A HYBRID METHOD FOR BLIND SIGNAL DE-NOISING VIA INDEPENDENT COMPONENT ANALYSIS

Alessandra Budillon, Francesco Palmieri and Rosario Varriale

Dip. di Ing. Elettronica e delle Telecomunicazioni
Universita' di Napoli Federico II, Italy
e-mail: {alebudil,frapalmi}@unina.it,rosvarri@tin.it

ABSTRACT

This paper presents a signal de-noising method based on the extraction of two independent components. Two FIR filters are adapted blindly and forced to produce outputs which are as independent as possible through time. Various cost functions are considered ranging from linear correlation to non linear correlation and joint entropy. The typical presence of a large number of local minima in the various cost functions has suggested a hybrid approach that trades computational complexity with solution accuracy. Two experiments on speech corrupted by independent noise and the separation of two AR processes are presented.

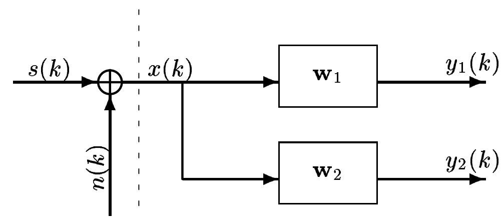

Figure 1: The de-noising scheme

1. INTRODUCTION

This paper explores a blind adaptive signal de-noising method based on Independent Component Analysis (ICA). The noise reduction problem using adaptive filtering is a classic one [5, 8] and has been approached typically through blind training of a predictor. Such a method is often referred to as Adaptive Line Enhancer (ALE) [8].

However, the success of Independent Component Analysis in separating instantaneous [3, 1, 2] and convolutive mixtures [7, 4] of independent sources suggests that such techniques may be applied to the classical problem of noise removal when not much knowledge about signal and noise is available. It is well-known that Wiener filtering that requires exact knowledge of cross correlation between observations and "desired" output, may have limited application when "clean" data is not available for training and when signals exhibit non stationary behaviour.

The idea we explore in this paper, is based on a de-noising system made up of two FIR filters that, operating on the same input are trained simultaneously to produce outputs which are as independent as possible through time and have unit variance.

In particular, if the input is the superposition of independent signal and noise, the problem becomes one of blind separation of an overcomplete mixture of two signals from only one sensor with a convolutive de-mixer [7, 4].

In this paper we report some preliminary experiments in which we train the two filters with various cost functions ranging from pure linear decorrelation to joint entropy pointing out to the major problems that arises in this scenario. In fact, just as in most overcomplete mixture cases, the presence of a large number of local minima requires typically sufficient annealing and solution selection.

A question that is also crucial to this problem is the sufficiency/insufficiency of decorrelation.

In this paper we consider different possible cost functions arriving to an hybrid algorithm: search for decorrelation and create a population of solutions out of which the configuration with the minimum higher-order independence is chosen.

Two experiments are reported on the blind restoration of a speech signal corrupted by band-pass noise and the separation of two band-pass AR processes. We compare the results of different algorithms derived by three cost functions with the ones obtained using the Wiener filter.

2. DE-NOISING ALGORITHM

2.1. Problem statement

Let $s(k)$ be the signal of interest and $x(k)$ the observed signal, given by $x(k) = s(k) + n(k)$, where $n(k)$ is the noise. Assume that $s(k)$ and $n(k)$ are zero mean, independent and stationary.

Suppose now that our signal de-noising is based on two FIR filters, $\mathbf{w}_1$ and $\mathbf{w}_2$, where $\mathbf{w}_i = [w_i(0), \ldots, w_i(q)]^T$ $i = 1, 2$ are the coefficient vectors of length $q+1$. Let $\mathbf{x}(k) = [x(k), \ldots, x(k-q)]^T$ the vector containing $q+1$ samples of the observed signal $x(k)$, so that the filters output can be written as $y_i(k) = \mathbf{w}_i^T \mathbf{x}(k)$, $i = 1, 2$.

The question is: can we find $\mathbf{w}_1$ and $\mathbf{w}_2$ to get (blindly) at the two outputs estimates of $s(k)$ and $n(k)$ simply by forcing independence? Obviously whether the result of this two-element filter bank is the decomposition of $x(k)$ into (typically distorted) versions of $s(k)$ and $n(k)$ will strongly depend on the structure of the two signals and their power relationship. In fact, if for example $n(k)$ has small power with respect to that of $s(k)$, the problem is basically a search for a decomposition of $s(k)$ only. Also if $x(k)$ is a more complicated superpositions of factors the problem is one of finding the most "separated" decomposition.

Therefore, to avoid possible ambiguities in the discussion that follows, we will focus on algorithms and problems in searching for a two-element Independent Component Analysis (ICA) filter bank, even though the signal-plus-noise scenario will be kept in mind as a key reference also in our simulations.

2.2. Forcing linear decorrelation

A first separation criterion may be to force the impulse response of the two FIR filters to be such that $r_{y_1 y_2}(l) = 0 \;\; \forall l$. Various algorithms have been proposed in the literature for such a task, both on-line and off-line [4]. The cost function is defined as the sum of the squares of the cross-correlations between y_1 and y_2 in a given time-lag interval (l_1, l_2) as:

$$\sum_{l=l_1}^{l_2} |E[y_1(k)y_2(k+l)]|^2 = \sum_{l=l_1}^{l_2} r_{y_1y_2}^2(l). \qquad (1)$$

The cross-correlation sequence can be written as

$$r_{y_1y_2}(l) = \mathbf{w}_1^T E[\mathbf{x}(k)\mathbf{x}^T(k+l)]\mathbf{w}_2 = \mathbf{w}_1^T \mathbf{R}_x(l)\mathbf{w}_2, \qquad (2)$$

where $\mathbf{R}_x(l)$ is a $(q+1) \times (q+1)$ Toeplitz correlation matrix of $\mathbf{x}(k)$:

$$\mathbf{R}_x(l) = \begin{pmatrix} r_x(l) & r_x(l-1) & \cdots & r_x(l-q) \\ r_x(l+1) & r_x(l+1) & \cdots & r_x(l-q+1) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(l+q) & r_x(l+q-1) & \cdots & r_x(l) \end{pmatrix} \qquad (3)$$

To avoid the trivial solution to the problem, i.e. both $\mathbf{w}_1$ and $\mathbf{w}_2$ equal to zero, some constraints are necessary. Using a constant output power condition, the cost function becomes:

$$C_1 = \sum_{l=l_1}^{l_2} (\mathbf{w}_1^T \mathbf{R}_x(l)\mathbf{w}_2)^2 + \sum_{i=1}^{2} \lambda_i (k - \mathbf{w}_i^T \mathbf{R}_x(0)\mathbf{w}_i). \qquad (4)$$

An effective off-line algorithm for solving this problem has been proposed in [4] for convolutive mixtures. The simple steps are outlined in the Appendix with specific reference to this two-filter scenario.

Unfortunately, is the nature of the cost function that creates problems, due to the presence of a large number of local minima, some of which correspond to trivial solutions. This is a common problem when we are dealing with an overcomplete mixture and proper restart and selection strategies must be devised for global convergence. Typically the iterative algorithm, that is basically a coordinate descent algorithm, converges to different solutions depending on the starting initial conditions. Furthermore: solutions that globally give minimum linear correlation are not necessarily good for separation. The simulations reported in the following will clearly demonstrate this effect.

2.3. Non linear decorrelation algorithms

The decorrelation criterion can be replaced by a stronger one, based on non linear versions of the filter outputs. The most general form of the cost function is:

$$\begin{aligned} C_2 \;\; = \;\; & \sum_{l=l1}^{l2} |E[f_1(y_1(k))f_2(y_2(k+l))]|^2 + \\ & + \lambda_1(1 - \mathbf{w}_1^T \mathbf{R}_x(0)\mathbf{w}_1)^2 + \\ & + \lambda_2(1 - \mathbf{w}_2^T \mathbf{R}_x(0)\mathbf{w}_2)^2. \end{aligned} \qquad (5)$$

where f_1 and f_2 are non linear functions, so that higher order statistics are involved. The second and the third terms are the power constraints. The coefficients λ_1 and λ_2, that typically must be chosen through cross-validation, have been set to $\lambda_1 = \lambda_2 = \lambda = l_2 - l_1$. We consider symmetrical source distributions and odd non linear functions, so that $E[f_i(y_i(k))] = 0$.

Unfortunately, this cost function is computationally more complex than the previous one and a gradient descend algorithm should be used since no closed form can be found for the filter coefficients. The gradients are easily computed:

$$\begin{aligned} \nabla_{\mathbf{w}_1} C_2 \;\; = \;\; & \sum_{l=l1}^{l2} 2E[f_1(y_1(k))f_2(y_2(k+l))] \\ & \cdot E[f_2(y_2(k+l))f_1'(y_1(k))\mathbf{x}(k)] + \\ & -4\lambda_1(1 - \mathbf{w}_1^T \mathbf{R}_x(0)\mathbf{w}_1)\mathbf{R}_x(0)\mathbf{w}_1; \end{aligned} \qquad (6)$$

$$\begin{aligned} \nabla_{\mathbf{w}_2} C_2 \;\; = \;\; & \sum_{l=l1}^{l2} 2E[f_1(y_1(k))f_2(y_2(k+l))] \\ & \cdot E[f_1(y_1(k))f_2'(y_2(k+l))\mathbf{x}(k+l)] + \\ & -4\lambda_2(1 - \mathbf{w}_1^T \mathbf{R}_x(0)\mathbf{w}_1)\mathbf{R}_x(0)\mathbf{w}_1. \end{aligned} \qquad (7)$$

A gradient search typically initializes the two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$ and at each iteration, minimizes the cost function w.r.t. one filter at each time.

Unfortunately, what seems to be the most obvious extension of the linear correlation method, experimentally presents two main difficulties:

a) C_2 in this overcomplete case appears generally even more critical than C_1 with respect to the presence of local minima and being more sensitive to the initial conditions.

b) At each iteration outputs statistics have to be evaluated.

A priori information about the signal spectra, so that the initial filter coefficients can be chosen appropriately, may lead to better results. Also the choice of the non linear functions, that is usually non critical (for instance in [6], it has been suggested to use x^3 and $\mathrm{atan}(x)$), can be done more accurately in analogy to the maximum entropy solution for the square case [3, 1]. For example using as a cost function the norm of the matrix:

$$H(\mathbf{y}) = E[\phi(\mathbf{y})\mathbf{y}^T - I], \tag{8}$$

with $\mathbf{y} = [y_1, y_2]^T$, and

$$\phi = [-\frac{d\log(p(y_1))}{dy_1}, -\frac{d\log(p(y_2))}{dy_2}]^T, \tag{9}$$

where $p_1(y_1)$ and $p_2(y_2)$ are the two desired pdfs. By imposing that the sum of the squared off diagonal entries of H, extended to different time lags, are to be minimized, the constrained cost function is written as:

$$\begin{aligned} C_3 \ = \ & \sum_{l=l1}^{l2} |E[f_1(y_1(k))y_2(k+l)]|^2 + \\ & + \sum_{l=l1}^{l2} |E[y_1(k)f_2(y_2(k+l))]|^2 + \\ & + \lambda_1(1 - \mathbf{w}_1^T\mathbf{R}_x(0)\mathbf{w}_1)^2 + \lambda_2(1 - \mathbf{w}_2^T\mathbf{R}_x(0)\mathbf{w}_2)^2. \end{aligned} \tag{10}$$

The gradients are easily derived in analogy to the above (6) as well constants λ_1 and λ_2.

Unfortunately, problems a) and b) outlined above, even if we have exact knowledge of the desired marginal distributions, are not significantly attenuated.

2.4. The proposed hybrid methods

It is clear from the above discussion that the solution to this problem must include a strategy for finding global minima among possibly many local minima. A method must be used to generate a sufficiently large number of initial conditions. Also a nonlinear independence criterion must be used to force higher-order independence. In consideration that:

a) the decorrelating algorithm, leads to a variety of local minima according to the initial conditions;

b) the decorrelating algorithm is very simple in its implementation since it is based on second order statistics that do not need to be computed at every iteration;

c) stronger independence may be necessary;

d) the non linear cost functions are also affected by local minima problems and are computationally more intensive in their moment estimation;

we propose the following strategy:

1. Search for a set of vectors $(\mathbf{w}_1, \mathbf{w}_2)$ minima of the decorrelation cost function C_1 by running many times the decorrelation algorithm starting at each time from different initial filter coefficients (eventually using an annealing strategy to generate the different initial conditions). Then store the different solutions.
2. Use such solutions as the initial conditions for a gradient algorithm from cost function C_2 (or C_3).
3. Choose as the final solution for the filters the one with the smallest value of C_2 (or C_3).

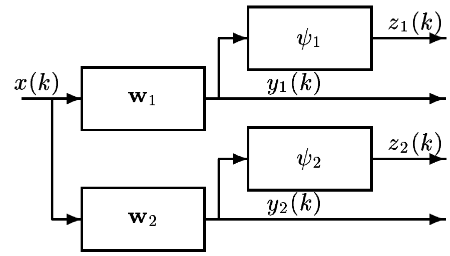

Figure 2: The two de-noising FIR filters scheme

Such a strategy works on the basic assumption that the an ICA solution should not bee too far from a decorrelating solution. This is reasonable if we think that if we had a really independent pair of output sequences, they would be also decorrelated. Moreover we have verified experimentally that the gradient algorithm based on the nonlinear criterion, typically converges very quickly to a solution in the neighborhood of the starting point. This has the consequence of keeping limited the computational complexity of the whole search. Furthermore, if we eliminate completely step 3 we may have may very limited effect on the accuracy of the solution. The few cases in which we obtained good results encouraged us to modify the above method by substituting C_3 (or C_4) with a more appropriate cost function.

Assuming that we have full knowledge of the target marginal densities for the outputs (even though this knowledge is not critical), a more appropriate measure of independence is given by the entropy (or equivalently the likelihood) [3, 1] computed at various time lags on $z_1(k)$ and $z_2(k)$. Figure 2 shows the ICA system with the non linear functions ψ_1 and ψ_2 chosen to be the unit variance cumulative distribution functions of the two target densities [1].

Indeed we adopt the following cost function:

$$-\sum_{l=l1}^{l2} h(z_1(k), z_2(k+l)), \tag{11}$$

where h is the joint differential entropy. More specifically $h(z_1(k), z_2(k+l))$ is evaluated empirically and the cost function is:

$$\begin{aligned} C_4 = \ & \textstyle\sum_{i=1}^{N}\sum_{j=1}^{N} P(z_1^i, z_2^j)\log P(z_1^i, z_2^j) \\ & -\log(\Delta_{z_1}\Delta_{z_2}), \end{aligned} \tag{12}$$

where N^2 are the bins of area $(\Delta_{z_1}\Delta_{z_2})$ and $P(z_1^i, z_2^j)$ the estimated probabilities.

Unfortunately, a search strategy based on C_4 would be hard because the gradient is not easily obtained. However C_4 can be used as a selection criterion. Therefore steps 2. and 3. are replaced by the step

2. Evaluate C_4 in the different solutions $(\mathbf{w}_1, \mathbf{w}_2)$ corresponding to the minima of C_2 and choose as the solution the one with the smallest value of C_4.

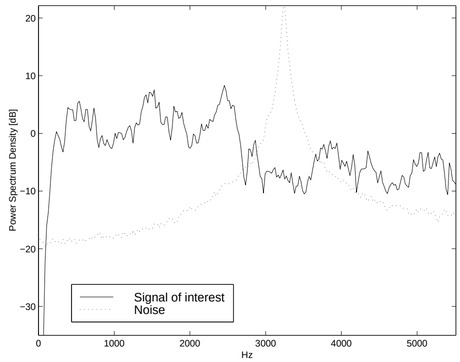

Figure 3: Signal of interest and noise spectra (experiment 1)

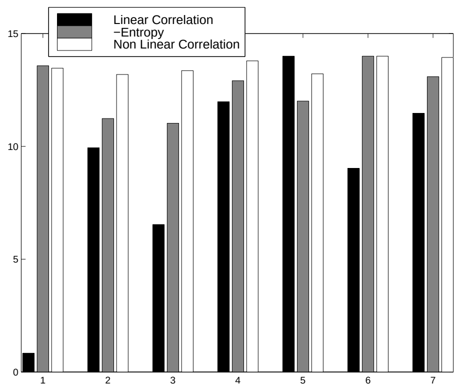

Figure 4: The values for cost functions (C_1, C_2, C_4) for seven stationary points of C_1 (logarithmic scale)

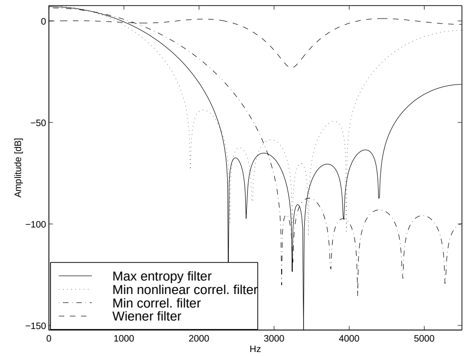

Figure 5: Frequency responses of the de-noising signal filters $\mathbf{w}_1$ correspondent to the best solutions for the three cost functions (C_1, C_2, C_4) and to the Wiener filter

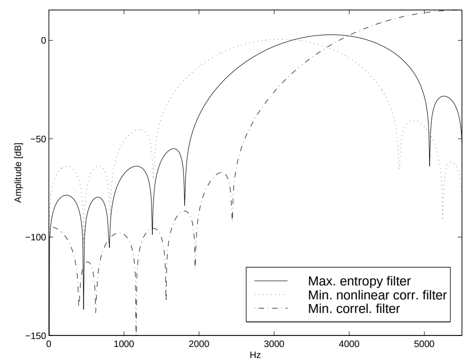

Figure 6: Frequency responses of the noise extracting filter $\mathbf{w}_2$ correspondent to the best solutions for the three cost functions (C_1, C_2, C_4)

3. SIMULATIONS RESULTS

Many simulations have been performed to test the different cost functions. We report here only two typical results.

Experiment 1:

A male speech signal $s(k)$ of length 12000, sampled at 11025kHz is corrupted by AR bandpass noise obtained as filtered uniform noise. The AR filter has two complex conjugate poles with equal module of 0.98 centered at 3000Hz. The signal power is 1 and the SNR is 1dB. Fig. 3 shows the noise power spectral density and an average power spectrum for the speech signal. The length of filters $\mathbf{w}_1$ and $\mathbf{w}_2$ is $q = 13$. Fig. 4 shows the normalized values of the cost functions C_1, C_2, e C_4, in some of the most relevant local minima of C_2 (seven). In C_2 the non linear functions $f_i(y_i)$ have been chosen equal to $\tanh(20y_i)$, in C_4, ψ_1 and ψ_2 are the sources CDFs, assumed known. Actually the CDFs were estimated off-line and the entropy computed using the exact empirical CDFs. Note how the different cost function would pick totally different solutions! This means that choosing the best decorrelation may be strongly misleading for independence. The best solution for C_4 is the third point; the optimum w.r.t C_2 is the second point and the best for C_1 is the first point.

For comparison we have also computed the Wiener filter (where obviously we have assumed for that solution to have access to the exact clean signal $s(k)$). In Table 1 the results in terms of square root MSE and signal-to-noise ratio (S/N) are reported. S/N is defined as the ratio between the power of the signal of interest and the noise power at the same output (the ambiguity about which output to associate to the useful signal has been solved by off-line selection). Note that S/N is a more appropriate measure of separation as it is not affected by possible signal distortion.

	C_4	C_2	C_1	C_{Wiener}
σ	0.7778	0.7945	0.8248	0.2834
S/N [dB]	25.02	24.83	24.88	16.01

Table 1: Square root MSE and separation S/N for the different cost functions (C_4, C_2, C_1)

Note that the solution with the highest entropy is the best and that the blind solution is only a few dB below the Wiener solution. The error is clearly due to the uncontrollable distortion introduced by the blind filters, which search for independent outputs only. In fact, in terms of S/N, which measure only the level of separation of the two input signal, the max entropy solution results even better

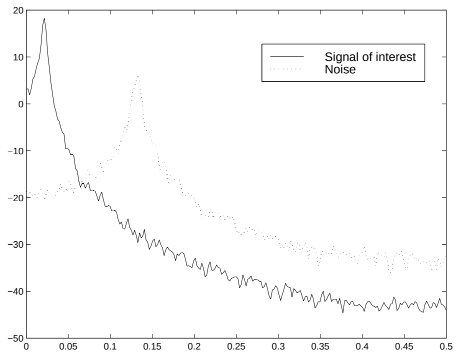

Figure 7: Spectra of the two AR signal (experiment 2)

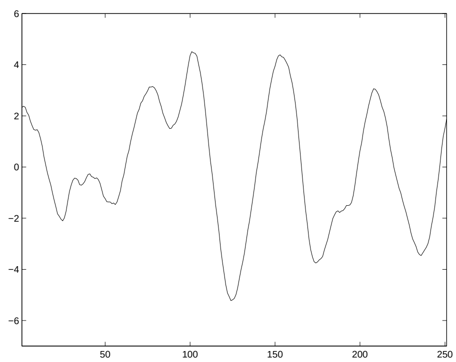

Figure 8: Signal of interest

than Wiener solution.

In Figs. 5 and 6 are reported the filter frequency responses of respectively the de-noising signal filter and the noise extracting filter, correspondent to the best solution for the three cost functions. In the figure the Wiener filter frequency response is also reported for comparison.

Note how the blind solutions tend to create a stronger null than the Wiener filter, and that the solution with minimum entropy is always better that the decorrelating solution and C_2.

Experiment 2:

Consider two AR(2) signals with poles characterized by $\rho_1 = 0.98$, $\omega_1 = 0.827$ and $\rho_2 = 0.99$, $\omega_2 = 0.132$ and with SNR=10dB. The processes are obtained from uniform i.i.d sequences. The first one is the useful signal and the second one is the noise. In this case the FIR length is 5. Fig. 7 shows the spectra of the two AR signals. Figs. 8, 9, 10 show a segment of the useful signal, the observed signal, and the de-noised signal respectively. We found in this case that the function C_1 has only one minimum. The square root MSE corresponding to this solution is 0.2085, while the value for the Wiener filter is 0.2061. Not much different is the filter bank in this case for best decorrelation and best independence. This is to be attributed to the quasi-gaussian nature of the two sequences.

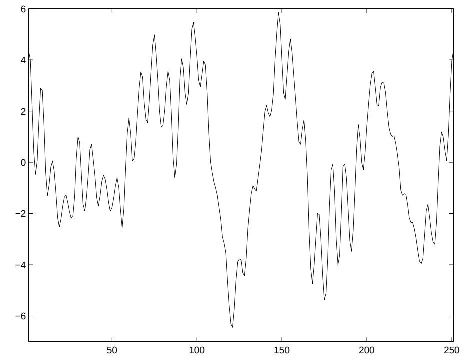

Figure 9: Signal corrupted by noise

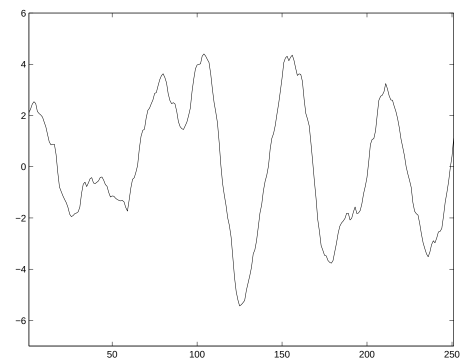

Figure 10: Filter output

4. CONCLUSIONS

In this paper we have investigated the possibility of training two FIR filters on the same input as independent component analyzers simply by forcing independence at their outputs. The problem is complicated by the presence of many local minima being an overcomplete problem. Selection and multistart method must be used to find good solutions. The reference application scenario is that of de-noising a signal corrupted by independent noise. A hybrid strategy has been proposed for finding global minima among possibly many local minima. It exploits the simplicity of the decorrelating algorithm in finding set of local minima that can be used as starting points for a nonlinear correlation algorithm, or as candidated for a more general information criterion to find the global solution. The simulation experiments sometimes reveal marked differences in solutions coming from the different cost functions and some other times best decorrelating solutions which are almost indistinguishable from the ones coming from the general information criterion. This results suggests that a hybrid method may be the best choice most of the time in de-noising problems.

A. APPENDIX

In the following we report the derivation of the decorrelating algorithm proposed in [4] for the square case applied to the two FIR filter bank. The typical coordinate descent approach to minimize C_1 with respect to $\mathbf{w}_1$ and $\mathbf{w}_2$ is: fix one filter and evaluate the other one to minimize C_1, then fix the other and find the first one; then iterate.

Denote with $\mathbf{r}_x(l)$, the vector obtained from the first row and the first column of the matrix $\mathbf{R}_x(l)$: $\mathbf{r}_x(l) = [r_x(l-q), \ldots, r_x(l+q)]^T$. It can be shown that

$$\mathbf{R}_x(l)\mathbf{w}_2 = \mathbf{A}_2(\mathbf{w}_2)\mathbf{r}_x(l), \tag{A.1}$$

where $\mathbf{A}_2(\mathbf{w}_2)$ is a function of $\mathbf{w}_2$:

$$\mathbf{A}_2(\mathbf{w}_2) =$$

$$\begin{pmatrix} w_2(q) & w_2(q-1) & \cdots & w_2(0) & 0 & \cdots & 0 \\ 0 & w_2(q) & \cdots & w_2(1) & w_2(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_2(q) & w_2(q-1) & \cdots & w_2(0) \end{pmatrix} \tag{A.2}$$

Equation (2) becomes

$$r_{y_1y_2}(l) = \mathbf{w}_1^T\mathbf{A}_2(\mathbf{w}_2)\mathbf{r}_x(l). \tag{A.3}$$

From (A.3) we can derive:

$$\mathbf{r}_{y_1y_2} = \mathbf{w}_1^T\mathbf{A}_2(\mathbf{w}_2)[\mathbf{r}_x(l_1), \ldots, \mathbf{r}_x(l_2)] = \mathbf{w}_1^T\mathbf{A}_2(\mathbf{w}_2)\mathbf{M}_x, \tag{A.4}$$

with $\mathbf{M}_x = [\mathbf{r}_x(l_1), \ldots, \mathbf{r}_x(l_2)]$ and $\mathbf{r}_{y_1y_2} = [r_{y_1y_2}(l_1), \ldots, r_{y_1y_2}(l_2)]$. Thus results:

$$\begin{aligned} C_2 &= \mathbf{w}_1^T\mathbf{A}_2(\mathbf{w}_2)\mathbf{M}_x\mathbf{M}_x^T\mathbf{A}_2^T(\mathbf{w}_2)\mathbf{w}_1 + \\ &\quad +\lambda_1(k - \mathbf{w}_1^T\mathbf{R}_x(0)\mathbf{w}_1) + \\ &\quad +\lambda_2(k - \mathbf{w}_2^T\mathbf{R}_x(0)\mathbf{w}_2), \\ &= \mathbf{w}_1^T\mathbf{G}_2\mathbf{w}_1 + \lambda_1(k - \mathbf{w}_1^T\mathbf{R}_x(0)\mathbf{w}_1) + \\ &\quad +\lambda_2(k - \mathbf{w}_2^T\mathbf{R}_x(0)\mathbf{w}_2), \end{aligned} \tag{A.5}$$

where $\mathbf{G}_2 = \mathbf{A}_2(\mathbf{w}_2)\mathbf{M}_x\mathbf{M}_x^T\mathbf{A}_2^T(\mathbf{w}_2)$. The gradient of (A.5) w.r.t $\mathbf{w}_1$ is:

$$\nabla_{\mathbf{w}_1}C_2 = 2\mathbf{G}_2\mathbf{w}_1 - 2\lambda_1\mathbf{R}_x(0)\mathbf{w}_1. \tag{A.6}$$

In a minimum it needs to be:

$$\nabla_{\mathbf{w}_1}C_2 = 0, \tag{A.7}$$

so that we have:

$$\mathbf{R}_x^{-1}(0)\mathbf{G}_2\mathbf{w}_1 = \lambda_1\mathbf{w}_1. \tag{A.8}$$

To reach a minimum of C_2, $\mathbf{w}_1$ is an eigenvector of $R_x^{-1}(0)G_2$ and λ_1 is the corresponding eigenvalue. Substituting in (A.5):

$$C_2 = \lambda_1\mathbf{w}_1^T\mathbf{R}_x(0)\mathbf{w}_1 + \lambda_1 k - \lambda_1\mathbf{w}_1^T\mathbf{R}_x(0)\mathbf{w}_1, \tag{A.9}$$

Therefore we choose λ_1 as the smallest eigenvalue.

In the same way for $\mathbf{w}_2$, having fixed $\mathbf{w}_1$, we have:

$$\begin{aligned} C_2 &= \mathbf{w}_2^T\mathbf{G}_1\mathbf{w}_2\lambda_1(k - \mathbf{w}_1^T\mathbf{R}_x(0)\mathbf{w}_1) + \\ &\quad +\lambda_2(k - \mathbf{w}_2^T\mathbf{R}_x(0)\mathbf{w}_2), \end{aligned} \tag{A.10}$$

where $G_1 = \mathbf{A}_1(\mathbf{w}_1)\mathbf{M}_x\mathbf{M}_x^T\mathbf{A}_1^T(\mathbf{w}_1)$ and

$$\mathbf{A}_1(\mathbf{w}_1) =$$

$$\begin{pmatrix} w_1(q) & w_1(q-1) & \cdots & w_1(0) & 0 & \cdots & 0 \\ 0 & w_1(q) & \cdots & w_1(1) & w_1(0) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & w_1(q) & w_1(q-1) & \cdots & w_1(0) \end{pmatrix} \tag{A.11}$$

Summary of the decorrelation algorithm:

1. Evaluate $r_x(l)$ with $l = \max(|l_1|, |l_2|) + q$ and M_x;
2. Initialize $\mathbf{w}_1$ and $\mathbf{w}_2$;
3. Construct $\mathbf{A}_1(\mathbf{w}1)$ and $\mathbf{G}_1$;
4. Update $\mathbf{w}_1$, replacing it by the eigenvector of $\mathbf{R}_x^{-1}(0)$ $\mathbf{G}_1$ with the smallest eigenvalue, subject to $\mathbf{w}_1\mathbf{R}_x(0)$ $\mathbf{w}_1 = k$;
5. Construct $\mathbf{A}_2(\mathbf{w}2)$ and $\mathbf{G}_2$;
6. Update $\mathbf{w}_2$, replacing it with the eigenvector of $\mathbf{R}_x^{-1}(0)$ $\mathbf{G}_2$ with the smallest eigenvalue, subject to $\mathbf{w}_2\mathbf{R}_x(0)$ $\mathbf{w}_2 = k$;
7. Stop if convergence is achieved, otherwise go to 4;

Despite other methods, in this case we do not have to evaluate outputs statistics after each updating.

B. REFERENCES

[1] A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7:1129 – 1159, 1995.

[2] A. Belouchrani, K.A. Meraim, J.F. Cardoso, and E. Moulines. A Blind Source Separation Technique Using Second-Order Statistics. *IEEE Transaction on Signal Processing*, 45(2):434 – 444, Fbr 1997.

[3] J.F. Cardoso. Blind Signal Separation: statistical principles. In *Proceedings of the IEEE, vol. 86, No. 10, October 1998*, pages 2009 – 2025. IEEE, 1998.

[4] D. Chan, S.J. Godsill, and P.J.W. Rayner. Multichannel Multitap Signal Separation by output decorrelation. Technical Report 0951-9211, Dept. of Engineering, Cambridge University, 1996.

[5] Simon Haykin. *Adaptive Filter Theory.* Prentice-Hall, 1996.

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

[7] K. Torkkola. Blind separation of convolved sources based on information maximization. In *IEEE Workshop on Neural Networks for Signal Processing*, Kyoto, Japan, 1996.

[8] B. Widrow and S. Stearns. *Adaptive Signal Processing.* Prentice-Hall, 1985.