Learning Complex-valued Latent Filters with Absolute Cosine Similarity

NANYANG TECHNOLOGICAL UNIVERSITY

www.ntu.edu.sg/eee

Anh H.T. Nguyen, V. G. Reju, Andy W. H. Khong, and Soon Ing Yann

School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore

nguyenha001@e.ntu.edu.sg, {reju, andykhong, eiysoon}@ntu.edu.sg.

ICASSP2017, New Orleans, USA

Introduction

- Goal: *Under-determined convolutive* blind source separation
- Objective: Improve the accuracy of mixing matrix estimation
- Existing algorithms: Directional clustering and sparse coding
- Challenges: *Complex-valued mixing matrix* and non-convexity

Background

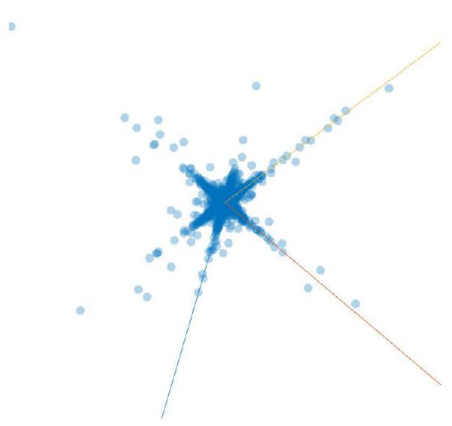
- Under-determined complex-valued instantaneous mixing model:

$$\mathbf{x}[k] = \mathbf{A}\mathbf{s}[k] = \sum_{j=1}^{N} \mathbf{a}_j s_j[k]$$

Observed complex-valued data — Unknown complex-valued filters — Unknown complex-valued sources

- **Assumption 1**: The sources are highly sparse so that the observed data concentrates around the directions specified by the columns of $\mathbf{A}$. E.g.,

$$\mathbf{x}[k] = \mathbf{A}\begin{bmatrix} s_1[k] \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix} \approx \mathbf{a}_1 s_1[k]$$



$\because s_1[k] \gg \epsilon_j$

There are *infinite number of unit vectors* that share the same direction in *complex* vector space. So the direction is measured using phase-invariant cosine distance:

$$D^2(\mathbf{x}[k], \mathbf{a}_1) = 1 - \cos^2\theta_H(\mathbf{x}[k], \mathbf{a}_1) = 1 - \frac{\left|\mathbf{a}_1^H \mathbf{x}[k]\right|^2}{\|\mathbf{a}_1\|_2^2 \,\|\mathbf{x}[k]\|_2^2} \approx 0$$

- **Assumption 2**: The sources are zero-mean and unit-variance. Thus, $\mathbf{A}$ becomes semi-unitary when $\mathbf{x}$[k] is whitened, i.e.,

$$\mathbf{A}\mathbf{A}^H = \mathbf{I}$$

In this case, for any $\mathbf{x}$[k], we always have

$$\sum_j \|\mathbf{a}_j\|_2^2 \cos^2\theta_H(\mathbf{x}[k], \mathbf{a}_j) = 1$$

Since $\mathbf{a}_j$ generally has different norm in under-determined case, sparsity penalty such as L1/L2 norm ratio is suboptimal for pre-whitened directional data.

Issues related to existing approaches

- Sparse filtering uses an unsuitable sparsity enforcer.

$$\min_{\hat{\mathbf{A}}} E\left\{\left\|\hat{\mathbf{A}}^H\mathbf{x}\right\|_1 / \left\|\hat{\mathbf{A}}^H\mathbf{x}\right\|_2\right\} \rightarrow \min_{\hat{\mathbf{A}}} E\left\{\sum_{j=1}^{N} \|\hat{\mathbf{a}}_j\|_2 \cos\theta_H(\mathbf{x}[k], \hat{\mathbf{a}}_j)\right\}$$

- K-hyperlines is only suitable for perfectly directional data.

$$\min_{\hat{\mathbf{A}}} E\left\{\min_{j=1,\ldots,N} D^2(\mathbf{x}[k], \mathbf{a}_j)\right\}$$

- "Soft" extensions of K-hyperlines are computationally expensive.
- Existing methods do not exploit the prior information of $\mathbf{A}$.

Proposed algorithm

- Proposed method: Minimize the expected "soft" minimum of phase-invariant cosine distance subject to semi-unitary constraint:

$$\min_{\hat{\mathbf{A}}} J(\hat{\mathbf{A}};\, r), \text{ s.t. } \hat{\mathbf{A}}\hat{\mathbf{A}}^H = \mathbf{I}_M.$$

where

$$J(\hat{\mathbf{A}};\, r) = E\left\{\left[\frac{1}{N}\sum_{j=1}^{N}\left(D^2(\mathbf{x}[k], \hat{\mathbf{a}}_j)\right)^r\right]^{1/r}\right\},\ r \in (-\infty, 1).$$

- Difficulty: Constrained non-convex optimization problem.
- Solution: Reparametrize semi-unitary constrained problems into *unconstrained ones in Euclidean space* that can be solved by any off-the-shelf tools such as L-BFGS, Nesterov's accelerated gradient, SGD, momentum, ADAM, etc.

$$\min_{\hat{\mathbf{A}}} f(\hat{\mathbf{A}}), \text{ s.t. } \hat{\mathbf{A}}\hat{\mathbf{A}}^H = \mathbf{I}_M \Rightarrow \min_{\mathbf{B}} f(\hat{\mathbf{A}}) \text{ s.t. } \hat{\mathbf{A}} = (\mathbf{B}\mathbf{B}^H)^{-1/2}\mathbf{B}$$

- $\hat{\mathbf{A}}$ is the nearest semi-unitary matrix of $\mathbf{B}$, thus always feasible
- The gradient w.r.t. $\boldsymbol{B}^*$ is parallel to the tangent space at $\hat{\mathbf{A}}$ since

all matrix belonged to normal space at $\hat{\mathbf{A}}$ yield the same cost

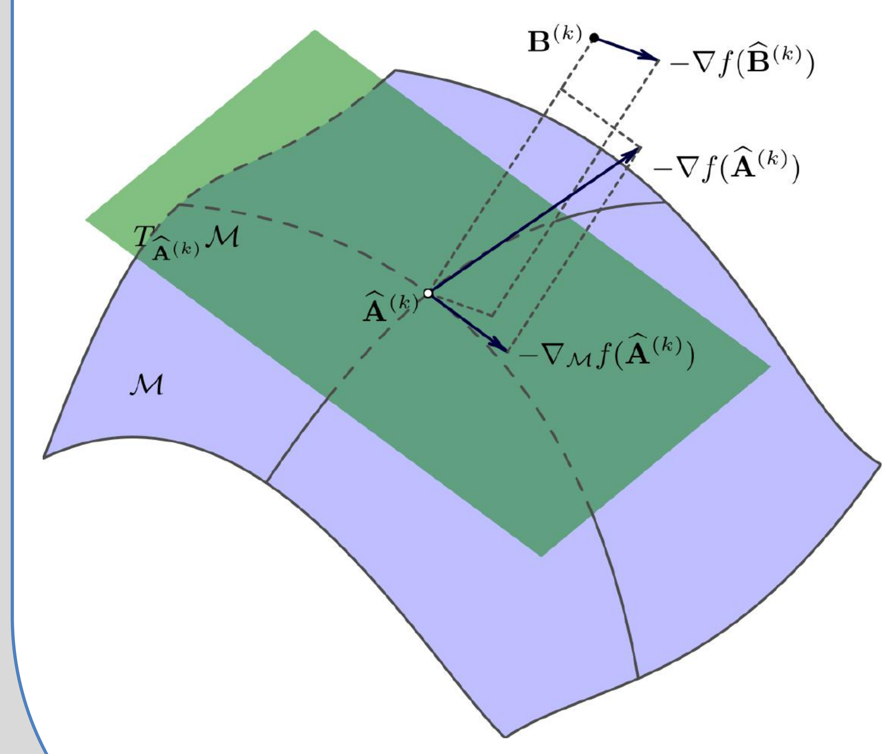


Fig.: Proposed reparameterization.

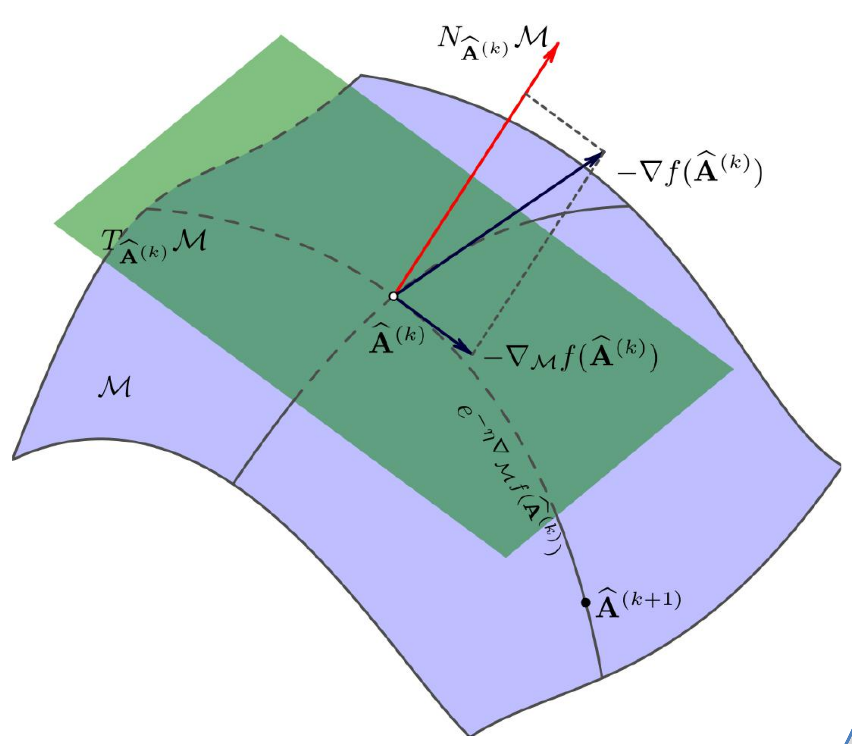


Fig.: Optimization on Stiefel manifold.

- The gradient w.r.t. $\boldsymbol{B}^*$ can be evaluated efficiently via eSVD

Algorithm 1 Gradient of in-line row-wise decoupling scheme

1: $\mathbf{U},\ \boldsymbol{\Sigma},\ \mathbf{V} \leftarrow \mathrm{SVD}(\mathbf{B})$
2: $\boldsymbol{\sigma} \leftarrow \mathrm{diag}(\boldsymbol{\Sigma})$
3: $\hat{\mathbf{A}} \leftarrow \mathbf{U}\mathbf{V}^H$
4: Find f and $\nabla_{\hat{\mathbf{A}}^*} f$ for a batch or minibatch
5: $\mathbf{C} \leftarrow -(\boldsymbol{\Sigma}^{-1}\mathbf{U}^H(\nabla_{\hat{\mathbf{A}}^*} f)\mathbf{V}) \oslash (\mathbf{1}\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}^T)$
6: $\nabla_{\mathbf{B}^*} f \leftarrow \mathbf{U}(\mathbf{C}^H + \mathbf{C})\boldsymbol{\Sigma}\mathbf{V}^H + \mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\nabla_{\hat{\mathbf{A}}^*} f$

- Applicable to *other semi-unitary/unitary constrained problems* such as ICA, orthogonal sparse PCA, unitary RNN, unitary beamforming, quadratic assignment problem, etc.

Simulation results

- We compare the performance of our proposed method (PM) to other techniques such as K-hyperlines (KHL), Gaussian mixture model (GMM), and sparse filtering (SF).

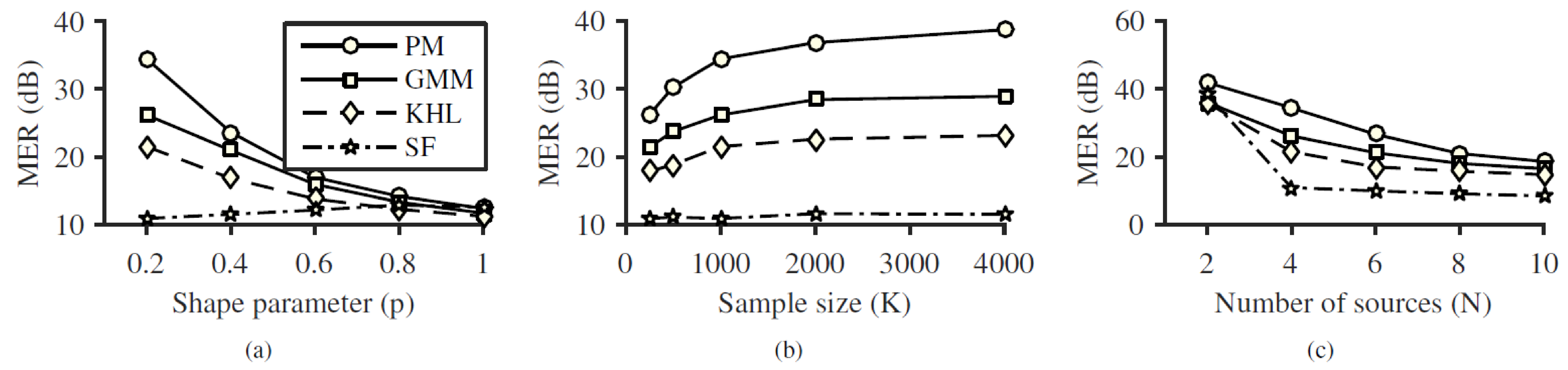


Fig. 1: Average MER in estimation of 2×4 mixing matrix w.r.t. : a) Sparseness. b) Sample size. c) Number of sources

- Application in blind separation of convolutive speech mixtures

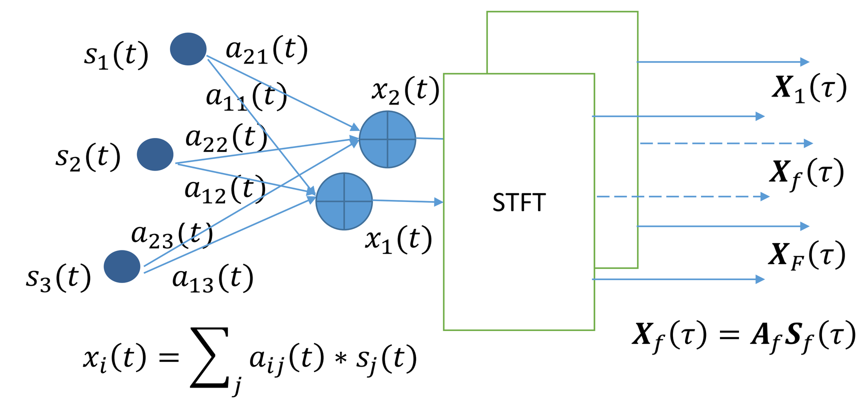


Table 1: Output SDR and SIR in dB for 2mic_4src_5cm subset of SiSEC dev1 dataset

RT60	130ms				250ms			
Source	4 males		4 females		4 males		4 females	
Perf. metric	SDR	SIR	SDR	SIR	SDR	SIR	SDR	SIR
PM	**4.55**	**8.27**	3.80	6.38	**3.67**	**6.06**	**3.57**	**5.36**
[22]	4.1	6.38	**4.47**	**6.48**	3.55	5.07	3.5	4.85
[9]	3.31	-	3.92	-	2.62	-	3.49	-
Input	-4.81	-4.60	-4.76	-4.68	-4.79	-4.64	-4.83	-4.71

- Improvement in SIR is 14% on SISEC dev1 dataset compared to the state of the art in [22].