CROSS-RELATION-BASED FREQUENCY-DOMAIN BLIND SYSTEM IDENTIFICATION USING ONLINE ADMM

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

KU Leuven
Department of Electrical Engineering (ESAT)
STADIUS Center for Dynamical Systems, Signal Processing and Data Analytics
3001 Leuven, Belgium

ABSTRACT

In this contribution, we propose a cross-relation-based adaptive algorithm for blind identification of single-input multiple-output (SIMO) systems in the frequency domain using the alternating direction method of multipliers (ADMM). The proposed algorithm exploits the separability of the cross-channel relations by splitting the multichannel identification problem into lower-dimensional sub-problems with lowered computational complexity and can be solved in parallel. Each sub-problem yields estimates for a subset of channel spectra, which then are combined into a consensus estimate per channel using general form consensus ADMM in an adaptive updating scheme. With numerical simulations, we show that it is possible to achieve estimation errors and convergence speeds comparable to low cost frequency-domain algorithms and high performing multichannel (Quasi-)Newton methods.

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

1. INTRODUCTION

The problem of blind system identification (BSI) has been subject of extensive research resulting in

2. PROBLEM STATEMENT

2.1. Signal Model

We define the acoustic SIMO system with the input signal $\mathbf{s}(n) = \begin{bmatrix} s(n) & s(k-1) & \dots & s(k-2L+2) \end{bmatrix}^{\mathrm{T}}$ and $i \in \mathcal{M}$ with $\mathcal{M} \triangleq \{1, \dots, M\}$ outputs $\mathbf{x}_i(n) = \begin{bmatrix} x_i(n) & x_i(k-1) & \dots & x_i(k-L+1) \end{bmatrix}^{\mathrm{T}}$. 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:

- The channel impulse responses have no common zeros, i.e. are not co-prime.
- The covariance matrix of the input signal $\mathbf{s}(n)$ is of full rank, i.e. the signal fully excites the channels.

The fundamental equality of this approach is

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

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 matrix $\mathbf{R}_{ij} = \mathrm{E}\left\{\mathbf{x}_i \mathbf{x}_j^{\mathrm{T}}\right\}$ instead of the signal vectors $\mathbf{x}_i$ or data matrices yields a more robust problem formulation. For the subsequent explanations the time index is dropped to make notation more compact.

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

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

*This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the SOUNDS European Training Network. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369.

†This work was supported in part by the Research Foundation - Flanders (FWO) grant 12ZD622N.

978-1-6654-6867-1/22/$31.00 ©2022 IEEE

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

This problem can also be written down in the frequency domain. The derivation is analogous to the time-domain one, using the frame-based overlap-save technique []. This gives the linear system of equations

$$\boldsymbol{\mathcal{R}}\underline{\boldsymbol{h}} = \mathbf{0} \tag{5}$$

where $\boldsymbol{\mathcal{R}}$ is analogous to (4), however the covariance matrices replaced by the cross-spectrum matrices

$$\boldsymbol{\mathcal{R}}_{ij} = \boldsymbol{\mathcal{S}}_i^{\mathrm{H}}(m)\boldsymbol{\mathcal{S}}_j(m) \tag{6}$$

with

$$\boldsymbol{\mathcal{S}}_i = \boldsymbol{\mathcal{W}}^{01}_{L\times 2L}\boldsymbol{\mathcal{D}}_i(m)\boldsymbol{\mathcal{W}}^{10}_{2L\times L} \tag{7}$$

where $\boldsymbol{\mathcal{D}}_i(m) = \mathrm{diag}\left\{\mathrm{FFT}_{2L}\left\{\mathbf{x}_{i,2L}(m)\right\}\right\}$ is a diagonal matrix of the signal spectrum and the overlap-save matrices

$$\mathbf{W}^{01}_{L\times 2L} = \begin{bmatrix}\mathbf{0}_{L\times L} & \mathbf{I}_{L\times L}\end{bmatrix} \tag{8}$$

$$\mathbf{W}^{10}_{2L\times L} = \begin{bmatrix}\mathbf{I}_{L\times L} & \mathbf{0}_{L\times L}\end{bmatrix}^{\mathrm{T}} \tag{9}$$

$$\boldsymbol{\mathcal{W}}^{01}_{2L\times L} = \mathbf{F}_{L\times L}\mathbf{W}^{01}_{L\times 2L}\mathbf{F}^{-1}_{2L\times 2L} \tag{10}$$

$$\boldsymbol{\mathcal{W}}^{10}_{2L\times L} = \mathbf{F}_{2L\times 2L}\mathbf{W}^{10}_{2L\times L}\mathbf{F}^{-1}_{L\times L} \tag{11}$$

where $\mathbf{F}$ as the DFT matrix for sizes L and 2L. The vector $\underline{\boldsymbol{h}}$ is the stacked vector of the complex-valued spectra of the impulse responses. In the presence of noise, the system of equations (5) is best solved by a minimization problem which seeks to minimize the squared error $\|\underline{\boldsymbol{e}}\|^2 = \|\boldsymbol{\mathcal{R}}\underline{\boldsymbol{h}}\|^2$ as

$$\text{minimize} \quad \underline{\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{\mathcal{R}}^{\mathrm{H}}\boldsymbol{\mathcal{R}}\underline{\boldsymbol{h}}, \tag{12}$$

$$\text{subject to} \quad \underline{\boldsymbol{h}}^{\mathrm{H}}\underline{\boldsymbol{h}} = a. \tag{13}$$

where the equality constraint is necessary to avoid the trivial zero solution. As this effectively seeks for the a-scaled eigenvector of the squared hermitian matrix $\boldsymbol{\mathcal{R}}^{\mathrm{H}}\boldsymbol{\mathcal{R}}$ corresponding to the smallest eigenvalue, we can avoid the squared matrix and replace it with its non-squared form $\boldsymbol{\mathcal{R}}$ as the eigenvectors in this case are the same. To compute the estimate of the spectra, we minimize the cost function

$$J(\underline{\boldsymbol{h}}) = \underline{\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{\mathcal{R}}\underline{\boldsymbol{h}} \tag{14}$$

as

$$\hat{\underline{\boldsymbol{h}}} = \arg\min_{\underline{\boldsymbol{h}}} \quad \underline{\boldsymbol{h}}^{\mathrm{H}}\boldsymbol{\mathcal{R}}\underline{\boldsymbol{h}}, \tag{15}$$

$$\text{s.t.} \quad \underline{\boldsymbol{h}}^{\mathrm{H}}\underline{\boldsymbol{h}} = a. \tag{16}$$

Attention: bit of rewriting necessary. not nice

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

Fig. 1: Problem splitting

3. 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 solution.

Attention: Show that!

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

3.1. Problem Splitting

In state-of-the-art algorithms, the minimization problem (15) is solved in its full form.

Note: State assumption that there as many nodes as there are channels/impulse earlier.

The problem is split into smaller sub-problems each defined by a subset of the full channel set $\mathcal{C}_i \subseteq \mathcal{M}$. Following from that, we define the sets $\bar{\mathcal{C}}_j = \{i | j \in \mathcal{C}_i\}$ for $i, j \in \mathcal{M}$ which is the set of sub-problems a channel j is part of. Further, $M_i = |\mathcal{C}_i|$ and $N_j = |\bar{\mathcal{C}}_j|$ with $i, j \in \mathcal{M}$. We replace the cost function (14) with the separable cost function

$$\tilde{J}(\underline{\boldsymbol{h}}) = \sum_{i=1}^{M}\tilde{J}_i(\underline{\boldsymbol{w}}_i) = \sum_{i=1}^{M}\underline{\boldsymbol{w}}_i^{\mathrm{T}}\boldsymbol{\mathcal{P}}_i\underline{\boldsymbol{w}}_i \tag{17}$$

where $\underline{\boldsymbol{w}}_i$ is defined as the stacked vector of estimated spectra analogous to $\underline{\boldsymbol{h}}$ (cf. subsection 2.2) but using the . Analogous, $\boldsymbol{\mathcal{P}}$ is constructed as defined in (4) with the reduced set of channels. In case the channel subsets are proper $\mathcal{C}_i \subset \mathcal{M}$, 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-eigenvalue-eigenvector comment be found?

3.2. 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 [], which is used to minimize a problem of form

$$\text{minimize} \quad \sum_{i=1}^{M} \tilde{J}_i(\underline{\boldsymbol{w}}_i) \tag{18}$$

$$\text{subject to} \quad (\underline{\boldsymbol{w}}_i)_j = \underline{\boldsymbol{h}}_{\mathcal{G}(i,j)}, \quad i \in \mathcal{M},\, j \in \mathcal{C}_i \tag{19}$$

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

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

The augmented Lagrangian, a real-valued function of a complex variable, for this particular general-form consensus problem is

$$\begin{aligned}\mathcal{L}_\rho(\underline{\boldsymbol{w}}, \underline{\boldsymbol{h}}, \underline{\boldsymbol{u}}) = \sum_{i\in\mathcal{M}} \Big(& \underline{\boldsymbol{w}}_i^{\mathrm{H}} \boldsymbol{\mathcal{P}}_i \underline{\boldsymbol{w}}_i + \underline{\boldsymbol{u}}_i^{\mathrm{H}} \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i\right) \\ & + \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i\right)^{\mathrm{H}} \underline{\boldsymbol{u}}_i + \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i\right)^{\mathrm{H}} \rho \mathbf{I} \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i\right)\Big).\end{aligned} \tag{20}$$

The ADMM then consists of the iterations

$$\begin{aligned}\underline{\boldsymbol{w}}_i^{k+1} = \underset{\underline{\boldsymbol{w}}_i}{\operatorname{argmin}} \Big\{ & \underline{\boldsymbol{w}}_i^{\mathrm{H}} \boldsymbol{\mathcal{P}}_i \underline{\boldsymbol{w}}_i + \underline{\boldsymbol{u}}_i^{k\,\mathrm{H}} \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i^k\right) \\ & + \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i^k\right)^{\mathrm{H}} \underline{\boldsymbol{u}}_i^k + \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i^k\right)^{\mathrm{H}} \rho \mathbf{I} \left(\underline{\boldsymbol{w}}_i - \tilde{\underline{\boldsymbol{h}}}_i^k\right)\Big\}\end{aligned} \tag{21}$$

$$\begin{aligned}\underline{\boldsymbol{h}}^{k+1} = \underset{\underline{\boldsymbol{h}}, \|\underline{\boldsymbol{h}}\|=a}{\operatorname{argmin}} \Big\{ \sum_{i\in\mathcal{M}} \Big(& \tilde{\underline{\boldsymbol{h}}}_i^{\mathrm{H}} \underline{\boldsymbol{u}}_i^k + \underline{\boldsymbol{u}}_i^{k\,\mathrm{H}} \tilde{\underline{\boldsymbol{h}}}_i \\ & + \left(\underline{\boldsymbol{w}}_i^{k+1} - \tilde{\underline{\boldsymbol{h}}}_i\right)^{\mathrm{H}} \rho \mathbf{I} \left(\underline{\boldsymbol{w}}_i^{k+1} - \tilde{\underline{\boldsymbol{h}}}_i\right)\Big)\Big\}\end{aligned} \tag{22}$$

$$\underline{\boldsymbol{u}}_i^{k+1} = \underline{\boldsymbol{u}}_i^k + \rho \left(\underline{\boldsymbol{w}}_i^{k+1} - \tilde{\underline{\boldsymbol{h}}}_i^{k+1}\right) \tag{23}$$

3.3. 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 (20) is time-dependent, which from here on will be denoted with the time index m in superscript $\underline{\boldsymbol{w}}_i^{\mathrm{H}} \boldsymbol{\mathcal{P}}_i^m \underline{\boldsymbol{w}}_i$.

The iterative form of the update steps as described in (21)-(23) can be transformed into an adaptive one by computing a (small) finite number of iterations at each time step m with the current data term. In this algorithm, one ADMM iteration is applied per time step, which simply allows us to replace the iteration index k with the time index m.

The minimization problem (21) can be solved by various methods, in this case however we perform a Newton update step of the form of

$$\underline{\boldsymbol{w}}_i^{m+1} = \underline{\boldsymbol{w}}_i^m - \mu \boldsymbol{\mathcal{V}}_i^m \left(\boldsymbol{\mathcal{P}}_i^m \underline{\boldsymbol{w}}_i^m + \underline{\boldsymbol{u}}_i^m + \rho\left(\underline{\boldsymbol{w}}_i^m - \tilde{\underline{\boldsymbol{h}}}_i^m\right)\right) \tag{24}$$

where $0 < \mu \leq 1$ is a step size and $\boldsymbol{\mathcal{V}}_i^m = (\boldsymbol{\mathcal{P}}_i^m + \rho \mathbf{I})^{-1}$ is the inverse Hessian of the problem. As this inverse is costly to compute, it is approximated by a diagonalized matrix

$$\tilde{\boldsymbol{\mathcal{V}}}_i^m = \operatorname{diag}\left\{(\operatorname{diag}\{\boldsymbol{\mathcal{P}}_i^m\} + \rho \mathbf{1})^{-1}\right\}, \tag{25}$$

similar to NMCFLMS [], which is straightforward to compute.

To compute the consensus $\underline{\boldsymbol{h}}$ we include the norm constraint $\|\underline{\boldsymbol{h}}\| = a$ with a Lagrange multiplier λ in the minimization problem and replace the average of penalties with a penalty of averages which takes the form of

Attention: poetic, cite our Boy[d]

$$\begin{aligned}\underline{\boldsymbol{h}}^{k+1} = \underset{\underline{\boldsymbol{h}}}{\operatorname{argmin}} \Big\{ & \frac{\lambda}{2}\left(\underline{\boldsymbol{h}}^{\mathrm{H}} \underline{\boldsymbol{h}} - a\right) \\ & + \frac{M\rho}{2} \left\| \underline{\boldsymbol{h}} - \bar{\underline{\boldsymbol{w}}}^{m+1} - \frac{1}{\rho} \bar{\mathbf{u}}^m \right\|^2 \Big\}\end{aligned} \tag{26}$$

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

$$(\bar{\underline{\boldsymbol{w}}}^{m+1})_g = \frac{1}{N_g} \sum_{\mathcal{G}(i,j)=g} (\underline{\boldsymbol{w}}_i^{m+1})_j, \quad g, i, j \in \mathcal{M}, \tag{27}$$

and

$$(\bar{\underline{\boldsymbol{u}}}^m)_g = \frac{1}{N_g} \sum_{\mathcal{G}(i,j)=g} (\underline{\boldsymbol{u}}_i^m)_j, \quad g, i, j \in \mathcal{M}. \tag{28}$$

Attention: UGLY AND CONFUSING??!!

Setting the derivative of (26) to zero and solving for the optimal Lagrange multiplier λ, we reach the update step

$$\underline{\boldsymbol{h}}^{k+1} = a \frac{\bar{\underline{\boldsymbol{w}}}^{m+1} + \frac{1}{\rho}\bar{\underline{\boldsymbol{u}}}^m}{\left\|\bar{\underline{\boldsymbol{w}}}^{m+1} + \frac{1}{\rho}\bar{\underline{\boldsymbol{u}}}^m\right\|}. \tag{29}$$

This results in a computationally inexpensive update step forcing the norm of the consensus to have value a.

4. NUMERICAL EVALUATION

First part: random impulse responses at different SNRs. Short like 64 tabs e.g. Error measure Normalized projection misalignment

$$\mathrm{NPM}(n) = 20 \log_{10}\left(\left\| \mathbf{h}(n) - \frac{\mathbf{h}^{\mathrm{T}}(n)\mathbf{h}_{\mathrm{t}}}{\mathbf{h}_{\mathrm{t}}^{\mathrm{T}}\mathbf{h}_{\mathrm{t}}} \right\| \cdot \|\mathbf{h}(n)\|^{-1}\right) \tag{30}$$

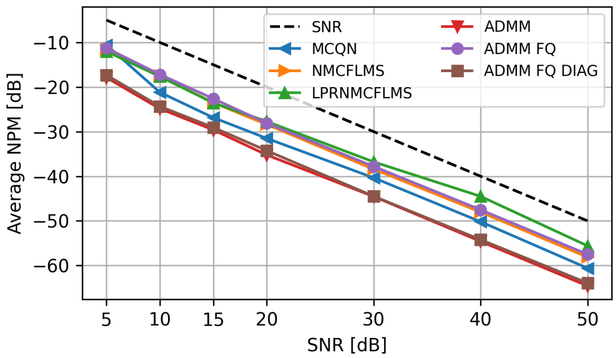

Fig. 2: NPM over SNR after convergence at 8000 samples signal length. $L = 16$. ADMM methods yield estimates as good or even better than the centralized Quasi Newton method which is the best of the rest.

where $(\mathbf{h})_i = \text{IFFT}\{(\underline{\boldsymbol{h}})_i\}$ with $i \in \mathcal{M}$ is the estimate of the stacked IRs and $\mathbf{h}_{\mathrm{t}}$ is the ground truth. Similar to [] the first experiment aims to evaluates the performance of the algorithm on randomly generated impulse responses under different SNRs using WGN as input signal and uncorrelated over channels as additive noise.

Attention: add SNR formula

The impulse responses of length $L = 16, 32, 64(?)$ are drawn from a Normal distribution with unit variance. The results are the median of 50 Monte-Carlo runs where the time-averaged NPM of the last 50 frames is taken as measurement.

Attention: Second part: Also short IRs now evaluated over different connectivity values (less to more connections in different sized graphs) Expectation: more nodes and more connections lead to faster convergence and better estimates (which is the case; plot measure for convergence speed). However it should show that even in not fully connected topology the estimates of ADMM method are as good as centralized methods.
Or if figured out, do longer simulated acoustic impulse responses.

5. CONCLUSIONS

LINE GO DONW

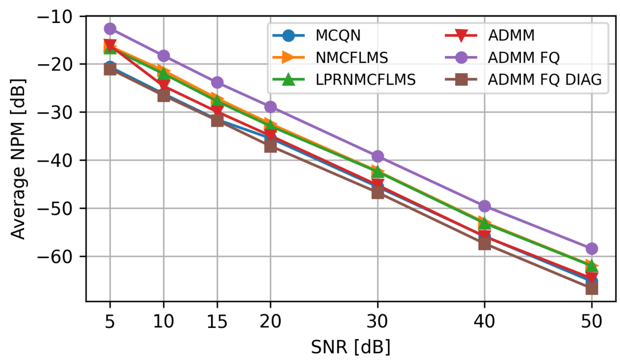

Fig. 3: NPM over SNR after convergence at 8000 samples signal length. $L = 16$ with 5 nodes and topology (DESCRIBE BY CONNECTION MATRIX OR FIGURE)