Master Thesis Presentation
21 February 2023

Generative Adversarial Networks for the Generation of Microphone Array Data

Gaspard Ulysse Fragnière

Abstract

This thesis investigates whether Generative Adversarial Network can be used to generate realistic Cross Spectral Matrix (CSM) through their eigendecomposition. Models to generate the eigenvalues' spectrum as well the strongest eigenvector are proposed. The GAN model used to generate those quantities is a Wasserstein GAN with penalized norm of gradient of the critic with respect to its input (WGAN-GP). The method shows that WGAN-GP are suited to generate the eigenvalues spectrum and the strongest eigenvector. Based on those models, a data augmentation scheme allowing to improve the realness of synthetic CSM is introduced. Moreover, this work shows that only a few measured source cases are needed in order to generate data with properties similar to experimental data. Based on the findings of this work, GAN could be a promising tool to achieve better generalization of deep learning models for source characterization.

Agenda

- Introduction
- Some Fundamentals
- Methods
- Results and Discussion
- Conclusion and Future Works

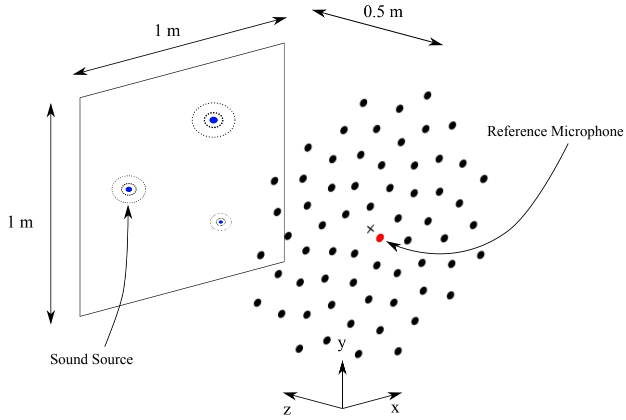
Introduction: Context

- We consider the problem of source characterization and more specifically how to solve it using DL-based approaches.
- Unfortunately, DL-based algorithm needs significative quantities of well-structured data to be trained with.
- Currently data is either real measurement, synthetic data or semi-synthetic data
- Therefore new ways to obtain realistic data swiftly are needed.

Introduction: Aim

- First, it is relevant to note that the data used for source characterization in states of the art papers is the Cross Spectral Matrix (CSM) (e.g. in [1], [2], [3], [4]).
- The CSM is a direct representation of the signals received in the array of microphone.
- It means that recording, simulating or generating raw microphone data is not necessary, if realistic CSM could be generated directly.
- Hence the goal of this thesis is to investigate the generation of the CSM directly or indirectly, as realistically as possible, using a GAN approach.

Fundamentals: Sound Model



- The sound model equation is given by:

$$\mathbf{p} = \mathbf{H}\mathbf{q} + \mathbf{n} \tag{1}$$

where $\mathbf{p} \in \mathbb{C}^{64}$ is the pressure in the 64 microphones of the array, $\mathbf{q}$ the J uncorrelated sources and $\mathbf{H} \in \mathbb{C}^{(64,J)}$ is the transfer function from the sources to the sensor. In this thesis, the case of a single source is considered. $\mathbf{n}$ models independent noise.

Fundamentals: CSM, Eigendecomposition and Rank I CSM

- The Cross Spectral Matrix (CSM) is a representation of the sound pressure in the different microphones. Using Welch's method, it can be approximated as:

$$\hat{\mathbf{C}} = \frac{1}{B}\sum_{b=1}^{B} \mathbf{p}\mathbf{p}^H \tag{2}$$

- A Cross spectral matrix $\hat{\mathbf{C}}$ of dimension $M \times M$ can be representated by its eigenvalues and eigenvectors using:

$$\hat{\mathbf{C}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H \tag{3}$$

- Using only one eigenvector $\mathbf{v}_i \in \mathbf{V}$ and the corresponding eigenvalue $\mathbf{\Lambda}_{ii}$, the Rank I CSM can be computed as:

$$\hat{\mathbf{C}}_i = \mathbf{v}_i\mathbf{\Lambda}_{ii}\mathbf{v}_i^H \tag{4}$$

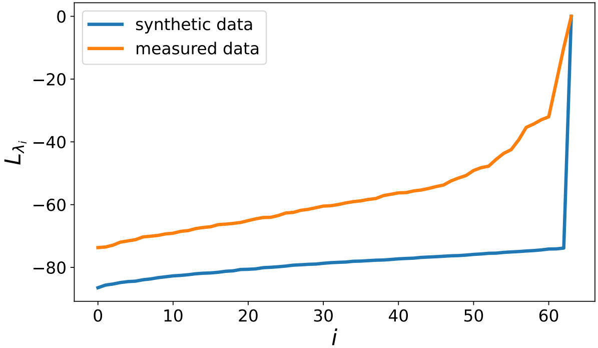
Fundamentals:GAN and WGAN-GP

- A Generative Adversial Network (GAN) is a network designed to generate realistic fake sample of some data it has been trained with.
- It consists of two networks competing against each other, namely a generator and a discriminator. The goal of the discriminator is to determine real from fake samples and the aim of the generator is to produce data realistic enough that the discriminator cannot determine that it has been generated.
- Both those inner-models are trained simultaneously. During training they both become gradually better at their tasks, until the generator becomes able to produce sufficiently realistic samples.
- The Wasserstein GAN with Gradient Penalty (WGAN-GP) is an improved version of the GAN. This version uses the Wasserstein distance as loss function, which makes WGAN-GPs better at learning distribution of training data than GANs.

Methods: Introduction

- This thesis investiguates the generation of CSMs through their eigendecomposition
- This approach was chosen since CSMs are complex and hermitian matrices, it is complicated for a network to learn their distribution directly.
- Since there is a lack of available data, the idea is to train the generative models first using Synthetic Data and then fine-tune them using real measurement.

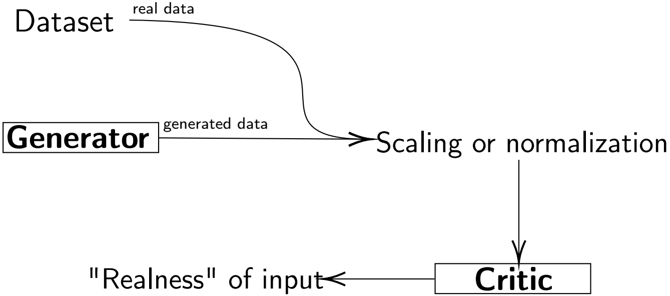
Methods: Data



- The synthetic data has been sampled from a Wishart distribution
- The measurement were performed in an anechoic chamber
- Both the synthetic data corresponds to the same position and same Helmotz number $He = 16$
- The decay of the eigenvalues of synthetic data and measured data is not similar.

Methods: Generating Eigenvalues

The first approach consisted in generating the scaled eigenvalues $[\lambda_0, \dots, \lambda_{63}] \in]0,1]^{64}$, with $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{63}$. This approach allowed to scale generated eigenavalues, before feeding them to the discriminator, in order to improve performances:



Methods: Generating Eigenvalues

- The second approach to generate the eigenvalues consisted in generating their level representation $L_{\lambda_0}, \ldots, L_{\lambda_{63}}$ defined as:

$$L_{\lambda_i} = 10 \log_{10}(\frac{\lambda_i}{\lambda_{63}}) \tag{5}$$

- Since all values L_{λ_i} are non-positive, the generator has been built such that the two last steps are first a ReLU, followed by a multiplication by -1.

- Because Leaky ReLU were used in the critic, the first layer of its network is also a multiplicative layer.

Methods: Generating Eigenvectors

- In order to generate the eigenvector, it was decided to start by only generating the strongest eigenvector, as a proof of concept. We defined this eigenvector as main eigenvector.
- Generating only the main eigenvector can be justified by the fact, it is the most meaningful eigenvector, since it contains all the information about the position of the source.
- Since the main eigenvector is $\in \mathbb{C}^{64}$, a network with a similar architecture as the one used for the eigenvalues can be used. But instead of scaling the generated sample, they are normalized before being fed to the discriminator.

Methods: Data Augmentation

For a received synthetic CSM $\hat{\mathbf{C}}$, its eigendecomposition $\hat{\mathbf{C}} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ is computed. The CSM $\hat{\mathbf{C}}$ can then be modified with

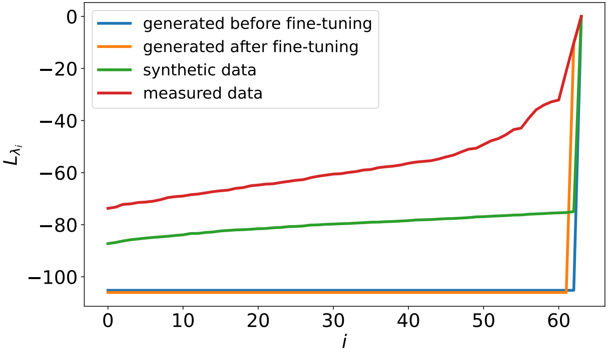
- generated eigenvalues $\hat{\mathbf{\Lambda}}$, with:

$$\hat{\mathbf{C}}_{\text{augm.}} = \mathbf{V}\hat{\mathbf{\Lambda}}\mathbf{V}^H \tag{6}$$

- a generated main eigenvector $\hat{\mathbf{v}}_M$ and corresponding semi-generated eigenvectors matrix $\hat{\mathbf{V}} = [\mathbf{v}_1^T, \dots, \hat{\mathbf{v}}_M^T]$, with:

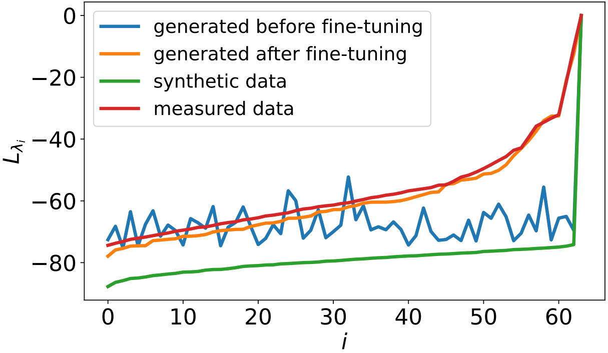
$$\hat{\mathbf{C}}_{\text{augm.}} = \hat{\mathbf{V}}\mathbf{\Lambda}\hat{\mathbf{V}}^H \tag{7}$$

Results and Discussion: Generating eigenvalues from scaled values



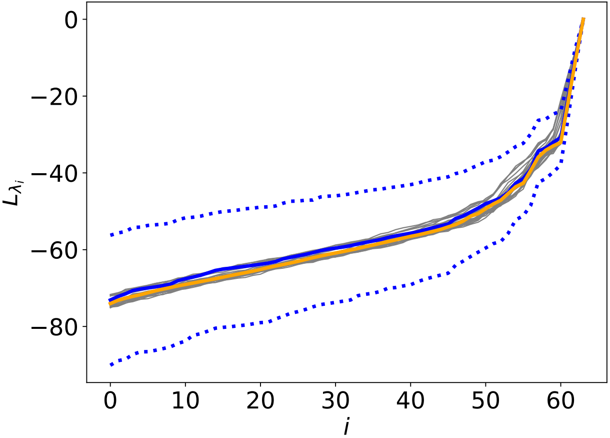
- Sudden and steep drop in value, reaching a level around 10^{-100}
- Both before and after fine-tuning
- the WGAN-GP cannot capture very well small numerical variation
- This approach is not suited.

Results and Discussion: Generating eigenvalues from levels



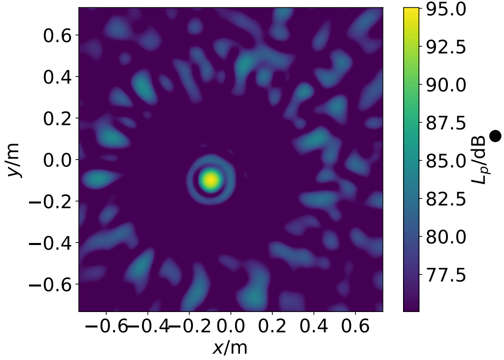
- The generated data looks more like the real data, especially after fine-tuning.
- We conclude that generating the eigenvalues from their spectrum is the right way

Results and Discussion: Generating eigenvalues from levels 2



- Moreover, it was observed that training the network first with synthetic data and then fine-tuning it with a single measurement allows to produce eigenvalues with a large variation.
- Hence, it can be concluded that this approach allows to generate eigenvalues samples that are representative not only of a single measurement, but of multiple ones

Results and Discussion: Generating strongest eigenvector



- When observing the beamforming map computed from generated main eigenvector, it can be concluded that WGAN-GP is able to generate sufficiently realistic sample.
- It was observed that all the generated main eigenvector were atually pretty similar. It was therefore concluded that the WGAN-GP cannot produce a great variety of sample, but it is important to note that it is representative of the data it was trained with.

Results and Discussion: Data Augmentation

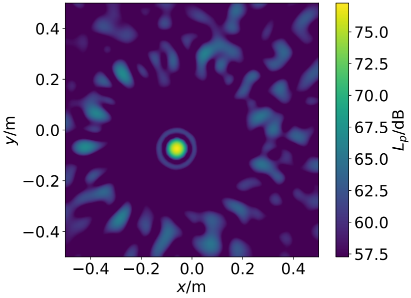


Figure 1: Synthetic

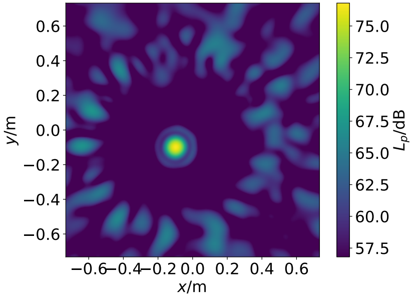
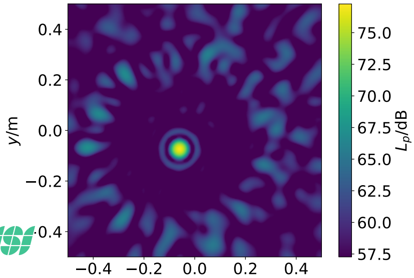
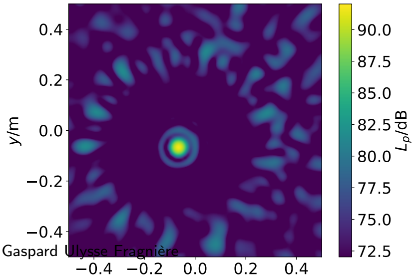
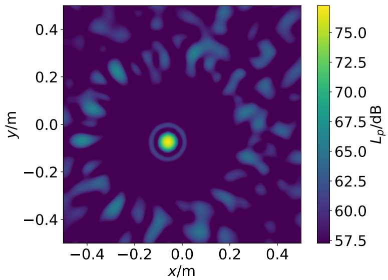


Figure 2: Measurement

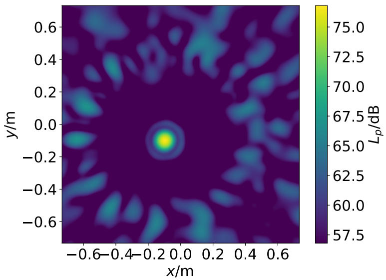




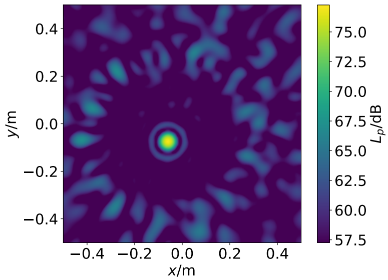
Results and Discussion: Data Augmentation



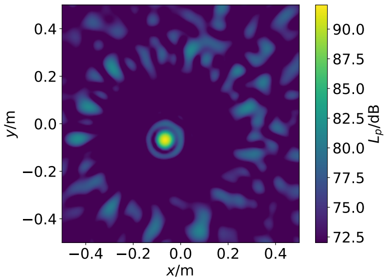
(a) Synthetic



(b) Measurement



(c) Augmented with Eigenvalues



(d) Augmented with Eigenvectors

Conclusion

- It has been shown that WGAN-GPs are suited both to generate eigenvalues and strongest eigenvector.
- Moreover, the data augmentation method using generated main eigenvector allows to improve how realistic synthetic CSM are.
- In conclusion, it can be stated that this thesis introduces new methods for learning matrix data, through eigendecomposition. It is shown that the eigedecomposition is a good representation of CSM to learn their distribution and hence generating them.
- The finding in this thesis are an important step toward solving the unavailability of real training data for source localization or characterization. Indeed, as shown in the literature survey, extensive researchs on this topic do not exist, which makes the contribution of this thesis especially relevant.

Future Works

- Extending the work to be able to generate CSM for different positions.
- More specifically, this would need to be done for the network to generate the main eigenvector. Indeed, the eigenvalue spectrum is not position dependent.
- More than simply training the eigenvectors for multiple positions, it should be investigated whether a Neural Network could be designed for generating eigenvectors also corresponding to a given source position not observed in the dataset.
- It also need to be investigated how to generate the remaining weakest eigenvectors. Extending the approach to generate the strongest eigenvector to all eigenvectors, has been tried but without compelling results.

Questions?

Bibliography

References

[1] P. Castellini, N. Giulietti, N. Falcionelli, A. F. Dragoni, and P. Chiariotti, "A neural network based microphone array approach to grid-less noise source localization," *Applied Acoustics*, vol. 177, p. 107947, 2021.

[2] S. Y. Lee, J. Chang, and S. Lee, "Deep learning-based method for multiple sound source localization with high resolution and accuracy," *Mechanical Systems and Signal Processing*, vol. 161, p. 107959, 2021.

[3] W. Ma and X. Liu, "Phased microphone array for sound source localization with deep learning," *Aerospace Systems*, vol. 2, no. 2, pp. 71–81, 2019.

[4] P. Xu, E. J. Arcondoulis, and Y. Liu, "Deep neural network models for acoustic source localization," in *Berlin Beamforming Conference*, 2021.

Appendix

- to fill if necessary ...