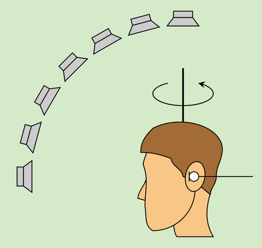
A Flexible Framework for Expectation Maximization-Based MIMO System Identification for Time-Variant Linear Acoustic Systems

Tobias Kabzinski and Peter Jax
Institute of Communication Systems (IKS)

1 Introduction



Measurement of thousands of impulse responses (IRs)

- Two approaches: static or **continuous**
- Example: HRTF measurements with rotating listener
- Unconstrained Expectation Maximization (EM) too complex in real world [1]

Core idea: impose **model structure** to reduce parameters and complexity

2 Signal Model

Problem: for acoustic MIMO system with T loudspeakers and R microphones, estimate $R{\cdot}T$ length-L IRs **offline** given inputs $x_t(k)$ and noisy outputs $y_r(k)$:

$$y_r(k) = \sum_{t=1}^{T} x_t(k) * h_{rt,k}(k) + \nu_r(k)\,, \quad r = 1, \dots, R$$

MIMO block observation model with block size N_o and $k = nN_o$

State-Space Model

$\mathbf{z}_n = [\mathbf{h}_{11,n}^{\mathrm{T}} \dots \mathbf{h}_{1T,n}^{\mathrm{T}} \dots \mathbf{h}_{RT,n}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{RTL}$ — state transition matrix $\mathbf{A}$ — process noise $\mathbf{q}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Gamma})$

$$\mathbf{z}_n = \mathbf{A}\mathbf{z}_{n-1} + \mathbf{q}_n \quad \text{(state equation)}$$

$$\mathbf{y}_n = \mathbf{C}_n\mathbf{z}_n + \boldsymbol{\nu}_n \quad \text{(observation equation)}$$

observation vector $\mathbf{y}_n = [\mathbf{y}_{1,n}^{\mathrm{T}} \dots \mathbf{y}_{R,n}^{\mathrm{T}}]^{\mathrm{T}} \in \mathbb{R}^{RN_o}$ with N_o samples of $y_r(k)$ per receiver

observation matrix $\mathbf{C}_n$ realizes convolution with input signals $x_t(k)$

observation noise $\boldsymbol{\nu}_n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$

3 EM-Based State and Parameter Estimation

Given observations $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$, jointly estimate states $\{\mathbf{z}_1, \dots, \mathbf{z}_N\}$ and parameters $\boldsymbol{\theta} = \{\mathbf{A}, \mathbf{\Gamma}, \mathbf{\Sigma}, \boldsymbol{\mu}_0, \mathbf{P}_0\}$ by maximizing expected log-likelihood [1, 2]

- **E-step**: maximum-likelihood estimates for state $\hat{\boldsymbol{\mu}}_n$ and error covariance $\hat{\mathbf{V}}_n$
  - Kalman filter recursion: $n = 1, \dots, N$

$$\mathbf{P}_{n-1} = \mathbf{A}\mathbf{V}_{n-1}\mathbf{A}^{\mathrm{T}} + \mathbf{\Gamma}, \qquad \boldsymbol{\mu}_n = \mathbf{A}\boldsymbol{\mu}_{n-1} + \mathbf{K}_n(\mathbf{y}_n - \mathbf{C}_n\mathbf{A}\boldsymbol{\mu}_{n-1})$$

$$\mathbf{K}_n = \mathbf{P}_{n-1}\mathbf{C}_n^{\mathrm{T}}\left(\mathbf{C}_n\mathbf{P}_{n-1}\mathbf{C}_n^{\mathrm{T}} + \mathbf{\Sigma}\right)^{-1}, \qquad \mathbf{V}_n = (\mathbf{I}_{N_z} - \mathbf{K}_n\mathbf{C}_n)\,\mathbf{P}_{n-1}$$

  - Kalman smoother recursion: $n = N, \dots, 1$

$$\mathbf{J}_n = \mathbf{V}_n\mathbf{A}^{\mathrm{T}}\mathbf{P}_n^{-1}, \quad \hat{\boldsymbol{\mu}}_n = \boldsymbol{\mu}_n + \mathbf{J}_n(\hat{\boldsymbol{\mu}}_{n+1} - \mathbf{A}\boldsymbol{\mu}_n), \quad \hat{\mathbf{V}}_n = \mathbf{V}_n + \mathbf{J}_n\left(\hat{\mathbf{V}}_{n+1} - \mathbf{P}_n\right)\mathbf{J}_n^{\mathrm{T}}$$

- **M-step**: parameter update rules use E-step results

$$\theta_i^\star \leftarrow f_i\left(\hat{\boldsymbol{\mu}}_1, \dots \hat{\boldsymbol{\mu}}_N, \hat{\mathbf{V}}_1, \dots \hat{\mathbf{V}}_N, \dots\right) \quad \text{for} \quad \theta_i \in \boldsymbol{\theta}$$

Excessive requirements for signal length $N_x = NN_o$ and state dim. $N_z = RTL$:

- Storage $\mathcal{O}(NN_z^2)$ and complexity[a]: $\mathcal{O}(NN_z^3)$

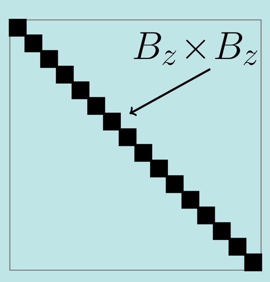
[a]dominant term for matrix-matrix multiplications

4 Specialized Model Structures

Ideas:

- Impose structure for parameters: scaled identity, diagonal, or block diagonal
  - Assumption: independence between particular states/observations
  - N_{B_z} state blocks of size B_z and N_{B_y} observation blocks of size B_y
- State transform $\mathbf{T}$, e.g., DFT, and permutation $\boldsymbol{\mathcal{P}}$: $\tilde{\mathbf{z}}_n = \boldsymbol{\mathcal{P}}(\mathbf{I}_{RT} \otimes \mathbf{T})\,\mathbf{z}_n$

Examples for block diagonal structure:



ID	coupling description	blocks N_{B_z}	B_z
1	full coupling between all states	1	RTL
2	R independent MISO systems	R	TL
8	fully independent coefficients	RTL	1
10	complex: within-receiver	$TL/2$	$2R$

Complete table → paper

Savings due to structure and block observation size N_o compared to [1]:

- Storage reduced by factor $\mathcal{O}(N_{B_z}N_o)$
- Complexity[a] reduced by factor $\mathcal{O}(N_{B_z}^2N_o)$

[a]dominant term for matrix-matrix multiplications

5 Example of M-Step Update Rule

Approach: set derivative of expected log-likelihood w.r.t. parameter to zero

- *Scaled identity* state transition matrix $\mathbf{A} = a\mathbf{I}$ with $a \in \mathbb{R}$ yields

$$a^\star = \operatorname{tr}\left\{\mathbf{\Gamma}^{-1}\sum_{n=2}^{N}\mathbb{E}\left\{\mathbf{z}_n\mathbf{z}_{n-1}^{\mathrm{T}}\right\}\right\} \Big/ \operatorname{tr}\left\{\mathbf{\Gamma}^{-1}\sum_{n=2}^{N}\mathbb{E}\left\{\mathbf{z}_{n-1}\mathbf{z}_{n-1}^{\mathrm{T}}\right\}\right\},$$

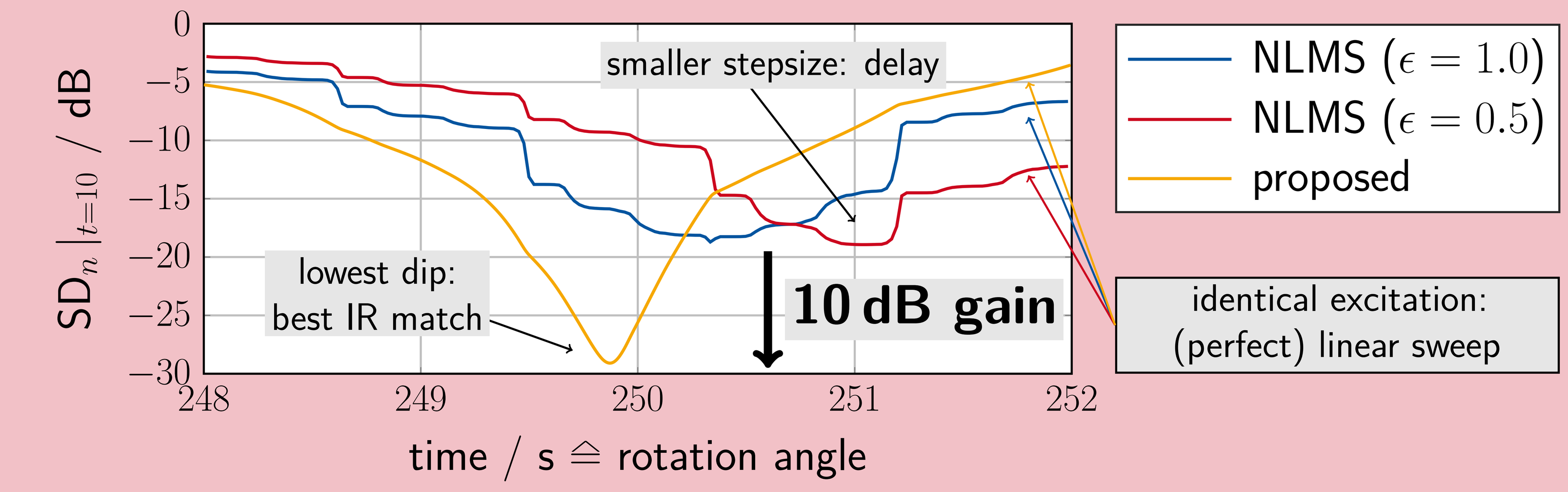
  corresponding to scalar fading factor in Kalman filtering literature
- E-step provides estimates for $\mathbb{E}\{\mathbf{z}_n\mathbf{z}_{n-1}^{\mathrm{T}}\}$ and $\mathbb{E}\{\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}\}$
- Complete set of update rules for $\mathbf{A}, \mathbf{\Gamma}, \mathbf{\Sigma}$ under various assumption → paper

6 Evaluation: Continuous HRTF Measurements

Compare each $\hat{\boldsymbol{\mu}}_n$ to *one* fixed-angle reference measurement $\mathbf{z}^{(\text{ref})}$:

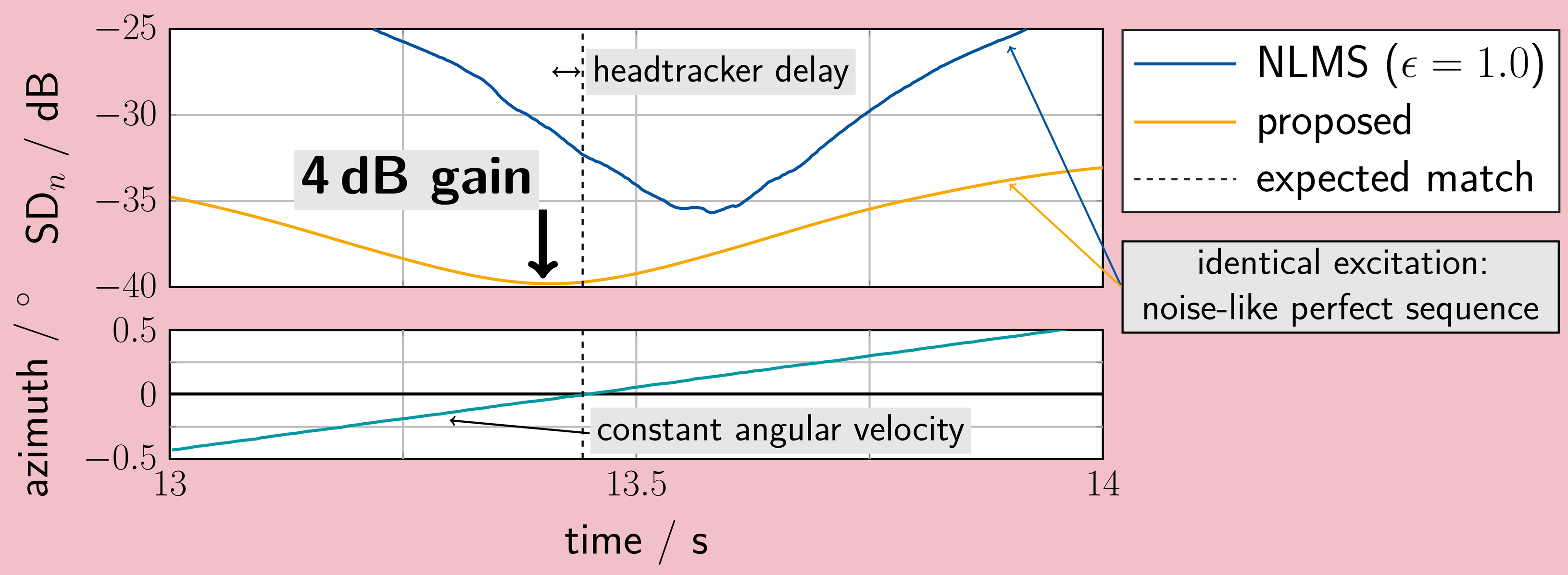
$$\text{SD}_n = 20\log_{10}\left(\frac{\left\|\hat{\boldsymbol{\mu}}_n - \mathbf{z}^{(\text{ref})}\right\|}{\left\|\mathbf{z}_n^{(\text{ref})}\right\|}\right)\text{dB} \qquad \text{(system distance)}$$

Example 1: HUTUBS Database [3]



- Anechoic setup: $R{=}1$, $T{=}37$, and $L{=}1024 \Rightarrow N_z{=}37888$
- To reduce complexity assume:
  - Independent coefficients (ID 8) for $\mathbf{A}$ and $\mathbf{\Gamma}$ in 4: $B_z = 1$
  - Block size $N_o = 256$ (≈ 5.8 ms)
  - ▶ storage reduced by factor 10^7, complexity reduced by factor 10^{11}

Example 2: Stereo Setup



- Semi-anechoic setup: $R{=}2$, $T{=}2$, and $L{=}4800 \Rightarrow N_z{=}19200$
- Assume coupling between frequency bins of receivers for $\mathbf{A}$ and $\mathbf{\Gamma}$:
  - ID 10 in 4 ($B_z = 4$), $\mathbf{T}$: DFT, and block size $N_o = 48$ (1 ms)
  - ▶ storage reduced by factor 10^5, complexity reduced by factor 10^9

7 Summary

▶ Flexible framework for improved EM-based offline MIMO system identification
▶ Choice of model structure and block sizes determines complexity
▶ Imposing model structure enables addressing wider range of applications

References

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Grant 509806277.

[1] T. Kabzinski and P. Jax, "Towards faster continuous multi-channel HRTF measurements based on learning system models," in *Proc. ICASSP 2022*, pp. 436–440.

[2] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[3] F. Brinkmann, M. Dinakaran, R. Pelzer, P. Grosche, D. Voss, and S. Weinzierl, "A cross-evaluated database of measured and simulated HRTFs including 3D head meshes, anthropometric features, and headphone impulse responses," *J. Audio Eng. Soc.*, vol. 67, no. 9, pp. 705–718, 2019.