ABSTRACT

Blind SIMO identification is challenging when additive noise is strong and for ill-conditioned/acoustic SIMO systems. A weighted cross relation (CR) algorithm presumably can be robust to noise but there lacks a practical way to define the weights. In this paper, the Pearson correlation coefficient (PCC) is used to develop an optimally weighted CR algorithm, which is validated by simulations.

Index Terms—Weighted cross relations, Pearson correlation coefficient, blind identification, acoustic SIMO system.

1. INTRODUCTION

Blind SIMO (single-input multiple-output) identification can find a variety of speech applications, e.g., time delay estimation for sound source localization [1] and speech dereverberation [2]. In these applications, acoustic impulse responses need to be known while the a priori knowledge of the source speech signal is unavailable, making the blind method a necessity.

This paper considers an acoustic SIMO system where the single input is a speech source and the outputs are microphone observations, as illustrated in Fig. 1. The n th system output y n (k) at time k is expressed as

\[ y_n(k) = g_n * s(k) + v_n(k) \]

(1)

where \( g_n \) is the channel impulse response from the source to the nth microphone, the symbol \( * \) denotes the linear convolution operator, \( s(k) \) is the source signal, and \( v_n(k) \) is the additive noise at the nth microphone. The channel impulse responses are delineated with finite impulse response (FIR) filters. The additive noise signals in different channels are assumed to be uncorrelated with the source signal and uncorrelated with each other.

In a vector/matrix form, the SIMO signal model (1) is written as

\[ \mathbf{y}_n(k) = \mathbf{G}_n \cdot \mathbf{s}(k) + \mathbf{v}_n(k) \]

(2)

where

\[ \mathbf{y}_n(k) = \begin{bmatrix} y_n(k) & y_n(k-1) & \cdots & y_n(k-L+1) \end{bmatrix}^T, \]

\[ \mathbf{G}_n = \begin{bmatrix} g_{n,0} & g_{n,1} & \cdots & g_{n,L-1} \\ 0 & g_{n,0} & \cdots & g_{n,L-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & g_{n,0} \end{bmatrix}_{L \times (2L-1)}, \]

\[ \mathbf{s}(k) = \begin{bmatrix} s(k) & s(k-1) & \cdots & s(k-2L+2) \end{bmatrix}^T, \]

\[ \mathbf{v}_n(k) = \begin{bmatrix} v_{n}(k) & v_{n}(k-1) & \cdots & v_{n}(k-L+1) \end{bmatrix}^T, \]

\[ \mathbf{x}_n(k) = \begin{bmatrix} x_{n}(k) & x_{n}(k-1) & \cdots & x_{n}(k-L+1) \end{bmatrix}^T, \]

\( [ ]^T \) denotes a vector/matrix transpose, and \( L \) is the length of the longest channel impulse response in such a SIMO system.

Therefore, the blind SIMO identification problem is to estimate \( \mathbf{g}_n = [g_{n,0} \ g_{n,1} \ \cdots \ g_{n,L-1}]^T, \quad n = 1, 2, \cdots, N \)

(3)

from the observations \( \mathbf{y}_n(k) \) without the knowledge of the source signal \( s(k) \).

The innovative idea of blind SIMO identification was first proposed by Sato in [3]. While higher (than second) order statistics (HOS) of the system outputs can be used (see [4] for a tutorial on the HOS-based approaches), second-order statistics (SOS) are sufficient to solve this problem [5]. The focus of the current blind SIMO identification research is primarily on the SOS-based methods. Celebrated work include the cross relation (CR) algorithm [6], [7], among many other variants (see [8] for a comprehensive survey on this subject).

According to [7], two conditions (one on the channel diversity and the other on the input signals) are necessary and sufficient to ensure blind SIMO identifiability, which are shared by all SOS-based methods:

1. The polynomials formed from \( g_{n} \) (\( n = 1, 2, \cdots, N \)) are co-prime, i.e., the channel transfer functions \( G_n(z) = \sum_{l=0}^{L-1} g_{n,l} z^{-l} \) do not share any common zeros;
2. The autocorrelation matrix of the input signal \( \mathbf{R}_s = E\{\mathbf{s}(k)\mathbf{s}^T(k)\} \) is of full rank, where \( E\{\cdot\} \) denotes mathematical expectation, such that the SIMO system can be fully excited.

It is already well known that existing blind SIMO identification algorithms are sensitive to additive noise, particularly for ill-conditioned (in terms of the assumption of no shared common zeros) SIMO systems, hence impairing the usefulness of such a technique in practice. In this paper, we intend to use the Pearson correlation coefficient (PCC) to develop an optimally weighted CR algorithm for use in noisy environments.
2. WEIGHTED AND TRADITIONAL CROSS RELATION ALGORITHMS

In the absence of noise (i.e., \( y_n = x_n \)), a SIMO system has the following cross relations (CRs):

\[
y_i * g_j = s * g_i * g_j = y_j * g_i, \quad i, j = 1, 2, \ldots, N. \tag{4}
\]

At time \( k \), we then have

\[
y_i^T(k)g_j = y_j^T(k)g_i, \quad i, j = 1, 2, \ldots, N. \tag{5}
\]

Multiplying (5) by \( y_i(k) \) from the left side and taking expectation yields

\[
R_{y_jy_i} = R_{y_iy_j}, \quad i, j = 1, 2, \ldots, N, \tag{6}
\]

where \( R_{y_jy_i} \triangleq E\{y_i(k)y_j^T(k)\} \).

When noise is present, for \( N \) model FIR filters \( h_n \) (\( n = 1, 2, \ldots, N \)), an error signal can be defined as follows

\[
e_i^2(k) \triangleq y_i^T(k)h_j - y_j^T(k)h_i. \tag{7}
\]

Accordingly, a weighted cost function can be formulated as

\[
J = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} w_{ij} \cdot E\{e_i^2(k)\}, \tag{8}
\]

where \( w_{ij} > 0 \) are weighting factors and should be symmetric, i.e., \( w_{ij} = w_{ji} \). We can use the expression

\[
E\{e_i^2(k)\} = h_i^T R_{y_jy_i} h_j + h_j^T R_{y_iy_j} h_i - 2 h_i^T R_{y_jy_i} h_i, \tag{9}
\]

and compute the gradient of (8) with respect to \( h_n \):

\[
\frac{\partial J}{\partial h_n} = \sum_{i=1}^{N-1} \sum_{j=n+1}^{N} \frac{\partial E\{e_i^2(k)\}}{\partial h_n} \cdot w_{ij} + \sum_{j=n+1}^{N} \frac{\partial E\{e_i^2(k)\}}{\partial h_n} \cdot w_{ij} \nonumber
\]

\[
= 2 \sum_{i=1, i \neq n}^{N} (R_{y_jy_i} h_i - R_{y_iy_j} h_i) w_{in}, \tag{10}
\]

Equating (10) to zero and putting the \( N \) expressions in a matrix form yields

\[
R_{y_jy_i} w_{ij} = 0_{N \times 1}, \tag{11}
\]

where

\[
R_{y_jy_i} = \begin{bmatrix}
\sum_{n \neq 1} w_{1n} \cdot R_{y_1y_n} & -w_{12} R_{y_2y_1} & \cdots & -w_{1N} R_{y_Ny_1} \\
-w_{21} R_{y_1y_2} & \sum_{n \neq 2} w_{2n} R_{y_ny_2} & \cdots & -w_{2N} R_{y_Ny_2} \\
\vdots & \ddots & \ddots & \vdots \\
-w_{N1} R_{y_1y_N} & -w_{N2} R_{y_2y_N} & \cdots & \sum_{n \neq N} w_{Nn} R_{y_Ny_n}
\end{bmatrix}
\]

\[
h = \begin{bmatrix}
h_1^T \\
h_2^T \\
\vdots \\
h_N^T
\end{bmatrix}^T.
\]

A weighted CR (WCR) algorithm is then deduced with the solution of (11), i.e., the eigenvector of \( R_{y_jy_i} w_{ij} \) corresponding to its smallest eigenvalue.

The error signal \( e_i^2(k) \) defined in (7) consists of two parts: a modeling error and the error caused by additive noise. The modeling error is what we want to minimize over the model filters. But the additive noise has a negative impact on this minimization procedure. Therefore, intuitively if the additive noise signals in the \( i \)th and \( j \)th channels are stronger than that in the other channels, then \( e_i^2(k) \) should be de-emphasized in the cost function and \( w_{ij} \) should be relatively smaller. From this perspective, the WCR algorithm is a neat idea for its robustness with noise. But there is no straightforward ways to quantify \( w_{ij} \) in practice. Therefore, \( w_{ij} \) has to be set as 1, leading to the traditional CR method. In the next sections, we will show how the Pearson correlation coefficient can help develop an optimally weighted CR algorithm.

3. BLIND SIMO IDENTIFICATION WITH THE SQUARED PEARSON CORRELATION COEFFICIENT

The Pearson correlation coefficient (PCC) of two zero-mean real-valued random variables \( y_1 \) and \( y_2 \) is defined as [9]:

\[
\rho(y_1, y_2) = \frac{E\{y_1y_2\}}{\sigma_{y_1}\sigma_{y_2}}, \tag{12}
\]

where \( \sigma_{y_1}^2 = E\{y_1^2\} \) and \( \sigma_{y_2}^2 = E\{y_2^2\} \) are the variances of the signals \( y_1 \) and \( y_2 \), respectively. In the context of blind SIMO identification, it will be more convenient to work with the squared Pearson correlation coefficient (SPCC):

\[
\rho^2(y_1, y_2) = \frac{E\{y_1y_2\}^2}{\sigma_{y_1}^2\sigma_{y_2}^2}. \tag{13}
\]

The SPCC gives an indication on the strength of the linear relationship between the two random variables \( y_1 \) and \( y_2 \). We always have \( 0 \leq \rho^2(y_1, y_2) \leq 1 \). If \( \rho^2(y_1, y_2) = 0 \), then \( y_1 \) and \( y_2 \) are said to be uncorrelated. The closer the value of \( \rho^2(y_1, y_2) \) is to 1, the stronger the correlation between the two variables.

The concept of SPCC can be generalized to the multichannel case. Let \( y_1, y_2, \ldots, y_N \) be \( N \) zero-mean real-valued random variables. One possible definition for the multichannel SPCC is

\[
\rho^2(y_1, y_2, \ldots, y_N) = \frac{2}{N(N - 1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho^2(y_i, y_j). \tag{14}
\]

This definition considers all possible two-channel SPCCs between \( y_i \) and \( y_j \), with \( i \neq j \), and counts the \( \rho^2(y_i, y_j) \) pair only once. It can be easily checked that \( 0 \leq \rho^2(y_1, y_2, \ldots, y_N) \leq 1 \). If all the signals are completely correlated, then \( \rho^2(y_1, y_2, \ldots, y_N) = 1 \). If all the signals are completely uncorrelated with each other, then \( \rho^2(y_1, y_2, \ldots, y_N) = 0 \).

In the problem of blind SIMO identification, instead of using the traditional mean square error (MSE) to define (7), we can measure the difference between the signals \( h_i^T y_1(k) \) and \( h_j^T y_1(k) \) with the SPCC:

\[
\rho^2[h_i^T y_1(k), h_j^T y_1(k)] = \frac{(h_i^T R_{y_jy_j} h_j)^2}{(h_i^T R_{y_jy_j} h_j)^2}. \tag{15}
\]

Then the channel impulse responses are determined by searching the model filters that maximize the multichannel SPCC

\[
\rho^2(h) \triangleq \frac{2}{N(N - 1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \rho^2[h_i^T y_1(k), h_j^T y_1(k)]. \tag{16}
\]
where
\[\alpha_{ij}(\mathbf{h}) = \frac{h_i^T \mathbf{R}_{yy,j} h_j}{h_i^T \mathbf{R}_{yy} h_i},\]
\[\beta_{ij}(\mathbf{h}) = \frac{h_i^T \mathbf{R}_{yy,j} h_j}{(h_i^T \mathbf{R}_{yy} h_i)(h_j^T \mathbf{R}_{yy} h_j)} = \frac{\alpha_{ij}(\mathbf{h})}{(h_j^T \mathbf{R}_{yy} h_j)} = \beta_{ij}(\mathbf{h}).\]

It can be checked that \(\beta^2(\mathbf{h}) = 1\) if and only if \(h_n = c_n g_n\), where \(c_n \neq 0\) \((n = 1, 2, \ldots, N)\) are arbitrary constants.

Taking the gradient of \(\beta^2(\mathbf{h})\) with respect to \(h_n\) produces
\[\frac{\partial \beta^2(\mathbf{h})}{\partial h_n} = \frac{4}{N(N - 1)} \left( \sum_{i=1,i\neq n}^N \alpha_{in}(\mathbf{h}) \beta_{in}(\mathbf{h}) \mathbf{R}_{yy,n} h_n - \sum_{i=1,i\neq n}^N \alpha_{jn}(\mathbf{h}) \beta_{jn}(\mathbf{h}) \mathbf{R}_{yy,j} h_n \right) + \alpha_{nn}(\mathbf{h}) \beta_{nn}(\mathbf{h}) \mathbf{R}_{yy,n} h_n - \frac{2}{N}\sum_{i=1,i\neq n}^N \alpha_{in}(\mathbf{h}) \beta_{in}(\mathbf{h}) \mathbf{R}_{yy,n} h_n - \frac{2}{N}\sum_{i=1,i\neq n}^N \alpha_{jn}(\mathbf{h}) \beta_{jn}(\mathbf{h}) \mathbf{R}_{yy,j} h_n,\]
\[n = 1, 2, \ldots, N.\]

Equating the gradient (17) to zero and putting the \(N\) expressions in a matrix form yields
\[\mathbf{R}_{yy}^{-1} \cdot \mathbf{h} = 0_{N \times 1},\]
(18)

where
\[\mathbf{R}_{yy}(\mathbf{h}) = \begin{bmatrix} D_1(\mathbf{h}) & -\beta_{12}(\mathbf{h}) \mathbf{R}_{yy,2} & \cdots & -\beta_{1N}(\mathbf{h}) \mathbf{R}_{yy,N} \\ -\beta_{12}(\mathbf{h}) \mathbf{R}_{yy,2} & D_2(\mathbf{h}) & \cdots & -\beta_{2N}(\mathbf{h}) \mathbf{R}_{yy,N} \\ \vdots & \vdots & \ddots & \vdots \\ -\beta_{1N}(\mathbf{h}) \mathbf{R}_{yy,N} & -\beta_{2N}(\mathbf{h}) \mathbf{R}_{yy,2} & \cdots & D_N(\mathbf{h}) \end{bmatrix},\]
and
\[D_i(\mathbf{h}) = \sum_{n=1,n\neq i}^N \alpha_{in}(\mathbf{h}) \beta_{in}(\mathbf{h}) \mathbf{R}_{yy,n}, \quad i = 1, 2, \ldots, N.\]

Equation (18) is highly nonlinear with respect to \(\mathbf{h}\), but a simple way to solve it is by iterations. The eigenvector of \(\mathbf{R}_{yy}\) corresponding to its smallest eigenvalue, i.e., the solution of the traditional CR algorithm, is taken as the initial estimate \(\mathbf{h}(0)\). Then, in the \(t\)th \((t \geq 1)\) iteration, \(\mathbf{h}(t)\) is updated by the eigenvector of \(\mathbf{R}_{yy}^{-1}[\mathbf{h}(t - 1)]\) corresponding to its smallest eigenvalue. This procedure proceeds until convergence or a specified maximum number of iterations \(T\) has been reached. This iterative procedure is summarized in Table 1.

When noise is weak and \(\mathbf{h} \approx g\) after convergence, we learn from the cross relation (6) that
\[\alpha_{ij}(\mathbf{h}) \approx 1,\]
\[\beta_{ij}(\mathbf{h}) \approx \frac{1}{h_i^T \mathbf{R}_{yy,j} h_j} = \frac{1}{h_i^T \mathbf{R}_{yy} h_i} \geq 0.\]

In this case, (18) evolves into (11) with \(w_{ij} = \beta_{ij}(\mathbf{h})\). So the use of SPCC provides an optimal way to define \(w_{ij}\) in the WCR algorithm. From (20), we see that \(\beta_{ij}\) is reverse proportional to the power of the \(j\)th and/or the \(i\)th channel outputs. This makes sense since a channel output with a higher power implies stronger additive noise in this channel (assuming the same gain over all channels of the SIMO system). Then, per the discussion at the end of last section, \(w_{ij}\) should be lower.

### 4. Simulations
In this section, we will evaluate the performance of the developed SPCC algorithm in comparison with the CR batch method by simulations. Similar to our earlier studies on this subject, we use the normalized projection misalignment (NPM) in dB as the performance measure, which is given by
\[\text{NPM}(\mathbf{h}, g) = 20 \log_{10} \frac{1}{N} \sum_{n=1}^N \left\| g_n - \frac{g_n^T \mathbf{h}_n}{\mathbf{h}_n^T \mathbf{h}_n} \cdot \mathbf{h}_n \right\| / \| g_n \|.\]
(21)

The first experiment is concerned with a simple three-channel SIMO system whose impulse responses are
\[g_1 = [1 - 2 \cos(\theta)]^T,\]
\[g_2 = [1 - 2 \cos(\theta + \vartheta)]^T,\]
\[g_3 = [1 - 2 \cos(\theta + 2\vartheta)]^T,\]
where \(\theta\) and \(\vartheta\) controls the positions of the zeros of the three channels. For \(\theta = \pi/10\) and \(\vartheta = 2\pi/3\), the zeros of the three channels are separated far away to each other. The system is definitely blindly identifiable and hence is deemed well-conditioned (WC). For \(\theta = \vartheta = \pi/10\), the SIMO system is regarded as ill-conditioned (IC) since the zeros of the two channels are quite close, which is about to invalidate the identifiability assumption of no common zeros. Such a SIMO system was first introduced in [10] (with only the first two channels) and was then widely employed in the studies of blind SIMO identification.

The source is an uncorrelated binary phase-shift-keying (BPSK) sequence and the additive noise is i.i.d. zero-mean Gaussian at a specified signal-to-noise ratio (SNR) defined as follows
\[\text{SNR}_n = 10 \log_{10} \frac{\sigma^2}{\sigma_n^2}, \quad n = 1, 2, \ldots, N.\]
(23)

We set the SNR of the first two channels always equal but 10 dB higher than that of the third channel in order to study whether the weights in the SPCC algorithm would vary with the channel SNR as expected from our analysis.

For each channel of the SIMO system, 500 output samples were used. The SPCC algorithm took less than \(T = 5\) iterations to converge.

For each set of specified SNRs, we averaged the NPMs of 100 Monte-Carlo trials. The results are presented in Fig. 2. We see that the SPCC performs better than the batch CR for low SNRs and their performances are comparable for high SNRs. This advantage of the SPCC is more evident for the ill-conditioned system.
CR for low SNRs. The results are presented in Fig. 5. The SPCC outperforms the batch new algorithm is more robust than the batch cross-relation method correlation coefficient (SPCC). Simulation results indicated that the identification algorithm was developed by using the squared Pearson An optimally weighted cross-relation algorithm for blind SIMO/CC 100 Monte-Carlo trials were conducted. For the SPCC, channels are equal but 10 dB higher than that of the third channel. signals are i.i.d. zero-mean Gaussian. The SNR’s of the first two the source. It is visualized in Fig. 4 (b). Again the additive noise speech signal of 5500 samples long (sampled at 8 kHz) was used as three channel impulse responses are shown in Fig. 4 (a). A female in a reverberant environment,” IEEE Trans. Speech Audio Process., vol. 13, pp. 882–896, Sept. 2005. 5. CONCLUSIONS An optimally weighted cross-relation algorithm for blind SIMO identification algorithm was developed by using the squared Pearson correlation coefficient (SPCC). Simulation results indicated that the new algorithm is more robust than the batch cross-relation method for ill-conditioned or acoustic SIMO systems and when the SNR is low.

6. REFERENCES


