Frobenius Norm-Based Alternating Least-Squares Approach for RETF Vector and Diffuse PSD Estimation

Marvin Tammen* Ina Kodrasi*† Simon Doclo*

* University of Oldenburg, Dept. Medical Physics and Acoustics and Cluster of Excellence Hearing4All, Germany

† Idiap Research Institute, Speech and Audio Processing Group, Switzerland

Problem Statement

- microphone signals corrupted by **reverberation and diffuse noise**
- multichannel Wiener filter (MWF) requires estimates of **relative early transfer functions** (RETFs) and **diffuse power spectral density** (PSD)
- many such estimators are **decoupled**
- goal: develop a **joint RETF and diffuse PSD estimator** under assumption of diffuse noise

Signal Model

- microphone signal model in STFT-domain:

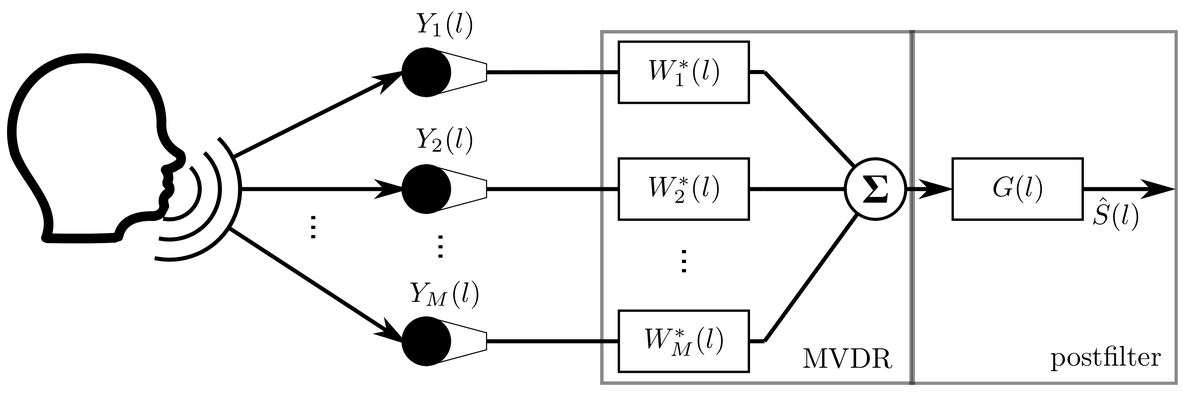
$$\mathbf{y}(l) = \mathbf{x}(l) + \mathbf{d}(l) \text{ with } \mathbf{y}(l) = \begin{bmatrix} Y_1(l) \\ \vdots \\ Y_M(l) \end{bmatrix}$$

  - M microphones, time frame l
  - independent processing in each sub-band
  - $\mathbf{x}(l)$: direct and early speech component (target)
  - $\mathbf{d}(l)$: diffuse noise and reverberation
- assumptions: $\mathbf{x}(l)$ and $\mathbf{d}(l)$ uncorrelated, stationary diffuse noise $\Rightarrow$ microphone PSD matrix:

$$\boldsymbol{\Phi}_{\boldsymbol{y}}(l) = \mathbb{E}\{\mathbf{y}(l)\mathbf{y}^H(l)\} = \phi_{\mathrm{s}}(l)\mathbf{a}(l)\mathbf{a}^H(l) + \phi_{\mathrm{d}}(l)\boldsymbol{\Gamma}$$

  - $\mathbf{a}(l)$: RETF vector, $\boldsymbol{\Gamma}$: diffuse coherence matrix
  - $\phi_{\mathrm{s}}(l), \phi_{\mathrm{d}}(l)$: target and diffuse PSD

Multichannel Wiener Filter



- two stages: minimum variance distortionless response **(MVDR) beamformer** and **single-channel postfilter**

$$\mathbf{w}_{\mathrm{MWF}}(l) = \underbrace{\left(\frac{\boldsymbol{\Gamma}^{-1}\mathbf{a}(l)}{\mathbf{a}^H(l)\boldsymbol{\Gamma}^{-1}\mathbf{a}(l)}\right)^H}_{\mathbf{w}_{\mathrm{MVDR}}(l)} \underbrace{\frac{\phi_{\mathrm{s}}(l)}{\phi_{\mathrm{s}}(l) + \phi_{\mathrm{d}}(l)/\left(\mathbf{a}^H(l)\boldsymbol{\Gamma}^{-1}\mathbf{a}(l)\right)}}_{G(l)}.$$

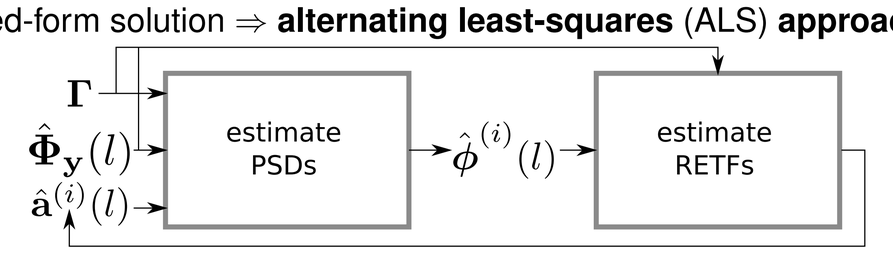
- **requires** estimates of **PSDs** $[\phi_{\mathrm{s}}(l), \phi_{\mathrm{d}}(l)]^T =: \boldsymbol{\phi}(l)$ **and RETFs** $\mathbf{a}(l)$

Proposed Method

- proposition: based on the **same cost function**, **jointly estimate PSDs and RETFs**:

$$\left(\hat{\mathbf{a}}_{\mathrm{LS}}, \hat{\boldsymbol{\phi}}_{\mathrm{LS}}\right) = \underset{\mathbf{a},\boldsymbol{\phi}}{\operatorname{argmin}} \left\| \hat{\boldsymbol{\Phi}}_{\mathbf{y}} - \left(\phi_{\mathrm{s}}\mathbf{a}\mathbf{a}^H + \phi_{\mathrm{d}}\boldsymbol{\Gamma}\right)\right\|_F^2 =: \underset{\mathbf{a},\boldsymbol{\phi}}{\operatorname{argmin}} \|\mathbf{E}\|_F^2$$

- no closed-form solution $\Rightarrow$ **alternating least-squares** (ALS) **approach**



(i) initialize RETFs $\hat{\mathbf{a}}_{\mathrm{LS}}^{(0)}$ (e.g., using DOA); set $i = 1$

(ii) estimate PSDs $\hat{\boldsymbol{\phi}}_{\mathrm{LS}}^{(i)}$ using (1), assuming $\mathbf{a} = \hat{\mathbf{a}}_{\mathrm{LS}}^{(i-1)}$

(iii) estimate RETFs $\hat{\mathbf{a}}_{\mathrm{LS}}^{(i)}$ using (2), assuming $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}_{\mathrm{LS}}^{(i)}$; set $i = i + 1$

(iv) repeat steps (ii) — (iii) until convergence ('**ALS iteration**')

Estimating PSDs and RETFs

- PSDs
  - assume $\mathbf{a} = \hat{\mathbf{a}}_{\mathrm{LS}}^{(i-1)} \Rightarrow$ estimate PSDs by minimizing Frobenius norm of error matrix $\mathbf{E}^{(i)} = \hat{\boldsymbol{\Phi}}_{\mathbf{y}} - \left(\hat{\phi}_{\mathrm{s,LS}}^{(i)}\hat{\mathbf{a}}_{\mathrm{LS}}^{(i-1)}\hat{\mathbf{a}}_{\mathrm{LS}}^{(i-1),H} + \hat{\phi}_{\mathrm{d,LS}}^{(i)}\boldsymbol{\Gamma}\right)$ [1]:

$$\hat{\boldsymbol{\phi}}_{\mathrm{LS}}^{(i)} = \underset{\boldsymbol{\phi}}{\operatorname{argmin}} \left\|\mathbf{E}^{(i)}\right\|_F^2 = \mathbf{A}^{-1}\mathbf{b}, \text{ where } \mathbf{A} = \begin{bmatrix} \left(\hat{\mathbf{a}}^H\hat{\mathbf{a}}\right)^2 & \hat{\mathbf{a}}^H\boldsymbol{\Gamma}\hat{\mathbf{a}} \\ \hat{\mathbf{a}}^H\boldsymbol{\Gamma}\hat{\mathbf{a}} & \mathrm{trace}\{\boldsymbol{\Gamma}^H\boldsymbol{\Gamma}\} \end{bmatrix}, \ \mathbf{b} = \begin{bmatrix} \mathrm{Re}\left\{\hat{\mathbf{a}}^H\hat{\boldsymbol{\Phi}}_{\mathbf{y}}\hat{\mathbf{a}}\right\} \\ \mathrm{Re}\{\mathrm{trace}\{\hat{\boldsymbol{\Phi}}_{\mathbf{y}}\boldsymbol{\Gamma}^H\}\} \end{bmatrix} \quad (1)$$

  - constrain PSDs to valid range: $0 \leq \{\hat{\phi}_{\mathrm{s,LS}}^{(i)}, \hat{\phi}_{\mathrm{d,LS}}^{(i)}\} \leq \frac{1}{M}\mathbf{y}^H\mathbf{y}$
- RETFs
  - assume $\boldsymbol{\phi} = \hat{\boldsymbol{\phi}}_{\mathrm{LS}}^{(i)}$
  - identify $\hat{\boldsymbol{\Phi}}_{\mathbf{y}} - \hat{\phi}_{\mathrm{d,LS}}^{(i)}\boldsymbol{\Gamma} =: \hat{\boldsymbol{\Phi}}_{\mathbf{x}}^{(i)}$

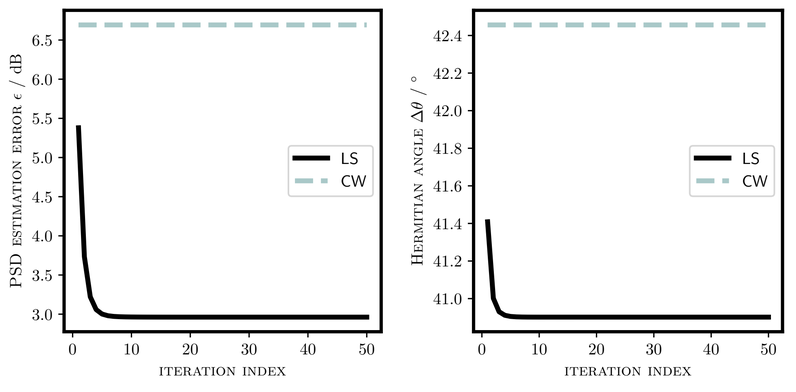
$$\Rightarrow \hat{\mathbf{a}}_{\mathrm{LS}}^{(i)} = \underset{\mathbf{a}}{\operatorname{argmin}} \left\|\hat{\boldsymbol{\Phi}}_{\mathbf{x}}^{(i)} - \hat{\phi}_{s,LS}^{(i)}\mathbf{a}\mathbf{a}^H\right\|_F^2 \triangleq \text{scaled rank-1 approx. of } \hat{\boldsymbol{\Phi}}_{\mathbf{x}}^{(i)} \text{ [2]}$$

$$= \sqrt{\frac{\lambda_1}{\hat{\phi}_{\mathrm{s,LS}}^{(i)}}}\mathbf{u}_1 \quad (2)$$

Simulations on Artificial Data

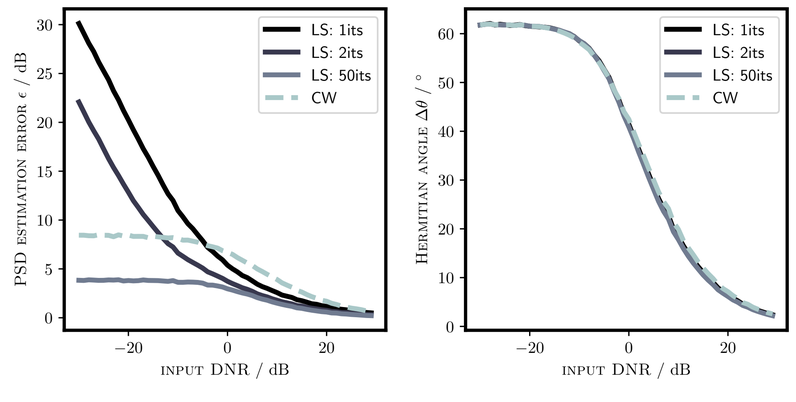
- **generated according to signal model** $\Rightarrow$ oracle information available
- **test robustness to model errors**: additional uncorrelated white noise
- evaluation measures: average diffuse PSD mismatch and Hermitian angle between estimate and oracle
- baseline: EVD-based PSD estimator combined with covariance whitening-based RETF estimator (denoted CW) [3]
- mismatches vs. number of iterations

more ALS iterations:
- no improved RETF estimate
- significantly smaller diffuse PSD mismatch



- mismatches vs. input DNR

- similar RETF estimation accuracy
- after convergence: overperformance at low DNR, same at high DNR



Simulations on Measured Data

- $M = 4$ microphones, $f_s = 16\,\mathrm{kHz}$
- **diffuse babble noise** [4] at 10 dB SNR wrt. reference microphone

	array geometry	mic. distance	θ	T_{60}
AS_1	linear	$d = 8\,\mathrm{cm}$	45°	0.61 s
AS_2	circular	$r = 10\,\mathrm{cm}$	45°	0.73 s
AS_3	linear	$d = 6\,\mathrm{cm}$	−15°	1.25 s

- additional **spatially uncorrelated white noise** at $\{0, 10, 20, 30\}$ dB DNR
- STFT: 64 ms frame length ($N_{\mathrm{FFT}} = 1024$), 16 ms shift
- $\boldsymbol{\Phi}_{\boldsymbol{y}}(l)$ estimated using recursive averaging, 40 ms smoothing constant
- target PSD $\phi_s(l)$ estimated using **decision-directed approach** [5] (LS estimate not used in postfilter)
- performance measures: **frequency-weighted segmental SNR, PESQ**

	ΔPESQ				ΔfwsSNR			
uncorrelated noise DNR / dB	0	10	20	30	0	10	20	30
CW	0.0547	0.1632	-0.0561	-0.1037	1.7112	2.2500	3.6697	3.4651
LS	**0.3823**	**0.5264**	**0.1818**	**0.2331**	**4.5457**	**4.9639**	**4.7897**	**3.7432**

References

[1] O. Schwartz, S. Gannot, and E. A. P. Habets. Joint estimation of late reverberant and speech power spectral densities in noisy environments using Frobenius norm. In *Proc. European Signal Processing Conference*, pages 1123–1127, Budapest, Hungary, September 2016.

[2] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1(3):211–218, 1936.

[3] I. Kodrasi and S. Doclo. EVD-based multi-channel dereverberation of a moving speaker using different RETF estimation methods. In *Proc. Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, pages 116–120, San Francisco, USA, March 2017.

[4] E. A. P. Habets, I. Cohen, and S. Gannot. Generating nonstationary multisensor signals under a spatial coherence constraint. *Journal of the Acoustical Society of America*, 124(5):2911–2917, November 2008.

[5] Y. Ephraim and D. Malah. Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 32(6):1109–1121, December 1984.