Structure from Sound with Incomplete Data

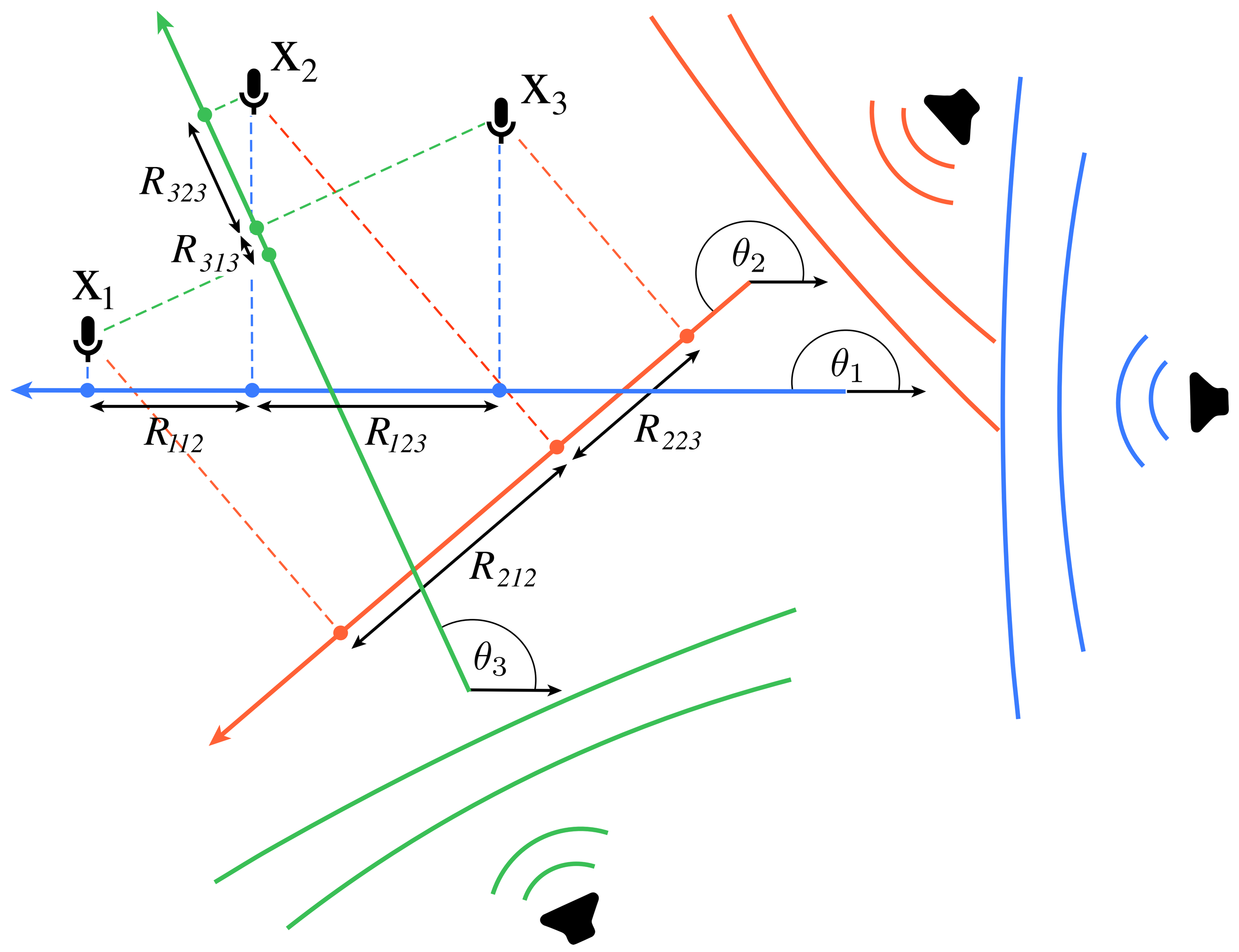
Miranda Kreković, Gilles Baechler, Ivan Dokmanić, and Martin Vetterli

LCAV - School of Computer and Communication Sciences - EPFL - Lausanne, Switzerland

Problem statement

Affine structure from sound [1]

- Joint localization of a microphone array and loudspeakers
- Far field assumption: plane accoustic waves



Problem formulation

$$\boldsymbol{D} = \boldsymbol{A}^\top \boldsymbol{X} + \boldsymbol{c}\boldsymbol{1}^\top$$

where

1. $\boldsymbol{D} \in \mathbb{R}^{K\times N}$ is the measurement matrix
2. $\boldsymbol{A} \in \mathbb{R}^{2\times K}$ is the projection matrix
$$\boldsymbol{A} = \begin{bmatrix} \cos\theta_1 & \ldots & \cos\theta_K \\ \sin\theta_1 & \ldots & \sin\theta_K \end{bmatrix}$$
3. $\boldsymbol{X} \in \mathbb{R}^{2\times N}$ is the coordinate matrix
$$\boldsymbol{X} = \begin{bmatrix} x_1 & \ldots & x_N \\ y_1 & \ldots & y_N \end{bmatrix}$$
4. $\boldsymbol{c} \in \mathbb{R}^{K}$ is the offset vector
$$\boldsymbol{c} = \begin{bmatrix} c_1 & \ldots & c_K \end{bmatrix}^\top$$

Incomplete and noisy $\boldsymbol{D}$

$$\widetilde{\boldsymbol{D}} = \boldsymbol{W} \circ (\boldsymbol{D} + \boldsymbol{Z})$$

where $\boldsymbol{W} \in \mathbb{R}^{K\times N}$ is a binary mask matrix

and $\boldsymbol{Z} \in \mathbb{R}^{K\times N}$ contains independent noise realizations

This work has been supported by the SNF grant: Inverse Problems Regularized by Sparsity, no. 20FP21151073
Ivan Dokmanic was supported by a Google Faculty Research Award.

Algorithms for missing entries

Acces to one full column in $\boldsymbol{D}$

- Subtract it from $\boldsymbol{D}$ to remove the influence of $\boldsymbol{c}$
- If $\boldsymbol{D}$ complete, factor it using SVD [1] or perform the alternating optimization [2] to estimate the unknown matrices

No full column in $\boldsymbol{D}$

- **Our solution:** consider pairwise differences between columns of $\boldsymbol{D}$

Our algorithm

Goal

Given a subset of noisy observations of distances $\widetilde{\boldsymbol{D}}$, jointly recover the points $\boldsymbol{X}$, the column-unitary matrix $\boldsymbol{A}$, and the translation vector $\boldsymbol{c}$, such that

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{A}}, \hat{\boldsymbol{c}} = \arg\min_{\boldsymbol{X},\boldsymbol{A}\in\mathcal{U}_c,\boldsymbol{c}} \left\| \widetilde{\boldsymbol{D}} - \boldsymbol{W} \circ \left(\boldsymbol{A}^\top \boldsymbol{X} + \boldsymbol{c}\boldsymbol{1}^\top\right) \right\|^2$$

Our formulation

1. Eliminate $\boldsymbol{c}$ by observing the differences between the columns of $\widetilde{\boldsymbol{D}}$
2. Measurement tensor $\widetilde{\boldsymbol{R}} \in \mathbb{R}^{K\times N\times N}$ with relative distances:
$$\widetilde{R}_{knm} = \widetilde{D}_{kn} - \widetilde{D}_{km}$$
3. Generalized mask $\boldsymbol{V} \in \mathbb{R}^{K\times N\times N}$:
$$V_{knm} = W_{kn} W_{km}$$
4. Reformulate the optimization problem as (*)

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{X},\boldsymbol{\theta}} \sum_{k=1}^{K} \sum_{n,m=1}^{N} \left(\widetilde{R}_{knm} - V_{knm}(\Delta_{x_{nm}} \cos\theta_k + \Delta_{y_{nm}} \sin\theta_k)\right)^2 \quad (*)$$

where $\Delta_{x_{nm}} = x_n - x_m$ and $\Delta_{y_{nm}} = y_n - y_m$.

Proposed algorithm:

Alternate between the estimates $\hat{\boldsymbol{X}}$ and $\hat{\boldsymbol{\theta}}$

- Find the global optimizer of (*) over $\boldsymbol{\theta}$ for fixed $\boldsymbol{X}$
- Find the global optimizer of (*) over $\boldsymbol{X}$ for fixed $\boldsymbol{\theta}$

Extension to 3D space

- Replace the polar coordinates of $\boldsymbol{A}$ with the spherical respresentation
- Add the third row to $\boldsymbol{X}$ corresponding to z-coordinates

$$\hat{\boldsymbol{X}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{X},\boldsymbol{\theta},\boldsymbol{\phi}} \sum_{k=1} \sum_{n,m=1} \widetilde{R}_{knm} - V_{knm}\big(\Delta_{x_{nm}} \cos\theta_k \sin\phi_k - \Delta_{y_{nm}} \sin\theta_k \sin\phi_k - \Delta_{z_{nm}} \cos\phi_k\big)^2$$

- For fixed $\boldsymbol{X}$, minimize over $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ jointly

{miranda.krekovic , gilles.baechler}@epfl.ch web: lcav.epfl.ch

Experimental results

Comparison of the proposed algorithm with

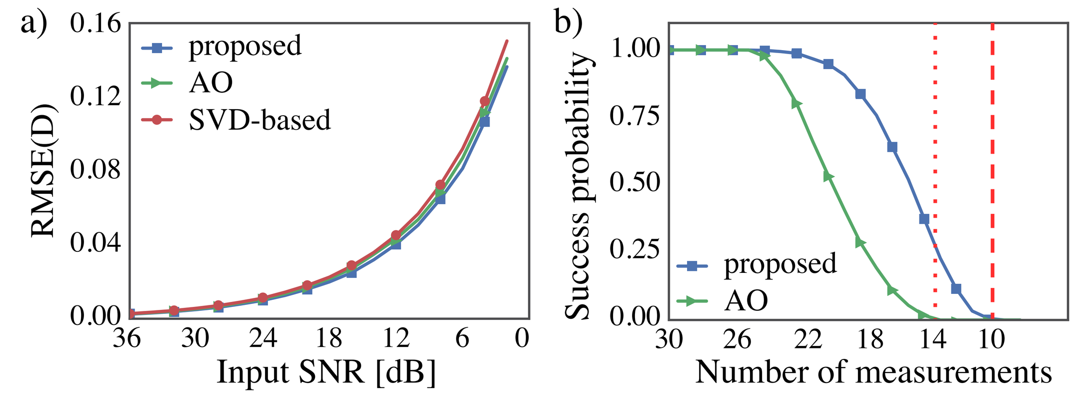
- the SVD-based estimator [1]
- the alternating optimization (AO) [2]

Peformance measurements:

- RMSE($\boldsymbol{D}$) (consistency) and RMSE($\boldsymbol{X}$) (accuracy)

Setup: $N = 6$ microphones, $K = 5$ acoustic events

Complete case: complete $\boldsymbol{D}$ corrupted with Gaussian noise



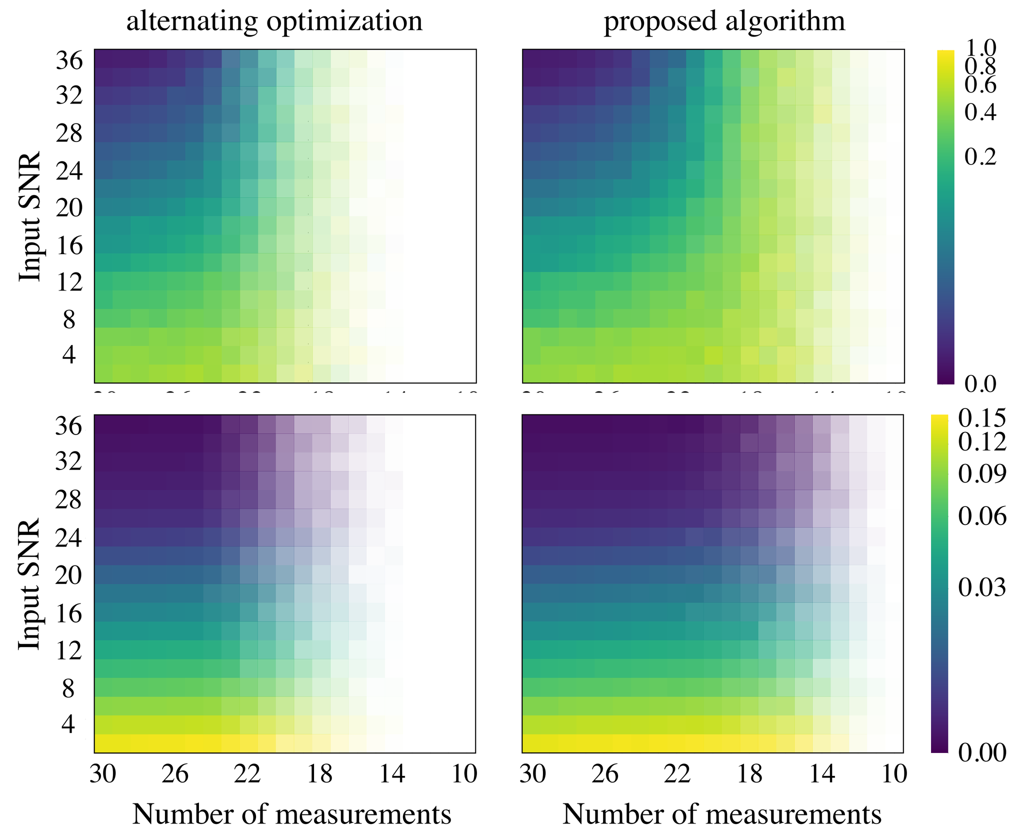
a) RMSE($\boldsymbol{D}$) vs input SNR

b) Likelihood of the algorithms to work for a randomly created mask $\boldsymbol{W}$ with a given number of measurements

Missing entries: incomplete conneced $\widetilde{\boldsymbol{D}}$ corrupted with Gaussian noise

a) RMSE($\boldsymbol{D}$)

b) RMSE($\boldsymbol{X}$)



The likelihood of the algorithms to work is depicted by transparency.

Conclusion

A novel algorithm for structure from sound with incomplete data which

- outperforms existing solutions
- allows larger number of missing entries

REFERENCES

[1] S. Thrun, "Affine structure from sound," Advances in Neural Information Processing Systems, pp. 1353–1360, 2006.
[2] Kuang, E. Ask, S. Burgess, and K. Åström, "Understanding TOA and TDOA network calibration using far field approximation as initial estimate."Int. Conf. on Pattern Recognition Applications and Methods, pp. 590–596, 2012.