

Once the parameters vectors, either $\boldsymbol{\theta}(l,i,j)$ or $\bar{\boldsymbol{\theta}}(l,i,j)$ are estimated for each i, j and l, the results can be plugged into the respective estimated matrices, obtaining $\hat{\boldsymbol{R}}_l^{(0)}$, $\hat{\boldsymbol{R}}_l^{(1)}$ and possibly $\hat{\boldsymbol{R}}_l^{(2)}$.

Note that while the true matrices $\boldsymbol{R}_l^{(0)}$, $\boldsymbol{R}_l^{(1)}$ and $\boldsymbol{R}_l^{(2)}$ are all symmetric, their estimated counterparts may be non-symmetric for $l \neq 0$. It is therefore proposed to "symmetrize" the estimates by replacing each $\hat{\boldsymbol{R}}_l^{(p)}$ with $\frac{1}{2}(\hat{\boldsymbol{R}}_l^{(p)} + \hat{\boldsymbol{R}}_l^{(p)T}), p = 0,1,2$.

Using these estimated matrices, we now proceed to obtain estimates of the unknown mixing parameters $\boldsymbol{A}_0$ and $\boldsymbol{\mathcal{E}}$.

3.2. Estimating $\boldsymbol{A}_0$ and $\boldsymbol{\mathcal{E}}$

Recalling the relations (7a,7b,7c) between $\boldsymbol{A}$ and $\boldsymbol{\mathcal{E}}$ and the matrices $\boldsymbol{R}_l^{(0)}$, $\boldsymbol{R}_l^{(1)}$ and $\boldsymbol{R}_l^{(2)}$, several approaches can be taken in extracting estimates of $\boldsymbol{A}$ and $\boldsymbol{\mathcal{E}}$ from the estimates of the correlation matrices. These approaches would differ in the different trade-offs they offer between computational simplicity and accuracy.

The "brute-force" LS approach would be to minimize the following, with respect to all the unknown parameters:

$$\min_{\boldsymbol{A}_0,\boldsymbol{\mathcal{E}},\boldsymbol{\Lambda}_0,\boldsymbol{\Lambda}_1,\ldots\boldsymbol{\Lambda}_L} \Bigg\{ w_0 \cdot \sum_{l=0}^{L} ||\hat{\boldsymbol{R}}_l^{(0)} - \boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T||_F^2$$
$$+ w_1 \cdot \sum_{l=0}^{L} ||\hat{\boldsymbol{R}}_l^{(1)} - \boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T - \boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T||_F^2$$
$$+ w_2 \cdot \sum_{l=0}^{L} ||\hat{\boldsymbol{R}}_l^{(2)} - \boldsymbol{\mathcal{E}}\boldsymbol{A}_0\boldsymbol{\Lambda}_l\boldsymbol{A}_0^T\boldsymbol{\mathcal{E}}^T||_F^2 \Bigg\}, \tag{11}$$

where the arbitrary weights w_0, w_1 and w_2 are essential in order to bring the scales of the three terms to compatible levels, or, in other words, to compensate for the fact that the three terms are of essentially different units. Moreover, an optimally weighted LS criterion can be used if the covariance of the estimation errors in $\hat{\boldsymbol{R}}_l^{(0)}$, $\hat{\boldsymbol{R}}_l^{(1)}$ and $\hat{\boldsymbol{R}}_l^{(2)}$ are known. For example, if the source signals are known to be Gaussian Moving Average processes of some known maximum order q, this covariance can be estimated from the estimated correlations up to lag q, as done, e.g., in [6] and [7]. Such estimates would in turn lead to asymptotically optimal weight matrices. However, in the current context the derivation of such weighting is fairly complex and is beyond the scope of this paper.

The resulting weighted LS problem is a highly involved nonlinear and non-convex minimization problem. While several minimization strategies can be considered, we choose to use one of the most simple (but suboptimal) possible, which in effect addresses the three terms separately. In addition, we shall employ the pre-whitening strategy, transforming the minimization with respect to $\boldsymbol{A}_0$ into a unitary minimization problem, which can be solved, e.g., using Jacobi rotations [8] (as opposed to using direct, non-orthogonal approximate joint diagonalization, as in [9, 10]).

Let $\hat{\boldsymbol{R}}_0^{(0)} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T$ denote the eigenvalue decomposition of $\hat{\boldsymbol{R}}_l^{(0)}$, where $\boldsymbol{U}$ is a unitary matrix and $\boldsymbol{D}$ is diagonal. Defining $\hat{\boldsymbol{W}} \triangleq \boldsymbol{D}^{-\frac{1}{2}}\boldsymbol{U}^T$ as the "estimated spatial whitening matrix", we construct the estimated whitened correlation matrices $\tilde{\hat{\boldsymbol{R}}}_l^{(0)}$ as follows:

$$\tilde{\hat{\boldsymbol{R}}}_l^{(0)} = \hat{\boldsymbol{W}}\hat{\boldsymbol{R}}_l^{(0)}\hat{\boldsymbol{W}}^T \quad , \quad l = 0, 1, ...L \qquad (12)$$

(leading, of course, to $\tilde{\hat{\boldsymbol{R}}}_0^{(0)} = \boldsymbol{I}$, which accounts for the term "spatial whitening"). In order to estimate $\boldsymbol{A}_0$ we now seek to minimize the following term with respect to a unitary matrix $\boldsymbol{V}$:

$$\min_{\boldsymbol{V},\boldsymbol{\Lambda}_0,\boldsymbol{\Lambda}_1,...\boldsymbol{\Lambda}_L} \left\{ \sum_{l=0}^{L} ||\tilde{\hat{\boldsymbol{R}}}_l^{(0)} - \boldsymbol{V}\boldsymbol{\Lambda}_l\boldsymbol{V}^T||_F^2 \right\}, \qquad (13)$$

which, as mentioned above, can be accomplished using sequential Jacobi rotations [8]. Denoting the minimizers as $\hat{\boldsymbol{V}}$ (and $\hat{\boldsymbol{\Lambda}}_0, \hat{\boldsymbol{\Lambda}}_1, ...\hat{\boldsymbol{\Lambda}}_L$), the estimate $\hat{\boldsymbol{A}}_0$ of $\boldsymbol{A}_0$ is given by

$$\hat{\boldsymbol{A}}_0 = \hat{\boldsymbol{W}}^{-1}\hat{\boldsymbol{V}}. \qquad (14)$$

An estimate of $\boldsymbol{\mathcal{E}}$ can be obtained from the second term of (11). Observing the expression for $\boldsymbol{R}_l^{(1)}$ in (7b), two options of the minimization are evident, both can be stated as

$$\min_{\boldsymbol{\mathcal{E}}} \left\{ \sum_{l=0}^{L} ||\hat{\boldsymbol{R}}_l^{(1)} - \boldsymbol{\mathcal{E}}\boldsymbol{Q}_l - \boldsymbol{Q}_l^T\boldsymbol{\mathcal{E}}^T||_F^2 \right\} \qquad (15)$$

where the symmetric matrices $\boldsymbol{Q}$ can either be given by

$$\boldsymbol{Q}_l = \hat{\boldsymbol{A}}_0\hat{\boldsymbol{\Lambda}}_l\hat{\boldsymbol{A}}_0^T \quad , \quad l = 0, 1, ...L \qquad (16a)$$

or by

$$\boldsymbol{Q}_l = \hat{\boldsymbol{R}}_l^{(0)} \quad , \quad l = 0, 1, ...L. \qquad (16b)$$

Since the estimated matrices cannot (almost surely) be exactly jointly diagonalized (for $L > 1$), the two expressions for $\boldsymbol{Q}_l$ are (almost surely) different. However, it has been observed empirically (in our preliminary tests), that the resulting performance is nearly insensitive to the choice of either (16a) or (16b) for $\boldsymbol{Q}_l$. Note only, that in a parallel computation setup, (16b) is somewhat preferable, since it does not require prior estimation of $\boldsymbol{A}_0$ and $\boldsymbol{\Lambda}_0, \boldsymbol{\Lambda}_1, ...\boldsymbol{\Lambda}_L$, and therefore allows estimation of $\boldsymbol{\mathcal{E}}$ in parallel.

Fortunately, the minimization criterion (15) is quadratic in $\boldsymbol{\mathcal{E}}$, and can therefore be minimized as follows. Denoting by $\boldsymbol{\varepsilon}_n$ the rows of $\boldsymbol{\mathcal{E}}$ such that $\boldsymbol{\mathcal{E}}^T = [\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2...\boldsymbol{\varepsilon}_M]$ and by $\boldsymbol{q}_n^l$ the columns of $\boldsymbol{Q}_l$ such that $\boldsymbol{Q}_l = [\boldsymbol{q}_1^l\boldsymbol{q}_2^l...\boldsymbol{q}_M^l]$, we have

$$\left(\boldsymbol{\mathcal{E}}\boldsymbol{Q}_l + \boldsymbol{Q}_l^T\boldsymbol{\mathcal{E}}^T\right)_{(n,m)} = \boldsymbol{\varepsilon}_n^T\boldsymbol{q}_m^l + \boldsymbol{q}_n^{lT}\boldsymbol{\varepsilon}_m. \qquad (17)$$

Denoting by vec($\bullet$) the operation of concatenating all the columns of the argument $M \times M$ matrix into one $M^2 \times 1$ column, we then have

$$\mathrm{vec}(\boldsymbol{\mathcal{E}}\boldsymbol{Q}_l) = \underbrace{\begin{bmatrix} \boldsymbol{q}_1^{lT} & & & \\ & \boldsymbol{q}_1^{lT} & & \\ & & \ddots & \\ & & & \boldsymbol{q}_1^{lT} \\ \boldsymbol{q}_2^{lT} & & & \\ & \boldsymbol{q}_2^{lT} & & \\ & & \ddots & \\ & & & \boldsymbol{q}_2^{lT} \\ & \vdots & & \\ & \vdots & & \\ & \vdots & & \\ \boldsymbol{q}_M^{lT} & & & \\ & \boldsymbol{q}_M^{lT} & & \\ & & \ddots & \\ & & & \boldsymbol{q}_M^{lT} \end{bmatrix}}_{\triangleq \boldsymbol{H}_1^l} \cdot \underbrace{\begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix}}_{\triangleq \boldsymbol{\varepsilon}} \qquad (18)$$

and, similarly,

$$\mathrm{vec}(\boldsymbol{Q}_l^T\boldsymbol{\mathcal{E}}^T) = \underbrace{\begin{bmatrix} \boldsymbol{q}_1^{lT} & & & \\ \boldsymbol{q}_2^{lT} & & & \\ \vdots & & & \\ \boldsymbol{q}_M^{lT} & & & \\ & \boldsymbol{q}_1^{lT} & & \\ & \boldsymbol{q}_2^{lT} & & \\ & \vdots & & \\ & \boldsymbol{q}_M^{lT} & & \\ & & \vdots & \\ & & \vdots & \\ & & & \boldsymbol{q}_1^{lT} \\ & & & \boldsymbol{q}_2^{lT} \\ & & & \vdots \\ & & & \boldsymbol{q}_M^{lT} \end{bmatrix}}_{\triangleq \boldsymbol{H}_2^l} \cdot \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{bmatrix} \tag{19}$$

Consequently, the linear LS minimization problem (15) can be restated as

$$\min_{\boldsymbol{\varepsilon}} \left\{ \sum_{l=0}^{L} || \,\mathrm{vec}(\hat{\boldsymbol{R}}_l^{(1)}) - [\boldsymbol{H}_1^l + \boldsymbol{H}_2^l]\boldsymbol{\varepsilon}||^2 \right\} \tag{20}$$

whose solution is

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}} \;\; &= \;\; \left[\sum_{l=0}^{L} \left[\boldsymbol{H}_1^l + \boldsymbol{H}_2^l\right]^T \left[\boldsymbol{H}_1^l + \boldsymbol{H}_2^l\right] \right]^{-1} \cdot \\ &\cdot \;\; \sum_{l=0}^{L} \left[\boldsymbol{H}_1^l + \boldsymbol{H}_2^l\right]^T \cdot \mathrm{vec}(\hat{\boldsymbol{R}}_1^{(1)}). \end{aligned} \tag{21}$$

Once the minimizing $\hat{\boldsymbol{\varepsilon}} = \mathrm{vec}(\hat{\boldsymbol{\mathcal{E}}}^T)$ is computed, $\hat{\boldsymbol{\mathcal{E}}}$ can be easily constructed by inverting the vec($\bullet$) operation. In our sub-optimal minimization scheme we deliberately ignore (or neglect) the last term of (11), especially since in practical situations the estimates of $\boldsymbol{R}_l^{(2)}$ are relatively inaccurate when the variation rate is slow.

The estimates $\hat{\boldsymbol{A}}_0$ and $\hat{\boldsymbol{\mathcal{E}}}$ can now be used for demixing the data by respectively inverting the time-dependent mixing:

$$\hat{\boldsymbol{s}}[n] = \hat{\boldsymbol{A}}_0^{-1}(\boldsymbol{I} + n\hat{\boldsymbol{\mathcal{E}}})^{-1}\boldsymbol{x}[n] \quad , \quad n = 1, 2, ...N \tag{22}$$

The signals $\hat{\boldsymbol{s}}[n]$ are the estimated source signals, up to the inherent scale and permutation ambiguities induces in any BSS problem. Note, however, that these ambiguities are absorbed in the estimated $\boldsymbol{A}_0$ only. There are no such ambiguities in the estimate of $\boldsymbol{\mathcal{E}}$.

4. SIMULATION RESULTS

We demonstrate the performance of the separation algorithm using a simulated example with $M = 2$ source signals, both zero-mean Gaussian MA processes. They were created by passing independent white Gaussian unit-variance noise sequences through the systems $H_1(z) = 1 + 2z^{-1} - 0.5z^{-2} - z^{-3} + z^{-4}$ (for $s_1[n]$) and $H_2(z) = 1 - z^{-1} + 3z^{-2} + 2z^{-3}$ (for $s_2[n]$).

The true mixing parameters were $\boldsymbol{A}_0 = \left[\begin{smallmatrix} 3 & 1 \\ -2 & 4 \end{smallmatrix}\right]$ and $\boldsymbol{\mathcal{E}} = \left[\begin{smallmatrix} -1 & 0.5 \\ -2 & 1 \end{smallmatrix}\right] \cdot 10^{-4}$. We demonstrate the performance using observation lengths with $N = 100$ to $N = 10000$. For the shorter observation lengths the mixture is nearly constant; for the longest observation length the mixture changes considerably from $\boldsymbol{A}(-N) = \left[\begin{smallmatrix} 7 & 0 \\ 6 & 2 \end{smallmatrix}\right]$ to $\boldsymbol{A}(N) = \left[\begin{smallmatrix} -1 & 2 \\ -10 & 6 \end{smallmatrix}\right]$. All algorithms were applied with five correlation lags up to $L = 4$.

In order to measure the performance of the algorithm, our measure of ISR is defined over the entire observation interval: Defining as $\boldsymbol{T}[n]$ the total (time-dependent) mixing-demixing effect, we have

$$\boldsymbol{T}[n] = \hat{\boldsymbol{A}}_0^{-1}(\boldsymbol{I}+n\hat{\boldsymbol{\mathcal{E}}})^{-1}(\boldsymbol{I}+n\boldsymbol{\mathcal{E}})\boldsymbol{A}_0, \; n = -N, -N+1, ...N \tag{23}$$

Then, in order to evaluate the total leakage of energy between reconstructed sources, we compute

$$\boldsymbol{T} = \sum_{n=-N}^{N} \boldsymbol{T}[n] \odot \boldsymbol{T}[n] \tag{24}$$

where $\odot$ denotes Hadamard's (element-wise) product (implying element-wise squaring in this case). Once $\boldsymbol{T}$ is calculated, the permutation ambiguity is resolved by selecting the permutation which maximizes $\boldsymbol{T}$'s trace, and then the average ISR is the average (over the two rows) of the ratios of off-diagonal to diagonal energy in each row.

For reference, we present results for the conventional SOBI algorithm, compared to the two versions of our "Time-Variable SOBI" (TV-SOBI) algorithm: TV-SOBI$_1$ using the "linear estimation model" (10) (estimating just $\hat{\boldsymbol{R}}_l^{(0)}$ and $\hat{\boldsymbol{R}}_l^{(1)}$) and TV-SOBI$_2$ using the "quadratic estimation model" (9) (estimating $\hat{\boldsymbol{R}}_l^{(2)}$ as well).

It is evident, that with short observation lengths, when the actual mixing variation is negligible, SOBI yields better performance, because the introduction of additional parameters into TV-SOBI unnecessarily increases the estimation variance. However, as N increases, the assumption of constant mixing departs from reality, and the performance of SOBI begins to deteriorate rapidly. This is where the compensating effect of the TV-SOBI algorithm(s) evolves. It is seen, in this

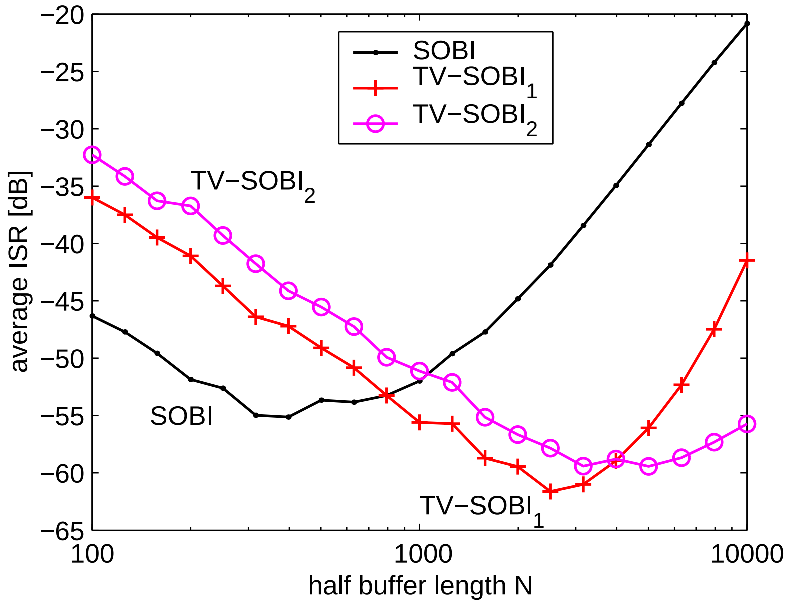

Figure 1: Performance in terms of average ISR [dB] for SOBI, TV-SOBI$_1$ (linear estimation model (10)) and TV-SOBI$_2$ (quadratic estimation model (9)) vs. N (half the observation length). Each point represents the average of 400 trials. All algorithms used the same data.

example, that the use of TV-SOBI$_1$ can enable prolonged observation intervals, offering improvement of over $20dB$ over conventional SOBI. When N is further increased, a similar trade-off is observed between TV-SOBI$_1$ and TV-SOBI$_2$, where the addition of quadratic parameters in TV-SOBI$_2$ increases the estimation variance at first, but eventually "pays off" by outperforming TV-SOBI$_1$ at large values of N.

All algorithms exhibit certain degradation in performance as N increases to extreme values, because with large N, small errors in the estimate of $\boldsymbol{\mathcal{E}}$ are strongly amplified towards the edges of the observation interval, and the resulting poor ISR at the edges overrides the good average at the middle.

5. CONCLUSION

We presented an expansion of the SOBI algorithm, named TV-SOBI, which addresses time-varying instantaneous mixtures using a linear model for the time-variation. We have shown that this expansion enables the use of longer observation interval, resulting in improved performance, when the mixtures are indeed time-varying. Various approaches can be taken in the different stages of the estimation procedure, and we have chosen to present in here just the most simple approach, capturing the essence of the model. Further improvement can be attained using more elaborate approaches, which are the subject of on-going research.

6. REFERENCES

[1] Andrzej Cichocki and Shun-ichi Amari, *Adaptive Blind Signal and Image Processing*, Wiley, 2002.

[2] Adel Belouchrani, Karim Abed-Meraim, Jean-François Cardoso, and Eric Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[3] Shun-ichi. Amari and Scott Douglas, "Why natural gradient," *Proc. ICASSP'98*, vol. 2, pp. 1213–1216, 1998.

[4] A. Cichocki, R. Unbehauen, L. Moszczynski, and E. Rummert, "A new on-line adaptive learning algorithm for blind separation of sources," *Proc. of ISANN'94*, pp. 406–411, 1994.

[5] Jean-François Cardoso, "Learning in manifolds: the case of source separation," *Proc. of SSAP'98*, 1998.

[6] Arie Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Processing Letters*, vol. 7, no. 7, pp. 197–200, 2000.

[7] Arie Yeredor and Eran Doron, "Using farther correlations to further improve the optimally-weighted sobi algorithm," *Proc. EUSIPCO'2002*, 2002.

[8] Jean-François Cardoso and Antoine Souloumiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, no. 1, pp. 161–164, 1996.

[9] Arie 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, 2002.

[10] Alle-Jan van der Veen, "Joint diagonalization via subspace fitting techniques," *Proceedings ICASSP 2001*, vol. 5, 2001, Salt Lake City.