

# CROSS-RELATION BASED BLIND SYSTEM IDENTIFICATION USING ONLINE ADMM

Matthias Blochberger\*, Filip Elvander†, Randall Ali, Toon van Waterschoot

KU Leuven  
ESAT - Department of Electrical Engineering  
STADIUS  
3001 Leuven

## ABSTRACT

The abstract should appear at the top of the left-hand column of text, about 0.5 inch (12 mm) below the title area and no more than 3.125 inches (80 mm) in length. Leave a 0.5 inch (12 mm) space between the end of the abstract and the beginning of the main text. The abstract should contain about 100 to 150 words, and should be identical to the abstract text submitted electronically. All manuscripts must be in English.

**Index Terms**— blind system identification, multichannel signal processing, ADMM, Online-ADMM

## 1. INTRODUCTION

Its my pleasure to introduce to you: Algorithm. The most accurate, the fastest, the one most robust to noise; Simply the best you can get. Please give a warm welcome and a big applause!

## 2. PROBLEM STATEMENT

### 2.1. Signal Model

We define the acoustic SIMO system with the input signal  $\mathbf{s}(n) = [s(n) \ s(k-1) \ \dots \ s(k-2L+2)]^T$  and  $i \in [M]$  outputs  $\mathbf{x}_i(n) = [x_i(n) \ x_i(k-1) \ \dots \ x_i(k-L+1)]^T$ . Each output  $\mathbf{x}_i$  is the convolution of  $\mathbf{s}$  with the respective channel impulse response  $\mathbf{h}_i$  with additive noise term  $\mathbf{v}_i$ , assumed to be zero-mean and uncorrelated. The signal model is described by

$$\mathbf{x}_i(n) = \mathbf{H}_i \mathbf{s}(n) + \mathbf{v}_i(n) \quad (1)$$

where  $\mathbf{H}_i$  is the  $L \times (2L-1)$  linear convolution matrix of the  $i$ th channel using the elements of  $\mathbf{h}_i$ .

**Note:** Add definition of  $\mathbf{H}_i$ ? Will take a lot of space.

### 2.2. Cross-relation approach

The cross-relation approach for BSI [] aims to only use output signals of the system to identify it. This is achieved by exploiting the relative channel information when more than one system output are available and certain identifiability conditions are satisfied:

- Condition 1
- Condition 2

The fundamental equality of this approach is

$$\mathbf{x}_i^T(n) \mathbf{h}_j = \mathbf{x}_j^T(n) \mathbf{h}_i, \quad i, j = 1, 2, \dots, M, i \neq j \quad (2)$$

which states that the output signal of one channel convolved with the impulse response of another is equal to the vice-versa. This follows from the commutative nature of the convolution operation. Using the (estimated) covariance  $\mathbf{R}_{ij} = E\{\mathbf{x}_i \mathbf{x}_j^T\}$  instead of the signal vectors  $\mathbf{x}_i$  yields a more convenient problem formulation. For the subsequent explanations the time index is dropped to make notation more compact, however the variables are time-variant.

We can combine all cross-relations (2) into a linear system of equations

$$\mathbf{R} \mathbf{h} = \mathbf{0} \quad (3)$$

where

$$\mathbf{R} = \begin{bmatrix} \sum_{m \neq 1} \mathbf{R}_{mm} & -\mathbf{R}_{21} & \dots & -\mathbf{R}_{M1} \\ -\mathbf{R}_{12} & \sum_{m \neq 2} \mathbf{R}_{mm} & \dots & -\mathbf{R}_{M2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{1M} & -\mathbf{R}_{2M} & \dots & \sum_{m \neq M} \mathbf{R}_{mm} \end{bmatrix}, \quad (4)$$

and  $\mathbf{h} = [\mathbf{h}_1^T \ \dots \ \mathbf{h}_M^T]^T$ .

**Note:** This only has a non-zero solution when  $\mathbf{R}$  is not full rank i.e. has at least one eigenvalue the is zero. It is full rank though, i think. So is this formulation weird to begin with?

In the presence of noise, the most convenient way to solve this problem by equating it to an error term  $\mathbf{R} \mathbf{h} = \mathbf{e}$  and minimizing this squared-error cost function

$$J = \|\mathbf{R} \mathbf{h}\|^2 \quad (5)$$

\*THANK EU!

†THANK FWQ?

which takes the form of

$$\begin{aligned} & \text{minimize} \quad \|\mathbf{R}\mathbf{h}\|^2, \\ & \text{s.t.} \quad \|\mathbf{h}\| = 1. \end{aligned} \quad (6)$$

The minimizer  $\hat{\mathbf{h}}$  of this problem is the eigenvector corresponding to the smallest eigenvalue of  $\mathbf{R}^T\mathbf{R}$ . This can be formulated in simpler form as

$$\begin{aligned} & \hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \mathbf{h}^T \mathbf{R} \mathbf{h}, \\ & \text{s.t.} \quad \|\mathbf{h}\| = 1, \end{aligned} \quad (7)$$

since the eigenvectors for the squared term  $\mathbf{R}^T\mathbf{R}$  and the non-squared one  $\mathbf{R}$  are the same.

**Attention:** Rewrite to have later section relate better.

### 3. REVIEW MULTICHANNEL NEWTON & QUASI-NEWTON

#### 4. PROPOSED METHOD

This method uses Online-ADMM to find a solution to the minimization problem posed in previous sections. To achieve this, we introduce reduced problem with the same minimizer.

**Attention:** Show that!

This problem can be split into smaller sub-problems which can be solved in parallel to reduce computation time.

##### 4.1. Problem Splitting

In state-of-the-art algorithms, the adaptive minimization problem (7) is solved in its full extend.

**Note:** Wording

We replace the cost function (5) with the separable cost function

$$\tilde{J} = \sum_{i=1}^M \tilde{J}_i = \sum_{i=1}^M \mathbf{w}_i^T \mathbf{P}_i \mathbf{w}_i \quad (8)$$

where  $\mathbf{w}_i$  is defined as the stacked vector of estimated impulse responses analogous to  $\mathbf{h}$  (cf. subsection 2.2) however only using a subset of the full channel set  $\mathcal{C}_i \subseteq [M]$ . Analogous,  $\mathbf{P}$  is constructed as defined in (4) with the reduced set of channels. In case the channel subsets are proper  $\mathcal{C}_i \subset [M]$ , this leads to sub-problems each with smaller dimensions than the original centralized problem, reducing complexity. It does not use all cross-relations available, however this does not impact identification performance significantly.

**Attention:** Show that!

In Figure 1 we try to visualize which information is used by the sub-problems compared to the centralized problem.

**Attention:** Write about how and why the solution is the same (why is it actually...). The optimal solution to the smaller problems is still the stacked impulse response of the subset of channel impulse responses. Can a relation to the smallest-eigenvector comment be found?

	$\mathbf{R}_{21}$	$\mathbf{R}_{31}$	$\mathbf{R}_{41}$	$\mathbf{R}_{51}$
$\mathbf{R}_{12}$		$\mathbf{R}_{31}$	$\mathbf{R}_{42}$	$\mathbf{R}_{52}$
$\mathbf{R}_{13}$	$\mathbf{R}_{23}$		$\mathbf{R}_{43}$	$\mathbf{R}_{53}$
$\mathbf{R}_{14}$	$\mathbf{R}_{24}$	$\mathbf{R}_{34}$		$\mathbf{R}_{54}$
$\mathbf{R}_{15}$	$\mathbf{R}_{25}$	$\mathbf{R}_{35}$	$\mathbf{R}_{45}$	

**Fig. 1:** Problem splitting

#### 4.2. General Consensus ADMM

The separability of the problem can be used to take advantage of the well known alternating direction method of multipliers (ADMM). Here more specifically, general consensus ADMM [] is used

$$\begin{aligned} & \text{minimize} \quad \sum_{i=1}^M \tilde{J}_i(\mathbf{w}_i) \\ & \text{subject to} \quad \mathbf{w}_i = \mathcal{G}_i(\mathbf{h}), \quad i = 1, \dots, M, j = 1, \dots, L \end{aligned} \quad (9)$$

$$\text{subject to} \quad \mathbf{w}_i = \mathcal{G}_i(\mathbf{h}), \quad i = 1, \dots, M, j = 1, \dots, L \quad (10)$$

Where  $\mathcal{G}_i$  is a selection operator linking the single-channel impulse response elements between the consensus variable with sub-problem estimates.

**Note:** Better description of variables and mapping; + maybe figure to explain. maybe as mapping matrix:

$$\mathbf{G}_i = \begin{bmatrix} \mathbf{0}_{L \times L} & \mathbf{0}_{L \times L} & \mathbf{I}_{L \times L} \\ \mathbf{I}_{L \times L} & \mathbf{0}_{L \times L} & \mathbf{0}_{L \times L} \end{bmatrix}_{M_i L \times M L}. \quad (11)$$

The augmented Lagrangian for this particular general-form consensus problem is

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{w}, \mathbf{h}, \mathbf{u}) = & \sum_{i=1}^M \left( \mathbf{w}_i^T \mathbf{P}_i \mathbf{w}_i + \mathbf{u}_i^T (\mathbf{w}_i - \mathcal{G}_i(\mathbf{h})) \right. \\ & \left. + \frac{\rho}{2} \|\mathbf{w}_i - \mathcal{G}_i(\mathbf{h})\|^2 \right). \end{aligned} \quad (12)$$

The ADMM then consists of the iterations

$$\begin{aligned} \mathbf{w}_i^{k+1} = \underset{\mathbf{w}_i}{\operatorname{argmin}} \left\{ & \mathbf{w}_i^T \mathbf{P}_i \mathbf{w}_i + \mathbf{u}_i^{kT} (\mathbf{w}_i - \mathcal{G}_i(\mathbf{h}^k)) \right. \\ & \left. + \frac{\rho}{2} \|\mathbf{w}_i - \mathcal{G}_i(\mathbf{h}^k)\|^2 \right\} \end{aligned} \quad (13)$$

$$\begin{aligned} \mathbf{h}^{k+1} = \underset{\mathbf{h}, \|\mathbf{h}\|=1}{\operatorname{argmin}} \left\{ & \sum_{i=1}^M \left( \mathbf{u}_i^{kT} \mathcal{G}_i(\mathbf{h}) \right. \right. \\ & \left. \left. + \frac{\rho}{2} \|\mathbf{w}_i^{k+1} - \mathcal{G}_i(\mathbf{h})\|^2 \right) \right\} \end{aligned} \quad (14)$$

$$\mathbf{u}_i^{k+1} = \mathbf{u}_i^k + \rho (\mathbf{w}_i^{k+1} - \mathcal{G}_i(\mathbf{h}^{k+1})) \quad (15)$$

### 4.3. Online ADMM

ADMM is originally an iterative method to solve optimization problems, however under certain conditions it can be applied as an adaptive algorithm, also referred to as *Online ADMM* [].

**Attention:** Give conditions and why they apply here.

The data term as part of (12) is time-dependent, which from here on will be denoted with the included time index  $n$  in superscript as  $\mathbf{w}_i^T \mathbf{P}_i^n \mathbf{w}_i$ .

The iterative form of the update steps as described in (13)-(15) can be transformed into an adaptive one by computing a finite number at each time step  $n$  with the current data term. In this algorithm, one ADMM iteration is applied per time step, which allows us to replace the iteration index  $k$  with the time index  $n$ .

The minimization problem (13) can be solved by different means, in this case we apply a Newton update of the form of

$$\mathbf{w}_i^{n+1} = \mathbf{w}_i^n - \mu \mathbf{V}^n (\mathbf{P}_i^n \mathbf{w}_i^n + \mathbf{u}_i^n + \rho (\mathbf{w}_i^n - \mathcal{G}_i(\mathbf{h}^n))) \quad (16)$$

where  $0 < \mu \leq 1$  is a step size and  $\mathbf{V}^n = (\mathbf{P}_i^n + \rho \mathbf{I})^{-1}$  is the inverse Hessian of the problem. This inverse can be updated directly by applying a set of rank-1 updates or a low-rank update using the matrix inversion lemma, which reduces computational complexity significantly.

**Attention:** Elaborate on this? Its probably too much to be explained in detail.

To compute the consensus  $\mathbf{h}$  we include the unit-norm constraint with a lagrange multiplier  $\lambda$  in the minimization problem and rewrite the average of penalties into a penalty of averages []

**Attention:** poetic, cite our Boy[d]

$$\mathbf{h}^{k+1} = \underset{\mathbf{h}}{\operatorname{argmin}} \left\{ \frac{\lambda}{2} (\mathbf{h}^T \mathbf{h} - 1) + \frac{M\rho}{2} \|\mathbf{h} - \bar{\mathbf{w}}^{n+1} - \rho^{-1} \bar{\mathbf{u}}^n\|^2 \right\} \quad (17)$$

where  $\bar{\mathbf{w}}^{n+1} = \sum_{i=1}^M \mathcal{G}_i^{-1}(\mathbf{w}_i^{n+1})$

**Attention:** Find a concise way to write down the partial consensus. Like in Boyd

## 5. PERFORMANCE EVALUATION

## 6. CONCLUSIONS