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 $\mathbf{s}(n) = \begin{bmatrix} s(n) & s(k-1) & \cdots & s(k-2L+2) \end{bmatrix}^\mathrm{T}$ and $i \in [M]$ outputs $\mathbf{x}_i(n) = \begin{bmatrix} x_i(n) & x_i(k-1) & \cdots & x_i(k-L+1) \end{bmatrix}^\mathrm{T}$. 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

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

where $\mathbf{H}_i$ is the $L \times (2L-1)$ linear convolution matrix of the ith 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

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

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 $\mathbf{R}_{ij} = \mathrm{E}\left\{\mathbf{x}_i\mathbf{x}_j^\mathrm{T}\right\}$ 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

$$\mathbf{R}\mathbf{h} = \mathbf{0} \tag{3}$$

where

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

and $\mathbf{h} = \begin{bmatrix} \mathbf{h}_1^\mathrm{T} & \cdots & \mathbf{h}_M^\mathrm{T} \end{bmatrix}^\mathrm{T}$.

    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

$$J = \|\mathbf{R}\mathbf{h}\|^2 \tag{5}$$

which takes the form of

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

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

$$\hat{\mathbf{h}} = \arg\min_{\mathbf{h}} \mathbf{h}^{\mathrm{T}}\mathbf{R}\mathbf{h}, \qquad (7)$$
$$\text{s.t. } \|\mathbf{h}\| = 1,$$

since the eigenvectors for the squared term $\mathbf{R}^{\mathrm{T}}\mathbf{R}$ 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: State assumption that there as many nodes as there are channels/impulse earlier. Maybe replace the word "channel", sounds very communication.

The problem is split into smaller sub-problems each defined by a subset of the full channel set $\mathcal{C}_i \subseteq [M]$. Following from that, we define the sets $\bar{\mathcal{C}}_j = \{i | j \in \mathcal{C}_i\}$ for $i, j \in [M]$ representing the sub-problems a channel is part of. Further, $M_i = |\mathcal{C}_i|$ and $N_j = |\bar{\mathcal{C}}_j|$ with $i, j \in [M]$.

Attention: Is notation $i \in [M]$ instead of $i = 1, ..., M$ okay?

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

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

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 . 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

Fig. 1: Problem splitting

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?

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

$$\text{minimize} \quad \sum_{i=1}^{M} \tilde{J}_i(\mathbf{w}_i) \qquad (9)$$
$$\text{subject to} \quad (\mathbf{w}_i)_j = \mathbf{h}_{\mathcal{G}(i,j)}, \quad i = 1, ..., M, j = 1, .., M_i \qquad (10)$$

where $\mathcal{G}(i, j) = g$ denotes the mapping of L local variable components $(\mathbf{w}_i)_j$, i.e. one of the impulse responses in the stacked vector, to the corresponding global variable components $(\mathbf{h})_g$ (cf. FIGURE). For brevity, a mapped global variable $\tilde{\mathbf{h}}_i$ with $(\tilde{\mathbf{h}}_i)_j = \mathbf{h}_{\mathcal{G}(i,j)}$ is defined.

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

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

$$\mathcal{L}_\rho(\mathbf{w}, \mathbf{h}, \mathbf{u}) = \sum_{i=1}^{M} \left(\mathbf{w}_i^{\mathrm{T}}\mathbf{P}_i\mathbf{w}_i + \mathbf{u}_i^{\mathrm{T}}\left(\mathbf{w}_i - \tilde{\mathbf{h}}_i\right) \right.$$
$$\left. + \frac{\rho}{2}\left\|\mathbf{w}_i - \tilde{\mathbf{h}}_i\right\|^2 \right). \qquad (11)$$

The ADMM then consists of the iterations

$$\mathbf{w}_i^{k+1} = \operatorname*{argmin}_{\mathbf{w}_i} \left\{ \mathbf{w}_i^{\mathrm{T}} \mathbf{P}_i \mathbf{w}_i + \mathbf{u}_i^{k\,\mathrm{T}} \left(\mathbf{w}_i - \tilde{\mathbf{h}}_i^k \right) \right.$$
$$\left. + \frac{\rho}{2} \left\| \mathbf{w}_i - \tilde{\mathbf{h}}_i^k \right\|^2 \right\} \qquad (12)$$

$$\mathbf{h}^{k+1} = \operatorname*{argmin}_{\mathbf{h}, \|\mathbf{h}\|=1} \left\{ \sum_{i=1}^{M} \left(\mathbf{u}_i^{k\,\mathrm{T}} \tilde{\mathbf{h}}_i \right. \right.$$
$$\left. \left. + \frac{\rho}{2} \|\mathbf{w}_i^{k+1} - \tilde{\mathbf{h}}_i\|^2 \right) \right\} \qquad (13)$$

$$\mathbf{u}_i^{k+1} = \mathbf{u}_i^k + \rho \left(\mathbf{w}_i^{k+1} - \tilde{\mathbf{h}}_i^{k+1} \right) \qquad (14)$$

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

The iterative form of the update steps as described in (12)-(14) 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 (12) can be solved by various methods, in this case however we perform a Newton update step of the form of

$$\mathbf{w}_i^{n+1} = \mathbf{w}_i^n - \mu \mathbf{V}^n \left(\mathbf{P}_i^n \mathbf{w}_i^n + \mathbf{u}_i^n + \rho \left(\mathbf{w}_i^n - \mathcal{G}_i(\mathbf{h}^n) \right) \right) \qquad (15)$$

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 $\|\mathbf{h}\| = 1$ with a Lagrange multiplier λ in the minimization problem and replace the average of penalties with a penalty of averages as

Attention: poetic, cite our Boy[d]

$$\mathbf{h}^{k+1} = \operatorname*{argmin}_{\mathbf{h}} \left\{ \frac{\lambda}{2} \left(\mathbf{h}^{\mathrm{T}} \mathbf{h} - 1 \right) \right.$$
$$\left. + \frac{M\rho}{2} \left\| \mathbf{h} - \bar{\mathbf{w}}^{n+1} - \frac{1}{\rho} \bar{\mathbf{u}}^n \right\|^2 \right\} \qquad (16)$$

where the $ML \times 1$ vectors $\bar{\mathbf{w}}^{n+1}, \bar{\mathbf{u}}^{n+1}$ are computed as the mapped averages

$$(\bar{\mathbf{w}}^{n+1})_g = \frac{1}{N_g} \sum_{\mathcal{G}(i,j)=g} (\mathbf{w}_i^{n+1})_j, \quad g, i, j \in [M], \quad (17)$$

and

$$(\bar{\mathbf{u}}^n)_g = \frac{1}{N_g} \sum_{\mathcal{G}(i,j)=g} \frac{1}{\rho} (\mathbf{u}_i^n)_j, \quad g, i, j \in [M]. \quad (18)$$

Attention: UGLY AND CONFUSING??!!

Setting the derivative of (16) to zero and solving for the optimal Lagrange multiplier λ that ensures the constraint, we reach the update step

$$\mathbf{h}^{k+1} = \frac{\bar{\mathbf{w}}^{n+1} + \frac{1}{\rho} \bar{\mathbf{u}}^n}{\left\| \bar{\mathbf{w}}^{n+1} + \frac{1}{\rho} \bar{\mathbf{u}}^n \right\|}. \qquad (19)$$

ALGORITHM 1 summarizes the procedure.

Algorithm 1 BSI-ADMM

---

for $n = 0, 1, 2, ...$ **do**
    for $i \in [M]$ **do**
        Acquire new data vector $x_i^{(t)}$
        Update $\mathbf{w}_i^{(n+1)} \leftarrow \mathbf{w}_i^{(n)}$
    end for
    Update $\mathbf{h}^{(n+1)} \leftarrow \mathbf{h}^{(n)}$
    for $i \in [M]$ **do**
        Update $\mathbf{u}_i^{(t+1)} \leftarrow \mathbf{u}_i^{(n)}$
    end for
end for

---

5. PERFORMANCE EVALUATION

First part: random impulse responses at different SNRs. Short like 64 tabs e.g. Second part: simulated IRs with WGN input and speech??? maybe not haha

6. CONCLUSIONS

LINE GO DONW